

# **Synthesis and Biophysical Characterization of New Glycosyltransferase Inhibitors Designed by Computer-Aided Strategies**

Thesis

submitted in fulfillment of the requirements  
of the degree Doctor rer. nat. of the Department of Chemistry  
Faculty of Sciences, University of Hamburg

submitted by  
Claas Strecker

**2019**

This thesis was conducted at the Institute of Organic Chemistry from Mai 2015 to Mai 2019.

I would like to thank Prof. Dr. Bernd Meyer for all the academic freedom granted during this PhD thesis.

I would like to thank Prof. Dr. Christian B. W. Stark for being the second reviewer.

1<sup>st</sup> Reviewer: Prof. Dr. Bernd Meyer

2<sup>nd</sup> Reviewer: Prof. Dr. Christian B. W. Stark


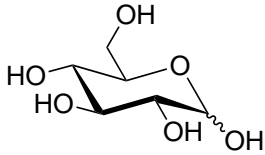

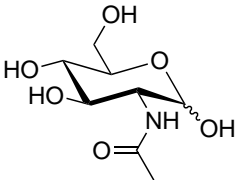

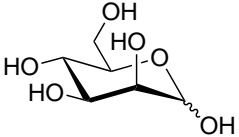

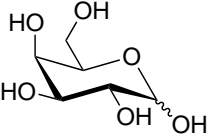

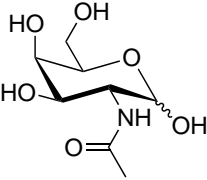

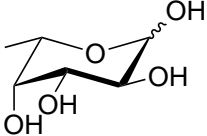
Date of defense: 19.07.2019

Date of print approval: 19.07.2019

## Abbreviations

|                |  |
|----------------|--|
| <b>Ac</b>      | Acetyl   |
| <b>AUC ROC</b> | Area Under the Curve of the Receiver Operating Curve                       |
| <b>BEDROC</b>  | Boltzmann-Enhanced Discrimination of the Receiver Operating Characteristic |
| <b>Bn</b>      | Benzyl   |
| <b>Boc</b>     | <i>tert</i> -Butyloxycarbonyl  |
| <b>DIPEA</b>   | <i>N,N</i> -Diisopropylethylamine  |
| <b>DMSO</b>    | Dimethyl Sulfoxide   |
| <b>DMF</b>     | <i>N,N</i> -Dimethylformamide  |
| <b>DUD-E</b>   | Directory of Useful Decoys-Enhanced  |
| <b>ESI</b>     | Electrospray Ionization  |
| <b>Et</b>      | Ethyl  |
| <b>FUT</b>     | Fucosyltransferase   |
| <b>FID</b>     | Free Induction Decay   |
| <b>GDP</b>     | Guanosine Diphosphate  |
| <b>GTA</b>     | Human Blood Group A Galactosyltransferase                                  |
| <b>GTB</b>     | Human Blood Group B Galactosyltransferase                                  |
| <b>HPLC</b>    | High-Performance Liquid Chromatography                                     |
| <b>HTS</b>     | High-Throughput Screening  |
| <b>LE</b>      | Ligand Efficiency  |
| <b>MD</b>      | Molecular Dynamics   |
| <b>MES</b>     | 2-( <i>N</i> -Morpholino)ethanesulfonic acid                               |
| <b>Me</b>      | Methyl   |
| <b>MM-GBSA</b> | Molecular Mechanics-Generalized Born Surface Area                          |
| <b>MS</b>      | Mass Spectrometry  |
| <b>NMR</b>     | Nuclear Magnetic Resonance   |
| <b>PDB</b>     | Protein Data Bank  |
| <b>RMSD</b>    | Root Mean Square Deviation   |
| <b>RP</b>      | Reversed Phase   |
| <b>STD</b>     | Saturation Transfer Difference   |
| <b>SPR</b>     | Surface Plasmon Resonance  |
| <b>TMSP</b>    | Trimethylsilylpropanoic acid   |
| <b>UDP</b>     | Uridine Diphosphate  |

## Abbreviations

| Monosaccharide          | Abbreviation | Symbol  | Structure   |
|-------------------------|--------------|---|---|
| D-Glucose               | Glc          |    |    |
| D-N-Acetylglucosamine   | GlcNAc       |    |    |
| D-Mannose               | Man          |    |    |
| D-Galactose             | Gal          |    |    |
| D-N-Acetylgalactosamine | GalNAc       |    |   |
| L-Fucose                | Fuc          |  |  |



### *Abbreviations*

---

| Amino acid    | Three letter code | One letter code |
|---------------|-------------------|-----------------|
| Alanine       | Ala               | A               |
| Arginine      | Arg               | R               |
| Asparagine    | Asn               | N               |
| Aspartic acid | Asp               | D               |
| Cysteine      | Cys               | C               |
| Glutamic acid | Glu               | E               |
| Glutamine     | Gln               | Q               |
| Glycine       | Gly               | G               |
| Histidine     | His               | H               |
| Isoleucine    | Ile               | I               |
| Leucine       | Leu               | L               |
| Lysine        | Lys               | K               |
| Methionine    | Met               | M               |
| Phenylalanine | Phe               | F               |
| Proline       | Pro               | P               |
| Serine        | Ser               | S               |
| Threonine     | Thr               | T               |
| Tryptophane   | Trp               | W               |
| Tyrosine      | Tyr               | Y               |
| Valine        | Val               | V               |

---

**Table of Contents**

|       |   |    |
|-------|---|----|
| 1     | Abstract . . . . .  | 1  |
| 2     | Kurzfassung . . . . .   | 4  |
| 3     | Introduction . . . . .  | 7  |
| 3.1   | Drug Discovery and Development . . . . .  | 7  |
| 3.1.1 | Fragment-Based Drug Discovery . . . . .   | 8  |
| 3.1.2 | Computer-Aided Drug Design . . . . .  | 10 |
| 3.2   | Glycosylation . . . . .   | 12 |
| 4     | General Objective . . . . .   | 14 |
| 5     | Results and Discussion . . . . .  | 15 |
| 5.1   | Development of New Inhibitors of Fucosyltransferase 8 . . . . .                                 | 15 |
| 5.1.1 | Introduction . . . . .  | 15 |
| 5.1.2 | Objective . . . . .   | 17 |
| 5.1.3 | Fragment-Based <i>De Novo</i> Design of Inhibitors of Fucosyltransferase 8 . . . . .            | 17 |
| 5.1.4 | Synthesis . . . . .   | 21 |
| 5.1.5 | Activity Assay . . . . .  | 25 |
| 5.1.6 | STD NMR . . . . .   | 26 |
| 5.1.7 | Identification of a New Putative Binding Site of Fucosyltransferase 8 . . . . .                 | 27 |
| 5.1.8 | Conclusion . . . . .  | 30 |
| 5.2   | Ensemble Docking of Renin . . . . .   | 32 |
| 5.2.1 | Introduction . . . . .  | 32 |
| 5.2.2 | Objective . . . . .   | 33 |
| 5.2.3 | Introduction of Tools and Metrics and Development of a Test System . . . . .                    | 34 |
| 5.2.4 | Evaluation of Clustering Approaches for Ensemble Docking . . . . .                              | 41 |
| 5.2.5 | Exhaustive Sampling of All Possible Ensembles . . . . .   | 48 |
| 5.2.6 | Conclusion . . . . .  | 50 |
| 5.3   | Optimized Inhibitors for Human Blood Group B Galactosyltransferase . . . . .                    | 52 |
| 5.3.1 | Introduction . . . . .  | 52 |
| 5.3.2 | Objective . . . . .   | 54 |
| 5.3.3 | Design of Inhibitors of Human Blood Group B Galactosyltransferase by Fragment Growing . . . . . | 54 |

*Table of Contents*

---

|       |  |     |
|-------|--|-----|
| 5.3.4 | Synthesis . . . . .  | 56  |
| 5.3.5 | Competitive STD NMR . . . . .  | 59  |
| 5.3.6 | Optimization of Ligand <b>37</b> . . . . .   | 62  |
| 5.3.7 | Activity Assay . . . . .   | 63  |
| 5.4   | Conclusion . . . . .   | 64  |
| 6     | Experimental Procedures . . . . .  | 65  |
| 6.1   | Development of New Inhibitors of Fucosyltransferase 8 . . . . .  | 65  |
| 6.1.1 | Molecular Modeling . . . . .   | 65  |
| 6.1.2 | Synthesis . . . . .  | 65  |
| 6.1.3 | Inhibition Assays . . . . .  | 75  |
| 6.2   | Ensemble Docking of Renin . . . . .  | 77  |
| 6.2.1 | Preparation of Crystal Structures, Active Ligands and Decoys . . . . .                                     | 77  |
| 6.2.2 | Construction of RMSD matrices and Multidimensional Scaling . . . . .                                       | 77  |
| 6.2.3 | Clustering . . . . .   | 78  |
| 6.2.4 | Docking . . . . .  | 78  |
| 6.2.5 | MD Simulations . . . . .   | 78  |
| 6.2.6 | Evaluation of Docking Performance . . . . .  | 79  |
| 6.3   | Optimized Inhibitors for Human Blood Group B Galactosyltransferase . . . . .                               | 80  |
| 6.3.1 | Molecular Modeling . . . . .   | 80  |
| 6.3.2 | Synthesis . . . . .  | 80  |
| 6.3.3 | Inhibition Assays . . . . .  | 90  |
| 7     | References . . . . .   | 92  |
| 8     | Hazards . . . . .  | 104 |
| 9     | Appendix . . . . .   | 109 |
| 9.1   | Chemotypes of Renin Inhibitors . . . . .   | 109 |
| 9.2   | Python Script for the Construction of RMSD Matrices . . . . .  | 115 |
| 9.3   | Docking Performance of Individual Crystal Structures of Renin . . . . .                                    | 116 |
| 9.4   | Docking Performance of Ensembles of Crystal Structures Constructed by Hierarchical Clustering . . . . .    | 118 |
| 9.5   | Docking Performance of Ensembles of Crystal Structures Constructed by <i>K</i> -means Clustering . . . . . | 119 |
| 9.6   | Docking Performance of Ensembles Constructed from MD-derived Structures . . . . .                          | 120 |
| 10    | Danksagung . . . . .   | 125 |

*Table of Contents*

---

11 Affidavits . . . . . 126

## 1 Abstract

This PhD thesis is composed of three projects that focus on new aspects of the computer-aided design and synthesis of glycosyltransferase inhibitors. The first project demonstrates the application of fragment-based *de novo* design to generate new chemical entities that inhibit fucosyltransferase 8. The second project illustrates how the success of docking can be dramatically improved by the rational selection of multiple protein structures to incorporate effects of protein flexibility as demonstrated on a test system of the aspartyl protease renin. The third project realizes the concept of fragment growing to successfully improve the binding affinity of a fragment, that has previously been shown to inhibit human blood group B galactosyltransferase, by a factor of three.

**1. Development of New Inhibitors of Fucosyltransferase 8:** Core fucosylation of *N*-glycans is catalyzed by fucosyltransferase 8 and has been associated with the progress of various types of cancer such as melanoma, non-small cell lung cancer, and aggressive prostate cancer. Hence, inhibitors of fucosyltransferase 8 represent an efficient tool for more detailed studies on the physiological role of fucosyltransferase 8 and might demonstrate therapeutic potential as well. Previously reported inhibitors of fucosyltransferase 8 feature non drug-like properties such as a high polarity or molecular weight that prevent cellular uptake. Therefore, a fragment-based strategy that allows for more control over molecular properties was embraced in this work. A fragment library of more than 700,000 fragments was docked against the donor binding site of fucosyltransferase 8. Two interesting fragments were identified and subsequently trimmed and linked *in silico*. The resulting ligand that features a pyrazole disubstituted with carboxamide functionality as a surrogate of the guanine moiety was synthesized by a convergent synthetic route starting from readily available materials. STD NMR revealed a  $K_d$  of 1.65 mM for this ligand. Even though the ligand is only of low-affinity, this represents the first low-molecular weight starting point for the development of inhibitors of fucosyltransferase 8 with drug-like properties. Furthermore, the discovery of a putative allosteric binding site is described. An *in silico* assessment indicates that this putative binding site might be more easy to target than the donor binding site of fucosyltransferase 8. Furthermore, an MD simulation indicates that this putative binding site is accessible from the acceptor site of fucosyltransferase 8. For ligands that bind to this putative binding site and extend into the acceptor site, disruption of the enzymatic activity of fucosyltransferase 8 is conceivable.

**2. Ensemble Docking of Renin:** Ensemble docking is a simple yet promising method to consider protein flexibility in docking studies: Ligands are docked into multiple protein structures to incorporate effects of protein flexibility, and the results are subsequently merged. Here, the potential of ensemble docking was studied using renin, an attractive target for antihypertensive drugs, as a test case. Renin is a suitable test system as a relevant number of crystal structures are available and the binding site of enzyme is known to feature significant flexibility. For this study, crystal structures as well as MD-derived structures from a total of 500 ns of simulation time were evaluated. Selection of protein structures is a significant factor for ensemble docking, e.g. it is obviously undesirable to use an ensemble of protein structures that are highly alike. In this work, two different clustering algorithms, *k*-means and hierarchical clustering with average linkage, were compared for ensemble definition. Three disciplines were evaluated to assess the docking performance: Binding pose prediction, the ability to rank actives above nonactives (screening utility), and scoring accuracy. Overall, *k*-means clustering yielded ensembles with superior performance. Interestingly, ensembles of MD-derived structures constructed by *k*-means clustering performed on average better than 75% of any individual crystal structure in terms of scoring accuracy at all inspected ensemble sizes. Finally, the full potential of ensemble docking was unraveled by computing the docking performance of all possible ensembles that are composed of up to four crystal structures of renin (more than 160,000 combinations). The resulting data makes a convincing case for the use of ensemble docking. For ensembles consisting of four crystal structures an average improvement of more than 70% in binding pose prediction, 11% in screening utility (as measured by the area under the curve of the receiver operating curve (AUC ROC) metric), and 55% in scoring accuracy (as measured by  $R^2$ ) was observed in comparison to the performance of individual crystal structures.

**3. Optimized Inhibitors for Human Blood Group B Galactosyltransferase:** Human blood group B galactosyltransferase (GTB) catalyzes the galactosylation of the H antigen and is responsible for the formation of the blood group antigen of phenotype B. Blood type subgroups have been repeatedly linked to an increased occurrence of diseases. For example, individuals with blood group phenotype B have a highly increased incidence rate for pancreatic cancer. 3-Phenyl-5-(piperazin-1-yl)-1,2,4-thiadiazole **35** has previously been described to inhibit GTB with a  $K_i$  of 800  $\mu$ M. In this work, a fragment growing approach was pursued to improve the binding affinity of fragment **35**. Enlarging the phenyl moiety of fragment **35** to a naphthyl moiety resulted in ligand **37** that showed a threefold improvement in binding affinity ( $K_i = 271 \mu$ M) as evaluated by competitive STD NMR. Substitution of the phenyl moiety of fragment **35** with

OBn-substituents in 2- and 3-position (as realized in ligands **38** and **39**) that were anticipated to exploit  $\pi$ - $\pi$  interactions with His233 and Trp300 proved not to be promising. A continuative attempt to substitute the piperazine moiety of ligand **37** resulted in ligand **62** that showed no measurable affinity for GTB under the given experimental conditions.

## 2 Kurzfassung

Diese Doktorarbeit ist in drei Projekte gegliedert, die sich mit dem computergestützten Design und der Synthese von Glycosyltransferase-Inhibitoren beschäftigen. Im ersten Projekt wird aufgezeigt, wie mittels einer fragment-basierten *de novo* Design-Strategie neuartige Inhibitoren der Fucosyltransferase 8 entwickelt werden können. Im zweiten Projekt wird dargelegt, wie der Erfolg von Docking-Studien dramatisch verbessert werden kann, indem mehrere Protein-Strukturen verwendet werden und somit Effekte der Protein-Flexibilität berücksichtigt werden. Dies wird beispielhaft an der Aspartyl-Protease Renin illustriert. Im dritten Projekt wird demonstriert, wie ein zuvor publizierter Fragment-Hit für die humane Galactosyltransferase B erfolgreich optimiert werden kann und somit eine Verbesserung der Bindungsaffinität um den Faktor drei erreicht werden konnte.

**1. Entwicklung Neuer Inhibitoren der Fucosyltransferase 8:** Die *core* Fucosylierung von *N*-Glycanen wird von der Fucosyltransferase 8 katalysiert und steht im Zusammenhang mit verschiedenen Krebsarten wie z.B. schwarzem Hautkrebs, nichtkleinzelligen Bronchialkarzinomen und aggressivem Prostata-Krebs. Folglich stellen Inhibitoren der Fucosyltransferase 8 ein wertvolles Werkzeug für tieferegehende Studien zur physiologischen Rolle der Fucosyltransferase 8 dar und besitzen weiterhin therapeutisches Potential. Zuvor publizierte Inhibitoren der Fucosyltransferase 8 weisen für Wirkstoffe unübliche Stoffeigenschaften auf wie z.B. eine hohe Polarität und molekulare Masse. Diese Eigenschaften verhindern die Zellaufnahme und mindern die Nützlichkeit dieser Inhibitoren signifikant ab. Aus diesen Gründen wurde in dieser Arbeit ein fragment-basierter Ansatz verfolgt. Ein solcher fragment-basierter Ansatz erlaubt eine stärkere Kontrolle über die zuvor erwähnten Stoffeigenschaften, die im Zusammenhang mit der Zellaufnahme stehen. Hierfür wurde eine Fragmentbibliothek mit über 700.000 Fragmenten in die Donorbindungstasche der Fucosyltransferase 8 gedockt. Zwei interessante Fragmente wurden identifiziert und *in silico* zu einem Liganden verbunden. Der resultierende Ligand, der als Surrogat für den Guanin-Rest einen zweifach mit Carboxamid-Gruppen substituierten Pyrazol-Kern präsentiert, wurde mittels einer konvergenten Route ausgehend von leicht verfügbaren Startmaterialien synthetisiert. Mittels STD NMR wurde für den Liganden **8** eine Bindungskonstante von 1.65 mM bestimmt. Obwohl der Ligand nur eine schwache Affinität aufweist, stellt dies den ersten Startpunkt für die Entwicklung zellgängiger Inhibitoren der Fucosyltransferase 8 dar. Weiterhin wird in dieser Arbeit eine potentielle allosterische Bindungstasche der Fucosyltransferase 8 diskutiert. Theoretische Studien legen nahe, dass die Entwicklung von hochaffinen Bindern für diese



Bindungstasche vielversprechender ist als für die zuvor als Ziel erkorene Donorbindungstasche der Fucosyltransferase 8. Eine MD-Simulation zeigt, dass diese potentielle Bindungstasche von der Akzeptorbindungstasche der Fucosyltransferase 8 zugänglich ist. Liganden, die in diese potentielle Bindungstasche binden und sich bis in die Akzeptorbindungstasche erstrecken, könnten die Enzymaktivität der Fucosyltransferase 8 erfolgreich hemmen.

**2. „Ensemble Docking“ von Renin:** „Ensemble Docking“ ist eine einfache und dennoch vielversprechende Methode zur Berücksichtigung der Flexibilität eines Proteins bei einem virtuellen *Screening*. Hierfür werden die Liganden bei dem virtuellen *Screening* in unterschiedliche Strukturen des Proteins gedockt, um Effekte der Protein-Flexibilität zu berücksichtigen, und die Ergebnisse anschließend zusammengeführt. Zur Feststellung des Leistungsvermögens von „Ensemble Docking“ wurde in dieser Arbeit das Protein Renin als Testsystem ausgewählt und untersucht. Renin ist ein geeignetes Testsystem, da eine ausreichende Anzahl von Kristallstrukturen existiert und bekannt ist, dass die Bindungstasche von Renin eine signifikante Flexibilität aufweist. Sowohl Kristallstrukturen als auch Strukturen aus MD-Simulationen mit einer Gesamtlänge von 500 ns wurden untersucht. Das Verfahren zur Auswahl der unterschiedlichen Proteinstrukturen ist ein signifikanter Faktor für den Erfolg von „Ensemble Docking“, z.B. ist es offensichtlich nicht wünschenswert mehrere Strukturen auszuwählen, die sich sehr ähnlich sind. In dieser Arbeit wurden zwei verschiedene Cluster-Algorithmen zur Definition der Ensembles untersucht: *K*-means und hierarchisches Clustering (mit *average linkage*). Drei Leistungskategorien wurden untersucht: Die Vorhersage des Bindungsmodus, die Fähigkeit in einem virtuellen *Screening* in der Ergebnis-Rangliste Liganden vor Nichtliganden einzuordnen und die Vorhersage der Bindungsaffinität. Ensembles, die mittels *k*-means Clustering zusammengestellt wurden, schneiden im Vergleich erfolgreicher ab. Außerdem wurden alle möglichen Ensemble-Kombinationen, die aus maximal vier Kristallstrukturen von Renin möglich sind, spezifiziert und ihr Leistungsfähigkeit bewertet. Die resultierenden Daten verdeutlichen das enorme Potential von „Ensemble Docking“. Für Ensembles, die sich aus vier Kristallstrukturen zusammensetzen, konnte im Mittelwert die korrekte Vorhersage des Bindungsmodus um mehr als 70% verbessert werden (im Vergleich zum Abschneiden von individuellen Kristallstrukturen). Weiterhin verbesserte sich die Ranking-Leistungsfähigkeit in einem virtuellen Screening-Szenario um 11% (gemessen am *area under the curve of the receiver operating curve*-Kennwert) und die Vorhersage der Bindungsaffinität um 55% (gemessen an  $R^2$ ).

**3. Optimierung von Inhibitoren der Humanen Galactosyltransferase B:** Die humane Galactosyltransferase B (GTB) katalysiert die Galactosylierung des H-Antigens und ist somit für die Bildung des Blutgruppen-Phenotyps B verantwortlich. Bestimmte Untergruppen des ABO-Blutgruppensystem sind wiederholt mit einer erhöhten Wahrscheinlichkeit verschiedener Krankheiten in Verbindung gebracht worden. Zum Beispiel haben Individuen des Blutgruppen-Phenotyps B eine deutlich erhöhte Wahrscheinlichkeit an Pankreas-Krebs zu erkranken. 3-Phenyl-5-(piperazin-1-yl)-1,2,4-thiadiazol **35** ist in einem dieser Arbeit vorhergehenden Fragment-Screening als Inhibitor der GTB mit einem  $K_i$  von  $800\ \mu\text{M}$  identifiziert worden. In dieser Arbeit wurde ein *fragment growing*-Ansatz verfolgt, um die Bindungsaffinität des Fragments **35** zu optimieren. Zunächst wurde der Phenylrest von Fragment **35** zu einem Naphthylrest vergrößert. Für den resultierenden Liganden **37** konnte mittels kompetitiver STD NMR eine Verbesserung der Bindungsaffinität um den Faktor 3 ( $K_i = 271\ \mu\text{M}$ ) festgestellt werden. Weiterhin wurde eine Substitution des Phenylrings von Fragment **35** mit einem OBn-Substituenten in 2- und 3-Position untersucht. Modeling-Studien zeigten, dass der OBn-Substituent vorteilhafte  $\pi$ - $\pi$ -Interaktionen mit den Seitenketten von His233 und Trp300 eingehen kann. In der Praxis erwiesen sich die entsprechenden Liganden **38** und **39** jedoch als nicht vielversprechend. Ein weiterführender Ansatz zur Substitution des Piperazin-Restes von Ligand **37** resultierte in Ligand **62**, der sich allerdings ebenfalls als nicht vielversprechend erwies.

## 3 Introduction

### 3.1 Drug Discovery and Development

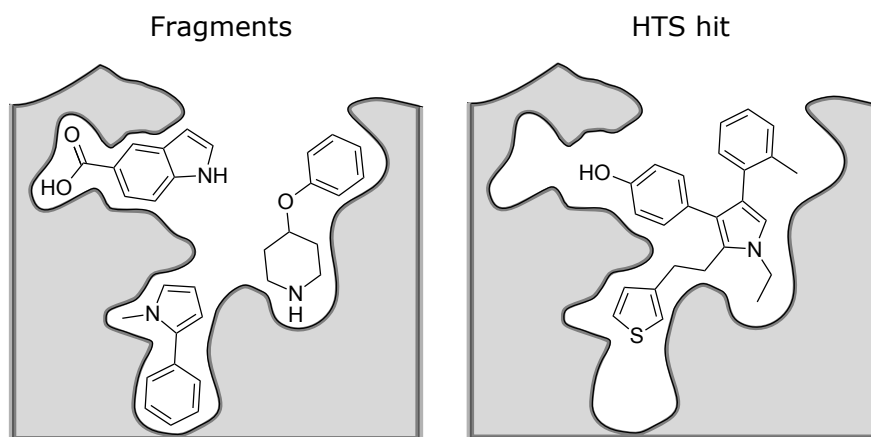
Life expectancy has increased drastically in the last century. This has been attributed to improved education, nutrition, life-style, and income rather than to improved medical care.<sup>[1]</sup> Nonetheless, a statistical estimate for the period 1982-2001 concludes that life expectancy has increased one week per year due to the launch of new drugs and that the associated price was 6750 US\$ per capita.<sup>[1]</sup> In a rapidly aging society, the significance of aging-associated diseases such as cancer and Alzheimer's disease is increasing. However, the pharmaceutical industry is struggling to find responses to these age-associated diseases. For example, for Alzheimer's disease many efforts towards a cure have been made (e.g. bapineuzumab, solanezumab, verubecestat, intepirdine, PBT2) yet they all failed in clinical trials so far. These failures are apparent symptoms of an industry that is struggling with a constant decline of research and development efficiency.<sup>[2]</sup> Since 1950, the number of new drugs approved per billion US\$ (in inflation adjusted terms) spent on research and development has halved roughly every nine years.<sup>[2]</sup> An estimate from the year 2010 concludes that launching a new drug to the market takes on average more than 10 years and costs more than 1.7 billion US\$ (including the cost of capital).<sup>[3]</sup> This makes pharmaceutical industry the most research intensive industry.<sup>[3]</sup> In parts, pharmaceutical industry has become unsustainable and therefore the call for „new paradigms“ is frequent.<sup>[4]</sup>

Drug discovery remained a serendipitous process for a long period of the 20th century. Advances in the field of genomics enabled the rapid identification of new targets and allowed for their recombinant expression.<sup>[5]</sup> The implementation of high-throughput screenings (HTS) allowed large substance libraries to be assayed at affordable pricing.<sup>[6]</sup> Combinatorial chemistry increased the number of new chemical entities that are accessible on a relevant time scale by a multitude.<sup>[2]</sup> Qualitative structure-activity relationship-models (QSAR) tried to span a theoretical framework for drug discovery, but their predictive power remained elusive.<sup>[7]</sup> Advances in X-ray crystallography allowed for structure-based drug design. Fragment-based drug discovery emerged as a new strategy to sample the vast chemical space more efficiently and was enabled by new biophysical screening methods like STD NMR and SPR.<sup>[8,9]</sup> Nowadays, computational methods are frequently employed in drug discovery. Docking enables *in silico* high-throughput screening, however there is still skepticism regarding its usefulness.<sup>[10]</sup> More rigorous methods, like free energy perturbation, allow for an accurate description of binding energies, but at

present their computational cost is too expensive.<sup>[11]</sup>

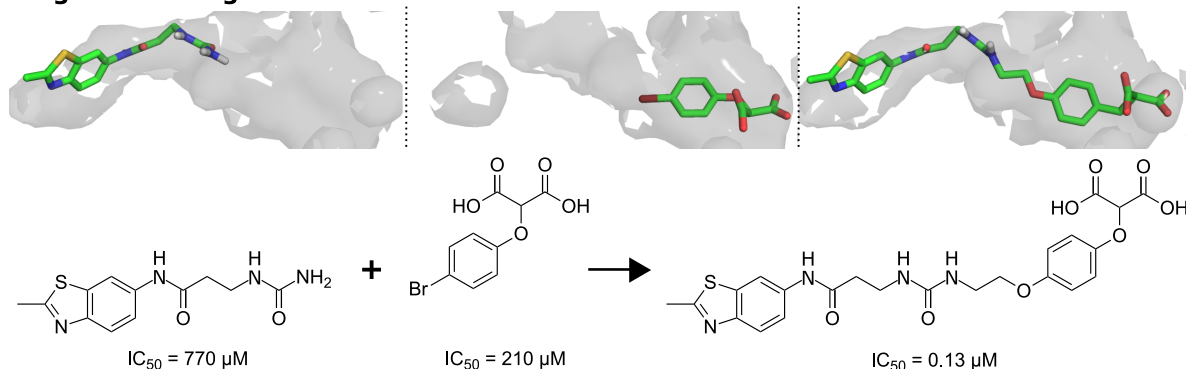
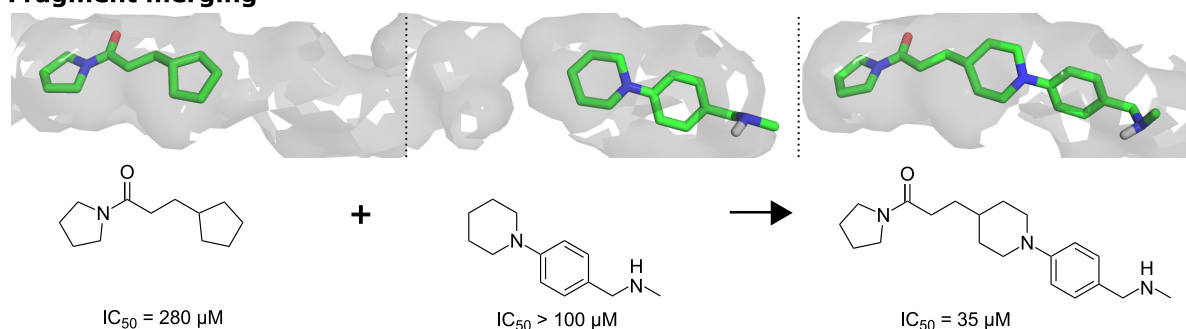
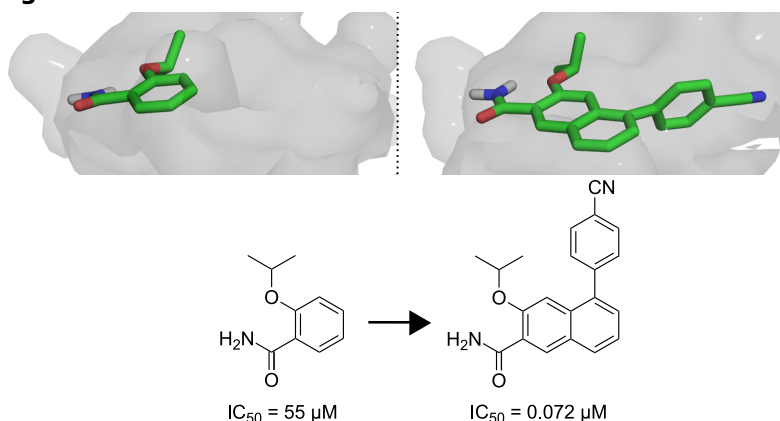
### 3.1.1 Fragment-Based Drug Discovery

The drug-like chemical space has been estimated to be about  $10^{60}$  molecules.<sup>[12]</sup> With the advent of HTS, quite large compound libraries have been assembled. Yet, these libraries cover only a fingertip of the drug-like chemical space and their maintenance is costly. For molecules of smaller molecular weight, the vastness of chemical space lessens significantly: Chemical space for molecules of less than 160 Da has been estimated to consist of only 14 million compounds.<sup>[13]</sup> This still represents a large number, however at this size, screening a relevant part of the chemical space becomes actually feasible. Additionally, it has been postulated that less complex molecules should exhibit higher hit rates.<sup>[14]</sup> These ideas provide the foundation for fragment-based drug discovery (FBDD). In analogy to Lipinski's rule of five, a rule of three has been proposed for fragments. The rule states that fragments should have a molecular weight of less than 300 Da, no more than three hydrogen bond donors or acceptors, and clogP values of less than three.<sup>[15]</sup> In comparison to HTS hits, fragments are typically low-affinity binders (range ~0.1-10 mM) and therefore require sensitive screening methods like STD NMR and SPR. They typically form only few but highly efficient interactions within the binding pocket (see **Fig. 1**).<sup>[9]</sup> This efficiency has been expressed in diverse „ligand efficiency“ metrics even though harsh criticism on these metrics has arisen.<sup>[16]</sup> On the contrary, HTS hits form many but suboptimal interactions within the binding pocket (see **Fig. 1**).<sup>[9]</sup> As a result, HTS hits frequently do not represent good starting points for further ligand optimization. Their affinity is prevalently entropy-driven and therefore more difficult to optimize than those of fragments that prevalently exhibit an enthalpy-driven binding affinity.<sup>[17,18]</sup>



**Fig. 1:** Left: Fragments constitute low-affinity binders and form only few yet highly efficient interactions. Right: HTS hits form many but suboptimal interactions.<sup>[9]</sup>

Fragment elaboration terms the transformation of a fragment into a drug candidate. This phase involves several cycles in which fragments are synthetically modified based on the analysis of their binding mode. The resulting change in binding affinity is evaluated and the cycle subsequently reiterated.<sup>[9]</sup> There are three strategies for fragment elaboration: Fragment linking, merging, and growing (see **Fig. 2**). For fragment linking, two fragments that bind to non-overlapping binding sites are joined into a single molecule. Fragment linking is an appealing strategy: The binding energy of the ligand obtained by ideal linking of two fragments can be higher than the sum of the individual binding energies of the the two fragments. This phenomenon has been coined „superadditivity“<sup>[19]</sup> and is a result of the loss of rigid body translational and rotational entropy upon ligand binding which is largely independent of the ligands molecular weight.<sup>[9,20]</sup> However, superadditivity is seldomly achieved because perfect linking of fragments is notoriously difficult to realize.<sup>[21]</sup> The fragment linking strategy also lacks general applicability because within the binding site fragments frequently rather bind to „hot spots“ than to distinct subpockets. For fragment merging, structural proportions of fragments that occupy overlapping binding sites are incorporated into a single ligand. For fragment growing, a single fragment serves as a starting point and new chemical functionalities are added subsequently.

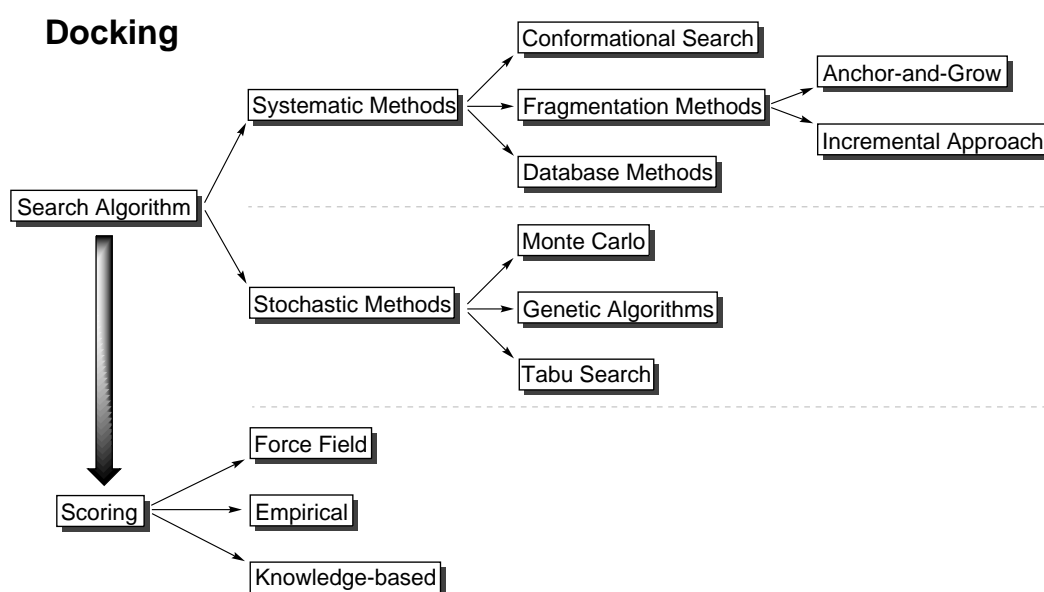
**Fragment linking****Fragment merging****Fragment growing**

**Fig. 2:** Strategies for fragment elaboration. Upper panel: Two fragments that bind to non-overlapping binding sites of lactate dehydrogenase A are joined into a single ligand.<sup>[22]</sup> Middle panel: Structural proportions of two fragments that bind to overlapping binding sites of *Mycobacterium tuberculosis* EthR are merged into a single ligand.<sup>[23]</sup> Lower panel: A fragment that binds to interleukin-1 receptor is grown by the addition of new structural proportions.<sup>[24]</sup> All binding sites are displayed in an inverted surface mode.

**3.1.2 Computer-Aided Drug Design**

X-ray crystallography has largely fulfilled the increased need of three-dimensional structural information for structure-based drug design. On this basis, computational methods raise the prospect of predicting the binding affinity of a ligand even before it has ever been synthesized. This capability would streamline the drug discovery process in an unprecedented manner.

Docking is a computational method that predicts the orientation and conformation of a ligand within a binding site (pose prediction) and the associated binding affinity (scoring).<sup>[25]</sup> A search algorithm is employed for the pose prediction. There are two main categories of search algorithms in docking: Systematic and stochastic algorithms.<sup>[26,27]</sup> Systematic algorithms can be subdivided into conformational search, fragmentation, and database methods.<sup>[27]</sup> Conformational search methods sample possible alterations of the ligand by a fixed increment. This represents an exhaustive approach with limited applicability because of its computational expenses.<sup>[27]</sup> Examples of fragmentation methods are anchor-and-grow and incremental construction algorithms.<sup>[26]</sup> Anchor-and-grow algorithms portion the ligand into fragments, then dock each of these fragments, and finally relink the fragments to recreate the original ligand.<sup>[27]</sup> Incremental construction algorithms identify a rigid fragment of the ligand, dock this fragment, and subsequently add the remaining flexible proportions.<sup>[27]</sup> Database methods employ libraries of pregenerated ligand conformations.<sup>[27]</sup> Stochastic methods apply random changes to the ligand which are then either accepted or rejected on the basis of a predefined criterion.<sup>[27]</sup> Examples of stochastic algorithms are Monte Carlo methods, genetic algorithms, and tabu search methods.<sup>[27]</sup>



**Fig. 3:** Search algorithms are employed to predict the pose of a ligand within a binding site. The associated binding affinity is predicted by a scoring function. The schematized tree diagram categorizes different approaches for search algorithms and scoring functions (cf. text).

Scoring functions predict the binding affinity of a ligand binding to a protein. However, they are oversimplifying their task as there is an inherent inverse relationship between the accuracy of a scoring function and its computational speed.<sup>[28]</sup> For example, scoring functions assume that the binding energy can be approximated by a linear combination of pairwise terms, however

many forces involved are non-additive.<sup>[26]</sup> There are three main categories of scoring functions: Force field-based, knowledge-based, and empirical scoring functions.<sup>[29]</sup> For force field-based scoring functions, the individual sum terms are calculated on the basis of interatomic potentials described in a force field.<sup>[29]</sup> Empirical scoring functions decompose the binding energy into individual energy terms as well, for example they typically include terms that reward hydrogen bonds or penalize frozen rotatable bonds.<sup>[28]</sup> However, the coefficients of these terms are derived by regression performed on binding affinity data contained in a training set of protein-ligand complexes with known three-dimensional structures<sup>[29]</sup> Empirical scoring functions are prone to failure if they are challenged to evaluate binding interactions that are significantly different to those displayed in the training set.<sup>[26]</sup> For knowledge-based scoring functions, the distance-dependent potential of atom pairs present in experimentally determined protein-ligand complexes is derived by inverse Boltzmann analysis.<sup>[28,29]</sup> The binding energy is estimated as the sum of these potentials.<sup>[28]</sup> Therefore, in contrast to force-field based and empirical scoring function, knowledge-based scoring functions are constructed without consideration of experimentally binding affinity data and solely rely on the statistical analysis of structural information.<sup>[28]</sup>

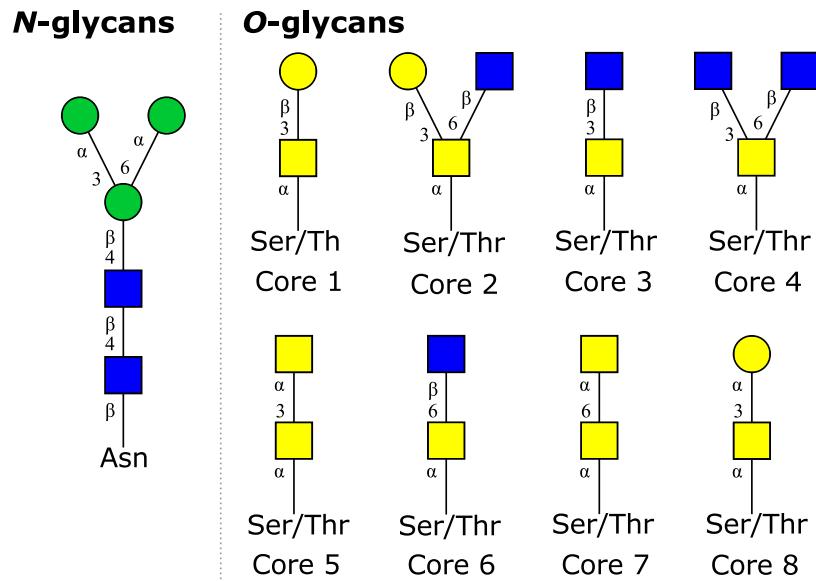
### 3.2 Glycosylation

Glycosylation is a frequent post-translational modification of proteins.<sup>[30]</sup> Its biological function can be manifold: Glycans have been shown to increase the solubility of proteins, to assist their folding, and to protect them against proteolytic cleavage.<sup>[31]</sup> Furthermore, glycans and their glycan-specific receptors play a crucial role in many cell-cell recognition and signaling events.<sup>[31]</sup> Therefore unsurprisingly, aberrant glycosylation patterns have been linked to a manifold of diseases including cancer.<sup>[31,32]</sup>

The majority of glycans is bound to proteins by either *N*- or *O*-glycosidic linkage. *N*-glycans are linked to asparagines located in the consensus sequence asparagine-X-serine/threonine ( $X \neq \text{Prolin}$ ).<sup>[33]</sup> All *N*-glycans feature the same pentasaccharide core shown in **Fig. 4** ( $\text{Man}\alpha 1-6(\text{Man}\alpha 1-3)\text{Man}\beta 1-4\text{GlcNAc}\beta 1-4\text{GlcNAc}\beta 1$ ).<sup>[33]</sup> In contrast, for *O*-glycans neither a specific consensus sequence nor a uniform core structure is known.<sup>[33]</sup> Mucine-type *O*-glycosylation that features a  $\alpha$ -linked *N*-acetylgalactosamine bound to a serine or threonine can be found most frequently. Eight different core structures are known for mucine-type *O*-glycosylation (see **Fig. 4**).<sup>[33]</sup> Even though no consensus sequence is known for the mucine-type *O*-glycosylation,



the substrate specificity of several polypeptide *N*-acetylgalactosamintransferases has been explored.<sup>[34]</sup>



**Fig. 4:** Left: All *N*-glycans share the same pentasaccharide core structure. Right: Mucine-type *O*-glycosylation is significantly more diverse: Eight different core structures are known.

## **4 General Objective**

Inhibitors of glycosyltransferases are efficient tools to study the physiological role of glycosylation. Given the fact that aberrant glycosylation patterns have been linked to a manifold of diseases, these inhibitors might demonstrate therapeutic potential as well. Unfortunately, most previous studies on glycosyltransferase inhibitors have focused on analogues of either donor or acceptor substrate. These substrate analogues have limited potential because they typically feature non drug-like properties, e.g. a high polarity, that diminish their cell viability.

The aim of this project was to find new strategies for the development of glycosyltransferase inhibitors with drug-like properties. In this context, two drug discovery strategies, namely fragment-based and computer-aided drug discovery, were identified as beneficial and were scheduled to find application in this thesis. Fragment-based strategies allow for a tight control over molecular properties and therefore allow to overcome the problems that are associated with glycosyltransferase inhibitors that are close substrate analogues (e.g. a high polarity). Computer-aided strategies allow to sample the vastness of chemical space efficiently and raise the prospect of streamlining the drug discovery process in a unprecedented manner. To demonstrate the effectiveness of these strategies for the design of glycosyltransferase inhibitors, the goal was to synthesize the designed inhibitors and assess their inhibitory potential by biophysical methods such as STD NMR.

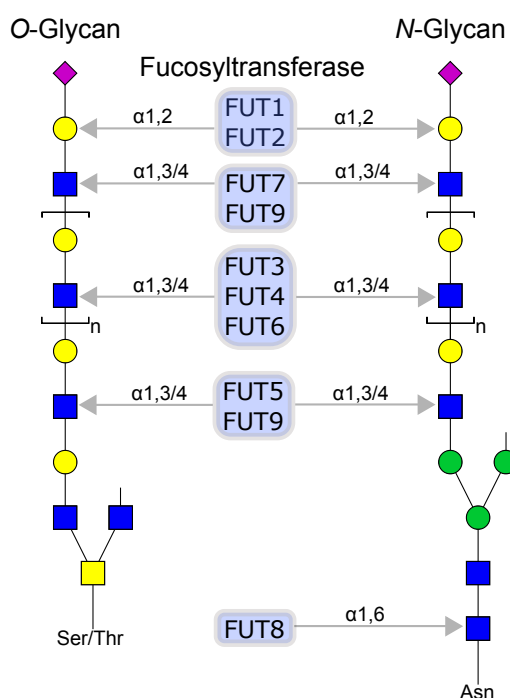
## 5 Results and Discussion

### 5.1 Development of New Inhibitors of Fucosyltransferase 8

Parts of this chapter have been submitted for publication: C. Strecker, M. Baerenfaenger, M. Mieke, E. Spillner, B. Meyer, *In silico* evaluation of the binding site of fucosyltransferase 8 and first attempts of synthesizing an inhibitor with drug-like properties, *ChemBioChem* **2019**, submitted.

#### 5.1.1 Introduction

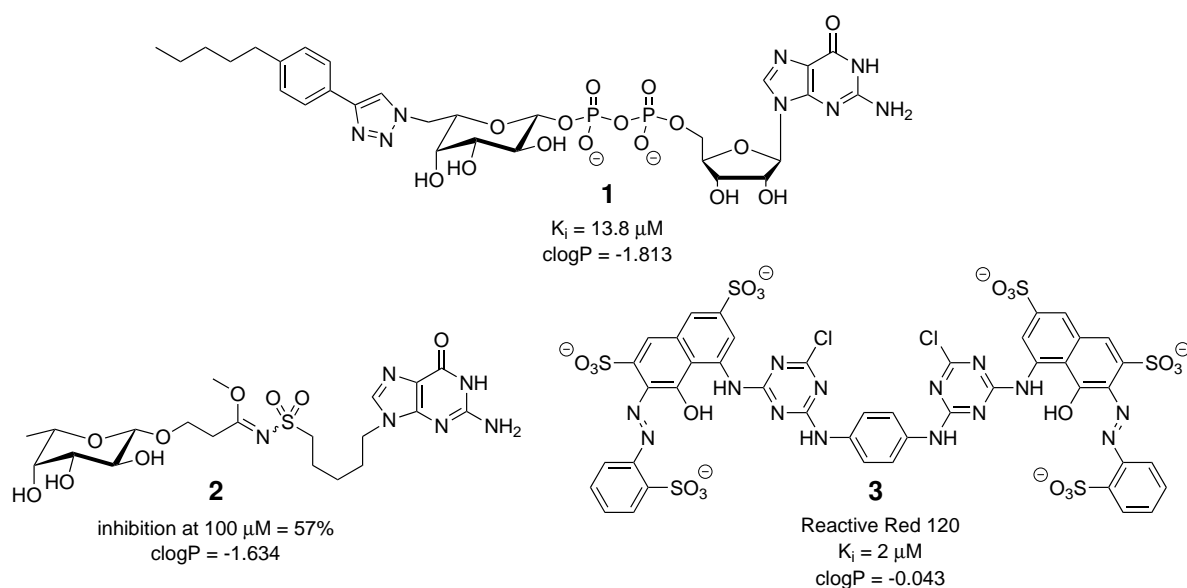
Fucosyltransferases catalyze the transfer of L-Fucose from the donor substrate GDP-Fucose to an acceptor substrate.<sup>[35]</sup> So far, 13 human fucosyltransferases are known.<sup>[36]</sup> POFUT1 and POFUT2 catalyze the transfer of L-Fucose to serine or threonine residues on proteins.<sup>[35]</sup> The other fucosyltransferases transfer L-Fucose to an acceptor glycan. Even though they show distinct substrate preferences as schematized in **Fig. 5**, a certain degree of redundancy exists. For example, FUT1 and FUT2 both yield  $\alpha$ 1,2-linked fucose and both are able to act on O- as well as on N-glycans.<sup>[35]</sup> In contrast, fucosyltransferase 8 (FUT8) is the only fucosyltransferase that aids the construction of  $\alpha$ 1,6-linked fucose and that acts exclusively on N-glycans.<sup>[35]</sup> FUT8 transfers fucose to the innermost N-acetylglucosamine of N-glycans, resulting in core fucosylation.



