HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Assay development and optimization for the validation and quantification of NHR2 tetramerization inhibitors and prediction of the binding mode

Inaugural-Dissertation

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

vorgelegt von
Tobias Kröger
aus Osnabrück

Düsseldorf, Dezember 18

Aus dem Institut für Pharmazeutische und Medizinische Chemie

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Holger Gohlke

2. Prof. Dr. Georg Groth

Tag der mündlichen Prüfung: 31.10.2018

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.

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

Düsseldorf, den 05.12.2018

Tobias Kröger

Somewhere, something incredible is waiting to be known

Carl Sagan

In loving memory of my father

Table of content

List of publications

This thesis is based on the publications **I** and **II** (Ref. [1] and [2], respectively). Texts and figures have been taken and modified from these manuscripts.

I **Kroeger, T.** (40%), Frieg, B., Zhang, T., Hansen, F.K., Marmann, A., Proksch, P., Nagel-Steger, L., Groth, G., Smits, S.H.J., Gohlke, H.
EDTA aggregates induce SYPRO Orange-based fluorescence in thermal shift assay.
PLOS ONE 2017, 12, e0177024.

My contribution to the publication: Preparation and provision of the samples for the experiments; planning and carrying out TSA experiments; analyzing the TSA results; planning and conducting the MD simulations with BF; analyzing the results from MD simulations with BF; writing of the manuscript with BF and HG; revision and editing of the manuscript with BF, FKH, PP, LNS, GG, SHJS and HG.

II **Kroeger, T.** (40%), Frieg, B., Viegas, A., Smits, S., Groth, G., Etzkorn, M., Gohlke, H.
Biophysical characterization of a small-molecule inhibitor of RUNX1/ETO tetramerization with anti-leukemic effects
in preparation 2018.

My contribution to the publication: Expression and purification of NHR2; preparation and provision of the samples for experiments; planning and carrying out the TSA experiments; analyzing the TSA results; planning and carrying out the MST experiments; analyzing the MST results; writing of the manuscript with BF and HG.

Contribution to further publications during the project (not used in the thesis):

Bhatia, S., Diedrich, D., Frieg, B., Ahlert , H., Stein, S., Bopp, B., Lang, F., Zang, T., **Kroeger, T.**, Ernst, T., Kögler, G., Krieg, A., Lüdeke, S., Kunkel, H., Rodrigues Moita, A.J., Kassack, M.U., Marquardt, V., Opitz, F.V., Oldenburg, M., Remke, M., Babor, F., Grez, M., Hochhaus, A., Borkhardt, A., Groth, G., Nagel-Steger, L., Jose, J., Kurz, T., Gohlke, H., Hansen, F.K., Hauer, J.
Targeting HSP90 dimerization via the C terminus is effective in imatinib-resistant CML and lacks the heat shock response.
Blood 2018, 132, 3.

Abbreviations

^3J	coupling constant
Å	Ångstrom (1 Å = 10^{-10} m)
ABC	ATP-binding-cassette
AML	acute myeloid leukemia
A_{STD}	STD amplification factor
ATP	adenosine triphosphate
AUC	analytical ultracentrifugation
c-Kit	cellular v-kit homolog
COM	center of mass
DMSO	dimethyl sulfoxide
E. coli	*Escherichia coli*
EC_{50}	half maximal effective concentration
EDTA	ethylene diamine tetraacetic acid
EGTA	ethylene glycol tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
Eq.	equation
ETO	eight-twenty one protein
FAB	French-American-British classification system for acute leukemia
FLT3	FMS-like tyrosine kinase-3 (FLT3)
g(r)	radial distribution function
Hsp90	heat shock protein 90 kDa
HSQC	Heteronuclear single quantum coherence spectroscopy
IC_{50}	half maximal inhibitory concentration
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
K_D	dissociation constant
K_{lig}	dissociation constant of NHR2/PPIM complexes
K_{OW}	octanol-water partition coefficient
K_{tet}	dissociation constant (for NHR2 tetramer)
LB	lysogeny broth
LC	liquid chromatography
MD	molecular dynamics
MST	microscale thermophoresis
NHR1-4	nervy homology region 1-4

NMR	nuclear magnetic resonance
OD_{600}	optical density (measured at 600 nm)
PBS	phosphate buffered saline
PDB	protein data bank
PPI	protein-protein interaction
PPIM	protein-protein interaction modulator
ppm	parts per million
residue	amino acid residue
RESP	restraint electrostatic potential
RHD	Runt homology domain
RMSD	root mean square deviation
rpm	rounds per minute
RT	room temperature
Runt	Runt domain of RUNX proteins
RUNX1	Runt-related transcription factor 1
RUNX1/ETO	RUNX1-ETO fusion protein caused by t(8;21)
$s_{20,w}$	sedimentation coefficient
SAR	structure-activity relationship
SASA	solvent accessible surface area
SD	standard deviation
SDS	sodium dodecyl sulfate
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	standard error of the mean
STD-NMR	saturation-transfer difference nuclear magnetic resonance
SV	sedimentation velocity
T_m	melting point
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation optimized spectroscopy
TSA	Thermal shift assay (also called: differential scanning fluorimetry)
v/v	volume per volume
w/v	weight per volume
WHO	World Health Organization

Zusammenfassung

Die akute myeloische Leukämie ist eine Unterform der Leukämie, die bei erwachsenen Patienten mehr als ein Viertel der Leukämien ausmacht. Diese Krankheit zeichnet sich durch ein schlechtes Behandlungsergebnis, niedrige Remissionsraten, ein kurzes krankheitsfreies Überleben und ein kurzes Gesamtüberleben aus. Bis heute gibt es keine wirksame Behandlung, sodass alle AML-Patienten eine schlimme Prognose haben. Das Fusionsprotein RUNX1/ETO erwies sich als attraktives Ziel für die Entwicklung einer Behandlung. Dieses Protein ist das Ergebnis einer Translokation, die die Chromosomen 8 und 21 betrifft. Es kann in ~ 12% *de novo* AML Fällen gefunden werden. Die Translokation wurde bereits 1973 entdeckt, und neue Erkenntnisse weisen darauf hin, dass es sich um das initiative Ereignis für akute myeloische Leukämie handelt. RUNX1/ETO ist als Homotetramer aktiv und enthält die α-helikale Oligomerisierungsdomäne nervy homology region 2 (NHR2), die sich für die Tetramerisierung als entscheidend erwiesen hat. Es wurden Strategien entwickelt, die darauf abzielten, kleine Moleküle zu identifizieren, die die Tetramerisierung von NHR2 inhibieren können. Mit experimentellen und computergestützten Methoden wurden die Hotspots identifiziert, Aminosäuren, die einen Großteil der freien Bindungsenergie tragen und somit essentiell für die Protein-Protein-Interaktion sind. In nachfolgenden Studien wurde der erste Inhibitor der NHR2-Tetramerisierung identifiziert: **7.44**. Studien haben gezeigt, dass **7.44** spezifisch mit NHR2 wechselwirkt und in der Lage ist, die Tetramerisierung von NHR2 im Zellkultur- und im Mausmodell zu verhindern.

Das Ziel dieser Arbeit war die Bestimmung der Stabilität des NHR2-Tetramers, der Affinität von **7.44** zu NHR2 und deren Bindemodus. Zuerst zeigte ich die Bildung von Aggregaten durch EDTA und die Wechselwirkung dieser Aggregate mit SYPRO Orange, während ich die Assays für die nachfolgenden Experimente entwickelte. MD-Simulationen lieferten ein molekulares Modell der Aggregation und der Interaktion. Diese SYPRO Orange/EDTA-Wechselwirkung hat weitreichende Konsequenzen, da EDTA und das Thermal Shift Assay weit verbreitet sind. Zweitens habe ich K_{tet} für das NHR2-Tetramer und K_{lig} für **7.44** vom NHR2-Dimer mit Mircoscale Thermophorese bestimmt. Diese Ergebnisse bezüglich der Stabilität des Tetramers und der Affinität des Liganden liefern neue Informationen, die als Referenz für potentielle Inhibitoren und die Optimierung dieser kleinen Moleküle dienen können. Drittens wurde die Bindungspose von **7.44** unter

Verwendung der STD-NMR bestimmt. Diese Experimente enthüllten die Teile von **7.44**, die bei einer Wechselwirkung NHR2 am nächsten kommen. Diese Information kann für eine effizientere Suche nach neuen potentiellen Inhibitoren der NHR2-Tetramerisierung verwendet werden. Viertens, unter Verwendung von MD-Simulationen konnte ein molekulares Bild der Interaktion zwischen **7.44** und NHR2 geliefert und eine Vorhersage der Bindestelle des Inhibitors der NHR2 Tetramerisierung erstellt werden.

Abstract

Acute myeloid leukemia is a subform of leukemia that accounts for more than a quarter of leukemias in adult patients. This disease is characterized by a poor treatment outcome, low remission rates, short disease-free survival and short overall survival. To this day, there is no efficient treatment available, leaving all patients of AML with a dire prognosis. The fusion protein RUNX1/ETO emerged as an attractive target for the development of a treatment. This protein is the result of a translocation involving chromosomes 8 and 21. It can be found in ~ 12% of *de novo* AML cases. The translocation was already discovered in 1973, and, to this day, new insights suggest that it is the initiative event for acute myeloid leukemia. RUNX1/ETO is active as a homotetramer and contains the α-helical oligomerization domain nervy homology region 2 (NHR2), which has been proven crucial for the tetramerization. Strategies arose that focused on identifying small molecules that are able to inhibit the tetramerization of NHR2. Using experimental and computational methods, hot spots, amino acids that contribute the majority of the binding free energy and thus are essential for the protein-protein interaction, were identified. In subsequent studies, the first inhibitor of NHR2 tetramerization was identified: **7.44**. Studies have shown that **7.44** specifically interacts with NHR2 and is able to prevent the tetramerization of NHR2 in cell cultures and a mouse model.

The aim of this thesis was to determine the stability of the NHR2 tetramer, the affinity of **7.44** towards NHR2 and their binding mode. First, I showed the formation of aggregates by EDTA and the interaction of these aggregates with SYPRO Orange while setting up the assays for the subsequent experiments. MD simulations provided a molecular model of the aggregation and the interaction. This SYPRO Orange/EDTA interaction has widespread consequences, since EDTA and the thermal shift assay are widely used. Second, I determined the K_{tet} for the NHR2 tetramer and the K_{lig} for **7.44** using microscale thermophoresis. These results regarding the stability of the tetramer and the affinity of the ligand provide new information that will serve as a reference point for other potential inhibitors and the optimization of these small molecules. Third, the binding pose of **7.44** was determined using STD-NMR. These experiments revealed the moieties of **7.44** that come closest to NHR2 upon interaction. This information can be used for a more efficient search for new potential inhibitors of the NHR2 tetramerization. Fourth, using MD simulations a molecular image of the interaction between **7.44** and NHR2 was provided and the binding site of the NHR2 tetramerization inhibitor was predicted.

1 Introduction

1.1 Leukemia

In the present work, I investigated a lead structure for a subtype of leukemia in order to provide a foundation for future optimization and development of an efficient drug to counteract this disease.

Leukemia is a neoplastic disease of the hematopoietic system and consequently affecting the blood system and immune system. The origin of this disease lies within the bone marrow, where genetic aberrations in affected cells lead to atypical proliferation and block in differentiation of white blood cells [3]. These cells cannot perform their main function of defense against pathogens, ultimately leading to a compromised immune system. Thereby, patients with leukemia are prone to infections. Degenerate leukocytes tend to spread rapidly and in an uncontrolled manner, displacing healthy cells in the bone marrow. Beside the effects of leukemia on the immune system the changes of the bone marrow have severe effects on the body. To highlight the effects of leukemia on the body following symptoms can arise in the course of the disease: (i) The displacement of erythrocyte precursors from the bone marrow will lead to anemia, leading to a diminishing performance and fatigue, (ii) the displacement of thrombocyte precursors lead to disorder in blood clotting and increased bleeding tendencies.

Leukemia, contrary to other forms of cancer, affects the entire body since the beginning. The degenerated cells are distributed throughout the organism through the blood stream. Leukemia can thereby be classified as a systemic disease [4].

Leukemia is divided into four subgroups based on the affected cells and the development rate: Acute myeloid leukemia (AML), chronic myeloid leukemia, acute lymphoblastic leukemia and chronic lymphoblastic leukemia. The term myeloid or lymphoid describes the affected cell type, and acute or chronic describes the progression rate of the affected cells, in which acute describes a faster and more aggressive progression of the disease. This project is focused on the AML.

1.2 Acute myeloid leukemia

AML (also referred to as acute non-lymphoblastic leukemia, acute myeloid leukemia or acute myelogenous leukemia) accounts for more than a quarter of leukemias in adult patients [5, 6]. A survey for cancer patients in Germany revealed 12,640 new patients

diagnosed with leukemia in 2012 [5]. Around 5% of the newly diagnosed patients were below 15 years old. The survey further uncovered that ~4.4% of all malignomas in children are AML, and it is the most prominent malignome at patients under 2 years [5]. The median age at diagnosis for AML is around 70 years [7, 8], showing that the risk of an AML onset increases significantly with age. The risk of AML peaks at ~12.6 per 100,000 adults 65 years of age or older [9]. In older patients, AML is characterized by a poor treatment outcome, low remission rates, short disease-free survival and short overall survival [10-16]. Untreated AML is typically fatal within days to weeks depending on the level of the blast count, the amount of abnormal immature cells (blasts), in the peripheral blood system and the presence of complications of bone marrow failures [6]. In their review of AML, Smith *et al.* describe that complete remission is achieved in only 40 to 80% of the recorded cases [6]. Of those patients undergoing the therapy, however, 10 to 20% of the patients will not survive AML. Further 10 to 20% will have a refractory disease, which means the treatment will have no effect. A significant portion of patients who achieved complete remission will have a relapse within 1 or 2 years after the therapy. For patients with a refractory disease and patients with a relapse after the previous therapy a remission is only rarely achieved. Furthermore these patients have an overall survival rate of 10% using current salvage therapies [6]. The term salvage therapy describes current medical procedures that may induce another remission in the presence of refractory AML. This leaves patients with AML with a dire prognosis.

Several factors are known to increase the likelihood of developing AML. These factors include: (i) Exposure to ionizing radiation, for example in medical procedures [17]; (ii) chemotherapy drugs like chlorambucil, cyclophosphamide, melphalan, thiotepa, treosulphan or etoposide [18]; (iii) non-ionizing electrical and magnetic fields [19, 20]; (iv) increased exposure to benzene, for example, in leather, rubber and chemical manufacturing processes can increase the risk for AML 2 to 10-fold [21]; (v) certain autosomal dominant and recessive cases of familial AML [22, 23], a rare inheritable form of AML, and (vi) genetic syndromes like trisomy 21 [24] or genetic defects characterized by impaired DNA reparation like the Bloom´s syndrome [25].

In the course of AML immature myeloid precursor, which undergo a differentiation block, in the bone marrow [6] (**Figure 1**). The differentiation block prevents the common myeloid progenitor cells from differentiating into a variety of required cell types, which are lacking in the process of AML. This differentiation block leads inevitably to an impairment of the normal function of the bone marrow. As time progresses, healthy myeloid progenitor

cells will be replaced by abnormal immature cells and, thus, these less differentiated cells will be released into the blood stream. At some point the remaining healthy bone marrow cannot compensate the loss of affected bone marrow, then, e.g., the missing erythrocytes cause systemic complications, since all kinds of tissues will not receive the needed oxygen via the blood stream. Thus, patients with AML often are diagnosed based on symptoms like persistent weakness, increased susceptibility to infections or easily invocable bruises or bleedings, caused by the lack of functional red blood cells, neutrophils and platelets.

The course of the disease and the prospect of complete remission or even cure may vary greatly depending on the form of AML (**Chapter 1.3**). This makes it even more important to study the causes and pathways of AML. However, AML can be caused by multiple different factors, and addressing them all in this project would not be possible. Hence, one needs to find a suitable starting point for investigation based on the classification of the targeted subtype of AML.

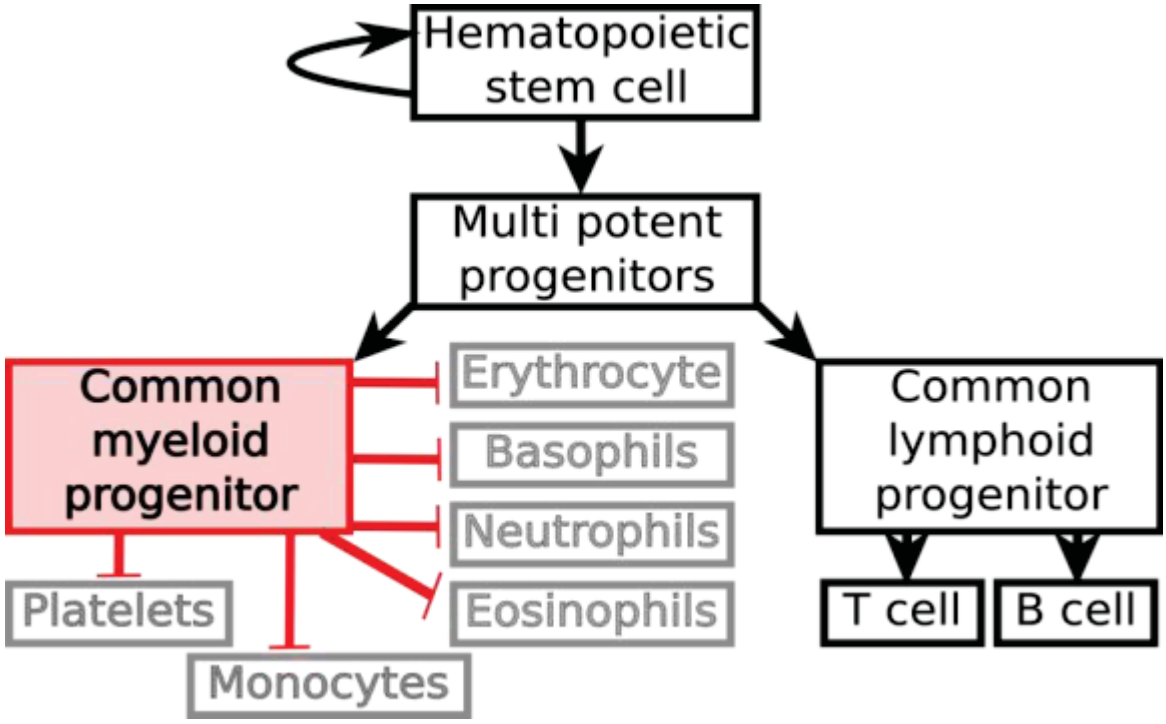

Figure 1: Schematic representation of the hematopoiesis during AML. A simplified model of the hematopoiesis considering the causes of AML. Red: block in differentiation, grey: lack of cells because of preceding differentiation block.

1.3 Classification of acute myeloid leukemia

AML can be divided into several subtypes depending on the underlying genetic aberrations that triggered AML. In order to distinguish different forms of AML, two common systems are used. The French-American-British (FAB) classification system was

proposed in 1976 [26]. The FAB system divides AML into eight subtypes depending on morphological and cytochemical characteristics of the diagnosed leukemia (**Table 1**).

Table 1: FAB classification of AML (modified from [26]).

Subtype	Name	Cytogenetics
M0	Acute myeloblastic leukemia (minimally differentiated)	
M1	Acute myeloblastic leukemia (without maturation)	
M2	Acute myeloblastic leukemia (with granulocytic maturation)	t(8;21)(q22;q22) [a], t(6;9)
M3	Promyelocytic or acute promyelocytic leukemia	t(15;17)
M4	Acute myelomonocytic leukemia	inv(16)(p13q22)[b], del(16q)[c]
M4eo	Myelomonocytic together with bone marrow eosinophilia	inv(16), t(16;16)
M5	Acute monoblastic Leukemia (M5a) Acute monocytic leukemia (M5b)	del (11q), t(9;11), t(11;19)
M6	Acute erythroid leukemias, including erythroleukemia (M6a) and very rare pure erythroid leukemia (M6b)	
M7	Acute megakaryoblastic leukemia	t(1;22)

[a] t(A;B)(q12;p34); balanced translocation (t); first bracket defines the affected chromosomes (here: chromosomes A and B (given in numerals)), second bracket defines the breakpoints (here: chromosome A, long arm (q), region 1, band 2 and chromosome B, short arm (p), region 3, band 4).

[b] inv(C)(p12q34); inversion; first bracket defines the affected chromosome (here: chromosomes A (given in numerals)), second bracket defines the breakpoints (here: chromosome C, short arm (p), region 1, band 2 and chromosome C, long arm (q), region 3, band 4).

[c] del(Dq); deletion; missing copy of genetic material on the long arm (q) of the affected chromosome D (given in numerals).

The second classification system was proposed by the World Health Organization (WHO), in collaboration Society for Hematopathology and the European Association of Haematopathology [27]. In addition to morphologic findings, the WHO classification system utilizes available information, including genetic, immunophenotypic, biologic, and

clinical features to define specific disease entities [28]. A list of the major AML and related neoplasm subgroups is shown in **Table 2** [29]. The WHO classification also includes eventual bone marrow dysfunction occurring in older patients that can onset acute leukemia. In summary, the WHO classification of AML defines 108 new diagnostic entities in hematopathology, including 50 new or provisional leukemia entries [30]. The AML subtype investigated in this study is the M2 subtype (FAB classification) or 'AML with recurrent genetic abnormalities' (WHO classification) with the t(8;21) translocation.

Table 2: WHO classification of AML and related neoplasms (modified from [29])

AML with recurrent genetic abnormalities
AML with t(8;21)(q22;q22.1) [a];*RUNX1-RUNX1T1*
AML with inv(16)(p13.1q22) [b] or t(16;16)(p13.1;q22);*CBFB-MYH11*
APL with *PML-RARA*
AML with t(9;11)(p21.3;q23.3);*MLLT3-KMT2A*
AML with t(6;9)(p23;q34.1);*DEK-NUP214*
AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM*
AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);*RBM15-MKL1*
Provisional entity: AML with BCR-ABL1
AML with mutated *NPM1*
AML with biallelic mutations of *CEBPA*
Provisional entity: AML with mutated RUNX1
AML with myelodysplasia-related changes
Therapy-related myeloid neoplasms
AML, not otherwise specified
AML with minimal differentiation
AML without maturation
AML with maturation
Acute myelomonocytic leukemia
Acute monoblastic/monocytic leukemia
Pure erythroid leukemia
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis

Myeloid sarcoma
Myeloid proliferations related to Down syndrome
Transient abnormal myelopoiesis (TAM)
Myeloid leukemia associated with Down syndrome

[a] t(A;B)(q12;p34); balanced translocation (t); first bracket defines the affected chromosomes (here: chromosomes A and B (given in numerals)), second bracket defines the breakpoints (here: chromosome A, long arm (q), region 1, band 2 and chromosome B, short arm (p), region 3, band 4).

[b] inv(C)(p12.3q45); inversion; first bracket defines the affected chromosome (here: chromosomes A (given in numerals)), second bracket defines the breakpoints (here: chromosome C, short arm (p), region 1, band 2, subband 3 and chromosome C, long arm (q), region 4, band 5).

1.4 PPIs: Promising targets in drug discovery

At the molecular level, cancer, like AML, is caused by genetic abnormalities which alter downstream signal transduction pathways and protein-protein interactions [31, 32]. From the perspective of system biology, the disease-perturbed protein and gene regulatory networks differ from the normal counterparts [33]. Therefore, protein-protein interactions (PPIs) are an attractive target for the identification of new drugs [34, 35]. They are involved in many biological processes and play an important role in signaling [36-38] and regulation [39, 40]. Some PPIs even form focal points by interacting with different proteins and thereby are so-called hub proteins [41]. With this level of involvement in various pathways spanning over different areas, malfunctions of even a few PPIs can result in serious consequences [42]. Accordingly, PPIs are involved in pathways leading to diseases [43-47]. Hence, the modulation of PPIs via small molecules holds great potential for possible treatment [42, 45, 48-51]. Yet, compared to the investigation of classical target, gathering information for the modulation of PPIs can be a difficult venture. These difficulties mainly arise from the size of the interaction interface and the absence of pronounced binding pockets [52]. Recent progress in the identification of PPI modulators (PPIM), however, shows that these difficulties can be overcome [35, 53, 54]. For these advances, progress in the investigation of PPI properties and PPI interfaces were utilized from interdisciplinary approaches combining biology, computational chemistry, biochemistry and biophysics [54-56]. To properly address PPIs efficiently the structural and functional characteristics of these interactions need to be determined.

1.5 PPIs: Functional and structural aspects

PPIs have distinct differences to "classical" targets. The "classical" target usually stands for small molecules or peptides binding to a well defined enzyme pocket, e.g. enzymes binding small molecules or peptides. This form of interaction is known as lock-and-key model [57]. The structural and chemical properties of the ligands are complementary to the binding site and thereby resulting in a high affinity and specificity to the targeted protein. An established approach for the inhibition of these classical targets is the identification and/or the design of small molecules with similar structural and chemical characteristics as the natural ligand. These small molecules bind to the binding site and thus prevent the interaction between the protein and the natural ligand. Information about the ligand or the characteristics of the interactions between the ligand and the protein gathered from one enzyme can be applied to members of the same protein family, if the interaction mechanisms and architecture of the binding site is conserved throughout the protein family [58, 59].

Established approaches on classical targets cannot simply be transferred onto PPIs, since PPIs differ in the size of the interaction interface, the shape of the interaction interface and the absence of well-defined pockets to classical targets. These structural differences may present the largest difficulty for the modulation of PPIs. The size of PPI interfaces is much larger than the binding site of classical targets. The size of binding sites in classical targets ranges from 300 to 1000 $Å^2$ [60-62], whereas the surface area of PPI interfaces ranges from ~1500 to 3000 $Å^2$ [63, 64]. As mentioned before, the interfaces of PPIs usually lack prominent pockets, thereby resulting in a rather flat contact area. The interaction via this flat interface can be compensated in three ways: (i) unlike deep binding pockets the flat interaction interface cannot wrap around the binding partner, which, however, is compensated by an enlarged interface area, (ii) irregularities in the surface, like bulges or indentations, result in a rough interface area [65, 66], leading to even bigger surface area. However, this compensation requires a complementary interaction partner [66, 67], and lastly, (iii) due to the matching accuracy of the interacting surfaces solvent molecules within the surface area will be repressed. These structural characteristics, especially the flat interaction interface, hamper the identification and design of PPIMs. On a flat surface the modulator cannot be enclosed, portions of the PPIM will not address the protein and cannot contribute to the binding affinity. This missing contribution can be assessed and compensated by an increased size of the PPIM [68] and thus an increased

contact area to the protein. The description of PPIs is often associated with so called hot spots as an important characteristic of the interaction interfaces [69-72].

1.6 PPIs: Hot spots

When looking at the classical targets one sees a more or less uniform principle in terms of molecular interactions. Good examples for this are ATP-binding proteins, like the ABC-transporter, which bind to the same molecule via a conserved ATP binding site [73], despite differences in their overall structure. Considering the structural characteristics of PPI, the interaction interface between the proteins is not as conserved as a binding site of a natural ligand in classical targets. Considering the large interaction area between multiple proteins a large number of surface amino acid residues come into close proximity and results in innumerably many combinations of interactions and connections between proteins. In contrast to a conserved binding site, like the ATP binding site in ABC-transporter, the chemical space within interaction interfaces of PPIs differs between targets, thus no structure-based general approach could be established [74].

Conversely, this means an approach for the identification of a PPIM for a targeted PPI needs to be tailored specifically on the target and cannot be transferred from a different target. On the first glance, this might seem like a very difficult task. Yet, successful studies already demonstrated that this challenge can be overcome [54, 74]. Furthermore, investigations of identified PPIMs revealed structural and chemical similarities, thus leading to the suggestion that there are shared and common features in the mechanism of the binding to the PPI interface [68, 74, 75], thus the area to be examined can be limited. One insight in PPIs was that many PPI interfaces of homodimers tend to be more hydrophobic within the interaction interface than the solvent accessible surface [76, 77]. This structural characteristic is quite similar to the cores of folded proteins, with hydrophobic residues buried within the protein [78, 79]. A good example for hydrophobic patches in the interface is the dimer of dimers of the Nervy homology two domain (NHR2), with its long hydrophobic interface [80].

Considering the PPI interface, it is notable that not all amino acid residues facing the interface contribute equally to the binding free energy [69, 71, 81]. Conversely, so called hot spots contribute the majority of the binding free energy and are thereby essential for the interaction [71, 81, 82]. Two definitions are used for the characterization of hot spots. First, in the original definition, hot spots are amino acid residues that contribute significantly to the binding and, if mutated to alanine, cause a drop of the binding free

energy of ≥ 2 kcal mol^{-1} [83]; second, according to the complementary definition, hot spots are regions within the surface of the protein interacting with ligands [84-86], other hot spots or anchor residues [87]. In many cases hot spots from one face of the interface interact with hot spots on the other face of the interface of the other binding partner [69, 88]. It has been reported, that hot spots form clusters in stable patches in the PPI interface [89]. These clusters of hot spots are called hot regions [90]. Hot spots within a hot region cooperate with the hot spots of the interaction partner and thereby stabilize the protein complex [90]. However, the contribution of different hot regions to the binding free energy is additive [91], indicating that hot regions are independent. The identification of hot spots or hot regions within the interface of a PPI can be a good starting point for the identification of PPIMs. Small molecules can be selected that can address and mimic the hot spots are a premise for PPIM-binding [92, 93]. In this study I work with a domain of the RUNX1/ETO fusion protein, which forms a homotetramer via PPI.

1.7 RUNX1/ETO fusion protein

The RUNX1/ETO fusion protein is the result of a translocation involving chromosome 8 and chromosome 21 (t(8;21)) [94-96] (**Figure 2**). This translocation is found in ~12% of *de novo* AML cases [97]. In 1973, the t(8;21) translocation was described for the first time when it was discovered in AML subtype M2 patient samples [98]. To this day, new insights have been gained, which suggest that the t(8;21) translocation is the initiative event of AML. The presence of RUNX1/ETO is an essential requirement for AML [99, 100], but not enough for the onset [101]. This assessment is supported by studies which found RUNX1/ETO transcripts in test persons who showed no signs of AML [102, 103]. Thus, further genetic mutations, e.g. FLT3-length mutation (FLT3-ITD) [104, 105], FLT3-TKD (D853) [104, 105], c-KIT (D816 (D>Y, D>V, D>H, D>I)) [104, 105] or c-KIT (N822K) [106], are required for the onset of AML [107-109]. Hence, AML can be explained as a multi-step process, in which course the RUNX1/ETO activity predisposes for AML and secondary mutations are required for the development of overt AML [101]. One effect of RUNX1/ETO on the cell is the repression of RUNX1 target genes [95]. As a result of the RUNX1/ETO activity, the myeloid precursor differentiation is blocked and promotes AML [100, 110, 111].

Structural aspects show that the monomer and dimer of the RUNX1/ETO fusion protein does not lead to the onset of AML [80]. However, the formation of a RUNX1/ETO homotetramer is required for the onset and the maintenance of AML [82]. To get a more

detailed depiction of the effects of the t(8;21) translocation and the resulting fusion protein, we have to take a closer look on the composition of this protein. The translocation brings together genes encoding for the RUNX1 protein, one key regulator in the hematopoietic cell differentiation [112], and genes encoding for the ETO protein [95, 100, 113]. RUNX1, a member of the RUNX-family of transcription factors, contains a Runt homology domain (RHD) at its N-terminus [114, 115]. The RHD is crucial for the binding to DNA and the formation of the core binding factor complex with the core binding factor β [116]. In turn, the formed complex interacts with hematopoietic genes responsible for myeloid and lymphoid development [117]. The importance of RUNX1 for the hematopoiesis was highlighted by knock-out experiments, where RUNX1$_{-/-}$ cells failed to contribute to hematopoiesis, and mice lacking RUNX1 die during midembryonic development [118].

The second component of the fusion protein is the ETO protein (**Figure 2**). ETO contains four conserved regions homolog to *Drosophila* Nervy, the nervy homology regions (NHR1, NHR2, NHR3 and NHR4) [100, 119, 120]. Hereinafter, I focus on NHR2 (PDB: 1WQ6) (**Figure 3**), which is an α-helical oligomerization domain within the fusion protein [80]. A sedimentation study for NHR2 revealed a distribution of the sedimentation coefficient with a peak around s = 2.7 S, thus suggesting that NHR2 is present as a tetramer in solution [80]. Thus, preventing the formation of a NHR2 tetramer, a dimer of dimers (**Figure 3A**), would be an appealing approach to prevent or stop the RUNX1/ETO activity and thereby cancel the onset and maintenance of AML. For this reason, the focus of this work has been put on the modulation of RUNX1/ETO tetramerization.

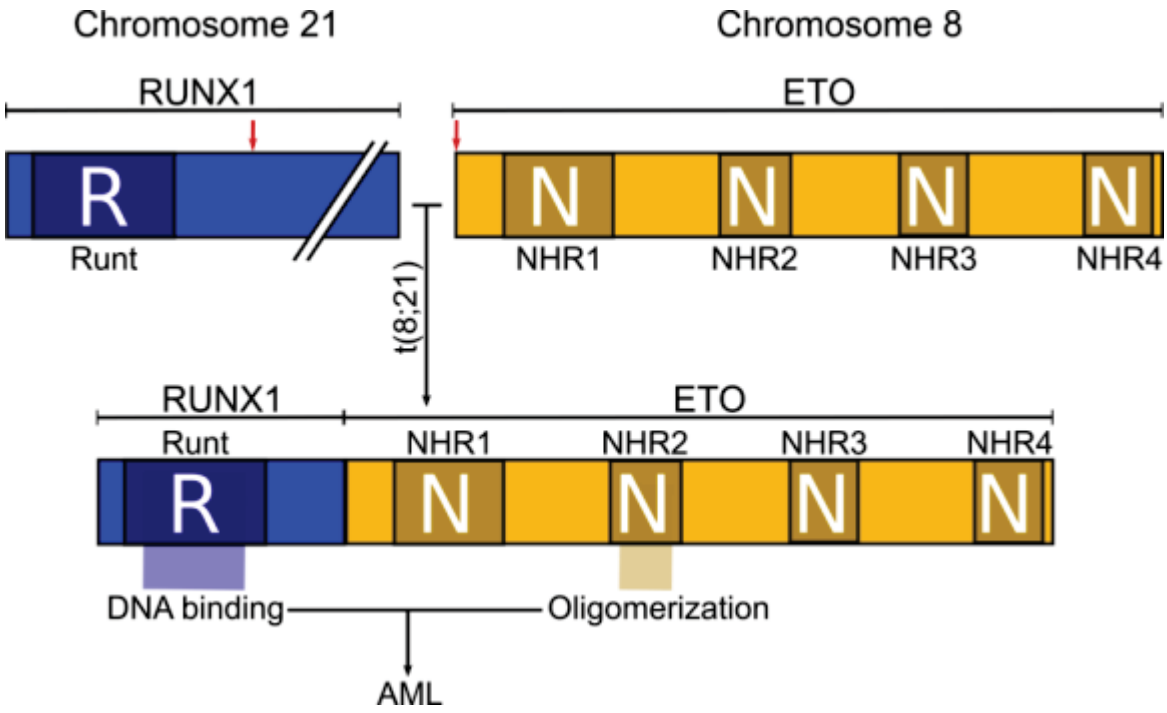

Figure 2: **Schematic representation of the formation and the composition of RUNX1/ETO.** Red arrows: breaking points on the chromosome arms and thereby fusion point of the different genes.

1.8 Modulation of the NHR2 PPI

The inhibition of the oncogenic function of RUNX1/ETO has been proven as a reasonable approach in previous studies [121]. One study describes polypeptides, generate by an Oridonin-mediated cleavage of RUNX1/ETO, interfering with the tetramerization of RUNX1/ETO [122]. Another study describes application of NHR2 domain-mimicking peptides resulting in an inhibition of the RUNX1/ETO tetramerization. The RUNX1/ETO activity requires a NHR2-mediated tetramerization of the fusion protein. Thus the inhibition of the NHR2-monomer tetramerization is an attractive approach for the treatment of AML. But due to the structural properties of NHR2 with its lack of prominent pockets and a total interaction area, containing all contact points of the four monomers with the other monomers, of around ~ 10,000 Å^2 [82], it is a good example for PPIs and thus is subject to the described challenge for the investigation of PPIs (**Chapter 1.5**)

In addition, the NHR2 tetramer (**Figure 3A**) is very stable, analyzing the thermal stability of the tetramer revealed a melting point (T_m) of ~85 °C [82]. In the past years extensive tests were performed on NHR2 for the identification of a suitable PPIM [54, 82]. The initial step was the identification of the hot spots responsible for the tetramerization of NHR2 [82], which led to five amino acids identified as hot spots (W498, W502, D533, E536 and W540) [123] (**Figure 3B**). Subsequently, the predicted hot spots were validated

with *in vitro* and *in vivo* experiments [82]. Due to the antiparallel, head-to-tail, orientation of the dimers the respective hot spots are arranged in close vicinity (**Figure 3C**). Furthermore the mutation of the hot spots to alanine prevented the tetramerization of NHR2 while retaining the helical structure of the protein [82], resulting in no onset of AML in the presence of the mutated RUNX1/ETO fusion protein. The PPI interface surrounding the hot spots of NHR2 is not completely flat, thus creating dents and shallow pockets on the surface of the NHR2 dimers. The hot spots are located at the ends of the protein close to a cavity which also is the largest pocket on the interface [82] (**Figure 3B,C**). In the following step a peptide was designed of 18 amino acids of the NHR2 wild type, in order to mimic the hot spots [54]. In an ELISA assay the inhibitory effects on the NHR2 tetramerization were validated, and a first EC_{50} (~390 μM) was determined [54]. In the next steps, results from the peptide and information about the hot spots were combined and utilized in a virtual screening. In a procedure, set up by Metz *et al.* [54], compounds from the "ZINC is not commercial" – database [124] were analyzed, scored depending on their similarity to a target template and ranked depending on their score. Top ranking compounds were selected for experimental validation. The experimental validation provided a promising small molecule inhibitor of NHR2 tetramerization, **7.18** [54] (**Figure 4**). A subsequent similarity search based on the fingerprint of compound **7.18** led to the identification of **7.44** [54] (**Figure 4**), a more effective inhibitor of the NHR2 tetramerization. *In vitro* experiments revealed an EC_{50} of ~630 μM, while in vivo experiments revealed a EC_{50} < 10 μM in human leukemic SKNO-1 cells [54]. *In vivo* experiments revealed that **7.44** significantly reduced the dissemination of leukemic cells in mice [121]. In this study, **7.44** was used as starting point for further investigation. In summary, using computational and experimental methods hot spots for the NHR2 tetramerization were determined. Thus, the approach to identify **7.44** is a good example of how the challenges concerning the investigation of PPIs can be overcome. With this strategy new PPIMs can be identified, and it was used as basis for this project.

