

Identification of Ligands with Tailored Selectivity: Strategies & Application

Dissertation

zur Erlangung des Doktorgrades der
Naturwissenschaften
(Dr. rer. nat.)

dem

Fachbereich Pharmazie der
Philipps-Universität Marburg
vorgelegt

von

Denis Schmidt

aus Darmstadt

Marburg/Lahn 2015

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Dr. Peter Kolb im Zeitraum von April 2011 bis Mai 2015 am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg durchgeführt.

Erstgutachter: Dr. Peter Kolb

Zweitgutachter: Prof. Dr. Gerhard Klebe

Eingereicht am 28.05.2015

Tag der mündlichen Prüfung: 13.07.2015

Hochschulkennziffer: 1180

Meiner Familie

Abstract

Docking is an established method in computer-aided drug design that tries to predict the binding and non-binding of a small molecule to a protein target. The docking against a single structure can be considered as established, in contrast to the docking to several structures in order to elucidate the binding profile of a ligand. This work explores the applicability of docking in such a scenario.

In a case study, ligands binding either to one or both receptors are identified for CXCR3 and CXCR4, a pair of chemokine GPCRs. This application proves the feasibility of docking to select molecules with tailored selectivity and yields ligands with excellent binding affinities.

The subsequent studies each utilizes different input structures of a common protein, also chemokine receptors. While again new ligands can be identified by docking, the complexity of GPCRs and functional GPCR assays is discussed, which can impair the identification of low-efficacy ligands. Finally, the limits of multi-target docking are explored by docking to a newly created dataset, containing several hundred kinase structures. The results indicate that the automatic prediction of very large binding profiles is beyond current possibilities. The prediction accuracy of docking can be improved by normalizing the docking scores across multiple ligands and multiple targets, which leads to the idea of “protein decoys” that might improve the robustness of future docking applications.

Contents

List of Figures	xi
List of Tables	xiii
Vorwort des Autors (deutsch)	xv
1 Introduction	1
2 Docking to GPCRs	3
2.1 From an Idea to a Lead Structure	3
2.2 Docking – Theory and Technique	4
2.3 G Protein-Coupled Receptors	6
2.3.1 Docking identifies novel GPCR ligands	9
2.3.2 Homology Models of GPCRs	11
2.3.3 Distinct Conformations Identify Different Ligands	11
2.4 Discussion	12
3 CXCR3-CXCR4 Ligands with Tailored Selectivity	15
3.1 Introduction	15
3.2 Results and Discussion	20
3.2.1 CXCR3 homology model	20
3.2.2 Docking	23
3.2.3 [³⁵ S]GTPγS Assay	26
3.2.4 Aggregator counter screening	34
3.2.5 Summary	36
3.3 Methods	40
3.3.1 Homology Modeling	40

3.3.2	Model refinement	40
3.3.3	Optimization of transmembrane binding pocket . . .	41
3.3.4	Docking & Re-Ranking	41
3.3.5	Ligand selection and validation	43
3.3.6	HEK membrane preparations	44
3.3.7	[³⁵ S]GTPγS incorporation assay	45
3.3.8	Functional study analysis	46
3.3.9	Ligand similarities and properties	47
3.3.10	Dynamic Light Scattering	47
3.3.11	Endothiapepsin Assay	47
4	Tool Compounds Targeting Zebrafish CXCR4 Isoforms	49
4.1	Introduction	49
4.2	Results	52
4.2.1	Homology Models and Docking	52
4.2.2	In vivo assay	53
4.3	Discussion	55
4.4	Material & Methods	59
4.4.1	Homology Modeling & Refinement	59
4.4.2	Docking	60
4.4.3	Compound Acquisition	60
4.4.4	<i>In vivo</i> screening	60
5	Docking to Multiple Structures of CCR5	63
5.1	Introduction	63
5.2	Results & Discussion	65
5.2.1	Homology Modeling of CCR5	65
5.2.2	Reshaping the CCR5 binding site	70
5.2.3	Virtual Screening against different CCR5 structures	75
5.2.4	Compound Validation	78
5.3	Summary & Outlook	86
5.4	Material & Methods	89
5.4.1	Structure Modeling	89

5.4.2	Model optimization	90
5.4.3	Binding Site Reshaping	90
5.4.4	Docking	91
5.4.5	Compound Acquisition	91
5.4.6	Second Messenger Assay	92
5.4.7	Error Propagation	93
6	Large-scale docking to kinases	95
6.1	Introduction	95
6.2	Results & Discussion	100
6.3	Summary & Perspective	111
6.4	Materials & Methods	113
6.4.1	Structure Selection and Preparation	113
6.4.2	Docking	113
6.4.3	Evaluation	114
6.4.4	Visualization	114
7	Summary	117
7.1	Summary (english)	117
7.2	Zusammenfassung (deutsch)	120
	Bibliography	125
	Acronyms	139
	About the author	143
	Publications	143
	Curriculum Vitae (deutsch)	145
	Eidesstattliche Erklärung	147

List of Figures

2.1	Scheme of a Docking Setup	5
2.2	Prototypic Structure of a GPCR	7
2.3	Betablocker-Pharmacophore	8
2.4	Binding Site of the β_2 -adrenergic Receptor	9
2.5	Hits from Screening to β_2 AR	10
3.1	Charge Distributions of CXCR3 and CXCR4 Binders	18
3.2	Physico-Chemical Properties of CXCR3 and CXCR4 Ligands	19
3.3	Ramachandran Plot of CXCR3 Homology Model	21
3.4	ROC Curve of Test Set Docked to CXCR3 Model	22
3.5	Overlay of CXCR3 Model and CXCR4 Crystal Structure	23
3.6	Comparison of Docking Poses in CXCR3 and CXCR4	25
3.7	Representative Dose-Effect Curves from $[^{35}\text{S}]\text{GTP}\gamma\text{S}$ Assay	32
3.8	RAMX3 Structure	32
3.9	Dose-Response Curves of Radioligand Displacement Assay	34
3.10	Effect of Substances in Absence of Endogenous Ligand	35
3.11	Results of DLS Experiments	37
3.12	Results of Endothiapepsin Assay	38
3.13	Comparison of Binding-Efficiency Indices	39
3.14	Plot of Rescoring Function	44
4.1	Comparison of CXCR4 Models and Xray Structure	54
4.2	Microscope Images of Treated Larvae	56
5.1	Known CCR5 Inhibitors	64
5.2	Overlay of CCR5 Crystal Structure and Homology Model	66

5.3	Turn Structure in CCR5 Crystal Structure	68
5.4	Sampled Ligand Conformations of TAK-779	71
5.5	Side View of Crystallized and Reshaped CCR5 Binding Site	72
5.6	Charge Distribution of Top-Ranked Molecules	76
5.7	Example of Dimissed Docking Pose	76
5.8	Results of Primary Screen	80
5.9	Sorted Results of Primary Screen	82
5.10	Selected Dose-Response Curves	84
6.1	Overview of the Protein Kinase Domain Fold	97
6.2	Comparison of Kinase Activity States	99
6.3	Number of Crystal Structures Per Kinase	101
6.4	Heatmap of Raw Docking Scores	103
6.5	Heatmap of Docking Scores for Selected Kinases	104
6.6	Heatmap of mMASC-Normalized Docking Scores	106
6.7	Number of Ligands per Kinase in Karaman Dataset	107
6.8	Performance of Docking Approaches (Structure Level)	108
6.9	Performance of Approaches (Kinase Level)	109

List of Tables

3.1	Sizes of Selected GPCR Binding Sites	17
3.2	Docking Ranks of Tested Compounds	26
3.3	Comprehensive Results for Identified Compounds	27
3.4	Chemical Structures of Identified Compounds	28
3.5	Chemical Structures of Inactive Molecules	29
3.6	Radioligand Displacement Data	31
3.7	Comprehensive Results for Second Set of Compounds	33
3.8	Residues Relevant in Model Refinement	42
3.9	Ligands used for Model Refinement	43
4.1	Structure Quality Assessment of Models and Template . . .	53
4.2	Compounds Selected for Testing	61
5.1	Chemical Structures of Tested Compounds	79
5.2	Comprehensive Results of Dose-Response Screening	83
5.3	Templates of CCR5 Homology Model	89
5.4	Raw Control Readouts in First Screening Round	93
5.5	Formulas for Propagation of Uncertainty	93

Vorwort des Autors (deutsch)

Mit dem Abschluss des Studiums schließt man gleichzeitig ein wichtiges Kapitel seines Lebens ab. Doch kaum ist das geschafft, findet man sich selbst vor der Frage nach der Promotion. Zwar stand für mich das “dass” fest, aber das “wo” ist immer eine gänzlich andere Frage. Wohl fühlen muss man sich am Ort der Promotion, wobei “Ort” sowohl die Geographie als auch das direkte Arbeitsumfeld einschließt. Dieses Wohlfühlen ist wichtig, schließlich verbringt man die nächsten drei Jahre dort. Zugegeben, meistens sind es mehr, so auch in meinem Fall. Die Arbeitsgruppe der Wahl muss außerdem gut vernetzt sein und hohen wissenschaftlichen Ansprüchen genügen, man möchte ja nicht “irgendwo” promovieren. Und selbstredend muss der Gruppenleiter/Betreuer die Sorgen und Frustrationen des Doktoranden verstehen, auf Augenhöhe diskutieren können und seinen Schützling anleiten ohne die eigenen Ideen des selbigen aus den Augen zu verlieren. Achja, der thematische Schwerpunkt der Promotion sollte einen auch ein wenig interessieren. Kurzum, vor der Promotion gilt es, ein nicht-triviales Optimierungsproblem zu lösen.

Bei der Lösung halfen mir unter anderem **Mireille Krier** und **Paul Czdrowski**. **Mireille** war zu damaligem Zeitpunkt die Betreuerin meiner Diplomarbeit bei Merck Serono und Paul ein Kollege, der selbst hier in Marburg promoviert hatte. Beide gaben mir dankenswerterweise Tipps, welche Arbeitskreise für mich interessant sein könnten und welche Konferenzen ich besuchen könnte, um jemanden aus diesen Gruppen zu treffen. Nur so konnte ich **Prof. Gerhard Klebe** kennen lernen. Die Beiträge von **Prof. Klebe** im Bereich von “computer-aided drug design” kannte ich aus dem Studium und ich konnte mir gut vorstellen in dieser Gruppe zu promovieren. Leider war zu dieser Zeit keine Stelle direkt in seiner Gruppe frei, aber

er “kannte da jemanden”, der für mich interessant sein könnte. Dieser “jemand” war **Peter Kolb**, mein späterer Betreuer. **Peter** hatte die Zusage zur Emmy-Noether-Förderung der DFG bekommen; einem Programm, dass es Nachwuchswissenschaftlern ermöglichen soll, eine eigene Arbeitsgruppe zu etablieren. Dem **DFG** gebührt daher der Dank für die Finanzierung der Gruppe und damit meiner Stelle. Der Aufbau dieser Nachwuchsgruppe sollte unter Anleitung eines Mentors geschehen. In diesem Fall natürlich **Prof. Klebe**, bei dem ich mich sehr herzlich bedanken möchte. Nicht nur dafür, dass er mich mit Peter bekannt gemacht hat, sondern auch für seine Funktion als mein Zweitbetreuer und dem wissenschaftlichen Beitrag zu dieser Arbeit. **Prof. Klebe** hat stets die wissenschaftlichen Kommunikation und das Teilen von Ressourcen ermöglicht und gefördert. Letztendlich war es mir nur so möglich die wissenschaftliche Erfahrung zu erlangen, die ich jetzt habe.

Vor seiner Karriere als Gruppenleiter war **Peter** Post-Doc in San Francisco. Dementsprechend erfolgte das erste Kennenlernen per Internet-Telefonie, erst das zweite Bewerbungsgespräch war von Angesicht zu Angesicht. In einem Café. **Peter** hatte kein Büro zu dieser Zeit, das (gemeinsame) bezogen wir erst an unserem (gemeinsamen) ersten Arbeitstag. Die Entscheidung in einer Arbeitsgruppe zu promovieren, die zu diesem Zeitpunkt noch nicht existiert, kann natürlich verunsichern. Allerdings vermittelte **Peter** den Eindruck genau jenes Mentors, der frische Ideen hat und gleichzeitig die nötige Erfahrung, auf die man als Doktorand angewiesen ist. Dieser Eindruck sollte nicht täuschen. Optimierungsproblem gelöst sozusagen. Eine ausführliche Aufzählung würde den Rahmen sprengen, deswegen bedanke ich mich bei **Peter** einfach nur sehr herzlich für seine exzellente Betreuung, durch die ich nicht nur wissenschaftliche Erfahrung gewonnen habe, sondern auch als Person gewachsen bin. Ich glaube, dass gerade diese Erfahrungen in einer Promotion sehr wichtig sind. Ich wünsche **Peter** alles Gute für seine wissenschaftliche Karriere, die genau wie meine vor einem neuen Scheideweg steht. Seine harte Arbeit und Zielstrebigkeit habe ich stets bewundert und ich bin mir sicher, dass sie ihm eine erfolgreiche akademische Laufbahn bescheren werden.

Ich möchte mich außerdem bedanken bei allen Personen, die direkt Beitrag zu meiner Arbeit geliefert haben. Das sind unsere Kollaborationspartner **Nuška Tschammer**, **Viachaslau Bernat**, **Regine Brox**, **Arndt Siekmann** sowie **Steffi Dörr**.

Des Weiteren danke ich **Lydia Hartleben** und **Christian Sohn**. Sie haben, meist still im Hintergrund, für reibungslose organisatorische und technische Abläufe gesorgt. Natürlich haben auch die einzelnen Mitglieder der AGs Klebe und Kolb einen Beitrag dazu geleistet. Indem jeder kleine Aufgaben übernommen hat, die in einer Arbeitsgruppe einfach anfallen und getan werden müssen, hat die Gruppe funktioniert. Dafür möchte ich mich bei allen Kollegen bedanken, genauso wie für den wissenschaftlichen Austausch in den gemeinsamen Seminaren und den Flur- und Sozialraum-Diskussionen. Hervorheben möchte ich dabei **Inna und Adam Biela**, **Felix Terwesten**, **Martin Stieler**, **Manuel Neeb**, **Johannes Schiebel** und **Prof. Andreas Heine** mit denen man über Wissenschaft, aber auch mal über die Welt jenseits der Forschung diskutieren konnte.

Obwohl größtenteils selbst Doktoranden, trugen auch die vielen Freunde in Marburg dazu bei, ab und zu mal dem Alltag zu entschwenden. Allen voran **Bettina Wilke**, **Alexandra Birk**, **Eva-Lisa Bodmann**, **Maren und Stephan Jakobi**, **Thomas Gerlach** und **Michael Leitner**. Vielen Dank für die schönen Ausflüge, die gemeinsamen Essen und die angenehme Zeit, die ich mit euch verbringen durfte. Und denen, die noch promovieren, drücke ich natürlich die Daumen.

Den Wenigen die diese Zeilen lesen danke ich, dass sie sich die Zeit nehmen, ein mehr als dreiseitiges Vorwort zu lesen.

Auch meiner Familie gebührt großer Dank. Meiner Mutter **Mafalda** und meinem Vater **Hans-Jürgen Schmidt** danke ich von ganzem Herzen für ihre Fürsorge. Ihre harte Arbeit, haben mir Möglichkeiten im Leben eröffnet die nicht selbstverständlich sind. Ich weiß das zu schätzen und möchte versuchen stets das Beste aus diesen Möglichkeiten zu machen um auf diese Weise etwas zurück zu geben. Ich weiß, dass ich stets auf euren Rückhalt zählen konnte und auch können werde. Danke. Ein Schritt ins Ungewisse fällt immer leichter, wenn man weiß, dass man umkehren kann. Auch mei-

ner Schwester **Marisa** und meiner Großmutter **Maria** will ich für ihren Beitrag in meinem Leben danken.

Barbara Wienen danke ich sehr herzlich für ihre Rolle in meinem Leben, für ihre Unterstützung, Geduld, Motivation und aufbauenden Worte während meiner gesamten Promotion. Vor allem in der letzten Phase der Promotion hast du im alltäglichen Leben viel Last von meinen Schultern genommen und zusätzlich zu deiner getragen. Ich hoffe, dass ich das in einem Jahr in gleichem Maße für dich tun kann.

1 Introduction

Life can be considered as a network of chemical interactions, a very complex one, indeed. Driven by the mystery of its own existence, mankind has spent an enormous effort trying to understand the human body. Today, we have well-defined knowledge about general biological processes in human cells. An important key concept are proteins, biological polymers comprised of amino acids. An enormous amount of different proteins is encoded in our genes. Through post-transcriptional modifications, the number of different manifestations of proteins increases even more. Enzymes catalyze chemical reactions to convert a variety of substrates, small peptides transfer signals by interacting with receptor proteins, transporters and channels regulate the distribution of all kinds of substances, and histones and transcription factors constantly control the translation of DNA (deoxyribonucleic acid), just to mention some key classes. Together, they all form a highly complex network of dependencies and regulations in which every protein fulfills its designated task.

Severe malfunction of this molecular machinery often expresses itself in health disorders. Pharmaceutical research tries to identify the proteins responsible for a certain dysfunction. Once the causes are understood, one tries to develop strategies to fight a disease or its symptoms. This process is commonly referred to as “target identification” and “target validation”. At its end often stands a protein target, the activity of which has to be either promoted or inhibited to compensate for imbalances in the protein interaction network. What follows is the “hit and lead identification” phase, which tries to nominate molecules that can modulate the protein target as desired and are candidates for future drugs.

Nowadays, computer-aided drug design is an essential part in pharma-

ceutical research and many methods have been developed in this field. One of these methods, called docking, is the topic of this work. With first descriptions dating back more than 30 years¹, it has steadily been developed and is widely applied today. This constant progress might have been coupled with the availability of three-dimensional protein structures, which serve as input for docking. The docking algorithms then try to fit a given ligand into this protein structure and to quantify this fitness. That is, they try to find the spatial orientation, and the associated value of the scoring function, that gives the best sterical and chemical complementarity between the ligand and the protein. These informations help to understand protein-ligand interactions. By docking many ligands, one can try to find molecules that fit into a protein better than others, which makes docking a valuable tool in the context of lead identification.

For a long time, drug design efforts focused on finding “magic bullets”², molecules that exclusively bind and modulate the desired target protein in order to avoid side effects. Starting about a decade ago, a new paradigm emerged from the finding that Gleevec, a promising anti-cancer drug, was no magic bullet but gets its full effect by inhibiting multiple kinases.³ In the meantime, several drugs with polypharmacological activity have been released. Often these are kinase inhibitors targeting cancer.⁴ Consequently, lead identification processes have to be re-evaluated to develop molecules that manipulate the protein interaction network at multiple sites.

In the present work, docking is challenged to identify ligands with tailored selectivity, that is, the binding or non-binding to different protein structures. The following chapters shall give a more detailed introduction of docking and GPCRs (G protein-coupled receptors). GPCRs form a particular protein receptor family which is considered relevant in the target identification and target validation phase of many disorders. Several case studies are presented, showing the successful application of docking to different GPCR targets and structures. Eventually, this approach is scaled up to several hundred different structures of kinases, a different protein family constituting yet another pharmaceutically important target class.

2 Docking to GPCRs

Reproduced with permission from Schmidt, D.; Kolb, P. Computer-Aided Design of Selective Ligands Binding to G Protein-Coupled Receptors. Dtsch. Med. Wochenschr. 2013, 138, 2260-2264.

2.1 From an Idea to a Lead Structure

Starting point of the development of new drugs usually is a target protein. The ultimate aim is to modulate the activity of this target. New targets can steadily be identified thanks to the increasing understanding of molecular mechanisms. The inhibition or activation of these targets may directly or indirectly cure a disease or alleviate its symptoms.

During the early phase of drug design, one or more lead structures have to be identified. Lead structures are chemical compounds that do already show the desired effect on the target protein, but have not yet been optimized regarding essential properties of a drug. For example, these properties may be solubility, drug safety and target specificity. Lead structures form an intermediate step from the “first idea” to a mature drug.

The methods that are employed in the area of computer-aided drug design can be roughly separated into ligand-based and (protein-)structure-based methods.⁵ The method of **ligand-based** design uses the knowledge about established binding molecules. By systematically testing chemical variations or structurally similar molecules it can be deduced, which of the properties of a molecule affect the interactions with the desired target in a positive or negative way. Basing on these insights, new molecules can be

developed and proposed for chemical synthesis.

The **structure-based** design, including docking, only relies on the knowledge of the three-dimensional structure of the target protein.⁶ This structure should thus be known with atomic detail, which can be achieved by techniques like X-ray crystallography. Given this three-dimensional structure of the target and a virtual library of chemical compounds, docking can be used *in silico* to test, which substance best “fits” (Lock-and-Key principle) the binding site of the target protein. After experimental verification or falsification of the binding event of this ligand, the gained insights can be used to improve the docking, i.e. by tailoring the ligand library. This iterative process culminates in the discovery of new lead structures which themselves form the basis for the development of new candidates for the preclinical phase.

In short, docking is a computer-based methods to identify chemical substances that bind to a given protein and, presumably, modulate its activity. By selection and rational modification, such substances can be further developed to clinical candidates.

2.2 Docking – Theory and Technique

For a docking experiment, two distinct inputs are necessary, which is the three-dimensional structure of a protein on the one hand side, as well as a database with a large number, up to several millions are possible, of small molecules. Typically these substances have a molecular mass between 300 g mol^{-1} – 500 g mol^{-1} , which approximately corresponds to the range of most approved drugs.⁷ During the docking calculations, each and every small molecule is placed into the binding site of the protein individually and its orientation is optimized. The process is computationally demanding, since the potential ligands can adopt numerous distinct conformations and each conformer can be placed at different positions and in various orientations. Each possible combination of position and orientation (each pose) is evaluated to identify the one that maintains the best pos-

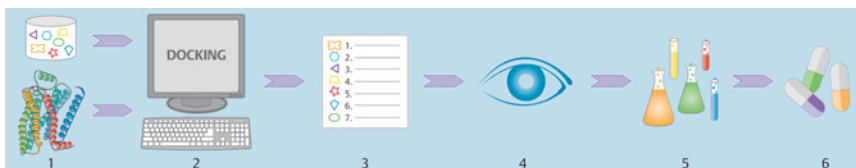


Figure 2.1: Typical application of docking in pharmaceutical research. Starting with a structure and a library of molecules (Step 1) the docking calculations are carried out *in silico* (Step 2). Each molecule is evaluated, scored and ranked (Step 3). After manual inspection (Step 4), single substances are selected and experimentally evaluated (Step 5). Ideally, new lead structures can be identified from these results (Step 6).

sible interactions with the target protein. Such evaluations commonly use physico-chemical potentials like the Lennard-Jones or the Coulomb potential or are described by so-called knowledge-based terms.⁸ The latter ones are based on the analysis of known three-dimensional protein-ligand structures, since atoms prefer certain spatial orientations towards each other, only dependent on their types.⁹ These relations can be described mathematically.

A “perfect” mathematical function for such evaluations is hard to realize since the calculations have to be very fast. This is usually achieved at the expense of accuracy. Finally, the relative orientation of ligand and protein towards each other is optimized by this mathematical function and associated with a score. That score represents the fineness of compatibility between the ligand and the binding site of the protein. The results of a docking run of a complete library thus is a, usually sorted, list of molecules and their corresponding scores, with the best-scoring molecule at the top (see Figure 2.1). At least in theory there should be a correlation between those values (or the rank within the list) and experimental affinities. Although this correlation is usually not strict^{10;11}, active molecules usually belong to the top scored molecules.

Docking as a method is very common today in computer-aided drug design, since it is fast compared to other structure-based methods and can help to reduce the number of required experimental tests. Furthermore,

docking is in many cases able to identify new ligands with chemical motifs that have not been described before, since the complementarity to the protein is the only requirement in contrast to ligand-based methods. Certainly docking also has a number of shortcomings. For example, the docked ligand is usually treated flexibly but the protein is mostly treated as rigid body, which obviously is not correct. Today, numerous docking programs exist that try to take this flexibility adequately into account^{12,13}, however, the number of possible poses increases dramatically, hence complicating the search for the correct pose. An additional complexity is imposed by single water molecules that interact with both, the ligand and the protein. The consideration of these water molecules likewise increases the docking complexity and the localization of such water molecules can hardly be predicted. Finally, docking suffers from conceptual limitations: Docking scores single, static poses of a ligand in a protein binding site. Certainly, the interactions between ligands and proteins *in vivo* are highly complex and dynamic processes that have not yet been finally understood. Considerably, we are only just beginning to glimpse the role of water on ligand binding affinity and kinetics.¹⁴ Hence, entropic effects, polarization or solvation effects will not be properly taken into account in a time-resolved manner in the near future. Nevertheless, docking is one of the most successful *in silico* tools, as shown by several studies.

In summary, docking is based on the “Lock-and-Key” principle. A molecule is systematically placed into the binding site of a macromolecule in various orientations. Every single pose is evaluated to identify the best-fitting one (by “trial-and-error”), which renders docking a very demanding method in terms of computer resources. Nevertheless, docking has become entrenched as tool in modern pharmaceutical research.

2.3 G Protein-Coupled Receptors

G protein-coupled receptors (GPCRs) are the most widely spread principle in nature to transduce extracellular signals over cellular membranes. This

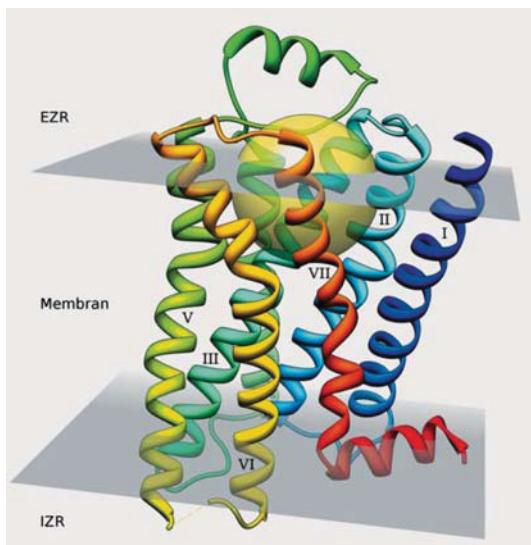


Figure 2.2: *Crystal structure (PDB¹⁷ 2rh1¹⁸) of the β_2 -adrenergic receptor (β_2AR) in ribbon representation. Amino acids are colored from blue (N-terminal, extracellular) to red (C-terminal, intracellular). Gray layers indicate the approximate position of the membrane. The typical binding site of small molecules is indicated as golden sphere. The seven trans-membrane helices are labeled by roman numbers (except IV). EZR: extra cellular space, IZR: intra cellular space.*

explains the high interest in pharmaceutical and academic research, met with this protein family. In 2012, the Nobel Prize in chemistry was awarded to Robert Lefkowitz and Brian Kobilka for their studies on GPCRs. By today, around one third of all approved drugs do bind to a GPCR.¹⁵ These include drugs like Aripiprazole for the treatment of schizophrenia or the antiplatelet agent Clopidogrel which both belong the most frequently sold drugs.¹⁶ Structurally all GPCRs of known geometry have their seven trans-membrane α -helices in common, that span the cell membrane in an anti-parallel fashion (see Figure 2.2). Upon the binding of a ligand on the extracellular side, the receptor conformation changes. This change allows the binding of a heterotrimeric G protein on the intracellular side. By the exchange of GDP (guanosine diphosphate) to GTP (guanosine triphos-

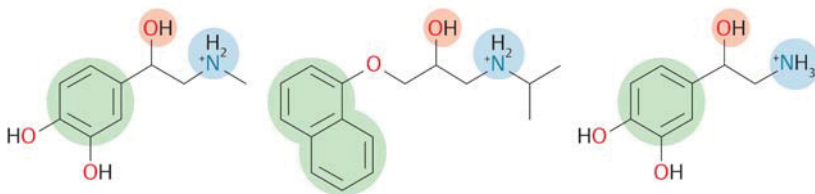


Figure 2.3: *Propanolol (middle) has a high structural similarity with the natural ligands of the adrenergic receptors Adrenaline (left) and Noradrenaline (right). The repeating moiety consists of a protonated amine (blue) in proximity to a hydroxy group (red) and a aromatic ring system (green).*

phate) in the G_α subunit of the G protein and the subsequent dissociation of this subunit the signal is transduced. Remarkably, many GPCRs do show a basal activity even in the absence of any bound ligand and are never completely muted. For these receptors, ligands are classified with respect to their influence on this basal activity. If an agent increases the basal activity it is classified as agonist, while an inverse agonist decreases basal activity. An antagonist does not induce any change in activity on its own, yet it competes with other molecules for their respective binding site.

By the arrangement of the seven transmembrane helices these receptors are highly flexible, allowing for a fine-tuned control of signal transduction. This flexibility together with the fact that GPCRs are membrane-bound receptors, however complicates the structure determination via protein X-ray crystallography. On that account, the development of GPCR modulators has been limited to ligand-based methods for a very long time. This becomes apparent in the structural similarity of the prototypic betablocker Propanolol, designed by Sir James Black in the 1960s, and many others with adrenaline and noradrenaline from which they were derived (see Figure 2.3). Only in 2000, the publication of the structure of rhodopsin allowed for the subsequent application of structure-based design methods. Nevertheless, it took another seven years until the structural elucidation of a pharmaceutically relevant GPCR, the β_2 -adrenergic receptor (β_2 AR) by the groups of Brian Kobilka¹⁹ and Ray Stevens.¹⁸ Since this point in time,

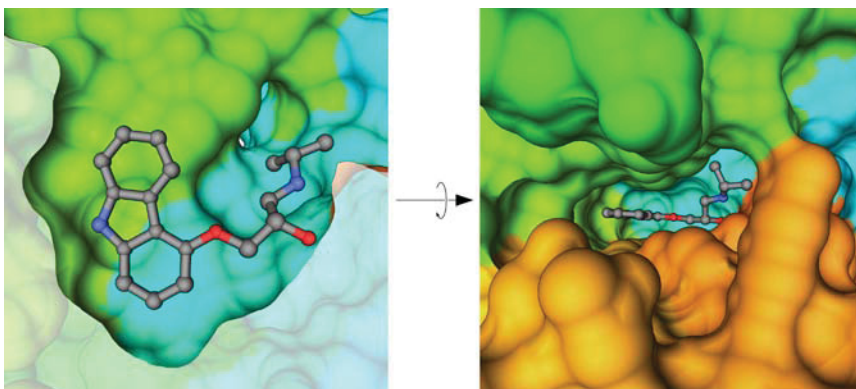


Figure 2.4: Surface of the binding site of the β_2 -adrenergic receptor bound to the inverse agonist Carazolol (carbon atoms in gray). **(Left)** Side view, parallel to membrane. Surface cut open to allow view into the binding site (shaded area). **(Right)** View turned by 90° , view perpendicular to membrane. Color code according to Figure 2.2.

a total of 76 additional structures could be solved (by the time of publication of the original article), leading to a downright rush to structure-based methods that shed light on a number of basic questions on GPCRs. Some of the corresponding publications shall be discussed in detail below.

2.3.1 Docking identifies novel GPCR ligands

The β_2 -adrenergic receptor has been the target of the first successful docking studies on GPCRs.^{20;21} Due to the deep and narrow binding cleft, this receptor is almost ideally suited for docking approaches. The background is that in such a binding site, less possible orientations exist for a ligand and in addition, the effects of water (solvation effects) are less pronounced, thereby simplifying the calculations (see Figure 2.4). The first crystal structure (PDB 2rh1) was elucidated in complex with Carazolol, an inverse agonist. Hence, the receptor was assumed to be in an inactive conformation.¹⁹ Consequently, it was investigated whether the given receptor structure would exclusively be able to identify new inverse agonists.

In addition, it has been of special interest, whether or not docking is able

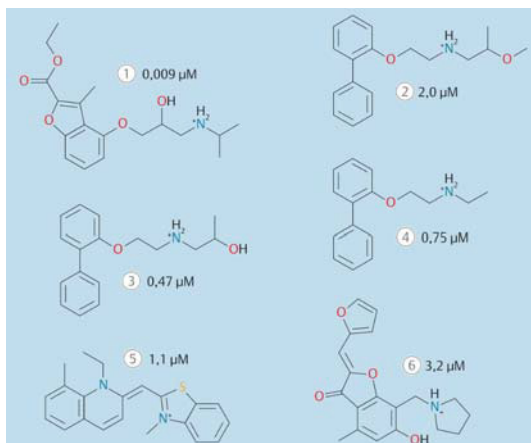


Figure 2.5: Six verified molecules modulating β_2 AR identified by Docking.

to identify novel ligands for this receptor which has been in the focus of research since the 1960s. Novel ligands, that is, molecules bearing chemical motifs that have not been described before as binding to this receptor. After successful docking and testing of 25 compounds, six molecules could be identified as binding ligands, corresponding to a very high hit rate of 24%. Two of these substances are based on chemical scaffolds that have not been described as binding to the β_2 -adrenergic receptor before (see Figure 2.5, **5+6**). The determination of the intrinsic activity classifies these molecules as inverse agonists. Remarkably, compound **1** in Figure 2.5 represents the most potent β_2 -adrenergic inverse agonist identified so far, with an affinity of 9 nmol L^{-1} . In summary this, as well as a second study by Sabio *et al.*²¹ could verify all assumptions made before. First of all, the binding site of the β_2 -adrenergic receptor is perfectly suited for docking approaches, as underlined by the high hit rate and the excellent affinities. As assumed, all ligands could be classified as inverse agonists, which might be biased by the fact that agonists are less frequent than inverse agonists. Finally, docking was able to identify several substances which can be considered new, in a way that they show only limited similarity to known ligands such as the Noradrenaline (see Figure 2.3).

