Computer-Aided Structure-Based Drug Discovery: CXCL12, \textit{P. aeruginosa} LpxA, and the Tiam1 PDZ Domain

Emmanuel William Smith

\textit{University of South Florida, emmanuelwsmith@gmail.com}

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the Medical Biochemistry Commons, and the Medical Molecular Biology Commons

Scholar Commons Citation


This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Computer-Aided Structure-Based Drug Discovery: 
CXCL12, *P. aeruginosa* LpxA, and the Tiam1 PDZ Domain

by

Emmanuel W. Smith

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
Department of Molecular Medicine
College of Medicine
University of South Florida

Major Professor: Yu Chen, Ph.D.
Robert J. Deschenes, Ph.D.
Vladimir Uversky, Ph.D.
Wayne C. Guida, Ph.D.

Date of Approval:
November 10, 2014

Keywords: Virtual Screening, Molecular Docking, X-ray Crystallography, Structural Biology

Copyright © 2014, Emmanuel W. Smith
Dedication

To my loving grandmother, Roula.
Acknowledgments

I would like to acknowledge those who helped me aim for my doctorate and those who helped me succeed: I thank my family for their support all those years. I especially thank Yu Chen for his mentorship that fueled my passion for science. I thank my lab members for making our work environment a fun and growing experience. I finally thank all my committee members for their guidance and advice.
Table of Contents

List of Tables ........................................................................................................ iv
List of Figures ........................................................................................................ v
List of Abbreviations ............................................................................................. vii
Abstract ................................................................................................................ ix

Chapter 1: Introduction to Structure-Based Drug Discovery ................................................................. 1
  Structure-based drug discovery .............................................................................. 1
    Fragment-based drug design ........................................................................... 2
  Molecular docking .............................................................................................. 4
    Protein structures in molecular docking ......................................................... 5
    DOCK ............................................................................................................. 7
  Virtual screening ............................................................................................... 8
    ZINC database ............................................................................................ 8

Chapter 2: Computer-Aided Structure-Based Drug Discovery Against Sulfotyrosine Recognition Sites on CXCL12 ................................................................................................. 10
  Note to reader #1............................................................................................ 10
  Note to reader #2............................................................................................ 10
  Overview ......................................................................................................... 11
  Introduction .................................................................................................... 11
    Chemokines ................................................................................................. 11
    CXCL12-CXCR4 ......................................................................................... 12
    Sulfotyrosine recognition .......................................................................... 13
    The sY21-binding site on CXCL12 ................................................................ 15
      Identification of ZINC310454 ................................................................. 16
  Results and Discussion .................................................................................... 18
    Hit-to-lead from ZINC310454 to tetrazole series ........................................ 18
      Tetrazole series design ........................................................................... 18
      Binding determination through 2D NMR ................................................ 19
      Chemotaxis inhibition assays ................................................................ 22
    Binding pose prediction of tetrazole derivative series ................................ 24
      Ortho-substituent tetrazole series ......................................................... 25
      Para-substituent and meta-substituent tetrazole series ......................... 26
      Discussion on docking studies ............................................................... 28
Chapter 3: Pseudomonas aeruginosa LpxA Structural Analysis and Virtual Screening ........................................66
Overview ............................................................................................................................................66
Introduction ......................................................................................................................................67
Results ..............................................................................................................................................71
P. aeruginosa LpxA apo structure .......................................................................................................71
P. aeruginosa LpxA/UDP-GlcNAc complex structure .......................................................................74
P. aeruginosa LpxA/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex structure .........................79
Virtual screening against P. aeruginosa LpxA ..................................................................................85
Hit screening through X-ray crystallography ....................................................................................87
Conclusion ........................................................................................................................................88
### Chapter 4: Virtual Screening Against the PDZ Domain of Tiam1

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>95</td>
</tr>
<tr>
<td>Introduction</td>
<td>95</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>97</td>
</tr>
<tr>
<td>Virtual screening</td>
<td>97</td>
</tr>
<tr>
<td>2D NMR screening</td>
<td>98</td>
</tr>
<tr>
<td>Conclusion</td>
<td>100</td>
</tr>
<tr>
<td>Experimental procedures</td>
<td>103</td>
</tr>
<tr>
<td>Virtual screening</td>
<td>103</td>
</tr>
</tbody>
</table>

### References

References .................................................................................................................. 105

### Appendix A: X-ray Crystallography Data Collection and Refinement Statistics

Appendix A: X-ray Crystallography Data Collection and Refinement Statistics............ 114


Appendix B: Permission to Use Manuscript: Smith 2014, J. Med. Chem......................... 118
List of Tables

Table 2.1: Tetrazole series based on hit compound ZINC310454.................................23

Table 2.2: CXCL12/BWA3111 complex crystallographic data collection and refinement
statistics.................................................................................................................................114

Table 3.1: LpxA apo crystallographic data collection and refinement statistics ..................115

Table 3.2: LpxA/UDP-GlcNAc crystallographic data collection and refinement statistics ......116

Table 3.3: UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc crystallographic data collection and
refinement statistics ..................................................................................................................117
List of Figures

Figure 1.1: Computer-aided structure-based drug discovery cycle ........................................3
Figure 2.1: CXCL12 binds to the receptor in a two-site/two-step process ...............................13
Figure 2.2: sY binding sites on CXCL12 ................................................................................15
Figure 2.3: Sulfotyrosine binding on sY21-binding site of CXCL12 .......................................17
Figure 2.4: New series of tetrazole-substituents based on hit compound ZINC310454 ..........19
Figure 2.5: Residue assignment from chemical shift perturbations for compounds of tetrazole series ..................................................................................................................20
Figure 2.6: Mapping of chemical shifts to CXCL12 monomeric model .................................21
Figure 2.7: Chemotaxis assay ..................................................................................................23
Figure 2.8: Docking predictions of tetrazole series (BWA3106-14) .......................................29
Figure 2.9: Complex crystallographic structure of CXCL12 dimer .......................................32
Figure 2.10: Complex crystallographic structure and characterization of the sY21 binding site ........................................................................................................................................35
Figure 2.11: C-series hit-to-lead optimization design scheme based on BWA3111 scaffold .................................................................................................................................................38
Figure 2.12: NMR titration assays showing binding of C-series compounds .......................40
Figure 2.13: Docking predictions for C-series compounds based on crystallographic CXCL12/BWA3111 complex structure .........................................................................................42
Figure 2.14: YL-series hit-to-lead optimization design scheme based on BWA3111 scaffold .................................................................................................................................................44
Figure 2.15: NMR titration assays showing binding of YL-series compounds .......................45
Figure 2.16: Docking predictions for YL-series compounds using an ensemble approach .......47
Figure 2.17: ZINC816454 2D NMR studies shows binding to CXCL12 ........................................50
Figure 2.18: sY12-binding site on CXCL12..................................................................................52
Figure 2.19: 2D NMR screening for sY12-binding site binders......................................................54
Figure 2.20: NMR screening of compounds against sY12-binding site, mapping of residues and docking poses .............................................................................................................56
Figure 2.21: NMR screening of compounds against sY21 and sY12-binding sites, mapping of residues and docking poses ...............................................................57
Figure 2.22: Identification of ZINC00095441 using ensemble virtual screening .........................58
Figure 2.23: Identification of weak binders to sY7-binding site using ensemble virtual screening .................................................................................................................................60
Figure 3.1: *P. aeruginosa* LpxA catalyzes the first and reversible step of Lipid A biosynthesis .................................................................................................................................68
Figure 3.2: *P. aeruginosa* LpxA apo crystal structure......................................................................73
Figure 3.3: UDP-GlcNAc bound to *P. aeruginosa* LpxA active site ............................................76
Figure 3.4: Structural comparisons and possible secondary site for UDP-GlcNAc .......................78
Figure 3.5: UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc bound to *P. aeruginosa* LpxA active site ........................................................................................................................................80
Figure 3.6: Structural comparisons to ortholog LpxA/product complexes ........................................82
Figure 3.7: Alternative conformation observed for highly conserved residues Gly139 and Ala138 ........................................................................................................................................84
Figure 3.8: Strategies in creating matching spheres for *P. aeruginosa* LpxA virtual screening .................................................................................................................................86
Figure 3.9: Unbiased Fo-Fc map suggesting presence of potential binders ..................................87
Figure 4.1: Syndecan1 binding site on Tiam1 PDZ domain............................................................99
Figure 4.2: ZINC compounds that bind to Tiam1 PDZ domain ......................................................101
Figure 4.3: Affinities of ZINC compounds binding to the Tiam1 PDZ domain .........................102
List of Abbreviations

NMR ............................................................................................... Nuclear Magnetic Resonance
HTS ................................................................................................. High Throughput Screening
SPR ................................................................................................. Surface Plasmon Resonance
CXCL12 ....................................................................................... CXC type chemokine Ligand 12
LpxA ............................................................................................... UDP-N-acetylglucosamine acyltrasferase
Tiam1 .............................................................................................. T-cell lymphoma Invasion And Metastasis gene 1
GEF ................................................................................................. Guanine Exchange Factor
GAP ................................................................................................. GTPase-Activating Protein
DH-PH .......................................................................................... Dbl homology - pleckstrin homology
Dvl ................................................................................................. Dishevelled
SDC1 ............................................................................................... Syndecan1
CXCR4 ........................................................................................... CXC type Receptor 4
sY ................................................................................................. sulfotyrosine
GPCR .............................................................................................. G-Protein Coupled-Receptor
SDF-1 .............................................................................................. Stromal-Cell-Derived Factor-1
CXCR7 ........................................................................................... CXC type Receptor 7
GAG ............................................................................................... Glycosaminoglycan
PPI ................................................................................................. Protein-Protein Interface
PDB ................................................................................................. Protein Data Bank
SAR.................................................................Structure Activity Relationship
CCL2.............................................................CC type chemokine Ligand 2
LE.................................................................Ligand Efficiency
HPLC ..........................................................High Performance Liquid Chromatography
HMQC .........................................................Heteronuclear Multiple Quantum Coherence
HSQC..........................................................Heteronuclear Single Quantum Coherence
ACP..............................................................Acyl Carrier Protein
LPS.............................................................Lipopolysaccharide
UDP.............................................................Uridine Diphosphate
LβH .............................................................Left-handed parallel beta-Helix
GlcNAc ........................................................N-acetylglucosamine
Abstract

For structure-based drug discovery, structural information of a target protein is necessary. NMR, or X-ray crystallography can provide necessary information on active site configuration that can lead a successful virtual screening campaign into identifying binders that may then be optimized into potent inhibitors. However, many challenges exist in the structure-based drug discovery cycle. For instance, structure determination of a protein of interest can many times be a daunting task. In addition, complex structure determination, which can allow essential characterization of protein-ligand interactions, is also challenging and many times impossible. Virtual screening heavily relies on such structural information, but hit-to-lead optimization schemes do as well. Furthermore, inherent protein characteristics such as conformational flexibility only add to the complexities in using structural information to identifying and optimizing inhibitors. In the scope of the work presented here, a structure-based drug discovery approach against three different protein targets is described. Each is presented with it’s own set of challenges, but each has successfully led to the identification of new ligands.

The drug discovery project against CXCL12 will first be described. CXCL12 is a small chemokine (~10KDa) that binds to the CXCR4 receptor promoting chemotaxis of lymphocytes but also metastasis of cancer cells. This interaction is further supported by sulfated tyrosines on CXCR4 that bind specific sites on the CXCL12 surface. The CXCL12-CXCR4 signaling axis has been a major focus of drug discovery, but efforts are mainly focused on CXCR4, since CXCL12 is a small protein lacking surface characteristics that are thought to be druggable. Yet,
through a combination of rigid, flexible, and ensemble docking in virtual screening studies, we have successfully identified compounds that bind each of the three sulfotyrosine recognition sites on CXCL12, which normally bind the sulfated tyrosines on CXCR4 (sY7, sY12, and sY21). Furthermore, we have led a hit-to-lead approach in optimizing compounds against the sY21-binding site, aided by trivial information gained through crystallographic complex structure determination of CXCL12 bound by such a compound. We aim to eventually link compounds against different sites together and greatly improve potency.

Next, the drug discovery project against *P. aeruginosa* LpxA will be described. In Gram-negative bacteria, the first step of lipid A biosynthesis is catalyzed by UDP-N-acetylglucosamine acyltrasferase (LpxA) through the transfer of a R-3-hydroxyacyl chain from the acyl carrier protein (ACP) to the 3'-hydroxyl group of UDP-GlcNAc. Acyl chain length selectivity varies between species of bacteria, but is highly specific and conserved within certain species. In *E. coli* and *L. interrogans* for example, LpxA is highly selective for longer R-3-hydroxyacyl chains (C14 and C12 respectively), while in *P. aeruginosa* the enzyme is highly selective for R-3-hydroxydecanoyl, a 10-hydrocarbon long acyl chain. Three *P. aeruginosa* LpxA crystal structures will be described here for the first time; the apo form, the complex with its substrate UDP-GlcNAc, and the complex with its product UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc. A comparison between the APO form and complexes identifies key residues that position UDP-GlcNAc appropriately for catalysis, and supports the role of His121 in generating the nucleophile by interacting with the UDP-GlcNAc 3'-hydroxyl group. Furthermore, the product-complex structure supports the role of Met169 as the “hydrocarbon ruler”, providing structural information on how *P. aeruginosa* LpxA is granted its exceptional selectivity for the 10-hydrocarbon long acyl
chain. Structural information of the active site was subsequently used in designing virtual screening experiments that led to the identification of two ligands, confirmed by X-ray crystallography screening to bind to the active site. We aim to continue application of X-ray crystallography into screening compound binding, and to also use a hit-to-lead approach in compound optimization.

Finally, the drug discovery project against the Tiam1 PDZ domain will be described. Tiam1 (T-cell lymphoma invasion and metastasis gene 1) is a GEF (guanine exchange factor) protein that activates Rac1 and initiates tumor formation. Tiam1 is regulated through its PDZ domain, which binds to syndecan1. We have successfully applied a virtual screening strategy to an existing crystallographic structure of the Tiam1 PDZ domain complexed to a syndecan1 peptide and identified four ligands that bind to the PDZ domain with low affinities. These compounds provide a starting point for future hit-to-lead optimization strategies.
Chapter 1

Introduction to Structure-Based Drug Discovery

Structure-based drug discovery

Structure-based drug discovery is a rapidly growing field that has witnessed numerous successes in recent years.\textsuperscript{1, 2} This is mostly due to major improvements in the fields of genomics, proteomics, structural biology and computer science. The technological innovations that have had the greatest impact include advancements in cloning techniques, protein production and purification, advancements in NMR technology, advancements in molecular docking programs, and advancements in X-ray crystallography, such as high-throughput screening of crystallization conditions, data collection technology, and strong synchrotron radiation.\textsuperscript{1, 3} Such innovations have made structure-based drug discovery an integral part of the drug discovery process.\textsuperscript{4}

The cycle of structure-based drug discovery starts with the identification of a suitable drug target mostly through biochemical and cellular techniques. The structure of the target protein can then be determined through X-ray crystallography, NMR, or even homology modeling. In addition, determination of the structure of the target protein complexed to a ligand in the active site of interest is extremely valuable since complex structures can provide crucial information on binding site characteristics. Such information can be subsequently used in the structure-based drug discovery process in which virtual (\textit{in silico}) screening can help identify ligands that may potentially bind to the active site.\textsuperscript{4} Alternatively, ligands can also be identified
through HTS (High Throughput Screening), where a library of compounds is tested experimentally for binding to a target protein. After identification of ligands that bind to the protein of interest, structural information on binding interactions between the protein and new ligand is needed to advance the structure-based approach. X-ray crystallography or NMR can provide valuable information on how proteins and ligands interact with each other, but such information may be hard to obtain in certain cases, so other complementary methods such as molecular modeling by docking can guide a hit-to-lead optimization strategy into increasing potency and efficacy. The final product of a successful structure-based drug discovery scheme is a high affinity ligand that can modulate the function of a human protein involved in pathogenesis, or inhibit the function of a protein that is essential to the survival of a pathogenic organism (Fig. 1.1).

**Fragment-based drug design.**

The use of fragments (MW < 200 Da) in hit identification is becoming increasingly popular because it offers many benefits; it can streamline the computational process greatly reducing computational time, it can help identify discrete pharmacophores that interact with diverse binding hot spots, provide a rational strategy for optimization, and help create potent lead compounds. SAR analysis in combination with NMR is widely used in identifying important functional groups, since NMR can detect small changes induced upon fragment binding. In addition, since the first report by Abbott et al on the use of X-ray crystallography in fragment screening, X-ray crystallography has also grown in popularity as a tool in identifying and studying fragments that bind active sites of interest.
Fragment-based methods have become more desirable than High Throughput Screening (HTS) methods since larger compounds used in HTS are not chemically diverse enough to explore the chemical space of a target-binding site. This is mostly because larger molecules have typically been pre-designed to be specific for specific targets. Smaller molecules on the other hand can capture more chemical space and result in higher hit-rates. Furthermore, fragment compounds tend to be more soluble and make experimental testing simpler and more precise.

![Figure 1.1. Computer-aided structure-based drug discovery cycle.](image)

The structure-based drug discovery cycle can start when a virtual database of compounds can be docked into the active site of a target protein. A docking program will score the poses of the compounds and sort the compounds based on score. The highest scoring compounds can be tested experimentally for binding. True binders can proceed into the structure-based optimization cycle and lead a hit-to-lead approach into increasing potency.
**Molecular docking**

Molecular docking is the computational tool used to predict the binding pose of a molecule in a macromolecular binding site. It has become an important component of the structure-based drug discovery process and as computational power continues to improve, molecular docking programs have been experiencing many advances in recent years as well.\(^3,^6\)

In molecular docking, numerous poses of a molecule are first modeled into a protein-binding site, and then an energy function scores each interaction.\(^6\) The final binding pose prediction can in many cases be a decent estimate of the actual biologically relevant binding pose, and can also be calculated in a very short time.\(^6\) In fact, docking programs now have the capacity to screen and calculate energy scores for millions of compounds in just a few days.\(^3,^{10}\)

On the other hand though, due to the high intrinsic complexity of protein-ligand interactions, there are many caveats with molecular docking programs. For instance, predicting affinity by docking programs is not practicable because energies calculated are typically not representative.\(^3,^{11}\) Additionally, considering how water molecules interact with both the protein and ligand during binding remains challenging. Furthermore, since proteins *in vivo* exist in highly dynamic states and are therefore inherently flexible, the rigid “lock and key” model to how proteins interact with ligands is not truly representative of the nature of biomolecular interactions. Yet, docking programs most often use the rigid “lock and key” model in order to reduce computational time.

Consequently, docking programs make many assumptions in order to predict binding poses in a timely manner, which is especially important during virtual screening. Assumptions differ from program to program, but they do increase the error rate. For instance, simplified algorithms are used in place of more accurate and complex algorithms, ionization states of amino
acids can be inaccurate, and side chain conformational changes are most times not even considered.\textsuperscript{3, 6}

As a result, the docking approach to virtual screening or to binding pose prediction of individual molecules is not always straightforward and careful examination is required in the design of each experiment, but also in the analysis of docking results. But even careful consideration of all those variables can lead to unsatisfactory results. Therefore, expertise in molecular docking and the tools to quickly validate and study protein-ligand interactions are needed to increase the success rate of a project involving molecular docking.\textsuperscript{12}

**Protein structures in molecular docking**

Molecular docking requires a protein structure as a template. In addition, structural information of a ligand bound to the active site greatly increases the accuracy of the docking prediction, since information on protein-ligand binding can be utilized making energy calculations more accurate. Protein structures can be determined through X-ray crystallography, NMR, or homology modeling, but there are limitations presented with each technique. Structures solved by X-ray crystallography are subject to resolution making docking pose predictions increasingly challenging with decreasing resolution. For high-resolution structures however, precise information on side-chain positions within the binding site can greatly improve docking pose predictions. Yet, crystal structures offer nothing but a snapshot; a single rigid conformation of a structure that may otherwise exist in a highly dynamic state. Consequently, the ideal conformation for binding a ligand may not be represented in the crystal structure since crystal packing may induce the crystal structure conformation. NMR structures on the other hand contain ensembles of multiple conformations that cumulatively may be more representative of a
dynamic protein *in vivo*. However, many times only one of these conformations is chosen for docking, and many times the chosen conformation may be the least favorable for binding.\(^4\) Therefore, the advantage of NMR structures lies in the ensemble combined, and as a result ensembles of NMR conformations have been used in docking to account for protein flexibility.\(^13\) Homology models have also seen rising success in virtual screening studies. Even though they have been considered the least desirable structures for molecular docking due to errors in residue positions, as homology modeling programs improve, the accuracy of the modeled structures improves as well, and with that the usefulness of homology models as templates for molecular docking is also increasing.\(^4\)

There is a general consensus in the molecular docking community on what structures provide the best results, measured in terms of enrichment. Enrichment in molecular docking is a qualitative measurement of performance that accounts for how many true ligands score favorably compared to non-binders. A high enrichment is the result of a successful docking strategy. Usually, docking into an ensemble of different conformations provides the highest enrichment of true binders, but it is computationally the most intensive.\(^13\) X-ray structures come second, especially when docking to deep rigid sites, and this is due to the high-resolution data they provide. Docking into single NMR conformations usually leads to lower enrichment than docking into single X-ray structures since the conformation chosen may be the least suitable for docking. Finally, homology models are the least desirable even though advancements in homology model docking are also occurring.
**DOCK®**

Many docking programs currently exist, but DOCK® was the first such program ever developed. The program works by adding hydrogens to the protein, and then establishes van der Waals parameters and partial charges to both the protein and the compound to be docked. It uses a geometrical matching algorithm to assess the surface area of the binding site and then creates a negative image of the binding site through the use of spheres for which an energy grid containing the computed van der Waals score and charge is calculated. It then matches the atoms of the docked ligand to the spheres, and uses an optimization algorithm to improve fitting. Finally the user can sort the docked compounds based on score.3, 14, 15

Several studies have compared various docking programs for performance and accuracy. In one such study, through the use of stringent metrics meant to reduce bias, DOCK, GLIDE, ICM, FlexX, Surflex, and PhDOCK, were all examined for docking pose accuracy and enrichment in virtual screening, suggesting that GLIDE can outperform DOCK in both.16 However, DOCK outperforms most programs in terms of speed and most importantly compatibility with Flexibase files, making screening the ZINC database fast and straightforward. It can sample a flexible ligand in just a few seconds, and can screen the ZINC leadlike database in just a few days using a modest computing cluster. Furthermore, it allows manipulation of many variables, making expert use very powerful. Finally, DOCK has led to more accurate predictions in binding poses compared to X-ray complex structures that were later solved, than all other programs combined.17
**Virtual screening**

Virtual (*in silico*) screening is a quick and efficient method of discovering ligands to proteins. It is a screening tool in which virtual libraries of compounds are evaluated for potential binding to a protein receptor through a docking program. It is fast and cost-effective, especially in comparison to High Throughput Screening (HTS), since it is using precise three-dimensional information of biologically relevant biomolecules in a virtual setting. Advancements and availability of cloud computing have made virtual screening a significant element in structure-based drug discovery.

