Computational Knowledge-Based Prediction of Protein-Protein Recognition

Inaugural-Dissertation

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

vorgelegt von

Dennis Manfred Krüger
aus Hannover

Düsseldorf, Mai 2014

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

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

Referent: Prof. Dr. Holger Gohlke
Korreferent: Prof. Dr. Martin Lercher

Datum der Disputation: 16.06.2014

Meinen Eltern

Danksagung

Die vorliegende Arbeit wurde unter Betreuung von Herrn Prof. Dr. Holger Gohlke im Arbeitskreis für Computergestützte Pharmazeutische Chemie und Molekulare Bioinformatik in der Zeit von April 2009 bis Dezember 2013 am Institut für Pharmazeutische und Medizinische Chemie der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf erstellt.

Mein Dank gilt allen, die mich bei der Anfertigung der Dissertation unterstützt haben.

Ich danke Herrn Prof. Dr. Holger Gohlke für die Möglichkeit diese Arbeit anzufertigen, sowie für die Betreuung und erstklassige Ausbildung.

Ich danke José Ignacio Garzón und Pablo Chacón vom Rocasolano Physical Chemistry Institute in Madrid (Spanien) für die erfolgreiche Kollaboration.

Ich danke Frau Gisela Jessen für die Hilfe bei der Datenaufbereitung.

Ich danke Andrea I. Ciplea für die seelische Unterstützung.

Abstract

Protein-protein complexes play key roles in all cellular signal transduction processes. Computational methods to predict the three-dimensional structures of such complexes are valuable tools in modern structural biology and drug discovery. In this work, a new approach to predict the underlying protein-protein interactions is presented. First, a distance-dependent knowledge-based scoring function was developed and adapted against experimental alanine scanning results to predict changes in the binding free energy upon alanine mutations in protein-protein interfaces. This approach was transferred to a web server to provide valuable information for guiding biological experiments and in the development of protein-protein interaction modulators. Second, the knowledge-based potentials were evaluated as a scoring and objective function for the structure prediction of bound and unbound protein-protein complexes. The results suggest that the potentials balance well several different types of interactions important for protein-protein recognition and are discussed regarding the influence of crystal packing and the type of protein-protein complex docked. Furthermore, a simple criterion is provided with which to estimate *a prior*i if unbound docking will be successful. Third, several methods were examined with their ability to consider local and global flexibility for protein-protein docking. In this regard, a normal mode-based geometric simulation method was used to sample conformational transitions that can be used as input to the docking approach. Finally, a large-scale validation study on docking small molecules into protein-protein interfaces was performed. Results obtained allow identifying those protein-protein interfaces that are amenable for molecular docking approaches. The research results presented in this work will support scientists to improve computational methods for protein complex prediction and to develop protein-protein interaction modulators.

Zusammenfassung

Proteinkomplexe spielen eine Schlüsselrolle bei den Prozessen der zellulären Signalübertragung. Computergestützte Methoden, um solche Komplexe vorherzusagen, sind wertvolle Werkzeuge in der modernen Wirkstoffentwicklung. In dieser Arbeit wird ein neuartiger Ansatz präsentiert, um die den Komplexen zugrunde liegenden Protein-Protein Interaktionen vorherzusagen. Zunächst wurde eine abstandsabhängige und wissensbasierte Bewertungsfunktion abgeleitet und an die Ergebnisse experimenteller „Alanine-Scanning" Untersuchungen adaptiert, um die Änderungen der freien Bindungsenergie bei Alanine-Mutationen in Proteinbinderegionen vorherzusagen. Dieser Ansatz wurde auf einem Webserver implementiert, um entsprechende Vorhersagen zur Unterstützung biologischer Experimente und für die Entwicklung von Protein-Interaktions-Modulatoren zur Verfügung zu stellen. Anschließend wurden die wissensbasierten Potentiale hinsichtlich ihrer Eignung als Ziel- und Bewertungsfunktion evaluiert, um Proteinkomplexe basierend auf gebundenen und ungebundenen Proteinstrukturen vorherzusagen. Die Ergebnisse dieser Vorhersagen zeigen auf, dass die Potentiale die verschiedenen Interaktionstypen, welche für die molekulare Erkennung von Proteinen von Bedeutung sind, gut abbilden. Die Ergebnisse werden weiterführend diskutiert bezüglich des Einflusses von Kristallpackungseffekten und der Art des Proteinkomplexes auf die Qualität der Vorhersage. Zudem wurde ein Kriterium entwickelt, welches erlaubt, *a prior*i abzuschätzen, ob eine Vorhersage erfolgreich sein kann. Weiterhin wurden verschiedene Verfahren hinsichtlich Ihrer Eignung untersucht, lokale und globale Flexibilität von Proteinen für Proteinkomplexvorhersagen zu berücksichtigen. Diesbezüglich wurde eine auf Normalmoden basierende, geometrische Simulationsmethode verwendet, um konformationelle Änderungen zu simulieren, welche für die Proteinstruktur der Komplexvorhersagen berücksichtigt werden können. Als letztes wurde eine groß angelegte Studie durchgeführt bezüglich der Vorhersage von Bindemodi kleiner Moleküle in Binderegionen auf der Oberfläche von Proteinen. Die dabei erzielten Ergebnisse erlauben es solche Binderegionen zu identifizieren, für welche die Methoden der Vorhersage von Bindemodi geeignet sind. Die in dieser Arbeit präsentierten Forschungsergebnisse werden Wissenschaftlern dabei unterstützen computergestützte Methoden zur Vorhersage von Proteinkomplexen zu verbessern und Protein-Interaktions-Modulatoren zu entwickeln.

Corpora non agunt nisi fixata.

Die Körper wirken nicht, wenn sie nicht gebunden sind.

Paul Ehrlich (1913)

Contents

List of Papers

1. **Krüger, D. M.**, Ignacio-Garzon, J., Chacon, P., Gohlke, H., DrugScore^{PPI} knowledge-based potentials used as scoring and objective function in protein-protein docking. PLoS ONE **2014**, 9, e89466. *
2. **Krüger, D. M.**, Rathi, P. C., Pfleger, C., Gohlke, H., CNA web server: rigidity theory-based thermal unfolding simulations of proteins for linking structure, (thermo-)stability, and function. Nucleic Acids Res. **2013**, 41, W340-8, DOI: 10.1093/nar/gkt292.
3. **Krüger, D. M.**, Jessen, G., Gohlke, H., How good are stat-of-the art docking tools in predicting ligand binding modes in protein-protein interfaces? J. Chem. Inf. Model. **2012**, 52, 2807-2811, DOI: 10.1021/ci3003599. *
4. **Krüger, D. M.**, Ahmed, A. Gohlke, H., NMSIM Web Server: Normal mode-based geometric simulations for exploring biologically relevant conformational transitions in proteins. Nucleic Acids Res. **2012**, 40, W310-6, DOI: 10.1093/nar/gks478.*
5. **Krüger, D. M.**, Bergs, J., Kazemi, S., Gohlke, H., Target flexibility in RNA-ligand docking modeled by elastic potential grids. ACS Med Chem Lett **2011**, 2, 489-493, DOI: 10.1021/ml100217h. Cover letter article in ACS Med Chem Lett 2012, 1.
6. **Krüger, D. M.**, Gohlke, H., DrugScore^{PPI} webserver: fast and accurate *in silico* alanine scanning for scoring protein-protein interactions. Nucleic Acids Res. **2010**, 38, W480-6, DOI: 10.1093/nar/gkq471. *
7. **Krüger, D. M.**, Evers, A., Comparison of structure- and ligand-based virtual screening protocols considering hit list complementarity and enrichment factors. ChemMedChem **2010**, 5, 148-158, DOI: 10.1002/cmdc.200900314. †
8. Kazemi, S., **Krüger, D. M.**, Sirockin, F., Gohlke, H., Elastic potential grids: accurate and efficient representation of intermolecular interactions for fully flexible docking. ChemMedChem **2009**, 4, 1264-1268, DOI: 10.1002/cmdc.200900146. †
9. **Krüger, D. M.**, Garzon, J.I., Chacon, P., Gohlke, H., Predicting protein-protein interactions with DrugScore^{PPI}: Docking, scoring, and *in silico* alanine scanning. in: "From Computational Biophysics to Systems Biology (CBSB11)", U.H.E. Hansmann, J. H. Meinke, S. Mohanty, W. Nadler, O. Zimmermann (eds.), NIC, IAS Series, Jülich **2011**, 8, 105-108.
10. **Krüger, D. M.**, Gohlke, H., Proteinstabilität webbasiert analysieren. Nachr. a. d. Chemie **2013**, 61, 909-911.
11. **Krüger, D. M.**, Gohlke, H., Proteindynamik webbasiert analysieren. Nachr. a. d. Chemie **2012**, 60, 1010-1012.
12. **Krüger, D. M.**, Gohlke, H., Protein-Protein-Interaktionen webbasiert analysieren. Nachr. a. d. Chemie **2011**, 59, 44-45.

* dedicated to this thesis † previous research

Abbreviations

ΔΔG	Difference in changes in Gibbs free energy
ΔG	Change in Gibbs free energy
ΔH	Change in enthalpy
ΔS	Change in entropy
3D	Three Dimensional
6D	Six Dimensional
Å	Ångström
ADRB2	Beta-2 Adrenergic Receptor
AMBER	Assisted Model Building with Energy Refinement
ASEdb	Alanine Scanning Energetics database
BCL2	B-Cell Lymphoma 2
BEI	Binding Efficiency Index
CAPRI	Critical Assessment of Predicted Interactions
CCDC	Cambridge Crystallographic Data Centre
CSD	Cambridge Structural Database
CYP450	Cytochrome P450
DHFR	Dihydrofolate Reductase
DNA	Deoxyribonucleic Acid
DOB	Degree Of Buriedness
e.g.	*exempli gratia* (latin for “for example”)
ENM	Elastic Network Model
Erα	Estrogen Receptor α
ERK	Extracellular Regulated Kinase
FFT	Fast Fourier Transform
FXa	Factor Xa
GPU	Graphics Processing Unit
HERG	Human Ether-à-go-go-Related Gene
HIV	Human Immunodeficiency Virus
i.e.	*id est* (latin for “that is”)
i_rmsd	interface root mean square deviation
JAK	Janus Kinase
l_rmsd	ligand root mean square deviation
MAPK	Mitogen-Activated Protein Kinase
MD	Molecular Dynamics

MDM2	Mouse Double Minute 2 homolog
MLR	Multiple Linear Regression
MM/GBSA	Molecular Mechanics Generalized Born Surface Area
MM/PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
MS	Molecular Surface
MW	Molecular Weight
NMR	Nuclear Magnetic Resonance
NMSim	Normal Mode-based Geometric Simulation
PDB	Protein Data Bank
PME	Particle Mesh Ewald
PPI	Protein-Protein Interactions
PPIM	Protein-Protein Interaction Modulator
QSAR	Quantitative Structure-Activity Relationship
r	Correlation coefficient
R	Gas constant
RalGDS	Ras-like Guanine Nucleotide Dissociation Stimulator
Ras	Rat sarcoma guanosine-nucleotide-binding protein
RCNMA	Rigid Cluster Normal Mode Analysis
RCSB	Research Collaboratory for Structural Bioinformatics
rmsd	root mean square deviation
rmsf	root mean square fluctuation
RNA	Ribonucleic Acid
ROG	Radius Of Gyration
SASA	Solvent Accessible Surface Area
SEI	Surface Efficiency Index
STAT	Signal Transducer and Activator of Transcription
STD	Standard Deviation
T	Temperature
vdW	van der Waals

The 20 Canonical Amino Acids

Amino Acid	Codes*	
Aliphatic (nonpolar)		
Glycin	GLY	G
Alanine	ALA	A
Valine	VAL	V
Leucine	LEU	L
Isoleucine	ILE	I
Hydroxyl-containing (polar)		
Serine	SER	S
Threonine	THR	T
Sulfur-containing ([a]polar/[b]nonpolar)		
Cysteine[a]	CYS	C
Methionine[b]	MET	M
Cyclic (nonpolar)		
Proline	PRO	P
Amide (polar)		
Glutamine	GLN	Q
Asparagine	ASN	N
Aromatic ([a]polar/[b]nonpolar)		
Phenylalanine[b]	PHE	F
Tyrosine[a]	TYR	Y
Tryptophane[b]	TRP	W
Basic (polar and charged)		
Histidine	HIS	H
Lysine	LYS	K
Arginine	ARG	R
Acidic (polar and charged)		
Glutamate	GLU	E
Aspartate	ASP	D

*Three- and one-letter abbreviation codes for amino acids adopted by the commission on Biochemical Nomenclature of the IUPAC-IUB

1. Introduction

1.1 The Nature of Proteins

Today, one of the most common research areas in many fields of natural sciences is the nature of proteins, most likely in the fields of biochemistry, biophysics, structural biology, and structural bioinformatics. Likewise, this is true for this thesis. For the understanding of subsequent chapters, knowledge about proteins is of fundamental importance. Thus, the following introduction provides a general overview about proteins, from sequence to structure to function, and their natural dynamic behaviour.

1.1.1 Structure

Proteins are biological macromolecules that are involved in nearly all cellular processes.[1] Their importance is already expressed in the term protein, which is derived from the greek word *proteios,* meaning "primary".[2] For most of the cells the amount of proteins to the oven-dry mass accounts for more than 50%, clearly pointing out their importance for a cell to exist.[3] The basis of all natural proteins is provided by 22 different L-α-amino acids whereas only 20 of them, the so-called canonical or standard amino acids, are directly encoded by the standard genetic code.[4, 5] Two non-standard amino acids, selenocysteine (Sec or U) and pyrrolysine (Pyl or O), are incorporated into proteins by particular biosynthetic mechanisms that extend the standard genetic code by substitution of the standard stop codon.[6, 7] Each protein has its own unique amino acid sequence that is specified by the nucleotide sequence of its encoding gene.[8] The cellular process of synthesizing proteins from a nucleotide sequence is known as translation.[9, 10]

The structural composition of proteins can be characterized by four distinct aspects: (1) Primary structure: the amino acid sequence that provides one or more polymers of amino acids, so-called polypeptide chains; (2) Secondary structure: regularly repeating structural elements within a polypeptide chain which are mainly stabilized by backbone-induced hydrogen bonds, namely α-helices, β-sheets, or turns (subdivided into tight turns, multiple turns, Ω-loops, and hairpins); (3) Tertiary structure: the 3D configuration of secondary structure elements defining the shape of a polypeptide chain; (4) Quaternary structure: the biological assembly arranged by several polypeptide chains, so-called subunits, defining a protein's overall conformation.[11-13] A protein conformation can consist of either one, or multiple identical (homomeric), or distinct subunits (heteromeric).[14] Protein conformations can be stabilized by disulfide bonds, which are covalent bonds that can be build up from two nearby cysteine residues within the same, or between two distinct subunits.[15] The overall formation of a protein's conformation is referred to as protein folding, a process that is either progressed unassisted, or supported by other proteins

(chaperones).[16] Based on their specific conformations, proteins are able to adopt multiple specific functions in the cell.[17] However, proteins are not entirely rigid molecules.[18] Consequently, the basic concept of molecular recognition is based on a protein's flexibility, i.e. the ability to change its conformational state.[19]

1.1.2 Flexibility

Proteins have highly diverse structures and can experience a wide range of conformational dynamics that range from almost stable folded to intrinsically disordered states.[20] Those conformational dynamics are well-grounded in the overall flexible nature of protein structures.[18] The flexibility of a protein is encoded within its amino acid sequence and thus an integral part of a protein's structure. Proteins may perform structural rearrangements, referred to as conformational changes, during binding to another macromolecule when passing into their bound state (referred to as induced fit).[21] However, even in their unbound state proteins are able to undergo structural variation when shifting between several related conformations (following a conformational selection mechanism).[22, 23] Even well folded structure elements (domains) own a certain inherent flexibility and, moreover, they are able to move relative to each other.[24] Conformational changes can range from relatively small movements of the backbone and/or side chains to large rearrangements of domains or loops, but can also involve compact folds where a protein undergoes a partial or complete change of its tertiary structure.[25] The ability of a protein to undergo conformational changes, i.e. to what extend such changes could be accomplished, is rooted in its underlying structural composition.[22] Nearly rigid, folded proteins are rather limited in plasticity in contrast to disordered proteins that remain unfolded and thus totally flexible in their free state. Accordingly, recent studies could show a significant correlation between free-state flexibility and conformational changes that occur upon binding.[25, 26]

Two paradigms about protein flexibility have been reassessed in the past decade. First, the lock-and-key fit according to Emil Fischer's hypothesis is not longer generally accepted.[27] Today, it is known that many proteins do not bind by distinct shape matching, but they are able to adapt themselves to the binding partner and vice versa.[28, 29] Second, the protein structure-function paradigm seems to be obsolete. The traditional concept that a protein's function depends on its underlying conformation is in contrast with the discovery of intrinsically disordered proteins, i.e. proteins that are fully functional but completely unfolded.[30, 31] Such proteins bind to other macromolecules following a coupled folding and binding mechanism.[28, 32]

In computational drug design, the intrinsic flexibility of proteins still makes their handling to be difficult.[33, 34] Here, much effort has been made in the last 30 years, but implementation of protein dynamics into molecular docking approaches is still one of the challenges, most notably in the

field of protein-protein complex prediction.[33, 35] Despite the difficulties, proteins are still the most important targets in drug discovery since they own a multitude of functions that are involved in almost all types of diseases. [36, 37]

1.1.3 Function

Their high structural diversity and plasticity allow proteins to adopt a vast array of functions within living organisms, thus representing the major molecular tools of cells.[38] Based on their physiological functions, proteins can be divided into several classes that will be described in brief in the following.

Structural proteins are mostly fibrous and responsible for rigidity in biological components, e.g. collagen and elastin are components of connective tissue and keratine can be found in hair and nails.[39, 40] In contrast, contractile proteins allow for elasticity of cells, such as proteins like actin and tubulin that polymerize to build up the cytoskeleton of a cell.[41] Storage proteins are used by cells as biological reserves, e.g. for metal ions or amino acids, and can be found in milk, egg white and plant seeds.[42-44] Accordingly, calsequestrin is a protein that stores calcium ions in the sarcoplasmic reticulum, and ovalbumin is a major source of amino acids for juvenile mammalians.[45, 46] Transporter proteins are responsible for moving substances within an organism.[47] Hence, hemoglobin that occurs in the blood of vertebrates carries oxygen from the lungs to other parts of the body.[48, 49] Membrane proteins are located within a cell's membrane and serve the transport of ions, ligands or other proteins through this barrier.[50, 51] Thus, they are included in all kinds of cell-signaling processes.[52] Examples are aquaporins that control the water passage of a cell, and calcium channels that are involved in the communication between neurons.[53-55] Hormonal proteins help to coordinate certain activities of a body, e.g. the glucose metabolism that is regulated by insulin.[56] Enzymes are proteins with catalytic activity to speed up chemical reactions.[57] Familiar examples are the digestive protein pepsin that breaks down proteins in food, and lactase that breaks down the sugar lactose found in milk.[58, 59] Antibodies, also known as immunoglobulins, are proteins with protective function that are utilized by the immune system in response to the presence of a substance that is foreign to the body (an antigen) such as bacteria or viruses.[60, 61]

So far, most research has been done on targeting the active sites of enzymes, receptors and ion channels with small molecule inhibitors. However, targeting protein-protein interaction sites has become a new focus in drug discovery during the last years since protein assemblies provide an important contribution to the molecular regulation of a variety of processes in the cell, concerning survival/apoptosis, proliferation, cell-cycle progression, cell shape, polarity, adhesion, migration and differentiation. Disregulation of cellular pathways is often causal of disease. For

instance, the MAP/ERK signaling pathway is involved in the development of many human diseases including Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis and various types of cancers; the JAK/STAT signaling pathway is involved in the development of human hematological malignancies as well as autoimmune and chronic inflammatory diseases; the integrin pathway is involved in the development of thrombosis, osteoporosis, tumor-induced angiogenesis, and cholestatic liver disease.[62-64] Moreover, cell signaling plays an important role in bacterial infections that can be treated by targeting specific bacterial pathways. Furthermore, human signaling pathways are often manipulated by viruses as they rely on the functions offered by cells for their propagation. For instance, the Wnt signaling pathway can become dysregulated through the actions of several oncogenic viruses like JC polyomavirus, human papillomavirus, or herpesvirus 4 and 8.[65]

The previous overview shows the variety of functions that proteins can adopt in cells confirmed by several clinically relevant examples and thus reveal why proteins are the most important targets in drug discovery.

1.2 The Determination of Protein Structures

The basis for structure-based drug design is knowledge about the 3D structures of proteins, ligands, or their complex structures. Structure determination is not a trivial problem, because molecules are too small to identify their structure just by light microscopy, even for proteins consisting of several thousands of atoms. Therefore, more expensive methods are needed to make a molecule's structure visible. The two most commonly-used methods to determine high-resolution structures from molecules or their complexes are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy.[66] X-ray crystallography is based on the diffraction of X-rays through a closely spaced grid of atoms in a crystal.[67] In contrast, NMR spectroscopy uses nuclear magnetic resonance spectra and can be applied in solution and solid state.[68] One has to note that 3D structures derived from these methods are finally models based on refined experimental data.[69, 70] The quality of those models depends on both the experimental conditions and the computational methods used.[71] Once a protein structure is determined, it is normally published in the RCSB protein data bank to make it publicly available to the research community.[72] All of the protein structures presented in this thesis were derived from the RCSB protein data bank.

Since most of the 3D structures used in this work are based on X-ray crystallography, it is important to gain insight into the accuracy of the atomic details of such structures. One important measure for the quality of an X-ray structure is the resolution reported in Ångström (Å).[73] 1 Å corresponds to 10^{-10} m. The resolution gives information about the uncertainty of the

atomic coordinates. For instance, if the resolution of an X-ray structure is 2.5 Å, the standard deviation (STD) in the atomic coordinates is about 0.4 Å, and it is about 0.1 Å, for a resolution of 1.5 Å.[74] In structure-based drug design, the resolution of a 3D structure is of fundamental importance.[75] Here, specific bond lengths play an important role to correctly identify interactions between the molecules of interest.[76] This becomes obvious considering that the average distance between oxygen and nitrogen atoms in hydrogen bonds across protein-protein interfaces is 2.93 Å ± 0.24 Å.[77] Consequently, if the atomic uncertainty exceeds 0.4 Å, consistent with a resolution lower than 2.5 Å, it will be more difficult to decide for a bond to exist. For this reason, only X-ray structures with a resolution higher than 2.5 Å were used in this work.

Beside an X-ray structure's resolution, another measure for the uncertainty of atomic coordinates is given by the B-factor (also called the Debye-Waller factor or temperature factor) reported in $Å^2$.[78, 79] While the resolution gives information about the uncertainty of atoms based on structure determination, the temperature factor measures static disorder caused by structural differences in different unit cells throughout the crystal, or dynamic disorder caused by thermal motions of atoms.[80] For instance, for a B-factor of 15 $Å^2$, the displacement of an atom from its equilibrium position is approximately 0.44 Å, and it is as much as 0.87 Å for a B-factor of 60 $Å^2$.[81, 82] Thus, B-factors below a value of 30 $Å^2$ correspond to well-defined parts of a structure, whereas B-factors higher than 60 $Å^2$ indicate disordered parts of the structure as it is often the case for highly flexible loop regions in proteins.[81, 83]

1.3 Molecular Interactions of Proteins

A major task in computational chemistry is the optimization or weakening of interactions between a ligand and another biomolecule.[84] This can be increasing the affinity of a ligand to improve binding to a receptor, or reducing the affinity of a ligand for an antitarget (e.g. HERG channel, CYP450) to prevent from undesirable side-effects.[85-87] The most important types of molecular interactions between biomolecules, this involves both protein-ligand and protein-protein interactions, are based on non-covalent (reversible) bonds.[88, 89] Those interactions include van der Waals (vdW) forces (Keesom-, Debye-, and London forces), hydrogen bonds, ionic interactions (salt bridges), π-system interactions, and metal interactions.[90] In contrast, covalent bonds form irreversible chemical bonds in between biomolecules, e.g. disulfide bonds.[90] However, covalent bonds between ligands and receptors are less desirable in drug discovery for reasons of toxicity.[91] Information about interactions occurring in protein-protein or protein-ligand complexes can be found in so-called interaction databases.[92-99]

The affinity between two biomolecules, i.e. the strength of binding of one molecule to another, is typically expressed in the Gibbs free energy of binding reported in kcal/mol (see Equation 1.1).[100-102]

$$\Delta G = \Delta H - T\Delta S \qquad \text{(Eq. 1.1)}$$

Here, ΔG is the change in Gibbs free energy, ΔH is the change in enthalpy, ΔS is the change in entropy, and T is the absolute temperature reported in Kelvin. Whereas the enthalpy implies changes in the binding affinity due to the formation or breaking of bonds described above, the entropy implies changes in the degrees of freedom for both of the binding partners and the solvent upon complex formation.[101, 103] When it comes to entropic contributions, water plays an important role.[101, 104] Since all biochemical reactions take place in aqueous solution molecules are coated by hydration shells.[101, 105] These hydration shells are (partially) destroyed during binding to another molecule thereby releasing water molecules, a process that is well-known as desolvation.[101, 106, 107] Desolvation effects always contribute significantly to the change in entropy upon binding.[101, 108] When it comes to enthalpic contributions, number and types of interactions are the crucial factors.[109] Usually, bonding forces are unevenly spread over the binding interface of a protein such that a few residues (so-called hotspots) contribute most to the free energy of binding.[1, 110, 111] Accordingly, much importance is attached to hotspots in lead finding and optimization.[112] In many chemical processes, the change in enthalpy is largely compensated by a corresponding change in entropy, a widely documented phenomenon called entropy-enthalpy compensation.[101, 113-115] This correlation between entropy and enthalpy often results in a small change of the free energy of binding, whereas changes in entropy and enthalpy can widely spread. Thus, when it comes to the prediction of binding affinities from the computational side, even a small difference in one of these terms can be crucial for the results.[116, 117]

As already mentioned, protein-ligand complexes can be built up from reversible or irreversible interactions. Similarly, protein-protein complexes are referred to be either transient (low affinity complexes) or permanent (high affinity complexes).[118] In contrast to a permanent interaction that is usually very stable, such that the protein only exists in its complexed form, proteins that form transient interactions associate and dissociate *in vivo*.[118] In this regard, protein assemblies may be differentiated whether their subunits can be observed independently *in vivo* (non-obligate), or not (obligate). Furthermore, interaction regions of these subunits may be characterized as a continuous epitope, where a continuous region of the sequence forms the interface, or a discontinuous epitope, where different regions of the sequence contribute. Note that irreversible interactions in protein-ligand complexes differ from permanent interactions in protein-protein

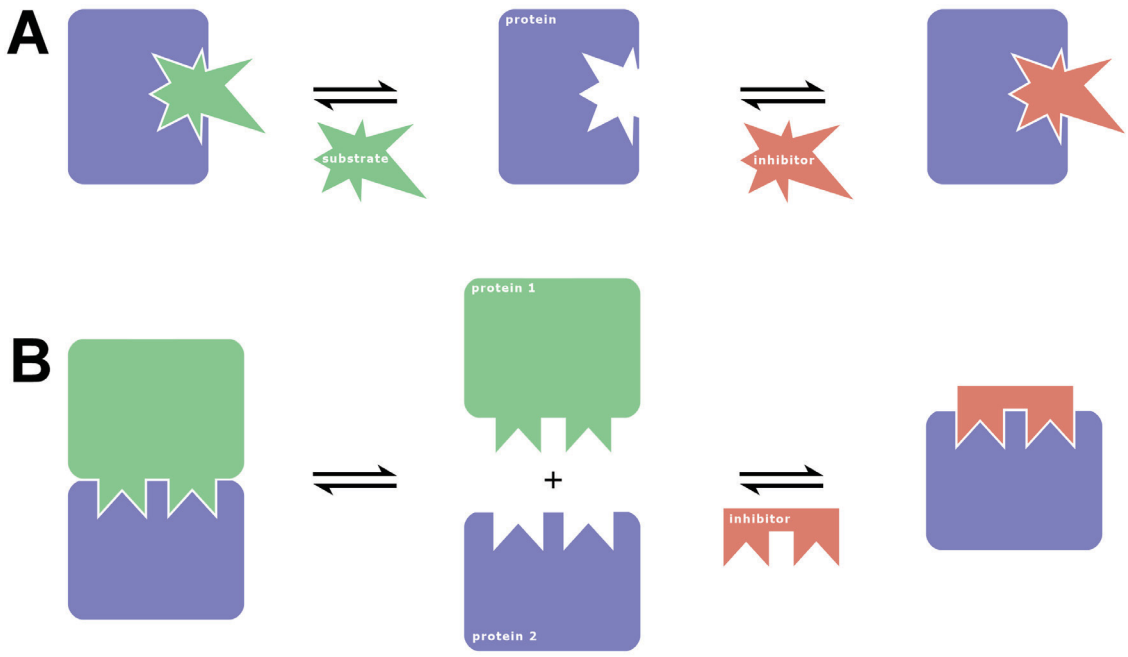


Figure 1.1: (A) Recognition and inhibition of enzymes. (B) Recognition and inhibition of protein-protein interactions.

complexes: Irreversible protein-ligand complexes are based on covalent interactions, whereas permanent protein-protein complexes are based on a multiplicity of strong non-covalent interactions.[118, 119]

In general, there are significant differences between interactions that occur in protein-protein and protein-ligand complexes.[120] Whereas a ligand can cover the whole chemical space, the formation of (human) protein-protein complexes is restricted to interactions between the 20 standard amino acids. These differences have already been investigated in several studies where occurrence frequencies of amino acids involved in interactions in protein-protein complexes were juxtaposed to those derived from protein-ligand complexes.[35, 121, 122] Differences in entropic and enthalpic contributions become clear when just considering the size of proteins and ligands: A protein owns a much higher solvent accessible surface area (SASA) and forms' plenty of interactions more compared to a small molecule.[123] Accordingly, the abundance of energetic contributions facilitates a wider scope of modulation possibilities during protein-protein binding.

Actually, protein-ligand complexes can be classified by two different types of interactions based on a ligand's location of binding to a receptor. One hast to distinguish between inhibitors that address a binding pocket of a protein, e.g. the catalytic site of an enzyme, or protein-protein interaction modulators (PPIMs) that bind to a particular position at the interface of a protein (see Figure 1.1).[124] Whereas "classical" ligands inhibit a protein by replacement of the natural substrate (another small molecule, Figure 1.1 A), PPIMs avoid binding of one protein to another

(Figure 1.1 B). Each class of inhibitors own its certain chemicophysical properties that have been reviewed in recent publications.[125-127]

Prediction of protein-protein and protein-ligand interactions just as binding affinities resulting from formation of such complexes are still major concerns in computational drug design.[128, 129] In this thesis, an approach to predict protein-protein complexes is presented (see Chapter 3) as well as an approach to predict the hotspots of such interactions (see Chapter 4). Furthermore, interactions of PPIMs to protein-protein interfaces are investigated (see Chapter 6).

1.4 Estimation of Molecular Interactions

The measurement of receptor-ligand interactions forms an important part of modern pharmaceutical development. Typically, a biological test system (assay) is used to experimentally quantify the binding affinity between a ligand and a receptor. In addition, interactions of molecules can be investigated qualitatively by x-ray crystallography or NMR spectroscopy (see Chapter 1.2). Numerous types of ligand binding assays have been developed, both radioactive and non-radioactive. An extensive overview about assay technologies and their advantages or rather disadvantages can be reviewed elsewhere.[130] The quantity to be measured by an assay is given in terms of a thermodynamically determined dissociation constant (K_d), or a kinetically determined inhibition constant (K_i) referred to enzyme-ligand complexes, reported in mol/l (M).[101, 131] The dissociation constant K_d is directly related to the Gibbs free energy of binding (see Equation 1.2).[13, 101]

$$\Delta G = \Delta H - T\Delta S = RT \ln K_d \qquad \text{(Eq. 1.2)}$$

Here, ΔG is the change in Gibbs free energy, ΔH is the change in enthalpy, ΔS is the change in entropy, T is the absolute temperature reported in Kelvin, R is the gas constant, and K_d is the dissociation constant. K_d describes the concentration of ligand, which is required to occupy half of the receptors.[13, 101] The lower the value of K_d, the lower the concentration of ligand needed, and vice versa. Thus, the lower the value of K_d, the higher is the binding affinity. Many times, instead of the inhibition constant K_d, an IC_{50} value is reported. The IC_{50} value describes the concentration of a ligand which is needed to inhibit a protein's activity by half.[132] It is commonly used to measure the potency of an antagonist. Respectively, the EC_{50} is used to measure the potency of agonistic effects. It refers to the concentration of ligand which induces a half maximal response.[132]

Furthermore, the change in binding free energy ($\Delta\Delta G$) is often calculated in pharmaceutical research (see Equation 1.2).[133, 134] For instance, it can be determined to identify residues that

contribute most to the formation of a complex (hotspots), or to separate strong from weak binding leads for further optimization of the scaffold (lead optimization).[135, 136] Hotspots can be identified by alanine scanning mutagenesis (see Chapter 3).[135] Here, the role of residues at specific positions is inferred from the change in binding free energy caused by alanine mutations. Lead optimization usually affords a series of ligands synthesized by substituting the moieties of a molecule. Those ligands are then quantified by their change in binding free energy.

If the 3D structure of a receptor exists, ligand binding modes and their affinities can be predicted by computational methods.[137-141] Such methods offer an advantageous alternative when experimental methods reach their limits. They are also useful to estimate whether a compound should be synthesized or not. A general approach to predict protein-ligand geometries and their affinities is molecular docking and scoring.[141, 142] A plenty of diverse docking tools and scoring functions have been developed in the last decade and will be introduced later on (see Chapter 1.5). While docking is used to sample possible positions of a ligand within a protein's binding site (see Chapter 1.5.1), scoring functions are used to guide the docking search and to predict binding affinities of protein-ligand complexes thereby obtained (see Chapter 1.5.3).[143, 144] Scoring functions can also be used to determine hotspots in protein-protein interfaces (see Chapter 3). Here, the change in binding free energy is estimated by computational alanine scanning mutagenesis.[135] Alanine scanning still represents a large experimental effort that cannot be applied easily and, consequently, there is a strong need for computational approaches to detect hot spots in protein-protein interfaces. Thus, a major part of this thesis was the development of a scoring function to predict protein-protein interactions to subsequently provide a basis for the implementation of a computational alanine scanning approach (see Chapter 3).

1.5 Molecular Docking

One of the widely used approaches in structure-based drug design to predict the binding mode of two biomolecules is molecular docking.[145] Docking approaches always imply a distinct type of scoring function, respectively, that is used to rank the predicted binding mode configurations by a calculated score.[146] Ideally, a score is in good correlation with the experimentally determined binding affinity.[147] However, whereas sampling of a biomolecule's conformational space is straightforward and just a question of computational resources available, binding affinity estimation based on scoring functions is still challenging.[148, 149]

In the last decade, molecular docking approaches have been adapted to a variety of diverse biomolecules, such as proteins, nucleic acids and small molecules (referred to as ligands).[150-152] In the next chapter, particular attention will be paid to protein-ligand and protein-protein docking approaches as well as their underlying scoring functions.

1.5.1 Protein-Ligand Docking

This paragraph is partially adopted from paper 3 and paper 8 (see List of Papers). Protein-ligand docking is one of the widely used approaches for structure-based lead finding and optimization in computational drug design.[145, 153] Predicted protein-ligand complex configurations are used for studying protein-ligand interactions, estimating binding affinities, and as a final filter step in virtual screening.[154] In the majority of the cases, methods on protein-ligand docking treat either both proteins and ligands as rigid molecules or allow for conformational flexibility of only the ligand, following a "rigid receptor hypothesis".[147] Furthermore, water molecules are usually not considered, even if it has been suggested in a few studies that inclusion of structural water can improve the docking accuracy.[155, 156] Docking accuracy and computational efficiency determine the scope and quality of a docking approach. The accuracy of docking approaches is normally tested by so-called redocking experiments.[157] Here, binding modes are predicted for a set of ligands and compared to the experimentally-observed (native) poses. A docking run is evaluated to be successful, if the method was able to reproduce the native pose within a certain threshold. For protein-ligand docking, typically a threshold of 2 Å root mean square deviation (rmsd) is used. The rmsd is a measure for the average deviation in the coordinates of heavy atoms between a predicted and the native ligand pose. Preserving computational efficiency is equally important, given the short timeframe usually available for a docking run. In particular, evaluating the interaction energy between protein and ligand is expensive. A widely used approach to increase the calculation speed is based on potential fields that are pre-calculated just once in the binding pocket region of the protein, by scanning interactions between the protein and ligand atom probes.[158-162] The potential field values are stored at the intersections of a regular 3D grid, providing a lookup table (see Figure 1.2). The approach is applicable to all distance-dependent pairwise interactions, such as electrostatic and van der Waals interactions and interactions described by statistical pair potentials (see Chapter 1.3).[154, 158] In subsequent docking runs, interaction energies between protein and ligand are then determined in constant time from the lookup table by means of interpolation. This provides a significant rate increase relative to individually evaluating the pair interactions. One of the most common docking softwares following such a grid-based approach is the AutoDock suite. AutoDock is the most widely used protein-ligand docking tool and has been applied in multiple studies and drug design projects since its first release in the late 1990s.[163, 164] Nowadays, three versions of AutoDock are available, version 3, 4 and AutoDock Vina, with the latter one being a new implementation where the docking procedure differs from the other versions.[165-167] While AutoDock3 and 4 use a force field-based scoring approach, AutoDock Vina uses a combination of knowledge-based potentials and empirical scoring.[165-167] The underlying sampling procedure used to explore the

conformational space of a ligand is performed by a Lamarckian genetic algorithm, respectively.[165] In previous studies, AutoDock3 was successfully adapted for use with the knowledge-based pair-potentials of DrugScore as an objective function.[152, 154] Here, the AutoDock3 results for identifying good binding geometries could be improved significantly when using DrugScore as an objective function.[154] AutoDock3 was also successfully used to predict binding modes of PPIMs in protein-protein interfaces (see Chapter 6).

For several pharmacologically important proteins, such as HIV-1 protease, aldose reductase, FK506 binding protein, renin, and dihydrofolate reductase (DHFR), pronounced plasticity upon ligand binding has been observed.[168-176] Protein plasticity comprises a range of possible movements, from single side chains to drastic structural rearrangements (see Chapter 1.1.2).[177] Not surprisingly, if docking is performed with the assumption of a rigid active site in those cases, a dramatic decrease in docking accuracy is observed.[178, 179] The drop in docking accuracy was found to be mirrored by the degree to which the protein moves upon ligand binding so that docking to an unbound form, so-called apo docking, usually shows the largest deterioration.[146, 178, 180] This clearly highlights the importance of developing strategies for taking protein plasticity into account in addition to the conformational flexibility of the ligand to prevent mis-dockings of ligands to flexible proteins. At present, three major routes to include protein plasticity during docking can be identified. The classification correlates with various types of protein movements observed upon ligand binding. First, plasticity is considered implicitly following a soft-docking strategy with attenuated repulsive forces between protein and ligand.[181, 182] While this is simple to implement and does not compromise docking efficiency, the range of

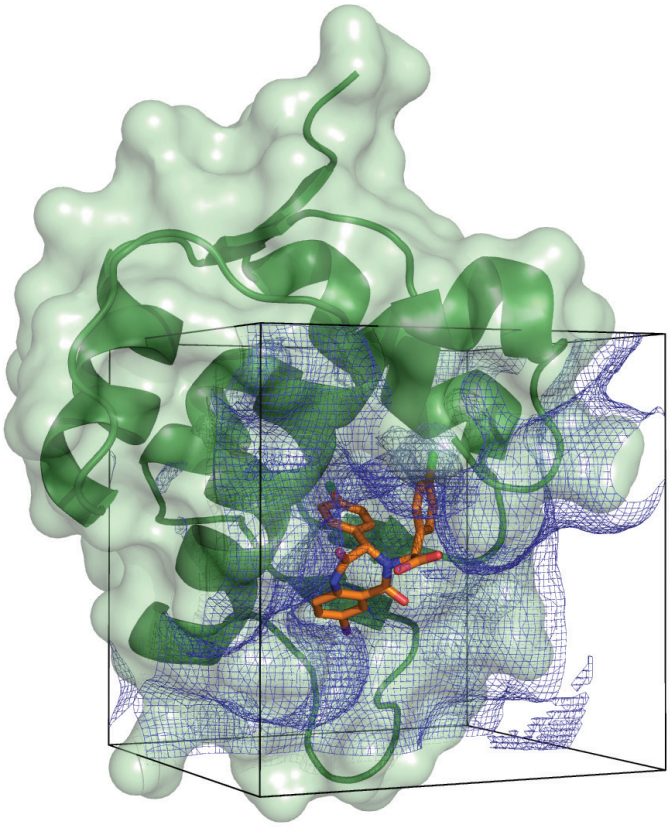


Figure 1.2: Regular 3D grid (blue) in the binding pocket region of human MDM2 (green) built from PDB code 1T4E that is used to store pre-calculated potential field values derived from interactions between aromatic carbons of a ligand and the protein. Benzodiazepine is shown as the ligand (orange).

possible movements that can be covered is rather limited. Second, only side chain conformational changes in the binding pocket are modeled.[183-185] These approaches assume that the protein has a rigid backbone structure, thus neglecting critical backbone shifts responsible for mis-docking of ligands.[179] Third, large-scale conformational changes including backbone motions are taken into account. There are several types of approaches in this category: perform parallel docking into multiple protein conformations; structurally combine multiple conformations; model protein motions in reduced coordinates; apply molecular dynamics (MD) or Monte Carlo based sampling to either generate protein-ligand configurations or to optimize pre-computed configurations; reproduce protein motions by elastic potential grids.[141, 186-194]

During the past 30 years, a plethora of protein-ligand docking tools have been developed, mostly aiming at predicting poses of ligands binding to "classical" targets, such as enzymes or receptors.[147] In contrast, much less effort has been devoted to predicting conformations of ligands that bind to protein-protein interfaces, so-called PPIMs. Protein-protein interfaces provide an important new class of drug targets because protein-protein interactions are involved in nearly all biological processes.[1] Up to date, there is still a lack of large-scale validation studies on docking into protein-protein interfaces, despite the fact that protein-protein interfaces provide major challenges for structure-based ligand design approaches, for at least two reasons:[112, 195] First, in contrast to "classical" targets, protein-protein interfaces are rather flat and usually lack a distinct binding pocket.[196] Second, due to the often large size of protein-protein interfaces (~1200 to ~4660 Å^2), interactions that are favorable for binding can be widely distributed over the interface.[197] Hence, it has remained elusive so far whether state-of-the-art docking tools are generally applicable for protein-protein interfaces. This provided the incentive for us to assess the predictive power of commonly used protein-ligand docking approaches with respect to pose prediction in protein-protein interfaces (see Chapter 6).

1.5.2 Protein-Protein Docking

Knowledge about binding configurations between proteins is used for hotspot identification, virtual screening, and classification of protein-protein interfaces.[112] Today, available 3D structures of protein-protein complexes just represent a small percentage of protein-protein interactions that are estimated to occur in living cells.[125] The size of the human interactome is estimated to involve approximately 650000 interactions but current databases of protein-protein interactions only report on approximately 1200 available 3D structures of human protein complexes.[198, 199] Thus, there is a need for computational methods to predict such interactions. Similar to protein-ligand docking, there are two main aspects in protein-protein docking: First, sampling the conformational space of the two binding partners, and second, assessing each predicted complex

with a scoring function.[200] However, the underlying search algorithms completely differ. Whereas the conformational space of a small molecule (and a rigid receptor) can be explored easily in reasonable time, the problem becomes more complex for proteins that can consist of several thousands of atoms. Therefore, protein-protein docking starts from a more simplified approach treating both binding partners as rigid molecules.[200] Despite this limitation, a complete sampling of the 6D search space (3D in translation and rotation, respectively) is computationally demanding. In the past decade, various protein-protein docking algorithms have been developed following such a rigid-body docking approach to predict the 3D structures of protein-protein complexes.[201] Based on their search algorithms, these approaches can be classified into three general categories: global search, local search, and randomized search methods.[202] Global search of all possible binding configurations can either be performed in real, or in transformed space using Fast Fourier Transform (FFT) algorithms.[203-209] Local search algorithms, involving distance geometry algorithms or geometric hashing, are based on matching local shape features.[210-213] Randomized search can be applied by genetic algorithms, or Monte Carlo methods.[214-219] An overview about currently available protein-protein docking tools and the underlying search algorithms can be reviewed elsewhere.[220] Most popular methods are based on FFT as they have been proven to show the best performance.[221] The FFT based approach goes back to the early work of Katchalski-Katzir *et al.* (1992).[205] Here, protein and ligand (the smaller protein) are projected onto regular 3D grids that carry information of the shape in terms of discrete functions. The number of grid cells depends on both the size of the protein structure and the grid spacing as a measure of concinnity of the grid, or rather, the accuracy of the shape representation. The translational space is then efficiently sampled using FFT to calculate the correlation between the two pre-calculated grids.[205] The sampling is completed by an implicit orientational search. Critical parameters for computational runtime and docking success are translational and orientational step size, typically 1-2 Å and around 6°.[150] A direct calculation of the correlation would be very costly for a complete sampling of the whole 6D search space, scaling at $O(N^6)$, where N is the number of values of the 3D correlation function.[205] However, due to the FFT approach, computational costs can be reduced to $O(N^3 \ln N)$.[205] In principle, such a geometric shape matching follows a soft docking approach. The match between the shapes of the proteins, or in other words, the correlation of the discrete functions representing these shapes, is not perfect. Protein-protein complexes always reveal gaps in between the molecules that, for instance, result from hydrogen atoms or water molecules which can be present in the interface.[205] To tolerate such imperfections, the algorithm is allowed to penetrate the surface due to a given threshold defining the surface thickness.[205] The surface thickness is another critical parameter, since increasing its value facilitates the chance of producing more false positive

predictions, whereas decreasing its value facilitates the chance of rejecting positive predictions. In newer protein-protein docking approaches the initial procedure has been enhanced by an extension of the correlation functions using several additive terms that not only consider geometric shape matching, but also desolvation, van der Waals forces, electrostatics, or statistical pair-potentials (see Chapter 4).[150] In addition to the measure of surface fit, those terms are used to calculate a docking score for each of the solutions and rank the predicted complexes accordingly (see Chapter 4). Instead of using FFT to accelerate sampling of the translational space, it can also be used to accelerate sampling of the rotational space. In this regard, some approaches make use of spherical harmonic functions to represent the surface shape and the interface properties of the interacting proteins.[150, 201, 222] Here, FFT is applied to calculate the correlation between spherical harmonic functions followed by an implicit translational search.[150]

Proteins are flexible biomolecules that are able to accomplish a wide range of possible movements (see Chapter 1.1.2). Consequently, rigid-body docking methods can only be successful in those cases where movements of proteins are negligibly small. Taking protein movements into account is still one of the major concerns in the field of protein-protein docking.[220] So far, several approaches have been attempted to include protein flexibility in docking approaches. A typical way to include protein flexibility is refinement of pre-calculated rigid-body docking solutions (see Chapter 5). Small conformational changes can be taken into account by side chain rotamer sampling or energy minimization.[193, 216, 218, 219] Furthermore, a few methods are able to treat medium/large conformational changes during docking. To account for global flexibility, a set of low-frequency normal modes, pre-calculated from the unbound protein, is used to guide a minimization procedure during docking.[217] This approach can be combined with side chain rotamer optimization to also account for local flexibility.[193] Despite much progress that has been made the last years, success rates for predicting protein-protein complexes where at least one of the binding partners shows large conformational changes are still disappointing.[223] In cases where the induced fit involves large conformational rearrangements, like domain movements, loop rearrangements, compact folds, or disordered to structured transitions, the problem is beyond current protein-protein docking methods. Even unrestricted MD simulations are not of practical use to describe long-range folding and binding mechanisms that occur on a microsecond timescale.[224, 225] Because there is still a need for such methods, we tried to tackle local flexibility using several distinct refinement procedures and global flexibility following a conformational selection approach (see Chapter 5).[27, 226]

The accuracy of protein-protein docking approaches is usually tested by redocking experiments (see Chapter 4), where protein structures are used in their bound conformations to predict the native protein-protein complex (bound-bound docking).[227] The same procedure is performed for

unbound protein structures, but unbound and native receptors are superimposed and the corresponding coordinate transformation is applied to superimpose the predicted and the native complex (unbound-unbound docking) to access the accuracy later on.[227] The superposition is required here to determine the correct position of the unbound protein structures in the native complex. Predicted complexes are evaluated with respect to their deviation in atomic coordinates from the native complex based on special assessment criteria that were defined by the authors of Critical Assessment of Predicted Interactions (CAPRI).[227] CAPRI is a worldwide experiment to

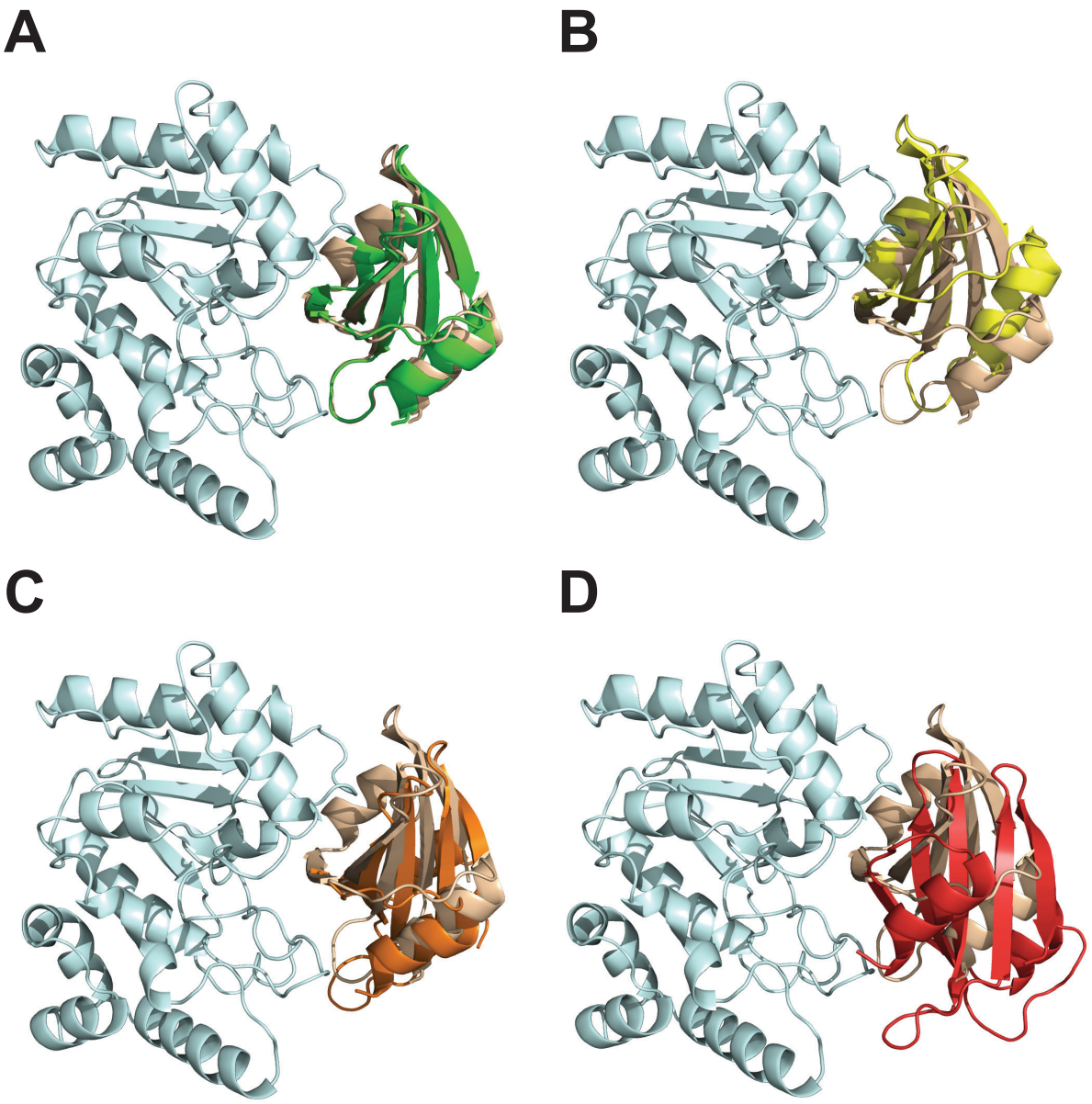


Figure 1.3: Various protein-protein docking predictions for the Uracil-DNA Glycosylase inhibitor protein (native position colored in wheat, predicted positions colored in green, yellow, orange and red) binding to Uracil-DNA Glycosylase (palecyan) created from PDB code 1UDI: (A) High quality prediction (l_rmsd = 0.8 Å, i_rmsd = 1.6 Å, f_{nat} = 0.83, f_{not} = 0.06); (B) Medium quality prediction (l_rmsd = 4.6 Å, i_rmsd = 4.1 Å, f_{nat} = 0.7, f_{not} = 0.33); (C) Acceptable quality prediction (l_rmsd = 7.6 Å, i_rmsd = 7.8 Å, f_{nat} = 0.3, f_{not} = 0.63); (D) Incorrect prediction (l_rmsd = 15.5 Å, i_rmsd = 14.9 Å, f_{nat} = 0.14, f_{not} = 0.79).

evaluate blind predictions of protein-protein docking tools.[228] In this regard, l_rmsd, i_rmsd, f_{nat} and f_{not} have been established to define the quality of protein-protein docking solutions: l_rmsd is the backbone rmsd of the ligand in the predicted versus the ligand in the native complex; i_rmsd is the rmsd of the backbone of the interface residues (defined by all residues that have at least one atom within 10 Å distance around the other protein) in the predicted versus the native complex; f_{nat} is the fraction of native residue-residue contacts defined by the number of native residue-residue contacts in the predicted complex divided by the number of contacts in the native complex; f_{not} is the fraction of non-native residue-residue contacts defined by the number of non-native residue-residue contacts in the predicted complex divided by the total number of contacts in the predicted complex.[227] Following these definitions, predicted complexes are divided into four classes, each describing the quality of the solution: high, medium, acceptable, and incorrect (see Figure 1.3).[227]

1.5.3 Scoring Functions

In molecular docking, scoring functions are used to assess the predicted docking poses by a calculated score such that the solution with the highest/lowest score is ideally identified as closest to the native one. Scoring represents the most challenging part in molecular docking.[229] Whereas sampling of a near-native conformation is successful in most of the cases, selection of the correct pose and estimation of binding affinities is still not satisfying.[229, 230] Consequently, there is still a need to develop new scoring functions to overcome these problems. This provided the incentive for the development of DrugScore^{PPI}, a knowledge-based scoring function to predict protein-protein interactions, described in this thesis (see Chapter 3). Today, a multitude of different scoring functions is available that account for different types of biomolecular interactions, such as protein-DNA, protein-RNA, protein-protein and protein-ligand interactions.[135, 151, 152, 230, 231] However, scoring functions were not only developed to predict interactions of particular types of biomolecules and their ligands, but also to predict target-specific interactions.[232] In this regard, several scoring functions were tailored to predict interactions of specific proteins (e.g. thrombin, ADRB2, ERα, FXa, kinases), or specific ligand molecules (e.g. carbohydrates, peptide ligands).[232-240] In general, all scoring functions can be roughly classified into three main categories: force-field based, empirical, and knowledge-based.[241, 242] Scoring functions for protein-protein complexes can be further classified into residue-level potentials and atomic potentials.[243-245] Residue-level (coarse-grained) potentials are computationally advantageous especially when applied to predict protein-protein complexes where the binding partners can undergo large conformational changes.[246-248] In contrast, atomic potentials are of higher resolution and are supposed to be most accurate and specific.[249] Atomic potentials are often knowledge-based; the reduced steepness of knowledge-

based potentials compared to force field-based or empirical scoring functions has been recognized as an advantage in docking.[146]

Force-field based scoring functions calculate docking energies from direct non-covalent interactions, such as van der Waals and electrostatic energies, using molecular mechanics force fields.[165] They are often augmented by physics-based terms derived from the molecular mechanics Poisson-Boltzmann or Generalized Born surface area (MM/PBSA or MM/GBSA) approaches, in order to also consider desolvation energies.[250] For instance, the inherent scoring function of AutoDock (see Chapter 1.5) is based on terms that are obtained from the AMBER force field (see Equation 1.3).[165]

$$\Delta G_{bind} = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{tor} + \Delta G_{sol} \qquad \text{(Eq. 1.3)}$$

Here, ΔG_{bind} is the estimated free energy of binding, ΔG_{vdw} is a van der Waals term, ΔG_{hbond} is a hydrogen bonding term, ΔG_{elec} is a Coulomb electrostatic potential, ΔG_{tor} is the change in torsional free energy, and ΔG_{sol} accounts for the desolvation energy of the ligand.[165] All terms were tailored to yield the best correlation to experimentally determined affinity data.

Empirical scoring functions decompose the overall binding free energy into several individual energetic contributions, such as conformational entropy, hydrogen bonds, metal interactions, or hydrophobic and hydrophilic surface areas, whereas coefficients of all these contributions are fitted to experimentally determined binding affinities by multiple linear regression (MLR), thus following a quantitative structure-activity relationship (QSAR) model.[251, 252] One of the most popular empirical scoring functions is ChemScore (see Equation 1.4).[253]

$$\Delta G_{bind} = \Delta G_{hbond} + \Delta G_{metal} + \Delta G_{lipo} + \Delta G_{rot} + \Delta G_{0} \qquad \text{(Eq. 1.4)}$$

Here, ΔG_{bind} is the estimated free energy of binding that is calculated by contributions from hydrogen bonds (ΔG_{hbond}), metal interactions (ΔG_{metal}), hydrophobic effects (ΔG_{lipo}), rotatable bonds of the ligand (ΔG_{rotor}), and a regression constant ΔG_{0}.[253] Each of the terms consists of a particular physical contribution to the binding free energy and a scale factor determined by MLR analysis. The final score is determined by adding in several clash penalty and internal torsion terms.[253]

Knowledge-based scoring functions are based on statistical occurrence frequencies of pairwise atomic interactions between the biomolecules of interest and are derived from known structures of complexes in a database.[152] The overall score is calculated as the sum of distance-dependent statistical potentials of all atom types within a defined distance cutoff. One of the most robust

knowledge-based scoring functions is DrugScore, following an inverse Boltzmann and modified Sippl approach (see Equation 1.5).[152, 254]

$$\Delta W = \sum_{l \in L} \sum_{n \in N} \Delta W_{\mathrm{T}(l),\mathrm{T}(n)}(d_{l,n}) \qquad \text{(Eq. 1.5)}$$

Initially developed to evaluate protein-ligand complexes, it was also adapted to protein-RNA complexes (DrugScore^{RNA}) in recent work, and to protein-protein complexes (DrugScore^{PPI}) as a part of this thesis (see Chapter 3 & 4).[135, 151] Here, ΔW is the docking score for a complex of a molecule L and a partner molecule N calculated as the sum of all occurring atom-atom interactions, whereas the term $\Delta W_{\mathrm{T}(l),\mathrm{T}(n)}(d_{l,n})$ defines the score for a specific interaction between atom l of type T(l) from molecule L and atom n of type T(n) from partner molecule N separated by the distance $d_{l,n}$.[152] For both the molecules a set of 17 atom types is defined based on the Sybyl atom type notation.[152] Crucial for the predictive power of knowledge-based scoring functions is choosing of a proper reference state. In the case of DrugScore, the reference state mimics a compact complex configuration with non-specific interactions. Thus, the reference state removes contacts from the distributions where interactions between the atoms are zero such that net potentials representing only specific interactions are obtained.[152]

The predictability of scoring functions is usually evaluated based on three criteria: (1) The ability to identify near-native predictions by ranking those poses on the top of a hit list; (2) The ability to correlate docking scores, or rather the predicted binding energies, to experimentally determined binding affinities; (3) The ability to construct funnel-shaped landscapes for the energy surfaces of the predicted poses when plotting docking scores versus corresponding rmsd values.[229, 230] Accordingly, we evaluated DrugScore^{PPI} following these steps (see Chapter 3 & 4).