2.3.2 Homology Models of GPCRs

As introduced, docking methods rely on the three-dimensional arrangement of the atoms of a protein. Since the experimental structure elucidation is especially difficult for GPCRs as membrane receptors, until today a comparatively low number of GPCRs have been characterized this way. To close this structural gap, a computer-based method exists, called homology modeling. This generic term stands for a whole number of techniques that can be reduced to a common denominator: The calculation of a three-dimensional structure of a protein from the structure of a similar (homologue) protein by comparing this two protein's sequences.

A GPCR of pharmaceutical interest is the chemokine C-X-C receptor 3 (CXCR3), for which so far no crystallographically determined structure exists. CXCR3 is especially expressed on lymphocytes²² and is involved in Multiple Sclerosis²³ and transplant rejection.²⁴ Based on the experimentally determined structure of the related receptor CXCR4 we built a homology model of CXCR3.²⁵ Subsequently, we tried to identify molecules that regulate the activity of this receptor using docking approaches. Out of seven selected compounds, biological activity against CXCR3 could be proven for four of those by *in vitro* assays. At the same time, no activity could be detected for these molecules against the related CXCR4, which speaks in favor of the relevance of the models used. In an additional step, a second docking was conducted to the crystal structure of the CXCR4. This time six substances have been selected and tested experimentally, out of which four were active but this time exclusively against CXCR4. Next to our own work, a second CXCR4 specific docking campaign was carried out by another group without showing any overlap with the ligands identified by our approach.²⁶

2.3.3 Distinct Conformations Identify Different Ligands

The structural flexibility of GPCRs can be exploited together with the sensitivity of docking towards subtle changes in the protein structure of GPCRs to use slightly different protein conformations to identify different

ligands. This has been demonstrated by two studies. On the one hand by a comparison of the crystal structure of the Dopamine D₃ receptor with a homology model, built before the publication of the former²⁷ and on the other hand by means of comparison of multiple homology models of the A₁-adenosine receptor.²⁸

In both cases, almost all protein structures used were able to identify potent ligands. The respective ligand sets did not overlap, hence, ligands that were well ranked in a given receptor conformation were not necessarily amongst the best-scored molecules in a second conformation. The purposeful generation of alternative receptor conformations could thus be a potential method to identify more target-specific ligands.

In summary, G protein-coupled receptors or GPCRs belong to the protein families with the highest current pharmaceutical relevance. Together with the increasing number of solved crystal structures of GPCRs, the application of docking studies has gained importance. Several studies were able to show that docking is well applicable to GPCRs in general. Chemically innovative and diverse ligands could be identified, predominantly inverse agonists, amongst others for the group of adrenergic receptors.^{27;28}

2.4 Discussion

The above summarized examples show that structure-based methods do allow the identification of ligands with chemically novel scaffolds. Even, if the target protein has been in the focus of pharmaceutical research for many decades. Presumably the high structural flexibility of GPCRs in particular are responsible of the driving force of the successful application of docking to this protein family. There is not a single “correct” receptor conformation but there exist several.²⁹ Each of these can be successfully used as docking target. It should not go unmentioned at this point that in general, small-molecule ligand databases tend to be enriched with GPCR binding compounds.³⁰ This rests upon the broad spectrum of ligands that can be bound by GPCRs, as well as the multitudinous shares of medicinal

chemistry to this field in the past decades. But as seen from the many ligands with chemical novelty this cannot be the only explanation of the high hit rates of docking-based virtual screenings. There is, however, room for improvement when docking to receptors that bind large ligands and hence have a broad binding site. Those spacious pockets and highly solvent accessible pockets render current docking approaches complicated.

It can nevertheless be assumed that protein structure-based methods will contribute especially to the field of GPCR research and support the utilization of the “golden age” of the structural biology of GPCRs for future pharmaceutical research.^{29;31}

As a consequence, with the increasing number of GPCR structures, rational drug design will likely lead to some new active agents with structural moieties unprecedented so far. It remains to see whether these molecules also offer improved mechanisms of action. Furthermore, by the ongoing elucidation of protein structure, one can expect an improving understanding of the mechanisms of signal transduction. This will likewise allow for a more specific design of molecules with desired properties. On these grounds, future drugs will be developed with specific pharmacological profiles of polypharmacological properties.

3 Identifying Modulators of CXC Receptors 3 and 4 with Tailored Selectivity using Multi-Target Docking

Reproduced with permission from Schmidt, D.; Bernat, V.; Brox, R.; Tschammer, N.; Kolb, P. Identifying Modulators of CXC Receptors 3 and 4 with Tailored Selectivity using Multi-Target Docking. ACS Chem. Biol. 2015, 10, 715-724, Copyright 2015 American Chemical Society.

Author Contributions: D.S., V.B., N.T., and P.K. designed the research. V.B. built the homology model. D.S. carried out docking, and D.S. and P.K. selected and acquired the ligands. D.S. carried out DLS and NMR experiments. R.B. and N.T. carried out biological assays. N.T. analyzed biological data. D.S., V.B., N.T., and P.K. wrote the manuscript.

3.1 Introduction

Most cells are delimited by membranes formed by lipid bilayers. For a cell to be able to react to its environment, communication across such membranes must be possible, however. Signal transduction is frequently achieved through G protein-coupled receptors (GPCRs), proteins that consist of seven membrane-spanning helices connected by intra- and extracellular loops. Binding of an agent to the outer part of the receptor elicits

a conformational change that can influence processes inside the cell in a multitude of ways.³² Historically, the function of many GPCRs has been defined by the ligands that bind to them, and also today, small organic molecules frequently serve as tool compounds to elucidate the mechanism and function of a particular receptor.³³

There are many possible ways to find such chemical modulators but one particularly effective method is small molecule docking. Docking of molecule libraries to X-ray structures or homology models of protein targets has become a well-established technique.^{6;30} Even the application to GPCRs has gained substantial momentum over the past years^{20;21;26–28;34–37}, owing to the steadily increasing number of X-ray structures of this pharmaceutically important receptor class. The focus of these and other studies on different protein classes has usually been the hit rate and whether or not chemotypes previously undescribed for a particular target were identified.²⁰ The selectivity of compounds — although crucial for the effects of a particular molecule — has rarely been included in such calculations and mainly been investigated *post hoc* by experimentally testing ligands on other (anti)targets.²⁸ Nonetheless, by docking to the structure of only one target, many selective compounds have been identified, but this fact was only established after the calculations. Outside of the realm of docking, the identification of ligands with polypharmacological profiles has recently been demonstrated using multi-objective optimization of virtually enumerated compounds.³⁸ In their ligand-based approach, the authors used Bayesian models trained on experimental affinity data to predict activity. Yet, also this approach might benefit from seed molecules with defined polypharmacological profile that have been identified with docking.

We therefore wanted to prospectively identify ligands for two proteins through docking and predict their binding pattern. For this purpose, we employed a re-ranking function that would allow us to reorder ligands according to a desired binding pattern. A system was chosen where the challenge was to find **non-selective** ligands. Such a pair are the CXC chemokine receptors 3 and 4 (CXCR3 and CXCR4, respectively). Signaling peptides that bind to these receptors share a specific CXC motif. This

Table 3.1: *Size comparison of exemplary GPCR binding sites.*^a

Receptor	PDB code	Pocket ID	Size (\AA^3)
Chemokine receptor CXCR4	3odu	110	3320
Chemokine receptor CCR5	4mbs ^b	34	3995
Neurotensin receptor 1	4buo ^b	37	1950
β_2 -adrenergic receptor	2rh1	49	2057
Adenosine A _{2A} receptor	3eml	52	1060

^aExemplary pocket volumes were taken from the CastP database³⁹^bStructures were not available at the time of analysis and had to be uploaded to the server

motif has two Cys (cysteine) residues flanking a third one (Cys-X-Cys). As shown for the CXCR4, the binding of its chemokine involves extracellular parts of the receptor as well as residues within the transmembrane binding pocket.³² With around 3000\AA^3 , this pocket is distinctly larger than those of non-chemokine class A GPCRs of known X-ray structure (see Table 3.1). This size presents a challenge when docking to this pocket, since it allows for many degrees of freedom. Furthermore, one can imagine a scenario where two ligands bind to different subpockets within this binding site. To properly account for this possibility, we used an allosteric ternary complex model to analyze our assay data. This model yields a binding constant K_B and a cooperativity factor $\alpha\beta$, the latter describing how the binding of a modulator influences the binding of a different ligand.⁴⁰

The receptors investigated here are involved in severe pathologies such as multiple sclerosis²³ and allograft rejection²⁴ in case of the CXCR3 or cancer⁴¹ and HIV (human immunodeficiency virus) infection⁴² in case of the CXCR4. More intriguingly, recent results in animal models indicate the potential joint role of these receptors in disease-related animal models. Kohler and coworkers, for example, showed that the concomitant modulation of CXCR3 and CXCR4 by synthetic peptide antagonists led to inhibition of experimental autoimmune encephalomyelitis, a mouse model of multiple sclerosis.⁴³ In addition, it was shown only recently that the CXCR3/CXCR4 double-knockdown reduced metastasis of colorectal can-

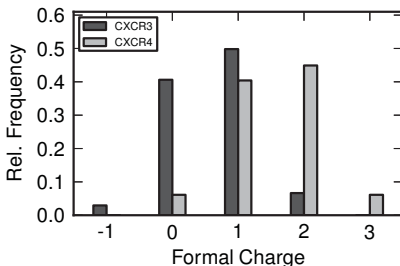


Figure 3.1: The net charges of CXCR3 (dark gray) and CXCR4 (light gray) ligands from ChEMBL are differently distributed. Charges were assigned using protonation states from the ZINC database.⁴⁶ The height of the bars indicate the relative frequency of occurrence of a particular formal charge.

cer into other tissues like lung and liver in the respective mouse model.⁴⁴ Notably, the authors postulate a synergistic effect of CXCR3 activation with CXCR4. This, in turn, means that simultaneous blocking of both receptors could have a higher pharmacological effect compared to the inhibition of either receptor.

Although the CXCR3 and the CXCR4 share 36.7% and 63.9% sequence identity and similarity, respectively, their small molecule ligand spaces differ: ChEMBL⁴⁵ (version 17) lists 858 and 484 ligands tested against the CXCR3 and the CXCR4, respectively, but none of them are categorized as active in the other receptor at a threshold of 10 μ M. A comparison of the properties of the ligands in each set shows that the CXCR4 tends to bind ligands with a stronger positive charge compared to the CXCR3 (see Figure 3.1). In addition, the ligand sets differ in the distribution of the number of hydrogen bond donors and the predicted lipophilicity (see Figure 3.2). Since a successful dual binder ideally is within the “allowed” range of both sets, this makes prediction more challenging. Finally, as only the structure of the CXCR4 has been determined crystallographically³², another challenge was that the CXCR3 structure had to be obtained by homology modeling.

For other chemokine targets, bioactive substances acting on more than

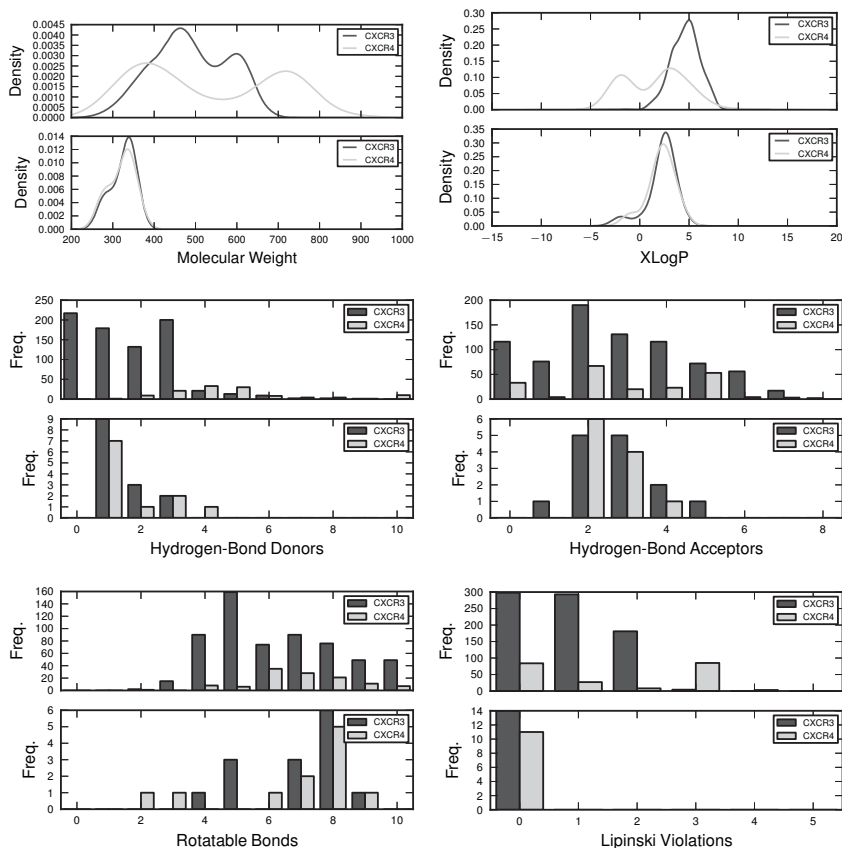


Figure 3.2: The distributions of selected physico-chemical properties are shown for ligands from ChEMBL (upper panel in each plot) and the new ligands described here (lower panel in each plot) separately for ligands for the CXCR3 (dark gray) and those for the CXCR4 (light gray). Dual binders have been counted once in each of this two sets. Properties with continuous values are shown as distributions with the y-axis showing the respective density (as calculated by a kernel density estimator), whereas discrete values are represented as histograms with the absolute frequency on the y-axis. Each upper panel shares its x-axis with the respective lower panel. All features were calculated using Openeye’s MolProp TK.⁴⁷

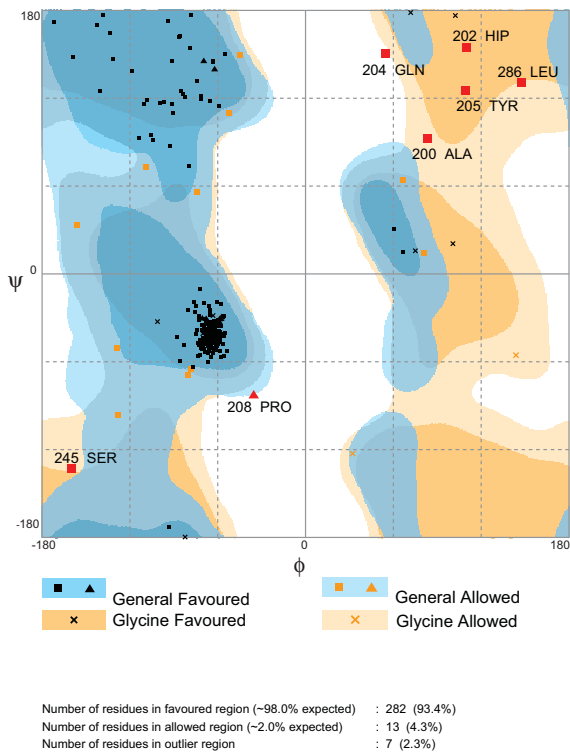
one receptor have been identified through optimization. AMD3451 is one such example, a cyclam-based nanomolar dual inhibitor which targets the CC chemokine receptor CCR5 and the CXCR4, thereby blocking entry of R5 and X4 human immunodeficiency virus in parallel.⁴⁸ UCB35625 also acts on two chemokine receptors, namely the CCR1 and the CCR3 and suppresses receptor-induced chemotaxis of transfected cells at nanomolar concentration. This effect can be used to treat allergic inflammatory diseases such as asthma, which involve the CCR1- and CCR3-based recruitment of eosinophiles.⁴⁹ These two examples, among others, highlight the importance of modulators targeting multiple chemokine receptors and the importance of prospective identification of such ligands.

Of course, trying to identify molecules that interact with multiple targets bears the risk of incorrectly annotating frequent hitters as dual binders. We therefore not only predicted dual binders, but also CXCR3-selective and CXCR4-selective compounds using our docking approach. The correct categorization of each ligand was verified by *in vitro* assays for both receptors. Multiple control experiments were used to rule out potential frequent hitters or aggregators. In a second round we support the identification of a dual binder by selecting chemically related molecules and verifying their action on both targets.

3.2 Results and Discussion

3.2.1 CXCR3 homology model

The validity of the CXCR3 homology model was assessed using a Ramachandran plot (see Figure 3.3). The majority of residues (97.7 %) lie in the allowed conformational space. The outlier residues are located in ECL2 (extracellular loop) and ICL2 and 3 (intracellular loop), thus hinting at the difficulty of loop modeling. To assess the predictive power of the model in ligand identification, we docked a subset of 511 small-molecule ligands and decoys of the CXCR3 with unambiguous IC₅₀ (half maximal inhibitory concentration) values from the ChEMBL GPCR SARfari database.⁴⁵ The



RAMPAGE by Paul de Bakker and Simon Lovell available at <http://www.crysl.bloc.cam.ac.uk/rampage/>
 Please cite: S.G. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Pisent, J.S. Richardson & D.C. Richardson (2002)
 Structure validation by Ca geometry, ϕ/ψ and C β deviation. *Protein: Structure, Function & Genetics* **50**: 437-450

Figure 3.3: Ramachandran plot for the residues of the CXCR3 homology model. Analysis was carried out using RAMPAGE.⁵⁰

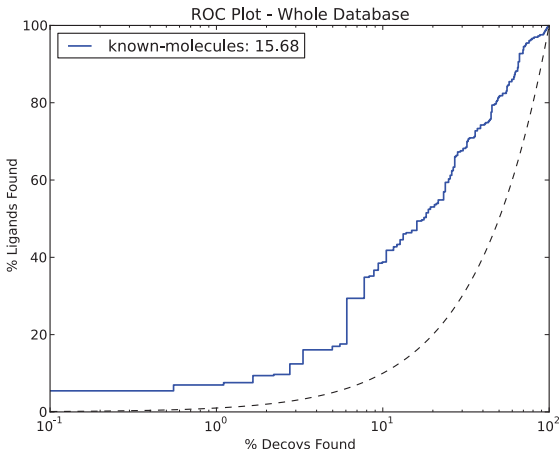


Figure 3.4: ROC curve of ligands and decoys from ChEMBL which were docked to the final homology model of the CXCR3. Blue line shows the enrichment of binders over non-binders on a logarithmic x-axis (focus on early enrichment). Dashed line shows the expected enrichment of a randomized set. The graph highlights that the final homology model is able to rank known binders higher than non-binders.

resulting ROC (receiver operating characteristic)⁵¹ plot demonstrates good early enrichment and a total logarithmic AUC (area under curve) of 15.68 (see Figure 3.4). A structural overlay of the CXCR3 model and the CXCR4 X-ray structure (see Figure 3.5) shows the position of conserved residues, and their similar orientation. The non-conserved residues include 7.39 (Ballesteros-Weinstein notation⁵²), where CXCR3 contains a serine while most other chemokines feature an aspartate⁵³, which has been shown to be a key interaction in CXCR4 for binding AMD3100.⁵⁴ Furthermore, two positions of the aromatic zipper⁵⁵, 3.32 and 7.43, show a Tyr (tyrosine)-Phe (phenylalanine) exchange and vice-versa. At position 3.29, CXCR4 has a histidine that has been shown to interact with the receptor’s peptidic ligands³², whereas the glycine that is at the same position in CXCR3 offers more room for ligands and has been proposed as potential selectivity site.⁵³ Finally, at positions 5.39 and 7.35, the CXCR3, in



Figure 3.5: *Overlay of the CXCR3 homology model in brick-red and the CXCR4 crystal structure in green in ribbon representation. Residues around the binding site are shown. Labeled residues are discussed in the text.*

contrast to the CXCR4, contains two basic amino acids, which are potential opposites for the negatively charged aspartates and glutamates in the vicinity. Altogether, our CXCR3 model resembles the conserved structural features of the CXCR4, while the sequence differences have been appropriately captured and are consistent with earlier experimental data.⁵³

3.2.2 Docking

More than 2 million compounds from the ZINC database were docked to the receptor structures of the CXCR3 and the CXCR4, respectively. After inspecting the best docked poses, we selected six and seven compounds we expected to be selective for the CXCR3 and the CXCR4, respectively. Through re-ranking all compounds with equation (3.1), four more compounds predicted to bind to both receptors were selected. In addition, we verified that all compounds did not contain any of the known problematic substructures as listed in the PAINS Filter A.⁵⁶ All 17 compounds were assayed for their ability to bind and inhibit CXCR3- and CXCR4-

mediated activation of G proteins, which was measured via [^{35}S]GTP γ S incorporation.

A detailed consideration of the predicted binding poses shows that several key interactions can be observed repeatedly. These are contacts to W109^{2,60}, D112^{2,63}, Q204 and R212 in the CXCR3 and D97^{2,63}, R188 and E288^{7,39} in the CXCR4. Several of these residues have previously been identified as key residues involved in ligand recognition and binding or signal transduction upon ligand binding.^{57–59} They are part of the so called “minor binding pocket”, which is formed by the transmembrane helices TM-I (transmembrane helix), -II, -III and -VII.⁶⁰

Figure 3.6 shows a comparative depiction of compounds **2** (see Figure 3.6a,b), **9** (Figure 3.6e,f) and **7** (see Figure 3.6c,d), which were thought to bind to the CXCR3 exclusively, the CXCR4 exclusively or to be a dual binder, respectively. In the CXCR3, compound **2** forms π -stacking interactions with W109^{2,60} as well as hydrogen bonds with Q204 and the backbone of C203. The carboxy function interacts with R212 (see Figure 3.6a). In the CXCR4, the quinazoline moiety does not form strong interactions with the receptor at all. The imidazole-ketone is charged and points towards E288^{7,39}, while the carboxy group cannot form an optimal interaction in this binding pose (see Figure 3.6b). Compound **2** ranks highly in the CXCR3 docking and below rank 485 000 in the CXCR4, resulting in a selectivity score S of 0.09 ($0 \leq S \leq 1$; at $S = 0$, a compound is non-selective; see Table 3.2 for comprehensive list; see Equation 3.1 for calculation of S). Because of this rank difference, taken together with the higher number of interactions with the CXCR3, we considered compound **2** to be CXCR3-selective. Compound **7** shows favorable interactions with both receptors, albeit in different poses. In the CXCR3, the ligand is anchored with a hydrogen bond between its amide and D112^{2,63} and the 2-(imidazol-1-yl)pyridine forms aromatic interactions with W109^{2,60} and fills a rather hydrophobic pocket close to A127^{3,28} and L190. The aniline motif is captured by F47 (see Figure 3.6c). In the CXCR4, the ligand is anchored in a comparable way to D97^{2,60}, but the entire ligand is bent and points to the bottom of the pocket with the imidazole ring, forming

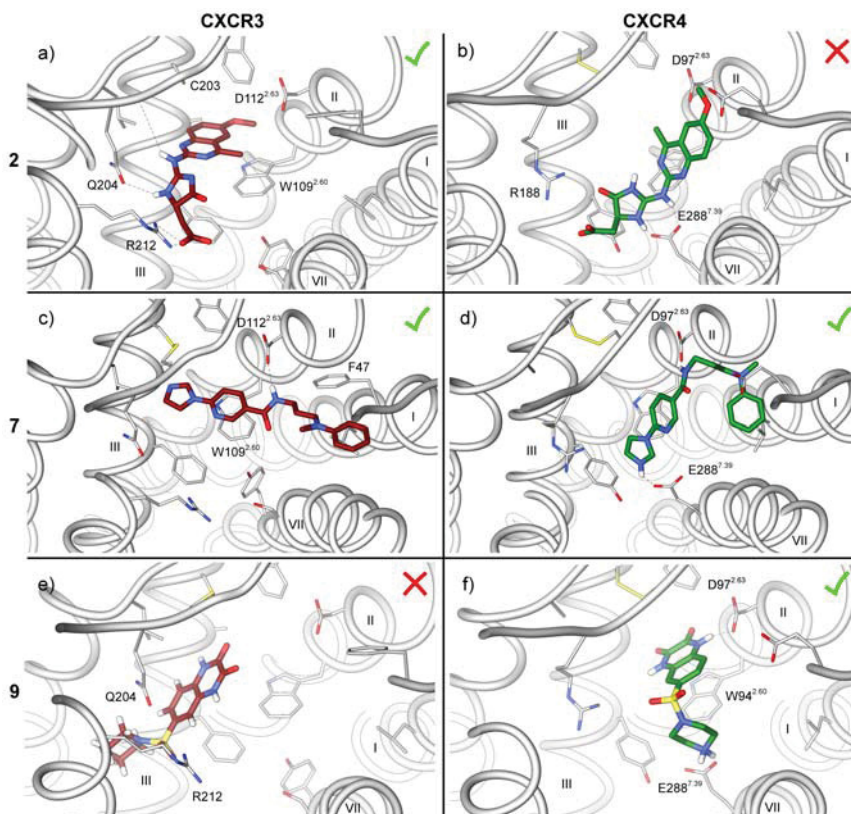


Figure 3.6: Comparison of docking poses for compound **2**, a selective CXCR3 binder (a,b), compound **7**, a dual binder (c,d) and compound **9**, an exclusive CXCR4 binder (e,f) in the CXCR3 homology model (a,c,e) and the CXCR4 crystal structure (b,d,f), respectively. Ligands in the CXCR3 structure are shown in brick-red and in green in the CXCR4 structure. The proteins are shown in light gray. Residues are labeled whenever discussed in the text. Roman numerals are used to label protein helices. 2D structures of compounds are depicted in Table 3.3. Checks and crosses indicate that the respective receptor-ligand combination was predicted to be active or inactive, respectively. All predictions were shown to be correct.

Table 3.2: *Docking ranks of compounds selected for testing.*^a

Cpd	Rank		S
	CXCR3	CXCR4	
1	54	2 891 970	0.946
2	114	485 407	0.09
3	456	2 863 174	0.930
4	157	2 791 072	0.897
5	25 526	40 768	0.011
6	5541	34 472	0.0069
7	12 488	39 455	0.009
8	2 374 085	263	0.934
9	868 801	49	0.236
10	783 666	489	0.207
11	—	301	—

^aFor every docked compound, the achieved absolute ranks in both dockings and the selectivity scores according to equation (1) are given. Dashes indicate that the compound was not scored in the respective docking at all.

electrostatic interactions with E288^{7,39} (see Figure 3.6d). In both receptors, compound **7** ranks within the top 1.5 % of the list ($S = 0.009$). Lastly, compound **9** does not form reasonable interactions in the CXCR3 at all (see Figure 3.6e), while it stacks with W94^{2,60} and can simultaneously form a hydrogen bond to D97^{2,63} in the CXCR4. In addition, a charge-assisted hydrogen bond to E288^{7,39} from the piperazine ring is possible, due to a kink introduced by the sulfone linker (see Figure 3.6f). It ranks among the 50 best compounds in the CXCR4 and worse than position 850 000 in the CXCR3, yielding a selectivity score of 0.236. Therefore, we expected this compound to bind to the CXCR4 exclusively.

3.2.3 [³⁵S]GTP γ S Assay

Compounds **1–4** and **18–20** were predicted to bind to the CXCR3 but not to the CXCR4 (see Table 3.3; Table 3.5). And indeed, four out of seven substances inhibited the CXCL11-mediated activation of G proteins with the best K_B at 12 nM and an $\alpha\beta$ value (which quantifies cooperativity)

Table 3.3: Comprehensive assay results for identified compounds.^a

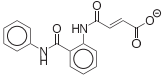
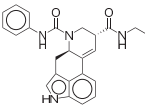
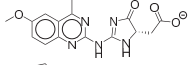
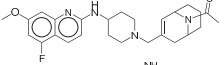
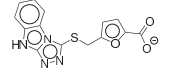
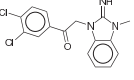
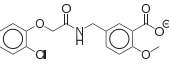
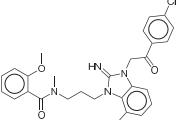
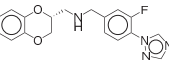
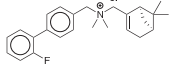
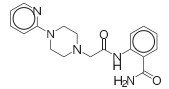
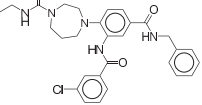
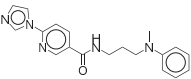
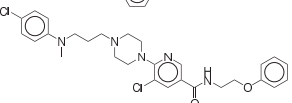
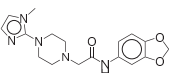
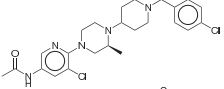
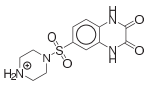
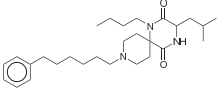
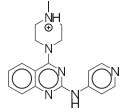
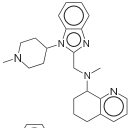
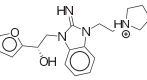
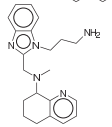
Prediction	Cpd	CXCR3-[³⁵ S]GTPγS	CXCR4-[³⁵ S]GTPγS	Displ.	BEI ^d
CXCR3	CXCR4	PK _B ± SEM		PK _B ± SEM	RAMX3 ^e
		<i>α</i> / <i>β</i> ^f	<i>α</i> / <i>β</i> ^f		
•	—	1	61 %@10 μM ^g	—	—/—
•	—	2	6.62 ± 0.36	—	20.18/—
•	—	3	7.91 ± 0.56	—	22.14/—
•	—	4	7.30 ± 0.72	—	21.03/—
•	•	5	7.34 ± 0.69	—	20.76/—
•	•	6	5.85 ± 0.74	6.34±0.66	21.50/18.65
•	•	7	7.30 ± 0.55	6.84±0.36	21.28/20.42
—	•	8	6.34 ± 0.45	—	20.32/—
—	•	9	—	7.18±0.69	—/23.09
—	•	10	—	6.10±0.33	—/18.94
—	•	11	—	5.81±0.16	—/16.99

^aEach compound that has correctly been predicted as a ligand by docking is shown. The further columns list the receptor it is predicted to be active against; the results from the [³⁵S]GTPγS assay ; the displacement of RAMX3 from the CXCR3 and the calculated binding efficiency index based on the measured pK_B values. Dashes indicate no activity of a compound. Shown enantiomers represent the one ranked highest in docking, whereas racemates have been used in all assays. ^bCooperativity Factor. ^cDisplacement of RAMX3 from CXCR3, (+) complete, (++) partial, (−) no displacement. ^dBinding efficiency index ($BEI = \frac{pIC_{50} \text{ or } pK}{MW}$) for CXCR3/CXCR4.

^eMost similar active compound found in respective ChEMBL set.

^gpK_B could not be determined. Instead % inhibition at maximal concentration is given.

Table 3.4: Chemical structures of identified compounds.^a

Cpd		SIM ^d	Closest active ^e
1		0.31	
2		0.22	
3		0.28	
4		0.32	
5		0.26 ^{g,h}	
6		0.30 ^{g,h}	
7		0.41 ^{g,h}	
8		0.27	
9		0.15	
10		0.25	
11		0.26	

^aThe structure is shown for each compound that has correctly been predicted. The further columns show the most similar known active from ChEMBL as well as the respective similarity.

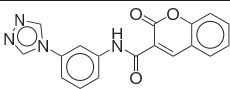
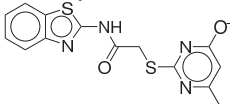
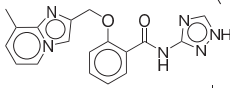
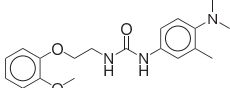
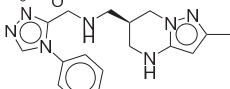
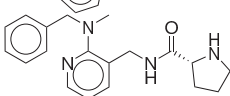
^dSimilarity expressed as Tanimoto coefficient using Morgan fingerprints.

^eMost similar active compound found in respective ChEMBL set.

^gMost similar active compound found in both ChEMBL sets.

^hMost similar compound is active on the CXCR3.

Table 3.5: Chemical structures of molecules selected for testing without showing an effect on the CXCR3 or the CXCR4.

Prediction		Cpd	
CXCR3	CXCR4		
•	—	18	
•	—	19	
•	—	20	
•	•	21	
—	•	22	
—	•	23	

of 0.37 (compound **3**). At the same time, none of the compounds showed detectable activity at the CXCR4.

Out of the six compounds with predicted CXCR4 activity (compounds **8–11**, **22**, **23**), three were confirmed in the CXCR4 [^{35}S]GTP γ S assay, with compound **9** modulating the activation of the CXCR4 by CXCL12 with nanomolar K_{B} (see Table 3.3). Compounds **10** and **11** acted as negative allosteric modulators, showing a strong negative cooperativity with $\alpha\beta$ values of 0.06 and 0.01, respectively. In contrast, compound **9** behaved as a positive allosteric modulator, with $\alpha\beta = 7.45$. Two more substances were completely inactive (compounds **22**, **23**), while the last one (compound **8**) unexpectedly showed activity in the CXCR3 [^{35}S]GTP γ S assay, independent of radioligand displacement. This indicates that the radioligand RAMX3 and compound **8** bind to different subpockets simultaneously. Overall, the results were supported by the absence of radioligand displacement in the CXCR3 (see Table 3.3; Table 3.6). In total, this corresponds to a 50 % hit rate in the CXCR4 [^{35}S]GTP γ S assay.

