Do All Roads Really Lead to Rome? Learnings from Comparative Analysis using SPR, NMR, & X-Ray Crystallography to Optimize Fragment Screening in Drug Discovery

Dissertation
zur

Erlangung des Doktorgrades
Der Naturwissenschaften
(Dr. rer. nat.)

dem

Fachbereich Pharmazie der Philipps-Universität Marburg vorgelegt von

> Engi Hassaan aus Kairo

Maburg/Lahn 2019

Erstgutachter: Prof. Dr. Gerhard Klebe

Institute für Pharmazeutische Chemie

Philipps-Universität Marburg

Zweitgutachter:

Dr. Wolfgang JahnkeChemical Biology & Therapeutics
Novartis Pharma AG

Eingereicht am **27.05.2019**

Tag der mündlichen Prüfung am 06.09.2019

Hochschulkennziffer: 1180

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. Gerhard Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von September 2015 bis September 2019 durchgeführt.

To my family For being my constant support, even thousands of kilometers away.

"What you seek is seeking you." - Rumi

Summary

There are several biophysical methods developed to rapidly identify weakly binding fragments to a target protein. X-ray crystallography provides structural information that is crucial for fragment optimization, however there are several criteria that must be met for a successful fragment screening including the production of soakable and well-diffracting crystals. Therefore, having a reliable cascade of screening methods to be used as pre-screens prior to labor-intensive X-ray crystallography would be extremely beneficial. This would allow the filtering of compounds as the screening progresses so that only the most promising hits remain. But which method should be the one to start the screening cascade? In this work, various sets of fragment libraries were screened against three different proteins; namely tRNA guanine transglycosylase (TGT) an important protein in Shigella, membrane associated protein peroxin 14 (PEX14) of *T. Brucei*, and endothiapepsin (EP), to investigate whether different screening methods will reveal similar collections of putative binders. The detailed comparative analysis of the findings obtained by the different methods is discussed in this thesis.

Shigellosis, an acute bacterial infection of the intestine, is caused by the gram-negative Shigella bacterium whose pathogenicity is reliant on virulence factors (VirF) required to invade epithelial cells. The expression of these VirF is modulated by TGT. Strategies developed to inhibit TGT include potent active-site inhibitors to block the binding of tRNA, thereby preventing the transcription of the virulence factors. Our 96-fragment library was screened against TGT using SPR, NMR, and X-ray crystallography, as described in Chapter 2. A total of 81 fragments were screened in SPR using a direct binding assay approach, revealing a hit rate of 12%. A total of 77 fragments were screened in NMR revealing a hit rate of 29%. High-resolution crystal structures were also collected for the entire fragment library by soaking, revealing a hit rate of 8%. Upon comparison of all discovered fragment hits no overlaps from all three methods were found. Several factors are responsible for this finding such as exclusion of fragments from individual screens due to technical reasons. In detail, four X-ray hits were excluded from the SPR and NMR screens, two SPR hits were discarded from the NMR screen, and five NMR hits were never subjected to the SPR screen. SPR and NMR are currently the most commonly applied primary fragment screening techniques, however, our results suggest that if they would have been applied as incipient methods of a screening cascade, they would have missed three binders discovered by a subsequently applied, more elaborate crystallographic screen. X-ray crystallography allows the detection of specific binders that may be too weak binders to be detected by SPR and even by NMR but can still provide valid structural information to support the search for appropriate starting points in lead discovery. Additionally, MD simulations of the apo wild type TGT have predicted the opening of a transient sub-pocket located above the guanine/preQ₁ pocket, which suggested a strategy to target this new binding site for the design of new inhibitors against TGT following a structure- based drug design concept which is also discussed in section 2.3.

The human African trypanosomiasis (HAT), also known as the sleeping sickness, is a vector-borne parasitic disease caused by *T. brucei* and transmitted to humans by bites of the tsetse fly. *T. brucei* lacks feedback allosteric regulation of early steps in glycolysis but compartmentalizes the relevant enzymes within organelles called glycosomes. PEX14, a peroxin protein essential for biogenesis of glycosomes, forms an important protein-protein interaction with PEX5, an import receptor that transports cytoplasmic glycosomal enzymes into the organelle. Disrupting the PEX14/PEX5 interaction leads to the accumulation of glycosomal enzymes in the cytosol, depletion of ATP, glucose toxicity, metabolic collapse and death of *T. brucei*. This disruption can be achieved through small molecules that bind to and block PEX14, preventing PEX5 binding. A previous NMR screening of a fragment library resulted in fragment hits that bind to the N-terminal domain (NTD) of *T. brucei* PEX14. In this project,

we attempted to validate these hits through X-ray crystallography by soaking, to allow visualization of the fragment interactions. The promising fragment hits would then be optimized into more potent lead compounds. Crystallization of the NTD PEX14 with a mutation in the first residue (E1W) revealed blocked binding pockets, as described in Chapter 3. The purpose of the added tryptophan was to render fluorescent properties to the short NTD construct which lacked fluorescent amino acids. However, this tryptophan was found to block the binding pockets of its neighboring crystal mates in the protein crystal, rendering a crystal form impossible to use for soaking. Attempts to find new crystal forms with free pockets were unsuccessful, as the small size of the protein and the hydrophobic nature of tryptophan rendered tightly packed protein crystals that block the binding pockets of neighboring crystal mates. Virtual Screening to discover novel ligands for co-crystallization revealed a ligand that aids the crystallization of the E1W PEX14 variant in the same space group but with a slightly different packing. This produced a crystal form that proved successful for fragment soaking as it enabled the binding of two additional fragment hits binding to further protein pockets. Additionally, the wild type form of PEX14 which lacks the tryptophan residue and thus has free binding pockets was crystallized. This enabled the soaking of a previously designed lead compound in different pockets of the PEX5 binding site. By obtaining a crystal structure of this complex at a resolution of 1.8 Å, the feasibility of using wild type PEX14 crystals for further fragment screening has been demonstrated.

Endothiapepsin is a member of the pepsin-like aspartic proteases responsible for the hydrolytic cleavage of peptide substrates. Owing to its high degree of similarity to other pharmacologically relevant aspartic proteases, it has served as the model enzyme for studying their mechanism and to discover first lead structures. In previous work done by other members from our group to identify and characterize endothiapepsin binders, X-ray crystallography was consulted as a primary fragment screening method and its hit identification potential was compared to several biochemical and biophysical screening methods. The fragment library screened was designed for general purposes and contained 361 entries. Comparison of the overlap in the hit rates of the different methods to that of X-ray crystallography revealed a low overlap, with the RDA having the highest overlap at 7% and MS having the lowest overlap at 1% followed by STD NMR at 3%. To understand the reason behind the low overlap, two of these screening techniques were prioritized for closer analysis as described in Chapter 4. The 71 X-ray detected fragment hits were selected and rescreened again with STD NMR under slightly different buffer conditions, in addition to WaterLOGSY NMR experiments. The second STD NMR screen detected almost double the amount of hits as the initial one, and the Water LOGSY screen had the highest correlation from the NMR methods to the X-ray hits at 69%. This comparative analysis also revealed the phenomena of active site fragment displacement by use of so-called reporter ligands and that non-deuterated water in STD NMR may lead to false negatives. The entire 361 fragment library was also screened with SPR using an inhibition in solution assay, adding another biophysical method for our comparative analysis to give us further insight of which conditions are crucial to maintain while transferring across different techniques. The resulting hit rate from SPR was 34%, correlating to an overlap of 11% with the X-ray hits - the highest correlation between screening methods reported by us thus far. Finally, we also studied fragment detection and cocktailing in crystallography in comparison to fragment cocktailing in NMR. From this we concluded that cocktailing in crystallography can also lead to false negatives due to fragment competitive behavior and can reveal a different binding mode for a given fragment compared to the adopted geometry found when soaked individually. As for NMR, despite the ability to detect competitive binding of fragments due to the temporary binding and unbinding events, the parallel binding and thus detection of fragments is not always guaranteed as seen in 20% of the fragments we screened, in addition to our observation that the detection of fragments in cocktail NMR may also depend on the comparison of the cocktail set they are a part of.

Zusammenfassung

Es gibt mehrere biophysikalische Methoden, um schwach bindende Fragmente an einem Zielprotein schnell zu identifizieren. Als eine Methode darunter liefert die Röntgenkristallographie strukturelle Informationen, die für die Optimierung von Fragmenten entscheidend sind. Dabei gibt es mehrere Kriterien, die für ein erfolgreiches Fragmentscreening erfüllt sein müssen, einschließlich der aufwändigen Herstellung von durchtränkbaren ("soakable") und gut streuenden Kristallen. Daher wäre es äußerst vorteilhaft, eine zuverlässige Kaskade von Screening-Methoden zu haben, die vor der arbeitsintensiven Röntgenkristallographie zur Vorselektion eingesetzt werden können. Diese würden das Aussortieren (oder Filtern) von Verbindungen im Laufe des Screenings ermöglichen, sodass nur die vielversprechendsten Treffer übrigblieben. Aber welche Methode sollte diejenige sein, mit der die Screening-Kaskade gestartet wird? In dieser Arbeit wurden verschiedene Varianten von Fragmentbibliotheken gegen drei verschiedene Proteine durchgemustert, und zwar tRNA-Guanin-Transglykosylase (TGT), ein wichtiges Zielprotein bei Shigellose, das membranassoziiertes Protein Peroxin 14 (PEX14) von T. Brucei und Endothiapepsin (EP), eine repräsentative Aspartylprotease, um zu untersuchen, ob verschiedene Screening-Methoden ähnliche Gruppen von vermeintlich bindenden Substanzen aufdecken können. Die detaillierte vergleichende Analyse der Ergebnisse der verschiedenen Methoden wird in dieser Arbeit diskutiert.

Die Shigellose, eine akute bakterielle Infektion des Darms, wird durch die Gram-negativen Shigella-Bakterien verursacht. Ihre Pathogenität hängt von Virulenzfaktoren (VirF) ab, deren ungestörte Funktion für die Invasion von Epithelzellen erforderlich ist. Die Expression dieser VirFs wird durch das Enzym TGT moduliert. Zu den Strategien, die zur Hemmung von TGT entwickelt wurden, gehören potente Inhibitoren des aktiven Zentrums, welche die Bindung von tRNA blockieren und so die Transkription der Virulenzfaktoren verhindern. Unsere 96-Fragment-Bibliothek wurde, wie in Kapitel 2 beschrieben, mittels SPR, NMR und Röntgenkristallographie auf TGT untersucht. Insgesamt 81 Fragmente wurden mit der SPR-Methode in einem "Direct Binding Assay"-Ansatz untersucht, was eine Trefferquote von 12% ergab. Insgesamt 77 Fragmente wurden mittels NMR untersucht, was zu einer Trefferquote von 29% führte. Hochauflösende Kristallstrukturen wurden auch für die gesamte Fragmentbibliothek durch Tränken der Kristalle mit den entsprechenden Fragmentlösungen ermittelt, was eine Trefferquote von 8% ergab. Beim Vergleich aller gefundenen Fragment-Treffer wurden keine übereinstimmenden Treffer für alle drei Methoden gefunden. Dafür sind mehrere Faktoren verantwortlich, wie z.B. der Ausschluss mancher Fragmente von einzelnen Screens aus technischen Gründen. Im Einzelnen wurden vier Röntgentreffer aus dem SPR- und NMR-Screening ausgeschlossen, zwei SPR-Treffer aus dem NMR-Screening verworfen und fünf NMR-Treffer wurden nie mit in das SPR-Screening einbezogen. SPR und NMR sind derzeit die am häufigsten angewandten primären Fragment-Screening-Techniken, aber unsere Ergebnisse deuten darauf hin, dass sie, wenn sie als einleitende Methoden einer Screening-Kaskade angewandt worden wären, drei potentiell bindende Substanzen übersehen hätten, die von einem später angewandten, aufwändigeren kristallographischen Filter entdeckt wurden. Die Röntgenkristallographie ermöglicht die Detektion spezifisch bindender Substanzen, die allerdings zu schwach binden, um durch SPR und NMR detektiert zu werden. Sie liefern aber dennoch valide Strukturinformationen, um geeigneten Ansatzpunkten für eine Leitstruktur-Entwicklung abzugeben. Darüber hinaus haben MD-Simulationen des apo-Wildtyp TGTs die Öffnung einer transienten Nebentasche oberhalb der Guanin/preQ1-Tasche vorhergesagt. Basierend darauf wurde eine Strategie zur Ausrichtung dieser neuen Bindungsstelle für das Design neuer Inhibitoren gegen TGT nach einem strukturbasierten Wirkstoffdesign-Konzept entwickelt, welches in Kapitel 2.3 diskutiert wird.

Die humane afrikanische Trypanosomiasis (HAT), auch bekannt als Schlafkrankheit, ist eine vektorübertragene parasitäre Krankheit, die durch T. brucei verursacht und durch Stiche der Tsetsefliege auf den Menschen übertragen wird. T. brucei fehlt es an einer rückgekoppelten allosterischen Regulation der frühen Schritte in der Glykolyse, sie verfügen aber über die relevanten Enzyme in abgetrennten Organellen, die Glykosome genannt werden. PEX14, ein Peroxinprotein, welches für die Biogenese der Glykosome essenziell ist, bildet eine wichtige Protein-Protein-Interaktion mit PEX5, einem Importrezeptor, der zytoplasmatische glykosomale Enzyme in das Organell transportiert. Die Störung der PEX14/PEX5-Interaktion führt zur Anhäufung von glykosomalen Enzymen im Zytosol, damit zum Abbau von ATP, zur Glukosetoxizität, zum metabolischen Zusammenbruch und am Ende zum Tod von T. brucei. Diese Unterbrechung der Interaktion kann durch kleine Moleküle erreicht werden, die an PEX14 binden, und so die Bindung von PEX5 an PEX14 verhindern. Ein initiales NMR-Screening einer Fragmentbibliothek führte zu Fragment-Treffern, die an die N-terminale Domäne (NTD) von T. brucei PEX14 binden. In der vorliegenden Arbeit wurde versucht, die NMR-Treffer durch Röntgenkristallographie durch Tränkexperimente zu validieren, um die räumlichen Wechselwirkungen der Fragmente sichtbar zu machen. Die vielversprechenden Fragment-Treffer könnten dann zu potenteren Leitsubstanzen optimiert werden. Die Kristallisation des NTD PEX14 mit einer mutierten Variante im ersten Rest (E1W) zeigte blockierte Bindetaschen, wie in Kapitel 3 beschrieben wird. Das hineinmutierte Tryptophans diente dem Zweck, das kurze NTD-Konstrukt mit fluoreszierende Eigenschaften zu versehen, denn dem nativen Protein fehlen fluoreszierende Aminosäuren. Es wurde jedoch festgestellt, dass dieses angefügte Tryptophan die Bindungstaschen seiner benachbarten Kristallpartner im Proteinkristall blockiert, was diese Kristallform für die Tränkungsversuche ungeeignet macht. Versuche, neue Kristallformen mit freien Taschen zu finden, blieben erfolglos, da die geringe Größe des Proteins und die hydrophobe Natur des Tryptophans dicht gepackte Proteinkristalle entstehen lässt, die wechselseitig die Bindungstaschen benachbarter Kristallpartner blockierten. Virtuelles Screening zur Entdeckung neuartiger Liganden für die Co-Kristallisation ergab einen Liganden, der die Kristallisation der E1W PEX14-Variante in eine leicht geänderte Packung unter Erhalt der Raumgruppensymmetrie überführt. Dadurch entstand eine Kristallform, die sich für die Tränkungsexperimente in Fragment-Lösungen geeignet erwies. Sie erlaubte die Bindung von zwei weiteren Fragment-Treffern in anderen Bindetaschen. Weiterhin konnte die Wildtypform von PEX14 kristallisiert werden, welcher der Tryptophanrest fehlt und welche somit in der Packung unbesetzte Bindungstaschen aufweist. Dies ermöglichte das Eindiffundieren eines bereits zuvor synthetisch dargestellten Leitstrukturvorschlags. Es wurde eine Kristallstruktur des Komplexes bei einer Auflösung von 1,8 Å erhalten, in der der Ligand mehrere Taschen der PEX5-Bindestelle besetzt. Dies unterstreicht die Machbarkeit von Wildtyp-PEX14-Kristallen und deren Eignung für ein weiteres Fragmentscreening.

Endothiapepsin ist ein Mitglied der pepsinähnlichen Aspartylproteasen, die für die hydrolytische Spaltung von Peptidsubstraten verantwortlich sind. Aufgrund seiner hohen Ähnlichkeit mit anderen pharmakologisch relevanten Aspartylproteasen dient es als Modellenzym für die Untersuchung ihres Mechanismus und das Auffinden erster Leitstrukturen. In früheren Arbeiten anderer Mitglieder unserer Arbeitsgruppe zur Identifizierung und Charakterisierung von Endothiapepsin-bindenden Fragmenten wurde die Röntgenkristallographie als primäre Screening-Methode herangezogen und ihr Potenzial zu identifizieren von Treffern mit mehreren biochemischen und biophysikalischen Screening-Methoden verglichen. Die untersuchte Fragmentbibliothek wurde für ein allgemeines Screening konzipiert und enthielt 361 Einträge. Der Vergleich der Trefferraten der verschiedenen Methoden mit denen der Röntgenkristallographie ergab eine nur geringe Überschneidung, wobei die RDA die höchste Überschneidung bei 7% und MS die niedrigste Überschneidung bei 1% aufweist, gefolgt von STD NMR mit 3%. Um die Gründe für die geringen Überschneidungen zu verstehen, wurden zwei dieser Screening-Techniken für eine genauere Analyse ausgewählt, wie in Kapital 4 beschrieben. Dazu wurden

die 71 röntgenkristallographisch detektierten Fragment-Treffer selektiert und mit STD-NMR unter leicht unterschiedlichen Pufferbedingungen sowie mit WaterLOGSY NMR-Experimenten erneut durchmustert. Dieser zweite STD NMR Screen erkannte fast die doppelte Anzahl von Treffern als die erste, und der Water LOGSY Screen hatte mit 69% die höchste Korrelation zwischen den NMR Methoden und den Vorschlägen aus der Kristallographie. Diese vergleichende Analyse zeigte auch, dass Fragmente durch den Einsatz so genannter Reporterliganden nicht, wie gewünscht, räumlicher verdrängt sondern teilweise verschoben werden und dass nicht-deuteriertes Wasser in STD NMR zu falschen Negativen führen kann. Die gesamte 361 Fragment-Bibliothek wurde mittels SPR unter Invertieren des Messprinzips (so genannter "inhibition solution assay") durchgemustert und als weitere biophysikalische Methode unserer vergleichenden Analyse hinzugefügt. Die erlaubt uns einen weiteren Einblick zu erkennen, welche Bedingungen beim Transfer zwischen verschiedenen Techniken unbedingt eingehalten werden müssen. Die resultierende Trefferrate aus SPR betrug 34%, was einer Überlappung von 11% mit den kristallographischen Treffern entspricht – somit der höchsten Korrelation zwischen den bisher von uns eingesetzten Screeningverfahren. Schließlich haben wir auch die Fragmenttreffer-Erkennung im Rahmen des so genannten "Cocktailing" für Ergebnisse mit der Kristallographie im Vergleich zu der NMR-Methode untersucht. Dabei ergab sich, dass das "Cocktailing" in der Kristallographie zu falschen negativen Treffern aufgrund der Kompetition von Fragmenten um den gleichen Bindeplatz führen kann oder abweichende Bindungsmodi für ein bestimmtes Fragment anzeigt werden als diese bei Einzelbestimmungen ermittelt wird. Was die NMR-Methode betrifft, so ist trotz der Fähigkeit, konkurrierende Bindungen von Fragmenten im dynamischen Prozess als Bindupngsereignisse nebeneinander zu beobachten, nicht immer gewährleistet. In ca. 20% der von uns untersuchten Beispiele lieferte diese Parallelbindung keine korrekten Hinweise. Weiterhin kann die Erkennung von Fragmenten in der NMR-Methode aus einem Fragment-"Cocktail" auch davon abhängen, wie der "Cocktail" zusammengesetzt war.

Table of Contents

Summary	I
Table of Contents	VII
Table of Figures	XI
Table of Tables	XV
Table of Equations	XVII
Table of Schemes	XIX
Abbreviations	XXI
Chapter 1 Introduction	1
1 Introduction	3
1.1 Fragment-Based Drug Design	5
1.2 Surface Plasmon Resonance	7
1.2.1 Concept of SPR	7
1.2.2 Direct Binding Assays	10
1.2.3 Inhibition in Solution Assay	10
1.3 Nuclear Magnetic Resonance	12
1.3.1 Concept of NMR	12
1.3.2 NMR Instrumentation	15
1.3.3 Ligand-Observed NMR (CPMG, STD NMR, Water LOGSY)	15
1.3.4 Protein-Observed NMR (SOFAST HMQC)	16
1.4 Crystallography	18
1.4.1 Concept of X-ray Crystallography	18
1.4.2 Methodology	19
1.4.3 Fragment Screening Using X-ray Crystallography	24
Chapter 2 tRNA-Guanine Transglycocylase	25
2.1 Introduction	27
2.1.1 Shigellosis	27
2.2 Fragment Screening	31
2.2.1 Fragment Screen by Surface Plasmon Resonance (SPR)	31
2.2.2 Fragment Screen by NMR	34
2.2.3 Fragment Screen by X-ray Crystallography	36
2.2.4 Discussion	45
2.2.5 Conclusions	47
2.3 Structure-based Drug Design	47
2.3.1 Definition	47

2.3.2 Background	47
2.3.3 Ligand Design	49
2.3.4 Affinity Determination	52
2.3.5 Crystal Structure Determinations	54
2.3.6 Expansion of the recognition pocket by a transient sub-pocket	58
2.3.7 Functional Role of the Transient Pocket and Outlook for Further Inhibito	r Design 60
2.4 Materials & Methods	61
2.4.1 Expression and Purification of the Z. mobilis TGT	61
2.4.2 Crystallization and Structure Refinement of Z. mobilis TGT	62
2.4.3 Applied Buffers for the Screening Experiments	63
2.4.4 SPR Conditions	64
2.4.5 NMR Conditions	65
2.4.6 MD Simulations and HSA Calculations.	66
Chapter 3 Peroxin 14	69
3.1 Introduction	71
3.1.1 Neglected Infectious Tropical Diseases	71
3.1.2 Membrane Associated Protein PEX14	73
3.2 Crystallization and Fragment Screening of Mutated PEX14	76
3.2.1 Expression Constructs	76
3.2.2 Crystallization	77
3.2.3 Fragment Screening	78
3.2.4 The Search for New Crystal Forms	80
3.2.5 Fragment Hit-to-Lead Generation	84
3.2.6 Soaking Co-Crystallized E1W PEX14	84
3.2.7 Discussion	93
3.3 Virtual Screening	94
3.3.1 Definition	94
3.3.2 Screening Methodology	94
3.3.3 Determining apparent K_D and IC_{50} using 2D NMR and AlphaScreen	98
3.3.4 Co-crystallization of Ligands	105
3.4 Crystallizing the Wild-Type	106
3.4.1 Crystallization Screening	106
3.4.2 Wild-Type Crystals Optimization	109
3.5 Wild Type with a Bound Ligand	114
3.5.1 Soaking of Wild Type Crystals	114
3.5.2 Wild Type PEX14 in Complex with a Lead Compound	114

3.6 Materials and Methods	117
3.6.1 Protein Expression and Purification	117
3.6.2 Crystallization	118
3.6.3 NMR Conditions	119
3.6.4 AlphaScreen-based Competition Assay Conditions	119
3.6.5 Trypanocidal Assay	120
Chapter 4 Endothiapepsin	121
4. 1 Introduction	123
4.1.1 Endothiapepsin	123
4.1.2 Characterization of Binding Using Orthogonal Biophysical Methods	124
4.2 Fragment Screening using Water LOGSY and STD NMR	125
4.2.1 Fragment Hit Rate with WaterLOGSY	126
4.2.2 Comparison of Fragment Hits and Hit Rate with and without Salt in STD NMR	127
4.2.3 Displacement of Fragments by Reporter Ligands and Vice Versa	131
4.2.4 Comparison of Light Water STD NMR and Heavy Water STD NMR	137
4.2.5 Discussion of Comparative Analysis	143
4.3 Fragment Screening Using Surface Plasmon Resonance	144
4.3.1 ISA Methodology	144
4.3.2 Fragment Hits	145
4.4 Comparative Analysis of Fragment Binding using X-ray Crystallography, SPR, and STE	
4.4.1 Discussion	147
4.5 Comparison of Cocktail Screening in X-Ray Crystallography vs NMR	151
4.5.1 Definition of Cocktail Screening	151
4.5.2 Results & Discussion	151
4.6 Materials and Methods	164
4.6.1 Purification of EP	164
4.6.2 Crystallization and Structure Refinement of EP	164
4.6.3 SPR Conditions	165
4.6.4 NMR conditions	166
Bibliography	169
Appendix	189
Acknowledgments	229
Conferences & Publications	231
Erklärung	233

Table of Contents

Table of Figures

Figure 1.1: HTS hits vs Fragment hits	5
Figure 1.2: The Kretschmann configuration	7
Figure 1.3: SPR sensorgram	8
Figure 1.4: MCK vs SCK	9
Figure 1.5: A sensorgram and corresponding steady state plot	10
Figure 1.6:Concept of DBA vs ISA	11
Figure 1.7:Atomic spins	12
Figure 1.8: FID and NMR spectrum	14
Figure 1.9: The setup of an NMR spectrometer	15
Figure 1.10: NMR spectra of relaxation-based method, STD NMR, and WaterLOGSY	16
Figure 1.11: A 2D NMR spectrum	17
Figure 1.12:Unit cell and crystal lattice	18
Figure 1.13: Bragg's law and the Ewald sphere	19
Figure 1.14: The phase diagram of protein crystallization	20
Figure 1.15: Hanging drop and sitting drop crystallization plates	20
Figure 1.16: The optical setup of an X-ray experiment.	21
Figure 1.17: A diffraction pattern recorded on a detector	22
Figure 1.18: An electron density map of a protein structure	23
Figure 2.1: Tertiary structure of TGT monomer and quaternary structure of TGT dimer	28
Figure 2.2: TGT dimer with a tRNA intermediate substrate bound to active site	
Figure 2.3: Substrates of TGT as they differ across different kingdoms	
Figure 2.4: Sequence alignment of TGT in different kingdoms	
Figure 2.5: Dose-response curve of J50	
Figure 2.6: CPMG and WaterLOGSY spectra of J50 binding to TGT.	
Figure 2.7: Fragments peak intensities difference after protein and inhibitor addition	
Figure 2.8: Overview of the spatial accommodation of the fragments detected in a crystallograph	
screen directly on protein crystals.	
Figure 2.9: Binding mode of fragment J41 (PDB: 5SW3) in the preQ ₁ pocket	
Figure 2.10: Binding mode of fragment J64 (PDB: 5N6F), guanine to the preQ ₁ pocket	
Figure 2.11: Binding mode of fragment J79 (PDB: 5UTJ) to the pre Q_1 pocket	
Figure 2.12: Binding mode of fragment J86 (PDB: 5V3C) to the preQ ₁ pocket	
Figure 2.13: Binding mode of fragment J72 (PDB: 5UTI) to the pre Q_1 pocket leading to the opening	
a sub-pocket next to the latter pocket	
Figure 2.14: Fragment J41 (PDB: 5SW3) binding mode to the surface of TGT	
Figure 2.15: Binding mode of fragment J14 (PDB: 6FSO) to a surface depression of TGT	
Figure 2.16: Binding mode of fragment J19 to the surface of TGT	
Figure 2.17: Binding mode of fragment J33 to the surface of TGT	
Figure 2.18: Event map generated by PAnDDA shows electron density for J33	
Figure 2.19: Overlap in hit rates between the different method and correlation between active-s	
and non-active-site binders as detected by the three methods	
Figure 2.20: The published structure of TGT in complex with J64 superimposed with one orientat	
of J50	
Figure 2.21: Schematic binding mode of lin-benzoguanine to TGT and crystallographically determ	
binding mode of a small fragment binding to TGT which opens a transient binding pocket	
	_

Figure 2.22: Crystallographically determined binding mode of ligands 2.1 and 2.5 in the recognition	
pocket of TGT, and superposition of 2.1 and 5.1	
Figure 2.23: Superposition of 2.5 and 5.1 in the recognition pocket of TGT, binding mode of 3.1, as	
superposition of 3.1 and 5.1	
Figure 2.24: Crystallographically determined binding mode of 4.1 in the recognition pockect of TG	
Figure 2.25: Crystallographically determined binding mode of 4.2 in the recognition pocket of TGT	· 58
Figure 2.26: Binding pose of 3.1 and and 4.2 with comparison of the spatial arrangement of	
computed hydration sites together with the crystallographically determined waters in TGT 4.1	59
Figure 3.1: Distribution of human African trypanosomiasis with incidences and risk for travelers	72
Figure 3.2: Role of glycosome, cytosol, and mitochondria in energy metabolism of blood stream fo	orm
of T. brucei	74
Figure 3.3: Structure of NTD of PEX14 bound with PEX5	74
Figure 3.4: Superimposition of the human PEX14 and the T. brucei PEX14	76
Figure 3.5: T. brucei Pex14 full sequence vs. the truncated N-terminal domain	77
Figure 3.6: Crystal form 1. Crystals of E1W PEX14	78
Figure 3.7: Fragment 5 (crystal form 2) hit in X-ray crystallography binding mode	80
Figure 3.8: Additional hits from the crystallization screen of mutated PEX14	82
Figure 3.9: Crystal form 5	85
Figure 3.10: Superimposition of the entire structure of the Crystal form 5 of the E1W PEX14 variar	nt
"VS ligand influenced" and the apo crystal of the E1W PEX14	86
Figure 3.11: Chemical structure of the hit F6 and resulting crystal structure after soaking	87
Figure 3.12: Superimposition of the "VS ligand influenced" crystal form 5 of the E1W PEX14 varian	nt
obtained by co-crystallization and all eight molecules in the asymmetric unit of the subsequently	
soaked crystal form obtained in complex wtih F6 (crystal form 6)	
Figure 3.13: Anomalous map of E1W PEX14 in complex with F6	89
Figure 3.14: The asymmetric unit of E1W PEX14 in complex with F6 and generated neighboring	
crystal mates within 4 Å of the asymmetric unit	
Figure 3.15: Electron densities in binding pockets of protein molecules A, B, D, F, G, and H	
Figure 3.16: Fragment binding modes in protein molecules A, B, C, and D	
Figure 3.17: Chemical structures of compound 2 (PDB:5L87) and compound 3 (PDB: 5N8V)	95
Figure 3.18: SOFAST HMQC spectra of labelled PEX14 with the VS compounds	
Figure 3.19: Scheme showing how the AlphaScreen functions and calculation of the IC ₅₀	99
Figure 3.20: A strong chemical shift observed in 2D correlation plot of lead compound 1 and VS 7 $^\circ$.100
Figure 3.21: Scatchard plot to determine the dissociation constant K _D	101
Figure 3.22: Images of crystallization hit of the wild-type PEX14 and reproduced crystals	106
Figure 3.23: Crystal of the wild-type PEX14 soaked in compound 1 (PDB: 6S9Y)	107
Figure 3.24: Screenshot from the output of XPREP	
Figure 3.25: Crystal structure of the wild-type PEX14 at a resolution of 1.93 Å	109
Figure 3.26: Melting plots produced from the TSA of wild-type PEX14	111
Figure 3.27: DLS measurements of the wild-type PEX14	112
Figure 3. 28: Polyhedral and ball and stick representation of [TeW ₆ O ₂₄] ⁶⁻	113
Figure 3.29: Image of wild-type PEX14 crystallization drops in the presence and absence of TEW \dots	113
Figure 3. 30: The space groups of higher symmetry suggested by XPREP for the wild-type dataset.	115
Figure 3.31: First crystal structure of the wild-type PEX14 in complex with a ligand – compound 1.	116
Figure 3.32: Superimposition of the wild-type compound 1 complex and the original fragment hit	F5
obtained from soaking the E1W PEX14 Crystal form 1	117