The virtual screening process includes various steps, such as processing of the receptor and of a compound library database. Each compound in the library database is then fitted into the receptor binding site and scored by a molecular docking program. Attention needs to be given when selecting an appropriate compound database to dock, but also in making sure that a library is quite diverse so as to capture as much chemical space possible while filtering out chemically undesirable compounds. Consideration of the methodology used to create the docking experiment is also important, since not doing so may result to binding modes that are neither realistic nor reliable. Usually, hits identified through virtual screening possess µM affinities.

**ZINC database**

The ZINC database is a virtual database of compounds that are readily available for purchase. It is a public resource and it is widely used in discovering ligands for proteins. The database contains over 35 million compounds in biologically relevant conformations and a ready-to-dock form.
The ZINC database was created to aid the structure-based drug discovery process. It is therefore large and diverse, containing annotated compounds from over 100 vendors and 150 catalogs. It includes many categories of biologically relevant representations, including compounds in different tautomeric and protonated forms, which can also be divided into four pH ranges. It is categorized into subsets that make searching easy and fast. Subsets are based on properties such as size, structure, activity, availability, vendor, delivery time, and name. It also allows the user to search for reactive groups, metabolites, drugs, building blocks, and analogs. The user can also upload compounds that are not on the ZINC database and process them through ZINC.

For the database to be effective, compounds should be able to be tested in a timely manner. Therefore, purchasing availability and quick delivery are important factors. Consequently, the database was created to implement the 90/90 rule stating that 90% of compounds are available within 90 days of purchase. For that to happen, ZINC is constantly curated and ever-evolving in order to be updated with newly purchasable compounds, but to also remove compounds that are no longer available for purchase.

The lead-like and fragment subsets of the ZINC database are the most commonly used subsets. The lead-like subset contains compounds that are between 250Da and 350Da in mass. These are better candidates when higher affinities are desirable such as when studying inhibition. The fragment-like subset has compounds that are less than 250Da. Those work better with techniques such as NMR, X-ray crystallography, and SPR (Surface Plasmon Resonance), where high compound concentrations can overcome limitations observed in activity assays.
Chapter 2

Computer-Aided Structure-Based Drug Discovery Against Sulfotyrosine Recognition Sites on CXCL12

Note to reader #1

Reprinted (adapted) with permission from Emmanuel W. Smith†, Yan Liu†, Anthony E. Getschman§, Francis C. Peterson§, Joshua J. Ziarek§, Rongshi Li‡, Brian F. Volkman§, and Yu Chen†, Journal of Medicinal Chemistry 2014, [Epub ahead of print]. Copyright 2014 American Chemical Society. (See Appendix B)

Note to reader #2

Members of the Volkman lab at the Medical College of Wisconsin performed the NMR and chemotaxis experiments described in this chapter. Members of the Rongshi Li lab at the Moffitt Cancer Center performed the chemical synthesis of the optimized compounds.

† Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs
‡ Department of Pharmaceutical Sciences, Center for Drug Discovery, College of Pharmacy and Cancer Genes and Molecular Regulation Program, Fred and Pamela Buffett Cancer Center, University of Nebraska Medical Center, 986805 Nebraska Medical Center, Omaha, Nebraska 68198, United States
§ Department of Biochemistry, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, Wisconsin 53226, United States
Overview

CXCL12 is a chemokine that binds to the CXCR4 receptor, promoting both chemotaxis of lymphocytes but also metastasis of cancer cells to areas of elevated chemokine concentration. Both CXCL12 and CXCR4 have become high priority targets because of their involvement in cancer metastasis. However efforts have been exclusively focused on CXCR4 since CXCL12 has a rather featureless surface and lacks the deep hydrophobic pocket that is present on CXCR4. NMR provided structural information on sulfated tyrosines (sY21, sY12, and sY7) of the N-terminus of CXCR4 bound to CXCL12, which increase affinity of CXCR4 for CXCL12. Through the application of a hybrid in silico/NMR screening strategy we have successfully identified ligands that bind to each of the three sulfotyrosine-binding sites. We have confirmed that ligands against the sY21-binding site act as inhibitors, and we have also provided structural information on the binding of such an inhibitor to the sY21-binding site through X-ray crystallography.

Introduction

Chemokines

Chemokines are small but potent chemotactic cytokines (8-14KDa). To date, about 50 chemokines have been identified and divided into four distinct families (C, CC, CXC, & CX3C) based on the arrangement of conserved cysteines in the N-terminus. These secreted proteins orchestrate homing of cells towards areas of high chemokine concentration through binding and activation of their cognate GPCRs (G-protein-coupled receptors) on the surface of cells.
Processes such as cell trafficking and adhesion greatly depend on the chemokine-receptor signaling axis.\textsuperscript{26-28}

**CXCL12-CXCR4**

CXCL12 (SDF-1a, Stromal-Cell-Derived Factor-1) is a CXC-type chemokine that binds to the CXCR4 and CXCR7 receptors attracting receptor-containing cells towards areas of elevated CXCL12 levels. Extracellular matrix glycosaminoglycans (GAGs) also bind CXCL12 and maintain a chemotactic concentration gradient.\textsuperscript{29} CXCL12 is constitutively expressed and essential during embryonic development but afterwards functions mainly in inflammatory response, immune surveillance and tissue homeostasis. This is done through trafficking of lymphocytes to where they are needed such as the lymph nodes, lung, and bone.\textsuperscript{30, 31}

Metastatic cancer cells exploit the same mechanism as lymphocytes by upregulating the expression of chemokine receptors.\textsuperscript{25, 26, 32} CXCR4, for example, is overexpressed in over 23 human cancers, allowing tumor cells to migrate to organs that produce CXCL12, leading to the formation of secondary colonies.\textsuperscript{32, 33} Since metastasis contributes the most to cancer mortality rates, preventing the migration of tumor cells is of paramount medical importance.\textsuperscript{34} As a result, novel inhibitors of the CXCR4-CXCL12 signaling axis have been under active development as potential cancer therapeutics.\textsuperscript{35, 36} Such efforts have mainly focused on the orthosteric site of CXCR4, a deep transmembrane pocket suitable for the binding of small molecule antagonists.\textsuperscript{37} For example, AMD3100 (Plerixafor), a CXCR4 antagonist, has been approved in promoting hematopoietic stem cell mobilization from the bone marrow to the blood in treating multiple myeloma and non-Hodgkin’s lymphoma.\textsuperscript{38} However, recent studies also suggest that neutralizing chemokines may prove to be a successful approach to cancer therapy as well.\textsuperscript{39-41} NOX-A12, an
RNA oligonucleotide in L-configuration that binds CXCL12 and blocks GAG binding is thought to increase the susceptibility of chronic lymphocytic leukemia cells to chemotherapy by interfering with chemokine-mediated cell motility.\textsuperscript{41}

**Sulfotyrosine recognition**

CXCL12 binds and activates CXCR4 in a two-step/two-site process (Fig. 2.1).\textsuperscript{42} First, CXCL12 is recognized by the extracellular N-terminal domain of the receptor (Site 1 binding) (Fig. 2.1B).\textsuperscript{43} Following recognition, the flexible N-terminus of CXCL12 docks into the receptor (Site 2 binding) (Fig. 2.1C), leading to receptor internalization and downstream signaling such as calcium influx and chemotaxis.

![Figure 2.1](image)

**Figure 2.1.** CXCL12 binds to the receptor in a two-site/two-step process: A) CXCR4 has a flexible extracellular N-terminal domain. B) In step-1/site-1, CXCL12 recognizes and binds N-terminal domain of CXCR4 aided by sulfotyrosine recognition. C) In step-2/site-2, the flexible N-terminal domain of CXCL12 docks into CXCR4 causing activation.
Previous studies have shown that the CXCR4 N-terminus is post-translationally sulfated at one or more tyrosines, including Y7, Y12 and Y21, which increases its affinity for CXCL12. Sulfation at Y21 not only contributes the most to enhancing binding affinity but also has the largest effect on downstream signaling.\textsuperscript{44-46} NMR structures of locked CXCL12 dimers, in complex with sulfated (only at Y21 or triply sulfated at Y7, Y12, & Y21) CXCR4\textsubscript{1-38} identified discrete binding pockets for each sulfotyrosine,\textsuperscript{44} suggesting potential target sites for which small molecule ligands can be engineered.

The NMR CXCL12-CXCR4\textsubscript{1-38} complex structure provides detailed information on the composition of these discrete sulfotyrosine-binding sites (sY7, sY12, and sY21-binding sites) (PDB ID: 2K05). All three sites are composed from an assortment of both polar and apolar residues, accommodating the negatively charged sulfate group and the benzene ring of tyrosine, respectively (Fig. 2.2). However, these sites are shallow, a characteristic common in Protein-Protein Interface (PPI) sites, but also rather flexible, as evident by comparison of the different conformations within the NMR ensemble (PDB ID: 2K05). These observations suggest that a structure-based drug discovery approach against such sites will be challenging, considering the lack of deep hydrophobic patches that confer high affinity to small ligands. It was therefore anticipated that a successful drug discovery approach would possibly involve targeting multiple sites, and eventually linking compounds that bind to separate sites together. Consequently, our drug discovery approach has been executed against each sulfotyrosine-binding site on CXCL12 independently, through the use of virtual screens and structural characterization of binding interactions. Our aim has been to identify binders against all sulfotyrosine-binding sites, with the purpose of linking them together and ultimately producing truly potent compounds.
Figure 2.2. sY binding sites on CXCL12. The complex CXCL12-CXCR4<sub>1-38</sub> complex NMR structure (PDB ID: 2K05) identified discrete binding sites on the surface of CXCL12 that bind sulfated tyrosines on the N-terminus of CXCR4 (sY7, sY12, and sY21). Closer examination of each site identifies polar and apolar residues that mediate the interaction between the three sulfotyrosines and their respective sites on CXCL12.

The sY21-binding site on CXCL12. It has been previously shown that the unsulfated CXCR4<sub>1-38</sub> peptide binds with a dissociation constant ($K_d$) of 4.5 ± 2.2 µM. The affinity increases with sulfonation of Y21, to a $K_d$ of 1.3 ± 0.5 µM. Furthermore, residues lining up the sY21-binding site on CXCL12 are conserved among other chemokines. Cumulatively, these
results highlight the overall importance and superiority of the sY21-binding site, when compared to the other two sY-binding sites. Consequently, since the interaction between sY21 and CXCL12 appears to provide most of the binding affinity and specificity between the chemokine and its receptor, our efforts were initially focused on that binding site first.

The NMR complex structure of the constitutive dimeric form of CXCL12, bound by the triply sulfated CXCR4\textsubscript{1-38} peptide (PDB ID: 2K05), provides detailed information on the sY21-binding site composition, and has also been the catalyst for our structure-based drug discovery approach against this particular binding site (Fig. 2.3A-C). The site is rather shallow and has a basic character. The sulfotyrosine is within hydrogen bond distance to Arg47, Asn22, and Asn46 (Fig. 2.3B) and even Asn45 in some conformations within the NMR ensemble. The backbone N atom of sY21 is within hydrogen bond distance to Glu15 (Fig. 2.3B). The aromatic ring of the sulfated tyrosine appears to also take advantage of the hydrophobic surface provided by binding site residues Val18, Val49 and Leu42 (Fig. 2.3C). It has also been suggested that a salt bridge aids in the binding of sY21 to CXCL12. This salt bridge is most likely occurring between Arg47 and the sulfate group (Fig. 2.3B).

**Identification of ZINC310454.** Since the only available structure of CXCL12 bound by any ligand at the sY21-binding site was that of the NMR structure complexed to the CXCR4\textsubscript{1-38} peptide (PDB ID: 2K05), the first conformation out of that NMR ensemble was used as a template for the preliminary virtual screening experiments. The program DOCK was used to screen the lead-like subset of the ZINC database, and a compound (ZINC310454) was identified to bind to CXCL12 with a $K_d$ of 64 µM (Table 2.1). Further testing identified that this inhibitor reduced both calcium flux and chemotaxis in THP-1 cells, indicating inhibition of CXCR4 signaling. Control studies were performed where the newly found inhibitor was tested against
CCL2, a chemokine that also activates Calcium flux in THP-1 cells, but no inhibitory effects were noticed. This further supported that ZINC310454 is a specific binder of CXCL12. Based on subsequent SAR analysis, chemical groups that mediated binding of ZINC310454 to CXCL12 were identified. NMR studies on five analog compounds demonstrated that the benzoic acid was especially important for affinity, possibly by mimicking the negatively charged sulfotyrosine group of the CXCR4 peptide.\textsuperscript{47}

![Figure 2.3. Sulfotyrosine binding on sY21-binding site of CXCL12. A. CXCR4\textsubscript{1-38} sulfated peptide (yellow) bound to CXCL12 dimer (green, blue). B. sY21 is within hydrogen bond distance to Glu15, Asn22, Asn46, and Asn47. C. Surface model shows benzene ring of sY21 may interact with apolar residues Val18, Leu42, and Val49.](image-url)
Results and discussion

**Hit-to-lead from ZINC310454 to tetrazole series**

Based on the SAR analysis done on ZINC310454, a series of new compounds was designed. These compounds were intended to retain important features of the ZINC310454 compound, while acquiring improved drug like properties. First of all, we aimed to increase Ligand Efficiency (LE) and bring it to an LE value comparable to other inhibitors that target PPIs. Compounds that successfully target PPIs typically have LE values that are just under 0.24. Getting an LE value better than that is unusual, since PPIs have a reduced number of possible interactions compared to non-PPI binding sites. Since ZINC310454 has a LE value of 0.23, our optimization efforts were therefore aimed to improve both affinity as well as the LE values. Consequently, in order to redefine the undesirable properties of ZINC310454, we trimmed down the molecule and used a fragment-based approach to built it back up.

**Tetrazole series design.** We first substituted the carboxyl group with its bioisostere; a 5-substituted-1H-tetrazole that is similar to the carboxyl group in size, acidity, and ionization potential, but also possesses improved drug-like properties since it is more stable metabolically. Also, we wanted to test what was the best position for the tetrazole group. We thus created analogs that positioned the tetrazole in the meta, para, and ortho positions, intending to find the ideal position that would efficiently anchor the compound tightly in the binding site. Finally, we wanted to test the importance of the acylthiourea linker of ZINC310454. This could be done through the creation of analogs that have a urea linker or an amide linker instead. If as effective as the acylthiourea linker, these linkers would provide better LE values by decreasing the number of hydrogen donor or acceptor atoms, while still satisfying the same number of polar
interactions between the compound and the sY21-binding site residues. Consequently, we designed and synthesized a series of nine new compounds in the 203-280Da MW range. The compounds were sorted into three groups based on the position of the tetrazole group (ortho, meta, and para positions). For each group, one compound was synthesized with just an amide group, the other with an amide group linking benzene, and the last compound with a urea group linking benzene (Fig. 2.4). Out of these compounds, we anticipated that the para and meta substituents were more likely to bind to the sY21-binding site, since due to their more linear nature, would possibly compliment the narrow sY21-binding site cleft better, especially when compared to the ortho substituents, which may be too bulky.

![Figure 2.4. New series of tetrazole-substituents based on hit compound ZINC310454.](image)

**Figure 2.4. New series of tetrazole-substituents based on hit compound ZINC310454.** The carboxyl group was replaced by a tetrazole group at the ortho, meta, and para positions. The naphthalene was completely removed or replaced by benzene. Amide and urea linkers replaced the acylthiourea linker.

**Binding determination through 2D NMR.** After synthesis of the new tetrazole series, binding was tested through 2D NMR. Compounds were dissolved in DMSO and titrated into solution containing dimeric CXCL12. Then, the perturbations were compared to a titration that
used DMSO without any compound, in order to identify which peaks were not due to DMSO. Surprisingly, all nine compounds produced chemical shift changes that were distinct from the DMSO control (Fig. 2.5). This suggested that all nine compounds bound to CXCL12. Affinities for the compounds that bound the tightest were then calculated. Out of the ortho-substituent series, BWA3107 bound with the highest affinity, highest out of all nine compounds in fact, with a Kd of 13±7 \(\mu\)M and possessed an LE value of 0.33. BWA3108, that has the urea linker, bound with a lower affinity with a Kd of 327±91.6 \(\mu\)M and has an LE value of 0.23. Out of the para-substituents, BWA3110 bound with a Kd of 64±23 \(\mu\)M and has an LE value of 0.29, and BWA3111 with a Kd of 24±28 \(\mu\)M and has an LE value of 0.30. Finally out of the meta-substituents, BWA3113 bound with a Kd of 50±16 \(\mu\)M and has an LE value of 0.29, and BWA3114 bound with a Kd of 126±23 \(\mu\)M and has an LE value of 0.25. The rest of the compounds bound too weakly and affinities could not be determined (Table 2.1).

![Figure 2.5. Residue assignment from chemical shift perturbations for compounds of tetrazole series. All compounds from the tetrazole series (BWA3106-14) caused distinct chemical shift perturbations, which were assigned to residues. For each compound titration, the residues perturbed the most are colored green.](image)
The chemical shifts were then mapped into a model of CXCL12 and indicated distinct patterns between the different analog types (Fig. 2.6); the *ortho*-substituents produced chemical shifts that gathered at the dimer interface, an area between beta-strands from adjacent monomers. That area also overlaps with the sY12-binding site. This further supported our proposition that the *ortho*-substituents, due to their bulkier conformation, may not fit in the narrow sY21-binding site. However, since binding was still observed at the area that overlaps with the sY12-binding site, the hypothesis that tetrazole moieties may occupy similar binding spots as the sulfated tyrosines on CXCR4 was further supported. On the other hand, the other six compounds, the *meta* and *para*-substituents, produced chemical shift perturbations that were consistent with binding to the sY21-binding site, possibly due to their more linear character.

Figure 2.6. Mapping of chemical shifts to CXCL12 monomeric model. Mapping of residues with the greatest chemical shift perturbations (green) suggest that *para* and *meta* substituents (BWA3109-14) may bind to the sY21-binding site while *ortho* substituents (BWA3106-08) may bind at the dimer interface.
Chemotaxis inhibition assays. The new tetrazole series was tested through a chemotaxis assay to show whether any of the compounds inhibited cell migration. The movement of cells through a permeable membrane towards a CXCL12 solution in the absence of any compound was first measured, and the value obtained was used as the baseline 100% chemotactic response. Each compound was then incubated with CXCL12 at a 100 µM concentration and then at a 250 µM concentration, and cell movement was again measured. Compared to the original ZINC310454 hit compound, it appears that most compounds had either negligible, or no improvement on chemotaxis inhibition, except compound BWA3111 (Fig. 2.7). BWA3111 shows significant improvement on chemotaxis inhibition over ZINC310454 at 100 µM, bringing it down to around 37±16%, and almost completely abolishes chemotaxis at 250 µM. These results are encouraging, since a correlation between the chemotaxis assay and NMR affinities has now been demonstrated as well. Finally, BWA3111 was also tested against CCL2 in a chemotaxis assay in order to test specificity. BWA3111 produced no chemotactic inhibition of CCL2, which further supports the specificity of BWA3111 to the CXCL12 chemokine. These results suggest that compound BWA3111 is the best lead compound so far, and that it specifically binds to the sY21-binding site inhibiting binding of CXCL12 to the CXCR4 receptor. Furthermore, these results also suggest that compounds that bind with high affinity to what we believe is the sY12-binding site area, as is believed to be the case for ortho-substituent BWA3106, do not influence the CXCL12-CXCR4 signaling axis. It is possible that the receptor N-terminus may still wrap around CXCL12, by overpassing the compound, or it may be possible that BWA3106 binds to a non-specific site, uninhibited by CXCR4 binding. In fact, it may be possible that BWA3106 may provide a good anchor point for extending a compound from that site outwards, or by extending
a compound from the adjacent and desirable sY21-binding site into the site where BWA3106 binds.