**Fig. 5:** Preferential fucosylation sites of FUT1-9.<sup>[35]</sup>

Core fucosylation has been implicated to play a role in various types of cancer. For example upregulation of FUT8 is a driver of melanoma metastasis,<sup>[37]</sup> associated with an unfavourable clinical outcome in patients with non-small cell lung cancer,<sup>[38]</sup> and with aggressive prostate cancer.<sup>[39]</sup> Furthermore, for therapeutic antibodies it has been shown that depletion of core fucosylation results in a 50-100 fold increase of Fc $\gamma$ -mediated cytotoxicity.<sup>[40]</sup> The important physiological role of FUT8 is underlined by the fact that 70% of FUT8-deficient knock-out mice die within three days after birth.<sup>[41]</sup> Inhibitors of FUT8 would therefore constitute an efficient tool for more detailed studies on the physiological role of FUT8 and might demonstrate therapeutic potential as well.

Therefore, FUT8 constitutes an attractive target for the development of novel therapeutic drugs. However, glycosyltransferases have proven to be difficult targets. Because the phosphate group(s) of the donor substrate account for a major portion of the binding affinity many studies have focused on modified pyrophosphate derivatives.<sup>[42]</sup> Similarly, acceptor-substrate analogues have attracted significant attention because they offer specificity for glycosyltransferases that resort on identical donor substrates.<sup>[42]</sup> However, close substrate analogues have limited potential due to their typically non drug-like properties such as their high polarity and limited chemical stability.<sup>[42]</sup> In special, this focus on substrate analogues has held back the development of fucosyltransferase inhibitors suitable for *in vivo* studies as reviewed by Tu *et al.*<sup>[43]</sup> For FUT8, so far three publications have reported inhibitors.<sup>[44–46]</sup> Manabe *et al.*<sup>[44]</sup> and Hosoguchi *et al.*<sup>[45]</sup> reported analogues of GDP-Fucose that inhibit FUT8. Their most potent inhibitors, **1** and **2** (as shown in **Fig. 6**), exhibit affinities in the lower  $\mu$ M range. However, the corresponding clogP values of less than -1.5 indicate that bioavailability is not be expected. Kamińska *et al.* reported a set of triazine dyes that inhibit FUT8.<sup>[46]</sup> Their most potent inhibitor reactive red 120 **3** carries six ionic sulfonates that will prevent membrane permeability. Also, the molecular mass of more than 1.3 kDa grossly violates Lipinski's rule. Furthermore, even though the reported  $K_i$  of 2  $\mu$ M for ligand **3** sounds promising on first sight, the concept of ligand efficiency (LE =  $\Delta G$ /no. of heavy atoms)<sup>[47]</sup> demasks ligand **3** to be a very inefficient binder with a ligand efficiency of only 0.09.



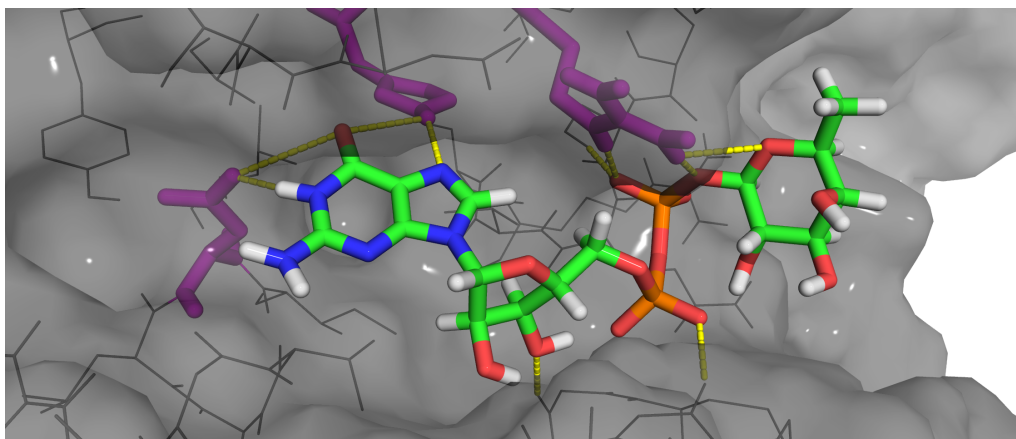
**Fig. 6:** Structures of the most potent FUT8 inhibitors reported so far.<sup>[44–46]</sup> All three inhibitors show traits that significantly impair their bioavailability (e.g.  $\text{clogP}$  values of less than  $-1.5$  for **1** and **2**).

### 5.1.2 Objective

The aim of this project was to generate new inhibitors of FUT8. As pointed out in the previous section, existing inhibitors of FUT8 suffer from non drug-like properties such as their high polarity. Therefore, a fragment-based strategy was embraced that allows for more control over molecular properties so that these non drug-like properties can be avoided. In the absence of a suited fragment library, it was resorted to computer-aided tools to create ligands by *de novo* design.

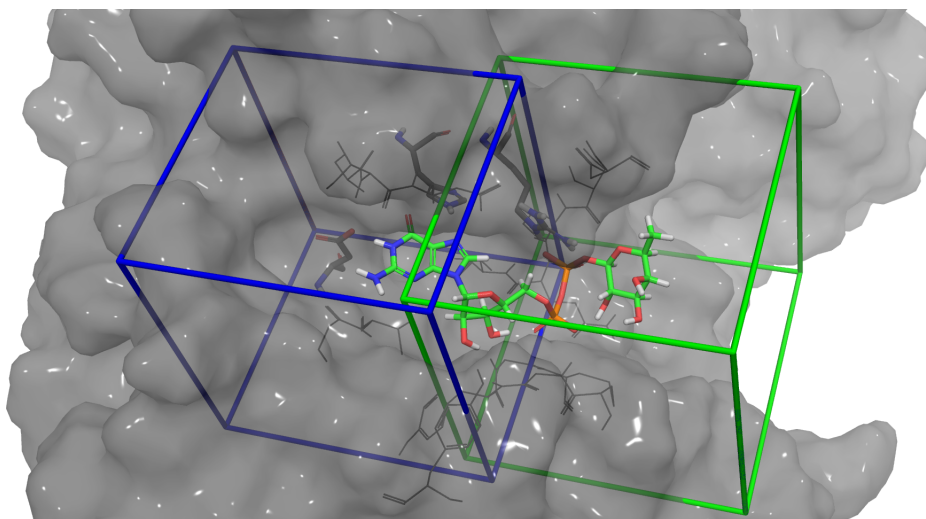
### 5.1.3 Fragment-Based *De Novo* Design of Inhibitors of Fucosyltransferase 8

A high-resolution crystal structure of human FUT8 exists (PDB: 2DE0).<sup>[48]</sup> Unfortunately, this is an *apo* crystal structure that is not suitable to start the design process because a key catalytic residue (Arg365) is collapsed into the donor binding site in this structure. However, a model of the FUT8-GDP-Fucose complex has been developed previously in our group on the basis of this crystal structure.<sup>[49]</sup> This model was recreated and subsequently refined by a MD simulation with a length of 1.5 ns. The final frame (as shown in **Fig. 7**) of this MD served as starting point for the subsequent docking studies. The efforts presented here concentrate on the donor binding site because the acceptor site of FUT8 is very shallow and solvent exposed and thus can be considered undruggable.



**Fig. 7:** Binding mode of GDP-Fucose. Hydrogen bonds are shown in yellow. Three amino acid residues are hypothesized to be of central significance for donor substrate binding: Asp453, His363, and Arg365 (shown as magenta sticks).

It was hypothesized that a potential ligand of the donor binding site should exhibit hydrogen bonding to three key amino acid residues: Asp453, His363, and Arg365 (displayed as magenta sticks in **Fig. 7**). Two separate docking campaigns were performed against the donor binding site of FUT8 employing a library of 700,000 fragments obtained from the ZINC12 database.<sup>[50]</sup> For this, two rather small grid boxes were used to ensure hits that cover the entire binding site. The grid box used in the first docking campaign was centered around the midpoint of the guanine moiety, the second centered on the  $\beta$ -phosphate (see **Fig. 8**).



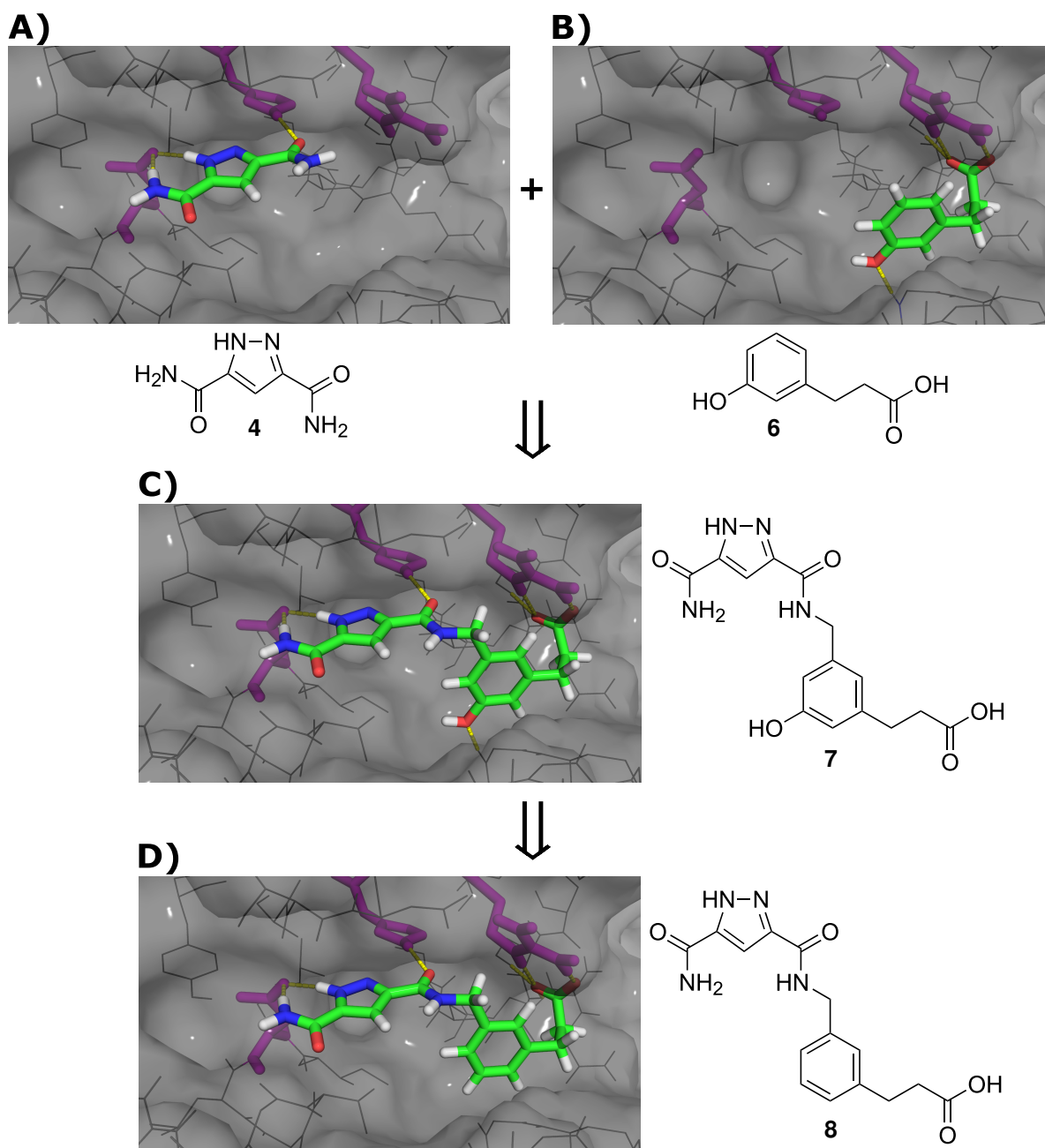
**Fig. 8:** Grid boxes used for the two described docking campaigns. The first grid box (in blue) is centered on the midpoint of the guanine moiety, the second (in green) centered on the  $\beta$ -phosphate.

In the first docking campaign, 1*H*-pyrazole-3,5-dicarboxamide **4** (ZINC95830521) was identified as an interesting hit ranked second place in the docking hit list (see panel A in **Fig. 9** for a binding pose, **Tab. 1** for a short summary of docking results). The NH of the pyrazole moiety and a NH of the primary carboxamide function form pincer-like hydrogen bonding to Asp453.

The carbonyl oxygen of the second carboxamide function is able to form a hydrogen bond to His363. In the second docking campaign, a large number of carboxylic acids hydrogen bonding to Arg365 was identified with high docking scores. In particular, 3,5-dihydroxyhydrocinnamic acid **5** (ZINC6091356) that was ranked 42nd place in the docking hit list stood out due to its structural simplicity. One of the phenolic OH functions of this fragment was removed to yield 3-(3-hydroxyphenyl)propionic acid **6** (ZINC156346, see panel B in **Fig. 9** for a binding pose). Both fragments, **4** and **6**, were linked *in silico* to yield ligand **7** (see panel C in **Fig. 9** for a binding pose). This ligand showed an improved calculated binding energy compared GDP-Fucose in MM-GBSA calculations (see **Tab. 1** for calculated binding energies). It was decided to remove the phenolic OH function of ligand **7** to reduce the synthetic effort. The resulting ligand **8** (see panel D in **Fig. 9** for a binding pose) still showed an improved calculated binding energy compared to GDP-Fucose (see **Tab. 1**). Importantly, ligand **8** exhibited stable hydrogen bonding to the three previously defined key amino acid residues (Asp453, His363, and Arg365) during a MD simulation of a length of 1.5 ns.

**Tab. 1:** Docking scores (from GlideScore SP5.0) and calculated binding energies (from Prime MM-GBSA) for the donor substrate GDP-Fucose, the docked fragments, and ligands after *in silico* linking. For structures, see **Fig. 9** on the next page.

| Ligand   | Docking Score | calc. $\Delta G$ [kcal·mol <sup>-1</sup> ] |
|----------|---------------|--|
| GDP-Fuc  | n.d.          | -48.5                                      |
| <b>4</b> | -6.116        | -27.7                                      |
| <b>6</b> | -6.413        | -33.4                                      |
| <b>7</b> | -9.182        | -60.3                                      |
| <b>8</b> | -8.363        | -53.8                                      |

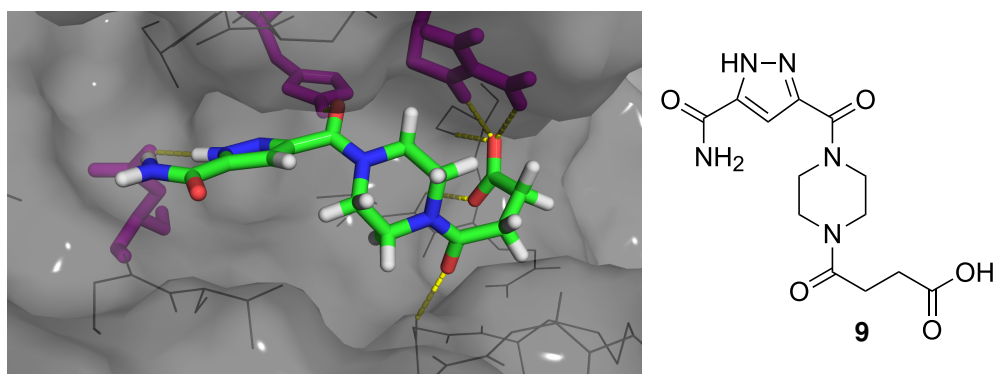


**Fig. 9:** Docking a fragment library into the donor site of FUT8 yielded two fragments, **4** and **6** (panel A and B), that were subsequently linked *in silico* to yield ligand **7** (panel C). After exemption of the phenolic OH function, ligand **8** (panel D) was obtained.

After the design of ligand **8**, it was realized that the ligand design process is significantly dependent on the input protein structure of FUT8. For the design of ligand **8**, the final frame of a MD simulation was used. This can be categorized as a rather arbitrary decision (and the discontent over this resulted in the exploration of ensemble docking as described in **Chapter 5.2**). By that time, it was observed that fragment 1*H*-pyrazole-3,5-dicarboxamide **4** can be successfully docked into many of the MD-derived structures of FUT8. In contrast, the region within the donor binding site that is responsible for binding of the ribose diphosphosphate moiety of GDP-Fucose seems to exhibit significantly more plasticity. As a result, fragment 3,5-



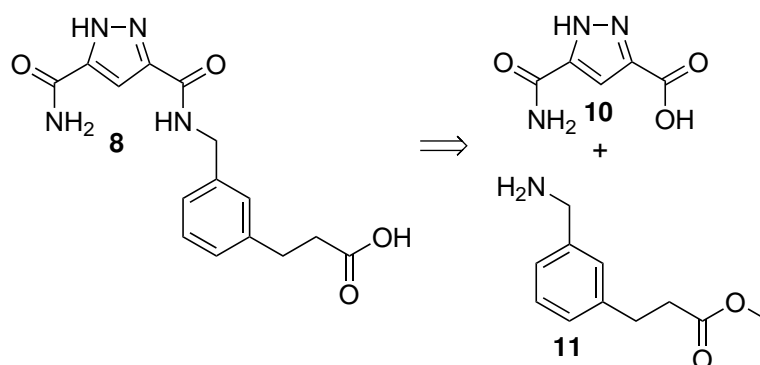
dihydroxyhydrocinnamic acid **5** can only occasionally be docked into the MD-derived structures of FUT8. To increase the chances of success, it was therefore decided to take a second approach. For this, fragment **4** served as a core structure as it was found to be a reasonable choice (see above). Furthermore, this choice offers synergism because the corresponding building block had to be synthesized for ligand **8** as well. However, the newly designed ligands should feature distinct chemotypes at the site responsible for binding of the ribose diphosphate moiety. For this, all amines from the previously used „FragmentsNow“ library were extracted and attached to the fragment **4** by employing CombiGlide, resulting in a library of approximately 50,000 ligands. These ligands were then docked into a MD-derived structure of FUT8 that is distinct from the structure that was used for design of ligand **8** (structure was chosen randomly). Ligand **9** was identified as an interesting hit (see **Fig. 10**) that showed a favorable calculated binding energy of  $-49.4 \text{ kcal}\cdot\text{mol}^{-1}$  in MM-GBSA calculations (see **Tab. 1** for a comparison).



**Fig. 10:** Docked binding pose of ligand **9**.

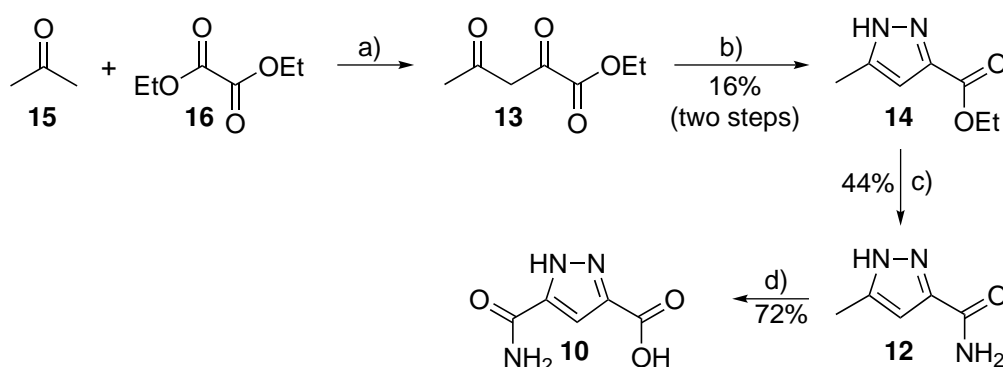
### 5.1.4 Synthesis

For the synthesis of ligand **8**, a convergent route in that building blocks **10** and **11** are coupled by amide linkage was envisioned (see **Fig. 11**). The ester function is subsequently deprotected.



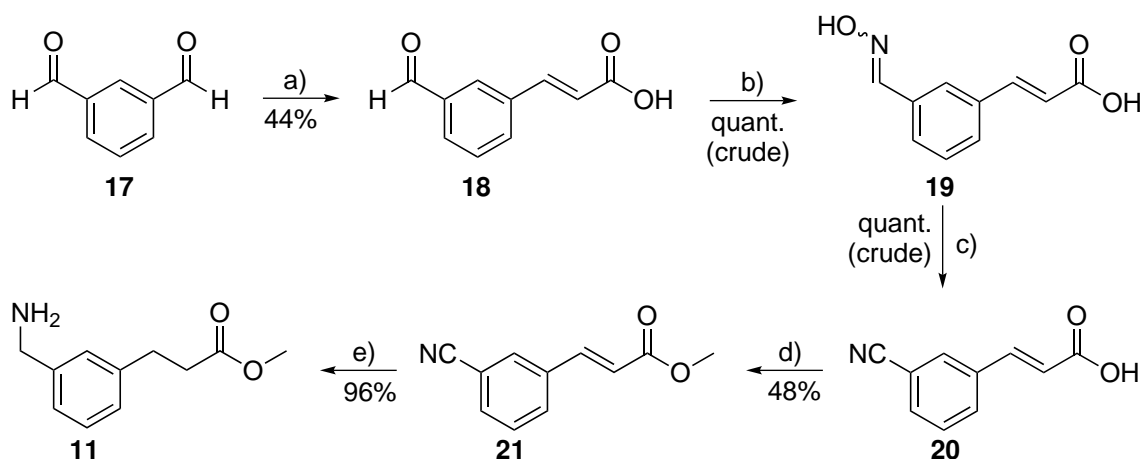
**Fig. 11:** Retrosynthesis of ligand **8**.

For the synthesis of building block **10**, protocols similar to those described by Skinner *et al.*<sup>[51]</sup> were used to access the intermediate product **12**. Acetone and diethyl oxalate were reacted in a mixed Claisen condensation to yield the 1,3-diketone **13** which was then converted to the pyrazole **14** without prior purification by the action of hydrazine. The ester function of **14** was ammonolysed to the primary amide **12** but in contrast to Skinner *et al.*<sup>[51]</sup>, who employed dry ammonia for this, a protocol by Jagdmann *et al.*<sup>[52]</sup> was employed in that ammonia is released by refluxing formamide in the presence of sodium methanoate. Finally, the aromatic methyl group of **12** was oxidized to give the carboxylic acid **10** by the action of potassium permanganate.



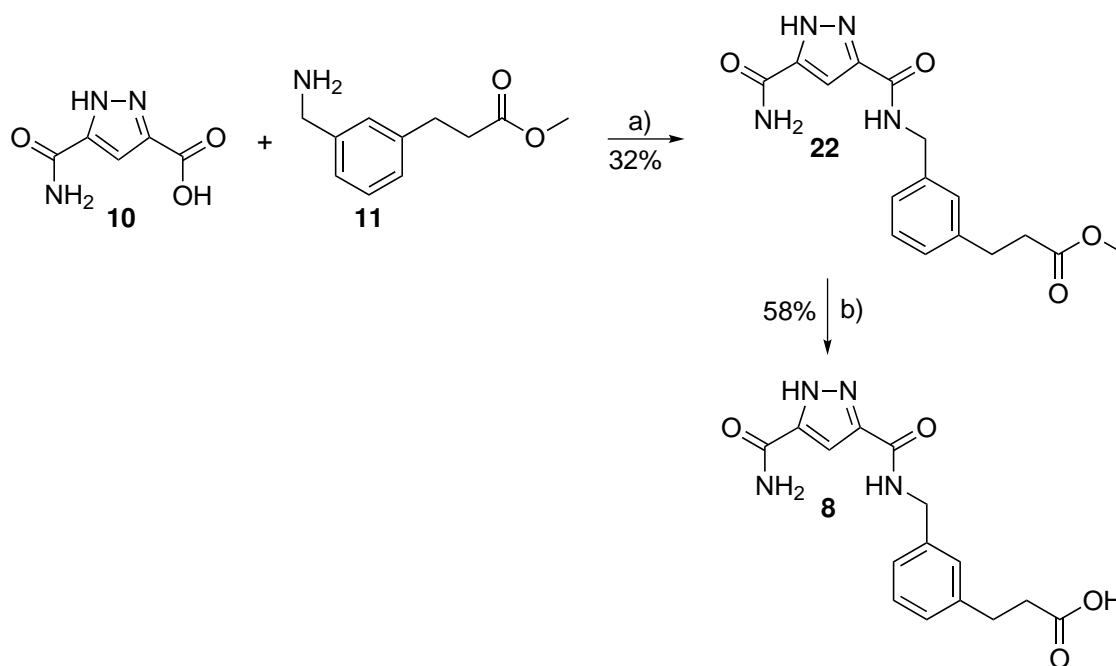
**Fig. 12:** Synthesis of building block **10**. Reaction conditions: a) 1.2 eq. NaOEt, EtOH, 15 min, 20 °C; b) 1.5 eq. N<sub>2</sub>H<sub>4</sub>·HCl, EtOH, 4 h, 78 °C; c) 4.0 eq. NaOMe, 6.0 eq. formamide, tetrahydrofuran, 5 h, 65 °C; d) 3.0 eq. KMnO<sub>4</sub>, H<sub>2</sub>O, 3 h, 95 °C.

Building block **11** became commercially available during the period of this PhD thesis. For this work, **11** was synthesized starting from isophthalaldehyde **17**. First, **17** was desymmetrized by a Knoevenagel condensation to yield **18** (as previously described by Hansen *et al.*)<sup>[53]</sup> The remaining aldehyde function of **18** was converted to oxime **19** by the action of hydroxylamine. The oxime **19** was subsequently dehydrated to nitrile **20** with cyanuric chloride in accordance to a protocol of De Luca *et al.*<sup>[54]</sup> Both **19** and **20** were converted without thorough purification because they showed similar elution behavior as **18**. After conversion of the carboxy function into the methyl ester **21** with methyl iodide, a significant change in retention factor occurred and the ester could be purified by column chromatography. Finally, both the nitrile and alkene function were reduced by palladium-catalyzed hydrogenation to yield the amine building block **11**. Hydrogenation was performed in slightly acidic medium to prevent secondary/tertiary amine formation.<sup>[55]</sup> In aqueous media reductive hydrolysis of the intermediate imine is possible.<sup>[55]</sup>



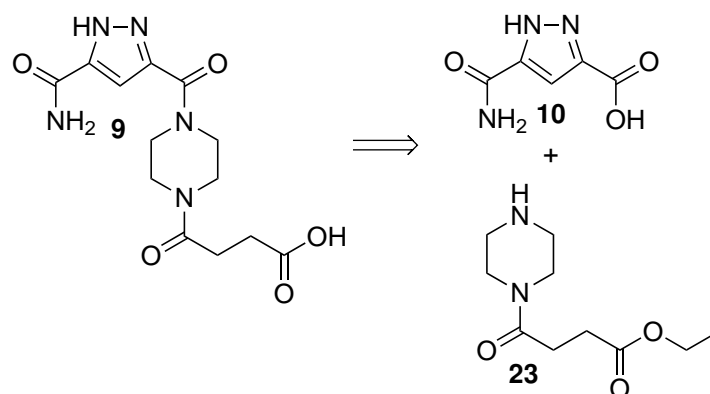
**Fig. 13:** Synthesis of building block **11**. Reaction conditions: a) 1.5 eq. malonic acid, 0.1 eq. piperidine, pyridine, 3 h, 100 °C; b) 3.0 eq.  $\text{NH}_2\text{OH}\cdot\text{HCl}$ , 3.0 eq.  $\text{NEt}_3$ , MeCN, 2 h, 20 °C; c) 2.0 eq. cyanuric chloride, DMF, 5 h, 20 °C; d) 1.2 eq. MeI, 1.2 eq.  $\text{CsCO}_3$ , DMF, 18 h, 20 °C; e)  $\text{H}_2$  (30 bar), Pd/C, dil. HCl in MeOH, 3 h, 20 h.

Both building blocks, **10** and **11**, were subsequently linked via amide coupling using propylphosphonic anhydride as coupling reagent yielding **22**. Finally, the ester function of **22** was hydrolyzed by the action of potassium hydroxide to yield ligand **8**.



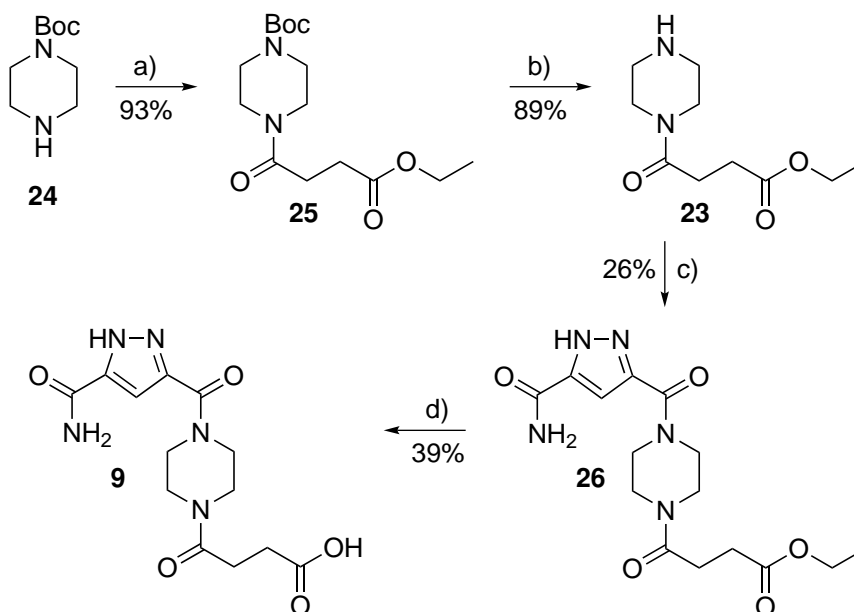
**Fig. 14:** Synthesis of ligand **8**. Reaction conditions: a) 2.0 eq. propylphosphonic anhydride, 2.0 eq. DIPEA, DMF, 24 h, 20 °C b) 3.0 eq. KOH,  $\text{H}_2\text{O}/\text{MeOH}$  1:4, 24 h, 20 °C.

For the synthesis of ligand **9**, a convergent route in that building blocks **10** and **23** are coupled by amide linkage was envisioned (see **Fig. 15**). Synthesis of building block **10** has already been described in the previous paragraphs.



**Fig. 15:** Retrosynthesis of ligand **9**.

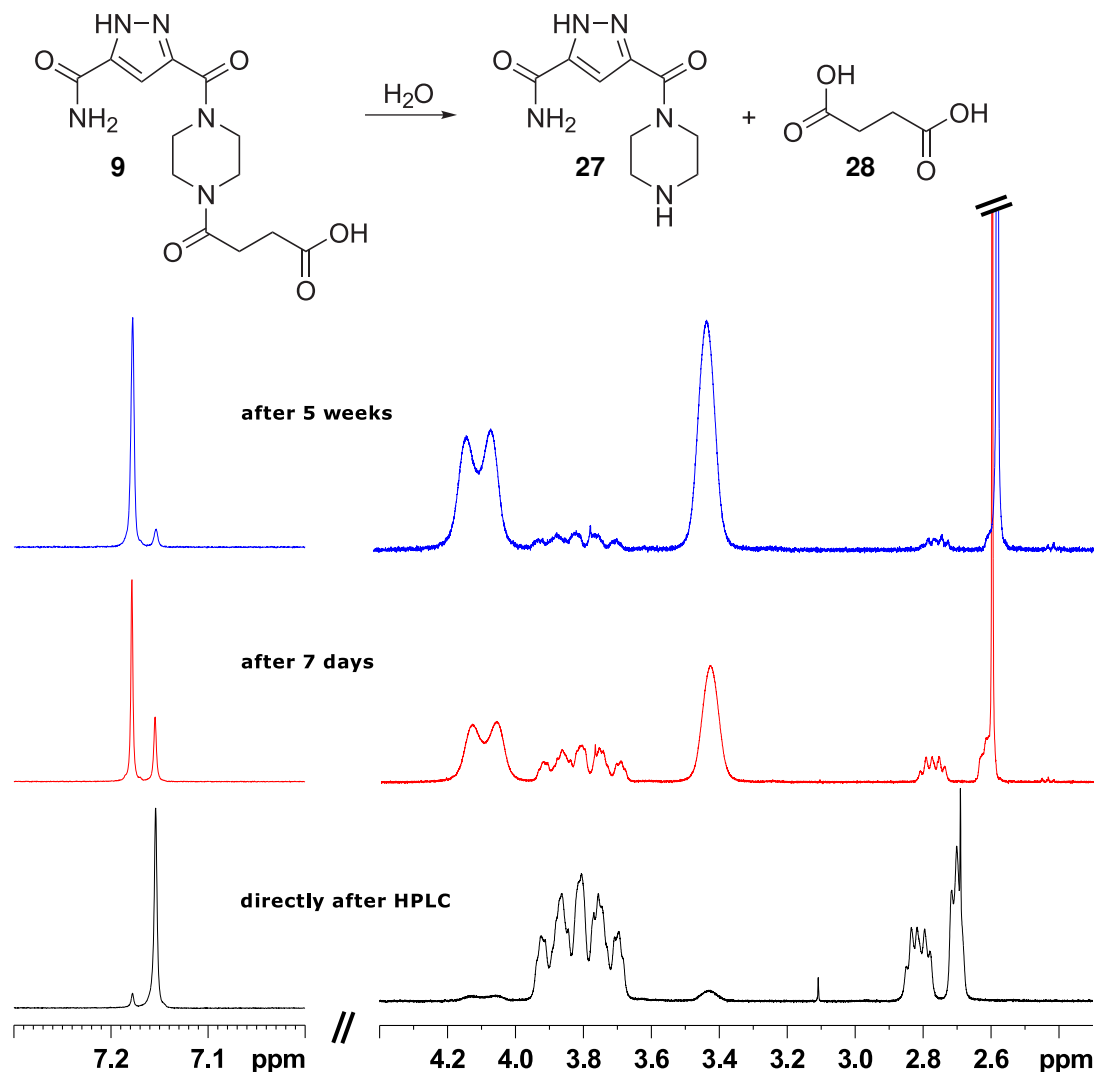
Synthesis of **9** was started from 1-boc-piperazine **24**. Reaction of 1-boc-piperazine **24** with ethyl succinyl chloride resulted in compound **25**. Subsequently, the Boc-group of compound **25** was cleaved off to yield the free amine **23**. Both building blocks, **10** and **23**, were subsequently linked via amide coupling using propylphosphonic anhydride as coupling reagent yielding **26**. Finally, the ester function of **26** was hydrolyzed by the action of potassium hydroxide to yield ligand **9**.



**Fig. 16:** Synthesis of ligand **9**. Reaction conditions: a) 1.15 eq. ethyl succinyl chloride, 1.2 eq. DIPEA, MeCN, 2 h, 0 to 20 °C; b) HCl, EtOH, 2 h, 20 °C; c) 0.66 eq. 3-carbamoyl-1H-pyrazole-5-carboxylic acid **10**, 2.0 eq. DIPEA, 2.0 eq. propylphosphonic anhydride, DMF, 1 h, 20 °C; d) 4 eq. KOH, MeOH/H<sub>2</sub>O 1:1, 24 h, 20 °C.

Despite repeated attempts to purify ligand **9** by RP-HPLC, NMR spectroscopy still revealed impurities. After several attempts, it was realized that even though the impurity can be separated by RP-HPLC it is regenerated by the decomposition of ligand **9** into compound **27** and succinic acid **28** (see **Fig. 17**). Decomposition of ligand **9** was monitored by <sup>1</sup>H-NMR spectroscopy (see

**Fig. 17).** In particular, the formation of a singlet at  $\sim 2.58$  ppm indicates the formation of succinic acid. After one week, more than 70% of ligand **9** had decomposed. Because of its instability, research on ligand **9** was abandoned.

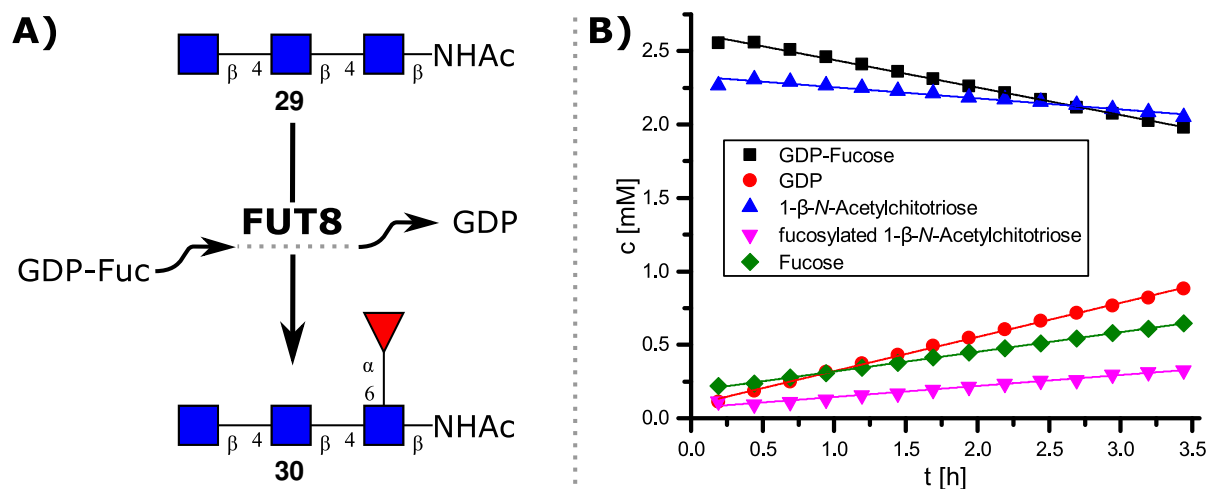


**Fig. 17:** Ligand **9** is hydrolyzed by water into compound **27** and succinic acid **28**. The decomposition of ligand **9** was monitored via  $^1\text{H-NMR}$  in unbuffered  $\text{D}_2\text{O}$  over a period of 5 weeks. Directly after HPLC purification of ligand **9** already  $\sim 5\%$  of decomposition products can be observed. After one week  $\sim 70\%$  of ligand **9** is decomposed and after 5 weeks decomposition is nearly quantitative.

### 5.1.5 Activity Assay

In order to assess the activity of recombinantly expressed FUT8, the FUT8-catalyzed fucosylation of 1- $\beta$ -*N*-acetylchitotriose **29** (see panel A of **Fig. 18**) was monitored via  $^1\text{H-NMR}$  spectroscopy over a course of 3.5 h. Noticeable, the fucosylation of 1- $\beta$ -*N*-acetylchitotriose **29** resulting in fucosylated 1- $\beta$ -*N*-acetylchitotriose **30** and the enzymatic cleavage of GDP-Fucose resulting in the release of fucose can be monitored simultaneously. The progress of the reaction is visualized in panel B of **Fig. 18**. From this period, initial rate constants were derived and are

summarized in **Tab. 2**. Fucosylated 1- $\beta$ -*N*-acetylchitotriose **30** was produced with an apparent initial rate constant of  $0.010\text{ s}^{-1}$ . This is in agreement with results of Ihara *et al.* who found a 5% activity for the core-fucosylation of chitotriose compared to the heptassacharide *N*-glycan.<sup>[56]</sup>



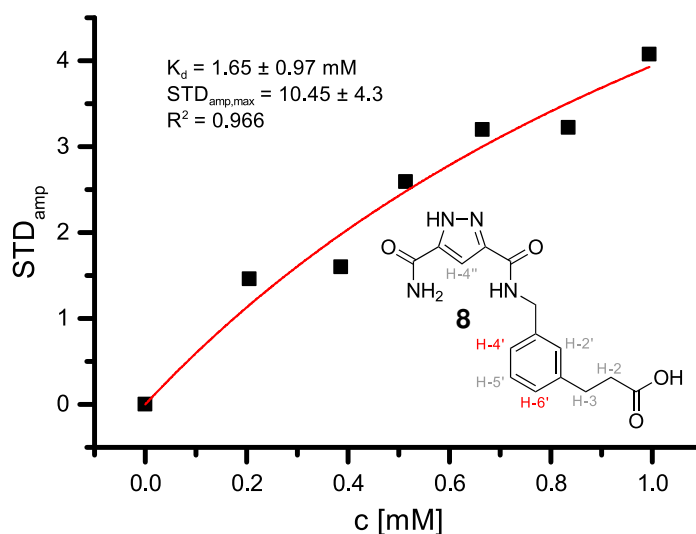
**Fig. 18:** Panel A: FUT8 catalyzes the fucosylation of 1- $\beta$ -*N*-acetylchitotriose **29** resulting in fucosylated 1- $\beta$ -*N*-acetylchitotriose **30**. Panel B: Reaction progress as monitored via  $^1\text{H-NMR}$  spectroscopy at 310 K.

**Tab. 2:** Initial rate constants of the FUT8-catalyzed fucosylation of 1- $\beta$ -*N*-acetylchitotriose **29** as monitored via  $^1\text{H-NMR}$  spectroscopy at 310 K.

| -                                | GDP-Fuc | GDP   | Chitotriose | Fuc-<br>Chitotriose | Fucose |
|----------------------------------|---------|-------|-------------|---------------------|--------|
| $k_{\text{ini}} [\text{s}^{-1}]$ | -0.026  | 0.032 | -0.011      | 0.010               | 0.019  |

### 5.1.6 STD NMR

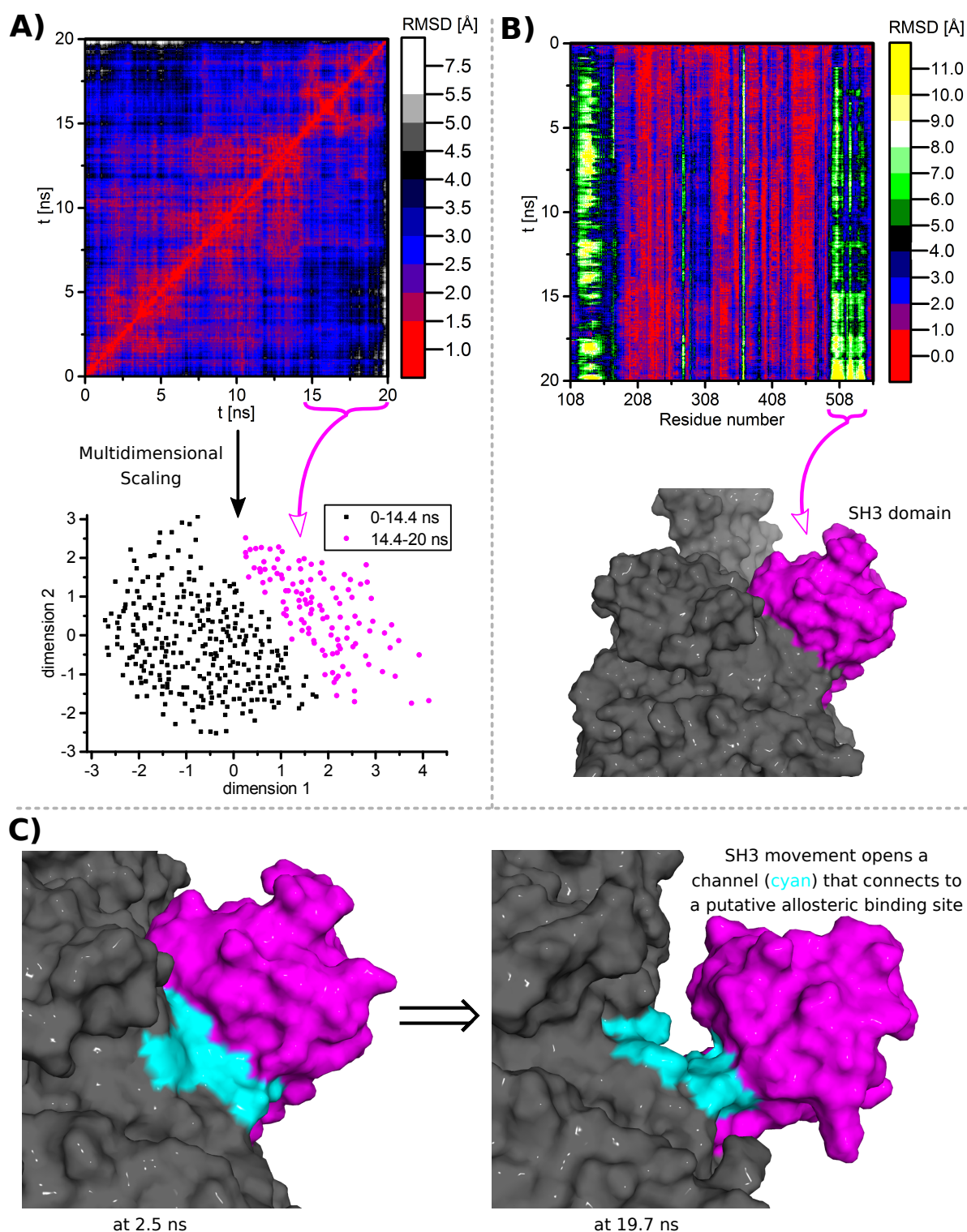
Finally, the dissociation constant for ligand **8** was evaluated. STD NMR revealed a  $K_d$  of  $1.65 \pm 0.97\text{ mM}$  (evaluating H-4'/H-6' (phenyl ring), see **Fig. 19**). A STD NMR spectrum acquired from a sample without added enzyme revealed no STD artifacts. Evaluation of the signal for the H-2' (phenyl ring) and H-4'' (pyrazole) yielded a higher  $K_d$  of 4.47 and 7.72 mM respectively. The signals of H-2 and H-3 (right next to the carboxy function) showed no STD effect. This is in accordance with the proposed binding mode in that all of the latter protons should be more solvent exposed than H-6'. Unfortunately, H-5' (phenyl ring) could not be evaluated because its signal is subsided by an impurity originating from the protein solution (imidazole). Similarly, the benzyl protons could not be evaluated as they are strongly affected by water suppression. Overall, the determined  $K_d$  is significantly higher than anticipated from MM-GBSA calculations (see subsection **5.1.3**).



**Fig. 19:** Determination of the dissociation constant of ligand **8** via STD NMR. The STD amplification factor of H-4'/H-6' (phenyl ring) is plotted against the concentration of ligand **8**. As a result, by employing a one-site binding model the  $K_d$  of ligand **8** was determined to be  $1.65 \pm 0.97$  mM.