Figure 4.1: Surface representation of endothiapepsin and its binding pockets	. 124
Figure 4.2: A pie of pie chart of the fragments screened in the first STD NMR	. 125
Figure 4.3: A pie chart indicating the binders and non-binders from the 361 fragments screened u	sing
X-ray crystallography	. 126
Figure 4.4: Venn Diagram depicting the number of fragment hits overlapping between both STD N	١МR
screens	. 127
Figure 4.5: The binding mode of F051 (PDB: 4Y4X) to endothiapepsin	. 131
Figure 4.6: Binding mode and chemical structure of the potent inhibitor ritonavir (PDB: 3PRS) to	
endothiapepsin	. 133
Figure 4.7: Electron density shows only binding of ritonavir in crystals simultaneously soaked in	
fragments F109, F236, F267, or F268, and ritonavir	. 134
Figure 4.8: Electron densities showing binding of both F216 or F218 and ritonavir	. 135
Figure 4.9: Simultaneous saking of endothiapepsin in an active site fragment and potent inhibitor	
SAP114 shows active site displacement of fragment, but preserved binding mode of second remo	
fragment copy	
Figure 4.10: Electron density shows only binding of fragments in crystals simultaneously soaked in	n
fragments F290, F306, or F063 and ritonavir	
Figure 4.11: ¹ H 1D NMR reference spectra of F039 and endothiapepsin and STDD NMR spectra in	
CH₃COOH and CD₃COOD	. 139
Figure 4.12: ¹ H 1D NMR reference spectra of F054 and endothiapepsin and STDD NMR spectra in	
CH₃COOH and CD₃COOD	. 139
Figure 4.13: ¹ H 1D NMR reference spectra of F063 and endothiapepsin and STDD NMR spectra in	
CH₃COOH and CD₃COOD.	
Figure 4.14: ¹ H 1D NMR reference spectra of F066 and endothiapepsin and STDD NMR spectra in	
CH₃COOH and CD₃COOD	
Figure 4.15: ¹ H 1D NMR reference spectra of F162 and endothiapepsin and STDD NMR spectra in	
CH₃COOH and CD₃COOD.	
Figure 4.16: ¹ H 1D NMR reference spectra of F189 and endothiapepsin and STDD NMR spectra in	
CH₃COOH and CD₃COOD	. 141
Figure 4.17: ¹ H 1D NMR reference spectra of F206 and endothiapepsin and STDD NMR spectra in	
CH₃COOH and CD₃COOD	. 142
Figure 4.18: ¹ H 1D NMR reference spectra of F286 and endothiapepsin and STDD NMR spectra in	
CH₃COOH and CD₃COOD	. 142
Figure 4.19: Venn diagram showing a comparison of fragment hit rates of water LOGSY NMR, the	غ د
initial STD NMR, and the second STD NMR screenrelative to the 71 fragment hits detected by	
crystallography	. 143
Figure 4.20: Concept of ISA and structure of peptide H-142 potent endothiapepsin inhibitor	
Figure 4.21: Venn Diagram depicting the number of fragment hits overlapping between the SPR a	
X-ray crystallography screens	. 145
Figure 4.22: Plot of log concentration of fragments against their % inhibition	
Figure 4.23: Chemical structures, binding affinities, and color-coded binding positions of all 16	
fragments that were detected as hits in crystallography, the first STD NMR, and SPR	. 150
Figure 4.24: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set A (F063, F267, F291	
	-
Figure 4.25: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set B (F224 and F236)	
Figure 4.26: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set C (F103 and F171)	
Figure 4.27: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set D (F308 and F333)	

Table of Figures

Figure 4.28: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set E (F052, F058, F063)	
	.54
Figure 4.29: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set F (F207 and F261) 1	.54
Figure 4.30: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set G (F216 and F338) 1	.55
Figure 4.31: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set H (F218 and F224) 1	.55
Figure 4.32: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set I (F227 and F274) 1	.56
Figure 4.33: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set J (F041 and F051) 1	.56
Figure 4.34: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set K (F162, F236, 338) 1	.57
Figure 4.35: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set L (F063 and F272) 1	.57
Figure 4.36: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set M (F063 and F268). 1	.58
Figure 4.37: Cocktail STD NMR vs cocktail X-ray crystallography of cocktail set N (F207 and F240) 1	.58
Figure 4.38: Difference between binding position of F063 when soaked as an individual fragment vs	š.
when soaked as part of a cocktail set 1	.60
Figure 4. 39: Calibration of BIAcore ISA experiments	.66

Table of Tables

Table of Tables

Table 4.1: Chemical structures of fragments that were detected as hits in STD NMR but not in Water LOGSY
Table 4.2: Chemical structures and predicted pKa values of fragment hits detected in either the first
or the second STD NMR screen
Table 4.3: Chemical structures, binding affinities, and endothiapepsin binding positions of all seven X-
ray fragment hits measured in both STD NMR screens but not be detected in either of them 130
Table 4.4: Chemical structures of catalytic dyad fragments found to bind in the competitive binding
experiments in the presence of ritonavir
Table 4.5: Fragments selected for STD NMR in heavy water and light water comparison
Table 4.6: Overview of the most potent fragment hits by SPR and their calculated $IC_{50}s$ and K_d 145
Table 4.7: Chemical structures of crystallographic fragment hits not detected as hits in SPR 147
Table 4.8: Results of fragment hit detection in fragment cocktails screened by STD NMR in
comparison to cocktailing in X-ray crystallography

Table of Equations

(Equation 1.1) 1:1 interaction model(Equation 1.2) Net rate of complex formation (association)	
(Equation 1.4) Larmor frequency (Equation 1.5) Energy difference between "parallel" and "anti-parallel" spin states	
(Equation 1.7) R factor	23
(Equation 3.1) Weighted chemical shifts of 2D correlation plot	100
(Equation 3.2) Scatchard plot equation	100

Table of Schemes

Scheme 2.1: Schematic representation of anticodon modification of tRNA by TGT	27
Scheme 2.2: Chemical structures of reporter ligands 1 and 2	34
Scheme 2.3: Design of pyrimidine ring-opened 5- (2) and 4-benzimidazole (3) derivatives and	tricyclic
dihydro-imidazoquinazoline analogs (4) of the parent lin-benzoguanine scaffold (1)	49
Scheme 2.4: Synthesis of 5-substituted Benzimidazole Ligands 2	50
Scheme 2.5: Synthesis of 4-substituted Benzimidazole Ligands 2	51
Scheme 2.6: Synthesis of the Tricyclic Dihydroimidazoquinazoline Ligand 4.1	51
Scheme 2.7: Synthesis of the Tricyclic Dihydroimidazoquinazoline Ligand 4.2	52
Scheme 3.1: Synthetic pathway of the lead compound 1 from the original fragment hit	84

Abbreviations

°C Temperature Celsius Å Ångström (1 Å = 10–10 m) A280 nm Adsorption at 280 nm

ACE Angiotensin converting enzyme

ACT Acetate
ALD Aldolase
Asp Aspartate

B₀ Magentic field strength

BACE1 ß-site amyloid precursor protein cleaving enzyme 1

CCD Charge-Coupled Device

clogP Calculated logarithm of the n-octanol/water partition coefficient

CPMG Carr-Purcell-Meiboom-Gill
CSD Cambridge Structural Database

Cys Cysteine

Da Dalton (1Da = 1g/mol)

DHAP Dihydroxy acetone phosphate

DMSO Dimethylsulfoxide
DNA Deoxyribonucleic acid
DTT 1,4-Dithiothreitol

e– Electron

E. coli Escherichia coli

ENO Enolase

EP Endothiapepsin

FBDD Fragment-based drug discovery
FBLD Fragment-based lead discovery

Fc Calculated Structure Factor Amplitudes

FID Free Induction Decay

Fo Observed Structure Factor Amplitudes

g Gram

g/mol Concentration in grams per mole

GAPDH Glyceraldehydes-3-phosphate dehydrogenase

GK Glycerol kinase

GOL Glycerol

GPDH Glycerol-3-phosphate dehydrogenase
GPO Mitochondrial glycerol phosphate oxidase

h Planck constant. a physical constant that relates the energy carried by a photon to its

frequency. The photon's energy is equal to its frequency multiplied by h (6.62607004

 \times 10-34 m2 kg / s)

H-bond Hydrogen bond

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HK Hexokinase hrs Hour (s)

HSA Hydration Site Analysis
HTS High Throughput Screening

Abbreviations

M

IC50 Half maximal inhibitory concentration
 IPTG Isopropyl β-D-1-thiogalactopyranoside
 ITC Isothermal Titration Calorimetry

K Kelvin

K_d Dissociation constant at equilibrium

Molarity (mol·L-1)

K_i Dissociation constant at equilibrium for an inhibitor

K_m Substrate concentration at which the reaction rate is half of Vmax

LB Lysogeny broth

Ligand Efficiency, defined as $-\Delta G$ in kilocalories per mole divided by the number of

LE heavy (nonhydrogen) atoms

m/v Mass per volume MCK **Multiple Cycle Kinetics** Multi-component reaction MCR MDMolecular dynamics MR Molecular Replacement MR Molecular replacement Mass Spectrometry MS MW Molecular weight

MWCO Molecular weight cut off

n.d Not determinedN/A Not applicableNaCl Sodium chloride

NAD(H)+ Nicotinamide adenine dinucleotide

NADP(H)+ Nicotinamide adenine dinucleotide phosphate

NaN3 Sodium azide

NMR Nuclear Magnetic Resonance

occ. Occupancy

OD₆₀₀ Optical density at 600nm

PDB Protein Data Bank
PEG Polyethylene glycol

PEX14 Peroxin 14
PEX5 Peroxin 5

PFK Phosphofructokinase
PGI Phosphoglucoisomerase
PGK Phosphoglycerate kinase
PGM Phosphoglycerate mutase

pH Negative decimal logarithm of the hydrogen ion activity

pl Isoelectric point PK Pyruvate Kinase

pKa Negative decimal logarithm of the acid dissociation constant

PPI Protein-protein interaction
RDA Reporter displacement assay
RMSD Root-mean-square deviation

RO3 Rule of three

RO5 Lipinski's rule of five rpm Rounds per Minute

Rsym Agreement between symmetry-related reflections

s seconds

S. flexneri Shigella flexneri

SAR Structure-Activity Relationship

SCK Single Cycle Kinetics

SDS page Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SPR Surface plasmon resonance
STD Saturation-transfer difference

STDD Saturation-transfer double difference

TDC Target Definition Compound
TGT tRNA-Guanin-Transglycosylase

Tm Melting temperature

TPI Triose phosphate isomerase

TRIS Tris(hydroxymethyl)aminomethane

tRNA Transfer-RNA
TSA Thermal Shift Assay

v Volume

vdW van der Waals W Water molecule

Water-Ligand Observed via Gradient SpectroscopY

WHO World Health Organization

Z. mobilis Zymomonas mobilis

Zn Zinc

γ Gyromagnetic ratio ΔE Difference in energy

λ Wavelength

 $\mu \qquad \qquad \text{Micro}$

Chapter 1 Introduction

Introductory Remarks

Publications that have resulted from this work:

- Hassaan, E. *et al.* Comparative Analysis of Fragment Binding to a tRNA-modifying Enzyme using SPR, NMR, and X-ray Crystallography, submitted to *ChemMedChem Journal* (June 2019).
- Hassaan, E. *et al.* Fragment Screening Hit Draws Attention to a Novel Transient Pocket Adjacent to the Recognition Site of the tRNA-Modifying Enzyme TGT, ready for submission.
- Hassaan, E. *et al.* Do All Roads Really Lead to Rome? Learnings from Comparative Analysis using SPR, NMR, & X-Ray Crystallography to Optimize Fragment Screening in Drug Discovery, manuscript in preparation
- Hassaan, E. *et al.* Overcoming Pitfalls from the Crystallographic Fragment Screening of the Trypanosomal Peroxisomal Membrane Protein PEX14, manuscript in preparation.
- Hassaan, E. *et al.* Combining Virtual Screening, SOFAST HMCQ NMR, and AlphaScreen Assay to Find Novel Inhibitors of the Trypanosoma Brucei PEX14, manuscript in preparation.
- Magari, F., <u>Hassaan, E.</u>, Gloeckner, S., Scanlan, J. H. W., Bertoletti, N., Metz, A., Huschmann, F. U., Siefker, C., Abazi, N., Heine, A., Weiss, M. S., Klebe, G. Exhaustive X-ray Crystallographic Fragment Screening of a Hit-Enriched 96 Fragment Library Against Diverse Targets, manuscript in preparation.

Chapter 1 Introduction

1 Introduction

This thesis focuses on the analysis of tools and methods used in fragment screening. Fragment screening has persistently proven in the past decade to be an effective tool in identifying first starting points for lead development. It started as back as 1909, with Ehrlich's notion that the chemical structure of a drug compound correlated with its activity ¹, which was later formulated by Gund into what we now know as a pharmacophore.² By the 1990s the concept of drug fragments was not a foreign one to medicinal chemists.³ Identifying small molecules by preparing and screening different analogues was the foundation of SAR (structure activity relationship)-driven medicinal chemistry but what was missing was the ability to screen weakly active fragment compounds. The Abbott SAR-by-NMR screening method was one of the first methods to address this.⁴ At that time, Mattos and Ringe had probed the binding sites on protein surfaces to identify hot spots by using multiple solvent crystal structures (MSCS) ⁵, a method explored computationally by Goodford by his program GRID in 1985 ⁶, and almost 2 decades after that by molecular dynamics by the group of Karplus using the multiple copy simultaneous search (MCSS).⁷

Since then, there have been several biophysical methods developed to rapidly identify weakly binding fragments to a target protein including Surface Plasmon Resonance (SPR), Nuclear Magnetic Resonance (NMR), and X-ray crystallography. SPR has the advantage of utilizing small amounts of protein and fragment, as well as allowing us to observe how molecules interact in real time. Meanwhile, ligand-based NMR is sensitive enough to pick up molecules that interact with low-affinity and also allows us to follow the binding and unbinding events, not only the final binding. Crystal structures from X-ray crystallography have an advantage in providing a three-dimensional structure of the fragment binding pose, which is crucial in visualizing the available space for potential fragment growing. Because fragments are used at high concentrations (limited by their solubility), X-ray crystallography can cope with a large range of ligand binding affinity (sub-nanomolar to millimolar) as well as various sizes of the target macromolecule (even up to megadaltons, depending on the robustness of the crystal and achieved resolution).⁸

Fragment screening is usually done in a cascade manner, but which method should be the one to start the screening cascade? Do these methods reveal similar collections of putative binders?

The goal of this work is to elucidate the differences between these biophysical methods, namely SPR, NMR, and X-ray crystallography, that lead to deviating hit rates. Several proteins have been studied in this context, being tRNA-guanine transglycosylase (TGT) an important protein in *Shigella flexneri*, membrane associated protein peroxin 14 (PEX14) of *Trypanosoma Brucei*, and endothiapepsin (EP) a fungal aspartic protease frequently used as a model protein for therapeutically important proteases.

The TGT project involves screening a 96-fragment library developed in collaboration with Jena Bioscience to validate its potential as a commercial product and suitability for an initial fragment-screening campaign. TGT was one of several enzymes the library was screened against. The library was screened using SPR, NMR, and X-ray crystallography. The X-ray crystallography screen revealed a small fragment, chemically related to the amino acid arginine, which opens a transient sub-pocket when bound to the recognition pocket of the target protein. This prompted us to target this transient sub-pocket also discovered by MD simulations and the above-mentioned fragment screen through structure-based drug design (SBDD). This sub-pocket is found above the preQ₁ recognition site and can provide a promising means of selectivity by developing covalent inhibitors that bind to the Cys158; the gatekeeper of this pocket.

Chapter 1 Introduction

The PEX14 project was part of the AEGIS (Accelerated Early stage drug dIScovery) consortium funded by the Marie Skłodowska-Curie actions fellowship. A previous NMR screening of a fragment library resulted in fragment hits that bind to the N-terminal domain (NTD) of *T. brucei* E1W PEX14. In this project, we attempt to validate these hits through X-ray crystallography by soaking, to allow visualization of the fragment interactions. The promising fragment hits would then be optimized into more potent lead compounds. Additionally, as the objective of the consortium was finding novel inhibitors to treat neglected infectious diseases, virtual screening was also utilized to discover novel compounds that block the PEX14/PEX5 protein-protein interaction.

Finally, the EP project entails an in-depth analysis of the deviating hit rates previously observed upon comparison of six different biophysical methods used to screen an inhouse 361- fragment library. The analysis was focused on two of these methods, X-ray crystallography and STD NMR, and explaining the low overlap between their discovered hits. Another goal of the project was screening of the 361-fragment library using SPR, a biophysical method that was not applied in the previous fragment screening due to complications in immobilizing the target protein. Instead the assay principal was inverted, and the protein was captured by an immobilized peptide decoy inhibitor. This additional biophysical method gives further insight of conditions crucial to maintain while transferring across different techniques. Finally, fragment detection and cocktailing in crystallography was compared to fragment cocktailing in NMR.

1.1 Fragment-Based Drug Design

1.1.1 What are Fragments?

The success of fragment screening is derived from deploying compounds of low molecular weights and simple structures – so called fragments - which enable a faster screening of libraries in comparison to more complex drug-like ligands. A typical fragment adheres to the 'rule of three' (Astex RO3) 9 which states that it should have a molecular weight of <300 Da, clogP \leq 3, and up to 3 hydrogen bond donors and acceptors. There have been many exceptions to this rule where libraries assembled have fragments that do not strictly observe the Astex "RO3", and were found to provide high hit rates. ¹⁰ Generally there are three steps in FBDD (fragment-based drug discovery), which include the design of the library, the screening of the library using several methods to detect binding to a target biomolecule, and finally the fragment hit-to-lead step which develops the fragment hit into a lead compound through synthetic chemistry guided by the structure information that could be provided by X-ray crystallography, *in silico* docking, and data on bioaffinity. ¹¹

Fragments are also applied to overcome the problem of limited chemical diversity of currently available compound libraries. This is because a fragment library has around 10^3 fragments adhering to the RO3, whereas the estimated $10^{60\text{-}200}$ possible druglike compounds synthesized for HTS (high-throughput screening) have 10^7 possible molecules that contain up to 11 atoms of C,N,O, and F that follow the "RO3". Tragments are also "ligand efficient" (defined as $-\Delta G$ in kilocalories per mole divided by the number of heavy (nonhydrogen) atoms) 14 relative to the compounds that make up HTS libraries, because they are weakly binding and thus must make optimal interactions with their targets to bind with a sufficient affinity for detection (Figure 1.1).

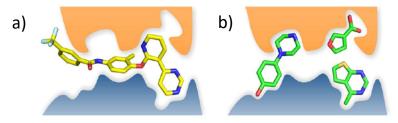


Figure 1.1: HTS hits (a) may bind through several suboptimal interactions. Fragment hits (b), on the other hand, are more ligand efficient as they involve fewer interactions, but they are more optimized. (Image taken from Scott et al., 2012.)

1.1.2 Fragment Libraries

As mentioned above, the design and assembly of a fragment library is the first step taken in FBDD. In our group, we designed a 361-fragment library from fragments that do not strictly adhere to the RO3 criteria but are well suited for crystallographic screening and follow-up chemistry. Commercially available compounds from several chemical suppliers (ASINEX, ChemBridge, MayBridge, InterBio Screen,

LifeChemicals, Enamine, Specs, Vitas M Laboratory) were selected and filtered based on the number of non-hydrogen atoms (between 8 and 20). This was done instead of filtering based on molecular weight simply because advantageous fragments with strong X-ray scatterers such as the electron-rich bromine tend to have large molecular weights. The fragments were then filtered based on undesirable moieties, log *P* values, number of rotatable bonds, number of hydrogen bond donors and acceptors, and the total polar surface area (TPSA). They were then clustered and selected by visual inspection to avoid accumulation of similar chemotypes to cover a range of chemical scaffolds. The physicochemical properties of the finally selected fragments in the library calculated using MOE (under the assumption of standard protonation states at physiological pH conditions) can be seen in Table 1.1. Considering the entire library, less than half of the selected compounds conform to the "RO3". All 361 fragments were validated by screening against endothiapepsin in a fluorescence-based competition assay. This

Chapter 1 Introduction

library was the library screened in the EP project.

Table 1.1: Physiochemical Parameters of the 361-Fragment Library.^a

Parameter	Min	Max	Average
No. of heavy atoms	8	20	15
MW (g/mol)	122	359	224
Lipinksi donor	0	4	1.3
Lipinski acceptor	1	7	3.7
Log P	-1.3	5.4	1.6
Free rotatable bonds	0	7	1.7
TPSA (Ų)	15	126	52

a) Copied from Koester et. al, 2011.

The 96-fragment library (also known as the Frag Xtal Screen) was assembled in collaboration with Jena Bioscience. It was developed as a commercial screening plate that contains 96 wells spotted with the 96 fragments, ready for solubilization and crystal soaking. Amongst the 96 fragments selected is a subset of the original 361-fragment library (54%) which are the fragments that demonstrated suitability for fragment screening when validated on endothiapepsin. The remainder of the library was assembled by filtering compounds from the PDBeChemdatabase ¹⁵ according to the number of containing C, N, and O atoms. Other criteria of selection included price, compound stability, and availability. Exceptions to these filtering criteria were compounds found to be important for protein-function elucidation (e.g NAD) and thus were included in the library. At the end most of the compounds were purchased from Sigma–Aldrich Co. LLC, Taufkirchen, Germany Merck KGaA, Darmstadt, Germany, and Biomol GmbH, Hamburg, Germany. ¹⁶ This library was the library screened in the TGT project.

The fragment library used to screen PEX14 was an inhouse library assembled at the Helmholtz Zentrum in Munich. The fragments were selected from the Maybridge library with the criteria of maximize structural diversity.

1.2 Surface Plasmon Resonance

1.2.1 Concept of SPR

Surface plasmon resonance (SPR) spectroscopy is a label-free technique that allows monitoring noncovalent molecular interactions like protein–peptide and small molecule–macromolecule (e.g., receptor–inhibitor complex) in real time. It has gained popularity in the past few decades as a tool for fragment screening as well as determining compound binding affinities. The power of SPR is that it is noninvasive and does not require labelling the protein with tags, dyes, or specialized reagents. SPR spectroscopy can provide answers regarding specificity of an interaction, the binding kinetics, and affinity of the compounds of interest.¹⁷

The theory behind SPR revolves around TIR, total internal reflection. In SPR, a monochromatic polarized light beam moves from a medium of higher refractive index to a medium of lower refractive index, where it is reflected rather than refracted. The light beam rebounds with an angle of reflection equal to the angle of incidence. TIR occurs once most of the light beam is entirely reflected and less of it crosses the boundary (a gold-glass dielectric layer) between the two mediums. In TIR, the photons of the reflected light beam react with and excite the free electrons of the metal layer (gold), leaking some electrical field intensity known as the evanescent wave. It is coined evanescent because the wave's amplitude decays exponentially as it travels further from the interface surface. This evanescent wave yields surface plasmons which are electromagnetic surface waves that spread parallel to the interface. As they penetrate the medium with the lower refractive index, time-dependent shifts in the value of the reflected "angle" of polarized light are recorded. The value of this reflected light angle is calculated as a function of the angle of incident light. 17-19

A typical SPR instrument consists of three major components; an optical light source, a sensor chip with a thin gold layer, and a detection system (Figure 1.2).

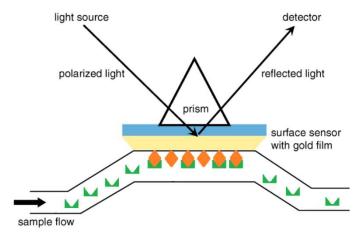


Figure 1. 2: A diagram showing the Kretschmann configuration which is the most common biosensor setup in SPR. The orange diamonds represent the ligand molecules immobilized on the gold surface while the green structures represent the analyte that binds to the ligand. The light source is a helium-neon (He -Ne) laser and the detector is usually a charge-coupled device (CCD). (Image adapted from Bakhtiar et al., 2013.)

The optical light source is usually a near-infrared light-emitting diode. The sensor chip is a glass surface covered with a thin layer of gold. This gold surface is often coated with an alkyl thiol monolayer which serves as a structural template for subsequent derivatizations and limits nonspecific interactions. After the alkyl thiol monolayer, a matrix of carboxymethylated dextran is often covalently attached. It has a thickness of about 25-100 nm and serves to promote a hydrophilic environment for biomolecular interaction. It is inert and does not interfere with SPR signals. To achieve TIR rather than dispersion,

the gold-glass dielectric layer is coupled to a prism. The detection system is most commonly a charge-coupled device (CCD).¹⁷

The two interacting partners in SPR are referred to as analyte and ligand. The partner that is immobilized onto the gold layer surface is the ligand, whilst the second interaction partner which flows over the chip surface is the analyte. The ligand in the context of a direct-binding assay (section 1.2.2) is in fact the protein, whilst the analyte is in fact the ligand or compound of interest. This nomenclature differs from other biophysical methods, but in the coming text we will stick to the SPR nomenclature. This analyte is injected into a continuously flowing buffer solution referred to as running buffer. The flow of this running buffer occurs on the gold surface of the sensor. Analyte-ligand noncovalent interactions occurring at the sensor surface cause a change in the mass, which in turn causes a change in the refractive index of the sensor surface and changes the resonance angle. ¹⁷ If no binding occurs between the analyte molecule in the running buffer and the immobilized ligand, the mass at the sensor surface will not change, and the resonance angle change will be zero.

The graphical output of an SPR experiment is a sensorgram, also known as a binding process curve (Figure 1.3). It plots the absolute response in resonance units (RU), which is the SPR angle change, against the time and 1000 RU is roughly a 0.1° change in SPR angle. A sensorgram shows the occurring interactions between the analyte and ligand in real time and can give information on the type of interaction occurring, whether it is a strong one or a weak one, and whether an interaction occurs at all. The positive response observed during injection of the analyte is known as the association phase when the analyte begins to interact with the ligand. Strong interactions will likely yield higher RUs at this phase than weaker ones, as more analyte molecules will bind to the ligand, leading to a greater change in the mass and thus the resonance angle. When the recorded response starts to decrease, this phase is known as the dissociation phase, when the analyte detaches from the ligand. After one cycle is complete, a regenerating solution is passed over the sensor chip to remove any bound analytes and prepare the chip for the next cycle. ²⁰

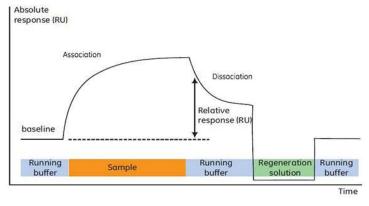


Figure 1.3: The output of an SPR experiment is a sensorgram which is a graphical representation of the interaction occurring between the analyte and the ligand. The bars below the curve show which solution is passed over the sensor surface at each given time. (Image taken from www.gelifesciences.com.)

For kinetic analysis experiments, the interaction is monitored as a function of time over a range of analyte concentrations. The sensorgrams are then fit (for the entire data set) using the software of the SPR instrument (e.g. BIAevaluation from Biacore®) to a mathematical model that describes the interaction. The phases of the interaction provide different information. The association phase during sample injection contains information on the association and dissociation processes, and the dissociation phase after sample injection once the running buffer has removed the dissociated analyte molecules, provides information on the dissociation process.²¹

A 1:1 interaction model can be described as the following ((Equation 1.1) 22 :

$$A + B \xrightarrow{k_a} AB$$
 (Equation 1.1)

where:

 k_a is the association rate constant (M⁻¹s⁻¹), or the rate of complex formation per second in M. and k_d is the dissociation rate constant (s⁻¹), or the rate of complex decay per second.

The net rate of complex formation (association) during injection is given by (Equation 1.2) ²²:

$$\frac{d[AB]}{dt} = k_a[A][B] - k_d[AB]$$
 (Equation 1.2)

and the rate of dissociation after the end of the injection is (Equation 1.3) ²²:

$$\frac{d[AB]}{dt} = -k_d[AB]$$
 (Equation 1.3)

There are two options for running the kinetic analysis experiments depending on the kinetic properties of the analyte. Multiple-cycle kinetics (MCK) is commonly used and it runs each analyte concentration in a separate cycle, where the surface is regenerated after each sample injection (Figure 1.4a). Single-cycle kinetics (SCK) is used when the analyte has a slow dissociation, and thus a series of analyte concentrations are run in one cycle with no regeneration between the sample injections (Figure 1.4b).²¹

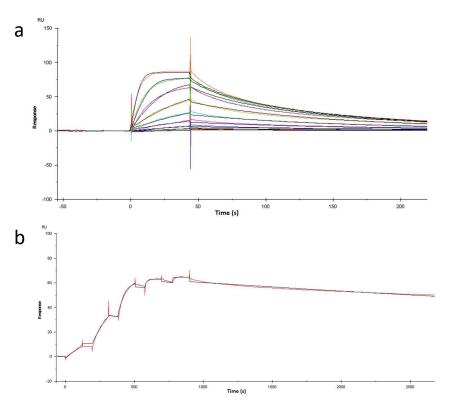


Figure 1.4: Images of sensorgrams obtained from a MCK (a) where the cycles at each analyte concentration are seen here displayed above one another and a SCK (b) which contains only one cycle with the a series of analyte concentrations.

Affinity can then be determined either through affinity constants from kinetics, or steady-state affinity.

The affinity constant for a 1:1 binding event in kinetics is equal to the ratio of the equilibrium association constant (K_A , where $K_A = k_a/k_d$) to the equilibrium dissociation constant (K_D , where $K_D = k_d/k_a$).

In steady-state affinity, the level of steady-state or equilibrium binding is analyzed as a function of interactant concentrations (Figure 1.5).

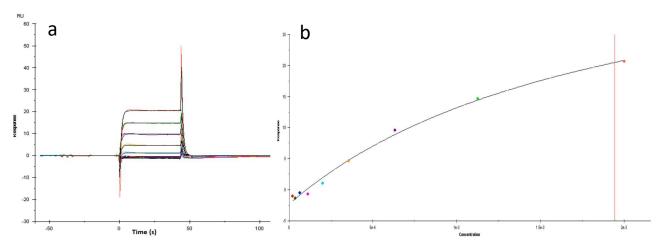


Figure 1.5: A sensorgram (a) and its corresponding plot of steady-state response against concentration (b) for binding affinity determination. The value of the calculated equilibrium dissociation constant K_D (b) is indicated by a red vertical line.