**Figure 2.7. Chemotaxis assay.** A. CXCL12 chemotaxis assay measuring chemotactic response in the absence of CXCL12 and inhibitor (just DMSO), the presence of CXCL12 without inhibitor (SDF1), and in the presence of both CXCL12 and inhibitor at 100 µM and 200 µM concentration (ZINC310454, BWA3106-14). B. CCL2 chemotaxis assay measuring chemotactic response in the absence of CCL2 and inhibitor (just DMSO), the presence of CCL2 without inhibitor (CCL2), and in the presence of both CCL2 and BWA3111 at 100 µM concentration.

**Table 2.1. Tetrazole series based on hit compound ZINC310454.**

<table>
<thead>
<tr>
<th>ID</th>
<th>Structure</th>
<th>MW</th>
<th>K_d µM</th>
<th>LE</th>
<th>% Chemotaxis 100 µM</th>
<th>% Chemotaxis 250 µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZINC310454</td>
<td><img src="image" alt="Structure" /></td>
<td>350</td>
<td>64±23</td>
<td>0.23</td>
<td>87±16%</td>
<td>75±18%</td>
</tr>
<tr>
<td>BWA3106 (ortho)</td>
<td><img src="image" alt="Structure" /></td>
<td>203.2</td>
<td>ND</td>
<td>ND</td>
<td>85±41%</td>
<td>87±33%</td>
</tr>
<tr>
<td>Compound</td>
<td>Position</td>
<td>Molecular Structure</td>
<td>Binding Energy</td>
<td>RMSD</td>
<td>Coefficient of Variation</td>
<td>SPI</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>---------------------</td>
<td>---------------</td>
<td>------</td>
<td>-------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>BWA3107</td>
<td>ortho</td>
<td><img src="image" alt="Molecular Structure" /></td>
<td>265.27</td>
<td>13±7</td>
<td>0.33</td>
<td>100±21%</td>
</tr>
<tr>
<td>BWA3108</td>
<td>ortho</td>
<td><img src="image" alt="Molecular Structure" /></td>
<td>280.28</td>
<td>327±91.6</td>
<td>0.23</td>
<td>90±31%</td>
</tr>
<tr>
<td>BWA3109</td>
<td>para</td>
<td><img src="image" alt="Molecular Structure" /></td>
<td>203.2</td>
<td>ND</td>
<td>ND</td>
<td>110±18%</td>
</tr>
<tr>
<td>BWA3110</td>
<td>para</td>
<td><img src="image" alt="Molecular Structure" /></td>
<td>265.27</td>
<td>64±23</td>
<td>0.29</td>
<td>101±24%</td>
</tr>
<tr>
<td>BWA3111</td>
<td>para</td>
<td><img src="image" alt="Molecular Structure" /></td>
<td>280.28</td>
<td>24±28</td>
<td>0.30</td>
<td>37±16%</td>
</tr>
<tr>
<td>BWA3112</td>
<td>meta</td>
<td><img src="image" alt="Molecular Structure" /></td>
<td>203.2</td>
<td>ND</td>
<td>ND</td>
<td>102±27%</td>
</tr>
<tr>
<td>BWA3113</td>
<td>meta</td>
<td><img src="image" alt="Molecular Structure" /></td>
<td>265.27</td>
<td>50±16</td>
<td>0.29</td>
<td>105±22%</td>
</tr>
<tr>
<td>BWA3114</td>
<td>meta</td>
<td><img src="image" alt="Molecular Structure" /></td>
<td>280.28</td>
<td>126±23</td>
<td>0.25</td>
<td>107±27%</td>
</tr>
</tbody>
</table>

**Binding pose prediction of tetrazole derivative series.** Molecular docking was employed to help predict the binding pose of each of the tetrazole derivative series compounds. Guided by the chemical mapping of the NMR perturbation data, an ensemble docking approach...
was employed in predicting the poses of the ortho-substituents (BWA3106-08) in the sY12-binding site, and the para and meta-substituents (BWA3109-3114) in the sY21-binding site. The ensemble docking approach was used in order to capture as many conformations as possible, aiming to increase the chances of capturing the correct binding pose. This strategy can help account for some of the inherent challenges of docking molecules to shallow chemokine surfaces, such as protein backbone and residue flexibility. Results of the ensemble approach were visually inspected, and poses were chosen that complemented the pocket the best but that also best agreed with the NMR chemical shifts perturbations.

*Ortho-substituent tetrazole series.* The ortho-substituent tetrazole series compounds (BWA3106, BWA3107, and BWA3108) were docked into the sY12-binding site in an ensemble of 20 different NMR conformations (from PDB ID: 2K05), and one X-ray structure conformation (PDB ID: 2NWG). For the NMR ensemble, the sY12 residue from the triply sulfated CXCR4₁-₃₈ peptide was used to create the matching spheres. For the crystallographic structure, a heparin sulfate ligand, that occupies the same area as sY12 does on CXCL12, was used to create the matching spheres.

The perturbations caused by compound BWA3106 were very weak, and therefore the Kd was not determined. However, the slight chemical shifts that were observed, still placed BWA3106 in the sY12-binding site. The best docking pose placed the compound in an orientation where the tetrazole may interact with His25, while the amide may interact with Gln48. The aromatic ring does not appear to lie against any hydrophobic residues, but extends out into the solution instead (Fig. 2.8). This might explain the low affinity observed. Compound BWA3107 bound to CXCL12 with a Kd of 13±7 µM. This was the strongest binder out of the entire tetrazole series. However, BWA3107 did not inhibit chemotaxis, and also the
perturbations placed it in the sY12-binding site area as well. The docking pose of BWA3107 placed the compound in an orientation where the tetrazole is close enough to interact with His25, the O atom of the amide linker within hydrogen bond distance to Lys27, and most importantly, the terminal benzene appears to complement the hydrophobic surface created by Leu29 fairly well (Fig. 2.8). This may account for this compound’s increased affinity. The next compound from the ortho-substituent, BWA3108, bound with a Kd of 327±91.6 µM, yet the perturbations placed the compound in the sY12-binding site area as well. Our best docking pose placed the compound in an orientation where the tetrazole and the linker carbonyl are both within hydrogen bond distance to Lys27, while the terminal benzene ring extends out in solution. The proximal benzene was placed within van der Walls distance to hydrophobic residues at the dimer interface. These reduced polar interactions coupled to the benzene ring extending out of the pocket and into the solvent, may possibly account for the reduced affinity observed (Fig. 2.8). The range of affinities observed in these three compounds along with the docking pose predictions suggest that there are binding hot spots in this sY12-binding site that may be exploitable in a similar computer-aided and structure-based drug approach, a topic that will be addressed later.

*Para-substituent and meta-substituent tetrazole series.* Since for the rest of the compounds, the para-substituents (BWA3109-11) and meta-substituents (BWA3112-14), the NMR chemical shift perturbations highlighted areas associated with the sY21-binding site, docking was focused on that site solely. An ensemble approach using all 20 conformations in the NMR complex structure (PDB ID: 2K05) was used, since it was the only available structure with a ligand in the sY21-binding site.

The first compound of the para series, BWA3109, only induced weak perturbations, but ones that corresponded to the target area nonetheless. Based on the predicted docking pose, the
tetrazole group may interact with the backbone N of Ala19 and side chain of Asn22. The amide
O atom may interact with the side chain of Arg47 while the amide N atom may interact with the
side chain of Glu15. Also, the benzene ring was predicted to lie against the non-polar surface
created by Val49 and Leu42 (Fig. 2.8). The next compound in the para substituent series,
BWA3110, bound with a Kd of 64±23 µM, similar to the Kd of the original hit compound,
ZINC310454. The predicted docked position resembles that of BWA3109, but has the terminal
benzene ring extending past the cleft (Fig. 2.8). Lastly, the third compound of the para series,
BWA3111 has the best affinity at a Kd of 24±28 µM from the compounds that caused chemical
shift perturbations at the sY21-binding site area. The docking prediction placed the tetrazole
within hydrogen bonding distance to the backbone N atom of Ala19, and the side chain of Asn22
(Fig. 2.8). Even though the polar interactions predicted by the docking studies were not the
greatest for BWA3111, increased non-polar interactions were observed between both the
proximal and terminal benzene rings. Yet, the docking predictions for the para-substituents lack
to fully explain the diverse affinities observed in these compounds, since BWA3109 appears to
complement the binding site the best, but still has the lowest affinity. It is likely then that the
terminal benzene improves the non-polar interactions, while the amide and urea linkers work as
anchors, anchoring the middle section of the compound, further supporting the hydrophobic
interactions caused by the terminal benzene.

The final three compounds, the meta-substituents, also demonstrated a range of affinities.
BWA3112, which produced the smallest chemical shifts, was predicted by docking to position
the tetrazole within hydrogen bonding distance to the backbone N of Ala19, while the amide O
atom within hydrogen bonding distance to Arg47 and Asn47 (Fig. 2.8). BWA3113, bound with a
Kd of 50±16 µM, and the docking prediction placed the tetrazole within hydrogen bonding
distance to the backbone N atom of Ala19, and the side chain of Asn22. No other polar interactions were predicted by the docking studies. Both aromatic rings were predicted to compliment the pocket well, which may account for the improved affinity (Fig. 2.8). Finally, the last compound of the series, BWA3114, bound with a Kd of 126±23 µM. The docking prediction placed the tetrazole within hydrogen bond distance to the backbone N atom of Ala19, but also the side chain of Asn22. Interestingly, the urea linker was predicted to possess favorable interactions with Glu15 and Arg47, while the terminal benzene extends past the cleft (Fig. 2.8).

Discussion on docking studies. In conclusion, the docking studies using the ensemble approach have provided useful information about characteristics of both the sY12 and sY21-binding sites and the possible ways the tetrazole derivatives can bind. This information can be used to rationally design the next series of lead compounds, by preserving characteristics that were deemed important by our docking studies. Even though not a lot can be acquired from our docking studies against the sY12-binding site, with the exception of identifying a few possibly important residues (His25, Lys27), a lot can be acquired about the sY21-binding site. Based on the docking predictions, it appears that a lot of the tetrazole groups interact with the backbone N atom of Ala19, and the side chain of Asn22. It also appears that the hydrophobic patch created by Val49 and Leu42 might compliment the proximal benzene ring in many of the compounds. Finally, Glu15, Asn46 and Arn47 may interact with the linker region, stabilizing the compounds, while the terminal group may extend past the cleft. Cumulatively the results suggest a very specific orientation for the tetrazole derivative compounds, and also suggest that extending these fragments past the cleft, by adding fragments to the terminal side, may be a reasonable strategy for improving affinities.
Figure 2.8. Docking predictions of tetrazole series (BWA3106-14). Ortho substituted compounds BWA3106-08 were docked into the sY12-binding site, while the rest of the compounds (para and meta substituted) BWA3109-14 were docked into the sY21-binding site.

X-ray crystallographic studies

Crystallographic complex structures can provide detailed information on the molecular interactions between ligands and their receptors. This information is invaluable in drug optimization efforts since it can help precisely describe all the interactions that mediate binding, but also help with subsequent optimization efforts. For the sY21-binding site target, a complex
structure could greatly help lead the design of a new series of tetrazole derivatives that would target the sY21-binding site, and extend past the site, further down the cleft, and towards the sY12-binding site.

CXCL12 had been crystallized previously. In the Protein Data Bank, there are three different crystal forms of CXCL12. The first form belongs to the P3₂12₁ space group with one monomer in each asymmetric unit and unit cell dimensions a = 55.5, b = 55.5, and c = 46.0 (eg PDB ID: 3GV3). This space group however would not be conducive to our complex structure determination efforts since side chains from crystal packing interface residues occupy the sY21-binding site and would make incorporation of compounds to that site practically impossible. The other space group in the PDB is a pP2₁2₁₂₁ with a dimer in each asymmetric unit. The unit cell dimensions are a = 36.5, b = 57.0, and c = 71.8 (PDB ID: 2J7Z). In this crystal form, the sY21-binding site is more accessible, but still partly inhibited by side chains from crystal packing residues. This may limit incorporation of bulkier compounds. The final space group in the PDB is pP2₁2₁₂₁ with 10 monomers per asymmetric unit, and unit cell dimensions a = 41.8, b = 117.5 and c = 134.5 (eg PDB ID: 3HP3). This crystal form would be the ideal form for our complex study purposes since the sY21-binding site is completely unconstrained and easily accessible.

Crystallization trials. After purifying the protein to over 95% purity, attempts to reproduce the most desirable crystal form followed, but those attempts were largely unsuccessful. In fact, attempts to reproduce any of the published conditions for any of the crystal forms were also unsuccessful. Consequently, crystal screen trials were performed in which over 1000 different conditions were tested.

The first crystal hit was observed in an acidic ammonium sulfate based crystallization condition (2.0M Ammonium Sulfate, 0.1M Sodium Acetate Trihydrate pH4.6). The crystals
appeared “needle”-like and after multiple rounds of optimization these crystals failed to reach diffractable quality. Next, crystals were observed in a basic condition (2.0M Ammonium Sulfate, 0.2M Lithium Sulfate, and 0.1M CAPS pH10.5), but these crystals were found to belong to the undesirable P3$_2$1 space group (PDB ID: 3GV3). Yet, soaking trials were performed, but as expected no compound was observed in the sY21-binding site. Finally, small crystals were observed in a condition (3.5M Ammonium Sulfate, 1%MPD, 0.1M MES Sodium Salt pH6.5), and the condition was then optimized (2.2M Ammonium Sulfate, 2%MPD, 0.1M MES Sodium Salt pH6.5) to consistently produce large single crystals. Surprisingly, crystals in this condition belonged to two different forms; the undesirable P3$_2$1 space group (e.g. PDB ID: 3GVR), but also the P$_2_1$2$_1$2$_1$ space group (e.g. PDB ID: 2J7Z). Subsequent crystallization trials through the use of seeding allowed to consistently reproduce the P$_2_1$2$_1$2$_1$ space group (e.g. PDB ID: 2J7Z) crystals, and achieve a diffraction up to 1.8 Angstrom resolution.

Next, compound ZINC310454 and all nine compounds from the tetrazole derivative series (BWA3106-3114) were all tested for complex crystal formation. Initial trials were performed by attempting to grow crystals in the presence of dissolved compounds, but crystals grew only in the presence of BWA3107 and BWA3108. Unfortunately, for those crystals, no density was observed in either the sY12 or the sY21-binding site. The rest of the compounds were then tested by soaking methods. Crystals were transferred into a crystallization solution that contained the dissolved compounds at 20mM concentration. However, density was only observed for BWA3111, and luckily the density was observed in the sY21-binding site.

**CXCL12-BWA3111 complex structure.** Since the quality of the electron density was ambiguous, we optimized the soaking method. We noticed that DMSO can non-specifically bind to the CXCL12 surfaces of interest and result in a noisy electron density map. As a result, no
DMSO was used in subsequent soaking experiments and the compound was dissolved to saturation directly in the stabilization buffer. The resulting complex structure was solved at 1.9 Å resolution, and has a primitive orthorhombic space group (P2₁2₁2₁) (Table 2.2) with a dimer in each asymmetric unit (Fig. 2.9). The overall structure of the CXCL12 dimer is nearly identical to that of the apo protein (PDB ID: 2J7Z) with an RMSD of 0.238 Å aligning 774 atoms. The only significant backbone movements between the complex structure and the previously determined apo structure are observed in the flexible N-terminus of each monomer.

Figure 2.9. Complex crystallographic structure of CXCL12 dimer. CXCL12 dimer (ribbon and surface model, blue and green) with compound BWA3111 (yellow) bound to the sY21-binding site.

Structural analysis. The unbiased Fo-Fc density unambiguously identifies the binding mode of compound BWA3111 (Fig. 2.10A), residing in the pocket that normally interacts with sY21 and adjacent residues, as seen in the NMR complex structure of CXCL12 bound by the CXCR4 sulfated N-terminal peptide (Fig. 2.10C). Compound BWA3111 was observed only in one of the two monomers in the asymmetric unit, because the corresponding binding pocket in the other monomer was partially blocked by crystal packing.

A number of protein main chain and side chain functional groups are involved in polar and non-polar interactions with the small molecule ligand. Ala19N and Asn22Nδ2 are in close
contact (~3.0 Å) to the tetrazole group, with the former establishing a favorable hydrogen bond and the latter an NH-π interaction (Fig. 2.10B). The urea linker is within hydrogen-bonding distance to Glu15, Asn45 and Arg47. Interestingly, compared to the apo structure (PDB ID: 2J7Z), Glu15 side chain swings into the sY21-binding site, and forms an ideal hydrogen-bonding network with compound BWA3111, with the two nitrogen atoms of the urea linker serving as hydrogen donors to the two oxygen atoms of the glutamate side chain (Fig. 2.10B). Furthermore, compared to the published apo structure, Arg47 undergoes conformational changes in order to form a hydrogen bond with the carbonyl oxygen of the urea linker while partially stacking against the terminal benzene ring of compound BWA3111 (Fig. 2.10B). The backbone atoms of residues 43-45 also move closer into the sY21-binding pocket by 0.5-0.8 Å, allowing a closer contact between Asn45 and the urea linker.

In addition to polar interactions, compound BWA3111 establishes extensive hydrophobic interactions with Val18, Leu42 and Val49 (Fig. 2.10B). The proximal benzene ring nestles in the hydrophobic pocket formed by Val18 and Leu42, making many non-polar contacts. The distal benzene ring is also within van der Waals contact distance with Val49δ1, although the interactions with Val49 can be further optimized. In fact, most of the terminal benzene ring seems to suspend in solution past the pocket, suggesting that this part of the ligand can indeed be further optimized to enhance binding to the chemokine (Fig. 2.18D). It also highlights the rigidity and superiority of the urea linker, which is further immobilized by the multiple hydrogen bonds as described above. These observations suggest additional carbon atoms may be introduced into the linker to allow adequate flexibility in future lead compounds.

As the sY21-binding site also resides close to the crystal-packing interface, the binding of compound BWA3111 has caused additional conformational changes in residues from
neighboring molecules in the crystal lattice. An alternative conformation is observed for His17 from a symmetry-related molecule, in order to relieve a possible steric clash with the tetrazole group by the original conformation. Arg8 from another adjacent molecule also becomes less ordered. We attempted to investigate whether crystal packing may influence the binding pose of compound BWA3111 by using crystals of R8A, H17A and R8A/H17A mutants for the soaking experiments. Unfortunately, compound BWA3111 was not observed in these crystals, possibly because the positive charges of these two residues can non-specifically increase the chances of incorporating the negatively charged compound BWA3111 into the binding pocket, an unintended but useful effect considering the challenges faced in crystallizing complexes with weak ligands. It should also be noted that His17 in monomer 1 is part of the sY21-binding site and may directly contribute to ligand binding through electrostatic interactions.

*Comparison with sY21 binding in CXCR4 interaction.* The binding of compound BWA3111 and the CXCR4_{1-38} peptide to the same sY21-binding site (PDB ID: 2K05) (Fig. 2.10B-C) share many similarities yet demonstrate key differences. Both compound BWA3111 and sY21 are negatively charged at physiological pH and interact with many of the same protein residues in a series of hydrogen bonds, including Glu15, Asn22 and Arg47. These interactions highlight the overall positively charged and highly polar environment of the binding site. However, the sulfate group of sY21 and the tetrazole group of compound BWA3111 reside in different areas of the binding pocket, with the former hydrogen bonding with Asn22, Asn46 and Arg47, and the latter with Asn22 and the backbone N atom of Ala19. The conformations of Glu15 and Arg47 also differ in the two complexes.

The most significant difference, however, lies in the more extensive hydrophobic interactions observed in the complex with compound BWA3111. The aromatic ring of sY21 is
placed perpendicularly and relatively superficially on the protein surface in a corner of the pocket formed by Val18, Leu42 and Val49, while the benzene ring of compound BWA3111 lays flat on the center of this small hydrophobic cavity. This increases hydrophobic contact in compound BWA3111 binding.