Paragraph 1.5.1 is partially reprinted (adapted) with permission from Kazemi, S., Krüger, D. M., Sirockin, F., Gohlke, H., Elastic potential grids: accurate and efficient representation of intermolecular interactions for fully flexible docking. ChemMedChem 2009, 4, 1264-1268. DOI: 10.1002/cmdc.200900146. URL : http://onlinelibrary.wiley.com/doi/10.1002/cmdc.200900146/abstract. Copyright 2009 Wiley-VCH Verlag GmbH& Co. KGaA, Weinheim.

Paragraph 1.5.1 is partially reprinted (adapted) with permission from Krüger, D. M., Jessen, G., Gohlke, H., How good are stat-of-the art docking tools in predicting ligand binding modes in protein-protein interfaces? J. Chem. Inf. Model. 2012, 52, 2807-2811. DOI: 10.1021/ci3003599. Copyright 2012 American Chemical Society.

2. Scope of the Thesis

In this thesis, protein-protein interactions shall be investigated, and the knowledge derived shall subsequently be used to develop a new approach to predict protein-protein complexes. In a first step, the DrugScore formalism from Gohlke *et al.* shall be used to derive a knowledge-based scoring function from a dataset of protein-protein complexes, namely DrugScore^{PPI}.[152] In a second step, this scoring function shall be evaluated by rescoring two non-redundant datasets for which bound and unbound protein complex predictions have been generated. The predictability shall be evaluated based on, first, the ability to identify near-native predictions by ranking those poses on the top of a hit list, and second, the ability to construct funnel-shaped landscapes for the energy surfaces of the predicted poses when plotting docking scores versus corresponding rmsd values. In a third step, DrugScore^{PPI} shall be tested in its ability to predict experimental alanine scanning results, i.e., changes in the binding free energy of protein-protein complexes upon alanine mutations in the interface, by correlating docking scores to experimentally determined changes in binding affinities.

Later on, DrugScore^{PPI} shall be applied as an objective function in combination with a protein-protein docking tool in order to predict 3D structures of protein-protein complexes. Here, parameters shall be validated to allow for an optimal interplay between the docking algorithm and the knowledge-based potentials. Afterwards, several methods shall be examined in order to incorporate local and global flexibility into the docking approach to account for protein flexibility. In this regard, normal mode-based geometric simulation methods shall be tested particularly. Furthermore, it shall be tested in how far the obtained protein-protein docking results can be improved by distinct refinement procedures.

Finally, it shall also be tested in how far the DrugScore^{PPI} potentials are amenable to predict the binding modes of small molecules in protein-protein interfaces.

3. *In Silico* Alanine Scanning for Scoring Protein-Protein Interactions (Paper I)

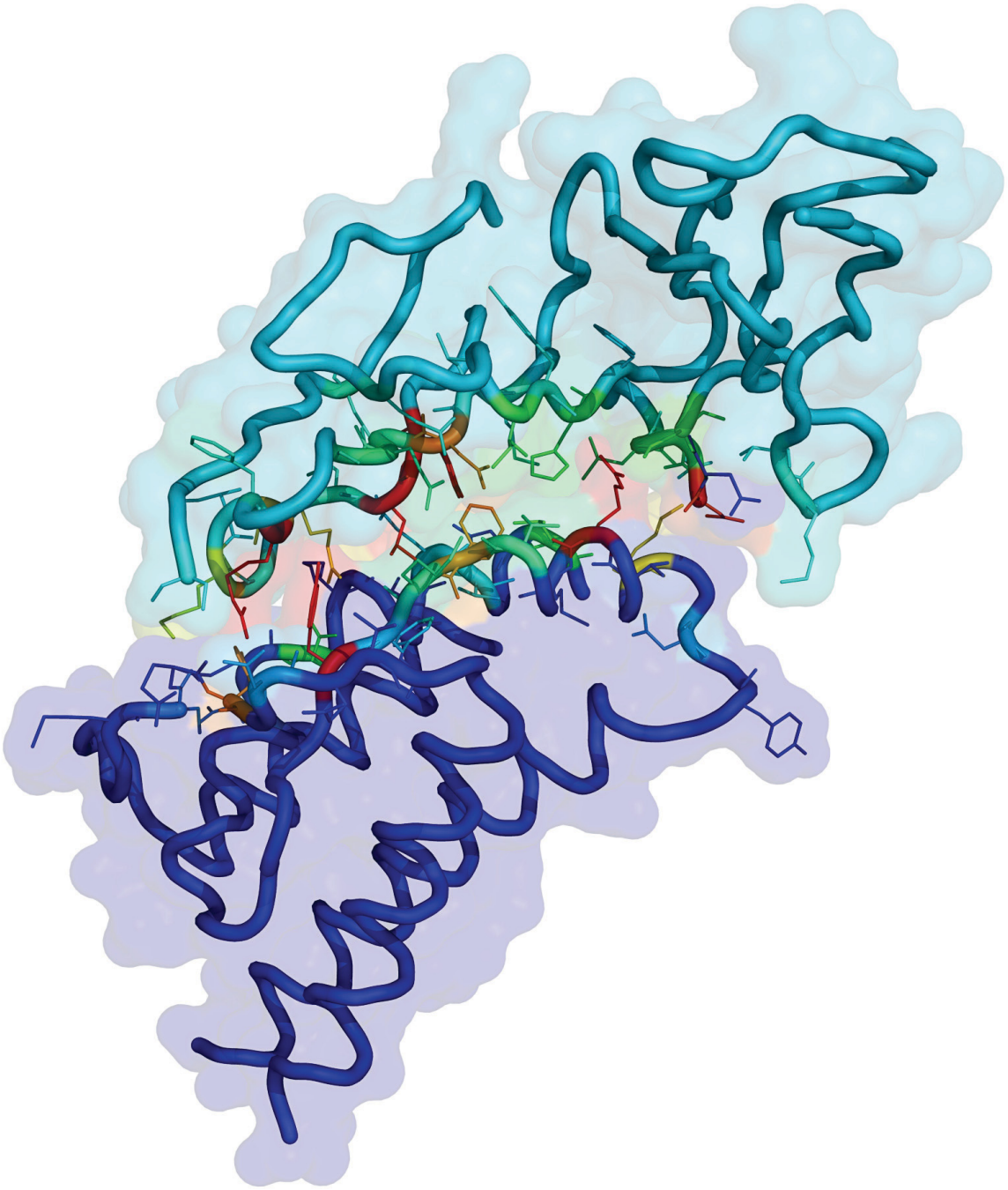
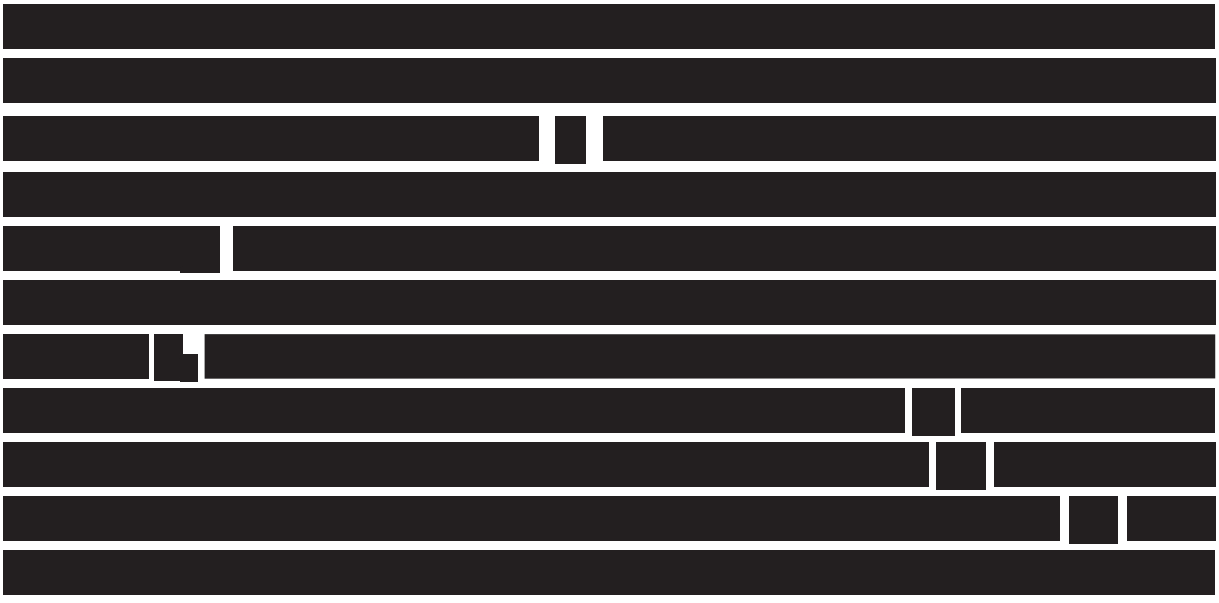
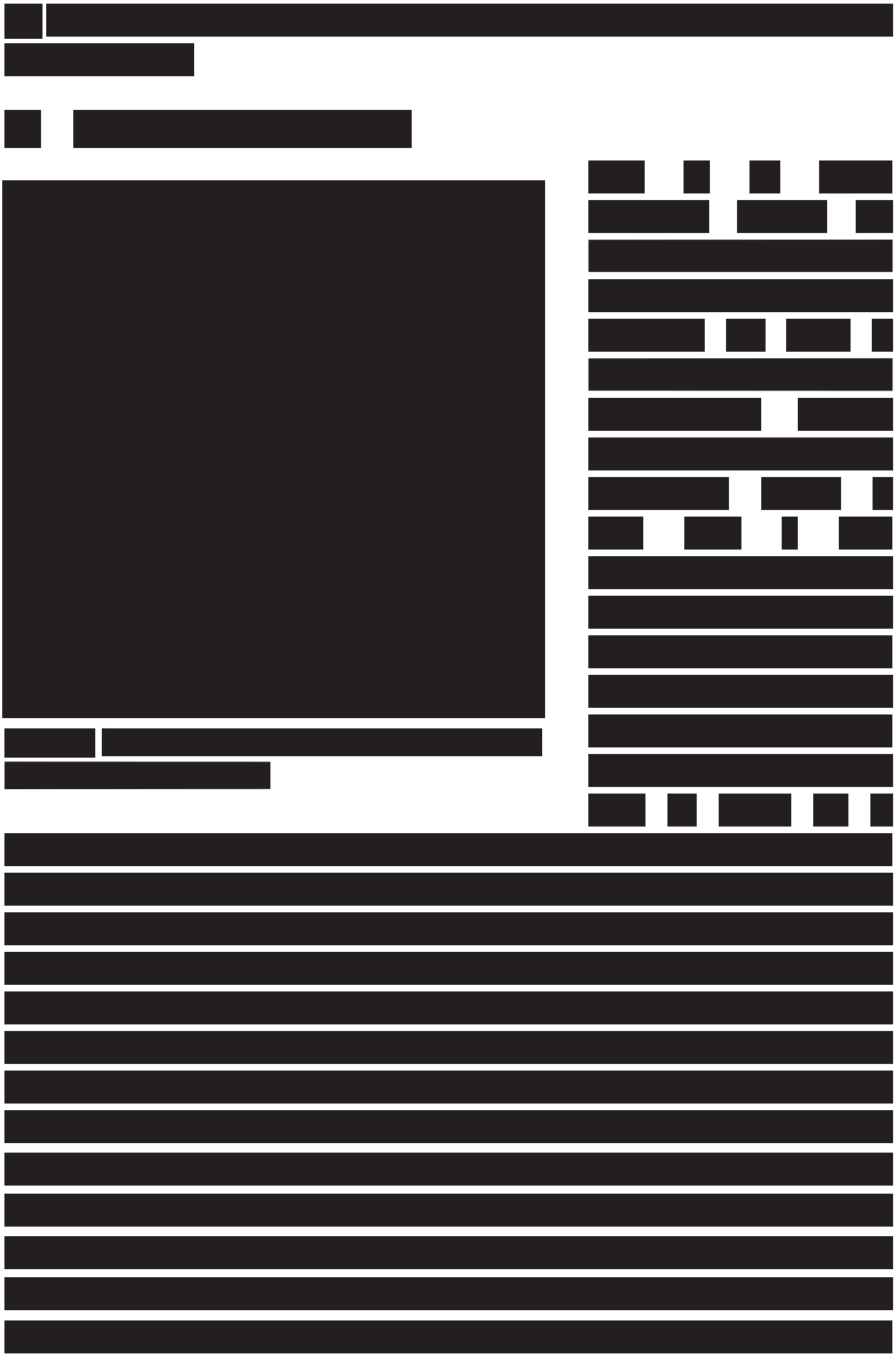


Figure 3.1: Tube representation of interleukin-2 (blue) complexed with its alpha receptor (cyan) created from PDB code 1Z92. Residues in the interface are represented by a rainbow gradient color code according to their side chain's predicted contribution to the binding free energy, respectively, with reddish colors indicating hot spot residues.

[redacted]





Reprinted (adapted) with permission from Krüger, D. M., Gohlke, H., DrugScore^{PPI} webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. Nucleic Acids Res. 2010, 38, W480-W486. DOI: 10.1093/nar/gkq471. Copyright The Author(s) 2010. Published by Oxford University Press. URL: http://www.oxfordjournals.org/.

4. Knowledge-based Prediction of Protein-Protein Complexes (Paper II)

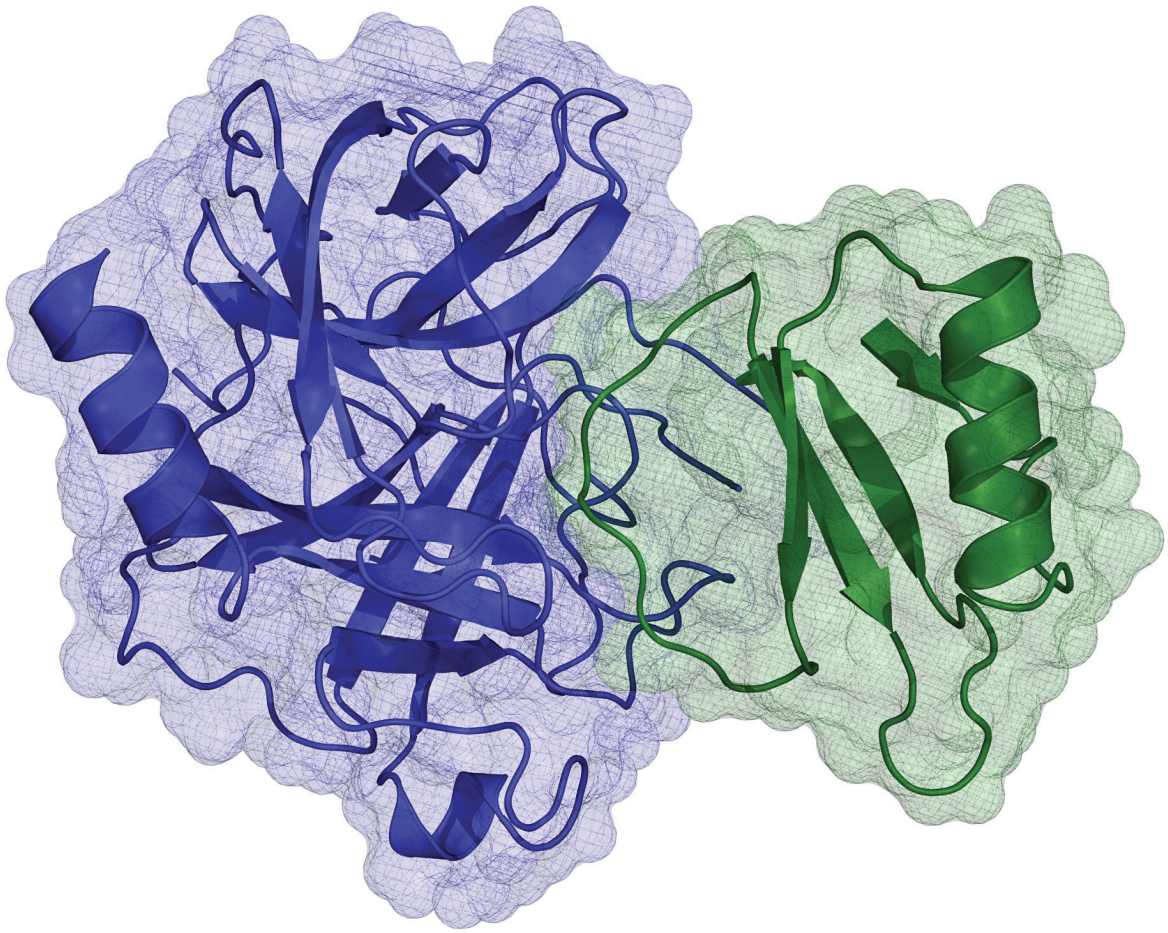
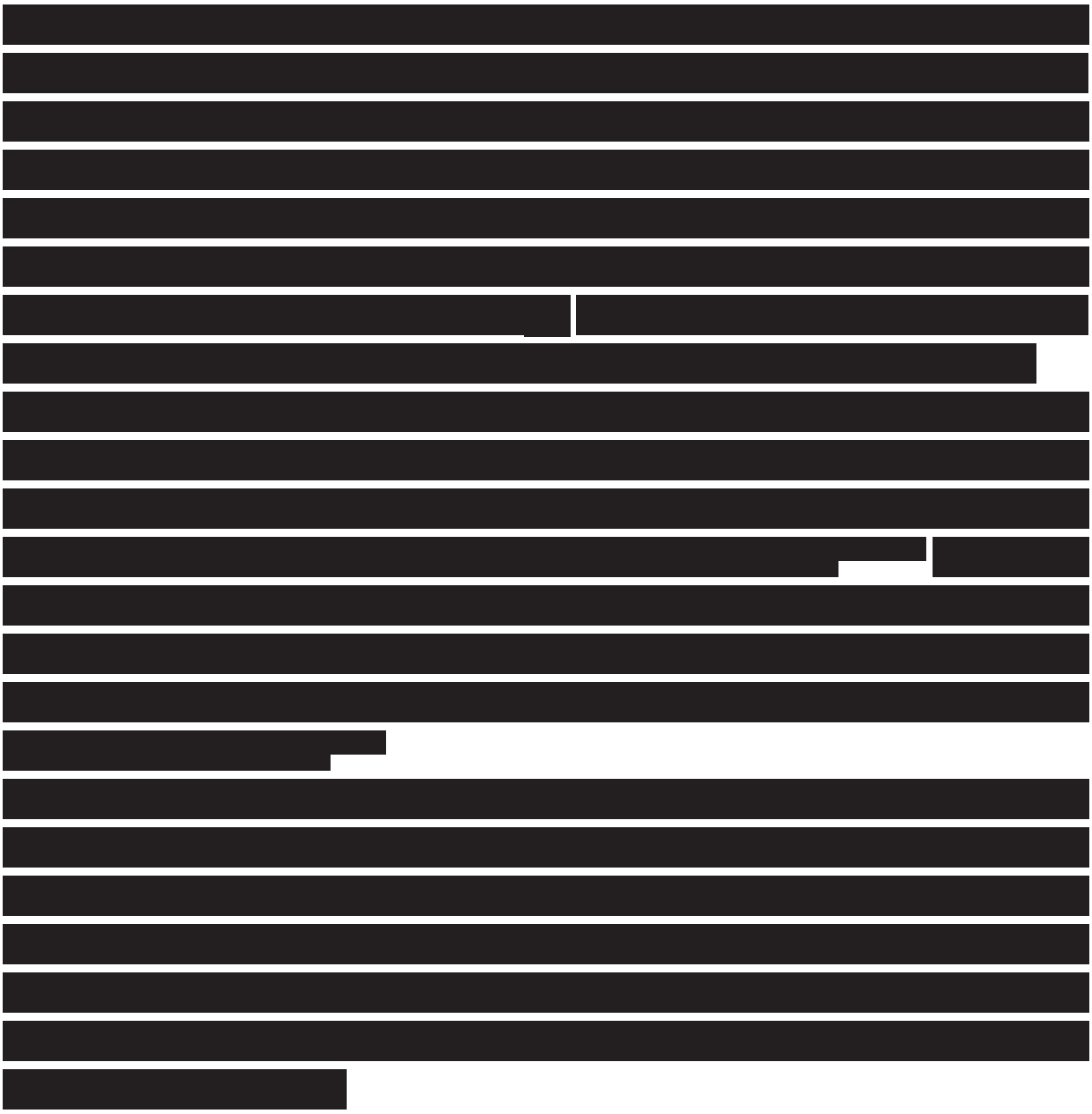


Figure 4.1: Illustration of a protein-protein complex given by bovine alpha-chymotrypsin (blue) binding to eglin c (green) created from PDB code 1ACB. Proteins are depicted in cartoon representation at which the respective van der Waals surface is given in the background.

[redacted]



Reprinted (adapted) with permission from Krüger, D. M., Ignacio-Garzon, J., Chacon, P., Gohlke, H., DrugScore^{PPI} knowledges-based potentials used as scoring and objective function in protein-protein docking. PLoS ONE, 2014, 9, e89466. URL: http://www.plosone.org/.

5. Accounting for Local and Global Flexibility in Proteins (Paper III)

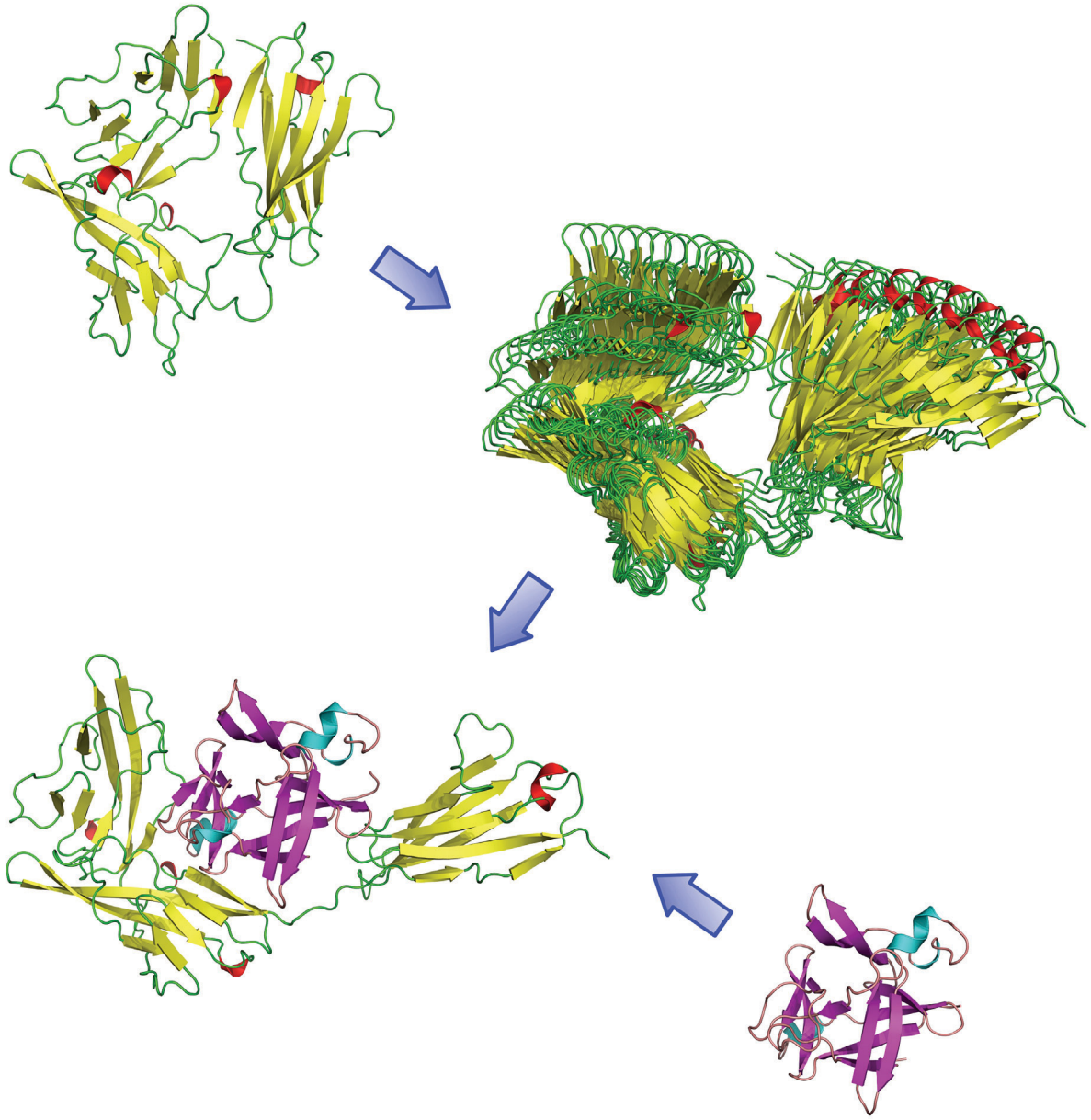


Figure 5.1: Scheme showing conformational changes of the interleukin-1 receptor (yellow/green/red) when binding to the interleukin-1 receptor antagonist (magenta/salmon/cyan); the conformations were obtained from a targeted geometric simulation using structural information from PDB code 1GOY and 1ILR (unbound structures), and PDB code 1IRA (complex structure). Proteins are depicted in cartoon representation.

Proteins are flexible biomolecules that are able to perform a wide range of possible global and local movements (see Chapter 1.1.2). Taking those movements into account is still one of the major concerns in the field of protein-protein docking.[201, 223] Indeed, in chapter 4.3 we showed that in comparison to the results from bound docking, DrugScore^{PPI} is not "soft" enough to compensate for the missing explicit treatment of protein flexibility. Thus, in this chapter, several approaches will be tested for their applicability to account for local or global flexibility in protein-protein docking. To this end, first, local protein movements have been explored *a posteriori* by several refinement procedures with the aim to improve pre-calculated rigid-body docking solutions. Second, global protein movements have been addressed *a priori* following a conformational selection approach according to which one should be able to predict bound protein conformations based on unbound ones for later use in protein-protein docking.

5.1 Refinement of Predicted Protein-Protein Complexes

Bound and unbound protein-protein docking solutions were generated from a dataset of 97 protein binding partners providing a hit list of 2000 protein complex predictions each (see Chapter 4). For unbound docking solutions, several refinement methods were applied afterwards in order to optimize the quality of the docking solutions for re-ranking of the hit lists as described in the following (see Table 5.1).

Table 5.1: List of methods and corresponding programs that were used for refinement of bound and unbound protein-protein complex predictions. Results achieved from these methods were either used to modify the DrugScore^{PPI} docking scores and/or to provide new hit lists thereby obtained.

Methods and corresponding programs used for refinement
Energy minimization with AMBER[287]
Side chain rotamer sampling with SCWRL[288]
Calculation of Δ*SASA* with Connolly's MS program[289]
Binding affinity estimation with DCOMPLEX[290]
Assessing stereochemical quality with PROCHECK[291]
Assessing stereochemical quality with MolProbity[292]
Interface prediction with ProMate[293]

Energy minimization with AMBER. The Amber (version 11) suite of programs together with the ff99SB modifications of the Cornell *et al.* force field was used to perform minimizations of pre-calculated protein-protein complexes.[287] Missing side chains were added manually according to paper 2, and waters were removed. Hydrogen atoms were automatically added by the LEaP program of the AMBER suite. Accordingly, for amino acid side chains a standard protonation state is assumed, i.e., Asp and Glu are

treated as deprotonated, Arg, Lys, and His as protonated. All complexes were then minimized by 100 steps of steepest descent minimization in order to remove clashes in the binding interface. Here, long-range electrostatic interactions were treated by the particle mesh Ewald (PME) method, and bond lengths were constrained using the SHAKE algorithm.[294] All minimized complexes were rescored with DrugScore^{PPI} and ranked accordingly.

Side chain rotamer sampling with SCWRL. The SCWRL (version 4.0) program was used to optimize the steric fitting of side chains in the interfaces of pre-calculated protein-protein complexes by side chain rotamer sampling.[288] SCWRL was initially developed to determine side chain conformations in protein structure prediction. In this regard, we tried to sample appropriate side chain conformations for the protein-protein complex predictions. All modified complexes were rescored with DrugScore^{PPI} and ranked accordingly.

Calculation of Δ*SASA* with Connolly's MS program. The change in the solvent accessible surface (Δ*SASA*) upon protein-protein binding was calculated by Connolly's molecular surface (MS) program (see Equation 5.1).[289] The term Δ*SASA* describes the buried surface area at the interface of a protein-protein complex and thus serves as an indicator for the size of the interface and the contribution of desolvation effects to the binding free energy. In this regard, it was used to modify the DrugScore^{PPI} docking scores (see Equation 5.2) to provide newly ranked hit lists.

$$\Delta SASA = SASA_{rec} + SASA_{lig} - SASA_{com} \qquad \text{(Eq. 5.1)}$$

Here, *SASA* is the solvent accessible surface of the receptor ($SASA_{rec}$), of the ligand ($SASA_{lig}$), and of the receptor-ligand complex ($SASA_{com}$).

$$\Delta W_{\text{mod}} = \frac{\Delta W}{\Delta SASA} \qquad \text{(Eq. 5.2)}$$

Here, Δ*W* is the docking score for a protein-protein docking solution as the sum of all occurring atom-atom interactions, and Δ*SASA* is the buried surface area at the interface of this complex.

Binding affinity estimation with DCOMPLEX. DCOMPLEX (no version number available) is a program that was established for the prediction of binding affinities and the energy evaluation of protein-protein complexes based on a statistical monomer-based potential.[290] The

program was used to rescore all complexes and to rank them according to the scores thereby obtained.

Assessing stereochemical quality with PROCHECK. The PROCHECK (version 3.5.4) suite of programs gives an assessment of the overall quality of a protein structure performing a detailed check on its stereochemistry.[291] Here, the covalent geometry of single residues in a given protein structure is compared to stereochemical parameters derived from high-resolution protein structures by Morris *et al.* as well as bond lengths and bond angles derived from well-refined, high-resolution structures of small-molecules in the Cambridge Structural Database (CSD) obtained by Engh *et al.*[291] For refinement, predicted protein-protein complexes were checked for their stereochemical quality and ranked accordingly. Furthermore, the top 100/500/1000 structures showing the best quality were ranked according to their DrugScore^{PPI} docking scores following the idea that a good stereochemical quality better reflects the distance-dependent nature of the pair potentials, as shown in chapter 4.

Assessing stereochemical quality with MolProbity. MolProbity (version 3.19) is a suite of programs that provides a broad-spectrum based evaluation of the quality of a protein structure at both the global and local levels.[292] It relies on optimized hydrogen placement, all-atom contact analysis, torsion angle analysis (including Ramachandran and rotamer analysis), and covalent-geometry analysis. MolProbity was initially developed to check for errors in X-ray structures. The output provides scores for clashes between atoms, unfavourable Ramachandran backbone torsion angles, and bad side chain rotamers. Predicted protein-protein complexes were checked for their overall structural quality and ranked accordingly. Furthermore, the top 100/500/1000 structures showing the best quality were ranked according to their DrugScore^{PPI} docking scores following the idea that a good structural quality better reflects the distance-dependent nature of the pair potentials (*vide supra*).

Interface prediction with ProMate. ProMate (version 2.0) is a structure-based program developed to predict the location of protein-protein interfaces of unbound proteins.[293] The method was validated on transient heteromeric complexes, whereas antibody-antigen interactions were excluded because of their specific nature (see Chapter 4.2). Interface locations can be predicted based on different structural properties that have been shown to be relevant for the distinction between binding and non-binding interfaces, namely, distribution of atoms, chemical character of atoms, distribution of single amino acids, distribution of amino acid pairs, evolutionary conservation, sequence distances within a circle, non-regular secondary structure

lengths, domains, hydrophobic patches, and crystallographic data (if available) in terms of bound water and the distribution of B-factors. For refinement, interfaces were predicted for unbound binding partners and all protein-protein docking solutions that do not correspond to at least one of the predicted interfaces were removed from the hit list, respectively.

5.2 Sampling the Conformational Space of Proteins in Free State

The conformational selection model describes the intrinsic ability of proteins to switch between conformationally distinct states under native conditions leading to conformational transitions that occur on a wide range of scales, both in time and space. Such transitions are well-known in the case of ligand binding to several pharmacologically important proteins, where ligand binding occurs upon selection of the "correct" conformational state of the protein out of an conformational ensemble.[33] A controversial debate has been conducted in the last years whether structural variation of a protein during ligand binding follows as selective stabilization of conformational states pre-existing independent of a ligand ("conformational selection model"), or an induced fit mechanism. At this time, both the mechanisms have been evidenced to play a role in protein-protein recognition. [27, 29] Based on the conformational selection model we applied three geometric simulation techniques, CONCOORD (version 2.1), FRODAN (version 1.0) and NMSim (version 1.0), for our attempts to sample bound protein conformations starting from unbound ones. The strategy was, first, trying to sample a near-native protein conformation that is very close to the native one by making use of a constrained network model (CNM), second, filtering this conformation out of the ensemble with respect to the C_α-i_rmsd, and third, using this conformation to perform a protein-protein docking. CONCOORD is a distance geometry-based approach that generates protein conformations by satisfying distance constraints derived from a starting structure.[295] FRODAN performs constrained geometric simulations by simulating diffusive motions of flexible regions and rigid clusters of proteins.[296] NMSim is a normal mode-based geometric simulation approach developed by Ahmed *et al.* for multiscale modeling of protein conformational changes.[297] NMSim is build upon a three-step approach (see Figure 5.2): First, the protein is decomposed into rigid clusters and flexible regions using the graph theoretical approach FIRST,[298] second, dynamical properties of the coarse-grained protein are revealed using an ENM representation by rigid cluster normal mode analysis (RCNMA),[297] and third, new protein conformations are generated by NMSim, whereas directions of backbone motions are biased by low-frequency normal modes and side chain motions are biased toward experimentally derived rotamer information.[297] Conformations generated thereby are optimized regarding steric clashes, constraint violations, and stereochemical accuracy. In contrast to FRODAN and CONCOORD, NMSim uses a directional guidance to sample biologically relevant

conformations. Each of the methods was applied to a data set of 15 unbound protein binding partners showing small-, medium-, and large-scale conformational changes upon protein-protein binding (see Table 5.2). In the case of NMSim, small scale, large scale, and radius of gyration (ROG)-guided motions were tested as provided by the NMSim web server (see Chapter 5.3).[299] All programs were applied using standard parameters as suggested by the authors. Subsequently, conformational ensembles were clustered by the kclust algorithm from the AMBER suite of programs using a threshold of 5 Å C_α-atom rmsd and superimposed onto the corresponding bound conformations. Finally, for each of the proteins the i_rmsd between a cluster conformation and the bound conformation was calculated (see Table 5.2).

5.3 Web Service Implementation

5.4 Results

In general, local refinement procedures (see Table 5.1) were neither able to improve the bound nor the unbound docking results, i.e. to increase the number of hits in the top 100/500/1000 predictions. Refined structures obtained from minimization with Amber or side chain sampling with SCWRL that were rescored using DrugScore^{PPI} and ranked according to the new score did not afford any improvements regarding the hit lists. The same is true when rescoring the structures with DCOMPLEX. $\Delta SASA$ values were used to modify the DrugScore^{PPI} docking scores (see Equation 5.2), but no enrichment of hits could be observed. The already implicit consideration of $\Delta SASA$ in the DrugScore^{PPI} scoring function could be an explanation for this.[152] PROCHECK was used to rank the predicted complexes by their stereochemical quality. Likewise, no enhancement of hits could be observed in the top 100/500/1000 predictions. Furthermore, Molprobity was used to calculate several scores that are related to contacts, overlaps, and hydrogen bonds, thus giving a hint about the overall structural quality of the

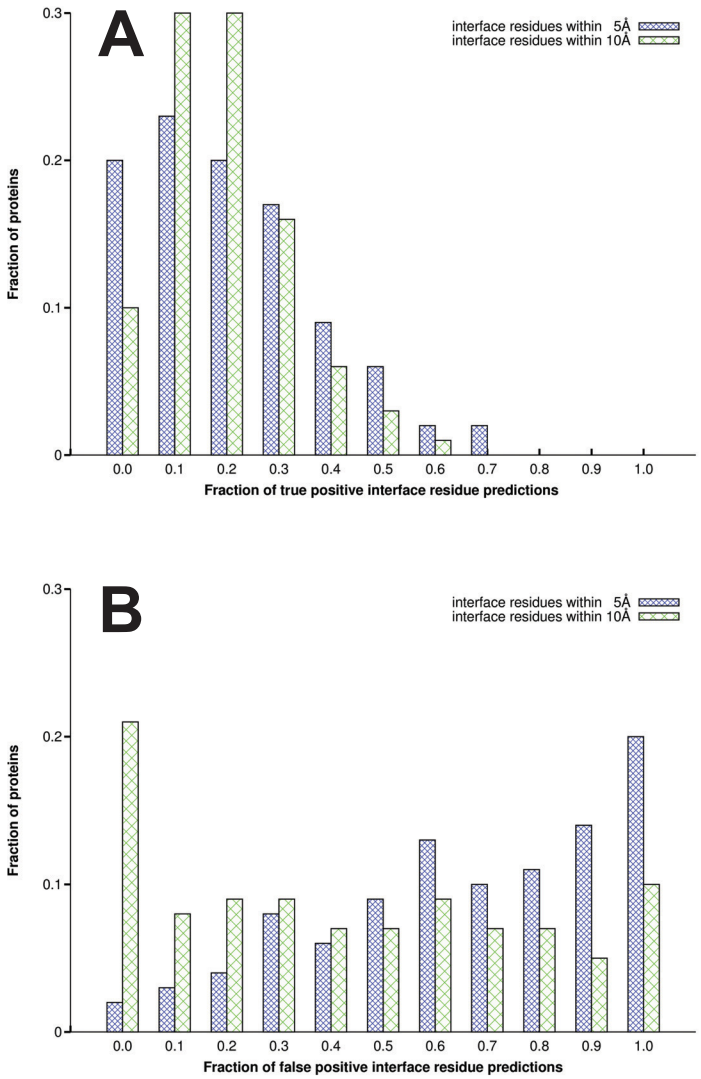


Figure 5.3: Fractions of true and false positive interface residue predictions obtained with ProMate were calculated for a threshold of 5 Å (blue, small checkered) and 10 Å (green, large checkered) distance to define the interface. (A) The fraction of correct interface residue predictions is defined by the number of correctly predicted interface residues divided by the total number of interface residues. (B) The fraction of false positive interface residue predictions is defined by the number of falsely predicted interface residues divided by the total number of predicted interface residues.

predicted complexes. These scores were both used to rank the complexes by their structural quality and to modify the DrugScore^{PPI} docking scores, however, without any success.

Finally, we applied ProMate to predict interface residues of the unbound binding partners. This interface information was used to filter the hit lists by removing all docking solutions that do not correspond to at least one of the predicted interfaces. A protein interface was defined as all residues within 5 Å (10 Å) distance of the partner protein, respectively. Results obtained from ProMate are depicted in Figure 5.3. When using a threshold of 5 Å (10 Å) to define the interface, ProMate was able to correctly predict at least one of the interface residues in 80% (90%) of the cases (see Figure 5.3 A). The fraction of correctly predicted interface residues ranges from 0.1 to 0.3 for the threshold of 5 Å (10 Å) for 60% (80%) of the proteins (see Figure 5.3 A), with the average fraction of correctly predicted interface residues being 0.17 (0.15) with STD = 0.16 (0.12). Critical for the filtering of docking solutions is the amount of false positive interface residue predictions that significantly increases the number of false positive complex predictions in the hit lists. The fraction of false positive interface residue predictions is 0.66 (0.42) with STD = 0.30 (0.33) for a threshold of 5 Å (10 Å) to define the interface (see Figure 5.3 B). The latter results show that a threshold of 10 Å

Table 5.2: Sampling the conformational space of proteins in free state by geometric simulation approaches.

PDB codes[a]		Motions[c]	i_rmsd					
bound	unbound		unbound[e]	CONCOORD[d]	FRODAN[d]	NMSim[d]		
complex	rec/lig[b]					small scale	large scale	ROG
1AY7	1RGH	small	0.48	0.93	0.51	1.32	1.02	1.32
	1A19		0.43	0.78	0.57	1.02	1.07	0.66
1DQJ	1DQQ		0.55	0.84	1.70	1.78	1.64	1.94
	3LZT		1.15	1.16	1.32	1.68	1.23	2.17
1EZU	1TRM		0.42	0.85	0.45	1.36	2.05	1.14
	1ECZ		0.43	0.34	2.65	1.08	0.47	0.90
2I25	2I21		1.94	1.87	2.01	2.21	2.35	2.53
	3LZT		1.00	1.32	1.06	2.39	3.05	2.71
2HLE	2BBA		2.07	2.02	2.15	2.42	1.99	2.91
	1IKO		0.86	0.94	1.03	1.33	1.47	2.44
1K5D	1RRP	medium	1.43	1.58	1.70	1.42	1.64	1.84
	1YRG		0.95	1.15	1.00	1.16	1.59	2.37
1M10	1AUQ		0.78	1.29	0.85	1.67	1.11	1.39
	1M0Z		3.53	3.30	3.88	3.77	3.49	4.30
1WQ1	6Q21		1.03	1.31	1.23	1.83	1.42	2.69
	1WER		2.10	2.14	2.12	3.04	2.80	2.69
1XQS	1XQR		2.27	2.26	2.42	2.18	1.99	2.34
	1S3X		0.43	0.87	0.45	0.73	0.60	0.74
2NZ8	1MH1		1.47	2.00	2.14	1.61	1.35	2.52
	1NTY		3.62	3.33	3.97	4.51	3.13	4.05
1EER	1ERN	large	1.85	1.98	2.28	2.51	2.61	3.10
	1BUY		2.43	3.11	2.62	2.59	2.43	2.48
1IBR	1F59		1.78	1.97	2.15	2.37	2.98	2.31
	1QG4		2.65	2.32	3.34	3.30	2.79	3.47
1IRA	1G0Y		8.82	8.98	8.73	7.92	8.79	8.91
	1ILR		0.92	1.41	1.57	1.42	2.11	2.19
1PXV	1X9Y		3.88	4.84	4.56	4.82	4.70	4.36
	1NYC		0.84	1.01	0.93	1.45	4.35	3.24
2OT3	1TXU		1.09	1.22	1.37	2.20	1.64	1.51
	1YZU		4.71	4.45	5.46	5.25	4.44	4.87
STD_{small}[f]			0.64	0.70	0.88	1.00	1.03	1.17
STD_{medium}[f]			1.18	1.21	1.32	1.43	1.22	1.56
STD_{large}[f]			2.14	2.25	2.31	2.25	2.41	2.40

[a] PDB codes for bound protein structures (complex structures) and unbound protein structures.

[b] Receptor and ligand top down in this order, respectively. The larger protein is considered to be the receptor.

[c] Degree of structural motions (conformational changes) that occur upon complexation.

[d] Lowest i_rmsd between a cluster conformation and the bound conformation. In Å.

[e] i_rmsd between bound and unbound protein structures. In Å.

[f] STDs of i_rmsd values for proteins involved in small-, medium-, and large-scale motions upon complexation.

is more appropriate than 5 Å because the fraction of false positive interface residue predictions is significantly lower in relation to the total number of interface residues (i.e., higher specificity), whereas the fraction of correctly predicted interface residues is almost similar (i.e., similar sensitivity). Nevertheless, even here the amount of false positive interface residue predictions was still too high to increase the number of hits in the top 100/500/1000 docking solutions.

To account for global flexibility, we tried to sample bound protein conformations starting from unbound ones by three different geometric simulation approaches. Results are shown in Table 5.2. In summary, none of the approaches was able to sample at least one protein conformation that is closer to the bound conformation than the unbound one. The i_rmsd values calculated for a sampled conformation and the corresponding unbound conformation are quite similar in each of the cases. STDs calculated for the i_rmsd values of motion-related subsets also show similar results. Note that i_rmsd values give information about the differences in the positions of atoms, but do not consider any type of movement that occurs upon binding. In this regard, the degree of structural motions describes the overall amount of structural rearrangements that occur in both of the proteins upon complexation. However, it can be observed that in many cases only one of the binding partners undergoes large conformational changes whereas the other one stays almost unchanged (see Figure 5.1). Therefore, structural motions classified as small upon complexation can be classified as medium considering only one of the binding partners, and vice versa (see Table 5.2). The reason why sampling of bound conformations failed is well-grounded in the characteristics of the underlying coarse-grained ENM representation of the proteins that precludes the formation or breaking of bonds between the atoms. Consequently, quite flexible structure elements, e.g. loops that often occur in protein-protein interfaces and undergo a reshaping upon binding in many cases, cannot be sampled successfully since the constraints between the atoms remain fixed. At least, geometric simulations can be used to reproduce such motions by targeted simulations (see Figure 5.1). A possibility to overcome the current limitations of geometric simulation methods is the use of MD simulations (see Chapter 8).

5.5 Conclusions

The local refinement procedures applied in this work were neither able to improve the bound nor the unbound docking predictions. In the case of bound docking, where convincing results were obtained already, additional knowledge-based pair potentials specifically for antigen-antibody complexes might lead to improvements of the results (see Chapter 4.2). In the case of unbound docking, the unbound backbone conformations of the binding partners lead to discrepancies in the predicted and the bound protein-protein complex configurations. Thus, local refinement methods might only be successful when global (backbone) flexibility has already been considered

successfully. In this regard, geometric simulation techniques have been shown to be a computationally efficient alternative to molecular dynamics simulations for conformational sampling of proteins.[297] However, these constrained simulations reach their limits when the conformational pathways contain disordered-to-structured transitions. This drawback could be overcome by more sophisticated sampling techniques such as MD simulations when the sampling problem is adequately addressed there (see Chapter 8).

Reprinted (adapted) with permission from Krüger, D. M., Ahmed, A. Gohlke, H., NMSIM Web Server: Normal mode-based geometric simulations for exploring biologically relevant conformational transitions in proteins. Nucleic Acids Res. 2012, 40, W310-W316. DOI: 10.1093/nar/gks478. Copyright The Author(s) 2012. Published by Oxford University Press. URL: http://www.oxfordjournals.org/.

6. Targeting Protein-Protein Interfaces with Small Molecules (Paper IV)

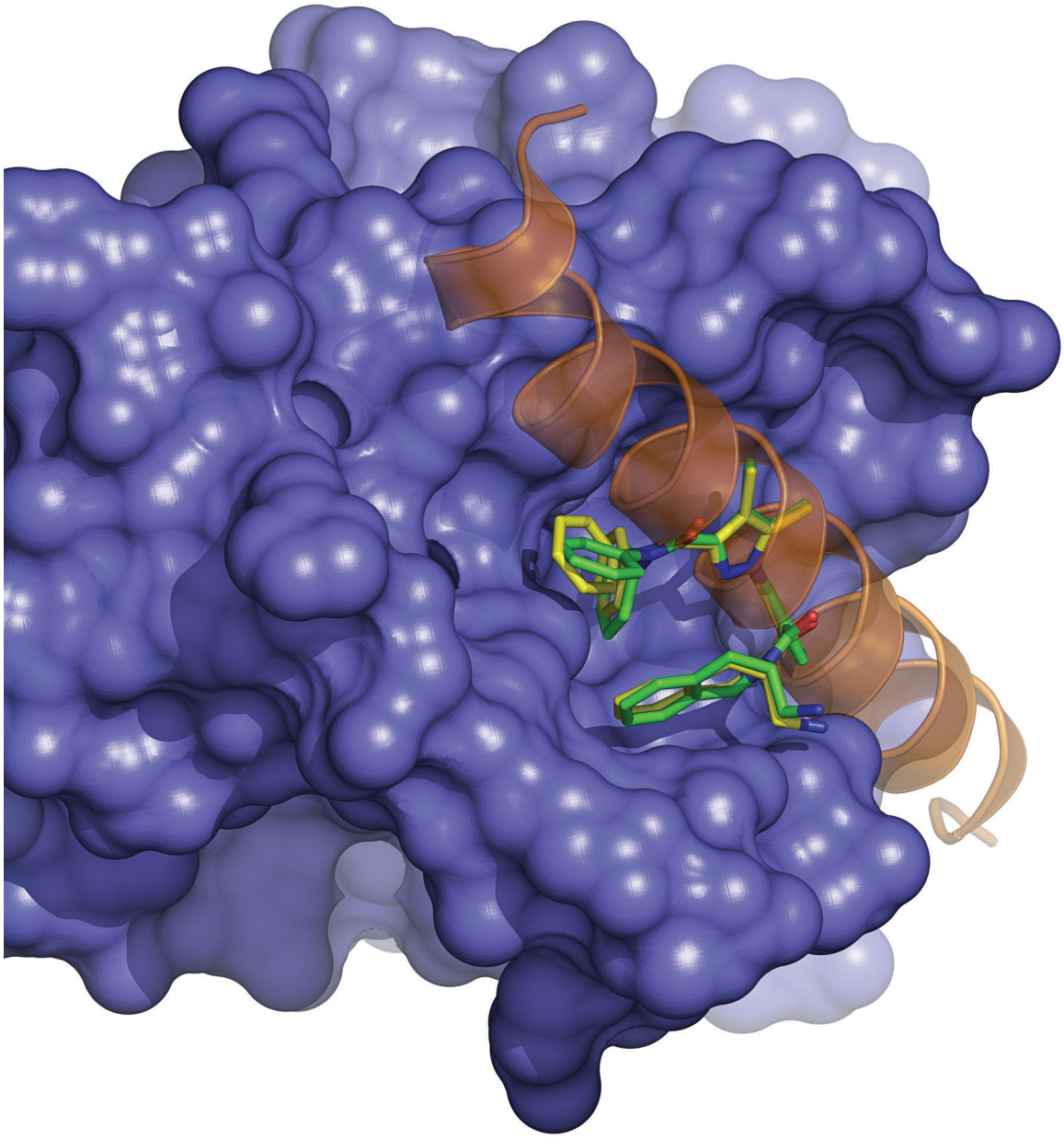
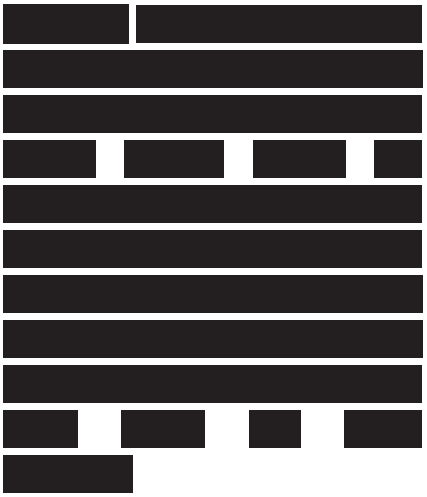


Figure 6.1: Phenyl pyrazole inhibitor (shown as sticks) binding to the apoptosis regulator BCL-2 (blue surface). The crystal structure configuration is depicted in green (PDB code 2W3L), a predicted solution from molecular docking is depicted in yellow. In orange cartoon representation, the crystal structure configuration of BAD (BCL-2 antagonist of cell death) is shown (PDB code 2BZW).

[redacted]

[redacted]



Reprinted (adapted) with permission from Krüger, D. M., Jessen, G., Gohlke, H., How good are stat-of-the art docking tools in predicting ligand binding modes in protein-protein interfaces? J. Chem. Inf. Model. 2012, 52, 2807-2811. DOI: 10.1021/ci3003599. Copyright (2012) American Chemical Society.

7. Summary

The main objective of this thesis was the investigation of protein-protein interactions and making use of this knowledge to develop a new approach to predict protein-protein complexes. As a basis for this research, a knowledge-based scoring function DrugScore^{PPI} was developed of which the pair-potentials were derived from a previously prepared dataset of protein-protein complexes (see Chapter 3 & 4). Based on these results, DrugScore^{PPI} was used for computational alanine scanning to predict changes in the binding free energy of protein-protein complexes upon mutations in the interface (see Chapter 3). Computed and experimental values showed good correlations and, thus, a QSAR-model was built to improve the predictive power. Based on these findings, the DrugScore^{PPI} web server was developed, which allows identifying hotspot residues in protein-protein interfaces and performing computational alanine scanning of a protein-protein interface within a few minutes (see Chapter 3). The results demonstrate that the DrugScore^{PPI} web server outperforms other state-of-the-art methods not only with respect to predictive power but also in terms of computational speed. DrugScore^{PPI} was successfully evaluated by rescoring two non-redundant datasets for which bound and unbound protein-protein complex predictions have been generated (see Chapter 4). Furthermore, DrugScore^{PPI} was applied as an objective function in combination with a fast spherical harmonics-based protein-protein docking tool in order to predict 3D structures of protein-protein complexes (see Chapter 4). For this, pre-calculated knowledge-based potential grids by DrugScore^{PPI} were used to sample protein-protein configurations and to successfully identify near-native configurations. The approach showed good results for bound and moderate results for unbound protein-protein complex predictions. A few parameters were identified to have an influence on the success of the protein-protein docking approach, such as the range of possible conformational changes of a protein, or crystal packing contacts (see Chapter 4). To account for protein flexibility, several methods were examined in order to incorporate local and global flexibility into the docking approach (see Chapter 5). In this regard, a normal mode-based geometric simulation (NMSim) method was used to sample input conformations of proteins following the idea of a conformational selection mechanism (see Chapter 5). Finally, a large-scale validation study on docking small molecules into protein-protein interfaces was performed (see Chapter 6). Results thereby obtained allow identifying those protein-protein interfaces that are amenable for molecular docking approaches.

8. Perspective

In recent years, targeting protein-protein interactions has become a major working point in structure-based drug design. Abundant information about sequences and structures of proteins afforded from large-scale genomic approaches poses new challenges in the field of systems biology. Proteins in a cell may have a multitude of potential different binding partners and thus may undergo many transient protein-protein interactions. The detailed investigation of disease-related cell-signaling pathways unravelled a huge quantity of associated proteins as possible drug targets. In this context, the prediction of protein-protein interactions by computational methods plays an important role because 3D structures of protein-protein complexes are often difficult or impossible to elucidate by experimental methods. Algorithmic strategies have been developed to dock proteins, to evaluate the resulting complexes by different scoring functions, and to detect residues in the protein-protein interface that are relevant for binding. However, good protein-protein docking predictions are only obtained for proteins that do not undergo large conformational changes. Conversely, protein-protein docking and scoring is effective in identifying near-native complex structures in cases where docking was performed using bound structures or structures that only show small conformational changes.

Consideration of protein flexibility still represents a major issue for structure-based drug design approaches. Despite many successful developments that have been made to consider large protein conformational changes, usually by reducing the problem to computationally efficient approaches, i.e., using normal-mode analysis, constrained networks, flexible grids, or structural ensembles, the problem is still unsolved. Especially when it comes to large biomolecules like proteins, efficient conformational sampling is considerably complex. Currently, the only possibility to efficiently sample protein conformational changes is to make use of unrestricted MD simulations. However, the problem is beyond standard methods when proteins undergo long-range folding mechanisms or disordered-to-structured transitions that occur on a microsecond timescale. Nevertheless, the computational and algorithmic development in the last decade gives reason to hope that computational costs to perform such simulations can be overcome in the near future.[304-306] GPU-based computing combined with enhanced sampling like accelerated MD (aMD),[307] or replica exchange molecular dynamics (REMD)[308] for enhanced sampling of conformational space were successfully applied to simulate pathways of small peptide folding and protein-ligand binding.[309-313] Furthermore, to reduce the computational demand of unbiased brute force MD simulations, Markov state models (MSMs) have been developed that make use of multiple shorter-timescale simulations to reach longer-timescale transitions of proteins.[314-316] In this regard, MD-based sampling of protein transition pathways

would allow for continuation of this work. It should be tested in how far MD simulations are suitable to sample bound protein conformations starting from unbound ones for later use in protein-protein docking. Thus, it should be at least possible to sample those protein conformations that occur on a nanosecond timescale and do not rely on induced fit mechanisms.

9. References

1. Wells, J. A.; McClendon, C. L., Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007,** 450, (7172), 1001-9.
2. Mulder, G. J., Ueber die Zusammensetzung einiger thierischen Substanzen. *Journal für praktische Chemie* **1839,** 16, 129-151.
3. Campbell, N. A., *Biologie*. Spektrum Akademischer Verlag: Heidelberg, Berlin, Oxford, 1997; p 80.
4. Crick, F. H., The genetic code--yesterday, today, and tomorrow. *Cold Spring Harb Symp Quant Biol* **1966,** 31, 1-9.
5. Atkins, J. F.; Gesteland, R., Biochemistry. The 22nd amino acid. *Science* **2002,** 296, (5572), 1409-10.
6. Xu, X. M.; Carlson, B. A.; Mix, H.; Zhang, Y.; Saira, K.; Glass, R. S.; Berry, M. J.; Gladyshev, V. N.; Hatfield, D. L., Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol* **2007,** 5, (1), e4.
7. Gaston, M. A.; Zhang, L.; Green-Church, K. B.; Krzycki, J. A., The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature* **2011,** 471, (7340), 647-50.
8. Brotman, M.; McGill, D. B., Protein biosynthesis and the genetic code. *Mayo Clin Proc* **1964,** 39, 777-91.
9. Caspersson, T., The relations between nucleic acid and protein synthesis. *Symp Soc Exp Biol* **1947**, (1), 127-51.
10. Watson, J. D., The synthesis of proteins upon ribosomes. *Bull Soc Chim Biol* **1964,** 46, 399-425.
11. Straub, F. B., Formation of the secondary and tertiary structure of enzymes. *Adv Enzymol Relat Areas Mol Biol* **1964,** 26, 89-114.
12. Anfinsen, C. B., The formation of the tertiary structure of proteins. *Harvey Lect* **1967,** 61, 95-116.
13. Berg, M. J., Tymoczoko, J. L., Stryer, L., *Biochemie*. Spektrum Akademischer Verlag: Heidelberg, Berlin, Oxford, 2003; Vol. 5, p 45-83.
14. Klotz, I. M.; Langerman, N. R.; Darnall, D. W., Quaternary structure of proteins. *Annu Rev Biochem* **1970,** 39, 25-62.
15. Wall, J. S., Disulfide bonds: determination, location, and influence on molecular properties of proteins. *J Agric Food Chem* **1971,** 19, (4), 619-25.
16. Creighton, T. E., Molecular chaperones. Unfolding protein folding. *Nature* **1991,** 352, (6330), 17-8.
17. Tokuriki, N.; Tawfik, D. S., Protein dynamism and evolvability. *Science* **2009,** 324, (5924), 203-7.
18. Huber, R., Conformational flexibility in protein molecules. *Nature* **1979,** 280, (5723), 538-9.
19. Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J., Conformational entropy in molecular recognition by proteins. *Nature* **2007,** 448, (7151), 325-9.
20. Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G., The energy landscapes and motions of proteins. *Science* **1991,** 254, (5038), 1598-603.
21. Marsh, J. A.; Teichmann, S. A.; Forman-Kay, J. D., Probing the diverse landscape of protein flexibility and binding. *Curr Opin Struct Biol* **2012,** 22, (5), 643-50.
22. Tobi, D.; Bahar, I., Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci U S A* **2005,** 102, (52), 18908-13.
23. Boehr, D. D.; Nussinov, R.; Wright, P. E., The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **2009,** 5, (11), 789-96.
24. Gerstein, M.; Lesk, A. M.; Chothia, C., Structural mechanisms for domain movements in proteins. *Biochemistry* **1994,** 33, (22), 6739-49.