1.2.2 Direct Binding Assays

In a direct binding assay (DBA), the ligand (i.e.: target protein) is immobilized onto the gold surface of the sensor chip (Figure 1.6a). There are several immobilization strategies available by which this is achieved. They utilize free amines, thiols, maleimide, or aldehyde moieties of the ligand.²³ The amine coupling for example is performed by utilizing the free primary amino groups in lysine amino acids, or even the N-terminus of the protein. It is important when immobilizing the target to allow sufficient density on the chip surface without overcrowding it and retain the native nature of the analyte molecule, in other words its biological and structural features. An excess of ligand immobilized onto the chip surface can lead to oligomerization or a phenomenon called "mass transport limitation" which is the rebinding of analyte molecules to the ligand after their dissociation. When this occurs, the measured kinetic constants will reflect the analyte transport process rather than the interaction between analyte and ligand. This can be avoided by reducing the amount of immobilized ligand to limit the amount of analyte binding or increasing the flow rate to remove the dissociated analyte and thus prevent it from rebinding.²¹ During immobilization it is important to consider the sample pH and isoelectric point (pI) of the protein to optimize the immobilization. The pH of immobilization should be at least one order of magnitude below the pI of the protein (so it can adopt an overall positive charge) but above a pH of 3.5 as the carboxylates of dextran on the chip surface have to remain negatively charged.

1.2.3 Inhibition in Solution Assay

In certain cases, immobilizing the ligand (i.e.: target protein) onto the surface of a sensor chip is not possible. A reason for this may be the difficulty in finding a pH suitable for immobilization, as will be discussed in Chapter 4.

An inhibition in solution assay (ISA) works by a different concept to the direct binding assay. In ISA, rather than immobilizing the ligand to the sensor chip, a TDC (target definition compound) is immobilized instead (Figure 1.6b). A TDC is a known active-site binder of the target protein, for example a peptide. The target protein and the compounds being tested for putative protein binding are then incubated together in a 96-well plate and the mixture is injected over the TDC immobilized surface. There is a competition between the compound in the running buffer and the immobilized TDC for the active site of the protein. The initial binding rate is used to determine the percentage of free protein in solution, in other words the unbound protein available to bind to the TDC. When the compound and the free protein in the running buffer bind together, this is reflected in a decrease of the signal recorded, accordingly the percentage of free protein is reduced. This is then recorded as a reduction of the percentage of bound protein to the TDC (as less free protein is now available to bind) and hence corresponds to a higher percentage of active-site binding of the compound to the protein. This allows a qualitative affinity ranking of the compounds since the experiments are all conducted at similar concentrations and under mass transport limitation.²⁴

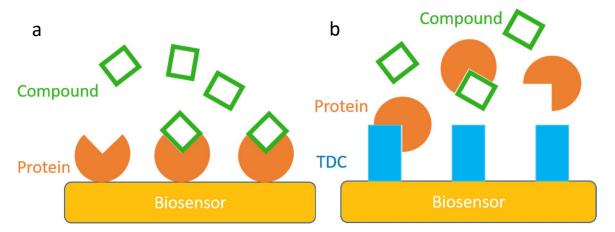


Figure 1.6: a) Concept of DBA. The ligand (protein) is immobilized onto the sensor chip and the running buffer carries the analyte (compound) over the sensor chip surface. b) Concept of ISA derived from Geschwindner, S. et al. (2013). Instead of the protein, the ligand immobilized is a TDC and a solution of fragment and protein is passed over the surface being carried in the running buffer so that the fragment and TDC are in competition for the binding site of the protein.

1.3 Nuclear Magnetic Resonance

1.3.1 Concept of NMR

1.3.1.1 Atomic Spins

NMR is a widely applied spectroscopic technique in chemical analytics and is also used in fragment screening. It observes atomic nuclei with a spin in a magnetic field. These atomic nuclei are "magnetic", which means that they have a "spin" and they perform, using a picture from classical physics, a precession movement in an externally applied magnetic field around the direction of the field. An example of such nuclei are the isotopes 1 H, 13 C, 15 N, 19 F that contain an odd number of protons and/or neutrons and possess a spin of 1/2. A spinning charge generates a magnetic field that has a magnetic moment (μ) proportional to the spin. Taking the hydrogen nuclei as an example, the hydrogen proton has a positive charge and can generate a magnetic moment. When placed in a magnetic field, the spin of the proton will either align "parallel" or "anti-parallel" relative to the direction of the applied magnetic field (Figure 1.7). 25

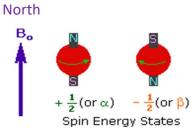


Figure 1.7: When a proton nucleus is exposed to an external magnetic field (B_0), the proton dipole adopts one of two possible spin states; either aligning parallel to B_0 to give +1/2, or anti-parallel to give -1/2. (Image taken from https://chem.libretexts.org.)

The rate of precession of the magnetic moment of the spinning proton around the external magnetic field (change in orientation of its axis) is known as the Larmor frequency (Equation 1.4). ²⁶

$$\omega = \gamma \beta_0$$
 (Equation 1.4)

Where ω is the precession frequency, β_0 is the strength of the applied external magnetic field, and γ is the gyromagnetic ratio which is a specific constant given to each nucleus or particle as seen in Table 1.2.

Table 1.2: Gyromagnetic ratios in MHz/Tesla of specific nucleus or particle.

Nucleus or Particle	Gyromagnetic Ratio (γ) in MHz/Tesla
¹ H	42.58
³He	-32.43
¹³ C	10.71
¹⁹ F	40.05
²³ Na	11.26
³¹ P	17.24
electron	-27,204

The "parallel" (low energy) and "anti-parallel" (high energy) spin states have a difference in energy (ΔE) proportional to the magnetic field strength (B_0), gyromagnetic ratio (γ), and Planck constant (h) (Equation 1.5).

$$\Delta E = \gamma * B_0 * h / (2 * \pi)$$
 (Equation 1.5)

If the atomic nuclei in the external magnetic field is subjected to radiofrequency (Rf) radiation (energy), it can <u>absorb</u> and <u>re-emit</u> electromagnetic radiation as energy.²⁷ The choice of the specific Rf pulse (containing a range of frequencies) to use depends on the strength of the magnetic field and the atomic nuclei present. For example, a ¹⁵N in a magnetic field of 11.74 Tesla will need a pulse at a frequency of 50MHz, as seen in Table 1.3.

Table 1.3: The NMR frequence	v required to disrupt spin alignment for	different isotopes in a magnetic field of 11.74 T.

Isotopes	NMR Frequency at Magnetic Field 11.74 T
H H	500 MHz
13 C	125 MHz
15 N	50 MHz
19 F	470 MHz

The summation of all the magnetic moments of the individual hydrogen nuclei (which are in precession in different phases) is known as the net magnetization vector (NMV), which is in the longitudinal direction or vertical in the direction of B_0 . If the applied Rf is equal to the Lamor frequency, we get what is known as resonance (practically induced by a 90° Rf pulse). A hallmark of resonance is the Rf inducing an in-phase precession of the protons, to give a transverse (NMV) and a transition of parallel to antiparallel spins known as spin flip. The amount of energy required to induce this flipping from one state to the other is equal to ΔE from (Equation 1.5).²⁵

When the Rf is removed, the NMV starts to recover, or relax, to its original vertical position in the direction of B₀.

The T_1 relaxation time, also known as the spin-lattice relaxation time, is a measure of how quickly this occurs. This return of excited nuclei from the high energy state to the lower energy ground state is associated with loss of energy (thermal) to the surrounding nuclei.

The T_2 relaxation time, also known as spin-spin relaxation time, is a measure of how quickly protons return from their in-phase precession to out of phase precession. T_2 is less than or equal to T_1 .

T₁ and T₂ are different in large molecules and small molecules, which is a property exploited in NMR when studying protein-ligand interactions as will be discussed below.

1.3.1.2 NMR Spectra

The energy absorbed by the nuclear spins after the Rf pulse induces a voltage in a recording coil (from the motion of the magnetic moments of the nuclei) that decays with time. This voltage is detected and amplified, and the signal is displayed as a free induction decay (FID), which is a time domain (Figure 1.8a). The FID is converted by a Fourier Transformation (FT) into the NMR spectrum which is the frequency domain (Figure 1.8b) and discloses the resonance frequencies absorbed by individual nuclei in a molecule.²⁸

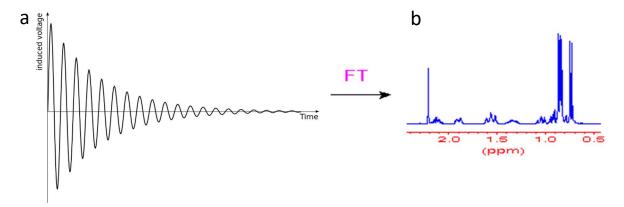


Figure 1.8: a) An FID is a time domain displaying the decay of the voltage induced by nuclei after a Rf pulse. b) Fourier Transformation is applied to the FID to convert it into the NMR spectrum in frequency domain.

The structures of molecules can be studied according to the positions, intensities and the splitting of the resonance signals in an NMR spectrum. The location of the resonance, or NMR signals, on a spectrum is called the "chemical shift". This location depends on the local chemical environment surrounding the nucleus under consideration, and once this environment changes due to chemical interactions (i.e: protein-ligand interactions) the chemical shift of the nuclei changes. The chemical shift is usually reported relative to a reference signal, mostly TMS (tetramethylsilane, $C_4H_{12}Si$). TMS is used as the internal standard as it contains twelve equivalent hydrogen atoms, and thus its 1H NMR spectrum consists of a single peak with a chemical shift assigned as δ =0, hence all chemical shifts are determined relative to it.

Depending on the applied screening protocol NMR provides different amount of information, particularly with respect to the 3D structure. In all cases, the 3D structure insight is less detailed as in crystallography but NMR has the advantage of not only determining protein-ligand binding but also quantifying binding affinities using special titration protocols.²⁹

1.3.2 NMR Instrumentation

A typical NMR spectrometer contains a super-conducting magnet that produces the external magnetic field and a spectrometer that transmits and receives the radio frequency waves used to make the NMR measurements. The magnet is cooled by a helium jacket followed by a nitrogen jacket. The sample probe is placed in between the poles of the magnet and the sample is immersed into this probe holder (Figure 1.9).

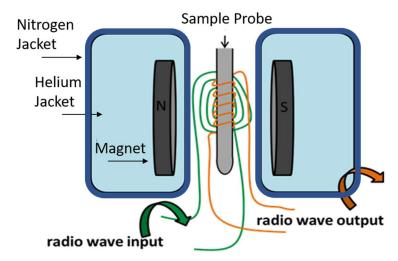


Figure 1.9: The setup of an NMR spectrometer. The spectrometer transmits and receives radio frequency waves and a strong magnet produces the magnetic field. The magnet is cooled by two jackets; one with liquid helium and the second liquid nitrogen. The sample is placed in between the magnet. (Image is adapted from chembam.com.)

1.3.3 Ligand-Observed NMR (CPMG, STD NMR, Water LOGSY)

Also known as ligand-based NMR, ligand-observed NMR is a method of determining protein-ligand binding by observing a spectrum of the ligand signals. There are several different protocols that can be used to perform a ligand-observed NMR, each one using a different property to differentiate binding ligands from non-binding ligands. An example of an NMR protocol that is used is relaxation-based methods such as inversion-recovery (which measures T_1) and the Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (which measures T_2). Large molecules such as protein exhibit significantly different NMR properties than small ligands in terms of relaxation, resulting in different T_1 and T_2 relaxation times. A protein has much shorter T_2 relaxation times than small ligand molecules, and thus ligands that bind to the protein adopt the relaxation properties of the target macromolecule. This is due to the interaction of the ligand with the large macromolecule during the residence time of the bound ligand on the protein, as well as the chemical exchange that occurs when binding (Figure 1.10a).

Another NMR protocol is STD NMR. In STD experiments, the protein is irradiated with a selective Rf and a spectrum is recorded. Ligands that are binding to the saturated protein will receive the Rf irradiation via NOEs (the Nuclear Overhauser Effect). A NOE occurs when a specific nucleus and a neighboring nucleus (dipole-dipole coupled) at equilibrium is magnetically excited and relaxation occurs between the two nuclei. This relaxation causes the intensity of the nucleus that was at equilibrium to increase, whilst the intensity of the excited nucleus in the on-resonance spectrum decreases. Another spectrum is recorded with an off-resonance irradiation. The two spectra are then subtracted (the difference spectrum is generated) and only signals from the binding ligands can be observed (Figure 1.10b). ³⁰

WaterLOGSY is also an NMR protocol that is widely applied. In WaterLOGSY experiments, rather than

Chapter 1 Introduction

irradiating the protein, water molecules are irradiated and transfer their magnetization to the small molecules also by NOE. The ligands that bind to the protein interact with water associated with the protein and receive irradiation from these waters, whereas the non-binding ligands are in the bulk solvent and receive irradiation from these "free" waters that are not associated with the protein. The NOE effect from associated water has an opposite sign to the one from free water. This results in two signals with opposite signs, one from the bound ligand, and one from the non-binding ligand (Figure 1.10c). ³¹ ¹H 1D spectra of small ligand can be generated using the pulse sequence zggpw5.

A benefit of using ligand-based NMR is that one does not need labelled protein, small amounts of protein are needed (10 μ M) as well as small amounts of ligand (100-300 μ M). However, it is a 1D NMR method meaning it only provides binding information in terms of "binding" or "not binding" but it cannot predict the binding site.

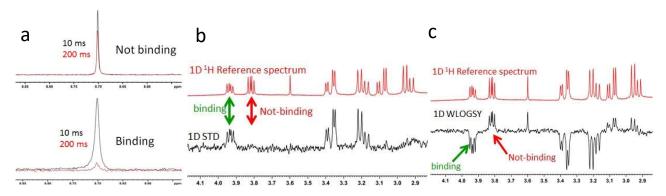


Figure 1.10: a) NMR spectra resulting from a relaxation-based method utilizes the difference in relaxation times between protein and ligand. Ligands that bind to protein adopt the slow relaxation properties of protein and their NMR signals are attenuated relative to non-binding ligands. b) STD NMR spectra is the difference between the on-resonance spectra and the off-resonance spectra. This results in a spectrum showing only the signals of the binding ligands. c) WaterLOGSY spectra display both the binding and non-binding ligands, but because their NOEs have opposite signals, it allows differentiating the binders from the non-binders. (All images are taken from www.theresonance.com.)

1.3.4 Protein-Observed NMR (SOFAST HMQC)

Protein-observed NMR is a method of determining protein-ligand binding by observing a spectrum of the protein signals. Unlike ligand-based NMR, protein-based NMR requires isotope labelled protein (13 C or 15 N). Additionally, larger amounts of both the protein (10 0 μ M) and the ligand (1 mM) are required but the spectra produced are 2D spectra and thus provide information not only on the binding of a ligand but also about the binding site. Titration of different concentrations of the ligand can also provide binding affinities (apparent K_d), which is further elaborated in Chapter 3.

An example of a protein-observed NMR experiment is SOFAST HMQC (selective optimized flip angle short transient heteronuclear multiple-quantum correlation) for which the technical details have been described elsewhere (Schanda et al., 2005). In SOFAST HMQC, not only ¹H nuclei are observed but also ¹⁵N or ¹³C nuclei (depending on the isotope used). This correlates the frequencies of both interacting nuclear spins which results in a 2D heteronuclear chemical shift correlation map with the ¹⁵N proton chemical shifts on the y-axis, and the ¹H proton chemical shifts on the x-axis (Figure 1.11). This allows the heteronuclear NMR analysis of protein. Each peak in the 2D spectrum represents a bonded N-H pair in the protein, and the two coordinates of the peak correspond to the chemical shifts of each of the N and H atoms. A chemical shifting in the peaks indicates ligand binding. If the sequence of the

protein is known, the assignment of the protein residues to each of the peaks allows determination of the binding site based on the residues that exhibit chemical shifts.

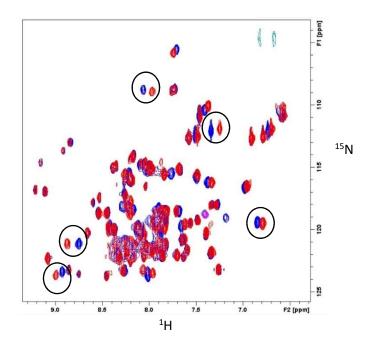


Figure 1.11: A 2D NMR spectrum of a protein before ligand addition (blue) and after ligand addition (red). The residues exhibiting the largest chemical shifts are indicative of the ligand binding site (encircled in black) and protein backbone assignment can provide information on the residues involved in the interaction.

1.4 Crystallography

1.4.1 Concept of X-ray Crystallography

X-ray crystallography is a widely used technique for structure determination of macromolecules and is increasingly used for fragment screening. In fragment screening it has the added benefit of not only determining whether a fragment binds to a protein or not, but also providing a three-dimensional (3D) molecular structure with the binding pose. This allows the visualization of the interacting protein residues and fragment moieties, which is crucial for molecular optimization and how to extend a molecular scaffold at the protein binding site.

1.4.1.1 Crystals and Unit Cells

A crystal can be described as a three-dimensional regular arrangement of molecules and the asymmetric unit of the unit cell is the smallest portion of the crystal lattice that can be used to generate the entire crystal packing, by applying translation and crystallographic symmetry (Figure 1.12a). Classification of the basic geometry of the unit cell in terms of axis, angles, and symmetry elements required to build up the unit cell from the asymmetric unit then leads to the space group.³²

The unit cell is defined by three repeating vectors \vec{a} (that aligns with the direction of the x-axis of the crystal lattice), \vec{b} (that aligns with the direction of the y-axis of the crystal lattice), and \vec{c} (that aligns with the direction of the z-axis of the crystal lattice) with the angles between them, α , β , and γ (Figure 1.12b). Different planes are constructed from the intersections of the crystal lattice and depending on the orientation of the plane it intersects one, two, or all three lattice axes. Parallel lattice planes are determined by the Miller indices, h, k, and l, which label the corresponding reflections and describe the intensity of the diffracted beam.³³

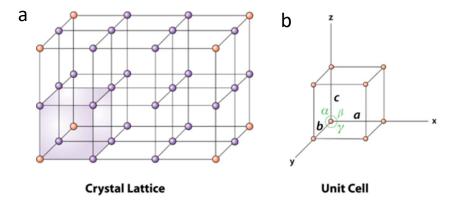


Figure 1.12: An illustration of the relationship of a unit cell (b) to the entire crystal lattice (b). A unit cell can be defined as the smallest repeating part of a crystal lattice, where the angles aren't necessarily 90° as depicted in the image. (Image taken from courses.lumenlearning.com.)

1.4.1.2 Bragg's Law

When an incoming beam hits the crystal, the lattice planes containing regularly arranged atoms scatter an array of waves. Through destructive interference, the reflected beam waves often cancel one another. They only add up if the Bragg's law is fulfilled, which occurs in a few specific directions (Figure 1.13a).

$$2d \cdot \sin\Theta = n \cdot \lambda$$
 (Equation 1.6)

To obtain the diffraction pattern, a series of parallel crystal planes must intersect the Ewald sphere which results in reflection spots on the detector (Figure 1.13b). The Ewald sphere is a geometric construction that describes the relationship between the wavevector of the incident and diffracted X-ray beams, the diffraction angle of a given reflection, and the reciprocal lattice of the crystal. ³⁴ A single reflection spot is the result of a scattered wave from a set of reciprocal lattice planes occupied by an array of atoms in the crystal.

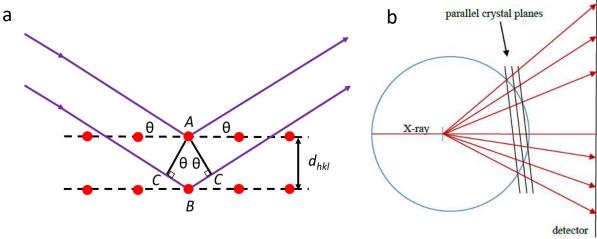


Figure 1.13: a) Reflected beam waves often cancel each other out (destructive interference) except when they fulfill Bragg's law, under which they add up (constructive interference). (Image adapted from Rhodes, G. 2006.) b) Parallel crystal planes must intersect the Ewald sphere for reflected beam waves to create a diffraction pattern. (Image taken from dissertation of Radeva, N. 2016.)

1.4.2 Methodology

The workflow in X-ray crystallography consists of several steps that have been elegantly described by M.S. Smyth and J.H.J. Martin (2000) and are detailed below.

1.4.2.1 Protein Crystallization

Not all proteins readily crystallize, which makes this step rate limiting in many protein crystallographic studies. The principle of crystallization is to place the purified protein under supersaturated conditions that would lead it to nucleate and form crystals, rather than precipitate. This can be demonstrated by the phase diagram for crystal growth in Figure 1.14.³⁵ To find the right conditions, this requires screening several different protein concentrations, precipitants, buffers, temperature, pH, salt and sometimes additives.³⁶ A common crystallization method is vapor diffusion where water vapor leaves the protein drop and ends up in the reservoir to achieve equilibrium. This leads to supersaturation, nucleation, and crystallization.³⁷ There are several commercially available crystallization screens which are 96-well or 384-well plates that contain solutions consisting of varying conditions. To find a new crystallization condition for a protein, the protein is added to such crystallization screens at varying

temperatures (most common are 4°C and 18°C) and the drops are monitored on a regular basis for crystal growth.

Once the crystallization conditions are found and proven to be reproducible, crystallization plates with the protein are set up. There are two choices of crystallization plates; hanging drop or sitting drop (Figure 1.15).

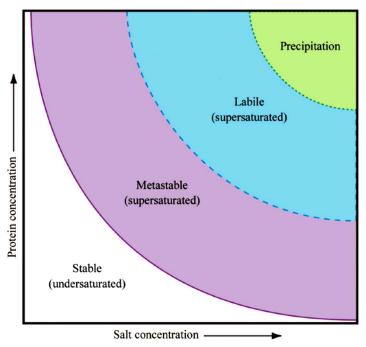


Figure 1.14: The phase diagram of protein crystallization is divided into undersaturated and supersaturated regions. Crystals will only grow from a supersaturated protein solution. The line between the stable (white) and metastable phase (purple) represents the equilibrium between the solid and free-molecule phase. Nuclei will develop into crystals in the metastable and the labile (blue) region, and the region with very high supersaturation (green) is where the protein solution might precipitate. (Image taken from McPherson, A and Gavira, J. A., 2013.)

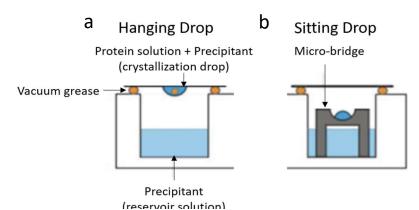


Figure 1.15: The components of a hanging drop (a) and sitting drop (b) crystallization plate. Crystallization occurs through vapor diffusion where the pressure gradient across the vapor space between the drop and the reservoir leads to a net loss of water in the drop. (Image taken from M.S. Smyth and J.H.J. Martin, 2000.)

Preparing crystals for X-ray measurements involves flash-freezing them in liquid nitrogen where they are also stored. This step is done to preserve the crystal structure for long-term storage. Additionally, X-ray measurements are done at cryogenic temperatures to avoid radiation damage to the crystals. To prevent the formation of ice rings in the crystal upon flash-freezing, a cryoprotecting step is added where the crystal is coated in a cryoprotectant such as PEG, MDP, or glycerol before it is immersed in liquid nitrogen.³⁸

1.4.2.2 Optical Setup

Synchrotrons are extremely strong sources of X-rays. X-rays are generated from accelerating electrons in a synchrotron storage ring and a monochromator is used so that a single X-ray wavelength is selected. Slits are used to focus the beam and ensure that it is parallel and is adjusted to 0.1-0.3 mm diameter. The crystal is mounted into the beam using a device called a goniometer where it rotates through the Ewald sphere to allow 360° images of the crystal to be taken. A detector is placed after the crystal and a small lead pellet called the backstop is suspended in the path of the direct beam to prevent the intense X-rays from over-exposure and damaging the detector. The detector commonly used in X-ray experiments is a CCD (charge coupled device) detector. A CCD is an integrated circuit engraved onto a silicon surface that forms pixels (light sensitive elements). The electrical charge generated from incident photons on this surface is read by electronics and converted to a digital copy of the light patterns falling on the device. More advance detectors such as Pixel Array Detectors (using the photoelectric effect) are now found at many synchrotrons and allow the direct detection of X-ray photons without the need to convert them first to visible light.

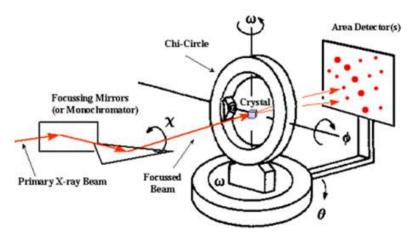


Figure 1.16: The optical setup of an X-ray experiment. (Image taken from publish.illinois.edu.)

1.4.2.3 Diffraction and Data Collection

The photons from the X-ray beam interact with the electrons around an atom which creates a diffraction of the 3D lattice and the various lattice planes produce scattered waves that are recorded on the detector (Figure 1.17). The different lattices generate different diffraction patterns that reflect the symmetry and the distance of the atoms to one another.⁴¹ Based on the diffraction pattern, the arrangement of the atoms in the macromolecule can be determined.

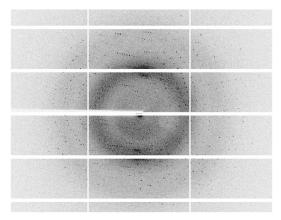


Figure 1.17: A diffraction pattern recorded on a detector is the result of the scattering of atomic electrons after colliding with photons of the X-ray beam.

Before data collection, test images of the crystal are recorded and softwares are used to assign the space group and to suggest the appropriate strategy for data collection based on the resulting diffraction pattern. These softwares predict the space group of the crystal and calculate how many degrees the crystal should be rotated and the number of images to be collected to ensure a complete data set. An example of such a software is iMosflm.⁴² The resolution of the collected dataset is determined by the minimum spacing of lattice planes in the crystal giving rise to a reflection spot, and the distance from the crystal to the detector is adjusted to allow the maximum resolution for collecting diffracted spots. The resolution of spots collected is proportional to the diffraction angle, so the highest resolution will be at the edge of the detector.³⁶

1.4.2.4 Data Processing and Calculating an Electron Density Map

After the diffraction images are collected, complex algorithms are used to process the diffraction data. The intensity of the individual reflection spots in the diffraction pattern holds information about the form of the molecule. Fourier transformation of this information translates the diffraction pattern to the generated electron density which represents the spatial distribution of the atoms in the crystal. The atoms of the crystallized molecules are then assigned to this electron density. The diffractometers used in X-ray crystallography register the intensity of the light that hits them, but this measurement is incomplete. The reason for this is that diffracted waves have only amplitudes (square root of the intensity) but the structure factors also have phases, which are not accessible by the measurement.

The phase information can be obtained by several methods, the simplest of which is molecular replacement (MR), which uses a previously solved structure of a homologous protein or the lower-resolution protein NMR structure of the same protein to calculate the structure factor amplitudes F_{calc} and the phases φ_{calc} . The model used is translated and rotated in the unit cell by computer simulations, producing a calculated diffraction pattern (calculated structure factors $F_{calculated}$) that is compared to

the observed diffraction pattern of the unsolved protein (observed structure factors $F_{observed}$) (Equation 1.7).³³

$$R = \frac{\sum |F_{observed} - F_{calculated}|}{\sum |F_{observed}|}$$
 (Equation 1.7)

Other methods for solving the phase problem involve reconstructing the phases for individual reflections by atom through heaving atom soaking. Heavy atoms are electron rich and they dominate the diffraction pattern which allows their position in the crystal lattice to be determined. Additionally, if the protein contains methionine residues, selenomethionines can be incorporated into the protein during expression and the phases of the resulting crystal structure can be determined through anomalous scattering. ⁴¹

The final electron density map combines the structural model (phases) and the experimentally collected data (amplitudes) and represents how well the selected model fits the experimental data. The two electron density maps that are most often used are the m|Fo|-|Fc| map and 2m|Fo|-|Fc| map. The m|Fo|-|Fc| map, known also as a difference or omit map, shows what has been overfit or not accounted for by the model, while the 2m|Fo|-|Fc| map includes both the m|Fo|-|Fc| map and the electron density around the model.⁴³

1.4.2.5 Data Refinement and Model Building

This is the final step that renders a 3D model of the protein structure. In this step, the 2m|Fo|-|Fc| and m|Fo|-|Fc| maps are used to correct the model and explain any unclear electron density. The m|Fo|-|Fc| map can either show *green blobs* which are the map at positive sigma values, or *red blobs* which is the map at the negative sigma value. A green blob indicates regions of the experimental density where the model is missing. In fragment screening or co-crystallization, this can belong to the molecule of interest. Red negative sigma density indicates regions where the model does not explain the experimental density, which could be a disordered region in the experimental data that was ordered in the model, such as a flexible loop. A common software used for display and model building is COOT ⁴⁴.

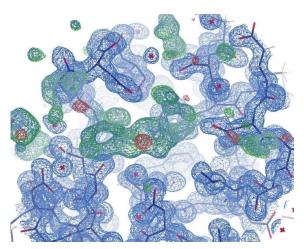


Figure 1.18: An electron density map of a protein structure where the 2m|Fo|-|Fc| (blue) is contoured at a sigma level of 1σ and the m|Fo|-|Fc| map (green and red) is contoured at a sigma level of 3σ . The green blobs indicate areas that are not found in or explained by the model, while the red blobs indicate missing parts of the experimental data. In this dataset, the green blob indicates a fragment binding.

1.4.3 Fragment Screening Using X-ray Crystallography

Fragment screening can be approached either by soaking or co-crystallization.

In soaking, apo crystals of the protein are soaked into solutions containing the fragment of interest at concentrations reaching up to 100 mM. The soaking solution used can be the reservoir solution in addition to a cryoprotectant. In co-crystallization the ligand is added to the protein to form a complex with the protein as it crystallizes. When studying larger ligands, one should keep in mind that there may be differences between soaked and co-crystallized structures. This has been reported with TGT, where crystal structures obtained by soaking did not display the conformational adaption seen in co-crystallized structures of the same complex. 46

Chapter 2 tRNA-Guanine Transglycocylase

Introductory Remarks

The first part of the following chapter (2.2 Fragment Screening) have been submitted to the ChemMedChem Journal.

Another manuscript is in preparation for the second part of the chapter (2.3 Structure-based Drug Design).

The expression and the purification of *Z. mobilis* TGT, the crystallization screen, and elucidation of the crystal structures was done by the author of this thesis.

The SPR and NMR screens were done by the author of this thesis in collaboration with Dr. Stefan Geschwindner and Dr. Per-Olof Eriksson, respectively, during an academic scientific secondment at Astrazeneca R&D in Gothenburg, Sweden

Design and synthesis of ligands 2.1- 2.6, 3.1, 3.2, 4.1, and 4.2 was done by Dr. Christoph Hohn, Dr. Levon Movsisyan, and Dr. F. Wieland Goetzke from the working group of Prof. Dr. François Diederich at ETH Zurich in Zurich, Switzerland.

Co-crystallization, elucidation of the crystal structures, and determination of binding affinities for ligands 2.1-2.6, 3.1, and 3.2 was done by Dr. Frederik R. Ehrmann from the working group of Prof. Dr. Gerhard Klebe at the Philipps University of Marburg in Marburg, Germany. Details can be found in his dissertation.

Co-crystallization, elucidation of the crystal structures, and determination of binding affinities for ligands 4.1 and 4.2 was done by the author of this thesis.