Figure 2.10. Complex crystallographic structure and characterization of the sY21 binding site. A) Stereo-image depiction of the unbiased Fo-Fc map of compound BWA3111 bound to the sY21-binding site of CXCL12. B) Complex crystal structure of CXCL12 (green) bound by compound BWA3111 (yellow) superimposed to the apo crystal structure (PDB ID: 2J7Z) (syan) shows conformational changes induced upon binding. C) NMR complex structure of CXCL12 bound to the D20-sY21-D22 segment of CXCR41-38 (PDB ID: 2K05) outlines the sY21-binding site as well as the possible hydrogen-bond interactions.
Some of the differences we have observed between the two complexes undoubtedly come from the structural ambiguity in NMR structures, particularly in terms of side chain conformations, as well as from the structural biases imposed by crystal packing, especially the lack of protein flexibility. However, although CXCL12 exhibits much flexibility in solution, its core domain including the majority of the sY21-binding site displays much less variation among different NMR conformations as well as various crystal structures. This is particularly true for the hydrophobic cavity and the peptide segment spanning His17 to Asn22. The most flexible region appears to come from the three asparagine residues 44-46. Asn45 or Asn46 each makes one hydrogen bond with the ligand in the respective complex. Although certain interactions involving these residues may have been missed in the NMR or crystal structures due to the limitations of each technique, it is unlikely that such experimental caveats have drastically distorted the ligand-binding mode.

Insights into structure-based inhibitor design. Even though we had thus far demonstrated that compound BWA3111 is a fragment that binds CXCL12, and that NMR perturbation studies place the compound in the sY21-binding site, we still depended on our docking studies to predict the binding pose. Molecular docking successfully predicted the correct orientation of the compound and assisted us with previous optimization efforts (data not shown). Docking however was unable to identify all the possible hydrogen bonds since it did not predict the protein conformational changes induced upon binding, such as changes in Glu15 and Arg47. The complex crystal structure therefore, has not only confirmed binding, but has shed light on crucial interactions for the first time, previously overlooked through other complimentary methods, such as NMR and molecular docking. This information is important for structure-based inhibitor design. It highlights the urea linker as an optimal anchor in hydrogen bonding interactions; the
rotation of Glu15, along with the arrangement of Asn45 and Arg47, creates an ideal hydrogen-bonding network with the urea linker of compound BWA3111. This also suggests that amide linkers are not as efficient as urea linkers since an amide linker can only form one hydrogen bond with Glu15 compared to the two hydrogen bonds that the urea linker is forming in the crystal structure. This difference has been exemplified through our docking studies that appropriately position amide-linker derivatives in the sY21-binding site, showing that only one hydrogen bond can form between the amide and Glu15.

The crystal structure not only identifies the important interactions between compound BWA3111 and CXCL12, but also elucidates unoccupied potential hot spots for ligand binding. The terminal benzene extends into a cleft that contains both polar and non-polar residues, characteristics that can be exploited in optimization efforts (Fig. 2.18D). Particularly, this cleft contains several non-polar residues including Pro10, Leu29 and Val39. It resides on the edge of the heparin-binding site and interacts with sY12 in the complex structure with the CXCR41-38 peptide (Fig. 2.18C). Whereas heparin fails to fully utilize the non-polar binding surface, more hydrophobic interactions are observed between sY12 and these protein side chains. In addition, in order to leverage these binding hot spots in our ongoing project of inhibitor design efforts, extra carbon atoms can be added after the urea linker of the current inhibitor scaffold for increased flexibility.

Although the sY12-binding site appears to be a more challenging target by itself, these observations suggest that small molecule ligands can be designed to span from the sY21-binding site into the sY12-binding site or even further into the heparin binding site. This would further improve the affinity and specificity of the novel ligand, increasing its utility in disrupting the CXCL12-CXCR4 signaling axis.
Computer-aided, structure-based optimization based on BWA3111

C-series. The complex structure of BWA3111 bound to CXCL12 identified key conformational changes that were not taken into consideration when previous computer aided optimization efforts were underway. The positioning of Glu15 to accommodate the urea linker along with the positions of Arg47 and Asn45 provide an ideal hydrogen-bonding network that can be utilized as a secondary anchor point, in conjunction to the tetrazole that together lock such a compound into position. It also suggests that the terminal benzene can be exploited in optimizations efforts. Therefore, in order to explore characteristics of the binding cleft past the sY21-binding site, where the terminal benzene of BWA3111 resides, we designed the C-series of compounds based on lead compound BWA3111; these compounds retained the BWA3111 scaffolds but differed through the addition of groups onto the terminal benzene. Then, through the help of docking, we prioritized compounds for synthesis that were more likely to fit in the pocket (data not shown) (Fig. 2.11). Finally, 17 compounds were synthesized to test binding through NMR (Fig. 2.12, C14-43).

![Figure 2.11. C-series hit-to-lead optimization design scheme based on BWA3111 scaffold. The BWA3111 scaffold was retained, and a variety of functional groups were added to the terminal benzene in order to increase apolar contacts.](image)
Design of C-series. From the 17 compounds that were synthesized, 10 compounds (C14, C17, C21, C23, C24, C25, C28, C29, C33, C34) retained the tetrazole in the *para* position, while the other 7 compounds (C22, C37, C39, C40, C41, C42, C43) retained the tetrazole in the *meta* position. Additional groups on the terminal benzene were added (Fig. 2.11). The compounds were subsequently tested for binding through NMR. Binding was compared to a DMSO control. In addition, diphenylurea was also used to test the efficiency of the urea linker in the absence of the tetrazole group. Out of the 17 compounds tested, 13 showed meaningful affinities that were then assigned by chemical shift mapping to residues in the sY21-binding site area (Fig. 2.12).

2D NMR testing. In addition to BWA3111 (24.0±28 µM), the compounds out of the C-series that showed the best affinities were C14 (26.7±9.6 µM), C20 (14.1±46 µM), C23 (29.9±19 µM), C33 (19.0±15 µM), C34 (26.4±24 µM), C37 (26.0±30 µM), and C42 (21.3±8.3 µM). Next, with slightly less affinity were compounds C39 (60.4±30 µM), C41 (97.6±108 µM), C43 (79.0±41 µM), then compounds C25 (171±147 µM), and C22 (304±320 µM). The remaining compounds from the C-series (C17, C24, C28, C29, C40) showed insignificant affinities and linear titration curves suggesting unspecific binding. In addition, diphenylurea showed very low affinity, further supporting the significance of the tetrazole in anchoring the compounds (Fig. 2.12).

Docking studies. In order to help explain the differences in affinities between the compounds in the C-series, stringent docking studies were applied using the CXCL12/BWA3111 X-ray complex structure as a template. The compound BWA3111 was used to generate the matching spheres, in order to bias the positioning of the docked compounds in adopting a pose similar to the BWA3111 compound as seen in the X-ray structure.
Figure 2.12. NMR titration assays showing binding of C-series compounds. DMSO titrations were used as a control to identify perturbations attributed to DMSO. Diphenylurea was titrated to test scaffold in absence of tetrazole. The C-series of compounds based on BWA3111 were screened through 2D NMR and residues perturbed were assigned.

Based on docking, compound C14 is predicted to bind very similarly to how compound BWA3111 binds. The chlorine atom added at the 3’ carbon on the terminal benzene appears to simply suspend in solution, providing no advantage or disadvantage to binding of the compound, as also evidenced by a similar binding affinity (26.7±9.6 µM) to BWA3111. The next compound in the series, C17, is predicted by docking to adopt an alternative conformation failing to utilize the urea linker efficiently. This might help explain the loss in binding seen by the NMR titration studies. C20 is predicted to adopt a conformation like BWA3111, and this is supported by its similar binding affinity (14.1±46 µM). C22 has a reduced affinity (304±320 µM), which might
be explained by its docking pose that positions the compound similarly to BWA3111, except that since it has the tetrazole in the \textit{meta} instead of the \textit{para} position, the proximal benzene ring is shifted slightly away from the hydrophobic patch, thus reducing hydrophobic contacts. C23 has similar affinity to BWA3111 (29.9±19 µM) and is predicted by docking to bind similarly to BWA3111 as well. The naphthalene ring, which replaces the benzene, is predicted to lie against the hydrophobic cleft, stabilizing the increased hydrophobic character of the compound. C24 is predicted to adopt a similar conformation to C17, which helps explain the loss of affinity through failing to utilize the urea linker efficiently. C25 is predicted to bind similarly to BWA3111, but this fails to explain its reduced affinity (171±147 µM), especially considering how similar these two compounds are. C28 loss of affinity might be explained by the unsatisfied interactions imposed by the added F₃C group. The same might apply for C29, which has the addition of a trimethyl group that doesn’t satisfy any hydrophobic interactions as displayed in the docking prediction. C33 on the other hand, may increase the hydrophobic interactions by the addition of one methyl group, which may interact with the additional hydrophobic surface in the cleft, as is supported by its binding affinity C33 (19.0±15 µM) and docking prediction. C34 is predicted to bind similarly to C33 and has similar affinity (26.4±24 µM). For C37, the docking prediction places the compound similarly to C33 and C34, but places the NO₂ group within hydrogen bond distance to Glu15. An increase in affinity however is not observed (26.0±30 µM); therefore it is unlikely that the NO₂ group interacts with Glu15. C39 has comparable affinity (60.4±30 µM), and even though is has a tetrazole in the \textit{meta} position, its affinity can be explained by re-orientation of the CF₃ group into the cleft due to the tilt caused by the \textit{meta}-substituted tetrazole group. C40 looses affinity but this can be explained by the docking prediction that positions the compound where the trimethyl group does not interact with a hydrophobic surface. C41 displays
decent affinity (97.6±108 µM) and this can be due to repositioning of the naphthalene towards the hydrophobic cleft due to the meta-substituted tetrazole group on the proximal benzene, as suggested by the docking prediction. Docking shows that compound C42 may retain good affinity (21.3±8.3 µM), despite the meta-substituted tetrazole induced slight shifting of the proximal benzene off the ideal flat position, by placing the proximal nitrogen within polar contact distance to Arg47. Finally, based on the docking prediction of compound C43 (79.0±41 µM), its affinity may be retained by shifting of the added terminal methyl group into the hydrophobic patch, caused by the global repositioning due to the meta-substituted tetrazole group (Fig. 2.13).

Figure 2.13. Docking predictions for C-series compounds based on crystallographic CXCL12/BWA3111 complex structure. Docking predictions of the C-series compounds tested through 2D NMR help explain affinities observed.
Discussion. The NMR results in combination to the docking predictions support the superiority of the tetrazole group in the para-substituted position, allowing the proximal benzene to lie flat against the hydrophobic surface presented by Val49. Loss of hydrophobic contacts, and therefore affinity, caused by shifting the tetrazole to the meta position can be rescued by addition of groups in the terminal side that may increase polar or non-polar contacts. It also appears that addition of groups in the 2’ position of the terminal benzene of para-substituted compounds may help increase affinity, but also help extend the compound towards the sY12-binding site, such as in the case of C33 and C44. However, compound C33 has proven to be much more soluble than any of its analogs, and has produced consistent affinities in multiple NMR experiments (data not shown). In addition, C33 has shown encouraging results in ongoing chemotaxis assays, and has therefore become our lead compound of interest. Unfortunately, crystallization attempts to complex C33 with CXCL12, either by co-crystallization or soaking methods, have been unsuccessful. This is probably due to the narrow space around the sY21-binding site in our crystal form, which is obstructed by the crystal-packing interface disallowing incorporation of bulkier molecules.

YL-series. Since we haven’t been able to significantly improve affinities for our tetrazole-based compounds, it may be possible that we have already exploited all the binding hot spots that can be exploited by such a pharmacophore. However, we had not yet attempted to extend compounds past the cleft branching out towards the adjacent sY12-binding site (Fig. 2.2). We therefore designed a series of compounds that included extended linkers in an attempt to extent compounds around the cleft and towards the sY12-binding site (Fig. 2.14).
**Design of YL-series.** Compounds were prioritized for synthesis through the help of docking studies (data not shown) and 10 compounds were finally synthesized and tested for binding through NMR (YL11145-66). These compounds were based on the BWA3111 scaffold by retaining the tetrazole in the para position, but differed through the insertion of additional atoms between the linker and the terminal benzene (Fig. 2.14). The purpose of these compounds was to test whether some analogs can wrap around the cleft while identifying and exploiting unutilized binding hot spots that lied further away from the sY21-binding site.

From the compounds synthesized, YL11145 has an additional chain of four carbons separating the terminal benzene from the urea linker, YL11147 has one additional carbon, YL11148 has two, and YL11149 has three. YL11163 has an additional chain of three carbons, but also has a chlorine atom at position 4’ of the terminal benzene. The rest of the series was designed with the purpose of identifying and utilizing any potential unutilized polar binding hot spots. Therefore, YL11156 has the addition of a chain of four carbons, but is also followed by an amide linker connecting the terminal benzene. YL11159 has the addition of a chain of two carbons followed by an amide group and benzene, and YL11162 has a chain of three carbons and
an amide. YL11165 has a chain of three carbons that also has a secondary one-carbon-chain ending in a carboxyl branching out of the first carbon. Finally, YL11166 has the addition of a chain of two carbons with a carboxyl group branching out of the second carbon (Fig. 2.15).

2D NMR testing. The 10 compounds were tested for binding through 2D NMR. Even though all of the compounds caused $^1$H-$^{15}$N chemical shift perturbations when titrated into the dimeric CXCL12 solution, only two showed saturation, suggesting specific binding. These were compounds YL11145 (Kd = 134±117 µM) and YL11163 (Kd = 876±217 µM) (Fig. 2.15).

Figure 2.15. NMR titration assays showing binding of YL-series compounds. The YL-series of compounds based on BWA3111 were screened through 2D NMR and residues perturbed were assigned. Affinities for only YL11145 and YL11163 were determined.
Docking studies. In order to investigate why some of the compounds bound while others did not, docking was used to help predict the binding poses. However, an ensemble docking approach was employed in order to increase the number of observable poses, in hope of identifying the most probable ones. Poses produced by docking were visually inspected, and the poses that appeared to complement the pocket the best were chosen as the most plausible (Fig. 2.16).

Compound YL11145, which has the highest affinity from the YL series (134±117 µM), appears to bind quite favorably to the sY21-binding site, as predicted by docking (Fig. 2.16). While the core compound binds similarly to BWA3111, it appears that the four-carbon long chain is the ideal length that allows the chain to efficiently wrap around the cleft and position the benzene appropriately against the hydrophobic area that lies between the sY21 and sY12-binding sites (Fig. 2.16). Compound YL11163 bound with a much lower affinity (876±217 µM). Docking placed the compound similarly to BWA3111 in the crystal structure, with the terminal benzene ring efficiently flipped and wrapping around the cleft (Fig. 2.16). However it appears, that the four-carbon chain wraps the cleft more efficiently than the two-carbon chain, as suggested by both the NMR and docking studies.

The rest of the compounds did not appear to bind specifically to the sY21-binding site, and some of those results can be discussed by the predicted docking poses (Fig. 2.16). YL11147, YL11148, and YL11165 are predicted to bind in a way that exposes the positions the terminal benzene out into the solution. Also, based on the docking poses, YL11149, YL11159, YL11162, and YL11166 do not appear to utilize the same efficient hydrogen bond network as the urea linker of BWA3111. This may explain why those compounds lost affinity. Finally, YL11156, which did not show specific binding either, appears in the docking prediction to adopt a rather
favorable conformation, with the terminal amide group within hydrogen bond distance to Arg47.
It is however likely that the amide group is not interacting with the flexible Arg47, but rather
interacting unfavorably with the apolar surface (Fig. 2.16).

Figure 2.16. Docking predictions for YL-series compounds using an ensemble approach. Docking predictions
of the YL-series compounds tested through 2D NMR help explain binding or lack of binding.
Discussion. The results from our hit-to-lead series so far suggest that the tetrazole in the para position with the urea as the linker is the best scaffold we have that can exploit both the polar and apolar nature of the sY21-binding site. Yet, extending compounds past the cleft and towards the sY12-binding site has proven to be challenging, mostly since we are unable to obtain complex crystals of any other compounds besides BWA3111, due to the nature of the sY21-binding site in our crystal form, which is obscured largely by the crystal packing interface. Docking studies however have provided useful information, which may shed some light on the differences in affinities seen in both the C-series and YL-series compounds. The docking and NMR results from testing these compounds suggest that hydrophobic patches in the cleft past the sY21-binding site can be exploited in increasing apolar interactions. However affinity is compromised as the length of the compound increases, in attempting to extend the compound towards the sY12-binding site. It is possible though that adding polar groups past the terminal benzene in the YL11145 scaffold may allow for polar interactions to take place, which may stabilize the compound and increase affinity.

Virtual screening against the sY21-binding site through a flexible approach

We decided to perform a complementary virtual screen trial in search of new hit compounds that may potentially have similar or better affinities but that could potentially lead an alternative hit-to-lead optimization scheme. However, our structure-based hit-to-lead approach, starting from ZINC310454 and leading to the tetrazole series, had proven to be quite challenging for the additional reason that the inherent flexibility of the sY21-binding site had created challenges in predicting the correct ligand poses when using docking. Therefore, in order to
partly account for some of the challenges faced from docking into shallow flexible sites, a flexible docking approach was employed to the complementary virtual screen.

For the flexible docking approach, the 1st NMR structure from the NMR ensemble (PDB ID: 2K05) was chosen as a template, but selected residues were manipulated; six different conformations were chosen for Arg47, and two conformations were chosen for Glu15. This resulted in 12 distinct binding pocket compositions. Sequential virtual screens of the lead-like subset of the ZINC database were then employed for each conformation. The results were combined; the compounds were matched to their corresponding protein conformation and then sorted by score. The top 2000 compounds were visually inspected and 10 compounds were selected for purchase and tested by 2D NMR. Out of the 10 compounds, compound ZINC816454 was the only compound to induce chemical shift perturbations. The Kd was determined at ~150 µM. Even though the binding may not be as strong as the original hit ZINC310454, ZINC816454 has a very different pharmacophore and also induced significant chemical shifts that were mapped in the helix region, suggesting a different mode of binding, as compared to the tetrazole series (Fig. 2.17). ZINC816454 therefore has the potential to lead a new series of compounds once SAR analysis sheds some light on the significant factional groups.

**Targeting the sY12-binding site on CXCL12**

The sY12-binding site on CXCL12 lays adjacent to the sY21-binding site in an area that overlaps with a larger pocket formed at the dimer interface. The two sulfotyrosine-binding sites are separated by a distance of <20 Angstroms (Fig. 2.18D). Sulftotyrosine binding to the sY12-binding site increases affinity for the CXCR4 N-terminus, but not as much as sulfation of Y21. Yet, it still makes an appealing target since structural information of that site suggests that
compounds may be designed to utilize some of the same interactions observed between CXCL12 and the sY12-binding site ligands.

Figure 2.17. ZINC816454 2D NMR studies shows binding to CXCL12. A. Compound ZINC816454 from the ZINC database chosen from a flexible docking virtual screen. B. Chemical shift perturbations caused by titration of ZINC816454. C. Chemical shift perturbations were assigned on residues of CXCL12. D. Docking pose prediction of ZINC816454 and mapping of residues (light blue) show areas affected by binding.

Currently in the PDB there are two complex structures that possess ligands in the sY12-binding site; one is the NMR CXCL12-CXCR4_{1-38} complex (PDB ID: 2K05), and the other is an X-ray crystallographic CXCL12-heparin sulfate complex structure (PDB ID: 2NWG). The NMR structure (PDB ID: 2K05), places the sY12 in that binding site (Fig. 2.18B), while the X-ray structure (PDB ID: 2NWG) places a heparin sulfate in that same site (Fig. 2.18C). Superimposition of both structures shows how both ligands occupy the same general area (Fig.
2.18A, D). Close examination of the ligand-protein interactions in both structures shows how some of the polar interactions with the sY12-binding site are the same between the sulfate groups of sY12 and heparin, such as in the case of Gln48, Arg41, and Lys27. These interactions can be emphasized when applying a virtual screening approach, so as to select ligands that may bind in similar ways to that pocket. Finally, since the known ligands for both sY-binding sites are about 13 Angstroms apart (Fig. 2.18D), successfully identifying novel ligands against the sY12-binding site can pave the way for an alternative strategy, where a fragment based approach can help extend compounds towards the direction of the sY21-binding site. Compounds extending towards each other from these two different sY-binding sites can be linked together via linkers, creating truly potent CXCL12 inhibitors.