25. Marsh, J. A.; Teichmann, S. A., Relative solvent accessible surface area predicts protein conformational changes upon binding. *Structure* **2011,** 19, (6), 859-67.
26. Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D., Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **2005,** 438, (7064), 117-21.
27. Changeux, J. P.; Edelstein, S., Conformational selection or induced fit? 50 years of debate resolved. *Biol Rep* **2011,** 3, 19.
28. Wright, P. E.; Dyson, H. J., Linking folding and binding. *Curr Opin Struct Biol* **2009,** 19, (1), 31-8.
29. Csermely, P.; Palotai, R.; Nussinov, R., Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* **2010,** 35, (10), 539-46.
30. Wright, P. E.; Dyson, H. J., Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **1999,** 293, (2), 321-31.
31. Chouard, T., Structural biology: Breaking the protein rules. *Nature* **2011,** 471, (7337), 151-3.
32. Mittag, T.; Kay, L. E.; Forman-Kay, J. D., Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* **2010,** 23, (2), 105-16.
33. Lin, J. H., Accommodating protein flexibility for structure-based drug design. *Curr Top Med Chem* **2011,** 11, (2), 171-8.
34. Spyrakis, F.; BidonChanal, A.; Barril, X.; Luque, F. J., Protein flexibility and ligand recognition: challenges for molecular modeling. *Curr Top Med Chem* **2011,** 11, (2), 192-210.
35. Tuffery, P.; Derreumaux, P., Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: limitations of current computational approaches. *J R Soc Interface* **2012,** 9, (66), 20-33.
36. Kenakin, T.; Miller, L. J., Seven transmembrane receptors as shapeshifting proteins: the impact of allosteric modulation and functional selectivity on new drug discovery. *Pharmacol Rev* **2010,** 62, (2), 265-304.
37. Wang, J.; Cao, Z.; Zhao, L.; Li, S., Novel strategies for drug discovery based on intrinsically disordered proteins (IDPs). *Int J Mol Sci* **2012,** 12, (5), 3205-19.
38. Bahar, I.; Lezon, T. R.; Yang, L. W.; Eyal, E., Global dynamics of proteins: bridging between structure and function. *Annu Rev Biophys* **2012,** 39, 23-42.
39. Hall, D. A.; Keech, M. K.; Reed, R.; Saxl, H.; Tunbridge, R. E.; Wood, M. J., Collagen and elastin in connective tissue. *J Gerontol* **1955,** 10, (4), 388-400.
40. Kitahara, T.; Ogawa, H., The extraction and characterization of human nail keratin. *J Dermatol Sci* **1991,** 2, (6), 402-6.
41. Webster, R. E.; Henderson, D.; Osborn, M.; Weber, K., Three-dimensional electron microscopical visualization of the cytoskeleton of animal cells: immunoferritin identification of actin- and tubulin-containing structures. *Proc Natl Acad Sci U S A* **1978,** 75, (11), 5511-5.
42. Choi, S. B.; Wang, C.; Muench, D. G.; Ozawa, K.; Franceschi, V. R.; Wu, Y.; Okita, T. W., Messenger RNA targeting of rice seed storage proteins to specific ER subdomains. *Nature* **2000,** 407, (6805), 765-7.
43. Muntz, K., Deposition of storage proteins. *Plant Mol Biol* **1998,** 38, (1-2), 77-99.
44. Chiancone, E.; Ceci, P.; Ilari, A.; Ribacchi, F.; Stefanini, S., Iron and proteins for iron storage and detoxification. *Biometals* **2004,** 17, (3), 197-202.
45. Sanchez, E. J.; Lewis, K. M.; Danna, B. R.; Kang, C., High-capacity Ca^{2+} binding of human skeletal calsequestrin. *J Biol Chem* **2012,** 287, (14), 11592-601.

46. Gettins, P. G., Serpin structure, mechanism, and function. *Chem Rev* **2002,** 102, (12), 4751-804.
47. Kelly, A.; Powis, S. H.; Kerr, L. A.; Mockridge, I.; Elliott, T.; Bastin, J.; Uchanska-Ziegler, B.; Ziegler, A.; Trowsdale, J.; Townsend, A., Assembly and function of the two ABC transporter proteins encoded in the human major histocompatibility complex. *Nature* **1992,** 355, (6361), 641-4.
48. Klotz, I. M.; Klotz, T. A., Oxygen-carrying proteins: a comparison of the oxygenation reaction in hemocyanin and hemerythrin with that in hemoglobin. *Science* **1955,** 121, (3145), 477-80.
49. Shikama, K., Nature of the FeO_2 bonding in myoglobin and hemoglobin: A new molecular paradigm. *Prog Biophys Mol Biol* **2006,** 91, (1-2), 83-162.
50. Doyle, D. A.; Morais Cabral, J.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R., The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **1998,** 280, (5360), 69-77.
51. Abbott, A., Membrane proteins: channel voyager makes waves. *Nature* **2003,** 426, (6968), 755.
52. Scott, J. D.; Pawson, T., Cell signaling in space and time: where proteins come together and when they're apart. *Science* **2009,** 326, (5957), 1220-4.
53. Brown, P. A.; Feinstein, M. B.; Sha'afi, R. I., Membrane proteins related to water transport in human erythrocytes. *Nature* **1975,** 254, (5500), 523-5.
54. Tournaire-Roux, C.; Sutka, M.; Javot, H.; Gout, E.; Gerbeau, P.; Luu, D. T.; Bligny, R.; Maurel, C., Cytosolic pH regulates root water transport during anoxic stress through gating of aquaporins. *Nature* **2003,** 425, (6956), 393-7.
55. Reuter, H., Calcium channel modulation by neurotransmitters, enzymes and drugs. *Nature* **1983,** 301, (5901), 569-74.
56. Saltiel, A. R.; Kahn, C. R., Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **2001,** 414, (6865), 799-806.
57. Eadie, G. S., On the evaluation of the constants V_m and K_m in enzyme reactions. *Science* **1952,** 116, (3025), 688.
58. Northrop, J. H., The mechanism of an enzyme reaction as exemplified by pepsin digestion. *Science* **1921,** 53, (1373), 391-3.
59. Crawford, M. A., Lactase deficiency. *Nature* **1969,** 223, (5207), 742.
60. Fagraeus, A., Plasma cellular reaction and its relation to the formation of antibodies in vitro. *Nature* **1947,** 159, (4041), 499.
61. Burton, D. R.; Poignard, P.; Stanfield, R. L.; Wilson, I. A., Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **2012,** 337, (6091), 183-6.
62. Kim, E. K.; Choi, E. J., Pathological roles of MAPK signaling pathways in human diseases. *Biochim Biophys Acta* **2010,** 1802, (4), 396-405.
63. Neurath, M. F.; Finotto, S., IL-6 signaling in autoimmunity, chronic inflammation and inflammation-associated cancer. *Cytokine Growth Factor Rev* **2011,** 22, (2), 83-9.
64. Schwartz, M. A.; Ginsberg, M. H., Networks and crosstalk: integrin signalling spreads. *Nat Cell Biol* **2002,** 4, (4), E65-8.
65. Shackelford, J.; Pagano, J. S., Tumor viruses and cell signaling pathways: deubiquitination versus ubiquitination. *Mol Cell Biol* **2004,** 24, (12), 5089-93.

66. Shaanan, B.; Gronenborn, A. M.; Cohen, G. H.; Gilliland, G. L.; Veerapandian, B.; Davies, D. R.; Clore, G. M., Combining experimental information from crystal and solution studies: joint X-ray and NMR refinement. *Science* **1992,** 257, (5072), 961-4.

67. Maugh, T. H., 2nd, X-ray crystallography: A refinement of technique. *Science* **1974,** 186, (4167), 913-4.

68. Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.; Futaki, S.; Ito, Y.; Hiroaki, H.; Shirakawa, M., High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **2009,** 458, (7234), 106-9.

69. Joosten, R. P.; Womack, T.; Vriend, G.; Bricogne, G., Re-refinement from deposited X-ray data can deliver improved models for most PDB entries. *Acta Crystallogr D Biol Crystallogr* **2009,** 65, (Pt 2), 176-85.

70. Jensen, L. H., Refinement and reliability of macromolecular models based on X-ray diffraction data. *Methods Enzymol* **1997,** 277, 353-66.

71. Rehr, J. J.; Ankudinov, A. L., Progress and challenges in the theory and interpretation of X-ray spectra. *J Synchrotron Radiat* **2001,** 8, (Pt 2), 61-5.

72. Bourne, P. E.; Addess, K. J.; Bluhm, W. F.; Chen, L.; Deshpande, N.; Feng, Z.; Fleri, W.; Green, R.; Merino-Ott, J. C.; Townsend-Merino, W.; Weissig, H.; Westbrook, J.; Berman, H. M., The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res* **2004,** 32, (Database issue), D223-5.

73. Larson, B. C.; Yang, W.; Ice, G. E.; Budai, J. D.; Tischler, J. Z., Three-dimensional X-ray structural microscopy with submicrometre resolution. *Nature* **2002,** 415, (6874), 887-90.

74. Kleywegt, G. J.; Jones, T. A., Model building and refinement practice. *Methods Enzymol* **1997,** 277, 208-30.

75. Jacobson, K. A.; Costanzi, S., New insights for drug design from the X-ray crystallographic structures of G-protein-coupled receptors. *Mol Pharmacol* **2012,** 82, (3), 361-71.

76. Davis, A. M.; St-Gallay, S. A.; Kleywegt, G. J., Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discov Today* **2008,** 13, (19-20), 831-41.

77. Xu, D.; Tsai, C. J.; Nussinov, R., Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng* **1997,** 10, (9), 999-1012.

78. Lipkin, H. J., Physics of Debye-Waller Factors. *Condensed Matter, Mesoscale and Nanoscale Physics* **2004**, (arXiv:cond-mat/0405023).

79. Abe, E.; Pennycook, S. J.; Tsai, A. P., Direct observation of a local thermal vibration anomaly in a quasicrystal. *Nature* **2003,** 421, (6921), 347-50.

80. Gao, H. X.; Peng, L. M., Parameterization of the temperature dependence of the Debye-Waller factors. *Acta Crystallogr A* **1999,** 55, (Pt 5), 926-932.

81. Campbell, I. D., *Biophysical Techniques*. Oxford University Press: 2012; p 130.

82. Makriyannis, A., Biegel, D., *Drug discovery strategies and methods*. CRC Press: 2003; p 10.

83. Petoukhov, M. V.; Eady, N. A.; Brown, K. A.; Svergun, D. I., Addition of missing loops and domains to protein models by X-ray solution scattering. *Biophys J* **2002,** 83, (6), 3113-25.

84. Paulsen, J. L.; Anderson, A. C., Scoring ensembles of docked protein-ligand interactions for virtual lead optimization. *J Chem Inf Model* **2009,** 49, (12), 2813-9.

85. Judd, A. S.; Souers, A. J.; Kym, P. R., Lead optimization of melanin concentrating hormone receptor 1 antagonists with low hERG channel activity. *Curr Top Med Chem* **2008,** 8, (13), 1152-7.

86. Zakharov, A. V.; Lagunin, A. A.; Filimonov, D. A.; Poroikov, V. V., Quantitative prediction of antitarget interaction profiles for chemical compounds. *Chem Res Toxicol* **2012,** 25, (11), 2378-85.

87. Galeazzi, R.; Massaccesi, L., Insight into the binding interactions of CYP450 aromatase inhibitors with their target enzyme: a combined molecular docking and molecular dynamics study. *J Mol Model* **2012,** 18, (3), 1153-66.

88. Klotz, I. M., Non-covalent bonds in protein structure. *Brookhaven Symp Biol* **1960,** 13, 25-48.

89. Baltz, J. M.; Cone, R. A., The strength of non-covalent biological bonds and adhesions by multiple independent bonds. *J Theor Biol* **1990,** 142, (2), 163-78.

90. Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., Darnell, J., *Molecular Cell Biology*. W. H. Freeman: New York, 2000; Vol. 4, p Section 2.2.

91. Guterman, L., Covalent drugs form long-lived ties. *Chemical & Engineering News* **2011,** 89, (36), 19 - 26.

92. Winter, C.; Henschel, A.; Kim, W. K.; Schroeder, M., SCOPPI: a structural classification of protein-protein interfaces. *Nucleic Acids Res* **2006,** 34, (Database issue), D310-4.

93. Franceschini, A.; Szklarczyk, D.; Frankild, S.; Kuhn, M.; Simonovic, M.; Roth, A.; Lin, J.; Minguez, P.; Bork, P.; von Mering, C.; Jensen, L. J., STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **2013,** 41, (D1), D808-15.

94. Wang, R.; Fang, X.; Lu, Y.; Wang, S., The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J Med Chem* **2004,** 47, (12), 2977-80.

95. Schreyer, A.; Blundell, T., CREDO: a protein-ligand interaction database for drug discovery. *Chem Biol Drug Des* **2009,** 73, (2), 157-67.

96. Reddy, A. S.; Amarnath, H. S.; Bapi, R. S.; Sastry, G. M.; Sastry, G. N., Protein ligand interaction database (PLID). *Comput Biol Chem* **2008,** 32, (5), 387-90.

97. Basse, M. J.; Betzi, S.; Bourgeas, R.; Bouzidi, S.; Chetrit, B.; Hamon, V.; Morelli, X.; Roche, P., 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res* **2013,** 41, (D1), D824-7.

98. Stein, A.; Russell, R. B.; Aloy, P., 3did: interacting protein domains of known three-dimensional structure. *Nucleic Acids Res* **2005,** 33, (Database issue), D413-7.

99. Finn, R. D.; Marshall, M.; Bateman, A., iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* **2005,** 21, (3), 410-2.

100. Gibbs, J. W., A method of geometrical representation of the thermodynamic properties of substances by means of surfaces. *Transactions of the Connecticut Academy of Arts and Sciences* **1873,** 2, 382-404.

101. Klebe, G., *Wirkstoffdesign*. Spektrum Akademischer Verlag: 2009; Vol. 2, p 49-66.

102. Zhou, H. X.; Gilson, M. K., Theory of free energy and entropy in noncovalent binding. *Chem Rev* **2009,** 109, (9), 4092-107.

103. Yu, H.; Rick, S. W., Free energy, entropy, and enthalpy of a water molecule in various protein environments. *J Phys Chem B* **2010,** 114, (35), 11552-60.

104. Oshima, H.; Yasuda, S.; Yoshidome, T.; Ikeguchi, M.; Kinoshita, M., Crucial importance of the water-entropy effect in predicting hot spots in protein-protein complexes. *Phys Chem Chem Phys* **2011,** 13, (36), 16236-46.

105. Sheu, S. Y.; Yang, D. Y., Determination of protein surface hydration shell free energy of water motion: theoretical study and molecular dynamics simulation. *J Phys Chem B* **2010,** 114, (49), 16558-66.

106. Kolar, M.; Fanfrlik, J.; Hobza, P., Ligand conformational and solvation/desolvation free energy in protein-ligand complex formation. *J Phys Chem B* **2011,** 115, (16), 4718-24.

107. Camacho, C. J.; Kimura, S. R.; DeLisi, C.; Vajda, S., Kinetics of desolvation-mediated protein-protein binding. *Biophys J* **2000,** 78, (3), 1094-105.

108. Shimokhina, N.; Bronowska, A.; Homans, S. W., Contribution of ligand desolvation to binding thermodynamics in a ligand-protein interaction. *Angew Chem Int Ed Engl* **2006,** 45, (38), 6374-6.

109. Connelly, P. R.; Aldape, R. A.; Bruzzese, F. J.; Chambers, S. P.; Fitzgibbon, M. J.; Fleming, M. A.; Itoh, S.; Livingston, D. J.; Navia, M. A.; Thomson, J. A.; et al., Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. *Proc Natl Acad Sci U S A* **1994,** 91, (5), 1964-8.

110. Thangudu, R. R.; Bryant, S. H.; Panchenko, A. R.; Madej, T., Modulating protein-protein interactions with small molecules: the importance of binding hotspots. *J Mol Biol* **2012,** 415, (2), 443-53.

111. Brinda, K. V.; Vishveshwara, S., Oligomeric protein structure networks: insights into protein-protein interactions. *BMC Bioinformatics* **2005,** 6, 296.

112. 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, (30), 4630-47.

113. Gallicchio, E., Kubo, M. M., Levy, R. M., Entropy-enthalpy compensation in solvation and ligand binding revisited. *J. Am. Chem. Soc.* **1998,** 120, 4526-4527.

114. Lumry, R.; Rajender, S., Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers* **1970,** 9, (10), 1125-227.

115. Olsson, T. S.; Ladbury, J. E.; Pitt, W. R.; Williams, M. A., Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein Sci* **2011,** 20, (9), 1607-18.

116. Cheng, T.; Liu, Z.; Wang, R., A knowledge-guided strategy for improving the accuracy of scoring functions in binding affinity prediction. *BMC Bioinformatics* **2010,** 11, 193.

117. Yang, C. Y.; Sun, H.; Chen, J.; Nikolovska-Coleska, Z.; Wang, S., Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction. *J Am Chem Soc* **2009,** 131, (38), 13709-21.

118. Nooren, I. M.; Thornton, J. M., Diversity of protein-protein interactions. *Embo J* **2003,** 22, (14), 3486-92.

119. Zhu, H.; Sommer, I.; Lengauer, T.; Domingues, F. S., Alignment of non-covalent interactions at protein-protein interfaces. *PLoS One* **2008,** 3, (4), e1926.

120. Lundberg, S.; Backman, L., Protein-protein and protein-ligand interactions. *Methods Enzymol* **1994,** 228, 241-54.

121. Kinjo, A. R.; Nakamura, H., Functional structural motifs for protein-ligand, protein-protein, and protein-nucleic acid interactions and their connection to supersecondary structures. *Methods Mol Biol* **2013,** 932, 295-315.

122. Burgoyne, N. J.; Jackson, R. M., Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* **2006,** 22, (11), 1335-42.

123. Mandell, J. G.; Baerga-Ortiz, A.; Falick, A. M.; Komives, E. A., Measurement of solvent accessibility at protein-protein interfaces. *Methods Mol Biol* **2005,** 305, 65-80.

124. Krüger, D. M.; Jessen, G.; Gohlke, H., 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), 2807-11.

125. Morelli, X.; Bourgeas, R.; Roche, P., Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol* **2011,** 15, (4), 475-81.

126. Gestwicki, J. E.; Marinec, P. S., Chemical control over protein-protein interactions: beyond inhibitors. *Comb Chem High Throughput Screen* **2007,** 10, (8), 667-75.

127. Cochran, A. G., Protein-protein interfaces: mimics and inhibitors. *Curr Opin Chem Biol* **2001,** 5, (6), 654-9.

128. Bienstock, R. J., Computational drug design targeting protein-protein interactions. *Curr Pharm Des* **2012,** 18, (9), 1240-54.

129. Ortiz, A. R.; Gomez-Puertas, P.; Leo-Macias, A.; Lopez-Romero, P.; Lopez-Vinas, E.; Morreale, A.; Murcia, M.; Wang, K., Computational approaches to model ligand selectivity in drug design. *Curr Top Med Chem* **2006,** 6, (1), 41-55.

130. de Jong, L. A.; Uges, D. R.; Franke, J. P.; Bischoff, R., Receptor-ligand binding assays: technologies and applications. *J Chromatogr B Analyt Technol Biomed Life Sci* **2005,** 829, (1-2), 1-25.

131. Meyer-Almes, F. J.; Auer, M., Enzyme inhibition assays using fluorescence correlation spectroscopy: a new algorithm for the derivation of kcat/KM and Ki values at substrate concentrations much lower than the Michaelis constant. *Biochemistry* **2000,** 39, (43), 13261-8.

132. Cheng, Y.-C.; Prusoff, W. H., Relationship between the inhibition constant (K_I) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochemical Pharmacology* **1973,** 22, (23), 3099–3108.

133. Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A., Calculation of the relative change in binding free energy of a protein-inhibitor complex. *Science* **1987,** 235, (4788), 574-6.

134. Kuhara, S.; Ezaki, E.; Fukamizo, T.; Hayashi, K., Estimation of the free energy change of substrate binding lysozyme-catalyzed reactions. *J Biochem* **1982,** 92, (1), 121-7.

135. Krüger, D. M.; Gohlke, H., DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. *Nucleic Acids Res* **2010,** 38, (Web Server issue), W480-6.

136. Badrinarayan, P.; Sastry, G. N., Rational approaches towards lead optimization of kinase inhibitors: The issue of specificity. *Curr Pharm Des* **2013,** 19, (26), 4714-38.

137. Tintelnot, M.; Andrews, P., Geometries of functional group interactions in enzyme-ligand complexes: guides for receptor modelling. *J Comput Aided Mol Des* **1989,** 3, (1), 67-84.

138. Paulsen, J. L.; Anderson, A. C., Scoring ensembles of docked protein:ligand interactions for virtual lead optimization. *J Chem Inf Model* **2009,** 49, (12), 2813-9.

139. Xiang, M.; Cao, Y.; Fan, W.; Chen, L.; Mo, Y., Computer-aided drug design: lead discovery and optimization. *Comb Chem High Throughput Screen* **2012,** 15, (4), 328-37.

140. Zhang, S., In silico lead identification and optimization for drug discovery. *Curr Pharm Des* **2012,** 18, (9), 1171-2.

141. Kazemi, S.; Krüger, D. M.; Sirockin, F.; Gohlke, H., Elastic potential grids: accurate and efficient representation of intermolecular interactions for fully flexible docking. *ChemMedChem* **2009,** 4, (8), 1264-8.

142. Krüger, D. M., Bergs, J. , Kazemi, S., Gohlke, H., Target flexibility in RNA-ligand docking modeled by elastic potential grids. *ACS Med. Chem. Lett.* **2011,** 2, (7), 489-493.

143. Kim, R.; Skolnick, J., Assessment of programs for ligand binding affinity prediction. *J Comput Chem* **2008,** 29, (8), 1316-31.

144. Moal, I. H.; Agius, R.; Bates, P. A., Protein-protein binding affinity prediction on a diverse set of structures. *Bioinformatics* **2011,** 27, (21), 3002-9.

145. Jorgensen, W. L., The many roles of computation in drug discovery. *Science* **2004,** 303, (5665), 1813-8.

146. Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., 3rd, Assessing scoring functions for protein-ligand interactions. *J Med Chem* **2004,** 47, (12), 3032-47.

147. 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 Engl* **2002,** 41, (15), 2644-76.

148. Feliu, E.; Oliva, B., How different from random are docking predictions when ranked by scoring functions? *Proteins* **2010,** 78, (16), 3376-85.

149. Huang, S. Y.; Grinter, S. Z.; Zou, X., Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys* **2010,** 12, (40), 12899-908.

150. Garzon, J. I.; Lopez-Blanco, J. R.; Pons, C.; Kovacs, J.; Abagyan, R.; Fernandez-Recio, J.; Chacon, P., FRODOCK: a new approach for fast rotational protein-protein docking. *Bioinformatics* **2009,** 25, (19), 2544-51.

151. Pfeffer, P.; Gohlke, H., DrugScoreRNA - knowledge-based scoring function to predict RNA-ligand interactions. *J Chem Inf Model* **2007,** 47, (5), 1868-76.

152. Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **2000,** 295, (2), 337-56.

153. Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J., Lead discovery using molecular docking. *Curr Opin Chem Biol* **2002,** 6, (4), 439-46.

154. Sotriffer, C. A.; Gohlke, H.; Klebe, G., Docking into knowledge-based potential fields: a comparative evaluation of DrugScore. *J Med Chem* **2002,** 45, (10), 1967-70.

155. Roberts, B. C.; Mancera, R. L., Ligand-protein docking with water molecules. *J Chem Inf Model* **2008,** 48, (2), 397-408.

156. Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R., Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem* **2005,** 48, (20), 6504-15.

157. Kazemi, S.; Krüger, D. M.; Sirockin, F.; Gohlke, H., Elastic potential grids: accurate and efficient representation of intermolecular interactions for fully flexible docking. *ChemMedChem* **2009,** 4, (8), 1264-8.

158. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E., A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **1982,** 161, (2), 269-88.

159. Goodsell, D. S.; Olson, A. J., Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990,** 8, (3), 195-202.

160. Totrov, M.; Abagyan, R., Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins* **1997,** Suppl 1, 215-20.

161. Trosset, J. Y.; Scheraga, H. A., Reaching the global minimum in docking simulations: a Monte Carlo energy minimization approach using Bezier splines. *Proc Natl Acad Sci U S A* **1998,** 95, (14), 8011-5.

162. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** 47, (7), 1739-49.

163. Morris, G. M.; Huey, R.; Olson, A. J., Using AutoDock for ligand-receptor docking. *Curr Protoc Bioinformatics* **2008,** Chapter 8, Unit 8 14.

164. Goodsell, D. S., Computational docking of biomolecular complexes with AutoDock. *Cold Spring Harb Protoc* **2009,** 2009, (5), pdb.prot5200.

165. Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Computational Chemistry* **1998,** 19, 1639-1662.

166. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S., A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* **2007,** 28, (6), 1145-52.

167. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010,** 31, (2), 455-61.

168. Heaslet, H.; Rosenfeld, R.; Giffin, M.; Lin, Y. C.; Tam, K.; Torbett, B. E.; Elder, J. H.; McRee, D. E.; Stout, C. D., Conformational flexibility in the flap domains of ligand-free HIV protease. *Acta Crystallogr D Biol Crystallogr* **2007,** 63, (Pt 8), 866-75.

169. Hornak, V.; Simmerling, C., Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discov Today* **2007,** 12, (3-4), 132-8.

170. Cheng, J. W.; Lepre, C. A.; Moore, J. M., 15N NMR relaxation studies of the FK506 binding protein: dynamic effects of ligand binding and implications for calcineurin recognition. *Biochemistry* **1994,** 33, (14), 4093-100.

171. May, A.; Zacharias, M., Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim Biophys Acta* **2005,** 1754, (1-2), 225-31.

172. Powell, N. A.; Clay, E. H.; Holsworth, D. D.; Bryant, J. W.; Ryan, M. J.; Jalaie, M.; Edmunds, J. J., Benzyl ether structure-activity relationships in a series of ketopiperazine-based renin inhibitors. *Bioorg Med Chem Lett* **2005,** 15, (21), 4713-6.

173. Benkoulouche, M.; Cotrait, M.; Maigret, B., Computer simulation of the conformational behaviour of angiotensinogen (6-13) renin substrate. *J Comput Aided Mol Des* **1992,** 6, (1), 79-91.

174. Bowman, A. L.; Lerner, M. G.; Carlson, H. A., Protein flexibility and species specificity in structure-based drug discovery: dihydrofolate reductase as a test system. *J Am Chem Soc* **2007,** 129, (12), 3634-40.

175. McElheny, D.; Schnell, J. R.; Lansing, J. C.; Dyson, H. J.; Wright, P. E., Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc Natl Acad Sci U S A* **2005,** 102, (14), 5032-7.

176. Schnell, J. R.; Dyson, H. J.; Wright, P. E., Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct* **2004,** 33, 119-40.

177. Barnard, D.; Diaz, B.; Hettich, L.; Chuang, E.; Zhang, X. F.; Avruch, J.; Marshall, M., Identification of the sites of interaction between c-Raf-1 and Ras-GTP. *Oncogene* **1995,** 10, (7), 1283-90.

178. 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.

179. Murray, C. W.; Baxter, C. A.; Frenkel, A. D., The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J Comput Aided Mol Des* **1999,** 13, (6), 547-62.

180. 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), 2214-25.

181. Schnecke, V.; Swanson, C. A.; Getzoff, E. D.; Tainer, J. A.; Kuhn, L. A., Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins* **1998,** 33, (1), 74-87.

182. Mizutani, M. Y.; Takamatsu, Y.; Ichinose, T.; Nakamura, K.; Itai, A., Effective handling of induced-fit motion in flexible docking. *Proteins* **2006,** 63, (4), 878-91.

183. Kallblad, P.; Dean, P. M., Efficient conformational sampling of local side-chain flexibility. *J Mol Biol* **2003,** 326, (5), 1651-65.

184. Zhao, Y.; Sanner, M. F., Protein-ligand docking with multiple flexible side chains. *J Comput Aided Mol Des* **2008,** 22, (9), 673-9.

185. Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M., Side-chain flexibility in proteins upon ligand binding. *Proteins* **2000,** 39, (3), 261-8.

186. Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A., Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc* **2005,** 127, (26), 9632-40.

187. Huang, S. Y.; Zou, X., Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007,** 66, (2), 399-421.

188. Nabuurs, S. B.; Wagener, M.; de Vlieg, J., A flexible approach to induced fit docking. *J Med Chem* **2007,** 50, (26), 6507-18.

189. Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T., FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol* **2001,** 308, (2), 377-95.

190. Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M., Molecular docking to ensembles of protein structures. *J Mol Biol* **1997,** 266, (2), 424-40.

191. Bahar, I.; Rader, A. J., Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol* **2005,** 15, (5), 586-92.

192. Mangoni, M.; Roccatano, D.; Di Nola, A., Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Proteins* **1999,** 35, (2), 153-62.

193. Andrusier, N.; Nussinov, R.; Wolfson, H. J., FireDock: fast interaction refinement in molecular docking. *Proteins* **2007,** 69, (1), 139-59.

194. Tatsumi, R.; Fukunishi, Y.; Nakamura, H., A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor. *J Comput Chem* **2004,** 25, (16), 1995-2005.

195. 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), 2607-25.

196. Jones, S.; Thornton, J. M., Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **1996,** 93, (1), 13-20.

197. Lo Conte, L.; Chothia, C.; Janin, J., The atomic structure of protein-protein recognition sites. *J Mol Biol* **1999,** 285, (5), 2177-98.

198. Stumpf, M. P.; Thorne, T.; de Silva, E.; Stewart, R.; An, H. J.; Lappe, M.; Wiuf, C., Estimating the size of the human interactome. *Proc Natl Acad Sci U S A* **2008,** 105, (19), 6959-64.

199. Kikugawa, S.; Nishikata, K.; Murakami, K.; Sato, Y.; Suzuki, M.; Altaf-Ul-Amin, M.; Kanaya, S.; Imanishi, T., PCDq: human protein complex database with quality index which summarizes different levels of evidences of protein complexes predicted from H-Invitational protein-protein interactions integrative dataset. *BMC Syst Biol* **2012,** 6 Suppl 2, S7.

200. Smith, G. R.; Sternberg, M. J., Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol* **2002,** 12, (1), 28-35.

201. Ritchie, D. W., Recent progress and future directions in protein-protein docking. *Curr Protein Pept Sci* **2008,** 9, (1), 1-15.

202. Schneidman-Duhovny, D.; Nussinov, R.; Wolfson, H. J., Predicting molecular interactions in silico: II. Protein-protein and protein-drug docking. *Curr Med Chem* **2004,** 11, (1), 91-107.

203. Jiang, F.; Lin, W.; Rao, Z., SOFTDOCK: understanding of molecular recognition through a systematic docking study. *Protein Eng* **2002,** 15, (4), 257-63.

204. Palma, P. N.; Krippahl, L.; Wampler, J. E.; Moura, J. J., BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins* **2000,** 39, (4), 372-84.

205. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A., Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A* **1992,** 89, (6), 2195-9.

206. Gabb, H. A.; Jackson, R. M.; Sternberg, M. J., Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* **1997,** 272, (1), 106-20.

207. Mandell, J. G.; Roberts, V. A.; Pique, M. E.; Kotlovyi, V.; Mitchell, J. C.; Nelson, E.; Tsigelny, I.; Ten Eyck, L. F., Protein docking using continuum electrostatics and geometric fit. *Protein Eng* **2001,** 14, (2), 105-13.

208. Chen, R.; Li, L.; Weng, Z., ZDOCK: an initial-stage protein-docking algorithm. *Proteins* **2003,** 52, (1), 80-7.

209. Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S., PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* **2006,** 65, (2), 392-406.

210. Shoichet, B. K.; Kuntz, I. D., Predicting the structure of protein complexes: a step in the right direction. *Chem Biol* **1996,** 3, (3), 151-6.

211. Fischer, D.; Lin, S. L.; Wolfson, H. L.; Nussinov, R., A geometry-based suite of molecular docking processes. *J Mol Biol* **1995,** 248, (2), 459-77.

212. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J., Geometry-based flexible and symmetric protein docking. *Proteins* **2005,** 60, (2), 224-31.

213. Bordner, A. J.; Gorin, A. A., Protein docking using surface matching and supervised machine learning. *Proteins* **2007,** 68, (2), 488-502.

214. Gardiner, E. J.; Willett, P.; Artymiuk, P. J., GAPDOCK: a Genetic Algorithm Approach to Protein Docking in CAPRI round 1. *Proteins* **2003,** 52, (1), 10-4.

215. Taylor, J. S.; Burnett, R. M., DARWIN: a program for docking flexible molecules. *Proteins* **2000,** 41, (2), 173-91.

216. Fernandez-Recio, J.; Totrov, M.; Abagyan, R., ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* **2003,** 52, (1), 113-7.

217. Zacharias, M., ATTRACT: protein-protein docking in CAPRI using a reduced protein model. *Proteins* **2005,** 60, (2), 252-6.

218. Dominguez, C.; Boelens, R.; Bonvin, A. M., HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **2003,** 125, (7), 1731-7.

219. Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D., Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* **2003,** 331, (1), 281-99.

220. Grosdidier, S.; Fernandez-Recio, J., Protein-protein docking and hot-spot prediction for drug discovery. *Curr Pharm Des* **2012,** 18, (30), 4607-18.

221. Wiehe, K.; Peterson, M. W.; Pierce, B.; Mintseris, J.; Weng, Z., Protein-protein docking: overview and performance analysis. *Methods Mol Biol* **2008,** 413, 283-314.

222. Ritchie, D. W.; Kozakov, D.; Vajda, S., Accelerating and focusing protein-protein docking correlations using multi-dimensional rotational FFT generating functions. *Bioinformatics* **2008,** 24, (17), 1865-73.

223. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J., Protein-protein docking dealing with the unknown. *J Comput Chem* **2010,** 31, (2), 317-42.

224. Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B., Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* **2003,** 68, (1), 91-109.

225. Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K., Challenges in protein folding simulations: Timescale, representation, and analysis. *Nat Phys* **2010,** 6, (10), 751-758.

226. Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R., Folding funnels and binding mechanisms. *Protein Eng* **1999,** 12, (9), 713-20.

227. Mendez, R.; Leplae, R.; De Maria, L.; Wodak, S. J., Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **2003,** 52, (1), 51-67.

228. Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J.; Vajda, S.; Vakser, I.; Wodak, S. J., CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* **2003,** 52, (1), 2-9.

229. Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R., Comparative assessment of scoring functions on a diverse test set. *J Chem Inf Model* **2009,** 49, (4), 1079-93.

230. Wang, R.; Lu, Y.; Wang, S., Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* **2003,** 46, (12), 2287-303.

231. Zhao, X.; Liu, X.; Wang, Y.; Chen, Z.; Kang, L.; Zhang, H.; Luo, X.; Zhu, W.; Chen, K.; Li, H.; Wang, X.; Jiang, H., An improved PMF scoring function for universally predicting the interactions of a ligand with protein, DNA, and RNA. *J Chem Inf Model* **2008,** 48, (7), 1438-47.

232. Seifert, M. H., Targeted scoring functions for virtual screening. *Drug Discov Today* **2009,** 14, (11-12), 562-9.

233. Fischer, B.; Fukuzawa, K.; Wenzel, W., Receptor-specific scoring functions derived from quantum chemical models improve affinity estimates for in-silico drug discovery. *Proteins* **2008,** 70, (4), 1264-73.

234. Knox, A. J.; Meegan, M. J.; Sobolev, V.; Frost, D.; Zisterer, D. M.; Williams, D. C.; Lloyd, D. G., Target specific virtual screening: optimization of an estrogen receptor screening platform. *J Med Chem* **2007,** 50, (22), 5301-10.

235. Breu, B.; Silber, K.; Gohlke, H., Consensus adaptation of fields for molecular comparison (AFMoC) models incorporate ligand and receptor conformational variability into tailor-made scoring functions. *J Chem Inf Model* **2007,** 47, (6), 2383-400.

236. Gozalbes, R.; Simon, L.; Froloff, N.; Sartori, E.; Monteils, C.; Baudelle, R., Development and experimental validation of a docking strategy for the generation of kinase-targeted libraries. *J Med Chem* **2008,** 51, (11), 3124-32.

237. Seifert, M. H., Optimizing the signal-to-noise ratio of scoring functions for protein--ligand docking. *J Chem Inf Model* **2008,** 48, (3), 602-12.

238. de Graaf, C.; Rognan, D., Selective structure-based virtual screening for full and partial agonists of the beta2 adrenergic receptor. *J Med Chem* **2008,** 51, (16), 4978-85.

239. Kerzmann, A.; Fuhrmann, J.; Kohlbacher, O.; Neumann, D., BALLDock/SLICK: a new method for protein-carbohydrate docking. *J Chem Inf Model* **2008,** 48, (8), 1616-25.

240. Hetenyi, C.; Paragi, G.; Maran, U.; Timar, Z.; Karelson, M.; Penke, B., Combination of a modified scoring function with two-dimensional descriptors for calculation of binding affinities of bulky, flexible ligands to proteins. *J Am Chem Soc* **2006,** 128, (4), 1233-9.

241. Jain, A. N., Scoring functions for protein-ligand docking. *Curr Protein Pept Sci* **2006,** 7, (5), 407-20.

242. Wang, J. C.; Lin, J. H., Scoring functions for prediction of protein-ligand interactions. *Curr Pharm Des* **2013**.

243. Liu, S.; Vakser, I. A., DECK: Distance and environment-dependent, coarse-grained, knowledge-based potentials for protein-protein docking. *BMC Bioinformatics* **2011,** 12, 280.

244. Rajgaria, R.; McAllister, S. R.; Floudas, C. A., A novel high resolution Calpha - Calpha distance dependent force field based on a high quality decoy set. *Proteins* **2006,** 65, (3), 726-41.

245. Chuang, G. Y.; Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S., DARS (Decoys As the Reference State) potentials for protein-protein docking. *Biophys J* **2008,** 95, (9), 4217-27.

246. Tovchigrechko, A.; Vakser, I. A., Development and testing of an automated approach to protein docking. *Proteins* **2005,** 60, (2), 296-301.

247. Lyskov, S.; Gray, J. J., The RosettaDock server for local protein-protein docking. *Nucleic Acids Res* **2008,** 36, (Web Server issue), W233-8.

248. Zacharias, M., Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci* **2003,** 12, (6), 1271-82.

249. Sippl, M. J., Knowledge-based potentials for proteins. *Curr Opin Struct Biol* **1995,** 5, (2), 229-35.

250. Massova, I., Kollman, P. A., Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000,** 18, 113-135.

251. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* **1997,** 11, (5), 425-45.

252. Prathipati, P.; Saxena, A. K., Evaluation of binary QSAR models derived from LUDI and MOE scoring functions for structure based virtual screening. *J Chem Inf Model* **2006,** 46, (1), 39-51.

253. Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D., Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* **1998,** 33, (3), 367-82.

254. Sippl, M. J., Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J Mol Biol* **1990,** 213, (4), 859-83.

255. De Genst, E.; Areskoug, D.; Decanniere, K.; Muyldermans, S.; Andersson, K., Kinetic and affinity predictions of a protein-protein interaction using multivariate experimental design. *J Biol Chem* **2002,** 277, (33), 29897-907.

256. Ozrin, V. D.; Subbotin, M. V.; Nikitin, S. M., PLASS: protein-ligand affinity statistical score - a knowledge-based force-field model of interaction derived from the PDB. *J Comput Aided Mol Des* **2004,** 18, (4), 261-70.

257. Huang, S. Y.; Zou, X., An iterative knowledge-based scoring function for protein-protein recognition. *Proteins* **2008,** 72, (2), 557-79.

258. Thorn, K. S.; Bogan, A. A., ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* **2001,** 17, (3), 284-5.

259. Kiel, C.; Serrano, L.; Herrmann, C., A detailed thermodynamic analysis of ras/effector complex interfaces. *J Mol Biol* **2004,** 340, (5), 1039-58.

260. 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, (Web Server issue), W382-8.

261. 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-RalGDS complexes. *J Mol Biol* **2003,** 330, (4), 891-913.

262. Benedix, A.; Becker, C. M.; de Groot, B. L.; Caflisch, A.; Bockmann, R. A., Predicting free energy changes using structural ensembles. *Nat Methods* **2009,** 6, (1), 3-4.

263. Kim, D. E.; Chivian, D.; Baker, D., Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **2004,** 32, (Web Server issue), W526-31.

264. Muthu, K.; Panneerselvam, M.; Jayaraman, M.; Topno, N. S.; Das, A. A.; Ramadas, K., Structural insights into interacting mechanism of ID1 protein with an antagonist ID1/3-PA7 and agonist ETS-1 in treatment of ovarian cancer: molecular docking and dynamics studies. *J Mol Model* **2012,** 18, (11), 4865-84.

265. Romano, V.; Raimondo, D.; Calvanese, L.; D'Auria, G.; Tramontano, A.; Falcigno, L., Toward a better understanding of the interaction between TGF-beta family members and their ALK receptors. *J Mol Model* **2012,** 18, (8), 3617-25.

266. Thakor, H.; Nicholas, S.; Porter, I. M.; Hand, N.; Stewart, R. C., Identification of an anchor residue for CheA-CheY interactions in the chemotaxis system of Escherichia coli. *J Bacteriol* **2011,** 193, (15), 3894-903.

267. Epa, V. C.; Dolezal, O.; Doughty, L.; Xiao, X.; Jost, C.; Pluckthun, A.; Adams, T. E., Structural model for the interaction of a designed Ankyrin Repeat Protein with the human epidermal growth factor receptor 2. *PLoS One* **2013,** 8, (3), e59163.

268. Ciglia, E., Vergin, J., Reimann, S., Smits, S.H.J., Schmitt, L., Groth, G., Gohlke, H., Resolving hot spots in the C-terminal dimerization domain that determine the stability of the molecular chaperone Hsp90. *PLoS One* **2014,** *in press*.

269. Pfeffer, P.; Gohlke, H., DrugScoreRNA--knowledge-based scoring function to predict RNA-ligand interactions. *J Chem Inf Model* **2007,** 47, (5), 1868-76.

270. Hwang, H.; Pierce, B.; Mintseris, J.; Janin, J.; Weng, Z., Protein-protein docking benchmark version 3.0. *Proteins* **2008,** 73, (3), 705-9.

271. Danley, D. E., Crystallization to obtain protein-ligand complexes for structure-aided drug design. *Acta Crystallogr D Biol Crystallogr* **2006,** 62, (Pt 6), 569-75.

272. Hinsen, K., Structural flexibility in proteins: impact of the crystal environment. *Bioinformatics* **2008,** 24, (4), 521-8.

273. Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R., A new test set for validating predictions of protein-ligand interaction. *Proteins* **2002,** 49, (4), 457-71.

274. Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J., A dissection of specific and non-specific protein-protein interfaces. *J Mol Biol* **2004,** 336, (4), 943-55.

275. Mian, I. S.; Bradwell, A. R.; Olson, A. J., Structure, function and properties of antibody binding sites. *J Mol Biol* **1991,** 217, (1), 133-51.

276. Jackson, R. M., Comparison of protein-protein interactions in serine protease-inhibitor and antibody-antigen complexes: implications for the protein docking problem. *Protein Sci* **1999,** 8, (3), 603-13.

277. Sundberg, E. J.; Mariuzza, R. A., Molecular recognition in antibody-antigen complexes. *Adv Protein Chem* **2002,** 61, 119-60.

278. Betts, M. J.; Sternberg, M. J., An analysis of conformational changes on protein-protein association: implications for predictive docking. *Protein Eng* **1999,** 12, (4), 271-83.

279. Goh, C. S.; Milburn, D.; Gerstein, M., Conformational changes associated with protein-protein interactions. *Curr Opin Struct Biol* **2004,** 14, (1), 104-9.

280. Ubbink, M., The courtship of proteins: understanding the encounter complex. *FEBS Lett* **2009,** 583, (7), 1060-6.

281. Fawzi, N. L.; Doucleff, M.; Suh, J. Y.; Clore, G. M., Mechanistic details of a protein-protein association pathway revealed by paramagnetic relaxation enhancement titration measurements. *Proc Natl Acad Sci U S A* **2010,** 107, (4), 1379-84.

282. Lalonde, S.; Ehrhardt, D. W.; Loque, D.; Chen, J.; Rhee, S. Y.; Frommer, W. B., Molecular and cellular approaches for the detection of protein-protein interactions: latest techniques and current limitations. *Plant J* **2008,** 53, (4), 610-35.

283. Pazos, F.; Helmer-Citterich, M.; Ausiello, G.; Valencia, A., Correlated mutations contain information about protein-protein interaction. *J Mol Biol* **1997,** 271, (4), 511-23.

284. Pierce, B.; Weng, Z., ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins* **2007,** 67, (4), 1078-86.

285. Pierce, B.; Weng, Z., A combination of rescoring and refinement significantly improves protein docking performance. *Proteins* **2008,** 72, (1), 270-9.

286. Mintseris, J.; Pierce, B.; Wiehe, K.; Anderson, R.; Chen, R.; Weng, Z., Integrating statistical pair potentials into protein complex prediction. *Proteins* **2007,** 69, (3), 511-20.

287. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.; Luo, R. C. W., W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg,; G. Seabra, I. K., K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell,; T. Steinbrecher, H. G., Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H.; Mathews, M. G. S., C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and; (2010), P. A. K. *AMBER 11*, University of California, San Francisco, 2010.

288. Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr., Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009,** 77, (4), 778-95.

289. Connolly, M. L., Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983,** 221, (4612), 709-13.

290. Liu, S.; Zhang, C.; Zhou, H.; Zhou, Y., A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. *Proteins* **2004,** 56, (1), 93-101.

291. Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M., PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.* **1993,** 26, 283-291.

292. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010,** 66, (Pt 1), 12-21.

293. Neuvirth, H.; Raz, R.; Schreiber, G., ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol* **2004,** 338, (1), 181-99.

294. Ryckaert, J. P., Ciccotti, G., Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977,** 23, 327−341.

295. de Groot, B. L.; van Aalten, D. M.; Scheek, R. M.; Amadei, A.; Vriend, G.; Berendsen, H. J., Prediction of protein conformational freedom from distance constraints. *Proteins* **1997,** 29, (2), 240-51.

296. Farrell, D. W.; Speranskiy, K.; Thorpe, M. F., Generating stereochemically acceptable protein pathways. *Proteins* **2010,** 78, (14), 2908-21.

297. 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, (7), 1604-22.

298. Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F., Protein flexibility predictions using graph theory. *Proteins* **2001,** 44, (2), 150-65.

299. Krüger, D. M.; Ahmed, A.; Gohlke, H., NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins. *Nucleic Acids Res* **2012,** 40, (Web Server issue), W310-6.

300. Korb, O.; Stutzle, T.; Exner, T. E., Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model* **2009,** 49, (1), 84-96.

301. Fuller, J. C.; Burgoyne, N. J.; Jackson, R. M., Predicting druggable binding sites at the protein-protein interface. *Drug Discov Today* **2009,** 14, (3-4), 155-61.

302. Abad-Zapatero, C.; Metz, J. T., Ligand efficiency indices as guideposts for drug discovery. *Drug Discov Today* **2005,** 10, (7), 464-9.

303. Morelli, X.; Bourgeas, R.; Roche, P., Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol* 15, (4), 475-81.

304. Borhani, D. W.; Shaw, D. E., The future of molecular dynamics simulations in drug discovery. *J Comput Aided Mol Des* **2011,** 26, (1), 15-26.

305. Schlick, T., Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules. *F1000 Biol Rep* **2009,** 1, 51.

306. Zwier, M. C.; Chong, L. T., Reaching biological timescales with all-atom molecular dynamics simulations. *Curr Opin Pharmacol* **2010,** 10, (6), 745-52.

307. Pierce, L. C.; Salomon-Ferrer, R.; Augusto, F. d. O. C.; McCammon, J. A.; Walker, R. C., Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J Chem Theory Comput* **2012,** 8, (9), 2997-3002.

308. Nadler, W.; Hansmann, U. H., Generalized ensemble and tempering simulations: a unified view. *Phys Rev E Stat Nonlin Soft Matter Phys* **2007,** 75, (2 Pt 2), 026109.

309. Piana, S.; Lindorff-Larsen, K.; Shaw, D. E., Atomic-level description of ubiquitin folding. *Proc Natl Acad Sci U S A* **2013,** 110, (15), 5915-20.

310. Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E., How does a drug molecule find its target binding site? *J Am Chem Soc* **2011,** 133, (24), 9181-3.

311. Kiefhaber, T.; Bachmann, A.; Jensen, K. S., Dynamics and mechanisms of coupled protein folding and binding reactions. *Curr Opin Struct Biol* **2012,** 22, (1), 21-9.

312. Rogers, J. M.; Wong, C. T.; Clarke, J., Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *J Am Chem Soc* **2014,** 136, (14), 5197-200.

313. Hofinger, S.; Almeida, B.; Hansmann, U. H., Parallel tempering molecular dynamics folding simulation of a signal peptide in explicit water. *Proteins* **2007,** 68, (3), 662-9.

314. Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C., Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J Chem Phys* **2007,** 126, (15), 155101.

315. Prinz, J. H.; Chodera, J. D.; Pande, V. S.; Swope, W. C.; Smith, J. C.; Noe, F., Optimal use of data in parallel tempering simulations for the construction of discrete-state Markov models of biomolecular dynamics. *J Chem Phys* **2011,** 134, (24), 244108.

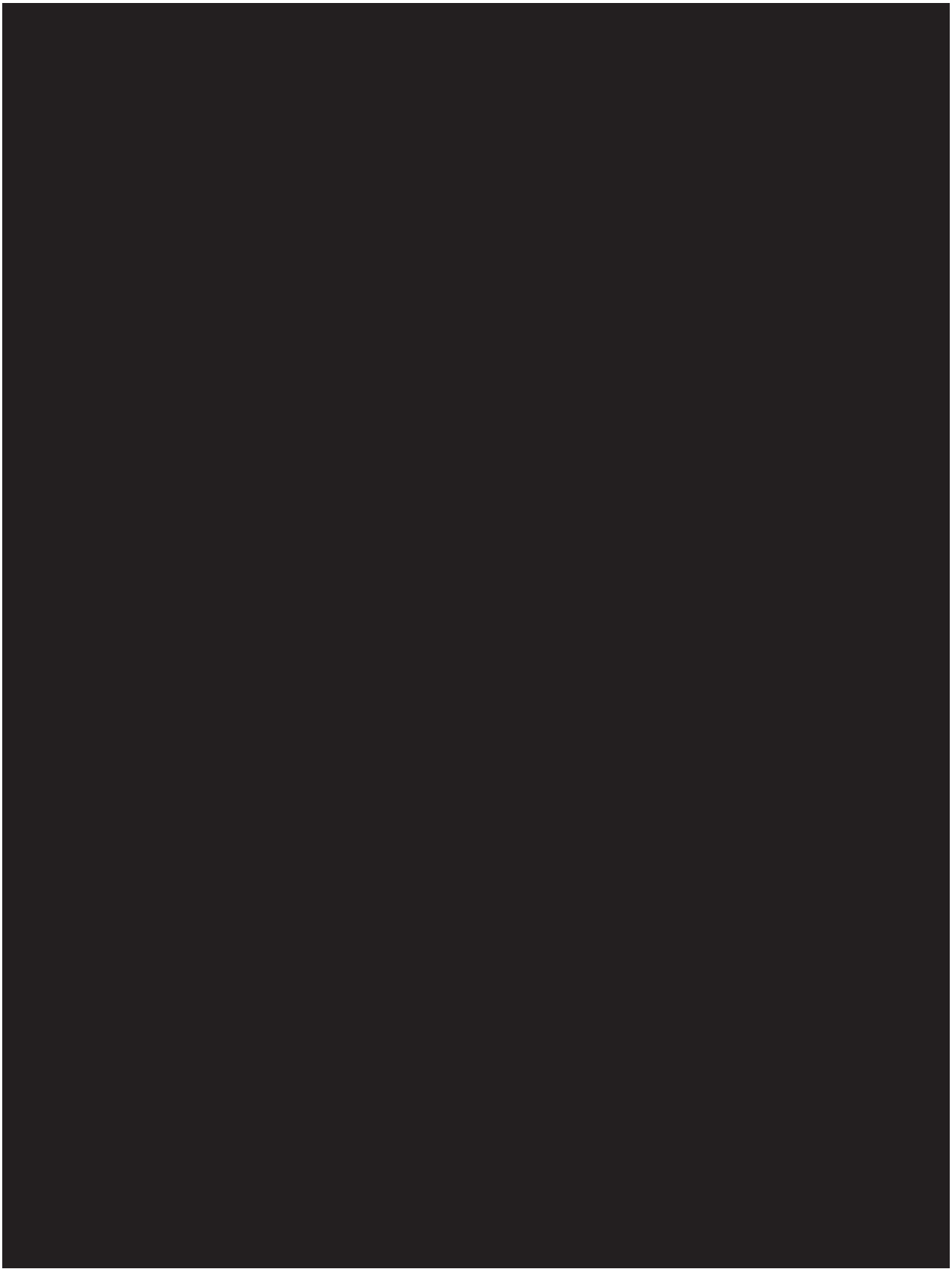
316. Noe, F.; Fischer, S., Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr Opin Struct Biol* **2008,** 18, (2), 154-62.

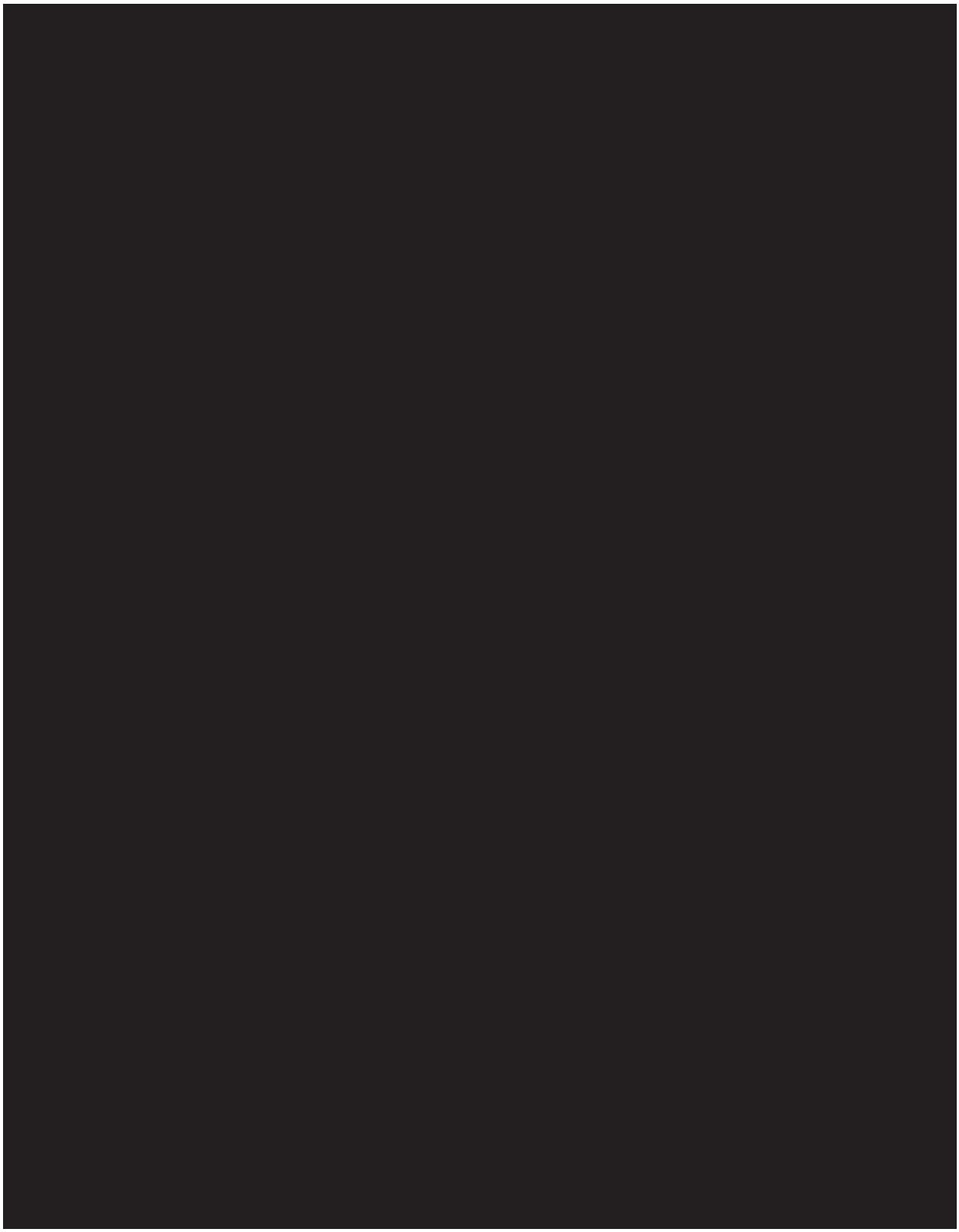
All figures in this work were generated with PyMOL version 0.99rc6.

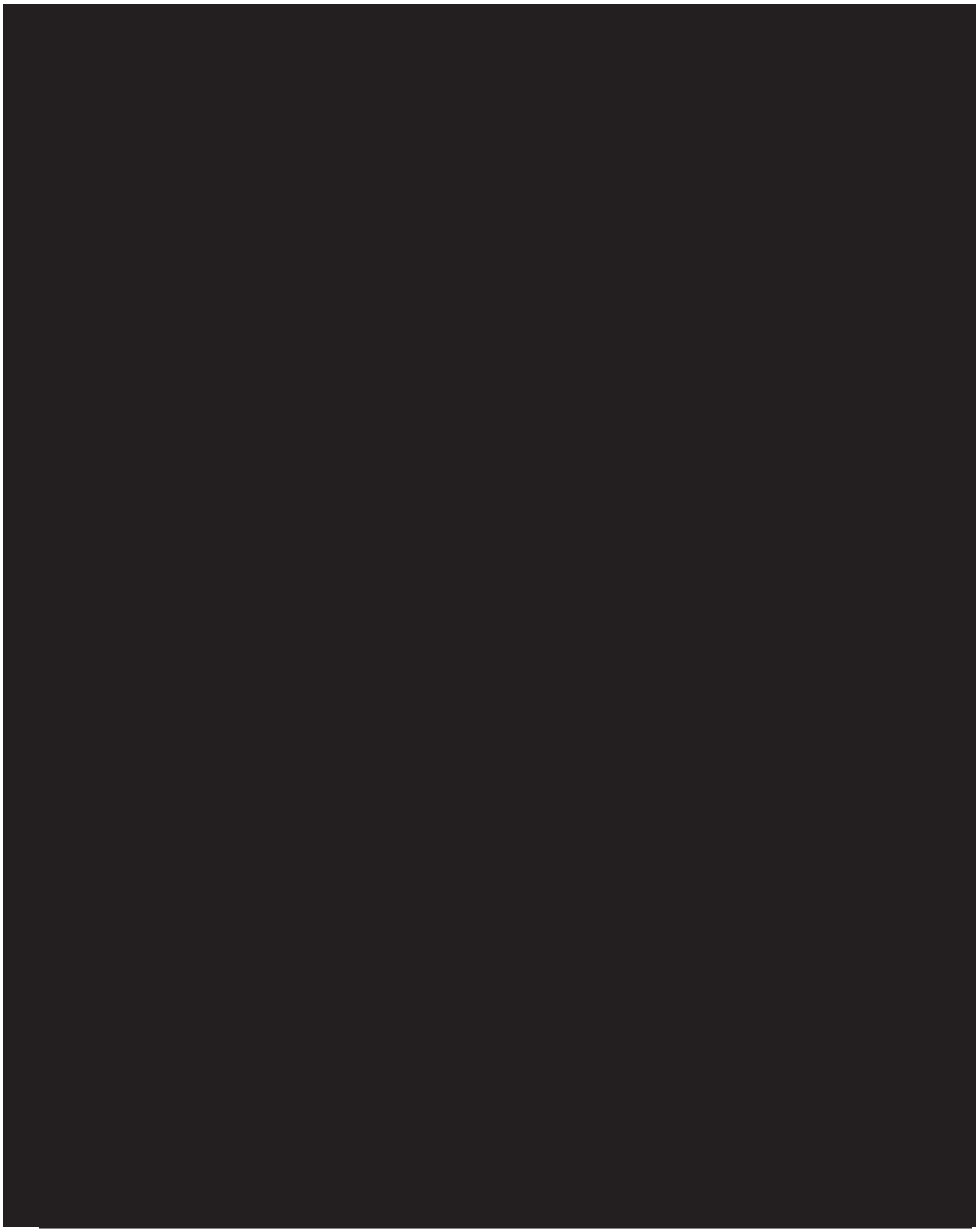
All graphs in this work were generated with Gnuplot version 4.2.