MD simulations were previously done by Naomi Tidten and STTMap calculations were done by Tobias Wulsdorf, both from the working group of Prof. Dr. Gerhard Klebe at the Philipps University of Marburg in Marburg, Germany

The comparative analysis of all methods was done by the author of this thesis.

Chapter 2 TGT

2.1 Introduction

2.1.1 Shigellosis

2.1.1.1 Epidemiology

Shigellosis is a world-wide endemic with 165 million cases reported per year, including 1.1 million deaths globally, mostly in children under five years of age. It is an acute bacterial infection of the intestine caused by the gramnegative Shigella. Symptoms of the disease include sudden abdominal cramping, nausea, fever, vomiting, and bloody stool. Sources of infection include contaminated food or water and therefore, it occurs in low income countries where poor sanitation and overcrowding can lead to the spread of the disease.⁴⁷ Furthermore, after sudden natural disasters (earthquakes, floods, etc.) or chaos of war the disease can also rapidly spread in countries of higher life standards.

2.1.1.2 Treatment

The mode of treatment of shigellosis is antibiotics. These include; beta-lactams such as ampicillin, amoxicillin, and third-generation cephalosporins (cefixime); quinolones such as nalidixic acid and norfloxacin; and finally, macrolides such as azithromycin. However, in addition to serious side effects that may be caused by some antibiotics such as hypersensitivity, dehydration - particularly of small children, anaphylactic shock and encephalopathy⁴⁸, there is a rising resistance to standard antibiotics such as sulfonamides, tetracyclines, ampicillin, and trimethoprim-sulfamethoxazole (TMP-SMX) reported worldwide, and therefore therapeutic options with a novel mode-of-action are necessary to target Shigellosis.⁴⁹

2.1.2 tRNA Guanine Transglycosylase

2.1.2.1 Role in Shigellosis

The pathogenicity of Shigella is dependent on virulence factors (VirF), which are required to invade epithelial cells. The expression of these virulence factors is modulated by the enzyme tRNA Guanine Transglycosylase (TGT, EC 2.4.2.29), which catalyzes an anticodon modification of tRNAs specific for Asn, Asp, His and Tyr, leading to the replacement of guanine-34 at the wobble position by the hypermodified base preQ₁ as seen in Scheme 2.1, which is further modified by other enzymes to queuine. As this modification is a prerequisite for the formation of Shigella virulence factors, the inhibition of TGT has a direct impact on reducing the pathogenicity of Shigella. This positive correlation between VirF and TGT was demonstrated by Durand and Björk, where a mutant of Shigella flexneri with an inactivated tgt gene could not invade host cells due to a reduction in translation of VirF but unchanged levels of virF mRNA. Additionally, transforming the aforementioned mutant with a plasmid containing functional shigella tgt gene restored queuine modification in the mutant as well as exhibiting VirF expression and virulence.

$$H_2N$$
 H_2N H_2N

Scheme 2.1: Schematic representation of anticodon modification of tRNA by TGT.

2.1.2.2 Structure of Zymomonas mobilis TGT

TGT is a homodimer of 86 kDa and adopts the highly populated triose-phosphate isomerase ($(B/a)_8$ - barrel (TIM) with a tightly attached C-terminal zinc-containing subdomain. The packing of the subdomain against the barrel is mediated by an α -helix, located close to the C-terminus, which displaces the eighth helix of the barrel. ⁵⁰ The interface between the two monomers is stabilized by a hot spot of four aromatic amino acids which form a well-organized hydrophobic patch as seen in Figure 2.1. One monomer contributes residues Trp326, Tyr330, and His333 while the Phe92' comes from the adjacent protein dimer. TGT is only functional as a homodimer, where one monomer carries out the catalytic reaction and the other stabilizes the dimer and positions the tRNA correctly for the base exchange reaction as seen in Figure 2.2. The active center of the *Z. mobilis* TGT is located in the area of the β -strands of the C-terminus of the TIM- *barrel* motifs and accepts tRNA molecules with the sequence U33 G34 U35 as a substrate. There are three sub-pockets in the active site of TGT:

- 1) The Ribose-33 pocket, formed by the residues Ala232, Gly261 to Lys264, Asp267, Cys281 to Leu283 and Arg286.
- 2) The Ribose-34 / phosphate 35 pocket, a hydrophobic pocket formed by residues Val282, Leu68, and Val45. A hydrophilic portion formed by Asp280 and Asp102. Additionally, Thr47, Asn70, and Gln107.
- 3) The Guanine-34 / $preQ_1$ pocket, where the base-exchange reaction takes place.

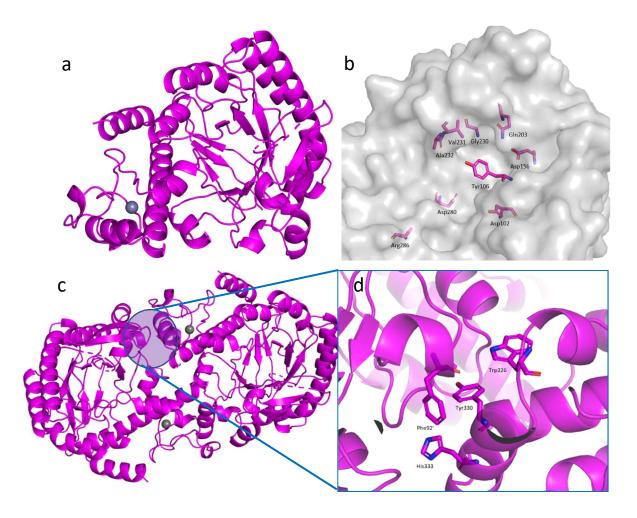


Figure 2.1: Tertiary structure of TGT monomer (a) displaying the TIM barrel and the described active site (b). Quaternary structure of TGT dimer (c) and the residues of the aromatic hotspot at the dimer interface (d).

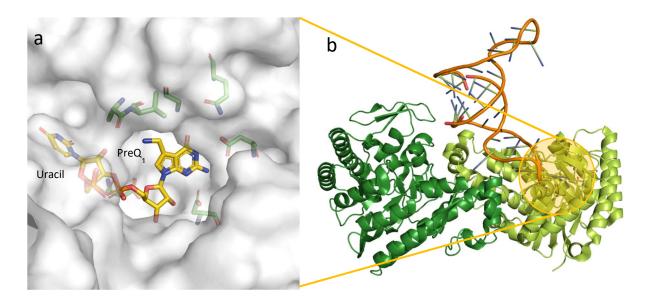


Figure 2.2: a) TGT dimer with a tRNA intermediate substrate in yellow bound to active site. b) One of the TGT monomers (light green) carries out the catalytic reaction while the other monomer (dark green) stabilizes the dimer and holds the tRNA molecule in position. (PDB:1Q2S). Residues Tyr106, Glu107, Val108, Met109, and Ser 110 have been omitted for image clarity.

2.1.2.3 Eukaryotic vs. Bacterial TGT

Eubacteria, eukaryotes, and archaebacteria all have the tRNA-modifying enzyme TGT, whereas the incorporation of the modified 7-substituted 7-deazaguanine bases into tRNAs occurs in a kingdom-specific manner. To start with, eubacterial TGTs catalyze the exchange of the guanine base by the pre-modified base 7-(aminomethyl)-7-deazaguanine (preQ₁) which is further modified into queuine in another two-step catalysis not involving TGT. Eukaryotic TGTs, which show a high sequence identity to eubacterial TGTs (43.2 % between *Homo sapiens* and *Shigella Flexneri*) directly incorporate queuine at the wobble position of tRNAs. Archaebacterial TGTs on the other hand show only about 20–25% sequence identity to eubacterial TGTs and address guanine at a different site, namely position 15 of the D-arm in the majority of archaeal tRNAs.⁵³ In addition to these substrate differences (as shown in Figure 2.3), there are some differences in the sequences of TGT across the different kingdoms (as shown in Figure 2.4). Sequence alignment done in Clustal Omega.⁵⁴

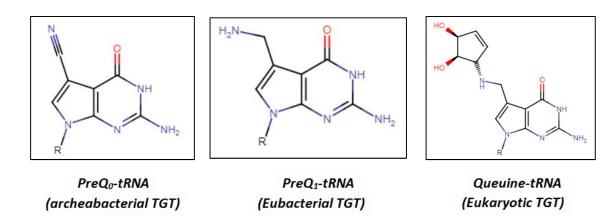


Figure 2.3: Substrates of TGT as they differ across different kingdoms.

sp Q9JMA2 TGT_MOUSE sp Q9BXR0 TGT_HUMAN sp P28720 TGT_ZYMMO sp P44594 TGT_HAEIN tr C3TM87 C3TM87_ECOLX sp Q54177 TGT_SHIFL	MAAVGSPGSLESAPRIMRLVAECSRSGARAGELRLPHGTVATPVFMPVGTQATMKG MAGAATQASLESAPRIMRLVAECSRSRARAGELWLPHGTVATPVFMPVGTQATMKG MVEATAQETDRPRFSFSIAAREGKARTGITIEMKRGVIRTPAFMPVGTAATVKAMKYELDKTSGNARRGRLVFERPQGTFSVETPAFMPVGTYGTVKGMKFELDTTDGRARRGRLVFDRGVVETPCFMPVGTYGTVKG : ** * : : : ** ****** * * * * * * * *	56 56 53 44 40 40
sp Q9JMA2 TGT_MOUSE sp Q9BXR0 TGT_HUMAN sp P28720 TGT_ZYMMO sp P44594 TGT_HAEIN tr C3TM87 C3TM87_ECOLX sp Q54177 TGT_SHIFL	ITTEQLDSLGCRICLGNTYHLGLRPGPELIRKAQGLHGFMNWPHNLLTDSGGFQMVSLFS ITTEQLDALGCRICLGNTYHLGLRPGPELIQKANGLHGFMNWPHNLLTDSGGFQMVSLVS LKPETVRATGADIILGNTYHLMLRPGAERIAKLGGLHSFMGMDRPILTDSGGYQVMSLSS MTPEEVRATGAEILLGNTFHLWLRPGQEVMRKHGDLHDFMQMHRPILTDSGGFQVFSLGK MTPEEVEATGAQIILGNTFHLWLRPGQEIMKLHGDLHDFMQMKGPILTDSGGFQVFSLGD MTPEEVEATGAQIILGNTFHLWLRPGQEIMKLHGDLHDFMQMKGPILTDSGGFQVFSLGD : * : : * * * * * * * * * * * * * * * *	116 116 113 104 100 100
sp Q9JMA2 TGT_MOUSE sp Q9BXR0 TGT_HUMAN sp P28720 TGT_ZYMMO sp P44594 TGT_HAEIN tr C3TM87 C3TM87_ECOLX sp Q54177 TGT_SHIFL	LSEVTEEGVHFRSPYDGEETLLSPERSVEIQNALGSDIIMQLDIVVSSTVTGPLVEEAMH LSEVTEEGVRFRSPYDGNETLLSPEKSVQIQNALGSDIIMQLDIVVSSTVTGPRVEEAMY LTKQSEEGVTFKSHLDGSRHMLSPERSIEIQHLLGSDIVMAFDECTPYPATPSRAASSME LRKITEEGVKFQNPINGERIFLSPEKSMEIQYDLGSDIVMIFDECTPYPATPDYAKKSME IRKITEQGVHFRNPINGDPIFLDPEKSMEIQYDLGSDIVMIFDECTPYPADWDYAKRSME IRKITEQGVHFRNPINGDPIFLDPEKSMEIQYDLGSDIVMIFDECTPYPADWDYAKRSME : : : *: ** * : . : * . : * . ** * * * *	176 176 173 164 160 160
sp Q9JMA2 TGT_MOUSE sp Q9BXR0 TGT_HUMAN sp P28720 TGT_ZYMMO sp P44594 TGT_HAEIN tr C3TM87 C3TM87_ECOLX sp Q54177 TGT_SHIFL	RSVRWLDRCIAAHKHPDKQNLFAIIQGGLNADLRTTCLKEMTKRDVPGFAIGGLSG RSIRWLDRCIAAHQRPDKQNLFAIIQGGLDADLRATCLEEMTKRDVPGFAIGGLSG RSMRWAKRSRDAFDSRKEQAENAALFGIQQGSVFENLRQQSADALAEIGFDGYAVGGLAV MSLRWAKRSRDRFDELGNKNALFGIIQGGVFEELRKVSLEGLVNIGFDGYAVGGLAV MSLRWAKRSRERFDSLGNKNALFGIIQGSTYEDLRDISVKGLVDIGFDGYAVGGLAV MSLRWAKRSRERFDSLGNKNALFGIIQGSTYEDLRDISVKGLVDIGFDGYAVGGLAV ************************************	232 232 233 221 217 217
sp Q9JMA2 TGT_MOUSE sp Q9BXR0 TGT_HUMAN sp P28720 TGT_ZYMMO sp P44594 TGT_HAEIN tr C3TM87 C3TM87_ECOLX sp Q54177 TGT_SHIFL	GESKAQFWKMVALSTSMLPKDKPRYLMGVGYATDLVVCVALGCDMFDCVYPTRTARFGSA GESKSQFWRMVALSTSRLPKDKPRYLMGVGYATDLVVCVALGCDMFDCVFPTRTARFGSA GEGQDEMFRVLDFSVPMLPDDKPHYLMGVGKPDDIVGAVERGIDMFDCVLPTRSGRNGQA GEPKEDMHRILEYICPQIPADKPRYLMGVGKPEDLVEGVRRGIDMFDCVMPTRNARNGHL GEPKADMHRILEHVCPQIPADKPRYLMGVGKPEDLVEGVRRGIDMFDCVMPTRNARNGHL GEPKADMHRILEHVCPQIPADKPRYLMGVGKPEDLVEGVRRGIDMFDCVMPTRNARNGHL ** ::::::::	292 292 293 281 277 277
sp Q9JMA2 TGT_MOUSE sp Q9BXR0 TGT_HUMAN sp P28720 TGT_ZYMMO sp P44594 TGT_HAEIN tr C3TM87 C3TM87_ECOLX sp Q54177 TGT_SHIFL	LVPTGNLQLKKKQYAKDFSPINPECPCPTCQTHSRAFLHALLHSDNTTALHHLTVHNIAY LVPTGNLQLRKKVFEKDFGPIDPECTCPTCQKHSRAFLHALLHSDNTAALHHLTVHNIAY FTWDGPINIRNARFSEDLTPLDSECHCAVCQKWSRAYIHHLIRAGEILGAMLMTEHNIAF FVTDGIVKIRNAKYRDDTSPLDPECDCYTCKNYTKAYLYHLDKCGEILGARLNTIHNLRY FVTDGVVKIRNAKYKSDTGPLDPECDCYTCRNYSRAYLHHLDRCNEILGARLNTIHNLRY FVTDGVVKIRNAKYKSDTGPLDPECDCYTCRNYSRAYLHHLDRCNEILGARLNTIHNLRY :	352 352 353 341 337 337
sp Q9JMA2 TGT_MOUSE sp Q9BXR0 TGT_HUMAN sp P28720 TGT_ZYMMO sp P44594 TGT_HAEIN tr C3TM87 C3TM87_ECOLX sp Q54177 TGT_SHIFL	QLQLLSAVRSSILEQRFPDFVRNFMRTMYGDHSLCPAWAVEALASVGIMLT 403 QLQLMSAVRTSIVEKRFPDFVRDFMGAMYGDPTLCPTWATDALASVGITLG 403 YQQLMQKIRDSISEGRFSQFAQDFRARYFARNS	

Figure 2.4: Sequence alignment of TGT in different kingdoms. Color coding is used to distinguish the common amino acids in all the sequences. Residue 158 in each organism is encircled in red. Sequence alignment was done in *Clustal Omega*.⁸

2.1.2.4 Disrupting TGT Function

Strategies developed to inhibit TGT include potent active-site inhibitors to block the binding of tRNA, thereby preventing the transcription of the virulence factors. Over the years, several of these active-site inhibitors have been developed and co-crystallized to yield high-resolution X-ray structures. 51,55,56,46 The most common and successful ligand series have been lin-benzoguanine derivatives addressing the ribose-33 or ribose-34 subpockets. As TGT is only functional as a homodimer, other strategies include the disruption of the homodimer at the interface. In addition to the ribose-33 and ribose-34 sub-pockets, MD simulations of the apo wild type TGT have predicted the opening of a transient sub-pocket located above the guanine/preQ₁ pocket.⁵⁷ These simulations showed a pronounced flexibility in the region between the crucial residues Val233 and Cys158, creating a small sub-pocket that allows the binding pocket to reach volumes even larger than the eukaryotic TGT models. The observation suggested a strategy to target in our project this new binding site for the design of new inhibitors against TGT following a structure- based drug design concept. The ligands developed to address this transient sub-pocket with a ligand will be described in further detail in this chapter (sections 2.3). The transient sub-pocket opens upon shifting of Cys158 backwards, which we can consider the gatekeeper of this sub-pocket. Ligands that putatively attach covalently to this Cys158 can also be a mean to achieve selective inhibition, as the amino acid sequence of eukaryotic and prokaryotic TGT reveals that the eukaryotic TGT lacks the Cys158 residue and has a Val145 residue instead.

2.2 Fragment Screening

The fragment library used in this project is an inhouse initial screening library consisting of 96 structurally diverse fragments developed in collaboration with the Helmholtz Zentrum in Berlin and Jena Bioscience to a commercially available screening kit. These fragments were selected specifically with a potential to harbor several assorted interactions to a large scope of target proteins. The 96 fragments contain a subset of a larger 361 fragment library that was designed to study whether a library with fragments non-adhering to the Astex rule of 39 could still provide hits that could be used as starting points for lead compounds. This library was tested against endothiapepsin and more than 50% of the resulting hits did not adhere to the rule of 3, which meant that, had the rule of 3 been applied, several fragment hits would have been missed. These fragment hits, in addition to others, were then selected to compile the 96-fragment library.

It is common for fragment-based hit finding projects to perform a screening campaign in a cascade manner, starting with the less demanding screening methods first and apply the more time and resource-requiring screens at a later stage. ^{59,11,60} In our work, we did not follow such an approach but rather screened the entire fragment library with three methods in parallel, particularly to make unbiased comparative analysis possible. The assessment of the results of the three different screening methods will allow us to prioritize which method to begin the screening cascade with.

In this section we present the results of our comparative analysis, and our findings for the low overlap in the hit rates between the applied screening methods.

2.2.1 Fragment Screen by Surface Plasmon Resonance (SPR)

A typical SPR-based fragment screening workflow to prioritize binders and select the most promising hits consists of the three steps, the Clean Screen, the Binding Level Screen and the Affinity Screen. Subsequent to the affinity screen, a competition screen is performed to validate the screening hits.

2.2.1.1 Clean Screen

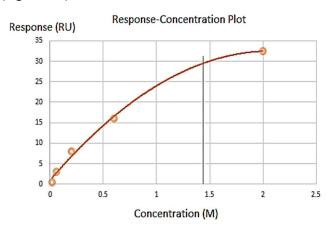
A low molecular weight clean screen was performed to identify and exclude compounds that bind non-specifically or show residual binding to the sensor chip matrix. It is important to remove such compounds because their residual binding can adversely affect the quality of the data of subsequent compound injections, as the biosensor

surface is typically used in an iterative fashion. All 96 fragments were screened against the dextran surface of the sensor chip without any immobilized protein, and there were 15 fragments that had to be excluded due to undesirable properties and technical reasons as mentioned earlier. Their chemical structures are shown in Table S1 of the Supplementary Information.

2.2.1.2 Binding Level Screen and Affinity Screen

The binding level screen was performed with a total of 81 fragments at a concentration of 1 mM and the resulting response units (RUs) were double-referenced against both ⁶¹, the reference channel and the buffer blanks. From the binders, 10 fragments were prioritized as hits (those that had an RU below 13 RUs were excluded, as this was the cut-off value chosen based on binding to reference channel by the negative control). Their structures and responses are given in Table 1 below and correspond to a quite high hit rate of 12.3%.

An affinity screen with a full dose-response was performed with the hits to estimate their binding affinity. From the resulting dose-response curves, the dissociation constant (K_D) could only be determined for J50 and amounts to 1.5 mM (+/- error limits) (Figure 2.5).



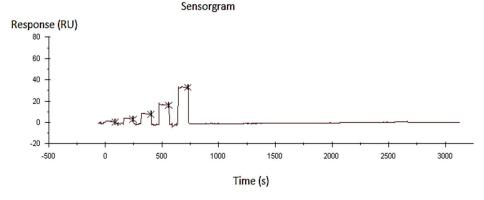


Figure 2.5: Dose-response curve of J50. K_D determined to be 1.5 mM.

2.2.1.3 Competition Screen

To determine whether the fragment hits were specific active-site binders or non-active-site binders, two separate competitive experiments were performed. In the first, compound **2** (Scheme 2.2) was measured at 40 nM in the presence and absence of the fragment hits and the responses were compared. If the response of compound 2 was attenuated in the presence of a fragment, this indicated competitive binding and hence the fragment was deemed as an active-site binder. Likewise, if the response of compound 2 was fortified in the presence of the

fragment, this indicated non-competitive binding and hence the fragment was deemed a non-active-site binder. Only the binding response of J50, J09, J79, and J92 caused a reduction of the potent inhibitor used for displacement, whereas J50 caused the strongest reduction in binding response. In the second experiment, doseresponse experiments were run in the presence of 8 nM of the potent active-site inhibitor compound 2^{62} (SPR $K_D = 7.7$ nM). These second experiments did not include J92, J09, J51, J55, and J79 because of limited availability of material. In these experiments the binding response of J50 was reduced in the presence of the potent inhibitor used for displacement.

Table 2.1: TGT fragment hits from a binding level screen.

Jena Plate ID	Fragment Chemical Structure	RU
J50	N NH NH	45.6
J36		21.3
J92	OH OH OH NO ₂	21.1
J09	O N N N	17.3
J51	N NH	16.7
J55	H ₃ C H ₂ N _{III} ,	15.3
J24		14.7
J07	O NH	14.3

J29	H ₂ N NH	13.7
J79	HN H ₂ O	13.5

Scheme 2.2: Chemical structures of reporter ligands **1**¹⁶: 6-amino-4-[2-(3,4-dihydroxy-5-methoxyoxolan-2-yl)ethyl]-2-(methylamino)-1H,5H-imidazo[4,5-g]quinazolin-8-one and **2**¹⁵: 6-amino-4-[2-[(cyclohexylmethyl)amino]ethyl]-2-(methylamino)-1H,5H,6H,7H-imidazo[4,5-g]quinazolin-8-one.

2.2.2 Fragment Screen by NMR

For binding studies and screening with NMR, the CPMG (Carr-Purcell-Meiboom-Gill)⁶³ and the WaterLOGSY (Water-Ligand Observed via Gradient SpectroscopY)⁶¹ techniques were used. Experimental details can be found in the Materials and Methods section.

2.2.2.1 Validation of NMR Setup

A binding test was performed using J50 (SPR K_D =1.5 mM) as a reporter and compound $\mathbf{1}^{46}$ (Scheme 2) (SPR K_D =68 \pm 5 nM) as a potent inhibitor. A sample with 200 μ M of the reporter compound in NMR buffer is measured first (Figure 2.6). Then a stepwise addition of sub-stochiometric amounts of the TGT protein follows and further spectra are collected. Upon addition of the TGT protein, the CPMG peak of J50 is attenuated and the WaterLOGSY peak turns positive, confirming the weak binding of J50 to the TGT protein. The attenuation of the CPMG peak can be explained by the exchange between free J50 in solution and J50 bound to TGT and indicates that the exchange rate is on an intermediate time scale, compatible with the affinity obtained from SPR (c.f. above). Upon addition of a more potent competitive inhibitor such as inhibitor $\mathbf{1}$ (Scheme 2.2) the bound J50 is displaced from the binding site of TGT and the intensity of the J50 peak is regained. The intensity changes in the WaterLOGSY spectrum can be explained in a similar fashion.

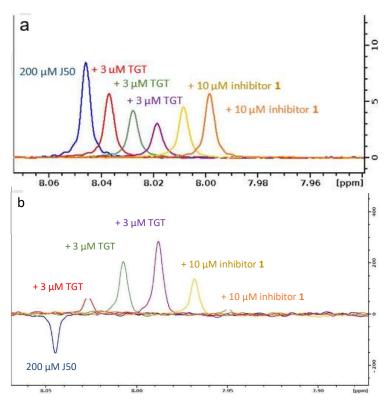


Figure 2.6: a) CPMG and b) WaterLOGSY spectra of J50 binding to TGT. The CPMG spectra in (a) show attenuation of the signal of the freely solvated J50 upon addition of TGT (red, green, violet), which is a proof of weak binding. Upon addition of inhibitor 1, an increase of the signal is detected, returning to that of the freely solvated J50, which is a proof for the displacement of J50 by the more potent inhibitor 1. The WaterLOGSY spectra (b) can be interpreted in a similar way.

2.2.2.2 Fragment Solubility

To ensure that the fragments are sufficiently soluble in the NMR buffer, samples of the fragments alone without protein were prepared and tested in NMR. The recorded spectra were analyzed. Amongst the 96 fragments, 12 had very limited solubility and were therefore excluded from the NMR screen. In addition to these 12 excluded fragments, J67 showed impurity (the spectrum shows signals in the typical range of aromatic portions, however J67 is aliphatic) and J69 has only exchangeable NH protons, thus it cannot be detected by proton NMR. Considering the requirement of rather large amounts of fragment material in NMR, five fragments could not be included due to limited availability of the compounds (J35, J53, J54, J55, J57). These fragments were also excluded from the screen (Table 19 in Appendix). In the end only 77 of the 96 fragments remained to be screened.

2.2.2.3 Binding Screen

The NMR screen was performed following the same protocol as applied for the initial binding test. Fragments were studied at a concentration of 200 μ M in the absence of TGT protein, and then 3 μ M of TGT were added to the same sample. If attenuation of the resonance signal of the studied fragment was experienced, the fragment was considered as a potential hit. Out of 77 fragments, 22 fragments showed attenuation in their NMR signals after TGT addition (Figure 2.7). To confirm specific binding of these hits, 20 μ M of the more potent competitive inhibitor 1 (Scheme 2.2) was added to the sample and a potential regain of the fragment's NMR resonance was recorded. If the fragment's signal intensity continued to become attenuated, it indicated that the fragment continued to bind in the presence of the competitive inhibitor and therefore it was classified as a non-active site binder, and vice versa. From the 22 fragment hits, only four fragments showed a significant regain (more than 4%) including the originally used reporter fragment J50 (Table 2.1). This corresponds to a hit rate of 5%. Figure 2.7 shows the relative intensity of the signals after protein addition and after addition of inhibitor 1.

Table 2. 2: Chemical structures of the fragment hits that showed active-site binding in the NMR-competitive experiments.

NMR-Specific Hit	Fragment Chemical Structure
J28	NH ₂
J50	NH NH
J66	CI
J93	O_2N N N N N N N N N N

2.2.3 Fragment Screen by X-ray Crystallography

All fragments of the 96 entry library¹⁶ were soaked at concentrations of 100 mM into apo crystals of TGT, for an exposure time ranging between three minutes and twenty hours depending on the crystal stability in the fragment solution. For the 96 fragments, eight hits were found to bind to TGT (PDB codes: 5SW3, 5N6F, 5UTI, 5UTJ, 5V3C, 6FS0) as listed in Table 2.3, five of which bind to the active site and three at the surface in the crystal packing (Figure 2.8). The structures were successfully refined to resolutions between 1.10 Å and 1.63 Å, giving clearly defined difference electron densities for the bound fragments. The interactions of the detected fragment hits are described below, where they have been classified based on their spatial locations.

Table 2.3: Chemical structures and resolutions of TGT fragment hits discovered by a direct crystallographic screening.

Fragment Jena ID	PDB Code	Chemical Structure	Resolution Å	Active-Site Binder
J14	6FSO	N NH	1.45	No
J41	5SW3	N N N N N N N N N N N N N N N N N N N	1.38	Yes (molecule 1) No (molecule 2)
J64	5N6F	HN N N N H	1.11	Yes

J72	5UTI	H ₂ N NH O O O O O O O O O O O O O O O O O O	1.36	Yes
J79	5UTJ	0 HN N H₂O	1.55	Yes
J86	5V3C	OH H ₂ N	1.42	Yes
J33	Not deposited	H ₂ N NH	1.25	No
J19	Not deposited	NH	1.37	No

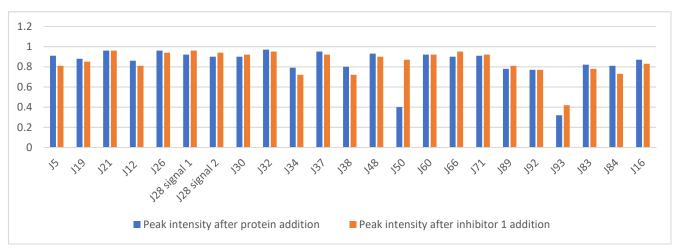


Figure 2.7: Fragments peak intensities difference after protein addition (blue) and after inhibitor 1 addition (orange).

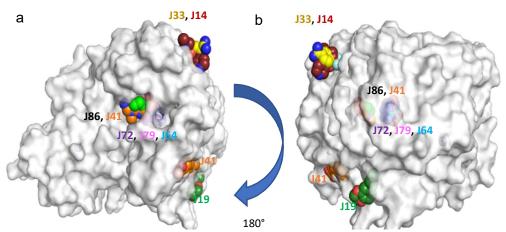


Figure 2.8: a) Overview of the spatial accommodation of the fragments detected in a crystallographic screen directly on protein crystals. Solvent accessible surface representation of TGT viewed toward the catalytic center with the discovered bound fragments in CPK representation, heteroatoms type-coded, carbon atoms of the individual fragments displayed with the same color as the corresponding fragment labels b) the same overview with the image rotated 180° to the left.

2.2.3.1 Fragments Binding to the Recognition Pocket Hosting the Nucleobases in the Catalytic Reaction

As described in the introduction, TGT catalyzes an anticodon modification of tRNAs by replacing guanine against the hypermodified base $preQ_1$. To accomplish this exchange, the nucleobase must be recognized in a deep pocket composed of two adjacent Asp residues (Asp103, Asp156). The site is complemented by the carboxamide terminus of Gln203 and the backbone NH of Gly230. The imidazole moiety of guanine and the exocyclic amino group of $preQ_1$ bind to the atoms of the peptide bond between Ala232 and Leu231. This peptide bond proved to be flexible and by a concerted backbone flip, it either exposed its H-bond donor or acceptor facility toward the bound substrate molecule.⁶⁴

Fragment J41. Two copies of fragment J41 bind to TGT. The first molecule binds with full occupancy to the preQ₁ recognition pocket where one oxygen atom of its carboxylate group forms two hydrogen bonds to the carboxamide nitrogen atom of Gln203 and backbone nitrogen atom of Gly230. The second oxygen atom of this carboxylate group forms one direct hydrogen bond to the carboxylate group of Asp156 and a second water-mediated interaction (W137) to the other oxygen atom of the same residue. J41 additionally forms a π - π stacking interaction to Tyr106 (Figure 2.9). The binding interactions of the second molecule of J41 will be described below.

