



**Predicting and Exploiting the
Determinants of Protein/Protein Interactions to Identify
Low-Molecular Inhibitors of RUNX1-ETO Tetramerization**

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Alexander Metz
aus Kassel

Düsseldorf, Mai 2014

aus dem Institut für Pharmazeutische und Medizinische Chemie
der Heinrich-Heine Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Holger Gohlke
Korreferent: Prof. Dr. Georg Groth

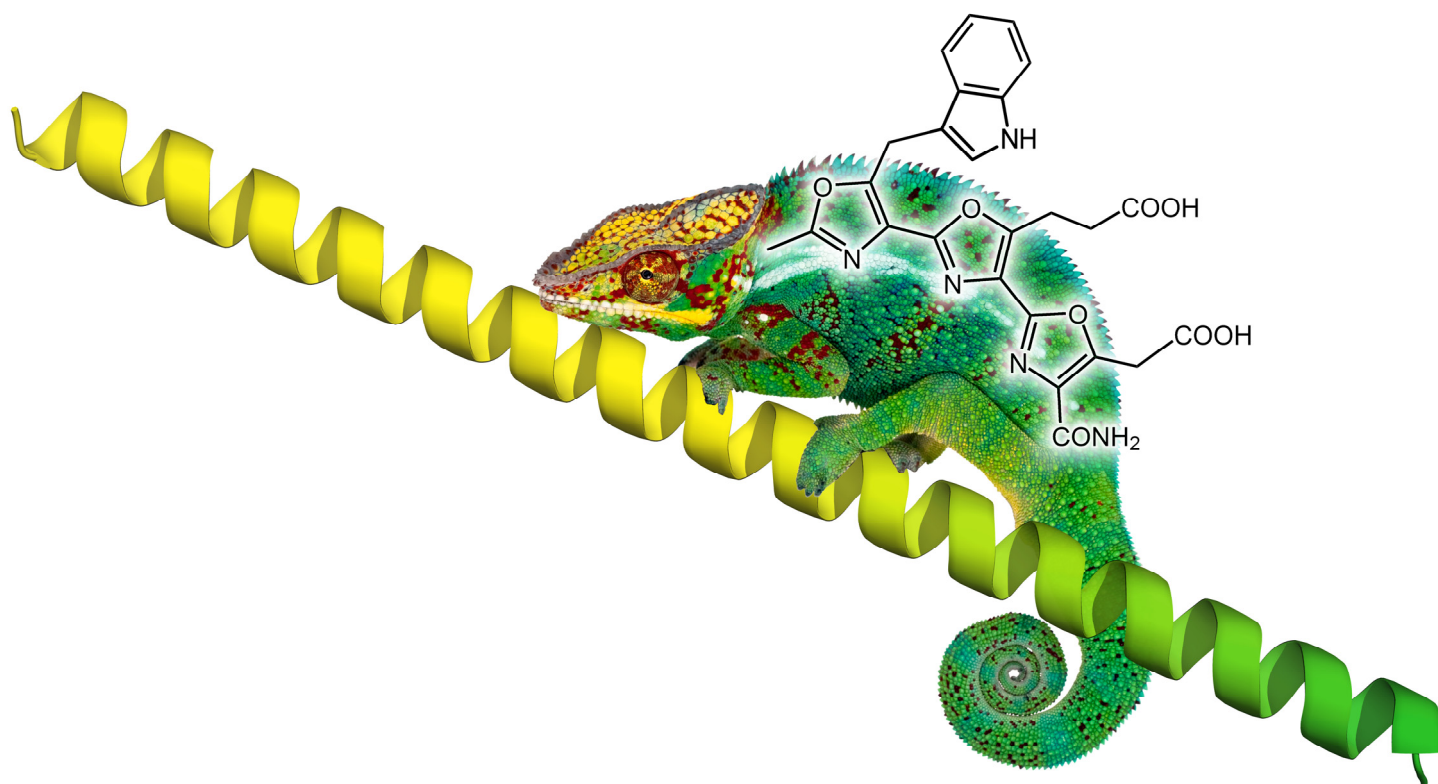
Tag der mündlichen Prüfung:

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Diese Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keinen erfolglosen Promotionsversuch unternommen.

Düsseldorf, den 20.05.2014

(Alexander Metz)



Cover picture: Artwork depicting the mimicry of an α -helix by a teroxazole α -helix mimetic.

The first principle is that you must not fool yourself – and you are the easiest person to fool. So you have to be very careful about that. After you've not fooled yourself, it's easy not to fool other scientists. You just have to be honest in a conventional way after that.

Richard Feynman

... it's like seeing a watch and wondering how actually it works. So in short, what we developed is a way, which requires a computer, to look, to take the structure of a protein and then to eventually understand how exactly it does what it does. Like if you have enzymes that digest food, and the structure exists, you want to understand how this is happening, and then you can use it for example to design drugs or just, like in my case, to satisfy your curiosity.

Arieh Warshel

TABLE OF CONTENTS

LIST OF PUBLICATIONS.....	viii
ABBREVIATIONS	ix
ZUSAMMENFASSUNG	xiii
ABSTRACT	xiv
1 INTRODUCTION AND BACKGROUND	1
1.1 PPIs: PROMISING TARGETS OR TOO MUCH RISK?.....	6
1.2 THE MOLECULAR NATURE OF PPIs.....	9
1.2.1 SURFACE SIZE AND SHAPE	9
1.2.2 ABSENCE OF POCKETS.....	11
1.2.3 STABILITY OF PPIs	15
1.2.4 MECHANISMS OF PPI AND HOT SPOTS	18
1.3 THE PPI OF INTERLEUKIN-2 AND ITS α -RECEPTOR	23
1.4 THE NHR2 OLIGOMERIZATION DOMAIN OF RUNX1-ETO	24
1.5 SELECTED COMPUTATIONAL METHODS	28
1.5.1 HOT SPOT PREDICTION BY THE MM-PBSA METHOD	28
1.5.2 3D SIMILARITY AND SUPERPOSITION BY ROCS.....	29
2 SCOPE OF THE THESIS	31
3 PPIs: RATIONAL APPROACHES REGARDING DRUGGABILITY, DETERMINANTS, AND MODULATORS (Publication I).....	33
3.1 FUNCTIONAL AND STRUCTURAL ASPECTS OF PPIs	33
3.2 DETERMINANTS OF DRUGGABILITY AND MODULATORS OF PPIs	34
3.3 CONCLUSIONS AND SIGNIFICANCE	35
4 INTERLEUKIN-2: HOT SPOTS, TRANSIENT POCKETS, AND PPIM IDENTIFICATION (PUBLICATION II)	36
4.1 A STRUCTURE-BASED STRATEGY TO IDENTIFY PPIMs.....	36
4.2 MM-PBSA FOR THE PREDICTION OF HOT SPOTS	37
4.3 PREDICTION OF TRANSIENT POCKETS	38
4.4 BINDING POSE PREDICTION BY DOCKING.....	39
4.5 MM-PBSA FOR THE RANKING OF BINDING POSES	40
4.6 DOCKING ENRICHMENT IN A LARGE SET OF DECOYS.....	41
4.7 CONCLUSIONS AND SIGNIFICANCE	42
5 TEROXAZOLES: A NEW HYDROPHILIC α-HELIX MIMETIC SCAFFOLD (Publication III)	43
5.1 COMPUTATIONAL INVESTIGATION OF TEROXAZOLES.....	43
5.2 CONCLUSIONS AND SIGNIFICANCE	44
6 RUNX1-ETO: IDENTIFICATION OF THE FIRST INHIBITORS OF NHR2-MEDIATED OLIGOMERIZATION (Publication IV).....	45
6.1 STRUCTURE-BASED IDENTIFICATION OF NHR2 INHIBITORS	45
6.2 CONCLUSIONS AND SIGNIFICANCE	49
7 SUMMARY AND PERSPECTIVES.....	50
8 ACKNOWLEDGMENTS	51

9	APPENDIX	52
9.1	MATHEMATICAL MODEL OF NHR2 EQUILIBRIUM	52
9.2	FIGURE CREATION.....	56
9.3	REPRINT PERMISSIONS FOR PUBLICATIONS	57
10	REFERENCES.....	58
11	PUBLICATIONS.....	97
	PUBLICATION I.....	97
	PUBLICATION II.....	116
	PUBLICATION II – SUPPORTING INFORMATION.....	131
	PUBLICATION III.....	146
	PUBLICATION III – SUPPORTING INFORMATION.....	155
	PUBLICATION IV.....	223
	PUBLICATION IV – SUPPORTING INFORMATION.....	230
12	Curriculum Vitae.....	245

LIST OF PUBLICATIONS

(Contribution in parentheses)

- I. Alexander Metz (37.5%),**[§] Emanuele Ciglia,[§] and Holger Gohlke.
Modulating Protein-Protein Interactions: From Structural Determinants of Binding to Druggability Prediction to Application.^[1]
Current Pharmaceutical Design, **2012**; 18: 4630-4647.
Impact factor reported for 2011: 3.870
- II. Alexander Metz (35.0%),**[§] Christopher Pfleger,[§] Hannes Kopitz, Stefania Pfeiffer-Marek, Karl-Heinz Baringhaus, and Holger Gohlke.
Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein-Protein Interface.^[2]
Journal of Chemical Information and Modeling, **2012**; 52: 120-133.
Impact factor reported for 2011: 4.675
- III. Cristiano Pinto Gomes,**[§] **Alexander Metz (32.5%),**[§] Jan W. Bats, Holger Gohlke, and Michael W. Göbel.
Modular Solid-Phase Synthesis of Teroxazoles as a Class of α -Helix Mimetics.^[3]
European Journal of Organic Chemistry, **2012**; 17: 3270-3277.
Impact factor reported for 2011: 3.329
- IV. Alexander Metz (30.0%),**[§] Julia Schanda,[§] Manuel Grez,⁺⁺ Christian Wichmann,⁺⁺ and Holger Gohlke.⁺⁺
From Determinants of RUNX1/ETO Tetramerization to Small-Molecule Protein-Protein Interaction Inhibitors Targeting Acute Myeloid Leukemia.^[4]
Journal of Chemical Information and Modeling **2013**, 53: 2197-2202.
Impact factor reported for 2012: 4.304
The here described “Inhibitors of NHR2 and/or RUNX1/ETO-tetramerization” have been filed for patenting at the European Patent Office (EP Appl.: 13165993.0).^[5]

[§] Both authors have contributed equally to the respective work.

⁺⁺ These authors share senior authorship.

ABBREVIATIONS

ΔG	change in (Gibbs) free energy
Å	Ångström ($1 \text{ Å} = 10^{-10} \text{ m}$)
ABCD assay	avidin–biotin complex DNA assay
AEL	acute erythroid leukemia
AML	acute myeloid leukemia
AML1	acute myeloid leukemia 1 protein (<i>also</i> CBFA2, RUNX1)
AML1-ETO	AML1-ETO fusion protein (<i>also</i> AML1-MTG8, RUNX1-ETO, RUNX1-RUNX1T1)
ATP	adenosine triphosphate
AUC	area under curve
BAD	Bcl-2-associated death promoter protein
BAK	Bcl-2 homologous antagonist/killer protein
BAX	Bcl-2-associated X protein
Bcl-2	B-cell lymphoma 2 protein
Bcl-X _L	B-cell lymphoma-extra large protein
BH3	Bcl-2 homology domain 3
BCR	breakpoint cluster region protein
BCR-ABL	BCR-ABL fusion protein (ABL: Abelson murine leukemia viral oncogene)
BIM	Bcl-2 interacting mediator of cell death
BSE	bovine spongiform encephalopathy
CADD	computer-aided drug design
CBF α 2	core-binding factor subunit α 2 protein
CD4	cluster of differentiation 4 glycoprotein
CDK	cyclin-dependent kinase
CBF	core-binding factor
CBFA2	core-binding factor subunit α 2 (<i>also</i> AML1, RUNX1)
CBFA2T1	CBFA2; translocated to, 1 ^[a] (<i>also</i> ETO or MTG8)
CBFA2T3	CBFA2; translocated to, 3 ^[a] (<i>also</i> ETO2, MTG16, MTGR2)
CBFA2T3-GLIS2	CBFA2T3-GLIS2 fusion protein; caused by inv(16)
CBF β	core-binding factor subunit β
CBP	CREB (cAMP response element-binding protein)-binding protein
CJD	Creutzfeldt-Jakob disease
CML	chronic myeloid leukemia

^[a] CBFA2; translocated to, 1 is a technical notation for *protein 1 translocated to CBFA2*, and accordingly.

CN-AML	cytogenetically normal AML
c-Jun	cellular Jun (Ju-nanna) homolog
c-Kit	cellular v-kit homolog (<i>also</i> Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog)
c-Myc	cellular myelocytomatosis protein
c-Raf	cellular rapidly accelerated fibrosarcoma protein
DLBCL	diffuse large B-cell lymphoma
DUD	directory of useful decoys
E2A	E-box binding protein 2 α
EC_{50}	half maximal effective concentration
EF	enrichment factor
ELISA	enzyme-linked immunosorbent assay
E protein	enhancer-box binding protein
ERK	extracellular-signal regulated kinase
ETO	eight-twenty one protein (<i>also</i> RUNX1T1, CBFA2T1)
ExoS	exoenzyme S protein
FAB	French–American–British classification system of acute leukemia
FDA	Food and Drug Administration
FKBP	FK506 binding protein
FRH	5-[2,3-Dichloro-4-(5-{1-[2-(2-guanidino-4-methyl-pentanoylamino)-acetyl]-piperidin-4-yl}-1-methyl-1 <i>H</i> -pyrazol-3-yl)-phoxymethyl]-furan-2-carboxylic acid
FRODA	framework rigidity optimized dynamics algorithm
FtsZ	filamenting temperature-sensitive mutant Z protein
GB	Generalized Born
GLIS2	glioma-associated oncogene (GLI)-similar protein 2
gp120	envelope glycoprotein GP120
G protein	guanine nucleotide-binding protein
GCPR	G protein-coupled receptor
HDAC	histone deacetylase
HDM2	human double minute 2 protein
HEB	human E-box-binding protein
HSP90	heat shock protein 90
IC_{50}	half maximal inhibitory concentration
<i>i.a.</i>	<i>inter alia</i> ; Latin for “among other things”
IGH	immune-globulin heavy chain
IGH-CBFA2T3	IGH-CBFA2T3 gene fusion; caused by t(14;16)
IL-2	interleukin-2
IL-2R α	interleukin-2 receptor subunit α

JAK	Janus kinase
JNK	c-Jun <i>N</i> -terminal kinase
kcal	1 kcal \approx 4.18 kJ
K_d	dissociation constant
$\log P$	logarithm of the octanol-water partition coefficient
LDK cells	Lineage ⁻ /Sca1 ⁺ /c-Kit ⁺ cells
MAPK	mitogen-activated protein kinases
Max	Myc-associated factor X protein
MD	molecular dynamics
MM	molecular mechanics
MM-PBSA	molecular mechanics Poisson-Boltzmann surface area
MM-GBSA	molecular mechanics Generalized Born surface area
mSIN3A	mammalian suppressor interacting 3 homolog A of transcriptional repressor Sin3
MTG8	myeloid transforming gene on chromosome 8 (<i>also</i> ETO, RUNX1T1)
MTG16	myeloid transforming gene on chromosome 16
MTGR1	myeloid translocation gene-related protein 1
MW	molecular weight (incorrectly used for <i>molecular mass</i>)
N-CoR	nuclear receptor co-repressor
NFIA	nuclear factor 1 A-type (a dimeric DNA-binding transcription factor)
NFIA-CBFA2T3	NFIA-CBFA2T3 fusion protein; caused by t(1;16)
NHR2	nervy homology region 2
NHR2 hot spots	hot spots of the tetramerization of NHR2 from dimers
NME	new molecular entity
NMR	nuclear magnetic resonance
non-DS-AMKL	non-Down syndrome pediatric acute megakaryoblastic leukemia
p300	co-activator protein p300
p53	tumor suppressor p53 protein
PIAS	protein inhibitor of activated STATs
PB	Poisson-Boltzmann
P/CAF	p300/CBP-associated factor
PDB	Protein Data Bank
PEBP2 α B	polyomavirus enhancer binding protein 2 α B protein
PKA	protein kinase A
PPIN	protein/protein interaction network
PPI	protein/protein interaction
PPI complex	protein/protein complex

PPI structure	protein/protein complex structure
PPIface	protein/protein interface
PPII	low-molecular protein/protein interaction inhibitor
PPIM	low-molecular protein/protein interaction modulator
PU.1 promoter	purine-rich box-1 promoter
R&D	research and development
Ras	rat sarcoma protein
Raf	rapidly accelerated fibrosarcoma protein
RE	RUNX1-ETO
RE-AML	RUNX1-ETO dependent AML
REtr	truncated RE lacking the NHR3 and NHR4 domains
REtr-m5	fivefold hot spot alanine mutant of REtr
residue	amino acid residue
<i>RMSD</i>	root mean square deviation
ROC	receiver operating characteristic
ROCS	rapid overlay of chemical structures
Runt	Runt domain of RUNX proteins (named after a mutant phenotype of <i>D. melanogaster</i>)
RUNX1	Runt-related transcription factor 1 (also AML1, CBF α 2, or PEBP2 α B)
RUNX1-ETO	RUNX1-ETO fusion protein; caused by t(8;21)
RUNX1-ETO2	RUNX1-ETO2 fusion protein; caused by t(16;21)
RUNX1T1	Runt-related transcription factor 1; translocated to, 1 ^[a] (also ETO, MTG8, CBFA2T1, AML1T1, CDR, MGC2796, MTG8b, or ZMYND2)
SA	surface area
SAR	structure-activity relationship
SBDD	structure-based drug design
SMARTS	SMILES arbitrary target specification
SMILES	simplified molecular-input line-entry system
SMMHC	smooth muscle myosin heavy chain protein
SON	DNA-binding protein SON
STAT	signal transducers and activators of transcription
TAT pathway	twin-arginine translocation protein export pathway
TAT-NHR2	trans-activator of transcription-NHR2 fusion peptide
teroxazoles	derivatives of the teroxazole scaffold (2'',5,5',5''-Tetrasubstituted-[2,4':2',4''-teroxazole]-4-carboxamide)
TNF α	tumor necrosis factor
vdW	van-der-Waals
VS	virtual screening
ZipA	Z-ring interacting protein A

ZUSAMMENFASSUNG

Protein/Protein Interaktionen (PPIs) sind essenziell für alle biologischen Prozesse. Durch ihre Funktion als zentrale Schalt- und Kontrollelemente ist die Fehlfunktion von PPIs eine der Hauptursachen von Krebs und anderen Erkrankungen. Daher ist ein zielgerichteter Eingriff in PPIs, speziell mit kleinen, für die pharmakologische Therapie geeigneten Molekülen, das Ziel intensiver Forschung.

Diese Forschung zu PPIs führte zu einer Reihe *niedermolekularer Protein/Protein Interaktions-Modulatoren* (PPIMs). Doch nur wenige PPIMs wurden mittels rationaler, strukturbasierter Ansätze entdeckt. Dies liegt an Schwierigkeiten bei der direkten Wiederverwendung computergestützter Methoden, die sich zur Identifizierung von Liganden konventioneller Targets bewährt haben.

Naturgemäß binden konventionelle Targets kleine Moleküle in ausgeprägten komplementären Bindetaschen. Dagegen sind *Protein/Protein Interaktionsflächen* (PPIfaces) oft groß und frei von markanten Taschen. Daher ist es schwer, kleine Moleküle mit adäquater Affinität und Spezifität zu finden, um eines der beiden Proteine zu verdrängen. Die vermehrte Entdeckung (wirkstoffähnlicher) PPIMs zeigt jedoch: diese Aufgaben sind lösbar, und PPIs sind nicht generell pharmakologisch unangreifbar.

Ziel dieser Arbeit war die Entwicklung einer computergestützten Strategie zur rationalen Identifizierung von PPIMs, speziell von NHR2-Inhibitoren, nur auf Grundlage der *Protein/Protein Komplex-Struktur* (PPI-Struktur). Diese Aufgabe umfasst die vier Schwerpunkte meiner Dissertation.

Zuerst überprüfte ich den Wissenstand zu PPIs, zu PPIMs, zu bestimmenden Faktoren ihrer Interaktionen als auch zur computergestützten Identifizierung von Wirkstoff-Bindestellen und PPIMs. **Als zweites** entwickelte ich eine Strategie, die mit Hilfe von aus PPI-Strukturen ermittelten Hot Spots und transienten Taschen ein struktur-basiertes virtuelles Screening leitet. Die Wiederauffindung bekannter Interleukin-2 Inhibitoren aus einer Vielzahl inaktiver Substanzen validierte diese Strategie. **Als drittes** analysierte ich die Eignung von Teroxazolen als neue Klasse hydrophiler α -Helix-Mimetika. Diese präsentieren Reste in ähnlicher räumlicher Anordnung wie eine α -Helix, allerdings in einer neuartigen Sequenzabfolge. **Als viertes** identifizierte ich die ersten mikromolaren Inhibitoren der NHR2-vermittelten Tetramerisierung von *RUNXI-ETO* (RE) aus Dimeren, einer Bedingung für Ausbruch und Fortbestand der *RE-abhängigen akuten myeloischen Leukämie* (RE-AML). Vorhergesagte Hot Spots nahe der größten Tasche im PPIface leiteten die computergestützte Identifizierung wirkstoffartiger PPIMs. Diese imitieren Hot Spots, zielen auf das PPIface und inhibieren die Dimer-Assoziation sowie die Proliferation RE-abhängiger Zellen.

Als Werkzeug zur Untersuchung von Auswirkungen der NHR2-Tetramerisierung sind diese PPIMs ein wichtiger Schritt in Richtung einer personalisierten RE-AML Therapie. Vor allem kann die vorgestellte Strategie als erster Schritt für vergleichbare Vorhaben zur Identifizierung oder zum Design von PPIMs dienen, selbst wenn nur eine PPI-Struktur mit eher flachem PPIface bekannt ist.

ABSTRACT

Protein/protein interactions (PPIs) are essential for all biological processes. Their role as central control switches and checkpoints in signaling and regulation makes their dysfunction an eminent cause of cancer and other diseases. Thus, a target-oriented intervention with PPIs, in particular by small molecules suitable for a pharmacological therapy, is the object of intense research.

In the recent years, this research of PPIs yielded a number of *low-molecular protein/protein interaction modulators* (PPIMs). However, only a few PPIMs were identified by rational or structure-based considerations. The reason is that it is difficult to directly re-use well-established computational methods to identify ligands of conventional targets, such as enzymes, receptors, or transporters.

Conventional targets naturally bind small molecules *via* pronounced complementary binding pockets. In contrast, many *protein/protein interfaces* (PPIfaces) are large and lack pronounced pockets. Thus, it is hard to find small molecules with an affinity and specificity that is adequate to displace one of the binding proteins. Nevertheless, the widespread discovery of (drug-like) PPIMs shows: these challenges can be overcome, and PPIs are not undruggable in general.

The goal of this thesis was to develop a computational strategy for the rational identification of PPIMs, notably of NHR2 inhibitors, starting only with a *protein/protein complex structure* (PPI structure). This task subdivides into the four core themes of this thesis.

First, I reviewed the knowledge about PPIs, PPIMs, the determinants of their interactions as well as computational methods for the identification of druggable sites and PPIMs. **Second**, I implemented a strategy that uses a PPI-structure based prediction of hot spots and transient pockets to guide a structure-based virtual screening. As a validation, I retrieved known PPIMs that bind to the PPIface of interleukin-2 from a large set of non-binders. **Third**, I analyzed the potential of teroxazoles as a new class of hydrophilic α -helix mimetics. These present side chains similar to α -helices and mimic a sequence pattern that has not yet been considered. **Fourth**, I identified the first micromolar inhibitors of the NHR2-mediated tetramerization of *RUNX1-ETO* (RE) from dimers, a prerequisite for the onset and maintenance of *RE-dependent acute myeloid leukemia* (RE-AML). Predicted tetramerization hot spots close to the largest pocket in the PPIface guided the computational identification of drug-like PPIMs. These PPIMs mimic NHR2 hot spots, aim at the PPIface, and inhibit dimer association as well as the proliferation of RE-dependent cells.

These PPIMs are valuable as probes and tools to study the effects of NHR2 tetramerization and are an important step towards a personalized therapy of RE-AML. Most importantly, however, the presented strategy can well be the first step in any comparable structure-based endeavor to identify or design PPIMs, even in cases where only a PPI structure with a rather flat PPIface is known.

1 INTRODUCTION AND BACKGROUND

The interaction of molecules is fundamental for all chemical and biological processes. In fact, living cells are densely packed with an immense number of molecules of a large variety of types.^[6] However, their interactions are tightly controlled to maintain the highly ordered and extremely complex system that a living cell is.^[7,8] That means, for any time and state, certain interactions occur due to the *affinity* between the binding molecules, which is absent between non-binding molecules, hence, causing *specificity*. Evolution has resulted in different mechanisms that ensure binding affinity and specificity and enable molecular recognition.

The most familiar of these mechanisms is the binding of a low-molecular *ligand* into the *pronounced* and *complementary* binding pocket of a protein. The underlying *lock-and-key model* was already suggested by Emil Fischer.^[9,10] Here, the interface area and, hence, the number and strength of physicochemical interactions is increased by enclosing the ligand in an accurately fitting cavity (Figure 1). Exclusively ligands whose shape and physicochemical properties are complementary to those of the cavity can bind with high affinity. Notably, the majority of enzymes, receptors, transporters, and ligand-gated ion channels^[11] specifically recognize and bind their low-molecular substrates and effectors by this mechanism.

Thus, binding to these *conventional targets* is also the point of intervention used by most low-molecular bioactive molecules, both naturally occurring effectors and xenobiotics, such as *pharmaceutical drugs* (hereinafter "drugs"). Ultimately, binding to a target is the premise for any drug to work as subsumed by the expression *Corpora non agunt nisi fixata* (drugs will not work unless they are bound) coined by Paul Ehrlich.^[13] In addition, small molecules are particularly suited as drugs, primarily due to their often favorable bioavailability.

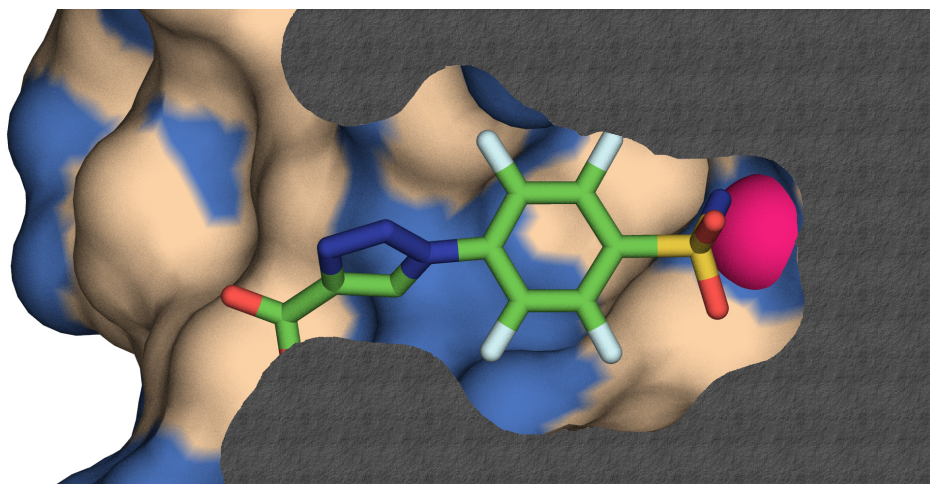


Figure 1. Cross section of the binding pocket of carbonic anhydrase (PDB code: 3P25^[12]).

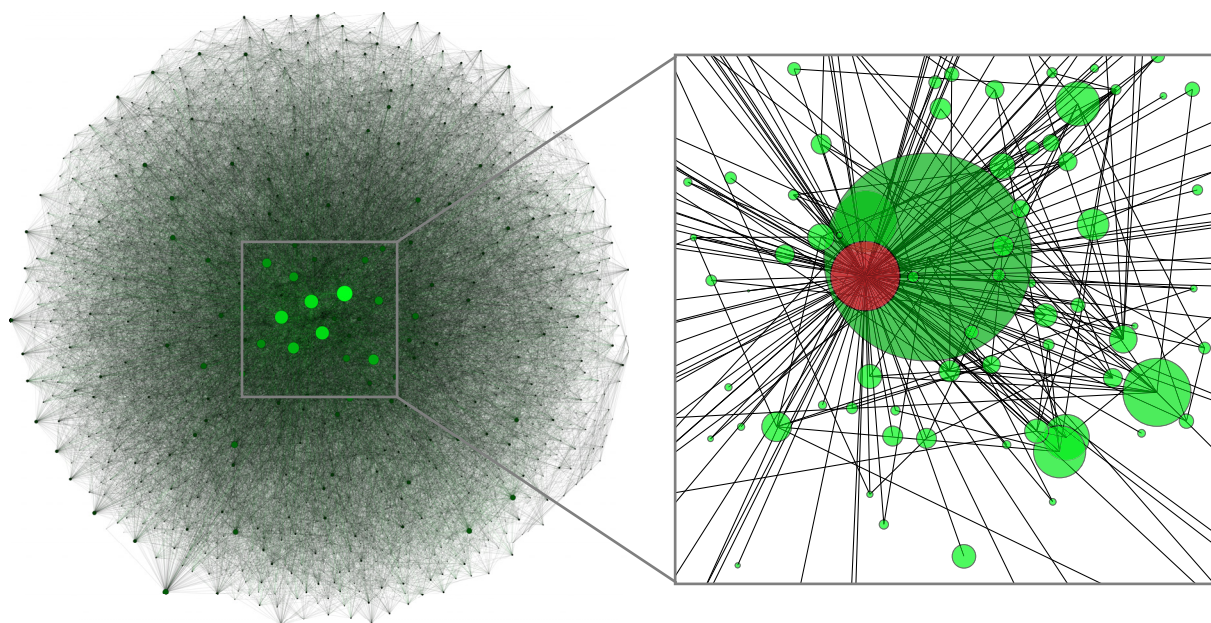


Figure 2. The human protein/protein interaction network. The left side shows the experimentally determined human PPIN with 9,376 proteins (nodes) and 31,021 PPIs (edges).^[14] The right side shows 133 direct interactions to p53 (red), indicating its role as a central control switch. The node size indicates the number of PPIs.

However, the number of conventional drug targets is limited.^[15] An estimated 7% of the protein coding human genes (1,620 out of 22,218) have explicit disease associations with only 105 used protein targets.^[16] The reported total number of drug targets varies roughly between 200 and 400,^[16-18] slightly exceeding 2,000 when including also experimental drug targets.^[19,20] On the drugs' side, the DrugBank data base lists a total of 7,681, amongst these 1,545 approved low-molecular drugs (as of May 2014).^[21]

Even when assuming that only a much lower proportion of the estimated 100,000 to 650,000 human *protein/protein interactions*^[22-24] (PPIs; Figure 2) are druggable,^[25] then the potential number of additional PPI drug targets would be in the thousands.^[22,26,27] In brief, it would be a waste to neglect the modulation of PPIs as a potentially valid therapeutic approach.^[28,29] This is all the more true in the light of stagnant drug development pipelines^[30-33] and the so-called “drug drought”.^[34-36]

Nevertheless, the development of a new drug is risky and costly. The reported R&D costs are often reported to have surpassed \$US 800 million^[35,37-39] reaching \$US 4 – 12 billion^[40] for individual marketed new molecular entities (NME), overall about \$US 50 billion per year (Figure 3).^[16,41] The bulk of these costs arises from drug candidates that fail in the later clinical stages of development. Only about one-fifth of the drugs that enter clinical trials ultimately get approved by the *Food and Drug Administration* (FDA).^[37] Nevertheless, still about

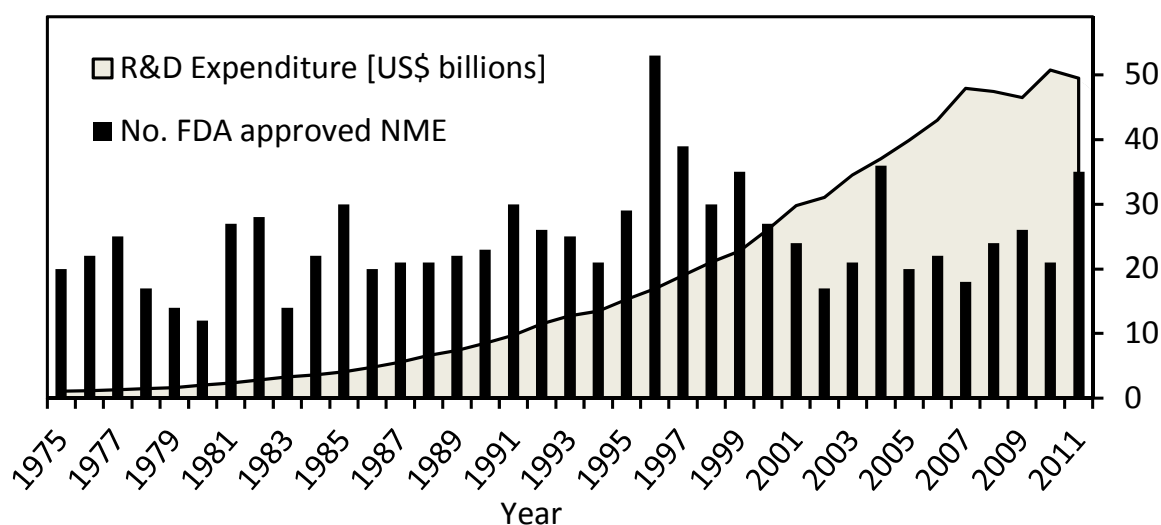


Figure 3. R&D expenditure of the PHARMACEUTICAL RESEARCH AND MANUFACTURERS OF AMERICA (PhRMA) companies and the number of NMEs, including biologics, approved by the FDA.^[41]

30% of the R&D costs are incurred even before entering clinical trials.^[37] Accordingly, it is important to select the most adequate targets early on. Hence PPIs, which have commonly been considered high risk targets or even undruggable,^[15] were for a long time neglected by the pharmaceutical industry.^[42,43] To access PPIs as drug targets, one needs reliable rational methods to predict which PPIs are druggable,^[42,44,45] whose modulation is pharmaceutically meaningful,^[18,46,47] and, of course, appropriate drugs to address these targets.^[48]

With this goal in mind, I introduce and discuss the aspects relevant for my endeavor to predict the determinants of PPIs in the context of the current state of the science. In particular, I focus on determinants of the tetramerization of the *nervy homology region 2* (NHR2) protein and how to use their prediction to identify *low-molecular protein/protein interaction modulators* or, more particularly, *inhibitors* (PPIMs or PPIIs; unless stated otherwise meaning low-molecular molecules). After a general explanation of why PPIs are worthwhile and promising targets (**Chapter 1.1**), I discuss the molecular nature and determinants of PPIs and why it is a challenging task to identify PPIMs that interfere with PPIs (**Chapter 1.2**). In spite of these challenges, numerous PPIMs (> 1,650^[49,50]) and PPI crystal structures (> 50^[51]) have been identified to date. Subsequently, I will discuss conclusions to be drawn from known PPIMs regarding the rational identification of new PPIMs and to assess the prospects for a PPIM-based pharmacological therapy. Finally, I introduce the two particular PPIs that I used as test cases in my studies: the well-investigated interaction between *interleukin-2* (IL-2) with the *IL-2 receptor subunit α* (IL-2R α),^[52] for which PPIMs are known (**Chapter 1.3**), and the

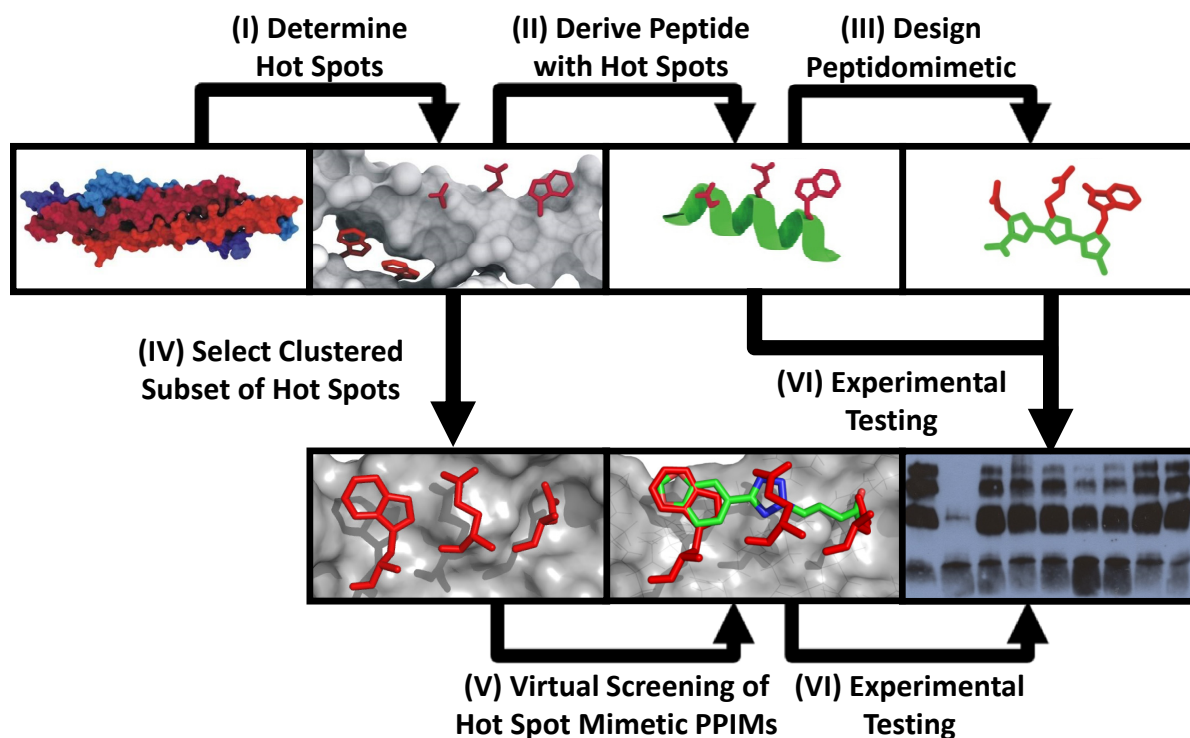


Figure 4. Strategy to identify PPIMs starting from a PPI structure. Successively, we (I) determine hot spots^[62] of the protein/protein association, (II) derive an inhibitory peptide that contains these hot spots, (III) design peptidomimetic PPIMs decorated with the essential functional groups of the hot spots, and (IV) use an adequate, clustered subset of these hot spots for (V) a virtual screening of conveniently available PPIMs that mimic these hot spots. (VI) In every step, we use biochemical and cellular experiments, in conjunction with hot spot alanine mutants, to validate the target and confirm the predicted hot spots and PPIMs. The illustrations depict the NHR2 tetramer^[63] as an example (**Chapter 1.4**).

NHR2 oligomerization domain of the RUNX1-ETO (RE) oncofusion protein, a new target in AML (**Chapter 1.4**).

Although I focus on the computational prediction of the determinants of PPIs with the goal to identify PPIMs, such predictions need prior knowledge and rigorous experimental verification. This is all the more true because PPIs are less established than conventional targets. The relevant experimental methods were reviewed exhaustively.^[45,53-56]

Turning to computational methods, various of these can help investigate and identify PPIs and PPIMs, *e.g.* *virtual screening* (VS)^[57-59] and *docking*.^[60,61] These and other computational methods were reviewed in **Publication I**. Nevertheless, I introduce the two methods MM-PBSA (**Chapter 1.5.1**) and ROCS (**Chapter 1.5.2**) in more detail, because they were crucial for my hot spot predictions and the structure-based identification of PPIMs.

After reasoning the scientific framework, I describe the key objectives and the thematic relation between the individual tasks that had to be accomplished (**Chapter 2**). For that, I detail how I addressed these tasks in the four publications compiled in this thesis (**Pubs. I to IV**)^[1-4] and summarize my findings (**Chapter 3 to 6**).

In essence, I reviewed the state-of-the-art in computational PPI analysis and PPIM identification and presented exemplary test cases for rational PPIM identification (**Chapter 3**).^[1] Based on this knowledge, I devised a new strategy for identifying PPIMs (Figure 4). To this end, I combined computational predictions and experimental results^[a] to identify PPI *hot spots*, *i.e.* those amino acids that form the core interactions of a PPI and account for the bulk of the binding free energy. Preferably, a druggable pocket is located, or can form, close to these hot spots. Hot spots and optionally pockets then guide a VS. Retrieving known PPIMs of IL-2 from a large number of non-binders demonstrated the reliability and performance of this strategy (**Chapter 4**).^[2]

Finally, I used predicted hot spots of the NHR2 tetramerization to guide the identification of PPIMs that target the *protein/protein interface* (PPIface) between the NHR2 dimers. Two routes were pursued simultaneously in order to identify such PPIMs. First, I analyzed the potential of teroxazoles as α -helix mimetics.^[a] These privileged scaffolds project side chains with launch vectors similar to an α -helix and, thus, allow them to mimic hot spots (**Chapter 5**).^[3] In addition, I identified conveniently available mimetics of the NHR2 hot spots by VS (**Chapter 6**).^[4] This led to the first drug-like PPIMs of NHR2 tetramerization, which prevent the proliferation of RE-dependent cells. More importantly, my strategy needs only a *protein/protein complex structure* (PPI structure) and can thus be the first step in any comparable structure-based endeavor to identify PPIMs.

^[a] All chemical syntheses as well as biochemical and cell biological experiments were performed by Cristiano Pinto Gomes, Jan. W. Bats, and Michael W. Göbel at the INSTITUTE OF ORGANIC CHEMISTRY AND CHEMICAL BIOLOGY of the GOETHE-UNIVERSITY (Frankfurt)^[3] and Julia Schanda, Christian Wichmann, and Manuel Grez at the INSTITUTE FOR BIOMEDICAL RESEARCH GEORG-SPEYER-HAUS (Frankfurt),^[4] respectively.

1.1 PPIs: Promising Targets or Too Much Risk?

PPIs are essential in all biological processes. In particular, PPIs play important roles as control switches and checkpoints in signaling and regulation.^[64] Some proteins are “hubs” that interact with dozens or hundreds of other proteins and/or are bottlenecks of important pathways.^[65,66] Thus, the dys- or malfunction of a few or even a single PPI^[67] and the dysregulated expression or aberrant localization of the involved proteins can have severe, if not fatal, consequences.^[68]

It was suggested that the size of an organism’s protein/protein interaction network (PPIN)^[22-24] may be a better measure for its apparent biological complexity^[22] than its genome size or number of proteins.^[69-71] For example, the number of roughly 20,000 protein coding human genes^[72] is only about equal or even lower than that of some plants, insects, and protozoa.^[69] In contrast, the estimated number of human PPIs (100,000 to 650,000)^[22-24] exceeds that of allegedly simpler organisms, such as *D. melanogaster* and *C. elegans*.^[22] More complex organisms also tend to have a higher proportion of multidomain proteins.^[73,74] By bringing together multiple domains with different functions in one protein and, in turn, multiple such proteins *via* PPIs, the grade of organization and complexity is greatly increased. Examples include complex signaling pathways, assembly lines, molecular machines, and cell adhesion. Ultimately, the NHR2 oligomerization domain described later also acts as an adapter module of the RE multidomain protein, which recruits other proteins that contain NHR2 and/or other adapter domains,^[75] such as transcription factors (**Chapter 1.4**).^[76,77]

Of course, there are additional mechanisms that add to this complexity, such as epigenetic regulation,^[78,79] differential expression, alternative splicing, somatic hyper mutation, posttranslational modification, degradation, sequestration,^[75,80,81] subcellular location,^[75,82,83] compartmentalization, molecular crowding,^[84] folding, and regulation by low-molecular effectors or by physicochemical changes in the proteins’ environment. All of these mechanisms are, in the end, mutually dependent on PPIs.

In this light, it is obvious that modulating aberrant or regular PPIs allows influencing complex biological PPINs.^[47] Although inevitably incomplete, the following list of disease-relevant PPIs may convey an idea of the wide range of potential applications for PPIMs.

Illustrative examples in which dysfunctional or dysregulated PPIs cause a disease^[68,85,86] are: (i) the mutation or overexpression of oncoproteins that are involved in PPIs that

contribute to causing cancer,^[87-89] such as Ras/Raf,^[90-93] c-Myc/Max,^[29,94-97] ERK,^[98] or other members of the MAPK/ERK pathway,^[99] the sequestration of tumor suppressor p53 by HDM2,^[28,100-103] antiapoptotic Bcl-2 proteins that act by binding proapoptotic proteins, such as BAK or BAX,^[28,104] or the oligomerization of multidomain fusion proteins,^[75] such as BCR-ABL and RE; (ii) protein aggregation, such as with amyloid- β or α -synuclein proteins that cause amyloidoses and neurodegenerative diseases (*e.g.* Alzheimer's or Parkinson's diseases)^[105,106] or as with hemoglobin in sickle cell anemia;^[107] (iii) cell-surface receptors that bind proteins or peptides, such as excessive human growth hormone in acromegaly, integrin/fibrinogen binding in coagulation and infarct, or as with hyperinsulinemic hypoglycemia; (iv) G protein-coupled receptors (GPCRs) that release an intracellular G protein upon binding an allosteric extracellular ligand;^[108] or (v) enzymes with protein or peptide substrates, such as proteases (*e.g.* caspases in apoptosis, blood clotting factor VIII, factor Xa, thrombin, and plasmin in hemostasis, and the renin-angiotensin system), protein kinases^[109] (*e.g.* tyrosine kinase receptors, MAP kinases, and cyclin-dependent kinases (CDKs) in oncogenesis), or histone acetyltransferases and deacetylases in chromatin remodeling.^[110,111]

In addition, PPIs are by no means restricted to interactions within a single organism or species. For example, some viruses corrupt or assume control of essential parts of their hosts' complex PPINs^[112-118] despite possessing only a few proteins themselves.^[28,65,119] In the most general sense, microorganisms and viruses cause or avoid all kinds of PPI moderated effects, such as immune response, infection, cellular internalization, and the action or modulation of enzymes with protein substrates.^[120] Finally, prions, like autogenic misfolded proteins, can transmit their misfolded state and cause diseases, such as bovine spongiform encephalopathy (BSE), scrapie, kuru, or Creutzfeldt-Jakob disease (CJD).^[121]

In summary, PPIs are essential for many diseases and indeed promising targets. Thus, it is desirable to modulate PPIs by a pharmaceutical therapy. Not without reason, inhibiting PPIs by antibodies is also an important aspect of how vertebrate immune systems fight diseases. Indeed, antibodies are currently the most successful class of PPI inhibiting drugs.^[122-124] They are conveniently accessible high-affinity PPIs and can also help to reveal druggable epitopes, understand binding mechanisms,^[125] and even inspire PPIM design^[126] by complementing insights from non-antibody PPIs.^[127] Antibodies, however, usually lack oral bioavailability and cell permeability and are, thus, mostly suited for extracellular targets.^[128]

The situation is further complicated because PPIs and the resulting complex PPINs^[112-115] (Figure 2) are, despite increased efforts, still far from being thoroughly understood, especially

when interactions with non-protein molecules are considered. Therefore, it is not obvious which PPIs may serve as drug targets.^[42,44,47,129] It is not sufficient that a particular PPI is an essential cause for a disease. Rather, one of the proteins that is involved in the PPI must be able to bind a drug-like PPIM molecule^[43,48] that has adequate pharmacokinetic properties,^[48,130-132] specificity, and causes a pharmacologically meaningful effect.^[18,46,47]

The identification of drug-like PPIMs is usually considered difficult. This is primarily because of the size of their PPIface, the absence of pronounced, *i.e.* large, deep, and buried, binding pockets,^[133] and the high stability of PPIs.^[43] Encouragingly, the widespread identification of PPIMs shows that this challenge can be overcome.^[129,134,135] Although the majority of these PPIMs are far from being ready-to-use drugs and non drug-like in Lipinski's sense,^[48,131] some of these PPIMs are bioavailable and others could be made so by new drug delivery^[136] and prodrug strategies.^[137] The disregard of PPIMs as drugs and the fact that most PPIMs originate from screening libraries that are assorted for conventional targets^[26,138] suggest that there is considerable space for optimizing their pharmacokinetics and pharmacodynamics. Indeed, the small but growing number of PPIMs that have entered clinical trials^[135,139-145] may still just be limited by the sparse efforts directed at modulating PPIs.^[146]

Nevertheless, also non-drug PPIMs are valuable as versatile probes or tools^[147] in chemical biology,^[148-150] assay development, as diagnostic agents, and to study effects of a graded PPI-modulation *in vivo* and *in vitro*. In this way, PPIMs can improve our understanding of PPIs, PPINs, and the underlying molecular processes. A fundamental understanding of the determinants and structure-activity relationships (SARs) of PPIs and PPIMs would allow a rational approach towards designed PPIs,^[151] personalized^[115,152,153] or multicomponent therapies,^[154,155] polypharmacology,^[16,47,156-160] avoiding adverse effects,^[161] identifying and understanding biologically active agents such as drugs, biocides, or toxins, as well as even more remote applications, such as biomaterials,^[162,163] biointerfaces,^[164-166] or biosensors.^[167]

However, the important question that remains is: Which particular disease-relevant PPIs are druggable? Undoubtedly, despite all promises, PPIs pose an increased risk of failure in drug development,^[28] as I discuss in **Chapter 1.2**. Nevertheless, the right solution is certainly not to neglect PPIs as drug targets. Instead, this risk of failure can be reduced by: (i) adapting rational methods of molecular modeling and *computer-aided drug design* (CADD) for the characteristics of PPIs and PPIMs in order to identify (ii) druggable PPIs, (iii) druggable sites in PPIfaces, and (iv) drug-like PPIMs (v) while concomitantly validating each prediction experimentally.

1.2 The Molecular Nature of PPIs

Four reasons usually given for why PPIs are hard and risky drug targets are that PPIfaces are large, they are flat, they lack pronounced binding pockets, and PPIs are stable. To partially disprove these points, I will address them in the given order, compare PPIfaces with binding sites of conventional targets, and point out the consequences for drug design at PPIfaces. In addition, I describe recurring characteristics of PPIs that may be exploited to identify PPIMs.

1.2.1 Surface Size and Shape

Surface size and shape pose a challenge for the modulation of PPIs by PPIMs. With respect to surface size, PPIfaces are usually larger than the interface between a conventional pharmacological receptor protein (hereinafter "receptor") and its low-molecular ligand (Figure 5). Protein/ligand interface areas are typically about 300 to 1,000 Å² in size,^[168-170] while PPIface areas usually range from ca. 1,500 to 3,000 Å² and can even be larger.^[51,171-174]

With respect to surface shape, PPIfaces usually lack the pronounced binding pockets of conventional receptors. Most conventional receptors naturally bind small molecules. Therefore, binding sites evolved in a way that enables binding with adequate affinity and specificity.^[170,177] Pockets are perfectly suited for this task for the following four reasons: (i) Surrounding a low-molecular ligand from all sides enlarges the interface area and, thus, the number of interactions that contribute to the binding free energy. (ii) The complementarity of these interactions, but also of the size and shape of the binding pockets and their ligands, prevents undesired and non-complementary ligands from binding and, thus, causes specificity. (iii) Excluding solvent molecules from the interface upon binding^[62] strengthens

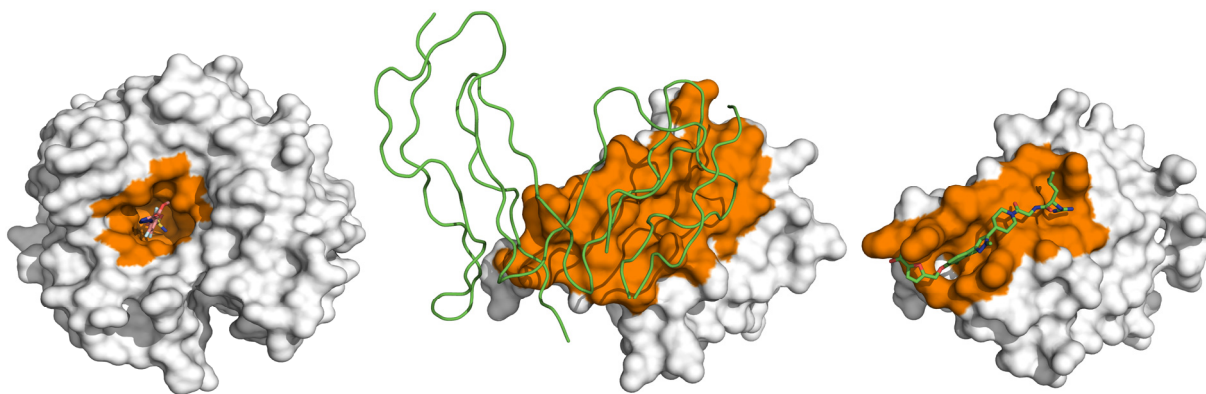


Figure 5. Size and shape of PPIfaces compared to protein/ligand interfaces. Crystal structures show that the interface (orange) of a protein receptor bound to a drug-like ligand lines a pronounced binding pocket and is much smaller (left, carbonic anhydrase^[12], compare Figure 1) than the relatively flat interface between the bound proteins of a typical PPI (center, IL-2/IL-2Rα complex^[175]). In comparison, the interface of the IL-2 protein to a PPIM at its PPIface has a shallow pocket and an intermediate interface area (right, IL-2/FRH complex^[176]).

polar and hydrophobic interactions with the ligand. (vi) Finally, replacing or displacing unfavorably confined water molecules can be enthalpically or entropically favorable.^[178,179]

In contrast, many PPIfaces lack pronounced binding pockets, although there are noteworthy exceptions that I will discuss later (**Chapter 1.2.2**). However, PPIs compensate for this lack as demonstrated by their affinity and specificity, which often even surpass those of small-molecule binding (**Chapter 1.2.3**). In particular, the apparent disadvantage of interacting through relatively flat surfaces is compensated in three ways: (i) the lack of wrapping by a binding pocket is compensated by an enlarged interface area; (ii) the surfaces' remaining roughness, caused by indentations and bulges, increases the interface area and requires geometrical complementarity of the binding partners,^[180,181] (iii) and also flat surfaces can exclude solvent molecules as described by the O-ring theory^[62] (**Chapter 1.2.4**).

Nevertheless, the flat shape of PPIfaces is a problem for the design of PPIMs targeting them. A relatively flat PPIface cannot completely enclose the PPIM. Consequently, a portion of the PPIM's surface, along with functional groups that otherwise could have contributed to affinity and specificity, are not addressed by the PPIM. To compensate for these missing contributions to binding affinity, PPIMs are usually somewhat larger than the drug-like ligands of conventional receptors.^[48] This, in turn, increases the interface area between the protein surface and the PPIM (Figure 5, right). Nevertheless, often some parts of a PPIM, in particular its solvent-facing backside, do not interact with the protein. Consequently, PPIMs have to be larger than ligands of conventional receptors to achieve the same affinity.^[133] This is reflected by the ligand's or PPIMs' *ligand efficiency* (*LE*), which is defined as the free energy of binding per non-hydrogen atom.^[182]

For tightly binding PPIMs, this *LE* is often lower ($LE \approx 0.24 \text{ kcal mol}^{-1} \text{ atom}^{-1}$) than that of ligands of conventional receptors ($LE \approx 0.25 \text{ to } 0.4 \text{ kcal mol}^{-1} \text{ atom}^{-1}$), but still of the same order of magnitude.^[28,48,133,183] In comparison, Wells *et al.* stated that the interface areas of six well-studied PPIs are about twice as large as those of corresponding PPIMs with comparable binding affinities.^[28] Hence, the *LE* of these PPIs is only about half of the *LE* of the corresponding PPIMs.^[28] In addition, Villoutreix *et al.* pointed out that a $LE \approx 0.3 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ would necessitate PPIMs with a molecular weight of $MW \geq 645 \text{ Da}$ in order to achieve a dissociation constant of $K_d \leq 10 \text{ nM}$.^[48] This violation of Lipinski's rules^[131] indicates that it might be hard to identify orally available PPIMs. However, although this implies that special attention is advisable, the oral availability of some non drug-like PPIMs shows that Lipinski's rules are no rigorous exclusion criteria.^[184]

1.2.2 Absence of Pockets

Binding pockets are cavities or pockets in the protein surface or interior that can bind small molecules.^[177] Most PPIfaces lack pronounced binding pockets and are relatively flat. The reason for this is that the binding partners are simply too large to completely wrap around each other. Some exceptional multiprotein complexes can temporarily accommodate an entire protein or large parts of it, such as the proteasome,^[185] chaperonins,^[186] nuclear pores,^[187] and the transporters of the TAT^[188] and the peroxisomal import pathway.^[189] However, such large cavities are not well-suited to tightly wrap around and bind a drug-like ligand.

In any case, the presence of a binding pocket on the protein surface is necessary for the development and binding of PPIMs, but not sufficient, as pointed out by Cheng *et al.*^[168] Consequentially, identifying such binding pockets is usually the first step in assessing target druggability and serves as a starting point for *structure-based drug design* (SBDD).^[42,56,190-197]

Fortunately, PPIfaces are not featureless (Figure 6), but often rough, angled, or provided with crevices.^[180,198] Nussinov *et al.* highlighted that PPIfaces do have pockets, mostly preformed in the unbound state, that upon binding are either left unfilled or complemented and, in this case, essential for the PPI.^[199] However, since pockets in PPIfaces are usually smaller than those of conventional targets^[42] the question is which of them are druggable.^[200]

Experimentally determined structures of protein/PPIM complexes reveal that preformed pockets often constitute PPIM binding sites. For this, PPIMs may also address multiple smaller pockets whose combined volume is comparable to conventional binding pockets.^[42]

More pronounced pockets exist in PPIfaces that bind uneven, protruding, or extended substructures of the protein ligand.^[201] Examples are antibodies, protein kinases, and proteases,^[42,202] which often have extended active site clefts. Accordingly, although a PPIface may extend beyond such a cleft, a small non-globular peptide is often sufficient for binding.

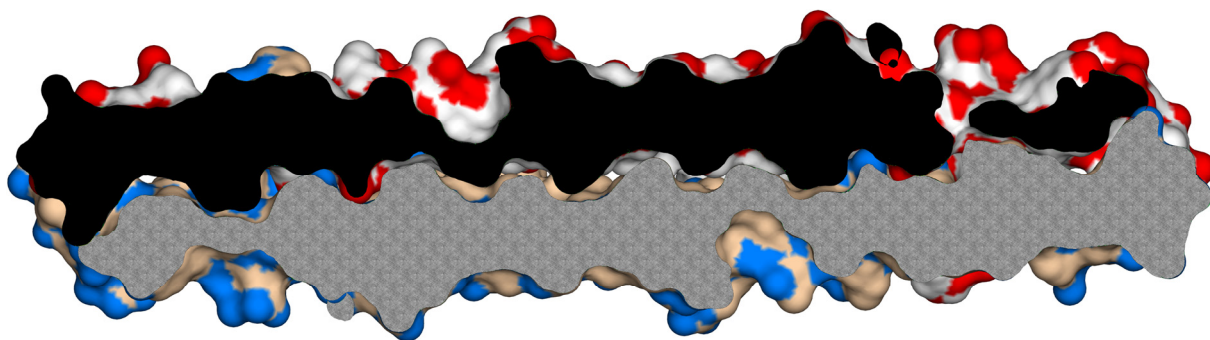


Figure 6. Roughness of a PPIface and complemented pockets. Depicted is a longitudinal section through PPIface between the two dimers of the NHR2 homotetramer (PDB code: 1WQ6^[63]).

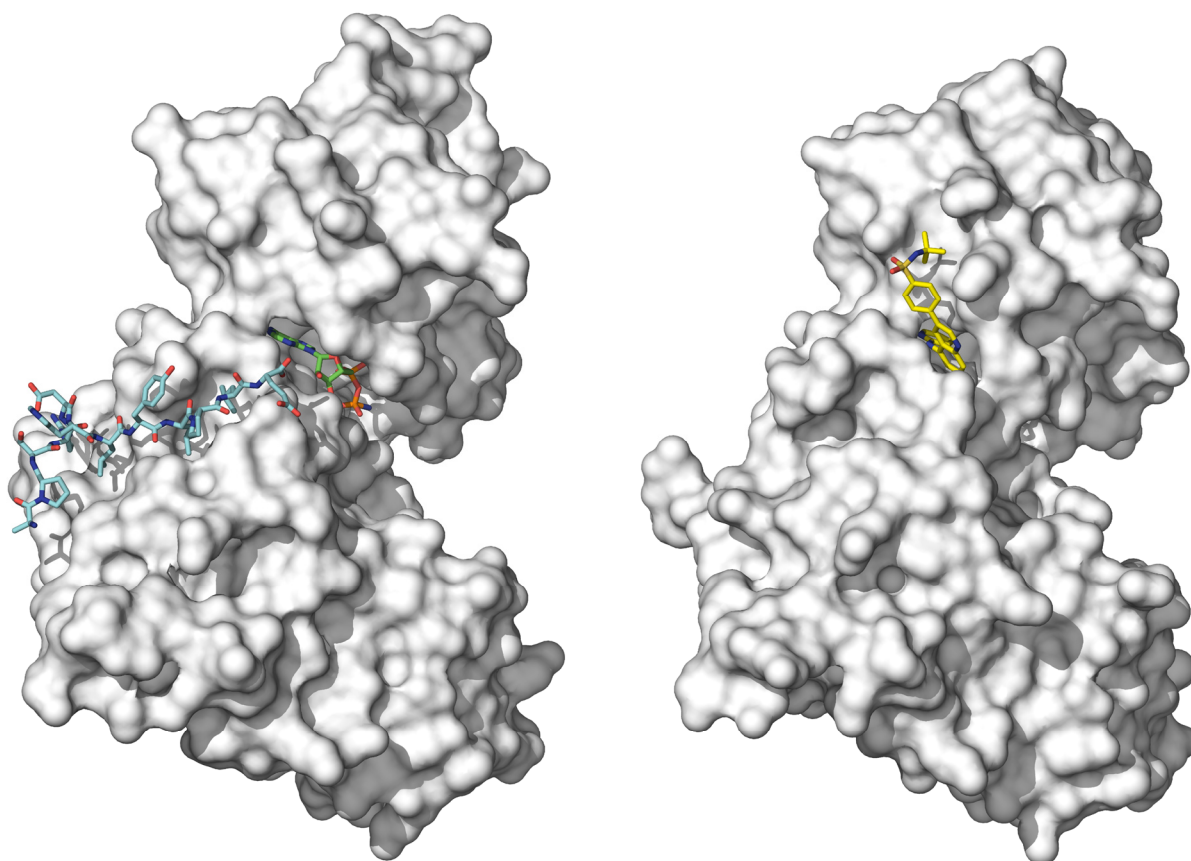


Figure 7. Binding sites of enzymes with protein substrates. Depicted is c-Jun *N*-terminal kinase 1 (JNK1; left; PDB code: 2XRW^[203]) with an ATP-competitive inhibitor in the active site (green sticks) and a co-crystallized inhibitory peptide in a shallow cleft outside of the active site (cyan sticks). An inhibitor of the JNK3 homolog exemplifies the case of ligands extending beyond the active site cleft (right; yellow sticks; PDB code: 2R9S^[204]).

Alternatively, drug-like ligands can efficiently bind to just a sub-region of such a cleft, usually in or close to the active site. Examples include marketed drugs, such as serine protease inhibitors^[205] and non-ATP^[a] competitive (type III) protein kinase inhibitors.^[206-208]

However, while the inhibition of enzymes that act on protein substrates is obviously a type of PPI-modulation, here the key interactions usually do not form between large surfaces of entire folded proteins or domains. Rather, only an exposed substructure of the substrate binds to the active site cleft, similar to a drug-like ligand that binds to the pronounced binding pocket of a conventional target. Thus, inhibitors that bind to active sites or native co-factor and effector binding sites, such as type I and II protein kinase inhibitors^[206] or inhibitors of the dimerization of nuclear receptors,^[209] are usually not considered PPIMs in the strict sense.

In contrast, conventional targets can be modulated by PPIMs that bind to “true” PPIfaces. First, PPIMs may exclusively target, or at least extend to,^[208,210,211] allosteric sites^[212-216] remote from the active site (Figure 7),^[204,217] either within or outside of the PPIface,^[218,219] such as exosites^[b] of proteases,^[205,220,221] protein kinases,^[203,207,210,217,222] or methyl

^[a] ATP: adenosine triphosphate; ^[b] exosite: region of the enzyme distant from the active site that is responsible for specific enzyme/substrate interactions.

transferases.^[223] Second, some PPIMs act by inhibiting the active oligomeric state of an enzyme.^[211,212,220,224,225] Finally, PPIMs can act by binding the macromolecular substrate instead of the enzyme,^[218,226-228] but there are only a few reported examples and structures.^[146] In any case, this raises the question of why one should inhibit enzymes through PPIfaces if there is already a pronounced binding pocket. The disadvantage of targeting such pockets is that they are often conserved across protein families and different genera, as required for binding similar substrates and co-factors. Thus, modulation *via* a PPIface may provide superior specificity compared to targeting active site binding pockets.^[208,210,229]

Nevertheless, there are also many cases of non-enzyme PPIs where an exposed substructure of the protein ligand forms key interactions to the receptor interface,^[42] often by means of hot spots.^[62] An examination by Bullock *et al.* revealed that 62% of the PPI complexes in the *Protein Data Bank* (PDB)^[236] have an α -helix in their interface^[237] and, thus, could potentially be targeted by α -helix mimetics^[184,238-242] (**Pub. II**). Figure 8 shows three disease relevant targets that demonstrate how exposed secondary structure elements that bind to a PPIface are accommodated in druggable pockets or clefts: (i) the antiapoptotic *B-cell lymphoma-extra large* (Bcl-X_L) protein, which either binds the α -helical *Bcl-2* homology domain 3 (BH3) of the proapoptotic *Bcl-2-associated death promoter* protein (BAD)^[230] or the

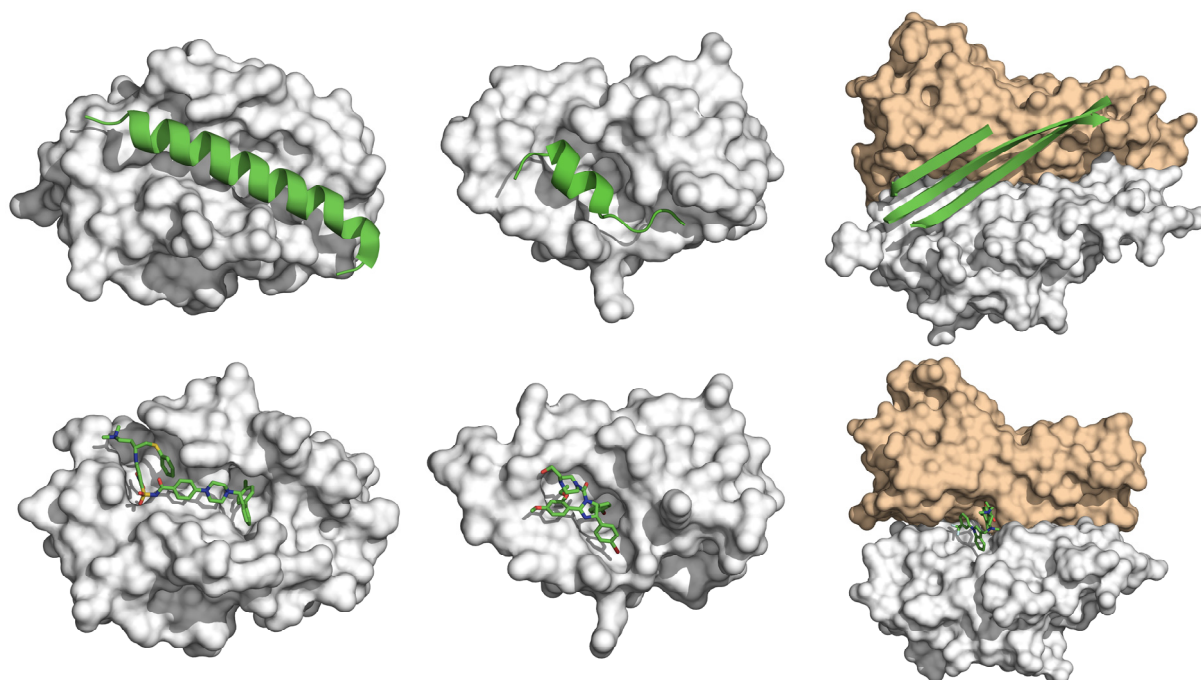


Figure 8. Binding sites of secondary structure elements in PPIfaces. Depicted are exemplary PPIfaces with bound secondary structure elements (top; cartoon) or PPIMs that bind in the respective binding cleft (bottom; sticks). Left: Bcl-X_L binding the interface α -helix of BAD (PDB code: 2BZW^[230]) or the PPIM ABT-737 (PDB code: 2YXJ^[231]). Center: HDM2 binding the interface α -helix of p53 (PDB code: 1YCR^[232]) or the PPIM Nutlin-2 (PDB code: 1RV1^[233]). Right: TNF α homotrimer with the TNF α dimer binding a β -sheet of the remaining TNF α monomer (PDB code: 1TNF^[234]) or a PPIM (PDB code: 2AZ5^[235])

PPIM ABT-737,^[231] (ii) the *human double minute 2* (HDM2) protein, which either binds an interface α -helix of the tumor suppressor *protein 53* (p53),^[232] thereby negatively regulating the latter, or the PPIM Nutlin-2,^[233] (iii) and *tumor necrosis factor α* (TNF α), an immunomodulatory regulator that is active as a homotrimer in which a TNF α dimer binds a β -sheet of the remaining TNF α monomer,^[234] or a competitive PPIM instead.^[235]

The structures of PPI and protein/PPIM complexes and the unbound proteins demonstrate that preformed druggable pockets exist in some PPIfaces. Bourgeas *et al.* investigated the structures of PPIs that are known to bind PPIMs. He pointed out that such PPIs actually have a below-average PPIface area, do not undergo large conformational changes upon PPIM binding, and that each respective PPIface has only a few pronounced pockets.^[174] To some degree, this resembles the situation of conventional targets and suggests that the PPIM targets reported so far may have been easier to target than stereotypically flat PPIfaces. Nevertheless, there exist many PPIfaces that exhibit pronounced pockets if unbound or in a PPI complex.

Notably, PPIfaces that bind PPIMs are more predisposed to pocket formation than the remaining protein surface.^[44,243-245] Transient pockets can form in previously flat PPIfaces (Figure 9). Preexisting pockets can reshape to accommodate and complement PPIMs.^[42,245,246] Due to this high mobility,^[28,44,247,248] binding at PPIfaces often deviates from the lock-and-key model,^[9,10] instead following an induced fit^[249] or conformational selection model.^[250,251] In this case, it is difficult to detect pockets based on a single non-PPIM-bound structure. Nevertheless, conformational sampling and pocket identification algorithms can identify transient pockets,^[194,244,245,252,253] *i.a.* in the PPIfaces of Bcl-X_L, IL-2, HDM2,^[247] and XIAP-BIR2.^[248]

By the example of the IL-2/IL-2R α complex^[175], I explored the application of SBDD to identify druggable sites and transient pockets in a flat PPIface (**Pub. II**). Predicted transient pockets guided the identification of PPIMs from a large number of non-binders. Thus, conformational sampling and pocket identification can help to identify PPIMs even for flat PPIfaces.

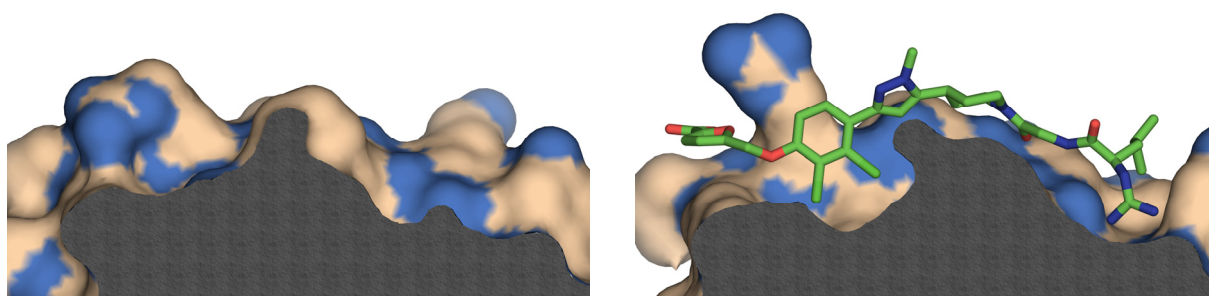


Figure 9. Cross section perpendicular to the PPIface of IL-2 with IL-2R α . Depicted are the PPIface between IL-2 and IL-R α in the unbound state (left; PDB code: 1M47^[243]) and the same interface with a bound PPIM in a binding pocket that was previously not, or not completely, present (right, PDB code: 1PY2^[176]).

1.2.3 Stability of PPIs

PPIs cover a wide range of affinities, with micro- to picomolar K_d values.^[254-256] This corresponds to free enthalpy changes upon binding of $\Delta G_{bind} = -6$ to -19 kcal mol⁻¹.^[198] The essential question is: are there PPIMs with sufficient affinity to compete with a given PPI at a reasonable, *i.e.* physiologically attainable and non-toxic, concentration?

The ability of a competitive inhibitor to efficiently prevent the interactions between two proteins depends on two factors: the binding affinity of the inhibitor, here a PPIM, and the concentration and affinity of the natural ligand, here the protein receptor and the protein ligand.^[254] Wells *et al.* already dispelled the misconception that all native PPI complexes have a generally higher affinity than protein/PPIM complexes and, thus, could not be displaced.^[28] Nevertheless, the high *MW* of most high-affinity PPIMs may affect their bioavailability.^[28,48] In fact, the affinity ranges of PPIs and low-molecular ligands largely overlap.^[28,182] However, the physiological concentrations of the interacting molecules can be very different.

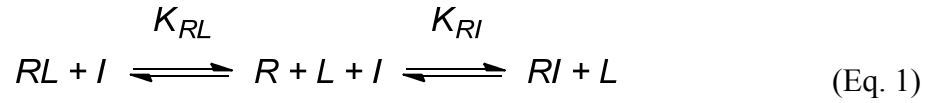
Regarding the proteins, their number ranges from less than 50 to more than 10⁶ molecules per cell.^[257] Of particular interest are receptors on the cell surface, which account for 60% of the exploited drug targets.^[16] Cell-surface receptors are typically expressed at low levels, ranging from 10² – 10³ molecules per cell.^[258,259] Protein ligands of such cell-surface receptors include protein hormones, which usually have nanomolar plasma concentrations that vary by less than one order of magnitude, and cytokines, which in contrast have picomolar plasma concentrations that can increase up to 1,000 fold during trauma or infection.^[260] The number of transcription factors in *Escherichia coli* is typically low (< 100 molecules per cell)^[261-263] but can be much higher in eukaryotic cells (10⁴ – 3 x 10⁵ molecules per cell).^[264] Maiwald *et al.* also reported numbers (and concentrations) for signaling proteins associated with the JAK–STAT pathway^[a] downstream of interferon receptors that range from 2 x 10³ molecules per cell (\cong 10 nM) for PIAS (protein inhibitor of activated STATs) to 4.74 x 10⁵ molecules per cell (\cong 1,500 nM) for STAT1.^[265]

Regarding the small molecules, either their plasma concentration or their cellular concentration is relevant, depending on the location of the target. The relevant plasma concentrations of drugs mostly depend on the administered dose, bioavailability, and the volume of distribution.^[266] Typical peak plasma concentration of anticancer agents after maximum tolerated doses are in the nanomolar^[267,268] to low micromolar range^[269] with similar values for other drugs,^[270-272] but even millimolar concentrations are attainable for less

^[a] JAK–STAT pathway: Janus kinase – signal transducer and activator of transcription pathway

toxic molecules.^[273] Effective free drug concentrations in cells, not to mention in live organisms, are much harder to measure.^[274] However, intracellular drug concentrations reaching the nanomolar^[275] to micromolar range^[276] have been reported.

The competitive inhibition of a PPI by a PPIM can be described as an equilibrium (Eq. 1): the protein receptor (R) binds either the protein ligand (L) to form the bimolecular PPI complex (RL) or to the competitive inhibitor (I) to form the protein/PPIM complex (RI), with the respective dissociation constants K_{RL} and K_{RI} .



This equilibrium can be described analytically and allows to determine molecular concentrations and IC_{50} values (Figure 10).^[277] An analysis of this equilibrium reveals that even PPIMs with moderate affinity ($K_{RI} = 1000$ nM) can inhibit strong PPIs ($K_{RL} = 1$ nM) at physiologically attainable inhibitor concentration ($IC_{50} = 10$ μ M) if the protein concentrations are low ($[R]_0 = [L]_0 = 10$ nM). In addition to this, PPIMs with affinities below the single-digit nanomolar range have been reported.^[278] High affinity inhibitors ($K_{RI} = 10$ nM) are sufficient to inhibit either strong PPIs ($K_{RL} = 1$ nM) at high protein concentrations

		$K_{RI} = 1$ nM				$K_{RI} = 10$ nM				$K_{RI} = 100$ nM				$K_{RI} = 1000$ nM			
		$[R]_0 \cdot \text{nM}^{-1}$				$[R]_0 \cdot \text{nM}^{-1}$				$[R]_0 \cdot \text{nM}^{-1}$				$[R]_0 \cdot \text{nM}^{-1}$			
		1	10	100	1000	1	10	100	1000	1	10	100	1000	1	10	100	1000
$K_{RL} = 1$	$[L]_0 \cdot \text{nM}^{-1}$	3.0E-09	2.1E-08	2.2E-07	2.2E-06	2.5E-08	1.2E-07	1.1E-06	1.1E-05	2.4E-07	1.1E-06	1.0E-05	1.0E-04	2.4E-06	1.1E-05	1.0E-04	1.0E-03
	10	1.1E-08	1.6E-08	2.1E-07	2.2E-06	1.1E-07	1.1E-07	1.1E-06	1.1E-05	1.1E-06	1.0E-06	9.7E-06	1.0E-04	1.1E-05	1.0E-05	9.6E-05	9.9E-04
	100	1.0E-07	1.0E-07	1.2E-07	2.1E-06	1.0E-06	9.7E-07	7.0E-07	1.1E-05	1.0E-05	9.6E-06	6.5E-06	9.6E-05	1.0E-04	9.6E-05	6.7E-05	9.5E-04
	1000	1.0E-06	1.0E-06	1.0E-06	1.1E-06	1.0E-05	1.0E-05	9.6E-06	5.9E-06	1.0E-04	1.0E-04	9.5E-05	5.8E-05	1.3E-03	1.0E-03	9.5E-04	7.1E-04
$K_{RL} = 10$	$[L]_0 \cdot \text{nM}^{-1}$	1.7E-09	8.3E-09	1.0E-07	1.1E-06	1.2E-08	2.7E-08	2.1E-07	2.2E-06	1.2E-07	2.1E-07	1.2E-06	1.1E-05	1.2E-06	2.0E-06	1.1E-05	1.0E-04
	10	2.5E-09	7.9E-09	9.8E-08	1.1E-06	2.1E-08	3.0E-08	2.1E-07	2.2E-06	2.0E-07	2.5E-07	1.2E-06	1.1E-05	2.0E-06	2.4E-06	1.1E-05	1.0E-04
	100	1.1E-08	1.6E-08	6.5E-08	1.1E-06	1.1E-07	1.1E-07	1.6E-07	2.1E-06	1.1E-06	1.1E-06	1.1E-06	1.1E-05	1.1E-05	1.1E-05	1.0E-05	9.7E-05
	1000	1.0E-07	1.1E-07	1.5E-07	5.8E-07	1.0E-06	1.0E-06	1.0E-06	1.2E-06	1.0E-05	1.0E-05	9.7E-06	7.0E-06	1.0E-04	1.0E-04	9.6E-05	6.8E-05
$K_{RL} = 100$	$[L]_0 \cdot \text{nM}^{-1}$	1.5E-09	6.0E-09	6.3E-08	9.0E-07	1.1E-08	1.6E-08	8.4E-08	1.0E-06	1.0E-07	1.2E-07	2.7E-07	2.1E-06	1.0E-06	1.1E-06	2.1E-06	1.2E-05
	10	1.6E-09	6.0E-09	6.2E-08	8.9E-07	1.2E-08	1.7E-08	8.3E-08	1.0E-06	1.1E-07	1.2E-07	2.7E-07	2.1E-06	1.1E-06	1.2E-06	2.1E-06	1.2E-05
	100	2.5E-09	6.8E-09	5.4E-08	8.5E-07	2.1E-08	2.5E-08	7.9E-08	9.8E-07	2.0E-07	2.1E-07	3.0E-07	2.1E-06	2.0E-06	2.0E-06	2.5E-06	1.2E-05
	1000	1.1E-08	1.6E-08	5.7E-08	5.4E-07	1.1E-07	1.1E-07	1.6E-07	6.5E-07	1.1E-06	1.1E-06	1.1E-06	1.6E-06	1.1E-05	1.1E-05	1.1E-05	1.1E-05
$K_{RL} = 1000$	$[L]_0 \cdot \text{nM}^{-1}$	1.5E-09	5.7E-09	4.8E-08	6.0E-07	1.1E-08	1.5E-08	6.0E-08	6.3E-07	1.0E-07	1.1E-07	1.6E-07	8.4E-07	1.0E-06	1.0E-06	1.2E-06	2.7E-06
	10	1.5E-09	5.7E-09	4.8E-08	6.0E-07	1.1E-08	1.5E-08	6.0E-08	6.2E-07	1.0E-07	1.1E-07	1.6E-07	8.3E-07	1.0E-06	1.0E-06	1.2E-06	2.7E-06
	100	1.6E-09	5.8E-09	4.8E-08	5.9E-07	1.2E-08	1.6E-08	6.0E-08	6.1E-07	1.1E-07	1.2E-07	1.7E-07	8.3E-07	1.1E-06	1.1E-06	1.2E-06	2.7E-06
	1000	2.5E-09	6.7E-09	4.8E-08	5.1E-07	2.1E-08	2.5E-08	6.8E-08	5.4E-07	2.0E-07	2.1E-07	2.5E-07	7.9E-07	2.0E-06	2.0E-06	2.1E-06	3.0E-06

Figure 10. Half maximal inhibitory concentrations (IC_{50}) of competitive PPIMs of heterodimeric PPIs in varied conditions: dissociation constant between the protein receptor and either the protein ligand (K_{RL}) or the PPIM inhibitor (K_{RI}) as well as the total concentration of the protein receptor $[R]_0$ and the protein ligand $[L]_0$. Colors indicate calculated^[277] $IC_{50} \leq 10$ nM (yellow), $IC_{50} \leq 1$ μ M (green), $IC_{50} \leq 10$ μ M (blue), $IC_{50} > 10$ μ M (red).

($[R]_0 = [L]_0 = 1000 \text{ nM}$) at a physiologically attainable inhibitor concentration ($IC_{50} = 5.9 \text{ }\mu\text{M}$) or slightly weaker PPIs ($K_{RL} = 10 \text{ nM}$) at moderate protein concentrations ($[R]_0 = [L]_0 = 100 \text{ nM}$) at nanomolar inhibitor concentration ($IC_{50} = 160 \text{ nM}$). Indeed, many potential PPI targets have a lower PPI affinity,^[279] including such with known PPIMs.^[51,174]

It should be added, however, that the physiological situation is more complicated in various aspects. If the PPIM does not completely prevent the PPI formation the result can be a partial competitive inhibition.^[254,280] There can be multiple binding sites, potentially connected allosterically,^[281] and non-specific binding.^[282] Macromolecular crowding affects PPI affinity and specificity.^[6,84,283-288] In addition, the interacting molecules are often distributed heterogeneously due to their association with subcellular compartments,^[83] membranes, lipid rafts,^[289] organelles, oligonucleotides, cytoskeleton, nuclear matrix,^[290] extracellular matrix, or other macromolecular complexes. In addition, cooperative PPIs can amplify the effect of small concentration changes of a bound PPI species on an interaction network.^[291-293] For this purpose, processes such as gene expression and repression,^[294-296] protein synthesis and degradation,^[297] enzymatic modification,^[298] sequestration,^[299-301] or localization^[302,303] can form (futile) cycles and cascade motifs that synergistically boost signal sensitivity.^[304,305]

In addition to affinity, the kinetics of PPIs and protein/PPIM interactions are relevant for PPI inhibition. Transient PPIs may dissociate sufficiently often to allow binding of a PPIM to the interface. However, because the dissociation constant is the ratio between off and on rates, in particular strong PPIs may dissociate slowly^[306,307] or rarely. The resulting hysteresis of the inhibition^[308] may affect the pharmacokinetics of PPIMs. Even permanent quaternary structures may form^[309] so that PPIMs cannot target the occupied PPIface^[310,311] and have to bind before the formation of the PPI complex.^[312] However, in some cases PPIMs have been found to actively disrupt PPIs in a concerted or pre-dissociation independent manner,^[28] as observed for the PPIM induced 600-fold accelerated disruption of the TNF α trimer (Figure 8).^[235,311]

Notably, not all PPIs need a high affinity to ensure specificity and can thus be reversed.^[313] Instead, complementarity, plasticity, and adaption of intrinsically unstructured regions^[96,246,314] can facilitate promiscuous yet selective binding.^[66,315] The similarity of protein folding and protein/protein binding has long been recognized.^[316-320] Moreover, the marginal stability of folded proteins ($\Delta G = -5 \text{ to } -15 \text{ kcal mol}^{-1}$)^[321-323] and the affinity of PPIs are on the same magnitude. Thus, the energetic cost for ordering disordered structures^[313,324] or forming transient pockets^[243,245] upon binding is naturally small. Consequentially, there are many scenarios where it is possible to inhibit PPIs by PPIMs under physiological conditions.

1.2.4 Mechanisms of PPI and Hot Spots

Although the fundamental principles of molecular interactions are universal, proteins have unique characteristics that determine the way they interact with each another. However, PPIs are not as uniform a class as other targets. For example, while all ATP binding pockets bind the same small molecule, PPIs occur between very diverse interfaces.

The combinatorial diversity of PPIfaces emerges from the assembly of complementary and cooperatively interacting functional groups into recognition motifs. In conjunction with their larger interface area, this suggests PPIfaces are more diverse than conventional targets. While this diversity allows highly specific binding at PPIfaces, it is also a reasonable explanation why there is, as yet, no unified structure-based approach for PPIM discovery.^[325]

Consequently, any PPIM, and likewise any attempt for its identification, may have to be tailored for the respective PPIface of interest. However, this is not a severe limitation for PPIM development, as demonstrated by hundreds of PPIMs for more than 40 different proteins that cover most folding classes.^[325] In fact, the structural and physicochemical similarities of known PPIMs (**Pub. I**) suggest that there are common features in their binding mechanisms.^[48,325,326] Conversely, this suggests that there are common features in the binding mechanisms of PPIs that are mimicked by the respective PPIMs.^[327]

The detailed structural characterization of an increasing number of PPIs is continuously improving our understanding of PPIs.^[43,56,328] Evidence suggests that proteins interact in preferred ways.^[198] The size, shape, and absence of pockets in PPIfaces have already been discussed (**Chapters 1.2.1 and 1.2.2**). As an additional aspect, many PPIfaces, in particular of homodimers, are more hydrophobic than the remaining protein surface.^[329,330] A good example is the elongated hydrophobic PPIface of leucine zippers, such as that of NHR2^[63] (Figure 11, **Pub. IV**). In this regard, PPIfaces resemble the hydrophobic core of folded proteins.^[316-320] In contrast, PPIfaces that mediate specific PPIs reportedly have more

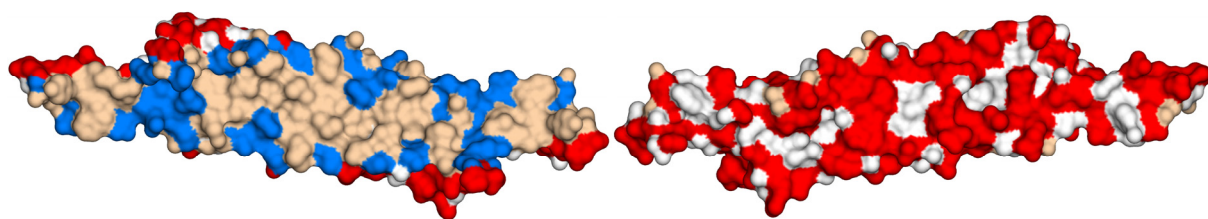


Figure 11. Hydrophobicity of PPIfaces. The rear (left) and frontal dimer (right) of the NHR2 tetramer crystal structure^[63] are laterally displaced by 110 Å. The revealed PPIface of the rear dimer is largely hydrophobic (wheat) with hydrophilic amino acids lining the elongated rim (blue). In contrast, the solvent exposed surface is largely hydrophilic (red) with scattered hydrophobic patches (white).

hydrogen bonds than non-specific PPIfaces and crystal contacts.^[329] In turn, this suggests that not purely hydrophobic epitopes may bind more specific PPIMs^[331] (**Pub. IV**).

Although the exclusion of water from a hydrophobic PPIface causes a favorable hydrophobic effect,^[329,332] hydrophobicity alone is usually not a distinctive mark for the identification of PPIfaces.^[329,333] Not all PPIfaces are *dry*.^[334,335] Rather, interfacial water molecules highlight their importance for mediating hydrophilic interactions in PPIs.^[179,336-339] Thus, water-mediated interactions,^[340] displacement of structurally bound water, and the action of water as a dielectric are critical for both PPIs and PPIM binding.

An even more important aspect for PPIM binding is that not all parts of a PPIface contribute equally to the binding free energy. Often, just few *hot spots* that form core interactions account for the bulk of binding free energy. **Hot spots** are *amino acid residues* (hereinafter "residues") that contribute significantly to binding and, in the original definition, cause a drop in the binding free energy of ≥ 2 kcal mol⁻¹ if they are mutated to alanine.^[62,173] In a complementary definition, **binding hot spots** are regions on protein surfaces that interact strongly with ligands,^[341-345] anchor residues,^[201] or hot spots. Naturally, hot spots on one face of the PPIface often pack against hot spots on the other protein's surface (Figure 12).^[62]

Unlike a situation where essential interactions are scattered across the PPIface, hot spots tend to cluster in stable patches in the center of PPIfaces.^[197] A narrow cluster of hot spots, a hot region,^[316,349] is an excellent starting point to identify PPIMs that address or mimic hot spots and may be a prerequisite for PPIM binding altogether.^[350,351] Notably, diverse mechanisms of PPIs make use of recurring topologies that are suited for drug design.^[352] For this reason, I will now introduce two important and frequently exploited mechanisms of PPI.

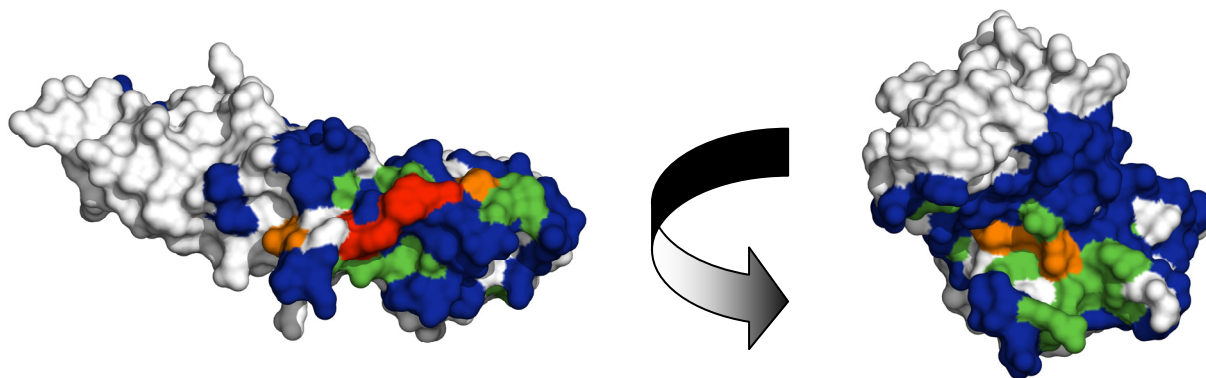


Figure 12. Complementary hot spots in the PPIface of a human growth hormone receptor monomer (left) with human growth hormone (right) that are enveloped by an O-ring. Amino acid residues are colored by their energetic contribution to binding energy, as measured by experimental alanine scanning (blue: ≤ 0.5 kcal mol⁻¹, green: $0.5 \leq 2.0$ kcal mol⁻¹, orange: $2.0 \leq 3.5$ kcal mol⁻¹, red: > 3.5 kcal mol⁻¹, white: no data).^[173,346-348]

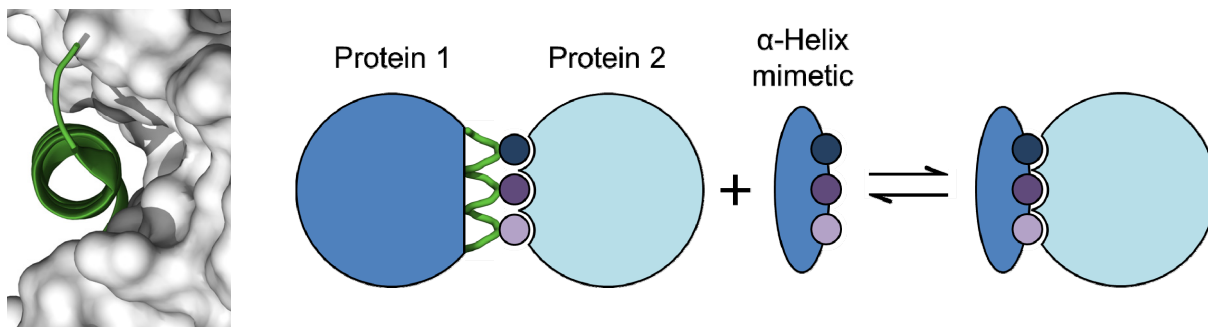


Figure 13. α -Helix of BIM (Bcl-2 interacting mediator of cell death) buried in a groove in the PPIface of Bcl-X_L (left)^[355] and the mode of action of α -helix mimetics (right): the arrangement and interactions of hot spots (small circles) on an α -helix (green) are mimicked by an α -helix mimetic. Depiction based on Ref. ^[184].

First, the *O-ring theory* describes a frequently encountered PPIface topology and an elaborate mechanism that explains the clustering of hot spots in the center of PPIfaces.^[62] Here, hot spots are occluded from the solvent by a ring of energetically less important but densely packed residues (Figure 12).^[340] This desolvation may progress towards the center of a PPIface.^[335] However, although water-free hot spots are favorable,^[353,354] some hot spots again require interactions that are mediated by structural water molecules.^[179,339,340]

Nevertheless, the lowered dielectric constant in the center of an O-ring strengthens electrostatic and hydrogen bonding interactions. In this regard, the ambiance within an O-ring resembles that in a hydrophobic binding pocket or the inside of a folded protein.^[317,318] Strikingly, hot spots are often occluded from solvent (> 70% burial on average), although occlusion alone is not a sufficient condition for a hot spot.^[62] O-rings can also strengthen hydrophobic interactions by hindering the interloping of water into a PPIface, thus decreasing the dissociation rate. Consequentially, O-rings favor hot spots that are capable to form both hydrogen bonding and aromatic interactions, such as tryptophan, tyrosine, and arginine.^[62,354]

Regarding PPIMs, there are two obvious consequences. On the one side, the spatial clustering of hot spots by itself is allegedly the very basis for PPIM binding at PPIfaces.^[356] On the other side, it is rather unlikely that a drug-like molecule could mimic an entire O-ring. However, quite recently a *molecular tweezer* has been reported that binds a single surface-exposed lysine in the PPIface of the 14-3-3 adapter protein and inhibits its PPIs with the c-Raf (cellular rapidly accelerated fibrosarcoma) protein and the ExoS (exoenzyme S) protein.^[357] In a way, this is the reversal of the situation with an O-ring, at which the center of a large PPIface is buried and occluded from solvent: here a large ligand buries a local protein feature, the exposed lysine side chain, in its hydrophobic interior. While this molecular tweezer certainly is not a good example for drug-likeness, it illustrates the importance of burial and solvent occlusion for hot spots and PPIM binding, in line with the O-ring theory.

As the second prevalent mechanism of PPI, I describe the binding of α -helices in PPIfaces.^[237,358,359] Such α -helices illustrate that burial and solvent occlusion can also be achieved by the binding of exposed compact substructures of one protein into a rather pronounced cavity of its PPI binding partner (Figure 8). There are many examples of non-flat PPIfaces, in particular if they bind α -helices (see **Chapter 1.2.2**). It is thus not surprising that such α -helix binding cavities in PPIfaces have been widely exploited for ligand identification. This led to what is likely the best-investigated class of PPIMs: *α -helix mimetics*, organic scaffolds that project chemical moieties similar to an α -helix (Figure 13; **Pub. III**).^[184,238-242]

While loops, β -sheets, and other surface exposed protein substructures may similarly form protruding epitopes that bind into PPIface cavities, α -helices and comparable helices offer at least four inherent advantages with respect to the identification of PPIMs.

First, α -helices are the most frequent secondary structure elements in PPIfaces,^[237] making their mimicry appealing for a wide variety of potential PPI targets. One reason for this prevalence is that (amphipathic) helices may direct their hydrophobic face^[184] towards the interior of their supporting protein while the remaining face is exposed to the solvent or binding partner. Such an amphiphilicity is less common for other secondary structure elements.

Second, α -helices are fairly stable by themselves because their backbone hydrogen bond donors and acceptors are internally saturated by forming interactions between consecutive α -helix turns.^[360,361] Actually, many proteins interact by a single detached α -helices,^[184] *i.a.* BH3-only proteins,^[81,362] leucine zippers,^[363] and also NHR2^[63,76] (**Pub. IV**). This facilitates deriving short, helical, pre-stabilized inhibitory peptides that can be excised from the PPIface region of a PPI protein. By presenting a continuous and ideally hot spot-containing epitope^[364] of the originating PPIface, such peptides are valuable for target validation, druggability assessment, and as starting points for the identification of PPIMs and CADD.^[359,365]

Third, α -helices are intrinsically compact^[361] (diameter $\approx 7 - 15 \text{ \AA}$ ^[366]) rod-like structures that can be enclosed by an extended crevice or groove (Figure 8 and Figure 13). Such grooves necessarily, but not exclusively, arise from a parallel arrangement of helices, like in helix-bundle and coiled-coil structures,^[367] even in the case of optimal hexagonal packing.^[368] Further, the similar dimension of drug-like molecules and helix-containing cavities can promote druggability and provide the foundation for more intricate details of the binding mechanisms. This facilitates rational PPIM discovery, as in case of α -helical nuclear-receptor box peptides that bind to nuclear receptors.^[369,370] Here, a *charge clamp* of oppositely charged

residues that stabilize the inherent helix dipole in the PPI is also exploited by antagonistic PPIMs,^[371] which were identified by a mixed structure- and peptide/ligand-based VS.^[372]

Finally, α -helices project amino acid side chains radially in a very determinate manner. The ca. 3.6 amino acids per α -helix turn^[361] place an amino acid triplet at sequence positions i , $i+3$ or $i+4$, and $i+7$, in direct succession on one face, and in parallel to the longitudinal axis, of their supporting α -helix^[239] (Figure 13, **Pub. III**). These side chains cluster in a continuous, often buried epitope on the α -helix that is parallel to the accommodating PPIface. In turn, α -helix mimetics are privileged scaffolds that mimic this regular side chain arrangement and have proven themselves as competitive inhibitors for various PPIs. This encouraged me to investigate the capability of teroxazoles as α -helix mimetics^[184,238-242] (**Pub. III**), also with the aim to design inhibitors of the tetramerization of the α -helical NHR2 protein (**Pub. IV**).

Aside from these quite general mechanisms, molecular details are crucial for PPIs and PPIM binding. At $T = 300$ K, a change of $1.4 \text{ kcal mol}^{-1}$ in ΔG_{bind} corresponds to a 10-fold change in K_d . Hence, even a single missing or suboptimal hydrogen bond, repulsive methyl group, or inadequately replaced water can make a big difference for binding or non-binding.

Additionally, proteins and their ligands are often not rigid and can mutually adjust to optimize their complementarity. In contrast, the most common structure determination method, X-ray crystallography, affords static conformations and very limited flexibility or mobility information. Nuclear magnetic resonance (NMR) spectroscopy affords structural ensembles and increasingly, but still limited, flexibility or mobility information. The plasticity of PPIfaces, however, allows the formation or widening of previously absent or shallow pockets^[44,246-248] and ligand binding in line with the induced fit^[249] or conformational selection model.^[250]

Thus, the rational identification of PPIMs must account for the important details of molecular structures, interactions, and plasticity while keeping a reasonable cost-benefit ratio (**Chapter 3**). To this end, I developed a computational strategy for PPIM identification that simultaneously considers aspects of energetics and plasticity in form of predicted hot spots and transient pockets. First, I validated this strategy by retrieving known PPIMs of IL-2 from a large number of non-binders (**Chapter 4**). Using this strategy, I identified the first PPIMs that inhibit the tetramerization of NHR2 (**Chapter 6**). This general strategy is applicable to a large number of targets because it was designed to identify PPIMs starting from only a PPI structure. In the following, I will discuss the relevance and interactions of IL-2 and NHR2, which served as test cases in my retrospective and prospective investigations, respectively.

1.3 The PPI of Interleukin-2 and its α -Receptor

IL-2 is a key cytokine involved in the regulation of the immune system.^[374] It is relevant for immunological diseases, transplant medicine, and cancer.^[52,375,376] Binding of IL-2^[377] to the trimeric IL-2 transmembrane receptor^[378,379] is initiated by the association of IL-2 to the extracellular α -helical domain of the receptor's α subunit (IL-2R α).^[175] Inhibiting the IL-2/IL-2R α interaction is potentially advantageous compared to the severe adverse effects associated with other intervention points in the IL-2 signaling pathway.^[380,381] Crystal structures of *apo* IL-2, IL-2 bound to five PPIMs, the IL-2/IL-2R α complex (Figure 14) as well as affinity data, also for 52 additional PPIMs, are known.^[382-384] Thus, IL-2 is an ideal test case for investigating the determinants of PPI and PPIM binding and for establishing and validating a rational strategy to identify competitive PPIMs starting from only the PPI structure (**Pub. II**).

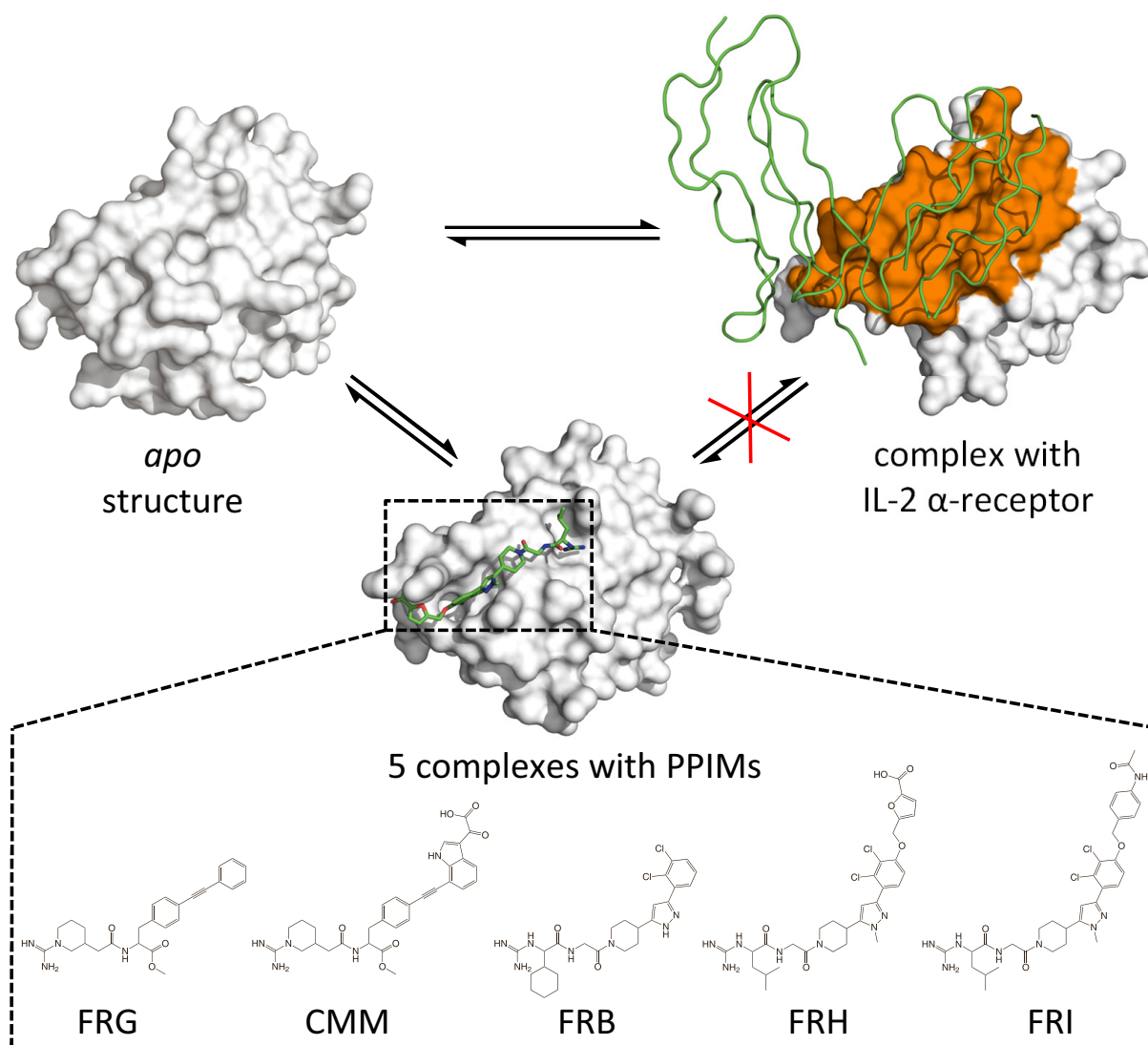


Figure 14. The crystal structures of IL-2,^[243] IL-2/IL-2R α ,^[175] and IL-2 in complex with one exemplary PPIM (FRH^[176]) in an schematic equilibrium that illustrates the competitive inhibition of IL-2/IL-2R α by the five PPIMs with known crystal structures in complex with IL-2 (bottom).^[176,243,373]

Before my investigation, alanine mutation experiments had confirmed the existence of shared hot spots for the binding of IL-2R α and PPIMs.^[176,373,382,385] In this case, the PPIMs of IL-2 mimic hot spots of IL-2R α .^[373] Given the large surface area of most PPIfaces, *e.g.* 2500 Å² for the IL-2/IL-2R α interaction, such hot spots are an excellent starting point for the identification of PPIMs. However, when dedicated knowledge about PPIMs is missing these hot spots must be predicted differently, *e.g.* based on the PPI structure.

Moreover, PPIM-bound IL-2 structures exhibit a binding pocket that is absent or shallow in non-PPIM-bound IL-2 structures (Figure 9). Naturally, the absence of this pocket is a major obstacle for any attempt to identify PPIMs based on a non-PPIM-bound IL-2 structure. Thus, it may not be surprising that the first PPIMs of IL-2 were not found in a structure-based design effort to identify IL-2 ligands but by serendipity, notably, in an empirical screening for IL-2R α ligands that mimic hot spots of IL-2.^[385,386] Only later, second generation PPIMs of IL-2 were designed using structural knowledge of IL-2/PPIM complexes partially obtained by covalent ligand tethering^[387-390] or fragment-based ligand design.^[229,382,383,391]

Nonetheless, Eyrisch and Helms identified, concurrently with to my own efforts (**Pub. II**), transiently opening pockets in the PPIfaces of unbound IL-2 and other PPIs.^[247] To this end, surface plasticity had to be considered adequately, *e.g.* by searching pockets in a conformational ensemble generated by MD simulations of the protein.^[247] Molecular docking into these transient pockets, in line with the conformational selection model,^[250] produced binding poses close ($RMSD < 2$ Å)^[a] to the respective PPIM-bound crystal structures.^[247]

These results indicate that hot spots and transient pockets are complementary determinants of PPIM binding at the IL-2 PPIface. The VS for PPIMs of IL-2 that I performed integrates both aspects by considering binding energetics and surface plasticity (**Chapter 4**).

1.4 The NHR2 Oligomerization Domain of RUNX1-ETO

NHR2 is the α -helical oligomerization domain of the RE fusion protein^[392] (hereinafter “RE”; abbreviations in footnotes). RE is present in ca. 12% of all *de novo* cases of AML (40% of FAB^[393] subtype M2)^[394] as a consequence of translocation t(8;21) that joins the hematopoietic^[395,396] master regulator gene RUNX1^[397-399] on chromosome 21 (*also* AML1, CBF α 2, or PEBP2 α B) and the ETO gene on chromosome 8 (*also* MTG8, RUNX1T1, or

^[a] *RMSD*: root mean square deviation; FAB: French–American–British classification system of acute leukemias; NHR2: neryv homology region 2; RUNX1: Runt-related transcription factor 1; ETO: eight twenty-one; AML: acute myeloid leukemia; CBF α 2: core-binding factor subunit α 2; PEBP2 α B: polyomavirus enhancer binding protein 2 α B; MTG8: myeloid transforming gene on chromosome 8; MTG16: myeloid transforming gene on chromosome 16; RUNX1T1: Runt-related transcription factor 1; translocated to, 1

CBFA2T1^[a]).^[400] Thus, RE contains the 177 *N*-terminal amino acids of RUNX1, including the DNA-binding Runt domain, fused to almost the entire ETO protein,^[401] including the four nery homology regions NHR1 to NHR4 (Figure 15).^[402,403] Notably, the structurally equal homolog RUNX1-ETO2^[404,405] is also found in *de novo*^[406] and therapy-related AML.^[407,408]

The regulatory network related to RE is only partially understood and a thorough description beyond the scope of this thesis.^[409-412] A short summary of the interactions associated with RE domains is given in the Appendix (**Chapter 9.2**). Briefly, different domains of RE interact with regulatory DNA,^[110,413] p53,^[414] PKA,^[415,416] and various nuclear proteins, *i.a.*, involved in transcriptional regulation including epigenetic mechanisms.^[63,76,400,410,417-440] Modulating the (epigenetic)^[79,441] action and PPIs of these RE-associated co-factors suggests various strategies against RE-dependent AML (RE-AML).^[216,442,443] However, although some co-factors bind RE monomers,^[63,400,444] only the α -helical NHR2 homotetramer bundle binds HEB,^[76] recruits NHR2-containing ETO homologs (ETO, ETO2, MTGR1),^[445] mediates RE self-association,^[63,400,402] and increases the affinity to preferably duplicate RUNX1 binding motifs.^[400,446]

The resultant high molecular weight complex ($MW > 2$ MDa)^[312,402] allows a cooperative regulation of gene expression.^[76,110,447] Here, RE generally, but not exclusively,^[76,410,448,449] represses RUNX1 target genes,^[76,418,430,431,439] *i.a.* by chromatin remodeling^[110,111] and block or sequestration of transcription factors.^[450,451] This blocks myeloid precursor cell differentiation and promotes RE-AML.^[400,402] Thus, the presence of RE, or its NHR2-containing splice variants,^[452,453] is an essential prerequisite for RE-AML,^[402,454,455] yet needs additional genetic aberrations^[456-458] in line with the multiple-hit hypothesis.^[459] Although RE-AML usually has a favorable prognosis,^{[460,461][b]} subsets are associated with poor prognoses and relapses after chemotherapy,^[462] *e.g.* if featuring c-Kit mutations^[409] or therapy-related RE-AML.^[463,464]

Various studies indicate the critical role of NHR2 and RE in leukemia. ETO2 (*also* CBFA2T3) is a hematopoietic regulator,^[465,466] tumor suppressor,^[467-469] affects cell metabolism *via* NHR2,^[470] and is part of two fusion proteins related to poor outcome in leukemia: (i) NFIA-CBFA2T3 in AEL^[471] and (ii) CBFA2T3-GLIS2 in pediatric CN-AML^[472] and linked to an aggressive subtype of non-DS-AMKL.^[473,474] An alternate RE fusion protein that

^[a] CBFA2T1, 2, 3: CBFA2; translocated to, homolog 1, 2, 3; HEB: HeLa E-box binding protein; c-Kit: cellular v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; NFIA: nuclear factor 1 A-type, a dimeric DNA-binding transcription factor; AEL: acute erythroid leukemia; GLIS2: glioma-associated oncogene (GLI)-similar protein 2; CN-AML: cytogenetically normal AML; non-DS-AMKL non-Down syndrome pediatric acute megakaryoblastic leukemia; ^[b] Still relapse rate $\approx 40\%$ and long-term survival $< 50\%$.

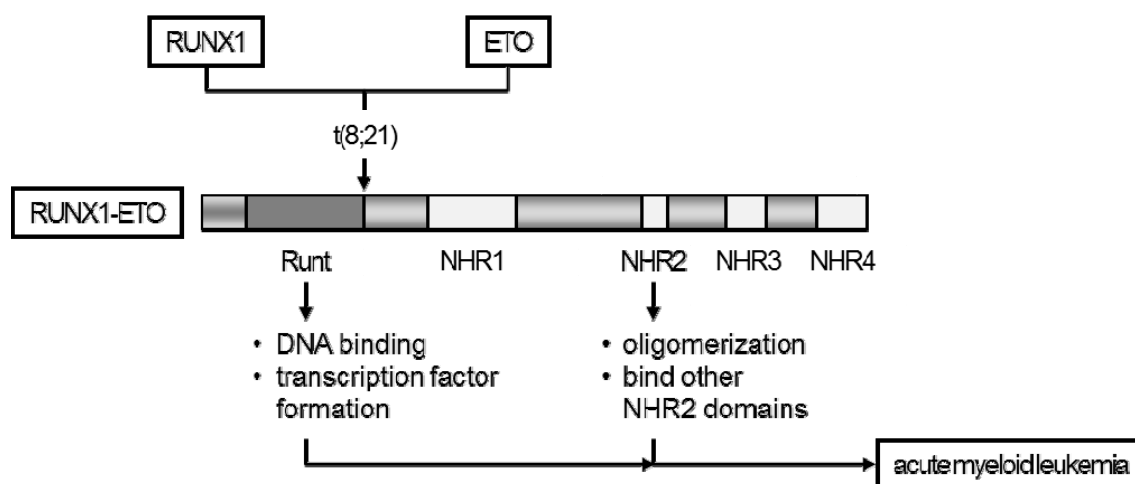


Figure 15. Domain structure and schematic mode of action of RE.

retains the PRMT1 target site synergizes with BCR-ABL in CML.^[475] Gene fusion with the IGH switch region dysregulates ETO2 in Burkitt lymphoma and DLBCL.^[476] Strikingly, replacing NHR2 by FKBP allowed a PPIM-induced disassembly of oligomeric RE constructs, which selectively interfered with leukemogenesis in transfected cells.^[477] Cleavage of RE by caspases-3 before NHR2^[478-481] or by cathepsin G^[482] causes anti-leukemic effects. Moreover, oncogenic RUNX1 mutants and fusion proteins exist,^[399,483-492] RE-AML even depends on native RUNX1,^[482,485] and ETO homologs are involved in non-leukemic cancer.^[467,493-495]

The homotetramerization of NHR2,^[63,110,400] more precisely the dimer-tetramer transition (hereinafter "NHR2 tetramerization"),^[400] is a key step in the onset and maintenance of RE-AML (Figure 15). Crucial to this finding was a computational hot spot prediction (Chapter 6),^[92,496,497] which revealed five residues in the NHR2 PPIface that are essential for NHR2 tetramerization (hereinafter "NHR2 hot spots"; W498, W502, D533, E536, and W540; Figure 16). Notably, the NHR2 hot spots are located close to the largest pocket in the NHR2 PPIface, not in the flat center of the hydrophobic leucine zipper (Figure 11) but clustered at both of the remote ends (Figure 16).^{[400][b]} Notably, alanine mutations of these NHR2 hot spots interfere with NHR2 tetramerization but preserve the α -helical structure of NHR2.^[400]

The NHR2 hot spots allowed validating NHR2 as a target as follows.^[400] RE lacking NHR3 and NHR4 (REtr) blocked myeloid differentiation *in vitro* and induced leukemia in mice,^[498] unlike the dimeric fivefold hot spot alanine mutant (REtr-m5).^[400] The intracellular expression of NHR2-containing peptides inhibited RE oligomerization, prevented the proliferation, and induced the differentiation of RE-transformed cells.^[400,402] Cell-penetrating

^[a] PRMT1: Protein arginine *N*-methyltransferase 1; BCR-ABL: BCR-ABL fusion protein (BCR breakpoint cluster region protein; ABL: Abelson murine leukemia viral oncogene) CML: chronic myeloid leukemia; IGH: immune-globulin heavy chain; DLBCL: diffuse large B-cell lymphoma; FKBP: FK506 binding protein;

^[b] The author contributed fruitful discussions to this study.

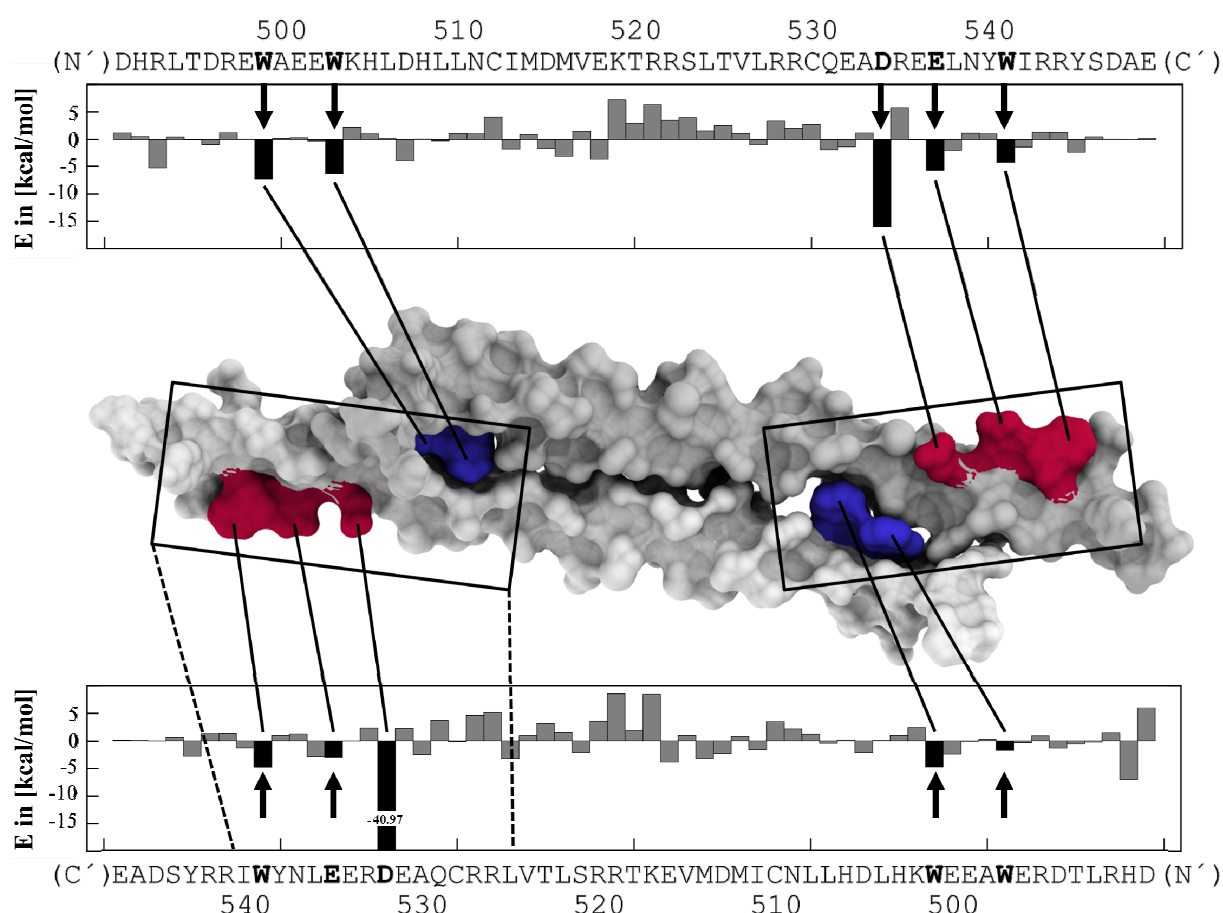


Figure 16. Hot spots of NHR2 tetramerization. Top view onto the PPIface on one NHR2 dimer in the conformation of the homotetrameric crystal structure (center).^[63] Contributions of individual amino acids to the computed binding free energy,^[92,400] calculated by the MM-GBSA method (**Chapter 1.5.1**), are plotted against the amino acid sequences of each monomer (top and bottom; RE sequence numbering).^[392] Hot spots are highlighted (bold and colored surface) and cluster at the tips of the NHR2 dimer (boxes).

TAT^[a]-NHR2 fusion peptides reduced the proliferation of RE-dependent Kasumi-1 cells and promoted their apoptosis.^[312] Finally, complementing *MTG16*^{-/-} cells^[b] by a MTG16 variant that lacked most of NHR2 except for the three crucial hot spots partially reconstituted T-cell development in mice.^[465]

In summary, inhibiting the NHR2-mediated tetramerization of oncoproteins is a valid strategy for molecular intervention towards a personalized therapy of RE-AML.^[76,400,402,460,499] The predicted hot spots are an excellent starting point for SBDD and encouraged me to search for the first NHR2 inhibitors, which aim at the NHR2 PPIface and prevent NHR2 tetramerization from dimers (**Chapter 6**).

^[a] TAT: trans-activator of transcription; ^[b] LSK (Lineage⁻/Sca1⁺/c-Kit⁺) cells

1.5 Selected Computational Methods

The mutual complementarity of computational and experimental methods can promote the efficiency and reliability of molecular investigations. Computational methods to analyze molecules and identify ligands of conventional targets are well-established.^[57-59,500] However, these methods must be adapted to the particular challenges of PPIs (**Chapter 1.2**). In **Publication I**, I reviewed the state-of-the-art in computational PPI analysis and PPIM identification. Here, I introduce two methods, MM-PBSA and ROCS, that were crucial for the later described hot spots detection and the structure-based identification of NHR2 inhibitors.

1.5.1 Hot Spot Prediction by the MM-PBSA Method

The MM-PBSA (MM: molecular mechanics, PB: Poisson-Boltzmann, SA: surface area) method performs a physics-based calculation of binding free energies (ΔG_{bind}).^[92,93,501-503] Recently, the method was reviewed in detail.^[504] In the following, I introduce its key aspects.

As an end-point free energy method, MM-PBSA calculates ΔG_{bind} (Eq. 2) as the difference of the free energies of bound ($G_{complex}$) and unbound species ($G_{receptor}$, G_{ligand}).

$$\Delta G_{bind} = \langle G_{complex} \rangle - \langle G_{receptor} \rangle - \langle G_{ligand} \rangle \quad (\text{Eq. 2})$$

To account for conformational flexibility and mobility, average free energies are calculated based on conformational ensembles ($\langle \dots \rangle$) from molecular dynamics (MD) simulations. Statistically independent conformations are extracted from the simulation trajectories and the solvent is discarded. The absolute free energy (G_{tot}) of each species is then calculated as the sum of four terms (Eq. 3): internal energy based on an MM force field (E_{MM}), electrostatic solvation free energy based on the PB continuum method (G_{PB}),^[505,506] non-polar solvation free energy as a function of SA (G_{SA}), and a configurational entropy contribution ($-TS$).

$$G^{tot} = E_{MM} + G_{PB} + G_{SA} - TS \quad (\text{Eq. 3})$$

Two simplifications are often made in MM-PBSA. First, neglecting the entropic term in Eq. 3 leaves an effective energy (G_{eff}). This allows satisfactory energy predictions and per-residue energy decompositions for similar ligands^[507] but avoids costly harmonic or slowly converging quasiharmonic entropy approximations.^[507-509] Second, extracting unbound conformations from the complex trajectory (1-trajectory approach) saves simulating the unbound states (3-trajectory approach). This neglects the reorganization energy upon association but leads to a cancellation of errors arising from non-converged sampling and noise.^[92,93,179,510]

Besides its value for predicting binding affinities, MM-PBSA allows decomposing ΔG_{bind} into residue-wise contributions (Figure 16) and, hence, predicting PPI hot spots as shown first for the interaction of Ras (rat sarcoma) with Ras effector proteins.^[92] Likewise, hot spots of NHR2 tetramerization were identified^[400] that guided my identification of NHR2 inhibitors (**Pub. IV**). Both studies used the MM-GBSA variant, in which the PB method is replaced by the more efficient Generalized Born (GB) approximation.^[501,502,511] However, the PB method showed better affinity predictions for PPIMs of IL-2. Thus, for the identification of IL-2 hot spots, we extended the energy decomposition scheme by the PB method (**Pub. II**).

As an alternative to MM-PBSA, computational alanine scanning^[512-514] predicts changes of ΔG_{bind} ($\Delta\Delta G_{bind}$) caused by virtual alanine mutations. However, the resultant perturbations may transcend specific localized interactions in a PPIface and may, thus, lead to misinterpretations.^[92] In contrast, MM-PBSA decompositions do not perturb the protein structure and, thus, reveal druggable sites as they are present in the PPIface of the native PPI complex.

1.5.2 3D Similarity and Superposition by ROCS

ROCS (rapid overlay of chemical structures) compares molecules by their shape and distribution of physicochemical properties (termed *colors*, Figure 17).^[515-518] Rigid conformers^[515,519,520] are rapidly superimposed by optimizing the overlap of atom-centered Gaussian functions^[517] that describe shape and colors. The 3D similarity is quantified by the Tanimoto index.^[521] Shape and colors (hydrogen bond donor/acceptor, anion/cation, hydrophobe, and ring) describe general molecular features and, thereby, allow identifying similar molecules but also distinct isofunctional molecular structures (scaffold- or lead-hopping).^[521-523]

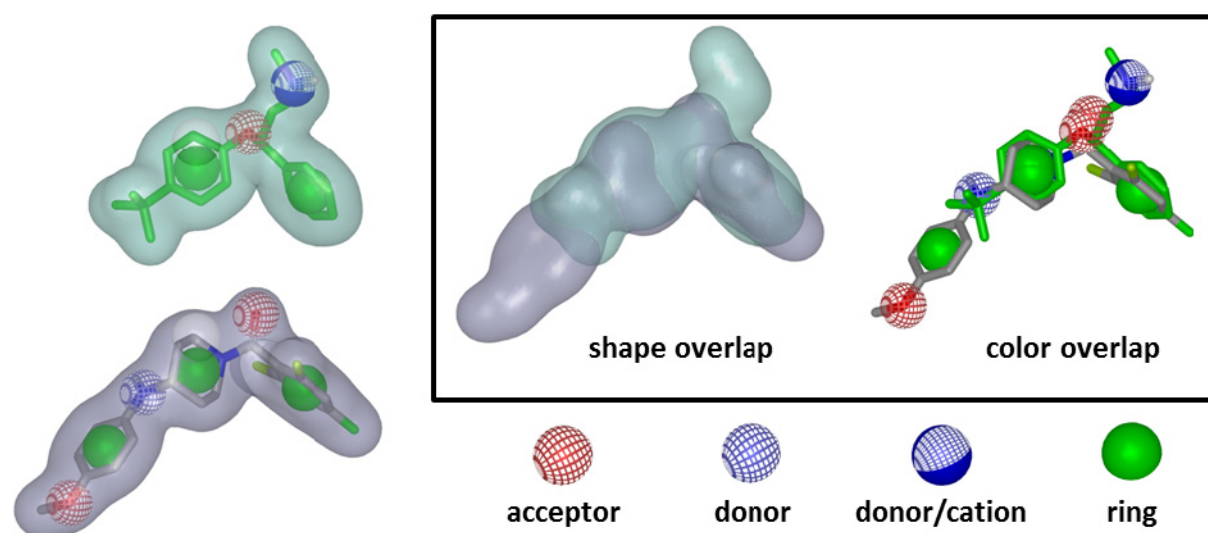


Figure 17. ROCS superimposition (box) of two molecules (left) by shape (surface) and physicochemical properties (spheres, for color types see legend below). Graphics by vROCS.^[515]

Methodically, conformers are pre-generated^[a] for a large set of molecules, which may be filtered by prior steps of VS.^[b] Each color is defined by a set of substructural patterns that describe (iso)functional groups in the context of a color force field (ImplicitMillsDean),^[524] which is later used to optimize the color overlap. These patterns are described in SMARTS^[c] line notation^[525] and can be customized. To describe shape and color distributions, spherical Gaussian functions are centered on the respective atoms or color features.

Gaussians allow rapidly calculating molecular volumes and optimizing their overlap. The smoothness of Gaussians prevents “getting stuck” in local minima, as occurs often with hard sphere models. In addition, the overlap between any number of atoms is calculated using, again, Gaussians,^[d] whose volume can be calculated analytically, tabulated, and retrieved efficiently.^[516,526-528] This allows processing thousands or even millions^[528,529] of conformers per second and large (virtual) combinatorial libraries ($> 10^8$ conformers) in a timely manner.

A similar shape and color distribution often indicates a similar biological function. ROCS calculates this similarity requiring no experimental structure.^[518,530,531] Searching by similarity to bound conformations, instead of to calculated low-energy conformers,^[515,519,520] can be beneficial.^[532-534] Complementary to SBDD and docking, also shape-similarity methods^[518,535] allow efficient VS^[536] and can identify new chemotypes, which may elucidate SARs,^[518,533] extend chemical accessibility, and improve the prospects for lead optimization.^[537,538]

In this manner, ROCS has been applied previously for the ligand-based identification of PPIMs, *e.g.* of ZipA/FtsZ,^[521] CD4/gp120,^[540] JNK/protein,^[541] and other PPIs.^[542-544] Likewise, a peptide ligand or a PPI structure can help to identify low-molecular ligands^[28] by computational methods,^[527,545-547] such as pharmacophore^[138,548-560] or shape-based VS.^[527,561] Peptide ligands can help to identify binding hot spots^[202] and PPI binding motifs^[550,562] to support receptor-based pharmacophore searches.^[545,563] So far, most cases addressed conventional targets.^[536] A single study used ROCS to directly identify competitive mimetics of a small ($MW = 534$ Da) cyclic peptide, a ligand of integrin receptor $\alpha_v\beta_3$,^[561] that was deduced as a minimum pharmacophore by alanine scanning and deletion experiments.^[564] Using a negative image of a peptide binding pocket as a ROCS shape query^[565] was reported for an undisclosed PPI target.^[527] In contrast, I used ROCS in a novel unprecedented fashion, to identify PPIMs by searching mimetics of hot spots as present in a PPI structure (**Pub. IV**).

^[a] *e.g.* using the OMEGA software;^[515,519,520] ^[b] *e.g.* according to physicochemical properties and functional groups using the FILTER software;^[515] ^[c] SMILES (Simplified molecular-input line-entry system)^[539] arbitrary target specification; ZipA/FtsZ: Z-ring interacting protein A/filamenting temperature-sensitive mutant Z; CD4/gp120: cluster of differentiation 4/envelope glycoprotein 120; ^[d] Gaussian Contraction formula^[515]

2 SCOPE OF THE THESIS

In this thesis, I pursued two key objectives. The first objective was to develop and test a general strategy that uses a prediction of the determinants of PPI to guide the identification of PPIMs. The second objective was to apply this strategy to identify low-molecular inhibitors of the tetramerization of the NHR2 domain of RE.

The original idea, to investigate protein/protein and protein/PPIM interactions in order to rationally identify PPIMs, was inspired by the increasing number of reports of structural and affinity data for PPIs and PPIMs. These cases, as well as previous approaches to identify PPIMs and druggable sites in PPIfaces, are reviewed in **Publication I**.

To address the first objective, I envisioned a general strategy to identify PPIMs. By decomposing the binding energy of a PPI complex into residue-wise contributions (**Chapter 1.5.1**), MM-PBSA reveals hot spots that form the core interactions of that PPI. We hypothesized that such PPI hot spots are also important for PPIM binding, so that predicted hot spots can guide the identification of PPIMs even if only a PPI structure is known.

Using IL-2 as a test case, I validated this concept as described in **Publication II**. Crystal structures of *apo* IL-2, IL-2 bound to five PPIMs, the IL-2/IL-2R α complex as well as affinity data, also for 52 additional PPIMs, were known (**Chapter 1.3**). Here, I revealed that common hot spots are determinants for binding both protein and PPIM ligands at the IL-2 PPIface to IL-2R α . The simultaneous consideration of binding energetics and plasticity allowed predicting a druggable site in this PPIface without using prior knowledge of PPIMs or PPIM-bound conformations. This is remarkable given the absence of a binding pocket in the unbound or protein-bound IL-2 structures. Above all, I could retrieve PPIMs of IL-2 from a large number of non-binders by a docking-based VS that was guided by predicted hot spots and pockets.^[a]

To address the second objective, I used previously predicted hot spots of NHR2 tetramerization from dimers (**Chapter 1.4**). These hot spots in the NHR2 PPIface had been identified using MM-PBSA.^[400] In contrast to wild type NHR2, structure-preserving alanine mutations of these hot spots prevented the tetramerization of NHR2 *in vitro*, the block of myeloid differentiation as well as the self-renewal of hematopoietic progenitor cells, and the induction of leukemia in a murine transplantation model.^[400] This validated these hot spots and established NHR2 inhibition as a strategy for treating RE-AML. Three hot spots clustered at the largest pocket mark two equivalent druggable sites at each end of the elongated PPIface of NHR2.

^[a] Pocket identification and virtual screening were performed in collaboration with Christopher Pfleger.

Hence, it was my objective to mimic and compete with key interactions of these clustered hot spots, as I had observed for PPIMs of IL-2. To this end, I suggested a hot spot-containing α -helical 18mer peptide, which inhibited NHR2 tetramerization. This finding demonstrated for the first time that NHR2 can be inhibited by a molecule that is significantly smaller than the NHR2 domain but carries the hot spots' relevant functional groups. In order to identify non-peptidic PPIMs that also mimic these hot spots, I followed two simultaneous approaches.

The first approach was to design α -helix mimetics, which project side chains in the same manner as an α -helix. Mimicking the arrangement of functional groups that form the hot spots' crucial interactions by equal or bioisosteric groups would effectively result in a competitive NHR2 inhibitor. In **Publication III**, I investigated the potential of teroxazoles as α -helix mimetics and demonstrated that teroxazoles preferably mimic positions i , $i+3$, and $i+6$. In doing so, teroxazoles adopt a minimum energy conformation that was subsequently confirmed by crystal structures of teroxazoles generated by a modular solid-phase synthesis. These results demonstrated that teroxazoles can mimic side chains that project from one face of an α -helix. In doing so, teroxazoles exhibit a new projection pattern, that complements previous α -helix mimetics and are at the same time conveniently more hydrophilic than the latter.

The second approach was to identify mimetics of NHR2 hot spots by VS as described in **Publication IV**. Here, I used the spatial arrangement of pharmacophoric features involved in the key interactions of the three clustered NHR2 hot spots in the context of a 3D similarity search with ROCS (**Chapter 1.5.2**). That way, I performed a ROCS-based VS of $\approx 6 \times 10^6$ molecules from the purchasable subset of the ZINC data base, which was prefiltered to match the physicochemical properties of the NHR2 hot spots and of PPIMs in general as outlined in **Publication I**. By this novel strategy, I effectively enriched hot spot mimetics that are potential inhibitors of NHR2, as was confirmed in subsequent experimental assays.

Finally, experimental testing of hand selected *in silico* hits led to the first drug-like PPIMs that mimic the key interactions formed by the NHR2 hot spots, aim at the PPIface between NHR2 dimers, inhibit their association, and thereby prevent the proliferation of RE-dependent cells. Most importantly, I demonstrated that the here proposed strategy can be the first step in any comparable structure-based endeavor to identify or design PPIMs, even in cases where only a PPI structure with a relatively flat PPIface is known.

3 PPIs: RATIONAL APPROACHES REGARDING DRUGGABILITY, DETERMINANTS, AND MODULATORS (Publication I)

The essential role of PPIs in biology is notorious but insufficiently understood.^[566-572] Recently, the interest in PPIs as drug targets increased^[28,129,135] because their malfunction is a frequent cause of disease.^[85,105,218,573,574] Despite the challenges of inhibiting PPIs, several studies provided a proof of concept that PPIs are valuable targets with therapeutic potential. Here, I reviewed the state-of-the-art in computational PPI analysis and PPIM identification.

3.1 Functional and Structural Aspects of PPIs

First, I summarized the characteristics that discriminate PPIs from conventional protein/ligand interactions and discussed the resulting challenges for drug discovery. In contrast to the small interfaces of conventional targets,^[168-170] many PPIfaces are large^[171,172] and lack preformed druggable binding pockets for low-molecular ligands (Figure 5). Consequentially, it is often hard to identify druggable sites in PPIfaces.^[28] The high affinity and promiscuity of some PPIs^[212,575] further complicates the identification of specific PPIMs.