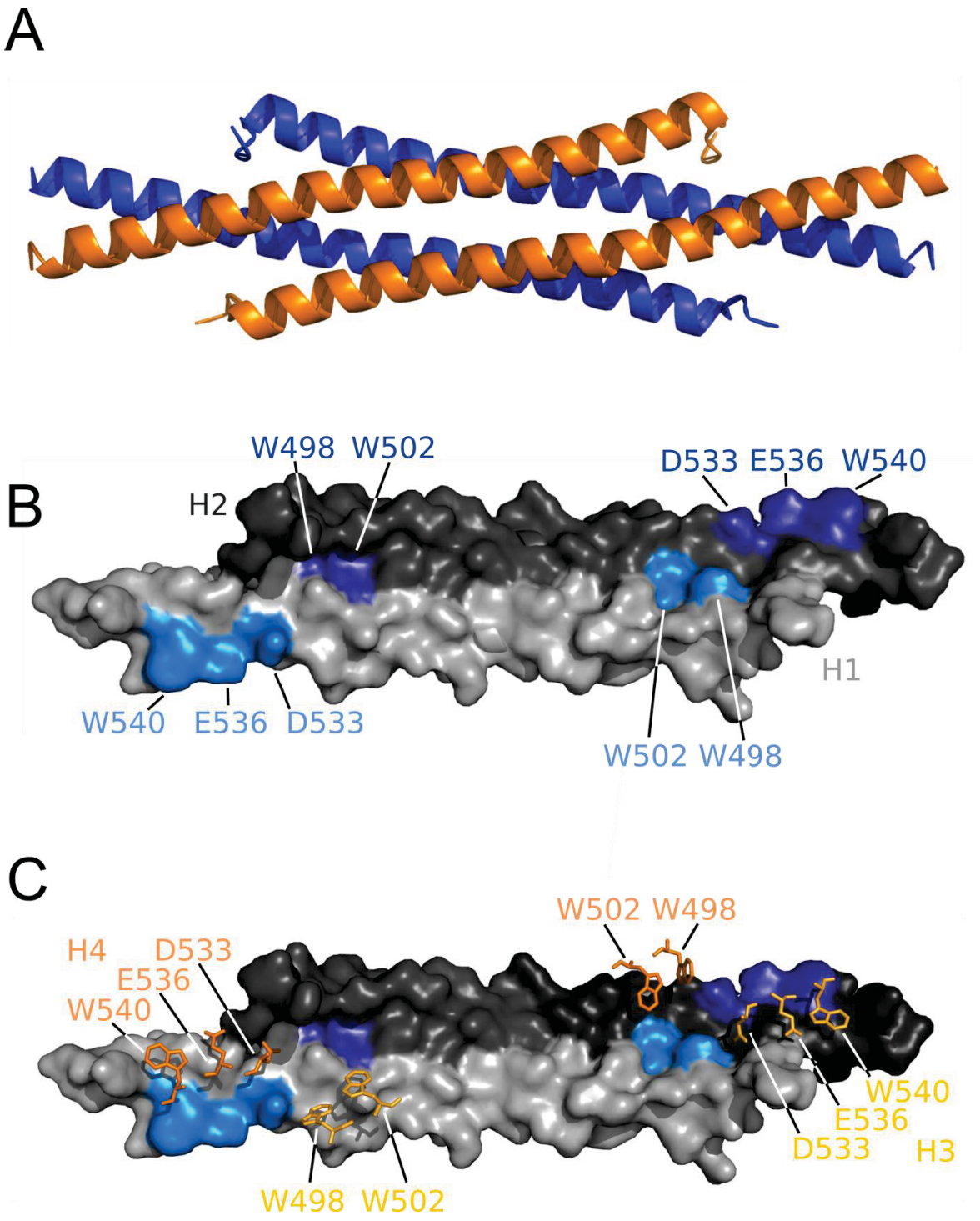

Figure 3: **Structure and hot spots of NHR2.** (**A**) Tetramerization of NHR2 by forming a dimer of dimers, blue = helix 1 of NHR2 (H1) and helix 2 of NHR2 (H2), orange= helix 3 of NHR2 (H3) and helix 4 of NHR2 (H4). (**B**) Hot Spots (W498, W502, D533, E536 and W540) of a NHR2 dimer, light grey= surface of H1, dark grey= surface of H2, light blue= hot spots of H1 (surface), dark blue= hot spots of H2 (surface). (**C**) Representation of the organization of hot spots on the interaction interface of the NHR2 tetramer: Dark orange= hot spots of H3, light orange= hot spots of H4.

1.9 Aim of this study

Until today, no effective treatment for AML has been published. However, new information about the necessity of RUNX1/ETO tetramerization for the onset of AML [82], and studies showing that the challenges about the investigation of PPIs can be overcome [54, 74], provide a better understanding and possible approaches for further investigation about potential PPIMs for RUNX1/ETO. Consequently, NHR2, as part of RUNX1/ETO, became an attractive target for the search of possible treatment options. In the course of this, the hot spots for the NHR2 tetramerization were identified, which enabled the identification of a PPIM [82].

Despite this progress, there is important information still to be gathered: (i) The dissociation constant of the NHR2 tetramer needs to be determined. This quantification of the stability of the NHR2 tetramer is crucial for the calculation of the dissociation constant of the ligands (K_{lig}). The equations for the calculation of K_{lig} and the relationship between K_D of the tetramer were described by Dr. Alexander Metz [125]. (ii) Due to the rather flat PPI interface surface, parts of the inhibitor will not face the PPI interface and are thereby irrelevant for the interaction with the protein. Thus, for a more detailed and selective optimization of the ligands the determination of the binding mode is needed, because this will reveal the essential parts of the inhibitor facing the PPI interface and interacting with the hot spots. Combining the calculated K_{lig}, as information of the efficiency of a ligand, and the information of the binding mode, as information which part of the inhibitor is responsible for the inhibition, will enable a targeted optimization of the inhibitor.

The aim of this project was the set-up of assays to gather the information about the stability of the NHR2 tetramer and the binding affinity of **7.44** in order to create a starting point for a rational improvement of this PPIM. To gather information about the stability of the NHR2 tetramer and selected potential inhibitors, the biophysical methods thermal shift assay and microscale thermophoresis were applied. To gather information about the binding mode of validated inhibitors NMR was selected as the method of choice. The experimental approach was accompanied by virtual screening to select new, untested compounds.

In summary, two main questions were addressed in this thesis: (i) How and with what affinity does **7.44** bind to NHR2? (ii) Which structural changes would make the inhibitor more effective?

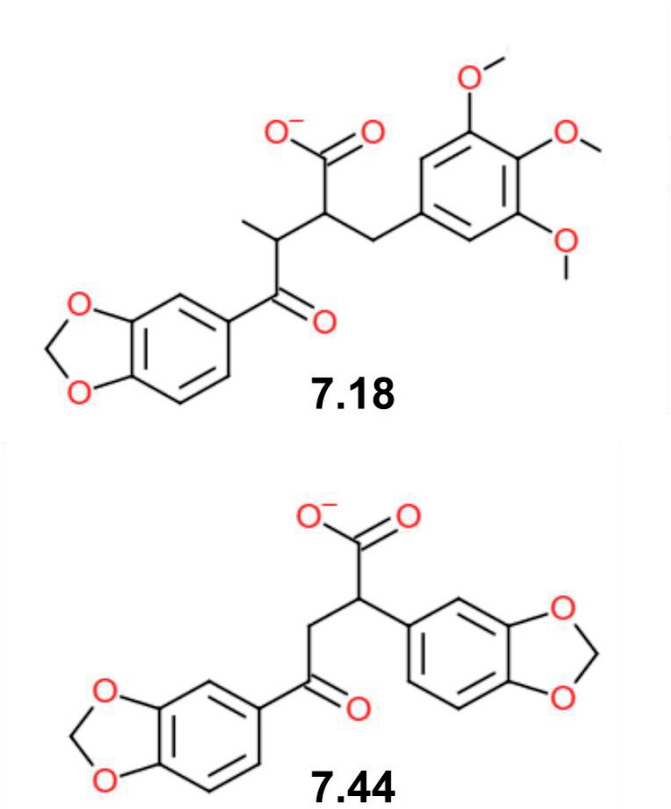

Figure 4: **2D structure of the identified PPIM of the NHR2 tetramerization, 7.18 and 7.44.**

2 Material and methods

2.1 Material

2.1.1 Chemicals

All chemicals were purchased from Sigma (Germany) or Analytik Jena AG (Germany). Isotopes were purchased from euriso-top (France). Compounds were purchased from Key organics (UK). **7.44** was provided by the National Cancer Institute (USA). Manufacturers/developers of the devices, kits or software are mentioned in the respective chapters.

2.1.2 Bacterial strains

The *E. coli*-strain BL21 (DE3) was used for the expression of the protein and stored as a permanent culture. For the preparation of plasmid DNA and consequently the sequencing of the plasmid DNA the *E. coli* strain DH5α was utilized.

Table 3: Utilized *E. coli* strains

***E. coli* strain**	**Genotype**	**Source**
BL21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	New England Biolabs
DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	Thermo Fisher Scientific

2.1.3 Plasmids

The plasmids used in this project were created and provided by Dr. Diana Kleinschrodt (Biochemistry I, Heinrich Heine University Düsseldorf, Germany) (**Figure 5**). The sequences of NHR2 and the attached tag are listed as appendix. The synthetic DNA corresponding to the coding region of residues 498 - 548 of RUNX1/ETO was cloned into expression vector pET19b in *E. coli* BL21 (DE3). A cysteine-free version of the NHR2

sequence was used where both cysteine residues were replaced by serines, to comply with the construct used in previous studies [126, 127]. At the N-terminus of the sequence a His-tag was added, thus allowing for a purification of NHR2 via immobilized metal ion affinity chromatography.

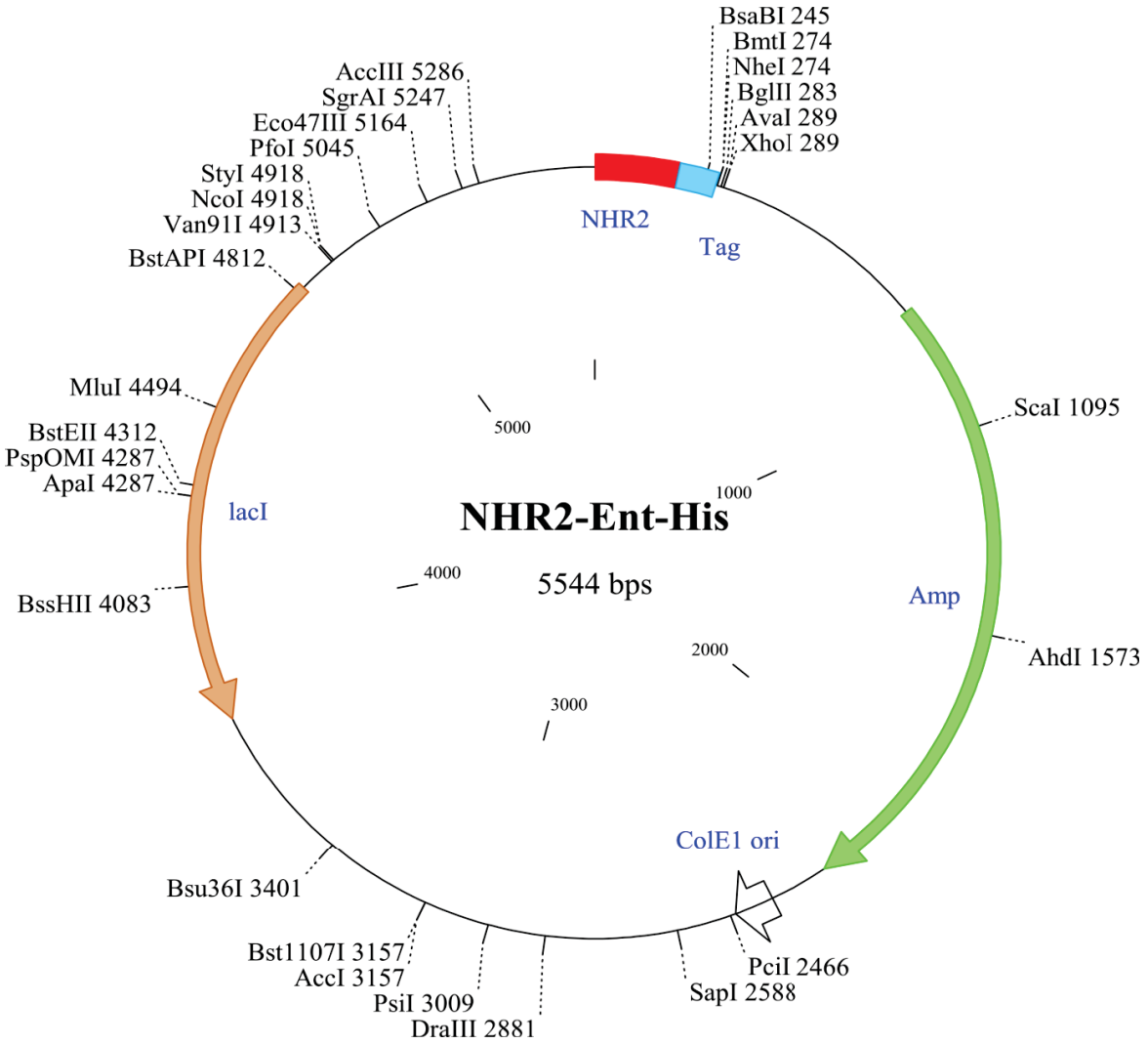

Figure 5: **Plasmid map of the pET19b vector containing NHR2.** Tag – enterokinase cleavage site, Myc-tag and His-tag attached to NHR2.

2.1.4 Enzymes

Table 4: Used Enzymes for His-Tag removal

Name	Manufacturer[a]
Recombinant enterokinase	Merck, Germany
Enterokinase, light chain	New England Biolabs, USA
Enterokinase	Abm, Canada

[a] sample composition and procedure for His-tag removal was performed according to the enclosed instructions of the manufacturers

2.1.5 Kits

Table 5: Used kits in this project

Name	Manufacturer	Usage [a]
NucleoSpin® Plasmid	Macherey-Nagel, Germany	Plasmid DNA purification
Protein labeling kit BLUE-NHS	NanoTemper Technologies, Germany	Thermophoresis
Enterokinase cleavage capture kit	Merck, Germany	His-Tag removal

[a] Proceed according to the enclosed instructions of the manufacturer.

2.1.6 Culture medium

Table 6: Composition of LB-medium

Medium	Components (per liter)	Usage
LB-medium	5 g yeast extract, 10 g tryptone/peptone, 5 g NaCl	Expression

2.1.7 Stock solution

Ampicillin and IPTG were prepared as 1000x stock solution with a stock concentration of 100 mg/ml (ampicillin) and 1 M (IPTG). The stock solutions were sterile-filtered (pore size 0.45 µm) and stored at -20 °C.

2.1.8 Buffer, solutions and gels

Table 7: Composition of regular used solutions

Buffer/solution	Composition	Usage
TFB I	30 mM potassium acetate, 100 mM $RbCl_2$, 50 mM MnCl2, 100 mM KCl, 10 mM CaCl2, 15% glycerine. The pH was adjusted to 7.0	Preparation of chemically competent cells
TFB II	10 mM MOPS, 75 mM CaCl2, 10 mM $RbCl_2$, 15% glycerine. pH was adjusted to 6.5.	
Resuspension buffer	100 mM Hepes (pH 8,0), 150 mM NaCl	Cell disruption
Li-buffer	100 mM Hepes (pH 8,0), 250 mM NaCl, 10 mM imidazole	Purification
Hi-buffer	100 mM Hepes (pH 8,0), 250 mM NaCl, 250 mM imidazole	
Stacking gel buffer	0.5 M Tris/HCl (pH 6.8), 0.4% (w/v) SDS	SDS-PAGE
4.5% Stacking gel (9 gels)	17.5 ml stacking gel buffer, 12 ml acrylamide stock solution (Carl-Roth), 42 ml H_2O, 440 µl Temed*, 140 µl 10% APS* *addition starts polymerization reaction	
Separation gel buffer	1.5 M Tris/HCl (pH 8.0), 0.4% (w/v) SDS	
15% Separation gel (9 gels)	18.75 ml separation gel buffer, 37.5 ml acrylamide stock solution (Carl-Roth), 18.75 ml H2O, 70 µl TEMED*, 300 µl 10% APS* *addition starts polymerization reaction	
Sample buffer for soluble proteins (5x)	10 ml stacking gel buffer, 10ml 10% SDS, 0.2% (w/v) bromophenol blue, 20 ml glycerine. Fill up to 50 ml with H_2O	
Running buffer (10x)	30 g Tris base, 10 g SDS, 144 g glycine (per liter)	

PBS	1,5 mM KH2PO4; 2,7 mM KCl; 8,1 mM Na2HPO4; 137 mM NaCl (pH 7.4)	Storage
PBS-EDTA	PBS + 10 mM EDTA (pH 7.4)	

2.1.9 Devices

Table 8: Devices of regular use

Device	Manufacturer	usage
Pipetman P10 Pipetman P20 Pipetman P200 Pipetman P1000	Gilson, USA	Pipetting
HT Ecotron Multitron Standard	Infors, Switzerland	Shakers
Sonopuls	Bandelin, Germany	Cell disruption
Centrifuge 5415D Megafuge 1.0R	Eppendorf, Germany Thermo Fisher Scientific, USA	Centrifugation (<50 ml)
Evolution RC	Thermo Fisher Scientific, USA	Centrifugation (>50 ml)
Optima L90K ultracentrifuge	Beckman Coultier, USA	Ultracentrifugation
Äkta Purifier (Box-900, UPC-900, P-900)	GE Healthcare, GB	Purification
Powerpak HC	Biorad, USA	Power Source for SDS-PAGE
Nanodrop ND-1000 Program: ND-1000 v3.8.1	Thermo Fisher Scientific, USA	Concentration determination
qTOWER 2.0 Program: qPCRsoft V2.0.37.0	Analytik Jena AG, Germany	Thermal Shift Assay
Monolith NT.115 Program: MO.Control; MO.Affinity Analysis	NanoTemper Technologies, Germany	Microscale Thermophoresis

2.2 Methods

2.2.1 Preparation of chemically-competent cells

5 ml LB-medium were inoculated with *E. coli* BL21 and incubated at 37 °C until an OD_{600} of ~ 0.5 was reached. The culture was centrifuged for 10 min at 2000 rpm and 4 °C. The supernatant was discarded and the pellet was carefully resuspended in 15 ml TFB I and incubated on ice for 10 min. After repeating the centrifugation, the pellet was resuspended in 2.5 ml TFB II and 100 µl aliquots frozen in liquid nitrogen. The chemically-competent cells were stored at -80 °C.

2.2.2 Transformation

Chemically-competent *E. coli*-cells were thawed on ice, 0.5 µl of the plasmid stock solution were added to 50 µl cells and then incubated on ice for 10 min. Afterwards, the transformation mix was heated to 42 °C for 60 s and then cooled down on ice again. 500 µl LB-medium was added, which was then incubated in a shaker at 37 °C and 200 rpm for 45 min. To harvest the cells, the mix was centrifugated at ~2000 rpm for 2 min. The supernatant was discarded by inversion; the cells were resuspended in the remaining supernatant and plated on agar plates containing ampicillin (100 µg/ml effective concentration). The plates were incubated at 37 °C over night. Colonies can be picked after the incubation and the picked cells were used for an expression of the protein or creation of a permanent culture.

2.2.3 Long-time storage of transformed cells

Instead of performing a transformation for every expression of NHR2, the harvested cells from the transformation were used to create permanent cultures. 5 ml LB-medium with ampicillin (100 µg/ml effective concentration) were added to transformed cells and incubated in a shaker at 37 °C and 200 rpm over night. Afterwards, 300 µl glycerin were added to 700 µl of the protein solution, thoroughly mixed and frozen in liquid nitrogen. Finally, the permanent culture was stored at -80 °C.

2.2.4 Plasmid purification and sequencing

5 ml LB_{Amp}-medium were inoculated with transformed *E. coli* XL1-Blue cells at 37 °C and 200 rpm over night. The plasmids were isolated using the NucleoSpin® Plasmid

kit following the instructions from the manufacturer. Using the Nanodrop to determine the concentration of the plasmid by measuring the UV absorption at 260 nm, the amount of plasmid in the sequencing sample was set to 80 – 100 ng/µl. The sequencing of the plasmids was performed by GATC Biotech AG (Germany). The pTI2-1-1 primer from GATC (GGCAAATATTCTGAAATGAGC) was used for the sequencing. The results were analyzed with Clone Manager 7 (Sci-Ed software). The protein was only accepted for further trials as there were no mutations in the sequence (data not shown).

2.2.5 Expression of NHR2

Prior to the expression of NHR2 100 ml LB_{Amp}-medium (0.1 mg/ml ampicillin) were inoculated with transformed *E. coli* – cells (either directly from the transformation or from a permanent culture) and incubated at 37 °C and 200 rpm over night. The cells were transferred from this pre-culture into 2 l LB_{Amp}-medium, reaching an OD_{600} of 0.1 and incubated at 37 °C and 180 rpm until the OD_{600} reached 0.6. The expression was induced by adding 2 ml 1 M IPTG (final concentration: 1 mM) to the medium and the mix was incubated at 37 °C and 180 rpm for 4 h. To harvest the cells the solution was centrifugated at 6000 rpm for 20 min. The supernatant was discarded and the cells were stored at -20 °C.

To express isotope labeled protein the cells were transferred from the pre-culture into 1 l M9 minimal medium (OD_{600} = 0.1). The expression was induced by adding 1 ml 1 M IPTG as the OD_{600} reached 0.3. The next steps were analogues to the expression in LB-medium.

2.2.6 Purification of NHR2

For the purification of NHR2 immobilized metal affinity chromatography (IMAC) was applied, using an Äkta purifier. *E. coli*-cells were resuspended in 15 ml resuspension buffer. The cells were disrupted via sonification at 50% power for three times 15 s each. The supernatant was centrifuged at 277.15K and 4000 rpm for 30 min. The resulting supernatant was centrifugated via ultracentrifugation at 277.15 K and 44000 rpm. The supernatant was transferred and mixed with Li-buffer. A HiTrap HP Chelating column (GE Healthcare) was loaded with Zn^{2+} ions and equilibribrated with the Li-buffer. The supernatant was applied to the column at a flow rate of 1 ml/min. Then, the column was washed with Li-buffer until the absorption went below 5% of the previous plateau-absorption. The protein was eluted by increasing the imidazole concentration, via linear

gradient setting, from 10 mM to 250 mM with a flow rate of 1.5 ml/min for 75 min. Fractions showing an absorption signal were analyzed with a SDS-PAGE to determine if the protein was present in the fractions and to determine the purity of the protein. Fractions containing the protein were pooled and the volume reduced to ~ 2.5 ml with an Amicon Ultra-15 centrifugal filter (30 kDa cut-off, Merck Millipore, USA). In order to exchange the buffer the protein solution was applied to a pre-equilibrated PD-10 desalting column (GE Healthcare, USA) and eluted by adding 3.5 ml PBS-EDTA to the column. The volume of the elute was reduced to ~ 1 ml (as described above) and the concentration of the protein was determined using a Nanodrop (**Chapter 2.1.9**; mode: Protein A280). The protein solution was stored at -20 °C.

2.2.7 SDS-PAGE

After expression and purification of NHR2 the amount and purity of the protein was inspected by SDS-PAGE. To investigate the success of the expression and the presence of NHR2 in the cells 1 ml was taken from the expression solution and centrifugated at 2000 rpm and 4 °C. At an OD_{600} of 1.0, 150 µl sample buffer for soluble protein were added to resuspend the pellet. The sample was mixed thoroughly and the supernatant was transferred to a new reaction tube. After heating the sample at 90 °C for 5 min, 20 µl were pipetted into a pocket of the gel. To investigate which fraction contains the protein and the purity of the protein after the purification 60 µl were taken from the fractions and mixed with 15 µl sample buffer. As a standard 5 µl PageRuler™ prestained protein ladder (Thermo Fisher Scientific, USA) were added to a separate pocket. A constant voltage of 150 V was applied until the dye front reached the end of the gel. The gel was stained for at least 1 h. Afterwards, the gel was destained for at least 1 h.

2.2.8 Thermal shift assay

The thermal stability of the protein was determined using the thermal shift assay (TSA), otherwise known as differential scanning fluorimetry [128] or thermofluor assay [129]. The method was previously described in ref. [70] and was adapted for this project. All samples contained an 1:1000 dilution of the commercially available stock solution of SYPRO Orange, a fluorescent dye (Analytik Jena AG, Germany), are gradually heated up 1 °C per circle (from 25 °C to 100 °C) while the fluorescence of the samples is monitored every circle using a real-time thermo-cycler qTOWER 2.0 (Analytik Jena AG, Germany).

The increase of the temperature destabilizes the protein and at a certain threshold temperature the protein begins to denaturate. Interactions between SYPRO Orange and the hydrophobic parts of the protein cause an increase of the quantum yield of the dye [130, 131]. The sample size was 20 µl and for each condition at least n = 8 samples were measured. The wavelengths for excitation and emission were 490 nm and 580 nm, respectively. The composition of samples containing SYPRO Orange an EDTA can be found in **Table 9** (**Chapter 6.1**).

A screening for potential buffer additives to counteract the SYPRO Orange-mediated fluorescence in the presence of ETDA was conducted in cooperation with Lukas Zubek and Neofitos Makridis (Students in the Department of Pharmacy, Heinrich Heine University Düsseldorf). The samples for the determination of suitable buffer additions, which should prevent the interference signal, contained 0.2 mg/ml of the described NHR2 construct, SYPRO Orange and one of the following buffer additions: 100 mM acetate, 100 mM glycine, 100 mM imidazole, 100 mM HEPES, 100 mM MES, 100 mM Tris or 200 mM citrate. The samples used for the investigation of the NHR2 tetramerization inhibition by **7.44** contained 0.2 mg/ml of the described NHR2 construct, SYPRO Orange, 200 mM citrate and **7.44** with concentrations ranging from 57.2 µM to 5.72 mM. The results of these experiments are shown as mean values ± standard error of the mean (SEM). The text for this method was taken and adapted from ref. [1, 2].

2.2.9 Microscale thermophoresis

The dissociation constant of the NHR2 tetramer and the dissociation constant of NHR2/PPIM complexes were analyzed using the microscale thermophoresis (MST) [132-135]. The thermophoresis of a molecule depends on the size, charge and hydration shell [135]. Thereby, molecules that interact with NHR2 and prevent the tetramerization of the protein complex lead to a change of the thermophoretic properties of NHR2. Prior to the experiment the protein was labeled using the BLUE-NHS (NanoTemper Technologies, Germany) protein labeling kit following the instructions of the manual. Alternatively the labeling was done using the Alexa Fluor® 488 NHS ester (Thermo Fisher Scientific, USA). 1 mg of the ester was dissolved in 100 µl DMSO resulting in a stock solution with a concentration of 15.5 mM. The labeling mix contained 20 µM protein and 500 µM Alexa Fluor® 488 NHS ester in 50 mM potassium phosphate buffer (pH 7.5) with a total volume of 200 µl. The mixture was incubated for 1 h. Free remaining Alexa Fluor® 488 NHS ester was removed using a PD MiniTrap G-25 column (GE Healthcare, USA). The tritrant was

prepared by making a 1:2 dilution series with a final volume of 10 µl in PCR cups. The analyte was prepared by making a stock solution with twice the final protein concentration. For the determination of the dissociation constant of the NHR2 tetramer into dimers (K_{tet}), 10 µl 100 nM labeled NHR2 in MST-buffer (included with the capillary set from NanoTemper Technologies) were mixed with 10 µl unlabeled NHR2 with concentrations ranging from 20.6 nM to 675.0 µM. For the determination of the EC_{50} of **7.44** 10 µl 100 nM labeled NHR2 in MST-buffer (included with the capillary set from NanoTemper Technologies) were mixed with 10 µl of **7.44** with concentrations ranging from 152 nM to 5.0 mM. For the investigation of the compound **T5** 10 µl 100 nM labeled NHR2 in MST-buffer were mixed with 10 µl of **T5** with concentrations ranging from 424.26 nM to 13.8 mM. For the investigation of the compound **T7** 10 µl 100 nM labeled NHR2 in MST-buffer were mixed with 10 µl of **T7** with concentrations ranging from 404.54 nM to 16.4 mM. For the investigation of small molecules as the titrant it was ensured that the final DMSO concentration never exceeded 10%. The samples were incubated for 30 min. Then, the samples were loaded into standard capillaries (K_D investigation) or premium coated capillaries (ligand investigation) and loaded into the Monolith NT.115. The experiment was controlled using the MO.Control software (NanoTemper Technologies, Germany). Prior to every thermophoresis measurement the Cap Scan was performed. This is a scan of the initial fluorescence in order to control the quality of the samples. The fluorescence was measured at 25 °C for 30 s at 50% LED power and 20% MST-power. The experiment was repeated at least 3 times independently for every sample. The K_{tet} (Eq. 1) or EC_{50} (Eq. 2) were calculated using the MO affinity analysis software (NanoTemper Technologies, Germany).

$$f(c) = [\text{unbound}] + \frac{[\text{bound}]-[\text{unbound}]}{2\left(\text{FluoConc}+c+K_D-\sqrt{(\text{FluoConc}+c+K_D)^2-4\,\text{FluoConc}\cdot c}\right)} \qquad \text{(Eq. 1)}$$

Here, f(c) stands for the fluorescence intensity, [unbound] stands for the concentration of unbound labeled and unlabeled NHR2, [bound] stands for the concentration of bound labeled and unlabeled NHR2, FluoConc stands for the concentration of labeled NHR2, c stands for the concentration of unlabeled NHR2.

$$f(c) = [\text{unbound}] + \left(\frac{[\text{bound}]-[\text{unbound}]}{1+\left(\left(\frac{EC_{50}}{c}\right)^n\right)}\right) \qquad \text{(Eq. 2)}$$

Here, f(c) stands for the fluorescence intensity, [unbound] stands for the concentration of unbound labeled NHR2, [bound] stands for the concentration of bound

labeled, c stands for the concentration of **7.44**, n stands for the Hill coefficient. The dissociation of **7.44** from an NHR2 dimer, K_{lig}, was then calculated using Eq. 4a to 15 (**Chapter 6.4**). The text for this method was taken and adapted from ref. [2].

2.2.10 LC-Mass spectrometry

The LC-Mass spectrometry was performed by Dr. Andreas Marmann (Institute for Pharmaceutical Biology and Biotechnology, Heinrich Heine University Düsseldorf, Germany). A sample was taken from the SYPRO Orange stock solution and chromatographically separated using an Agilent 1100 series Liquid Chromatography system (Agilent Technologies, USA) equipped with a quadrupole ion trap LCQ Deca mass spectrometer (Thermo Fisher Scientific, USA). The mobile phase consisted of water (A) and methanol (B) and the elution conditions were: 0-5 min, 10% B; 30 min, 10-100% B; 10 min, 100% B; 15 min, 10% B. The SYPRO Orange molecules were positively ionized by protons and negatively ionized by formate. The data were analyzed with the software Xcalibur V6.0 (ThermoFisher Scientific, USA). The text for this method was taken and adapted from ref. [1].

2.2.11 NMR spectroscopy

2.2.11.1 Investigation of SYPRO Orange

The NMR spectroscopy was performed by Dr. Finn Hansen (Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, Germany). A sample of the SYPRO Orange stock solution was lyophilized and subsequently dissolved in 0.8 ml DMSO-*d6*. A proton (^{1}H) NMR spectrum was recorded using a Bruker Avance III – 600 (600.22 MHz for ^{1}H) (Bruker, USA). Chemical shifts are given in parts per million (ppm, δ relative to residual solvent peak) or in Hertz (Hz). The text for this method was taken and adapted from ref. [1].

2.2.11.2 Investigation of **7.44** and NHR2

The NMR spectroscopy of **7.44** and NHR2 was performed by Dr. Aldino Viegas and Dr. Manuel Etzkorn (Institute for Physical Biology, Heinrich Heine University Düsseldorf, Germany). The experiments were performed at 24.85 °C using a Bruker Avance III HD spectrometer (Bruker, USA) operating at 600 MHz, equipped with a 5 mm

triple resonance TCI (^{1}H, ^{13}C, ^{15}N) cryoprobe and shielded z-gradients. Data was processed with TopSpin 3.2 (Bruker BioSpin, USA).

2.2.11.3 Assignment of ligand **7.44**

The assignment of ligand **7.44** and NHR2 was performed by Dr. Aldino Viegas and Dr. Manuel Etzkorn (Institute for Physical Biology, Heinrich Heine University Düsseldorf, Germany). For the assignment of the ^{1}H resonances a solution containing 2 mM **7.44**, 100 mM NaCl and 10% (v/v) D2O in 20 mM Na-PO_4 buffer was prepared. The assignment of the protons of the ligand was achieved by the analysis of the ^{1}H and 2D ^{1}H,^{1}H-TOCSY spectra and aided by ^{1}H chemical shift prediction as performed by the ChemNMR package, contained within the software ChemDraw (PerkinElmer, USA). The ^{1}H-NMR spectrum was acquired in a spectral window of 8417.51 Hz centered at 2823.00 Hz with 128 transients, 32k data points (k is the equivalent to multiplying the number by 1024) and a relaxation delay of 1.0 second. The solvent suppression was performed using an excitation sculpting scheme with gradients [136] in which the solvent signal was irradiated with a selective pulse (Squa100.1000) with a length of 2 ms. The 2D ^{1}H,^{1}H-TOCSY spectrum was acquired with 8 transients in a matrix with 2k data points in F2 in a spectral window of 7812.50 Hz, centered at 2823.00 Hz and 256 increments in F1 with a relaxation delay of 1.0 s. The text for this method was taken and adapted from ref. [2].

2.2.11.4 Saturation transfer difference NMR

The STD-NMR was performed by Dr. Aldino Viegas and Dr. Manuel Etzkorn (Institute for Physical Biology, Heinrich Heine University Düsseldorf, Germany). The Saturation Transfer Difference NMR (STD-NMR) was previously described in ref. [137] and was adapted for this project. For the measurement of the reference spectrum a sample containing 20 µM NHR2 and 0 µM **7.44** was prepared. For the mixture sample a sample containing 20 µM NHR2 and 2000 µM **7.44** was prepared. The spectra were acquired with 128 transients in a matrix with 32k data points in t_2 in a spectral window of 8417.51 Hz centered at 2819.34 Hz. The water proton signals were suppressed by using an excitation sculpting with gradients. A spin lock filter ($T_{1\rho}$) with a 2 KHz field and a length of 15 ms was applied to suppress protein background. Selective saturation of protein resonances (on-resonance spectrum) was performed by irradiating at 463 Hz using a series of 40 Eburp2.1000 shaped 90° pulses (50 ms, 1 ms delay between pulses) for a total saturation

time of 2.0 s. For the reference spectrum (off-resonance) the samples were irradiated at 50000 Hz. To ensure that the ligand signals were not affected, proper control experiments were performed with the reference samples in order to optimize the frequency for protein saturation and off-resonance irradiation. The STD effect was calculated using $(I_0 - I_{STD})/I_0$, in which $(I_0 - I_{STD})$ is the peak intensity in the STD spectrum and I_0 is the peak intensity in the off-resonance spectrum. The STD intensity of the largest STD effect was set to 100% as a reference, and the relative intensities were determined [138, 139]. The binding epitope is created by the comparison of the STD intensity relative to the reference one and this is described by the STD amplification factor (A_{STD}). The differences in A_{STD} for the different protons can be quantitatively expressed by analyzing the relative STD effects at a given saturation time and reflect the relative proximity of that proton to the receptor binding site. The text for this method was taken and adapted from ref. [2].

2.2.11.5 Determination of the binding mode of **7.44** and NHR2

The NMR experiments and the assignment of the residues were performed by Dr. Aldino Viegas and Dr. Manuel Etzkorn (Institute for Physical Biology, Heinrich Heine University Düsseldorf, Germany). Two-dimensional ^{15}N-^{1}H- and ^{13}C-^{1}H-edited heteronuclear single quantum coherence (HSQC) and three-dimensional HNCA, HNCO, HN(CA)CO, HN(CO)CACB and HNCACB experiments were performed to obtain the chemical shift assignments of backbone atoms. NMR experiments were performed at 298 K on a Bruker Avance III HD spectrometer (Bruker, USA) operating either at 600 or 700 MHz, equipped with a 5 mm triple resonance TCI (^{1}H, ^{13}C, ^{15}N) cryoprobe and shielded *z*-gradients. Data was processed with TopSpin 3.2 (Bruker BioSpin, USA).

2.2.12 Sedimentation velocity analysis

The sedimentation velocity (SV) analysis was performed by Dr. Luitgard Nagel-Steger and Tao Zhang (Institute for Physical Biology, Heinrich Heine University Düsseldorf, Germany). The SV analysis was used for the determination of the size distributions of EDTA aggregates in the presence of sodium ions and SYPRO Orange. The experiments were performed with a Beckman Optima XL-A ultracentrifuge (Beckman-Coulter, USA), equipped with a fluorescence detection system (Aviv, USA) and a four-hole rotor. Fluorophores could be exited at 488 nm and the resulting emission fluorescence detected in a range from 505 nm to 565 nm. The sample consisted of 40 mM EDTA, 136 mM Na^+ and 0.4% SYPRO Orange in H_2O (pH 11). The samples (100 µl each) were

loaded into a 3 mm titanium double sector cell with quartz glass windows and centrifuged at 60,000 rpm and 20 °C for 20 h and the data were collected every 90 sec. The data were evaluated using the software package Sedfit (Version 15.01b) [140]. Fitting parameters were calculated with Sednterp (Version 20130813 BETA, http://bitcwiki.sr.unh.edu/index.php/Downloads) [141], the partial specific volumes calculated according to the method of Durchschlag and Zipper [142], and the graphical outputs generated by GUSSI (Version 1.2.1) (http://biophysics.swmed.edu/MBR/software.html) [143]. The text for this method was taken and adapted from ref. [1].

2.2.13 Molecular dynamics simulations

Molecular dynamics simulations were used in this project to (a) investigate interactions between SYPRO Orange and EDTA on an atomic level and (b) investigate the binding mode of the NHR2 tetramerization inhibitor **7.44**.

2.2.13.1 MD simulation for the investigation of the SYPRO Orange – EDTA interaction

In order to set up MD simulations to investigate the interaction between SYPRO Orange and EDTA their structure had to be generated using MAESTRO 9.5 [144]. The molecular structure of SYPRO Orange was generated according to the elucidated 2D structure (Figure 20): Regarding EDTA, two variants of this molecule were generated in order to address two different protonation states of EDTA that had been determined to be of importance for further investigation. First, an EDTA variant addressing the $EDTA^{3-}$-protonation state was generated with one protonated amine group (from here on after referred to as $EDTA^{3-}$). Second, I prepared a variant with four deprotonated carboxyl groups and two deprotonated amine groups ($EDTA^{4-}$). All generated structures underwent quantum mechanical geometry optimization using Gaussian 09 [145] with the HF/6-31G* basis set. These structures were later utilized as initial structures for the MD simulations.

In the initial MD simulations, with respect to SYPRO Orange and EDTA referred to as 'initial simulations', systems were prepared containing 20 $EDTA^{3-}$ or $EDTA^{4-}$ molecules, according to a pH < 9 or pH > 10, respectively, and two SYPRO Orange molecules. To place the structures randomly in the simulation box I used *PACKMOL* [146]. Using the *LEaP* program [147] of the AMBER 12 program suite [148] TIP3P water [149] was added to the simulation box to solvate the structures. The final volume of the

simulation box resulted in a concentration of 125 mM for EDTA and 12 mM for SYPRO Orange. The force field parameters were taken from the General AMBER Force Field GAFF [150] the atomic partial charges were derived according to the restraint electrostatic potential fit (RESP) procedure [151] and previous single point calculations were conducted with Gaussian 09 [145]. The last preparation step is the neutralization of the system, this was achieved by adding sodium ions (Na^+) to the simulation box using *LEaP* [147]. The parameters for the Na^+ ions were taken from the AMBER force field ff12SB from the AMBER 12 suite of programs [148]. After completing the preparation of the simulation boxes minimization, thermalization and production calculations were performed with the *pmemd.cuda* module [152] in Amber 12 [148]. In summary, three individual rounds of energy minimization with high, low, and no positional restraints applied to all solute atoms were performed. The NVT production simulations were performed at 300.0 K (~ 27° C) for 1.0 µs. Long-range electrostatic interactions were treated by applying the particle mesh Ewald method [153, 154]. The calculated conformations were stored in a trajectory file every 20 ps.

Besides the initial simulations I prepared sets of simulation with various changes regarding the present conditions in order to investigate the influence of these changes on the SYPRO Orange/EDTA interaction: i) a simulation box only containing 16 EDTA molecules ($EDTA^{4-}$ or $EDTA^{3-}$), reducing the EDTA concentration to 100 mM, and 1 SYPRO Orange molecule , reducing the SYPRO Orange concentration to 6 mM, in order to create concentrations similar to the experimental concentration, ii) a simulation box prepared exactly like the initial simulation, however, leaving away the Na^+ ions and neutralizing the system by applying a uniform neutralizing plasma [155] (present in particle mesh Ewald simulations with periodic boundaries in Amber 12 [148]) in order to investigate whether the SYPRO Orange/EDTA interaction depends on the presence of Na^+ ions, iii) 20 previously generated octahedral $EDTA^{4-}$-Ca^{2+} complexes, derived by short MD simulations of one $EDTA^{4-}$ molecule and two Ca^{2+} ions, whose parameter were taken from Bradbrook *et al.* [156] in TIP3P water [149] beforehand, were placed in the simulation box instead of the usual EDTA molecules in order to investigate if the SYPRO Orange/EDTA interaction is affected by EDTA complexation of Ca^{2+} (the remaining charge of the system was neutralized by adding Na^+ ions), iv) a simulation system was prepared where the ion parameter for the Na^+ were used from Joung and Cheatham [157] that underwent an improvement in terms of ions' solution properties, in order to exclude a potential influence of the ion parameters on the SYPRO Orange/EDTA interaction and v)

two production simulations were performed at 333.15 K (60 °C) and 353.15 K (80 °C), which are relevant for TSA experiments.