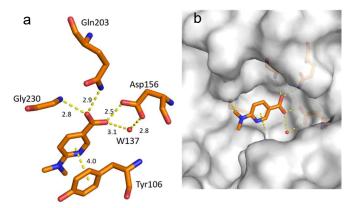


Figure 2.9: a) Binding mode of fragment J41 (PDB: 5SW3) in the preQ₁ pocket. Hydrogen bonds and π - π stacking interactions are shown as yellow dotted lines, water molecules as red spheres, all distances in Å. b) Protein displayed by the gray solvent accessible surface. Residues Tyr106, Cys158, and Val159 have been omitted for image clarity.

Fragment J64. Guanine is present as a search fragment in the screened library. As a matter of fact, our screen also retrieved the natural substrate of TGT as a bound fragment hit. It binds with an occupancy of 100% to the recognition pocket and forms two separate pairs of bidentate hydrogen bonds to the carboxylate groups of Asp156 and Asp102. To accomplish this interaction pattern, one of the two carboxylic acid groups must be assumed to be protonated. Additionally, the carbonyl group of its pyrimidine ring forms a hydrogen bond with the carboxamide nitrogen atom of Gln203 and the backbone nitrogen atom of Gly230. J64 also forms the π - π stacking interaction with Tyr106 at a distance of 3.3 Å (not shown) (Figure 2.).

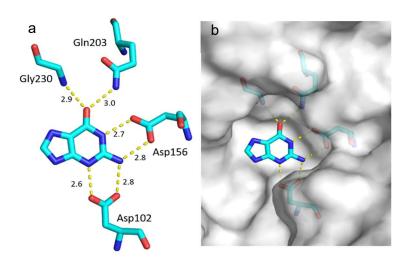


Figure 2.10: a) Binding mode of fragment J64 (PDB: 5N6F), guanine to the preQ₁ pocket. Hydrogen bonds are shown as yellow dotted lines, water molecules as red spheres, all distances in Å. b) Protein displayed by the gray solvent accessible surface. Residues Tyr106 and Met109 have been omitted for image clarity.

Fragment J79. The fragment refines to an occupancy of 94% in the preQ₁ pocket where its triazole ring forms four hydrogen bonds with the enzyme; one with the backbone nitrogen atom of Gly230, one with the carboxamide nitrogen atom of Gln203, and a bidentate hydrogen bond (strong at 2.7 Å and weak at 3.8 Å) with the carboxylate oxygen atoms of Asp156. Additionally, the pyrimidinedione ring of J79 forms a hydrogen bond via its 5-carbonyl group to the nitrogen of Gly261 (weak H-bond 3.8 Å) and a hydrogen bond via its 7-carbonyl group to the carboxamide nitrogen of Gln107. The 4-nitrogen of J79 forms a water-mediated interaction (W28) to the nitrogen atoms of both, Ala232 and Gly230. J79 also forms the π - π stacking interaction with Tyr106 at a distance of 3.5 Å (data not shown) (Figure 2.).

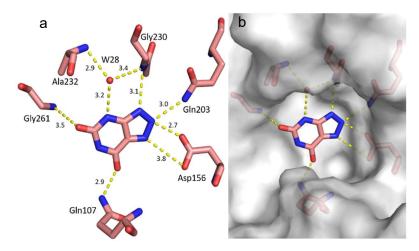


Figure 2.11: Binding mode of fragment J79 (PDB: 5UTJ) to preQ $_1$ pocket. Hydrogen bonds are shown as yellow dotted lines, water molecules as red spheres, all distances in Å. b) Protein displayed by the gray solvent accessible surface. Residues Cys158 and Met109 have been omitted for image clarity.

Fragment J86. This fragment binds with an occupancy of 100% to the preQ₁ recognition pocket and forms a hydrogen bond to the carboxamide nitrogen atom of Gln203 and backbone nitrogen atom of Gly230 via one oxygen atom of its carboxylate group. The second oxygen atom of the acid group faces in monodentate fashion the carboxylate of Asp156. Thus, one of the facing carboxylic acid groups must be assumed to be protonated. Additionally, the terminal primary amine of J86 forms direct hydrogen bonds with the carbonyl oxygen atoms of Leu231 and Ala232, as well as a water-mediated interaction (W120) to the carbonyl oxygen atom of Gly261 (Figure 2.10).

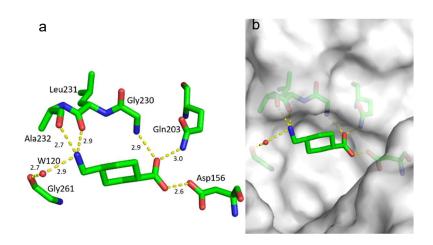


Figure 2.10: a) Binding mode of fragment J86 (PDB: 5V3C) to the $preQ_1$ pocket. Hydrogen bonds as yellow dotted lines, water molecules as red spheres, all distances in Å. b) Protein displayed by the gray solvent accessible surface. Residues Tyr106 and Ser110 have been omitted for image clarity.

Fragment J72. This fragment, a derivative of the amino acid arginine, is amongst the most interesting fragment hits as it induces the opening of a sub-pocket next to the $preQ_1$ recognition site. This opening was previously only reported in co-crystallized structure of mutated variants of TGT in complex with the nucleobase queuine (PDB code 3BLO). The latter substrate is related to $preQ_1$ but comprises an attached sidechain bearing an unsaturated five-membered carbo-cycle. The opening of the $preQ_1$ sub-pocket is brought on by the shifting of the Cys158 gatekeeper residue by 4 Å from its original position in the uncomplexed enzyme. J72 binds with full occupancy to the $preQ_1$ recognition pocket where the fragment's guanidinium moiety forms two bidentate hydrogen-bond interaction with the carboxylate groups of both, Asp156 and Asp102. The fragment adopts a bend, back-folded geometry and forms with its carboxylate group a two-membered water-water chain (W263 and W302) that mediates a H-bond interaction to the carboxamide nitrogen of Gln203 and two direct hydrogen bonds with the nitrogen of Gly230. The amino nitrogen of J72 forms a hydrogen bond with the amide oxygens of Leu231 and Met260. Additionally, this likely charged and protonated amino group forms a water-mediated interaction (W80) with the amide oxygens of both Gly261 and Ala232 (Figure 2.11).

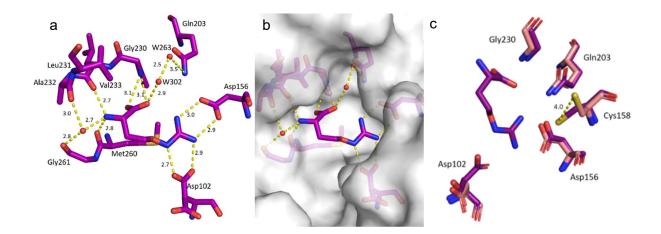


Figure 2.11: a) Binding mode of fragment J72 (PDB: 5UTI) to the $preQ_1$ pocket leading to the opening of a sub-pocket next to the latter pocket. b) The opened pocket, filled by water molecules, is indicated, protein displayed by the gray solvent accessible surface. c) A spatial movement of the Cys158 residue (carbons violet) by 4 Å is detected compared to the original position of this amino acid in the apo TGT structure (PDB: 1PUD, carbons pink) which indicates the opening of the sub-pocket. Residues Tyr106 and Ser110 have been omitted for image clarity. Hydrogen bonds are shown as yellow dotted lines, water molecules as red spheres, all distances in Å.

2.2.3.2 Fragments Binding to the Surface of the Protein or into the Interface of Other Crystal Mates.

Fragment J41. Some fragment hits bind to exposed pockets and depressions on the surface of TGT rather than to the active site pocket. The second molecule of fragment J41 binds with an occupancy of 100% to the surface of TGT via its carboxylate group forming hydrogen bonds with the imidazole side chain of His133 and backbone nitrogen atom of Met134. Upon comparison with PDB codes 5N6F, 5UTI, 5UTJ, 5V3C, and 6FS0, the Arg139 side chain is always found oriented in the same position as fragment J41. The flexibility of the Arg139 sidechain is what allows J41 to bind in its place. Additionally, J41 forms a water-mediated interaction (W34) to the backbone nitrogen atom of Ala80 and carbonyl oxygen of Met75 (Figure 2.12).

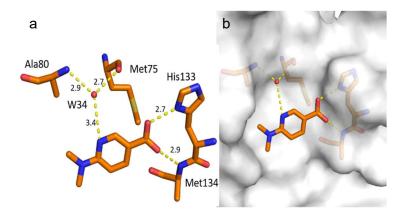


Figure 2.12: a) Fragment J41 (PDB: 5SW3) binding mode to the surface of TGT. Hydrogen bonds are shown as yellow dotted lines, water molecules as red spheres, all distances in Å. b) Protein displayed by the gray solvent accessible surface.

Fragment J14. The fragment binds at an occupancy of 100% to the surface of TGT. Its furan moiety establishes a water-mediated contact (W157) to the backbone nitrogen atom of Ala217 and carbonyl oxygen atom of Gln213 (Figure 2.13). Its secondary and likely charged amino group addresses both, the carboxamide oxygen atom of Gln213 as well as a carboxylate oxygen atom of Glu173 through hydrogen bonds (2.6, 2.7 Å). Additionally, the nitrogen atom of the pyridine moiety of J14 accepts a hydrogen bond from one nitrogen atom of the terminal Arg177 guanidinium head group.

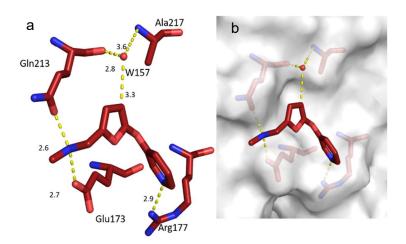


Figure 2.13: a) Binding mode of fragment J14 (PDB: 6FSO) to a surface depression of TGT where the fragment forms some polar contacts to the protein. Hydrogen bonds are shown as yellow dotted lines, water molecules as red spheres, all distances in Å. b) Protein displayed by the gray solvent accessible surface.

Fragment J19. This fragment binds with a refined occupancy of 56% to the surface of TGT via a weak hydrogen bond (3.8 Å) between its benzodioxine ring and the backbone nitrogen atom of Lys85 (Figure 2.14b). Additionally, it forms two water-mediated interactions, one by the nitrogen atom of its pyrrolidine ring via W102 to the carbonyl oxygen atom of Leu146, and a second between the pyrrolidine ring via W269 to the carbonyl oxygen atom of His145. The entire pyrrolidine moiety could not be resolved in the difference electron density (Figure 2.14c). The target protein is only active as a homodimer. Elaborate mutational studies showed that a cluster of four aromatic residues is important for the stability of the dimer interface. Therestingly, while generating the symmetry mate to complete the dimer, it becomes obvious that although J19 does not form direct contacts with the aromatic hot spot formed by residues Trp326, Tyr330, His333 and Phe92' from the other crystal mate, it binds to the interface of the crystallographic symmetry mate in direct contact to residues Ser188, Arg189, and Lys190, which shift by 3.3 Å, 3.8 Å, and 3.7 Å respectively in comparison to the structure of TGT in complex with J41 (PDB: 5SW3) (Figure 2.14d). Obviously, the fragment shifts the adjacent residues in space to create sufficient space for its accommodation. This shifting also causes the residues of the aromatic cluster to shift by at least 3.8 Å in space compared to their positions in the interface for the unperturbed dimer (Figure 2.14g).

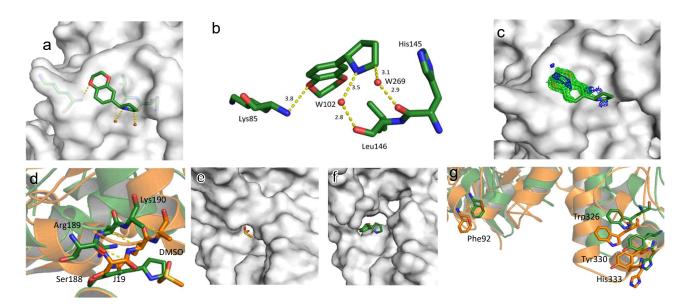


Figure 2.14: a) Binding mode of fragment J19 to the surface of TGT. b) The fragment is held in position by some weak polar interactions. c) In the refinement, the pyrrolidine moiety of J19 could not be fully resolved due to a partially defined electron density. The m|Fo|-|Fc| density map (blue) is contoured at a sigma level of 3σ and the 2m|Fo|-|Fc| density map (green) is contoured at a sigma level of 1σ . d) Fragment J19 binds where usually a DMSO molecule can be found in other soaked crystals like the structure of TGT in complex with J41 (orange, PDB: 5SW3). e) When J19 binds it induces a shifting of the residues at the interface of the crystallographic symmetry mate to accommodate itself. These residues are Lys190, Arg189, and Ser188, which are shifted by 3.3 Å, 3.8 Å, and 3.7 Å respectively. f) This shifting can also be seen in the gap produced at this interface. g) Additionally, the residues that contribute to the aromatic hotspot at this interface are also shifted by at least 3.8 Å. In green: dimer with J19, in orange: dimer without J19. The symmetry mates are not shown.

Fragment J33. The fragment binds with a refined occupancy of 73% to the surface of TGT via a bidentate salt bridge to Glu31' from a neighboring crystal packing mate using its likely charged amidino function. Additionally, the p-trifluoromethyl group forms van der Waals contacts with the methyl group of Ala217 and carbonyl oxygen atom of Gln213. The benzene ring of J33 is involved in a water-mediated contact (W212) to the side chain of Arg177 (Figure 2.15). This fragment was detected by running the program PanDDA (Pan-Dataset Density Analysis) ^{68, 69} designed to analyze multiple datasets and identify ligand binding. When checking the generated "event maps" we took notice of this fragment that otherwise could not be identified in the usual m|Fo|-|Fc| density map contoured at a sigma level of 3 σ . This prompted us to reevaluate the structure and we could see more of the fragment when reducing the m|Fo|-|Fc| density map contouring to a sigma level of 2.3. After refinement was complete, the density for the fragment becomes more apparent as seen in Figure 2.16. The B-factors of the refined J33 are also listed (Figure 2.16d).

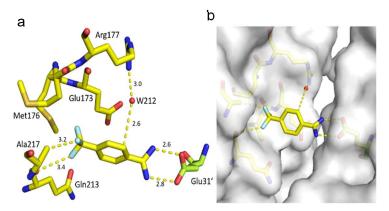


Figure 2.15: a) Binding mode of fragment J33 to the surface of TGT. b) The fragment bridges a contact between two crystal mates in the packing by interacting with Glu31' of an adjacent TGT molecule in the packing.

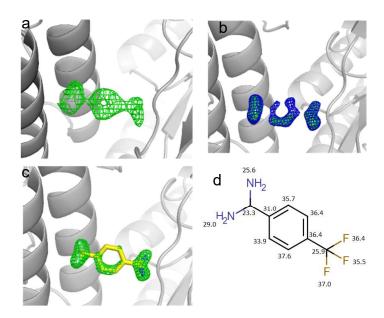


Figure 2.16: a) Event map generated by PAnDDA shows electron density for J33. b) The m|Fo|-|Fc| density map (blue) is contoured at a sigma level of 2.3 σ and the 2m|Fo|-|Fc| density map (green) is contoured at a sigma level of 1σ . c) J33 after refinement, the 2m|Fo|-|Fc| density map (green) is contoured at a sigma level of 1σ . d) B-factors $[\mathring{A}^2]$ of the refined J33 fragment of the non-hydrogen atoms.

Before comparing the results of the three biophysical fragment screenings, we must consider that one obvious reason for the observed deviating hit rates by the distinct methods originates from differences in the biophysical principles of the methods as well as their sensitivity. The concentrations by which the fragments are exposed to TGT deviate in each method. While X-ray crystallography typically screens at very high fragment concentrations (up to 100 mM), fragment concentrations used in SPR are limited to 2 mM due to the risk of unspecific adhesion on the dextran surface chips, and in NMR, fragment concentration were tested at 200 μ M to ensure significant binding effect (see Materials and Methods). This makes detecting of very weak binders more likely at the high concentrations applied in X-ray crystallography. It should also be noted that the NMR binding experiment, contrary to SPR, is able to detect binding at significantly lower concentrations than K_D which warrants the NMR technique being designated as a "spying technique". However, in a recent comparative analysis performed by us, the higher hit rate obtained by crystallography is not simply explained by the fact that the additional hits were only weak binders which show up in the crystals due to the applied high concentrations, as the measured binding affinities of the hits showed strong binders as well. Another obvious deviation between the applied assays may be attributed to the use of different buffers and conditions, as discussed in the following discussion.

2.2.4 Discussion

2.2.4.1 Overlap in Hit Rates of X-ray Crystallography and SPR

There is only one overlapping fragment hit, namely J79, between the crystallographic and SPR fragment screens (Figure 2.17a). In crystallography, J79 binds to the preQ₁ pocket with an occupancy of 94%. In SPR, the binding response of J79 was only 13.5 RU, attributing it with very low response in comparison to the other detected fragment hits. J79 was amongst the fragments that were active-site binders when measured in the SPR competitive experiments. Amongst these fragments were also J09, J50, and J92, which are not from the crystallographically documented hits, even after attempts to soak crystals in the same solution used for crystallization at a pH of 5.5. A possible reason for J50 not being detected by X-ray crystallography may be the clashing of the piperidine ring of J50 with Ala232 which prevents it from binding as seen in Figure 2.18. A similar observation has been reported before, where the affinities of several pterin derivatives discovered through virtual screening showed that a compound similar to J50 had a reduced affinity due to the unfavorable conformation the compound would need to adopt in the binding pocket.⁷¹ In contrast, while J33 is a fragment hit in crystallography, it was among the fragments excluded from the SPR assay due to residual binding to the dextran matrix. Therefore, excluding fragments from specific screens for technical reasons, is also one explanation for missing overlap in the mutual hit rates.

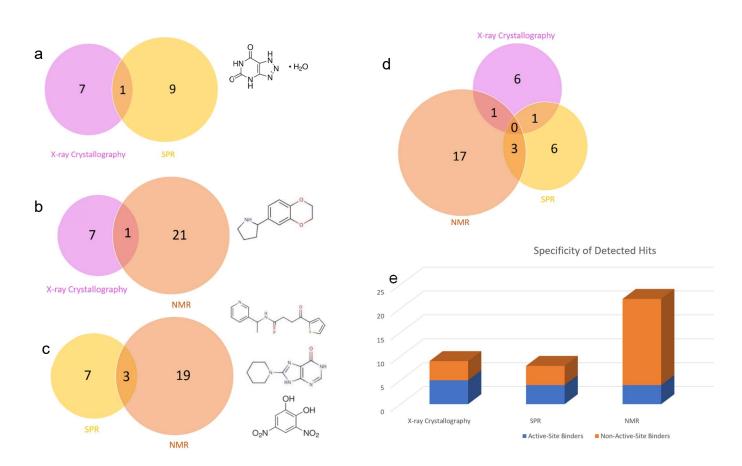


Figure 2.17: a) Overlap in hit rates between crystallography (left, magenta) and SPR (right, yellow) along with the formula of the overlapping hit J79. b) Overlap in hit rates between crystallography (left, magenta) and NMR (right, orange) and formulae of the overlapping hit J19. c) Overlap in hit rates between SPR (left, yellow) and NMR (right, orange) and formulae of the overlapping hits J24, J50, J92. d) Zero overlap in hit rates between crystallography (magenta), SPR (yellow) and NMR (orange). e) Correlation between active-site and non-active-site binders as detected by the three methods.

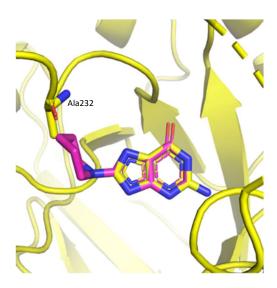


Figure 2.18: The published structure of TGT in complex with J64 (yellow, PDB: 5N6F) superimposed with one orientation of J50 (pink). The piperadine ring clashes with Ala232, demonstrating the unfavorable binding of the fragment in the binding pocket of TGT.

2.2.4.2 Overlap in Hit Rates of X-ray Crystallography and NMR

There is one overlapping hit between the crystallographic and the NMR-based fragment screens, namely J19 (Figure 2.17b). It was picked up by the two screening methods and was confirmed as a non-active-site binder. J19 binds with an occupancy of 56% to the crystal packing interface of TGT. In NMR, the difference in signal peak intensity of J19 was positive, which indicates that it continued to bind in the presence of the potent active-site inhibitor 1. It is also worth mentioning that the crystallographic fragment hits J14, J64, J79 were excluded from the NMR screening due to insufficient solubility.

2.2.4.3 Overlap in Hit Rates between SPR and NMR

There are three overlapping fragment hits between the SPR and NMR fragment screens: J24, J50, and J92 (Figure 2.17c). Among them, J50 is confirmed by both approaches to be a specific active-site binder. In SPR, J50 produces the highest amount of response units from all discovered hits (RU of 45.6). In NMR, it also shows the largest difference in signal intensity for the experiments after protein and after inhibitor addition (+0.48 (48%) relative intensity). Furthermore, it has to be remembered that the SPR hit J79 was excluded from the NMR screening due to insufficient solubility. J24 and J92 did not come up as hits neither after soaking at a pH of 7 nor 5.5. A possible explanation is that J24 may be a surface binder, which makes it difficult to bind in the crystal packing. In solution with the solvated protein molecule however, it does not face this obstacle of packing and thus can bind. Additionally, the pH at which the SPR and NMR were done is 7.4. It has been reported before that a peptide bond flip occurs between Leu231 and Ala232 at this pH, which allows the amide of Ala232-Leu231 to present its donor moiety. 64 This peptide flip does not occur at a pH of 5.5 which may hinder binding.

2.2.4.4 Overlap in Hit Rates of Crystallography, SPR, and NMR

Upon comparison of all fragment hits, there were no overlapping fragment hits from all three methods (Figure 2.17d). As mentioned, several factors are responsible for this finding, at first exclusion of fragments from individual screens due to technical reasons, e.g. insufficient solubility, undesirable

adhesion to sensor chip, or resolving of crystals upon soaking. In detail, four X-ray hits (J33 and J14, J64, J79) were excluded from the SPR and NMR screens, respectively. Additionally, two SPR hits (J55 and J79) were discarded from the NMR screen and five NMR hits (J16, J21, J26, J48, J83) were never subjected to the SPR screen.

Considering the hits in detail, crystallography had the highest correlation between fragment hits detected and percentage of specific active-site binders (Figure 2.17e) whereas NMR suggested a large amount of non-active-site binders.

2.2.5 Conclusions

SPR and NMR are currently the most commonly applied primary fragment screening techniques⁷², however, our results suggest that, if they would have been applied as incipient methods of a screening cascade, they would have missed three specific binders J41, J72, and J86 discovered by a subsequently applied, more elaborate crystallographic screen.

In summary, X-ray crystallography allows the detection of specific binders that may be too weak binders to be detected by SPR and even by NMR. Albeit weak, these specific binders can still provide valid structural information to support the search for appropriate starting points in lead discovery. Additionally, SPR and NMR detected a significantly higher percentage of non-active-site binders relative to crystallography, which shows the importance of also running competition based screenings in SPR and NMR, a requirement not needed in crystallography. NMR proved to be a more sensitive method relative to SPR, as it appeared to pick up the highest numbers of putative binders to the TGT protein.

2.3 Structure-based Drug Design

2.3.1 Definition

Structure-based drug design is a technique adopted when a 3D-structure or model of the target is available. It focuses on the optimization of a molecule that fits into the binding pocket of the target protein and forms interactions that are energetically favorable. Appropriate molecules are either found through virtual screening or chemical design, using the 3D structure of the target to guide the search. ⁴¹

2.3.2 Background

Fragment-based lead discovery is a rather novel approach to generate first hits primarily as promising starting points for drug development projects. However, general purpose fragment libraries usually compile a broad range of chemotypes and can therefore probe protein binding sites efficiently. They are able to provide surprising insights into unexpected features of the studied protein. The crystallographic fragment screening on TGT brought about a small fragment, chemically related to the amino acid arginine, which opens a new, obviously transient sub-pocket when bound to the recognition pocket of the target protein (PDB: 5UTI). Via its guanidinium group the discovered fragment binds to the Asp102/Asp156 motif in the active center and perturbs the spatial arrangement of the remaining recognition pattern next to Gly230 and Gln203 (Figure 2.19b). As a result, a small transient pocket opens, which increases the volume of the recognition site significantly. Two water molecules, stabilized by the nearby carboxylate group of the fragment, are found in the transient pocket. The observed structure immediately provokes some interesting questions. Is the opening of

the transient pocket just an uncommon and reinforced but functionally useless adaptation of the protein or does it indicate a functionally essential feature of the enzyme? Furthermore, can a ligand successfully accommodate this transient pocket and does this support the development of more potent or even more selective ligands?

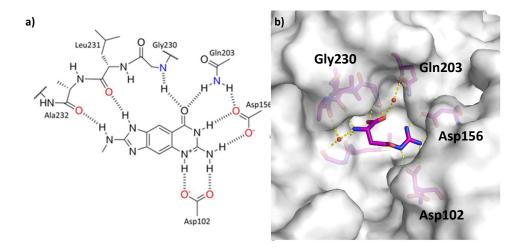


Figure 2.19: a) Schematic binding mode of lin-benzoguanine (1) to TGT. b) Crystallographically determined binding mode of a small fragment binding to TGT. The fragment succeeds in opening a transient binding pocket next to Gly230 and Gln203, which is solvated by water molecules.

Here we study parameters which feature the opening of the transient recognition pocket and elucidate whether the observed adaptations relate to any functional properties of the enzyme family that could be exploited for the development of potent and selective anti-infective drugs. Starting with the well-characterized *lin*-benzoguanine scaffold 1, appropriate to inhibit eubacterial TGT, we first designed some ligands capable to perturb the recognition pattern next to the residues Gly230/Gln203 as suggested by our fragment structure. By chemical synthesis of suitable model compounds, determination of their binding affinity to the enzyme and crystal structure analysis along with a molecular dynamics-based hydration site analysis, we succeeded to trace the role and trigger of the opening of the transient pocket and its potential impact on the development of selective inhibitors of eubacterial TGTs.

2.3.3 Ligand Design

In the past, the design of our ligands aimed at inhibiting bacterial TGT was focused on lin-benzoguanine derivatives as this parent scaffold perfectly occupies the recognition pocket (Figure 2.19a). Without further decoration, it already shows an inhibition with a $K_i = 58 \pm 36$ nM⁷⁴ and provides multiple options for exit vectors to efficiently address the various sub-pockets surrounding the central recognition site. Thus, sub-nanomolar potency could be achieved. 75 Stimulated by the binding geometry of the abovementioned fragment complex and the surprising opening of a transient pocket, we decided, either to open the aminopyrimidine ring of the lin-benzoguanine scaffold or to reduce the carbonyl group in the pyrimidine ring of the parent scaffold. The resulting dihydro-imidazoquinazolines allow the attachment of substituents at C8 that may penetrate into the opened transient pocket (s. Scheme). Thereby, the carbonyl function of the pyrimidine ring, interacting with the backbone NH of Gly230 and the carboxamide nitrogen of Gln203, would be lost. As indicated by the fragment complex, some inherent flexibility with respect to the latter residues might be given. In order to start our ring opening exercise conservatively and to study the impact of the abandoned hydrogen-bonding interactions of the core scaffold to Gly230 and Gln203, we decided to also investigate benzimidazole derivatives with substituents attached to the 5-position. For this purpose, we considered amides and hydrazides. With respect to R1 in Scheme 1, we studied all ligands with an attached methyl group. For some selected examples, also the morpholino ethyl derivatives were investigated.

R1 = Me,
$$O$$

R1 = Me, O

R3

R2 = H, Me, NH₂

R2 = H, Me, NH₂

R3 + H, O

R3 + NH₂

R3 + H, O

R4 + CH₃

R3 + H, O

R4 + CH₃

R5 + H, O

R1 = Me, O

R1 = Me, O

R3 + NH₂

R3 + H, O

R4 + CH₃

R5 + H, O

R6 + CH₃

R7 + CH₃

R8 + H, O

R1 + NH₂

R1 + NH₂

R2 + H, O

R3 + H, O

R4 + CH₃

R5 + H, O

R6 + CH₃

R7 + CH₃

R8 + H, O

R1 + CH₃

R1 + CH₃

R3 + H, O

R1 + CH₃

R3 + H, O

R4 + CH₃

R5 + CH₃

R6 + CH₃

R7 + CH₃

R7 + CH₃

R8 + CH₃

R9 + CH₃

R9

Scheme 2.3: Design of pyrimidine ring-opened 5- (2) and 4-benzimidazole (3) derivatives and tricyclic dihydro-imidazoquinazoline analogs (4) of the parent lin-benzoguanine scaffold (1).

2.3.3.1 Synthesis

The synthesis of 2-substituted *lin*-benzoguanine **1**⁷⁶ and *lin*-benzohypoxanthines **5**⁷⁷ was previously reported by the group of Prof. Dr. François Diederich of ETH Zurich. The 5-substituted benzimidazole ligands were synthesized in 5 steps from 5-benzimidazole carboxylic acid (**6**) (Scheme 2.4). The synthesis commenced with an esterification to give **7**. The benzimidazole nitrogen was sulfamoyl protected to give an inseparable, tautomeric mixture of **8** and **8**′, followed by selective bromination at the 2-position. Partial separation of the tautomers was possible by column flash chromatography and the enriched tautomer **9** was subjected to nucleophilic aromatic substitution with primary amines to afford 2-aminobenzimidazoles **10.1** and **10.2**. Functional group interconversion of the ester with aq. ammonia, methanolic methylamine or aq. hydrazine under microwave irradiation afforded the

protected carboxamides **11.1-11.6**. Final deprotection with HCl and reverse phase flash chromatography purification yielded the ligands **2.1-2.6**.

Scheme 2.4: **Synthesis of 5-substituted Benzimidazole Ligands 2.** Conditions: a) H_2SO_4 , MeOH, 65 °C, 18 h, quant.; b) Me_2SO_2Cl , NEt₃, toluene, 23 °C, 24 h, 66 % (1/0.6 mixture with 1N tautomer **8'**); c) LiN(SiMe₃)₂, THF, -78 °C, 1h, then NBS, THF, -78 °C to 23 °C, 15 min, 56 % (1/0.05 mixture with 1N tautomer **9'**); d) for **10.1**: MeNH₂, EtOAc/MeOH, 23 °C, 14 h, quant.; for **10.2**; 2-morpholinoethylamine, iPr₂NEt, EtOAc/MeOH, 23 °C, 15 h, 54%; e) for **11.1** and **11.2**: aq. NH₄OH, MeOH, 100 °C (MW), 15-30 min, **11.1**: 37%; **11.2**: 31%; for **11.3** and **11.4**: MeNH₂, MeOH 140 °C (MW), 1.5-2.5 h, **11.3**: 67%; **11.4**: 75%; for **11.5** and **11.6**: N_2H_4 , H_2O , 100 °C (MW), 30 min, **11.5**: 80%; **11.6**: 47%; e) HCl, dioxane, reflux, 14-45 h, **2.1**: 23%; **2.2**: 64%; **2.3**: 14 %; **2.4**: 19%; **2.5**: 9%; **2.6**: 52%. THF = tetrahydrofuran, NBS = N-bromosuccinimide, MW = microwave.