**Virtual screening against the CXCL12 sY12-binding site.** The complex crystallographic structure of CXCL12 with the unsaturated heparin disaccharide, revealed structural details of the interactions between glycosaminoglycans and CXCL12. Two copies of the disaccharide were observed in the crystal structure. Only one of them, residing at the dimer interface, was deemed biologically relevant whereas the other one was most likely a crystallization artifact. The heparin-binding site at the dimer interface is characterized by a high concentration of positively charged residues, including His25, Lys27 and Arg41 from both CXCL12 monomers (Fig. 2.18C). It is not surprising that the disaccharide, with four negatively charged functional groups, binds to this area. Most of the intermolecular interactions are mediated through a series of hydrogen bonds involving the charged moieties. Only one non-polar contact is observed between the disaccharide and a hydrophobic residue, Leu29 (Fig. 2.18C).
Figure 2.18. sY12-binding site on CXCL12. A. Superimposing of two complex structures show how sY12 (yellow) and heparin sulfate (pink) occupy the same space on CXCL12. B. NMR complex CXCL12/CXCR41-18 structure (PDB ID: 2K05) reveals detailed information of sY12 binding to CXCL12. C. CXCL12-heparin sulfate crystal structure (PDB ID: 2NWG) positions heparin sulfate (yellow) in the region where sY12 of CXCR41-38 normally binds. D. CXCL12-compound BWA3111 crystal structure with heparin sulfate (yellow) and sY12 (purple) superimposed in sY12 site suggests that compounds specific to both sites could potentially be linked together via linkers.

Since the crystallographic structural information of the sY12-biding site binding to heparin sulfate was available, and since a lot of the interactions observed between heparin sulfate and sY12 are the same, preliminary virtual screening experiments were performed against the X-ray structure (PDB ID: 2NWG) only. An additional rational for the strategy applied was that the conformation observed in the X-ray structure can be easily reproduced in our crystallographic trials since we have been able to re-create this crystal form previously. Therefore, discovering compounds that bind to this conformation may be a better strategy for complex structure
determination and subsequent structure-based hit-to-lead optimization. Furthermore, since more interactions were identified between heparin sulfate and CXCL12, than with sY12 and CXCL12, information from these interactions may provide a good starting point for our docking studies.

Consequently, the X-ray structure in the dimer form was used as the receptor template and heparin sulfate was used to create the matching spheres. Both the fragment and lead like subsets of the ZINC database were screened, and the top 1000 compounds from each subset were visually inspected. After the first round of docking studies however, it was observed that very few compounds supporting hydrophobic interactions scored high on the list. As a result, another round of virtual screening was performed, but this time matching spheres were created by a composite of docked compounds; three favorable docked compounds from the previous virtual screens were chosen, which complimented different areas of the surface though both polar and apolar interactions. These where then combined into one file, and the unfavorable or unnecessary parts were trimmed off. The matching spheres were subsequently created by this sum of compounds, and virtual screens were repeated as previously. This time, more desirable compounds scored favorably as expected. 30 compounds were chosen as the best candidates and out of these compounds, eight were purchased to be tested through NMR (Fig. 2.19).

2D NMR studies. Out of the eight compounds purchased and tested, three induced chemical shift perturbations (ZINC238469, ZINC94774, and ZINC2497769). ZINC2497769 however induced the largest chemical shift perturbations. Consequently, this new potential hit became the focus of our drug discovery efforts against the sY12-binding site (Fig. 2.19).

The Kd for ZINC2497769 was determined at ~170 µM. Titrations produced a subset of peaks that led to defining signatures for residues V23, K24, I28, V39, A40 Q48, and V49. These residues cumulatively define the sY12-binding site pocket. Combined chemical shift mapping of
both amide-proton and carbon-proton for V39 and Q48 showed that there were significant chemical shifts for both. These results suggest that ZINC2497769 binds in a pose similarly to what was predicted by docking (Fig. 2.20A). Similar titration studies were performed on CXCL12₂, a constant dimeric form construct of CXCL12, in order to study binding at the dimer interface. Since the sY12-binding site lies at the dimer interface, it was of interest to see whether one ZINC2497769 molecule interacts with just one or both monomers. The largest (blue) and intermediate (orange) chemical shift changes were mapped into a larger area highlighting respective sites in both subunits, and the Kd was slightly improved (~150 µM), suggesting that there were additional contacts made between ZINC2497769 and residues from the neighboring CXCL12 subunit (Fig. 2.20B).

Figure 2.19. 2D NMR screening for sY12-binding site binders. Eight compounds purchased from the ZINC database were tested through 2D NMR and ZINC2497769 induced significant chemical shift perturbations in an area mapping the sY12-binding site.
In order to proceed with optimization efforts on ZINC2497769, we applied a SAR analysis approach; in order to investigate the importance of the sulfo-carboxylate arms of the pyrimidine, we performed a restrictive subset search of the ZINC database that identified a homolog compound, ZINC11768885, which lacks one of the sulfo-carboxylate arms. This compound was purchased and tested via 2D NMR and was found to also bind to CXCL12 but with a reduced affinity (~ 800 µM). Nonetheless, the residues perturbed were the ones highlighting the sY12-binding site, suggesting binding to that site, but also highlighting the importance of the sulfo-carboxylate arm in binding affinity (Fig. 2.20C).

We wanted to further investigate the localization and binding of ZINC2497769 to CXCL12. Since heparin binds to the sY12-binding site as well, a heparin binding competition assay was performed to study competition between ZINC2497769 and heparin. ZINC2497769 was indeed found to compete with heparin binding with an IC$_{50}$ of ~1 mM and a K$_i$ of ~ 950 µM (data not shown). Lastly, a chemotaxis assay was performed to study inhibition, but ZINC2497769 did not inhibit chemotaxis at ~500 µM or below, suggesting that CXCR4 may still wrap around the chemokine (data not shown). This further supports our previous reasoning that sY12-binding site binders may work best as extensions of sY21-binding site inhibitors, rather than inhibitors themselves. But for compounds against both sites to be linked together, simultaneous binding against both the sY12 and sY21-binding sites by their respective ligands should first be confirmed. Therefore, in order to investigate whether compounds against both binding sites bind simultaneously, NMR titrations were repeated, first with original ZINC310454 hit compound, which binds to the sY21-binding site, and then with our new ZINC2497769 hit compound, which binds to the sY12-binding site (Fig. 2.21A,B). Then titrations were performed with both compounds together (Fig. 2.21C). The chemical shifts induced were mapped onto the
protein, demonstrating the co-titration of both compounds together produced peaks representative to each compound’s respective binding features, indicating that both compounds bound to their respective binding sites, non-competitively. These results suggest that indeed our fragments show high promise in eventually being linked together, producing highly potent compounds.

Figure 2.20. NMR screening of compounds against sY12-binding site, mapping of residues and docking poses. A. ZINC2497769 tested with monomeric CXCL12 perturbed residues around the sY12-binding site. B. ZINC2497769 tested with dimeric CXCL12 perturbed residues around the sY12-binding site. C. ZINC11768885 tested with monomeric CXCL12 perturbed residues around the sY12-binding site.
Virtual screening using ensemble docking. Finally, we wanted to perform additional virtual screening experiments in search of binders that either have a different pharmacophore, or conversely that would bind with higher affinity. In order to sample a larger chemical space we employed an ensemble docking approach, rather than docking into one rigid conformation. In this approach we screened the lead-like and fragment subsets of the ZINC database against our own database of 21 different CXCL12 conformations (20 NMR conformations from PDB ID: 2K05, and one X-ray conformation from PDB ID: 2NWG) (Fig. 2.22A). We combined all the results together, ranked the compounds according to binding score, and sorted the compound...
database by matching each compound pose to its corresponding protein conformations. Redundant compounds were removed.

2D NMR studies. The top 1000 compounds were visually inspected and 23 compounds were prioritized for testing. Out of these 23 compounds tested by 2D NMR, only ZINC00095441 was found to bind (Fig. 2.22B,C). Surprisingly this compound was paired to the crystallographic conformation from the ensemble, suggesting that the conformation from the X-ray structure is the best candidate for virtual screening experiments. Furthermore, ZINC00095441 bound with an affinity comparable to compound ZINC2497769 (~ 200 µM). This compound also shares a similar pharmacophore to ZINC2497769, including the extended ring system and two arms with a terminal carboxylate (Fig. 2.22B). More tests will are needed to deduce which out of the two hits is a better candidate for our optimization efforts.

Figure 2.22. Identification of ZINC00095441 using ensemble virtual screening. A. Representation of sY12-binding site ensemble containing 20 CXCL12-sY12 (purple) and one CXCL12-heparin sulfate (yellow) conformations. B. ZINC00095441 has a similar pharmacophore as ZINC2497769. C. 2D NMR chemical shift signature for ZINC00095441 is similar to ZINC2497769 and has a comparative Kd of 200 µM.
Targeting the sY7-binding site on CXCL12

The sY7-binding site on CXCL12 lies at a dimer interface, on the other side of where sY21 binds (Fig. 2.23A). In the dimeric CXCL12 NMR structure complexed to CXCR4_{1-38} (PDB ID: 2K05), sY7 binds at a site on the opposite subunit from where sY21 binds. The role of the sY7-binding site on CXCL12 is not yet clear. It appears that when complexed to a monomeric CXCL12 form, sY7 binds to the same subunit as sY21, and still at a different sY7 site. Whether sulfation of Y7 functions in regulation, or dimerization, remains unknown. It is therefore of interest to see whether a similar method of computer-aided structure-based drug discovery could be used to identify binders specific to the sY7-binding site. Such binders could then be used to study the functional role of the sY7-binding site in more detail, or even help the extension of compounds around the protein, and possibly linking fragments against each sY-binding site together in future optimization efforts.

**Virtual screening using ensemble approach.** The sY7-binding site appears rather flexible due to the presence of multiple arginine side chains. Therefore, in order to account for protein flexibility we used an ensemble docking approach in our virtual screening efforts. In this approach we screened the lead-like and fragment subsets of the ZINC database against a database of 20 protein conformations (NMR structure provided by the Volkman lab) (Fig. 2.23A). The results were combined and the compounds were ranked according to binding score. The compounds were then matched to their corresponding protein conformation, and redundant molecules were removed. The final top 1500 compounds were visually inspected and 50 lead-like compounds and 41 fragment compounds were prioritized for testing by 2D NMR.

**2D NMR studies.** From the 41 candidate compounds, 13 were tested by 2D NMR using the locked CXCL12 monomer. From the 13, three compounds (ZINC16480048, ZINC44900490,
and ZINC04181455) induced minor chemical shift perturbations, to residues that suggest that the compounds bind to the sY7-binding site area. The determined Kds showed weak binding at 800-1000 µM Kd (Fig. 2.23B).

**Figure 2.23. Identification of weak binders to sY7-binding site using ensemble virtual screening.**

A. Representation of sY7-binding site ensemble containing 20 CXCL12-sY12 (yellow) conformations. B. 2D NMR data for three compounds (ZINC16480048, ZINC44900490, and ZINC04181455) highlighting residues perturbed (green) and suggests binding to sY7-binding site.

**Conclusion**

The CXCL12-CXCR4 signaling axis has been a key target for drug discovery due to its involvement in various diseases, particularly cancer. Previous efforts focused on CXCR4 solely, since chemokines had been deemed “undruggable” due to their small size and shallow surfaces. However, as most drug discovery efforts against CXCR4 have failed in clinical trials due to toxicity issues, the need for alternative approaches has become apparent, such as directing efforts
to targeting CXCL12. We have currently demonstrated, through our efforts and effective collaboration that allowed application of tools such as NMR, X-ray crystallography, molecular docking, cell assays, and chemical synthesis, that a computer-aided structure-based approach can successfully identify compounds that bind to sulfotyrosine-binding sites on CXCL12.

We have so far identified and validated the role of ZINC310454 and various tetrazole-based analogs as CXCL12 inhibitors, which bind specifically to the sY21-binding site. We also have evidence that ZINC816454, a compound with a different pharmacophore identified by a flexible virtual screening approach, also binds to the sY21-binding site as well. Additionally, we have identified and verified ZINC2497769 as a specific binder of the sY12-binding site. We also have evidence that ZINC00095441, a compound identified through an ensemble screening approach, also binds to the sY12-binding site. Finally we have evidence that ZINC16480048, ZINC44900490, and ZINC04181455, compounds identified through an ensemble screening approach, may bind to the sY7-binding site with weak affinities.

In studying these compounds, we have confirmed that sY21-binding site binders act as inhibitors of the CXCL12-CXCR4 axis. Our data also suggest that sY12-binding site binders may function best as extensions of sY21-binding site inhibitors. This is important especially since we have shown that ZINC310454 and ZINC2497769 can bind simultaneously to their respective sites, suggesting that compounds against both the sY21 and sY12-binding sites can effectively be linked together producing highly potent inhibitors. Also, optimization of sY7-binding site compounds may shed light on the role of sY7 by studying inhibition of that particular interaction.

Finally, we have solved a CXCL12 complex crystal structure with BWA3111 bound to the sY21-binding site. This is the first structure of any inhibitor bound to any chemokine.
Besides the NMR CXCL12-CXCR4\textsubscript{1-38} structure (PDB ID: 2K05), the only structure of a CXCL12-ligand complex is the crystal structure of CXCL12 bound by a heparin disaccharide (PDB ID: 2NWG),\textsuperscript{50} a fragment of the natural ligand that binds at the same pocket as sY12 of CXCR4. Therefore, our complex structure provides vital information that can be used by others in designing inhibitors that bind the sY21-binding site.

Targeting CXCL12 in disrupting the CXCL12-CXCR4 signaling axis is of great interest to the chemokine community. Apart from one previous study identifying Chalcone-4 as a neutralizing ligand for CXCL12, but for which structural information was lacking,\textsuperscript{40} our virtual screening hits and the tetrazole derivatives are the only reported CXCL12-specific small molecule inhibitors. We have confirmed that sulfotyrosine binding sites on CXCL12 are achievable and practical targets. And since sulfotyrosine recognition by chemokines is predicted to contribute to most chemokine-receptor interactions, successfully targeting such sites could potentially extend to targeting other chemokines as well paving the way for chemokine targeted drug discovery.\textsuperscript{51}

**Experimental Procedures**

**Purification**

Smt3-CXCL12 was cloned into the pET 28a plasmid and transformed into the BL21 (DE3) expression strain. Cells were grown at 37°C in TB media to an OD\textsubscript{600} of 0.8 and induced with 1.0mM IPTG for 5 hours. Cells were pelleted by centrifugation, resuspended in 50mM NaH\textsubscript{2}PO\textsubscript{4} pH7.4, 300mMNaCl, 10mM Imidazole, and 0.1% BME, and lysed in a French pressure cell. CXCL12 was expressed exclusively in insoluble inclusion bodies. The inclusion
body pellet was isolated by centrifugation at 15,000 x g for 30 minutes, and resuspended in 6M Guanidine HCl, 50mM NaH2PO4 pH7.4, 300mM NaCl, 10mM Imidazole, and 0.1% BME. Following centrifugation at 15,000 x g for 30 minutes, the supernatant containing soluble, unfolded CXCL12, was loaded onto a Ni2+-NTA resin gravity column and eluted through a decrease in pH (6M guanidine HCl, 100mM NaOAc pH 4.5, 300mM NaCl, 10mM Imidazole). Refolding was accomplished through drop-wise serial dilution in a refolding buffer (100mM Tris pH 8.0, 10mM Cysteine, 0.5mM Cystine) followed by overnight SUMO cleavage at 30°C. Separation of CXCL12 from the His-tag and SUMO protease was performed through binding of cleaved CXCL12 to a strong cation-exchange column. After separation, the protein was further purified through reverse-phase HPLC with an Acetonitrile gradient of 30-60%. Pure CXCL12 was lyophilized and re-dissolved in water.

**NMR Spectroscopy**

NMR data was collected on a Bruker Avance 600 MHz spectrometer equipped with a TCI cryoprobe at 298 K. Lyophilized [U-15N]-CXCL12WT was reconstituted in a 25 mM deuterated MES, 10% (v/v) D2O, 0.02% (w/v) NaN3 buffer, pH 6.8. NMR samples of 50 µM CXCL12WT and 0-1600 µM compound were made with a LEAP PAL robot where each sample contained 2.0% deuterated DMSO. A Bruker SampleJet was used for automated sample handling and 1H-15N Heteronuclear Multiple Quantum Coherence (HMQC) spectra were collected for each titration point. Spectral data was processed with in house scripts and chemical shift changes were tracked using CARA software. Total 1H and 15N chemical shift perturbations were calculated as previously shown.
Crystallization

CXCL12 was concentrated to 8 mg ml\(^{-1}\) for crystallization trials. The Qiagen JCSG suites were screened using the Phoenix Microdispenser and hits were optimized and tested for space group determination. Hits that led to crystals with a trigonal space group (e.g. PDB ID: 3GV3) were unsuccessful in binding compound 1 since the sY21-binding site was completely blocked by crystal packing residues. Hundreds of small crystals appeared in a condition containing 3.5 AmSO\(_4\), 1%MPD, and 0.1 M MES Sodium Salt pH 6.5. These were optimized (2 M AmSO4, 2% MPD, and 0.1 M MES Sodium Salt pH 6.5) to consistently produce a few and large single crystals. Surprisingly, two forms of crystals appeared in this condition. One form belonged to the ineffectual trigonal space group, while the other to an orthorhombic space group (e.g. PDB ID: 2J7Z).

Since crystals belonging to two different space groups would form in the same crystallization drop, seeding was used to induce crystallization of the orthorhombic space group instead of the trigonal space group form. Crystals belonging to the P\(2_12_12_1\) space group would grow to their full size in 3-4 days, measure up to 400 µm in each dimension and diffract up to 1.8 Å.

Attempts to grow complex CXCL12 crystals in the presence of compound 1 were unsuccessful. We therefore used a soaking method. Since compound 1 was insoluble at high concentrations we resorted to introducing it in its solid powder form at various amounts directly on the crystal drop. Soaking was also performed for only 30 minutes since compound 1 would quickly degrade the crystals upon introduction. Crystals were then flash-frozen in 2.2 M AmSO4, 25% Glycerol, 2% MPD, and 0.1 M MES Sodium Salt pH 6.5. Data sets were collected
for many crystals in search for the best candidate, since solubility issues with compound 1 produced inconsistent results with compound density, while damaging a lot of the crystals.

**Data collection and processing**

Data was collected at the Moffitt Cancer Center home source with a Rigaku Raxis detector. Data was processed with XDS\textsuperscript{54} and the structure was solved through molecular replacement (using PDB ID: 2J7Z) with MOLREP\textsuperscript{55} and refinement with Refmac5\textsuperscript{56} and PDB_REDO.\textsuperscript{57} Figures were made using PYMOL (www.pymol.org). The structure has been deposited into the Protein Data Bank with accession code 4UAI.

**Molecular docking**

Computational docking studies were performed with the program DOCK 3.5.54. To prepare each protein receptor file for docking, all hydrogens, ions and waters were removed when present. Matching spheres were created by the respective ligands in each pocket, and a default total of 60 atom-derived spheres were used. Pre-generated ligand conformations were sampled in each binding pocket and scored. Ligand partial atomic charges and disolvation were calculated by AMSOL. The ligand and receptor bin size/overlap were set at the default setting of 0.4/0.3 and 0.4/0.3 respectively. The energy scoring grids were calculated by summing the energies calculated by DISTMAP (excluded volume grid), CHEMGRID (van der Waals potential), and DelPhi (desolvation).
Chapter 3

Pseudomonas aeruginosa LpxA Structural Analysis and Virtual Screening

Overview

In Gram-negative bacteria, the first step of lipid A biosynthesis is catalyzed by UDP-N-acetylglucosamine acyltransferase (LpxA) through the transfer of a R-3-hydroxyacyl chain from the acyl carrier protein (ACP) to the 3’-hydroxyl group of UDP-GlcNAc. Acyl chain length selectivity varies between species of bacteria, but is highly specific and conserved within certain species. In E. coli and L. interrogans for example, LpxA is highly selective for longer R-3-hydroxyacyl chains (C14 and C12 respectively), while in P. aeruginosa the enzyme is highly selective for R-3-hydroxydecanoyl, a 10-hydrocarbon long acyl chain. Here we report three P. aeruginosa LpxA crystal structures; the apo form, the complex with its substrate UDP-GlcNAc, and the complex with its product UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc. The unit cell of each structure belongs to the same P2 space group with six monomers per asymmetric unit, three of which form the functional trimer. A comparison between the apo form and complexes identifies key residues that position UDP-GlcNAc appropriately for catalysis, and supports the role of His121 in generating the nucleophile by interacting with the UDP-GlcNAc 3’-hydroxyl group. Furthermore, the product-complex structure supports the role of Met169 as the “hydrocarbon ruler”, providing structural information on how P. aeruginosa LpxA is granted its exceptional selectivity for the 10-hydrocarbon long acyl chain. Lastly, structural analysis of the
active site aided in developing an effective hybrid \textit{in silico}/X-ray crystallography screening strategy, which led to the identification of two binders. Such information can lead to the development of LpxA specific inhibitors, a newly recognized target in combating drug resistant bacteria such as \textit{P. aeruginosa}.