The total simulation time of all described systems in this chapter accumulates to 27 µs. The analysis of the resulting MD trajectory was performed with *cpptraj* [158] of AmberTools 14 [159]. The trajectories were analyzed with regard to the following aspects: i) the total solvent accessible surface area as an indicator for the compactness of a resulting EDTA aggregate, ii) the so-called 'EDTA shell', the amount of EDTA molecules around SYPRO Orange (cut-off 5 Å), iii) the distance between the center of mass (COM) of the SYPRO Orange core region and the COM of the EDTA ethylene group was calculated as a measure of interactions between SYPRO Orange and EDTA molecules, iv) the radial distribution function $g(r)$ of the carbon atoms of the EDTA ethylene groups was calculated, while ignoring intramolecular distances, to provide a measure for the probability of finding an ethylene group of another EDTA molecule at a distance r away from the reference ethylene group. The text for this method was taken and adapted from ref. [1].

2.2.13.2 MD simulation for the prediction of the binding mode **7.44** and NHR2

The MD simulations in this chapter were performed by Benedikt Frieg (Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, Germany). To investigate the interaction between **7.44** and NHR2 in full atomic detail, unbiased molecular dynamics (MD) simulations of **7.44** binding to NHR2 were performed. A dimeric NHR2 structure (residues D11 – E69 for both subunits) was extracted from the tetrameric crystal structure of NHR2 (PDB-ID: 1WQ6 [80]; from here on the residue numbering of the whole fusion protein RUNX1-ETO was applied, as described in PDB-ID: 4JOL [160]). The protein structure was prepared using Maestro [161] for a pH ≈ 7.4 to mimic experimental conditions, such that all titratable residues are present in the ionic state, except for histidines, which were assigned to the HID state. The initial 3D structure of **7.44** was prepared according to the 2D structural formula in **Figure 27B** (**Chapter 3.9**). The initial 3D structures were subjected to quantum mechanical (QM) geometry optimization at the HF/6-31G* level of theory using Gaussian 09 [162]. The optimized 3D structure was used for subsequent binding simulations. Therefore, the dimeric NHR2 structure, two **7.44** molecules, and sodium ions were randomly placed using PACKMOL [163]. The systems were solvated with the TIP3P solvent model, also using PACKMOL [163]. All relevant system files for subsequent simulations were generate using LEaP of Amber17 [164]. A set of 35 independent initial configurations for MD simulations of direct

7.44 binding to NHR2 were prepared (**Figure 37A**). To describe the protein dynamics the Amber ff14SB force field was applied [165]. To describe the dynamics of **7.44** the GAFF force field was applied [166]. Missing atomic partial charges of **7.44** were derived according to the restraint electrostatic potential fit procedure [167, 168]. Ion parameters were taken from ref. [169]. The detailed minimization, thermalization and equilibration protocol is reported in ref. [170], which was successfully applied to study direct ligand binding processes [171, 172]. In summary, the solvated systems were subjected to three rounds of energy minimization to eliminate any bad contacts. Subsequently, systems were heated to 300 K and the pressure was adapted such that a density of 1 g/cm^3 was obtained. During thermalization and density adaptation, the solute was kept fixed by positional restraints of 1 kcal mol^{-1} Å^{-2}, which were gradually removed. Following, the systems were subjected to unbiased production simulations of 350 ns length each, to study **7.44** binding to NHR2 (**7.44** concentration was ≈ 1.2 mM). Noteworthy, during all binding simulations all molecules were always free to move and not biased by any artificial guiding force. All minimization, equilibration, and production simulations were performed with the pmemd.cuda module [173, 174] of Amber17 [164]. During production simulations, the time step for the integration of Newton's equation of motion was set to 4 fs by following the hydrogen mass repartitioning strategy [175]. Coordinates were stored into a trajectory file every 200 ps. Thus, all stably bound **7.44** poses were initially determined by calculating the **7.44** RMSD (only considering all non-hydrogen atoms) after superimposing the protein structure (only considered residues M513 – K518 of the dimeric interface). That way the **7.44** RMSD becomes a measure for the spatial displacement between two snapshots. All **7.44** poses with RMSD < 1.0 Å were defined as stably bound poses (**Figure 32A**) and were considered for a hierarchical clustering. The minimum distance ε between the clusters was used as cluster criterion. ε was gradually increased in 0.5 Å intervals until the population of the largest cluster remained unchanged (ε = 5.5 Å). The cluster representatives were then analyzed towards the interaction between **7.44** and NHR2. The text for this method was taken and adapted from ref. [2].

2.2.14 SAR by catalogue

Fingerprints were calculated for small molecules from the commercially available subset of the ZINC Database [124]. The similarities of the small molecules to **7.44** were determined using the MOLPRINT2D software (http://cheminformatics.org/molprint_download/) to calculate the Tanimoto coefficient

[176]. The small molecules were ranked by their score and selected by hand for further investigation. Selected small molecules were purchased from Key Organics (UK) and then tested using the MST assay (**Chapter 2.2.9**). A list all compounds, including their Key Organics and ZINC database ID, can be found in **Table 10** (**Chapter 6.7**).

3 Results

3.1 Getting started: Protein expression and -purification

At the start of the project the provided plasmids (**Chapter 2.1.3**) were transformed into *E. coli* BL 21. Before focusing on the development and optimization of biochemical or biophysical assays, the main focus lied upon the question if standard conditions of expression and purification of soluble proteins are sufficient for the preparation of NHR2, and if the expression and purification generates enough pure NHR2 to perform biochemical and biophysical investigations. In order to answer this, the protein was expressed in *E. coli* BL21, using LB-medium as media, at 37 °C for 4 hours.

After the expression of the protein the target protein was purified using the IMAC procedure for His-tagged protein [177]. As shown in the chromatogram (**Figure 6A**), after initiation of the gradual change of the buffer composition, the UV-absorption presents a local maximum at around 175 ml. Judging by the size of the peak and under the assumption that the peak is the result of the elution of the target protein, this spectrum indicates a good amount of protein sufficient for the planned experiments. The fractions of the peak were selected for the buffer exchange and the "quality control" of the solution.

During the buffer exchange, samples were taken from the solution in order to control the purity of the protein using SDS-PAGE. The resulting gel shows clear bands in the size range of the NHR2 monomer of ~ 11.3 kDa (**Figure 6B**); while a control fraction from the buffer-exchange flow-through contains no protein. This result validates the performed procedure as sufficient to generate enough pure NHR2 to carry out the following investigations.

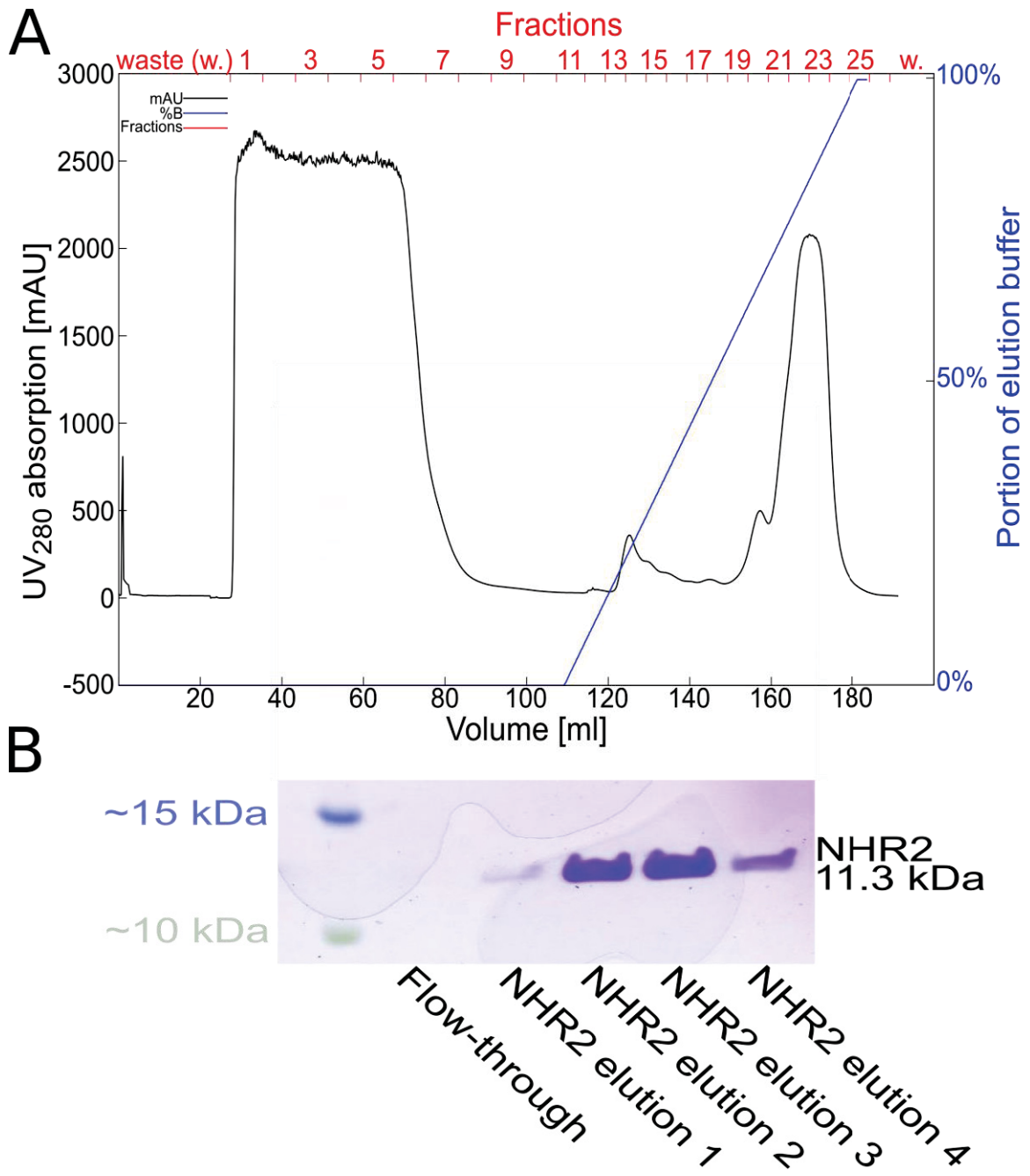

Figure 6: Purification of NHR2 after expression in *E. coli* BL21. (**A**) Chromatogram of the purification of NHR2 using IMAC (black: milli absorption units, blue: percentage of elution buffer B in buffer composition, red: collected fractions); Fractions 21-24 used in buffer exchange. (**B**) Purity control of NHR2 purification after final buffer exchange with SDS-PAGE.

3.2 First thermal shift assay experiments to probe thermal stability of NHR2

Investigating the thermal properties of the NHR2 protein, Wichmann *et al.* found that the NHR2 tetramer has a T_m at 85 °C [82]. Additionally, a mutant variant of NHR2 not able to form tetramer, however, presented a T_m at 60 °C. Taking this reduction of the T_m of around 25 °C into consideration, adding an inhibitor of the NHR2 tetramerization to the protein would have the same effect on the T_m due to the interruption or prevention of the tetramerization. The TSA enables the investigation of several different conditions at the same time at relatively low cost and with a short preparation time of less than an hour. In order to set up this assay, the ideal conditions for the system needed to be determined. Using a PBS buffer for NHR2 the main concern of condition optimization was the pH of the system, since no information concerning potential complications with the assay were known at the time. The pH screening revealed, on the one hand, typical TSA curves [178, 179] of the relative fluorescence in the presence of NHR2 at a pH ≥ 10 with a determined T_m of 70 °C but no increase in the fluorescence intensity at a pH < 10 (**Figure 7A**). Unexpectedly, the screening also revealed untypical TSA curves in the absence of the protein at a pH ≥ 10 with a determined "T_m" at ~78 °C (**Figure 7B**). A signal in the absence of protein was not described in literature and contradicted the general description of the mechanism of SYPRO Orange.

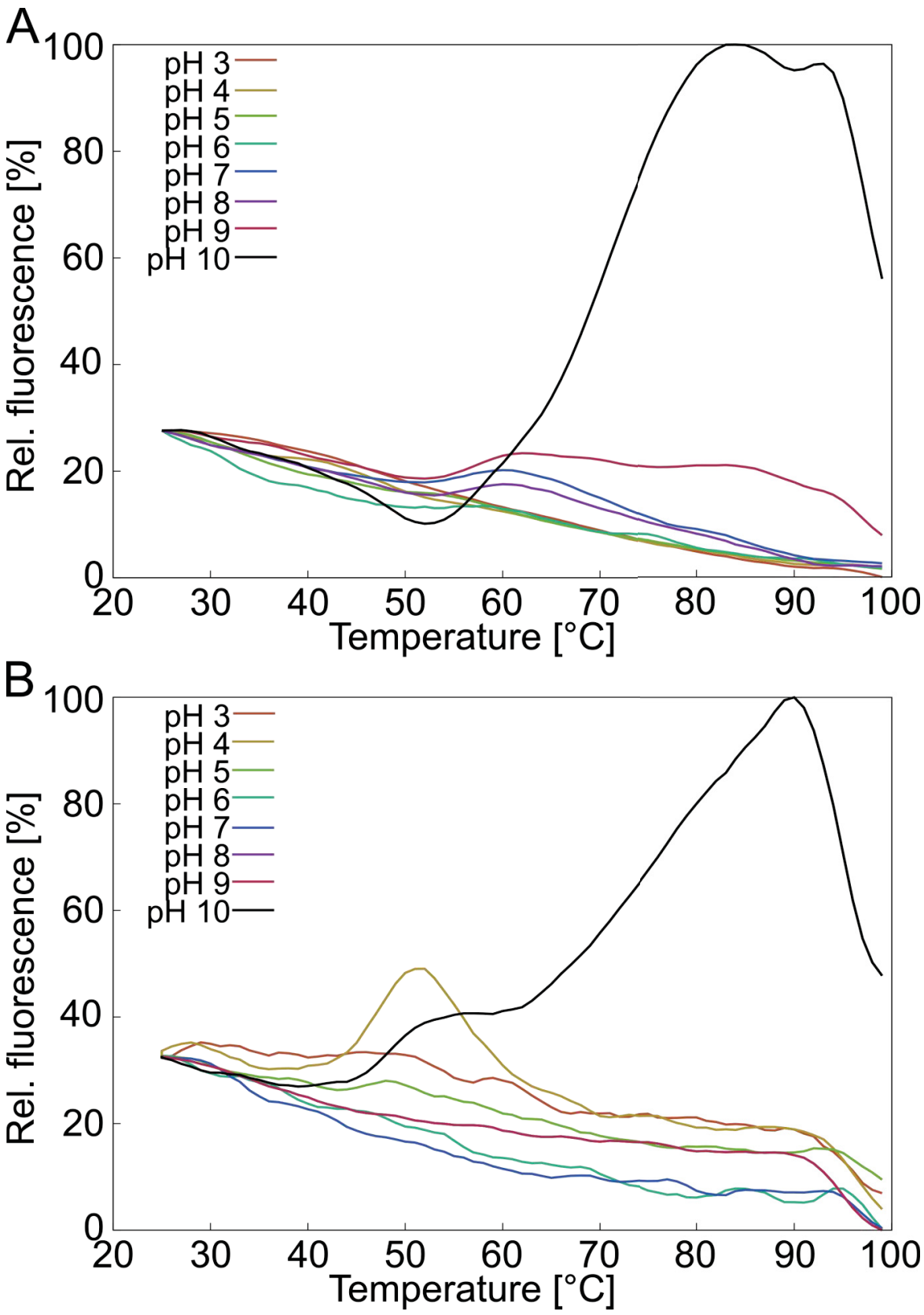

Figure 7: Thermofluor assay signal of NHR2 at different pH. (**A**) Rel. Fluorescence signal in the presence of NHR2 in PBS buffer. (**B**) Rel. Fluorescence signal of the PBS buffer (control).

For the following Chapters (**Chapter 3.3 - 3.5** (**3.5.1- 3.5.3**)), the text and figures have been taken and modified from the publication:

"**Kroeger, T.**, Frieg, B., Zhang, T., Hansen, F.K., Marmann, A., Proksch, P., Nagel-Steger, L., Groth, G., Smits, S.H.J., Gohlke, H. *EDTA aggregates induce SYPRO Orange-based fluorescence in thermal shift assay.* **PLOS ONE** 2017, 12, e0177024."

3.3 $EDTA^{4-}$ and $EGTA^{4-}$ interfere in SYPRO Orange based thermal shift assay

The increased fluorescence intensity in the absence of a protein and the lowered "melting point" in the presence of the NHR2 tetramer were challenges whose origin had to be identified and investigated in order to proceed with the set up of this assay. To identify the cause of this unexpected fluorescence signal of SYPRO Orange (1:1000 dilution according to previously performed TSA studies [70]) in water, PBS buffer (1.5 mM KH_2PO_4, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 137 mM NaCl) or PBS buffer + EDTA-Na^+ (25 mM) at pH 10 (adjusted with NaOH) without protein were performed. Similar to the initial TSA test this experiment revealed typical TSA curves with a calculated "T_m" at ~68 °C, only in the presence of EDTA (**Figure 8**). This result indicates an interference of EDTA in the SYPRO Orange based fluorescence. This temperature-dependent change of the fluorescence intensity, caused by the interference, results in a signal similar to a fluorescence signal caused by an unfolding protein.

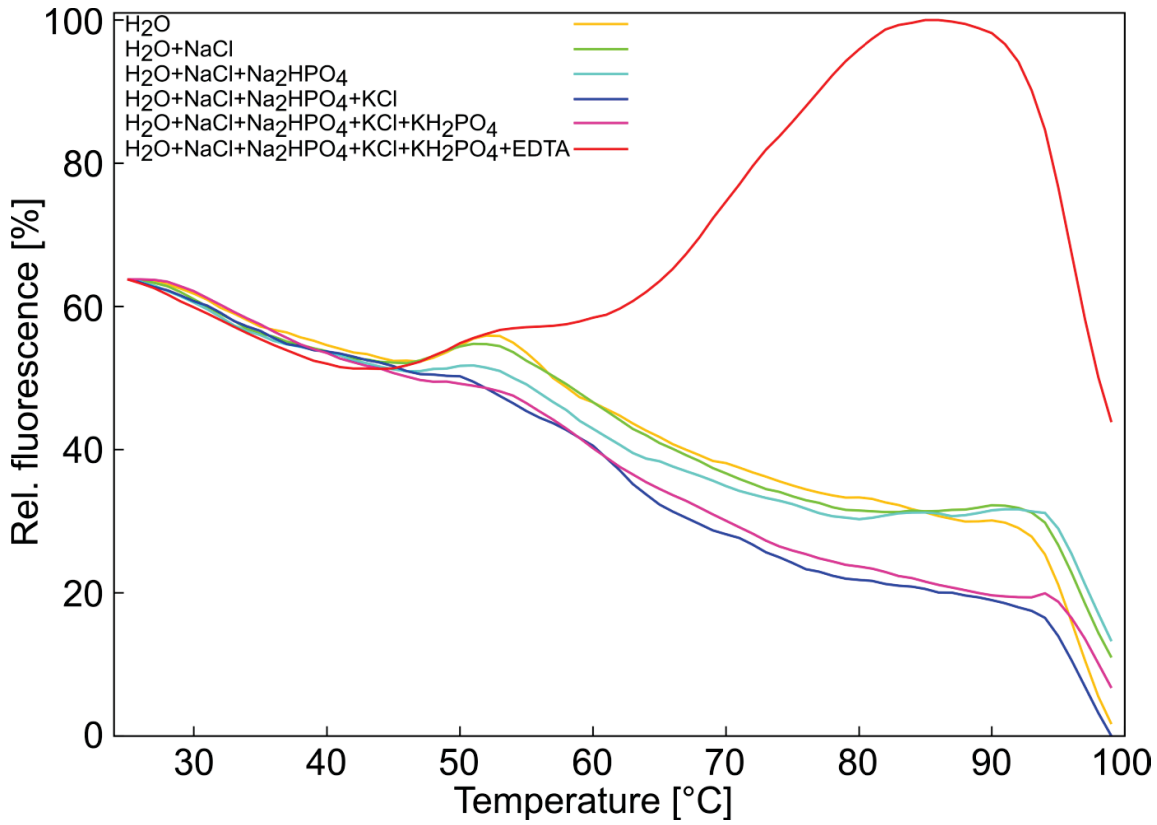

Figure 8: Thermofluor assay signal in the absence of protein. Fluorescence signal of SYPRO Orange in: H2O (red); H2O + 137 mM NaCl (green); H2O + 137 mM NaCl + 1.5 mM KH2PO4 (blue); H2O + 137 mM NaCl + 1.5 mM KH2PO4 + 2.7 mM KCl (dark blue); H2O + 137 mM NaCl + 1.5 mM KH2PO4 + 2.7 mM KCl + 1.5 mM KH2PO4 (purple); H2O + 137 mM NaCl + 1.5 mM KH2PO4 + 2.7 mM KCl + 1.5 mM KH2PO4 + 100 mM EDTA (red); the signals were normalized to the overall highest detected signal. Figure taken from ref. [1].

To scrutinize the interference of EDTA and SYPRO Orange, a TSA experiment with SYPRO Orange (1:1000 dilution of the commercial available stock solution) and various concentrations of EDTA-Na^+ in water at pH 10 was performed. With increasing EDTA concentration higher changes of the fluorescence intensity were measured, leading to an EDTA-concentration depending curve of the relative fluorescence. For this concentration dependence an EC_{50} = 36.3 ± 0.6 mM EDTA was calculated (**Figure 9A**). At a concentration of 36.3 mM a molar ratio between SYPRO Orange and EDTA of 1:3.6 is reached which indicates that multiple EDTA molecules must interfere with a SYPRO Orange molecule in order to trigger the change in fluorescence intensity. A possible explanation for this might be the formation of a supramolecular EDTA aggregate [180]. These results confirm that EDTA is the cause of the interference. Considering the properties of EDTA, especially the pH-dependent protonation states of this molecule, one can suggest that a specific protonation state causes the interference. In our case of interest

are $EDTA^{3-}$, prevailing at 6.1 < pH < 10.3, and $EDTA^{4-}$, prevailing at pH > 10.3 [181]. To determine which of these EDTA species triggers the change of fluorescence intensity, I performed TSA experiments with SYPRO Orange (1:1000 dilution of the stock solution) and EDTA (100 mM) in water with varying pH (between 7.0 and 12.0). At pH > 9 an increase in the relative change of the fluorescence intensity occurs, reaching a half-maximal relative change of the intensity at pH 10 (**Figure 9B**). At this pH the $EDTA^{4-}$ concentration is ~27 mM in the sample solution. The maximal relative change in fluorescence intensity is reached at pH 12, reaching an $EDTA^{4-}$:$EDTA^{3-}$ ratio in the solution of ~70:1. These results show that the EDTA-subspecies $EDTA^{4-}$ is causing the fluorescence in presence of SYPRO Orange.

Another property of EDTA is its ability to complex metal ions; $EDTA^{4-}$ binds mainly ions with a charge ≥ 2. The interaction is usually carried out by the four carboxylates and two amines, thus forming a stable complex with an octahedral geometry [182]. In order to determine whether this capability to form octahedral structures of $EDTA^{4-}$ is causing or impedes the interference, I performed a TSA experiment with SYPRO Orange (1:1000 dilution of the stock solution), EDTA (100 mM) and varying concentration of Ca^{2+}-ions (0 - 500 mM) in aqueous solution at pH 10. This titration of Ca^{2+} led to a quenching of the fluorescence signal with an EC_{50} = 100 ± 1.4 mM (**Figure 9C**). The result demonstrates that $EDTA^{4-}$ when forming a complex with ions cannot trigger the interference. In addition, the Ca^{2+} titration showed that the $EDTA^{4-}$-caused interference can be reversed, thus, suggesting a non-covalent interaction between SYPRO Orange and EDTA.

Lastly, in order to investigate whether this interference is $EDTA^{4-}$-specific or can be caused by similar molecules, I performed TSA experiments with SYPRO Orange (1:1000 dilution of the stock solution) and EGTA (0 -60 mM) in aqueous solution at pH 10. Structurally EGTA is closely related to EDTA, where a bis(2-aminoethoxy)ethylene moiety in the former replaces a 1,2-diaminoethylene moiety. EGTA has its largest pK_a at 9.4 [183] resulting in a $EGTA^{4-}$:$EGTA^{3-}$ ratio of ~3.33:1 at pH 10. $EGTA^{4-}$ has the capability to bind Ca^{2+}-ions in a similar fashion as $EDTA^{4-}$ [184]. Adding EGTA to SYPRO Orange in an aqueous solution led to an increase of the relative fluorescence (**Figure 9D**). At the highest applied EGTA concentration the SYPRO Orange:EGTA ratio was 1:6. Similar to the TSA experiments with EDTA, the addition of Ca^{2+}-ions to the sample quenches the detected fluorescence signal (**Figure 9D**). Thus,

EGTA, presumably the $EGTA^{4-}$ species, can trigger the SYPRO Orange-based fluorescence in a comparable way as $EDTA^{4-}$.

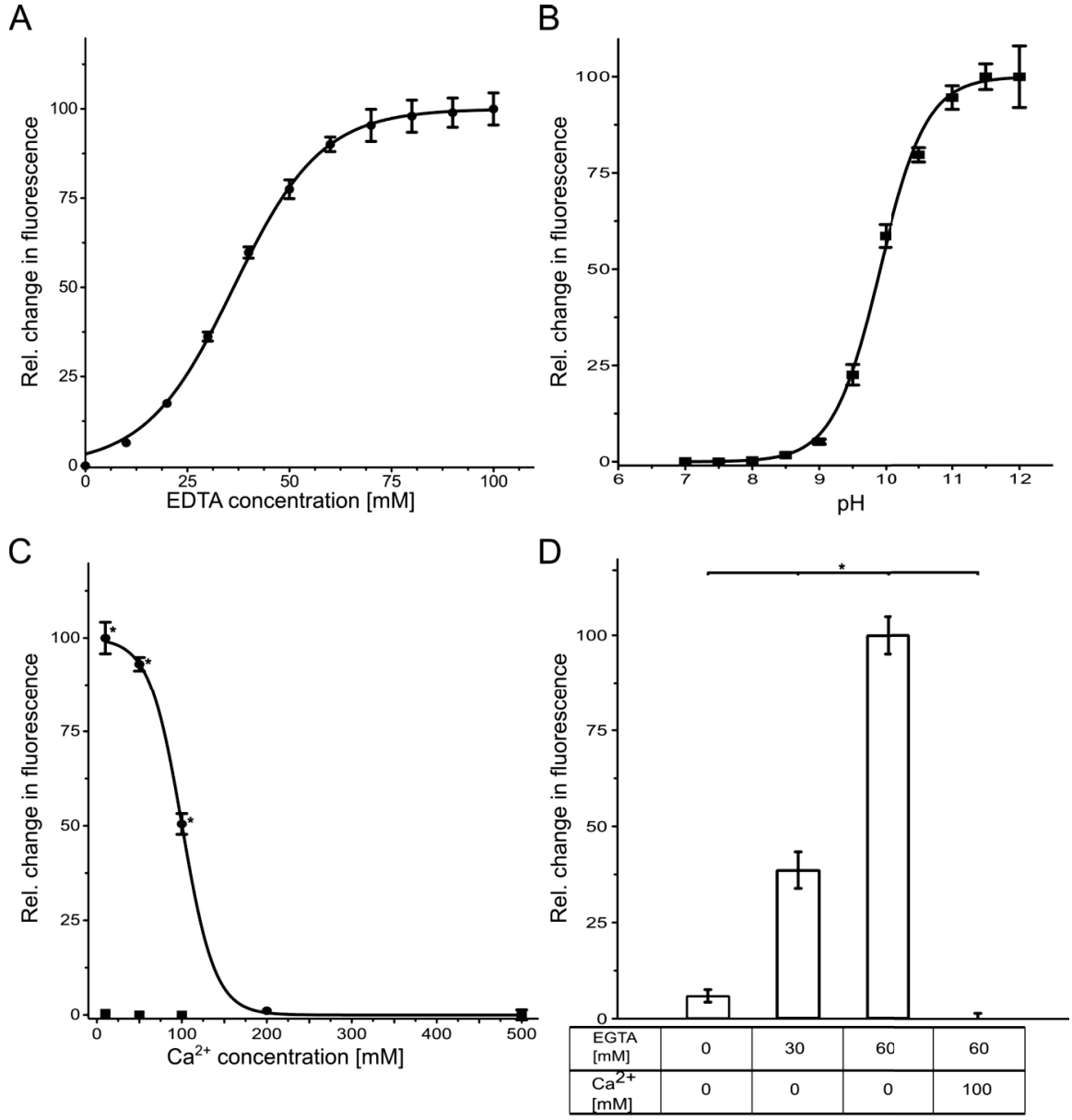

Figure 9: Influence of EDTA and EGTA on the SYPRO Orange-based fluorescence in a Thermofluor assay. (**A**) Dependence of the relative change in fluorescence intensity on the EDTA concentration (pH = 10); EC_{50} = 36.3 mM EDTA. (**B**) Dependence of the relative change in fluorescence intensity of the SYPRO Orange/EDTA system on the pH (100 mM EDTA); half-maximal change of fluorescence at pH = 9.9. (**C**) Dependence of the relative change in fluorescence intensity of the SYPRO Orange/EDTA system on the addition of Ca^{2+} (pH 10). Circles represent results for a sample in the presence of 100 mM EDTA, EC_{50} = 100 mM Ca^{2+}; squares represent results for a sample in the absence of EDTA (negative control). (**D**) Dependence of the relative change in fluorescence intensity of the EGTA – SYPRO Orange system (pH = 10) on EGTA concentration and the absence or presence of Ca^{2+}. (**A**)-(**D**): Values are normalized with respect to the minimal (0%) and maximal (100%) change detected. The error bars show the SEM. *: $p < 0.0001$. Figure taken from ref. [1].

3.4 Sedimentation velocity analysis of EDTA aggregates

From the previous results the idea arose that EDTA forms a supramolecular aggregate in order to cause the interference in the SYPRO Orange based TSA. The formation of such aggregate was already described by Müller and Haeberli [180].

In order to analyze if the formation of an EDTA aggregate increases the fluorescence intensity, sedimentation velocity (SV) experiments were performed by Dr. Luitgard Nagel-Steger and Tao Zhang (Department of Physical Biology, Heinrich Heine University Düsseldorf). The SV experiments were used with a detection system, which allows the excitation of SYPRO Orange at a wavelength of 488 nm. The collected concentration profiles show the sedimentation boundaries ranging from the beginning of the experiments, displayed with purple lines, to the end of the experiments, displayed with red lines (**Figure 10**). These sedimentation boundaries, depicting the distribution of SYPRO Orange in the solution, are shifting from the meniscus of the sample cell (low radius) to the bottom of the cell (large radius) over the course of the experiments. The shift of the distribution of SYPRO Orange suggests that this fluorescent dye sediments over time in the centrifugal force field.

A c(s) distribution analysis revealed an average $s_{20,w}$-value (sedimentation coefficient value) of the predominant species that sedimented in solution of 0.6 S. Additionally, at 2.15 S a small fraction of larger species was found. The weighted average frictional ratio f/f_0 of the $c(s)$ analysis was 2.2, implying a non-globular, extended shape for the s-value species. The control sample, containing only SYPRO Orange, had an insufficient fluorescence signal. An accurate evaluation for this sample was not possible (data not shown). The results of these experiments proved that EDTA forms supramolecular aggregates at high pH, and thus, which may lead to an enhanced fluorescence signal of SYPRO Orange.

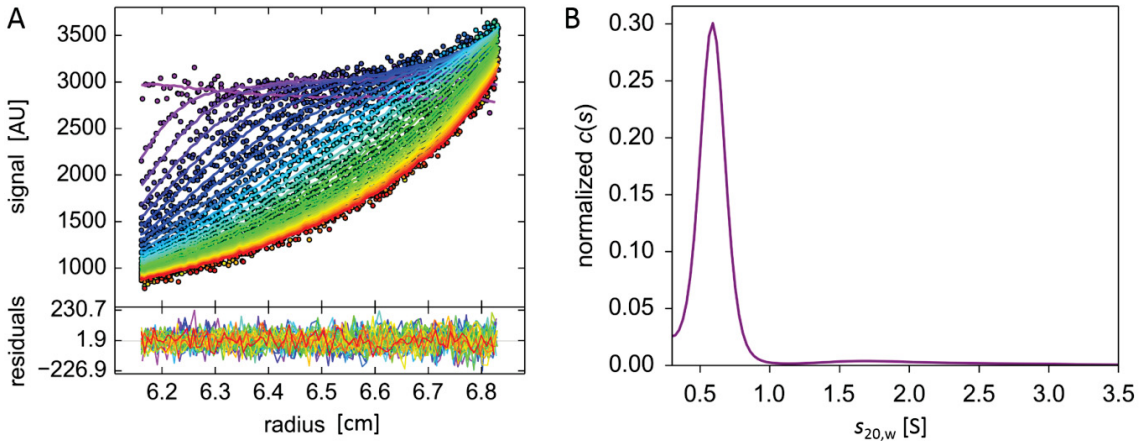

Figure 10: SV analysis on EDTA in the presence of Na+ and SYPRO Orange at pH 11. The analyzed sample consisted of 40 mM EDTA, 136 mM Na^+ and 0.4% SYPRO Orange. (**A**) Original sedimentation profile and *c*(*s*) fitting output of the sample. Real time data acquired by the detector were shown in colored dots, while the deconvolution results were displayed in colored curves; the colors are ranging from purple (beginning of the experiment) to red (end of the experiment). Fitting residuals were shown at the bottom of the graph. (**B**) The sedimentation coefficient distribution of the sample determined by *c*(*s*) model. Data were normalized according to the area under the curve and expressed as $s_{20,w}$-values. Figure taken from ref. [1].

3.5 Predicting the molecular structure of EDTA aggregates

With the gathered results from the biophysical experiments described in the previous chapters and the previous description of the formation of supramolecular EDTA aggregates [180] the next set goal was to depict the molecular structure of an EDTA aggregate and the resulting interaction with SYPRO Orange.

3.5.1 Structure elucidation of SYPRO Orange

Regarding the structure of SYPRO Orange two structures of orange-colored fluorescent SYPRO® stains with a different length of the alkyl chains at the aniline nitrogen have been published [185]. Since no information were available about which structure was present in the SYPRO Orange used in this project, we used liquid chromatography-mass spectroscopy to determine the mass of the molecules (**Figure 11**) and NMR spectroscopy to determine the configuration of the double bond (**Figure 12**). The experiments in this chapter were performed by Dr. Andreas Marmann (Institute for

Pharmaceutical Biology and Biotechnology, Heinrich Heine University Düsseldorf, Germany) and Dr. Finn Hansen (Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, Germany).

The determination of the mass revealed a molecular mass of 487.3 Da (positive ionization with a proton; **Figure 11A**) and 531.1 Da (negative ionization with formate, **Figure 11B**), resulting in a molecular mass of ~486.3 Da and thus pointing to the described structure 304. The determination of the double bond configuration (**Figure 12A, B**), using NMR spectroscopy, revealed a coupling constant of the protons at the double bond of ^{3}J = 16 Hz, thus, pointing to a trans configuration of the double bond (**Figure 20**). Using these information the actual structure of SYPRO Orange was prepared for MD simulation.

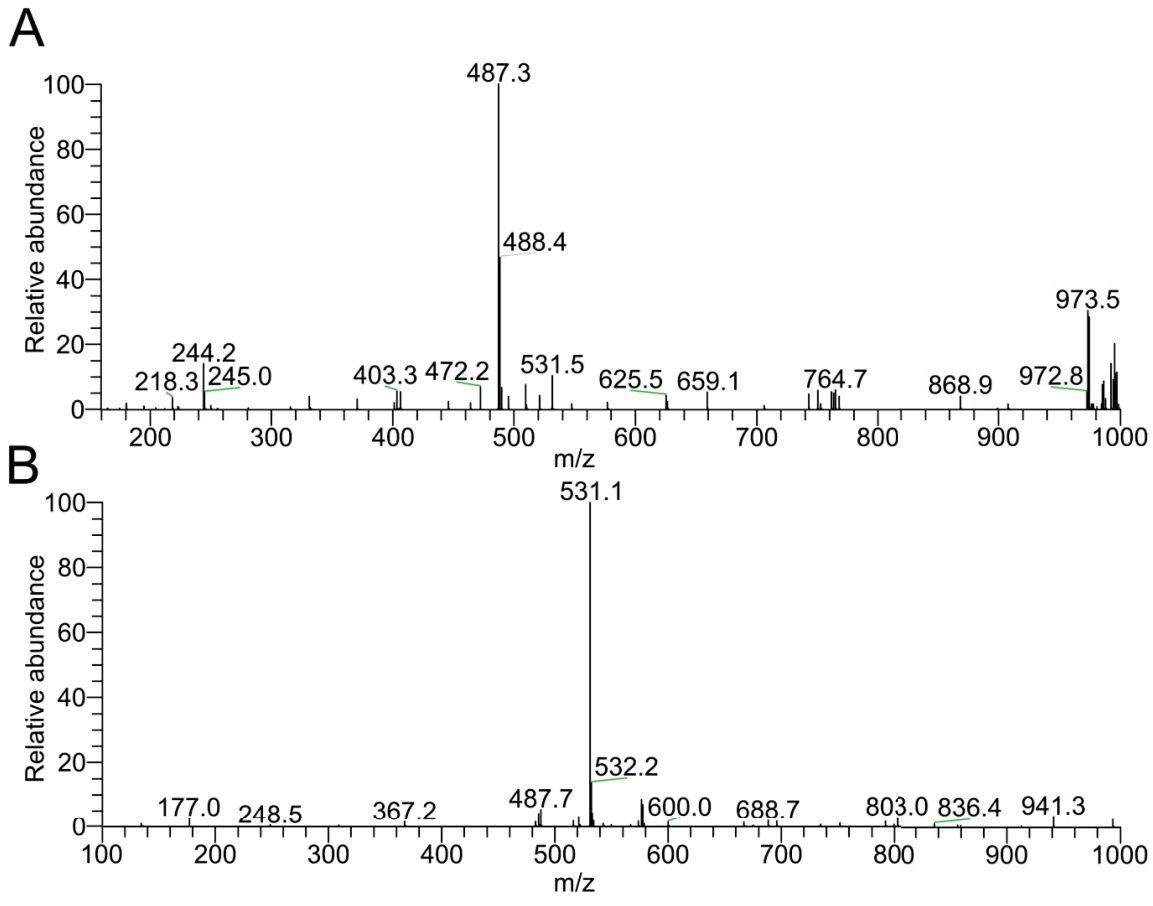

Figure 11: Mass spectra of SYPRO Orange. (**A**) Results for the pseudo molecular ion obtained by positive ionization with a proton; 487.3 Da. (**B**) Results with negative ionization with formate; 531.1 Da. The calculated mass for the investigated SYPRO Orange ion is 486.3 Da, in agreement with the structural formula shown in **Figure 20**. Figure taken from ref. [1].