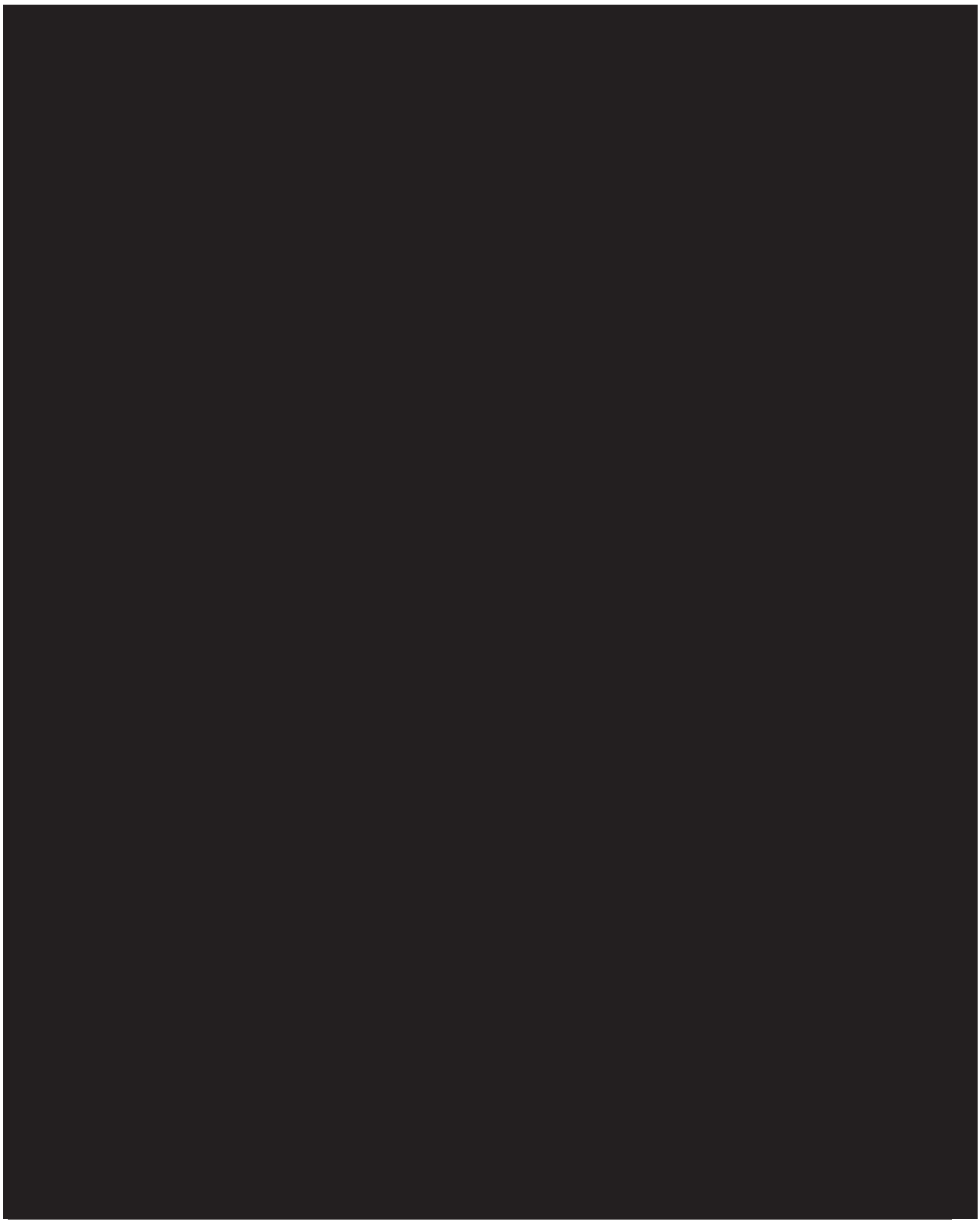
10. Papers

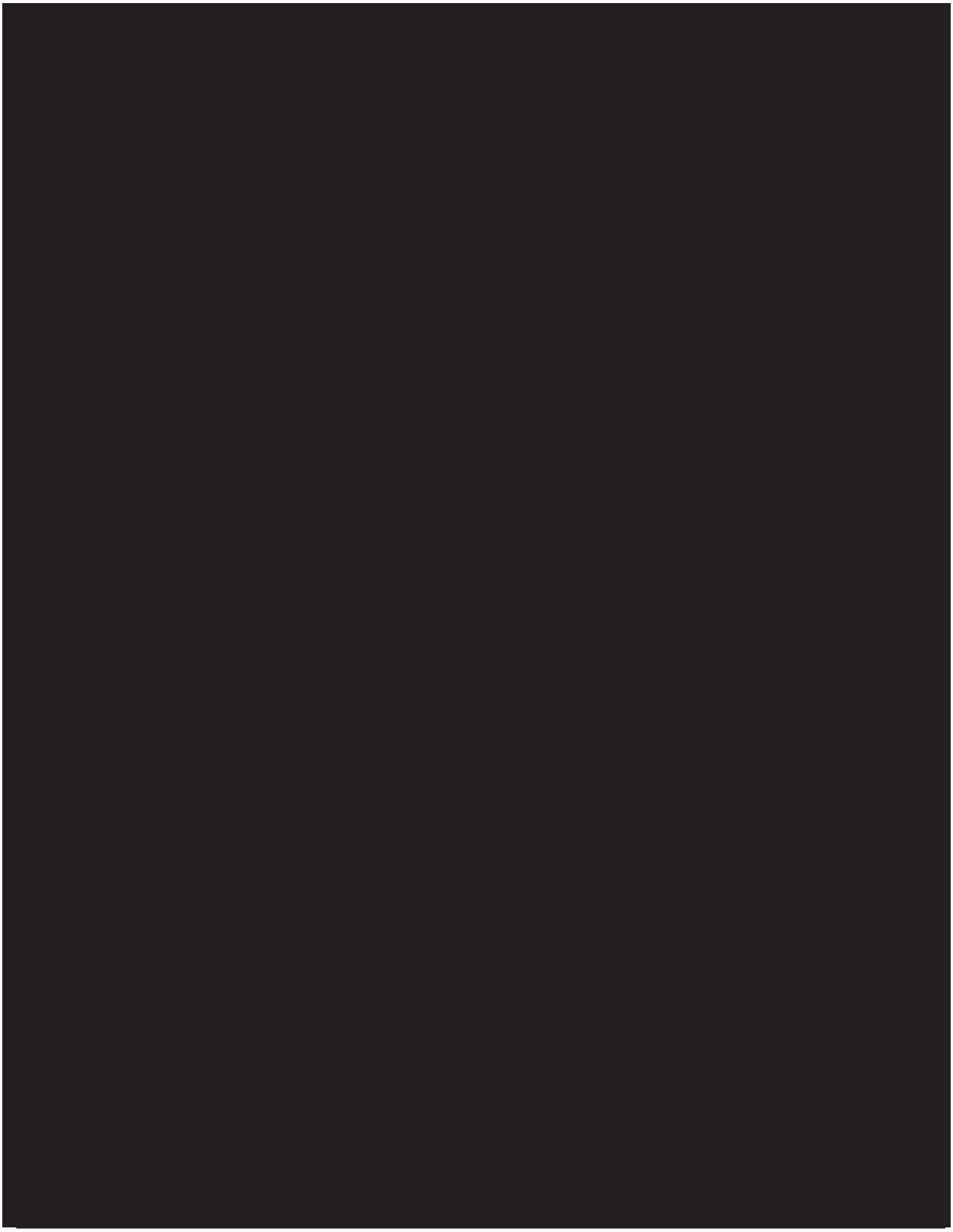
10.1 Paper I

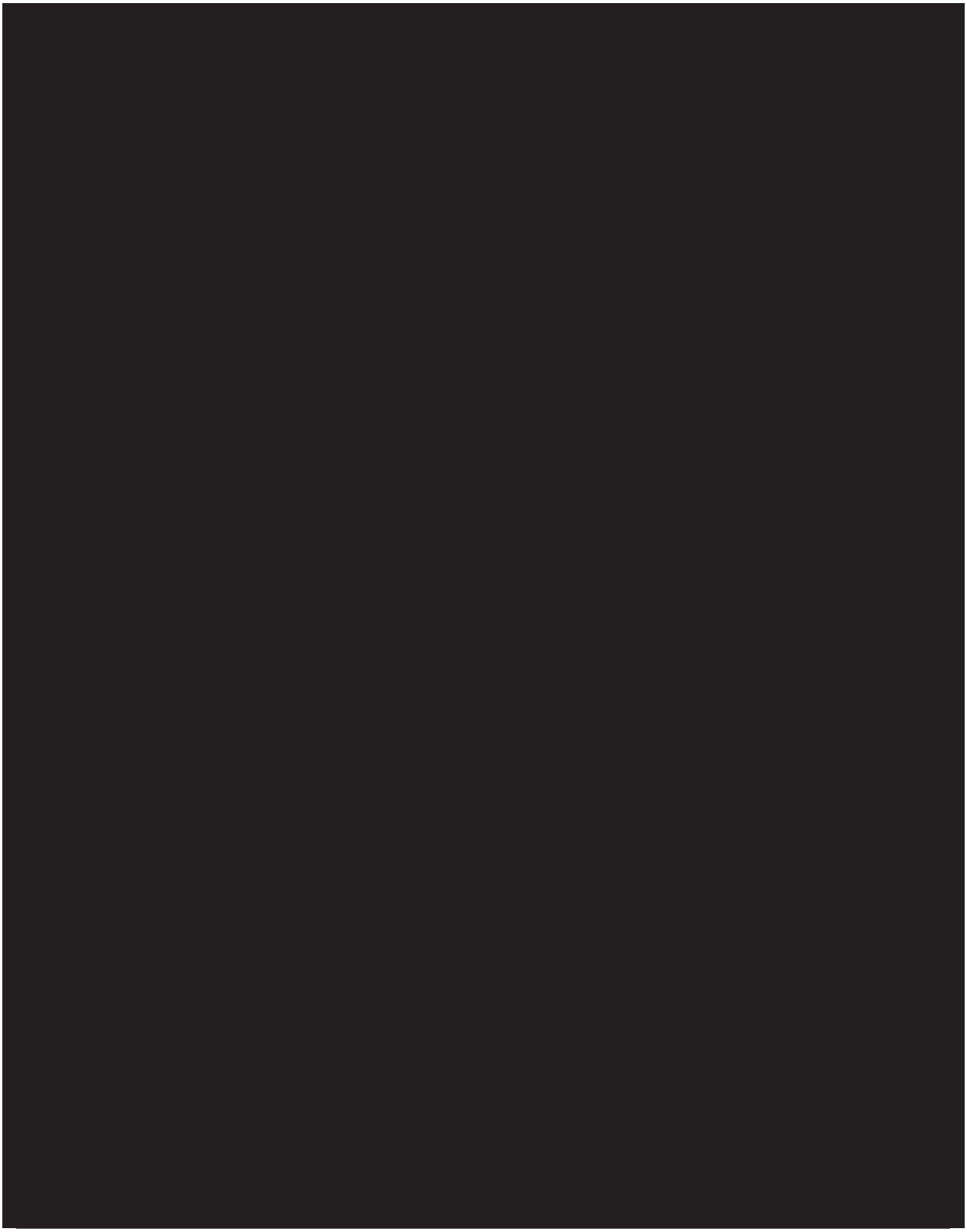


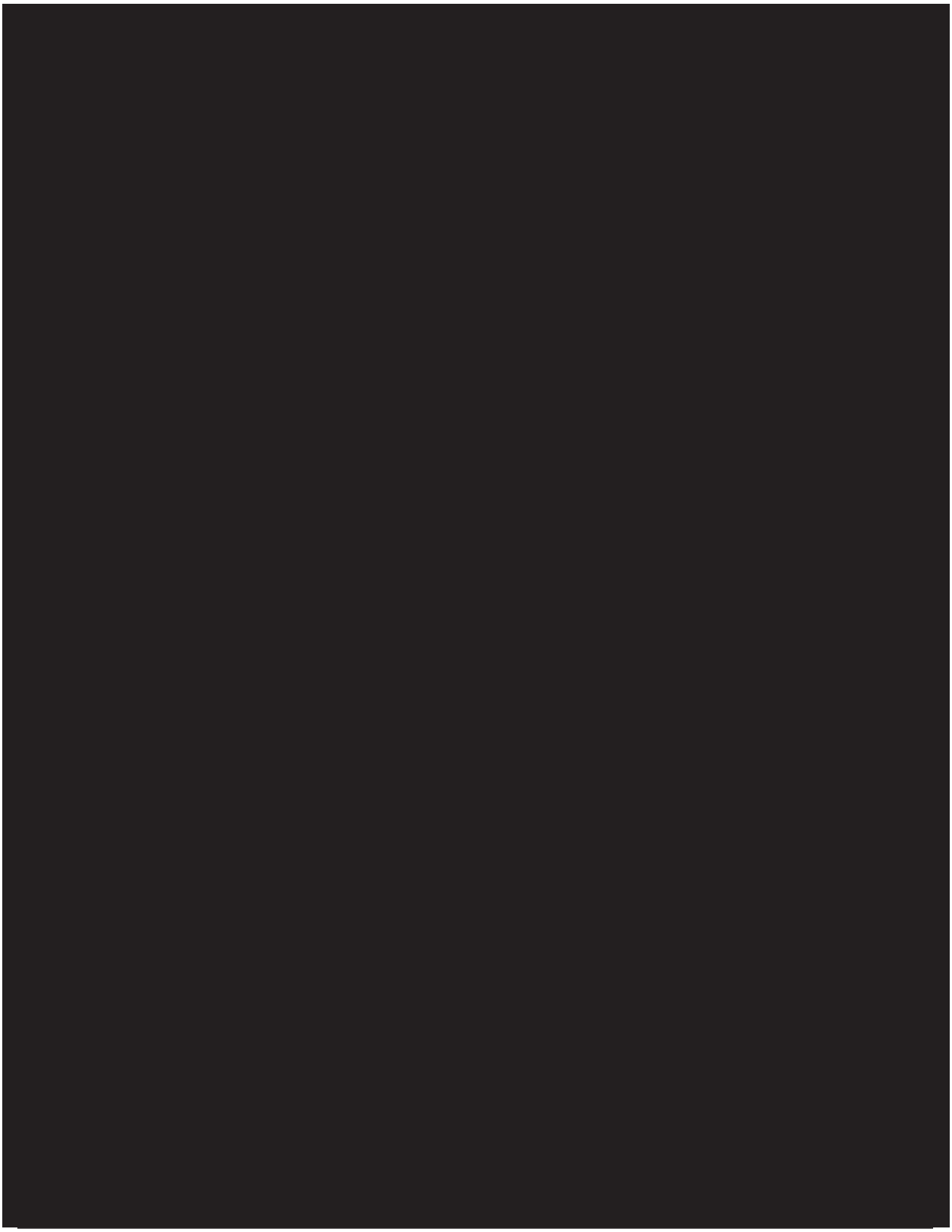












Reprinted (adapted) with permission from (Krüger, D. M., Gohlke, H., DrugScore^{PPI} webserver: fast and accurate *in silico* alanine scanning for scoring protein-protein interactions. Nucleic Acids Res. 2010, 38, W480-W486). DOI: 10.1093/nar/gkq471. © The Author(s) 2010. Published by Oxford University Press. http://www.oxfordjournals.org/.

Nucleic Acids Res., 8.026*, 65%, 1. Author

(Name of the journal, Impact factor, Contribution to this work, Authorship)

*last reported at the date of submission of this work

Supplementary data

DrugScore^{PPI} webserver:

Fast and accurate *in silico* alanine scanning for scoring protein-protein interactions

Dennis M. Krüger, Holger Gohlke[*]

Department of Mathematics and Natural Sciences, Heinrich-Heine-University, 40225 Düsseldorf, Germany

[(*)]To whom correspondence should be addressed: Universitätsstr. 1, 40225 Düsseldorf; Phone: +49-211-81-13662; Fax: +49-211-13847; E-mail: gohlke@uni-duesseldorf.de

Table S1: Alanine scanning dataset used for training[a]

PDB code	Mutated partner	Residue number	Amino acid[b]	$\Delta\Delta G_{exp}$ [c]	$\Delta\Delta G_{calc}$ [d]
1A22	hGH	21	H	0.2	0.48
1A22	hGH	22	Q	-0.2	0.22
1A22	hGH	25	F	-0.4	0.83
1A22	hGH	45	L	1.2	0.87
1A22	hGH	46	Q	0.1	0.38
1A22	hGH	56	E	0.4	0.66
1A22	hGH	62	S	0.1	0.46
1A22	hGH	63	N	0.3	0.61
1A22	hGH	64	R	1.6	3.21
1A22	hGH	65	E	-0.5	0.49
1A22	hGH	68	Q	0.6	0.87
1A22	hGH	164	Y	0.3	0.53
1A22	hGH	167	R	0.3	0.48
1A22	hGH	168	K	-0.2	1.03
1A22	hGH	171	D	0.8	1.45
1A22	hGH	172	K	2	1.02
1A22	hGH	174	E	-0.9	0.58
1A22	hGH	175	T	2	0.65
1A22	hGH	176	F	1.9	0.49
1A22	hGH	178	R	2.4	1.02
1A22	hGH	179	I	0.8	1.63
1A22	hGHbp	243	R	2.12	1.11
1A22	hGHbp	244	E	1.69	0.52
1A22	hGHbp	271	R	0.54	1.04
1A22	hGHbp	275	E	-0.1	0.26
1A22	hGHbp	276	W	0.51	2.33
1A22	hGHbp	298	S	-0.05	0.36
1A22	hGHbp	302	S	-0.2	0.64
1A22	hGHbp	303	I	1.61	0.89
1A22	hGHbp	305	I	1.94	0.69
1A22	hGHbp	320	E	-0.19	0.53
1A22	hGHbp	321	K	0.08	0.37
1A22	hGHbp	324	S	0.28	0.55
1A22	hGHbp	326	D	0.99	0.61
1A22	hGHbp	327	E	0.97	0.66
1A22	hGHbp	364	D	1.49	1.46
1A22	hGHbp	365	I	2.13	0.42
1A22	hGHbp	366	Q	0.02	0.10
1A22	hGHbp	367	K	-0.02	0.79
1A22	hGHbp	371	V	-0.64	0.74
1A22	hGHbp	416	Q	0.89	0.39
1A22	hGHbp	417	R	0.28	-0.18
1A22	hGHbp	418	N	0.3	0.91
1A22	hGHbp	419	S	0.03	0.55
1A22	hGHbp	301	T	1.76	0.53
1A4Y	Rnase Inh	261	W	0.1	0.77
1A4Y	Rnase Inh	263	W	1.2	2.15
1A4Y	Rnase Inh	289	S	0	0.51
1A4Y	Rnase Inh	318	W	1.5	2.01
1A4Y	Rnase Inh	320	K	-0.3	0.69
1A4Y	Rnase Inh	344	E	0.2	0.59
1A4Y	Rnase Inh	375	W	1	1.66

1A4Y	Rnase Inh	401	E	0.9	0.62
1A4Y	Rnase Inh	434	Y	3.3	3.41
1A4Y	Rnase Inh	435	D	3.5	3.82
1A4Y	Rnase Inh	437	Y	0.8	2.61
1A4Y	Rnase Inh	459	I	0.7	0.64
1A4Y	Angiogenin	5	R	2.3	2.03
1A4Y	Angiogenin	8	H	0.9	0.53
1A4Y	Angiogenin	12	Q	0.3	0.68
1A4Y	Angiogenin	13	H	-0.3	0.45
1A4Y	Angiogenin	31	R	0.2	0.94
1A4Y	Angiogenin	32	R	0.9	1.35
1A4Y	Angiogenin	68	N	0.2	0.41
1A4Y	Angiogenin	84	H	0.2	0.55
1A4Y	Angiogenin	89	W	0.2	1.59
1A4Y	Angiogenin	108	E	-0.3	0.73
1A4Y	Angiogenin	114	H	0.65	0.53
1AHW	TF	167	T	0	0.35
1AHW	TF	170	T	1	0.47
1AHW	TF	176	L	1	0.36
1AHW	TF	178	D	-0.5	0.49
1AHW	TF	197	T	1.3	0.34
1AHW	TF	198	V	-0.3	0.47
1AHW	TF	199	N	1.1	0.45
1AIE	P53	352	D	0.55	0.89
1AIE	P53	326	E	0.4	-0.09
1AIE	P53	336	E	-0.75	0.31
1AIE	P53	339	E	0.45	0.40
1AIE	P53	343	E	0.28	0.26
1AIE	P53	346	E	0.1	0.34
1AIE	P53	349	E	0.95	0.56
1AIE	P53	328	F	1.8	1.02
1AIE	P53	338	F	1.8	1.05
1AIE	P53	351	K	-0.62	1.16
1AIE	P53	350	L	0.53	1.09
1AIE	P53	340	M	1.6	0.46
1AIE	P53	345	N	0.93	1.84
1AIE	P53	331	Q	0.35	0.06
1AIE	P53	333	R	0.97	0.64
1AIE	P53	335	R	0.53	0.75
1AIE	P53	337	R	2.33	1.57
1AIE	P53	342	R	0.35	0.69
1AIE	P53	329	T	0.9	0.42
1BRS	Barnase	54	D	-0.8	0.35
1BRS	Barnase	59	R	5.2	2.93
1BRS	Barnase	60	E	-0.2	0.81
1BRS	Barnase	73	E	2.8	0.56
1BRS	Barstar	29	Y	3.4	5.03
1BRS	Barstar	35	D	4.5	3.70
1BRS	Barstar	42	T	1.8	0.49
1BRS	Barstar	76	E	1.3	0.54
1BRS	Barstar	80	E	0.5	0.37
1BXI	Im9	23	C	0.92	0.38
1BXI	Im9	24	N	0.14	0.02
1BXI	Im9	26	D	0.34	0.31
1BXI	Im9	27	T	0.73	0.57
1BXI	Im9	28	S	0.17	0.40

1BXI	Im9	29	S	0.96	0.47
1BXI	Im9	30	E	1.41	0.80
1BXI	Im9	33	L	3.42	0.96
1BXI	Im9	34	V	2.58	0.90
1BXI	Im9	37	V	1.66	1.03
1BXI	Im9	38	T	0.9	0.55
1BXI	Im9	41	E	2.08	0.69
1BXI	Im9	48	S	0.01	0.36
1BXI	Im9	50	S	2.19	0.57
1BXI	Im9	53	I	0.85	0.59
1BXI	Im9	54	Y	4.83	4.04
1BXI	Im9	55	Y	4.63	4.48
1CBW	BPTI	11	T	0.2	0.53
1CBW	BPTI	15	K	2	2.17
1CBW	BPTI	17	R	0.5	1.11
1CBW	BPTI	19	I	0.1	0.99
1CBW	BPTI	34	V	0	0.44
1CBW	BPTI	39	R	0.2	0.93
1DAN	TF	17	T	0.1	0.45
1DAN	TF	18	N	0.2	0.79
1DAN	TF	20	K	2.6	1.34
1DAN	TF	21	T	-0.2	0.42
1DAN	TF	22	I	0.7	1.34
1DAN	TF	24	E	0.7	0.57
1DAN	TF	37	Q	0.55	0.96
1DAN	TF	41	K	0.35	0.35
1DAN	TF	42	S	-0.1	0.28
1DAN	TF	44	D	0.7	2.23
1DAN	TF	46	K	0.25	0.69
1DAN	TF	47	S	0.05	0.39
1DAN	TF	48	K	0.4	0.62
1DAN	TF	50	F	0.4	1.44
1DAN	TF	58	D	2.18	2.21
1DAN	TF	68	K	-0.1	0.33
1DAN	TF	99	E	-0.2	0.37
1DAN	TF	128	E	0.1	0.46
1DAN	TF	133	L	0	1.22
1DAN	TF	135	R	0.55	1.17
1DAN	TF	140	F	1.5	0.81
1DAN	TF	163	S	0	0.40
1DAN	TF	203	T	0.1	0.43
1DAN	TF	207	V	-0.2	1.42
1DAN	TF	208	E	0	0.45
1DFJ	Rnase Inh	202	E	1	0.55
1DFJ	Rnase Inh	257	W	1.3	1.30
1DFJ	Rnase Inh	259	W	2.2	2.01
1DFJ	Rnase Inh	283	E	1.3	0.41
1DFJ	Rnase Inh	285	S	0.8	0.39
1DFJ	Rnase Inh	314	W	1	0.76
1DFJ	Rnase Inh	316	K	1.3	0.48
1DFJ	Rnase Inh	397	E	1.3	0.62
1DFJ	Rnase Inh	453	R	0.8	0.49
1DFJ	Rnase Inh	455	I	0.3	1.06
1DFJ	Rnase Inh	430	Y	5.9	4.50
1DFJ	Rnase Inh	431	D	3.6	2.45
1DFJ	Rnase Inh	433	Y	2.6	2.38

1DVF	D1.3	H30	T	0.9	0.45
1DVF	D1.3	H32	Y	1.8	0.01
1DVF	D1.3	H52	W	4.2	2.89
1DVF	D1.3	H54	D	4.3	3.46
1DVF	D1.3	H56	N	1.2	1.53
1DVF	D1.3	H58	D	1.6	0.82
1DVF	D1.3	H99	R	1.9	0.23
1DVF	D1.3	H100	D	2.8	1.16
1DVF	D1.3	L30	H	1.7	0.44
1DVF	D1.3	L32	Y	2	1.47
1DVF	D1.3	L49	Y	1.7	0.77
1DVF	D1.3	L50	Y	0.7	1.04
1DVF	D1.3	L92	W	0.3	1.16
1DVF	D1.3	L93	S	1.2	0.42
1DVF	E5.2	30	K	1	0.33
1DVF	E5.2	33	H	1.9	0.62
1DVF	E5.2	52	D	1.7	1.03
1DVF	E5.2	54	N	1.9	1.25
1DVF	E5.2	97	I	2.7	1.81
1DVF	E5.2	98	Y	4.7	4.58
1DVF	E5.2	100	Q	1.6	1.68
1DVF	E5.2	49	Y	1.9	1.12
1F47	FTSZ fragm.	4	D	0.7	0.00
1F47	FTSZ fragm.	5	Y	0.9	2.35
1F47	FTSZ fragm.	6	L	0.9	1.70
1F47	FTSZ fragm.	7	D	1.8	0.89
1F47	FTSZ fragm.	8	I	2.5	1.74
1F47	FTSZ fragm.	11	F	2.5	1.11
1F47	FTSZ fragm.	12	L	2.3	1.05
1F47	FTSZ fragm.	14	K	0	0.42
1F47	FTSZ fragm.	15	Q	0	0.19
1FCC	Protein G	25	T	0.24	0.43
1FCC	Protein G	28	K	1.3	1.55
1FCC	Protein G	31	K	3.5	1.10
1FCC	Protein G	35	N	2.4	2.45
1FCC	Protein G	40	D	0.3	0.94
1FCC	Protein G	42	E	0.4	0.52
1FCC	Protein G	43	W	3.8	1.56
1GC1	CD4	23	S	0.29	0.47
1GC1	CD4	25	Q	0.03	0.41
1GC1	CD4	27	H	0.28	0.51
1GC1	CD4	29	K	0.59	0.99
1GC1	CD4	31	S	0.1	0.46
1GC1	CD4	32	N	0.18	0.11
1GC1	CD4	33	Q	0.1	0.21
1GC1	CD4	35	K	0.32	1.42
1GC1	CD4	40	Q	-0.41	0.45
1GC1	CD4	42	S	0	0.59
1GC1	CD4	44	L	1.04	0.51
1GC1	CD4	45	T	-0.15	0.50
1GC1	CD4	52	N	0.7	1.07
1GC1	CD4	56	D	-0.07	0.42
1GC1	CD4	59	R	1.16	0.59
1GC1	CD4	60	S	-0.09	0.28
1GC1	CD4	63	D	-0.32	1.07
1GC1	CD4	64	Q	0.44	0.76

1IAR	IL4	19	E	-0.32	0.40
1IAR	IL4	82	F	-0.08	0.53
1IAR	IL4	11	I	0.07	0.41
1IAR	IL4	5	I	1.17	1.66
1IAR	IL4	77	K	0.15	0.37
1IAR	IL4	84	K	0.35	0.42
1IAR	IL4	15	N	-0.03	0.42
1IAR	IL4	89	N	1.56	1.91
1IAR	IL4	78	Q	0.13	0.50
1IAR	IL4	8	Q	-0.02	0.38
1IAR	IL4	81	R	0.48	1.16
1IAR	IL4	85	R	0.43	1.09
1IAR	IL4	88	R	3.75	1.79
1IAR	IL4	16	S	-0.18	0.41
1IAR	IL4	13	T	0.98	0.53
1IAR	IL4	6	T	-0.1	0.62
1JCK	SEC3	20	T	1.4	0.60
1JCK	SEC3	26	Y	1.7	1.92
1JCK	SEC3	60	N	1.3	1.08
1JCK	SEC3	91	V	2.1	1.58
1JCK	SEC3	103	K	0.4	0.44
1JCK	SEC3	176	F	1.9	0.60
1JRH	A6	L27	E	0.54	0.43
1JRH	A6	L28	D	0.44	0.54
1JRH	A6	L30	Y	1.1	1.62
1JRH	A6	L91	Y	0.58	0.91
1JRH	A6	L92	W	2.8	2.75
1JRH	A6	L93	S	-0.65	0.43
1JRH	A6	L94	T	0.38	0.51
1JRH	A6	L96	W	1.7	0.82
1JRH	A6	H32	Y	1.4	1.85
1JRH	A6	H52	W	2.7	1.60
1JRH	A6	H53	W	2.4	0.67
1JRH	A6	H54	D	1.9	1.12
1JRH	A6	H56	D	1.8	0.91
1JRH	A6	H58	Y	1.2	2.54
1JRH	A6	H95	R	0.54	0.37
1JRH	Interferon	48	N	-0.3	0.48
1JRH	Interferon	51	V	1.9	1.57
1JRH	Interferon	52	K	3	1.81
1JRH	Interferon	53	N	3.9	2.84
1JRH	Interferon	54	S	0.3	0.43
1JRH	Interferon	79	N	-0.4	0.65
1JRH	Interferon	84	R	-0.3	0.28
1JRH	Interferon	98	K	0	0.83
1VFB	D1.3	L30	H	0.8	0.52
1VFB	D1.3	L32	Y	1.3	2.17
1VFB	D1.3	L49	Y	0.8	1.28
1VFB	D1.3	L50	Y	0.4	1.80
1VFB	D1.3	L53	T	-0.23	0.54
1VFB	D1.3	L92	W	1.71	1.34
1VFB	D1.3	L93	S	0.11	0.37
1VFB	D1.3	H30	T	0.09	0.42
1VFB	D1.3	H32	Y	0.5	0.99
1VFB	D1.3	H52	W	1.23	1.59
1VFB	D1.3	H99	R	0.47	0.00

1VFB	D1.3	H100	D	3.1	3.61
1VFB	HEL	18	D	0.3	1.85
1VFB	HEL	19	N	0.3	1.39
1VFB	HEL	23	Y	0.4	0.41
1VFB	HEL	24	S	0.8	0.51
1VFB	HEL	116	K	0.7	0.95
1VFB	HEL	118	T	0.8	0.38
1VFB	HEL	119	D	1	1.56
1VFB	HEL	120	V	0.9	0.85
1VFB	HEL	121	Q	2.9	1.95
1VFB	HEL	124	I	1.2	0.73
1VFB	HEL	125	R	1.8	2.19
1VFB	HEL	129	L	0.2	0.41
3HFM	HYHEL-10	L50	Y	4.6	3.28
3HFM	HYHEL-10	L53	Q	1	0.49
3HFM	HYHEL-10	L96	Y	2.8	1.37
3HFM	HYHEL-10	H31	S	0.2	0.53
3HFM	HYHEL-10	H32	D	2	0.77
3HFM	HYHEL-10	H33	Y	6	4.95
3HFM	HYHEL-10	H53	Y	3.29	2.75
3HFM	HYHEL-10	H58	Y	1.7	3.09
3HFM	HEL	15	H	-0.5	0.23
3HFM	HEL	20	Y	5	2.84
3HFM	HEL	21	R	1	1.93
3HFM	HEL	63	W	0.3	0.73
3HFM	HEL	73	R	-0.2	1.07
3HFM	HEL	75	L	1.25	1.04
3HFM	HEL	89	T	0	0.58
3HFM	HEL	93	N	0.6	0.97
3HFM	HEL	98	I	-0.1	0.50
3HFM	HEL	100	S	0.25	0.58
3HFM	HEL	101	D	1.50	1.77

[a] 309 protein-protein interface alanine mutations derived from the Alanine Scanning Energetics Database (ASEdb).[1]
[b] One letter code.
[c] Experimental ΔΔG values for wildtype-to-Ala mutations in kcal mol^{-1} derived from ASEdb.
[d] ΔΔG values in kcal mol^{-1} computed by adapted Drugscore^{PPI} potentials.

Table S2: External Ras/RalGDS test set[a]

PDB code	Mutated partner	Residue number[b]	Amino acid[c]	$\Delta\Delta G_{exp}$ [d]	$\Delta\Delta G_{calc}$ [e]
1LFD	RalGDS-RBD	14	I	1.45	1.01
1LFD	RalGDS-RBD	16	R	2.30	0.51
1LFD	RalGDS-RBD	23	N	0.95	-0.31
1LFD	RalGDS-RBD	25	N	0.50	1.03
1LFD	RalGDS-RBD	27	Y	3.60	3.16
1LFD	RalGDS-RBD	28	K	2.50	0.83
1LFD	RalGDS-RBD	29	S	0.95	0.47
1LFD	RalGDS-RBD	44	K	0.45	1.27
1LFD	RalGDS-RBD	47	D	-0.30	0.2
1LFD	RalGDS-RBD	48	K	2.80	0.87
1LFD	RalGDS-RBD	52	E	-0.20	0.47
1LFD	Ras	25	Q	0.90	0.19
1LFD	Ras	29	V	0.50	0.06
1LFD	Ras	31	E	0.25	0.47
1LFD	Ras	33	D	1.10	1.83
1LFD	Ras	37	E	1.20	-0.32
1LFD	Ras	38	D	3.90	2.62
1LFD	Ras	39	S	-0.75	0.23
1LFD	Ras	40	Y	3.70	1.23
1LFD	Ras	41	R	0.85	0.81
1LFD	Ras	62	E	0.20	0.03
1LFD	Ras	63	E	0.10	0.24

[a] 22 alanine mutations in the Ras/RalGDS interface derived from Kiel *et al.*[2]
[b] Numbering according to Vetter *et al.*[3]
[c] One letter code.
[d] Experimental $\Delta\Delta G$ values in kcal mol^{-1} derived from Kiel *et al.*[2]
[e] $\Delta\Delta G$ values in kcal mol^{-1} computed by adapted Drugscore^{PPI} potentials.

Table S3: Weighting coefficients according to eq. 2

Coefficient[a]	Value[b]
$c_{T(r)}$	
Val_C.3	-19.40
Leu_C.3	-12.78
Ile_C.3	-21.92
Trp_C.2	-30.73
Asn_N.am	-28.89
Gln_C.2	-49.04
Gln_O.2	4.43
Gln_N.am	20.35
Tyr_C.3	28.19
Tyr_C.ar	-27.73
Asp_C.3	-48.80
Asp_C.2	90.97
Asp_O.co2	-72.96
Glu_C.2	24.25
Glu_O.co2	-19.57
His_C.2	2.56
Lys_C.3	-10.15
Arg_C.3	-24.39
Arg_C.cat	-112.79
Arg_N.pl3	38.07
Backbone_C.2	-10.59
Backbone_O.2	-35.13
Backbone_N.am[c]	-7.85
s	
Met_C.3	
Met_S.3	
Phe_C.3	
Phe_C.ar	
Cys_C.3	
Cys_S.3	
His_C.3	
Trp_C.3	
Trp_C.ar	
Ser_C.3	
Ser.O.3	-6.30
Thr_C.3	
Thr_O.3	
Asn_C.3	
Asn_C.2	
Asn_O.2	
Gln_C.3	
Tyr_O.3	
Glu_C.3	
Lys_N.3	
Backbone_C.3	
k	0.03
a	0.89

[a] Amino acid three letter code (except for 'Backbone') followed by Tripos mol2 type descriptor.
[b] In 10^6 kcal mol^{-1}.
[c] This includes Trp_N.am and His_N.am for the sidechain nitrogen atoms of Trp and His, respectively. Following the Tripos mol2 type definition, Trp_N.pl3 and His_N.pl3 would have been chosen. However,

because these residues form H-bonds in the majority of the cases and not salt-bridges, the N.pl3 atomtype was reserved for Arg and replaced by N.am in Trp and His.

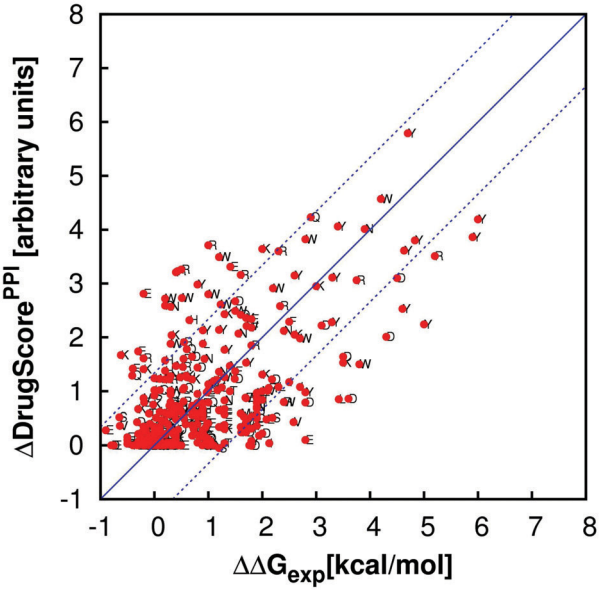


Figure S1: Calculated ΔΔG values using original Drugscore^{PPI} pair potentials versus experimentally determined ΔΔG values for the alanine scanning dataset (Table S1): r = 0.58, STD = 1.06 kcal mol^{-1}, p-value < 0.05, F = 158.48, N = 309.

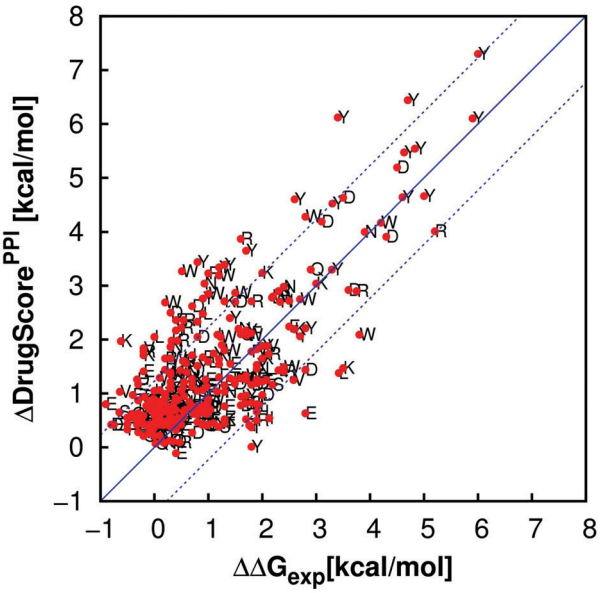


Figure S2: Calculated ΔΔG values using adapted Drugscore^{PPI} potentials versus experimentally determined ΔΔG values for the alanine scanning dataset (Table S1): r_{train} = 0.73, STD = 0.84 kcal mol^{-1}, p-value < 0.05, F = 345.16, N = 309.

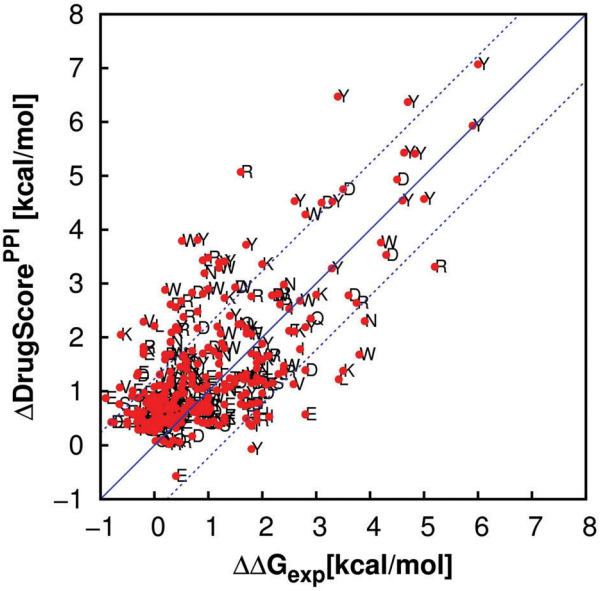


Figure S3: Calculated ΔΔG values obtained by leave-one-mutation-out cross-validation using adapted Drugscore[PPI] potentials versus experimentally determined ΔΔG values for the alanine scanning dataset (Table S1). r_{LOO} = 0.64, *STD* = 0.94 kcal mol^{-1}, *N* = 309.

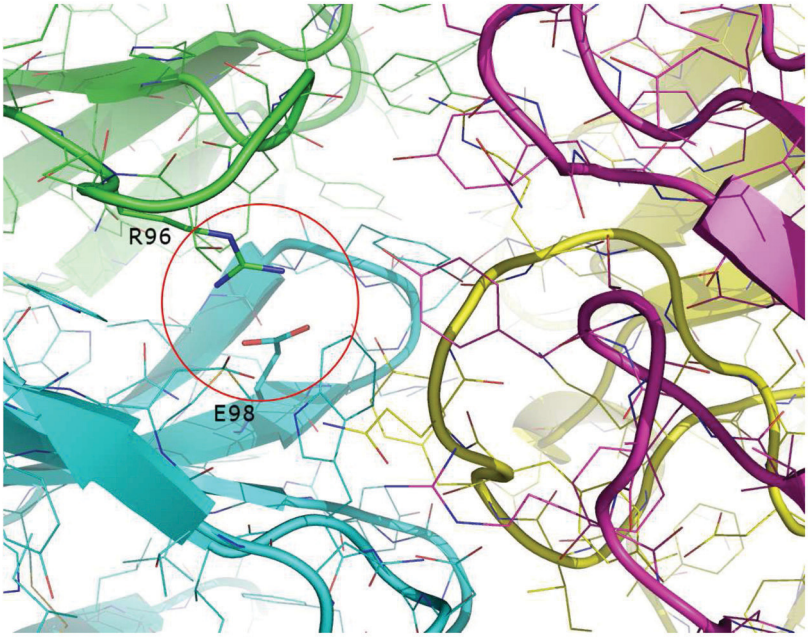
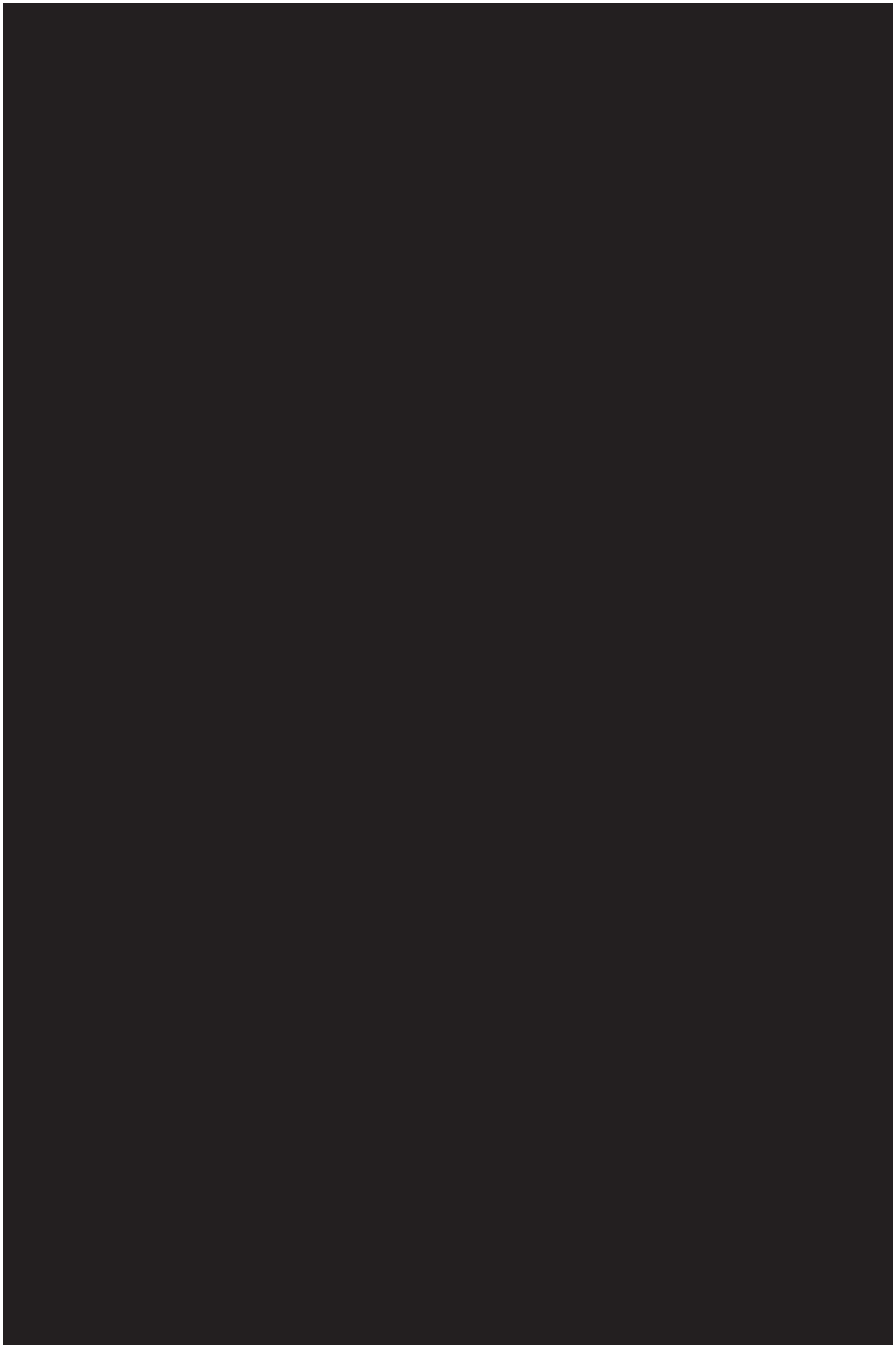


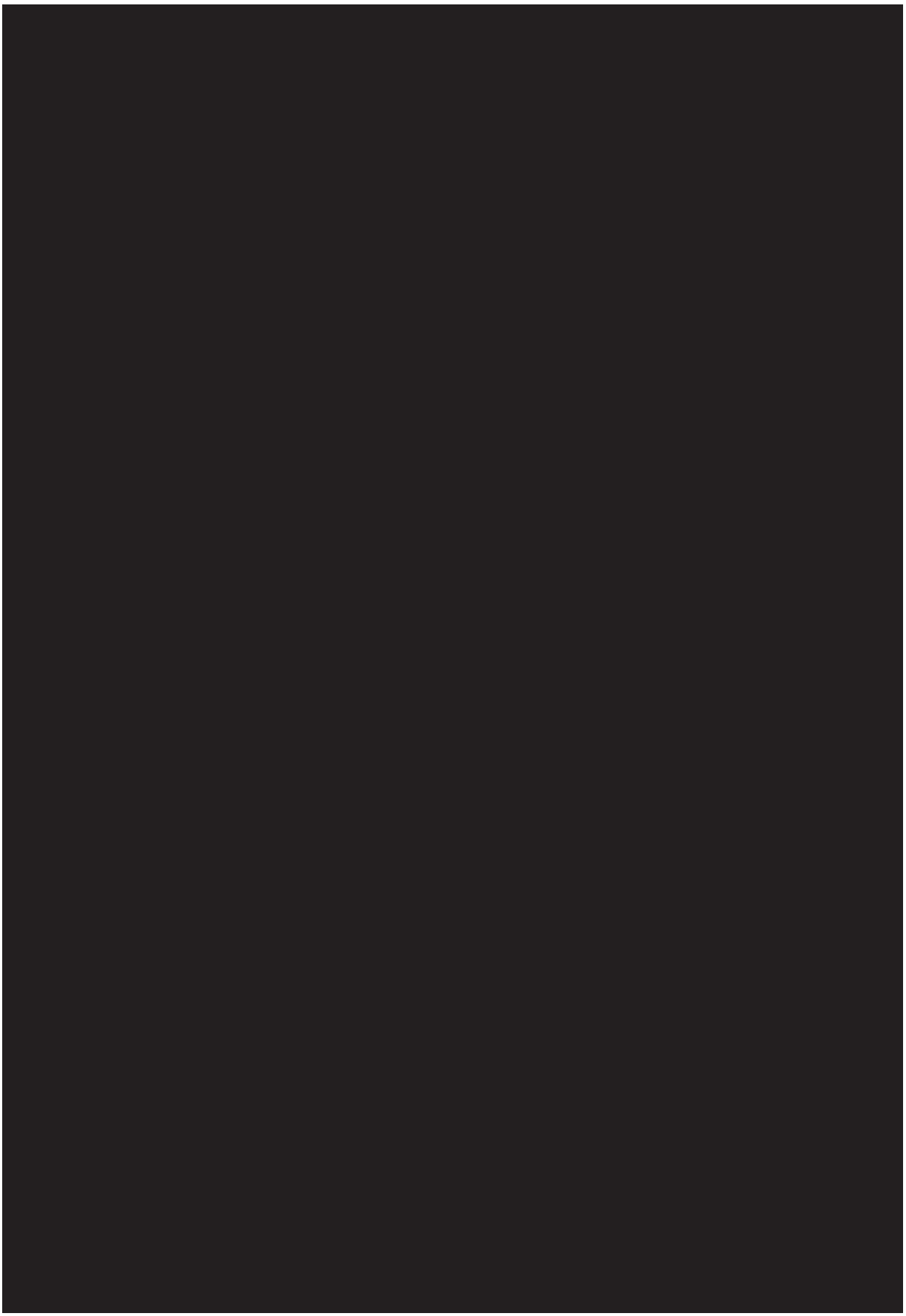
Figure S4: Example for mutations not included in the dataset of 309 alanine mutations. Close to the protein-protein interface of PDB entry 1DVF, formed by fragments from antibodies D1.3 (chain A: green; chain B: cyan) and E5.2 (chain C :magenta; chain D: yellow), residues R96 on chain A and E98 on chain B form an intramolecular saltbridge. ΔΔG values associated with the alanine mutations of these residues will very likely report on the stabilization or destabilization of the structure of D1.3 rather than on changes in the interactions with E5.2.

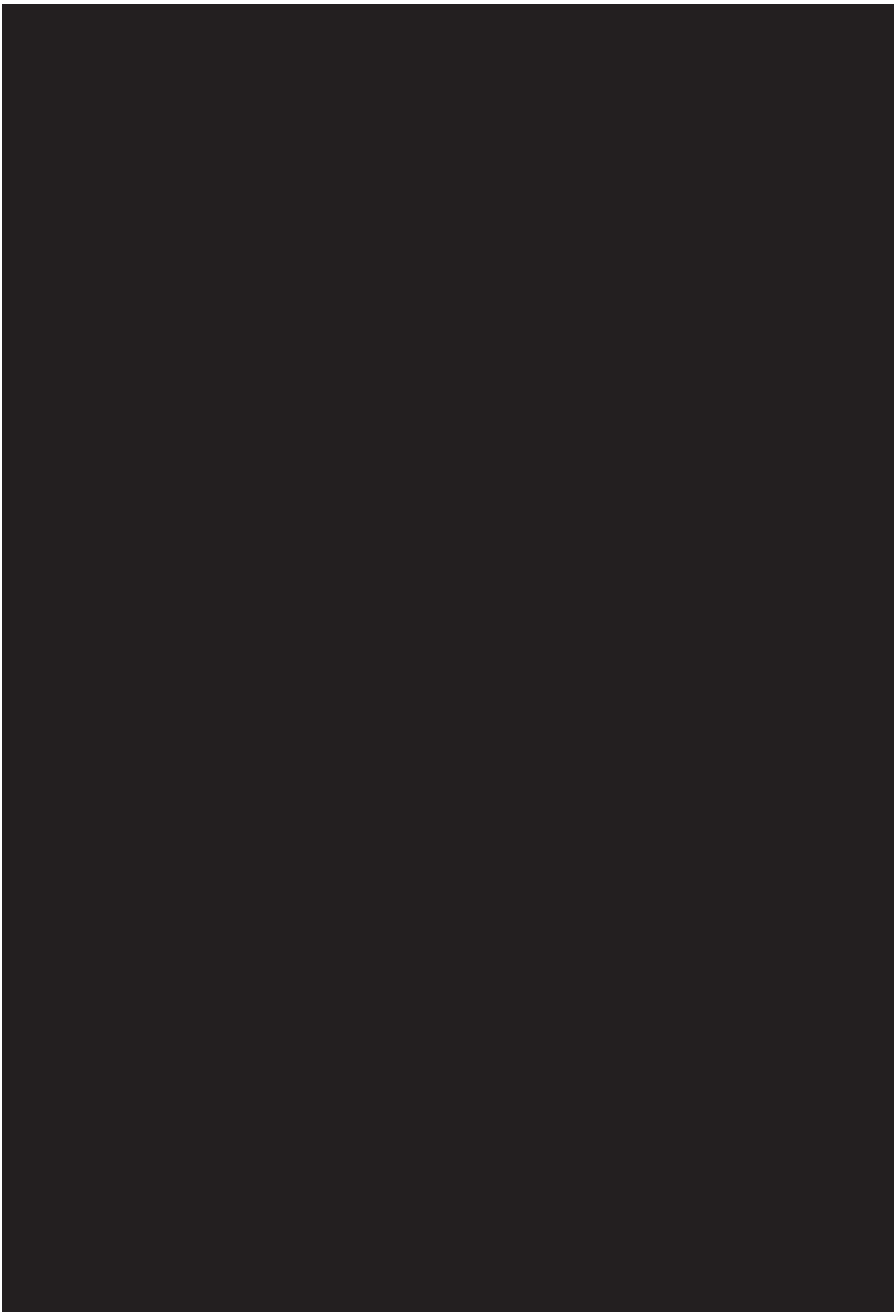
References

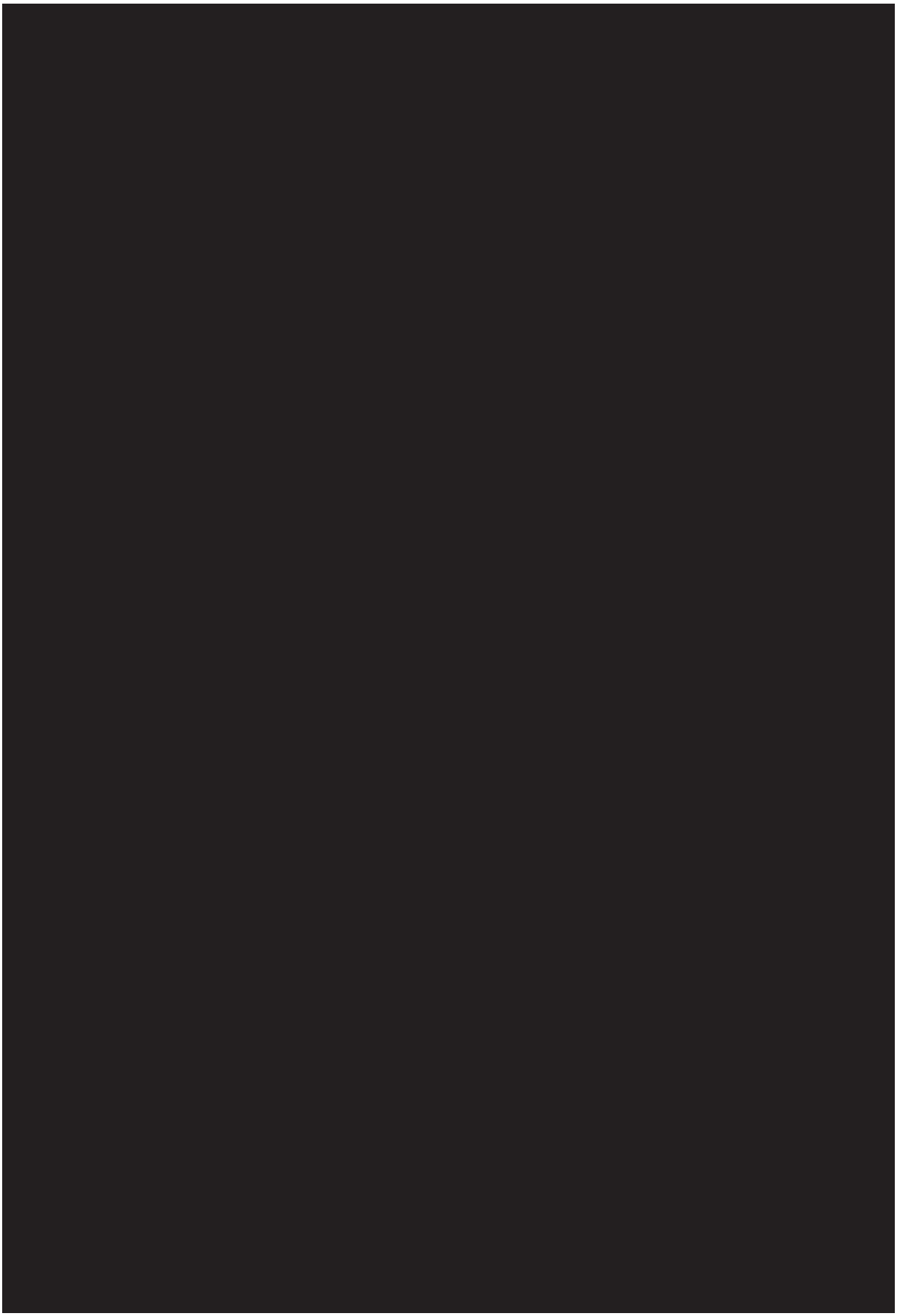
1. Thorn, K. S. & Bogan, A. A. (2001). ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* **17**, 284-5.
2. Kiel, C., Serrano, L. & Herrmann, C. (2004). A detailed thermodynamic analysis of ras/effector complex interfaces. *J Mol Biol* **340**, 1039-58.
3. Vetter, I. R., Linnemann, T., Wohlgemuth, S., Geyer, M., Kalbitzer, H. R., Herrmann, C. & Wittinghofer, A. (1999). Structural and biochemical analysis of Ras-effector signaling via RalGDS. *FEBS Lett* **451**, 175-80.