PPIs need adequate specificity and affinity, so that in a crowded cellular environment only the proper PPIs occur at the right time and condition.^[576,577] Knowing the determinants of this specificity and affinity is crucial to identify PPIMs. In the following, I will outline recurring mechanisms by which PPIs achieve affinity and specificity.^[198]

The hydrophobic effect is a strong driving force for most PPIs.^[332,578-580] Nevertheless, the existence of water-mediated interactions in PPIfaces^[181,581,582] and the elevated propensity of tryptophan, tyrosine, and arginine^[583] highlights the importance of polar interactions.^[327]

Affinity and specificity are achieved by shape and physicochemical complementarity of a ligand and its binding pocket^[341] or of the interacting PPIfaces.^[199] Even if permanent binding-compatible pockets are absent in individual unbound experimental structures, proteins exist as an ensemble of conformational states. This plasticity can allow the opening of transient pockets that bind a ligand as described by the conformational selection model.^[198]

Hot spots are residues that form important interactions in a PPIface and that, if mutated to alanine, cause a drop in binding free energy of $\geq 2 \text{ kcal mol}^{-1}$.^[173,584] Reportedly, hot spots are often spatially clustered, preordered, rigidified, conserved, excluded from solvent by pockets or an O-ring, and complement one another across PPIfaces.^[62,199,212,585-588]

3.2 Determinants of Druggability and Modulators of PPIs

Next, I described how the characteristics of PPIs can be exploited for drug design. The presented applications showed that computational methods are able to successfully predict the druggability of PPIs, their druggable sites, and PPIMs. Moreover, I presented a survey of the PPIMs discovered so far, discussed their drug-likeness, and what can be learned from them.

The druggability of PPIs had traditionally been rated poor due to setbacks in early screening attempts. Besides the structural challenges of PPIs, this failure was likely related to the mismatch between the physicochemical properties of PPIMs and molecules in screening libraries that were assembled following conventional drug-likeness criteria.^[131] On average, PPIMs have a higher *MW*, hydrophobicity, rigidity, and more aromatic moieties than conventional drugs.^[325] My representative survey of known PPIMs confirmed this finding (Figure 2 in **Pub. I**). However, the druggability of at least some PPIs was demonstrated by a number of PPIMs in advanced clinical trials and even some marketed drugs.^[566,589-593]

Notably, PPIs that bind PPIMs often have a pronounced preformed binding pocket (Figure 8).^[28,43,212,574] Such pockets naturally bind exposed substructures of the protein ligand, frequently α -helices, which exist in $\approx 62\%$ of the reported PPI structures.^[237] Often, clustered hot spots, preferably located close to a pocket, form a narrow epitope with a large interface area that forms strong interactions and can thus be addressed or mimicked by PPIMs.^[28,583] For this reason, the (computational) identification of **hot spots and pockets that indicate druggable sites** is a crucial prerequisite to guide the rational identification of PPIMs.^[193,194]

Algorithms for **identifying pockets** were reviewed exhaustively before but mostly dealt with conventional targets.^[45,194,195,594] Geometry-^[595-602] and energy-based algorithms^[603-609] require a structure, but there are also some sequence-based algorithms.^[610-614] However, only few applications considered protein dynamics or plasticity.^[45,247,615-618] Reportedly, geometry-based algorithms are hampered by the flatness of PPIface pockets.^[247] Nevertheless, Q-SiteFinder^[42,605] and PocketFinder,^[609] which cluster favorable interaction points, identified up to 90% of the investigated pockets correctly^[605] including pockets in unbound PPIfaces.^[42] In combination with molecular simulations, newly formed or widened pockets were identified,^[619,620] even those undetectable in unbound PPIfaces, *i.a.* of Bcl-X_L, IL-2, and HDM2.^[247] MD simulations with co-solvent^[244,621] or small probe molecules^[622-624] allowed identifying favorable binding sites and plastic interface regions. A pocket, however, may not be sufficient for PPIM binding^[168] if important interactions, such as those formed by hot spots, are missing.

For **identifying hot spots**, X-ray crystallography^[625,626] and NMR spectroscopy^[56,627-629] are very valuable methods, but provide only limited details about flexibility or mobility.^[630] Tethering,^[388-390,631,632] co-crystallization,^[625,626] SAR by NMR,^[627,628] and alanine scanning^[173,512] can identify hot spots and binding sites of low-molecular and protein ligands. Such data is stored in data bases.^[55,328,584,633-635] However, complementary hot spot predictions suggest the appropriate experiments, thereby enhance efficiency, and improve understanding.

Computational alanine scanning^[512-514] and MM-PBSA^[504] (**Chapter 1.5.1**) are prominent methods for predicting hot spot. The main difference between both methods is that computational alanine scanning predicts $\Delta\Delta G_{bind}$ due to *in silico* alanine mutations whereas MM-PBSA predicts a residue's contribution to ΔG_{bind} without perturbing the structure, thus avoiding relaxation issues. Both methods need a PPI structure, or preferably a conformational ensemble of such, to predict the contribution of each residue to ΔG_{bind} . Different methods are used for conformational sampling, *i.a.* MD^[92,93] or constrained geometric simulations,^[636] and for calculating ΔG , *i.a.* physics-based force fields^[501] and empirical^[637-641] or knowledge-based^[60,641-643] scoring functions. Convincingly, several studies predicted hot spots with high sensitivity and with a good agreement between predicted and known hot spots.^[54,60,92,93,644]

Finally, I illustrated **three case studies**. First, I described how computational analyses and predictions contributed to the identification of PPIMs of HDM2,^[645-649] which has a pronounced binding cleft similar to conventional targets (Figure 8). The remaining case studies dealt with PPIMs of IL-2 and NHR2 and are described in **Publication II** and **IV**.

3.3 Conclusions and Significance

The review of computational methods for PPI analysis, PPIM identification, and self-conducted case studies of their application led me to conclude that:

- The characteristics of PPIs are a challenge for PPIM identification, but known PPIMs demonstrate: PPIs are not undruggable, however PPIMs differ from classical drugs
- Computer-aided drug design can help to identify PPIMs but must be adapted for PPIs
- Hot spots close to transient or permanent pockets indicate druggable sites in PPIfaces

Based on these findings, I proposed a strategy for identifying PPIMs that requires only a PPI structure. The validation (**Pub. II**) and prospective application (**Pub. IV**) of this strategy will be discussed in **Chapter 4** and **Chapter 6**, respectively.

4 INTERLEUKIN-2: HOT SPOTS, TRANSIENT POCKETS, AND PPIM IDENTIFICATION (Publication II)

Although the number of PPIMs is increasing, few of them were found due to rational considerations. Their large PPIfaces,^[171,173] missing pockets,^[172] and stability led to considering PPIs hard to target.^[28] Methods for the computational prediction of druggable sites and ligands of conventional targets have to be adapted for the characteristics of PPIs. Here, I presented a strategy that uses structure-based hot spot and pocket predictions to guide the identification, binding pose prediction, and ranking of PPIMs. IL-2 served as a validation example with known structures and affinities for the binding of IL-2R α and five PPIMs (**Chapter 1.3**).

4.1 A Structure-Based Strategy to Identify PPIMs

Methodologically, the strategy was implemented as follows (Figure 18). Starting from PPI structures, we generated conformational ensembles by (Ia.) MD and (Ib.) constrained geometric simulations with FRODA (framework rigidity optimized dynamics algorithm).^[618] Based on these conformations, we predicted (IIa.) hot spots by MM-PBSA per-residue decomposition of ΔG_{bind} and (IIb.) transient pockets by structural investigations using the

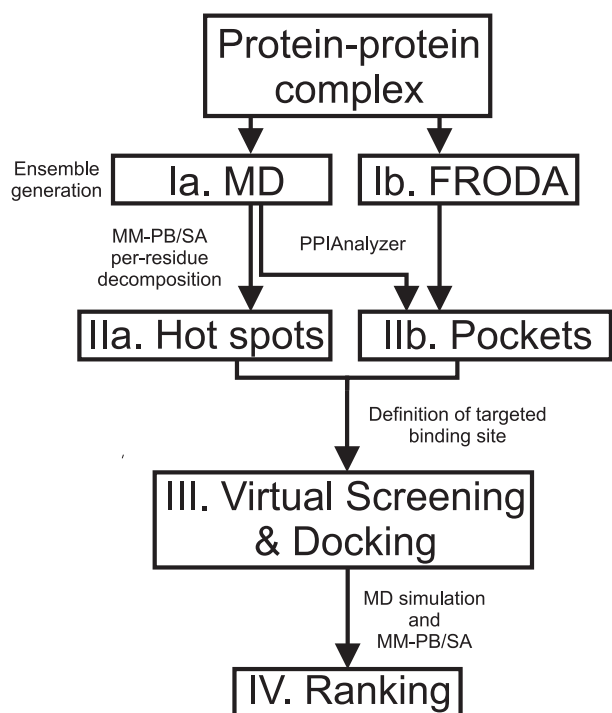


Figure 18. Outline of the presented strategy for hot spot and pocket identification, docking, and ranking of PPIMs using only a PPI structure as the starting point.

PocketAnalyzer program.^[650] (III.) With these hot spots and pockets, we defined the druggable site, guided a VS for PPIMs, and predicted their binding poses by docking. (IV.) Finally, the identified PPIMs were ranked by their MM-PBSA^[501] binding effective energy (ΔG_{eff} , **Chapter 1.5.1**) that was calculated based on the conformational ensembles from MD simulations started from docked binding poses.

To test this strategy at conditions as similar as possible to a “real-life” scenario, I retrieved known PPIMs of IL-2 from a large set of physicochemically similar non-binders (decoys).

4.2 MM-PBSA for the Prediction of Hot Spots

Using MM-PBSA per-residue decompositions of ΔG_{bind} ,^[93] I predicted hot spots of the different IL-2 interactions (Figure 19). To account for mobility and the opening of transient pockets in the PPIface, I performed short ($t \geq 6$ ns) MD simulations of unbound IL-2 and IL-2 bound to either IL-2R α or five PPIMs. These simulations were started from crystal structures using established procedures,^[92] force fields,^[651-653] and the AMBER 9 software.^[92,654]

To compare MM-PBSA variants, I performed MM-PBSA and MM-GBSA calculations using the 1- and 3-trajectory approaches (**Chapter 1.5.1**). Hannes Kopitz therefore extended the decomposition scheme^[92] to consider PB electrostatic solvation free energy.^[506,655]

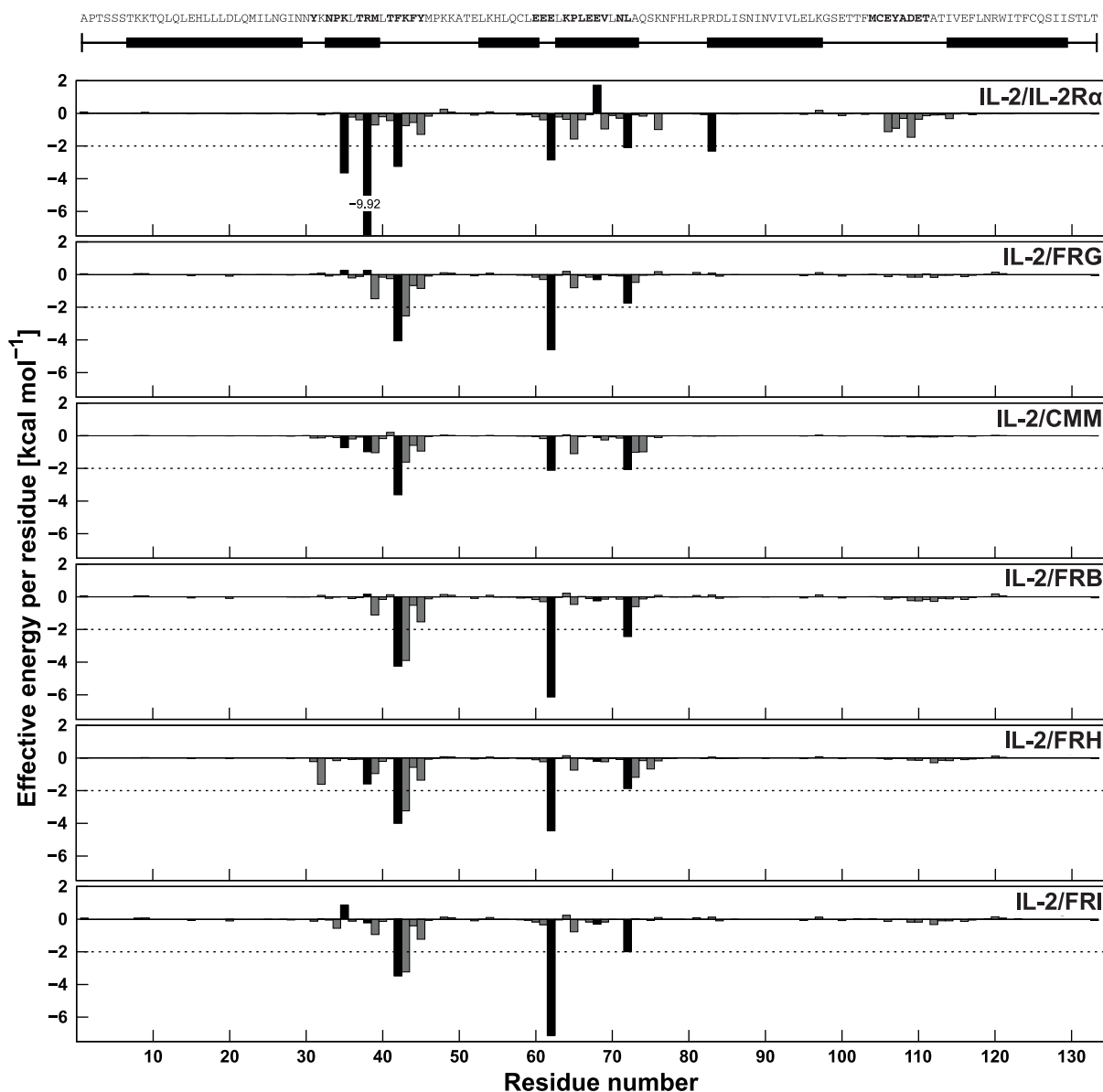


Figure 19. MM-PBSA per-residue contributions to ΔG_{bind} for binding of IL-2R α (top) and five PPIMs to IL-2. Interface residues (bold) and hot spots of IL-2/IL-2R α with $\Delta\Delta G_{bind} \geq 2$ kcal mol⁻¹ are highlighted (black bars).

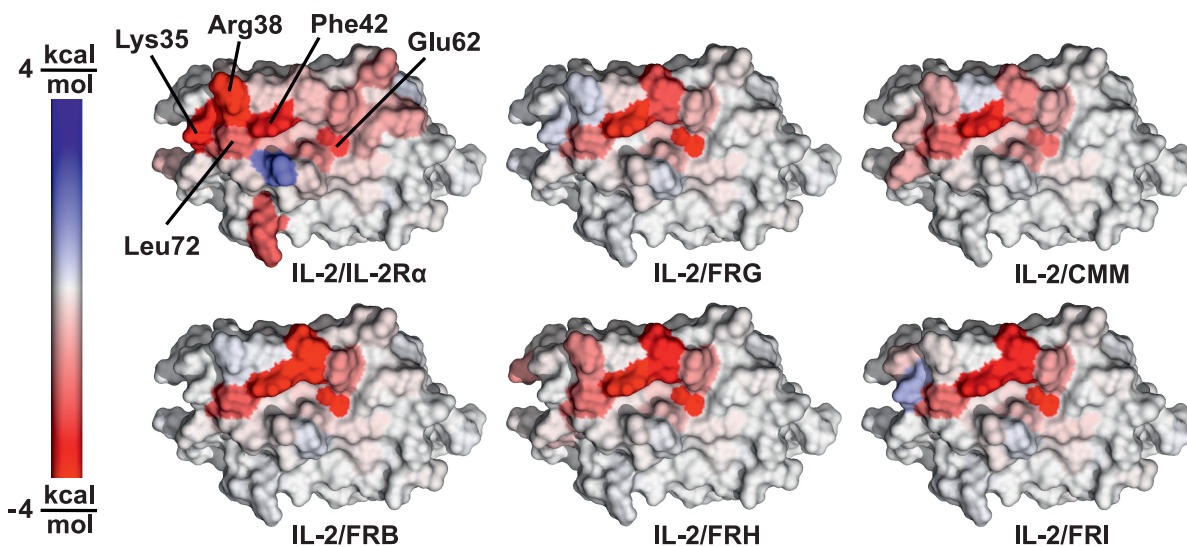


Figure 20. MM-PBSA per-residue contributions to ΔG_{bind} for binding of IL-2R α and five PPIMs to IL-2 (see Figure 19) mapped onto the crystal structure of IL-2 bound to IL-2R α . Color-code has a linear scale.

Notably, ΔG_{bind} was stable over the simulation time if calculated with the 1-trajectory approach, but not for the 3-trajectory approach. The reason was a drift of G_{tot} for the short MD simulations, as had been observed earlier.^[92,93,656-659] Extracting unbound conformations from the complexes with the 1-trajectory approach led to a drastic reduction of the statistical uncertainty of the free energy components.^[54,93,660] Similarly, the drift of G_{tot} cancelled out in the calculation of G_{bind} (Eq. 2 in **Chapter 1.5.1**). This suggested that the 1-trajectory approach is suited for identifying hot spots even with short MD simulations.

Using the MM-PBSA 1-trajectory approach, I identified five hot spots of the IL-2/IL-2R α interaction (Lys35, Arg38, Phe42, Glu62, and Leu72) that agreed with results from mutation experiments.^[373,382] One remote hot spot (Arg83) was neglected for guiding the subsequent docking. However, three of the clustered hot spots were also important for binding the five PPIMs (Figure 19). In doing so, the PPIMs mimicked interactions that occur between IL-2 hot spots and IL-2R α . The identified hot spots cluster together (Figure 20) and form a functional epitope that covers only $\approx 20\%$ (500 Å²) of the PPIface area. This suggested that hot spot predictions based on a PPI structure can define druggable sites and guide PPIM identification.

4.3 Prediction of Transient Pockets

Analyzing the PPIfaces of non-PPIM-unbound IL-2 conformations confirmed the absence of a binding pocket. However, crystal structure suggested that PPIMs bind to IL-2 *via* a conformational selection mechanism.^[250] Thus, we set out to identify interface pockets in conformational ensembles generated by MD and FRODA simulations that started from unbound IL-2.

To this end, Christopher Pfleger developed the PPIAnalyzer method, which uses the geometry-based PocketAnalyzer algorithm,^[596,650] which identifies interface pockets in structurally diverse conformations with high stereochemical quality. He found that conformations from FRODA simulations came closer to the PPIM-bound crystal structures and had more adequately located transient pockets compared to the conformations from MD simulations.

We attribute the success of FRODA to an appropriate rigidity-based coarse-graining algorithm,^[618] which identified residues not involved in pocket opening. Thus, motions were focused to residues that are involved in pocket opening, such as the gate keeper hot spot Phe42.^[243,373] In the MD simulations, friction and the exposure of a hydrophobic pocket to explicit solvent impeded the opening of an adequately located pocket.

For the subsequent docking experiments, we selected representative conformations with the 10 largest pocket volumes for each of the MD- or FRODA-generated ensembles.

4.4 Binding Pose Prediction by Docking

The five PPIMs with known bound crystal structures were docked into each selected conformation. The docking was performed using established procedures (AutoDock 3.05^[661] with DrugScore^[662,663]). To exclude any bias due to the knowledge of the IL-2/PPIM crystal structures, we placed the potential grids for docking solely based on (I) the five identified clustered hot spots and (II) the amino acids that lined the identified interface pockets. A

docking was considered successful if the ligand pose in the largest cluster with the lowest intermolecular docking energy had an *RMSD* < 2.0 Å to the native pose.

As a control, re-docking into PPIM-bound crystal structures was successful in all cases while *apo*-docking failed due to the absent pocket. Docking into the MD-generated conformations also failed because our unbiased method selected misplaced pockets. In contrast, docking into at least one selected FRODA-generated conformation was successful for four out of five PPIMs (Figure 21).

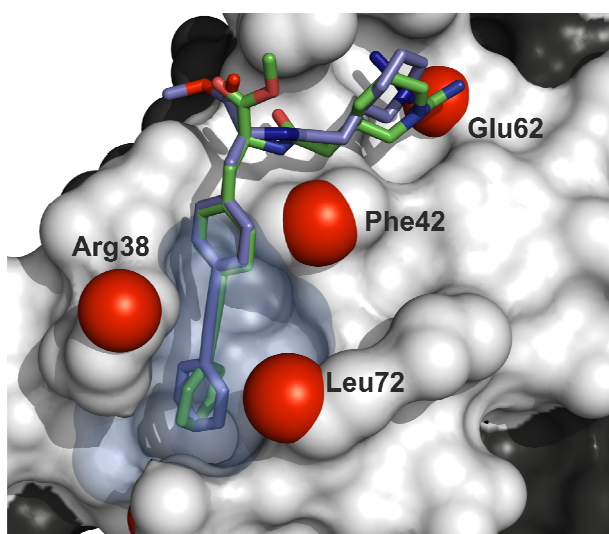


Figure 21. Predicted binding pose of a PPIM (FRG, blue sticks) docked into a FRODA snapshot containing an identified transient pocket (blue volume) close to the hot spots (red). The *RMSD* between the predicted and crystallographic binding pose (green sticks) is 1.28 Å.

In contrast to re-docking, docking into unbound *apo* conformations usually leads to deteriorated docking poses^[664] depending on the degree to which a protein conformation changes upon ligand binding.^[665,666] Thus, being able to start from the *apo* IL-2 structure and identify transient pockets in trajectories of computationally inexpensive FRODA simulations that are adequate for ligand docking is a valuable achievement.

4.5 MM-PBSA for the Ranking of Binding Poses

Based on **poses from crystal structures and docking**, I predicted ΔG_{bind} of the five PPIMs using the MM-PBSA 1-trajectory approach. The PPIMs' similarity justified neglecting conformational strain and configurational entropy changes upon binding. The calculated binding effective energies ($\Delta G_{eff,calc}$) were stable throughout the MD simulations.

My MM-PBSA calculations based on **IL-2/PPIM crystal structures** correctly ranked four out of five PPIMs. The linear correlation of computed and experimental ΔG_{bind} was good and significant ($R^2 = 0.81$; $p < 0.05$; Figure 22). This was particularly remarkable because the range of experimentally determined binding affinities was only 3 kcal mol⁻¹. So far, this was one of only a few successful applications of MM-PBSA for the ranking of PPIMs.^[667,668]

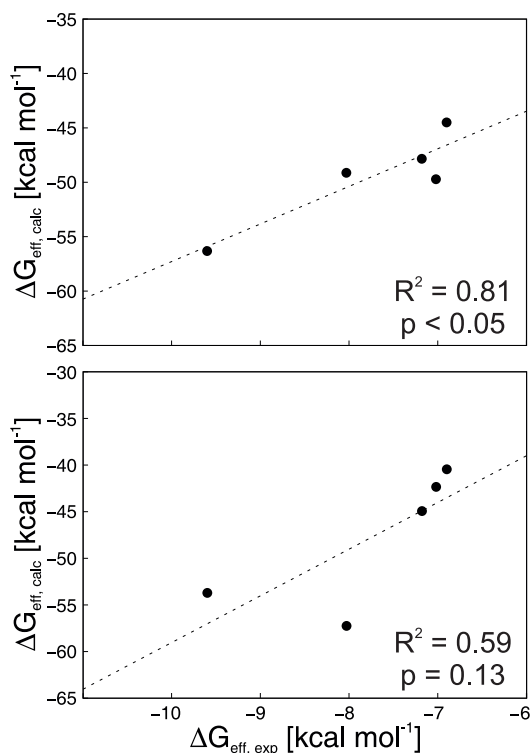


Figure 22. Correlation of experimentally determined free enthalpies of binding ($\Delta G_{eff,exp}$) and effective energies computed by MM-PBSA ($\Delta G_{eff,calc}$) based on crystal structures (top) or docked binding poses (bottom).

MM-PBSA calculations for **docked IL-2/PPIM structures** were based on selected poses from docking into FRODA-generated conformations, where the PPIM formed a salt bridge with hot spot Glu62. The linear correlation of computed $\Delta G_{eff,calc}$ and experimentally determined $\Delta G_{eff,exp}$ was fair and weakly significant ($R^2 = 0.59$; $p < 0.13$). Larger deviations of $\Delta G_{eff,calc}$ were observed for PPIMs with large structural deviations between docked and native binding poses. These results demonstrated that MM-PBSA calculations require at least good starting structures ($RMSD < 2.5$ Å) to allow ranking PPIMs correctly. Still, the quality of the generated docking poses was sufficient to successfully discriminate subgroups of high- and low-affinity ligands (Figure 22).

4.6 Docking Enrichment in a Large Set of Decoys

To show that my prediction of hot spots close to a pocket can guide the identification of PPIMs, I performed a retrospective VS for PPIMs of IL-2. To this end, I retrieved known PPIMs from a large set of decoys using the same docking procedure as for pose predictions.

As known binders, I selected the five PPIMs from bound crystal structures (Figure 14) and 52 PPIMs with known IC_{50} .^[382-384] As decoys, I selected 996 molecules from the “purchasable subset” of the ZINC data base^[669] following the *directory of useful decoys* (DUD)^[670] procedure. This procedure maximized the physicochemical similarity between known binders and decoys in order to prevent any bias due to trivial dissimilarities, such as of $\log P$, MW , the number of rotatable bonds, hydrogen bond donors/acceptors, and specific functional groups.^[670] Still, the known PPIMs were optimized for binding to IL-2 and partially violate Lipinski’s rules^[131] so that the properties of PPIMs and decoys did not match perfectly. To correct the often observed size-related bias to the ranking (Figure 23), I divided each ligand’s docking score by the square root of its MW .^[671]

Docking the 57 (5) PPIMs (with known bound crystal structures) into the 10 largest FRODA-generated pockets led to good enrichment factors at 1% of the ranked datasets of $EF_1 = 13.6$ to 23.8 ($EF_1 = 37.2$ to 92.2) and receiver operating characteristics (ROC) with area under the curve values of $AUC \geq 0.89$ ($AUC \geq 0.93$). Thus, I used predicted hot spots and transient pockets to successfully guide the identification of PPIMs in a large set of decoys.

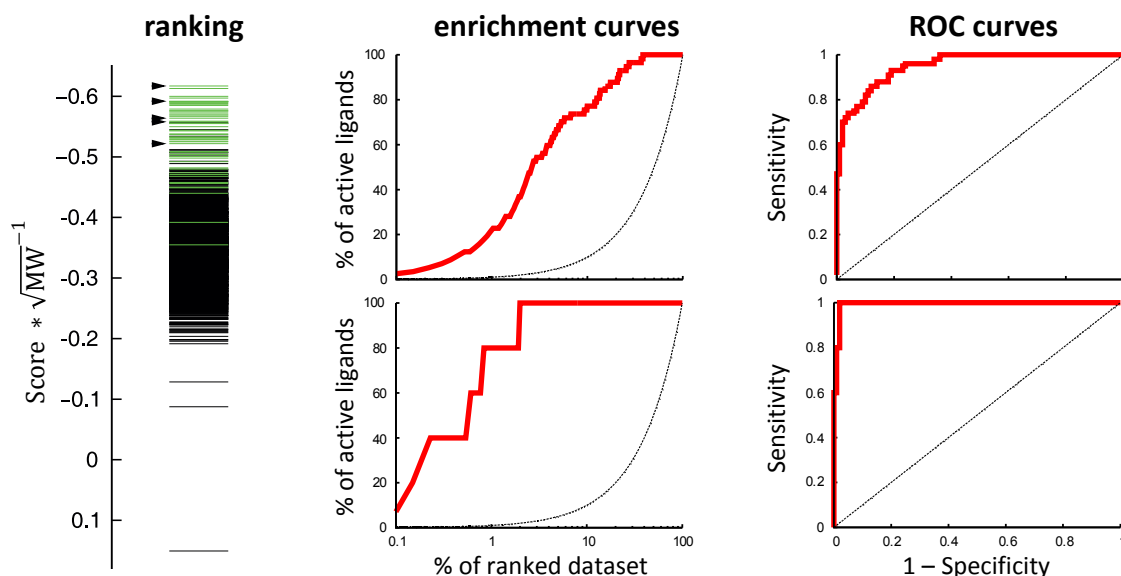


Figure 23. Performance of docking PPIMs and decoys into one of the largest FRODA-generated pockets. In the ranking (left), the 57 PPIMs (green) and the five PPIMs with bound crystal structures (►) ranked higher than the decoys (black). In the enrichment (center) and ROC curves (right, red), the 57 PPIMs (top) and the five PPIMs with known bound crystal structures (bottom) were selected much earlier than with a random selection (black).

4.7 Conclusions and Significance

In this study, I presented the first generally applicable computational strategy that simultaneously considers aspects of energetics and plasticity in order to identify PPIMs that bind to a PPIface. Predicted hot spots and transient pockets guided the identification of PPIMs starting from only a PPI structure. The major results of my computational investigations are:

- Calculated per-residue contributions to the effective ΔG_{bind} from MM-PBSA revealed that most hot spots of the IL-2/IL-2R α interaction are also hot spots of PPIM binding
- A geometry-based algorithm revealed transient pockets that formed during molecular simulations started from a pocketless non-PPIM-bound IL-2 structure
- Docking into transient pockets, guided by hot spots, led to correct PPIM binding poses and allowed identifying PPIMs in a large set of physicochemically similar decoys
- Ranking docked poses by MM-PBSA discriminated high- from low-affinity PPIMs

To mimic a “real-life” scenario, I did not use any knowledge of the PPIMs or PPIM-bound conformations for other purposes than comparing them to my predictions. Because my strategy for identifying PPIMs requires only a PPI structure, it will be applicable in a prospective manner in a wide variety of cases. It can thus be the first step in any comparable structure-based endeavor to identify PPIMs. To sustain this claim, I set out to identify the first inhibitors of NHR2 tetramerization. For this purpose, I pursued two complementary strategies simultaneously. On the one side, I analyzed the capability of teroxazoles to act as a new class of hydrophilic α -helix mimetics (**Chapter 5**). This was also encouraged by the location of three clustered NHR2 hot spots on a contiguous α -helix.^[400] On the other side, I performed a prospective structure-based VS for small molecules that mimic the arrangement of relevant functional groups of these three clustered hot spots (**Chapter 6**).

5 TEROXAZOLES: A NEW HYDROPHILIC α -HELIX MIMETIC SCAFFOLD (Publication III)

The majority of the PPI complexes in the PDB (62%) have an α -helix in their PPIface.^[237] Such α -helices often present linearly arranged hot spot side chains that project from one of their faces into a cavity of the interacting protein (Figure 13). Imitating such an arrangement of side chains by α -helix mimetics^[184,238-242] is a frequently pursued strategy for inhibiting α -helix mediated PPIs.

5.1 Computational Investigation of Teroxazoles

A review of α -helix mimetics scaffolds^[242] revealed that they are still limited regarding synthetic access to diverse substitutions, available projection patterns, and hydrophobicity. Most α -helix mimetics present substituents on linearly connected and often coplanar aromatic rings. By computational investigations, I found that teroxazoles suggested by M. W. Göbel can act as polar α -helix mimetics with a novel projection pattern mimicking α -helix positions i , $i+3$, and $i+6$. C. Pinto Gomes and M. W. Göbel performed the modular solid-phase synthesis of diversely substituted teroxazole derivatives as a new class of α -helix mimetics.

Using MM-PBSA,^[504] (Chapter 1.5.1) I calculated the effective energy of teroxazole **1** as a function of the inter-ring torsion angles ν and ω (Figure 24). In compliance with the general AMBER force field, I parameterized torsion angle potentials and atomic partial charges of **1** based on quantum chemical computations.^[653] This revealed a preference of the coplanar orientation of the oxazole rings in agreement with crystal structures solved by J. W. Bats.

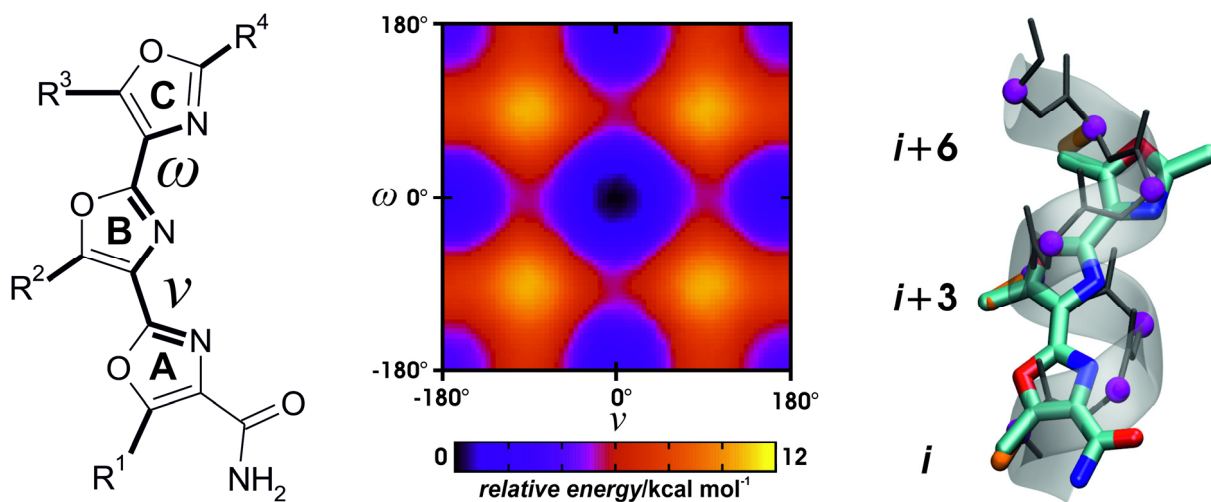


Figure 24. Markush structure of the teroxazole scaffold with highlighted inter-ring torsion angles ν and ω (left). Relative MM-PBSA effective energies of teroxazole **1** (R^1 - R^4 = CH₃) as a function of ν and ω (center). Superimposition of C $_{\beta}$ atoms of side chains at positions i , $i+3$, and $i+6$ of a canonical α -helix onto the corresponding substituent atoms of R¹-R³ of **1** in a low-energy conformation (ν = 0°, ω = 0°).

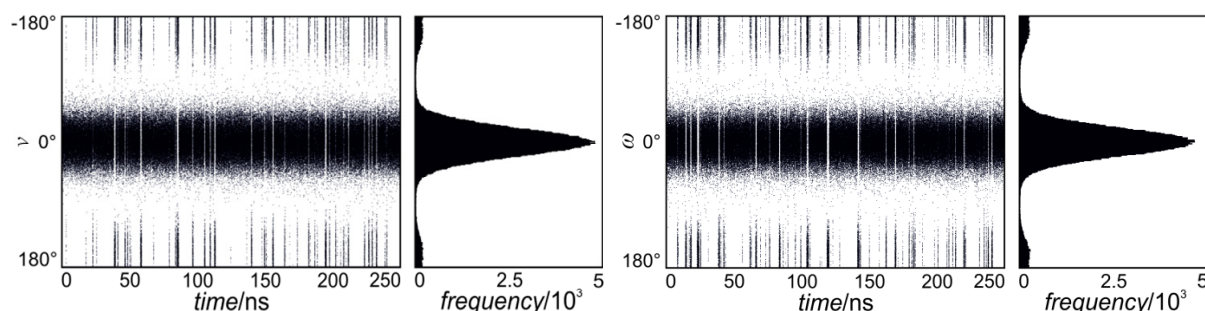


Figure 25. Torsion angles ν (left) and ω (right) of **1** as a function of simulation time and as histograms.

The substitution sites of the minimum effective energy conformer superimpose excellently with C_β atoms at helix positions i , $i+3$, and $i+6$ ($RMSD = 0.35 \text{ \AA}$), covering a broad region on one face of an α -helix that until then had not been addressed by other α -helix mimetics.^[672,673] Rotating one or both torsion angles by 180° increased the effective energy by ≈ 2 or 4 kcal mol^{-1} due to a less favorable solvation of antiparallel oxazole dipoles. These flipped ring conformations, however, may also use R^4 and mimic helix positions i , $i+2$, $i+6$, and $i+7$.

During a 250 ns long MD simulation of teroxazole **1**, the rings rotated repeatedly (Figure 25) in agreement with a rotational barrier of $\approx 5.5 \text{ kcal mol}^{-1}$ calculated by MM-PBSA. The rings deviated by up to 50° from coplanarity, corresponding to a calculated energetic cost of $\approx 3.0 \text{ kcal mol}^{-1}$. This limited flexibility improves the ability to project side chains with distances and angles as they are found in α -helices and allows induced-fit upon binding.^[674]

Finally, computed $\log P$ values^[675] suggested that teroxazoles are more polar ($\log P = 0.34$) and may offer increased solubility compared to the popular terphenyl^[676-679] ($\log P = 3.34$) and oxazole-pyridazine-piperazine^[680,681] ($\log P = 0.96$) α -helix mimetics.

5.2 Conclusions and Significance

The major results of my investigation of teroxazoles as potential α -helix mimetics are:

- Teroxazoles can mimic α -helix positions i , $i+3$, and $i+6$ without conformational strain
- The limited flexibility of the inter-ring torsion angles of the otherwise coplanar oxazoles allows adjusting the arrangement of substituents for optimal α -helix mimicry
- Computed $\log P$ values show teroxazoles are polar and may offer increased solubility

This led to the synthesis of diversely substituted teroxazoles as hydrophilic α -helix mimetics, the first to mimic positions i , $i+3$, and $i+6$. Varied linker length may allow mimicking NHR2 hot spots in positions i , $i+3$, and $i+7$, but has not been tested. Importantly, however, my computational analysis of a chemical scaffold's capability to present side chains like an α -helix can serve as a general approach to investigate the conformational properties of new PPIMs.

6 RUNX1-ETO: IDENTIFICATION OF THE FIRST INHIBITORS OF NHR2-MEDIATED OLIGOMERIZATION (Publication IV)

The RE fusion protein, result of translocation t(8;21), connects the DNA-binding Runt domain of RUNX1 and most of ETO including its NHR2 homotetramerization domain (**Chapter 1.4**). Hot spots close to the largest pocket,^[400] predicted just as in **Publication II**, constitute an **essential structural motif** in the NHR2 PPIface (Figure 26). Alanine mutations of these hot spots had demonstrated that NHR2 tetramerization from dimers is essential for the onset and maintenance of RE-AML, thus validating the NHR2 PPIface as a target.^[400] Here, I presented a general strategy and its prospective application for the identification of PPIMs that mimic clustered hot spots of NHR2 in order to inhibit its tetramerization.

6.1 Structure-Based Identification of NHR2 Inhibitors

Structure-based approaches have helped identifying PPIMs,^[196,682] but mostly started from peptide- or ligand-bound structures. In contrast, I identified NHR2 inhibitors based on the essential structural motif as present in the NHR2 tetramer (Figure 26).^[63] The hot spots D533, E536, and W540 were chosen as the most suitable **template motif** for a VS for the following reasons: these hot spots cover one face of each α -helix of an NHR2 dimer, project side chains in sequence positions i , $i+3$, and $i+7$ just as known α -helix mimetics,^[240,683,684] are clustered

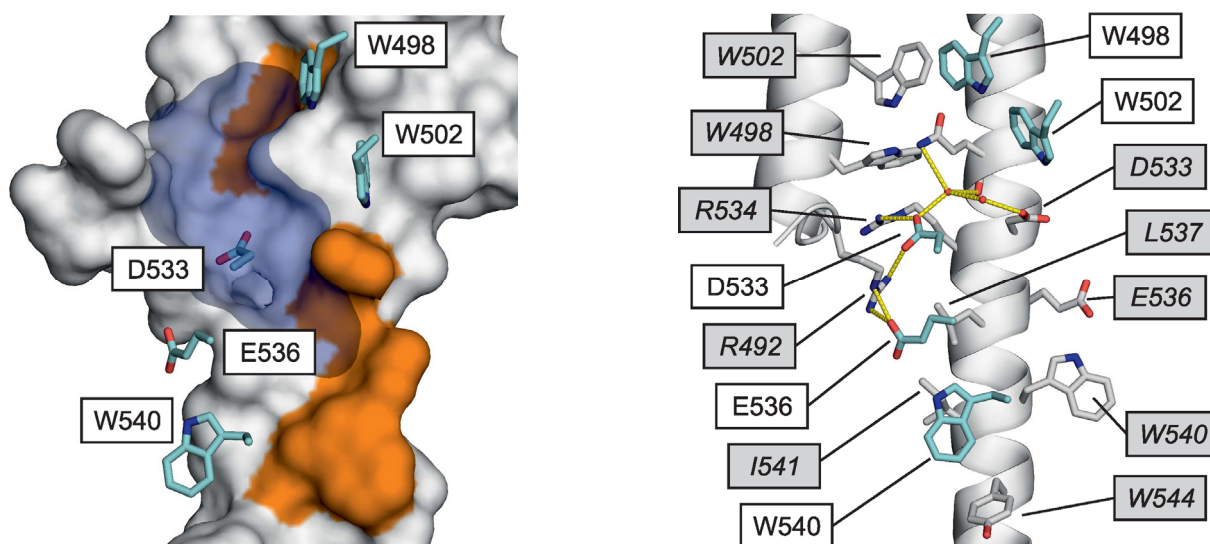


Figure 26. Essential structural motif in the NHR2 interface. Left: Five hot spots on one dimer^[400] (cyan sticks) bind to the targeted NHR2 dimer (surface; PDB code: 1WQ6^[63]). Hot spots on the targeted PPIface are highlighted (orange). Hot spot D533 extends into the deepest pocket (transparent blue; identified by PocketAnalyzer^[650]). Right: The polar (yellow) and hydrophobic interactions of three hot spots form the template motif used for VS (D533, E536, and W540) at the targeted NHR2 interface (cartoon). All five hot spots (cyan sticks), important residues on the targeted NHR2 interface (grey sticks, italic labels with grey background), and two water molecules mediating interactions between D533 and the targeted NHR2 interface are highlighted.

sufficiently close to allow mimicry by PPIMs (**Pub. I**), and extend into the deepest interface pocket (D533). The mix of charged/acceptor (D533, E536) and aromatic/donor (W540) interactions should lead to PPIMs with lower hydrophobicity than common α -helix mimetic scaffolds^[239,240,683] and higher specificity than provided by purely hydrophobic interactions.

Based on this template motif, I suggested the α -helical 18mer peptide **P1** as an NHR2 inhibitor. **P1** was derived from the NHR2 domain (Figure 1 in **Pub. IV**) and comprises the template motif. Convincingly, chemical cross-linking experiments revealed that **P1** inhibits the NHR2 tetramerization with $IC_{50} \approx 250 \mu\text{M}$ (Figure 2 in **Pub. IV**). This moderate IC_{50} reflects the competition of **P1** against the strong tendency of NHR2 dimers to self-associate.^[400] Still, **P1** provided the first proof-of-principle that the NHR2 tetramerization can be inhibited by a molecule significantly smaller than the NHR2 domain itself. Replacing the hot spots in **P1** by alanines abolished the inhibitory effect, confirming the hot spots' importance. Thus, I used the known location and orientation of these hot spots' side chains as present in the NHR2 crystal structure for a VS of PPIMs that mimic the hot spot interactions.

To this end, I designed three queries (Figure 27) containing (I) the carboxylic and indole functional groups of D533, E536, and W540, (II) the functional groups from (I) decorated on a terpyridine α -helix mimetic scaffold^[684] that overlaps with the helix that should be replaced, or (III) the functional groups from (I) plus custom repulsive Gaussian potentials centered on adjacent target atoms ($\leq 4 \text{ \AA}$ distance). These repulsive Gaussian potentials were implemented as a modification of the color force field to penalize molecules that clashed with the PPIface. With these queries, I searched the refined purchasable subset of the ZINC 11 data base^[669] ($\approx 6 \times 10^6$ molecules) using ROCS^[515] (**Chapter 1.5.2**). ROCS superimposes pre-

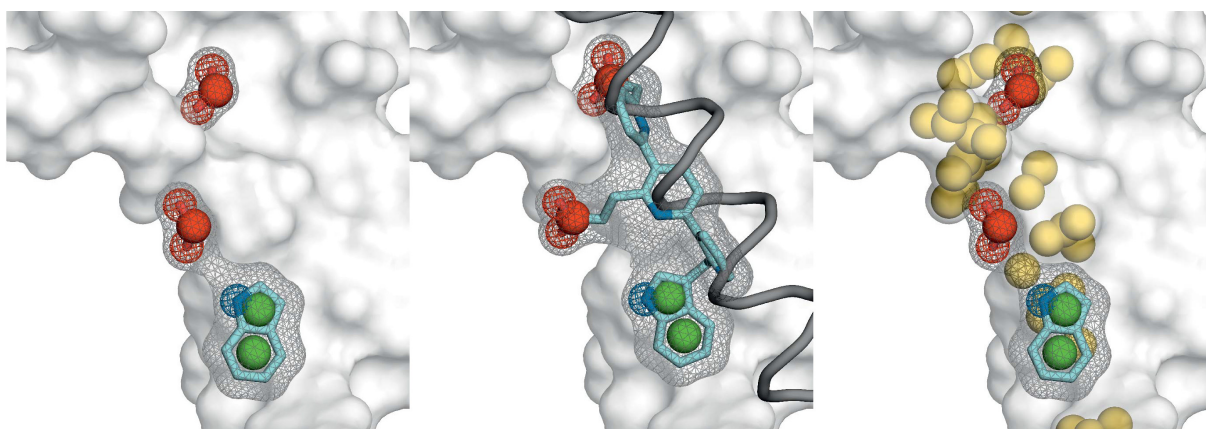


Figure 27. The three ROCS queries. Query I (left), query II (center), and query III (right) are explained in the main text. The used ROCS colors are highlighted as follows: donor (blue mesh), acceptor (red meshes), anionic (red spheres), ring (green spheres), repulsive Gaussian potentials (golden spheres). The functional groups of the template motif and the terphenyl scaffold (cyan sticks) are enclosed by a shape volume (grey surface mesh). The targeted NHR2 interface is depicted for clarity (transparent surface) and is not part of any ROCS query.

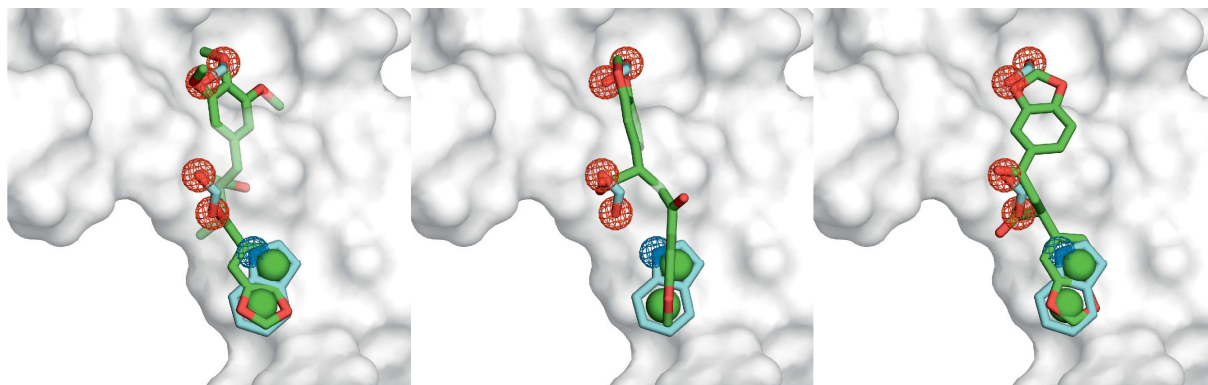


Figure 28. Exemplary ROCS poses of the most active inhibitors. ROCS poses of **7.18** generated with query III (left), of **7.44** generated with query I (center), and of **7.44** generated with query III (right).

generated^[515] conformers based on the spatial overlap of molecular shape and pharmacophoric color features. The ZINC data base was refined by selecting drug-like (OpenEye default filter^[515]), non-positively charged, purchasable ZINC entries with a $MW \leq 650$ Da. That way, I accounted for the twofold negative charge and large stretch (14–16 Å) of the template motif.

To also account for conformational variability, I performed two ROCS searches for each query (I–III) that used the different hot spot configurations ($RMSD = 0.3$ Å) on either end of the NHR2 dimer (Figure 29). To favor compounds that mimicked the hot spots' functional groups, I re-ranked all ROCS poses by their color overlap. For each query, I combined the top scoring ZINC entries obtained for both hot spot configurations and removed duplicates.

The resulting ROCS poses vividly illustrate the overlap of each ZINC entry and the hot spots' functional groups (Figure 28). We assessed the 1000 top-scoring ROCS poses for each query visually with respect to interaction complementarity, steric overlap with the PPiface as well as the competing α -helix, conformational strain, and structural diversity. That way, we selected 80 compounds (Table S1 in **Pub. IV**) for experimental testing.^[a]

ELISA and ABCD assays^[b] revealed seven compounds with consistent and selective inhibition in both assays (Table S2 in **Pub. IV**) and activities comparable to **P1** (42.8% inhibition at $c = 500$ μ M and $IC_{50} = 390 \pm 30$ μ M by ELISA). The most promising molecule **7.18** (Figure 29) exhibited an NHR2 inhibition of 80.3% ($c = 1$ mM) in the ABCD assay and of 47.7% ($c = 2$ mM) in the ELISA assay but was inactive in cells. The PPIMs' selectivity for NHR2 was demonstrated by a reduced inhibition when the NHR2 domain of the reporter construct was replaced with the BCR tetramerization domain (Figure S2 of **Pub. IV**). This finding also demonstrated that the NHR2 inhibition was not caused by non-specific binding.

^[a] All non-computational experiments of **Publication IV** were performed by Julia Schanda, Christian Wichmann, and Manuel Grez at the INSTITUTE FOR BIOMEDICAL RESEARCH GEORG-SPEYER-HAUS (Frankfurt). Experimental results are presented in Figure 2, Figure 3, and the Supplementary Information of **Publication IV**.

^[b] ELISA: enzyme-linked immunosorbent assay; ABCD assay: avidin biotin complex DNA assay

By a fingerprint-based similarity search,^[685] I identified **7.44**, a structural analog of **7.18** (Figure 29). The improved activity of **7.44** ($IC_{50} = 630 \pm 24 \mu\text{M}$ by ELISA) was only 1.6-fold lower than that of **P1**. Notably, Hill coefficients of **7.44** and **P1** of 1.8 and 1.9 show that only after binding a first PPIM, the second, equivalent binding site on the NHR2 dimer becomes accessible more easily. This cooperative effect is in accord with a mathematical model of the NHR2 equilibrium (**Chapter 9.1**). In addition, **7.44** selectively reduced the viability of RE-dependent human SKNO-1 cells ($EC_{50} < 10 \mu\text{M}$) in agreement with the cellular effect caused by expressing an inhibitory NHR2-containing protein construct.^[402] Negative controls with an *in vitro* inactive molecule or of RE-independent U937 cells with **7.44** showed no such effects.

Structurally, **7.18** contains a central 4-oxobutanoic acid scaffold substituted by a 4-(1,3-benzodioxolyl) and a 2-(3,4,5-trimethoxy-benzyl) moiety (Figure 29). The ROCS poses show (Figure 28) that the carboxylate group of **7.18** mimics that of E536, while the benzodioxolyl and trimethoxybenzyl groups of **7.18** mimic the indole of W540 and the carboxylate of D533, respectively. The similar ROCS pose of **7.44** clashed less with the PPIface, thus explaining its improved activity. Moreover, replacing the trimethoxybenzyl of **7.18** by the shorter, more compact, and less flexible second benzodioxolyl of **7.44** is entropically favorable. Finally, ROCS suggested preferred binding of the 2*S* stereoisomers of **7.18** and **7.44** and an additional reversed binding mode for **7.44**. Crystal structures^[225,686] and data from the SwissBioisostere data base^[687] confirmed that carboxylate and indole can be mimicked by 1,3-benzodioxoles.

The predicted binding modes of **7.18** and **7.44** are good starting points for lead optimization. This is because of the simple chemical structures of **7.18** and **7.44**, their low *MW* ($MW_{7.18} = 415 \text{ Da}$; $MW_{7.44} = 341 \text{ Da}$), and a ligand efficiency on a par with PPIMs employed in the clinics^[1,28] ($LE_{7.44} = -0.18 \text{ kcal mol}^{-1}$ related to IC_{50}).

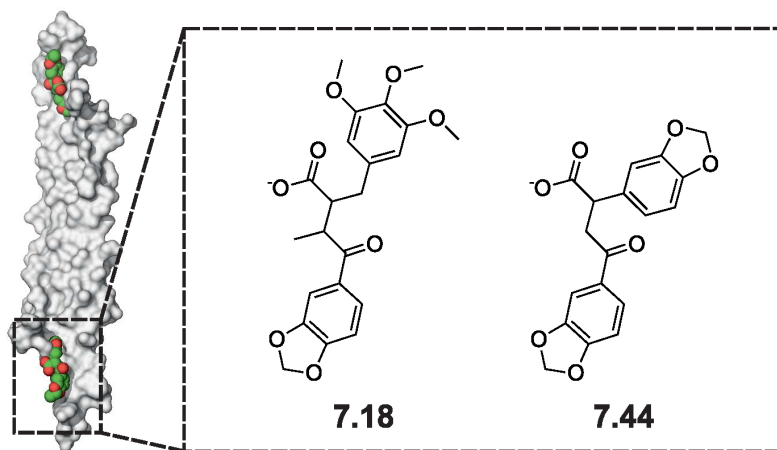


Figure 29. The most potent NHR2 inhibitors **7.18** and **7.44** aim at the two equivalent binding regions in the PPIface between two NHR2 dimers.

Finally, I discussed possible alternative mechanisms that can explain the effects of our PPIMs. Kaould *et al.* excluded unspecific effects of **7.44** due to aggregation or redox cycling but showed that **7.44** inhibits JNK and affects the JNK-pathway in cells.^[541] This may explain why **7.44** affected RE-dependent SKNO-1 cells with an EC_{50} that was lower than the IC_{50} of the *in vitro* inhibition of NHR2. In turn, I suggested that NHR2 inhibition may prevent the RE-mediated activation of the JNK-pathway,^[688,689] as suggested by the absence of c-Jun up-regulation by RE with hot spot alanine mutations.^[400] Although **7.44** is similar to (epi)podophyllotoxins, non-intercalating inhibitors of topoisomerase II^[686] and microtubule formation,^[690] the absence of these targets in our *in vitro* assays contradicts such a mode of action.

6.2 Conclusions and Significance

Hot spot and pocket prediction^[650] (**Pub. II**) had previously revealed a structural motif that is essential for NHR2 tetramerization from dimers and allowed to validate NHR2 as a target.^[400] On this basis, I suggested peptide **P1** that contained three clustered hot spots and inhibits NHR2 tetramerization from dimers. This cluster of hot spots that are located close to the largest pocket in the NHR2 PPIface was used as a template motif for the structure-based VS of hot spot mimicking NHR2 inhibitors. The major results of my investigations are:

- Using computational predictions, I identified the first low-molecular NHR2 inhibitors, which mimic clustered hot spots in the NHR2 tetramerization interface
- Seven out of 80 tested compounds selectively inhibited NHR2 tetramerization *in vitro*
- Using a fingerprint-based similarity search for analogs of the most potent primary hit (**7.18**), I identified an even more potent NHR2 inhibitor (**7.44**) that selectively also inhibited the proliferation of RE-dependent cells at low micromolar concentration

In summary, my computational predictions led to the first NHR2 inhibitors. These can be used as, or developed into, diagnostic agents or therapeutics for a personalized therapy of RE-AML. For this purpose, only 80 molecules that were hand selected from the top ranked 0.05% of a total of $\approx 6 \times 10^6$ molecules had to be tested. Two factors were essential to this success. First, we integrated complementary computational predictions and experimental validations in every step. Second, we predicted hot spots and pockets to guide the identification of PPIMs, a strategy that I had validated in **Publication II**.

Most importantly, however, these results demonstrated that my computational strategy can identify PPIMs in cases where nothing else than a PPI structure is known. Even for the still rather shallow and narrow PPIface of NHR2, specific PPIMs were identified. Hence, this strategy can be the first step in any comparable structure-based endeavor to identify PPIMs.

7 SUMMARY AND PERSPECTIVES

In this thesis, I developed and validated a general strategy that uses the prediction of hot spots and pockets based on a PPI structure to guide the computational identification of PPIMs. As variations of a common theme, docking identified PPIMs that **address** predicted hot spots while ROCS identified PPIMs that **mimic** them. Using the latter variant, I identified the first *in vivo* active low-molecular inhibitors of the leukemia associated tetramerization of NHR2.

Recently, I developed a structure- and ligand-based consensus strategy to identify new NHR2 inhibitors, *i.a.* analogs of **7.44**. Ranking a large screening library ($\geq 10^5$ compounds) by this strategy suggested 600 further compounds; preliminary experimental evaluations show three low-micro-molar NHR2 inhibitors. Ongoing experiments aim at discovering a potent lead and will reveal the efficiency of the ROCS-based VS (**Pub. IV**). A mid-term goal is to resolve inhibitor-bound NHR2 structures to aid SBDD and lead optimization.^[691] To improve our understanding of the basis of affinity at the NHR2 PPIface, I suggested variants of **P1** based on affinity-maturated NHR2 mutants.^[692] Ultimately, NHR2 inhibitors can serve as tools to study NHR2-dependent effects and as therapeutics for a personalized therapy of RE-AML. Already, NHR2 inhibitors are used to study RE-dependent epigenetic regulation.^[a]

With regard to my analysis of teroxazoles (**Pub. III**), their suggested action as α -helix mimetics (**Pub. III**), in particular as NHR2 inhibitors, remains to be validated experimentally. Beyond that, however, similar analyses have already led to the synthesis of trispyrimidonamides as another promising new class α -helix mimetics.^[693]

As emphasized throughout this thesis, I presented a general strategy for VS of hot spot mimetic PPIMs that can be applied even if only a PPI structure is known (**Pubs. II and IV**). Only recently, this has been backed by analyses of the disease-related homooligomerization of human heat shock protein 90 (HSP90). Similar to my studies of IL-2 and NHR2, hot spot predictions revealed a druggable site in the HSP90 PPIface. In this case, hot spot predictions based on a homology model were later confirmed by similar predictions based on a new crystal structure and mutation experiments.^[694,695] The implied usability of homology models, simplified access to free energy calculations by a new workflow tool,^[503] and knowledge-based scoring functions for predicting PPI hot spots^[60] further extended the application area.

Most importantly, my investigations point out a rational access to understanding PPIs and PPIMs on a molecular level, to predict PPIMs, and to efficiently assist experimental investigations of a large number of PPI targets that were previously thought undruggable.

^[a] Manuel Grez *et al.*, unpublished results.

8 ACKNOWLEDGMENTS

First and foremost, I thank my supervisor Prof. Dr. Holger Gohlke. Without him this work would not have been possible. During the last years, he supported me with his expertise, valuable suggestions and discussions, as well as guidance and encouragement.

In particular, I thank my family and friends for their continued support.

I am very grateful to my dear colleagues in the CPC Lab group and beyond for their friendship, willingness to help, the great working atmosphere, and many stimulating discussions. In particular, I thank Christopher Pflieger, Emanuele Ciglia, Tobias Kröger, and Hannes Kopitz for advancing our joint research and authorships to a success. In addition, I thank Dr. Nadine Homeyer for her assistance with free energy calculations.

I am grateful for the fruitful collaboration with the groups of Dr. Manuel Grez, Dr. Christian Wichmann, and Dr. Joachim Koch at the INSTITUTE FOR BIOMEDICAL RESEARCH GEORG-SPEYER-HAUS (Frankfurt). In particular, I thank Dr. Julia Schanda and Dr. Yvonne Bartel who performed experiments related to NHR2.

I am grateful for the fruitful collaboration with the group of Prof. Dr. Michael W. Göbel at the INSTITUTE OF ORGANIC CHEMISTRY AND CHEMICAL BIOLOGY at the GOETHE-UNIVERSITY (Frankfurt). In particular, I thank Cristiano Pinto Gomes who performed experiments related to oligo-oxazoles and Dr. Jan Bats who solved their crystal structures.

I am grateful for the fruitful collaboration with the LEAD DISCOVERY CENTER and to Dr. Karl-Heinz Baringhaus and Dr. Stefania Pfeiffer-Marek at SANOFI-AVENTIS DEUTSCHLAND (LGCR DRUG DESIGN).

I am grateful to the NCI/DTP OPEN CHEMICAL REPOSITORY for compound samples and to OPENEYE SCIENTIFIC SOFTWARE for a no-cost academic software license.

I am grateful for reprint permissions by BENTHAM SCIENCE PUBLISHERS (**Pub. I**), by the AMERICAN CHEMICAL SOCIETY (**Pub. II**), and by JOHN WILEY & SONS (**Pubs. III and IV**).

Finally, I am very grateful to Emanuele Ciglia, Dr. John Farmer, Christoph Gertzen, Daniela Grimme, Dr. Alexandra Hamacher, Dr. Nadine Homeyer, Doris Klein, Christopher Pflieger, Jagmohan Saini, Dr. Julia Schanda, and Ido Ben-Shalom for critical discussions and reading of this manuscript.

9 APPENDIX

9.1 Mathematical Model of NHR2 Equilibrium

In all experimental assays, we obtained ligand effectivities as IC_{50} values or relative activities at specific concentrations. These values depend on experimental parameters. To estimate the dissociation constant of the NHR2/PPIM complex (K_{lig}) I developed a mathematical model of the equilibrium of NHR2 and the NHR2-containing complexes (Eq. S1).



All predictions by this model are based on three assumptions: (i) The only relevant molecular species are: NHR2 tetramer (T), unbound bivalent NHR2 dimer (D), NHR2 dimer bound to a single PPIM (DL), and NHR2 dimer bound to two PPIMs (DLL). (ii) All ligand binding sites are independent, *i.e.* the dissociation constants K_{lig} for the first and second ligand binding event are equal. (iii) The experimentally measured effect in each assay is only proportional to the concentration of the NHR2 tetramer ($[T]$).

Naturally, the calculated K_{lig} depends on the dissociation constant K_{tet} of the NHR2 tetramer. However, K_{tet} could not be measured experimentally. Due to the stability of the NHR2 tetramer T , no dimer D could be detected. A lower limit of $K_{tet} = 2.8 - 4 \mu\text{M}$ may be defined by measurements of the self-association of an NHR2 peptide^[696] lacking all hot spots but W502.^[400] However, I expect a much higher affinity for native NHR2 with K_{tet} in the nano- to picomolar range.

In order to determine K_{lig} as a function of IC_{50} and K_{tet} , I set up equations that describe the equilibria in Eq. S1. This set of equations was derived by (i) defining the equilibrium constants and defining the mass balance equations for (ii) the NHR2 dimer D and (iii) the PPIM L . In the following, I describe how this set of equations was derived.

For any given concentration, the equilibrium of T and D is described by K_{tet} (Eq. S2).

$$K_{tet} = \frac{[D]^2}{[T]} \quad (\text{Eq. S2})$$

In the absence of a PPIM L , the concentration of unbound NHR2 dimer D is $[D]_0$ and the, in this case maximal, concentration of NHR2 tetramer T is $[T]_0$; the respective concentrations at IC_{50} are $[D]_{50}$ and $[T]_{50}$. The total concentration of NHR2 dimer $[D]_{tot}$ is (Eq. S3):

$$[D]_{tot} = 2[T]_0 + [D]_0 \quad (\text{Eq. S3})$$

Substituting Eq. S3 into Eq. S2 and subsequent rearrangement leads to Eq. S4.

$$[D]_0^2 + \frac{K_{tet}[D]_0}{2} - \frac{K_{tet}[D]_{tot}}{2} = 0 \quad (\text{Eq. S4})$$

Solving the quadratic Eq. S4 with respect to $[D]$ leads to Eq. S5. The negative solution of Eq. S4 has no physical meaning because of $[D]_0 \geq 0$ and was omitted.

$$[D]_0 = -\frac{K_{tet}}{4} + \sqrt{\left(\frac{K_{tet}}{4}\right)^2 + \frac{K_{tet}[D]_{tot}}{2}} \quad (\text{Eq. S5})$$

Substituting Eq. S5 into Eq. S3 and rearrangement leads to Eq. S6,

$$[T]_0 = \frac{[D]_{tot} + \frac{K_{tet}}{4} - \sqrt{\left(\frac{K_{tet}}{4}\right)^2 + \frac{K_{tet}[D]_{tot}}{2}}}{2} \quad (\text{Eq. S6})$$

which allows to calculate the maximal concentration of tetramer $[T]_0$ as a function of K_{tet} and $[D]_{tot}$. K_{tet} is constant and $[D]_{tot}$ is the summed concentration of all NHR2-containing species.

In the presence of a total PPIM concentration of $[L]_{tot} = IC_{50}$, the total concentration of NHR2 dimer $[D]_{tot}$ is defined by Eq. S7,

$$[D]_{tot} = 2[T]_{50} + [D]_{50} + [DL]_{50} + [DLL]_{50} \quad (\text{Eq. S7})$$

in which $[T]_{50}$, $[D]_{50}$, $[DL]_{50}$, and $[DLL]_{50}$ are the concentrations of the respective molecular species at a total PPIM concentration of $[L]_{tot} = IC_{50}$.

In this case, $[T]_{50}$ is half of the maximal tetramer concentration $[T]_0$ at $[L] = 0$ (Eq. S8),

$$[T]_{50} = \frac{[T]_0}{2} \quad (\text{Eq. S8})$$

if one assumes that the experimentally measured effect is proportional to $[T]$.

The concentration of the unbound NHR2 dimer $[D]_{50}$ is obtained by substituting Eq. S8 into Eq. S2 and rearrangement leading to Eq. S9. The negative solution of the quadratic equation has no physical meaning because of $[D]_{50} \geq 0$ and was omitted.

$$[D]_{50} = \sqrt{\frac{K_{tet}[T]_0}{2}} \quad (\text{Eq. S9})$$

Due to the definition of the dissociation constant of the PPIM-bound complexes K_{lig} , the concentrations of the PPIM-bound complexes are given by Eq. S10 and by Eq. S11,

$$[DL]_{50} = 2 \frac{[D]_{50}[L]_{50}}{K_{lig}} \quad (\text{Eq. S10})$$

$$[DLL]_{50} = \frac{[D]_{50}[L]_{50}^2}{K_{lig}^2} \quad (\text{Eq. S11})$$

if one assumes that the dissociation constants for both ligand binding events are equal, in agreement the remote ligand binding sites and our model (Eq. S1) that depends on the pre-dissociation of T . $[L]_{50}$ is the concentration of the unbound PPIM at IC_{50} . The factor of two in Eq. S10 is a result of the two equivalent ligand binding sites of the NHR2 dimer.

Substituting Eq. S10 and Eq. S11 into Eq. S7 and rearrangement leads to Eq. S12.

$$K_{lig}^2 + \frac{2[D]_{50}[L]_{50}}{(2[T]_{50} + [D]_{50} - [D]_{tot})} K_{lig} + \frac{[D]_{50}[L]_{50}^2}{(2[T]_{50} + [D]_{50} - [D]_{tot})} = 0 \quad (\text{Eq. S12})$$

Solving the quadratic Eq. S12 with respect to K_{lig} leads to Eq. S13. The negative solution of Eq. S12 has no physical meaning because of $K_{lig} \geq 0$ and was omitted.

$$K_{lig} = -\frac{[D]_{50}[L]_{50}}{(2[T]_{50} + [D]_{50} - [D]_{tot})} + \sqrt{\left(\frac{[D]_{50}[L]_{50}}{(2[T]_{50} + [D]_{50} - [D]_{tot})}\right)^2 - \frac{[D]_{50}[L]_{50}^2}{(2[T]_{50} + [D]_{50} - [D]_{tot})}} \quad (\text{Eq. S13})$$

$$K_{lig} = -\frac{[D]_{50}[L]_{50}}{([D]_{50} - [D]_{tot} + 2[T]_{50})} + \sqrt{\frac{[D]_{50}[L]_{50}^2([D]_{50} - 2[T]_{50})}{([D]_{50} - [D]_{tot} + 2[T]_{50})}} \quad \text{simplified (Eq. S13)}$$

According to Eq. S8 and Eq. S9 the terms $[D]_{50}$ and $[T]_{50}$ in Eq. S12 are constants for given $[T]_0$ and K_{tet} . Thus, Eq. S12 is a function of $[L]_{50}$, which remains to be defined.

For $[L]_{tot} = IC_{50}$, the total concentration of PPIM $[L]_{tot}$ is defined by Eq. S14.

$$[L]_{tot} = IC_{50} = [L]_{50} + [DL]_{50} + 2[DLL]_{50} \quad (\text{Eq. S14})$$

Substituting Eq. S10 and Eq. S11 into Eq. S14 and rearrangement leads to Eq. S15.

$$[L]_{50}^2 + \frac{\left(2 \frac{[D]_{50}}{K_{lig}} + 1\right)}{2 \frac{[D]_{50}}{K_{lig}^2}} [L]_{50} - \frac{IC_{50}}{2 \frac{[D]_{50}}{K_{lig}^2}} = 0 \quad (\text{Eq. S15})$$

Solving the quadratic Eq. S15 with respect to $[L]_{50}$ leads to Eq. S16. The negative solution of Eq. S15 has no physical meaning because of $[L]_{50} \geq 0$ and was omitted.

$$[L]_{50} = -\frac{1}{2} \frac{\left(2 \frac{[D]_{50}}{K_{lig}} + 1\right)}{\frac{2 [D]_{50}}{K_{lig}^2}} + \sqrt{\left(\frac{1}{2} \frac{\left(2 \frac{[D]_{50}}{K_{lig}} + 1\right)}{\frac{2 [D]_{50}}{K_{lig}^2}}\right)^2 + \frac{IC_{50}}{2 \frac{[D]_{50}}{K_{lig}^2}}} \quad (\text{Eq. S16})$$

$$[L]_{50} = \frac{1}{4} \left(-2K_{lig} - \frac{K_{lig}^2}{[D]_{50}} + \sqrt{\frac{K_{lig}^2 (4[D]_{50}^2 + 8[D]_{50} IC_{50} + 4[D]_{50} K_{lig} + K_{lig}^2)}{[D]_{50}^2}} \right) \text{ simplified (Eq. S16)}$$

According to Eq. S9, $[D]_{50}$ in Eq. S16 is constant for given $[T]_0$.

Finally, solving the system of Eq. 13 and Eq. 16 allows calculating K_{lig} for given $[D]_{tot}$, K_{tet} , and IC_{50} . For this purpose, $[D]_{tot}$ is the total concentration of NHR2 dimer in the assay. In this case $[D]_{50}$ and $[T]_{50}$ are constants calculated by Eq. S8 and Eq. S9, respectively. Of course K_{lig} can also be calculated for any relative activity of a PPIM at a specific concentration in an analogous manner.

An estimated $K_{tet} = 1$ nM and measured $IC_{50} = 630$ mM or $IC_{50} = 390$ mM for **7.44** or **P1** at $[D]_{tot} = 95.9$ nM for the NHR2 dimer in the ELISA leads to $K_{lig}(\mathbf{7.44}) = 274$ μ M or $K_{lig}(\mathbf{P1}) = 170$ μ M. At these conditions, the equilibrium model leads to dose-response curves (Figure S1) with Hill coefficients of 2.0 for both cases. This demonstrates the emergent cooperativity caused by the two equivalent binding sites in the NHR2 PPIface and is in good agreement with the experimentally determined Hill coefficients of 1.8 for **7.44** and of 1.9 for **P1** (Pub. IV).^[4]

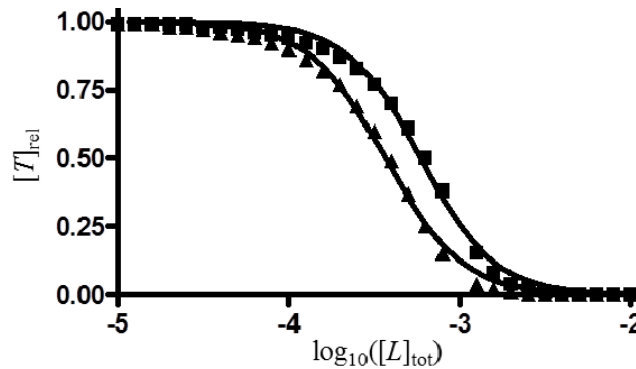


Figure S1. Dose-response curves of **7.44** (■) and **P1** (▲) according to the equilibrium model. The relative concentration of NHR2 ($[T]_{rel}$) is assumed proportional to the assay signal. Conditions described in main text.

9.2 Interactions of RE Domains

Different RE domains interact with regulatory DNA (such as RUNX and PU.1 promoters),^[110,413] p53,^[414] PKA,^[415,416] and recruit nuclear proteins,^[410,417-419] such as E proteins (*i.a.* HEB and E2A),^[76,420,421] the vitamin D₃ receptor,^[422] the apoptosis-related protein SON,^[423] DNA methyltransferases,^[63,424,425] ubiquitin-ligases,^[426] protein arginine methyltransferases,^[427,428] histone deacetylases (such as HDAC1),^[429,430] and nuclear co-repressors (such as N-CoR, SMRT, mSIN3A, and MTGR1).^[400,431-436] The trans-activation domain of RUNX1^[437] that binds histone acetyltransferases (such as p300, CBP, and P/CAF)^[427,438,439] is missing in RE,^[418] but p300 may still bind *via* NHR1.^[438,440]

9.3 Figure Generation

All 3D representations of molecules were generated with PyMOL^[697] with the following exceptions. Figure 16 and Figure 24 (right) were generated with VMD.^[698] Figure 17 was generated with vROCS.^[515]

All 2D data plots were generated with gnuplot.^[699]

Figure 2 was generated with Gephi v0.8.2.^[700]

Figure 5 was adapted from a design by Christopher Pfleger.

^[a] PKA: protein kinase A; HEB: HeLa E-box binding protein; SON: SON DNA-binding protein; HDAC1: histone deacetylases 1; N-CoR: nuclear receptor co-repressor; SMRT: silencing mediator for retinoid or thyroid-hormone receptors; mSIN3A: mammalian suppressor interacting 3 homolog A of transcriptional repressor Sin3; MTGR1: myeloid translocation gene-related 1; CBP: CREB (cAMP response element-binding protein)-binding protein; P/CAF: p300/CBP-associated factor; CBF β : core-binding factor subunit β ; SMMHC: smooth muscle myosin heavy chain

9.4 Reprint Permissions for Publications

I am grateful to Bentham Science Publishers, Ltd. for the permission to reprint **Publication I**.
Reprinted (adapted) with permission from:

Alexander Metz, Emanuele Ciglia, and Holger Gohlke.
Modulating Protein-Protein Interactions: From Structural Determinants of Binding to Druggability Prediction to Application.
Current Pharmaceutical Design, 2012; 18: 4630-4647.
DOI: 10.2174/138161212802651553
Link: <http://dx.doi.org/10.2174/138161212802651553>

Copyright (2012) Bentham Science Publishers, Ltd.

I am grateful to the American Chemical Society for the permission to reprint **Publication II**.
Reprinted (adapted) with permission from:

Alexander Metz, Christopher Pfleger, Hannes Kopitz, Stefania Pfeiffer-Marek, Karl-Heinz Baringhaus, and Holger Gohlke.
Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein-Protein Interface.
Journal of Chemical Information and Modeling, 2012; 52: 120-133.
DOI: 10.1021/ci200322s
Link: <http://pubs.acs.org/doi/abs/10.1021/ci200322s>

Copyright (2011) American Chemical Society.

I am grateful to John Wiley & Sons, Inc. for the permission to reprint **Publication III**.
Reprinted (adapted) with permission from:

Cristiano Pinto Gomes, Alexander Metz, Jan W. Bats, Holger Gohlke, and Michael W. Göbel.
Modular Solid-Phase Synthesis of Teroxazoles as a Class of α -Helix Mimetics.
European Journal of Organic Chemistry, 2012; 17: 3270-3277.
DOI: 10.1002/ejoc.201200339
Link: <http://onlinelibrary.wiley.com/doi/10.1002/ejoc.201200339/abstract>

Copyright (2012) WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

I am grateful to the American Chemical Society for the permission to reprint **Publication IV**.
Reprinted (adapted) with permission from:

Alexander Metz, Julia Schanda, Manuel Grez, Christian Wichmann, and Holger Gohlke.
From Determinants of RUNX1/ETO Tetramerization to Small-Molecule Protein/Protein Interaction Modulators Targeting Acute Myeloid Leukemia.
Journal of Chemical Information and Modeling, 2013; 53: 2197-2202.
DOI: 10.1021/ci400332e
Link: <http://pubs.acs.org/doi/abs/10.1021/ci400332e>

Copyright (2013) American Chemical Society.

10 REFERENCES

- [1] A. Metz, E. Ciglia, H. Gohlke, Modulating Protein-Protein Interactions: From Structural Determinants of Binding to Druggability Prediction to Application, *Curr. Pharm. Des.* **2012**, 18(30): 4630-4647.
- [2] A. Metz, C. Pfleger, H. Kopitz, S. Pfeiffer-Marek, K.H. Baringhaus, H. Gohlke, Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein-Protein Interface, *J. Chem. Inf. Model.* **2012**, 52(1): 120-133.
- [3] C.P. Gomes, A. Metz, J.W. Bats, H. Gohlke, M.W. Göbel, Modular Solid-Phase Synthesis of Teroxazoles as a Class of α -Helix Mimetics, *Eur. J. Org. Chem.* **2012**, 2012(17): 3270–3277.
- [4] A. Metz, J. Schanda, M. Grez, C. Wichmann, H. Gohlke, From Determinants of RUNX1/ETO Tetramerization to Small-Molecule Protein-Protein Interaction Inhibitors Targeting Acute Myeloid Leukemia, *J. Chem. Inf. Model.* **2013**, 53(9): 2197-2202.
- [5] A. Metz, Institute for Biomedical Research Georg-Speyer-Haus (Frankfurt), H. Gohlke, Inhibitors of NHR2 and/or RUNX1/ETO-tetramerization (EP Appl. 13165993.0), **2013**: 1-69.
- [6] G.J. Pielak, A model of intracellular organization, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102(17): 5901-5902.
- [7] X.W. Zhu, M. Gerstein, M. Snyder, Getting connected: analysis and principles of biological networks, *Genes Dev.* **2007**, 21(9): 1010-1024.
- [8] M. Vidal, M.E. Cusick, A.L. Barabasi, Interactome Networks and Human Disease, *Cell* **2011**, 144(6): 986-998.
- [9] E. Fischer, Einfluss der Configuration auf die Wirkung der Enzyme, *Ber. Dtsch. Chem. Ges.* **1894**, 27: 2985-2993.
- [10] F.W. Lichtenthaler, 100 Years "Schlüssel-Schloss-Prinzip": What Made Emil Fischer Use This Analogy?, *Angew. Chem., Int. Ed. Engl.* **1995**, 33(23-24): 2364-2374.
- [11] D. Lemoine, R.T. Jiang, A. Taly, T. Chataigneau, A. Specht, T. Grutter, Ligand-Gated Ion Channels: New Insights into Neurological Disorders and Ligand Recognition, *Chem. Rev.* **2012**, 112(12): 6285-6318.
- [12] M. Aggarwal, R. McKenna, PDB ID: 3P25, Human Carbonic Anhydrase II complexed with 2,3,4,5-tetrafluorophenyl-1-sulfonamido-4-(methyl-1,2,3-triazole-4-carboxylate).
- [13] P. Ehrlich, Address in Pathology on Chemotherapeutics: Scientific Principles, Methods, and Results, *Lancet* **1913**: 445-451.
- [14] S. Kerrien, B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, M. Duesbury, M. Dumousseau, M. Feuermann, U. Hinz, *et al.*, The IntAct molecular interaction database in 2012, *Nucleic Acids Res.* **2012**, 40(D1): D841-D846.
- [15] A.L. Hopkins, C.R. Groom, The druggable genome, *Nat. Rev. Drug Discovery* **2002**, 1(9): 727-730.
- [16] J.P. Overington, B. Al-Lazikani, A.L. Hopkins, How many drug targets are there?, *Nat. Rev. Drug Discovery* **2006**, 5(12): 993-996.
- [17] P. Imming, C. Sinning, A. Meyer, Drugs, their targets and the nature and number of drug targets, *Nat. Rev. Drug Discovery* **2006**, 5(10): 821-834.
- [18] M. Rask-Andersen, M.S. Almen, H.B. Schioth, Trends in the exploitation of novel drug targets, *Nat. Rev. Drug Discovery* **2011**, 10(8): 579-590.
- [19] F. Zhu, B.C. Han, P. Kumar, X.H. Liu, X.H. Ma, X.N. Wei, L. Huang, Y.F. Guo, L.Y. Han, C.J. Zheng, *et al.*, Update of TTD: Therapeutic Target Database, *Nucleic Acids Res.* **2010**, 38: D787-D791.

- [20] F. Zhu, Z. Shi, C. Qin, L. Tao, X. Liu, F. Xu, L. Zhang, Y. Song, X.H. Liu, J.X. Zhang, *et al.*, Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery, *Nucleic Acids Res.* **2012**, 40(D1): D1128-D1136.
- [21] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, *et al.*, DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs, *Nucleic Acids Res.* **2011**, 39: D1035-D1041.
- [22] M.P.H. Stumpf, T. Thorne, E. de Silva, R. Stewart, H.J. An, M. Lappe, C. Wiuf, Estimating the size of the human interactome, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105(19): 6959-6964.
- [23] L. Bonetta, Interactome under construction, *Nature* **2010**, 468(7325): 851-854.
- [24] Q.C. Zhang, D. Petrey, L. Deng, L. Qiang, Y. Shi, C.A. Thu, B. Bisikirska, C. Lefebvre, D. Accili, T. Hunter, *et al.*, Structure-based prediction of protein-protein interactions on a genome-wide scale, *Nature* **2012**, 490(7421): 556-561.
- [25] L.N. Makley, J.E. Gestwicki, Expanding the Number of 'Druggable' Targets: Non-Enzymes and Protein-Protein Interactions, *Chem. Biol. Drug Des.* **2013**, 81(1): 22-32.
- [26] C. Reynès, H. Host, A.C. Camproux, G. Laconde, F. Leroux, A. Mazars, B. Deprez, R. Fahraeus, B.O. Villoutreix, O. Sperandio, Designing Focused Chemical Libraries Enriched in Protein-Protein Interaction Inhibitors using Machine-Learning Methods, *PLoS Comput. Biol.* **2010**, 6(3): e1000695.
- [27] M.E. Bunnage, Getting pharmaceutical R&D back on target, *Nat. Chem. Biol.* **2011**, 7(6): 335-339.
- [28] J.A. Wells, C.L. McClendon, Reaching for high-hanging fruit in drug discovery at protein-protein interfaces, *Nature* **2007**, 450(7172): 1001-1009.
- [29] L. Vassilev, D. Fry (eds.) *Small-Molecule Inhibitors of Protein-Protein Interactions* **2011**, (Springer-Verlag, Berlin Heidelberg), 1 Ed., Vol. 1, p. 174.
- [30] F.J. Cohen, Macro trends in pharmaceutical innovation, *Nat. Rev. Drug Discovery* **2005**, 4(1): 78-84.
- [31] K.I. Kaitin, Deconstructing the Drug Development Process: The New Face of Innovation, *Clin. Pharmacol. Ther. (N. Y., NY, U. S.)* **2010**, 87(3): 356-361.
- [32] K.I. Kaitin, J.A. DiMasi, Pharmaceutical Innovation in the 21st Century: New Drug Approvals in the First Decade, 2000-2009, *Clin. Pharmacol. Ther. (N. Y., NY, U. S.)* **2011**, 89(2): 183-188.
- [33] B.H. Munos, W.W. Chin, How to Revive Breakthrough Innovation in the Pharmaceutical Industry, *Sci. Transl. Med.* **2011**, 3(89).
- [34] M. Herper, Drug Drought, *Forbes* **2007**, 180(9): 44.
- [35] S.M. Paul, D.S. Mytelka, C.T. Dunwiddie, C.C. Persinger, B.H. Munos, S.R. Lindborg, A.L. Schacht, How to improve R&D productivity: the pharmaceutical industry's grand challenge, *Nat. Rev. Drug Discovery* **2010**, 9(3): 203-214.
- [36] E.E. Slater, K.I. Kaitin, Explaining the Drug Drought, *Pharmaceut. Executive* **2010**, 30: 22-24.
- [37] J.A. DiMasi, R.W. Hansen, H.G. Grabowski, The price of innovation: new estimates of drug development costs, *J. Health Econ.* **2003**, 22(2): 151-185.
- [38] C.P. Adams, V.V. Brantner, Estimating the cost of new drug development: Is it really \$802 million?, *Health Affairs* **2006**, 25(2): 420-428.
- [39] C.P. Adams, V.V. Brantner, Spending on New Drug Development, *Health Economics* **2010**, 19(2): 130-141.
- [40] M. Herper, The Truly Staggering Cost Of Inventing New Drugs, *Forbes* **2012**, 189(4): 38-38.
- [41] Pharmaceutical Research and Manufacturers of America, Pharmaceutical Industry Profile 2012, **2012**: 1-58.

-
- [42] J.C. Fuller, N.J. Burgoyne, R.M. Jackson, Predicting druggable binding sites at the protein-protein interface, *Drug Discovery Today* **2009**, 14(3-4): 155-161.
- [43] S. Surade, T.L. Blundell, Structural biology and drug discovery of difficult targets: the limits of ligandability, *Chem. Biol. (Oxford, U. K.)* **2012**, 19(1): 42-50.
- [44] O. Ulucan, S. Eyrisch, V. Helms, Druggability of Dynamic Protein-protein Interfaces, *Curr. Pharm. Des.* **2012**, 18(30): 4599-4606.
- [45] D. Gonzalez-Ruiz, H. Gohlke, Targeting protein-protein interactions with small molecules: Challenges and perspectives for computational binding epitope detection and ligand finding, *Curr. Med. Chem.* **2006**, 13(22): 2607-2625.
- [46] J. Bultinck, S. Lievens, J. Tavernier, Protein-protein Interactions: Network Analysis and Applications in Drug Discovery, *Curr. Pharm. Des.* **2012**, 18(30): 4619-4629.
- [47] G. Kar, G. Kuzu, O. Keskin, A. Gursoy, Protein-protein Interfaces Integrated into Interaction Networks: Implications on Drug Design, *Curr. Pharm. Des.* **2012**, 18(30): 4697-4705.
- [48] B.O. Villoutreix, C.M. Labbe, D. Lagorce, G. Laconde, O. Sperandio, A Leap into the Chemical Space of Protein-Protein Interaction Inhibitors, *Curr. Pharm. Des.* **2012**, 18(30): 4648-4667.
- [49] A.P. Higuero, H. Jubb, T.L. Blundell, TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions, *Database* **2013**: bat039.
- [50] C.M. Labbe, G. Laconde, M.A. Kuenemann, B.O. Villoutreix, O. Sperandio, iPPI-DB: a manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions, *Drug Discovery Today* **2013**: 958-968.
- [51] M.J. Basse, S. Betzi, R. Bourgeois, S. Bouzidi, B. Chetrit, V. Hamon, X. Morelli, P. Roche, 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions, *Nucleic Acids Res.* **2013**, 41(D1): D824-D827.
- [52] T.R. Malek, The biology of interleukin-2, *Annu. Rev. Immunol.* **2008**, 26: 453-479.
- [53] S.J. Wodak, J. Vlasblom, A.L. Turinsky, S. Pu, Protein-protein interaction networks: the puzzling riches, *Curr. Opin. Struct. Biol.* **2013**, 23: 941-953.
- [54] R.T. Bradshaw, B.H. Patel, E.W. Tate, R.J. Leatherbarrow, I.R. Gould, Comparing experimental and computational alanine scanning techniques for probing a prototypical protein-protein interaction, *Protein Eng., Des. Sel.* **2011**, 24(1-2): 197-207.
- [55] B.A. Shoemaker, A.R. Panchenko, Deciphering protein-protein interactions. Part I. Experimental techniques and databases, *PLoS Comput. Biol.* **2007**, 3(3): 337-344.
- [56] H. Jubb, A.P. Higuero, A. Winter, T.L. Blundell, Structural biology and drug discovery for protein-protein interactions, *Trends Pharmacol. Sci.* **2012**, 33(5): 241-248.
- [57] Y. Tanrikulu, B. Kruger, E. Proschak, The holistic integration of virtual screening in drug discovery, *Drug Discovery Today* **2013**, 18(7-8): 358-364.
- [58] G. Schneider, Virtual screening: an endless staircase?, *Nat. Rev. Drug Discovery* **2010**, 9(4): 273-276.
- [59] B.K. Shoichet, Virtual screening of chemical libraries, *Nature* **2004**, 432(7019): 862-865.
- [60] D.M. Krüger, H. Gohlke, DrugScore^{PPI} webserver: fast and accurate *in silico* alanine scanning for scoring protein-protein interactions, *Nucleic Acids Res.* **2010**, 38: W480-W486.
- [61] D.M. Krüger, G. Jessen, H. Gohlke, How Good Are State-of-the-Art Docking Tools in Predicting Ligand Binding Modes in Protein-Protein Interfaces?, *J. Chem. Inf. Model.* **2012**, 52(11): 2807-2811.

-
- [62] A.A. Bogan, K.S. Thorn, Anatomy of hot spots in protein interfaces, *J. Mol. Biol.* **1998**, 280(1): 1-9.
- [63] Y. Liu, M.D. Cheney, J.J. Gaudet, M. Chruszcz, S.M. Lukasik, D. Sugiyama, J. Lary, J. Cole, Z. Dauter, W. Minor, *et al.*, The tetramer structure of the Nrvy homology two domain, NHR2, is critical for AML1/ETO's activity, *Cancer Cell* **2006**, 9(4): 249-260.
- [64] T. Pawson, P. Nash, Protein-protein interactions define specificity in signal transduction, *Genes Dev.* **2000**, 14(9): 1027-1047.
- [65] M.D. Dyer, T.M. Murali, B.W. Sobral, The Landscape of Human Proteins Interacting with Viruses and Other Pathogens, *PLoS Pathog.* **2008**, 4(2): e32.
- [66] C.J. Oldfield, J. Meng, J.Y. Yang, M.Q. Yang, V.N. Uversky, A.K. Dunker, Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners, *BMC Genomics* **2008**, 9(S1).
- [67] A. David, R. Razali, M.N. Wass, M.J.E. Sternberg, Protein-Protein Interaction Sites are Hot Spots for Disease-Associated Nonsynonymous SNPs, *Hum. Mutat.* **2012**, 33(2): 359-363.
- [68] M.W. Gonzalez, M.G. Kann, Chapter 4: Protein Interactions and Disease, *PLoS Comput. Biol.* **2012**, 8(12): e1002819.
- [69] L. Pray, Eukaryotic genome complexity, *Nature Education* **2008**, 1(1): 96.
- [70] J.M. Claverie, Gene number: What if there are only 30,000 human genes?, *Science* **2001**, 291(5507): 1255-1257.
- [71] M. Pertea, S.L. Salzberg, Between a chicken and a grape: estimating the number of human genes, *Genome Biol.* **2010**, 11(5): 206.
- [72] E. Pennisi, ENCODE Project Writes Eulogy For Junk DNA, *Science* **2012**, 337(6099): 1159-1161.
- [73] C. Vogel, M. Bashton, N.D. Kerrison, C. Chothia, S.A. Teichmann, Structure, function and evolution of multidomain proteins, *Curr. Opin. Struct. Biol.* **2004**, 14(2): 208-216.
- [74] D. Ekman, A.K. Bjorklund, J. Frey-Skott, A. Elofsson, Multi-domain proteins in the three kingdoms of life: Orphan domains and other unassigned regions, *J. Mol. Biol.* **2005**, 348(1): 231-243.
- [75] C.W. So, M.L. Cleary, Dimerization: a versatile switch for oncogenesis, *Blood* **2004**, 104(4): 919-922.
- [76] X.J. Sun, Z. Wang, L. Wang, Y. Jiang, N. Kost, T.D. Soong, W.Y. Chen, Z. Tang, T. Nakadai, O. Elemento, *et al.*, A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis, *Nature* **2013**, 500: 93-97.
- [77] J.H.A. Martens, A. Mandoli, F. Simmer, B.J. Wierenga, S. Saeed, A.A. Singh, L. Altucci, E. Vellenga, H.G. Stunnenberg, ERG and FLI1 binding sites demarcate targets for aberrant epigenetic regulation by AML1-ETO in acute myeloid leukemia, *Blood* **2012**, 120(19): 4038-4048.
- [78] R. Jaenisch, A. Bird, Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals, *Nature Genetics* **2003**, 33: 245-254.
- [79] F. Andreoli, A.J.M. Barbosa, M.D. Parenti, A. Del Rio, Modulation of Epigenetic Targets for Anticancer Therapy: Clinicopathological Relevance, Structural Data and Drug Discovery Perspectives, *Curr. Pharm. Des.* **2013**, 19(4): 578-613.
- [80] J. Qiu, J. Wong, D.J. Tweardy, S. Dong, Decreased intranuclear mobility of acute myeloid leukemia 1-containing fusion proteins is accompanied by reduced mobility and compartmentalization of core binding factor β , *Oncogene* **2006**, 25(28): 3982-3993.

-
- [81] L.D. Walensky, Protein-Protein Interactions: A PUMA Mechanism Unfolds, *Nat. Chem. Biol.* **2013**, 9(3): 141-143.
- [82] A. Disanza, E. Frittoli, A. Palamidessi, G. Scita, Endocytosis and spatial restriction of cell signaling, *Mol. Oncol.* **2009**, 3(4): 280-296.
- [83] K. Tanaka, T. Tanaka, M. Kurokawa, Y. Imai, S. Ogawa, K. Mitani, Y. Yazaki, H. Hirai, The AML1/ETO(MTG8) and AML1/Evi-1 leukemia-associated chimeric oncoproteins accumulate PEBP2 β (CBF β) in the nucleus more efficiently than wild-type AML1, *Blood* **1998**, 91(5): 1688-1699.
- [84] Y.C. Kim, R.B. Best, J. Mittal, Macromolecular crowding effects on protein-protein binding affinity and specificity, *J Chem Phys* **2010**, 133(20): 205101.
- [85] D.P. Ryan, J.M. Matthews, Protein-protein interactions in human disease, *Curr. Opin. Struct. Biol.* **2005**, 15(4): 441-446.
- [86] L. Sam, Y. Liu, J. Li, C. Friedman, Y.A. Lussier, Discovery of protein interaction networks shared by diseases, *Pac. Symp. Biocomput. 2007* **2007**: 76-87.
- [87] A. Garner, K. Janda, Protein-Protein Interactions and Cancer: Targeting the Central Dogma, *Curr. Top. Med. Chem.* **2011**, 11(3): 258-280.
- [88] D. Hanahan, R.A. Weinberg, The Hallmarks of Cancer, *Cell* **2000**, 100(1): 57-70.
- [89] D. Hanahan, R.A. Weinberg, Hallmarks of Cancer: The Next Generation, *Cell* **2011**, 144(5): 646-674.
- [90] A. Wittinghofer, H. Waldmann, Ras - A Molecular Switch Involved in Tumor Formation, *Angew. Chem., Int. Ed.* **2000**, 39(23): 4193-4214.
- [91] L.T. Vassilev, p53 Activation by Small Molecules: Application in Oncology, *Journal of Medicinal Chemistry* **2005**, 48(14): 4491-4499.
- [92] H. Gohlke, C. Kiel, D.A. Case, Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes, *J. Mol. Biol.* **2003**, 330(4): 891-913.
- [93] H. Gohlke, D.A. Case, Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf, *J. Comput. Chem.* **2004**, 25(2): 238-250.
- [94] H.B. Wang, D.I. Hammoudeh, A.V. Follis, B.E. Reese, J.S. Lazo, S.J. Metallo, E.V. Prochownik, Improved low molecular weight Myc-Max inhibitors, *Mol. Cancer Ther.* **2007**, 6(9): 2399-2408.
- [95] H.B. Wang, J. Chauhan, A. Hu, K. Pendleton, J.L. Yap, P.E. Sabato, J.W. Jones, M. Perri, J.S. Yu, E. Cione, *et al.*, Disruption of Myc-Max Heterodimerization with Improved Cell-Penetrating Analogs of the Small Molecule 10074-G5, *Oncotarget* **2013**, 4(6): 936-947.
- [96] S.J. Metallo, Intrinsically disordered proteins are potential drug targets, *Curr. Opin. Chem. Biol.* **2010**, 14(4): 481-488.
- [97] J.L. Yap, J. Chauhan, K.Y. Jung, L.J. Chen, E.V. Prochownik, S. Fletcher, Small-molecule inhibitors of dimeric transcription factors: Antagonism of protein-protein and protein-DNA interactions, *MedChemComm* **2012**, 3(5): 541-551.
- [98] F.M. Chen, C.N. Hancock, A.T. Macias, J. Joh, K. Still, S.J. Zhong, A.D. MacKerell, P. Shapiro, Characterization of ATP-independent ERK inhibitors identified through in silico analysis of the active ERK2 structure, *Bioorg. Med. Chem. Lett.* **2006**, 16(24): 6281-6287.
- [99] J.L. Yap, S. Worlikar, A.D. MacKerell, P. Shapiro, S. Fletcher, Small-Molecule Inhibitors of the ERK Signaling Pathway: Towards Novel Anticancer Therapeutics, *Chemmedchem* **2011**, 6(1): 38-48.
- [100] B. Vogelstein, D. Lane, A.J. Levine, Surfing the p53 network, *Nature* **2000**, 408(6810): 307-310.

-
- [101] K. Suzuki, H. Matsubara, Recent Advances in p53 Research and Cancer Treatment, *J. Biomed. Biotechnol.* **2011**: Art. No. 978312.
- [102] D.P. Lane, C.F. Cheok, S. Lain, p53-based Cancer Therapy, *Cold Spring Harbor Perspect. Biol.* **2010**, 2(9): a001222.
- [103] P.M. Fischer, Peptide, peptidomimetic, and small-molecule antagonists of the p53-HDM2 protein-protein interaction, *Int. J. Pept. Res. Ther.* **2006**, 12(1): 3-19.
- [104] J.M. Adams, S. Cory, The Bcl-2 apoptotic switch in cancer development and therapy, *Oncogene* **2007**, 26(9): 1324-1337.
- [105] L.L. Blazer, R.R. Neubig, Small Molecule Protein-Protein Interaction Inhibitors as CNS Therapeutic Agents: Current Progress and Future Hurdles, *Neuropsychopharmacology* **2009**, 34(1): 126-141.
- [106] L. Jiang, C. Liu, D. Leibly, M. Landau, M. Zhao, M.P. Hughes, D.S. Eisenberg, J. Kuriyan, Structure-based discovery of fiber-binding compounds that reduce the cytotoxicity of amyloid beta, *eLife* **2013**, 2: e00857.
- [107] S.J. Watowich, L.J. Gross, R. Josephs, Analysis of the Intermolecular Contacts within Sickle Hemoglobin Fibers: Effect of Site-Specific Substitutions, Fiber Pitch, and Double-Strand Disorder, *J. Struct. Biol.* **1993**, 111(3): 161-179.
- [108] R.C. Stevens, V. Cherezov, V. Katritch, R. Abagyan, P. Kuhn, H. Rosen, K. Wuthrich, The GPCR Network: a large-scale collaboration to determine human GPCR structure and function, *Nat. Rev. Drug Discovery* **2013**, 12(1): 25-34.
- [109] P. Cohen, Protein kinases - the major drug targets of the twenty-first century?, *Nat. Rev. Drug Discovery* **2002**, 1(4): 309-315.
- [110] A. Ptasinska, S.A. Assi, D. Mannari, S.R. James, D. Williamson, J. Dunne, M. Hoogenkamp, M. Wu, M. Care, H. McNeill, *et al.*, Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding, *Leukemia* **2012**, 26: 1829-1841.
- [111] A. Maiques-Diaz, F.S. Chou, M. Wunderlich, G. Gomez-Lopez, F.V. Jacinto, S. Rodriguez-Perales, M.J. Larrayoz, M.J. Calasanz, J.C. Mulloy, J.C. Cigudosa, *et al.*, Chromatin modifications induced by the AML1-ETO fusion protein reversibly silence its genomic targets through AML1 and Sp1 binding motifs, *Leukemia* **2012**, 26(6): 1329-1337.
- [112] H. Ruffner, A. Bauer, T. Bouwmeester, Human protein-protein interaction networks and the value for drug discovery, *Drug Discovery Today* **2007**, 12(17-18): 709-716.
- [113] I.W. Taylor, J.L. Wrana, Protein interaction networks in medicine and disease, *Proteomics* **2012**, 12(10): 1706-1716.
- [114] R.K. Nibbe, S.A. Chowdhury, M. Koyuturk, R. Ewing, M.R. Chance, Protein-protein interaction networks and subnetworks in the biology of disease, *Wiley Interdiscip. Rev.: Syst. Biol. Med.* **2011**, 3(3): 357-367.
- [115] T. Ideker, R. Sharan, Protein networks in disease, *Genome Res.* **2008**, 18(4): 644-652.
- [116] N.E. Davey, G. Trave, T.J. Gibson, How viruses hijack cell regulation, *Trends Biochem. Sci.* **2011**, 36(3): 159-169.
- [117] J.P. Donahue, M.L. Vetter, N.A. Mukhtar, R.T. D'Aquila, The HIV-1 Vif PPLP motif is necessary for human APOBEC3G binding and degradation, *Virology* **2008**, 377(1): 49-53.
- [118] K. Busschots, J. De Rijck, F. Christ, Z. Debyser, In search of small molecules blocking interactions between HIV proteins and intracellular cofactors, *Mol. BioSyst.* **2009**, 5(1): 21-31.
- [119] S.D. Shapira, I. Gat-Viks, B.O.V. Shum, A. Dricot, M.M. de Grace, L.G. Wu, P.B. Gupta, T. Hao, S.J. Silver, D.E. Root, *et al.*, A Physical and Regulatory Map of Host-

- Influenza Interactions Reveals Pathways in H1N1 Infection, *Cell* **2009**, 139(7): 1255-1267.
- [120] M.D. Dyer, C. Neff, M. Dufford, C.G. Rivera, D. Shattuck, J. Bassaganya-Riera, T.M. Murali, B.W. Sobral, The Human-Bacterial Pathogen Protein Interaction Networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*, *PLoS One* **2010**, 5(8): e12089.
- [121] E.D. Belay, L.B. Schonberger, The public health impact of prion diseases, *Annu. Rev. Public Health* **2005**, 26: 191-212.
- [122] R.K. Oldham, R.O. Dillman, Monoclonal antibodies in cancer therapy: 25 years of progress, *J. Clin. Oncol.* **2008**, 26(11): 1774-1777.
- [123] L.H. Stockwin, S. Holmes, Antibodies as therapeutic agents: *vive la renaissance!*, *Expert Opin. Biol. Ther.* **2003**, 3(7): 1133-1152.
- [124] C. Vogel, M.A. Cobleigh, D. Tripathy, J.C. Gutheil, L.N. Harris, L. Fehrenbacher, D.J. Slamon, M. Murphy, W.F. Novotny, M. Burchmore, *et al.*, First-line, single-agent Herceptin^(R) (trastuzumab) in metastatic breast cancer: a preliminary report, *Eur. J. Cancer* **2001**, 37: S25-S29.
- [125] V. Lafont, M. Schaefer, R.H. Stote, D. Altschuh, A. Dejaegere, Protein-protein recognition and interaction hot spots in an antigen-antibody complex: Free energy decomposition identifies "efficient amino acids", *Proteins: Struct., Funct., Bioinf.* **2007**, 67(2): 418-434.
- [126] L. Griffin, A. Lawson, Antibody fragments as tools in crystallography, *Clin. Exp. Immunol.* **2011**, 165(3): 285-291.
- [127] A.P. Higuero, H. Jubb, T.L. Blundell, Protein-protein interactions as druggable targets: recent technological advances, *Curr. Opin. Pharmacol.* **2013**, 13(5): 791-796.
- [128] K. Imai, A. Takaoka, Comparing antibody and small-molecule therapies for cancer, *Nat. Rev. Cancer* **2006**, 6(9): 714-727.
- [129] G. Zinzalla, D.E. Thurston, Targeting protein-protein interactions for therapeutic intervention: a challenge for the future, *Future Med. Chem.* **2009**, 1(1): 65-93.
- [130] J.H. Lin, A.Y.H. Lu, Role of pharmacokinetics and metabolism in drug discovery and development, *Pharmacological Reviews* **1997**, 49(4): 403-449.
- [131] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Delivery Rev.* **2001**, 46(1-3): 3-26.
- [132] A. Ruiz-Garcia, M. Bermejo, A. Moss, V.G. Casabo, Pharmacokinetics in drug discovery, *J. Pharm. Sci.* **2008**, 97(2): 654-690.
- [133] R. Gowthaman, E.J. Deeds, J. Karanicolas, The structural properties of non-traditional drug targets present new challenges for virtual screening, *J. Chem. Inf. Model.* **2013**, 53(8): 2073-2081.
- [134] A. Whitty, G. Kumaravel, Between a rock and a hard place?, *Nat. Chem. Biol.* **2006**, 2(3): 112-118.
- [135] A. Mullard, Protein-protein interaction inhibitors get into the groove, *Nat. Rev. Drug Discovery* **2012**, 11(3): 172-174.
- [136] T.M. Allen, P.R. Cullis, Drug delivery systems: Entering the mainstream, *Science* **2004**, 303(5665): 1818-1822.
- [137] J. Rautio, H. Kumpulainen, T. Heimbach, R. Oliyai, D. Oh, T. Jarvinen, J. Savolainen, Prodrugs: design and clinical applications, *Nat. Rev. Drug Discovery* **2008**, 7(3): 255-270.
- [138] D. Fry, K.S. Huang, P. Di Lello, P. Mohr, K. Muller, S.S. So, T. Harada, M. Stahl, B. Vu, H. Mauser, Design of Libraries Targeting Protein-Protein Interfaces, *ChemMedChem* **2013**, 8(5): 726-732.

- [139] B. Vu, P. Wovkulich, G. Pizzolato, A. Lovey, Q.J. Ding, N. Jiang, J.J. Liu, C.L. Zhao, K. Glenn, Y. Wen, *et al.*, Discovery of RG7112: A Small-Molecule MDM2 Inhibitor in Clinical Development, *ACS Med. Chem. Lett.* **2013**, 4(5): 466-469.
- [140] A.J. Souers, J.D. Levenson, E.R. Boghaert, S.L. Ackler, N.D. Catron, J. Chen, B.D. Dayton, H. Ding, S.H. Enschede, W.J. Fairbrother, *et al.*, ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets, *Nat. Med.* **2013**, 19(2): 202-208.
- [141] M.D. Wendt (ed.) *Protein-Protein Interactions* **2012**, (Springer-Verlag, Berlin Heidelberg), 1 Ed., Vol. 1, p. 174.
- [142] Y. Zhao, D. Bernard, S. Wang, Small Molecule Inhibitors of MDM2-p53 and MDMX-p53 Interactions as New Cancer Therapeutics, *BioDiscovery* **2013**, 8(4): 1-15.
- [143] J.C. Reed, Drug insight: cancer therapy strategies based on restoration of endogenous cell death mechanisms, *Nat. Clin. Pract. Oncol.* **2006**, 3(7): 388-398.
- [144] U. Fischer, K. Schulze-Osthoff, Apoptosis-based therapies and drug targets, *Cell Death Differ.* **2005**, 12: 942-961.
- [145] C.M. Rudin, C.L. Hann, E.B. Garon, M.R. de Oliveira, P.D. Bonomi, D.R. Camidge, Q. Chu, G. Giaccone, D. Khaira, S.S. Ramalingam, *et al.*, Phase II Study of Single-Agent Navitoclax (ABT-263) and Biomarker Correlates in Patients with Relapsed Small Cell Lung Cancer, *Clin. Cancer Res.* **2012**, 18(11): 3163-3169.
- [146] J. Rudolph, Inhibiting transient protein-protein interactions: lessons from the Cdc25 protein tyrosine phosphatases, *Nat. Rev. Cancer* **2007**, 7(3): 202-211.
- [147] L.T. Vassilev, Small-molecule antagonists of p53-MDM2 binding - Research tools and potential therapeutics, *Cell Cycle* **2004**, 3(4): 419-421.
- [148] D. Rauh, Chemical Biology - Current and Next Challenges, *ACS Chem. Biol.* **2013**, 8(1): 1-2.
- [149] C. Lipinski, A. Hopkins, Navigating chemical space for biology and medicine, *Nature* **2004**, 432(7019): 855-861.
- [150] F.S. Liang, W.Q. Ho, G.R. Crabtree, Engineering the ABA Plant Stress Pathway for Regulation of Induced Proximity, *Sci. Signaling* **2011**, 4(164): rs2.
- [151] G. Schreiber, S.J. Fleishman, Computational design of protein-protein interactions, *Curr. Opin. Struct. Biol.* **2013**, 23: 903-910.
- [152] K.K. Jain, Role of pharmacoproteomics in the development of personalized medicine, *Pharmacogenomics* **2004**, 5(3): 331-336.
- [153] L.A. Liotta, E.C. Kohn, E.F. Petricoin, Clinical proteomics: Personalized molecular medicine, *JAMA, J. Am. Med. Assoc.* **2001**, 286(18): 2211-2214.
- [154] C.T. Keith, A.A. Borisy, B.R. Stockwell, Multicomponent therapeutics for networked systems, *Nat. Rev. Drug Discovery* **2005**, 4(1): 71-U10.
- [155] M.L. Miller, E.J. Molinelli, J.S. Nair, T. Sheikh, R. Samy, X. Jing, Q. He, A. Korkut, A.M. Crago, S. Singer, *et al.*, Drug Synergy Screen and Network Modeling in Dedifferentiated Liposarcoma Identifies CDK4 and IGF1R as Synergistic Drug Targets, *Sci. Signal.* **2013**, 6(294): ra85.
- [156] Z.A. Knight, H. Lin, K.M. Shokat, Targeting the cancer kinome through polypharmacology, *Nat. Rev. Cancer* **2010**, 10(2): 130-137.
- [157] A.L. Hopkins, Network pharmacology: the next paradigm in drug discovery, *Nat. Chem. Biol.* **2008**, 4(11): 682-690.
- [158] A.L. Hopkins, J.S. Mason, J.P. Overington, Can we rationally design promiscuous drugs?, *Curr. Opin. Struct. Biol.* **2006**, 16(1): 127-136.
- [159] J. Besnard, G.F. Ruda, V. Setola, K. Abecassis, R.M. Rodriguiz, X.P. Huang, S. Norval, M.F. Sassano, A.I. Shin, L.A. Webster, *et al.*, Automated design of ligands to polypharmacological profiles, *Nature* **2012**, 492(7428): 215-220.

-
- [160] A. Cortes-Cabrera, G.M. Morris, P.W. Finn, A. Morreale, F. Gago, Comparison of ultra-fast 2D and 3D ligand and target descriptors for side effect prediction and network analysis in polypharmacology, *Br. J. Pharmacol.* **2013**, 170(3): 557-567.
- [161] H.B. Engin, O. Keskin, R. Nussinov, A. Gursoy, A Strategy Based on Protein-Protein Interface Motifs May Help in Identifying Drug Off-Targets, *J. Chem. Inf. Model.* **2012**, 52(8): 2273-2286.
- [162] S.G. Zhang, Fabrication of novel biomaterials through molecular self-assembly, *Nat. Biotechnol.* **2003**, 21(10): 1171-1178.
- [163] A. Fegan, B. White, J.C.T. Carlson, C.R. Wagner, Chemically Controlled Protein Assembly: Techniques and Applications, *Chem. Rev.* **2010**, 110(6): 3315-3336.
- [164] A.E. Nel, L. Madler, D. Velegol, T. Xia, E.M.V. Hoek, P. Somasundaran, F. Klaessig, V. Castranova, M. Thompson, Understanding biophysicochemical interactions at the nano-bio interface, *Nat. Mater.* **2009**, 8(7): 543-557.
- [165] J. Xu, C. Zhu, Y. Zhang, N. Jiang, S. Li, Z. Su, T. Akaike, J. Yang, hE-cadherin-Fc fusion protein coated surface enhances the adhesion and proliferation of human mesenchymal stem cells, *Colloids Surf., B* **2013**, 109C: 97-102.
- [166] N.A. Kotov, J.O. Winter, I.P. Clements, E. Jan, B.P. Timko, S. Campidelli, S. Pathak, A. Mazzatenta, C.M. Lieber, M. Prato, *et al.*, Nanomaterials for Neural Interfaces, *Adv. Mater.* **2009**, 21(40): 3970-4004.
- [167] J. Homola, Surface plasmon resonance sensors for detection of chemical and biological species, *Chem. Rev.* **2008**, 108(2): 462-493.
- [168] A.C. Cheng, R.G. Coleman, K.T. Smyth, Q. Cao, P. Soulard, D.R. Caffrey, A.C. Salzberg, E.S. Huang, Structure-based maximal affinity model predicts small-molecule druggability, *Nat. Biotechnol.* **2007**, 25(1): 71-75.
- [169] P. Chakrabarti, J. Janin, Dissecting protein-protein recognition sites, *Proteins: Struct., Funct., Genet.* **2002**, 47(3): 334-343.
- [170] R.D. Smith, L.G. Hu, J.A. Falkner, M.L. Benson, J.P. Nerothin, H.A. Carlson, Exploring protein-ligand recognition with binding MOAD, *J. Mol. Graphics Modell.* **2006**, 24(6): 414-425.
- [171] L. Lo Conte, C. Chothia, J. Janin, The atomic structure of protein-protein recognition sites, *J. Mol. Biol.* **1999**, 285(5): 2177-2198.
- [172] S. Jones, J.M. Thornton, Principles of protein-protein interactions, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, 93(1): 13-20.
- [173] T. Clackson, J.A. Wells, A Hot-Spot of Binding-Energy in a Hormone-Receptor Interface, *Science* **1995**, 267(5196): 383-386.
- [174] R. Bourgeas, M.J. Basse, X. Morelli, P. Roche, Atomic Analysis of Protein-Protein Interfaces with Known Inhibitors: The 2P2I Database, *PLoS One* **2010**, 5(3): e9598.
- [175] M. Rickert, X.Q. Wang, M.J. Boulanger, N. Goriatcheva, K.C. Garcia, The Structure of Interleukin-2 Complexed with Its Alpha Receptor, *Science* **2005**, 308(5727): 1477-1480.
- [176] C.D. Thanos, M. Randal, J.A. Wells, Potent small-molecule binding to a dynamic hot spot on IL-2, *J. Am. Chem. Soc.* **2003**, 125(50): 15280-15281.
- [177] A. Kahraman, R.J. Morris, R.A. Laskowski, J.M. Thornton, Shape variation in protein binding pockets and their ligands, *J. Mol. Biol.* **2007**, 368(1): 283-301.
- [178] L. Wang, B.J. Berne, R.A. Friesner, Ligand binding to protein-binding pockets with wet and dry regions, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108(4): 1326-1330.
- [179] A. Metz, Entwicklung eines Hybrid-Modells der Solvation für MM-PB/SA Berechnungen, *Diploma Thesis (Goethe Universität, Frankfurt am Main)* **2006**: 1-102.
- [180] R. Norel, S.L. Lin, H.J. Wolfson, R. Nussinov, Shape Complementarity at Protein-Protein Interfaces, *Biopolymers* **1994**, 34(7): 933-940.

-
- [181] M.C. Lawrence, P.M. Colman, Shape Complementarity at Protein-Protein Interfaces, *J. Mol. Biol.* **1993**, 234(4): 946-950.
- [182] I.D. Kuntz, K. Chen, K.A. Sharp, P.A. Kollman, The maximal affinity of ligands, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96(18): 9997-10002.
- [183] P. Buchwald, Small-Molecule Protein-Protein Interaction Inhibitors: Therapeutic Potential in Light of Molecular Size, Chemical Space, and Ligand Binding Efficiency Considerations, *IUBMB Life* **2010**, 62(10): 724-731.
- [184] V. Azzarito, K. Long, N.S. Murphy, A.J. Wilson, Inhibition of α -helix-mediated protein-protein interactions using designed molecules, *Nat. Chem.* **2013**, 5(3): 161-173.
- [185] R.J. Tomko Jr, M. Hochstrasser, Molecular Architecture and Assembly of the Eukaryotic Proteasome, *Annu. Rev. Biochem.* **2013**, 82: 415-445.
- [186] F.U. Hartl, M. Hayer-Hartl, Molecular chaperones in the cytosol: from nascent chain to folded protein, *Science* **2002**, 295(5561): 1852-1858.
- [187] A. Hoelz, E.W. Debler, G. Blobel, The Structure of the Nuclear Pore Complex, *Annu. Rev. Biochem.* **2011**, 80: 613-643.
- [188] T. Palmer, B.C. Berks, The twin-arginine translocation (Tat) protein export pathway, *Nat. Rev. Microbiol.* **2012**, 10(7): 483-496.
- [189] W. Schliebs, W. Girzalsky, R. Erdmann, Peroxisomal protein import and ERAD: variations on a common theme, *Nat. Rev. Mol. Cell Biol.* **2010**, 11(12): 885-890.
- [190] I.D. Kuntz, Structure-Based Strategies for Drug Design and Discovery, *Science* **1992**, 257(5073): 1078-1082.
- [191] T.L. Blundell, Structure-based drug design, *Nature* **1996**, 384(6604): 23-26.
- [192] S. Kalyaanamoorthy, Y.P.P. Chen, Structure-based drug design to augment hit discovery, *Drug Discovery Today* **2011**, 16(17-18): 831-839.
- [193] E.B. Fauman, B.K. Rai, E.S. Huang, Structure-based druggability assessment - identifying suitable targets for small molecule therapeutics, *Curr. Opin. Chem. Biol.* **2011**, 15(4): 463-468.
- [194] S. Henrich, O.M.H. Salo-Ahen, B. Huang, F. Rippmann, G. Cruciani, R.C. Wade, Computational approaches to identifying and characterizing protein binding sites for ligand design, *J. Mol. Recognit.* **2010**, 23(2): 209-219.
- [195] A.T.R. Laurie, R.M. Jackson, Methods for the prediction of protein-ligand binding sites for Structure-Based Drug Design and virtual ligand screening, *Curr. Protein Pept. Sci.* **2006**, 7(5): 395-406.
- [196] T. Geppert, S. Bauer, J.A. Hiss, E. Conrad, M. Reutlinger, P. Schneider, M. Weisel, B. Pfeiffer, K.H. Altmann, Z. Waibler, *et al.*, Immunosuppressive Small Molecule Discovered by Structure-Based Virtual Screening for Inhibitors of Protein-Protein Interactions, *Angew. Chem., Int. Ed.* **2012**, 51(1): 258-261.
- [197] Y.Y. Kuttner, S. Engel, Protein Hot Spots: The Islands of Stability, *J. Mol. Biol.* **2012**, 415(2): 419-428.
- [198] O. Keskin, A. Gursoy, B. Ma, R. Nussinov, Principles of protein-protein interactions: What are the preferred ways for proteins to interact?, *Chem. Rev.* **2008**, 108(4): 1225-1244.
- [199] X. Li, O. Keskin, B.Y. Ma, R. Nussinov, J. Liang, Protein-protein interactions: Hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: Implications for docking, *J. Mol. Biol.* **2004**, 344(3): 781-795.
- [200] M. Weisel, E. Proschak, J.M. Kriegel, G. Schneider, Form follows function: Shape analysis of protein cavities for receptor-based drug design, *Proteomics* **2009**, 9(2): 451-459.

- [201] L.M. Meireles, A.S. Domling, C.J. Camacho, ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery, *Nucleic Acids Res.* **2010**, 38(Web Server issue): W407-411.
- [202] S. Fulle, C. Withers-Martinez, M.J. Blackman, G.M. Morris, P.W. Finn, Molecular Determinants of Binding to the Plasmodium Subtilisin-like Protease 1, *J. Chem. Inf. Model.* **2013**, 53(3): 573-583.
- [203] A. Garai, A. Zeke, G. Gogl, I. Toro, F. Fordos, H. Blankenburg, T. Barkai, J. Varga, A. Alexa, D. Emig, *et al.*, Specificity of Linear Motifs That Bind to a Common Mitogen-Activated Protein Kinase Docking Groove, *Sci. Signaling* **2012**, 5(245): ra74.
- [204] R. Jiang, D. Duckett, W. Chen, J. Habel, Y.Y. Ling, P. LoGrasso, T.M. Kamenecka, 3,5-Disubstituted quinolines as novel c-Jun N-terminal kinase inhibitors, *Bioorg. Med. Chem. Lett.* **2007**, 17(22): 6378-6382.
- [205] M. Drag, G.S. Salvesen, Emerging principles in protease-based drug discovery, *Nat. Rev. Drug Discovery* **2010**, 9(9): 690-701.
- [206] A.C. Dar, K.M. Shokat, The Evolution of Protein Kinase Inhibitors from Antagonists to Agonists of Cellular Signaling, *Annu. Rev. Biochem.* **2011**, 80: 769-795.
- [207] A.C. Backes, B. Zech, B. Felber, B. Klebl, G. Muller, Small-molecule inhibitors binding to protein kinase. Part II: the novel pharmacophore approach of type II and type III inhibition, *Expert Opin. Drug Discovery* **2008**, 3(12): 1427-1449.
- [208] K.C. Han, S.Y. Kim, E.G. Yang, Recent Advances in Designing Substrate-Competitive Protein Kinase Inhibitors, *Curr. Pharm. Des.* **2012**, 18(20): 2875-2882.
- [209] A. Tamrazi, K.E. Carlson, J.R. Daniels, K.M. Hurth, J.A. Katzenellenbogen, Estrogen receptor dimerization: Ligand binding regulates dimer affinity and dimer dissociation rate, *Mol. Endocrinol.* **2002**, 16(12): 2706-2719.
- [210] A. Licht-Murava, H. Eldar-Finkelman, Exploiting substrate recognition for selective inhibition of protein kinases, *Curr. Pharm. Des.* **2012**, 18(20): 2914-2920.
- [211] F. Immekus, L.J. Barandun, M. Betz, F. Debaene, S. Petiot, S. Sanglier-Cianferani, K. Reuter, F. Diederich, G. Klebe, Launching Spiking Ligands into a Protein Protein Interface: A Promising Strategy To Destabilize and Break Interface Formation in a tRNA Modifying Enzyme, *ACS Chem. Biol.* **2013**, 8(6): 1163-1178.
- [212] M.R. Arkin, J.A. Wells, Small-molecule inhibitors of protein-protein interactions: Progressing towards the dream, *Nat. Rev. Drug Discovery* **2004**, 3(4): 301-317.
- [213] M. Shimaoka, T.A. Springer, Therapeutic antagonists and conformational regulation of integrin function, *Nat. Rev. Drug Discovery* **2003**, 2(9): 703-716.
- [214] A.B. Yongye, M. Vivoli, I. Lindberg, J.R. Appel, R.A. Houghten, K. Martinez-Mayorga, Identification of a Small Molecule That Selectively Inhibits Mouse PC2 over Mouse PC1/3: A Computational and Experimental Study, *PLoS One* **2013**, 8(2): e56957.
- [215] G. Collier, V. Ortiz, Emerging computational approaches for the study of protein allostery, *Arch. Biochem. Biophys.* **2013**, 538(1): 6-15.
- [216] M.J. Gorczynski, J. Grembecka, Y.P. Zhou, Y. Kong, L. Roudaia, M.G. Douvas, M. Newman, I. Bielnicka, G. Baber, T. Corpora, *et al.*, Allosteric inhibition of the protein-protein interaction between the leukemia-associated proteins Runx1 and CBF β , *Chem. Biol. (Oxford, U. K.)* **2007**, 14(10): 1186-1197.
- [217] R. Nussinov, C.J. Tsai, Allostery in Disease and in Drug Discovery, *Cell* **2013**, 153(2): 293-305.
- [218] M.R. Arkin, A. Whitty, The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions, *Curr. Opin. Chem. Biol.* **2009**, 13(3): 284-290.

- [219] D. Cardinale, G. Guaitoli, D. Tondi, R. Luciani, S. Henrich, O.M.H. Salo-Ahen, S. Ferrari, G. Marverti, D. Guerrieri, A. Ligabue, *et al.*, Protein-protein interface-binding peptides inhibit the cancer therapy target human thymidylate synthase, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108(34): E542-E549.
- [220] J. Pokorna, L. Machala, P. Rezacova, J. Konvalinka, Current and Novel Inhibitors of HIV Protease, *Viruses* **2009**, 1(3): 1209-1239.
- [221] M. Coppens, J.W. Eikelboom, D. Gustafsson, J.I. Weitz, J. Hirsh, Translational Success Stories Development of Direct Thrombin Inhibitors, *Circ. Res.* **2012**, 111(7): 920-929.
- [222] L.O. Kirkland, C. McInnes, Non-ATP competitive protein kinase inhibitors as anti-tumor therapeutics, *Biochem. Pharmacol.* **2009**, 77(10): 1561-1571.
- [223] A.D. Ferguson, N.A. Larsen, T. Howard, H. Pollard, I. Green, C. Grande, T. Cheung, R. Garcia-Arenas, S. Cowen, J.Q. Wu, *et al.*, Structural Basis of Substrate Methylation and Inhibition of SMYD2, *Structure* **2011**, 19(9): 1262-1273.
- [224] T. Shahian, G.M. Lee, A. Lazic, L.A. Arnold, P. Velusamy, C.M. Roels, R.K. Guy, C.S. Craik, Inhibition of a viral enzyme by a small-molecule dimer disruptor, *Nat. Chem. Biol.* **2009**, 5(9): 640-646.
- [225] D.D. Davey, M. Adler, D. Arnaiz, K. Eagen, S. Erickson, W. Guilford, M. Kenrick, M.M. Morrissey, M. Ohlmeyer, G. Pan, *et al.*, Design, Synthesis, and Activity of 2-Imidazol-1-ylpyrimidine Derived Inducible Nitric Oxide Synthase Dimerization Inhibitors, *J. Med. Chem.* **2007**, 50(6): 1146-1157.
- [226] T.L. Kukar, T.B. Ladd, M.A. Bann, P.C. Fraering, R. Narlawar, G.M. Maharvi, B. Healy, R. Chapman, A.T. Welzel, R.W. Price, *et al.*, Substrate-targeting γ -secretase modulators, *Nature* **2008**, 453(7197): 925-929.
- [227] T. Kodadek, Biochemistry: Molecular cloaking devices, *Nature* **2008**, 453(7197): 861-862.
- [228] M. Pavlaki, K. Debeli, I.E. Triantaphyllidou, N. Klouras, E. Giannopoulou, A.J. Aletras, A proposed mechanism for the inhibitory effect of the anticancer agent titanocene dichloride on tumour gelatinases and other proteolytic enzymes, *JBIC, J. Biol. Inorg. Chem.* **2009**, 14(6): 947-957.
- [229] A. Winter, A.P. Higuero, M. Marsh, A. Sigurdardottir, W.R. Pitt, T.L. Blundell, Biophysical and computational fragment-based approaches to targeting protein-protein interactions: applications in structure-guided drug discovery, *Q. Rev. Biophys.* **2012**, 45(4): 383-426.
- [230] K.-H. Lee, W.-D. Han, K.-J. Kim, B.-H. Oh, PDB ID: 2BZW, The crystal structure of BCL-X_L in complex with full-length BAD, **2005**.
- [231] E.F. Lee, P.E. Czabotar, B.J. Smith, K. Deshayes, K. Zobel, P.M. Colman, W.D. Fairlie, Crystal structure of ABT-737 complexed with Bcl-x(L): implications for selectivity of antagonists of the Bcl-2 family, *Cell Death Differ.* **2007**, 14(9): 1711-1713.
- [232] P.H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A.J. Levine, N.P. Pavletich, Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain, *Science* **1996**, 274(5289): 948-953.
- [233] L.T. Vassilev, B.T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, *et al.*, In vivo activation of the p53 pathway by small-molecule antagonists of MDM2, *Science* **2004**, 303(5659): 844-848.
- [234] M.J. Eck, S.R. Sprang, The Structure of Tumor Necrosis Factor- α at 2.6 Å Resolution: Implications for Receptor-Binding, *J. Biol. Chem.* **1989**, 264(29): 17595-17605.

- [235] M.M. He, A.S. Smith, J.D. Oslob, W.M. Flanagan, A.C. Braisted, A. Whitty, M.T. Cancilla, J. Wang, A.A. Lugovskoy, J.C. Yoburn, *et al.*, Small-molecule inhibition of TNF- α , *Science* **2005**, 310(5750): 1022-1025.
- [236] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, *Nucleic Acids Res.* **2000**, 28(1): 235-242.
- [237] B.N. Bullock, A.L. Jochim, P.S. Arora, Assessing Helical Protein Interfaces for Inhibitor Design, *J. Am. Chem. Soc.* **2011**, 133(36): 14220-14223.
- [238] A.L. Jochim, P.S. Arora, Assessment of helical interfaces in protein-protein interactions, *Mol. BioSyst.* **2009**, 5(9): 924-926.
- [239] C.G. Cummings, A.D. Hamilton, Disrupting protein-protein interactions with non-peptidic, small molecule α -helix mimetics, *Curr. Opin. Chem. Biol.* **2010**, 14(3): 341-346.
- [240] H. Yin, A.D. Hamilton, Strategies for targeting protein-protein interactions with synthetic agents, *Angew. Chem., Int. Ed.* **2005**, 44(27): 4130-4163.
- [241] A.L. Jochim, P.S. Arora, Systematic Analysis of Helical Protein Interfaces Reveals Targets for Synthetic Inhibitors, *ACS Chem. Biol.* **2010**, 5(10): 919-923.
- [242] M. Lanning, S. Fletcher, Recapitulating the α -helix: nonpeptidic, low-molecular-weight ligands for the modulation of helix-mediated protein-protein interactions, *Future Med. Chem.* **2013**, 5(18): 2157-2174.
- [243] M.R. Arkin, M. Randal, W.L. DeLano, J. Hyde, T.N. Luong, J.D. Oslob, D.R. Raphael, L. Taylor, J. Wang, R.S. McDowell, *et al.*, Binding of small molecules to an adaptive protein-protein interface, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100(4): 1603-1608.
- [244] S. Eyrisch, V. Helms, What induces pocket openings on protein surface patches involved in protein-protein interactions?, *J. Comput.-Aided Mol. Des.* **2009**, 23(2): 73-86.
- [245] D.K. Johnson, J. Karanicolas, Druggable Protein Interaction Sites Are More Predisposed to Surface Pocket Formation than the Rest of the Protein Surface, *PLoS Comput. Biol.* **2013**, 9(3): e1002951.
- [246] M. Bista, S. Wolf, K. Khoury, K. Kowalska, Y. Huang, E. Wrona, M. Arciniega, Grzegorz M. Popowicz, Tad A. Holak, A. Dömling, Transient Protein States in Designing Inhibitors of the MDM2-p53 Interaction, *Structure* **2013**, 21(12): 2143-2151.
- [247] S. Eyrisch, V. Helms, Transient pockets on protein surfaces involved in protein-protein interaction, *J. Med. Chem.* **2007**, 50(15): 3457-3464.
- [248] S. Eyrisch, J.L. Medina-Franco, V. Helms, Transient pockets on XIAP-BIR2: toward the characterization of putative binding sites of small-molecule XIAP inhibitors, *J. Mol. Model.* **2012**, 18(5): 2031-2042.
- [249] D.E. Koshland, The Key-Lock Theory and the Induced Fit Theory, *Angew. Chem., Int. Ed.* **1994**, 33(23-24): 2375-2378.
- [250] D.D. Boehr, R. Nussinov, P.E. Wright, The role of dynamic conformational ensembles in biomolecular recognition, *Nat. Chem. Biol.* **2009**, 5(11): 789-796.
- [251] R. Nussinov, B. Ma, C.-J. Tsai, Multiple conformational selection and induced fit events take place in allosteric propagation, *Biophys. Chem.* **2014**, 186: 22-30.
- [252] D.B. Kokh, S. Richter, S. Henrich, P. Czodrowski, F. Rippmann, R.C. Wade, TRAPP: A Tool for Analysis of Transient Binding Pockets in Proteins, *J. Chem. Inf. Model.* **2013**, 53(5): 1235-1252.
- [253] X.L. Zheng, L.F. Gan, E.K. Wang, J. Wang, Pocket-Based Drug Design: Exploring Pocket Space, *AAPS J.* **2013**, 15(1): 228-241.

-
- [254] A. Schön, S.Y. Lam, E. Freire, Thermodynamics-based drug design: strategies for inhibiting protein-protein interactions, *Future Med. Chem.* **2011**, 3(9): 1129-1137.
- [255] P.L. Kastiris, I.H. Moal, H. Hwang, Z.P. Weng, P.A. Bates, A.M.J.J. Bonvin, J. Janin, A structure-based benchmark for protein-protein binding affinity, *Protein Sci.* **2011**, 20(3): 482-491.
- [256] K.N. Houk, A.G. Leach, S.P. Kim, X.Y. Zhang, Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes, *Angew. Chem., Int. Ed.* **2003**, 42(40): 4872-4897.
- [257] S. Ghaemmaghami, W. Huh, K. Bower, R.W. Howson, A. Belle, N. Dephoure, E.K. O'Shea, J.S. Weissman, Global analysis of protein expression in yeast, *Nature* **2003**, 425(6959): 737-741.
- [258] T. Taniguchi, A. Takaoka, A weak signal for strong responses: Interferon- α/β revisited, *Nat. Rev. Mol. Cell Biol.* **2001**, 2(5): 378-386.
- [259] M. Tsudo, R.W. Kozak, C.K. Goldman, T.A. Waldmann, Demonstration of a non-Tac peptide that binds interleukin 2: A potential participant in a multichain interleukin 2 receptor complex, *Proc. Natl. Acad. Sci. U. S. A.* **1986**, 83(24): 9694-9698.
- [260] J.G. Cannon, Inflammatory cytokines in nonpathological states, *News Physiol. Sci.* **2000**, 15: 298-303.
- [261] R. Hussein, H.N. Lim, Direct comparison of small RNA and transcription factor signaling, *Nucleic Acids Res.* **2012**, 40(15): 7269-7279.
- [262] W. Gilbert, B. Muller-Hill, Isolation of the lac repressor, *Proc. Natl. Acad. Sci. U. S. A.* **1966**, 56(6): 1891-1898.
- [263] Y. Taniguchi, P.J. Choi, G.W. Li, H.Y. Chen, M. Babu, J. Hearn, A. Emili, X.S. Xie, Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells, *Science* **2010**, 329(5991): 533-538.
- [264] W.W. Fisher, J.J. Li, A.S. Hammonds, J.B. Brown, B.D. Pfeiffer, R. Weizmann, S. MacArthur, S. Thomas, J.A. Stamatoyannopoulos, M.B. Eisen, *et al.*, DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109(52): 21330-21335.
- [265] T. Maiwald, A. Schneider, H. Busch, S. Sahle, N. Gretz, T.S. Weiss, U. Kummer, U. Klingmuller, Combining theoretical analysis and experimental data generation reveals IRF9 as a crucial factor for accelerating interferon α -induced early antiviral signalling, *FEBS Journal* **2010**, 277(22): 4741-4754.
- [266] T. Rodgers, M. Rowland, Mechanistic approaches to volume of distribution predictions: Understanding the processes, *Pharm. Res.* **2007**, 24(5): 918-933.
- [267] S.J. Sollott, L. Cheng, R.R. Pauly, G.M. Jenkins, R.E. Monticone, M. Kuzuya, J.P. Froehlich, M.T. Crow, E.G. Lakatta, E.K. Rowinsky, *et al.*, Taxol Inhibits Neointimal Smooth-Muscle Cell Accumulation after Angioplasty in the Rat, *J. Clin. Invest.* **1995**, 95(4): 1869-1876.
- [268] C. Benson, S. Kaye, P. Workman, M. Garrett, M. Walton, J. de Bono, Clinical anticancer drug development: targeting the cyclin-dependent kinases, *Br. J. Cancer* **2005**, 92(1): 7-12.
- [269] B.A. Chabner, D.D.L. Longo (2010) *Cancer Chemotherapy and Biotherapy: Principles and Practice* (Wolters Kluwer Health/Lippincott Williams & Wilkins) 5 Ed p 812.
- [270] M. Pichler, Z.Y. Wang, C. GrabnerWeiss, D. Reimer, S. Hering, M. Grabner, H. Glossmann, J. Striessnig, Block of P/Q-type calcium channels by therapeutic concentrations of aminoglycoside antibiotics, *Biochemistry* **1996**, 35(46): 14659-14664.