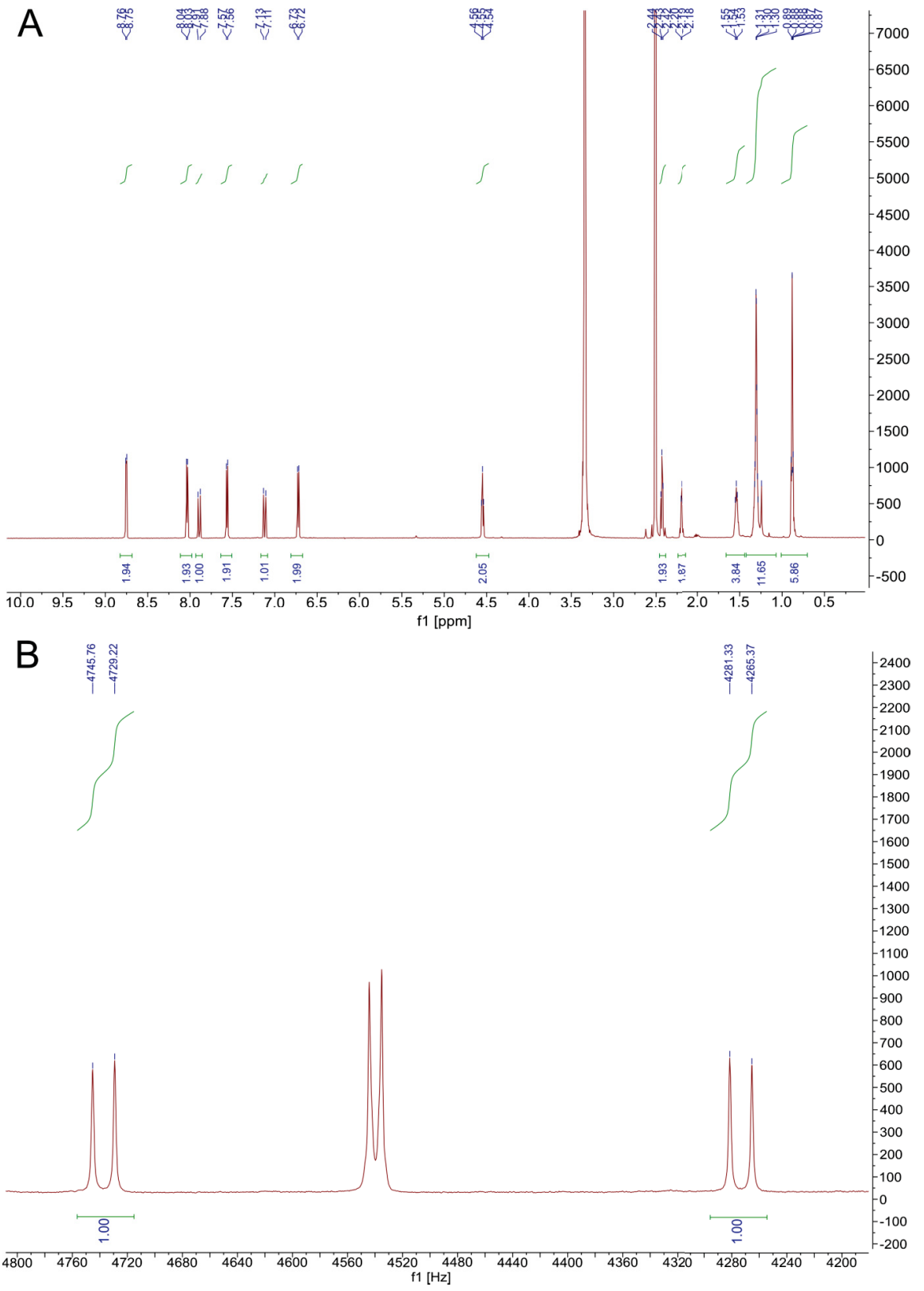

Figure 12: NMR spectroscopy to determine the configuration of the double bond in SYPRO Orange. (**A**) Full proton (^{1}H) NMR spectrum of SYPRO Orange in DMSO-d_6. (**B**) Zoom into the SYPRO Orange-specific part of the spectrum in panel A (between 7.0 and 8.0 ppm); the coupling constant of the protons at the double bond is ^{3}J = 16 Hz, revealing a trans configuration of that bond. Figure taken from ref. [1].

3.5.2 Structural model of the supramolecular EDTA aggregates

As mentioned earlier, EDTA forms aggregates, however information about the configuration of such an aggregate on the atomic level was missing. In order to visualize the structural model and to clear up how this aggregate can interfere with SYPRO Orange, I performed all-atom MD simulations of EDTA-Na^+ and SYPRO Orange in explicit solvent. First, I performed MD simulations with a length of 1 μs each, containing 20 $EDTA^{4-}$ molecules and two SYPRO Orange molecules, resulting in a concentration of 125 mM, and 12.5 mM, in the simulation box, respectively (**Figure 13**). In a control set up the $EDTA^{4-}$ molecules were replaced with $EDTA^{3-}$ molecules. In any case, sodium counter ions, as they are part of the PBS buffer and are often combined with EDTA in commercial batches (EDTA-$2Na^+$), were added to the system to neutralize the overall charge of the system. Inspecting the trajectories of the MD simulations I observed an aggregation of $EDTA^{4-}$ but no aggregation of $EDTA^{3-}$ which is in agreement with the experimental results from Müller and Haeberli [180].

To quantify the observed events, the solvent accessible surface area (SASA) for all EDTA molecules was measured. As $EDTA^{4-}$ aggregate the SASA decreases from ~7000 $Å^2$ by ~40% to ~4400 $Å^2$ after 500 ns (**Figure 13A**). After additional 200 ns, the SASA briefly increases again, indicating a collapse of the aggregate with reformation shortly later. The SASA measured for $EDTA^{3-}$ fluctuates around the initial value of ~6300 $Å^2$ and remains > 5300 $Å^2$ for the duration of the MD simulations (**Figure 13A**). Furthermore the frequency distribution of the measured SASA revealed two major populations in the case of $EDTA^{4-}$ (peaks at ~5000 $Å^2$ and ~4500 $Å^2$) and one for $EDTA^{3-}$ (peak at ~ 6300 $Å^2$) (**Figure 13A**). The results from these MD simulations point out that the $EDTA^{4-}$ species forms aggregates, but not the $EDTA^{3-}$ species. Further visual inspections of the aggregate unveiled the formation of a bilayer-like core in the center of the aggregate, containing a layer of sodium ions separating two layers of $EDTA^{4-}$ (**Figure 13B**), with their negatively charged carboxylic function pointing towards the central sodium ion layer (**Figure 13C**). This position of EDTA results in a bent structure of the molecules, leading to the protrusion of the ethylene moiety to the outside of the bilayer.

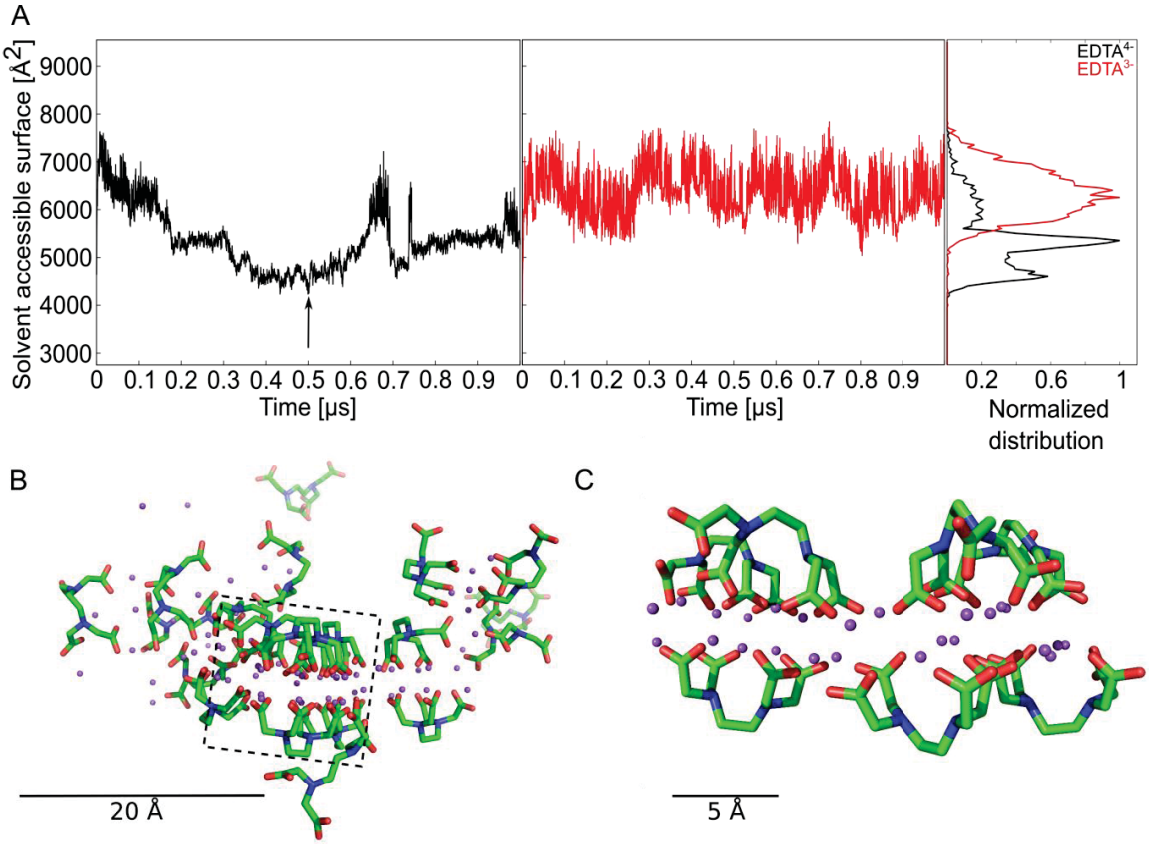

Figure 13: MD simulations to investigate EDTA aggregation and interaction with SYPRO Orange. (**A**) SASA and frequency distribution of the SASA over all 20 EDTA molecules in the simulation box; $EDTA^{4-}$ (black), $EDTA^{3-}$ (red), the arrow shows which snapshot was used to visualize the EDTA aggregate. (**B**) Snapshot of an MD trajectory showing an $EDTA^{4-}$-Na^{+} aggregate; the snapshot was taken at 500 ns (see arrow in panel A); $EDTA^{4-}$: green: C atoms, blue: N atoms, red: O atoms; magenta: Na^{+} ions. Box: This part of the image was zoomed to get a better view on the bilayer within the aggregate. For clarity, some EDTA molecules in front of the bilayer were discarded. (**C**) Zoomed view on an $EDTA^{4-}$-Na^{+} aggregate showing a bilayer formation. Figure taken from ref. [1].

Next, I investigated whether this aggregation occurs because of the applied force field in the simulation. To investigate this force field dependence, the previously described set up was used, but now applying force field parameters for ions from the work of Joung and Cheatham [186]. Compared to the previously used parameters, from Åqvist [187], the parameters from Joung and Cheatham underwent an improvement with respect to the hydration behavior of ions [186]. The measured SASA of the $EDTA^{4-}$ molecules decreases from initially ~6000 $Å^2$ by ~40% to ~3700 $Å^2$ indicating $EDTA^{4-}$ aggregation (**Figure 14**). Thus, an influence of the force field parameter upon aggregation of $EDTA^{4-}$ molecules can be ruled out.

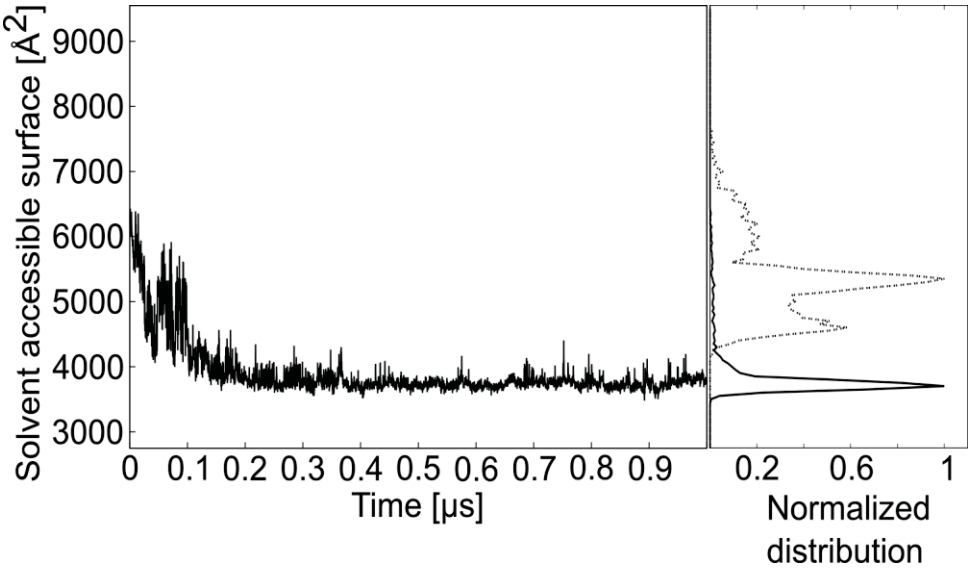

Figure 14: Dependence of EDTA aggregation on the force field parameters for metal ions. SASA and frequency distribution of the SASA over all 20 $EDTA^{4-}$ molecules in the simulation box (solid line). Here parameters for metal ions from Joung *et al.* [169] were applied. For comparison, the results obtained with the parameters for metal ions from the ions94 library of Amber12 [188] are shown with a dashed line in the frequency distribution (**Figure 13A**). Figure taken from ref. [1].

In the initial MD simulations I used 20 EDTA molecules, which results in a higher concentration compared with the experiments. In order to mimic the conditions of the experiments I reduced the amount of EDTA molecules in the system by 20%, thus ending up with 16 $EDTA^{4-}$ or $EDTA^{3-}$ in the simulation box.

The measured SASA of $EDTA^{4-}$ decreases from initially ~8000 $Å^2$ by ~40% to ~4800 $Å^2$ with a peak on the frequency distribution at 5400 $Å^2$ (**Figure 15**), whereas the SASA of $EDTA^{3-}$ molecules remained > 5500 $Å^2$ throughout the MD simulations, resulting in a distribution function that peaks at 6900 $Å^2$ (**Figure 15**). Thus, an aggregation purely based upon the initially increased concentration, compared to the experimental conditions, can be ruled out.

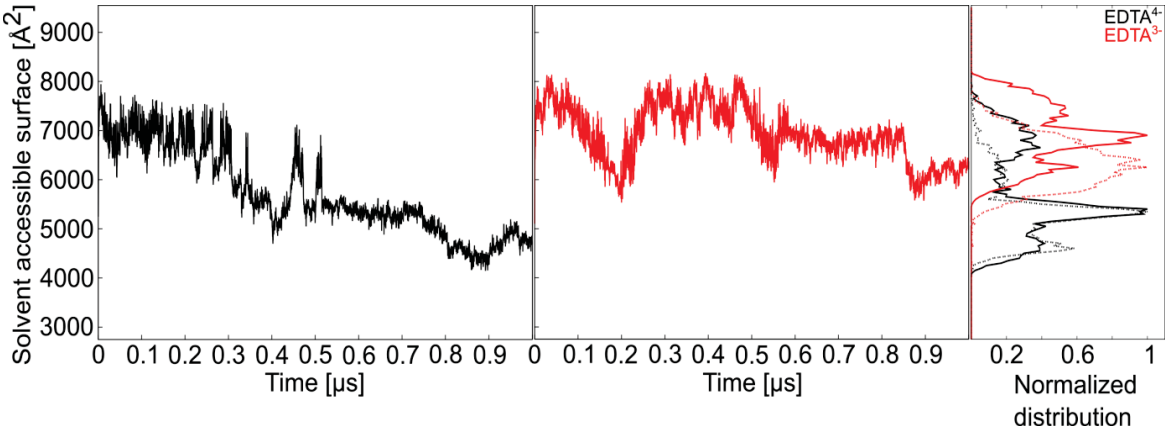

Figure 15: Dependence of EDTA aggregation on the EDTA concentration. SASA and frequency distribution of the SASA over all 16 EDTA molecules in the simulation box; $EDTA^{4-}$ (black), $EDTA^{3-}$ (red). For comparison, the results obtained with the original concentration are shown with a dashed line (**Figure 13A**). Figure taken from ref. [1].

During the experiments and the previous MD simulations, besides EDTA and SYPRO Orange, the only common components (except water) were sodium ions. To investigate if the presence of sodium ions is a prerequisite for the aggregation of EDTA I set up MD simulations with 20 $EDTA^{4-}$ molecules and two SYPRO Orange molecules as before, but in this case the sodium counter ions were not added to the system; instead, a uniform neutralizing plasma was applied [155]. In contrast to the previous MD simulations, the measured SASA of $EDTA^{4-}$ does not decrease and remains at ~ 9200 $Å^2$ during the MD simulations (**Figure 16**). This result indicates that the presence of sodium ions is required for an $EDTA^{4-}$ aggregation.

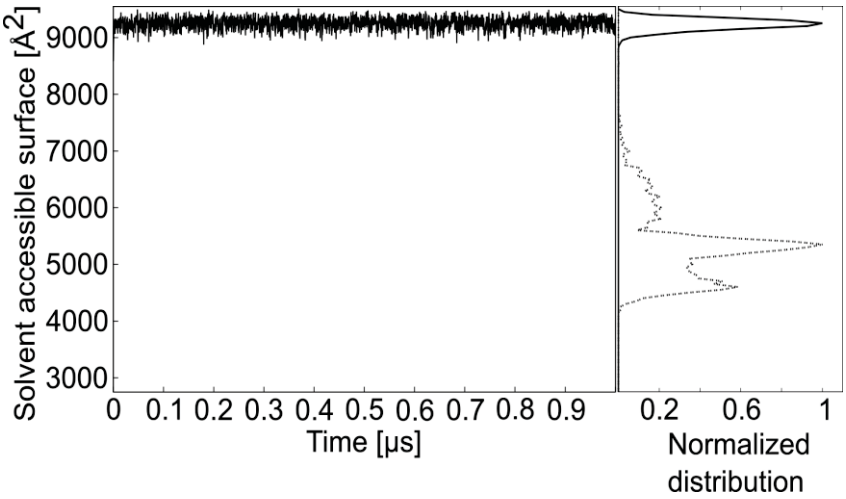

Figure 16: EDTA aggregation does not occur in the absence of Na^+ ions. SASA and frequency distribution of the SASA over all 20 $EDTA^{4-}$ molecules in the simulation box in the absence of Na^+ counter ions. The dashed line in the frequency distribution has been added for comparison (**Figure 13A**) and shows the results obtained in the presence of the counter ions. Figure taken from ref. [1].

As our experiments have shown, the effects of EDTA upon SYPRO Orange were impeded in the presence of Ca^{2+}; hence, the effects of these ions were investigated in a series of MD simulations. Therefore, I created 20 $EDTA^{4-}$-Ca^{2+} complexes with an octahedral coordination of Ca^{2+} [182] in the simulation box and performed MD simulations of 1 µs length. To keep these simulations comparable to our initial set up I added 80 sodium ions, this however, made the application of a uniform neutralizing plasma [155] necessary. The measured SASA of $EDTA^{4-}$ in the presence of Ca^{2+} starts at ~6100 $Å^2$ due to the compact octahedral shape of the $EDTA^{4-}$-Ca^{2+} complexes (**Figure 17A**). In the course of the simulations, the measured SASA decreases to ~4500 $Å^2$, resulting in a decrease of ~25%, only ~60% relative to the decrease observed for $EDTA^{4-}$-Na^{+} without Ca^{2+}. Noteworthy, the $EDTA^{4-}$:Ca^{2+} ratio, each 125 mM, is equal to the experimental conditions at the experimental determined EC_{50} = 100 mM Ca^{2+} (**Figure 9C**). Thus, only about half of the EDTA molecules were expected to form $EDTA^{4-}$-Ca^{2+} complexes in our MD simulations. Strong electrostatic and polarization interactions with its surroundings due to divalent nature of Ca^{2+} ions makes an accurate descriptions of these ions complicated [189]. Still, these data indicate that complex formation with Ca^{2+} can strongly reduce aggregate formation of $EDTA^{4-}$, depending on the molar ratio of $EDTA^{4-}$:Ca^{2+}.

Additionally, visual inspection of the MD trajectory did not reveal a bilayer-like configuration of the $EDTA^{4-}$ molecules but rather a fluffier configuration (**Figure 17B**).

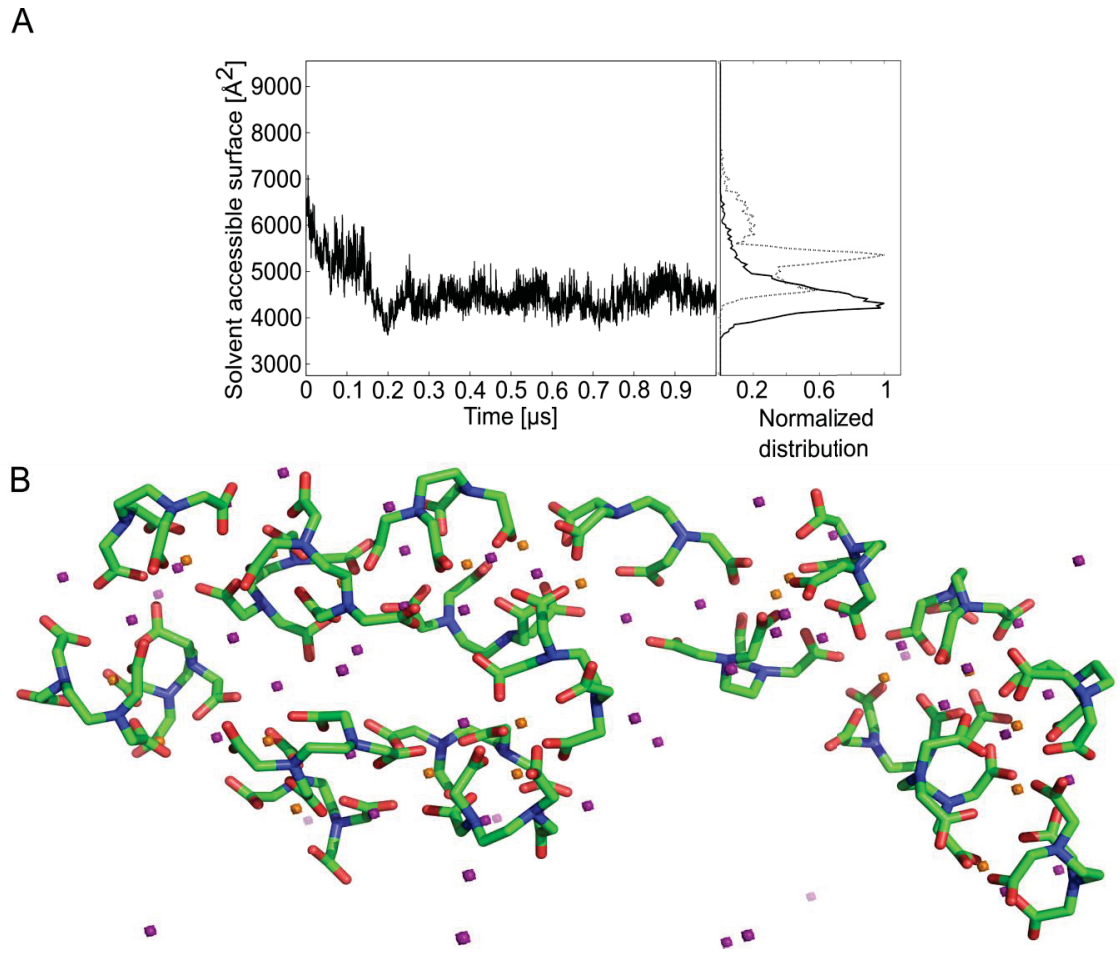

Figure 17: EDTA aggregation in the presence of Ca^{2+}. (**A**) SASA and frequency distribution of the SASA over all 20 $EDTA^{4-}$ molecules in the simulation box. The dashed line in the frequency distribution has been added for comparison (**Figure 13A**) and shows the results obtained without calcium ions but in the presence of Na^{+}. (**B**) Snapshot of an MD trajectory showing an $EDTA^{4-}/Ca^{2+}$ – Na^{+} aggregate; the snapshot was taken at 500 ns; $EDTA^{4-}$: green: C atoms, blue: N atoms, red: O atoms; magenta: Na^{+} ions; orange: Ca^{2+} ions. Figure taken and adapted from ref. [1].

The measured change in fluorescence in my TSA experiments occurred at higher temperatures (**Figure 8**). In order to investigate the influence of the temperature on the aggregation of EDTA, MD simulations with the same set up as the initial simulations were prepared, but this time the temperature was raised to 333.15 K (60 °C) and 353.15 K (80 °C), which was 33.15 and 53.15 K higher as in the previous simulations. Compared to the TSA experiments the lower temperature in the MD simulation set up (333.15 K) is 8 K lower than the "denaturation temperature" of ~68 °C (= 341.15 K) . In contrast, the higher temperature in the MD simulation set up (353.15 K) is 12 K higher. At 333.15 K, the measured SASA of $EDTA^{4-}$ molecules decreases from initially ~6000 $Å^2$ by ~50% to ~3100 $Å^2$, whereas the SASA of $EDTA^{3-}$ molecules fluctuates around ~5500 $Å^2$ (**Figure**

18A). The results from the MD simulations at 353.15 K are very similar (**Figure 18B**). The frequency distributions of SASA for $EDTA^{4-}$ peak at ~3750 $Å^2$ (333.15 K) and ~3200 $Å^2$ (353.15 K), > 1600 $Å^2$ lower than in the case of the MD simulations at 300 K (**Figure 18A, B**). Similar to previous results, the formation of an $EDTA^{4-}$-Na^+ aggregate led to a bilayer-like structure (**Figure 18C**). However, compared to the results gathered at lower temperatures (300 K), the aggregate appeared more compact at 353.15 K. Noteworthy, higher temperatures could have adverse effects on the TIP3P water model [190, 191], thus, making it more difficult to reproduce experimental values. Still, the gathered information indicates that $EDTA^{4-}$ aggregation is promoted by increasing temperatures, without influencing the behavior of $EDTA^{3-}$ in solution.

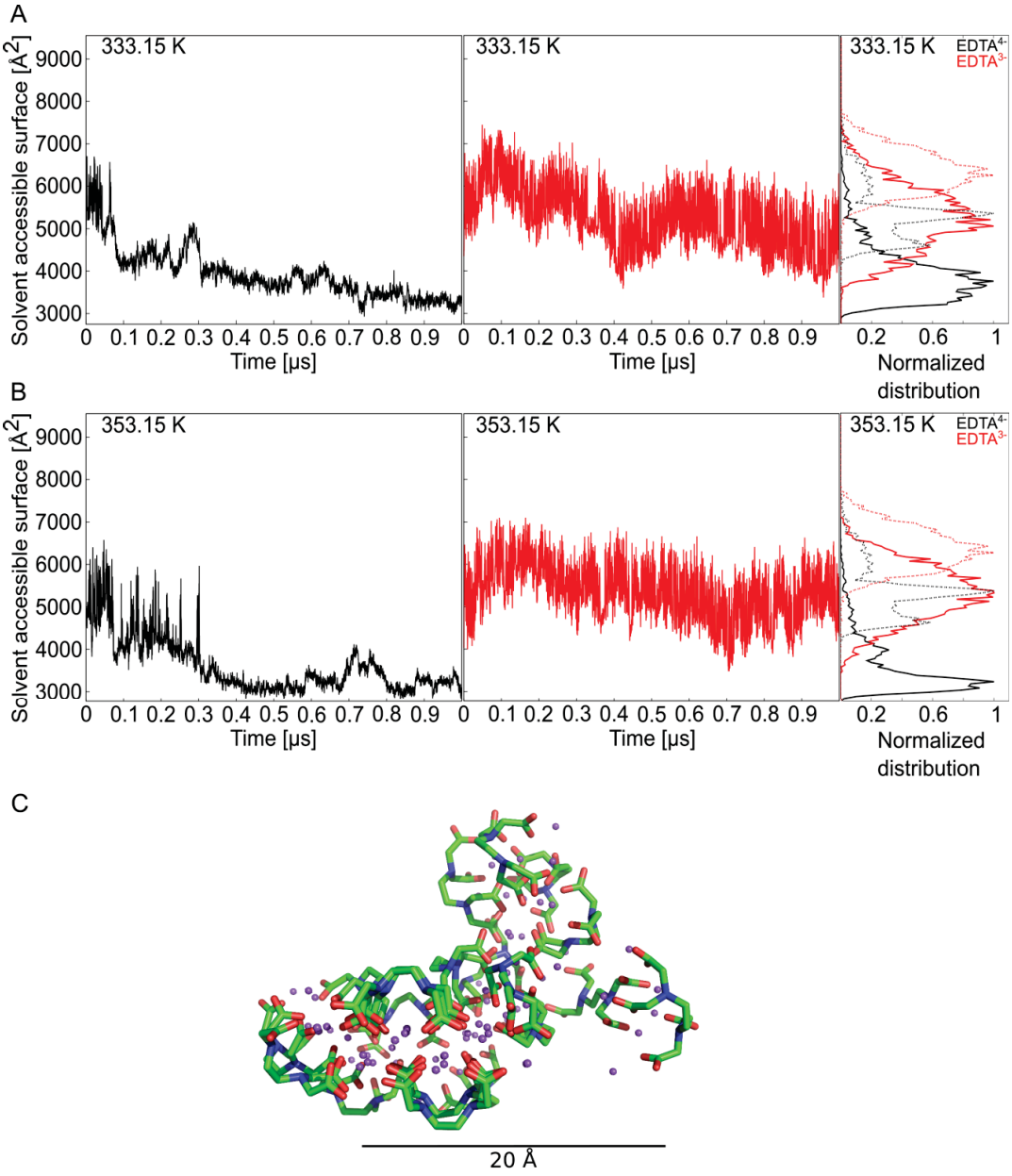

Figure 18: Temperature dependence of EDTA aggregation. (**A**) SASA and frequency distribution of the SASA over all 20 EDTA molecules in the simulation box at 333.15 K; $EDTA^{4-}$ (black), $EDTA^{3-}$ (red). The dashed line in the frequency distribution has been added for comparison (**Figure 13A**) and shows the results obtained at 300 K (**B**) SASA and frequency distribution of the SASA over all 20 EDTA molecules in the simulation box at 353.15 K; $EDTA^{4-}$ (black), $EDTA^{3-}$ (red). The dashed lines in the frequency distribution have been added for comparison (**Figure 13A**) and show the results obtained at 300 K. (**C**) Snapshot of an MD trajectory showing an $EDTA^{4-}$ – Na^{+} aggregate; the snapshot was taken from the MD simulations performed at 353.15 K at 500 ns; $EDTA^{4-}$: green: C atoms, blue: N atoms, red: O atoms; magenta: Na^{+} ions. Figure taken and adapted from ref. [1].

Lastly, to get a further insight into the structural composition of the EDTA aggregate under the different investigated conditions, the radial distribution function $g(r)$ of carbon atoms from the ethylene moiety of EDTA was calculated. The intramolecular distances were ignored. $g(r)$ describes how density varies as a function of the distance r from a reference particle. In the case of $EDTA^{4-}$ in the presence of Na^+ at 300 K, $g(r)$ shows two prominent peaks at ~ 6 Å and ~11 Å (**Figure 19**). The first peak relates to the carbon atoms of ethylene groups of two neighboring EDTA molecules, whereas the second peak relates to the carbon atoms of ethylene groups of opposing EDTA molecules (on different sides of the central sodium layer; **Figure 13B, C**). These results indicate that $EDTA^{4-}$ molecules are mostly found within the bilayer structure in the presence of Na^+.

At higher temperatures (353.15 K) the determined $g(r)$ showed more pronounced peaks at ~ 6 Å and ~11 Å (**Figure 19**), supporting the idea that higher temperatures foster the aggregation of EDTA. In the presence of Ca^{2+} only the latter peak at ~11 Å remained prominent (**Figure 19**). The missing peak at 6 Å indicates that the ethylene moieties of $EDTA^{4-}$ do not come close to each other, in line with the fluffier configuration exemplarily shown in **Figure 17B**. With no Na^+ ions present in the system, a broad peak can be seen between 15 and 20 Å (**Figure 19**), thus, showing the lack of a close-ranged, ordered structure. In summary, these results are the first detailed insight on the atomic level into the $EDTA^{4-}$-Na^+ aggregation and provide a possible structural model of such an aggregate.

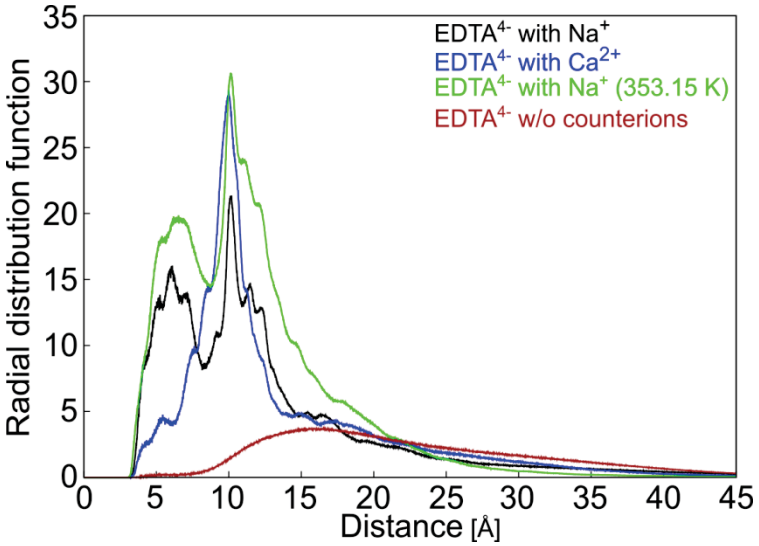

Figure 19: Analysis of the distribution of EDTA molecules under various conditions. Radial distribution function of carbon atoms of the $EDTA^{4-}$ ethylene moiety, ignoring intramolecular distances, at 300 K in the presence of Na^+ (black), in the presence of Ca^{2+} and Na^+ (blue),or in the absence of counterions (brown), and at 353.15 K in the presence of Na^+ (green). Figure taken from ref. [1].

3.5.3 Structural model of the interaction between EDTA and SYPRO Orange

The aggregation of EDTA itself is not sufficient to trigger an enhancement of the fluorescence of SYPRO Orange. In order for this interference to occur the aggregate has to interact with SYPRO Orange. For this reason, the initial MD simulation set up was used to determine and quantify interactions between EDTA and SYPRO Orange. To quantify these interactions the distance distribution between the center of mass of the ethylene groups of $EDTA^{4-}$ or $EDTA^{3-}$ molecules and the center of mass of the SYPRO Orange "core region" (for a definition of the "core region", see **Figure 20**) was calculated and the number of $EDTA^{4-}$ or $EDTA^{3-}$ molecules around SYPRO Orange (termed "EDTA shell" hereafter) was determined.

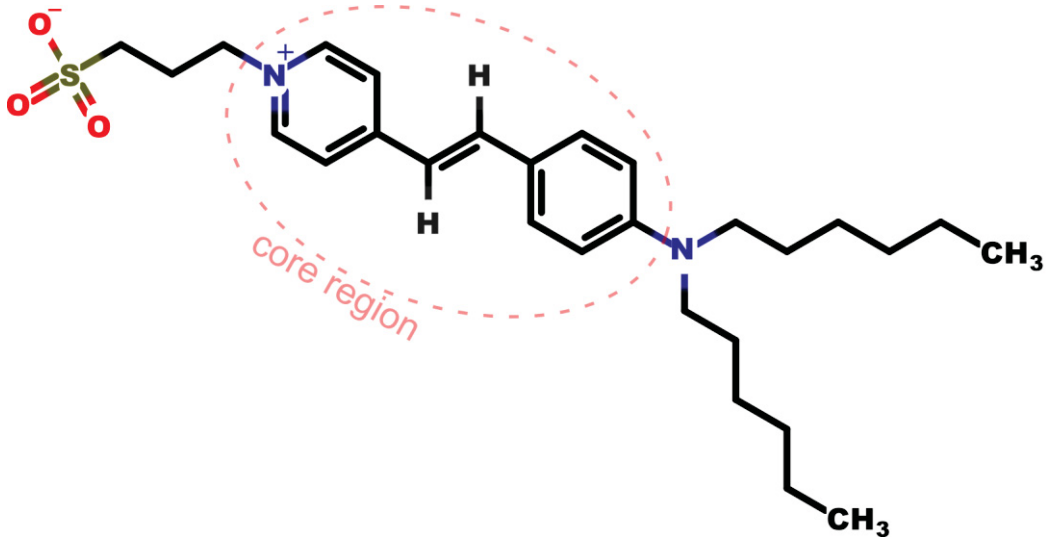

Figure 20: 2D structure of SYPRO Orange. Starting from the published structures (303/304) [185], the length of the alkyl chains and the configuration at the double bond were assessed by mass spectrometry and NMR, respectively (**Figure 11** and **Figure 12**), here we defined the core region of SYPRO Orange as the aromatic rings and the central double bond. Figure taken from ref. [1].

For an EDTA molecule to be part of the EDTA shell around SYPRO Orange the distance between any atom of the SYPRO Orange and the EDTA ethylene group had to be ≤ 5 Å. After 200 ns, an EDTA shell is established consisting of ~4 $EDTA^{4-}$ molecules (**Figure 21A**). The resulting SYPRO Orange:$EDTA^{4-}$ ratio of 1:4 is in a good agreement with SYPRO Orange:$EDTA^{4-}$ of 1:3.6 at the EC_{50} from the experimental investigations. For $EDTA^{3-}$ no formation of an EDTA shell was measured. At 353.15 K, the established EDTA shell consists of ~7 $EDTA^{4-}$ molecules (**Figure 21B**). The increased amount of $EDTA^{4-}$ molecules in the shell around SYPRO Orange can be attributed to the increased

compactness of the EDTA aggregate at higher temperatures (**Figure 18C**). Similar to the MD simulations at 300 K, no formation of an EDTA shell was measured for $EDTA^{3-}$ at 353.15 K. In the presence of Ca^{2+}, an EDTA shell is established consisting of ~4 $EDTA^{4-}$ molecules forms briefly after 300 ns, from which point the number of EDTA4 molecules in the shell decreases continuously (**Figure 21C**).

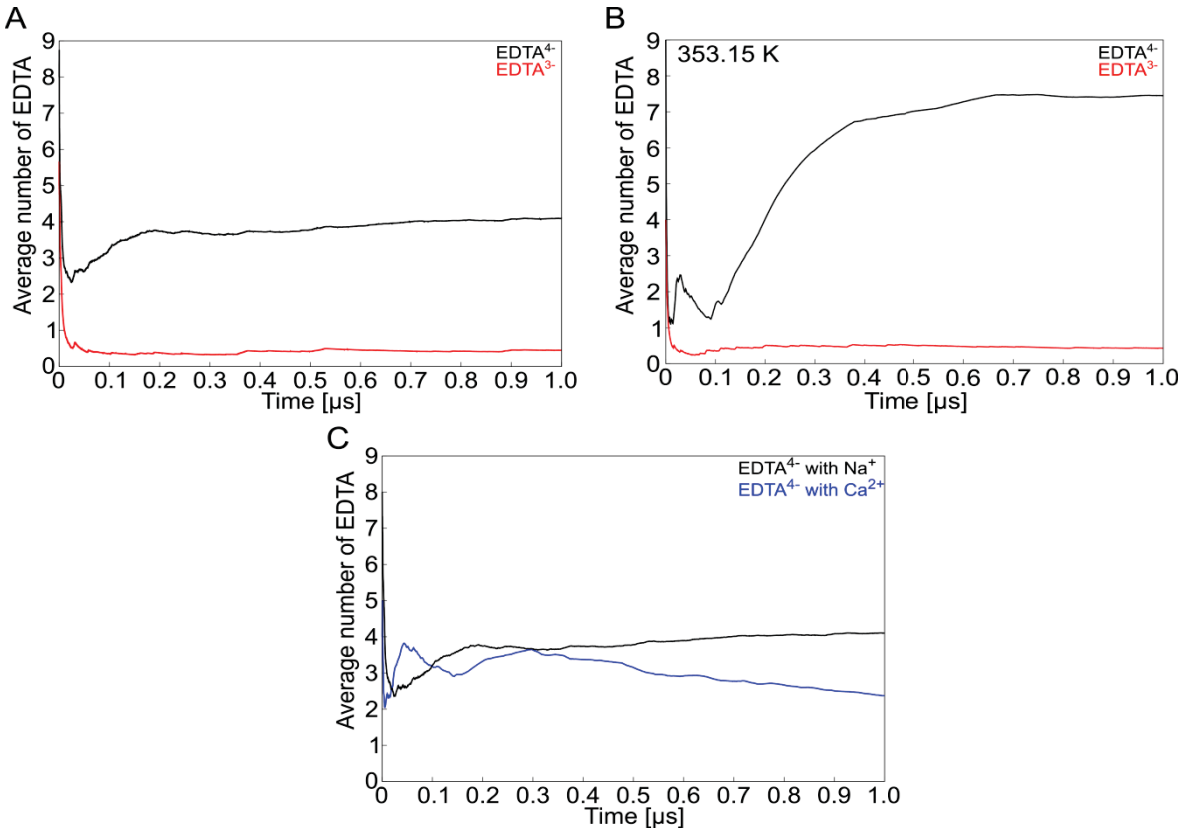

Figure 21: Formation of an EDTA shell around SYPRO Orange under various conditions. (**A**) Average number of EDTA molecules within 5 Å around SYPRO Orange for $EDTA^{4-}$ (black) and $EDTA^{3-}$ (red) at 300 K. (**B**) Average number of EDTA molecules within 5 Å around SYPRO Orange for $EDTA^{4-}$ (black) and $EDTA^{3-}$ (red) at 353.15 K. (**C**) Average number of EDTA molecules within 5 Å around SYPRO Orange for $EDTA^{4-}$ in the presence of Na^{+} (black) and in the presence of Ca^{2+} and Na^{+} (blue). The core region is defined as the two aromatic rings with the linker (**Figure 20**). Figure taken and adapted from ref. [1].