\textbf{Introduction}

\textit{Pseudomonas aeruginosa} is an opportunistic pathogen that can thrive in many environments and infect many hosts.\textsuperscript{58} Of particular susceptibility to infection are immunocompromised humans such as in the case of people who suffer from cystic fibrosis, impaired immunity, or severe burns.\textsuperscript{59} Serious infections can lead to death and are predominantly hospital acquired. Furthermore, the pathogen has the ability to acquire resistance to multiple drugs through various mechanisms and is therefore of major concern to clinicians since it is making infections increasingly difficult to treat.\textsuperscript{60-63} Consequently, \textit{P. aeruginosa} has entered the category of Superbugs, further highlighting the importance of identifying and studying new druggable targets.\textsuperscript{62,64}

In \textit{P. aeruginosa}, just like in most Gram-negative bacteria, the lipopolysaccharide (LPS) is a major component of the outer membrane that protects the bacterium from its environment.\textsuperscript{65,66} It is highly immunogenic causing virulence upon infection,\textsuperscript{62,67,68} and as a result it is referred to as “endotoxin”.\textsuperscript{65,67} The LPS structure may differ slightly from one bacterium to another, but it is mainly comprised of three components; the O-antigen, the Core oligosaccharide, and Lipid A.\textsuperscript{66,69-71} Lipid A is a glucosamine-based phospholipid that anchors the lipopolysaccharide to the outer monolayer of the outer membrane.\textsuperscript{65,69,70,72} In \textit{P. aeruginosa}, Lipid A is made out of a diglucosamine biphosphate backbone that is linked by O- and N- primary and secondary fatty
This component is integral to the bacterium’s survival, and thus has become a desirable target for drug discovery.\(^73\)

The pathway leading to the biosynthesis of Lipid A has been well documented in \textit{E. coli}.\(^65,74\) It takes place in a series of steps involving 9 enzymes.\(^65\) The first step is mediated by UDP-N-acetylglucosamine (UDP-GlcNAc) acyltransferase (LpxA), which catalyzes the reversible transfer of a R-3-hydroxyacyl from the R-3-hydroxyacyl-acyl carrier protein (ACP) to the glucosamine 3-OH group of UDP-GlcNAc.\(^73,75\) This first step is thermodynamically unfavorable (\(K_{eq}\) of \(\sim 0.01\))\(^75\) and therefore reversible, but LpxC, the second enzyme in the pathway, catalyzes the committed step through the irreversible deacetylation of UDP-3-O-(R-3-hydroxyacyl)-GlcNAc.\(^76\) The third step is catalyzed by LpxD through the addition of a second \(\beta\)-hydroxyacyl chain to make UDP-2,3-diacyl-GlcNAc. A series of additional steps result in the production of Kdo\(_2\)-Lipid A, which is then attached onto the Core component of the LPS (Fig. 3.1).\(^65,77\)

\[\text{UDP-GlcNAc} + \text{3-OH-acyl-ACP} \rightleftharpoons \text{UDP-acyl-GlcNAc} + \text{ACP}\]

\[\text{UDP-acyl-GlcNAc} + \text{UDP-GlcNAc} \rightleftharpoons \text{UDP-2,3-diacyl-GlcNAc} + \text{UDP}\]

\[\text{UDP-2,3-diacyl-GlcNAc} \rightarrow \text{Kdo}\(_2\)-Lipid A\]

\(\text{Figure 3.1. } P.\text{ aeruginosa} \text{ LpxA catalyzes the first and reversible step of Lipid A biosynthesis. LpxA catalyzes the reversible transfer of 3-OH-C\(_{10}\) from the ACP protein to the 3’-OH position (highlighted in yellow) of UDP-GlcNAc. A series of additional enzymes carry on the remaining steps that lead to the biosynthesis of Kdo}\(_2\)-Lipid A.}\]
LpxA is a soluble cytoplasmic protein that forms a functional homotrimer.\textsuperscript{76, 78} It shares sequence homology with LpxD, which also forms a homotrimer, unlike LpxC which is a Zn\textsuperscript{2+}-dependent enzyme that shares no sequence homology with other deacetylases.\textsuperscript{65} Previous crystal structures have revealed that LpxA forms a left-handed parallel $\beta$-helix generated by 30 hexapeptide repeats.\textsuperscript{79-84} Every turn in the $\beta$-helix is made up of a single hexapeptide.\textsuperscript{76}

Conserved residues, identified previously through sequence alignments, assisted in the identification of the active site, which is located at a cleft formed by the dimer interface. The catalytic residue, which is conserved across all species, is a histidine that primes the 3-OH of GlcNAc for nucleophilic attack through deprotonation.\textsuperscript{85} For this to occur, the substrate, UDP-GlcNAc, is appropriately oriented in the active site through the formation of multiple hydrogen bonds with other conserved residues. Finally, the acyl chain, carried by the Acyl Carrier Protein (ACP), docks into the hydrophobic cleft and gets displaced from the ACP and subsequently attached to the GlcNAc forming the product, UDP-3-O-(R-3-hydroxyacyl)-GlcNAc. In \textit{E. coli}, the catalytic residue is His125,\textsuperscript{85} in \textit{L. interrogans} it is His120,\textsuperscript{79} while in \textit{P. aeruginosa} it has been identified as His121.\textsuperscript{76}

LpxA enzymes are highly selective of their substrate’s acyl chain length, but this selectivity varies across species, with some bacteria being highly selective of longer chain hydrocarbons while other bacteria being highly selective of shorter chain hydrocarbons.\textsuperscript{86, 87} In \textit{E. coli} for example, LpxA catalyzes the transfer of R-3-hydroxymyristoyl, a 14C-long fatty acyl chain, from the R-3-hydroxymyristoyl-acyl carrier protein (ACP) to GlcNAc. \textit{E. coli} LpxA selects this 14C-long chain over a 10C-long chain by a factor of $\sim$1000.\textsuperscript{70, 87} In \textit{L. interrogans}, LpxA is highly selective for 3-R-hydroxylauroyl, a 12C-long fatty acyl chain.\textsuperscript{79} In \textit{P. aeruginosa} however, LpxA is highly selective for R-3-hydroxydecanoyl, a 10C-long R-3-hydroacetyl chain.\textsuperscript{88}
Lipid A isolates from *E. coli* and *P. aeruginosa* have confirmed incorporation of these highly specific chains. Structural studies have shown that in *E. coli* LpxA, His191 might be implicated in restricting chains longer than 14C to fit, while in *L. interrogans* LpxA, structural studies have shown that Lys171 serves a similar purpose. In *P. aeruginosa* LpxA, Met169 has been shown to function as the “hydrocarbon ruler”, denying longer chain hydrocarbons from binding. Furthermore, it has been shown that this high selectivity can be reversed *in vitro* and *in vivo* between *E. coli* LpxA and *P. aeruginosa* LpxA through reciprocal mutations of G173M and M169G respectively. The evolution of this variation in high selectivity between species is still a matter of debate, but progress has been made in identifying key residues. Until now however, the structure of *P. aeruginosa* has been lacking, so a comprehensive understanding on how this selectivity is accomplished by Met169 was incomplete.

In this chapter the first structures of *P. aeruginosa* LpxA are presented in three forms; in the apo form at 2.17 Å, in complex with the substrate UDP-GlcNAc at 2.07 Å, and in complex with the product UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc at 2.52 Å. The unit cell of each structure belongs to the same P2 space group; there are six monomers in each asymmetric unit, three of which are forming the biologically relevant homotrimer. *P. aeruginosa* LpxA contains high structural homology to other previously solved LpxA structures, such as *E. coli* and *L. interrogans*, but also contains variation in residue composition and side chain positions, especially in flexible loop regions. Furthermore, the structures unambiguously identify the active site and provide detailed information on the residues that orient GlcNAc for catalysis, the role of H121 as the catalytic residue, and on how M169 confers *P. aeruginosa* LpxA with its exceptional selectivity for R-3-hydroxydecanoyl, the 10C-long hydrocarbon chain, thus serving as the “hydrocarbon ruler”. This information helped to better understand substrate recognition and
catalysis in this class of enzymes, but also led to successful virtual screens and the identification of two small molecule binders, confirmed through X-ray crystallography to bind to the active site. These results will pave the way for a structure-based approach in developing \textit{P. aeruginosa} LpxA specific inhibitors, since LpxA has been validated as a promising target for antibiotics.\textsuperscript{76}

\textbf{Results}

\textit{P. aeruginosa} LpxA apo structure

The \textit{P. aeruginosa} LpxA apo crystal diffracted up to 2.17 Angstrom resolution and belongs to the $P2_12_12_1$ space group (Table 3.1). Each asymmetric unit contains six monomers. Three of those monomers are forming the biologically relevant homotrimer through a non-crystallographic 3-fold symmetry (Fig. 3.2A). The remaining three monomers in the asymmetric unit are also forming functional homotrimers, but with monomers from adjacent asymmetric units. The conformation of each monomer is virtually identical. Superimposition of each monomer in the biologically relevant trimer shows negligible structural variation, mainly in the flexible loop regions (Fig. 3.2B), and with an average RMSD value of 0.337 aligning an average of 1519 atoms.

All 258 amino acids in the apo structure were clearly defined. The structure of \textit{P. aeruginosa} LpxA, which highly resembles orthologs in other bacteria, is made up of two distinct domains; an N-terminal $\beta$-strand domain (purple) and a C-terminal $\alpha$-helical domain (blue) (Fig. 3.2A-C). Ten $\beta$-helical coils form the typical L$\beta$H (Left-handed parallel beta-Helix) domain of the N-terminus, a motif common in enzymes with acyltransferase activity. Starting at Met1 and ending at Ser192, this coiled motif consists of 29 distinct $\beta$-strands, 27 of which are
hexapeptides. Every three hexapeptides form one complete β-helical coil, creating the appearance of an equilateral triangle. Side chains in the coil hexapeptides alternate between those faced internally, and those faced externally. Those faced internally establish polar and non-polar interactions further rigidifying the structure. Finally, the last coil, which precedes the α-helical domain, is shorter compared to the rest.

Two loops (L1 and L2) extend out of the LβH domain at coils C3-C4 and C4-C5, interrupting the motif. These coils show slight conformational variation between different monomers in the asymmetric unit (Fig. 3.2B), and structural variation between orthologs in other bacteria (Fig. 3.2C). The first loop, which extends out of C3 and into C4, is 15 residues long (Gly65-Arg79), and the second loop extends out of C4 and into C5 (Arg96-Glu104) and is 9 residues long.

Immediately following the LβH domain, a series of 4 consecutive α-helices of different sizes and tilts are formed, starting at Asn194 and extending all the way to the C-terminus. The α-helical domain shows some conformational variation among ortholog proteins from other bacterial and plant species. While P. aeruginosa LpxA is most similar to the E. coli LpxA (PDB ID: 1LXA), there seems to be some noticeable variation at the C-terminal end of the third α-helix, involving dissimilar tilts compared to bacterial orthologs A. baumannii (PDB ID: 4E6U), L. interrogans (PDB ID: 3HSQ), C. jejuni (PDB ID: 3ROS), B. thailandensis (PDB ID: 4EQY), H. pylori (PDB ID: 1J2Z), and even plant ortholog A. thaliana (plant) (PDB ID: 3T57), (Fig. 3.2C).

The catalytic site in the biologically relevant trimer lies in the dimer interface where coils C6-C9 from adjacent monomers meet. Therefore, each biologically relevant trimer contains three catalytic sites. While there is sequence similarity in active site residues between orthologs, there
are also differences. Yet, catalytic residues His121 and His140 in *P. aeruginosa* LpxA (Fig. 3.2D) are highly conserved across all species (Fig. 3.2E). This suggests the same mode of catalysis, but perhaps variation in ligand positioning between orthologs.

Figure 3.2. *P. aeruginosa* LpxA apo crystal structure. A. The biologically relevant homotrimer contained within the asymmetric unit seen from two side views and one top view. B. Superimposition of all three monomers from the biologically relevant homotrimer. C. *P. aeruginosa* LpxA monomer superimposed to seven ortholog LpxA monomeric structures (*E. coli* (PDB ID: 1LXA), *B. thailandensis* (PDB ID: 4EQY), *A. baumannii* (PDB ID: 4E6U), *L. interrogans* (PDB ID: 3HSQ), *H. pylori* (PDB ID: 1J2Z), *A. thaliana* (PDB ID: 3T57), and *C. jejuni* (PDB ID: 3ROS)). Superimposition shows high structural similarity, with variation in flexible loop regions and alpha-helical domains. D. *P. aeruginosa* catalytic site highlighting conserved catalytic residue His121 and conserved His140. E. *P. aeruginosa* LpxA superimposed to all seven ortholog LpxA structures available to date in the PDB shows highly conserved histidine residues in the active site.
**P. aeruginosa LpxA/UDP-GlcNAc complex structure**

The *P. aeruginosa* LpxA/UDP-GlcNAc complex structure, obtained from the LpxA apo crystal soak, was solved at 2.05 Ångstrom resolution (Table 3.2). The resulting electron density unambiguously identified the presence and localization of six UDP-GlcNAc molecules within the asymmetric unit. Three of these molecules appear at the dimer interface active site. However, only one of the three is at a dimer interface active site contained within the asymmetric unit (Fig. 3.3A), while the other two are at active sites formed at the crystal-packing interface by monomers from adjacent asymmetric units. For these two ligands, applying symmetry related operations yields the complete active site and identifies all ligand-protein interactions, confirming that all three molecules bind to the active site in virtually the same way (Fig. 3.3B). Therefore, the focus hereinafter will be on the active site formed within the asymmetric unit at the dimer interface between monomers A and B (Fig. 3.3A-F). The remaining three UDP-GlcNAc molecules are bound to non-specific sites. One of them, seems to loosely interact with residues of monomer C (green) right at the cleft where α-helix 1 and the β-strand domain meet. This interaction is not supported by many hydrogen bonds, and the electron density for the GlcNAc moiety is not well defined, probably as a result of conformational disorder. This suggests that this binding pose is a crystal-packing artifact. The other two molecules however bind at the dimer interface, but at a secondary site situated below the active site, on the N-terminal side of loop L1. One of these two molecules binds at the dimer interface within the asymmetric unit (below the active site), while the other is interacting with monomers from adjacent asymmetric units, but once symmetry operations are applied, the binding poses of those two molecules are nearly identical. The focus hereafter will be on the molecule contained within
the asymmetric unit and since the interactions are with dimer interface residues (chain A and B), it is of interest to postulate on any potential function this secondary site might have.

UDP-GlcNAc interacts with the active site mostly through polar interactions with residues from both monomer A (blue) and B (green), as well as three water molecules (Fig. 3.3C). The uracil moiety and certain atoms of the diposphate moiety hydrogen bond to monomer A (blue), while the ribose, GlcNAc, and other atoms of the diphosphate moieties interact with monomer B (green) (Fig. 3.3C). On the one end of the UDP-GlcNAc molecule, the 3-N atom of uracil is within hydrogen bond distance to Asn194/OD1 (Chain A), while the 4-O atom of uracil is within hydrogen bond distance to Asn194/ND2 (Chain A) (Fig. 3.3D). The remaining non-polar atoms of the uracil ring appear to lie against the aromatic side chain of Phe166 (chain A) (Fig. 3.3D). The 3-OH of the ribose moiety hydrogen bonds with His156/NE2 (chain B), while the remaining polar atoms of that moiety appears to interact only with the solvent (Fig. 3.3E). Arg200/NH2 (chain A) interacts with the linker-O atom of α-phosphate and Arg200/NH1 with an O atom of the β-phosphate (Fig. 3.3D). An α-phosphate O atom and the O atom linking α-phosphate and β-phosphate are within hydrogen bond distance to Gln157/NE2 (chain B) (Fig. 3.3E). Meanwhile, the GlcNAc moiety appears to implicate an abundance of hydrogen bonds with residues of chain B. The GlcNAc 6-OH forms hydrogen bonds with a series of polar atoms, such as Tyr158/OH, His140/NE2, Asp70/OD2, and Lys72/NZ (Fig. 3.3F). In addition, a water molecule appears to mediate interactions between the 6-OH of GlcNAc, and Lys72/NZ, Asp70/OD2, and Thr98/OG1 of chain B (Fig. 3.3F). The 3-OH and the 2-OH groups of GlcNAc are within hydrogen bond distance to the catalytic base His121/NE2 (chain B), which functions in deprotonating 3-OH, priming it for catalysis (Fig. 3.3F). Lastly, the acetyl N atom is within
hydrogen bond distance to Asp70/OD1, while the acetyl O atom interacts with the backbone Leu71/N atom (Fig. 3.3F).

Figure 3.3. UDP-GlcNAc bound to *P. aeruginosa* LpxA active site. A. The Fo-Fc map unambiguously identifies the binding pose of UDP-GlcNAc in the active site. B. Superimposition of three UDP-GlcNAc molecules observed in biologically relevant active sites within the asymmetric unit shows that the binding pose is nearly identical. C. Hydrogen bonding interactions in active site formed at dimer interface where monomer A (blue) and monomer B (green) meet. D. Uracil and phosphate moiety interacts with monomer A (blue). E. Ribose and phosphate moieties interact with residues of monomer B (green). F. Glucosamine moiety interacts with residues of monomer B (green).

Comparison of the LpxA apo structure to the LpxA/UDP-GlcNAc complex shows the conformational changes induced upon ligand binding in order to better accommodate the ligand (Fig. 3.4A). Apparent changes are observed in Gln157 (chain B), which flips away from the pocket to make room for the ligand and also hydrogen bond to the α-phosphate. Conformational changes are also observed in Lys72 (chain B), which better positions itself to interact with the GlcNAc moiety. Asp70 (chain B) also moves away from the pocket, to better complement the ligand, and interact with both 6-OH and the acetyl group. On the A chain, Arg200, which appears in two alternative conformations in the apo structure, adopts a single rigid conformation in the
complex structure, probably due to the reinforced hydrogen bonding with the biphosphate linker (Fig. 3.4A).

Currently in the PDB, there is only one ortholog LpxA structure complexed with UDP-GlcNAc, and that is of *E. coli* (PDB ID: 2JF3). The *E. coli*/UDP-GlcNAc complex structure contains one monomer per asymmetric unit. By applying symmetry related operations and superimposing the *E. coli* complex to the *P. aeruginosa* LpxA complex structure, direct comparison of the binding pose of UDP-GlcNAc between these two orthologs demonstrates that UDP-GlcNAc binds quite differently in these two structures (Fig. 3.4B). The biggest difference lies in the positioning of the uridine moiety, which is flipped around the Phosphate-O linker bond, when comparing the *E. coli* and *P. aeruginosa* complexes (Fig. 3.4B). Hydrophobic interactions, like the ones observed between UDP-GlcNAc and Phe166 (chain A) (Fig. 3.3D), are not observed in the *E. coli* complex. Also, even though His156 (chain B), which interacts with the ribose moiety in *P. aeruginosa* LpxA, is conserved between these two orthologs, in the *E. coli* complex it does not interact with ribose and it also flipped away from the ligand (Fig. 3.4B, His160). However, while the position of the α-phosphates is different in these two binding poses, the position of the β-phosphates is quite similar, and even more so is the positioning of the GlcNAc moiety. Therefore, the hydrogen bond network between the GlcNAc moieties in these two complexes is very similar. An exception is that Tyr158 in *P. aeruginosa* LpxA is Phe162 in *E. coli* LpxA (Fig. 3.4B), and is thus incapable of hydrogen bonding to 3-OH of GlcNAc. Also, the water molecule mediator seen in the *P. aeruginosa* LpxA complex is not seen in the *E. coli* LpxA complex.

The UDP-GlcNAc molecule observed at the secondary site below the active site also establishes multiple favorable interactions (Fig. 3.4C). The 3-N of uracil is within hydrogen bond
distance to Asp5/OD2 (chain A), while 2-OH is within hydrogen bond distance to Arg7/NH2 (chain A). The ribose 2-OH is within hydrogen bond distance to Asp11/OD1 (chain B) and a β-phosphate O atom is interacting with the backbone Ala30/N atom (chain B). The 2-OH of GlcNAc is within hydrogen bond distance to the backbone Gly48/O atom and the 3-OH within hydrogen bond distance to Asp67/OD1 (chain B). A water molecule is within hydrogen bond distance to Ala30/O and the backbone Gly48/N atom of chain B, and the acetyl N atom, the linker-O of β-phosphate, and the O atom of β-phosphate of the ligand (Fig. 3.4C).

Figure 3.4. Structural comparisons and possible secondary site for UDP-GlcNAc. A. Comparison of P. aeruginosa Lpxa complex (blue for monomer A, green for monomer B) to the apo P. aeruginosa Lpxa (purple for monomer A, pink for monomer B) shows conformational changes induced upon binding of UDP-GlcNAc (yellow). B. Comparison of P. aeruginosa Lpxa/UDP-GlcNAc complex (blue for monomer A, green for monomer B) to the E. coli Lpxa/UDP-GlcNAc complex (monomer A - light pink) shows difference in monomer A amino acid composition, but also difference in UDP-GlcNAc orientation (P. aeruginosa ligand (yellow)/E. coli ligand (dark pink)). C. UDP-GlcNAc binds at a secondary site on the same dimer interface as the active site, but on the N-terminal side of loop L1. A rather extensive hydrogen bond network supports this binding pose.
**P. aeruginosa** LpxA/ UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex structure

The *P. aeruginosa* LpxA/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex structure, obtained from a LpxA apo crystal soak, was solved at 2.52 Angstrom resolution (Table 3.3). Surprisingly, electron density for the product was observed in all 6 active sites contained within the asymmetric unit, compared to only three active sites occupied in the LpxA/UDP-GlcNAc complex structure. Furthermore, it appears that for chain A and C, loop L1 changes conformation from that observed in the apo and UDP-GlcNAc complex, in order to accommodate the product. However, electron density was ambiguous in certain areas and for certain ligands. Therefore, the focus hereafter will be at the active site formed between chain A and B.