-
- [271] T. Kobayashi, K. Ikeda, T. Kumanishi, Inhibition by various antipsychotic drugs of the G-protein-activated inwardly rectifying K^+ (GIRK) channels expressed in *Xenopus* oocytes, *Br. J. Pharmacol.* **2000**, 129(8): 1716-1722.
- [272] M. Rowland, C. Peck, G. Tucker, Physiologically-Based Pharmacokinetics in Drug Development and Regulatory Science, *Annu. Rev. Pharmacol. Toxicol.* **2011**, 51: 45-73.
- [273] Q. Chen, M.G. Espey, A.Y. Sun, J.H. Lee, M.C. Krishna, E. Shacter, P.L. Choyke, C. Pooput, K.L. Kirk, G.R. Buettner, *et al.*, Ascorbate in pharmacologic concentrations selectively generates ascorbate radical and hydrogen peroxide in extracellular fluid *in vivo*, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104(21): 8749-8754.
- [274] C.T. Dollery, Intracellular Drug Concentrations, *Clin. Pharmacol. Ther.* **2013**, 93(3): 263-266.
- [275] R. Matesanz, I. Barasoain, C.G. Yang, L. Wang, X. Li, C. De Ines, C. Coderch, F. Gago, J.J. Barbero, J.M. Andreu, *et al.*, Optimization of taxane binding to microtubules: Binding affinity dissection and incremental construction of a high-affinity analog of paclitaxel, *Chem. Biol.* **2008**, 15(6): 573-585.
- [276] G. Chiosis, H. Huezo, N. Rosen, E. Mimnaugh, L. Whitesell, L. Neckers, 17AAG: Low target binding affinity and potent cell activity - Finding an explanation, *Mol. Cancer Ther.* **2003**, 2(2): 123-129.
- [277] Z.X. Wang, An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule, *FEBS Lett.* **1995**, 360(2): 111-114.
- [278] G. Lessene, P.E. Czabotar, B.E. Sleebs, K. Zobel, K.N. Lowes, J.M. Adams, J.B. Baell, P.M. Colman, K. Deshayes, W.J. Fairbrother, *et al.*, Structure-guided design of a selective BCL-X_L inhibitor, *Nat. Chem. Biol.* **2013**, 9(6): 390-397.
- [279] M.D.S. Kumar, M.M. Gromiha, PINT: Protein-protein interactions thermodynamic database, *Nucleic Acids Res.* **2006**, 34: D195-D198.
- [280] C.G. Whiteley, Enzyme kinetics: Partial and complete competitive inhibition, *Biochem. Educ.* **1997**, 25(3): 144-146.
- [281] S. Mitternacht, I.N. Berezovsky, Coherent Conformational Degrees of Freedom as a Structural Basis for Allosteric Communication, *PLoS Comput. Biol.* **2011**, 7(12): e1002301.
- [282] K.D. Wilkinson, Quantitative analysis of protein-protein interactions, *Methods Mol. Biol. (N. Y.)* **2004**, 261: 15-32.
- [283] A.P. Minton, The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media, *J. Biol. Chem.* **2001**, 276(14): 10577-10580.
- [284] H.X. Zhou, G. Rivas, A.P. Minton, Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences, *Annu. Rev. Biophys.* **2008**, 37: 375-397.
- [285] P.L. Kastiris, A.M. Bonvin, On the binding affinity of macromolecular interactions: daring to ask why proteins interact, *J. R. Soc., Interface* **2013**, 10(79): 20120835.
- [286] D. Bray, Signaling complexes: Biophysical constraints on intracellular communication, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, 27: 59-75.
- [287] M. Lewitzky, P.C. Simister, S.M. Feller, Beyond 'furballs' and 'dumpling soups' - towards a molecular architecture of signaling complexes and networks, *FEBS Lett.* **2012**, 586(17): 2740-2750.
- [288] S.R. McGuffee, A.H. Elcock, Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm, *PLoS Comput. Biol.* **2010**, 6(3): e1000694.

-
- [289] K. Simons, D. Toomre, Lipid rafts and signal transduction, *Nat. Rev. Mol. Cell Biol.* **2000**, 1(1): 31-39.
 - [290] S. Nagel, L. Hambach, J. Krauter, L. Venturini, O. Heidenreich, A. Ganser, G. Heil, Analysis of the nuclear distribution of the translocation t(8;21)-derived fusion protein AML1/ETO by confocal laser scanning microscopy, *J. Hematother. Stem Cell Res.* **2002**, 11(2): 401-408.
 - [291] D.E. Koshland, K. Hamadani, Proteomics and models for enzyme cooperativity, *J. Biol. Chem.* **2002**, 277(49): 46841-46844.
 - [292] R.A. Veitia, A sigmoidal transcriptional response: cooperativity, synergy and dosage effects, *Biol. Rev. Cambridge Philos. Soc.* **2003**, 78(1): 149-170.
 - [293] T. Shibata, K. Fujimoto, Noisy signal amplification in ultrasensitive signal transduction, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102(2): 331-336.
 - [294] J. Momand, G.P. Zambetti, D.C. Olson, D. George, A.J. Levine, The Mdm-2 Oncogene Product Forms a Complex with the P53 Protein and Inhibits P53-Mediated Transactivation, *Cell* **1992**, 69(7): 1237-1245.
 - [295] S. Hooshangi, S. Thiberge, R. Weiss, Ultrasensitivity and noise propagation in a synthetic transcriptional cascade, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102(10): 3581-3586.
 - [296] G. Chalancon, C.N.J. Ravarani, S. Balaji, A. Martinez-Arias, L. Aravind, R. Jothi, M.M. Babu, Interplay between gene expression noise and regulatory network architecture, *Trends Genet.* **2012**, 28(5): 221-232.
 - [297] S.M. Millard, S.A. Wood, Riding the DUBway: regulation of protein trafficking by deubiquitylating enzymes, *J. Cell. Biol.* **2006**, 173(4): 463-468.
 - [298] S.B. Van Albada, P.R. Ten Wolde, Enzyme localization can drastically affect signal amplification in signal transduction pathways, *PLoS Comput. Biol.* **2007**, 3(10): 1925-1934.
 - [299] F. Ricci, A. Vallee-Belisle, K.W. Plaxco, High-Precision, *In Vitro* Validation of the Sequestration Mechanism for Generating Ultrasensitive Dose-Response Curves in Regulatory Networks, *PLoS Comput. Biol.* **2011**, 7(10): e1002171.
 - [300] N.E. Buchler, F.R. Cross, Protein sequestration generates a flexible ultrasensitive response in a genetic network, *Mol. Syst. Biol.* **2009**, 5(1): 272.
 - [301] B. Graves, T. Thompson, M.X. Xia, C. Janson, C. Lukacs, D. Deo, P. Di Lello, D. Fry, C. Garvie, K.S. Huang, *et al.*, Activation of the p53 pathway by small-molecule-induced MDM2 and MDMX dimerization, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109(29): 11788-11793.
 - [302] C. Gomez-Urbe, G.C. Verghese, L.A. Mirny, Operating regimes of signaling cycles: Statics, dynamics, and noise filtering, *PLoS Comput. Biol.* **2007**, 3(12): 2487-2497.
 - [303] B.N. Kholodenko, J.F. Hancock, W. Kolch, Signalling ballet in space and time, *Nat. Rev. Mol. Cell Biol.* **2010**, 11(6): 414-426.
 - [304] B.N. Kholodenko, Cell-signalling dynamics in time and space, *Nat. Rev. Mol. Cell Biol.* **2006**, 7(3): 165-176.
 - [305] A. Goldbeter, D.E. Koshland, An Amplified Sensitivity Arising from Covalent Modification in Biological-Systems, *Proc. Natl. Acad. Sci. U. S. A.* **1981**, 78(11): 6840-6844.
 - [306] H.J. Bai, K. Yang, D.Q. Yu, C.S. Zhang, F.J. Chen, L.H. Lai, Predicting kinetic constants of protein-protein interactions based on structural properties, *Proteins: Struct., Funct., Bioinf.* **2011**, 79(3): 720-734.
 - [307] A.V. Veselovsky, Y.D. Ivanov, A.S. Ivanov, A.I. Archakov, P. Lewi, P. Janssen, Protein-protein interactions: mechanisms and modification by drugs, *J. Mol. Recognit.* **2002**, 15(6): 405-422.

-
- [308] C. Frieden, Kinetic Aspects of Regulation of Metabolic Processes: THE HYSTERETIC ENZYME CONCEPT, *J. Biol. Chem.* **1970**, 245(21): 5788-5799.
 - [309] P. Walter, O. Ulucan, J. Metzger, V. Helms, Bioinformatics of Protein-Protein Interfaces and Small Molecule Effectors, *Curr. Bioinf.* **2012**, 7(2): 159-172.
 - [310] M.D. Holdom, A.M. Davies, J.E. Nettleship, S.C. Bagby, B. Dhaliwal, E. Girardi, J. Hunt, H.J. Gould, A.J. Beavil, J.M. McDonnell, *et al.*, Conformational changes in IgE contribute to its uniquely slow dissociation rate from receptor FcεRI, *Nat. Struct. Mol. Biol.* **2011**, 18(5): 571–576.
 - [311] B. Kim, A. Eggel, S.S. Tarchevskaya, M. Vogel, H. Prinz, T.S. Jardetzky, Accelerated disassembly of IgE-receptor complexes by a disruptive macromolecular inhibitor, *Nature* **2012**, 491(7425): 613-619.
 - [312] Y. Bartel, M. Grez, C. Wichmann, Interference with RUNX1/ETO Leukemogenic Function by Cell-Penetrating Peptides Targeting the NHR2 Oligomerization Domain, *Biomed. Res. Int.* **2013**, 2013: Article ID 297692.
 - [313] A.K. Dunker, C.J. Oldfield, J.W. Meng, P. Romero, J.Y. Yang, J.W. Chen, V. Vacic, Z. Obradovic, V.N. Uversky, The unfoldomics decade: an update on intrinsically disordered proteins, *BMC Genomics* **2008**, 9(Suppl 2): S1.
 - [314] J. Gsponer, M.M. Babu, The rules of disorder or why disorder rules, *Prog. Biophys. Mol. Biol.* **2009**, 99(2-3): 94-103.
 - [315] A.K. Dunker, M.S. Cortese, P. Romero, L.M. Iakoucheva, V.N. Uversky, Flexible nets: The roles of intrinsic disorder in protein interaction networks, *FEBS J.* **2005**, 272(20): 5129-5148.
 - [316] O. Keskin, B.Y. Ma, K. Rogale, K. Gunasekaran, R. Nussinov, Protein-protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach, *Phys. Biol.* **2005**, 2(2): S24-S35.
 - [317] T. Haliloglu, O. Keskin, B.Y. Ma, R. Nussinov, How similar are protein folding and protein binding nuclei? Examination of vibrational motions of energy hot spots and conserved residues, *Biophys. J.* **2005**, 88(3): 1552-1559.
 - [318] C.J. Tsai, D. Xu, R. Nussinov, Protein folding via binding and *vice versa*, *Folding Des.* **1998**, 3(4): R71-R80.
 - [319] A. Nicholls, K.A. Sharp, B. Honig, Protein Folding and Association: Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons, *Proteins: Struct., Funct., Genet.* **1991**, 11(4): 281-296.
 - [320] J. Janin, R.P. Bahadur, P. Chakrabarti, Protein-protein interaction and quaternary structure, *Q. Rev. Biophys* **2008**, 41(2): 133-180.
 - [321] D.M. Taverna, R.A. Goldstein, Why are proteins marginally stable?, *Proteins: Struct., Funct., Genet.* **2002**, 46(1): 105-109.
 - [322] P.D. Williams, D.D. Pollock, R.A. Goldstein, Functionality and the evolution of marginal stability in proteins: Inferences from lattice simulations, *Evol. Bioinform.* **2006**, 2: 91-101.
 - [323] A.R. Fersht, L. Serrano, Principles of Protein Stability Derived from Protein Engineering Experiments, *Curr. Opin. Struct. Biol.* **1993**, 3(1): 75-83.
 - [324] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol.* **2005**, 6(3): 197-208.
 - [325] X. Morelli, R. Bourgeas, P. Roche, Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I), *Curr. Opin. Chem. Biol.* **2011**, 15(4): 475-481.
 - [326] O. Sperandio, C.H. Reynes, A.C. Camproux, B.O. Villoutreix, Rationalizing the chemical space of protein-protein interaction inhibitors, *Drug Discovery Today* **2010**, 15(5-6): 220-229.

- [327] A. Voet, F. Berenger, K.Y. Zhang, Electrostatic Similarities between Protein and Small Molecule Ligands Facilitate the Design of Protein-Protein Interaction Inhibitors, *PLoS One* **2013**, 8(10): e75762.
- [328] G.R. Bickerton, A.P. Higuero, T.L. Blundell, Comprehensive, atomic-level characterization of structurally characterized protein-protein interactions: the PICCOLO database, *BMC Bioinf.* **2011**, 12: 313.
- [329] R.P. Bahadur, P. Chakrabarti, F. Rodier, J. Janin, A Dissection of Specific and Non-specific Protein-Protein Interfaces, *J. Mol. Biol.* **2004**, 336(4): 943-955.
- [330] S. Dey, A. Pal, P. Chakrabarti, J. Janin, The Subunit Interfaces of Weakly Associated Homodimeric Proteins, *J. Mol. Biol.* **2010**, 398(1): 146-160.
- [331] M.J. Vicent, E. Perez-Paya, M. Orzaez, Discovery of inhibitors of protein-protein interactions from combinatorial libraries, *Curr. Trends Med. Chem.* **2007**, 7(1): 83-95.
- [332] C.J. Tsai, S.L. Lin, H.J. Wolfson, R. Nussinov, Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect, *Protein Sci.* **1997**, 6(1): 53-64.
- [333] L. Ezkurdia, L. Bartoli, P. Fariselli, R. Casadio, A. Valencia, M.L. Tress, Progress and challenges in predicting protein-protein interaction sites, *Briefings Bioinf.* **2009**, 10(3): 233-246.
- [334] J. Janin, Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition, *Structure* **1999**, 7(12): R277-R279.
- [335] Z.H. Li, Y. He, L. Wong, J.Y. Li, Progressive dry-core-wet-rim hydration trend in a nested-ring topology of protein binding interfaces, *BMC Bioinf.* **2012**, 13.
- [336] Y. Levy, J.N. Onuchic, Water mediation in protein folding and molecular recognition, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, 35: 389-415.
- [337] M. Ahmad, W. Gu, V. Helms, Mechanism of fast peptide recognition by SH3 domains, *Angew. Chem., Int. Ed. Engl.* **2008**, 47(40): 7626-7630.
- [338] M. Ahmad, W. Gu, T. Geyer, V. Helms, Adhesive water networks facilitate binding of protein interfaces, *Nat. Commun.* **2011**, 2: 261.
- [339] M.F. Lensink, I.H. Moal, P.A. Bates, P.L. Kastiris, A.S. Melquiond, E. Karaca, C. Schmitz, M. van Dijk, A.M. Bonvin, M. Eisenstein, *et al.*, Blind prediction of interfacial water positions in CAPRI, *Proteins* **2014**, 82(4): 620-632.
- [340] I.S. Moreira, J.M. Martins, R.M. Ramos, P.A. Fernandes, M.J. Ramos, Are Hot-Spots Occluded from Water?, *Biophys. J.* **2013**, 104(2): 505a-505a.
- [341] H. Gohlke, G. Klebe, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, *Angew. Chem., Int. Ed. Engl.* **2002**, 41(15): 2645-2676.
- [342] M.S. Golden, S.M. Cote, M. Sayeg, B.S. Zerbe, E.A. Villar, D. Beglov, S.L. Sazinsky, R.M. Georgiadis, S. Vajda, D. Kozakov, *et al.*, Comprehensive Experimental and Computational Analysis of Binding Energy Hot Spots at the NF- κ B Essential Modulator/IKK β Protein-Protein Interface, *J. Am. Chem. Soc.* **2013**, 135(16): 6242-6256.
- [343] B.S. Zerbe, D.R. Hall, S. Vajda, A. Whitty, D. Kozakov, Relationship between Hot Spot Residues and Ligand Binding Hot Spots in Protein-Protein Interfaces, *J. Chem. Inf. Model.* **2012**, 52(8): 2236-2244.
- [344] H. Gohlke, M. Hendlich, G. Klebe, Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function, *Perspect. Drug Discovery Des.* **2000**, 20(1): 115-144.
- [345] S. Grüneberg, B. Wendt, G. Klebe, Subnanomolar inhibitors from computer screening: A model study using human carbonic anhydrase II, *Angew. Chem., Int. Ed. Engl.* **2001**, 40(2): 389-393.

-
- [346] B.C. Cunningham, J.A. Wells, Comparison of a Structural and a Functional Epitope, *J. Mol. Biol.* **1993**, 234(3): 554-563.
- [347] S.H. Bass, M.G. Mulkerrin, J.A. Wells, A Systematic Mutational Analysis of Hormone-Binding Determinants in the Human Growth-Hormone Receptor, *Proc. Natl. Acad. Sci. U. S. A.* **1991**, 88(10): 4498-4502.
- [348] B.C. Cunningham, J.A. Wells, High-Resolution Epitope Mapping of hGH-Receptor Interactions by Alanine-Scanning Mutagenesis, *Science* **1989**, 244(4908): 1081-1085.
- [349] O. Keskin, B.Y. Ma, R. Nussinov, Hot regions in protein-protein interactions: The organization and contribution of structurally conserved hot spot residues, *J. Mol. Biol.* **2005**, 345(5): 1281-1294.
- [350] D.R. Koes, C.J. Camacho, PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure, *Nucleic Acids Res.* **2012**, 40(W1): W387-W392.
- [351] N.I. Vasilevich, I.I. Afanasyev, E.A. Rastorguev, D.V. Genis, V.S. Kochubey, Dual mode of action of phenyl-pyrazole-phenyl (6-5-6 system)-based PPI inhibitors: alpha-helix backbone *versus* alpha-helix binding epitope, *MedChemComm* **2013**, 4: 1597-1603.
- [352] Y. Li, Z. Liu, L. Han, C. Li, R. Wang, Mining the Characteristic Interaction Patterns on Protein-Protein Binding Interfaces, *J. Chem. Inf. Model.* **2013**, 53(9): 2437-2447.
- [353] J.Y. Li, Q. Liu, 'Double water exclusion': a hypothesis refining the O-ring theory for the hot spots at protein interfaces, *Bioinformatics* **2009**, 25(6): 743-750.
- [354] I.S. Moreira, J.M. Martins, R.M. Ramos, P.A. Fernandes, M.J. Ramos, Understanding the importance of the aromatic amino-acid residues as hot-spots, *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, 1834(1): 404-414.
- [355] E.F. Lee, J.D. Sadowsky, B.J. Smith, P.E. Czabotar, K.J. Peterson-Kaufman, P.M. Colman, S.H. Gellman, W.D. Fairlie, High-Resolution Structural Characterization of a Helical α/β -Peptide Foldamer Bound to the Anti-Apoptotic Protein Bcl-x_L, *Angew. Chem., Int. Ed. Engl.* **2009**, 48(24): 4318-4322.
- [356] R.R. Thangudu, S.H. Bryant, A.R. Panchenko, T. Madej, Modulating Protein-Protein Interactions with Small Molecules: The Importance of Binding Hotspots, *J. Mol. Biol.* **2012**, 415(2): 443-453.
- [357] D. Bier, R. Rose, K. Bravo-Rodriguez, M. Bartel, J.M. Ramirez-Anguita, S. Dutt, C. Wilch, F.G. Klarner, E. Sanchez-Garcia, T. Schrader, *et al.*, Molecular tweezers modulate 14-3-3 protein-protein interactions, *Nat. Chem.* **2013**, 5(3): 234-239.
- [358] C.M. Bergey, A.M. Watkins, P.S. Arora, HippDB: a database of readily targeted helical protein-protein interactions, *Bioinformatics* **2013**, 29(21): 2806-2807.
- [359] L.G. Milroy, L. Brunsveld, Pharmaceutical implications of helix length control in helix-mediated protein-protein interactions, *Future Med. Chem.* **2013**, 5(18): 2175-2183.
- [360] A.S. Yang, B. Honig, Free-Energy Determinants of Secondary Structure Formation: I. α -Helices, *J. Mol. Biol.* **1995**, 252(3): 351-365.
- [361] L. Pauling, R.B. Corey, H.R. Branson, The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain, *Proc. Natl. Acad. Sci. U. S. A.* **1951**, 37(4): 205-211.
- [362] J.M. Adams, S. Cory, The Bcl-2 protein family: Arbiters of cell survival, *Science* **1998**, 281(5381): 1322-1326.
- [363] W.H. Landschulz, P.F. Johnson, S.L. Mcknight, The Leucine Zipper: A Hypothetical Structure Common to a New Class of DNA Binding Proteins, *Science* **1988**, 240(4860): 1759-1764.

-
- [364] N. London, B. Raveh, O. Schueler-Furman, Druggable protein-protein interactions - from hot spots to hot segments, *Curr. Opin. Chem. Biol.* **2013**, 17(6): 952–959.
 - [365] N. London, B. Raveh, D. Movshovitz-Attias, O. Schueler-Furman, Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions?, *Proteins: Struct., Funct., Bioinf.* **2010**, 78(15): 3140-3149.
 - [366] M. Althage, T. Bizouarn, B. Kindlund, J. Mullins, J. Ålander, J. Rydström, Cross-linking of transmembrane helices in proton-translocating nicotinamide nucleotide transhydrogenase from *Escherichia coli*: implications for the structure and function of the membrane domain, *Biochim. Biophys. Acta, Bioenerg.* **2004**, 1659(1): 73-82.
 - [367] A.N. Lupas, M. Gruber, The structure of α -helical coiled coils, *Adv. Protein Chem.* **2005**, 70: 37-78.
 - [368] T.J. Richmond, F.M. Richards, Packing of α -Helices: Geometrical Constraints and Contact Areas, *J. Mol. Biol.* **1978**, 119(4): 537-555.
 - [369] L. Caboni, B. Egan, B. Kelly, F. Blanco, D. Fayne, M.J. Meegan, D.G. Lloyd, Structure-Activity Relationships in Non-Ligand Binding Pocket (Non-LBP) Diarylhydrazide Antiandrogens, *J. Chem. Inf. Model.* **2013**, 53(8): 2116-2130.
 - [370] H. Gronemeyer, J.A. Gustafsson, V. Laudet, Principles for modulation of the nuclear receptor superfamily, *Nat. Rev. Drug Discovery* **2004**, 3(11): 950-964.
 - [371] F. Blanco, B. Egan, L. Caboni, J. Elguero, J. O'Brien, T. McCabe, D. Fayne, M.J. Meegan, D.G. Lloyd, Study of *E/Z* Isomerization in a Series of Novel Non-ligand Binding Pocket Androgen Receptor Antagonists, *J. Chem. Inf. Model.* **2012**, 52(9): 2387-2397.
 - [372] L. Caboni, G.K. Kinsella, F. Blanco, D. Fayne, W.N. Jagoe, M. Carr, D.C. Williams, M.J. Meegan, D.G. Lloyd, "True" Antiandrogens-Selective Non-Ligand-Binding Pocket Disruptors of Androgen Receptor-Coactivator Interactions: Novel Tools for Prostate Cancer, *J. Med. Chem.* **2012**, 55(4): 1635-1644.
 - [373] C.D. Thanos, W.L. DeLano, J.A. Wells, Hot-spot mimicry of a cytokine receptor by a small molecule, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, 103(42): 15422-15427.
 - [374] O. Boyman, J. Sprent, The role of interleukin-2 during homeostasis and activation of the immune system, *Nat. Rev. Immunol.* **2012**, 12(3): 180-190.
 - [375] W.J. Murphy, J.L.M. Ferrara, T. Malek, A delicate balance: tweaking IL-2 immunotherapy, *Nat. Med.* **2012**, 18(2): 208-209.
 - [376] T.R. Malek, I. Castro, Interleukin-2 Receptor Signaling: At the Interface between Tolerance and Immunity, *Immunity* **2010**, 33(2): 153-165.
 - [377] K.A. Smith, Interleukin-2: Inception, Impact, and Implications, *Science* **1988**, 240(4856): 1169-1176.
 - [378] D.J. Stauber, E.W. Debler, P.A. Horton, K.A. Smith, I.A. Wilson, Crystal structure of the IL-2 signaling complex: Paradigm for a heterotrimeric cytokine receptor, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, 103(8): 2788-2793.
 - [379] A.M. Levin, D.L. Bates, A.M. Ring, C. Krieg, J.T. Lin, L. Su, I. Moraga, M.E. Raeber, G.R. Bowman, P. Novick, *et al.*, Exploiting a natural conformational switch to engineer an interleukin-2 'superkine', *Nature* **2012**, 484(7395): 529-535.
 - [380] S. Kalsoom, U. Rashid, A. Shaikat, O.M. Abdalla, K. Hussain, W. Khan, S. Nazir, M.A. Mesaik, Zaheer-ul-Haq, F.L. Ansari, In vitro and in silico exploration of IL-2 inhibition by small drug-like molecules, *Med. Chem. Res.* **2013**, 22: 5739-5751.
 - [381] K.M. Murphy (2011) Manipulation of the Immune Response. *Janeway's Immunobiology*, (Garland Science, New York), 8 Ed, pp 669-716.
 - [382] B.C. Raimundo, J.D. Oslob, A.C. Braisted, J. Hyde, R.S. McDowell, M. Randal, N.D. Waal, J. Wilkinson, C.H. Yu, M.R. Arkin, Integrating fragment assembly and biophysical methods in the chemical advancement of small-molecule antagonists of

- IL-2: An approach for inhibiting protein-protein interactions, *J. Med. Chem.* **2004**, 47(12): 3111-3130.
- [383] A.C. Braisted, J.D. Oslob, W.L. Delano, J. Hyde, R.S. McDowell, N. Waal, C. Yu, M.R. Arkin, B.C. Raimundo, Discovery of a potent small molecule IL-2 inhibitor through fragment assembly, *J. Am. Chem. Soc.* **2003**, 125(13): 3714-3715.
- [384] M. Arkin, J.D. Lear, A new data analysis method to determine binding constants of small molecules to proteins using equilibrium analytical ultracentrifugation with absorption optics, *Anal. Biochem.* **2001**, 299(1): 98-107.
- [385] K. Sauve, M. Nachman, C. Spence, P. Bailon, E. Campbell, W.H. Tsien, J.A. Kondas, J. Hakimi, G. Ju, Localization in Human Interleukin-2 of the Binding-Site to the α -Chain (P55) of the Interleukin-2 Receptor, *Proc. Natl. Acad. Sci. U. S. A.* **1991**, 88(11): 4636-4640.
- [386] J.W. Tilley, L. Chen, D.C. Fry, S.D. Emerson, G.D. Powers, D. Biondi, T. Varnell, R. Trilles, R. Guthrie, F. Mennona, *et al.*, Identification of a Small Molecule Inhibitor of the IL-2/IL-2R α Receptor Interaction Which Binds to IL-2, *J. Am. Chem. Soc.* **1997**, 119(32): 7589-7590.
- [387] C.G. Wilson, M.R. Arkin, Probing structural adaptivity at PPI interfaces with small molecules, *Drug Discovery Today: Technol.* **2013**, 10(4): e501-e508.
- [388] D.A. Erlanson, J.A. Wells, A.C. Braisted, Tethering: Fragment-based drug discovery, *Annu. Rev. Biophys. Biomol. Struct.* **2004**, 33: 199-223.
- [389] D.A. Erlanson, A.C. Braisted, D.R. Raphael, M. Randal, R.M. Stroud, E.M. Gordon, J.A. Wells, Site-directed ligand discovery, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97(17): 9367-9372.
- [390] T. Berg, Use of "tethering" for the identification of a small molecule that binds to a dynamic hot spot on the interleukin-2 surface, *ChemBioChem* **2004**, 5(8): 1051-1053.
- [391] D.C. Fry, C. Wartchow, B. Graves, C. Janson, C. Lukacs, U. Kammlott, C. Belunis, S. Palme, C. Klein, B. Vu, Deconstruction of a Nutlin: Dissecting the Binding Determinants of a Potent Protein-Protein Interaction Inhibitor, *ACS Med. Chem. Lett.* **2013**, 4(7): 95-100.
- [392] H. Miyoshi, T. Kozu, K. Shimizu, K. Enomoto, N. Maseki, Y. Kaneko, N. Kamada, M. Ohki, The t(8;21) Translocation in Acute Myeloid-Leukemia Results in Production of an AML1-MTG8 Fusion Transcript, *EMBO J.* **1993**, 12(7): 2715-2721.
- [393] J.M. Bennett, D. Catovsky, M.T. Daniel, G. Flandrin, D.A. Galton, H.R. Gralnick, C. Sultan, Proposals for the classification of the acute leukaemias. French-American-British (FAB) Co-operative Group, *Br. J. Haematol.* **1976**, 33(4): 451-458.
- [394] S.E. Langabeer, H. Walker, J.R. Rogers, A.K. Burnett, K. Wheatley, D. Swirsky, A.H. Goldstone, D.C. Linch, Incidence of AML1/ETO fusion transcripts in patients entered into the MRC AML trials, *Br. J. Haematol.* **1997**, 99(4): 925-928.
- [395] S. Doulatov, F. Notta, E. Laurenti, J.E. Dick, Hematopoiesis: A Human Perspective, *Cell Stem Cell* **2012**, 10(2): 120-136.
- [396] S.H. Orkin, L.I. Zon, Hematopoiesis: An evolving paradigm for stem cell biology, *Cell* **2008**, 132(4): 631-644.
- [397] M. Ichikawa, T. Asai, S. Chiba, M. Kurokawa, S. Ogawa, Runx1/AML-1 ranks as a master regulator of adult hematopoiesis, *Cell Cycle* **2004**, 3(6): 722-724.
- [398] M. Lichtinger, R. Ingram, R. Hannah, D. Muller, D. Clarke, S.A. Assi, M. Lie-A-Ling, L. Noailles, M.S. Vijayabaskar, M.C. Wu, *et al.*, RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis, *EMBO J.* **2012**, 31(22): 4318-4333.
- [399] T. Yamagata, K. Maki, K. Mitani, Runx1/AML1 in Normal and Abnormal Hematopoiesis, *Int. J. Hematol.* **2005**, 82(1): 1-8.

-
- [400] C. Wichmann, Y. Becker, L. Chen-Wichmann, V. Vogel, A. Vojtkova, J. Herglotz, S. Moore, J. Koch, J. Lausen, W. Mantele, *et al.*, Dimer-tetramer transition controls RUNX1/ETO leukemogenic activity, *Blood* **2010**, 116(4): 603-613.
 - [401] J.N. Davis, L. McGhee, S. Meyers, The ETO (*MTG8*) gene family, *Gene* **2003**, 303: 1-10.
 - [402] C. Wichmann, L.P. Chen, M. Heinrich, D. Baus, E. Pfitzner, M. Zornig, O.G. Ottmann, M. Grez, Targeting the oligomerization domain of ETO interferes with RUNX1/ETO oncogenic activity in t(8;21)-positive leukemic cells, *Cancer Res.* **2007**, 67(5): 2280-2289.
 - [403] G. Nucifora, J.D. Rowley, Aml1 and the 8;21 and 3;21 Translocations in Acute and Chronic Myeloid Leukemia, *Blood* **1995**, 86(1): 1-14.
 - [404] A. Athanasiadou, E. Stalika, V. Sidi, M. Papaioannou, M. Gaitatzi, A. Anagnostopoulos, RUNX1-MTG16 fusion gene in *de novo* acute myeloblastic leukemia with t(16;21)(q24;q22), *Leuk. Lymphoma* **2011**, 52(1): 145-147.
 - [405] E. De Braekeleer, N. Douet-Guilbert, M.J. Le Bris, F. Morel, C. Ferec, M. De Braekeleer, RUNX1-MTG16 fusion gene in acute myeloblastic leukemia with t(16;21)(q24;q22): case report and review of the literature, *Cancer Genet. Cytogenet.* **2008**, 185(1): 47-50.
 - [406] A. Zatkova, C. Fonatsch, W.R. Sperr, P. Valent, A patient with *de novo* AML M1 and t(16;21) with karyotype evolution, *Leuk. Res.* **2007**, 31(9): 1319-1321.
 - [407] R. Berger, M. LeConiat, S.P. Romana, P. Jonveaux, Secondary acute myeloblastic leukemia with t(16;21)(q24;q22) involving the AML1 gene, *Hematol. Cell Ther.* **1996**, 38(2): 183-186.
 - [408] T. Gamou, E. Kitamura, F. Hosoda, K. Shimizu, K. Shinohara, Y. Hayashi, T. Nagase, Y. Yokoyama, M. Ohki, The partner gene of *AML1* in t(16;21) myeloid malignancies is a novel member of the *MTG8(ETO)* family, *Blood* **1998**, 91(11): 4028-4037.
 - [409] H. Reikvam, K.J. Hatfield, A.O. Kittang, R. Hovland, O. Bruserud, Acute Myeloid Leukemia with the t(8;21) Translocation: Clinical Consequences and Biological Implications, *J. Biomed. Biotechnol.* **2011**, (2011): 104631.
 - [410] L.F. Peterson, D.E. Zhang, The 8;21 translocation in leukemogenesis, *Oncogene* **2004**, 23(24): 4255-4262.
 - [411] T.A. Gruber, J.R. Downing (2012) Molecular genetics of acute myeloid leukemia. *Childhood Leukemias*, ed. C.-H. Pui (Cambridge University Press, New York, USA), 3 Ed.
 - [412] J.N. Nichol, S. Assouline, W.H. Miller (2013) The Etiology of Acute Leukemia. *Neoplastic Diseases of the Blood*, eds, Wiernik P.H., Goldman J.M., Dutcher J.P., Kyle R.A. (Springer New York), pp 177-198.
 - [413] R.K. Vangala, M.S. Heiss-Neumann, J.S. Rangatia, S.M. Singh, C. Schoch, D.G. Tenen, W. Hiddemann, G. Behre, The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia, *Blood* **2003**, 101(1): 270-277.
 - [414] T. Ozaki, A. Nakagawara, H. Nagase, RUNX Family Participates in the Regulation of p53-Dependent DNA Damage Response, *Int. J. Genomics* **2013**, 2013: 271347.
 - [415] T. Fukuyama, E. Sueoka, Y. Sugio, T. Otsuka, Y. Niho, K. Akagi, T. Kozu, MTG8 proto-oncoprotein interacts with the regulatory subunit of type II cyclic AMP-dependent protein kinase in lymphocytes, *Oncogene* **2001**, 20(43): 6225-6232.
 - [416] T. Corpora, L. Roudaia, Z.M. Oo, W. Chen, E. Manuylova, X. Cai, M.J. Chen, T. Cierpicki, N.A. Speck, J.H. Bushweller, Structure of the AML1-ETO NHR3-PKA(RII α) complex and its contribution to AML1-ETO activity, *J. Mol. Biol.* **2010**, 402(3): 560-577.

- [417] K. Lam, D.E. Zhang, RUNX1 and RUNX1-ETO: roles in hematopoiesis and leukemogenesis, *Front. Biosci., Landmark Ed.* **2012**, 17: 1120-1139.
- [418] D. Hildebrand, J. Tiefenbach, T. Heinzel, M. Grez, A.B. Maurer, Multiple regions of ETO cooperate in transcriptional repression, *J. Biol. Chem.* **2001**, 276(13): 9889-9895.
- [419] M.E. Engel, H.N. Nguyen, J. Mariotti, A. Hunt, S.W. Hiebert, Myeloid Translocation Gene 16 (*MTG16*) Interacts with Notch Transcription Complex Components To Integrate Notch Signaling in Hematopoietic Cell Fate Specification, *Mol. Cell. Biol.* **2010**, 30(7): 1852-1863.
- [420] J.S. Zhang, M. Kalkum, S. Yamamura, B.T. Chait, R.G. Roeder, E protein silencing by the leukemogenic AML1-ETO fusion protein, *Science* **2004**, 305(5688): 1286-1289.
- [421] C.H. Gow, C. Guo, D. Wang, Q. Hu, J. Zhang, Differential involvement of E2A-corepressor interactions in distinct leukemogenic pathways, *Nucleic Acids Res.* **2013**.
- [422] E. Puccetti, D. Obradovic, T. Beissert, A. Bianchini, B. Washburn, F. Chiaradonna, S. Boehrer, D. Hoelzer, O.G. Ottmann, P.G. Pelicci, *et al.*, AML-associated Translocation Products Block Vitamin D₃-induced Differentiation by Sequestering the Vitamin D₃ Receptor, *Cancer Res.* **2002**, 62(23): 7050-7058.
- [423] E.Y. Ahn, M. Yan, O.A. Malakhova, M.C. Lo, A. Boyapati, H.B. Ommen, R. Hines, P. Hokland, D.E. Zhang, Disruption of the NHR4 domain structure in AML1-ETO abrogates SON binding and promotes leukemogenesis, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105(44): 17103-17108.
- [424] F. Fazi, G. Zardo, V. Gelmetti, L. Travaglini, A. Ciolfi, L. Di Croce, A. Rosa, I. Bozzoni, F. Grignani, F. Lo-Coco, *et al.*, Heterochromatic gene repression of the retinoic acid pathway in acute myeloid leukemia, *Blood* **2007**, 109(10): 4432-4440.
- [425] T. Schoofs, W.E. Berdel, C. Muller-Tidow, Origins of aberrant DNA methylation in acute myeloid leukemia, *Leukemia* **2014**, 28: 1-14.
- [426] O.H. Krämer, R.H. Stauber, G. Bug, J. Hartkamp, S.K. Knauer, SIAH proteins: critical roles in leukemogenesis, *Leukemia* **2013**, 27(4): 792-802.
- [427] J. Herglotz, O.N. Kuvardina, S. Kolodziej, A. Kumar, H. Hussong, M. Grez, J. Lausen, Histone arginine methylation keeps RUNX1 target genes in an intermediate state, *Oncogene* **2013**, 32(20): 2565-2575.
- [428] W.J. Shia, A.J. Okumura, M. Yan, A. Sarkeshik, M.C. Lo, S. Matsuura, Y. Komeno, X.Y. Zhao, S.D. Nimer, J.R. Yates, *et al.*, PRMT1 interacts with AML1-ETO to promote its transcriptional activation and progenitor cell proliferative potential, *Blood* **2012**, 119(21): 4953-4962.
- [429] B.A. Hug, M.A. Lazar, ETO interacting proteins, *Oncogene* **2004**, 23(24): 4270-4274.
- [430] V. Gelmetti, J.S. Zhang, M. Fanelli, S. Minucci, P.G. Pelicci, M.A. Lazar, Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO, *Mol. Cell. Biol.* **1998**, 18(12): 7185-7191.
- [431] B. Lutterbach, J.J. Westendorf, B. Linggi, A. Patten, M. Moniwa, J.R. Davie, K.D. Huynh, V.J. Bardwell, R.M. Lavinsky, M.G. Rosenfeld, *et al.*, ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors, *Mol. Cell. Biol.* **1998**, 18(12): 7176-7184.
- [432] J. Lausen, S.G. Cho, S.H. Liu, M.H. Werner, The nuclear receptor co-repressor (N-CoR) utilizes repression domains I and III for interaction and co-repression with ETO, *J. Biol. Chem.* **2004**, 279(47): 49281-49288.
- [433] R.C. DeKever, M. Yan, E.Y. Ahn, W.J. Shia, N.A. Speck, D.E. Zhang, Attenuation of AML1-ETO cellular dysregulation correlates with increased leukemogenic potential, *Blood* **2013**, 121(18): 3714-3717.

-
- [434] A. Melnick, G.W. Carlile, M.J. McConnell, A. Polinger, S.W. Hiebert, J.D. Licht, AML-1/ETO fusion protein is a dominant negative inhibitor of transcriptional repression by the promyelocytic leukemia zinc finger protein, *Blood* **2000**, 96(12): 3939-3947.
 - [435] L. McGhee, J. Bryan, L. Elliott, H.L. Grimes, A. Kazanjian, J.N. Davis, S. Meyers, Gfi-1 Attaches to the Nuclear Matrix, Associates With ETO (MTG8) and Histone Deacetylase Proteins, and Represses Transcription Using a TSA-Sensitive Mechanism, *J. Cell. Biochem.* **2003**, 89(5): 1005-1018.
 - [436] D. Salat, R. Liefke, J. Wiedenmann, T. Borggrefe, F. Oswald, ETO, but Not Leukemogenic Fusion Protein AML1/ETO, Augments RBP-J κ /SHARP-Mediated Repression of Notch Target Genes, *Mol. Cell. Biol.* **2008**, 28(10): 3502-3512.
 - [437] L.S.H. Chuang, K. Ito, Y. Ito, RUNX family: Regulation and diversification of roles through interacting proteins, *Int. J. Cancer* **2013**, 132(6): 1260-1271.
 - [438] X.N. Gao, J. Lin, Q.Y. Ning, L. Gao, Y.S. Yao, J.H. Zhou, Y.H. Li, L.L. Wang, L. Yu, A Histone Acetyltransferase p300 Inhibitor C646 Induces Cell Cycle Arrest and Apoptosis Selectively in AML1-ETO-Positive AML Cells, *PLoS One* **2013**, 8(2): e55481.
 - [439] J.X. Wang, T. Hoshino, R.L. Redner, S. Kajigaya, J.M. Liu, ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, 95(18): 10860-10865.
 - [440] L. Wang, A. Gural, X.J. Sun, X.Y. Zhao, F. Perna, G. Huang, M.A. Hatlen, L. Vu, F. Liu, H.M. Xu, *et al.*, The Leukemogenicity of AML1-ETO Is Dependent on Site-Specific Lysine Acetylation, *Science* **2011**, 333(6043): 765-769.
 - [441] S.K. Zaidi, A.J. Van Wijnen, J.B. Lian, J.L. Stein, G.S. Stein, Targeting deregulated epigenetic control in cancer, *J. Cell. Physiol.* **2013**, 228(11): 2103-2108.
 - [442] C. Wichmann, M. Grez, J. Lausen, Molecular Targeting of Aberrant Transcription Factors in Leukemia: Strategies for RUNX1/ETO, *Curr. Drug Targets* **2010**, 11(9): 1181-1191.
 - [443] J.H. Bushweller, J. Grembecka, A. Illendula, L. Dixon, INHIBITORS OF INV(16) LEUKEMIA (Pub. No.: US 2012/0059003 A1), **2012**: 1-146.
 - [444] A. Olsson, I. Olsson, R.S. Dhanda, Transcriptional repression by leukaemia-associated ETO family members can be independent of oligomerization and coexpressed hSIN3B and N-CoR, *Biochim. Biophys. Acta, Gene Regul. Mech.* **2008**, 1779(10): 590-598.
 - [445] I. Kitabayashi, K. Ida, F. Morohoshi, A. Yokoyama, N. Mitsuhashi, K. Shimizu, N. Nomura, Y. Hayashi, M. Ohki, The AML1-MTG8 leukemic fusion protein forms a complex with a novel member of the MTG8(ETO/CDR) family, MTGR1, *Mol. Cell. Biol.* **1998**, 18(2): 846-858.
 - [446] A.J. Okumura, L.F. Peterson, F. Okumura, A. Boyapati, D.E. Zhang, t(8;21)(q22;q22) fusion proteins preferentially bind to duplicated AML1/RUNX1 DNA-binding sequences to differentially regulate gene expression, *Blood* **2008**, 112(4): 1392-1401.
 - [447] M.C. Lo, L.F. Peterson, M. Yan, X.L. Cong, F.L. Jin, W.J. Shia, S. Matsuura, E.Y. Ahn, Y. Komeno, M. Ly, *et al.*, Combined gene expression and DNA occupancy profiling identifies potential therapeutic targets of t(8;21) AML, *Blood* **2012**, 120(7): 1473-1484.
 - [448] R.C. DeKelver, B. Lewin, K. Lam, Y. Komeno, M. Yan, C. Rundle, M.-C. Lo, D.-E. Zhang, Cooperation between RUNX1-ETO9a and Novel Transcriptional Partner KLF6 in Upregulation of *Alox5* in Acute Myeloid Leukemia, *PLoS Genet.* **2013**, 9(10): e1003765.

-
- [449] A.M. Abdul-Nabi, E.R. Yassin, N. Varghese, H. Deshmukh, N.R. Yaseen, *In Vitro Transformation of Primary Human CD34+ Cells by AML Fusion Oncogenes: Early Gene Expression Profiling Reveals Possible Drug Target in AML*, *PLoS One* **2010**, 5(8): e12464.
- [450] T. Heinzl, O.H. Krämer, Pharmacodynamic markers for histone deacetylase inhibitor development, *Drug Discovery Today: Dis. Mech.* **2007**, 4(4): 277-283.
- [451] S. Minucci, P.G. Pelicci, Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer, *Nat. Rev. Cancer* **2006**, 6(1): 38-51.
- [452] M. Yan, E.Y. Ahn, S.W. Hiebert, D.E. Zhang, RUNX1/AML1 DNA-binding domain and ETO/MTG8 NHR2-dimerization domain are critical to AML1-ETO9a leukemogenesis, *Blood* **2009**, 113(4): 883-886.
- [453] D. Mannari, D. Gascoyne, J. Dunne, T. Chaplin, B. Young, A novel exon in AML1-ETO negatively influences the clonogenic potential of the t(8;21) in acute myeloid leukemia, *Leukemia* **2010**, 24(4): 891-894.
- [454] K.L. Rhoades, C.J. Hetherington, N. Harakawa, D.A. Yergeau, L.M. Zhou, L.Q. Liu, M.T. Little, D.G. Tenen, D.E. Zhang, Analysis of the role of AML1-ETO in leukemogenesis, using an Inducible transgenic mouse model, *Blood* **2000**, 96(6): 2108-2115.
- [455] O. Heidenreich, J. Krauter, H. Riehle, P. Hadwiger, M. John, G. Heil, H.P. Vornlocher, A. Nordheim, AML1/MTG8 oncogene suppression by small interfering RNAs supports myeloid differentiation of t(8;21)-positive leukemic cells, *Blood* **2003**, 101(8): 3157-3163.
- [456] Y.Z. Yuan, L.M. Zhou, T. Miyamoto, H. Iwasaki, N. Harakawa, C.J. Hetherington, S.A. Burel, E. Lagasse, I.L. Weissman, K. Akashi, *et al.*, AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98(18): 10398-10403.
- [457] J.R. Downing, The core-binding factor leukemias: lessons learned from murine models, *Curr. Opin. Genet. Dev.* **2003**, 13(1): 48-54.
- [458] L.F. Peterson, A. Boyapati, E.Y. Ahn, J.R. Biggs, A.J. Okumura, M.C. Lo, M. Yan, D.E. Zhang, Acute myeloid leukemia with the 8q22;21q22 translocation: secondary mutational events and alternative t(8;21) transcripts, *Blood* **2007**, 110(3): 799-805.
- [459] A.G. Knudson, Hereditary Cancer, Oncogenes, and Antioncogenes, *Cancer Res.* **1985**, 45(4): 1437-1443.
- [460] G. Marcucci, K. Mrozek, A.S. Ruppert, K. Maharry, J.E. Kolitz, J.O. Moore, R.J. Mayer, M.J. Pettenati, B.L. Powell, C.G. Edwards, *et al.*, Prognostic Factors and Outcome of Core Binding Factor Acute Myeloid Leukemia Patients With t(8;21) Differ From Those of Patients With inv(16): A Cancer and Leukemia Group B Study, *J. Clin. Oncol.* **2005**, 23(24): 5705-5717.
- [461] H.H. Zhu, X.H. Zhang, Y.Z. Qin, D.H. Liu, H. Jiang, H. Chen, Q. Jiang, L.P. Xu, J. Lu, W. Han, *et al.*, MRD-directed risk stratification treatment may improve outcomes of t(8;21) AML in the first complete remission: results from the AML05 multicenter trial, *Blood* **2013**, 121(20): 4056-4062.
- [462] A. Shimada, T. Taki, K. Tabuchi, A. Tawa, K. Horibe, M. Tsuchida, R. Hanada, I. Tsukimoto, Y. Hayashi, *KIT* mutations, and not *FLT3* internal tandem duplication, are strongly associated with a poor prognosis in pediatric acute myeloid leukemia with t(8;21): a study of the Japanese Childhood AML Cooperative Study Group, *Blood* **2006**, 107(5): 1806-1809.
- [463] S.A. Gustafson, P. Lin, S.S. Chen, L. Chen, L.V. Abruzzo, R. Luthra, L.J. Medeiros, S.A. Wang, Therapy-Related Acute Myeloid Leukemia With t(8;21) (q22;q22) Shares

- Many Features With De Novo Acute Myeloid Leukemia With t(8;21)(q22;q22) but Does Not Have a Favorable Outcome, *Am. J. Clin. Pathol.* **2009**, 131(5): 647-655.
- [464] L.A. Godley, M.M. LeBeau (2007) Therapy-Related AML. *Acute Myelogenous Leukemia*, ed. Karp J.E. (Humana Press Inc., Totowa, NJ, USA), pp 71-95.
- [465] A. Hunt, M. Fischer, M.E. Engel, S.W. Hiebert, Mtg16/Eto2 Contributes to Murine T-Cell Development, *Mol. Cell. Biol.* **2011**, 31(13): 2544-2551.
- [466] M.A. Fischer, I. Moreno-Miralles, A. Hunt, B.J. Chyla, S.W. Hiebert, Myeloid translocation gene 16 is required for maintenance of haematopoietic stem cell quiescence, *EMBO J.* **2012**, 31(6): 1494-1505.
- [467] M. Kochetkova, O.L. McKenzie, A.J. Bais, J.M. Martin, G.A. Secker, R. Seshadri, J.A. Powell, S.J. Hinze, A.E. Gardner, H.E. Spendlove, *et al.*, *CBFA2T3 (MTG16)* is a putative breast tumor suppressor gene from the breast cancer loss of heterozygosity region at 16q24.3, *Cancer Res.* **2002**, 62(16): 4599-4604.
- [468] R. Kumar, J. Manning, H.E. Spendlove, G. Kremmidiotis, R. McKirdy, J. Lee, D.N. Millband, K.M. Cheney, M.R. Stampfer, P.P. Dwivedi, *et al.*, ZNF652, a novel zinc finger protein, interacts with the putative breast tumor suppressor CBFA2T3 to repress transcription, *Mol. Cancer Res.* **2006**, 4(9): 655-665.
- [469] Z. Saif, Functional analysis of CBFA2T3: a breast cancer tumour suppressor from chromosome band 16q24.3, *Dissertation (The University of Adelaide)* **2009**: 1-270.
- [470] P. Kumar, V.V. Sharoyko, P. Spegel, U. Gullberg, H. Mulder, I. Olsson, R. Ajore, The Transcriptional Co-Repressor Myeloid Translocation Gene 16 Inhibits Glycolysis and Stimulates Mitochondrial Respiration, *PLoS One* **2013**, 8(7): e68502.
- [471] F. Micci, J. Thorsen, I. Panagopoulos, K.B. Nyquist, B. Zeller, A. Tierens, S. Heim, High-throughput sequencing identifies an *NFIA/CBFA2T3* fusion gene in acute erythroid leukemia with t(1;16)(p31;q24), *Leukemia* **2013**, 27(4): 980-982.
- [472] R. Masetti, M. Pigazzi, M. Togni, A. Astolfi, V. Indio, E. Manara, R. Casadio, A. Pession, G. Basso, F. Locatelli, *CBFA2T3-GLIS2* fusion transcript is a novel common feature in pediatric, cytogenetically normal AML, not restricted to FAB M7 subtype, *Blood* **2013**, 121(17): 3469-3472.
- [473] T.A. Gruber, A.L. Gedman, J.H. Zhang, C.S. Koss, S. Marada, H.Q. Ta, S.C. Chen, X.P. Su, S.K. Ogden, J.J. Dang, *et al.*, An Inv(16)(p13.3q24.3)-Encoded *CBFA2T3-GLIS2* Fusion Protein Defines an Aggressive Subtype of Pediatric Acute Megakaryoblastic Leukemia, *Cancer Cell* **2012**, 22(5): 683-697.
- [474] C. Thiollier, C.K. Lopez, B. Gerby, C. Ignacimouttou, S. Poglio, Y. Duffourd, J. Guegan, P. Rivera-Munoz, O. Bluteau, V. Mabialah, *et al.*, Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models, *J. Exp. Med.* **2012**, 209(11): 2017-2031.
- [475] L. Solari, T. Bauer, F. Dicker, C. Haferlach, M. Griesshammer, S. Schnittger, H. Becker, M. Lubbert, A novel recurrent *AML1-ETO* fusion: tight *in vivo* association with *BCR-ABL1*, *Leukemia* **2013**, 27(6): 1397-1400.
- [476] I. Salaverria, T. Akasaka, S. Gesk, M. Szczepanowski, B. Burkhardt, L. Harder, C. Damm-Welk, I. Oschlies, W. Klapper, M.J.S. Dyer, *et al.*, The *CBFA2T3/ACSF3* Locus Is Recurrently Involved in IGH Chromosomal Translocation t(14;16)(q32;q24) in Pediatric B-Cell Lymphoma with Germinal Center Phenotype, *Genes, Chromosomes Cancer* **2012**, 51(4): 338-343.
- [477] C. Kwok, B.B. Zeisig, J.H. Qiu, S.O. Dong, C.W.E. So, Transforming activity of AML1-ETO is independent of CBF β and ETO interaction but requires formation of homo-oligomeric complexes, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106(8): 2853-2858.

-
- [478] T. Zhen, C.F. Wu, P. Liu, H.Y. Wu, G.B. Zhou, Y. Lu, J.X. Liu, Y. Liang, K.K. Li, Y.Y. Wang, *et al.*, Targeting of AML1-ETO in t(8;21) Leukemia by Oridonin Generates a Tumor Suppressor-Like Protein, *Sci. Transl. Med.* **2012**, 4(127): ra38.
 - [479] G.B. Zhou, H. Kang, L. Wang, L. Gao, P. Liu, J. Xie, F.X. Zhang, X.Q. Weng, Z.X. Shen, J. Chen, *et al.*, Oridonin, a diterpenoid extracted from medicinal herbs, targets AML1-ETO fusion protein and shows potent antitumor activity with low adverse effects on t(8;21) leukemia in vitro and in vivo, *Blood* **2007**, 109(8): 3441-3450.
 - [480] G.S. Zhou, Z. Hu, H.T. Fang, F.X. Zhang, X.F. Pan, X.Q. Chen, A.M. Hu, L. Xu, G.B. Zhou, Biologic activity of triptolide in t(8;21) acute myeloid leukemia cells, *Leukemia Res.* **2011**, 35(2): 214-218.
 - [481] H.T. Fang, B. Zhang, X.F. Pan, L. Gao, T. Zhen, H.X. Zhao, L. Ma, J. Xie, Z. Liu, X.J. Yu, *et al.*, Bortezomib interferes with C-KIT processing and transforms the t(8;21)-generated fusion proteins into tumor-suppressing fragments in leukemia cells, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109(7): 2521-2526.
 - [482] W. Jin, K. Wu, Y.Z. Li, W.T. Yang, B. Zou, F. Zhang, J. Zhang, K.K. Wang, AML1-ETO targets and suppresses cathepsin G, a serine protease, which is able to degrade AML1-ETO in t(8;21) acute myeloid leukemia, *Oncogene* **2013**, 32(15): 1978-1987.
 - [483] M. Osato, Point mutations in the *RUNX1/AML1* gene: another actor in RUNX leukemia, *Oncogene* **2004**, 23(24): 4284-4296.
 - [484] The Cancer Genome Atlas Network, Comprehensive molecular portraits of human breast tumours, *Nature* **2012**, 490(7418): 61-70.
 - [485] O. Ben-Ami, D. Friedman, D. Leshkowitz, D. Goldenberg, K. Orlovsky, N. Pencovich, J. Lotem, A. Tanay, Y. Groner, Addiction of t(8;21) and inv(16) Acute Myeloid Leukemia to Native RUNX1, *Cell Rep.* **2013**, 4(6): 1131-1143.
 - [486] S.A. Shurtleff, A. Buijs, F.G. Behm, J.E. Rubnitz, S.C. Raimondi, M.L. Hancock, G.C.F. Chan, C.H. Pui, G. Grosveld, J.R. Downing, TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis, *Leukemia* **1995**, 9(12): 1985-1989.
 - [487] E. De Braekeleer, N. Douet-Guilbert, F. Morel, M.J. Le Bris, C. Ferec, M. De Braekeleer, *RUNX1* translocations and fusion genes in malignant hemopathies, *Future Oncol.* **2011**, 7(1): 77-91.
 - [488] K. Paulsson, A.N. Bekassy, T. Olofsson, F. Mitelman, B. Johansson, I. Panagopoulos, A novel and cytogenetically cryptic t(7;21)(p22;q22) in acute myeloid leukemia results in fusion of *RUNX1* with the ubiquitin-specific protease gene *USP42*, *Leukemia* **2006**, 20(2): 224-229.
 - [489] A. Zelent, M. Greaves, T. Enver, Role of the *TEL-AML1* fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia, *Oncogene* **2004**, 23(24): 4275-4283.
 - [490] A. Giguere, J. Hebert, Identification of a Novel Fusion Gene Involving *RUNX1* and the Antisense Strand of *SV2B* in a *BCR-ABL1*-Positive Acute Leukemia, *Genes, Chromosomes Cancer* **2013**, 52: 1114-1122.
 - [491] I. Taniuchi, M. Osato, Y. Ito, Runx1: no longer just for leukemia, *EMBO J.* **2012**, 31(21): 4098-4099.
 - [492] J.K. Mangan, N.A. Speck, *RUNX1* Mutations in Clonal Myeloid Disorders: From Conventional Cytogenetics to Next Generation Sequencing, A Story 40 Years in the Making, *Crit. Rev. Oncog.* **2011**, 16(1-2): 77-91.
 - [493] T. Sjöblom, S. Jones, L.D. Wood, D.W. Parsons, J. Lin, T.D. Barber, D. Mandelker, R.J. Leary, J. Ptak, N. Silliman, *et al.*, The consensus coding sequences of human breast and colorectal cancers, *Science* **2006**, 314(5797): 268-274.

- [494] L.D. Wood, D.W. Parsons, S. Jones, J. Lin, T. Sjoblom, R.J. Leary, D. Shen, S.M. Boca, T. Barber, J. Ptak, *et al.*, The genomic landscapes of human breast and colorectal cancers, *Science* **2007**, 318(5853): 1108-1113.
- [495] Z.Y. Kan, B.S. Jaiswal, J. Stinson, V. Janakiraman, D. Bhatt, H.M. Stern, P. Yue, P.M. Haverty, R. Bourgon, J.B. Zheng, *et al.*, Diverse somatic mutation patterns and pathway alterations in human cancers, *Nature* **2010**, 466(7308): 869-875.
- [496] G. Archontis, T. Simonson, M. Karplus, Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase, *J. Mol. Biol.* **2001**, 306(2): 307-327.
- [497] Z.S. Hendsch, B. Tidor, Electrostatic interactions in the GCN4 leucine zipper: Substantial contributions arise from intramolecular interactions enhanced on binding, *Protein Sci.* **1999**, 8(7): 1381-1392.
- [498] M. Yan, S.A. Burel, L.F. Peterson, E. Kanbe, H. Iwasaki, A. Boyapati, R. Hines, K. Akashi, D.E. Zhang, Deletion of an AML1-ETO C-terminal NcoR/SMRT-interacting region strongly induces leukemia development, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101(49): 17186-17191.
- [499] A.V. Rulina, P.V. Spirin, V.S. Prassolov, Activated Leukemic Oncogenes *AML1-ETO* and *c-kit*: Role in Development of Acute Myeloid Leukemia and Current Approaches for Their Inhibition, *Biochemistry (Moscow)* **2010**, 75(13): 1650-1666.
- [500] D.E. Clark, What has virtual screening ever done for drug discovery?, *Expert Opin. Drug Discovery* **2008**, 3(8): 841-851.
- [501] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S.H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, *et al.*, Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, *Acc. Chem. Res.* **2000**, 33(12): 889-897.
- [502] J. Srinivasan, T.E. Cheatham, P. Cieplak, P.A. Kollman, D.A. Case, Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices, *J. Am. Chem. Soc.* **1998**, 120(37): 9401-9409.
- [503] N. Homeyer, H. Gohlke, FEW: A workflow tool for free energy calculations of ligand binding, *J. Comput. Chem.* **2013**, 34(11): 965-973.
- [504] N. Homeyer, H. Gohlke, Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method, *Mol. Inf.* **2012**, 31(2): 114-122.
- [505] B. Honig, K. Sharp, A.S. Yang, Macroscopic Models of Aqueous-Solutions: Biological and Chemical Applications, *J. Phys. Chem.* **1993**, 97(6): 1101-1109.
- [506] Q. Lu, R. Luo, A Poisson-Boltzmann dynamics method with nonperiodic boundary condition, *J. Chem. Phys.* **2003**, 119(21): 11035-11047.
- [507] V. Zoete, M.B. Irving, O. Michielin, MM-GBSA binding free energy decomposition and T cell receptor engineering, *J. Mol. Recognit.* **2010**, 23(2): 142-152.
- [508] H. Kopitz, D.A. Cashman, S. Pfeiffer-Marek, H. Gohlke, Influence of the Solvent Representation on Vibrational Entropy Calculations: Generalized Born Versus Distance-Dependent Dielectric Model, *J. Comput. Chem.* **2012**, 33(9): 1004-1013.
- [509] H. Gohlke, L.A. Kuhn, D.A. Case, Change in Protein Flexibility Upon Complex Formation: Analysis of Ras-Raf Using Molecular Dynamics and a Molecular Framework Approach, *Proteins: Struct., Funct., Bioinf.* **2004**, 56(2): 322-337.
- [510] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, *J. Chem. Inf. Model.* **2011**, 51(1): 69-82.

- [511] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics, *J. Am. Chem. Soc.* **1990**, 112(16): 6127-6129.
- [512] W.L. DeLano, Unraveling hot spots in binding interfaces: progress and challenges, *Curr. Opin. Struct. Biol.* **2002**, 12(1): 14-20.
- [513] I.S. Moreira, P.A. Fernandes, M.J. Ramos (2007) Computational Determination of the Relative Free Energy of Binding - Application to Alanine Scanning Mutagenesis. *Molecular Materials with Specific Interactions – Modeling and Design*, ed. Sokalski W.A. (Springer), Vol 4, pp 305-339.
- [514] I. Massova, P.A. Kollman, Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies, *J. Am. Chem. Soc.* **1999**, 121(36): 8133-8143.
- [515] ROCS version 3.0.0, vROCS version 3.1.1, FILTER version 2.0.2, and OMEGA version 2.2.1. OpenEye Scientific Software. Santa Fe, NM, USA, <http://www.eyesopen.com>.
- [516] J.A. Grant, M.A. Gallardo, B.T. Pickup, A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape, *J. Comput. Chem.* **1996**, 17(14): 1653-1666.
- [517] J.A. Grant, B.T. Pickup, A Gaussian Description of Molecular Shape, *J. Phys. Chem.* **1995**, 99(11): 3503-3510.
- [518] P.C.D. Hawkins, A.G. Skillman, A. Nicholls, Comparison of shape-matching and docking as virtual screening tools, *J. Med. Chem.* **2007**, 50(1): 74-82.
- [519] P.C.D. Hawkins, A. Nicholls, Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures, *J. Chem. Inf. Model.* **2012**, 52(11): 2919-2936.
- [520] P.C.D. Hawkins, A.G. Skillman, G.L. Warren, B.A. Ellingson, M.T. Stahl, Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database, *J. Chem. Inf. Model.* **2010**, 50(4): 572-584.
- [521] T.S. Rush, J.A. Grant, L. Mosyak, A. Nicholls, A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction, *J. Med. Chem.* **2005**, 48(5): 1489-1495.
- [522] G. Schneider, W. Neidhart, T. Giller, G. Schmid, "Scaffold-hopping" by topological pharmacophore search: A contribution to virtual screening, *Angew. Chem., Int. Ed.* **1999**, 38(19): 2894-2896.
- [523] H.-J. Böhm, A. Flohr, M. Stahl, Scaffold hopping, *Drug Discovery Today: Technol.* **2004**, 1(3): 217-224.
- [524] J.E.J. Mills, P.M. Dean, Three-dimensional hydrogen-bond geometry and probability information from a crystal survey, *J. Comput.-Aided Mol. Des.* **1996**, 10(6): 607-622.
- [525] SMARTS - A Language for Describing Molecular Patterns. Daylight Chemical Information Systems. Laguna Niguel, CA, USA, <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>.
- [526] A.J. Grant, B.T. Pickup (1998) Gaussian Shape Methods. *Computer Simulations of Biomolecular Systems*, eds, Van Gunsteren W., Weiner P., Wilkinson A.W. (Kluwer/Escom, Dordrecht, The Netherlands), pp 150-176.
- [527] A. Nicholls, G.B. McGaughey, R.P. Sheridan, A.C. Good, G. Warren, M. Mathieu, S.W. Muchmore, S.P. Brown, J.A. Grant, J.A. Haigh, *et al.*, Molecular Shape and Medicinal Chemistry: A Perspective, *J. Med. Chem.* **2010**, 53(10): 3862-3886.
- [528] I.S. Haque, V.S. Pande, PAPER-Accelerating Parallel Evaluations of ROCS, *J. Comput. Chem.* **2010**, 31(1): 117-132.

- [529] FastROCS version 1.0. OpenEye Scientific Software. Santa Fe, NM, USA, <http://www.eyesopen.com/fastrocs>.
- [530] S.R. Vasudevan, J.B. Moore, Y. Schymura, G.C. Churchill, Shape-Based Reprofile of FDA-Approved Drugs for the H₁ Histamine Receptor, *J. Med. Chem.* **2012**, 55(16): 7054-7060.
- [531] M. Brickelmaier, A. Lugovskoy, R. Kartikeyan, M.M. Reviriego-Mendoza, N. Allaire, K. Simon, R.J. Frisque, L. Gorelik, Identification and Characterization of Mefloquine Efficacy against JC Virus In Vitro, *Antimicrob. Agents Chemother.* **2009**, 53(5): 1840-1849.
- [532] J.O. Ebalunode, W.F. Zheng, Molecular Shape Technologies in Drug Discovery: Methods and Applications, *Curr. Top. Med. Chem.* **2010**, 10(6): 669-679.
- [533] W.J. Geldenhuys, M.O. Funk, C.J. Van der Schyf, R.T. Carroll, A scaffold hopping approach to identify novel monoamine oxidase B inhibitors, *Bioorg. Med. Chem. Lett.* **2012**, 22(3): 1380-1383.
- [534] J. Kirchmair, S. Distinto, P. Markt, D. Schuster, G.M. Spitzer, K.R. Liedl, G. Wolber, How To Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information, *J. Chem. Inf. Model.* **2009**, 49(3): 678-692.
- [535] G.M. Sastry, S.L. Dixon, W. Sherman, Rapid Shape-Based Ligand Alignment and Virtual Screening Method Based on Atom/Feature-Pair Similarities and Volume Overlap Scoring, *J. Chem. Inf. Model.* **2011**, 51(10): 2455-2466.
- [536] G.B. McGaughey, R.P. Sheridan, C.I. Bayly, J.C. Culberson, C. Kreatsoulas, S. Lindsley, V. Maiorov, J.F. Truchon, W.D. Cornell, Comparison of topological, shape, and docking methods in virtual screening, *J. Chem. Inf. Model.* **2007**, 47(4): 1504-1519.
- [537] K. Grabowski, K.H. Baringhaus, G. Schneider, Scaffold diversity of natural products: inspiration for combinatorial library design, *Nat. Prod. Rep.* **2008**, 25(5): 892-904.
- [538] G. Schneider, P. Schneider, S. Renner, Scaffold-Hopping: How Far Can You Jump?, *QSAR Comb. Sci.* **2006**, 25(12): 1162-1171.
- [539] D. Weininger, SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules, *J. Chem. Inf. Comput. Sci.* **1988**, 28(1): 31-36.
- [540] J.M. LaLonde, M.A. Elban, J.R. Courter, A. Sugawara, T. Soeta, N. Madani, A.M. Princiotto, Y. Do Kwon, P.D. Kwong, A. Schon, *et al.*, Design, synthesis and biological evaluation of small molecule inhibitors of CD4-gp120 binding based on virtual screening, *Bioorg. Med. Chem.* **2011**, 19(1): 91-101.
- [541] T.S. Kaoud, C. Yan, S. Mitra, C.C. Tseng, J. Jose, J.M. Taliaferro, M. Tuohetahuntala, A. Devkota, R. Sammons, J. Park, *et al.*, From in Silico Discovery to intra-Cellular Activity: Targeting JNK-Protein Interactions with Small Molecules, *ACS Med. Chem. Lett.* **2012**, 3(9): 721-725.
- [542] R.P. Fitzgerald, Small Molecule Inhibitors of the p53-MDM2 Protein-Protein Interaction, *PhD Thesis (University of Nottingham)* **2009**: 1-251.
- [543] L. Ganesan, Small Molecule Inhibition of CD40-CD154 Co-stimulatory Interaction, *PhD Thesis (University of Miami)* **2012**: 1-122.
- [544] L. Ganesan, D. Vidovic, S.C. Schurer, P. Buchwald, Exploratory computational assessment of possible binding modes for small molecule inhibitors of the CD40-CD154 co-stimulatory interaction, *Pharmazie* **2012**, 67(5): 374-379.
- [545] M. Floris, S. Moro, Mimicking Peptides... In Silico, *Mol. Inf.* **2012**, 31(1): 12-20.
- [546] R.P. Sheridan, S.B. Singh, E.M. Fluder, S.K. Kearsley, Protocols for bridging the peptide to nonpeptide gap in topological similarity searches, *J. Chem. Inf. Comput. Sci.* **2001**, 41(5): 1395-1406.

- [547] H.M. Sun, G. Tawa, A. Wallqvist, Classification of scaffold-hopping approaches, *Drug Discovery Today* **2012**, 17(7-8): 310-324.
- [548] M.D. Kiran, N.V. Adikesavan, O. Cirioni, A. Giacometti, C. Silvestri, G. Scalise, R. Ghiselli, V. Saba, F. Orlando, M. Shoham, *et al.*, Discovery of a Quorum-Sensing Inhibitor of Drug-Resistant Staphylococcal Infections by Structure-Based Virtual Screening, *Mol. Pharmacol.* **2008**, 73(5): 1578-1586.
- [549] S.M. Massa, Y.M. Xie, F.M. Longo, Alzheimer's Therapeutics: Neurotrophin Domain Small Molecule Mimetics, *J. Mol. Neurosci.* **2003**, 20(3): 323-326.
- [550] L. Parthasarathi, F. Casey, A. Stein, P. Aloy, D.C. Shields, Approved Drug Mimics of Short Peptide Ligands from Protein Interaction Motifs, *J. Chem. Inf. Model.* **2008**, 48(10): 1943-1948.
- [551] M. Floris, J. Masciocchi, M. Fanton, S. Moro, Swimming into peptidomimetic chemical space using pepMMsMIMIC, *Nucleic Acids Res.* **2011**, 39: W261-W269.
- [552] M.D. Miller, R.P. Sheridan, S.K. Kearsley, SQ: A program for rapidly producing pharmacophorically relevant molecular superpositions, *J. Med. Chem.* **1999**, 42(9): 1505-1514.
- [553] L.H. Yang, L.Q. Guo, A. Pasternak, R. Mosley, S. Rohrer, E. Birzin, F. Foor, K. Cheng, J. Schaeffer, A.A. Patchett, Spiro[1*H*-indene-1,4'-piperidine] Derivatives As Potent and Selective Non-Peptide Human Somatostatin Receptor Subtype 2 (sst₂) Agonists, *J. Med. Chem.* **1998**, 41(13): 2175-2179.
- [554] G.D. Hartman, M.S. Egbertson, W. Halczenko, W.L. Laswell, M.E. Duggan, R.L. Smith, A.M. Naylor, P.D. Manno, R.J. Lynch, G. Zhang, *et al.*, Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors, *J. Med. Chem.* **1992**, 35(24): 4640-4642.
- [555] L. De Luca, M.L. Barreca, S. Ferro, F. Christ, N. Iraci, R. Gitto, A.M. Monforte, Z. Debyser, A. Chimirri, Pharmacophore-Based Discovery of Small-Molecule Inhibitors of Protein-Protein Interactions between HIV-1 Integrase and Cellular Cofactor LEDGF/p75, *ChemMedChem* **2009**, 4(8): 1311-1316.
- [556] F. Caporuscio, A. Tafi, E. Gonzalez, F. Manetti, J.A. Este, M. Botta, A dynamic target-based pharmacophoric model mapping the CD4 binding site on HIV-1 gp120 to identify new inhibitors of gp120-CD4 protein-protein interactions, *Bioorg. Med. Chem. Lett.* **2009**, 19(21): 6087-6091.
- [557] A. Voet, K.Y. Zhang, Pharmacophore modelling as a virtual screening tool for the discovery of small molecule protein-protein interaction inhibitors, *Curr. Pharm. Des.* **2012**, 18(30): 4586-4598.
- [558] A. Voet, E.F. Banwell, K.K. Sahu, J.G. Heddle, K.Y. Zhang, Protein interface pharmacophore mapping tools for small molecule protein: protein interaction inhibitor discovery, *Curr. Top. Med. Chem.* **2013**, 13(9): 989-1001.
- [559] D.R. Koes, C.J. Camacho, ZINCPharmer: pharmacophore search of the ZINC database, *Nucleic Acids Res.* **2012**, 40(W1): W409-414.
- [560] X. Xue, J.-L. Wei, L.-L. Xu, M.-Y. Xi, X.-L. Xu, F. Liu, X.-K. Guo, L. Wang, X.-J. Zhang, M.-Y. Zhang, *et al.*, Effective Screening Strategy Using Ensembled Pharmacophore Models Combined with Cascade Docking: Application to p53-MDM2 Interaction Inhibitors, *J. Chem. Inf. Model.* **2013**, 53(10): 2715-2729.
- [561] P.R. Hall, A. Leitao, C.Y. Ye, K. Kilpatrick, B. Hjelle, T.I. Oprea, R.S. Larson, Small molecule inhibitors of hantavirus infection, *Bioorg. Med. Chem. Lett.* **2010**, 20(23): 7085-7091.
- [562] V. Vacic, C.J. Oldfield, A. Mohan, P. Radivojac, M.S. Cortese, V.N. Uversky, A.K. Dunker, Characterization of molecular recognition features, MoRFs, and their binding partners, *J. Proteome. Res.* **2007**, 6(6): 2351-2366.

-
- [563] C. Barillari, G. Marcou, D. Rognan, Hot-Spots-Guided Receptor-Based Pharmacophores (HS-Pharm): A Knowledge-Based Approach to Identify Ligand-Anchoring Atoms in Protein Cavities and Prioritize Structure-Based Pharmacophores, *J. Chem. Inf. Model.* **2008**, 48(7): 1396-1410.
- [564] P.R. Hall, L. Malone, L.O. Sillerud, C.Y. Ye, B.L. Hjelle, R.S. Larson, Characterization and NMR solution structure of a novel cyclic pentapeptide inhibitor of pathogenic hantaviruses, *Chem. Biol. Drug Des.* **2007**, 69(3): 180-190.
- [565] J.O. Ebalunode, Z. Ouyang, J. Liang, W.F. Zheng, Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques, *J. Chem. Inf. Model.* **2008**, 48(4): 889-901.
- [566] P.M. Fischer, Protein-Protein Interactions in Drug Discovery, *Drug Des. Rev. Online* **2005**, 2(3): 179-207.
- [567] A.I. Archakov, V.M. Govorun, A.V. Dubanov, Y.D. Ivanov, A.V. Veselovsky, P. Lewi, P. Janssen, Protein-protein interactions as a target for drugs in proteomics, *Proteomics* **2003**, 3(4): 380-391.
- [568] P. Pagel, S. Kovac, M. Oesterheld, B. Brauner, I. Dunger-Kaltenbach, G. Frishman, C. Montrone, P. Mark, V. Stumpflen, H.W. Mewes, *et al.*, The MIPS mammalian protein-protein interaction database, *Bioinformatics* **2005**, 21(6): 832-834.
- [569] T. Beuming, L. Skrabanek, M.Y. Niv, P. Mukherjee, H. Weinstein, PDZBase: a protein-protein interaction database for PDZ-domains, *Bioinformatics* **2005**, 21(6): 827-828.
- [570] S. Cory, J.M. Adams, The BCL2 family: Regulators of the cellular life-or-death switch, *Nat. Rev. Cancer* **2002**, 2(9): 647-656.
- [571] A. Villunger, C. Scott, P. Bouillet, A. Strasser, Essential role for the BH3-only protein Bim but redundant roles for Bax, Bcl-2, and Bcl-w in the control of granulocyte survival, *Blood* **2003**, 101(6): 2393-2400.
- [572] D. Kozakov, D.R. Hall, G.Y. Chuang, R. Cencic, R. Brenke, L.E. Grove, D. Beglov, J. Pelletier, A. Whitty, S. Vajda, Structural conservation of druggable hot spots in protein-protein interfaces, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108(33): 13528-13533.
- [573] J.A. Gerrard, C.A. Hutton, M.A. Perugini, Inhibiting protein-protein interactions as an emerging paradigm for drug discovery, *Mini Rev. Med. Chem.* **2007**, 7(2): 151-157.
- [574] P. Chene, Drugs targeting protein-protein interactions, *ChemMedChem* **2006**, 1(4): 400-411.
- [575] W.L. DeLano, M.H. Ultsch, A.M. de Vos, J.A. Wells, Convergent solutions to binding at a protein-protein interface, *Science* **2000**, 287(5456): 1279-1283.
- [576] E.J. Deeds, O. Ashenberg, J. Gerardin, E.I. Shakhnovich, Robust protein-protein interactions in crowded cellular environments, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104(38): 14952-14957.
- [577] I.M.A. Nooren, J.M. Thornton, Diversity of protein-protein interactions, *EMBO J.* **2003**, 22(14): 3486-3492.
- [578] N. Tuncbag, A. Gursoy, E. Guney, R. Nussinov, O. Keskin, Architectures and functional coverage of protein-protein interfaces, *J. Mol. Biol.* **2008**, 381(3): 785-802.
- [579] C.J. Tsai, D. Xu, R. Nussinov, Structural motifs at protein-protein interfaces: Protein cores versus two-state and three-state model complexes, *Protein Sci.* **1997**, 6(9): 1793-1805.
- [580] L. Young, R.L. Jernigan, D.G. Covell, A Role for Surface Hydrophobicity in Protein-Protein Recognition, *Protein Sci.* **1994**, 3(5): 717-729.
- [581] T.M. Raschke, Water structure and interactions with protein surfaces, *Curr. Opin. Struct. Biol.* **2006**, 16(2): 152-159.

-
- [582] F. Rodier, R.P. Bahadur, P. Chakrabarti, J. Janin, Hydration of protein-protein interfaces, *Proteins: Struct., Funct., Bioinf.* **2005**, 60(1): 36-45.
- [583] I.S. Moreira, P.A. Fernandes, M.J. Ramos, Hot spots-A review of the protein-protein interface determinant amino-acid residues, *Proteins: Struct., Funct., Bioinf.* **2007**, 68(4): 803-812.
- [584] K.S. Thorn, A.A. Bogan, ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions, *Bioinformatics* **2001**, 17(3): 284-285.
- [585] B.Y. Ma, T. Elkayam, H. Wolfson, R. Nussinov, Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100(10): 5772-5777.
- [586] D. Rajamani, S. Thiel, S. Vajda, C.J. Camacho, Anchor residues in protein-protein interactions, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101(31): 11287-11292.
- [587] O.N. Yagurtcu, S.B. Erdemli, R. Nussinov, M. Turkay, O. Keskin, Restricted mobility of conserved residues in protein-protein interfaces in molecular simulations, *Biophys. J.* **2008**, 94(9): 3475-3485.
- [588] A. Shulman-Peleg, M. Shatsky, R. Nussinov, H.J. Wolfson, Spatial chemical conservation of hot spot interactions in protein-protein complexes, *BMC Biol.* **2007**, 5.
- [589] R.J. Owellen, C.A. Hartke, R.M. Dickerson, F.O. Hains, Inhibition of Tubulin-Microtubule Polymerization by Drugs of Vinca Alkaloid Class, *Cancer Res.* **1976**, 36(4): 1499-1502.
- [590] G. Liu, Small molecule antagonists of the LFA-1/ICAM-1 interaction as potential therapeutic agents, *Expert Opin. Ther. Pat.* **2001**, 11(9): 1383-1393.
- [591] L.O. Sillerud, R.S. Larson, Design and structure of peptide and peptidomimetic antagonists of protein-protein interaction, *Curr. Protein Pept. Sci.* **2005**, 6(2): 151-169.
- [592] E.J. Topol, T.V. Byzova, E.F. Plow, Platelet GPIIb-IIIa blockers, *Lancet* **1999**, 353(9148): 227-231.
- [593] D. Kuritzkes, S. Kar, P. Kirkpatrick, Fresh from the pipeline: Maraviroc, *Nat. Rev. Drug Discovery* **2008**, 7(1): 15-16.
- [594] S. Perot, O. Sperandio, M.A. Miteva, A.C. Camproux, B.O. Villoutreix, Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery, *Drug Discovery Today* **2010**, 15(15-16): 656-667.
- [595] D.G. Levitt, L.J. Banaszak, POCKET: Computer-Graphics Method for Identifying and Displaying Protein Cavities and Their Surrounding Amino-Acids, *J. Mol. Graphics* **1992**, 10(4): 229-234.
- [596] M. Hendlich, F. Rippmann, G. Barnickel, LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins, *J. Mol. Graphics Modell.* **1997**, 15(6): 359-363.
- [597] B.D. Huang, M. Schroeder, LIGSITE^{CSC}: predicting ligand binding sites using the Connolly surface and degree of conservation, *BMC Struct. Biol.* **2006**, 6: 19.
- [598] R.A. Laskowski, SURFNET: a Program for Visualizing Molecular-Surfaces, Cavities, and Intermolecular Interactions, *J. Mol. Graphics* **1995**, 13(5): 323-330.
- [599] J. Liang, H. Edelsbrunner, C. Woodward, Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design, *Protein Sci.* **1998**, 7(9): 1884-1897.
- [600] M. Weisel, E. Proschak, G. Schneider, PocketPicker: analysis of ligand binding-sites with shape descriptors, *Chem. Cent. J.* **2007**, 1: 7.
- [601] P. Schmidtke, V. Le Guilloux, J. Maupetit, P. Tuffery, fpocket: online tools for protein ensemble pocket detection and tracking, *Nucleic Acids Res.* **2010**, 38: W582-W589.

-
- [602] G.P. Brady, P.F.W. Stouten, Fast prediction and visualization of protein binding pockets with PASS, *J. Comput.-Aided Mol. Des.* **2000**, 14(4): 383-401.
- [603] P.J. Goodford, A Computational-Procedure for Determining Energetically Favorable Binding-Sites on Biologically Important Macromolecules, *J. Med. Chem.* **1985**, 28(7): 849-857.
- [604] J. Ruppert, W. Welch, A.N. Jain, Automatic identification and representation of protein binding sites for molecular docking, *Protein Sci.* **1997**, 6(3): 524-533.
- [605] A.T.R. Laurie, R.M. Jackson, Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites, *Bioinformatics* **2005**, 21(9): 1908-1916.
- [606] M. Morita, S. Nakamura, K. Shimizu, Highly accurate method for ligand-binding site prediction in unbound state (apo) protein structures, *Proteins: Struct., Funct., Bioinf.* **2008**, 73(2): 468-479.
- [607] C. Mattos, D. Ringe, Locating and characterizing binding sites on proteins, *Nat. Biotechnol.* **1996**, 14(5): 595-599.
- [608] M. Clark, F. Guarnieri, I. Shkurko, J. Wiseman, Grand canonical Monte Carlo simulation of ligand-protein binding, *J. Chem. Inf. Model.* **2006**, 46(1): 231-242.
- [609] J.H. An, M. Totrov, R. Abagyan, Pocketome via Comprehensive Identification and Classification of Ligand Binding Envelopes, *Mol. Cell. Proteomics* **2005**, 4(6): 752-761.
- [610] D. Kuhn, N. Weskamp, S. Schmitt, E. Hullermeier, G. Klebe, From the similarity analysis of protein cavities to the functional classification of protein families using Cavbase, *J. Mol. Biol.* **2006**, 359(4): 1023-1044.
- [611] R. Najmanovich, N. Kurbatova, J. Thornton, Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites, *Bioinformatics* **2008**, 24(16): I105-I111.
- [612] T. Pupko, R.E. Bell, I. Mayrose, F. Glaser, N. Ben-Tal, Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues, *Bioinformatics* **2002**, 18(Suppl. 1): S71-S77.
- [613] F. Glaser, T. Pupko, I. Paz, R.E. Bell, D. Bechor-Shental, E. Martz, N. Ben-Tal, ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information, *Bioinformatics* **2003**, 19(1): 163-164.
- [614] M. Brylinski, J. Skolnick, A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105(1): 129-134.
- [615] J.R. Schames, R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H.H. Ni, J.A. McCammon, Discovery of a novel binding trench in HIV integrase, *J. Med. Chem.* **2004**, 47(8): 1879-1881.
- [616] M. Lei, M.I. Zavodszky, L.A. Kuhn, M.F. Thorpe, Sampling protein conformations and pathways, *J. Comput. Chem.* **2004**, 25(9): 1133-1148.
- [617] M.I. Zavodszky, L. Ming, M.F. Thorpe, A.R. Day, L.A. Kuhn, Modeling correlated main-chain motions in proteins for flexible molecular recognition, *Proteins: Struct., Funct., Bioinf.* **2004**, 57(2): 243-261.
- [618] S.A. Wells, S. Menor, B. Hespeneide, M.F. Thorpe, Constrained geometric simulation of diffusive motion in proteins, *Phys. Biol.* **2005**, 2(4): S127-S136.
- [619] G.R. Smith, M.J.E. Sternberg, P.A. Bates, The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking, *J. Mol. Biol.* **2005**, 347(5): 1077-1101.
- [620] M.R. Landon, R.E. Amaro, R. Baron, C.H. Ngan, D. Ozonoff, J.A. McCammon, S. Vajda, Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by

- computational solvent mapping of a reduced and representative receptor ensemble, *Chem. Biol. Drug Des.* **2008**, 71(2): 106-116.
- [621] C.Y. Yang, S.M. Wang, Hydrophobic Binding Hot Spots of Bcl-xL Protein-Protein Interfaces by Cosolvent Molecular Dynamics Simulation, *ACS Med. Chem. Lett.* **2011**, 2(4): 280-284.
- [622] A. Miranker, M. Karplus, Functionality Maps of Binding Sites: a Multiple Copy Simultaneous Search Method, *Proteins: Struct., Funct., Genet.* **1991**, 11(1): 29-34.
- [623] C.M. Stultz, M. Karplus, MCSS functionality maps for a flexible protein, *Proteins: Struct., Funct., Genet.* **1999**, 37(4): 512-529.
- [624] J. Zeng, T. Nheu, A. Zorzet, B. Catimel, E. Nice, H. Maruta, A.W. Burgess, H.R. Treutlein, Design of inhibitors of Ras-Raf interaction using a computational combinatorial algorithm, *Protein Eng.* **2001**, 14(1): 39-45.
- [625] D.R. Davies, D.W. Begley, R.C. Hartley, B.L. Staker, L.J. Stewart (2011) Predicting the Success of Fragment Screening by X-Ray Crystallography. *Fragment-Based Drug Design: Tools, Practical Approaches, and Examples*, Methods in Enzymology, ed. Juo L. (Elsevier Academic Press), Vol 493, pp 91-114.
- [626] G. Buhrman, C. O'Connor, B. Zerbe, B.M. Kearney, R. Napoleon, E.A. Kovrigina, S. Vajda, D. Kozakov, E.L. Kovrigin, C. Mattos, Analysis of Binding Site Hot Spots on the Surface of Ras GTPase, *J. Mol. Biol.* **2011**, 413(4): 773-789.
- [627] S.B. Shuker, P.J. Hajduk, R.P. Meadows, S.W. Fesik, Discovering high-affinity ligands for proteins: SAR by NMR, *Science* **1996**, 274(5292): 1531-1534.
- [628] T. Oltersdorf, S.W. Elmore, A.R. Shoemaker, R.C. Armstrong, D.J. Augeri, B.A. Belli, M. Bruncko, T.L. Deckwerth, J. Dinges, P.J. Hajduk, *et al.*, An inhibitor of Bcl-2 family proteins induces regression of solid tumours, *Nature* **2005**, 435(7042): 677-681.
- [629] P.J. Hajduk, J.C. Mack, E.T. Olejniczak, C. Park, P.J. Dandliker, B.A. Beutel, SOS-NMR: A saturation transfer NMR-based method for determining the structures of protein-ligand complexes, *J. Am. Chem. Soc.* **2004**, 126(8): 2390-2398.
- [630] J. Fernandez-Recio, Prediction of protein binding sites and hot spots, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, 1(5): 680-698.
- [631] D.A. Erlanson, S.K. Hansen, Making drugs on proteins: site-directed ligand discovery for fragment-based lead assembly, *Curr. Opin. Chem. Biol.* **2004**, 8(4): 399-406.
- [632] G. Tóth, K. Mukhyala, J.A. Wells, Computational approach to site-directed ligand discovery, *Proteins: Struct., Funct., Bioinf.* **2007**, 68(2): 551-560.
- [633] T.B. Fischer, K.V. Arunachalam, D. Bailey, V. Mangual, S. Bakhru, R. Russo, D. Huang, M. Paczkowski, V. Lalchandani, C. Ramachandra, *et al.*, The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces, *Bioinformatics* **2003**, 19(11): 1453-1454.
- [634] C. Rohl, Y. Price, T.B. Fischer, M. Paczkowski, M.F. Zette, J. Tsai, Cataloging the relationships between proteins, *Mol. Biotechnol.* **2006**, 34(1): 69-93.
- [635] E. Guney, N. Tuncbag, O. Keskin, A. Gursoy, HotSprint: database of computational hot spots in protein interfaces, *Nucleic Acids Res.* **2008**, 36: D662-D666.
- [636] B.L. deGroot, D.M.F. vanAalten, R.M. Scheek, A. Amadei, G. Vriend, H.J.C. Berendsen, Prediction of protein conformational freedom from distance constraints, *Proteins: Struct., Funct., Genet.* **1997**, 29(2): 240-251.
- [637] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, The FoldX web server: an online force field, *Nucleic Acids Res.* **2005**, 33: W382-W388.
- [638] D.E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server, *Nucleic Acids Res.* **2004**, 32: W526-W531.

- [639] T. Kortemme, D. Baker, A simple physical model for binding energy hot spots in protein-protein complexes, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99(22): 14116-14121.
- [640] T. Kortemme, D.E. Kim, D. Baker, Computational alanine scanning of protein-protein interfaces, *Sci. Signaling* **2004**, 2004(219): pl2.
- [641] N. Tuncbag, O. Keskin, A. Gursoy, HotPoint: hot spot prediction server for protein interfaces, *Nucleic Acids Res.* **2010**, 38: W402-W406.
- [642] O. Keskin, I. Bahar, A.Y. Badretdinov, O.B. Ptitsyn, R.L. Jernigan, Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions, *Protein Sci.* **1998**, 7(12): 2578-2586.
- [643] Q. Liu, S.C.H. Hoi, C.T.T. Su, Z.H. Li, C.K. Kwok, L. Wong, J.Y. Li, Structural analysis of the hot spots in the binding between H1N1 HA and the 2D1 antibody: do mutations of H1N1 from 1918 to 2009 affect much on this binding?, *Bioinformatics* **2011**, 27(18): 2529-2536.
- [644] S. Huo, I. Massova, P.A. Kollman, Computational alanine scanning of the 1:1 human growth hormone-receptor complex, *J. Comput. Chem.* **2002**, 23(1): 15-27.
- [645] C.L. Zhuang, Z.Y. Miao, L.J. Zhu, Y.Q. Zhang, Z.Z. Guo, J.Z. Yao, G.Q. Dong, S.Z. Wang, Y. Liu, H. Chen, *et al.*, Synthesis and biological evaluation of thio-benzodiazepines as novel small molecule inhibitors of the p53-MDM2 protein-protein interaction, *Eur. J. Med. Chem.* **2011**, 46(11): 5654-5661.
- [646] C.Q. Hu, X. Li, W.S. Wang, L. Zhang, L.L. Tao, X.W. Dong, R. Sheng, B. Yang, Y.Z. Hu, Design, synthesis, and biological evaluation of imidazoline derivatives as p53-MDM2 binding inhibitors, *Bioorg. Med. Chem.* **2011**, 19(18): 5454-5461.
- [647] K. Ding, Y.P. Lu, Z. Nikolovska-Coleska, G.P. Wang, S. Qiu, S. Shangary, W. Gao, D.G. Qin, J. Stuckey, K. Krajewski, *et al.*, Structure-based design of spiro-oxindoles as potent, specific small-molecule inhibitors of the MDM2-p53 interaction, *J. Med. Chem.* **2006**, 49(12): 3432-3435.
- [648] P.S. Galatin, D.J. Abraham, QSAR: Hydropathic analysis of inhibitors of the p53-mdm2 interaction, *Proteins: Struct., Funct., Genet.* **2001**, 45(3): 169-175.
- [649] D.K. Agrafiotis, Stochastic algorithms for maximizing molecular diversity, *J. Chem. Inf. Comput. Sci.* **1997**, 37(5): 841-851.
- [650] I.R. Craig, C. Pflieger, H. Gohlke, J.W. Essex, K. Spiegel, Pocket-Space Maps To Identify Novel Binding-Site Conformations in Proteins, *J. Chem. Inf. Model.* **2011**, 51(10): 2666-2679.
- [651] J.M. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, *J. Comput. Chem.* **2000**, 21(12): 1049-1074.
- [652] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, *Proteins: Struct., Funct., Bioinf.* **2006**, 65(3): 712-725.
- [653] J.M. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and Testing of a General Amber Force Field, *J. Comput. Chem.* **2004**, 25(9): 1157-1174.
- [654] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, *et al.*, AMBER 9. University of California, San Francisco, **2006**.
- [655] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, B. Honig, Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects, *J. Comput. Chem.* **2002**, 23(1): 128-137.

- [656] Q. Cui, T. Sulea, J.D. Schrag, C. Munger, M.N. Hung, M. Naim, M. Cygler, E.O. Purisima, Molecular dynamics-solvated interaction energy studies of protein-protein interactions: the MP1-p14 scaffolding complex, *J. Mol. Biol.* **2008**, 379(4): 787-802.
- [657] N.J. Deng, P. Cieplak, Insights into affinity and specificity in the complexes of α -lytic protease and its inhibitor proteins: binding free energy from molecular dynamics simulation, *Phys. Chem. Chem. Phys.* **2009**, 11(25): 4968-4981.
- [658] V. Zoete, O. Michielin, Comparison Between Computational Alanine Scanning and Per-Residue Binding Free Energy Decomposition for Protein-Protein Association Using MM-GBSA: Application to the TCR-p-MHC Complex, *Proteins* **2007**, 67(4): 1026-1047.
- [659] A. Tuncel, I.H. Kavakli, O. Keskin, Insights into subunit interactions in the heterotetrameric structure of potato ADP-glucose pyrophosphorylase, *Biophys. J.* **2008**, 95(8): 3628-3639.
- [660] T. Hou, R. Yu, Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance, *J. Med. Chem.* **2007**, 50(6): 1177-1188.
- [661] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.* **1998**, 19(14): 1639-1662.
- [662] H. Gohlke, M. Hendlich, G. Klebe, Knowledge-based scoring function to predict protein-ligand interactions, *J. Mol. Biol.* **2000**, 295(2): 337-356.
- [663] C.A. Sotriffer, H. Gohlke, G. Klebe, Docking into knowledge-based potential fields: a comparative evaluation of DrugScore, *J. Med. Chem.* **2002**, 45(10): 1967-1970.
- [664] J.A. Erickson, M. Jalaie, D.H. Robertson, R.A. Lewis, M. Vieth, Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy, *J. Med. Chem.* **2004**, 47(1): 45-55.
- [665] P. Ferrara, H. Gohlke, D.J. Price, G. Klebe, C.L. Brooks, 3rd, Assessing scoring functions for protein-ligand interactions, *J. Med. Chem.* **2004**, 47(12): 3032-3047.
- [666] M.L. Verdonk, P.N. Mortenson, R.J. Hall, M.J. Hartshorn, C.W. Murray, Protein-ligand docking against non-native protein conformers, *J. Chem. Inf. Model.* **2008**, 48(11): 2214-2225.
- [667] G. Hu, D. Wang, X. Liu, Q. Zhang, A computational analysis of the binding model of MDM2 with inhibitors, *J. Comput.-Aided Mol. Des.* **2010**, 24(8): 687-697.
- [668] S. Ahmed, R.P. Metpally, S. Sangadala, B.V. Reddy, Virtual screening and selection of drug-like compounds to block noggin interaction with bone morphogenetic proteins, *J. Mol. Graphics Modell.* **2010**, 28(7): 670-682.
- [669] J.J. Irwin, B.K. Shoichet, ZINC - A Free Database of Commercially Available Compounds for Virtual Screening, *J. Chem. Inf. Model.* **2005**, 45(1): 177-182.
- [670] N. Huang, B.K. Shoichet, J.J. Irwin, Benchmarking sets for molecular docking, *J. Med. Chem.* **2006**, 49(23): 6789-6801.
- [671] M. Jacobsson, A. Karlen, Ligand bias of scoring functions in structure-based virtual screening, *J. Chem. Inf. Model.* **2006**, 46(3): 1334-1343.
- [672] J.S. Albert, M.S. Goodman, A.D. Hamilton, Molecular Recognition of Proteins: Sequence-Selective Binding of Aspartate Pairs in Helical Peptides, *J. Am. Chem. Soc.* **1995**, 117(3): 1143-1144.
- [673] J.S. Albert, M.W. Peczu, A.D. Hamilton, Design, Synthesis and Evaluation of Synthetic Receptors for the Recognition of Aspartate Pairs in an α -helical Conformation, *Bioorg. Med. Chem.* **1997**, 5(8): 1455-1467.

- [674] M.J. Adler, A.D. Hamilton, Oligophenylenaminones as Scaffolds for α -Helix Mimicry, *J. Org. Chem.* **2011**, 76(17): 7040-7047.
- [675] Molinspiration. Molinspiration Cheminformatics. Slovenský Grob, Slovak Republic, <http://www.molinspiration.com/cgi-bin/properties>, **2008**.
- [676] O. Kutzki, H.S. Park, J.T. Ernst, B.P. Orner, H. Yin, A.D. Hamilton, Development of a potent Bcl-x_L antagonist based on α -helix mimicry, *J. Am. Chem. Soc.* **2002**, 124(40): 11838-11839.
- [677] H. Yin, G.I. Lee, K.A. Sedey, O. Kutzki, H.S. Park, B.P. Orner, J.T. Ernst, H.G. Wang, S.M. Sebt, A.D. Hamilton, Terphenyl-based Bak BH3 α -helical proteomimetics as low-molecular-weight antagonists of Bcl-x_L, *J. Am. Chem. Soc.* **2005**, 127(29): 10191-10196.
- [678] H. Yin, G.I. Lee, H.S. Park, G.A. Payne, J.M. Rodriguez, S.M. Sebt, A.D. Hamilton, Terphenyl-based helical mimetics that disrupt the p53/HDM2 interaction, *Angew. Chem., Int. Ed. Engl.* **2005**, 44(18): 2764-2767.
- [679] B.P. Orner, J.T. Ernst, A.D. Hamilton, Toward proteomimetics: Terphenyl derivatives as structural and functional mimics of extended regions of an α -helix, *J. Am. Chem. Soc.* **2001**, 123(22): 5382-5383.
- [680] L. Moisan, S. Odermatt, N. Gombosuren, A. Carella, J. Rebek, Synthesis of an oxazole-pyrrole-piperazine scaffold as an α -helix mimetic, *Eur. J. Org. Chem.* **2008**, (10): 1673-1676.
- [681] S.M. Biros, L. Moisan, E. Mann, A. Carella, D. Zhai, J.C. Reed, J. Rebek, Heterocyclic α -helix mimetics for targeting protein-protein interactions, *Bioorg. Med. Chem. Lett.* **2007**, 17(16): 4641-4645.
- [682] P. Mukherjee, P. Desai, Y.D. Zhou, M. Avery, Targeting the BH3 Domain Mediated Protein-Protein Interaction of Bcl-x_L through Virtual Screening, *J. Chem. Inf. Model.* **2010**, 50(5): 906-923.
- [683] I.C. Kim, A.D. Hamilton, Diphenylindane-based proteomimetics reproduce the projection of the *i*, *i*+3, *i*+4, and *i*+7 residues on an α -helix, *Org. Lett.* **2006**, 8(9): 1751-1754.
- [684] J.M. Davis, A. Truong, A.D. Hamilton, Synthesis of a 2,3';6',3"-Terpyridine Scaffold as an α -Helix Mimetic, *Org. Lett.* **2005**, 7(24): 5405-5408.
- [685] A. Bender, H.Y. Mussa, R.C. Glen, S. Reiling, Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): Evaluation of performance, *J. Chem. Inf. Comput. Sci.* **2004**, 44(5): 1708-1718.
- [686] C.C. Wu, T.K. Li, L. Farh, L.Y. Lin, T.S. Lin, Y.J. Yu, T.J. Yen, C.W. Chiang, N.L. Chan, Structural Basis of Type II Topoisomerase Inhibition by the Anticancer Drug Etoposide, *Science* **2011**, 333(6041): 459-462.
- [687] M. Wirth, V. Zoete, O. Michielin, W.H.B. Sauer, SwissBioisostere: a database of molecular replacements for ligand design, *Nucleic Acids Res.* **2013**, 41(D1): D1137-D1143.
- [688] A. Elsasser, M. Franzen, A. Kohlmann, M. Weisser, S. Schnittger, C. Schoch, V.A. Reddy, S. Burel, D.E. Zhang, M. Ueffing, *et al.*, The fusion protein AML1-ETO in acute myeloid leukemia with translocation t(8;21) induces c-jun protein expression via the proximal AP-1 site of the c-jun promoter in an indirect, JNK-dependent manner, *Oncogene* **2003**, 22(36): 5646-5657.
- [689] F.H. Gao, Q. Wang, Y.L. Wu, X. Li, K.W. Zhao, G.Q. Chen, c-Jun N-terminal kinase mediates AML1-ETO protein-induced connexin-43 expression, *Biochem. Biophys. Res. Commun.* **2007**, 356(2): 505-511.

-
- [690] V. Srivastava, A.S. Negi, J.K. Kumar, M.M. Gupta, S.P.S. Khanuja, Plant-based anticancer molecules: A chemical and biological profile of some important leads, *Bioorg. Med. Chem.* **2005**, 13(21): 5892-5908.
- [691] N. Homeyer, F. Stoll, A. Hillisch, H. Gohlke, Binding free energy calculations for lead optimization: Assessment of their accuracy in an industrial drug design context, *submitted* **2014**.
- [692] J. Speck, C. Rauber, T. Kukenshoner, C. Niemoller, K.J. Mueller, P. Schleberger, P. Dondapati, J. Hecky, K.M. Arndt, K.M. Muller, TAT hitchhiker selection expanded to folding helpers, multimeric interactions and combinations with protein fragment complementation, *Protein Eng., Des. Sel.* **2013**, 26(3): 225-242.
- [693] L. Spanier, E. Ciglia, F.K. Hansen, K. Kuna, W. Frank, H. Gohlke, T. Kurz, Design, synthesis, and conformational analysis of trispyrimidonamides as α -helix mimetics, *J. Org. Chem.* **2014**, 79: 1582–1593.
- [694] E. Ciglia, J. Vergin, S. Reimann, S.H.J. Smits, L. Schmitt, G. Groth, H. Gohlke, Resolving Hot Spots in the C-Terminal Dimerization Domain that Determine the Stability of the Molecular Chaperone Hsp90, *PLoS One* **2014**, 9(4): e96031.
- [695] A. Metz, E. Ciglia, T. Kröger, H. Gohlke, Protein-Protein-Interaktionen nachgeahmt: Strukturbasierte Identifizierung niedermolekularer Inhibitoren, *GIT - Labor-Fachzeitschrift (in press)* **2014**.
- [696] H. Porumb, Optical spectroscopy investigation of peptides issued from the AML1-ETO-E-protein complex relevant to acute myeloid leukemia, *Spectroscopy* **2008**, 22(4): 251-260.
- [697] W.L. DeLano, The PyMOL Molecular Graphics System. DeLano Scientific, CA, USA. Palo Alto, **2002**.
- [698] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, *J. Mol. Graphics Modell.* **1996**, 14(1): 33-38.
- [699] T. Williams, C. Kelley, H.B. Broeker, E.A. Merritt, *gnuplot - An Interactive Plotting Program*, **2007**.
- [700] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, *International AAAI Conference on Weblogs and Social Media* **2009**.

Publication I

Modulating Protein-Protein Interactions: From Structural Determinants of Binding to Druggability Prediction to Application

Alexander Metz,[§] Emanuele Ciglia,[§] and Holger Gohlke

Current Pharmaceutical Design, **2012**; 18: 4630-4647.

[§]Both authors have contributed equally to the respective work.

Modulating Protein-Protein Interactions: From Structural Determinants of Binding to Druggability Prediction to Application

Alexander Metz[§], Emanuele Ciglia[§] and Holger Gohlke*

Institute for Pharmaceutical and Medicinal Chemistry, Department of Mathematics and Natural Sciences, Heinrich-Heine-University, Düsseldorf, Germany

Abstract: During the last decades, a large amount of evidence has been gathered on the importance of protein-protein interactions in tuning and regulating most important biological processes. Since many of these pathways are deeply involved in diseases, extensive research efforts have been undertaken towards the modulation of protein-protein interactions. At the early stage of this challenge most of the attention was drawn to the drawbacks of such a therapeutic approach. Encouragingly, however, several recent studies provided a proof of concept that protein-protein interactions are actually valuable targets and that they do have a promising therapeutic potential.

This review is divided into three sections. In the first section we summarize the general features of protein-protein interfaces, focusing on the characteristics that make them different from classical protein-ligand binding sites, as well as on problematic aspects that hamper the application of classical drug discovery approaches. In the second section, we present how some of the characteristics of protein-protein interactions can be exploited fruitfully in drug design, hence focusing on the druggability of protein-protein interfaces. Methods successfully applied to protein-protein interactions will be introduced, giving special attention to the computational ones. In the third section, three case studies are presented. First, we describe protein-protein interaction modulators targeting HDM2 and the computational methods applied to identify them. Next, we present the retrospective application of the discussed approaches on the well-examined target IL-2. We conclude with a prospective application to the NHR2 protein, a target just recently validated experimentally with the aid of computational methods.

Keywords: Alanine scanning, binding pocket detection, druggability, DrugScore, high throughput screening, hot spots, IL-2, MM-GBSA, NHR2, virtual screening.

INTRODUCTION

Recently, multiprotein complexes have become attractive targets for drug discovery [1, 2] due to the essential role of non-covalent association of proteins in the communication of cell components [3]. This is highlighted by the importance of these systems in signaling [4-7] and the regulation of, e.g., cellular growth [1] and apoptosis [8, 9]. It does not come as a surprise then that protein-protein interactions (PPIs) are involved in many diseases, such as cancer, neurodegenerative diseases [10], and viral and bacterial infections [11]. For this reason, interfering with PPIs has a great therapeutic potential, providing attractive opportunities for pharmacological intervention [3, 11-14]. However, modulating PPIs is a daunting task. First, in contrast to “classical” targets such as enzymes or receptors, much less experience has been gained so far due to the novelty of many protein-protein targets. Second, the intrinsic complexity of PPIs requires innovative methodological approaches. Encouragingly, extensive investigations have proved the general feasibility of interfering with PPIs as a valuable approach for treating a number of diseases [3, 11-15]. Here, the most important goal is to identify small molecules protein-protein interaction modulators (PPIMs) that efficiently and selectively affect processes involving protein-protein binding. These successes have benefited from remarkable steps towards an understanding of PPI properties, the determinants of binding to protein-protein interfaces, as well as the implications of modulating PPIs for biological systems. This knowledge originates from an interdisciplinary approach, including the fields of structural biology, biochemistry, genomics, medicinal chemistry, and computational chemistry.

In this review, first, we present those characteristics of PPIs that are important for molecular recognition with a special emphasis on how to identify the determinants of binding of PPIMs, which

is important from a drug discovery point of view. Second, we focus on the question of how to estimate the druggability of protein-protein interfaces. Finally, we describe case studies that elucidate the application of the approaches discussed in the first two parts.

PROTEIN-PROTEIN INTERACTIONS: FUNCTIONAL AND STRUCTURAL ASPECTS

“Classical” targets versus PPIs. In the case of protein-ligand binding, an enzyme or receptor (hereafter together referred to as “receptor”) interacts with a small molecule or a peptide within a relatively small and well-defined binding site located in a cavity on the receptor surface. When there are no conformational changes on the binding site of the receptor, this situation can be described by the simplistic “lock-and-key” model already suggested by Emil Fischer [16]. According to this model, high affinity and specificity are achieved through shape and chemical complementarity, leading to a compact and tight fit between the binding partners [17]. When trying to interfere with such a system, the most direct and obvious approach is to develop small molecules resembling the natural ligand, i.e., bearing chemical groups that can be accommodated by and form interactions with the binding site of the receptor. Especially for enzyme targets, it is possible to identify protein families that share the same biological function [14]. Usually, members of the same protein family have common interaction mechanisms and binding pocket architectures. This allows exploiting information gained on one enzyme when trying to identify small-molecule ligands for other targets of the same family [18-22]. As discussed in the following, PPI targets are intrinsically different from “classical” targets, such as enzymes and G protein-coupled receptors. This makes it difficult to target protein-protein interfaces by approaches established for classical targets.

Surface size and shape of PPIs. Structural characteristics provide the biggest challenge when aiming at modulating PPIs. First, on a global level, protein-protein interfaces are generally much larger than binding site regions of classical targets. In fact, ligand-receptor contact areas are typically about 300 to 1000 Å² in

*Address correspondence to this author at the Universitätsstr. 1, 40225 Düsseldorf, Germany; Tel: (+49) 211 81-13662, Fax: (+49) 211 81-13847, E-mail: gohlke@uni-duesseldorf.de

[§]Both authors contributed equally to this work.

size [23-25], while protein-protein contact areas can range from ~ 1500 to 3000 \AA^2 or even be larger [26, 27]. Second, protein-protein interfaces are often shallow and lack deep grooves or indentations (Fig. 1a) that are usually present in classical targets (Fig. 1b). Third, interactions between protein binding partners often occur through several, not necessarily sequentially connected spots, thus leading to a discontinuous epitope. All of the above make identification of a spatially defined region within the interface that is responsible for binding a difficult task. Encouragingly, counterexamples have been presented that benefited from a deep knowledge of the respective protein-protein interface [11]. Finally, proteins are usually promiscuous molecules [28, 29] that are able to bind more than one binding partner, possibly even at the same site. While this allows proteins to take part in intricate interaction networks, it increases the level of difficulty for finding a small molecule that modulates a specific protein-protein interaction only.

Specificity and complementarity. Cells are crowded environments and, hence, potentially all molecules populating the same cellular compartment can contact each other [30, 31]. Accordingly, it is especially important for proteins that essential interactions maintain a high degree of specificity and occur only when needed, limiting the myriad of possible contacts [30, 32]. Thus, identifying the determinants of binding at protein-protein interfaces is an important goal in molecular biology with high relevance also in related fields, notably in pharmacology, genomics, and biological chemistry. Although no common strategy can be devised to achieve binding affinity and specificity in PPIs, one can nevertheless identify some mechanisms that occur preferentially in PPIs. First, proteins are marginally stable molecules [33] forming an ensemble of conformational states, each of which could potentially interact with a binding partner [31]. These conformational changes can result in the formation of cavities in the interfaces that could not be detected by visual inspection of the static representation of a crystal structure [34, 35]. That way, proteins can exhibit grooves that allow for molecular recognition and binding [36]. Therefore, it is worth investigating conformational ensembles in solution by analyzing the dynamics of the protein of interest in detail. Several tools can assist in this task, among them NMR and scattering techniques [37, 38] for determining protein structures in solution and molecular dynamics (MD) simulations for exploring the dynamic behavior of the system by computational means [34, 39-45]. The importance of accounting for receptor flexibility to identify adequate receptor conformations complementary to a PPIM is demonstrated in a study by Isvoran *et al.* [46] combining both experimental and computational approaches. Here, docking into multiple crystallographic and NMR receptor structures in connection with complex relaxation and rescoring identified binding poses of a terphenyl PPIM with calmodulin and human centrin 2 that are considerably closer to the native one than those from docking into individual, non-relaxed, and non-complementary structures.

Nussinov *et al.* pointed out that in protein-protein interfaces *unfilled pockets* and *complemented pockets* can be distinguished [47]. Unfilled pockets are present both before and after protein-protein association. They are not crucial for complex formation, but are important for the overall flexibility. In contrast, complemented pockets are detectable at the interface before binding, but disappear after association. These pockets are then filled by the binding partner, being responsible for tight and highly complementary binding of the proteins involved. The same authors also demonstrated that pre-existing pockets do not undergo significant rearrangement after binding. This means that complemented pockets offer a favorable setting for binding interactions. Interestingly, they also found that there is a weak correlation between the conservation of residues and their frequency of occurrence in complemented pockets [47]. Such residues are often also hot spots because of their enlarged contact area and the exclusion from solvent [47, 48]. Conversely, this implies that it should be possible to identify hot spots

and, hence, complemented pockets through the identification of conserved amino acids [49].

Hot spots. A fundamental characteristic of protein-protein interfaces is their energetic non-homogeneity [50]. Evidence from alanine scanning experiments shows that the binding energy is not equally distributed among all amino acids participating in the interaction [51-54]. Within the large surfaces involved in the interaction, generally some patches suffice for complex formation, the so-called *hot regions* [55]. These often have a conserved residue composition for binding similar proteins but can also differ in composition for promiscuous binding by the same interface [56]. Furthermore, only some of the residues belonging to these regions account for most of the binding energy. These amino acids are called *hot spots* if, by definition, a mutation to alanine leads to a change in the binding free energy of $\geq 2 \text{ kcal mol}^{-1}$ [57]. Hot spot amino acids on one face of the complex are usually located in correspondence to hot spots on the other face, forming interactions that lead to complex stabilization [29]. Within the hot regions, there is a very tight geometric and energetic complementarity between the binding partners. Thus, bulky side chains on one protein are accommodated in indentations on the other protein, hydrophobic groups on one protein form close contacts with hydrophobic groups on the other protein, and polar residues establish hydrogen bonds or salt bridges between the two proteins. Rajamani *et al.* showed that *anchor residues*, which are highly buried, preordered in the unbound state, (structurally) conserved, and often energetic hot spots of PPIs, are present in many protein-protein interfaces and can possibly be exploited as starting points for PPIM development [58]. Similarly, Yogurtcu *et al.* found that hot spots are more rigid than the surrounding interface in MD simulations [59]. Hot spots within one hot region work together in a cooperative fashion, thereby stabilizing the complex [31, 55, 60]. In contrast, energetic contributions from different patches are additive [61-63], suggesting that hot regions are independent from each other. As a consequence, protein-protein interfaces appear to have a hierarchical and modular architecture being formed by separate patches, within which each hot spot amino acid strongly depends on the other close-by amino acids for an efficient interaction [31]. Interestingly, hot spots are among the most conserved residues [47, 48, 64, 65]. This relation has also been proposed to be a way to distinguish between binding interfaces and otherwise exposed protein surfaces [48]. This hypothesis is strengthened by the observation that no residue conservation was found within solvent exposed surfaces [48]. Overall, this highlights the importance of hot spots for protein-protein complex formation and explains why evolutionary changes rarely lead to a significant modification in hot spot composition [64].

Although the leading role in driving the interaction between protein binding partners relies on hot spots, the surrounding amino acids are also important. According to the *O-ring theory* [51], surrounding residues have the function to protect hot spots from solvent molecules, favoring hydrophilic or even hydrophobic interactions that would be otherwise disturbed by the presence of water. A high degree of complementarity between the binding partners is sometimes also achieved through water mediated interactions [66-68]. Such structural water molecules are particularly important in regulating hydrogen bond networks within the interface: I) by bridging interactions between the binding partners or II) by favoring the formation of a dry core in the interface that maximizes the interactions between hot spots surrounded by a rim of amino acids and water molecules [24].

Interaction types and amino acid composition. Given that protein-protein interfaces have considerable areas of hydrophobic residues, resembling cores of globular proteins [69], it has been suggested that the hydrophobic effect is the driving force leading to protein-protein association [70-72]. However, a careful analysis shows a situation similar to protein folding: the hydrophobic effect is a leading force but the proteins do not necessarily adopt a con-

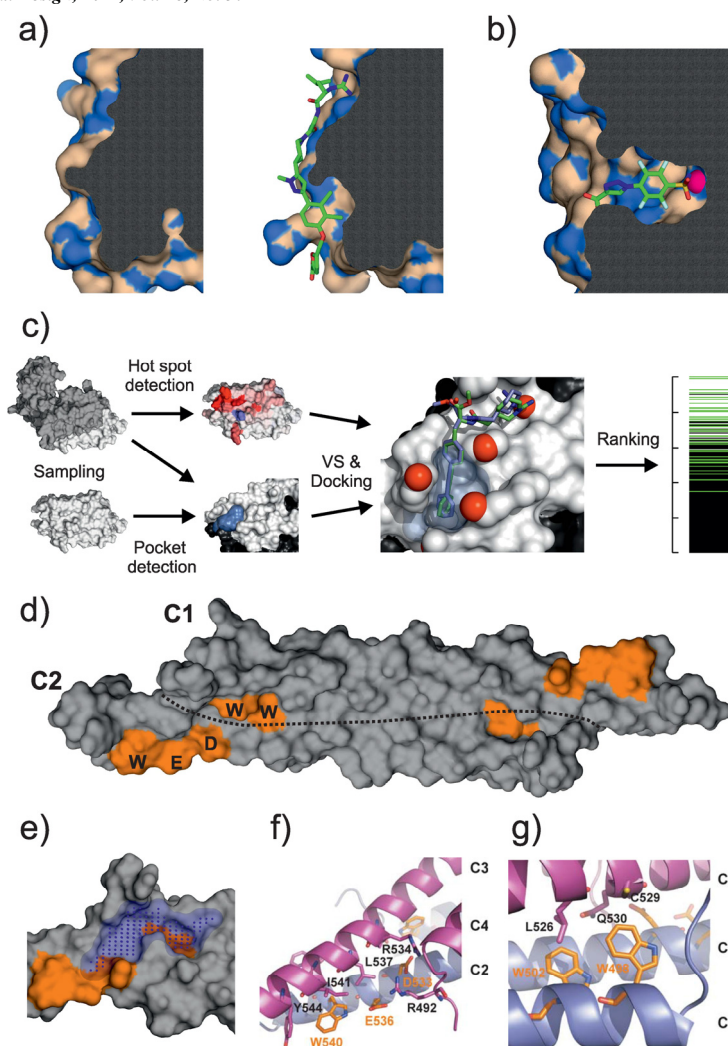


Fig. (1). Druggable binding sites at protein-protein interfaces. (a) Protein-protein interfaces usually lack deep pockets in the unbound state, as depicted for unbound IL-2 (left panel, PDB code: 1m47). Still, surface flexibility can allow for the formation of druggable pockets, as depicted for a PPIM-bound conformation of IL-2 (right panel, PDB code: 1py2). (b) In contrast classical targets have narrow and deep binding pockets, as depicted for carbonic anhydrase in complex with an inhibitor (PDB code: 3p25). The surface coloring in (a) and (b) highlights polar (blue) and non-polar (beige) atoms; Zn^{2+} highlighted in purple. (c) Strategy for PPIM identification on protein-protein interfaces. Surface coloring of hot spots (red) on a linear color scale as calculated by per-residue free energy decomposition; pocket volume depicted as blue surface. (d) Hot spots (W498, W502, D533, E536, and W540) in the dimer/dimer interface of NHR2 (PDB code: 1wq6) calculated by per-residue MM-PBSA free energy decomposition. The dotted line represents the border between the two antiparallel α -helices C1 and C2. (e) Potentially druggable pocket in the dimer/dimer interface of NHR2. The blue dots mark the location of the largest indentation in the binding epitope. Intermolecular contacts involving (f) hot spots W540, D533, and E536 on helix C2 as well as (g) hot spots W502 and W498 on helix C1.

formation with optimally buried non-polar surface area [73-76]. This hints at further mechanisms being involved. In fact, even though hydrophobicity is important in this context, the role of electrostatic interactions cannot be neglected [77-80]. In fact, the hydrophobicity of protein-protein interfaces is usually intermediate between the one found for a protein core and a solvent exposed protein surface. The amino acid composition in the hot spots, which has been shown to be non-random [81], reflects this situation. In fact, it has been observed that hot spots are enriched in tryptophan, tyrosine, arginine, and, to a lesser extent, isoleucine [50], whereas leucine, serine, threonine, and valine were found slightly depleted [48, 51]. One could argue that large side chains just contribute more, but functional considerations prevail. In particular, tyrosine and tryptophan allow establishing stacking and hydrophobic

interactions owing to their aromatic, non-polar side chains, but at the same time offer the possibility to create hydrogen bonds due to the phenolic OH group and the indolic nitrogen. On the contrary, arginine, being a polar amino acid bearing a charged guanidinium group, is mostly involved in hydrogen bonds and salt bridges across the interface although the electron delocalization of the guanidinium π -system also confers a pseudo-aromatic character [51]. This dual side chain behavior exemplifies the two-faced chemical nature of protein-protein interfaces. As a word of caution, even though the mentioned residues are the most frequent ones in PPIs, this knowledge should neither lead to neglecting the importance of other amino acids for binding nor to uncritically considering these residues hot spots just because of their occurrence in an interface.

DRUGGABILITY OF PROTEIN-PROTEIN INTERACTIONS

PPIs are far from being widely exploited targets in drug development. Even though there are some examples of marketed small-molecule drugs acting on PPIs [3, 82-86] and some further molecules are in advanced clinical trials (Fig. 2) [1, 87-92], PPIs are usually considered high risk targets by pharmaceutical companies [2, 93]. This is for two reasons: First, initial attempts to identify PPIMs by high-throughput screening (HTS) were mostly unsuccessful, particularly when using chemical libraries designed for traditional targets [11]. Second, the wideness of protein-protein interfaces, the lack of defined binding pockets, and the stability of PPIs led to PPIs being considered difficult to target if not undruggable [93, 94]. Also due to this overgeneralization, there is still a large gap between the knowledge gathered on these systems [1, 3, 12-14] and its actual use in the development of therapeutics. Yet, some prominent counterexamples, such as the well-studied systems p53/MDM2 [11, 14] and Bcl-x_i/Bac [11], have contributed significantly to expose the myth depicting PPIs as undruggable systems [29, 94].

A big challenge associated with PPIs is the high degree of diversity in terms of the molecular recognition properties encountered. Each interface is unique, bearing its own particular characteristics and, thus, requiring a specifically tailored approach. In fact, binding sites at protein-protein interfaces are often not well conserved, which is different from enzymes that bind the same type of substrate and, therefore, share many common features in the binding regions if they belong to the same family [14]. Nevertheless, as the amount of structural data of bound PPIMs increases approaches that exploit PPIM binding information from homologues [95] will become increasingly applicable. Furthermore, there are differences between protein-protein interfaces and the non-interacting surface of a protein that allow the sequence- and structure-based prediction of residues in the interface and an enrichment of hot spots, which often stand out in such analyses [53, 96-100].

For establishing the suitability of a protein-protein interface as a target for drug discovery, first, one needs to define what is meant by "druggability" in this context. In the straightforward definition of Egner and Hillig, druggability can be considered as the likelihood of finding a selective, low molecular weight molecule that binds with high affinity to the target [101]. But what are the characteristics of a PPI that allow targeting the interface? Due to the inherent complexity of the issue, it seems impossible to answer this question unambiguously. Aside from the particular characteristics of protein-protein interfaces, as presented earlier, it is important to consider that druggability is not an absolute property of a target molecule such as chemical class, molecular weight, or logP, but always refers to a specific application. Accordingly, authors have provided different concepts for assessing druggability both qualitatively and quantitatively [14, 23, 101-106]. Utilizing computational techniques to assess a target's druggability is appealing. An important reason for this is that it should permit to cut down research costs relating to experimental investigations that otherwise must be carried out in a more extensive fashion. However, despite large research efforts, initial progress is only emerging in this field [107].

Even though a unified approach to unarguably establish the druggability for a certain PPI is not available yet, there are some general considerations valid for all PPIs, which can be used for a preliminary assessment. An interesting approach to select protein-protein interfaces suitable for drug discovery is the decision tree proposed by Chene [14]. The author showed that considerations on the physicochemical properties of an interface allow assessing whether a PPI could be a suitable target for the design of modulators. A first point concerns the natural binding state of the protein of interest, i.e., whether it falls within the *obligate* or *non-obligate* class of protein-protein complexes. In the former class, the monomers involved do not exist in the non-associated form in the cell, while in the latter class the protein binding partners can be bound or

dissociated at different times or conditions. Consequently, targeting a permanent PPI should be much harder than a transient one. Other important factors to be considered are the availability of structural information, the presence of cavities, the degree of interface hydrophobicity, and the size and complementarity of the interface. In an ideal case, there is a detailed characterization of the PPI by structural studies that clarifies the determinants of binding. Next, there should be cavities on the surface with appropriate sizes to accommodate PPIMs and to allow specific targeting. In addition, the overall hydrophobic character of the interface should be intermediate, permitting to develop molecules with an adequate trade-off between optimal binding and favorable pharmacokinetic properties.

Another important factor influencing the druggability of PPIs is the presence of helices at the interface. With α -helices being the most frequently occurring type of secondary structure both in the protein core and in exposed regions [108, 109], helices located on accessible surfaces are often responsible for molecular recognition. Along these lines, a survey on the Protein Data Bank [110] by Arora *et al.* revealed that 62% of the protein-protein complexes present in the database have helical interfaces [105, 106]. Furthermore, the authors divided these interfaces into three categories according to the helical character: I) receptors containing a cleft for helix binding, where a minimum of two close amino acids contribute importantly to the interaction (as in the p53/MDM2 complex [11, 14, 111]); II) extended interfaces where strong binding is achieved through multiple contacts between two- to five-turn helices and a higher number of residues; III) proteins featuring both of the described characteristics and showing quite weak interactions [106, 108]. From a drug discovery point of view, it appears obvious that complexes belonging to the first category offer better chances for developing PPIMs than complexes falling in the second and third categories. In addition, knowledge about how amino acids are arranged within interfacial helices can guide the design of α -helix mimetics with different chemical scaffolds [108]. This may be a first step in the development of PPIMs.

Any analysis of a protein-protein interface should take these aspects into account in order to assess the druggability of the system. In addition, (computer-aided) binding pocket and hot spot detection have a great impact for characterizing the PPI and assessing the druggability of a protein-protein interface. These methods will therefore be presented in the following.

BINDING POCKET DETECTION

Identifying binding pockets in protein-protein interfaces.

Identifying binding pockets is often the first step in assessing target druggability and has important implications for docking and structure-based drug design [102, 112]. In fact, when the goal is to develop a PPIM with drug-like characteristics (e.g., as compliant as possible with Lipinski's rules [113]) it is necessary to figure out where such a molecule can efficiently bind to the target interface. In an ideal situation, knowledge on experimentally validated binding sites is available from the literatures but this is not always the case. Additionally, proteins are usually part of complex interaction networks [114] such that multiple binding interfaces can be present [112], which may even be interlinked allosterically. Therefore, the choice of the correct target site is affecting the entire drug discovery pipeline, and caution should be taken in identifying and evaluating this site.

Binding is a complex event arising from several factors of which shape and physicochemical complementarity are of major importance [112]. Accordingly, binding pocket detection algorithms have been developed that can be sub-divided into two major classes [36], geometry-based [115-122] and energy-based ones [123-130]. Methods using structure and sequence comparison [131-135] or techniques taking into account the dynamics of protein structures [34, 35, 107, 136-139] have been reported less frequently. Several authors reviewed the available methods [36, 107,

112, 140], especially those applied to the identification of protein-ligand binding sites in classical targets. In this review, we focus on methods applied to detect binding pockets at protein-protein interfaces.

Pocket detection algorithms applicable to classical targets might not be readily applicable to protein-protein interfaces due to the size and flatness of the latter. Fuller *et al.* compared protein-ligand and protein-protein binding interfaces with Q-SiteFinder [93, 125]. Q-SiteFinder scans the protein surface with hydrophobic probes, clusters positions of favorable interaction energy, and ranks these clusters by their accumulated interaction energy to identify the most important binding sites. This procedure confirmed that classical targets exhibit larger pocket volumes than pockets located in protein-protein interfaces. Thus, inhibitors of classical targets tend to target a single high-volume pocket, while PPI inhibitors target multiple smaller pockets [93, 125]. The immediate consequence is that the identification of binding sites in a protein-protein interface solely based on geometrical considerations is challenging [34]. Q-SiteFinder could also successfully identify the actual PPIM binding pocket in the unbound conformation of PPIM targets [93]. Even when employing a rather stringent scheme to assess the accuracy of found pockets, penalizing overly large pockets reaching beyond the ligand, Q-SiteFinder was able to find the correct binding pocket in the top three predicted sites for 90% of the investigated proteins [125]. As an alternative, PocketFinder defines pockets enveloping grid points with attractive Lennard-Jones potential [130]. With PocketFinder, from 5,616 ligand bound pockets 95% were detected successfully with comparable results for *apo* structures; likewise, pockets involved in PPIs were successfully detected. Bourgeois *et al.* compared binding pockets/interfaces of PPIMs and the corresponding protein-protein interfaces found in the 2P2I database to those of heterodimeric protein-protein complexes without known PPIM and identified discriminating properties [141, 142]. In addition, the authors classified those protein-protein interfaces with known PPIMs into two groups, either with a single important secondary structure element or a more globular domain in the interface, by principal component analysis and clustering with respect to the discriminating physicochemical descriptors of the proteins [141]. The physicochemical properties of PPIMs, in comparison to common drugs, were found to be shifted towards higher molecular weights, hydrophobicity, and rigidity as well as towards an increased occurrence of aromatic moieties [142].

Molecular dynamics simulations for binding pocket detection. As it is often not possible to identify well-defined binding pockets at the interface present in the crystal structure of a protein-protein complex, it is important to go beyond analyzing the static structure of PPIs and take into account protein dynamics [29, 44, 45]. This yields a more detailed view of the conformational space accessible to a protein-protein interface. For example, it was shown that nanosecond MD simulation started from an unbound conformation can sample a bound conformation in many cases [34, 35, 143]. Accordingly, Eyrisch and Helms successfully applied a pocket detection protocol that makes use of MD simulation-derived conformations and the PASS (putative active sites with spheres) algorithm [122]. Starting from the crystal structure of unbound Bcl-x_L, IL-2, and MDM2, the authors were able to detect binding pockets that would have been missed when just applying any of the available algorithms to the unbound conformation because of the transient nature of these pockets [34]. Mimicking the experimental multiple solvent crystal structure (MSCS) [144] and NMR solvent mapping experiments, cosolvent MD simulations favor the opening of hydrophobic binding pockets that would be unfavorable in purely aqueous environment. Yang *et al.* recently demonstrated the feasibility of this approach for Bcl-x_L [43]. Miranker and Karplus presented the multiple copy simultaneous search (MCSS) where thousands of probe molecules are positioned on a protein interface and energy minimized to identify favorable binding sites [145] and

potentially plastic interface regions [146]. MCSS has been applied to identify anchor residues in PPIs and to design peptidic Ras/Raf inhibitors [147]. Landon *et al.* applied the CS-MAP computational solvent mapping method [148] to a clustered structural ensemble of H5N1 avian influenza neuraminidase generated by MD simulation to identify novel hot spots while accounting for target flexibility [149]. Eyrisch *et al.* presented a comparative study showing that more and larger pockets open in methanol cosolvent MD simulations than when performing conformational sampling based on normal mode analysis or by (t)CONCOORD [42]. Coarse-grained simulations have been shown to be capable of sampling the bound state of proteins starting from an unbound one, too [150, 151]. Accordingly, the grid-based pocket detection algorithm PocketAnalyzer^{PCA} [152] was successfully applied to identify PPIM binding sites in ensembles of IL-2 generated by a constrained geometric simulation method; these sites provided an entry point for a subsequent virtual screening (see also below) [35].

Machine learning approaches. Machine learning-based and empirical scoring functions have been applied as binding site prediction methods for PPIs. These methods make use of the characteristic differences between binding sites in protein-protein interfaces and the remaining protein surface, e.g., with respect to sequence conservation, amino-acid occurrence, secondary structure, solvent accessibility, and side chain conformational entropy [153]. For a description of the numerous tools and web servers available we refer to more detailed reviews [53, 153]. Recently, Li *et al.* used ray casting to identify pockets and protrusions in protein surfaces, which can be used as a filter for detecting surface shape complementarity and help speeding up protein-protein docking [154]. Tan *et al.* presented a pocket detection algorithm based on *depth* [155], i.e., the distance of a residue to the solvent, which has been proposed to be superior to using the solvent accessible surface area (SASA) for the prediction of shallow binding pockets [156]. Ertekin *et al.* showed that residues near cavities exhibit high frequency vibrations (HFVs) that can be identified using an elastic network model [157]. Fragment docking has been presented as a valuable tool to identify pockets and evaluate their druggability [158].

A characterization of a protein-protein interface that aims at identifying potential binding sites for small molecules is definitely an important starting point for any drug discovery project targeting a PPI. Nevertheless, as already pointed out by Cheng *et al.*, the presence of a pocket on the protein surface is “necessary but not sufficient” for the development of PPIMs [23], and additional investigations of the interface are in order.

HOT SPOT DETECTION

Spatially clustered hot spots are crucial for the binding of small drug-like PPIMs in a large protein-protein interface [11, 50]. Thus, methods for the detection of hot spots [96] do not only provide a more detailed understanding of the energetic determinants of binding but yield information that complements the one derived from binding pocket detection. Initially, we will briefly introduce experimental methods for hot spot determination, which is followed by a more detailed discussion of computational methods.

Experimental hot spot detection. Mutagenesis of interface amino acids is the most significant method to detect and validate hot spots. Mutating selected or, seldom, all such amino acids to alanine is called *alanine scanning* and yields a finger print of the amino acids important for a PPI [54, 159]. A mutation to alanine is usually chosen because it has a small neutral side chain devoid of polar interactions and does not significantly influence the protein backbone as, e.g., glycine would do. Still, even a mutation to alanine can potentially introduce larger structural changes in the complex or influence the unbound state of a protein such that changes in relative binding free energies observed between wild-type and mutant complex do not necessarily originate from interac-

tions lost in the interface [159]. Also, if alanine partially carries over interactions of the original amino acid, e.g., in terms of backbone interactions or because the original amino acid is similar to alanine, the change of affinity upon mutation will be less than the total contribution of the original amino acid. Furthermore, alanine scanning is very laborious because it requires protein purification and analysis. This bottleneck can be alleviated by combinatorial alanine scanning using phage display [160, 161] or combinatorial solid-supported peptide libraries [162]. Alternatively, methods of fragment-based drug design, including covalent tethering [163-166], co-crystallization [167], SAR by NMR [168, 169], and SOS-NMR [170], can identify binding fragments of rather low affinity and, thereby, probe druggability [171]. Also, solvent mapping by MSCS [172] and chemical shift perturbation (CSP) NMR experiments [173, 174] are methods that suggest where organic molecules will preferably bind and so have been exploited in data-driven docking [107, 175]. All these methods can help identify a smaller, druggable, and hot spot-containing sub-region of the interface, even if there is no open binding pocket detectable in the unbound state of the receptor [176]. Information about experimentally determined hot spots are available in several databases [57, 177-181], although the coverage is low when compared to the number of PPIs considered to be interesting drug targets.

Since experimental methods for detecting hot spots are laborious, there is a high demand for computational prediction methods. Methods for performing *in silico* hot spot detection can be categorized into: *in silico* alanine scanning, non-perturbing fully atomistic approaches, machine learning approaches, and approaches using nothing but unbound protein structures.

***In silico* alanine scanning.** Among the computational alternatives for hot spot detection that require experimentally determined or modeled structures of protein-protein complexes as input, *in silico* alanine scanning [159, 182, 183] is the most straightforward analogue of the above described experimental method. Here, a relative binding (free) energy is calculated for a wild-type complex and one with alanine mutants in the interface. Usually, intermolecular energy and (de)solvation free energy terms are considered for this, sometimes also intramolecular energies and entropic contributions. *In silico* alanine scanning often uses simple physical models or empirical (scoring) functions for assessing the energy change [182]. Therefore, *in silico* alanine scanning is usually fast and computationally modest, allowing a rapid detection of binding determinants. As a downside, these methods rely on approximations that often reduce their accuracy. FoldX [184] and Robetta [185-187] are widely used implementations of this approach. The alanine mutants are generated by side chain truncation with subsequent structural relaxation of the environment or the whole complex. The energy change is determined by an energy function whose terms have been parameterized based on experimental data. Kiel *et al.* performed an alanine scanning for Ras/RalGDS and Ras/Raf-RBD complexes and found very good correlations ($R > 0.95$) between experimentally measured changes in the free energy of binding and the prediction of FoldX [188]. In a similar study on various members of the ubiquitin domain superfold family, experimentally found hot spots could be determined by FoldX elucidating the basis of binding specificity, even though using homology models as input structures [189]. Carbonell *et al.* investigated the distribution of hot spots in protein-protein complexes of the non-redundant yeast interactome by FoldX alanine scanning and found that hot spots of promiscuous binding are located in independent modules while those of specific binding are arranged predominantly in one module [190]. Ivanov *et al.* predicted alanine mutations that disrupt the rabies virus phosphoprotein dimerization by FoldX, which agreed with results from a yeast two-hybrid assay [191]. Kortemme *et al.* correctly identified 79% out of 233 experimentally validated hot spots from 19 protein-protein complexes using the Robetta alanine scanning method [187]. Jochim *et al.* applied Robetta to all helix-

mediated PPIs in the PDB and analyzed the distribution of hot spots in these helices to propose new PPI targets and assess their druggability, e.g., by helix mimetics [105, 108]. Donald *et al.* predicted hot spots by Robetta that are located in the interface of the extracellular stalk region of the β_3 and the complementary α_{IIb} and α_v integrin subunits whose mutation lead to destabilization and thereby activation *in vivo*, although the energy threshold for predicting a hot spot was alleviated to ≥ 0.3 kcal mol⁻¹ [192]. Recently, Liu *et al.* applied a consensus strategy exploiting FoldX, Robetta, KFC, and their Z-score approach, a measure for the significance of the contribution to binding of a residue derived from knowledge-based pairwise potentials and available surface area, to compare the determinants of binding of H1N1 hemagglutinin antigen variants to an antibody [193]. Perez *et al.* predicted hot spots of the betaine transporter BetP membrane protein by applying FoldX alanine scanning to crystal structures and a structural ensemble generated by MD simulation; the hot spots were later found to disrupt the trimer [194]. We recently developed a webservice (<http://cpclab.uni-duesseldorf.de/dsppi>) for hot spot prediction in PPIs by *in silico* alanine scanning that uses the knowledge-based scoring function DrugScore^{PPI}. DrugScore^{PPI} consists of pair potentials derived from atom type-specific pair distribution functions from 851 experimental protein-protein complex structures. The weights of the pair potentials have been adapted by partial least squares regression on relative binding affinities for the so far largest set of 309 alanine scanning results [195]. DrugScore^{PPI} efficiently predicted affinity changes for an external set of 22 alanine mutants of the Ras/RalGDS complex showing higher correlation to experiment ($R = 0.66$) than FoldX ($R = 0.52$), Robetta ($R = 0.43$), and CCPBSA ($R = 0.23$). Tuncbag and Keskin *et al.* presented the HotPoint method [196], an empirical model based on knowledge-based residue pair-distribution potentials [197] and solvent occlusion to predict hot spots. Notably, many methods for hot spot prediction are applied to single experimental or modeled structures of a protein-protein complex. However, caution is needed because the hot spot detection outcome from a single complex structure may be less representative if the proteins are flexible *in vivo*. Therefore, it is preferable to perform calculations on conformational ensembles of the proteins, e.g., obtained from MD or coarse-grained simulations.