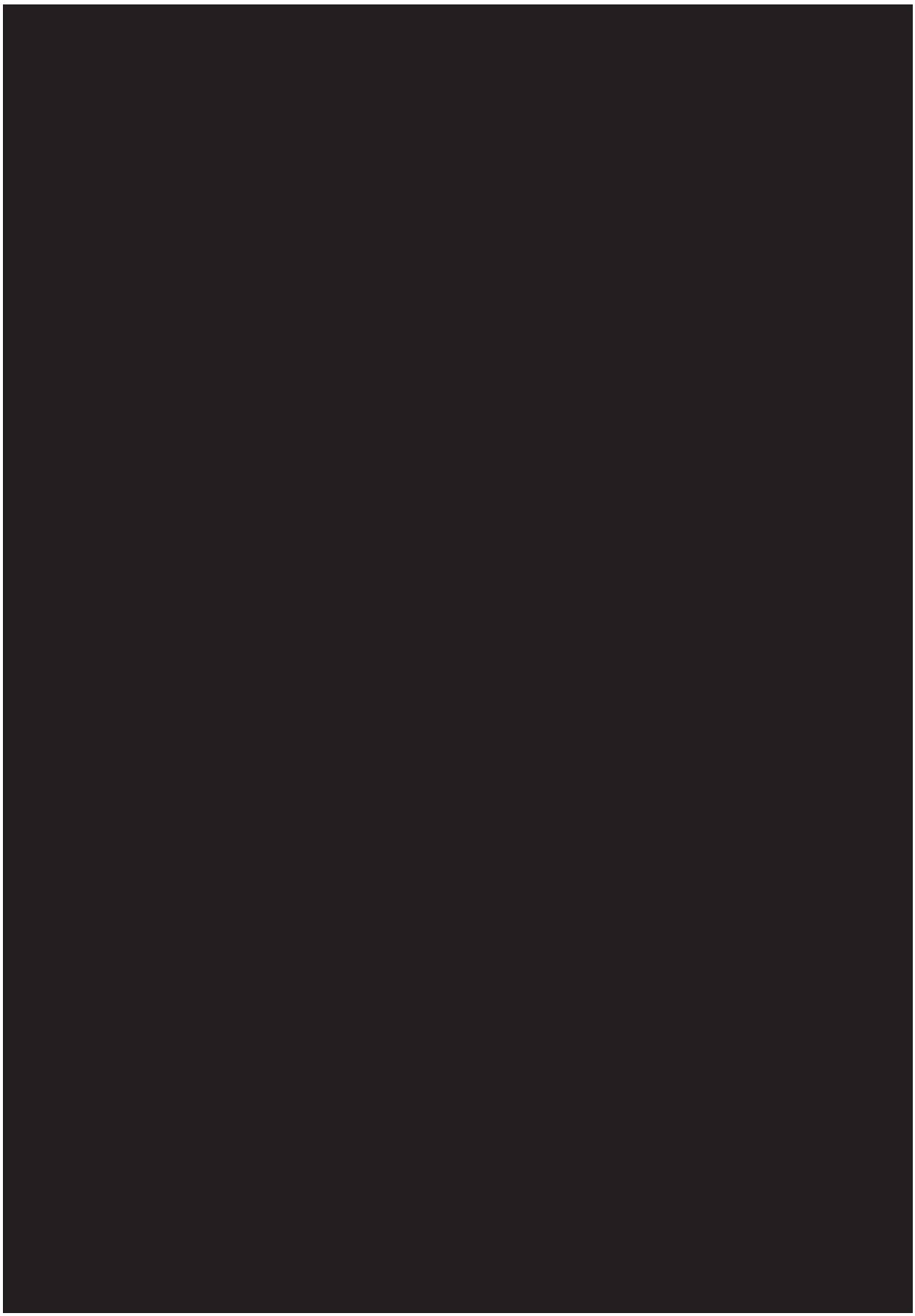
10.2 Paper II

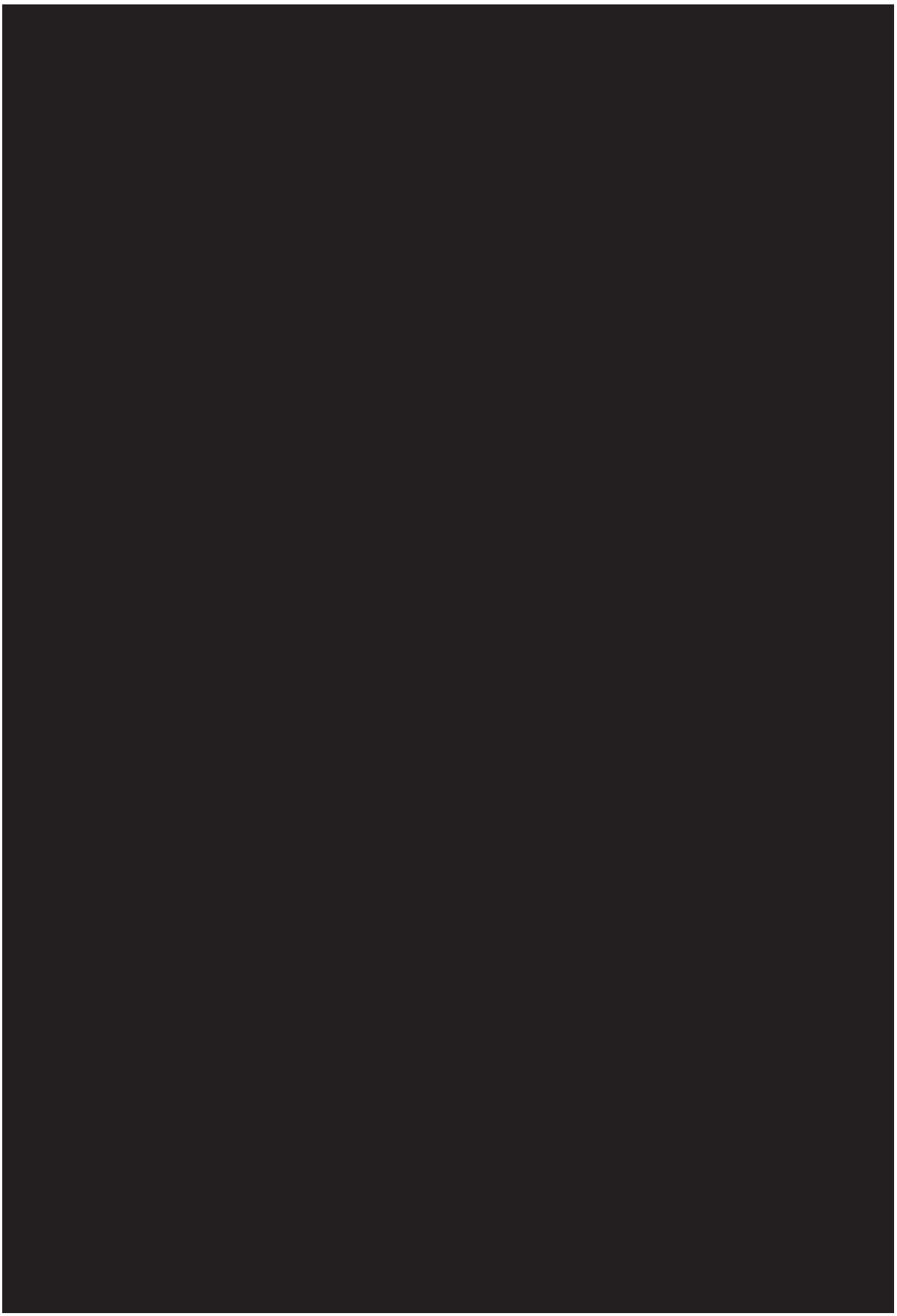


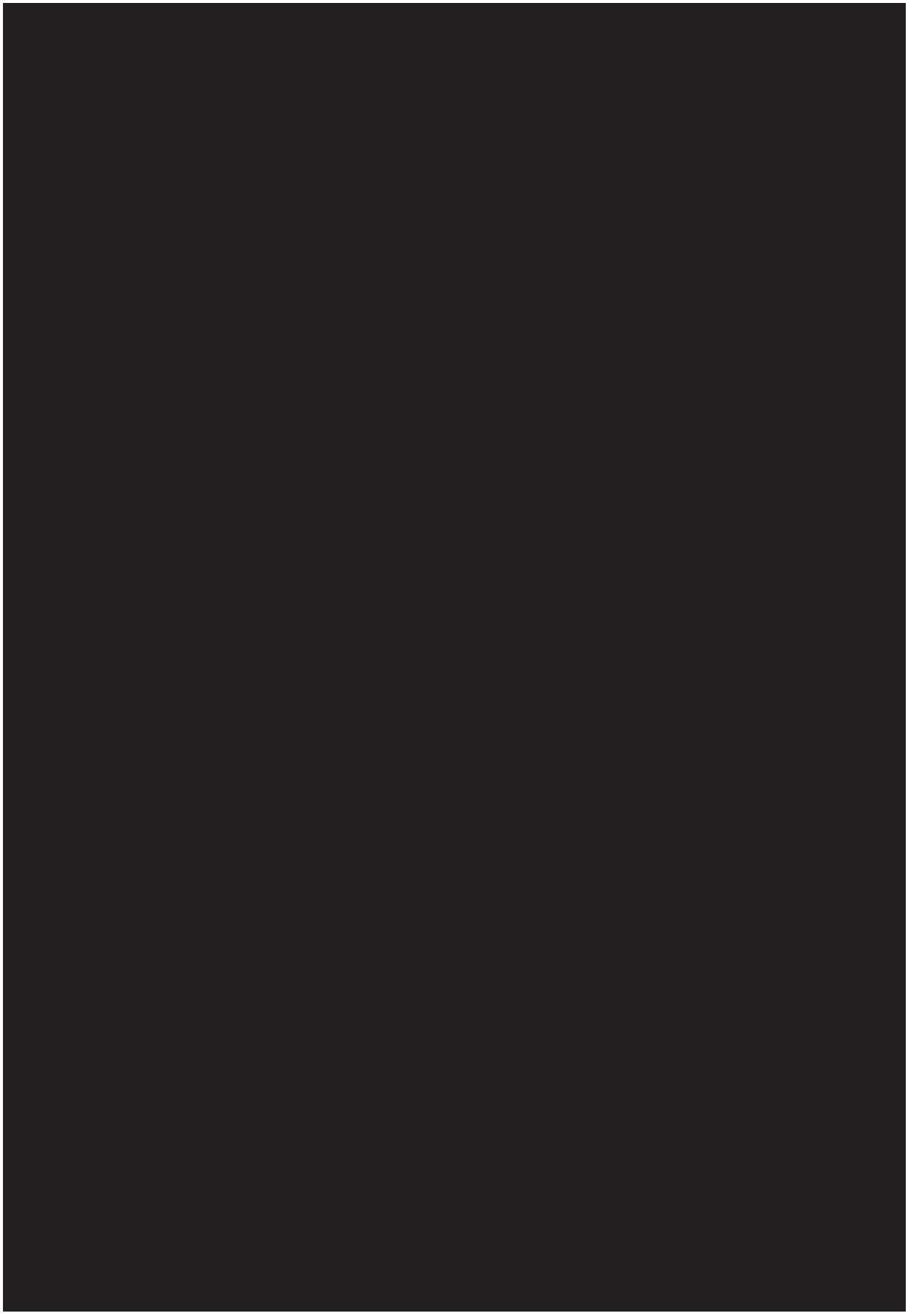


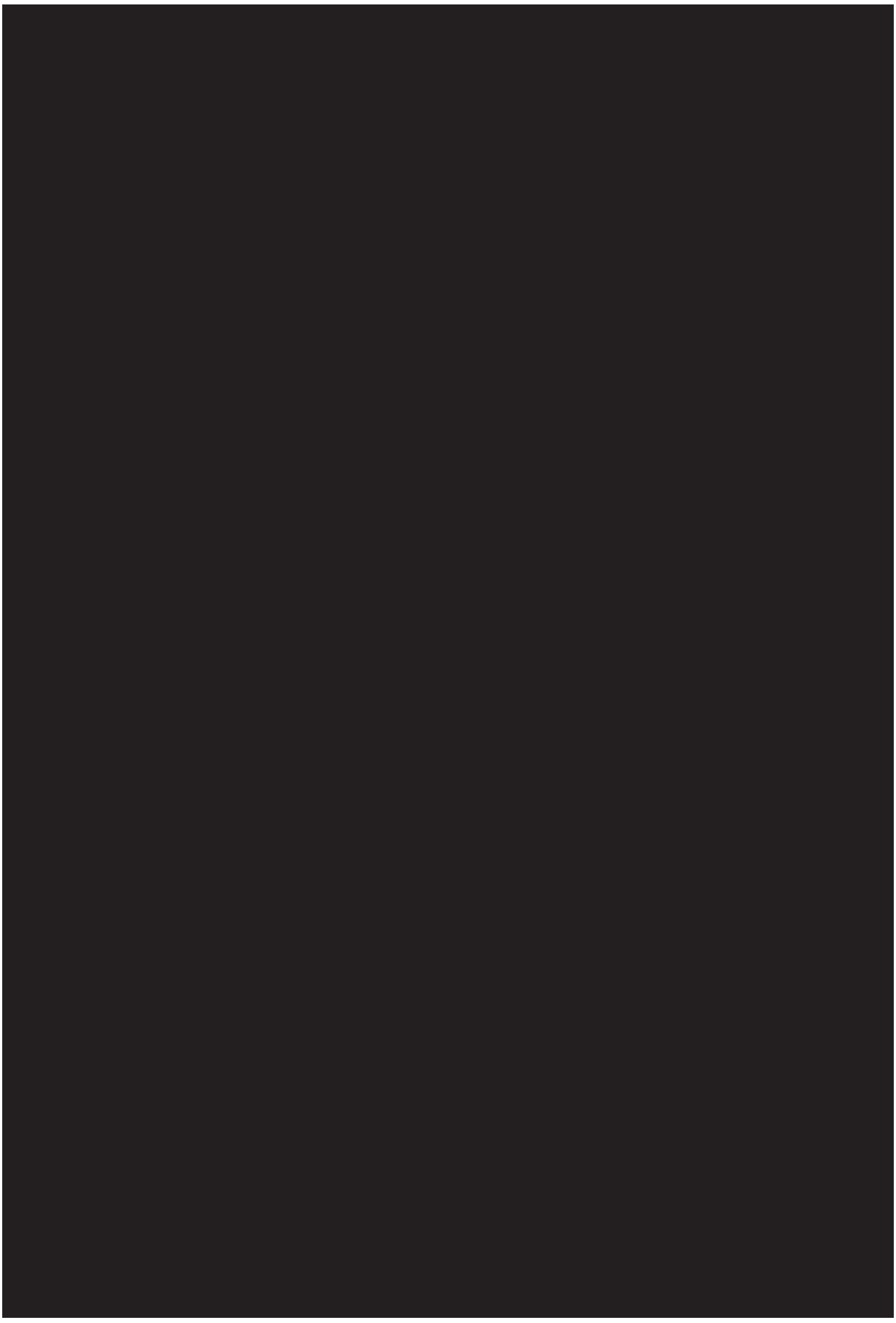


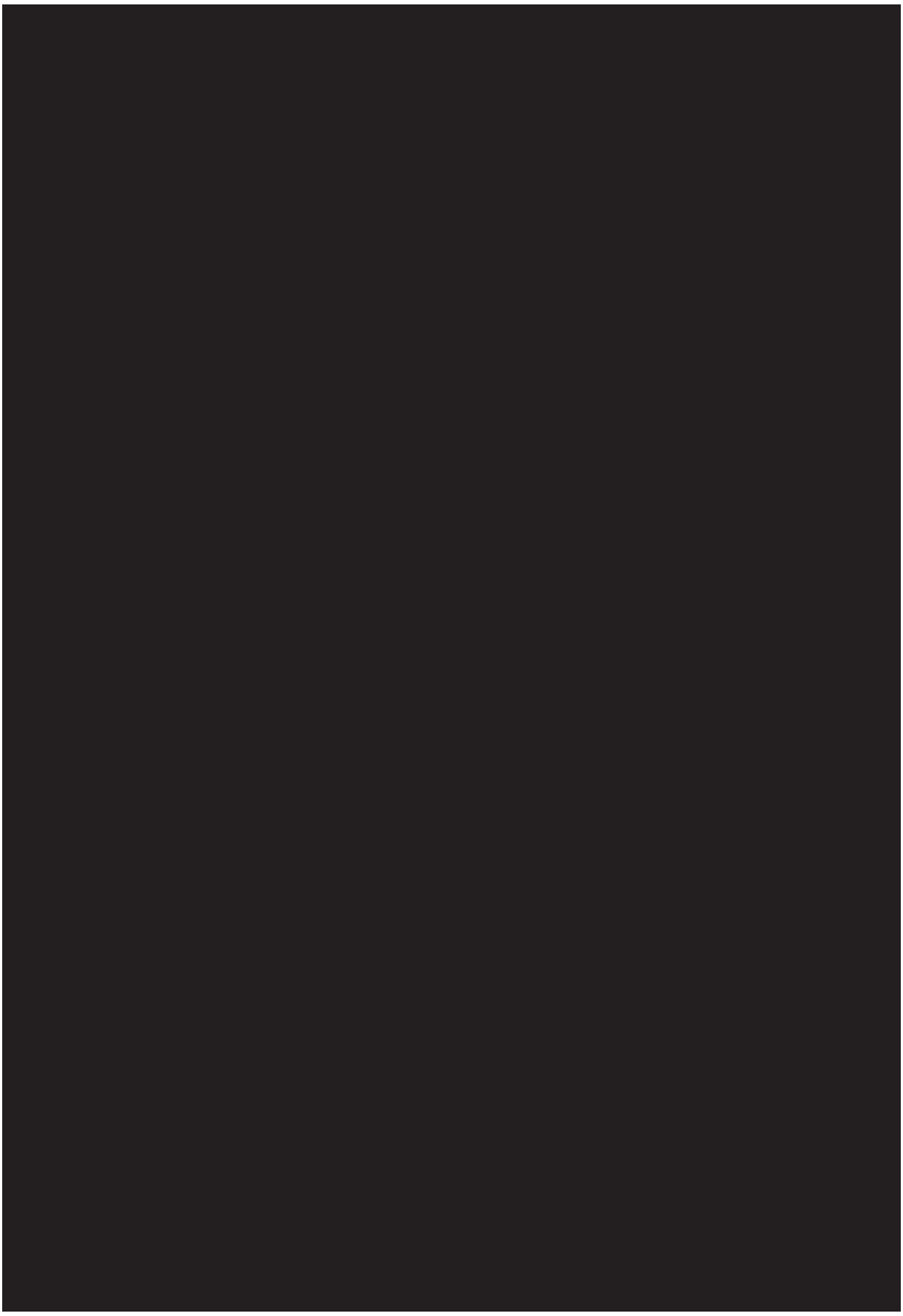


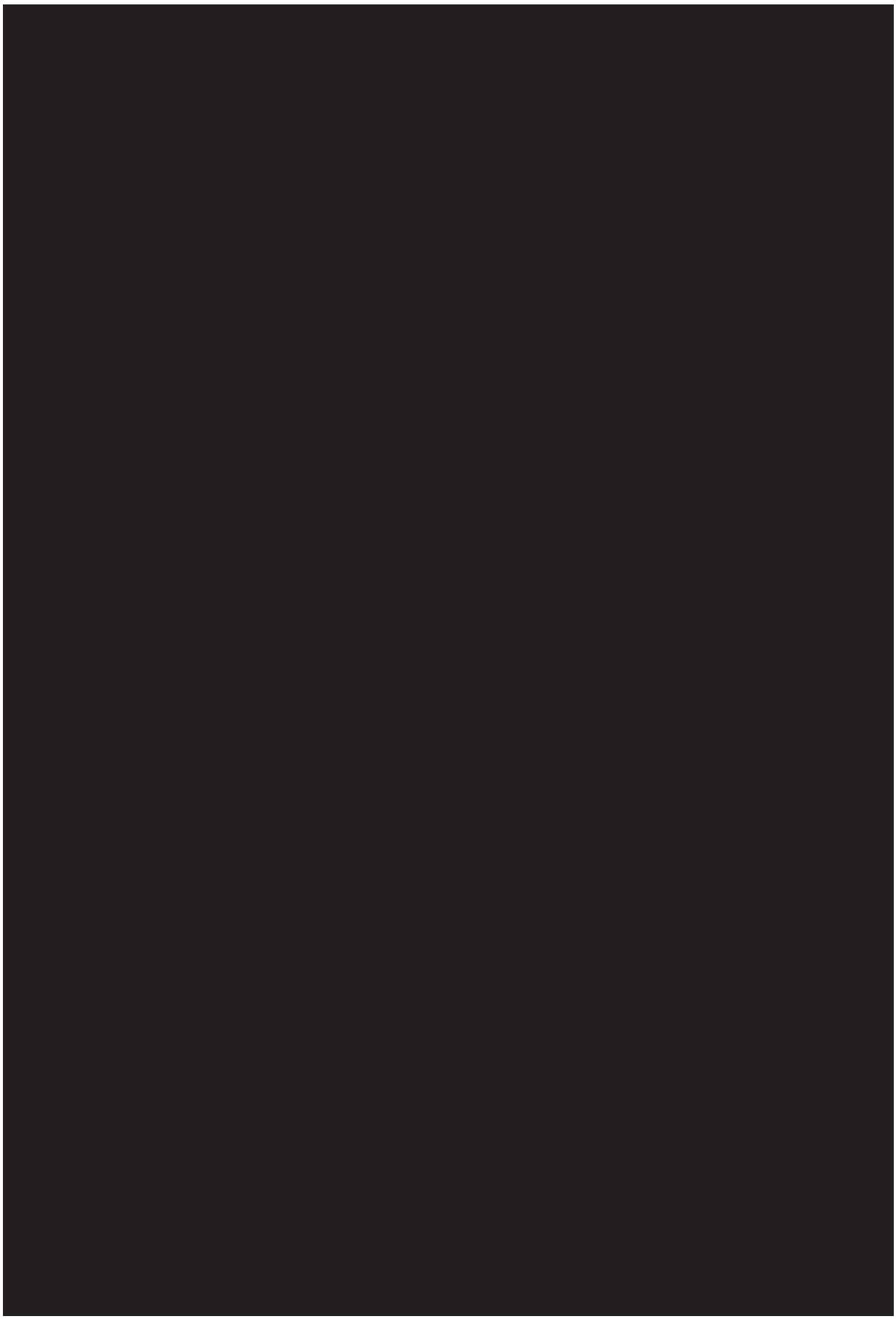


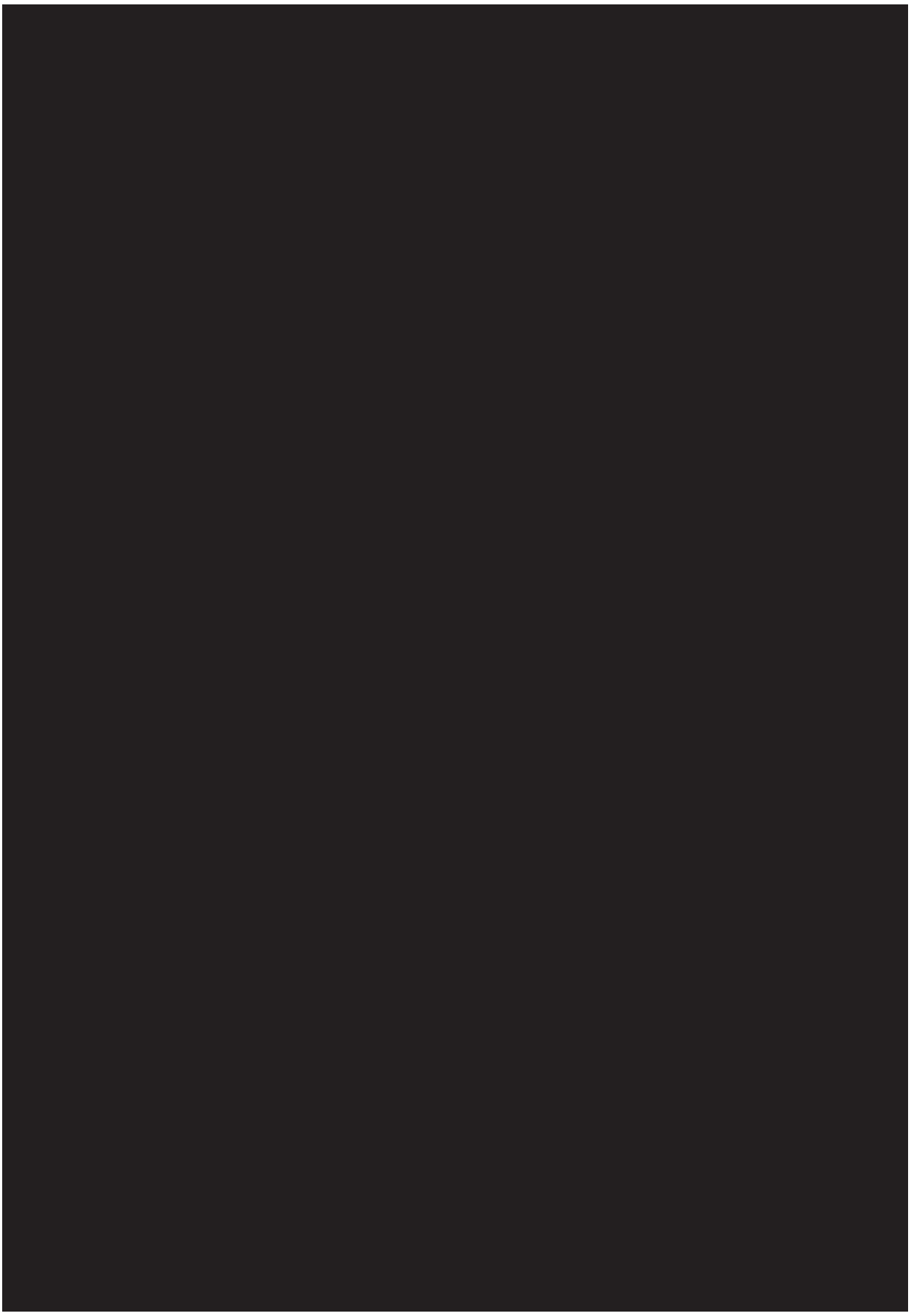












PLoS ONE, 3.73*, 60%, 1. Author

(Name of the journal, Impact factor, Contribution to this work, Authorship)

*last reported at the date of submission of this work

Supporting Information

DrugScore^{PPI} Knowledge-Based Potentials Used as Scoring and Objective Function in Protein-Protein Docking

Dennis M. Krüger[1], José Ignacio Garzón[2], Pablo Chacón[2], Holger Gohlke[1*]

[1]Institute for Pharmaceutical and Medicinal Chemistry, Heinrich-Heine-University, Düsseldorf, Germany

[2]Rocasolano Physical Chemistry Institute, Consejo Superior de Investigaciones Científicas, Madrid, Spain

[*]Universitätsstr.1, 40225 Düsseldorf (Germany), Fax: (+49) 211-8113847, E-mail: gohlke@uni-duesseldorf.de

Supplemental Tables

Table S1: 27 targets skipped from the ZDOCK benchmark 3.0 [1].

PDB ID	Difficulty	Category[a]
1A2K	Easy	O
1ACB	Medium	E
1ATN	Difficult	O
1AZS	Easy	O
1E4K	Difficult	A
1E6E	Easy	E
1E96	Easy	O
1EFN	Easy	O
1F34	Easy	E
1FAK	Difficult	O
1FC2	Easy	O
1FQ1	Difficult	E
1GHQ	Easy	O
1GP2	Medium	O
1GRN	Medium	O
1H1V	Difficult	O
1HE8	Medium	O
1IB1	Medium	O
1IJK	Medium	E
1J2J	Easy	O
1JMO	Difficult	O
1KLU	Easy	O
1NW9	Medium	E
1Z0K	Easy	O
1Z5Y	Easy	O
2O8V	Easy	E
2OOB	Easy	O

[a] Complex category labels: E = Enzyme/inhibitor or enzyme/substrate, A = Antigen-antibody, AB = Antigen-bound antibody, O = Others.

Table S2: Computational efficiency of docking with DrugScore^{PPI}/FRODOCK and the original FRODOCK implementation.[a]

Runtime [b]		PDB ID	no. of atoms			max. diameter [c]
DrugScore^{PPI}/ FRODOCK	Original FRODOCK		complex	receptor	ligand	
18	15	1ACB	2291	1768	522	48
23	9	2PCC	3218	2371	847	55
330	130	1AHW	4916	3304	1612	76

[a] Using 16 cores on dual CPU compute servers equipped with 2 GHz Intel Xeon Quadcore CPUs, 24 GB of RAM, and Infiniband interconnect.

[b] In min.

[c] Maximum diameter of the receptor, which influences the size of the search space. In Å.

Table S3: DrugScore^{PPI} potential maps and corresponding Sybyl atomtypes.[a]

Potential map	Considered atomtype(s)[a]	Definition
C_2	C.2	sp^2 carbon
C_3	C.3	sp^3 carbon
C_ar	C.ar	aromatic carbon
C_cat	C.cat	carbocation (C+) in a guanidinium group
N_3	N.2, N.3, N.4	nitrogen sp^2, sp^3, sp^3 positively charged
N_am	N.am	amide nitrogen
N_pl3	N.pl3	trigonal planar nitrogen
O_2	O.2	sp^2 oxygen
O_3	O.3	sp^3 oxygen
O_co2	O.co2	oxygen in a carboxylate group
S_3	S.3	sp^3 sulfur

[a] Sybyl atomtype notation from Tripos, available at http://tripos.com/mol2/atom_types.html.

Table S4: Distribution of the decoy quality in the "unbound perturbation" dataset.

PDB ID[a]	< 2 Å [b]	< 5 Å [b]	< 10 Å [b]	≥ 10 Å [b]
1A0O	0.00	0.60	27.76	72.24
1ACB	0.10	6.00	25.70	74.30
1AHW	0.00	1.23	14.93	85.07
1ATN	4.60	9.91	28.60	71.40
1AVW	0.00	5.44	29.91	70.09
1AVZ	0.00	0.70	15.72	84.28
1BQL	0.70	5.73	20.32	79.68
1BRC	1.00	5.00	29.60	70.40
1CGI	0.40	8.21	27.63	72.37
1BRS	0.10	4.30	35.14	64.86
1CHO	0.10	6.30	30.50	69.50
1BTH	0.00	0.00	17.90	82.10
1BVK	0.00	1.31	23.84	76.16
1DFJ	0.00	2.42	20.14	79.86
1CSE	0.00	8.08	33.23	66.77
1DQJ	0.00	0.30	16.28	83.72
1EFU	0.00	0.20	6.53	93.47
1FBI	0.00	6.23	25.83	74.17
1EO8	0.00	0.60	15.88	84.12
1FIN	0.00	0.00	5.23	94.77
1FQ1	0.00	0.60	22.61	77.39
1FSS	0.40	14.70	31.50	68.50
1GLA	0.10	4.20	32.23	67.77
1GOT	0.00	0.20	9.18	90.82
1IAI	0.00	4.00	15.40	84.60
1IGC	0.70	2.30	31.30	68.70
1JHL	0.00	1.70	27.40	72.60
1MAH	0.71	21.01	37.17	62.83
1MDA	0.61	8.99	30.10	69.90
1MEL	0.00	5.04	35.75	64.25
1MLC	0.40	7.16	28.15	71.85
1NCA	1.03	10.70	30.66	69.34
2BTF	0.61	9.11	29.96	70.04
1NMB	6.28	19.13	33.20	66.80
1PPE	4.92	17.47	48.59	51.41
1QFU	0.30	9.50	28.01	71.99
1TAB	0.10	16.78	34.87	65.13
1SPB	5.91	13.21	32.53	67.47
1TGS	2.02	6.98	30.16	69.84
1STF	12.63	21.44	32.57	67.43
1UDI	0.80	20.26	42.23	57.77
2JEL	0.20	6.90	42.50	57.50
1UGH	3.61	17.23	38.98	61.02
2KAI	0.80	19.90	41.00	59.00
1WEJ	0.00	2.22	24.65	75.35
1WQ1	0.00	5.20	34.15	65.85
2PCC	0.00	1.91	28.74	71.26
2PTC	1.61	5.83	29.05	70.95

3HHR	0.00	0.20	10.61	89.39
2SIC	0.50	17.39	33.67	66.33
2SNI	0.40	9.85	29.95	70.05
2TEC	4.21	13.04	31.90	68.10
2VIR	0.70	6.32	37.81	62.19
4HTC	0.30	10.91	28.13	71.87
Average	**1.05**	**7.48**	**27.88**	**72.12**

[a] Dataset from Baker and coworkers [2].
[b] Percentage of decoys that are within the given all-atom rmsd range.

Table S5: Distribution of the decoy quality in the "unbound docking" dataset.[a]

PDB ID[b]	< 2 Å [c]	< 5 Å [c]	< 10 Å [c]	≥ 10 Å [c]
1A0O	0.00	0.50	19.00	81.00
1ACB	*0.00*	*0.50*	*3.00*	*97.00*
1AHW	0.00	0.00	5.50	94.50
1ATN	2.50	4.50	7.00	93.00
1AVW	0.00	0.50	7.50	92.50
1AVZ	*0.00*	*0.00*	*1.00*	*99.00*
1BQL	2.50	10.50	17.50	82.50
1BRC	*0.00*	*0.00*	*0.00*	*100.00*
1CGI	0.50	10.00	20.50	79.50
1BRS	0.00	1.00	15.00	85.00
1CHO	0.00	3.00	20.00	80.00
1BTH	*0.00*	*0.00*	*0.00*	*100.00*
1BVK	0.00	0.00	75.70	24.30
1DFJ	*0.00*	*0.00*	*0.00*	*100.00*
1CSE	0.00	0.00	9.50	90.50
1DQJ	*0.00*	*0.00*	*1.00*	*99.00*
1EFU	*0.00*	*0.00*	*0.50*	*99.50*
1FBI	*0.00*	*1.50*	*1.50*	*98.50*
1EO8	*0.00*	*0.00*	*0.00*	*100.00*
1FIN	*0.00*	*0.00*	*0.00*	*100.00*
1FQ1	*0.00*	*0.00*	*1.00*	*99.00*
1FSS	0.00	3.00	3.50	96.50
1GLA	*0.00*	*0.00*	*0.00*	*100.00*
1GOT	*0.00*	*0.00*	*0.00*	*100.00*
1IAI	*0.00*	*0.00*	*0.00*	*100.00*
1IGC	*0.00*	*0.00*	*0.00*	*100.00*
1JHL	0.00	0.00	6.12	93.88
1MAH	0.50	7.00	7.50	92.50
1MDA	*0.00*	*0.00*	*0.00*	*100.00*
1MEL	0.00	5.00	20.00	80.00
1MLC	0.50	3.50	4.50	95.50
1NCA	*0.00*	*0.00*	*0.00*	*100.00*
2BTF	*0.00*	*0.00*	*0.00*	*100.00*
1NMB	*0.00*	*0.00*	*0.00*	*100.00*
1PPE	7.50	24.50	83.00	17.00
1QFU	*0.00*	*1.00*	*1.50*	*98.50*
1TAB	0.00	12.50	22.50	77.50
1SPB	3.00	4.50	7.50	92.50
1TGS	0.50	6.50	25.50	74.50
1STF	4.50	8.00	9.00	91.00
1UDI	0.00	5.35	10.16	89.84
2JEL	2.50	16.50	41.50	58.50
1UGH	3.00	18.00	35.50	64.50
2KAI	1.50	30.50	41.50	58.50
1WEJ	*0.00*	*0.00*	*0.00*	*100.00*
1WQ1	0.00	0.00	4.00	96.00
2PCC	*0.00*	*0.00*	*0.00*	*100.00*
2PTC	0.50	2.00	3.50	96.50

3HHR	*0.00*	*0.00*	*0.00*	*100.00*
2SIC	0.00	7.50	9.00	91.00
2SNI	1.00	3.00	10.00	90.00
2TEC	2.50	8.00	11.50	88.50
2VIR	*0.00*	*0.00*	*0.00*	*100.00*
4HTC	0.00	5.00	20.00	80.00
Average[d]	**1.10 (0.00)**	**6.68 (0.13)**	**19.10 (0.40)**	**80.90 (99.60)**

[a] 24 targets where 97% or more of the decoys have an all-atom rmsd > 10Å are marked in grey. For 17 out of these 24 decoys no solution with an all-atom rmsd < 10Å was generated. For 7 decoys only one solution with an all-atom rmsd < 10Å was found in the whole decoy set.
[b] Dataset from Baker and coworkers [2].
[c] Percentage of decoys that are within the given all-atom rmsd range.
[d] Numbers not in brackets refer to decoy sets where at least two decoys with an all-atom rmsd < 10Å are available; numbers in brackets consider decoy sets where less than two decoys with rmsd < 10Å are available.

Table S6: Results from rescoring "unbound perturbation" decoys with DrugScore^{PPI}.

PDB ID[a]	N10Å[b]		Best rmsd[c]	
	This work	Baker and coworkers	This work	Baker and coworkers
1A0O	1	1	8.35	8.22
1ACB	0	2	11.14	9.08
1AHW	3	5	7.40	6.17
1ATN	5	5	2.37	0.96
1AVW	5	5	5.14	6.11
1AVZ	0	0	10.12	10.14
1BQL	1	5	1.57	1.40
1BRC	4	1	3.60	3.76
1BRS	6	4	8.42	3.89
1BTH	6	0	5.15	18.23
1BVK	3	5	6.70	6.25
1CGI	6	4	3.24	3.15
1CHO	1	3	5.25	5.61
1CSE	4	2	6.11	8.65
1DFJ	6	4	4.30	3.92
1DQJ	3	2	6.65	5.86
1EFU	0	0	19.31	10.64
1EO8	0	1	12.35	7.38
1FBI	2	3	3.58	2.71
1FIN	0	0	14.82	16.56
1FQ1	4	2	7.34	9.61
1FSS	6	5	4.48	2.98
1GLA	0	1	15.17	5.95
1GOT	0	0	16.60	19.27
1IAI	4	0	3.96	14.11
1IGC	4	2	1.92	1.92
1JHL	3	1	6.11	8.56
1MAH	6	5	4.60	2.36
1MDA	2	3	3.00	7.86
1MEL	4	5	4.43	5.38
1MLC	0	0	11.16	18.89
1NCA	6	5	1.24	1.24
1NMB	1	5	2.66	0.65
1PPE	6	5	1.38	0.61
1QFU	0	5	14.28	2.00
1SPB	6	5	0.88	0.91
1STF	5	5	1.57	1.31
1TAB	4	5	3.37	3.67
1TGS	5	5	0.62	1.39
1UDI	5	5	2.08	1.16
1UGH	6	5	1.68	1.69
1WEJ	4	0	6.83	10.65
1WQ1	2	3	5.81	3.99
2BTF	2	4	2.40	1.90
2JEL	6	5	4.71	4.71
2KAI	6	4	1.21	2.04
2PCC	4	3	8.76	7.28

2PTC	4	2	5.20	0.98
2SIC	6	5	2.64	2.41
2SNI	6	4	1.61	1.61
2TEC	5	5	1.97	2.25
2VIR	3	4	4.99	6.12
3HHR	0	0	12.53	10.75
4HTC	6	5	3.64	3.81
Totals	**31**[d]	**34**[d]	**44**[e]	**45**[e]

[a] Dataset from Baker and coworkers [2].
[b] No of decoys that have all-atom rmsd < 10Å in the top 5.
[c] Best all-atom rmsd in the top 5. In Å.
[d] Totals count the number of targets that have at least three solutions with an all-atom rmsd < 10Å in the top 5.
[e] Totals count the number of targets that have an all-atom rmsd < 10Å in the top 5.

Table S7: Results from rescoring "unbound docking" decoys with DrugScore^{PPI}.

PDB ID[a]	Best rmsd[b]	
	This work	Baker and coworkers
1A0O	5.08	5.08
1ACB	12.29	11.84
1AHW	7.57	7.24
1ATN	2.3	2.30
1AVW	6.47	5.76
1AVZ	19.84	20.03
1BQL	1.24	2.09
1BRC	15.96	15.96
1BRS	6.59	8.67
1BTH	17.37	16.96
1BVK	6.32	6.15
1CGI	3.88	3.27
1CHO	4.82	6.88
1CSE	8.59	8.87
1DFJ	20.59	22.09
1DQJ	17.75	17.75
1EFU	29.51	29.51
1EO8	29.43	56.51
1FBI	20.99	22.25
1FIN	27.15	26.84
1FQ1	25.59	24.64
1FSS	2.34	3.19
1GLA	53.04	25.66
1GOT	55.46	55.84
1IAI	34.04	33.96
1IGC	28.09	24.19
1JHL	9.4	14.91
1MAH	2.4	2.40
1MDA	54.18	53.71
1MEL	3.56	5.80
1MLC	6.52	2.52
1NCA	31.28	31.47
1NMB	72.82	72.48
1PPE	3.66	0.59
1QFU	28.55	3.20
1SPB	0.9	1.32
1STF	2.32	0.56
1TAB	3.78	3.35
1TGS	4.48	3.31
1UDI	3.75	3.75
1UGH	1.28	2.15
1WEJ	17.9	17.90
1WQ1	5.08	44.03
2BTF	11.97	27.52
2JEL	3.42	1.45
2KAI	3.09	2.40
2PCC	18.75	19.07

2PTC	9.69	10.35
2SIC	4.93	2.98
2SNI	1.36	1.36
2TEC	0.87	2.09
2VIR	56.65	41.47
3HHR	33.95	33.90
4HTC	4.8	3.84
R10Å[c]	**24 (6)**	**22 (6)**
R5Å[d]	**17 (5)**	**15 (5)**

[a] Dataset from Baker and coworkers [2].
[b] Best all-atom rmsd in the top 10. In Å. Decoys were clustered according to Baker and coworkers [2].
[c] Number of targets that have at least one solution with all-atom rmsd < 10Å in the first solutions of the top 10 clusters. Numbers not in brackets refer to decoy sets where at least two decoys with rmsd < 10Å are available; numbers in brackets refer to decoy sets where less than two decoys with rmsd < 10Å are available.
[d] Number of targets that have at least one solution with all-atom rmsd < 5Å in the first solutions of the top 10 clusters. Numbers not in brackets refer to decoy sets where at least two decoys with rmsd < 10Å are available; numbers in brackets refer to decoy sets where less than two decoys with rmsd < 10Å are available.

Table S8: Protein-protein complexes where multiple ligand binding modes have to be considered for rmsd calculations.

Difficulty	Category[a]	PDB ID
Easy	E	1EZU
Easy	O	1F51
Easy	AB	1I9R
Easy	AB	1QFW
Easy	O	1RLB
Easy	E	2PCC
Easy	O	1SBB
Medium	E	1KKL
Medium	O	1N2C
Medium	O	2H7V
Difficult	O	1DE4

[a] Complex category labels: E = Enzyme/inhibitor or enzyme/substrate, A = Antigen-antibody, AB = Antigen-bound antibody, O = Others.

Table S9: Results for bound docking with the DrugScore^{PPI}/FRODOCK approach.[a]

Difficulty	Category[b]	PDB ID[c]	Quality_1[d]	Quality_2[e]	Quality_3[f]	i_rmsd[g]	l_rmsd[h]	f_{nat}[i]	f_{not}[j]
Easy	A	1AHW	-	260	260	1.58	2.19	0.92	0.25
Easy	O	1AK4	-	91	91	1.43	3.73	0.83	0.18
Easy	O	1AKJ	-	21	21	2.11	4.34	0.85	0.32
Easy	E	1AVX	-	1	1	1.27	2.96	0.94	0.26
Easy	E	1AY7	-	1	1	1.09	1.31	0.96	0.16
Easy	O	1B6C	-	1	1	1.53	1.63	0.97	0.25
Medium	A	1BGX	-	1	1	2.45	4.72	0.82	0.26
Easy	AB	1BJ1	4	4	4	0.84	2.02	0.93	0.05
Difficult	O	1BKD	1	1	1	0.48	0.57	0.92	0.03
Easy	O	1BUH	-	7	7	1.12	1.06	0.83	0.04
Easy	A	1BVK	10	10	10	0.97	1.19	0.92	0.18
Easy	E	1BVN	1	1	1	0.39	0.41	0.96	0.12
Easy	E	1CGI	1	1	1	0.92	0.94	0.82	0.03
Easy	E	1D6R	1	1	1	0.62	0.56	0.94	0.03
Difficult	O	1DE4	-	803	803	1.55	2.50	0.97	0.27
Easy	E	1DFJ	-	1	1	1.44	1.70	0.86	0.24
Easy	A	1DQJ	1	1	1	1.03	1.06	0.97	0.27
Easy	A	1E6J	-	-	-	-	-	-	-
Easy	E	1EAW	-	1	1	1.18	1.39	0.95	0.14
Difficult	O	1EER	1	1	1	0.72	0.66	0.90	0.03
Easy	E	1EWY	-	943	2	9.60	9.68	0.18	0.89
Easy	E	1EZU	1	1	1	0.72	0.59	0.93	0.08
Easy	O	1F51	-	1	1	1.31	1.08	0.85	0.15
Easy	O	1FQJ	-	3	3	1.20	1.50	0.88	0.11
Easy	AB	1FSK	-	1	1	1.09	1.15	0.93	0.18
Easy	O	1GCQ	-	1	1	1.59	1.52	0.93	0.23
Easy	O	1GLA	-	-	-	-	-	-	-
Easy	O	1GPW	-	1	1	1.56	2.12	0.85	0.23
Easy	O	1HE1	1	1	1	0.59	0.69	0.89	0.02
Easy	E	1HIA	1	1	1	0.74	1.07	0.91	0.14
Medium	O	1I2M	-	1	1	1.16	1.16	0.91	0.18
Easy	O	1I4D	-	-	-	-	-	-	-
Easy	AB	1I9R	-	-	-	-	-	-	-
Difficult	O	1IBR	-	1	1	1.08	1.08	0.91	0.10
Easy	AB	1IQD	-	1	1	1.15	1.16	0.85	0.15
Difficult	O	1IRA	-	1	1	1.41	1.44	0.87	0.27
Easy	A	1JPS	-	8	8	1.57	1.55	0.95	0.34
Easy	AB	1K4C	-	2	2	1.09	1.06	0.90	0.15
Medium	O	1K5D	-	9	9	1.61	2.14	0.83	0.16
Easy	O	1K74	-	2	2	1.51	2.78	0.82	0.19
Easy	O	1KAC	-	184	81	7.17	9.56	0.31	0.78
Medium	E	1KKL	-	12	12	1.21	2.04	0.89	0.19
Easy	O	1KTZ	-	491	491	3.78	6.81	0.69	0.62
Easy	O	1KXP	-	3	3	1.81	3.21	0.80	0.13
Easy	AB	1KXQ	-	1	1	1.13	1.13	0.95	0.25
Medium	E	1M10	-	2	2	2.12	3.24	0.80	0.12
Easy	E	1MAH	-	1	1	1.72	2.20	0.86	0.24

Easy	O	1ML0	-	2	2	1.39	1.34	0.96	0.23
Easy	A	1MLC	18	18	18	1.00	1.09	0.93	0.16
Medium	O	1N2C	1	1	1	0.95	0.95	0.88	0.23
Easy	E	1N8O	-	1	1	1.50	3.10	0.92	0.15
Easy	AB	1NCA	-	3	3	2.43	4.25	0.75	0.19
Easy	AB	1NSN	-	250	246	8.07	9.97	0.31	0.80
Easy	E	1OPH	-	4	4	1.95	3.45	0.91	0.18
Easy	E	1PPE	-	1	1	1.99	1.95	0.89	0.31
Difficult	E	1PXV	-	1	1	1.61	2.79	0.88	0.19
Easy	O	1QA9	-	988	469	9.24	9.85	0.21	0.83
Easy	AB	1QFW	-	18	18	1.45	1.29	0.93	0.21
Easy	E	1R0R	-	1	1	1.57	4.96	0.92	0.24
Difficult	O	1R8S	-	1	1	3.75	5.60	0.78	0.27
Easy	O	1RLB	-	-	609	7.71	9.83	0.30	0.76
Easy	O	1S1Q	28	28	28	0.96	0.94	0.94	0.19
Easy	O	1SBB	-	-	-	-	-	-	-
Easy	O	1T6B	-	3	3	1.77	2.76	0.85	0.24
Easy	E	1TMQ	-	1	1	1.75	1.73	0.97	0.38
Easy	E	1UDI	-	1	1	1.43	1.74	0.95	0.22
Easy	A	1VFB	9	9	9	0.72	0.88	0.88	0.06
Easy	A	1WEJ	312	15	15	1.80	3.72	0.79	0.22
Medium	O	1WQ1	1	1	1	0.92	1.44	0.91	0.08
Easy	O	1XD3	-	1	1	5.71	5.47	0.91	0.16
Medium	O	1XQS	-	1	1	1.84	4.35	0.79	0.17
Difficult	O	1Y64	-	40	40	1.50	2.68	0.88	0.32
Easy	E	1YVB	2	1	1	2.25	5.96	0.75	0.35
Easy	O	1ZHI	-	65	65	1.15	1.18	0.96	0.29
Easy	O	2AJF	-	644	155	6.81	8.03	0.47	0.71
Easy	E	2B42	-	1	1	1.11	1.48	0.96	0.24
Easy	O	2BTF	-	1	1	1.30	1.45	0.71	0.06
Difficult	O	2C0L	-	1	1	2.43	4.14	0.95	0.27
Medium	O	2CFH	-	1	1	1.67	1.91	0.93	0.31
Easy	A	2FD6	-	666	55	8.93	9.54	0.25	0.82
Medium	O	2H7V	-	79	79	1.08	1.37	0.86	0.06
Easy	O	2HLE	1	1	1	0.85	5.50	0.94	0.12
Difficult	AB	2HMI	-	-	-	-	-	-	-
Easy	O	2HQS	-	1	1	1.51	3.41	0.97	0.20
Medium	O	2HRK	6	6	6	0.89	1.00	0.92	0.02
Easy	A	2I25	1	1	1	0.67	0.94	0.95	0.03
Easy	AB	2JEL	-	1	1	1.30	1.80	0.85	0.10
Easy	E	2MTA	-	105	105	1.66	2.55	0.88	0.38
Medium	O	2NZ8	-	1	1	1.23	1.18	0.93	0.22
Difficult	O	2OT3	-	1	1	1.12	1.12	0.94	0.09
Easy	E	2PCC	-	-	226	6.93	9.83	0.44	0.73
Easy	AB	2QFW	1	1	1	0.98	1.13	0.96	0.15
Easy	E	2SIC	-	1	1	1.96	2.48	0.83	0.24
Easy	E	2SNI	-	1	1	1.09	6.01	0.92	0.28
Easy	E	2UUY	-	1944	1944	3.00	9.00	0.77	0.35
Easy	A	2VIS	1	1	1	1.05	1.58	0.91	0.10
Easy	E	7CEI	-	260	260	1.58	2.19	0.92	0.25

[a] Subset of 97 structures from the ZDOCK benchmark 3.0.
[b] Complex category labels: E = Enzyme/inhibitor or enzyme/substrate, A = Antigen-antibody, AB = Antigen-bound antibody, O = Others.
[c] Subset of 27 structures skipped from the ZDOCK benchmark 3.0. For details see Materials and Methods section.
[d] Rank of the first solution that has a high accuracy. The lower the rank, the better is the prediction. "-" denotes that no solution was found.
[e] Rank of the first solution that has at least medium accuracy. The lower the rank, the better is the prediction. "-" denotes that no solution was found.
[f] Rank of the first solution that has at least acceptable accuracy. The lower the rank, the better is the prediction. "-" denotes that no solution was found. Additional details for this case are given in the subsequent columns.
[g]: All-atom interface RMSD of the docked ligand and its bound crystal structure conformation.
[h]: All-atom RMSD of the docked ligand and its bound crystal structure conformation.
[i]: Fraction of native residue-residue contacts of the docked ligand conformation compared to its bound crystal structure conformation.
[j]: Fraction of non-native residue-residue contacts of the docked ligand conformation compared to its bound crystal structure conformation.

Table S10: Results for bound docking with the original FRODOCK implementation.[a]

Difficulty	Category	PDB ID	Quality_1	Quality_2	Quality_3	i_rmsd	l_rmsd	f_{nat}	f_{not}
Easy	A	1AHW	-	10	10	1.71	1.89	0.82	0.19
Easy	O	1AK4	-	546	546	7.47	9.54	0.73	0.45
Easy	O	1AKJ	-	665	665	3.48	5.61	0.68	0.37
Easy	E	1AVX	-	1	1	2.12	3.56	0.97	0.26
Easy	E	1AY7	-	1	1	1.63	2.19	0.88	0.18
Easy	O	1B6C	-	1	1	3.16	3.73	0.89	0.25
Medium	A	1BGX	-	1	1	2.02	2.16	0.82	0.25
Easy	AB	1BJ1	-	6	6	1.29	2.95	0.83	0.13
Difficult	O	1BKD	1	1	1	0.88	0.97	0.85	0.08
Easy	O	1BUH	27	27	27	0.81	0.87	0.94	0
Easy	A	1BVK	-	24	24	2.11	2.46	0.69	0.08
Easy	E	1BVN	-	1	1	2.99	3.36	0.78	0.23
Easy	E	1CGI	1	1	1	0.91	0.94	0.87	0.03
Easy	E	1D6R	-	27	27	1.91	2.95	0.88	0.1
Difficult	O	1DE4	23	23	23	0.89	1.14	0.91	0.04
Easy	E	1DFJ	-	329	52	6.32	6.41	0.45	0.77
Easy	A	1DQJ	-	12	12	1.24	1.54	0.85	0.15
Easy	A	1E6J	-	-	-	-	-	-	-
Easy	E	1EAW	-	1	1	2.09	2.15	0.91	0.35
Difficult	O	1EER	-	1	1	2.01	2.52	0.82	0.22
Easy	E	1EWY	-	3	3	2.96	2.91	0.83	0.5
Easy	E	1EZU	-	2	2	2.02	2.55	0.8	0.17
Easy	O	1F51	-	4	4	1.59	1.87	0.89	0.11
Easy	O	1FQJ	-	119	74	5.21	6.64	0.44	0.55
Easy	AB	1FSK	-	1	1	1.95	2.88	0.72	0.23
Easy	O	1GCQ	-	20	20	1.67	2.03	0.9	0.21
Easy	O	1GLA	-	89	89	5.1	6	0.6	0.38
Easy	O	1GPW	-	1	1	3.59	5.9	0.71	0.55
Easy	O	1HE1	-	1	1	1.75	1.63	0.65	0.08
Easy	E	1HIA	-	1	1	1.35	1.42	0.94	0.21
Medium	O	1I2M	-	1	1	1.72	1.89	0.77	0.09
Easy	O	1I4D	-	-	-	-	-	-	-
Easy	AB	1I9R	-	-	-	-	-	-	-
Difficult	O	1IBR	-	1	1	1.56	1.82	0.91	0.15
Easy	AB	1IQD	-	6	6	1.56	2.34	0.89	0.11
Difficult	O	1IRA	-	1	1	2.17	2.38	0.78	0.31
Easy	A	1JPS	-	11	11	2.41	3	0.9	0.27
Easy	AB	1K4C	-	499	499	1.35	2.31	0.81	0.09
Medium	O	1K5D	1	1	1	1.05	1.09	0.91	0.09
Easy	O	1K74	-	3	3	5.27	9.31	0.71	0.55
Easy	O	1KAC	51	20	20	2.68	4.6	0.8	0.18
Medium	E	1KKL	16	16	16	0.98	1.08	0.95	0.08
Easy	O	1KTZ	-	158	158	2.32	3.19	0.9	0.38
Easy	O	1KXP	-	1	1	1.59	2.65	0.86	0.12
Easy	AB	1KXQ	-	1	1	2.01	2.7	0.89	0.25
Medium	E	1M10	-	1	1	2.01	2.55	0.93	0.19
Easy	E	1MAH	-	1	1	1.89	1.93	0.92	0.26
Easy	O	1ML0	-	2	2	2.31	2.54	0.85	0.2

Easy	A	1MLC	-	4	4	1.99	3.32	0.82	0.25
Medium	O	1N2C	-	3	3	2.26	3.14	0.71	0.08
Easy	E	1N8O	-	1	1	2.14	3.14	0.91	0.16
Easy	AB	1NCA	-	28	28	1.32	2.45	0.82	0.11
Easy	AB	1NSN	-	160	160	1.87	1.95	0.78	0.28
Easy	E	1OPH	-	51	51	2.56	3.48	0.87	0.11
Easy	E	1PPE	1	1	1	0.82	0.83	0.96	0.15
Difficult	E	1PXV	-	1	1	1.37	1.76	0.9	0.11
Easy	O	1QA9	-	-	-	-	-	-	-
Easy	AB	1QFW	-	186	186	1.68	2.77	0.92	0.24
Easy	E	1R0R	-	2	2	4.49	4.88	0.93	0.14
Difficult	O	1R8S	-	6	4	7.43	8.11	0.43	0.37
Easy	O	1RLB	-	728	208	7.06	8.83	0	1
Easy	O	1S1Q	-	153	153	4.27	5.06	0.63	0.45
Easy	O	1SBB	-	-	-	-	-	-	-
Easy	O	1T6B	-	26	26	1.98	2.61	0.96	0.34
Easy	E	1TMQ	-	1	1	1.52	1.55	0.98	0.29
Easy	E	1UDI	-	1	1	1.56	1.75	0.83	0.06
Easy	A	1VFB	-	24	24	1.63	2.08	0.75	0.05
Easy	A	1WEJ	-	80	80	3.32	5.1	0.82	0.29
Medium	O	1WQ1	-	-	36	5.12	6.05	0.4	0.5
Easy	O	1XD3	-	1	1	6.25	5.43	0.88	0.09
Medium	O	1XQS	-	20	1	1.07	1.89	0	1
Difficult	O	1Y64	-	2	2	2.91	4.37	0.78	0.52
Easy	E	1YVB	-	1	1	3.76	6.6	0.63	0.49
Easy	O	1ZHI	117	62	62	3.57	5.72	0.93	0.36
Easy	O	2AJF	-	55	55	1.14	1.46	0.85	0.18
Easy	E	2B42	-	1	1	2.84	3.66	0.78	0.22
Easy	O	2BTF	1	1	1	0.86	0.9	0.84	0.01
Difficult	O	2C0L	-	2	2	3.17	4.66	0.95	0.33
Medium	O	2CFH	-	1	1	2.01	2.68	0.9	0.26
Easy	A	2FD6	-	161	2	7.88	8.51	0.15	0.9
Medium	O	2H7V	-	82	82	1.5	2.76	0.89	0.11
Easy	O	2HLE	-	1	1	3.08	6.76	0.78	0.11
Difficult	AB	2HMI	-	1575	1575	1.79	3.52	0.91	0.07
Easy	O	2HQS	-	1	1	4.07	4.26	0.81	0.22
Medium	O	2HRK	-	19	19	1.58	1.7	0.85	0.05
Easy	A	2I25	-	1	1	1.43	1.59	0.89	0.04
Easy	AB	2JEL	32	32	32	0.88	1.02	0.91	0.12
Easy	E	2MTA	-	2	2	1.39	2.37	0.94	0.23
Medium	O	2NZ8	-	1	1	1.78	1.99	0.95	0.29
Difficult	O	2OT3	-	1	1	1.68	1.91	0.78	0.05
Easy	E	2PCC	-	30	30	9.25	8.65	0.89	0.11
Easy	AB	2QFW	-	1	1	1.55	2.79	0.94	0.23
Easy	E	2SIC	-	1	1	1.75	2.32	0.93	0.23
Easy	E	2SNI	-	2	2	3.69	5.97	0.9	0.07
Easy	E	2UUY	-	-	-	-	-	-	-
Easy	A	2VIS	-	10	10	1.73	2.99	0.83	0.14
Easy	E	7CEI	-	10	10	1.71	1.89	0.82	0.19

[a] Subset of 97 structures from the ZDOCK benchmark 3.0. See footnotes of Table S9 for further explanations.