The Fo-Fc map unambiguously identifies the binding pose of UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc at the A-B chain interface, which appears for the most part to bind similarly to how UDP-GlcNAc binds (Fig. 3.5A). The UDP-GlcNAc moiety uses the same hydrogen bond network as seen in the substrate complex (Fig. 3.5B). The 3-OH of the GlcNAc moiety is still within hydrogen bond distance to the catalytic base His121/NE2, and the acyl chain extends from then forth. His118/NE2 appears within hydrogen bond distance to the acyl chain carboxyl, which might help anchor the chain into the hydrophobic cavity (Fig. 3.5C). The acyl chain appears to possess multiple hydrophobic contacts with the protein. Apolar atoms of the acyl chain complement the hydrophobic tunnel formed by chain A (blue) residues Val132, Tyr152, Met169, the hydrophobic side of Asn133, and the backbone of Gly151, and chain B (green) residues Ala136, Ala138, and Leu154 (Fig. 3.5C-D).

The function of Met169 is now evident in the structure. It acts as a “hydrocarbon ruler” by allowing incorporation of acyl chains that are up to 10 hydrocarbons long (Fig. 3.5E). Chains
longer 10 hydrocarbons would clash with Met169 (Fig. 3.5F-G), while shorter hydrocarbon chains would have reduced hydrophobic interactions with the pocket and therefore result in reduced affinity.

Figure 3.5. UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc bound to P. aeruginosa LpxA active site. A. The Fo-Fc map unambiguously identifies the binding pose of UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc in the active site. B. UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc binds to the active site utilizing many of the same interactions as UDP-GlcNAc. C. Residues from adjacent monomers form a hydrophobic cleft that accommodates the acyl chain of UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc. D. Surface model shows great complementarity between the acyl chain and hydrophobic pocket in the P. aeruginosa LpxA/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex structure. E. P. aeruginosa LpxA/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex structure shows the role of Met169 as the hydrocarbon ruler. F. The 14C-long acyl chain product (purple) specific to E. coli LpxA would clash with Met169 in P. aeruginosa. G. The 12C-long acyl chain product specific to L. interrogans LpxA (pink) would clash with Met169 in P. aeruginosa.

Structures of E. coli and L. interrogans LpxA complexed to acetylated UDP-GlcNac products have been previously solved; one E. coli LpxA is in complex with UDP-3-O-(R-3-
hydroxymyristoyl)-GlcNAc, a 14C-long chain product of *E. coli* LpxA (PDB ID: 2QIA) (Fig. 3.6A), and the other is in complex with UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc (PDB ID: 2QIA) (Fig. 3.6B), which is the *P. aeruginosa* LpxA product. The *L. interrogans* LpxA is in complex with its 12C-long acyl chain product (Fig. 3.6C). Superimposition of these structures onto the *P. aeruginosa* LpxA/UDP-3-O-(R-3-hydroxyacyl)-GlcNAc complex structure reveals many similarities but also demonstrates key differences (Fig. 3.6A-D).

UDP-3-O-(R-3-hydroxymyristoyl)-GlcNAc binds to *E. coli* LpxA in an orientation similar to our *P. aeruginosa* LpxA/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex structure. Only minor differences are observed, which are probably due to minor differences in binding site residue composition (Fig. 3.6A). Interestingly, the orientation of UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc in the second *E. coli* LpxA complex is quite different from the *E. coli* LpxA UDP-3-O-(R-3-hydroxymyristoyl)-GlcNAc complex. While the acetylated GlcNAc moieties and β-phosphate of both ligands are oriented virtually the same, the α-phosphate and uridine moieties are reversed through rotational reorientation of the P-O bonds. In fact, the positioning of the uridine moiety in UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc in *E. coli* is the same as UDP-GlcNAc in the *E. coli* complex structure (Fig. 3.6B). Even though such an orientation was not observed in any of our complex structures, this still suggests that there might be a secondary conformation in vivo, which aids in catalysis. Superimposition to the previously solved *L. interrogans* LpxA/product complex structure also suggests that acylated products bind similarly in both *L. interrogans* LpxA and *P. aeruginosa* LpxA. The alternative uridine conformation observed in the *E. coli* LpxA/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc (Fig. 3.6B) or *E. coli* LpxA/UDP-GlcNAc (Fig. 3.4B) complex structures, is not observed in the *L. interrogans* LpxA complexes. Yet some variation in the GlcNAc moiety positioning is observed.
when comparing *P. aeruginosa* LpxA and *L. interrogans* LpxA complexes (Fig. 3.6C). The variations seen between all the ligand poses mentioned (Fig. 3.6D) are probably due to the slightly different active site residue composition between the three orthologs.

**Figure 3.6. Structural comparisons to ortholog LpxA/product complexes.** A. Comparison of *P. aeruginosa* Lpxa/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex (blue for monomer A, green for monomer B) to the *E. coli* Lpxa/UDP-3-O-(R-3-hydroxymyristoyl)-GlcNAc complex (monomer A - light pink) shows difference in monomer A amino acid composition, but also difference in ligand orientation (*P. aeruginosa* ligand (yellow)/*E. coli* ligand (purple)). B. Comparison of *P. aeruginosa* Lpxa/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex (blue for monomer A, green for monomer B) to the *E. coli* Lpxa/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex (monomer A - light pink) shows difference in monomer A amino acid composition, but also difference in ligand orientation (*P. aeruginosa* ligand (yellow)/*E. coli* ligand (dark pink)). C. Comparison of *P. aeruginosa* Lpxa/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex (blue for monomer A, green for monomer B) to the *L. interrogans* Lpxa/product complex (monomer A - orange) shows difference in monomer A amino acid composition, but also difference in ligand orientation (*P. aeruginosa* ligand (yellow)/*L. interrogans* ligand (pink)). D. Products from all structures superimposed (*P. aeruginosa* ligand (yellow)/*E. coli* 14C ligand (purple)/*E. coli* 10C ligand (dark pink)/*L. interrogans* ligand (pink)). E. *E. coli* complex shows His191 “hydrocarbon ruler”. F. *L. interrogans* complex shows Lys171 “hydrocarbon ruler”.

82
A major difference however between the three orthologs lies in the hydrophobic tunnel that accommodates the acyl chains. *E. coli* LpxA accommodates a 14 hydrocarbon-long acyl chain, and anything longer than that is restricted by His191, which acts as the “hydrocarbon ruler” (Fig. 3.6E). Likewise, in the *L. interrogans* LpxA, anything longer than a 12 hydrocarbon-long acyl chain would be restricted by Lys171, its corresponding “hydrocarbon ruler” (Fig. 3.6F).

Finally, the proposed orientation for the oxyanion hole in the *E. coli* and *L. interrogans* complex structures (residues Glu143 and Glu138 respectively) is not observed in the *P. aeruginosa* complex structure. The corresponding residue in *P. aeruginosa* LpxA is Glu139. However the backbone N of Glu139 is flipped away from the ligand site, and it appears that the acyl chain carbonyl of UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc is facing the opposite direction into the solvent (Fig. 3.7A). Our proposed carbonyl orientation could be incorrect due to the poor resolution (2.52 Angstrom), but regardless of what direction it is truly facing, the carbonyl would still be too far to interact with any residues that could potentially function as an oxyanion hole. The unbiased Fo-Fc maps for the apo (2.17 Angstrom), substrate-complex (2.05 Angstrom), and product-complex (2.52 Angstrom), all suggest that the backbone carbonyl of Ala138 is oriented towards the active site, while the backbone N of Gly139 is orientated away from the active site, making bonding to the acyl chain difficult (Fig. 3.7C-E). In the *L. interrogans* complex structure, the acyl chain carbonyl is positioned within hydrogen bond distance to the respective Gly residue, supporting the role of the conserved Gly as the oxyanion hole (Fig. 3.7B). The orientation of highly conserved Gly139 and also Ala138 is contradictory to what has been previously observed in other ortholog LpxA structures (Fig. 3.7F).
Figure 3.7. Alternative conformation observed for highly conserved residues Gly139 and Ala138. A. The orientations of Gly139, Ala138, and the product acyl chain carbonyl are not supporting the role of Gly139 as the oxyanion hole. B. The P. aeruginosa LpxA/UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex structure superimposed to the L. interrogans/product complex structure shows opposing orientations for highly conserved residues Gly and Ala. C-E. Unbiased Fo-Fc maps for P. aeruginosa LpxA/substrate at 2.05Å (C), Apo at 2.17Å (D), and product complex at 2.52Å (E) all suggest that the backbone N of Gly139 is orientated away from the active site, while the backbone carbonyl of Ala138 is pointing towards the active site. F. Superimposition of seven ortholog LpxA structures (E. coli (PDB ID: 1LXA), B. thailandensis (PDB ID: 4EQY), A. baumannii (PDB ID: 4E6U), L. interrogans (PDB ID: 3HSQ), H. pylori (PDB ID: 1JZ2), A. thaliana (PDB ID: 3T57), and C. jejuni (PDB ID: 3ROS)) show an alternative orientation for highly conserved Gly and Ala, suggesting the role of Gly as the oxyanion hole.
Virtual screening against *P. aeruginosa* LpxA

Virtual screens were performed after structure determination in search of potential inhibitors. The *P. aeruginosa* LpxA/UDP-GlcNAc complex structure was used to perform preliminary virtual screens, since it was the complex structure that possessed the highest resolution (2.05 Å). The leadlike and substrate subsets of the ZINC database were screened and the 1000 highest scoring compounds from both rounds were visually inspected for complementarity. However, few compounds (<10%) possessed hydrophobic scaffolds that fit the hydrophobic cleft that normally contains the acyl chain, a characteristic that was desired in potential binders. Next, virtual screens were repeated by replacing the UDP-GlcNAc ligand with the superimposed UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc from the LpxA/product complex structure in generating the matching spheres (Fig. 3.8A, D). The protein from the LpxA/substrate structure was used again for the receptor file since it possessed the highest resolution. The leadlike and substrate subsets of the ZINC database were screened and the 1000 highest scoring compounds from both rounds were visually inspected for complementarity. Again, few compounds (<10%) possessed hydrophobic scaffolds that entered the hydrophobic cleft, since most compounds bound to the same area as UDP-GlcNAc binds, and extended out of the pocket. This was not surprising, since very few compounds would possess a scaffold resembling the long acyl chain used to generate the matching spheres (Fig. 3.8D). Consequently, a third round of virtual screening experiments was performed utilizing a strategy to enrich the frequency of compounds that possessed hydrophobic scaffolds that would better complement the hydrophobic cleft that normally fits the acyl chain. This strategy included creating enhanced matching spheres by using an ensemble of compounds previously identified from our virtual screens. The compounds selected possessed apolar scaffolds that complemented the acyl-chain binding site.
From the ensemble, atoms that were not necessary were truncated, and atoms that were complementing the cavity were retained (Fig. 3.8B). From this ensemble, matching spheres were generated that would bias enrichment of compounds with increased hydrophobic scaffolds (Fig. 3.8C, D, E). Next the leadlike and substrate subsets of the ZINC database were screened and the 1000 highest scoring compounds from both rounds were visually inspected for complementarity. As expected, there was indeed enrichment of more desirable compounds with scaffolds that exploited the hydrophobic cleft (<80%). There was also very high enrichment of negatively charged compounds, which can be expected since that site has many positively charged residues. From these results, 10 compounds were purchased for testing through X-ray crystallography.

Figure 3.8. Strategies in creating matching spheres for *P. aeruginosa* LpxA virtual screening. A. UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc fitted into the receptor site from the LpxA/UDP-GlcNAc complex structure through superimposition. B. Ensemble of docked compounds with desirable hydrophobic scaffolds retained, while atoms that would favor enrichment of undesirably characteristics were truncated. C. Superimposition of UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc and the contrived ensemble shows how the “ligands” exploit different characteristics of the site. D. Matching spheres generated by UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc (purple). E. Matching spheres generated by contrived ensemble (grey). F. Superimposition of matching spheres generated by UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc (purple) and the contrived ensemble (grey) shows how the matching spheres generated from the later favors the hydrophobic pocket.
**Hit screening through X-ray crystallography**

*(Compounds in this section cannot be presented due to disclosure issues. Therefore, single colored spheres will take the position of both polar and apolar atoms in showing complimentarily between compounds and electron density map)*

X-ray crystallography was used to screen for potential binders selected from the virtual screens especially since soaking experiments had previously worked suggesting that the crystals were suitable for complex structure determination by soaking methods. The 10 compounds selected from the virtual screens and purchased were dissolved in crystallization buffer, and crystals were then soaked in the solution overnight. Resulting electron densities confirmed the presence of two compounds in the active site. However, these crystals diffracted to ~3 Angstrom resolution, and the unbiased Fo-Fc was slightly ambiguous making the precise position of the ligands difficult to accurately predict (Fig. 3.9A, B). Nonetheless, the presence of the electron density was in the predicted site, normally occupied by the hydrocarbon chain, conforming that our biased structure based virtual screen approach was successful in predicting compounds that bound to the apolar site.

![Figure 3.9. Unbiased Fo-Fc map suggesting presence of potential binders. A-B. Fo-Fc maps suggesting binding poses of potential compounds that complement the hydrophobic pocket normally occupied by the acyl chain of UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc.](image)
Conclusion

Due to its wide range of antibiotic resistance mechanisms, *P. aeruginosa* has recently entered the Superbug category, making it one of the most feared nosocomial infections. Consequently, research has focused on identifying new druggable targets, and LpxA has become one of them. But even though seven ortholog LpxA structures had been solved, only two of which were in complex with the glucosamine ligands, the *P. aeruginosa* LpxA structure was still lacking. Due to high sequence similarity between orthologs, it could have been safely presumed that *P. aeruginosa* LpxA would contain similar domains and form similar trimers as its orthologs. However, active site configuration would still be hard to accurately predict, since there are residues in the active site that are not conserved across different orthologs. For rational drug design efforts, it is necessary to have precise structural characterization of the active site residues, and detailed information on how they interact with the ligands. Our X-ray structures have thus provided extensive information on the *P. aeruginosa* LpxA enzyme, including structural analysis of the different domains, but most importantly, precise structural characterization of binding sites, including the catalytic residues, and the “hydrocarbon ruler”.

The mode of catalysis appears to be the conserved between orthologs. In *P. aeruginosa* LpxA, conserved residues His118 and His140 aid in anchoring the substrate while His121 aids in catalysis. However, the orientation for the proposed oxyanion hole observed in ortholog LpxA structures was not observed for Gly139 in our *P. aeruginosa* Lpxa structures. It has been previously proposed that the backbone N of this residue stabilizes the transition state by promoting deacylation of the ACP protein, and acylation of the UDP-GlcNAc. However, in order for that to occur, the amide linker should be positioned in a way where the carbonyl from the
preceding alanine residue is orientated away from the active site so that the backbone N of Gly is
flipped towards the active site. This favorable glycine conformation has not only been observed
in every other LpxA ortholog structure, but has even been observed in every LpxD structure
solved to date. Our results therefore suggest that we might have observed the alternative
conformation due to the inherent flexibility of the Gly139 residue, and that the favorable
conformation may be adopted during catalysis, especially since our purified \textit{P. aeruginosa} LpxA
has been shown to be biologically active in our activity assays (data not shown).

In Lipid A biosynthesis, precise incorporation of specific hydrocarbon length acyl chains
is highly conserved within a species, but also divergent between different species. “Hydrocarbon
rulers” in LpxA were first identified by Wyckoff et al in \textit{E. coli} and \textit{P. aeruginosa}.\textsuperscript{88} These are
amino acids that act as precise measuring tools, allowing the incorporation of hydrocarbon
chains of very specific lengths; longer chains are disallowed, while shorter chains do not provide
sufficient hydrophobic interactions and therefore lose significant affinity. In \textit{E. coli} LpxA,
His191 acts as the “hydrocarbon ruler”, and in \textit{L. interrogans} LpxA, it is Lys171.

Previous mutational studies had shown how reciprocal mutations can completely reverse
this selectivity between \textit{E. coli} and \textit{P. aeruginosa}. The \textit{E. coli} LpxA G173M mutant lost its
affinity for the 14C-long product while gaining activity for the 10C-long product. Conversely, it
was also shown that the M169G mutation in \textit{P. aeruginosa} LpxA switched its specificity with the
14C-long product.\textsuperscript{88} In \textit{E. coli} LpxA, Gly173 lies in the same position as Met169 in \textit{P.
aeruginosa} LpxA. Our structure now provides detailed information on the function of Met169
and supports all previous findings. Superimposition of the \textit{P. aeruginosa} LpxA with \textit{E. coli} and
\textit{L. interrogans} LpxA complexes makes it obvious that 14 and 12 hydrocarbon-long chains would
clash with Met169 in the \textit{P. aeruginosa} Lpxa (Fig. 3.5F, G). Interestingly, the “hydrocarbon
ruler” Met169 in *P. aeruginosa* LpxA is provided by one monomer (chain A), while the “hydrocarbon ruler” His191 and Lys171 from the *E. coli* and *L. interrogans* LpxAs respectively, is provided by the other monomer (chain B) in the dimer interface. Finally, since *P. aeruginosa* LpxA has a Pro187 in the place of the *E. coli* LpxA His191, no steric hindrance would be expected at that site. It would therefore be interesting to test what is the largest hydrocarbon chain that can fit in the *P. aeruginosa* LpxA M169G mutant.

Increased hydrophobic contacts in the acyl chain site and Met169 may help explain why the UDP-GlcNAc was not seen in certain sites in the UDP-GlcNAc complex. The UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc product, due to its increased hydrophobic nature has increased apolar interactions with the hydrocarbon binding cavity. This possibly favors binding to all possible sites due to increased affinity. The conformational changes induced upon binding also helps explain why resolution of these crystals was compromised.

Ligands were found to bind at a secondary site bellow the active site (Fig. 3.4C). Even though this binding could be the result of crystal packing, the ligands interact and possess many hydrogen bonds with residues from both monomers. Cumulatively, these interactions are fewer than the interactions observed at the biologically relevant active site, but are still plentiful enough to possibly contribute decent affinity *in vivo*. Consequently, it raises the question of whether this could be a biologically relevant secondary site, which could possess a reduced affinity but serve a biological function, such as homing ligands close to the active site and make catalysis more efficient.

These structures not only provide information on the protein as a whole and on the function of important residues, but they also provide detailed information on the conformations of residues in the *P. aeruginosa* LpxA active site, both in the apo and complex forms. This
information is invaluable in employing a structure based drug discovery approach that can lead to the discovery of new inhibitors against \textit{P. aeruginosa} LpxA. Both the highly basic nature of the exposed active site but also the hydrophobic nature of the hydrocarbon chain cavity can be exploited simultaneously in designing molecules that will compete with the acylation of UDP-GlcNAc. Consequently, our virtual screen studies were performed utilizing that precise information. Through the use and manipulation of the binding site environment, a successful strategy led to the identification of two possible binders. However, higher resolution data is still needed to accurately identify the binding pose of these compounds, information that will greatly facilitate a hit-to-lead approach in searching for additional scaffolds that can interact with unutilized hotspots. Also, additional virtual screening experiments through the manipulation of parameters such as partial charge modification of charged residues and correct re-orientation of receptor hydrogen atoms can help identify additional binders. Finally, considering the restricted depth of the hydrocarbon cavity orchestrated by Met169, the unique composition of certain active side residues, and our identification of two possible binders so far, the design of potent molecules specific to the \textit{P. aeruginosa} strain is indeed promising.