Non-perturbing fully atomistic approaches. As a complementary alternative to *in silico* alanine scanning, there are methods that calculate the contribution of individual amino acids to the free energy of binding without mutating them. In one of the first studies, Novotny *et al.* correctly predicted residues important for binding in antigen-antibody complexes by a physics-based energy approximation [198]. Similarly, the nowadays most widely used molecular mechanics-generalized Born surface area (MM-GBSA) [35, 199, 200] and the molecular mechanics-Poisson Boltzmann surface area (MM-PBSA) methods allow a per-residue decomposition of the binding free energy. A critical review of these methods also covering PPI applications has been published recently [201]. Both methods predict the total binding free energy by means of endpoint free energy calculations. The underlying energy function consists of electrostatic and van-der-Waals terms from the molecular mechanics (MM) force field, which are complemented by polar (based on the generalized Born (GB) or Poisson-Boltzmann (PB) continuum models) and non-polar (surface area-dependent (SA)) contributions to the (de)solvation free energy. Entropy changes upon binding can be determined from normal mode analysis or quasi-harmonic analysis. All of these terms can be decomposed into contributions of individual residues, which allows revealing hot spots. It is also possible to further decompose the binding free energy into pairwise contributions, which highlights important interactions between pairs of amino acids. The method is usually used for post-processing ensembles from MD simulation trajectories. If the unbound proteins and the protein-protein complex are sampled individually, this leads to the conceptually rigorous *three trajectory* approach, which takes into account energetic differences caused by

conformational changes upon complex formation but is also computationally demanding. A widely used alternative is the *single trajectory* approach in which the unbound structures are extracted from the trajectory of the complex without further relaxation. The latter approach, besides being faster, was shown to accurately reproduce experimental alanine scanning data. Indeed, the single trajectory approach often proved to be superior to the three trajectory alternative due to the cancellation of errors [35, 52]. However, it has to be mentioned that the MM-PB(GB)SA energy function has also been applied for *in silico* alanine scanning on structural ensembles from MD simulation [183, 202]. In this context, Moreira *et al.* found improved predictions when using different dielectric constants to account for the varying extent of relaxation upon mutating charged, polar, and non-polar residues to alanine [203].

CC-PBSA [204] is a conceptually related method using similar terms to calculate binding free energies but CONCOORD [205] for conformational sampling. CONCOORD generates conformational ensembles by iteratively satisfying geometrical constraints starting from a random structure. Here, weighting factors for the energy terms had to be derived by fitting to experimental data. Furthermore, the linear interaction energy approach (LIE) [206] has been applied to compare the binding energy of a large set of mutated proteases and their inhibitor proteins [207]. In the LIE approach, the binding free energy is calculated as the weighted sum of intermolecular energies and the (de)solvation energy of the ligand, and the weights are derived by fitting to experimental data.

Machine learning approaches. In addition to merely training an empirical physical model on experimental data, machine learning approaches can use structural, physicochemical, or sequence descriptors for hot spot prediction without the need for a model based on first principles. Chen *et al.* compared various hot spot prediction methods to support vector machine (SVM) models trained on different sets of sequence and structure-based descriptors and found a sequence-based SVM to outperform FoldX and Robetta amongst others on an independent test set [208]. Cho *et al.* used decision-tree based feature selection to identify properties that discriminate hot spots with their SVM model MINERVA [209]. They found hydrophobicity and π - π interactions to be hallmarks of hot spots. Additionally, they found the atomic packing density weighted by the fraction of the available surface area buried upon complex formation to be highly predictive in contrast to the raw coordination number. Even in the crystal structure of the unbound proteins, the weighted atomic packing density of hot spots was found to be significantly higher than for the remaining surface residues, although calculating the surface area buried upon complex formation still required a complex structure. Kosloff *et al.* predicted hot spots of various G protein-RGS (regulator of G protein signaling) protein interactions [210] by a combination of structure-based sequence alignment, continuum electrostatics-based per-residue electrostatics, and buried surface area-dependent non-polar interaction energy [211]. In contrast to Robetta alanine scanning, they could identify hot spots with important backbone or long-range electrostatic interactions. Combining sequence conservation with hot spots energetics they distinguish *significant & conserved* from *modulatory* residues, the former ones being essential for overall binding affinity and the latter ones for binding specificity of complexes. The authors could redesign low-affinity RGS proteins to higher affinity ones by mutating modulatory residues. A comparison of experimental hot spots from the alanine scanning energy database (ASEdb) [57] and those predicted by decision trees and SVM based on buried SASA, sequence conservation, and hot spot propensity can be found in the HotSprint database [181]. Darnell *et al.* created the decision-tree based KFC model using a combination of shape specificity and descriptors of interaction types; both descriptors show a lower accuracy when used alone [212]. Here, the decision-tree machine-learning algorithm preformed slightly better than SVM and Bayesian networks or Robetta alanine scanning

when applied to the (training) dataset. Nevertheless, the successor model KFC2, which comes in two variants based on different descriptors, is a SVM that was recently reported to outperform KFC and other machine learning methods [213]. Shulman-Peleg *et al.* investigated the spatial distribution of 3D alignments of physicochemical interactions (hydrogen-bond donor and/or acceptor, aromatic) in protein-protein interfaces of functionally similar PPIs and found conserved distribution patterns to predominantly contain hot spots; conversely, this allows hot spot prediction without the necessity of sequence conservation [64].

***In silico* hot spot detection without a complex structure.** The above methods require a complex structure as input. As this may often be not available, methods have been developed that aim at predicting hot spots solely based on protein sequence information or unbound protein structures. ISIS [214], a neural network based method using only sequence information, predicted hot spots with high precision [100] at the cost of a low sensitivity due to a strict threshold for the discrimination of hot spots [96]. To predict hot spots based on the unbound structure of a protein, the principle behind NMR solvent mapping and MSCS has been transferred into computational algorithms [215]. Cosolvent MD simulations, previously mentioned in the context of inducing pocket opening, provide the location of sites occupied by different cosolvent probes. The site's population with a cosolvent can be used to approximate a maximal binding affinity of residues at this site by applying the inverse Boltzmann principle [216]. Although other methods for sampling probe populations could be considered, e.g., Monte Carlo simulation, docking, or MCSS, these methods do not account for plasticity or do not generate a Boltzmann ensemble, thus aggravating the estimation of energy and entropy of binding. Nevertheless, Grosdidier *et al.* presented an approach to predict hot spots by protein-protein docking [217]. Here, the normalized interface propensity (NIP) of individual amino acids calculated from docking poses could predict hot spots with a high accuracy without the need to approximate an energy [218]. Recently, Geppert *et al.* presented iPred [219], which predicts protein-protein interfaces and hotspots based on the difference of local intramolecular atom- and residue-specific pair potentials between interface and non-interface residues. In conjunction with the geometry based pocket detection algorithm PocketPicker [120], iPred identified druggable sites in the interface of the unbound structure of interferon IFN- α to its receptor IFNAR. These sites could subsequently be addressed by pharmacophore based screening and docking to yield an *in vivo* active PPIM of the IFN- α /IFNAR interaction from a set of only six tested compounds [220]. In a retrospective study on more than 15 PPIs with known PPIM structures, Kozakov *et al.* predicted hot spots and binding sites to assess interface druggability by computational solvent mapping of 16 different probe molecules with FTMAP [221]. They found druggable sites to be structurally conserved between bound and unbound structures and that local side chain rearrangements, implemented by a rotamer search followed by energy minimization, suffice to accommodate for most adaptations in PPIM binding.

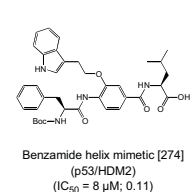
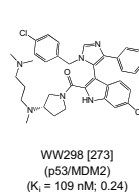
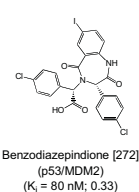
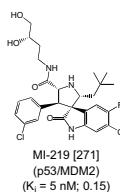
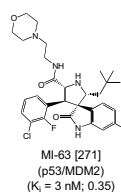
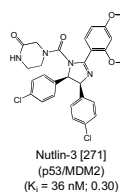
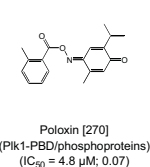
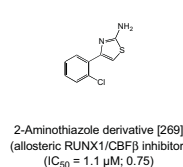
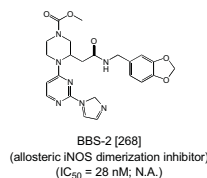
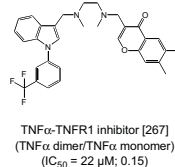
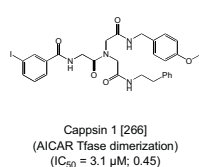
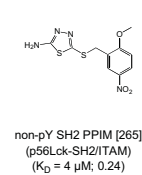
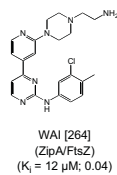
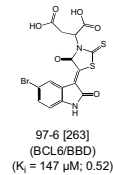
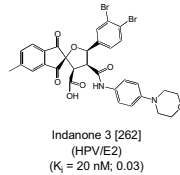
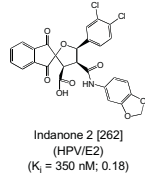
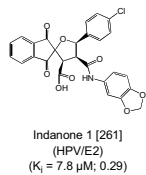
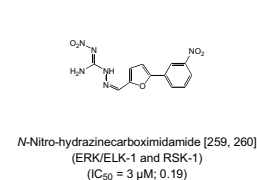
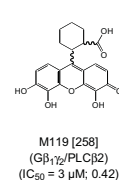
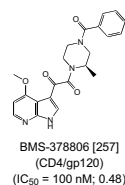
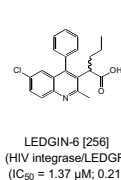
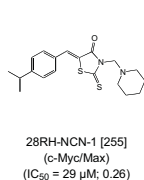
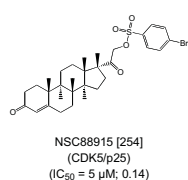
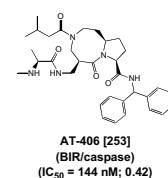
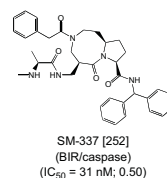
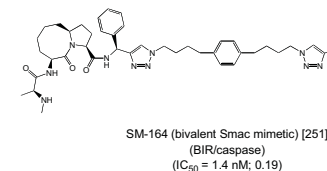
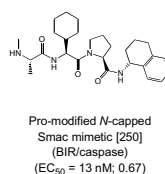
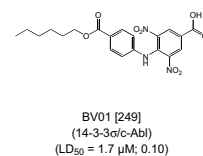
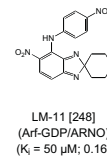
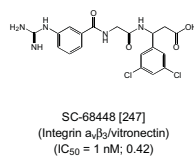
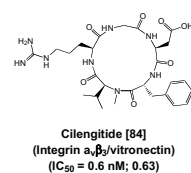
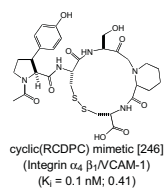
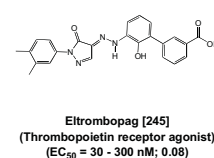
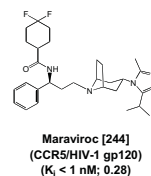
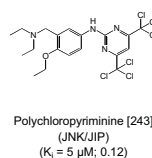
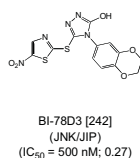
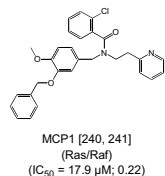
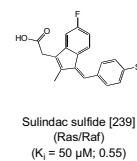
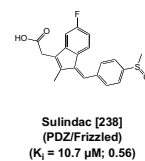
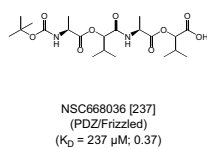
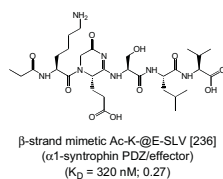
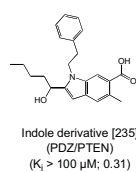
In summary, several comparative studies showed encouraging results in terms of agreement between experimental and computed results for hot spot detection [52, 195, 199, 200, 202].

CASE STUDIES

During the last two decades many studies have investigated PPIs and identified PPIMs using both experimental and computational approaches. The available experimental data has been integrated into PPI-specific databases [6, 141, 222]. Antibodies are currently the most successful class of drugs [223] inhibiting PPIs [224, 225]. As conveniently accessible high affinity PPIs, they can help reveal druggable epitopes, understand binding mechanisms [226], and may even inspire PPIM design [227] by complementing insights from non-antibody PPIs. As a drawback, antibod-

Modulating Protein-Protein Interactions

Current Pharmaceutical Design, 2012, Vol. 18, No. 30 4637



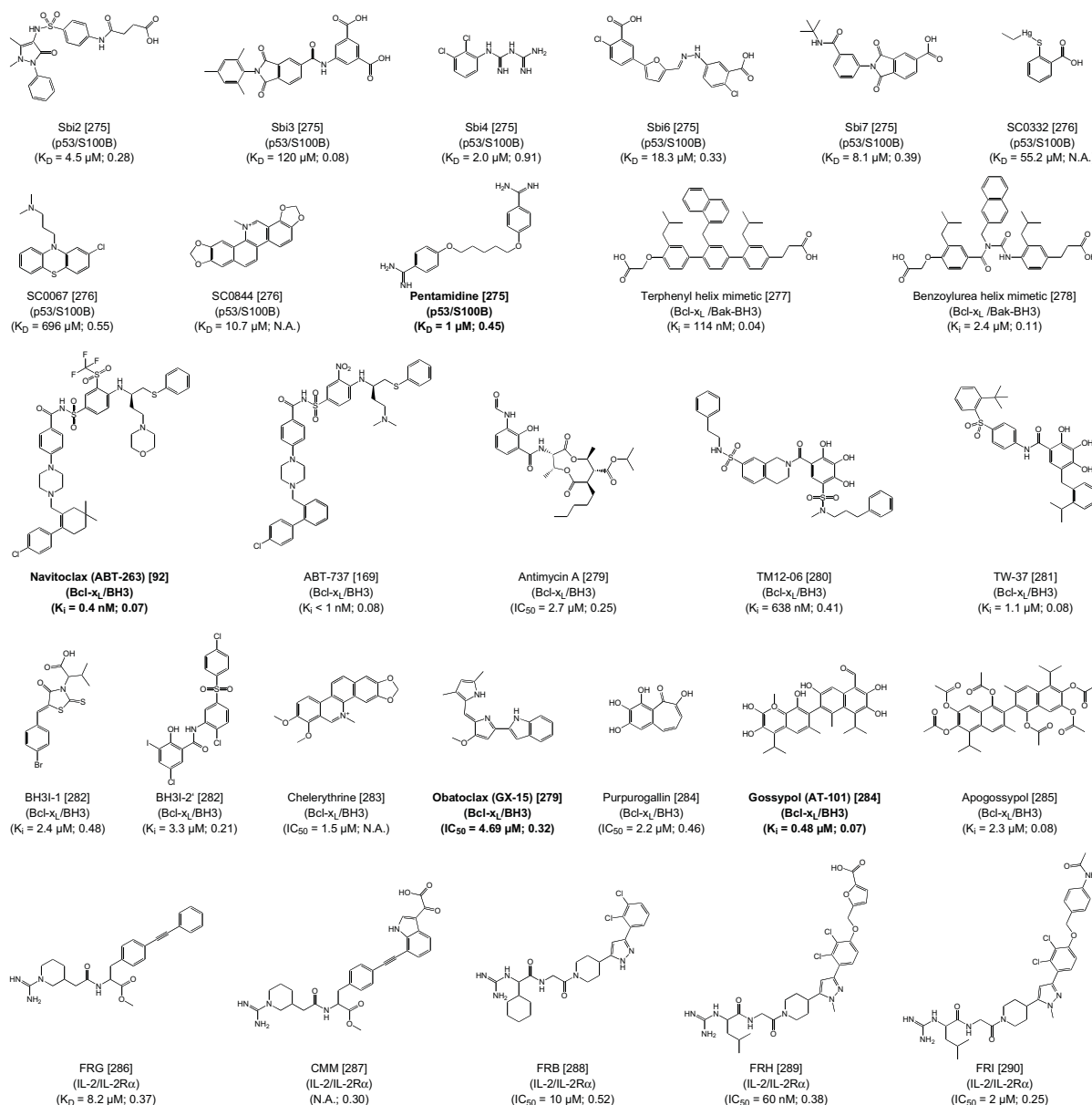


Fig. (2). A survey of protein-protein interaction modulators. Labels contain the PPIM name followed by a reference detailing the potency. Below, (PPI target/PPIM competitor) and (potency; Drug-Score) in the first and second parentheses, respectively. Labels of marketed drugs or compounds that have been the subject of clinical trials are highlighted in bold. The Drug-Score was calculated by OSIRIS Property Explorer (<http://www.organic-chemistry.org/prog/peo/druglikeness.html>).

ies are not cell permeable and lack oral bioavailability [228]. We thus focus this section on small-molecule PPIMs. As related studies and discovered PPIMs have been the subject of many detailed reviews already [3, 11, 87, 107, 229-234], we solely give a survey of the so far identified PPIMs with their reported potency (Fig. 2) without claiming completeness. In this list of PPIMs we also added a druglikeness measure, the so-called Drug-Score calculated by the OSIRIS druglikeness server. The score highlights that not all of the so far discovered PPIMs are drug-like; this also includes those be-

ing marketed or tested in clinical trials for their potential pharmacological relevance.

In addition, we describe three case studies in more detail. First, we describe PPIM design for p53/HDM2, one of the most thoroughly investigated PPI systems, as an example where a preformed binding pocket exists in the interface. Then, we describe a retro- and a prospective study from our lab concerning interleukin-2 (IL-2) and neryl homology region 2 (NHR2) proteins.

HDM2

In many tumors, p53 acts as a tumor-suppressor protein [291-293]. However, binding of the human double minute 2 (HDM2) protein (or the mouse analog MDM2), which is overexpressed in many tumors, blocks transactivation by p53 and increases p53's degradation. Thus, the p53/HDM2 interaction is an important pharmaceutical target for cancer treatment. Crystallographic structures revealed that the key interaction in the p53/MDM2 complex arises from the binding of a 15-residue α -helix of p53 into a hydrophobic cleft [111]. Furthermore, alanine-scanning revealed three hot spots on the helix [294]. Notably, this binding site suits most of the criteria in the decision tree proposed by Chene [14]. Initially, PPIMs binding to HDM2, e.g., nutlins [87, 295, 296], benzodiazepinediones [272], and others (Fig. 2) [229, 297-299], were identified via HTS. Structure-based design and molecular modeling were then used for ligand screening and optimization [300-304] leading to *in vitro* activities down to $IC_{50} = 3$ nM [302]. Several computational techniques have helped designing and screening for ligands of HDM2 or MDM2. (I) MD simulation and computational alanine scanning could accurately predict the hot spots of the p53/MDM2 interaction already by the efficient post-processing of wild-type trajectories [183]. In the same study, also the change in binding affinity due to other covalent modifications, e.g., methylation of the hot spot tryptophane of p53, could be confirmed in good agreement with experimental data. Finally, the opening or widening of the binding pocket into a PPIM binding-competent conformation could be sampled by MD and detected computationally [34, 42]. (II) The molecular diversity of compound libraries (e.g., benzodiazepinediones) was maximized to optimize molecules for HTS and synthesis strategies [304, 305]. (III) Molecular docking [271, 306], also in combination with *de novo* design [299, 302, 306], was applied to predict binding modes and optimize the design of PPIMs. (IV) Virtual Screening [303, 307], QSAR [303], and receptor-based pharmacophore models using ensembles of receptor structures [308] were also applied. From this, the following general strategy emerges: If there is no binding-competent pocket in the *apo* or protein-bound structure of a PPI target, such structures can potentially be found if multiple receptor conformations from NMR ensembles or crystallography are available [46] or if an ensemble can be generated by molecular simulations, e.g., MD simulation, preferentially in solvent less polar than water [42], or constrained geometric simulation [35, 42] (Fig. 1c). Furthermore, post-processing schemes including complex relaxation and rescoring have been demonstrated to improve the ranking and identification of native like binding poses [46]. In summary, many computational methods used in conventional computer-assisted drug design could be applied successfully to HDM2, also as a consequence of the deep binding cleft that is already preformed in the unbound HDM2 structure.

INTERLEUKIN-2

IL-2 is a key cytokine involved in the regulation of the immune system with relevance for immunological diseases, transplant medicine, and cancer [309]. Binding of the α -helical IL-2 to the trimeric IL-2 receptor is initiated by the association of IL-2 to the extracellular domain of the receptor's α subunit (IL-2R α). The IL-2/IL-2R α complex has been the subject of extensive studies that provided crystallographic structures and thermodynamic characterization of the protein-protein complex and five IL-2/PPIM complexes (Fig. 2), rendering this system a perfect test case for structure-based computational methods on PPIM design. For the binding of a PPIM, a pocket in the flat but flexible interface of unbound (or receptor-bound) IL-2 has to open (Fig. 1a-b). The absence of such a pocket is a major obstacle for structure-based design if based solely on the unbound or receptor-bound structure of IL-2. Additionally, it is difficult to decide which part of the large IL-2/IL-2R α interface (~ 2500 Å²) to address with a small molecule. Thus, it is not surprising that the first known PPIMs binding to IL-2 were not found by

structure-based design but rather by high-throughput screening. Later, IL-2 PPIMs were designed using structural knowledge obtained by tethering experiments and/or fragment-based ligand design [163]. This resulted in PPIMs with affinities down to the nanomolar range [289].

As to the question of transient pockets, we were able to show that conformational sampling of the unbound IL-2 structure by a constrained geometric simulation method resulted in the opening of such pockets, whereas MD simulations in explicit solvent failed in doing so, probably due to the pockets being rather hydrophobic [35]. Not using any knowledge about known IL-2 PPIMs, we were then able to identify these pockets from the ensemble structures based on geometric criteria as provided by PocketAnalyzer^{PCA} [152]. Notably, molecular docking into these pockets closely reproduced the bound state of the known IL-2 PPIMs as could these PPIMs be successfully ranked by MM-PB(GB)SA calculations and enriched in a large set of decoys. As to the question of the hot region, a narrow cluster of hot spots was predicted [35] by MM-PB(GB)SA effective binding free energy decomposition [199] starting from the IL-2/IL-2R α complex. Performing such decomposition for IL-2/PPIM complexes showed that essentially the same hot spots are also used for PPIM binding, pointing to mimicry of the PPI by the small molecules. Notably, two recent studies applying the energy based pocket detection algorithm Q-SiteFinder [93] and the solvent mapping algorithms FTMAP [221] showed that the druggable site could also be identified in the unbound structure of IL-2. Still, the need to account for interface flexibility persists, because *apo*-docking into the unbound structure of IL-2 was unsuccessful. Being able to accurately predict transient pockets from an unbound structure, and hot spot positions and binding energetics from complex structures, strongly suggests that the strategy and methods used here (Fig. 1c) will also be applicable in a prospective manner where nothing else than a protein-protein complex structure is known. Hence, this approach can well be the first step in a structure-based endeavor to identify PPIMs.

NHR2

NHR2 (nervy homology region 2) is the α -helical oligomerization domain of the RUNX1-ETO fusion protein present in approximately 12% of all acute myeloid leukemia (AML) [310]. The formation of NHR2 homotetramers from dimers has been shown to be essential for the leukemogenic activity of RUNX1-ETO [311]. In a prospective study, spatially clustered hot spots in the tetramer interface were predicted by MM-GBSA free energy decomposition (Fig. 1d) and were subsequently validated by *in vitro* and *in vivo* experiments [311]. The results reveal that alanine mutants of the hot spots prevent tetramerization of NHR2 and abolish AML formation in a mouse transplant model, thereby validating NHR2 as a promising target. As for the druggability of this PPI, a shallow elongated cavity was detected next to the hot spots (Fig. 1e). The anti-parallel orientation of helices C1 and C2 in the NHR2 dimer places D533, E536, and W540 (Fig. 1f) in close proximity to residues W498 and W502 (Fig. 1g), which results in a spatially compact arrangement of the hot spot residues. Furthermore, these residues are not located in the center of the interface, which is rather flat, but at its edges. These findings provided the incentive to develop a short peptide derived from the wild-type NHR2 sequence as an initial NHR2 tetramerization inhibitor (unpublished results). Based on this proof-of-principle, a virtual screening for small molecules was performed on the ZINC database [312] exploiting the knowledge about the predicted and validated hot spots. Encouragingly, some of the top-ranking small molecules from this screening exhibit *in vitro* PPIM activity in NHR2 tetramerization assays (unpublished results).

In summary, it was possible to (I) identify hot spots of the tetramerization of NHR2 that could be confirmed experimentally. These hot spots were (II) transferred to a peptide that is currently further optimized. After identifying a potent peptidic PPI modulator

(III) virtual screening for molecules exhibiting an arrangement of pharmacophoric groups as found in the peptide was carried out. These results lead us to conclude that the mutual integration of experimental and computational techniques is a promising approach to cope with the challenges of protein-protein interfaces in PPIM identification and design.

CONCLUSIONS AND PERSPECTIVES

Targeting protein-protein interfaces is currently a topic of outstanding interest in drug discovery. Since these targets offer great opportunities to interfere with PPI networks and, hence, for new therapeutics considerable effort has been undertaken for the development of PPIMs. As a result, the detailed characterization of many PPIs brought us remarkably closer towards an understanding of PPIs and their druggability [94]. While many PPIMs have been discovered by HTS, the structural insight into PPIs from experimentally determined protein-protein complexes and the experimental and computational methods for the identification of clustered hot spots and binding pockets has accelerated the rational design of PPIMs. Indeed, there are already a few examples of marketed small-molecule drugs acting on PPIs [3, 82-86, 225]. However, PPIs are different from classical targets in that binding pockets are often less pronounced, and hot spots are not in all cases arranged in a manner that they can easily be addressed by a small molecule. As pointed out by Morelli *et al.* along with the observation that there is not yet a unified approach for PPIM discovery, it appears that any such attempt has to be tailored for a specific PPI [142]. Nevertheless, the wealth of reported PPIMs shows that many PPIs are at least ligandable [94] (Fig. 2). With respect to the druggability of PPIs, it has to be mentioned that many of the so far developed PPIMs address PPIs that are predisposed by having preformed pockets and clustered hot spots and, accordingly, are more druggable than other PPIs. Also, it has to be mentioned that many of the reported PPIMs are not drug-like in the sense of Lipinski's rules, leaving considerable space for improvement and optimization to achieve the desired specificity and ADME properties. With the increasing number of known PPIMs, it is becoming clear that their chemical space is not identical to that of the majority of marketed drugs [313]. In fact, many of the PPIMs with pharmacological and clinical relevance do not exhibit the characteristics classically considered to be preferable for a drug-like molecule. Consequently, most currently available compound libraries, predominantly comprised of molecules with characteristics appropriate for classical targets, are not ideal for the identification of PPIMs, and methods for tailoring libraries for PPIM identification are being developed [314].

Note, however, that the trend in recently approved new molecular entities shows that the traditional criteria for drug-like properties, though desirable, are not a strict criterion for exclusion [313]. In addition, there are several examples for the optimization of non-drug-like molecules [315] and novel drug delivery approaches [316] with which some of the barriers for non-drug-like PPIMs may be overcome.

Here, we reviewed methods applicable to PPI druggability prediction and provided case studies of their successful application on PPI targets and for PPIM development. Many of these methods originate in druggability prediction of classical targets and have been adapted for application to PPIs where it is crucial to identify less pronounced potential binding sites to discern well druggable targets. Furthermore, methods for the detection of hot spots, based on the structure of the protein-protein complex, the unbound protein, the sequence, or a combination thereof, enable the identification of regions in which a small-molecule PPIM can efficiently bind. Additionally, experimental evidence shows that the flexibility of protein surfaces and protein-protein interfaces enables the opening of druggable pockets. Such pockets cannot be easily identified in the absence of a bound PPIM and, consequently, require an ade-

quate treatment of the protein's flexibility, e.g., by molecular simulation methods.

We think that any attempt to identify or optimize PPIMs can greatly benefit from integrating computational and experimental methods of pocket and hot spot detection, screening, and rational design. However, even though the success of several such attempts has been reported, it is hard to decide which computational methods will work best for a specific PPI because many of the presented methods have only been applied to one or a few targets. Furthermore, the performance of general strategies for the prediction of pockets and hot spots is hardly comparable, for two reasons. First, the datasets used to validate many methods vary considerably, often as a consequence of the prerequisites each individual method has. Second, the definition of pockets and hot spots often varies, thus complicating a statistical comparison of the prediction performance. To overcome this situation there is a high demand for common benchmarking datasets and a comparative database with experimental data as well as predictions from the various methods for enabling a comparison amongst subsets of known targets and to extrapolate to new ones. Furthermore, adapting the content of (virtual) screening libraries in order to cover the chemical space of PPIMs [142, 314], e.g., by including large but preorganized scaffolds containing hydrophobic/aromatic groups as often found in PPIMs and privileged scaffolds such as peptidomimetics, will facilitate the identification of new PPIMs. In fact, the amount of available data on PPIs is still very low in comparison to classical targets. However, with the expected progress in experimentally determined PPI structures, targets, and affinity data thereof and of PPIMs it will eventually be possible to compare PPI targets, transfer successful strategies, and exploit the potential of modulating PPIs to its full extent.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

We acknowledge a fruitful collaboration with the group of Manuel Grez at Georg-Speyer-Haus, Frankfurt, on modulating NHR2 tetramerization. We also thank D. Grimme and D. Krüger for critically reading the manuscript. We are grateful for financial support by the "Strategischer Forschungs-Fonds" and computational support by the "Zentrum für Informations- und Medientechnologie" (ZIM) at the Heinrich Heine University.

ABBREVIATIONS

PPI	=	Protein-protein interaction
PPIM	=	Small-molecule protein-protein interaction modulator
RMSD	=	Root mean-square deviation
MD	=	Molecular dynamics
IL-2	=	Interleukin-2
IL-2R α	=	α -subunit of the interleukin-2 receptor
GB	=	Generalized Born
PB	=	Poisson-Boltzmann
MM-PBSA	=	Molecular mechanics Poisson-Boltzmann surface area
MM-GBSA	=	Molecular mechanics generalized Born surface area

REFERENCES

- [1] Zinzalla G, Thurston DE. Targeting protein-protein interactions for therapeutic intervention: a challenge for the future. *Future Med Chem* 2009; 1: 65-93.

Modulating Protein-Protein Interactions

- [2] Mullard A. Protein-protein interaction inhibitors get into the groove. *Nat Rev Drug Discov* 2012; 11: 173-5.
- [3] Fischer PM. Protein-protein Interactions in Drug Discovery. *Drug Design Reviews - Online* 2005; 2: 179-207.
- [4] Xenarios I, Eisenberg D. Protein interaction databases. *Curr Opin Biotechnol* 2001; 12: 334-9.
- [5] Archakov AI, Govorun VM, Dubanov AV, *et al.* Protein-protein interactions as a target for drugs in proteomics. *Proteomics* 2003; 3: 380-91.
- [6] Pagel P, Kovac S, Oesterheld M, *et al.* The MIPS mammalian protein-protein interaction database. *Bioinformatics* 2005; 21: 832-4.
- [7] Beumung T, Skrabanek L, Niv MY, Mukherjee P, Weinstein H. PDZBase: a protein-protein interaction database for PDZ-domains. *Bioinformatics* 2005; 21: 827-8.
- [8] Cory S, Adams JM. The Bcl2 family: regulators of the cellular life-or-death switch. *Nat Rev Cancer* 2002; 2: 647-56.
- [9] Villunger A, Scott C, Bouillet P, Strasser A. Essential role for the BH3-only protein Bim but redundant roles for Bax, Bcl-2, and Bcl-w in the control of granulocyte survival. *Blood* 2003; 101: 2393-400.
- [10] Blazer LL, Neubig RR. Small Molecule Protein-Protein Interaction Inhibitors as CNS Therapeutic Agents: Current Progress and Future Hurdles. *Neuropsychopharmacol* 2009; 34: 126-41.
- [11] Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 2007; 450: 1001-9.
- [12] Ryan DP, Matthews JM. Protein-protein interactions in human disease. *Curr Opin Struct Biol* 2005; 15: 441-6.
- [13] Gerrard JA, Hutton CA, Perugini MA. Inhibiting protein-protein interactions as an emerging paradigm for drug discovery. *Mini Rev Med Chem* 2007; 7: 151-7.
- [14] Chene P. Drugs targeting protein-protein interactions. *Chemmedchem* 2006; 1: 400-11.
- [15] Arkin MR, Whitty A. The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions. *Curr Opin Chem Biol* 2009; 13: 284-90.
- [16] Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Dtsch Chem Ges* 1894; 2985.
- [17] Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed* 2002; 41: 2645-76.
- [18] Bon RS, Waldmann H. Bioactivity-Guided Navigation of Chemical Space. *Acc Chem Res* 2010; 43: 1103-14.
- [19] Scheck M, Koch MA, Waldmann H. Synthesis of a dysidiolide-inspired compound library and discovery of acetylcholinesterase inhibitors based on protein structure similarity clustering (PSSC). *Tetrahedron* 2008; 64: 4792-802.
- [20] Leung CH, Chan DSH, Kwan MHT, *et al.* Structure-Based Repurposing of FDA-Approved Drugs as TNF-alpha Inhibitors. *Chemmedchem* 2011; 6: 765-8.
- [21] Moriaud F, Richard SB, Adcock SA, *et al.* Identify drug repurposing candidates by mining the Protein Data Bank. *Brief Bioinform* 2011; 12: 336-40.
- [22] Doppelt-Azeroual O, Delfaud F, Moriaud F, de Brevern AG. Fast and automated functional classification with MED-SuMo: An application on purine-binding proteins. *Protein Sci* 2010; 19: 847-67.
- [23] Cheng AC, Coleman RG, Smyth KT, *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* 2007; 25: 71-5.
- [24] Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. *Proteins-Structure Function and Genetics* 2002; 47: 334-43.
- [25] Smith RD, Hu L, Falkner JA, Benson ML, Nerothin JP, Carlson HA. Exploring protein-ligand recognition with Binding MOAD. *J Mol Graph Model* 2006; 24: 414-25.
- [26] Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol* 1999; 285: 2177-98.
- [27] Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci USA* 1996; 93: 13-20.
- [28] DeLano WL, Ultsch MH, de Vos AM, Wells JA. Convergent solutions to binding at a protein-protein interface. *Science* 2000; 287: 1279-83.
- [29] Arkin MR, Wells JA. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* 2004; 3: 301-17.
- [30] Deeds EJ, Ashenberg O, Gerardin J, Shakhnovich EI. Robust protein-protein interactions in crowded cellular environments. *P Natl Acad Sci USA* 2007; 104: 14952-7.
- [31] Keskin O, Gursoy A, Ma B, Nussinov R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem Rev* 2008; 108: 1225-44.
- [32] Nooren IMA, Thornton JM. Diversity of protein-protein interactions. *Embo J* 2003; 22: 3486-92.
- [33] Taverna DM, Goldstein RA. Why are proteins marginally stable? *Proteins* 2002; 46: 105-9.
- [34] Eyrich S, Helms V. Transient pockets on protein surfaces involved in protein-protein interaction. *J Med Chem* 2007; 50: 3457-64.
- [35] Metz A, Pfeiffer C, Kopitz H, Pfeiffer-Marek S, Baringhaus KH, Gohlke H. Hot spots and transient pockets: Predicting the determinants of small-molecule binding to a protein-protein interface. *J Chem Inf Model* 2012; 52: 120-33.
- [36] Perot S, Sperandio O, Miteva MA, Camproux AC, Villoutreix BO. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov Today* 2010; 15: 656-67.
- [37] Dong YD, Boyd BJ. Applications of X-ray scattering in pharmaceutical science. *Int J Pharm* 2011; 417: 101-11.
- [38] Svergun DI, Koch MHJ. Small-angle scattering studies of biological macromolecules in solution. *Rep Prog Phys* 2003; 66: 1735-82.
- [39] Shaw DE, Maragakis P, Lindorff-Larsen K, *et al.* Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* 2010; 330: 341-6.
- [40] Acoca S, Cui QZ, Shore GC, Purisima EO. Molecular dynamics study of small molecule inhibitors of the Bcl-2 family. *Proteins-Structure Function and Bioinformatics* 2011; 79: 2624-36.
- [41] Brown SP, Hajduk PJ. Effects of conformational dynamics on predicted protein druggability. *Chemmedchem* 2006; 1: 70-2.
- [42] Eyrich S, Helms V. What induces pocket openings on protein surface patches involved in protein-protein interactions? *J Comput Aided Mol Des* 2009; 23: 73-86.
- [43] Yang CY, Wang SM. Hydrophobic Binding Hot Spots of Bcl-xL Protein-Protein Interfaces by Cosolvent Molecular Dynamics Simulation. *Acs Med Chem Lett* 2011; 2: 280-4.
- [44] Cozzini P, Kellogg GE, Spyraakis F, *et al.* Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *J Med Chem* 2008; 51: 6237-55.
- [45] Ahmed A, Kazemi S, Gohlke H. Protein Flexibility and Mobility in Structure-Based Drug Design. *Frontiers in Drug Design and Discovery* 2007; 3: 455-76.
- [46] Isvoran A, Badel A, Craescu CT, Miron S, Miteva MA. Exploring NMR ensembles of calcium binding proteins: Perspectives to design inhibitors of protein-protein interactions. *Bmc Struct Biol* 2011; 11.
- [47] Li X, Keskin O, Ma B, Nussinov R, Liang J. Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. *J Mol Biol* 2004; 344: 781-95.
- [48] Ma B, Elkayam T, Wolfson H, Nussinov R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci USA* 2003; 100: 5772-7.
- [49] Ahmad S, Keskin O, Mizuguchi K, Sarai A, Nussinov R. CCRXP: exploring clusters of conserved residues in protein structures. *Nucleic Acids Res* 2010; 38: W398-W401.
- [50] Moreira IS, Fernandes PA, Ramos MJ. Hot spots-A review of the protein-protein interface determinant amino-acid residues. *Proteins-Structure Function and Bioinformatics* 2007; 68: 803-12.
- [51] Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J Mol Biol* 1998; 280: 1-9.
- [52] Bradshaw RT, Patel BH, Tate EW, Leatherbarrow RJ, Gould IR. Comparing experimental and computational alanine scanning techniques for probing a prototypical protein-protein interaction. *Protein Eng Des Sel* 2011; 24: 197-207.
- [53] Tuncbag N, Kar G, Keskin O, Gursoy A, Nussinov R. A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Brief Bioinform* 2009; 10: 217-32.
- [54] Clackson T, Wells JA. A Hot-Spot of Binding-Energy in a Hormone-Receptor Interface. *Science* 1995; 267: 383-6.

- [55] Keskin O, Ma B, Nussinov R. Hot regions in protein-protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol* 2005; 345: 1281-94.
- [56] Cukuroglu E, Gursoy A, Keskin O. Analysis of Hot Region Organization in Hub Proteins. *Ann Biomed Eng* 2010; 38: 2068-78.
- [57] Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* 2001; 17: 284-5.
- [58] Rajamani D, Thiel S, Vajda S, Camacho CJ. Anchor residues in protein-protein interactions. *Proc Natl Acad Sci USA* 2004; 101: 11287-92.
- [59] Yagurtcu ON, Erdemli SB, Nussinov R, Turkey M, Keskin O. Restricted mobility of conserved residues in protein-protein interfaces in molecular simulations. *Biophys J* 2008; 94: 3475-85.
- [60] Keskin O, Ma B, Rogale K, Gunasekaran K, Nussinov R. Protein-protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach. *Phys Biol* 2005; 2: S24-35.
- [61] Reichmann D, Cohen M, Abramovich R, *et al.* Binding hot spots in the TEM1-BLIP interface in light of its modular architecture. *J Mol Biol* 2007; 365: 663-79.
- [62] Reichmann D, Rahat O, Albeck S, Meged R, Dym O, Schreiber G. The modular architecture of protein-protein binding interfaces. *Proc Natl Acad Sci USA* 2005; 102: 57-62.
- [63] Schreiber G, Fersht AR. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J Mol Biol* 1995; 248: 478-86.
- [64] Shulman-Peleg A, Shatsky M, Nussinov R, Wolfson HJ. Spatial chemical conservation of hot spot interactions in protein-protein complexes. *Bmc Biol* 2007; 5: 43.
- [65] Fromer M, Linial M. Exposing the co-adaptive potential of protein-protein interfaces through computational sequence design. *Bioinformatics* 2010; 26: 2266-72.
- [66] Raschke TM. Water structure and interactions with protein surfaces. *Curr Opin Struct Biol* 2006; 16: 152-9.
- [67] Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. *J Mol Biol* 1993; 234: 946-50.
- [68] Rodier F, Bahadur RP, Chakrabarti P, Janin J. Hydration of protein-protein interfaces. *Proteins* 2005; 60: 36-45.
- [69] Tuncbag N, Gursoy A, Guney E, Nussinov R, Keskin O. Architectures and functional coverage of protein-protein interfaces. *J Mol Biol* 2008; 381: 785-802.
- [70] Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci* 1997; 6: 53-64.
- [71] Tsai CJ, Xu D, Nussinov R. Structural motifs at protein-protein interfaces: protein cores versus two-state and three-state model complexes. *Protein Sci* 1997; 6: 1793-805.
- [72] Young L, Jernigan RL, Covell DG. A role for surface hydrophobicity in protein-protein recognition. *Protein Sci* 1994; 3: 717-29.
- [73] Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL. Evaluation of energetic and dynamic coupling networks in a PDZ domain protein. *J Mol Biol* 2006; 364: 337-51.
- [74] James LC, Roversi P, Tawfik DS. Antibody multispecificity mediated by conformational diversity. *Science* 2003; 299: 1362-7.
- [75] Lindner AB, Eshhar Z, Tawfik DS. Conformational changes affect binding and catalysis by ester-hydrolysing antibodies. *J Mol Biol* 1999; 285: 421-30.
- [76] Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. *Protein Eng* 1999; 12: 713-20.
- [77] Norel R, Sheinerman F, Petrey D, Honig B. Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci* 2001; 10: 2147-61.
- [78] Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* 2000; 10: 153-9.
- [79] Sheinerman FB, Honig B. On the role of electrostatic interactions in the design of protein-protein interfaces. *J Mol Biol* 2002; 318: 161-77.
- [80] Xu D, Lin SL, Nussinov R. Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. *J Mol Biol* 1997; 265: 68-84.
- [81] Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* 1996; 257: 342-58.
- [82] Owellen RJ, Hartke CA, Dickerson RM, Hains FO. Inhibition of Tubulin-Microtubule Polymerization by Drugs of Vinca Alkaloid Class. *Cancer Res* 1976; 36: 1499-502.
- [83] Liu G. Small molecule antagonists of the LFA-1/ICAM-1 interaction as potential therapeutic agents. *Expert Opin Ther Pat* 2001; 11: 1383-93.
- [84] Sillerud LO, Larson RS. Design and structure of peptide and peptidomimetic antagonists of protein-protein interaction. *Curr Protein Pept Sci* 2005; 6: 151-69.
- [85] Topol EJ, Byzova TV, Plow EF. Platelet GPIIb-IIIa blockers. *Lancet* 1999; 353: 227-31.
- [86] Kuritzkes D, Kar S, Kirkpatrick P. Maraviroc. *Nat Rev Drug Discov* 2008; 7: 15-6.
- [87] Popowicz GM, Domling A, Holak TA. The Structure-Based Design of Mdm2/Mdmx-p53 Inhibitors Gets Serious. *Angew Chem Int Ed* 2011; 50: 2680-8.
- [88] Smith J, Stewart BJ, Glaysher S, *et al.* The effect of pentamidine on melanoma ex vivo. *Anticancer Drugs* 2010; 21: 181-5.
- [89] Cheng G, Saleh MN, Marcher C, *et al.* Eltrombopag for management of chronic immune thrombocytopenia (RAISE): a 6-month, randomised, phase 3 study. *Lancet* 2011; 377: 393-402.
- [90] Cheng G. Eltrombopag for the treatment of immune thrombocytopenia. *Expert Rev Hematol* 2011; 4: 261-9.
- [91] Domling A. Small molecular weight protein-protein interaction antagonists: an insurmountable challenge? *Curr Opin Chem Biol* 2008; 12: 281-91.
- [92] Tse C, Shoemaker AR, Adickes J, *et al.* ABT-263: A potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res* 2008; 68: 3421-8.
- [93] Fuller JC, Burgoyne NJ, Jackson RM. Predicting druggable binding sites at the protein-protein interface. *Drug Discov Today* 2009; 14: 155-61.
- [94] Surade S, Blundell TL. Structural biology and drug discovery of difficult targets: the limits of ligandability. *Chem Biol* 2012; 19: 42-50.
- [95] Wass MN, David A, Sternberg MJE. Challenges for the prediction of macromolecular interactions. *Curr Opin Struct Biol* 2011; 21: 382-90.
- [96] Fernandez-Recio J. Prediction of protein binding sites and hot spots. *Wiley Interdiscip Rev Comput Mol Sci* 2011; 1: 680-98.
- [97] Leis S, Schneider S, Zacharias M. In Silico Prediction of Binding Sites on Proteins. *Curr Med Chem* 2010; 17: 1550-62.
- [98] de Vries SJ, Bonvin AMJJ. How proteins get in touch: Interface prediction in the study of biomolecular complexes. *Curr Protein Pept Sci* 2008; 9: 394-406.
- [99] Ezkurdia L, Bartoli L, Fariselli P, Casadio R, Valencia A, Tress ML. Progress and challenges in predicting protein-protein interaction sites. *Brief Bioinform* 2009; 10: 233-46.
- [100] Ofrian Y, Rost B. Protein-protein interaction hotspots carved into sequences. *Plos Comput Biol* 2007; 3: 1169-76.
- [101] Egner U, Hillig RC. A structural biology view of target drugability. *Expert Opin Drug Discov* 2008; 3: 391-401.
- [102] Fauman EB, Rai BK, Huang ES. Structure-based druggability assessment--identifying suitable targets for small molecule therapeutics. *Curr Opin Chem Biol* 2011; 15: 463-8.
- [103] Hajduk PJ, Huth JR, Tse C. Predicting protein druggability. *Drug Discov Today* 2005; 10: 1675-82.
- [104] Hajduk PJ, Huth JR, Fesik SW. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* 2005; 48: 2518-25.
- [105] Bullock BN, Jochim AL, Arora PS. Assessing helical protein interfaces for inhibitor design. *J Am Chem Soc* 2011; 133: 14220-3.
- [106] Jochim AL, Arora PS. Assessment of helical interfaces in protein-protein interactions. *Mol Biosyst* 2009; 5: 924-6.
- [107] Gonzalez-Ruiz D, Gohlke H. Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr Med Chem* 2006; 13: 2607-25.
- [108] Jochim AL, Arora PS. Systematic Analysis of Helical Protein Interfaces Reveals Targets for Synthetic Inhibitors. *Acs Chem Biol* 2010; 5: 919-23.
- [109] Jones S, Thornton JM. Protein-Protein Interactions - a Review of Protein Dimer Structures. *Prog Biophys Mol Bio* 1995; 63: 31-65.
- [110] Berman HM, Westbrook J, Feng Z, *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000; 28: 235-42.

Modulating Protein-Protein Interactions

Current Pharmaceutical Design, 2012, Vol. 18, No. 30 4643

- [111] Kussie PH, Gorina S, Marechal V, *et al.* Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 1996; 274: 948-53.
- [112] Henrich S, Salo-Ahen OMH, Huang B, Rippmann F, Cruciani G, Wade RC. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J Mol Recognit* 2010; 23: 209-19.
- [113] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 2001; 46: 3-26.
- [114] Kiel C, Beltrao P, Serrano L. Analyzing protein interaction networks using structural information. *Annu Rev Biochem* 2008; 77: 415-41.
- [115] Levitt DG, Banaszak LJ. Pocket - a Computer-Graphics Method for Identifying and Displaying Protein Cavities and Their Surrounding Amino-Acids. *J Mol Graphics* 1992; 10: 229-34.
- [116] Hendlich M, Rippmann F, Barnickel G. LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model* 1997; 15: 359-+.
- [117] Huang BD, Schroeder M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *Bmc Struct Biol* 2006; 6.
- [118] Laskowski RA. Surfnet a Program for Visualizing Molecular Surfaces, Cavities, and Intermolecular Interactions. *J Mol Graphics* 1995; 13: 323-&.
- [119] Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci* 1998; 7: 1884-97.
- [120] Weisel M, Proschak E, Schneider G. PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem Cent J* 2007; 1.
- [121] Schmidke P, Le Guilloux V, Maupetit J, Tuffery P. fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res* 2010; 38: W582-W9.
- [122] Brady GP, Stouten PFW. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* 2000; 14: 383-401.
- [123] Goodford PJ. A Computational Procedure for Determining Energetically Favorable Binding-Sites on Biologically Important Macromolecules. *J Med Chem* 1985; 28: 849-57.
- [124] Ruppert J, Welch W, Jain AN. Automatic identification and representation of protein binding sites for molecular docking. *Protein Sci* 1997; 6: 524-33.
- [125] Laurie ATR, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* 2005; 21: 1908-16.
- [126] Morita M, Nakamura S, Shimizu K. Highly accurate method for ligand-binding site prediction in unbound state (apo) protein structures. *Proteins-Structure Function and Bioinformatics* 2008; 73: 468-79.
- [127] Molecular Discovery, Molecular Discovery Ltd., Perugia, Italy, Available from: http://www.moldiscovery.com/soft_grid.php.
- [128] Mattos C, Ringe D. Locating and characterizing binding sites on proteins. *Nat Biotechnol* 1996; 14: 595-9.
- [129] Clark M, Guarnieri F, Shkurko I, Wiseman J. Grand canonical Monte Carlo simulation of ligand-protein binding. *J Chem Inf Model* 2006; 46: 231-42.
- [130] An JH, Totrov M, Abagyan R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol Cell Proteomics* 2005; 4: 752-61.
- [131] Kuhn D, Weskamp N, Schmitt S, Hüllermeier E, Klebe G. From the similarity analysis of protein cavities to the functional classification of protein families using Cavbase. *J Mol Biol* 2006; 359: 1023-44.
- [132] Najmanovich R, Kurbatova N, Thornton J. Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics* 2008; 24: 1105-111.
- [133] Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* 2002; 18 Suppl 1: S71-7.
- [134] Glaser F, Pupko T, Paz I, *et al.* ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* 2003; 19: 163-4.
- [135] Brylinski M, Skolnick J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc Natl Acad Sci USA* 2008; 105: 129-34.
- [136] Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni HH, McCammon JA. Discovery of a novel binding trench in HIV integrase. *J Med Chem* 2004; 47: 1879-81.
- [137] Lei M, Zavodszky MI, Kuhn LA, Thorpe MF. Sampling protein conformations and pathways. *J Comput Chem* 2004; 25: 1133-48.
- [138] Zavodszky MI, Ming L, Thorpe MF, Day AR, Kuhn LA. Modeling correlated main-chain motions in proteins for flexible molecular recognition. *Proteins-Structure Function and Bioinformatics* 2004; 57: 243-61.
- [139] Wells SA, Menor S, Hespeneide B, Thorpe MF. Constrained geometric simulation of diffusive motion in proteins. *Phys Biol* 2005; 2: S127-S36.
- [140] Laurie ATR, Jackson RM. Methods for the prediction of protein-ligand binding sites for Structure-Based Drug Design and virtual ligand screening. *Curr Protein Pept Sci* 2006; 7: 395-406.
- [141] Bourgeois R, Basse MJ, Morelli X, Roche P. Atomic Analysis of Protein-Protein Interfaces with Known Inhibitors: The 2P2I Database. *Plos One* 2010; 5.
- [142] Morelli X, Bourgeois R, Roche P. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol* 2011; 15: 475-81.
- [143] Smith GR, Sternberg MJE, Bates PA. The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. *J Mol Biol* 2005; 347: 1077-101.
- [144] Allen KN, Bellamacina CR, Ding XC, *et al.* An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem* 1996; 100: 2605-11.
- [145] Miranker A, Karplus M. Functionality Maps of Binding-Sites: A Multiple Copy Simultaneous Search Method. *Proteins* 1991; 11: 29-34.
- [146] Stultz CM, Karplus M. MCSS functionality maps for a flexible protein. *Proteins* 1999; 37: 512-29.
- [147] Zeng J, Nheu T, Zorzet A, *et al.* Design of inhibitors of Ras-Raf interaction using a computational combinatorial algorithm. *Protein Eng* 2001; 14: 39-45.
- [148] Dennis S, Kortvelyesi T, Vajda S. Computational mapping identifies the binding sites of organic solvents on proteins. *Proc Natl Acad Sci USA* 2002; 99: 4290-5.
- [149] Landon MR, Amaro RE, Baron R, *et al.* Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. *Chem Biol Drug Des* 2008; 71: 106-16.
- [150] Ahmed A, Gohlke H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins-Structure Function and Bioinformatics* 2006; 63: 1038-51.
- [151] Ahmed A, Rippmann F, Barnickel G, Gohlke H. A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins. *J Chem Inf Model* 2011; 51: 1604-22.
- [152] Craig IR, Pfleger C, Gohlke H, Essex JW, Spiegel K. Pocket-Space Maps To Identify Novel Binding-Site Conformations in Proteins. *J Chem Inf Model* 2011; 51: 2666-79.
- [153] Zhou HX, Qin SB. Interaction-site prediction for protein complexes: a critical assessment. *Bioinformatics* 2007; 23: 2203-9.
- [154] Li Y, Cortes J, Simeon T. Enhancing systematic protein-protein docking methods using ray casting: Application to ATTRACT. *Proteins-Structure Function and Bioinformatics* 2011; 79: 3037-49.
- [155] Tan KP, Varadarajan R, Madhusudhan MS. DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic Acids Res* 2011; 39: W242-8.
- [156] Chakravarty S, Varadarajan R. Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure* 1999; 7: 723-32.
- [157] Ertekin A, Nussinov R, Haliloglu T. Association of putative concave protein-binding sites with the fluctuation behavior of residues. *Protein Sci* 2006; 15: 2265-77.
- [158] Ward RA. Using protein-ligand docking to assess the chemical tractability of inhibiting a protein target. *J Mol Model* 2010; 16: 1833-43.
- [159] DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol* 2002; 12: 14-20.

- [160] Weiss GA, Watanabe CK, Zhong A, Goddard A, Sidhu SS. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci USA* 2000; 97: 8950-4.
- [161] Sidhu SS, Fairbrother WJ, Deshayes K. Exploring protein-protein interactions with phage display. *ChemBiochem* 2003; 4: 14-25.
- [162] Katz C, Levy-Beladev L, Rotem-Bamberger S, Rito T, Rudiger SGD, Friedler A. Studying protein-protein interactions using peptide arrays. *Chem Soc Rev* 2011; 40: 2131-45.
- [163] Erlanson DA, Wells JA, Braisted AC. Tethering: Fragment-based drug discovery. *Annu Rev Biophys Biomol Struct* 2004; 33: 199-223.
- [164] Erlanson DA, Hansen SK. Making drugs on proteins: site-directed ligand discovery for fragment-based lead assembly. *Curr Opin Chem Biol* 2004; 8: 399-406.
- [165] Toth G, Mukhyala K, Wells JA. Computational approach to site-directed ligand discovery. *Proteins-Structure Function and Bioinformatics* 2007; 68: 551-60.
- [166] Erlanson DA, Braisted AC, Raphael DR, et al. Site-directed ligand discovery. *Proc Natl Acad Sci USA* 2000; 97: 9367-72.
- [167] Davies DR, Begley DW, Hartley RC, Staker BL, Stewart LJ. Predicting the Success of Fragment Screening by X-Ray Crystallography. In: Juo L, Ed. *Fragment-Based Drug Design: Tools, Practical Approaches, and Examples*, Elsevier Academic Press; 2011. pp. 91-114.
- [168] Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 1996; 274: 1531-4.
- [169] Oltsersdorf T, Elmore SW, Shoemaker AR, et al. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 2005; 435: 677-81.
- [170] Hajduk PJ, Mack JC, Olejniczak ET, Park C, Dandliker PJ, Beutler BA. SOS-NMR: A saturation transfer NMR-based method for determining the structures of protein-ligand complexes. *J Am Chem Soc* 2004; 126: 2390-8.
- [171] Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* 2007; 6: 211-9.
- [172] Buhman G, C OC, Zerbe B, et al. Analysis of Binding Site Hot Spots on the Surface of Ras GTPase. *J Mol Biol* 2011; 413: 773-89.
- [173] Bernini A, Venditti V, Spiga O, Niccolai N. Probing protein surface accessibility with solvent and paramagnetic molecules. *Prog Nucl Magn Reson Spectrosc* 2009; 54: 278-89.
- [174] Weigelt J, van Dongen M, Uppenberg J, Schultz J, Wikstrom M. Site-selective screening by NMR spectroscopy with labeled amino acid pairs. *J Am Chem Soc* 2002; 124: 2446-7.
- [175] Grimme D, Gonzalez-Ruiz D, Gohlke H. Computational strategies and challenges for targeting protein-protein interactions with small molecules. In: Luque FJ, Barril X, Ed. *Physico-chemical and Computational Approaches to Drug Discovery*. London, UK, Royal Society of Chemistry; 2012.
- [176] Berg T. Use of "tethering" for the identification of a small molecule that binds to a dynamic hot spot on the interleukin-2 surface. *ChemBiochem* 2004; 5: 1051-3.
- [177] Fischer TB, Arunachalam KV, Bailey D, et al. The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces. *Bioinformatics* 2003; 19: 1453-4.
- [178] Rohl C, Price Y, Fischer TB, Paczkowski M, Zette MF, Tsai J. Cataloging the relationships between proteins: a review of interaction databases. *Mol Biotechnol* 2006; 34: 69-93.
- [179] Bickerton GR, Higuero AP, Blundell TL. Comprehensive, atomic-level characterization of structurally characterized protein-protein interactions: the PICCOLO database. *Bmc Bioinformatics* 2011; 12.
- [180] Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *Plos Comput Biol* 2007; 3: 337-44.
- [181] Guney E, Tuncbag N, Keskin O, Gursoy A. HotSpring: database of computational hot spots in protein interfaces. *Nucleic Acids Res* 2008; 36: D662-D66.
- [182] Moreira IS, Fernandes PA, Ramos MJ. Computational Determination of the Relative Free Energy of Binding - Application to Alanine Scanning Mutagenesis. In: Sokalski WA, Ed. *Molecular Materials with Specific Interactions - Modeling and Design*, Springer; 2007. pp. 305-39.
- [183] Massova I, Kollman PA. Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. *J Am Chem Soc* 1999; 121: 8133-43.
- [184] Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res* 2005; 33: W382-8.
- [185] Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 2004; 32: W526-31.
- [186] Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci USA* 2002; 99: 14116-21.
- [187] Kortemme T, Kim DE, Baker D. Computational alanine scanning of protein-protein interfaces. *Sci STKE* 2004; 2004: pl2.
- [188] Kiel C, Serrano L, Herrmann C. A detailed thermodynamic analysis of Ras/effector complex interfaces. *J Mol Biol* 2004; 340: 1039-58.
- [189] Kiel C, Serrano L. The ubiquitin domain superfold: Structure-based sequence alignments and characterization of binding epitopes. *J Mol Biol* 2006; 355: 821-44.
- [190] Carbonell P, Nussinov R, del Sol A. Energetic determinants of protein binding specificity: Insights into protein interaction networks. *Proteomics* 2009; 9: 1744-53.
- [191] Ivanov I, Crepin T, Jamin M, Ruigrok RWH. Structure of the Dimerization Domain of the Rabies Virus Phosphoprotein. *J Virol* 2010; 84: 3707-10.
- [192] Donald JE, Zhu H, Litvinov RI, DeGrado WF, Bennett JS. Identification of Interacting Hot Spots in the beta 3 Integrin Stalk Using Comprehensive Interface Design. *J Biol Chem* 2010; 285: 38658-65.
- [193] Liu Q, Hoi SCH, Su CTT, et al. Structural analysis of the hot spots in the binding between H1N1 HA and the 2D1 antibody: do mutations of H1N1 from 1918 to 2009 affect much on this binding? *Bioinformatics* 2011; 27: 2529-36.
- [194] Perez C, Khafizov K, Forrest LR, Kramer R, Ziegler C. The role of trimerization in the osmoregulated betaine transporter BetP. *Embo Rep* 2011; 12: 804-10.
- [195] Kruger DM, Gohlke H. DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. *Nucleic Acids Res* 2010; 38: W480-6.
- [196] Tuncbag N, Keskin O, Gursoy A. HotPoint: hot spot prediction server for protein interfaces. *Nucleic Acids Res* 2010; 38: W402-6.
- [197] Keskin O, Bahar I, Badretidinov AY, Pitsyn OB, Jernigan RL. Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. *Protein Sci* 1998; 7: 2578-86.
- [198] Novotny J, Bruccoleri RE, Saul FA. On the Attribution of Binding-Energy in Antigen-Antibody Complexes Mcpc-603, D1.3, and Hyhel-5. *Biochemistry* 1989; 28: 4735-49.
- [199] Gohlke H, Kiel C, Case DA. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes. *J Mol Biol* 2003; 330: 891-913.
- [200] Gohlke H, Case DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comput Chem* 2004; 25: 238-50.
- [201] Homeyer N, Gohlke H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol Inform* 2012; 31: 114-22.
- [202] Huo S, Massova I, Kollman PA. Computational alanine scanning of the 1: 1 human growth hormone-receptor complex. *J Comput Chem* 2002; 23: 15-27.
- [203] Moreira IS, Fernandes PA, Ramos MJ. Computational alanine scanning mutagenesis - An improved methodological approach. *J Comput Chem* 2007; 28: 644-54.
- [204] Benedix A, Becker CM, de Groot BL, Caflisch A, Bockmann RA. Predicting free energy changes using structural ensembles. *Nat Methods* 2009; 6: 3-4.
- [205] deGroot BL, vanAalten DMF, Scheek RM, Amadei A, Vriend G, Berendsen HJC. Prediction of protein conformational freedom from distance constraints. *Proteins* 1997; 29: 240-51.
- [206] Aqvist J, Luzhkov VB, Brandsdal BO. Ligand binding affinities from MD simulations. *Acc Chem Res* 2002; 35: 358-65.
- [207] Almlof M, Aqvist J, Smalas AO, Brandsdal BO. Probing the effect of point mutations at protein-protein interfaces with free energy calculations. *Biophys J* 2006; 90: 433-42.

Modulating Protein-Protein Interactions

Current Pharmaceutical Design, 2012, Vol. 18, No. 30 4645

- [208] Chen RY, Chen WJ, Yang SX, *et al.* Rigorous assessment and integration of the sequence and structure based features to predict hot spots. *Bmc Bioinformatics* 2011; 12.
- [209] Cho KI, Kim D, Lee D. A feature-based approach to modeling protein-protein interaction hot spots. *Nucleic Acids Res* 2009; 37: 2672-87.
- [210] Kosloff M, Travis AM, Bosch DE, Siderovski DP, Arshavsky VY. Integrating energy calculations with functional assays to decipher the specificity of G protein-RGS protein interactions. *Nat Struct Mol Biol* 2011; 18: 846-U128.
- [211] Sheinerman FB, Al-Lazikani B, Honig B. Sequence, structure and energetic determinants of phosphopeptide selectivity of SH2 domains. *J Mol Biol* 2003; 334: 823-41.
- [212] Darnell SJ, Page D, Mitchell JC. An automated decision-tree approach to predicting protein interaction hot spots. *Proteins-Structure Function and Bioinformatics* 2007; 68: 813-23.
- [213] Zhu XL, Mitchell JC. KFC2: A knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins-Structure Function and Bioinformatics* 2011; 79: 2671-83.
- [214] Ofra Y, Rost B. ISIS: interaction sites identified from sequence. *Bioinformatics* 2007; 23: E13-6.
- [215] Landon MR, Lancia DR, Yu J, Thiel SC, Vajda S. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J Med Chem* 2007; 50: 1231-40.
- [216] Seco J, Luque FJ, Barril X. Binding Site Detection and Druggability Index from First Principles. *J Med Chem* 2009; 52: 2363-71.
- [217] Grosdidier S, Fernandez-Recio J. Docking and scoring: applications to drug discovery in the interactomics era. *Expert Opin Drug Discov* 2009; 4: 673-86.
- [218] Grosdidier S, Fernandez-Recio J. Identification of hot-spot residues in protein-protein interactions by computational docking. *Bmc Bioinformatics* 2008; 9: Art. No. 447.
- [219] Geppert T, Hoy B, Wessler S, Schneider G. Context-Based Identification of Protein-Protein Interfaces and "Hot-Spot" Residues. *Chem Biol* 2011; 18: 344-53.
- [220] Geppert T, Bauer S, Hiss JA, *et al.* Immunosuppressive Small Molecule Discovered by Structure-Based Virtual Screening for Inhibitors of Protein-Protein Interactions. *Angew Chem Int Ed Engl* 2011.
- [221] Kozakov D, Hall DR, Chuang GY, *et al.* Structural conservation of druggable hot spots in protein-protein interfaces. *Proc Natl Acad Sci USA* 2011; 108: 13528-33.
- [222] Higuero AP, Schreyer A, Bickerton GRJ, Pitt WR, Groom CR, Blundell TL. Atomic Interactions and Profile of Small Molecules Disrupting Protein-Protein Interfaces: the TIMBAL Database. *Chem Biol Drug Des* 2009; 74: 457-67.
- [223] Oldham RK, Dillman RO. Monoclonal antibodies in cancer therapy: 25 years of progress. *J Clin Oncol* 2008; 26: 1774-7.
- [224] Stockwin LH, Holmes S. Antibodies as therapeutic agents: vive la renaissance! *Exp Opin Biol Ther* 2003; 3: 1133-52.
- [225] Vogel C, Cobleigh MA, Tripathy D, *et al.* First-line, single-agent Herceptin (R) (trastuzumab) in metastatic breast cancer: a preliminary report. *Eur J Cancer* 2001; 37: S25-S9.
- [226] Lafont V, Schaefer M, Stote RH, Altschuh D, Dejaegere A. Protein-protein recognition and interaction hot spots in an antigen-antibody complex: Free energy decomposition identifies "efficient amino acids". *Proteins-Structure Function and Bioinformatics* 2007; 67: 418-34.
- [227] Griffin L, Lawson A. Antibody fragments as tools in crystallography. *Clin Exp Immunol* 2011; 165: 285-91.
- [228] Imai K, Takaoka A. Comparing antibody and small-molecule therapies for cancer. *Nat Rev Cancer* 2006; 6: 714-27.
- [229] Berg T. Small-molecule inhibitors of protein-protein interactions. *Curr Opin Drug Discov Devel* 2008; 11: 666-74.
- [230] Zhong SJ, Macias AT, MacKerell AD. Computational identification of inhibitors of protein-protein interactions. *Curr Top Med Chem* 2007; 7: 63-82.
- [231] Buchwald P. Small-Molecule Protein-Protein Interaction Inhibitors: Therapeutic Potential in Light of Molecular Size, Chemical Space, and Ligand Binding Efficiency Considerations. *Iubmb Life* 2010; 62: 724-31.
- [232] Sperandio O, Reynes CH, Camproux AC, Villoutreix BO. Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov Today* 2010; 15: 220-9.
- [233] Vassilev L, Fry D, editors. *Small-Molecule Inhibitors of Protein-Protein Interactions*. 1 ed. Berlin Heidelberg: Springer-Verlag; 2011.
- [234] Vogler M, Dinsdale D, Dyer MJS, Cohen GM. Bcl-2 inhibitors: small molecules with a big impact on cancer therapy. *Cell Death Differ* 2009; 16: 360-7.
- [235] Fujii N, Haresco JJ, Novak KAP, *et al.* Rational design of a nonpeptide general chemical scaffold for reversible inhibition of PDZ domain interactions. *Bioorg Med Chem Lett* 2007; 17: 549-52.
- [236] Hammond MC, Harris BZ, Lim WA, Bartlett PA. Beta strand peptidomimetics as potent PDZ domain ligands. *Chem Biol* 2006; 13: 1247-51.
- [237] Shan JF, Shi DL, Wang JM, Zheng J. Identification of a specific inhibitor of the dishevelled PDZ domain. *Biochemistry* 2005; 44: 15495-503.
- [238] Lee HJ, Wang NX, Shi DL, Zheng JJ. Sulindac Inhibits Canonical Wnt Signaling by Blocking the PDZ Domain of the Protein Dishevelled. *Angew Chem Int Edit* 2009; 48: 6448-52.
- [239] Herrmann C, Block C, Geisen C, *et al.* Sulindac sulfide inhibits Ras signaling. *Oncogene* 1998; 17: 1769-76.
- [240] Kato-Stankiewicz J, Hakimi I, Zhi G, *et al.* Inhibitors of Ras/Raf-1 interaction identified by two-hybrid screening revert Ras-dependent transformation phenotypes in human cancer cells. *P Natl Acad Sci USA* 2002; 99: 14398-403.
- [241] Lu YC, Sakamuri S, Chen QZ, *et al.* Solution phase parallel synthesis and evaluation of MAPK inhibitory activities of close structural analogues of a Ras pathway modulator. *Bioorg Med Chem Lett* 2004; 14: 3957-62.
- [242] Stebbins JL, De SK, Machleidt T, *et al.* Identification of a new JNK inhibitor targeting the JNK-JIP interaction site. *P Natl Acad Sci USA* 2008; 105: 16809-13.
- [243] Chen T, Kablaoui N, Little J, *et al.* Identification of small-molecule inhibitors of the JIP-JNK interaction. *Biochem J* 2009; 420: 283-94.
- [244] Dorri P, Westby M, Dobbs S, *et al.* Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob Agents Ch* 2005; 49: 4721-32.
- [245] Erickson-Miller CL, Delorme E, Tian SS, *et al.* Preclinical Activity of Eltrombopag (SB-497115), an Oral, Nonpeptide Thrombopoietin Receptor Agonist. *Stem Cells* 2009; 27: 424-30.
- [246] Jackson DY, Quan C, Artis DR, *et al.* Potent alpha 4 beta 1 peptide antagonists as potential anti-inflammatory agents. *J Med Chem* 1997; 40: 3359-68.
- [247] Carron CP, Meyer DM, Pegg JA, *et al.* A peptidomimetic antagonist of the integrin alpha(v)beta3 inhibits Leydig cell tumor growth and the development of hypercalcemia of malignancy. *Cancer Res* 1998; 58: 1930-5.
- [248] Viaud J, Zeghouf M, Barelli H, *et al.* Structure-based discovery of an inhibitor of Arf activation by Sec7 domains through targeting of protein-protein complexes. *P Natl Acad Sci USA* 2007; 104: 10370-5.
- [249] Corradi V, Mancini M, Santucci MA, *et al.* Computational techniques are valuable tools for the discovery of protein-protein interaction inhibitors: The 14-3-3 sigma case. *Bioorg Med Chem Lett* 2011; 21: 6867-71.
- [250] Oost TK, Sun CH, Armstrong RC, *et al.* Discovery of potent antagonists of the antiapoptotic protein XIAP for the treatment of cancer. *J Med Chem* 2004; 47: 4417-26.
- [251] Sun HY, Nikolovska-Coleska Z, Lu JF, *et al.* Design, synthesis, and characterization of a potent, nonpeptide, cell-permeable, bivalent smac mimetic that concurrently targets both the BIR2 and BIR3 domains in XIAP. *Journal of the American Chemical Society* 2007; 129: 15279-94.
- [252] Peng Y, Sun H, Nikolovska-Coleska Z, *et al.* Potent, Orally Bioavailable Diazabicyclic Small-Molecule Mimetics of Second Mitochondria-Derived Activator of Caspases. *J Med Chem* 2008; 51: 8158-62.
- [253] Cai Q, Sun HY, Peng YF, *et al.* A Potent and Orally Active Antagonist (SM-406/AT-406) of Multiple Inhibitor of Apoptosis Proteins (IAPs) in Clinical Development for Cancer Treatment. *J Med Chem* 2011; 54: 2714-26.

- [254] Zhang B, Corbel C, Gueritte F, Couturier C, Bach S, Tan VBC. An in silico approach for the discovery of CDK5/p25 interaction inhibitors. *Biotechnol J* 2011; 6: 871-81.
- [255] Wang HB, Hammoudeh DI, Follis AV, Reese BE, Lazo JS, Metallo SJ, *et al.* Improved low molecular weight Myc-Max inhibitors. *Mol Cancer Ther* 2007; 6: 2399-408.
- [256] Christ F, Voet A, Marchand A, *et al.* Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat Chem Biol* 2010; 6: 442-8.
- [257] Wang T, Zhang ZX, Wallace OB, *et al.* Discovery of 4-benzoyl-1-[(4-methoxy-1H-pyrrolo[2,3-b]pyridin-3-yl)oxoacetyl]-2-(R)-methylpiperazine (BMS-378806): A novel HIV-1 attachment inhibitor that interferes with CD4-gp120 interactions. *J Med Chem* 2003; 46: 4236-9.
- [258] Bonacci TM, Mathews JL, Yuan CJ, *et al.* Differential targeting of G beta gamma-subunit signaling with small molecules. *Science* 2006; 312: 443-6.
- [259] Yap JL, Worlikar S, MacKerell AD, Shapiro P, Fletcher S. Small-Molecule Inhibitors of the ERK Signaling Pathway: Towards Novel Anticancer Therapeutics. *Chemmedchem* 2011; 6: 38-48.
- [260] Chen FM, Hancock CN, Macias AT, *et al.* Characterization of ATP-independent ERK inhibitors identified through in silico analysis of the active ERK2 structure. *Bioorg Med Chem Lett* 2006; 16: 6281-7.
- [261] Yoakim C, Ogilvie WW, Goudreau N, *et al.* Discovery of the first series of inhibitors of human papillomavirus type 11: Inhibition of the assembly of the E1-E2-origin DNA complex. *Bioorg Med Chem Lett* 2003; 13: 2539-41.
- [262] White PW, Titolo S, Brault K, *et al.* Inhibition of human papillomavirus DNA replication by small molecule antagonists of the E1-E2 protein interaction. *J Biol Chem* 2003; 278: 26765-72.
- [263] Cerchietti LC, Ghetu AF, Zhu X, *et al.* A Small-Molecule Inhibitor of BCL6 Kills DLBCL Cells *In vitro* and *In vivo*. *Cancer Cell* 2010; 17: 400-11.
- [264] Rush TS, Grant JA, Mosyak L, Nicholls A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J Med Chem* 2005; 48: 1489-95.
- [265] Huang N, Nagarsekar A, Xia GJ, Hayashi J, MacKerell AD. Identification of non-phosphate-containing small molecular weight inhibitors of the tyrosine kinase p56 Lck SH2 domain via in silico screening against the pY+3 binding site. *J Med Chem* 2004; 47: 3502-11.
- [266] Capps KJ, Humiston J, Dominique R, Hwang I, Boger DL. Discovery of AICAR Tfase inhibitors that disrupt requisite enzyme dimerization. *Bioorg Med Chem Lett* 2005; 15: 2840-4.
- [267] He MM, Smith AS, Oslob JD, *et al.* Small-molecule inhibition of TNF-alpha. *Science* 2005; 310: 1022-5.
- [268] Paige JS, Jaffrey SR. Pharmacologic manipulation of nitric oxide signaling: Targeting NOS dimerization and protein-protein interactions. *Curr Top Med Chem* 2007; 7: 97-114.
- [269] Gorczynski MJ, Grembecka J, Zhou YP, *et al.* Allosteric inhibition of the protein-protein interaction between the leukemia-associated proteins Runx1 and CBF beta. *Chem Biol* 2007; 14: 1186-97.
- [270] Reindl WG, Yuan JP, Kramer A, Strebhardt K, Berg T. Inhibition of polo-like kinase 1 by blocking Polo-box domain-dependent protein-protein interactions. *Chem Biol* 2008; 15: 459-66.
- [271] Shangary S, Qin DG, McEachern D, *et al.* Temporal activation of p53 by a specific MDM2 inhibitor is selectively toxic to tumors and leads to complete tumor growth inhibition. *Proc Natl Acad Sci USA* 2008; 105: 3933-8.
- [272] Grasberger BL, Lu TB, Schubert C, *et al.* Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *J Med Chem* 2005; 48: 909-12.
- [273] Popowicz GM, Czarna A, Wolf S, *et al.* Structures of low molecular weight inhibitors bound to MDMX and MDM2 reveal new approaches for p53-MDMX/MDM2 antagonist drug discovery. *Cell Cycle* 2010; 9: 1104-11.
- [274] Shaginian A, Whitby LR, Hong S, *et al.* Design, Synthesis, and Evaluation of an alpha-Helix Mimetic Library Targeting Protein-Protein Interactions. *J Am Chem Soc* 2009; 131: 5564-72.
- [275] Markowitz J, Chen J, Gitti R, *et al.* Identification and characterization of small molecule inhibitors of the calcium-dependent S100B-p53 tumor suppressor interaction. *J Med Chem* 2004; 47: 5085-93.
- [276] Wilder PT, Charpentier TH, Liriano MA, *et al.* *In vitro* screening and structural characterization of inhibitors of the S100B-p53 interaction. *Int J High Throughput Screen* 2010; 2010: 109-26.
- [277] Yin H, Lee GI, Sedey KA, *et al.* Terphenyl-based bak BH3 alpha-helical proteomimetics as low-molecular-weight antagonists of Bcl-XL. *J Am Chem Soc* 2005; 127: 10191-6.
- [278] Rodriguez JM, Hamilton AD. Benzoylurea oligomers: Synthetic foldamers that mimic extended alpha helices. *Angew Chem Int Ed* 2007; 46: 8614-7.
- [279] Zhai D, Jin C, Satterthwait AC, Reed JC. Comparison of chemical inhibitors of antiapoptotic Bcl-2-family proteins. *Cell Death Differ* 2006; 13: 1419-21.
- [280] Tang GZ, Yang CY, Nikolovska-Coleska Z, *et al.* Pyrogallol-based molecules as potent inhibitors of the antiapoptotic Bcl-2 proteins. *J Med Chem* 2007; 50: 1723-6.
- [281] Wang GP, Nikolovska-Coleska Z, Yang CY, *et al.* Structure-based design of potent small-molecule inhibitors of anti-apoptotic Bcl-2 proteins. *J Med Chem* 2006; 49: 6139-42.
- [282] Degterev A, Lugovskoy A, Cardone M, *et al.* Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. *Nat Cell Biol* 2001; 3: 173-82.
- [283] Chan SL, Lee MC, Tan KO, *et al.* Identification of chelerythrine as an inhibitor of BclXL function. *J Biol Chem* 2003; 278: 20453-6.
- [284] Kitada S, Leone M, Sareth S, Zhai D, Reed JC, Pellicchia M. Discovery, characterization, and structure-activity relationships studies of proapoptotic polyphenols targeting B-cell lymphocyte/leukemia-2 proteins. *J Med Chem* 2003; 46: 4259-64.
- [285] Becattini B, Kitada S, Leone M, *et al.* Rational design and real time, in-cell detection of the proapoptotic activity of a novel compound targeting Bcl-XL. *Chem Biol* 2004; 11: 389-95.
- [286] Arkin MR, Randal M, DeLano WL, *et al.* Binding of small molecules to an adaptive protein-protein interface. *Proc Natl Acad Sci USA* 2003; 100: 1603-8.
- [287] Thanos CD, Randal M, Wells JA. Potent small-molecule binding to a dynamic hot spot on IL-2. *J Am Chem Soc* 2003; 125: 15280-1.
- [288] Raimundo BC, Oslob JD, Braisted AC, *et al.* Integrating fragment assembly and biophysical methods in the chemical advancement of small-molecule antagonists of IL-2: An approach for inhibiting protein-protein interactions. *J Med Chem* 2004; 47: 3111-30.
- [289] Braisted AC, Oslob JD, Delano WL, *et al.* Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J Am Chem Soc* 2003; 125: 3714-5.
- [290] Thanos CD, DeLano WL, Wells JA. Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc Natl Acad Sci USA* 2006; 103: 15422-7.
- [291] Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. *Nature* 2000; 408: 307-10.
- [292] Suzuki K, Matsubara H. Recent Advances in p53 Research and Cancer Treatment. *J Biomed Biotechnol* 2011; Art. No. 978312.
- [293] Lane DP, Cheok CF, Lain S. p53-based Cancer Therapy. *Cold Spring Harb Perspect Biol* 2010; 2: a001222.
- [294] Picksley SM, Vojtesek B, Sparks A, Lane DP. Immunochemical Analysis of the Interaction of P53 with Mdm2 - Fine Mapping of the Mdm2 Binding-Site on P53 Using Synthetic Peptides. *Oncogene* 1994; 9: 2523-9.
- [295] Fry DC, Emerson SD, Palme S, Vu BT, Liu CM, Podlaski F. NMR structure of a complex between MDM2 and a small molecule inhibitor. *J Biomol Nmr* 2004; 30: 163-73.
- [296] Vassilev LT, Vu BT, Graves B, *et al.* *In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 2004; 303: 844-8.
- [297] Khoury K, Popowicz GM, Holak TA, Domling A. The p53-MDM2/MDMX axis - A chemotype perspective. *Medchemcomm* 2011; 2: 246-60.
- [298] Galatin PS, Abraham DJ. A nonpeptidic sulfonamide inhibits the p53-mdm2 interaction and activates p53-dependent transcription in mdm2-overexpressing cells. *J Med Chem* 2004; 47: 4163-5.
- [299] Hardcastle IR, Ahmed SU, Atkins H, *et al.* Isoindolinone-based inhibitors of the MDM2-p53 protein-protein interaction. *Bioorg Med Chem Lett* 2005; 15: 1515-20.
- [300] Zhuang C, Miao Z, Zhu L, *et al.* Synthesis and biological evaluation of thio-benzodiazepines as novel small molecule inhibitors of the p53-MDM2 protein-protein interaction. *Eur J Med Chem* 2011; 46: 5654.

Modulating Protein-Protein Interactions

- [301] Hu CQ, Li X, Wang WS, *et al.* Design, synthesis, and biological evaluation of imidazoline derivatives as p53-MDM2 binding inhibitors. *Bioorg Med Chem* 2011; 19: 5454-61.
- [302] Ding K, Lu YP, Nikolovska-Coleska Z, *et al.* Structure-based design of spiro-oxindoles as potent, specific small-molecule inhibitors of the MDM2-p53 interaction. *J Med Chem* 2006; 49: 3432-5.
- [303] Galatin PS, Abraham DJ. QSAR: Hydropathic analysis of inhibitors of the p53-mdm2 interaction. *Proteins* 2001; 45: 169-75.
- [304] Agrafiotis DK. Stochastic algorithms for maximizing molecular diversity. *J Chem Inf Comp Sci* 1997; 37: 841-51.
- [305] Kalinski C, Umkehrer M, Weber L, Kolb J, Burdack C, Ross G. On the industrial applications of MCRs: molecular diversity in drug discovery and generic drug synthesis. *Mol Divers* 2010; 14: 513-22.
- [306] Ding K, Lu Y, Nikolovska-Coleska Z, Qiu S, Ding YS, Gao W, *et al.* Structure-based design of potent non-peptide MDM2 inhibitors. *J Am Chem Soc* 2005; 127: 10130-1.
- [307] Rothweiler U, Czarna A, Krajewski M, *et al.* Isoquinolin-1-one inhibitors of the MDM2-p53 interaction. *Chemmedchem* 2008; 3: 1118-28.
- [308] Bowman AL, Nikolovska-Coleska Z, Zhong HZ, Wang SM, Carlson HA. Small molecule inhibitors of the MDM2-p53

Current Pharmaceutical Design, 2012, Vol. 18, No. 30 4647

- interaction discovered by ensemble-based receptor models. *Journal of the American Chemical Society* 2007; 129: 12809-14.
- [309] Malek TR. The biology of interleukin-2. *Annu Rev Immunol* 2008; 26: 453-79.
- [310] Langabeer SE, Walker H, Rogers JR, *et al.* Incidence of AML1/ETO fusion transcripts in patients entered into the MRC AML trials. *Brit J Haematol* 1997; 99: 925-8.
- [311] Wichmann C, Becker Y, Chen-Wichmann L, *et al.* Dimer-tetramer transition controls RUNX1/ETO leukemogenic activity. *Blood* 2010; 116: 603-13.
- [312] Irwin JJ, Shoichet BK. ZINC - A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005; 45: 177-82.
- [313] Faller B, Ottaviani G, Ertl P, Berellini G, Collis A. Evolution of the physicochemical properties of marketed drugs: can history foretell the future? *Drug Discov Today* 2011; 16: 976-84.
- [314] Reynes C, Host H, Camproux AC, *et al.* Designing Focused Chemical Libraries Enriched in Protein-Protein Interaction Inhibitors using Machine-Learning Methods. *Plos Comput Biol* 2010; 6.
- [315] Zhao HY. Lead optimization in the nondrug-like space. *Drug Discov Today* 2011; 16: 158-63.
- [316] Allen TM, Cullis PR. Drug delivery systems: Entering the mainstream. *Science* 2004; 303: 1818-22.

Received: April 1, 2012

Accepted: April 11, 2012

Publication II

Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein-Protein Interface

Alexander Metz,[§] Christopher Pflieger,[§] Hannes Kopitz, Stefania Pfeiffer-Marek, Karl-Heinz Baringhaus, and Holger Gohlke

Journal of Chemical Information and Modeling, **2012**; 52: 120-133.

[§]Both authors have contributed equally to the respective work.

Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein–Protein Interface

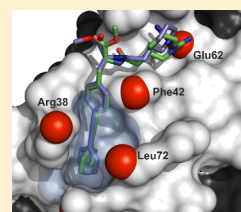
Alexander Metz,^{†,§} Christopher Pfeiffer,^{†,§} Hannes Kopitz,[†] Stefania Pfeiffer-Marek,[†] Karl-Heinz Baringhaus,[†] and Holger Gohlke^{*,†}

[†]Institute for Pharmaceutical and Medicinal Chemistry, Department of Mathematics and Natural Sciences, Heinrich-Heine-University, Düsseldorf, Germany

[‡]Sanofi-Aventis Deutschland GmbH, LGCR Drug Design, Frankfurt am Main, Germany

[§] Supporting Information

ABSTRACT: Protein–protein interfaces are considered difficult targets for small-molecule protein–protein interaction modulators (PPIMs). Here, we present for the first time a computational strategy that simultaneously considers aspects of energetics and plasticity in the context of PPIM binding to a protein interface. The strategy aims at identifying the determinants of small-molecule binding, hot spots, and transient pockets, in a protein–protein interface in order to make use of this knowledge for predicting binding modes of and ranking PPIMs with respect to their affinity. When applied to interleukin-2 (IL-2), the computationally inexpensive constrained geometric simulation method FRODA outperforms molecular dynamics simulations in sampling hydrophobic transient pockets. We introduce the PPIAnalyzer approach for identifying transient pockets on the basis of geometrical criteria only. A sequence of docking to identified transient pockets, starting structure selection based on hot spot information, RMSD clustering and intermolecular docking energies, and MM-PBSA calculations allows one to enrich IL-2 PPIMs from a set of decoys and to discriminate between subgroups of IL-2 PPIMs with low and high affinity. Our strategy will be applicable in a prospective manner where nothing else than a protein–protein complex structure is known; hence, it can well be the first step in a structure-based endeavor to identify PPIMs.



INTRODUCTION

Protein–protein interactions (PPIs) are involved in nearly all biological processes. Due to their universal occurrence, protein–protein interfaces provide an important, yet neglected, new class of drug targets.¹ At present, the design of small-molecule protein–protein interaction modulators (PPIMs) encounters at least two challenges. First, in contrast to enzymes, protein–protein interfaces are rather flat and usually lack a distinct binding pocket.² Second, due to the often large size of protein–protein interfaces (~ 1200 to ~ 4660 Å²),^{3,4} interactions that are favorable for binding can be widely distributed over the interface.

Experimental evidence suggests that these challenges can be overcome.^{5–8} Most strikingly, residues participating in important interactions have been shown to be spatially clustered in protein–protein interfaces, forming so-called “hot spot” regions.^{4,9–11} Mimicking localized interactions at these hot spots provides a possibility for PPIM development.^{1,4,12–14} Furthermore, an opening of so-called transient pockets was observed in protein–protein interfaces.⁵ In fact, binding of several PPIMs to transient pockets in protein–protein interfaces has been reported.⁷ The most prominent example is given by small-molecule inhibitors binding to interleukin-2 (IL-2). These PPIMs inhibit the interaction with the IL-2 α -receptor (IL-2R α).^{5,15,16} Notably, the small molecule-bound IL-2 exhibits pockets in the interface region that are present neither in the unbound nor in the IL-2R α -bound crystal structure.^{5,17}

Computational methods can aid in finding and in the design of PPIMs if they are able to provide an accurate description of the energetics and dynamics of small-molecule binding to protein–protein interfaces.^{7,13,18–21} While many computational studies have been reported that deal with the energetics^{13,22–33} and dynamics^{7,13,19,34–39} of protein–protein interfaces *per se*, only a few have focused on the aspect of small-molecule binding to protein–protein interface regions,^{7,14,20,40–46} and none has considered aspects of energetics and interface plasticity simultaneously in this context so far. Thus, in the present study, we set out to evaluate the capability of state-of-the-art computational methods to predict small-molecule binding to protein–protein interfaces. In particular, we address five major questions that consider aspects of structure, dynamics, and energetics important for binding to protein–protein interfaces: (i) Can one identify hot spots in a protein–protein interface based on a protein–protein complex structure and make use of these hot spots for predicting the binding mode of small-molecule ligands? (ii) Can one sample the opening of transient pockets comparable to those observed in the protein–protein interface of a bound protein, as suggested by the conformational selection model,⁴⁷ starting from an unbound protein conformation? (iii) Is it possible to identify protein conformations with transient pockets by energetic or

Received: July 13, 2011

Published: November 17, 2011

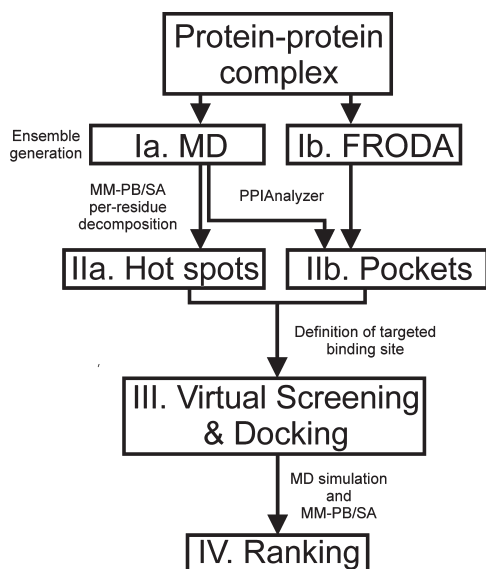


Figure 1. Outline of the strategy for hot spot and transient pocket identification, docking, and ranking of PPIMs using only the structure of a protein–protein complex as a starting point. A similarity-based virtual screening was not performed in this study. See text for details.

geometrical criteria? (iv) By docking to transient pockets, can one reproduce binding modes of known PPIMs and enrich PPIMs by virtual screening? (v) Can one rank known PPIMs with respect to their affinity?

To answer these questions, we chose IL-2 as a model system because of the wealth of information available for this system in terms of crystal structures of unbound IL-2, IL-2 bound to IL-2R α , and IL-2 bound to five PPIMs as well as experimental binding and inhibition data for the wild type and mutant protein.^{5,17,48} As to the methodological strategy (Figure 1), we apply and compare molecular dynamics (MD) and constrained geometric (FRODA) simulations for generating structural ensembles, introduce the PPIAnalyzer approach for investigating structural properties of protein–protein interfaces within these ensembles, and apply the MM-PB(GB)SA (molecular mechanics Poisson–Boltzmann (generalized Born) surface area) approach to identify hot spots and rank PPIMs. We note that when pursuing this strategy, we paid particular attention to mimicking a “real-life” scenario in structure-based ligand design. Thus, our strategy will be applicable also in a prospective study where nothing else than a protein–protein complex structure is known at the beginning.

MATERIALS AND METHODS

Overall Strategy. As to the methodological strategy (Figure 1), we pursued the following steps:

- Ia/b. Starting from a given protein–protein complex structure, conformational ensembles are generated on the basis of MD and FRODA simulations.
- IIa. Hot spot residues of a protein–protein complex structure are identified by MM-PBSA free energy decomposition on the basis of the structural ensemble generated by MD simulation.

- IIb. Transient pockets in the protein–protein interface are identified by energetic or geometrical criteria in conformational ensembles generated either by MD or by FRODA.
- III. PPIM binding poses are predicted by molecular docking using the hot spot and transient pocket information for guidance. This docking setup is also used to enrich PPIMs from a large set of decoys, thus performing virtual screening.
- IV. PPIMs are ranked by their MM-PBSA binding effective energies (ΔG_{eff} = gas phase energy + solvation free energy according to a continuum solvent model) calculated for conformational ensembles from MD simulations that were started from the docked binding poses.

These steps will now be described in more detail. Detailed information about structure preparation and protocols for molecular dynamic simulations and docking experiments is provided in the Supporting Information.

FRODA Simulations. FRODA is a geometrical simulation method that explores the internal mobility of biomolecular systems. Details of the algorithm can be found in ref 49. The FRODA simulation was performed with the FIRST 6.2 suite of programs. For the rigid cluster decomposition with FIRST, a hydrogen bond energy cutoff of $-1.0 \text{ kcal mol}^{-1}$ was applied together with the “H 1” function for hydrophobic interactions. For each system, 10 000 000 conformations were sampled, of which every 10 000th conformation was stored. The random displacement distance of the mobile atoms was set to 0.1 Å, and the continuous motion (CM) method was applied.

MM-PB(GB)SA Calculations. MM-PBSA^{34,50} calculations were carried out according to the “multiple trajectory method” and the “single trajectory method”. For the multiple trajectory method, snapshots of all of the IL-2 complexes, the unbound IL-2 structures, and the IL-2R α subunit as well as the small-molecule ligands were extracted from independent MD trajectories. Alternatively, for the single trajectory method, snapshots of the binding partners were extracted from MD trajectories of the complexes only. All counterions and water molecules were stripped from the snapshots. Snapshots were extracted every 10 ps. Autocorrelation analysis of the effective energy of snapshots revealed that this time interval is sufficient for generating statistically independent snapshots. The gas phase energy was calculated on the basis of the ff99SB force field⁵¹ without applying any non-bonded cutoff. The polar part of the solvation free energy was determined by solving the linearized Poisson–Boltzmann (PB) equation⁵² or by applying the “OBC” generalized Born (GB) method ($igb = 5$) using mbondi2 radii.⁵³ A dielectric constant of 1 and 80 for the interior and exterior of the solute was applied, respectively. The polar contributions were computed at 100 mM ionic strength. Nonpolar solvation energies were calculated by a solvent-accessible surface area (SASA) dependent term, using a surface tension proportionality constant of $\gamma = 0.0072 \text{ kcal mol}^{-1} \text{ Å}^{-2}$. Contributions from vibrational entropy were neglected,⁵⁴ which can be justified with the small-molecule ligands being very similar. For calculating per residue contributions, the decomposition scheme²² implemented in the SANDER and MM-PBSA code of AMBER 10 was extended to also consider the PB reaction field energy. This is done on the basis of the concept of induced surface charges on the dielectric boundary.⁵⁵ The contribution of a residue k to the reaction field energy E_{rf} is then calculated without additional computational costs as the sum of Coulomb interactions of all of

its atomic charges q_i with all induced surface charges q_p with r being the distance between the two charges (eq 1):

$$E_{\text{eff}}(k) = \sum_{i \in k} \sum_j \frac{q_i q_j}{4\pi\epsilon_0 r} \quad (1)$$

Identification of Residues in the Interface of IL-2/IL-2R α
Residues that are ≤ 5 Å apart from IL-2R α in the IL-2/IL-2R α complex structure (PDB code: 1z92) were chosen as interface residues of IL-2. Residues pointing toward the interior of IL-2 and thus not contributing to the binding of IL-2R α or small ligands were excluded upon visual inspection. This resulted in a set of 31 IL-2 interface residues: Tyr31, Asn33, Pro34, Lys35, Thr37, Arg38, Met39, Thr41, Phe42, Lys43, Phe44, Tyr45, Glu60, Glu61, Glu62, Lys64, Pro65, Leu66, Glu67, Glu68, Val69, Asn71, Leu72, Met104, Cys105, Glu106, Tyr107, Ala108, Asp109, Glu110, and Thr111.

Structural Analysis of Protein–Protein Interfaces. For investigating structural properties of protein–protein interface regions from conformational ensembles, we developed the PPIAnalyzer method (Figure S1). The method comprises three steps:

- I. Structural changes of the interface are determined in terms of root mean-square deviations (RMSD) of backbone and side chain atoms.
- IIa. The steric quality of the generated conformations is assessed.
- IIb. Distinct interface conformations are selected on the basis of a clustering with respect to the RMSD of heavy atoms of interface residues.
- III. Transient pockets are identified in those conformations.

In more detail, heavy atom RMSD values of interface residues are calculated with respect to the small molecule-bound (PDB codes: 1m48, 1m49, 1pw6, 1py2, and 1qvn) and unbound crystal structures (PDB code: 1m47) after structurally aligning the interface regions. We note that this RMSD calculation was only done to retrospectively assess the sampling of bound interface conformations; at no time of the study was knowledge of bound conformations applied for the identification of protein conformations that were subsequently used in docking experiments. The stereochemical quality of each snapshot was assessed using PROCHECK.⁵⁶ Snapshots not satisfying all of the following stereochemical criteria were excluded from further investigation:

- I. At most, two bad contacts are present.
- II. Less than 5% of the amino acids are in disallowed regions of the Ramachandran plot.
- III. At most two unfavorable main chain or side chain parameters are present. Of the remaining snapshots, 100 representative structures were selected using a k -medoids clustering algorithm⁵⁷ with respect to the RMSD of the interface residues.

Finally, potential binding pockets were detected using the PocketAnalyzer program⁵⁸ that implements a pocket identification strategy similar to the one proposed by Hendlich et al.⁵⁹ Details of the algorithm can be found in ref 58. Here, the following parameters were used: minimal degree of buriedness: 9; minimal number of neighbors: 9; minimal cluster size: 50; grid spacing: 1.0 Å. Snapshots for the subsequent docking experiments were chosen with respect to their identified pocket volume.

Data Set of Useful Decoys. Following the procedure described to generate the directory of useful decoys (DUD),⁶⁰ we selected compounds from the “purchasable subset” of the ZINC

database⁶¹ (as of May 19, 2010) that are similar with respect to physicochemical properties to the five IL-2 ligands in complex structures (Table 1). Descriptors for the ZINC compounds were downloaded from the ZINC Web site or were calculated for the IL-2 ligands using Molinspiration.⁶² The number of functional groups was calculated using the OpenEye FILTER program.⁶³ The pairwise dissimilarity between compounds was calculated as the *weighted root mean-cubed difference* (RMCD; eq 2) of the differences of normalized descriptors X_i (weights w_i in parentheses) logP (8); molecular weight (4); number of hydrogen bond donors (4) and acceptors (4); number of rotatable bonds (4); number of amide (1), amino (1), and carboxylic acid (1) groups; and the sum of the numbers of amidino and guanidino groups (1).

$$\text{RMCD} = \sqrt[3]{\frac{\sum_i w_i |X_i|^3}{\sum_i w_i}} \quad (2)$$



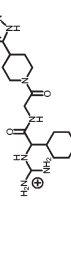
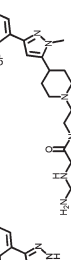

The ZINC compounds were then sorted by their pairwise RMCDs to all five IL-2 ligands in complex structures. The 10 000 most similar ZINC compounds were clustered into 1000 clusters on the basis of the pairwise RMCD using hierarchical clustering according to Ward's method as implemented in the hclust module of R.⁶⁴ Out of each cluster, the compound with the smallest RMSD of pairwise RMCD to the reference ligands was selected. Four compounds were not retrievable from the ZINC. For the remaining 996 unique decoy structures, a total of 1297 protonation and tautomerization states, as stored in the ZINC database, were considered.

Ranking of Docked Structures. For ranking docked structures by MM-PBSA, appropriate starting structures for the MD-based snapshot generation must be chosen initially. For this, consider that 100 docking runs were performed for each of the 10 structures with the largest interface pocket volumes obtained by a FRODA simulation of the unbound IL-2 structure and for each ligand. The starting structure selection was based on hot spot information, RMSD clustering, and intermolecular docking energy. First, it was required that a ligand's guanidinium group be within 5 Å of the side chain heavy atoms of the hot spot amino acid Glu62. Second, all of the remaining docking poses were clustered with respect to heavy atom RMSD of the PPIMs in the docked complexes after aligning only the proteins. Hierarchical complete linkage clustering was performed with R⁶⁴ with a cluster distance of 5 Å. Compared to clustering docking results above, a larger clustering distance of 5 Å was chosen to account for the fact that the clustering is performed over complex structures with different receptor conformations. Finally, the structure with the lowest intermolecular docking energy from the largest cluster was chosen as a starting structure for MD simulation and subsequent MM-PB-(GB)SA binding effective energy calculation. Equilibrations, production runs, and MM-PB(GB)SA calculations were performed as described in the Supporting Information for the crystal structures. To allow the complex structures to relax after geometrical FRODA simulations and subsequent docking, we performed 20 ns of unrestrained MD simulation, with only the last 10 ns being used for snapshot extraction.

RESULTS AND DISCUSSION

Structures from MD Simulations. In order to investigate the opening of transient pockets in the IL-2 interface and to calculate effective energies by MM-PB(GB)SA, conformational ensembles of

Table 1. Structures of IL-2, IL-2 Complexes, and IL-2 Binding Partners with Respective Structural, Experimental, and Computed Data

PDB code	1m47 ^a	1m4c ^a	1m48 ^b	1m49 ^b	1pw6 ^b	1py2 ^b	1qvn ^b	1z92 ^c
Ligand	—	—						IL-2R α
Ligand abbreviation	—	—	FRG	CMM	FRB	FRH	FRI	IL-2R α
Resolution ^d	1.99	2.40	1.95	2.00	2.60	2.80	2.70	2.80
Experimental affinity ^e	—	—	-6.90 ^j	-7.18 ^k	-7.02 ^l	-9.60 ^m	-8.03 ⁿ	-10.97 ^o
Calculated affinity	—	—	-44.50	-47.84	-49.72	-56.32	-49.13	-77.19
(SEM) ^j	—	—	(± 0.14)	(± 0.16)	(± 0.13)	(± 0.16)	(± 0.17)	(± 0.37)
MD length ^g	11.21	12.47	9.11	13.79	6.97	6.17	9.15	10.00
Drift ^h	—	—	0.20	0.19	-0.27	-0.63	0.74	0.49
IL-2 energy	-3181.50	-3182.45	-3203.77	-3211.09	-3212.32	-3210.79	-3188.06	-3190.96
(SEM) ^j	(± 1.06)	(± 1.02)	(± 1.25)	(± 1.19)	(± 1.36)	(± 1.47)	(± 1.36)	(± 1.33)

^a Unbound IL-2. ^b IL-2 in complex with a small molecule. ^c IL-2 in complex with the extracellular IL-2R α . ^d In Ångströms. ^e In kilocalories per mole. ^f Binding effective energy calculated by the MM-PBSA single trajectory method, in kilocalories per mole. Standard error of the mean (SEM) in parentheses. ^g In nanoseconds. ^h Average drift of the binding effective energy in the MD trajectory, in kilocalories per mole. Standard error of the mean (SEM) in parentheses. ⁱ Calculated average absolute effective energy of IL-2 in the unbound state or the bound state as extracted from the complex ensemble calculated by linear regression in kilocalories per mole per nanosecond. ^j Derived by van't Hoff analysis of surface plasmon resonance data. ^k Derived from the average of $K_d = 7.5 \mu\text{M}$ and $K_d = 4.2 \mu\text{M}$ as determined by equilibrium analytical ultracentrifugation. ^l Calculated from $K_d = 7 \mu\text{M}$ as determined by surface plasmon resonance. ^m Calculated from $K_d = 100 \text{ nM}$ as determined by surface plasmon resonance. ⁿ Calculated from $K_d = 1.4 \mu\text{M}$, which was calculated by $K_{d,A} = K_{d,B} (\text{IC}_{50,A}/\text{IC}_{50,B})$ from the reported IC_{50} values of FRB ($\text{IC}_{50} = 10 \mu\text{M}$)⁶ and FRI ($\text{IC}_{50} = 2 \mu\text{M}$)¹⁵, the latter two being measured by ELISA under identical conditions.⁴⁵ and in agreement with the calculation from K_d 1 nM as denoted by Rickert et al.⁴⁶

unbound IL-2 and IL-2 bound to either IL-2R α or five PPIMs were generated by MD simulations of at least 6 ns in length (Table 1).

For all systems, the RMSD of heavy atoms with respect to structures at the end of the equilibration procedure (see Supporting Information) was determined (Table S1, Supporting Information). Over all trajectories, IL-2 showed mean RMSD values of 2.55–3.46 Å. These values are in good agreement with those generally found during other MD simulations.⁶⁵ The interface regions of IL-2 showed generally lower RMSD values of 1.94–2.88 Å, with the interface region of the two unbound IL-2 structures showing the largest structural deviations (2.44 and 2.88 Å). This agrees well with observations from X-ray crystallography that show an opening of transient pockets in this region (see below). The small-molecule ligands bound to IL-2 showed RMSD values between 0.89 and 2.20 Å and, thus, stayed close to the initial binding region. Overall, after an initial rise during the first 2–4 ns, the RMSD values remain constant for the remainder of the MD simulations (Table S1).

To investigate the mobility of IL-2 and its interface region, we calculated root-mean-square fluctuations (RMSFs) of heavy atoms of protein residues. Starting structures of MD simulations with RMSF values mapped in a color-coded fashion are shown in Figure S2 (Supporting Information). Unsurprisingly, the largest fluctuations up to 9.58 Å were found for flexible loop regions and the termini. Many of these mobile regions have not been resolved in several of the crystallographic structures,^{5,66} which already provides a hint as to their mobility. In contrast, all of the interface residues of IL-2 show RMSF values <2.50 Å. Interestingly, the mobility of Phe42 of IL-2, whose conformational transition is crucial for the opening of a transient pocket (see below), was found to be significantly higher in the unbound structure (RMSF = 1.86 Å (1.60 Å) for 1m47 (1m4c)) than in the bound structures (RMSF between 0.57 and 1.15 Å).

MM-PB(GB)SA calculations, which make use of a continuum electrostatic model for evaluating (de)solvation effects, may fail if structural waters are present in or close to the binding interface.^{67,68} To identify such water molecules, we calculated the RMSFs of all water molecules and, subsequently, investigated the residence times of waters with low RMSFs by visual inspection. First, the analysis did not reveal any long-lasting (residence time >1 ns) water molecule on the outer surface of the IL-2 interface except in the case of the IL-2/IL-2R α complex. However, none of these water molecules formed strong interactions with the protein for the complete simulation time. Second, in the interior of IL-2 adjacent to the binding interface, long-lasting (residence time >1 ns) water molecules were found at three distinct sites (I, in the interior of IL-2 close to Glu62; II, inside of a loop region enclosed by Tyr45, Ala108, Asp109, and Glu110; III, at the N-terminal end of helix D enclosed by Met39, Phe42, and Leu114). However, none of these waters is in direct contact with any of the IL-2 ligands. Furthermore, these waters are conserved in almost all simulations of unbound and bound IL-2 so that potential effects on MM-PB(GB)SA results should cancel. Overall, these findings lead us to expect only minor influences due to structural waters on MM-PB(GB)SA results for hot spot prediction and ligand ranking.

Identification of Hot Spots by MM-PBSA Free Energy Decomposition. Mimicking localized interactions in hot spot regions of protein–protein interfaces has proven valuable for PPIM development.⁴ We thus set out to computationally identify

hot spots in the protein–protein interface of IL-2/IL2R α and IL-2/small-molecule complexes (Figure 2). For this, we implemented and applied the MM-PBSA per residue effective energy decomposition,^{25,69} which complements the MM-GBSA effective energy decomposition introduced by us.²² Here, we applied the MM-PBSA single trajectory method. While this method neglects energetic contributions due to conformational changes of the binding partners, it leads to a drastic reduction in the statistical uncertainty of the free energy components.^{34,70,71}

For validation, computed effective energy components were compared to experimentally determined changes in the binding free energy of IL-2/IL-2 R α and IL-2/FRH complexes upon mutations of IL-2 interface residues to alanine.^{6,15} The experiments showed that for both IL-2R α and FRH binding was strongly disrupted ($EC_{50,Ala}/EC_{50,WT} \geq 100$, equivalent to $\Delta G \geq 2.8$ kcal mol⁻¹ at 37 °C) when Phe42, Tyr45, or Glu62 were mutated to alanine. Encouragingly, the effective energy decomposition also identified Phe42 and Glu62 as hot spot residues ($\Delta G_{eff} = -2.84$ to -4.45 kcal mol⁻¹) and only slightly underestimated the contribution of Tyr45 ($\Delta G_{eff} \approx -1.3$ kcal mol⁻¹). Thus, experimental and computational predictions of hot spots are in good agreement. Moderate changes in the binding affinity ($EC_{50,Ala}/EC_{50,WT} \geq 10$, equivalent to $\Delta G \geq 1.4$ kcal mol⁻¹ at 37 °C) were observed for IL-2/IL-2R α when IL-2 residues Thr41, Lys43, or Phe44 were mutated to alanine.¹⁵ Computed ΔG_{eff} values are in the range of -0.4 to -0.8 kcal mol⁻¹ in these cases, demonstrating that smaller effects on the binding affinity could be well identified by the MM-PBSA effective energy decomposition, too. Finally, residues Lys35, Arg38, and Leu72 were also identified as hot spots by the MM-PBSA effective energy decomposition. However, except for Leu72, which moderately disrupted the IL-2/FRH complex ($EC_{50,Ala}/EC_{50,WT} \geq 10$) when mutated, all others did only show a weak disrupting effect on IL-2R α and FRH binding ($EC_{50,Ala}/EC_{50,WT} = 3-5$).

The seemingly prominent interactions of Arg83 ($\Delta G_{eff} \approx -2.3$ kcal mol⁻¹) are dominated by an intermittent salt bridge to Asp56 (IL-2R α residues are highlighted in italics, whereas IL-2 residues are depicted in “normal” font), for which the proteins needed to undergo structural changes during the MD simulation. As Arg83 is far apart from the localized cluster formed by the other hot spots, it was neglected for the guidance of the subsequent docking. Finally, our calculations predict Glu68 to contribute unfavorably ($\Delta G_{eff} = +1.72$ kcal mol⁻¹) to the binding of IL-2R α ; Glu68 thus is a “cold spot”. In summary, the identified hot spots cluster together and form a functional epitope localized on helices A' and B' with an approximate area of 500 Å² corresponding to 20% of the total protein–protein interface area.

Mimicry of Localized Interactions in Hot Spot Regions by PPIMs. Next, we investigated to what extent the PPIMs mimic IL2R α as an interaction partner. If such mimicry existed, hot spots identified in a protein–protein complex could be used for guiding PPIM identification and development. Figure 2 reveals that the energetic fingerprint of IL-2/IL-2R α is indeed highly similar to the energetic fingerprints of the IL-2/small-molecule complexes. Three amino acids stand out in that respect: I. Phe42 is the center of a hydrophobic core forming contacts to other IL-2 residues (Met39, Val69, Leu72) as well as Met25, Asn27, Leu42, and His120, which explains its hot spot character in the IL-2/IL-2R α case. Moreover, Phe42 forms favorable interactions with the piperidine (FRI, FRH), pyrazole (FRB), or central phenyl

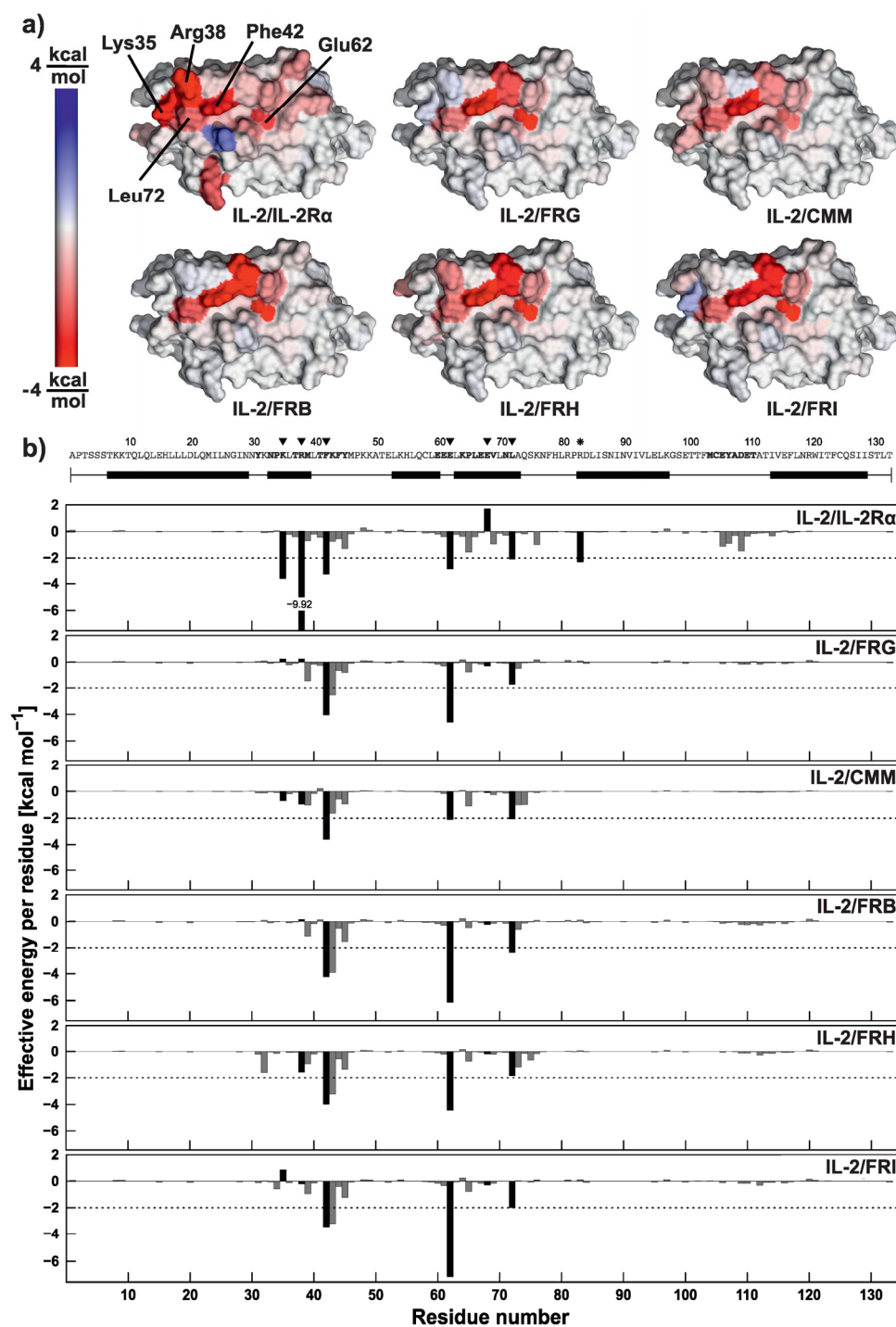


Figure 2. Per-residue contributions to the binding effective energy as calculated by MM-PBSA decomposition. The per-residue contributions were calculated by applying the single trajectory MM-PBSA method to the MD trajectories of IL-2 in complex with IL-2R α (PDB code: 1z92) and the five PPIMs FRG, CMM, FRB, FRH, and FRI (PDB codes: 1m48, 1m49, 1pw6, 1py2, and 1qvn). (a) The per-residue contribution is mapped onto the crystal structure of IL-2 bound to IL-2R α using a color-code with a linear scale for IL-2/IL-2R α , IL-2/FRG, IL-2/CMM, IL-2/FRB, IL-2/FRH, and IL-2/FRI. In (b) per-residue contributions of the six complexes are depicted as bar plots. At the top, the IL-2 sequence is depicted in single letter code with all interface residues highlighted (bold). α -helices are marked as horizontal boxes in the line below. Hot spots (highlighted as black bars in the energy plot and marked by black triangles (▼) above the sequence) were selected based on the per-residue decomposition of the IL-2/IL-2R α trajectory by applying an energy cutoff of 2 kcal mol⁻¹ (dashed line). See text for details on Arg83, which is marked by an asterisk (*) above the sequence.

ring of the tolane moiety (FRG, CMM) of the PPIMs and their adjacent amide moieties, which makes Phe42 also a hot spot for the binding of these PPIMs. II. Glu62 forms a stable salt bridge with Arg36 in the protein–protein complex. Glu62 is also the strongest anchor for the binding of the five PPIMs by salt-bridge formation with the guanidinium groups present in all ligands. III. The hydrophobic interaction of Leu72 with Met25 and Leu2 is replaced by the phenyl and 1,2-dichlorophenyl moieties of FR1, the 1,2-dichlorophenyl moieties of FRH or FRB, the indolyl moiety of CMM, or the terminal phenyl ring of the tolane moiety of FRG. In contrast, interactions involving Lys35 and Arg38, which are important for IL-2/IL-2R α affinity, are not or only weakly mimicked by the PPIM.

In summary, these findings demonstrate that three out of five computationally identified hot spots of the IL-2/IL-2R α complex are equally important for small-molecule binding to IL-2. Furthermore, all five computed hot spots cluster in a subregion of the structural epitope of IL-2/IL-2R α . Together, this strongly suggests that hot spots computed from protein–protein complexes can be used for guiding the identification and optimization of PPIM.

Opening of a Transient Pocket during Simulations Started from the Unbound State. PPIMs have been found to be particularly effective when they bind to well-defined clefts or grooves in the protein–protein interface.^{5,9,72,73} Here, we investigate whether an opening of transient pockets in the rather flat protein–protein interface of IL-2 can be observed when sampling the conformational space of unbound IL-2, following the “conformational selection” model.⁴⁷ Two conformational sampling techniques were used. First, as a state-of-the-art method, we applied all-atom MD simulations in explicit solvent of 10 ns in length. This setup is similar to a study by Helms and Eyrich.⁷ Second, as a computationally cheaper alternative, we applied the geometrical simulation method FRODA.⁴⁹ FRODA has already been successfully applied for identifying spontaneous and relevant *apo-to-holo* conformational transitions of HIV-1 TAR RNA.⁷⁴ FRODA relies upon a decomposition of a biomacromolecule into rigid and flexible regions.⁷⁵ In the unbound structure of IL-2, helices A, A', B, B', C, and D form discrete rigid clusters that are interconnected by flexible hinges (Figure S3, Supporting Information). With respect to the interface region, only 25% of the atoms belong to rigid clusters. This ensures that the majority of the interface atoms can move freely during the FRODA simulation.

The interface heavy atom RMSD of all snapshots of MD and FRODA simulations were calculated with respect to the PPIM-bound and unbound IL-2 structures (Table S2, Supporting Information). The overall structural changes between bound and unbound IL-2 structures are small: in the best case, an interface conformation coming as close as 1.42 Å RMSD to a bound state was found, starting from a structural deviation between bound and unbound structures of 1.50 Å. This can be explained in that structural deviations between bound and unbound structures are uniformly distributed over the interface rather than caused by a large-scale collective movement. Interestingly, when comparing the performance of FRODA and MD simulations, FRODA snapshots were generally found to be more similar to four out of five bound IL-2 structures than were MD snapshots. We attribute this fact to an appropriate coarse-graining of the unbound IL-2 structure prior to the FRODA simulation. Apparently, residues were correctly identified to be part of a rigid cluster that is not involved in pocket opening, leading to a focusing of movements to that region where a pocket opens up (Figure S4,

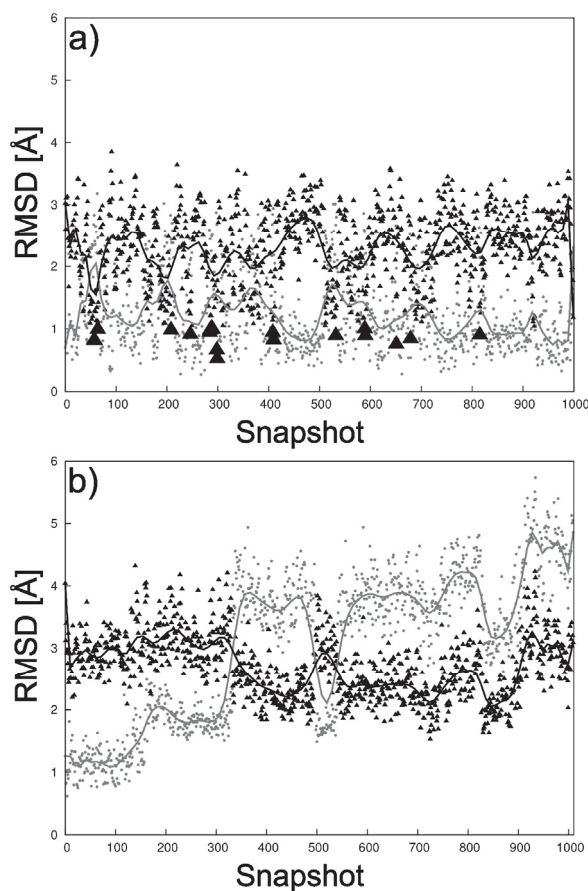


Figure 3. Conformational analysis of the interface residue Phe42. Consecutive snapshots generated by (a) FRODA and (b) MD simulation of unbound IL-2 are compared to the bound (black, PDB code: 1m48) and unbound IL-2 structures (gray, PDB code: 1m47) regarding the RMSD of Phe42 based on a structural alignment of heavy atoms of the interface. For the FRODA simulation, the RMSD shows a clear antidromic character that indicates the flipping of the aromatic ring. In contrast, for the MD simulation, the antidromic character of the RMSD curve is much less pronounced. Phe42 conformations that come closer to the bound IL-2 state than 1 Å RMSD are marked by black triangles.

Supporting Information). In contrast, in the MD simulations, all residues are allowed to move freely, leading to larger overall structural deviations in the protein–protein interface that do not necessarily lead to a structure close to a bound conformation.

Of the residues that line the PPIM binding pocket, Phe42 has been described as functioning as a gate keeper by flipping its phenyl ring.^{5,15} In Figure 3, RMSD time series calculated from all heavy atoms of Phe42 with respect to both unbound and bound IL-2 structures are shown for FRODA and MD simulations. The FRODA simulation shows an antidromic behavior of both time series, and the simulated conformations repeatedly approach the bound structure and depart from it. This leads to Phe42 approaching the small molecule bound conformation to <1.0 Å RMSD in 18 cases during the simulation. In contrast, the RMSD time series of the MD simulation showed a much less pronounced antidromic behavior, and a conformation of Phe42 with an RMSD <1.0 Å with respect to bound IL-2 was never detected during the simulation.

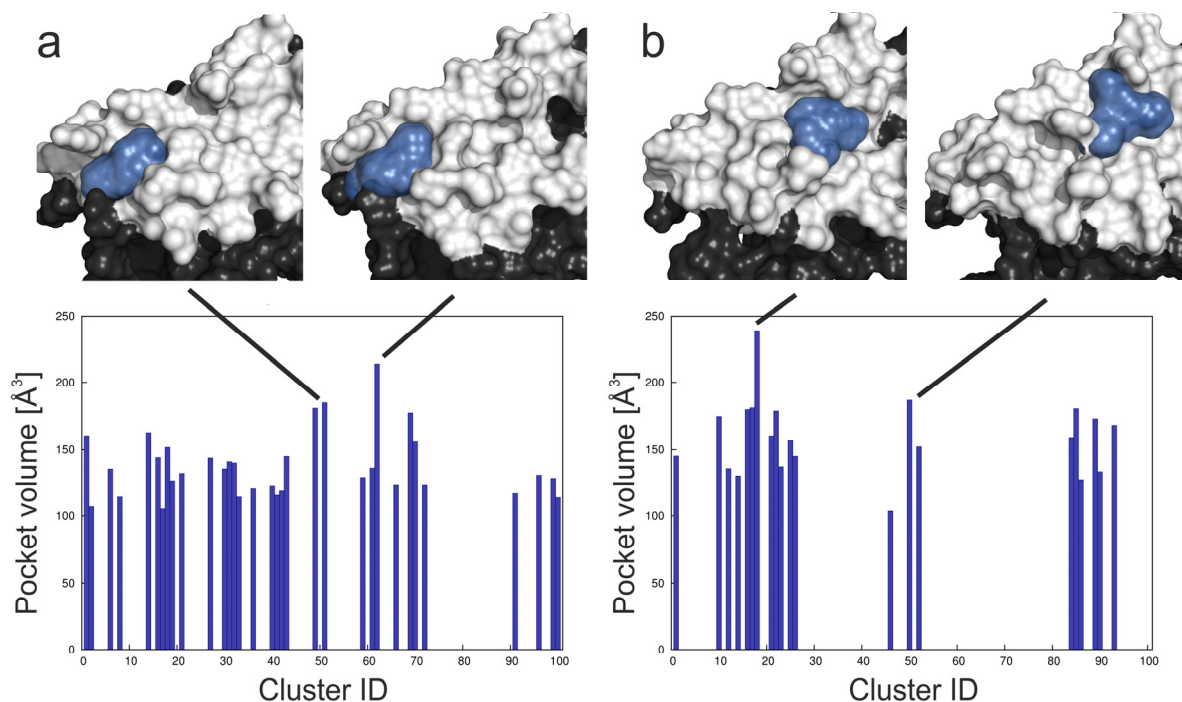


Figure 4. Detection of interface pockets in the cluster representatives of IL-2 structures generated by (a) FRODA and (b) MD simulation. The box plots depict pocket volumes computed by PocketAnalyzer. In addition, the two largest pockets found in IL-2 structures generated by either simulation method are shown. Notably, the locations of these pockets differ for both methods.

Identification of Transient Pockets in Structural Ensembles. Next, the question needs to be addressed as to how one can identify binding-competent conformations from the generated ensembles without already knowing the bound conformation from experimental results. An evaluation based on energetic criteria appears to be difficult^{34,76} (see the Supporting Information on “Statistical significance of MM-PBSA results” and Figures S5 and S8) because the expected change in free energy accompanying the opening of a transient pocket is on the order of $k_B T$, which is several orders of magnitude smaller than the total conformational energy of the protein.⁷⁷ Instead, we resorted to a sequential scheme that involved checking the stereochemical quality of the simulated conformations, clustering of similar conformations, and identification of transient interface pockets based on volume and the degree of “buriedness”. Thus, we only considered geometrical parameters for the identification of transient pockets.

Conformations sampled by either MD or FRODA simulations generally showed a high degree of stereochemical quality, as determined by PROCHECK.⁵⁶ In fact, none of the FRODA conformations had to be excluded from further investigations, whereas only 45 MD conformations were discarded because they had more than two unfavorable main chain parameters. In the next step, 100 structurally varying interface conformations were selected as representatives from each simulation by *k*-medoids clustering. The interface RMSD of the selected representatives ranges from 0.85 to 3.43 Å for MD-generated conformations and from 0.78 to 2.14 Å for FRODA-generated conformations. Finally, the PocketAnalyzer program⁸⁸ was applied for pocket detection. Pockets embraced by at least 70% of the interface residues were identified as interface pockets. As for the crystal

structures, all small-molecule bound structures displayed interface pockets with volumes ranging from 107 to 234 Å³, with all of these pockets being located between residues Lys35, Arg38, and Phe42. No pocket was present in the unbound IL-2 structure as well as the IL-2/IL-2R α complex.

In 33% of the selected FRODA conformations, an interface pocket was detected (Figure 4). The average volume of these pockets is 138 Å³, with minimal (maximal) values of 104 Å³ (215 Å³). Similar to the crystal structures, all interface pockets from FRODA-generated conformations were located between residues Lys35, Arg38, and Phe42. As for MD-generated conformations, an interface pocket was identified in 22% of the conformations, with an average volume of 159 Å³ (min., 103 Å³; max., 240 Å³). Notably, 2/3 of these interface pockets are located between Lys43, Tyr45, and Phe42 and, thus, deviate in position from the pockets found in the bound crystal structures.

Docking into Transient Pockets. We then investigated whether simulated IL-2 conformations with transient pockets in the protein–protein interface can be used as receptor structures for docking. Therefore, we selected those 10 representative structures from each simulation that showed the largest interface pocket volume (Table S3, Supporting Information). Each of the five known IL-2 ligands (Table 1) was then docked into this set of conformations. To exclude any bias due to the knowledge of the experimentally determined complex structures, the placement of the potential grids for docking (Figure 5) was solely based on (I) all hot spots identified by MM-PBSA except Arg83 (see above, Figure 2) and (II) all amino acids lining the identified interface pocket (Table S4, Supporting Information). Docking was considered successful when the ligand pose with the lowest

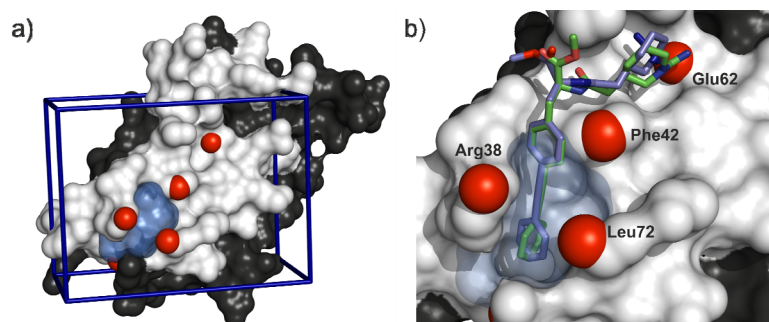


Figure 5. Definition of the potential grid and exemplary docking result for IL-2. The interface region of IL-2/IL-2R α is colored white. (a) The hot spots identified by MM-PBSA decomposition (red spheres) and the transient pocket (blue surface) were used to define the location and size of the potential grid (see Materials and Methods). The grid dimensions are represented by a blue box. (b) Predicted binding pose of the ligand FRG (blue sticks) docked into a FRODA snapshot containing an identified transient pocket. The RMSD between the predicted and crystallographic binding pose (PDB code: 1m48, green sticks) is 1.28 Å.

Table 2. Results of Redocking and *apo*-Docking

redocking					<i>apo</i> -docking				
ligand	PDB code ^a	RMSD ^b	score ^c	cluster size	ligand	PDB code ^a	RMSD ^b	score ^c	cluster size
FRG	1m48	1.50	−13.84	84	FRG	1m47	2.58	−10.93	19
CMM	1m48	1.37	−15.69	82	CMM	1m47	3.08	−12.53	40
FRB	1pw6	0.59	−13.01	25	FRB	1m47	4.63	−10.82	51
FRH	1py2	0.87	−15.38	53	FRH	1m47	3.80	−14.10	9
FRI	1qvn	1.14	−14.65	27	FRI	1m47	8.10	−13.30	9

^a PDB code of the IL-2 complex structure. ^b RMSD of the ligand pose with the lowest energy in the largest cluster with respect to the native structure, in Ångströms. ^c In kilocalories per mole.

intermolecular docking energy in the largest cluster had an RMSD < 2.0 Å to the native pose.

For comparison, we first redocked the five IL-2 ligands into the corresponding IL-2 complex structure. Likewise, we also performed docking of these ligands to an *apo* structure of IL-2 (PDB code: 1m47). Since no interface pocket could be detected in this *apo* structure, the same potential grid definition as for the redocking approach was used for the *apo*-docking. The redocking was successful in all cases, whereas *apo*-docking failed (Table 2). The latter is not unexpected due to the absence of any pronounced indentation in the protein–protein interface. Additionally, the success and convergence of the redocking is demonstrated by the occurrence of ≥ 25 poses in the largest cluster for IL-2/FRB, IL-2/FRI, and IL-2/FRH and >80 poses in the largest cluster for IL-2/FRG and IL-2/CMM.

The results of docking into transient pockets of simulated conformations are summarized in Table 3. Notably, docking into MD-generated conformations was not successful because in no case was a docked pose with RMSD < 2 Å obtained, and in only 2 out of 50 dockings was a pose with RMSD < 2.5 Å identified. This failure is a result of the interface pockets being located at a different position than the pockets found in the bound crystal structures. A more detailed inspection of the MD-generated conformations showed that correctly localized transient pockets do exist (data not shown). However, these pockets are much less pronounced than those occurring in FRODA-generated snapshots and, hence, are not among the 10 largest pockets chosen for the docking experiments. This is because a hydrophobic channel

Table 3. Number of Successful Attempts of Docking into Snapshots with Identified Transient Pockets

ligand	RMSD ^a							
	MD				FRODA			
	<2.0	<2.5	<3.5	≥ 3.5	<2.0	<2.5	<3.5	≥ 3.5
FRG	0	2	1	7	2	6	1	1
CMM	0	0	4	6	2	4	4	0
FRB	0	0	3	7	0	0	7	2
FRH	0	0	0	10	1	2	4	3
FRI	0	0	0	10	1	1	0	8

^a RMSD of the ligand pose with the lowest energy in the largest cluster with respect to the native structure, in Ångströms.

embraced by Lys35, Arg38, and Phe42 must open for the transient pockets to be correctly localized.⁵ Such an opening tends to be less pronounced in MD-generated ensembles, as demonstrated by the lower hydrophobicity of these pockets when compared to pockets found in crystal structures.⁷

In contrast, we were able to identify the hydrophobic channel in all of the selected FRODA-generated snapshots, possibly due to the absence of solvent in the simulation process. As a consequence, docking to at least one transient pocket was successful for all IL-2 ligands but FRB (Table 3). This is exemplarily shown for the ligand FRG in Figure 5b.

These results are encouraging in that a drop in docking accuracy compared to redocking is often found to be mirrored by the degree to which a protein moves upon ligand binding.^{78,79} Thus, docking to an *apo* form usually shows the largest deterioration.⁸⁰ Being able to start from the *apo* IL-2 structure and identify transient pockets in trajectories of computationally inexpensive FRODA simulations that are adequate for ligand docking thus is a valuable achievement.

Docking Enrichment in a Large Set of Decoys. In order to demonstrate that the identified hot spots and transient pockets

Table 4. Docking Enrichment of Known IL-2 Ligands^a

FRODA snapshot ^b	EF _{max} ^{c,d}	EF ₂₀ ^{c,d}	EF ₃ ^{c,d}	EF ₁ ^{c,d}	AUC ^{d,e}
6	23.7 (260.2)	4.6 (5.0)	17.4 (26.0)	22.0 (74.3)	0.95 (0.99)
117	23.8 (173.6)	4.4 (5.0)	18.0 (32.5)	22.1 (74.4)	0.95 (0.99)
169	23.7 (260.0)	4.8 (5.0)	22.0 (32.5)	23.7 (92.9)	0.98 (1.00)
301	23.8 (260.4)	4.9 (5.0)	19.7 (26.0)	18.7 (55.8)	0.99 (0.99)
418	23.8 (260.4)	4.8 (5.0)	10.4 (26.0)	11.9 (37.2)	0.94 (0.98)
514	19.0 (97.6)	5.0 (5.0)	16.2 (32.5)	17.0 (55.8)	0.98 (0.99)
534	23.8 (86.8)	4.4 (5.0)	11.0 (26.0)	13.6 (37.2)	0.92 (0.98)
657	16.2 (57.8)	4.0 (4.0)	10.4 (19.5)	15.3 (55.8)	0.89 (0.93)
698	23.8 (260.4)	4.1 (5.0)	16.2 (32.5)	23.8 (55.8)	0.94 (0.99)
729	23.8 (86.8)	4.6 (5.0)	12.2 (26.0)	13.6 (55.8)	0.95 (0.99)

^a The set of known IL-2 ligands consists of five IL-2 ligands with available complex crystal structures as well as 52 structures with similar scaffolds and known IC₅₀.^{6,17,48} The set of decoys consists of 996 unique structures with a total of 1297 protonation and tautomerization states.