Finally, four compounds (compounds **5–7**, **21**) were expected to exhibit affinity towards both receptors, the most challenging prediction category in this study. Two of these four, namely compounds **6** and **7**, are active in both receptor assays – as predicted. Interestingly, compound **6** behaves as positive allosteric modulator in both receptors (see Figure 3.7) while not detectably displacing the CXCR3 radioligand. This is in accordance with the high $\alpha\beta$ value of 47. Compound **7**, on the other hand, shows a negative allosteric modulator effect throughout both receptors, accompanied by the efficient displacement of the allosteric CXCR3 radioligand in the binding assay (see Figure 3.9). For the further investigation of the dual binder chemical space, we selected six additional compounds by fingerprint similarity to compound **7** (substances **12–17**). This second set of compounds consistently inhibited both receptors with K_{B} s ranging from 300 nM down to the one-digit nanomolar range (see Table 3.7), indicating that this scaffold entertains efficient interactions with the receptor. Of the remaining two compounds that we had predicted as dual binders, compound **5** showed an effect in the CXCR3 while **21** showed no effect in either receptor (see

Table 3.6: *Displacement of RAMX3 from the CXCR3 for denoted compounds.*

Cpd	pK_B ± SEM	α
1	5.15 ± 0.19	0.41
2	4.59 ± 0.19	0.09
3	4.44 ± 0.51	0.29
4	5.80 ± 0.28	0.56
5	5.14 ± 0.14	0.00
6	—	—
7	5.45 ± 0.34	0.07
8	—	—
9	—	—
10	—	—
11	—	—
12	85 %@10 μM ^a	—
13	61 %@10 μM ^a	—
14	74 %@10 μM ^a	—
15	53 %@10 μM ^a	—
16	40 %@10 μM ^a	—
17	44 %@10 μM ^a	—
18	4.86 ± 0.10	0.00
19	4.53 ± 0.26	0.15
20	5.54 ± 0.25	0.26
21	5.66 ± 0.27	0.05
22	—	—
23	—	—

^apK_B could not be determined. Instead % inhibition at maximal concentration is given.

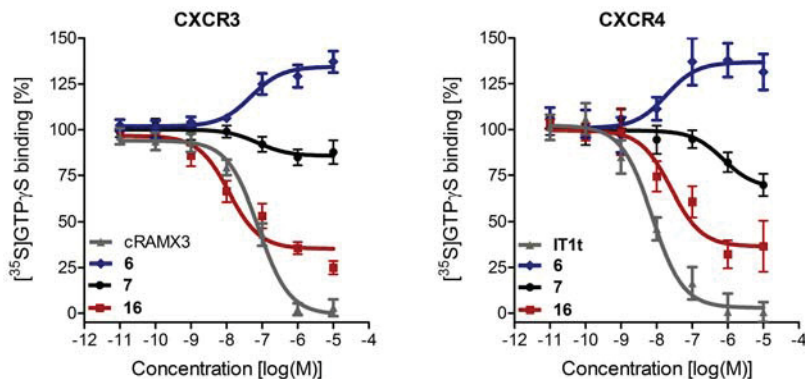


Figure 3.7: Representative dose-effect curves from the $[^{35}\text{S}]$ GTP γ S assay. The dose-dependent receptor activity is plotted for known nanomolar receptor antagonists as reference (gray) and selected dual binders identified in this study. An endogenous chemokine is used for each target to achieve initial receptor activity (100%). cRAMX3 denotes the “cold” (unlabeled) variant of RAMX3 and is depicted in Figure 3.8.

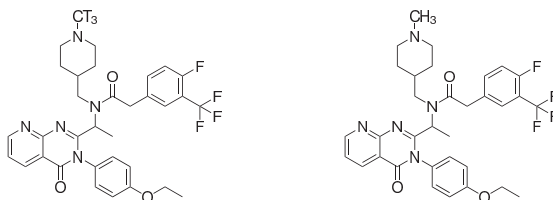
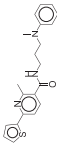
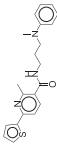
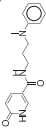
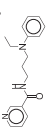
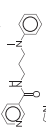
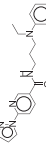
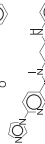


Figure 3.8: The structure of labeled (“hot”) and unlabeled (“cold”) RAMX3.⁶¹

Table 3.7: For compounds selected by 2D-similarity to compound **7**, the results of the [³⁵S] GTPγS assay and the derived binding efficiency indices are shown.

Cpd		CXCR3-[³⁵ S]GTPγS		CXCR4-[³⁵ S]GTPγS		BEI ^b
		pK_b	σ_{obs}	pK_b	σ_{obs}	
12		7.29±0.20	0.16	8.25±0.25	0.06	19.97/22.60
13		7.20±0.29	0.11	7.71±0.30	0.03	25.26/27.05
14		6.96±0.39	0.38	7.74±0.31	0.02	24.59/27.35
15		6.53±0.37	0.26	7.84±0.37	0.08	24.28/29.14
16		8.42±0.16	0.12	8.72±0.13	0.04	24.13/24.99
17		7.62±0.25	0.14	7.75±0.22	0.03	22.75/23.13

^aCooperativity Factor;

^bBinding efficiency index ($BEI = \frac{pIC_{50} \text{ or } pK_i}{MW}$) for CXCR3/CXCR4;

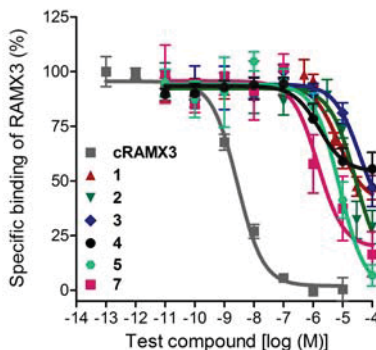


Figure 3.9: Dose-response curves of radioligand displacement assay. For the initial set of compounds, the dose-dependent displacement of RAMX3 from the CXCR3 is shown. The curves indicate that the compounds interfere with RAMX3 binding but cannot displace it completely. The dose-dependent binding of RAMX3 to the CXCR3 is shown as reference.

Table 3.5).

Lastly, we underlined the allosteric nature of our ligands by running the [^{35}S]GTP γ S assay in the absence of endogenous ligand, showing that the presented ligands have no effect in this case (see Figure 3.10).

In summary, we correctly predicted CXCR3- and CXCR4-selective binders with a very high success rate of 57 % and 50 %, respectively. Even more notably, two dual binders were correctly identified by docking. An additional similarity search around compound **7** lead to six more modulators.

3.2.4 Aggregator counter screening

To reduce the risk of false positive hits by compound aggregation^{62;63}, we used DLS (Dynamic Light Scattering) to identify potential aggregators. Moreover, we performed a fluorescence-based endothiapepsin assay as an orthogonal counter screen. At concentrations well above those used in the chemokine receptor assays, the vast majority of the compounds including, notably, the dual binders showed no sign of aggregation. Only compounds **8**, **9** and **10** and, to a mild degree, compound **6** showed autocorrelation

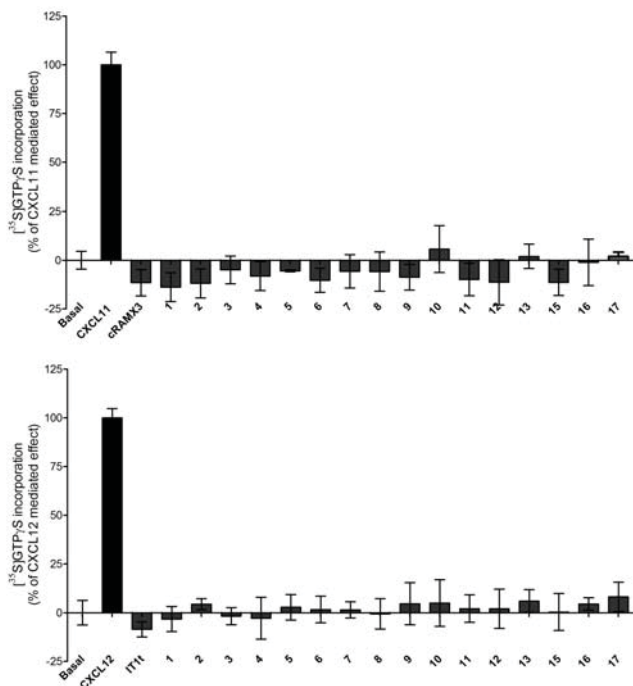


Figure 3.10: Effect of substances in absence of endogenous ligands. Compounds that have shown activity are screened in the [³⁵S]GTPγS assay again in the absence of endogenous ligands against the CXCR3 (upper panel) and the CXCR4 (lower panel). Additionally, the basal activity, the activity induced by the endogenous ligand (CXCL11 for CXCR3 and CXCL12 for CXCR4 and the activity induced by a known modulator (cRAMX3 for CXCR3 and IT1t for CXCR4) are shown as reference. Assays were carried out at 10 μM as described in the “Methods”.

curves and count rates that might be associated with aggregational behavior (see Figure 3.11). At the same time, only compounds **10** and **11** showed inhibition of endothiapepsin (see Figure 3.12). These findings would speak in favor of aggregation-based promiscuous inhibition. To investigate this in further detail, protein crystals of endothiapepsin were soaked with compounds **10** and **11**. While some difference electron density was visible in the binding site, it was not sufficiently resolved to place the ligands, which speaks for incomplete population of the ligands in the crystals. Altogether, except for compounds **10** and **11**, we can rule out that our ligands act via an aggregation-based mechanism under the investigated assay conditions. However, even for **10** and **11** this seems unlikely in the present assay, as neither compound shows any activity against the CXCR3 under the same assay conditions used for the CXCR4 and compound **11** shows no sign of aggregation in the DLS assay.

3.2.5 Summary

While docking to single targets works well⁶, the identification of ligands with predefined selectivity has only rarely been investigated. Knowing the binding patterns of ligands, however, is important to clearly understand their biological effects. In this study, we docked a large database of molecules to two targets and systematically selected candidates that would bind preferentially to only one or both targets.

The main result is that we were successful both in finding selective as well as non-selective ligands. Overall, we identified eleven novel ligands for the CXCR3/CXCR4 receptor pair. Given the high hit rates of 57 % (compounds **1–4**), 50 % (compounds **9–11**), and 50 % (compounds **6** and **7**), respectively, serendipitous discovery of ligands in each category can be regarded as unlikely.

At the same time, it can be concluded that the CXCR3 model we built does not suffer from template bias. Otherwise, we would have found many more potential dual binders, but a low hit rate in that category. It has to be emphasized again that finding a dual binder for this protein pair is

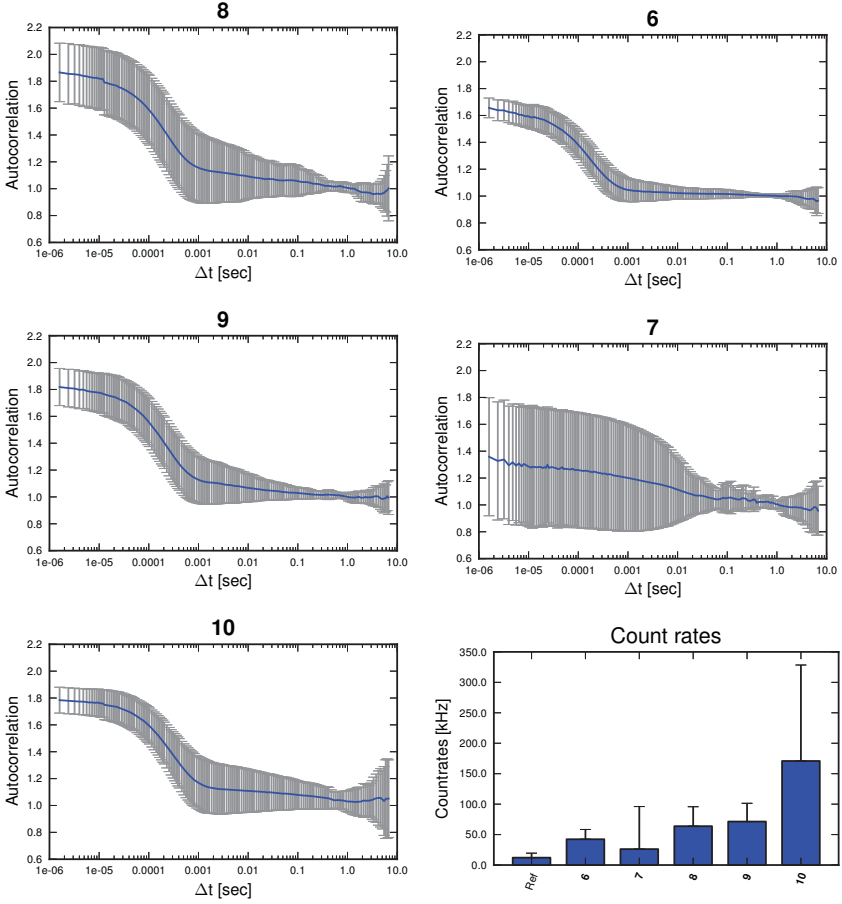


Figure 3.11: Results of DLS experiments. For selected compounds, the autocorrelation curves of the Dynamic Light Scattering experiments are shown in blue with error bars in gray. The corresponding absolute count rates are shown as barplot.

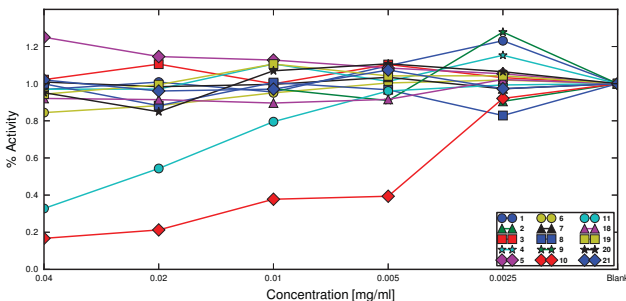


Figure 3.12: Inhibition of endothiapepsin in dependence of compound concentration as determined by a fluorescence assay. Values have been scaled by reference measurements containing no compound.

rather unlikely, as evidenced by the ligand sets obtained from ChEMBL and remarked previously by Pease & Horuk.⁶⁴

The identified ligands, furthermore, show remarkable binding constants mostly in the two-digit nanomolar range. In combination with the rather low molecular weights (see Figure 3.2), these yield very good binding efficiency indices⁶⁵ with an average of around 22, clearly above-average compared to the ChEMBL sets of known ligands (see Figure 3.13).

As mentioned in the Introduction, the CXCR4 tends to bind ligands with a stronger positive charge compared to the CXCR3. Since for a given molecule all precalculated protonation states from the ZINC database were docked, we can state that in the CXCR3, negatively charged or neutral species tend to be scored highest, while in the CXCR4, the stronger positively charged species of a molecule is preferred. As shown in Figure 3.6, this also applies for compound **7**, where the neutral species is the best-scoring one in the CXCR3 whereas it is the positively charged one in the CXCR4. Interestingly, this change in protonation is predicted to occur at the imidazole moiety and thus is chemically reasonable under assay conditions at pH 7.4, assuming no significant pK_a shifts due to local charge distributions in the binding site. Apart from the charges, the physico-chemical properties of the ligands presented here lie well within the distributions of

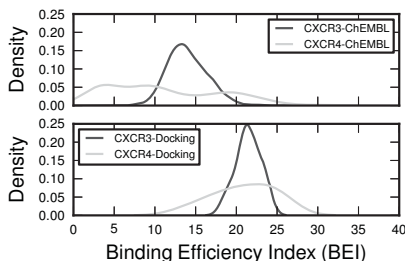


Figure 3.13: Distribution of BEIs (binding efficiency indices) for ligands with annotated IC_{50} s or K_i s taken from ChEMBL (upper panel) and ligands identified here (lower panel), separately for the CXCR3 (dark gray) and the CXCR4 (light gray). The relative frequency of BEIs is expressed as density, calculated by a kernel density estimator. For dual binders, the BEI values for both receptors were considered.

known ligands.

Additionally, we can draw very similar conclusions as in previous studies^{20;26;27} with respect to the general suitability of docking against GPCRs. First, in each category, high hit rates were obtained, sometimes even higher than in other studies investigating peptide-binding GPCRs.²⁶ Homology models also continue to be a good source of ligands, with the hit rate for the model of the CXCR3 being higher than the one for the CXCR4 in our study. Second, docking is able to identify compounds with chemistry previously undescribed for these targets. All but one compound identified herein have a Tanimoto similarity below 0.4 to any known CXCR3 or CXCR4 binder from ChEMBL (see Table 3.3).

Lastly, the importance of careful verification of the binding specificity of each ligand has to be emphasized. The risk of being overwhelmed by frequent hitters or aggregators is particularly high in settings like the present one, where compounds with a particular **non**selectivity for two receptors are desired. On the other hand, compounds should not be ruled out just because of failure in only one control assay, as shown by the fact that compounds that were flagged as potential aggregators in the DLS assay are in fact likely specific binders in our assay conditions.

The computational identification of ligands with tailored selectivity patterns represents a promising avenue towards the discovery of effective medications⁴ and precise tool compounds in chemical biology applications. For the receptors investigated, CXCR3 and CXCR4, the compounds described here might serve as seeds to further investigate the potential of dual binders as therapeutics. Future studies will show to what extent the panels of targets to be docked against can be expanded.

3.3 Methods

3.3.1 Homology Modeling

The sequence of the human CXCR3 (Uniprot: P49682, P42–Q345) was aligned to the sequence of the crystal structure of the CXCR4 (PDB: 3odu, P27–Q328)³² with T4-lysozyme residues (G900–S1201) removed, using the alignment service of uniprot.org. The sequence alignment was used as input for Modeller 9.10.⁶⁶ Homology Modeling was performed using multiple templates of the CXCR4 bound to the small-molecule modulator It1t (PDB: 3odu, 3oe6, 3oe8, 3oe9). One hundred models were generated and evaluated using the Structural Assessment service provided by Swissmodel⁶⁷ with QMEAN scoring. The eleven top-scoring models were selected for a subsequent loop and binding site remodeling step. For each of these eleven models, extracellular loops 2 and 3 (ECL2: A200–F207, ECL3: A285–C290) were resampled, resulting in 110 new models. Resampling was restricted in order to conserve disulfide bridges C124–C203 and C43–C290. These models were again evaluated using the Swissmodel server and the best-scored model of the overall 121 models was used for further optimization steps.

3.3.2 Model refinement

Protonation states were assigned using YASARA⁶⁸ to correspond to pH 7.0, optimizing the rotameric states for histidine, asparagine, and glu-

tamine residues at the same time. Using PLOP^{69;70}, unfavorable interactions were identified and automatically curated. We rebuilt ECL2, employing a two-step procedure: first, rebuilding A200–C203, followed by the larger part L190–C203. Spatial restraints were used to maintain the overall shape of the loop and to prevent it from collapsing into the transmembrane pocket. To optimize electrostatic interactions, selected residues were subjected to a local energy minimization (see Table 3.8 (“EM”)).

3.3.3 Optimization of transmembrane binding pocket

After global optimization of the homology model, the allosteric binding pocket of the receptor was optimized in more detail using six negative allosteric modulators with subnanomolar affinity for the CXCR3 (compounds **24–29**, Table 3.9). These molecules were docked to the model using Autodock Vina 1.1.1⁷¹, calculating nine binding poses per ligand. The superposition of all 54 binding poses was inspected and the two most abundant orientations were selected for further investigation. This excluded **27–29** from further analysis. For the remaining six models (three ligands with two binding poses each), the binding pocket was optimized using PLOP. The residues involved in minimization for each binding pose are listed in Table 3.8 (“BS1”, “BS2”). The final model was selected by cross-docking all three ligands to each of the six models and choosing the one with the best average score. We evaluated the final model by extracting ligands ($IC_{50} < 1 \mu M$) and decoys ($IC_{50} > 1 \mu M$) from ChEMBL and docking them to the model.

3.3.4 Docking & Re-Ranking

Docking was performed using DOCK 3.5.54.^{1;72–74} The docking spheres generated by sphgen were manually inspected and modified where necessary to achieve a more uniform distribution across the binding site. The ‘lead-like’ subset of ZINC version 11⁴⁶ was used, comprising around 2.4 M molecules at the time of download. The model of the CXCR3 and the crystal structure of the CXCR4 (PDB: 3odu) were prepared for docking,

Table 3.8: *Residues of CXCR3 that were subject to different refinement steps (“BS1”, “BS2”, “EM”) during model building.*

Residue	BW	BS1	BS2	EM
Gln45	—	x	-	x
Phe47	—	x	x	x
Ser48	—	x	x	-
Asp52	1.31	x	x	-
Arg53	1.32	x	x	x
Leu56	1.35	x	x	-
Tyr60	1.39	x	x	-
Thr105	2.56	x	x	-
Leu108	2.59	x	x	-
Trp109	2.60	x	x	-
Asp112	2.63	x	x	-
Ala113	2.64	x	x	-
Trp117	—	x	x	-
Cys124	3.25	x	x	-
Ala127	3.28	x	x	-
Gly128	3.29	x	x	-
Ala129	3.30	x	x	-
Phe131	3.32	x	x	-
Asn132	3.33	x	x	-
Phe135	3.36	x	x	-
Asp186	4.60	x	x	-
Leu190	4.64	x	x	-
His202	4.98	x	x	-
Cys203	—	x	x	-
Gln204	—	x	x	-
Tyr205	—	x	x	-
Arg212	—	x	x	-
Leu215	5.38	x	x	-
Arg216	5.39	-	x	x
Gln219	5.42	x	x	-
Trp268	6.48	-	x	-
Tyr271	6.51	x	x	-
Val275	6.55	-	x	-
Asp278	6.58	-	-	x
Asp297	7.32	-	-	x
Val298	7.33	x	-	-
Lys300	7.35	x	x	x
Ser301	7.36	x	x	-
Ser304	7.39	x	x	-
Gly305	7.40	x	x	-
Tyr308	7.43	x	x	-

Table 3.9: ChEMBL IDs of CXCR3 ligands used to refine the binding pocket of the homology model.

Cpd	ChEMBL ID
24	CHEMBL1077831
25	CHEMBL254083
26	CHEMBL403036
27	CHEMBL576097
28	CHEMBL578187
29	CHEMBL578192

optimizing hydrogen placement with CHARMM.⁷⁵ In order to identify dual binders, we developed a re-ranking procedure yielding the selectivity score S . S is defined as follows: given two dockings, the selectivity score of a compound only takes the relative ranking of this compound in both dockings into account. The relative rank is based on the rank of the compound (r_x) and the number of compounds docked in this docking campaign (m_x).

$$S_{R1,R2} = \frac{(R_1 - R_2)^2 + R_1 + R_2}{2} \quad (3.1)$$

with

$$R_x = \left(\frac{r_x - 1}{m_x - 1} \right) \quad (3.2)$$

where x denotes one of the two docking campaigns. The selectivity score penalizes unfavorable rankings in each docking as well as a high difference in rank and is scaled between 0 and 1 (see Figure 3.14).

3.3.5 Ligand selection and validation

Candidate molecules were selected from the 500 best-ranked molecules after repeated visual inspection in order to remove molecules that achieve artificially high ranks because of one of several deficiencies of docking scoring functions.²⁸ For all compounds, a ¹H-NMR (nuclear magnetic resonance) spectra of 3 mg compound in 1 mL DMSO (dimethyl sulfoxide)-D6 was determined using a JEOL ECA-500 machine after receiving them from their

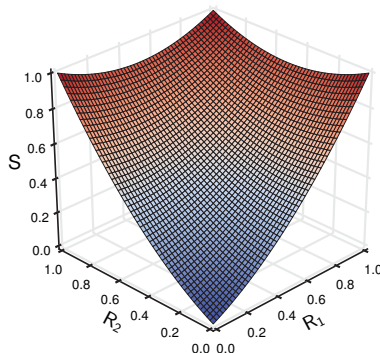


Figure 3.14: 3D plot of the rescoring function used to identify dual binders. Axis units show relative ranks and range from 0 (low rank) to 1 (high rank). Dual binders should have a selectivity score close to zero.

respective vendors (Vitas-M Laboratory, Ltd [compounds **1**, **2** and **19**]; Life Chemicals Inc. [compound **8**]; Princeton Biomolecular Research Inc. [compounds **10** and **11**]; Ukrorgsyntez Ltd. (UORSY) [compounds **12**, **13** and **20**]; ChemBridge Corporation [compound **23**]; Enamine Ltd. [compounds **3–7**, **14–18** and **21–23**]). Each spectrum was checked for consistency with theoretically expected peak distributions.

3.3.6 HEK membrane preparations

HEK (human embryonic kidney) cells were cultured in a 150-mm cell culture plate in DMEM (Dulbecco’s Modified Eagle’s Medium)/F-12 supplemented with 10 % (v/v) FBS (fetal bovine serum), 2 mM L-glutamine, 1 % (w/v) penicillin-streptomycin and incubated at 37 °C in humid atmosphere with 5 % CO₂. At 50 %–70 % confluency, cells were transiently transfected with 30 µg of CXCR3 or CXCR4 cDNA, using TransIT-293 transfection reagent (Mirus Corporation) and harvested 48 h after transfection. The medium was aspirated, cells washed once with ice cold phosphate-buffered saline (PBS, pH 7.4) and detached with harvest buffer (10 mM Tris-HCl, 0.5 mM EDTA (ethylenediaminetetraacetic acid), 5.4 mM KCl,

140 nM NaCl, pH 7.4). Cells were scraped, collected in a centrifuge tube and spun at 220 G for 8 min. The resulting pellet was resuspended in 5 mL of homogenate buffer (50 mM Tris-HCl, 5 mM EDTA, 1.5 mM CaCl_2 , 5 mM MgCl_2 , 5 mM KCl, 120 mM NaCl, pH 7.4) and subsequently lysed with an Ultraturrax. After additional centrifugation at 50 000 G, membranes were resuspended in binding buffer (50 mM Tris, 1 mM EDTA, 5 mM MgCl_2 , 100 $\mu\text{g mL}^{-1}$ bacitracin, 5 $\mu\text{g mL}^{-1}$ soybean trypsin inhibitor) and homogenized ten times with a glass-Teflon homogenizer at 4 °C. The homogenized membranes were shock-frozen in liquid nitrogen and stored at -80 °C. The protein concentration was determined with the Lowry method, using bovine serum albumin as standard.

3.3.7 [^{35}S]GTP γS incorporation assay

The [^{35}S]GTP γS incorporation assay was performed on membrane preparations of transiently transfected HEK293 cells that expressed either the CXCR3 or the CXCR4. The assay was carried out in 96-well plates at a final volume of 200 μL . The incubation buffer contained 20 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 10 mM MgCl_2 -hexahydrate, 100 mM NaCl and 70 mg L^{-1} saponin (pH 7.4). Membranes (30 $\mu\text{g mL}^{-1}$ of membrane protein), compounds and 10 μM GDP were preincubated for 30 min at 37 °C together with either 5 nM CXCL11 (CXCR3-specific chemokine) or 10 nM CXCL12 (CXCR4-specific chemokine). After the addition of 0.10 nM [^{35}S]GTP γS , membranes were incubated for an additional 30 min at 37 °C. Incubation was terminated by filtration through Whatman GF/B filters soaked with ice-cold PBS. The filter-bound radioactivity was measured as described above. Three to four experiments per compound were performed with each concentration in triplicate. K_b and $\alpha\beta$ values were obtained from compound dose-response curves at constant chemokine concentration.

3.3.8 Functional study analysis

To characterize the allosteric profile of our novel ligands, we applied the ternary complex model of allosterism to analyze the data obtained from the functional assays. The assumptions were that the allosteric modulators do not cause a depression of maximal response or a suppression of basal activity, as these are not accounted for in this model. Importantly, even if these assumptions did not hold entirely true for all the novel allosteric modulators, this analysis enables a first approximation and a semi-empirical estimate of cooperativity.⁷⁶ The data from functional studies, where discrete concentrations of agonists were used, were fitted to the following equations using Prism 5.0:

$$K_{app} = \frac{K_A \left(1 + \frac{[B]}{K_B}\right)}{1 + \frac{\alpha\beta[B]}{K_B}} \quad (3.3)$$

$$Y = \frac{Y_0 (1 + K_A)}{[c] + K_{app}} \quad (3.4)$$

where K_{app} describes the occupancy of the orthosteric site, K_A was the EC_{50} (half maximal effective concentration) value of CXCL11 or CXCL12 for CXCR3 and CXCR4, respectively, Y_0 the amount of agonist binding in the absence of modulator, Y is the agonist binding, $[c]$ the concentration of CXCL11 or CXCL12 used, $[B]$ the concentration of novel allosteric modulator, K_B the equilibrium dissociation constant of modulator binding, and $\alpha\beta$ the ternary complex constant, which denotes the cooperativity factor.^{40;76} Importantly, the analysis of activity data requires considering the cooperativity of allosteric and orthosteric ligand binding α as well as the cooperativity against effectors like G proteins β , although these cannot be separated. In contrast, in binding data analysis the cooperativity factor only takes α into account. Values of $\alpha\beta > 1$ denote positive cooperativity, whereas $\alpha\beta < 1$ denotes negative cooperativity. Values of $\alpha\beta$ approaching 0 are indistinguishable from competitive antagonism. In that case, the K_B values approaches the K_i value.⁴⁰ In the absence of cooperativity, $\alpha\beta = 1$

applies. In the [^{35}S]GTP γ S accumulation assay, the K_A value for CXCR3 was set to 1.55 nM and to 0.70 nM for CXCR4. The concentration of the chemokines CXCL11 and CXCL12 were set to 5 nM and 10 nM, respectively.

3.3.9 Ligand similarities and properties

Ligand similarities (SIM) presented throughout the chapter represent the Tanimoto coefficient between the ECFP4-like (extended-connectivity fingerprint) Morgan fingerprints as implemented in RDkit.⁷⁷ Ligand properties were calculated using OpenEye’s MolProp toolkit.⁴⁷

3.3.10 Dynamic Light Scattering

DMSO stocks of each compound were diluted with filtered binding buffer to a final compound concentration of $4 \times 10^{-3} \text{ mg mL}^{-1}$ ($>10 \mu\text{M}$) and 1 % DMSO. Light scattering was measured using a Xtal-Concepts Spectro Size 300 machine at a scattering angle of 90° with an emission wavelength of 660 nm. For each compound, a maximum of 20 samples was recorded with a measurement length of 10 s each.

3.3.11 Endothiapepsin Assay

The endothiapepsin inhibition assay was carried out as described elsewhere.⁷⁸ A DMSO stock solution with 4 mg mL^{-1} was used, leading to a maximum compound concentration of $4 \times 10^{-2} \text{ mg mL}^{-1}$ ($>100 \mu\text{M}$) in the well.

4 Development of Tool

Compounds Targeting CXCR4

Isoforms in Zebrafish (*Danio Rerio*)

In collaboration with Dr. Arndt Siekmann, Max-Planck Institute for Molecular Biomedicine, Münster, Germany.

Contributions: Arndt Siekmann contributed the initial idea and carried out all experiments in zebrafish.

4.1 Introduction

As has been described before, the CXC receptor 4 or its malfunction are involved in infections or diseases, making the receptor an interesting target for pharmaceutical research. However, CXCR4 also plays important roles in embryogenic processes as extensively studied in different model animals.

In mice, for example, CXCR4 and its ligand CXCL12 (SDF-1), in concert with other signaling pathways, are responsible for different hematopoietic processes such as B-cell lymphopoiesis and bone-marrow myelopoiesis⁷⁹ or for growth-related processes such as the vascularization of the gastrointestinal tract.⁸⁰

Another widely used *in vivo* model organism, besides the mouse, is the zebrafish (*danio rerio*). It has desirable attributes such as the fast embry-

onic development and the optical clarity, which allows the application of examination techniques such as fluorescent labeling and staining in live animals. As a consequence, numerous experimental procedures for genomic analysis and alteration have been developed for zebrafish, further increasing the experimental value of this model organism.

Interestingly, due to a potential gene duplication event, zebrafish possess two isoforms of the CXCR4 and consequently two ligand isoforms, termed CXCR4a and CXCR4b and SDF-1b (CXCL12b) and SDF-1a (CXCL12a), respectively. Both isoforms are highly similar, with CXCL12a (99 amino acids, 11.3 kDa) and CXCL12b (97 amino acids, 11.3 kDa) sharing more than 73% identical residues. Similarly, their receptors (360 & 353 amino acids, 40.2 & 39.4 kDa) have a sequence identity higher than 72%. Interestingly, the major contribution to receptor-ligand selectivity can be assigned to a single position, position 33, which is an asparagine (N33, CXCL12a) and a serine (S33, CXCL12b), respectively. Yet, these subtle differences in concert with different expression patterns, predominantly contribute to the subfunctionalization of the two receptor/ligand copies⁸¹, which has been investigated in several studies.