Looking at the distance distribution in more detail, one can see two major groups of $EDTA^{4-}$: The first group (55% of all $EDTA^{4-}$ molecules) had a peak in the distribution at ~15 Å; whereas the remaining 45% had a peak at ~35 Å (**Figure 22A**). For $EDTA^{3-}$, only one group can be found with a peak in the distance distribution at ~33 Å (**Figure 22B**). A list of all frequency distributions for each EDTA can be found in **Figure 35** (**Chapter 6.3**). In the absence of any interactions between SYPRO Orange and EDTA, the average distance between the center of mass of the SYPRO Orange core region and the center of

mass of the EDTA ethylene group is expected to be ~ 28 Å, equivalent to half of the size of the simulation box.

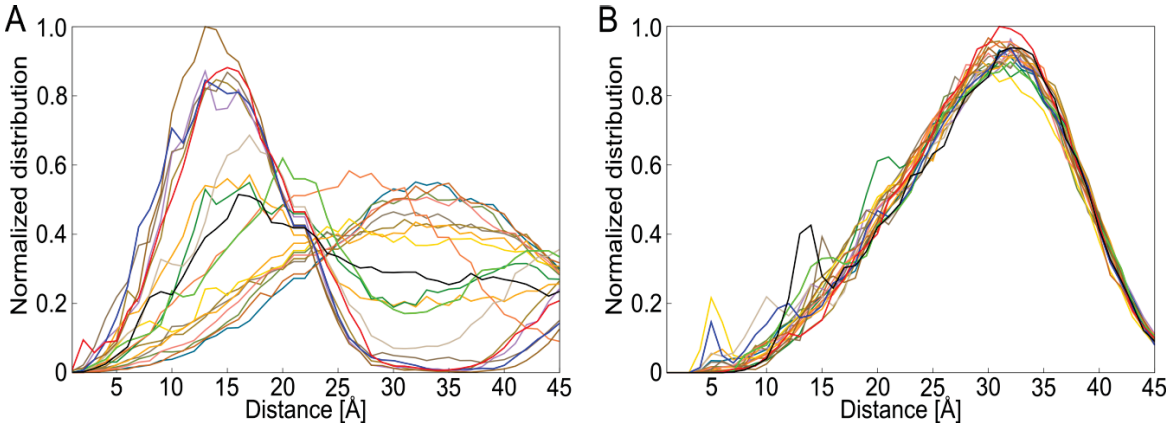

Figure 22: Analysis of the distance between EDTA molecules and SYPRO Orange. (**A**) Frequency distribution of the distance of the center of the ethylene group of every $EDTA^{4-}$ molecule to the center of mass of the core region of SYPRO Orange (**Figure 20**). (**B**) Frequency distribution of the distance of the center of the ethylene group of every $EDTA^{3-}$ molecule to the center of mass of the core region of SYPRO Orange (**Figure 20**). The force field ff12SB was applied in both MD simulations. Figure taken and adapted from ref. [1].

Thus, the presented results indicate $EDTA^{4-}$ interacts with SYPRO Orange, but $EDTA^{3-}$ does not form interactions with SYPRO Orange. SYPRO Orange is neutral at pH > 10, thus, making different electrostatic interactions between the EDTA species and SYPRO Orange implausible. The explanation for this differential behavior must originate from the formation of the $EDTA^{4-}$ aggregates. An example for how SYPRO Orange interacts with an $EDTA^{4-}$ aggregate is depicted in **Figure 23A**. Here, the surface of the aggregate and SYPRO Orange were colored according to the local partial atomic charges. SYPRO Orange binds with its core region to the non-polar regions of the aggregate formed by the ethylene moieties of $EDTA^{4-}$; these non-polar interactions are expanded by the propyl and hexyl chains of the dye. A very similar binding mode is revealed by the exemplary aggregate configuration from the MD simulations at 353.15 K (**Figure 23B**). In summary, the presented results provide evidence that indicate the formation of a bilayer-like configuration within the $EDTA^{4-}$-Na^{+} aggregates and the interaction between these aggregates and SYPRO Orange via the non-polar regions of the aggregates (**Figure 23C**).

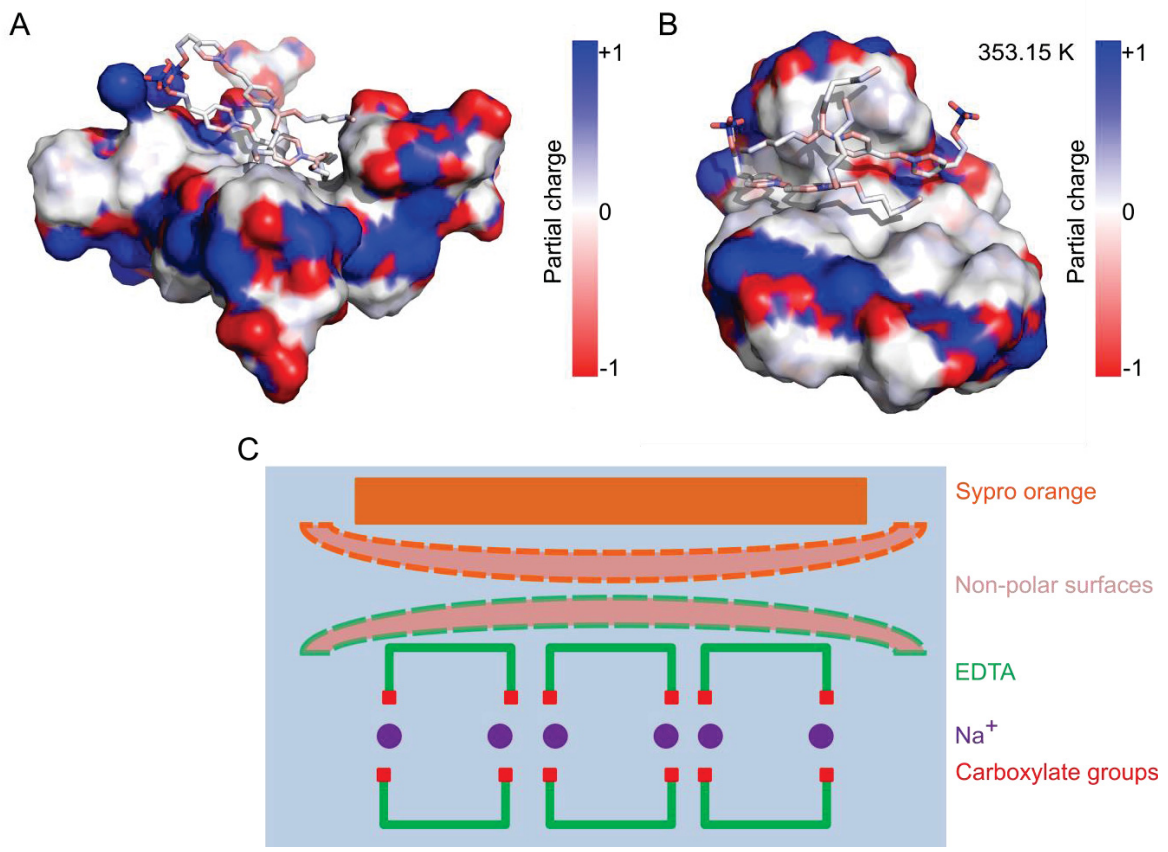

Figure 23: Structural model from MD simulations of EDTA aggregates and SYPRO Orange forming a complex with the hydrophobic region of EDTA.(**A**) Snapshot of an MD trajectory showing the $EDTA^{4-}$-Na^{+} aggregate from **Figure 13** (surface representation) to which two SYPRO Orange molecules bind non-covalently. The coloring of the surface is according to the local partial atomic charge. (**B**) Snapshot of an MD trajectory showing the $EDTA^{4-}$ – Na^{+} aggregate (353.15 K) from **Figure 18** (surface representation) to which two SYPRO Orange molecules non-covalently bind. The coloring of the surface is according to the local partial atomic charge. (**C**) Schematic figure showing the association of SYPRO Orange with an $EDTA^{4-}$-Na^{+} aggregate. The two non-polar surfaces of the aggregate and SYPRO Orange are highlighted by dashed borders. Figure taken and adapted from ref. [1].

3.6 Utilizing new information in thermal shift assay experiments

Despite identifying the cause of the fluorescence signal in the absence of a protein and uncovering the molecular mechanism behind this interference, a protein-specific fluorescence signal could not be measured at pH ~7.4. Thus, the experiments were performed in cooperation with Lukas Zubek and Neofitos Makridis (Students in the Department of Pharmacy, Heinrich Heine University Düsseldorf). The goal of this practical was to find a substance, that is commonly used in biochemical or biophysical experiments and that could lead to the occurrence of a NHR2-specific fluorescence signal. Eight substances were chosen for these experiments: Acetate, glycine, imidazole, HEPES,

MES, Tris and citrate, as all of these substances are very commonly used in biochemical laboratories. The substances were separately added to the initial TSA samples mixture. In most cases no increase in relative fluorescence intensity could be observed (**Figure 24A**). Despite the presence of protein and EDTA at pH ~7.4 no increase of fluorescence was detected, suggesting that these newly introduced buffer additives might impede the interactions between SYPRO Orange and the hydrophobic amino acids of the protein. However, since no improvements could be made, these approaches were not followed up. Adding Tris (100 mM) or citrate (200 mM) resulted in the detection of typical TSA curves [178, 179] (**Figure 24B**). An increase in the fluorescence intensity was only detected in the presence of protein, thus, suggesting that these components prevent a premature aggregation of unfolded NHR2 allowing SYPRO Orange to interact with the hydrophobic amino acids. In order to exclude an influence of Tris or citrate on the thermal stability of NHR2 TSA experiments were performed to determine the T_m of NHR2 in the presence of these buffer additives. The calculation of the change rate of the relative fluorescence and the determination of the global maximum of each curve resulted in melting points for NHR2 in the presence of citrate $T_m = 83 \pm 1.49$ °C (citrate) and NHR2 in presence of Tris $T_m = 80 \pm 1.36$ °C (**Figure 24C**). Compared with previous results for NHR2 from Wichmann *et al.* ($T_m = 85$ °C) [82] I measured a difference to the determined melting point of 2 °C (citrate) and 5 °C (Tris).

In summary, the introduction of citrate or Tris, fairly common buffering agents in biochemical and biophysical laboratories, allow detecting a NHR2-specific fluorescence signal and the determination of the thermal stability of NHR2.

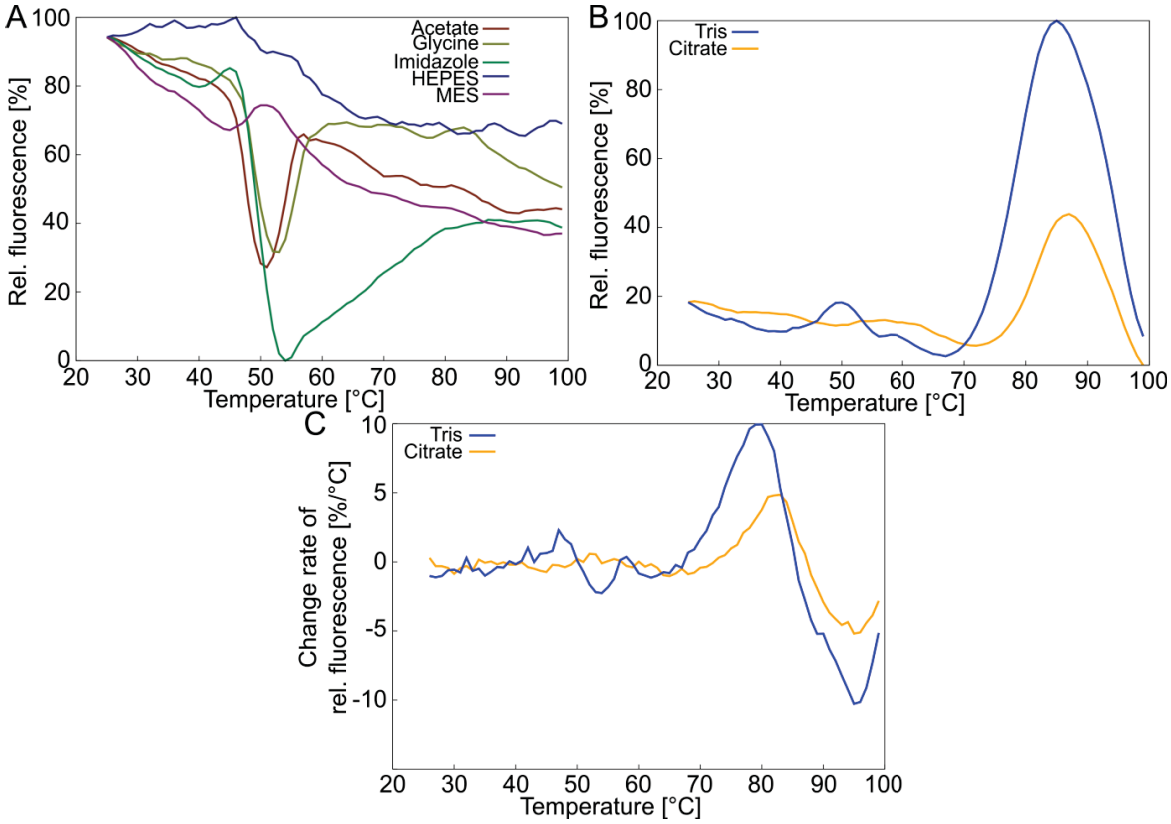

Figure 24: The thermal shift assay signal of NHR2 with different buffer compositions. (**A**) Relative fluorescence signal in the presence of NHR2 and with addition of acetate, glycine, imidazole, HEPES or MES. (**B**) Relative fluorescence signal in the presence of NHR2 and with addition of Tris or citrate. (**C**) Calculated first derivative of the relative fluorescence signal from panel B.

For the following Chapters (**Chapter 3.7 - 3.9 and 3.11**), the text and figures have been taken and modified from the publication:

" **Kroeger, T.**, Frieg, B., Viegas, A., Smits, S., Groth, G., Etzkorn, M., Gohlke, H. *Biophysical characterization of a small-molecule inhibitor of RUNX1/ETO tetramerization with anti-leukemic effects*. **In preparation** 2018."

3.7 **7.44**-mediated interference in NHR2 tetramerization

The stability of NHR2 in the presence of **7.44** was analyzed by TSA [1, 70]. Using this assay the thermally induced protein unfolding can be monitored by the binding of the fluorescent dye SYPRO Orange [192] to the hydrophobic core of a protein that becomes exposed upon unfolding. This interaction leads to a related increase in fluorescence emission. The temperature at the midpoint of the unfolding transition is defined as the melting temperature (T_m) of the protein [48]. Usually, if a ligand interferes with the protein a shift of the T_m occurs [49], which is described as monophasic melting [193]. However, a

previous study showed that the NHR2 tetramer and NHR2 dimer have very different melting temperatures (respectively ~85 °C and ~60 °C) [82]. Thereby, the proportions of the two species should be altered in a dose-dependent manner towards the dimer by a **7.44**-mediated inhibition of the NHR2 tetramerization. In such cases, a biphasic melting behavior should occur, with NHR2 presenting melting points at both temperatures, rather than a monophasic melting behavior with one melting point at an intermediate melting temperature [193]. As the results show, in the presence of **7.44** with a concentrations range of a factor of ~100, a melting point for NHR2 was detected at ~85 °C (**Figure 25**), thus indicating that the monitored melting point relates to the NHR2 tetramer. The intensity of the detected fluorescence associated with the melting of NHR2 decreases with increasing concentration of **7.44** (**Figure 25A**), thus, indicating that the portion of NHR2 tetramer in the equilibrium diminishes. Regarding the melting of the NHR2 dimer at ~60 °C, an increased fluorescence intensity was observed with increasing **7.44** concentration at this temperature, indicating that the portion of NHR2 dimer in the equilibrium rises. However, it was not possible to determine a clear melting point due to an overlay of the NHR2 melting-induced fluorescence signal with a strong fluorescence signal that decreases with increasing temperature. This fluorescence signal also roughly increases with the applied **7.44** concentration. This signal can be explained in that the NHR2 dimer exposes its hydrophobic PPI interface [82] to which SYPRO Orange is able to bind. The decrease of the fluorescence signal > 90 °C can be explained by a post-peak aggregation of denaturated protein as well as protein:SYPRO Orange complexes [194, 195]. Due to the lack of a clear melting point at ~60 °C, a quantitative analysis of the gathered data was not possible [193]. Therefore, information regarding the maxima of the fluorescence derivatives (dF/dT) at ~85 °C (**Figure 36**) were used as an indicator representing how much NHR2 tetramer is present [193]. Using the Ricketts and Head equation with two "slopes" [196] for fitting dF/dT versus the concentration of **7.44** (**Figure 25B**) revealed an EC_{50} = 1.09 ± 0.09 mM for the effect of **7.44** titrated to a constant concentration of NHR2 of 0.2 mg/ml. The applied protein concentration was necessary, since at lower NHR2 concentrations the signal intensity was too weak and the background signal impaired the analysis.

Noteworthy, the determined EC_{50} value contains information regarding the dissociation of the NHR2 tetramer and the association of **7.44** with NHR2. Ultimately, the results show that **7.44** interferes with the NHR2 tetramerization and thus leading to a reduced concentration of the NHR2 tetramer with increasing **7.44** concentration.

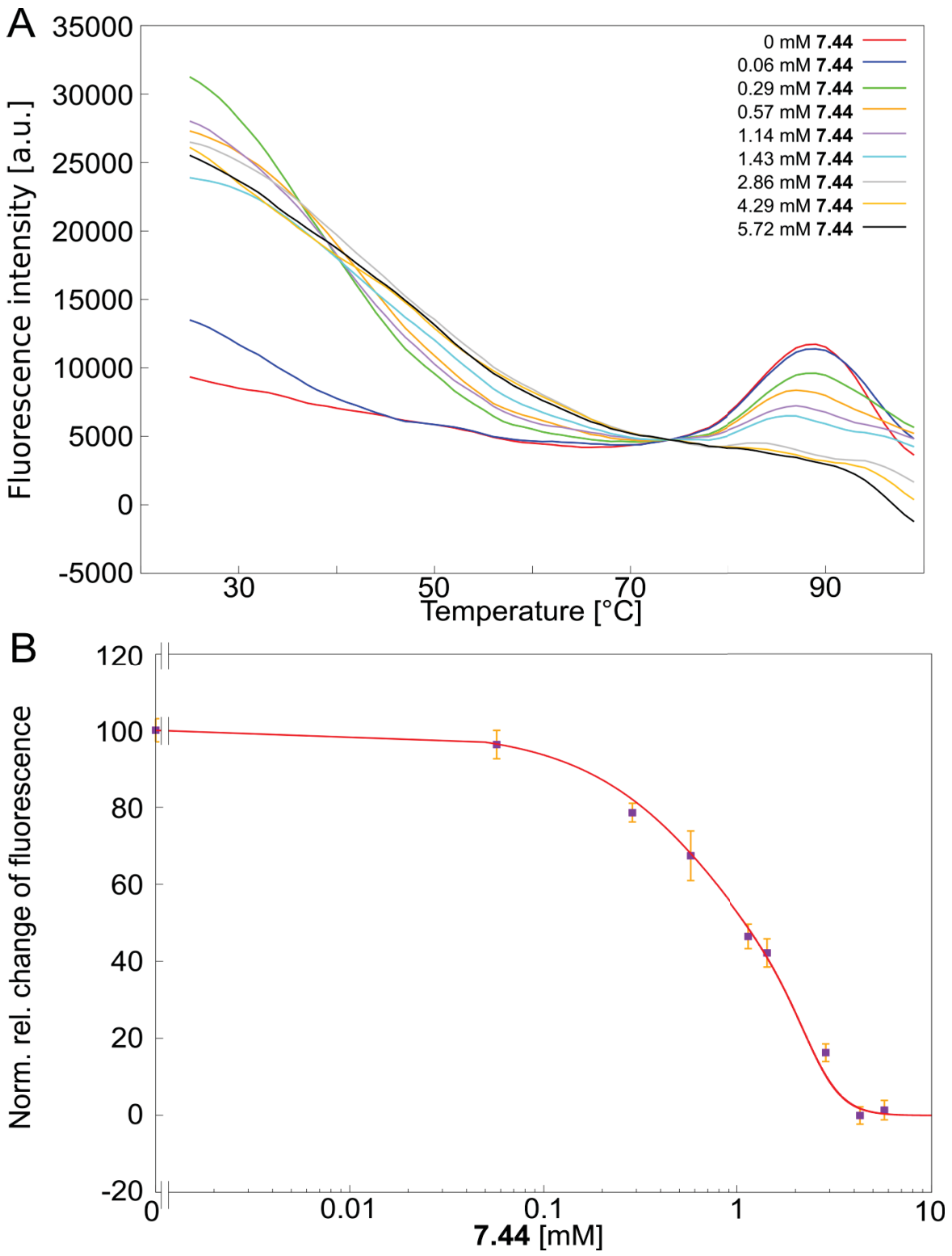

Figure 25: Validation of the interaction between 7.44 and NHR2 with TSA. (**A**) Fluorescence intensity in the presence of **7.44** at different concentrations; a.u. = arbitrary units; for better visualization, the curves were shifted along the y-axis to match at the point of strongest convergence. Note that this shifting does not change d*F*/d*T* values (**Figure 36**). (**B**) Effects of a titration of **7.44** to a constant concentration of NHR2 on the fluorescence derivative d*F* / d*T*. The error bars show the SEM for $N = 16$. Figure taken from ref. [2].

3.8 Microscale thermophoresis measurements reveal the dissociation constants of the NHR2 tetramer, and of **7.44** with respect to the NHR2 dimer

Until now, the dissociation constant of the NHR2 tetramer into NHR2 dimers, K_{tet}, could not be determined and, thus the dissociation constant of **7.44**, K_{lig}, remained unknown. The NHR2 tetramer (T) is in equilibrium with its dimer (D). The tetramerization inhibitor **7.44** (L) can only bind to NHR2 in its dimeric state (D), as depicted in Eq. 3. The expected binding sites of **7.44** are far apart from each other [54], thus suggesting that the K_{lig} of the first and second ligand binding event are identical. This assumption is in line with the predicted binding modes of **7.44** (**Chapter 3.11**).

$$T \overset{K_{tet}}{\rightleftharpoons} D \overset{K_{lig}}{\rightleftharpoons} DL \overset{K_{lig}}{\rightleftharpoons} DLL \qquad \text{(Eq. 3)}$$

The aim of the microscale thermophoresis measurements was the determination of the K_{lig}. This method detects the directed movements of molecules along a generated temperature difference (thermophoresis). The movements depend on the size, charge, and solvation shell of the molecules [135]. Here, the K_{lig} as a function of K_{tet} and EC_{50} of **7.44** binding to NHR2 was determined according to Eq. 4a-15 [55, 125] (**Chapter 6.4**), because the MST signal contains both the dissociation of the NHR2 tetramer and the association of **7.44** with the NHR2 dimer if **7.44** is titrated to NHR2. For both the measurements of K_{tet} and EC_{50}, varying concentrations of unlabeled NHR2 or **7.44** were titrated to a constant concentration of fluorescently labeled NHR2. With increasing concentration of the titrant, an increasing fraction of labeled NHR2 will form complexes with the titrant until all labeled NHR2 will participate in these $protein_{label}$-titrant complexes. In order to determine K_{tet}, the MST was recorded for labeled NHR2 with a concentration of 50 nM in the presence of a varying concentration of unlabeled NHR2 ranging from 337.5 µM to 10.3 nM. A non-linear regression curve was fit to the recorded MST values with the K_D Fit equation (Eq. 1 (**Chapter 2.2.9**), **Figure 26A**). The inflection point of the curve revealed a K_{tet} of 11.3 ± 1.81 µM. Next, the MST was recorded for 50 nM labeled NHR2 in the presence of varying concentrations of **7.44** ranging from 2.5 mM to 76.29 nM. Again, a non-linear regression curve was fit to the recorded MST values (Eq. 2 (**Chapter 2.2.9**), **Figure 26B**). The inflection point of the curve revealed an EC_{50} of 6.65 µM ± 0.717 µM, which demonstrates the binding of **7.44** to NHR2. The determination of the EC_{50} revealed an average Hill coefficient of $n = 1.5$ for **7.44** binding to NHR2. Using Eq. 14 and Eq. 15 (**Chapter 6.4**), a K_{lig} of 34.3 ± 3.76 µM was then calculated for **7.44** binding to the NHR2 dimer. In summary, to my knowledge this is the first time that the equilibrium constant for

the dissociation of the NHR2 tetramer to NHR2 dimers and the K_{lig} for **7.44** binding to the NHR2 dimer were determined.

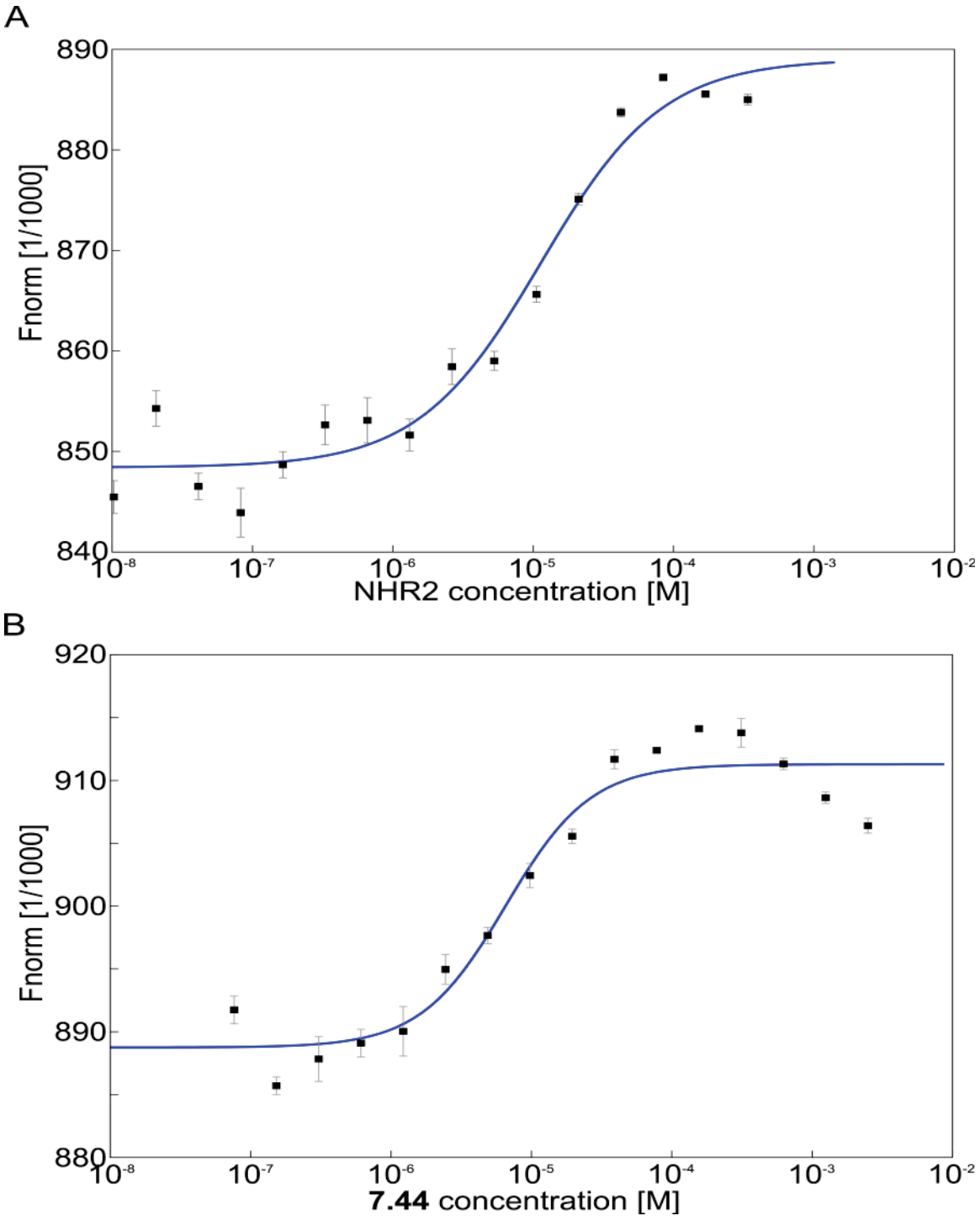

Figure 26: MST analysis of NHR2 tetramer dissociation and 7.44 binding to NHR2. (**A**) Titration of unlabeled NHR2 to a constant concentration of labeled NHR2 induces a change in thermophoresis. n = 5 for each concentration; the error bars show the SEM. K_{tet} = 11.3 ± 1.81 µM was calculated using Eq. 1 (**Chapter 2.2.9**). (**B**) Titration of **7.44** to a constant concentration of labeled NHR2 induces a change in thermophoresis. n = 3; the error bars show the SEM. EC_{50} = 6.65 ± 0.717 µM was calculated for **7.44** binding to NHR2 using Eq. 2 (**Chapter 2.2.9**). Figure taken from ref. [2].

3.9 Saturation transfer difference NMR: Epitope mapping of **7.44**

The STD-NMR was performed by Dr. Aldino Viegas and Dr. Manuel Etzkorn (Institute for Physical Biology, Heinrich Heine University Düsseldorf, Germany). The STD-NMR [138, 139] was conducted in order to provide information regarding the binding pose of **7.44** to the effect that the atoms of the ligand which are closer to the protein will be identified, when an interaction occurs (ligand epitope mapping). A reference ^{1}H spectrum of ligand **7.44** and the corresponding STD-NMR spectrum in the presence of NHR2 and 100 molar excess of ligand was acquired (**Figure 27**). The difference spectrum was generated by subtracting the spectrum containing the saturated protein from the spectrum without saturation [139], thereby it contains only signals of protons from **7.44** that received saturation from the protein depending on their proximity to the protein. Since a signal in the STD-NMR spectrum indicates an interaction between a protein and a ligand, this can be seen as a verification of the interaction between NHR2 and **7.44**, which is in agreement with previous studies [54, 121]. Since not all protons of **7.44** received the same amount of saturation the difference can be used to epitope map the ligand. The determination of the amplification factor (A_{STD}) values, shown in the STD-NMR spectrum and as overlay on the structure of **7.44**, show the highest determined A_{STD} values for the proton H7´´ (100%). The A_{STD} value of proton H7´´ indicates that this proton is the closest proton of **7.44** to NHR2 upon interaction. The remaining protons of the two aromatic rings show relative intensities between 46.2% and 72.7%, thus indicating that both ring systems of **7.44** are involved in the binding to NHR2. Furthermore, the A_{STD} determination revealed the lowest detectable value (29%) for the proton H2. In addition, the proton H3 is not showing up in the STD-NMR spectrum, indicating that this proton is > 5 Å away from the protein upon binding to the protein. Some protons of the buffer also appear in the difference spectra, within the spectral region between 3.1 and 3.6 ppm (marked with *), indicating that the surfactant also interacts with the proteins. However, the surfactant is added to the sample to aid in the stabilization and solubilization of NHR2 and interactions between NHR2 and the surfactant can occur, thus these signals had to be expected. In summary, the STD-NMR results reveal that both heteroaromatic moieties of **7.44** are important for the binding to NHR2.

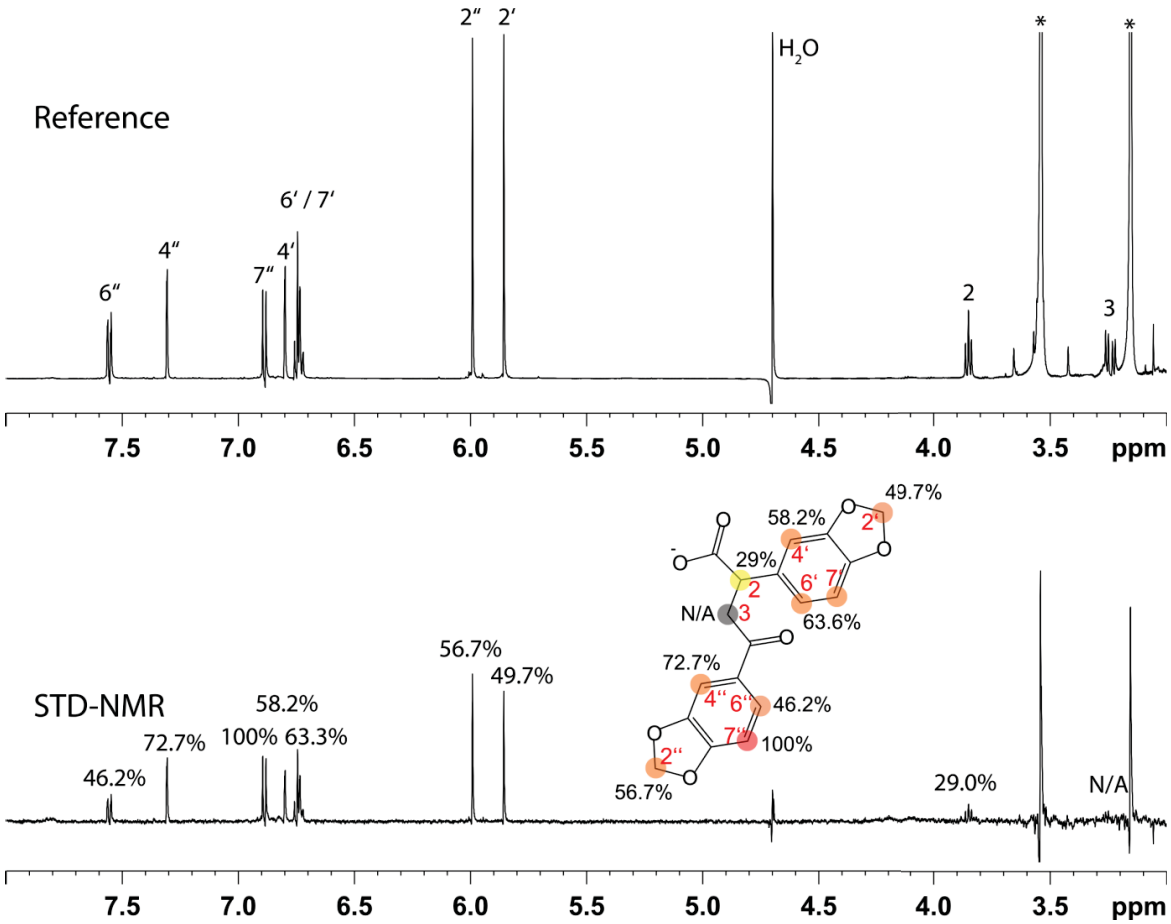

Figure 27: ^{1}H STD-NMR experiment of 7.44 binding to NHR2. (**Top**) Reference spectrum of **7.44**. Assignment of the individual peaks is indicated (see **7.44** structure in bottom panel). Solvent peaks are indicated by *. (**Bottom**) ^{1}H STD-NMR spectrum of ligand **7.44** in the presence of NHR2. Relative intensities are indicated on top of the corresponding peaks and mapped onto the structure of **7.44**. The circles are colored by the relative intensity of the protons, which is correlated with the proximity to the protein: red – 100% (the closest), orange - 46.2% to 72.7%, yellow – 29%, and grey – not detectable (meaning > 5 Å away from the protein). The proton numbering is indicated by the red numbers. Data was acquired at 600 MHz and 298 K, with a 100-fold molar excess of **7.44**. Figure taken from ref. [2].

3.10 NMR experiments to determine the binding mode of **7.44** and NHR2

The NMR spectroscopy of **7.44** and NHR2 was performed by Dr. Aldino Viegas and Dr. Manuel Etzkorn (Institute for Physical Biology, Heinrich Heine University Düsseldorf, Germany). First, we acquired data on the NHR2 protein alone in buffer (experiment: ^{15}N-^{1}H-TROSY-HSQC) (**Figure 28**). Considering the total number of amino acids, we did not see as many signals as expected. This may be due to the fact that it is a tetramer and the total molecular weight was too high. The higher the total molecular weight of a protein the slower is the rotation, i.e. tumbling, of the protein. A decreased rotation leads to a decrease in the T_2 relaxation time (transverse relaxation), thus relaxation

is achieved faster. Consequently, the faster relaxation leads to line broadening [197] and, eventually, to the disappearing of some peaks. For more flexible parts of the protein, because they are not so constrained, they have a different tumbling time then the core of the protein, i.e., they rotate faster. This is why we can see the peaks for the more flexible regions.

The majority of the signals we do see probably belong to the His-tag. On the top right, we could see the signals corresponding to the Asn (x3) and Gln (x2) side chains. The signals were doubled because we were seeing the NH_2 groups, with a horizontal duplication of signals because of the exchange of one of these H with the deuterium in the solvent. ^{2}H itself would result in no signal. However, the chemical environment for a ^{1}H slightly differs depending if a neighboring ^{1}H is replaced with a ^{2}H, thus giving rise to two signals to the ^{15}N resonance [198]. The signal at around 110 ppm (for ^{15}N) was from one (or all) Gly. Since they are only in the tag, this clearly originated from the His-tag.

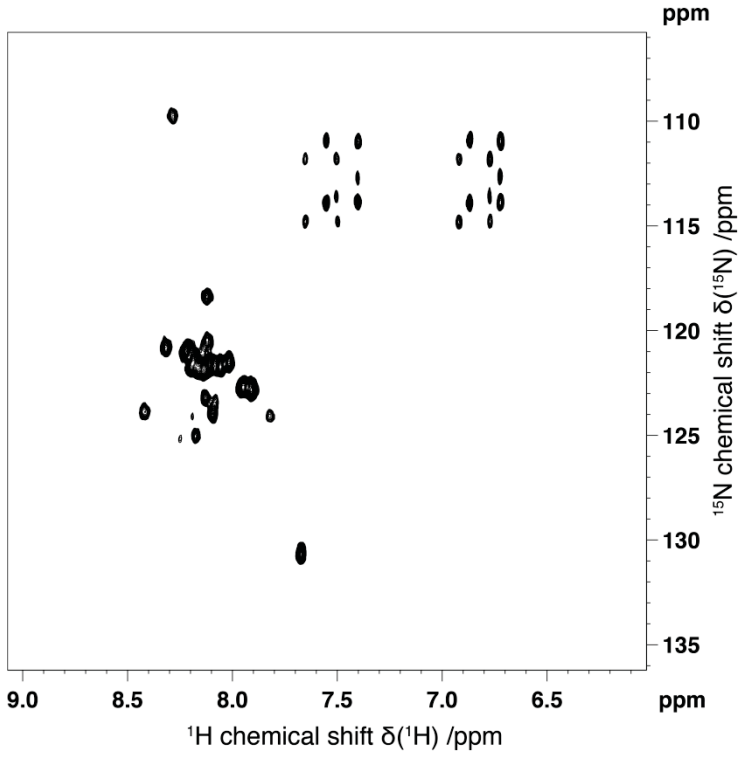

Figure 28: NHR spectrum of ^{15}N-labeled NHR2 in PBS buffer. 200 µM NHR2; PBS buffer, 100 mM NaCl, pH= 7.4, 10% D_2O (v/v); T = 25 ºC; The 2D ^{15}N,^{1}H-TROSY-HSQC spectrum was acquired with 8 transients in a matrix with 2k data points in F2 in a spectral window of 8417.51 Hz, centered at 2816.94 Hz and 128 increments in F1 in a spectral window of 2188.46 Hz, centered at7263.60 Hz.