^b Consecutive number of the snapshot from a total of 1000 snapshots uniformly extracted from the 10 000 000 FRODA-generated structures starting from the unbound IL-2 structure (PDB code: 1m47). The 10 snapshots with the largest pocket volumes were used. ^c EF₁, EF₃, EF₂₀, and EF_{max} correspond to the enrichment factors at 1%, 3%, and 20% of the ranked database and the maximal enrichment factors over the whole data set. ^d Values correspond to all 57 known IL-2 ligands, while values in parentheses correspond to the five IL-2 ligands with available complex crystal structures only. ^e Area under the receiver operator curve (ROC).

could also be used for structure-based virtual screening (VS), we performed a retrospective VS for IL-2 PPIMs. As known binders, the five IL-2 ligands in complex structures (Table 1) and 52 structures with similar scaffolds and known IC₅₀ values were used.^{6,17,48} Decoys were selected from the “purchasable subset” of the ZINC database⁶¹ following the procedure described for creating the directory of useful decoys (DUD).⁶⁰ The DUD procedure aims at selecting decoy structures that are physicochemically similar to known binders in order to avoid any bias in enrichment calculations. This led to 996 unique decoy structures with a total of 1297 protonation and tautomerization states. We note that during docking, the docking scores were normalized by the square root of the molecular weight of a ligand in order to correct for any size related bias, too.⁸¹

For the 57 (5) known IL-2 ligands (in complex structures) we found good enrichments for the individual transient binding pockets (Table 4, Figure 6, Figure S6, Supporting Information) with EF_{max} = 16.2–23.8 (EF_{max} = 57.8–260.4) and EF₁ = 13.6–23.8 (EF₁ = 37.2–92.2), and area under the curve values of receiver operator curves of AUC ≥ 0.89 (AUC ≥ 0.93).

We note that these enrichments may be too optimistic compared to a real-life scenario and, hence, should be interpreted cautiously because the VS has likely benefitted from the fact that the known IL-2 ligands were structurally optimized for binding to IL-2 and partially violate Lipinski’s rules.⁸² Hence, even though the decoys were selected following the DUD procedure, in some cases, a perfect match of the property distribution curves between binders and decoys could not be achieved (Figure S7, Supporting Information). This is particularly true for the properties “molecular weight”, “no. of hydrogen bond donors”, and “no. of amidino and guanidino groups”. Still, with respect to the aim of this study, our results demonstrate that known IL-2 ligands could successfully be screened from a set of decoys using only information about hot spots and transient pockets on the protein side.

Ranking of IL-2 Ligands. Binding effective energies calculated by the MM-PBSA single trajectory method appear to be converged and remain stable throughout simulation lengths of 6–14 ns (Table S5 and Figure S8, Supporting Information).

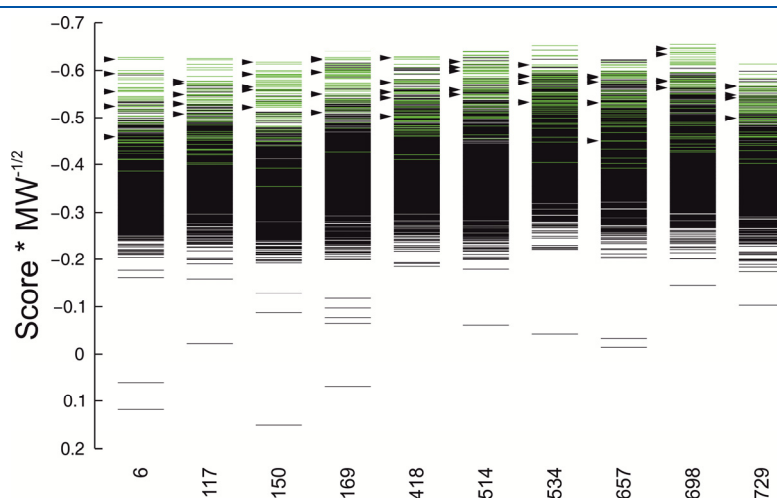


Figure 6. Ranking of docked structures. The best poses (as defined in Materials and Methods) of docked binders and decoy structures were ranked by intermolecular energy divided by the square root of the molecular weight. The 10 FRODA structures with the largest transient pocket are indicated by their snapshot number at the bottom. The 57 known IL-2 ligands and the decoys are depicted with green and black lines, respectively. In addition, the five IL-2 ligands with available complex crystal structures are highlighted by arrows.

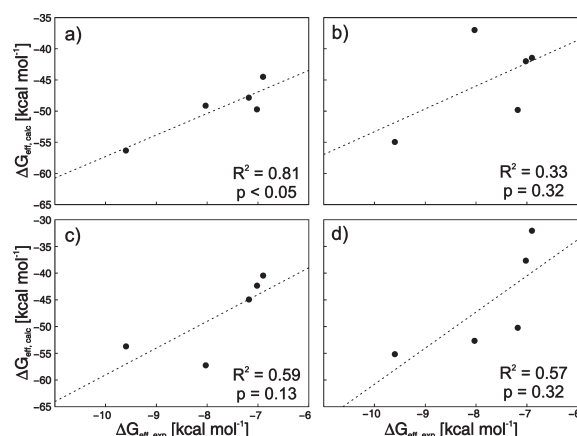


Figure 7. Correlation of computed binding effective energies ($\Delta G_{\text{eff,calc}}$) with respect to experimental free enthalpies of binding ($\Delta G_{\text{eff,exp}}$). Binding effective energies were calculated using the Poisson–Boltzmann continuum solvation model (a and c) and the generalized Born continuum solvation model (b and d) within the MM-PB(GB)SA single trajectory method. MM-PB(GB)SA calculations were based on either crystallographic (a and b) or docked (c and d) starting structures.

Absolute binding effective energies computed for the five IL-2 ligands starting from *crystal structures* of the complexes are about 45 kcal mol^{-1} more negative than the experimentally determined free enthalpies of binding (Figure 7, Table 1). Two reasons account for this. First, unfavorable energetic contributions due to conformational strain of the binding partners are not taken into account in the single trajectory method. These contributions can be as high as 36 kcal mol^{-1} .⁸³ Second, we neglect any changes in the configurational entropy of the binding partners, which accounted for contributions to the free energy of $20\text{--}30 \text{ kcal mol}^{-1}$ at 300 K in related studies.^{22,84} However, these two points do not have a major impact on relative binding effective energies as demonstrated by a good ($R^2 = 0.81$) and significant ($p < 0.05$) linear correlation of computed and experimentally determined binding energies, allowing for a successful ranking of four out of the five ligands (Figure 7a). This result is all the more remarkable in that the range of experimentally determined binding affinities is only 3 kcal mol^{-1} . To our knowledge, this is one of only a few reports so far of successfully applying MM-PBSA to rank PPIMs.^{69,85} When using the GB model, no significant correlation is obtained. In particular, the binding effective energies of ligands FRI and CMM show strong deviations from the correlation line (Figure 7b). A per residue decomposition of the binding effective energies using the GB model did not allow one to assign the origin of these deviations to contributions by a particular set of residues (data not shown).

For ranking *docked* IL-2/small-molecule complex structures (Figure 7), first, reasonable poses obtained by docking into FRODA-generated structures with transient pockets were selected without making use of any knowledge of the bound crystal structures (Table S6, Supporting Information). Then, MD simulations and MM-PB(GB)SA calculations were applied as in the case of the crystal structures. The linear correlation of computed MM-PBSA binding effective energies with respect to experimentally determined binding free energies is fair ($R^2 = 0.59$) and weakly significant ($p = 0.13$; Figure 7c). Again, no significant correlation was obtained in the case of MM-GBSA. The largest deviations from the correlation line are observed for ligands

FRH and FRI in the case of MM-PBSA, which also show the largest structural deviations from the native pose in the starting structures (Table S6). These results demonstrate that for MM-PBSA calculation to be successful in ranking PPIMs, (at least) good starting structures ($\text{RMSD} < 2.5 \text{ \AA}$) are required. Nevertheless, it is encouraging to note that the quality of the generated docking poses was still sufficient to successfully discriminate between the subgroups of high and low affinity ligands.

CONCLUSION

We have presented for the first time a computational strategy that simultaneously considers aspects of energetics and plasticity in the context of PPIM binding to a protein interface. In particular, our strategy aims at identifying the determinants of small-molecule binding, hot spots and transient pockets, in a protein–protein interface in order to make use of this knowledge for predicting binding modes of and ranking PPIMs with respect to their affinity. Although performed in a retrospective manner on the well-investigated system of IL-2, we note that at no point in the study did we utilize information from the experiment about the binding mode and affinity of PPIMs. Thus, our strategy will be applicable also in a prospective manner where nothing other than a protein–protein complex structure is known; hence, it can well be the first step in a structure-based endeavor to identify PPIMs.

Perhaps the most surprising result from a methodological point of view is that the computationally much cheaper constrained geometric simulation method FRODA outperforms state-of-the-art MD simulations in *sampling* transient pockets in the IL-2 interface. Apparently, the neglect of solvent in FRODA not only leads to a reduced computational burden but also facilitates the opening of a hydrophobic channel. Although applied to only one protein–protein interface in the present study, we note that the good performance of FRODA in sampling relevant conformational transitions is in line with results obtained by this⁷⁴ and a related method⁸⁶ on other systems.

It is encouraging that geometrical parameters summarized in the PPIAnalyzer method sufficed to successfully *identify* transient pockets. On the one hand, this finding alleviates the need to resort to conformational free/effective energies for identifying such pockets. Using energetic criteria is hampered by the demand for very precise computations due to the fact that small differences in conformational energies must be calculated from large absolute values. On the other hand, this finding reconfirms the FRODA results, as it demonstrates that the most pronounced pockets only opened up where expected. As FRODA strongly depends on a preceding flexibility analysis of the protein, it is thus tempting to speculate that regions that are prone to open transient pockets could be identified by such a flexibility analysis. This knowledge could then be used to focus other locally enhanced sampling schemes on that particular region.⁸⁷

Finally, we consider it a valuable achievement that the sequence of, first, docking to identified transient pockets; second, starting structure selection based on hot spot information, RMSD clustering, and intermolecular docking energies; and third, MM-PBSA calculations allowed one to discriminate between subgroups of IL-2 PPIMs with low and high affinity. Also, we obtained good enrichments for the individual transient binding pockets in a docking-based, retrospective virtual screening for IL-2 PPIMs. Together with the fact that the known PPIMs of IL-2 were identified to mimic many of the interactions also found in the

IL-2/IL-2R α region, this suggests that current computational methods can assist the knowledge-driven process of PPIM identification when starting from a given protein–protein complex.

■ ASSOCIATED CONTENT

Supporting Information. Tables with the heavy atom RMSD of IL-2 and its complexes during MD simulations (Table S1), heavy atom RMSD of the IL-2 interface region of experimentally determined bound IL-2 conformations (Table S2), the largest pocket volumes for selected FRODA and MD snapshots (Table S3), selected pocket residues of IL-2 for the definition of the potential energy grids (Table S4), the drift of the effective energies during MD simulations (Table S5), and the selection of the best docking poses in the FRODA snapshots with the largest pocket volume (Table S6) as well as graphical representations of the workflow of the PPIAnalyzer method (Figure S1); the RMSF values of IL-2 during MD simulations of the unbound and bound states (Figure S2); the rigid cluster decomposition of unbound IL-2 obtained by FIRST (Figure S3); the overlay of the protein–protein interface region of IL-2 in unbound, bound, and FRODA sampled conformation (Figure S4); the calculated mean absolute effective energies G_{eff} of IL-2 in its unbound and bound conformations (Figure S5); the enrichment of known IL-2 ligands (Figure S6); the property distributions of known IL-2 ligands and decoys (Figure S7); the time series of the effective energies (Figure S8); and the multiple sequence alignment of sequences of IL-2 crystal structures (Figure S9). This information is available free of charge via the Internet at <http://pubs.acs.org/>.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (+49) 211 81-13662. Fax: (+49) 211 81-13847. E-mail: gohlke@uni-duesseldorf.de.

Author Contributions

[§]Both authors contributed equally to this work.

■ ACKNOWLEDGMENT

We are grateful to Teresa Jiménez Vaquero for providing an initial version of the grid-based pocket identification algorithm and to Dr. Simone Fulle for critically reading the manuscript. This work was supported by Sanofi-Aventis Deutschland GmbH (LGCR Drug Design). We are grateful to the “Zentrum fuer Informations- und Medientechnologie” (ZIM) at the Heinrich-Heine-University, Düsseldorf for computational support. We are grateful to OpenEye Scientific Software for granting a no-cost academic license to us. Figures were generated by gnuplot and PyMOL.

■ ABBREVIATIONS:

PPI, protein–protein interaction; PPIM, small-molecule protein–protein interaction modulator; RMSD, root mean-square deviation; RMCD, root mean-cubed difference; MD, molecular dynamics; IL-2, interleukin-2; IL-2R α , α -subunit of the interleukin-2 receptor; GB, generalized Born; PB, Poisson–Boltzmann; MM-PBSA, molecular mechanics Poisson–Boltzmann surface area; MM-GBSA, molecular mechanics generalized Born; DUD, directory of useful

decoys; ROC, receiver operator curve; AUC, area under the curve; EF_X, enrichment factor at X% of the data set.

■ REFERENCES

- (1) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* **2007**, *450* (7172), 1001–1009.
- (2) Jones, S.; Thornton, J. M. Principles of protein–protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (1), 13–20.
- (3) Lo Conte, L.; Chothia, C.; Janin, J. The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **1999**, *285* (5), 2177–2198.
- (4) Clackson, T.; Wells, J. A. A Hot-Spot of Binding-Energy in a Hormone-Receptor Interface. *Science* **1995**, *267* (5196), 383–386.
- (5) Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C. Binding of small molecules to an adaptive protein–protein interface. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (4), 1603–1608.
- (6) Raimundo, B. C.; Oslob, J. D.; Braisted, A. C.; Hyde, J.; McDowell, R. S.; Randal, M.; Waal, N. D.; Wilkinson, J.; Yu, C. H.; Arkin, M. R. Integrating fragment assembly and biophysical methods in the chemical advancement of small-molecule antagonists of IL-2: An approach for inhibiting protein–protein interactions. *J. Med. Chem.* **2004**, *47* (12), 3111–3130.
- (7) Eyrisch, S.; Helms, V. What induces pocket openings on protein surface patches involved in protein–protein interactions? *J. Comput.-Aided Mol. Des.* **2009**, *23* (2), 73–86.
- (8) Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; Joseph, M. K.; Kitada, S.; Korsmeyer, S. J.; Kunzer, A. R.; Letai, A.; Li, C.; Mitten, M. J.; Nettesheim, D. G.; Ng, S.; Nimmer, P. M.; O'Connor, J. M.; Oleksijew, A.; Petros, A. M.; Reed, J. C.; Shen, W.; Tahir, S. K.; Thompson, C. B.; Tomaselli, K. J.; Wang, B.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **2005**, *435* (7042), 677–681.
- (9) Keskin, Z.; Gursoy, A.; Ma, B.; Nussinov, R. Principles of protein–protein interactions: What are the preferred ways for proteins to interact? *Chem. Rev.* **2008**, *108* (4), 1225–1244.
- (10) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot spots-A review of the protein–protein interface determinant amino-acid residues. *Proteins* **2007**, *68* (4), 803–812.
- (11) Ozbabacan, S. E.; Gursoy, A.; Keskin, O.; Nussinov, R. Conformational ensembles, signal transduction and residue hot spots: application to drug discovery. *Curr. Opin. Drug Discovery Dev.* **2010**, *13* (5), 527–537.
- (12) Berg, T. Small-molecule inhibitors of protein–protein interactions. *Curr. Opin. Drug Discovery Dev.* **2008**, *11* (5), 666–674.
- (13) Gonzalez-Ruiz, D.; Gohlke, H. Targeting protein–protein interactions with small molecules: Challenges and perspectives for computational binding epitope detection and ligand finding. *Curr. Med. Chem.* **2006**, *13* (22), 2607–2625.
- (14) Zhong, S.; Macias, A. T.; MacKerell, A. D., Jr. Computational identification of inhibitors of protein–protein interactions. *Curr. Top. Med. Chem.* **2007**, *7* (1), 63–82.
- (15) Thanos, C. D.; DeLano, W. L.; Wells, J. A. Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (42), 15422–15427.
- (16) Wilson, C. G.; Arkin, M. R. Small-molecule inhibitors of IL-2/IL-2R: lessons learned and applied. *Curr. Top. Microbiol. Immunol.* **2011**, *348*, 25–59.
- (17) Braisted, A. C.; Oslob, J. D.; DeLano, W. L.; Hyde, J.; McDowell, R. S.; Waal, N.; Yu, C.; Arkin, M. R.; Raimundo, B. C. Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J. Am. Chem. Soc.* **2003**, *125* (13), 3714–3715.
- (18) Fuller, J. C.; Burgoyne, N. J.; Jackson, R. M. Predicting druggable binding sites at the protein–protein interface. *Drug Discovery Today* **2009**, *14* (3–4), 155–161.

- (19) Brown, S. P.; Hajduk, P. J. Effects of conformational dynamics on predicted protein druggability. *ChemMedChem* **2006**, *1* (1), 70–72.
- (20) Czarna, A.; Beck, B.; Srivastava, S.; Popowicz, G. M.; Wolf, S.; Huang, Y.; Bista, M.; Holak, T. A.; Domling, A. Robust generation of lead compounds for protein-protein interactions by computational and MCR chemistry: p53/Hdm2 antagonists. *Angew. Chem., Int. Ed. Engl.* **2010**, *49* (31), 5352–5356.
- (21) Corradi, V.; Mancini, M.; Manetti, F.; Petta, S.; Santucci, M. A.; Botta, M. Identification of the first non-peptidic small molecule inhibitor of the c-Abl/14–3-3 protein-protein interactions able to drive sensitive and Imatinib-resistant leukemia cells to apoptosis. *Bioorg. Med. Chem. Lett.* **2010**, *20* (20), 6133–6137.
- (22) Gohlke, H.; Kiel, C.; Case, D. A. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes. *J. Mol. Biol.* **2003**, *330* (4), 891–913.
- (23) Gohlke, H.; Kuhn, L. A.; Case, D. A. Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins* **2004**, *56* (2), 322–337.
- (24) Ababou, A.; van der Vaart, A.; Gogonea, V.; Merz, K. M., Jr. Interaction energy decomposition in protein-protein association: a quantum mechanical study of barnase-barstar complex. *Biophys. Chem.* **2007**, *125* (1), 221–236.
- (25) Lafont, V.; Schaefer, M.; Stote, R. H.; Altschuh, D.; Dejaegere, A. Protein-protein recognition and interaction hot spots in an antigen-antibody complex: free energy decomposition identifies "efficient amino acids". *Proteins* **2007**, *67* (2), 418–434.
- (26) Wichmann, C.; Becker, Y.; Chen-Wichmann, L.; Vogel, V.; Vojtkova, A.; Herglotz, J.; Moore, S.; Koch, J.; Lausen, J.; Mantele, W.; Gohlke, H.; Grez, M. Dimer-tetramer transition controls RUNX1/ETO leukemogenic activity. *Blood* **2010**, *116* (4), 603–613.
- (27) Zoete, V.; Michielin, O. Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCR-p-MHC complex. *Proteins* **2007**, *67* (4), 1026–1047.
- (28) Krueger, D. M.; Gohlke, H. DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. *Nucleic Acids Res.* **2010**, *38* (Suppl), W480–W486.
- (29) Cole, D. J.; Skylaris, C. K.; Rajendra, E.; Venkitaraman, A. R.; Payne, M. C. Protein-protein interactions from linear-scaling first-principles quantum-mechanical calculations. *Europhys. Lett.* **2010**, *91* (3), 37004.
- (30) Dastidar, S. G.; Madhumalar, A.; Fuentes, G.; Lane, D. P.; Verma, C. S. Forces mediating protein-protein interactions: a computational study of p53 "approaching" MDM2. *Theor. Chem. Acc.* **2010**, *125* (3–6), 621–635.
- (31) Wong, S.; Amaro, R. E.; McCammon, J. A. MM-PBSA Captures Key Role of Intercalating Water Molecules at a Protein-Protein Interface. *J. Chem. Theory Comput.* **2009**, *5* (2), 422–429.
- (32) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Protein-protein recognition: a computational mutagenesis study of the MDM2-P53 complex. *Theor. Chem. Acc.* **2008**, *120* (4–6), 533–542.
- (33) Cui, Q. Z.; Sulea, T.; Schrag, J. D.; Munger, C.; Hung, M. N.; Naim, M.; Cygler, M.; Purisima, E. O. Molecular dynamics-solvated interaction energy studies of protein-protein interactions: The MP1-p14 scaffolding complex. *J. Mol. Biol.* **2008**, *379* (4), 787–802.
- (34) Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **2004**, *25* (2), 238–250.
- (35) Joce, C.; Stahl, J. A.; Shridhar, M.; Hutchinson, M. R.; Watkins, L. R.; Fedichev, P. O.; Yin, H. Application of a novel in silico high-throughput screen to identify selective inhibitors for protein-protein interactions. *Bioorg. Med. Chem. Lett.* **2010**, *20* (18), 5411–5413.
- (36) Neumann, J.; Gottschalk, K. E. The Effect of Different Force Applications on the Protein-Protein Complex Barnase-Barstar. *Biophys. J.* **2009**, *97* (6), 1687–1699.
- (37) Andrusier, N.; Mashiah, E.; Nussinov, R.; Wolfson, H. J. Principles of flexible protein-protein docking. *Proteins* **2008**, *73* (2), 271–289.
- (38) Chaudhury, S.; Gray, J. J. Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles. *J. Mol. Biol.* **2008**, *381* (4), 1068–1087.
- (39) Chang, C. E. A.; McLaughlin, W. A.; Baron, R.; Wang, W.; McCammon, J. A. Entropic contributions and the influence of the hydrophobic environment in promiscuous protein-protein association. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (21), 7456–7461.
- (40) Enyedy, I. J.; Ling, Y.; Nacro, K.; Tomita, Y.; Wu, X. H.; Cao, Y. Y.; Guo, R. B.; Li, B. H.; Zhu, X. F.; Huang, Y.; Long, Y. Q.; Roller, P. P.; Yang, D. J.; Wang, S. M. Discovery of small-molecule inhibitors of bcl-2 through structure-based computer screening. *J. Med. Chem.* **2001**, *44* (25), 4313–4324.
- (41) Lugovskoy, A. A.; Degterev, A. I.; Fahmy, A. F.; Zhou, P.; Gross, J. D.; Yuan, J. Y.; Wagner, G. A novel approach for characterizing protein ligand complexes: Molecular basis for specificity of small-molecule Bcl-2 inhibitors. *J. Am. Chem. Soc.* **2002**, *124* (7), 1234–1240.
- (42) Koehler, N. K. U.; Yang, C. Y.; Varady, J.; Lu, Y. P.; Wu, X. W.; Liu, M.; Yin, D. X.; Bartels, M.; Xu, B. Y.; Roller, P. P.; Long, Y. Q.; Li, P.; Kattah, M.; Cohn, M. L.; Moran, K.; Tilley, E.; Richert, J. R.; Wang, S. M. Structure-based discovery of nonpeptidic small organic compounds to block the T cell response to myelin basic protein. *J. Med. Chem.* **2004**, *47* (21), 4989–4997.
- (43) Nikolovska-Coleska, Z.; Xu, L.; Hu, Z. J.; Tomita, Y.; Li, P.; Roller, P. P.; Wang, R. X.; Fang, X. L.; Guo, R. B.; Zhang, M. C.; Lippman, M. E.; Yang, D. J.; Wang, S. M. Discovery of embelin as a cell-permeable, small-molecular weight inhibitor of XLAP through structure-based computational screening of a traditional herbal medicine three-dimensional structure database. *J. Med. Chem.* **2004**, *47* (10), 2430–2440.
- (44) Gao, Y.; Dickerson, J. B.; Guo, F.; Zheng, J.; Zheng, Y. Rational design and characterization of a Rac GTPase-specific small molecule inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (20), 7618–7623.
- (45) Fujii, N.; Haresco, J. J.; Novak, K. A. P.; Stokoe, D.; Kuntz, I. D.; Guy, R. K. A selective irreversible inhibitor targeting a PDZ protein interaction domain. *J. Am. Chem. Soc.* **2003**, *125* (40), 12074–12075.
- (46) Debnath, A. K.; Radigan, L.; Jiang, S. B. Structure-based identification of small molecule antiviral compounds targeted to the gp41 core structure of the human immunodeficiency virus type 1. *J. Med. Chem.* **1999**, *42* (17), 3203–3209.
- (47) Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5* (11), 789–796.
- (48) Arkin, M.; Lear, J. D. A new data analysis method to determine binding constants of small molecules to proteins using equilibrium analytical ultracentrifugation with absorption optics. *Anal. Biochem.* **2001**, *299* (1), 98–107.
- (49) Wells, S.; Menor, S.; Hespeneide, B.; Thorpe, M. F. Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.* **2005**, *2* (4), S127–S136.
- (50) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889–897.
- (51) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65* (3), 712–725.
- (52) Lu, Q.; Luo, R. A Poisson-Boltzmann dynamics method with nonperiodic boundary condition. *J. Chem. Phys.* **2003**, *119* (21), 11035–11047.
- (53) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55* (2), 383–394.
- (54) Zoete, V.; Irving, M. B.; Michielin, O. MM-GBSA binding free energy decomposition and T cell receptor engineering. *J. Mol. Recognit.* **2010**, *23* (2), 142–152.

- (55) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23* (1), 128–137.
- (56) Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- (57) de Hoon, M. J. L.; Imoto, S.; Nolan, J.; Miyano, S. Open source clustering software. *Bioinformatics* **2004**, *20* (9), 1453–1454.
- (58) Craig, I. R.; Pfleger, C.; Gohlke, H.; Essex, J. W.; Spiegel, K. Pocket-Space Maps To Identify Novel Binding-Site Conformations in Proteins. *J. Chem. Inf. Model.* **2011**, *51* (10), 2666–2679.
- (59) Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graphics Modell.* **1997**, *15* (6), 359–363.
- (60) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49* (23), 6789–6801.
- (61) Irwin, J. J.; Shoichet, B. K. ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177–182.
- (62) Molinspiration, version 1; Molinspiration Cheminformatics: Slovensky Grob, Slovak Republic, 2008. www.molinspiration.com (accessed Nov. 2011).
- (63) Filter, version 2.0.2; OEChem, version 1.4.2; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2010. www.eyesopen.com (accessed Nov. 2011).
- (64) R: A Language and Environment for Statistical Computing, version 2.6.2; R Development Core Team: Vienna, Austria, 2008.
- (65) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330* (6002), 341–346.
- (66) Rickert, M.; Wang, X. Q.; Boulanger, M. J.; Goriatcheva, N.; Garcia, K. C. The structure of interleukin-2 complexed with its alpha receptor. *Science* **2005**, *308* (5727), 1477–1480.
- (67) Jiang, L.; Kuhlman, B.; Kortemme, T. A.; Baker, D. A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. *Proteins* **2005**, *58* (4), 893–904.
- (68) Mobley, D. L.; Dill, K. A. Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". *Structure* **2009**, *17* (4), 489–498.
- (69) Hu, G. D.; Wang, D. Y.; Liu, X. G.; Zhang, Q. G. A computational analysis of the binding model of MDM2 with inhibitors. *J. Comput.-Aided Mol. Des.* **2010**, *24* (8), 687–697.
- (70) Bradshaw, R. T.; Patel, B. H.; Tate, E. W.; Leatherbarrow, R. J.; Gould, I. R. Comparing experimental and computational alanine scanning techniques for probing a prototypical protein-protein interaction. *Protein Eng., Des. Sel.* **2011**, *24* (1–2), 197–207.
- (71) Hou, T. J.; Yu, R. Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: Mechanism for binding and drug resistance. *J. Med. Chem.* **2007**, *50* (6), 1177–1188.
- (72) Chene, P. Drugs targeting protein-protein interactions. *Chem-MedChem* **2006**, *1* (4), 400–411.
- (73) Li, Y.; Huang, Y.; Swaminathan, C. P.; Smith-Gill, S. J.; Mariuzza, R. A. Magnitude of the hydrophobic effect at central versus peripheral sites in protein-protein interfaces. *Structure* **2005**, *13* (2), 297–307.
- (74) Fulle, S.; Christ, N. A.; Kestner, E.; Gohlke, H. HIV-1 TAR RNA spontaneously undergoes relevant apo-to-holo conformational transitions in molecular dynamics and constrained geometrical simulations. *J. Chem. Inf. Model.* **2010**, *50* (8), 1489–1501.
- (75) Gohlke, H.; Thorpe, M. F. A natural coarse graining for simulating large biomolecular motion. *Biophys. J.* **2006**, *91* (6), 2115–2120.
- (76) Zoete, V.; Meuwly, M.; Karplus, M. Study of the insulin dimerization: Binding free energy calculations and per-residue free energy decomposition. *Proteins* **2005**, *61* (1), 79–93.
- (77) Ballet, T.; Boulange, L.; Brechet, Y.; Bruckert, F.; Weidenhaupt, M. Protein conformational changes induced by adsorption onto material surfaces: an important issue for biomedical applications of material science. *Bull. Pol. Acad. Sci.: Tech. Sci.* **2010**, *58* (2), 303–315.
- (78) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47* (12), 3032–3047.
- (79) Verdonk, M. L.; Mortenson, P. N.; Hall, R. J.; Hartshorn, M. J.; Murray, C. W. Protein-Ligand Docking against Non-Native Protein Conformers. *J. Chem. Inf. Model.* **2008**, *48* (11), 2214–2225.
- (80) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47* (1), 45–55.
- (81) Jacobsson, M.; Karlen, A. Ligand bias of scoring functions in structure-based virtual screening. *J. Chem. Inf. Model.* **2006**, *46* (3), 1334–1343.
- (82) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23* (1–3), 3–25.
- (83) Ahmed, A.; Kazemi, S.; Gohlke, H. Protein flexibility and mobility in structure-based drug design. *Front. Drug Des. Discovery* **2007**, *3*, 455–476.
- (84) Kongsted, J.; Ryde, U. An improved method to predict the entropy term with the MM/PBSA approach. *J. Comput.-Aided Mol. Des.* **2009**, *23* (2), 63–71.
- (85) Ahmed, S.; Metpally, R. P. R.; Sangadala, S.; Reddy, B. V. B. Virtual screening and selection of drug-like compounds to block noggin interaction with bone morphogenetic proteins. *J. Mol. Graphics Modell.* **2010**, *28* (7), 670–682.
- (86) Gohlke, H.; Ahmed, A.; Rippmann, F.; Barnickel, G. A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins. *J. Chem. Inf. Model.* **2011**, *51* (7), 1604–1622.
- (87) Simmerling, C.; Elber, R. Hydrophobic Collapse in a Cyclic Hexapeptide - Computer-Simulations of Chdlfc and Caaaac in Water. *J. Am. Chem. Soc.* **1994**, *116* (6), 2534–2547.

Publication II – Supporting Information

Structure Preparation

Starting structures for the simulations of human IL-2 and its complexes were taken from the Protein Data Bank¹ (PDB codes: 1m47, 1m4c, 1m48, 1m49, 1pw6, 1py2, 1qvn, and 1z92). These structures were modified to achieve consistency with respect to the sequence and number of amino acids. Solvent and buffer molecules were removed except for crystal waters bound to protein chains, which were considered in the MD simulations. Histidine protonation and rotation states were assigned manually such that all IL-2 chains have the same constitution and that histidines can form optimal local interactions. In the case of multiple identical chains, the one with the lowest number of unresolved residues was chosen. Missing residues (Figure S9) were modeled with MODELLER 7v7² using other IL-2 structures as templates, as was done for the Ala69Val mutation in the structure with PDB code 1QVN. The flexible loop between *Ser64* and *Leu100* (all IL-2 α residues are highlighted in italics, whereas all IL-2 residues are depicted in “normal” font) of IL-2 α was not resolved in the crystal structure (PDB code: 1z92).³ As the loop does not contact the binding interface,³ it was not considered any further. This should not influence the structural integrity of IL-2 α during MD simulations because either end of the loop is bound to the residual IL-2 α structure by a disulfide bond. Ligand structures were extracted from the complexes. For docking, the ligands were converted to MOL2 files using the PRODRG2⁴ server. Atom types were corrected manually if necessary. Flexible torsions were determined by AutoTors from the AutoDock suite of programs.⁵

Molecular Dynamics Simulations

MD simulations were performed with the AMBER 9 package of molecular simulation programs⁶ using the Cornell *et al.* force field⁷ with modifications introduced by Hornak *et al.* (ff99SB)⁸ and the general amber force field (GAFF)⁹ for proteins and small molecules, respectively. Partial charges of small molecules were generated according to the RESP procedure.⁹⁻¹⁰ The structures were solvated in a truncated octahedron of TIP3P water¹¹ such that the distance between the edges of the box and the closest solute atom was at least 11 Å. Periodic boundary conditions were applied using the particle mesh Ewald (PME) method¹² to treat long-range electrostatic interactions. Bond lengths involving bonds to hydrogen atoms were constrained by SHAKE.¹³⁻¹⁴ The time step for all MD simulations was 2 fs, and a direct-space non-bonded cutoff of 8 Å was applied. After minimization the system was heated from 100 K to 300 K using canonical ensemble (NVT) MD. Then, the solvent density was adjusted using isothermal-isobaric ensemble (NPT) MD. Positional restraints applied during equilibration were reduced in a stepwise manner over 50 ps followed by 50 ps of unrestrained canonical ensemble (NVT) MD at 300 K with a time constant of 2 ps for heat bath coupling. Snapshots were extracted every 10 ps from production runs for further analysis (Table 1).

Docking

All docking runs were performed with AutoDock 3.05⁵ using DrugScore pair potentials¹⁵ as a scoring function.¹⁶⁻¹⁷ The docking protocol for flexible ligand docking comprised 100 independent runs per ligand using an initial population size of 100 individuals, 5.0×10^3 generations, a maximum number of 10.0×10^6 energy evaluations, a mutation rate of 0.02, a crossover rate of 0.8, and an elitism value of 1. For the enrichment evaluation the maximum number of energy evaluations and the population size were reduced to 3.0×10^6 and 50, respectively. Before calculating the DrugScore potential grids, all structures were aligned to

the x/y -plane of the Cartesian coordinate system such that the rms distance between the interface amino acids and the plane is minimal. By doing so, the potential grids are optimally positioned for the mainly flat interface region of IL-2. The dimensions of the grids were chosen such that the grids extend beyond all hot spots as well as amino acids lining the identified interface pockets by at least 2.5 Å. In the case of *apo*-docking where no transient pocket is available, the same potential grid definition was chosen as for the re-docking approach. We note that this way no information about the known binding modes of the PPIM was considered for setting up the docking. The grid spacing was set to 0.375 Å. Similar docking poses (RMSD < 1 Å) were clustered, and the intermolecular docking energy was calculated. As the final docking result, the ligand pose with the lowest intermolecular docking energy from the largest cluster was chosen. A docking experiment was considered successful when this ligand pose had an RMSD < 2.0 Å to the native pose.

Statistical Significance of MM-PB/SA Results

To investigate the energetics of IL-2/IL-2R α and IL-2/small-molecule complex formation, the MM-PB/SA method was applied to compute effective energies as the sum of gas-phase energies and solvation free energies. Entropic terms resulting from translational, rotational, and vibrational contributions of the solutes were omitted. The gas-phase and solvation free energy values were averaged over 617 – 1379 snapshots (Table 1) taken at 10 ps intervals from the trajectories of the MD simulations. The correlation time for relaxation of effective energy fluctuations was computed to < 10 ps (data not shown), in agreement with related studies.¹⁸ Hence, the extracted snapshots should be uncorrelated, and mean values of binding effective energies computed by the single trajectory MM-PB/SA method can be estimated to within a standard error of the mean (SEM) between 0.13 – 0.37 kcal mol⁻¹ (Table 1).

Time-series of effective energies computed using the MM-PB/SA method are displayed in Figure S8 for snapshots of the unbound solutes and the IL-2/IL-2R α and IL-2/small-molecule complexes. In all cases, significant drifts and fluctuations in the absolute effective energies were found, which demonstrates the sensitivity of these values to conformational details and reflects structural variations throughout the MD trajectories. The observed energy drift (Table S5) depends on the size and conformational complexity of the solutes (Table 1) with IL-2/IL-2R α showing the largest drift (-11.02 kcal mol⁻¹ ns⁻¹), unbound IL-2 and the IL-2/small-molecule complexes showing drifts of -0.56 – -6.03 kcal mol⁻¹ ns⁻¹, and the small molecules showing negligible drifts of -0.30 – 0.20 kcal mol⁻¹ ns⁻¹.

These analyses indicate as to why MM-PB/SA binding effective energies computed by the multiple trajectory method for IL-2/small-molecule complexes do not correlate with experimental results ($R^2 < 0.1$, data not shown). In contrast, in the case of the single trajectory method, binding effective energies show a much smaller drift (-0.63 – 0.74 kcal mol⁻¹ ns⁻¹, Table S5) due to a cancellation of internal energies.¹⁸ These results also provide an explanation as to why differentiating between conformational states of IL-2 based on absolute effective energies is not successful (Figure S8), in addition to the error introduced by neglecting changes in the solute's configurational entropy. As the energy drifts are mainly caused by conformational transitions of the solute that occur, in particular, in modeled regions, loops, and termini, much longer simulation times would be required to obtain mean absolute effective energies that are stable over time. However, even when simulating for up to 10 ns in related studies,¹⁸⁻²³ this problem could not be alleviated, and comparable drifts were observed.

Tables

Table S1: Heavy atom RMSD during MD simulation

PDB code	RMSD ^a				
	Complex	IL-2 ^b	IL-2 interface ^c	Bound ligand ^d	Unbound ligand ^d
1m47	—	3.17 (3.63)	2.88 (3.70)	—	—
1m4c	—	3.16 (3.83)	2.44 (3.13)	—	—
1m48	2.57 (3.08)	2.55 (3.03)	2.42 (3.00)	1.46 (2.02)	2.66 (4.21)
1m49	2.66 (3.06)	2.68 (3.09)	2.13 (2.60)	1.06 (1.89)	2.64 (4.29)
1pw6	2.70 (3.25)	2.72 (3.27)	2.20 (2.86)	0.89 (1.46)	2.01 (4.76)
1py2	2.59 (3.09)	2.61 (3.12)	1.94 (2.36)	1.55 (2.03)	2.92 (4.94)
1qvn	2.88 (3.21)	2.81 (3.17)	2.27 (3.03)	2.20 (3.94)	2.81 (5.52)
1z92	3.50 (4.52)	3.46 (4.32)	2.49 (3.18)	2.98 (4.07)	3.77 (5.73)

^a Mean heavy atom RMSD with respect to the equilibrated structure; in Å. Five N-terminal amino acids of IL-2 were omitted. Maximum RMSD in parentheses.

^b Unbound IL-2 or IL-2 extracted from the MD trajectory of the complex.

^c IL-2 residues Tyr31, Asn33-Lys35, Thr37-Met39, Thr41-Tyr45, Glu60-Glu62, Lys64-Val69, Asn71, Leu72, and Met104-Thr111.

^d Aligned with respect to the ligand.

Table S2: RMSD of the IL-2 interface region of experimentally determined bound IL-2 conformations^a

IL-2 structure	IL-2/FRG	IL-2/CMM	IL-2/FRB	IL-2/FRH	IL-2/FRI
Unbound structure ^b	1.77	1.69	3.08	1.50	1.69
MD ^c	1.83	1.74	2.99	1.74	1.78
FRODA ^c	1.66	1.58	2.98	1.42	1.51

^a All heavy atoms of the interface region (IL-2 residues Tyr31, Asn33-Lys35, Thr37-Met39, Thr41-Tyr45, Glu60-Glu62, Lys64-Val69, Asn71, Leu72, Met104-Thr111) are considered; in Å.

^b RMSD from the unbound IL-2 conformation (PDB code: 1m47).

^c Minimal RMSD obtained from snapshots generated by either MD or FRODA simulation starting from the unbound IL-2 conformation (PDB code: 1m47).

Table S3: Ten largest pocket volumes of selected FRODA and MD snapshots

FRODA snapshot^a	Volume^b	MD snapshot^c	Volume^b
6	159	123	175
117	157	199	182
169	148	215	180
301	145	234	240
418	148	269	181
514	185	498	184
534	188	782	163
657	215	794	180
698	176	843	173
729	156	940	167

^a Consecutive number of the snapshot from a total of 1,000 snapshots uniformly extracted from the 10,000,000 FRODA-generated snapshots starting from the unbound IL-2 structure (PDB code: 1m47).

^b In Å³.

^c Consecutive number of the snapshot from a total of 1,021 snapshots 10 ps apart that were generated by MD starting from the unbound IL-2 structure (PDB code: 1m47).

Table S4: Pocket residues of IL-2 selected for the definition of the potential energy grids

Protein structure	Residues
Crystal structures ^a	Ile28, Tyr31-Tyr45, Cys58, Glu61-Pro65, Glu68-Lys76, Tyr107, Ile114
MD	Tyr31, Pro34-Lys35, Arg38-Met39, Thr41-Tyr45, Glu60-Glu62, Lys64-Val69, Asn71-Leu72, Cys105, Tyr107, Thr111
FRODA	Tyr31, Pro34-Lys35, Thr37-Met39, Thr41-Tyr45, Glu61-Glu62, Lys64-Val69, Asn71-Leu72

^a Pocket residues identified in PDB codes 1m48 (chain A,B), 1m49 (chain A,B), 1pw6 (chain A), 1py2 (chain A,B,C,D), and 1qvn (chain B,C,D).

Table S5: Drift of the effective energy

PDB code	Drift of effective energy ^a					
	Multiple trajectory method ^b			Single trajectory method ^c		Binding effective energy from single trajectory method ^c
	IL-2	ligand	complex	IL-2	ligand	—
1m47	-1.45	—	—	—	—	—
1m4c	-0.56	—	—	—	—	—
1m48	—	0.03	-2.38	-2.33	0.13	0.20
1m49	—	-0.09	-3.08	-3.22	-0.05	0.19
1pw6	—	0.20	-4.55	-4.05	-0.23	-0.27
1py2	—	-0.30	-5.33	-4.96	0.16	-0.63
1qvn	—	-0.04	-6.03	-6.78	0.00	0.74
1z92	—	0.16	-11.02	-8.24	-3.27	0.49

^a In kcal mol⁻¹ ns⁻¹.^b Structures of IL-2, ligand, and complex were generated by separate MD simulations.^c Structures of IL-2 and ligand were extracted from the MD trajectory of the respective complexes.**Table S6:** Selection of poses from docking into FRODA snapshots

Ligand	PDB code ^a	RMSD ^b	Score ^c	Clustered poses ^d	Cluster size ^e
FRG	1m48	2.08	-13.20	213	204
CMM	1m48	2.59	-14.81	335	327
FRB	1pw6	2.48	-12.78	105	94
FRH	1py2	4.30	-15.77	272	78
FRI	1qvn	3.22	-14.92	133	46

^a PDB code of the corresponding IL-2 complex structure.^b RMSD of the ligand pose with the lowest intermolecular docking energy in the largest cluster with respect to the native pose; in Å.^c In kcal mol⁻¹.^d The number of poses out of 1,000 docked poses (10 FRODA simulated structures with largest pocket volume times 100 docking runs) where the ligand's guanidinium group is within 5 Å of the side chain heavy atoms of Glu62 that were subjected to hierarchical complete linkage clustering with R²⁴ with a cluster distance of 5 Å.^e Number of ligand poses in the largest cluster.

Figures

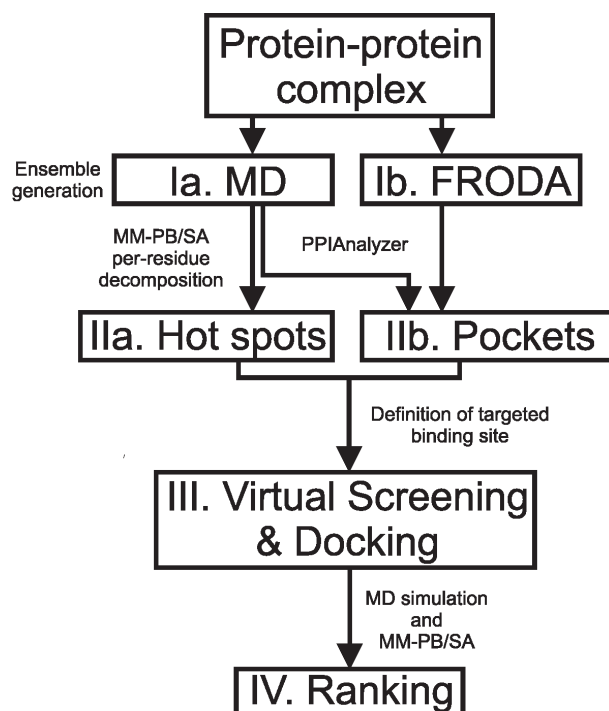


Figure S1: Workflow of the PPIAnalyzer method. The method contains three main steps: I. Analysis of geometrical properties in terms of root mean-square deviations (RMSD) and rotamer analysis. II. Reduction of the dataset by assessing the steric quality of the generated conformations and clustering with respect to the RMSD of heavy atoms of interface residues. III. Identification of transient pockets in the remaining conformations. Representative structures that show the largest interface pocket volume are then selected for the subsequent docking experiments.

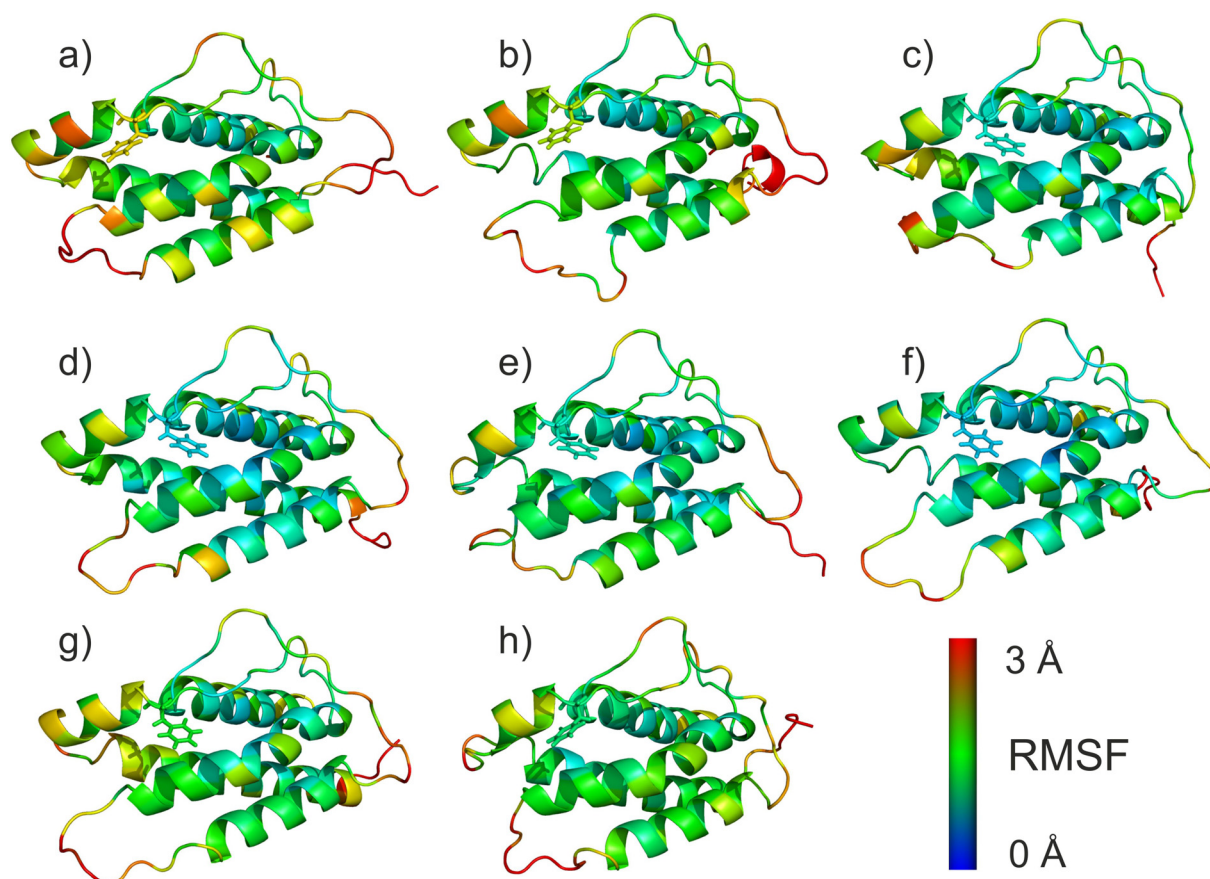


Figure S2: RMSF values of IL-2 residues obtained by MD simulations of the unbound and bound states. The RMSF value of each residue is calculated as the average over all atoms. RMSF values are color-coded onto the respective starting structure of the MD simulations: a) 1m47 and b) 1m4c for unbound IL-2; c) 1m48, d) 1m49, e) 1pw6, f) 1py2, and (g) 1qvn for IL-2 bound to PPIM; h) 1z92 for IL-2 bound to IL-2R α . The RMSF values were calculated for snapshots 10 ps apart. Prior to the RMSF calculations, all snapshots were structurally aligned to the starting structure of the MD simulation considering all heavy protein atoms. The highly mobile five N-terminal residues of IL-2 were neglected in the structural alignment. The protein is depicted in cartoon representation. Phe42 is depicted in stick representation to indicate the location of the small-molecule binding pocket. Figures were generated by PyMOL.²⁵

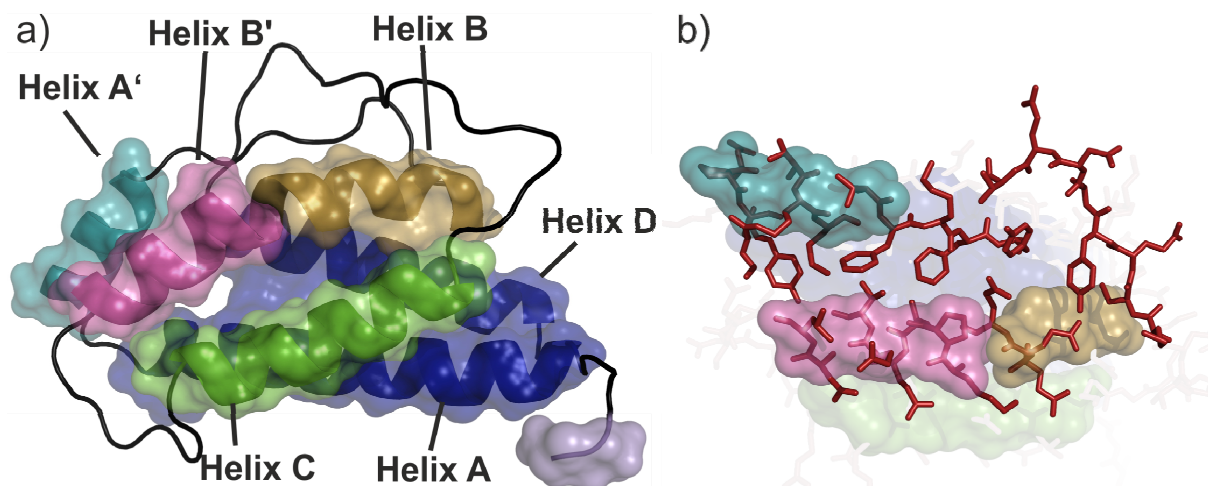


Figure S3: Rigid cluster decomposition obtained by FIRST. (a) The rigid clusters (transparent surfaces) are denominated RC1-6 in the order of decreasing size. RC1 (blue) covers helices A and D, RC2 (green) covers helix C, RC3 (magenta) covers parts of helix B', RC4 (turquoise) covers helix A', RC5 (gold) covers parts of helix B, and RC6 (light blue) is located at the N-terminus of IL-2. (b) 25.5% of all interface atoms are part of the rigid clusters RC3, RC4, and RC5. All flexible atoms (red) can move freely in FRODA simulations. Figures were generated by PyMOL.²⁵

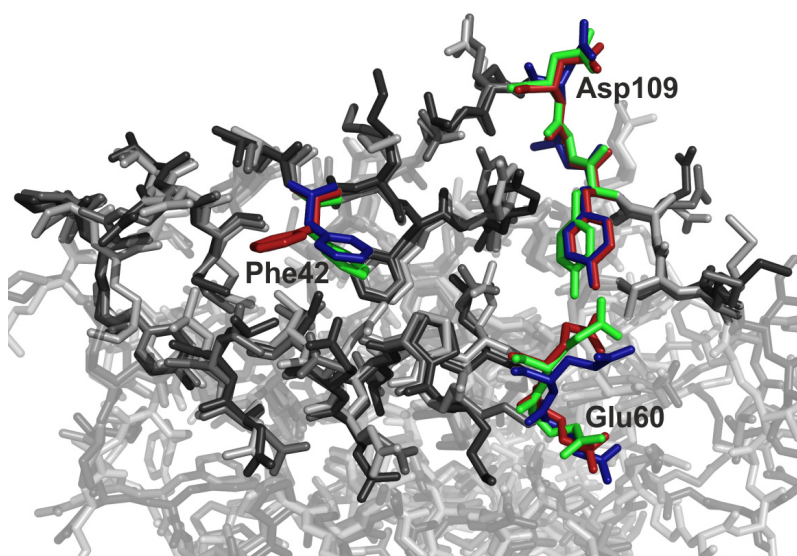


Figure S4: Overlay of the protein/protein interface region of IL-2 in unbound (red) and bound (green) conformation. Exemplarily, one snapshot from a FRODA simulation started from the unbound state is shown (blue), demonstrating that the movement of Phe42 can even be observed in the absence of the ligand, leading to a transient pocket opening. Regions for which no movements were observed by experiment (around Glu60 and Asp109) also remain immobile during the simulation. Figure was generated by PyMOL.²⁵

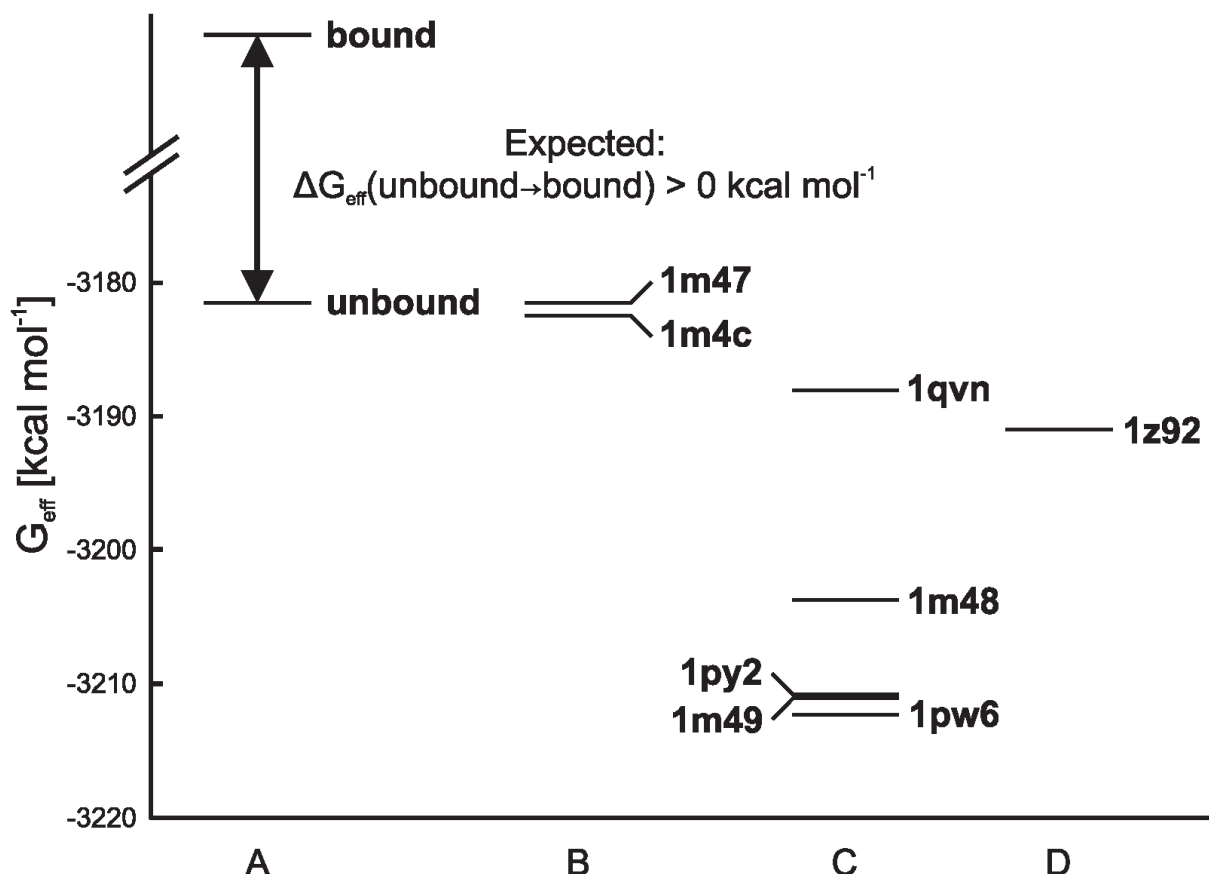


Figure S5: Mean absolute effective energies G_{eff} of IL-2 in its unbound and bound conformations. Conformational stress and changes in solvation and configurational entropy are expected to increase the free energy of a bound conformation over an unbound one (lane A). In contrast, computed G_{eff} of IL-2 extracted from MD trajectories of IL-2/small-molecule complexes (lane C) or from the IL-2/IL-2R α complex (lane D) are lower than G_{eff} of unbound IL-2 (lane B). We attribute this observation to neglecting changes in configurational entropy upon the conformational transitions and the occurrence of significant drifts of G_{eff} over time (see Table S5 and Figure S8). Figure was generated by gnuplot.²⁶

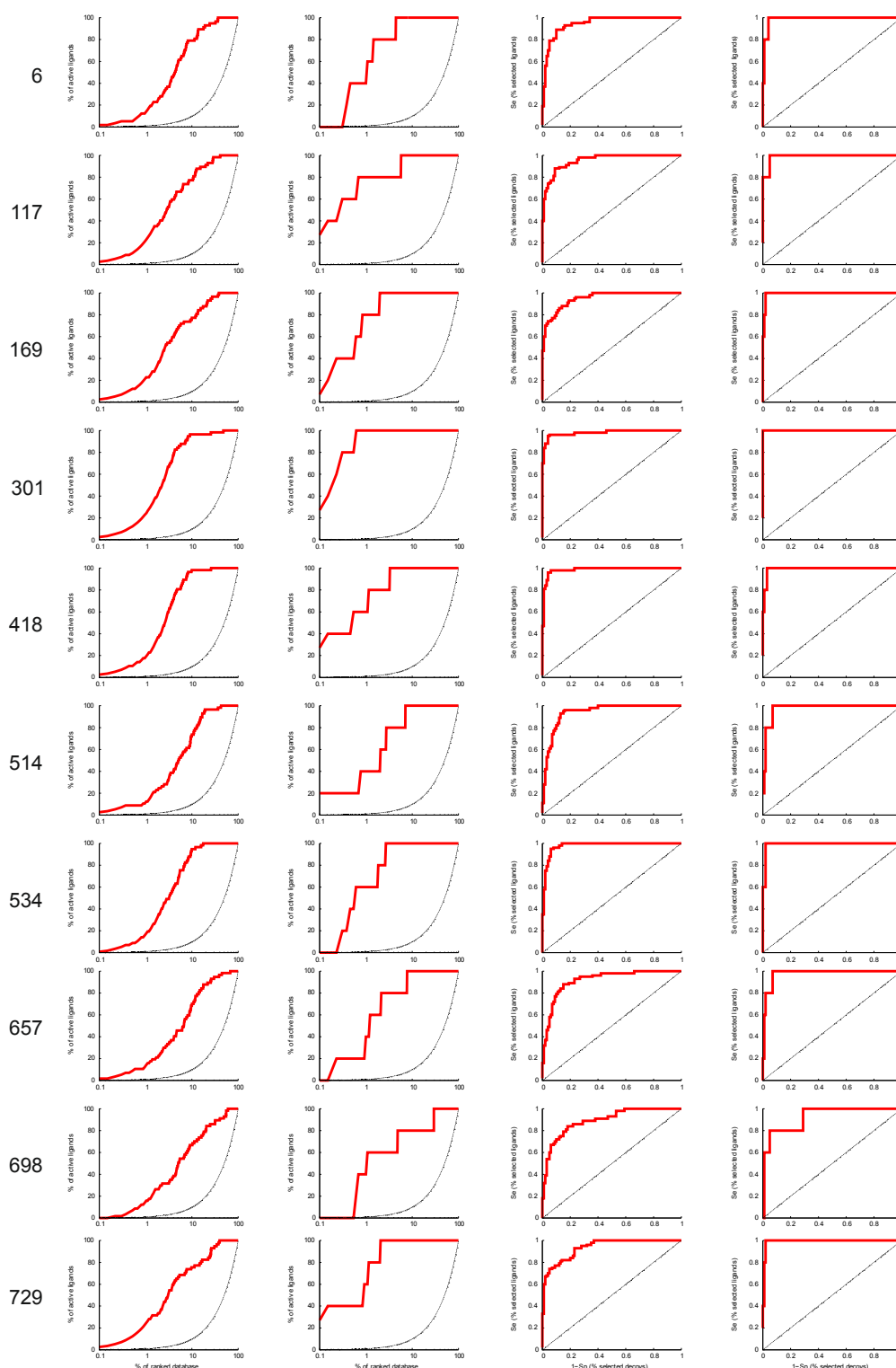


Figure S6: Docking enrichment of known IL-2 ligands. The number of the FRODA snapshot with a transient pocket used for docking is indicated in the left row. Enrichment plots for all 57 IL-2 ligands (1st vertical lane) and the five IL-2 ligands with available complex crystal structure (2nd lane) as well as ROC curves for all 57 IL-2 ligands (3rd lane) and the five IL-2 ligands with available complex crystal structure (4th lane) are given.

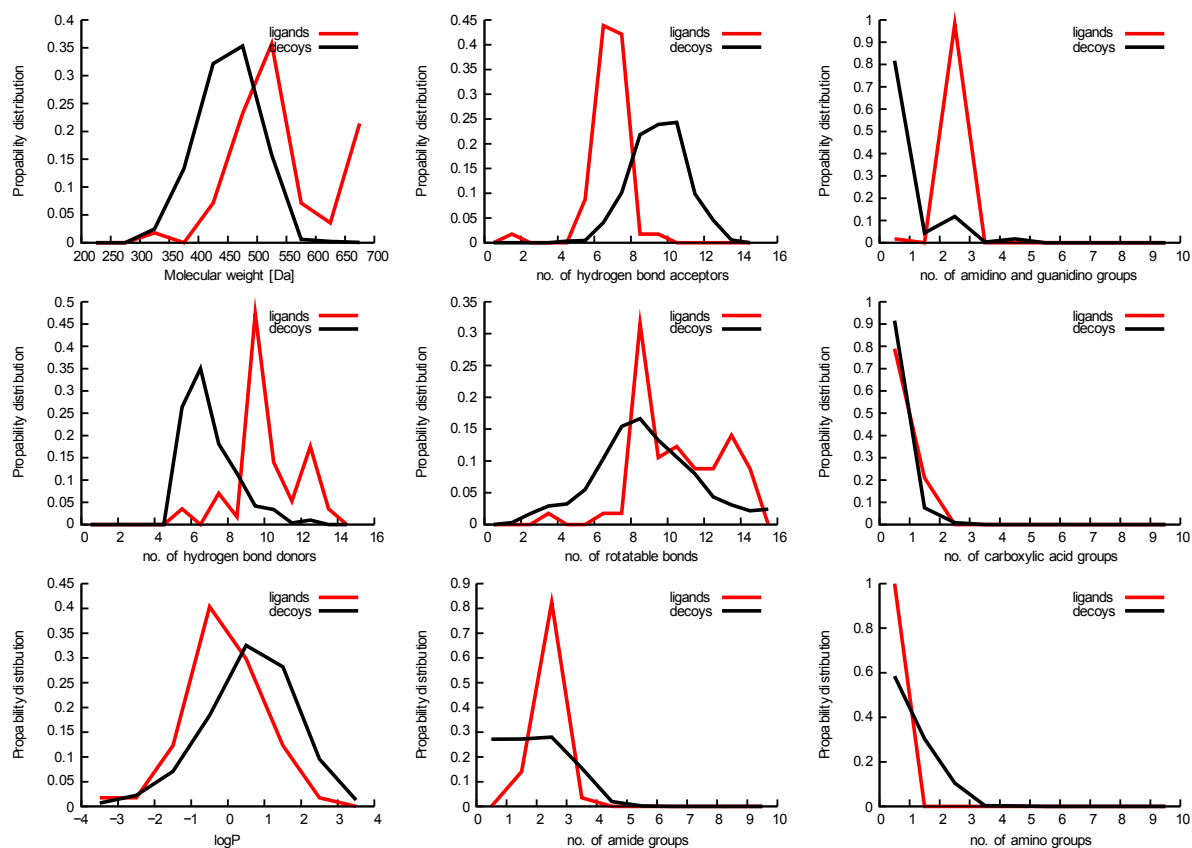


Figure S7: Property distribution of the known IL-2 ligands and decoys. The red line represents all 57 IL-2 ligands. The black line represents the decoy set generated with the aim of similar physicochemical properties to the five IL-2 ligands with available complex crystal structures following the DUD procedure. Figures were generated by gnuplot.²⁶

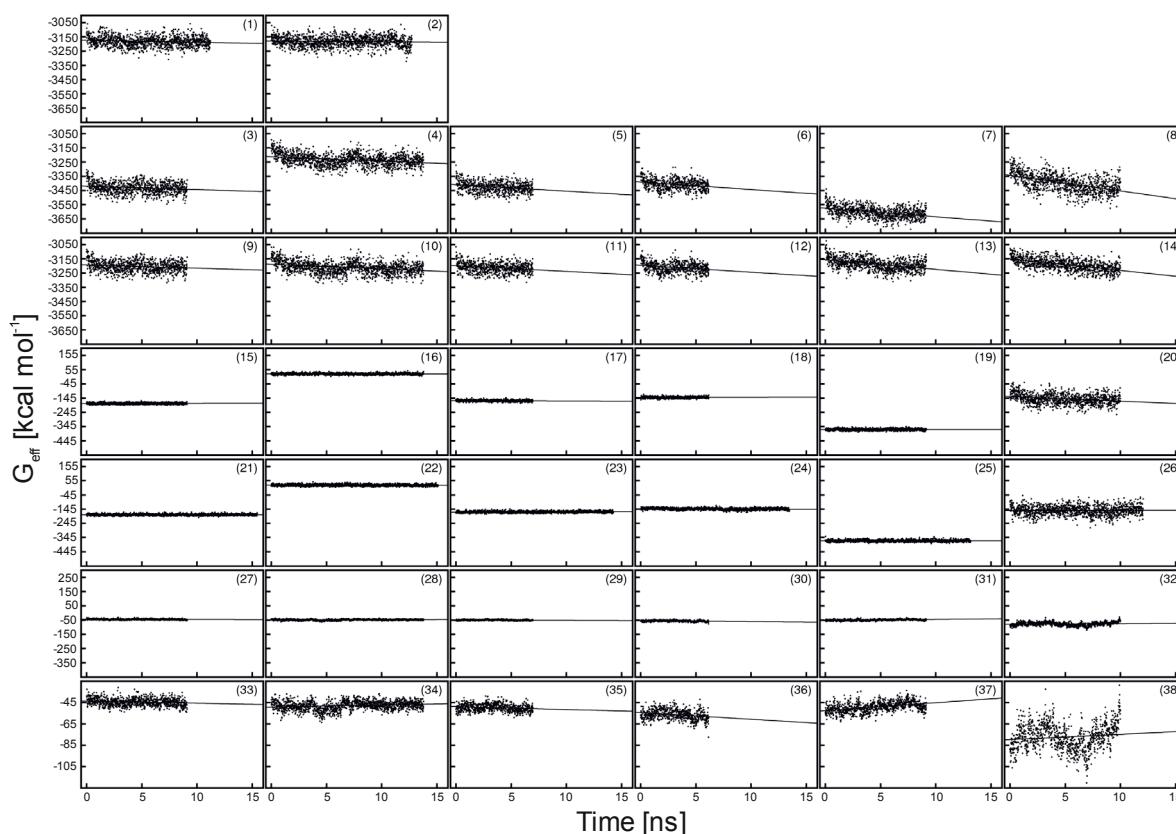


Figure S8: Time series of effective energies. The effective energies were calculated by applying the MM-PB/SA method to snapshots extracted every 10 ps from MD trajectories for: (1) unbound IL-2 [PDB-code: 1m47]; (2) unbound IL-2 [1m4c]; IL-2 in complex with (3) FRG [1m48], (4) CMM [1m49], (5) FRB [1pw6], (6) FRH [1py2], (7) FRI [1qvn], and (8) IL-2R α [1z92]; IL-2 extracted from the trajectories of the complexes of IL-2 with (9) FRG [1m48], (10) CMM [1m49], (11) FRB [1pw6], (12) FRH [1py2], (13) FRI [1qvn], and (14) IL-2R α [1z92]; IL-2 ligands extracted from the trajectories of the complexes of IL-2 with (15) FRG [1m48], (16) CMM [1m49], (17) FRB [1pw6], (18) FRH [1py2], (19) FRI [1qvn], and (20) IL-2R α [1z92]; unbound ligands of IL-2 (21) FRG [1m48], (22) CMM [1m49], (23) FRB [1pw6], (24) FRH [1py2], (25) FRI [1qvn], and (26) IL-2R α [1z92]. In addition, MM-PB/SA single trajectory binding effective energies are depicted for the complexes of IL-2 with (27) FRG [1m48], (28) CMM [1m49], (29) FRB [1pw6], (30) FRH [1py2], (31) FRI [1qvn], and (32) IL-2R α [1z92]. The range of the ordinate values is identical in all plots (1) – (32). For reasons of clarity, MM-PB/SA single trajectory binding effective energies are depicted again with a magnified ordinate scale for the complexes of IL-2 with (33) FRG [1m48], (34) CMM [1m49], (35) FRB [1pw6], (36) FRH [1py2], (37) FRI [1qvn], and (38) IL-2R α [1z92]. Figures were generated by gnuplot.²⁶

```

wt      APTSSSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKIFYMPKKATELKHLQCLEEEELKPLI
1m47    -----STKKTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKIFYMPKKATELKHLQCLEEEELKPLI
1m4C    -----STKKTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKIFYMPKKATELKHLQCLEEEELKPLI
1m48    ---SSSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKIFYMPKKATELKHLQCLEEEELKPLI
1m49    ---SSSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKIFYMPKKATELKHLQCLEEEELKPLI
1pw6    -----SSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKIFYMPKKATELKHLQCLEEEELKPLI
1py2    -----STKKTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKIFYMPKKATELKHLQCLEEEELKPLI
1qvn    ---SSSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKIFYMPKKATELKHLQCLEEEELKPLI
1z92    -----STKKTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKIFYMPKKATELKHLQCLEEEELKPLI
          *****

wt      EVLNLAQSKNFHLRPRDLISNINIVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT
1m47    EVLNLAQ--NFHLRPRDLISNINIVIVLELKG----FMCEYADETATIVEFLNRWITFCQSIISTLT
1m4C    EVLNLA-----RDLISNINIVIVLELKG---FMCEYADETATIVEFLNRWITFCQSIISTLT-
1m48    EVLNLAQSK---NFRDLISNINIVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT
1m49    EVLNLAQ-----RPRDLISNINIVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT-
1pw6    EVLNLAQSKNFHLRPRDLISNINIVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT
1py2    EVLNLAQ-----RPRDLISNINIVIVLELKG-ETTFMCEYADETATIVEFLNRWITFCQSIISTLT-
1qvn    EALNLAQ-----RPRDLISNINIVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT-
1z92    EVLNLA-----RPRDLISNINIVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT
          * . ****          *****          *****

```

Figure S9: Multiple sequence alignment of sequences of IL-2 crystal structures (PDB codes: 1m47, 1m4C, 1m48, 1m49, 1pw6, 1py2, 1qvn, and 1z92). Residues that have not been resolved are indicated by a dash (-) and were modeled using MODELLER 7v7³ to match the full length wild-type sequence (wt). Ala69 of one of the crystal structures (PDB code: 1qvn) was mutated to alanine using MODELLER 7v7 to match the wt sequence. The multiple sequence alignment was created using CLUSTAL-W.²⁷

References

- (1) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235-242.
- (2) Sali, A.; Blundell, T. L., Comparative Protein Modeling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779-815.
- (3) Rickert, M.; Wang, X. Q.; Boulanger, M. J.; Goriatcheva, N.; Garcia, K. C., The structure of interleukin-2 complexed with its alpha receptor. *Science* **2005**, *308* (5727), 1477-1480.
- (4) Schuettelkopf, A. W.; van Aalten, D. M. F., PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1355-1363.
- (5) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19* (14), 1639-1662.
- (6) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A., AMBER 9, University of California, San Francisco. **2006**.
- (7) Wang, J. M.; Cieplak, P.; Kollman, P. A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21* (12), 1049-1074.
- (8) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65* (3), 712-725.
- (9) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157-1174.
- (10) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the Multimolecule and Multiconformational Resp Methodology to Biopolymers - Charge Derivation for DNA, Rna, and Proteins. *J. Comput. Chem.* **1995**, *16* (11), 1357-1377.
- (11) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926-935.
- (12) Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089-10092.
- (13) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327-341.
- (14) Miyamoto, S.; Kollman, P. A., Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952-962.

-
- (15) Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, 295 (2), 337-356.
- (16) Sotriffer, C. A.; Gohlke, H.; Klebe, G., Docking into knowledge-based potential fields: A comparative evaluation of DrugScore. *J. Med. Chem.* **2002**, 45 (10), 1967-1970.
- (17) Pfeffer, P.; Gohlke, H., DrugScore(RNA) - Knowledge-based scoring function to predict RNA-ligand interactions. *J. Chem. Inf. Model.* **2007**, 47 (5), 1868-1876.
- (18) Gohlke, H.; Kiel, C.; Case, D. A., Insights into protein/protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes. *J. Mol. Biol.* **2003**, 330 (4), 891-913.
- (19) Gohlke, H.; Case, D. A., Converging free energy estimates: MM-PB(GB)SA studies on the protein/protein complex Ras-Raf. *J. Comput. Chem.* **2004**, 25 (2), 238-250.
- (20) Cui, Q. Z.; Sulea, T.; Schrag, J. D.; Munger, C.; Hung, M. N.; Naim, M.; Cygler, M.; Purisima, E. O., Molecular dynamics-solvated interaction energy studies of protein/protein interactions: The MP1-p14 scaffolding complex. *J. Mol. Biol.* **2008**, 379 (4), 787-802.
- (21) Deng, N. J.; Cieplak, P., Insights into affinity and specificity in the complexes of alpha-lytic protease and its inhibitor proteins: binding free energy from molecular dynamics simulation. *Phys. Chem. Chem. Phys.* **2009**, 11 (25), 4968-4981.
- (22) Zoete, V.; Michielin, O., Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein/protein association using MM-GBSA: application to the TCR-p-MHC complex. *Proteins* **2007**, 67 (4), 1026-47.
- (23) Tuncel, A.; Kavakli, I. H.; Keskin, O., Insights into subunit interactions in the heterotetrameric structure of potato ADP-glucose pyrophosphorylase. *Biophys. J.* **2008**, 95 (8), 3628-3639.
- (24) R Development Core Team, *A Language and Environment for Statistical Computing*, 2.6.2; Vienna, Austria. **2008**.
- (25) DeLano, W. L. *PyMOL Molecular Graphics System*, 0.99rc6; DeLano Scientific LLC: **2006**.
- (26) Williams T.; Kelley C.; Broeker H.B.; E.A., M., *gnuplot - An Interactive Plotting Program*. **2007**.
- (27) Thompson, J. D.; Higgins, D. G.; Gibson, T. J., Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res.* **1994**, 22 (22), 4673-4680.

Publication III

Modular Solid-Phase Synthesis of Terroxazoles as a Class of α -Helix Mimetics

Cristiano Pinto Gomes,[§] Alexander Metz,[§] Jan W. Bats,
Holger Gohlke, and Michael W. Goebel

European Journal of Organic Chemistry, **2012**; 17: 3270-3277.

[§]Both authors have contributed equally to the respective work.

FULL PAPER



DOI: 10.1002/ejoc.201200339

Modular Solid-Phase Synthesis of Teroxazoles as a Class of α -Helix MimeticsCristiano Pinto Gomes,^{[a]†} Alexander Metz,^{[b]†} Jan W. Bats,^[a] Holger Gohlke,^[b] and Michael W. Göbel^{*[a]}**Keywords:** Protein–protein interactions / Conformation analysis / Molecular dynamics / Peptidomimetics / Helical structures

α -Helices are ubiquitous structural elements of proteins and are important in molecular recognition. Small molecules mimicking α -helices have proven to be valuable biophysical probes or modulators of protein–protein interactions. Here, we present modeling studies and the modular solid-phase synthesis of teroxazole derivatives as a new class of α -helix mimetics. The synthesis is compatible with a variety of functional groups and should thus be generally applicable for

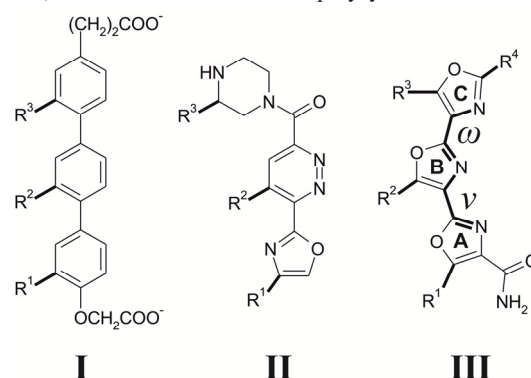
generating diversely substituted oligo-oxazole scaffolds. The teroxazole scaffold is predicted to be polar and to project peptidomimetic side chains at positions *i*, *i*+3, and *i*+6 of an α -helix, which complements projection patterns of existing helix mimetics. The scaffold retains sufficient conformational flexibility to conform to induced-fit models of protein–protein interaction inhibition.

Introduction

α -Helices are ubiquitous structural elements of proteins and are important in molecular recognition.^[1] Accordingly, small molecules mimicking α -helices have proven to be valuable biophysical probes or modulators of protein–protein interactions (PPI).^[2] Typically, only some of the side chains of an α -helix form interaction “hot spots” in PPI.^[3] Imitating these interactions by using α -helix mimetics thus offers a route to the rational development of PPI modulators.^[4]

Initial approaches to mimic α -helices with non-peptidic compounds began more than two decades ago.^[2a,2c,5] Small α -helix mimetics, such as derivatives of allenes, alkylidene cycloalkanes, spiranes, biphenyls,^[6] bicyclic indanes,^[7] and benzodiazepinediones^[8] often mimic α -helix positions that are at close range. A pioneering achievement was the design and synthesis of terphenyl derivatives (Scheme 1, **I**) by the Hamilton group.^[9] Such extended scaffolds are substituted with R groups that mimic the position and orientation of C $_{\alpha}$ –C $_{\beta}$ launch vectors of side chains on one “face” of an α -helix. Terphenyl derivatives have been shown to inhibit Bak/Bcl-X_L^[10] and p53/HDM2^[11] interactions at submicromolar

concentrations.^[9,12] However, terphenyls are rather hydrophobic, which led to the development of other scaffolds that are more hydrophilic and/or amphiphilic including oligo-pyridines,^[13] phenylpyridals,^[14] phenylenaminones,^[15] benzoylureas,^[16] oxazole-pyridazine-piperazines (Scheme 1, **II**) and oxazole-pyrrole-piperazines,^[17] 1,4-dipiperazino benzenes,^[18] 5-6-5 imidazole-phenyl-thiazoles, terphthalimides,^[19] biphenyl 4,4'-dicarboxamides,^[20] oligobenzamides,^[21] and 6/6/6/6 *trans*-fused polycyclic ethers.^[22]



Scheme 1. Terphenyl (**I**), oxazole-pyridazine-piperazine (**II**), and teroxazole (**III**) scaffolds for α -helix mimicry. R¹–R⁴ = CH₃ or peptidomimetic side chain. The inter-ring torsion angles, ν and ω , of **III** are highlighted in bold. Oxazole rings of **III** are labeled **A**, **B**, and **C**.

Here, we describe molecular modeling studies and the modular solid-phase synthesis of teroxazole derivatives (Scheme 1, **III**)^[23] as a new class of α -helix mimetics. The modeling studies suggest that substituted teroxazoles are hydrophilic and preferentially project R groups with launch vectors similar to those of side chains at positions *i*, *i*+3,

[a] Institute of Organic Chemistry and Chemical Biology, Department of Biochemistry, Chemistry, and Pharmacy, Goethe-University, Max-von-Laue-Str. 7, 60438 Frankfurt, Germany
Fax: +49-69-798-29464
E-mail: m.gobel@chemie.uni-frankfurt.de

[b] Institute for Pharmaceutical and Medicinal Chemistry, Department of Mathematics and Natural Sciences, Heinrich-Heine-University, Universitätsstr. 1, 40225 Düsseldorf, Germany

† Both authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/ejoc.201200339>.

and $i+6$ of an α -helix. This suggestion was confirmed by single-crystal structures. This projection pattern complements that of previous scaffolds and covers a broader region on the α -helix surface, which may be advantageous when it comes to mimicking interactions between two helices that wrap around each other and/or are not arranged in a collinear way.

Results and Discussion

Modeling Studies

Computation of relative molecular mechanics Poisson–Boltzmann surface area (MM-PB/SA)^[24] effective energies as a function of inter-ring torsion angles ν and ω reveal that teroxazoles can adopt low-energy conformations with an α -helix-like arrangement of side chains (Scheme 1, III; Figure 1a). The conformation with both torsion angles eclipsed ($\nu = 0^\circ$, $\omega = 0^\circ$) has the lowest energy. Rotating one or both of the torsion angles by 180° increases the effective energy by approximately 2 and 4 kcal mol⁻¹, respectively. The prevalence of the ($\nu = 0^\circ$, $\omega = 0^\circ$) conformation can be explained by a parallel orientation of the ring dipoles leading to a more favorable solvation contribution (data not shown). The rotation around ν or ω is hindered by an energy barrier of ≈ 5.5 kcal mol⁻¹, but nonetheless occurs multiple times during a molecular dynamics (MD) simulation of 250 ns length (Figure 2). The torsion angle distributions from the MD trajectory also reveal significant deviations of ν and ω of up to 50° from a coplanar orientation of the rings. This observation is in good agreement with the MM-PB/SA computations that yield an energetic cost of ≈ 3 kcal mol⁻¹ for such a deviation. The apparent torsional flexibility is expected to enable the teroxazole scaffold to

project peptidomimetic side chains fulfilling distance and angular requirements of an α -helix, and to conform to induced-fit models of protein-protein interaction inhibition.^[15]

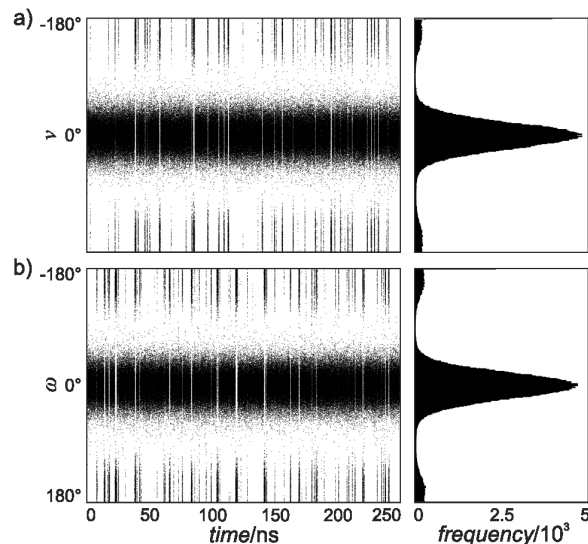


Figure 2. Mutual orientation of teroxazole rings in aqueous solution. The inter-ring torsion angles, ν and ω , of **13a** in explicit solvent were calculated from a MD trajectory of 250 ns length at intervals of 1 ns. Depicted are the inter-ring torsion angles (a) ν and (b) ω as a function of the simulation time (left panels) and in terms of histograms (right panels) with bins of 1° . The inter-ring torsion angles are defined in Scheme 1, III.

Superimposing the substituents R^1 – R^4 of coplanar teroxazole conformations onto C_β atoms of a canonical α -helical octapeptide reveals that the teroxazole scaffold can

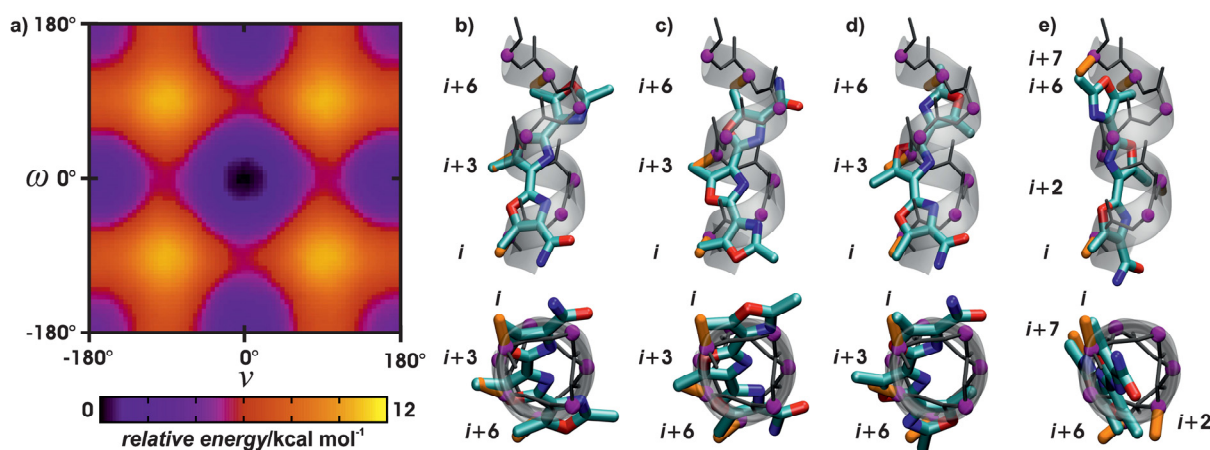


Figure 1. Preferred conformation of a teroxazole scaffold in aqueous solution. (a) Relative MM-PB/SA^[36] effective energies of teroxazole **13a** as a function of the inter-ring torsion angles ν and ω (Scheme 1, III). (b) Superimposition of C_β atoms of side chains at positions i , $i+3$, and $i+6$ of a canonical α -helical octapeptide onto the corresponding substituent atoms of a teroxazole in a low-energy conformation ($\nu = 0^\circ$, $\omega = 0^\circ$). (c) Analogous alignments with the teroxazole scaffold reversed with respect to the α -helix axis. (d) Making use of the substituent R^4 of a rotated ring C ($\nu = 0^\circ$, $\omega = 180^\circ$). (e) If ring A is rotated instead ($\nu = 180^\circ$, $\omega = 0^\circ$), it is possible to align all four substituents R^1 – R^4 with α -helix side chains at positions i , $i+2$, $i+6$, and $i+7$. Below, the α -helix/ α -helix-mimetics superimpositions were rotated by 90° . The N-terminus of the α -helix is oriented towards the viewer. The relative MM-PB/SA effective energies of the conformations of **13a** used for (b) or (c), (d), and (e) are 0.30, 2.48 and 1.96 kcal mol⁻¹, respectively. Graphics by gnuplot^[39] and VMD.^[40]

FULL PAPER

C. Pinto Gomes, A. Metz, J. W. Bats, H. Gohlke, M. W. Göbel

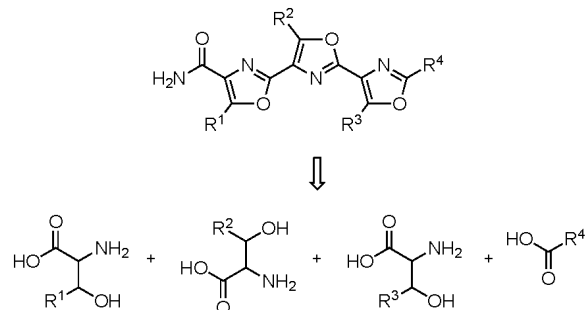
closely mimic the arrangement of peptide side chains (Figure 1b–e). Good agreement is found for the minimum energy conformation ($\nu = 0^\circ$, $\omega = 0^\circ$) with ring A pointing towards the N-terminus of the α -helix, mimicking α -helix positions i , $i+3$, and $i+6$ [root mean square deviation (RMSD) = 0.35 Å; Figure 1b]. Reversing the orientation of the teroxazole scaffold with respect to the α -helix axis does not change the positional agreement between substituent atoms and C_β atoms but impairs the orientational agreement between the respective bonds (Figure 1c). Rotating ring C by 180° ($\nu = 0^\circ$, $\omega = 180^\circ$; Figure 1d) allows the substituent at R^4 to be used. In this conformation the distance between the substituents in ring A or B and ring C is increased, which results in a poorer superimposition with respect to C_β atoms (RMSD = 1.03 Å). Because R^3 and R^4 are located on opposite edges of ring C, it is not possible to align them to amino-acid side chains located on one “face” of an α -helix. Nevertheless, in situations where the mimicked α -helix is deeply buried, one could imagine making use of all four substituents arranged as side chains in positions i , $i+2$, $i+6$, and $i+7$ (RMSD = 0.84 Å, Figure 1e). The side chains now form two pairs [(i , $i+7$) and ($i+2$, $i+6$)], each of which is located on one “face” of an α -helix. Alternatively, if substituent R^3 or R^4 is not required to mimic an important interaction site, it can be used to fine-tune physicochemical and pharmacokinetic properties.

Terphenyl^[10–12] and oxazole-pyridazine-piperazine^[17] derivatives mimic side chain positions i , $i+3/i+4$, and $i+7$. In both cases, the side chains are located on one “face” of an α -helix. On the contrary, the side chain positions i , $i+3$, and $i+6$ addressed by coplanar teroxazole derivatives cover a broader region on the α -helix surface that is rarely addressed by other α -helix mimetics.^[25] This broader region may be advantageous to mimic interactions between two helices that wrap around each other and/or are not arranged in a collinear way. α -Helix mimetics based on terphenyl scaffolds are rather hydrophobic (Scheme 1, **I** with R^1 – $R^4 = \text{CH}_3$; $\log P = 3.34$) that may compromise the water solubility and, hence, the biocompatibility. Among others^[13,19] oxazole-pyridazine-piperazine scaffolds^[17a] have been developed as more polar alternatives to enhance solubility (Scheme 1, **II** with R^1 – $R^4 = \text{CH}_3$; $\log P = 0.96$). Likewise, the teroxazole scaffold introduced here is polar and should offer increased solubility (Scheme 1, **III** with R^1 – $R^4 = \text{CH}_3$; $\log P = 0.34$). These modeling studies suggest teroxazoles as a new class of α -helix mimetics, the R-groups of which point in the directions of side chains of the α -helix that are preferentially located on one face of the helix.

Chemistry

We envisioned a fast and reliable procedure to prepare substituted teroxazoles starting from easily available precursors. Owing to the fact that naturally occurring oligo-oxazoles bearing a C2–C4 linkage pattern are derived from serine containing substrates,^[26] substituents at C5 can be obtained from functionalized β -hydroxy- α -amino acids.

Starting from these precursors, construction of a peptide followed by further oxazole building transformations is expected to result in the desired heterocycles (Scheme 2).



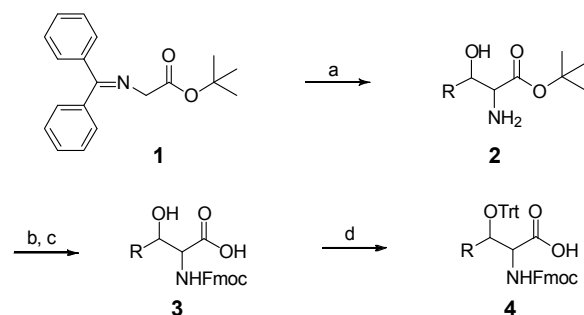
Scheme 2. Retrosynthetic disconnection of the teroxazole scaffold.

Wipf and Miller,^[27] showed that β -hydroxy amides can be efficiently converted to C5-substituted oxazole subunits. The reaction conditions are very mild and therefore compatible with a variety of functional groups [e.g., amides, esters, 9H-fluoren-9-ylmethoxycarbonyl (Fmoc), Boc, Alloc, and Cbz]. Kessler et al.^[28] extended this concept by applying it to solid-phase synthesis to offer a valuable tool for the rapid construction of highly-substituted oxazole moieties. Based on these findings we developed a general solid-phase strategy for the preparation of a teroxazole scaffold. This procedure allows incorporation of structurally different side chains into the framework employing a modular concept. Moreover, the method should be appropriate for use in parallel synthesis. The retrosynthetic approach is shown in Scheme 2. The strategy includes incorporation of four components to introduce structural diversity. Substituted β -hydroxy- α -amino acids and a simple carboxylic acid are required as building blocks for the synthesis. These monomers can be readily connected by using standard amide bond-forming reactions [1-hydroxybenzotriazole (HOBt)/diisopropylcarbodiimide (DIC)], allowing rapid and efficient access to peptidic precursors for further transformations. We chose a solid-phase approach that benefits from the facile purification and reaction work-up owing to the immobilized substrate. The key step involves a modified Gabriel–Robinson cyclization of the peptidic precursor.^[27] The required β -ketoamide is generated by oxidation of the side chain alcohol and subsequently converted into the oxazole by cyclodehydration. Repeating the cycle of peptide coupling, oxidation, and cyclodehydration led to the desired teroxazole compounds.

Synthesis of the Building Blocks

The building blocks were prepared by aldol addition of protected glycine derivative **1**^[29] and corresponding aldehydes to yield β -hydroxy- α -amino acid *tert*-butyl esters **2** as racemic mixtures of diastereomers (Scheme 3). At this stage both diastereomers could be separated by column chromatography for analytical purposes. However, because the newly generated stereocenters are lost in the oxidation-

cyclodehydration sequence to form the resin-bound oxazole moiety, there is no need for a stereoselective aldol reaction or a separation of stereoisomers. Subsequent cleavage of the *tert*-butyl esters under strongly acidic conditions afforded the hydrochloride salt of the amino acids that were subsequently protected resulting in **3**. Finally, the secondary hydroxy groups were converted into the trityl ethers in the presence of in situ generated trityl triflate^[30,31] to provide the orthogonally protected β -hydroxy- α -amino acids **4**.

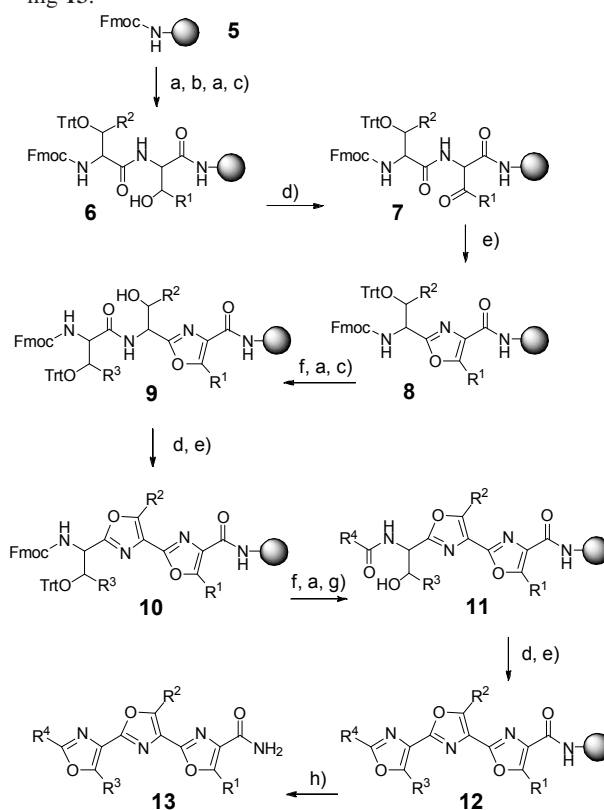


Scheme 3. General synthesis of orthogonally protected β -hydroxy- α -amino acids bearing nonpolar residues: (a) (i) 1 M Lithium hexamethyldisilazide (LHMDS), THF, $-78\text{ }^{\circ}\text{C}$; (ii) RCHO, THF, $-78\text{ }^{\circ}\text{C}$; (iii) 1 M HCl, THF, $0\text{ }^{\circ}\text{C}$. (b) 6 M HCl, reflux. (c) *N*-(9H-Fluoren-9-ylmethoxycarbonyloxy)succinimide, Na_2CO_3 , 1,4-dioxane/water (1.2:1), $0\text{ }^{\circ}\text{C}$ to room temp. (d) TrtCl, AgOTf, 2,6-lutidine, CH_2Cl_2 , $0\text{ }^{\circ}\text{C}$ to room temp. (**2a**: R = $\text{CH}_2\text{CH}_2\text{NHCbz}$; **2b** and **3b**: R = Et; **2c**, **3c**, and **4**: R = *i*Bu; **2d** and **3d**: R = CH_2 -1-naphthyl)

Solid-Phase Synthesis of Terloxazole Derivatives

The synthesis was performed on Rink amide 4-methylbenzhydrylamine hydrochloride salt (MBHA) resin (**5**) at a scale of 0.25 mmol by using standard solid-phase techniques and Fmoc/triphenylmethyl (Trt) synthetic procedures (Scheme 4). The resin-bound Fmoc group was removed with piperidine in dimethylformamide (DMF; 20%, v/v), followed by attachment of Fmoc-protected β -hydroxy- α -amino acids **3** to the solid support in the presence of HOBt/DIC. Treatment of the resin with piperidine and coupling of Fmoc/Trt-protected β -hydroxy- α -amino acids **4** with HOBt/DIC formed the immobilized dipeptides **6**. The corresponding β -ketoamides **7** were obtained by oxidation of the hydroxy group with Dess–Martin periodinane (DMP). Subsequent treatment of the ketone with PPh_3/I_2 afforded oxazole derivatives **8** by a Robinson–Gabriel cyclodehydration.^[27] Cleavage of the trityl ether with 1% trifluoroacetic acid (TFA) in dichloromethane in the presence of a scavenger and removal of the Fmoc group with piperidine were followed by coupling of **4** with HOBt/DIC to provide resin-bound β -hydroxyamides **9**. After oxidation to the β -ketoamide with Dess–Martin periodinane, cyclodehydration with PPh_3/I_2 resulted in resin-bound bioxazole derivatives **10**. Removal of the Fmoc protecting group with piperidine was followed by coupling of the terminal carboxylic acid (HOBt/DIC). After cleavage of the trityl ether with diluted TFA, peptides **11** were subjected to oxidation and

cyclodehydration as before. Resin-bound terloxazole derivatives **12** were finally cleaved from the support under acidic conditions and purified by column chromatography affording **13**.



Scheme 4. Solid-phase synthesis of the terloxazole scaffold: (a) Piperidine/DMF (1:4). (b) Fmoc-protected amino acid **3**, HOBt, DIC, NMP. (c) Fmoc/Trt-protected amino acid **4**, HOBt, DIC, NMP. (d) Dess–Martin periodinane, CH_2Cl_2 . (e) PPh_3 , I_2 , DIPEA, CH_2Cl_2 . (f) TFA/TIPS/ CH_2Cl_2 (1:5:94). (g) Carboxylic acid, HOBt, DIC, NMP or acetic anhydride, DIPEA, NMP. (h) TFA/TIPS/ H_2O (95:2.5:2.5). (**13a**: $\text{R}^1\text{--R}^4$ = Me; **13b**: R^1 = Et, R^2 = *i*Bu, R^3 = Me, R^4 = Bn; **13c**: $\text{R}^1\text{--R}^3$ = *i*Bu)

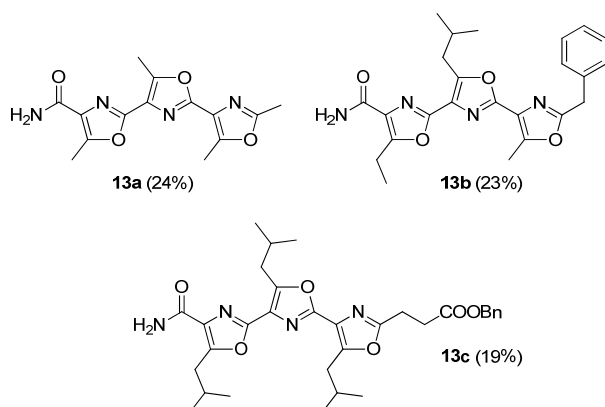
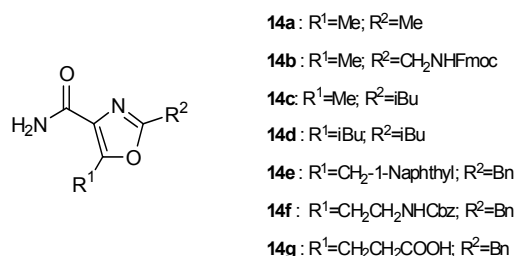
Applying this methodology, terloxazole derivatives **13a–c** (Scheme 5) were synthesized starting from protected β -hydroxy- α -amino acids in overall yields of 19–24%, corresponding to average yields per step above 90%. Solid-phase syntheses are normally conducted on a small scale providing only small quantities of target compounds. However, **13a–c** could be obtained in quantities of around 20 mg, sufficient for most scientific purposes (0.25 mmol of resin-bound NH_2).

We also synthesized a set of monooxazoles bearing different types of side chains and functional groups to demonstrate the structural diversity accessible by our method (Scheme 6). The products obtained comprise alkyl, aryl, basic, and acidic residues.

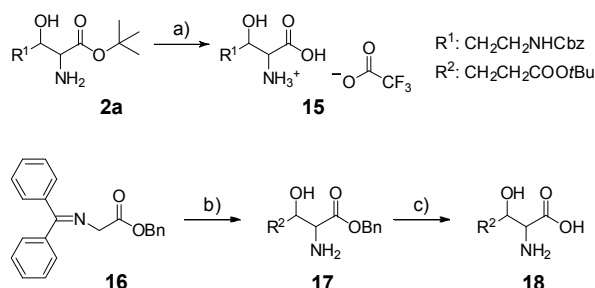
To obtain the precursors of **14f** and **14g**, we had to modify some reaction conditions and protecting groups (Scheme 7). In the case of **14f** it was necessary to use milder

FULL PAPER

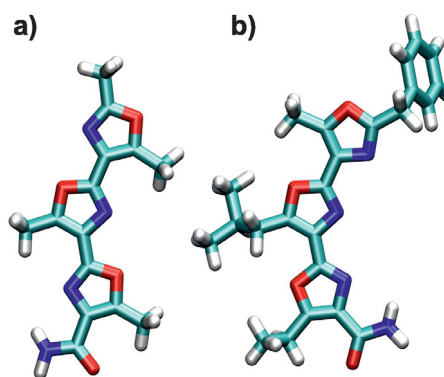
C. Pinto Gomes, A. Metz, J. W. Bats, H. Gohlke, M. W. Göbel

Scheme 5. Teroxazole derivatives **13a–c** synthesized in this study.Scheme 6. Structures of the substituted monooxazoles **14a–g**.

conditions for the cleavage of the *tert*-butyl ester **2a** to prevent premature loss of the Cbz group. The deprotection was achieved with a mixture of TFA/CH₂Cl₂ (1:1) at room temperature affording **15**. The *tert*-butyl ester of O'Donnell imine **1** needed to be replaced during the preparation of **14g** because the carboxyl group at the side chain of the aldehyde already contained this protecting group. When benzyl ester **16** was used in the aldol reaction yielding **17**, subsequent hydrogenolysis at atmospheric pressure removed the benzyl group furnishing free β-hydroxy-α-amino acid **18** that was further protected as shown in Scheme 3.

Scheme 7. Modified preparation of unprotected β-hydroxy-α-amino acids: (a) TFA/CH₂Cl₂ (1:1), room temp. (b) (i) 1 M LHMDS, THF, −78 °C; (ii) R²CHO, THF, −78 °C; (iii) 1 M HCl, THF, 0 °C. (c) H₂, Pd/C, 1 bar, room temp.

Single crystals of **13a** and **13b** suitable for X-ray crystallographic analysis were obtained by slow solvent evaporation (ethyl acetate). These structures confirm the coplanar orientation of the oxazole rings found to be energetically favorable in the modeling studies, with ν/ω values of 178.6/179.4° and 6.9/−6.1° for **13a** and **13b**, respectively (Figure 3).

Figure 3. Single-crystal structures of (a) **13a** and (b) **13b**.

Conclusions

A modular solid-phase synthesis of a series of α-helix mimetics based on a teroxazole scaffold is presented. The synthesis starts from substituted β-hydroxy-α-amino acids, is compatible with a variety of functional groups, and should thus be applicable for generating diversely substituted oligo-oxazole scaffolds. Molecular modeling studies demonstrate that teroxazoles are able to mimic the side chains *i*, *i*+3, and *i*+6 of an α-helix, which cover a broader region on the α-helix surface than peptidomimetic side chains presented by terphenyl or oxazole-pyridazine-piperazine scaffolds. Furthermore, the teroxazole scaffold is found to be more hydrophilic than the terphenyl scaffold and as hydrophilic as the oxazole-pyridazine-piperazine scaffold, which should result in improved solubility and biocompatibility. At present, a series of teroxazoles synthesized with the methods presented above is being evaluated for its potential to act as protein-protein interaction modulators.

Experimental Section

Conformational and Physicochemical Properties: The inter-ring torsion angles ν and ω (Scheme 1, scaffold **III** and Figure 1a) are crucial for the conformational properties of the teroxazole scaffold. Hence, the respective torsion angle potential was parameterized for use with other parameters of the molecular mechanics general AMBER force field (GAFF).^[32] Conformations of the bioxazole derivative 2',5,5'-trimethyl-[2,4'-bioxazole]-4-carboxamide (**28**, see Supporting Information) were optimized at the MP2/6-31G* level by using Gaussian 03,^[33] with the inter-ring torsion angles ν and ω constrained at intervals of 15° over a range of 360°. Likewise, the conformations were optimized – constrained to the same inter-ring torsion angles – by using the GAFF force field within

AMBER 9.^[34] Torsion potential energies for each torsion angle position were obtained by subtracting the GAFF molecular mechanics energy, devoid of the contribution of the inter-ring torsion, from the MP2/6-31G* energy of the corresponding optimized structures. A new molecular mechanics torsion angle potential was then determined by fitting the appropriate GAFF term to these torsion potential energies. This torsion angle potential was used for both inter-ring torsions ν and ω of teroxazole **13a** (Figure 1a with $R^{1-4} = CH_3$) in the calculations below. Partial charges for the bi- and teroxazoles were derived by multiconformational RESP fitting^[35] to the HF/6-31G* electrostatic potentials of the optimized conformations with coplanar rings.

To determine the conformational preferences of **13a**, relative MM-PB/SA^[36] effective energies were calculated for conformations that had been optimized by using the GAFF force field with the inter-ring torsion angles constrained at intervals of 5° over a range of 360° (Figure 1b). Next, minimization, equilibration, and MD simulations of **13a** in explicit solvent was carried out with the AMBER 9 program package by using the GAFF force field and standard procedures (TIP3P water model, PBC, PME, SHAKE, time step of 2 ps). The distribution of torsion angles was analyzed over a MD trajectory of 250 ns length (Figure 2).

$\log P$ values of α -helix mimetics in Scheme 1 were calculated by Molinspiration MiLogP,^[37] which has been shown to have good predictive power (root mean square error = 1.10) in a recent study on a dataset of ≈ 96000 compounds.^[38]

General Procedures for Solid-Phase Reactions

Coupling of Carboxylic Acids or Protected Amino Acids: A solution of the corresponding carboxylic acid or protected amino acid (3 equiv.), HOBt·H₂O (3 equiv.) and DIC (3 equiv.) in *N*-methyl-2-pyrrolidone (NMP; 3 mL) was stirred (10 min) at room temperature and added to the resin-bound amine. The suspension was agitated until the reaction was complete (monitored by the Kaiser test). The resin was washed with NMP (5 \times).

Acylation with Anhydrides: A mixture of the corresponding anhydride (2 \times 5 equiv.) and *N,N*-diisopropylethylamine (DIPEA; 2 \times 5 equiv.) in NMP (5 mL) was added to the resin-bound amine and agitated (2 \times 30 min) at room temperature. After the reaction was complete (monitored by the Kaiser test) the resin was washed with NMP (5 \times).

Removal of Fmoc Protecting Groups: The resin-bound Fmoc-protected amine was treated with 20% piperidine in DMF (1 \times 5 min, 1 \times 20 min, 1 \times 10 min) and washed with NMP (5 \times).

Removal of Triphenylmethyl Protecting Groups: A mixture of TFA/triisopropylsilane (TIPS)/CH₂Cl₂ (1:5:94) was added to the resin-bound Trt-protected alcohol, agitated (2 \times 10 min, 3 \times 5 min) at room temperature and washed with CH₂Cl₂ (5 \times).

Cleavage from the Resin: The resin-bound oxazole derivative was treated with a mixture of TFA/TIPS/H₂O (95:2.5:2.5) at room temperature (2 h). After washing the solid support with CH₂Cl₂ (3 \times), the filtrate and washings were combined and concentrated in vacuo.

Preparation of Teroxazoles

Typical Procedure: 2'-Benzyl-5-ethyl-5'-isobutyl-5''-methyl-[2,4':2',4''-teroxazole]-4-carboxamide (**13b**): Rink amide MBHA resin (391 mg, 0.25 mmol, 0.64 mmol/g loading) was swelled in NMP (1 h), followed by removal of the Fmoc protecting group. After coupling of amino acid **3b** (R = ethyl, mixture of isomers, 267 mg, 0.75 mmol), the Fmoc protecting group was removed and compound **4** (R = isobutyl, mixture of isomers, 469 mg, 0.75 mmol)

was coupled to the resin. Washing with CH₂Cl₂ (5 \times) was followed by treatment (2 h) of the resin with DMP (0.1 M, 7.5 mL, 0.75 mmol) in CH₂Cl₂. The resin was washed with CH₂Cl₂ (5 \times), NMP (5 \times) and CH₂Cl₂ (5 \times), and a solution of PPh₃ (656 mg, 2.50 mmol), iodine (635 mg, 2.50 mmol) and DIPEA (871 μ L, 5.00 mmol) in CH₂Cl₂ (17 mL) was added. After shaking (12 h), the Trt protecting group was cleaved, and the resin was washed with NMP (5 \times). The Fmoc protecting group was removed and Fmoc-*O*-trityl-L-threonine (438 mg, 0.75 mmol) was coupled to the resin-bound amine. The resin was washed with CH₂Cl₂ (5 \times), treated (2 h) with a solution of DMP (0.1 M, 7.5 mL, 0.75 mmol) in CH₂Cl₂ and washed with CH₂Cl₂ (5 \times), NMP (5 \times) and CH₂Cl₂ (5 \times). A solution of PPh₃ (656 mg, 2.50 mmol), iodine (635 mg, 2.50 mmol) and DIPEA (871 μ L, 5.00 mmol) in CH₂Cl₂ (17 mL) was added and agitated (12 h), followed by washing with CH₂Cl₂ (5 \times) and NMP (5 \times). After removal of the Fmoc protecting group, phenylacetic acid (102 mg, 0.75 mmol) was attached to the resin-bound amine. The resin was washed with CH₂Cl₂ (5 \times), the trityl protecting group was removed and a solution of DMP (0.1 M, 7.5 mL, 0.75 mmol) in CH₂Cl₂ was added. The suspension was agitated (2 h), washed with CH₂Cl₂ (5 \times), NMP (5 \times) and CH₂Cl₂ (5 \times) and treated (12 h) with a solution of PPh₃ (656 mg, 2.50 mmol), iodine (635 mg, 2.50 mmol) and DIPEA (871 μ L, 5.00 mmol) in CH₂Cl₂ (17 mL). The resin was washed with CH₂Cl₂ (5 \times), NMP (5 \times) and CH₂Cl₂ (5 \times), followed by cleavage of the resin. The brown residue was purified by column chromatography (silica gel; *n*-hexane/ethyl acetate, 1:1) and recrystallized (cyclohexane), providing **13b** (25 mg, 23%) as a colorless solid. M.p. 185–187 °C. *R*_f = 0.52 (ethyl acetate/*n*-hexane, 9:1). IR (KBr): $\tilde{\nu}$ = 3471 (m), 3350 (w), 3276 (w), 3131 (m), 2957 (m), 2871 (w), 1689 (s), 1649 (m), 1628 (m), 1586 (m), 1551 (w), 1496 (w), 1455 (m), 1419 (m), 1386 (w), 1368 (w), 1350 (w), 1248 (w), 1198 (m), 1161 (w), 1096 (w), 1064 (s), 1034 (s), 985 (w), 946 (w), 801 (w), 773 (w), 729 (m), 696 (m), 669 (w), 567 (w) cm⁻¹. ¹H NMR (250 MHz, CDCl₃): δ = 0.94 [d, *J* = 6.7 Hz, 6 H, (CH₃)₂CH], 1.27 (t, *J* = 7.6 Hz, 3 H, CH₃CH₂), 2.13 [m, 1 H, (CH₃)₂CH], 2.61 (s, 3 H, ArCH₃), 2.95 (d, *J* = 7.2 Hz, 2 H, *i*PrCH₂), 3.12 (q, *J* = 7.6 Hz, 2 H, CH₂CH₃), 4.08 (s, 2 H, CH₂Ph), 5.43 (br. s, 1 H, NH₂), 6.85 (br. s, 1 H, NH₂), 7.17–7.30 (m, 5 H, C₆H₅) ppm. ¹³C NMR (101 MHz, CDCl₃): δ = 11.8, 12.1, 19.4, 22.4, 28.3, 34.5, 34.5, 124.7, 126.1, 127.2, 128.7, 128.7, 128.8, 135.0, 150.8, 153.1, 153.1, 155.3, 158.2, 161.8, 163.8 ppm. C₂₄H₂₆N₄O₄ (434.49): calcd. C 66.34, H 6.03, N 12.89; found C 66.11, H 5.95, N 12.88.

2'',5,5',5''-Tetramethyl-[2,4':2',4''-teroxazole]-4-carboxamide (13a**):** Monomers used: Fmoc-L-threonine, Fmoc-*O*-trityl-L-threonine and acetic anhydride. The reaction gave **13a** as colorless solid (24%). M.p. 254–255 °C from ethyl acetate/*n*-hexane. *R*_f = 0.12 (ethyl acetate/*n*-hexane, 9:1). IR (KBr): $\tilde{\nu}$ = 3455 (m), 3338 (w), 3272 (w), 3190 (m), 2929 (w), 1674 (s), 1648 (m), 1634 (w), 1622 (m), 1592 (m), 1575 (w), 1552 (m), 1429 (m), 1372 (w), 1334 (m), 1307 (w), 1291 (w), 1244 (m), 1206 (m), 1180 (w), 1122 (m), 1053 (s), 983 (m), 945 (w), 936 (w), 799 (w), 768 (w), 745 (w), 735 (w), 696 (w), 680 (m), 630 (w) cm⁻¹. ¹H NMR (250 MHz, CDCl₃): δ = 2.48 (s, 3 H, CH₃), 2.69 (s, 3 H, CH₃), 2.72 (s, 3 H, CH₃), 2.73 (s, 3 H, CH₃), 5.91 (br. s, 1 H, NH₂), 6.97 (br. s, 1 H, NH₂) ppm. ¹³C NMR (63 MHz, CDCl₃): δ = 11.67, 11.69, 11.8, 13.7, 124.2, 125.4, 129.1, 149.9, 150.2, 153.3, 154.1, 155.0, 160.5, 165.0 ppm. C₁₄H₁₄N₄O₄ (302.29): calcd. C 55.63, H 4.67, N 18.53; found C 55.71, H 4.78, N 18.25.

Benzyl 3-(4''-Carbamoyl-5,5',5''-triisobutyl-[2',4:2'',4'-teroxazole]-2-yl)propanoate (13c**):** Monomers used: Fmoc-protected β -hydroxy- α -amino acid **3c** (R = isobutyl, mixture of isomers), Fmoc/Trt-protected β -hydroxy- α -amino acid **4** (R = isobutyl, mixture of isomers)

FULL PAPER

C. Pinto Gomes, A. Metz, J. W. Bats, H. Gohlke, M. W. Göbel

and monobenzyl succinate (**23**). The reaction gave **13c** as a colorless semi-solid (19%). $R_f = 0.42$ (*n*-hexane/ethyl acetate, 1:1). IR (KBr): $\tilde{\nu} = 3471$ (m), 3353 (w), 3279 (w), 3142 (w), 2958 (s), 2926 (m), 2870 (m), 1734 (s), 1686 (s), 1647 (w), 1624 (m), 1588 (m), 1546 (w), 1466 (m), 1430 (w), 1387 (w), 1368 (w), 1340 (w), 1196 (w), 1166 (m), 1095 (w), 1031 (m), 801 (w), 792 (w), 751 (w), 698 (m), 645 (w), 582 (w), 512 (m) cm^{-1} . ^1H NMR (400 MHz, CDCl_3): $\delta = 0.96$ [d, $J = 6.7$ Hz, 6 H, $(\text{CH}_3)_2\text{CH}$], 0.99 [d, $J = 6.5$ Hz, 6 H, $(\text{CH}_3)_2\text{CH}$], 1.00 [d, $J = 6.5$ Hz, 6 H, $(\text{CH}_3)_2\text{CH}$], 2.03–2.22 [m, 3 H, $(\text{CH}_3)_2\text{CH}$], 2.92 (t, $J = 7.6$ Hz, 2 H, CH_2CH_2), 2.94 (d, $J = 7.2$ Hz, 2 H, $i\text{PrCH}_2$), 3.00 (d, $J = 7.1$ Hz, 2 H, $i\text{PrCH}_2$), 3.04 (d, $J = 7.1$ Hz, 2 H, $i\text{PrCH}_2$), 3.14 (t, $J = 7.4$ Hz, 2 H, CH_2CH_2), 5.14 (s, 2 H, CH_2Ph), 5.65 (br. s, 1 H, NH_2), 6.95 (br. s, 1 H, NH_2), 7.28–7.40 (m, 5 H, Ph) ppm. ^{13}C NMR (101 MHz, CDCl_3): $\delta = 22.29, 22.30, 22.4, 23.4, 28.06, 28.07, 28.3, 31.0, 34.3, 34.4, 34.5, 66.6, 125.1, 126.1, 128.2, 128.3, 128.5, 129.9, 135.7, 152.9, 153.2, 153.6, 155.3, 156.5, 162.0, 163.8, 171.6$ ppm. HRMS (MALDI): calcd. for $\text{C}_{32}\text{H}_{40}\text{N}_4\text{O}_6 + \text{H}^+$ [M + H $^+$] 577.3021; found 577.3020.

CCDC-866342 (for **13a**) and -866343 (for **13b**) contain the supplementary crystallographic data for this paper. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

Supporting Information (see footnote on the first page of this article): Synthetic procedures, characterization data for all new compounds, ^1H and ^{13}C NMR spectra and the crystal structure of bi-oxazole **28**.

Acknowledgments

Financial support by the Förderfonds der Goethe-Universität is gratefully acknowledged. H.G. is grateful to the Zentrum für Informations- und Medientechnologie of the Heinrich-Heine-University, Düsseldorf, for computational support and to the Strategischer Forschungsfond of the Heinrich-Heine-University, Düsseldorf, for financial support.

- [1] a) M. P. H. Stumpf, T. Thorne, E. de Silva, R. Stewart, H. J. An, M. Lappe, C. Wiuf, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6959–6964; b) A. G. Cochran, *Curr. Opin. Chem. Biol.* **2001**, *5*, 654–659.
- [2] a) J. M. Davis, L. K. Tsou, A. D. Hamilton, *Chem. Soc. Rev.* **2007**, *36*, 326–334; b) M. R. Arkin, M. Randal, W. L. DeLano, J. Hyde, T. N. Luong, J. D. Oslob, D. R. Raphael, L. Taylor, J. Wang, R. S. McDowell, J. A. Wells, A. C. Braisted, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1603–1608; c) J. A. Wells, C. L. McClendon, *Nature* **2007**, *450*, 1001–1009.
- [3] a) T. Clackson, J. A. Wells, *Science* **1995**, *267*, 383–386; b) M. J. Ramos, I. S. Moreira, P. A. Fernandes, *Proteins Struct., Funct., Bioinf.* **2007**, *68*, 803–812; c) R. Nussinov, Z. J. Hu, B. Y. Ma, H. Wolfson, *Proteins Struct., Funct., Genet.* **2000**, *39*, 331–342; d) A. A. Bogan, K. S. Thorn, *J. Mol. Biol.* **1998**, *280*, 1–9; e) R. Nussinov, S. E. A. Ozbabacan, A. Gursoy, O. Keskin, *Curr. Opin. Drug Discovery Dev.* **2010**, *13*, 527–537; f) A. Metz, C. Pfleger, H. Kopitz, S. Pfeiffer-Marek, K.-H. Baringhaus, H. Gohlke, *J. Chem. Inf. Comput. Sci.* **2012**, *52*, 120–133.
- [4] D. Gonzalez-Ruiz, H. Gohlke, *Curr. Med. Chem.* **2006**, *13*, 2607–2625.
- [5] a) M. W. Pecuh, A. D. Hamilton, *Chem. Rev.* **2000**, *100*, 2479–2493; b) E. Ko, J. Liu, K. Burgess, *Chem. Soc. Rev.* **2011**, *40*, 4411–4421; c) E. Ko, J. Liu, L. M. Perez, G. L. Lu, A. Schaefer, K. Burgess, *J. Am. Chem. Soc.* **2011**, *133*, 462–477.
- [6] E. Jacoby, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 891–893.
- [7] D. C. Horwell, W. Howson, G. S. Ratcliffe, H. M. G. Willems, *Bioorg. Med. Chem.* **1996**, *4*, 33–42.
- [8] B. L. Grasberger, T. B. Lu, C. Schubert, D. J. Parks, T. E. Carver, H. K. Koblish, M. D. Cummings, L. V. LaFrance, K. L. Milkiewicz, R. R. Calvo, D. Maguire, J. Lattanze, C. F. Franks, S. Y. Zhao, K. Ramachandren, G. R. Bylebyl, M. Zhang, C. L. Manthey, E. C. Petrella, M. W. Pantoliano, I. C. Deckman, J. C. Spurlino, A. C. Maroney, B. E. Tomczuk, C. J. Molloy, R. F. Bone, *J. Med. Chem.* **2005**, *48*, 909–912.
- [9] J. T. Ernst, O. Kutzki, A. K. Debnath, S. Jiang, H. Lu, A. D. Hamilton, *Angew. Chem.* **2002**, *114*, 288–291; *Angew. Chem. Int. Ed.* **2002**, *41*, 278–281.
- [10] a) H. Yin, G. I. Lee, K. A. Sedey, O. Kutzki, H. S. Park, B. P. Orner, J. T. Ernst, H. G. Wang, S. M. Sebt, A. D. Hamilton, *J. Am. Chem. Soc.* **2005**, *127*, 10191–10196; b) O. Kutzki, H. S. Park, J. T. Ernst, B. P. Orner, H. Yin, A. D. Hamilton, *J. Am. Chem. Soc.* **2002**, *124*, 11838–11839.
- [11] H. Yin, G. I. Lee, H. S. Park, G. A. Payne, J. M. Rodriguez, S. M. Sebt, A. D. Hamilton, *Angew. Chem.* **2005**, *117*, 2764–2767; *Angew. Chem. Int. Ed.* **2005**, *44*, 2704–2707.
- [12] B. P. Orner, J. T. Ernst, A. D. Hamilton, *J. Am. Chem. Soc.* **2001**, *123*, 5382–5383.
- [13] J. M. Davis, A. Truong, A. D. Hamilton, *Org. Lett.* **2005**, *7*, 5405–5408.
- [14] G. T. Bourne, D. J. Kuster, G. R. Marshall, *Chem. Eur. J.* **2010**, *16*, 8439–8445.
- [15] M. J. Adler, A. D. Hamilton, *J. Org. Chem.* **2011**, *76*, 7040–7047.
- [16] a) J. M. Rodriguez, N. T. Ross, W. P. Katt, D. Dhar, G. I. Lee, A. D. Hamilton, *ChemMedChem* **2009**, *4*, 649–656; b) J. M. Rodriguez, A. D. Hamilton, *Angew. Chem.* **2007**, *119*, 8768–8771; *Angew. Chem. Int. Ed.* **2007**, *46*, 8614–8617.
- [17] a) L. Moisan, S. Odermatt, N. Gombosuren, A. Carella, J. Rebek, *Eur. J. Org. Chem.* **2008**, 1673–1676; b) S. M. Biros, L. Moisan, E. Mann, A. Carella, D. Zhai, J. C. Reed, J. Rebek, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4641–4645.
- [18] P. Maity, B. König, *Org. Lett.* **2008**, *10*, 1473–1476.
- [19] H. Yin, A. D. Hamilton, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1375–1379.
- [20] J. M. Rodriguez, L. Nevola, N. T. Ross, G. I. Lee, A. D. Hamilton, *ChemBioChem* **2009**, *10*, 829–833.
- [21] I. Saraogi, C. D. Incavito, A. D. Hamilton, *Angew. Chem.* **2008**, *120*, 9837–9840; *Angew. Chem. Int. Ed.* **2008**, *47*, 9691–9694.
- [22] H. Oguri, S. Tanabe, A. Oomura, M. Umetsu, M. Hirama, *Tetrahedron Lett.* **2006**, *47*, 5801–5805.
- [23] C. Pinto Gomes, Diploma Thesis, Goethe-University Frankfurt (D), **2007**.
- [24] a) H. Gohlke, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 238–250; b) H. Gohlke, C. Kiel, D. A. Case, *J. Mol. Biol.* **2003**, *330*, 891–913; c) J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, D. A. Case, *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409; d) P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham, *Acc. Chem. Res.* **2000**, *33*, 889–897; e) N. Homeyer, H. Gohlke, *Mol. Informatics* **2012**, *31*, 114–122.
- [25] a) J. S. Albert, M. S. Goodman, A. D. Hamilton, *J. Am. Chem. Soc.* **1995**, *117*, 1143–1144; b) J. S. Albert, M. W. Pecuh, A. D. Hamilton, *Bioorg. Med. Chem.* **1997**, *5*, 1455–1467.
- [26] R. S. Roy, A. M. Gehring, J. C. Milne, P. J. Belshaw, C. T. Walsh, *Nat. Prod. Rep.* **1999**, *16*, 249–263.
- [27] P. Wipf, C. P. Miller, *J. Org. Chem.* **1993**, *58*, 3604–3606.
- [28] E. Biron, J. Chatterjee, H. Kessler, *Org. Lett.* **2006**, *8*, 2417–2420.
- [29] M. J. O'Donnell, R. L. Polt, *J. Org. Chem.* **1982**, *47*, 2663–2666.
- [30] R. M. Burk, T. S. Gac, M. B. Roof, *Tetrahedron Lett.* **1994**, *35*, 8111–8112.
- [31] J. T. Lundquist IV, A. D. Satterfield, J. C. Pelletier, *Org. Lett.* **2006**, *8*, 3915–3918.

- [32] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- [33] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, *Gaussian 03*, Gaussian, Inc., Wallingford, CT, **2004**.
- [34] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- [35] P. Cieplak, W. D. Cornell, C. Bayly, P. A. Kollman, *J. Comput. Chem.* **1995**, *16*, 1357–1377.
- [36] I. Massova, P. A. Kollman, *Perspect. Drug Discovery Des.* **2000**, *18*, 113–135.
- [37] Molinspiration Cheminformatics “Molinspiration” to be found under <http://www.molinspiration.com/cgi-bin/properties> **2008**.
- [38] a) R. Mannhold, G. I. Poda, C. Ostermann, I. V. Tetko, *J. Pharm. Sci.* **2009**, *98*, 861–893; b) R. Mannhold, G. I. Poda, C. Ostermann, I. V. Tetko, *Chem. Biodiversity* **2009**, *6*, 1837–1844.
- [39] T. Williams, C. Kelley, *gnuplot 4.2, Vol. 4.2*, **2009**.
- [40] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.

Received: February 13, 2012
Published Online: May 8, 2012

Publication III – Supporting Information

Eur. J. Org. Chem. **2012** • © WILEY-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2012 • ISSN 1434–193X

SUPPORTING INFORMATION

DOI: 10.1002/ejoc.201200339

Title: Modular Solid-Phase Synthesis of Teroxazoles as a Class of α -Helix Mimetics

Author(s): Cristiano Pinto Gomes, Alexander Metz, Jan W. Bats, Holger Gohlke, Michael W. Göbel*

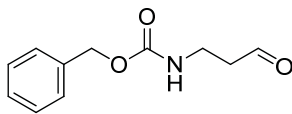
Table of Contents

General remarks	S2
Preparation of β -Hydroxy- α -amino Acid Esters.....	S4
Preparation of Fmoc-protected β -Hydroxy- α -amino Acids.....	S9
Preparation of Fmoc- and Trityl-protected β -Hydroxy- α -amino Acids.....	S14
Preparation of Thioxooxazolidines	S15
Preparation of Free Amino Acids	S18
Preparation of Oxazoles	S19
Preparation of Bioxazole	S22
References	S24
^1H and ^{13}C NMR Spectra.....	S25

General remarks: All NMR spectra were recorded on a Bruker AM 250 (^1H : 250 MHz, ^{13}C : 63 MHz), Bruker AV 300 (^1H : 300 MHz, ^{13}C : 75 MHz) or Bruker AV 400 (^1H : 400 MHz, ^{13}C : 101 MHz). ^1H chemical shifts (δ) are reported in ppm relative to residual protons of CDCl_3 (7.26 ppm), $\text{DMSO}-d_6$ (2.50 ppm) or acetic acid- d_4 (2.04 ppm) as internal standards. Multiplicities are indicated as followed: s (singlet), br s (broad singlet), d (doublet), dd (doublet of doublets), ddd (doublet of doublet of doublets), t (triplet), q (quartet) or m (multiplet). ^{13}C chemical shifts are reported with complete proton decoupling in ppm relative to CDCl_3 (77.00 ppm), $\text{DMSO}-d_6$ (39.43 ppm) or acetic acid- d_4 (20.00 ppm) as an internal standard. FT-IR spectra were recorded on a Perkin-Elmer 1600 Series spectrometer. Peaks are reported in cm^{-1} and intensities are classified as s (strong), m (medium) or w (weak). Elemental analyses were performed on a Heraeus CHN Rapid instrument. HRMS spectra were recorded on a MALDI LTQ Orbitrap XL mass spectrometer from Thermo Fisher Scientific (Bremen, Germany). Melting points (uncorrected) were determined using a Kofler hot plate microscope. Analytical thin layer chromatography (TLC) was performed on alumina plates precoated with silica gel 60 F_{254} indicator (Merck); visualisation by UV light (254 nm) or aqueous KMnO_4 . Flash column chromatography was performed using Merck silica gel (230-400 mesh). All commercially available reagents were used without further purification. All solvents were used after distillation. THF was refluxed over and distilled from sodium/benzophenone. 2,6-lutidine was refluxed over and distilled from CaH_2 . Manual solid-phase synthesis was performed by using 20 mL polyethylene syringe reactors that were equipped with a fritted disc. Rink amide MBHA resin (100-200 mesh) was obtained from Novabiochem (Switzerland).

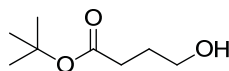
2-(Diphenylmethylene)glycine *tert*-butyl ester (**1**)^[1], 2-(diphenylmethylene)glycine benzyl ester (**16**)^[2], diphenylmethanimine (**19**)^[3], glycine benzyl ester 4-toluenesulfonic acid salt (**20**)^[4], 3-benzyloxycarbonylamino-1-propanol (**21**)^[5], succinic acid mono-*tert*-butyl ester (**22**)^[6] and succinic acid monobenzyl ester (**23**)^[7] were synthesized as described previously.

Abbreviations: DIC: diisopropylcarbodiimide, DIPEA: *N,N*-diisopropylethylamine, Fmoc: 9*H*-fluoren-9-ylmethoxycarbonyl, FmocOSu: *N*-(9*H*-fluoren-9-ylmethoxycarbonyloxy)succinimide, HOBt: 1-hydroxybenzotriazole, LHMDs: lithium hexamethyldisilazide, NMP: *N*-methyl-2-pyrrolidone, Tf: trifluoromethylsulfonyl, Trt: triphenylmethyl.

3-(Benzyloxycarbonylamino)propanal (24)

The oxidation of alcohol **21** was performed using a method similar to a published procedure.^[8] 3-Benzyloxycarbonylamino-1-propanol (**21**) (8.37 g, 40 mmol) and TEMPO (0.19 g, 1.2 mmol) in dichloromethane (80 mL) were added to a solution of NaIO₄ (10.27 g, 48 mmol) and NaBr (0.41 g, 4.0 mmol) in water (100 mL). After stirring for 12 h at rt, the organic layer was separated, washed with conc. Na₂S₂O₃ and dried over MgSO₄. The solvent was removed in vacuo and the residue was recrystallized from Et₂O/*n*-hexane providing **24** (7.23 g, 87%) as a colorless solid.

M.p. 52-54 °C (ref.^[9] 57-58 °C). *R*_f = 0.36 (*n*-hexane/ethyl acetate 1:1). ¹H NMR (300 MHz, CDCl₃): δ = 2.72 (t, *J* = 5.7 Hz, 2H, CH₂CH₂CHO), 3.48 (dt, *J* = 6.0, 6.0 Hz, 2H, CH₂CH₂CHO), 5.08 (s, 2H, OCH₂Ph), 5.21 (s, 1H, NH), 7.27-7.40 (m, 5H, Ph), 9.78 ppm (s, 1H, CHO). ¹³C NMR (75 MHz, CDCl₃): δ = 34.5, 44.0, 66.7, 128.0, 128.1, 128.5, 136.4, 156.3, 201.0 ppm. IR (KBr): $\tilde{\nu}$ = 3321 (s), 3060 (w), 2950 (w), 2900 (w), 2851 (w), 2736 (w), 1712 (s), 1683 (s), 1540 (s), 1454 (w), 1420 (w), 1386 (m), 1310 (m), 1284 (m), 1253 (m), 1142 (m), 967 (w), 905 (w), 842 (w), 785 (w), 752 (m), 723 (m), 696 (m), 648 (w), 576 cm⁻¹ (w). Elemental analysis calcd (%) for C₁₁H₁₃NO₃ (207.23): C 63.76, H 6.32, N 6.76; found: C 63.67, H 6.35, N 6.81.

4-Hydroxybutanoic Acid *tert*-Butyl Ester (25)

The reduction of carboxylic acid **22** was performed similarly to a published procedure.^[10] Succinic acid mono-*tert*-butyl ester (**22**) (10.86 g, 62.3 mmol) and *N*-methylmorpholine (6.9 mL, 62.3 mmol) were dissolved in 1,2-dimethoxyethane (100 mL) and cooled to 0 °C. Afterwards, isobutyl chloroformate (8.1 mL, 62.3 mmol) was added dropwise and the resulting suspension was stirred for 5 min at 0 °C. The reaction mixture was filtered and the residue washed with cold 1,2-dimethoxyethane. The combined filtrates were added slowly to a freshly prepared solution of NaBH₄ (3.54 g, 93.5 mmol) in water (50 mL) at -10 °C and stirred for 30 min. The ice bath was removed and the reaction mixture was stirred for further 30 min. After addition of conc. NaCl (100 mL) the solution was extracted with ethyl acetate, dried over MgSO₄ and concentrated. The residue was filtered over a short column of alumina with dichloromethane/methanol (5:1) and the solvents were removed in vacuo from the filtrate providing **25** (9.50 g, 95%) as a colorless oil.

*R*_f = 0.34 (*n*-hexane/ethyl acetate 1:1). ¹H NMR (300 MHz, CDCl₃): δ = 1.45 (s, 9H, *t*Bu), 1.84 (tt, *J* = 7.1, 7.1 Hz, 2H, HOCH₂CH₂CH₂), 1.87 (br s, 1H, OH), 2.35 (t, *J* = 7.1 Hz, 2H, HOCH₂CH₂CH₂), 3.62-3.73 ppm (m, 2H, HOCH₂CH₂CH₂). ¹³C NMR (75 MHz, CDCl₃): δ = 27.8, 28.0, 32.4, 62.1, 80.5, 173.4 ppm. IR (film): $\tilde{\nu}$ = 3434 (m), 2978 (m), 2934 (m), 2878 (w), 1731 (s), 1479 (w), 1457 (m), 1421

(m), 1393 (m), 1368 (s), 1322 (m), 1255 (m), 1153 (s), 1060 (m), 954 (w), 915 (w), 845 (m), 752 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_8\text{H}_{16}\text{O}_3$ (160.21): C 59.97, H 10.07; found: C 59.78, H 10.16.

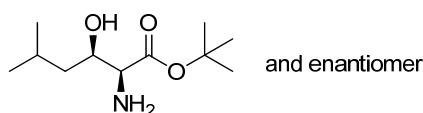
Preparation of β -Hydroxy- α -amino Acid Esters

Typical Procedure for the Aldol reaction.

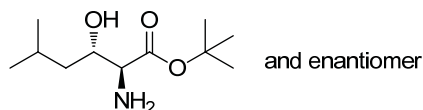
2-Amino-3-hydroxy-5-methylhexanoic Acid *tert*-Butyl Ester (**2c**)

A solution of 2-(diphenylmethylene)glycine *tert*-butyl ester (**1**) (8.06 g, 27 mmol) in THF (80 mL) was added to a solution of LHMDs in THF (1.0 M, 30 mL, 30 mmol) at -78°C . After the yellow mixture was stirred for 30 min, isovaleraldehyde (4.4 mL, 41 mmol) was added dropwise over 30 min. The solution was stirred at -78°C for 2 h, quenched by the addition of saturated aqueous NaHCO_3 (50 mL) and extracted with ethyl acetate (3×60 mL). The combined organic layers were washed with water (50 mL), dried over MgSO_4 , and the solvents were removed in vacuo. The yellow oil was dissolved in THF (100 mL) and cooled to 0°C . To this solution was added aq. HCl (0.5 M, 63 mL, 30 mmol) dropwise over 30 min. After stirring for 1 h at 0°C , the mixture was washed with diethyl ether (4×50 mL), neutralized with saturated aqueous NaHCO_3 and extracted with ethyl acetate (3×60 mL). The combined organic extracts were dried over MgSO_4 , and the solvents were evaporated in vacuo. The light yellow residue was purified by column chromatography (silica gel; ethyl acetate/methanol, 19:1), affording (\pm)-*syn*-**2c** (2.30 g, 39%) and (\pm)-*anti*-**2c** (1.77 g, 30%) as colorless oils.