As in many other species, it could be determined that these CXCR4 isoforms and their respective ligands are involved in the guidance of migrating cells by chemotaxis. In case of the latter receptor-ligand pair, CXCR4b/CXCL12a, the cells guided can be primordial germ cells, which are directed to the somatic site of the latter gonads.⁸² This differentiation process is independent of CXCR4a and SDF-1b, which, in contrast, plays a major role in the growth of blood vessels starting from existing ones.

The formation of the so-called LDA (lateral dorsal aorta), forms an important step in zebrafish embryogenesis and is achieved in a bidirectional growth process, where the anterior LDA grows in posterior direction, while the posterior LDA grows anteriorly, finally fusing and forming a continuous vessel, as revealed by *in vivo* fluorescent techniques.⁸³ During this process, CXCL12b is expressed in the endoderm, a structure ventral to the LDA, while its receptor is solely expressed in the anterior LDA, emphasizing the complexity of expression patterns and networks involved during

such developmental steps.

In parallel to the growth of the LDA, a second sprouting process can be observed that forms the bilateral PHBCs (primordial hindbrain channels), a venous structure in the zebrafish hindbrain. This process seems to be independent of CXCR4 signaling, since knockout mutants show a normal phenotype. However, at a later stage of embryogenesis, starting at 30 hpf (hours post-fertilization), another process of angiogenic sprouting in zebrafish hindbrain is evidently under the control of CXCR4.⁸⁴ Originating from the bilateral PHBCs, several CtAs (central arteries) start growing and branching in an arch-shaped manner. These CtAs can form interconnections before finally fusing with the medial-located BA (basilar artery). During this cross-linking of venous and arterial structures, CXCR4a is expressed in the tips of sprouting CtAs while SDF-1b is secreted near the BA, thus guiding the direction of growth. Upon connection of the CtAs to the BA, the resulting blood flow down-regulates CXCR4 expression and in turn, stops further growing. In knockdown mutants lacking CXCR4a or SDF-1b, the CtAs emerge from the PHBCs the same way, but the growth is undirected and stochastically driven, leading to an increased number of CtA interconnections and a decreased number of drains into the BA.

These examples from zebrafish nicely underline that CXCR4, in either of its isoforms, has an important function in the proper development during embryogenesis. Of course, CXCR4 signaling is not solely responsible for these mechanisms but this receptor is only part of a complex network and interplays with other targets such as VEGF and Notch.⁸⁴ The complexity of these signaling networks not only arises from the number of participating structures, but also spatial (CXCR4 has to be expressed in the anterior LDA but not in the posterior LDA) and temporal (LDA formation and BA connection occur at different times in developmental cycle) factors, do play a critical role.

One major drawback of the genomic methods used to investigate the above described mechanisms is that usually these techniques only allow gene silencing without further spatio-temporal control, i.e. they do not allow silencing at a certain developmental stage or in a tissue-specific manner

or if they do, the respective experimental setup is very complex. Consequently, certain phenotypes could be the result of a secondary effect instead of a direct effect related to the knockout of that gene. A small-molecule tool compound might, in contrast, allow the rapid and reversible inhibition and activation of specific targets and in turn allow a more fine-tuned control over the experimental setup.

In collaboration with the group of Dr. Siekmann (Max-Planck Institute for Molecular Biomedicine, Münster, Germany), I started the development of such tool compounds modulating CXCR4, preferably selective for either CXCR4a or CXCR4b of the zebrafish. Since neither of these structures had been determined experimentally, two homology models had to be built based on solved GPCR crystal structures. Subsequently, we docked the ZINC^{46;85} lead-like subset to these models and selected molecules whose docked poses showed favorable interactions with the receptor models. These molecules were tested *in vivo* by incubating the zebrafish embryos together with the respective compound and monitoring the phenotype during the following developmental stages, with the aim to identify bioactive compounds showing a CXCR4 subtype-related phenotype.

4.2 Results

4.2.1 Homology Models and Docking

The final homology models were assessed in an unbiased way using the ModEval Server, which calculates the GA341⁸⁶ and the normalized DOPE (z-DOPE) score⁸⁷ (Table 4.1). Both scores try to score the quality of a structure based on pairwise potentials, but while the first one positively correlates with the model quality, a better model is indicated by a more negative z-DOPE score. The homology models perform slightly worse than the crystal structure but lie within a comparable range with respect to these quality scores. Furthermore, a manual comparison of the crystal structure of human CXCR4 that was used as template with the built homology mod-

Table 4.1: *Structure quality assessment of CXCR4a, CXCR4b and the used template.^a*

	hCXCR4 template	CXCR4a model	CXCR4b model
GA341	0.998	0.825	0.869
z-DOPE	-0.592	-0.461	-0.364

^aAssessment scores were calculated using the ModEval server. For GA341, scores > 0 indicate good models while for z-DOPE scores < 0 are preferred. For non-membrane proteins, scores above 1 or below -1, respectively, indicate a native-like structure.

els clearly shows the similarity between the structures and that the homology models successfully resemble the main features of the template (see Figure 4.1). These features include position 7.39 (Ballesteros-Weinstein notation⁵²), where the conserved glutamic acid is responsible for forming a key interaction with a bound ligand.⁵⁴ Furthermore, the respective aromatic zippers⁵⁵ have positions 3.32 and 6.51 in common, while the tyrosine at position 7.43 in zebrafish variants offers an additional hydrogen bonding capability compared to the phenylalanine in human CXCR4. Additionally, the aromatic zipper is extended by a phenylalanine in CXCR4a and CXCR4b. After building and validating the homology models of both isoforms, the ZINC lead-like subset^{46;85} was docked to each of the models as described in the “Methods” section. From each of the two docking runs, compounds were selected from the top-scoring poses based on manual inspection. In total, 42 substances were independently selected for subsequent experimental testing, 21 for each of the models (see Table 4.2).

4.2.2 In vivo assay

All compounds were applied to zebrafish during embryogenesis and larvae were monitored for abnormal phenotypes. For all but one compound, the larvae either underwent a normal embryogenetic cycle or died. Given the latter case, the applied doses of test compound were lowered until a non-lethal dose was reached. In all of these cases, no phenotypic differences could be observed once such a non-lethal dose was reached. No attempt was made

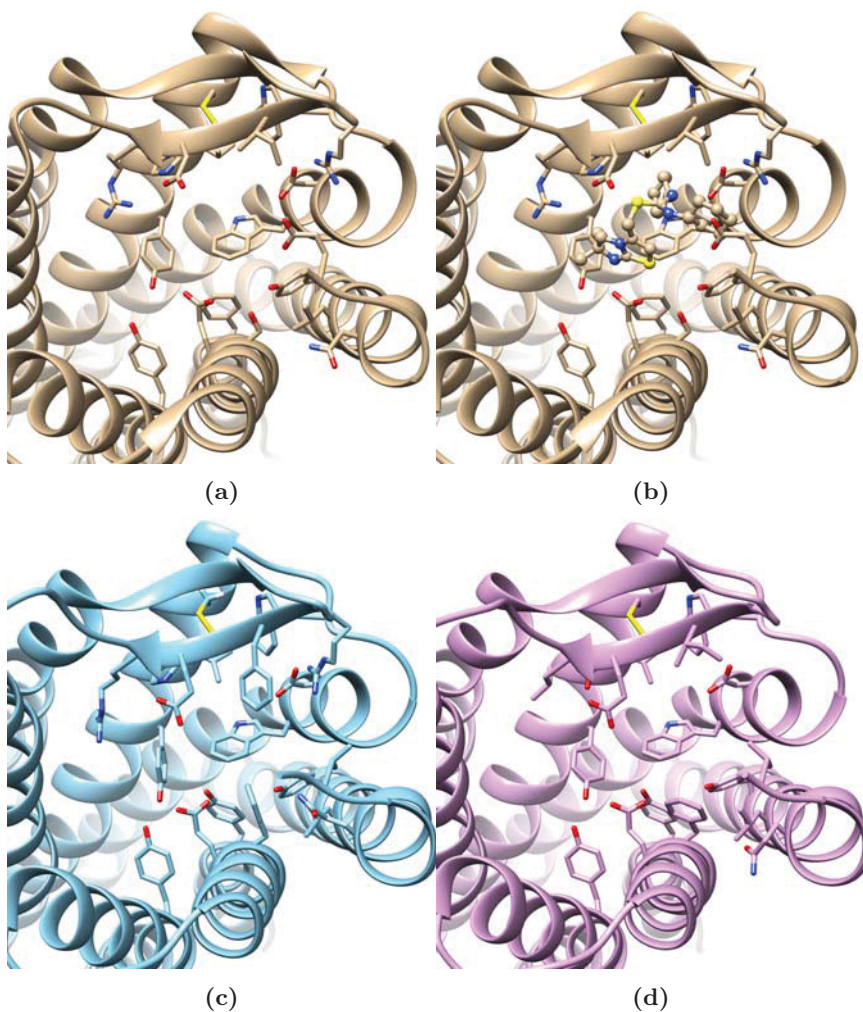


Figure 4.1: Comparison of the human CXCR4 crystal structure (a+b, light brown) and the created homology models of zebrafish CXCR4a (c, cyan) and CXCR4b (d, magenta), respectively. Receptors in ribbon representation. Residues surrounding binding site in stick representation. (a) Human CXCR4 (PDB 3odu). (b) Human CXCR4 (PDB 3odu) with co-crystallized ligand ITD as stick&stick. (c) Homology model of zebrafish CXCR4a. (d) Homology model of zebrafish CXCR4b.

to further investigate these lethal compounds, since these effects could well be target-mediated, i.e. by blocking off-targets, as well as of chemical nature, i.e. the change of pH value or osmotic pressure or similar at concentrations as high as 100 μ M. One single compound, C72447712, reproducibly introduced cranial bleeding in zebrafish larvae as shown in Figure 4.2. However, this specific phenotype differs from the phenotype introduced by CXCR4a/b knockdown.

4.3 Discussion

As described in the results, the models built here do resemble the main features of the (human) CXCR4 binding site. However, it has to be noted here that no ligand information, that is, information about known binders, were used to refine the receptor binding site around the poses of these compounds. It has been shown that the use of such additional information can substantially improve the quality, in terms of retrieval rate, of a homology model, even when the binding poses used in such a process are not experimentally supported but purely modeled.⁸⁸ Unfortunately, the amount of ligand data on zebrafish CXCR4a/b is sparse, at best. In ChEMBL⁴⁵ for example, not a single compound is listed binding to either CXCR4a or CXCR4b (as of Dec 2014).

This fact might be the reason that only one compound out of 42 showed a biological activity, which corresponds to a low hit rate, compared to other studies applying docking techniques to GPCRs.²⁵ This hit rate must be considered however within the context of the setup used herein. All compounds which were predicted to be affine to either of the receptors were predicted based on homology models, which is more challenging than using an experimentally determined structure, despite the fact that homology models can be successfully used within such a setup. The above described lack of ligand data that could be used to refine the model has to be taken into account as well.

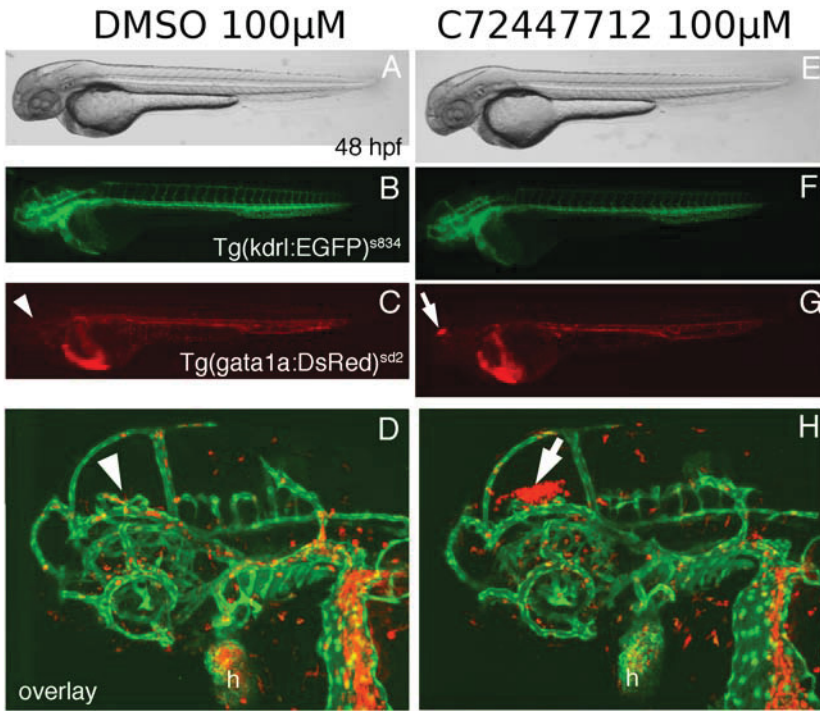


Figure 4.2: Substance C72447712 leads to cranial bleedings in zebrafish embryos. Embryos were treated with 100 μ M DMSO as control (left column) or with substance C72447712 (right column). (A) Brightfield image of control embryo at 48 hpf. (B) *Tg(kdrl:EGFP)^{s843}* transgene shows blood vessels in green. (C) *Tg(gata1a:DsRed)^{sd2}* transgene shows red blood cells. Arrowhead marks brain blood vessels. (D) Confocal zoom in image of brain blood vessels marked by *Tg(kdrl:EGFP)^{s843}* transgene and red blood cells marked by *Tg(gata1a:DsRed)^{sd2}*. Arrowhead indicates anterior brain blood vessels. (E) Brightfield image of C72447712 treated embryo at 48 hpf. (F) *Tg(kdrl:EGFP)^{s843}* transgene shows blood vessels in green. (G) *Tg(gata1a:DsRed)^{sd2}* transgene shows red blood cells. Arrow indicates cranial bleeding. (H) Confocal zoom in image of brain blood vessels marked by *Tg(kdrl:EGFP)^{s843}* transgene and red blood cells marked by *Tg(gata1a:DsRed)^{sd2}*. Arrow indicates cranial hemorrhage. Hpf: hours post fertilization

Secondly, and maybe more importantly, the *in vivo* experiments used to verify the predictions made have some disadvantages compared to *in vitro* assay. First of all, although GPCRs are membrane receptors and as such do not require the ligand to diffuse through the membrane, it must be assumed that the ligand concentration throughout the tissue of the zebrafish larvae is non-uniformly distributed. The respective diffusion constants, of course, may vary between the single test substances. In turn, this means that the final bioactive ligand concentration, i.e. the number of ligands interacting with the respective receptor, is undefined and may be way smaller than the concentration applied. The same limitations, of course, apply for the update rates of the test substances into the larvae. Altogether, it is possible that a compound cannot interact with the receptor in reasonable high concentrations although the *ex vivo* concentration is very high. By design, the molecule library used for docking should contain structures that have a high probability of being bioavailable. However, this assumption is based on the calculation of simple physico-chemical descriptors and thus bioavailability cannot be guaranteed.

Furthermore, while there’s usually a defined readout in *in vitro* experiments, which can be detected in a quantitative manner, the definition of such a quantitative measure can be difficult in *in vivo* models. For example considering the fusion of the anterior and posterior lateral dorsal aorta as described above.

All these aspects do contribute to the difficulty in predicting bioactive substances and proving such bioactivity. In the given case, however, we do see a biological effect, but this effect has not been described before to be associated with CXCR4 knockdown. The observed bleeding is located in the hindbrain, anterior to the bilateral PHBCs (see Figure 4.2), while the latter as well as the fusion of the LDA seems to be unaffected, in contrast to a CXCR4 knockdown. Of note, the fact that C72447712 shows an effect different from a CXCR4 knockdown does not exclude the possibility that this effect is mediated via a directed CXCR4 interaction. In fact, the heterogeneous ligand distribution (spatially and temporally) and the (potentially) incomplete down-regulation compared to a knockout might

well lead to the exhibition of a different effect, which is supported by the fact that the observed phenotype involves the integrity of blood vessels, for which a correctly functional CXCR4 is crucial. Moreover, C72447712 could act as an agonist, for which the phenotypic effect has not been described.

Nevertheless, the possible interaction of C72447712 with an off-target is plausible and has not been neglected. Instead, SEA⁸⁹ has been used to predict potential targets other than CXCR4, yielding the murine GABA (γ -Aminobutyric acid) transporter 1 as potential off-target with an E-value of 2.7×10^{-4} using ChEMBL⁴⁵ version 16 as database and ECFP4 as fingerprint metric (as of Dec 2014). Since no ligands are annotated for zebrafish targets in ChEMBL, SEA cannot yield any zebrafish-specific targets. Searching the ZFIN database⁹⁰ did however not yield a GABA transporter 1 with a phenotype comparable to the one observed.

In summary, the application of modeling and virtual screening to two zebrafish variants of CXCR4 lead to a single compound C72447712 which induced a cranial bleeding in the zebrafish hindbrain. This might well be the result of an interaction with CXCR4 despite the fact that this new phenotype differs from those induced by CXCR4 knockdowns. The achieved hit rate is low and might be the result of a lack of ligand information that could be used to shape the binding site of the built models. It could improve iteratively by feeding back experimental data into the model, thereby refining the models. Furthermore, even correct predictions might be considered false negatives if the interaction is hindered by differing spatial-temporal distribution of CXCR4 and respective test substance or the insufficient uptake of the latter into the larvae. If an existing interaction is not observed since the morphological effect cannot be detected by the experimental setup used, the application of different gene markers or staining techniques might be required. Consequently, an in-depth analysis might yield additional hits than the one observed.

Although the desired small-molecule tool compound could not be identified in this first round of screening, the detailed analysis of the selected compounds, the iterative improvement of the used models, the potential off-target interaction and the new, before described phenotype pose inter-

esting challenges for subsequent research projects.

4.4 Material & Methods

4.4.1 Homology Modeling & Refinement

The sequence of CXCR4a (entry F1QCB2) and CXCR4b (entry Q9PTF7) were retrieved from the Uniprot database.⁹¹ These sequences were aligned against the sequence of the crystal structure of the human CXCR4 (PDB 3ODU) using Muscle v3.8.31.⁹² The resulting alignment was manually truncated to those amino acids which had been crystallographically resolved. Then, three homology models were built of each of the two receptors using Modeller⁶⁶ v9.10. During modeling, artificial restraints were applied to residues 201–205 of CXCR4a and residues 193–197 of CXCR4b, which were defined to be part of an α -helix (TM-V).

The initial homology models were all protonated using PropKa^{93;94} 3.0 as implemented in PDB2PQR⁹⁵ 1.8 at pH 7.0. These hydrogens were minimized using PLOP^{69;70}, followed by the calculation of energies and the identification of structural clashes as defined by PLOP. In a second minimization step, the sidechains of these clashing residues were unconstrained, while all remaining atoms were kept fixed. Subsequently, the best model was manually selected for each of the two targets. The initial model of CXCR4a was used for a first round of docking and experimental testing and due to the lack of results, both models were separately inspected and further optimized.

For CXCR4a several sidechains close to the binding site were selected, for which the orientations were resampled using PLOP. These residues were Leu47 (leucine), Tyr51, Val118 (valine), Tyr122, Leu126, Phe191, Glu193 (glutamic acid), Arg194 (arginine), Arg209, Tyr265, Ile269 (isoleucine), Asp272 (aspartic acid), Leu290, Gln291 (glutamine) and Phe295. The resulting structure underwent two final minimization steps, the first one involving residues 196–204, forming the N-terminal part of TM-V and a final all-atom relaxation using a high RMSG (root-mean-square gradient)

of 0.5 as termination criterion.

For CXCR4b, as for CXCR4a, several sidechains were manually selected subsequently and their conformations automatically resampled using PLOP. These were Leu28, Val30, Gln35, Asn111 (asparagine), Asp169, Ile183, Glu185, Leu186, Thr187 (threonine), Lys197 (lysine), Arg201, Asp265, Glu284 and Phe288. The final minimization steps included Tyr188 followed by an all-atom relaxation as described for CXCR4a.

4.4.2 Docking

Virtual screening was carried out using DOCK 3.6^{1;72-74} with the ZINC^{46;85} lead-like subset as input. The necessary input spheres for DOCK were generated from sphgen. Since for homology models no crystallographic ligand information are available, an artificial ligand was built to define the binding site for docking. For this artificial ligand, the crystallographic data of Arg2, Aln3 and Dpr16 (D-proline) of the cyclic 16-mer peptide (PDB 3oe0) and the small-molecule ITD (PDB 3odu) were used. These atomic positions were solely selected to represent the complete spatial extent of the binding pocket. This artificial ligand did not influence the sampling or scoring process in any kind. Consequently, sphgen was run using the “uselig sph=no” setting.

From the respective docking runs, the top scoring molecules were manually inspected and evaluated and promising candidates were selected.

4.4.3 Compound Acquisition

All compounds were acquired from Enamine Ltd. and ChemBridge Corporation as indicated in Table 4.2.

4.4.4 *In vivo* screening

Embryos were dechorionated at 26 hours post fertilization (hpf) and incubated in either 100 μ M DMSO (vehicle) or 100 μ M of test substance at standard conditions. Solutions were prepared from 50 mM stocks in

Table 4.2: *Compounds selected for experimental testing.*^a

	Zinc-ID	Vendor	Vendor-ID	Model
1	C14541589	Chembridge	37918222	CXCR4b
2	C19733246	Chembridge	56561462	CXCR4b
3	C19810479	Chembridge	64204583	CXCR4b
4	C23368589	Chembridge	48301499	CXCR4b
5	C32859998	Enamine	Z203876164	CXCR4a
6	C32860075	Enamine	Z203880752	CXCR4a
7	C45814912	Enamine	Z507420516	CXCR4a
8	C47676702	Enamine	Z595833238	CXCR4a
9	C48018216	Enamine	Z433968290	CXCR4a
10	C56890235	Enamine	Z748134498	CXCR4a
11	C58304919	Enamine	Z1097113944	CXCR4a
12	C58329494	Enamine	Z1104856873	CXCR4b
13	C65435638	Chembridge	67775968	CXCR4a
14	C67688459	Chembridge	31651906	CXCR4a
15	C67712788	Chembridge	59060042	CXCR4b
16	C67724182	Chembridge	36244854	CXCR4b
17	C67774860	Chembridge	67542592	CXCR4b
18	C67884655	Chembridge	54121352	CXCR4b
19	C67921334	Chembridge	86052064	CXCR4b
20	C67983484	Chembridge	98321975	CXCR4a
21	C69390116	Enamine	Z1139549946	CXCR4b
22	C69716774	Enamine	Z1088496416	CXCR4a
23	C69848777	Enamine	Z1139513069	CXCR4a
24	C69952155	Enamine	Z1102775653	CXCR4a
25	C70457919	Chembridge	56912524	CXCR4a
26	C71405573	Chembridge	12508662	CXCR4b
27	C71768012	Chembridge	91924514	CXCR4a
28	C72148370	Chembridge	62916307	CXCR4a
29	C72155873	Chembridge	72062625	CXCR4b
30	C72156022	Chembridge	72312322	CXCR4b
31	C72162162	Chembridge	52465319	CXCR4a
32	C72162456	Chembridge	53236042	CXCR4a
33	C72165221	Chembridge	81726638	CXCR4b
34	C72166712	Chembridge	85293936	CXCR4a
35	C72280919	Enamine	Z1183935251	CXCR4a
36	C72285240	Enamine	Z1207958007	CXCR4b
37	C72447712	Chembridge	63825925	CXCR4b
38	C73271416	Enamine	Z1314157363	CXCR4b
39	C77285475	Enamine	Z1252568433	CXCR4b
40	C77505538	Chembridge	85126517	CXCR4b
41	C78491589	Enamine	Z103788666	CXCR4a
42	C79074287	Enamine	Z446174310	CXCR4b

^a Shown is the compound ID, the vendor the substance was acquired from, the vendor's compound id and the model the compound was selected from.

DMSO. Imaging was performed on a Leica stereomicroscope or on a Leica SP5 confocal microscope. Confocal images were analyzed using Imaris software (Bitplane). For images from the stereomicroscope, Volocity software was used.

5 Docking to Crystallized and Modeled CCR5 Receptor Conformations

Contributions: Second messenger assays were carried out by the contract research organization DiscoverX Corporation, Fremont, CA, USA.

5.1 Introduction

The β -chemokine receptor CCR5 is another prominent example of a receptor from the chemokine receptor family with high pharmaceutical relevance. As chemokine receptor it is highly expressed on leukocytes and is responsible for the triggering of chemotaxis. Four major endogenous chemokine agonists are known, which are RANTES (“regulated on activation, normal T cell expressed and secreted”, CCL5), the macrophage inflammatory proteins-1 α and -1 β (MIP-1 α and MIP-1 β , CCL3 and CCL4) and the monocyte chemoattractant protein-2 (MCP-2, CCL8), however, evidence exists for interactions with other signaling agents.⁹⁶

CCR5 has a proven role in rheumatoid arthritis, graft rejection, neurodegenerative diseases and asthma, but the main reason it has been in the spotlight of research for several years now is its ability to trigger internalization, which can be exploited by HI virions to enter and infect the host cell.⁹⁷ As a result of the intense research activities on this target, several high-affinity ligands binding to CCR5 are known, including, but not limited

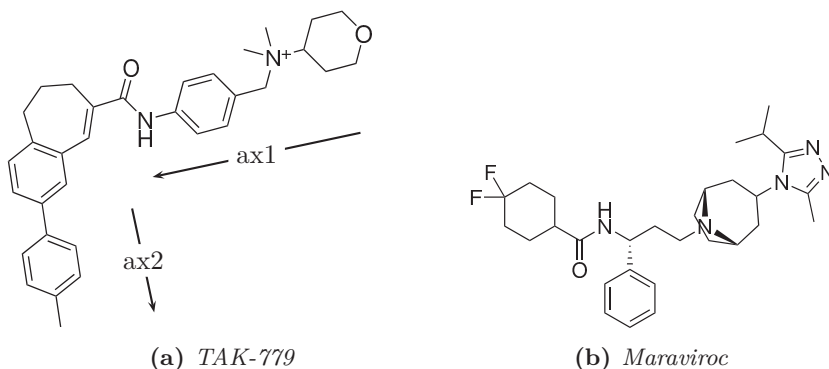


Figure 5.1: Chemical structures of known CCR5 inhibitors a) TAK-779, shown with axes for orientation. Implications of those axes are discussed in the text. b) Maraviroc

to, TAK-779, developed by Takeda Inc.⁹⁸ (see Figure 5.1a) and UK-427857 (see Figure 5.1b), which was developed from a high-throughput screening hit at Pfizer.⁹⁹ Since 2007, UK-427857 is approved by the FDA (U.S. Food And Drug Administration) as HIV entry inhibitor named Maraviroc.

At the beginning of the project outlined here, no crystal structure of CCR5 was publicly available. Consequently, the aim of this work was to develop a homology model of CCR5 using one or several known binders to model a highly accurate binding site and to use this model for subsequent virtual screenings and potential SAR (structure-activity relationship) studies. However, close to the finalization of the model, the CCR5 crystal structure was solved and published with Maraviroc, an inverse agonist, bound. As a consequence, the outline of this project was realigned.

Instead of considering the building of a homology model as obsolete, this new situation was considered as a chance to directly compare an unbiased homology model with the respective crystal structure. This comparison is not limited to structural differences between model and crystal structure but also the docking runs, i.e. the ranking of the ligands was compared.

In addition, the binding site of the crystal structure was remodeled, resulting in a third structure, which was able to host TAK-779. This was based on the observation that the crystal of CCR5 structure bound to Maraviroc was unlikely to be able to bind TAK-779 in the observed receptor conformation. Through detailed inspection of the crystal structure combined with proposed interaction modes of TAK-779 derived from experimental data, a single phenylalanine side-chain could be identified as gatekeeper residue, closing a large subpocket located deeper within the helix bundle. By flipping this residue to a new conformation, the respective subpocket was opened and able to host part of TAK-779 upon binding. To refine this third structure, TAK-779 was manually placed into the structure, the complex was modified where considered necessary, and subsequently minimized. The whole cycle was iterated several times to create a final complex as near as possible to a native conformation of the receptor bound to TAK-779.

Each of the three structures was then subjected to large-scale virtual screening, during which the ZINC lead-like subset was docked to every receptor. Subsequently, promising candidates were selected from each docking run. Selected compounds were finally evaluated for their *in vitro* activity against the CCR5.

5.2 Results & Discussion

5.2.1 Homology Modeling of CCR5

The homology model of CCR5 was built using three distinct crystal structures as templates. All three templates had a sequence identity above 30 %, which can be considered a lower threshold of sequence identity for which the respective structures can be expected to have to same fold¹⁰⁰ (see Table 5.3). Initially, 20 models were generated using Modeller⁶⁶ which were manually inspected and the most promising one selected. This model was then manually refined in an iterative process as described in detail in the “Methods” section. An unbiased homology model was created, i.e. no

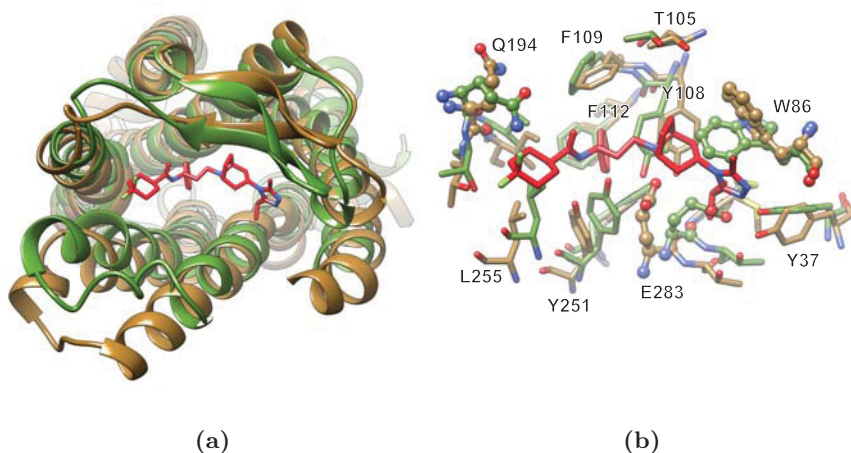


Figure 5.2: *Overlay of the CCR5 crystal structure (brown) and the homology model (green). The crystal ligand Maraviroc is shown in red. View from the extracellular side onto the plane of the membrane (a) Ribbon representation (b) Stick representation of the binding site. Trp86^{2,60} (tryptophan), Glu283^{7,39} and Gln194^{5,40} are highlighted by ball-and-stick representation. Residues are labeled where feasible.*

ligand structure was used throughout the refinement process of the model, although this is known to potentially improve the quality of a homology model.^{34;88} The reasoning was that any ligand information used would introduce a bias towards (using Maraviroc) or away (using any other ligand) from the crystal structure. Nevertheless, the final homology model nicely resembled the overall structure of CCR5 as shown in Figure 5.2a, with a calculated RMSD (root-mean-square distance) of the backbone atoms of 2.6 Å. Major differences include the N-terminus (not depicted in Figure 5.2) which is not forming a distinct secondary structure in the homology model, due to the lack of an appropriate template. A typical issue in homology models of GPCRs. Furthermore, the β -hairpin structure of ECL2 is more “closed” in the model, thereby covering the transmembrane binding site. At its C-terminal end, the crystallized conformation of this loop shows an

unusual turn structure without internal hydrogen bonds that could stabilize the turn (see Figure 5.3a). A closer inspection of the electron density of the crystal structure highlights that the structure assignment is not unambiguous due to the experimental uncertainty at the given resolution of 2.7 Å. An alternative loop conformation can easily be modeled that fits the electron density equally well and at the same time brings the carbonyl oxygen of Pro183 (proline) and the amide nitrogen of Gln186 in hydrogen bonding distance (see Figure 5.3b). Also, the overall shape of the loop might well be influenced by the crystal packing with the symmetry related unit (see Figure 5.3, light brown), leading to a bend conformation of the loop in contrast to the extended conformation in the homology model (see Figure 5.3c). Due to the in-between distance, a strong influence of this loop region on the binding pocket is unlikely. Yet, these two aspects nicely highlight that have to be considered when working with crystal structures.

The most apparent structural deviation is at the extracellular parts of TM-VI and TM-VII, where TM-VI is slightly more ordered and forms a more ideal α -helix in the model, while the first turn of TM-VII is strongly disordered. Also, the modeled helices tend to be bend towards the center of the structure, leading to a smaller binding site.

A more detailed comparison of the binding sites (see Figure 5.2b) underlines the successful modeling, since most residues have comparable side-chain orientations in the model and the crystal structure. Notable exceptions are highlighted in Figure 5.2b by a ball-and-stick representation and include Trp86^{2,60}, Glu283^{7,39} and Gln194^{5,40}. In the experimentally determined structure, Trp86^{2,60} is in an “upright” conformation while the modeled conformation is flipped by around 90°, i.e. the indole moiety is close to parallel to the layer of the membrane. The latter conformation would interfere with the binding of Maraviroc but is the observed conformation in CXCR4 (PDB 3odu³²), thus both conformations seem possible. This turned conformation closes a small subpocket hosting the isopropyl moiety of Maraviroc and thus changes the overall shape of the pocket.