### 5.1.7 Identification of a New Putative Binding Site of Fucosyltransferase 8

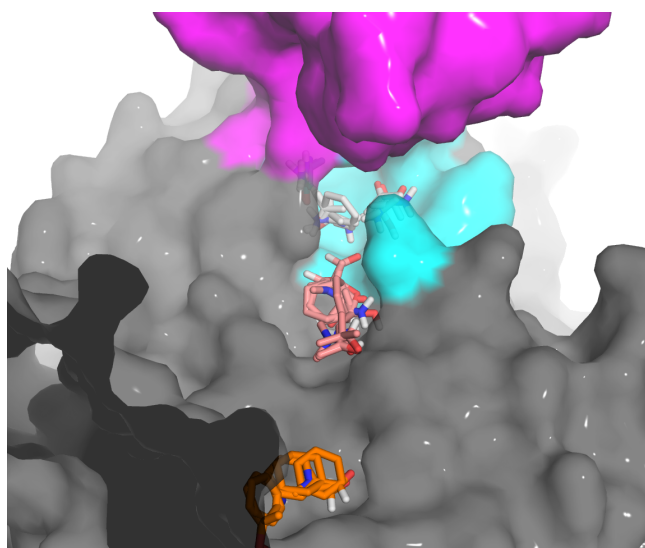
The analysis of an additional MD simulation of a length of 20 ns of the complex of FUT8 with GDP-Fucose showed evidence of a putative allosteric binding site that will be discussed in the following. Panel A of **Fig. 20** displays a matrix of pairwise RMSD values (for all backbone atoms) of structures taken from this MD simulation. The RMSD matrix shows a significant conformational change occurring after approximately 14.4 ns. Next, the origin of this conformational change was analyzed. For this, for all investigated structures the RMSD (of the  $C\alpha$ -atom) of each residue was calculated in reference to the initial frame of this MD simulation. The resulting data is plotted in panel B of **Fig. 20**. The plot reveals that the conformational change is caused by residues of the SH3 domain of FUT8 (colored in magenta in panel B of **Fig. 20**). The SH3 domain of FUT8 is located at the C-terminus.<sup>[48]</sup> SH3 domains are known to play a role in protein-protein interactions e.g. in signal-transduction networks.<sup>[57]</sup> However, for the SH3 domain of FUT8 no such function has been described so far.<sup>[48]</sup> Yet, the SH3 of FUT8 is responsible for the accommodation of the 6-branch of the acceptor *N*-glycan.<sup>[58]</sup> It has been noted previously that FUT8 is significantly more tolerant to modifications of the 6-branch of the acceptor *N*-glycan than to modifications of the 3-branch.<sup>[59]</sup> The significant flexibility of the SH3 domain observed in this MD simulations can explain this acceptor specificity well. Even more interestingly, the observed movement of the SH3 domain of FUT8 opens a channel (see panel C of **Fig. 20**) that connects to a putative binding site (cf. **Fig. 21** and the appendant paragraph).



**Fig. 20:** A MD simulation of FUT8 with a length of 20 ns was performed. **A)** A RMSD matrix (for all backbone atoms) of this MD simulation reveals a significant conformational change occurring after  $\sim 14.4$  ns. Multidimensional scaling (a form of dimensionality reduction) of this RMSD matrix highlights the conformational change in an easily comprehensible fashion.<sup>[60]</sup> **B)** RMSD of the  $C\alpha$ -atom of each residue (as referenced to the initial frame of the MD simulation) plotted against the simulation time. The plot reveals that the conformational change is caused by residues of the SH3 domain (colored in magenta). **C)** Two structures illustrate the previous findings: A significant snap back of the SH3 domain occurs after  $\sim 14.4$  ns. Interestingly, this opens a channel (colored in cyan) that connects to a putative allosteric binding site (cf. **Fig. 21** and the next paragraph).



The druggability of this putative binding site was studied *in silico*. For this FTMap, an application for the identification of binding „hot spots“, was used. Prior work e.g. from Ringe *et al.*, who crystallized proteins in the presence of organic solvent molecules to investigate binding „hot spots“, laid the experimental foundation for *in silico* tools such as FTMap.<sup>[61]</sup> FTMap places small molecular probes (such as acetamide, ethanol and urea) with varying functionality and size onto the surface of the investigated protein and finds likely binding sites for these probes by the use of energy functions.<sup>[62]</sup> The MD-derived structure shown on the right side of panel C in **Fig. 20** was taken to FTMap. Interestingly, on the back side (viewing from the acceptor site) of the channel (that opens in the described MD simulation) four binding „hot spots“ were identified by FTMap (as shown in **Fig. 21**). These four „hot spots“ accommodate a total of 45 probe molecules. In comparison, for the donor binding site of FUT8 (of the same MD-derived structure) FTMap identified four „hot spots“ as well. However, these accommodate only a total of 23 probe molecules. This indicates that this putative binding site might be more easy to target than the donor binding site of FUT8. A ligand binding to this putative binding site and extending into the channel towards the acceptor binding site might lock this conformation of the acceptor site and therefore alter the acceptor preferences of FUT8. For a ligand that extends even further from the channel into the acceptor binding site, disruption of the enzymatic activity of FUT8 is conceivable. Additionally, ligands of this putative binding site might offer specificity: Ligands that target the donor binding site of FUT8 will likely also inhibit other fucosyltransferases as all of them use GDP-Fucose as a donor substrate. The presented hypothesis was not experimentally verified in this thesis due to lack of time.



**Fig. 21:** FTMap identifies four binding „hot spots“ on the back side (viewing from the acceptor site) of the channel (colored in cyan) that opens in the MD simulation. The MD-derived structure shown on the right side of panel C in **Fig. 20** was used. This new putative binding site might be more easy to target than the donor binding site of FUT8 and might offer specificity over other fucosyltransferases.

### 5.1.8 Conclusion

In this chapter, approaches towards the fragment-based *de novo* design of inhibitors of FUT8 were presented. Fragment-based approaches allow for a tight control over molecular properties. As a result, this allows to circumvent non drug-like properties of ligands that have been evident in previous works on inhibitors of FUT8. Starting from a docking regimen of over 700,000 fragments, two interesting fragments were identified and subsequently linked *in silico*. The resulting ligand **8** was synthesized and the corresponding dissociation constant was evaluated to  $K_d = 1.65 \pm 0.97$  mM via STD NMR. In future work, binding of ligand **8** to FUT8 should be analyzed by orthogonal binding assays for two reasons: First, the true  $K_d$  of ligand **8** might be lower than experimentally determined as the ligands most tightly binding atoms (as predicted from the docking pose shown in **Fig. 9**), e.g. the pyrazole NH, cannot be evaluated via STD NMR. Secondly, at this point it remains elusive if ligand **8** binds competitively to the donor site as anticipated. Even though it was certainly hoped for a more affine ligand, ligand **8** can be used as starting point for the development of inhibitors of FUT8 with drug-like properties. Taking the concept of ligand efficiency ( $LE = -\Delta G/\text{no. of heavy atoms}$ ) into consideration,<sup>[47]</sup> ligand **8** features a ligand efficiency of  $LE = 0.17$ . In comparison, the most potent inhibitor of FUT8 known to date, reactive red 120 **3** with a  $K_i$  of  $2 \mu\text{M}$ ,<sup>[46]</sup> exhibits only a ligand efficiency of  $LE = 0.09$  (due to its high molecular weight of  $> 1.3$  kDa). Surely, an efficient binder should exhibit a  $LE > 0.3$  but it should be noted that potent inhibitors have emerged from low-affinity starting points more than once before. For example, Chessari *et al.* reported the development of nanomolar inhibitors of cIAP1 starting from a fragment with an  $IC_{50} > 5$  mM and a ligand efficiency of  $LE < 0.21$ .<sup>[63]</sup> A second ligand, ligand **9**, was synthesized but its binding to FUT8 could not be analyzed due to the ligands surprising decomposition in aqueous media. A take-home message for future success in a computer-aided ligand design process as presented in this chapter is to put a stronger emphasis on synthetic accessibility. Synthesis of ligand **8** and **9** proved to be somewhat lengthy. Drug discovery is a game of attrition and spreading the risk on a higher number of ligands is certainly recommendable.

Furthermore, hints of an allosteric binding site of FUT8 were presented in this chapter. A MD simulation revealed a significant conformational flexibility of the SH3 domain of FUT8 that is known to be responsible for the accommodation of the 6-branch of the acceptor *N*-glycan. In this MD simulation, movement of the SH3 domain opened up a channel leading towards a putative binding site. An *in silico* assessment indicates that this putative binding site might be more easy to target than the donor binding site of FUT8. For ligands that bind to this putative

binding site and extend through the channel into the acceptor site, disruption of the enzymatic activity of FUT8 might be feasible. This very interesting hypothesis certainly calls for future experimental validation.

## 5.2 Ensemble Docking of Renin

Parts of this chapter have been published previously: C. Strecker, B. Meyer, Plasticity of the Binding Site of Renin: Optimized Selection of Protein Structures for Ensemble Docking, *J. Chem. Inf. Model.* **2018**, *58*, 1121-1131.

### 5.2.1 Introduction

X-ray crystallography has proven to be a very powerful tool and is the general accepted gold standard for the representation of protein structures at atomic resolution.<sup>[64]</sup> Unfortunately, X-ray crystal structures provide merely a static picture of proteins that are in truth in motion and can feature astonishing flexibility. Understanding this flexibility of proteins is frequently vital to understand their function as famously condensed by Richard P. Feynman: „[...] *everything that living things do can be understood in terms of the jiggings and wiggings of atoms.*“ Especially for protein-ligand binding events, the classic key-lock principle introduced by Emil Fischer<sup>[65]</sup> proved to be insufficient and has subsequently been replaced by the induced fit and the conformational selection model. The induced fit model proposes that the initial encounter of ligand and protein is accompanied by the formation of weak interactions. These weak interactions induce conformational changes within the binding site that subsequently allow for the formation of stronger protein-ligand interactions.<sup>[66]</sup> The conformational selection model proposes that a protein samples multiple conformations. The ligand binds to one specific conformation of these.<sup>[67]</sup> As a consequence, a protein conformation that is only subordinate in free solution can become the predominant conformation upon ligand binding in this mechanism. The conformational selection model has been proven to be a dominant mechanism of ligand binding.<sup>[68]</sup>

Leads for the development of new drugs are classically obtained by high-throughput screenings of large compound libraries. Advances in automation have made HTS comparably fast and cost-efficient. However, virtual screening methods such as docking still raise the prospect of speeding up the process dramatically while reducing the associated costs.<sup>[69]</sup> Docking is high-throughput while there is no need to maintain a physical compound library. Virtual designed compound libraries can cover a much larger chemical space than combinatorial chemistry could ever reach within a reasonable time scale.<sup>[69]</sup> Docking campaigns can be started long before a robust protein expression protocol has been established. Even in the absence of X-ray crystal structures docking campaigns may be performed by employing homology models.<sup>[70]</sup> Binding poses obtained from docking allow for a straight forward optimization of ligands. While

individual success stories appear time and time again, blind prediction assessments highlight challenges associated with docking.<sup>[71,72]</sup> Issues like the treatment of water molecules, protein flexibility, and the oversimplification of physical interactions by scoring functions add severe complexity to docking.<sup>[29,73]</sup> Various approaches have been employed to account for protein flexibility in docking. „Soft docking“ is a simple approach in that the Lennard-Jones potential is softened so that minor steric clashes do not result in major penalties.<sup>[74]</sup> In a more refined approach rotamer libraries for defined rotatable bonds are employed.<sup>[75]</sup> However, in this approach it is computationally too expensive to take backbone flexibility into account. A very simple approach seems more promising: Ensemble docking, in which ligands are docked into multiple structures and the results are subsequently merged.<sup>[76]</sup> Ensemble docking has been implicated to give enriched docking results in multiple publications.<sup>[77–79]</sup> The “relaxed complex method” from McCammon *et al.* represents an early approach to ensemble docking.<sup>[80]</sup> In this approach structures were extracted from MD trajectories at periodic intervals typically yielding large numbers of considered structures.<sup>[80]</sup> However, the multitude of structures increases the computational cost significantly.<sup>[81]</sup> Additionally, for X-ray crystal structures it has been shown that including too many structures into an ensemble increases the probability of false positives.<sup>[82,83]</sup> Composite grids merge diverse structures into a single representation and as a consequence reduce the multitude of structures efficiently.<sup>[84]</sup> This has been implemented in the docking program FlexE.<sup>[84]</sup> However, this strategy was shown to fail for test systems that exhibit a larger conformational diversity.<sup>[85]</sup> Clustering approaches allow to reduce the number of structures while still covering the significant variations of binding sites.<sup>[77,79,86]</sup> In the presence of sufficient experimental data ensembles can be constructed in a result-oriented manner. This has been achieved by exhaustive combinatorial search (only feasible for small ensemble sizes)<sup>[83]</sup> or by the construction of performance indices.<sup>[87]</sup>

### 5.2.2 Objective

The aim of this project was to systematically analyze the potential and limitations of ensemble docking. In special, the ambition was to analyze the impact of different approaches of ensemble definition. Furthermore, an assessment of the utility of MD-derived structures for ensemble docking was intended. Crystal structures represent the current gold standard for docking, however their value is limited by the challenges associated with obtaining larger numbers of crystal structures that sufficiently reflect protein flexibility. In contrast, MD simulations allow for the quick and convenient exploration of protein flexibility. Previous studies on ensemble docking

that utilized MD-derived structures are of limited meaningfulness because they typically used single, unrepeated MD simulations (see for example the work of Osguthorpe *et al.*).<sup>[78]</sup> These short and unrepeated MD simulations can be heavily dependent on the initial velocities that are randomly assigned. Their conclusions might therefore be unreproducible.<sup>[88]</sup>

### 5.2.3 Introduction of Tools and Metrics and Development of a Test System

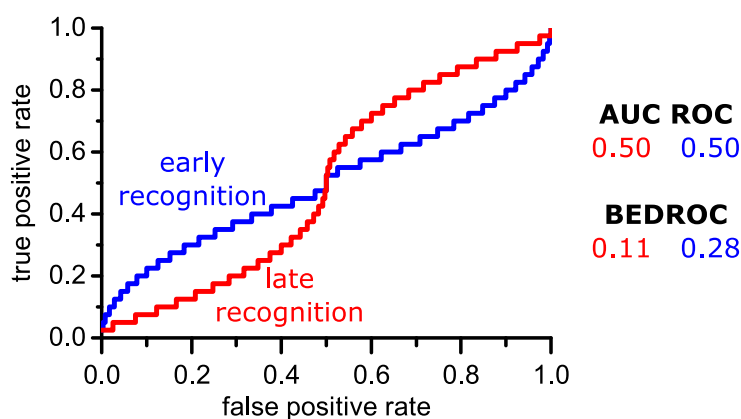
In the following, renin is established as a test case to study ensemble docking. Renin is a suitable test system as a relevant number of crystal structures are available and the binding site of the enzyme is known to feature significant flexibility.

There are three ways to assess the performance of a docking algorithm:<sup>[89]</sup>

1. **Pose prediction:** The ability to correctly reproduce an experimentally determined ligand binding pose. The binding site of the docked protein-ligand complex is aligned to an experimentally determined structure of the protein-ligand complex. Then, the RMSD between the docked and „real“ ligand binding pose is computed. Docking is declared successful if the RMSD is below a cut-off value (typically 2 Å). Docking algorithms frequently perform quite well in this discipline.
2. **Screening utility:** The ability to rank experimentally known binders (actives) above non-binders (nonactives, also coined decoys in this context). This ability can be captured by metrics such as AUC ROC or BEDROC that are introduced in the next paragraph. The decoys used for this assessment can originate from experimental screening results. Frequently however, sufficiently large enough libraries of experimentally determined decoys are not publicly available and therefore studies resort on *in silico* tools to generate decoys. In this work, the „directory of useful decoys-enhanced“ (DUD-E) was used to generate decoys.<sup>[90]</sup> DUD-E generates decoys that are topological dissimilar to the active ligands (by using 2D similarity fingerprints) but have similar physiochemical properties.<sup>[90]</sup> This is of importance because docking algorithms have been shown to discriminate actives from decoys on the basis of physiochemical properties such as charge or molecular weight.<sup>[90]</sup> However, the occurrence of potentially active ligands among the generated decoys can obviously not be ruled out. But chances are low, as for any known target the subspace of binders represents only a fingertip of all chemical matter.
3. **Scoring accuracy:** The ability to correctly rank a set of active ligands. Frequently, the respective docking scores are plotted against experimental values of affinity for this

purpose. After linear regression, the coefficient of determination ( $R^2$ ) is used as a metric. In this context, an ideal performance would result in a  $R^2$  of 1. Docking algorithms frequently perform very poorly in this discipline because the underlying scoring functions oversimplify the complexity of physical interactions.<sup>[29]</sup>

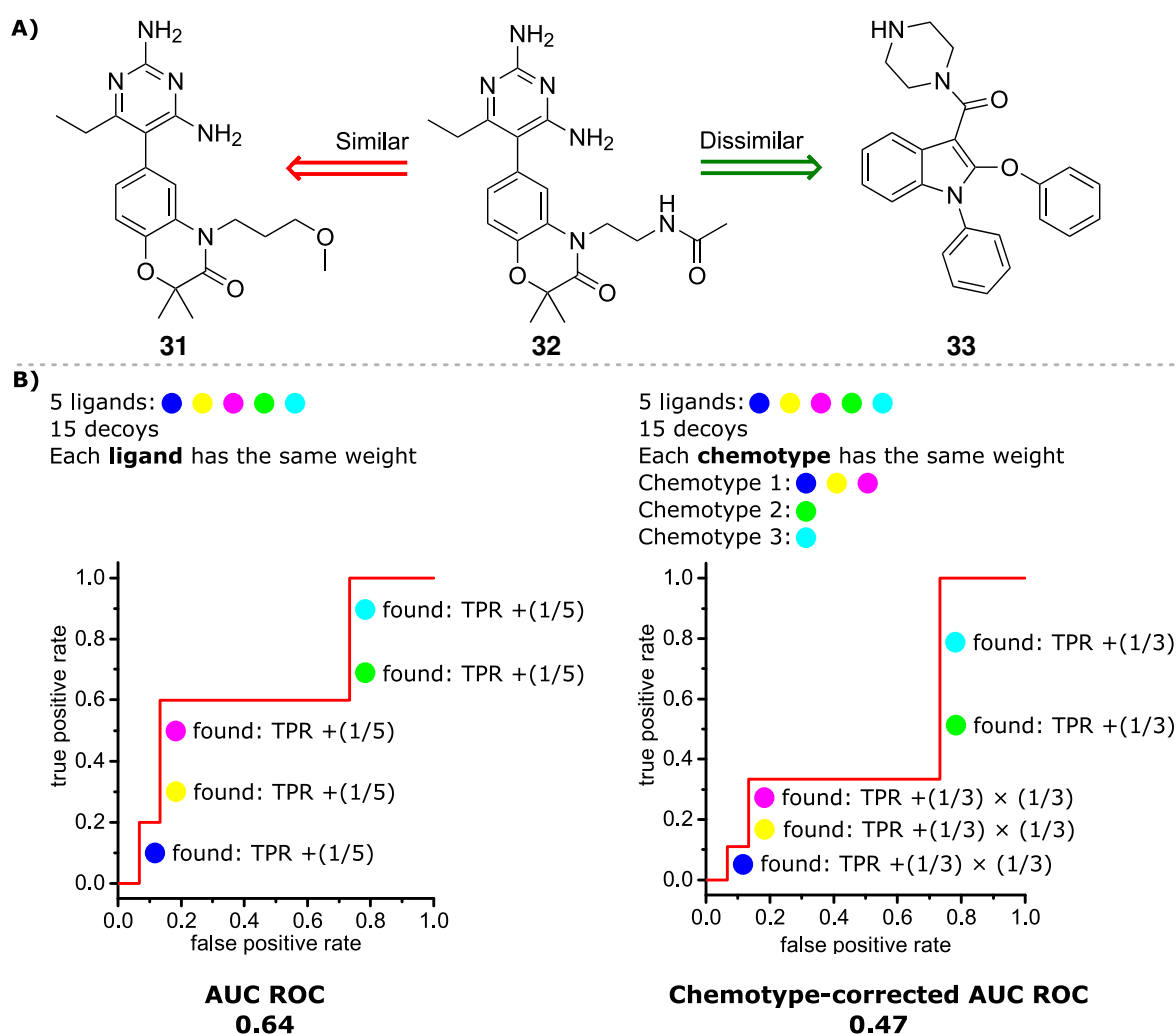
As stated above, the area under the curve of the receiver operating curve (AUC ROC) is a metric of the screening utility. For the receiver operating curve, the true positive rate is plotted against the false positive rate. This means that the correct recognition of an active ligand adds  $1/(\text{total number of actives})$  to the y-value. In contrast, the incorrect identification of a decoy as an active ligand adds  $1/(\text{total number of decoys})$  to the x-value. The AUC ROC is bound between 0 and 1 with the latter corresponding to a perfect performance. A value of 0.5 corresponds to random performance. However, the AUC ROC does not reward early recognition of active ligands (as illustrated in **Fig. 22**). Early recognition of ligands is worth more than late recognition because in a real life scenario only an initial scrap from the top of the docking results will be investigated experimentally due to limited time and budget. To account for early recognition, the BEDROC score (Boltzmann-enhanced discrimination of the receiver operating characteristic) applies an exponential weight onto the receiver operating curve.<sup>[91]</sup> The BEDROC score is still bound between 0 and 1 with the latter corresponding to a perfect performance.<sup>[91]</sup> It exhibits a free parameter  $\alpha$  which was set to 20 in this work. This means that the top 8% of the docking hits account for 80% of the BEDROC score.



**Fig. 22:** Two hypothetical receiver operating curves. The blue curve simulates an early recognition scenario, the red curve a late recognition scenario. The AUC ROC metric fails to discriminate the divergent nature of both curves: Both perform with a random performance of 0.5. In contrast, the BEDROC score (calculated with  $\alpha = 20$ ) rewards the early recognition scenario.

Frequently, test systems used in docking studies feature larger groups of ligands with high structural similarity (in this context termed „chemotype“). In panel A of **Fig. 23** an example of two ligands, **31** and **32**, belonging to the same chemotype is shown. A docking algorithm that

is able to correctly identify ligand **32** as a ligand can be expected to identify ligand **31** as well. Consequently, results from docking studies that feature test systems with very similar ligands can be expected to be heavily biased. Therefore, a metric for the screening utility (additional to the AUC ROC metric) that rewards the identification of different chemotypes is needed. In this work, the chemotype-corrected AUC ROC metric proposed by Clark *et al.* is used.<sup>[92]</sup> While the AUC ROC metric rewards the correct identification of an active by adding  $+(1/\text{total number of actives})$  to the true positive rate, the chemotype-corrected AUC ROC metric adds  $+(1/\text{total number of chemotypes}) \times (1/(\text{total number of members in the chemotype group of the identified active}))$ . A comparison of both metrics is shown in panel B of **Fig. 23**.

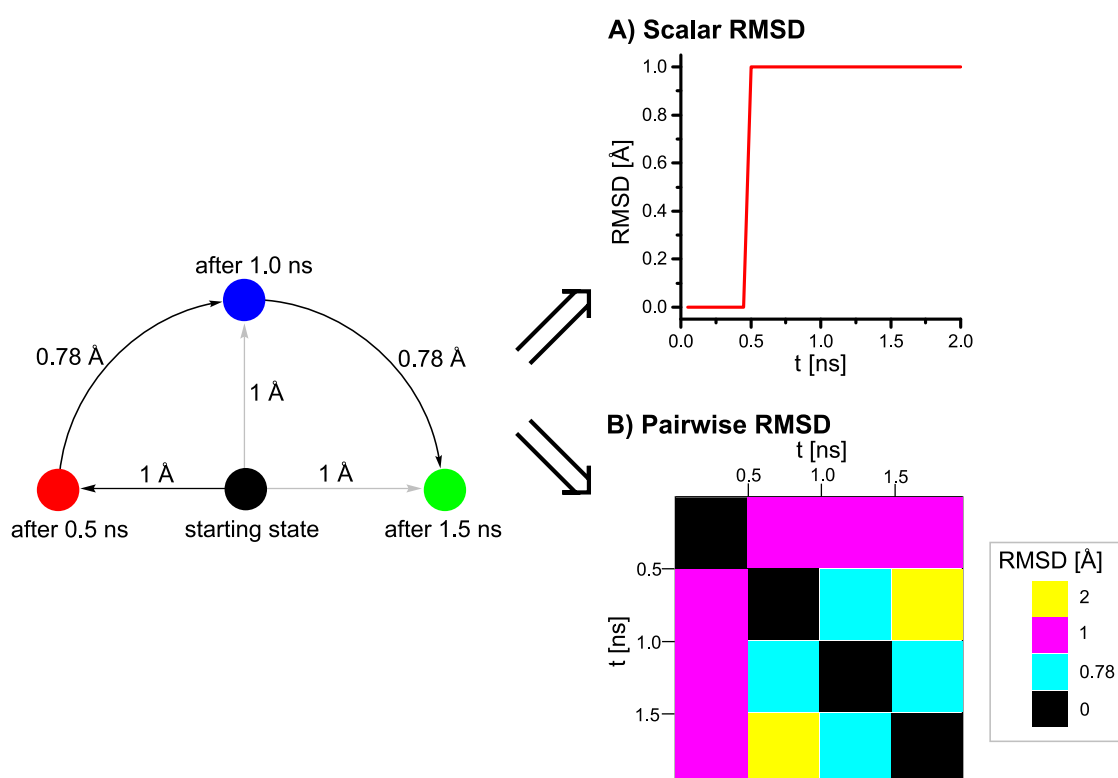


**Fig. 23: A)** Ligand **31** and **32** are structural similar and belong to the same chemotype. In contrast, ligand **33** is structurally dissimilar. **B)** Results from a hypothetical docking study. On the left side, the AUC ROC metric puts the same weight on every identified ligand. On the right side, the chemotype-corrected AUC ROC metric puts the same weight on every identified chemotype.

For ensemble docking, ligands are docked into multiple structures (=ensemble) to account for effects of protein flexibility. It is obviously undesirable to use an ensemble of protein



structures that are highly alike. As a result, descriptors that discriminate between different protein structures are required. This can be a single observable like a distance between two atoms or a dihedral angle. However, these can obviously only describe very local changes. On a less local scale, the root mean square deviation (RMSD) between sets of multiple atoms can be calculated. In the analysis of MD simulations, trajectories that plot RMSD values against the simulation time (like the one schematized in panel A of **Fig. 24**) can be frequently encountered. For this, the RMSD value is calculated in reference to a particular protein structure (usually at  $t = 0$ ). However, these scalar RMSD values are of very limited utility because behind a single RMSD value a broad range of different protein structures may be hidden (as illustrated in **Fig. 24**).<sup>[93]</sup> Matrices of pairwise RMSD values between all protein structures are able to reveal these hidden conformational changes (see panel B of **Fig. 24**).<sup>[93]</sup>



**Fig. 24:** A hypothetical scenario of conformational states sampled during an MD simulation. The protein remains in the conformational starting state (black circle) for 0.5 ns and then immediately jumps to a second conformational state (red circle) that is 1 Å distant. Then, after 1.0 and 1.5 ns subsequent jumps to conformational states (blue and green circle) occur that correspond to a circular motion around the starting state (black circle). **A)** The RMSD value is calculated in reference to the starting state. Only the first conformational jump is detected (black to red circle). The other jumps remain hidden. **B)** A matrix of pairwise RMSD values reveals all the conformational states.

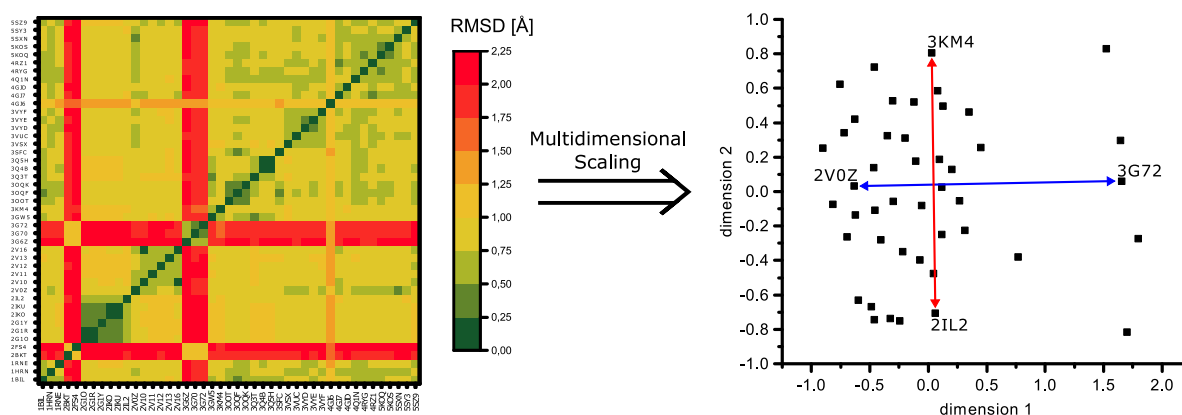
Renin is an aspartyl protease that catalyzes the conversion of angiotensinogen to angiotensin I and hence part of the renin-angiotensin-aldosterone system (RAAS). RAAS controls blood pressure and is targeted by antihypertensive drugs. 45 inhibitors of renin were selected for that

both a crystal structure and an IC<sub>50</sub> value (of the respective stereoisomer present in the crystal structure) were available. The potency of these renin inhibitors spans from pIC<sub>50</sub> 4.37 to 9.70. The ligands were manually partitioned into 11 different chemotype subgroups on the basis of common chemical substructures (see **Appendix 9.1**). This wealth of chemotypes indicates a challenging test system. For the subsequent evaluation of the screening utility, 50 decoys were generated for each renin inhibitor using the DUD-E function.<sup>[90]</sup>

To classify the 45 renin crystal structures according to their structural similarity, a matrix of pairwise RMSD values of their binding sites was generated. In **Fig. 25** the corresponding RMSD matrix is shown. This matrix was generated in PyMOL using a simple Python script (see **Appendix 9.2**). In this context, the binding site of renin was defined to cover residues that are within 4 Å of the primary ligand in any of the 45 minimized crystal structures. In later MD simulations it was found that the side chains of four residues, that do not contribute to ligand binding, dominate structural heterogeneity. For that reason Glu189, Phe319, and Asp320 were completely omitted from the binding site definition, and for Arg148, only the side chain was removed. Because interpretation of pairwise RMSD matrices is rather confusing, multidimensional scaling was used to visualize the structural heterogeneity of the crystal structures. In **Fig. 25** the result of multidimensional scaling (in this case 2D) of the pairwise RMSD matrix of all crystal structures is shown. Multidimensional scaling is a method for the visualization of distance matrices and represents a form of dimensionality reduction.<sup>[60]</sup> For each object given in the distance matrix initially a data point in a space of predefined dimensionality is created. The stress value, which compares the Euclidian distances between the data points created to the distances given in the distance matrix, is calculated. In this work Sammon's definition of the stress value is used:<sup>[94]</sup>

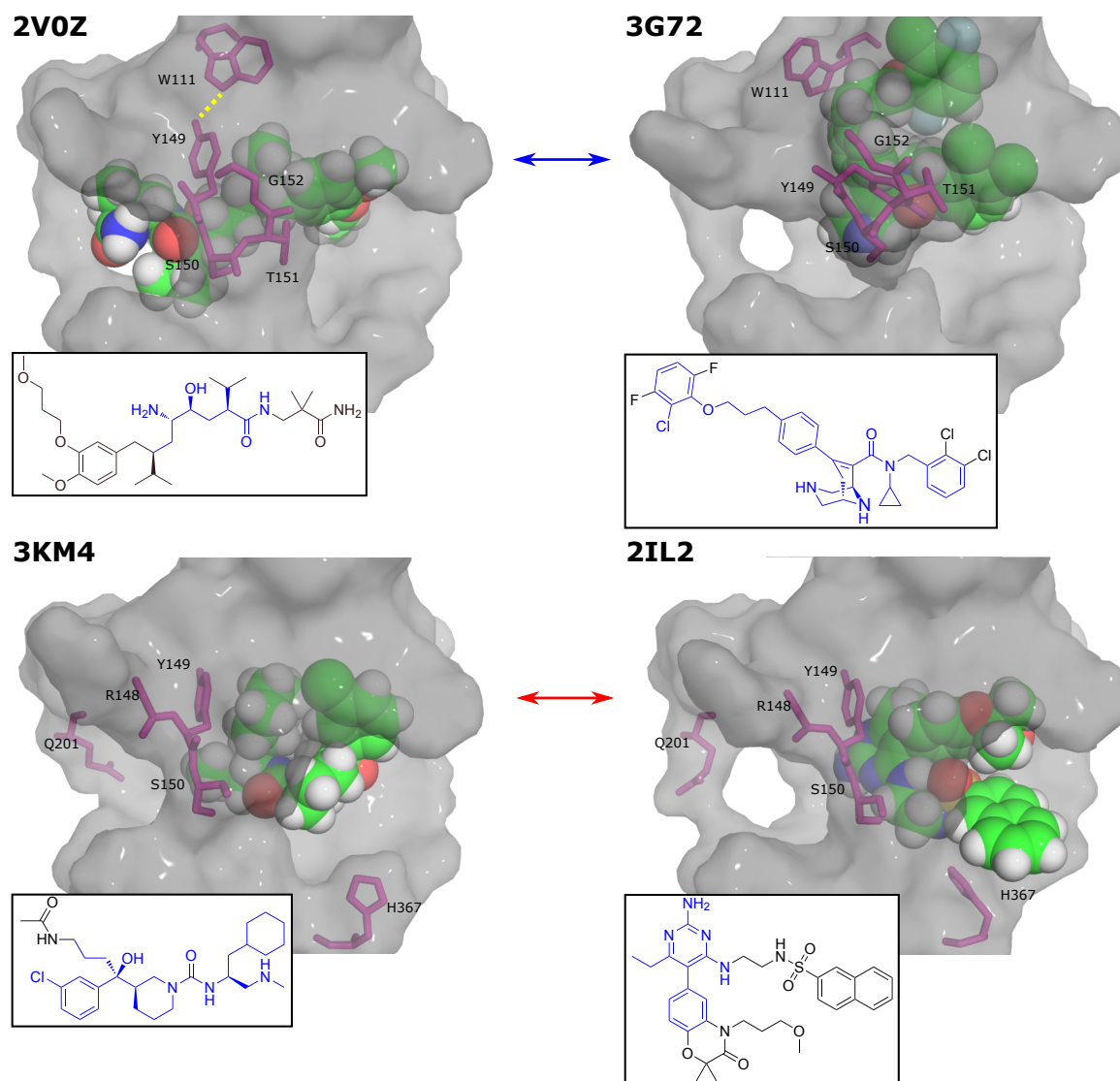
$$stress = \frac{1}{\sum_{i < j} [d_{ij}^*]} \sum_{i < j} \frac{[d_{ij}^* - D_{ij}]^2}{d_{ij}^*}$$

where  $d_{ij}^*$  denotes the distance between i-th and j-th object in the distance matrix and  $D_{ij}$  the distance between data points in the representation of the dimensionality defined above. Data points are then rearranged iteratively in order to minimize the stress value. Multidimensional scaling was performed using Matlab's "mdscale" command using "Sammon" as criterion.



**Fig. 25:** Left: RMSD matrix of the binding sites of the 45 renin crystal structures in the described test system. Right: Data points obtained by multidimensional scaling (with two dimensions) performed onto the RMSD matrix. Residues contributing mainly to RMSD between the crystal structures connected by the colored arrows are (blue) 1. Thr151 (6.174 Å) 2. Tyr149 (5.737 Å) 3. Ser150 (5.001 Å) 4. Gly152 (4.791 Å) 5. Trp111 (3.967 Å); (red) 1. Gln201 (2.916 Å) 2. Ser150 (2.694 Å) 3. Arg148 (2.290 Å) 4. His367 (2.266 Å) 5. Tyr149 (2.214 Å).

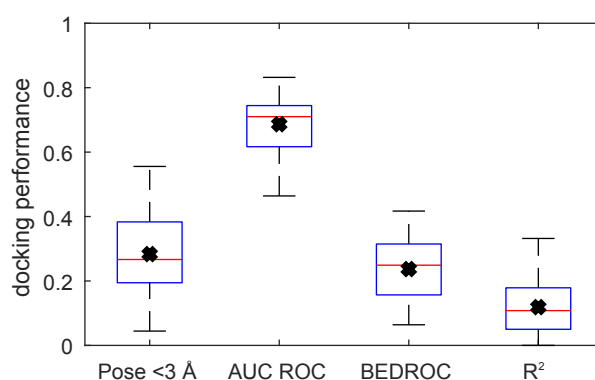
Next, residues of the binding pocket that are major contributors to structural heterogeneity were analyzed. For this purpose, the crystal structures at the endpoints of the arrows depicted in **Fig. 25** were aligned, which were chosen to be approximately parallel to the x- and the y-axis, respectively. The blue arrow represents the opening of the flap indicated by movements of residues Thr151, Tyr159, Ser150, and Gly152. This is the major cause of structural heterogeneity in this dimension. Flap opening is also accompanied by a major repositioning of the non flap residue Trp111.<sup>[95]</sup> Along the red arrow depicted in **Fig. 25** Gln201 undergoes a major repositioning. Flap residues contribute also significantly to structural heterogeneity in this dimension. The binding sites of the four crystal structures corresponding to the endpoints of the arrows are visualized in **Fig. 26**. The displayed structural changes clearly impact ligand binding. In the upper left panel of **Fig. 26** (structure of 2V0Z) a hydrogen bond between Trp111-N $\epsilon$  and Tyr149-OH is present. This hydrogen bond is disrupted in the 3G72 structure (upper right panel) which opens a sub-pocket that is filled by the respective ligand. The orientation of Gln201 displayed in the 3KM4 structure (bottom left) is clearly unfavorable for binding of aliskiren (shown in the 2V0Z structure – top left). In the 2IL2 structure (bottom right) His367 has an orientation that allows for  $\pi$ - $\pi$  interaction with the ligand's naphthyl moiety. In contrast, the orientation of His367 in the 3KM4 structure (bottom left) would be unfavorable in this context.



**Fig. 26:** Binding sites of the four renin crystal structures at the arrow endpoints in **Fig. 25**. Ligands are displayed as ball-and-stick model. Residues that contribute mainly to the structural heterogeneity between the pairs are displayed in magenta: (upper panel) Thr151, Tyr149, Ser150, Gly152, and Trp111; (lower panel) Gln201, Ser150, Arg148 (backbone only), His367, and Tyr149. It is clearly visible that the changes of the protein structures impact ligand binding (cf. text). The chemical structures of the respective ligands are shown. The substructures used for the definition of chemotypes (cf. **Fig. 32** and the appendant paragraph) are highlighted in blue.

Next, the individual docking performance of the respective 45 crystal structures was evaluated. This allows to pigeonhole the performance of ensembles that will be evaluated in the next section. First, pose prediction was evaluated (cut-off value of 3 Å). On average, docking into an individual crystal structure yields a correctly docked pose for 28% of the ligands (with a standard deviation of 12.5%). The distribution of individual values is shown in a boxplot in **Fig. 27** (see **Appendix 9.3** for raw data). Subsequently, the screening utility was evaluated. For the test system used in this work, the mean AUC ROC value of all 45 crystal structures is 0.686 with a standard deviation of 0.08. The screening utility in terms of AUC ROC can on average be assessed as satisfactory: The performance is above random for all but one of the crystal

structures but there is certainly room for improvement. To account for early recognition the BEDROC score was evaluated as well. The mean BEDROC score of all 45 crystal structures is 0.237 with a standard deviation of 0.10. This leaves massive space for improvement of early recognition. The distribution of individual BEDROC scores is reflected as a boxplot in **Fig. 27**. Finally, the scoring accuracy was evaluated. After linear regression, the mean coefficient of determination of all 45 crystal structures is 0.119 with a standard deviation of 0.083. This can be considered as a very bad performance and is reflecting well-known issues associated with scoring functions.<sup>[29]</sup> It is disillusioning that for some individual crystal structures (e.g., 2G1R, 5SXN) there is no correlation between  $pC_{50}$  and docking score at all. Again the distribution of individual  $R^2$  values is shown in a boxplot (cf. **Fig. 27**).



**Fig. 27:** Docking performance of the individual crystal structures in the test system used in this work. Docking into three quarters of the crystal structures yields correctly docked poses for less than 40% of the ligands. The overall scoring accuracy is disillusioning. The additional black cross indicates the respective mean value.

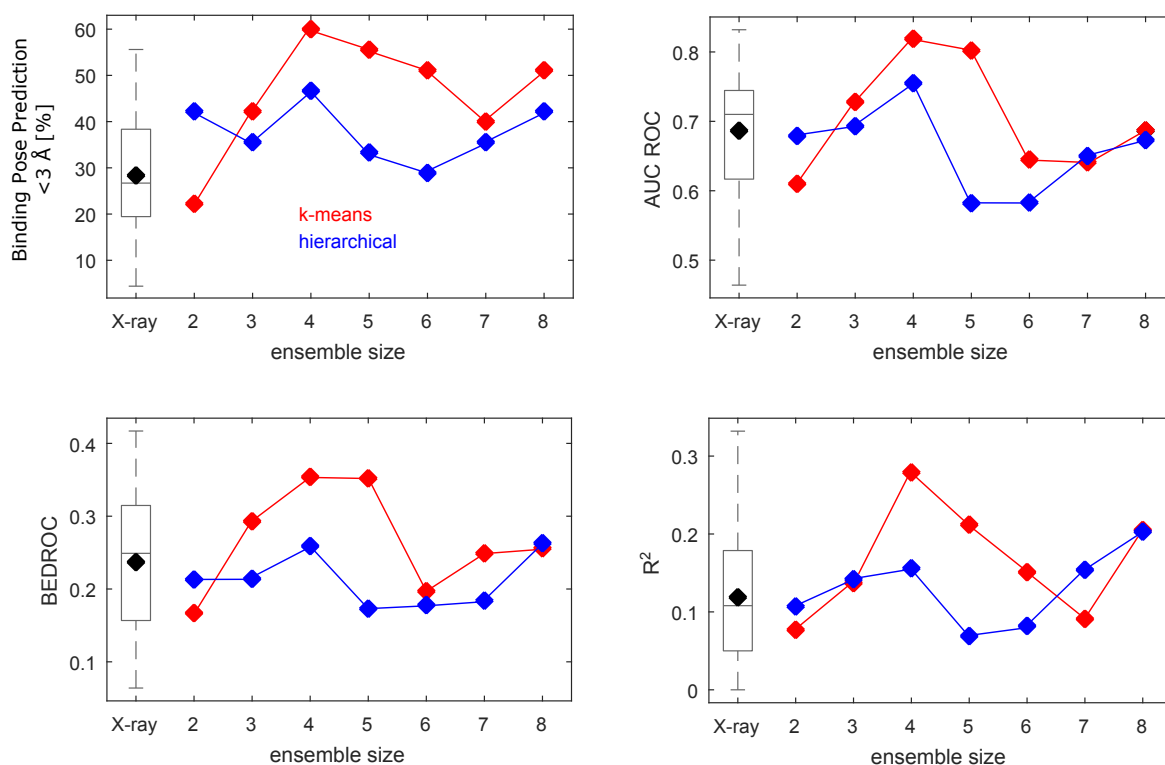
#### 5.2.4 Evaluation of Clustering Approaches for Ensemble Docking

In the next step, the docking performance of ensembles constructed from the 45 crystal structures in the test system was analyzed. Previous studies of ensemble docking have employed various protocols to select structures that should be included in ensembles. Structures have been randomly selected,<sup>[96]</sup> while others have used the clustering function implemented in GROMOS<sup>[77,85,97]</sup> that was developed by Daura *et al.*<sup>[98]</sup> Osguthorpe *et al.* employed hierarchical clustering with average linkage.<sup>[78]</sup> Tian *et al.* used phylogenetic trees for structure selection.<sup>[79]</sup> However, there are no comparative studies that evaluate the impact of structure selection on ensemble docking. In this work, the docking performance of ensembles constructed from  $k$ -means and hierarchical clustering with average linkage was evaluated. These two algorithms have been described as useful for clustering of MD trajectories by Shao *et al.*<sup>[99]</sup>  $K$ -means is a clustering algorithm that partitions data into a predefined number ( $k$ ) of

clusters. Following steps are iterated during *k*-means clustering:

1. *k* cluster centers (means) are generated randomly.
2. Each data point is assigned to the cluster center that has the least squared Euclidean distance.
3. The centroid of all data points belonging to one cluster becomes the new cluster center.
4. Step 2 and 3 are repeated until a convergence criterion is met. Because the results are dependent on the initial generation of cluster centers (that is random, see step 1), it is common to do multiple runs with independent starting conditions.

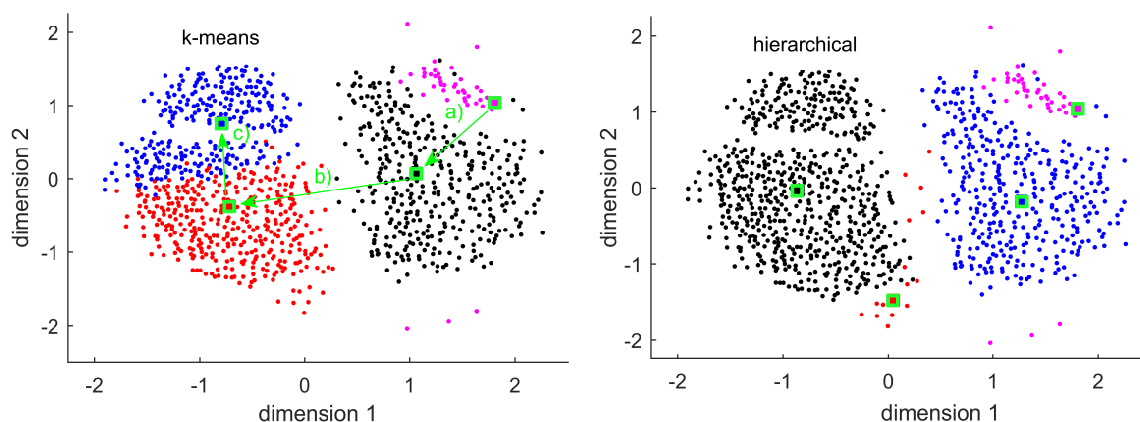
*K*-means clustering is sometimes met with precaution due to its tendency to produce evenly sized clusters. However, the working hypothesis of this study is that the sensitivity to outliers observed for hierarchical clustering with average linkage is far more critical. The docking performance of ensembles of crystal structures constructed by *k*-means and hierarchical clustering with average linkage in terms of binding pose prediction, screening utility and scoring accuracy is depicted in **Fig. 28**. Ensemble sizes ranging from 2 to 8 were evaluated. From **Fig. 28** it is evident that improved binding pose prediction correlates with improved AUC ROC values. This indicates that an erroneous enrichment of active ligands does not take place. Furthermore, the AUC ROC and the BEDROC metric seem strongly coupled indicating that no improvement/deterioration of early recognition is introduced by the construction of ensembles. At various points it can be seen that addition of new structures to the ensemble is not always improving the docking performance and might actually decrease it at times. This behavior has been reported before.<sup>[83]</sup> It can further be seen that only for a relatively small share of the ensembles the performance is better than the mean performance of individual crystal structures. No ensemble performs better than the best performing crystal structure in terms of screening utility as well as scoring accuracy. This is in accordance to previous findings of Craig *et al.*, who concluded that ensembles rarely perform better than the best performing crystal structure.<sup>[100]</sup> However, nobody knows upfront which is the best performing crystal structure. In this study, the best performance in terms of all four evaluated metrics is observed with *k*-means clustering and an ensemble size of four. The respective ensemble is able to reproduce more native ligand poses correctly than any individual crystal structure and performs better than 88% of all crystal structures in the test system in terms of screening utility as well as scoring accuracy. Here, 60% of the docked poses are correct, the respective AUC ROC value is 0.819 (better than 97% of the crystal structures), the BEDROC score 0.354 (better than 88% of the crystal structures), and  $R^2$  is 0.279 (better than 97% of the crystal structures). It should be noted though that there is no general, prospective method to predict the optimal ensemble size.