Next, we repeated the experiment in the presence of DMSO (10% v/v), using the same buffer (red spectrum) (**Figure 29**). We did this because for the STD experiments the ligand was solubilized in DMSO and we needed to see if this has some effect in the protein. DMSO affects the viscosity of the solution; therefore affecting the overall tumbling of the proteins [199]. Furthermore, it can interact transiently and non-specifically with the residues of the protein ,thus changing their chemical shifts [199]. As we can see, the DMSO had some effect on the protein, i.e., we saw some shifts in the signals (**Figure 29**). Since **7.44** was solubilized in DMSO the presence of DMSO could not be prevented. However, the DMSO concentration was kept as low as possible in order to minimize the effect of DMSO on the protein. Since the effect is DMSO concentration dependent, the same concentration of DMSO was added to the samples containing only NHR2. Nevertheless, these shifts were not very strong and are expected. As shown in **Figure 29**, the bottom peak presents biggest chemical shift. We assume that this chemical shift represents a terminal residue since these residues are not flanked by other residues; thus, they do not have the same general chemical environment as flanked residues.

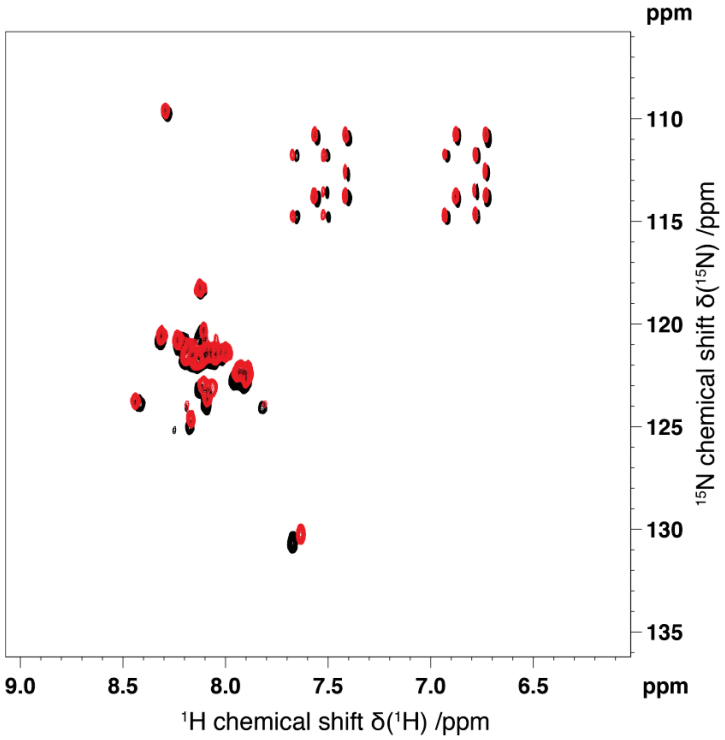

Figure 29: NHR spectrum of ^{15}N-labeled NHR2 in PBS buffer and 10% DMSO. 100 µM NHR2; PBS buffer, 100 mM NaCl, pH = 7.4, 10% D_2O (v/v); 10% DMSO (v/v) T = 25ºC, the 2D ^{15}N,^{1}H-TROSY-HSQC spectrum was acquired with 16 transients in a matrix with 2k data points in F2 in a spectral window of 8417.51 Hz, centered at 2816.94 Hz and 128 increments in F1 in a spectral window of 2188.46 Hz, centered at 7263.60 Hz.

Next, we repeated the experiments in the presence of the ligand. First we noticed the appearance of new signals (red circles). This is a good sign as it showed that the presence of **7.44** has an influence on the chemical environment of the residues of NHR2 (**Figure 30**). However, at least one of the signals that appeared (around 111 ppm in ^{15}N) was also from a Gly, which can only be found at the His-tag.

Whether these signals were caused by the dissociation of the NHR2 tetramer or an interaction between **7.44** and the His-tag of NHR2 could not be determined by the obtained spectrum. In order to determine the part of NHR2 that interacted with **7.44** and caused the signals, an assignment of the residues of NHR2 was needed.

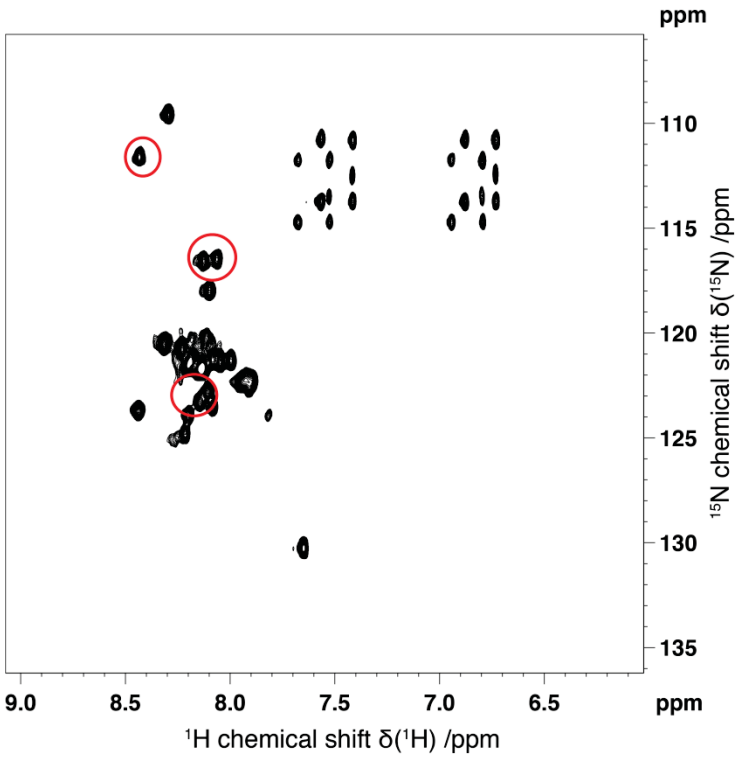

Figure 30: NMR spectrum of ^{15}N-labeled NHR2 in the presence of 7.44. 50 µM NHR2; 5 mM **7.44**; PBS buffer, 100 mM NaCl, pH = 7.4, 10% D_2O (v/v); T = 25ºC, (the ligand was prepared in DMSO and in a way that we would add 20uL of it; in this way the amount of DMSO in the final solution is also 10% DMSO). The 2D ^{15}N,^{1}H-TROSY-HSQC spectrum was acquired with 6 transients in a matrix with 2k data points in F2 in a spectral window of 8417.51 Hz, centered at 2816.94 Hz and 128 increments in F1 in a spectral window of 2188.46 Hz, centered at 7263.60 Hz.

Next, we tried to do the assignment. For this, I attempted to produce a triple-labeled (^{13}C, ^{15}N, ^{2}H) protein. However, when the data were acquired and no ^{13}C signal was detected, we noticed that the protein was not properly labeled. At this point we changed the pH from 7.4 to 6.5. We tried to decrease the pH in order to decrease the amide proton

exchange rate of the exposed residues [200]. If the exchange rate is too high, these exchanges cannot be detected. The spectra we tried to acquire were: (i) ^{15}N,^{1}H-TROSY-HSQC (sufficient ^{15}N labeling but insufficient concentration), (ii) 3D TrHNCO (no detectable signal), (iii) 3D TOCSY-TrHSQC (no signal due to low concentration), and (iv) ^{13}C,^{1}H-HSQC (no signal).

Next, a double-labeled (^{15}N, ^{13}C) sample was produced to see if we could get at least some assignments. This time, measured at 700 MHz, we were able to assign some residues (**Figure 31A**). The spectrum shows the free NHR2 (black spectrum) and bound NHR2 (red spectrum). Concerning the acquired spectra, we believe that the only peaks we saw belong to the His-tag. We counted 40 peaks in the HSQC, but in principle all amino acids, except prolines, should be represented by a peak. With respect to the sequence of the His-tagged NHR2 (**Figure 31B**), from the residues of the His-tag (red), those highlighted yellow could be assigned by following a standard assignment procedure, which focuses on characteristic and typical peaks that represent specific amino acids as starting points for assignment. The remaining peaks representing the residues of the His-tag are too overlapped to properly assign them. As for the peak labeled "**1**", unfortunately an assignment of this peak was not possible. However, the peak is consistent with typical peaks detected for serines. Furthermore, in the presence of ligand the peak shifts. We assume this peak represents either S42 or S58 due to their proximity to the hot spots, but no definitive assignment to a serine of NHR2 could be made. It is, however, certain that it is not caused by the His-tag since all serine residues from the His-tag were already assigned. As for "**2**" and "**3**", no assignment was possible. The ASN and QLN side chains are not affected by binding as they do not shift (blue circle) (**Figure 31A**).

Lastly, we removed the His-tag from labeled NHR2 with a recombinant enterokinase. However, during the IMAC purification of the supposedly tagless NHR2, by immobilizing all His-tagged NHR2, only a very small proportion of the enterokinase-treated NHR2 flowed through the column. Subsequently, a control SDS-PAGE showed no visual detectable protein bands. The result of the purification indicated that the applied cleavage procedure or the utilized enterokinase was not suitable for NHR2 under the present conditions. Thus, we were not able to obtain a sufficient quantity of labeled tagless NHR2 for NMR measurements.

In summary, the conducted NMR experiments described in this chapter did not allow for a clear resonance assignment of the binding relevant amino acids.

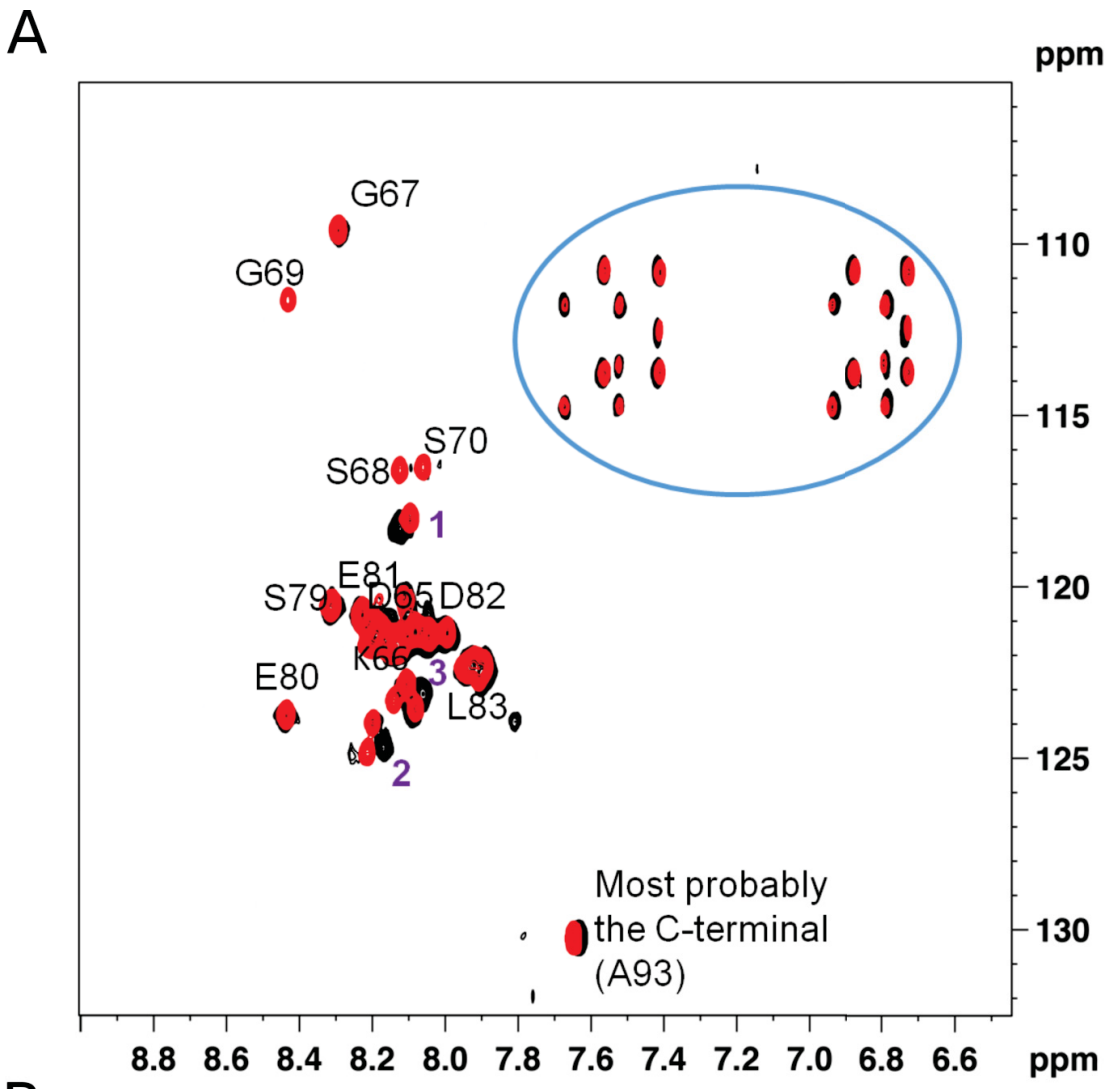

Figure 31: Assignment of NHR2 residues. (**A**) $^{15}N,^{1}H$-TROSY spectra of NHR2 in the absence and the presence of ligand (black and red, respectively). The obtained assignments and regions of interest (purple numbers and blue circle) are indicated in the figure. 50 µM NHR2; 5 mM **7.44**; PBS buffer, 100 mM NaCl, pH = 6.5, 10% D_2O (v/v); T = 25°C; measured at 700 MHz. The numbering of the assigned residues was applied with regard to the amino acid sequence of the His-tagged NHR2 construct (**Chapter 2.1.3**). (**B**) Amino acid sequence of the His-tagged NHR2 construct; hot spots (green boxes), His-tag (red letters) and assigned residues (yellow boxes).

3.11 Binding mode prediction of **7.44**

In order to elucidate the interaction between **7.44** and NHR2, a series of 35 MD simulations of free **7.44** diffusion around a NHR2 dimer were performed by Benedikt Frieg (Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, Germany). In the simulations, no guiding force was applied, thus all molecules remained unbiased. The visual inspection of the MD trajectories revealed multiple events of the formation and breaking of interactions between **7.44** molecules and the NHR2 dimer (**Figure 37B**). However, stable interactions between **7.44** and NHR2 were only prevalent at the NHR2 tetramerization interface (**Figure 32A**). A possible explanation is that the NHR2 tetramerization interface is largely hydrophobic (**Figure 37C**). An interaction between NHR2 and **7.44** is most likely promoted by this hydrophobicity. Subsequently, all **7.44** conformations that form stable interactions with NHR2 were clustered according to their structural similarity. The top six ranked clusters, with a cumulative cluster population of ~57.9%, revealed binding poses at the NHR2 tetramerization interface (**Figure 37D**). Noteworthy, the clustering exposed that the two most populated clusters reveal two adjacent **7.44** binding sites at the NHR2 dimer (**Figure 32B**). These binding sites are related by rotational symmetry and are almost equally populated (population of c1 = 13.5% vs. c2 = 13.1%). Two binding sites would allow for two ligands to bind at NHR2 simultaneously, which is in agreement with the determined Hill coefficient (**Chapter 3.8**) and previous studies [54].

In summary, the MD simulations suggest that up to two **7.44** molecules can bind to the NHR2 tetramerization interface.

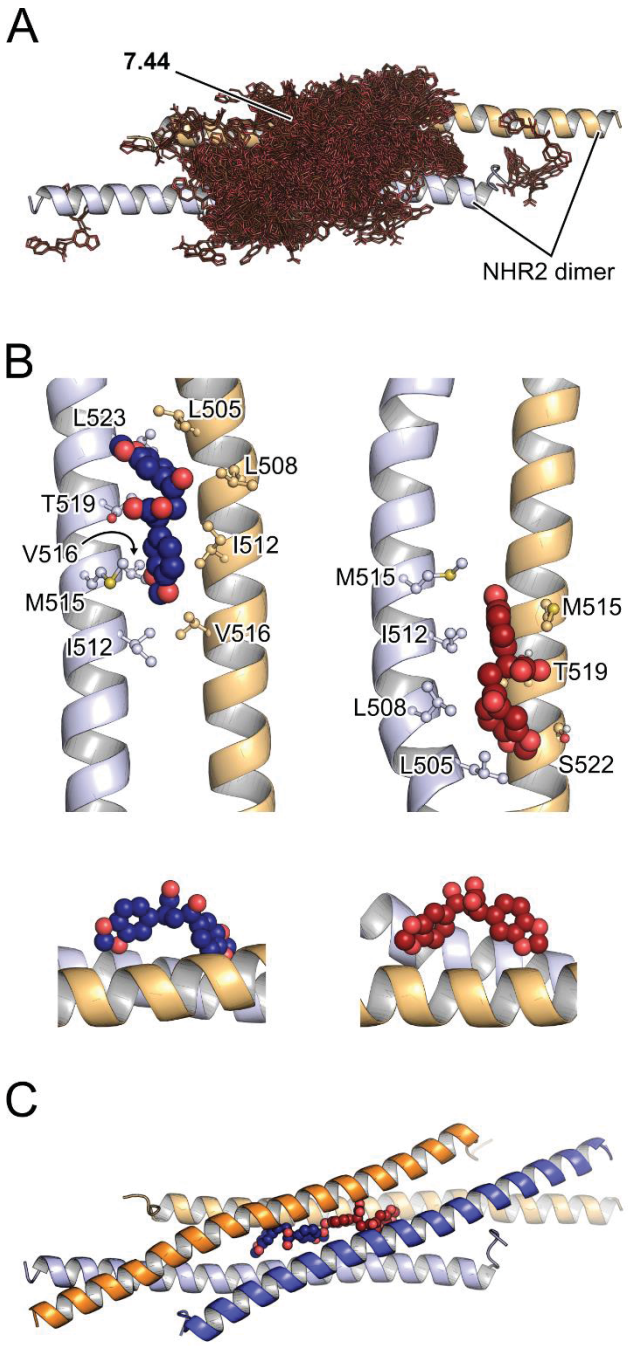

Figure 32: Interactions between 7.44 and the NHR2 dimer. (**A**) All stably bound **7.44** conformations observed during MD simulations projected onto the initial NHR2 dimeric structure. (**B**) Representative binding modes from the two largest populated clusters. The interacting residues are explicitly shown as ball-stick model and **7.44** as spheres. The bottom panels show the binding poses in side view orientation. (**C**) Projection of the cluster representatives onto the tetrameric NHR2 crystal structure [80]. In A – C, the protein structure is always shown in cartoon representation with each subunit colored differently and **7.44** is either shown as sticks (A) or spheres (B, C). All hydrogen atoms are omitted for presentation purposes. Figure taken from ref. [2].

3.12 Investigation of structurally similar substances

One goal of this project was to provide information that can be used for the optimization or the identification of new inhibitors of the NHR2 tetramerization. At the time of these investigations I used **7.44** as a template to search new and untested compounds with similar structural properties. In order to find suitable compounds which fingerprints were calculated for the commercially available section of the ZINC database [124]. As a measure of the 2D similarity with **7.44** the Tanimoto coefficient [176] was calculated for every compound, and out of the best ranking 100 compounds 30 were chosen by hand. The aim was, as far as possible, the selection of similar structures with small structural differences, in order to be able to precisely attribute an influence on the binding affinity to a defined structural change. Out of the 30 selected compounds, 16 compounds were available for purchase and were acquired from Key Organics for investigation. The compounds were analyzed by MST (**Chapter 2.2.9**) as done for **7.44** (**Chapter 3.8**). For 14 out of the 16 compounds no change in the thermophoresis of NHR2 was measured; their structures can be seen in **Figure 38**. In the presence of **T5** and **T7,** a change of the thermophoresis was detected (**Figure 33**). In the case of **T7,** the EC_{50} could not be determined, since the data points for a concentration above 10 mM were missing and this concentration was exceeding the maximal obtainable concentration I could achieve with the available compound sample. For **T5,** an EC_{50} of ~ 3.5 mM was determined. This EC_{50} is ~520 times higher than the EC_{50} for **7.44** (6.65 µM). Despite providing no improvement compared to **7.44** the gathered information can be used as an example of unfavorable changes to the structure of **7.44** for future approaches of optimization.

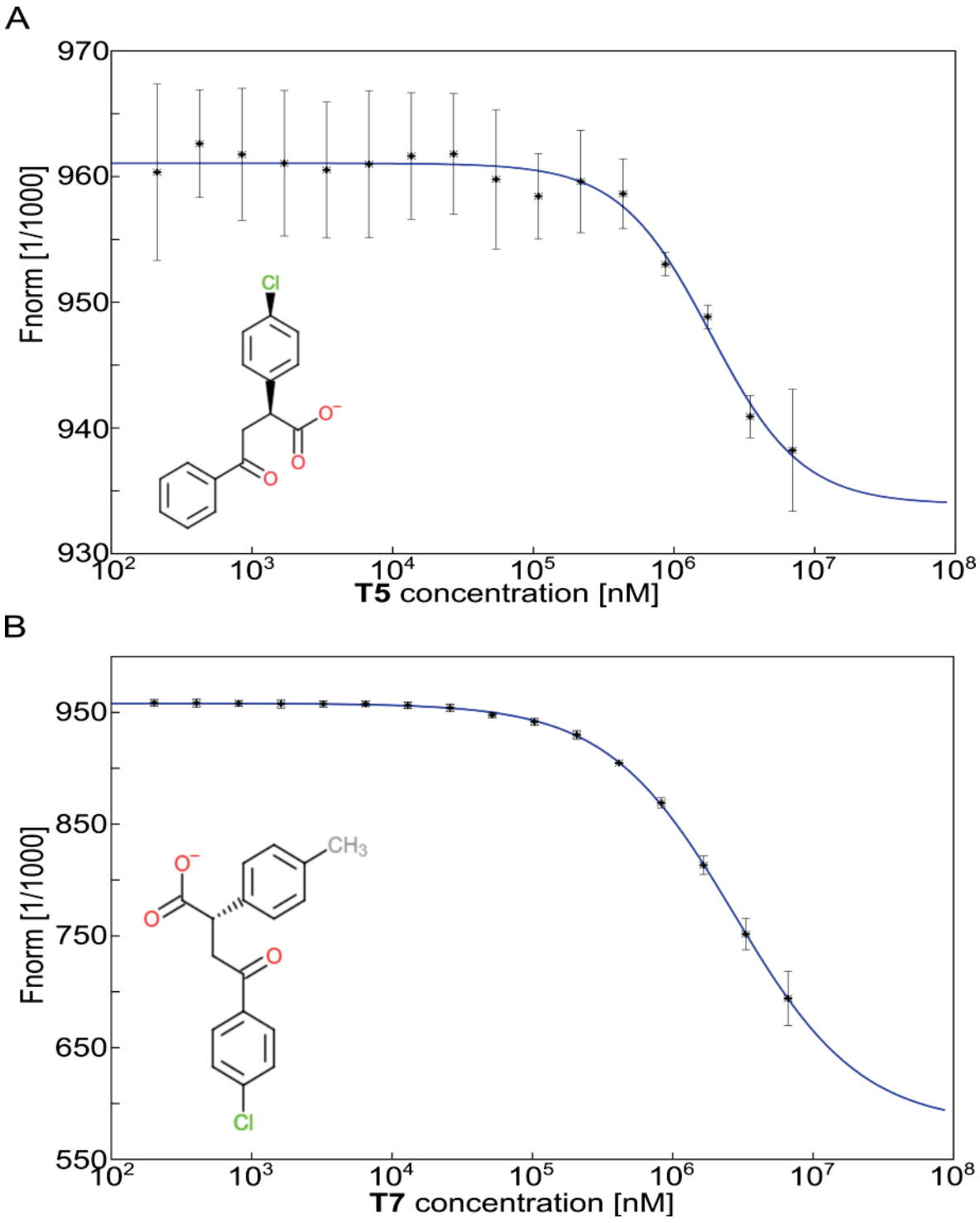

Figure 33: MST analysis of T5 and T7. (**A**) Effects of a titration of **T5** to a constant concentration of NHR2 on the thermophoresis of the NHR2 tetramer. n = 3; the error bars show the SD. EC_{50} ≈ 3.5 mM was calculated for **T5** using the MO Affinity Analysis software (NanoTemper Technologies, Germany). (**B**) Effects of a titration of **T7** to a constant concentration of NHR2 on the thermophoresis of the NHR2 tetramer. n = 3; the error bars show the SD. An EC_{50} could not be determined.

4 Discussion

In my project I examined the inhibitor of the NHR2 tetramerization **7.44**. Aiming to provide information regarding the affinity of the inhibitor as well as the binding mode I conducted biochemical and biophysical experiments. These experiments uncovered the K_{tet} of the NHR2 tetramer, the K_{lig} for **7.44** and the binding pose of **7.44**. MD simulations were performed to gain further insights into processes on the atomic level. However, during my experiments I encountered challenges regarding the thermal shift assay I tried to set up for my system.

For the following Chapter, the text has been taken and modified from the publication:

"**Kroeger, T.**, Frieg, B., Zhang, T., Hansen, F.K., Marmann, A., Proksch, P., Nagel-Steger, L., Groth, G., Smits, S.H.J., Gohlke, H. *EDTA aggregates induce SYPRO Orange-based fluorescence in thermal shift assay.* **PLOS ONE** 2017, 12, e0177024."

4.1 EDTA-mediated SYPRO Orange-based fluorescence in thermal shift assay

For testing the inhibitory potential of novel ligands and verifying the inhibitory potential of **7.44** I used the TSA method, which was previously used to study the stability of HSP90 in our working group [70]. The basis of the TSA is a fluorescent dye, such as SYPRO Orange, which in combination with interactions with hydrophobic amino acids of the sample protein experiences a shift in its quantum yield [130]. The shift of the quantum yield increases the fluorescence output of the dye and increases the fluorescence intensity during temperature-induced unfolding of the protein. Under native conditions, hydrophobic residues are mainly buried in the core region of soluble proteins. Hence, hydrophobic interactions between NHR2 and SYPRO Orange are prevented and a low fluorescence signal will be detected by the TSA method. As temperature increases, the protein loses its structural stability and starts to unfold. Due to the unfolding, SYPRO Orange can now interact with hydrophobic residues, causing an increase in the intensity of the fluorescence signal.

The conditions applied in this assay were derived from the previous work in our working group [70]. However, the EDTA in the buffer that ensured the solubility of NHR2 caused unplanned fluorescent signals. Subsequently, these effects of EDTA on the SYPRO Orange-based fluorescence in the TSA were further investigated. An interference of EDTA in the TSA was previously not mentioned and thus not considered in the experiments. With the results from my PhD project I uncovered that the "melting points" gathered in the presence of SYPRO Orange and EDTA might be erroneous. In order to investigate the connection between EDTA and a fluorescence signal in the absence of a protein, I set up experiments depending on different properties of EDTA. First, to verify that EDTA caused the interference, I tested if the effect on the SYPRO Orange-based fluorescence is depending on the concentration of EDTA (**Figure 9A**). An EC_{50} of 36.3 mM was determined, which is a commonly used EDTA concentration in experiments [201-203]. This EDTA concentration-dependency of the effect indicates that EDTA can cause an increase in the fluorescence intensity, even in the absence of any protein and other components in the used buffer. Next, EDTA is sensible to the pH [204], thus, different protonated subpopulations of EDTA will prevail at different pH intervals. Since the interference occurred at a pH ~10 two subpopulations of EDTA would be present in the solution ($EDTA^{3-}$ and $EDTA^{4-}$). A pH screening ranging from pH 7 to 12 would show a potential pH dependency of the interference and thus uncover, if a specific subpopulation caused the interference (**Figure 9B**). The increase of the relative change of fluorescence above pH 8 with a half maximal relative change of fluorescence at pH 10 and the maximal relative change of fluorescence above pH 12 corresponds with the proportion of the $EDTA^{4-}$ species in the total EDTA population. The $EDTA^{4-}$ concentration at the point of half maximal relative change of fluorescence equals ~27 mM, which lies near the EC_{50} from the concentration dependency of the SYPRO Orange-based fluorescence. The chosen pH maximum (pH 12) lies within the practically applied range of pH, as, for example, lipases and proteases can have a pH optimum above pH = 10 [205, 206]. Next, EDTA is known to complex metal ions in a stable octahedral conformation, which might cause the interference of EDTA. In order to probe this assumption, I performed Ca^{2+} titration on a solution containing EDTA and SYPRO Orange. The results from the titration indicates a decrease of the fluorescence (**Figure 9C**), which suggests that the effects of $EDTA^{4-}$ on SYPRO Orange can be reversed. Based on the results themselves it is unclear if this quenching of the signal is based on the formation of an octahedral $EDTA^{4-}$-Ca^{2+} complex [182] or the masking of the charge of EDTA. Another possible explanation for the

quenching of the fluorescence signal is based on the work of Müller and Haeberli [180] who found that EDTA can form supramolecular aggregates in solutions with basic pH. The addition of Ca^{2+} ions might impair the formation of such aggregates. With retrospect to the previous investigation of a concentration dependence of the detected fluorescence signal, my results of a SYPRO Orange:EDTA ratio of 1:3.6 for at EC_{50} = 36.3 mM EDTA (**Figure 9A**) supports this result.

In order to verify the aggregation of EDTA we analyzed EDTA and SYPRO Orange with the analytical ultracentrifugation (AUC). The AUC is widely used to quantitatively analyze biomacromolecules in solution [207]. The included fluorescence detection system allows the evaluation of sedimentation of the fluorescent sample [208]. Thus, should an aggregation of EDTA molecules occur and additionally be connected to SYPRO Orange, a change of the sedimentation coefficient for SYPRO Orange should be detectable. In our AUC experiments, we analyzed the movement of the sedimenting material at 60,000 rpm over time. Our results verify that the EDTA in the solution aggregates (**Figure 10**).

Finally, I tested the structurally similar EGTA in combination with SYPRO Orange to investigate whether the observed interference only occurs in the presence of EDTA or if it can be triggered with other molecules as well. Very similar results obtained for SYPRO Orange and EGTA, in the absence of protein, with respect to the EGTA concentration dependence, pH, and Ca^{2+} influence (**Figure 9D**) suggest that changes in fluorescence intensity in TSA experiments with SYPRO Orange at basic pH might occur also for other members of the EGTA family of molecules.

Even though our results are in agreement with previous studies [180] and in line with each other, no information about a structural model of an EDTA aggregate was proposed [180]. Thus, a molecular origin of the EDTA aggregation and the resulting effect on the SYPRO Orange-based fluorescence remained unknown. My aim was to provide a structural model for the aggregation of EDTA and the interaction between SYPRO Orange and EDTA using all-atom MD simulations. Previous studies have shown that MD simulations can successfully be used to elucidate binding modes, equilibria, and kinetics of protein-ligand [209-214] and host-guest systems [215, 216]. However, challenges regarding the force field accuracy [217, 218] and the exceeding of events beyond the simulated timescale [219, 220] are associated with MD simulations.

As described in ref. [1], these challenges described in this section were addressed on four levels: First, state-of-the-art parameterizations in terms of the ff12SB force field

from AMBER 12 [148] and the applied TIP3P water model [149] were applied. With regard to the Na^+ ions I applied the ion parameters from Åqvist [187] and Joung and Cheatham [186]. Despite the difference in the ions' hydration properties from both cases the simulation of the aggregation of $EDTA^{4-}$ showed similar results (**Figure 14**), thus, indicating that the $EDTA^{4-}$ aggregation occurs independently from specific ion parameters. Second, the simulation time for each MD simulation was set to 1 μs. Within this timescale I observed the collapse and reformation of an aggregate (**Figure 13**). In addition, the SASA values remained nearly constant within the last 250 ns of the simulation, thus, indicating that the chosen timeframe is sufficient to observe and analyze the aggregation of EDTA and the interaction between EDTA and SYPRO Orange. Third, my aim was to validate my simulation setup by semi-quantitatively comparing the aggregation of $EDTA^{4-}$ in the presence and absence of Ca^{2+} (**Figure 17**) and at different temperatures (**Figure 18**) with respect to experiment. As to the simulations in the presence or absence of Ca^{2+}, the MD simulations revealed a decrease in the aggregation formation of 40% in the presence of a Ca^{2+} concentration equivalent to that in TSA experiments that led to only 50% of the relative change in fluorescence intensity. As to the latter, the MD simulations indicate that increasing temperature fosters $EDTA^{4-}$ aggregation and that this effect is higher above the experimental "denaturation temperature" of SYPRO Orange/EDTA solutions than below. These results correspond to a higher fluorescence intensity found in TSA above the experimental "denaturation temperature" than below. Fourth, rather than focusing on absolute changes in molecular properties, I focused on relative ones in order to exploit effects of cancellation of errors regarding the used force fields. Consequently, the properties of the aggregation of $EDTA^{4-}$ were compared to properties of the aggregation of $EDTA^{3-}$ (**Figure 13**, **Figure 15** and **Figure 18**), the properties of the aggregation of $EDTA^{4-}$ at different concentration were analyzed (**Figure 13** and **Figure 15**) and the properties of the aggregation of $EDTA^{4-}$ in the presence or absence of Na^+ were analyzed (**Figure 13** and **Figure 16**). Using this approach, my simulations indicate that $EDTA^{4-}$ forms aggregates, the aggregation can occur at concentration similar to the concentration used in my experiments and that the simulation is relying on the presence of Na^+ ions.

The obtained model of the supramolecular EDTA aggregates from my MD simulations reveal that the $EDTA^{4-}$ molecules, in a U-form shape, arrange so that their carboxylate groups are facing a central layer of Na^+ ions, thereby exposing their ethylene groups to the solvent (**Figure 13B,C**). This structure of the aggregate resembles an inverse

bilayer. Inverse bilayer, similar to the one observed, have been observed for salts of amphiphilic anions in the solid state [221].

As described in ref. [1], the occurrence of EDTA aggregate bilayer can be attributed to (i) the exposed ethylene moiety only forming a small hydrophobic surface and being flanked by two polar tertiary amine groups, this results in a sufficient solubility in water, as inferred from the related compound *N*,*N*,*N*',*N*'-tetramethylethylendiamine, which is soluble in water and shows a log K_{OW} = 0.3 [222] and (ii) consistent with the "law of matching water affinities" [223] Na^+ forms contact ion pairs with carboxylate groups especially well, that way displaying specific binding properties in addition to nonspecific charge screening [224-227]. This results in Na^+ stabilizing α-poly-L-glutamate with ionized side chains in a helical conformation [225]. Moreover, by way of bridging carboxylates at strongly conserved positions, Na^+ stabilizes micelles [228] and superlattices [229] of amphiphilic carboxylates. There, $RCOO^-$ … Na+ … ^-OOCR ion triplets were found to be particularly important for aggregation [229]. As indicated by **Figure 13B**, contact ion triplets can be found in the $EDTA^{4-}$ aggregate. Paradoxically, the formation of contact ion triplets reduces electrostatic repulsion when more carboxylic groups are deprotonated. This is due to strong ionic bridging interactions replacing weaker hydrogen bonds involving non-ionized carboxylic groups [229].

The information gathered from the structural model of the EDTA aggregates give a possible explanation why $EDTA^{4-}$ aggregates but for $EDTA^{3-}$ no aggregation can be observed: $EDTA^{3-}$ has a positive charge at the amino group, which might cause repulsive forces within the bilayer and thereby destabilize the aggregate. Both the TSA experiments and the MD simulations are in agreement that the observed aggregation of EDTA is fostered by increasing temperatures. This can be explained by the effect of higher temperatures on the hydration free energy of carboxylate side chains [230] and the formation of contact ion pairs involving Na^+, which become more favorable [231]. Finally, due to the limitation with respect to the maximal amount of atoms within a MD simulation only a certain amount of EDTA could be placed within the simulation box, thus limiting the maximal size the EDTA aggregate that can be generated. However, the largest longitudinal axis of the observed aggregates of ~38 Å (**Figure 13B**) and ~31 Å (**Figure 18C**) are in agreement with the measured radius of 20 – 40 Å found by Müller and Haeberli for EDTA aggregates by cryo-electron microscopy [180].

Furthermore, the MD simulations revealed that SYPRO Orange binds with its mostly hydrophobic fluorophore moiety to the non-polar surface region of the EDTA

aggregate formed by the ethylene moieties of $EDTA^{4-}$ (**Figure 23A,B**). This interaction might explain the increased fluorescence intensity observed in TSA due to a solvatochromic effect [130, 131].

The observed EDTA/SYPRO Orange interaction seems to correspond to the SDS/SYPRO Orange interaction, which describes the occurrence of a fluorescence signal at protein bands in SDS gels stained with SYPRO Orange [192]. In addition, higher temperatures are beneficial for the aggregation of $EDTA^{4-}$, thus improving the interactions between EDTA and SYPRO Orange. This can also be seen in an increased EDTA shell around SYPRO Orange at higher temperatures (**Figure 21B**). Applying these information onto the TSA results would explain why the increased fluorescence intensity was observed at higher temperatures ("T_{mEDTA}"= 68 °C).

The TSA experiments showed that, in the presence of Ca^{2+} ions, the fluorescence signal can be quenched (**Figure 9C**). This is another agreement with the MD simulation, which showed that the EDTA/Ca^{2+} complexes do not form the bilayer structure (**Figure 17B**), thus decreasing the number of $EDTA^{4-}$ within the shell around SYPRO Orange (**Figure 21C**). The more Ca^{2+} is in the system the more $EDTA^{4-}$ molecules form a complex with these ions and cannot contribute to the aggregate anymore, thus decreasing the hydrophobic surface of the aggregate and thereby decreasing the hydrophobic interactions with SYPRO Orange. This will result in a decreased solvatochromic effect and therefore in a reduction of the fluorescence intensity.

In summary, these experiments and simulations explain that $EDTA^{4-}$ can form interactions with SYPRO Orange at increased temperatures, thereby resulting in a fluorescence signal that resembles TSA curves of unfolding proteins. Noteworthy, since EDTA has a widespread use in biology, pharmacy and food technology, and TSA is widely used to determine the thermal stability of proteins, the results of these experiments highlight that TSA experiments in the presence of EDTA, or other members of the EGTA family, might be affected by the interactions with SYPRO Orange.

Regarding the purpose of the TSA to identify potential inhibitors of the NHR2 tetramerization EDTA, leaving EDTA away would impair the solubility of the protein. The resulting precipitation might be due to ZN^{2+} ions that did not remain in the column matrix during the elution step in the purification process. Accordingly, leaving EDTA out would have resulted in a precipitation of NHR2 and thereby in the loss of most of the generated protein. Even several sequential buffer exchanges did not produce a significant improvement of the solubility of NHR2 in the absence of EDTA. EDTA had no influence

on the relative fluorescence at the targeted pH of ~7.4, but no increase of fluorescence was detected at this pH. In order to detect a fluorescence signal, I analyzed potential buffer additives and was able to identify Tris and citrate as fitting buffer components. For both, Tris and citrate, the TSA resulted in an increase of the relative fluorescence intensity at pH ~7.4 (**Figure 24B**). The estimated T_m of NHR2 derived from the curves is ~83 °C in the presence of citrate and ~80 °C in the presence of Tris (**Figure 24C**). Both T_m are in agreement with results from Wichmann *et al.* of ~85 °C [82] with a difference of 2 °C for citrate and 5 °C for Tris. Furthermore, both cases resulted in typical TSA curves [178, 179] with no increase of the relative fluorescence intensity in the absence of the protein.

For the following Chapter, the text has been taken and modified from the publication:

"**Kroeger, T.**, Frieg, B., Viegas, A., Smits, S., Groth, G., Etzkorn, M., Gohlke, H. *Biophysical characterization of a small-molecule inhibitor of RUNX1/ETO tetramerization with anti-leukemic effects*. **In preparation** 2018."