As a side note, the salt bridge between Glu283^{7,39} and the azabicyclo

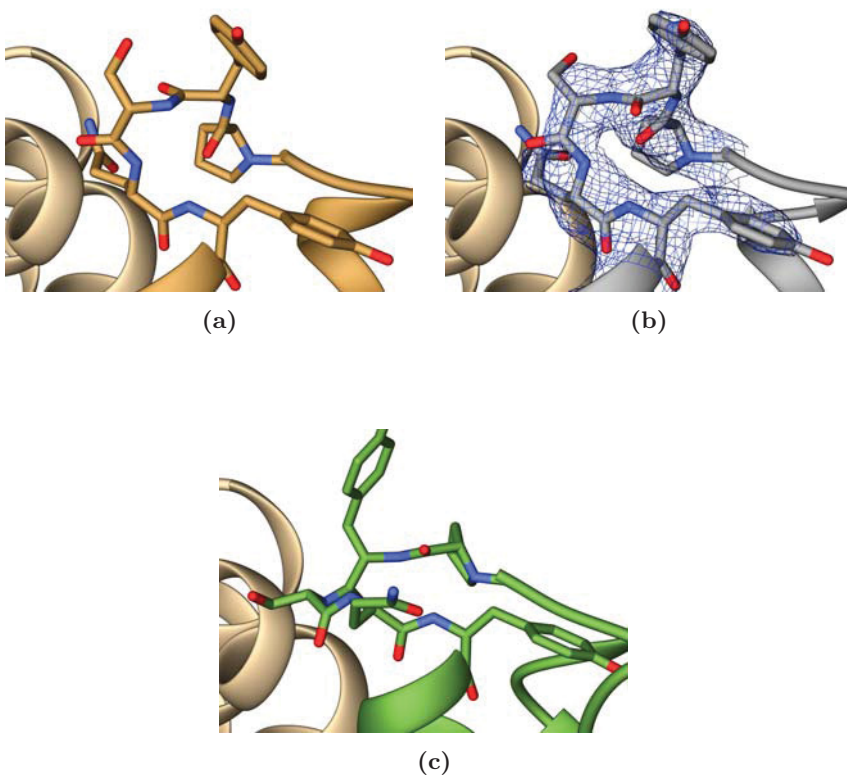


Figure 5.3: *The CCR5 crystal structure shows an unusual turn structure between ECL2 and TM-V without stabilizing hydrogen bonds. (a) Crystal structure of CCR5 (PDB 4mbs, brown) with symmetry related unit (light brown). (b) Alternative loop conformation (gray) brings Pro183 and Gln186 in hydrogen-bonding distance. Electron density (blue) from PDB 4mbs at a σ -level of 1. (c) Extended loop conformation in homology model (green) might interfere with symmetry related unit (light brown).*

(tropanyl) moiety is a nice example of an electrostatic interaction which is embedded in a hydrophobic environment, in the given case the isopropyl and the benzyl moiety of Maraviroc. The hydrophobic environment shields the interaction from other polar groups, thereby decreasing the dielectric constant ϵ and consequently increasing the enthalpic contribution compared to a salt bridge which is solvent exposed.¹⁰¹ Such a shielded salt bridge may be beneficial for a long residence time, which has been measured to be around 16 h for Maraviroc bound to CCR5.¹⁰²

This specific glutamate adopts a different side-chain orientation in the model which is closer to Tyr37, presumably caused by the minimization protocol used, where the glutamate was allowed to move freely, i.e. unconstrained, during the last minimization cycle. Since no ligand was used within this refinement, the final structure is biased towards a conformation where the negative charge is at least partly compensated, achieved by the hydrogen bond to Tyr37-OH. Finally, also Gln194^{5,40} was modeled pointing into the binding pocket and thus offers additional hydrogen bonding possibilities. In combination with the aforementioned inwards movement of TM-VI and TM-VII, this conformation contributes to the size reduction of the binding pocket.

Altogether, it can be assumed that the templates and the sequence alignment used for model building were in general correct, as concluded from the successful overlay of the binding site residues. Furthermore, the model is structurally close to the experimentally determined one in terms of RMSD. Using Maraviroc as ligand might have increased the similarity even more, given that the binding mode of Maraviroc would have been correctly predicted. The model also reproduces the positions of the majority of the binding site residues with few exceptions including the known key residue Glu283^{7,39} as well as Trp86^{2,60}. The “ligand-free” optimization protocol used led to a shrunk binding site, which might be the major drawback of the developed model.

5.2.2 Reshaping the CCR5 binding site

Besides Maraviroc as FDA-approved CCR5 inverse agonist, a number of CCR5 modulating lead structures have been identified, including TAK-779, a CCR5 antagonist which has been developed already 15 years ago (see Figure 5.1a). Until now, several works have been published proposing binding modes of TAK-779 in complex with CCR5 based on experimental data like site-directed mutagenesis.^{103–106} However, crystallographic data have not been available at this point in time. Remarkably, although the sets of proposed key residues involved in ligand binding overlap throughout the mentioned articles, the authors propose different binding modes. With the availability of the crystal structure, a new opportunity emerges to derive a more accurate model of the bound complex. Based on previous publications and a comparison with the crystal structure it was assumed that (i) the tertiary amine binds to Glu283^{7,39} (ii) the oxane moiety points into the minor binding pocket between TM-I, TM-II, TM-III and TM-VII, comparable to the 3-methyl-5-isopropyl-1,2,3-triazol function in Maraviroc and (iii) the position of the amide bonds are comparable between Maraviroc and TAK-779. Interestingly, the conformational flexibility of TAK-779 is restrained by the cycloheptene ring to have a nearly 90° kink in the three-dimensional geometry of the molecule. This is indicated by the axes in Figures 5.1a & 5.4 and the superposition of sampled ligand conformations (see Figure 5.4). This, in concert with the assumptions made before, lead to the conclusion that the hydrophobic tail of this ligand points either out of the binding site towards the extracellular side of the cell (as it has indeed been proposed before¹⁰⁴) or deeper into the transmembrane helix bundle. While it seems unlikely that a high-affinity ligand has a hydrophobic group pointing towards the solvent, the latter orientation is hindered by the limited depth of the pocket which is formed by Tyr108^{3,32}, Phe112^{3,36} and Trp248^{6,48}. This depression does not offer more space than for the phenyl moiety of Maraviroc. Delving into the buried volumes within the solved structure reveals a small, hidden cavity capped by Phe112^{3,36} in the extracellular direction. By choosing a different phenylalanine side-

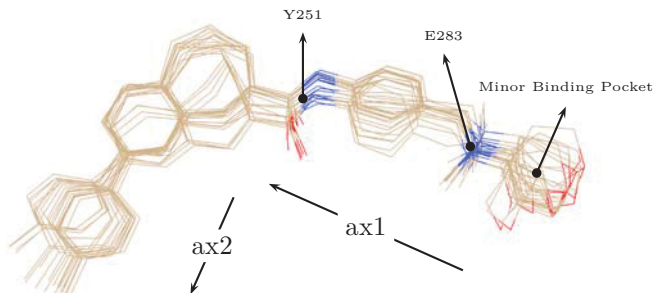


Figure 5.4: Sampled ligand conformations of TAK-779 are superimposed to highlight its kinked overall conformation. Axes as in Figure 5.1. Arrows indicate potential interactions between ligand moieties and CCR5. These assumptions were used to model TAK-779 bound to CCR5. Ligand conformations calculated using Corina^{107–109}.

chain rotamer, this pocket opens towards the ligand binding site and allows TAK-779 to bind deeply within the transmembrane helix bundle (see Figure 5.5). According to the Dunbrack library¹¹⁰ used, the selected rotamer ($\chi_1 \approx -80^\circ$, $\chi_2 \approx 100^\circ$) has a considerably lower probability of occurrence with 13% compared to the original rotamer ($\chi_1 \approx 180^\circ$, $\chi_2 \approx 80^\circ$) with 76%. Still, it is the second most abundant orientation of Phenylalanine observed, indicating that this newly created conformation is near-native. Concurrently, while the side chain flip did cause minor movements of the neighboring Phe79^{2,53}, it did not introduce a severe sterical clash. The completely remodeled and minimized structure shows an **all-atom** RMSD of only 2.35 Å, accentuating that this conformation could potentially be adopted by CCR5 *in vivo* and thus might have a biological relevance. Finally, the resulting pocket is not only able to host TAK-779, but there is still unoccupied space in the newly formed pocket which explains studies on chemical derivatives of TAK-779, showing that even extended 4-methyl-biphenyl motifs can bind to CCR5 within a nanomolar affinity range.¹¹¹

The generated model of CCR5 in a TAK-779-bound conformation was subsequently used for a virtual screening campaign as described below,

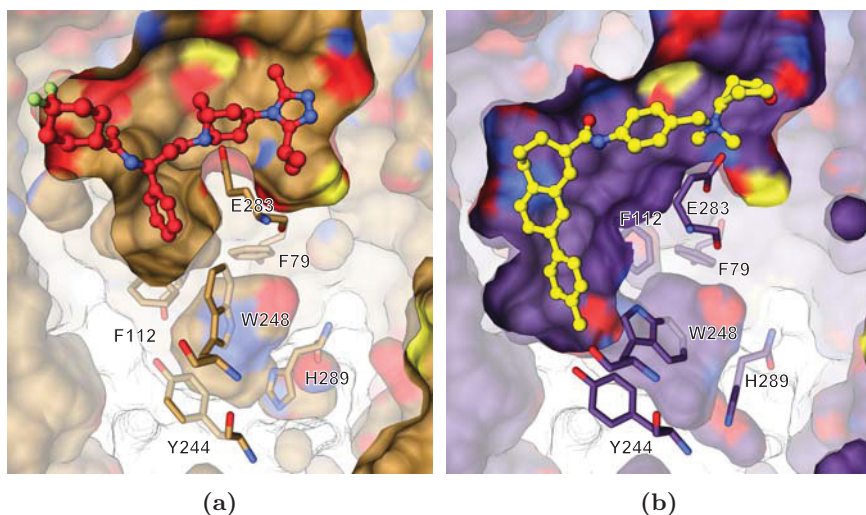


Figure 5.5: Comparison of the CCR5 crystal structure (brown) with the reshaped model (purple). Side view parallel to the membrane. Receptor in surface representation cut open perpendicular to the line of sight. Ligands in ball-and-stick representation. Light areas indicate where surface is hidden by the cut plane. (a) Crystal structure (brown) bound to Maraviroc (red), (b) Reshaped model (purple) bound to TAK-779 (yellow). By turning the F112 side chain, a new binding pocket becomes accessible.

convinced that this model will enrich different chemotypes upon docking compared to the remaining two structures used. However, although the complex was modeled very carefully and likely resembles the overall biological interaction between CCR5 and TAK-779, it has to be emphasized that there is room for further improvement. A shift of TAK-779 along axis “ax1” and simultaneously along axis “ax2” would lead to a counterclockwise rotation of TAK-779 within the binding site, thereby resolving minor issues. These issues are the slightly stranded tetrahydropyran moiety, which has been forced into a boat conformation during the minimization. The above described motion of the ligand could allow it to adopt a more relaxed conformation and potentially flip into a chair conformation. As an additional side effect, the quaternary amine would move closer to the center of the binding pocket and thus allow Glu283^{7,39} to adopt a side-chain orientation closer to the one found in the crystal structure. The benzene ring of the 6,7-dihydro-5H-benzocycloheptene substructure would then form π -stacking interactions with Phe112^{3,36} instead of the terminal 4-methyl-phenyl group. The latter one would move deeper into the transient pocket and could form aromatic interactions with Tyr244^{6,44} and Trp248^{6,48}. Though they have of course been considered, these proposed interactions are purely hypothetical and could not be reproduced using the minimization protocol described herein. This protocol consisted of the manual movement of the ligand in the bound complex alternating with an energy minimization. These steps were repeatedly iterated until convergence was reached. At this point, the energy minimization caused the ligand to move back into the initial orientation after a manual movement of the ligand towards the desired binding mode. Consequently, it has to be assumed that (i) the minimization procedure converged at the stable receptor-ligand complex or (ii) the minimization trajectory is trapped at a local optimum and further improvements require major side-chain reorientations or moves of the backbone structure. These major structural changes might be necessary, although the overall change in orientation of the ligand within the binding site might be small. For the further use of the structure, the first case is assumed but the possible existence of a spatially

close but energetically preferred structure must not be neglected. However, overcoming the structural barrier to find this potential structure might require more sophisticated simulation techniques such as steered molecular dynamics simulations. Such a simulation would allow to apply an artificial force along the axes “ax1” and “ax2”, beyond the possibilities of the simple minimization protocol used here. This could allow the ligand to overcome the energy barrier of the local minimum and to move into the above described position. Also, these simulations can take explicit solvent effects and the influence of an artificial membrane into account that is required to elucidate significant changes on receptor geometry.^{112;113} Due to the complex nature of such simulations, the correct setup is time-consuming and requires future work.

In summary, a CCR5 model was developed bound to its antagonist TAK-779 starting from the crystal structure of CCR5, solved in complex with Maraviroc, an inverse agonist. Phe112^{3.36} was identified as a gatekeeper residue, capping a hidden subpocket deep within the transmembrane helix bundle. By flipping the phenylalanine aside, this hidden pocket opens and allows the long hydrophobic tail of TAK-779 to bind to the CCR5. This newly identified void also explains how chemical derivatives of TAK-779 with extended hydrophobic tails can bind to the receptor with nanomolar affinities as determined by Junker and coworkers.¹¹¹ The development of a receptor conformation that is presumably able to bind these ligands consequently allows the future development of quantitative SARs to better understand which chemical entities strengthen and weaken the ligand-receptors interactions. Molecular simulations would furthermore help to understand the dynamics of receptor-ligand interactions beyond this static view and might eventually yield even better models of the CCR5-TAK-779 complex, which could not be developed using the modeling procedure outlined here. Finally, sufficiently sophisticated simulations might give an insight into the mechanisms causing TAK-779 to act as antagonist and Maraviroc being an inverse agonist. Altogether, a number of new and exciting questions arose from this project for future research topics, starting with a virtual screening approach as presented in this work.

5.2.3 Virtual Screening against different CCR5 structures

Each of the three above-described structures, i.e. the crystal structure, the homology model and the crystal structure with a reshaped binding site, were used for large-scale virtual screening. The complete ZINC lead-like subset was used as ligand set and all dockings were run in an unbiased fashion and independently of each other, i.e. no filtering rules or pharmacophore models were used to reduce the number of ligands. The whole set of necessary calculations required a large amount of computational resource, in total more than 23 300 CPU (Central Processing Unit) hours were spent on the docking calculations alone, distributed over about 250 CPU cores.

After docking, the best 500 scored molecules for each single docking run were inspected. For the docking to the model and the reshaped structure, candidate molecules were selected from these sets of molecules. For the docking to the crystal structure, however, the poses were post-processed after the docking calculations and molecules were filtered as described below. As depicted in Figure 5.6, around 70 out of the 100 best-scoring molecules from the docking to the crystal structure have a formal charge of +2 or higher. Furthermore, the corresponding calculated binding modes do show an overemphasized charge-based interaction with Glu283^{7,39}, as exemplarily shown in Figure 5.7, where four hydrogen bonds are formed with Glu283^{7,39} as acceptor, two of which are charge-assisted. One possible reason might be inherent in the docking process or more accurately in the treatment of charges in general. Starting from the prepared protein, electrostatic grids are precalculated to allow a fast evaluation of sampled binding poses.

Consequently, polarization effects are not taken into account, i.e. given a certain charge on the protein side, a docking score increases with additional opposing charges placed close by on the ligand side, without taking into account the influence of each single charge on the whole electrostatic environment. This holds true for docking in general and the aforementioned drawbacks and inaccuracies are accepted in the interest of a gain in docking

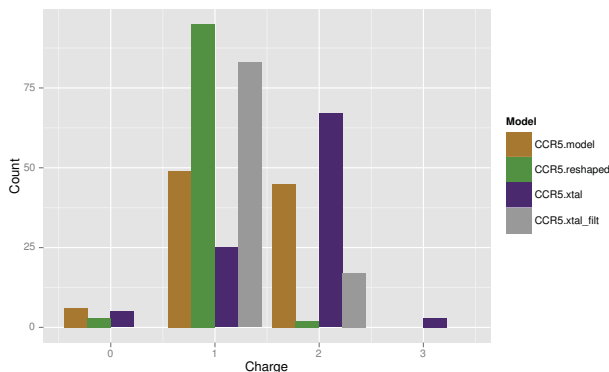


Figure 5.6: Distribution of formal charges for the top scoring molecules for each of the docking runs. The crystal structure tends to enrich more strongly charged molecules compared to the other receptor conformations used.

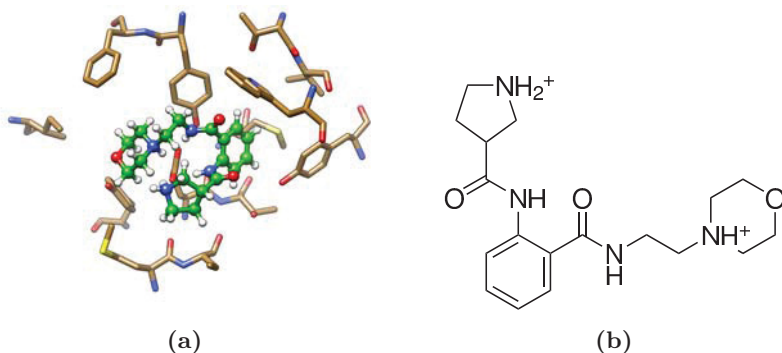


Figure 5.7: Example of a docking pose that has been dismissed during the post filtering. (a) Close-up view of binding site of the CCR5 crystal structure (brown, stick representation) with a exemplarily selected docking pose (green, ball-and-stick representation). The docked molecule shows an overemphasized, charge-based interactions with E283^{7,39}. The pyrrolidine and the morpholino moiety are both charged and form hydrogen bonds with E283^{7,39}, as do both amide binds. (b) Chemical structure of the molecule shown in (a).

speed, usually without causing major problems. In case of CCR5, however, the binding site offers a spatially exposed charge, Glu283^{7,39}, which can be addressed by ligands from various directions at the same time. In contrast, the glutamate has moved in the homology model and the reshaped crystal structure during the minimization process as described before (see Figures 5.2 & 5.5). In fact, the summed solvent-accessible surface of oxygen atoms in ϵ -position is about 39 Å² in the crystal structure, but only 14 Å² and 8 Å² in the reshaped crystal structure and the homology model, respectively.

This specific exposed glutamate of the crystal structure might be the reason for the, presumably artificial, enrichment of strongly positively charged ligands. Without compensating penalties on the docking score, ligands with multiple protonatable groups and their corresponding poses are enriched during docking. There are different possible solution approaches for that. In theory, one could change the partial point charges on the protein side or the dielectric properties during the grid calculations. However, this would also lower the score of a single charged ligand and effectively change the overall weighting of polar and apolar docking scores. Instead, the generated poses were post-processed in a filtering step. Using Chimera, the docking pose of every docked ligand was discarded if it maintained more than one hydrogen bond with Glu283^{7,39}.

It can be debated whether such a kind of filtering should occur during the pose sampling step in contrast to filtering the final (single) pose(s). For a given molecule there might exist a docking pose fulfilling the defined filtering criteria that is not stored due to a score just slightly worse than a pose retained from sampling. When filtering final poses, this molecule would be completely discarded but when filtering already during the generation of poses, only a certain pose would be discarded and a new, possibly more valid one, in lieu could be sampled. Thus, when filtering final poses, potential hits might not be identified as such, hence the alternative approach might be beneficial. It does on the other hand side massively increase the computational effort since much more poses have to be evaluated and furthermore it can be reasoned that in a large-scale virtual screening one is

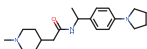
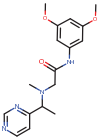
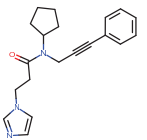
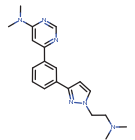
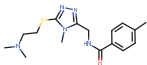
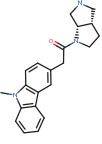
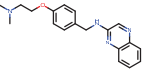
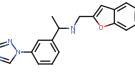
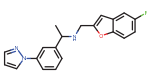
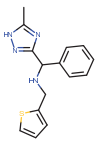
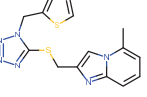
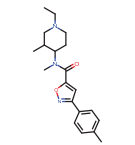
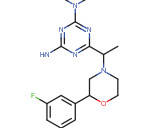
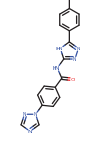
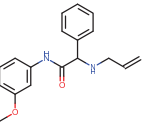
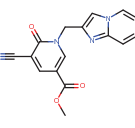
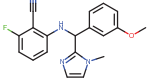
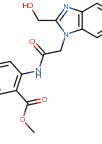
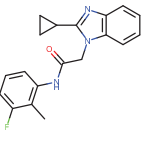
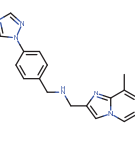
not interested in a particular ligand, but in the identification of any binding molecule. As such, an increased false negative rate is not problematic as long as sufficient true positives are identified.

5.2.4 Compound Validation

After the finalization of the docking process, 20 compounds were selected for evaluation in biological experiments (see Table 5.1). Of these, five have been selected from the docking to the crystal structure, eight from the docking to the reshaped crystal structure which offers a new binding site deep in the transmembrane helix bundle and another seven from the homology model. All substances were subjected to an initial screening at a single concentration of 10 μ M using a commercially screening service. Using a β -galactosidase complementation assay, which traces changes in cAMP (cyclic adenosine monophosphate) level caused by Adenylate Cyclase activity, each compound was investigated for its effect as agonist as well as antagonist. Based on the data of the plate control, the standard deviation can be estimated separately for the 0% and the 100% values of the assay, as discussed in detail in the “Methods” section. For further analysis, the higher and thus more conservative $\sigma_{100\%}$ uncertainties will be used, which are around 6.4% for the agonist assay and 6.8% for the antagonist assay. Importantly, this uncertainty expressed as propagated standard deviation is solely estimated from the experimental repeatability, without taking any biological effects into account.

The results of this primary screen are depicted in Figure 5.8. Overall, the ligands tested in this first screening approach do not exceed a 30% efficacy. However, several ligands do show a distinct biological effect beyond the noise in the data. The antagonist screen reveals seven substances that exceed the 1σ level and of these, three have an average measured efficacy beyond the more strict and expressive 2σ value. These three are DS49232, DS50192 and DS83957, with 19.5%, 19.1% and 29.9% efficacy, respectively. These efficacies correspond to 2.9σ , 2.8σ and 4.4σ above 0%, rendering a random effect very unlikely. More specifically, these σ devia-

Table 5.1: Structures selected from docking for experimental evaluation

1	DS49232 ccr5.resaped	2	DS10403 ccr5.xtal	3	DS57404 ccr5.resaped	4	DS55622 ccr5.resaped
							
5	DS62705 ccr5.resaped	6	DS67459 ccr5.resaped	7	DS05596 ccr5.resaped	8	DS11585 ccr5.xtal
							
9	DS50192 ccr5.xtal	10	DS89674 ccr5.xtal	11	DS81651 ccr5.model	12	DS83957 ccr5.resaped
							
13	DS84666 ccr5.model	14	DS55435 ccr5.resaped	15	DS00794 ccr5.xtal	16	DS75993 ccr5.model
							
17	DS29983 ccr5.model	18	DS62535 ccr5.model	19	DS23844 ccr5.model	20	DS50263 ccr5.model
							

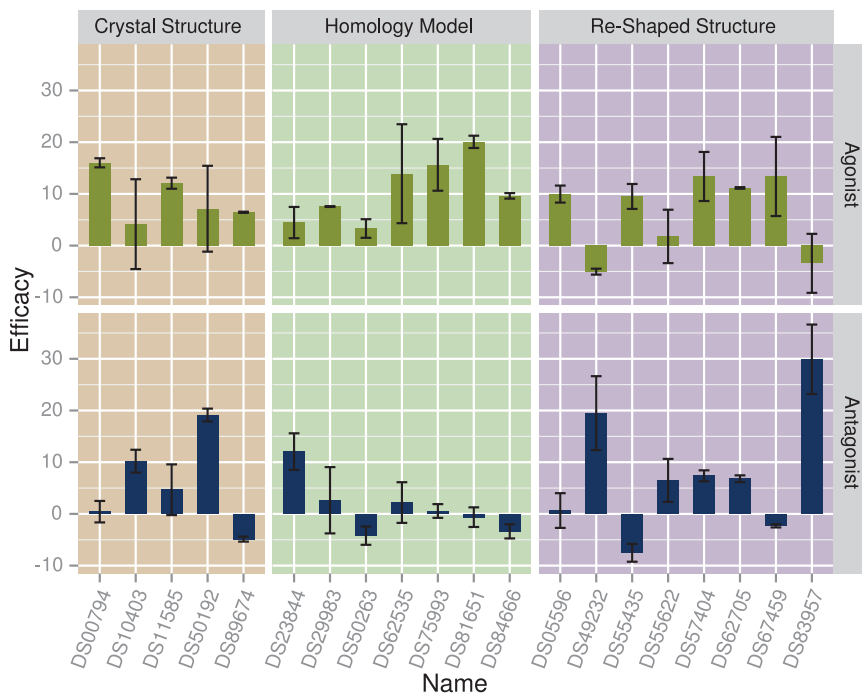


Figure 5.8: Graphical representation of the results of the primary screen. For every substance tested, measured efficacy is shown for the agonist screening (upper panel, yellow) and the antagonist screening (lower panel, blue), separately for compounds selected from the docking to the crystal structure (left panel, brown), the homology model (middle panel, green) and the re-shaped model (right panel, purple).

tions indicate that the probability of a measurement to show such a high effect without an underlying biological effect is extremely small. In addition, the measured effects as agonists for these compounds are very close to 1σ (DS50192) or even negative (DS49232, DS83957) which means that these results do not contradict those from the antagonist assay.

Compared to the former results, the measured efficacies of the agonist apparently tend to be higher (thirteen compounds show efficacies above 1σ) and deviate stronger from the mean, which is also reflected by the higher $\sigma_0\%$ value. Out of the thirteen compounds mentioned, six show an average effect higher than twice the background standard deviation with DS81651, exhibiting the strongest agonistic effect of 20.1 % ($>3\sigma$). At the same time, this compound does not show an antagonistic effect and the single measurements are close to each other, i.e. the measurement shows a high repeatability, similar to DS00794. In contrast to that, DS75993 and DS57404 show a high, DS62535 and DS67459 even a very high spread between the two measurements. In case of DS57404, the compound in addition shows a slight antagonistic effect which is sufficiently small to be considered false positive, yet it might also contradict the agonistic mode of action. In total, the results of the agonist assay are less conclusive than those of the antagonistic screening because of the higher baseline activity and the stronger diverging measurements. Still, even when dismissing DS62535 and DS67459 because of the high deviation between the repeats of the measurements, four substances selected from the virtual screening campaigns show an agonistic effect as proven by the single point measurement discussed above. Three additional compounds present antagonistic behaviors which are much more pronounced when compared to the effects of the remaining compounds than in the agonist assay.

Interestingly, there are indications for a certain correlation between the mode of action of a compound and the virtual screening run it was selected from. In Figure 5.9 the experimentally determined efficacies from Figure 5.8 have been rearranged to highlight this fact. The data of the agonist screening (Figure 5.8, top row & Figure 5.9, left column) and the antagonist screening (Figure 5.8, bottom row & Figure 5.9, right column)

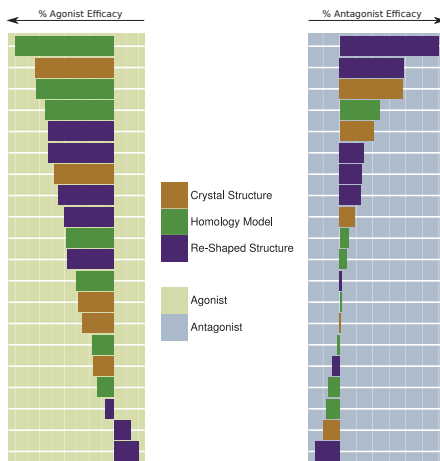


Figure 5.9: Graphical representation of the results of the primary screen, re-organized from Figure 5.8. Measured efficacies are shown for the agonist screening (left panel, yellow) and the antagonist screening (right panel, blue). Within each panel, efficacies are sorted in decreasing order and colored by the docking strategy they were identified with.

have been separately sorted by value and colored according to the structure/docking run the compound was selected from. This sort of data presentation nicely shows that three out of the four highest values in the agonist assay have been measured for compounds selected from the docking to the homology model. Again, this model has been built and refined without using any ligand structure. In contrast, the two most potent antagonists have been selected from the docking to the re-shaped crystal structure, which bears ligand information of TAK-779, a CCR5 antagonist. Of course, this assumption has to be considered with caution, since the number of compounds tested is too small to be able to draw final conclusions and, of course, there is a certain amount of experimenter's bias introduced through the process of ligand selection. Still, this observation strongly underlines the demand for subsequent experiments, both *in silico* and *in vitro*, addressing questions on receptor conformations and the mode of action as pointed out earlier.

Table 5.2: *Comprehensive results of the second screening (dose-response screening)*

Compound	Type	RC ₅₀ [$\mu\text{mol L}^{-1}$]	Hill	Max. Response [%]
DS00794	Agonist	> 100	—	23.7
DS55435	Agonist	> 100	—	15.6
DS67459	Agonist	59.16	1.9	83.4
DS75993	Agonist	> 100	—	16.9
DS81651	Agonist	32.98	1.3	23.4
DS49232	Antagonist	> 100	—	0.0
DS50192	Antagonist	> 100	—	0.1
DS83957	Antagonist	> 100	—	0.1

From the first round of screening, which used single point measurements, the most promising candidates were selected to undergo a second screening campaign. This time, we recorded dose-response curves using the same assay system as described before. DS49232, DS50192 and DS89674 were selected for a screening in an antagonist assay system. For the corresponding agonistic screen, DS00794, DS55435, DS67459, DS75993 and DS81651 were selected. The screening of DS57404 and DS62535 was waived in favor of DS55435. The reasoning is that, although DS55435 is only an averagely strong agonist, it simultaneously shows the strongest negative antagonistic effect, vindicating a more detailed investigation of that substance. A negative antagonistic effect indicates that the receptor activity is higher as the reference activity used (80%), corresponding an agonistic effect, whereas a negative agonistic effect corresponds to an inverse agonist. The comprehensive results from this second screening are listed in Table 5.2. Merely for DS67459 and DS81651 dose response curves could be fitted, depicted in Figure 5.10, and subsequent EC₅₀ values derived. Those values are located in the range of 30 $\mu\text{mol L}^{-1}$ –60 $\mu\text{mol L}^{-1}$, which is the order of magnitude one has to expect when dealing with primary hits from a virtual screening campaign, especially considering (i) the peptidic nature of endogenous chemokine receptor modulators, (ii) the fact that these substances are partly derived from homology models and (iii) the

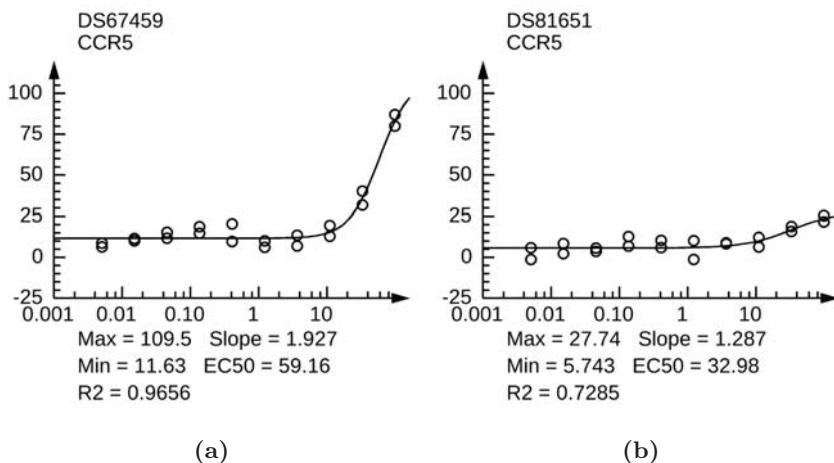


Figure 5.10: Selected dose-response curves of second screening for (a) DS67459 and (b) DS81651.

size of the compounds, which is limited to 350 Da by the screening library used. The same arguments apply when discussing the maximum biological effect, which is around 83 % for DS67459 and 23 % for DS81651. Additionally, the agonistic effect does not necessarily correlate with the binding affinity. Instead, even nanomolar binders can behave as partial agonists. Moreover, the measured response in general not only depends on the nature of the ligands and their concentrations but also on various equilibrium constants and concentrations within the whole signaling pathway. Examples include the equilibrium dissociation constant of the signaling pathway activation, i.e. the concentration of receptor-ligand complex that causes 50 % response.¹¹⁴ This value can be interpreted as strength of the signal amplification of incoming stimulus, i.e. ligand binding, and is strongly correlated with the density of expressed GPCRs on the cell surface. The ratio of these two values directly, and strongly, influence the apparent potency and maximum response.¹¹⁴ In other words, the measured efficacy might change using a different batch of cells. In the given case it has to be added that full saturation has not been achieved using the maximum concentra-

tion of $100\text{ }\mu\text{mol L}^{-1}$ according to the dose response curves. Thus, the final effect could be higher, which would influence the recorded EC_{50} accordingly. The remaining potential agonists show an overall maximum effect between 16 and 24 %, which is comparable to the primary screening but without a dose-dependent behavior. The potential antagonists tested do not show any effect at all.