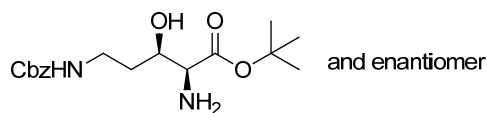
($2S^*,3R^*$)-2-Amino-3-hydroxy-5-methylhexanoic Acid *tert*-Butyl Ester ((\pm)-*syn*-**2c**)



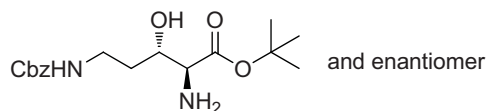
R_f = 0.63 (ethyl acetate/methanol 4:1, 1% NEt_3). ^1H NMR (400 MHz, CDCl_3): δ = 0.87 (d, J = 6.6 Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 0.89 (d, J = 6.7 Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 1.20 (ddd, J = 12.7, 9.0, 3.8 Hz, 1H, $i\text{PrCH}_2$), 1.36–1.46 (m, 1H, $i\text{PrCH}_2$), 1.43 (s, 9H, $t\text{Bu}$), 1.76 (m, 1H, $(\text{CH}_3)_2\text{CH}$), 2.22 (br s, 3H, CHNH_2 , CHOH), 3.14 (d, J = 4.8 Hz, 1H, CHNH_2), 3.72 ppm (ddd, J = 8.8, 4.9, 3.8 Hz, 1H, CHOH). ^{13}C NMR (63 MHz, CDCl_3): δ = 21.9, 23.5, 24.6, 28.0, 43.1, 59.2, 70.2, 81.6, 173.4 ppm. IR (film): $\tilde{\nu}$ = 3377 (m), 2956 (s), 2870 (m), 1591 (m), 1469 (m), 1393 (m), 1368 (m), 1332 (s), 1252 (m), 1156 (m), 1075 (w), 1002 (w), 983 (w), 955 (w), 849 (m), 754 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{11}\text{H}_{23}\text{NO}_3$ (217.31): C 60.80, H 10.67, N 6.45; found: C 60.52, H 10.72, N 6.24. The relative configuration was assigned by comparison with previously reported analytical data of (\pm)-*syn*-**2c**.^[11]

(2S*,3S*)-2-Amino-3-hydroxy-5-methylhexanoic Acid *tert*-Butyl Ester ((±)-*anti*-2c)

R_f = 0.45 (ethyl acetate/methanol 4:1, 1% NEt_3). ^1H NMR (400 MHz, CDCl_3): δ = 0.89 (d, J = 6.8 Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 0.91 (d, J = 6.8 Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 0.97 (ddd, J = 13.7, 9.7, 2.5 Hz, 1H, $i\text{PrCH}_2$), 1.34 (ddd, J = 14.3, 10.5, 4.2 Hz, 1H, $i\text{PrCH}_2$), 1.45 (s, 9H, $t\text{Bu}$), 1.74-1.89 (m, 1H, $(\text{CH}_3)_2\text{CH}$), 2.01 (br s, 3H, CHNH_2 , CHOH), 3.45 (d, J = 4.4 Hz, 1H, CHNH_2), 3.86 ppm (ddd, J = 10.4, 4.2, 2.6 Hz, 1H, CHOH). ^{13}C NMR (63 MHz, CDCl_3): δ = 21.5, 23.8, 24.3, 28.0, 40.9, 59.2, 70.1, 81.6, 173.1 ppm. IR (film): $\tilde{\nu}$ = 3374 (m), 2956 (s), 2870 (m), 1732 (s), 1594 (m), 1469 (m), 1394 (m), 1368 (m), 1251 (m), 1156 (s), 1069 (m), 1008 (w), 850 (m), 785 (w), 750 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{11}\text{H}_{23}\text{NO}_3$ (217.31): C 60.80, H 10.67, N 6.45; found: C 61.02, H 10.69, N 6.55. The relative configuration was assigned by comparison with previously reported analytical data of (±)-*anti*-2c.^[11]

(2S*,3R*)-2-Amino-5-benzyloxycarbonylamino-3-hydroxypentanoic Acid *tert*-Butyl Ester ((±)-*syn*-2a)

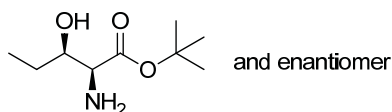
Yield: 39%; colorless oil. R_f = 0.51 (ethyl acetate/methanol 5:1, 1% NEt_3). ^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ = 1.40 (s, 9H, $t\text{Bu}$), 1.50-1.68 (m, 4H, $\text{CH}_2\text{CH}_2\text{NH}$, NH_2), 2.93-3.21 (m, 2H, $\text{CH}_2\text{CH}_2\text{NH}$), 3.05 (d, J = 4.0 Hz, 1H, CHNH_2), 3.66 (m, 1H, CHOH), 4.57 (d, J = 5.2 Hz, 1H, CHOH), 5.00 (m, 2H, OCH_2Ph), 7.23 (dd, J = 5.5, 5.5 Hz, 1H, $\text{CH}_2\text{CH}_2\text{NH}$), 7.27-7.41 ppm (m, 5H, Ph). ^{13}C NMR (75 MHz, $\text{DMSO}-d_6$): δ = 27.6, 33.9, 37.5, 59.3, 65.0, 69.8, 79.7, 127.5, 127.6, 128.2, 137.2, 156.0, 173.4 ppm. IR (film): $\tilde{\nu}$ = 3368 (m), 3034 (w), 2977 (m), 2935 (m), 1722 (s), 1533 (m), 1456 (m), 1394 (w), 1368 (m), 1256 (s), 1156 (s), 1027 (w), 847 (m), 776 (w), 752 cm^{-1} (m). Elemental analysis calcd (%) for $\text{C}_{17}\text{H}_{26}\text{N}_2\text{O}_5$ (338.40): C 60.34, H 7.74, N 8.28; found: C 60.34, H 7.78, N 8.24. The relative configuration was assigned by derivatization to thioxooxazolidine (±)-*trans*-26a.

(2S*,3S*)-2-Amino-5-benzyloxycarbonylamino-3-hydroxypentanoic Acid *tert*-Butyl Ester ((±)-*anti*-2a)

Yield: 18%; colorless oil. R_f = 0.46 (ethyl acetate/methanol 5:1, 1% NEt_3). ^1H NMR (300 MHz, CDCl_3): δ = 1.46 (s, 9H, $t\text{Bu}$), 1.50-1.62 (m, 2H, $\text{CH}_2\text{CH}_2\text{NHCbz}$), 2.31 (br s, 3H, CHOH , CHNH_2), 3.25-3.37

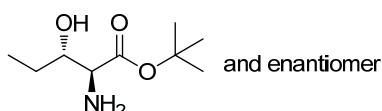
(m, 1H, CH₂CH₂NHCbz), 3.38-3.53 (m, 1H, CH₂CH₂NHCbz), 3.46 (d, *J* = 4.9 Hz, 1H, CHNH₂), 3.85 (m, 1H, CHOH), 5.09 (m, 2H, OCH₂Ph), 5.26 (m, 1H, NHCbz), 7.30-7.37 ppm (m, 5H, Ph). ¹³C NMR (75 MHz, CDCl₃): δ = 28.0, 31.9, 38.3, 58.8, 66.7, 70.5, 82.0, 128.1, 128.5, 136.6, 156.7, 172.7 ppm. IR (film): $\tilde{\nu}$ = 3363 (m), 2977 (m), 2929 (m), 1722 (s), 1532 (m), 1455 (m), 1394 (w), 1369 (m), 1254 (s), 1154 (s), 1084 (w), 1027 (m), 846 (m), 777 (w), 740 (m), 698 cm⁻¹ (m). Elemental analysis calcd (%) for C₁₇H₂₆N₂O₅ (338.40): C 60.34, H 7.74, N 8.28; found: C 60.14, H 7.55, N 8.23. The relative configuration was assigned by derivatization to thioxooxazolidine (\pm)-*cis*-**26a**.

(2*S,3*R**)-2-Amino-3-hydroxypentanoic Acid *tert*-Butyl Ester (\pm)-*syn*-2b)**

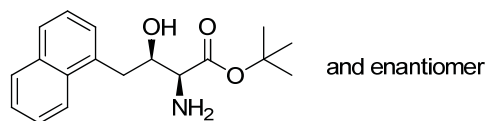


Yield: 27%; colorless oil. *R*_f = 0.46 (ethyl acetate/methanol 4:1, 1% NEt₃). ¹H NMR (300 MHz, DMSO-*d*₆): δ = 0.86 (t, *J* = 7.4 Hz, 3H, CH₃CH₂), 1.32-1.56 (m, 2H, CH₃CH₂), 1.40 (s, 9H, *t*Bu), 1.53 (br s, 2H, CHNH₂), 3.07 (d, *J* = 3.9 Hz, 1H, CHNH₂), 3.51 (m, 1H, CHOH), 4.46 ppm (d, *J* = 6.0 Hz, 1H, CHOH). ¹³C NMR (75 MHz, DMSO-*d*₆): δ = 10.2, 26.5, 27.6, 58.6, 73.3, 79.5, 173.7 ppm. IR (film): $\tilde{\nu}$ = 3375 (m), 2976 (s), 2935 (m), 2878 (m), 1728 (s), 1595 (m), 1459 (m), 1394 (m), 1369 (s), 1284 (m), 1252 (m), 1156 (s), 1084 (w), 1050 (w), 1019 (w), 978 (m), 847 (m), 779 (w), 770 (w), 753 (w), 602 cm⁻¹ (w). Elemental analysis calcd (%) for C₉H₁₉NO₃ (189.25): C 57.12, H 10.12, N 7.40; found: C 56.89, H 9.90, N 7.34. The relative configuration was assigned by comparison with previously reported analytical data of free amino acid (\pm)-*syn*-**27**.^[12]

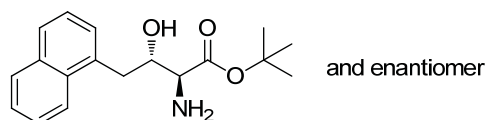
(2*S,3*S**)-2-Amino-3-hydroxypentanoic Acid *tert*-Butyl Ester (\pm)-*anti*-2b)**



Yield: 20%; colorless oil. *R*_f = 0.31 (ethyl acetate/methanol 4:1, 1% NEt₃). ¹H NMR (250 MHz, DMSO-*d*₆): δ = 0.87 (t, *J* = 7.4 Hz, 3H, CH₃CH₂), 1.24-1.49 (m, 2H, CH₃CH₂), 1.40 (s, 9H, *t*Bu), 1.61 (br s, 2H, CHNH₂), 3.12 (d, *J* = 5.3 Hz, 1H, CHNH₂), 3.40 (m, 1H, CHOH), 4.58 ppm (d, *J* = 6.1 Hz, 1H, CHOH). ¹³C NMR (75 MHz, DMSO-*d*₆): δ = 10.2, 25.1, 27.7, 59.7, 74.1, 79.4, 173.1 ppm. IR (film): $\tilde{\nu}$ = 3379 (m), 2976 (s), 2934 (m), 2878 (m), 1732 (s), 1596 (m), 1458 (m), 1394 (m), 1384 (m), 1368 (m), 1310 (w), 1249 (m), 1156 (s), 1048 (m), 980 (m), 948 (w), 912 (w), 849 (m), 787 (w), 751 (w), 714 cm⁻¹ (w). Elemental analysis calcd (%) for C₉H₁₉NO₃ (189.25): C 57.12, H 10.12, N 7.40; found: C 57.24, H 9.98, N 7.43. The relative configuration was assigned by comparison with previously reported analytical data of free amino acid (\pm)-*anti*-**27**.^[12]

(2*S,3*R**)-2-Amino-3-hydroxy-4-(1-naphthyl)butanoic Acid *tert*-Butyl Ester ((±)-*syn*-2d)**

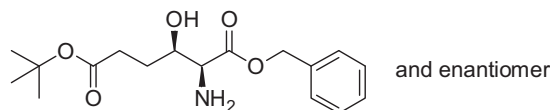
Yield: 23%; colorless semi-solid. R_f = 0.70 (ethyl acetate/methanol 4:1, 1% NEt_3). ^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ = 1.36 (s, 9H, *t*Bu), 1.75 (br s, 2H, CHNH_2), 3.14 (d, J = 3.2 Hz, 1H, CHNH_2), 3.16 (dd, J = 13.5, 7.3 Hz, 1H, CH_2 -1-Naphthyl), 3.37 (dd, J = 13.6, 6.1 Hz, 1H, CH_2 -1-Naphthyl), 4.10 (m, 1H, CHOH), 4.78-4.97 (m, 1H, CHOH), 7.39-7.61 (m, 4H, Ar), 7.78 (m, 1H, Ar), 7.87-7.95 (m, 1H, Ar), 8.13-8.21 ppm (m, 1H, Ar). ^{13}C NMR (75 MHz, $\text{DMSO}-d_6$): δ = 27.6, 37.2, 58.3, 72.3, 79.6, 123.8, 125.29, 125.35, 125.6, 126.4, 127.4, 128.4, 131.8, 133.3, 135.3, 173.7 ppm. IR (KBr): $\tilde{\nu}$ = 3368 (w), 3312 (w), 3069 (w), 2982 (w), 2932 (w), 2866 (w), 2368 (w), 1730 (s), 1654 (w), 1596 (m), 1508 (w), 1475 (w), 1458 (w), 1396 (m), 1367 (m), 1274 (w), 1250 (m), 1156 (s), 1130 (w), 1080 (w), 1056 (w), 1019 (w), 998 (w), 949 (w), 851 (m), 801 (m), 777 (m), 734 (w), 671 (w), 581 (w), 578 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{18}\text{H}_{23}\text{NO}_3$ (301.38): C 71.73, H 7.69, N 4.65; found: C 71.60, H 7.68, N 4.64. The relative configuration was assigned by derivatization to thioxooxazolidine (±)-*trans*-**26c**.

(2*S,3*S**)-2-Amino-3-hydroxy-4-(1-naphthyl)butanoic Acid *tert*-Butyl Ester ((±)-*anti*-2d)**

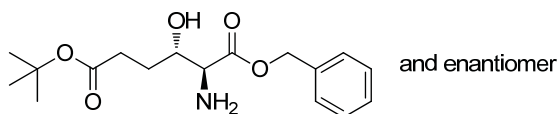
Yield: 35%; colorless oil. R_f = 0.66 (ethyl acetate/methanol 4:1, 1% NEt_3). ^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ = 1.44 (s, 9H, *t*Bu), 1.87 (br s, 2H, CHNH_2), 3.01 (dd, J = 13.9, 9.0 Hz, 1H, ArCH_2), 3.30 (d, J = 5.2 Hz, 1H, CHNH_2), 3.38 (dd, J = 14.0, 3.6 Hz, 1H, ArCH_2), 3.88 (m, 1H, CHOH), 4.83 (d, J = 6.6 Hz, 1H, CHOH), 7.38-7.57 (m, 4H, Ar), 7.78 (dd, J = 6.9, 2.5 Hz, 1H, Ar), 7.86-7.95 (m, 1H, Ar), 8.09-8.16 ppm (m, 1H, Ar). ^{13}C NMR (75 MHz, $\text{DMSO}-d_6$): δ = 27.7, 35.9, 60.2, 73.4, 79.8, 123.8, 125.2, 125.3, 125.4, 126.3, 127.6, 128.4, 131.9, 133.3, 135.5, 173.0 ppm. IR (KBr): $\tilde{\nu}$ = 3375 (w), 3048 (w), 2977 (w), 2932 (w), 1728 (s), 1654 (w), 1597 (w), 1509 (w), 1475 (w), 1458 (w), 1395 (w), 1368 (m), 1250 (m), 1154 (s), 1070 (w), 846 (w), 796 (m), 777 (m), 735 (w), 672 (w), 550 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{18}\text{H}_{23}\text{NO}_3$ (301.38): C 71.73, H 7.69, N 4.65; found: C 71.50, H 7.95, N 4.61. The relative configuration was assigned by derivatization to thioxooxazolidine (±)-*cis*-**26c**.

Preparation of 2-Amino-3-hydroxyhexanedioic Acid 1-Benzyl-6-*tert*-butyl Ester **17**

4-Hydroxybutanoic acid *tert*-butyl ester **25** was oxidized to the aldehyde as described for **24**. The subsequent aldol reaction was performed according to a procedure mentioned earlier (see preparation of **2c**) using benzyl ester **16** instead of *tert*-butyl ester **1**.

(2*S,3*R**)-2-Amino-3-hydroxyhexanedioic Acid 1-Benzyl-6-*tert*-butyl Ester ((±)-*syn*-17)**

Yield: 25%; colorless oil. R_f = 0.60 (ethyl acetate/methanol 4:1, 1% NEt_3). ^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ = 1.37 (s, 9H, *t*Bu), 1.61-1.74 (m, 4H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$, CHNH_2), 2.24 (m, 2H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 3.31 (d, J = 3.2 Hz, 1H, CHNH_2), 3.70-3.84 (m, 1H, CHOH), 4.75 (d, J = 6.3 Hz, 1H, CHOH), 5.11 (s, 2H, CH_2Ph), 7.27-7.43 ppm (m, 5H, Ph). ^{13}C NMR (75 MHz, $\text{DMSO}-d_6$): δ = 27.7, 28.2, 31.2, 59.7, 65.3, 71.8, 79.2, 127.7, 127.8, 128.2, 136.1, 172.3, 173.9 ppm. IR (film): $\tilde{\nu}$ = 3386 (m), 3066 (w), 3034 (w), 2977 (m), 2933 (m), 1728 (s), 1588 (w), 1498 (w), 1456 (m), 1392 (m), 1368 (m), 1257 (m), 1215 (w), 1153 (s), 1082 (w), 1029 (w), 952 (w), 848 (m), 752 (m), 699 cm^{-1} (m). Elemental analysis calcd (%) for $\text{C}_{17}\text{H}_{25}\text{NO}_5$ (323.38): C 63.14, H 7.79, N 4.33; found: C 63.09, H 8.02, N 4.26. The relative configuration was assigned by derivatization to thioxooxazolidine (±)-*anti*-26b.

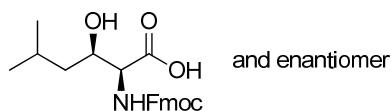
(2*S,3*S**)-2-Amino-3-hydroxyhexanedioic Acid 1-Benzyl-6-*tert*-butyl Ester ((±)-*anti*-17)**

Yield: 20%; colorless oil. R_f = 0.52 (ethyl acetate/methanol 4:1, 1% NEt_3). ^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ = 1.38 (s, 9H, *t*Bu), 1.48-1.61 (m, 1H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$, CHNH_2), 1.67-1.80 (m, 1H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$, CHNH_2), 1.80 (br s, 2H, CHNH_2), 2.13-2.36 (m, 2H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 3.27 (d, J = 6.1 Hz, 1H, CHNH_2), 3.43-3.60 (m, 1H, CHOH), 4.88 (d, J = 5.0 Hz, 1H, CHOH), 5.11 (s, 2H, CH_2Ph), 7.27-7.46 ppm (m, 5H, Ph). ^{13}C NMR (75 MHz, $\text{DMSO}-d_6$): δ = 27.6, 29.1, 31.4, 58.5, 65.5, 70.7, 79.3, 127.6, 127.8, 128.3, 136.1, 172.2, 174.3 ppm. IR (film): $\tilde{\nu}$ = 3373 (m), 3066 (w), 3034 (w), 2977 (m), 2933 (m), 1729 (s), 1588 (w), 1498 (w), 1456 (m), 1420 (w), 1392 (m), 1367 (m), 1257 (m), 1215 (w), 1153 (s), 1074 (w), 953 (w), 848 (m), 753 (m), 698 cm^{-1} (m). Elemental analysis calcd (%) for $\text{C}_{17}\text{H}_{25}\text{NO}_5$ (323.38): C 63.14, H 7.79, N 4.33; found: C 63.36, H 8.02, N 4.31. The relative configuration was assigned by derivatization to thioxooxazolidine (±)-*syn*-26b.

Preparation of Fmoc-protected β -Hydroxy- α -amino Acids.

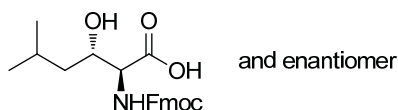
Typical Procedure for the Fmoc-protection.

(2*S**,3*R**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-5-methylhexanoic Acid ((\pm)-*syn*-**3c**)

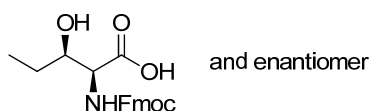


To (2*S**,3*R**)-2-amino-3-hydroxy-5-methylhexanoic acid *tert*-butyl ester **2c** (0.85 g, 3.9 mmol) was added aq. HCl (6 M, 6 mL) and the mixture was heated for 2 h under reflux. After evaporation of the solvent under reduced pressure, the residue was dissolved in water (5 mL) and lyophilized, affording a colorless solid in quantitative yield. The crude product was used without further purification. The resulting solid was added to a solution of Na₂CO₃ (0.62 g, 5.9 mmol) in water (20 mL) and cooled to 0 °C. A solution of FmocOSu (1.32 g, 3.9 mmol) in 1,4-dioxane (15 mL) was added dropwise over 30 min. The suspension was stirred for 1 h at 0 °C and for 20 h at room temperature. The resulting solution was acidified by the addition of HCl (6 M, 20 mL) and extracted with ethyl acetate (2 × 50 mL). The combined organic layers were washed with water (3 × 30 mL) and dried over MgSO₄. After evaporation of the solvent under reduced pressure, the residue was recrystallized (ethyl acetate/*n*-hexane), providing (\pm)-*syn*-**3c** (1.38 g, 92%) as a colorless solid.

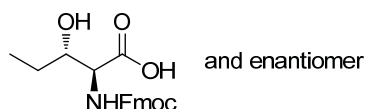
M.p. 161-163 °C. *R*_f = 0.45 (ethyl acetate/*n*-hexane 5:1, 1% HOAc). ¹H NMR (300 MHz, DMSO-*d*₆): δ = 0.85 (d, *J* = 6.5 Hz, 3H, (CH₃)₂CH), 0.88 (d, *J* = 6.7 Hz, 3H, (CH₃)₂CH), 1.18 (ddd, *J* = 13.0, 7.9, 4.7 Hz, 1H, *i*PrCH₂), 1.38 (ddd, *J* = 14.5, 8.7, 5.9 Hz, 1H, *i*PrCH₂), 1.55-1.75 (m, 1H, (CH₃)₂CH), 3.93-4.07 (m, 2H, CHNH, CHOH), 4.18-4.40 (m, 3H, OCH₂CH, OCH₂CH), 4.62 (br s, 1H, CHOH), 7.04 (d, *J* = 8.9 Hz, 1H, NHFmoc), 7.33 (dd, *J* = 7.4, 7.4 Hz, 2H, Ar), 7.42 (dd, *J* = 7.4, 7.4 Hz, 2H, Ar), 7.74 (d, *J* = 7.4 Hz, 2H, Ar), 7.90 (d, *J* = 7.4 Hz, 2H, Ar), 12.62 ppm (br s, 1H, COOH). ¹³C NMR (75 MHz, DMSO-*d*₆): δ = 21.9, 22.9, 23.9, 42.6, 46.5, 58.8, 65.6, 68.5, 120.0, 125.1, 125.2, 126.9, 127.5, 140.58, 140.61, 143.6, 143.7, 156.2, 172.3 ppm. IR (KBr): $\tilde{\nu}$ = 3444 (m), 3301 (m), 3068 (w), 2957 (m), 2866 (w), 1731 (s), 1695 (s), 1540 (s), 1468 (w), 1450 (m), 1420 (w), 1343 (w), 1316 (m), 1275 (m), 1246 (m), 1178 (w), 1120 (w), 1082 (m), 1075 (m), 981 (w), 956 (w), 939 (w), 851 (w), 778 (w), 757 (m), 740 (s), 692 (m), 622 cm⁻¹ (w). Elemental analysis calcd (%) for C₂₂H₂₅NO₅ (383.44): C 68.91, H 6.57, N 3.65; found: C 68.87, H 6.60, N 3.56.

(2*S,3*S**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-5-methylhexanoic Acid ((±)-*anti*-3c)**

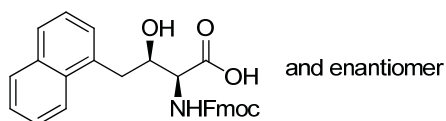
Yield 88%. M.p. 178-180 °C. R_f = 0.45 (ethyl acetate/*n*-hexane 5:1, 1% HOAc). ^1H NMR (300 MHz, DMSO- d_6): δ = 0.86 (d, J = 6.6 Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 0.89 (d, J = 6.9 Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 1.19 (ddd, J = 12.6, 9.5, 2.6 Hz, 1H, $i\text{PrCH}_2$), 1.47 (ddd, J = 14.0, 10.3, 4.4 Hz, 1H, $i\text{PrCH}_2$), 1.68-1.88 (m, 1H, $(\text{CH}_3)_2\text{CH}$), 3.72-3.93 (m, 1H, CHOH), 4.04 (dd, J = 8.7, 5.6 Hz, 1H, CHNH), 4.17-4.40 (m, 3H, OCH_2CH , OCH_2CH), 4.88 (br s, 1H, CHOH), 7.33 (dd, J = 7.4, 7.4 Hz, 2H, Ar), 7.42 (dd, J = 7.2, 7.2 Hz, 2H, Ar), 7.48 (d, J = 8.8 Hz, 1H, NHFmoc), 7.75 (d, J = 7.4 Hz, 2H, Ar), 7.88 (d, J = 7.4 Hz, 2H, Ar), 12.55 ppm (br s, 1H, COOH). ^{13}C NMR (75 MHz, DMSO- d_6): δ = 21.4, 23.6, 23.8, 41.8, 46.6, 59.9, 65.7, 68.6, 120.0, 125.20, 125.22, 127.0, 127.5, 140.62, 140.64, 143.72, 143.74, 156.0, 172.1 ppm. IR (KBr): $\tilde{\nu}$ = 3325 (m), 3065 (w), 2955 (m), 1719 (m), 1696 (s), 1545 (m), 1467 (w), 1450 (m), 1420 (w), 1384 (w), 1370 (w), 1310 (m), 1286 (w), 1263 (m), 1240 (w), 1144 (w), 1105 (w), 1081 (w), 1056 (w), 1040 (w), 1008 (w), 993 (w), 932 (w), 850 (w), 774 (w), 756 (w), 736 (m), 692 (w), 648 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{22}\text{H}_{25}\text{NO}_5$ (383.44): C 68.91, H 6.57, N 3.65; found: C 68.66, H 6.64, N 3.50.

(2*S,3*R**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxypentanoic Acid ((±)-*syn*-3b)**

Yield: 74%; colorless solid. M.p. 147-148 °C. R_f = 0.56 (ethyl acetate/methanol 4:1, 1% HOAc). ^1H NMR (300 MHz, DMSO- d_6): δ = 0.85 (t, J = 7.4 Hz, 3H, CH_3CH_2), 1.41 (m, 2H, CH_3CH_2), 3.82 (m, 1H, CHOH), 4.06 (dd, J = 9.3, 3.0 Hz, 1H, CHNH), 4.18-4.39 (m, 3H, OCH_2CH , OCH_2CH), 4.68 (br s, 1H, CHOH), 7.02 (d, J = 9.3 Hz, 1H, NHFmoc), 7.33 (dd, J = 7.4, 7.4 Hz, 2H, Ar), 7.42 (dd, J = 7.2, 7.2 Hz, 2H, Ar), 7.74 (d, J = 7.4 Hz, 2H, Ar), 7.89 (d, J = 7.4 Hz, 2H, Ar), 12.60 ppm (br s, 1H, COOH). ^{13}C NMR (63 MHz, DMSO- d_6): δ = 10.1, 26.7, 46.6, 57.9, 65.7, 72.0, 120.02, 120.04, 125.21, 125.25, 127.0, 127.6, 140.6, 140.7, 143.7, 143.8, 156.3, 172.5 ppm. IR (KBr): $\tilde{\nu}$ = 3387 (m), 3276 (m), 3063 (m), 3038 (m), 2963 (m), 2934 (m), 2895 (w), 1735 (m), 1700 (s), 1545 (m), 1478 (w), 1450 (m), 1420 (w), 1384 (w), 1341 (m), 1324 (m), 1295 (m), 1262 (m), 1108 (m), 1074 (m), 1036 (w), 996 (w), 968 (w), 811 (w), 758 (m), 738 (m), 698 (w), 621 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{20}\text{H}_{21}\text{NO}_5$ (355.38): C 67.59, H 5.96, N 3.94; found: C 67.63, H 6.03, N 4.10.

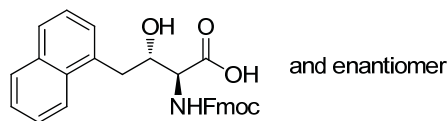
(2S*,3S*)-2-(9H-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxypentanoic Acid ((±)-anti-3b)

Yield: 91%; colorless solid. M.p. 197-199 °C. R_f = 0.56 (ethyl acetate/methanol 4:1, 1% HOAc). ^1H NMR (250 MHz, DMSO- d_6): δ = 0.88 (t, J = 7.3 Hz, 3H, CH_3CH_2), 1.31-1.62 (m, 2H, CH_3CH_2), 3.55-3.71 (m, 1H, CHOH), 3.99 (dd, J = 8.6, 6.1 Hz, 1H, CHNH), 4.16-4.37 (m, 3H, OCH_2CH , OCH_2CH), 4.89 (br s, 1H, CHOH), 7.33 (dd, J = 7.2, 7.2 Hz, 2H, Ar), 7.42 (dd, J = 7.1, 7.1 Hz, 2H, Ar), 7.45 (d, J = 8.7 Hz, 1H, NHFmoc), 7.74 (d, J = 7.3 Hz, 2H, Ar), 7.89 (d, J = 7.4 Hz, 2H, Ar), 12.48 ppm (br s, 1H, COOH). ^{13}C NMR (75 MHz, DMSO- d_6): δ = 10.0, 25.8, 46.5, 58.9, 65.6, 71.8, 120.0, 125.2, 126.9, 127.5, 140.58, 140.59, 143.67, 143.71, 155.9, 172.2 ppm. IR (KBr): $\tilde{\nu}$ = 3307 (m), 3070 (w), 3018 (w), 2958 (w), 2936 (w), 2918 (w), 1726 (m), 1693 (s), 1554 (m), 1478 (w), 1451 (w), 1444 (w), 1420 (w), 1384 (w), 1310 (m), 1266 (m), 1160 (w), 1106 (w), 1082 (w), 1043 (m), 980 (w), 938 (w), 895 (w), 754 (w), 738 (m), 696 (w), 648 (w), 621 (w), 586 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{20}\text{H}_{21}\text{NO}_5$ (355.38): C 67.59, H 5.96, N 3.94; found: C 67.86, H 6.15, N 3.95.

(2S*,3R*)-2-(9H-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-4-(1-naphthyl)butanoic Acid ((±)-syn-3d)

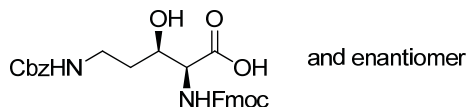
Yield: 77%; colorless solid. M.p. 122-125 °C. R_f = 0.39 (ethyl acetate/methanol 9:1, 1% HOAc). ^1H NMR (400 MHz, DMSO- d_6): δ = 3.22 (d, J = 7.1 Hz, 2H, CH_2 -1-Naphthyl), 4.04 (dd, J = 9.3, 2.4 Hz, 1H, CHNH), 4.28-4.46 (m, 4H, CHOH , OCH_2CH , OCH_2CH), 5.16 (br s, 1H, CHOH), 7.28 (d, J = 9.4 Hz, 1H, NHFmoc), 7.29-7.58 (m, 8H, Ar), 7.77-7.87 (m, 3H, Ar), 7.89-7.95 (m, 3H, Ar), 8.09-8.17 (m, 1H, Ar), 12.66 ppm (br s, 1H, COOH). ^{13}C NMR (75 MHz, DMSO- d_6): δ = 37.2, 46.6, 57.6, 65.8, 71.2, 120.0, 123.7, 125.2, 125.5, 125.9, 126.8, 127.0, 127.4, 127.6, 128.5, 131.6, 133.4, 134.3, 140.7, 143.7, 143.8, 156.4, 172.3 ppm. IR (KBr): $\tilde{\nu}$ = 3553 (m), 3348 (m), 3065 (m), 3042 (m), 2955 (w), 1719 (s), 1527 (m), 1478 (w), 1420 (w), 1398 (w), 1340 (w), 1293 (w), 1258 (w), 1233 (m), 1103 (w), 1070 (m), 1009 (w), 961 (w), 879 (w), 853 (w), 778 (m), 751 (m), 742 (w), 671 (w), 640 (w), 621 (w), 569 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{29}\text{H}_{25}\text{NO}_5$ (467.51): C 74.50, H 5.39, N 3.00; found: C 74.69, H 5.64, N 2.79.

(2*S,3*S**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-4-(1-naphthyl)butanoic Acid ((±)-*anti*-3d)**



Yield: 63%; colorless solid. M.p. 238-240 °C (dec.). R_f = 0.39 (ethyl acetate/methanol 9:1, 1% HOAc). ^1H NMR (400 MHz, DMSO- d_6): δ = 3.12 (dd, J = 14.1, 9.1 Hz, 1H, CH_2 -1-Naphthyl), 3.34 (dd, J = 14.0, 3.7 Hz, 1H, CH_2 -1-Naphthyl), 4.14 (m, 1H, CHOH), 4.18-4.36 (m, 4H, CHNH , OCH_2CH , OCH_2CH), 5.10 (br s, 1H, CHOH), 7.26-7.36 (m, 1H, Ar), 7.36-7.48 (m, H, Ar), 7.52 (m, H, Ar), 7.73-7.84 (m, H, CHOH , Ar), 7.85-7.96 (m, H, Ar), 8.16 (d, J = 7.6 Hz, H, Ar), 12.76 ppm (br s, 1H, COOH). ^{13}C NMR (75 MHz, DMSO- d_6): δ = 36.1, 46.5, 59.5, 65.7, 71.1, 120.0, 124.0, 125.2, 125.3, 125.6, 126.5, 127.0, 127.5, 127.6, 128.4, 131.8, 133.3, 135.0, 140.6, 143.7, 156.1, 172.1 ppm. IR (KBr): $\tilde{\nu}$ = 3293 (m), 3063 (w), 3038 (w), 2963 (w), 2930 (w), 1702 (s), 1542 (m), 1509 (w), 1450 (w), 1414 (w), 1398 (w), 1376 (w), 1340 (w), 1294 (m), 1276 (m), 1235 (m), 1185 (w), 1149 (w), 1100 (w), 1082 (w), 1035 (m), 1020 (m), 1000 (m), 936 (w), 801 (m), 778 (m), 756 (m), 736 (m), 687 (m), 621 (w), 564 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{29}\text{H}_{25}\text{NO}_5$ (467.51): C 74.50, H 5.39, N 3.00; found: C 74.54, H 5.58, N 3.08.

(2*S,3*R**)-5-Benzoyloxycarbonylamino-2-(9*H*-fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-pentanoic Acid ((±)-*syn*-3a)**



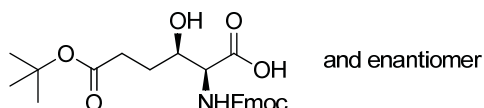
Due to the presence of the Cbz-group, cleavage of the ester was carried out using mild conditions. *tert*-Butyl ester ((±)-*syn*-2a (0.76 g, 2.25 mmol) was dissolved in dichloromethane (10 mL) and cooled to 0 °C. After addition of TFA (10 mL) the cooling bath was removed and the solution was stirred for 3 h at rt. Subsequently, the solvents were removed and the residue was Fmoc-protected as described for **3c**, affording ((±)-*syn*-3a (0.77 g, 68%) as a colorless solid.

M.p. 116-117 °C. R_f = 0.43 (ethyl acetate/methanol 10:1, 1% HOAc). ^1H NMR (400 MHz, DMSO- d_6): δ = 1.57 (m, 2H, $\text{CH}_2\text{CH}_2\text{N}$), 2.98-3.20 (m, 2H, $\text{CH}_2\text{CH}_2\text{N}$), 4.00 (m, 1H, CHOH), 4.07 (dd, J = 9.2, 2.9 Hz, 1H, CHNHFMoc), 4.16-4.38 (m, 3H, OCH_2CH , OCH_2CH), 4.76 (s, 1H, CHOH), 5.01 (m, 2H, OCH_2Ph), 7.12 (d, J = 9.2 Hz, 1H, NHFMoc), 7.22-7.46 (m, 10H, Ar, $\text{CH}_2\text{CH}_2\text{NH}$), 7.75 (d, J = 7.4 Hz, 2H, Ar), 7.89 (d, J = 7.5 Hz, 2H, Ar), 12.67 ppm (br s, 1H, COOH). ^{13}C NMR (75 MHz, DMSO- d_6): δ = 33.9, 37.3, 46.6, 58.8, 65.1, 65.7, 68.2, 120.0, 125.17, 125.22, 127.0, 127.5, 127.6, 128.2, 137.2, 140.60, 140.62, 143.6, 143.7, 156.0, 156.3, 172.1 ppm. IR (KBr): $\tilde{\nu}$ = 3407 (m), 3312 (m), 3067 (w), 2953 (w), 1720 (m), 1691 (s), 1673 (m), 1549 (s), 1478 (w), 1450 (m), 1384 (w), 1338 (w), 1271 (s),

1150 (w), 1108 (w), 1087 (w), 1061 (w), 1044 (w), 1014 (w), 778 (w), 758 (w), 738 (m), 696 (m), 621 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{28}\text{H}_{28}\text{N}_2\text{O}_7$ (504.53): C 66.66, H 5.59, N 5.55; found: C 66.75, H 5.66, N 5.57.

Typical Procedure for the Fmoc-protection in presence of a *tert*-butyl ester.

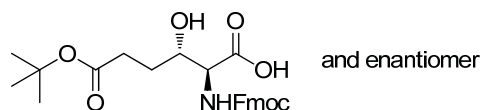
(2*S,3*R**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxyhexanedioic Acid 6-*tert*-Butyl Ester ((\pm)-*syn*-3e)**



To a solution of benzyl ester (\pm)-*syn*-17 (1.65 g, 5.1 mmol) in methanol (30 mL) was added Pd/C (10% palladium; 0.17 g). The reaction mixture was stirred for 4 h under a hydrogen atmosphere (1 bar). Afterwards, the suspension was filtered over a short plug of Celite and concentrated. According to the procedure described for (\pm)-*syn*-3c, the residue was further converted to the Fmoc-derivative, affording (\pm)-*syn*-3e (1.86 g, 80%) as a colorless semi-solid.

R_f = 0.62 (ethyl acetate/methanol 5:1, 1% HOAc). ^1H NMR (400 MHz, $\text{DMSO}-d_6$): δ = 1.38 (s, 9H, *t*Bu), 1.62 (m, 2H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 2.25 (m, 2H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 3.92 (td, J = 6.7, 3.2 Hz, 1H, CHOH), 4.03 (dd, J = 9.1, 3.1 Hz, 1H, CHNH), 4.22-4.35 (m, 3H, OCH_2CH , OCH_2CH), 7.09 (d, J = 9.1 Hz, 1H, NHFmoc), 7.33 (td, J = 7.4, 0.8 Hz, 2H, Ar), 7.42 (t, J = 7.4 Hz, 2H, Ar), 7.74 (d, J = 7.4 Hz, 2H, Ar), 7.89 (d, J = 7.5 Hz, 2H, Ar), 11.82 ppm (br s, 1H, COOH). ^{13}C NMR (101 MHz, $\text{DMSO}-d_6$): δ = 27.6, 29.1, 31.2, 46.6, 58.5, 65.7, 69.5, 79.4, 120.00, 120.02, 125.18, 125.24, 127.0, 127.5, 140.61, 140.63, 143.67, 143.73, 156.3, 172.0, 172.2 ppm. IR (KBr): $\tilde{\nu}$ = 3420 (m), 3067 (w), 2977 (m), 2934 (w), 1727 (s), 1526 (m), 1478 (w), 1450 (m), 1420 (w), 1394 (w), 1368 (m), 1323 (w), 1252 (m), 1152 (s), 1079 (m), 952 (w), 844 (w), 759 (m), 740 (m), 671 (w), 621 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{25}\text{H}_{29}\text{NO}_7$ (455.50): C 65.92, H 6.42, N 3.08; found: C 65.89, H 6.20, N 2.96.

(2*S,3*S**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxyhexanedioic Acid 6-*tert*-Butyl Ester ((\pm)-*anti*-3e)**



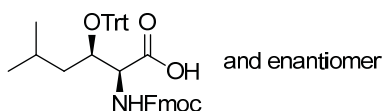
Yield: 79%; colorless semi-solid. R_f = 0.62 (ethyl acetate/methanol 5:1, 1% HOAc). ^1H NMR (400 MHz, $\text{DMSO}-d_6$): δ = 1.38 (s, 9H, *t*Bu), 1.63-1.73 (m, 2H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 2.15-2.26 (m, 1H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 2.26-2.37 (m, 1H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 3.73 (m, 1H, CHOH), 4.01 (dd, J = 8.7, 5.8 Hz, 1H, CHNH), 4.19-4.32 (m, 3H, OCH_2CH , OCH_2CH), 7.33 (td, J = 7.4, 0.8 Hz, 2H, Ar), 7.42 (t, J = 7.4 Hz, 2H, Ar), 7.50 (d, J = 8.8 Hz, 1H, NHFmoc), 7.74 (d, J = 7.4 Hz, 2H, Ar), 7.89 (d, J = 7.5 Hz, 2H, Ar), 12.36 ppm (br s, 1H, COOH). ^{13}C NMR (101 MHz, $\text{DMSO}-d_6$): δ = 27.7, 28.4, 31.3, 46.5, 59.2, 65.7, 69.5, 79.4, 120.0, 125.21, 125.24, 127.0, 127.6, 140.61, 140.63, 143.7, 156.0, 172.0, 172.2 ppm. IR (KBr): $\tilde{\nu}$ = 3406 (m), 3064 (w), 3040 (w), 2975 (m), 2935 (w), 1748 (w), 1716 (s), 1523 (m), 1478

(w), 1450 (m), 1398 (m), 1367 (m), 1324 (w), 1279 (m), 1234 (m), 1214 (w), 1199 (w), 1176 (m), 1162 (m), 1083 (m), 1050 (w), 1008 (w), 956 (w), 920 (w), 849 (w), 774 (w), 759 (m), 740 (m), 622 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{25}\text{H}_{29}\text{NO}_7$ (455.50): C 65.92, H 6.42, N 3.08; found: C 65.77, H 6.19, N 3.06.

Preparation of Fmoc- and Trityl-protected β -Hydroxy- α -amino Acids

Typical Procedure for the Tritylation.

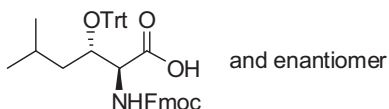
(2*S**,3*R**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-5-methyl-3-(trityloxy)hexanoic Acid ((\pm)-*syn*-4)



To a mixture of AgOTf (1.16 g, 4.5 mmol) and 2,6-lutidine (524 μL , 4.5 mmol) in CH_2Cl_2 (5 mL) was added a solution of TrtCl (1.25 g, 4.5 mmol) in CH_2Cl_2 (4 mL) at 0 $^\circ\text{C}$. To the resulting yellow suspension was added a solution of secondary alcohol (\pm)-*syn*-3c (0.58 g, 1.5 mmol) in CH_2Cl_2 (4 mL) dropwise over 15 min. The mixture was stirred for 1 h at 0 $^\circ\text{C}$ and for 12 h at room temperature. After addition of methanol (1 mL) the reaction mixture was stirred for 15 min and the solvents were evaporated in vacuo. The crude brown oil was purified by column chromatography (silica gel; *n*-hexane/ethyl acetate, 4:1). The solvents were evaporated resulting in (\pm)-*syn*-4 (0.71 g, 75%) as a colorless foam.

R_f = 0.39 (*n*-hexane/ethyl acetate 3:2, 1% HOAc). ^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ = 0.25 (d, J = 6.5 Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 0.52 (d, J = 6.5 Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 0.80-0.94 (m, 1H, $i\text{PrCH}_2$), 1.12-1.25 (m, 1H, $(\text{CH}_3)_2\text{CH}$), 1.54 (m, 1H, $i\text{PrCH}_2$), 3.67 (m, 1H, CHOTrt), 4.22 (dd, J = 9.8, 2.2 Hz, 1H, CHNH), 4.27-4.42 (m, 3H, OCH_2CH , OCH_2CH), 7.16-7.58 (m, 19H, Ph, Ar), 7.83 (d, J = 7.9 Hz, 1H, Ar), 7.87 (d, J = 7.9 Hz, 1H, Ar), 7.92 (d, J = 7.9 Hz, 2H, Ar), 8.08 (d, J = 9.8 Hz, 1H, NHFmoc), 12.53 ppm (br s, 1H, COOH). ^{13}C NMR (75 MHz, $\text{DMSO}-d_6$): δ = 20.2, 23.8, 23.9, 46.6, 56.2, 65.8, 73.5, 85.9, 120.0, 125.3, 126.7, 127.0, 127.2, 127.4, 127.6, 128.8, 140.60, 140.62, 143.6, 143.8, 144.9, 156.7, 172.2 ppm. IR (KBr): $\tilde{\nu}$ = 3442 (m), 3059 (m), 3034 (w), 2956 (m), 2869 (w), 1725 (s), 1508 (m), 1491 (m), 1449 (s), 1414 (w), 1319 (m), 1217 (m), 1118 (w), 1057 (m), 1033 (w), 1001 (m), 934 (w), 898 (w), 758 (m), 741 (s), 705 (s), 633 (m), 621 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{41}\text{H}_{39}\text{NO}_5$ (625.75): C 78.70, H 6.28, N 2.24; found: C 78.84, H 6.24, N 2.06.

(2*S**,3*S**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-5-methyl-3-(trityloxy)hexanoic Acid ((\pm)-*anti*-4)

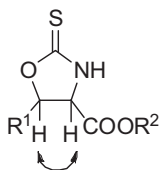


Yield: 53%; colorless semi-solid. $R_f = 0.39$ (*n*-hexane/ethyl acetate 3:2, 1% HOAc). ^1H NMR (300 MHz, DMSO- d_6): $\delta = 0.36$ (d, $J = 6.4$ Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 0.61 (d, $J = 6.5$ Hz, 3H, $(\text{CH}_3)_2\text{CH}$), 1.21 (m, 1H, *i*PrCH₂), 1.39 (m, 1H, $(\text{CH}_3)_2\text{CH}$), 1.54 (m, 1H, *i*PrCH₂), 3.70 (m, 1H, CHOTrt), 4.04 (dd, $J = 7.3, 3.0$ Hz, 1H, CHNHFmoc), 4.18–4.32 (m, 3H, OCH₂CH, OCH₂CH), 7.18–7.50 (m, 20H, Ph, Ar, NHFmoc), 7.78 (d, $J = 7.4$ Hz, 1H, Ar), 7.79 (d, $J = 7.5$ Hz, 1H, Ar), 7.92 (d, $J = 7.5$ Hz, 2H, Ar), 12.69 ppm (br s, 1H, COOH). ^{13}C NMR (75 MHz, DMSO- d_6): $\delta = 22.1, 22.4, 23.7, 46.5, 57.4, 66.0, 71.7, 86.4, 120.1, 125.25, 125.31, 127.0, 127.6, 128.5, 140.59, 140.62, 143.7, 143.8, 144.4, 155.9, 171.4$ ppm. IR (KBr): $\tilde{\nu} = 3429$ (m), 3060 (m), 2955 (m), 1723 (s), 1501 (m), 1448 (m), 1340 (w), 1222 (m), 1126 (w), 1046 (m), 994 (w), 937 (w), 900 (w), 743 (m), 704 (m), 634 (m), 539 (w), 506 (w), 423 cm^{-1} (w). Elemental analysis calcd (%) for C₄₁H₃₉NO₅ (625.75): C 78.70, H 6.28, N 2.24; found: C 78.83, H 6.41, N 2.18.

Preparation of Thioxooxazolidines

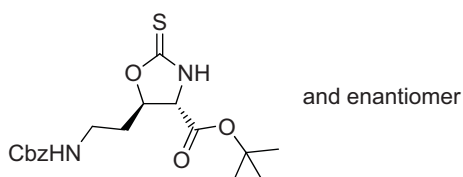
General Procedure.

β -Hydroxy- α -amino acid ester **2** or **17** (0.2 mmol) and 1,1'-thiocarbonyldiimidazol (0.2 mmol) were solved in dichloromethane (5 mL). After stirring for 2 h at rt the solvent was removed in vacuo and the residue was purified by column chromatography.



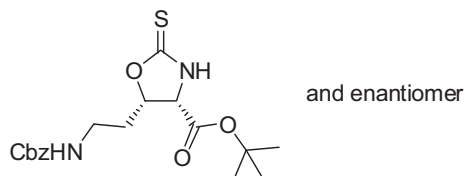
The relative stereochemistry was assigned by analysis of the coupling constant between CHOC=S and CHNHC=S.

(4*S**,5*R**)-5-(2-Benzyloxycarbonylaminoethyl)-2-thioxooxazolidine-4-carboxylic Acid tert-Butyl Ester ((±)-*trans*-26a)



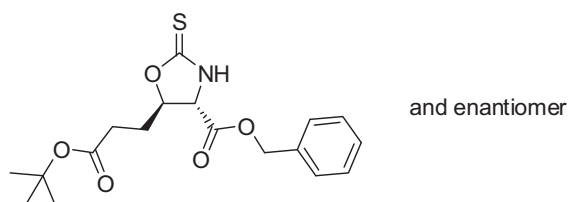
Yield: 89%; colorless oil. $R_f = 0.39$ (*n*-hexane/ethyl acetate 1:1). ^1H NMR (300 MHz, CDCl₃): $\delta = 1.46$ (s, 9H, *t*Bu), 1.98–2.22 (m, 2H, CH₂CH₂NHCbz), 3.39 (dd, $J = 6.4, 6.4$ Hz, 2H, CH₂CH₂NHCbz), 4.17 (d, $J = 6.7$ Hz, 1H, CHNHC=S), 4.88–4.98 (m, 1H, CHOC=S), 5.09 (s, 2H, CH₂Ph), 5.21 (m, 1H, NHCbz), 7.27–7.39 (m, 5H, Ph), 8.22 ppm (br s, 1H, NHC=S). ^{13}C NMR (75 MHz, CDCl₃): $\delta = 27.8, 34.7, 36.9, 62.6, 66.8, 83.2, 84.5, 128.0, 128.1, 128.5, 136.2, 156.4, 166.9, 188.6$ ppm. IR (KBr): $\tilde{\nu} = 3328$ (m), 2979 (m), 1711 (s), 1517 (s), 1369 (m), 1250 (s), 1176 (s), 1156 (s), 1021 (m), 926 (w), 840 (m), 741 (m), 699 (m), 456 cm^{-1} (w). HRMS (MALDI): m/z calcd for C₁₈H₂₄N₂O₅S+H⁺: 381.1479 [M+H⁺]; found: 381.1473.

(4S*,5S*)-5-(2-Benzylloxycarbonylaminoethyl)-2-thioxooxazolidine-4-carboxylic Acid *tert*-Butyl Ester (\pm)-*cis*-26a)



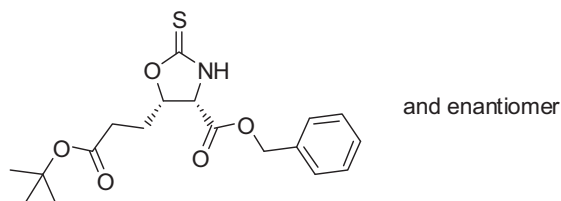
Yield: 87%; colorless semi-solid. R_f = 0.36 (*n*-hexane/ethyl acetate 1:1). ^1H NMR (300 MHz, CDCl_3): δ = 1.47 (s, 9H, *t*Bu), 1.76-1.90 (m, 1H, $\text{CH}_2\text{CH}_2\text{NHCbz}$), 2.04-2.22 (m, 1H, $\text{CH}_2\text{CH}_2\text{NHCbz}$), 3.32-3.53 (m, 2H, $\text{CH}_2\text{CH}_2\text{NHCbz}$), 4.48 (d, J = 8.9 Hz, 1H, $\text{CHNHC}=\text{S}$), 5.02-5.20 (m, 2H, $\text{CHOC}=\text{S}$, NHCbz), 5.10 (s, 2H, CH_2Ph), 7.28-7.39 (m, 5H, Ph), 7.70 ppm (br s, 1H, $\text{NHC}=\text{S}$). ^{13}C NMR (75 MHz, CDCl_3): δ = 27.9, 30.2, 37.5, 61.1, 66.8, 82.2, 84.5, 128.1, 128.2, 128.5, 136.3, 156.5, 166.5, 189.7 ppm. IR (KBr): $\tilde{\nu}$ = 3394 (m), 3186 (m), 2983 (m), 1730 (s), 1698 (s), 1523 (s), 1458 (w), 1369 (m), 1260 (s), 1228 (m), 1179 (s), 1152 (s), 1080 (m), 1022 (w), 983 (m), 932 (w), 855 (m), 833 (m), 748 (m), 695 (m), 637 (w), 551 (w), 457 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{18}\text{H}_{24}\text{N}_2\text{O}_5\text{S}$ (380.46): C 56.82, H 6.36, N 7.36, S 8.43; found: C 56.96, H 6.17, N 7.26, S 8.56.

(4S*,5R*)-5-(3-(*tert*-Butoxy)-3-oxopropyl)-2-thioxooxazolidine-4-carboxylic Acid Benzyl Ester (\pm)-*trans*-26b)



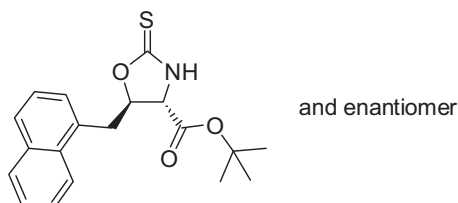
Yield: 92%; colorless oil. R_f = 0.42 (*n*-hexane/ethyl acetate 2:1). ^1H NMR (300 MHz, CDCl_3): δ = 1.43 (s, 9H, *t*Bu), 2.13 (m, 2H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 2.46 (ddd, J = 7.5, 7.1, 1.7 Hz, 2H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 4.29 (d, J = 6.3 Hz, 1H, $\text{CHNHC}=\text{S}$), 4.99 (m, 1H, $\text{CHOC}=\text{S}$), 5.18 (d, J = 12.0 Hz, 1H, CH_2Ph), 5.25 (d, J = 12.0 Hz, 1H, CH_2Ph), 7.28-7.48 (m, 5H, Ph), 7.92 ppm (br s, 1H, $\text{NHC}=\text{S}$). ^{13}C NMR (75 MHz, CDCl_3): δ = 28.0, 30.0, 30.2, 62.1, 68.3, 81.1, 84.3, 128.6, 128.8, 128.9, 134.2, 167.8, 171.3, 188.9 ppm. IR (film): $\tilde{\nu}$ = 3311 (m), 2977 (m), 2932 (w), 1727 (s), 1498 (m), 1456 (w), 1392 (w), 1368 (m), 1254 (m), 1192 (m), 1153 (s), 1103 (w), 1030 (w), 972 (w), 943 (w), 921 (w), 845 (w), 750 (m), 698 cm^{-1} (m). Elemental analysis calcd (%) for $\text{C}_{18}\text{H}_{23}\text{NO}_5\text{S}$ (365.44): C 59.16, H 6.34, N 3.83, S 8.77; found: C 59.29, H 6.45, N 3.67, S 9.03.

(4*S,5*S**)-5-(3-(*tert*-Butoxy)-3-oxopropyl)-2-thioxooxazolidine-4-carboxylic Acid Benzyl Ester ((±)-*cis*-26b)**

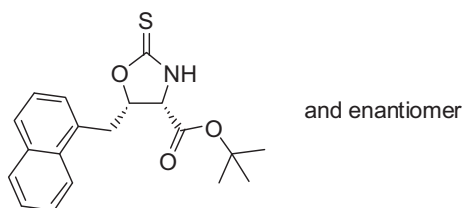


Yield: 88%; colorless solid. M.p. 94-95 °C from ethyl acetate/*n*-hexane. R_f = 0.34 (*n*-hexane/ethyl acetate 2:1). ^1H NMR (300 MHz, CDCl_3): δ = 1.41 (s, 9H, *t*Bu), 1.66-1.79 (m, 1H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 1.92-2.03 (m, 1H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 2.46 (dd, J = 7.4, 6.7 Hz, 2H, $\text{CH}_2\text{CH}_2\text{COO}t\text{Bu}$), 4.63 (d, J = 9.1 Hz, 1H, $\text{CHNHC}=\text{S}$), 5.09 (m, 1H, $\text{CHOC}=\text{S}$), 5.21 (s, 2H, CH_2Ph), 7.29-7.40 (m, 5H, Ph), 8.03 ppm (br s, 1H, $\text{NHC}=\text{S}$). ^{13}C NMR (75 MHz, CDCl_3): δ = 25.4, 28.0, 30.6, 60.7, 68.1, 80.9, 83.2, 128.7, 128.81, 128.83, 134.2, 167.5, 171.3, 189.6 ppm. IR (KBr): $\tilde{\nu}$ = 3176 (m), 3034 (w), 2984 (m), 2942 (w), 1746 (s), 1731 (s), 1523 (s), 1457 (w), 1438 (w), 1417 (w), 1392 (w), 1363 (m), 1342 (w), 1277 (w), 1250 (m), 1213 (s), 1179 (m), 1149 (s), 1092 (m), 1072 (w), 1022 (w), 989 (w), 945 (w), 909 (m), 852 (m), 756 (m), 730 (w), 699 (m), 635 (w), 574 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{18}\text{H}_{23}\text{NO}_5\text{S}$ (365.44): C 59.16, H 6.34, N 3.83, S 8.77; found: C 59.43, H 6.45, N 3.55, S 8.73.

(4*S,5*R**)-5-(1-Naphthylmethyl)-2-thioxooxazolidine-4-carboxylic Acid *tert*-Butyl Ester ((±)-*trans*-26c)**



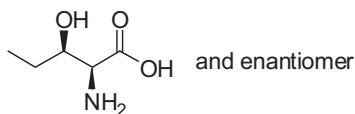
Yield: 92%; colorless solid. M.p. 135-137 °C from ethyl acetate/*n*-hexane. R_f = 0.26 (*n*-hexane/ethyl acetate 3:1). ^1H NMR (300 MHz, CDCl_3): δ = 1.26 (s, 9H, *t*Bu), 3.52 (dd, J = 14.5, 6.6 Hz, 1H, CH_2 -1-Naphthyl), 3.76 (dd, J = 14.5, 6.5 Hz, 1H, CH_2 -1-Naphthyl), 4.25 (d, J = 6.5 Hz, 1H, $\text{CHNHC}=\text{S}$), 5.15 (ddd, J = 6.5, 6.0, 6.0 Hz, 1H, $\text{CHOC}=\text{S}$), 7.41-7.61 (m, 5H, Ar, $\text{NHC}=\text{S}$), 7.81 (m, 1H, Ar), 7.85-7.92 (m, 1H, Ar), 8.01 ppm (dd, J = 8.2, 0.9 Hz, 1H, Ar). ^{13}C NMR (75 MHz, CDCl_3): δ = 27.6, 37.5, 61.8, 84.2, 84.8, 122.9, 125.6, 125.9, 126.6, 128.4, 129.1, 130.1, 131.9, 134.0, 166.9, 188.9 ppm. IR (KBr): $\tilde{\nu}$ = 3173 (m), 2973 (m), 2928 (w), 1750 (s), 1596 (w), 1517 (s), 1384 (w), 1368 (m), 1352 (w), 1343 (w), 1293 (w), 1244 (s), 1179 (s), 1151 (m), 1100 (w), 1050 (w), 1018 (w), 983 (w), 953 (w), 922 (w), 840 (w), 806 (w), 791 (m), 782 (m), 734 (w), 656 (w), 633 (w), 584 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{19}\text{H}_{21}\text{NO}_3\text{S}$ (343.44): C 66.45, H 6.16, N 4.08; found: C 66.32, H 6.29, N 4.15.

(4*S,5*S**)-5-(1-naphthylmethyl)-2-thioxooxazolidine-4-carboxylic Acid *tert*-Butyl Ester ((±)-*cis*-26c)**

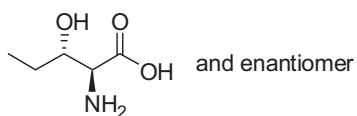
Yield: 96%; colorless solid. M.p. 168-171 °C from ethyl acetate/*n*-hexane. R_f = 0.17 (*n*-hexane/ethyl acetate 3:1). ^1H NMR (300 MHz, CDCl_3): δ = 1.56 (s, 9H, *t*Bu), 3.34 (dd, J = 15.0, 10.1 Hz, 1H, CH_2 -1-Naphthyl), 3.68 (dd, J = 15.0, 3.2 Hz, 1H, CH_2 -1-Naphthyl), 4.67 (d, J = 8.9 Hz, 1H, $\text{CHNHC}=\text{S}$), 5.38 (ddd, J = 10.1, 8.9, 3.2 Hz, 1H, $\text{CHOC}=\text{S}$), 7.40-7.60 (m, 5H, Ar, $\text{NHC}=\text{S}$), 7.76-7.83 (m, 1H, Ar), 7.86-7.92 ppm (m, 1H, Ar). ^{13}C NMR (75 MHz, CDCl_3): δ = 28.1, 33.1, 61.6, 83.9, 84.8, 122.6, 125.6, 125.7, 126.5, 127.8, 128.1, 129.2, 131.1, 131.5, 133.9, 166.5, 189.8 ppm. IR (KBr): $\tilde{\nu}$ = 3265 (m), 2977 (m), 1732 (s), 1597 (w), 1496 (s), 1395 (w), 1371 (m), 1303 (m), 1262 (s), 1175 (s), 1157 (s), 1082 (m), 1068 (m), 1044 (w), 1027 (w), 976 (w), 932 (w), 894 (w), 873 (w), 854 (w), 835 (w), 799 (m), 782 (m), 766 (w), 741 (w), 718 (w), 626 (w), 596 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{19}\text{H}_{21}\text{NO}_3\text{S}$ (343.44): C 66.45, H 6.16, N 4.08; found: C 66.19, H 6.28, N 4.17.

Preparation of Free Amino Acids**General Procedure.**

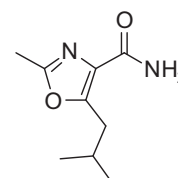
A small amount of *tert*-butyl ester **2b** was hydrolyzed with HCl (6 M). After removal of volatiles the residue was dissolved in water and filtrated over a short column of Amberlite IR-120 using aqueous NH_3 (5%). The filtrate was concentrated and the residue was recrystallized from water/acetone.

(2*S,3*R**)-2-Amino-3-hydroxypentanoic Acid ((±)-*syn*-27)**

Colorless solid. M.p. 227-229 °C (dec.) (ref.^[13] 229-230 °C (dec.)). ^1H NMR (300 MHz, D_2O): δ = 0.98 (t, J = 7.4 Hz, 3H, CH_3), 1.48-1.71 (m, 2H, CH_2), 3.65 (d, J = 4.4 Hz, 1H, CHNH_2), 3.98 ppm (ddd, J = 8.5, 5.3, 4.5 Hz, 1H, CHOH). ^{13}C NMR (75 MHz, D_2O): δ = 11.8, 28.8, 61.2, 73.7, 175.5 ppm. IR (KBr): $\tilde{\nu}$ = 3444 (m), 3092 (m), 2971 (m), 2621 (m), 1659 (m), 1594 (s), 1520 (m), 1464 (w), 1410 (m), 1367 (m), 1274 (w), 1173 (w), 1118 (w), 1058 (w), 1039 (w), 978 (m), 896 (w), 688 (w), 630 (w), 520 cm^{-1} (m). Elemental analysis calcd (%) for $\text{C}_5\text{H}_{11}\text{NO}_3$ (133.15): C 45.10, H 8.33, N 10.52; found: C 45.19, H 8.49, N 10.36.

(2S*,3S*)-2-Amino-3-hydroxypentanoic Acid ((±)-anti-27)

Colorless solid. M.p. 223-225 °C (dec.) (ref.^[13] 223-224 °C (dec.)). ¹H NMR (300 MHz, D₂O): δ = 0.99 (t, J = 7.4 Hz, 3H, CH₃), 1.45-1.60 (m, 2H, CH₂), 3.83 (d, J = 3.7 Hz, 1H, CHNH₂), 3.98 ppm (ddd, J = 7.5, 5.9, 3.7 Hz, 1H, CHOH). ¹³C NMR (75 MHz, D₂O): δ = 12.4, 26.8, 61.8, 73.7, 174.3 ppm. IR (KBr): $\tilde{\nu}$ = 3448 (m), 3088 (m), 2962 (m), 2876 (m), 2590 (w), 1654 (m), 1613 (s), 1582 (s), 1428 (m), 1352 (m), 1321 (w), 1189 (w), 1118 (w), 1084 (w), 1060 (w), 996 (w), 968 (w), 910 (w), 831 (w), 619 (w), 490 cm⁻¹ (w). Elemental analysis calcd (%) for C₅H₁₁NO₃ (133.15): C 45.10, H 8.33, N 10.52; found: C 44.97, H 8.25, N 10.38.

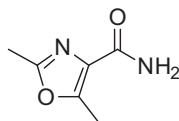
Preparation of Oxazoles**Typical Procedure: 5-Isobutyl-2-methyl-1,3-oxazole-4-carboxamide (14c)**

Rink amide MBHA resin (313 mg, 0.20 mmol, 0.64 mmol/g loading) was swelled in NMP (1 h). After cleavage of the Fmoc protecting group, 2-(9H-fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-5-methylhexanoic acid (**3c**) (230 mg, 0.60 mmol) was coupled to the resin bound amino group. The Fmoc protecting group was cleaved, followed by acylation with acetic anhydride (120 μ L, 2.00 mmol). Washing of the solid support with CH₂Cl₂ (5 \times) was followed by treatment (2 h) with a solution of DMP (0.1 M, 6.0 mL, 0.60 mmol) in CH₂Cl₂. After washing with CH₂Cl₂ (5 \times), NMP (5 \times) and CH₂Cl₂ (5 \times) a solution of PPh₃ (525 mg, 2.00 mmol), iodine (508 mg, 2.00 mmol) and DIPEA (698 μ L, 2.00 mmol) in CH₂Cl₂ (14 mL) was added to the resin and shaken (12 h). The resin was washed with CH₂Cl₂ (5 \times), NMP (5 \times) and CH₂Cl₂ (5 \times) and cleaved from the resin. The brown residue was purified by column chromatography (silica gel; *n*-hexane/ethyl acetate, 1:1), affording **14c** (26 mg, 71%) as a colorless solid.

M.p. 133-135 °C from ethyl acetate/*n*-hexane. R_f = 0.39 (ethyl acetate/*n*-hexane 3:1). ¹H NMR (250 MHz, CDCl₃): δ = 0.95 (d, J = 6.7 Hz, 6H, (CH₃)₂CH), 2.05 (m, 1H, (CH₃)₂CH), 2.41 (s, 3H, ArCH₃), 2.92 (d, J = 7.2 Hz, 2H, *i*PrCH₂), 5.41 (br s, 1H, NH₂), 6.75 ppm (br s, 1H, NH₂). ¹³C NMR (75 MHz, CDCl₃): δ = 13.7, 22.2, 27.9, 34.2, 129.0, 156.7, 158.4, 163.9 ppm. IR (KBr): $\tilde{\nu}$ = 3376 (s), 3297 (m), 3228 (m), 2960 (m), 2929 (w), 2872 (w), 1689 (s), 1662 (s), 1628 (m), 1616 (m), 1596 (m), 1461 (w), 1437 (m), 1391 (m), 1342 (w), 1317 (w), 1286 (w), 1244 (m), 1192 (m), 1131 (w), 1087 (w), 1054 (w), 1013 (w), 962 (w), 874 (w), 806 (w), 792 (w), 703 (m), 688 (w), 670 (w), 639 cm⁻¹ (m).

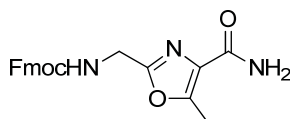
Elemental analysis calcd (%) for $C_9H_{14}N_2O_2$ (182.22): C 59.32, H 7.74, N 15.37; found: C 59.51, H 7.82, N 15.16.

2,5-Dimethyl-1,3-oxazole-4-carboxamide (14a)



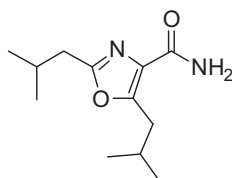
Monomers used: Fmoc-L-Thr-OH and acetic anhydride. Yield: 63%; colorless solid. M.p. 159-161 °C from H_2O (ref.^[14] 234 °C from H_2O). R_f = 0.18 (ethyl acetate/*n*-hexane 3:1). 1H NMR (300 MHz, acetic acid- d_4): δ = 2.43 (s, 3H, CH_3), 2.57 ppm (s, 3H, CH_3). ^{13}C NMR (75 MHz, acetic acid- d_4): δ = 11.9, 13.4, 128.8, 156.2, 161.4, 167.2 ppm. IR (KBr): $\tilde{\nu}$ = 3357 (m), 3216 (m), 2928 (w), 1683 (s), 1662 (s), 1628 (m), 1595 (w), 1439 (m), 1404 (w), 1390 (w), 1368 (m), 1306 (w), 1247 (m), 1201 (m), 1118 (m), 1085 (m), 1048 (w), 986 (m), 938 (w), 794 (w), 780 (m), 696 (s), 666 cm^{-1} (m). Elemental analysis calcd (%) for $C_6H_8N_2O_2$ (148.20): C 51.42, H 5.75, N 19.99; found: C 51.58, H 5.87, N 20.15.

2-(9H-Fluoren-9-ylmethoxycarbonylaminoethyl)-5-methyl-1,3-oxazole-4-carboxamide (14b)



Monomers used: Fmoc-L-Thr-OH and Fmoc-Gly-OH. Yield: 71%; colorless solid. M.p. 226-227 °C from ethyl acetate/*n*-hexane. R_f = 0.40 (ethyl acetate/*n*-hexane 8:1). 1H NMR (300 MHz, DMSO- d_6): δ = 2.52 (s, 3H, CH_3), 4.20-4.38 (m, 5H, $CH_2NHFmoc$, OCH_2CH , OCH_2CH), 7.29 (m, 6H, Ar, NH_2), 7.70 (d, J = 7.4 Hz, 2H, Ar), 7.89 (d, J = 7.5 Hz, 2H, Ar), 8.00 ppm (t, J = 5.9 Hz, 1H, $NHFmoc$). ^{13}C NMR (75 MHz, DMSO- d_6): δ = 11.2, 37.4, 46.5, 65.6, 120.0, 125.0, 126.9, 127.5, 140.6, 143.7, 152.4, 156.1, 158.2, 163.0 ppm. IR (KBr): $\tilde{\nu}$ = 3429 (m), 3323 (m), 3167 (m), 3069 (w), 3016 (w), 2963 (w), 2924 (w), 1701 (s), 1685 (s), 1627 (m), 1582 (w), 1547 (m), 1478 (w), 1467 (w), 1445 (m), 1384 (w), 1348 (w), 1329 (w), 1268 (m), 1193 (m), 1156 (m), 1104 (w), 1083 (w), 1046 (w), 963 (w), 917 (w), 800 (m), 758 (m), 738 (m), 697 (w), 644 (w), 622 (w), 570 cm^{-1} (m). Elemental analysis calcd (%) for $C_{21}H_{19}N_3O_4$ (377.39): C 66.83, H 5.07, N 11.13; found: C 66.80, H 5.25, N 10.91.

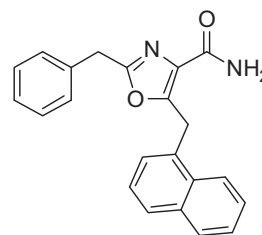
2,5-Isobutyl-1,3-oxazole-4-carboxamide (14d)



Monomers used: Fmoc-protected β -hydroxy- α -amino acid **3c** and isovaleric anhydride. Yield: 62%; colorless solid. M.p. 48-50 °C from ethyl acetate/*n*-hexane. R_f = 0.53 (ethyl acetate/*n*-hexane 3:1).

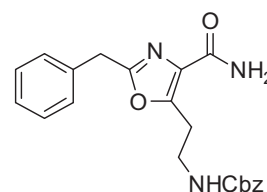
^1H NMR (250 MHz, CDCl_3): δ = 0.94 (t, J = 7.0 Hz, 12H, $(\text{CH}_3)_2\text{CH}$), 1.94-2.22 (m, 2H, $(\text{CH}_3)_2\text{CH}$), 2.57 (d, J = 7.1 Hz, 2H, $i\text{PrCH}_2$), 2.92 (d, J = 7.2 Hz, 2H, $i\text{PrCH}_2$), 5.72 (br s, 1H, NH_2), 6.81 ppm (br s, 1H, NH_2). ^{13}C NMR (63 MHz, CDCl_3): δ = 22.21, 22.22, 27.4, 27.9, 34.2, 36.8, 128.8, 156.4, 161.4, 164.2 ppm. IR (KBr): $\tilde{\nu}$ = 3474 (s), 3384 (w), 3352 (w), 3274 (w), 3139 (m), 2960 (s), 2928 (m), 2892 (w), 2870 (m), 1689 (s), 1630 (s), 1582 (m), 1466 (w), 1433 (w), 1414 (m), 1390 (w), 1363 (w), 1338 (w), 1317 (w), 1289 (w), 1231 (w), 1212 (w), 1187 (m), 1133 (w), 1082 (w), 1044 (m), 975 (w), 926 (w), 870 (w), 823 (w), 792 (w), 762 (w), 700 (m), 586 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{12}\text{H}_{20}\text{N}_2\text{O}_2$ (224.30): C 64.26, H 8.99, N 12.49; found: C 64.12, H 9.00, N 12.32.

2-Benzyl-5-(1-naphthylmethyl)-1,3-oxazole-4-carboxamide (14e)



Monomers used: Fmoc-protected β -hydroxy- α -amino acid **3d** and phenylacetic acid. Yield: 64%; off-white solid. M.p. 128-132 $^\circ\text{C}$ from ethyl acetate/*n*-hexane. R_f = 0.24 (*n*-hexane/ethyl acetate 1:1). ^1H NMR (300 MHz, CDCl_3): δ = 3.95 (s, 2H, CH_2 -1-Naphthyl), 4.87 (s, 2H, CH_2Ph), 5.84 (br s, 1H, NH_2), 6.89 (br s, 1H, NH_2), 7.13-7.34 (m, 5H, Ph), 7.37-7.60 (m, 4H, Ar), 7.78 (d, J = 7.9 Hz, 1H, Ar), 7.86 (dd, J = 7.3, 1.9 Hz, 1H, Ar), 8.28 ppm (d, J = 7.8 Hz, 1H, Ar). ^{13}C NMR (75 MHz, CDCl_3): δ = 29.3, 34.3, 124.2, 125.5, 125.7, 126.3, 127.1, 127.5, 127.7, 128.5, 128.6, 128.7, 128.8, 131.9, 132.8, 133.8, 134.7, 155.0, 160.6, 164.0 ppm. IR (KBr): $\tilde{\nu}$ = 3377 (m), 3220 (m), 3054 (w), 1679 (s), 1658 (s), 1624 (m), 1581 (m), 1512 (w), 1494 (m), 1454 (w), 1430 (m), 1411 (m), 1398 (m), 1339 (m), 1293 (m), 1258 (w), 1241 (m), 1215 (m), 1184 (m), 1156 (w), 1139 (m), 1099 (w), 1065 (m), 1029 (w), 1003 (w), 987 (m), 966 (w), 942 (w), 918 (w), 903 (w), 864 (w), 841 (w), 828 (m), 790 (s), 773 (m), 735 (m), 717 (m), 695 (s), 674 (w), 651 (w), 624 (m), 566 cm^{-1} (m). Elemental analysis calcd (%) for $\text{C}_{22}\text{H}_{18}\text{N}_2\text{O}_2$ (342.39): C 77.17, H 5.30, N 8.18; found: C 76.93, H 5.22, N 8.01.

2-Benzyl-5-(2-benzyloxycarbonylaminoethyl)-1,3-oxazole-4-carboxamide (14f)

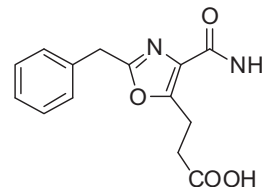


Monomers used: Fmoc-protected β -hydroxy- α -amino acid (\pm)-*syn*-**3a** and phenylacetic acid. Yield: 54%; colorless solid. M.p. 110-112 $^\circ\text{C}$ from ethyl acetate/*n*-hexane. R_f = 0.40 (ethyl acetate/*n*-hexane 5:1). ^1H NMR (300 MHz, CDCl_3): δ = 3.20 (t, J = 6.4 Hz, 2H, $\text{CH}_2\text{CH}_2\text{NH}$), 3.47 (dd, J = 11.9, 6.0 Hz, 2H, $\text{CH}_2\text{CH}_2\text{NH}$), 4.03 (s, 2H, OCH_2Ph), 5.04 (s, 2H, ArCH_2Ph), 5.70 (br s, 1H, NH_2), 6.13 (br s, 1H, NHCbz), 6.81 (br s, 1H, NH_2), 7.15-7.42 ppm (m, 10H, Ph). ^{13}C NMR (75 MHz, CDCl_3): δ = 26.0, 34.3, 39.6, 66.5, 127.3, 127.9, 128.0, 128.4, 128.7, 128.8, 129.8, 134.7, 136.6, 154.6, 156.4, 160.8, 164.3 ppm. IR (KBr): $\tilde{\nu}$ = 3411 (m), 3375 (m), 3280 (w), 3061 (w), 3033 (w), 2936 (w), 1696 (s), 1685 (s), 1624 (m), 1583 (m), 1531 (m), 1497 (m), 1453 (w), 1424 (m), 1413 (m), 1366 (w), 1311 (w), 1249 (m), 1218 (m), 1182 (w), 1146 (w), 1124 (w), 1093 (w), 1077 (w), 1039 (m), 1027 (m), 975

(w), 903 (w), 804 (w), 773 (m), 740 (m), 725 (m), 697 (m), 609 (m), 583 cm^{-1} (m). Elemental analysis calcd (%) for $\text{C}_{21}\text{H}_{21}\text{N}_3\text{O}_4$ (379.41): C 66.48, H 5.58, N 11.08; found: C 66.35, H 5.65, N 10.95.

3-(2-Benzyl-4-carbamoyl-1,3-oxazole-5-yl)propanoic Acid (14g)

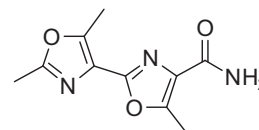
Monomers used: Fmoc-protected β -hydroxy- α -amino acid **3e** and phenylacetic acid. Yield: 60%; colorless solid. M.p. 212-215 °C from ethyl acetate/*n*-hexane. R_f = 0.68 (ethyl acetate/methanol 1:1, 1% HOAc).



^1H NMR (400 MHz, $\text{DMSO}-d_6$): δ = 2.55 (t, J = 7.6 Hz, 2H, $\text{CH}_2\text{CH}_2\text{COOH}$), 3.17 (t, J = 7.6 Hz, 2H, $\text{CH}_2\text{CH}_2\text{COOH}$), 4.11 (s, 2H, CH_2Ph), 7.17-7.54 (m, 7H, NH_2 , Ph), 12.27 ppm (br s, 1H, COOH). ^{13}C NMR (75 MHz, $\text{DMSO}-d_6$): δ = 20.9, 31.4, 33.3, 126.8, 128.49, 128.54, 129.1, 135.4, 154.6, 159.7, 162.9, 173.0 ppm. IR (KBr): $\tilde{\nu}$ = 3357 (m), 3159 (m), 2927 (w), 2691 (w), 2595 (w), 1712 (s), 1682 (s), 1625 (m), 1574 (m), 1497 (w), 1422 (m), 1291 (m), 1250 (w), 1191 (m), 1077 (m), 1011 (w), 918 (w), 848 (w), 818 (w), 760 (w), 706 (m), 672 (m), 632 (w), 607 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{14}\text{H}_{14}\text{N}_2\text{O}_4$ (274.27): C 61.31, H 5.14, N 10.21; found: C 61.16, H 5.21, N 9.99.

Preparation of a Bioxazole

2',5,5'-Trimethyl-[2,4'-bioxazole]-4-carboxamide (28)



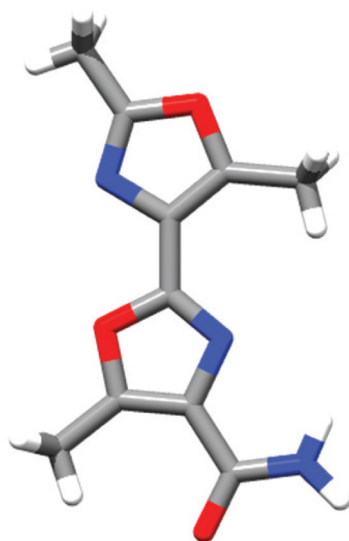
Rink amide MBHA resin (391 mg, 0.25 mmol, 0.64 mmol/g loading) was swelled in NMP (1 h). The resin bound Fmoc protecting group was removed and Fmoc-L-Thr-OH (256 mg, 0.75 mmol) was coupled to the free amino group. After cleavage of the Fmoc protecting group, Fmoc-L-Thr(Trt)-OH (438 mg, 0.75 mmol) was coupled to the resin. The resin was washed with CH_2Cl_2 (5 \times), treated (2 h) with a solution of DMP (0.1 M, 7.5 mL, 0.75 mmol) in CH_2Cl_2 and washed with CH_2Cl_2 (5 \times), NMP (5 \times) and CH_2Cl_2 (5 \times). A solution of PPh_3 (656 mg, 2.50 mmol), iodine (635 mg, 2.50 mmol) and DIPEA (871 μL , 5.00 mmol) in CH_2Cl_2 (17 mL) was added, and the suspension was shaken (12 h). The resin was washed with CH_2Cl_2 (5 \times) and NMP (5 \times), followed by cleavage of the Fmoc protecting group. After acylation with acetic anhydride (396 μL , 3.75 mmol), the resin was washed with CH_2Cl_2 (5 \times) and the Trt protecting group was removed. A solution of DMP (0.1 M, 7.5 mL, 0.75 mmol) in CH_2Cl_2 was added and the suspension was agitated (2 h), after which the resin was washed with CH_2Cl_2 (5 \times), NMP (5 \times) and CH_2Cl_2 (5 \times). Treatment (12 h) with a solution of PPh_3 (656 mg, 2.50 mmol), iodine (635 mg, 2.50 mmol) and DIPEA (871 μL , 5.00 mmol) in 17 mL CH_2Cl_2 was followed by washing with CH_2Cl_2 (5 \times), NMP (5 \times) and CH_2Cl_2 (5 \times). After cleavage of the resin, the brown residue was purified by column chromatography (silica gel; *n*-hexane/ethyl acetate, 3:2), affording **28** (26 mg, 47%) as a colorless solid.

M.p. 202-204 °C from ethyl acetate/*n*-hexane. R_f = 0.32 (ethyl acetate/*n*-hexane 9:1). ^1H NMR (250 MHz, CDCl_3): δ = 2.49 (s, 3H, CH_3), 2.63 (s, 3H, CH_3), 2.70 (s, 3H, CH_3), 6.07 (br s, 1H, NH_2), 7.02 ppm (br s, 1H, NH_2). ^{13}C NMR (63 MHz, CDCl_3): δ = 11.6, 11.8, 13.7, 124.3, 129.1, 149.9, 153.4,

153.7, 160.6, 164.6 ppm. IR (KBr): $\tilde{\nu}$ = 3485 (m), 3307 (m), 3247 (m), 3150 (m), 3023 (w), 2953 (w), 2921 (w), 2854 (w), 1687 (m), 1673 (s), 1646 (m), 1625 (m), 1596 (m), 1564 (m), 1449 (m), 1420 (m), 1375 (m), 1343 (m), 1307 (w), 1280 (w), 1213 (m), 1200 (w), 1120 (m), 1062 (m), 984 (w), 952 (w), 933 (w), 800 (w), 785 (w), 770 (w), 744 (w), 713 (w), 675 (m), 620 (w), 576 cm^{-1} (w). Elemental analysis calcd (%) for $\text{C}_{10}\text{H}_{11}\text{N}_3\text{O}_3$ (221.21): C 54.29, H 5.01, N 19.00; found: C 54.51, H 5.20, N 18.71.

A single crystal of **28** was obtained by slow solvent evaporation (dichloromethane).

Crystal data for **2',5,5'-Trimethyl-[2,4'-bioxazole]-4-carboxamide (28)**: $\text{C}_{10}\text{H}_{11}\text{N}_3\text{O}_3$, $M_r = 221.21$, colorless rod, dimensions = 0.07 x 0.08 x 0.70 mm, orthorhombic, space group $Pnma$, $a = 15.482(4)$, $b = 6.6553(14)$, $c = 10.512(3)$ Å, $\alpha = \beta = \gamma = 90^\circ$, $V = 1083.2(4)$ Å³, $T = 163(2)$ K, $\lambda = 0.71073$, $Z = 4$, $\rho_{\text{calcd}} = 1.357$ Mg m⁻³, $\theta_{\text{max}} = 30.05^\circ$, $F(000) = 464$, $\mu(\text{MoK}\alpha) = 0.103$ mm⁻¹. 14894 reflections measured, 1604 unique ($R_{\text{int}} = 0.1194$), final $R1 = 0.0545$, $wR2 = 0.1039$ for 911 observed reflections ($I > 2\sigma(I)$), GOF = 1.040. Data/restraints/parameters 1604/0/124. Largest difference Fourier peak and hole 0.231 and -0.269 e Å⁻³. CCDC-866341 contains the supplementary crystallographic data for this paper. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

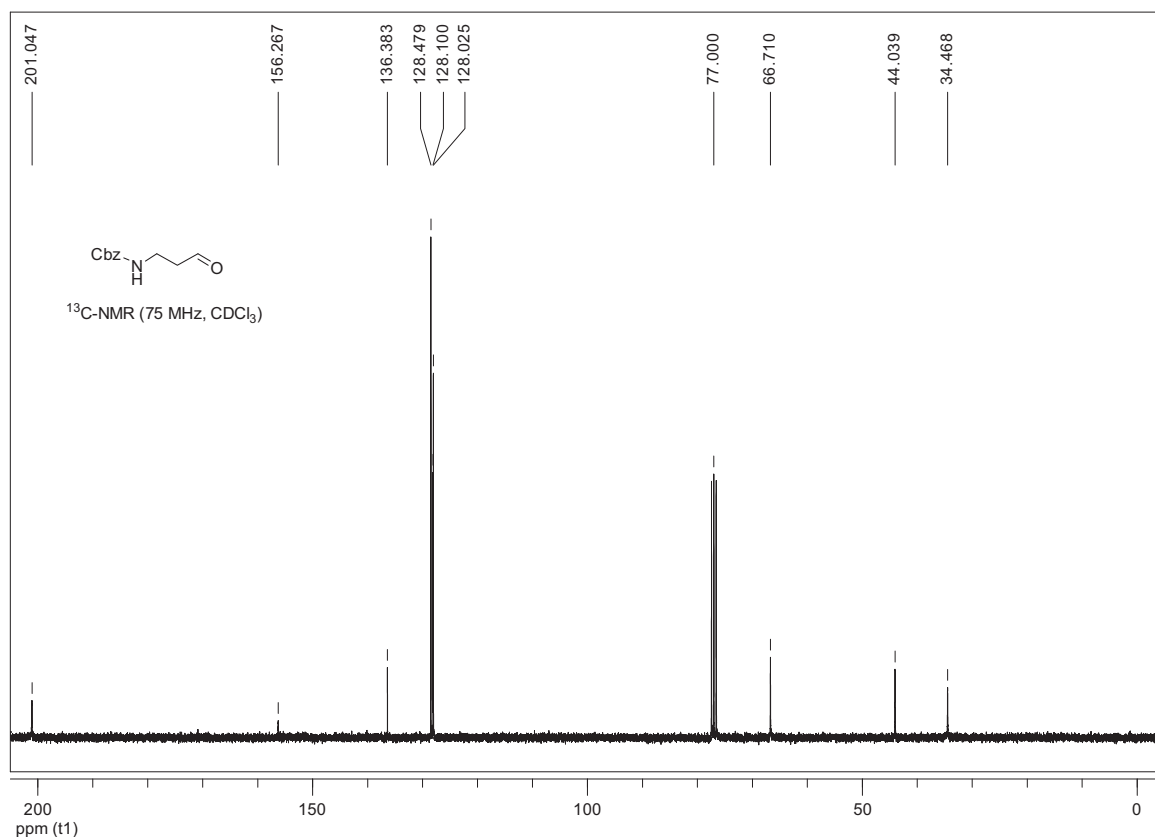
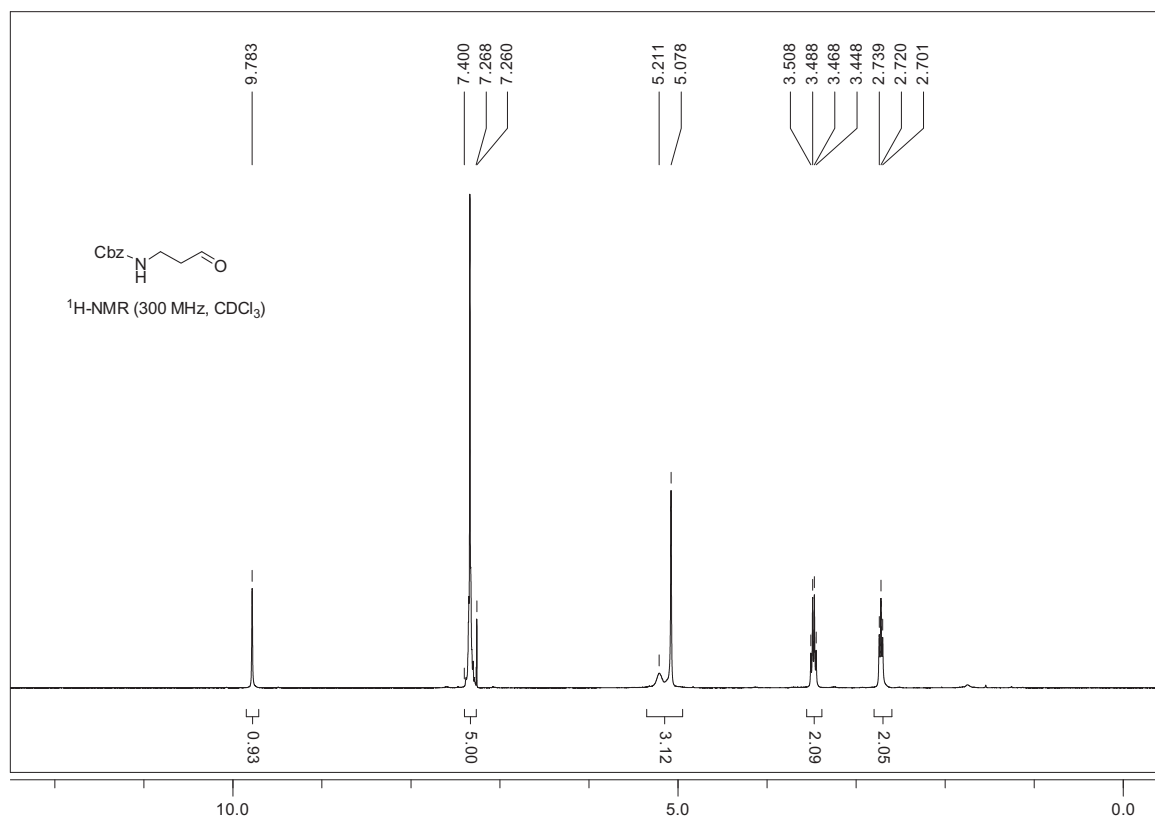


Single-crystal structure of bioxazole **28**.

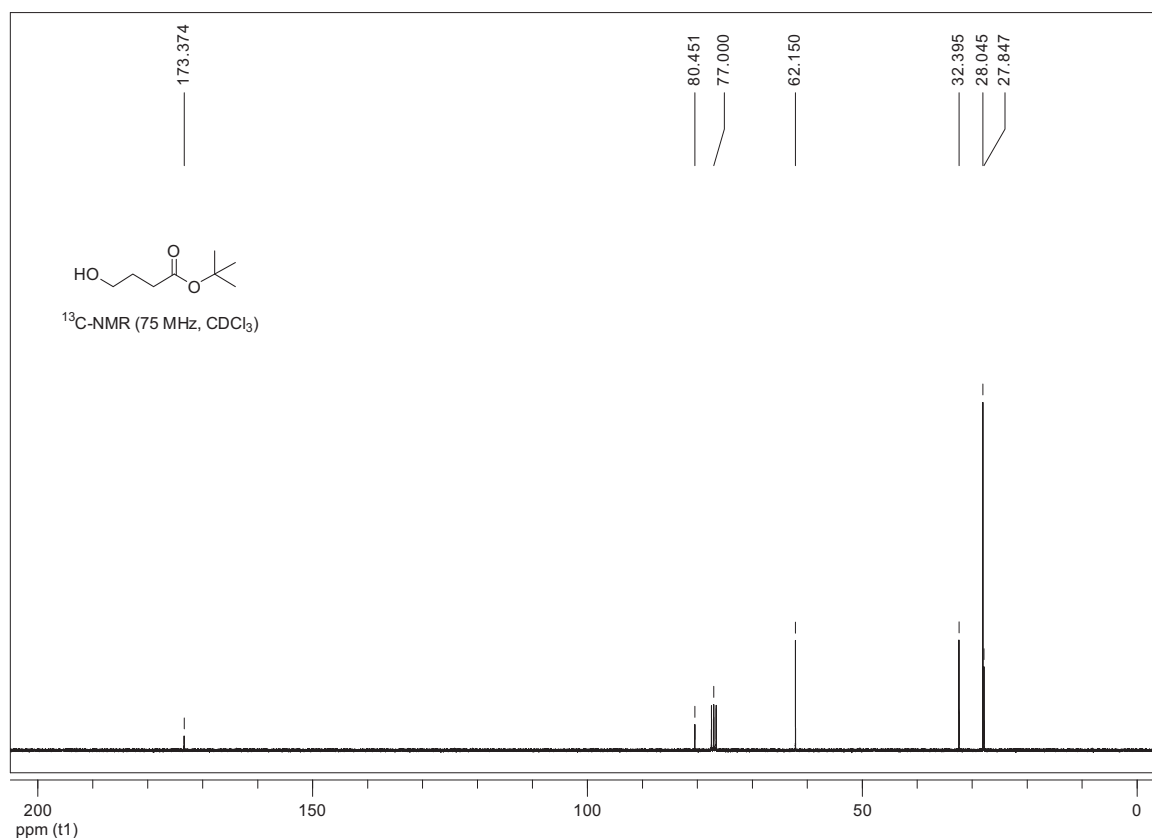
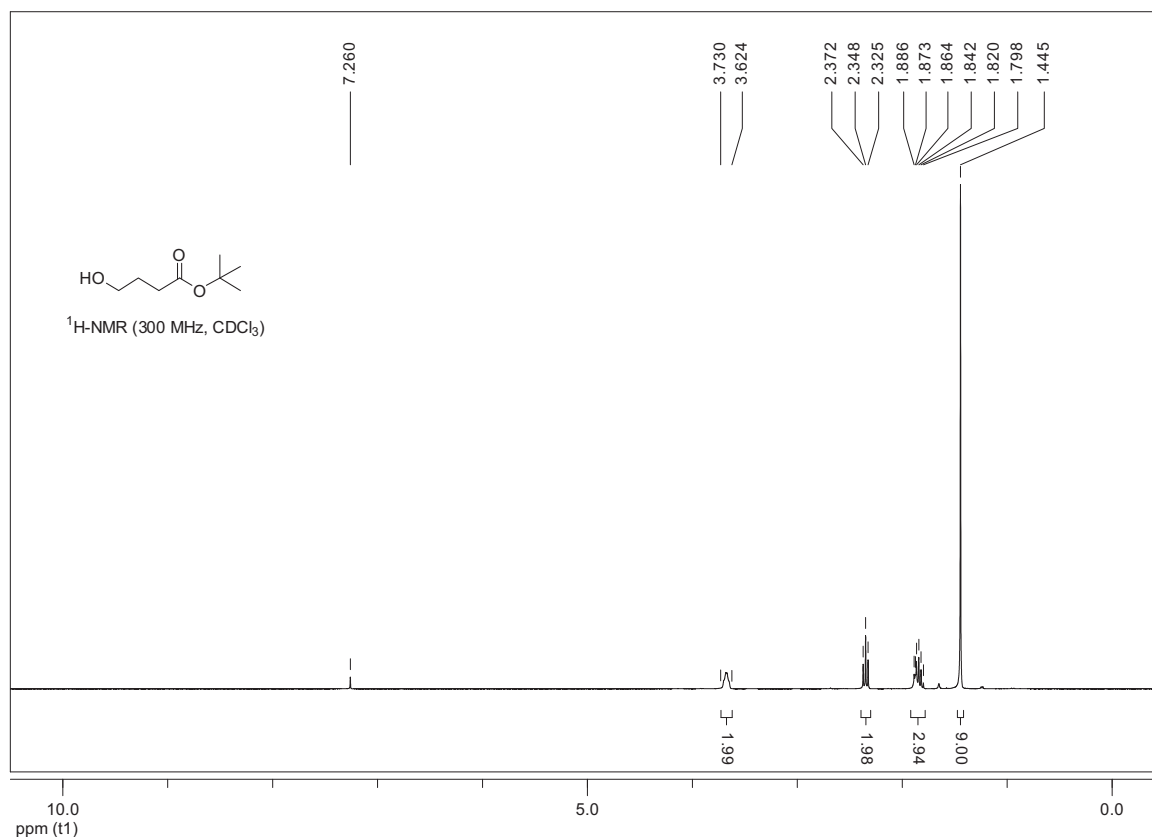
References

- [1] P. Danner, M. Bauer, P. Phukan, Martin E. Maier, *Eur. J. Org. Chem.* **2005**, 2005, 317-325.
- [2] M. J. O'Donnell, R. L. Polt, *J. Org. Chem.* **1982**, 47, 2663-2666.
- [3] P. L. Pickard, T. L. Tolbert, *Org. Synth.* **1973**, Coll. Vol. 5, 520-522; **1964**, Vol. 44, 51-52.
- [4] I. Arai, I. Muramatsu, *J. Org. Chem.* **1983**, 48, 121-123.
- [5] D. J. Miller, M. Bashir-Uddin Surfraz, M. Akhtar, D. Gani, R. K. Allemann, *Org. Biomol. Chem.* **2004**, 2, 671-688.
- [6] F. Liu, H.-Y. Zha, Z.-J. Yao, *J. Org. Chem.* **2003**, 68, 6679-6684.
- [7] S. Isomura, P. Wirsching, K. D. Janda, *J. Org. Chem.* **2001**, 66, 4115-4121.
- [8] M. Lei, R.-J. Hu, Y.-G. Wang, *Tetrahedron* **2006**, 62, 8928-8932.
- [9] A. J. Geall, I. S. Blagbrough, *Tetrahedron* **2000**, 56, 2449-2460.
- [10] M. Rodriguez, M. Llinares, S. Doulut, A. Heitz, J. Martinez, *Tetrahedron Lett.* **1991**, 32, 923-926.
- [11] T. Ooi, M. Kameda, M. Taniguchi, K. Maruoka, *J. Am. Chem. Soc.* **2004**, 126, 9685-9694.
- [12] M. A. Blaskovich, G. Evindar, N. G. W. Rose, S. Wilkinson, Y. Luo, G. A. Lajoie, *J. Org. Chem.* **1998**, 63, 3631-3646.
- [13] J. M. Andrés, N. de Elena, R. Pedrosa, *Tetrahedron* **2000**, 56, 1523-1531.
- [14] A. Treibs, W. Sutter, *Chem. Ber.* **1951**, 84, 96-100.

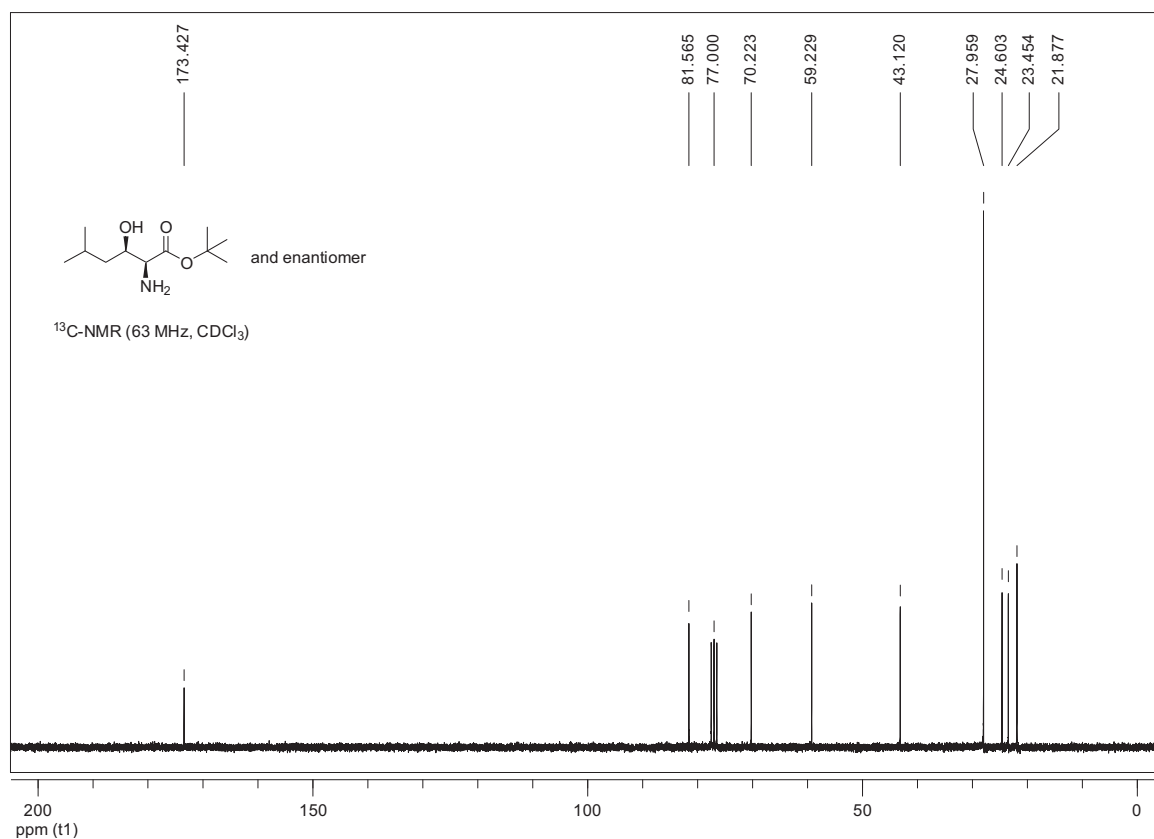
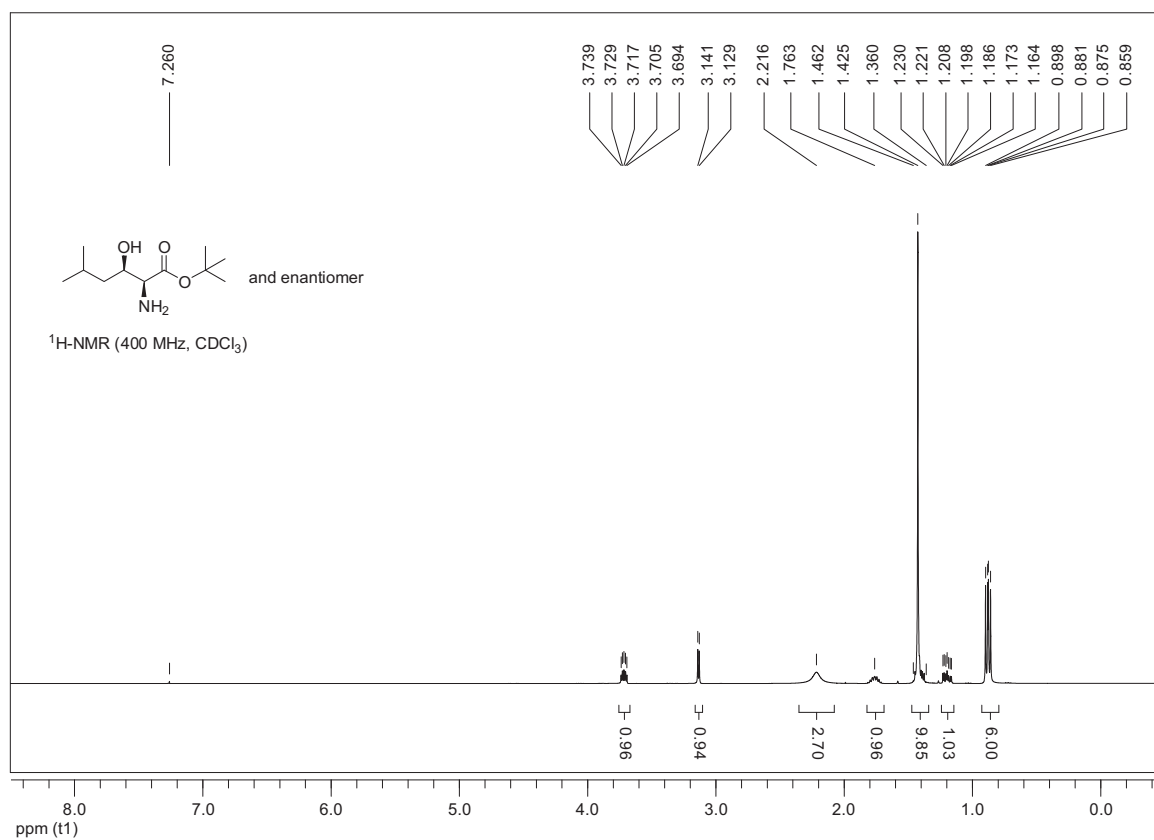
3-(Benzyloxycarbonylamino)propanal (24)

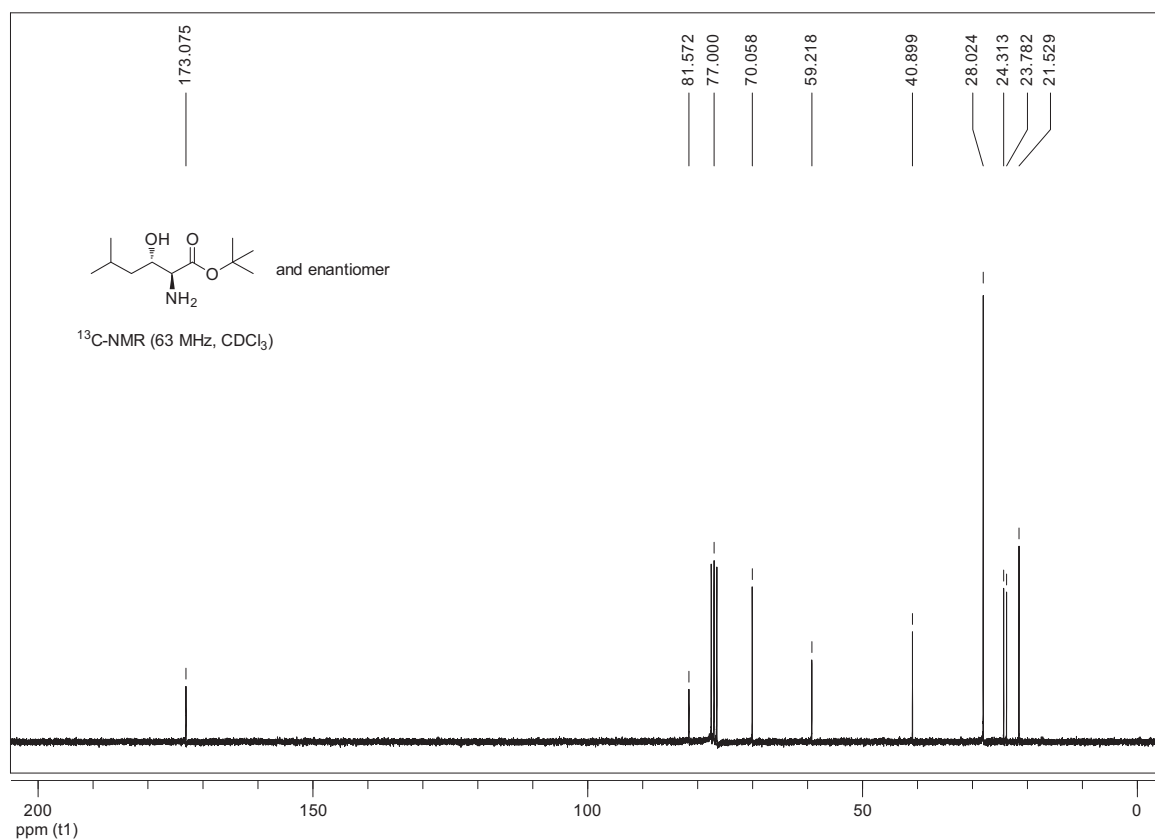
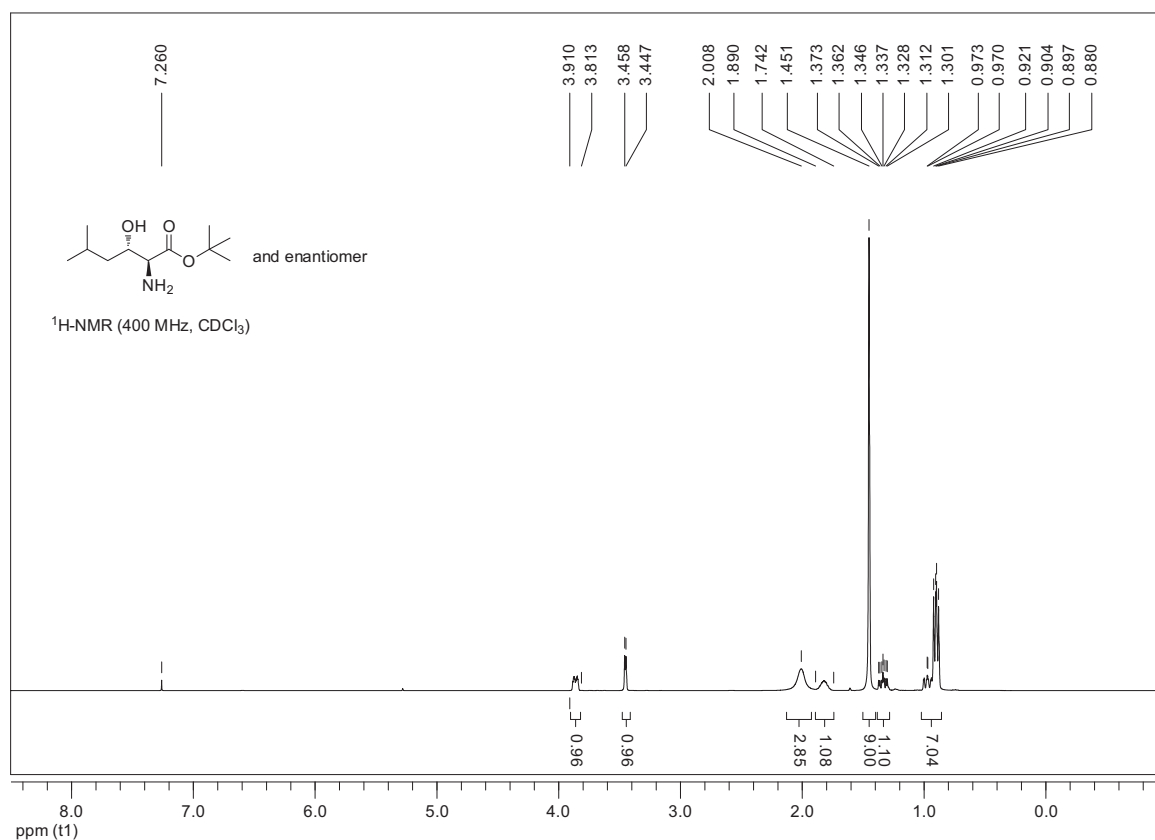


4-Hydroxybutanoic Acid *tert*-Butyl Ester (25)

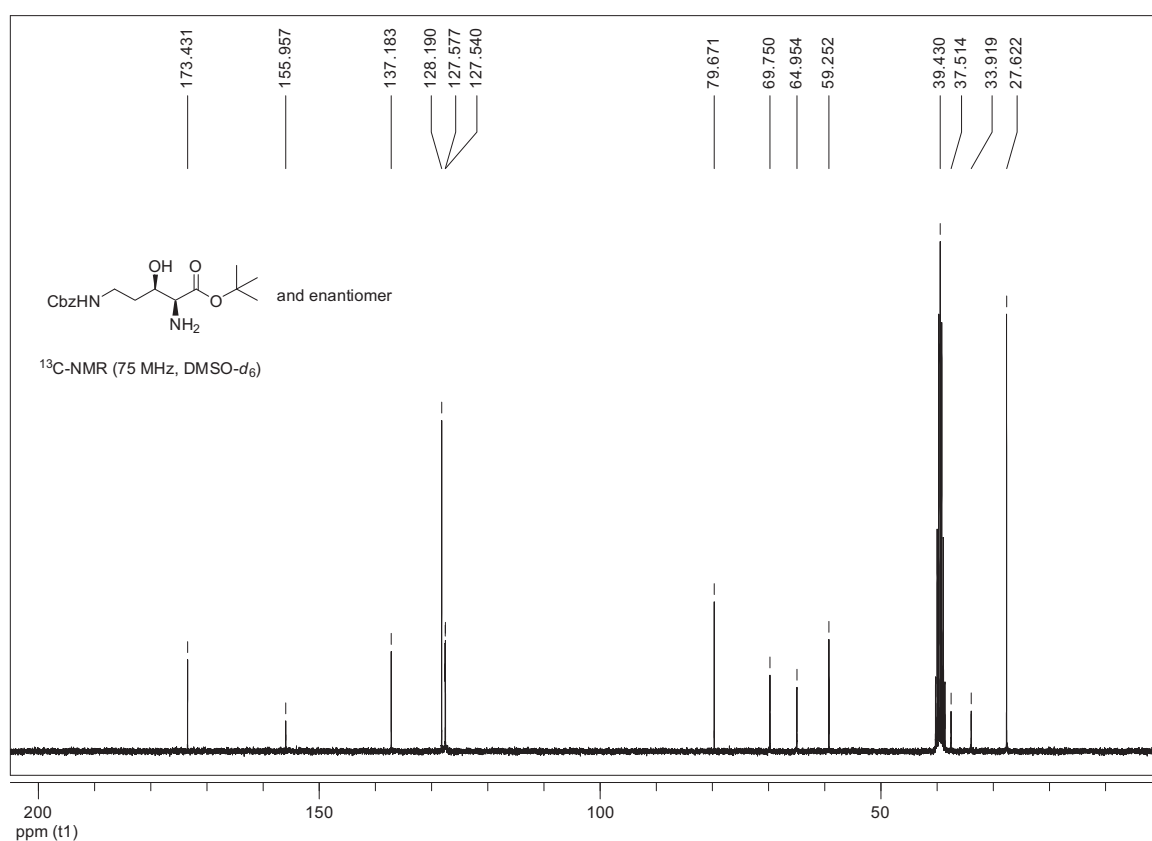
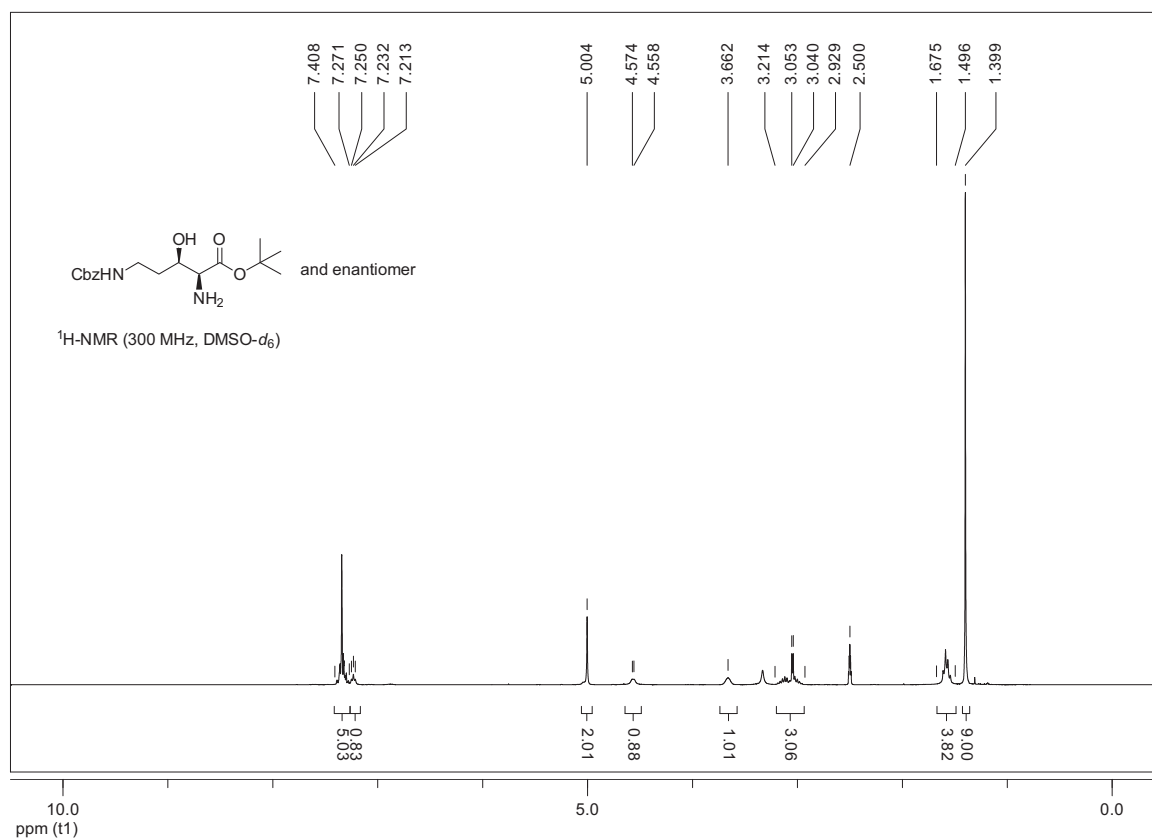


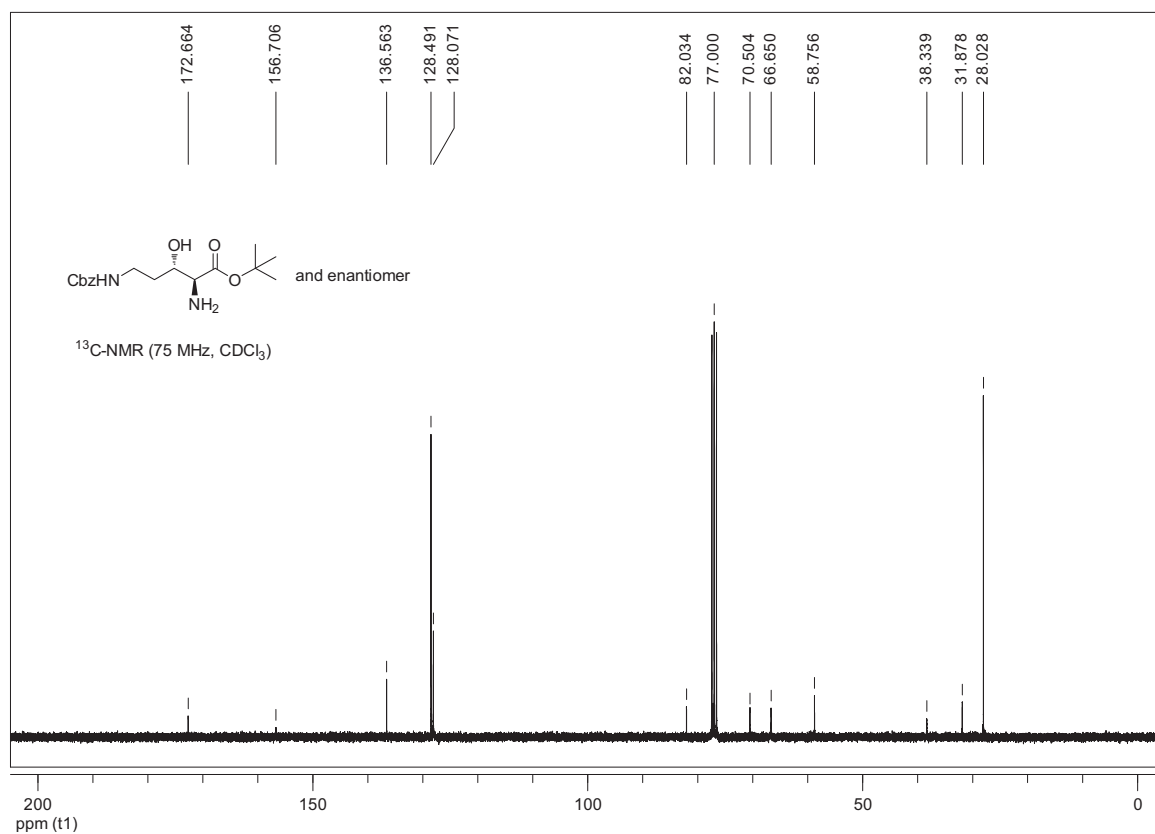
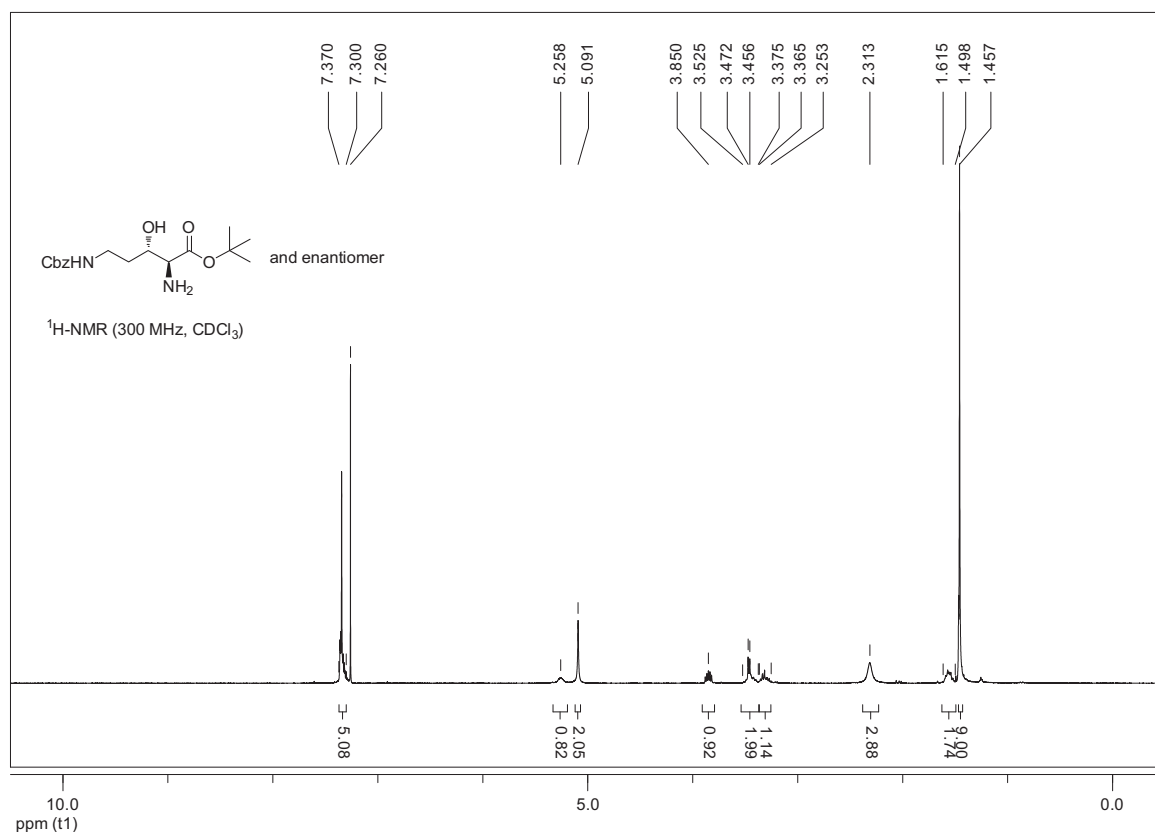
(2S*,3R*)-2-Amino-3-hydroxy-5-methylhexanoic Acid *tert*-Butyl Ester (±)-*syn*-2c

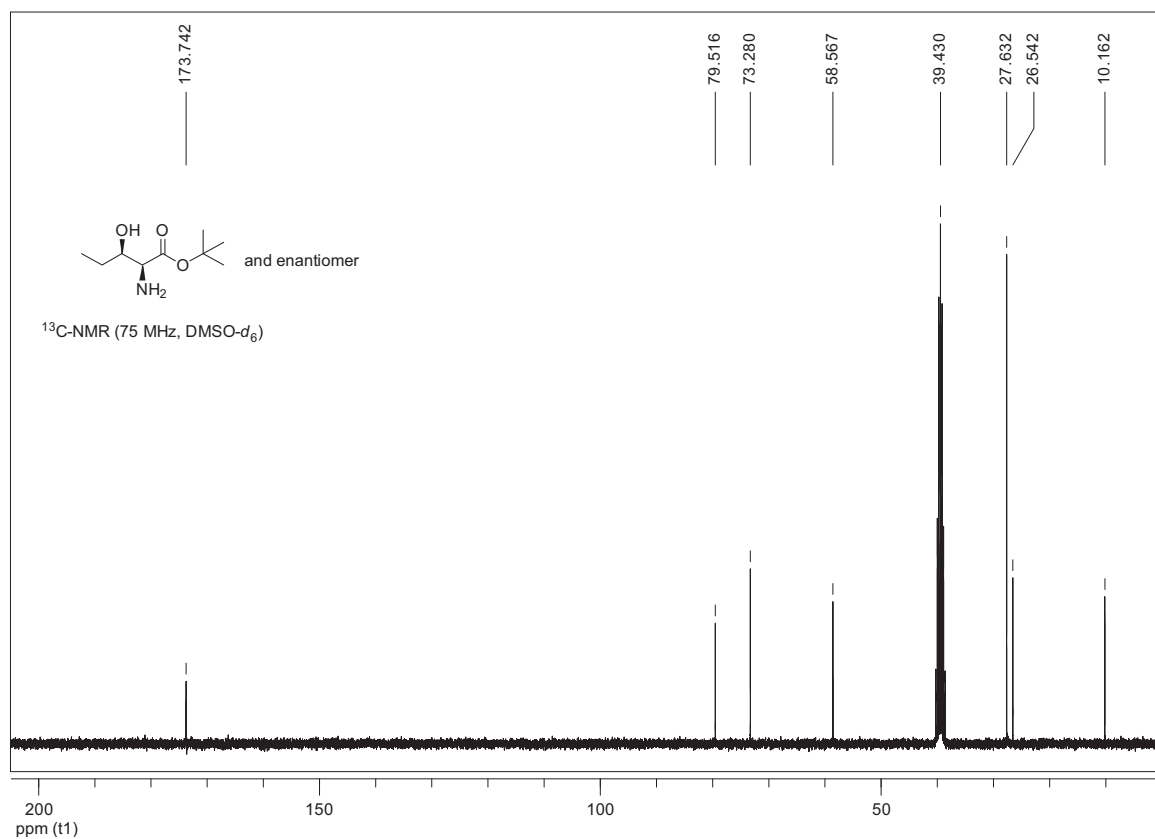
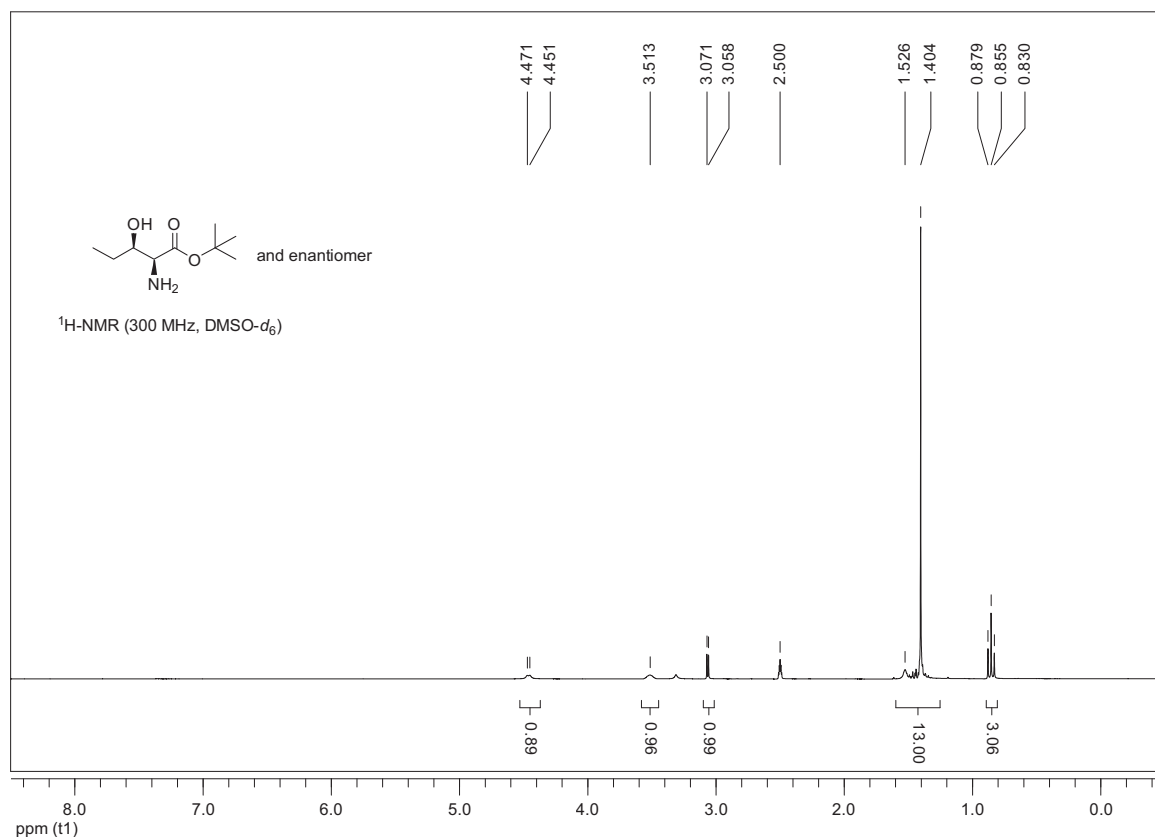


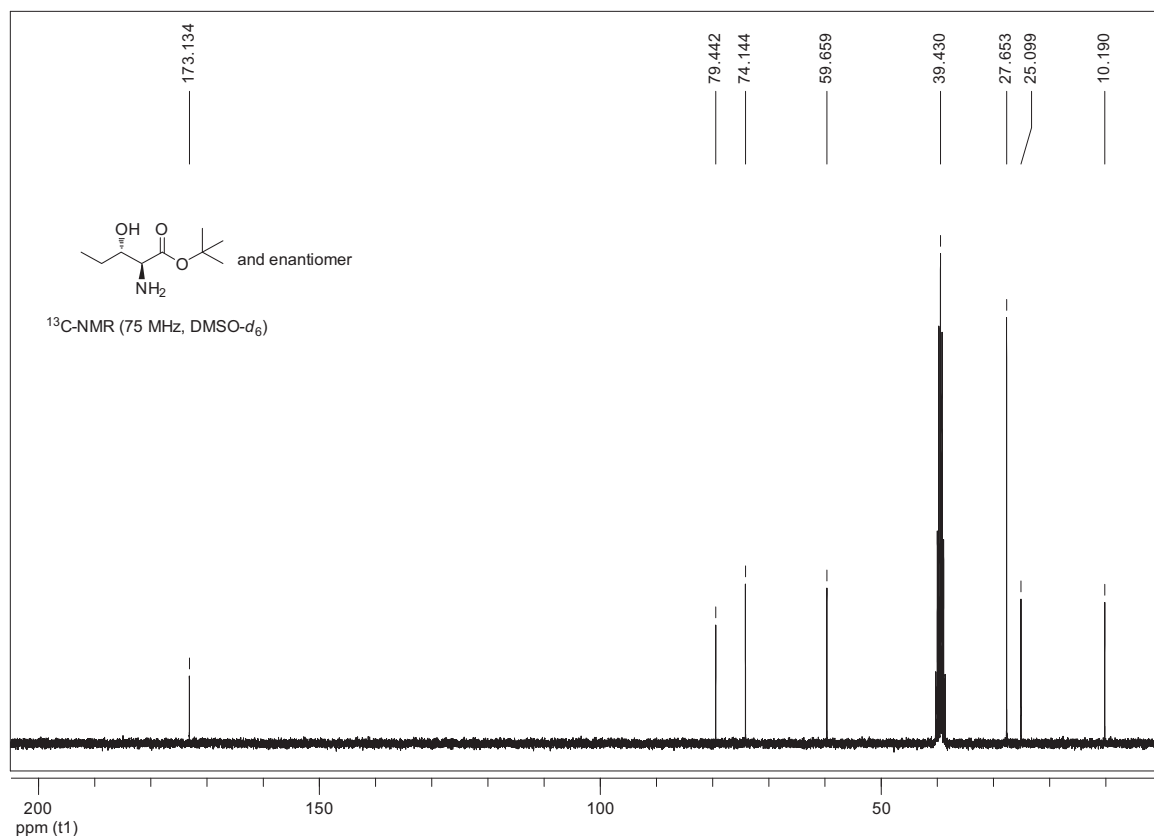
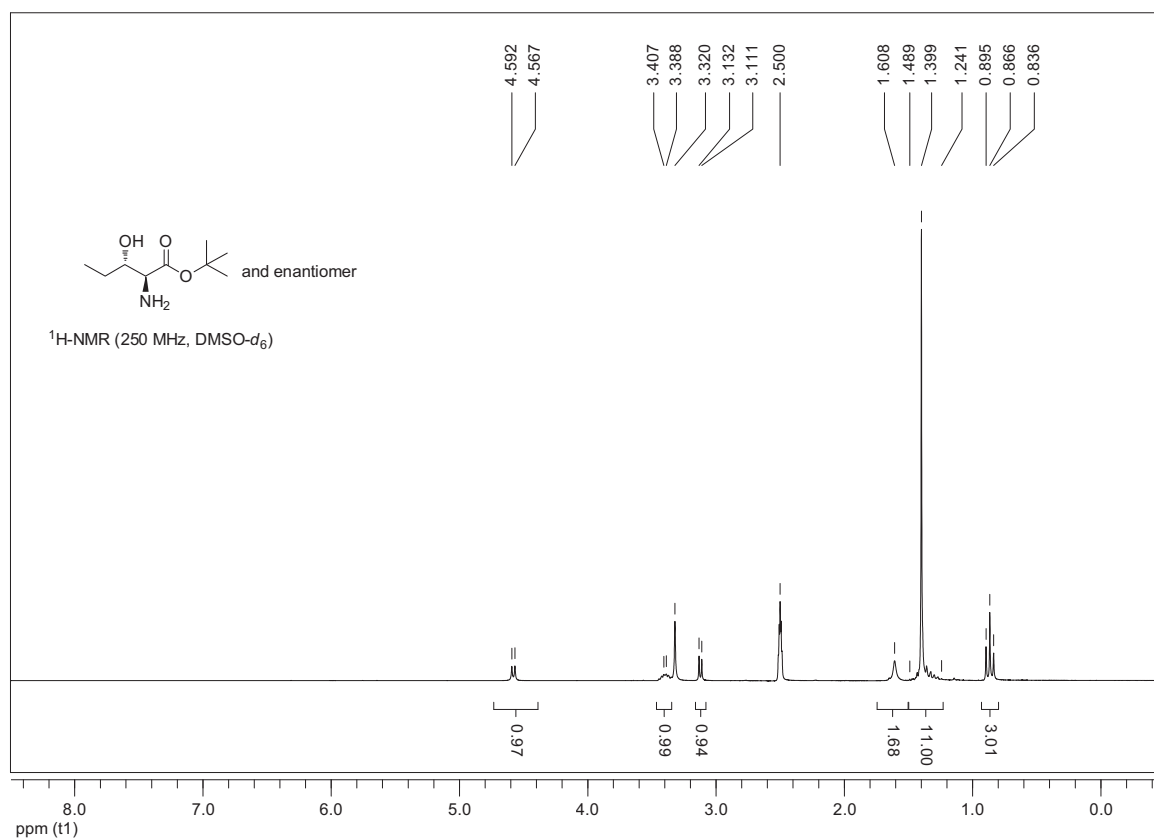
(2S*,3S*)-2-Amino-3-hydroxy-5-methylhexanoic Acid *tert*-Butyl Ester ((±)-*anti*-2c)