**Experimental Procedures**

**Materials**

All chemicals and chromatography supplies were purchased from Fisher Scientific. UDP-GlcNAc was purchased from Sigma-Aldrich. Crystal screens were purchased from Qiagen.
Purification of recombinant LpxA

The plasmid pET280 containing the His-tagged *P. aeruginosa* LpxA sequence was transformed into Rosetta (DE3) cells. The cells were incubated in LB media supplemented with 35 μg/ml chloramphenicol and 50 μg/ml kanamycin at 37°C overnight. The overnight culture was then diluted into 1L LB media containing 35 μg/ml chloramphenicol and 50 μg/ml kanamycin at 1:500 and incubated at 37°C until the OD$_{600}$ reached 0.6-0.8. The protein expression was initiated with 0.5 mM IPTG and incubation continued at 37°C for an additional 4 hours. The cells were harvested by centrifugation at 5,000g for 10 min. The pellet was resuspended in buffer A (20 mM Tris-HCl, pH 8.4, 250 mM NaCl, 20 mM imidazole, and 10 % glycerol). The cells were disrupted by sonication followed by centrifugation at 35,000g for 40 min to remove debris. The supernatant was then loaded to a HisTrap affinity column. The protein was eluted with a linear gradient of imidazole. The fractions containing the tagged LpxA were pooled and concentrated. The sample was loaded to a HiLoad 16/60 Superdex 75 column for further purification in Thrombin cleavage buffer (20 mM Tris pH 8.4, 150 mM NaCl and 10 % glycerol). The peak fractions containing the His-tagged LpxA were pooled, and the concentration of the protein was determined by OD$_{280}$. The Thrombin protease was added at a ratio of one unit per mg of protein. After overnight incubation at room temperature, the samples were then loaded to a HisTrap column to remove any His-tagged protein contamination. The flow-through was collected and concentrated, followed by gel filtration purification. The protein eluted at a peak consistent with the size of the trimeric form (~84 kDa). The untagged LpxA was stored at -80°C at 14.3 mg/ml concentration in a buffer containing 20 mM potassium phosphate (pH 8.6) and 250 mM NaCl. The purity of the protein was determined by SDS-PAGE at >95%.
LpxA crystallization

Qiagen crystallization screens JCSG suites I-IV, AmSO₄, MPD, Core I & II, were screened using the Phoenix nanodispenser, and 0.2 and 0.4 µl aliquots of protein solution (14.3 mg/ml) were used to search for crystallization conditions. *P. aeruginosa* LpxA readily crystallized in many conditions, producing cuboidal crystals that did not diffract X-rays. Crystals with an “almond”-like morphology emerged in 0.1M imidazole pH 8.0, 20% (w/v) PEG1,000, and 0.2M Ca acetate, which were of diffractable quality. The diffracting crystals appeared within 2 to 4 days and measured around 0.03-0.04 mm in diameter.

The apo crystal was soaked in crystallization buffer containing 25% glycerol and cryo-cooled in liquid nitrogen. The UDP-GlcNAc and UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complexes were obtained by transferring apo crystals in crystallization solution containing 600mM of UDP-GlcNAc and 200mM of UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc respectively. The ten compounds selected from the virtual screens were dissolved in crystallization buffer at a 200mM concentration, and then apo crystals were transferred to the compound-containing drops. All crystals were left in the solution overnight, and were briefly soaked in crystallization buffer containing 25% glycerol and ligand and then immediately cryo-cooled in liquid nitrogen.

Data collection and structure determination

X-ray diffraction data for the apo and UDP-GlcNAc complex LPXA crystals was collected at the 8.3.1 beamline at the Advanced Light Source (ALS) at the Lawrence Berkeley National Laboratory. X-ray diffraction data for the UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc complex was collected at the GM/CA beamline at the Advanced Photon Source (APS) at
Argonne National Laboratory. For all three datasets, processing and scaling was done through HKL2000. The apo structure was solved via molecular replacement in MolRep (CCP4 suite) using a homology model constructed through Modeller, which was based on the *E. coli* LpxA structure (PDB ID: 2AQ9). The complex structures were subsequently solved via molecular replacement using the apo structure as a template. Ambiguous regions in electron density were removed and rebuilt using the program COOT.

**Molecular docking**

Virtual screens were performed using the program DOCK 3.5.54. The highest resolution structure obtained (LpxA/UDP-GlcNAc complex, 2.05 Angstrom) was used as the receptor file. To prepare receptor file for docking, all ions and waters were removed. UDP-GlcNAc, UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc, and a superimposed and modified ensemble of docked compounds were used in generation of the matching spheres through the program startdockblaster5c and a default total of 60 atom-derived spheres was used each time. Pre-generated ligand conformations were sampled in each binding pocket and scored when screening the leadlike and fragment subset of the ZINC database. Ligand partial atomic charges and disolvation were calculated by AMSOL. The ligand and receptor bin size/overlap were set at the default setting of 0.4/0.3 and 0.4/0.3 respectively. The energy scoring grids were calculated by summing the energies calculated by DISTMAP (excluded volume grid), CHEMGRID (van der Waals potential), and DelPhi (desolvation).
Chapter 4

Virtual Screening Against the PDZ Domain of Tiam1

Overview

Tiam1 (T-cell lymphoma invasion and metastasis gene 1) is a guanine exchange factor that specifically binds and activates Rac1, a Rho-family GTPase. Aberrant activation of Rac1 can lead to various forms of tumor progression. The Tiam1 PDZ domain regulates Tiam1 activity through an interaction with another protein, syndecan1. In this study, we have successfully applied a hybrid in silico/NMR screening strategy to identify four compounds that bind with low affinities to a binding site on the PDZ domain of Tiam1, where the C-terminus of syndecan1 normally binds.

Introduction

Rho-family GTPases are proteins that are highly conserved in eukaryotes, and control a diverse set of cellular events such as actin cytoskeletal rearrangement, cell adhesion, and cell cycle progression. Rac1 GTPase is such a protein. GEF (guanine nucleotide exchange factor) and GAP (GTPase-activating protein) proteins regulate the activity of GTPases through the exchange between the GTP and GDP bound states. A GEF protein that specifically binds and
activates Rac1 is Tiam1 (T-cell lymphoma and metastasis 1).\textsuperscript{91} The catalytic domain of Tiam1, which catalyzes the exchange between the GTP and GDP bound states, is a DH-PH (Dbl homology - pleckstrin homology) domain combination, while the remaining domains work in regulating the protein through additional protein-protein interactions.\textsuperscript{92} One of these regulatory domains is a PDZ domain, a small globular domain that mediates assembly of proteins into a larger complex by recognition and binding of C-terminal residues in their binding partners.\textsuperscript{93} Tiam1 is also homologous to Tiam2 and contains similar domains, including similar PDZ domains.\textsuperscript{91}

Studies have shown that the signaling between Tiam1 and Rac1 is associated with another protein, syndecan1, via the Tiam1 PDZ domain, which binds a C-terminal peptide of syndecan1.\textsuperscript{94} Syndecans are cell surface heparin sulfate proteoglycans involved in cell-cell junctions and cell adhesion.\textsuperscript{95} It has since been suggested that signal interference between Tiam1 and syndecan1 can be beneficial in combating epithelia-to-mesenchymal transition in the progression of tumors, mainly by perturbing cell polarity, in which Tiam1 and syndecan1 have important roles.\textsuperscript{96, 97} This makes the interaction between the PDZ domain of Tiam1 and syndecan1 an appealing target for structure-based drug discovery. Targeting the PDZ domain of Tiam1 could provide an alternative strategy in the treatment of cancer, especially considering the adverse effects associated with inhibitors that target Rac1 or the catalytic domain of GEFs.\textsuperscript{98} Similar approaches have previously proven to be successful. For example, studies have shown that a C-terminal NMDA Receptor peptide can disrupt the interaction between PSD-95 PDZ2 and NMDAR and reduce stroke-associated damage in ischemic neurons.\textsuperscript{99} Furthermore, optimization of the C-terminal NMDA Receptor peptide led to improved potency and efficacy in subsequent studies.\textsuperscript{100-102}
However, peptide inhibitors have limitations, such as metabolic instability leading to short half-lives, and unfavorable physiochemical properties that make it challenging for them to pass through the blood-brain barrier and cell membranes.\textsuperscript{103} For these reasons, there is a need for small molecule inhibitors that can overcome the limitations that come with using peptide mimetics. Additionally, small molecule inhibitors can be used as chemical probes and help elucidate details of Tiam1 signaling through its PDZ domain.

As structures of PDZ domains complexed with peptides become available in the PDB, a structure-based virtual screening approach is becoming an attractive tool in identifying small molecule inhibitors that can disrupt signaling. Virtual screening methods have previously been successful in identifying small molecule inhibitors of the PDZ domain of Dvl (Dishevelled) protein, which is involved in Wnt signaling.\textsuperscript{104, 105} In this study, we have successfully applied a structure-based virtual screening approach in identifying four ligands verified by 2D NMR to bind to the PDB domain of Tiam1, two of which are specific to Tiam1 and not Tiam2. Using a hit-to-lead approach, these ligands can be used to develop specific and potent inhibitors of the Tiam1 PDZ domain.

**Results and discussion**

**Virtual screening**

An X-ray crystallographic structure of the Tiam1 PDZ domain has been previously solved in complex with a SDC1 (TKQEEFYA) (PDB ID: 4GVD) peptide (Fig. 4.1A). The SDC1
peptide belongs to the C-terminus of syndecan1 and has been shown to mediate the interaction with Tiam1 by binding to the PDZ domain. The crystal structure therefore identified the specific residues that mediated these interactions. In the Tiam1 PDZ/SDC1 structure, the Ala8 of SDC1 is interacting with the backbone of Tyr858, Phe860, and Ser861, but also with the apolar side chains of Leu915 and Leu920 (Fig. 4.1A). The backbone of the SDC1 Tyr7 appears to interact with the Leu862 backbone. The Phe6 of SDC1 is interacting with Leu911 and Lys912 through apolar interactions (Fig. 4.1A). The remaining peptide appears to possess fewer interactions since it is more exposed to the solvent (Fig. 4.1A).

Therefore, our structure based drug discovery approach was focused in the pocket bound by the three last peptides (FYA) (Fig. 4.1B-C). The program DOCK3.5.54 was used to screen the fragment and lead-like subsets of the ZINC database in various docking trials. The matching spheres were created in the area of the binding site that corresponded to the three residues from the SDC1 peptide (FYA) (Fig. 4.1B-C). Compounds from the docking studies were visually inspected and 42 compounds were chosen as potential binders. From these compounds nine were purchased and tested by 2D NMR.

2D NMR screening

Note to reader: Members of the Fuentes lab at the University of Iowa performed the following NMR experiments.

Nine compounds were purchased from the ZINC database and tested via NMR (ZINC IDs: 18203043, 20057225, 18276277, 13942845, 19093798, 18207615, 01648367, 01639663, and 01502107). $^{15}$N- $^1$H HSQC spectra of the labeled Tiam1 PDZ domain were measured at
various concentrations to identify compounds that induced perturbations of PDZ domain residues, which would indicate binding to the PDZ domain. Surprisingly, out of the nine compounds, four induced chemical shift perturbations (ZINC IDs: 18203043, 20057225, 18276277, 13942845) (Fig. 4.2). The chemical shift perturbations were then mapped onto the Tiam1 PDZ structure and highlighted an area surrounding the pocket, suggesting that the compounds did indeed bind to the pocket in a conformation resembling the docking pose predictions (Fig. 4.2). For example, ZINC18203043 caused chemical shift perturbations mapped at binding cleft residues, including Thr857, alpha helix residues Phe914 and Lys912, and beta strand residues Phe860, Ser861, Leu862 and Ser863. The other three compounds (ZINC IDs: 20057225, 18276277, 13942845) also perturbed residues associated with the same site in addition to Thr853, Phe860, and Leu920 (Fig. 4.2).

Figure 4.1. Syndecan1 binding site on Tiam1 PDZ domain. A. X-ray crystallographic structure of Tiam1 PDZ domain complexed to TKQEEFYA peptide identifies polar and apolar interactions between the protein and peptide. B. Shorter peptide (FYA) used to create the docking environment in order to bias ligands that will occupy the deeper hydrophobic pocket. C. Matching spheres created in the deeper hydrophobic cleft of the Tiam1 PDZ domain.
Affinities were measured through non-linear fitting of compound concentration. The compounds appeared to bind weakly; ZINC18203043 and ZINC20057225 appeared to bind the tightest with \( K_d \) of 1.344 mM and 0.984 mM, respectively. ZINC18276277, and ZINC13942845 bound with affinities of 1.593 mM and 2.136 mM, respectively (Fig. 4.3).

To test specificity, NMR titration experiments were performed against the PDZ domain of Tiam2, which is highly homologous to the Tiam1 PDZ domain. Two of the four compounds (ZINC IDs: 18203043, 20057225) bound with even lower affinities (2.420 mM and 1.960 mM, respectively), while mapping of chemical shift perturbations indicated binding to the same pocket (data not shown). The affinities of these two compounds are two-fold less that what was observed for the Tiam1 PDZ domain. This suggests preference of these compounds for Tiam1 PDZ over the Tiam2 PDZ. The remaining two compounds did not show binding to Tiam2 PDZ, suggesting specificity for only Tiam1 PDZ.

**Conclusion**

Several steps of cancer progression, including proliferation, invasion, and even metastasis, appear to be influenced by abnormal signaling of the Rac1 protein.\(^\text{106}\) In addition, abnormal signaling has been traced back to abnormal activity of GEF proteins, consequently making GEF proteins attractive drug targets in the fight against cancer.\(^\text{107}\) Tiam1, a GEF protein for Rac1, has been shown to be overexpressed in many cancers including prostate cancer, breast cancer, colon cell carcinoma, and melanoma.\(^\text{97,108-110}\) Consequently, Tiam1 has been selected as an important and appealing target for drug discovery.
Figure 4.2. ZINC compounds that bind to Tiam1 PDZ domain. Four compounds chosen through virtual screening from the ZINC database induced chemical shift perturbations to the Tiam1 PDZ domain (ZINC IDs: 18203043, 20057225, 18276277, 13942845). Mapping of chemical perturbations identifies residues associated with the peptide-binding cleft, suggesting compounds bind similarly to the docking prediction.
The Tiam1 PDZ domain, which interacts with Syndecan1, has recently also become an appealing drug target. The interaction between syndecan1 and the Tiam1 PDZ domain mediates Rac1-Tiam1 signaling by subsequent activation of Rac1. Syndecan1, a cell surface sulfate proteoglycan, has been shown to be involved in various types of cancer progression. Inhibitors that target communication between Tiam1 and Syndecan1 may therefore contribute to cancer therapy but also help study signaling of these proteins in more detail.

The Tiam1 PDZ domain interacts with Syndecan1 through a protein-protein interface in which a C-terminal TKQEEFYA peptide of Syndecan1 binds to the narrow cleft formed at the surface of the Tiam1 PDZ domain. The crystallographic complex structure of Tiam1 bound to TKQEEFYA, has allowed us to use a structure based virtual screening approach that could help identify binders of the Tiam1 PDZ domain. Binders that can be optimized and developed into

102
inhibitors that interfere with the cross talk between Tiam1 and syndecan1, and influence the Tiam1 mediated activation of Rac1, may slow down the progression of various types of cancers.

In this study we have successfully identified four new compounds, purchased from the ZINC database of compounds, which bind to the Tiam1 PDZ domain with low affinities. Two of the compounds show specific binding to the Tiam1 PDZ domain and not to the Tiam2 PDZ domain. Furthermore, an additional two compounds show preference for the Tiam1 PDZ domain over the Tiam2 PDZ domain. These results suggest that specific inhibitors of the Tiam1 PDZ domain are indeed possible.

Future SAR analysis will help identify the important functional groups that mediate the interaction between each compound and the Tiam1 PDZ domain SDC1-binding site. Then, a hit-to-lead approach will help improve potency and efficacy, combined with the use of X-ray crystallography to determine complex structures of the Tiam1 PDZ domain with lead inhibitors. If successful, potent compounds against this new target may provide a new strategy in the treatment of cancer but also help acquire a better understanding of the signaling axis between Rac1, Tiam1, and syndecan1.

**Experimental procedures**

**Virtual screening**

Virtual screening against the binding pocket of the Tiam1 PDZ domain was performed using DOCK 3.5.54. The structure of the Tiam1 PDZ domain with the model TKQEEFYA
peptide was used as a template in the docking study. To prepare the structure for docking, all
waters were removed.

For the Tiam1 PDZ/SDC1 complex structure, residues Y6, A8 and part of Y7 of the ligand
were used to generate the matching spheres. A total of 45 atom-derived spheres were generated
and used. Three sets of docking rounds were performed using: 1) standard atomic partial
charges; 2) increased partial charges (by +0.4 or -0.4) for polar atoms of binding site residues
Y858, G859, F860, S861, and L862 to favor electrostatic interactions with these residues; and 3)
increased partial charges for the aforementioned residues plus L915.

The fragment and lead-like subset of the ZINC small-molecule database were docked
into the active site of the Tiam1 structure. The top 1000 compounds from each round were
visually inspected and a final 42 compounds were prioritized for testing.
References


Appendix A

X-ray Crystallography Data Collection and Refinement Statistics

Table 2.2. CXCL12/BWA3111 complex crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2_1_2_1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>36.93, 57.71, 72.53</td>
</tr>
<tr>
<td>α, β, γ, (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.5418</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.00-1.90</td>
</tr>
<tr>
<td>I/σI</td>
<td>21.77(3.25)*</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.8%</td>
</tr>
<tr>
<td>R_merge</td>
<td>0.061 (0.626)*</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.89 (6.85)*</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>Number of reflections used</td>
<td>12099</td>
</tr>
<tr>
<td>R-factor</td>
<td>0.1793</td>
</tr>
<tr>
<td>R_free</td>
<td>0.2194</td>
</tr>
<tr>
<td>Number of atoms</td>
<td>1194</td>
</tr>
<tr>
<td>Protein</td>
<td>1087</td>
</tr>
<tr>
<td>Ligand</td>
<td>21</td>
</tr>
<tr>
<td>Water</td>
<td>41</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>59.3</td>
</tr>
<tr>
<td>Ligand</td>
<td>78.9</td>
</tr>
<tr>
<td>Water</td>
<td>55.0</td>
</tr>
<tr>
<td>Root mean square deviations</td>
<td></td>
</tr>
<tr>
<td>Bond Lengths (Å)</td>
<td>0.0152</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.8017</td>
</tr>
</tbody>
</table>

* Numbers in parentheses are for the highest resolution shell.
Table 3.1. LpxA apo crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>80.09, 82.97, 222.04</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.1271</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>111.20-2.17</td>
</tr>
<tr>
<td>I/σI</td>
<td>13.16 (1.86)*</td>
</tr>
<tr>
<td>Completeness</td>
<td>98.9%</td>
</tr>
<tr>
<td>R_{merge}</td>
<td>0.060 (0.610)*</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.4 (4.3)*</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>Number of reflections used</td>
<td>70389</td>
</tr>
<tr>
<td>R_{factor}</td>
<td>0.1917</td>
</tr>
<tr>
<td>R_{free}</td>
<td>0.2371</td>
</tr>
<tr>
<td>Number of atoms</td>
<td>23908</td>
</tr>
<tr>
<td>Root mean square deviations</td>
<td></td>
</tr>
<tr>
<td>Bond Lengths (Å)</td>
<td>0.0138</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.6155</td>
</tr>
</tbody>
</table>

* Numbers in parentheses are for the highest resolution shell.
Table 3.2. LpxA/UDP-GlcNAc crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2_12_1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>80.04, 83.56, 222.43</td>
</tr>
<tr>
<td>α, β, γ, (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.1271</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>111.21-2.06</td>
</tr>
<tr>
<td>I/σI</td>
<td>12.26(1.78)*</td>
</tr>
<tr>
<td>Completeness</td>
<td>98.6%</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>0.070 (0.898) *</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.2 (6.3) *</td>
</tr>
</tbody>
</table>

| Refinement            |                   |
| Number of reflections used | 87335            |
| R<sub>factor</sub>     | 0.1938            |
| R<sub>free</sub>       | 0.2284            |
| Number of atoms       | 24446             |

<table>
<thead>
<tr>
<th>Root mean square deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond Lengths (Å)</td>
</tr>
<tr>
<td>Bond angles (°)</td>
</tr>
</tbody>
</table>

* Numbers in parentheses are for the highest resolution shell.
Table 3.3. UDP-3-O-(R-3-hydroxydecanoyl)-GlcNAc crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>80.04, 83.56, 222.43</td>
</tr>
<tr>
<td>α, β, γ, (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.066</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.00-2.52</td>
</tr>
<tr>
<td>I/σI</td>
<td>13.34(2.01)*</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.8%</td>
</tr>
<tr>
<td>R_merge</td>
<td>0.068 (0.789)*</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.3 (4.5) *</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>Number of reflections used</td>
<td>52497</td>
</tr>
<tr>
<td>R_factor</td>
<td>0.2101</td>
</tr>
<tr>
<td>R_free</td>
<td>0.2495</td>
</tr>
<tr>
<td>Number of atoms</td>
<td>23759</td>
</tr>
<tr>
<td>Root mean square deviations</td>
<td></td>
</tr>
<tr>
<td>Bond Lengths (Å)</td>
<td>0.0105</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.5524</td>
</tr>
</tbody>
</table>
* Numbers in parentheses are for the highest resolution shell.
Appendix B

Permission to Use Manuscript: Smith 2014, J. Med. Chem

Title: Structural Analysis of a Novel Small Molecule Ligand Bound to the CXCL12 Chemokine
Author: Emmanuel W. Smith, Yan Liu, Anthony E. Getschman, et al
Publication: Journal of Medicinal Chemistry
Publisher: American Chemical Society
Date: Nov 1, 2014

Permission is granted for your request in both print and electronic formats, and translations.
If figures and/or tables were requested, they may be adapted or used in part.
Please print this page for your records and send a copy of it to your publisher/graduate school.
Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.