In summary, out of the five potential agonists tested, three could not be confirmed by dose response screening, including DS55435, which was selected for the second screening campaign despite the rather low agonistic effect in the primary screen. Conversely, two substances detected in the primary screen show dose-dependent effects and can thus be considered as preliminary hits.

At the same time, none of the potential antagonists could be verified, although the exact same stock solutions were used for the subsequent dose response screening as for the primary screen. This is rather unexpected since the primary antagonist screen had a lower baseline and the potential hits were more distinct from this baseline compared to the agonist screening. Furthermore, the most promising candidate of all, DS83957, induced a response of almost 30 % in the primary screening, which is highly unlikely to be observed without any biological interaction of the ligand with the assay system, based on the standard deviations calculated from the plate controls. If such a biological interaction exists, be it specific CCR5 binding or of unspecific nature, it should be reproducible using the same assay conditions. Since this is not the case here, it is assumed that the response is masked by inherent changes in the assay system. These could be changes in expression density of CCR5 on the surface as well as the amount of downstream actors like G proteins of Adenylate Cyclase, but also changes in concentrations of chemicals and endogenous ligand. To rule out simple human mistakes during the handling of chemicals and the assay system as a whole or similar errors, the antagonist dose response screening was repeated. Again, no response could be detected which supports the conclusion that the difference between the response of the primary screen and those of the dose-response screening must be due to different batches of

research materials used in the assay.

5.3 Summary & Outlook

The CC chemokine receptor CCR5 is and has been in the focus of pharmaceutical research due to its role in several inflammatory diseases but also, maybe most importantly, due to the role in the host cell invasion process of HIV. This research momentum has lead to the development of Maraviroc, an FDA-approved HIV entry inhibitor. The corresponding crystal structure was the first one of CCR5 to be published. This crystal structure finally allows virtual screening approaches to CCR5, which was one part of the presented project.

Besides, two models of the CCR5 were built to explore more receptor protein conformations and exploit these for virtual screening, in order to find new and different molecules than with the crystal structure alone. The first model was an ordinary homology model, which was built on existing GPCR structure templates. Since the crystal structure was already known, a comparison with the unbiased homology model was possible, highlighting the overall similarity of the model with the known structure whilst showing subtle, but important, differences in the binding site.

Finally, a third structure was developed by re-shaping the binding site. The rationale for the development of this third structure was the observation that the crystal structure could not explain the binding of TAK-779 and its derivatives synthesized by Junker and coworkers to CCR5. These molecules, as well as those molecules identified by the docking to this new structure, share a common, simple pharmacophore consisting of a potential positive charge, presumably interacting with Glu283^{7,39}; a linker moiety that can contain an amide function or a similar group; and finally an extended aromatic portion. The crystal structure could not explain the binding of this aromatic portion to the receptor, while the modeled structure shows a completely new binding pocket, deep within the protein. As for the two structures introduced before, a virtual screening campaign was

carried out against this modeled conformation.

The computational effort spent on the identification of new CCR5 modulators exceeded those of many other virtual screening campaigns against a single protein target by far. This originated from the use of multiple input structures, the use of a large ligand database and the relinquishment of any pre-screening methods such as pharmacophore models or property-based filtering. A post-docking filtering step was conducted to abolish potentially artificially enriched molecules showing over-emphasized charge interactions with Glu283^{7,39} from the docking to the CCR5 crystal structure. Apart from that, no additional computational post-processing was used. Instead, the top-scoring poses of each single docking run were manually inspected, judged and iteratively narrowed down, culminating in a set of 20 potential CCR5 modulators.

This set was tested using a commercially available functional assay service. In a primary screen using a fixed ligand concentration, three substances could be classified as potential antagonists and five more as potential agonists. These classification are based on the measured activity in the respective assay relative to the background variance of that assay. The primary hits show a low overall efficacy, not exceeding 30 %. However, in a virtual screening campaign using docking, compounds are selected based on sterical and chemical complementarity. Unlike an experimental high-throughput screening, there is no selection bias towards high-efficacy ligands. Still, low-efficacy ligands can exhibit high binding affinity towards their target. Worthwhile to mention, each of the three structures used add to these seven ligands.

The potential hits from the first round of screening were validated in a dose-response testing using the same assay setup. In this second round of screening, two of the agonist candidate compounds could be verified by showing a full dose response curve. The derived EC₅₀ values and the maximum responses are well within the scope to be expected in a virtual screening setup such as the one presented here, arising from the complex nature of chemokine receptors and their multifarious interactions with upstream and downstream regulators. The remaining three agonist candidates showed a

low background response without a dose-dependent behavior and thus have to be considered false positive hits from the first screening, at least in the assay system used to this point. Of the three antagonist candidates, none could be verified although these molecules were the most promising candidates in the first round of screening. Two possible scenarios exist, (i) those compounds were false positive hits in the first screening or (ii) those compounds were false negatives in the second screening. Due to the overall favorable results from the first screening, one may assume that the second scenario applies. Several explanations exist why a receptor response might remain undetected. One particularly conclusive one already mentioned is the receptor expression density on the cell surface. In contrast to other factors like concentrations of assay chemicals or comparable variables, the receptor density can not easily be controlled nor measured effectively. Also, it could change over time under different environments or through certain treatments of the cells (i.e. freeze-thaw cycles). Hence, these compounds should, and will, be screened again using different, complementary assay systems.

Given the two confirmed hits, DS67459 and DS81651, the identification of new chemical modulators of CCR5 can be considered successful with a hit rate of 10 %, which is a respectable result. Nonetheless, future efforts will be expended on screening the potential antagonists again to ensure no hit is missed. Furthermore, the hit rate might improve even more by using complementary assays. So far only the signaling via the G protein-coupled signaling pathway has been considered, but the ligands presented here might well induce biased signaling¹¹⁵ that is, they signal via the β -arrestin pathway more strongly or even exclusively. Finally, a binding assay might even reveal hits that do bind to CCR5 as predicted but do not induce signaling at all. Beyond the evident achievement of hit identification, several general conclusions about the docking to GPCRs and homology models of those could be strengthened.

First of all, docking to GPCRs works, as has been shown by numerous studies before, for example the docking to the A₁ Adenosine receptor²⁸, the β_2 -adrenergic receptor^{20;21} or the CXCR4 receptor 4.²⁶ Finally, the just re-

Table 5.3: *Template structures of CCR5 homology model*

PDB	Description	% identity	Resolution
4grv-A	Neurotensin receptor	30 %	2.8 Å
4ea3-B	N/OFQ opioid receptor	31 %	3.0 Å
3odu-A	Chemokine receptor CXCR4	36 %	2.5 Å

cently published OX₂ orexin receptor structure nicely showed that docking is able to reproduce the quite surprising binding mode of suvorexant.¹¹⁶ Furthermore, docking to homology models works well. In fact, the so far proven binders in this study have been selected from docking to the two models, not from docking to the crystal structure. Yet, this might of course change with future results. Finally, this also underlines the plausibility of the built models and thus allows the further use of those to address other issues and in turn deepens the understanding of CCR5, its structure, its dynamics, its ligands and the conjunction thereof.

5.4 Material & Methods

5.4.1 Structure Modeling

Three different crystal structures were used to model the CCR5, namely (i) PDB¹⁷ 3odu³², a crystal structure of the CXCR4 receptor, (ii) PDB 4grv¹¹⁷, the structure of the neurotensin receptor NTS1 and (iii) PDB 4ea3¹¹⁸, the crystallized nociceptin/orphanin FQ receptor (see Table 5.3). The T4-lysozyme insertion of 3odu and 4grv was removed manually. The respective multiple sequence alignment was derived from the structural alignment using the “salign” function as implemented in Modeller⁶⁶ 9.10. The alignment was augmented by including all human sequences of the CC, the CXC, the neurotensin- and the opioid receptor family, downloaded from GPCRDB.¹¹⁹ Using the “automodel” function of Modeller, 20 models were built and by manual inspection, one of these was selected for in-detail investigation and further optimization.

5.4.2 Model optimization

The initial model was inspected and for several residues around the binding site, new side-chain rotamers were selected from the Dunbrack side-chain rotamer library¹¹⁰, to improve sterical fit and hydrogen bonding patterns. Residues concerned were Leu36, Trp86, Leu104, Thr105, Leu107, Ile198, Asn252, Thr284 and Thr288. Subsequently, these residues as well as all residues within a 3.5 Å radius were energy minimized using PLOP.^{69;70} After the optimization of the binding site, ECL2 was minimized around the disulfide bridge between Cys101 and Cys178. The C-terminal end of the receptor (residues 318 et seqq.) was removed, due to the lack of an appropriate template. For residues 301–317, 25 new models were sampled using Modeller, where residues 301–305 (C-terminal end of TM-VII) and 308–317 (helix 8) were restricted to have an α -helical secondary structure. Also for ECL3, 25 new models were sampled, applying the α -helical constraint to residues 269–273. The resulting models were inspected after a short energy-minimization using PLOP and the best one was selected. Using PropKa^{93;94} 3.0 implemented in PDB2PQR⁹⁵ 1.8, protonation states were assigned at pH 7.0. Since no ligand was considered during this calculation, Glu283 was assigned a high pKa (7.6) and thus protonated. For further calculations, Glu283 was deprotonated manually, since it is known to form important charged interactions with CCR5 ligands. Finally, the whole structure was relaxed using PLOP’s energy-minimization procedure with an RMSG of 0.1.

5.4.3 Binding Site Reshaping

The CCR5 crystal structure 4mbs was used as starting point for the design of a receptor able to bind TAK-779. An initial orientation of the ligand was achieved by manually placing it in the binding site and minimizing the ligand together with its surrounding residues. To form a deeper binding pocket, a different side-chain orientation was selected for Phe112 from the Dunbrack side-chain rotamer library¹¹⁰ and the complex was manually refined in several iterative steps, which included swapping the donor

and acceptor functions of the hydroxy group of Tyr37 and Tyr251 and the corresponding amide function of TAK-779. To reduce the bias of manual curation, the complex structure was created by re-docking TAK-779 into this final receptor structure using AutoDock Vina⁷¹ v1.1.2. Input conformations were pre-generated using Corina^{107–109} 3.491, allowing the sampling of ring conformations (“rc” keyword). For each conformation, a maximum of ten poses was created by AutoDock Vina and the most promising one selected manually.

5.4.4 Docking

Virtual screening was carried out using DOCK 3.5.54^{1;72–74} with the ZINC⁴⁶ lead-like subset as input. The necessary input spheres for DOCK were generated by sphgen. For the crystal structure, the necessary seed coordinates were taken from the respective ligand, whereas for the modified crystal structure the docking pose of TAK-779 was used. For the homology model, no ligand information was available. To allow an even sampling of the binding site, a two-step procedure was applied. First, sphgen was used to calculate spheres for the whole receptor structure and the cluster representing the binding site was taken as input ligand for a second run of sphgen, while making sure that the ligand coordinates were not used to generate spheres (“useligsph=no”). From the respective docking runs, the top scoring molecules were manually inspected and evaluated and promising candidates were selected.

5.4.5 Compound Acquisition

Compounds were purchased from the following vendors: Enamine [DS84666, DS89674, DS55435, DS50263, DS83957, DS50192, DS23844, DS11585, DS29983, DS62535, DS81651, DS75993, DS00794, DS05596] and Chembridge [DS49232, DS10403, DS57404, DS55622, DS62705, DS67459]. For stock solutions, compounds were dissolved in DMSO and mass spectra were recorded to verify compound identity. Compound purity was stated to be higher than 90 % and was not assessed.

5.4.6 Second Messenger Assay

The cAMP second messenger assays “HitHunter cAMP XS+” were carried out by DiscoverRx Corporation, Fremont, CA, USA. Samples were shipped as 10 mmol L⁻¹ DMSO stocks. Briefly, cAMP Hunter cell lines were expanded from freezer stocks according to standard procedures, before incubating them at 37 °C in a total volume of 20 µL into white-walled, 384-well microplates. Cells were pre-incubated with sample in the presence of forskolin, to induce a response. For agonist screening, an EC₈₀, for antagonist screening an EC₂₀ of forskolin was used. Medium was aspirated from cells and replaced with 15 µL 2:1 HBSS (Hank’s Balanced Salt Solution)/10 nmol L⁻¹ HEPES:cAMP XS+ Ab reagent. Intermediate dilution of sample stocks was performed to generate 4X sample in assay buffer containing 4X EC₈₀ (for agonist screening) or EC₂₀ (for antagonist screening) forskolin. 5 µL of 4X sample was added to cells and incubated at 37 °C. The final (highest) compound concentration used in the assay was 10 µmol L⁻¹ in the single point measurements and 100 µmol L⁻¹ for the dose response measurements. Assay signal was generated through incubation with 20 µL cAMP XS+ ED/CL lysis cocktail for one hour followed by incubation with 20 µL cAMP XS+ EA reagent for three hours at room temperature. Microplates were read following signal generation with a PerkinElmer Envision™ instrument for chemiluminescent signal detection. Compound activity was analyzed using CBIS data analysis suite (ChemInnovation, CA, USA). The relative effect was calculated using Formula 5.1 for agonist screening and Formula 5.2 for antagonist screening, where $RLU_{vehicle\ control}$ and $RLU_{forskolin\ positive\ control}$ correspond to the positive assay controls and define 100% while $RLU_{max\ control}$ and $RLU_{EC_{80}\ control}$ represent the negative controls and set the 0% baseline.

$$\% \text{ Activity} = \frac{100\% \times (1 - (RLU_{sample} - RLU_{max\ control}))}{RLU_{vehicle\ control} - RLU_{max\ control}} \quad (5.1)$$

$$\% \text{ Inhibition} = \frac{100\% \times (RLU_{sample} - RLU_{EC_{80}\ control})}{RLU_{forskolin\ positive\ control} - RLU_{EC_{80}\ control}} \quad (5.2)$$

Table 5.4: *Raw control readouts in first round of screening*

Assay Type	Pos. Control		Neg. Control	
	μ	σ	μ	σ
Agonist	12690	1331	64947	1951
Antagonist	62640	1662	17385	1401

Table 5.5: *Formulas for propagation of uncertainty*

Function	Approx. Standard Deviation
$f = aA - bB$	$\sigma_f = \sqrt{a^2\sigma_A^2 + b^2\sigma_B^2 - 2ab\sigma_{AB}}$
$f = \frac{A}{B}$	$\sigma_f = \frac{A}{B} \sqrt{\frac{\sigma_A^2}{A^2} + \frac{\sigma_B^2}{B^2} - 2\frac{\sigma_{AB}}{AB}}$

5.4.7 Error Propagation

The mean readout for the positive and negative plate controls as well as their respective standard deviations as reported by the contract researcher are shown in Table 5.4. At least three of each controls were used per plate. Given the formulas for the propagation of uncertainty¹²⁰ as denoted in Table 5.5, the standard deviation σ of efficacies as calculated by Equations 5.1 or 5.2 can be calculated. For the σ value of the positive control, these formulas can be simplified to Formula 5.3, under the assumption that the errors of the measurements are uncorrelated, i.e. their covariance σ_{AB} is zero. The resulting $\sigma_{100\%}$ are 6.4 % for the agonist screening and 6.8 % for the antagonist screening. By considering limit values, also the propagated standard deviations of the negative controls $\sigma_{0\%}$ can be approximated, yielding values of $\sigma_{0\%, Agonist} \approx 5.3\%$ and $\sigma_{0\%, Antagonist} \approx 4.4\%$.

$$\sigma_{100\%} = \sqrt{\frac{2(\sigma_{pos.cont.}^2 + \sigma_{neg.cont.}^2)}{(\sigma_{pos.cont.} - \sigma_{neg.cont.})^2}} \quad (5.3)$$

6 A large-scale docking approach to kinases

6.1 Introduction

In prior examples of docking scenarios addressing multiple targets, the number of structures used was small, at least compared to the number of potential protein drug targets. Specifically, two to three different receptors were used. The question arises to which extent docking can be used for the prediction of “selectivity profiles”, that is, the prediction of binding or non-binding of a compound against a panel of protein targets. Addressing this question requires the docking of one compound to many structures, sometimes referred to as “inverse docking”.¹²¹ Obviously, a generalization using a single ligand is not possible, which leads to a setup where multiple ligands are docked to many targets. For a retrospective evaluation, experimental data are required for all protein-ligand combinations. These data form the “ground truth”, with which the docking results, i.e. scores or ranks, can be compared. Together, a dataset is required containing protein crystal structures and ligands for which experimental binding data are available. The development of such a dataset and the subsequent large-scale docking assessment is described here.

The protein targets suitable for such an application have to be available in high numbers in the PDB database¹⁷ to serve as target structures for docking. Furthermore, the experimental data required have to be complete in a way that all ligands have been tested against all targets, i.e. the corresponding protein-ligand matrix must be (at least almost) completely

populated. In addition, the data have to be comparable and should ideally originate from the same source. Together, these prerequisites render the usage of GPCRs highly difficult, mainly due to the lack of crystal structure data. This can be generalized to membrane proteins. In contrast, the catalytic domains of protein kinases are globular protein targets, although they can be indirectly bound to the membrane via additional domains.¹²² Catalytic protein kinase domains are highly represented in the PDB with currently more than 3200 structures (structures with PFAM¹²³ classification PF00069, as of Apr 2015).

Kinases, in general, form a very large and heterogeneous group of proteins, sharing the ability to transfer a phosphate group from a phosphate donor to a phosphate acceptor. The phosphate donor is usually formed by ATP (adenosine triphosphate), which donates a phosphate group that is then transferred to the acceptor. The latter is very often another protein; however, kinases exist that phosphorylate other molecular species such as lipids. An example for such a kinase is the sphingosine kinase that catalyzes the reaction of sphingosine to sphingosine-1-phosphate.¹²⁴ The phosphorylated lipid then acts as lipid mediator within the cell or on other cells, via the sphingosine-1-phosphate receptor, a GPCR.¹²⁵ Other chemical entities exist that are phosphorylated by kinases, but the vast majority of kinases target other proteins and are consequently referred to as protein kinases. The phosphate group is transferred to a specific residue on the target protein, mostly serine, threonine (serine/threonine kinases) or tyrosine (tyrosine kinases). The catalytic domains of protein kinases share a common fold (see Figure 6.1). This fold has two lobes, an N-terminal and a C-terminal one. The N-terminal lobe consists of β -sheets and an α -helix, the C-helix¹²⁷, whereas the C-terminal lobe is mainly α -helical. The lobes are connected by a so-called “hinge region”, which forms hydrogen bonds with the adenine ring of ATP that binds in the flat pocket between the two lobes.

Using phosphorylation, kinases are able to modify the activity of downstream regulators and thus play a very central role in the regulation of cellular processes¹²⁸, highlighted by the existence of more than 500 ki-

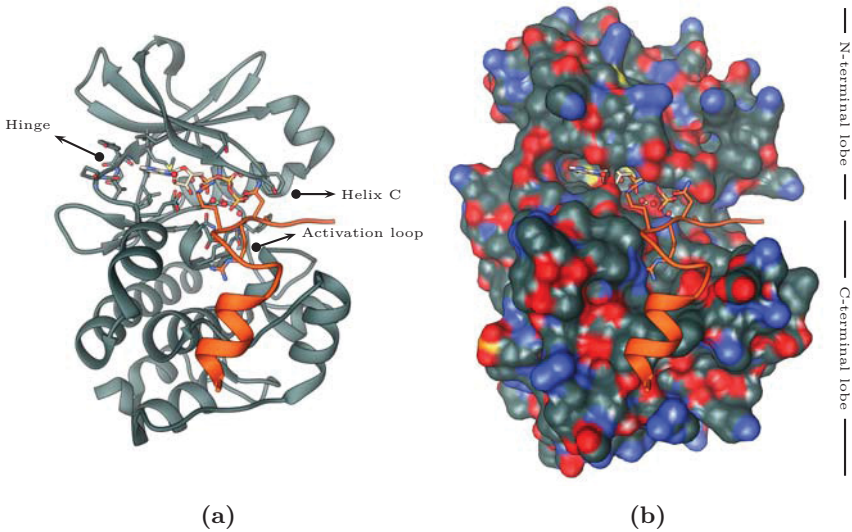


Figure 6.1: Overview of the structure of the catalytic domain of a protein kinase (dark gray) with ATP bound (light brown, sticks). A peptidic, competitive inhibitor is bound to the site of phosphorylation (orange, ribbon). Structure from PDB 4wb5.¹²⁶ (a) Kinase domain in ribbon representation. (b) Kinase domain in surface representation. The surface highlights the flat ATP binding pocket between the two lobes.

nases in the human kinome.¹²⁹ A dysfunction or -regulation is hence often associated with a physiological disorder. Especially for countering various forms of cancer, many kinases are considered valuable drug targets.^{4;126;128} This medical relevance has obviously stimulated the research effort spent on kinases and has lead not only to the aforementioned number of crystal structures but also to a number of new drugs like Flavopiridol, Roscovitine and Gleevec.¹³⁰ These pharmaceuticals bind, at least partly, to the main binding pocket of kinases and thus compete with ATP. This is the common mechanism of small molecules that inhibit kinase activity. On the one hand, this allows, to a limited extend, the transfer of knowledge from one target to another, since core scaffolds can bind to different kinases (“hinge binding motifs”).¹³¹ This binding site similarity can be desired for polypharmacological kinase inhibitors.¹³² On the other hand, the design of selective kinase inhibitors is an ongoing issue and has been long focused on the exploitation of differences in residues flanking the ATP binding site. However, new crystal structures allowed insights into the dynamics of kinases, revealing two different mechanisms that would open new binding pockets^{130;133} (see Figure 6.2). The first mechanism is the transition from the active conformation to the inactive conformation of the kinase via the movement of the activation segment, or loop. The two related states are commonly referred to as the “DFG-in” and the “DFG-out” motif.¹³³ The DFG motif is a segment of three highly conserved amino acids, aspartate, phenylalanine and glycine at the N-terminal end of the activation segment. This loop can switch into a conformation where the binding of the substrate is blocked. In this conformation, the phenylalanine residue of the DFG motif is pointing outwards into the direction of the solvent (see Figure 6.2b). This allows a ligand to bind to the pocket that would be occupied by the phenylalanine in the active conformation. Another dynamic process of kinases is related with an outward shift of helix C, which requires the breaking of a salt bridge, acting as an “ionic lock” (see Figure 6.2c). Together, these moves give room for the binding of larger ligands that extend from the ATP binding pocket deeper into the protein. This spatial area is termed the “back pocket” or “deep pocket” and is considered as

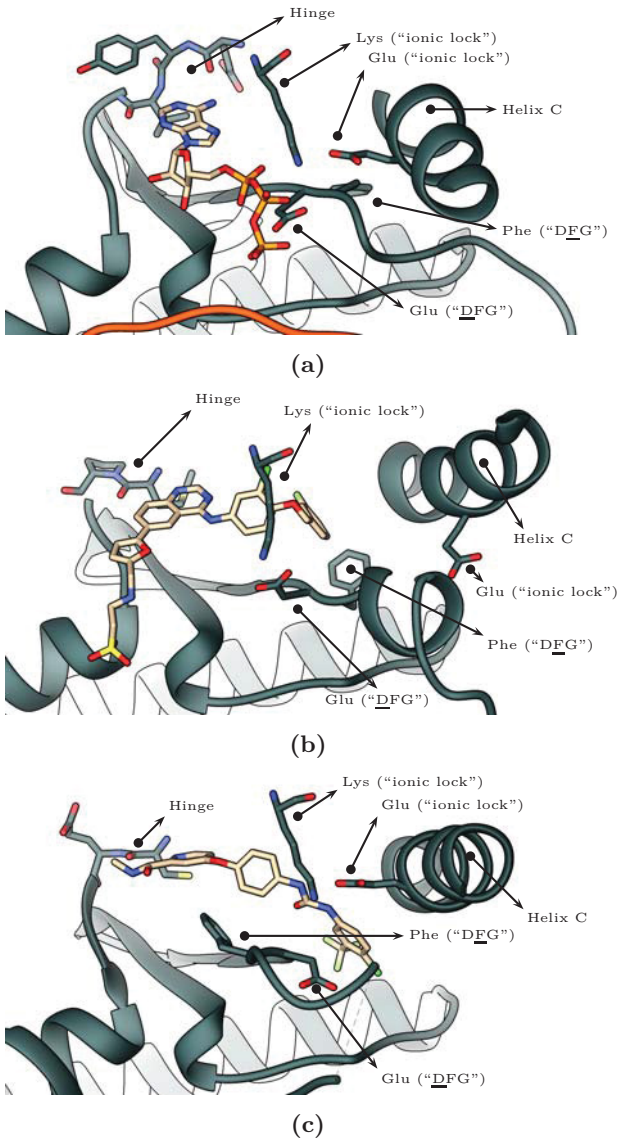


Figure 6.2: View of the binding site of different kinase (gray, ribbon) and ligand (light brown, stick) combinations. β -sheets removed for clarity. (a) ATP bound to PKA (PDB 4wb5¹²⁶). DFG motif and α -helix C are in "In" state. (b) Lapatinib bound to EGFR (PDB 1xkk¹³⁴). Helix C in "Out" conformation, the locking salt bridge is disrupted. (c) Sorafenib bound to B-Raf (PDB 1uw¹³⁵). DFG motif in "Out" state.

an important aspect in the design of selective kinase inhibitors (“type II inhibitors”).¹³⁶

To investigate the selectivity of a molecule to multiple kinases, assay systems are required that allow the screening of compounds against a panel of kinases at a reasonable effort. Consequently, such systems have been developed and have lead to the availability of large experimental datasets. Karaman and co-workers were amongst the first to systematically screen a compound set against a broad panel of kinases and publish the corresponding data.¹³⁶ The authors tested 38 kinase inhibitors that were developed against different targets. For every compound, the affinity was evaluated against a panel of 317 kinases in total, consisting of 287 distinct protein kinases (~55% of the human kinome), a number of disease-related mutants thereof and some lipid kinases. This satisfies the requirements for an experimental dataset as defined before.

Together, kinases are ideally suited for an assessment of docking performance across many targets. Consequently, a dataset suitable for the intended large-scale docking approach has been built by mapping the kinases from the dataset of Karaman *et al.* to the PDB¹⁷. The ligands from the experimental dataset were then docked to each retrieved structure. Again using the experimental dataset, ligands were classified as known active or inactive binder for every structure and used to quantify the retrieval rate achieved by docking by means of ROC⁵¹ and enrichment. In addition, different procedures were selected from literature that aim at improving docking accuracy by normalizing docking scores between multiple pockets. The performance of such normalized scores has then been compared to that of raw docking scores.

6.2 Results & Discussion

After initial retrieval and the subsequent selection process, a total of 650 crystal structures was considered as set of target structures for docking. These 650 structures represented 129 different kinases out of the 317 ki-

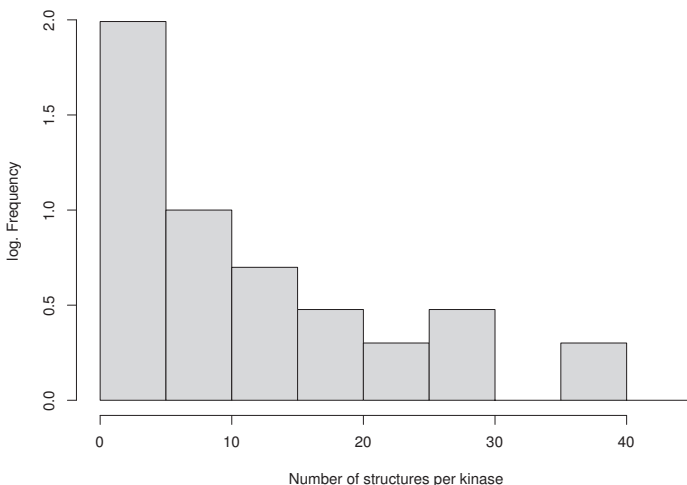


Figure 6.3: Distribution of the number of crystal structures per kinase for the 650 kinase structures in the final dataset. *y*-axis logarithmic.

nases in the experimental dataset from Karaman and co-workers. For the remaining kinases, no structure was available in the first place or they failed to fulfil the quality criteria described in the “Methods” section. The structure/kinase ratio implies that, on average, every kinase is represented by about five different structures. However, as Figure 6.3 shows, this value differs strongly between different kinases. Around 100 kinases are each represented by five structures or less. More detailed, 56 kinases are represented by a single structure only. In other words, more than 40 % of the kinases constitute less than 9 % of the structures. On the other hand, some kinases are highly overrepresented. The five (4 %) most abundant kinases (CDK2, MAPK14, CHEK1, PIM1 & PDPK1) contribute a total of 178 structures (27 %). In theory, a high number of structures for a given kinase is favorable, since the conformational space of the kinase and its binding site is covered to a larger extent. Consequently, it is more likely that the conformation necessary to bind a certain ligand is available. Accordingly, kinases for which only few structures are available, are in theory more likely to generate false-negatives, i.e. *in vivo* binders with an unfav-

vorable docking score, caused by the non-compatibility of the ligand with the few available binding site conformations. However, dismissing kinases with few structures to balance the distribution of structures per kinase would dramatically decrease the size of the dataset and was therefor not considered.

The initially selected and prepared structures were then used in the large-scale docking approach, where all ligands from the experimental dataset were docked to the 650 structures. The complete matrix with the docking scores of all structure-ligand combinations is depicted in Figure 6.4. Docking scores are represented by a color scale ranging from orange (negative, favorable) to purple (positive, unfavorable). Gray indicates structure-ligand combinations without assigned docking score. Except for few cases, no scoreable ligand poses were found for these combinations. The remaining few cases are outliers that have been removed for proper visualization (see “Methods” section for details). Three columns, i.e. ligands, stand out in that depiction with a high number of white and purple fields compared to the remaining columns. These ligands with mainly unfavorable scores are staurosporine and two chemical derivatives of it. Staurosporine is known to bind to a large number of kinases¹³⁶, hence one might expect a high number of favorable docking scores when docking to numerous kinases. An explanation that this is not observed here might be the large and bulky structure of staurosporine which requires an open binding site conformation. Obviously, the binding sites of most kinase structures used here are adapted to a smaller ligand and are thus in a more closed conformation. Therefor, staurosporine is a convincing example pointing to the necessity of a suitable receptor conformation for successful docking. Besides this observation, the general glance of the matrix indicates certain trends in the scores. First, some ligands (in columns) tend to get constantly higher/lower scores than other ligands. It is generally accepted that the absolute value of docking scores depends on the nature of the docked ligands¹³⁷, which is nicely underscored here. But not only the ligands, also the structures, respectively kinases, seem to have an overall different level of docking scores, visible as blocks of rows with a similar color pattern in Figure 6.4. As an example,

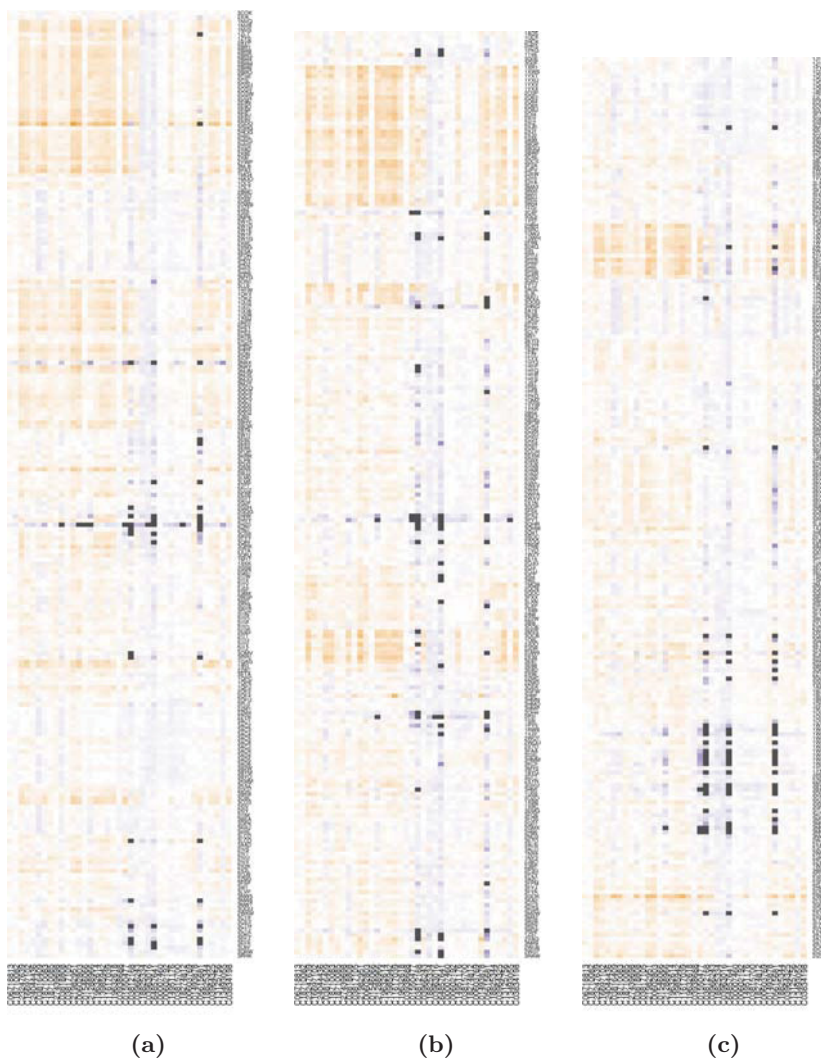


Figure 6.4: Score heatmap of all structure-ligand combinations. Structures are grouped by kinase. Raw docking scores are depicted by color scale ranging from orange (negative, favorable) over white (average) to purple (positive, unfavorable). Gray fields indicate protein-ligand combinations without score (see “Methods” section for details). For visualization, the matrix has been split. (a) Upper part. (b) Middle part. (c) Lower part.