**Fig. 28:** Docking performance of ensembles constructed from crystal structures. The boxplots labeled “X-ray” reflect the docking performance of individual crystal structures, the black diamond their mean value. The red diamond reflects the performance of ensembles constructed from *k*-means clustering, the blue diamond the performances of ensembles constructed from hierarchical clustering with average linkage. For docking, *k*-means clustering yields better performing ensembles than hierarchical clustering with average linkage.

Finally, the docking performance of ensembles constructed from MD-derived structures was evaluated. Crystal structures are regarded to be the gold standard for docking. However, especially at the beginning of a lead discovery project their availability is often limited and thus MD simulations have a huge potential in this scenario because they allow for the quick and convenient exploration of protein flexibility. MD simulations with a length of 50 ns were performed starting from the 2G1R crystal structure after deletion of the ligand. The 2G1R crystal structure was chosen as starting point as it exhibits a screening utility close to the mean of all crystal structures (AUC ROC: 0.640; BEDROC: 0.197). In contrast, the 2G1R structure exhibits no scoring accuracy at all in this test system ( $R^2 = 0.001$ ). In order to increase reproducibility, MD simulations were repeated ten times from the same input structure with randomized initial velocities. For each MD run, docking was performed into ensembles constructed by *k*-means and hierarchical clustering with average linkage. In **Fig. 29** the result of multidimensional scaling onto a RMSD matrix of structures obtained from one particular MD run (plus the crystal structures of this test system) is shown. It should be noted that this MD run exhibits particular well-separated states. First of all, it is apparent that the structural

overlap between crystal structures and MD-derived structures is rather low. This statement is true for all 10 MD runs. Graphical inspection roughly divides the depicted MD-derived structures into three clusters. Therefore,  $k$ -means clustering (with  $k = 3$ ) and hierarchical clustering with average linkage (maxclust = 3) was performed onto this data set and each structure was colored according to its respective cluster membership. For  $k$ -means clustering it is apparent that the blue cluster extends too far into the red cluster. However, the cluster representatives fall into the clusters that graphical inspection would have constructed for  $k$ -means clustering. To the contrary, hierarchical clustering fails to divide the two clusters on the left-hand side of the distribution and rather creates a third cluster in red that represents a small number of outliers. These traits of hierarchical clustering with average linkage are certainly unfavorable for the selection of structures from MD simulations for ensemble docking. Then, it was analyzed which features set the MD-derived structures apart from the crystal structures. For this purpose, the structures at the end points of the green arrows depicted in **Fig. 29**, which correspond to the cluster representatives, were aligned. It can be seen that the MD-derived structures are dominantly set apart from the crystal structures by reorientation of flap residues (cf. legend of **Fig. 29**).

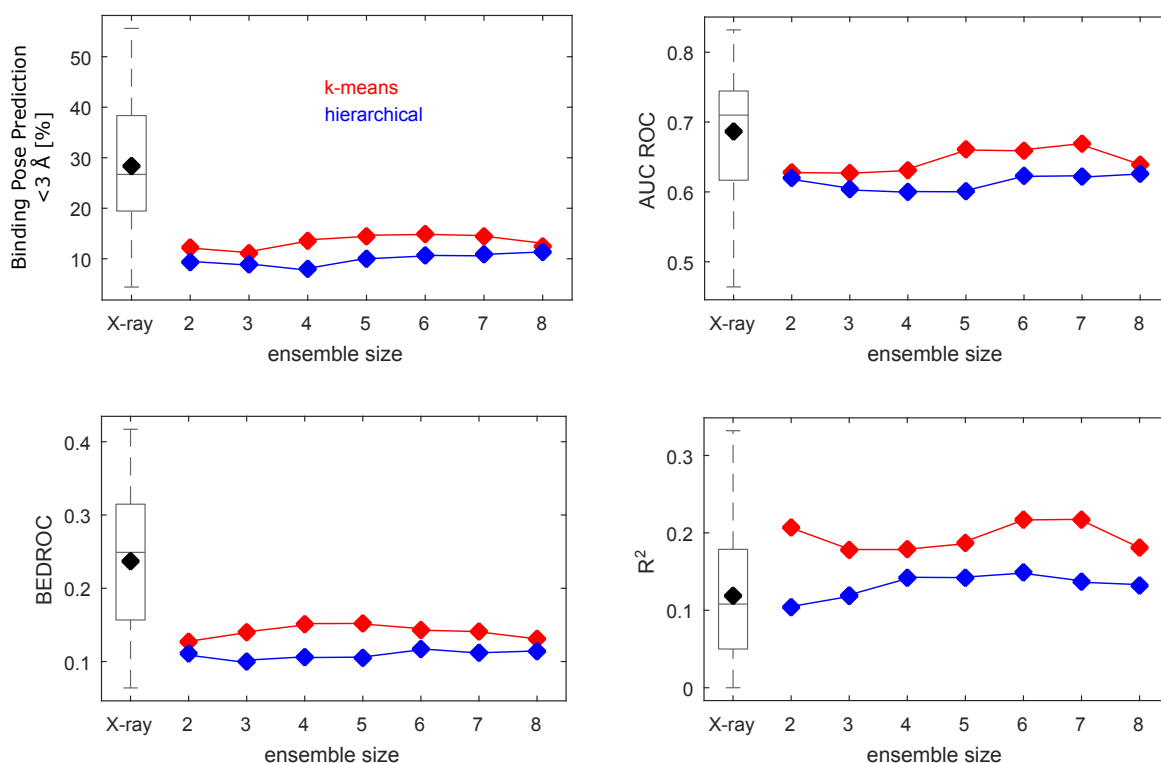


**Fig. 29:** Data points obtained by multidimensional scaling (with two dimensions) performed for the pairwise RMSD matrix of structures derived from a MD simulation of renin with a length of 50 ns. Crystal structures are added and colored in magenta. Data points are colored to represent their cluster membership: On the top clustering was performed using  $k$ -means ( $k = 3$ ); in the bottom panel hierarchical clustering with average linkage (maxclust = 3) was employed. Hierarchical clustering creates unfavorable clusters of some few outliers (cf. text). Data points framed by green squares indicate the cluster representatives (plus the MD starting point). Residues contributing majorly to RMSD along the indicated green arrows are (a) Gln201 (3.326 Å), Ile203 (3.101 Å), Thr151 (3.079 Å), His127 (2.384 Å), Pro372 (2.181 Å); (b) Ser150 (5.675 Å), Thr151 (4.905 Å), Arg148 (4.580 Å), Gly152 (4.426 Å), Tyr149 (4.246 Å); (c) Arg148 (3.474 Å), Gly152 (3.092 Å), Ser150 (3.062 Å), Ile203 (2.819 Å), Thr151 (2.718 Å).

The docking performance of ensembles of MD-derived structures constructed by  $k$ -means and hierarchical clustering with average linkage is depicted in **Fig. 30**. The values represent the average of 10 repeated MD runs with randomized initial velocities. Ensemble sizes ranging

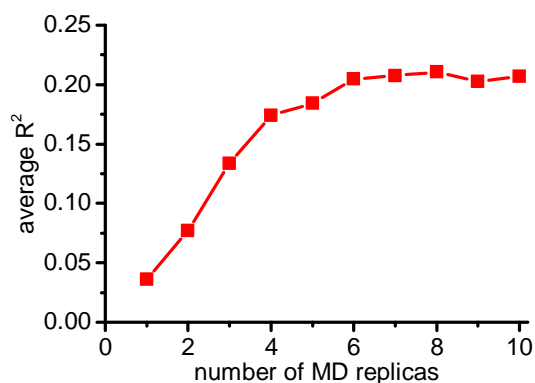


from 2 to 8 were evaluated. Compared to docking into crystal structures it is evident that docking into MD-derived ensembles yields about 50% of correctly docked poses (cf. **Fig. 30**). In terms of screening utility MD-derived ensembles perform not as good as the mean values obtained for the performance of individual crystal structures at all ensemble sizes. Structure selection seems to be of great importance for ensemble docking. For ensembles constructed by hierarchical clustering the AUC ROC metric does not change significantly with the size of the ensembles. A slight increase is found at higher ensemble sizes. For ensembles constructed by *k*-means clustering the AUC ROC has its maximum at an ensemble size from 5 to 7 with the AUC ROC peaking at an ensemble size of 7 (AUC ROC = 0.669). This is close to the mean performance of individual crystal structures. The maximum for the BEDROC score is reached at an ensemble size of five with BEDROC = 0.152. Again, ensembles constructed by hierarchical clustering with average linkage perform worse than those constructed by *k*-means clustering. Finally, it is surprising that the scoring accuracy of ensembles constructed from MD-derived structures is superior to the mean performance of individual crystal structures (see **Fig. 30** bottom right). Again, ensembles constructed by *k*-means clustering perform better than ensembles constructed by hierarchical clustering. It is worth noting that ensembles of MD-derived structures constructed by *k*-means clustering perform on average better than 75% of all crystal structures for all ensemble sizes evaluated in this study. This is particularly astonishing because the 2G1R crystal structure that served as the starting point for subsequent MD simulations exhibits no correlation between docking scores and experimental pIC<sub>50</sub> values at all.



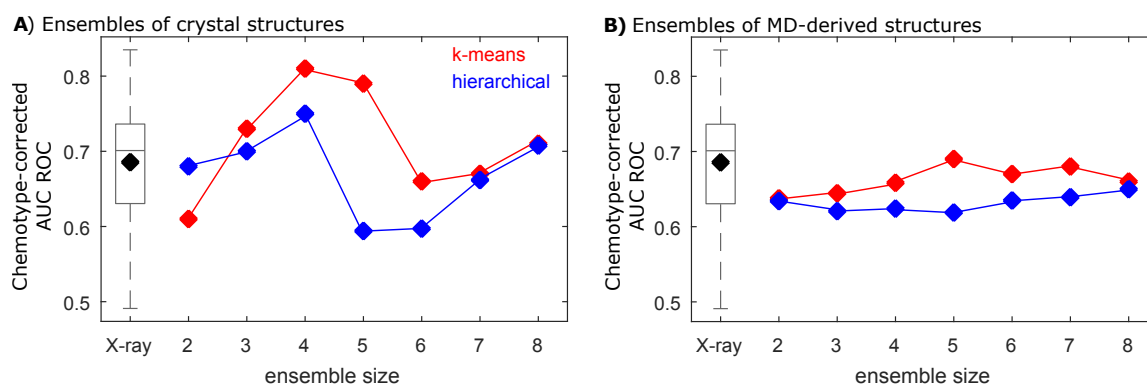
**Fig. 30:** Docking performance of ensembles constructed from MD-derived structures. The boxplots labeled „X-ray“ reflect the docking performance of individual crystal structures, the black diamond their mean value. The red diamond reflects the performance of ensembles obtained by *k*-means clustering, and the blue diamond, the performances of ensembles obtained by hierarchical clustering with average linkage. The values for MD-derived ensembles represent the average of 10 repeated MD runs with randomized initial velocities. In terms of screening utility MD-derived ensembles perform not quite as well as crystal structures. However, MD-derived ensembles are superior to crystal structures in terms of scoring accuracy. Overall, *k*-means clustering seems to yield better performing ensembles than hierarchical clustering with average linkage.

As previously stated, many preceding studies on ensemble docking that utilized MD-derived structures have used single, unrepeated MD simulations (see for example the work of Osguthorpe *et al.*).<sup>[78]</sup> These short and unrepeated MD simulations can be heavily dependent on the initial velocities that are randomly assigned. Their conclusions might therefore be unreproducible.<sup>[88]</sup> In this work, MD simulations were repeated ten times (resulting in a total 500 ns simulation time) and the values that are reported in the previous paragraph (and that are shown in **Fig. 30**) are averages of these ten replicas. In **Fig. 31** the importance of these replicas is highlighted. Drawing conclusions from single, unrepeated MD simulations can result in grossly wrong insights. Importantly, **Fig. 31** illustrates that the values that are reported in this work appear to have converged within reasonable limits. All data obtained from individual MD replicas is summarized in the appendix (see **Appendix 9.6**).



**Fig. 31:** Shown is the average  $R^2$  (as a metric of scoring accuracy) as a function of the number of MD replicas for ensembles constructed from MD-derived structures by  $k$ -means clustering (ensemble size = 2). In this work, MD simulations were repeated ten times with randomized initial velocities. In this case, performing only a single, unrepeated MD simulation would have resulted in a very wrong estimate of  $R^2$ . Importantly, after six MD replicas the average value of  $R^2$  appears to have converged within a reasonable limit.

The identification of diverse chemotypes is increasingly recognized as an important challenge in docking studies. For the evaluation of a chemotype bias, a chemotype-corrected AUC ROC with arithmetic weighting of individual chemotype group members (as introduced in the previous section) was employed. No significant differences between AUC ROC and chemotype-corrected AUC ROC values were observed for ensembles derived from crystal structure and for MD-derived structures (cf. **Fig. 32**). Importantly, this indicates no bias for specific chemotypes. It has to be noted that this does not imply that all chemotypes are recognized with equal probability. For example, even though the best performing ensemble in this study (consisting of four crystal structures) has an AUC ROC value of 0.819 and a chemotype-corrected AUC ROC value of 0.806 indicating no significant chemotype bias, this ensemble fails to correctly reproduce the native ligand pose of any member of chemotypes 7, 10, or 11 (see **Appendix 9.1**). Despite a significant improvement in docking different chemotypes by using ensemble docking, the limited number of members in the ensemble cannot guarantee a complete coverage of the chemical space of ligands.



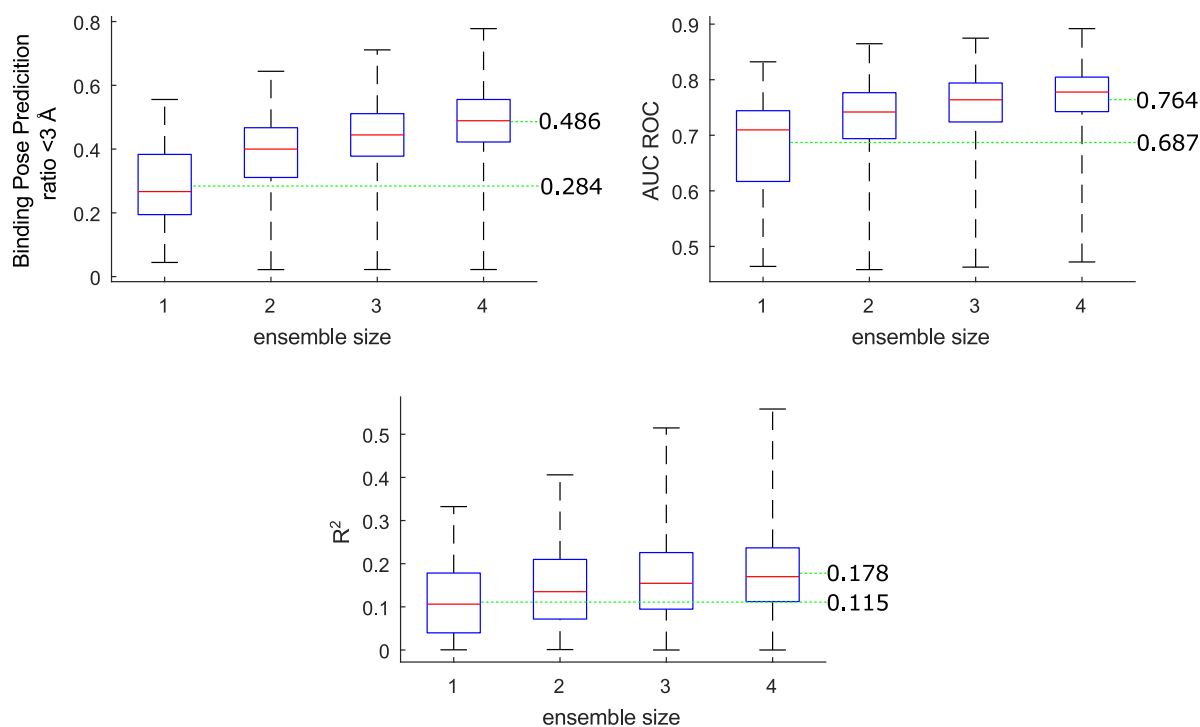
**Fig. 32:** Docking performance of ensembles constructed from crystal structures (top) and MD-derived structures (bottom) in terms of chemotype-corrected AUC ROC. The boxplots labeled „X-ray“ reflect the docking performance of all 45 crystal structures, and the black diamond, their mean value. The red diamonds reflect the performance of ensembles obtained by *k*-means clustering, and the blue diamonds, the performances of ensembles obtained by hierarchical clustering (with average linkage). In comparison to the AUC ROC values (cf. **Fig. 28** and **Fig. 30**) no significant differences between AUC ROC and chemotype-corrected AUC ROC values are observed.

### 5.2.5 Exhaustive Sampling of All Possible Ensembles

Finally, the docking performance of all possible ensembles that are composed of up to four crystal structures of renin was computed. This unravels the full potential of ensemble docking and sets the performance of ensembles constructed by clustering into perspective. Such an exhaustive sampling of all possible ensembles has previously been conducted by Korb *et al.* who investigated the docking performance of ensembles for a limited number of targets in terms of binding pose prediction and screening utility (AUC ROC).<sup>[83]</sup> In this work, besides binding pose prediction and screening utility, the scoring accuracy ( $R^2$ ) was evaluated as well. It is to be noted that such an exhaustive analysis of possible ensembles is limited by the obvious combinatorial explosion. A total number of 148,995 ensembles exists for ensembles composed of four crystal structures of renin (considering a total of 45 crystal structures of renin in this test system). For ensembles composed of five crystal structures of renin already > 1.2 million combinations are possible. With the available CPU power, investigating all possible ensembles of five crystal structures of renin was not feasible (bearing in mind that for each ensemble a receiver operating curve of more than 2,000 data points has to be calculated).

In **Fig. 33** the docking performance in terms of binding pose prediction, screening utility, and scoring accuracy of all possible ensembles, that are composed of up to four crystal structure of renin, are summarized as boxplots. In terms of binding pose prediction, a significant improvement can be achieved by employing ensemble docking. Docking the test ligands into individual crystal structures of renin yields on average a correctly docked pose in only 28.4%

of the cases. In contrast, docking into an ensemble consisting of four crystal structures of renin yields on average a correctly docked pose in 48.6% of the cases. This constitutes an astonishing improvement of more than 70% in binding pose prediction that is well worth the additional computational cost. Furthermore, it is apparent that the standard deviation decreases significantly by the use of ensemble docking. The improved binding pose prediction of ensembles can be seen to result in an improvement of screening utility that is, however, less pronounced. Docking the test ligands into individual crystal structures of renin results in an average AUC ROC of 0.687. In contrast, docking into an ensemble consisting of four crystal structures of renin results in an average AUC ROC of 0.764. This is equivalent to an improvement of 11%. In terms of scoring accuracy, a more significant effect can be observed. Docking the test ligands into individual crystal structures of renin results in an average  $R^2$  of 0.115. In contrast, docking into an ensemble consisting of four crystal structures of renin results in an average  $R^2$  of 0.178. This corresponds to an improvement of 55%. This data makes a convincing case for the use of ensemble docking. However, from the data presented here it may be extrapolated that the potential additional improvement gained by a further increase of ensemble size levels off and might not be worth the additional computational cost.



**Fig. 33:** Docking performance (in terms of binding pose prediction, screening utility and scoring accuracy) of all possible ensembles that are composed of up to four crystal structure of renin summarized as boxplots (cf. above paragraph). The mean values of the performance of individual crystal structures and ensembles consisting of four crystal structures are labeled by green dotted lines (red lines indicate median performance).

## 5.2.6 Conclusion

In this chapter, the potential of ensemble docking was analyzed. Ensemble docking is a simple yet promising method to consider protein flexibility in docking studies: Ligands are docked into multiple structures, and the results are subsequently merged. Three disciplines of docking, binding pose prediction, screening utility (AUC ROC & BEDROC) as well as scoring accuracy ( $R^2$ ) were evaluated. For this, a test system including 45 inhibitors of renin and 2240 decoys was analyzed. The docking performance of the corresponding 45 X-ray crystal structures as well as structures derived from a total of 500 ns of MD simulation time were analyzed.

The full potential of ensemble docking was unraveled by computing the docking performance of all possible ensembles that are composed of up to four crystal structures of renin (more than 160,000 combinations). The resulting data makes a convincing case for the use of ensemble docking. For ensembles consisting of four crystal structures an average improvement of >70% in binding pose prediction, 11% in screening utility (as measured by the AUC ROC metric), and 55% in scoring accuracy (as measured by  $R^2$ ) was observed in comparison to the performance of individual crystal structures.

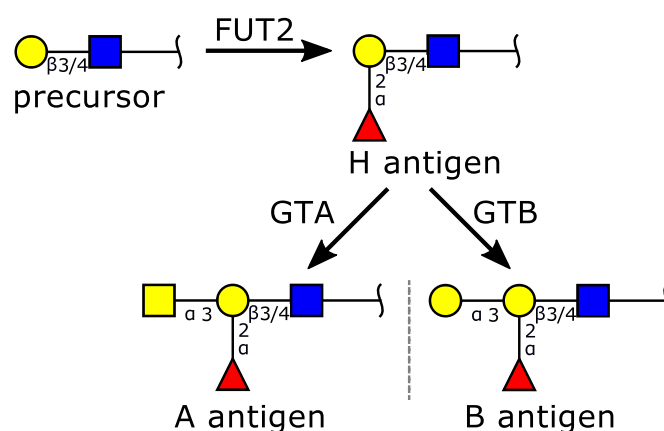
Furthermore, the impact of different approaches of ensemble definition was evaluated. For this, the performance of ensembles constructed by *k*-means and hierarchical clustering with average linkage of crystal structures as well as MD-derived structures was evaluated. Overall, *k*-means clustering yields ensembles that show improved performance compared to hierarchical clustering with average linkage. Unsurprisingly, ensembles constructed from MD-derived structures perform worse than the mean performance of individual crystal structures in terms of binding pose prediction and screening utility. However surprisingly, in this study ensembles of MD-derived structures constructed by *k*-means clustering perform on average better than 75% of any individual crystal structure in terms of scoring accuracy at all inspected ensembles sizes. Importantly, in contrast to previous studies on ensemble docking with MD-derived structures that picked their structures from unrepeated MD simulations, in this study values are reported as averages of ten repeated MD simulations and are shown to have converged within reasonable limits. This represents an important step towards reproducibility and meaningfulness. Finally, the ability of ensemble docking to recognize diverse chemotypes was evaluated. No bias for specific chemotypes is introduced by the use of ensemble docking as evaluated by the use of a chemotype-corrected AUC ROC metric.

### 5.3 Optimized Inhibitors for Human Blood Group B Galactosyltransferase

Parts of this chapter have been submitted for publication: C. Strecker, H. Peters, T. Hackl, T. Peters, B. Meyer, Fragment Growing to Design Optimized Inhibitors for Human Blood Group B Galactosyltransferase (GTB), *ChemMedChem* **2019**, submitted.

#### 5.3.1 Introduction

For human blood transfusion, the ABO system represents the most important out of more than 30 blood classification systems. It has four different phenotypes, A, B, AB, and O, that differ in the carbohydrate antigens that are dominantly presented on the cell wall of erythrocytes. The O phenotype only presents the H antigen that is formed by the action of FUT2 on a precursor antigen (see **Fig. 34**).<sup>[101]</sup> In the A and B phenotype, the H antigen is modified by addition of an *N*-acetylgalactosamine or a galactose that are each transferred by the action of GTA or GTB, respectively.<sup>[101]</sup> Both, GTA and GTB, are highly homologous and differ only in four amino acids yet they still exhibit distinct specificity. The mechanistic basis of their specificity has therefore been intensively studied.<sup>[102–104]</sup>



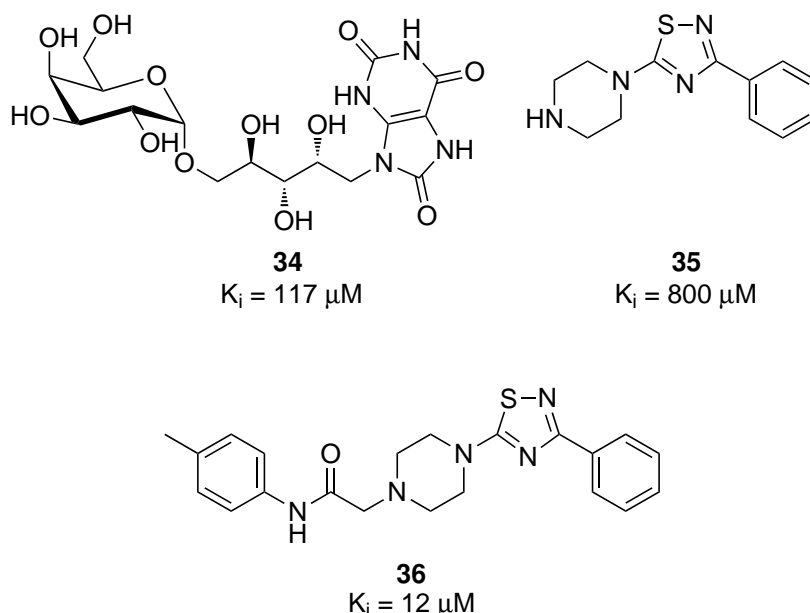
**Fig. 34:** Biosynthesis of blood group antigens. The H antigen (shown are type I ( $\beta$ 1-3) and type II ( $\beta$ 1-4)) is present in individuals of phenotype O. For individuals of the A or B phenotype an *N*-acetylgalactosamine or a galactose is added to the H antigen by the action of GTA or GTB, respectively.

Even though the ABO blood group system is known for over one hundred years and can be routinely serotyped, its biological function remains elusive.<sup>[101]</sup> However, a large body of work has linked blood group, in particular non-O phenotypes, to an increased incidence of various cancer types and their progression even though no underlying biological mechanism has been established for this yet.<sup>[105,106]</sup> Individuals of phenotype B have a 1.5-fold higher incidence rate for pancreatic cancer compared to individuals of phenotype O.<sup>[107]</sup> Additionally, for pancreatic cancer it has been shown that individuals of phenotype O have better survival



rates compared to individuals of non-O phenotype.<sup>[108]</sup> Similarly, for hepatocellular carcinoma it has been demonstrated that individuals of phenotype B have a median overall survival of only 34 months compared to 55 months for individuals of phenotype O.<sup>[109]</sup>

Based on the correlation of blood type with cancer incidence and progression it would be very desirable to be able to modulate the amount and the nature of blood group antigens in cancer patients. Therefore, design and synthesis of GTB or GTA inhibitors with potential drug like properties is very interesting. In the past, attempts to inhibit glycosyltransferases have frequently relied on substrate analogues.<sup>[42]</sup> One example of such a substrate analogue for the inhibition of GTB from our group is shown in **Fig. 35** (ligand **34**).<sup>[110]</sup> However, such substrate analogues have limited potential due to their non drug-like properties such as their high polarity.<sup>[42]</sup> For GTB, a fragment screening revealed fragment **35** (see **Fig. 35**), which is the starting point for the work described in this study.<sup>[111,112]</sup> Fragment **35** competes with acceptor binding with a  $K_i$  of 800  $\mu\text{M}$ . The fragment exhibits a >5 fold specificity for GTB compared to GTA.<sup>[112]</sup> Furthermore, X-ray crystallography revealed that the fragment binds to an open conformation of GTB and displaces  $\text{Mn}^{2+}$  from the active site.<sup>[112]</sup> In a previous PhD thesis from this lab, fragment **35** was advanced by a fragment growing approach to yield ligand **36** with a reported  $K_i$  of 12  $\mu\text{M}$  as evaluated by competitive STD NMR.<sup>[113]</sup>



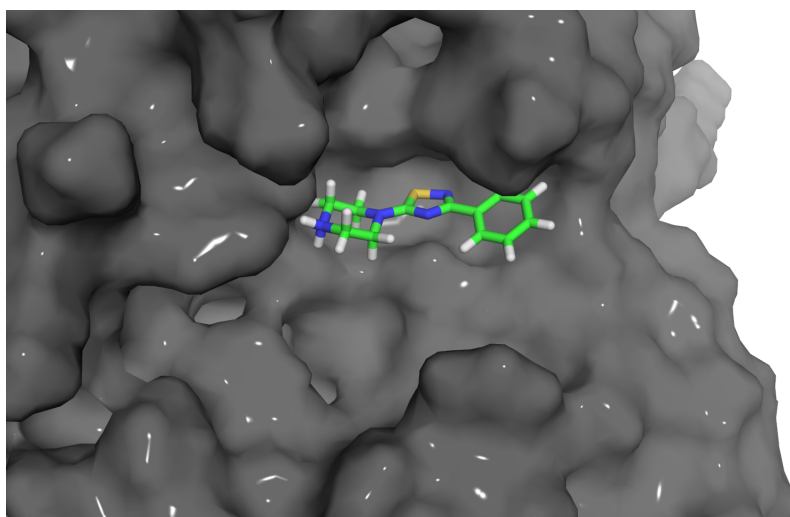
**Fig. 35:** Top left: Example of a donor substrate analogue that inhibits GTB. Top right: GTB inhibitor that binds to the acceptor and donor site derived from a fragment screen. Bottom: GTB inhibitor obtained by fragment growing approach.

### 5.3.2 Objective

The aim of this sub project was to optimize fragment **35** (see **Fig. 35**) into a more potent inhibitor of GTB. Because substitution of the piperazine moiety of the fragment has been investigated previously in our group by Leccese,<sup>[113]</sup> this work focused on the optimization of fragment **35** by replacement of its piperazine moiety and thiadiazole core, and growing of its phenyl moiety.

### 5.3.3 Design of Inhibitors of Human Blood Group B Galactosyltransferase by Fragment Growing

Basis to the structure-based ligand design process described in the next paragraphs is the 3U0X crystal structure of GTB that accommodates fragment **35** in its active binding site (see **Fig. 36**). This crystal structure displays an „open“ conformation of GTB.<sup>[112]</sup>

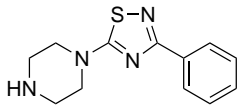
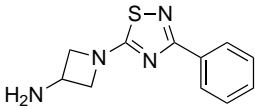
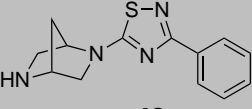
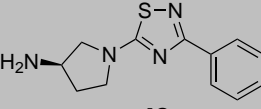
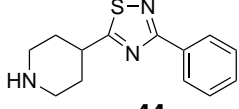
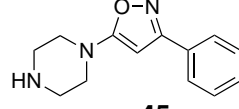
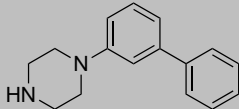
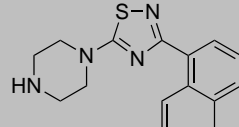
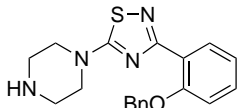
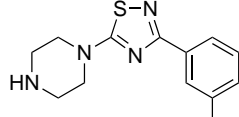


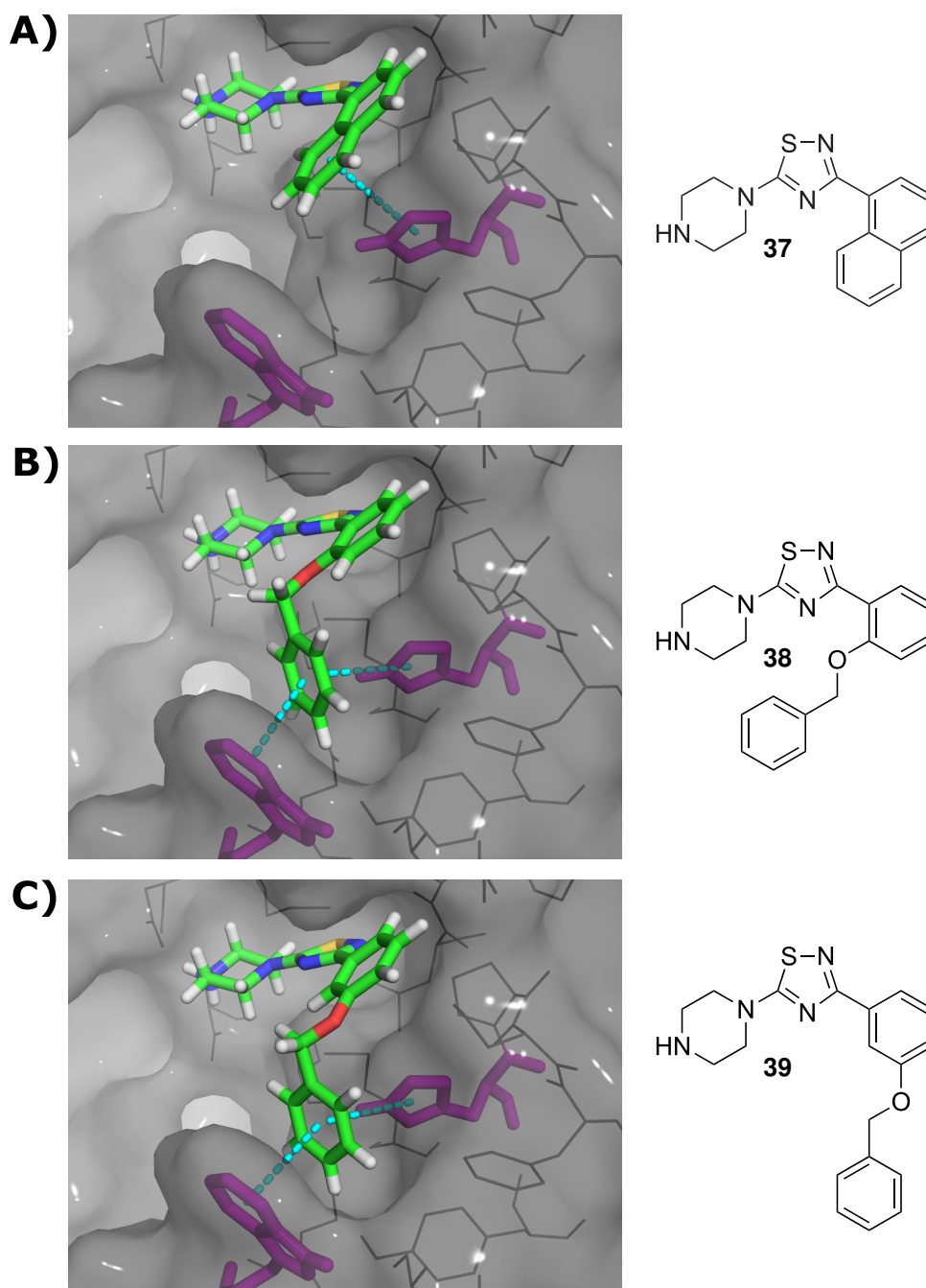
**Fig. 36:** Crystal structure (PDB: 3U0X) of GTB with fragment **35** in its active binding site.

Several modifications of fragment **35** were put to the test. For this, the intended ligands were docked into the 3U0X crystal structure and the protein-ligand complexes were subjected to MD simulations. Snapshots were extracted from this MD simulation at monotonic intervals and subjected to MM-GBSA calculations. Displacement of the 1,2,4-thiadiazole core (e.g. by a phenyl or isoxazole ring) resulted in no improvement of computed binding energies (see **Tab. 3**). For the piperazine moiety, it has been shown previously that displacement with a 1,4-diazepan moiety diminishes binding affinity.<sup>[111]</sup> Therefore, smaller options such as azetidin-3-amine were probed. However, they did not result in improved computed binding energies (see **Tab. 3**). Finally, it was realized that growing of the phenyl moiety of fragment **35** yields improved computed binding affinities. Substitution of the phenyl for a naphthyl moiety (resulting in ligand

**37**) yields an improvement in binding affinity (see **Tab. 3** and **Fig. 37** for a binding pose). The added aromatic ring allows an additional edge-to-face  $\pi$ - $\pi$  interaction to be formed with His233 as shown in **Fig. 37**. It was recognized that placing the additional aromatic moiety closer to Trp300 should allow for a simultaneous parallel-displaced  $\pi$ - $\pi$  interaction with Trp300 and an edge-to-face  $\pi$ - $\pi$  interaction with His233. This is realized by the 2- and 3-benzyloxy-substituted phenyl moieties in ligand **38** and **39** (see **Fig. 37** for the respective binding poses). Both show improved computed binding affinities but the 3-benzyloxy-substituted ligand **39** still significantly excels the 2-benzyloxy-substituted ligand **38** (see **Tab. 3**).

**Tab. 3:** Modifications of fragment **35** for that binding energies were calculated (MM-GBSA). Growing of the phenyl moiety of fragment **35** yields improved calculated binding energies.

| Ligand   | $\Delta G$<br>[kcal·mol <sup>-1</sup> ] | Ligand  | $\Delta G$<br>[kcal·mol <sup>-1</sup> ] |
|--|---|---|---|
| <br><b>40</b>   | -36.8<br>± 3.1                          | <br><b>41</b>   | -33.8<br>± 2.7                          |
| <br><b>42</b> | -36.0<br>± 2.8                          | <br><b>43</b> | -33.5<br>± 3.3                          |
| <br><b>44</b> | -36.6<br>± 3.4                          | <br><b>45</b> | -34.2<br>± 3.5                          |
| <br><b>46</b> | -35.1<br>± 3.4                          | <br><b>37</b> | -41.6<br>± 3.0                          |
| <br><b>38</b> | -46.8<br>± 4.9                          | <br><b>39</b> | -56.2<br>± 6.0                          |

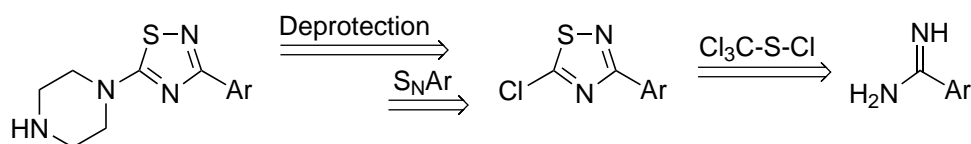


**Fig. 37:** Binding poses of ligands **37**, **38**, and **39** obtained by docking into the 3U0X crystal structure. For the orientation of the core fragment, all three ligands show an identical binding mode. Substitution of the phenyl moiety of fragment **35** for a naphthyl moiety (**37**, see panel A) allows an additional edge-to-face  $\pi$ - $\pi$  interaction (represented by a cyan dotted line) to be formed with His233 (colored magenta). It was recognized that placing the additional aromatic moiety closer to Trp300 (colored in magenta) should allow for a simultaneous parallel-displaced  $\pi$ - $\pi$  interaction with Trp300 and an edge-to-face  $\pi$ - $\pi$  interaction with His233. This is realized by the 2- and 3-benzyloxy-substituted phenyl moieties in ligand **38** and **39** (see panel B and C respectively).

### 5.3.4 Synthesis

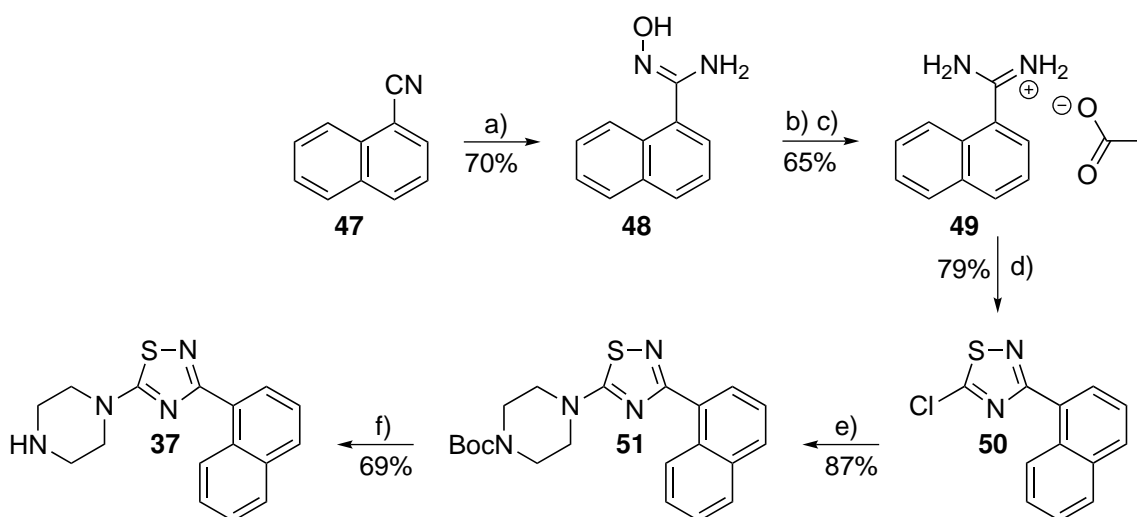
Based on the design process laid out in the previous section, it was decided to synthesize ligands **37**, **38**, and **39**. The general synthetic strategy was planned to involve the conversion of

amidines to 5-chloro-3-aryl-1,2,4-thiadiazoles and their subsequent nucleophilic substitution as key steps (see **Fig. 38**). The corresponding amidines are accessible from nitriles (via amidoximes or directly). Conversion of amidines to 5-chloro-3-aryl-1,2,4-thiadiazoles can be achieved by the action of perchloromethyl mercaptan. Nucleophilic substitution of the 5-chloro-1,2,4-thiadiazole moiety proceeds smoothly as the 5-position of the thiadiazole ring is strongly activated due to the electron withdrawing nature of the heteroatoms.



**Fig. 38:** The synthetic strategy for the synthesis of ligands **37**, **38**, and **39** was envisioned to involve the corresponding amidines as key intermediates. These are converted to 5-chloro-3-aryl-1,2,4-thiadiazoles by the action of perchloromethyl mercaptan. Subsequent nucleophilic substitution with protected piperazines and deprotection yields the ligands.

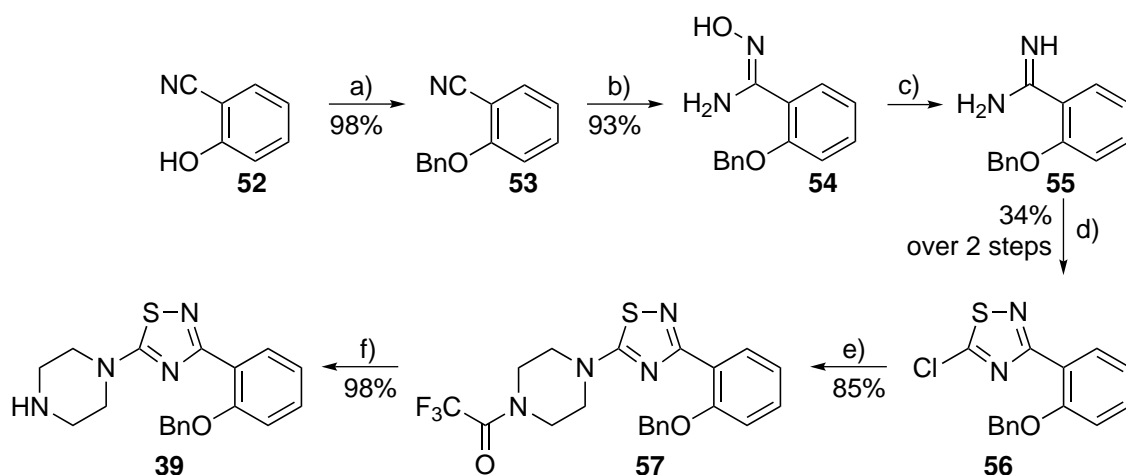
Synthesis of **37** was started from 1-cyanonaphthalene **47** and is summarized in **Fig. 39**. First, the nitrile **47** was converted to the amidoxime **48** by the action of hydroxylamine. Then, the amidoxime **48** was reduced to the amidine **49** by catalytic hydrogenation employing a protocol by Judkins *et al.*<sup>[114]</sup> The thiadiazole core was established by the action of perchloromethyl mercaptan. For the subsequent nucleophilic substitution 1-boc-piperazine was used. Finally, after acidic deprotection the ligand **37** was obtained in an overall yield of 22% over five steps.



**Fig. 39:** Synthesis of ligand **37**. Reaction conditions: a) 8.0 eq. NH<sub>2</sub>OH·HCl, 8.8 eq. Na<sub>2</sub>CO<sub>3</sub>, H<sub>2</sub>O/EtOH 2:1, 20 h, reflux; b) 1.5 eq. Ac<sub>2</sub>O, AcOH, 10 min, 20 °C; c) H<sub>2</sub> (1 atm), Pd/C, AcOH, 4 h, 20 °C; d) 1.0 eq. perchloromethyl mercaptan, 4.0 eq. NaOH in H<sub>2</sub>O, 2 h, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C → 20 °C; e) 1.0 eq. 1-boc-piperazine, 4.0 eq. NEt<sub>3</sub>, DMF, 2 h, 20 °C; f) 3 M HCl in EtOAc, 2 h, 20 °C.

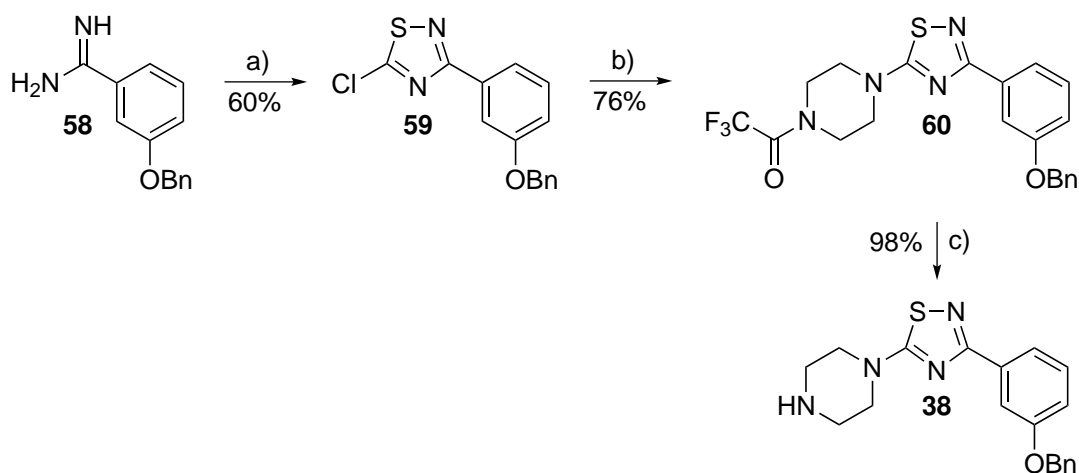
Synthesis of **39** was started from 2-cyanophenol **52** and is summarized in **Fig. 40**. After benzylation, the corresponding nitrile **53** was converted to the amidoxime **54** in high yield.

The previously used protocol from Judkins *et al.* for the catalytic reduction of amidoximes to amidines is obviously unsuited for the synthetic route towards ligand **39** due to the instability of the OBn-substituent towards hydrogenation. Even though non-hydrogenative methods for the direct conversion of nitriles to amidines exist,<sup>[115,116]</sup> they were not considered for reasons such as prolonged reaction times. Amidoximes constitute very convenient precursors for amidines, however the only published non-hydrogenative protocol for the reduction of amidoximes to amidines employs tin(II) chloride but is plagued by the use of elevated temperatures, prolonged reaction times, a large reagent excess, and the difficulties of tin salt removal.<sup>[117]</sup> Therefore, the reduction of **54** by samarium(II) iodide was explored: Thin layer chromatography indicated quantitative conversion of the educt within minutes and MS verified the formation of amidine **55**. However, separation of the amidine **55** from samarium salts was unsuccessful. Therefore, it was successfully proceeded with the crude product. Given the short reaction time at room temperature the  $\text{SmI}_2$ -promoted reduction of amidoximes is certainly a promising route to access amidines. Subsequent steps are analogous to the synthesis of ligand **37**. For the nucleophilic substitution step however, trifluoroacetyl piperazine was used and the following deprotection was hence performed by the action of base. Ligand **39** was obtained in an overall yield of 25% over six steps.