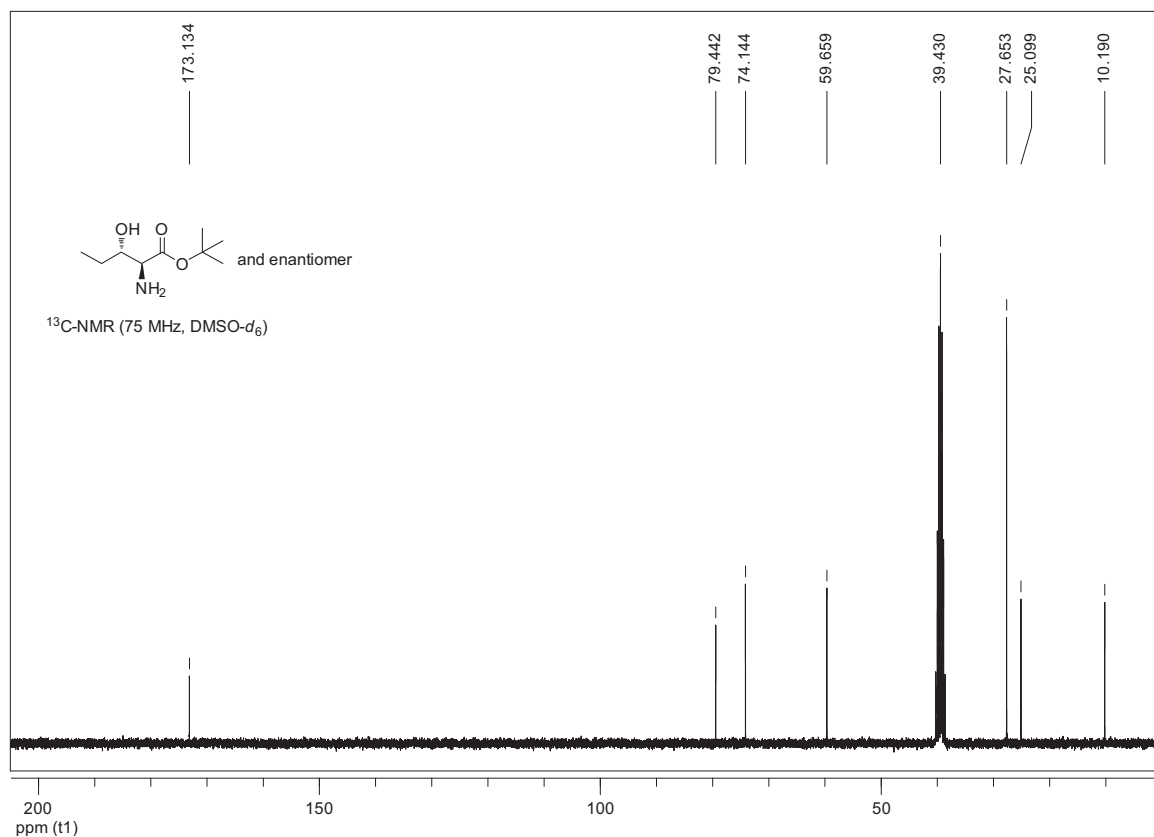
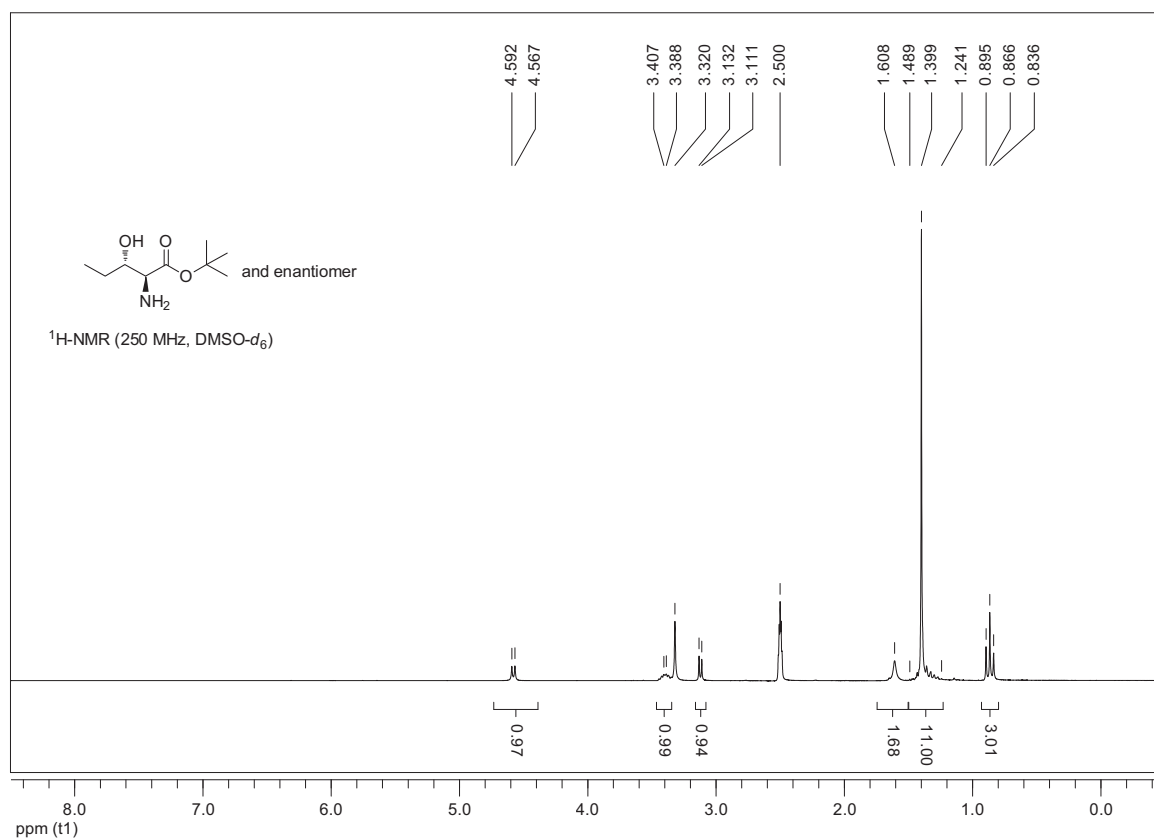
(2*S,3*R**)-2-Amino-5-benzyloxycarbonylamino-3-hydroxypentanoic Acid *tert*-Butyl Ester (±)-*syn*-2a**

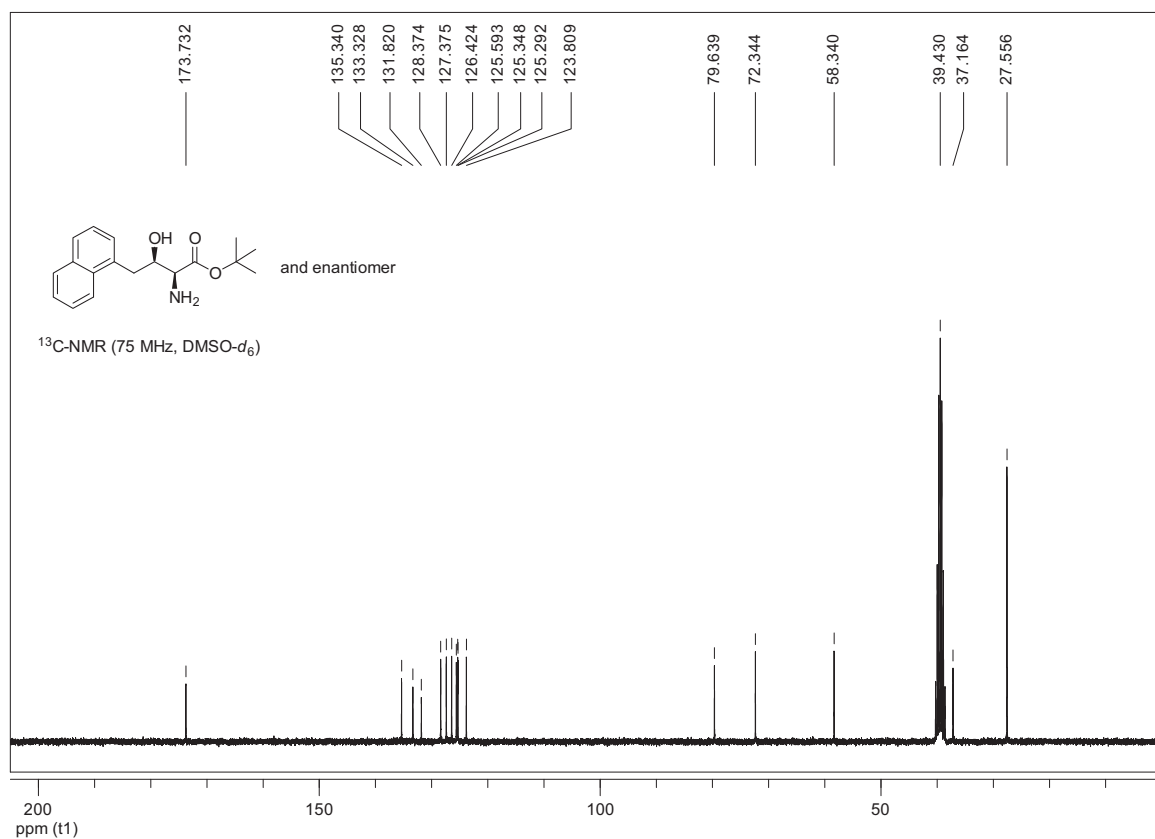
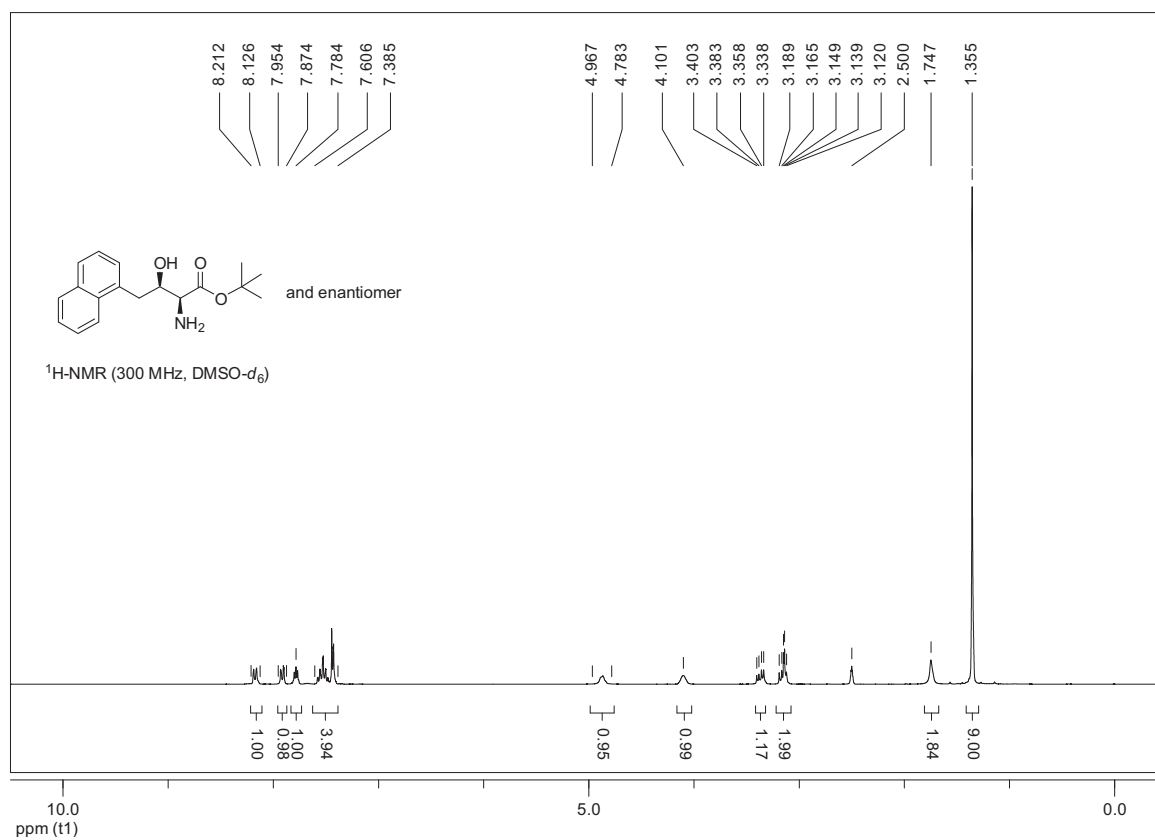


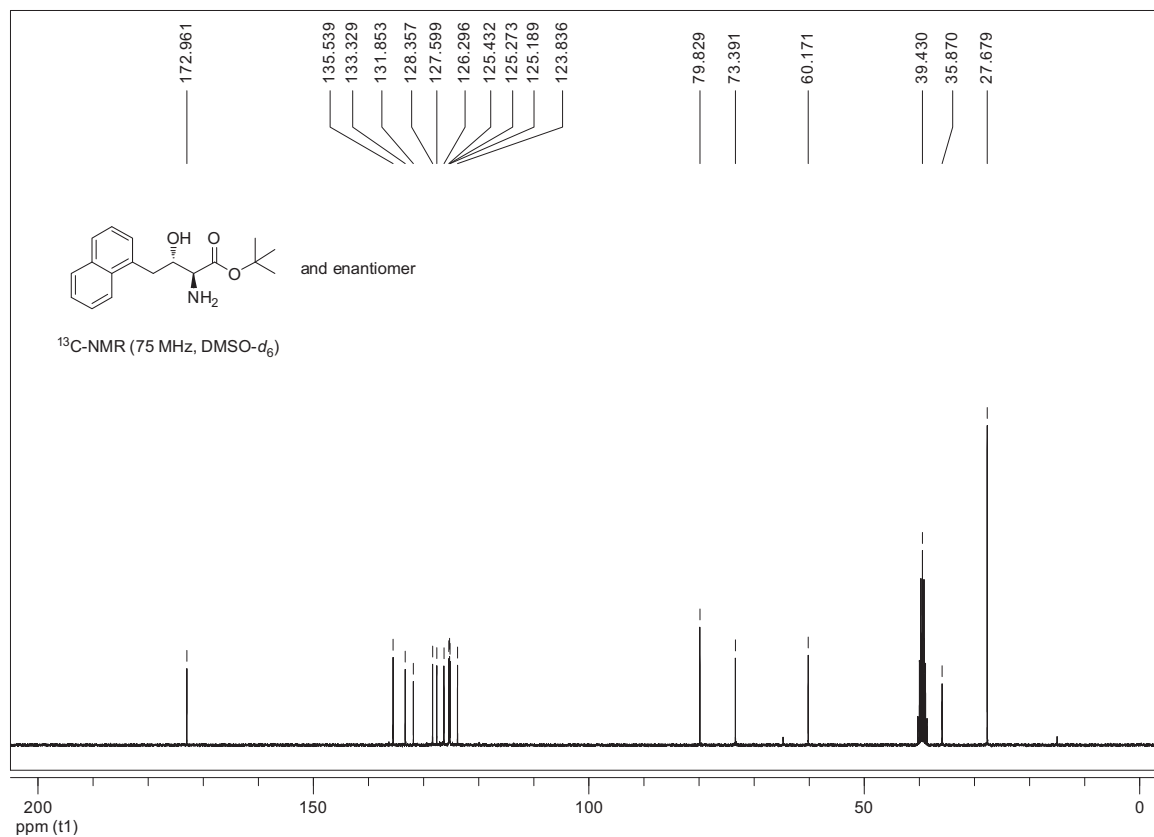
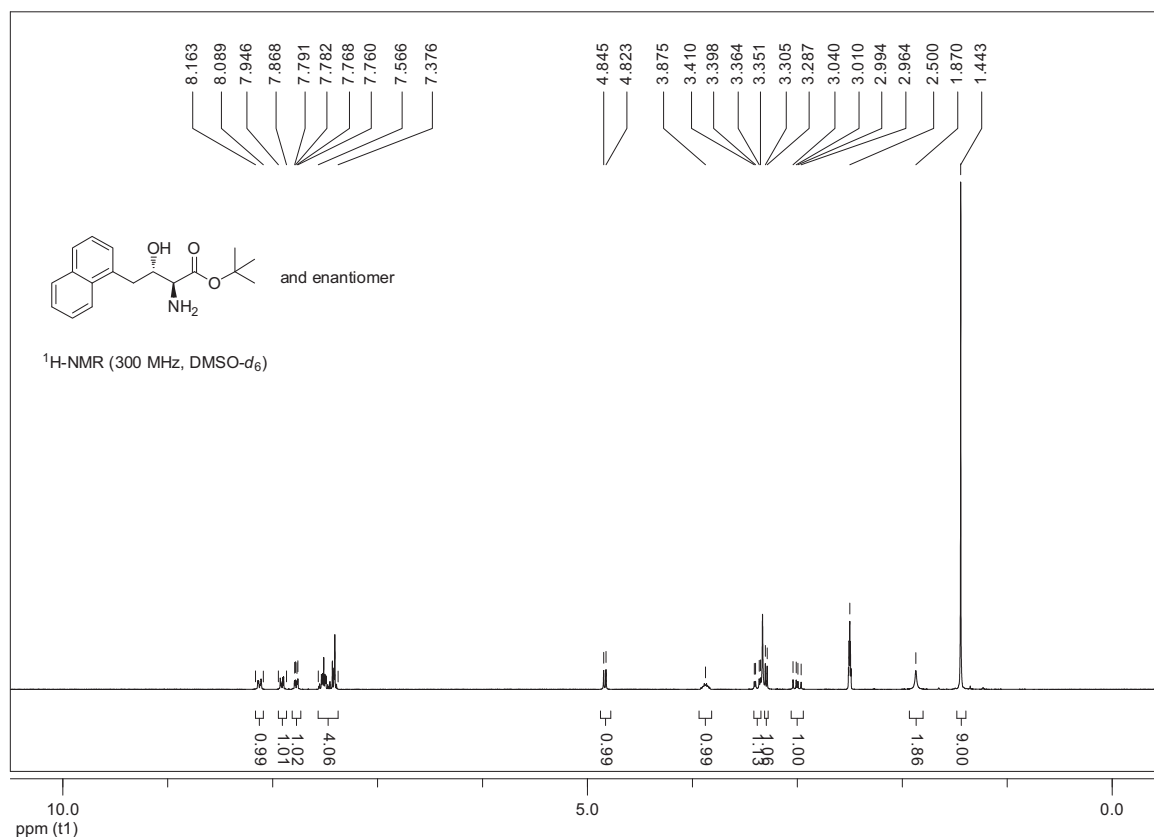
(2S*,3S*)-2-Amino-5-benzoyloxycarbonylamino-3-hydroxypentanoic Acid *tert*-Butyl Ester ((±)-*anti*-2a)

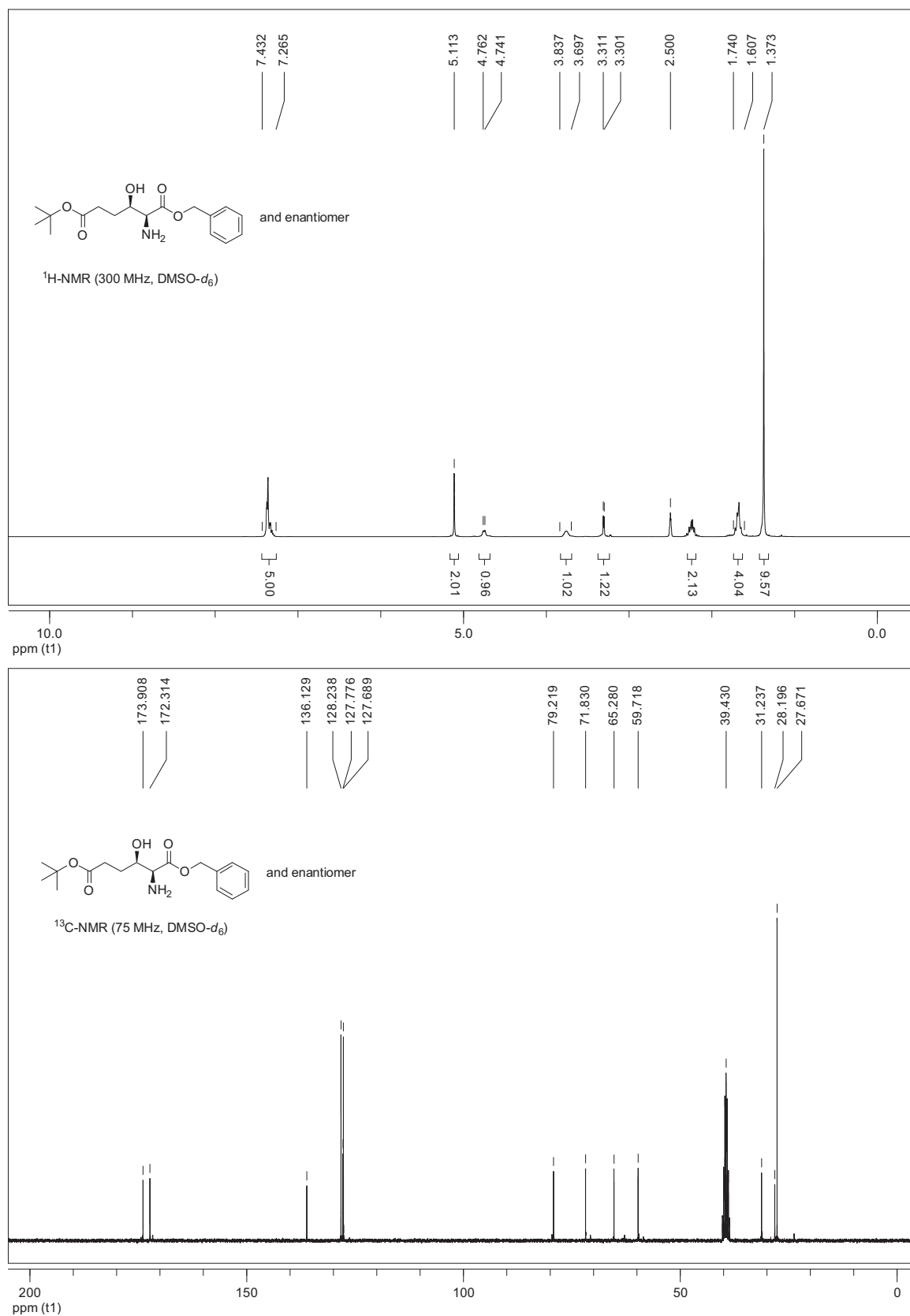
(2S*,3R*)-2-Amino-3-hydroxypentanoic Acid *tert*-Butyl Ester (±)-*syn*-2b)


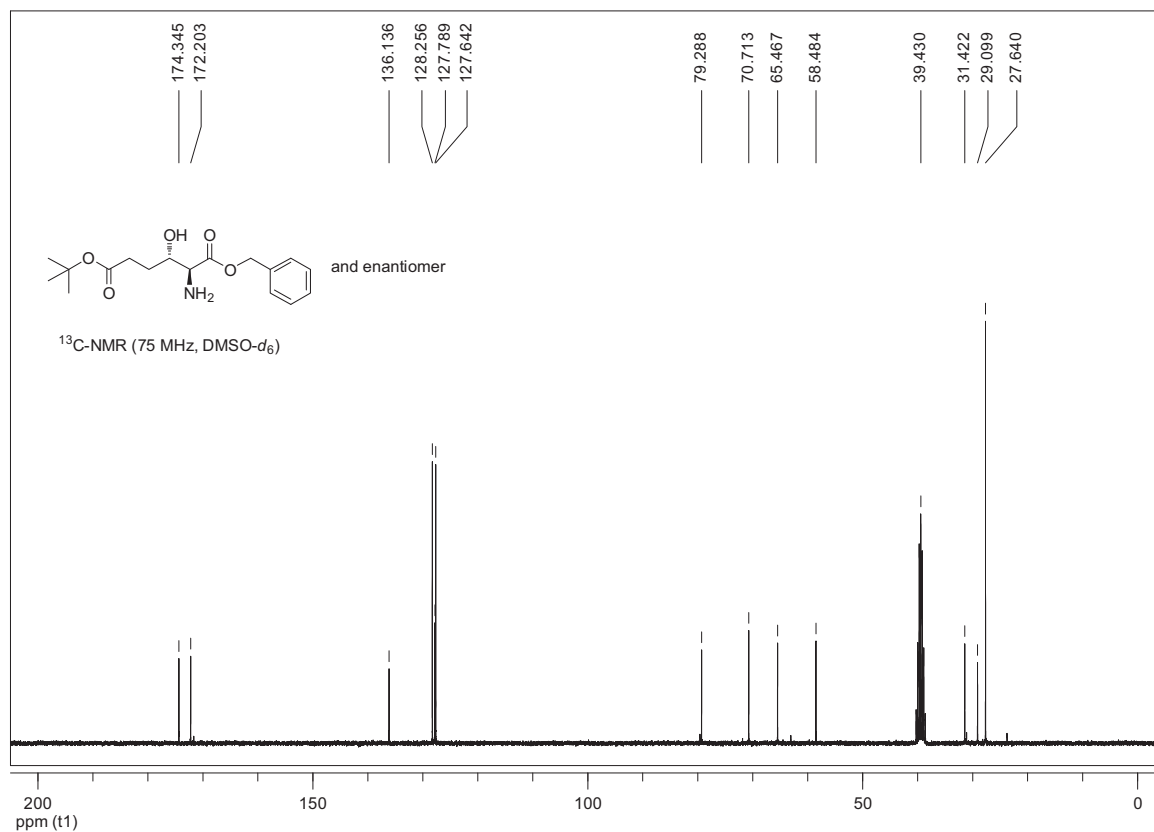
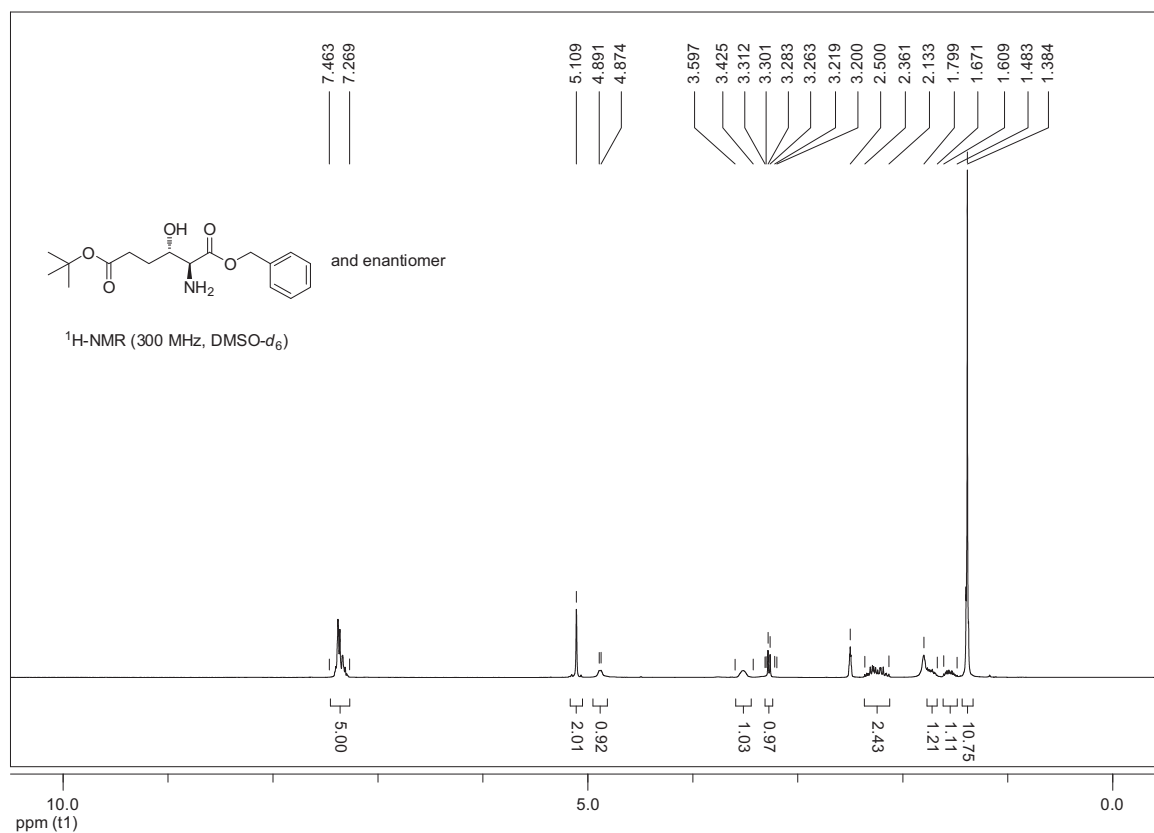
(2S*,3S*)-2-Amino-3-hydroxypentanoic Acid *tert*-Butyl Ester (±)-*anti*-2b)


(2S*,3S*)-2-Amino-3-hydroxypentanoic Acid *tert*-Butyl Ester (±)-*anti*-2b)


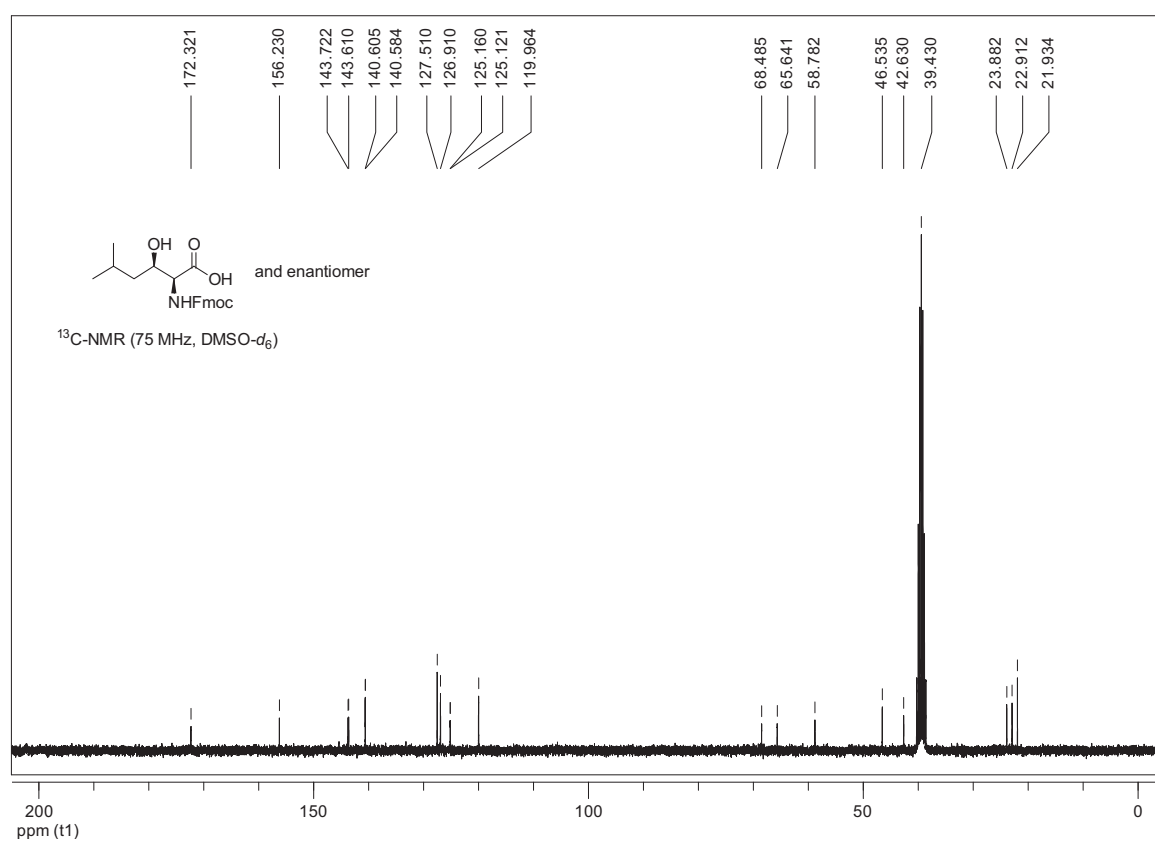
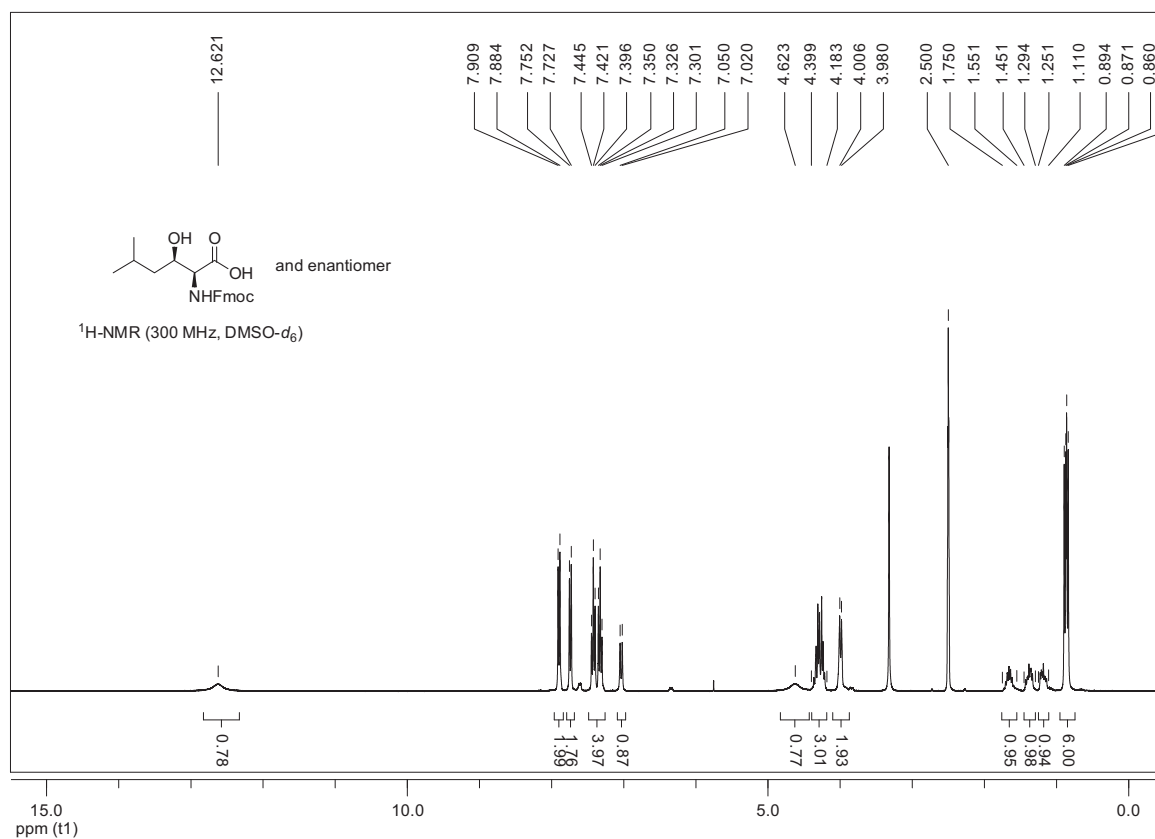
(2S*,3R*)-2-Amino-3-hydroxy-4-(1-naphthyl)butanoic Acid *tert*-Butyl Ester ((±)-*syn*-2d)

(2S*,3S*)-2-Amino-3-hydroxy-4-(1-naphthyl)butanoic Acid *tert*-Butyl Ester ((±)-*anti*-2d)

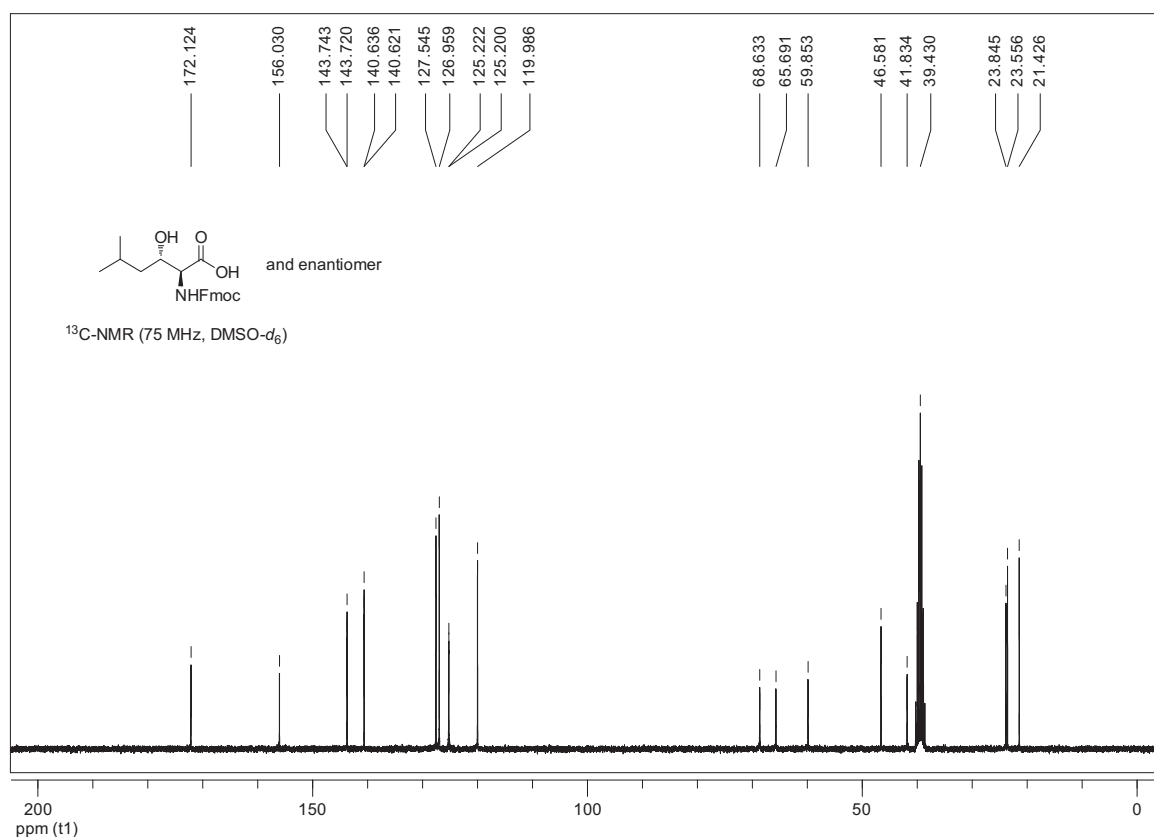
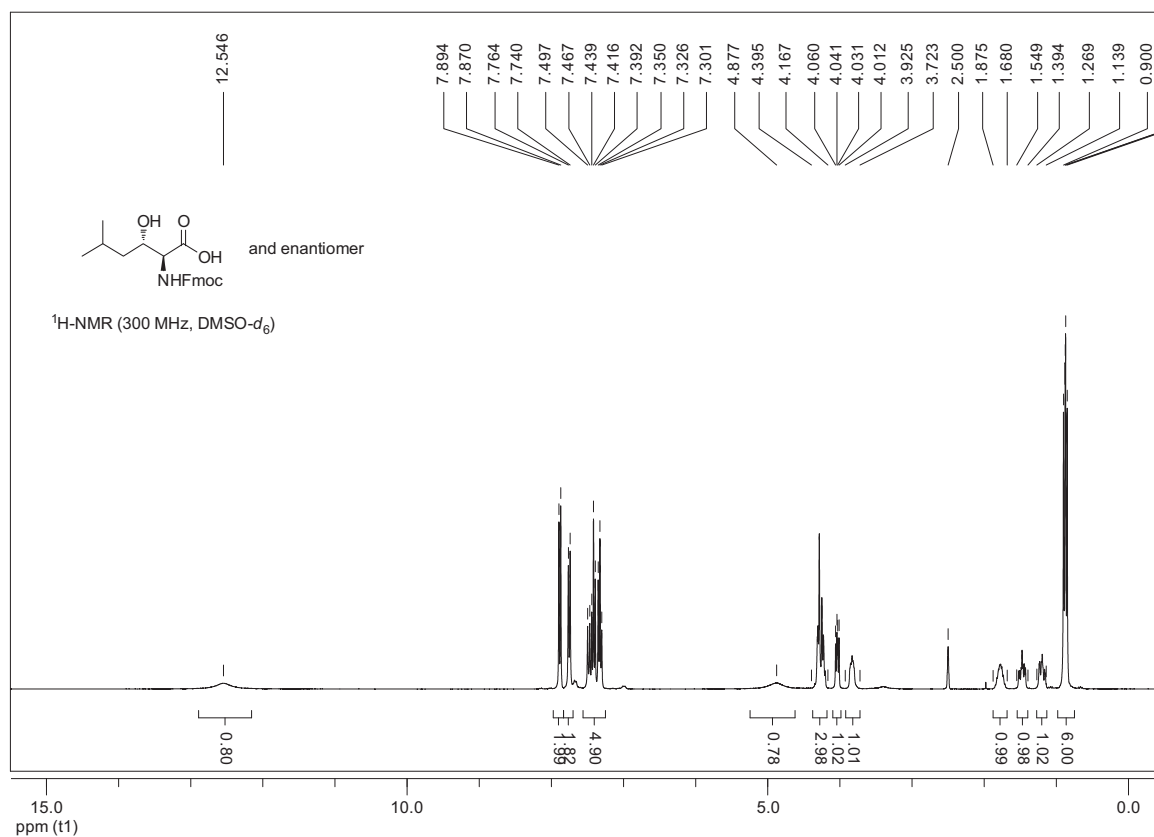
(2*S,3*R**)-2-Amino-3-hydroxyhexanedioic Acid 1-Benzyl-6-*tert*-butyl Ester ((±)-*syn*-17)**


(2S*,3S*)-2-Amino-3-hydroxyhexanedioic Acid 1-Benzyl-6-*tert*-butyl Ester ((±)-*anti*-17)


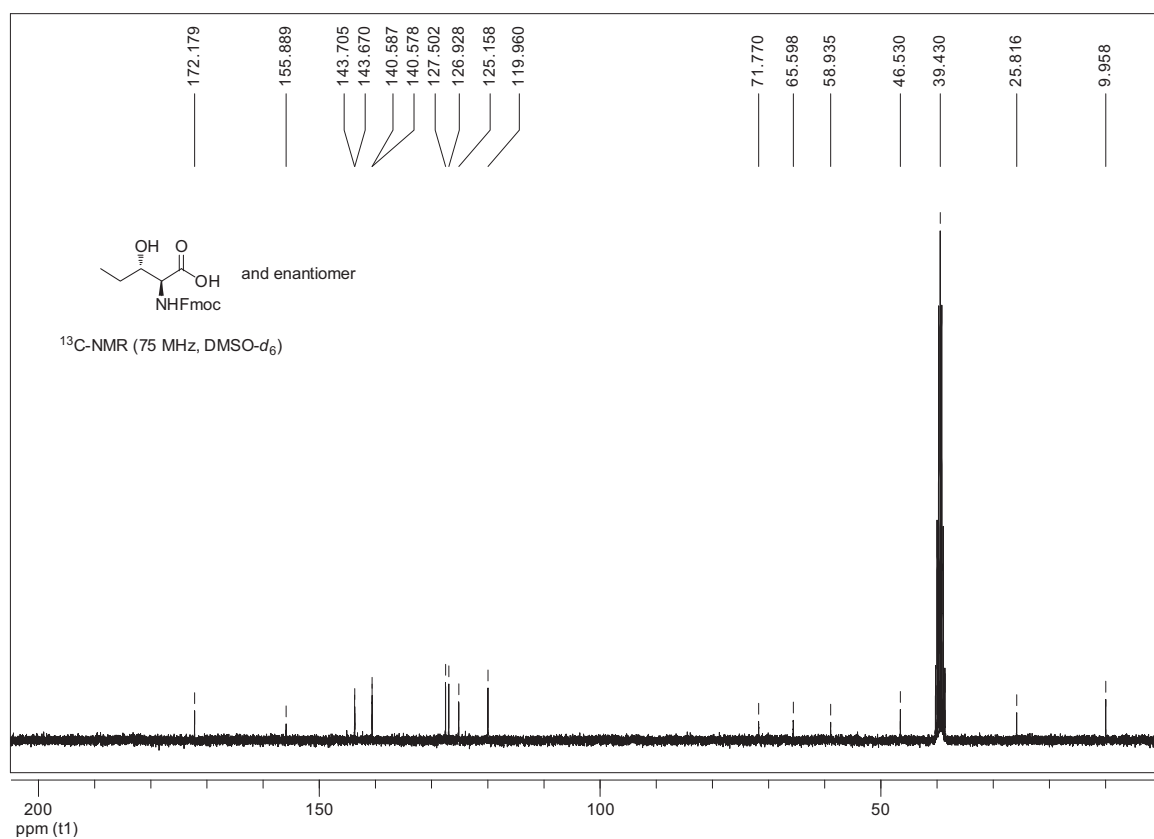
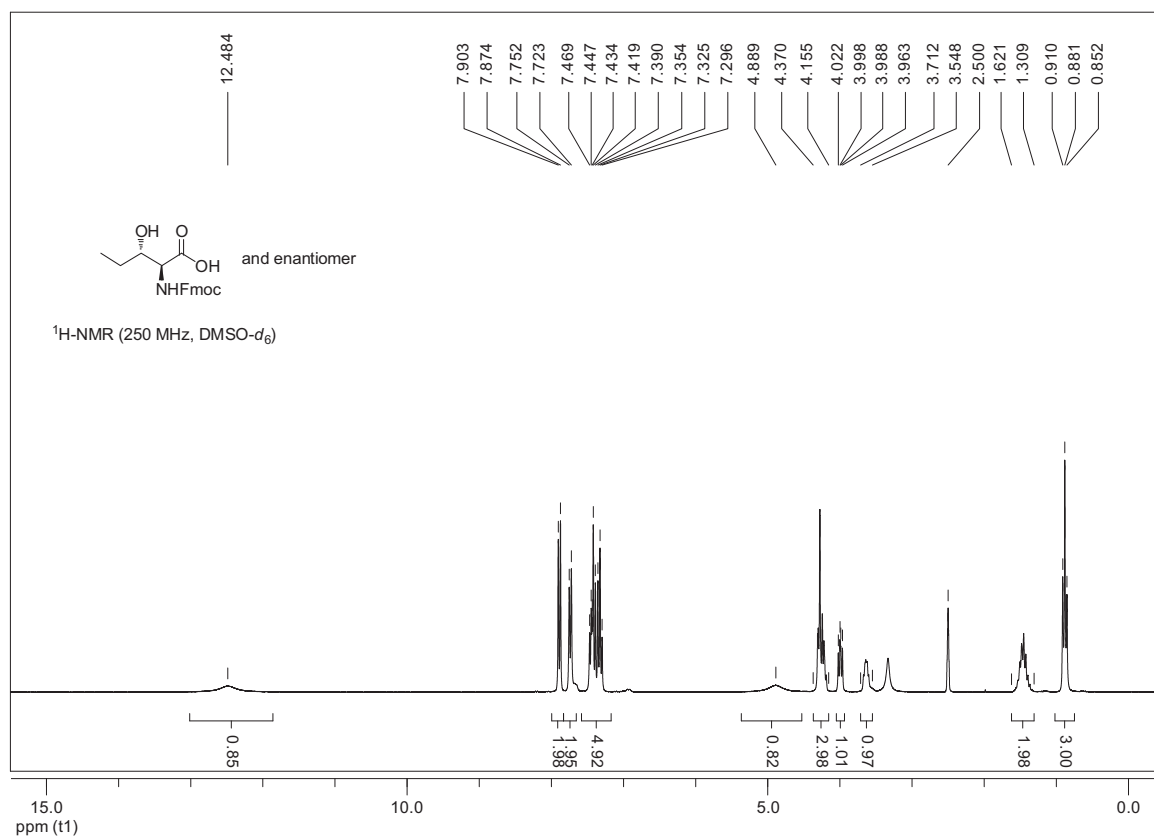
(2*S,3*R**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-5-methylhexanoic Acid ((±)-*syn*-3c)**

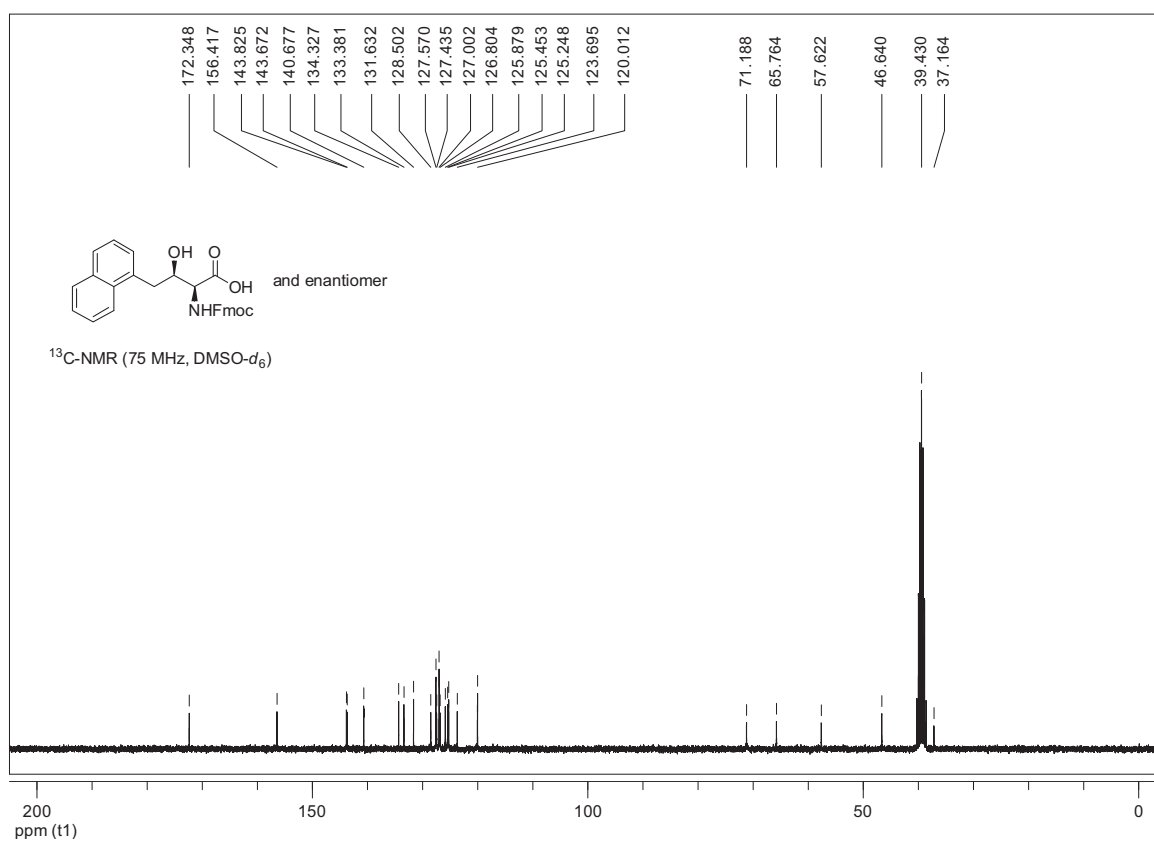
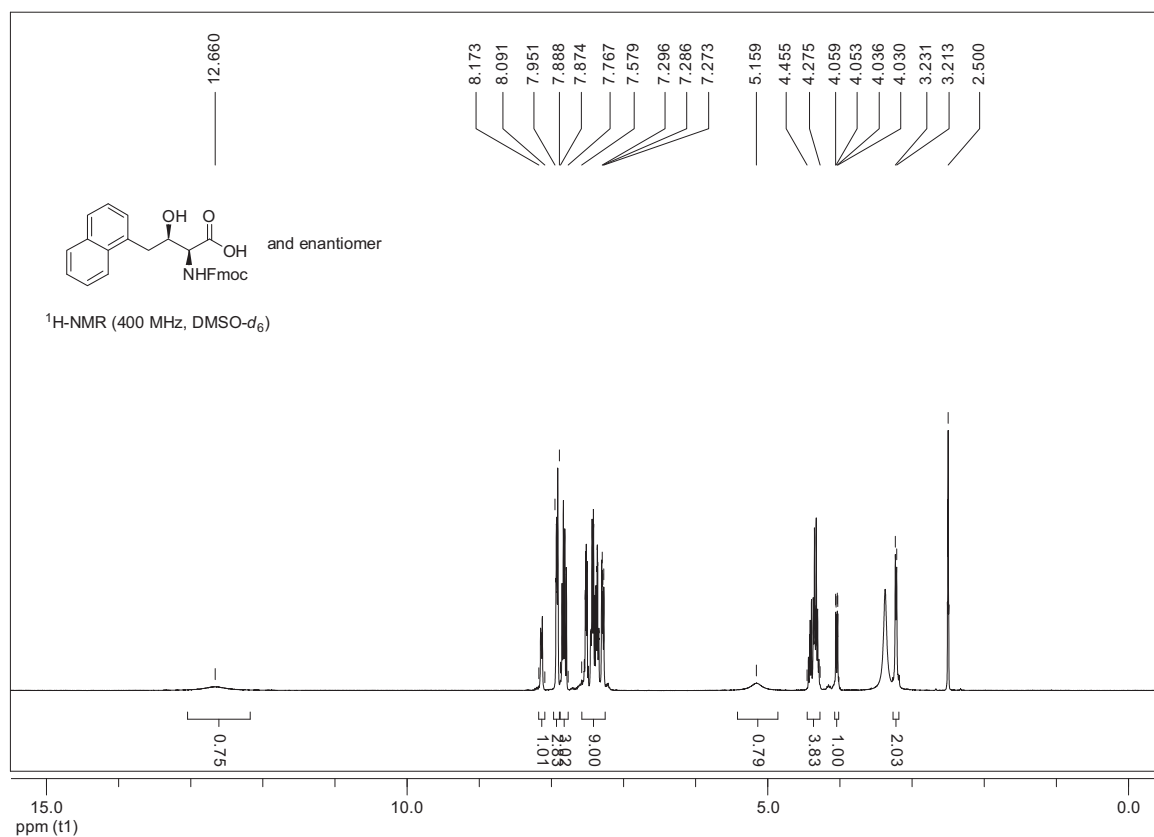


(2*S,3*S**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-5-methylhexanoic Acid
(±)-*anti*-3c**

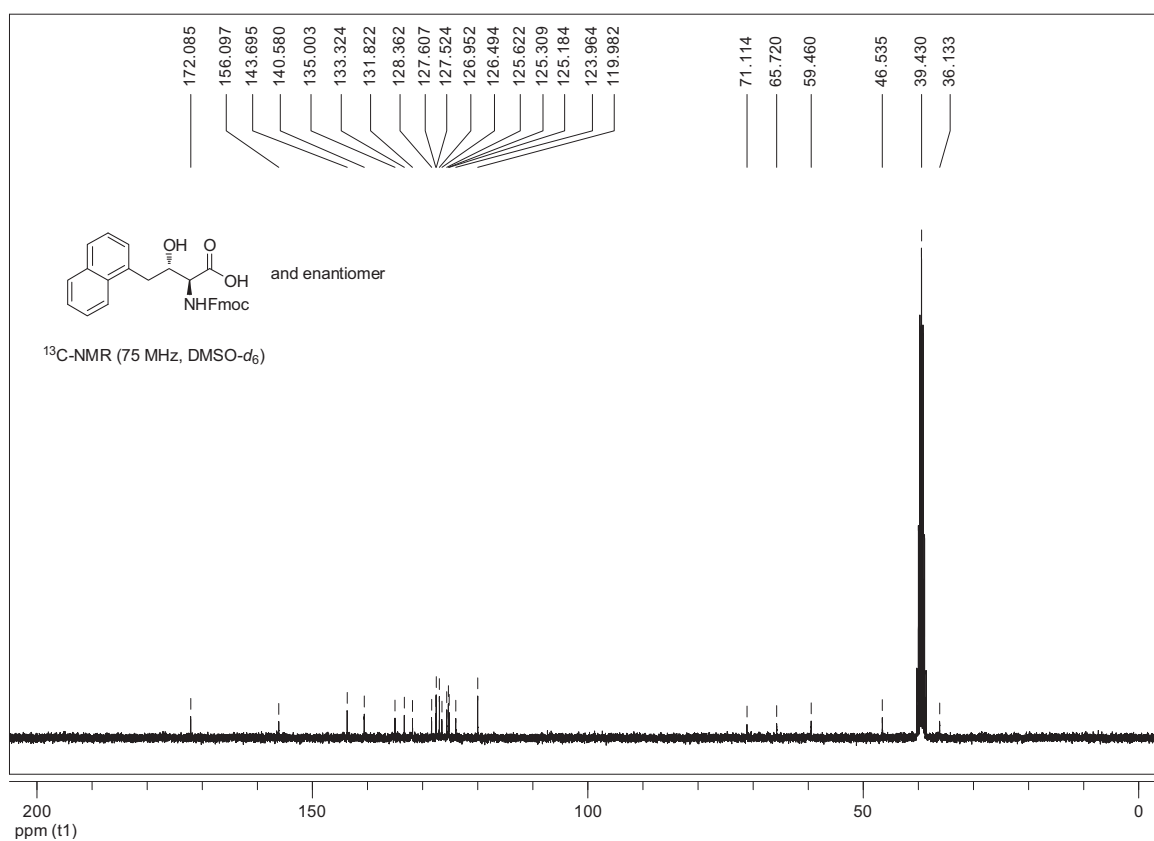
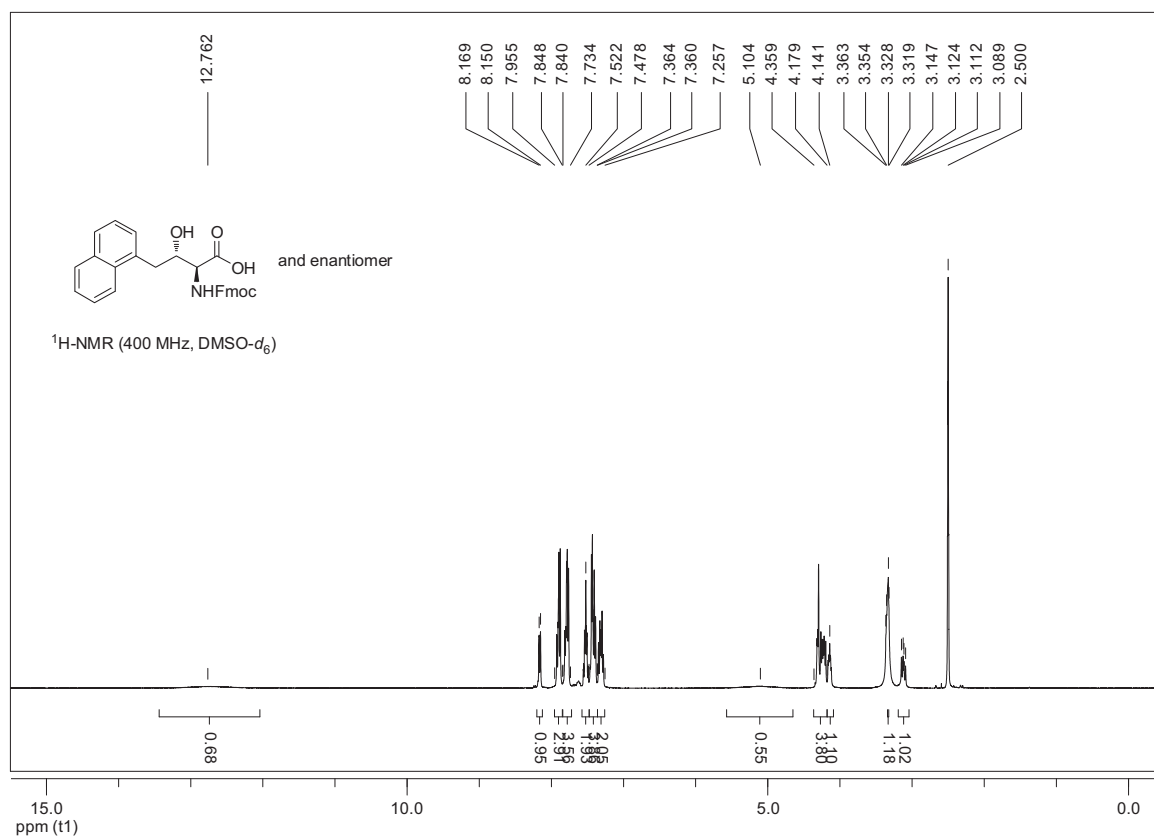


(2*S,3*S**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxypentanoic Acid ((±)-*anti*-3b)**

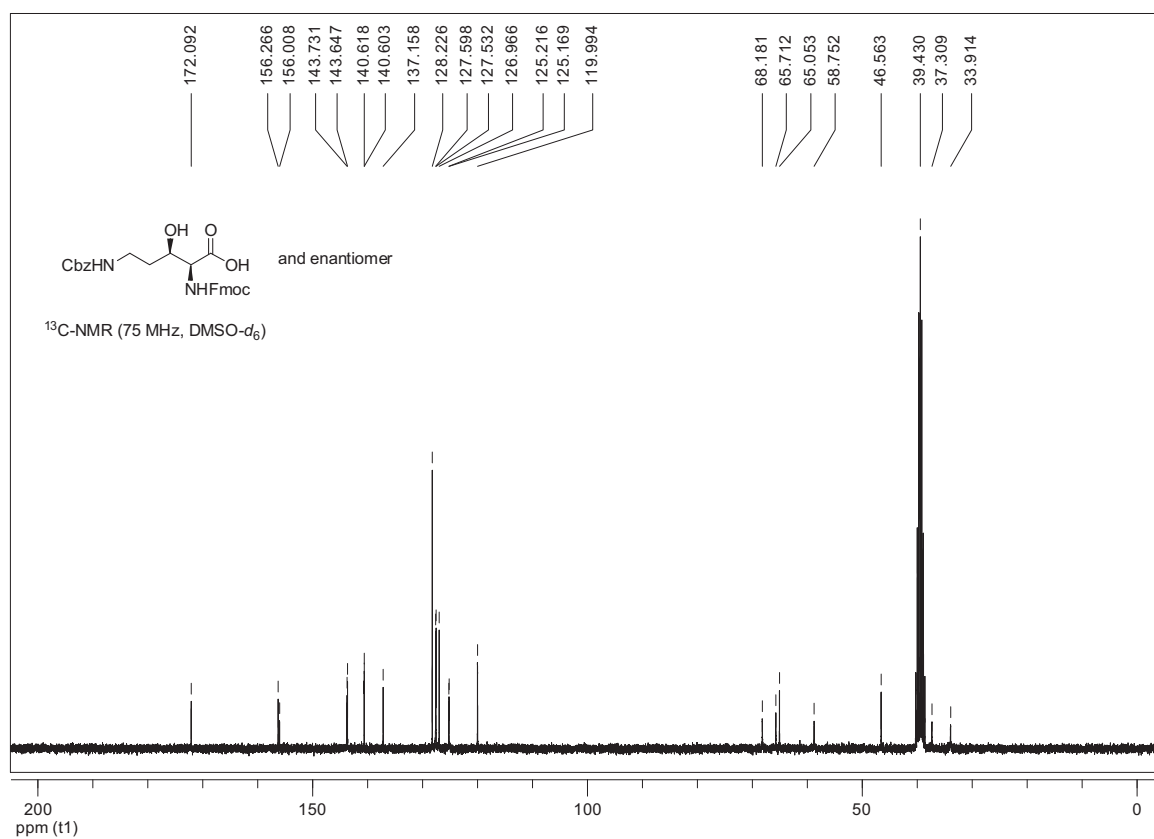
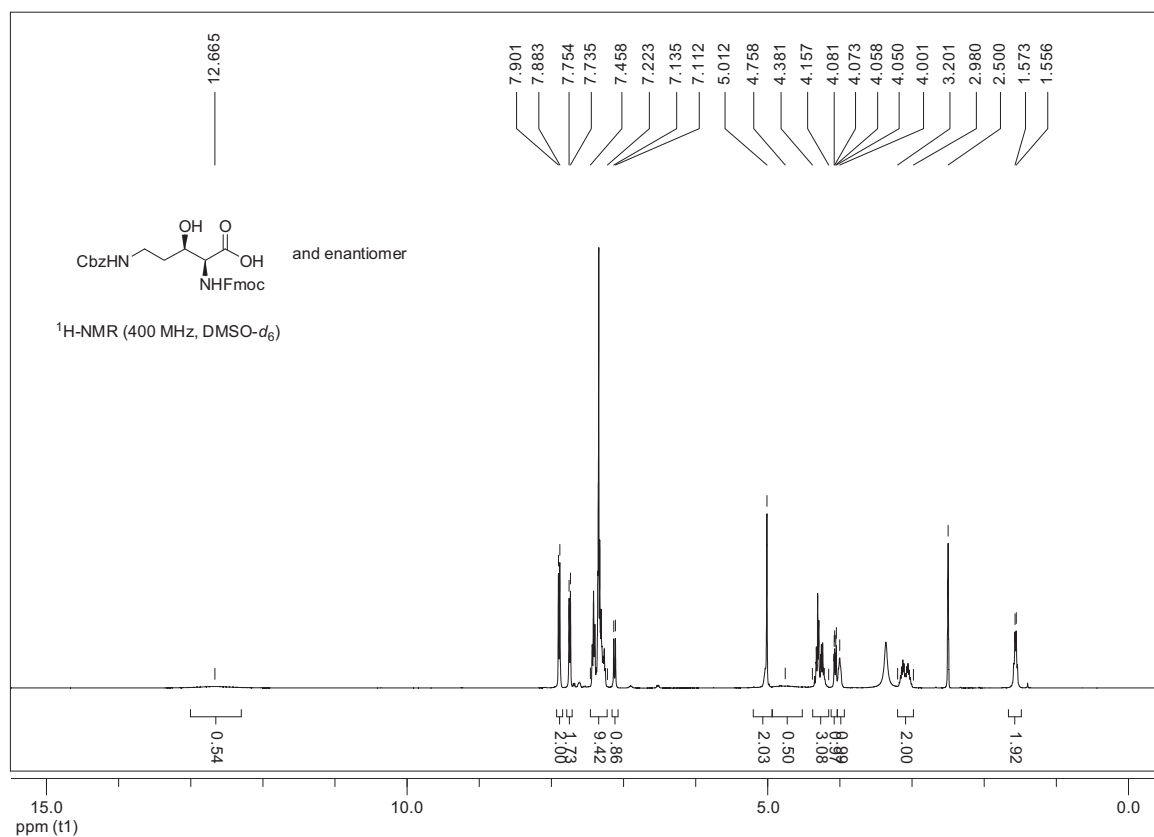


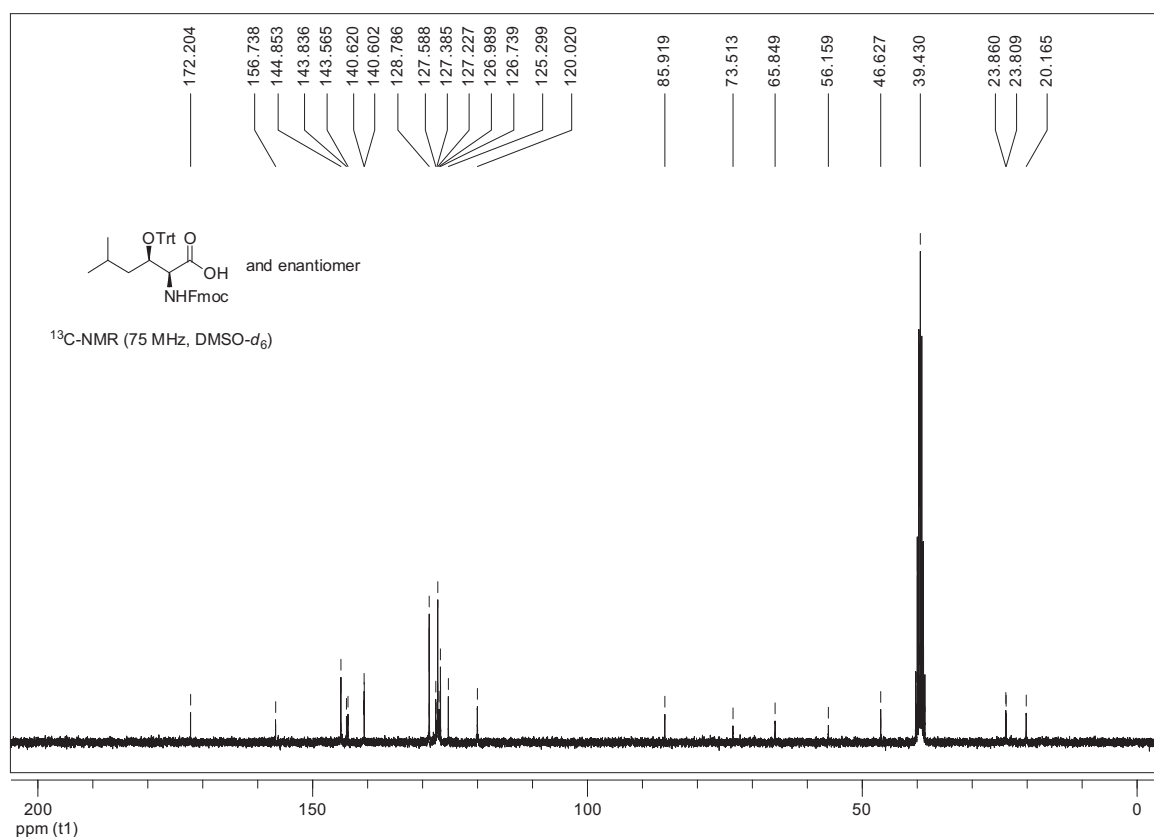
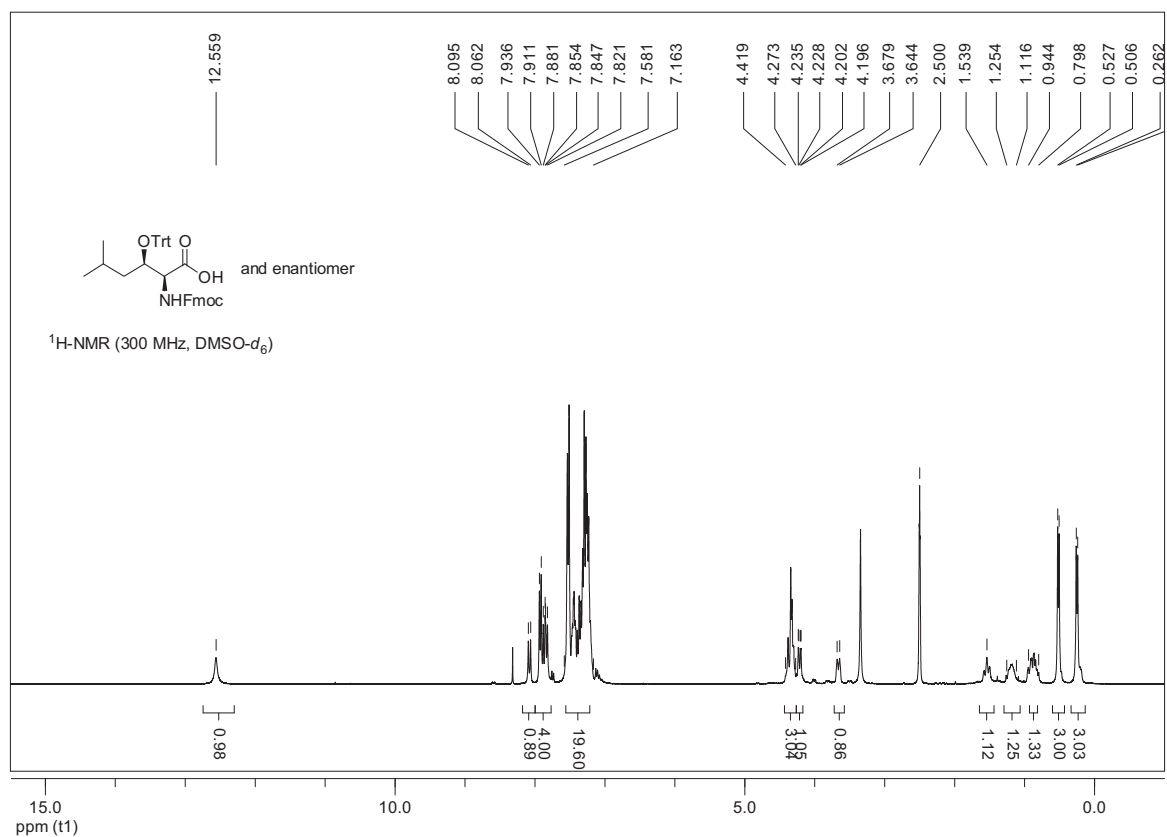
(2*S,3*R**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-4-(1-naphthyl)butanoic Acid (±)-*syn*-3d**

(2*S,3*S**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-3-hydroxy-4-(1-naphthyl)butanoic Acid ((±)-*anti*-3d)**

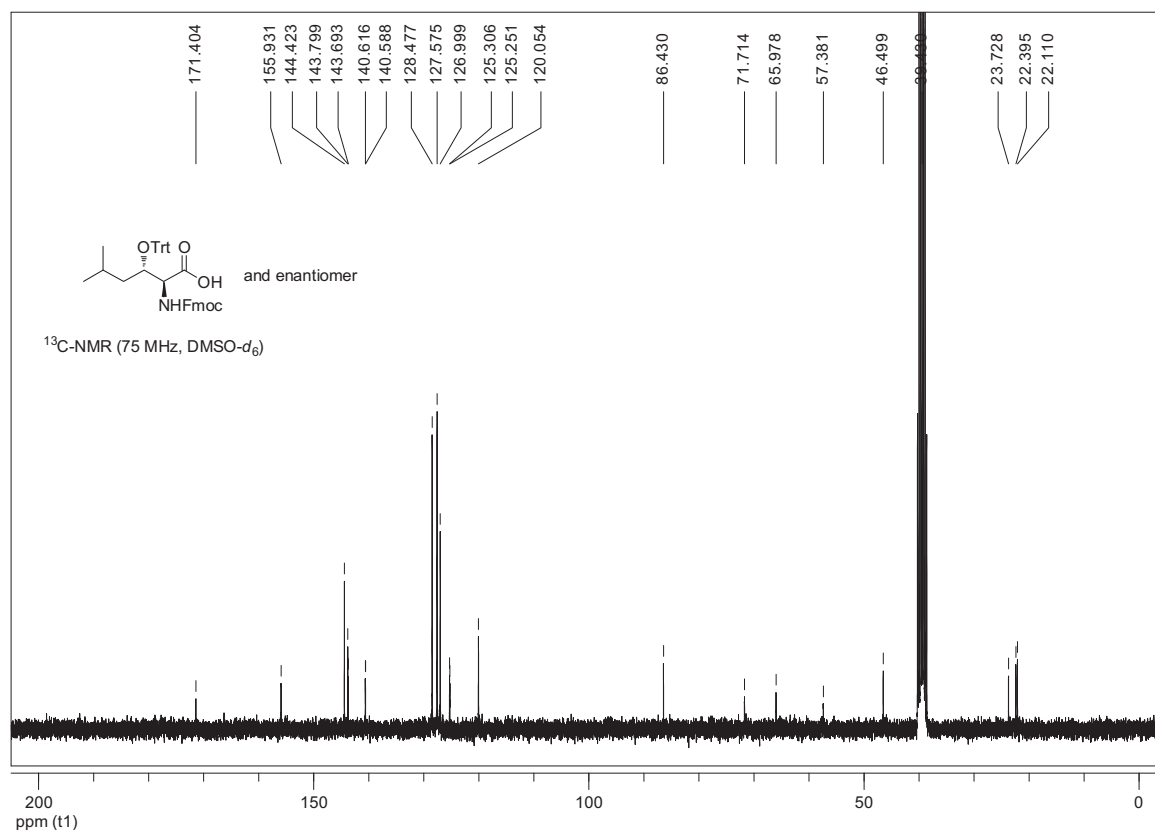
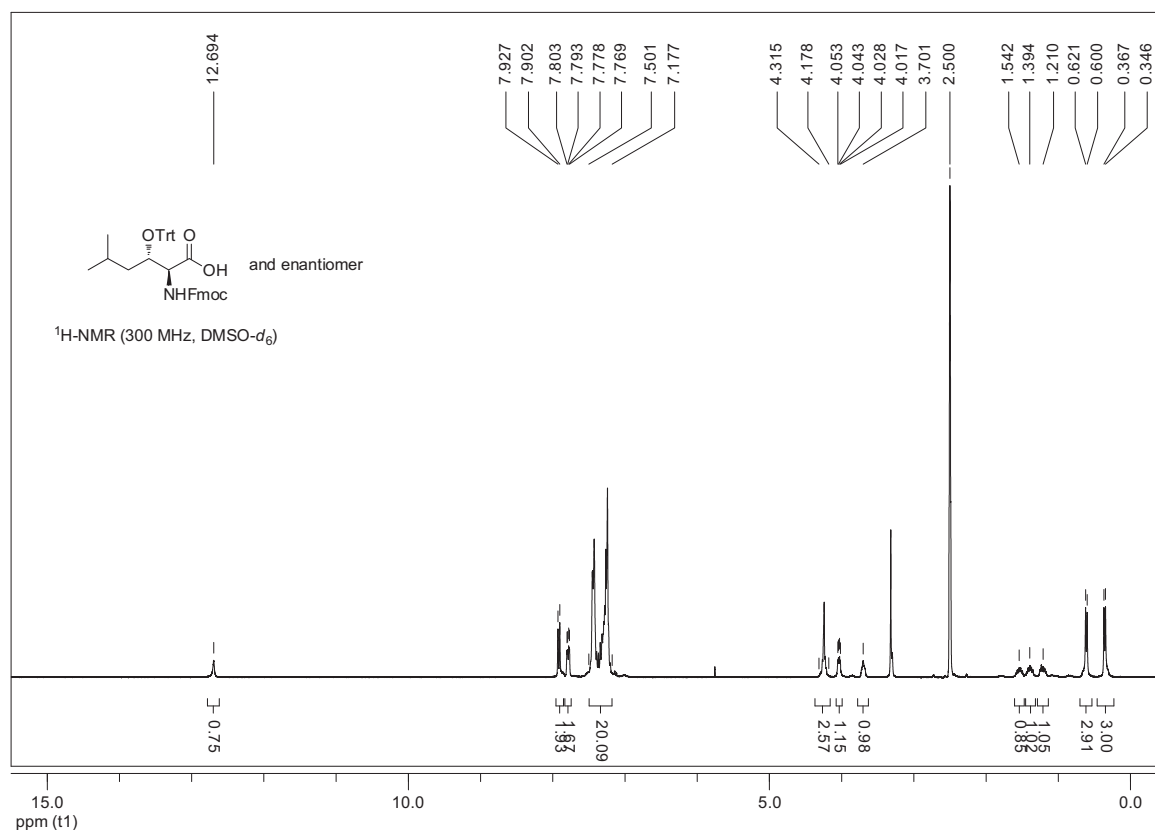


(2*S,3*R**)-5-Benzylloxycarbonylamino-2-(9*H*-fluoren-9-ylmethoxycarbonylamino)-3-hydroxypentanoic Acid ((±)-*syn*-3a)**

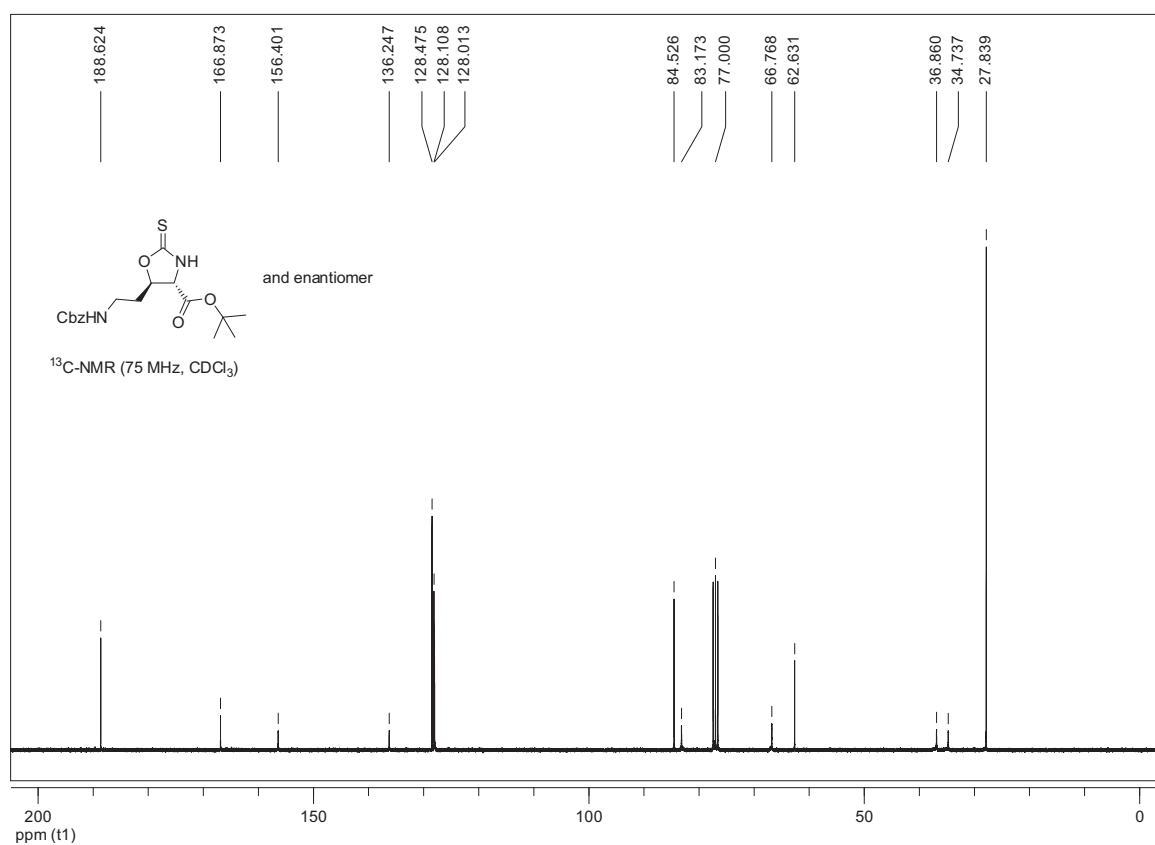
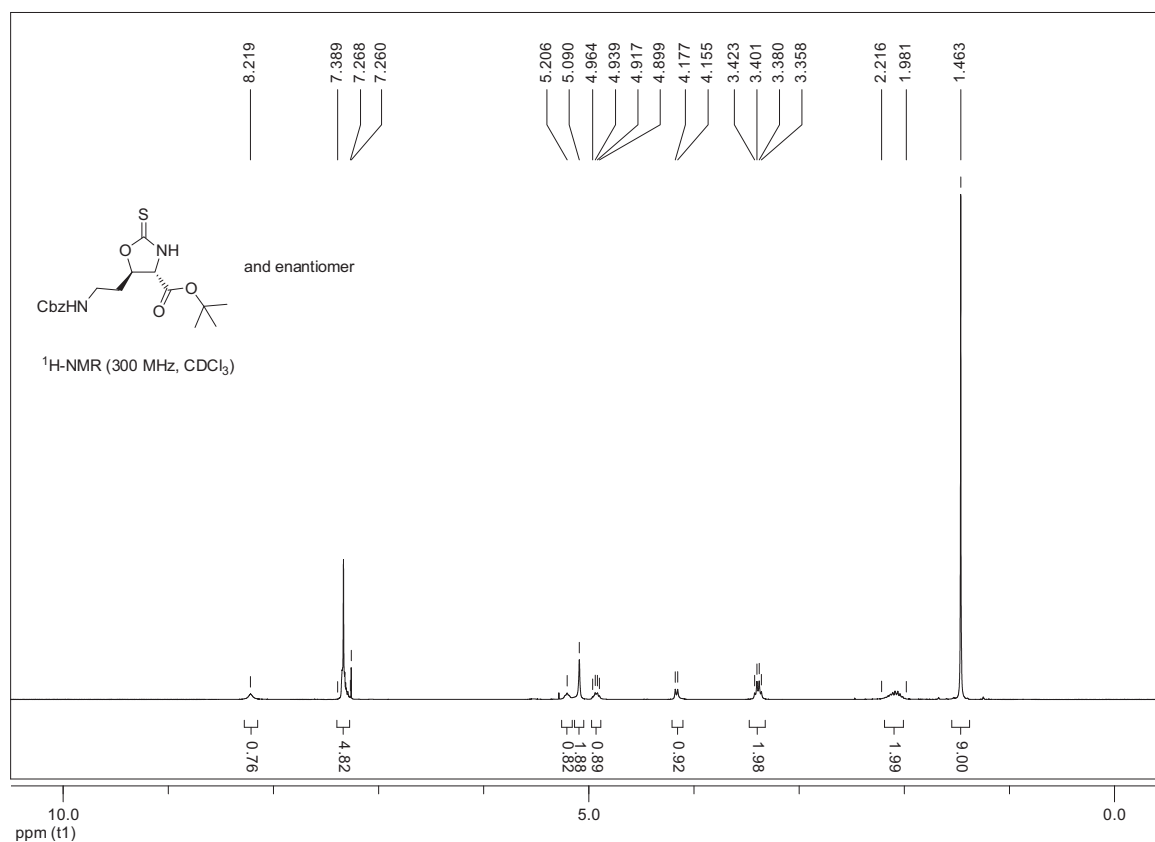


(2*S,3*R**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-5-methyl-3-(trityloxy)hexanoic Acid ((±)-*syn*-4)**

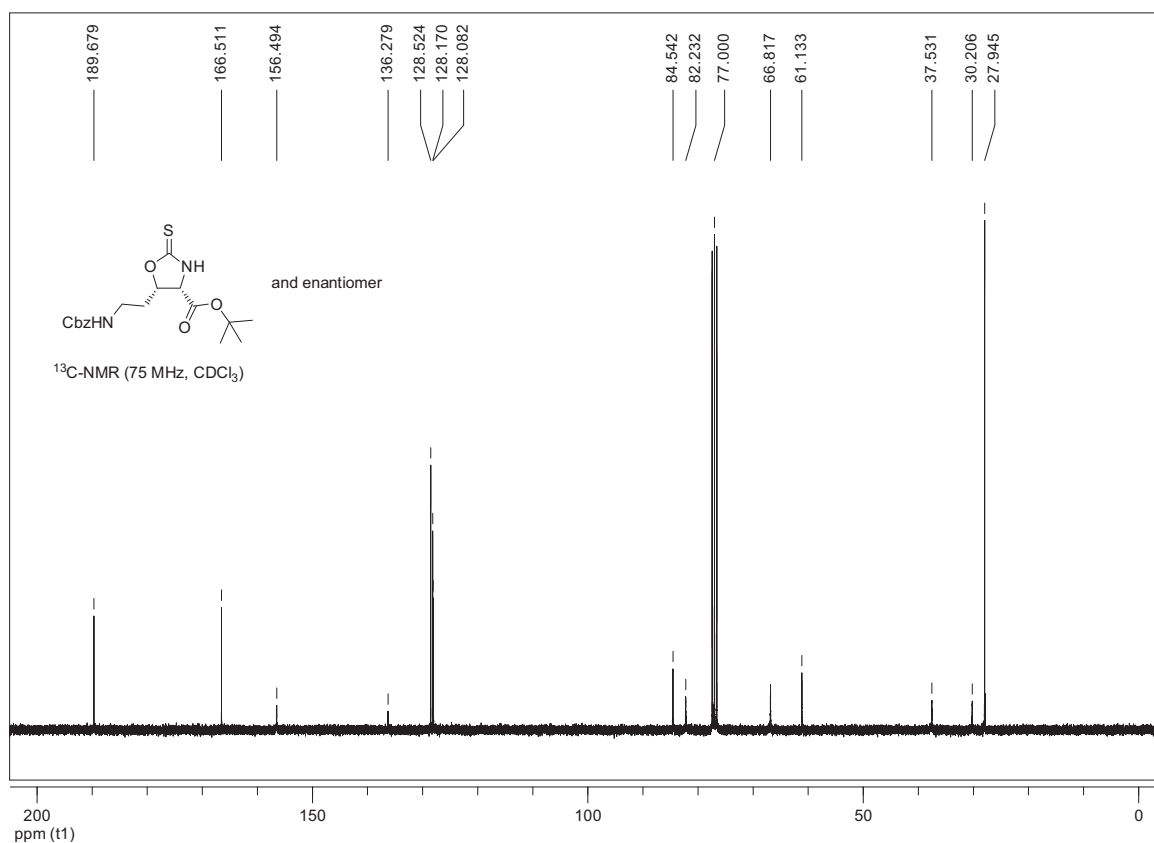
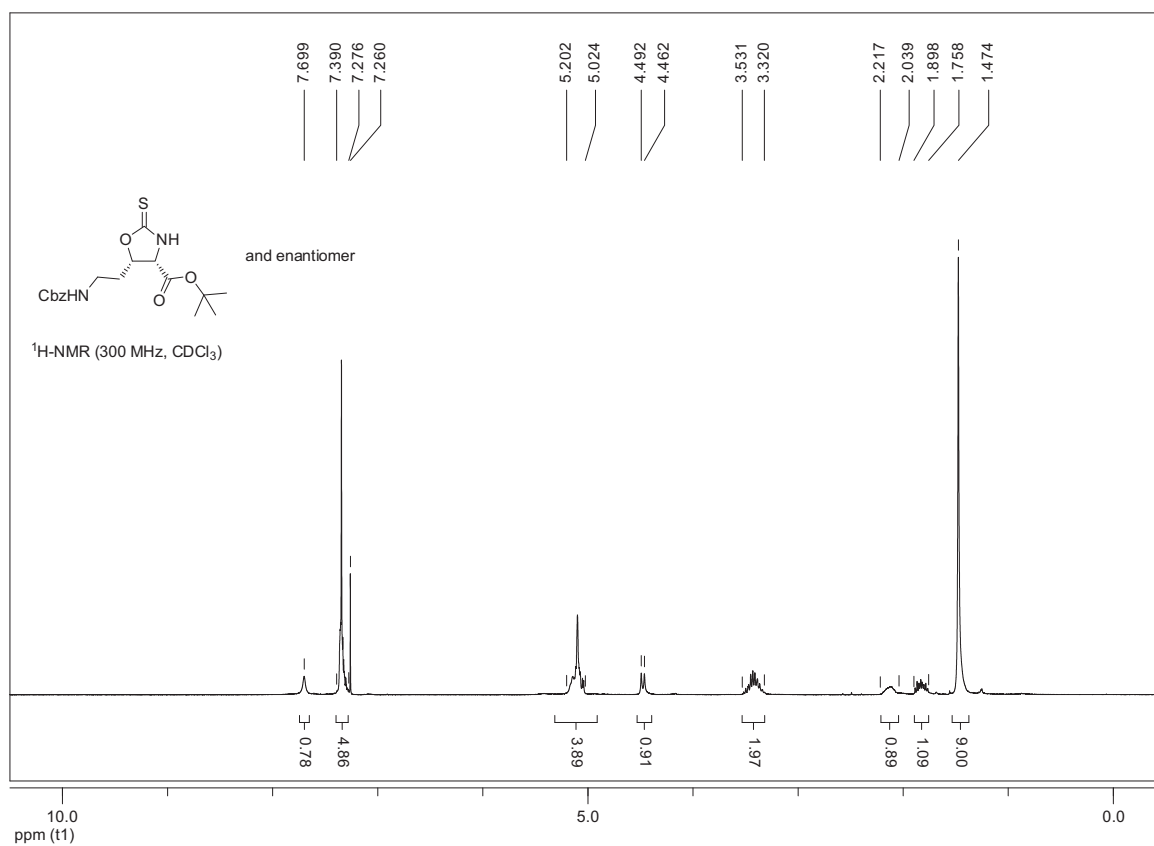
(2*S,3*S**)-2-(9*H*-Fluoren-9-ylmethoxycarbonylamino)-5-methyl-3-(trityloxy)hexanoic Acid ((±)-*anti*-4)**

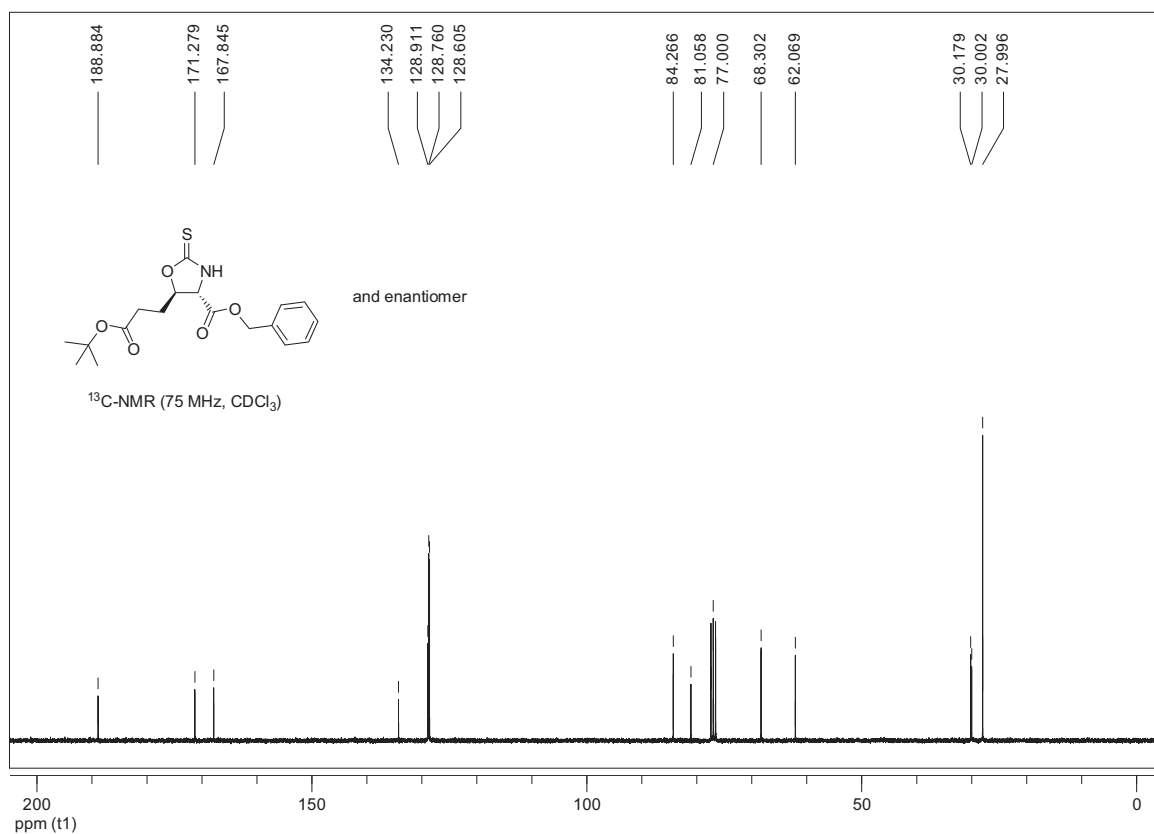
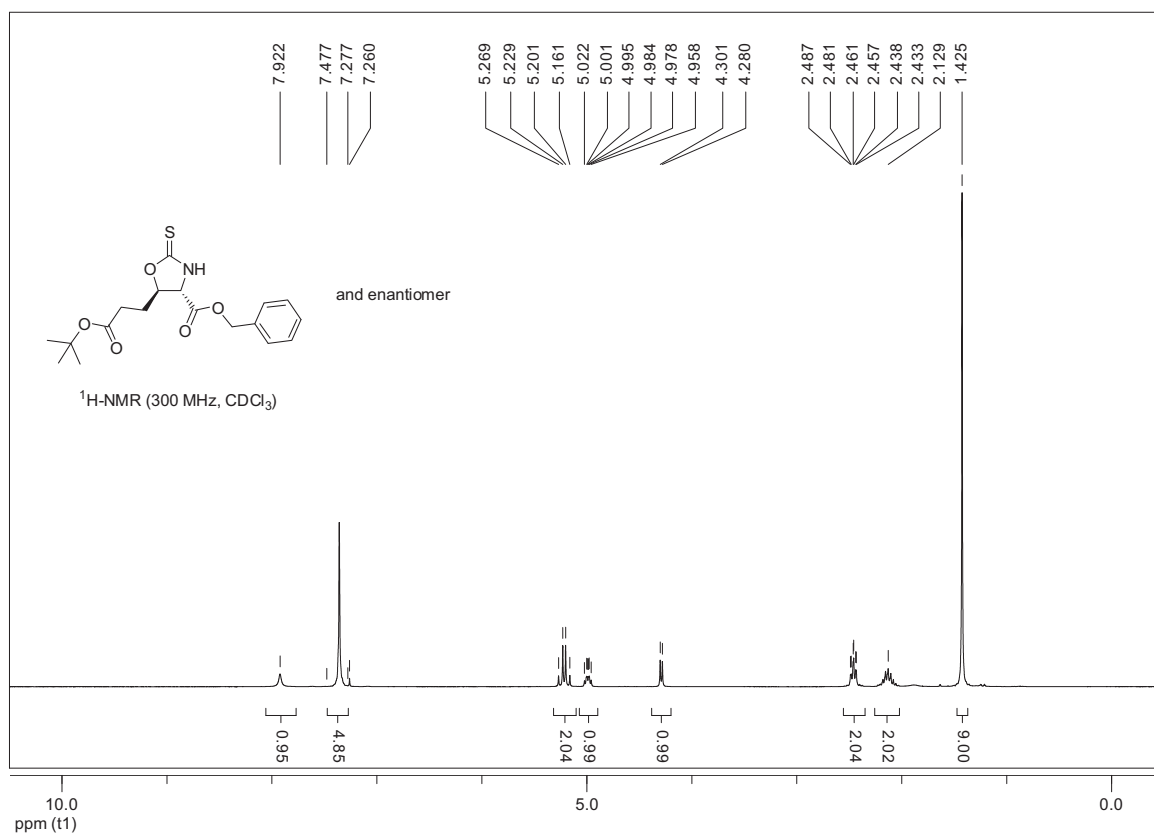


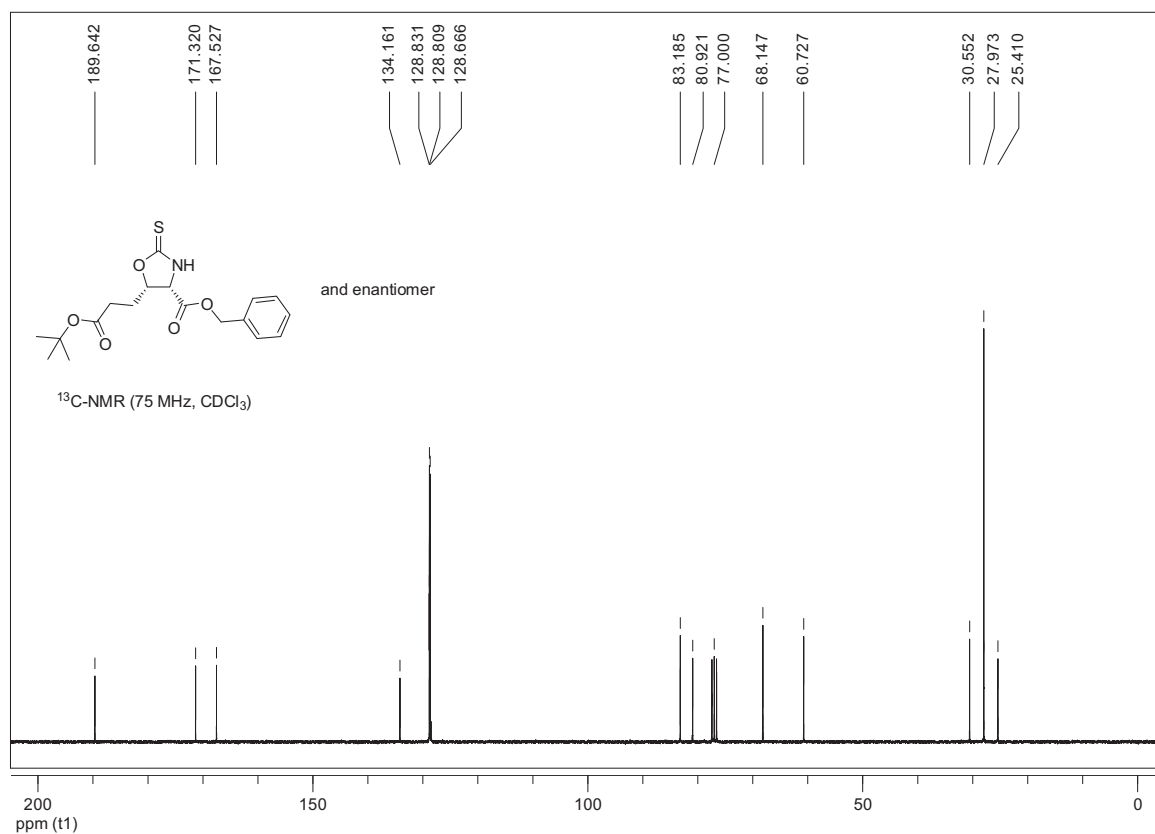
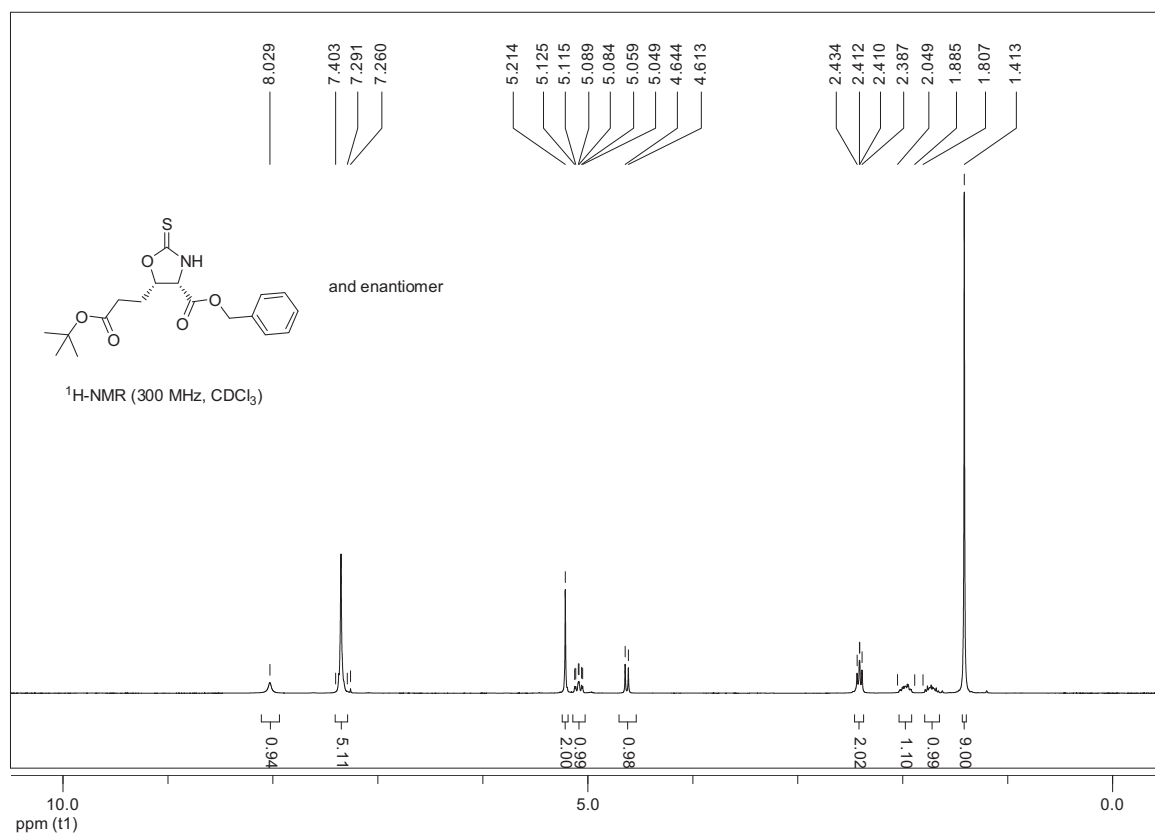
(4*S,5*R**)-5-(2-Benzoyloxycarbonylaminoethyl)-2-thioxooxazolidine-4-carboxylic Acid
tert-Butyl Ester ((±)-*trans*-26a)**



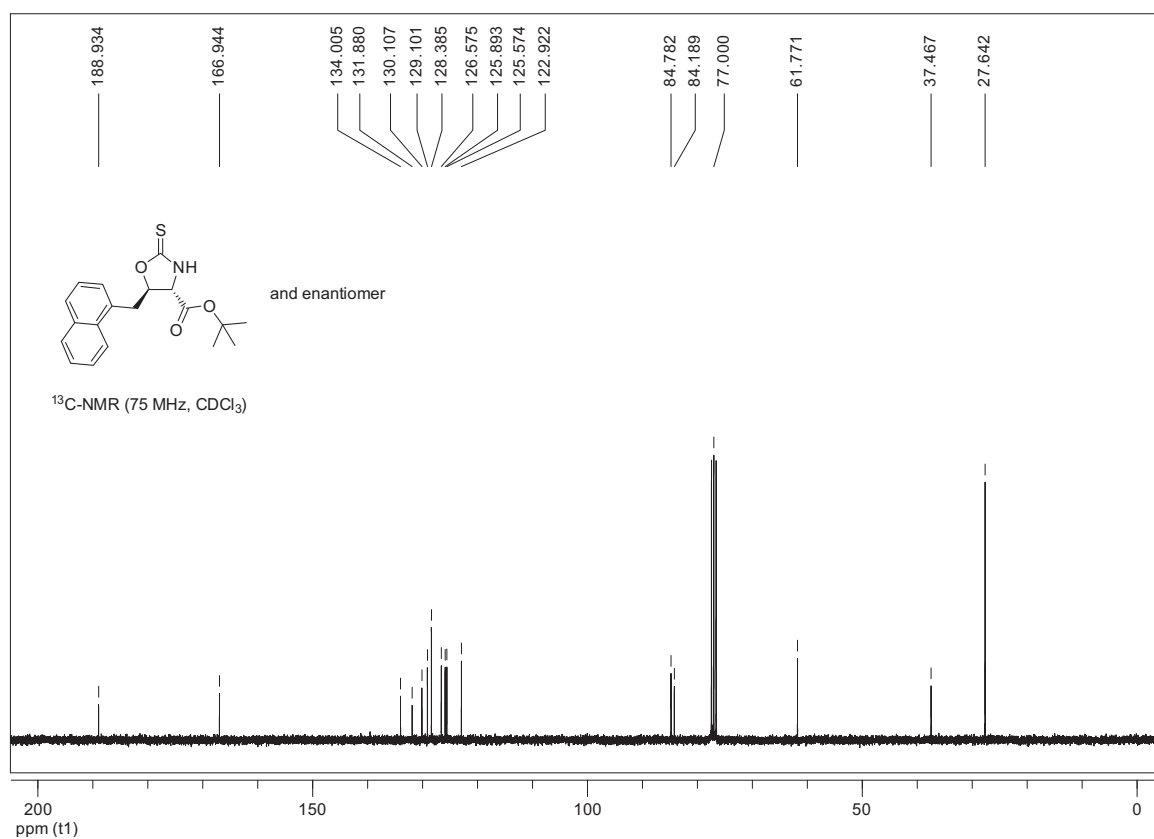
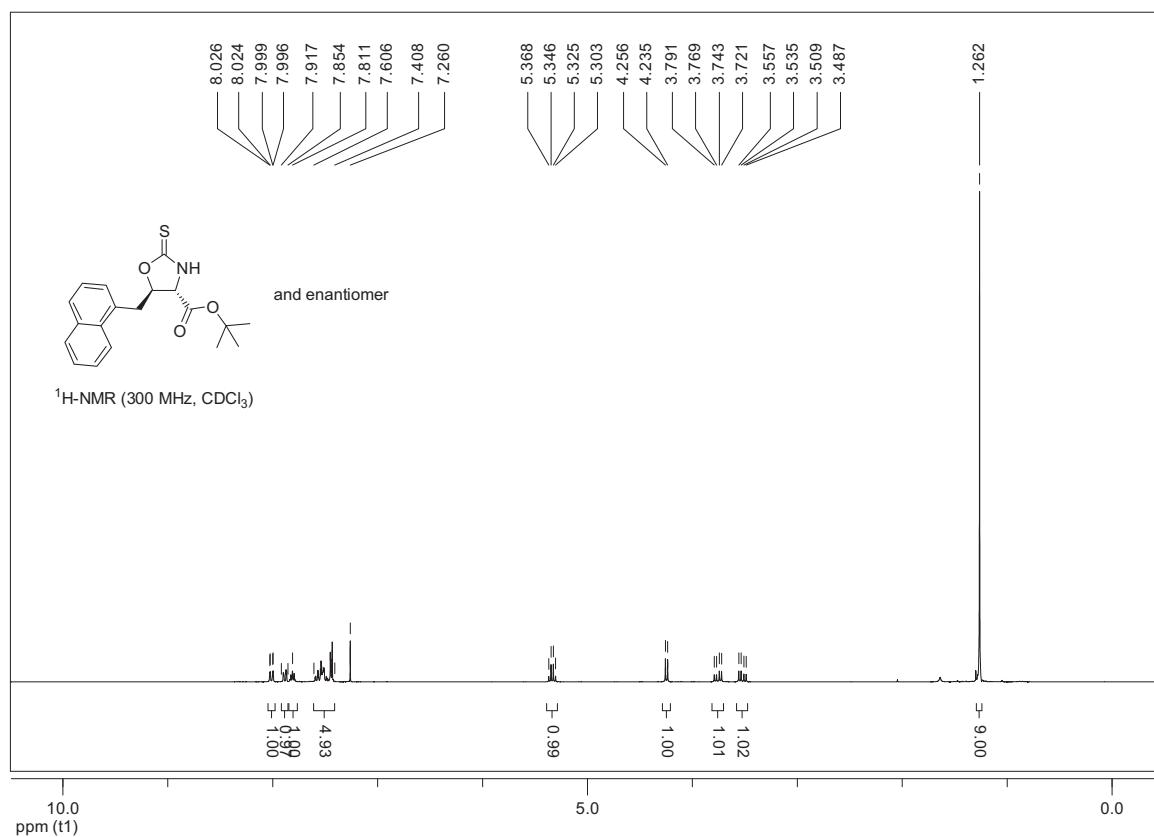
(4*S,5*S**)-5-(2-Benzoyloxycarbonylaminoethyl)-2-thioxooxazolidine-4-carboxylic Acid
tert-Butyl Ester (\pm)-*cis*-26a)**

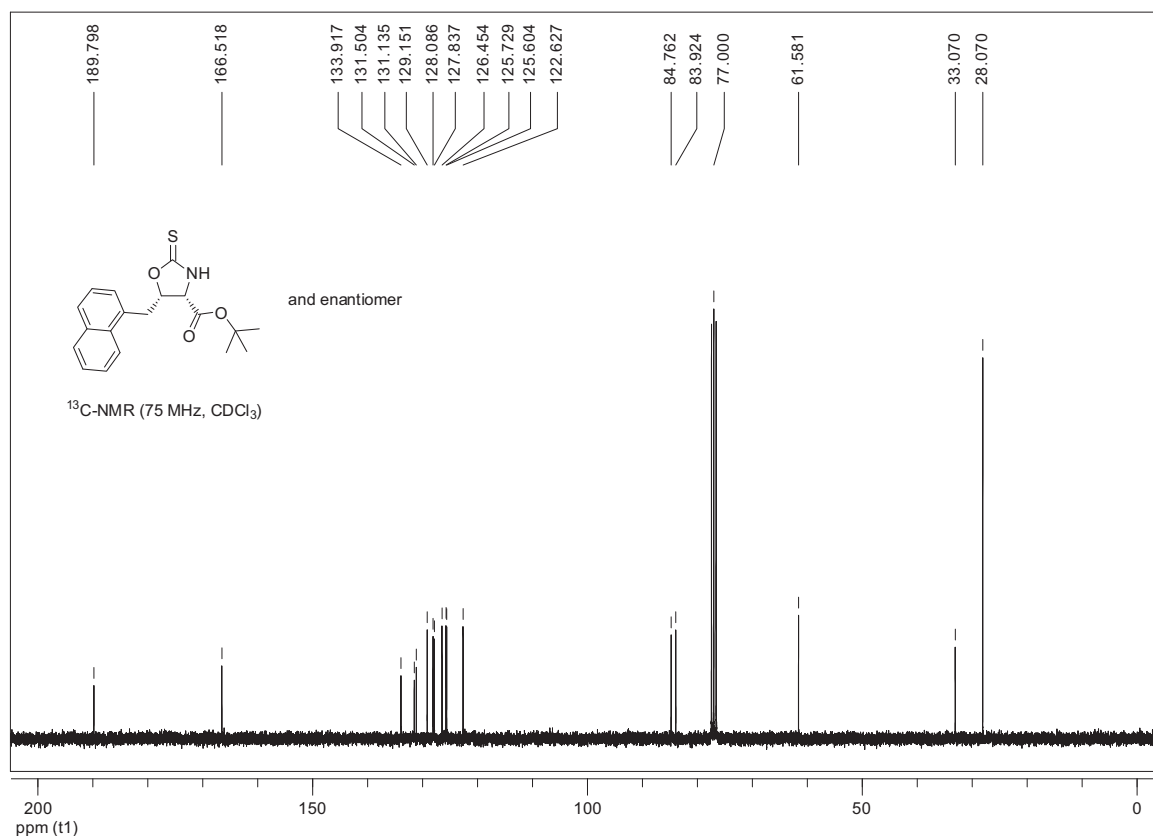
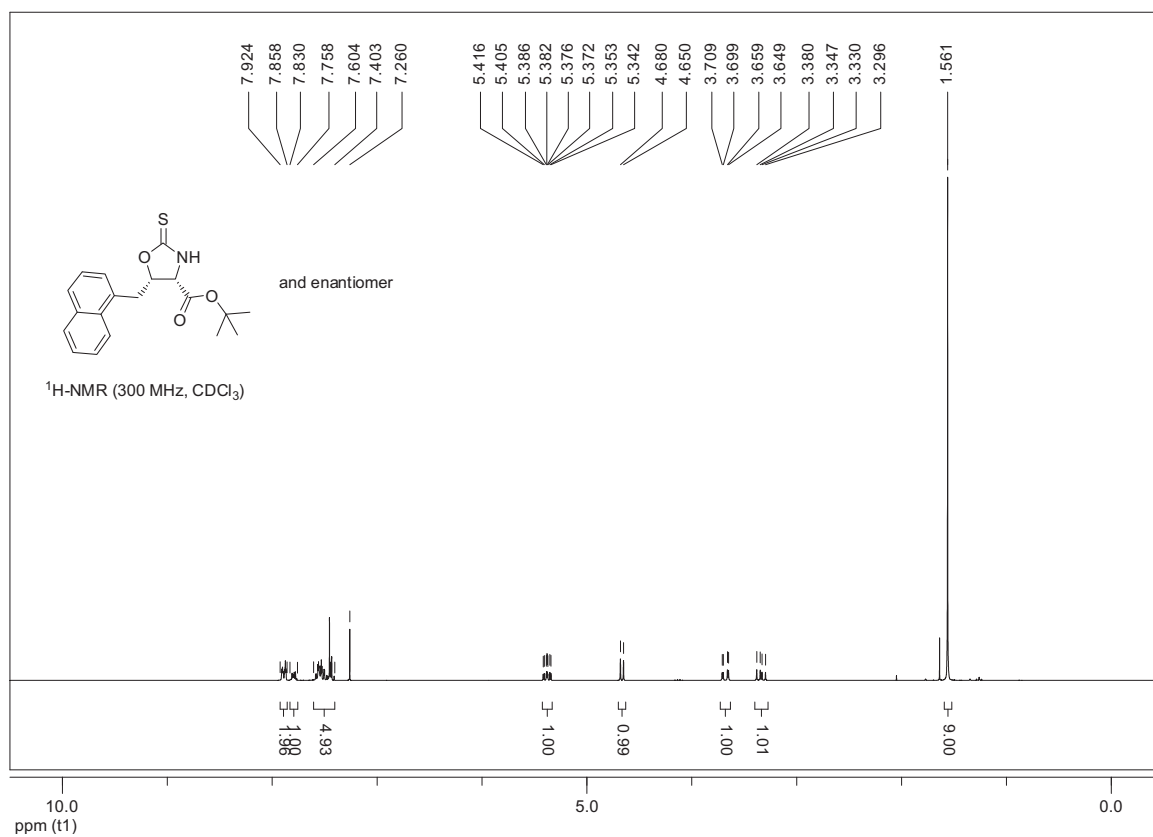


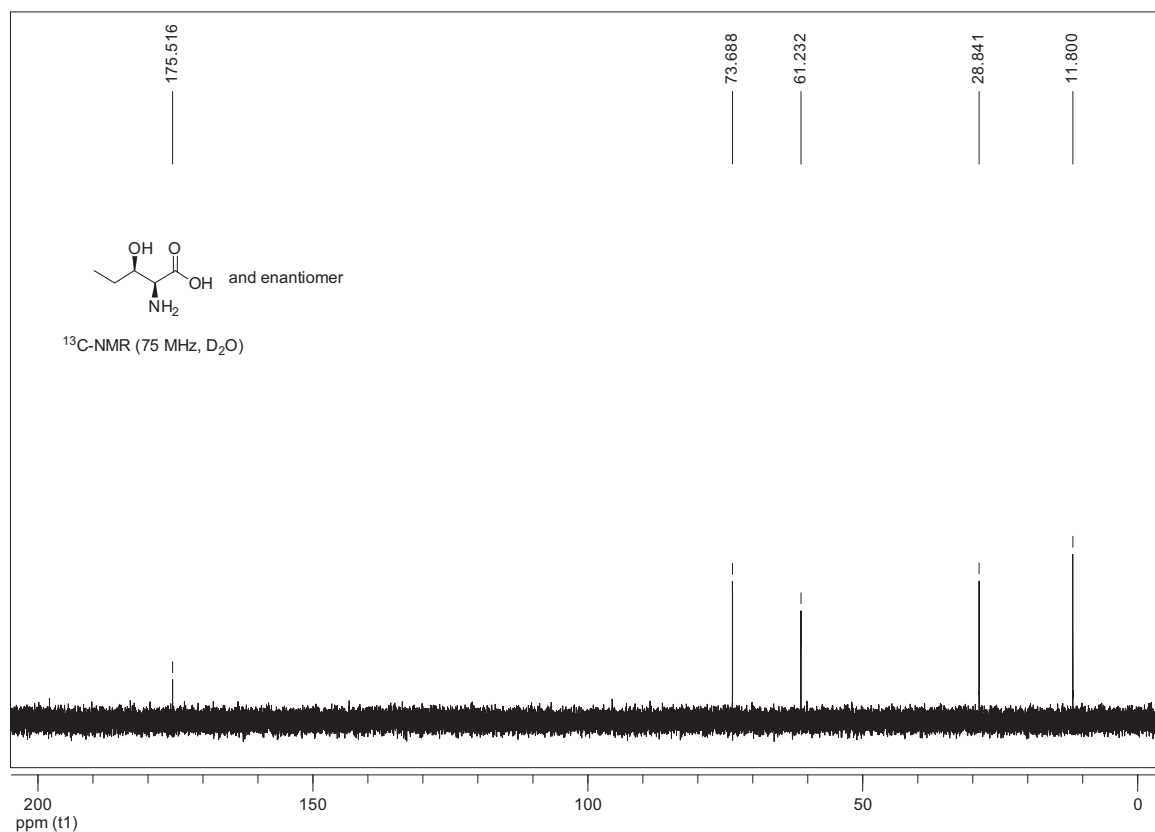
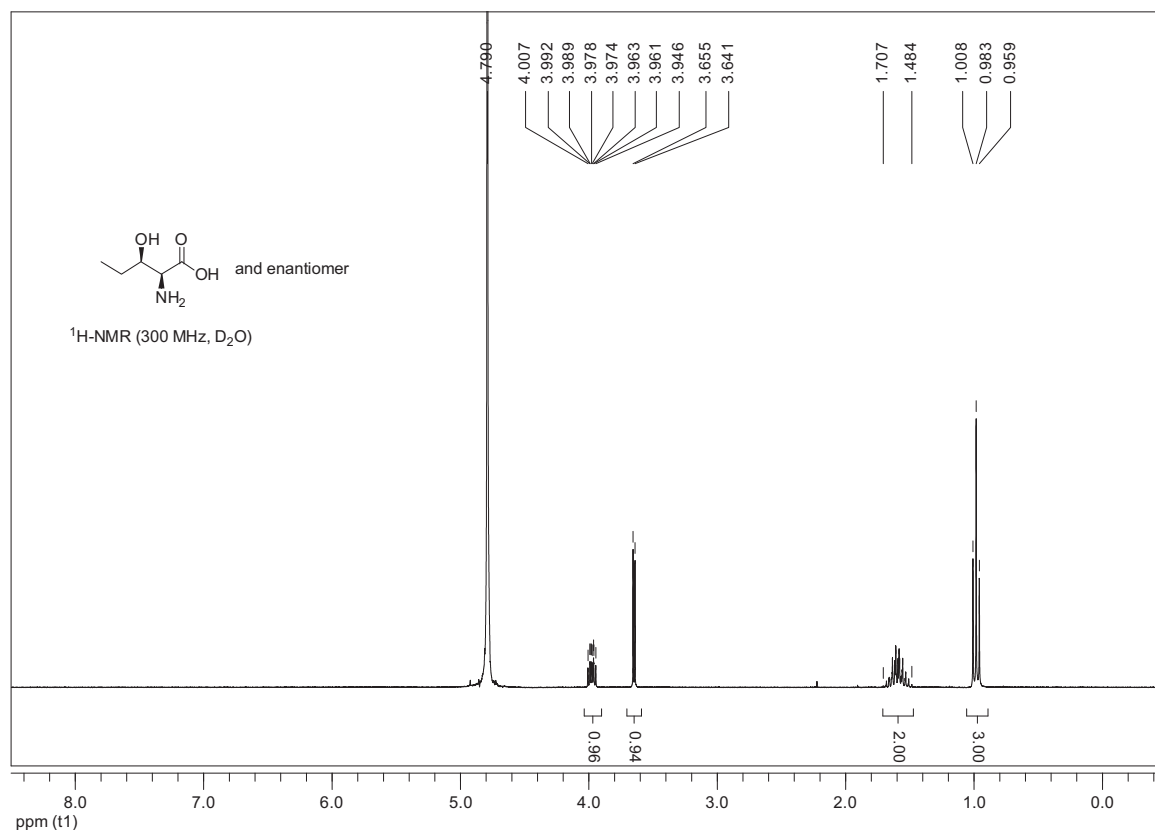
(4*S,5*R**)-5-(3-(*tert*-Butoxy)-3-oxopropyl)-2-thioxooxazolidine-4-carboxylic Acid Benzyl Ester (±)-*trans*-26b)**

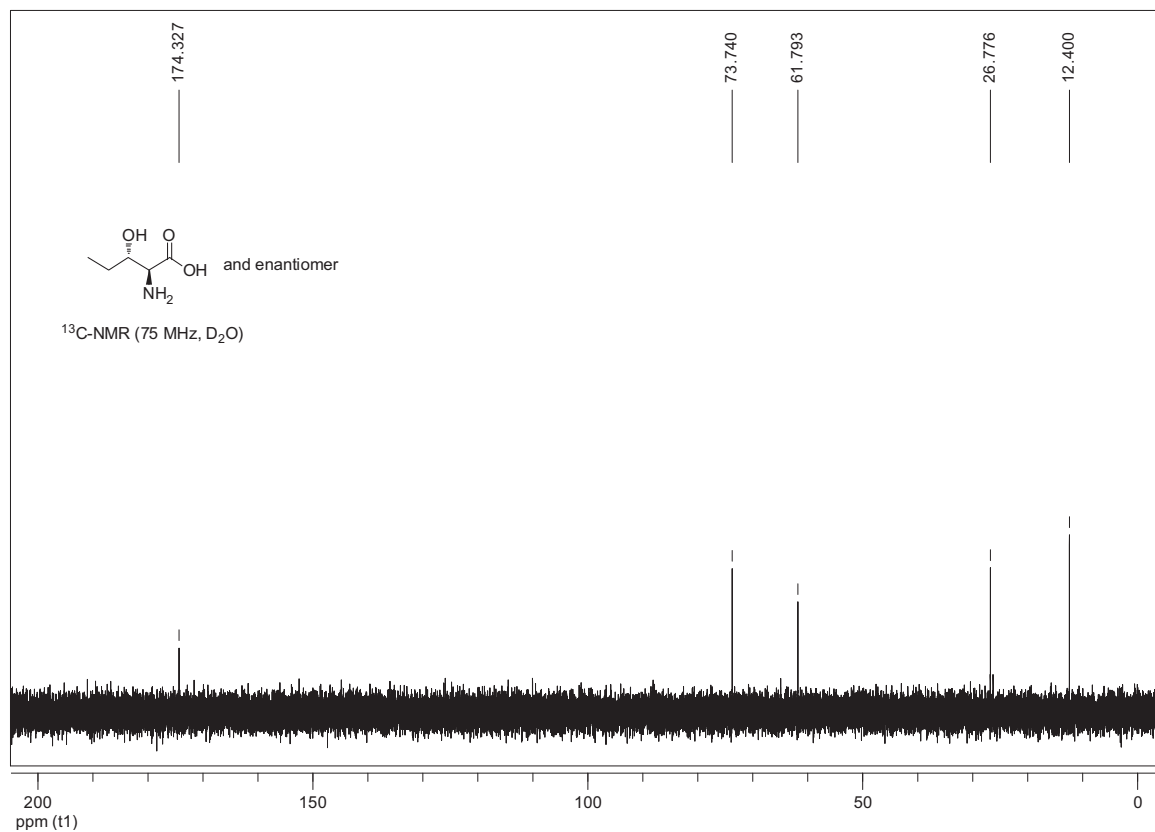
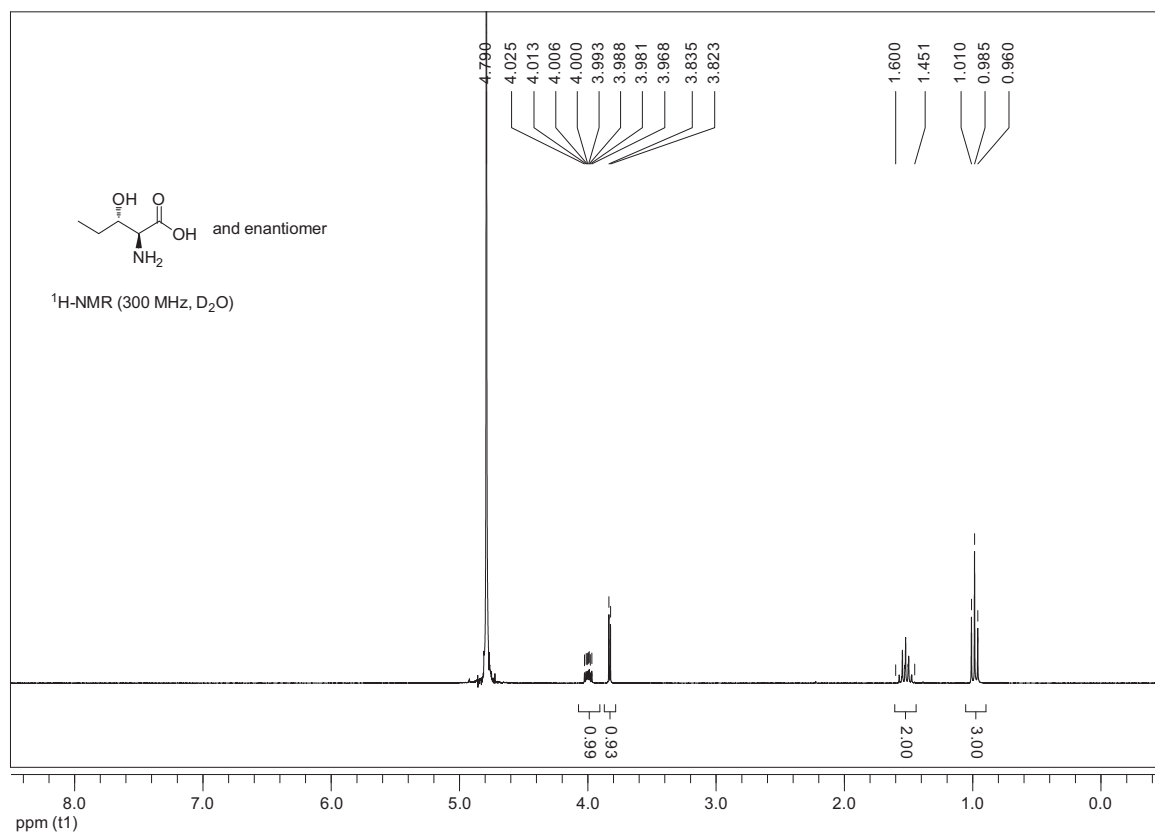
(4*S,5*S**)-5-(3-(*tert*-Butoxy)-3-oxopropyl)-2-thioxooxazolidine-4-carboxylic Acid Benzyl Ester (±)-*cis*-26b)**

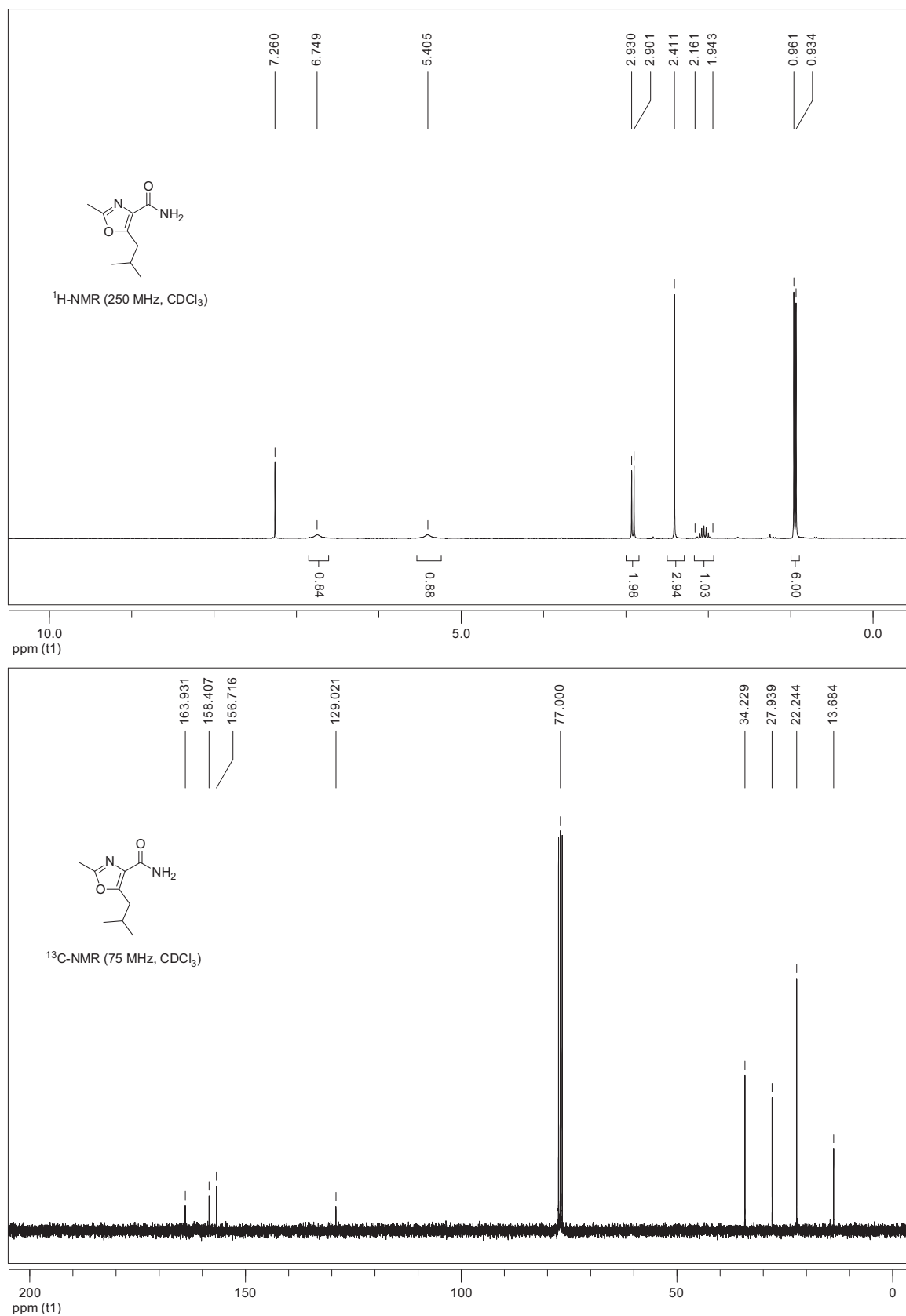
(4*S,5*R**)-5-(1-Naphthylmethyl)-2-thioxooxazolidine-4-carboxylic Acid *tert*-Butyl Ester ((±)-*trans*-26c)**

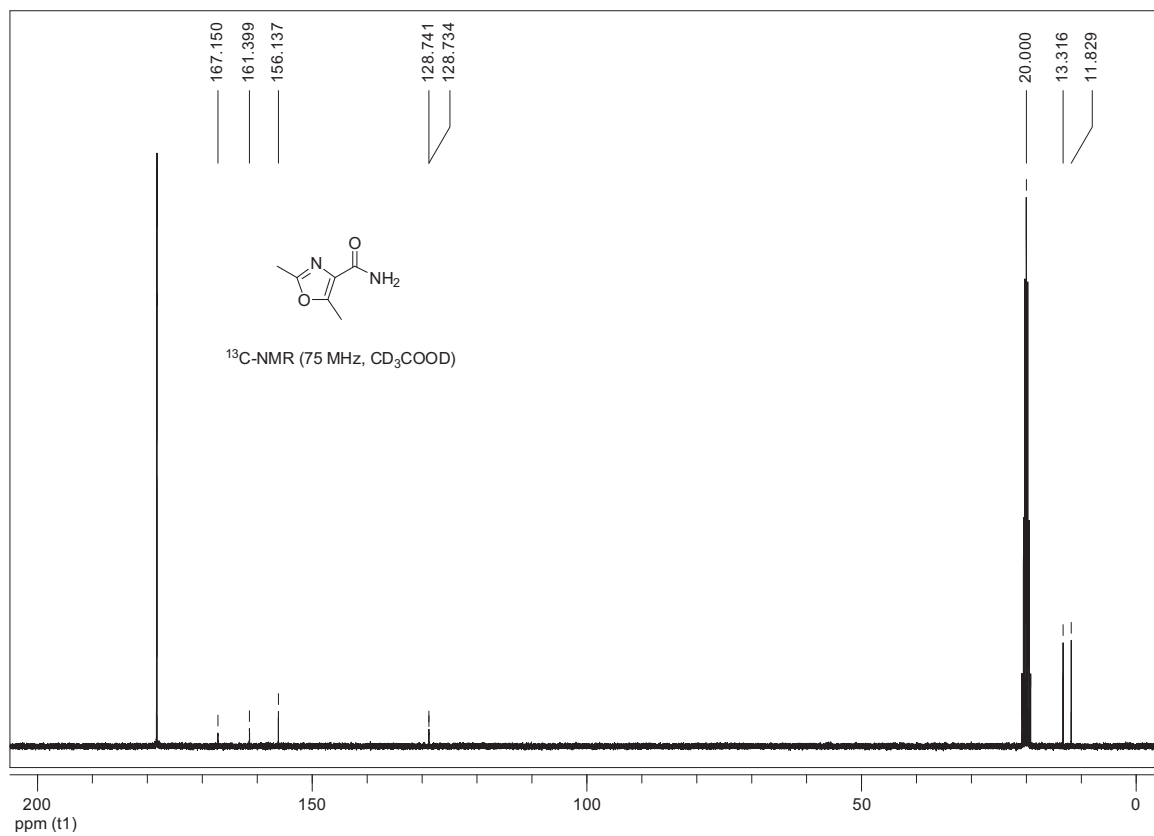
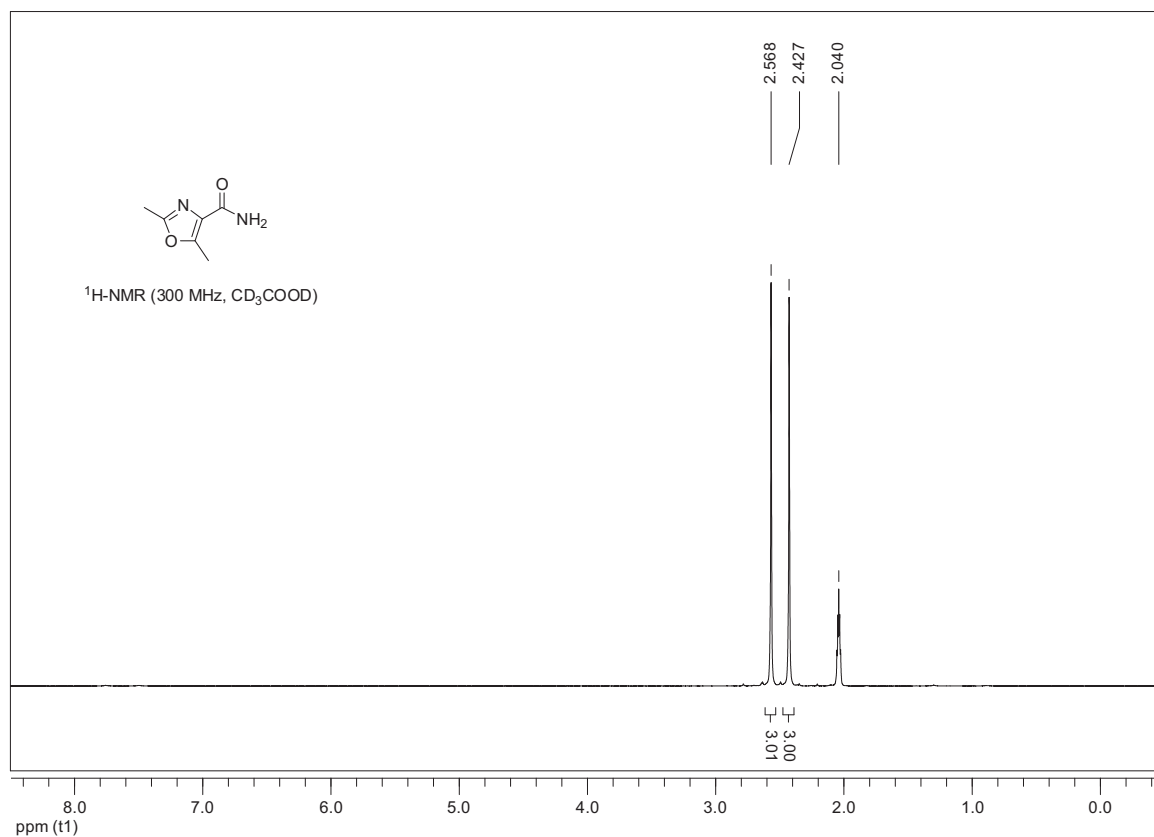