Table S11: Results for unbound global docking with the DrugScore^{PPI}/FRODOCK approach.[a]

Difficulty	Category	PDB ID	Quality_1	Quality_2	Quality_3	i_rmsd	l_rmsd	f_{nat}	f_{not}
Easy	A	1AHW	-	-	504	8.33	9.65	0.23	0.53
Easy	O	1AK4	-	-	-	-	-	-	-
Easy	O	1AKJ	-	-	110	5.31	8.67	0.64	0.69
Easy	E	1AVX	-	6	6	1.69	3.17	0.75	0.12
Easy	E	1AY7	-	562	322	7.69	6.55	0.17	0.84
Easy	O	1B6C	-	43	43	3.6	3.58	0.74	0.48
Medium	A	1BGX	-	-	-	-	-	-	-
Easy	AB	1BJ1	-	195	195	1.79	3.29	0.79	0.06
Difficult	O	1BKD	-	-	12	6.26	5.74	0.17	0.54
Easy	O	1BUH	-	-	1136	5.73	6.4	0.36	0.59
Easy	A	1BVK	-	-	70	7.93	8.62	0.1	0.95
Easy	E	1BVN	-	64	64	2.21	2.36	0.62	0.14
Easy	E	1CGI	-	-	162	9.7	9.11	0.13	0.78
Easy	E	1D6R	-	-	79	6.4	9.62	0.33	0.68
Difficult	O	1DE4	-	-	-	-	-	-	-
Easy	E	1DFJ	-	-	4	5.68	5.52	0.28	0.78
Easy	A	1DQJ	-	-	530	5.68	7.75	0.57	0.37
Easy	A	1E6J	-	438	438	3.46	3.13	0.69	0.6
Easy	E	1EAW	-	1	1	3.47	4.33	0.76	0.29
Difficult	O	1EER	-	-	392	8.24	9.35	0.14	0.7
Easy	E	1EWY	-	524	19	7.33	7.35	0.25	0.85
Easy	E	1EZU	-	-	-	-	-	-	-
Easy	O	1F51	-	-	207	8.38	9.27	0.14	0.82
Easy	O	1FQJ	-	-	892	4.98	6.13	0.52	0.44
Easy	AB	1FSK	-	1	1	1.61	2.03	0.96	0.24
Easy	O	1GCQ	-	-	695	4.72	5.98	0.59	0.54
Easy	O	1GLA	-	-	-	-	-	-	-
Easy	O	1GPW	-	120	120	3.82	4.95	0.85	0.59
Easy	O	1HE1	-	-	21	7.66	7.84	0.18	0.79
Easy	E	1HIA	-	-	89	7.89	9.01	0.26	0.61
Medium	O	1I2M	-	-	255	5.72	6	0.22	0.38
Easy	O	1I4D	-	-	-	-	-	-	-
Easy	AB	1I9R	-	-	-	-	-	-	-
Difficult	O	1IBR	-	-	-	-	-	-	-
Easy	AB	1IQD	-	-	467	5.86	9.9	0.89	0.11
Difficult	O	1IRA	-	-	-	-	-	-	-
Easy	A	1JPS	-	-	-	-	-	-	-
Easy	AB	1K4C	-	-	-	-	-	-	-
Medium	O	1K5D	-	-	263	7.3	9.92	0.17	0.44
Easy	O	1K74	-	311	81	6.44	6.98	0.46	0.59
Easy	O	1KAC	-	1056	383	7.86	9.33	0.19	0.86
Medium	E	1KKL	-	-	-	-	-	-	-
Easy	O	1KTZ	-	-	-	-	-	-	-
Easy	O	1KXP	-	-	358	6.64	6.63	0.28	0.61
Easy	AB	1KXQ	-	-	37	3.44	5.66	0.7	0.29
Medium	E	1M10	-	-	-	-	-	-	-
Easy	E	1MAH	-	13	13	2.7	3.49	0.81	0.39
Easy	O	1ML0	-	39	39	2.47	2.86	0.77	0.34

Easy	A	1MLC	-	23	23	2.17	2.87	0.85	0.49
Medium	O	1N2C	-	-	1724	8.64	9.98	0.2	0.54
Easy	E	1N8O	-	23	23	4	4.9	0.6	0.2
Easy	AB	1NCA	-	1	1	3.03	4.94	0.84	0.25
Easy	AB	1NSN	-	650	650	4.34	4.85	0.63	0.47
Easy	E	1OPH	-	-	-	-	-	-	-
Easy	E	1PPE	-	1	1	2.41	2.55	0.86	0.27
Difficult	E	1PXV	-	-	-	-	-	-	-
Easy	O	1QA9	-	-	-	-	-	-	-
Easy	AB	1QFW	-	292	292	1.72	2.03	0.91	0.18
Easy	E	1R0R	-	87	87	1.43	1.91	0.92	0.21
Difficult	O	1R8S	-	-	-	-	-	-	-
Easy	O	1RLB	-	-	387	3.43	5.17	0.7	0.34
Easy	O	1S1Q	-	415	70	4.68	5.33	0.65	0.6
Easy	O	1SBB	-	-	-	-	-	-	-
Easy	O	1T6B	-	-	901	5.68	8.19	0.48	0.39
Easy	E	1TMQ	-	1746	1746	4.14	4.62	0.51	0.46
Easy	E	1UDI	-	-	75	5.5	5.84	0.36	0.64
Easy	A	1VFB	-	1024	122	8.61	6.84	0.16	0.87
Easy	A	1WEJ	-	649	41	7.33	8.9	0.24	0.8
Medium	O	1WQ1	-	-	154	6.84	9.29	0.28	0.6
Easy	O	1XD3	-	1036	518	8.05	9.69	0.32	0.4
Medium	O	1XQS	-	-	-	-	-	-	-
Difficult	O	1Y64	-	-	-	-	-	-	-
Easy	E	1YVB	-	1	1	2.3	4.8	0.86	0.25
Easy	O	1ZHI	-	-	1095	7.22	9.4	0.35	0.77
Easy	O	2AJF	-	-	-	-	-	-	-
Easy	E	2B42	-	30	30	2.52	3.61	0.73	0.28
Easy	O	2BTF	-	-	630	9.27	9.88	0.14	0.78
Difficult	O	2C0L	-	-	-	-	-	-	-
Medium	O	2CFH	-	178	178	2.43	2.36	0.86	0.42
Easy	A	2FD6	-	-	-	-	-	-	-
Medium	O	2H7V	-	-	-	-	-	-	-
Easy	O	2HLE	-	1077	118	6.21	7.34	0.27	0.57
Difficult	AB	2HMI	-	-	-	-	-	-	-
Easy	O	2HQS	-	-	163	5.12	7.51	0.52	0.52
Medium	O	2HRK	-	-	-	-	-	-	-
Easy	A	2I25	-	-	-	-	-	-	-
Easy	AB	2JEL	-	-	605	8.77	8.99	0.2	0.74
Easy	E	2MTA	-	-	1080	6.59	8.23	0.17	0.86
Medium	O	2NZ8	-	-	-	-	-	-	-
Difficult	O	2OT3	-	-	-	-	-	-	-
Easy	E	2PCC	-	-	1118	7.6	9.86	0.28	0.8
Easy	AB	2QFW	-	-	12	3.51	7.01	0.81	0.29
Easy	E	2SIC	-	215	5	4.56	6.61	0.66	0.45
Easy	E	2SNI	-	-	63	4.88	8.63	0.51	0.32
Easy	E	2UUY	-	-	-	-	-	-	-
Easy	A	2VIS	-	1166	95	5.45	6.22	0.42	0.54
Easy	E	7CEI	-	-	504	8.33	9.65	0.23	0.53

[a] Subset of 97 structures from the ZDOCK benchmark 3.0. See footnotes of Table S9 for further explanations.

Table S12: Results for unbound global docking with the FRODOCK approach.[a]

Difficulty	Category	PDB ID	Quality_1	Quality_2	Quality_3	i_rmsd	l_rmsd	f_{nat}	f_{not}
Easy	A	1AHW	-	288	288	1.41	1.41	0.78	0.07
Easy	O	1AK4	-	-	-	-	-	-	-
Easy	O	1AKJ	-	486	322	8.77	8.11	0.52	0.81
Easy	E	1AVX	-	11	10	3.53	6.24	0.65	0.22
Easy	E	1AY7	-	1	1	3.13	4.45	0.83	0.14
Easy	O	1B6C	-	5	5	3.31	3.33	0.69	0.43
Medium	A	1BGX	-	-	-	-	-	-	-
Easy	AB	1BJ1	-	49	49	1.44	2.51	0.67	0.08
Difficult	O	1BKD	-	-	82	6.07	5.86	0.19	0.51
Easy	O	1BUH	-	1688	476	5.44	5.65	0.37	0.5
Easy	A	1BVK	-	336	13	7.55	8.2	0.24	0.86
Easy	E	1BVN	-	2	2	2.22	2.36	0.62	0.21
Easy	E	1CGI	-	38	19	7.48	7.62	0.21	0.59
Easy	E	1D6R	-	-	96	7.14	8.82	0.14	0.85
Difficult	O	1DE4	-	1597	939	2.99	7.51	0.5	0.43
Easy	E	1DFJ	-	2	1	4.88	5.7	0.66	0.57
Easy	A	1DQJ	-	-	-	-	-	-	-
Easy	A	1E6J	-	17	10	8.94	8.45	0.23	0.82
Easy	E	1EAW	-	41	41	3.26	4.15	0.79	0.24
Difficult	O	1EER	-	-	191	7.18	8.43	0.35	0.58
Easy	E	1EWY	-	1	1	2.65	2.55	0.61	0.48
Easy	E	1EZU	-	197	197	3.37	3.21	0.49	0.13
Easy	O	1F51	-	83	83	4.01	4.01	0.66	0.47
Easy	O	1FQJ	-	-	437	4.52	5.58	0.57	0.35
Easy	AB	1FSK	-	1	1	2.35	2.68	0.85	0.29
Easy	O	1GCQ	-	-	157	6.75	7.98	0.24	0.76
Easy	O	1GLA	-	327	327	4.36	4.72	0.68	0.42
Easy	O	1GPW	-	1	1	2.78	4.04	0.83	0.53
Easy	O	1HE1	-	-	17	8.09	9.23	0.22	0.84
Easy	E	1HIA	-	-	3	8.08	8.79	0.11	0.88
Medium	O	1I2M	-	-	129	4.45	4.83	0.29	0.2
Easy	O	1I4D	-	-	-	-	-	-	-
Easy	AB	1I9R	-	-	-	-	-	-	-
Difficult	O	1IBR	-	-	-	-	-	-	-
Easy	AB	1IQD	-	285	241	6.95	7.42	0.85	0.15
Difficult	O	1IRA	-	-	-	-	-	-	-
Easy	A	1JPS	-	895	562	3.8	7.89	0.57	0.57
Easy	AB	1K4C	-	-	-	-	-	-	-
Medium	O	1K5D	-	-	1307	4.83	6.82	0.26	0.26
Easy	O	1K74	-	214	1	4.95	5.63	0.64	0.52
Easy	O	1KAC	-	529	14	5.6	8.93	0.67	0.64
Medium	E	1KKL	-	-	300	6.74	7.73	0.43	0.55
Easy	O	1KTZ	-	557	557	2.24	2.51	0.74	0.21
Easy	O	1KXP	-	223	167	4.73	7.14	0.4	0.36
Easy	AB	1KXQ	-	4	4	2.92	3.91	0.76	0.23
Medium	E	1M10	-	-	1798	6.28	8.32	0.29	0.32
Easy	E	1MAH	-	1	1	1.8	2.29	0.79	0.26
Easy	O	1ML0	-	10	10	1.99	2.42	0.75	0.28

Easy	A	1MLC	-	46	46	2.17	2.74	0.9	0.48
Medium	O	1N2C	-	-	6	7.59	9.90	0.22	0.51
Easy	E	1N8O	-	101	17	4.43	5.04	0.58	0.26
Easy	AB	1NCA	-	700	700	1.9	2.56	0.68	0.11
Easy	AB	1NSN	-	471	426	5.13	5.84	0.44	0.55
Easy	E	1OPH	-	-	-	-	-	-	-
Easy	E	1PPE	-	1	1	1.67	1.75	0.91	0.24
Difficult	E	1PXV	-	-	42	7.8	6.62	0.1	0.85
Easy	O	1QA9	-	-	-	-	-	-	-
Easy	AB	1QFW	-	376	331	5.87	9.87	0.57	0.48
Easy	E	1R0R	-	29	8	5.68	7.65	0.53	0.53
Difficult	O	1R8S	-	-	-	-	-	-	-
Easy	O	1RLB	-	-	547	4.72	8.26	0.57	0.34
Easy	O	1S1Q	-	303	55	5.89	7.14	0.54	0.65
Easy	O	1SBB	-	-	-	-	-	-	-
Easy	O	1T6B	-	886	646	6.32	8.61	0.45	0.43
Easy	E	1TMQ	-	22	22	3.32	2.82	0.72	0.49
Easy	E	1UDI	-	13	1	6.75	7.08	0.35	0.62
Easy	A	1VFB	-	303	25	6.84	7.93	0.33	0.78
Easy	A	1WEJ	-	36	36	2.01	2.61	0.7	0.14
Medium	O	1WQ1	-	1630	1630	3.95	4.95	0.57	0.38
Easy	O	1XD3	-	40	5	5.64	6.84	0.42	0.34
Medium	O	1XQS	-	-	-	-	-	-	-
Difficult	O	1Y64	-	-	-	-	-	-	-
Easy	E	1YVB	-	1	1	2.11	2.81	0.68	0.31
Easy	O	1ZHI	-	594	153	4.75	8.58	0.61	0.34
Easy	O	2AJF	-	621	621	3.89	3.94	0.55	0.48
Easy	E	2B42	-	1712	2	4.53	6.21	0.55	0.45
Easy	O	2BTF	-	-	457	5.72	6.24	0.29	0.57
Difficult	O	2C0L	-	-	-	-	-	-	-
Medium	O	2CFH	-	6	6	3.55	4.29	0.82	0.43
Easy	A	2FD6	-	-	560	3.93	9.01	0.51	0.47
Medium	O	2H7V	-	-	1615	6.29	8.78	0.6	0.4
Easy	O	2HLE	-	13	13	3.38	3.37	0.53	0.37
Difficult	AB	2HMI	-	-	-	-	-	-	-
Easy	O	2HQS	-	541	59	5.11	5.65	0.47	0.61
Medium	O	2HRK	-	-	308	6.77	7.36	0.6	0.4
Easy	A	2I25	-	-	36	6.52	8.53	0.46	0.31
Easy	AB	2JEL	-	164	55	7.89	8.59	0.22	0.69
Easy	E	2MTA	-	72	20	6.66	7.46	0.38	0.75
Medium	O	2NZ8	-	-	30	9.59	9.62	0.13	0.89
Difficult	O	2OT3	-	-	165	9.12	8.29	0.14	0.62
Easy	E	2PCC	-	734	299	7.4	9.64	0.33	0.75
Easy	AB	2QFW	-	189	113	4.19	9.27	0.69	0.35
Easy	E	2SIC	-	94	19	3.83	5.35	0.65	0.22
Easy	E	2SNI	-	-	-	-	-	-	-
Easy	E	2UUY	-	-	-	-	-	-	-
Easy	A	2VIS	-	465	186	2.77	5.28	0.74	0.2
Easy	E	7CEI	-	288	288	1.41	1.41	0.78	0.07

[a] Subset of 97 structures from the ZDOCK benchmark 3.0. See footnotes of Table S9 for further explanations.

Table S13: Results for unbound knowledge-driven docking with the DrugScore^{PPI}/FRODOCK approach.[a]

Difficulty	Category	PDB ID	Quality_1	Quality_2	Quality_3	i_rmsd	l_rmsd	f_{nat}	f_{not}
Easy	A	1AHW	-	-	177	8.45	9.65	0.24	0.53
Easy	O	1AK4	-	-	-	-	-	-	-
Easy	O	1AKJ	-	-	33	5.42	8.67	0.64	0.68
Easy	E	1AVX	-	3	3	1.67	3.17	0.74	0.13
Easy	E	1AY7	-	116	54	7.76	6.55	0.18	0.83
Easy	O	1B6C	-	5	5	3.62	3.58	0.74	0.48
Medium	A	1BGX	-	-	-	-	-	-	-
Easy	AB	1BJ1	-	43	43	1.83	3.29	0.79	0.07
Difficult	O	1BKD	-	-	-	-	-	-	-
Easy	O	1BUH	-	772	217	5.72	6.4	0.36	0.59
Easy	A	1BVK	-	1182	16	8.03	8.62	0.13	0.92
Easy	E	1BVN	-	21	21	2.29	2.36	0.62	0.15
Easy	E	1CGI	-	-	69	9.04	9.11	0.14	0.77
Easy	E	1D6R	-	-	81	5.34	7.53	0.6	0.42
Difficult	O	1DE4	-	-	-	-	-	-	-
Easy	E	1DFJ	-	-	3	5.7	5.52	0.29	0.77
Easy	A	1DQJ	-	-	78	5.7	7.75	0.56	0.38
Easy	A	1E6J	-	21	21	3.56	3.13	0.68	0.6
Easy	E	1EAW	-	1	1	3.76	4.33	0.75	0.3
Difficult	O	1EER	-	1999	356	8.36	9.35	0.15	0.69
Easy	E	1EWY	-	143	12	7.54	7.35	0.26	0.85
Easy	E	1EZU	-	1993	20	5.13	8.75	0.39	0.33
Easy	O	1F51	-	627	85	8.7	9.27	0.15	0.8
Easy	O	1FQJ	-	-	347	5.09	6.13	0.52	0.44
Easy	AB	1FSK	-	1	1	1.69	2.03	0.94	0.25
Easy	O	1GCQ	-	-	575	4.72	5.98	0.59	0.54
Easy	O	1GLA	-	-	1645	8.61	9.25	0.31	0.63
Easy	O	1GPW	-	44	44	3.85	4.95	0.84	0.59
Easy	O	1HE1	-	1303	5	7.67	7.84	0.19	0.78
Easy	E	1HIA	-	-	19	8.05	9.01	0.27	0.6
Medium	O	1I2M	-	-	81	5.93	6	0.23	0.39
Easy	O	1I4D	-	-	-	-	-	-	-
Easy	AB	1I9R	-	-	-	-	-	-	-
Difficult	O	1IBR	-	-	-	-	-	-	-
Easy	AB	1IQD	-	-	62	5.85	9.9	0.82	0.18
Difficult	O	1IRA	-	-	-	-	-	-	-
Easy	A	1JPS	-	1935	1161	3.47	6.54	0.59	0.37
Easy	AB	1K4C	-	-	-	-	-	-	-
Medium	O	1K5D	-	-	83	7.32	9.92	0.17	0.44
Easy	O	1K74	-	76	18	6.42	6.98	0.47	0.59
Easy	O	1KAC	-	273	97	8.16	9.33	0.2	0.85
Medium	E	1KKL	-	446	70	8.22	8.43	0.1	0.91
Easy	O	1KTZ	-	1101	580	6.66	7.46	0.58	0.49
Easy	O	1KXP	-	-	112	6.68	6.63	0.29	0.61
Easy	AB	1KXQ	-	247	4	3.45	5.66	0.7	0.3
Medium	E	1M10	-	-	-	-	-	-	-
Easy	E	1MAH	-	3	3	2.81	3.49	0.8	0.39
Easy	O	1ML0	-	4	4	2.48	2.86	0.76	0.35

Easy	A	1MLC	-	1	1	2.28	2.87	0.84	0.49
Medium	O	1N2C	-	-	174	8.71	9.98	0.21	0.54
Easy	E	1N8O	-	12	12	4.15	4.9	0.59	0.21
Easy	AB	1NCA	-	1	1	3.17	4.94	0.83	0.26
Easy	AB	1NSN	-	77	77	4.55	4.85	0.62	0.47
Easy	E	1OPH	-	-	-	-	-	-	-
Easy	E	1PPE	-	1	1	2.44	2.55	0.85	0.28
Difficult	E	1PXV	-	-	699	7.65	8.87	0.25	0.58
Easy	O	1QA9	-	-	-	-	-	-	-
Easy	AB	1QFW	-	47	47	1.84	2.03	0.9	0.19
Easy	E	1R0R	-	30	30	5.52	1.91	0.91	0.22
Difficult	O	1R8S	-	-	-	-	-	-	-
Easy	O	1RLB	-	790	52	3.54	5.17	0.69	0.35
Easy	O	1S1Q	-	83	12	5.53	5.33	0.65	0.6
Easy	O	1SBB	-	-	-	-	-	-	-
Easy	O	1T6B	-	539	69	6.02	8.19	0.48	0.39
Easy	E	1TMQ	-	216	216	4.17	4.62	0.51	0.47
Easy	E	1UDI	-	1920	9	5.73	5.84	0.36	0.64
Easy	A	1VFB	-	335	45	8.64	6.84	0.18	0.86
Easy	A	1WEJ	-	67	4	7.61	8.9	0.25	0.78
Medium	O	1WQ1	-	-	27	7.17	9.29	0.28	0.59
Easy	O	1XD3	-	151	66	8.25	9.69	0.32	0.4
Medium	O	1XQS	-	-	1261	7.43	8.81	0.22	0.53
Difficult	O	1Y64	-	-	-	-	-	-	-
Easy	E	1YVB	-	1	1	2.31	4.8	0.85	0.25
Easy	O	1ZHI	-	-	288	7.22	9.4	0.36	0.76
Easy	O	2AJF	-	521	373	6.25	6.88	0.26	0.69
Easy	E	2B42	-	12	12	2.63	3.61	0.73	0.28
Easy	O	2BTF	-	-	37	9.37	9.88	0.15	0.76
Difficult	O	2C0L	-	-	843	9.73	9.39	0.3	0.78
Medium	O	2CFH	-	59	59	3.52	2.36	0.85	0.42
Easy	A	2FD6	-	-	714	4.18	8.55	0.45	0.54
Medium	O	2H7V	-	-	726	7.66	9.32	0.33	0.67
Easy	O	2HLE	-	433	59	6.24	7.34	0.28	0.57
Difficult	AB	2HMI	-	-	-	-	-	-	-
Easy	O	2HQS	-	1427	54	5.14	7.51	0.52	0.52
Medium	O	2HRK	-	-	571	10.37	9.91	0.33	0.67
Easy	A	2I25	-	-	1095	9.17	8.07	0.28	0.46
Easy	AB	2JEL	-	-	37	9.35	8.99	0.21	0.73
Easy	E	2MTA	-	-	126	6.68	8.23	0.18	0.85
Medium	O	2NZ8	-	866	678	7.79	8.97	0.26	0.67
Difficult	O	2OT3	-	-	530	6.67	9.19	0.2	0.29
Easy	E	2PCC	-	-	96	10.02	8.42	0.12	0.94
Easy	AB	2QFW	-	1702	8	3.67	7.01	0.8	0.29
Easy	E	2SIC	-	85	4	4.49	6.61	0.66	0.45
Easy	E	2SNI	-	-	35	7.5	8.63	0.51	0.32
Easy	E	2UUY	-	-	-	-	-	-	-
Easy	A	2VIS	-	393	34	5.69	6.22	0.42	0.54
Easy	E	7CEI	-	-	177	8.45	9.65	0.24	0.53

[a] Subset of 97 structures from the ZDOCK benchmark 3.0. Results from one of the three docking runs (see main text for explanation) are shown only. See footnotes of Table S9 for further explanations.

Table S14: Calculation of A_{rel}.

PDB ID[a]	all rmsd [b]	log all rmsd [c]	MW [d]	$A_s^{observed}$ [e]	$A_s^{predicted}$ [f]	A_{rel} [g]
1AHW_l	1.25	0.10	23005	11914	9996	1.19
1AHW_r	0.75	-0.13	46947	19131	17190	1.11
1AK4_l	0.84	-0.07	15126	8585	7268	1.18
1AK4_r	0.37	-0.43	17887	7394	8256	0.90
1AKJ_l	0.62	-0.21	25652	10918	10859	1.01
1AKJ_r	1.13	0.05	43649	18692	16264	1.15
1AVX_l	0.55	-0.26	17865	8308	8248	1.01
1AVX_r	0.52	-0.28	23483	9238	10154	0.91
1AY7_l	0.68	-0.17	10175	5016	5377	0.93
1AY7_r	0.44	-0.35	10571	5619	5536	1.02
1B6C_l	0.34	-0.47	11824	5876	6027	0.97
1B6C_r	0.94	-0.03	37434	15518	14472	1.07
1BGX_l	0.98	-0.01	92091	36273	28685	1.26
1BGX_r	1.62	0.21	46611	18950	17096	1.11
1BJ1_l	0.62	-0.21	22013	11079	9667	1.15
1BJ1_r	-	-	-	-	-	-
1BKD_l	1.34	0.13	18833	8203	8586	0.96
1BKD_r	2.11	0.32	54436	24586	19236	1.28
1BUH_l	1.02	0.01	9190	5528	4977	1.11
1BUH_r	0.99	-0.01	33490	14320	13298	1.08
1BVK_l	0.70	-0.15	14307	6307	6967	0.91
1BVK_r	0.73	-0.13	24762	10110	10571	0.96
1BVN_l	0.55	-0.26	7959	4337	4462	0.97
1BVN_r	0.47	-0.33	55212	17440	19444	0.90
1CGI_l	1.03	0.01	6339	4056	3753	1.08
1CGI_r	0.71	-0.15	25672	10724	10865	0.99
1D6R_l	1.05	0.02	6211	4287	3695	1.16
1D6R_r	0.47	-0.33	23315	9545	10098	0.95
1DE4_l	0.82	-0.09	43032	18569	16089	1.15
1DE4_r	1.21	0.08	143072	46707	40093	1.16
1DFJ_l	0.65	-0.19	13699	6878	6741	1.02
1DFJ_r	1.38	0.14	49009	18385	17760	1.04
1DQJ_l	0.69	-0.16	14307	6306	6967	0.91
1DQJ_r	0.66	-0.18	46107	19039	16955	1.12
1E6J_l	1.48	0.17	8204	5021	4566	1.10
1E6J_r	0.91	-0.04	46576	19161	17086	1.12
1EAW_l	0.60	-0.22	6524	3924	3836	1.02
1EAW_r	0.47	-0.33	26439	10633	11111	0.96
1EER_l	2.38	0.38	45757	22573	16857	1.34
1EER_r	2.78	0.44	18447	10312	8452	1.22
1EWY_l	0.82	-0.08	10685	5085	5581	0.91
1EWY_r	3.32	0.52	62754	26062	21431	1.22
1EZU_l	2.15	0.33	32169	16220	12897	1.26
1EZU_r	0.41	-0.38	23738	9389	10237	0.92
1F51_l	0.80	-0.10	13290	6016	6588	0.91
1F51_r	0.86	-0.06	35319	16402	13846	1.18
1FQJ_l	0.91	-0.04	15679	8722	7470	1.17
1FQJ_r	0.51	-0.30	36206	15058	14110	1.07

1FSK_l	0.67	-0.17	17405	8654	8086	1.07
1FSK_r	-	-	-	-	-	-
1GCQ_l	0.80	-0.10	24495	12638	10484	1.21
1GCQ_r	0.55	-0.26	7740	4187	4368	0.96
1GLA_l	0.50	-0.30	15713	6479	7482	0.87
1GLA_r	0.54	-0.27	55287	18157	19464	0.93
1GPW_l	0.55	-0.26	22339	9022	9775	0.92
1GPW_r	0.89	-0.05	27686	11152	11507	0.97
1HE1_l	0.59	-0.23	19609	9194	8853	1.04
1HE1_r	0.96	-0.02	13889	7152	6812	1.05
1HIA_l	2.34	0.37	5177	3990	3218	1.24
1HIA_r	0.53	-0.27	25586	10370	10837	0.96
1I2M_l	0.42	-0.37	42788	15891	16020	0.99
1I2M_r	1.86	0.27	22833	10397	9939	1.05
1I4D_l	0.68	-0.17	19609	9194	8853	1.04
1I4D_r	0.85	-0.07	45682	21804	16837	1.30
1I9R_l	-	-	-	-	-	-
1I9R_r	0.77	-0.12	47415	16155	17320	0.93
1IBR_l	2.70	0.43	49107	20679	17787	1.16
1IBR_r	1.41	0.15	20925	9732	9301	1.05
1IQD_l	0.35	-0.46	18049	8316	8313	1.00
1IQD_r	-	-	-	-	-	-
1IRA_l	0.53	-0.28	16311	7909	7697	1.03
1IRA_r	6.01	0.78	35612	16766	13933	1.20
1JPS_l	0.87	-0.06	20831	10937	9270	1.18
1JPS_r	0.68	-0.17	46003	18834	16926	1.11
1K4C_l	0.60	-0.22	39854	17153	15178	1.13
1K4C_r	-	-	-	-	-	-
1K5D_l	0.56	-0.25	38223	14480	14703	0.98
1K5D_r	1.44	0.16	38841	16318	14884	1.10
1K74_l	0.77	-0.11	32422	14390	12974	1.11
1K74_r	0.57	-0.24	26329	11490	11076	1.04
1KAC_l	0.71	-0.15	13397	6431	6628	0.97
1KAC_r	0.44	-0.36	19936	8607	8965	0.96
1KKL_l	0.55	-0.26	8938	4579	4873	0.94
1KKL_r	1.08	0.03	48858	22865	17719	1.29
1KTZ_l	0.38	-0.42	11963	6175	6081	1.02
1KTZ_r	1.16	0.07	12724	6975	6373	1.09
1KXP_l	1.85	0.27	50566	23502	18188	1.29
1KXP_r	0.81	-0.09	39636	16562	15114	1.10
1KXQ_l	-	-	-	-	-	-
1KXQ_r	0.43	-0.37	55383	17407	19490	0.89
1M10_l	0.61	-0.21	29549	12817	12091	1.06
1M10_r	0.63	-0.20	23676	9910	10217	0.97
1MAH_l	0.88	-0.06	6764	4106	3943	1.04
1MAH_r	0.62	-0.21	58791	19586	20395	0.96
1ML0_l	1.09	0.04	8151	5508	4543	1.21
1ML0_r	1.52	0.18	81489	31623	26139	1.21
1MLC_l	0.67	-0.18	14307	6306	6967	0.91
1MLC_r	0.84	-0.07	46789	18995	17146	1.11

1N2C_l	4.10	0.61	58716	20887	20375	1.03
1N2C_r	0.45	-0.35	224238	57197	56414	1.01
1N8O_l	0.69	-0.16	15828	8802	7523	1.17
1N8O_r	0.39	-0.41	24830	9863	10593	0.93
1NCA_l	0.29	-0.54	43681	14481	16273	0.89
1NCA_r	-	-	-	-	-	-
1NSN_l	0.94	-0.03	15505	7708	7406	1.04
1NSN_r	-	-	-	-	-	-
1OPH_l	0.30	-0.52	23315	9117	10098	0.90
1OPH_r	0.82	-0.09	41913	17026	15770	1.08
1PPE_l	0.59	-0.23	3257	2630	2262	1.16
1PPE_r	0.40	-0.40	23315	9124	10098	0.90
1PXV_l	0.72	-0.15	13063	7133	6502	1.10
1PXV_r	0.81	-0.09	19991	9086	8984	1.01
1QA9_l	0.80	-0.10	19914	9992	8958	1.12
1QA9_r	0.79	-0.10	20604	10736	9193	1.17
1QFW_l	-	-	-	-	-	-
1QFW_r	1.26	0.10	21252	11053	9412	1.17
1R0R_l	0.62	-0.21	5580	3553	3407	1.04
1R0R_r	0.31	-0.51	27278	9826	11378	0.86
1R8S_l	1.12	0.05	22536	11048	9841	1.12
1R8S_r	0.86	-0.07	20484	9186	9152	1.00
1RLB_l	0.62	-0.21	20055	9681	9006	1.07
1RLB_r	0.74	-0.13	49328	19388	17848	1.09
1S1Q_l	0.74	-0.13	7936	4280	4452	0.96
1S1Q_r	0.66	-0.18	16418	8342	7736	1.08
1SBB_l	0.50	-0.30	28033	11668	11616	1.00
1SBB_r	0.82	-0.08	25850	12380	10922	1.13
1T6B_l	0.50	-0.30	19821	8496	8926	0.95
1T6B_r	1.19	0.08	74897	31161	24515	1.27
1TMQ_l	0.76	-0.12	12686	5994	6359	0.94
1TMQ_r	0.37	-0.44	51095	15955	18332	0.87
1UDI_l	0.88	-0.05	9334	5361	5036	1.06
1UDI_r	0.47	-0.33	25876	10713	10931	0.98
1VFB_l	0.60	-0.22	14307	6627	6967	0.95
1VFB_r	0.43	-0.37	24700	10035	10551	0.95
1WEJ_l	0.44	-0.36	11714	6288	5985	1.05
1WEJ_r	0.75	-0.13	47139	18785	17243	1.09
1WQ1_l	0.78	-0.11	19490	8443	8813	0.96
1WQ1_r	0.85	-0.07	36965	16587	14334	1.16
1XD3_l	0.59	-0.23	7936	4280	4452	0.96
1XD3_r	0.56	-0.25	23585	9734	10187	0.96
1XQS_l	0.65	-0.19	21361	10875	9449	1.15
1XQS_r	2.29	0.36	28343	13037	11714	1.11
1Y64_l	0.98	-0.01	40038	16078	15231	1.06
1Y64_r	10.45	1.02	47317	25630	17293	1.48
1YVB_l	1.33	0.12	12192	6993	6170	1.13
1YVB_r	0.59	-0.23	27152	11055	11338	0.98
1ZHI_l	0.88	-0.06	13302	7494	6592	1.14
1ZHI_r	0.72	-0.14	23502	12050	10160	1.19

2AJF_l	0.95	-0.02	20821	10317	9266	1.11
2AJF_r	0.47	-0.33	69075	25433	23053	1.10
2B42_l	0.54	-0.27	20386	7956	9119	0.87
2B42_r	0.42	-0.37	37016	15229	14349	1.06
2BTF_l	0.56	-0.25	14930	6952	7197	0.97
2BTF_r	1.11	0.05	39636	16563	15114	1.10
2C0L_l	1.23	0.09	13183	6845	6547	1.05
2C0L_r	0.80	-0.10	33183	13554	13205	1.03
2CFH_l	1.43	0.15	16030	8103	7596	1.07
2CFH_r	0.64	-0.19	17536	8613	8133	1.06
2FD6_l	2.73	0.44	29526	14313	12084	1.18
2FD6_r	1.23	0.09	46278	18855	17003	1.11
2H7V_l	0.67	-0.17	19609	9194	8853	1.04
2H7V_r	1.64	0.22	30882	15555	12503	1.24
2HLE_l	0.80	-0.10	15938	8214	7563	1.09
2HLE_r	0.72	-0.15	20449	9809	9140	1.07
2HMI_l	-	-	-	-	-	-
2HMI_r	3.48	0.54	113417	47483	33605	1.41
2HQS_l	0.52	-0.28	12189	5939	6168	0.96
2HQS_r	1.85	0.27	42119	16605	15829	1.05
2HRK_l	0.69	-0.16	11635	6332	5954	1.06
2HRK_r	0.63	-0.20	19560	9784	8837	1.11
2I25_l	0.48	-0.32	14322	6317	6973	0.91
2I25_r	0.49	-0.31	12439	6545	6264	1.04
2JEL_l	0.64	-0.19	9119	4832	4948	0.98
2JEL_r	-	-	-	-	-	-
2MTA_l	0.38	-0.42	11491	5281	5898	0.90
2MTA_r	0.30	-0.52	53042	18653	18861	0.99
2NZ8_l	0.99	0.00	19609	9194	8853	1.04
2NZ8_r	1.69	0.23	35622	16417	13936	1.18
2OT3_l	1.26	0.10	17763	8064	8212	0.98
2OT3_r	0.91	-0.04	28536	14533	11775	1.23
2PCC_l	0.48	-0.32	12063	6395	6120	1.04
2PCC_r	0.34	-0.47	33011	12421	13153	0.94
2SIC_l	0.60	-0.22	11044	5938	5723	1.04
2SIC_r	0.24	-0.61	27538	9891	11460	0.86
2SNI_l	0.49	-0.31	7400	4510	4221	1.07
2SNI_r	0.27	-0.58	27453	9801	11433	0.86
2UUY_l	1.39	0.14	6030	4230	3613	1.17
2UUY_r	0.27	-0.56	23315	9027	10098	0.89
2VIS_l	0.50	-0.30	28952	12724	11905	1.07
2VIS_r	5.08	0.71	46343	19411	17021	1.14
7CEI_l	0.66	-0.18	15032	7741	7234	1.07
7CEI_r	0.62	-0.21	9887	5705	5261	1.08

[a] Subset of 88 structures from the ZDOCK benchmark 3.0. "_r" refers to the receptor protein, "_l" refers to the ligand protein. "-" indicates that no unbound structure was given in the dataset.

[b] All-atom rmsd. Calculated with PyMol [3].

[c] Logarithm of the all-atom rmsd.

[d] Molecular weight in kD. Calculated with Maestro [4].

[e] Observed solvent accessible surface area, for details see ref. [5]. Calculated with Maestro [4].

[f] Predicted solvent accessible surface area calculated as given in eq. 1 from ref. [5].

[g] Relative solvent accessible surface area calculated as given in eq. 2 from ref. [5].

Supplemental Figures

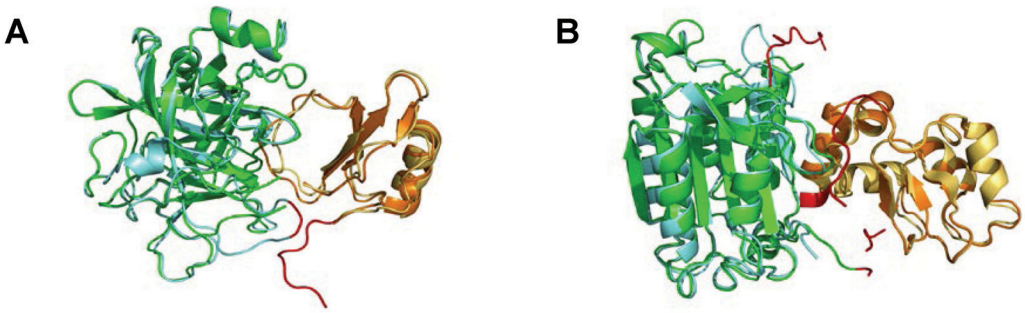


Figure S1: Crystal structures of protein-protein complexes where interfaces are incomplete, i.e., residues are missing: (A) Bovine alpha chymotrypsin in complex with eglin c, (B) Caspase-9 in complex with XIAP-BIR3. Depicted are bound and unbound complexes including the bound receptor (green), unbound receptor (cyan), bound ligand (orange), unbound ligand (yellow), and missing (incomplete) interface parts (red), respectively.

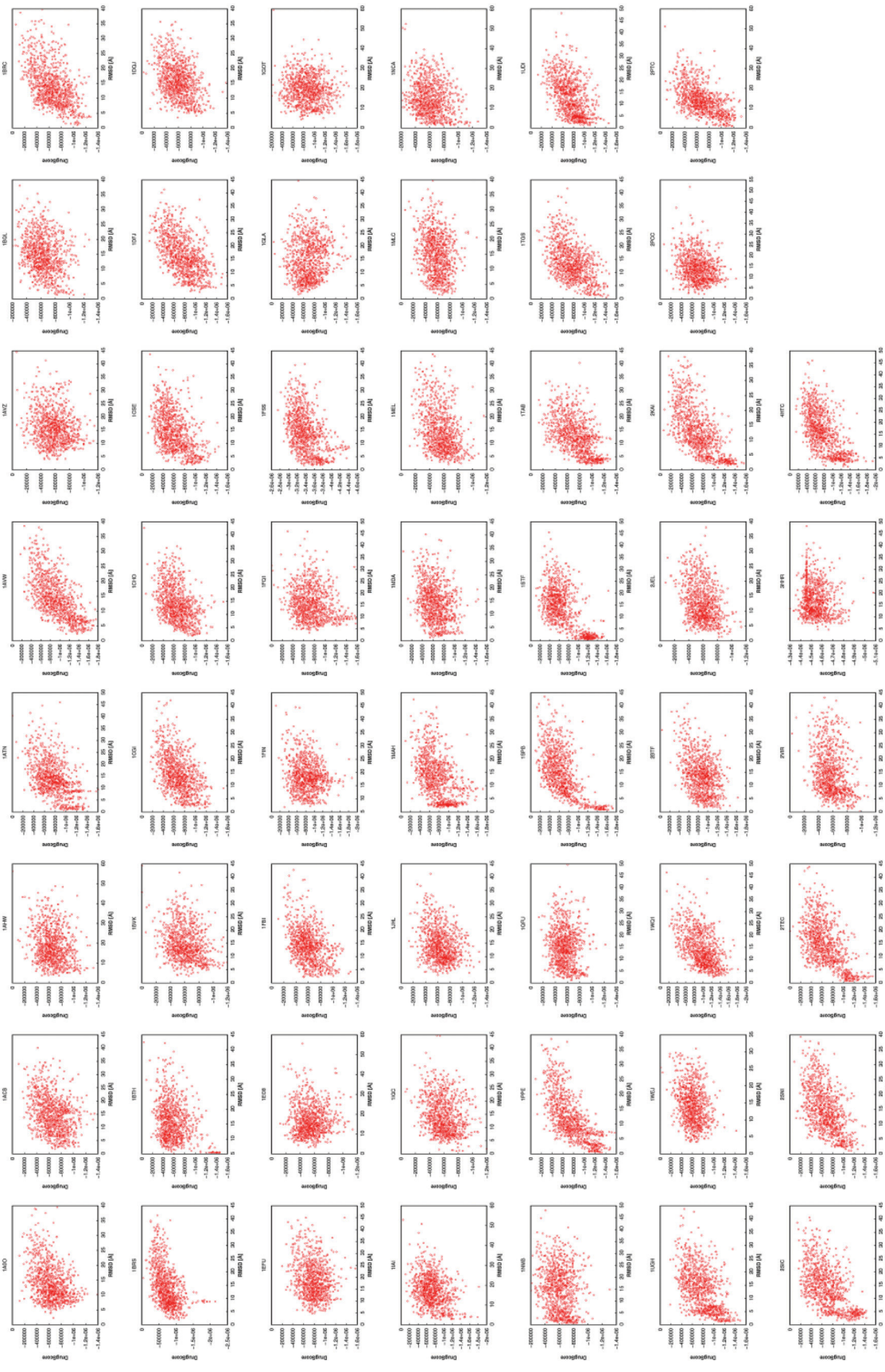


Figure S2: Energy landscapes depicting DrugsScore^{PPI} scores of the decoys of the "unbound perturbation" dataset versus the root mean square deviation of all atoms with respect to the native structure.

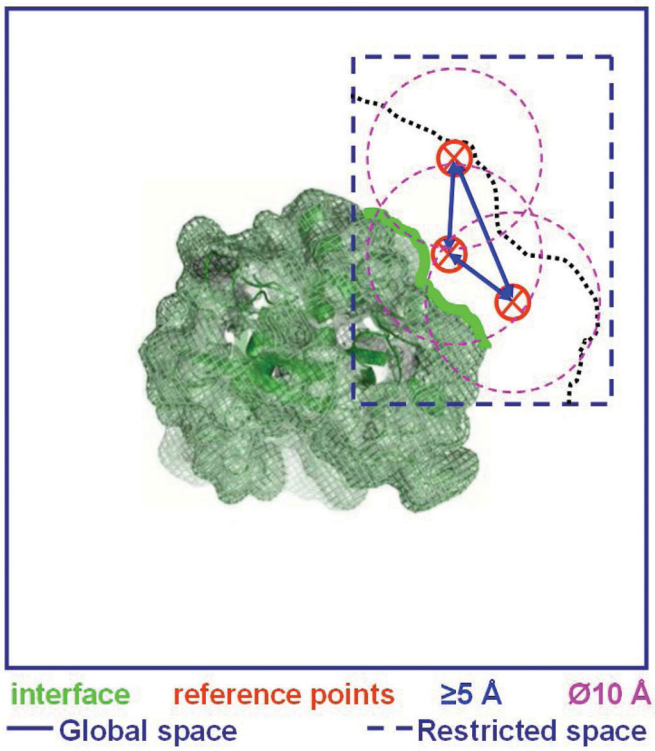


Figure S3: Restriction of the ligand search space for knowledge-driven docking (dark blue dashed line) compared to the global search space (dark blue straight line). For this, a reference point (red) within 5 Å distance (black dotted line) of the receptor interface (green line) is chosen. The search space was then restricted to 10 Å around this reference point (magenta). To minimize the bias by the selected reference point on the docking results, each docking run was repeated three times using a randomly selected reference point where each of the points must be at least 5 Å away (blue arrows) from the other two points.

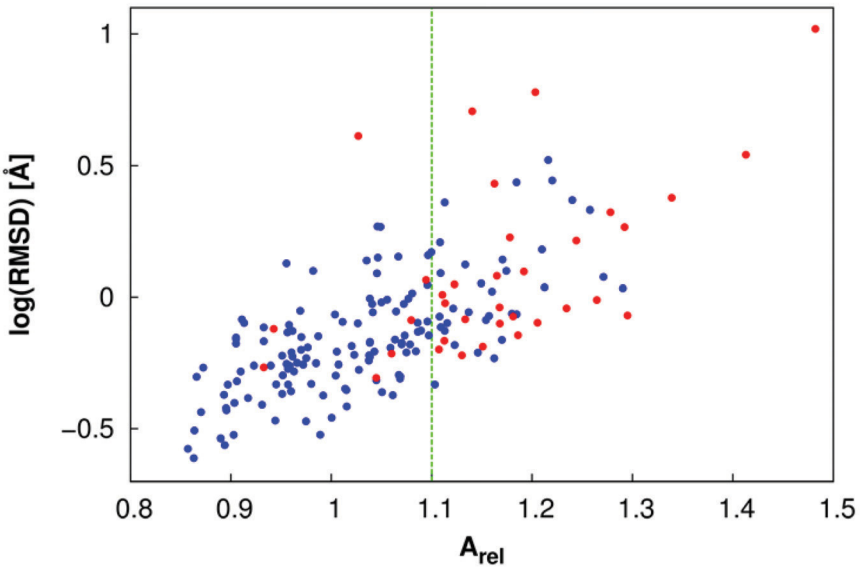


Figure S4: Logarithm of the all-atom rmsd versus calculated A_{rel} (see eq. 2 in ref. [5]) for 88 binding partners in the cleaned version of the ZDOCK benchmark 3.0 where both of the proteins are in the unbound state. The log(rmsd) ~ A_{rel} relation was chosen according to eq. 5 in ref. [5]. Of those unbound knowledge-driven dockings that failed to get a near-native solution with i_rmsd < 10 Å in the top 100, the binding partner with the larger A_{rel} is depicted in red. In 80% of these dockings, at least one of the binding partners had $A_{rel} > 1.1$ (green dashed line).

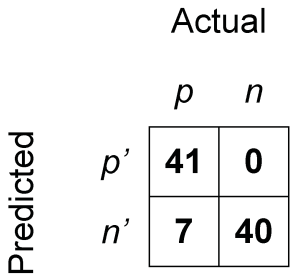
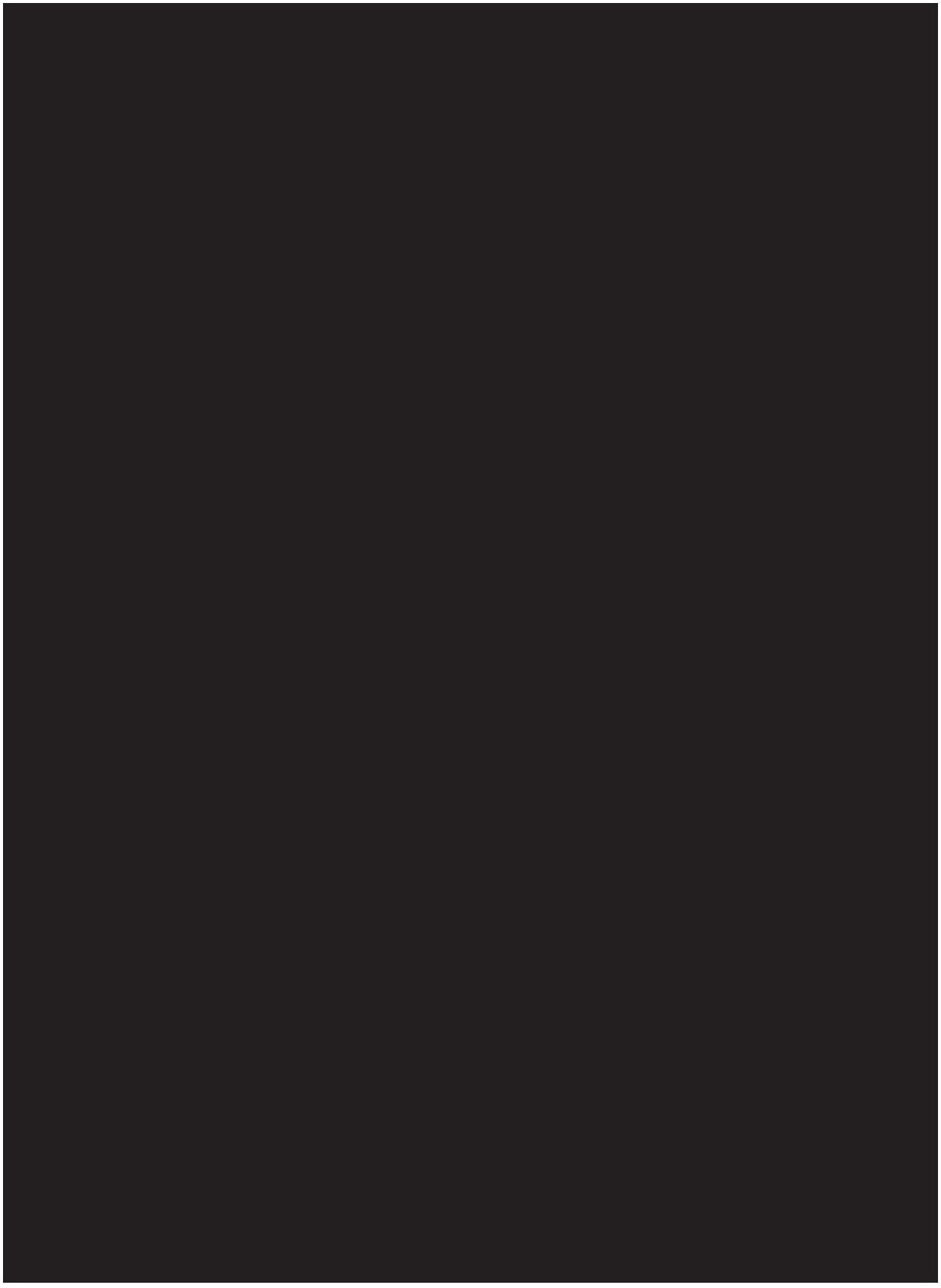


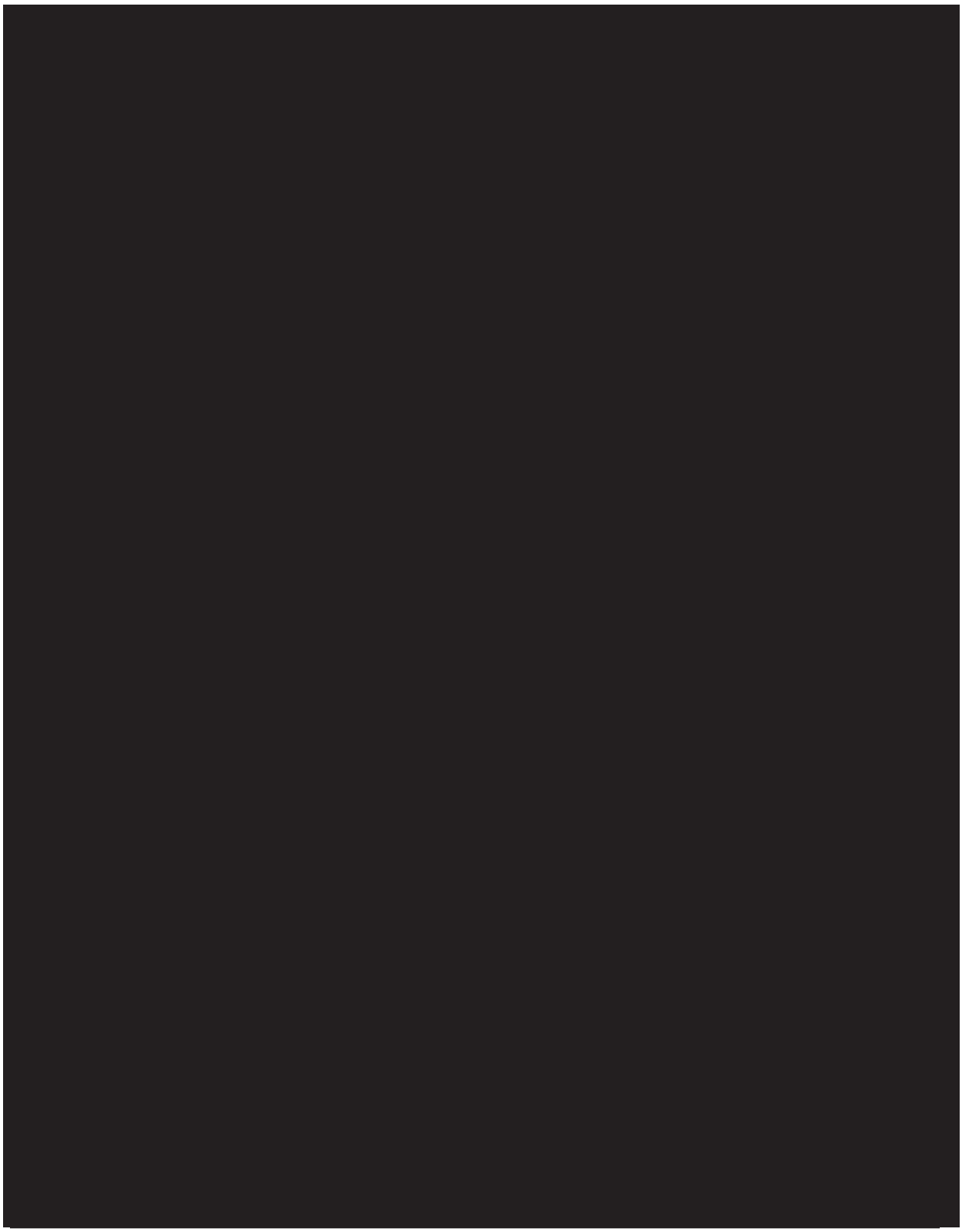
Figure S5: Confusion matrix for two possible outcomes *p* (positive) and *n* (negative) for classifying 88 unbound complexes with respect to whether they can be successfully docked with at least acceptable accuracy in the top 100 predictions using as a criterion that both binding partners have $A_{rel} < 1.1$. No. of true positives: 41, no. of false positives: 0, no. of false negatives: 7, no. of true negatives: 40.

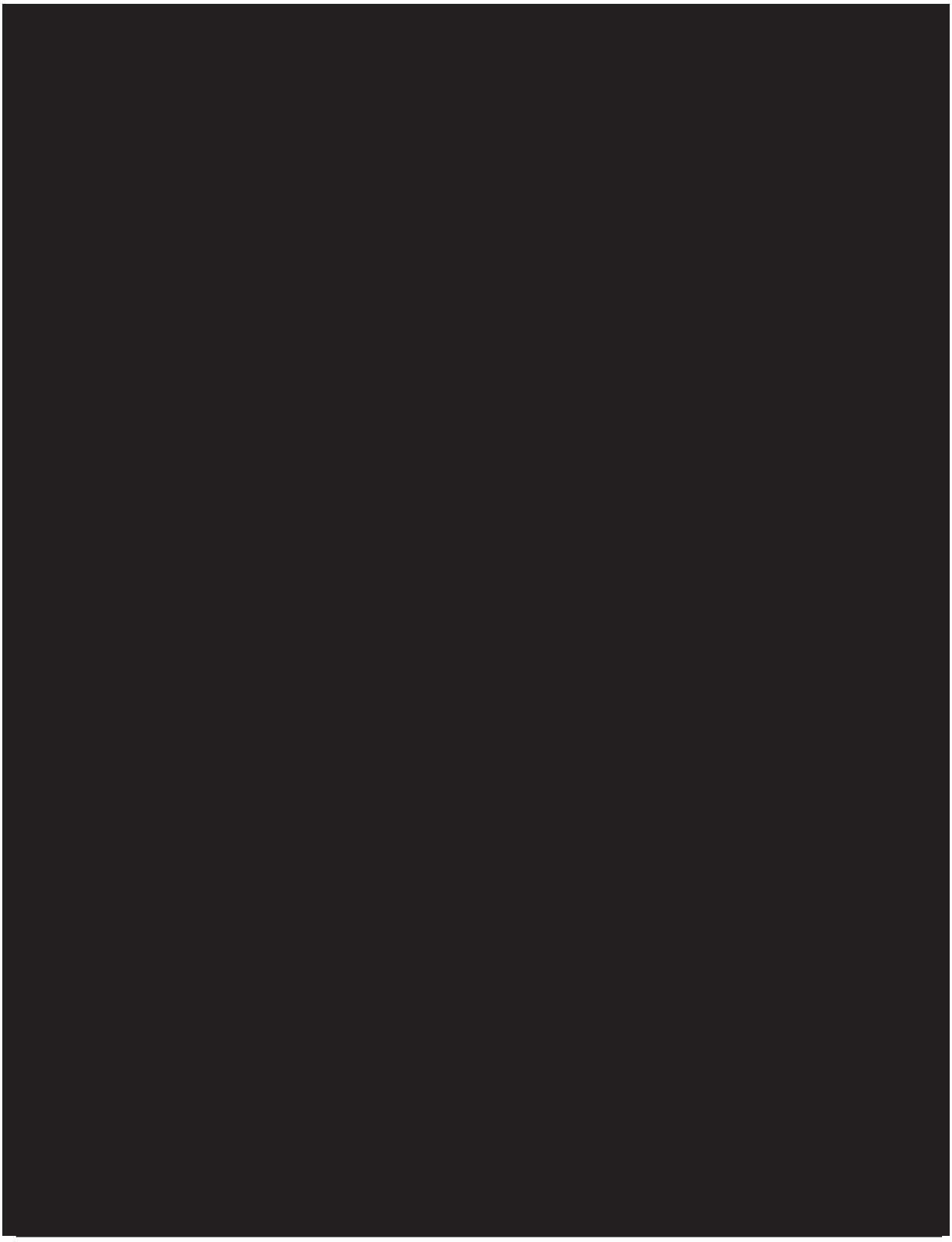
Supplemental References

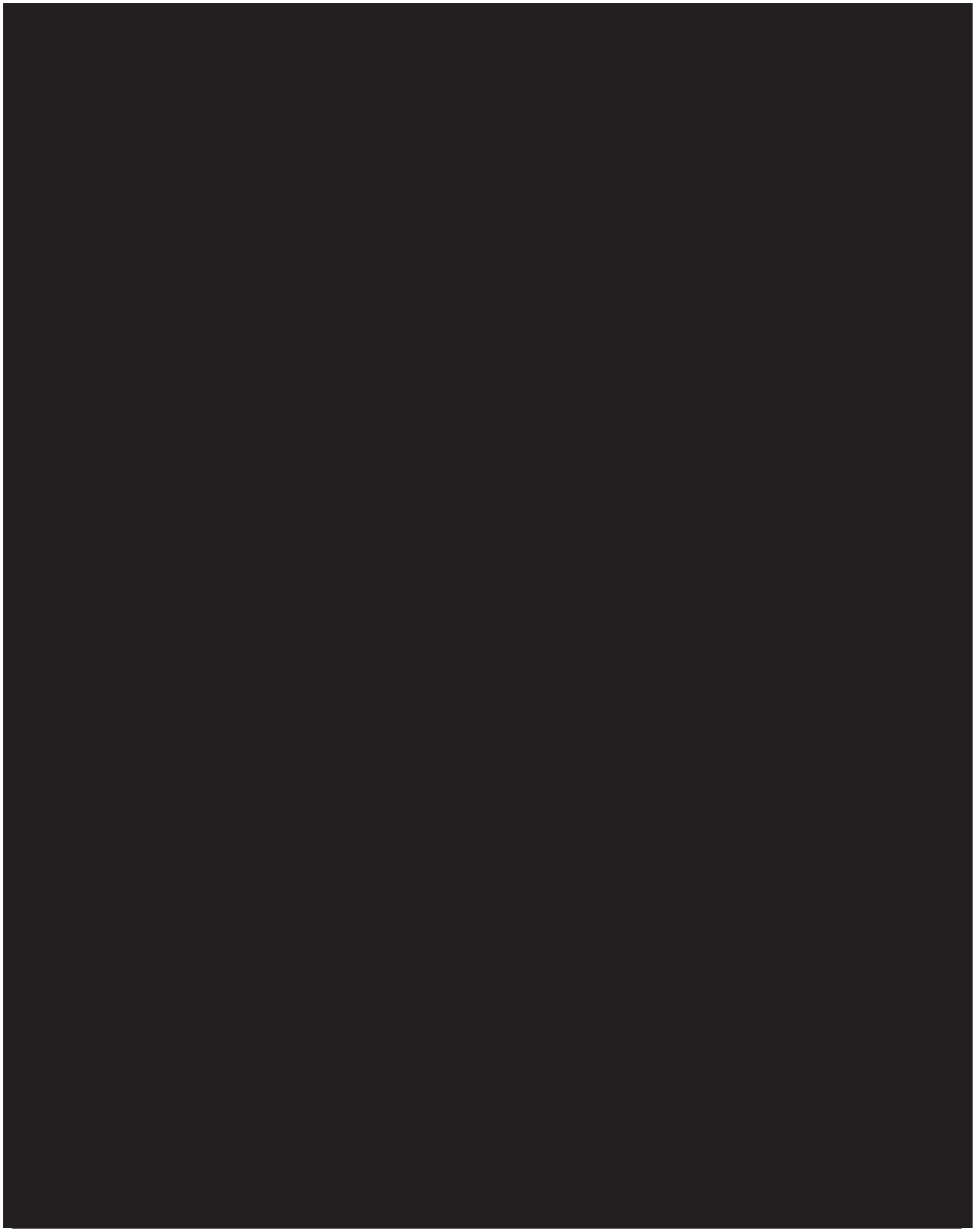
1. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z (2008) Protein-protein docking benchmark version 3.0. Proteins 73: 705-709.
2. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, et al. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 331: 281-299.
3. The PyMOL Molecular Graphics System, version 1.5.0.4, Schrödinger, LLC.
4. Maestro, version 9.1, Schrödinger, LLC.
5. Marsh JA, Teichmann SA (2011) Relative solvent accessible surface area predicts protein conformational changes upon binding. Structure 19: 859-867.