**Fig. 40:** Synthesis of ligand **39**. Reaction conditions: a) 1.2 eq.  $\text{K}_2\text{CO}_3$ , 2.0 eq. BnBr, 3 h, DMF, 50 °C; b) 8.0 eq.  $\text{NH}_2\text{OH}\cdot\text{HCl}$ , 8.8 eq.  $\text{Na}_2\text{CO}_3$ ,  $\text{H}_2\text{O}/\text{EtOH}$  2:1, 20 h, reflux; c) 2.4 eq.  $\text{SmI}_2$ , 30 min, THF/MeOH 1:1, 20 °C; d) 1.0 eq. perchloromethyl mercaptan, 4.0 eq. NaOH in  $\text{H}_2\text{O}$ , 2 h,  $\text{CH}_2\text{Cl}_2$ , 0 °C  $\rightarrow$  20 °C; e) 1.5 eq. trifluoroacetyl piperazine, 6.0 eq.  $\text{NEt}_3$ , DMF, 2 h, 20 °C; f) 5.0 eq.  $\text{Ba}(\text{OH})_2$ , MeOH, 1 h, 20 °C.

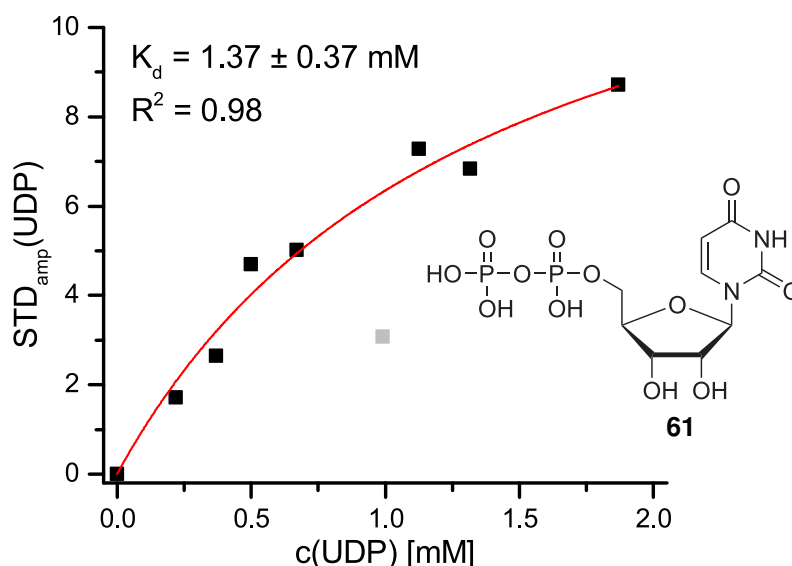
Synthesis of **38** was started from the commercially available amidine **58** and is summarized in **Fig. 41**. All steps are analogous to the synthesis of ligand **38**. Ligand **39** was obtained in an overall yield of 45% over three steps.



**Fig. 41:** Synthesis of ligand **38**. Reaction conditions: a) 1.0 eq. perchloromethyl mercaptan, 4.0 eq. NaOH in H<sub>2</sub>O, 2 h, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C → 20 °C; b) 1.5 eq. trifluoroacetyl piperazine, 6.0 eq. NEt<sub>3</sub>, DMF, 2 h, 20 °C; c) 5.0 eq. Ba(OH)<sub>2</sub>, MeOH, 1 h, 20 °C.

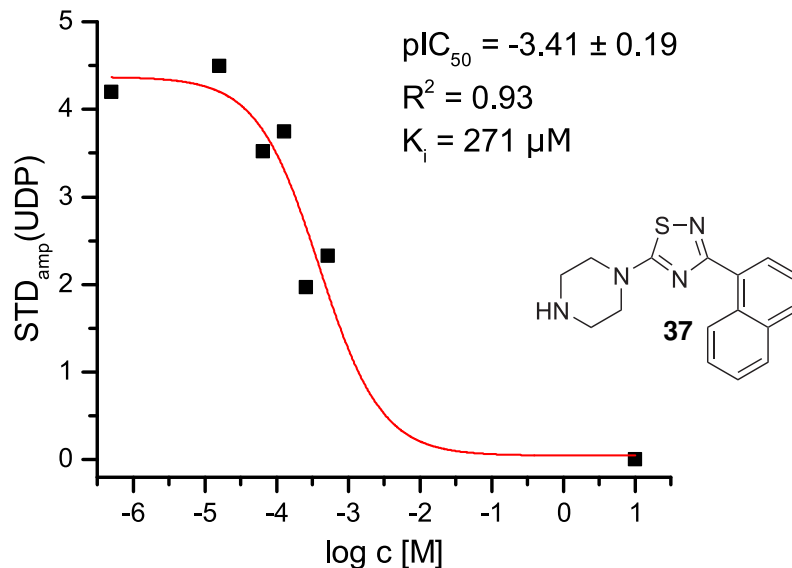
### 5.3.5 Competitive STD NMR

STD NMR has proven to be a very powerful tool for the determination of dissociation constants.<sup>[118]</sup> In contrast to enzymatic assays, STD NMR does not allow for the direct discrimination of competitive and non- or uncompetitive binding events. However, STD NMR can be performed in the presence of a known substrate to prove competitive binding. In this work, competitive STD NMR of ligands **37**, **38**, and **39** was performed by employing UDP as a competitor. UDP is known to bind the donor site of GTB with reported  $K_d$ 's ranging from 192-910  $\mu$ M.<sup>[102,119]</sup> In this work, measurements were performed at a pD of 5.8 to raise the solubility of the ligands. This is in minor deviation to the above reported  $K_d$ 's of UDP that were performed at GTB's optimum pH of 6.7. First, the  $K_d$  of the competitor UDP was determined by STD NMR. At the slightly more acidic pD of 5.8 evaluation of H-5/H-1' protons of UDP yielded a  $K_d$  of 1.37 mM (see **Fig. 42**).



**Fig. 42:** STD NMR of UDP binding to GTB. A GTB concentration of  $4.5 \mu\text{M}$  was used. The data points represent ligand excesses of 49 to 415. The STD effect of H-5/H-1' of UDP was evaluated. Fitting data points according to an one-site binding model resulted in a  $K_d$  of  $1.37 \text{ mM}$ .

In the next step, competitive STD NMR was performed for ligand **37**. A  $K_i$  of  $271 \mu\text{M}$  was determined (see **Fig. 43**). This represents a threefold better binding affinity compared to fragment **35**.

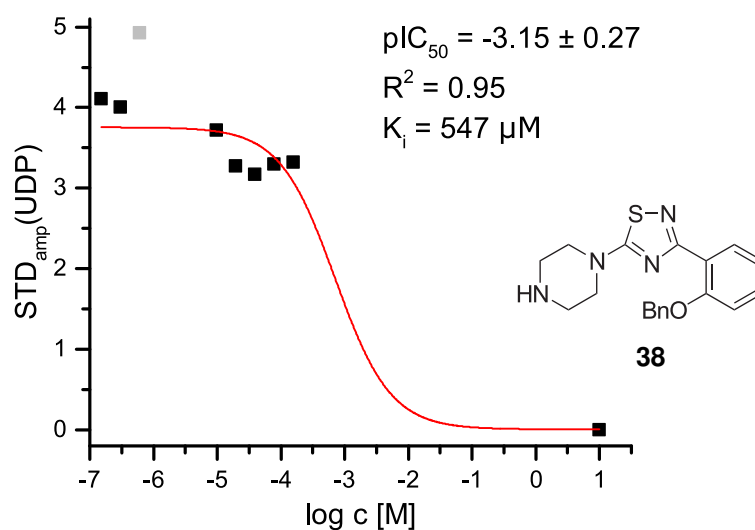


**Fig. 43:** Competitive STD NMR of ligand **37**. A GTB concentration of  $4.5 \mu\text{M}$  and a competitor (UDP) concentration of  $620 \mu\text{M}$  was used. The STD effect of H-5/H-1' of UDP was evaluated. Fitting data points according to an one-site competition model resulted in a  $pIC_{50}$  of  $-3.41$ . According to the Cheng-Prusoff equation, this corresponds to a  $K_i$  of  $271 \mu\text{M}$ .<sup>[120]</sup> An artificial zero point was added to the data at a concentration of  $10 \text{ M}$ . Shifting of the zero point to higher concentrations does not affect the result.

For ligand **38**, a  $K_i$  of  $574 \mu\text{M}$  was measured. However, caution is advisable for this value as only the initial part of the dose-response curve could be observed due to the limited solubility

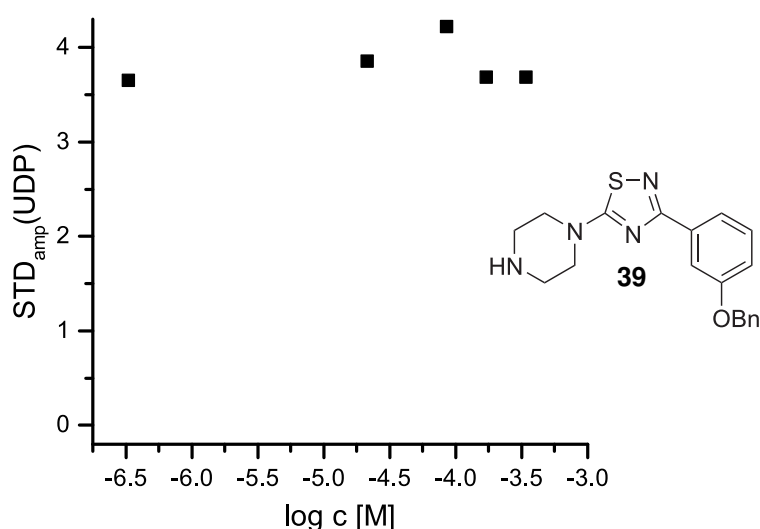


of the ligand (see **Fig. 44**).



**Fig. 44:** Competitive STD NMR of ligand **37**. A GTB concentration of  $4.5 \mu\text{M}$  and a competitor (UDP) concentration of  $404 \mu\text{M}$  was used. The STD effect of H-5/H-1' of UDP was evaluated. Fitting data points according to an one-site competition model resulted in a  $pIC_{50}$  of  $-3.15$ . According to the Cheng-Prusoff equation this corresponds to a  $K_i$  of  $547 \mu\text{M}$ .<sup>[120]</sup> An artificial zero point was added to the data at a concentration of  $10 \text{ M}$ . Shifting of the zero point to higher concentrations does not affect the result. It is advisable to use the  $K_i$  value cautiously as only the initial part of the dose-response curve could be observed due to the limited solubility of the ligand.

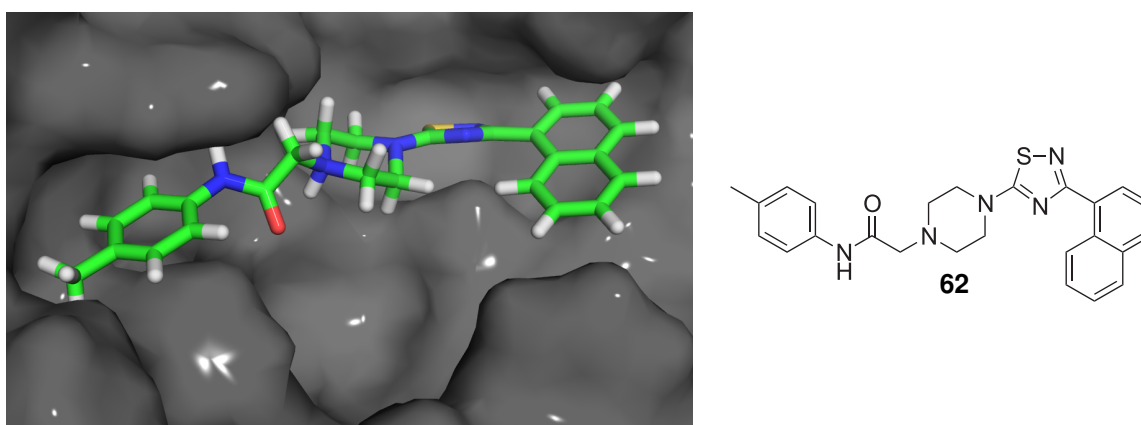
For ligand **39**, no competitive displacement of UDP could be observed under the given experimental conditions (see **Fig. 45**).



**Fig. 45:** Competitive STD NMR of ligand **39**. A GTB concentration of  $4.5 \mu\text{M}$  and a competitor (UDP) concentration of  $398 \mu\text{M}$  was used. The STD effect of H-5/H-1' of UDP was evaluated. No competitive displacement of UDP could be observed under the given experimental conditions.

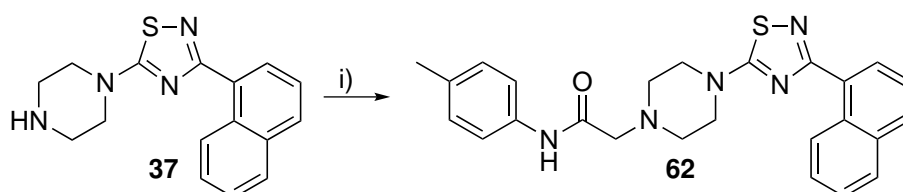
### 5.3.6 Optimization of Ligand 37

Previous work from this group has indicated that substitution of the piperazine moiety of fragment **35** with 2-bromo-*N*-(4-methylphenyl)acetamide yields the potent inhibitor **36** (see **Fig. 35**) that inhibits GTB with a  $K_i$  of  $12\ \mu\text{M}$  as evaluated by competitive STD NMR.<sup>[113]</sup> Therefore, it was obvious to consider substitution of fragment **37** with 2-bromo-*N*-(4-methylphenyl)acetamide as well. Docking studies indicated that the resulting ligand **62** is able to adopt an identical pose as fragment **35** within the GTB binding site (see **Fig. 46**) and MM-GBSA studies hinted a significantly improved binding affinity for **62** ( $\Delta G$  [kcal/mol] (calc.) =  $-57.9$ , see **Tab. 3** for a comparison).



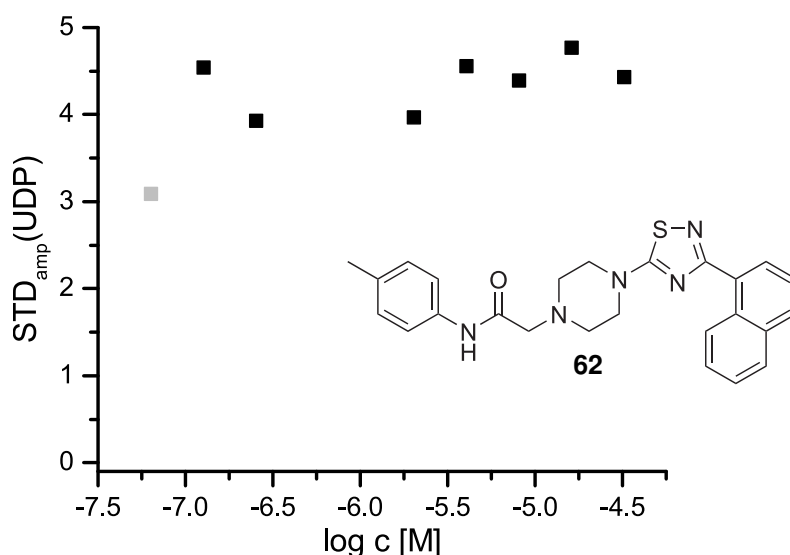
**Fig. 46:** Binding pose of **62**.

Ligand **62** was synthesized from **37** in a single step (see **Fig. 47**) with a yield of 56%.



**Fig. 47:** Synthesis of **62**. Reaction conditions: i) 2-bromo-*N*-(4-methylphenyl)acetamide,  $\text{CsCO}_3$ , MeCN.

However, no competitive displacement of UDP by ligand **62** could be observed by STD NMR under the given experimental conditions (see **Fig. 48**).

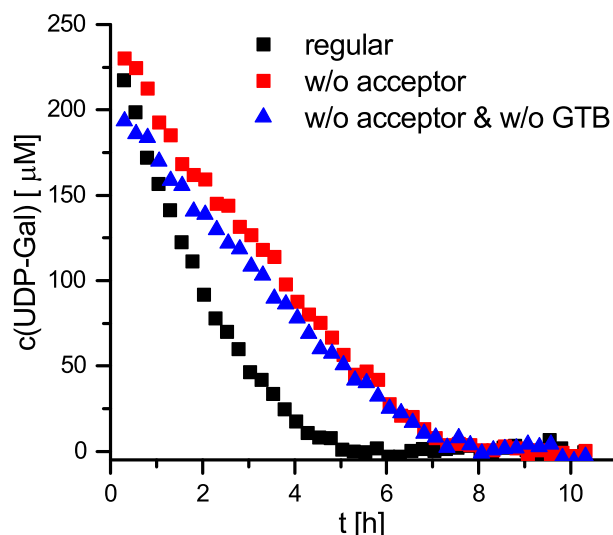


**Fig. 48:** Competitive STD NMR of ligand **62**. A GTB concentration of  $3\ \mu\text{M}$  and a competitor (UDP) concentration of  $348\ \mu\text{M}$  was used. The STD effect of H-5/H-1' of UDP was evaluated. No competitive displacement of UDP could be observed under the given experimental conditions. At a concentration of  $65\ \mu\text{M}$  of **62** a clouding of the sample could be observed indicating micelle formation. Therefore, the highest concentration point measured was  $32.5\ \mu\text{M}$ .

### 5.3.7 Activity Assay

Classically for the determination of  $K_m$  and  $v_{max}$  within the framework of the Michaelis-Menten theory, the initial reaction rate of an enzyme reaction has to be characterized at multiple substrate concentrations. This is time and resource intensive. However, as laid out by Schnell *et al.* and Goudar *et al.* it is possible to determine  $K_m$  and  $v_{max}$  in a single experiment from the time-dependent observation of the progress of an enzymatic reaction.<sup>[121,122]</sup> Our group has shown that NMR spectroscopy is a powerful tool for such progress curve analysis<sup>[123]</sup> and was able to determine inhibition constants for inhibitors of GTB from such experiments before.<sup>[110]</sup> For the enzymatic reaction of GTB, the transfer of galactose from the donor substrate UDP-Gal onto an acceptor substrate results in the formation of UDP. Because UDP is an inhibitor of GTB, it has to be removed from the reaction solution in order to achieve correct results. This has previously been achieved by the use of alkaline phosphatase by Schaefer *et al.*<sup>[110]</sup> In contrast to Schaefer *et al.*, in this work GTB was studied at a slightly more acidic pD value of 5.8 in order to raise the solubility of the examined ligands. Therefore, alkaline phosphatase was substituted with acidic phosphatase from wheat germ. Using  $\alpha\text{-Fuc-(1,2)-}\beta\text{-Gal-octyl}$  as an acceptor substrate the depletion of UDP-Gal could be monitored successfully via  $^1\text{H-NMR}$  spectroscopy in the presence of GTB (see **Fig. 49**). However, control experiments in the absence of acceptor substrate (and acceptor substrate plus GTB) revealed that to a significant extent UDP-Gal is problematically cleaved by acidic phosphatase (see **Fig. 49**). Therefore, the described attempts

to determine inhibition constants for ligand **37**, **38**, **39**, and **62** by means of progress curve analysis were unsuccessful. In conclusion, acidic phosphatase (as supplied by TCI) is unsuited for this type of enzymatic assay.



**Fig. 49:** Progress curve (monitored via  $^1\text{H-NMR}$  at 310 K showing the decay of UDP-Gal during the GTB catalyzed transfer of galactose onto the acceptor substrate  $\alpha\text{-Fuc-(1,2)-}\beta\text{-Gal-octyl}$  (black). However, a control in the absence of acceptor substrate (red) and a control in the absence of both acceptor substrate and GTB (blue) revealed that (to a significant extent) UDP-Gal is problematically cleaved by acidic phosphatase.

## 5.4 Conclusion

In this chapter, a fragment growing strategy was employed to discover inhibitors of GTB with improved affinity. This endeavor started from fragment **35** that originates from a previous fragment screening and inhibits GTB with a  $K_i$  of  $800\ \mu\text{M}$ . Enlarging the phenyl moiety of fragment **35** to a naphthyl moiety resulted in ligand **37** that showed a threefold improvement in binding affinity ( $K_i = 271\ \mu\text{M}$ ). Substitution of the phenyl moiety of fragment **35** with OBn-substituents in 2- and 3-position (as realized in ligands **38** and **39**) that were anticipated to exploit  $\pi\text{-}\pi$  interactions with His233 and Trp300 proved not to be promising. This indicates that the acceptor binding site of GTB might be narrower than suggested by the co-crystal structure of fragment **35** and GTB (PDB: 3U0X). An alternative explanation may be the entropic penalty for the additional three rotatable bonds. A continuative attempt to substitute the piperazine moiety of **37** led to ligand **62** that showed no measurable affinity for GTB under the given experimental conditions. Nonetheless, further attempts to substitute the piperazine moiety are certainly the most promising next step, both from a viewpoint of synthetic accessibility as well as available unoccupied space within the GTB binding site.

## 6 Experimental Procedures

### 6.1 Development of New Inhibitors of Fucosyltransferase 8

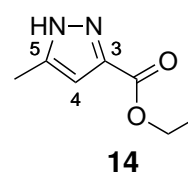
#### 6.1.1 Molecular Modeling

Schrödinger's Maestro was used for the ligand design process. Schrödinger includes the software modules „Prime“, „LigPrep“, „Glide“ and „Desmond“. The default force field is OPLS2005. On the basis of the only available X-ray crystal structure of human FUT8 (PDB: 2DE0) a model for donor substrate binding that has previously been developed in our working group was recreated.<sup>[49]</sup> The FUT8-GDP-Fucose complex was taken to „Desmond“ and fitted into an orthorhombic water box (SPC model) expanding 10 Å in each direction from the complex. Then, a 1.5 ns long MD simulation („Desmond v3“) was performed using the NVT ensemble at 310 K. The final frame of this MD simulation was used for subsequent docking campaigns. For this, the „FragNow“ subset of the ZINC12 databank was used.<sup>[50]</sup> This library contains fragments with a molecular weight of less than 250 Da, a clogP of less than 3.5, and less than 5 rotatable bonds. This library was prepared at pH 7 ± 0.2 using Maestro's „LigPrep“ (with default options except pH) totaling about 700,000 ligands. Two independent docking campaigns were performed: The first one using a grid box centered around guanine (expanding 15, 17 and 17 Å in X, Y and Z direction) and the second one centered on the β-phosphate of GDP-Fucose (expanding 17 Å in X, Y, and Z direction). Docking was performed using „Glide“ (using default options). Initially, the „HTVS“ scoring function was used. Top hits were then redocked using the „Standard Precision“ scoring function (GlideScore SP5.0). For MM-GBSA calculations, „Prime“ (v3.0, OPLS3 force field) was used.

#### 6.1.2 Synthesis

##### 6.1.2.1 Ethyl 5-methyl-1H-pyrazole-3-carboxylate **14**

Synthesis of ethyl 5-methyl-1H-pyrazole-3-carboxylate **14** has been described previously, e.g. by Skinner *et al.*<sup>[51]</sup> A sodium ethanoate solution was prepared by the addition of sodium (2.64 g, 114.8 mmol) to ethanol (80 mL). To this solution, a mixture of acetone (7.0 mL, 5.5 g, 95 mmol) and



diethyl oxalate (13.0 mL, 14.0 g, 96.0 mmol) was added dropwise causing the formation of a yellow precipitate. After complete addition, the reaction mixture was taken up in water and the pH value was adjusted to 3 with dilute sulfuric acid. The aqueous phase was extracted with

dichloromethane. The combined organic phases were dried over sodium sulfate, filtered and freed of the solvent *in vacuo*. The residue was dissolved in ethanol (80 mL) and hydrazine monohydrochloride (2.85 g, 41.7 mmol) was added. The reaction mixture was heated to reflux for 4 h and then freed of the solvent *in vacuo*. The residue was taken up in water and extracted with dichloromethane. The combined organic phases were dried over sodium sulfate, filtered and freed of the solvent *in vacuo*. The product was purified by column chromatography on silica gel (elution gradient: petrol ether/ethyl acetate 7:3 to 1:1).

Yield: 2.328 g (15.10 mmol, 16%) yellowish solid; C<sub>7</sub>H<sub>10</sub>N<sub>2</sub>O<sub>2</sub>; molecular weight: 154.17 g/mol; R<sub>f</sub> = 0.23 (petrol ether/ethyl acetate 1:1).

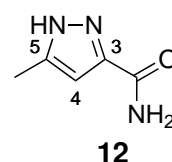
MS (ESI<sup>+</sup>): expt. 177.064 ([M+Na]<sup>+</sup>), calc. 177.063 ([M+Na]<sup>+</sup>).

<sup>1</sup>H-NMR (400 MHz, CDCl<sub>3</sub>): δ [ppm] = 11.72 (bs, 1H, NH), 6.57 (s, 1H, H-4), 4.36 (q, 2H, <sup>3</sup>J = 7.1 Hz, COOCH<sub>2</sub>CH<sub>3</sub>), 2.36 (s, 3H, CH<sub>3</sub>), 1.35 (t, 3H, <sup>3</sup>J = 7.2 Hz, COOCH<sub>2</sub>CH<sub>3</sub>).

<sup>13</sup>C-NMR (101 MHz, CDCl<sub>3</sub>): δ [ppm] = 162.2 (COOCH<sub>2</sub>CH<sub>3</sub>), 143.1, 141.9 (C-3, C-5), 107.3 (C-4), 60.9 (COOCH<sub>2</sub>CH<sub>3</sub>), 14.4 (COOCH<sub>2</sub>CH<sub>3</sub>), 11.5 (CH<sub>3</sub>).

#### 6.1.2.2 5-Methyl-1H-pyrazole-3-carboxamide 12

The synthesis was performed in accordance to a protocol of Jagdmann *et al.*<sup>[52]</sup> Ethyl 5-methyl-1H-pyrazole-3-carboxylate **14** (2.336 g, 15.15 mmol) was dissolved in tetrahydrofuran (40 mL). Formamide (3.62 mL, 4.09 g, 90.9 mmol) was added. Then, a 2.5 M solution of sodium methanoate (24.5 mL, 61.2 mmol) in methanol was added. The reaction mixture was heated to reflux for 5 h and after cooling neutralized with dilute hydrochloric acid and subsequently freed of the solvent *in vacuo*. The crude product was desalted by filtration over silica gel (eluent: dichloromethane/methanol 7:1). The residue was purified by column chromatography on silica gel (eluent: dichloromethane/methanol 7:1).



Yield: 841 mg (6.72 mmol, 44%) colourless solid; C<sub>5</sub>H<sub>7</sub>N<sub>3</sub>O; molecular weight: 125.13 g/mol; R<sub>f</sub> = 0.23 (dichloromethane/methanol 7:1).

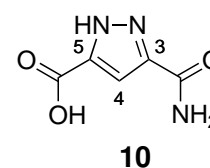
MS (ESI<sup>+</sup>): expt. 148.047 ([M+Na]<sup>+</sup>), calc. 148.049 ([M+Na]<sup>+</sup>).

<sup>1</sup>H-NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 12.83 (s, 1H, NH), 7.37 (s, 1H, CONH<sub>2</sub>), 7.10 (s, 1H, CONH<sub>2</sub>), 6.37 (s, 1H, H-4), 2.22 (s, 3H, CH<sub>3</sub>).

<sup>13</sup>C-NMR (126 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 163.8 (CONH<sub>2</sub>), 146.9 (C-3), 139.7 (C-5), 104.2 (C-4), 10.4 (CH<sub>3</sub>).

### 6.1.2.3 3-Carbamoyl-1*H*-pyrazole-5-carboxylic acid **10**

5-Methyl-1*H*-pyrazole-3-carboxamide **12** (100 mg, 800 μmol) was suspended in water (10 mL). Potassium permanganate (379 mg, 2.40 mmol) was added. The reaction mixture was stirred at 95 °C for 3 h and then filtrated while still hot. The filter cake was washed with hot water twice. The filtrate was treated with sodium sulfite until discoloration and then the pH value was adjusted to 1 with concentrated sulfuric acid. The solution was stored overnight at 4 °C during which product precipitated. The product was filtered off.



Yield: 89 mg (570 μmol, 72%) colourless solid; C<sub>5</sub>H<sub>5</sub>N<sub>3</sub>O<sub>3</sub>; molecular weight: 155.11 g/mol.

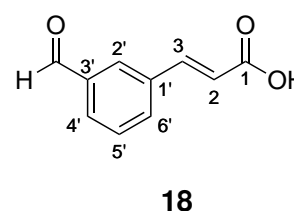
MS (ESI<sup>-</sup>): expt. 154.017 ([M-H]<sup>-</sup>), calc. 154.026 ([M-H]<sup>-</sup>).

<sup>1</sup>H-NMR (400 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 10.11 (bs, 2H, COOH, NH), 7.84 (s, 1H, CONH<sub>2</sub>), 7.46 (s, 1H, CONH<sub>2</sub>), 7.17 (s, 1H, H-4).

<sup>13</sup>C-NMR (125.7 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 161.6, 161.3 (CONH<sub>2</sub>, COOH), 142.9, 140.0 (C-3, C-5), 108.3 (C-4).

### 6.1.2.4 (*E*)-3-(3-Formylphenyl)acrylic acid **18**

Synthesis of (*E*)-3-(3-formylphenyl)acrylic acid **18** has been described previously, e.g. by Hansen *et al.*<sup>[53]</sup> Malonic acid (5.214 g, 50.11 mmol) was dissolved in pyridine (25 mL). Piperidine (330 μL, 0.284 g, 3.34 mmol) and isophthalaldehyde (4.480 g, 33.40 mmol) were added. The reaction mixture was



stirred at 100 °C for 3 h. The reaction mixture was freed of the solvent *in vacuo* and the

residue was then coevaporated with toluene. The crude product was purified by column chromatography on C<sub>18</sub>-reversed phase material (eluent: water/acetonitrile 90:10 + 0.5% NH<sub>3</sub> (aq.) (25%)). Product fractions were combined and acidified with formic acid. A precipitate was formed, filtered off and lyophilized from water.

Yield: 2.587 g (14.69 mmol, 44%) colourless solid; C<sub>10</sub>H<sub>8</sub>O<sub>3</sub>; molecular weight: 176.05 g/mol; R<sub>f</sub> = 0.28 (petrol ether/ethyl acetate 1:1 + 0.5% formic acid).

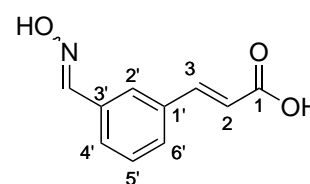
MS (ESI<sup>-</sup>): expt. 175.092 ([M-H]<sup>-</sup>), calc. 175.040 ([M-H]<sup>-</sup>).

<sup>1</sup>H-NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 12.55 (bs, 1H, COOH), 10.04 (s, 1H, C(=O)-H), 8.22 (m, 1H, H-2'), 8.04-8.02 (m, 1H, <sup>3</sup>J = 7.7 Hz, H-6'), 7.93-7.91 (m, 1H, <sup>3</sup>J = 7.7 Hz, H-4'), 7.68 (d, 1H, <sup>3</sup>J = 16.1 Hz, H-3), 7.65 (dd, 1H, <sup>3</sup>J = 7.7 Hz, <sup>3</sup>J = 7.7 Hz, H-5'), 6.66 (d, 1H, <sup>3</sup>J = 16.1 Hz, H-2).

<sup>13</sup>C-NMR (126 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 192.8 (C(=O)-H), 167.3 (COOH), 142.5 (C-3), 136.7 (C-3'), 135.2 (C-1'), 133.8 (C-6'), 130.2 (C-5'), 129.8 (C-4'), 129.5 (C-2'), 121.0 (C-2).

#### 6.1.2.5 (*E*)-3-(3-((Hydroxyimino)methyl)phenyl)acrylic acid **19**

(*E*)-3-(3-Formylphenyl)acrylic acid **18** (763 mg, 4.33 mmol) was dissolved in acetonitrile (40 mL). Triethylamine (900 μL, 657 mg, 6.50 mmol) and hydroxylamine hydrochloride (451 mg, 6.50 mmol) were added. The reaction mixture was stirred at room temperature for a total of 2 h. After 60 and 90 min respectively, additional portions of triethylamine and hydroxylamine hydrochloride were added. The reaction mixture was freed of the solvent *in vacuo* and purified by filtration over silica gel (eluent: petrol ether/ethyl acetate 1:1 + 0.5% formic acid).



**19**

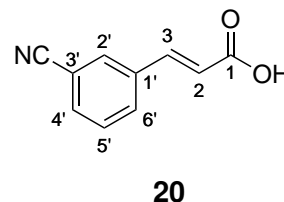
Yield: quantitatively (crude). Colourless solid; C<sub>10</sub>H<sub>9</sub>NO<sub>3</sub>; molecular weight: 191.18 g/mol; R<sub>f</sub> = 0.33 (petrol ether/ethyl acetate 1:1 + 0.5% formic acid).



MS (ESI<sup>-</sup>): expt. 190.104 ([M-H]<sup>-</sup>), calc. 190.050 ([M-H]<sup>-</sup>).

#### 6.1.2.6 (*E*)-3-(3-Cyanophenyl)acrylic acid **20**

Synthesis was performed in accordance to a protocol of De Luca *et al.*<sup>[54]</sup> Cyanuric chloride (1.210 g, 6.567 mmol) was dissolved in *N,N*-dimethylformamide (15 mL) and the mixture was stirred at room temperature for 15 min during which a precipitate was formed. Crude (*E*)-3-(3-((hydroxyimino)methyl)phenyl)acrylic acid **19** (627 mg, 3.28 mmol) was dissolved in *N,N*-dimethylformamide and added to the above mixture. The reaction mixture was stirred at room temperature for 5 h. Then water (2.0 mL) was added and the reaction mixture was freed of the solvent *in vacuo*. The residue was purified by filtration over silica gel (eluent: petrol ether/ethyl acetate 1:1 + 0.5% formic acid).

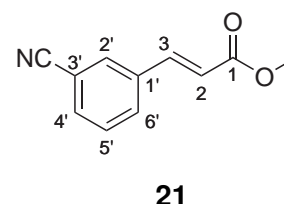


Yield: quantitatively (crude). Colourless solid; C<sub>10</sub>H<sub>7</sub>NO<sub>2</sub>; molecular weight: 173.17 g/mol; R<sub>f</sub> = 0.28 (petrol ether/ethyl acetate 1:1 + 0.5% formic acid).

MS (ESI<sup>-</sup>): expt. 172.089 ([M-H]<sup>-</sup>), calc. 172.040 ([M-H]<sup>-</sup>).

#### 6.1.2.7 (*E*)-Methyl 3-(3-cyanophenyl)acrylate **21**

Crude (*E*)-3-(3-cyanophenyl)acrylic acid **20** (996 mg, 5.75 mmol) was dissolved in *N,N*-dimethylformamide (15 mL). Cesium carbonate (2.248 g, 6.902 mmol) and subsequently methyl iodide (430 μL, 980 mg, 6.90 mmol) were added. The reaction mixture was stirred at room temperature for 18 h and then freed of the solvent *in vacuo*. The residue was purified by column chromatography on silica gel (eluent: petrol ether/ethyl acetate 4:1).



Yield: 522 mg (2.79 mmol, 48%) colourless solid; C<sub>11</sub>H<sub>9</sub>NO<sub>2</sub>; molecular weight: 187.19 g/mol; R<sub>f</sub> = 0.28 (petrol ether/ethyl acetate 4:1).

MS (ESI<sup>+</sup>): expt. 188.070 ([M+H]<sup>+</sup>), calc. 188.071 ([M+H]<sup>+</sup>).

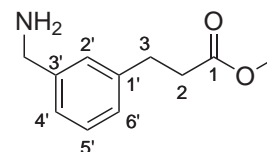
<sup>1</sup>H-NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 8.27-8.26 (m, 1H, **H-2'**), 8.09-8.06 (m, 1H, <sup>3</sup>J = 8.1 Hz, **H-6'**), 7.89-7.86 (m, 1H, <sup>3</sup>J = 8.1 Hz, **H-4'**), 7.69 (d, 1H, <sup>3</sup>J = 16.1 Hz, **H-3**), 7.63

(dd, 1H,  $^3J = 7.9$  Hz,  $^3J = 7.9$  Hz, **H-5'**), 6.82 (d, 1H,  $^3J = 16.1$  Hz, **H-2**), 3.74 (s, 3H, COOCH<sub>3</sub>).

<sup>13</sup>C-NMR (125.7 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 166.3 (COOCH<sub>3</sub>), 142.3 (**C-3**), 135.3 (**C-1'**), 133.5 (**C-4'**), 132.8 (**C-6'**), 131.8 (**C-2'**), 130.1 (**C-5'**), 120.3 (**C-2**), 118.4 (**CN**), 112.1 (**C-3'**), 51.6 (COOCH<sub>3</sub>).

#### 6.1.2.8 Methyl 3-(3-(aminomethyl)phenyl)propanoate **11**

(*E*)-Methyl 3-(3-cyanophenyl)acrylate **21** (100 mg, 0.534 mmol) was dissolved in 1.5% hydrochloric acid in methanol (10 mL) and then Pd/C (51 mg) was added. The reaction mixture was stirred at room temperature for 3 h over an atmosphere of H<sub>2</sub> (30 bar).



**11**

The reaction mixture was filtered over Celite and freed of the solvent *in vacuo*. The residue was taken up in water and the pH was adjusted to 8 with sodium bicarbonate. The aqueous phase was extracted with dichloromethane. The combined organic phases were dried over sodium sulfate, filtered and freed of the solvent *in vacuo*.

Yield: 99 mg (0.51 mmol, 96%) slightly yellowish oil; C<sub>11</sub>H<sub>15</sub>NO<sub>2</sub>; molecular weight: 193.24 g/mol; R<sub>f</sub> = 0.30 (dichloromethane/methanol 9:1 + 0.5% ammonium hydroxide (25wt%)).

MS (ESI<sup>+</sup>): expt. 194.1134 ([M+H]<sup>+</sup>), calc. 194.1176 ([M+H]<sup>+</sup>).

<sup>1</sup>H-NMR (500 MHz, MeOD-*d*<sub>4</sub>): δ [ppm] = 7.24 (dd, 1H,  $^3J = 7.6$  Hz,  $^3J = 7.6$  Hz, **H-5'**), 7.20-7.18 (m, 1H, **H-2'**), 7.18-7.15 (m, 1H,  $^3J = 7.6$  Hz, **H-4'**), 7.11-7.08 (m, 1H,  $^3J = 7.6$  Hz, **H-6'**), 3.76 (s, 2H, CH<sub>2</sub>NH<sub>2</sub>), 3.64 (s, 3H, COOCH<sub>3</sub>), 2.92 (t, 2H,  $^3J = 7.6$  Hz, **H-3**), 2.64 (t, 2H,  $^3J = 7.6$  Hz, **H-2**).

<sup>13</sup>C-NMR (126 MHz, MeOD-*d*<sub>4</sub>): δ [ppm] = 175.1 (COOCH<sub>3</sub>), 143.6 (**C-3'**), 142.3 (**C-1'**), 129.7 (**C-5'**), 128.5 (**C-2'**), 128.0 (**C-6'**), 126.4 (**C-4'**), 52.0 (COOCH<sub>3</sub>), 46.6 (CH<sub>2</sub>NH<sub>2</sub>), 36.6 (**C-2**), 31.9 (**C-3**).

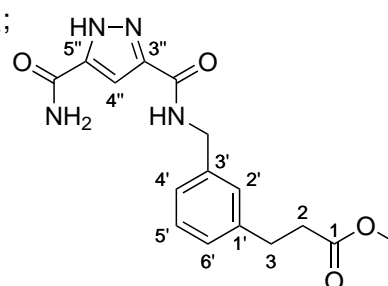
#### 6.1.2.9 Methyl 3-(3-((5-carbamoyl-1*H*-pyrazole-3-carboxamido)methyl)phenyl)propanoate **22**

3-Carbamoyl-1*H*-pyrazole-5-carboxylic acid **10** (69 mg, 447 μmol) and methyl 3-(3-(aminomethyl)phenyl)propanoate **11** (95 mg, 492 μmol) were dissolved in *N,N*-dimethylformamide (2 mL).

Then, *N,N*-diisopropylethylamine (152  $\mu$ L, 116 mg, 894  $\mu$ mol) was added and subsequently a solution of propylphosphonic anhydride (284 mg, 894  $\mu$ mol) in *N,N*-dimethylformamide (50wt%, 522  $\mu$ L). The reaction mixture was stirred at room temperature for 24 h and then freed of the solvent *in vacuo*. The residue was taken up in 5% hydrochloric acid and extracted with ethyl acetate. The combined organic phases were dried over sodium sulfate, filtered and freed of the solvent *in vacuo*. The residue was purified via RP-HPLC (column: Nucleodur C<sub>18</sub> Isis; solvent A: 95% H<sub>2</sub>O + 5% MeCN; solvent B: 95% MeCN + 5% H<sub>2</sub>O; gradient: 0-5 min, 35% B; 5-13 min, 100% B; 13-15 min, 100% B; 15-18 min, 100% B; 18-20 min, 35% B; flow rate: 20 mL/min; R<sub>t</sub> = 5.2 min).

Yield: 51.5 mg (156  $\mu$ mol, 32%) colourless solid; C<sub>16</sub>H<sub>18</sub>N<sub>4</sub>O<sub>4</sub>; molecular weight: 330.34 g/mol; R<sub>f</sub> = 0.37 (dichloromethane/methanol 9:1).

MS (ESI<sup>+</sup>): expt. 331.136 ([M+H]<sup>+</sup>), calc. 331.140 ([M+H]<sup>+</sup>).



**22**

<sup>1</sup>H-NMR (500 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  [ppm] = 13.92 (bs, 1H, NH (pyrazole)), 8.88 (bs, 1H, R<sub>1</sub>CONHCH<sub>2</sub>R<sub>2</sub>), 7.86 (bs, 1H, CONH<sub>2</sub>), 7.44 (bs, 1H, CONH<sub>2</sub>), 7.25 (s, 1H, H-4''), 7.23 (dd, 1H, <sup>3</sup>J = 7.7 Hz, <sup>3</sup>J = 7.7 Hz, H-5'), 7.16-7.14 (m, 1H, H-2'), 7.14-7.11 (m, 1H, <sup>3</sup>J = 7.7 Hz, H-4'/H-6'), 7.10-7.07 (m, 1H, <sup>3</sup>J = 7.5 Hz, H-4'/H-6'), 4.40 (d, 2H, <sup>3</sup>J = 5.9 Hz, R<sub>1</sub>CONHCH<sub>2</sub>R<sub>2</sub>), 3.56 (s, 3H, COOCH<sub>3</sub>), 2.83 (t, 2H, <sup>3</sup>J = 7.6 Hz, H-3), 2.60 (t, 2H, <sup>3</sup>J = 7.7 Hz, H-2).

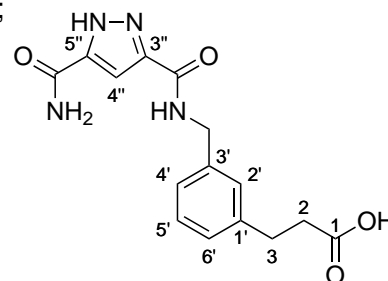
<sup>13</sup>C-NMR (126 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  [ppm] = 172.6 (COOCH<sub>3</sub>), 140.5, 139.6 (C-1', C-3'), 128.3 (C-5'), 127.1 (C-2'), 126.6, 125.1 (C-4', C-6'), 105.7 (C-4''), 51.3 (COOCH<sub>3</sub>), 41.9 (R<sub>1</sub>CONHCH<sub>2</sub>R<sub>2</sub>), 34.8 (C-2), 30.2 (C-3).

#### 6.1.2.10 3-(3-((5-Carbamoyl-1*H*-pyrazole-3-carboxamido)methyl)phenyl)propanoic acid **8**

Methyl 3-(3-((5-carbamoyl-1*H*-pyrazole-3-carboxamido)methyl)phenyl)propanoate **22** (41 mg, 120  $\mu$ mol) was dissolved in methanol (8 mL). Then, potassium hydroxide (21 mg, 370  $\mu$ mol) was dissolved in water (2 mL) and added to the above solution. The reaction mixture was stirred at room temperature for 24 h and then neutralized with dilute hydrochloric acid and freed of the solvent *in vacuo*. The residue was desalted by filtration over silica gel (eluent: dichloromethane/methanol 9:1 + 0.5% formic acid). The crude product was purified via RP-

HPLC (column: Nucleodur C<sub>18</sub> Isis; solvent A: H<sub>2</sub>O + 0.01% NH<sub>3</sub>; solvent B: MeCN; gradient: 0-10 min, 10% B; 10-20 min, 40% B; 20-22 min, 90% B; 22-25 min, 90% B; 25-27 min, 10% B; 27-30 min, 10% B; flow rate: 1 mL/min; R<sub>t</sub> = 5.0-7.0 min).

Yield: 22.7 mg (71.9 μmol, 58%) colourless solid; C<sub>15</sub>H<sub>16</sub>N<sub>4</sub>O<sub>4</sub>; molecular weight: 316.31 g/mol; R<sub>f</sub> = 0.17 (dichloromethane/methanol 9:1 + 0.5% formic acid).



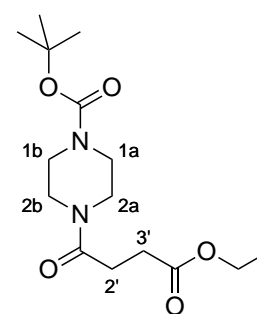
**8**

<sup>1</sup>H-NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 8.95 (bs, 1H, R<sub>1</sub>CONHCH<sub>2</sub>R<sub>2</sub>), 7.88 (bs, 1H, CONH<sub>2</sub>), 7.42 (bs, 1H, CONH<sub>2</sub>), 7.23 (s, 1H, H-4''), 7.21 (dd, 1H, <sup>3</sup>J = 7.6 Hz, <sup>3</sup>J = 7.6 Hz, H-5'), 7.17-7.15 (m, 1H, H-2'), 7.12-7.07 (m, 2H, H-4', H-6'), 4.40 (d, 2H, <sup>3</sup>J = 6.1 Hz, R<sub>1</sub>CONHCH<sub>2</sub>R<sub>2</sub>), 2.78 (t, 2H, <sup>3</sup>J = 7.6 Hz, H-3), 2.43 (t, 2H, <sup>3</sup>J = 7.6 Hz, H-2).

<sup>13</sup>C-NMR (126 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 174.5 (COOH), 161.3 (CONH<sub>2</sub>), 160.1 (R<sub>1</sub>CONHCH<sub>2</sub>R<sub>2</sub>), 141.5 (C-1'), 139.4 (C-3'), 128.2 (C-5'), 127.4 (C-2'), 126.6, 124.8 (C-4', C-6'), 105.8 (C-4''), 42.0 (R<sub>1</sub>CONHCH<sub>2</sub>R<sub>2</sub>), 36.2 (C-2), 30.8 (C-3).

#### 6.1.2.11 *tert*-Butyl 4-(4-ethoxy-4-oxobutanoyl)piperazine-1-carboxylate **25**

1-Boc-piperazine **24** (2.00 g, 10.7 mmol) was suspended in acetonitrile (20 mL) and *N,N*-diisopropylethylamine (2.19 mL, 1.67 g, 12.9 mmol) was added. The mixture was cooled down to 0 °C. To this mixture, a solution of ethyl succinyl chloride (1.75 mL, 2.03 g, 12.4 mmol) in acetonitrile (20 mL) was added dropwise. The reaction mixture was stirred for 2 h at room temperature and subsequently freed of the solvent *in vacuo*. The product was purified by column chromatography on silica gel (elution gradient: dichloromethane (100%) to dichloromethane/methanol 24:1).