For the 4-substituted benzimidazole ligands **3.1** and **3.2** we persued an alternative synthetic strategy (Scheme 2.5).⁷⁸ 2-Aminoaniline **12** was converted into 2-aminobenzimidazoles **13.1** and **13.2** via a two-step protocol with isothiocyanates followed by treatment with EDCI.⁷⁹ The nitro group was reduced using PtO₂ and Pd/C under H₂ and the obtained aniline was guanylated with N,N'-di-boc-1H-pyrazole-1-carboxamidine to afford **14.1** and **14.2**. The N-Boc protecting groups were cleaved with HCl and recrystallization afforded the ligands **3.1** and **3.2**.

$$H_2N$$
 NO_2 $a)$ H_2 NO_2 $b), c)$ H_2 NO_2 $b), c)$ H_3 $NBoc$ N

Scheme 2.5: Synthesis of 4-substituted Benzimidazole Ligands 2. Conditions: a) for 13.1 and 13.2: R^1NCS , DMF, 90 °C, 2 h, then EDCI, 90 °C, 2 h; 13.1: 69%; 13.2: 63% b) for 14.1 and 14.2: H_2 , Pd/C, PtO_2 , EtOH, 4 h, 23 °C, c) N, N'-di-boc-1H-pyrazole-1-carboxamidine, iPr_2NEt , DMF, 23 °C, 16 h; 14.1: 55%; 14.2: 50%; d) for 3.1 and 3.2: HCI, Et_2O/CH_2CI_2 , 23 °C, 10 min, 3.1: 59%; 3.2: 52%. DMF = N, N-dimethylformamide, EDCI = 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide. Boc = tert-butyloxycarbonyl.

Our synthetic strategy for the tricyclic dihydroimidazoquinazoline analogs **4.1** and **4.2** relied on an intramolecular cyclization of *N*-Boc protected guanidines on activated benzylic alcohols (Scheme 2.6 and Scheme 2.7). The synthesis of precursor **15** was previously described by our group. The methyl ester of **15** was reduced with LiAlH₄ to afford benzylic alcohol **16**. Chemoselective *N*-guanylation gave *N*-Boc protected **17**. Tosyl chloride was used to activate the benzylic alcohol for cyclization. Final cleavage of the SEM and the *N*-Boc protecting groups of **18** was achieved with HBF₄ and recrystallization yielded the ligand **4.1**.

Scheme 2.6: Synthesis of the Tricyclic Dihydroimidazoquinazoline Ligand 4.1. Conditions: a) LiAlH₄, THF, 0 °C, 1 h, 58%; b) N'-di-boc-1H-pyrazole-1-carboxamidine, iPr2NEt, DMF, 23 °C, 2 h, 44%; c) TsCl, KOH, THF, reflux, 1 h, 52%; d) HBF₄·Et₂O, CH₂Cl₂, 23 °C, 30 min, 57%. SEM = 2-(trimethylsilyl)ethoxymethyl, Ts = toluenesulfonyl.

We applied a similar synthetic strategy to construct the 2-amino-benzimidazole core of **4.2** as for **3** starting from 4-bromo-5-nitrobenzene-1,2-diamine (**19**). Regioselective nitration via a three-step protocol,⁸¹ followed by benzimidazole formation gave **20**. Introduction of the SEM protecting group (**21**) and carbonylation via a sequence of a palladium catalyzed Heck reaction with ethyl acrylate followed by oxidative cleavage with catalytic OsO_4 afforded a mixture of N-tautomeric nitrobenzaldehydes **22** and **22'** which was separated by careful column flash chromatography. Alkynylation with *in situ* generated propynyllithium gave the racemic propargylic alcohol (±)-**23**. The nitro group was reduced selectively with zinc and aq. NH_4Cl and subsequently the activated guanylation reagent N,N'-di-boc-1H-pyrazole-1-carboxamidine was used to prepare (±)-**24**. Intramolecular cyclization of the guanidine (±)-**24** on the propagylic alcohol was again achieved by

activation with tosyl chloride under basic conditions to yield (\pm)-25. Final deprotection with SnCl₄ and subsequent purification by reverse-phase HPLC afforded the hydrochloride salt of racemic (\pm)-4.2.

Scheme 2.7: Synthesis of the Tricyclic Dihydroimidazoquinazoline Ligand 4.2. Conditions: a) TsCl, pyridine, 65 °C, 17 h, 83%; b) HNO₃, AcOH, 50 °C, 1 h, 69%; c) H_2SO_4 , H_2O , 85 °C, 2 h, 87%; d) MeNCS, $(CH_3)_2SO$, 70 °C, 21 h, then EDCl, 70 °C, 1 h, 44%; e) SEM chloride, NaH, DMF, 0 °C, 1.5 h, 79% (1.2/1 mixture with 1N tautomer 22'); f) $CH_2=CHCO_2Et$, [Pd(OAc)₂], PPh₃, toluene, 110 °C, 48 h; g) OsO₄, NaIO₄, THF/H₂O, 23 °C, 36 h, 52% of 23 and 32% of 23' (over 2 steps). h) $H_3CCH=CHBr$, n-BuLi, THF/hexane, -78 °C to 23 °C, 3 h, 83%; i) Zn, aq. NH₄Cl, MeOH, 0 °C, 2 h; j) N'-di-boc-1H-pyrazole-1-carboxamidine, iPr₂NEt, CH_2Cl_2 , 21 h, 50 °C, 57% (over 2 steps); k) TsCl, CS_2CO_3 , 23 °C, 4 h, 58%; l) SnCl₄, CH_2Cl_2 , 0 °C to 23 °C, 6 h, 34%.

2.3.4 Affinity Determination

Table 2.4: Chemical formulas, binding constants in µM and pdb-codes of the complexes of the studied compounds.^a

Ligand	Parent Scaffold	R	<i>K</i> _i [μM]	PDB code
1.1	I =0	Me	0.058 ± 0.036 ⁷⁶	4PUK ⁷⁴
1.2	R NH 16 NH ₂			
		0 🗸	0.006 ± 0.006 ⁷⁶	4PUJ ⁷⁴
2.1	0	Me	300 ± 37	5J9M
2.2	H ₂ H ₄ NH ₂ NH ₂	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\		
	, ,	0 🗸	544 ± 43	5JT5
2.3	H 4 II	Me	270 ± 50	
2.4	R N 7 6	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\		
		0 ,	264 ± 40	

2.5	O NH2	Me	283 ± 40	5J9N
2.6	NH2 NH2	0 N	282 ± 18	5JT6
3.1	H 4	Me	19 ± 2	5J9O
3.2	R NH NH ₂	O N ∩	58 ± 2	5JT7
a a b	Ŗ3			
4.1 ^b	R NH ₂	Me	29 ± 21 (K _D)	6RKT
	4.1 R3= H, 4.2 R3= ———————————————————————————————————		8.9 ± 1.3 (K _D)	6RKQ
5.1	12.	Me	6.5 ± 2.9 ⁷⁷	3S1G ⁷⁷
5.2	H2 NH 6	0 N	4.1 ⁷⁷	4Q4R ⁸²

a) Binding constants were determined via a radioactive assay with the exception of ligands 4.1 and 4.2 for which SPR measurements were used. b) Reported values are only rough estimates of the K_D, further details are found in the experimental methods.

2.3.5 Crystal Structure Determinations

The crystal structures were obtained by co-crystallizing the corresponding ligands with the enzyme. In all cases, structures with the common twofold dimer packing in space group C2 could be determined at a resolution between 1.21-1.82 Å. In some cases, both a structure with R1= Me and R1= morpholinoethyl could be resolved. However, as previously observed, the morpholino substituent showed pronounced positional disorder in all examples apart 1.2^{76} and no difference electron density could be attributed to this part of the structures. Furthermore, as a structural comparison of the corresponding complexes with the R1 = Me and morpholino ethyl showed, identical binding modes are adopted in all cases. Accordingly, in the following only the crystals structures of the R3 = Me derivatives will be described. However, the crystallographic tables of all here investigated complexes can be found in the Appendix and the coordinates of all complexes have been deposited with the Protein Data Bank (PDB; www.rcsb.org).

2.3.5.1 Crystal Structures with the Ring Opened 5-amide and 5-hydrazide Benzimidazoles 2.1, 2.5

As observed in previous studies, the bicyclic benzimidazole core scaffold forms the expected H-bonding pattern to the backbone atoms of the amino acids Ala232, Leu231 and Gly230 (Figure 2.20a,b). 75,76,83 The carboxamide terminus of Gln203 acts as H-bond donor to the carbonyl groups of the attached 5-amide or 5-hydrazide substituents of **2.1** and **2.5**. Furthermore, the bicyclic core structure establishes π -stacking interactions with the adjacent amino acids Tyr106 and Met260. The side chain of the catalytically essential Asp102 adopts two alternative, equally populated conformations in the complex with **2.1**, whereas in the complex with the hydrazide substituent **2.5**, only the conformer with the carboxylate group oriented into the recognition pocket is observed. The geometry with the side chain rotated off the catalytic center is also found in the structure of the apo protein⁸⁴ and complexes accommodating *lin*-benzohypoxanthines (**5.1**, **5.2**). The alternative orientation towards the ligand-binding site is generally found for complexes with *lin*-benzoguanines. In TGT-**2.1**, the carboxylate group of Asp102 binds to the primary amide function of the ligand (2.9 and 3.0 Å) via the interstitial water molecule W1. The second oxygen of this carboxylate is solvated by additional water molecules found in the ribose-34 pocket (2.9 - 3.1 Å).

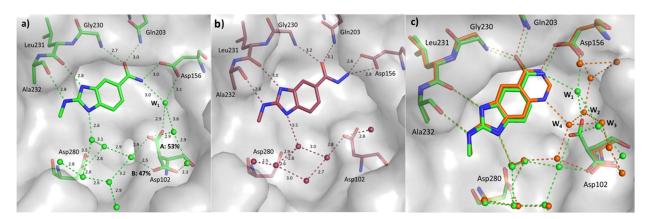


Figure 2.20: Crystallographically determined binding mode of a) **2.1**, b) **2.5** in the recognition pocket of TGT, heteroatoms type-coded, water molecules small spheres, all distances in Å, hydrogen-bonds as dashed lines; c) superposition of **2.1** (carbon atoms green) and **5.1** (carbon atoms orange, PDB: 3S1G ³⁶).

The carboxylate group of Asp156 also forms an H-bond to the terminal amide nitrogen of **2.1** (3.0 Å) and water W1 additionally solvates the neighboring oxygen of the Asp156 carboxylate (3.1 Å). In the complex with the hydrazide-based benzimidazole **2.5** (Figure 2.20b), the terminal nitrogen of the hydrazide protrudes directly toward the carboxylic acid carbon of the Asp156 side chain and forms two H-bonds to its oxygens (2.3 - 2.9 Å). The hydrazide group adopts trans geometry (179°) and pushes the central carbon of Asp156 by approx. 0.9 Å (C_{α} ---- C_{α}) out of the recognition pocket compared to the complex with **5.1**. This push back of Asp156 is associated with a reorganization of further neighboring amino acids and leads to a change in the secondary structure of the loop between the amino acids Glu157 and Thr161 (Figure 2.21a).

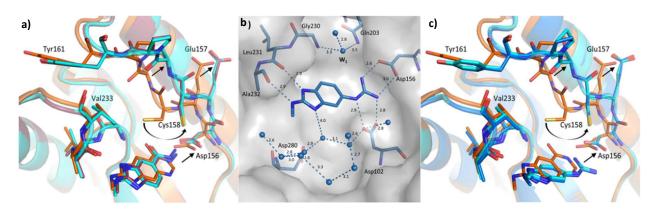


Figure 2.21: a) Superposition of **2.5** (carbon atoms cyan) and **5.1** (carbon atoms orange, PDB: 3S1G³⁶) in the recognition pocket of TGT, heteroatoms type-coded, water molecules small spheres, all distances in Å, hydrogen-bonds as dashed lines; b) binding mode of **3.1**, c) superposition of **3.1** (carbon atoms cyan) and **5.1** (carbon atoms orange, PDB: 3S1G³⁶).

As mentioned above, the carboxylate group of Asp102 refines to one fully populated orientation toward the ligand. Nevertheless, its oxygens do not build a water-mediated interaction network with the ligand, as found for **2.1**. No sufficiently well-defined difference electron density for a water molecule could be detected. In contrast, the interactions to the water molecules in the ribose-34 pocket are also established here (2.7 - 3.2 Å).

The affinities of the amide **2.1** (K_i = 300 ± 37 μ M) and the morpholino analog **2.2** (K_i = 544 ± 43 μ M) decrease by a factor of 46 and 132, respectively, compared to the analog *lin*-benzohypoxanthines **5.1** (K_i = 6.5 ± 2.9 μ M) and **5.2** (K_i = 4.1 μ M). In the crystal structures with the latter ligands a well-defined water network between the ligand and Asp102 is formed (Figure 2.20c).⁷⁷ A similar network is impossible to establish with **2.1**, which likely explains the reduced affinity of the 5-amide derivatives **2.1** and **2.2**. The affinity loss compared to the parent *lin*-benzoguanine scaffold (**1.1**, **1.2**) is even more dramatic and accounts for nearly five orders of magnitude with respect to **2.1** (Table 2.4).

The structural rearrangement of the loop extending from Glu157 to Thr161 along with the ruptured or newly formed interactions and the reorganization of water molecules makes it difficult to compare the hydrazide-based ligands **2.5** and **2.6** with the 5-amide analogs **2.1** and **2.2**. Nevertheless, it appears surprising that for **2.5** (K_i = 283 ± 40 μ M) and **2.6** (K_i = 282 ± 18 μ M) the affinity falls into the same range. This suggests that energetically nearly no costs are involved in the structural rearrangement of the protein, which may represent the transformation to a conformation linked to a protein function that stabilizes the bound ligands.

2.3.5.2 Crystal Structures with the Ring Opened 4-guanidino-benzimidazole 3.1

Compared to the *lin*-benzoguanine scaffold, the 4-guanidino-benzimindazole **3.1** lacks the endocyclic carbonyl function to establish direct H-bonding interactions with Gly230 and Gln203. The corresponding crystal structure shows (Figure 2.21b) that these amino acids are now solvated and stabilized (3.3 - 3.5 Å) via hydrogen bonds to an interstitial water molecule (W1). The carboxylate groups of Asp102 and Asp156 each adopt an orientation which is analogously observed with the bound *lin*-benzoguanine scaffold **1.1.**⁷⁴ The basic guanidine group (pKa > 12) likely binds in charged state. Consequently, salt bridges are expected to form with both aspartates. Furthermore, the 2-aminoimidazole ring interacts with the carbonyl groups of Ala232 and Leu231 (2.9 Å) via H-bonds.

At first sight, the loss of the H-bonds to Gly230 and Gln203 also leads to a drop in affinity as observed for the 5-amide analogs **2.1** and **2.5** with respect to the corresponding *lin*-benzoguanine **1.1** or *lin*-benzohypoxanthine **5.1**. With $K_i = 19 \pm 2 \mu M$ for **3.1** and $K_i = 58 \pm 2 \mu M$ for **3.2**, the 4-guanidine-substituted ligands bind slightly stronger to TGT than the studied 5-benzimidazoles. This suggests that the gained charge-assisted interactions to Asp102 and Asp156 can only partially compensate for the lost interactions to the nitrogen atoms of Gly230 and Gln203. However, also a second, perhaps even more important effect has to be considered. By opening the pyrimidine ring of the *lin*-benzoguanine scaffold, the ligand scaffold loses its correct preorganization. Instead, the guanidine portion, attached to the benzimidazole, must adopt one single conformation, which supposedly is not the most favored one. This, along with the loss of conformational degrees of freedom, will have a detrimental effect on binding affinity. In other studies we saw that this effect can reduce the binding constant by two to three orders of magnitude.⁸⁵

Interestingly, as with hydrazide-substituted benzimidazoles **2.5** and **2.6**, the structural transformation of the secondary structural element (Glu157 to Thr161) is also observed with the guanidine-substituted ligands **3.1** and **3.2** (Figure 2.21c). Also here, a spatial shift of Asp156 is observed, as seen in TGT·**5.1** (C_{α} --- C_{α} = 0.3 Å). However, a second aspect may be important in this complex to transfer the enzyme into the alternative state with the opened transient pocket. The above-mentioned water W1, which binds in the vicinity of the original position of carbonyl function of the *lin*-benzoguanine scaffold, solvates the polar groups of Gly230 and Gln203, but in addition, it fills the created void and exerts steric pressure onto the flexible loop of Glu157 to Thr161 (Figure 2.21c). Accordingly, W1 supports the expansion of the opening pocket.

To study the impact of steric pressure in this region of the binding pocket by a bound ligand in more detail, we designed chemically reduced tricyclic analogs of the *lin*-benzoguanine scaffold bearing an sp³ carbon atom at the former spatial position of the carbonyl group in pyrimidine ring (4 in Scheme 2.3). Through the attachment of substituents at this carbon atom, a chiral center is created. Therefore, we first decided to study the achiral unsubstituted derivative 4.1 and subsequently the propinyl analog (4.2). Modeling considerations strongly supported the hypothesis that only the *R*-enantiomers will orient the attached substituents in a way to fit into the possibly opened transient pocket. The compounds were synthesized and subjected to co-crystallization.

<u>2.3.5.3 Crystal Structures with Dihydro-imidazoquinazolines</u> <u>**4.1** and **4.2** of the *Lin*-benzoguanine <u>Scaffold</u></u>

Ligand **4.1** binds with an occupancy of 87% to the recognition pocket. The guanidine moiety in the six-membered ring forms two bidentate hydrogen bonds with the carboxylate groups of Asp156 and Asp102. The second guanidine moiety in the imidazole ring interacts via a hydrogen bond with the carbonyl oxygen of Leu231, while the exocyclic amino group hydrogen-bonds the carbonyl oxygen of Ala232. Furthermore, **4.1** establishes a π - π stacking with Tyr106 and Met260. Thus, a very similar recognition pattern as seen for *lin*-benzoguanine **1.1** ⁷⁴ is found in this area. Nevertheless, the ligand

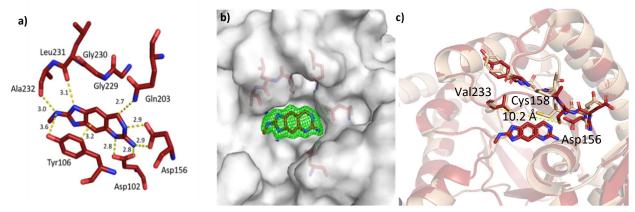


Figure 2.22: Crystallographically determined binding mode of a) **4.1** in the recognition pocket of TGT, heteroatoms type-coded, all distances in \mathring{A} , hydrogen-bonds as dashed lines; b) **4.1** difference electron density (|Fo|-|Fc| omit map) contoured at 3σ , protein displayed with the solvent accessible surface showing the recognition site together with the opened transient pocket; c) binding mode of **4.1**.

also induces the structural transformation of the secondary structural element (loop Glu157 to Thr161), which leads to an opening of the transient sub-pocket (Figure 2.22c) and the carboxamide group of Gln203 forms a van der Waals contact to C8 of the dihydro-imidazoquinazoline scaffold. Due to the pocket opening, the distance between Val233 C_{α} and Cys158 S_{ν} expands in TGT·**4.1** to 10.2 Å.

The complex formed with **4.2** shows very similar geometry as TGT-**4.1** with nearly identical recognition pattern. In contrast to the latter complex, Gln203 adopts an altered conformation providing additional space to accommodate the needle-like propinyl side-chain. Even though the complex was co-crystallized from a racemic mixture, the enzyme only picks the *R*-enantiomer. Supposedly, the *S*-enantiomer does not bind as it would have to poke into the protein and create substantial steric clashes. The difference density allows placement of all atoms of the propinyl side chain, even though the less-well defined density next to the terminal methyl group indicates enhanced residual mobility of the extended side chain. As for the other complex, **4.2** induces a spatial shift of the loop comprising Glu157 to Thr161. Asp156 is, compared to TGT-**1.1**, pushed by 0.6 Å out of the recognition pocket and helps to trigger the structural transformation of the loop (Figure 2.23c). The opening of the transient pocket results in a mutual distance between Val233 C_{α} and Cys158 S_{γ} of 10.8 Å.

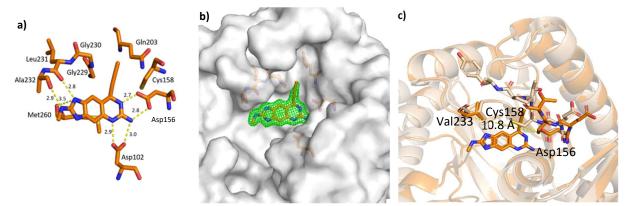


Figure 2.23: Crystallographically determined binding mode of a) 4.2 in the recognition pocket of TGT, heteroatoms type-coded, all distances in Å, hydrogen-bonds as dashed lines; b) 4.2 together the difference electron density (m|Fo|-|Fc| omit map) contoured at 3σ , protein displayed with the solvent accessible surface showing the recognition site together with the opened transient pocket; c) binding mode of 4.2.

2.3.6 Expansion of the recognition pocket by a transient sub-pocket

The enzyme TGT is capable of performing large conformational adaptations upon ligand binding. An opening of a small transient pocket is not only observed in the initially described fragment structure (Figure 2.19) but also in the here described subset of complexes, that push Asp156 slightly out of the recognition pocket and create sufficient steric pressure onto residues Gly230 and Gln203. Transient pockets may open as intermediate states under dynamic conditions in conformational equilibrium and they can be stabilized by a well-fitting and energetically favorably binding ligand. To study the occurrence of such transient pockets, usually molecular dynamics simulations are consulted. Already earlier on, we speculated about the existence of such a transient pocket based on the results of an MD simulation of TGT.⁵⁷ This former study revealed transient openings of a pocket along the trajectory of the protein. By analyzing individual frames along the trajectory in terms of the mutual distance between S_V of Cys158 and C_{α} of Val233, values between 6 Š(closed pocket) to 12 Š(opened pocket) reflecting a binding site volume which varies between 250 and 750 ų were recorded. Remarkably, the various complexes described here, also show a strong variation of this distance, which matches with the range indicated by the MD simulation (Table 2.5). Thus, the incipient simulations suggest that, under physiological conditions, the apo-enzyme is capable to open such a transient pocket.

Table 2.5: Mutual Cys158S $_{\gamma}$ -Val233C $_{\alpha}$ distances and backwards push of Asp156 C $_{\gamma}$ in different TGT-ligand complexes with respect to TGT-1.1.

Complex	Cys158S $_{\gamma}$ -Val233C $_{\alpha}$ distance [Å]	Asp156 push compared to TGT·1.1 [Å] ^{a)}	PDB code
TGT- 1.1	6.5	0.0	4PUK
TGT·frag	9.8	0.4	5UTI
TGT- 2.1	6.4	0.2	5J9M
TGT- 2.2	6.2	0.1	5JT5
TGT- 2.5	10.2	0.9	5J9N
TGT- 2.6	10.2	0.6	5JT6

TGT- 3.1	10.5	0.7	5J9O
TGT- 3.2	10.6	0.7	5JT7
TGT- 5.1	6.2	0.3	3S1G
TGT- 4.1	10.2	0.3	6RKT
TGT- 4.2	10.8	0.6	6RKQ

a) The amount by which Asp156 is pushed out of the recognition pocket was determined by performing an overall RMSD C_{α} atom superposition of the studied complex with TGT·1.1 and measuring the mutual distances of Asp156 C_{ν} s in both structures.

Obviously, the reorganization of the protein, leading to the expansion of the binding pocket, correlates with the properties of the substituents in 2.1-2.6, 3.1, and 3.2 attached to the benzimidazole or the dihydro-imidazoquinazoline scaffold (4.1 and 4.2). The primary amide stabilizes the closed conformation found in most TGT-ligand complexes. The extension of the amide by an additional amino group to a hydrazide leads to the structural transformation. The attached terminal amino group of the hydrazide exerts steric pressure on Asp156 and thus induces the protein rearrangement leading to its stabilization. The guanidine-substituted ligands show the same reorganization as a combined effect of steric pressure and altered solvation of the amino acids Gly230 and Gln203. The water molecule W1 and the slight spatial shift of Asp156 are responsible for this effect. In the complexes with the dihydro-imidazoquinazolines 4.1 and 4.2 the steric pressure is exerted by the attachment at C8. Obviously, solely a hydrogen as substituent is sufficient to trigger the pocket opening.

The newly formed cavity is occupied and stabilized by several water molecules, which indicates the hydrophilic character of the created pocket (Figure 2.24). A complex water network with short distances between 2.5 and 3.5 Å is detected involving the carbonyl groups of the Glu157, Thr159, Ala168 and Gly230 and the backbone amide nitrogens of Gly204, Ser205 and Thr159 which protrude toward the expanded pocket. In addition, the side chains of Gln203, Ser171 and Ser205 form H-bonds to the accommodated water molecules.

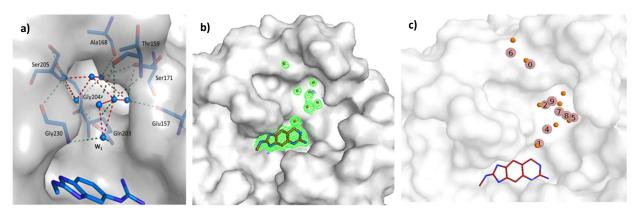


Figure 2.24: a) Binding pose of **3.1** together with the opened transient pocket, which is filled with a cluster of water molecules. The protein is displayed by its solvent accessible surface and the residues involved in interactions with the water cluster are indicated, hydrogen bonds as dashed lines. b) **4.2** together with the difference electron density (m|Fo|-|Fc| omit map) contoured at 3σ , protein displayed with the solvent accessible surface showing the recognition site together with the opened transient pocket which is filled by water molecules; c) comparison of the spatial arrangement of computed hydration sites (numbered pink balls, Table 2.6) together with the crystallographically determined waters in TGT-**4.1** (orange spheres, cf. b).

Depending on the achieved resolution, the amount of water molecules captured by crystallography will differ. ⁸⁶ Furthermore, crystal structures do not provide information about the thermodynamic properties of the water molecules found in such the transient pocket. However, upon ligand binding to the protein, the solvent molecules occupying the pocket will be displaced to the surrounding bulk

solvent phase. To estimate the contribution of the water displacement to the overall inventory of ligand binding, the energetics displacing the waters are important. Waters displaced from hydration sites that are energetically less favorable than in the bulk, will contribute to ligand binding. While displacement from hydration sites that are more favorable in the protein than in the bulk, will counteract binding. We therefore performed MD simulations to perform a hydration site analysis (SSTMap calculation)⁸⁷ to determine the enthalpic and entropic contributions of the bound water molecules and the corresponding Gibbs free energy contribution for their release to the bulk phase (Table 2.6). Our calculations suggest that individual water molecules cluster at representative sites which reasonably well reproduce the crystallographically determined water sites (Figure 2.24b,c). Thus, the crystal structure of TGT-4.1 reveals the presence of eight water molecules, while MD simulations suggests a similar cluster nine of water molecules. The distances between the assigned crystallographic water molecules and the centers of the computed water clusters are listed in Table 2.6.

Table 2.6: Hydration site analysis of the 9 most prominent water clusters found in the transient recognition pocket by molecular dynamics.

Hydration Site	Distance to closest crystal water [Å]	ΔE _{solv} [kJ·mol ⁻¹]	T∆S _{solv} [kJ·mol ⁻¹]	ΔG _{solv} [kJ·mol ⁻¹]
9	0.5	12.51	-16.40	28.91
1	0.89	5.23	-22.64	27.87
4	1.37	1.42	-19.00	20.42
0	0.68	-1.17	-21.34	20.13
2	0.72	-3.18	-22.38	19.16
5	0.73	-2.85	-20.29	17.49
7	1.08	-3.31	-16.95	13.64
8	2.3	-5.27	-18.24	12.97
6	1.55	0.63	-12.30	12.93

The HSA calculations suggest that all the waters in the transient pocket have unfavorable Gibbs free binding energies with respect to the bulk phase. Hydration sites 9 and 1 are energetically highly depleted, suggesting unfavorable interactions in the transient binding pocket compared to bulk solvent. We therefore expect that their displacement favorably contributes to ligand binding. This hypothesis is supported by the observation that the ligands which open the transient pocket and displace some of the water molecules in the pocket (e.g. **4.2**) are surprisingly potent compared to the members of series **2** and **3**.

2.3.7 Functional Role of the Transient Pocket and Outlook for Further Inhibitor Design

We inspected the *Z. mobilis* TGT complex structures, deposited in the protein data bank (PDB, www.rcsb.org). ⁸⁸ The multiple structures show that even though the loop between Glu157 and Thr161 is structurally in most complexes spatially highly conserved, the side chain of Cys158 occurred already in a few structures with alternative conformations. In these examples, the Cys158S $_{\gamma}$ - Val233C $_{\alpha}$ distances vary from 6.6 Å to 9.8 Å. However, the examples with expanded distances are predominantly mutated variants of TGT in the guanine/preQ $_{1}$ recognition pocket. These variants were produced to investigate the differences of the eubacterial and eukaryotic TGTs. ⁶⁵ The latter enzymes recognize queuine as significantly expanded substrate. In the human enzyme as in most other eukaryotic species,

Cys158 is replaced by Val and the opposing Val233 is exchanged to Gly. Remarkably, the crystal structure of queuine (PDB code: 3BLO⁶⁵) with this variant shows a similar expansion of the recognition pocket as observed in complexes studied here. Possibly, the latter complexes indicate an intrinsic conformational flexibility the architecture of TGT enzymes exhibits in this area that is functionally important for the eukaryotic species to recognize its extended substrate.

The pronounced similarity of the recognition site in eubacterial variant, which hosts the nucleobases guanine and preQ₁, and the eukaryotic one means a real challenge developing species-selective inhibitors. However, one observation with our expanded inhibitors might provide a promising opportunity to approach this problem. Remarkably, the Cys158 residue is highly conserved across all eubacterial enzymes but lacking in the eukaryotic species (Figure 2.4). The thiol group of a cysteine residue is a potential anchor to covalently attach an inhibitor, e.g. via a Michael acceptor or a potentially alkylating group. For the development of a potent anti-infective, irreversible inhibitors may constitute a desired strategy to follow. The ligands studied here based on the 4-guanidino benzimidazoles or the C(8)-functionalized dihydro-imidazoquinazoline scaffold induce the opening of the transient pocket and make the thiol group of the crucial cysteine accessible for chemical modification. Via appropriate design, it may be possible to develop inhibitors, preferentially based on the latter likely more potent dihydro-imidazoquinazoline scaffold, that selectively and irreversibly block the bacterial TGT isoforms via a warhead capable to covalently bind to the thiol group of Cys158.