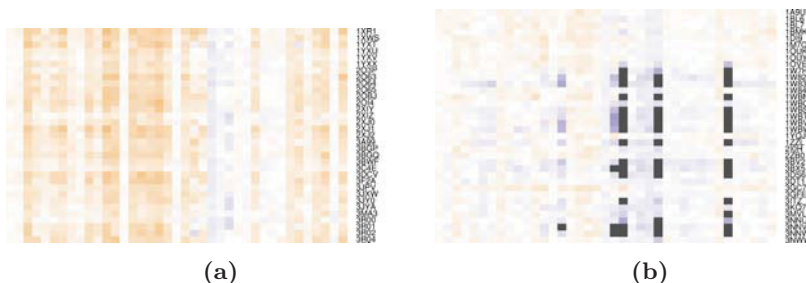


Figure 6.5: Heatmap of docking scores for selected kinases. Excerpt from Figure 6.4. (a) Heatmap for all PIM1 structures. (b) Heatmap for all p38 α structures.

Figure 6.5 shows two of such blocks, representing the structures of PIM1 and p38 α , respectively. Normalization procedures have been discussed in literature to equalize the overall score levels between different structures or ligands and thus improve the prediction accuracy.

Three different approaches were selected from literature and the prediction accuracy was compared between these normalized and the raw docking scores. The normalization procedures are hereafter referred to as “MASC”, “mMASC” and “iterative normalization”. The MASC (“multiple active site correction”) approach was introduced by Vigers and co-workers.¹³⁸ When docking to multiple targets, they proposed to normalize the docking scores of a ligand against all targets. More specifically, for each score, the mean determined across all structures is subtracted, followed by a division by the standard deviation. This corresponds to a Z-normalization of all scores of a ligand (Z-normalization of the columns when referring to the matrix in Figure 6.4). The idea behind this strategy is that the non-specific part of the docking score, e.g. due to the molecular size of the molecule, is reduced. Jacobsson *et al.* developed this method further and proposed the mMASC¹ approach. They subtracted a ligand’s mean from each of its scores but did not divide it by the standard deviation. Such a division would penalize ligands with a high spread in score amongst different tar-

¹In the original publication, the authors used the term “MASC”, just as Vigers and Rizzo. “mMASC” has been introduced here for clarity.

gets and the necessity of such a penalty is not obvious. Before calculating the mean score of a ligand, they applied a Z-normalization along all scores within a structure to make the scores comparable between different structures. This seems reasonable, since the target can bias the docking score the same way as a ligand does. And indeed, as mentioned before, this behavior has been observed here (see Figure 6.5). In short, the mMASC approach is the Z-normalization over each structure (normalizing each row when referring to Figure 6.4), followed by shifting each score by the mean of the respective ligand (column mean when referring to Figure 6.4). Figure 6.6 shows the heatmap of docking scores after normalization using the mMASC approach. Compared to the unnormalized scores in Figure 6.4, the aforementioned blocks of patterns seem to be less pronounced, which was the goal of normalization. Finally, Casey and co-workers¹³⁹ proposed an iterative approach where the scores are Z-normalized first for each structure and then for each ligand. These steps are repeated until the scores converge.

Using the raw docking scores and the three sets of transformed scores, the prediction power of docking was assessed retrospectively using two different measures, the ROC, more accurately the AUC of the ROC, and the enrichment factor. While the ROC is a global assessment of a ranking, the enrichment is calculated at a given point, or percentage, of the ranked list. Importantly, within the Karaman dataset the number of active ligands out of the 38 used differs strongly between the individual kinases as shown in Figure 6.7. If the enrichment would be calculated for a fixed number of top scoring poses, the enrichment factor between kinases would be scaled to a different range and could thus not be compared. For a fair comparison, an “adaptive” enrichment was calculated, which takes the number of actual actives into account (see “Methods”).

For one, these assessments were calculated for each structure and furthermore for each kinase, i.e. groups of structures. The docking scores of all structures of a certain kinase were consolidated, meaning that the docking score of a ligand to a kinase equals the best (smallest) score of that ligand to all structures of that kinase (see Equation 6.1).

The ROC and enrichment values are presented in Figure 6.8 on the struc-

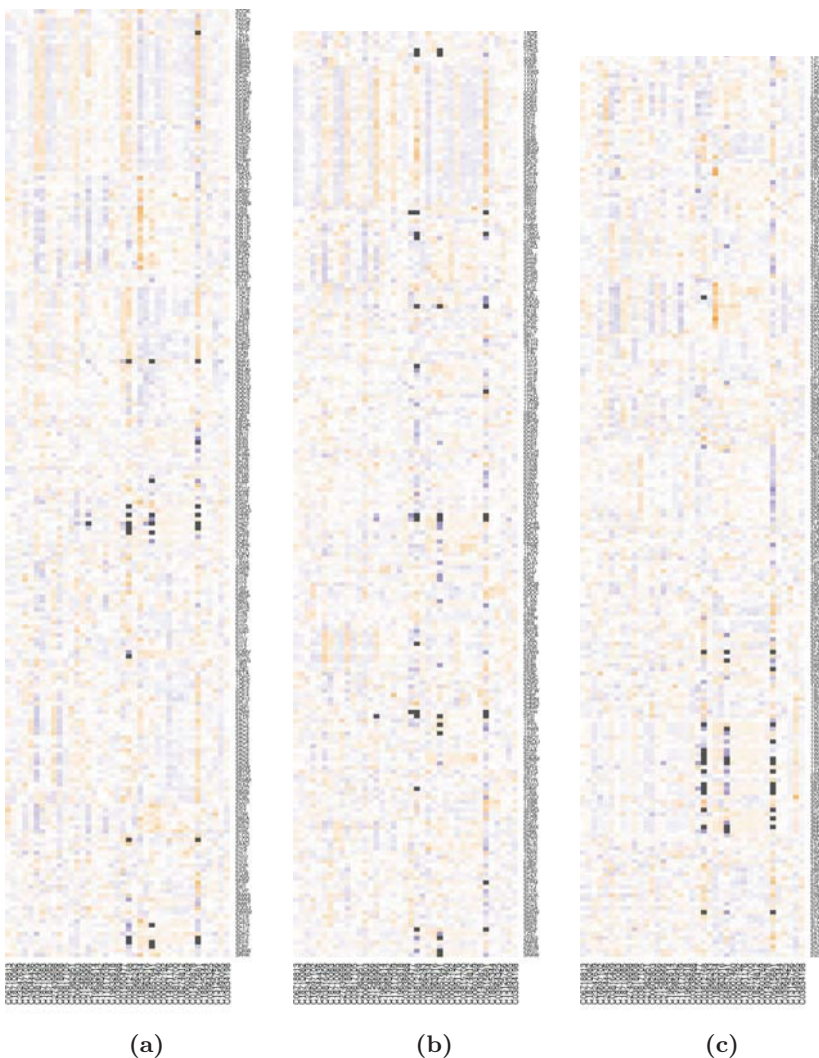


Figure 6.6: Score heatmap of all structure-ligand combinations. Structures are grouped by kinases. mMASC-normalized docking scores are depicted by color scale ranging from orange (negative, favorable) over white (average) to purple (positive, unfavorable). Gray fields indicate protein-ligand combinations without score (see “Methods” section for details). For visualization, the matrix has been split. (a) Upper part. (b) Middle part. (c) Lower part.

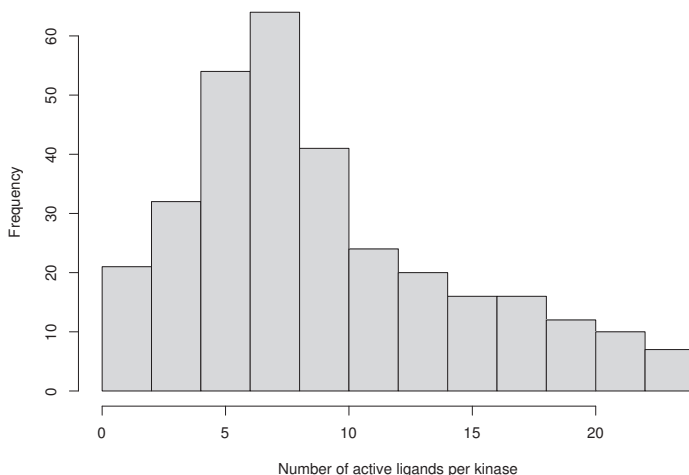
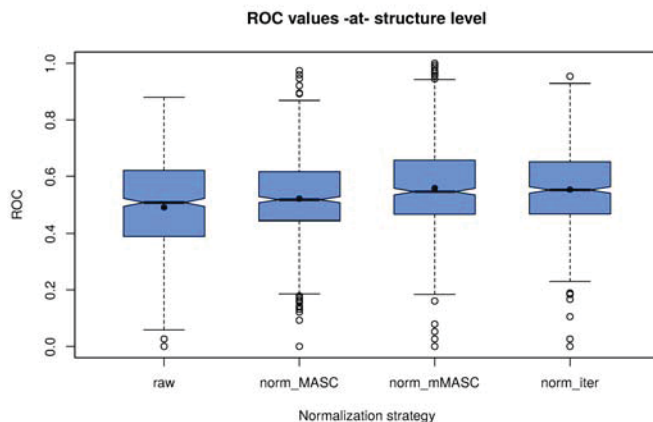
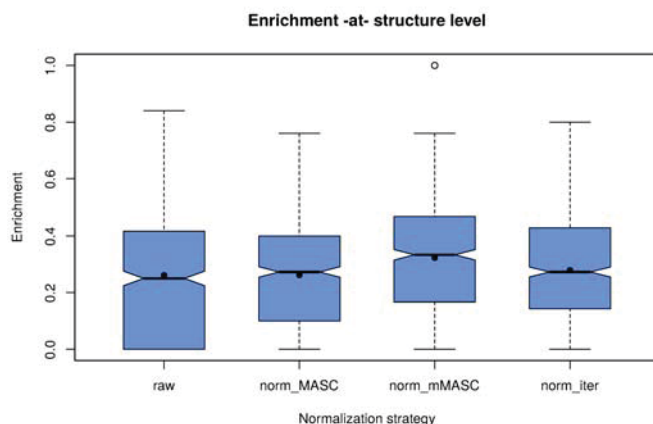


Figure 6.7: *Distribution of the number of active ligands per kinase for each kinase in the Karaman dataset¹³⁶*

ture level and Figure 6.9 on the kinase level. For the raw docking scores, the overall performance of the large-scale docking approach presented here is very mediocre. The ROC AUC has a mean around 0.5, on the structure level, corresponding to a randomly sorted ranking (see Figure 6.8a). The distribution is only slightly skewed, which means it ranges to both sides of the average. This implies that in some structures the active ligands are on average ranked worse than the non-active ones. However, in the opposite direction, the result is almost equally frequent. The enrichment is more promising than the ROC AUC, indicating that on average, 25 % of all known active ligands can be found among the top-scoring molecules (see Figure 6.8b). In comparison, the average ROC AUC indicates that in general, active ligands are not ranked better than known inactive ligands. The enrichment, however, suggests the opposite. This discrepancy might indicate a scenario where some actives are well enriched, leading to a reasonable enrichment value, but others are ranked low. These low ranked actives compensate the highly ranked ones in the ROC AUC calculation. Such a scenario might relate to the problem of “availability of a suitable



(a)



(b)

Figure 6.8: Docking performance in the large-scale docking approach for the raw and normalized docking scores. Evaluation on the level of structures. Box spans range from first to third quartile. Whiskers span 1.5 times inter-quartile range above and below median. Black line indicates median. Black dot indicates mean. Notches indicate 95% confidence interval. (a) Performance expressed as ROC AUC. (b) Performance expressed as enrichment.

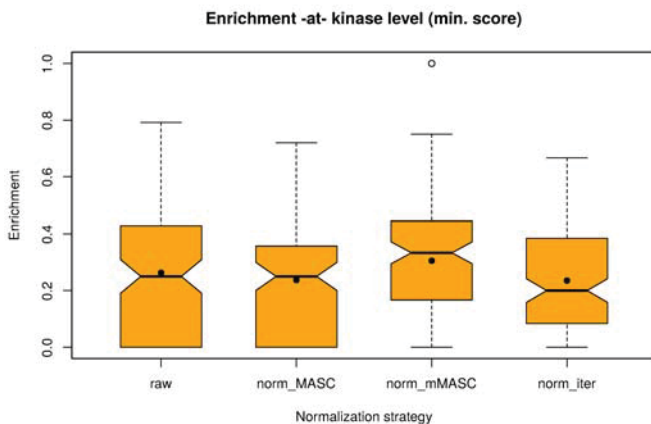
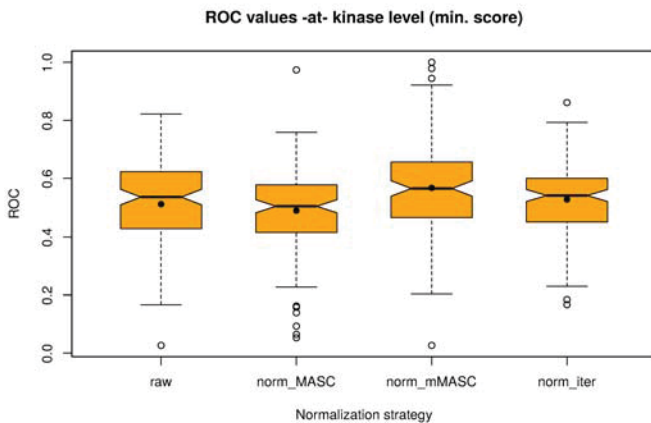


Figure 6.9: Docking Performance in the large-scale docking approach for the raw and normalized docking scores. Evaluation on the level of kinases. Box spans range from first to third quartile. Whiskers span 1.5 times inter-quartile range above and below median. Black line indicates median. Black dot indicates mean. Notches indicate 95 % confidence interval. **(a)** Performance expressed as ROC AUC. **(b)** Performance expressed as enrichment.

binding site”, i.e. every structure can bind only those ligands that fit its conformation.

When consolidating the docking scores for the structures to the respective kinase, the median ROC AUC as well as the enrichment remain almost unaltered, however, the interquartile range becomes smaller. Hence, the prediction becomes more stable and has smaller extreme values, yet in both directions. In short, the average prediction accuracy does not improve when using multiple structures for docking. This finding is counterintuitive, but can eventually be explained by the fact that the majority of all kinases are represented by only a few structures (see Figure 6.3). Thus, the distribution of ROC AUCs and enrichment values for kinases is strongly dominated by those kinases that have a small chance to improve. In addition, the similarities between multiple structures belonging to the same kinase have not been assessed. Consequently, structural diversity cannot be taken for granted even if numerous structures exist for a kinase.

Up to a certain extent, the claims made about the ROC AUC values and enrichments on the structure and the kinase level for raw docking scores can be transferred to each set of transformed docking scores. However, a comparison between the different sets of docking scores indicates that each of the three normalization procedures can improve the docking prediction accuracy, at least slightly. Exceptions are the ROC value of the MASC approach and the enrichment of the iterative approach on the kinase level (see Figure 6.9), which are both slightly worse than the respective performance using the raw docking scores. The mMASC approach is the procedure that performs best among all methods. When comparing ROC AUC values, the iterative approach does not perform significantly worse, since the 95 % confidence intervals (indicated by the notches) are overlapping. Still, the average values speak in favor of the mMASC method. This observation clearly holds true when moving from the ROC evaluation to the enrichment. Here, mMASC on average outperforms all other methods and increases the amount of active ligands in the top scoring molecules to 30 %.

6.3 Summary & Perspective

Kinases distinguish themselves from many other protein families by the availability of large, consecutive, experimental inhibition datasets as well as a high number of crystal structures. These features render kinases an ideal protein target family for a large-scale docking setup, together with a retrospective analysis of the docking performance. A dataset of crystal structures has been built here, containing 650 structures from 129 different kinases, which has been used for such a docking assessment. Of note, it can be used as input dataset for a wide range of other computational methods. For the 129 kinases, experimental binding data are available for 38 ligands, allowing the calculation of ROC AUC and enrichment values for the 650 docking runs. The assessment of docking performance has been accomplished using the raw docking scores as well as the transformed scores from three different normalization methods. All three post-processing functions can improve the overall performance with the mMASC methods being the most successful one. Still, the predictive power is limited. For mMASC, around one third of all actives can be retrieved as the top scoring molecules, indicated by the enrichment factor.

Despite the overall performance, the improvement upon normalization clearly shows that there is information available not only in the single docking score itself, but also in the distribution of docking scores over different structures and ligands. These informations can be used to augment docking predictions, as done here using normalization. Another idea of normalizing ligand docking scores is by using ligand decoys. Wallach and co-workers¹⁴⁰, for example, used virtually generated decoy ligands to normalize the scores of the ligands of interest. Using decoys for normalization has the advantage of being independent of the composition of the ligand library used for docking. Herein, a small and strongly biased ligand library was used. Diverse decoy sets might be beneficial for the normalization in future docking experiments. Decoys can be virtually generated¹⁴⁰ or selected from a library of existing molecules²⁶; however, a virtual library can cover a larger chemical space. Like the ligand set, the set of protein

structures is strongly biased since it is comprised of kinase domains only. It is very large, indeed, yet it might profit from a more diverse selection of protein structures. One might image a protein decoy set, where for a given binding pocket, new binding pockets are selected from the PDB¹⁷ which are similar in terms of physico-chemical properties, i.e. size, electrostatic properties, but are structurally unrelated. Cavbase¹⁴¹ and the scoring functions implemented therein might serve as similarity measure for such an approach.

In summary, a structure dataset has been compiled which can be used for future assessment of docking and other retrieval methods. It is inhomogeneous in a way that the number of available structures per kinase differs greatly among different kinases. Together with the small number of ligands and their chemical similarity as kinase inhibitors, this makes the dataset challenging for docking retrievals.

Score normalization has shown to be able to improve the retrieval rate but might still benefit from ligand decoy sets for normalization. This approach could also be applied to protein decoys, an idea that has not yet been described before and might be an interesting future development. So far, the prediction of a “selectivity profile” for that very large number of targets is not reliably possible.

In contrast to the preparation of only a few structures for docking, the distinct characteristics of a structure cannot be taken into account in such a large dataset. This includes, for example, the presence and position of water molecules, ions or additives as well as structural resolution and quality. Such aspects can clearly influence the docking results. The resulting mediocre docking performance achieved here is in contrast to the numerous success stories of docking and highlights the importance of a proper structure selection, preparation and docking setup by an experienced investigator.

6.4 Materials & Methods

6.4.1 Structure Selection and Preparation

The accession codes given by Karaman et al.¹³⁶ were mapped to PDB codes using the Uniprot database.⁹¹ For every PDB code obtained, the respective PDB structure was downloaded (2012, Feb) and verified to possess a kinase domain since the sequence accession code might cover additional domains. Furthermore, every structure was ensured to contain a ligand in the ATP binding site. Other ligands and crystallization additives were removed from the structure. The amino acid sequences were extracted from the structures and aligned using the MUSCLE multiple sequence alignment tool¹⁴², separately for every accession code, i.e. same sequences were aligned. The alignments were inspected and checked for mutations and deletions. These were then evaluated and classified as either significantly influencing ligand binding or not. If so, the respective structure was dismissed. This classification was based on manual inspection and evaluation of the structure not on simple distance-based criteria. In addition, these alignments were used to trim the structures, i.e. elongated and mostly disordered N- and C-termini were truncated. The first and last residue of the kinase domain, as annotated in Uniprot, were used as reference. Truncated versions of the structure contained the complete kinase domain, where possible. Structures where the kinase domain was not completely resolved, were again manually evaluated and retained were tenable. Structures were protonated using the “AddHyd” command implemented in YASARA⁶⁸ version 11.9.18 and the resulting hydrogen bonding network was minimized using CHARMM⁷⁵ version 31b2. Kinase inhibitors were taken from the ZINC database^{46;85} using full-text and structure-based searches.

6.4.2 Docking

DOCK 3.5.54^{1;72–74} was used to dock the downloaded inhibitors to the prepared structures. The necessary input spheres for DOCK were generated from sphgen¹, using the crystal ligand as reference for defining the binding

site.

6.4.3 Evaluation

The normalization procedures used^{138;139;143} were implemented as functions in R. The calculation of ROC AUC values and enrichment values was implemented likewise. To assess the ROC AUC and the enrichment for a kinase, i.e. a set of structures in contrast to a single structure, a new vector of ligand docking scores was calculated using the best (smallest) score for each ligand i over every structure j representing the given kinase k (see Equation 6.1).

$$S^{ki} = \min_j (S^{ji}) \quad (6.1)$$

The enrichment values were calculated using variable thresholds. The number of top scoring compounds considered was not fixed at a certain percentage of the database size. Instead, the number of ligands considered was set to the number of true actives in the experimental dataset. Consequently, the enrichment is scaled to 0%–100% and can thus be compared between different structures and kinases, although the absolute number of true active differs. Ligands were classified as actives against a certain kinase/structure, if the K_d value was reported to be below $10\text{ }\mu\text{mol L}^{-1}$ by Karaman *et al.*. Otherwise, they were classified as inactives.

6.4.4 Visualization

For proper visualization, selected scores were removed in Figure 6.4 & Figure 6.6. Some few protein-ligand combinations lead to extremely high docking scores, i.e. a docking pose was found but the associated score was extremely unfavorable. The heatmap visualization used applies a linear color gradient along the distribution of scores. The aforementioned outliers lead to a very broad score range. By removing positive outliers, the score distribution is balanced around the average and so is the distribution of colors in the heatmap. More specifically, every score S with a positive value exceeding the absolute value of the smallest, negative value of all

scores was removed. This removal of outliers was only applied during visualization. Consequently, the protein-ligand combinations without assigned scored (gray fields) slightly differ between the heatmap representations.

7 Summary

7.1 Summary (english)

The first chapters of this work have focused on the identification of ligands for GPCR targets with an emphasis on the use of multiple structures. This could be homology models or crystal structures. These structures represented either (i) different states of the same receptor as described for CCR5, (ii) different isoforms of a receptor such as CXCR4a and CXCR4b or (iii) actually different receptors. The last instance is assuredly the most interesting one since the binding of a molecule to multiple, well-defined, protein targets forms the basis of polypharmacology.²⁻⁴

The identification of ligands binding to CXCR3 and CXCR4 and the prediction of the selectivity of these ligands represents a successful example with respect to modeling and docking. In this specific example, eleven substances were correctly predicted to bind to either one or both of the receptors. The most challenging category has assuredly been the identification of ligands binding to both receptors, due to the different charge preferences of these two receptors. Yet, two molecules could be selected from both docking applications and experimentally verified to match the expected dual-binding profile. Furthermore, the compounds showed very good binding affinities despite a limited efficacy. These results substantiate the general applicability of docking to identify substances that bind to multiple targets.

An effort into the opposite direction is the identification of ligands selective for one receptor over, one or more, others. In such a scenario the receptors are very similar and tend to bind the same ligands. This can

be different isoforms of the same receptor, such as CXCR4 in zebrafish or different subtypes like those of the muscarinic¹⁴⁴ or serotonin¹⁴⁵ receptors. These two studies highlight the topicality of the general question of selectivity prediction of a ligands against several targets. Especially the work from Kruse *et al.*¹⁴⁴ additionally shows the high difficulty of a successful prediction in such a setup, as it was addressed here by the docking to isoforms of CXCR4 subtypes of zebrafish. Compounds that selectively regulate a desired receptor isoform can be utilized to investigate this receptor's role in zebrafish development. This underscores the applicability of the prediction of tailored selectivity.

The zebrafish *in vivo* system could not verify the efficacy of the predicted ligands and accordingly, the correctness of the selectivity predictions made could not be assessed. Chemokine GPCRs form a challenging group of targets due to the large and solvated binding site. Yet, new GPCR-binding ligands have been predicted and verified in this work. However, ligands predicted by docking are selected on the base of spatial and chemical complementarity to the binding site, not on the response they induce in the receptor. As a consequence, the effect of these ligands can be very weak, as it has been observed in the CXCR3-CXCR4 and the CCR5 example. This also applies to ligands with high binding affinity. Thus, the zebrafish *in vivo* system might be inadequate to detect the modulation of the CXCR4 isoforms.

Also the CCR5 example shows the problems that can be associated with GPCR assays, since the reproduction of initial screenings turned out to be problematic. Assay protocols might not be optimized for compounds with weak efficacy or may not be sufficiently sensitive in a reliable fashion. However, the initial screening successfully identified compounds that were selected from different input structures. The fact that different structures bind different ligands is an interesting consideration in the realm of docking to multiple target proteins. One can imagine different structures of a particular protein that bind different sets of ligands. These sets might be partly overlapping or not. Thus, to predict whether or not a compound binds to different proteins, one has to carefully think about choosing the

right structures for each of the proteins.

For the protein kinase family, many more crystal structures are available than for GPCRs. This allowed to build a dataset of crystal structures and kinase inhibitors for which experimental binding data are available. An unsupervised docking against this dataset suggested that the prediction of binding profiles on a large-scale basis is not yet feasible with the given methods. Instead, careful structure selection and preparation seems to be key for an successful docking to multiple targets. Several ideas originated from the discussed experiment to improve docking accuracy, e.g. by using protein decoys next to ligand decoys. New methods likes these will help to improve the outcome of such complex docking challenges and allow the wider application of docking for the prediction of ligands with tailored selectivity.

In summary, the CXCR3-CXCR4 case study proves that docking can predict dual-binding ligands. However, GPCRs have certain peculiarities. Depending on where they bind and which conformational change they induce, ligands can act as allosteric modulators, antagonists or agonists or bind without inducing a response. If there is a response, this can be forwarded via different pathways and lead to different signals. Altogether, GPCR assays might be too complex to expand the experiments to a higher number and more divergent structures. Kinases might be better suited for future studies with more targets, due to the availability of crystal structures and experimental setups. Nevertheless, the large-scale docking setup to kinases highlights the necessity of a proper docking setup by an experienced experimenter and methods like normalization using ligand and protein decoys.

The presented work shows the feasibility of docking approaches to identify ligands with tailored selectivity. Future experiments are going to extend the number of target structures, presumably moving to another protein family like kinases. With increasing experience and maybe improved algorithms, upcoming docking setups will hopefully allow the prediction of selectivity profiles over a larger panel of targets, even across multiple protein families.

7.2 Zusammenfassung (deutsch)

Die ersten Kapitel dieser Arbeit haben sich auf die Identifizierung von Liganden für GPCRs konzentriert. Der Fokus lag dabei auf der Nutzung von mehreren Strukturen, seien es Homologiemodelle oder Kristallstrukturen. Diese Strukturen konnten entweder (i) verschiedene Zustände des gleichen Rezeptors darstellen, wie für CCR5 beschrieben oder (ii) verschiedene Isoformen des gleichen Rezeptors wie CXCR4a und CXCR4b oder aber (iii) wirklich unterschiedliche Rezeptoren. Der letzte Fall ist mit Sicherheit der interessanteste, da die Bindung eines Moleküls an mehrere, definierte Zielproteine die Grundlage von Polypharmakologie ist.²⁻⁴ Sowohl die Entdeckung von Liganden die an CXCR3 und CXCR4 binden als auch die Vorhersage der Selektivität dieser Liganden ist ein gelungenes Beispiel dafür.

In diesem speziellen Beispiel wurde für elf Substanzen korrekt vorhergesagt, ob sie entweder an einen oder an beide Rezeptoren binden. Die größte Herausforderung war mit Sicherheit die letzte Kategorie, da gezeigt wurde, dass diese Rezeptoren die Bindung von Liganden mit unterschiedlichen Ladungen präferieren. Nichtsdestotrotz, konnten zwei Moleküle von den beiden Docking Anwendungen ausgewählt werden, für die die erwartete duale Bindung nachgewiesen werden konnte. Zusätzlich zeigten diese Substanzen sehr gute Bindeaffinitäten trotz der begrenzten Bindungseffekte. Diese Ergebnisse beweisen, dass Docking generell angewendet werden kann, um Liganden zu identifizieren, die an mehrere Zielproteine binden.

Ein Vorstoß in die entgegengesetzte Richtung ist die Identifizierung von Liganden die selektiv sind für einen Rezeptor gegenüber einem oder mehreren anderen. In so einem Szenario sind sich die Rezeptoren sehr ähnlich und binden tendenziell die gleichen Liganden. Das können verschiedene Isoformen sein, wie die von CXCR4 im Zebrafisch oder verschiedene Subtypen wie die des muskarinischen¹⁴⁴ oder des Serotonin-Rezeptors.¹⁴⁵ Diese beiden Studien zeigen die Aktualität der Fragestellung nach der Vorhersage der Selektivität eines Liganden gegen mehrere Zielproteine. Vor allem die Arbeit von Kruse *et al.*¹⁴⁴ zeigt zusätzlich auf, wie schwierig die erfolg-

reiche Vorhersage in so einem Versuchsaufbau ist, wie er hier angegangen wurde durch das Docking gegen CXCR4 Isoformen aus dem Zebrafisch. Substanzen die gezielt die Isoform eines Rezeptors regulieren können, können genutzt werden um die Rolle dieses Rezeptors in der Entwicklung von Zebrafischen zu verstehen. Diese Anwendung unterstreicht die Relevanz der Vorhersage von maßgeschneiderten Selektivitäten.

Das Zebrafisch *in vivo* Systems war nicht in der Lage die Wirkung der vorhergesagten Liganden zu zeigen, dementsprechend konnte die Korrektheit der Vorhersage von Selektivitäten nicht ausgewertet werden. Chemokine GPCRs stellen eine herausfordernde Gruppe von Zielproteinen dar, durch ihre große Bindetasche, die zugänglich ist für Wasser. Trotzdem konnten an anderer Stelle in dieser Arbeit Liganden entdeckt und verifiziert werden, die an chemokine GPCRs binden. Grundsätzlich werden Liganden beim Docking selektiert aufgrund ihrer räumlichen und chemischen Komplementarität zur Bindetasche, nicht aufgrund der Antwort, die sie im Rezeptor induzieren. Als Konsequenz kann der induzierte Effekt solcher Liganden sehr klein sein, wie es auch beim CXCR3-CXCR4 und CCR5 Beispiel ersichtlich war. Das gilt auch, wenn die Bindeaffinität hoch ist. Folglich ist das Zebrafisch *in vivo* System möglicherweise ungeeignet um die Modulation der CXCR4 Isoformen nachzuweisen.

Auch beim Beispiel von CCR5 werden mögliche Probleme von GPCR Assaysystemen deutlich, da die Wiederholbarkeit der initialen Screenings problematisch war. Assayprotokolle sind unter Umständen nicht optimiert für Substanzen mit nur schwach ausgeprägten Effekten und im unteren Bereich der Sensitivität nicht ausreichend verlässlich. Nichtsdestotrotz konnten beim ersten Screening erfolgreich Substanzen entdeckt werden, die mit verschiedenen Ausgangsstrukturen ausgewählt wurden. Die Tatsache, dass verschiedene Strukturen verschiedene Liganden binden können ist von Interesse bei der Anwendung von Docking gegen mehrere Zielproteine. Man kann sich verschiedene Strukturen eines bestimmten Proteins vorstellen, die verschiedene Gruppen von Liganden binden. Diese Gruppen können teilweise überlappen oder nicht. Für die Vorhersage, ob ein Ligand an verschiedene Proteine bindet oder nicht, muss entsprechend wohl überlegt

sein, welche Struktur für jedes der Proteine gewählt wird.

Für die Familie der Proteinkinasen existieren deutlich mehr Kristallstrukturen als für GPCRs. Das machte es möglich einen Datensatz zusammen zu stellen mit Kristallstrukturen und Kinaseinhibitoren, für die experimentelle Bindungsdaten existieren. Ein nicht-überwachtes Docking gegen alle Strukturen dieses Datensatzes legt nahe, dass die Vorhersage von Bindeprofilen in einem großen Maßstab nicht möglich ist mit existierenden Methoden. Im Gegenteil scheint die überlegte Auswahl und Präparierung von Strukturen der Schlüssel zu einem erfolgreichen Docking gegen mehrere Strukturen zu sein. Mehrere Ideen sind aus diesem Experiment entstanden um die Vorhersagegenauigkeit von Docking zu verbessern, zum Beispiel Proteindecoys zusätzlich zu Ligandendecoys zu benutzen. Neue Ansätze wie diese werden künftig helfen, den Erfolg von solch herausfordernden Docking-Szenarien zu verbessern. Dadurch wird es möglich sein, Docking breiter einzusetzen um Liganden mit maßgeschneiderten Selektivitäten zu entwickeln.