**25**

Yield: 3.13 g (9.97 mmol, 93%) slightly yellowish oil; C<sub>15</sub>H<sub>26</sub>N<sub>2</sub>O<sub>5</sub>; molecular weight: 314.37 g/mol; R<sub>f</sub> = 0.18 (dichloromethane).

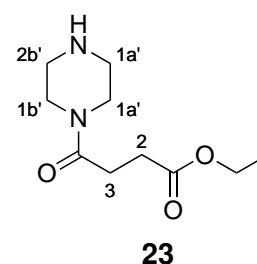
MS (ESI<sup>+</sup>): expt. 337.173 ([M+Na]<sup>+</sup>), calc. 337.173 ([M+H]<sup>+</sup>); expt. 259.129 ([M-C<sub>4</sub>H<sub>9</sub><sup>+</sup>+2H]<sup>+</sup>), calc. 259.129 ([M-C<sub>4</sub>H<sub>9</sub><sup>+</sup>+2H]<sup>+</sup>); expt. 213.087 ([M-C<sub>4</sub>H<sub>9</sub><sup>+</sup>-EtOH+H]<sup>+</sup>), calc. 213.087 ([M-C<sub>4</sub>H<sub>9</sub><sup>+</sup>-EtOH+H]<sup>+</sup>).

<sup>1</sup>H-NMR (400 MHz, CDCl<sub>3</sub>): δ [ppm] = 4.14 (q, 2H, <sup>3</sup>J = 7.1 Hz, COOCH<sub>2</sub>CH<sub>3</sub>), 3.61-3.54 (m, 2H, **H-2'/H-3'**), 3.49-3.42 (m, 4H, **H-2a-b**), 3.42-3.36 (m, 2H, **H-2'/H-3'**), 2.68-2.59 (m, 4H, **H-1a-b**), 1.46 (s, 9H, NCOOC(CH<sub>3</sub>)<sub>3</sub>), 1.25 (t, 3H, <sup>3</sup>J = 7.1 Hz, COOCH<sub>2</sub>CH<sub>3</sub>).

<sup>13</sup>C-NMR (101 MHz, CDCl<sub>3</sub>): δ [ppm] = 173.2, 170.1 (R<sub>1</sub>NC(=O)CH<sub>2</sub>CH<sub>2</sub>COOEt), 154.7 (NCOOC(CH<sub>3</sub>)<sub>3</sub>), 80.4 (NCOOC(CH<sub>3</sub>)<sub>3</sub>), 60.7 (COOCH<sub>2</sub>CH<sub>3</sub>), 45.3 (**C-2a-b**), 41.7 (**C-2'/C-3'**), 29.4 (**C-1a-b**), 28.5 (NCOOC(CH<sub>3</sub>)<sub>3</sub>), 28.0 (**C-1a-b**), 14.3 (COOCH<sub>2</sub>CH<sub>3</sub>).

#### 6.1.2.12 Ethyl 4-oxo-4-(piperazin-1-yl)butanoate **23**

A constant stream of dry HCl gas was produced by dropwise addition of sulfuric acid to solid sodium chloride. The HCl stream was passed into a solution of *tert*-butyl 4-(4-ethoxy-4-oxobutanoyl)piperazine-1-carboxylate **25** (1.21 g, 3.84 mmol) in ethanol (20 mL). The reaction mixture was stirred for 2 h at room temperature. The reaction mixture was neutralized by addition of a saturated sodium bicarbonate solution.



The aqueous solution was extracted with dichloromethane. The combined organic phases were dried over sodium sulfate, filtered and freed of the solvent *in vacuo*. Even though the NMR spectrum reveals impurities, no additional purification was pursued.

Yield: 730 mg (3.41 mmol, 89%) slightly yellowish oil; C<sub>10</sub>H<sub>18</sub>N<sub>2</sub>O<sub>3</sub>; molecular weight: 214.26 g/mol; R<sub>f</sub> = 0.26 (dichloromethane/methanol 9:1 + 0.5% NH<sub>3</sub>).

MS (ESI<sup>+</sup>): expt. 215.1390 ([M+H]<sup>+</sup>), calc. 215.1390 ([M+H]<sup>+</sup>); expt. 169.0973 ([M-EtOH]<sup>+</sup>), calc. 169.0972 ([M-EtOH]<sup>+</sup>).

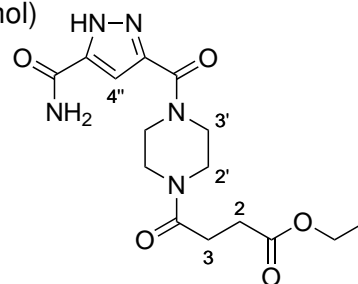
<sup>1</sup>H-NMR (600 MHz, MeOD-*d*<sub>4</sub>): δ [ppm] = 4.12 (q, 2H, <sup>3</sup>J = 7.1 Hz, COOCH<sub>2</sub>CH<sub>3</sub>), 3.55-3.52 (m, 4H, **H-1a'/H-1b'**), 2.86-2.82 (m, 2H, **H-2/H-3**), 2.79-2.76 (m, 2H, **H-2/H-3**), 2.69-2.58 (m, 4H, **H-2a'/H-2b'**), 1.25 (t, 3H, <sup>3</sup>J = 7.1 Hz, COOCH<sub>2</sub>CH<sub>3</sub>).

$^{13}\text{C}$ -NMR (151 MHz,  $\text{MeOD-}d_4$ ):  $\delta$  [ppm] = 174.7, 172.3 ( $\text{R}_1\text{NC(=O)CH}_2\text{CH}_2\text{COOEt}$ ), 61.6 ( $\text{COOCH}_2\text{CH}_3$ ), 47.2 (**C-1a'/H-1b'**), 46.5 (**C-2/C-3**), 46.2 (**C-2/C-3**), 43.5 (**C-1a'/H-1b'**), 30.2, 28.7 (**C-2a'/H-2b'**), 14.5 ( $\text{COOCH}_2\text{CH}_3$ ).

### 6.1.2.13 Ethyl 4-(4-(5-carbamoyl-1H-pyrazole-3-carbonyl)piperazin-1-yl)-4-oxobutanoate **26**

3-Carbamoyl-1H-pyrazole-5-carboxylic acid **10** (10.0 mg, 64.5  $\mu\text{mol}$ ) and ethyl 4-oxo-4-(piperazin-1-yl)butanoate **23** (20.7 mg, 96.7  $\mu\text{mol}$ ) were dissolved in *N,N*-dimethylformamide (1.5 mL).

Then *N,N*-diisopropylethylamine (32.9  $\mu\text{L}$ , 25.0 mg, 193  $\mu\text{mol}$ ) was added and subsequently a solution of propylphosphonic anhydride (61.5 mg, 193  $\mu\text{mol}$ ) in *N,N*-dimethylformamide (50wt%, 113  $\mu\text{L}$ ). The reaction mixture was stirred at room



**26**

temperature for 60 min. Then, 5% hydrochloric acid (20 mL) was added. The aqueous phase was extracted with dichloromethane. The combined organic phases were dried over sodium sulfate, filtered and freed of the solvent *in vacuo*. The residue was purified via RP-HPLC (column: Nucleodur  $\text{C}_{18}$  Isis; solvent A: 100%  $\text{H}_2\text{O}$  + 0.1% formic acid; solvent B: 100% MeCN + 0.1% formic acid; gradient: 0-5 min, 10% B; 5-15 min, 50% B; 15-16 min, 90% B; 16-18 min, 90% B; 18-19 min, 10% B; 19-23 min, 10% B; flow rate: 1.5 mL/min;  $R_t$  = 10.8 min).

Yield: 5.8 mg (16.5  $\mu\text{mol}$ , 26%) colourless solid;  $\text{C}_{15}\text{H}_{21}\text{N}_5\text{O}_5$ ; molecular weight: 351.36 g/mol.

MS (ESI<sup>+</sup>): expt. 352.167 ([M+H]<sup>+</sup>), calc. 352.162 ([M+H]<sup>+</sup>).

$^1\text{H}$ -NMR (400 MHz,  $\text{DMSO-}d_6$ ):  $\delta$  [ppm] = 13.94 (bs, 1H, NH (pyrazole)), 7.95, 7.52 (s, 1H, CONH<sub>2</sub>), 7.16 (s, 1H, H-4''), 4.04 (q, 2H,  $^3J = 7.2$  Hz,  $\text{COOCH}_2\text{CH}_3$ ), 4.00-3.44 (m, 8H, H-2', H-3'), 2.65-2.56 (m, 2H, H-3), 1.17 (t, 3H,  $^3J = 7.1$  Hz,  $\text{COOCH}_2\text{CH}_3$ ). H-2 signal is overlapping with the solvent signal.

### 6.1.2.14 Ethyl 4-(4-(5-carbamoyl-1H-pyrazole-3-carbonyl)piperazin-1-yl)-4-oxobutanoate **9**

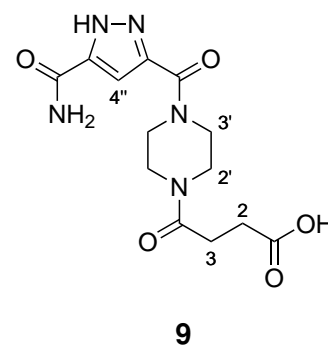
4-(4-(5-Carbamoyl-1H-pyrazole-3-carbonyl)piperazin-1-yl)-4-oxobutanoate **26** (5.8 mg, 16.5  $\mu\text{mol}$ ) and potassium hydroxide (3.7 mg, 66  $\mu\text{mol}$ ) were dissolved in a 1:1 mixture of methanol and water (2.0 mL). The reaction mixture was stirred at room temperature for 24 h and lyophilized after neutralization. The residue was purified via RP-HPLC (column: Phenomenex Luna NH<sub>2</sub>; solvent A: 100%  $\text{H}_2\text{O}$  + 0.1% formic acid; solvent B: 100% MeCN + 0.1% formic acid; gradient:

0-12 min, 2% B; 12-15 min, 20% B; 15-16 min, 90% B; 16-18 min, 90% B; 18-19 min, 2% B; 19-23 min, 2% B; flow rate: 1 mL/min;  $R_t \sim 7.5$  min).

Yield: 2.1 mg (6.5  $\mu$ mol, 39%) colourless solid;  $C_{13}H_{17}N_5O_5$ ; molecular weight: 323.30 g/mol.

MS (ESI<sup>-</sup>): expt. 322.110 ([M-H]<sup>-</sup>), calc. 322.116 ([M-H]<sup>-</sup>).

<sup>1</sup>H-NMR (400 MHz, D<sub>2</sub>O):  $\delta$  [ppm] = 7.15 (s, 1H, **H-4''**), 3.96-3.66 (m, 8H, **H-2'**, **H-3'**), 2.88-2.76 (m, 2H, **H-3**), 2.74-2.66 (m, 2H, **H-2**).



### 6.1.3 Inhibition Assays

The activity assay and STD NMR measurements were performed on a Bruker Avance 700 MHz spectrometer in 3 mm NMR tubes. Samples were prepared in D<sub>2</sub>O containing MES-*d*<sub>13</sub> (50 mM) and TMSP-*d*<sub>4</sub> (1 mM) at pD 7.0. The FUT8 solution was rebuffed to the buffer specified above by using Amicon Ultra-4 cellulose filter (molecular weight cut-off 5 kDa) and the protein concentration was determined by using a nanodrop at 280 nm.

#### 6.1.3.1 Activity assay

The fucosylation of 1- $\beta$ -*N*-acetylchitotriose was monitored via <sup>1</sup>H-NMR over a course of 3.5 h. Every 15 min <sup>1</sup>H-NMR spectra with excitation sculpting were acquired at 310 K using a pseudo 2D pulse program. The first data point was acquired after 11.5 min. The sample contained a FUT8 concentration of 2  $\mu$ M, a GDP-Fucose concentration of 2.6 mM and a 1- $\beta$ -*N*-acetylchitotriose concentration of 2.3 mM. Additionally, the sample contained 10 U of alkaline phosphatase (EC 3.1.3.1) and 1 mg/mL bovine serum albumin. Spectra were acquired with 32,768 data points and a total of 64 scans. FIDs were multiplied with an exponential function (line broadening 0.2) before Fourier transformation. The concentration of assay components was determined from H-8 of GDP-Fuc and GDP, H-1 of (fucosylated) 1- $\beta$ -*N*-acetylchitotriose and the H-6 methyl group of fucose.

#### 6.1.3.2 STD NMR

The standard pulse program „stddiffesgp2d“ was used. On resonance irradiation was applied at 0 ppm. Saturation was achieved by a cascade of 40 Gaussian pulses with a duration of 50 ms.

A spinlock pulse of 15 ms length was used to reduce the protein background. Experiments were performed at 300 K. All samples contained a FUT8 concentration of 5  $\mu$ M and ligand excesses of ligand **8** ranging from 41 to 199. Spectra were acquired with 24,576 data points and a total of 512 scans. FIDs were multiplied with an exponential function (line broadening 2) before Fourier transformation. For the determination of dissociation constants STD amplification factors were plotted against the ligand concentration. The data points were then fitted using an one-site binding model in Origin2016G.



## 6.2 Ensemble Docking of Renin

All work was performed using Schrödinger Maestro 2017-1 version, PyMOL 1.5.0.4 and Matlab R2016b. Schrödinger includes the software modules „Protein Preparation Wizard“, „Prime“, „Epik“, „LigPrep“, „Glide“, and „Desmond“. The default force field is OPLS3.

### 6.2.1 Preparation of Crystal Structures, Active Ligands and Decoys

A set of 45 crystal structures of renin in complex with an inhibitor was downloaded from the Protein Data Bank (PDB). Their PDB codes are: 1BIL, 1HRN, 1RNE, 2BKT, 2FS4, 2G1O, 2G1R, 2G1Y, 2IKO, 2IKU, 2IL2, 2V0Z, 2V10, 2V11, 2V12, 2V13, 2V16, 3G6Z, 3G70, 3G72, 3GW5, 3KM4, 3OOT, 3OQF, 3OQK, 3Q3T, 3Q4B, 3Q5H, 3SFC, 3VSX, 3VUC, 3VYD, 3VYE, 3VYF, 4GJ6, 4GJ7, 4GJD, 4Q1N, 4RYG, 4RZ1, 5KOQ, 5KOS, 5SXN, 5SY3, and 5SZ9. The crystal structures were preprocessed using Maestro's „Protein Preparation Wizard“. Bond orders were assigned using the chemical component dictionary, hydrogens were added, missing side chains were added using „Prime“, and heteroatom states were generated using „Epik“ at pH  $7.4 \pm 0.1$ . All water molecules, ions, and additives used for crystallization (e.g. triethylene glycol or dimethyl sulfoxide) were deleted. For crystal structures that contained more than one renin chain only one chain was kept. Orientation of hydroxyl groups, asparagines, glutamines, and histidines was optimized using PROPKA at pH 7.4. Finally, the structures were subjected to a restrained minimization. The sequence numbering was manually changed to comply with the numbering given in the Uniprot database. Ligands were extracted from the minimized crystal structures and prepared by employing Maestro's „LigPrep“. Ionization of the ligands was not changed and no tautomers were generated. For each ligand 50 decoys were generated with the DUD-E system.<sup>[90]</sup> 10 decoys appeared multiple times and were deleted so that in total 2240 decoys are in the test system. Decoys were prepared employing „LigPrep“ with the same settings as for the ligands.

### 6.2.2 Construction of RMSD matrices and Multidimensional Scaling

Structures were exported to PyMol. All hydrogen atoms were removed. The residues that form the binding site correspond to residues that are within 4 Å of the primary ligand in any of the 45 minimized crystal structures. In MD studies it was found that the side chains of four residues do not contribute to ligand binding but dominate structural heterogeneity. For that reason Glu189, Phe319 and Asp320 were completely omitted from the binding epitope and for Arg148 only the side chain was removed. This results in the definition of the binding site

as: Thr84-Tyr87, Val102, Asp104, Gly106-Asn109, Trp111-Pro113, Ala123, His127, Leu147, Arg148 (backbone only), Tyr149-Gly152, Val154, Val177, Met180, Pro184, Phe185, Leu187-Ala188, Phe190-Val193, Gln201-Ile203, Tyr228, Leu290, Asp292-Ser299, His367, Met369, Asp370-Pro372, Thr375, Thr378, Asp380, Gly382, Ala383. RMSD matrices were constructed by iterating PyMOL's „align“ command with the cycles option set to zero employing a basic Python script (see **Appendix 9.2**). For clustering of MD-derived structures, frames were extracted at a stepsize of 50 ps (yielding 1000 frames per MD). Multidimensional scaling was performed using Matlab's „mdscale“ command using „Sammon“ as criterion.

### 6.2.3 Clustering

*Hierarchical clustering:* Matlab's „linkage“ command was applied to the respective RMSD matrix with the method set to „average“. Subsequently, Matlab's „cluster“ command was applied to the resulting matrix. The number of clusters was controlled by using the „maxclust“ option. Cluster representatives were selected as follows: A new RMSD matrix of all cluster members was constructed. The sum of RMSD values for each cluster member to all other cluster members was calculated. The structure with the smallest sum was determined to be the cluster representative.

*K-means clustering:* To the respective RMSD matrix multidimensional scaling with three dimensions was applied. The resulting data points were clustered employing Matlab's „kmeans“ command with the number of replicates and the number of maximum iterations set to 10,000.

### 6.2.4 Docking

Docking was performed using „Glide“. For the grid, the center of the docking box was defined as the centroid of the residues Asp104 and Asp292. A cubic docking box with an axis length of 40 Å was used (35 Å for inner box). The docking mode „Standard Precision“ was employed. To allow „soft docking“ van der Waals radii of atoms with a partial atomic charge of less than 0.15 were scaled by a factor of 0.8.<sup>[74]</sup> Only one pose per ligand was generated.

### 6.2.5 MD Simulations

The ligand of the 2G1R crystal structure (that was prepared as described above) was deleted. The structure was solvated with the „System Builder“ of „Desmond“. An orthorhombic solvent box extending 10 Å beyond the protein in all directions was used. The system was neutralized by adding 7 Na<sup>+</sup> ions and solvated with water (SPC model). Additional sodium chloride was added in a concentration of 0.15 M to represent the physiological concentration. In total the

system contained 9936 water molecules, 35 Na<sup>+</sup> ions and 28 Cl<sup>-</sup> ions in a volume of 381697 Å<sup>3</sup>. 50 ns long MD simulations were performed in the NPT ensemble at 310 K and a pressure of 1.01325 bar with recording intervals of 0.1 ps for energy and 0.5 ps for the trajectory. Long range electrostatic interactions were calculated using a smooth particle-mesh Ewald approximation with a cut-off radius of 9 Å. The system was coupled to a Noose-Hoover chain thermostat with a relaxation time of 1 ps and to an isotropic Martyna-Tobias-Klein barostat with a relaxation time of 2 ps. The RESPA integrator (reference system propagation algorithm) time step was set to 2 fs for bonded and short-range interactions within the 9.0 Å cutoff and 6 fs for long-range non-bonded interactions. The system was relaxed using „Desmond's“ default relaxation protocol. The MD simulation was repeated 10 times from the same input file, each time with a random seed in order to randomize initial velocities.

### 6.2.6 Evaluation of Docking Performance

For evaluation of ensembles the docking results of the respective cluster representatives were mixed, sorted after their docking score and multiple entries of the same ligand were reduced so that only the entry with the best docking score remained („merge & shrink“) by employing a python script. For evaluation of binding poses, the binding site of the native crystal structure and the docked structure were aligned using the previously defined binding epitope. Then, the RMSD between the native ligand and docked pose was calculated using heavy atoms only. The docked pose was evaluated to be correct if the RMSD is less than 3 Å. For evaluation of screening utility, the AUC ROC and BEDROC score (with  $\alpha = 20$ ) were calculated.<sup>[91]</sup> For evaluation of scoring accuracy, experimental pIC<sub>50</sub> values of active ligands were plotted against the docking score. A linear regression was performed and the coefficient of determination was utilized as the characteristic. For evaluation of enrichment of complexes of chemically closely related ligands (here defined as chemotypes), in accordance with Clark *et al.* a chemotype-corrected AUC ROC with arithmetic weighting of individual chemotype group members was calculated.<sup>[92]</sup> Ligands were assigned to 11 different chemotype groups on the basis of common chemical substructures.

## 6.3 Optimized Inhibitors for Human Blood Group B Galactosyltransferase

### 6.3.1 Molecular Modeling

All work was performed using Schrödinger Maestro 2017-1 version. Schrödinger includes the software modules „Protein Preparation Wizard“, „Prime“, „LigPrep“, „Glide“, and „Desmond“. The default force field is OPLS3. Unless otherwise stated default options of these modules were employed. The 3U0X crystal structure was preprocessed using the „Protein Preparation Wizard“. Only the A chain of the dimer was used. Bond orders were assigned, hydrogens added and heteroatom states were generated using „Epik“ at pH  $5.0 \pm 0.2$ . All water molecules, ions, and additives used for crystallization were deleted. Orientation of hydroxyl groups, asparagines, glutamines, and histidines was optimized using PROPKA at pH 5.0. Finally, the structure was subjected to a restrained minimization. For docking, the ligands were prepared by employing „LigPrep“ at pH  $5.0 \pm 0.2$ . The grid was created with „Glide“. The center of the docking box was defined as the centroid of the co-crystallized ligand. A cubic docking box with an axis length of 30 Å was used (25 Å for inner box). For docking in „Glide“, the „Standard Precision“ mode was used. For MD simulations, the docked ligand/GTB complex was solvated with the „System Builder“ of „Desmond“. An orthorhombic solvent box extending 10 Å beyond the protein in all directions was used. The system was neutralized by adding Cl<sup>-</sup> ions and solvated with water (SPC model). Additional sodium chloride was added to reach a concentration of 0.15 M. MD simulations were performed for 10 ns in the NPT ensemble at 310 K and a pressure of 1.01325 bar with recording intervals of 0.1 ps for energy and 0.5 ps for the trajectory. A random seed for initial velocities was used. From these MD simulations every 200<sup>th</sup> frame was extracted (totaling 100 structures), stripped from water and the energy of ligand binding was calculated using the MM-GBSA function of „Prime“. In **Tab. 3** averages of these calculations are reported.

### 6.3.2 Synthesis

#### 6.3.2.1 General Procedure A: Synthesis of Amidoximes from Nitriles

The nitrile (1.0 eq.) was suspended in a solution of ethanol and water (1:2). A portion of hydroxylamine hydrochloride (2.0 eq.) and sodium carbonate (2.2 eq.) was added. The reaction mixture was heated to reflux for the specified time span and during this, additional portions of hydroxylamine hydrochloride and sodium carbonate were added. After cooling, the reaction mixture was extracted with dichloromethane. The combined organic phases were dried over sodium sulfate, filtered and freed of the solvent *in vacuo*. The residue was purified as specified.

### 6.3.2.2 General Procedure B: Synthesis of 5-chloro-3-aryl-1,2,4-thiadiazoles from Amidines

The amidine (1.0 eq.) was suspended in dichloromethane and cooled to 0 °C. Then, perchloromethyl mercaptan (1.0 eq.) was added. After 2 min, an aqueous solution of sodium hydroxide (4.0 eq.) was added. The reaction mixture was stirred at 0 °C for 1 h and then 1 h at room temperature. Water was added to the reaction mixture and then extracted with dichloromethane. The combined organic phases were dried over sodium sulfate, filtered and freed of the solvent *in vacuo*. The residue was purified by column chromatography on silica gel.

### 6.3.2.3 General Procedure C: Nucleophilic Substitution of 5-chloro-3-aryl-1,2,4-thiadiazoles with Piperazine Derivatives

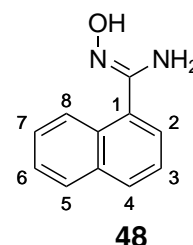
The 5-chloro-3-aryl-1,2,4-thiadiazole (1.0-1.5 eq.) was dissolved in *N,N*-dimethylformamide. Then, triethylamine (4.0-6.0 eq.) and the mono-protected piperazine derivative (1.0 eq.) were added. The reaction mixture was stirred at room temperature for 2 h and then freed of the solvent *in vacuo*. The residue was purified by column chromatography on silica gel.

### 6.3.2.4 General Procedure D: Deprotection of Trifluoroacetamide-protected Piperazines

The trifluoroacetamide-protected piperazine (1.0 eq.) and barium hydroxide (5.0 eq.) were dissolved in methanol and stirred at room temperature for one hour. The reaction mixture was freed of the solvent *in vacuo* and the residue was purified by column chromatography on silica gel.

### 6.3.2.5 *N*-Hydroxy-1-naphthimidamide 48

Synthesis of *N*-hydroxy-1-naphthimidamide **48** has been described previously, e.g. by Kivrak *et al.*<sup>[124]</sup> 1-Cyanonaphthalene **47** (2.00 g, 13.1 mmol) was treated according to general procedure A. The reaction time was 20 h. In total, 4 portions of hydroxylamine hydrochloride and sodium carbonate were added. The crude product was recrystallized from a mixture of hexane and chloroform (1:1).



Yield: 1.70 g (9.11 mmol, 70%) colourless solid; C<sub>11</sub>H<sub>10</sub>N<sub>2</sub>O; molecular weight: 186.21 g/mol.

MS (ESI<sup>+</sup>): expt. 187.088 ([M+H]<sup>+</sup>), calc. 187.087 ([M+H]<sup>+</sup>).

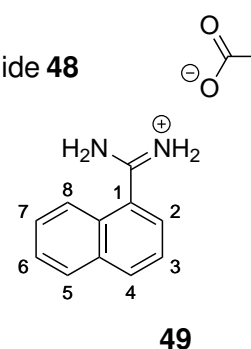
<sup>1</sup>H-NMR (400 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 9.56 (s, 1H, =N-OH), 8.35-8.29 (m, 1H, H<sub>aryl</sub>), 7.97-7.92 (m, 2H, H<sub>aryl</sub>), 7.57-7.50 (m, 4H, H<sub>aryl</sub>), 5.96 (bs, 2H, -NH<sub>2</sub>).

<sup>13</sup>C-NMR (100.6 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 151.6 (C(=NOH)-NH<sub>2</sub>), 133.2, 132.2, 131.1 (C-1/C-4a/C-8a), 128.8, 128.0, 126.5, 126.1, 125.9, 125.2 (C<sub>aryl</sub>).

### 6.3.2.6 1-Naphthamidinium acetate **49**

Synthesis of 1-naphthamidinium acetate **49** has been described previously, e.g. by Alonso-Alija *et al.*<sup>[125]</sup> *N*-hydroxy-1-naphthimidamide **48**

(1.656 g, 8.893 mmol) was dissolved in acetic acid (15 mL). Then, acetic anhydride (1.26 mL, 1.36 g, 13.3 mmol) was added and the reaction mixture was stirred for 10 min at room temperature. After this, Pd/C (250 mg) was added and the reaction mixture was stirred over an atmosphere of hydrogen for 4 h. The solution was filtered through



Celite and the filtrate freed of the solvent *in vacuo*. The crude product was recrystallized from a mixture of hexane/chloroform/methanol 1:10:1.

Yield: 1.335 g (5.881 mmol, 65%) beige solid; C<sub>13</sub>H<sub>14</sub>N<sub>2</sub>O<sub>2</sub>; molecular weight: 230.26 g/mol.

MS (ESI<sup>+</sup>): expt. 171.092 ([M+H]<sup>+</sup>), calc. 171.092 ([M+H]<sup>+</sup>).

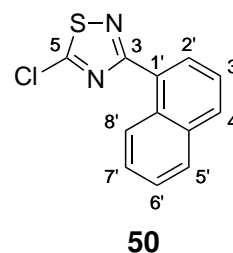
<sup>1</sup>H-NMR (400 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 10.65 (bs, 4H, C(=NH<sub>2</sub><sup>+</sup>)-NH<sub>2</sub>), 8.19-8.15 (m, 1H, H<sub>aryl</sub>), 8.10-8.06 (m, 1H, H<sub>aryl</sub>), 8.04-8.00 (m, 1H, H<sub>aryl</sub>), 7.74-7.62 (m, 4H, H<sub>aryl</sub>), 1.73 (s, 3H, acetate).

<sup>13</sup>C-NMR (100.6 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 176.8 (COO<sup>-</sup> (acetate)), 167.1 (C(=NH<sub>2</sub><sup>+</sup>)-NH<sub>2</sub>), 132.9 (C-1/C-4a/C-8a), 131.1 (C<sub>aryl</sub>), 129.5, 129.0 (C-1/C-4a/C-8a), 128.5 (C<sub>aryl</sub>), 127.7 (C<sub>aryl</sub>), 126.8 (C<sub>aryl</sub>), 126.2 (C<sub>aryl</sub>), 125.1 (C<sub>aryl</sub>), 124.2 (C<sub>aryl</sub>), 24.9 (CH<sub>3</sub> (acetate)).

### 6.3.2.7 5-Chloro-3-(naphthalen-1-yl)-1,2,4-thiadiazole **50**

1-Naphthamidinium acetate **49** (682 mg, 2.96 mmol) was treated according to general procedure B. Eluent for chromatography: hexane/ethyl acetate 1:1.

Yield: 580 mg (2.35 mmol, 79%) colourless oil;  $C_{12}H_7ClN_2S$ ; molecular weight: 246.71 g/mol;  $R_f = 0.39$  (hexane/ethyl acetate 2:1).

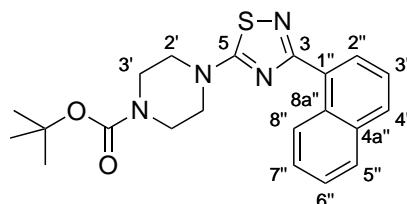


$^1H$ -NMR (500 MHz,  $CDCl_3$ ):  $\delta$  [ppm] = 9.07-9.04 (m, 1H, **H-8**), 8.34-8.31 (m, 1H, **H-2**), 8.01-7.98 (m, 1H, **H<sub>Aryl</sub>**), 7.94-7.91 (m, 1H **H<sub>Aryl</sub>**), 7.66-7.62 (m, 1H, **H<sub>Aryl</sub>**), 7.60-7.54, (m, 2H, **H<sub>Aryl</sub>**).

$^{13}C$ -NMR (125 MHz,  $CDCl_3$ ):  $\delta$  [ppm] = 172.9, 172.4 (**C-3**, **C-5**), 134.1 (**C-1'**), 131.8 (**C<sub>Aryl</sub>**), 130.8 (**C-4a/C-8a**), 130.4 (**C<sub>Aryl</sub>**), 128.83 (**C-4a/C-8a**), 128.75, 127.6, 126.3 (**C<sub>Aryl</sub>**), 126.0 (**C-8**), 125.1 (**C<sub>Aryl</sub>**).

### 6.3.2.8 *tert*-Butyl 4-(3-(naphthalen-1-yl)-1,2,4-thiadiazol-5-yl)piperazine-1-carboxylate **51**

5-Chloro-3-(naphthalen-1-yl)-1,2,4-thiadiazole **50** (477 mg, 1.93 mmol) and 1-boc-piperazine (359 mg, 1.93 mmol, 1.0 eq.) were reacted according to general procedure C using triethylamine (1072  $\mu$ L, 783 mg, 7.73 mmol, 4.0 eq.) as base. Eluent for chromatography: hexane/ethyl acetate 2:1 to 1:2).



Yield: 718 mg (1.81 mmol, 94%) colourless solid;  $C_{21}H_{24}N_2O_2S$ ; molecular weight: 396.51 g/mol,  $R_f = 0.54$  (hexane/ethyl acetate 1:1).

MS (ESI<sup>+</sup>): expt. 397.167 ([M+H]<sup>+</sup>), calc. 397.169 ([M+H]<sup>+</sup>); expt. 341.105 ([M-C<sub>4</sub>H<sub>9</sub><sup>+</sup>+2H]<sup>+</sup>), calc. 341.107 ([M-C<sub>4</sub>H<sub>9</sub><sup>+</sup>+2H]<sup>+</sup>).

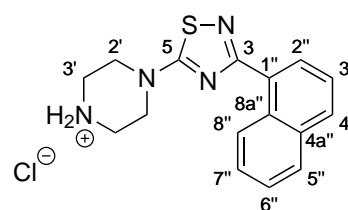
$^1H$ -NMR (500 MHz,  $CDCl_3$ ):  $\delta$  [ppm] = 8.98-8.94 (m, 1H, **H-8''**), 8.18-8.15 (m, 1H, **H-2''**), 7.94-7.90 (m, 1H, **H-4''**), 7.90-7.86 (m, 1H, **H-5''**), 7.58-7.48 (m, 3H, **H-3''**, **H-6''**, **H-7''**), 3.63 (s, 8H, **H-2'**, **H-3'**), 1.50 (s, 9H, *t*Bu (Boc)).

$^{13}C$ -NMR (125 MHz,  $CDCl_3$ ):  $\delta$  [ppm] = 184.8 (**C-5**), 171.2 (**C-3**), 154.6 (NC(=O)*t*Bu), 134.1 (**C-1''**), 131.2 (**C-8a''**), 130.8 (**C-4a''**), 130.6 (**C-4''**), 129.3 (**C-2''**), 128.5 (**C-5''**), 126.9 (**C-7''**), 126.5 (**C-8''**), 126.0 (**C-6''**), 125.2 (**C-3''**), 80.8 (**C(CH<sub>3</sub>)<sub>3</sub>**), 48.9 (**C-2'**, **C-3'**), 28.5 (**C(CH<sub>3</sub>)<sub>3</sub>**).

### 6.3.2.9 3-(Naphthalen-1-yl)-5-(piperazin-1-yl)-1,2,4-thiadiazole hydrochloride **37**

*tert*-Butyl 4-(3-(naphthalen-1-yl)-1,2,4-thiadiazol-5-yl)piperazine-1-carboxylate **51** (42 mg, 105  $\mu$ mol) was dissolved in a 3 M solution (4.0 mL) of hydrochloric acid in ethyl acetate. The reaction mixture was stirred for 2 h at room temperature and then neutralized with a saturated solution of sodium bicarbonate.

The aqueous phase was extracted with ethyl acetate. The combined organic phases were dried over sodium sulfate, filtered and freed of the solvent *in vacuo*. The residue was lyophilized from a dilute solution of hydrochloric acid. Finally, the residue was suspended in a solution (5.0 mL) of hexane/chloroform 1:1 and filtered. The product was washed with hexane (2.0 mL) twice.



**37**

Yield: 24 mg (73  $\mu$ mol, 69%) colourless solid; C<sub>16</sub>H<sub>17</sub>ClN<sub>4</sub>S; molecular weight: 332.85 g/mol.

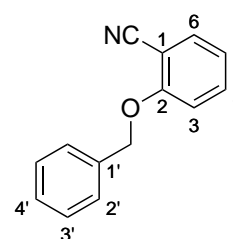
MS (ESI<sup>+</sup>): expt. 297.117 ([M+H]<sup>+</sup>), calc. 297.117 ([M+H]<sup>+</sup>).

<sup>1</sup>H-NMR (400 MHz, MeOD-d<sub>4</sub>):  $\delta$  [ppm] = 8.89-8.84 (m, 1H, **H-8''**), 8.13-8.09 (m, 1H, **H-2''**), 8.01-7.97 (m, 1H, **H-5''**), 7.96-7.91 (m, 1H, **H-4''**), 7.58-7.50 (m, 3H, **H-7''**, **H-6''**, **H-3''**), 3.97-3.91 (m, 4H, **H-2'**), 3.48-3.42 (m, 4H, **H-3'**).

<sup>13</sup>C-NMR (101 MHz, MeOD-d<sub>4</sub>):  $\delta$  [ppm] = 186.0 (**C-5**), 172.3 (**C-3**), 135.5 (**C-1''**), 132.3 (**C-8a''**), 131.8 (**C-5''**), 131.7 (**C-4a''**), 130.2 (**C-2''**), 129.5 (**C-4''**), 127.9, 127.3 (**C-7''/C-6''/C-3''**), 127.1 (**C-8''**), 126.0 (**C-7''/C-6''/C-3''**), 46.8 (**C-2'**), 43.8 (**C-3'**).

### 6.3.2.10 2-(Benzyloxy)benzonitrile **53**

Synthesis of 2-(benzyloxy)benzonitrile **53** has been described previously, e.g. by Loughlin *et al.*<sup>[126]</sup> 2-Cyanophenol **52** (2.095 g, 17.56 mmol) and potassium carbonate (4.856 g, 35.13 mmol) were suspended in *N,N*-dimethylformamide (20 mL). Then, benzyl bromide (2.504 mL, 3.606 g, 21.08 mmol) was added. The reaction mixture was stirred at 50 °C for 3 h and subsequently filtered. The filtrate was freed of the solvent *in vacuo* and the resulting residue was purified by column chromatography on silica gel (eluent: hexane/ethyl acetate 4:1).



**53**



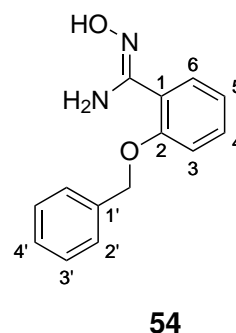
Yield: 3.605 g (17.23 mmol, 98%) colourless solid; C<sub>14</sub>H<sub>11</sub>NO; molecular weight: 209.24 g/mol, R<sub>f</sub> = 0.47 (hexane/ethyl acetate 2:1).

<sup>1</sup>H-NMR (500 MHz, CDCl<sub>3</sub>): δ [ppm] = 7.59-7.56 (m, 1H, **H-6**), 7.51-7.44 (m, 3H, **H-4**, **H-2'**), 7.42-7.37 (m, 2H, **H-3'**), 7.36-7.31 (m, 1H, **H-4'**), 7.03-6.99 (m, 2H, **H-3**, **H-5**), 5.22 (s, 2H, benzyl-CH<sub>2</sub>).

<sup>13</sup>C-NMR (125 MHz, CDCl<sub>3</sub>): δ [ppm] = 160.4 (**C-2**), 135.8 (**C-1'**), 134.4 (**C-4**), 134.0 (**C-6**), 128.8 (**C-2'**), 128.3 (**C-4'**), 127.1 (**C-3'**), 121.2 (**C-5**), 116.6 (-CN), 113.0 (**C-3**), 102.5 (**C-1**), 70.7 (benzyl-CH<sub>2</sub>).

### 6.3.2.11 (Z)-2-(Benzyloxy)-N'-hydroxybenzimidamide **54**

Synthesis of 2-(benzyloxy)benzonitrile **53** has been described previously, e.g. by Tarasenko *et al.*<sup>[127]</sup> 2-(Benzyloxy)benzonitrile **53** (3.543 g, 16.93 mmol) was treated according to general procedure A. The reaction time was 20 h. In total, 4 portions of hydroxylamine hydrochloride and sodium carbonate were added. The crude product was purified by column filtration over silica gel (eluent: hexane/ethyl acetate 2:1).



Yield: 3.827 g (15.79 mmol, 93%) colourless solid; C<sub>14</sub>H<sub>14</sub>N<sub>2</sub>O<sub>2</sub>; molecular weight: 242.27 g/mol, R<sub>f</sub> = 0.10 (hexane/ethyl acetate 2:1).

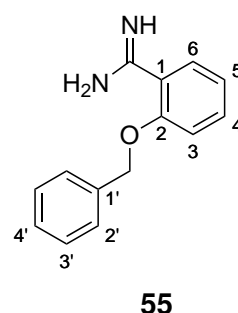
MS (ESI<sup>+</sup>): expt. 243.114 ([M+H]<sup>+</sup>), calc. 243.113 ([M+H]<sup>+</sup>).

<sup>1</sup>H-NMR (400 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 9.41 (s, 1H, N-OH), 7.52-7.48 (m, 2H, **H-2'**), 7.43-7.30 (m, 5H, **H-3'**, **H-4'**, **H-4**, **H-6**), 7.15-7.11 (m, 1H, **H-3**), 6.98-6.92 (m, 1H, **H-5**), 5.66 (bs, 2H, -NH<sub>2</sub>), 5.16 (s, 2H, benzyl-CH<sub>2</sub>).

<sup>13</sup>C-NMR (101 MHz, DMSO-*d*<sub>6</sub>): δ [ppm] = 156.2 (**C-2**), 150.8 (**C(-NOH)-NH<sub>2</sub>**), 137.0 (**C-1'**), 130.1, 129.8 (**C-4/C-6**), 128.4 (**C-3'**), 127.8 (**C-4'**), 127.5 (**C-2'**), 123.2 (**C-1**), 120.5 (**C-5**), 113.2 (**C-3**), 69.7 (benzyl-CH<sub>2</sub>).

### 6.3.2.12 2-(Benzyloxy)benzimidamide **63**

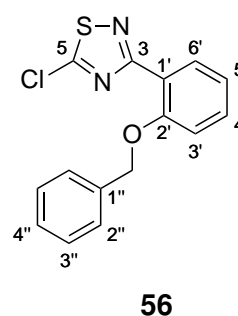
(*Z*)-2-(Benzyloxy)-*N'*-hydroxybenzimidamide **54** (130 mg, 0.537 mmol) was dissolved in methanol (12.9 mL). Then a 0.1 M solution (12.9 mL) of samarium(II) iodide (1.29 mmol, 2.4 eq.) in tetrahydrofuran was added. The reaction mixture was stirred at room temperature for 30 min and then freed of the solvent *in vacuo*. The residue was purified by column filtration over silica gel (eluent: dichloromethane/methanol 9:1 + 0.5% NH<sub>3</sub>). Formation of 2-(benzyloxy)benzimidamide **63** was verified by MS and all of the crude product was used for the next reaction without further purification.



### 6.3.2.13 3-(2-(Benzyloxy)phenyl)-5-chloro-1,2,4-thiadiazole **56**

All of the crude 2-(benzyloxy)benzimidamide **63** from the reaction described above was treated according to general procedure B. Eluent for chromatography: hexane/ethyl acetate 4:1.

Yield: 56 mg (0.18 mmol, 34% over two step) colourless oil; C<sub>15</sub>H<sub>11</sub>ClN<sub>2</sub>OS; molecular weight: 302.78 g/mol, R<sub>f</sub> = 0.49 (hexane/ethyl acetate 4:1).



<sup>1</sup>H-NMR (400 MHz, CDCl<sub>3</sub>): δ [ppm] = 8.00-7.97 (m, 1H, **H-6'**), 7.53-7.50 (m, 2H, **H-2''**), 7.45-7.34 (m, 3H, **H-3''**, **H-4'**), 7.33-7.27 (m, 1H, **H-4''**), 7.11-7.05 (m, 2H, **H-5'**), 5.24 (s, 2H, benzyl-CH<sub>2</sub>).

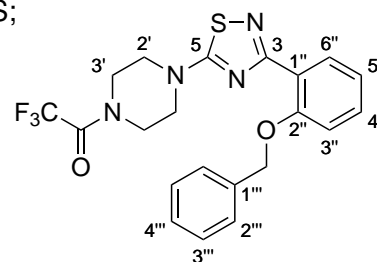
<sup>13</sup>C-NMR (101 MHz, CDCl<sub>3</sub>): δ [ppm] = 172.1 (**C-5**), 170.9 (**C-3**), 157.2 (**C-2'**), 137.0 (**C-1''**), 132.2, 132.1 (**C-4'/C-6'**), 128.6 (**C-3''**), 127.8 (**C-4''**), 127.0 (**C-2''**), 122.3 (**C-1'**), 121.2 (**C-5'**), 114.2 (**C-3'**), 71.0 (benzyl-CH<sub>2</sub>).

### 6.3.2.14 1-(4-(3-(2-(Benzyloxy)phenyl)-1,2,4-thiadiazol-5-yl)piperazin-1-yl)-2,2,2-trifluoroethanone **57**

3-(2-(Benzyloxy)phenyl)-5-chloro-1,2,4-thiadiazole **56** (82 mg, 0.27 mmol) and trifluoroacetyl-piperazine trifluoroacetate (120 mg, 0.41 mmol, 1.5 eq.) were reacted according to general procedure C using triethylamine (225 μL, 164 mg, 1.62 mmol, 6.0 eq.) as base. Eluent for

chromatography: hexane/ethyl acetate 4:1 to 1:1).

Yield: 103 mg (0.230 mmol, 85%) colourless oil; C<sub>21</sub>H<sub>19</sub>F<sub>3</sub>N<sub>4</sub>O<sub>2</sub>S;  
molecular weight: 448.46 g/mol, R<sub>f</sub> = 0.11 (hexane/ethyl  
acetate 4:1).



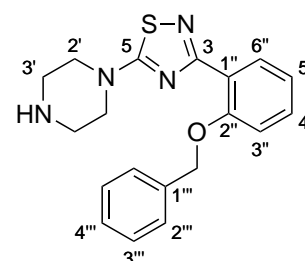
**57**

<sup>1</sup>H-NMR (500 MHz, CDCl<sub>3</sub>): δ [ppm] = 7.93-7.90 (m, 1H, H-6''), 7.51-7.47 (m, 2H, H-2'''), 7.40-7.32 (m, 3H, H-4'', H-3'''), 7.31-7.27 (m, 1H, H-4'''), 7.08-7.02 (m, 2H, H-3'', H-5''), 5.19 (s, 2H, benzyl-CH<sub>2</sub>), 3.83-3.67 (m, 4H, H-3'), 3.64-3.58 (m, 4H, H-2').

<sup>13</sup>C-NMR (125 MHz, CDCl<sub>3</sub>): δ [ppm] = 184.5 (C-5), 169.3 (C-3), 157.0 (C-2'), 155.8 (NC(=O)CF<sub>3</sub>), 137.3 (C-1'''), 131.9 (C-6''), 131.4 (C-4''), 128.4 (C-3'''), 127.8 (C-4'''), 127.3 (C-2'''), 123.4 (C-1''), 121.2 (C-5''), 116.3 (NC(=O)CF<sub>3</sub>), 114.2 (C-3''), 70.9 (benzyl-CH<sub>2</sub>), 48.7, 48.4 (C-2'), 45.0, 42.6 (C-3').

### 6.3.2.15 3-(2-(Benzyloxy)phenyl)-5-(piperazin-1-yl)-1,2,4-thiadiazole 39

1-(4-(3-(2-(Benzyloxy)phenyl)-1,2,4-thiadiazol-5-yl)piperazin-1-yl)-2,2,2-trifluoroethanone **57** (30 mg, 67 μmol) was treated according to general procedure D. Eluent for chromatography: dichloromethane/methanol 15:1 + 0.5% NH<sub>3</sub>.



**39**

Yield: 23 mg (66 μmol, 98%) colourless oil; C<sub>19</sub>H<sub>20</sub>N<sub>4</sub>OS; molecular weight: 352.45 g/mol, R<sub>f</sub> = 0.29 (dichloromethane/methanol 15:1 + 0.5% NH<sub>3</sub>).

MS (ESI<sup>+</sup>): expt. 353.152 ([M+H]<sup>+</sup>), calc. 353.143 ([M+H]<sup>+</sup>).

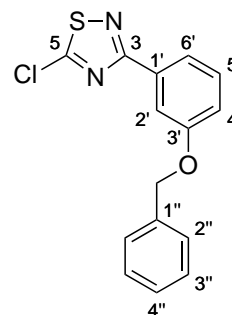
<sup>1</sup>H-NMR (400 MHz, MeOD-*d*<sub>4</sub>): δ [ppm] = 7.74-7.71 (m, 1H, H-6''), 7.47-7.43 (m, 2H, H-2'''), 7.42-7.37 (m, 1H, H-4''), 7.35-7.30 (m, 2H, H-3'''), 7.29-7.24 (m, 1H, H-4'''), 7.17-7.13 (m, 1H, H-3''), 7.05-7.00 (m, 1H, H-5''), 5.15 (s, 2H, benzyl-CH<sub>2</sub>), 3.54-3.50 (m, 4H, H-2'), 2.93-2.89 (m, 4H, H-3').