2.4 Materials & Methods

2.4.1 Expression and Purification of the Z. mobilis TGT

The expression and purification protocol was adapted as previously described by Jakobi et al.⁶⁷ The E. coli cells BL21-CodonPlus (DE3) -RIPL (Cam r), transformed with the plasmid vector pPR-IBA2-ZM10 (Amp r), were incubated in a pre-culture of 100 mL LB medium containing 100 mg · L⁻¹ ampicillin and 34 mg \cdot L⁻¹ chloramphenicol for 17 h at 37 °C and 220 rpm. In addition to ampicillin and chloramphenicol resistance, the plasmid contained a sequence encoding for an N-terminal Strep-tag II® separated from the tgt start codon by a spacer sequence and a sequence encoding a thrombin cleavage site. The preculture was added to 2×2 L main culture (LB medium including 100 mg \cdot L ⁻¹ ampicillin and 34 mg \cdot L⁻¹ chloramphenicol) which is incubated at 37 °C and 220 rpm until the OD_{600} = 0.7. This main culture was then cooled to 15 °C and the protein expression induced by addition of IPTG (final concentration 1 mM). The main culture was then incubated at 15 °C and 220 rpm for a further 16-18 h. Afterwards, the cell pellets were harvested by centrifugation (10,000 rpm at 4 °C). They were subsequently resuspended in 100 mL lysis buffer (20 mM TRIS pH 7.8, 10 mM EDTA, 1 mM DTT and 2 cOmplete™-Protease Inhibitor Cocktail Tablets (Roche) per 4 L of bacterial culture) and cell disruption was achieved via three rounds of sonification using a Branson Sonifier 250, with 90 seconds intervals (duty cycle 30%, output control 7). Alternatively, cell disruption was achieved via an EmulsiFlex-C5™ high-pressure homogenizer (Avestin Europe GmbH). The soluble protein in the supernatant was then separated by centrifugation from the insoluble cell constituents in the pellets (centrifugation speed 19,000 rpm, 45 min, 4 °C). Purification of the protein was achieved at room temperature by two FPLC steps using an ÄKTA Purifier LC system. A Q-Sepharose Fast Flow Anion Exchange Column (XK 26/15; GE Healthcare) was equilibrated with buffer A (10 mM TRIS pH 7.8, 1 mM EDTA, 1 mM DTT) and the clear cell lysate was passed through the column. The protein was then eluted by a buffer B containing 10 mM TRIS pH 7.8, 1 mM EDTA, 1 mM DTT, 1 M NaCl through gradient elution (a linear increase in the proportion of buffer B (from 0 - 100% at 4 mL· min -1 column flow) and fractioning of the eluted portion. Fractions containing the target protein Z. mobilis TGT with the Strep tag II were determined by SDS-PAGE. A Strep-Tactin Superflow column (XK 16/10, IBA) was equilibrated with buffer W (100 mM TRIS pH 7.8, 1M NaCl, 1 mM EDTA) and the corresponding fractions were passed through the column. The target protein was eluted by buffer E (100 mM TRIS pH 7.8, 1 M NaCL, 1 mM EDTA, 2.5 mM D-desthiobiotin). The fractions containing TGT protein were then concentrated in a VIVASPIN $^{\circ}$ 20 centrifugal concentrator (Sartorius, MWCO = 30,000) to a concentration of approximately 2 mg \cdot mL $^{-1}$ in a high salt buffer containing 10 mM TRIS pH 7.8, 2M NaCl, 1mM EDTA. Subsequently, the Strep-tag $^{\circ}$ II was cleaved off and separated from the TGT protein via a Thrombin Cleavage Capture Kit (Novagen $^{\circ}$) following the manufacturer's instructions where 2.5 U of biotinylated thrombin per mg TGT protein was incubated with the TGT protein for 16-18 h at 20 °C. The cleaved TGT protein was separated from the Strep-tag $^{\circ}$ II, the biotinylated thrombin from the kit, and the streptavidin-agarose beads by filtration using the filters of the kit. The separated TGT protein was then dialyzed against high salt buffer and concentrated via VIVASPIN $^{\circ}$ 20 centrifugal concentrator until a final protein concentration of 12 mg \cdot mL $^{-1}$. Finally, the protein was flash frozen into aliquots of 70 µL and stored at -80 °C.

SDS-PAGE. The purity of *Z. mobilis* TGT was checked with SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis). The SDS gels consist of two components; the Collecting Gel and the Separating Gel.

Composition of the Collecting Gel and Separating Gel (amount adds up to 9 SDS gels) are shown in Table 2.7.

Table 2.7 Composition of SDS Separating and Collecting gels.	Table 2.7	Composition	of SDS S	eparating	and Colle	cting gels.
--	-----------	-------------	----------	-----------	-----------	-------------

Reagent volume in Separating Gel	Reagent volume in Collecting Gel
Demineralized water 6.26 mL	Demineralized water 18.7 mL
1M Tris pH 8.8 13.33 mL	1M Tris pH 6.8 6 mL
10% SDS solution 0.4 mL	10% SDS solution 0.3 mL
Rotiphorese® Gel 30 (Acrylamides: N,	Rotiphorese® Gel 30 (Acrylamides: N,
NMethylenbisacrylamide 37.5: 1) 20 mL	NMethylenbisacrylamide 37.5: 1) 5 mL
TEMED 40 μL	TEMED 30 μL
10% APS 400 μL	10% APS 300 μL

The Separating Gel was prepared first and poured into an SDS gel cast. 1 mL of isopropanol was added to cover each gel. After the Separating Gel had hardened, the isopropanol was removed and the Collecting Gel was poured onto the Separating Gel and the combs were inserted. After the Collecting Gel also hardened, the SDS gels were separated and individually wrapped in moist paper towels and stored in the fridge. To run an SDS-PAGE, 2.5 μL of 4x SDS sample buffer was added to 10μL sample solution, vortexed, and heated at 95°C for 5 min. The samples were centrifuged and loaded onto the SDS gel in a Bio-Core electrophoresis chamber. As a reference, PageRuler™ Prestained Protein Ladder (SM0671 / 2) was used. LaemmLi buffer was used to cover the SDS gel, 130V was applied, and the gel left to run. Afterwards the gel was left to stain overnight in Coomassie Blue staining solution. The gel was de-stained the next day in Coomassie Blue decolorization solution, whereupon the protein bands became visible.

2.4.2 Crystallization and Structure Refinement of Z. mobilis TGT

Z. mobilis TGT apo crystals were grown at 18 °C using the sitting-drop and hanging-drop vapor diffusion methods. 1.5 μ L of 12 mg \cdot mL⁻¹ were mixed with 1.5 μ L of reservoir solution containing 100 mM MES pH 5.5, 1 mM DTT, 13% (w/v) PEG8000, 10% (v/v) DMSO in the wells of a crystallization plate containing 650 μ L reservoir solution. Crystals could be seen within three days. Crystals were soaked into a solution containing 100 mM of fragment solution (dissolved in 100% DMSO), 17.5% PEG3350, 25% PEG400, 200

mM NaCl, for a time ranging between 3 minutes and 20 hours depending on crystal stability to the fragment solution. Crystals did not need a cryo-buffer as the amount of PEG was sufficient, so they were directly flash-frozen in liquid nitrogen.

Z. mobilis TGT co-crystals were grown at 18°C using the sitting-drop vapor diffusion method. A solution of 0.23 mM of *Z. mobilis* TGT (stored in high salt buffer) was mixed with 6.6 mM of the respective inhibitor (stock solution in 100% DMSO) and incubated at 18°C for 1 hour. The solution was then centrifuged and 1.5 μL of this mixture was mixed with 1.5 μL of reservoir solution (100 mM MES pH 5.5, 1 mM DTT, 13% (w/v) PEG8000, 10% (v/v) DMSO) in the wells of a crystallization plate containing 650 μL reservoir solution. Co-crystals could be seen within three days up to one week. They were transferred into cryo-buffer (50 mM MES pH 5.5, 0.5mM DTT, 300mM NaCl, 2% (v/v) DMSO, 4% (w/v) PEG 8000, 30% (v/v) glycerol) for several seconds and flash-frozen in liquid nitrogen.

Data collection and refinement statistics are shown in the Appendix (Table 1- Table 5). The crystallographic tables for PDB codes 5J9M, 5JT5, 5J9N, 5JT6, 5J9O, and 5JT7 were taken from the dissertation of Dr. Frederik R. Ehrmann, 2016. The diffraction data were indexed, scaled, and merged using XDS⁹⁰ and XDSAPP.⁹¹ Molecular replacement from the program PHASER MR⁹² from the CCP4 suite⁹³ was used to determine all crystal structures. The structure 4LBU was used as a search model. In the refinement, a 5% subset of all reflections was omitted during refinement to be used for Rfree calculation. Model building was achieved in COOT⁴⁴ and refinement using PHENIX.refine version 1.10.1-2155.94 Cartesian simulated annealing with default parameters was used as a first refinement step for all the structures. This was followed by refinement of XYZ coordinates and occupancies of protein residues and fragments (with the exception of water molecules whose occupancies were fixed). In the case of protein residues that gave additional density, they were refined in double confirmation and kept if their refined occupancy was ≥ 20%. The structure of TGT in complex with J14 (PDB: 6FSO) was refined isotropically with 6 TLS groups. The structures of TGT in complex with J41 (PDB: 5SW3), TGT in complex with J72 (PDB: 5UTI), TGT in complex with J79 (PDB: 5UTJ), and TGT in complex with J86 (PDB: 5V3C) were all refined anisotropically except for water molecules. The structure of TGT in complex with J64 (PDB: 5N6F) was refined anisotropically. Hydrogen atoms were added to the refined structures in the last refinement step in PHENIX.refine. For two of the structures; TGT in complex with J19 and TGT in complex with J33, the resulting m|Fo|-|Fc| density maps could not identify the fragments. Only after observing the event maps created by PanDDA (Pan-Dataset Density Analysis) could the fragments be identified. As these two structures do not meet the criteria for PDB depositions, they were not deposited. For the structure of TGT in complex with J79 (PDB: 5UTJ) the average B-factor of the ligand remains high despite refining occupancy. The structure of TGT in complex with ligand 4.1 (PDB: 6RKT) and TGT in complex with ligand 4.2 (PDB: 6RKQ) was refined isotropically with 6 TLS groups. Details of structural refinements of TGT in complex with ligand 2.1 -2.6, 3.1, and 3.2 can be found in the dissertation of Dr. Frederik R. Ehrmann, 2016. 5 Chemicalize 6 developed by ChemAxon⁹⁷ was used for name-to-structure generation and SMILES code notation. The ligand PDB and restraint files were generated with the Grade Web Server.98

2.4.3 Applied Buffers for the Screening Experiments

The conditions applied in the experiments were all close to neutral pH (pH 7 in soaking, Hepes and dTris buffers pH of 7.4 in SPR and NMR, respectively). TCEP is used in SPR and NMR as a reducing agent to prevent the formation of disulfide bridges in protein-protein interfaces. During crystal growths, TCEP was also added to the crystallization condition. Since soaking involves the use of pre-formed apo crystals, TCEP was not present in the soaking solutions. All screening conditions had high salt concentration. In SPR, the salt concentration is important to prevent non-specific binding to the dextran surface in the reference channels⁹⁹ while in NMR, this salt concentration is important to prevent nonspecific binding through non-specific ionic interactions.¹⁰⁰ A notable difference between the screening buffers is the concentration of DMSO. The soaking conditions had a concentration of

10% DMSO, in comparison to 1% DMSO in SPR and 0.2% in NMR. Higher DMSO concentrations will enhance fragment solubility and reduces the risk of precipitation. The applied SPR conditions require 0.05% Tween, which the soaking conditions did not contain. This may bring about a similar effect, even though Tween and DMSO unlikely function in the same way. While DMSO is used as co-solvent to enhance compound solubility, Tween is used to reduce compound aggregation.¹⁰¹ The fact that the NMR buffer did not contain a high percentage of DMSO explains why several fragments had to be excluded from the screen for limited solubility. PEG is used in the soaking conditions as a cryoprotectant, it is neither used in the SPR nor NMR screening buffers. Despite PEG being noninvasive and thus less likely to interact within the interior of protein crystals 102,35 it can still be found in some structures if the soaking conditions apply PEG in high concentration and PEG molecules have suitable size. The drawback of this is that a PEG molecule may even bind in the position where a fragment could accommodate, thus interfering with ligand binding. We found such a displacement of small PEG molecules in a study using crystals of endothiapepsin. Fragments need a certain potency to displace the PEGs. EDTA is a chelating agent added to coordinate metal ions. In SPR and crystallography it is not necessary to add EDTA but in NMR it is primarily added to chelate paramagnetic impurities. 104 The sodium salt of trimethylsilylpropanoic acid deuterated in the propionic part (TMSP-d4) is used as internal NMR reference. It contains nine equivalent methyl hydrogens and therefore its ¹H NMR spectrum consists of a sharp singlet which is set as chemical shift of 0 ppm. As most compounds studied by ¹H NMR spectroscopy show resonances downfield of the TMSP signal (especially organic compounds), there is usually no overlap with the signals of the samples.

2.4.4 SPR Conditions

The SPR studies have been performed at 20 °C using a BIAcore 3000 instrument for the clean screen, and a BIAcore T200 instrument for the fragment screen and determining binding affinities of ligands **4.1** and **4.2**. The *Z. mobilis* TGT was immobilized via amine coupling to a dextran sensor chip at a pH of 5.5. For the BIAcore 3000 experiments, a Xantec CMD 500L sensor chip was used and for the BIAcore T200 experiments a BIAcore CM7 sensor chip was used. The functional groups of the Xantec CMD 500L chip surface were first activated by injecting for 10 min at 10 μ L/min with a 1:1 mixture of 0.5 M 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and 0.5 M N-hydroxysuccinimide (NHS), immediately followed by injecting the TGT protein for 7 min at a flow rate of 10 μ L/min to 2623 RU. Remaining activated carboxyl groups on the surface were blocked with 4 x 1.5 min pulses of 0.5 M ethanolamine or 150 mM ethylenediamine to overcome sticking of compounds to the reference channel. Immobilization levels achieved on the CM7 sensor chip were 9400 RU for the fragment screen and 4200 RU for the binding affinities measurements. The running buffer used for the immobilization was 10 mM Hepes buffer pH 7.4, 0.05% Tween 20, 150 mM NaCl, 1 mM TCEP, 1% DMSO.

Before the final immobilization of the protein, a pH scouting was performed to find the appropriate pH conditions for TGT coupling onto the surface (the pH of the buffer in which the protein will be diluted). The running buffer used for pH scouting was 10 mM Hepes buffer pH 7.4, 0.05% Tween 20, 150 mM NaCl, 1 mM TCEP, 1% DMSO. To find the appropriate pH for TGT immobilization, the pI of the protein was considered. The selected pH must be below the TGT pI of 6.3 but above the pKa values of 3.5 of the carboxylates in the dextran matrix. A 50 μ g · mL⁻¹ stock solution of TGT was prepared in buffers of 10 mM sodium acetate at various pHs of 5.5, 5.0, 4.5, and 4.0, respectively. The protein samples in different buffers were injected onto the activated sensor chip surface separately in 20 μ L volumes at a flow rate of 10 μ L · min ⁻¹. Before each injection the sensor chip was preconditioned with a solution of 50 mM NaOH and 1 M NaCl at 10 μ L · min ⁻¹ for 2 minutes. The corresponding curves and response values were monitored and the suitable pH of 5.5 was selected. A tool compound is used to test the correct immobilization of a protein to the sensor chip, and to also test the protein activity after consecutive uses of the sensor chip. A tool compound should have a known binding affinity that gives a measurable response on the sensorgram. Inhibitor 1 with a K_D of 68 ± 5nM was used as the tool

compound.

Clean Screen

A clean screen with the fragments alone in the running buffer without protein was done to exclude the problematic fragments from the binding level screen. For this clean screen, the fragments were prepared in a 96-well plate in 2 mM concentrations by diluting them in the running buffer. The running buffer was 10 mM Hepes, pH 7.4, 0.05% Tween 20, 150 mM NaCl, 1 mM TCEP, and 1% DMSO. The plates were spun down, and half of the volume was transferred into a new plate and diluted again 1:1 with the running buffer to give a final fragment concentration of 1 mM. The fragments were then passed over the CMD 500L sensor chip at a flow rate of 20 μ L/min for 1.5 min. To remove compounds from the biosensor surfaces, the flow channels were regenerated by a 90 s pulse of 0.5% SDS. Fragments with irregular curves (very slow dissociation rate, jump in RU) were excluded from the fragment screening.

Binding Level Screen (Fragment Screen)

The fragments were screened at 1 mM concentrations, in a fragment binding level screen, and a standard solvent correction procedure was performed to correct for potential deviations in DMSO concentration between samples and running buffer. The running buffer used was 10 mM Hepes, pH 7.4, 0.05% Tween 20, 150 mM NaCl, 1 mM TCEP, and 1% DMSO. To remove compounds from the biosensor surfaces the flow channels were regenerated after each cycle with 50% DMSO (a 1:1 dilution of running buffer and DMSO).

Affinity Screen (Determining Binding Kinetics)

For fragments, the affinity screen was done using 9 different concentrations and 3 blanks as follows: 0 μ M, 0 μ M, 0 μ M, 20 μ M, 35.6 μ M, 63.2 μ M, 112 μ M, 200 μ M, 356 μ M, 632 μ M, 1125 μ M, 2000 μ M. The running buffer used was 10 mM Hepes buffer pH 7.4, 0.1% Tween 20, 250 mM NaCl, 1 mM TCEP, and 2% DMSO.

Single Cell Kinetics was used to determine the binding affinity of ligands **4.1** and **4.2**. The direct binding screen for was done at 1/3 log concentrations of 1 μ M, 3.16 μ M, 10 μ M, 31.6 μ M, and 100 μ M. Due to the sticky nature of the ligands **4.1** and **4.2**, an exaggerated response could be observed at a concentration of 100 μ M so kinetic analysis was not possible. Instead, the apparent K_D value was extrapolated from the saturation points, where the last point at 100 μ M was not considered. For ligand **4.1**, another direct binding screen was done at concentrations of 3 μ M, 10 μ M, 30 μ M, 100 μ M and 300 μ M to reach saturation but in both experiments secondary effects could be seen at the higher concentrations and complete saturation could not be achieved, thus the exact K_D value cannot be determined. Hence, we only report from the SPR measurements that the K_D value for ligand **4.1** lies in the double-digit micromolar range. The running buffer used was 10mM Hepes buffer pH 7.4, 0.1% Tween 20, 250 mM NaCl, 1 mM TCEP, and 2% DMSO.

2.4.5 NMR Conditions

NMR experiments for fragment screening were performed at 25 °C using a Bruker Avance Neo 600 MHz spectrometer equipped with a cryogenically cooled probe-head. Automatic sample changing was accomplished with a Bruker SampleJet system. For the initial validation experiments (Figure 2) a Bruker

Avance III HD 800 MHz spectrometer equipped with a cryogenically cooled probe-head, was used. NMR samples contained 200 μ M fragment in aqueous buffer with 20 mM deuterated TRIS at pH 7.4, 1 mM EDTA, 150 mM NaCl, 10% (vol/vol) D₂O, 3 mM TCEP and 10 μ M TMSP-d11. A TECAN EVO 100 liquid handling robot was used to fill 5 mm O.D (outer tube diameter) NMR tubes in a SampleJet rack. The interscan delay was 3 s and 64 scans were accumulated. Solvent suppression was accomplished with an excitation sculpting scheme¹⁰⁵ using 2 ms sinc shape flip back pulses. In the CPMG experiment the DMSO peak was suppressed by off resonance irradiation during the interscan delay. For each compound the binding experiment was run in two or three steps. First CPMG and WaterLOGSY experiments were run on a sample containing the ligand alone to check solubility and integrity. Once the structure was confirmed from matching the spectra peaks to the predicted fragment peaks in the ACD/Labs software, 3 μ M protein was added to check binding. For the samples with fragments that did bind, 10 μ M of a potent inhibitor, inhibitor 1 (SPR K_D of 68 ± 5 nM), was gradually added to the samples and a third set of NMR experiment were recorded. Inhibitor 1 binds to the active site and can therefore be used to identify specific active site binders. Spectra were processed and analyzed with TopSpin*.

2.4.6 MD Simulations and HSA Calculations.

MD Simulations. For the MD simulations we used the protein-ligand complex of TGT-4.1 (pdb-code 6RKT). The structure was loaded into MOE¹⁰⁶ and the ligand molecule was removed. Then missing atoms were built and protonation states were assigned using the protonate3D utility of MOE. The cysteine residues coordinating the Zn²⁺ ion, were modelled in their deprotonated form. Then, the protonated protein structure, crystallographic water molecules and the zinc ion were loaded into *tLEaP* from the AmberTools17 package.¹⁰⁷ In *tLEaP*, parameters from the amber *FF99SB*¹⁰⁸ force field were assigned to the protein atoms and TIP4P-Ew¹⁰⁹ parameters were assigned to the water molecules. Two sodium counter ions were added using the *addions2* utility in order ensure net neutrality through the MD simulation. Finally, the protein was embedded into a truncated octahedron simulation box filled with TIP4P-Ew water molecules. The box was built such that the minimum distance between each solute or crystallographically determined water molecule and any box edge was at minimum 10 Å, resulting in a simulation box with 12756 water molecules in total.

In the following, all minimization operations were carried out using *pmemd* and all MD simulations were calculated using the *pmemd.cuda*^{110–112} for use with GPUs. Both programs were, as well as all other programs from the Amber program package, used from Amber16 together with AmberTools17.¹⁰⁷

Initially, the system energy was minimized with 2500 steps of steepest descent minimization followed by 2500 steps of conjugate gradient minimization. During this minimization operation, all solute heavy atoms were restrained to their crystallographically determined positions using a harmonic restraining potential with a force constant of 25 kcal·mol⁻¹·Å⁻². Subsequently, the force constant was set to 2 kcal·mol⁻¹·Å⁻² and the minimization was repeated analogously to the first one. In the next step, the system was heat to 300 K within 25 ps using an integration time step of 1 fs. At this point, the solute heavy atoms are again positionally restrained with a force constant of 25 kcal·mol⁻¹·Å⁻². Now, the integration time step was switched to 2 fs and the system equilibrated under NPT conditions for 5 ns using the Berendsen barostat¹¹³ and a target pressure of 1 bar. The temperature was regulated at 300 K using a Langeving dynamics thermostat with a collision frequency value of $\gamma = 2$ ps⁻¹. In the final equilibration step, the system is run under NVT conditions for an additional 5 ns.

A final production run was carried out for 30 ns and coordinates were saved do disk every 2 ps.

During all MD runs, the system was treated using periodic boundary conditions and the particle-mesh Ewald technique together with a 9.0 Å real-space distance cutoff for the electrostatic interactions. Furthermore, all bonds involving hydrogen atoms were constrained using the SHAKE algorithm.¹¹³

Hydration Site Calculations. For the HSA calculations, we used the program SSTMap^{87,114} (Version 1.1.1). We used ligand molecule TGT·**4.1** in its bound pose in order to define the binding site for the HSA analysis. After the HSA calculation, the hydration site energy values were manually referenced to the TIP4P-Ew bulk energy value.

Chapter 2 TGT

Chapter 3 Peroxin 14

Introductory Remarks

The following chapter is the preparation of a manuscript to be submitted.

The expression and the purification of both constructs of the *T. brucei* PEX14 were optimized and done by the author of this thesis. The set up and screening for crystallization conditions was done by Ralf Poeschke from the MarXtal facility at the Philipps University of Marburg in Marburg, Germany. The additive screening with the compound TEW was done in collaboration with Dr. Anna Aagaard during an academic scientific secondment at Astrazeneca R&D in Gothenburg, Sweden.

The virtual screening for novel compounds to inhibit PEX14 was done in collaboration with Ryan Byrne from the working group of Prof. Dr. Gisbert Schneider during an academic scientific secondment at ETH Zurich in Zurich, Switzerland.

The SOFAST HMQC 2D NMR experiments on the virtually screened ligands and the AlphaScreen assay were done in collaboration with Charlotte Softley and Roberto Fino, respectively, from the working group of Prof. Dr. Michael Sattler during an academic scientific secondment at the Helmholtz Zentrum in Munich, Germany.

The TSA (thermal shift assay) experiment trials were done by the author of this thesis.

The DLS (dynamic light scattering) experiment was done by Patrick Walter from the working group of Prof. Dr. Helene Munier Lehmann at Louis Pasteur in Paris, France.

Compound 1, designed from growing the fragment hit F5, was synthesized by Khang Ngo from the working group of Prof. Dr. Gerhard Klebe at the Philipps University of Marburg in Marburg, Germany.

The trypanocoidal assay was done by Vishal Kalel from the Ruhr-Universität Bochum in Bochum, Germany.

The crystallization of PEX14 and elucidation of the crystal structures was done by the author of this thesis.

Chapter 3 PEX14

3.1 Introduction

3.1.1 Neglected Infectious Tropical Diseases

Neglected tropical diseases (NTDs) are a group of chronic conditions affecting primarily the poorest 500 million people living in sub-Saharan Africa (SSA), where the extreme poverty helps spread the diseases.

These diseases include helminth infections, Schistosomiasis, and the protozoan infections; human African trypanosomiasis and visceral leishmaniasis, which affect around 100,000 people in areas of conflict in SSA and cause high mortality.¹¹⁵

3.1.1.1 Human African Trypanosomiasis

Epidemiology

The human African trypanosomiasis (HAT), also known as the sleeping sickness, is endemic in 36 sub-Saharan Africa countries, threatening millions of people. It is a vector-borne parasitic disease transmitted to humans by bites of the tsetse fly (*Glossina* genus). The parasite responsible for the infection is a protozoan parasite *T. brucei* belonging to the genus *Trypanosoma*. The tsetse fly can acquire the infection either from humans or animals that harbor the parasite. Only certain species of the tsetse fly transmit the disease and for reasons unknown, some regions where the tsetse flies are found do not have the sleeping sickness. However rural populations that depend on agriculture, fishing, animal husbandry or hunting are those most exposed to the tsetse fly and thus the disease. Other ways of disease transmission include transmission of the trypanosome from mother-to-child, mechanical transmission through other blood-sucking insects, accidental infections due to pricks with contaminated needles, as well as through sexual contact.

Since the last epidemic in 1970 until the late 1990s, the efforts of WHO, national control programs and non-governmental organizations (NGOs) have helped control the disease. The number of new HAT cases reported between 2000 and 2012 dropped by 73%. The estimated number of actual cases is below 10,000 and the estimated population at risk is 65 million people. Over the last 10 years, the Democratic Republic of Congo (DRC) has reported over 70% of the cases, and other countries have reported less than 100 new cases per year including Cameroon, Central African Republic, Nigeria, South Sudan, and United Republic of Tanzania.¹¹⁹

Forms of HAT

Depending on the parasite involved, HAT takes 2 forms¹¹⁸, their distribution is shown in Figure 3.1: 1- *Trypanosoma brucei gambiense* which accounts for 98% of reported cases, is found in 24 countries in West and Central Africa. It may take months or even years for a person with this chronic infection to show symptoms of the disease. This makes it difficult to cure as the patient is often then already in the advanced stage of the disease where the central nervous system is affected.¹¹⁹

2- *Trypanosoma brucei rhodesiense* is the less common form accounting for less than 2% of reported cases and is found in 13 countries in Eastern and Southern Africa. Contrary to the first form, the first symptoms in an infected person appear after a few weeks or months and the disease develops rapidly on from there, invading the central nervous system.¹¹⁹

The disease is also found in cattle, where it is termed *Nagana*. 116 Nagana is caused by other sub-species of the *Trypanosoma* genus that are pathogenic to both wild and domesticated animals. Animals can also host the human pathogen parasites. 119

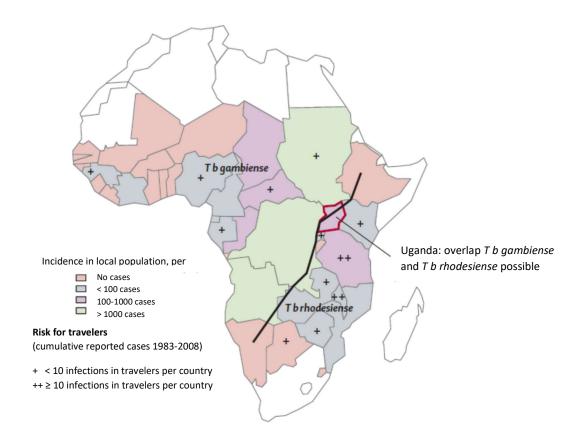


Figure 3.1: Distribution of human African trypanosomiasis with incidences and risk for travelers. The area above the black line is where *Trypanosoma brucei gambiense* predominates, whilst the area below the black line is where *Trypanosoma brucei rhodesiense* predominates (J Blum, Swiss Tropical Institute).

Stages of the disease

There are two major stages of the disease:

- 1- The haemo-lymphatic stage which involves fever, headaches, joint pains, and itching, occurs when multiplying trypanosomes are still in the subcutaneous tissues, blood, and lymph. 120,119
- 2- The neurological or meningo-encephalic stage which occurs when parasites cross the blood-brain barrier and infect the central nervous system. ¹²⁰ It is in this stage when the obvious symptoms of the disease begin to appear including confusion, behavioral changes, sensory disturbances, poor coordination, and of course the disturbance in the sleep cycle which coins the disease its name. ¹¹⁹ If left untreated it could lead to coma and death. ¹²⁰

Treatment

Treatment of HAT depends on the stage of the disease.

First stage treatments are safer and easier to administer than the second stage treatments, as the second stage treatments must cross the blood-brain barrier to reach the parasite and thus are toxic and complicated to administer. Assessment of treatment outcome requires a 24-month follow-up of the patient which includes laboratory examinations of cerebrospinal fluid by lumbar puncture to ensure there are no viable parasites that could reproduce the disease after the treatment.¹¹⁹

There are five drugs that have been used for treatment depending on the stage of the disease, in addition to fexinidazole which is a new oral treatment for *gambiense* HAT. It is already registered in

the Democratic Republic of Congo and has been received positively by the European Medicines Agency's Committee for Medicinal Products for Human Use in November 2018. 119

Drugs used in treatment of first stage:

- Pentamidine: used for treatment of *T.b. gambiense* sleeping sickness. Has undesirable side effects but well tolerated by patients. 119
- Suramin: used for treatment of *T.b. rhodesiense*. Has undesirable side effects including urinary tract and allergic reactions. ¹¹⁸

Drugs used in treatment of second stage:

- Melarsoprol: recommended as a second line-treatment of *gambiense* and first line-treatment of *rhodesiense* infections. It is derived from arsenic and so has several undesirable side effects including encephalopathic syndrome which can be fatal. There has been an observed increase in resistance especially in Central Africa.¹¹⁸
- Eflornithine: this complex treatment is only effective against *T.b. gambiense* and is difficult to apply. 118
- Nifurtimox: a combination treatment of nifurtimox and effornithine is used against *T.b. gambiense.* This combination makes it easier to use effornithine as it reduces treatment duration and number of necessary IV perfusions.¹¹⁹

3.1.2 Membrane Associated Protein PEX14

3.1.2.1 Role in Trypanosomes

Peroxisomes are membrane-enclosed organelles that compartmentalize a variety of metabolic reactions. Glycolysis in trypanosomes occurs within glycosomes, which are peroxisome-related organelles that contain metabolic enzymes including those required for glucose metabolism PEX1 as seen in Figure 3.2. PEX14 is a peroxin protein that is essential for biogenesis of glycosomes. PEX14 forms an essential protein-protein interaction with PEX5, an import receptor which transports cytoplasmic glycosomal enzymes into the organelle. PEX5 has a C-terminal tetratricopeptide repeat (TPR) domain that recognizes a peroxisomal targeting signal (PTS) peptide motif in cargo proteins and binds to it. The N-terminal of PEX5 on the other hand has diaromatic WxxxF peptide motifs that PEX14 recognizes through its globular N-terminal domain. PEX5 docks onto PEX14 in the peroxisomal transmembrane and together they form a complex which imports cargo protein.