4.2 Binding mode prediction and affinity determination of NHR2 tetramerization inhibitor **7.44**

This chapter discusses the investigation, and the ensuing results of the interaction between the NHR2 domain of the fusion protein RUNX1/ETO, the product of the t(8;21) chromosomal translocation involved in the t(8;21) acute myeloid leukemia (**Chapter 1.2**) [82], and the inhibitor of the NHR2 tetramerization, **7.44** [54, 121]. Biophysical assays based on TSA and MST were established to verify the **7.44**-mediated inhibition of the NHR2 tetramerization and determine the affinity of **7.44** with respect to binding to NHR2. As part of this project, I show that the dissociation constant of **7.44** from the NHR2 dimer, K_{lig}, is 34.3 µM. Furthermore, by the combination of results gathered using STD-NMR [138, 139] and MD simulations of free ligand diffusion [232, 233] a binding mode model was set up that is based on integrative modeling. The binding mode model suggests that **7.44** binds within the NHR2 tetramerization interface and reveals moieties of **7.44** that are important for binding to the protein.

The conducted investigations delivered three independent biophysical results verifying the interaction between **7.44** and NHR2: First, the TSA revealed typical melting

curves [1, 178, 179] (**Figure 25**), while applying commonly used protein concentrations [70, 234-236]. The increase of the **7.44** concentration caused a decrease of the fluorescence intensity at ~85 °C, which is the previously determined melting point of the NHR2 tetramer [82]. In addition, an increasing **7.44** concentration causes a general increase of the fluorescence intensity at ~60 °C, which is the previously determined melting point of the NHR2 dimer [82]. These results provide a biophysical evidence for the inhibition of the NHR2 tetramerization via **7.44**, since the findings expose an alteration of the proportions of the two NHR2 species in a dose-dependent manner towards the dimer. However, due to the lack of a clear melting point at ~60 °C, a subsequent analysis of the melting curves with regard to a biphasic melting behavior [193] was not possible. Rather, we used the maxima of the fluorescence derivatives at ~85 °C as an indicator representing how much NHR2 tetramer is present, as described previously [193], which yielded an EC_{50} = 1.09 ± 0.09 mM for the effect of **7.44** titrated to NHR2.

Second, using MST I quantified the dissociation constant of the NHR2 tetramer into dimeric NHR2 and the dissociation constant of **7.44** with respect to the NHR2 dimer. This method exploits that the measured molecular movements of molecules along a temperature gradient is depending on the size, charge, and solvation shell of these molecules [135]. Thus, the dissociation of the NHR2 tetramer and the subsequent association of **7.44** to the NHR2 can be detected [133] since both events will at least result in a change of the size. The gathered data revealed a K_{tet} = 11.3 ± 1.81 µM and an EC_{50} = 6.65 ± 0.717 µM for **7.44** (**Figure 26**). Assuming that (i) two **7.44** molecules can bind to NHR2 simultaneously and that (ii) K_{lig} is identical for both **7.44** molecules, K_{lig} was then calculated to be 34.3 ± 3.76 µM. These assumptions are supported by the C_2-symmetry of the NHR2 dimer and by the fact that the two binding sites are not overlapping (see below). The determined, average Hill coefficient for **7.44** binding to NHR2 of n = 1.5 additionally supports the assumed inhibitory mechanism: Upon the binding of one **7.44** molecule to one of the binding sites in the tetramerization interface of NHR2, the second binding site becomes more easily accessible, which in turn increases the affinity for a second **7.44** molecule. This result agrees with previous studies [54].

Third, STD-NMR was used for an epitope mapping of **7.44** while it interacts with NHR2. In the case of multiple protons across **7.44** a signal was observable in the STD-NMR, thus indicating that **7.44** binds to NHR2. However, a difference in the amount of saturation across the protons is observable, indicating that the two heteroaromatic moieties of **7.44** are especially important for the binding to NHR2 (**Figure 27**). These results are in

good agreement with the model of the **7.44** epitope proposed by Metz *et al.* [54]. Supplementary solution NMR to determine the relevant amino acids of NHR2 for the binding of **7.44** remained unsuccessful. The gathered results did not allow for a clear resonance assignment. The reason for the insufficient signals could not be determined as part of this project, thus further trials to determine ideal conditions are required in order to determine the binding site of **7.44**.

The STD-NMR provided information regarding the epitope mapping of the ligand site, but no information regarding the protein site could be obtained. In order to close this information gap, a set of 35 MD simulations of free **7.44** molecules diffusion around the NHR2 dimer were performed. Similar MD simulations have already been used successfully for the identification of binding modes of small-molecules at proteins [171, 233]. Prior to the actual analysis of the MD simulation trajectories, the overall structure of the NHR2 dimer was examined. It is indicated in **Figure 32** and **Figure 37** that the NHR2 dimer maintained its helical character like the NHR2 tetramer (**Figure 3A**). The analysis of the trajectory revealed multiple events of bonds forming and breaking up between **7.44** and NHR2. This would indicate the achievement of an equilibrium, at least at non-specific binding sites, which in turn would be in agreement with a weak binding affinity of **7.44**. Furthermore, the MD simulations revealed two adjacent **7.44** binding sites at the NHR2 dimer that are almost equally populated. Due to their spatial separation both binding sites can be occupied simultaneously. This information supports the above assumption regarding a 2:1 stoichiometry of binding between **7.44** and NHR2. This in turn is consistent with a Hill coefficient > 1 and confirms that the binding process has reached equilibrium. The analysis of the binding pose of **7.44** at the two binding sites showed that the heteroaromatic moieties of **7.44** point to the tetramerization interface formed by the α - helices in NHR2. Hydrophobic interactions between NHR2 and **7.44** are stabilizing the complex. During the interaction with the protein the carboxyl group of **7.44** are facing away from NHR2 (**Figure 32B**), which is in perfect agreement with the epitope mapping from the STD-NMR (**Figure 27**).

It was unexpected that the predicted binding sites are located at the center of the NHR2 tetramerization interface (**Figure 32C**). On the one hand, this result and the associated binding modes of **7.44** indicate that the interaction between **7.44** and NHR2 is able to interfere with the NHR2 tetramerization, which is in agreement with the TSA and MST experiments (**Figure 25** and **Figure 26**). On the other hand, the location of the predicted binding sites differs from the previously predicted binding sites, a subset of hot

spots (D533, E536, and W540) [54]. Further investigation, especially experimental determination, of the binding site of **7.44** is required.

4.3 The search for more efficient inhibitors of the NHR2 tetramerization

The last part of my PhD project was the identification of compounds with similar structure compared to **7.44**. Out of the 16 analyzed structures only two showed an inhibitory effect on the NHR2 tetramerization (**T5** and **T7**, **Figure 33**). The determination of the EC_{50} revealed a significantly weaker efficiency compared to **7.44**. Thus, these structures would not be considered as a step towards optimization. A determination of the influence of the structural changes on the affinity of the investigated compounds is not possible due to the limited information available. However, combining the information regarding the binding pose of **7.44** (**Chapter 3.9**), the prediction of the binding sites (**Chapter 3.11**) and the structural differences between **7.44** and the new compounds, **T5** and **T7**, indicates why the binding affinity of **T5** and **T7** was significantly lower compared to the binding affinity of **7.44**.

Here, the heteroaromatic moieties of **7.44** interact with the hydrophobic NHR2 tetramerization interface. Unlike **T5** and **T7**, **7.44** has four hydrogen-bond acceptors with the four oxygen atoms in the heteroaromatic moieties that can form hydrogen bonds with the protein in addition to the hydrophobic interactions between NHR2 and **7.44**. Hydrogen bonds between a ligand and a protein in combination with hydrophobic interactions can mutually increase strength of these interactions [237].

In addition, the STD-NMR has shown that the heteroaromatic moieties of **7.44** are important for the interaction with the NHR2. Reducing the size of these moieties would result in a reduction of the potential “contact area” between the inhibitor and NHR2; thus, reducing the buried hydrophobic surface area of the NHR2 tetramerization interface. This reduction of buried hydrophobic area results in a reduction in binding affinity [238]. Both aspects would explain why **7.44** has a significantly higher binding affinity than **T5** and **T7**. Since none of the 16 structures achieved an EC_{50} in the same magnitude as **7.44**, the structural differences might be seen as unfavorable.

However, based on the available data, it is not yet possible to definitively assess the influence of the individual chemical changes between the new substances and with respect to **7.44** on the potential to inhibit the NHR2 tetramerization or the binding affinity to NHR2. A current barrier to the efficient selection of substances for further testing is that the binding site on the protein has not yet been determined experimentally. Although the

information on the binding pose of **7.44** shows which parts of the molecule are of particular importance for an interaction with NHR2, it does not reveal which residues of the protein interact with **7.44**.

5 Summary and perspective

In this thesis, I investigated the NHR2 domain of the RUNX1/ETO fusion protein, aiming to set up an assay to determine the stability of the tetramer and the binding mode of **7.44**, as a basis for a purposeful improvement of this PPIM. Using a strategy consisting of experiment and computations, my first aim was setting up the TSA as an assay allowing the identification and verification of potential inhibitors of the NHR2 tetramerization. In this context, an atypical fluorescence signal interfered with the intent to validate the interaction between **7.44** and NHR2. Subsequently, a comprehensive investigation of the underlying origin of this interference revealed an interaction between SYPRO Orange and EDTA aggregates as the cause. Since EDTA is widely used in biology, pharmacy and food technology and the TSA is a widely used technique for the investigation of the thermal stability of proteins [1], researchers conducting TSA experiments have to ensure that their results are not influenced by the presence of EDTA.

During these investigations, the MST was successfully established as a method to determine the stability of the NHR2 tetramer (K_{tet} = 11.3 μM) and the affinity of **7.44** (K_{lig} = 34.338 μM). These results provide the first biophysical evidence of **7.44** binding to NHR2 and interfering with the NHR2 tetramerization.

My next aim was the determination of the binding mode of **7.44** and NHR2. The STD-NMR revealed the moieties of **7.44** that are closer to the protein upon interaction, thus indicating atoms of the ligand that are important for the interaction. The MD simulations provided a molecular image of the interaction of **7.44** and NHR2 and allowed for a prediction of the binding site. This information can be used for further searches of new potential inhibitors of the NHR2 tetramerization or for the optimization of inhibitors.

There is still no treatment for AML, and NHR2 has been shown to be an important target, which highlights the importance of the identification and optimization of inhibitors of the NHR2 tetramerization. This study adds new information about the stability of the NHR2 tetramer and the affinity of the NHR2 tetramerization inhibitor **7.44**. Subsequent to this study, the binding mode of **7.44** and NHR2 needs to be experimentally determined. Furthermore, a better understanding of the influence of the chemical space surrounding **7.44** on the NHR2 tetramerization will be crucial for an efficient target-oriented ligand optimization.

Ultimately I could show with this work that combining the TSA, for the identification of NHR2 tetramerization inhibitors, the MST assay, for the quantification of

the binding affinity of the identified inhibitor, and the STD-NMR, for the identification of the binding-relevant epitope, provides an approach that can help in future studies to identify and develop structurally related inhibitors of the NHR2 tetramerization, which possess increased binding affinity, improved bioavailability, and enhanced anti-leukemic effects to inhibit RUNX1/ETO oncogenic function in t(8;21) acute myeloid leukemia.

6 Appendix

6.1 Sample composition for EDTA-mediated SYPRO Orange-based fluorescence

Table 9: Composition of samples for Thermal shift assay [a] [b].

EDTA[c]	EGTA[c]	Ca^{2+}[c]	pH	Figure
0	0	0	10	
1	0	0	10	
10	0	0	10	
20	0	0	10	
30	0	0	10	
40	0	0	10	2A
50	0	0	10	
60	0	0	10	
70	0	0	10	
80	0	0	10	
90	0	0	10	
100	0	0	10	
100	0	0	7	
100	0	0	7.5	
100	0	0	8	
100	0	0	8.5	2B
100	0	0	9	
100	0	0	9.5	
100	0	0	10	

100	0	0	10.5	
100	0	0	11	
100	0	0	11.5	
100	0	0	12	
100	0	10	10	
100	0	50	10	
100	0	100	10	
100	0	200	10	
100	0	500	10	2C
0	0	10	10	
0	0	50	10	
0	0	100	10	
0	0	500	10	
0	0	0	10	
0	30	0	10	2D
0	60	0	10	
0	60	100	10	

[a] Technical details of Thermofluor assay are described in Materials and Methods.

[b] Components were solved in H_2O.

[c] Concentration, in mM.

6.2 DNA – and protein sequence

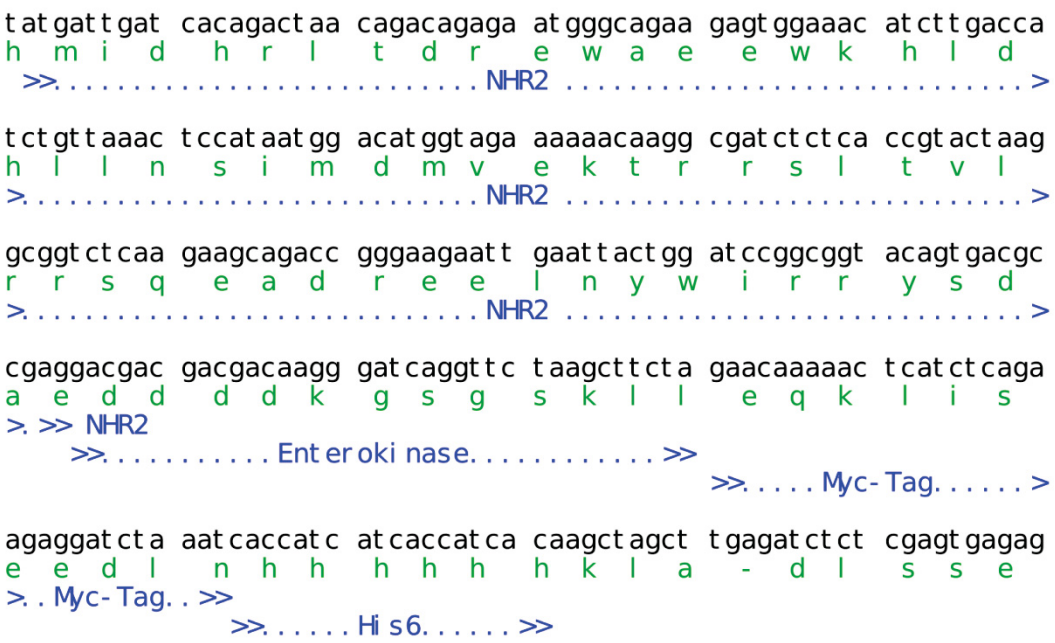

Figure 34: **Sequence of NHR2-His-tag construct from the pET19-vector**. Black – DNA sequence on the plasmid, green – amino acid sequence of the resulting protein, blue – resulting regions of the constructed protein.

6.3 Proximity of EDTA and SYPRO Orange molecules

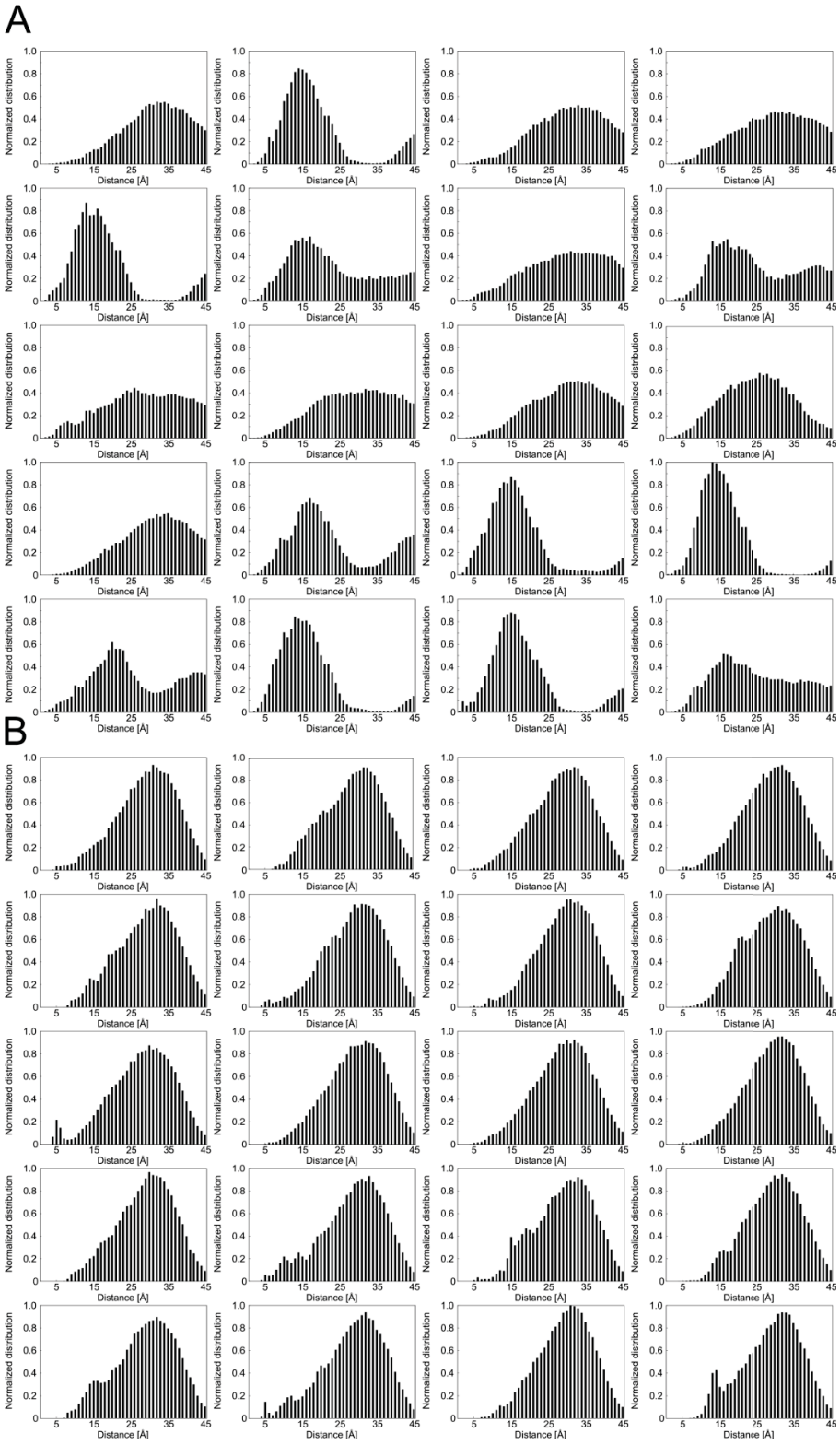

Figure 35: Proximity of EDTA and SYPRO Orange molecules. Frequency distributions of distances between the center of ethylene groups of each $EDTA^{4-}$ (**A**) or $EDTA^{3-}$ (**B**) molecule and the core region (**Figure 20**) of the nearest SYPRO Orange molecule. Figure taken from ref. [1].

6.4 Determination of K_{tet} and K_{lig} from MST experiments

This part was taken from ref. [125] and slightly adapted in the text. The EC_{50} value determined by TSA and MST in this study as well as the EC_{50} values determined by BS^3 cross-linking assay and ELISA before [54] are composites of the dissociation constant of the NHR2 tetramer (K_{tet}) and the association constant of **7.44** to NHR2 dimers. As the NHR2 tetramer (T) is in an equilibrium with its dimeric form (D), the ligand (L) can only bind to the NHR2 dimer (D) after dissociation of T. Given the symmetry of a NHR2 dimer, a second ligand can bind to the DL complex, resulting in a DLL complex (Eq. 3, **Chapter 3.8**). The respective equilibrium constants are defined as:

$$K_{tet} = [D]^2 / [T] \quad \text{(Eq. 4a)}$$

and

$$K_{lig} = [D] \cdot [L] / [DL] \quad \text{(Eq. 4b)}$$

K_{lig} could not be determined directly. Instead, K_{lig} was determined as a function of EC_{50} and K_{tet} according to the following equations [55, 125]. The following derivation was taken from ref. [125] and has been applied in a related context in ref. [55].

In the absence of L, the concentration of unbound NHR2 dimer D is $[D]_0$ and the concentration of the NHR2 tetramer T is $[T]_0$. The total concentration of NHR2 dimer $[D]_{tot}$ is (Eq. 5):

$$[D]_{tot} = 2[T]_0 + [D]_0 \quad \text{(Eq. 5)}$$

Substituting Eq. 5 into Eq. 4a and solving the obtained quadratic equation with respect to $[D]_0$ leads to Eq. 6:

$$[D]_0 = -\frac{K_{tet}}{4} + \sqrt{\left(\frac{K_{tet}}{4}\right)^2 + \frac{K_{tet}\,[D]_{tot}}{2}} \quad \text{(Eq. 6)}$$

Substituting Eq. 6 into Eq. 5 and rearranging leads to Eq. 7, allowing the calculation of the maximal NHR2 tetramer concentration $[T]_0$ as a function of K_{tet} and $[D]_{tot}$.

$$[T]_0 = \frac{[D]_{tot} + \frac{K_{tet}}{4} - \sqrt{\left(\frac{K_{tet}}{4}\right)^2 + \frac{K_{tet}\,[D]_{tot}}{2}}}{2} \quad \text{(Eq.7)}$$

In the presence of a total ligand concentration of $[L]_{tot} = EC_{50}$, the total NHR2 dimer concentration $[D]_{tot}$ is defined by Eq. 8:

$$[D]_{tot} = 2[T]_{50} + [D]_{50} + [DL]_{50} + [DLL]_{50} \quad \text{(Eq. 8)}$$

in which $[T]_{50}$, $[D]_{50}$, $[DL]_{50}$ and $[DLL]_{50}$ are the concentrations of the respective molecular species at $[L]_{tot} = EC_{50}$. In this case, $[T]_{50}$ is half of the maximal tetramer concentration $[T]_0$ at [L] = 0 (Eq. 9).

$$[T]_{50} = \frac{[T]_0}{2} \qquad \text{(Eq. 9)}$$

The concentration of the unbound NHR2 dimer $[D]_{50}$ is obtained by substituting Eq. 9 into Eq. 4a (Eq. 10):

$$[D]_{50} = \sqrt{\frac{K_{tet}\,[T]_0}{2}} \qquad \text{(Eq. 10)}$$

Based on the definition of the dissociation constant of the PPIM-bound complexes K_{lig}, the concentrations of the ligand-bound complexes are given by Eq. 11 and Eq. 12:

$$[DL]_{50} = 2\,\frac{[D]_{50}\,[L]_{50}}{K_{lig}} \qquad \text{(Eq. 11)}$$

$$[DLL]_{50} = \frac{[D]_{50}\,[L]_{50}^2}{K_{lig}^2} \qquad \text{(Eq. 12)}$$

if one assumes that the dissociation constants for both ligand binding events are equal, in agreement with the remote ligand binding sites and our model (Eq. 3) that depends on the pre-dissociation of T. $[L]_{50}$ is the concentration of the unbound PPIM at EC_{50}. The factor of two in Eq. 11 is a result of the two equivalent ligand binding sites of the NHR2 dimer.

Substituting Eq. 11 and Eq. 12 into Eq. 8 and solving the resulting quadratic equation with respect to K_{lig} leads to Eq. 13. The negative solution of Eq. 13 has no physical meaning because of $K_{lig} \geq 0$ and was omitted.

$$K_{lig} = -\frac{[D]_{50}\,[L]_{50}}{(2[T]_{50}+[D]_{50}-[D]_{tot})} + \sqrt{\left(\frac{[D]_{50}\,[L]_{50}}{(2[T]_{50}+[D]_{50}-[D]_{tot})}\right)^2 - \frac{[D]_{50}\,[L]_{50}^2}{(2[T]_{50}+[D]_{50}-[D]_{tot})}} \qquad \text{(Eq. 13)}$$

According to Eq. 9 and Eq. 10, the terms $[D]_{50}$ and $[T]_{50}$ in Eq. 13 are constants for given $[T]_0$ and K_{tet}. Thus, Eq. 13 is a function of $[L]_{50}$, which remains to be defined. For $[L]_{tot} = EC_{50}$, the total ligand concentration $[L]_{tot}$ is defined by Eq. 14.

$$[L]_{tot} = IC_{50} = [L]_{50} + [DL]_{50} + 2\,[DLL]_{50} \qquad \text{(Eq. 14)}$$

Substituting Eq. 11 and Eq. 12 into Eq. 14 and solving the resulting quadratic equation with respect to $[L]_{50}$ leads to Eq. 15. The negative solution of Eq. 15 has no physical meaning because of $[L]_{50} \geq 0$ and was omitted.

$$[L]_{50} = -\frac{1}{2}\frac{\left(2\frac{[D]_{50}}{K_{lig}}+1\right)}{2\frac{[D]_{50}}{K_{lig}^2}} + \sqrt{\left(\frac{1}{2}\frac{\left(2\frac{[D]_{50}}{K_{lig}}+1\right)}{2\frac{[D]_{50}}{K_{lig}^2}}\right)^2 + \frac{IC_{50}}{2\frac{[D]_{50}}{K_{lig}^2}}} \quad \text{(Eq. 15)}$$

According to Eq. 10, $[D]_{50}$ in Eq. 15 is constant for given $[T]_0$. Finally, solving the system of Eq. 13 and Eq. 15 allows calculating K_{lig} for given $[D]_{tot}$, K_{tet}, and EC_{50}. For this purpose, $[D]_{tot}$ is the total concentration of NHR2 dimer in the assay.

6.5 First derivative of fluorescence signal from TSA

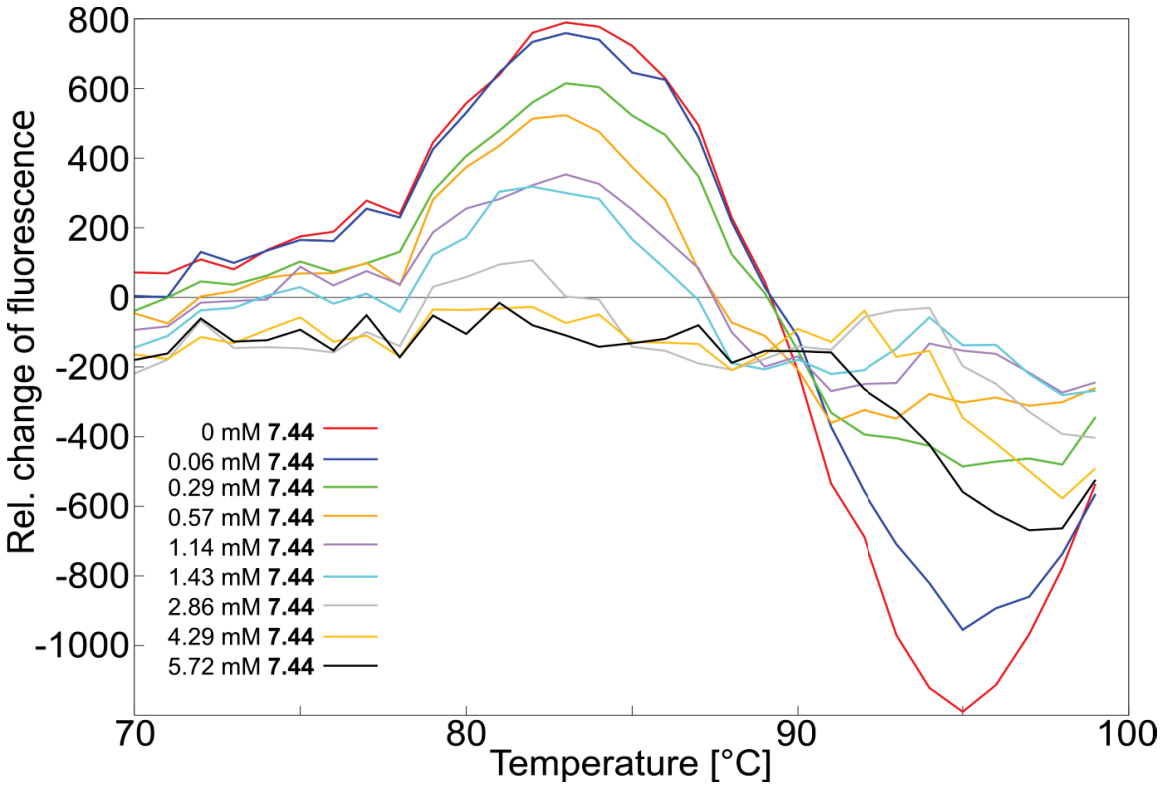

Figure 36: First derivative of fluorescence signal. The d*F*/d*T* was determined for 9 samples containing a constant concentration of NHR2 and a varying concentration (provided in the legend) of **7.44**. Figure taken from ref. [2].

6.6 Molecular dynamics (MD) simulations of **7.44** binding to the NHR2 dimer

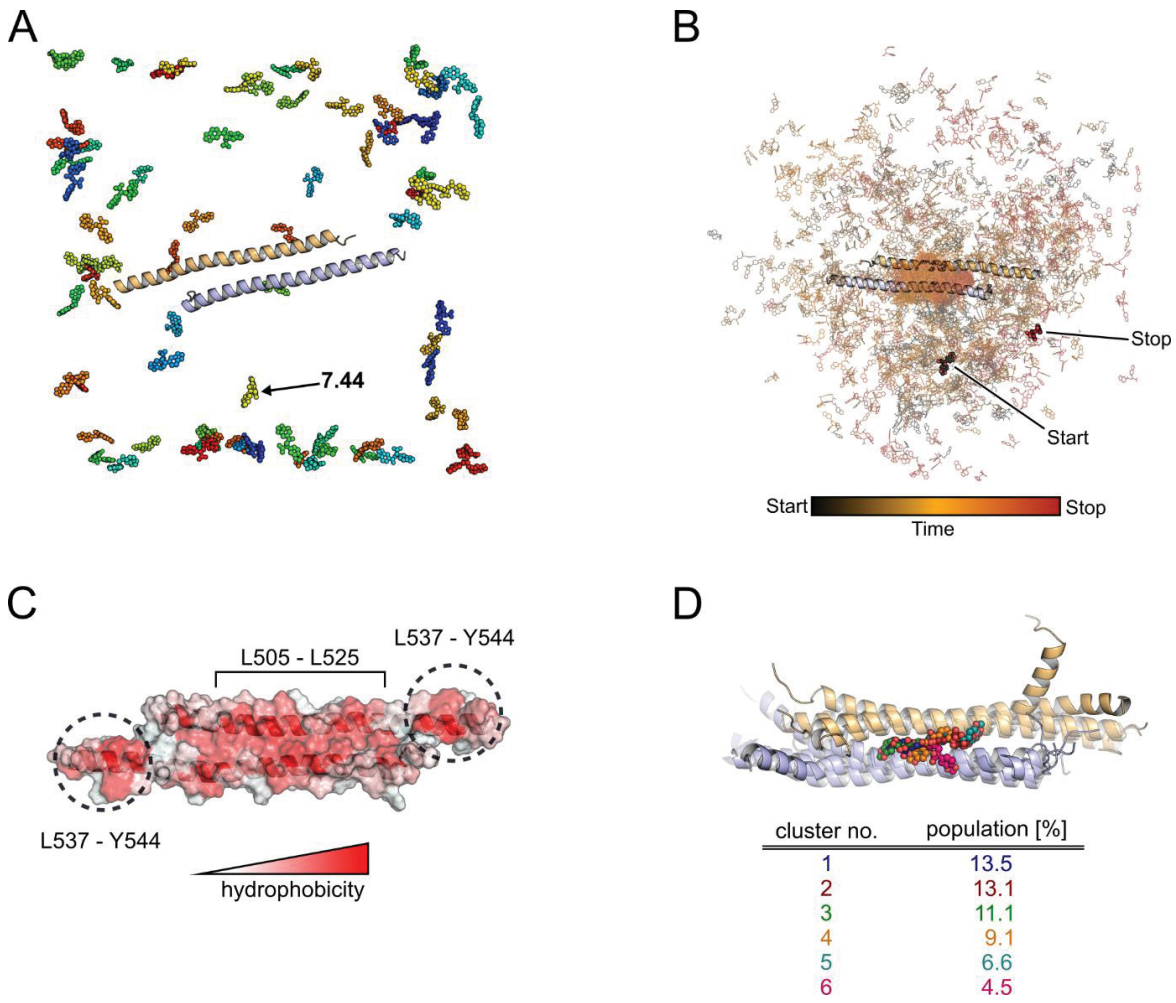

Figure 37: Molecular dynamics (MD) simulations of 7.44 binding to the NHR2 dimer. (**A**) Distribution of **7.44** (depicted as sphere) around the dimeric NHR2 (depicted as cartoon representation). **7.44** is always colored differently with each color denoting an independent starting structure for the binding simulations. (**B**) Traces of one **7.44** molecule during (MD) simulations of 350 ns length. The initial (start) and the final (stop) positions of **7.44** are shown as spheres and the intermediate states are shown as sticks colored according to the elapsed time (see scale on the bottom). (**C**) The NHR2 dimer structure used for MD simulations is shown as surface-cartoon representation colored according to the hydrophobicity spectrum. The labels depict residues that form the hydrophobic tetramerization interfaces of the NHR2 dimer. (**D**) Overlay of the top six ranked clusters (cluster population provided in table) with **7.44** always colored differently. Figure taken from ref. [2].

6.7 Compounds from SAR by catalogue

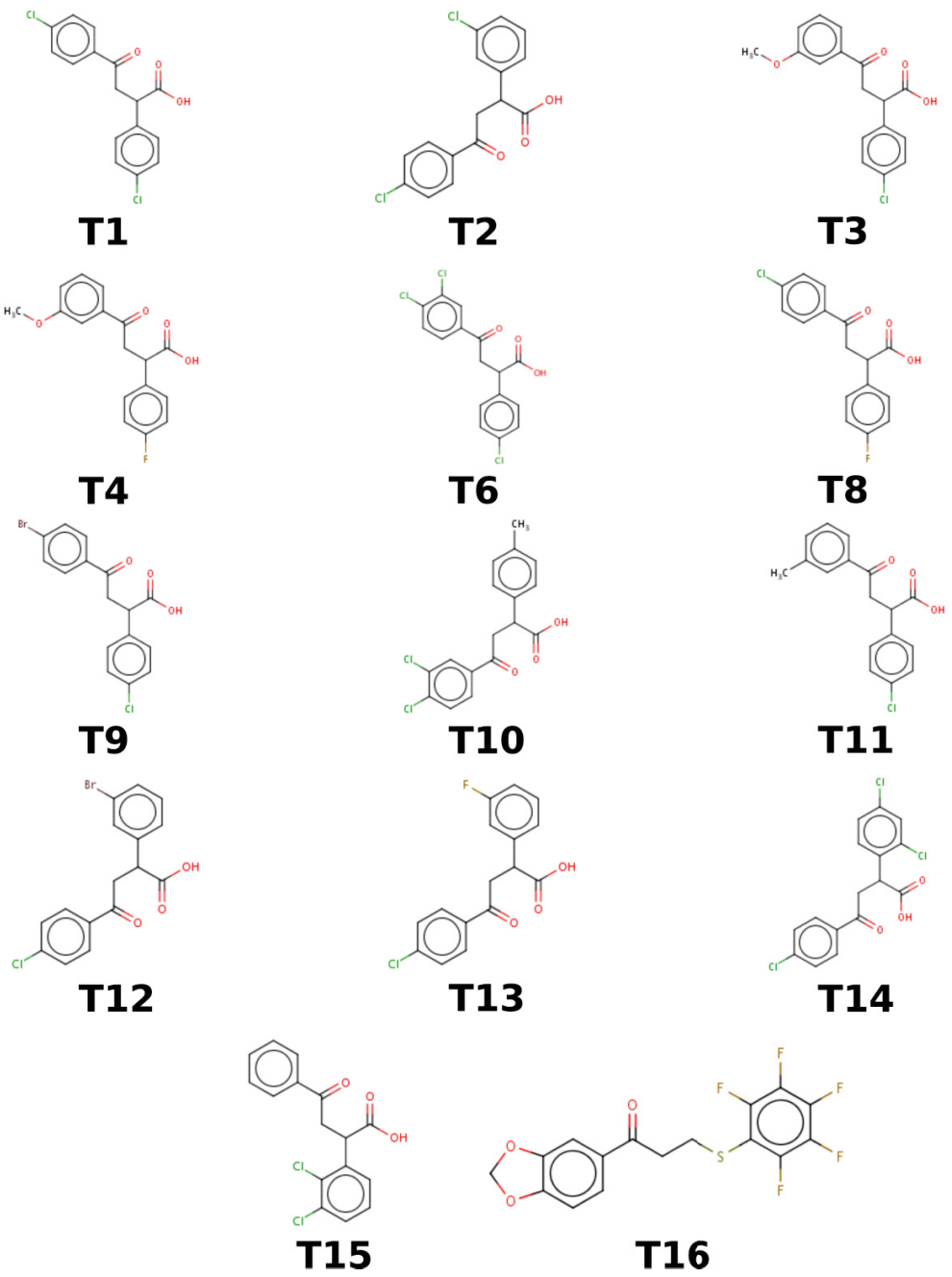

Figure 38: Resulting structures of the SAR by catalogue. Structures obtained from ChemDB [239].

Table 10: Compound IDs and ZINC IDs for selected and purchased compounds.

Name	Compound ID (Key Organics, UK)	ZINC ID
T1	7J-563S	ZINC05344495
T2	7J-529S	ZINC01397581
T3	8J-550S	ZINC01400835
T4	7J-537S	ZINC01397593
T5	7J-560S	ZINC01397623
T6	7J-624S	ZINC05344493
T7	7J-525S	ZINC01397576
T8	7J-549S	ZINC05344496
T9	7J-620S	ZINC05344494
T10	7J-628S	ZINC01397690
T11	7J-571S	ZINC01397637
T12	7J-587S	ZINC01397658
T13	7J-557S	ZINC01397620
T14	7J-541S	ZINC05344397
T15	6J-577S	ZINC03048603
T16	JS-2635	ZINC02531692

7 Acknowledgements

First of all, I want to thank Prof. Dr. Holger Gohlke for giving me the opportunity to pursue the PhD at the Heinrich Heine University, and for his support, his guidance and his teaching throughout my work.

Thanks to my colleagues and friends of the CPC lab. In particular, I would like to thank: Stephan Schott Verdugo, Dr. Christopher Pfleger, Dr. Giulia Pagani, and Neha Verma, for their constant, invaluable and patient help in calculations, scripting, reading this thesis, and much more; Dr. Emanuele Ciglia and Dr. Ido Ben-Shalom for their invaluable support in the beginning of my project. Thank you for all the wonderfull memories I was able to collect with you.

Thanks to Dr. Aleksandra Zivkovic for the supervision in the instrumental analytics practical course; Dr. Alexander Metz, to whom I will always be grateful for introducing me into this project and supporting me during the early days of my work; and Benedikt Frieg for his support and help with everything revolving around MD simulations.

Thanks to Prof. Dr. Georg Groth, Dr. Sander Smits, Dr. Luitgard Nagel-Steger, Dr. Finn Hansen, Dr. Aldino Viegas, Dr. Manuel Etzkorn, Dr. Andreas Marmann and Tao Zhang for the collaboration that brought our joint efforts to a success in this project.

Thanks to Dr. Michael Lenders, Dr. Christian Schwarz, Dr. Sven Reimann, Dr. Alexander Minges, Mareike Berleth, Kerstin Förster, Jens Reiners, Olivia Spitz, and Martina Wesemann for their support and fruitful discussion in the laboratory.

Thanks to my mother and my siblings, who were always there for me.

Thanks to my father. I will never forget you!