$^{13}\text{C}$ -NMR (101 MHz,  $\text{MeOD-}d_4$ ):  $\delta$  [ppm] = 186.0 (**C-5**), 170.6 (**C-3**), 158.3 (**C-2''**), 138.7 (**C-1'''**), 132.4 (**C-6''**), 132.3 (**C-4''**), 129.3 (**C-3'''**), 128.6 (**C-4'''**), 128.2 (**C-2'''**), 125.1 (**C-1''**), 121.9 (**C-5''**), 115.1 (**C-3''**), 71.7 (benzyl-**CH**<sub>2</sub>), 50.6 (**C-2'**), 45.6 (**C-3'**).

### 6.3.2.16 3-(3-(Benzyloxy)phenyl)-5-chloro-1,2,4-thiadiazole 59

3-(Benzyloxy)benzene-1-carboximidamide hydrochloride **58** (95 mg, 0.36 mmol) was treated according to general procedure B. Eluent for chromatography: hexane/ethyl acetate 4:1.

Yield: 66 mg (0.22 mmol, 60%) colourless oil;  $\text{C}_{15}\text{H}_{11}\text{ClN}_2\text{OS}$ ; molecular weight: 302.78 g/mol,  $R_f$  = 0.58 (hexane/ethyl acetate 4:1).



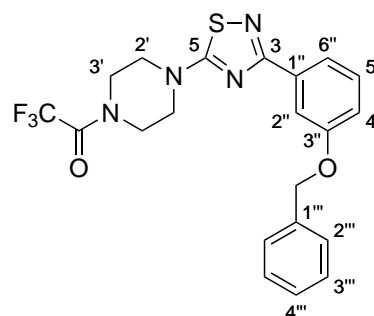
**59**

$^1\text{H}$ -NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  [ppm] = 7.92-7.89 (m, 1H, **H-6'**), 7.89-7.86 (m, 1H, **H-5'**), 7.50-7.45 (m, 2H, **H-2''**), 7.44-7.32 (m, 4H, **H-2'**, **H-3''**, **H-4''**), 7.14-7.10 (m, 1H, **H-4'**), 5.16 (s, 2H, benzyl-**CH**<sub>2</sub>).

$^{13}\text{C}$ -NMR (101 MHz,  $\text{CDCl}_3$ ):  $\delta$  [ppm] = 173.0 (**C-5**), 172.1 (**C-3**), 159.2 (**C-3'**), 136.8 (**C-1'''**), 133.3 (**C-1'**), 130.1 (**C-2'**), 128.7 (**C-3'''**), 128.2 (**C-4'''**), 127.7 (**C-2'''**), 121.0 (**C-5'**), 118.4 (**C-4'**), 113.8 (**C-6''**), 70.3 (benzyl-**CH**<sub>2</sub>).

### 6.3.2.17 1-(4-(3-(3-(Benzyloxy)phenyl)-1,2,4-thiadiazol-5-yl)piperazin-1-yl)-2,2,2-trifluoroethanone 60

3-(3-(Benzyloxy)phenyl)-5-chloro-1,2,4-thiadiazole **59** (66 mg, 0.22 mmol) and trifluoroacetyl piperazine trifluoroacetate (97 mg, 0.33 mmol, 1.5 eq.) were reacted according to general procedure C using triethylamine (181  $\mu\text{L}$ , 132 mg, 1.30 mmol, 6.0 eq.) as base. Eluent for chromatography: hexane/ethyl acetate 4:1 to 1:1).



**60**

Yield: 74 mg (0.17 mmol, 76%) colourless oil;  $\text{C}_{21}\text{H}_{19}\text{F}_3\text{N}_4\text{O}_2\text{S}$ ; molecular weight: 448.46 g/mol,  $R_f$  = 0.17 (hexane/ethyl acetate 4:1).

MS (ESI<sup>+</sup>): expt. 449.123 ([M+H]<sup>+</sup>), calc. 449.125 ([M+H]<sup>+</sup>).

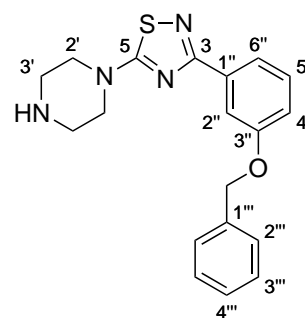
<sup>1</sup>H-NMR (400 MHz, MeOD-*d*<sub>4</sub>): δ [ppm] = 7.85-7.83 (m, 1H, **H-6''**), 7.82-7.78 (m, 1H, **H-5''**), 7.49-7.45 (m, 2H, **H-2'''**), 7.43-7.30 (m, 4H, **H-3'''**, **H-4'''**, **H-2''**), 7.08-7.03 (m, 1H, **H-4''**), 5.13 (s, 2H, benzyl-**CH**<sub>2</sub>), 3.88-3.75 (m, 4H, **H-3'**), 3.72-3.64 (m, 4H, **H-2'**).

<sup>13</sup>C-NMR (101 MHz, MeOD-*d*<sub>4</sub>): δ [ppm] = 185.1 (**C-5**), 170.3 (**C-3**), 159.0 (**C-3''**), 155.9 (NC(=O)CF<sub>3</sub>), 137.0 (**C-1'''**), 134.5 (**C-6''**), 129.7 (**C-2''**), 128.7 (**C-3'''**), 128.1 (**C-4'''**), 127.7 (**C-2'''**), 120.9 (**C-5''**), 117.3 (**C-4''**), 116.4 (NC(=O)CF<sub>3</sub>), 114.1 (**C-6''**), 70.2 (benzyl-**CH**<sub>2</sub>), 48.7, 48.3 (**C-2'**), 45.0, 42.6 (**C-3'**).

### 6.3.2.18 3-(3-(Benzyloxy)phenyl)-5-(piperazin-1-yl)-1,2,4-thiadiazole **38**

1-(4-(3-(3-(Benzyloxy)phenyl)-1,2,4-thiadiazol-5-yl)piperazin-1-yl)-2,2,2-trifluoroethanone **60** (23.5 mg, 52.4 μmol) was treated according to general procedure D. Eluent for chromatography: dichloromethane/methanol 30:1 + 0.5% NH<sub>3</sub>.

Yield: 18.1 mg (51.4 μmol, 58%) colourless oil; C<sub>19</sub>H<sub>20</sub>N<sub>4</sub>OS; molecular weight: 352.45 g/mol, R<sub>f</sub> = 0.13 (dichloromethane/methanol 30:1 + 0.5% NH<sub>3</sub>).



**38**

MS (ESI<sup>+</sup>): expt. 353.160 ([M+H]<sup>+</sup>), calc. 353.143 ([M+H]<sup>+</sup>).

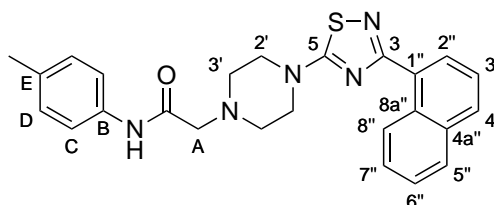
<sup>1</sup>H-NMR (400 MHz, MeOD-*d*<sub>4</sub>): δ [ppm] = 7.79-7.76 (m, 1H, **H-2''**), 7.75-7.71 (m, 1H, **H-6''**), 7.48-7.44 (m, 2H, **H-2'''**), 7.40-7.28 (m, 4H, **H-3'''**, **H-4'''**, **H-5''**), 7.09-7.04 (m, 1H, **H-4''**), 5.13 (s, 2H, benzyl-**CH**<sub>2</sub>), 3.60-3.55 (m, 4H, **H-2'**), 2.98-2.93 (m, 4H, **H-3'**).

<sup>13</sup>C-NMR (101 MHz, MeOD-*d*<sub>4</sub>): δ [ppm] = 186.6 (**C-5**), 171.4 (**C-3**), 160.3 (**C-3''**), 138.6 (**C-1'''**), 135.9 (**C-1''**), 130.6 (**C-5''**), 129.5 (**C-3'''**), 128.9 (**C-4'''**), 128.6 (**C-2'''**), 121.7 (**C-6''**), 117.9 (**C-4''**), 115.3 (**C-2''**), 71.1 (benzyl-**CH**<sub>2</sub>), 50.6 (**C-2'**), 45.7 (**C-3'**).

### 6.3.2.19 2-(4-(3-(Naphthalen-1-yl)-1,2,4-thiadiazol-5-yl)piperazin-1-yl)-*N*-(*p*-tolyl)acetamide **62**

Ligand **37** (32.1 mg, 96.4  $\mu\text{mol}$ ), 2-bromo-*N*-(4-methylphenyl)acetamide (22.0 mg, 96.4  $\mu\text{mol}$ ) and cesium carbonate (62.8 mg, 192.8  $\mu\text{mol}$ ) were suspended in acetonitrile (20 mL). The reaction mixture was heated to reflux for 20 h and then filtered. The filtrate was freed of the solvent *in vacuo* and the residue was purified by column chromatography on silica gel (eluent: hexane/ethyl acetate 3:2) and subsequently by RP-HPLC (column: Nucleodur C<sub>18</sub> Isis; solvent A: H<sub>2</sub>O + 0.1% formic acid; solvent B: MeCN + 0.1% formic acid; gradient: 0-8 min, 40 to 60% B; 8-15 min, 60 to 95% B; 15-18 min, 90% B; 18-20 min, 95 to 40% B; 20-22 min, 40% B; flow rate: 1.5 mL/min; R<sub>t</sub> ~ 10.2 min).

Yield: 23.8 mg (53.6  $\mu\text{mol}$ , 56%) colourless solid; C<sub>25</sub>H<sub>25</sub>N<sub>5</sub>OS; molecular weight: 443.56 g/mol, R<sub>f</sub> = 0.15 (hexane/ethyl acetate 3:2).



**62**

MS (ESI<sup>+</sup>): expt. 444.197 ([M+H]<sup>+</sup>), calc. 444.185 ([M+H]<sup>+</sup>).

<sup>1</sup>H-NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  [ppm] = 8.98 (s, 1H, NHC(=O)), 8.95-8.91 (m, 1H, **H-8''**), 8.17-8.13 (m, 1H, **H-2''**), 8.02 (s, 1H, formate), 7.95-7.91 (m, 1H, **H-3''/H-4''/H-5''/H-6''/H-7''**), 7.91-7.86 (m, 1H, **H-3''/H-4''/H-5''/H-6''/H-7''**), 7.59-7.48 (m, 3H, **H-3''/H-4''/H-5''/H-6''/H-7''**), 7.48-7.43 (m, 2H, **H-C**), 7.18-7.14 (m, 2H, **H-D**), 3.78-3.72 (m, 4H, **H-2'**), 3.31 (s, 2H, **H-A**), 2.88-2.83 (m, 4H, **H-3'**), 2.33 (s, 3H, Ph-CH<sub>3</sub>).

<sup>13</sup>C-NMR (101 MHz, CDCl<sub>3</sub>):  $\delta$  [ppm] = 184.7 (**C-5**), 171.3 (**C-3**), 167.5 (NHC(=O)), 163.9 (formate), 134.7, 134.6 (**C-E** + **C-1''/C-3''-C-7''/C-4a''/C-8a''/C-B**), 134.2, 131.2, 130.7, 130.7 (**C-1''/C-3''-C-7''/C-4a''/C-8a''/C-B**), 129.8 (**C-D**), 129.3 (**C-2''**), 128.6, 127.0 (**C-1''/C-3''-C-7''/C-4a''/C-8a''/C-B**), 126.5 (**C-8''**), 125.2 (**C-1''/C-3''-C-7''/C-4a''/C-8a''/C-B**), 119.9 (**C-C**), 61.9 (**C-A**), 52.6 (**C-3'**), 48.9 (**C-2'**), 21.0 (Ph-CH<sub>3</sub>).

### 6.3.3 Inhibition Assays

All inhibition assays were measured on a Bruker Avance 700 MHz spectrometer in 3 mm NMR tubes. Chemical shifts and concentrations of all spectra were referenced to TMSP-*d*<sub>4</sub>. Samples

were prepared in D<sub>2</sub>O containing sodium acetate-*d*<sub>3</sub> (50 mM), sodium chloride (100 mM), magnesium chloride (5 mM) and TMS-*d*<sub>4</sub> (1 mM) at pD 5.8. The GTB solution was rebuffed to the buffer specified above by using Amicon Ultra-4 cellulose filter (molecular weight cut-off 5 kDa) and the protein concentration was determined by using a nanodrop at 280 nm.

### 6.3.3.1 Competitive STD NMR

On resonance irradiation was applied at 0.2 ppm. Saturation was achieved by a cascade of 40 Gaussian pulses with a duration of 50 ms. A spinlock pulse of 25 ms length was used to reduce the protein background. Experiments were performed at 300 K. All samples contained a GTB concentration of 4.5 μM (except for ligand **62** for that only 3 μM were used) and additional 5% DMSO-*d*<sub>6</sub>. Following UDP concentrations were employed for the measurements of ligands **37**, **38**, **39**, and **62** respectively: 620 μM, 404 μM, 398 μM, and 348 μM. For every ligand, an artefact spectrum in the absence of GTB was obtained for the highest ligand concentration. No significant artefacts were observed. Spectra were acquired with 32,768 data points and a total of 512 or 1024 scans. FIDs were multiplied with an exponential function (line broadening 3) before Fourier transformation. For the determination of pIC<sub>50</sub> values, STD amplification factors for the H-5/H-1' proton of UDP were plotted against the logarithmic ligand concentration. The data points were then fitted using a one-site competition model in Origin2016G.

### 6.3.3.2 Progress curve analysis

Samples contained α-Fuc-(1,2)-β-Gal-octyl (2.0 mM), UDP-Gal (250 μM), GTB (1.0 μM), BSA (1 mg/mL), acidic phosphatase (0.8 mg/mL, from wheat germ, EC: 3.1.3.2, vendor: TCI). The reaction was monitored at 310 K via <sup>1</sup>H-NMR with excitation sculpting employing a pseudo 2D pulse program. A spectrum with 32,768 data points and 128 scans was acquired every 15 min. UDP-Gal was added last to the reaction mixture and the NMR tube was immediately transferred into the probe head. The first data point was usually obtained after 10-15 min. FIDs were multiplied with an exponential function (line broadening 3) before Fourier transformation. The decline of UDP-Gal concentration is reported as evaluated by the H-5 and H-1' signal. A control experiment in the absence of acceptor substrate (and both acceptor substrate and GTB) was performed to determine the extent of UDP-Gal depletion by acidic phosphatase.

## 7 References

- [1] F. R. Lichtenberg, The Impact of New Drug Launches on Longevity: Evidence from Longitudinal, Disease-Level Data from 52 Countries, 1982–2001, *Int. J. Health Care Finance Econ.* **2005**, *5*, 47–73.
- [2] J. W. Scannell, A. Blanckley, H. Boldon, B. Warrington, Diagnosing the decline in pharmaceutical R&D efficiency, *Nat. Rev. Drug Discov.* **2012**, *11*, 191–200.
- [3] S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, A. L. Schacht, How to improve R&D productivity: the pharmaceutical industry's grand challenge, *Nat. Rev. Drug Discov.* **2010**, *9*, 203–214.
- [4] E. David, T. Tramontin, R. Zimmel, Pharmaceutical R&D: the road to positive returns, *Nat. Rev. Drug Discov.* **2009**, *8*, 609–610.
- [5] J. Drews, Drug Discovery: A Historical Perspective, *Science* **2000**, *287*, 1960–1964.
- [6] R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. S. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, U. Schopfer, G. S. Sittampalam, Impact of high-throughput screening in biomedical research, *Nat. Rev. Drug Discov.* **2011**, *10*, 188–195.
- [7] A. M. Doweiko, QSAR: dead or alive?, *J. Comput. Aided Mol. Des.* **2008**, *22*, 81–89.
- [8] P. J. Hajduk, J. Greer, A decade of fragment-based drug design: strategic advances and lessons learned, *Nat. Rev. Drug Discov.* **2007**, *6*, 211–219.
- [9] D. E. Scott, A. G. Coyne, S. A. Hudson, C. Abell, Fragment-Based Approaches in Drug Discovery and Chemical Biology, *Biochemistry* **2012**, *51*, 4990–5003.
- [10] G. Schneider, Virtual screening: an endless staircase?, *Nat. Rev. Drug Discov.* **2010**, *9*, 273–276.
- [11] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyran, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, R. Abel, Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field, *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.



- [12] R. S. Bohacek, C. McMartin, W. C. Guida, The art and practice of structure-based drug design: A molecular modeling perspective, *Med. Res. Rev.* **1996**, *16*, 3–50.
- [13] T. Fink, H. Bruggesser, J.-L. Reymond, Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons, *Angew. Chem. Int. Ed.* **2005**, *44*, 1504–1508.
- [14] M. M. Hann, A. R. Leach, G. Harper, Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856–864.
- [15] M. Congreve, R. Carr, C. Murray, H. Jhoti, A ‘Rule of Three’ for fragment-based lead discovery?, *Drug Discov. Tod.* **2003**, *8*, 876–877.
- [16] P. W. Kenny, A. Leitão, C. A. Montanari, Ligand efficiency metrics considered harmful, *J. Comput. Aided Mol. Des.* **2014**, *28*, 699–710.
- [17] G. Williams, G. G. Ferenczy, J. Ulander, G. M. Keserű, Binding thermodynamics discriminates fragments from druglike compounds: a thermodynamic description of fragment-based drug discovery, *Drug Discov. Tod.* **2017**, *22*, 681–689.
- [18] E. Freire, Do Enthalpy and Entropy Distinguish First in Class From Best in Class?, *Drug Discov. Tod.* **2008**, *13*, 869–874.
- [19] W. P. Jencks, On the attribution and additivity of binding energies, *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4046–4050,.
- [20] C. W. Murray, M. L. Verdonk, The consequences of translational and rotational entropy lost by small molecules on binding to proteins, *J. Comput. Aided Mol. Des.* **2002**, *16*, 741–753.
- [21] S. Chung, J. B. Parker, M. Bianchet, L. M. Amzel, J. T. Stivers, Impact of linker strain and flexibility in the design of a fragment-based inhibitor, *Nat. Chem. Biol.* **2009**, *5*, 407–413.
- [22] R. A. Ward, C. Brassington, *et al.*, Design and Synthesis of Novel Lactate Dehydrogenase A Inhibitors by Fragment-Based Lead Generation, *J. Med. Chem.* **2012**, *55*, 3285–3306.
- [23] P. O. Nikiforov, S. Surade, M. Blaszczyk, V. Delorme, P. Brodin, A. R. Baulard, T. L. Blundell, C. Abell, A fragment merging approach towards the development of small molecule inhibitors of *Mycobacterium tuberculosis* EthR for use as ethionamide boosters, *Org. Biomol. Chem.* **2016**, *14*, 2318–2326.

- [24] K. L. Lee, C. M. Ambler, *et al.*, Discovery of Clinical Candidate 1-[(2S,3S,4S)-3-Ethyl-4-fluoro-5-oxopyrrolidin-2-yl]methoxy-7-methoxyisoquinoline-6-carboxamide (PF-06650833), a Potent, Selective Inhibitor of Interleukin-1 Receptor Associated Kinase 4 (IRAK4), by Fragment-Based Drug Design, *J. Med. Chem.* **2017**, *60*, 5521–5542.
- [25] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949.
- [26] N. Brooijmans, I. D. Kuntz, Molecular recognition and docking algorithms, *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
- [27] S. F. Sousa, P. A. Fernandes, M. J. Ramos, Protein–Ligand Docking: Current Status and Future Challenges, *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 15–26.
- [28] J. Liu, R. Wang, Classification of Current Scoring Functions, *J. Chem. Inf. Model.* **2015**, *55*, 475–482.
- [29] S.-Y. Huang, S. Z. Grinter, X. Zou, Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12899–12908.
- [30] R. Apweiler, H. Hermjakob, N. Sharon, On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochim. Biophys. Acta* **1999**, *1473*, 4–8.
- [31] A. Varki, Biological roles of oligosaccharides: all of the theories are correct, *Glycobiology* **1993**, *3*, 97–130.
- [32] S. S. Pinho, C. A. Reis, Glycosylation in cancer: mechanisms and clinical implications, *Nat. Rev. Cancer* **2015**, *15*, 540–555.
- [33] A. Varki, *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, 2. Auflage, **2009**.
- [34] F.-G. Hanisch, C. A. Reis, H. Clausen, H. Paulsen, Evidence for glycosylation-dependent activities of polypeptide N-acetylgalactosaminyltransferases rGalNAc-T2 and -T4 on mucin glycopeptides, *Glycobiology* **2001**, *11*, 731–740.
- [35] B. Ma, J. L. Simala-Grant, D. E. Taylor, Fucosylation in prokaryotes and eukaryotes, *Glycobiology* **2006**, *16*, 158–184.

- [36] D. J. Becker, J. B. Lowe, Fucose: biosynthesis and biological function in mammals, *Glycobiology* **2003**, *13*, 41–53.
- [37] P. Agrawal, B. Fontanals-Cirera, E. Sokolova, S. Jacob, C. A. Vaiana, D. Argibay, V. Davalos, M. McDermott, S. Nayak, F. Darvishian, M. Castillo, B. Ueberheide, I. Osman, D. Fenyő, L. K. Mahal, E. Hernando, A Systems Biology Approach Identifies FUT8 as a Driver of Melanoma Metastasis, *Cancer Cell* **2017**, *31*, 804–819.
- [38] R. Honma, I. Kinoshita, E. Miyoshi, U. Tomaru, Y. Matsuno, Y. Shimizu, S. Takeuchi, Y. Kobayashi, K. Kaga, N. Taniguchi, H. Dosaka-Akita, Expression of Fucosyltransferase 8 Is Associated with an Unfavorable Clinical Outcome in Non-Small Cell Lung Cancers, *Oncology* **2015**, *88*, 298–308.
- [39] X. Wang, J. Chen, Q. K. Li, S. B. Peskoe, B. Zhang, C. Choi, E. A. Platz, H. Zhang, Overexpression of  $\alpha(1,6)$  fucosyltransferase associated with aggressive prostate cancer, *Glycobiology* **2014**, *24*, 935–944.
- [40] R. L. Shields, J. Lai, R. Keck, L. Y. O'Connell, K. Hong, G. Meng, S. H. A. Weikert, L. G. Presta, Lack of Fucose on Human IgG1 N-Linked Oligosaccharide Improves Binding to Human Fc $\gamma$ RIII and Antibody-dependent Cellular Toxicity, *J. Biol. Chem.* **2002**, *277*, 26733–26740.
- [41] X. Wang, S. Inoue, J. Gu, E. Miyoshi, K. Noda, W. Li, Y. Mizuno-Horikawa, M. Nakano, M. Asahi, M. Takahashi, N. Uozumi, S. Ihara, S. H. Lee, Y. Ikeda, Y. Yamaguchi, Y. Aze, Y. Tomiyama, J. Fujii, K. Suzuki, A. Kondo, S. D. Shapiro, C. Lopez-Otin, T. Kuwaki, M. Okabe, K. Honke, N. Taniguchi, Dysregulation of TGF- $\beta$ 1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15791–15796.
- [42] L. Tedaldi, G. K. Wagner, Beyond substrate analogues: new inhibitor chemotypes for glycosyltransferases, *Med. Chem. Commun.* **2014**, *5*, 1106–1125.
- [43] Z. Tu, Y.-N. Lin, C.-H. Lin, Development of fucosyltransferase and fucosidase inhibitors, *Chem. Soc. Rev.* **2013**, *42*, 4459–4475.
- [44] Y. Manabe, S. Kasahara, Y. Takakura, X. Yang, S. Takamatsu, Y. Kamada, E. Miyoshi, D. Yoshidome, K. Fukase, Development of  $\alpha$ 1,6-fucosyltransferase inhibitors through the diversity-oriented syntheses of GDP-fucose mimics using the coupling between alkyne and sulfonyl azide, *Bioorg. Med. Chem.* **2017**, *25*, 2844–2850.

- [45] K. Hosoguchi, T. Maeda, J.-i. Furukawa, Y. Shinohara, H. Hinou, M. Sekiguchi, H. Togame, H. Takemoto, H. Kondo, S.-I. Nishimura, An Efficient Approach to the Discovery of Potent Inhibitors against Glycosyltransferases, *J. Med. Chem.* **2010**, *53*, 5607–5619.
- [46] J. Kamińska, J. Dzięciol, J. Kościelak, Triazine dyes as inhibitors and affinity ligands of glycosyltransferases, *Glycoconj. J.* **1999**, *16*, 719–723.
- [47] A. L. Hopkins, G. M. Keserü, P. D. Leeson, D. C. Rees, C. H. Reynolds, The role of ligand efficiency metrics in drug discovery, *Nat. Rev. Drug Discov.* **2014**, *13*, 105–121.
- [48] H. Ihara, et. al, Crystal structure of mammalian  $\alpha$ 1,6-fucosyltransferase, FUT8, *Glycobiology* **2007**, *5*, 455–466.
- [49] M. P. Kötzler, S. Blank, F. I. Bantleon, E. Spillner, B. Meyer, Donor substrate binding and enzymatic mechanism of human core  $\alpha$ 1,6-fucosyltransferase (FUT8), *Biochim. Biophys. Acta* **2012**, *1820*, 1915–1925.
- [50] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, R. G. Coleman, ZINC: A Free Tool to Discover Chemistry for Biology, *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.
- [51] P. J. Skinner, M. C. Cherrier, P. J. Webb, Y.-J. Shin, T. Gharbaoui, A. Lindstrom, V. Hong, S. Y. Tamura, H. T. Dang, C. C. Pride, R. Chen, J. G. Richman, D. T. Connolly, G. Semple, Fluorinated pyrazole acids are agonists of the high affinity niacin receptor GPR109a, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5620–5623.
- [52] G. E. Jagdmann, H. R. Munson, T. W. Gero, A Mild Efficient Procedure for the Conversion of Carboxylic Acid Esters to Primary Amides Using Formamide/Methanolic Sodium Methoxide, *Synth. Comm.* **1990**, *20*, 1203–1208.
- [53] S. W. Hansen, M. N. Erichsen, B. Fu, W. E. Bjørn-Yoshimoto, B. Abrahamsen, J. C. Hansen, A. A. Jensen, L. Bunch, Identification of a New Class of Selective Excitatory Amino Acid Transporter Subtype 1 (EAAT1) Inhibitors Followed by a Structure–Activity Relationship Study, *J. Med. Chem.* **2016**, *59*, 8757–8770.
- [54] L. De Luca, G. Giacomelli, A. Porcheddu, Beckmann Rearrangement of Oximes under Very Mild Conditions, *J. Org. Chem.* **2002**, *67*, 6272–6274.
- [55] P. N. Rylander, *Hydrogenation Methods*, Academic Press, 1. Auflage, **1985**.

- [56] H. Ihara, S. Hanashima, T. Okada, R. Ito, Y. Yamaguchi, N. Taniguchi, Y. Ikeda, Fucosylation of chitooligosaccharides by human  $\alpha$ 1,6-fucosyltransferase requires a nonreducing terminal chitotriose unit as a minimal structure, *Glycobiology* **2010**, *20*, 1021–1033.
- [57] B. J. Mayer, SH3 domains: complexity in moderation, *J. Cell Sci.* **2001**, *114*, 1253–1263.
- [58] M. P. Kötzer, S. Blank, F. I. Bantleon, M. Wienke, E. Spillner, B. Meyer, Donor Assists Acceptor Binding and Catalysis of Human  $\alpha$ 1,6-Fucosyltransferase, *ACS Chem. Biol.* **2013**, *8*, 1830–1840.
- [59] A. D. Calderon, Y. Liu, X. Li, X. Wang, X. Chen, L. Li, P. G. Wang, Substrate specificity of FUT8 and chemoenzymatic synthesis of core-fucosylated asymmetric *N*-glycans, *Org. Biomol. Chem.* **2016**, *14*, 4027–4031.
- [60] J. B. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika* **1964**, *29*, 1–27.
- [61] C. Mattos, D. Ringe, Locating and characterizing binding sites on proteins, *Nature Biotechnol.* **1996**, *14*, 595–599.
- [62] D. Kozakov, D. R. Hall, R. L. Napoleon, C. Yueh, A. Whitty, S. Vajda, New Frontiers in Druggability, *J. Med. Chem.* **2015**, *58*, 9063–9088.
- [63] G. Chessari, I. M. Buck, J. E. H. Day, P. J. Day, A. Iqbal, C. N. Johnson, E. J. Lewis, V. Martins, D. Miller, M. Reader, D. C. Rees, S. J. Rich, E. Tamanini, M. Vitorino, G. A. Ward, P. A. Williams, G. Williams, N. E. Wilsher, A. J.-A. Woolford, Fragment-Based Drug Discovery Targeting Inhibitor of Apoptosis Proteins: Discovery of a Non-Alanine Lead Series with Dual Activity Against cIAP1 and XIAP, *J. Med. Chem.* **2015**, *58*, 6574–6588.
- [64] Y. Shi, A Glimpse of Structural Biology through X-Ray Crystallography, *Cell* **2014**, *159*, 995–1014.
- [65] E. Fischer, Einfluss der Configuration auf die Wirkung der Enzyme, *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985–2993.
- [66] D. E. Koshland Jr., Application of a Theory of Enzyme Specificity to Protein Synthesis, *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98–104.
- [67] D. D. Boehr, R. Nussinov, P. E. Wright, The role of dynamic conformational ensembles in biomolecular recognition, *Nat. Chem. Biol.* **2009**, *5*, 789–796.

- [68] A. D. Vogt, E. Di Cera, Conformational Selection Is a Dominant Mechanism of Ligand Binding, *Biochemistry* **2013**, *52*, 5723–5729.
- [69] R. D. Cramer, D. E. Patterson, R. D. Clark, F. Soltanshahi, M. S. Lawles, Virtual Compound Libraries: A New Approach to Decision Making in Molecular Discovery Research, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1010–1023.
- [70] S. L. McGovern, B. K. Shoichet, Information Decay in Molecular Docking Screens against Holo, Apo, and Modeled Conformations of Enzymes, *J. Med. Chem.* **2003**, *46*, 2895–2907.
- [71] I. Kufareva, V. Katritch, R. C. Stevens, R. Abagyan, Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges, *Structure* **2014**, *22*, 1120–1139.
- [72] H. A. Carlson, R. D. Smith, K. L. Damm-Ganamet, J. A. Stuckey, A. Ahmed, M. A. Convery, D. O. Somers, M. Kranz, P. A. Elkins, G. Cui, C. E. Peishoff, M. H. Lambert, J. Dunbar, J. B., CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma, *J. Chem. Inf. Model.* **2016**, *56*, 1063–1077.
- [73] J. D. Durrant, J. A. McCammon, Computer-Aided Drug-Discovery Techniques that Account for Receptor Flexibility, *Curr. Opin. Pharmacol.* **2010**, *10*, 770–774.
- [74] A. M. Ferrari, B. Q. Wei, L. Costantino, B. K. Shoichet, Soft Docking and Multiple Receptor Conformations in Virtual Screening, *J. Med. Chem.* **2004**, *47*, 5076–5084.
- [75] A. Yi-Ching Yang, P. Källblad, R. L. Mancera, Molecular Modeling Prediction of Ligand Binding Site Flexibility, *J. Comput.-Aided Mol. Des.* **2004**, *18*, 235–250.
- [76] R. M. A. Knegtel, I. D. Kuntz, C. M. Oshiro, Molecular Docking to Ensembles of Protein Structures, *J. Mol. Biol.* **1997**, *266*, 424–440.
- [77] L. S. Cheng, R. E. Amaro, D. Xu, W. W. Li, P. W. Arzberger, J. A. McCammon, Ensemble-Based Virtual Screening Reveals Potential Novel Antiviral Compounds for Avian Influenza Neuraminidase, *J. Med. Chem.* **2008**, *51*, 3878–3894.
- [78] D. J. Osguthorpe, W. Sherman, A. T. Hagler, Exploring Protein Flexibility: Incorporating Structural Ensembles From Crystal Structures and Simulation into Virtual Screening Protocols, *J. Phys. Chem. B* **2012**, *116*, 6952–6959.

- [79] S. Tian, H. Sun, P. Pan, D. Li, X. Zhen, Y. Li, T. Hou, Assessing an Ensemble Docking-Based Virtual Screening Strategy for Kinase Targets by Considering Protein Flexibility, *J. Chem. Inf. Model.* **2014**, *54*, 2664–2679.
- [80] J.-H. Lin, A. L. Perryman, J. R. Schames, J. A. McCammon, The Relaxed Complex Method: Accommodating Receptor Flexibility for Drug Design with an Improved Scoring Scheme, *Biopolymers* **2003**, *68*, 47–62.
- [81] R. E. Amaro, R. Baron, J. A. McCammon, An Improved Relaxed Complex Scheme for Receptor Flexibility in Computer-Aided Drug Design, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 693–705.
- [82] M. Rueda, G. Bottegoni, R. Abagyan, Recipes for the Selection of Experimental Protein Conformations for Virtual Screening, *J. Chem. Inf. Model.* **2010**, *50*, 186–193.
- [83] O. Korb, T. S. G. Olsson, S. J. Bowden, R. J. Hall, M. L. Verdonk, L. J. W., J. C. Cole, Potential and Limitations of Ensemble Docking, *J. Chem. Inf. Model.* **2012**, *52*, 1262–1274.
- [84] H. Claußen, C. Buning, M. Rarey, T. Lengauer, FlexE: Efficient Molecular Docking Considering Protein Structure Variations, *J. Mol. Biol.* **2001**, *308*, 377–395.
- [85] T. Polgár, G. M. Keserü, Ensemble Docking into Flexible Active Sites. Critical Evaluation of FlexE against JNK-3 and  $\beta$ -Secretase, *J. Chem. Inf. Model.* **2006**, *46*, 1795–1805.
- [86] T. L. Offutt, R. V. Swift, R. E. Amaro, Enhancing Virtual Screening Performance of Protein Kinases with Molecular Dynamics Simulations, *J. Chem. Inf. Model.* **2016**, *56*, 1923–1935.
- [87] Z. Huang, C. F. Wong, Inexpensive Method for Selecting Receptor Structures for Virtual Screening, *J. Chem. Inf. Model.* **2016**, *56*, 21–34.
- [88] B. Knapp, L. Luis Ospina, C. M. Deane, Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas, *J. Chem. Theory Comput.* **2018**, *14*, 6127–6138.
- [89] A. N. Jain, N. A., Recommendations for evaluation of computational methods, *J. Comput. Aided Mol. Des.* **2008**, *22*, 133–139.
- [90] M. M. Mysinger, M. Garchia, J. J. Irwin, B. K. Shoichet, Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking, *J. Med. Chem.* **2012**, *55*, 6582–6594.

- [91] J.-F. Truchon, C. I. Bayly, Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem, *J. Chem. Inf. Model.* **2007**, *47*, 488–508.
- [92] R. D. Clark, D. J. Webster-Clark, Managing Bias in ROC Curves, *J. Comput. Aided Mol. Des.* **2008**, *22*, 141–146.
- [93] A. Grossfield, D. M. Zuckermann, Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations, *Annu. Rep. Comput. Chem.* **2009**, *5*, 23–48.
- [94] J. W. Sammon, A Nonlinear Mapping for Data Structure Analysis, *IEEE Trans. Comput.* **1969**, *C 18*, 401–409.
- [95] R. L. Webb, N. Schiering, R. Sedrani, J. Maibaum, Direct Renin Inhibitors as a New Therapy for Hypertension, *J. Med. Chem.* **2010**, *53*, 7490–7520.
- [96] X. Barril, S. D. Morley, Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures, *J. Med. Chem.* **2005**, *48*, 4432–4443.
- [97] S. R. Ellingson, Y. Miao, J. Baudry, J. C. Smith, Multi-Conformer Ensemble Docking to Difficult Protein Targets, *J. Phys. Chem. B* **2015**, *119*, 1026–1034.
- [98] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, Peptide Folding: When Simulation Meets Experiment, *Angew. Chem. Int. Ed.* **1999**, *38*, 236–240.
- [99] J. Shao, S. W. Tanner, N. Thompson, T. E. Cheatham III, Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms, *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.
- [100] I. R. Craig, J. W. Essex, K. Spiegel, Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments, *J. Chem. Inf. Model.* **2010**, *50*, 511–524.
- [101] P. Greenwell, Blood group antigens: molecules seeking a function?, *Glyconj. J.* **1997**, *14*, 159–173.
- [102] L. L. Grimm, S. Weissbach, F. Flügge, N. Begemann, M. M. Palcic, T. Peters, Protein NMR Studies of Substrate Binding to Human Blood Group A and B Glycosyltransferases, *ChemBioChem* **2017**, *18*, 1260–1269.
- [103] J. A. Alfaro, R. B. Zheng, M. Persson, J. A. Letts, R. Polakowski, Y. Bai, S. N. Borisova, N. O. L. Seto, T. L. Lowary, M. M. Palcic, S. V. Evans, ABO(H) Blood Group A and B











- Glycosyltransferases Recognize Substrate via Specific Conformational Changes, *J. Biol. Chem.* **2008**, *283*, 10097–10108.
- [104] S. M. L. Gagnon, P. J. Meloncelli, R. B. Zheng, O. Haji-Ghassemi, A. R. Johal, S. N. Borisova, T. L. Lowary, S. V. Evans, High Resolution Structures of the Human ABO(H) Blood Group Enzymes in Complex with Donor Analogs Reveal That the Enzymes Utilize Multiple Donor Conformations to Bind Substrates in a Stepwise Manner, *J. Biol. Chem.* **2015**, *290*, 27040–27052.
- [105] S.-i. Hakomori, Antigen structure and genetic basis of histo-blood groups A, B and O: their changes associated with human cancer, *Biochim. Biophys. Acta* **1999**, *1473*, 247–266.
- [106] S. K. Rummel, R. E. Ellsworth, The role of the histoblood ABO group in cancer, *Future Sci. OA* **2016**, *2*, FSO107.
- [107] B. M. Wolpin, P. Kraft, M. Gross, K. Helzlsouer, H. B. Bueno-de Mesquita, E. Steplowski, R. Z. Stolzenberg-Solomon, A. A. Arslan, E. J. Jacobs, A. LaCroix, G. Petersen, W. Zheng, D. Albanes, N. E. Allen, L. Amundadottir, G. Anderson, M.-C. Boutron-Ruault, J. E. Buring, F. Canzian, S. J. Chanock, S. Clipp, J. M. Gaziano, E. L. Giovannucci, G. Hallmans, S. E. Hankinson, R. N. Hoover, D. J. Hunter, A. Hutchinson, K. Jacobs, C. Kooperberg, S. M. Lynch, J. B. Mendelsohn, D. S. Michaud, K. Overvad, A. V. Patel, A. Rajkovic, M.-J. Sánchez, X.-O. Shu, N. Slimani, G. Thomas, G. S. Tobias, D. Trichopoulos, P. Vineis, J. Virtamo, J. Wactawski-Wende, K. Yu, A. Zeleniuch-Jacquotte, P. Hartge, C. S. Fuchs, Pancreatic Cancer Risk and ABO Blood Group Alleles: Results from the Pancreatic Cancer Cohort Consortium, *Cancer Res.* **2010**, *70*, 1015–1023.
- [108] N. N. Rahbari, U. Bork, U. Hinz, A. Leo, J. Kirchberg, M. Koch, M. W. Büchler, J. Weitz, ABO blood group and prognosis in patients with pancreatic cancer, *BMC Cancer* **2012**, *12*, 1–9.
- [109] T. Wu, X.-A. Ma, G.-Q. Wang, Q. Li, M.-J. Li, J.-Y. Guo, X. Liang, Z.-P. Ruan, T. Tian, K.-J. Nan, L.-N. Liu, H. Guo, ABO blood type correlates with survival in hepatocellular carcinoma following hepatectomy, *Sci. Rep.* **2017**, *7*, 4412.
- [110] K. Schaefer, N. Sindhuwinata, T. Hackl, M. P. Kötzler, F. C. Niemeyer, M. M. Palcic, T. Peters, B. Meyer, A Nonionic Inhibitor with High Specificity for the UDP-Gal Donor









- Binding Site of Human Blood Group B Galactosyltransferase: Design, Synthesis, and Characterization, *J. Med. Chem.* **2013**, *56*, 2150–2154.
- [111] C. Rademacher, J. Landström, N. Sindhuwinata, M. M. Palcic, G. Widmalm, T. Peters, NMR-based exploration of the acceptor binding site of human blood group B galactosyltransferase with molecular fragments, *Glycoconj. J.* **2010**, *27*, 349–358.
- [112] R. Jørgensen, L. L. Grimm, N. Sindhuwinata, T. Peters, M. M. Palcic, A Glycosyltransferase Inhibitor from a Molecular Fragment Library Simultaneously Interferes with Metal Ion and Substrate Binding, *Angew. Chem. Int. Ed.* **2012**, *51*, 4171–4175.
- [113] P. Leccese, *Design, Synthese und Analyse von Inhibitoren der humanen Galactosyltransferase B (GTB)*, PhD thesis, **2016**.
- [114] B. D. Judkins, D. G. Allen, A. Cook, Tracey, B. Evans, T. E. Sardharwala, A Versatile Synthesis of Amidines from Nitriles Via Amidoximes, *Synth. Commun.* **1996**, *26*, 4351–4367.
- [115] R. S. Garigipati, An efficient conversion of nitriles to amidines, *Tetrahedron Lett.* **1990**, *31*, 1969–1972.
- [116] U. E. W. Lange, B. Schäfer, A new mild method for the synthesis of amidines, *Tetrahedron Lett.* **1999**, *40*, 7067–7071.
- [117] J. Cesar, K. Nadrah, M. Sollner Dolenc, Solid-phase synthesis of amidines by the reduction of amidoximes, *Tetrahedron Lett.* **2004**, *45*, 7445–7449.
- [118] M. Mayer, B. Meyer, Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.
- [119] N. Sindhuwinata, L. L. Grimm, S. Weißbach, S. Zinn, E. Munoz, M. M. Palcic, T. Peters, Thermodynamic signature of substrates and substrate analogs binding to human blood group B galactosyltransferase from isothermal titration calorimetry experiments, *Biopolymers* **2013**, *99*, 784–795.
- [120] Y.-C. Cheng, W. H. Prusoff, Relationship between the inhibition constant ( $K_i$ ) and the concentration of inhibitor which causes 50 per cent inhibition ( $I_{50}$ ) of an enzymatic reaction, *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.

- [121] S. Schnell, C. Mendoza, Closed Form Solution for Time-dependent Enzyme Kinetics, *J. Theor. Biol.* **1997**, *187*, 207–212.
- [122] C. T. Goudar, J. R. Sonnad, R. G. Duggleby, Parameter estimation using a direct solution of the integrated Michaelis-Menten equation, *Biochim. Biophys. Acta* **1999**, *1429*, 377–383.
- [123] F. Exnowitz, B. Meyer, T. Hackl, NMR for direct determination of  $K_m$  and  $V_{max}$  of enzyme reactions based on the Lambert W function-analysis of progress curves, *Biochim. Biophys. Acta* **2012**, *1824*, 443–449.
- [124] A. Kivrak, M. Zora, A Novel Synthesis of 1,2,4-Oxadiazoles and Isoxazoles, *Tetrahedron* **2014**, *70*, 817–831.
- [125] C. Alonso-Alija, H. Gielen, M. Hendrix, U. Niewöhner, D. Schauss, H. Bischoff, N. Burkhardt, V. Geiss, K.-H. Schlemmer, N. J. Cuthbert, M. Fitzgerald, R. G. Sturton, M. Michels, Preparation of 5-ethylimidazotriazinone PDE 4 inhibitors for treatment and/or prophylaxis of inflammatory processes and/or immune diseases, *WO 02/098880* **2002**.
- [126] W. A. Loughlin, S. A. Knevitt, R. E. Hosking, R. L. Marshall, Approaches to the High-Throughput Synthesis of Analogues of Dihydroaeruginic Acid, *Austr. J. Chem.* **2000**, *53*, 457–462.
- [127] M. Tarasenko, N. Duderin, T. Sharonova, S. Baykov, A. V. Smirnov, Room-Temperature Synthesis of Pharmaceutically Important Carboxylic Acids Bearing the 1,2,4-Oxadiazole Moiety, *Tetrahedron Lett.* **2017**, *58*, 3672–3677.