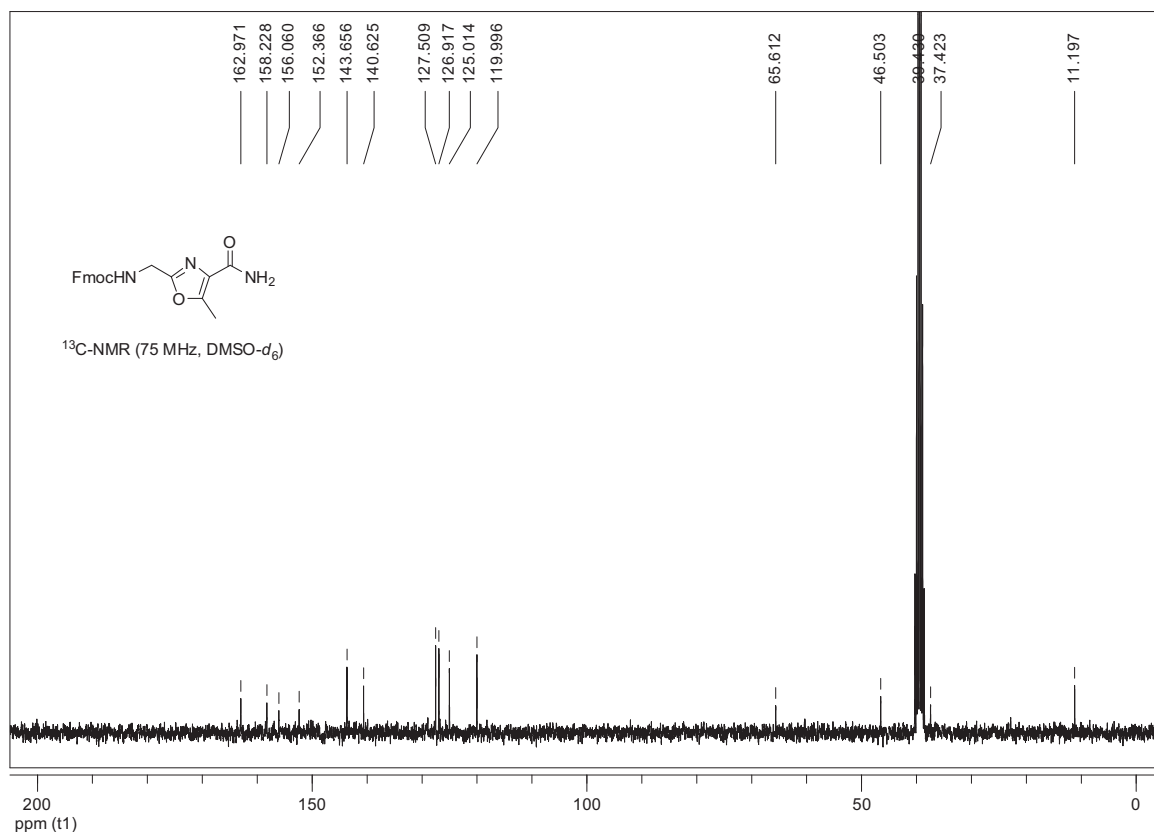
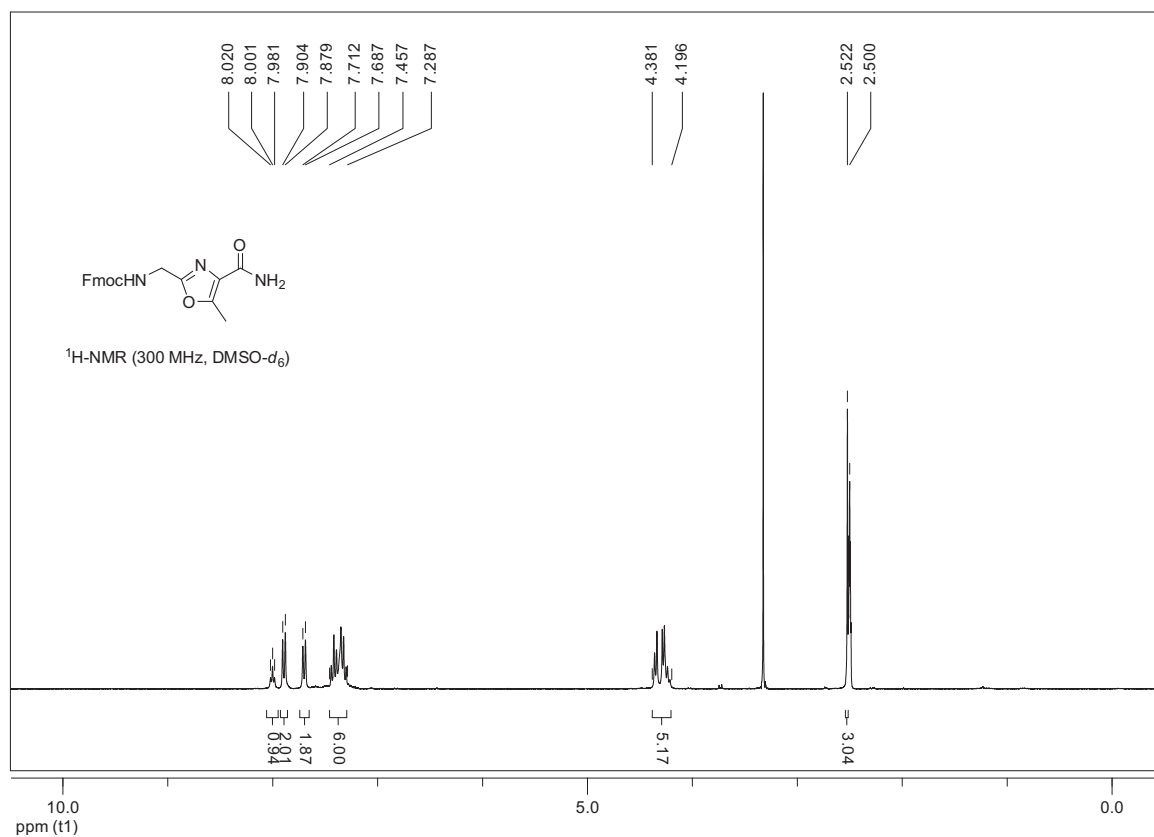
(4*S,5*S**)-5-(1-naphthylmethyl)-2-thioxooxazolidine-4-carboxylic Acid *tert*-Butyl Ester ((±)-*cis*-26c)**

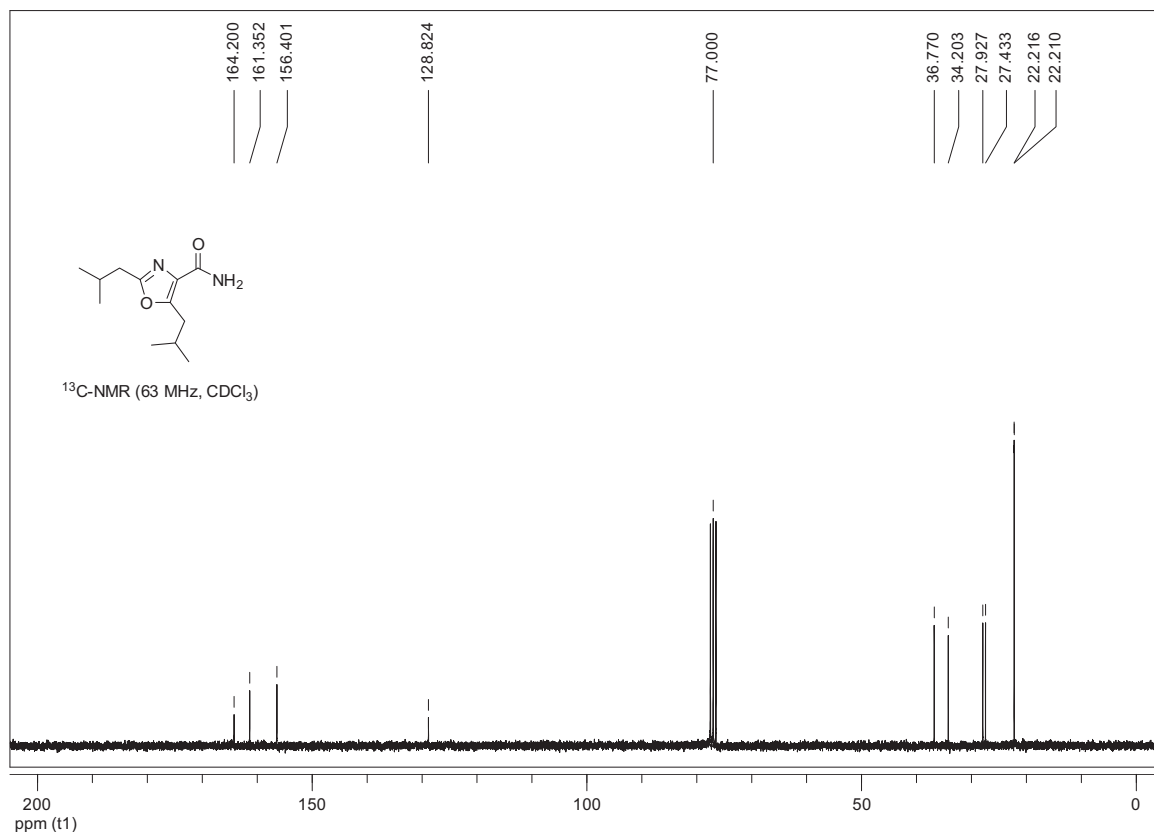
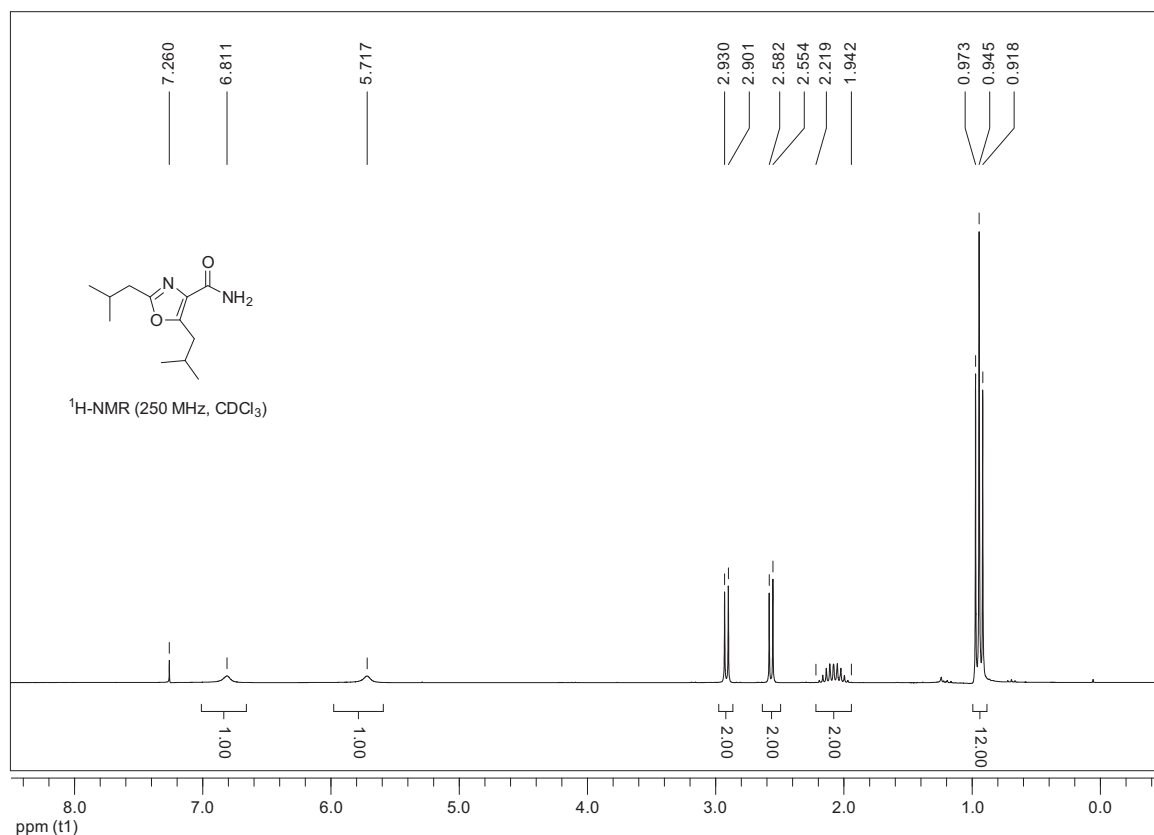
(2*S,3*R**)-2-Amino-3-hydroxypentanoic Acid ((±)-*syn*-27)**

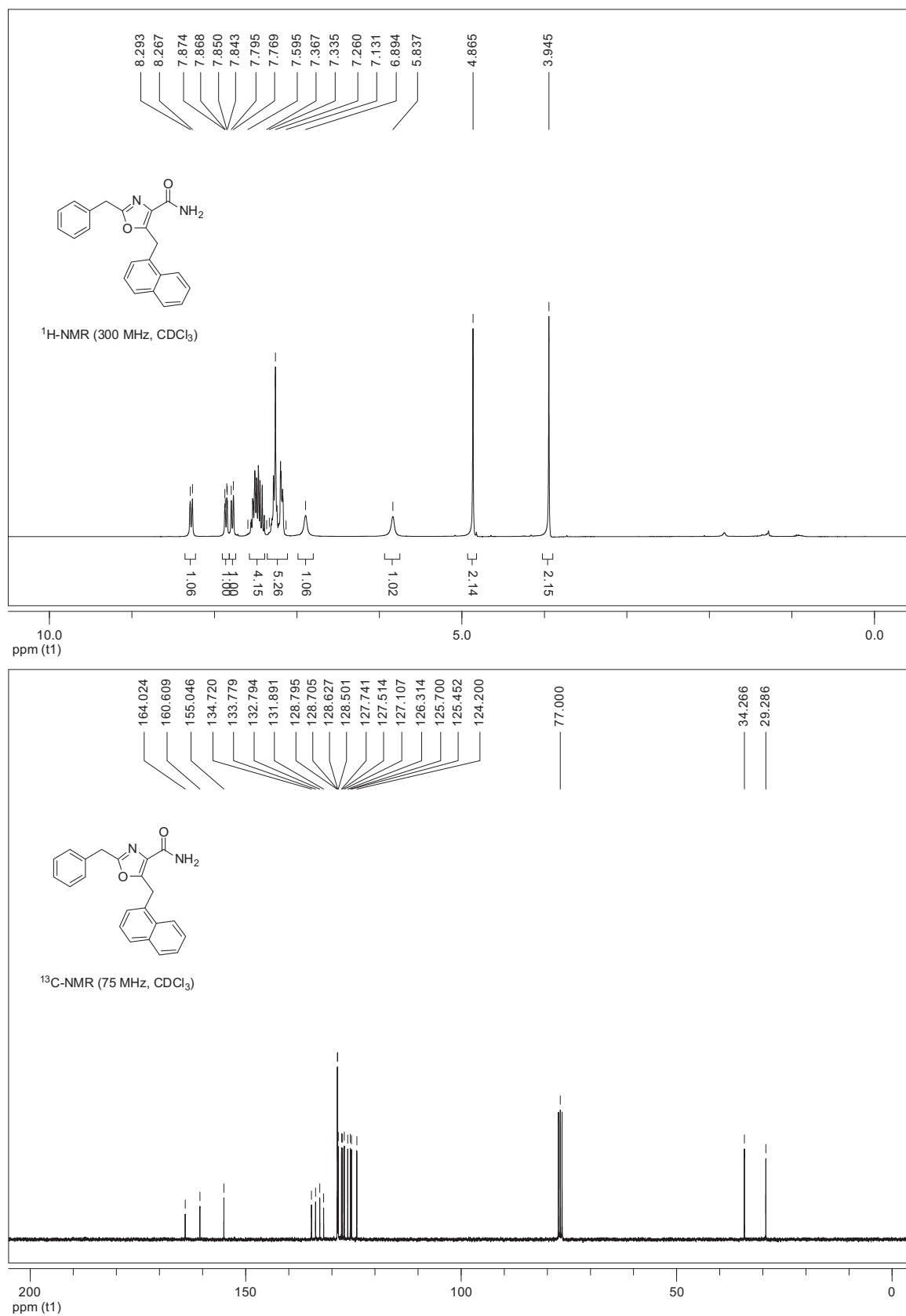
(2S*,3S*)-2-Amino-3-hydroxypentanoic Acid ((±)-anti-27)


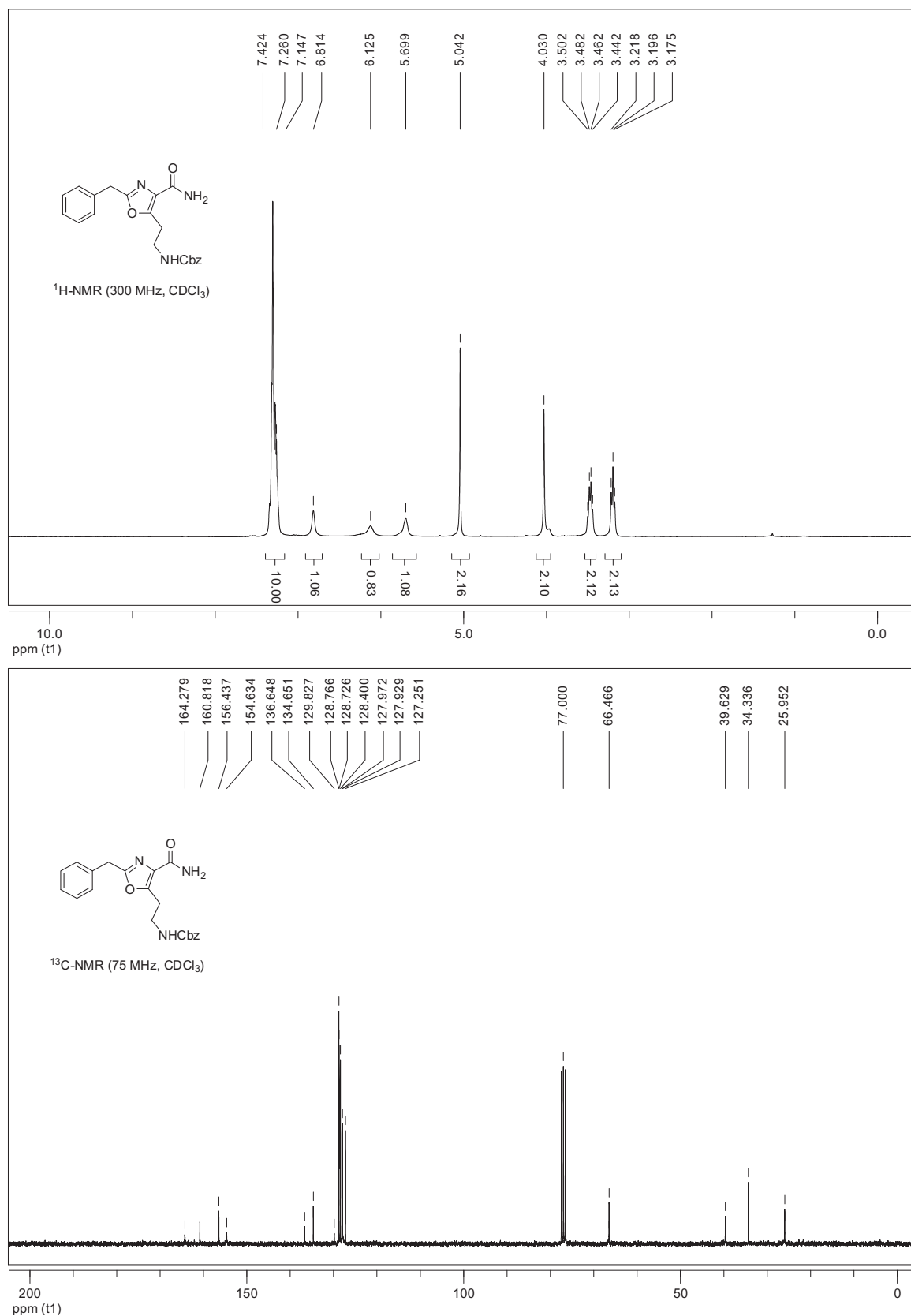
5-Isobutyl-2-methyl-1,3-oxazole-4-carboxamide (14c)

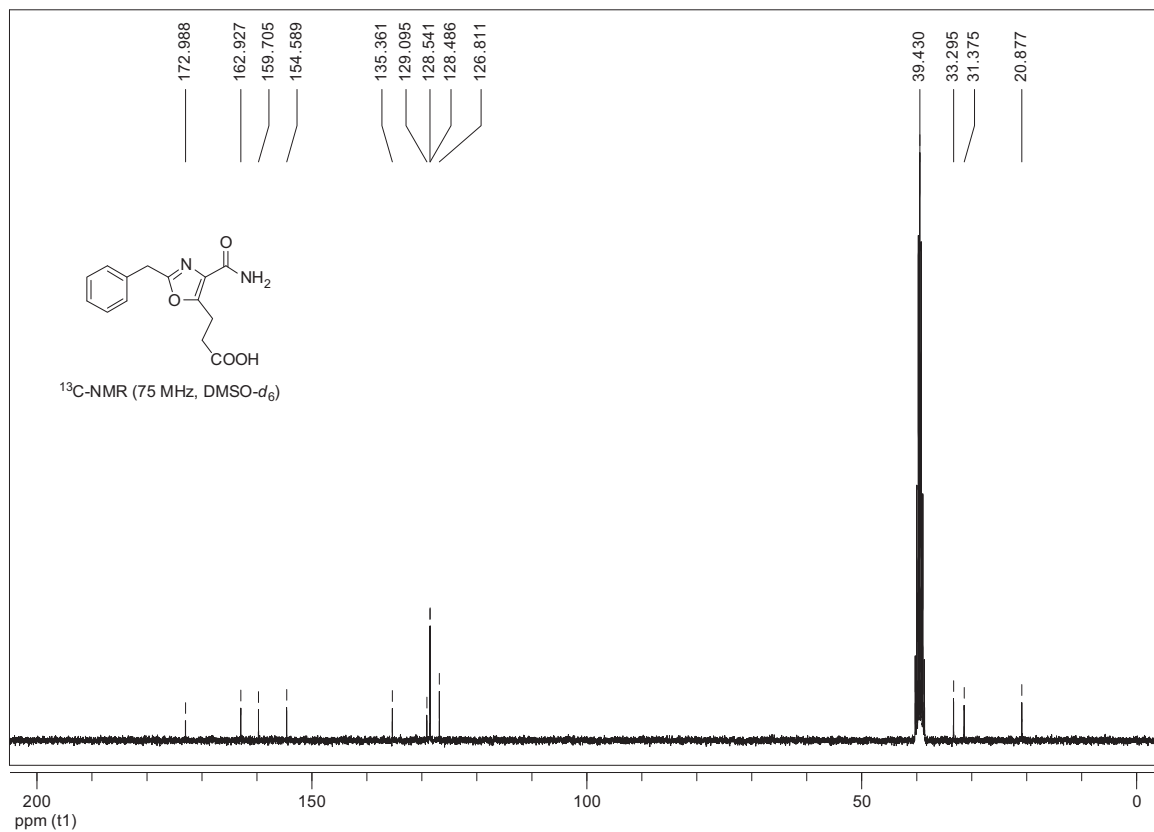
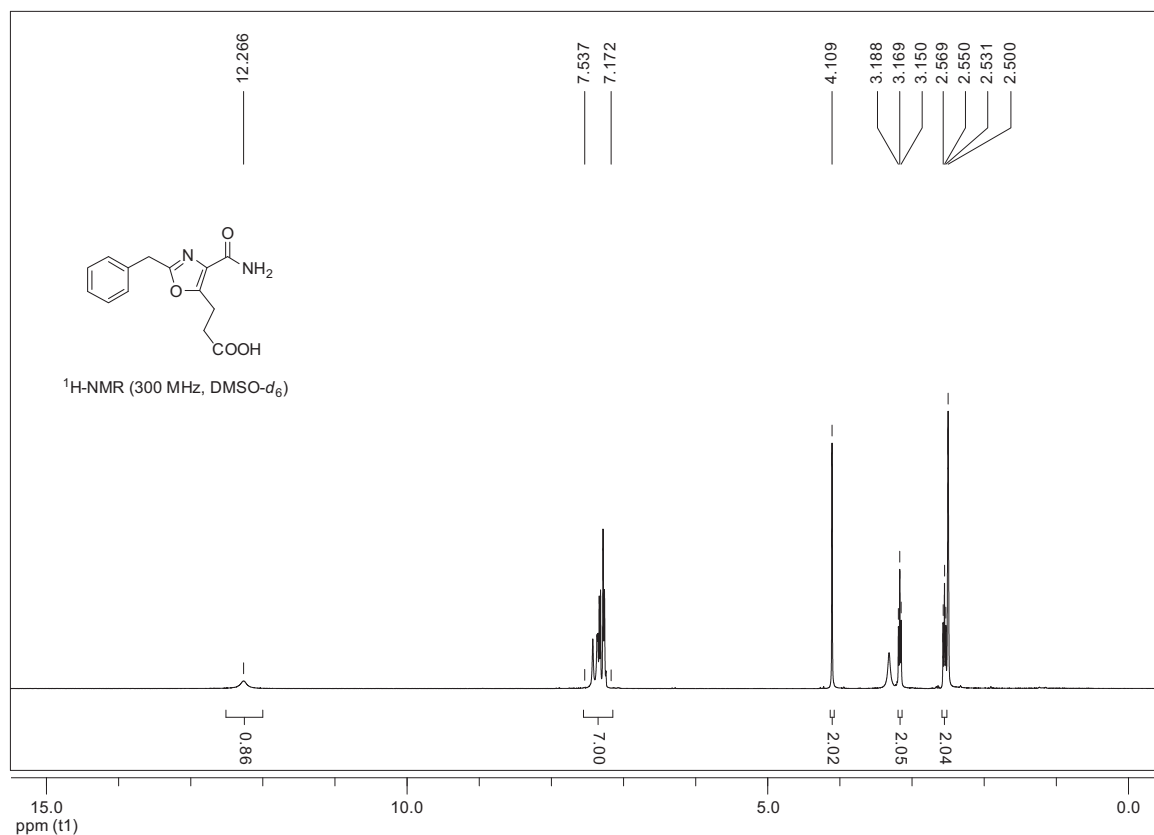
2,5-Dimethyl-1,3-oxazole-4-carboxamide (14a)


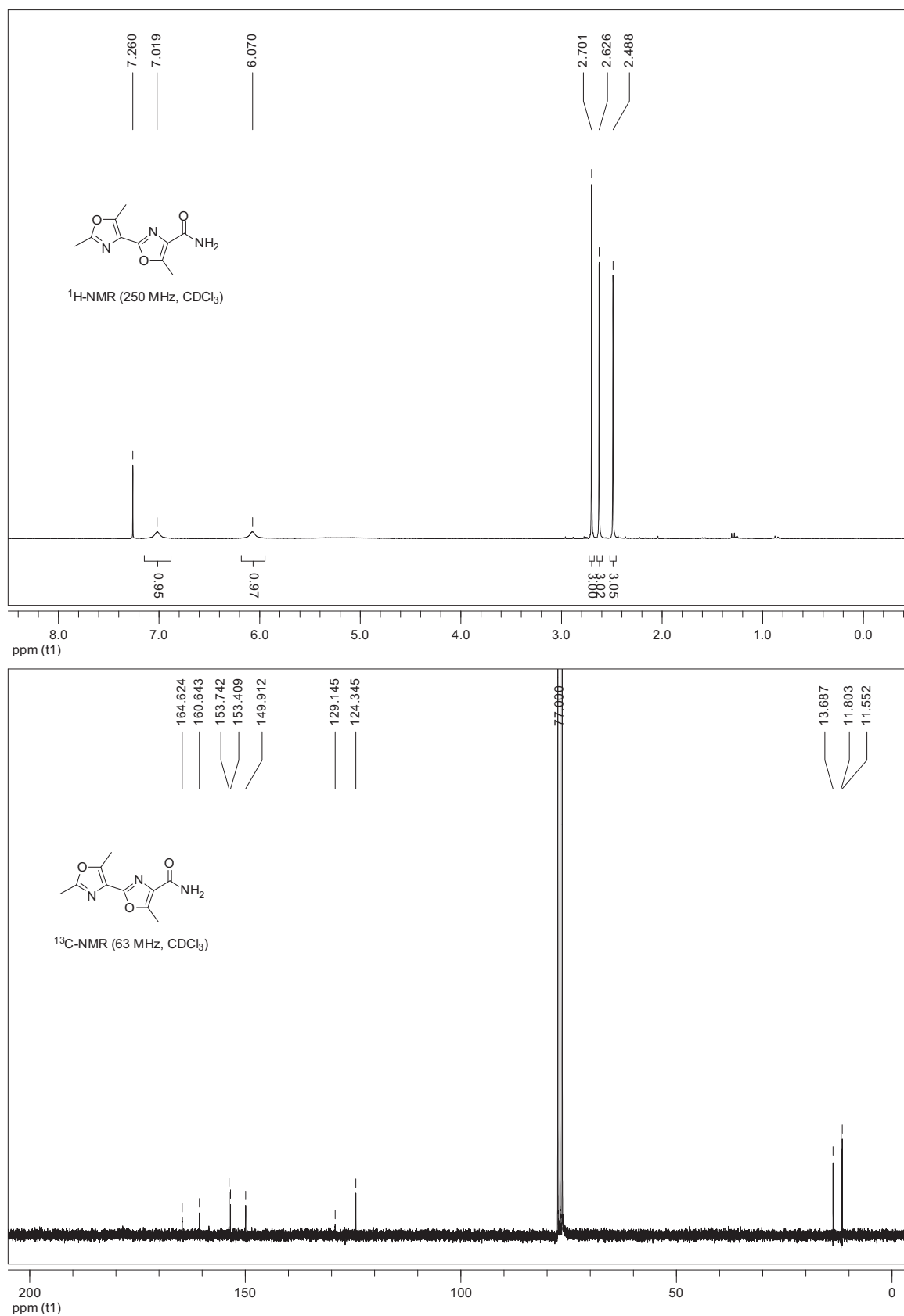
2-(9*H*-Fluoren-9-ylmethoxycarbonylaminomethyl)-5-methyl-1,3-oxazole-4-carboxamide (14b)

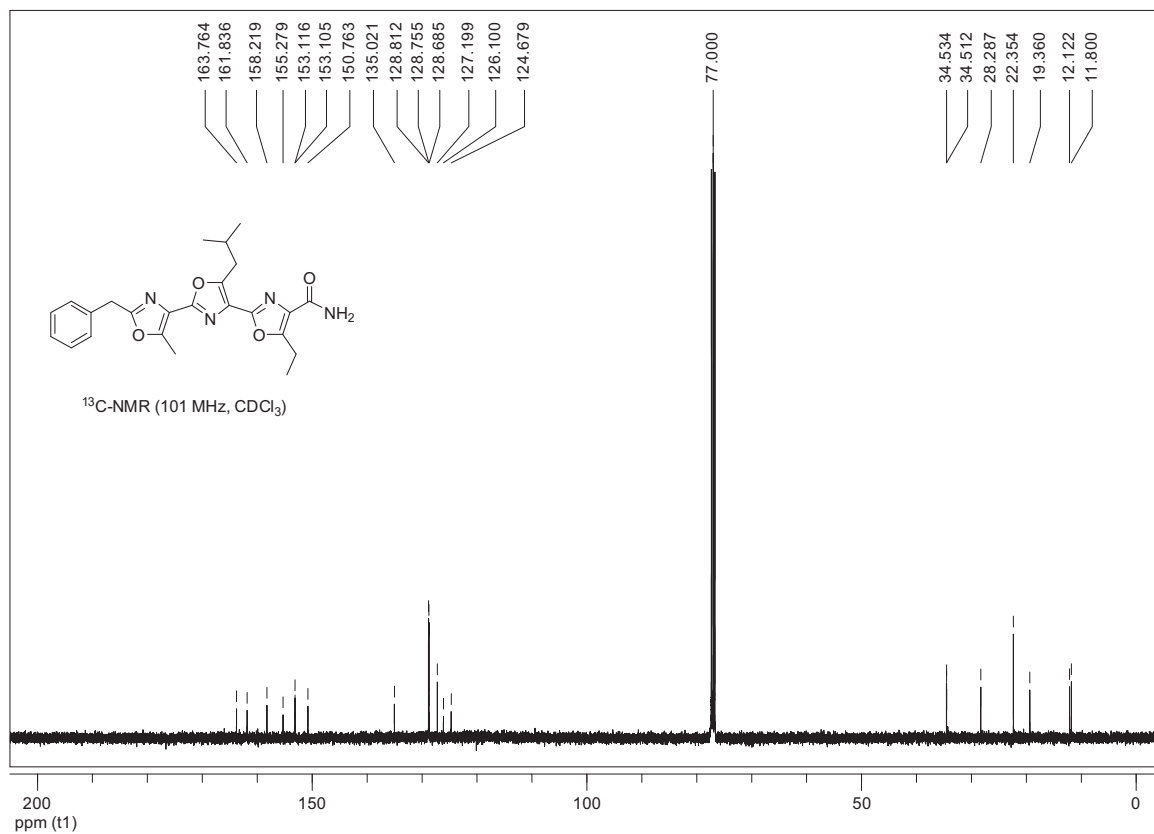
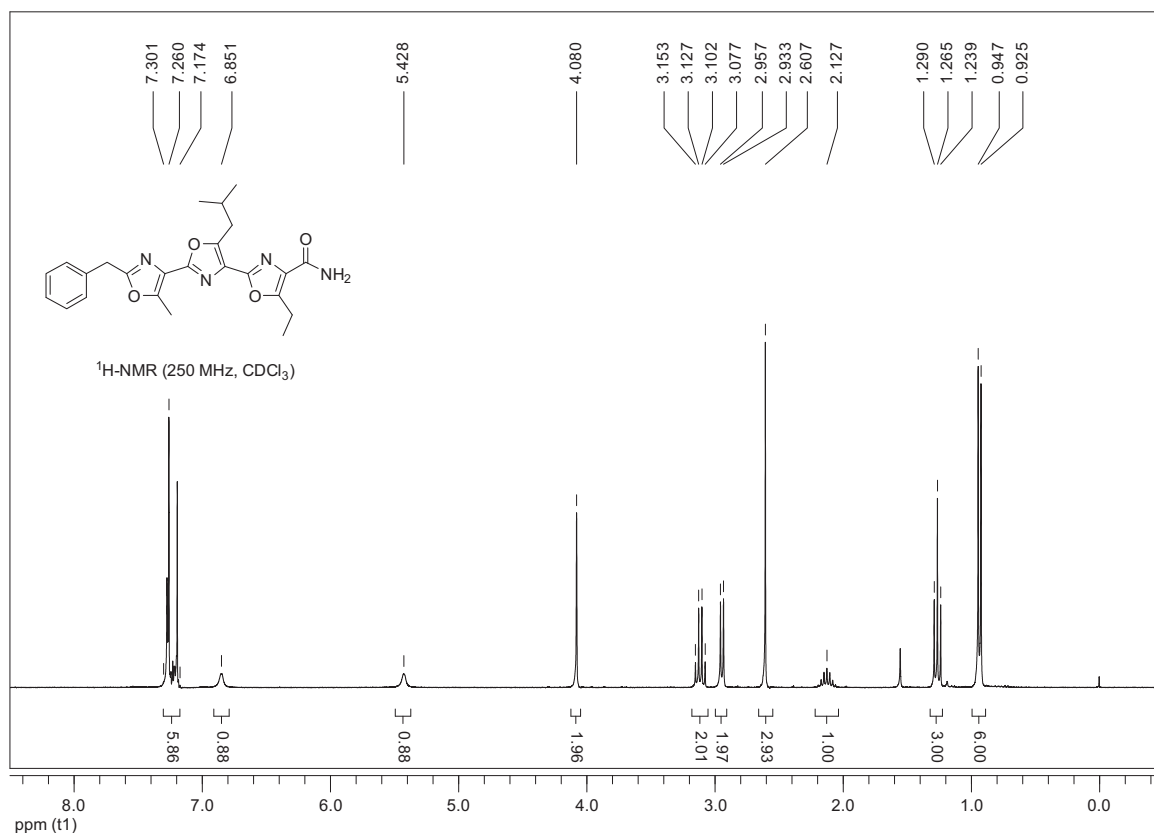
2,5-Isobutyl-1,3-oxazole-4-carboxamide (14d)

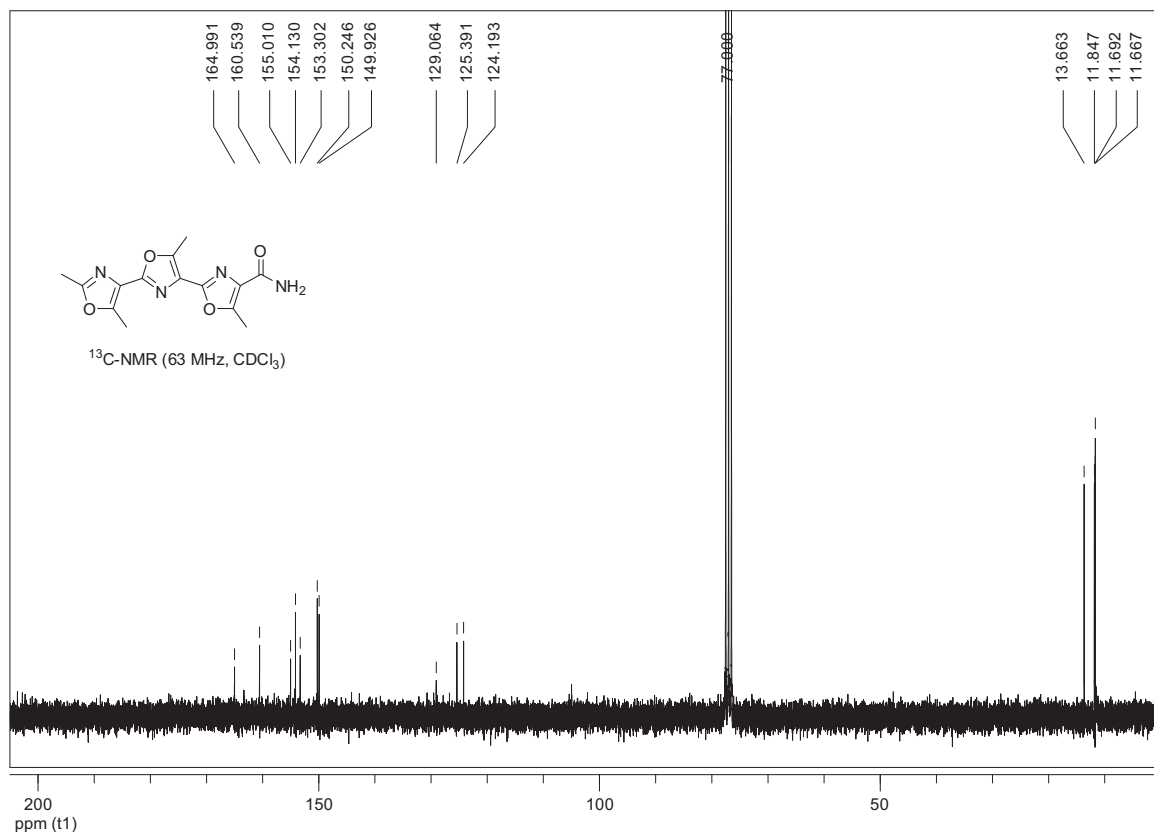
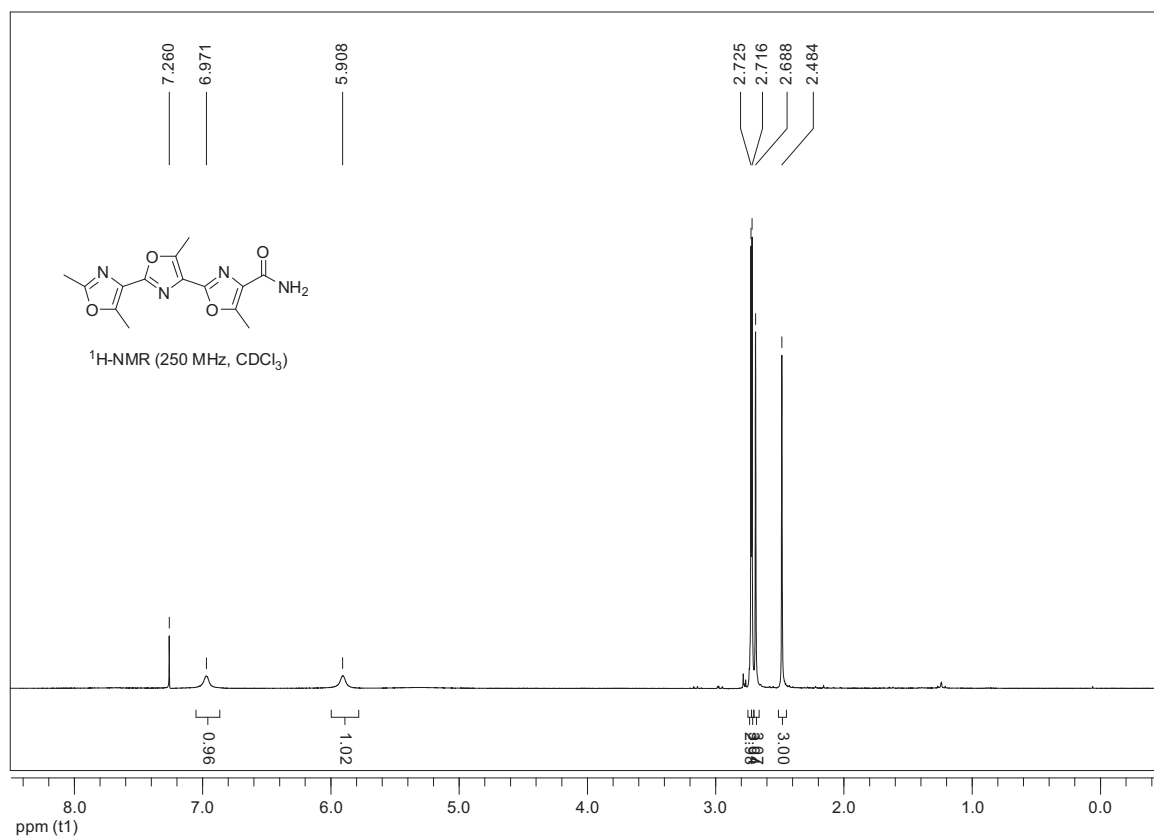
2-Benzyl-5-(1-naphthylmethyl)-1,3-oxazole-4-carboxamide (14e)


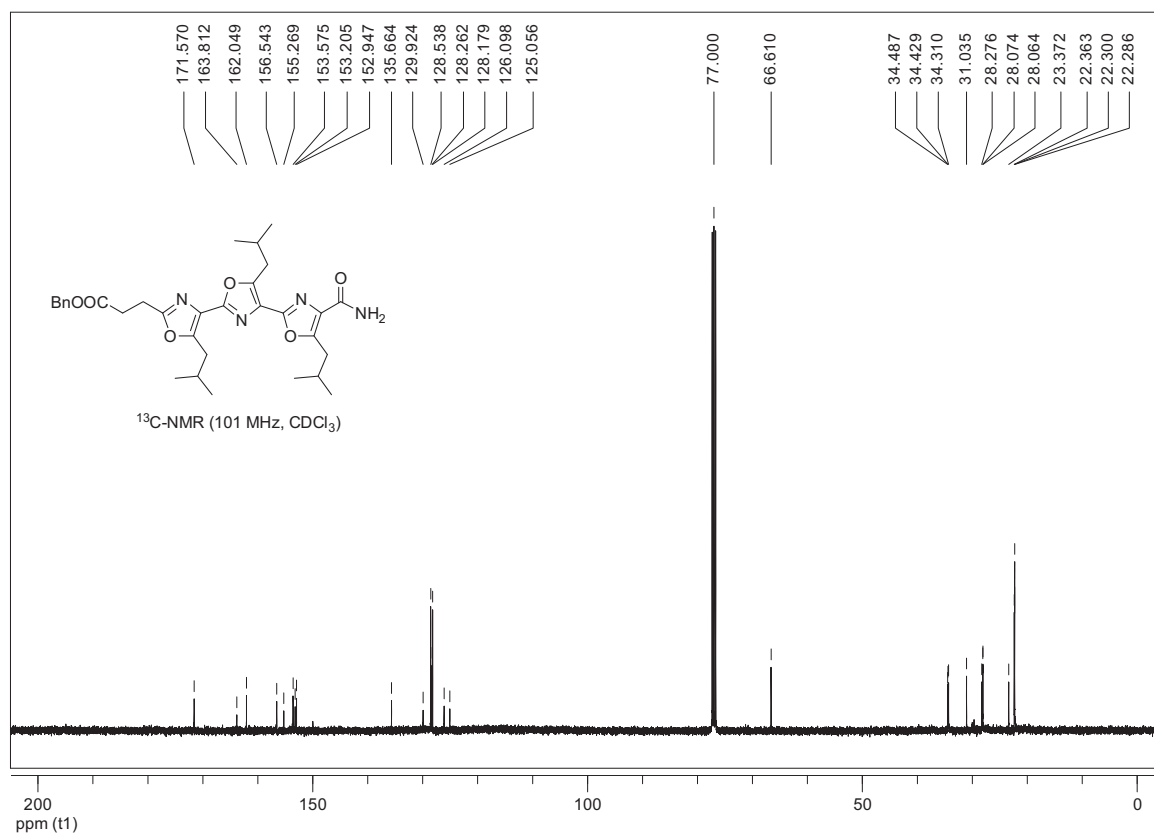
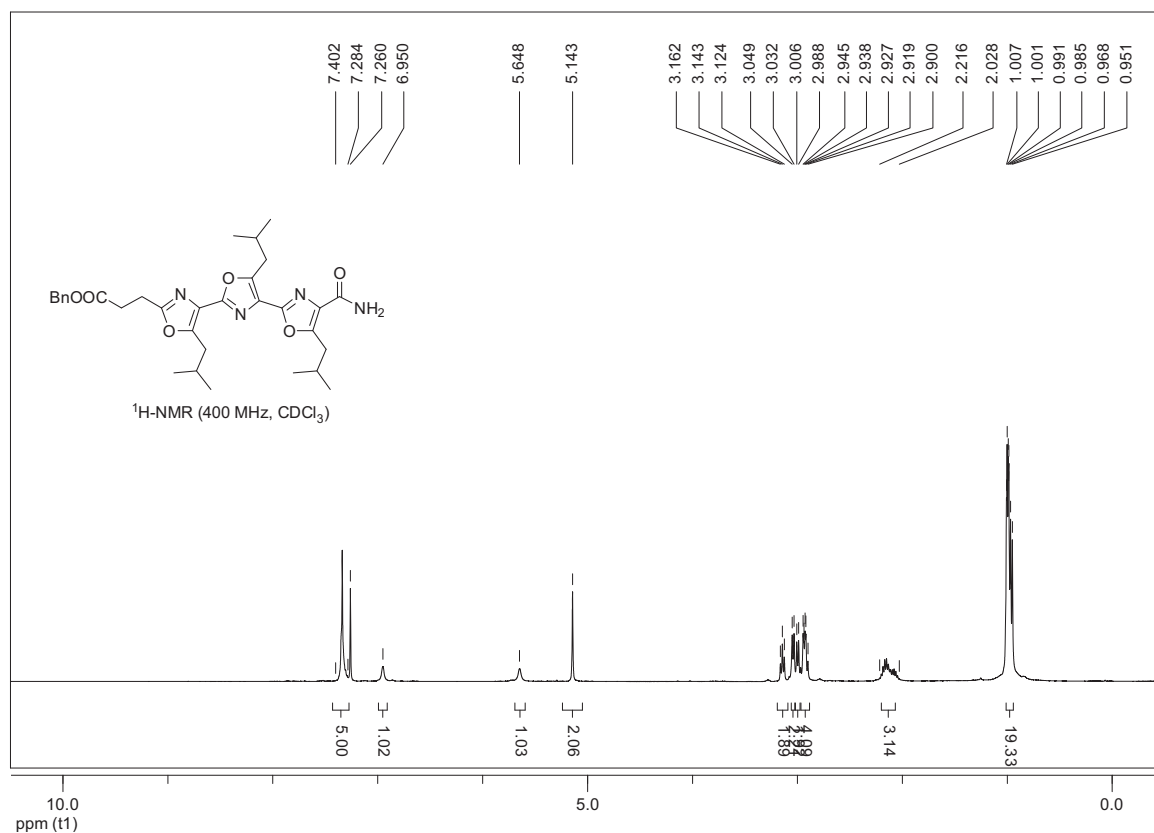
2-Benzyl-5-(2-benzyloxycarbonylaminoethyl)-1,3-oxazole-4-carboxamide (14f)

3-(2-Benzyl-4-carbamoyl-1,3-oxazole-5-yl)propanoic Acid (14g)


2',5,5'-Trimethyl-[2,4'-bioxazole]-4-carboxamide (28)


2''-Benzyl-5-ethyl-5'-isobutyl-5''-methyl-[2,4':2',4''-teroxazole]-4-carboxamide (13b)

2'',5,5',5''-Tetramethyl-[2,4':2',4''-teroxazole]-4-carboxamide (13a)


3-(4''-Carbamoyl-5,5',5''-triisobutyl-[2',4:2'',4'-teroxazole]-2-yl)propanoic Acid Benzyl Ester (13c)

Publication IV

From Determinants of RUNX1/ETO Tetramerization to Small-Molecule Protein-Protein Interaction Inhibitors Targeting Acute Myeloid Leukemia

Alexander Metz,[§] Julia Schanda,[§]
Manuel Grez,⁺⁺ Christian Wichmann,⁺⁺ and Holger Gohlke⁺⁺
Journal of Chemical Information and Modeling **2013**, 53: 2197-2202.

[§]Both authors have contributed equally to the respective work.

⁺⁺These authors share senior authorship.

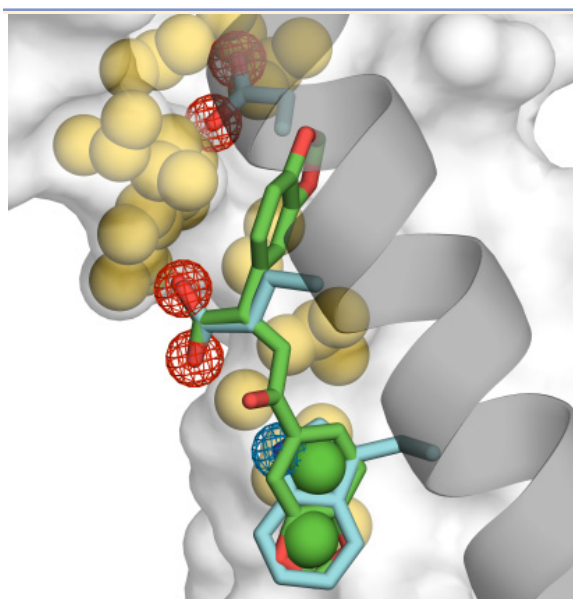
From Determinants of RUNX1/ETO Tetramerization to Small-Molecule Protein–Protein Interaction Inhibitors Targeting Acute Myeloid Leukemia

Alexander Metz,^{†,§} Julia Schanda,^{‡,§} Manuel Grez,^{*,‡,||} Christian Wichmann,^{‡,||,*,} and Holger Gohlke^{*,†,||}

[†]Institute for Pharmaceutical and Medicinal Chemistry, Department of Mathematics and Natural Sciences, Heinrich-Heine-University, Universitätsstr. 1, 40225 Düsseldorf, Germany

[‡]Institute for Biomedical Research, Georg-Speyer-Haus, Paul-Ehrlich-Str. 42-44, 60596 Frankfurt, Germany

S Supporting Information



ABSTRACT: We identified the first small-molecule protein–protein interaction inhibitors of RUNX1/ETO tetramerization applying structure-based virtual screening guided by predicted hot spots and pockets in the interface. A 3D similarity screening revealed specific hot spot mimetics, one of which prevents the proliferation of RUNX1/ETO-dependent SKNO-1 cells at low micromolar concentration. Using solely a protein–protein complex structure to start with, this strategy can be the first step in any comparable structure-based endeavor to identify protein–protein interaction inhibitors.

The chromosomal translocation t(8;21) is frequently found in acute myeloid leukemia (AML).² This translocation involves the *RUNX1* gene, a key regulator of hematopoietic cell differentiation,³ and the *ETO* gene, containing a novel homology region 2 (NHR2) oligomerization domain.⁴ Subsets of RUNX1/ETO-dependent AML are associated with unfavorable prognoses and high relapse rates after chemotherapy,⁵ e.g., when coinciding with c-KIT mutations.² We and others

demonstrated that homotetramerization of the α -helical NHR2 domain of the RUNX1/ETO fusion protein (also AML1/ETO) is an essential prerequisite for the onset and maintenance of AML.^{1,4,6–8} Recently, the NHR2 containing CBFA2T3/GLIS2 fusion protein has been linked to an aggressive subtype of pediatric acute megakaryoblastic leukemia.^{9,10} Thus, interfering with NHR2-mediated tetramerization in oncoproteins is an attractive strategy for molecular intervention and personalized therapy.^{11,12}

We reported a 128mer fusion protein containing the complete NHR2 domain (NC128) that disrupts RUNX1/ETO tetramerization and counteracts leukemic cell characteristics.¹¹ Likewise, a cell-penetrating TAT-NHR2 fusion protein interferes with the leukemogenic function of RUNX1/ETO.¹³ While providing a proof-of-principle for the molecular disruption of RUNX1/ETO tetramerization, these fusion proteins lack favorable ADME properties. Thus, we set out to identify small-molecule protein–protein interaction inhibitors (PPII) with the same mode of action but better pharmacokinetic properties.

The NHR2 tetramer is a symmetric dimer of dimers, each of which contains two extended α -helical monomers that associate in a head-to-tail orientation to form a four-helix bundle (Figure 1a).⁴ Targeting such protein–protein interactions (PPI) is considered difficult because of the size, lack of deep binding pockets, and stability of PPI.¹⁴ However, the widespread identification of (drug-like) PPII demonstrates that this challenge can be overcome.^{15,16} Often PPII act by targeting a subregion of the interface that contains hot spots and pockets.^{16,17}

Here, we report the first PPII of RUNX1/ETO tetramerization that prevent the proliferation of RUNX1/ETO-dependent cells and were identified by a combination of computational and experimental methods. Key to this success was a computational strategy we introduced recently (Figure 1).¹⁸ This strategy starts from the structure of a protein–protein complex and simultaneously considers aspects of energetics and plasticity relevant for PPII binding to an interface. Hence, we analyzed the energetic contribution of individual amino acids to the NHR2 dimer–tetramer transition (tetramerization) by a structural decomposition of a computed binding free energy¹⁹ and predicted a cluster of five amino acids (W498, W502,

Published: August 19, 2013

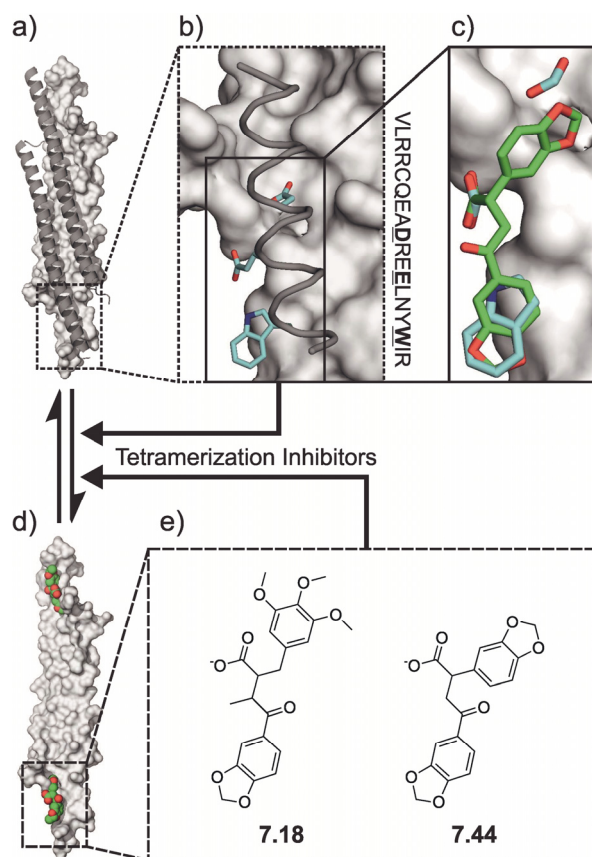


Figure 1. Identification of inhibitors of NHR2 tetramerization. (a) Starting from the NHR2 tetramer (PDB: 1wq6⁴) we derived an (b) inhibitory peptide (**P1**, ribbon) carrying three previously identified hot spots of NHR2 tetramerization (cyan; highlighted in the sequence).¹ (c) Small molecules were screened by ROCS for their ability to mimic the hot spot pattern (**7.44** superimposed with query (**III**)). (d) This led to the identification of inhibitors of NHR2 tetramerization that are proposed to bind to either one of the two equivalent binding regions in the interface of the NHR2 dimer. (e) The most potent inhibitors of NHR2 tetramerization **7.18** and **7.44**.

D533, E536, W540; RUNX1/ETO sequence numbering; Figure S1, Supporting Information) with strong contributions to the stability of the tetramer ("hot spots").¹

Mutating the hot spots to alanine abolishes tetramer formation without affecting the dimer formation or helicity of NHR2.¹ Moreover, RUNX1/ETO dimers do not block myeloid differentiation, are unable to enhance the self-renewal capacity of hematopoietic progenitors, and fail to induce leukemia in a murine transplantation model.¹ Scanning the NHR2 dimer–dimer interface²⁰ revealed the deepest and largest pocket (269 Å³; Figure S1a, Supporting Information), which is still rather shallow compared to binding sites in conventional^{21,22} and other protein–protein²³ targets. Yet, pockets are considered important for the high-affinity binding of PPII.^{16,24} Considering that this pocket is close to the cluster of identified hot spots and even binds the most important hot spot D533,¹ our analyses revealed an essential structural motif in the NHR2 interface that is suitable for intervention in t(8;21) leukemia.¹

While structure-based drug design approaches previously helped identifying PPII,^{25,26} mostly starting from peptide or ligand-bound structures, we decided to proceed with the structure-based identification of inhibitors of the NHR2 tetramerization based on the essential structural motif. Further analyses of the five hot spots suggested D533, E536, and W540 as the most suitable template motif (Figure 1b) for the following reasons: these hot spots decorate one face of each α -helix of an NHR2 dimer, project side chains from sequence positions *i*, *i*+3, and *i*+7 similar to known helix mimetics,^{27–29} cluster sufficiently close together to be replaced by a drug-like molecule,¹⁶ and extend into the deepest pocket of the NHR2 interface. In addition, they contact the NHR2 interface by a balanced mix of charged/acceptor (D533, E536) and aromatic/donor (W540) interactions. This would allow mimetics of the hot spots to be less hydrophobic than the common helix mimetic scaffolds^{27,28,30} and offer more specificity for binding than purely hydrophobic interactions.

On the basis of this template motif, we suggested the 18mer peptide **P1** as an inhibitor of NHR2 tetramerization. **P1** was derived from the NHR2 domain (Figure 1b; Figure S2a, Supporting Information) and encloses the template motif. The N-terminal 12mer peptide (Ac-EADREELNYWIR-NH₂) has an α -helical content of ~35% as determined by circular dichroism spectroscopy (Figure S3, Supporting Information). Convincingly, BS³ cross-linking experiments revealed that **P1** inhibits the NHR2 tetramerization with IC₅₀ \approx 250 μ M (Figure 2a–c). The moderate IC₅₀ reflects the competition of **P1** against the strong tendency of NHR2 dimers to self-associate. The NHR2 tetramer has a melting point of 85 °C,¹ and we could not detect unbound dimers for any construct containing the wild type NHR2 domain by either size-exclusion chromatography or analytical ultracentrifugation.¹ Still, this data provides the first proof-of-principle that the NHR2 tetramerization can be inhibited by a molecule significantly smaller than the NHR2 domain itself.

Further BS³ cross-linking experiments revealed that alanine mutations of the three hot spot residues in **P1** (resulting in peptide **P2**) or another two interface residues projecting toward the interface (peptide **P3**) abolish the inhibitory effect (Figure 2a). As these results reconfirmed the importance of the hot spot residues, we used this information together with the structural knowledge of the location and orientation of these residues' side chains in the crystal structure in a virtual screening (VS) for PPII mimicking the hot spot interactions.

To this end, we built three queries (Figure S4, Supporting Information) containing (I) only the carboxylic and indole functional groups of D533, E536, and W540, (II) a terpyridine helix mimetic²⁹ decorated with the functional groups from (I) and with the terpyridine modeled in the position of the helix that should be replaced, or (III) the functional groups from (I) plus repulsive Gaussian potentials centered on all adjacent protein interface atoms (≤ 4 Å distance) to penalize molecules clashing with the interface. With these queries, we searched the refined (Supporting Information) purchasable subset of the ZINC 11 database.³¹ We screened $\sim 6 \times 10^6$ molecules with ROCS (Rapid Overlay of Chemical Structures)³² that superimposes molecules based on a 3D overlap of molecular shape and donor, acceptor, anionic, cationic, hydrophobic, and aromatic (ring) properties, referred to as colors.

We performed three ROCS searches on the conformers, one for each query. Also, to account for the conformational variability of the target and the template motif, each search was

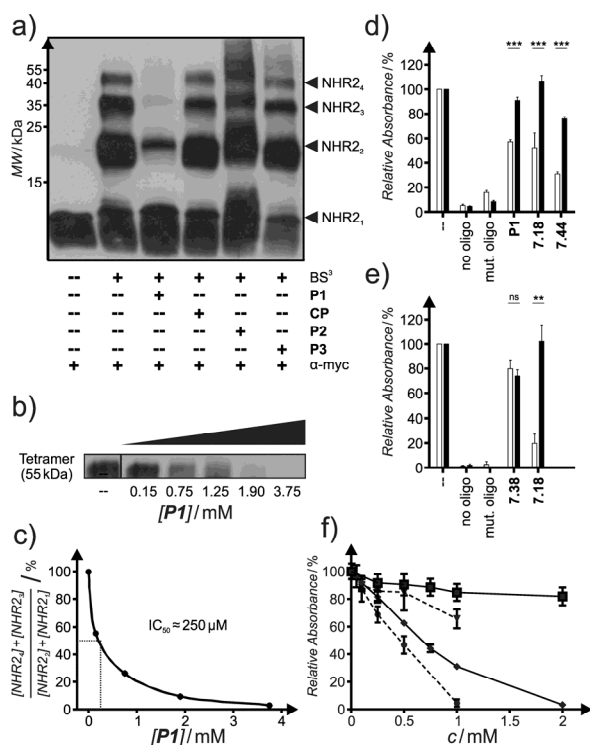


Figure 2. Inhibition of NHR2 tetramerization. (a) Peptide **P1**, but neither **P2**, **P3**, nor the unrelated peptide **CP** inhibit NHR2 tetramerization as shown by the reduction of BS^3 cross-linked NHR2 oligomers ($c = 1100 \mu\text{M}$). (b) The inhibition by **P1** is dose-dependent (c) with $\text{IC}_{50} \approx 250 \mu\text{M}$. (d) ELISA and (e) ABCD experiments show that **P1**, **7.18**, and **7.44**, but not **7.38**, selectively inhibit tetramer-dependent binding of the RUNX1-NHR2 protein (white bars) to an immobilized RUNX3 oligonucleotide while binding of the RUNX1-BCR protein (black bars) is not inhibited ($[\text{P1}] = 500 \mu\text{M}$ and $[\text{PPII}] = 2000 \mu\text{M}$ in ELISA; $[\text{PPII}] = 1000 \mu\text{M}$ in ABCD assay). (f) Dose-dependent inhibition of RUNX1-NHR2 tetramerization in the ELISA by **P1** (●; $\text{IC}_{50} = 390 \pm 30 \mu\text{M}$) and **7.44** (◆; $\text{IC}_{50} = 630 \pm 24 \mu\text{M}$) but neither by **P3** (▼) nor by **7.38** (■). Standard deviations for $n \geq 2$. ns: not significant. **: $0.01 > p \geq 0.001$. ***: $p < 0.0001$ (unpaired *t*-test).

performed independently using side chain conformations of the hot spots from either end of the NHR2 interface. These interface conformations differ by a root-mean-square deviation (RMSD) of 0.3 \AA (Figure S1b,c, Supporting Information). All ROCS results were reranked exclusively by color overlap to favor compounds that optimally mimic the hot spot functional groups. For each search, the top scoring ZINC entries of either set of side chain conformations were combined; duplicates were removed.

The resulting poses vividly illustrate to what extent substructures of compounds overlap with the functional groups of the hot spots (Figure 1c; Figure S4d–f, Supporting Information). We assessed the 1000 top scoring compound poses for each query visually based on interaction complementarity, steric overlap with the targeted interface and the helix to be replaced, conformational strain, and structural diversity. Finally, we selected 80 compounds (Table S1, Supporting Information) for experimental testing.

We measured the inhibition of NHR2 tetramerization in vitro by tetramer-dependent binding of a RUNX1-NHR2

protein construct (Figure 2; Figure S2b, Supporting Information) to an immobilized oligonucleotide, which was derived from the RUNX3 promoter sequence, in ELISA and ABCD assays³³ (Figure 2d–f; Figures S5 and S6 and Table S2, Supporting Information). Seven compounds showed consistent and selective inhibition in both assays (Table S2, Supporting Information) with activities comparable to **P1** (42.8% inhibition at $c = 500 \mu\text{M}$ and $\text{IC}_{50} = 390 \pm 30 \mu\text{M}$ by ELISA; Figure 2f; Figure S7, Supporting Information). The most promising compound **7.18** (Figure 1e) showed an NHR2 inhibition of 80.3% ($c = 1 \text{ mM}$) in the ABCD and 47.7% ($c = 2 \text{ mM}$) in the ELISA assay (Table S2, Supporting Information) but was inactive in cells. Selectivity for NHR2 was demonstrated by the reduced inhibition when replacing the NHR2 domain of the construct with the homologous BCR tetramerization domain (RUNX1-BCR protein; Figure 2; Figure S2b, Supporting Information); this finding also suggests that the observed inhibition of NHR2 tetramerization in RUNX1-NHR2 does not occur due to nonspecific protein binding.

Structurally, **7.18** contains a central 4-oxobutanoic acid scaffold substituted by a 4-(1,3-benzodioxolyl) and a 2-(3,4,5-trimethoxy-benzyl) moiety (Figure 1e). The ROCS pose shows how **7.18** mimics the template motif. The PPII's carboxylate group mimics that of E536, while the benzodioxolyl and trimethoxybenzyl moieties mimic the indole of W540 and the carboxylate of D533, respectively (Figure S4d, Supporting Information).

Next, a fingerprint similarity search³⁴ for structural analogs of **7.18** in the refined database³¹ revealed **7.44** with improved in vitro activity ($\text{IC}_{50} = 630 \pm 24 \mu\text{M}$ by ELISA; Figure 2; Figure S7, Supporting Information). This activity is only 1.6-fold lower than that of **P1**. The related Hill coefficient for **7.44** is 1.8, which does not contradict the assumed inhibitory mechanism: Once a first PPII interferes with NHR2 tetramer formation, the second symmetry-related binding site on the NHR2 dimer becomes more easily accessible such that the affinity for the second PPII increases. Notably, for **P1** a similar Hill coefficient of 1.9 is found (Figure S7, Supporting Information), suggesting a common mechanism for both inhibitors as expected. Furthermore, **7.44** selectively reduced the viability of RUNX1/ETO-dependent human leukemic SKNO-1 cells ($\text{EC}_{50} < 10 \mu\text{M}$) in accord with the effect of NC128,¹¹ whereas treatment of these cells with inactive **7.38** or of RUNX1/ETO-independent U937 cells with **7.44** had no effect (Figure 3).

The improved activity can be explained by comparing the ROCS poses of **7.18** and **7.44** (Figure S4d–f, Supporting Information). **7.44** overlaps less with the protein by replacing the trimethoxybenzyl of **7.18** by a shorter, more compact, less flexible, and hence also entropically more favorable second benzodioxolyl moiety that mimics the carboxylate of D533 (Figure 1c). The ROCS ranks suggest a preferred binding of the 2S stereoisomers of **7.18** and **7.44**. In addition, a reversed binding mode is found for **7.44** by ROCS (Figure S4f, Supporting Information). Crystal structure analyses^{35,36} and a search in the SwissBioisostere database³⁷ support this mimicry of both carboxylate and indole groups by 1,3-benzodioxoles. The proposed binding modes will help suggesting further modifications of these lead compounds. In that respect, both PPII are good starting points for optimization due to their low molecular weight (**7.18**: 415 Da; **7.44**: 341 Da), simple chemical structure, and ligand efficiency ($\text{LE}_{7.44} = -0.18$

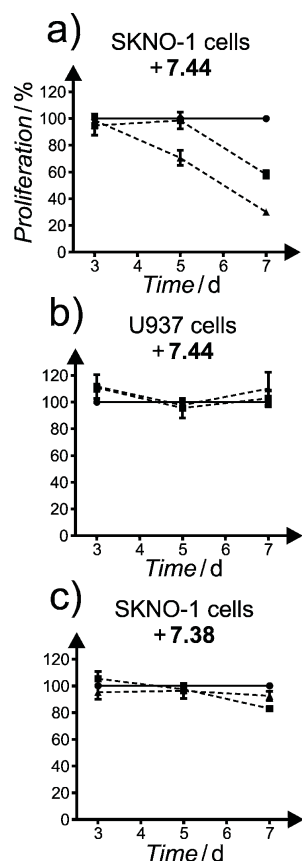


Figure 3. NHR2 inhibitor 7.44 specifically reduces proliferation of RUNX1/ETO-dependent cells. SKNO-1 cell or U937 cells were treated with 1 μ M (■) or 10 μ M (▲) of 7.44 or 7.38 or no PPII (●). (a) RUNX1/ETO-dependent SKNO-1 cells were severely affected by 7.44 (b) while RUNX1/ETO-independent U937 cells were unaffected. 7.38, inactive in vitro, did not affect SKNO-1 cells. Proliferation measured by XTT assay. Standard deviations for $n \geq 3$.

kcal mol⁻¹ related to the IC₅₀) that is already similar to PPII employed in the clinics.^{16,38}

7.44 was recently shown to inhibit c-Jun N-terminal kinase (JNK) and affect the JNK-pathway in cells.³⁹ This may provide an explanation as to why 7.44 reduces the viability of RUNX1/ETO-dependent SKNO-1 cells at an EC₅₀ value much lower than the IC₅₀ value found in in vitro experiments on the inhibition of NHR2 tetramerization (see above). In turn, our results indicate that 7.44 could exert an additional effect on the JNK-pathway in that RUNX1/ETO-mediated activation of the JNK-pathway^{40,41} can also be inhibited by inhibition of NHR2 tetramerization, as demonstrated by the absence of c-Jun upregulation in the case of RUNX1/ETO with hot spot residues mutated to alanine.¹ Finally, the study excluded the possibility of unspecific protein binding of 7.44 in terms of aggregates as did it exclude an unspecific action of 7.44 as a redox cycling compound.³⁹ 7.44 is structurally similar to (epi)podophyllotoxins; the latter are non-intercalating inhibitors of topoisomerase II³⁶ and microtubule formation.⁴² We can exclude that 7.44 acts that way because of the absence of the respective targets in our ELISA and ABCD assays.

In summary, from the determinants of NHR2 tetramerization¹ previously revealed by computational hot spot¹⁸ and pocket prediction,²⁰ a molecular recognition pattern at the NHR2 dimer interface was defined. This pattern successfully guided VS resulting in the first small-molecule inhibitors of NHR2 tetramerization, one of which prevents the proliferation of RUNX1/ETO-dependent SKNO-1 cells at a low micromolar concentration. These results could guide further efforts to intervene with RUNX1/ETO-positive AML and other NHR2 tetramerization-dependent mechanisms. Our results furthermore demonstrate that by this computational strategy small-molecule PPII can be identified even in cases where nothing else than a protein–protein complex structure is known. Hence, this strategy can be the first step in any comparable structure-based endeavor to identify PPII.

■ ASSOCIATED CONTENT

§ Supporting Information

Tables with the tested substances (Table S1) and the in vitro inhibition of NHR2 tetramerization (Table S2) as well as graphical representations of the hot spot containing region of the NHR2 interface (Figure S1), experimentally tested peptides and proteins (Figure S2), helical content of the inhibitory 12mer peptide (Figure S3), ROCS queries used for VS and exemplary ROCS poses of the most active inhibitors (Figure S4), biochemical assays for measuring inhibition of NHR2 tetramerization (Figure S5), ELISA screening results for inhibitors of RUNX1/ETO tetramerization (Figure S6), and the dose-dependent inhibition of RUNX1-NHR2 tetramerization in the ELISA (Figure S7). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: grez@gsh.uni-frankfurt.de (M.G.).

*E-mail: christian.wichmann@med.uni-muenchen.de (C.W.).

*E-mail: gohlke@uni-duesseldorf.de (H.G.). Fax: (+49) 211-8113847 (H.G.).

Present Address

[#](C.W.) Department of Transfusion Medicine, Cell Therapeutics and Hemostasis, Ludwig-Maximilian University Hospital, Munich, Germany.

Author Contributions

[§]These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

^{||}These authors share senior authorship.

■ ACKNOWLEDGMENTS

We are grateful for financial support from the NGFNplus Cancer Network (Grant 01GS0879), Deutsche Krebshilfe (Grant 102362, TP7), LOEWE Center for Cell and Gene Therapy Frankfurt (HMWK III L 4-518/17.004 [2010]), "Strategischer Forschungsfonds" at Heinrich-Heine-University, and "Kind-Philipp-Stiftung"; for computational support by the "Zentrum für Informations und Medientechnologie" at the Heinrich-Heine-University; to the NCI/DTP Open Chemical Repository (<http://dtp.cancer.gov/>) for compound samples; and to OpenEye for an academic license.

■ ABBREVIATIONS

ABCD assay, avidin–biotin complex DNA assay; AML, acute myeloid leukemia; ELISA, enzyme-linked immunosorbent assay; JNK, c-Jun N-terminal kinase; NHR2, nervy homology region 2; PPI, protein–protein interaction; PPII, protein–protein interaction inhibitor; RMSD, root-mean-square deviation; ROCS, rapid overlay of chemical structures; VS, virtual screening

■ REFERENCES

- (1) Wichmann, C.; Becker, Y.; Chen-Wichmann, L.; Vogel, V.; Vojtkova, A.; Herglotz, J.; Moore, S.; Koch, J.; Lausen, J.; Mantele, W.; Gohlke, H.; Grez, M. Dimer-tetramer transition controls RUNX1/ETO leukemogenic activity. *Blood* **2010**, *116* (4), 603–613.
- (2) Reikvam, H.; Hatfield, K. J.; Kittang, A. O.; Hovland, R.; Bruserud, O. Acute myeloid leukemia with the t(8;21) Translocation: Clinical consequences and biological implications. *J. Biomed. Biotechnol.* **2011**, *104631*, 1–23.
- (3) Tenen, D. G. Disruption of differentiation in human cancer: AML shows the way. *Nat. Rev. Cancer* **2003**, *3* (2), 89–101.
- (4) Liu, Y.; Cheney, M. D.; Gaudet, J. J.; Chruszcz, M.; Lukasik, S. M.; Sugiyama, D.; Lary, J.; Cole, J.; Dauter, Z.; Minor, W.; Speck, N. A.; Bushweller, J. H. The tetramer structure of the nervy homology two domain, NHR2, is critical for AML1/ETO's activity. *Cancer Cell* **2006**, *9* (4), 249–260.
- (5) Shimada, A.; Taki, T.; Tabuchi, K.; Tawa, A.; Horibe, K.; Tsuchida, M.; Hanada, R.; Tsukimoto, I.; Hayashi, Y. KIT mutations, and not FLT3 internal tandem duplication, are strongly associated with a poor prognosis in pediatric acute myeloid leukemia with t(8;21): A study of the Japanese Childhood AML Cooperative Study Group. *Blood* **2006**, *107* (5), 1806–1809.
- (6) Ptasińska, A.; Assi, S. A.; Mannari, D.; James, S. R.; Williamson, D.; Dunne, J.; Hoogenkamp, M.; Wu, M.; Care, M.; McNeill, H.; Cauchy, P.; Cullen, M.; Toose, R. M.; Tenen, D. G.; Young, B. D.; Cockerill, P. N.; Westhead, D. R.; Heidenreich, O.; Bonifer, C. Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia* **2012**, *26* (1), 1829–1841.
- (7) Rhoades, K. L.; Hetherington, C. J.; Harakawa, N.; Yergeau, D. A.; Zhou, L. M.; Liu, L. Q.; Little, M. T.; Tenen, D. G.; Zhang, D. E. Analysis of the role of AML1-ETO in leukemogenesis, using an Inducible transgenic mouse model. *Blood* **2000**, *96* (6), 2108–2115.
- (8) Heidenreich, O.; Krauter, J.; Riehle, H.; Hadwiger, P.; John, M.; Heil, G.; Vormlocher, H. P.; Nordheim, A. AML1/MTG8 oncogene suppression by small interfering RNAs supports myeloid differentiation of t(8;21)-positive leukemic cells. *Blood* **2003**, *101* (8), 3157–3163.
- (9) Gruber, T. A.; Larson Gedman, A.; Zhang, J.; Koss, C. S.; Marada, S.; Ta, H. Q.; Chen, S. C.; Su, X.; Ogden, S. K.; Dang, J.; Wu, G.; Gupta, V.; Andersson, A. K.; Pounds, S.; Shi, L.; Easton, J.; Barbato, M. I.; Mulder, H. L.; Manne, J.; Wang, J.; Rusch, M.; Ranade, S.; Ganti, R.; Parker, M.; Ma, J.; Radtke, I.; Ding, L.; Cazzaniga, G.; Biondi, A.; Kornblau, S. M.; Ravandi, F.; Kantarjian, H.; Nimer, S. D.; Dohner, K.; Dohner, H.; Ley, T. J.; Ballerini, P.; Shurtleff, S.; Tomizawa, D.; Adachi, S.; Hayashi, Y.; Tawa, A.; Shih, L. Y.; Liang, D. C.; Rubnitz, J. E.; Pui, C. H.; Mardis, E. R.; Wilson, R. K.; Downing, J. R. An Inv(16)(p13.3q24.3)-encoded CBF A2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell* **2012**, *22* (5), 683–697.
- (10) Thiollier, C.; Lopez, C. K.; Gerby, B.; Ignacimoutou, C.; Poglio, S.; Duffourd, Y.; Guegan, J.; Rivera-Munoz, P.; Bluteau, O.; Mabalal, V.; Diop, M.; Wen, Q.; Petit, A.; Bauchet, A. L.; Reinhardt, D.; Bornhauser, B.; Gautheret, D.; Lecluse, Y.; Landman-Parker, J.; Radford, I.; Vainchenker, W.; Dastugue, N.; de Botton, S.; Dessen, P.; Bourquin, J. P.; Crispino, J. D.; Ballerini, P.; Bernard, O. A.; Pflumio, F.; Mercher, T. Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models. *J. Exp. Med.* **2012**, *209* (11), 2017–2031.
- (11) Wichmann, C.; Chen, L. P.; Heinrich, M.; Baus, D.; Pfützner, E.; Zornig, M.; Ottmann, O. G.; Grez, M. Targeting the oligomerization domain of ETO interferes with RUNX1/ETO oncogenic activity in t(8;21)-positive leukemic cells. *Cancer Res.* **2007**, *67* (5), 2280–2289.
- (12) Sun, X. J.; Wang, Z.; Wang, L.; Jiang, Y.; Kost, N.; Soong, T. D.; Chen, W. Y.; Tang, Z.; Nakada, T.; Elemento, O.; Fischle, W.; Melnick, A.; Patel, D. J.; Nimer, S. D.; Roeder, R. G. A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis. *Nature* **2013**, *500* (7460), 93–97.
- (13) Bartel, Y.; Grez, M.; Wichmann, C. Interference with RUNX1/ETO leukemogenic function by cell-penetrating peptides targeting the NHR2 oligomerization domain. *Biomed. Res. Int.* **2013**, *2013*, 297692.
- (14) Surade, S.; Blundell, T. L. Structural biology and drug discovery of difficult targets: The limits of ligandability. *Chem. Biol.* **2012**, *19* (1), 42–50.
- (15) Zinzalla, G.; Thurston, D. E. Targeting protein–protein interactions for therapeutic intervention: A challenge for the future. *Future Med. Chem.* **2009**, *1* (1), 65–93.
- (16) Metz, A.; Ciglia, E.; Gohlke, H. Modulating protein–protein interactions: From structural determinants of binding to druggability prediction to application. *Curr. Pharm. Des.* **2012**, *18*, 4630–4647.
- (17) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot spots: A review of the protein–protein interface determinant amino-acid residues. *Proteins: Struct., Funct., Bioinf.* **2007**, *68* (4), 803–812.
- (18) Metz, A.; Pfeiffer, C.; Kopitz, H.; Pfeiffer-Marek, S.; Baringhaus, K. H.; Gohlke, H. Hot spots and transient pockets: Predicting the determinants of small-molecule binding to a protein–protein interface. *J. Chem. Inf. Model.* **2012**, *52* (1), 120–133.
- (19) Gohlke, H.; Kiel, C.; Case, D. A. Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes. *J. Mol. Biol.* **2003**, *330* (4), 891–913.
- (20) Craig, I. R.; Pfeiffer, C.; Gohlke, H.; Essex, J. W.; Spiegel, K. Pocket-space maps to identify novel binding-site conformations in proteins. *J. Chem. Inf. Model.* **2011**, *51* (10), 2666–2679.
- (21) Kahraman, A.; Morris, R. J.; Laskowski, R. A.; Thornton, J. M. Shape variation in protein binding pockets and their ligands. *J. Mol. Biol.* **2007**, *368* (1), 283–301.
- (22) Perot, S.; Sperandio, O.; Miteva, M. A.; Camproux, A. C.; Villoutreix, B. O. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discovery Today* **2010**, *15* (15–16), 656–667.
- (23) Eyrisch, S.; Helms, V. Transient pockets on protein surfaces involved in protein–protein interaction. *J. Med. Chem.* **2007**, *50* (15), 3457–3464.
- (24) Higuero, A. P.; Schreyer, A.; Bickerton, G. R. J.; Pitt, W. R.; Groom, C. R.; Blundell, T. L. Atomic interactions and profile of small molecules disrupting protein–protein interfaces: The TIMBAL Database. *Chem Biol Drug Des* **2009**, *74* (5), 457–467.
- (25) Geppert, T.; Bauer, S.; Hiss, J. A.; Conrad, E.; Reutlinger, M.; Schneider, P.; Weisel, M.; Pfeiffer, B.; Altmann, K. H.; Waibler, Z.; Schneider, G. Immunosuppressive small molecule discovered by structure-based virtual screening for inhibitors of protein–protein interactions. *Angew. Chem., Int. Ed.* **2012**, *51* (1), 258–261.
- (26) Mukherjee, P.; Desai, P.; Zhou, Y. D.; Avery, M. Targeting the BH3 domain mediated protein–protein interaction of Bcl-xL through virtual screening. *J. Chem Inf Model* **2010**, *50* (5), 906–923.
- (27) Kim, I. C.; Hamilton, A. D. Diphenylindane-based proteomimetics reproduce the projection of the i, i+3, i+4, and i+7 residues on an alpha-helix. *Org. Lett.* **2006**, *8* (9), 1751–1754.
- (28) Yin, H.; Hamilton, A. D. Strategies for targeting protein–protein interactions with synthetic agents. *Angew. Chem., Int. Ed.* **2005**, *44* (27), 4130–4163.
- (29) Davis, J. M.; Truong, A.; Hamilton, A. D. Synthesis of a 2,3,6'-terpyridine scaffold as an alpha-helix mimetic. *Org. Lett.* **2005**, *7* (24), 5405–5408.
- (30) Cummings, C. G.; Hamilton, A. D. Disrupting protein–protein interactions with non-peptidic, small molecule alpha-helix mimetics. *Curr. Opin. Chem. Biol.* **2010**, *14* (3), 341–346.

- (31) Irwin, J. J.; Shoichet, B. K. ZINC—A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177–182.
- (32) OEChem, version 1.7.2; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2009.
- (33) McKay, I. A.; Kirby, L.; Volyanik, E. V.; Kumar, V.; Wong, P. W. Y.; Bustin, S. A. An enzyme-linked immunosorbent assay for the detection of agents which interfere with the DNA binding activities of transcription factors - Exemplified by NF-IL6. *Anal. Biochem.* **1998**, *265* (1), 28–34.
- (34) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): Evaluation of performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1708–1718.
- (35) Davey, D. D.; Adler, M.; Arnaiz, D.; Eagen, K.; Erickson, S.; Guilford, W.; Kenrick, M.; Morrissey, M. M.; Ohlmeyer, M.; Pan, G.; Paradkar, V. M.; Parkinson, J.; Polokoff, M.; Saionz, K.; Santos, C.; Subramanyam, B.; Vergona, R.; Wei, R. G.; Whitlow, M.; Ye, B.; Zhao, Z. S.; Devlin, J. J.; Phillips, G. Design, synthesis, and activity of 2-imidazol-1-ylpyrimidine derived inducible nitric oxide synthase dimerization inhibitors. *J. Med. Chem.* **2007**, *50* (6), 1146–1157.
- (36) Wu, C. C.; Li, T. K.; Farh, L.; Lin, L. Y.; Lin, T. S.; Yu, Y. J.; Yen, T. J.; Chiang, C. W.; Chan, N. L. Structural basis of type II topoisomerase inhibition by the anticancer drug etoposide. *Science* **2011**, *333* (6041), 459–462.
- (37) Wirth, M.; Zoete, V.; Michielin, O.; Sauer, W. H. B. SwissBioisostere: A database of molecular replacements for ligand design. *Nucleic Acids Res.* **2013**, *41* (D1), D1137–D1143.
- (38) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* **2007**, *450* (7172), 1001–1009.
- (39) Kaoud, T. S.; Yan, C.; Mitra, S.; Tseng, C. C.; Jose, J.; Taliaferro, J. M.; Tuohetahuntala, M.; Devkota, A.; Sammons, R.; Park, J.; Park, H.; Shi, Y.; Hong, J.; Ren, P.; Dalby, K. N. From in silico discovery to intra-cellular activity: Targeting JNK-protein interactions with small molecules. *ACS Med. Chem. Lett.* **2012**, *3* (9), 721–725.
- (40) Elsasser, A.; Franzen, M.; Kohlmann, A.; Weissner, M.; Schnittger, S.; Schoch, C.; Reddy, V. A.; Burel, S.; Zhang, D. E.; Ueffing, M.; Tenen, D. G.; Hiddemann, W.; Behre, G. The fusion protein AML1-ETO in acute myeloid leukemia with translocation t(8;21) induces c-jun protein expression via the proximal AP-1 site of the c-jun promoter in an indirect, JNK-dependent manner. *Oncogene* **2003**, *22* (36), 5646–5657.
- (41) Gao, F. H.; Wang, Q.; Wu, Y. L.; Li, X.; Zhao, K. W.; Chen, G. Q. c-Jun N-terminal kinase mediates AML1-ETO protein-induced connexin-43 expression. *Biochem. Biophys. Res. Commun.* **2007**, *356* (2), 505–511.
- (42) Srivastava, V.; Negi, A. S.; Kumar, J. K.; Gupta, M. M.; Khanuja, S. P. S. Plant-based anticancer molecules: A chemical and biological profile of some important leads. *Bioorg. Med. Chem.* **2005**, *13* (21), 5892–5908.

Publication IV – Supporting Information

Supporting Information

From Determinants of RUNX1/ETO Tetramerization to Small Molecule Protein-Protein Interaction Inhibitors Targeting Acute Myeloid Leukemia

Alexander Metz,¹ Julia Schanda,² Manuel Grez,² Christian Wichmann,² Holger Gohlke¹

¹Heinrich-Heine-University Düsseldorf, Institute for Pharmaceutical and Medicinal Chemistry, Department of Mathematics and Natural Sciences, Universitätsstr. 1, 40225 Düsseldorf (Germany)

²Georg-Speyer-Haus, Institute for Biomedical Research, Paul-Ehrlich-Str. 42-44, 60596 Frankfurt (Germany)

Table of Contents

Materials and Methods	2
Ligand Data	2
Protein Data	2
Hot Spot Interactions	2
ROCS Queries and Searches	2
Fingerprint Similarity Search	3
Peptides and Small Molecule Inhibitors	3
Protein Preparation	4
Cross-Linking Assay	4
ELISA Assay	4
ABCD Assay	4
Cell Survival Assays	5
Supplementary Tables	6
Supplementary Figures	8
References in Supporting Information	15

Materials and Methods

Ligand Data

For prospective VS and ROCS searches we used the purchasable subset of the ZINC 11 database (as of May 2010)¹ including the provided protonation states.

For the ROCS searches, this database was further refined. First, we selected $\sim 6 \times 10^6$ drug-like (OpenEye default filter²), non-positively charged ZINC entries from the purchasable subset, allowing for a molecular weight up to 650 Da. That way, we account for the twofold negative charge and large size (14-16 Å in diameter) of the template motif. Second, we generated up to 100 conformers for each of these entries using OMEGA.² This yielded a total of $\sim 447 \times 10^6$ conformers that were submitted to subsequent ROCS searches.

Ligand-based similarity searches (hit expansion) using MOLPRINT fingerprint similarity³ were conducted on the same refined database.

Protein Data

The crystallographic protein structure of the NHR2 tetramer (PDB: 1wq6⁴) was obtained from the Protein Data Bank PDB.⁵

Hot Spot Interactions

The template motif formed by three of the five hot spots of NHR2 tetramerization (D533, E536, and W540) addresses the target NHR2 interface by forming a dense, complementary, and zipper-like network of polar and non-polar interactions (Figure S1b-c).⁶ In particular, D533 is a buried anchor residue⁴ that forms salt-bridges with *R534* and *R492* (residues of the target NHR2 interface in italics). Also, *R492* forms a solvent-exposed salt-bridge with E536 that, in turn, interacts by a charge-assisted hydrogen bond with the indole N-H group of W540. Moreover, E536 is stabilized by hydrophobic packing with *L537*. W540 forms a hydrophobic contact with *I541* and an aromatic edge-to-face interaction with *Y544*. Notably, D533 also forms a buried water mediated interaction with the backbone carbonyl and side chain amide of *Q530* and, mediated by another water molecule, to the approximately symmetric *D533*. In sum, the key recognition pattern that addresses the NHR2 interface consists of the carboxylate and indole groups of the side chains of the template motif D533, E536, and W540.

ROCS Queries and Searches

We built three ROCS queries, (I), (II), and (III), as outlined in the main text (Figure S4). All of these queries contain the carboxylic and indole functional groups of D533, E536, and W540.

For query (I) and query (III), we placed suitable colors at the crystallographic positions of the corresponding atoms of these functional groups in the template motif: acceptor colors were placed at the positions of each carboxylate oxygen atom, anion colors were placed at the positions of each carboxylate carbon atom, a donor color was placed at the position of the indole nitrogen, and two ring colors were placed at the center of each of the indole rings. All colors were positioned using the vROCS graphical user interface (Version 3.0.0) of ROCS.²

The terpyridine helix scaffold⁷ used for query (II) was manually modeled onto the crystallographic side chain conformation of D533, E536, and W540 in place of the NHR2 backbone using the Moloc molecular design suite⁸ (as of April 2008); the protein backbone and all other amino acids of the NHR2

dimer that carry the template motif were removed. The modeled terpyridine scaffold was then minimized using a two-step procedure. First, the scaffold was partially minimized within Moloc using the MAB all atom force field⁹ in the presence of the fixed structure of the target NHR2 dimer while also holding the functional carboxylic and indole groups fixed. Second, the decorated terpyridine structure was minimized in the presence of the fixed structure of the target NHR2 dimer. All minimizations were carried out using the Moloc default convergence criteria. Finally, we removed the target NHR2 dimer and placed suitable colors at the minimized positions of the corresponding atoms of these functional groups as described for the other queries.

For query (III), we defined all atoms of the target NHR2 dimer within 4 Å of the template motif, including hydrogen atoms added to the crystal structure with the xLEaP module of the AMBER suite of programs,¹⁰ to be gold atoms. Gold was then added as a new color to the Implicit Mills Dean color force field¹¹ and was defined to be repulsive to any other color that may be contained in a ligand during a ROCS search. The weight of the repulsive potential was set equal to all other color interactions.

All atoms of these queries, except the repulsive gold atoms of the target NHR2 interface, were enclosed by a shape according to the ROCS default. In order to also account for the conformational variability of the target and the template motif, we created ROCS queries using side chain conformations of the hot spot residues from either end of the approximately symmetrical NHR2 dimer. These conformations structurally differ by a root mean-square deviation (RMSD) of 0.3 Å (Figure S4).

ROCS searches were then carried out on all conformers from the refined ZINC 11 database (see above) using the described 3 x 2 ROCS queries. The Implicit Mills Dean color force field¹¹ was used and, in the case of query (III), was modified to include the repulsive gold color described above. Poses were optimized using the gradients of the color force field. The top scoring conformers of each entry, ranked by ComboScore, were retained. For each query type, respectively, the conformers screened by the ROCS searches for either end of the NHR2 dimer were combined and re-ranked by ColorScore. In the case of duplicate ZINC entries the top-ranked pose for each ZINC entry was selected. The overall top-ranked 1,000 poses of each query type were assessed visually based on color overlap, steric overlap with the targeted interface and the helix to be replaced, conformational strain, and structural diversity; 80 compounds were finally selected for experimental testing.

Fingerprint Similarity Search

Initially, we tested 80 compounds identified by the ROCS-based VS. To identify structurally related compounds of the initial hits **7.18** we performed a fingerprint-based similarity search.¹² For this, we calculated Tanimoto coefficients, based on MOLPRINT³ 2D fingerprints, between both of these compounds and all ZINC entries from the refined purchasable subset of the ZINC 11 database. We selected the 60 ZINC entries most similar to each of the two initial hits. All selected ZINC entries were combined; duplicates and compounds already tested were removed. The ROCS poses of the remaining compounds were assessed visually and 14 compounds were selected for a second round of experimental testing (Table S1). One of these compounds (**7.41**) was not contained in the refined ZINC database but was identified using the similarity search of the ZINC homepage and was also tested experimentally.

Peptides and Small Molecule Inhibitors

Peptides were synthesized by GenScript and Dr. Diana Imhof (University of Bonn). Control peptides were kindly provided by Dr. Joachim Koch (Georg-Speyer-Haus). Peptides were dissolved in dH₂O. Compounds were obtained from: **1** (Sigma Aldrich), **2a-2e** (Enamine), **3a-3b** (Chembridge), **4** (TimTec),

5.1-5.28 (Princeton), **6.1-6.3** (Otava), **7.1-7.44** (National Cancer Institute¹³), and **8.1-8.10** (Chemonaut). Compounds were dissolved in DMSO only or in dH₂O followed by dropwise addition of ammonia solution to a final concentration of 25%.

Protein Preparation

For protein expression, chemically competent BL21(DE3) *E. coli* (Invitrogen) were used. An overnight pre-culture containing ampicillin (100 µg ml⁻¹) and glucose (0.8% w/v) was used the next day to inoculate a fresh culture at a ratio of 1:10. At an OD₆₀₀ value of 0.7, protein expression was induced by IPTG (250 µM for the NHR2 protein, 100 µM for the RUNX1-NHR2 protein, and 500 µM for the RUNX1-BCR protein; Figure S2b), and the culture was incubated for 4 h at 37°C. For protein lysis, the bacterial pellet was resuspended in IMAC buffer (20 mM Tris-HCl, 500 mM NaCl, 10% glycerine, 20 mM imidazole, pH 8.0) in the presence of a protease inhibitor cocktail (P8849, Sigma-Aldrich). Cell lysis was performed by the addition of lysozyme (1 mg ml⁻¹) and subsequent sonication. For protein purification, the HisTrap HP column (GE Healthcare) was used.

Cross-Linking Assay

Purified NHR2 protein (5 µM) was incubated with the peptides at 4°C for 1 h. The BS³ (Bis[sulfosuccinimidyl] suberate) crosslink-reaction was performed at a final concentration of 1 mM for 30 min at RT. Tris-HCl (0.05 M, pH 7.4) was added to the reaction mix and incubated for 10 min to stop the reaction. Protein-oligomerization was analyzed by western blotting. For protein detection, the membrane was incubated with a primary anti-myc antibody (9E10, Santa Cruz) and a secondary HRP-coupled antibody (goat anti-mouse IgG-HRP, Santa Cruz).

ELISA Assay

Wells of a 96 well plate were coated with 100 ng streptavidin (Dianova) in PBS over night at 4°C. A double-stranded biotinylated RUNX3 oligonucleotide (100 ng, R3 = 5'-AGG GCC TGG CCT TGT GGT TCT GTG GTT GAG GGA CCA GGC-3') was bound to streptavidin in PBS containing 0.1% BSA and 0.05% Tween-20 for 2 h at RT on the 96 well plate. RUNX1-containing proteins were incubated with synthetic peptides or chemical compounds in binding buffer (25 mM HEPES, pH 7.5, 50 mM KCl, 1 mM EDTA, pH 8.0, 10 mM MgCl₂, 5% glycerol, 1 mM DTT) in the presence of 1 µg salmon sperm and 1% IGEPAL CA-630 for 5 h and subsequently added to the streptavidin-bound R3 oligonucleotide for 15 min on the 96 well plate. Binding of RUNX1-containing proteins to the R3 oligonucleotide was detected with a primary anti-myc antibody (9E10, Santa Cruz) and a secondary HRP-coupled antibody (donkey anti-mouse IgG-HRP, Santa Cruz). ELISA was developed using the Sure Blue TMB Microwell Peroxidase substrate (KLP), and the reaction was stopped with sulfuric acid (1 N). Absorption at 450 nm and 650 nm was measured with the Spectra Max 340, Molecular Devices.

ABCD Assay

The sequence of the 5'-biotinylated oligonucleotide corresponding to the RUNX1 binding sequences within RUNX3 (R3) and its mutant (R3mut) used in this study were: R3 = 5' AGG GCC TGG CCT TGT GGT TCT GTG GTT GAG GGA CCA GGC; R3mut = 5' AGG GCC TGG CCT TGT TAG TCT GTT AGT GAG GGA CCA GGC. Oligonucleotides were annealed to the corresponding unmodified antisense oligonucleotide to generate the RUNX1 binding site. Purified RUNX1 proteins were pre-incubated with a

compound for 1 h at 4°C. Similarly the biotinylated double-stranded oligonucleotides were pre-incubated with streptavidin coated magnetic beads (Dynabeads M-280 Streptavidin, Invitrogen) for 1 h at 4°C. Protein/compound samples and beads with bound oligonucleotide were incubated in binding buffer (25 mM HEPES, pH 7.5, 50 mM KCl, 1 mM EDTA, pH 8.0, 10 mM MgCl₂, 5% glycerol, 1 mM DTT) in the presence of 2% IGEPAL CA-630, 1 µg of salmon sperm, and 0.1% BSA for 1½ h. After washing three times, the bound proteins were eluted from beads in SDS-PAGE sample buffer and resolved on SDS-PAGE followed by immunoblotting with a primary anti-myc (9E10, Santa Cruz) antibody and a secondary HRP-coupled antibody (goat anti-mouse IgG-HRP, Santa Cruz).

Cell Survival Assays

SKNO-1 cells were cultured in RPMI 1640 + 10% FCS and 7 ng ml⁻¹ human GM-CSF. U937 cells were cultured in RPMI 1640 + 10% FCS. Cell proliferation and viability were measured with the cell proliferation kit II (XTT, Roche Applied Science). 3000 cells per well were cultured in a 96 well plate and daily treated with compounds at different concentrations. The XTT assay was performed at days 3, 5, and 7.

Supplementary Tables

Table S1. Tested substances

Comp. number	ZINC ID ^[a]	ROCS rank ^[b]			xlogP ^[c]
		(I)	(II)	(III)	
1	00056609	6	39451	1490153	-0.50
2a	12597398	1	113	7	-0.25
2b	03888554	997	1155	169	3.14
2c	04151299	556	25988	38514	-0.60
2d	32626810	950	5147	163	2.54
2e	14190348	32	44	1	3.54
3a	02855282	896	45897	3430	2.38
3b	03002645	88	35	14	-1.00
4	10335979	587	10610	171	1.11
5.1	04680006	1925	219	217	2.55
5.2	01218040	401	175	193186	1.66
5.3	19854471	72	134741	123345	0.67
5.4	03899737	8213	676	332253	3.33
5.5	02139078	1037	411296	649	0.67
5.6	02078987	9086	532	61033	3.14
5.7	12397407	361	14296	39412	-0.89
5.8	04180643	892	1154	162725	2.10
5.9	04719088	700	196	192745	1.99
5.10	12898352	2647	634	1270	0.17
5.11	02123682	82214	63908	830	2.55
5.12	11867206	97103	873	51564	-0.15
5.13	02155007	2362	950	1076	2.43
5.14	02102851	1679	639	2745	1.12
5.15	20412010	1903	551	2804473	2.28
5.16	02123746	51104	260816	871	2.91
5.17	02120017	1312	301	570	3.41
5.18	13628649	240	14263	81	3.12
5.19	12900927	396	41192	50096	2.29
5.20	04025278	376	5331	383312	2.36
5.21	02100522	1159	922	387	2.16
5.22	00137368	104	78119	2655660	-0.21
5.23	06645563	535	26	144	-0.18
5.24	02091382	1448	794	687	1.30
5.25	00352826	813	17728	11886	1.65
5.26	11865713	572956	22172	438	2.65
5.27	00519227	127	71	64648	0.51
5.28	02087529	4708	795	688	1.51
6.1	04136906	10105	785	19454	2.06
6.2	20342417	175	50	71162	1.47
6.3	20376469	525	114503	136	1.22
7.1	00062059	368	88220	35454	0.95
7.2	00067364	271	618120	1319066	1.82
7.3	00156500	270	377	952262	0.52
7.4	00332363	862	73747	932106	0.52
7.5	00393862	237	44878	426047	2.36
7.6	01559245	543	1091808	1174573	2.64
7.7	01574263	67	108	357158	0.96
7.8	01581098	479	318	114538	-2.34
7.9	01586785	1387	442	298607	1.71
7.10	01591215	215	626887	911272	1.59
7.11	01598799	1740	13422	689	-0.94
7.12	01617871	20264	245	277245	0.53
7.13	01619030	9571	1339	667	3.19
7.14	01627262	317	1091628	1849035	1.29
7.15	01642167	7225	488	45395	1.61
7.16	01664780	462	117581	9093	3.96
7.17	01668736	51120	268	41958	2.67
7.18	01668741	660	166	33109	3.15
7.19	01671398	542	31018	1489609	0.67
7.20	01685026	440	1091516	1339674	1.81
7.21	01687659	346	99328	1318493	-3.61
7.22	01687668	335	33688	357125	1.81
7.23	01708620	365	859	201448	1.12
7.24	01715996	575	77937	730729	-0.11
7.25	01721300	34	482404	136750	3.15
7.26	01727327	703	10714	1535586	1.26
7.27	01742186	3855811	158829	967	0.58
7.28	01868549	3235	191	411438	2.73
7.29	03306723	30	637	454267	0.41
7.30	03589671	459	82433	569898	3.72
7.31	03894527	267	11551	931038	0.10
7.32	04353551	441668	320454	266	2.52
7.33	04964947	256	420223	218949	0.52
7.34	05640994	453	204744	308847	-0.77
7.35	05742396	476	196665	62575	1.89
7.36	06272493	5626	780	4150	0.45
7.37	06576285	265933	556	15064	-0.18
7.38	08682289	888	1082430	2526491	0.22
7.39 ^[d]	13520633	10511	334	600302	-2.27
7.40 ^[d]	16951989	16	3	263069	-2.27
8.1 ^[e]	02152406	- ^[f]	- ^[f]	- ^[f]	1.21
8.2 ^[e]	02152499	- ^[f]	- ^[f]	- ^[f]	2.37
8.3 ^[e]	03863604	- ^[f]	- ^[f]	- ^[f]	0.97
8.4 ^[e]	02100423	- ^[f]	- ^[f]	- ^[f]	3.46
8.5 ^[e]	02104070	- ^[f]	- ^[f]	- ^[f]	2.95
8.6 ^[e]	02112529	- ^[f]	- ^[f]	- ^[f]	3.68
8.7 ^[e]	02113251	- ^[f]	- ^[f]	- ^[f]	3.03
8.8 ^[e]	02120756	- ^[f]	- ^[f]	- ^[f]	2.01
8.9 ^[e]	02124731	- ^[f]	- ^[f]	- ^[f]	3.30
8.10 ^[e]	06624199	- ^[f]	- ^[f]	- ^[f]	2.15
7.41 ^[e]	01667072	- ^[f]	- ^[f]	- ^[f]	1.62
7.42 ^[e]	01583437	- ^[f]	- ^[f]	- ^[f]	2.58
7.43 ^[e]	06411667	- ^[f]	- ^[f]	- ^[f]	2.58
7.44 ^[e]	06513489	- ^[f]	- ^[f]	- ^[f]	2.83

[a] The ZINC ID of the stereoisomer according to which the compound was selected for experimental testing. The actual configuration of the compound samples was not redetermined by us. 11 compounds tested in more detail are highlighted in bold. [b] Rank by ColorScore according to the three ROCS queries (I), (II), and (III) outlined in the main text and in Figure S4. [c] xlogP values taken from the ZINC 11

database.¹ [d] Diastereomers with different NCI numbers, however, with unspecified stereochemistry for the compound samples obtained from the NCI/DTP. [e] Compounds selected by fingerprint similarity search. [f] Not applicable.

Table S2. *In vitro* inhibition of NHR2 tetramerization^[a]

Comp. Number	ELISA			ABCD	
	<i>IC</i> ₅₀ [μM]	Inhibition [%] ^[b]		Inhibition [%] ^[c]	
		NHR2 ^[d]	BCR ^[e]	NHR2 ^[d]	BCR ^[e]
P1	390 ± 30	42.8	9.2	N.A.	N.A.
P3	>> 2000	N.A.	N.A.	N.A.	N.A.
2b	N.A.	83.1	41.5	49.4	11.8
5.4	N.A.	89.7	35.4	63.9	-3.3
5.8	N.A.	70.3	12.0	68.2	56.1
5.9	N.A.	51.7	20.7	52.5	14.6
5.17	N.A.	41.7	15.9	89.5	95.3
5.18	N.A.	80.5	3.9	83.0	52.6
7.3	N.A.	94.4	31.5	57.7	28.5
7.4	N.A.	71.6	10.0	47.4	6.1
7.18	N.A.	47.7	-6.7	80.2	-2.2
7.38	>> 2000	4.0	N.A.	20.1	26.1
7.44	630 ± 24	69.0	24.0	N.A.	N.A.

[a] Technical details of ELISA and ABCD assays are described in Materials and Methods. [b] Relative inhibition at 1 mM inhibitor concentration. Inhibition by **2b** was measured at 50 μM concentration; inhibition by **P1** was measured at 500 μM concentration; inhibition by **7.3**, **7.4**, **7.18**, and **7.38** was measured at 2 mM concentration. [c] Relative inhibition at 1 mM inhibitor concentration. [d] Inhibition of tetramerization of the RUNX1-NHR2 protein (Figure S2b). [e] Inhibition of tetramerization of the RUNX1-BCR protein (Figure S2b). “N.A.” means data have not been measured for one of the following reasons: 1) *IC*₅₀ values were only measured for the most promising compounds. 2) For the peptide with mutated hot spots **P3** only *IC*₅₀ was measured demonstrating inactivity. 3) Compound **7.44** was not tested in the ABCD assay because it was obtained in the second round of screening (hit expansion) after the ABCD study had been completed. 4) **P1** and **P3** could not be measured in ABCD assay. 5) **7.38** was not tested for inhibition of tetramerization of the RUNX1-BCR protein because it did not inhibit tetramerization of the RUNX1-NHR2 protein.

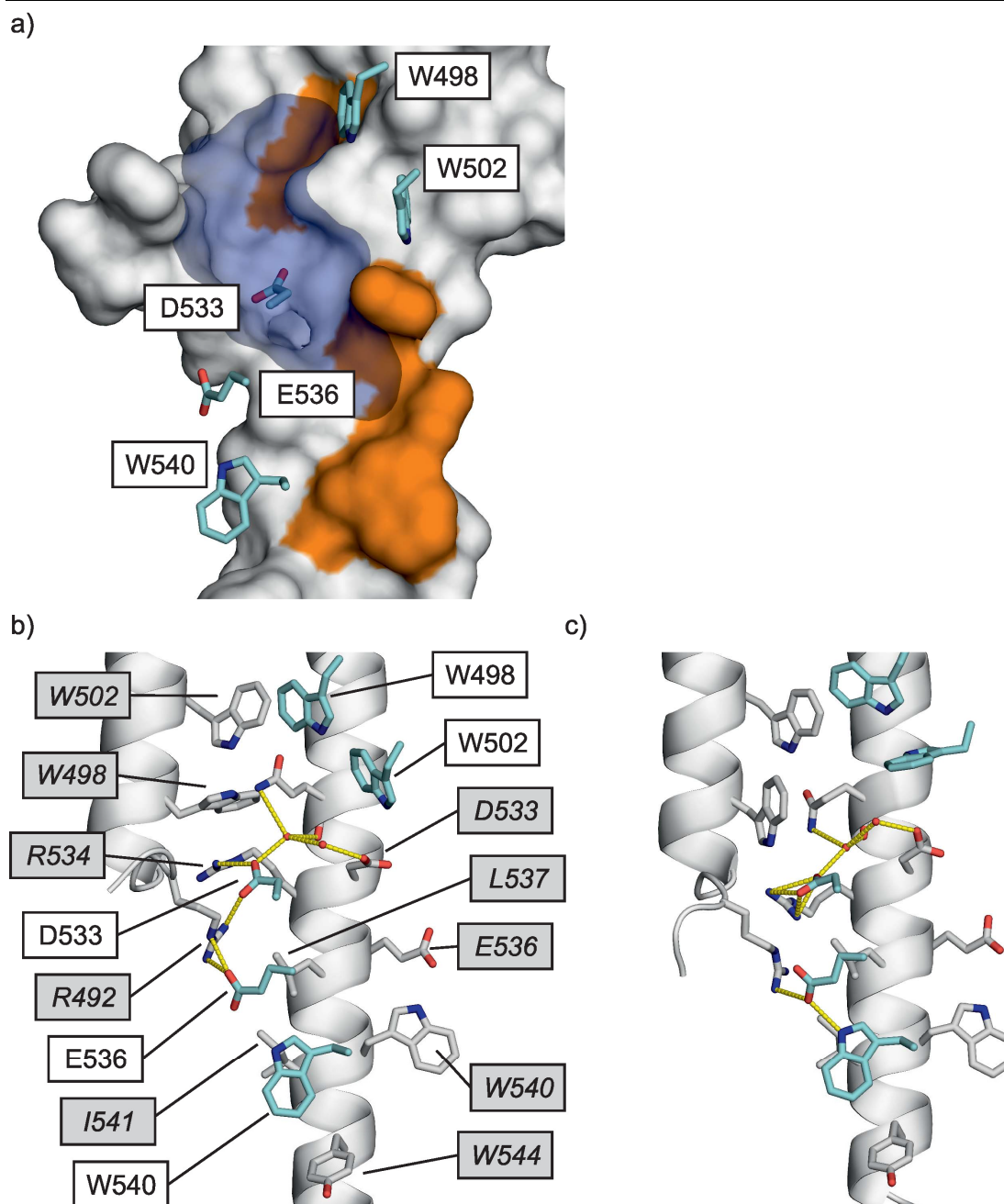


Figure S1. Hot spot containing region of the NHR2 interface. a) Depicted are the five hot spots of one dimer⁶ (cyan sticks) binding to the other NHR2 dimer (surface) in the crystallographic configuration (PDB: 1wq6⁴); the equivalent hot spots on the targeted NHR2 interface are highlighted by an orange surface. The hot spot D533 extends into the deepest pocket of the targeted NHR2 interface (transparent blue surface; identified by PocketAnalyzer¹⁴). In b) the polar (yellow dashes) and hydrophobic interactions of the three hot spots chosen as template residues for virtual screening (D533, E536, and W540) with the targeted NHR2 interface (cartoon) are shown. All five hot spots (cyan sticks), important residues on the targeted NHR2 interface (grey sticks, italic labels with grey background), and two water molecules mediating interactions between D533 and the targeted NHR2 interface are highlighted. c) Likewise, the other, almost identical binding site in the targeted NHR2 interface is shown with residue numbers omitted for clarity.

a)

NHR2	V L R R C Q E A D R E E L N Y W I R
	aa in RUNX1/ETO:	525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542	
P1	Ac-	V L R R C Q E A D R E E L N Y W I R	-NH ₂
P2	Ac-	V L R R C Q E A A R E A L N Y A I R	-NH ₂
P3	Ac-	V A R R C A E A A R E A L N Y A I R	-NH ₂
CP	Ac-	V L Q E L Q R L E S R L Q P F L Q R	-NH ₂

b)

NHR2

RUNX1-NHR2

RUNX1-NHR2-m5

RUNX1-BCR


Figure S2. Experimentally tested peptides and proteins. a) Sequences of the peptides used in NHR2 inhibition assays aligned to the wild type NHR2 sequence (alanine mutations highlighted). Corresponding sequence positions according to RUNX1/ETO numbering are denoted for the NHR2 sequence. All peptides were C-terminally acetylated and N-terminally amidated. b) Protein constructs used in ELISA and ABCD assays. The wild type NHR2 tetramerization domain or the homologous BCR tetramerization domain (highlighted) are optionally enclosed by an N-terminal RUNX1 domain and C-terminal c-myc and His6 tags. The alanine mutations in the RUNX1-NHR2-m5 protein (indicated by the “A”s above the NHR2 domain) correspond to RUNX1/ETO positions 526, 530, 533, 536, and 540, in accord with the alanine mutations of **P2** depicted in a).

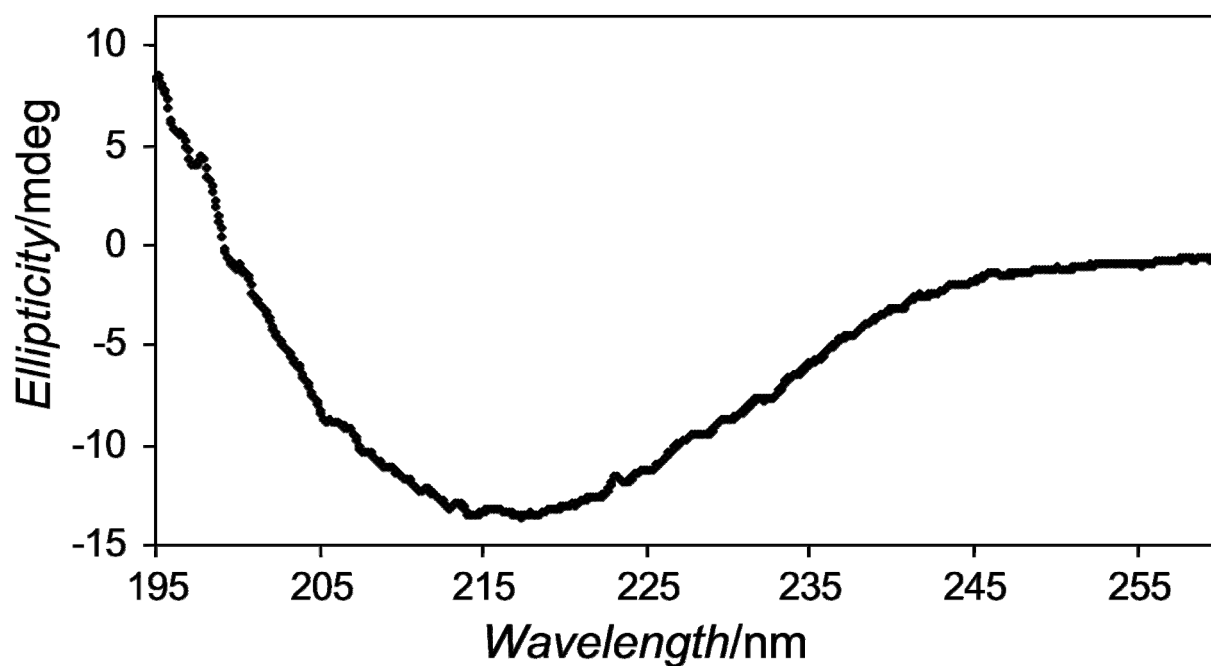


Figure S3. Helical content of the inhibitory 12mer peptide. Depicted is the CD spectrum of the 12mer peptide Ac-EADREELNYWIR-NH₂. The helical content was calculated to be $34\% \pm 4\%$ ($n = 2$) with the procedure of Rohl and Baldwin.¹⁵ For the calculation a peptide length of 13 amino acids was assumed to account for the N- and C-terminal modifications.

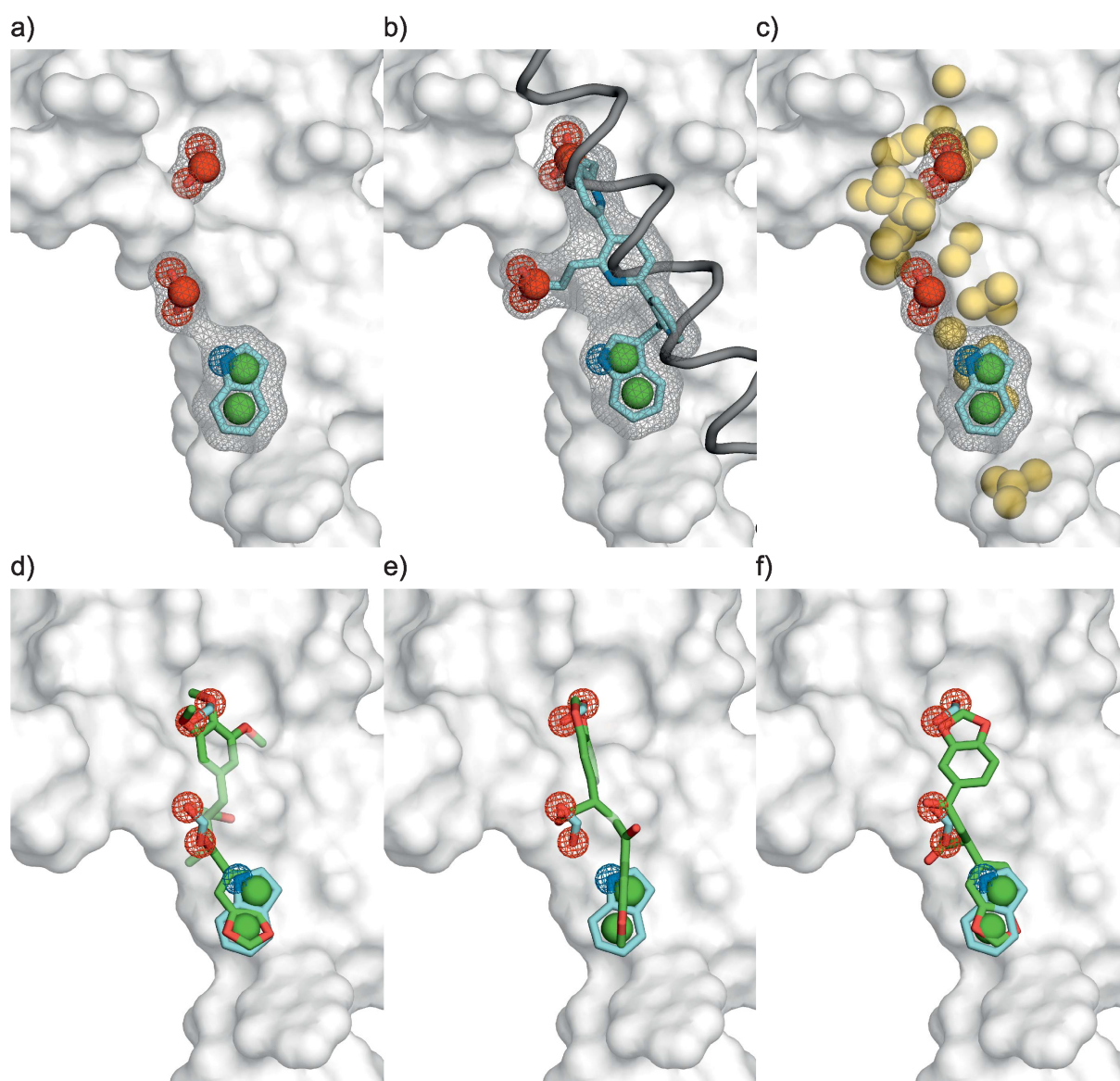


Figure S4. The three ROCS queries used for VS and exemplary ROCS poses of the most active inhibitors. a) Query (I) contains only the carboxylic and indole functional groups of D533, E536, and W540. b) Query (II) contains a terpyridine helix mimetic⁷ decorated with the functional groups from (I) and a terpyridine helix mimetic scaffold modeled in place of the helix that is to be replaced. Query (III) contains the functional groups from (I) plus repulsive Gaussian potentials centered on all adjacent protein interface atoms (≤ 4 Å distance, golden spheres) in order to penalize molecules clashing with the interface. The ROCS algorithm generates ligand poses that maximize the combined color and shape overlap (ComboScore) with these queries.² The used colors are highlighted as follows: donor (blue mesh), acceptor (red meshes), anionic (red spheres), and ring (green spheres). The functional groups of the template motif and the terphenyl scaffold (cyan sticks) are enclosed by a shape volume (grey surface mesh). The repulsive Gaussian potentials are modeled with the same weight as the conventional colors but with an opposite algebraic sign, thus repelling each color type. The targeted NHR2 interface is depicted for clarity (transparent surface) and is not part of any ROCS query. The bottom row shows ROCS poses of d) **7.18** generated with query (III), e) **7.44** generated with query (I), and f) **7.44** generated with query (III).

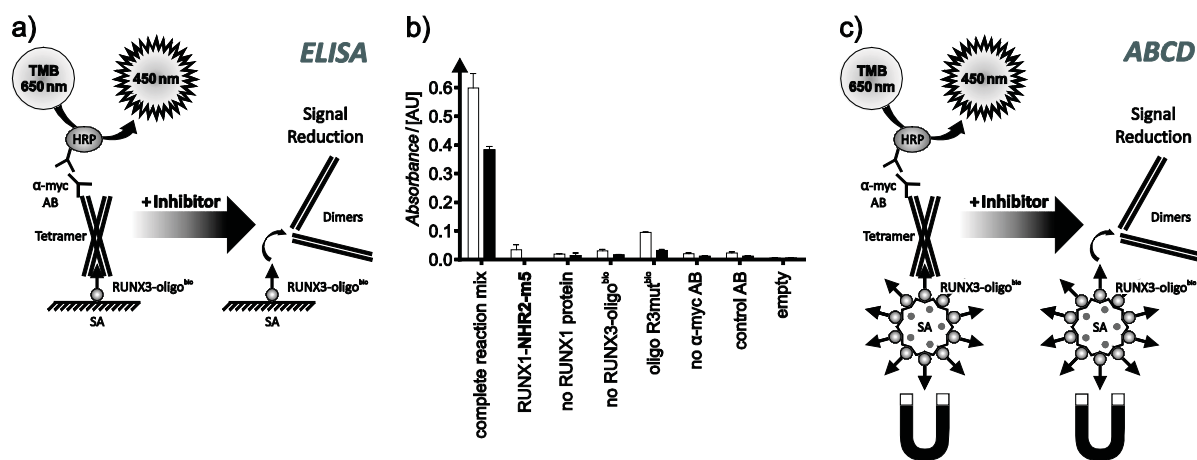


Figure S5. Biochemical assays for measuring inhibition of NHR2 tetramerization. NHR2 tetramerization and its inhibition was measured in a purpose-built sandwich ELISA (enzyme-linked immunosorbent assay). a) The RUNX1-NHR2 tetramer specifically binds to the biotinylated RUNX3-oligonucleotide (RUNX3-oligo^{bio}) that is immobilized on a streptavidin coated surface (SA). Disrupting the RUNX1-NHR2 tetramer into dimers leads to a release of the RUNX1-NHR2 protein from the surface. Antibody-conjugated horseradish peroxidase (HRP) binds to a primary anti-myc antibody (α-myc AB) that binds to the immobilized myc-labeled RUNX1-NHR2 tetramer. HRP catalyzes the conversion of 3,3',5,5'-tetramethylbenzidine (TMB) to 3,3',5,5'-tetramethylbenzidine diimine. The ELISA signal was quantified photometrically as the difference of absorbance at 650 nm and 450 nm relative to the corresponding difference of absorbance in the absence of an inhibitor (--). b) The specificity of this ELISA assay was validated by comparing signal reduction related to tetramer dependent surface binding of the RUNX1-NHR2 protein (white bars) or the RUNX1-BCR protein (black bars; both Figure S2) under the following conditions: (i) presence of the complete reaction mix but no inhibitor; (ii) presence of hot spot alanine mutations in the NHR2 domain (RUNX1-NHR2-m5; N.A. for RUNX1-BCR); (iii) absence of a RUNX1-containing protein (no RUNX1 protein); (iv) absence of RUNX3-oligo^{bio} (no RUNX3-oligo^{bio}); (v) presence of a mutated RUNX3-oligo^{bio} (oligo R3mut^{bio}); (vi) absence of the primary anti-myc antibody (no α-myc AB); (vii) replacement of the anti-myc antibody by a control antibody (control AB), or (viii) absence of all components from the reaction mix (empty). c) The inhibitory effect of selected compounds was confirmed with a complementary ABCD (avidin, biotin, complex, DNA) assay that differs from the ELISA in that the SA coated surface is replaced by SA magnetic beads for segregation of bound tetramer followed by SDS-PAGE, immunoblotting, and signal quantification consistent with the ELISA.

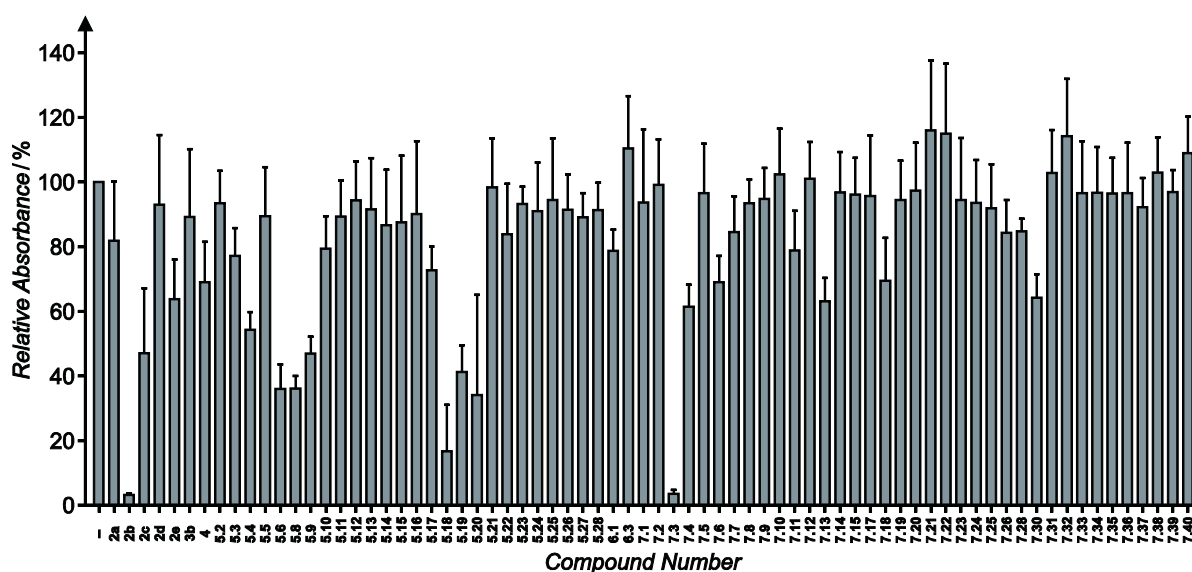


Figure S6. ELISA screening for inhibitors of RUNX1/ETO tetramerization. PPII are labeled by their compound number (Table S1). The ELISA signal was quantified photometrically as the difference of absorbance at 650 nm and 450 nm relative to the corresponding difference of absorbance in the absence (--) of a PPII. Standard deviations for $n = 6$.

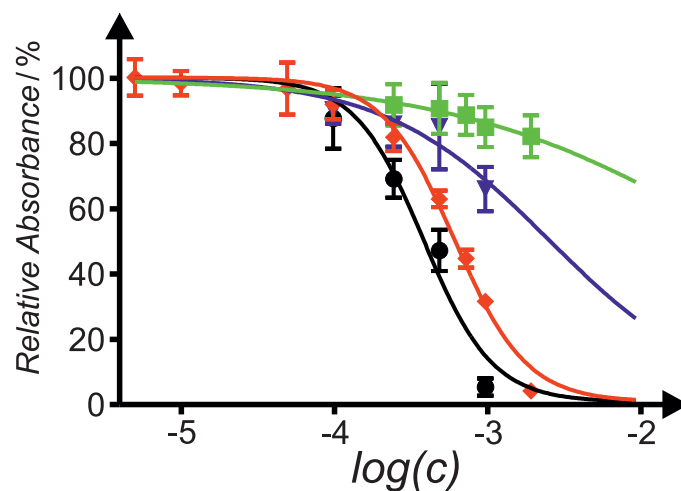


Figure S7. Dose-dependent inhibition of RUNX1-NHR2 tetramerization (fitted to a four-parameter logistics function) in the ELISA by **P1** (●; $IC_{50} = 390 \pm 30 \mu\text{M}$; Hill coefficient: 1.9 ± 0.3 (SE)) and **7.44** (◆; $IC_{50} = 630 \pm 24 \mu\text{M}$; Hill coefficient: 1.8 ± 0.1 (SE)) but neither by **P3** (▼) nor by **7.38** (■). c given in M; standard deviations for $n \geq 3$.

References in Supporting Information

1. Irwin, J. J.; Shoichet, B. K., ZINC - A free database of commercially available compounds for virtual screening. *J Chem Inf Model* **2005**, *45* (1), 177-182.
2. OEChem, version 1.7.2, OpenEye Scientific Software, Inc.: Santa Fe, NM, USA, 2009.
3. Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S., Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): Evaluation of performance. *J Chem Inf Comput Sci* **2004**, *44* (5), 1708-1718.
4. Liu, Y.; Cheney, M. D.; Gaudet, J. J.; Chruszcz, M.; Lukasik, S. M.; Sugiyama, D.; Lary, J.; Cole, J.; Dauter, Z.; Minor, W.; Speck, N. A.; Bushweller, J. H., The tetramer structure of the Nervy homology two domain, NHR2, is critical for AML1/ETO's activity. *Cancer Cell* **2006**, *9* (4), 249-260.
5. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000**, *28* (1), 235-242.
6. Wichmann, C.; Becker, Y.; Chen-Wichmann, L.; Vogel, V.; Vojtkova, A.; Herglotz, J.; Moore, S.; Koch, J.; Lausen, J.; Mantele, W.; Gohlke, H.; Grez, M., Dimer-tetramer transition controls RUNX1/ETO leukemogenic activity. *Blood* **2010**, *116* (4), 603-613.
7. Davis, J. M.; Truong, A.; Hamilton, A. D., Synthesis of a 2,3',6',3''-terpyridine scaffold as an alpha-helix mimetic. *Org Lett* **2005**, *7* (24), 5405-5408.
8. Gerber, P. R. Moloc - A Molecular Design Software Suite <http://www.moloc.ch>: 1986.
9. Gerber, P. R.; Muller, K., Mab, a Generally Applicable Molecular-Force Field for Structure Modeling in Medicinal Chemistry. *J Comput-Aided Mol Des* **1995**, *9* (3), 251-268.
10. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J Comput Chem* **2005**, *26* (16), 1668-1688.
11. Mills, J. E. J.; Dean, P. M., Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J Comput-Aided Mol Des* **1996**, *10* (6), 607-622.
12. Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S., Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier. *J Chem Inf Model* **2004**, *44* (1), 170-178.
13. Monga, M.; Sausville, E. A., Developmental Therapeutics Program at the NCI: molecular target and drug discovery process. *Leukemia* **2002**, *16* (4), 520-526.
14. Craig, I. R.; Pflieger, C.; Gohlke, H.; Essex, J. W.; Spiegel, K., Pocket-Space Maps To Identify Novel Binding-Site Conformations in Proteins. *J Chem Inf Model* **2011**, *51* (10), 2666-2679.
15. Rohl, C. A.; Baldwin, R. L., Comparison of NH exchange and circular dichroism as techniques for measuring the parameters of the helix-coil transition in peptides. *Biochemistry-Us* **1997**, *36* (28), 8435-8442.

12 Curriculum Vitae

Personal Information

Name Alexander Metz
Date of birth 28.11.1978 in Kassel, Germany

Education and Professional Experience

- 2014 – to date **Research Fellow** (Philipps-University Marburg, Germany)
Fragment-Based and Computer-Aided Drug Discovery
- 2009 – 2014 **PhD Student** (Heinrich-Heine-University Düsseldorf, Germany)
Topic: Predicting and Exploiting the Determinants of Protein/Protein Interactions to Identify Low-Molecular Inhibitors of RUNX1-ETO Tetramerization
Advisor: Prof. Dr. H. Gohlke
- 2008 – 2009 **Research Assistant** (Christian-Albrechts-University Kiel, Germany)
- 2007 – 2008 **Research Assistant** (Goethe-University Frankfurt, Germany)
- 2007 – 2007 **Research Project at Sanofi-Aventis** (Frankfurt, Germany)
Topic: MD Simulation and Free-Energy Calculations Regarding Ligand Binding at the Protein-Protein Interface of Interleukin-2
- 1999 – 2006 **Diploma in Biochemistry** (Goethe-University Frankfurt, Germany)
Topic: Developing a Hybrid Solvation Model for MM-PB/SA calculations
Advisors: Prof. Dr. H. Gohlke and Prof. Dr. C. Glaubitz
- 1998 – 1999 **Basic Military Service**
- 1995 – 1998 **Abitur** (Burggymnasium Friedberg, Germany)

- Publications** Classen, T., Kullartz, I., Metz, A., Gohlke, H., Pietruszka, J., *Changing the stereoselectivity of 2-deoxyribose-6-phosphate aldolase: Learning from pyruvate-dependent aldolase*, submitted **2014**.
- Metz, A., Ciglia, E., Kröger, T., Gohlke, H., *Protein-Protein-Interaktionen nachgeahmt: Strukturbasierte Identifizierung niedermolekularer Inhibitoren*. GIT - Labor-Fachzeitschrift **2014**, 8, 33.
- Jakobi, S., Nguyen, T.X.P., Debaene, F., Metz, A., Sanglier-Cianférani, S., Reuter, K., Klebe, G., *Hot Spot Analysis to Dissect the Functional Protein-Protein Interface of a tRNA-modifying Enzyme*, *Proteins: Struct., Funct., Bioinf.* **2014**, 82, 2713.
- Metz, A., Schanda, J., Grez, M., Wichmann, C., Gohlke, H., *From determinants of RUNX1/ETO tetramerization to small-molecule protein-protein interaction inhibitors targeting acute myeloid leukemia*. *J. Chem. Inf. Model.* **2013**, 53, 2197.
- Metz, A., Pfleger, C., Kopitz, H., Pfeiffer-Marek, S., Baringhaus, K.-H., Gohlke, H., *Hot spots and transient pockets: predicting the determinants of small-molecule binding to a protein-protein interface*, *J. Chem. Inf. Model.* **2012**, 52, 120.
- Gomes, C.P., Metz, A., Gohlke, H., Göbel, M.W., *Modular solid-phase synthesis of teroxazoles as a new class of α -helix mimetics*, *Eur. J. Org. Chem.* **2012**, 2012, 3270.
- Metz, A., Ciglia, E., Gohlke, H., *Modulating protein-protein interactions: From structural determinants of binding to druggability prediction to application*, *Curr. Pharm. Des.* **2012**, 18, 4630.

Patents *Inhibitors of NHR2 and/or RUNX1/ETO-tetramerization*. EP Appl. 13165993.0, **2013**.

Awards Best poster award at Frontiers in Medicinal Chemistry - Joint German-Swiss Meeting on Medicinal Chemistry **2011**
Lecture award of the Erlangen Molecular Modeling Workshop **2009**