Zusammengefasst zeigt das CXCR3-CXCR4 Fallbeispiel dass Docking Liganden mit dualem Bindungsprofil vorhersagen kann. Allerdings haben GPCRs einige Besonderheiten. Abhängig davon wo sie binden und welche Konformationsänderung sie hervorrufen, können Liganden als allosterische Modulatoren, Antagonisten oder Agonisten wirken, oder binden ohne eine Signalantwort zu induzieren. Wenn es eine Antwort gibt, kann diese auf verschiedenen Pfaden weitergeleitet werden und zu verschiedenen Signalen führen. Alles in allem sind GPCRs vermutlich zu komplex um die Experimente auszuweiten auf eine größere Zahl von unähnlicheren Strukturen. Kinasen könnten hingegen besser geeignet sein für zukünftige Studien mit mehr Zielproteinen, durch die Verfügbarkeit von Kristallstrukturen und experimentellen Möglichkeiten. Trotzdem hebt der Hochdurchsatz-Dockingansatz gegen den Kinasedatensatz hervor, dass eine korrekte Anwendung von Docking durch einen erfahrenen Experimentator nötig ist, genauso wie Methoden wie die Normalisierung von Dockingwerten durch Liganden- und Proteindecoys.

Die vorgestellte Arbeit zeigt die Machbarkeit, Dockingansätze zu nutzen

um Liganden mit maßgeschneiderten Selektivitäten zu entwickeln. Zukünftige Experimente werde die Zahl der genutzten Strukturen erweitern, vermutlich durch die Nutzung einer neuen Klasse von Zielproteinen, wie die Kinasen. Mit steigender Erfahrung und eventuell verbesserten Algorithmen werden kommende Docking-Ansätze hoffentlich die Vorhersage von Selektivitätsprofilen gegen eine größere Auswahl von Zielproteinen erlauben, sogar über verschiedene Proteinfamilien hinweg.

Bibliography

- [1] Kuntz, ID, Meng, EC, Oatley, SJ, Langridge, R, *et al.* A geometric approach to macromolecule-ligand interactions. *J Mol Biol*, **161**:269–288 (1982).
- [2] Roth, BL, Sheffler, DJ, and Kroeze, WK. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov*, **3**(4):353–359 (2004).
- [3] Frantz, S. Drug discovery: playing dirty. *Nature*, **437**(7061):942–943 (2005).
- [4] Knight, ZA, Lin, H, and Shokat, KM. Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer*, **10**(2):130–137 (2010).
- [5] Scior, T, Bender, A, Tresadern, G, Medina-Franco, JL, *et al.* Recognizing Pitfalls in Virtual Screening: A Critical Review. *J Chem Inf Model*, **52**(4):867–881 (2012).
- [6] Kolb, P and Irwin, J. Docking Screens: Right for the Right Reasons? *Curr Top Med Chem*, **9**(9):755–770 (2009).
- [7] Lipinski, CA, Lombardo, F, Dominy, BW, and Feeney, PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, **64**(SUPPL.):4–17 (2012).
- [8] Leach, AR, Shoichet, BK, and Peishoff, CE. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J Med Chem*, **49**(20):5851–5855 (2006).

- [9] Neudert, G and Klebe, G. DSX: A Knowledge-Based Scoring Function for the Assessment of Protein-Ligand Complexes. *J Chem Inf Model*, **51**(10):2731–2745 (2011).
- [10] Gohlke, H and Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chemie - Int Ed*, **41**(15):2644–2676 (2002).
- [11] Enyedy, IJ and Egan, WJ. Can we use docking and scoring for hit-to-lead optimization? *J Comput Aided Mol Des*, **22**(3-4):161–168 (2008).
- [12] Claußen, H, Buning, C, Rarey, M, Lengauer, T, *et al.* FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol*, **308**(2):377–395 (2001).
- [13] Österberg, F, Morris, GM, Sanner, MF, Olson, AJ, *et al.* Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins Struct Funct Genet*, **46**(1):34–40 (2002).
- [14] Krimmer, SG, Betz, M, Heine, A, and Klebe, G. Methyl, ethyl, propyl, butyl: Futile but not for water, as the correlation of structure and thermodynamic signature shows in a congeneric series of thermolysin inhibitors. *ChemMedChem*, **9**(4):833–846 (2014).
- [15] Overington, JP, Al-Lazikani, B, and Hopkins, AL. How many drug targets are there? *Nat Rev Drug Discov*, **5**(12):993–996 (2006).
- [16] McGrath, NA, Brichacek, M, and Njardarson, JT. A Graphical Journey of Innovative Organic Architectures That Have Improved Our Lives. *J Chem Educ*, **87**(12):1348–1349 (2010).
- [17] Berman, HM, Westbrook, J, Zukang, F, Gilliland, G, *et al.* The Protein Data Bank. *Nucleic Acids Res*, **28**(1):235–242 (2000).
- [18] Cherezov, V, Rosenbaum, DM, Hanson, MA, Rasmussen, SGF, *et al.* High-resolution crystal structure of an engineered human β_2 -adrenergic G protein-coupled receptor. *Science*, **318**(5854):1258–1265 (2007).

- [19] Rosenbaum, DM, Cherezov, V, Hanson, MA, Rasmussen, SGF, *et al.* GPCR engineering yields high-resolution structural insights into β_2 -adrenergic receptor function. *Science*, **318**(5854):1266–1273 (2007).
- [20] Kolb, P, Rosenbaum, DM, Irwin, JJ, Fung, JJ, *et al.* Structure-based discovery of β_2 -adrenergic receptor ligands. *Proc Natl Acad Sci U S A*, **106**(16):6843–6848 (2009).
- [21] Sabio, M, Jones, K, and Topiol, S. Use of the X-ray structure of the β_2 -adrenergic receptor for drug discovery. Part 2: Identification of active compounds. *Bioorg Med Chem Lett*, **18**(20):5391–5395 (2008).
- [22] Qin, S, Rottman, JB, Myers, P, Kassam, N, *et al.* The Chemokine Receptors CXCR3 and CCR5 Mark Subsets of T Cells Associated with Certain Inflammatory Reactions. *J Clin Invest*, **101**(4):746–754 (1998).
- [23] Sørensen, TL, Tani, M, Jensen, J, Pierce, V, *et al.* Expression of specific chemokines and chemokine receptors in the central nervous system of multiple sclerosis patients. *J Clin Invest*, **103**(6):807–815 (1999).
- [24] Hancock, WW, Lu, B, Gao, W, Csizmadia, V, *et al.* Requirement of the Chemokine Receptor CXCR3 for Acute Allograft Rejection. *J Exp Med*, **192**(10):1515–1520 (2000).
- [25] Schmidt, D, Bernat, V, Brox, R, Tschammer, N, *et al.* Identifying Modulators of CXC Receptors 3 and 4 with Tailored Selectivity Using Multi-Target Docking. *ACS Chem Biol*, **10**(3):715–724 (2015).
- [26] Mysinger, MM, Weiss, DR, Ziarek, JJ, Gravel, S, *et al.* Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4. *Proc Natl Acad Sci U S A*, **109**(14):5517–5522 (2012).
- [27] Carlsson, J, Yoo, L, Gao, ZG, Irwin, JJ, *et al.* Structure-Based Discovery of A_{2A} Adenosine Receptor Ligands. *J Med Chem*, **53**(9):3748–3755 (2010).
- [28] Kolb, P, Phan, K, Gao, ZG, Marko, AC, *et al.* Limits of ligand selectivity from docking to models: in silico screening for A₁ adenosine receptor antagonists. *PLoS One*, **7**(11):e49910 (2012).
- [29] Granier, S and Kobilka, B. A new era of GPCR structural and chemical biology. *Nat Chem Biol*, **8**(8):670–673 (2012).

- [30] Kolb, P, Ferreira, RS, Irwin, JJ, and Shoichet, BK. Docking and chemoinformatic screens for new ligands and targets. *Curr Opin Biotechnol*, **20**(4):429–436 (2009).
- [31] Kolb, P and Klebe, G. The Golden Age of GPCR Structural Biology: Any Impact on Drug Design? *Angew Chem Int Ed*, **50**(49):11573–11575 (2011).
- [32] Wu, B, Chien, EYT, Mol, CD, Fenalti, G, *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*, **330**(6007):1066–1071 (2010).
- [33] Keiser, MJ, Irwin, JJ, and Shoichet, BK. The chemical basis of pharmacology. *Biochemistry*, **49**(48):10267–10276 (2010).
- [34] Cavasotto, CN, Orry, AJW, Murgolo, NJ, Czarniecki, MF, *et al.* Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J Med Chem*, **51**(3):581–588 (2008).
- [35] Carlsson, J, Coleman, RG, Setola, V, Irwin, JJ, *et al.* Ligand discovery from a dopamine D₃ receptor homology model and crystal structure. *Nat Chem Biol*, **7**(11):769–778 (2011).
- [36] Akdemir, A, Edink, E, Thompson, AJ, Lummis, SCR, *et al.* Identification of novel $\alpha 7$ nicotinic receptor ligands by in silico screening against the crystal structure of a chimeric $\alpha 7$ receptor ligand binding domain. *Bioorg Med Chem*, **20**(19):5992–6002 (2012).
- [37] Sirci, F, Istyastono, EP, Vischer, HF, Kooistra, AJ, *et al.* Virtual fragment screening: Discovery of histamine H₃ receptor ligands using ligand-based and protein-based molecular fingerprints. *J Chem Inf Model*, **52**(12):3308–3324 (2012).
- [38] Besnard, J, Ruda, GF, Setola, V, Abecassis, K, *et al.* Automated design of ligands to polypharmacological profiles. *Nature*, **492**(7428):215–220 (2012).
- [39] Dundas, J, Ouyang, Z, Tseng, J, Binkowski, A, *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res*, **34**(Web Server issue):W116–118 (2006).

- [40] Christopoulos, A and Kenakin, T. G protein-coupled receptor allostereism and complexing. *Pharmacol Rev*, **54**(2):323–374 (2002).
- [41] Balkwill, F. Cancer and the chemokine network. *Nat Rev Cancer*, **4**(7):540–550 (2004).
- [42] Endres, MJ, Clapham, PR, Marsh, M, Ahuja, M, *et al.* CD4-Independent Infection by HIV-2 Is Mediated by Fusin/CXCR4. *Cell*, **87**(4):745–756 (1996).
- [43] Kohler, RE, Comerford, I, Townley, S, Haylock-Jacobs, S, *et al.* Antagonism of the chemokine receptors CXCR3 and CXCR4 reduces the pathology of experimental autoimmune encephalomyelitis. *Brain Pathol*, **18**(4):504–516 (2008).
- [44] Murakami, T, Kawada, K, Iwamoto, M, Akagami, M, *et al.* The role of CXCR3 and CXCR4 in colorectal cancer metastasis. *Int J Cancer*, **132**(2):276–287 (2013).
- [45] Gaulton, A, Bellis, LJ, Bento, AP, Chambers, J, *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*, **40**(D1):D1100–D1107 (2012).
- [46] Irwin, JJ, Sterling, T, Mysinger, MM, Bolstad, ES, *et al.* ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model*, **52**(7):1757–1768 (2012).
- [47] OpenEye Toolkits 2014.Feb.4.
- [48] Princen, K, Hatse, S, Vermeire, K, Aquaro, S, *et al.* Inhibition of human immunodeficiency virus replication by a dual CCR5/CXCR4 antagonist. *J Virol*, **78**(23):12996–13006 (2004).
- [49] Sabroe, I, Peck, MJ, Van Keulen, BJ, Jorritsma, A, *et al.* A small molecule antagonist of chemokine receptors CCR1 and CCR3. Potent inhibition of eosinophil function and CCR3-mediated HIV-1 entry. *J Biol Chem*, **275**(34):25985–25992 (2000).
- [50] Lovell, SC, Davis, IW, Arendall, WB, De Bakker, PIW, *et al.* Structure validation by $C\alpha$ geometry: ϕ, ψ and $C\beta$ deviation. *Proteins Struct Funct Genet*, **50**(3):437–450 (2003).

- [51] Triballeau, N, Acher, F, Brabet, I, Pin, JP, *et al.* Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem*, **48**(7):2534–2547 (2005).
- [52] Ballesteros, JA and Weinstein, H. *Receptor Molecular Biology*, volume 25 of *Methods in Neurosciences*. Elsevier (1995).
- [53] Scholten, DJ, Roumen, L, Wijtmans, M, Verkade-Vreeker, MCA, *et al.* Identification of overlapping but differential binding sites for the high-affinity CXCR3 antagonists NBI-74330 and VUF11211. *Mol Pharmacol*, **85**(1):116–126 (2014).
- [54] Rosenkilde, MM, Gerlach, LO, Jakobsen, JS, Skerlj, RT, *et al.* Molecular mechanism of AMD3100 antagonism in the CXCR4 receptor: transfer of binding site to the CXCR3 receptor. *J Biol Chem*, **279**(4):3033–3041 (2004).
- [55] Rosenkilde, MM, Andersen, MB, Nygaard, R, Frimurer, TM, *et al.* Activation of the CXCR3 Chemokine Receptor through Anchoring of a Small Molecule Chelator Ligand between TM-III, -IV, and -VI. *Mol Pharmacol*, **71**(3):930–941 (2007).
- [56] Baell, JB and Holloway, Ga. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem*, **53**(7):2719–2740 (2010).
- [57] Colvin, Ra, Campanella, GSV, Manice, La, and Luster, AD. CXCR3 requires tyrosine sulfation for ligand binding and a second extracellular loop arginine residue for ligand-induced chemotaxis. *Mol Cell Biol*, **26**(15):5838–5849 (2006).
- [58] BreLOT, A, Heveker, N, Adema, K, Hosie, MJ, *et al.* Effect of mutations in the second extracellular loop of CXCR4 on its utilization by human and feline immunodeficiency viruses. *J Virol*, **73**(4):2576–2586 (1999).
- [59] BreLOT, A, Heveker, N, Montes, M, and Alizon, M. Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem*, **275**(31):23736–23744 (2000).

- [60] Rosenkilde, MM, Benned-Jensen, T, Frimurer, TM, and Schwartz, TW. The minor binding pocket: a major player in 7TM receptor activation. *Trends Pharmacol Sci*, **31**(12):567–574 (2010).
- [61] Bernat, V, Heinrich, MR, Baumeister, P, Buschauer, A, *et al.* Synthesis and application of the first radioligand targeting the allosteric binding pocket of chemokine receptor CXCR3. *ChemMedChem*, **7**(8):1481–1489 (2012).
- [62] Coan, KED and Shoichet, BK. Stoichiometry and physical chemistry of promiscuous aggregate-based inhibitors. *J Am Chem Soc*, **130**(29):9606–9612 (2008).
- [63] Sassano, MF, Doak, AK, Roth, BL, and Shoichet, BK. Colloidal aggregation causes inhibition of G protein-coupled receptors. *J Med Chem*, **56**(6):2406–2414 (2013).
- [64] Pease, JE and Horuk, R. Chemokine receptor antagonists: part 2. *Expert Opin Ther Pat*, **19**(2):199–221 (2009).
- [65] Abad-Zapatero, C and Metz, JT. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov Today*, **10**(7):464–469 (2005).
- [66] Sali, A and Blundell, TL. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Biol*, **234**:779–815 (1993).
- [67] Arnold, K, Bordoli, L, Kopp, J, and Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**(2):195–201 (2006).
- [68] Krieger, E, Koraimann, G, and Vriend, G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins*, **47**(3):393–402 (2002).
- [69] Jacobson, MP, Friesner, RA, Xiang, Z, and Honig, B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J Mol Biol*, **320**(3):597–608 (2002).
- [70] Jacobson, MP, Kaminski, GA, Friesner, RA, and Rapp, CS. Force Field Validation Using Protein Side Chain Prediction. *J Phys Chem B*, **106**(44):11673–11680 (2002).

- [71] Trott, O and Olson, AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*, **31**(2):455–461 (2010).
- [72] Meng, EC, Shoichet, BK, and Kuntz, ID. Automated docking with grid-based energy evaluation. *J Comp Chem*, **13**:505–524 (1992).
- [73] Shoichet, BK and Kuntz, ID. Matching chemistry and shape in molecular docking. *Protein Eng*, **6**(7):723–732 (1993).
- [74] Shoichet, BK, Leach, AR, and Kuntz, ID. Ligand solvation in molecular docking. *Proteins*, **34**(1):4–16 (1999).
- [75] Brooks, BR, Bruccoleri, RE, Olafson, BD, States, DJ, *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*, **4**:187–217 (1983).
- [76] Bernat, V, Admas, TH, Brox, R, Heinemann, FW, *et al.* Boronic acids as probes for investigation of allosteric modulation of the chemokine receptor CXCR3. *ACS Chem Biol*, **9**(11):2664–2677 (2014).
- [77] Landrum, G. RDKit: Open-source chemoinformatics.
- [78] Köster, H, Craan, T, Brass, S, Herhaus, C, *et al.* A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J Med Chem*, **54**(22):7784–7796 (2011).
- [79] Nagasawa, T, Hirota, S, Tachibana, K, Takakura, N, *et al.* Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature*, **382**(6592):635–638 (1996).
- [80] Tachibana, K, Hirota, S, Iizasa, H, Yoshida, H, *et al.* The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract. *Nature*, **393**(6685):591–594 (1998).
- [81] Boldajipour, B, Doitsidou, M, Tarbashevich, K, Laguri, C, *et al.* Cxcl12 evolution–subfunctionalization of a ligand through altered interaction with the chemokine receptor. *Development*, **138**(14):2909–2914 (2011).

- [82] Doitsidou, M, Reichman-Fried, M, Stebler, J, Köprunner, M, *et al.* Guidance of Primordial Germ Cell Migration by the Chemokine SDF-1. *Cell*, **111**(5):647–659 (2002).
- [83] Siekmann, AF, Standley, C, Fogarty, KE, Wolfe, SA, *et al.* Chemokine signaling guides regional patterning of the first embryonic artery. *Genes Dev*, **23**(19):2272–2277 (2009).
- [84] Bussmann, J, Wolfe, SA, and Siekmann, AF. Arterial-venous network formation during brain vascularization involves hemodynamic regulation of chemokine signaling. *Development*, **138**(9):1717–1726 (2011).
- [85] Irwin, JJ and Shoichet, BK. ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model*, **45**(1):177–182 (2005).
- [86] Melo, F, Sánchez, R, and Šali, A. Statistical potentials for fold assessment. *Protein Sci*, **11**(2):430–448 (2002).
- [87] Shen, MY and Šali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci*, **15**(11):2507–2524 (2006).
- [88] Evers, A and Klebe, G. Ligand-supported homology modeling of G protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew Chem Int Ed Engl*, **43**(2):248–251 (2004).
- [89] Keiser, MJ, Roth, BL, Armbruster, BN, Ernsberger, P, *et al.* Relating protein pharmacology by ligand chemistry. *Nat Biotechnol*, **25**(2):197–206 (2007).
- [90] Bradford, Y, Conlin, T, Dunn, N, Fashena, D, *et al.* ZFIN: enhancements and updates to the Zebrafish Model Organism Database. *Nucleic Acids Res*, **39**(Database issue):D822–D829 (2011).
- [91] The Uniprot Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res*, **42**(Database issue):D191–D198 (2014).
- [92] Edgar, RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, **32**(5):1792–1797 (2004).

- [93] Li, H, Robertson, AD, and Jensen, JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, **61**(4):704–721 (2005).
- [94] Olsson, MHM, Søndergaard, CR, Rostkowski, M, and Jensen, JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput*, **7**(2):525–537 (2011).
- [95] Dolinsky, TJ, Nielsen, JE, McCammon, JA, and Baker, NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*, **32**(Web Server issue):W665–W667 (2004).
- [96] Blanpain, C, Migeotte, I, Lee, B, Vakili, J, *et al.* CCR5 binds multiple CC-chemokines: MCP-3 acts as a natural antagonist. *Blood*, **94**(6):1899–1905 (1999).
- [97] Gerard, C and Rollins, BJ. Chemokines and disease. *Nat Immunol*, **2**(2):108–115 (2001).
- [98] Baba, M, Nishimura, O, Kanzaki, N, Okamoto, M, *et al.* A small-molecule, nonpeptide CCR5 antagonist with highly potent and selective anti-HIV-1 activity. *Proc Natl Acad Sci U S A*, **96**(May):5698–5703 (1999).
- [99] Dorr, P, Westby, M, Dobbs, S, Griffin, P, *et al.* Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob Agents Chemother*, **49**(11):4721–4732 (2005).
- [100] Chothia, C and Lesk, AM. The relation between the divergence of sequence and structure in proteins. *EMBO J*, **5**(4):823–826 (1986).
- [101] Bissantz, C, Kuhn, B, and Stahl, M. A medicinal chemist’s guide to molecular interactions. *J Med Chem*, **53**:5061–5084 (2010).
- [102] Swinney, DC, Beavis, P, Chuang, KT, Zheng, Y, *et al.* A study of the molecular mechanism of binding kinetics and long residence times of human CCR5 receptor small molecule allosteric ligands. *Br J Pharmacol*, **171**(14):3364–3375 (2014).

- [103] Seibert, C, Ying, W, Gavrilov, S, Tsamis, F, *et al.* Interaction of small molecule inhibitors of HIV-1 entry with CCR5. *Virology*, **349**:41 – 54 (2006).
- [104] Maeda, K, Das, D, Ogata-Aoki, H, Nakata, H, *et al.* Structural and molecular interactions of CCR5 inhibitors with CCR5. *J Biol Chem*, **281**(18):12688–12698 (2006).
- [105] Kondru, R, Zhang, J, Ji, C, Mirzadegan, T, *et al.* Molecular interactions of CCR5 with major classes of small-molecule anti-HIV CCR5 antagonists. *Mol Pharmacol*, **73**(3):789–800 (2008).
- [106] Hall, SE, Mao, A, Nicolaidou, V, Finelli, M, *et al.* Elucidation of binding sites of dual antagonists in the human chemokine receptors CCR2 and CCR5. *Mol Pharmacol*, **75**:1325–1336 (2009).
- [107] Hiller, C and Gasteiger, J. Ein automatisierter Molekülbaukasten. In J Gasteiger (editor), *Software-Entwicklung der Chemie*, pages 53–66. Springer, Berlin, 1st edition (1987).
- [108] Gasteiger, J, Rudolph, C, and Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput Methodol*, **3**(6):537–547 (1990).
- [109] Sadowski, J, Rudolph, C, and Gasteiger, J. The generation of 3D models of host-guest complexes. *Anal Chim Acta*, **265**(2):233–241 (1992).
- [110] Dunbrack, RL. Rotamer Libraries in the 21st Century. *Curr Opin Struct Biol*, **12**(4):431–440 (2002).
- [111] Junker, A, Schepmann, D, Yamaguchi, J, Itami, K, *et al.* Diverse modifications of the 4-methylphenyl moiety of TAK-779 by late-stage Suzuki-Miyaura cross-coupling. *Org Biomol Chem*, **12**:177–186 (2014).
- [112] Nygaard, R, Zou, Y, Dror, RO, Mildorf, TJ, *et al.* The dynamic process of β_2 -adrenergic receptor activation. *Cell*, **152**(3):532–542 (2013).
- [113] Dror, RO, Green, HF, Valant, C, Borhani, DW, *et al.* Structural basis for modulation of a G protein-coupled receptor by allosteric drugs. *Nature*, **503**(7475):295–299 (2013).

- [114] Kenakin, T. Principles: Receptor Theory in Pharmacology. *Trends Pharmacol Sci*, **25**(4):186–192 (2004).
- [115] Rajagopal, S, Rajagopal, K, and Lefkowitz, RJ. Teaching old receptors new tricks: biasing seven-transmembrane receptors. *Nat Rev Drug Discov*, **9**(5):373–386 (2010).
- [116] Yin, J, Mobarec, JC, Kolb, P, and Rosenbaum, DM. Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. *Nature*, **519**(7542):247–250 (2014).
- [117] White, JF, Noinaj, N, Shibata, Y, Love, J, *et al.* Structure of the agonist-bound neurotensin receptor. *Nature*, **490**(7421):508–513 (2012).
- [118] Thompson, AA, Liu, W, Chun, E, Katritch, V, *et al.* Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature*, **485**(7398):395–399 (2012).
- [119] Vroiling, B, Sanders, M, Baakman, C, Borrmann, A, *et al.* GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Res*, **39**(Database issue):D309–D319 (2011).
- [120] Ku, H. Notes on the use of propagation of error formulas. *J Res Natl Bur Stand Sect C Eng Instrum*, **70C**(4):263 (1966).
- [121] Chen, YZ and Zhi, DG. Ligand - Protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins Struct Funct Genet*, **43**(2):217–226 (2001).
- [122] Hubbard, SR and Till, JH. Protein Tyrosine Kinase Structure and Function. *Annu Rev Biochem*, **69**:373–398 (2000).
- [123] Finn, RD, Bateman, A, Clements, J, Coghill, P, *et al.* Pfam: The protein families database. *Nucleic Acids Res*, **42**(D1):222–230 (2014).
- [124] Olivera, A and Spiegel, S. Sphingosine kinase: A mediator of vital cellular functions. *Prostaglandins Other Lipid Mediat*, **64**(1-4):123–134 (2001).
- [125] Hanson, MA, Roth, CB, Jo, E, Griffith, MT, *et al.* Crystal Structure of a Lipid G Protein-Coupled Receptor. *Science*, **335**(6070):851–855 (2012).

- [126] Cheung, J, Ginter, C, Cassidy, M, Franklin, MC, *et al.* Structural insights into mis-regulation of protein kinase A in human tumors. *Proc Natl Acad Sci*, **112**(5):1374–1379 (2015).
- [127] Jacobs, MD, Caron, PR, and Hare, BJ. Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: Structure of Ick/imatinib complex. *Proteins Struct Funct Genet*, **70**(4):1451–1460 (2008).
- [128] Yarden, Y and Slivkowski, MX. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol*, **2**(2):127–137 (2001).
- [129] Manning, G, Whyte, DB, Martinez, R, Hunter, T, *et al.* The protein kinase complement of the human genome. *Science*, **298**(5600):1912–1934 (2002).
- [130] Noble, MEM, Endicott, Ja, and Johnson, LN. Protein Kinase Inhibitors: Insights into Drug Design from Structure. *Science*, **303**(5665):1800–1805 (2004).
- [131] Bamborough, P, Brown, MJ, Christopher, JA, Chung, CW, *et al.* Selectivity of kinase inhibitor fragments. *J Med Chem*, **54**(14):5131–5143 (2011).
- [132] Peters, JU. Polypharmacology - Foe or friend? *J Med Chem*, **56**(22):8955–8971 (2013).
- [133] Liu, Y and Gray, NS. Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol*, **2**(7):358–364 (2006).
- [134] Wood, ER, Truesdale, AT, McDonald, OB, Yuan, D, *et al.* A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): Relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res*, **64**(18):6652–6659 (2004).
- [135] Wan, PTC, Garnett, MJ, Roe, SM, Lee, S, *et al.* Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell*, **116**(6):855–867 (2004).
- [136] Karaman, MW, Herrgard, S, Treiber, DK, Gallant, P, *et al.* A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol*, **26**(1):127–132 (2008).

- [137] Huang, N, Shoichet, BK, and Irwin, JJ. Benchmarking Sets for Molecular Docking. *J Med Chem*, **49**(23):6789–6801 (2006).
- [138] Vigers, GPA and Rizzi, JP. Multiple Active Site Corrections for Docking and Virtual Screening. *J Med Chem*, **47**(1):80–89 (2004).
- [139] Casey, FP, Pihan, E, and Shields, DC. Discovery of Small Molecule Inhibitors of Protein–Protein Interactions Using Combined Ligand and Target Score Normalization. *J Chem Inf Model*, **49**(12):2708–2717 (2009).
- [140] Wallach, I, Jaitly, N, Nguyen, K, Schapira, M, *et al.* Normalizing molecular docking rankings using virtually generated decoys. *J Chem Inf Model*, **51**(8):1817–1830 (2011).
- [141] Schmitt, S, Kuhn, D, and Klebe, G. A new method to detect related function among proteins independent of sequence and fold homology. *J Mol Biol*, **323**(2):387–406 (2002).
- [142] Konagurthu, AS, Whisstock, JC, Stuckey, PJ, and Lesk, AM. MUSTANG: A Multiple Structural Alignment Algorithm. *Proteins*, **64**:559–574 (2006).
- [143] Jacobsson, M and Karlén, A. Ligand Bias of Scoring Functions in Structure-Based Virtual Screening. *J Chem Inf Model*, **46**(3):1334–1343 (2006).
- [144] Kruse, AC, Weiss, DR, Rossi, M, Hu, J, *et al.* Muscarinic receptors as model targets and antitargets for structure-based ligand discovery. *Mol Pharmacol*, **84**(4):528–40 (2013).
- [145] Rodríguez, D, Brea, J, Loza, MI, and Carlsson, J. Structure-based discovery of selective serotonin 5-HT_{1B} receptor ligands. *Structure*, **22**(8):1140–1151 (2014).

Acronyms

Arg arginine. 59, 60

Asn asparagine. 60, 90

Asp aspartic acid. 59, 60

ATP adenosine triphosphate. 96–99, 113

AUC area under curve. 22, 105, 107–111, 114

BA basilar artery. 51

cAMP cyclic adenosine monophosphate. 78, 92

CPU Central Processing Unit. 75

CtA central artery. 51

Cys cysteine. 17, 90

DLS Dynamic Light Scattering. 34, 36, 37, 39

DMEM Dulbecco's Modified Eagle's Medium. 44

DMSO dimethyl sulfoxide. 43, 47, 56, 60, 62, 91, 92

DNA deoxyribonucleic acid. 1

Dpr D-proline. 60

EC₅₀ half maximal effective concentration. 46, 83, 85, 87

ECFP extended-connectivity fingerprint. 47, 58

ECL extracellular loop. 20, 40, 41, 66, 68, 90

EDTA ethylenediaminetetraacetic acid. 44, 45

FBS fetal bovine serum. 44

FDA U.S. Food And Drug Administration. 64, 70, 86

GABA γ -Aminobutyric acid. 58

GDP guanosine diphosphate. 7, 45

Gln glutamine. 59, 60, 66–69

Glu glutamic acid. 59, 60, 66, 67, 69, 70, 73, 75, 77, 86, 87, 90

GPCR G protein-coupled receptor. 2

GTP guanosine triphosphate. 7

HBSS Hank's Balanced Salt Solution. 92

HEK human embryonic kidney. 44, 45

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid. 45, 92

HIV human immunodeficiency virus. 17, 64, 86

IC₅₀ half maximal inhibitory concentration. 20, 39

ICL intracellular loop. 20

Ile isoleucine. 59, 60, 90

LDA lateral dorsal aorta. 50, 51, 57

Leu leucine. 59, 60, 90

Lys lysine. 60

NMR nuclear magnetic resonance. 43

PHBC primordial hindbrain channel. 51, 57

Phe phenylalanine. 22, 59, 60, 70, 71, 73, 74, 90

Pro proline. 67, 68

RMSD root-mean-square distance. 66, 69, 71

RMSG root-mean-square gradient. 59, 90

ROC receiver operatoring characteristic. 22, 100, 105, 107, 110, 111, 114

SAR structure-activity relationship. 64, 74

Thr threonine. 60, 90

TM transmembrane helix. 24, 59, 67–70, 90

Trp tryptophan. 66, 67, 69, 70, 73, 90

Tyr tyrosine. 22, 59, 60, 69, 70, 73, 91

Val valine. 59, 60

About the author

Publications

Journal Articles

- (3) D Schmidt, V Bernat, R Brox, N Tschammer, P Kolb
Identifying Modulators of CXCR3 and CXCR4 with Tailored Selectivity using Multi-Target Docking.
ACS Chem. Biol., **2015**, *10*(3), 715-724.
- (2) D Schmidt, P Kolb
Computergestützte Entwicklung selektiver Liganden G Protein-gekoppelter Rezeptoren.
Dtsch. Med. Wochenschr., **2013**, *138*, 2260-2264. [review, german]
- (1) M Vockenhuber, CM Sharma, MG Statt, D Schmidt, Z Xu, S Dietrich, H Liesegang, DH Mathews & B Suess
Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*.
RNA Biol., **2011**, *8*, Epub May 2011.

Presentations

- 04/2015** “Identifying modulators of CXCR3 and CXCR4 with tailored selectivity using multi-target docking.”
GLISTEN Allschwil 2015 Conference
Invited Speaker
- 08/2014** “Virtual Screening Using X-Ray and Modeled Structures of the Chemokine Receptor CCR5”

EuroQSAR 2014, St. Petersburg, Russia

Poster Presentation

11/2013 “Large-scale Docking Approaches to the Kinome”

German Conference on Chemoinformatics, Fulda, Germany

Poster Presentation

03/2013 “Identification of Novel Modulators of CXC Receptors 3 and 4
with Tailored Selectivity”

GRC Computer-Aided Drug Design, Boston, USA

Poster Presentation & Oral Poster Presentation

07/2012 “Computer-Aided Discovery of Novel Chemotypes Addressing
the Chemokine Receptor CXCR3”

Summer School -Drug Design-, Vienna, Austria

Poster Presentation

Curriculum Vitae (deutsch)

Aus Gründen des Persönlichkeitsschutzes wird von der elektronischen
Veröffentlichung des Lebenslaufes abgesehen.

Eidesstattliche Erklärung

Gemäß §10 der Promotionsordnung vom 22.04.2009 versichere ich, dass ich meine Dissertation mit dem Titel

**“Identification of Ligands with Tailored
Selectivity: Strategies & Application”**

selbstständig und ohne unerlaubte Hilfe angefertigt habe. Es wurden dabei keine anderen als die von mir ausdrücklich bezeichneten Quellen genutzt. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 28.05.2015

.....
Denis Schmidt