## 8 Hazards

| Substance        | GHS pictogram   | Hazard statement             | Precautionary statement  |
|------------------|---|------------------------------|--|
| Acetic acid      |    | H226-H314                    | P280-P305 + P351 + P338-P310   |
| Acetic anhydride |    | H226-H302-H314-H330          | P210-P260-P280-P304 + P340 + P310-P305 + P351 + P338-P370 + P378                             |
| Acetone          |    | H225-H319-H336               | P210-P280-P304 + P340 + P312-P305 + P351 + P338-P337 + P313-P403 + P235                      |
| Acetonitrile     |  | H225-H302 + H312 + H332-H319 | P210-P261-P280-P305 + P351 + P338-P370 + P378-P403 + P235                                    |
| Barium hydroxide |  | H302-H314                    | P260-P280-P301 + P312 + P330-P303 + P361 + P353-P304 + P340 + P310-P305 + P351 + P338 + P310 |
| Benzyl bromide   |  | H315-H319-H335               | P261-P305 + P351 + P338  |
| 1-Boc-piperazine |  | H315-H319-H335               | P261-P305 + P351 + P338  |
| Cesium carbonate |  | H318-H335                    | P280-P305 + P351 + P338 + P310   |









Hazards

| Substance                              | GHS pictogram   | Hazard statement                                 | Precautionary statement  |
|--|---|--|--|
| Chloroform                             |    | H302-H315-H319-<br>H331-H336-H351-<br>H361d-H372 | P201-P260-P264-<br>P280-P304 + P340<br>+ P312-P403 +<br>P233   |
| 1-Cyanonaphthalene                     |    | H302-H312-H315-<br>H319-H332-H335                | P261-P280-P305 +<br>P351 + P338  |
| 2-Cyanophenol                          |    | H302-H317-H318                                   | P280-P301 + P312<br>+ P330-P305 +<br>P351 + P338 +<br>P310   |
| Cyanuric chloride                      |   | H302-H314-H317-<br>H330                          | P260-P280-P301 +<br>P312 + P330-P303<br>+ P361 +<br>P353-P304 + P340<br>+ P310-P305 +<br>P351 + P338 |
| Dichloromethane                        |  | H315-H319-H335-<br>H336-H351-H371                | P260-P280-P305 +<br>P351 + P338  |
| Diethyl oxalate                        |  | H302-H314-<br>H360FD-H371-<br>H372               | P201-P260-P280-<br>P303 + P361 +<br>P353-P305 + P351<br>+ P338-P308 +<br>P313                        |
| <i>N,N</i> -Diisopropylethyl-<br>amine |  | H225-H302-H318-<br>H331-H335                     | P210-P261-P280-<br>P304 + P340 +<br>P312-P305 + P351<br>+ P338 +<br>P310-P403 + P233                 |
| <i>N,N</i> -Dimethyl-<br>formamide     |  | H226-H312 +<br>H332-H319-H360D                   | P201-P210-P261-<br>P280-P308 +<br>P313-P370 + P378   |









*Hazards*

| Substance                   | GHS pictogram | Hazard statement  | Precautionary statement   |
|-----------------------------|---------------|---|---|
| Ethanol                     |               | H225-H319   | P210-P305 + P351<br>+ P338-P370 +<br>P378-P403 + P235   |
| Ethyl acetate               |               | H225-H319-H336  | P210-P305 + P351<br>+ P338-P370 +<br>P378-P403 + P235   |
| Ethyl succinyl chloride     |               | H314  | P280-P305 + P351<br>+ P338-P310   |
| Formamide                   |               | H351-H360D-H373   | P201-P260-P280-<br>P308 +<br>P313   |
| Formic acid                 |               | H226-H302-H314-<br>H331                                   | P210-P280-P303 +<br>P361 + P353-P304<br>+ P340 +<br>P310-P305 + P351<br>+ P338-P403 +<br>P233 |
| Hexane                      |               | H225-H304-H315-<br>H336-H361f-H373-<br>H411               | P201-P210-P273-<br>P301 + P310-P308<br>+ P313-P331  |
| Hydrazine monohydrochloride |               | H301 + H311 +<br>H331-H317-H350-<br>H410                  | P201-P261-P273-<br>P280-P301 +<br>P310-P311   |
| Hydrochloric acid           |               | H290-H314-H335  | P261-P280-P305 +<br>P351 + P338-P310  |
| Hydroxylamine hydrochloride |               | H290-H302 +<br>H312-H315-H317-<br>H319-H351-H373-<br>H410 | P273-P280-P305 +<br>P351 + P338-P501  |
| Isophthalaldehyde           | -             | -   | -   |

*Hazards*

| Substance                                    | GHS pictogram   | Hazard statement                                  | Precautionary statement   |
|--|---|---|---|
| Methyl iodide                                |    | H301 + H331-<br>H312-H315-H317-<br>H334-H335-H351 | P261-P280-P301 +<br>P310-P311   |
| Methanol                                     |    | H225-H301 + H311<br>+ H331-H370                   | P210-P280-P302 +<br>P352 + P312-P304<br>+ P340 +<br>P312-P370 +<br>P378-P403 + P235           |
| Palladium on carbon                          | -   | -   | -   |
| Perchloromethyl mercaptan                    |    | H301 +<br>H311-H314-H330                          | P260-P280-P284-<br>P301 + P310-P305<br>+ P351 +<br>P338-P310                                  |
| Potassium carbonate                          |  | H315-H319-H335                                    | P305 + P351 +<br>P338   |
| Potassium permanganate                       |  | H272-H302-H410                                    | P220-P273-P501  |
| Piperidine                                   |  | H225-H302-H311 +<br>H331-H314                     | P210-P280-P304 +<br>P340 + P310-P305<br>+ P351 +<br>P338-P370 +<br>P378-P403 + P235           |
| Propylphosphonic acid anhydride (50% in DMF) |  | H290-H314-H360                                    | P201-P280-P303 +<br>P361 + P353-P304<br>+ P340 +<br>P310-P305 + P351<br>+ P338-P308 +<br>P313 |
| Pyridine                                     |  | H225-H302 + H312<br>+ H332-H315-H319              | P210-P280-P305 +<br>P351 + P338   |

Hazards

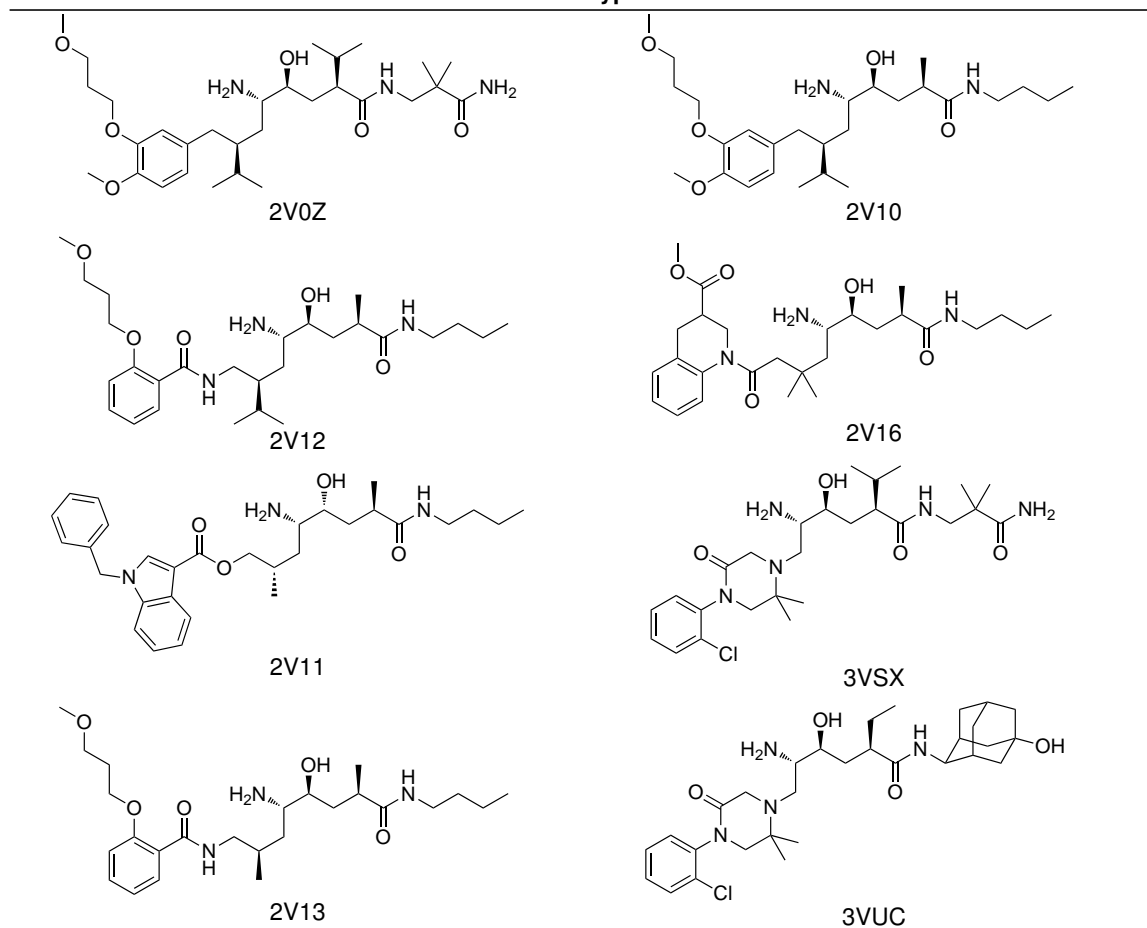
| Substance                                   | GHS pictogram   | Hazard statement                | Precautionary statement   |
|---|---|---------------------------------|---|
| Samarium(II) iodide solution (0.1 M in THF) |    | H225-H319-H335-H351-H373        | P210-P260-P280-P305 + P351 + P338-P370 + P378-P403 + P235               |
| Sodium                                      |    | H260-H314                       | P223-P231 + P232-P280-P305 + P351 + P338-P370 + P378-P422               |
| Sodium carbonate                            |    | H319                            | P305 + P351 + P338  |
| Sodium hydroxide                            |    | H290-H314                       | P280-P303 + P361 + P353-P304 + P340 + P310-P305 + P351 + P338           |
| Sulfuric acid                               |  | H290-H314                       | P260-P280-P303 + P361 + P353-P304 + P340 + P310-P305 + P351 + P338      |
| Triethylamine                               |  | H225-H302-H311 + H331-H314-H335 | P210-P261-P280-P303 + P361 + P353-P305 + P351 + P338-P370 + P378        |
| Trifluoroacetic acid                        |  | H314-H332-H412                  | P273-P280-P305 + P351 + P338-P310                                       |
| Tetrahydrofuran                             |  | H225-H302-H319-H335-H351        | P210-P280-P301 + P312 + P330-P305 + P351 + P338-P370 + P378-P403 + P235 |



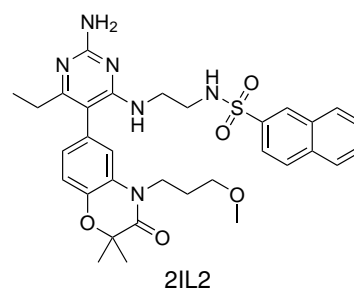
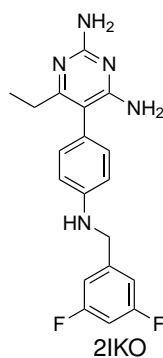
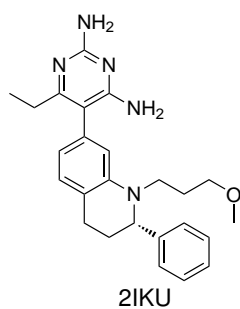
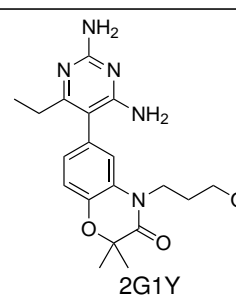
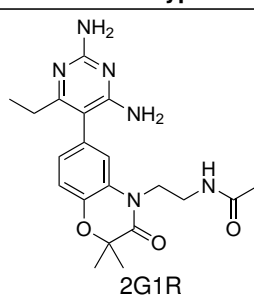
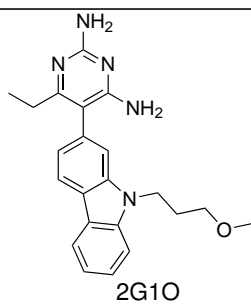
## 9 Appendix

### 9.1 Chemotypes of Renin Inhibitors

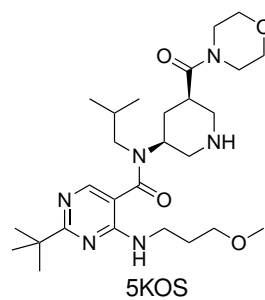
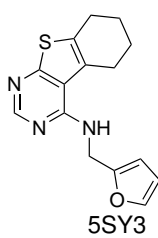
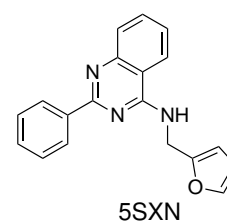
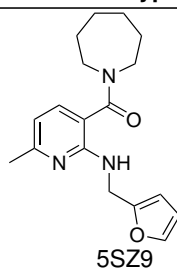
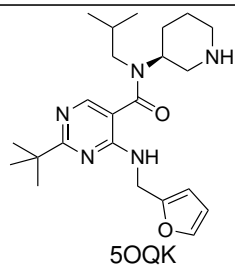
Chemotype 1



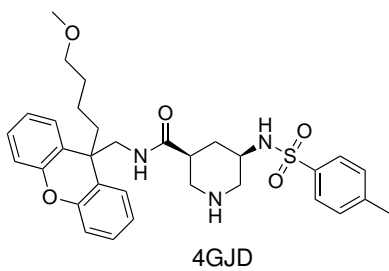
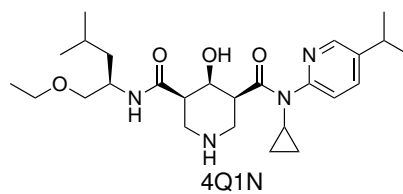
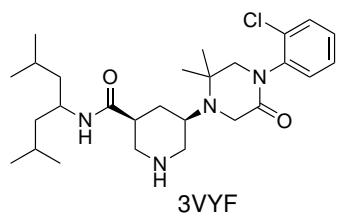
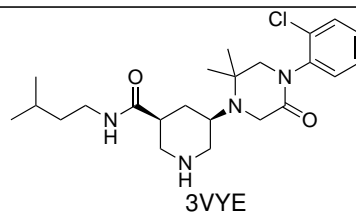
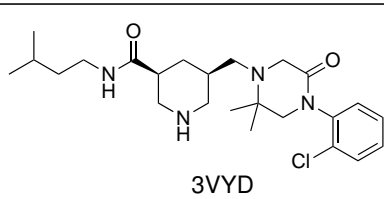
**Chemotype 2**



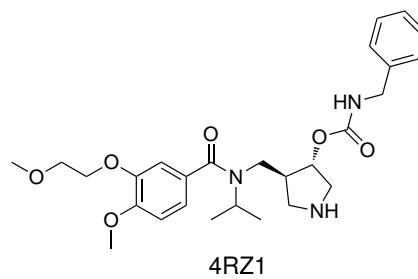
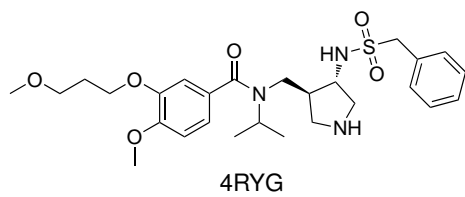
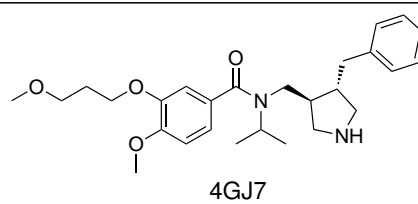
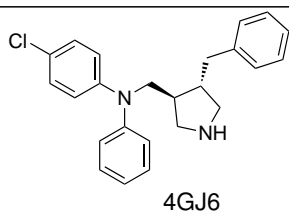
**Chemotype 3**



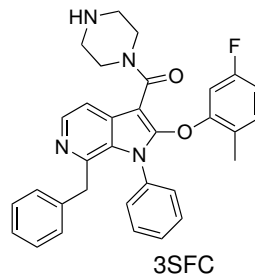
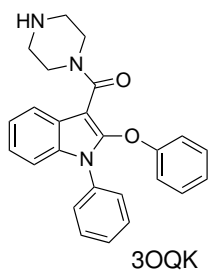
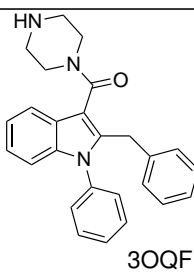
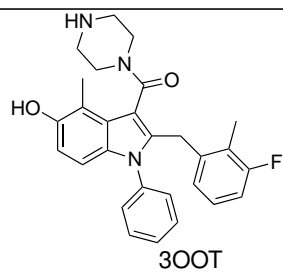
## Chemotype 4



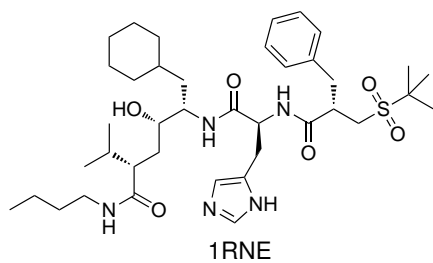
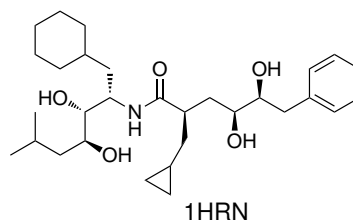
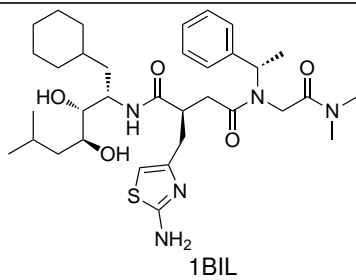
## Chemotype 5



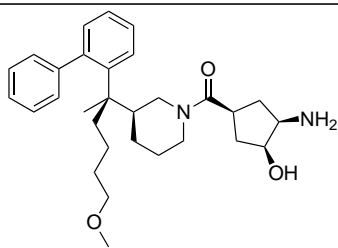
Chemotype 6



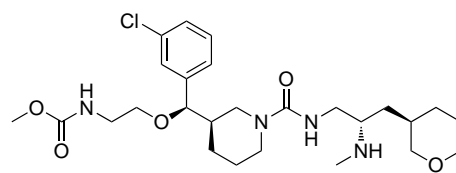
Chemotype 7



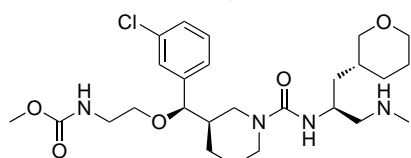
Chemotype 8



3Q3T

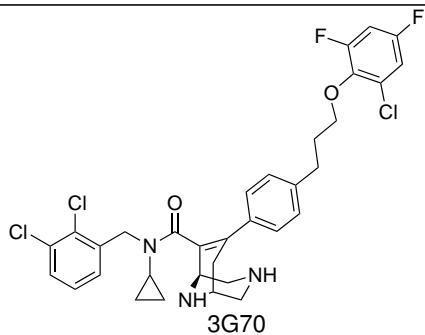


3Q4B

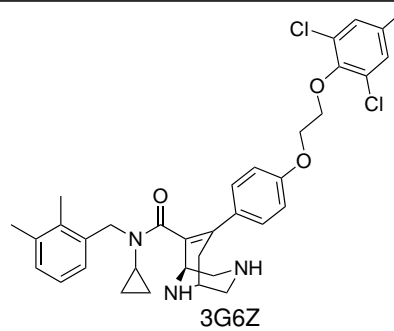


3Q5H

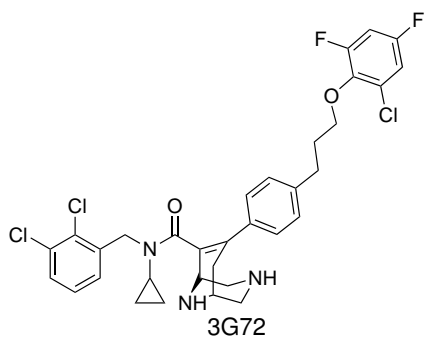
Chemotype 9



3G70

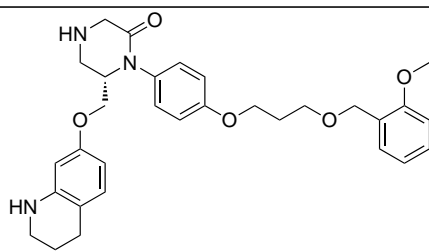


3G6Z

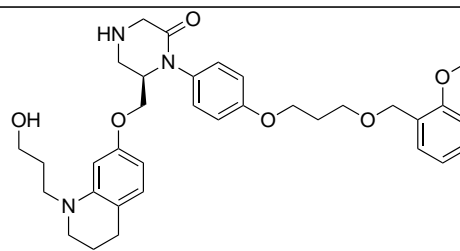


3G72

Chemotype 10

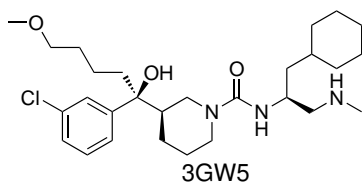


2BKT

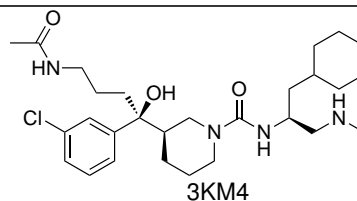


2FS4

Chemotype 11



3GW5



3KM4

## 9.2 Python Script for the Construction of RMSD Matrices

Python script that calculates pairwise RMSD values between all PyMol entries:

```
from pymol import cmd
def RMSDmatrix():
    entry = cmd.get_names()
    for i in range(0, len(entry) ):
        for j in range(i+1, len(entry) ):
            RMSD = cmd.align('%s' % (entry[i]), '%s' % (entry[j]), cycles=0)
            RMSDstring = '%6.3f' % (RMSD[0])
            print RMSDstring
        print "Script has finished"
cmd.extend('RMSDmatrix',RMSDmatrix)
```

### 9.3 Docking Performance of Individual Crystal Structures of Renin

**Tab. 16:** Docking performance of individual crystal structures in our test system in terms of binding pose prediction (expressed as ratio of binding poses with RMSD <3 Å), screening utility (AUC ROC and BEDROC (with  $\alpha = 20$ )), scoring accuracy ( $R^2$ ) and chemotype-corrected AUC ROC.

| PDB  | Binding Pose<br><3 Å [%] | AUC ROC | BEDROC | $R^2$ | Chemotype-<br>corrected AUC<br>ROC |
|------|--------------------------|---------|--------|-------|------------------------------------|
| 1BIL | 31.1                     | 0.607   | 0.094  | 0.069 | 0.655                              |
| 1HRN | 22.2                     | 0.600   | 0.133  | 0.035 | 0.587                              |
| 1RNE | 22.2                     | 0.667   | 0.166  | 0.016 | 0.681                              |
| 2BKT | 4.4                      | 0.464   | 0.150  | 0.041 | 0.491                              |
| 2FS4 | 11.1                     | 0.642   | 0.111  | 0.028 | 0.661                              |
| 2G1O | 20.0                     | 0.612   | 0.165  | 0.011 | 0.582                              |
| 2G1R | 26.7                     | 0.640   | 0.197  | 0.001 | 0.620                              |
| 2G1Y | 26.7                     | 0.723   | 0.251  | 0.007 | 0.718                              |
| 2IKO | 8.9                      | 0.639   | 0.099  | 0.053 | 0.641                              |
| 2IKU | 22.2                     | 0.658   | 0.201  | 0.014 | 0.647                              |
| 2IL2 | 42.2                     | 0.746   | 0.275  | 0.118 | 0.736                              |
| 2V0Z | 46.7                     | 0.789   | 0.395  | 0.171 | 0.761                              |
| 2V10 | 35.6                     | 0.744   | 0.347  | 0.242 | 0.741                              |
| 2V11 | 35.6                     | 0.748   | 0.327  | 0.226 | 0.730                              |
| 2V12 | 40.0                     | 0.783   | 0.417  | 0.244 | 0.781                              |
| 2V13 | 42.2                     | 0.724   | 0.276  | 0.125 | 0.719                              |
| 2V16 | 44.4                     | 0.724   | 0.307  | 0.207 | 0.725                              |
| 3G6Z | 13.3                     | 0.545   | 0.105  | 0.131 | 0.571                              |
| 3G70 | 13.3                     | 0.577   | 0.130  | 0.108 | 0.602                              |
| 3G72 | 17.8                     | 0.574   | 0.089  | 0.086 | 0.579                              |
| 3GW5 | 51.1                     | 0.736   | 0.313  | 0.225 | 0.733                              |
| 3KM4 | 42.2                     | 0.727   | 0.249  | 0.074 | 0.700                              |
| 3OOT | 20.0                     | 0.738   | 0.300  | 0.089 | 0.726                              |
| 3OQF | 35.6                     | 0.742   | 0.244  | 0.197 | 0.737                              |
| 3OQK | 28.9                     | 0.698   | 0.183  | 0.106 | 0.688                              |
| 3Q3T | 55.6                     | 0.757   | 0.350  | 0.088 | 0.750                              |
| 3Q4B | 28.9                     | 0.654   | 0.241  | 0.148 | 0.665                              |
| 3Q5H | 33.3                     | 0.657   | 0.279  | 0.159 | 0.685                              |
| 3SFC | 11.1                     | 0.613   | 0.175  | 0.033 | 0.634                              |
| 3VSX | 24.4                     | 0.746   | 0.288  | 0.212 | 0.747                              |
| 3VUC | 26.7                     | 0.710   | 0.284  | 0.193 | 0.701                              |
| 3VYD | 37.8                     | 0.739   | 0.320  | 0.332 | 0.728                              |
| 3VYE | 37.8                     | 0.738   | 0.281  | 0.174 | 0.713                              |
| 3VYF | 22.2                     | 0.709   | 0.380  | 0.273 | 0.713                              |
| 4GJ6 | 15.6                     | 0.684   | 0.177  | 0.139 | 0.699                              |
| 4GJ7 | 35.6                     | 0.766   | 0.351  | 0.100 | 0.786                              |
| 4GJD | 17.8                     | 0.618   | 0.159  | 0.142 | 0.615                              |
| 4Q1N | 24.4                     | 0.748   | 0.326  | 0.226 | 0.748                              |
| 4RYG | 40.0                     | 0.813   | 0.386  | 0.098 | 0.800                              |
| 4RZ1 | 42.2                     | 0.737   | 0.252  | 0.059 | 0.723                              |
| 5KOQ | 40.0                     | 0.832   | 0.381  | 0.111 | 0.835                              |
| 5KOS | 35.6                     | 0.771   | 0.247  | 0.143 | 0.769                              |



*Appendix*

---

| PDB  | Binding Pose<br><3 Å [%] | AUC ROC | BEDROC | R <sup>2</sup> | Chemotype-<br>corrected AUC<br>ROC |
|------|--------------------------|---------|--------|----------------|------------------------------------|
| 5SXN | 6.7                      | 0.576   | 0.113  | 0.000          | 0.570                              |
| 5SY3 | 15.6                     | 0.608   | 0.064  | 0.026          | 0.605                              |
| 5SZ9 | 20.0                     | 0.571   | 0.084  | 0.064          | 0.550                              |

---

## 9.4 Docking Performance of Ensembles of Crystal Structures Constructed by Hierarchical Clustering

**Tab. 17:** Docking performance of ensembles of crystal structures constructed by hierarchical clustering with average linkage.

| Ensemble size | PDB  | Binding Pose <3 Å [%] | AUC ROC | BEDROC | R <sup>2</sup> | Chemotype-corrected AUC ROC |
|---------------|--|-----------------------|---------|--------|----------------|-----------------------------|
| 2             | 3G70<br>4RZ1   | 42.2                  | 0.679   | 0.213  | 0.107          | 0.682                       |
| 3             | 4GJ6<br>4RZ1<br>3G70   | 35.6                  | 0.693   | 0.214  | 0.142          | 0.702                       |
| 4             | 2G1R<br>3OQF<br>4GJ6<br>3G70                                 | 46.7                  | 0.755   | 0.259  | 0.156          | 0.750                       |
| 5             | 3G70<br>2BKT<br>2G1R<br>3OQF<br>4GJ6                         | 33.3                  | 0.582   | 0.173  | 0.069          | 0.594                       |
| 6             | 2BKT<br>2FS4<br>3G70<br>2G1R<br>3OQF<br>4GJ6                 | 28.9                  | 0.583   | 0.178  | 0.082          | 0.597                       |
| 7             | 2V10<br>3OQF<br>2BKT<br>2FS4<br>3G70<br>2G1R<br>4GK6         | 35.6                  | 0.650   | 0.184  | 0.154          | 0.662                       |
| 8             | 4RZ1<br>3Q5J<br>2V10<br>2BKT<br>2FS4<br>3G70<br>2G1R<br>4GJ6 | 42.2                  | 0.673   | 0.263  | 0.203          | 0.708                       |

## 9.5 Docking Performance of Ensembles of Crystal Structures Constructed by K-means Clustering

**Tab. 18:** Docking performance of ensembles of crystal structures constructed by *k*-means clustering.

| Ensemble size | PDB  | Binding Pose <3 Å [%] | AUC ROC | BEDROC | R <sup>2</sup> | Chemotype-corrected AUC ROC |
|---------------|--|-----------------------|---------|--------|----------------|-----------------------------|
| 2             | 3G72<br>1HRN   | 22.2                  | 0.610   | 0.167  | 0.077          | 0.609                       |
| 3             | 3G72<br>1HRN<br>5KOQ   | 42.2                  | 0.728   | 0.293  | 0.137          | 0.728                       |
| 4             | 3OQF<br>2V0Z<br>2G1Y<br>3G72                                 | 60.0                  | 0.819   | 0.354  | 0.279          | 0.806                       |
| 5             | 2V16<br>3OQF<br>3G72<br>3OQK<br>2G1R                         | 55.6                  | 0.802   | 0.352  | 0.212          | 0.793                       |
| 6             | 2V16<br>3OQK<br>3G70<br>3OQF<br>2BKT<br>2G1R                 | 51.1                  | 0.645   | 0.197  | 0.151          | 0.665                       |
| 7             | 2BKT<br>5SY3<br>3G70<br>3OQF<br>2G1R<br>3Q4B<br>4RZ1         | 40.0                  | 0.641   | 0.249  | 0.091          | 0.667                       |
| 8             | 3OQF<br>3G70<br>3VYD<br>2G1R<br>2V11<br>4RZ1<br>3Q4B<br>2BKT | 51.1                  | 0.687   | 0.256  | 0.205          | 0.712                       |

## 9.6 Docking Performance of Ensembles Constructed from MD-derived Structures

**Tab. 19:** Docking performance in terms of binding pose prediction (expressed as ratio of binding poses with RMSD < 3 Å) of ensembles constructed from MD-derived structures by *k*-means clustering. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| <i>k</i> -means        | ensemble size |      |      |      |      |      |      |      |
|------------------------|---------------|------|------|------|------|------|------|------|
|                        | 2             | 3    | 4    | 5    | 6    | 7    | 8    |      |
| Binding Pose < 3 Å [%] | T1            | 22.2 | 20.0 | 24.4 | 22.2 | 15.6 | 13.3 | 11.1 |
|                        | T2            | 4.4  | 2.2  | 8.9  | 20.0 | 17.8 | 24.4 | 17.8 |
|                        | T3            | 17.8 | 22.2 | 20.0 | 17.8 | 22.2 | 13.3 | 22.2 |
|                        | T4            | 6.7  | 4.4  | 4.4  | 11.1 | 11.1 | 4.4  | 4.4  |
|                        | T5            | 17.8 | 11.1 | 13.3 | 8.9  | 17.8 | 11.1 | 8.9  |
|                        | T6            | 11.1 | 15.6 | 4.4  | 6.7  | 11.1 | 13.3 | 11.1 |
|                        | T7            | 6.7  | 15.6 | 17.8 | 17.8 | 15.6 | 24.4 | 17.8 |
|                        | T8            | 15.6 | 4.4  | 15.6 | 17.8 | 13.3 | 11.1 | 13.3 |
|                        | T9            | 17.8 | 8.9  | 20.0 | 17.8 | 17.8 | 17.8 | 13.3 |
|                        | T10           | 2.2  | 6.7  | 6.7  | 4.4  | 6.7  | 11.1 | 4.4  |
| Mean                   | 12.2          | 11.1 | 13.6 | 14.4 | 14.9 | 14.4 | 12.4 |      |
| STD                    | 6.9           | 7.0  | 7.1  | 6.1  | 4.4  | 6.2  | 5.8  |      |

**Tab. 20:** Docking performance in terms of binding pose prediction (expressed as ratio of binding poses with RMSD < 3 Å) of ensembles constructed from MD-derived structures by hierarchical clustering with average linkage. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| hierarchical           | ensemble size |      |      |      |      |      |      |      |
|------------------------|---------------|------|------|------|------|------|------|------|
|                        | 2             | 3    | 4    | 5    | 6    | 7    | 8    |      |
| Binding Pose < 3 Å [%] | T1            | 11.1 | 11.1 | 11.1 | 17.8 | 20.0 | 20.0 | 20.0 |
|                        | T2            | 4.4  | 6.7  | 8.9  | 13.3 | 8.9  | 6.7  | 4.4  |
|                        | T3            | 2.2  | 4.4  | 2.2  | 2.2  | 8.9  | 8.9  | 11.1 |
|                        | T4            | 4.4  | 4.4  | 4.4  | 4.4  | 6.7  | 6.7  | 6.7  |
|                        | T5            | 26.7 | 22.2 | 11.1 | 17.8 | 6.7  | 6.7  | 4.4  |
|                        | T6            | 2.2  | 2.2  | 2.2  | 4.4  | 4.4  | 11.1 | 8.9  |
|                        | T7            | 13.3 | 13.3 | 8.9  | 8.9  | 6.7  | 6.7  | 8.9  |
|                        | T8            | 24.4 | 15.6 | 17.8 | 17.8 | 8.9  | 8.9  | 15.6 |
|                        | T9            | 2.2  | 8.9  | 8.9  | 8.9  | 8.9  | 11.1 | 11.1 |
|                        | T10           | 2.2  | 0.0  | 4.4  | 4.4  | 26.7 | 22.2 | 22.2 |
| Mean                   | 9.3           | 8.9  | 8.0  | 10.0 | 10.7 | 10.9 | 11.3 |      |
| STD                    | 9.4           | 6.8  | 4.8  | 6.2  | 7.0  | 5.7  | 6.2  |      |

**Tab. 21:** Docking performance (in terms of AUC ROC) of ensembles constructed from MD-derived structures by *k*-means clustering. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| <b>k-means</b> |       | <b>ensemble size</b> |       |       |       |       |       |       |
|----------------|-------|----------------------|-------|-------|-------|-------|-------|-------|
|                |       | 2                    | 3     | 4     | 5     | 6     | 7     | 8     |
| AUC ROC        | T1    | 0.646                | 0.648 | 0.647 | 0.744 | 0.710 | 0.620 | 0.626 |
|                | T2    | 0.517                | 0.509 | 0.529 | 0.622 | 0.620 | 0.690 | 0.632 |
|                | T3    | 0.653                | 0.631 | 0.662 | 0.618 | 0.656 | 0.673 | 0.649 |
|                | T4    | 0.581                | 0.654 | 0.664 | 0.622 | 0.627 | 0.647 | 0.649 |
|                | T5    | 0.636                | 0.652 | 0.624 | 0.676 | 0.649 | 0.630 | 0.599 |
|                | T6    | 0.630                | 0.691 | 0.629 | 0.647 | 0.674 | 0.660 | 0.657 |
|                | T7    | 0.637                | 0.662 | 0.708 | 0.684 | 0.640 | 0.720 | 0.636 |
|                | T8    | 0.663                | 0.650 | 0.653 | 0.673 | 0.690 | 0.718 | 0.721 |
|                | T9    | 0.711                | 0.570 | 0.622 | 0.712 | 0.668 | 0.700 | 0.647 |
|                | T10   | 0.607                | 0.599 | 0.575 | 0.609 | 0.656 | 0.636 | 0.577 |
|                | Mean  | 0.628                | 0.627 | 0.631 | 0.661 | 0.659 | 0.669 | 0.639 |
| STD            | 0.052 | 0.053                | 0.050 | 0.045 | 0.028 | 0.037 | 0.038 |       |

**Tab. 22:** Docking performance (in terms of AUC ROC) of ensembles constructed from MD-derived structures by hierarchical clustering with average linkage. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| <b>hierarchical</b> |       | <b>ensemble size</b> |       |       |       |       |       |       |
|---------------------|-------|----------------------|-------|-------|-------|-------|-------|-------|
|                     |       | 2                    | 3     | 4     | 5     | 6     | 7     | 8     |
| AUC ROC             | T1    | 0.622                | 0.627 | 0.638 | 0.650 | 0.655 | 0.655 | 0.655 |
|                     | T2    | 0.587                | 0.590 | 0.647 | 0.621 | 0.628 | 0.663 | 0.675 |
|                     | T3    | 0.589                | 0.509 | 0.525 | 0.566 | 0.577 | 0.578 | 0.584 |
|                     | T4    | 0.553                | 0.573 | 0.618 | 0.617 | 0.631 | 0.630 | 0.627 |
|                     | T5    | 0.676                | 0.651 | 0.608 | 0.625 | 0.587 | 0.564 | 0.570 |
|                     | T6    | 0.556                | 0.575 | 0.595 | 0.575 | 0.587 | 0.623 | 0.638 |
|                     | T7    | 0.650                | 0.657 | 0.539 | 0.538 | 0.606 | 0.606 | 0.616 |
|                     | T8    | 0.685                | 0.620 | 0.613 | 0.612 | 0.647 | 0.636 | 0.639 |
|                     | T9    | 0.664                | 0.658 | 0.610 | 0.593 | 0.598 | 0.607 | 0.605 |
|                     | T10   | 0.613                | 0.584 | 0.606 | 0.617 | 0.710 | 0.657 | 0.654 |
|                     | Mean  | 0.620                | 0.604 | 0.600 | 0.601 | 0.623 | 0.622 | 0.626 |
| STD                 | 0.048 | 0.047                | 0.039 | 0.033 | 0.041 | 0.033 | 0.033 |       |

**Tab. 23:** Docking performance (in terms of BEDROC) of ensembles constructed from MD-derived structures by *k*-means clustering. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| <b>k-means</b> |       | <b>ensemble size</b> |       |       |       |       |       |       |
|----------------|-------|----------------------|-------|-------|-------|-------|-------|-------|
|                |       | 2                    | 3     | 4     | 5     | 6     | 7     | 8     |
| BEDROC         | T1    | 0.115                | 0.124 | 0.204 | 0.232 | 0.187 | 0.101 | 0.094 |
|                | T2    | 0.082                | 0.083 | 0.124 | 0.157 | 0.145 | 0.163 | 0.151 |
|                | T3    | 0.140                | 0.154 | 0.159 | 0.135 | 0.161 | 0.135 | 0.152 |
|                | T4    | 0.065                | 0.120 | 0.157 | 0.137 | 0.138 | 0.179 | 0.177 |
|                | T5    | 0.065                | 0.134 | 0.111 | 0.146 | 0.136 | 0.111 | 0.102 |
|                | T6    | 0.097                | 0.242 | 0.149 | 0.087 | 0.127 | 0.106 | 0.096 |
|                | T7    | 0.225                | 0.173 | 0.168 | 0.188 | 0.167 | 0.193 | 0.130 |
|                | T8    | 0.191                | 0.168 | 0.174 | 0.138 | 0.146 | 0.220 | 0.239 |
|                | T9    | 0.199                | 0.093 | 0.158 | 0.188 | 0.164 | 0.131 | 0.117 |
|                | T10   | 0.094                | 0.111 | 0.109 | 0.115 | 0.057 | 0.074 | 0.053 |
|                | Mean  | 0.127                | 0.140 | 0.151 | 0.152 | 0.143 | 0.141 | 0.131 |
| STD            | 0.058 | 0.047                | 0.029 | 0.041 | 0.035 | 0.046 | 0.052 |       |

**Tab. 24:** Docking performance (in terms of BEDROC) of ensembles constructed from MD-derived structures by hierarchical clustering with average linkage. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| <b>hierarchical</b> |       | <b>ensemble size</b> |       |       |       |       |       |       |
|---------------------|-------|----------------------|-------|-------|-------|-------|-------|-------|
|                     |       | 2                    | 3     | 4     | 5     | 6     | 7     | 8     |
| BEDROC              | T1    | 0.121                | 0.137 | 0.139 | 0.134 | 0.145 | 0.137 | 0.136 |
|                     | T2    | 0.112                | 0.112 | 0.136 | 0.136 | 0.153 | 0.179 | 0.180 |
|                     | T3    | 0.110                | 0.076 | 0.076 | 0.101 | 0.078 | 0.082 | 0.084 |
|                     | T4    | 0.077                | 0.100 | 0.139 | 0.139 | 0.134 | 0.142 | 0.140 |
|                     | T5    | 0.156                | 0.118 | 0.119 | 0.114 | 0.110 | 0.106 | 0.103 |
|                     | T6    | 0.049                | 0.071 | 0.118 | 0.101 | 0.102 | 0.103 | 0.120 |
|                     | T7    | 0.110                | 0.107 | 0.047 | 0.047 | 0.064 | 0.064 | 0.064 |
|                     | T8    | 0.144                | 0.090 | 0.129 | 0.133 | 0.156 | 0.153 | 0.161 |
|                     | T9    | 0.098                | 0.098 | 0.059 | 0.033 | 0.050 | 0.058 | 0.058 |
|                     | T10   | 0.136                | 0.088 | 0.095 | 0.115 | 0.181 | 0.094 | 0.090 |
|                     | Mean  | 0.111                | 0.100 | 0.106 | 0.105 | 0.117 | 0.112 | 0.114 |
| STD                 | 0.032 | 0.020                | 0.034 | 0.037 | 0.044 | 0.040 | 0.041 |       |

**Tab. 25:** Docking performance (in terms of  $R^2$ ) of ensembles constructed from MD-derived structures by  $k$ -means clustering. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| <b>k-means</b> |       | <b>ensemble size</b> |       |       |       |       |       |       |
|----------------|-------|----------------------|-------|-------|-------|-------|-------|-------|
|                |       | 2                    | 3     | 4     | 5     | 6     | 7     | 8     |
| $R^2$          | T1    | 0.036                | 0.046 | 0.168 | 0.280 | 0.288 | 0.235 | 0.145 |
|                | T2    | 0.118                | 0.233 | 0.251 | 0.112 | 0.109 | 0.283 | 0.281 |
|                | T3    | 0.247                | 0.076 | 0.101 | 0.157 | 0.271 | 0.251 | 0.268 |
|                | T4    | 0.296                | 0.279 | 0.214 | 0.279 | 0.212 | 0.156 | 0.160 |
|                | T5    | 0.225                | 0.107 | 0.038 | 0.133 | 0.221 | 0.200 | 0.181 |
|                | T6    | 0.308                | 0.250 | 0.273 | 0.285 | 0.260 | 0.172 | 0.124 |
|                | T7    | 0.222                | 0.267 | 0.294 | 0.228 | 0.266 | 0.269 | 0.171 |
|                | T8    | 0.235                | 0.308 | 0.235 | 0.199 | 0.307 | 0.313 | 0.296 |
|                | T9    | 0.134                | 0.067 | 0.179 | 0.103 | 0.144 | 0.150 | 0.105 |
|                | T10   | 0.249                | 0.146 | 0.040 | 0.100 | 0.097 | 0.137 | 0.080 |
|                | Mean  | 0.207                | 0.178 | 0.179 | 0.187 | 0.217 | 0.217 | 0.181 |
| STD            | 0.085 | 0.100                | 0.093 | 0.076 | 0.076 | 0.062 | 0.076 |       |

**Tab. 26:** Docking performance (in terms of  $R^2$ ) of ensembles constructed from MD-derived structures by hierarchical clustering with average linkage. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| <b>hierarchical</b> |       | <b>ensemble size</b> |       |       |       |       |       |       |
|---------------------|-------|----------------------|-------|-------|-------|-------|-------|-------|
|                     |       | 2                    | 3     | 4     | 5     | 6     | 7     | 8     |
| $R^2$               | T1    | 0.088                | 0.100 | 0.144 | 0.105 | 0.115 | 0.097 | 0.096 |
|                     | T2    | 0.162                | 0.190 | 0.271 | 0.272 | 0.127 | 0.047 | 0.081 |
|                     | T3    | 0.091                | 0.002 | 0.005 | 0.076 | 0.100 | 0.064 | 0.068 |
|                     | T4    | 0.094                | 0.131 | 0.012 | 0.012 | 0.030 | 0.041 | 0.045 |
|                     | T5    | 0.109                | 0.181 | 0.133 | 0.114 | 0.123 | 0.138 | 0.147 |
|                     | T6    | 0.124                | 0.160 | 0.280 | 0.211 | 0.271 | 0.297 | 0.220 |
|                     | T7    | 0.119                | 0.146 | 0.093 | 0.084 | 0.184 | 0.184 | 0.181 |
|                     | T8    | 0.150                | 0.141 | 0.321 | 0.319 | 0.294 | 0.293 | 0.284 |
|                     | T9    | 0.058                | 0.058 | 0.033 | 0.055 | 0.072 | 0.076 | 0.075 |
|                     | T10   | 0.041                | 0.083 | 0.124 | 0.172 | 0.174 | 0.133 | 0.127 |
|                     | Mean  | 0.104                | 0.119 | 0.142 | 0.142 | 0.149 | 0.137 | 0.132 |
| STD                 | 0.038 | 0.059                | 0.115 | 0.099 | 0.083 | 0.095 | 0.076 |       |

**Tab. 27:** Docking performance (in terms of chemotype-corrected AUC ROC) of ensembles constructed from MD-derived structures by *k*-means clustering. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| <b>k-means</b>              |       | <b>ensemble size</b> |          |          |          |          |          |          |
|-----------------------------|-------|----------------------|----------|----------|----------|----------|----------|----------|
|                             |       | <b>2</b>             | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> |
| Chemotype-corrected AUC ROC | T1    | 0.668                | 0.661    | 0.656    | 0.750    | 0.725    | 0.628    | 0.611    |
|                             | T2    | 0.503                | 0.502    | 0.549    | 0.641    | 0.637    | 0.710    | 0.659    |
|                             | T3    | 0.684                | 0.657    | 0.691    | 0.655    | 0.675    | 0.710    | 0.663    |
|                             | T4    | 0.603                | 0.658    | 0.701    | 0.645    | 0.653    | 0.663    | 0.665    |
|                             | T5    | 0.642                | 0.651    | 0.627    | 0.681    | 0.636    | 0.639    | 0.605    |
|                             | T6    | 0.654                | 0.702    | 0.670    | 0.677    | 0.692    | 0.675    | 0.684    |
|                             | T7    | 0.631                | 0.667    | 0.719    | 0.701    | 0.670    | 0.726    | 0.658    |
|                             | T8    | 0.662                | 0.665    | 0.654    | 0.680    | 0.709    | 0.732    | 0.732    |
|                             | T9    | 0.732                | 0.601    | 0.616    | 0.710    | 0.675    | 0.707    | 0.646    |
|                             | T10   | 0.625                | 0.597    | 0.554    | 0.612    | 0.682    | 0.650    | 0.592    |
|                             | Mean  | 0.638                | 0.644    | 0.658    | 0.686    | 0.674    | 0.683    | 0.657    |
| STD                         | 0.058 | 0.060                | 0.074    | 0.052    | 0.028    | 0.036    | 0.043    |          |

**Tab. 28:** Docking performance (in terms of chemotype-corrected AUC ROC) of ensembles constructed from MD-derived structures by hierarchical clustering with average linkage. T1-T10 reflect the ten trajectories of MD simulations performed in this work. Their mean value and standard deviation (STD) is calculated.

| <b>hierarchical</b>         |       | <b>ensemble size</b> |          |          |          |          |          |          |
|-----------------------------|-------|----------------------|----------|----------|----------|----------|----------|----------|
|                             |       | <b>2</b>             | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> |
| Chemotype-corrected AUC ROC | T1    | 0.609                | 0.622    | 0.632    | 0.667    | 0.670    | 0.671    | 0.671    |
|                             | T2    | 0.595                | 0.605    | 0.660    | 0.642    | 0.654    | 0.676    | 0.693    |
|                             | T3    | 0.625                | 0.508    | 0.522    | 0.600    | 0.596    | 0.595    | 0.608    |
|                             | T4    | 0.586                | 0.610    | 0.636    | 0.636    | 0.644    | 0.642    | 0.640    |
|                             | T5    | 0.685                | 0.648    | 0.611    | 0.622    | 0.592    | 0.570    | 0.570    |
|                             | T6    | 0.557                | 0.580    | 0.605    | 0.593    | 0.611    | 0.642    | 0.656    |
|                             | T7    | 0.648                | 0.665    | 0.568    | 0.567    | 0.633    | 0.632    | 0.647    |
|                             | T8    | 0.695                | 0.626    | 0.638    | 0.636    | 0.655    | 0.647    | 0.668    |
|                             | T9    | 0.659                | 0.653    | 0.608    | 0.605    | 0.612    | 0.617    | 0.616    |
|                             | T10   | 0.633                | 0.608    | 0.635    | 0.646    | 0.720    | 0.677    | 0.674    |
|                             | Mean  | 0.634                | 0.621    | 0.624    | 0.619    | 0.635    | 0.639    | 0.650    |
| STD                         | 0.045 | 0.050                | 0.057    | 0.029    | 0.039    | 0.034    | 0.040    |          |



## 10 Danksagung

Ich danke:

**Prof. Dr. Bernd Meyer:** Für die Betreuung dieser Arbeit und alle gewährten Freiheiten.

**Prof. Dr. Christian B. W. Stark:** Für die Übernahme des Zweitgutachtens.

**Prof. Dr. Thomas Peters:** Für die Bereitstellung von GTB-Enzym.

**Prof. Dr. Edzard Spillner:** Für die Bereitstellung von FUT8-Enzym.

**Dr. Thomas Hackl:** Für die Unterstützung in NMR-Fragen.

den verbliebenen und ehemaligen **AKBM's**.

meinen **Eltern & Familie** für die Unterstützung.

„Für die Gang“:

**Melissa Bärenfänger & Anna-Lena Giertz:** Für alles. Alle Lady-Dinge, schöne Stunden beim Kochen, im Wildpark oder Grillen im Stadtpark.

**Raphael „Ralf“ Schuster:** Für alle Mittwoch, SCP-Auswärtsfahrten und dafür, dass du mich beim Tennis immer gewinnen lässt.

**Jan Seddig:** Für eine anhaltende Freundschaft - auch über Kontinente hinweg.

**Simone Wolter:** Für das *dolce vita* am Würz-Main und die bedingungslose Unterstützung.

## **11 Affidavits**

### **Erklärung**

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig angefertigt und keine anderen als die von mir angegebenen Hilfsmittel und Quellen verwendet habe. Ich versichere weiterhin, dass die vorliegende Dissertation weder in gleicher noch in veränderter Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Hamburg, den 22.08.2019

Claas Strecker