8 List of figures

9 List of tables

10 Curriculum vitae

Personal information:

Name:	Tobias Kröger
Date of birth:	08.12.1987 in Osnabrück (Germany)

Education:

1998 – 2007	*High school degree*
	Kardinal-von-Galen-Gymnasium, Mettingen
2007 – 2010	*University degree (Bachelor of Science)*
	Universität Osnabrück, Osnabrück
2010 – 2012	*University degree (Master of Science)*
	Universität Osnabrück, Osnabrück
2012 – now	*PhD student*
	Heinrich-Heine-Universität Düsseldorf, Düsseldorf

Work experience:

2012 – 2016	*Scientific coworker*
	Heinrich-Heine-Universität Düsseldorf, Düsseldorf

11 References

1. Kroeger, T., et al., *EDTA aggregates induce SYPRO orange-based fluorescence in thermal shift assay.* PLoS One, 2017. **12**(5): p. e0177024.
2. Kroeger, T., et al., *Biophysical characterization of a small-molecule inhibitor of RUNX1/ETO tetramerization with anti-leukemic effects.* in preparation, 2018.
3. Warner, J.K., et al., *Concepts of human leukemic development.* Oncogene, 2004. **23**(43): p. 7164-77.
4. Seufert, W. and W.D. Seufert, *The Recognition of Leukemia as a Systemic Disease.* Journal of the History of Medicine and Allied Sciences, 1982. **XXXVII**(1): p. 34-50.
5. Robert Koch institute. *Krebs in Deutschland 2011/2012*. 2015 17.12.2015 [cited 2017 07.04.2017]; 10th edition: Available from: http://www.krebsdaten.de.
6. Smith, M., et al., *Adult acute myeloid leukaemia.* Crit Rev Oncol Hematol, 2004. **50**(3): p. 197-222.
7. Estey, E. and H. Dohner, *Acute myeloid leukaemia.* Lancet, 2006. **368**(9550): p. 1894-907.
8. Phekoo, K.J., et al., *The incidence and outcome of myeloid malignancies in 2,112 adult patients in southeast England.* Haematologica, 2006. **91**(10): p. 1400-4.
9. Lowenberg, B., J.R. Downing, and A. Burnett, *Acute myeloid leukemia.* N Engl J Med, 1999. **341**(14): p. 1051-62.
10. Burnett, A., M. Wetzler, and B. Lowenberg, *Therapeutic advances in acute myeloid leukemia.* J Clin Oncol, 2011. **29**(5): p. 487-94.
11. Derolf, A.R., et al., *Improved patient survival for acute myeloid leukemia: a population-based study of 9729 patients diagnosed in Sweden between 1973 and 2005.* Blood, 2009. **113**(16): p. 3666-72.
12. Lichtman, M.A., *A historical perspective on the development of the cytarabine (7days) and daunorubicin (3days) treatment regimen for acute myelogenous leukemia: 2013 the 40th anniversary of 7+3.* Blood Cells Mol Dis, 2013. **50**(2): p. 119-30.
13. Pfirrmann, M., et al., *Prediction of post-remission survival in acute myeloid leukaemia: a post-hoc analysis of the AML96 trial.* Lancet Oncol, 2012. **13**(2): p. 207-14.
14. Rai, K.R., et al., *Treatment of acute myelocytic leukemia: a study by cancer and leukemia group B.* Blood, 1981. **58**(6): p. 1203-12.
15. Yates, J., et al., *Cytosine arabinoside with daunorubicin or adriamycin for therapy of acute myelocytic leukemia: a CALGB study.* Blood, 1982. **60**(2): p. 454-62.
16. Yates, J.W., et al., *Cytosine arabinoside (NSC-63878) and daunorubicin (NSC-83142) therapy in acute nonlymphocytic leukemia.* Cancer Chemother Rep, 1973. **57**(4): p. 485-8.
17. IARC, *IARC MONOGRAPHS ON THE EVALUATION OF CARCINOGENIC RISKS TO HUMANS*. Ionizing Radiation, Part 1: X- and Gamma (γ)-Radiation, and Neutrons. 2000, Lyon: IARC Press.
18. Stewart, B.W. and P. Kleihues, eds. *World Cancer report*. 2003, IARC Press: Lyon.
19. Savitz, D.A. and E.E. Calle, *Leukemia and occupational exposure to electromagnetic fields: review of epidemiologic surveys.* J Occup Med, 1987. **29**(1): p. 47-51.

20. Theriault, G., et al., *Cancer risks associated with occupational exposure to magnetic fields among electric utility workers in Ontario and Quebec, Canada, and France: 1970-1989.* Am J Epidemiol, 1994. **139**(6): p. 550-72.
21. IARC, *Wood, leather, and some associated industries*. IARC: monographs on the evaluation of the carcinogenic risk of chemicals to man. Vol. 25. 1981, Lyon: IARC Press.
22. Horwitz, M., *The genetics of familial leukemia.* Leukemia, 1997. **11**(8): p. 1347-59.
23. Song, W.J., et al., *Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia.* Nat Genet, 1999. **23**(2): p. 166-75.
24. Hasle, H., I.H. Clemmensen, and M. Mikkelsen, *Risks of leukaemia and solid tumours in individuals with Down's syndrome.* Lancet, 2000. **355**(9199): p. 165-9.
25. Taylor, A.M., *Chromosome instability syndromes.* Best Pract Res Clin Haematol, 2001. **14**(3): p. 631-44.
26. Bennett, J.M., et al., *Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group.* Br J Haematol, 1976. **33**(4): p. 451-8.
27. Vardiman, J.W., et al., *The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes.* Blood, 2009. **114**(5): p. 937-51.
28. Vardiman, J.W., N.L. Harris, and R.D. Brunning, *The World Health Organization (WHO) classification of the myeloid neoplasms.* Blood, 2002. **100**(7): p. 2292-302.
29. Arber, D.A., et al., *The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia.* Blood, 2016. **127**(20): p. 2391-405.
30. Betz, B.L. and J.L. Hess, *Acute myeloid leukemia diagnosis in the 21st century.* Arch Pathol Lab Med, 2010. **134**(10): p. 1427-33.
31. Chu, L.H. and B.S. Chen, *Construction of a cancer-perturbed protein-protein interaction network for discovery of apoptosis drug targets.* BMC Syst Biol, 2008. **2**: p. 56.
32. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
33. Hood, L., et al., *Systems biology and new technologies enable predictive and preventative medicine.* Science, 2004. **306**(5696): p. 640-3.
34. Mullard, A., *Protein-protein interaction inhibitors get into the groove.* Nat Rev Drug Discov, 2012. **11**(3): p. 173-5.
35. Zinzalla, G. and D.E. Thurston, *Targeting protein-protein interactions for therapeutic intervention: a challenge for the future.* Future Med Chem, 2009. **1**(1): p. 65-93.
36. Buck, E. and R. Iyengar, *Organization and functions of interacting domains for signaling by protein-protein interactions.* Sci STKE, 2003. **2003**(209): p. re14.
37. Virkamaki, A., K. Ueki, and C.R. Kahn, *Protein-protein interaction in insulin signaling and the molecular mechanisms of insulin resistance.* J Clin Invest, 1999. **103**(7): p. 931-43.
38. Zhang, X., et al., *Protein-protein interactions among signaling pathways may become new therapeutic targets in liver cancer (Review).* Oncol Rep, 2016. **35**(2): p. 625-38.
39. Pellicena, P. and J. Kuriyan, *Protein-protein interactions in the allosteric regulation of protein kinases.* Curr Opin Struct Biol, 2006. **16**(6): p. 702-9.
40. Rao, V.S., et al., *Protein-protein interaction detection: methods and analysis.* Int J Proteomics, 2014. **2014**: p. 147648.

41. Ekman, D., et al., *What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae?* Genome Biol, 2006. **7**(6): p. R45.
42. Gonzalez, M.W. and M.G. Kann, *Chapter 4: Protein interactions and disease.* PLoS Comput Biol, 2012. **8**(12): p. e1002819.
43. Arkin, M., *Protein-protein interactions and cancer: small molecules going in for the kill.* Curr Opin Chem Biol, 2005. **9**(3): p. 317-24.
44. Blazer, L.L. and R.R. Neubig, *Small molecule protein-protein interaction inhibitors as CNS therapeutic agents: current progress and future hurdles.* Neuropsychopharmacology, 2009. **34**(1): p. 126-41.
45. Ivanov, A.A., F.R. Khuri, and H. Fu, *Targeting protein-protein interactions as an anticancer strategy.* Trends Pharmacol Sci, 2013. **34**(7): p. 393-400.
46. Kar, G., A. Gursoy, and O. Keskin, *Human cancer protein-protein interaction network: a structural perspective.* PLoS Comput Biol, 2009. **5**(12): p. e1000601.
47. Oti, M., et al., *Predicting disease genes using protein-protein interactions.* J Med Genet, 2006. **43**(8): p. 691-8.
48. Gerrard, J.A., C.A. Hutton, and M.A. Perugini, *Inhibiting protein-protein interactions as an emerging paradigm for drug discovery.* Mini Rev Med Chem, 2007. **7**(2): p. 151-7.
49. Ryan, D.P. and J.M. Matthews, *Protein-protein interactions in human disease.* Curr Opin Struct Biol, 2005. **15**(4): p. 441-6.
50. Wells, J.A. and C.L. McClendon, *Reaching for high-hanging fruit in drug discovery at protein-protein interfaces.* Nature, 2007. **450**(7172): p. 1001-9.
51. Petta, I., et al., *Modulation of Protein-Protein Interactions for the Development of Novel Therapeutics.* Mol Ther, 2016. **24**(4): p. 707-18.
52. Gowthaman, R., E.J. Deeds, and J. Karanicolas, *Structural properties of non-traditional drug targets present new challenges for virtual screening.* J Chem Inf Model, 2013. **53**(8): p. 2073-81.
53. Falchi, F., F. Caporuscio, and M. Recanatini, *Structure-based design of small-molecule protein-protein interaction modulators: the story so far.* Future Med Chem, 2014. **6**(3): p. 343-57.
54. Metz, A., et al., *From determinants of RUNX1/ETO tetramerization to small-molecule protein-protein interaction inhibitors targeting acute myeloid leukemia.* J Chem Inf Model, 2013. **53**(9): p. 2197-202.
55. Bopp, B., et al., *Design and biological testing of peptidic dimerization inhibitors of human Hsp90 that target the C-terminal domain.* Biochim Biophys Acta, 2016. **1860**(6): p. 1043-55.
56. Golden, M.S., et al., *Comprehensive experimental and computational analysis of binding energy hot spots at the NF-kappaB essential modulator/IKKbeta protein-protein interface.* J Am Chem Soc, 2013. **135**(16): p. 6242-56.
57. Fischer, E., *Einfluss der Configuration auf die Wirkung der Enzyme.* Ber. Dtsch. Chem. Ges., 1894. **27**(3): p. 2985-2993.
58. Doppelt-Azeroual, O., et al., *Fast and automated functional classification with MED-SuMo: an application on purine-binding proteins.* Protein Sci, 2010. **19**(4): p. 847-67.
59. Moriaud, F., et al., *Identify drug repurposing candidates by mining the protein data bank.* Brief Bioinform, 2011. **12**(4): p. 336-40.
60. Chakrabarti, P. and J. Janin, *Dissecting protein-protein recognition sites.* Proteins, 2002. **47**(3): p. 334-43.
61. Cheng, A.C., et al., *Structure-based maximal affinity model predicts small-molecule druggability.* Nat Biotechnol, 2007. **25**(1): p. 71-5.

62. Smith, R.D., et al., *Exploring protein-ligand recognition with Binding MOAD.* J Mol Graph Model, 2006. **24**(6): p. 414-25.
63. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions.* Proc Natl Acad Sci U S A, 1996. **93**(1): p. 13-20.
64. Lo Conte, L., C. Chothia, and J. Janin, *The atomic structure of protein-protein recognition sites.* J Mol Biol, 1999. **285**(5): p. 2177-98.
65. Keskin, O., et al., *Principles of protein-protein interactions: what are the preferred ways for proteins to interact?* Chem Rev, 2008. **108**(4): p. 1225-44.
66. Norel, R., et al., *Shape complementarity at protein-protein interfaces.* Biopolymers, 1994. **34**(7): p. 933-40.
67. Lawrence, M.C. and P.M. Colman, *Shape complementarity at protein/protein interfaces.* J Mol Biol, 1993. **234**(4): p. 946-50.
68. Villoutreix, B.O., et al., *A leap into the chemical space of protein-protein interaction inhibitors.* Curr Pharm Des, 2012. **18**(30): p. 4648-67.
69. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces.* J Mol Biol, 1998. **280**(1): p. 1-9.
70. Ciglia, E., et al., *Resolving hot spots in the C-terminal dimerization domain that determine the stability of the molecular chaperone Hsp90.* PLoS One, 2014. **9**(4): p. e96031.
71. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface.* Science, 1995. **267**(5196): p. 383-6.
72. Metz, A., et al., *Hot spots and transient pockets: predicting the determinants of small-molecule binding to a protein-protein interface.* J Chem Inf Model, 2012. **52**(1): p. 120-33.
73. Wilkens, S., *Structure and mechanism of ABC transporters.* F1000Prime Rep, 2015. **7**: p. 14.
74. Morelli, X., R. Bourgeas, and P. Roche, *Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I).* Curr Opin Chem Biol, 2011. **15**(4): p. 475-81.
75. Sperandio, O., et al., *Rationalizing the chemical space of protein-protein interaction inhibitors.* Drug Discov Today, 2010. **15**(5-6): p. 220-9.
76. Dey, S., et al., *The subunit interfaces of weakly associated homodimeric proteins.* J Mol Biol, 2010. **398**(1): p. 146-60.
77. Larsen, T.A., A.J. Olson, and D.S. Goodsell, *Morphology of protein-protein interfaces.* Structure, 1998. **6**(4): p. 421-7.
78. Haliloglu, T., et al., *How similar are protein folding and protein binding nuclei? Examination of vibrational motions of energy hot spots and conserved residues.* Biophys J, 2005. **88**(3): p. 1552-9.
79. Janin, J., R.P. Bahadur, and P. Chakrabarti, *Protein-protein interaction and quaternary structure.* Q Rev Biophys, 2008. **41**(2): p. 133-80.
80. Liu, Y., et al., *The tetramer structure of the Nervy homology two domain, NHR2, is critical for AML1/ETO's activity.* Cancer Cell, 2006. **9**(4): p. 249-60.
81. Moreira, I.S., P.A. Fernandes, and M.J. Ramos, *Hot spots--a review of the protein-protein interface determinant amino-acid residues.* Proteins, 2007. **68**(4): p. 803-12.
82. Wichmann, C., et al., *Dimer-tetramer transition controls RUNX1/ETO leukemogenic activity.* Blood, 2010. **116**(4): p. 603-13.
83. Thorn, K.S. and A.A. Bogan, *ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions.* Bioinformatics, 2001. **17**(3): p. 284-5.

84. Gohlke, H., M. Hendlich, and G. Klebe, *Predicting binding modes, binding affinities and `hot spots' for protein-ligand complexes using a knowledge-based scoring function.* Perspectives in Drug Discovery and Design, 2000. **20**(1): p. 115-144.
85. Gohlke, H. and G. Klebe, *Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors.* Angew Chem Int Ed Engl, 2002. **41**(15): p. 2644-76.
86. Zerbe, B.S., et al., *Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces.* J Chem Inf Model, 2012. **52**(8): p. 2236-44.
87. Meireles, L.M., A.S. Domling, and C.J. Camacho, *ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W407-11.
88. Arkin, M.R. and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing towards the dream.* Nat Rev Drug Discov, 2004. **3**(4): p. 301-17.
89. Kuttner, Y.Y. and S. Engel, *Protein hot spots: the islands of stability.* J Mol Biol, 2012. **415**(2): p. 419-28.
90. Keskin, O., B. Ma, and R. Nussinov, *Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues.* J Mol Biol, 2005. **345**(5): p. 1281-94.
91. Reichmann, D., et al., *The modular architecture of protein-protein binding interfaces.* Proc Natl Acad Sci U S A, 2005. **102**(1): p. 57-62.
92. Koes, D.R. and C.J. Camacho, *Small-molecule inhibitor starting points learned from protein-protein interaction inhibitor structure.* Bioinformatics, 2012. **28**(6): p. 784-91.
93. Vasilevich, N.I., et al., *Dual mode of action of phenyl-pyrazole-phenyl (6-5-6 system)-based PPI inhibitors: alpha-helix backbone versus alpha-helix binding epitope.* MedChemComm, 2013. **4**(12): p. 1597-1603.
94. Downing, J.R., *The AML1-ETO chimaeric transcription factor in acute myeloid leukaemia: biology and clinical significance.* Br J Haematol, 1999. **106**(2): p. 296-308.
95. Elagib, K.E. and A.N. Goldfarb, *Oncogenic pathways of AML1-ETO in acute myeloid leukemia: multifaceted manipulation of marrow maturation.* Cancer Lett, 2007. **251**(2): p. 179-86.
96. Reikvam, H., et al., *Acute myeloid leukemia with the t(8;21) translocation: clinical consequences and biological implications.* J Biomed Biotechnol, 2011. **2011**: p. 104631.
97. Langabeer, S.E., et al., *Incidence of AML1/ETO fusion transcripts in patients entered into the MRC AML trials. MRC Adult Leukaemia Working Party.* Br J Haematol, 1997. **99**(4): p. 925-8.
98. Rowley, J.D., *Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia.* Ann Genet, 1973. **16**(2): p. 109-12.
99. Schanda, J., et al., *Suppression of RUNX1/ETO oncogenic activity by a small molecule inhibitor of tetramerization.* Haematologica, 2017.
100. Wichmann, C., et al., *Targeting the oligomerization domain of ETO interferes with RUNX1/ETO oncogenic activity in t(8;21)-positive leukemic cells.* Cancer Res, 2007. **67**(5): p. 2280-9.
101. Kelly, L.M. and D.G. Gilliland, *Genetics of myeloid leukemias.* Annu Rev Genomics Hum Genet, 2002. **3**: p. 179-98.

102. Miyamoto, T., I.L. Weissman, and K. Akashi, *AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation.* Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7521-6.
103. Nucifora, G., R.A. Larson, and J.D. Rowley, *Persistence of the 8;21 translocation in patients with acute myeloid leukemia type M2 in long-term remission.* Blood, 1993. **82**(3): p. 712-5.
104. Kuchenbauer, F., et al., *Identification of additional cytogenetic and molecular genetic abnormalities in acute myeloid leukaemia with t(8;21)/AML1-ETO.* Br J Haematol, 2006. **134**(6): p. 616-9.
105. Schessl, C., et al., *The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice.* J Clin Invest, 2005. **115**(8): p. 2159-68.
106. Wang, Y.Y., et al., *AML1-ETO and C-KIT mutation/overexpression in t(8;21) leukemia: implication in stepwise leukemogenesis and response to Gleevec.* Proc Natl Acad Sci U S A, 2005. **102**(4): p. 1104-9.
107. Downing, J.R., *The core-binding factor leukemias: lessons learned from murine models.* Curr Opin Genet Dev, 2003. **13**(1): p. 48-54.
108. Peterson, L.F., et al., *Acute myeloid leukemia with the 8q22;21q22 translocation: secondary mutational events and alternative t(8;21) transcripts.* Blood, 2007. **110**(3): p. 799-805.
109. Yuan, Y., et al., *AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10398-403.
110. Burel, S.A., et al., *Dichotomy of AML1-ETO functions: growth arrest versus block of differentiation.* Mol Cell Biol, 2001. **21**(16): p. 5577-90.
111. Ptasinska, A., et al., *Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding.* Leukemia, 2012. **26**(8): p. 1829-41.
112. Tenen, D.G., *Disruption of differentiation in human cancer: AML shows the way.* Nat Rev Cancer, 2003. **3**(2): p. 89-101.
113. Erickson, P.F., et al., *The ETO portion of acute myeloid leukemia t(8;21) fusion transcript encodes a highly evolutionarily conserved, putative transcription factor.* Cancer Res, 1994. **54**(7): p. 1782-6.
114. De Braekeleer, E., C. Ferec, and M. De Braekeleer, *RUNX1 translocations in malignant hemopathies.* Anticancer Res, 2009. **29**(4): p. 1031-7.
115. Ito, Y., *Oncogenic potential of the RUNX gene family: 'overview'.* Oncogene, 2004. **23**(24): p. 4198-208.
116. Speck, N.A. and D.G. Gilliland, *Core-binding factors in haematopoiesis and leukaemia.* Nat Rev Cancer, 2002. **2**(7): p. 502-13.
117. Link, K.A., F.S. Chou, and J.C. Mulloy, *Core binding factor at the crossroads: determining the fate of the HSC.* J Cell Physiol, 2010. **222**(1): p. 50-6.
118. Okuda, T., et al., *AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis.* Cell, 1996. **84**(2): p. 321-30.
119. Donghui, W., et al., *Oligomerization study of NHR3 and NHR4 domains from ETO protein involved in t(8;21)-associated acute myeloid leukemia.* Chinese Science Bulletin, 2005. **50**(9): p. 875-879.
120. Duque-Afonso, J., M. Lübbert, and M.L. Cleary, *Epigenetic Modifications Mediated by the AML1/ETO and MLL Leukemia Fusion Proteins*, in *Epigenetic Therapy of Cancer: Preclinical Models and Treatment Approaches*, M. Lübbert

and P.A. Jones, Editors. 2014, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 121-144.
121. Schanda, J., et al., *Suppression of RUNX1/ETO oncogenic activity by a small molecule inhibitor of tetramerization.* Haematologica, 2017. **102**(5): p. e170-e174.
122. Zhen, T., et al., *Targeting of AML1-ETO in t(8;21) leukemia by oridonin generates a tumor suppressor-like protein.* Sci Transl Med, 2012. **4**(127): p. 127ra38.
123. Metz, A., E. Ciglia, and H. Gohlke, *Modulating protein-protein interactions: from structural determinants of binding to druggability prediction to application.* Curr Pharm Des, 2012. **18**(30): p. 4630-47.
124. Irwin, J.J. and B.K. Shoichet, *ZINC--a free database of commercially available compounds for virtual screening.* J Chem Inf Model, 2005. **45**(1): p. 177-82.
125. Metz, A., *Predicting and Exploiting the Determinants of Protein/Protein Interactions to Identify Low-Molecular Inhibitors of RUNX1-ETO Tetramerization (PhD Thesis)*. 2014, Düsseldorf: Heinrich Heine Universität.
126. Bartel, Y., M. Grez, and C. Wichmann, *Interference with RUNX1/ETO leukemogenic function by cell-penetrating peptides targeting the NHR2 oligomerization domain.* Biomed Res Int, 2013. **2013**: p. 297692.
127. Becker, Y., *Inhibierung der leukämischen Funktion des Fusionsproteins AML1/ETO durch Interferenz mit der Oligomerisierungsdomäne.* 2011.
128. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. **2**(9): p. 2212-21.
129. Ericsson, U.B., et al., *Thermofluor-based high-throughput stability optimization of proteins for structural studies.* Anal Biochem, 2006. **357**(2): p. 289-98.
130. Forneris, F., et al., *ThermoFAD, a Thermofluor-adapted flavin ad hoc detection system for protein folding and ligand binding.* FEBS J, 2009. **276**(10): p. 2833-40.
131. Joubert, M.K., et al., *Classification and characterization of therapeutic antibody aggregates.* J Biol Chem, 2011. **286**(28): p. 25118-33.
132. Duhr, S. and D. Braun, *Why molecules move along a temperature gradient.* Proc Natl Acad Sci U S A, 2006. **103**(52): p. 19678-82.
133. Jerabek-Willemsen, M., et al., *Molecular interaction studies using microscale thermophoresis.* Assay Drug Dev Technol, 2011. **9**(4): p. 342-53.
134. Seidel, S.A., et al., *Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions.* Methods, 2013. **59**(3): p. 301-15.
135. Wienken, C.J., et al., *Protein-binding assays in biological liquids using microscale thermophoresis.* Nat Commun, 2010. **1**: p. 100.
136. Hwang, T.L. and A.J. Shaka, *Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients.* Journal of Magnetic Resonance Series A, 1995. **112**(2): p. 275-279.
137. Otrelo-Cardoso, A.R., et al., *Biochemical, stabilization and crystallization studies on a molecular chaperone (PaoD) involved in the maturation of molybdoenzymes.* PLoS One, 2014. **9**(1): p. e87295.
138. Mayer, M. and B. Meyer, *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor.* Journal of the American Chemical Society, 2001. **123**(25): p. 6108-6117.
139. Viegas, A., et al., *Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding.* Journal of Chemical Education, 2011. **88**(7): p. 990-994.
140. Schuck, P., *Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling.* Biophys J, 2000. **78**(3): p. 1606-19.

141. Hurton, T., et al. *SEDNTERP* 2016 [cited 2016 25 November]; Available from: http://bitcwiki.sr.unh.edu/index.php/Main_Page.
142. Durchschlag, H. and P. Zipper, *Calculation of the partial volume of organic compounds and polymers*, in *Ultracentrifugation*, M.D. Lechner, Editor. 1994, Steinkopff: Darmstadt. p. 20-39.
143. Brautigam, C.A., *Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data.* Methods Enzymol, 2015. **562**: p. 109-33.
144. Maestro, v., Schrodinger, LLC, New York, 2013, *Maestro, version 9.5, Schrodinger, LLC, New York, 2013*.
145. Frisch, M.J., et al., *Gaussian 09.* Wallingford, CT, USA, 2009.
146. Martinez, L., et al., *PACKMOL: a package for building initial configurations for molecular dynamics simulations.* Journal of Computational Chemistry, 2009. **30**(13): p. 2157-2164.
147. Schafmeister, C.E.A.F., W.S. Ross, and V. Romanovski, *LEaP.* University of California, San Francisco, 1995.
148. Case, D.A., et al., *AMBER 12.* University of California, San Francisco., 2012.
149. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water.* Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
150. Wang, J.M., et al., *Development and testing of a general amber force field.* Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174.
151. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP Model.* Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
152. Salomon-Ferrer, R., et al., *Routine microsecond molecular dynamics simulations with Amber on GPUs. 2. Explicit solvent particle mesh ewald.* Journal of Chemical Theory and Computation, 2013. **9**(9): p. 3878-3888.
153. Darden, T., D.M. York, and L.G. Pedersen, *Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems.* Journal of chemical physics, 1993. **98**(12): p. 10089-10092.
154. York, D.M., T.A. Darden, and L.G. Pedersen, *The effect of long-range electrostatic interactions in simulations of macromolecular crystals - a comparison of the ewald and truncated list methods.* Journal of Chemical Physics, 1993. **99**(10): p. 8345-8348.
155. Darden, T., D. Pearlman, and L.G. Pedersen, *Ionic charging free energies: spherical versus periodic boundary conditions.* Journal of Chemical Physics, 1998. **109**(24): p. 10921-10935.
156. Bradbrook, G.M., et al., *X-ray and molecular dynamics studies of concanavalin- A glucoside and mannoside complexes - Relating structure to thermodynamics of binding.* Journal of the Chemical Society, Faraday Transactions, 1998. **94**(11): p. 1603-1611.
157. Joung, I.S. and T.E. Cheatham, *Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations.* Journal of Physical Chemistry B, 2008. **112**(30): p. 9020-9041.
158. Roe, D.R. and T.E. Cheatham, *PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data.* Journal of Chemical Theory and Computation, 2013. **9**(7): p. 3084-3095.
159. Case, D., et al., *Amber 14.* 2014.
160. Sun, X.J., et al., *A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis.* Nature, 2013. **500**(7460): p. 93-7.
161. *Schrödinger Release 2018-2: Maestro*. 2018, Schrödinger, LLC: New York, NY.
162. Frisch, M.J., et al., *Gaussian 16 Rev. B.01*. 2016: Wallingford, CT.

163. Martinez, L., et al., *PACKMOL: a package for building initial configurations for molecular dynamics simulations.* J Comput Chem, 2009. **30**(13): p. 2157-64.
164. Case, D., et al., *Amber 2017, University of California, San Francisco*. 2017.
165. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J Chem Theory Comput, 2015. **11**(8): p. 3696-713.
166. Wang, J., et al., *Development and testing of a general amber force field.* J Comput Chem, 2004. **25**(9): p. 1157-74.
167. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model.* The Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
168. Cornell, W.D., et al., *Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation.* Journal of the American Chemical Society, 1993. **115**(21): p. 9620-9631.
169. Joung, I.S. and T.E. Cheatham, 3rd, *Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations.* J Phys Chem B, 2008. **112**(30): p. 9020-41.
170. Frieg, B., et al., *Molecular Mechanisms of Glutamine Synthetase Mutations that Lead to Clinically Relevant Pathologies.* PLoS Comput Biol, 2016. **12**(2): p. e1004693.
171. Bhatia, S., et al., *Targeting HSP90 dimerization via the C-terminus is effective in imatinib resistant CML and lacks heat shock response.* Blood, 2018.
172. Frieg, B., D. Haussinger, and H. Gohlke, *Towards Restoring Catalytic Activity of Glutamine Synthetase With a Clinically Relevant Mutation.*
173. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems.* The Journal of chemical physics, 1993. **98**(12): p. 10089-10092.
174. Salomon-Ferrer, R., et al., *Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald.* Journal of Chemical Theory and Computation, 2013. **9**(9): p. 3878-3888.
175. Hopkins, C.W., et al., *Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning.* J Chem Theory Comput, 2015. **11**(4): p. 1864-74.
176. Flower, D.R., *On the properties of bit string-based measures of chemical similarity.* Journal of Chemical Information and Computer Sciences, 1998. **38**(3): p. 379-386.
177. Bornhorst, J.A. and J.J. Falke, *[16] Purification of proteins using polyhistidine affinity tags*, in *Methods in enzymology*. 2000, Elsevier. p. 245-254.
178. Pantoliano, M.W., et al., *High-density miniaturized thermal shift assays as a general strategy for drug discovery.* J Biomol Screen, 2001. **6**(6): p. 429-40.
179. Wan, K.F., et al., *Differential scanning fluorimetry as secondary screening platform for small molecule inhibitors of Bcl-XL.* Cell Cycle, 2009. **8**(23): p. 3943-52.
180. Muller, M. and A. Haeberli, *pH-dependent formation of ethylenediaminetetraacetic acid supramolecular aggregates.* FEBS Lett, 1994. **340**(1-2): p. 17-21.
181. Han, S., E. Mathias, and Y. Ba, *Proton NMR determination of Mg2+ and Ca2+ concentrations using tetrasodium EDTA complexes.* J. Chem, 2007. **1**: p. 1-5.
182. Anastassopoulou, J., et al., *The Role of β-Antagonists on the Structure of Human Bone-A Spectroscopic Study.* Infrared Spectroscopy-Life and Biomedical Sciences (Ed. T. Theophanides), 2012: p. 259-271.
183. Raaflaub, J., *Applications of metal buffers and metal indicators in biochemistry.* Methods Biochem Anal, 1956. **3**: p. 301-25.

184. Harrison, D. and C. Long, *The calcium content of human erythrocytes.* The Journal of physiology, 1968. **199**(2): p. 367.
185. Haugland, R.P., et al., *Merocyanide dye protein stains*. 1996, Molecular Probes Inc., assignee: World patent WO 9636882.
186. Joung, I.S. and T.E. Cheatham, *Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations.* The Journal of Physical Chemistry B, 2008. **112**(30): p. 9020-9041.
187. Åqvist, J., *Ion-water interaction potentials derived from free energy perturbation simulations.* The Journal of Physical Chemistry, 1990. **94**(21): p. 8021-8024.
188. Case, D. and P. Kollman, *AmberTools 12.* University of California, San Francisco, 2012.
189. Kohagen, M., P.E. Mason, and P. Jungwirth, *Accurate description of calcium solvation in concentrated aqueous solutions.* J Phys Chem B, 2014. **118**(28): p. 7902-9.
190. Corradini, D., M. Rovere, and P. Gallo, *The Widom line and dynamical crossover in supercritical water: Popular water models versus experiments.* J Chem Phys, 2015. **143**(11): p. 114502.
191. Paschek, D., *Temperature dependence of the hydrophobic hydration and interaction of simple solutes: an examination of five popular water models.* J Chem Phys, 2004. **120**(14): p. 6674-90.
192. Steinberg, T.H., et al., *SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein.* Anal Biochem, 1996. **239**(2): p. 223-37.
193. Vivoli, M., et al., *Determination of protein-ligand interactions using differential scanning fluorimetry.* J Vis Exp, 2014(91): p. 51809.
194. Huynh, K. and C.L. Partch, *Analysis of protein stability and ligand interactions by thermal shift assay.* Curr Protoc Protein Sci, 2015. **79**: p. 28 9 1-14.
195. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nature protocols, 2007. **2**(9): p. 2212.
196. Ricketts, J.H. and G.A. Head, *A five-parameter logistic equation for investigating asymmetry of curvature in baroreflex studies.* Am J Physiol, 1999. **277**(2 Pt 2): p. R441-54.
197. Wink, D.J., *Spin-lattice relaxation times in 1H NMR spectroscopy.* Journal of Chemical Education, 1989. **66**(10): p. 810.
198. Abildgaard, J., et al., *Deuterium isotope effects on 15N backbone chemical shifts in proteins.* J Biomol NMR, 2009. **44**(3): p. 119-26.
199. Wallerstein, J. and M. Akke, *What's in Your Buffer? Minute additions of DMSO affect protein dynamics measurements by NMR relaxation experiments via significant changes in solvent viscosity.* Chemphyschem, 2018.
200. Zhang, Y.Z., Y. Paterson, and H. Roder, *Rapid amide proton exchange rates in peptides and proteins measured by solvent quenching and two-dimensional NMR.* Protein Sci, 1995. **4**(4): p. 804-14.
201. Kang, I., et al., *Mannan-binding lectin (MBL)-associated plasma protein present in human urine inhibits calcium oxalate crystal growth.* FEBS Lett, 1999. **462**(1-2): p. 89-93.
202. Roti Roti, J.L., et al., *DNase I sensitivity of nuclear DNA measured by flow cytometry.* Cytometry, 1985. **6**(2): p. 101-8.
203. Sando, J.J., et al., *Activation of thymocyte glucocorticoid receptors to the steroid binding form. The roles of reduction agents, ATP, and heat-stable factors.* J Biol Chem, 1979. **254**(11): p. 4779-89.

204. Hovey, J.K., L.G. Hepler, and P.R. Tremaine, *Thermodynamics of aqueous EDTA systems: Apparent and partial molar heat capacities and volumes of aqueous EDTA4-, HEDTA3-, H2EDTA2-, NaEDTA3-, and KEDTA3-at 25° C. Relaxation effects in mixed aqueous electrolyte solutions and calculations of temperature dependent equilibrium constants.* Canadian journal of chemistry, 1988. **66**(4): p. 881-896.
205. Cherif, S., et al., *A newly high alkaline lipase: an ideal choice for application in detergent formulations.* Lipids Health Dis, 2011. **10**: p. 221.
206. Saeki, K., et al., *Detergent alkaline proteases: enzymatic properties, genes, and crystal structures.* J Biosci Bioeng, 2007. **103**(6): p. 501-8.
207. Patel, T.R., D.J. Winzor, and D.J. Scott, *Analytical ultracentrifugation: A versatile tool for the characterisation of macromolecular complexes in solution.* Methods, 2016. **95**: p. 55-61.
208. MacGregor, I.K., A.L. Anderson, and T.M. Laue, *Fluorescence detection for the XLI analytical ultracentrifuge.* Biophys Chem, 2004. **108**(1-3): p. 165-85.
209. Chen, C., et al., *Dynamic Equilibria of Short-Range Electrostatic Interactions at Molecular Interfaces of Protein-DNA Complexes.* J Phys Chem Lett, 2015. **6**(14): p. 2733-7.
210. Cheng, S.Y., et al., *Data supporting beta-amyloid dimer structural transitions and protein-lipid interactions on asymmetric lipid bilayer surfaces using MD simulations on experimentally derived NMR protein structures.* Data Brief, 2016. **7**: p. 658-72.
211. Mobasheri, M., et al., *Solubilization Behavior of Polyene Antibiotics in Nanomicellar System: Insights from Molecular Dynamics Simulation of the Amphotericin B and Nystatin Interactions with Polysorbate 80.* Molecules, 2015. **21**(1): p. E6.
212. Mountain, R.D. and A.H. Harvey, *Molecular Dynamics Evaluation of Dielectric-Constant Mixing Rules for H2O-CO2 at Geologic Conditions.* J Solution Chem, 2015. **44**(11): p. 2179-2193.
213. Panneerselvam, S., et al., *A Combined Molecular Docking/Dynamics Approach to Probe the Binding Mode of Cancer Drugs with Cytochrome P450 3A4.* Molecules, 2015. **20**(8): p. 14915-35.
214. Satoh, M., et al., *Multiple binding modes of a small molecule to human Keap1 revealed by X-ray crystallography and molecular dynamics simulation.* FEBS Open Bio, 2015. **5**: p. 557-70.
215. Boonyarattanakalin, K., et al., *Influence of Ethanol as a Co-Solvent in Cyclodextrin Inclusion Complexation: A Molecular Dynamics Study.* Sci Pharm, 2015. **83**(2): p. 387-99.
216. Keshwani, N., et al., *The role of cross-chain ionic interactions for the stability of collagen model peptides.* Biophys J, 2013. **105**(7): p. 1681-8.
217. Carballo-Pacheco, M. and B. Strodel, *Advances in the Simulation of Protein Aggregation at the Atomistic Scale.* J Phys Chem B, 2016. **120**(12): p. 2991-9.
218. Wildman, J., et al., *General Force-Field Parametrization Scheme for Molecular Dynamics Simulations of Conjugated Materials in Solution.* J Chem Theory Comput, 2016. **12**(8): p. 3813-24.
219. Dror, R.O., et al., *Biomolecular simulation: a computational microscope for molecular biology.* Annu Rev Biophys, 2012. **41**: p. 429-52.
220. Elber, R., *Long-timescale simulation methods.* Curr Opin Struct Biol, 2005. **15**(2): p. 151-6.

221. Dorn, T., A.-C. Chamayou, and C. Janiak, *Hydrophilic interior between hydrophobic regions in inverse bilayer structures of cation–1, 1'-binaphthalene-2, 2'-diyl phosphate salts.* New journal of chemistry, 2006. **30**(2): p. 156-167.
222. *GESTIS Substance Database*. Insitute for Occupational Safety and Health of the German Social Accident Insurance: 1. February 2016.
223. Collins, K.D., *Charge density-dependent strength of hydration and biological structure.* Biophys J, 1997. **72**(1): p. 65-76.
224. Aziz, E.F., et al., *Cation-specific interactions with carboxylate in amino acid and acetate aqueous solutions: X-ray absorption and ab initio calculations.* J Phys Chem B, 2008. **112**(40): p. 12567-70.
225. Fedorov, M.V., J.M. Goodman, and S. Schumm, *To switch or not to switch: the effects of potassium and sodium ions on alpha-poly-L-glutamate conformations in aqueous solutions.* J Am Chem Soc, 2009. **131**(31): p. 10854-6.
226. Hess, B. and N.F. van der Vegt, *Cation specific binding with protein surface charges.* Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13296-300.
227. Vrbka, L., et al., *Quantification and rationalization of the higher affinity of sodium over potassium to protein surfaces.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15440-4.
228. Jager, C.M., et al., *Self-assembly of structurally persistent micelles is controlled by specific-ion effects and hydrophobic guests.* Langmuir, 2010. **26**(13): p. 10460-6.
229. Rosenlehner, K., et al., *Sodium effect on self-organization of amphiphilic carboxylates: formation of structured micelles and superlattices.* Chemistry, 2010. **16**(31): p. 9544-54.
230. Elcock, A.H., *The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins.* J Mol Biol, 1998. **284**(2): p. 489-502.
231. Dang, L.X., *Temperature dependence of interactions of an ion pair in water: A molecular dynamics study.* The Journal of Chemical Physics, 1992. **97**(3): p. 1919-1921.
232. Bhatia, S., et al., *Targeting HSP90 dimerization via the C-terminus is effective in imatinib resistant CML and lacks heat shock response induction.* Blood, 2018: p. blood-2017-10-810986.
233. Gohlke, H., et al., *Binding region of alanopine dehydrogenase predicted by unbiased molecular dynamics simulations of ligand diffusion.* J Chem Inf Model, 2013. **53**(10): p. 2493-8.
234. DeSantis, K., et al., *Use of differential scanning fluorimetry as a high-throughput assay to identify nuclear receptor ligands.* Nucl Recept Signal, 2012. **10**: p. e002.
235. Pilkington S, et al., *Amyloid fibrils as a nanoscaffold for enzyme immobilization.* Biotechnology Progress, 2010. **26**(1): p. 93-100.
236. Rice, L.B., et al., *Structural and Regulatory Changes in PBP4 Trigger Decreased beta-Lactam Susceptibility in Enterococcus faecalis.* MBio, 2018. **9**(2).
237. Muley, L., et al., *Enhancement of hydrophobic interactions and hydrogen bond strength by cooperativity: synthesis, modeling, and molecular dynamics simulations of a congeneric series of thrombin inhibitors.* J Med Chem, 2010. **53**(5): p. 2126-35.
238. Chen, J., N. Sawyer, and L. Regan, *Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area.* Protein Science, 2013. **22**(4): p. 510-515.
239. J. Chen, E.L., S. J. Swamidass, D. Wang, P. Baldi, *ChemDB Update-Full-Text Search and Virtual Chemical Space.* Bioinformatics, 2007. **23**: p. 2348-2351.