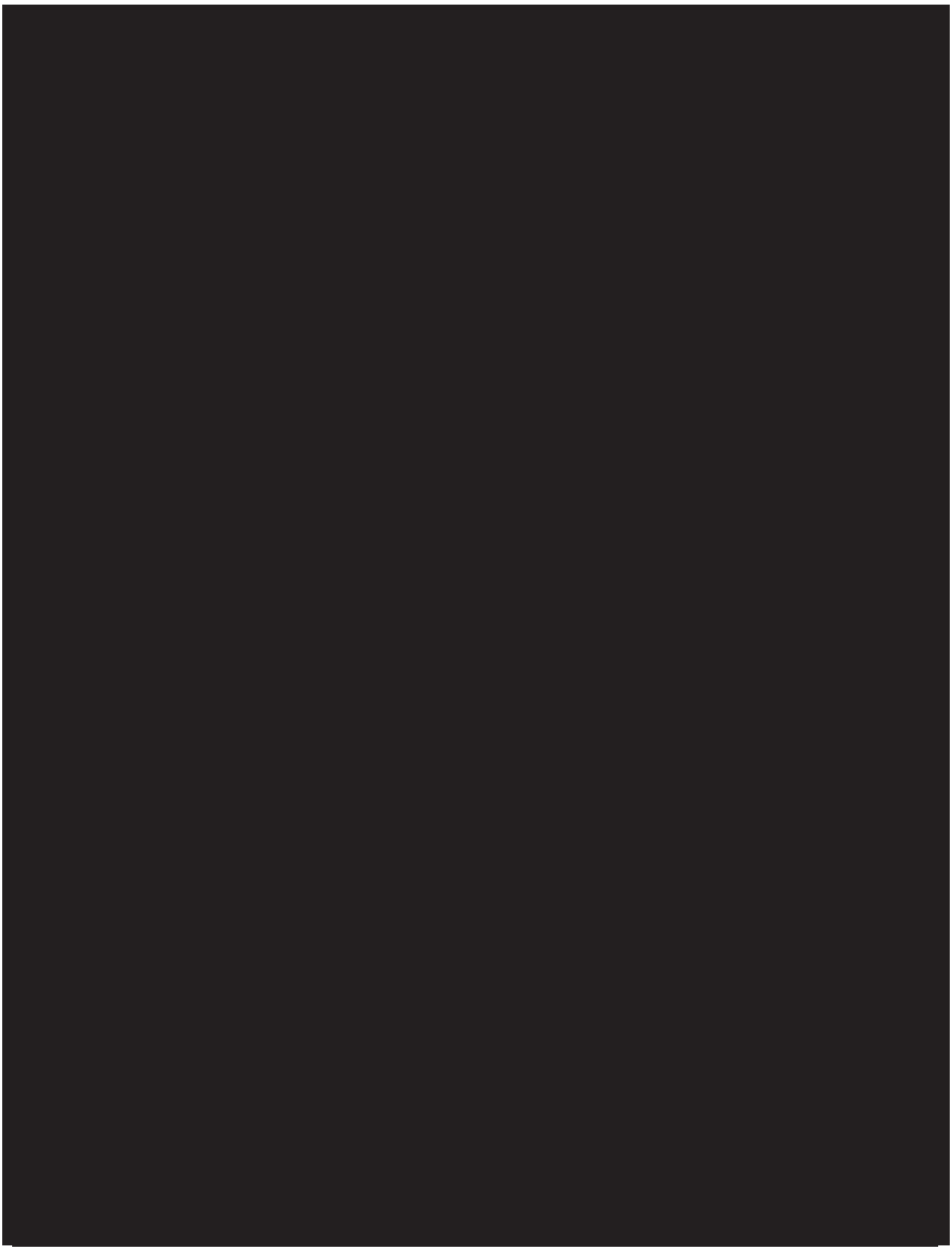
10.3 Paper III

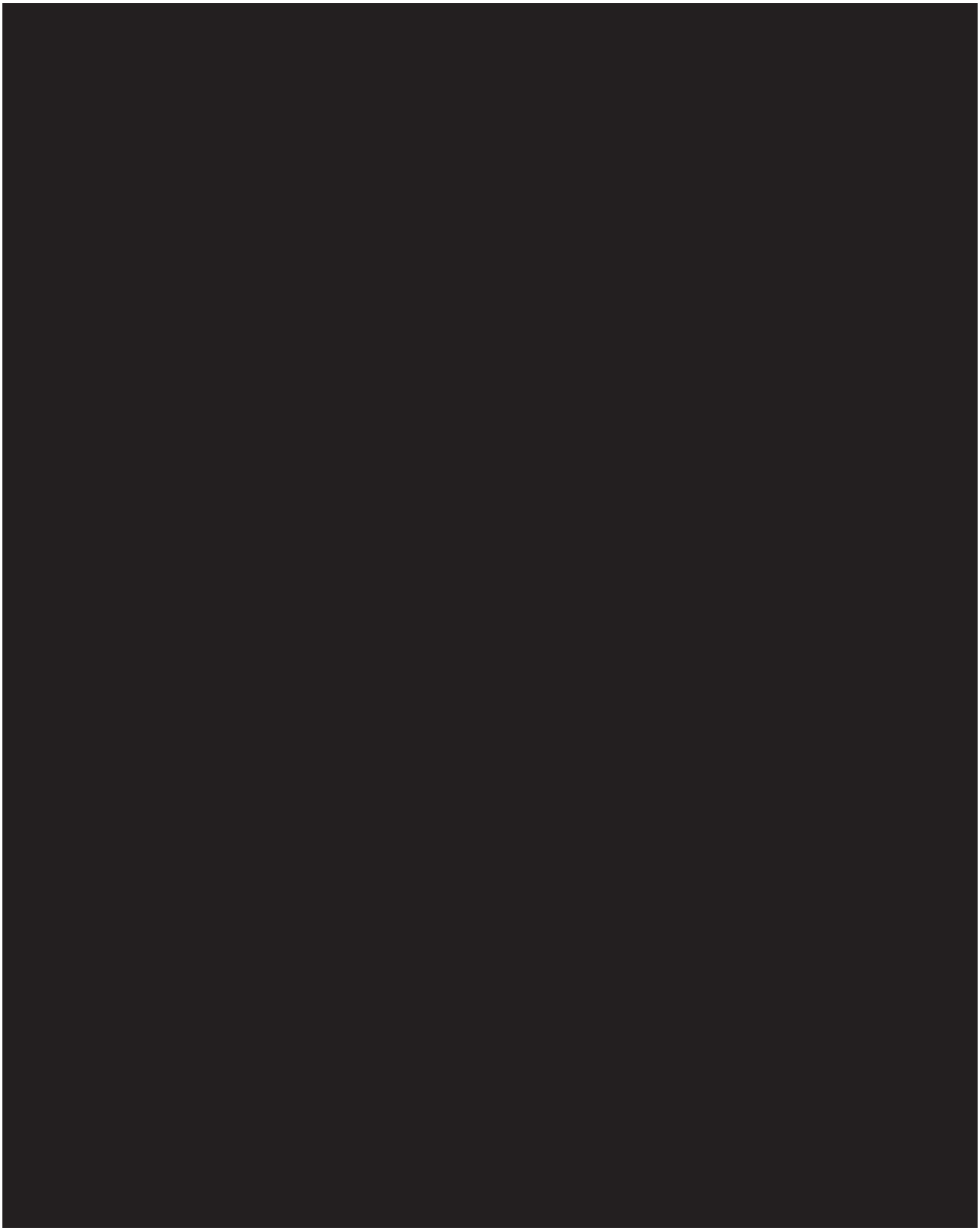


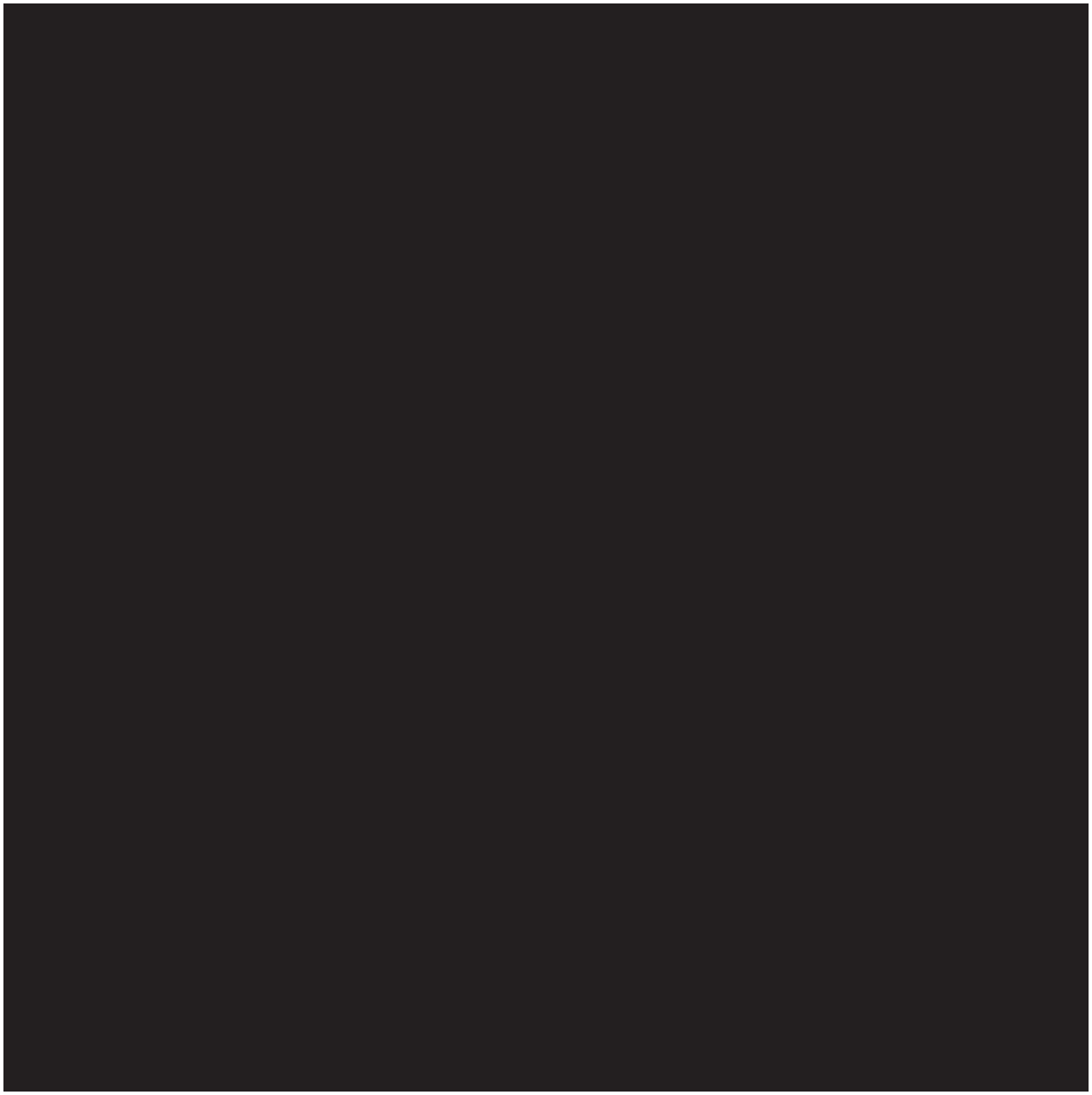












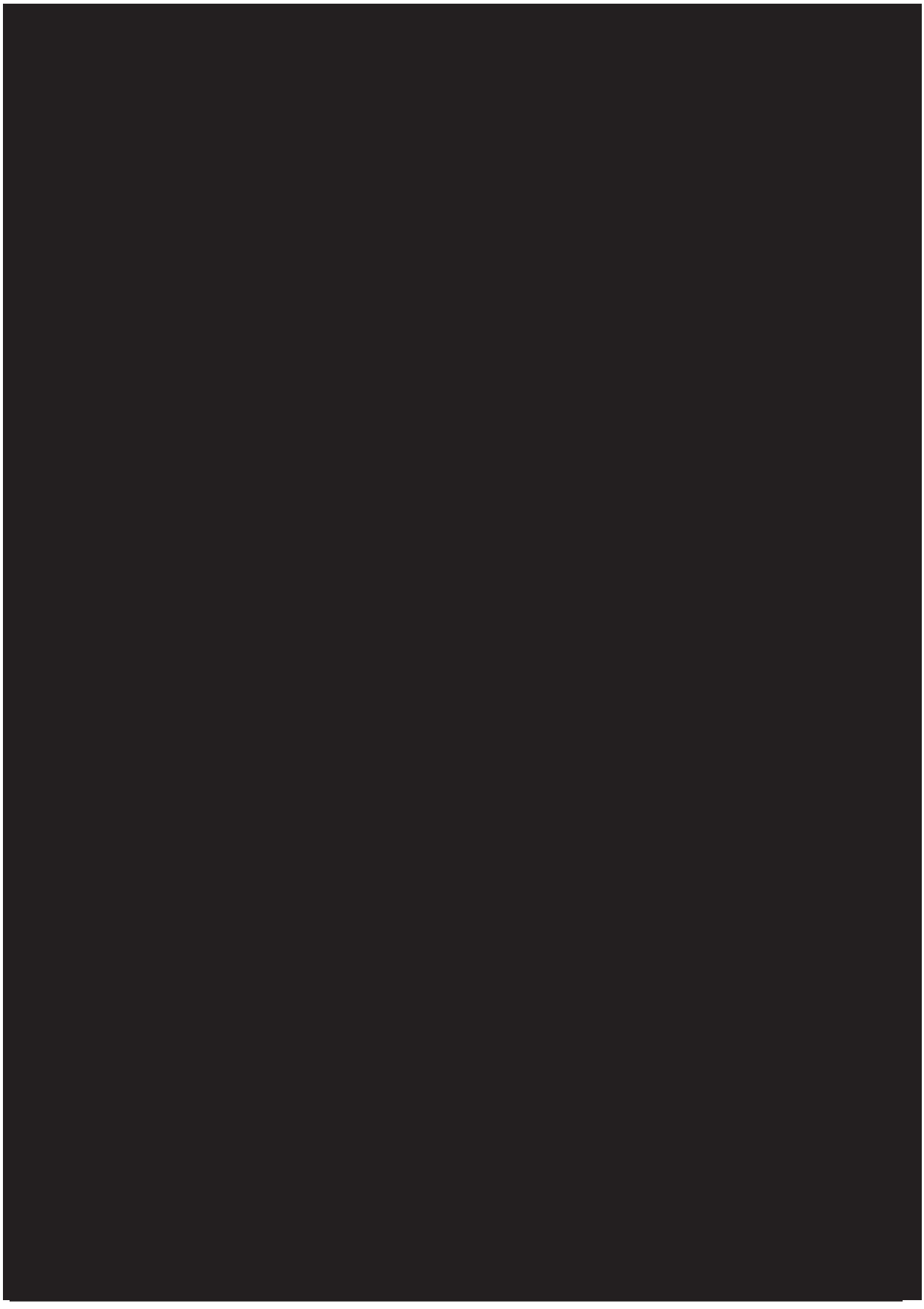
Reprinted (adapted) with permission from (Krüger, D. M., Ahmed, A. Gohlke, H., NMSIM Web Server: Normal mode-based geometric simulations for exploring biologically relevant conformational transitions in proteins. Nucleic Acids Res. 2012, 40, W310-W316). *DOI: 10.1093/nar/gks478*. URL: http://www.oxfordjournals.org/. © The Author(s) 2012. Published by Oxford University Press.

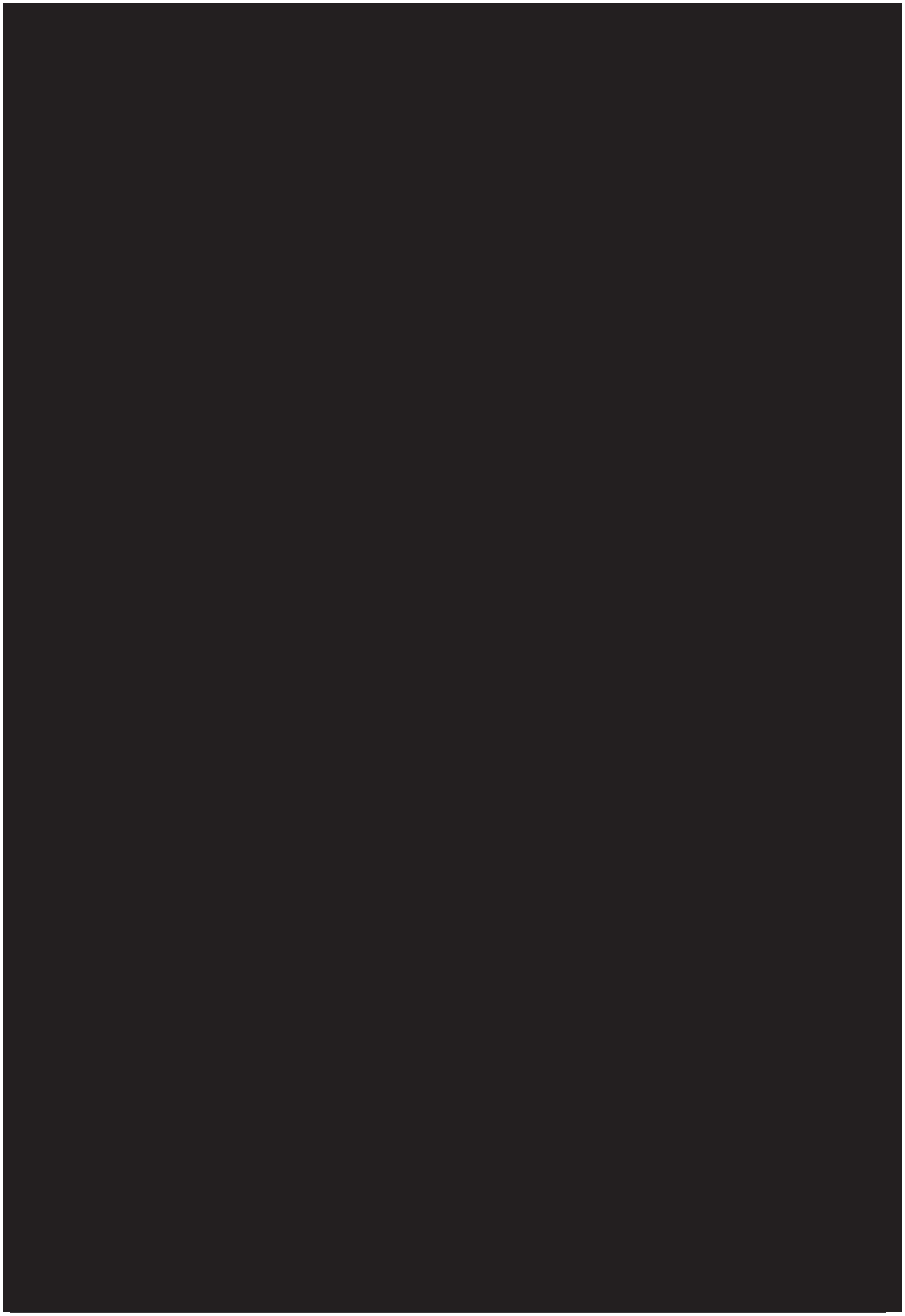
Nucleic Acids Res., 8.026*, 65%, 1. Author

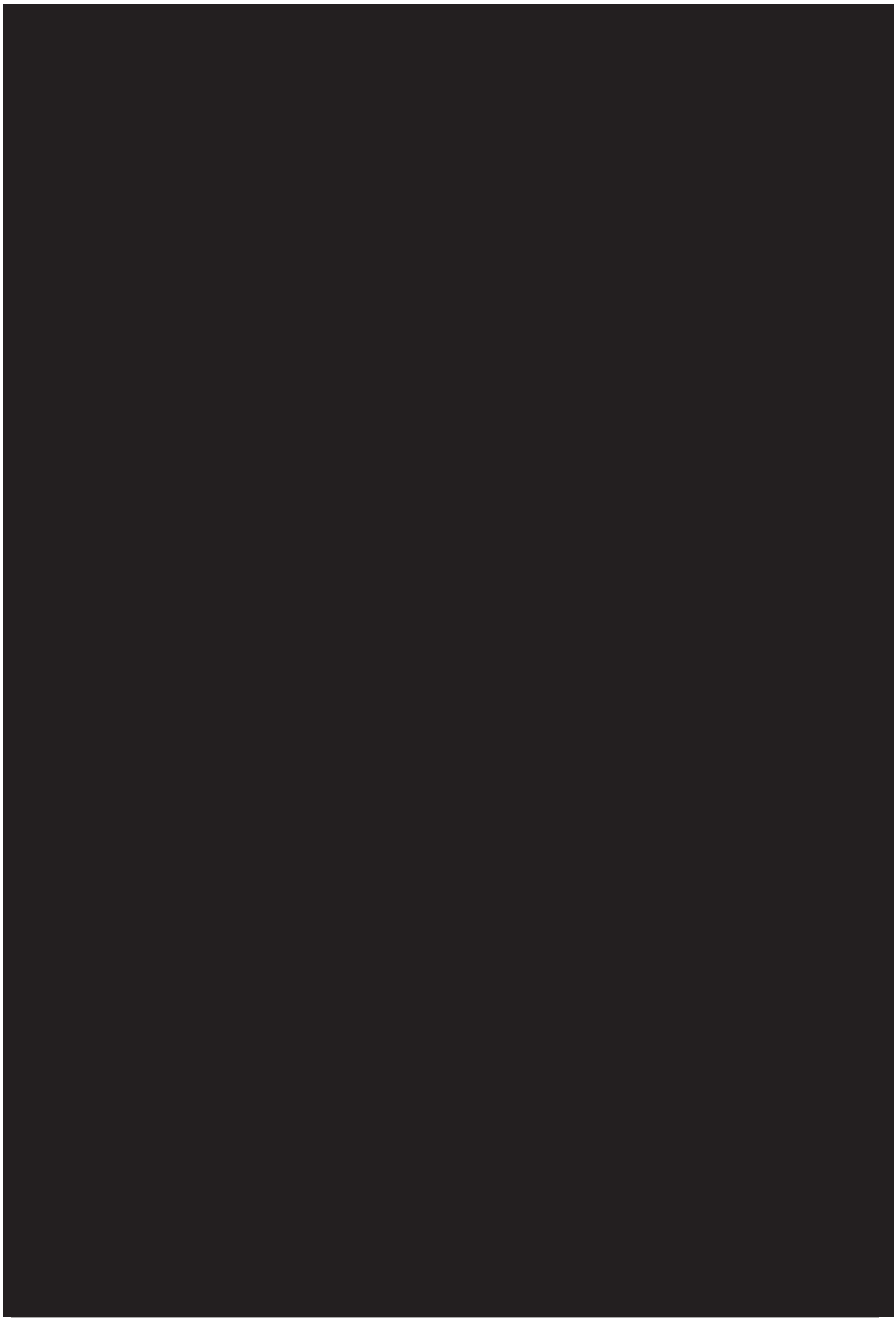
(Name of the journal, Impact factor, Contribution to this work, Authorship)

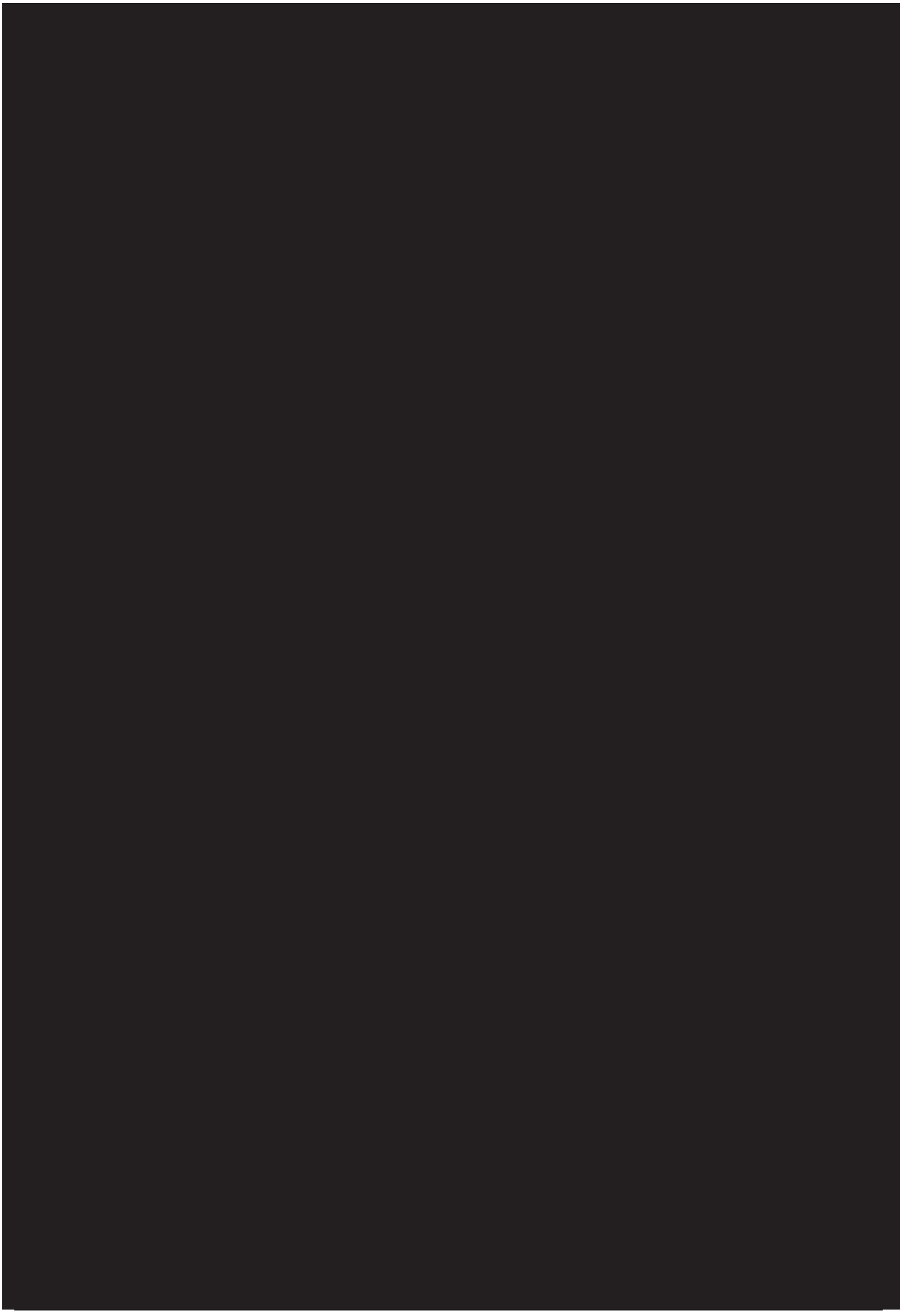
*last reported at the date of submission of this work

10.4 Paper IV









Reprinted (adapted) with permission from (Krüger, D. M., Jessen, G., Gohlke, H., How good are stat-of-the art docking tools in predicting ligand binding modes in protein-protein interfaces? J. Chem. Inf. Model. 2012, 52, 2807-2811). DOI: 10.1021/ci3003599. Copyright (2012) American Chemical Society.

J. Chem. Inf. Model., 4.675*, 70%, 1. Author

(Name of the journal, Impact factor, Contribution to this work, Authorship)

*last reported at the date of submission of this work

Supporting Information

How good are state-of-the-art docking tools in predicting ligand binding modes in protein-protein interfaces?

Dennis M. Krüger[a], Gisela Jessen[a], Holger Gohlke[a]*

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

*Universitätsstr.1, 40225 Düsseldorf (Germany), Fax: (+49) 211-8113847, E-mail: gohlke@uni-duesseldorf.de

Preparation of the PPIM dataset

In order to investigate the ability of the docking tools to predict near-native ligand geometries in protein-protein interfaces we manually selected a dataset of 22 different target proteins consisting of 63 crystal- and 17 NMR-structures, a total of 80, from the PDB database (see Table S4). The dataset contains 26 (modified) peptide ligands. All but two crystal structures have a resolution ≤ 3 Å. PPIM complex structures were selected from the literature, and it was required that either the target protein interface has been described or that the corresponding protein-protein complex has been resolved. These references are given in Table S1.

Protein structure and ligand preparation for docking

The coordinates from the protein complexes were retrieved from the Protein Data Bank (PDB).[1] For each complex structure, the ligand was removed, as were all water molecules. Ions were considered as heteroatoms and included into the docking. Ligand coordinates were extracted from the Relibase database together with the assignment of Sybyl atom types. When using the AutoDock forcefield for docking, for each protein, hydrogens were added using AutoDock-Tools (ADT) 1.5.4 except for nonpolar hydrogen atoms, and charges were added as suggested by ADT.[2] When using the DrugScore potentials for docking, hydrogens and charges are not needed and, therefore, not considered. The ligands were further prepared by ADT by assigning aromatic carbon atoms and rotatable bonds.

For Glide, all protein structures were prepared with the PrepWizard in Maestro version 9.1.[3] using two different settings: (1) Default settings using full setup including H-bond assignment and minimization; (2) Proteins were prepared with the PrepWizard by just adding hydrogens. For investigating the structural differences that arise from the minimization in the case of setup (1), we calculated the interface rmsd (defined by all heavy atoms within 5 Å distance around the ligand) with respect to the crystal structure. The results show an interface rmsd < 0.4 Å for 84% of the systems. This small difference in receptor coordinates is within the experimental uncertainty of the X-ray experiment. For the remaining 16% of the systems, an interface rmsd between 0.4 and 0.7 Å is found. The latter fraction might explain why a small (~ 4%) increase in the docking success rate is observed (see Figure S4) if hydrogens are added only during the receptor setup. All ligands were prepared with LigPrep version 25223 using default settings.[4] That way, the preparation of protein and ligand was done independently of each other, i.e., neglecting any knowledge about the correct binding mode as in a “real-life” case.

Alternatively, protein and ligand structures were prepared with the PrepWizard while still being in the complex, as described in ref. [5]. The latter procedure does require knowledge of the complex crystal structure and, hence, precludes the use of the docking tool in a predictive manner.

Docking engine, scoring function, and docking experiments

For the evaluation of various AutoDock versions, we used AutoDock3.0.5,[6] AutoDock4.2.3,[7] and AutoDock Vina1.1.1[8] as a docking engine. The program Glide[5] was used with Glide Script version 56107. In addition to the respective AutoDock scoring functions of version 3 and 4, we used the distance-dependent pair potentials of DrugScore.[9] In these cases, DrugScore potential grids were pre-calculated. For all AutoDock and Drugscore potential grids, a grid spacing of 0.375 Å and a grid dimension of 61x61x61 grid points (resulting in a grid size of 22.5x22.5x22.5 $Å^3$) was used. The same was applied for Glide grid calculations. Grid dimensions were chosen to be identical for the different docking tools to provide for the same search space, thus enabling a comparability of the results.

Docking parameters

Docking with AutoDock3 and 4 was performed using the Lamarkian genetic algorithm.[6] Each docking experiment was repeated 100 times, yielding 100 docked protein-ligand configurations. The following parameters were used for docking: A population size of 200, a random starting position and conformation, a maximal effect of mutation of 2 Å for translation and 50° for rotation, a mutation rate of 0.02 and a crossover rate of 0.80, a local search rate of 0.02, a maximum number of consecutive successes or failures before a change is made to the stepsize of the local search of 4. All docking simulations were performed with a maximum of 50.000 generations. The number of energy evaluations was set to $3*10^6$. Finally, all docked conformations were clustered using a tolerance of 0.5 Å rmsd. Docking with AutoDock Vina and Glide was performed using default settings based on the assumption that the standard parameters identified by the authors perform best.

Evaluation of docking accuracy

The docking accuracy was evaluated by calculating the rmsd between docked ligand poses and the experimental ligand configuration. For AutoDock3 and 4, the ligand pose found on the first scoring rank of the largest cluster generated by AutoDock was chosen. This follows the idea carried over from protein folding that the narrow well of the native state is

surrounded by a broad, shallower minimum in which near-native conformations can be found. Accordingly, it has been hypothesized that native-like structures reside in a broader energy minimum than non-native structures.[10] This then implies that instead of focusing solely on the conformation of best scoring, the best-scoring structure of the largest cluster of structurally related solutions should be considered the most probable pose.[10-12] For AutoDock Vina and Glide, the ligand pose found on the first scoring rank was considered. We note that AutoDock Vina results are already clustered internally.[13]

Tables

Table S1: Results for re-docking on 80 protein-ligand complexes from the PPIM dataset for DrugScore-adapted AutoDock3 and GlideSP.

Target protein	PDB ID	AD3DS [a]	GlideSP SEP[b]	iRMSD ub[c]	Peptide [d]	Complex PDB ID[e]	Unbound PDB ID[f]	Reference [g]
BcL-2	4AQ3	0.98	10.11	2.04	no	n.d.	1G5M	14-16
BcL-2	2O22	2.74	5.71	2.23	no			
BcL-2	2O2F	0.73	11.83	2.27	no			
BcL-2	1YSW	1.51	10.99	2.24	no			
BcL-2 / BcL-XL chimera	2W3L	0.71	6.48	-	no	n.d.	1R2D	14-16
BcL-XL	2O2M	3.04	11.45	2.44	no	1BXL	1R2D	14-16
BcL-XL	2O2N	1.71	3.22	2.70	no	(bak)		
BcL-XL	3INQ	1.47	6.34	2.51	no	2BZW		
BcL-XL	3QKD	1.26	n.d.	2.44	no	(bad)		
BcL-XL	1YSG	8.32	3.10	2.76	no			
BcL-XL	1YSI	1.75	10.49	2.39	no			
BcL-XL	1YSN	1.69	12.42	2.62	no			
BcL-XL	2YXJ	1.67	12.68	2.48	no			
Calmodulin	1A29	1.41	2.45	9.73	no	1CFF	1CCL	17, 18
Calmodulin	1CTR	4.38	5.28	9.77	no			
Calmodulin	1LIN	0.99	1.35	8.17	no			
Calmodulin	1MUX	5.94	3.13	1.80	no			
Calpain	1ALW	5.21	1.06	2.29	no	n.d.	1AJ5	19, 20
Calpain	1NX0	4.98	4.14	2.63	yes			
DIAP1-BIR1	1SE0	1.80	4.12	-	yes	3SIP	n.d.	21
DIAP1-BIR1	1SDZ	0.99	11.13	-	yes			
DIAP1-BIR2	1JD5	2.39	3.95	-	yes	1JD4	n.d.	22
DIAP1-BIR2	1JD6	3.27	5.65	-	yes			
DIAP1-BIR2	1Q4Q	4.86	11.15	-	yes			
Grb2-SH2	1BM2	0.74	0.95	1.51	yes	n.d.	3EAC	23-25
Grb2-SH2	1JYR	4.40	10.81	1.43	yes			
Grb2-SH2	1ZFP	4.79	6.29	1.35	yes			
HexA	2GK1	0.97	4.49	0.68	no	n.d.	2GJX	26
hGST P1-1	22GS	3.65	5.52	4.93	no	n.d.	2LJR	27

HIV Integrase	3LPT	12.06	7.09	1.01	no	2B4J	2ITG	[15, 28]
HPV11-E2	1R6N	0.54	4.19	1.25	no	1TUE	1QQH	[15, 16]
Human Cardiac Troponin C	1DTL	3.53	5.57	2.69	no	1OZS	3SD6	[29, 30]
Human Cardiac Troponin C	1IH0	6.44	9.51	4.69	no			
Human Cardiac Troponin C	2KDH	4.42	7.09	5.70	no			
Human Cardiac Troponin C	2KFX	3.97	4.26	3.34	no			
IL2	1M48	1.05	0.64	1.73	no	1Z92	1M47	[15, 16, 31]
IL2	1M49	1.62	3.77	1.76	no			
IL2	1PW6	0.45	0.88	1.80	no			
IL2	1PY2	0.81	4.51	1.55	no			
IL2	1QVN	2.45	n.d.	1.86	no			
INTEGRIN Alpha (Iib) beta3	2VC2	0.98	3.38	0.98	no	n.d.	3FCS	[32-34]
INTEGRIN Alpha (Iib) beta3	2VDM	1.13	1.43	0.98	no			
INTEGRIN Alpha (Iib) beta3	2VDN	2.36	3.83	1.39	yes			
INTEGRIN Alpha (Iib) beta3	2VDO	5.60	n.d.	0.94	yes			
INTEGRIN Alpha (Iib) beta3	2VDP	3.50	4.22	0.93	yes			
INTEGRIN Alpha (Iib) beta3	2VDR	5.07	8.34	0.97	yes			
MDM2	2AXI	1.15	n.d.	2.81	yes	1YCQ	1Z1M	[15, 16, 35]
MDM2	1T4E	0.66	0.60	2.42	no			
MDM2	2GV2	4.23	n.d.	2.58	yes			
MDM2	1RV1	1.32	9.21	2.43	no			
MDM2	1TTV	4.73	0.69	2.60	no			
MDM4	3FE7	5.08	n.d.	-	yes	3DAB	n.d.	[35, 36]
MDM4	3FEA	4.61	n.d.	-	yes			
ML-IAP-BIR	1OY7	1.11	1.84	-	yes	n.d.	n.d.	[37, 38]
ML-IAP-BIR	1OXN	3.38	3.34	-	yes			
ML-IAP-BIR	1OXQ	0.67	1.20	-	yes			
ML-IAP-BIR	3F7G	0.68	0.60	-	no			
ML-IAP-BIR / XIAP-BIR3	2I3H	0.61	0.78	-	yes	n.d.	n.d.	[38, 39]
ML-IAP-BIR / XIAP-BIR3	2I3I	1.02	0.61	-	yes			
ML-IAP-BIR / XIAP-BIR3	1TW6	0.56	0.83	-	yes			
ML-IAP-BIR / XIAP-BIR3	3F7H	0.66	2.18	-	no			
ML-IAP-BIR / XIAP-BIR3	3F7I	0.71	0.45	-	no			
TNF-alpha	2AZ5	2.48	10.67	5.03	no	1TNF	2E7A	[15, 40]
TNFRc1	1FT4	7.60	7.09	1.23	no	1TNR	1EXT	[15, 41]

XIAP-BIR3	1TFQ	1.16	1.17	4.15	yes	1NW9 (caspas-9) 1G73 (smac)	1F9X	[15, 16, 42, 43]
XIAP-BIR3	1TFT	1.13	1.25	3.99	yes			
XIAP-BIR3	2JK7	1.52	1.46	4.22	no			
XIAP-BIR3	3CM2	1.36	1.19	4.19	no			
XIAP-BIR3	3CM7	0.65	1.87	4.29	no			
XIAP-BIR3	3CLX	0.95	0.99	4.24	no			
XIAP-BIR3	2OPY	5.05	4.59	4.42	no			
XIAP-BIR3	2OPZ	0.55	0.98	4.29	yes			
XIAP-BIR3	3G76	5.74	7.67	4.19	no			
XIAP-BIR3	3EYL	0.77	0.79	4.22	no			
XIAP-BIR3	3HL5	3.36	3.75	4.57	no			
XIAP-BIR3	2VSL	1.75	1.37	4.41	no			
ZipA	1S1J	3.68	5.40	0.83	no	1F47	1F46	[15, 16, 44]
ZipA	1S1S	5.27	4.75	0.96	no			
ZipA	1Y2F	8.74	4.26	1.13	no			
ZipA	1Y2G	5.30	8.33	0.98	no			
% ≤2Å		**53.8 (62.1)***	**30.0 (45.5)***					

* Ligands with > 20 rotatable bonds were neglected.

[a] Rmsd values (in Å) given for docking with AutoDock3.0.5 using DrugScore as objective function.

[b] Rmsd values (in Å) given for docking with GLIDE Script version 56107. Protein and ligand were seperated and prepared independently: The protein was prepared with PrepWizard, and the ligand was prepared with LigPrep from Maestro version 9.2.112. “n.d.” indicates that no results could be determined due to scoring penalty reasons.

[c] Interface-RMSD between the unbound protein structure and the ligand-bound protein structure. In Å. The interface was defined as all residues within 5Å distance of the ligand.

[d] Indicates whether the ligand is a (modified) peptide or not.

[e] “n.d.” indicates that no unbound protein structure is available.

[f] “n.d.” indicates that no protein-protein complex structure is available.

[g] For each target protein a reference is given where the protein interface and the corresponding protein-protein complex that should be inhibited is described.

Table S2: Results for re-docking on the CCDC/Astex clean set[45] with three versions of AutoDock, two DrugScore-adapted AutoDock versions, and GLIDE.[a]

PDB ID	AD3[b]		AD4[e]		AD3_DS[f]		AD4_DS[g]		VINA[h]		GLIDE[k]	
	first[c]	clust[d]	first	clust	first	clust	first	clust	u.b.[i]	l.b.[j]	first[l]	first[m]
1a28	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	1.2	1.2	0.8	1.2
1a42	2.8	3.4	2.9	2.9	2.9	2.9	2.9	2.9	1.4	1.4	1.3	5.1
1a4g	1.1	1.1	1.0	1.0	1.1	1.0	1.0	1.0	2.5	2.5	2.6	3.2
1a4q	1.0	0.5	2.1	2.1	0.8	0.8	2.1	2.1	3.0	3.0	0.6	0.7
1a6w	1.5	1.5	1.1	1.6	1.2	1.2	1.1	1.6	2.4	2.4	1.3	3.7
1a9u	5.2	0.4	5.4	5.4	5.5	5.5	5.4	5.4	4.1	4.1	8.4	6.9
1aaq	1.7	1.7	2.0	2.0	1.7	1.7	2.0	2.0	2.0	2.0	3.6	4.8
1abe	1.0	1.0	0.8	0.8	0.9	0.9	0.8	0.8	1.1	1.1	3.4	0.3
1abf	0.9	0.9	0.6	0.6	0.9	0.9	0.6	0.6	1.2	1.2	1.2	0.3
1acj	3.3	3.3	0.9	0.9	0.9	0.9	0.9	0.9	0.4	0.4	1.2	0.6
1acl	4.0	4.0	3.7	3.7	3.7	3.7	3.7	3.7	0.7	0.7	2.3	0.9
1acm	9.6	9.6	0.8	0.8	0.9	0.9	0.8	0.8	2.2	2.2	1.1	3.4
1aco	10.4	10.8	0.5	0.5	1.3	0.5	0.5	0.5	1.3	1.3	3.1	1.3
1aec	4.0	4.0	3.2	3.2	3.1	1.1	3.2	3.2	3.5	3.5	7.0	3.4
1ai5	2.4	2.4	8.0	7.7	7.7	0.8	8.0	7.7	2.6	2.6	5.4	4.5
1aoe	0.9	0.9	4.4	4.4	4.4	4.4	4.4	4.4	4.1	4.1	7.6	0.7
1apt	0.9	0.9	1.3	1.3	1.8	1.8	1.3	1.3	1.5	1.5	9.7	1.7
1apu	1.7	1.6	1.4	1.4	1.7	2.0	1.4	1.4	1.8	1.8	4.0	9.0
1aqw	4.3	2.7	4.1	6.4	4.1	2.6	4.1	6.4	10.4	10.4	0.9	0.8
1ase	3.5	3.5	2.4	2.4	2.4	2.4	2.4	2.4	5.3	5.3	8.5	1.9
1atl	3.1	0.9	3.0	1.2	2.8	1.2	3.0	1.2	3.1	3.1	3.6	4.1
1azm	3.0	2.2	4.0	4.0	4.1	4.1	4.0	4.0	2.1	2.1	1.0	1.7
1b58	2.0	0.8	1.0	1.0	1.3	1.1	1.0	1.0	1.8	1.8	1.5	0.7
1b59	2.7	1.6	4.2	2.4	4.1	1.2	4.2	2.4	1.8	1.8	3.4	1.8
1b9v	1.4	1.2	4.1	0.9	4.0	0.8	4.1	0.9	2.2	2.2	1.2	1.0
1baf	4.1	4.1	4.4	4.4	4.6	4.6	4.4	4.4	2.2	2.2	7.5	7.5
1bbp	0.8	1.0	0.7	0.8	0.9	0.9	0.7	0.8	1.7	1.7	1.6	1.7
1bgo	1.5	1.5	1.6	1.6	4.1	2.0	1.6	1.6	1.6	1.6	2.4	2.3
1bl7	8.5	8.5	9.0	9.0	8.8	8.8	9.0	9.0	1.2	1.2	0.8	0.5
1blh	5.9	6.3	4.8	5.0	4.3	4.3	4.8	5.0	0.3	0.3	4.7	3.2
1bma	0.8	0.8	0.7	0.7	0.8	0.8	0.7	0.7	3.2	3.2	1.4	1.4
1bmq	1.9	1.7	1.6	1.6	1.6	1.6	1.6	1.6	1.3	1.3	4.8	3.9
1byb	0.8	1.0	1.0	1.0	1.1	1.2	1.0	1.0	1.5	1.5	4.8	2.0
1byg	1.0	1.0	0.5	0.5	0.4	0.4	0.5	0.5	2.3	2.3	0.6	0.6
1c12	5.4	3.7	4.5	4.5	4.5	3.9	4.5	4.5	2.0	2.0	9.3	9.1
1c1e	0.3	0.3	9.8	9.8	9.8	9.8	9.8	9.8	3.3	3.3	1.3	6.6

1c5c	0.9	0.9	3.0	4.5	4.6	4.6	3.0	4.5	2.5	2.5	6.4	0.3
1c5x	4.7	4.7	0.9	0.9	0.8	0.8	0.9	0.9	2.5	2.5	1.7	0.3
1c83	0.7	0.7	0.4	0.4	0.6	0.6	0.4	0.4	3.9	3.9	0.6	0.9
1cbs	9.7	5.1	6.1	6.1	6.3	6.3	6.1	6.1	0.7	0.7	3.0	2.0
1cbx	1.9	5.6	0.6	0.6	0.7	0.7	0.6	0.6	3.1	3.1	0.8	0.9
1cdg	2.6	2.9	2.7	3.0	3.0	2.6	2.7	3.0	1.5	1.5	2.5	2.8
1cil	4.2	1.9	3.0	2.7	3.0	2.8	3.0	2.7	2.4	2.4	4.7	4.0
1ckp	5.8	5.8	4.9	0.7	0.7	0.7	4.9	0.7	3.2	3.2	2.0	1.7
1cle	2.4	2.4	5.2	5.2	4.4	4.4	5.2	5.2	3.1	3.1	12.9	11.3
1com	0.7	0.7	0.5	1.5	0.4	0.4	0.5	1.5	1.1	1.1	1.2	3.3
1coy	7.2	2.9	0.8	0.8	0.8	0.8	0.8	0.8	1.1	1.1	1.0	0.5
1cps	3.1	1.1	5.9	6.1	4.7	5.9	5.9	6.1	2.7	2.7	1.9	1.1
1cqp	2.5	1.2	3.1	3.1	3.1	2.0	3.1	3.1	2.2	2.2	2.3	6.6
1cvu	5.3	1.2	5.8	5.8	5.4	0.8	5.8	5.8	1.1	1.1	7.1	7.4
1cx2	1.6	1.6	6.2	6.1	6.1	1.3	6.2	6.1	0.3	0.3	1.2	0.9
1d0l	1.8	1.5	3.1	3.2	1.8	1.6	3.1	3.2	1.7	1.7	3.4	7.0
1d3h	1.6	1.6	2.0	5.7	5.6	5.6	2.0	5.7	9.0	9.0	2.0	1.5
1d4p	4.1	0.9	0.6	0.6	0.7	0.7	0.6	0.6	1.6	1.6	0.3	1.6
1dbb	4.0	3.8	0.7	0.7	0.7	0.7	0.7	0.7	1.6	1.6	1.7	2.6
1dbj	5.0	5.0	0.9	0.9	0.9	0.9	0.9	0.9	2.2	2.2	4.4	2.3
1dd7	5.2	4.7	3.7	3.4	3.4	0.6	3.7	3.4	3.3	3.3	7.7	6.3
1dg5	6.7	5.1	5.3	5.3	5.7	5.7	5.3	5.3	0.4	0.4	4.4	5.9
1dhf	4.8	1.1	6.6	3.7	6.9	1.0	6.6	3.7	1.7	1.7	1.3	5.0
1did	1.5	1.7	8.7	1.1	1.2	1.2	8.7	1.1	3.7	3.7	4.0	4.5
1dmp	0.8	0.8	0.7	0.7	0.6	0.6	0.7	0.7	0.5	0.5	1.2	0.8
1dog	2.1	1.5	1.0	1.6	1.0	1.0	1.0	1.6	1.1	1.1	0.8	0.4
1dr1	2.4	2.4	1.1	1.1	1.4	1.4	1.1	1.1	1.8	1.8	4.6	6.4
1dwb	2.0	2.0	9.2	9.2	9.3	9.3	9.2	9.2	0.0	0.0	1.9	0.3
1dwc	4.1	4.1	4.9	4.2	4.3	4.5	4.9	4.2	2.4	2.4	2.7	4.2
1dwd	2.1	2.3	4.1	3.0	2.9	0.9	4.1	3.0	2.0	2.0	5.1	6.4
1dy9	6.6	6.6	7.2	7.2	3.1	3.1	7.2	7.2	2.8	2.8	3.0	3.6
1eap	4.7	0.9	1.4	4.7	4.9	1.2	1.4	4.7	1.7	1.7	1.8	2.4
1ebg	12.7	12.7	1.8	1.8	1.6	1.6	1.8	1.8	0.9	0.9	2.0	2.5
1eed	3.0	3.0	1.7	1.7	1.3	1.3	1.7	1.7	3.2	3.2	11.4	11.7
1ei1	0.9	0.8	3.1	2.8	0.4	0.4	3.1	2.8	1.2	1.2	2.1	4.8
1ejn	1.2	1.2	0.6	1.2	1.1	1.1	0.6	1.2	1.2	1.2	2.2	0.8
1eoc	1.1	1.1	0.5	0.5	0.8	0.8	0.5	0.5	1.8	1.8	4.8	2.3
1epb	2.0	1.7	2.3	2.4	2.0	2.0	2.3	2.4	1.9	1.9	2.7	2.9
1epo	1.1	1.1	1.3	1.3	0.8	0.6	1.3	1.3	1.9	1.9	10.4	7.1
1eta	8.1	5.7	9.2	8.4	8.8	2.3	9.2	8.4	5.0	5.0	8.3	4.6
1etr	3.6	1.0	5.2	5.2	4.8	1.5	5.2	5.2	1.7	1.7	6.6	2.7

1ets	1.8	1.4	2.7	1.1	1.0	1.0	2.7	1.1	3.4	3.4	6.5	3.9
1ett	3.0	0.8	2.8	3.8	5.6	2.4	2.8	3.8	2.2	2.2	17.4	2.0
1f0r	1.7	1.7	1.1	1.1	1.2	1.2	1.1	1.1	2.8	2.8	2.7	4.2
1f0s	1.7	1.9	2.3	2.3	1.5	1.5	2.3	2.3	1.5	1.5	8.9	3.3
1f3d	1.0	1.0	0.3	0.3	1.0	0.4	0.3	0.3	1.3	1.3	1.4	1.2
1fax	1.8	1.4	2.9	2.5	2.7	1.8	2.9	2.5	1.8	1.8	1.7	1.6
1fen	1.2	1.2	0.9	0.9	2.0	1.7	0.9	0.9	1.2	1.2	1.2	9.0
1fgi	4.1	4.1	1.8	1.8	1.8	1.8	1.8	1.8	1.9	1.9	1.0	1.6
1fkg	1.1	0.9	1.8	0.9	0.8	0.8	1.8	0.9	2.1	2.1	1.8	4.6
1fki	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	1.6	1.6	1.6	1.1
1fl3	1.9	1.5	1.7	1.7	1.8	1.8	1.7	1.7	1.5	1.5	1.3	1.5
1flr	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	n.d.	n.d.	4.8	0.8
1frp	9.8	11.7	0.9	2.9	1.1	1.1	0.9	2.9	1.9	1.9	1.4	1.4
1glp	9.1	7.0	6.0	6.0	5.9	3.9	6.0	6.0	5.8	5.8	1.3	1.8
1glq	1.5	1.7	8.2	8.2	6.6	3.5	8.2	8.2	3.3	3.3	1.3	4.8
1hak	5.1	1.5	4.1	5.0	5.0	5.0	4.1	5.0	4.6	4.6	9.9	8.9
1hdc	2.1	1.5	5.8	3.3	3.4	3.3	5.8	3.3	6.2	6.2	3.6	3.3
1hfc	2.0	1.9	4.8	4.6	1.7	1.7	4.8	4.6	2.1	2.1	3.0	2.8
1hiv	1.8	1.8	3.9	3.9	1.7	1.7	3.9	3.9	1.3	1.3	11.4	3.7
1hos	2.0	2.0	2.3	2.3	1.7	1.9	2.3	2.3	0.3	0.3	5.5	2.7
1hpv	3.5	3.0	1.4	1.4	0.9	0.9	1.4	1.4	1.5	1.5	4.4	4.0
1hri	2.2	2.3	11.6	6.0	3.1	3.0	11.6	6.0	4.9	4.9	2.3	2.0
1hsb	2.5	1.0	1.6	1.6	2.9	1.6	1.6	1.6	2.5	2.5	1.2	4.7
1hsl	2.0	0.6	0.6	1.6	1.4	1.4	0.6	1.6	1.4	1.4	2.7	0.9
1htf	2.8	1.8	3.7	3.4	3.4	1.7	3.7	3.4	4.7	4.7	9.5	4.0
1hvr	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.1	0.1	1.1	0.9
1hyt	5.2	5.2	0.8	2.0	0.9	0.9	0.8	2.0	5.4	5.4	1.1	2.0
1ibg	4.5	1.8	1.4	1.4	1.4	1.4	1.4	1.4	0.8	0.8	3.3	9.0
1ida	1.1	1.1	2.6	2.6	3.0	1.2	2.6	2.6	2.7	2.7	4.9	10.8
1imb	5.6	5.6	6.2	6.2	5.9	5.9	6.2	6.2	4.0	4.0	1.8	5.1
1ivb	9.2	9.2	3.1	3.1	3.4	3.4	3.1	3.1	2.7	2.7	5.4	0.6
1ivq	1.0	1.0	2.0	2.0	0.7	1.2	2.0	2.0	1.3	1.3	2.0	4.3
1jap	5.2	5.2	7.6	6.9	5.9	5.9	7.6	6.9	1.0	1.0	2.2	6.0
1kel	1.4	1.4	1.1	1.1	1.2	1.2	1.1	1.1	1.9	1.9	1.7	4.7
1lah	1.3	0.3	0.5	0.5	0.5	0.5	0.5	0.5	1.6	1.6	1.3	0.4
1lcp	2.8	2.8	1.3	1.3	1.5	1.5	1.3	1.3	1.0	1.0	3.1	3.8
1ldm	8.3	8.3	8.4	8.4	8.7	8.7	8.4	8.4	2.0	2.0	1.6	2.2
1lic	7.3	7.3	4.8	4.8	4.7	4.7	4.8	4.8	1.4	1.4	3.5	11.4
1lna	6.9	4.7	3.3	3.5	3.5	3.5	3.3	3.5	1.9	1.9	3.5	4.4
1lpm	1.8	2.3	1.2	1.2	1.2	1.2	1.2	1.2	1.6	1.6	9.8	2.6
1lst	3.0	1.5	1.1	1.1	1.1	1.1	1.1	1.1	1.4	1.4	1.5	1.0

1lyb	1.6	1.6	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.6	4.0
1lyl	1.4	1.4	1.3	2.4	1.3	1.3	1.3	2.4	8.9	8.9	1.7	1.3
1mbi	14.2	8.0	14.6	6.7	14.6	6.7	14.6	6.7	2.0	2.0	8.6	5.8
1mcq	5.9	4.8	5.9	5.2	5.9	5.9	5.9	5.2	2.2	2.2	5.9	5.1
1mdr	9.4	10.2	9.9	9.9	9.8	9.8	9.9	9.9	9.1	9.1	3.2	2.1
1mld	13.3	8.3	1.4	6.0	9.5	0.7	1.4	6.0	1.3	1.3	2.2	3.1
1mmq	1.0	1.0	1.2	1.2	1.1	1.1	1.2	1.2	1.6	1.6	8.6	7.9
1mrg	2.8	2.8	0.4	0.4	0.5	0.5	0.4	0.4	1.3	1.3	0.5	0.5
1mrk	2.6	2.3	2.1	2.1	2.1	2.1	2.1	2.1	1.2	1.2	1.9	3.0
1mts	1.2	1.0	0.9	0.9	1.9	0.9	0.9	0.9	1.5	1.5	1.4	1.9
1mup	3.2	3.2	3.1	3.1	2.7	2.7	3.1	3.1	2.3	2.3	5.1	4.1
1nco	3.5	1.2	2.6	1.2	1.3	1.3	2.6	1.2	2.6	2.6	9.3	8.4
1ngp	1.3	0.8	0.8	0.8	3.1	3.1	0.8	0.8	1.8	1.8	4.9	4.3
1nis	11.0	11.0	1.2	1.4	0.9	0.9	1.2	1.4	1.9	1.9	5.8	0.9
1okl	1.1	1.1	2.8	2.8	2.7	2.7	2.8	2.8	1.1	1.1	5.1	0.7
1okm	3.0	2.7	5.7	3.5	5.7	5.7	5.7	3.5	2.7	2.7	3.8	2.3
1pbd	6.3	6.3	0.3	0.3	0.7	0.7	0.3	0.3	5.0	5.0	0.9	2.3
1pdz	16.6	11.7	10.5	1.9	10.4	2.0	10.5	1.9	1.8	1.8	6.3	4.8
1phd	9.0	9.0	5.1	5.1	5.1	5.1	5.1	5.1	2.1	2.1	1.9	7.9
1phg	6.5	6.5	0.5	1.9	0.6	0.6	0.5	1.9	2.3	2.3	8.3	6.2
1poc	2.5	2.5	1.3	1.3	1.3	1.3	1.3	1.3	1.9	1.9	2.9	1.4
1ppc	1.7	1.5	5.5	3.4	1.1	1.1	5.5	3.4	4.3	4.3	3.7	1.8
1pph	5.2	2.5	3.8	3.8	5.8	5.8	3.8	3.8	2.9	2.9	4.0	5.1
1ppi	1.0	1.0	1.1	1.1	0.8	0.8	1.1	1.1	1.2	1.2	9.5	9.4
1pso	2.0	2.0	2.0	2.0	2.8	2.8	2.0	2.0	1.6	1.6	3.2	3.5
1ptv	4.6	4.9	5.1	5.1	4.6	1.9	5.1	5.1	0.2	0.2	2.5	2.0
1qbr	0.9	0.9	0.9	0.9	0.8	0.8	0.9	0.9	0.2	0.2	0.9	0.8
1qbu	0.5	0.5	0.8	0.8	0.9	0.9	0.8	0.8	1.7	1.7	1.1	0.8
1qcf	0.6	0.6	3.9	3.9	4.0	4.0	3.9	3.9	0.1	0.1	0.5	0.3
1qpe	4.8	0.6	0.6	0.6	0.5	0.5	0.6	0.6	1.6	1.6	1.1	0.4
1qpq	9.0	9.0	8.8	1.2	9.4	1.2	8.8	1.2	0.9	0.9	3.3	1.4
1rds	1.2	1.0	1.1	0.9	1.1	1.5	1.1	0.9	2.3	2.3	2.3	0.9
1rne	1.4	1.4	1.1	1.1	1.0	1.3	1.1	1.1	1.5	1.5	1.6	1.0
1rnt	2.3	4.1	1.1	1.1	1.0	1.0	1.1	1.1	2.9	2.9	3.6	2.4
1rob	2.7	2.7	1.4	1.4	1.1	1.1	1.4	1.4	4.6	4.6	2.9	4.5
1rt2	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	n.d.	n.d.	0.4	0.5
1slt	1.6	1.6	4.5	4.3	4.2	4.2	4.5	4.3	3.9	3.9	1.2	1.2
1snc	3.4	4.8	3.1	3.5	4.0	3.7	3.1	3.5	1.6	1.6	2.3	2.6
1srj	3.3	1.9	0.9	4.2	4.1	4.1	0.9	4.2	1.8	1.8	0.6	0.7
1tdb	6.1	1.5	6.7	6.7	8.2	6.9	6.7	6.7	4.3	4.3	8.6	8.8
1tlp	3.6	3.6	4.6	4.6	4.6	2.4	4.6	4.6	1.4	1.4	2.6	5.1

1tmn	4.6	2.7	5.5	5.5	3.0	3.0	5.5	5.5	2.2	2.2	1.9	8.4
1tng	3.5	3.5	1.0	1.0	3.6	3.6	1.0	1.0	0.8	0.8	0.4	0.2
1tnh	3.8	3.8	1.0	1.0	1.0	1.0	1.0	1.0	3.1	3.1	0.3	0.6
1tni	2.1	2.0	2.0	1.8	2.0	2.0	2.0	1.8	1.7	1.7	2.9	2.6
1tnl	4.8	4.8	4.8	4.8	4.9	4.9	4.8	4.8	1.0	1.0	2.4	2.1
1tpp	3.6	2.1	1.9	1.9	5.6	2.1	1.9	1.9	1.5	1.5	3.1	3.1
1trk	1.0	0.7	0.8	1.6	0.9	0.9	0.8	1.6	1.0	1.0	1.2	1.8
1tyl	0.9	0.9	4.3	4.3	4.3	4.3	4.3	4.3	1.6	1.6	1.8	1.1
1ukz	9.2	4.6	5.2	5.2	5.3	1.2	5.2	5.2	1.2	1.2	1.8	3.5
1ulb	6.0	6.0	0.5	0.5	0.5	0.5	0.5	0.5	6.5	6.5	0.5	0.4
1uvs	2.6	3.9	3.4	3.7	3.3	3.2	3.4	3.7	2.9	2.9	2.3	1.3
1uvt	5.7	1.1	3.9	3.9	3.8	1.0	3.9	3.9	1.7	1.7	2.8	4.3
1vgc	9.2	9.0	6.0	5.8	5.8	5.8	6.0	5.8	2.0	2.0	3.0	2.2
1wap	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	1.6	1.6	0.8	0.3
1xid	2.3	2.3	2.7	2.7	2.2	2.2	2.7	2.7	1.4	1.4	4.1	5.2
1xie	6.5	6.9	1.8	1.8	9.1	1.8	1.8	1.8	2.2	2.2	4.3	4.5
1ydr	1.5	1.5	1.4	1.4	1.4	1.4	1.4	1.4	2.3	2.3	2.6	1.7
1ydt	0.7	0.7	2.8	2.8	3.3	3.2	2.8	2.8	4.7	4.7	2.8	1.1
1yee	5.9	5.9	5.7	5.6	2.8	3.1	5.7	5.6	2.6	2.6	3.3	3.8
25c8	1.4	1.8	1.8	1.6	2.0	1.6	1.8	1.6	1.6	1.6	1.3	2.0
2aad	0.9	0.9	0.8	0.7	0.6	0.6	0.8	0.7	1.6	1.6	1.0	1.4
2ack	1.3	1.3	4.0	4.0	3.7	3.7	4.0	4.0	2.6	2.6	3.6	4.1
2ada	0.5	0.5	0.6	0.6	0.8	0.6	0.6	0.6	1.5	1.5	2.9	0.5
2ak3	6.6	1.9	2.2	2.2	2.2	2.1	2.2	2.2	2.2	2.2	4.0	4.2
2cht	0.4	0.4	7.2	0.5	0.4	0.4	7.2	0.5	1.5	1.5	4.1	0.6
2cmd	10.4	1.2	0.9	0.9	1.2	1.2	0.9	0.9	1.4	1.4	2.3	3.2
2cpp	5.5	5.5	12.2	6.0	12.2	6.0	12.2	6.0	1.1	1.1	2.8	2.1
2ctc	2.2	2.2	0.7	0.7	0.5	1.3	0.7	0.7	0.1	0.1	0.9	0.6
2dbl	1.1	1.1	0.7	1.0	0.8	0.8	0.7	1.0	3.3	3.3	1.2	1.0
2fox	0.5	1.8	0.5	0.5	0.6	0.6	0.5	0.5	2.3	2.3	3.2	1.2
2gbp	0.5	0.5	0.8	0.8	0.7	0.7	0.8	0.8	1.2	1.2	3.2	2.8
2h4n	4.7	4.7	4.8	4.6	5.0	5.0	4.8	4.6	2.4	2.4	5.1	5.2
2ifb	5.3	5.3	3.1	3.1	0.9	0.9	3.1	3.1	1.8	1.8	1.6	0.9
2lgs	2.4	2.4	1.7	1.7	1.8	1.2	1.7	1.7	1.4	1.4	5.9	2.1
2mcp	2.8	3.0	10.1	4.4	10.4	10.4	10.1	4.4	1.0	1.0	3.6	4.0
2pcp	0.6	0.6	2.6	2.6	0.5	0.5	2.6	2.6	0.8	0.8	0.9	0.7
2phh	5.8	5.8	0.8	0.8	0.9	0.9	0.8	0.8	1.9	1.9	0.7	1.8
2pk4	1.0	1.0	12.2	2.1	12.1	1.9	12.2	2.1	1.3	1.3	2.0	1.4
2qwk	2.0	2.0	4.0	0.8	2.2	2.2	4.0	0.8	0.3	0.3	1.4	0.9
2r07	1.7	1.5	6.1	6.0	6.2	1.0	6.1	6.0	5.6	5.6	1.7	11.3
2tmn	4.4	6.4	0.9	0.9	0.7	0.7	0.9	0.9	2.1	2.1	3.3	3.7

2tsc	3.2	3.2	4.5	2.3	2.4	2.4	4.5	2.3	2.2	2.2	9.9	4.9
2yhx	5.4	2.4	5.8	4.9	3.8	3.0	5.8	4.9	1.8	1.8	7.9	7.4
2ypi	11.5	11.5	1.0	1.0	1.1	1.1	1.0	1.0	0.4	0.4	5.1	1.0
3cla	4.2	6.3	9.1	8.2	9.0	6.8	9.1	8.2	6.2	6.2	8.4	1.5
3cpa	2.9	4.0	0.6	5.1	5.1	5.1	0.6	5.1	2.8	2.8	2.4	1.3
3erd	1.2	0.8	1.2	0.8	0.5	0.5	1.2	0.8	0.1	0.1	1.2	0.3
3ert	1.6	1.6	1.4	1.5	1.3	1.3	1.4	1.5	0.4	0.4	1.6	1.1
3gpb	9.7	2.3	10.2	10.2	10.3	1.0	10.2	10.2	8.6	8.6	6.5	8.9
3hvt	0.4	0.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	0.7	4.4
3tpi	1.3	1.3	0.5	0.5	0.7	0.5	0.5	0.5	2.2	2.2	1.4	1.0
4aah	12.2	0.8	0.7	0.7	0.8	0.8	0.7	0.7	1.0	1.0	0.7	0.7
4cox	1.0	1.0	1.2	1.2	1.6	1.6	1.2	1.2	9.2	9.2	1.0	1.1
4cts	8.7	1.1	0.7	0.7	0.6	0.6	0.7	0.7	1.6	1.6	1.0	0.3
4dfr	1.6	1.2	6.9	6.7	6.8	6.8	6.9	6.7	3.2	3.2	1.3	1.3
4est	3.2	3.2	5.7	4.4	2.8	3.3	5.7	4.4	4.7	4.7	9.9	6.0
4fbp	4.0	4.0	0.9	1.6	0.9	0.9	0.9	1.6	1.0	1.0	2.4	2.9
4lbd	1.3	1.3	0.4	0.4	0.4	0.4	0.4	0.4	0.7	0.7	0.5	0.3
4phv	2.6	3.3	0.8	0.8	0.6	0.6	0.8	0.8	0.3	0.3	2.3	1.3
5abp	0.9	0.9	0.8	0.8	0.6	0.6	0.8	0.8	1.1	1.1	1.3	0.4
5cpp	8.0	8.0	2.3	2.3	2.3	2.3	2.3	2.3	0.1	0.1	2.2	1.7
5er1	1.8	1.8	5.8	5.8	4.7	4.7	5.8	5.8	2.8	2.8	2.5	8.7
6rnt	5.4	5.4	5.7	5.8	5.8	5.8	5.7	5.8	5.2	5.2	7.4	8.0
6rsa	1.0	1.0	1.3	1.3	1.2	1.2	1.3	1.3	2.3	2.3	n.d.	-
7tim	2.8	3.4	0.9	1.3	1.2	1.0	0.9	1.3	2.2	2.2	2.4	1.2
% ≤2Å	**42.0**	**54.0**	**58.5**	**57.1**	**52.2**	**64.3**	**49.6**	**50.9**	**14.3**	**57.1**	**42.2**	**46.0**

[a] In addition to these results, we also tested the recently developed scoring function DSX.[46] When using DSX for re-scoring DrugScore-adapted AutoDock3 docking results for the CCDC/Astex clean set, we achieved a success rate of 54.9% for identifying a solution with rmsd ≤ 2Å. Just considering cases where at least one near-native solution could be sampled (79.5%, i.e. 178 complexes out of 224 from the CCDC/Astex clean set), DSX achieved a success rate of 69.1% in recognizing such a solution, whereas DrugScore adapted AutoDock3 achieved a success rate of 80.9% in doing so.

[b] Rmsd values (in Å) given for docking with AutoDock3.0.5.[6]

[c] Rmsd between the docking pose found on the first scoring rank and the native solution.

[d] Rmsd between the docking pose found on the first scoring rank of the largest cluster and the native solution.

[e] Rmsd values (in Å) given for docking with AutoDock4.2.3.[7]

[f] Rmsd values (in Å) given for docking with AutoDock3.0.5 using DrugScore as objective function.

[g] Rmsd values (in Å) given for docking with AutoDock4.2.3 using DrugScore as objective function.

[h] Rmsd values (in Å) given for docking with AutoDock Vina1.1.1.[13]

[i] Rmsd/ub matches each atom in one conformation with itself in the other conformation, ignoring any symmetry.[8, 13]

[j] Rmsd/lb of two conformations is defined as follows:

rmsd/lb(c_1, c_2) = max(rmsd'(c_1, c_2), rmsd'(c_2, c_1)),where rmsd' matches each atom in one conformation with the closest atom of the same element type in the other conformation (rmsd' cannot be used directly, because it is not symmetric).[8, 13]

[k] Rmsd values (in Å) given for docking with GLIDE Script version 56107.

[l] Rmsd between the docking pose found on the first scoring rank and the native solution. Proteins were prepared with the PrepWizard using full setup including H-bond assignment and minimization.

[m] Rmsd between the docking pose found on the first scoring rank and the native solution. Proteins were prepared with the PrepWizard by just adding hydrogens.

Table S3: Results for re-docking using differently prepared proteins and ligands of a subset of 135 complexes from the CCDC/Astex clean set[45] used in a study by Friesner *et al.*[5].

PDB ID	GlideSP SEP_F[a]	GlideSP SEP_H[b]	GlideSP REF[c]	GlideSP COM[d]
1aaq	3.6	4.8	1.3	1.2
1abe	3.4	0.3	0.2	0.1
1abf	1.2	0.3	0.2	0.3
1acj	1.2	0.6	0.3	0.1
1acm	1.1	3.4	0.3	1.0
1aco	3.1	1.3	1.0	3.9
1apt	9.7	1.7	0.6	1.5
1apu	4.0	9.0	1.2	1.8
1atl	3.6	4.1	0.9	4.0
1azm	1.0	1.7	1.9	1.0
1baf	7.5	7.5	0.8	5.7
1bbp	1.6	1.7	5.0	0.4
1bma	1.4	1.4	9.3	0.4
1byb	4.8	2.0	10.5	0.4
1c83	0.6	0.9	0.1	0.7
1cbs	3.0	2.0	2.0	2.4
1cbx	0.8	0.9	0.4	0.2
1cdg	2.5	2.8	4.0	4.8
1cil	4.7	4.0	3.8	0.3
1com	1.2	3.3	3.6	0.3
1coy	1.0	0.5	0.3	0.3
1cps	1.9	1.1	3.0	0.8
1dbb	1.7	2.6	0.4	0.5
1dbj	4.4	2.3	0.2	0.2
1did	4.0	4.5	3.8	3.1
1dog	0.8	0.4	3.7	0.3
1dr1	4.6	6.4	1.5	2.7
1dwb	1.9	0.3	0.3	1.1
1dwc	2.7	4.2	0.9	1.4
1dwd	5.1	6.4	1.3	4.5
1eap	1.8	2.4	2.3	0.9
1ebg	2.0	2.5	0.2	1.3
1eed	11.4	11.7	5.9	6.0
1ejn	2.2	0.8	0.7	0.5
1epb	2.7	2.9	1.8	1.6
1eta	8.3	4.6	2.9	1.6

1etr	6.6	2.7	1.5	5.4
1ets	6.5	3.9	1.4	2.9
1ett	17.4	2.0	0.6	0.6
1f0r	2.7	4.2	1.4	1.4
1f0s	8.9	3.3	1.6	1.0
1fen	1.2	9.0	0.7	0.6
1fkg	1.8	4.6	1.3	0.7
1fki	1.6	1.1	1.9	1.1
1frp	1.4	1.4	0.3	0.5
1glp	1.3	1.8	0.3	0.2
1glq	1.3	4.8	0.3	3.6
1hdc	3.6	3.3	0.6	0.8
1hfc	3.0	2.8	2.2	2.7
1hpv	4.4	4.0	0.9	0.5
1hri	2.3	2.0	1.6	1.2
1hsl	2.7	0.9	1.3	1.5
1htf	9.5	4.0	3.0	2.5
1hvr	1.1	0.9	1.5	0.2
1hyt	1.1	2.0	0.3	0.4
1ida	4.9	10.8	11.9	2.0
1imb	1.8	5.1	0.9	5.4
1ivb	5.4	0.6	5.0	0.8
1lah	1.3	0.4	0.1	0.2
1lcp	3.1	3.8	2.0	2.5
1ldm	1.6	2.2	0.3	0.3
1lic	3.5	11.4	4.9	2.0
1lna	3.5	4.4	1.0	5.9
1lst	1.5	1.0	0.1	0.3
1mbi	8.6	5.8	1.7	0.4
1mdr	3.2	2.1	0.5	0.9
1mld	2.2	3.1	0.3	1.6
1mmq	8.6	7.9	0.9	0.2
1mrg	0.5	0.5	0.3	0.2
1mrk	1.9	3.0	1.2	1.7
1mup	5.1	4.1	4.4	3.0
1nco	9.3	8.4	7.0	7.4
1nis	5.8	0.9	1.0	3.6
1okl	5.1	0.7	2.8	0.2
1pbd	0.9	2.3	0.2	0.2
1phd	1.9	7.9	1.2	1.1
1phg	8.3	6.2	4.3	5.5

1poc	2.9	1.4	5.1	8.1
1ppc	3.7	1.8	7.9	1.1
1pph	4.0	5.1	4.3	4.5
1ppi	9.5	9.4	6.2	3.4
1pso	3.2	3.5	13.1	4.3
1rds	2.3	0.9	3.8	0.9
1rne	1.6	1.0	10.1	0.5
1rnt	3.6	2.4	0.7	0.5
1rob	2.9	4.5	1.9	0.8
1slt	1.2	1.2	0.5	0.6
1snc	2.3	2.6	1.9	3.0
1srj	0.6	0.7	0.6	0.2
1tdb	8.6	8.8	1.5	1.9
1tlp	2.6	5.1	1.9	3.4
1tmn	1.9	8.4	2.8	1.7
1tng	0.4	0.2	0.2	0.2
1tnh	0.3	0.6	0.3	0.2
1tni	2.9	2.6	2.2	2.2
1tnl	2.4	2.1	0.2	0.2
1tpp	3.1	3.1	1.1	1.5
1trk	1.2	1.8	1.6	1.5
1tyl	1.8	1.1	1.1	1.1
1ukz	1.8	3.5	0.4	0.8
1ulb	0.5	0.4	0.3	0.3
1wap	0.8	0.3	0.1	0.2
1xid	4.1	5.2	4.3	5.0
1xie	4.3	4.5	3.9	0.5
2ack	3.6	4.1	1.0	0.9
2ada	2.9	0.5	0.5	0.2
2ak3	4.0	4.2	0.7	0.8
2cht	4.1	0.6	0.4	0.6
2cmd	2.3	3.2	0.7	1.3
2cpp	2.8	2.1	0.2	0.2
2ctc	0.9	0.6	1.6	0.2
2dbl	1.2	1.0	0.7	1.1
2gbp	3.2	2.8	0.2	0.2
2ifb	1.6	0.9	1.4	0.9
2lgs	5.9	2.1	7.6	4.7
2mcp	3.6	4.0	1.3	1.3
2phh	0.7	1.8	0.4	0.2
2pk4	2.0	1.4	0.9	0.9

2r07	1.7	11.3	0.5	1.1
2tmn	3.3	3.7	0.6	0.7
2yhx	7.9	7.4	3.8	2.3
2ypi	5.1	1.0	0.3	1.1
3cla	8.4	1.5	8.5	6.8
3cpa	2.4	1.3	2.4	0.5
3hvt	0.7	4.4	0.8	0.8
3tpi	1.4	1.0	0.5	0.3
4aah	0.7	0.7	0.3	0.6
4cts	1.0	0.3	0.2	0.1
4dfr	1.3	1.3	1.1	1.0
4fbp	2.4	2.9	0.6	0.3
4phv	2.3	1.3	0.4	0.4
5abp	1.3	0.4	0.2	0.2
5cpp	2.2	1.7	0.6	0.8
6rnt	7.4	8.0	2.2	2.6
7tim	2.4	1.2	0.1	0.8
% ≤2Å	**40.7**	**44.4**	**72.6**	**75.6**

[a] Rmsd values (in Å) given for docking with GLIDE Script version 56107. Protein and ligand were separated and prepared independently: The protein was prepared with PrepWizard, and the ligand was prepared with LigPrep from Maestro version 9.2.112. Proteins were prepared with the PrepWizard using full setup including H-bond assignment and minimization.

[b] Rmsd values (in Å) given for docking with GLIDE Script version 56107. Protein and ligand were separated and prepared independently: The protein was prepared with PrepWizard, and the ligand was prepared with LigPrep from Maestro version 9.2.112. Proteins were prepared with the PrepWizard by just adding hydrogens.

[c] Docking results taken from Friesner *et al.*[5] Rmsd values in Å.

[d] Rmsd values (in Å) given for docking with GLIDE Script version 56107. Protein and ligand were prepared while being in the complex with PrepWizard from Maestro version 9.2.112. Proteins were prepared with the PrepWizard using full setup including H-bond assignment and minimization.

Table S4: Properties of ligands that bind to protein-protein interfaces.

Target protein	PDB ID	Resolution [a]	Ligand torsions[b]	% Lig. in solvent[c]	MW [g/mol] [d]	PSA/100 [Å²][e]	Binding constant [μM][f]	BEI[i]	SEI[i]
BcL-2	4AQ3	2.4	15	50.1	865	1.22	0.037 [h]	8.59	6.11
BcL-2	2O22	-	10	9.8	596	1.24	0.067 [g]	12.04	5.77
BcL-2	2O2F	-	13	25.0	688	1.34	n.d.	n.d.	n.d.
BcL-2	1YSW	-	13	22.9	694	1.34	n.d.	n.d.	n.d.
BcL-2 / BcL-XL chimera	2W3L	2.10	7	38.1	577	0.86	0.1 [h]	12.13	8.13
BcL-XL	2O2M	-	13	27.1	688	1.34	n.d.	n.d.	n.d.
BcL-XL	2O2N	-	13	42.3	731	1.28	n.d.	n.d.	n.d.
BcL-XL	3INQ	2.0	17	29.9	780	1.33	n.d.	n.d.	n.d.
BcL-XL	3QKD	2.0	16	32.8	839	1.42	0.009 [h]	9.60	5.67
BcL-XL	1YSG	-	2	18.2	215	0.40	300	16.37	8,78
BcL-XL	1YSI	-	10	29.0	551	1.21	0.036 [g]	13.52	6.15
BcL-XL	1YSN	-	13	6.3	694	1.34	n.d.	n.d.	n.d.
BcL-XL	2YXJ	2.20	16	44.6	814	1.33	0.0005 [g]	11.42	6.98
Calmodulin	1A29	2.74	5	28.6	409	0.11	1	14.69	54.95
Calmodulin	1CTR	2.45	5	46.4	403	0.10	1	14.87	61.73
Calmodulin	1LIN	2.00	5	39.3	409	0.11	1	14.69	54.95
Calmodulin	1MUX	-	9	27.3	342	0.74	31 [h]	13.19	6.11
Calpain	1ALW	2.03	3	15.4	307	0.40	0.3 [g]	21.24	16.25
Calpain	1NX0	2.30	17	31.3	459	1.89	n.d.	n.d.	n.d.
DIAP1-BIR1	1SE0	1.75	22	51.8	779	2.31	0.12	8.88	3.00
DIAP1-BIR1	1SDZ	1.78	21	50.9	765	2.31	0.27	8.59	2.85
DIAP1-BIR2	1JD5	1.90	26	48.4	893	3.00	0.24	7.41	2.21
DIAP1-BIR2	1JD6	2.70	25	53.0	919	2.91	0.04	8.05	2.54
DIAP1-BIR2	1Q4Q	2.10	22	36.4	772	2.63	n.d.	n.d.	n.d.
Grb2-SH2	1BM2	2.10	9	56.1	839	3.10	0.07 [h]	8.53	2.30
Grb2-SH2	1JYR	1.55	32	60.8	1069	5.14	0.018	7.24	1.51
Grb2-SH2	1ZFP	1.80	26	40.0	861	4.13	0.026 [h]	8.81	1.84
HexA	2GK1	3.25	4	14.3	222	0.87	n.d.	n.d.	n.d.
hGST P1-1	22GS	1.90	3	16.7	194	0.70	3.7	27.97	7.80
HIV Integrase	3LPT	2.00	3	54.5	313	0.73	12.2 [h]	15.71	6.73
HPV11-E2	1R6N	2.40	5	39.0	607	1.38	0.04	12.18	5.35
Human Cardiac Troponin	1DTL	2.15	10	22.2	368	0.17	n.d.	n.d.	n.d.
Human Cardiac Troponin	1IH0	-	4	20.0	426	0.80	10	11.75	6.23
Human Cardiac Troponin	2KDH	-	12	42.4	458	1.97	1100	6.45	1.50
Human Cardiac Troponin	2KFX	-	9	31.8	342	0.74	100	11.70	5.42
IL2	1M48	1.95	10	36.4	448	1.10	8.2	11.36	4.61
IL2	1M49	2.00	12	31.7	558	1.83	n.d.	n.d.	n.d.

IL2	1PW6	2.60	5	38.9	535	1.42	6 [h]	9.75	3.68
IL2	1PY2	2.80	12	31.1	671	1.87	0.06 [h]	10.77	3.86
IL2	1QVN	2.70	12	38.3	687	1.69	n.d.	n.d.	n.d.
INTEGRIN Alpha (lib)	2VC2	3.10	10	29.4	523	1.45	0.00007	19.43	7.01
INTEGRIN Alpha (lib)	2VDM	2.90	14	30.0	441	1.12	0.0019	19.79	7.78
INTEGRIN Alpha (lib)	2VDN	2.90	10	49.1	832	3.28	n.d.	n.d.	n.d.
INTEGRIN Alpha (lib)	2VDO	2.51	32	51.2	1189	5.56	n.d.	n.d.	n.d.
INTEGRIN Alpha (lib)	2VDP	2.80	27	40.4	745	3.82	n.d.	n.d.	n.d.
INTEGRIN Alpha (lib)	2VDR	2.40	29	51.9	774	4.17	n.d.	n.d.	n.d.
MDM2	2AXI	1.40	18	65.7	1379	4.25	0.14 [h]	4.97	1.61
MDM2	1T4E	2.60	4	43.8	580	0.90	0.08	12.23	7.93
MDM2	2GV2	1.80	32	54.2	1207	3.92	0.005 [h]	6.88	2.12
MDM2	1RV1	2.30	9	51.2	690	0.83	0.14 [h]	9.93	8.23
MDM2	1TTV	-	6	32.3	458	0.47	0.16 [h]	14.83	14.43
MDM4	3FE7	1.35	32	65.9	1172	3.95	0.075	6.08	1.80
MDM4	3FEA	1.33	32	57.8	1207	3.95	0.036	6.17	1.88
ML-IAP-BIR	1OY7	2.70	13	25.0	400	1.72	4.4 [g]	13.38	3.11
ML-IAP-BIR	1OXN	2.20	13	23.1	372	1.72	2.1 [g]	15.25	3.30
ML-IAP-BIR	1OXQ	2.30	10	25.9	384	1.23	0.5 [g]	16.43	5.11
ML-IAP-BIR	3F7G	2.30	10	41.2	466	1.06	0.043 [g]	15.82	6.94
ML-IAP-BIR / XIAP-BIR3	2I3H	1.62	10	20.6	472	1.62	0.03 [g]	15.95	4.64
ML-IAP-BIR / XIAP-BIR3	2I3I	2.30	6	25.7	498	1.13	0.05 [g]	14.67	6.46
ML-IAP-BIR / XIAP-BIR3	1TW6	1.71	10	25.9	384	1.23	0.35 [g]	16.83	5.24
ML-IAP-BIR / XIAP-BIR3	3F7H	1.80	9	36.8	520	0.95	0.08 [g]	13.66	7.46
ML-IAP-BIR / XIAP-BIR3	3F7I	1.90	7	20.0	480	0.95	0.046 [g]	15.30	7.71
TNF-alpha	2AZ5	2.10	9	45.0	549	0.43	22 [h]	8.49	10.88
TNFRc1	1FT4	2.90	6	71.0	458	0.85	0.27 [h]	14.36	7.75
XIAP-BIR3	1TFQ	-	7	34.4	444	0.95	0.012	17.86	8.33
XIAP-BIR3	1TFT	-	9	33.3	536	1.04	0.005	15.50	7.95
XIAP-BIR3	2JK7	2.82	8	38.9	488	0.95	0.067 [g]	14.71	7.54
XIAP-BIR3	3CM2	2.50	10	50.0	494	1.34	0.34 [g]	13.10	4.83
XIAP-BIR3	3CM7	3.10	10	22.2	494	1.26	0.25 [g]	13.37	5.22
XIAP-BIR3	3CLX	2.70	10	41.7	494	1.26	0.25 [g]	13.37	5.22
XIAP-BIR3	2OPY	2.80	8	28.1	439	1.59	30	10.29	2.84
XIAP-BIR3	2OPZ	3.00	10	32.3	433	1.46	0.5	14.57	4.31
XIAP-BIR3	3G76	3.00	27	38.2	975	2.38	0.23 [h]	6.81	2.79
XIAP-BIR3	3EYL	3.00	11	43.2	508	1.34	0.22 [g]	13.11	4.98
XIAP-BIR3	3HL5	1.80	8	51.4	508	1.21	0.034 [g]	14.71	6.18
XIAP-BIR3	2VSL	2.10	14	30.3	462	1.40	0.004 [g]	18.19	6.01
ZipA	1S1J	2.18	0	44.4	241	0.37	4.9 [g]	22.00	14.24
ZipA	1S1S	2.10	6	32.1	405	0.73	4.9 [g]	13.13	7.31

ZipA	1Y2F	2.00	6	36.7	425	0.84	12	11.58	5.83
ZipA	1Y2G	1.90	3	26.9	344	0.66	83	11.85	6.19

[a] Resolution of the crystal structure. In Å. For NMR structures, no value is given.

[b] The ligand's number of torsions as indicated by AutoDock.

[c] Percentage of the ligand exposed to solvent, i.e., where there is no contact of a ligand atom to the protein interface. A ligand atom was defined to be not in the interface when there is no protein atom within a distance of 4 Å.

[d] Molecular weight

[e] Polar surface area

[f] Protein-ligand binding constants (K^d, K_i or IC_{50}); "n.d." indicates that no binding constant was found.

[g] K_i

[h] IC_{50}

[i] Binding efficiency index; BEI = (pK^d or pK_i or pIC_{50})/MW; "n.d." indicates that no value for BEI could be determined.

[j] Surface efficiency index; SEI = (pK^d or pK_i or pIC_{50})/PSA; "n.d." indicates that no value for SEI could be determined.

Table S5: Properties of protein-protein interfaces.

Target protein	PDB ID	No. of arom. res. [a]	No. of charged res.[b]	No. of polar res. [c]	No. of hydro. res.[d]	No. of salt-briges[e]	Pocket volume Å²[f]	Hotspots [g]
BcL-2	4AQ3	5	3	0	5	0	827.1	n.d.
BcL-2	2O22	11	5	4	19	0	239.42	
BcL-2	2O2F	9	8	3	9	0	257.52	
BcL-2	1YSW	8	8	5	13	0	350.74	
BcL-2 / BcL-XL chimera	2W3L	6	8	1	9	0	186.52	n.d.
BcL-XL	2O2M	8	6	3	13	0	205.12	
BcL-XL	2O2N	10	9	5	13	0	342.76	TYR101
BcL-XL	3INQ	5	5	1	4	1	993.3	GLN125 VAL126
BcL-XL	3QKD	7	4	2	6	1	1042.9	LEU130
BcL-XL	1YSG	5	7	6	10	0	291.24	ASN136
BcL-XL	1YSI	8	8	5	17	0	471.22	ARG 139
BcL-XL	1YSN	8	5	6	12	0	285.06	TYR195
BcL-XL	2YXJ	10	9	9	14	0	325.96	
Calmodulin	1A29	5	17	7	29	0	1207.46	None
Calmodulin	1CTR	2	7	1	11	0	144.26	
Calmodulin	1LIN	5	18	5	29	0	1169.82	
Calmodulin	1MUX	2	5	1	14	0	152.40	
Calpain	1ALW	4	3	1	8	1	83.24	
Calpain	1NX0	4	4	2	1	0	61.60	n.d.
DIAP1-BIR1	1SE0	2	5	1	2	0	35.64	
DIAP1-BIR1	1SDZ	2	5	1	2	0	32.94	ILE87 Asp94 TRP103 ASN106 ARG111 ASP135 ILE136 ASP141
DIAP1-BIR2	1JD5	4	6	4	6	0	103.66	
DIAP1-BIR2	1JD6	2	5	3	4	0	55.22	n.d.
DIAP1-BIR2	1Q4Q	2	4	4	5	0	53.58	
Grb2-SH2	1BM2	3	4	3	2	2	44.56	n.d.
Grb2-SH2	1JYR	4	5	6	4	2	83.72	
Grb2-SH2	1ZFP	3	5	3	3	1	49.30	
HexA	2GK1	7	9	1	1	0	85.36	n.d.
hGST P1-1	22GS	2	4	1	3	0	23.32	n.d.
HIV Integrase	3LPT	1	0	3	3	0	114.42	TRP131

HPV11-E2	1R6N	5	3	7	6	0	289.76	TRP446 ARG447 ARG454 TYR455 ILE461 ARG622
Human Cardiac Troponin	1DTL	7	28	11	38	0	1036.38	
Human Cardiac Troponin	1IH0	3	5	0	13	0	171.06	V160
Human Cardiac Troponin	2KDH	2	4	0	14	0	124.34	
Human Cardiac Troponin	2KFX	4	2	3	16	0	223.78	
IL2	1M48	3	5	2	5	1	56.02	ARG38
IL2	1M49	3	5	4	5	1	48.80	PHE42
IL2	1PW6	8	22	9	18	1	967.86	TYR45
IL2	1PY2	2	2	1	4	1	32.38	TYR107
IL2	1QVN	2	2	1	4	1	23.94	
INTEGRIN Alpha (lib)	2VC2	8	9	7	8	0	326.52	n.d.
INTEGRIN Alpha (lib)	2VDM	8	9	6	8	0	301.40	
INTEGRIN Alpha (lib)	2VDN	9	9	7	8	1	336.66	
INTEGRIN Alpha (lib)	2VDO	9	10	7	8	1	313.68	
INTEGRIN Alpha (lib)	2VDP	8	10	6	9	1	291.82	
INTEGRIN Alpha (lib)	2VDQ	10	14	8	10	1	323.94	
INTEGRIN Alpha (lib)	2VDR	9	10	6	5	1	129.40	
MDM2	2AXI	7	3	3	11	0	195.50	ILE57
MDM2	1T4E	7	2	4	13	0	183.66	TYR63
MDM2	2GV2	8	0	2	11	0	177.22	TYR96
MDM2	1RV1	7	0	2	9	0	101.20	
MDM2	1TTV	6	0	1	13	0	80.64	
MDM4	3FE7	5	1	3	10	0	134.04	ILE60
MDM4	3FEA	5	1	2	14	0	141.14	TYR66 GLN71 VAL92 TYR99
ML-IAP-BIR	1OY7	2	3	2	1	0	24.62	n.d.
ML-IAP-BIR	1OXN	2	3	2	2	0	26.12	
ML-IAP-BIR	1OXQ	2	2	2	1	0	17.20	
ML-IAP-BIR	3F7G	3	5	3	3	0	56.50	
ML-IAP-BIR / XIAP-BIR3	2I3H	2	3	2	1	1	10.02	
ML-IAP-BIR / XIAP-BIR3	2I3I	2	3	2	1	0	12.32	n.d.
ML-IAP-BIR / XIAP-BIR3	1TW6	2	3	2	1	0	23.80	
ML-IAP-BIR / XIAP-BIR3	3F7H	2	4	2	1	0	21.76	
ML-IAP-BIR / XIAP-BIR3	3F7I	2	2	2	1	0	21.80	

TNF-alpha	2AZ5	1	0	2	3	0	1134.60	TYR59 ILE97 ARG103 LYS112 TYR115 TYR119 VAL123 ILE155
TNFRc1	1FT4	6	0	4	6	0	33.02	None
XIAP-BIR3	1TFQ	2	5	2	4	0	25.14	
XIAP-BIR3	1TFT	1	3	2	1	0	12.56	LEU307 GLU319
XIAP-BIR3	2JK7	2	5	2	4	0	43.22	TRP323
XIAP-BIR3	3CM2	2	3	2	1	0	46.00	TYR324
XIAP-BIR3	3CM7	2	5	2	4	0	42.80	ASN340
XIAP-BIR3	3CLX	6	9	7	7	0	426.88	LEU344
XIAP-BIR3	2OPY	6	7	7	7	0	162.44	
XIAP-BIR3	2OPZ	3	6	2	4	0	53.82	
XIAP-BIR3	3G76	13	17	8	24	0	754.94	
XIAP-BIR3	3EYL	2	3	2	1	2	21.14	
XIAP-BIR3	3HL5	2	5	2	4	0	40.34	
XIAP-BIR3	2VSL	2	3	2	1	0	17.56	
ZipA	1S1J	2	1	1	2	0	11.72	
ZipA	1S1S	1	0	2	7	0	30.68	
ZipA	1Y2F	1	3	4	8	0	89.76	VAL65
ZipA	1Y2G	1	2	2	3	0	15.60	LYS66

[a] Number of *aromatic* residues in the protein binding interface. The interface is defined as all residues within 5 Å distance around the ligand.

[b] Number of *charged* residues in the protein binding interface. The interface is defined as all residues within 5 Å distance around the ligand.

[c] Number of *polar* residues in the protein binding interface. The interface is defined as all residues within 5 Å distance around the ligand.

[d] Number of *hydrophobic* residues in the protein binding interface. The interface is defined as all residues within 5 Å distance around the ligand.

[e] Number of salt bridges between protein and ligand.

[f] Pockets were determined with PocketAnalyzer[47], corresponding volumes were calculated with McVol[48].

[g] Hotspots were determined with DrugScore^{PPI} [49]. A residue was considered to be a *medium* hotspot if $\Delta\Delta G_{calc} \geq 1.0$ kcal/mol and < 1.5 kcal/mol and to be a *strong* hotspot if $\Delta\Delta G_{calc} \geq$

1.5 kcal/mol. Residue identifiers for *strong* hotspots are underlined. “n.d.” indicates that no protein-protein complex structure is available. “None” indicates that no hotspot could be determined.

Figures

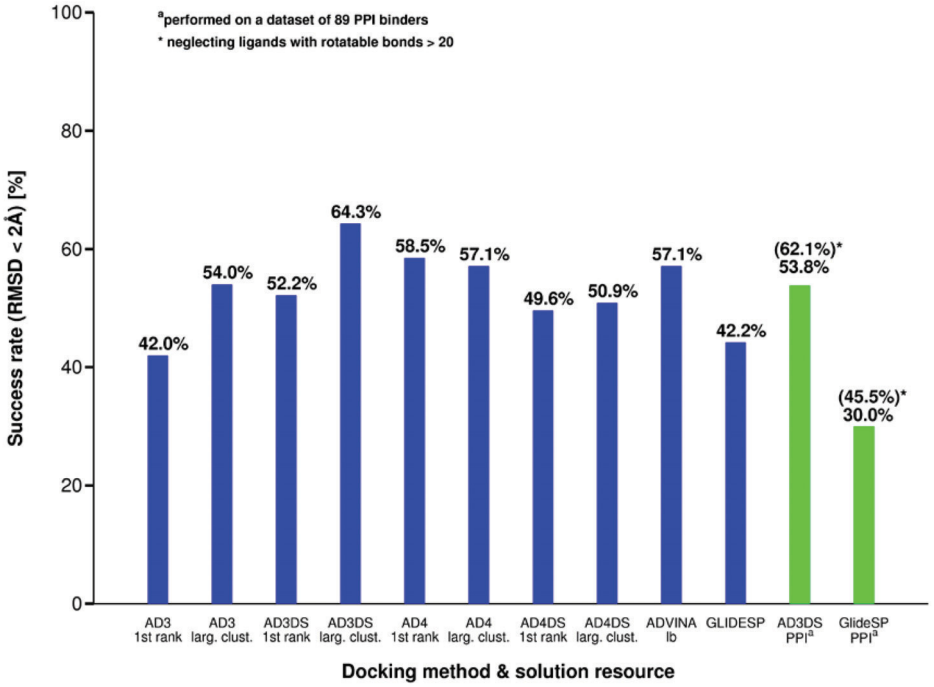


Figure S1: Success rates for re-docking on the CCDC/Astex clean set[45] with three versions of AutoDock, two DrugScore-adapted AutoDock versions, and GlideSP are depicted in blue. When "1st" rank is stated, the solution from the first rank with the lowest energy was considered as the result. When "larg. clust." is stated, the solution from the the first rank with the lowest energy from the largest cluster was considered as the result. In addition, success rates for re-docking on the PPIM dataset are depicted in green for DrugScore-adapted AutoDock3 and GlideSP.

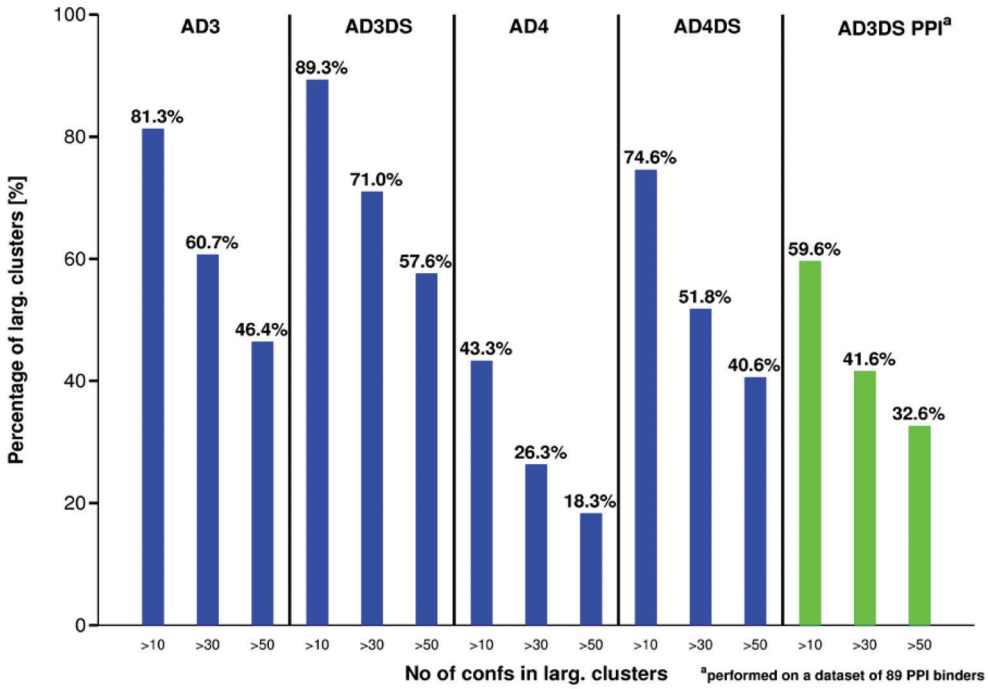


Figure S2: Percentage of largest clusters for a given minimal number of ligand conformations. For example, DrugScore-adapted AutoDock3 generated largest clusters with 89.3% of them containing more than 10 conformations, 71.0% containing more than 30 conformations, and 57.6% containing more than 50 conformations. Results are shown for docking with AutoDock version 3 and 4 to the CCDC/Astex clean set[45] and with DrugScore-adapted AutoDock3 to the PPIM dataset.

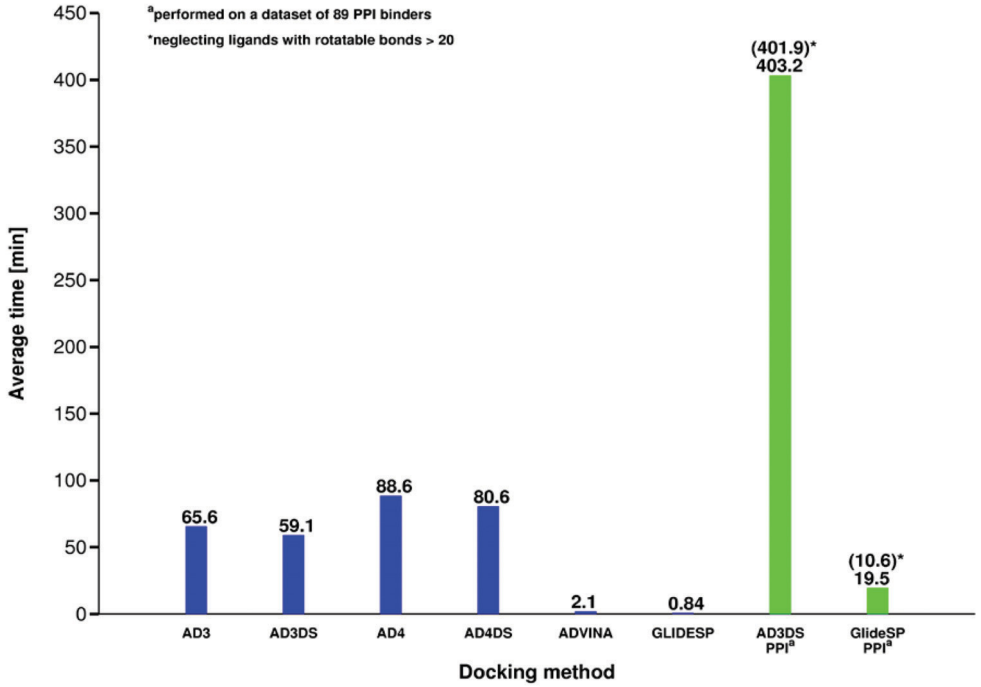
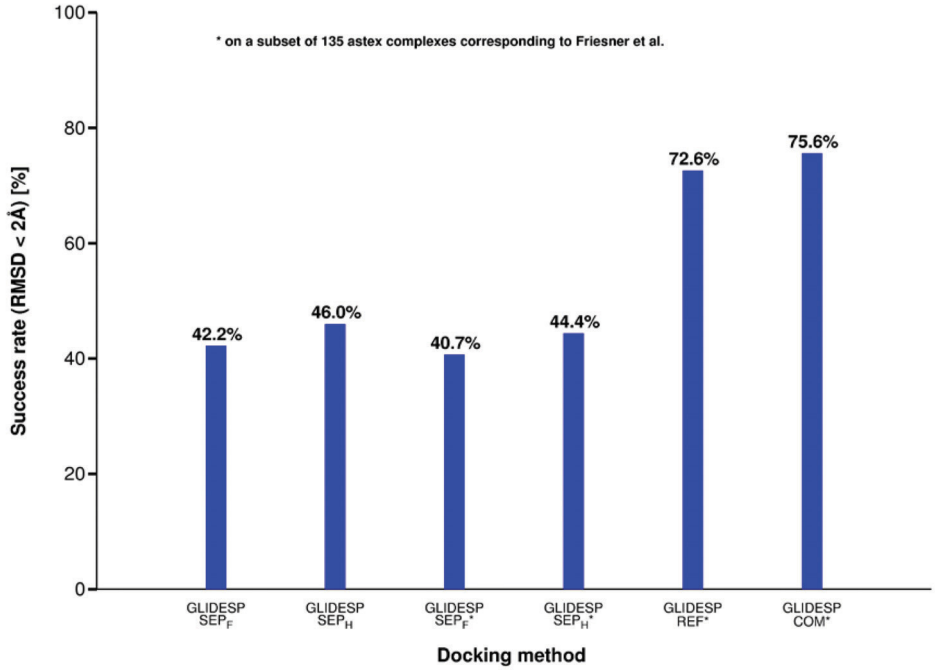


Figure S3: Average runtimes per compound for re-docking on the CCDC/Astex clean set[45] are depicted in blue for three versions of AutoDock, two DrugScore-adapted AutoDock versions, and GlideSP. The same is shown for re-docking on the PPIM dataset, depicted in green for DrugScore-adapted AutoDock3 and GlideSP. In the case of AutoDock, average runtimes were calculated over 100 docking simulations each. In the case of Glide, the average runtime is reported per compound per docking job.



ure S4: Success rates for re-docking on a subset of 135 protein-ligand complexes from the DC/Astex clean set[45] with GlideSP based on differently prepared proteins and ligands. The set chosen agrees with the dataset from Friesner *et al.*[5] I) "SEP" indicates that the protein and and were separated and, subsequently, prepared independently from each other, i.e., the protein s prepared with PrepWizard, and the ligand was prepared with LigPrep. "$_F$" indicates that pWizard was used with full setup including H-bond assignment and minimization, whereas ' indicates that PrepWizard was only used to add hydrogens. II) "COM" indicates that the tein-ligand complex as a whole was prepared with PrepWizard using full setup including H- d assignment and minimization. For comparison the results from Friesner *et al.*[5] are denoted "REF". The success rates for the whole CCDC/Astex clean set[45] are given as the first two bars.

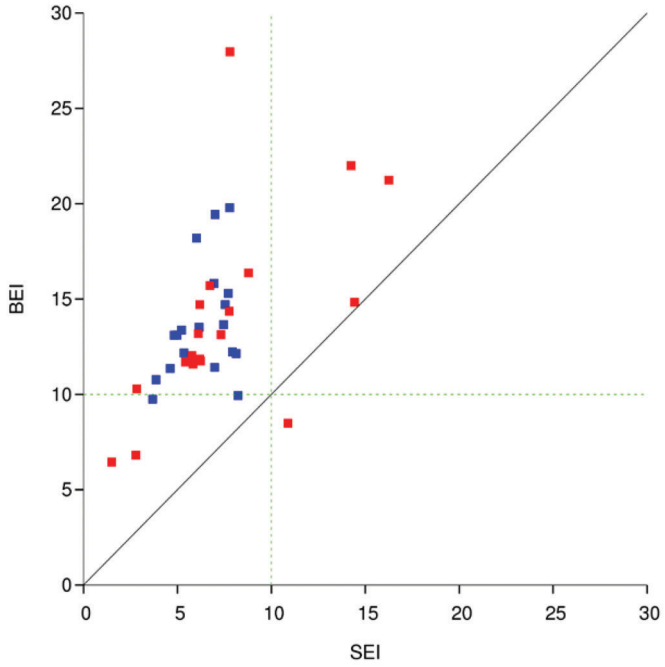


Figure S5: Mapping of binding (BEI) and surface (SEI) efficiency indexes for the 54 PPIM from the PPIM dataset. Docking successes obtained with Autodock3 and DrugScore are depicted in blue, failures in red. The dashed green lines mark the lower (BEI) and upper (SEI) limits of the efficiency indices within which most of the PPIMs (61.3%) were docked successfully.

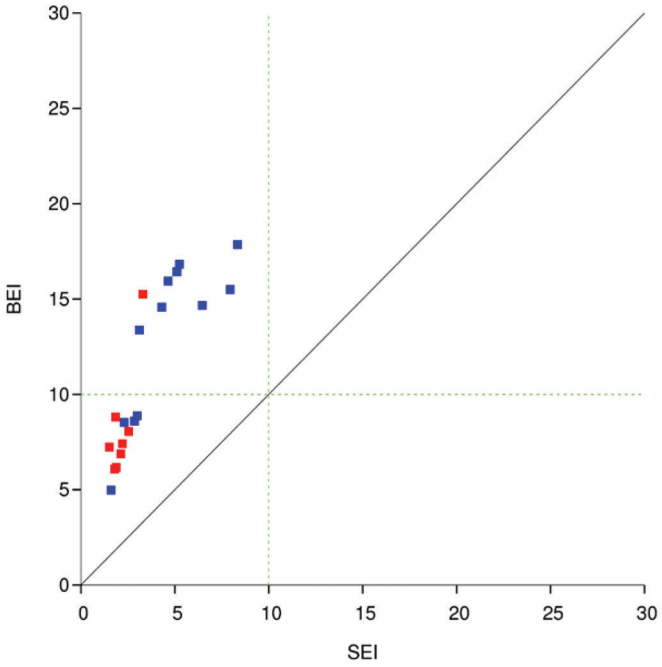


Figure S6: Mapping of binding and surface efficiency indexes for the 26 (modified) peptide ligands from the PPIM dataset. Hits achieved with Autodock3 and DrugScore are depicted in blue, failures in red. The dashed green lines mark the lower (BEI) and upper (SEI) limits of the efficiency indices within which most of the PPIMs (90.9%) were docked successfully.

References

1. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M., The Protein Data Bank: a computer-based archival file for macromolecular structures. *Arch Biochem Biophys* **1978,** 185, (2), 584-91.
2. http://autodock.scripps.edu/resources/adt, In.
3. *Maestro, version 9.1, Schrödinger, LLC, New York, NY, 2009.*
4. LigPrep, version 38108 , Schrödinger, LLC, New York, NY, 2009.
5. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** 47, (7), 1739-49.
6. Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J, Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. *J. Computational Chemistry* **1998,** 19, 1639-1662.
7. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S., A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* **2007,** 28, (6), 1145-52.
8. http://vina.scripps.edu/manual.html
9. Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **2000,** 295, (2), 337-356.
10. Shortle, D.; Simons, K. T.; Baker, D., Clustering of low-energy conformations near the native structures of small proteins. *Proc Natl Acad Sci U S A* **1998,** 95, (19), 11158-62.
11. Zhang, Y.; Skolnick, J., SPICKER: a clustering approach to identify near-native protein folds. *J Comput Chem* **2004,** 25, (6), 865-71.
12. Wendel, C.; Gohlke, H., Predicting transmembrane helix pair configurations with knowledge-based distance-dependent pair potentials. *Proteins* **2008,** 70, (3), 984-999.
13. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31, (2), 455-61.
14. Almerico, A. M.; Tutone, M.; Lauria, A., In-silico screening of new potential Bcl-2/Bcl-xl inhibitors as apoptosis modulators. *J Mol Model* **2009,** 15, (4), 349-55.
15. Morelli, X.; Bourgeas, R.; Roche, P., Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol* 15, (4), 475-81.
16. Fuller, J. C.; Burgoyne, N. J.; Jackson, R. M., Predicting druggable binding sites at the protein-protein interface. *Drug Discov Today* **2009,** 14, (3-4), 155-61.
17. Clapperton, J. A.; Martin, S. R.; Smerdon, S. J.; Gamblin, S. J.; Bayley, P. M., Structure of the complex of calmodulin with the target sequence of calmodulin-dependent protein kinase I: studies of the kinase activation mechanism. *Biochemistry* **2002,** 41, (50), 14669-79.
18. Kovacs, E.; Liliom, K., Sphingosylphosphorylcholine as a novel calmodulin inhibitor. *Biochem J* **2008,** 410, (2), 427-37.
19. Stuart, B. G.; Coxon, J. M.; Morton, J. D.; Abell, A. D.; McDonald, D. Q.; Aitken, S. G.; Jones, M. A.; Bickerstaffe, R., Molecular modeling: a search for a calpain inhibitor as a new treatment for cataractogenesis. *J Med Chem* 54, (21), 7503-22.
20. Kiss, R.; Kovacs, D.; Tompa, P.; Perczel, A., Local structural preferences of calpastatin, the intrinsically unstructured protein inhibitor of calpain. *Biochemistry* **2008,** 47, (26), 6936-45.
21. Li, X.; Wang, J.; Shi, Y., Structural mechanisms of DIAP1 auto-inhibition and DIAP1-mediated inhibition of drICE. *Nat Commun* 2, 408.

22. Wu, J. W.; Cocina, A. E.; Chai, J.; Hay, B. A.; Shi, Y., Structural analysis of a functional DIAP1 fragment bound to grim and hid peptides. *Mol Cell* **2001,** 8, (1), 95-104.
23. Clements, J. H.; DeLorbe, J. E.; Benfield, A. P.; Martin, S. F., Binding of flexible and constrained ligands to the Grb2 SH2 domain: structural effects of ligand preorganization. *Acta Crystallogr D Biol Crystallogr* 66, (Pt 10), 1101-15.
24. Chen, H. H.; Chen, C. W.; Chang, Y. Y.; Shen, T. L.; Hsu, C. H., Preliminary crystallographic characterization of the Grb2 SH2 domain in complex with a FAK-derived phosphotyrosyl peptide. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 66, (Pt 2), 195-7.
25. Ogura, K.; Shiga, T.; Yokochi, M.; Yuzawa, S.; Burke, T. R., Jr.; Inagaki, F., Solution structure of the Grb2 SH2 domain complexed with a high-affinity inhibitor. *J Biomol NMR* **2008,** 42, (3), 197-207.
26. Lemieux, M. J.; Mark, B. L.; Cherney, M. M.; Withers, S. G.; Mahuran, D. J.; James, M. N., Crystallographic structure of human beta-hexosaminidase A: interpretation of Tay-Sachs mutations and loss of GM2 ganglioside hydrolysis. *J Mol Biol* **2006,** 359, (4), 913-29.
27. Prade, L.; Huber, R.; Manoharan, T. H.; Fahl, W. E.; Reuter, W., Structures of class pi glutathione S-transferase from human placenta in complex with substrate, transition-state analogue and inhibitor. *Structure* **1997,** 5, (10), 1287-95.
28. Christ, F.; Voet, A.; Marchand, A.; Nicolet, S.; Desimmie, B. A.; Marchand, D.; Bardiot, D.; Van der Veken, N. J.; Van Remoortel, B.; Strelkov, S. V.; De Maeyer, M.; Chaltin, P.; Debyser, Z., Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat Chem Biol* 6, (6), 442-8.
29. Lindhout, D. A.; Sykes, B. D., Structure and dynamics of the C-domain of human cardiac troponin C in complex with the inhibitory region of human cardiac troponin I. *J Biol Chem* **2003,** 278, (29), 27024-34.
30. Oleszczuk, M.; Robertson, I. M.; Li, M. X.; Sykes, B. D., Solution structure of the regulatory domain of human cardiac troponin C in complex with the switch region of cardiac troponin I and W7: the basis of W7 as an inhibitor of cardiac muscle contraction. *J Mol Cell Cardiol* 48, (5), 925-33.
31. Wang, X.; Rickert, M.; Garcia, K. C., Structure of the quaternary complex of interleukin-2 with its alpha, beta, and gammac receptors. *Science* **2005,** 310, (5751), 1159-63.
32. Blamey, C. J.; Ceccarelli, C.; Naik, U. P.; Bahnson, B. J., The crystal structure of calcium- and integrin-binding protein 1: insights into redox regulated functions. *Protein Sci* **2005,** 14, (5), 1214-21.
33. Yuan, W.; Leisner, T. M.; McFadden, A. W.; Wang, Z.; Larson, M. K.; Clark, S.; Boudignon-Proudhon, C.; Lam, S. C.; Parise, L. V., CIB1 is an endogenous inhibitor of agonist-induced integrin alphaIIbbeta3 activation. *J Cell Biol* **2006,** 172, (2), 169-75.
34. Kubota, D.; Ishikawa, M.; Yamamoto, M.; Murakami, S.; Hachisu, M.; Katano, K.; Ajito, K., Tricyclic pharmacophore-based molecules as novel integrin alpha(v)beta3 antagonists. Part 1: design and synthesis of a lead compound exhibiting alpha(v)beta3/alpha(IIb)beta3 dual antagonistic activity. *Bioorg Med Chem* **2006,** 14, (7), 2089-108.
35. Huang, L.; Yan, Z.; Liao, X.; Li, Y.; Yang, J.; Wang, Z. G.; Zuo, Y.; Kawai, H.; Shadfan, M.; Ganapathy, S.; Yuan, Z. M., The p53 inhibitors MDM2/MDMX complex is required for control of p53 activity in vivo. *Proc Natl Acad Sci U S A* 108, (29), 12001-6.
36. Kallen, J.; Goepfert, A.; Blechschmidt, A.; Izaac, A.; Geiser, M.; Tavares, G.; Ramage, P.; Furet, P.; Masuya, K.; Lisztwan, J., Crystal Structures of Human MdmX (HdmX) in Complex with p53 Peptide Analogues Reveal Surprising Conformational Changes. *J Biol Chem* **2009,** 284, (13), 8812-21.
37. Liu, B.; Han, M.; Wen, J. K.; Wang, L., Livin/ML-IAP as a new target for cancer treatment. *Cancer Lett* **2007,** 250, (2), 168-76.

38. Vucic, D.; Deshayes, K.; Ackerly, H.; Pisabarro, M. T.; Kadkhodayan, S.; Fairbrother, W. J.; Dixit, V. M., SMAC negatively regulates the anti-apoptotic activity of melanoma inhibitor of apoptosis (ML-IAP). *J Biol Chem* **2002,** 277, (14), 12275-9.
39. Zobel, K.; Wang, L.; Varfolomeev, E.; Franklin, M. C.; Elliott, L. O.; Wallweber, H. J.; Okawa, D. C.; Flygare, J. A.; Vucic, D.; Fairbrother, W. J.; Deshayes, K., Design, synthesis, and biological activity of a potent Smac mimetic that sensitizes cancer cells to apoptosis by antagonizing IAPs. *ACS Chem Biol* **2006,** 1, (8), 525-33.
40. He, M. M.; Smith, A. S.; Oslob, J. D.; Flanagan, W. M.; Braisted, A. C.; Whitty, A.; Cancilla, M. T.; Wang, J.; Lugovskoy, A. A.; Yoburn, J. C.; Fung, A. D.; Farrington, G.; Eldredge, J. K.; Day, E. S.; Cruz, L. A.; Cachero, T. G.; Miller, S. K.; Friedman, J. E.; Choong, I. C.; Cunningham, B. C., Small-molecule inhibition of TNF-alpha. *Science* **2005,** 310, (5750), 1022-5.
41. Carter, P. H.; Scherle, P. A.; Muckelbauer, J. K.; Voss, M. E.; Liu, R. Q.; Thompson, L. A.; Tebben, A. J.; Solomon, K. A.; Lo, Y. C.; Li, Z.; Strzemienski, P.; Yang, G.; Falahatpisheh, N.; Xu, M.; Wu, Z.; Farrow, N. A.; Ramnarayan, K.; Wang, J.; Rideout, D.; Yalamoori, V.; Domaille, P.; Underwood, D. J.; Trzaskos, J. M.; Friedman, S. M.; Newton, R. C.; Decicco, C. P., Photochemically enhanced binding of small molecules to the tumor necrosis factor receptor-1 inhibits the binding of TNF-alpha. *Proc Natl Acad Sci U S A* **2001,** 98, (21), 11879-84.
42. Crisostomo, F. R.; Feng, Y.; Zhu, X.; Welsh, K.; An, J.; Reed, J. C.; Huang, Z., Design and synthesis of a simplified inhibitor for XIAP-BIR3 domain. *Bioorg Med Chem Lett* **2009,** 19, (22), 6413-8.
43. Liu, Z.; Sun, C.; Olejniczak, E. T.; Meadows, R. P.; Betz, S. F.; Oost, T.; Herrmann, J.; Wu, J. C.; Fesik, S. W., Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature* **2000,** 408, (6815), 1004-8.
44. Tsao, D. H.; Sutherland, A. G.; Jennings, L. D.; Li, Y.; Rush, T. S., 3rd; Alvarez, J. C.; Ding, W.; Dushin, E. G.; Dushin, R. G.; Haney, S. A.; Kenny, C. H.; Malakian, A. K.; Nilakantan, R.; Mosyak, L., Discovery of novel inhibitors of the ZipA/FtsZ complex by NMR fragment screening coupled with structure-based design. *Bioorg Med Chem* **2006,** 14, (23), 7953-61.
45. Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R., A new test set for validating predictions of protein-ligand interaction. *Proteins* **2002,** 49, (4), 457-71.
46. Neudert, G.; Klebe, G., DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. *J Chem Inf Model* 51, (10), 2731-45.
47. 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* 51, (10), 2666-79.
48. Till, M. S.; Ullmann, G. M., McVol - a program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm. *J Mol Model* 16, (3), 419-29.
49. Kruger, D. M.; Gohlke, H., DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. *Nucleic Acids Res* 38, (Web Server issue), W480-6.

Eidesstattliche Erklärung

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. Ferner versichere ich, dass die Arbeit in gleicher oder ähnlicher Form bisher keiner anderen Fakultät vorgelegt wurde.

Düsseldorf, Mai 2014

Dennis M. Krüger

Non est ad astra mollis e terris via.
Lucius Annaeus Seneca, Hercules Furens, 437

Finis coronat opus.

Ovid, Heroides 2, 85