Dependence of trypanosomes on glycolysis and thus the importance of the formed PEX14/PEX5 complex depends on the host of the deadly *T. brucei* parasite. When *T. brucei* is in its mammalian host - or its bloodstream form, it receives a constant millimolar level of glucose as it lives extracellularly in the bloodstream and cerebrospinal fluid. Therefore ATP is generated through glycolysis and the majority of proteins in the glycosome are thus glycolytic enzymes. The exact process of glycolysis in this stage is described in Figure 3.2. ATP is generated not in the glycosome, but rather the cytosol where pyruvate kinase transfers the high energy phosphate from phosphoenolpyruvate to ADP. When *T. brucei* is in its insect host - or its procyclic form, glucose of the digesting bloodmeal of the tsetse flies metabolizes rapidly, leaving amino acids instead as the primary nutrient for the parasites. The procyclic form is procyclic form, glucose of the digesting bloodmeal of the tsetse flies metabolizes rapidly, leaving amino acids instead as the primary nutrient for the parasites.

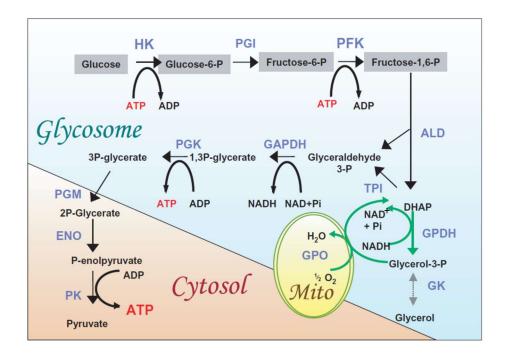


Figure 3.2: The glycosome, cytosol, and mitochondria (Mito) are all involved in the energy metabolism of bloodstream form *T. brucei*. The flow of metabolites of glucose is indicated by the black arrows. ATP generated in the glycosome is in balance as two equivalents of ATPs are required to start the metabolic process, whilst the net ATP is generated in the cytosol. The mitochondria is also involved in the glycerophosphate shunt where dihydroxyacetone phosphate (DHAP) is cycled to glycerol-3-phosphate and back, allowing the glycosome to maintain the NAD+/NADH balance. (Parsons, M. et al. *Molecular Microbiology* (2004)).

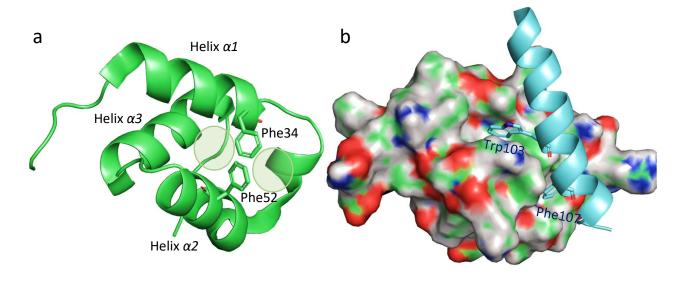


Figure 3.3: a) PDB: 2W84 Structure of NTD of PEX14 consists of a three helical bundle, where the PEX5 binding site is formed by helices $\alpha 1$ and $\alpha 2$ and has two hydrophobic pockets, shown as green circles, separated by two aromatic residues (Phe35 and Phe52) shown as sticks. b) PEX5 (blue) binds diagonally across helices $\alpha 1$ and $\alpha 2$ in PEX14 and buries its conserved aromatic residues Trp103 and Phe107 into the two hydrophobic pockets in PEX14.

3.1.2.2 Quaternary Structure of PEX14

The N-terminal domain of PEX14 is a 7 kDa monomer consisting of a three helical bundle (Figure 3.3a). The $\alpha 1$ and $\alpha 2$ helices are antiparallel, whilst the $\alpha 3$ helix forms a diagonal scaffold. The binding site of PEX14 with PEX5 is formed by helices $\alpha 1$ and $\alpha 2$ and has two hydrophobic pockets, separated by two

aromatic residues (Phe35 and Phe52). The PEX14/PEX5 complex is formed when PEX5 binds diagonally across helices $\alpha 1$ and $\alpha 2$ in PEX14 (Figure 3.3b), pointing its conserved aromatic residues Trp103 and Phe107 into the two hydrophobic pockets in PEX14, where they are deeply buried.

3.1.2.3 How to Disrupt Function of PEX14

T. brucei lacks feedback allosteric regulation of early steps in glycolysis but compartmentalizes the relevant enzymes within low permeable organelles called glycosomes. This compartmentalization separates enzymes and metabolites in the cytosol and glycosome from one another, which prevents toxic accumulation of intermediates. 121,127 This hypothesis has been proven through PEX14 RNA interference studies which confirmed that glucose becomes toxic to T. brucei upon glycosomal import disruption. ^{127,128} This was done by showing that in PEX14 deficient trypanosomes, addition of glucose led to accumulation of glucose-6-phosphate due to the impairment of glycosomal protein import in these trypanosomes. Additionally, loss of glycosomal compartmentation led to dramatic increases of glycerol-3-phosphate upon addition of glycerol due to the depletion of glycerol kinase which also saved the cells from glycerol toxicity.¹²⁷ It has been reported before that disrupting the PEX14/PEX5 interaction leads to the accumulation of glycosomal enzymes in the cytosol, depletion of ATP, and glucose toxicity which leads to the metabolic collapse and ultimately death of *T. brucei*. ¹²⁹ This disruption can be achieved through small molecules that bind to PEX14 and block it, preventing the binding of PEX5. One way of finding such molecules is through fragment screening. A previous NMR screening of a fragment library resulted in fragment hits that bind to PEX14.¹³⁰ In this project, we attempt to validate these hits through X-ray crystallography which will allows us to visualize the interactions of these fragments. The promising fragment hits can then be optimized into more potent lead compounds.

Selectivity can be achieved for *T. brucei* PEX14 as there are structural differences between the human PEX14 and the trypanosomal PEX14, as seen in Figure 3.4. The overall folds are very similar, except that the PEX14 of *T. brucei* exhibits an additional C-terminal helix. There are also some amino-acid differences in the PEX5 binding pockets of both species, whereas the human PEX14 has Leu28, Thr31, Lys34, and Asn38, the trypanosome PEX14 has instead an Arg, Asn, Glu, and Asp, respectively. This suggests that specific polar interactions can be used to selectively target *T. brucei* PEX14.¹²⁹ The hydrophobic binding pockets in both species however, are separated by two phenylalanine residues.

a

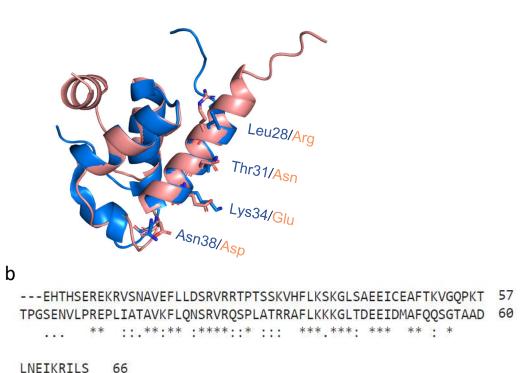


Figure 3.4: a) Superimposition of the human PEX14 (PDB: 2W84) in blue and the *T. brucei* PEX14 in salmon shows structural differences including amino acid residues surrounding PEX5 binding pocket in addition to the additional C-terminal helix shown in *T. brucei* PEX14. This indicates that selective inhibitors can be designed for *T. brucei* PEX14. b) Sequence alignment of human and *T. brucei* PEX14.

3.2 Crystallization and Fragment Screening of Mutated PEX14

3.2.1 Expression Constructs

EPSSLG---

66

Wild-Type Vs. Mutated PEX14 T. Brucei

Given the nature of PEX14 being a transmembrane protein, the entire construct is difficult to express and crystallize due to the high flexibility of the amino acid residues. ¹³¹ Additionally, our PEX5 interaction of interest occurs at the N-terminal domain, and for these reasons we have used a truncated version of the full construct where we have taken the residues 19-84 of the PEX14 N-terminus (Figure 3.5a). In our project, we worked with two different constructs of this N-terminal domain; the wild-type construct with a glutamate residue as the first residue, and a mutated construct E1W with a tryptophan residue as the first residue (Figure 3.5b). The purpose of the tryptophan was to render 280 nm absorbance properties to the construct, as the short 66 residues did not contain any 280 nm absorbing amino acids which makes it difficult to handle the protein in purification and standard protein concentration measurements. After purification a GAM expression tag (Gly-2, Ala-1, Met0) sometimes remains.

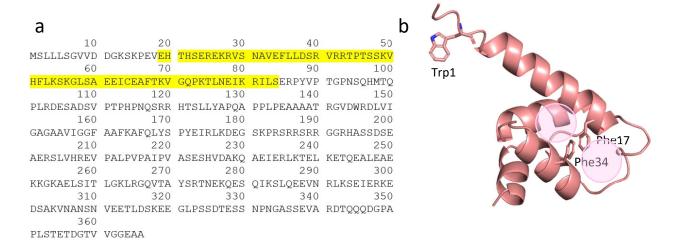


Figure 3.5: a) *T. brucei* Pex14 full sequence consists of 366 amino acids, whereas the truncated N-terminal domain consist of only 66 amino acids, as highlighted in yellow. b) First residue was mutated from a glutamate to a tryptophan. The first residue was selected as it is far from the encircled binding sites where the phenylalanine residues separate the binding pockets.

3.2.2 Crystallization

After expression and purification of the mutated construct, crystallization screens using Hampton Research plates were set up at a protein concentration of 17 mg/mL at 18°C and 4°C. From the two subsequent hits, one of them was reproducible inhouse. To obtain single crystals that did not grow in clusters 5% of glycerol was added as suggested by literature to reduce the nucleation rate and crystal growth¹³² (Figure 3.6).

Table 3.1: Crystallization Conditions from screening for the mutated PEX14 construct. Only one of the conditions was reproducible.

Protein	Concentration	Buffer	Salt	Precipitant	Temp	Reproducible
Mutated	17mg/ml	nono	0.2M Ammonium	30%(w/v)	4°C	Yes
PEX14	171118/1111	none	Sulphate	PEG 8000	4 0	res
Mutatod		0.1M	0.01M Cobalt	1.8M		
Mutated PEX14	17mg/ml	MES pH	Chloride	Ammonium	4°C	No
FEX14		6.5	Chioride	Sulphate		





Unit Cell Dimensions						
Α	b	С	α	β	γ	
43.2 43.2 81.0 90.0 90.0 90.0						

1 molecule/ asymmetric unit tetragonal P4,2,2

Figure 3.6: **Crystal form 1**. Crystals of E1W PEX14 grown in the a) absence of glycerol where they grow in clusters and in the b) presence of 5% glycerol where they grow as single formed crystals, which are more suitable for soaking experiments. Unit cell dimensions of the crystal are also shown in addition to the space group.

3.2.3 Fragment Screening

In a previous STD NMR fragment screen done at the Helmholtz Zentrum in Munich, a set of fragments were discovered that bind to E1W PEX14¹³⁰ (Table 3.2). Apo crystals of the protein (Figure 3.6b) were soaked into these fragments at a final fragment concentration of 100 mM for a time ranging between 20-24 hours. Among these 11 fragments, only one was found to bind in the crystal structure (crystal form 2, Figure 3.7a). Additionally, this fragment does not address either of the two known hydrophobic pockets of PEX14. To further evaluate the reason behind this we ran the structure through a script developed by Dr. Alexander Metz (personal communication) called "XTunnel", which visualizes the crystal packing in a protein crystal by generating crystal mates within 4 Å of the asymmetric crystal unit. This allows us to also visualize the channels available in the protein crystal and assess whether the binding pocket of the protein crystal is accessible by the fragments or not. This visualization of the entire crystal packing revealed blocked binding pockets in the crystal structure (Figure 3.7b). It was clear that the additional tryptophan residue at the N-terminus of the protein construct binds into and blocks one of the hydrophobic pockets of the neighboring crystal mate, while the second hydrophobic pocket of the crystal mate is also blocked by a threonine residue of another neighboring crystal (Figure 3.7c). This explained why the fragments were not able to bind into the binding pockets, and why the only bound fragment managed to squeeze itself into the available space found in the protein crystal. Accordingly, for fragments to bind to the other pockets, they would have to displace the residues of adjacent packing mates already binding into both hydrophobic pockets.

Table 3.2: Fragment hits for T. brucei PEX14 detected by NMR screening.

Fragment Number	Chemical Name	Chemical Structure
F1	1-benzothiophene-3-carboxylic acid	- OH
F2	1-benzothiophene-5-carboxylic acid	HO
F3	2'-Hydroxy-1'-acetonaphthone	O CH₃ OH
F4	1,8-Diaminonaphthalene	NH ₂ NH ₂
F5	1H-indole-7-carboxylic acid	HOO
F6	benzo[b]thiophene-7-carboxylic acid	HOO
F8	3-benzofuranacetic acid	ОН
F9a	alpha, alpha-dimethyl-6-methoxy-2- napthalenemethanol	OH
F10	4H-Thieno[3,2-b]pyrrole-5-carboxylic acid	S OH
F11	Ethyl 3-indoleacetate	O_CH ₃
F12	3-Indoleacetic acid	P ST

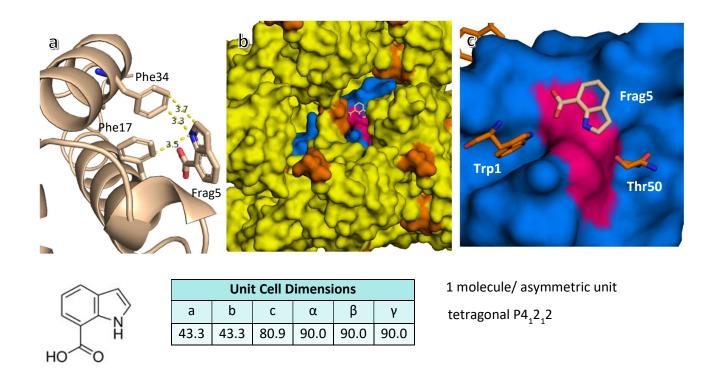


Figure 3.7: a) Fragment 5 (**crystal form 2**) was a hit in X-ray crystallography, but it is shown to stack against the phenylalanine residues separating the binding pockets without addressing either pocket. b) XTunnel revealed a dense packing of the protein crystal (asymmetric unit is shown in blue, crystal mates within 4 Å are shown in yellow, Phe34 and Phe17 are highlighted in pink, and the first four residues of the construct including the E1W mutation are highlighted in orange) and a blocked binding pocket where c) a tryptophan residue from a neighboring crystal mate blocks the pocket where the tryptophan of PEX5 binds, and another threonine residue blocks the second pocket where the phenylalanine residue of PEX5 binds. PDB: 6S6R. The fragment chemical structure, unit cell dimensions, space group, and number of molecules per asymmetric unit is also shown.

3.2.4 The Search for New Crystal Forms

Since fragment diffusion into free binding pockets is a prerequisite for fragment screening by soaking, we set up more crystallization screens using the MarXtal Screening Facility and evaluated over 1,500 conditions to find new crystal packing with more promising spatial arrangements of the protein molecules to study ligand binding. From the discovered crystallization hits (Figure 3.8), two conditions were reproducible. We collected datasets for these two crystals, D5 MBCII was collected using our inhouse MAR345 detector, and F12 Memgold was collected at BL14.1 at the BESSY II electron storage ring operated by the Helmholtz-Zentrum Berlin. MBCII was refined to a resolution of 2.6 Å while F12 Memgold was refined to a resolution of 2.3 Å. Despite crystallizing under different conditions and in two different space groups, both crystal forms also revealed a blocked binding pocket (

Table 3.3).

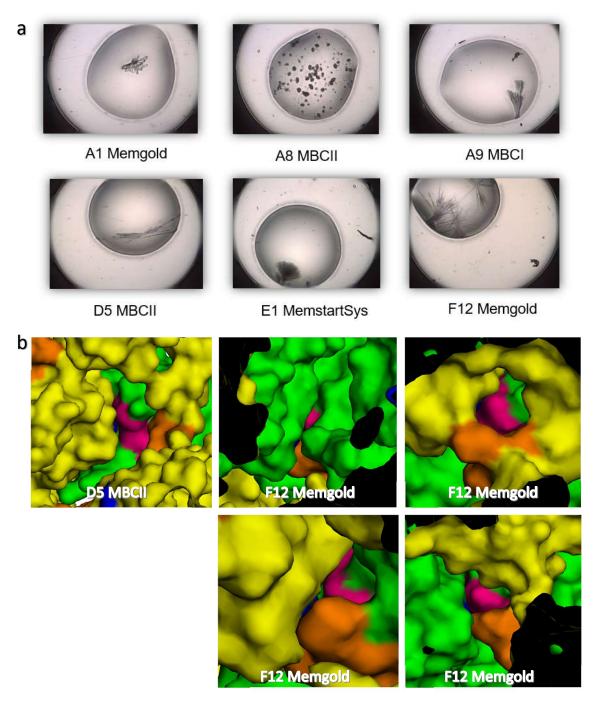


Figure 3.8: a) Additional hits from the crystallization screen of mutated PEX14. Only D5 MBCII and F12 Memgold were reproducible, however both revealed blocked binding pockets (phenylalanine residues separating the two pockets are highlighted in pink. b). The binding pockets were blocked by a neighboring crystal mate or another molecule in the asymmetric unit, where the Thr50 residue (highlighted in blue) or Trp1 residue (highlighted in orange) block the pocket.

Table 3.3: Crystallization conditions of the crystallization hits from mutated PEX14 crystallization screen. The images of the crystal hits are also shown.^a

Crystallization Conditions	Crystal Image	Unit Cell Dimensions and Space Group ^a		
D5 MBCII – Crystal form 3 Temperature: 18°C Concentration: 28mg/ml Salt: 0.1M Lithium Sulphate Buffer: 0.1M Trisodium Citrate pH 5.6 Precipitant: 30% PEG 400		$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$		
F12 Memgold – Crystal form 4 Temperature: 18°C Concentration: 28mg/ml Salt: 0.07M Sodium Chloride Buffer: 0.05M Trisodium Citrate pH 4.5 Precipitant: 22% PEG 400		$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$		

a) Both crystal forms crystallize in different space groups, however both of them have blocked binding pockets.

Co-crystallization of the fragments was also attempted by drying 0.5 μ L of a 1M stock solution of fragment (approx. 142mM) onto a crystallization well of a hanging drop plate and crystallizing the protein into the drop. We attempted all three crystallization conditions found to crystallize this mutated PEX14 construct, including published conditions¹²⁹ as seen in Table 3.4. However, no co-crystals grew under these conditions.

Table 3.4: Co-crystallization trials attempted with the NMR fragment hits from Table 3.2 did not produce any crystals.

Protein Concentration	Fragment Concentration	Buffer	Salt	Precipitant	Temp.
2mM	20mM	0.1M Tris HCl pH 8.5	0.2M MgCl ₂	30% PEG 4000	4°C
2mM	20mM	0.1M Tris HCl pH 8.5	0.22M Lithium Sulphate	29% PEG 4000	18°C
2mM	20mM	0.1M Bis Tris Propane pH 7.5	0.2M Sodium Sulphate	20% PEG 3350	18°C
2mM	20mM		0.2M Ammonium Sulphate	30% PEG 8000 5% Glycerol	4°C
1mM	100mM	0.1M Tris HCl pH 8.5	0.2M MgCl ₂	30% PEG 4000	4°C
1mM	100mM	0.1M Tris HCl pH 8.5	0.22M Lithium Sulphate	29% PEG 4000	18°C
1mM	100mM	0.1M Bis Tris Propane pH 7.5	0.2M Sodium Sulphate	20% PEG 3350	18°C
1mM	100mM		0.2M Ammonium Sulphate	30% PEG 8000 5% Glycerol	4°C

3.2.5 Fragment Hit-to-Lead Generation

With the initial fragment hit we had obtained from soaking (Figure 3.7a) we decided to initiate a lead generation by fragment growing, which was done by Khang Ngo (Philipps University of Marburg, personal communication). The synthetic pathway is shown below (Scheme 3.1). We further tested this compound in SOFAST HMQC NMR and in AlphaScreen to determine its K_D and IC₅₀ values, respectively (to be discussed in section 3.3). The co-crystallization of this lead compound with the mutated PEX14 in a crystallization screen (0.17 M ammonium acetate, 0.085 M sodium citrate pH 5.6, 25.5% PEG 4000, 15% glycerol) yielded a crystal structure with a resolution of 1.15 Å, but the synthesized ligand could not be resolved in the structure. However, we did manage to collect a crystal structure of this compound soaked in the wild-type PEX14 as will be discussed in the coming section 3.5.2.

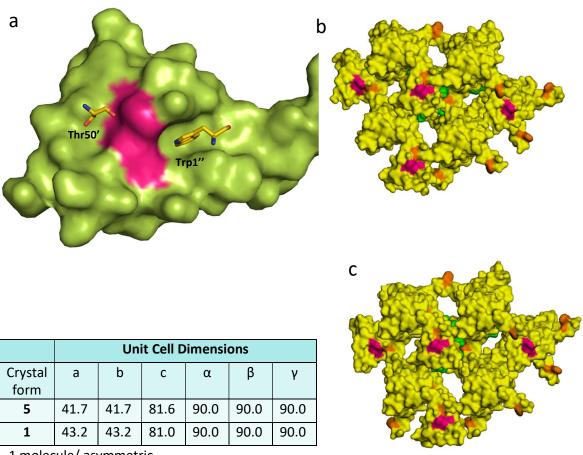
Scheme 3.1: Synthetic pathway of the lead compound 1 from the original fragment hit.

3.2.6 Soaking Co-Crystallized E1W PEX14

The E1W PEX14 construct proved difficult to crystallize in a crystal form with an accessible binding pocket likely due to the highly directional impact of filling of the hydrophobic pocket next to Phe17 and Phe34 with the accessible and spatially less restricted Trp1 residue of a second PEX14 molecule which was artificially introduced into parent protein for analytical reasons. Nevertheless, published structures of co-crystallized complexes with the E1W construct suggest that potent ligands can bind to the hydrophobic pockets before formation of the crystal contacts. Likely they have sufficient affinity to avoid displacement from the parent protein upon nucleation of the crystal packing. Some fragments with lower potency will likely be displaced by the contact of Trp1 in the binding pocket when the first protein molecules assemble to form a crystallization nucleus. With this in mind, we utilized computational methods to screen libraries for novel ligands as potential PEX14 inhibitors that would bind upon co-crystallization. Details of these methods and the virtual screening cascade implemented will be discussed in section 3.3.

Screening for co-crystallization conditions of E1W PEX14 with the discovered virtual screen (VS) hit ligands (described in section 3.3) led to a new crystallization condition (1 mM protein, 20 mM ligand, 0.17 M ammonium acetate, 0.085 M sodium citrate pH 5.6, 25.5% PEG 4000, 15% glycerol, 18°C) yielding a crystal in the P4₁2₁2 space group but slightly different unit cell dimensions (Figure 3.9, **crystal form 5**). Solving the structure of these crystals and generating the crystal packing of the new co-crystal form revealed a packing similar to the **crystal form 1** as seen in Figure 3.9b and Figure 3.9c and pockets that were still blocked by their neighboring crystal mates. It seems that the applied VS ligands somehow supported crystallization of the E1W PEX14 variant, as the crystallization conditions without adding the VS ligands did not yield any crystals. It may also be that the VS ligands are bound to the crystal but populated at only 10% or less, and thus cannot be seen in the structure. Superimposition of

the initial **crystal form 1** of E1W PEX14 variant and the new **crystal form 5** revealed a large shift in the positions of several residues along the protein structure including the first six residues of the sequence Gly-2, Ala-1, Met0, Trp1, His2, Thr3 that shift more than 0.6 Å. In total 25 residues in the protein structure are shifted by more than 0.4 Å, which already makes up 38% of the entire protein (Table 3.5). This perturbation in over a third of the protein could explain why the new **crystal form 5** obtained by these co-crystallization attempts could accommodate the fragments while the apo crystal form could not (Figure 3.10).



1 molecule/ asymmetric

tetragonal P4₁2₁2

Figure 3.9: **Crystal form 5** a) Central asymmetric unit cell is in green, and the phenylalanine residues separating the two hydrophobic pockets are highlighted in pink. One pocket is blocked by a Thr50 residue of the neighboring crystal mate and the other pocket is blocked by a Trp1 residue from another neighboring crystal mate. b) The generated crystal mates, 5 Å from the asymmetric unit shown in yellow, produce a crystal packing that is of similar density of **crystal form 1** (c) and should allow the fragments to diffuse into the into the binding pocket.

Table 3.5: The measured distances indicating the extent of shifting of the residues in the **crystal form 5** (crystallized in presence of VS ligands) of the E1W PEX14 variant relative to initial **crystal form 1** apo crystal.

Residue	Distance Å
Gly-2 Cα	0.85
Ala-1 Cα	0.89
Met0 Cα, Cε	0.32, 1.07
Trp1 Cα, Cγ	0.85, 1.47
His2 Cα, Cγ	0.68, 0.97
Thr3 Cα	0.66
His4 Cα	0.49
Ser5 Cα	0.43
Asp20 Cα	0.69
Ser21 Cα	1.29
Arg22 Cα	1.87
Val23 Cα	0.81
Arg24 Cα	0.81
Arg25 Cα	0.49
Thr26 Cα	0.77
Pro27 Cα	1.06
Thr28 N	0.94
Ser29 Cα	0.99
Ser30 Cα	0.44
Gly53 Cα	0.43
Gln54 Cα	0.42
Lys56 Cα	0.77
Thr57 Cα	0.83
Leu58 Cα	0.80
Asn59 Cα	0.44

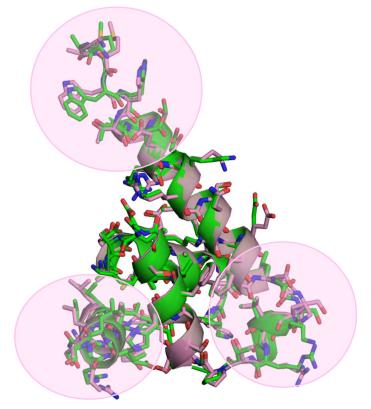


Figure 3.10: Superimposition of the entire structure of the **crystal form 5** of the E1W PEX14 variant "VS ligand influenced" (pink) and the apo crystal of the E1W PEX14 (green). The residues that encounter the most perturbation in their positions are encircled in pink. Together these residues make up over one-third of the protein structure, and these structural differences could explain why one crystal form could accommodate the fragment hits over the other.

Nevertheless, the novel **crystal form 5** crystals were used to rescreen the NMR fragment hits previously discussed (Table 3.2). Crystals were soaked in fragment solutions of 100 mM concentration for four hours. This resulted in a second E1W PEX14 fragment hit as seen below. The fragment F6 results in a crystal with eight independent molecules in the asymmetric unit for which the diffraction data suggested a setting of the crystal symmetry in the monoclinic space group $P2_1$. The crystals that generated this hit were crystallized upon addition of the ligand VS 7 (section 3.3). F6 seems to have altered the packing of the crystal, so that the symmetry of the diffraction data corresponded best to a different space group than the original crystals before soaking. This is likely due to the relatively small size of the protein, and the high concentration of the fragment solution used which overall perturbs the symmetry of the protein arrangement in a way that a symmetry reduction is experienced from a tetragonal $P4_12_12$ to a monoclinic $P2_1$. Superimposition of the crystal packing in $P4_12_12$ before soaking with the resulting crystal form $P2_1$ after soaking reveals perturbations of the residues at the N-terminus as seen in Figure 3.12. The measured distances indicating the extent of shifting of the residues in the "VS ligand influenced" **crystal form 5** of the E1W PEX14 variant relative to the crystal obtained by soaking F6 **crystal form 6** can be seen in Table 3.6.

Fragment Hit F6 crystal form 6



Unit Cell Dimensions						
а	а b с α β γ					
79.7	39.7	81.1	90.0	90.0	90.0	

8 molecules/ asymmetric unit monoclinic P2₁

Figure 3.11: Chemical structure of F6. This crystal structure after soaking gives 8 molecules per asymmetric unit corresponding to a different crystallographic setting in space group $P2_1$ compared to the symmetry of the un-soaked crystal which results in a crystal with a lower symmetry (**crystal form 6**). The unit cell dimensions are also displayed. The crystal structure was resolved to 1.76 Å.

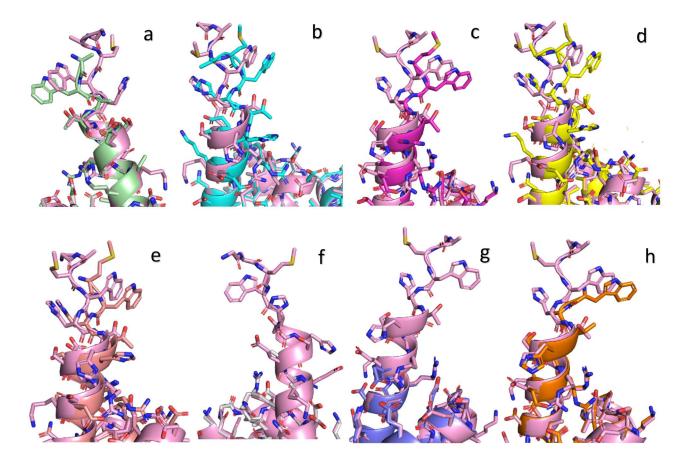


Figure 3.12: Superimposition of the "VS ligand influenced" **crystal form 5** of the E1W PEX14 variant obtained by co-crystallization (pink) with each of the eight molecules in the asymmetric unit (a - h) of the subsequently soaked crystal form obtained in complex with F6 (**crystal form 6**) (PDB: 6S7M). The residues that encounter the largest perturbations in their positions are the residues at the N-terminus namely the first 11 residues Ala-1, Met0, Trp1, His2, Thr3, His4, Ser5, Glu6, Arg7, Glu8, Lys9. In some of the eight copies, not all of these residues are resolved. These perturbations could explain the change in the crystal forms.

Table 3.6: The measured distances indicating the extent of shifting of the residues in the "VS ligand influenced" **crystal form 5** of the E1W PEX14 variant relative to the subsequently soaked crystal of the fragment hit F6.

F6 Hit crystal form 6			
Molecule A	Distance Å	Molecule D (cont.)	Distance Å
Met 0 Cα	2.8	His4 Cα, Cβ	1.5, 2.0
Trp1 Cα	1.7	Ser5 Cα, Cβ	1.1, 1.4
His2 Cα	1.0	Glu6 Cα, Cγ	0.9, 1.2
Thr3 Cα	0.6	Arg7 Cα, Cγ	0.8, 1.1
His4 Cα	0.4	Glu8 Cα, Cγ	0.7, 1.4
Ser5 Cα	0.4	Lys9 Cα, Cγ	0.5, 3.0
Glu6 Cα, Cγ	0.3, 0.6		
Arg7 Cα, Cβ	0.2, 0.3	Molecule E	Distance Å
Glu8 Cα, Cγ	0.3, 1.0	Met 0 Cα, Cγ, Sδ	2.8, 6.3, 7.8
Lys9 Cα, Cγ	0.3, 2.5	Trp1 Cα	2.3
		His2 Cα	1.3
Molecule B	Distance Å	Thr3 Cα	1.2
Ala -1 Cα	5.2	His4 Cα, Cβ	1.1, 1.4
Met 0 Cα, Cγ, Sδ	3.2, 8.1, 8.8	Ser5 Cα, Cβ	0.7, 0.9
Trp1 Cα	2.7	Glu6 Cα, Cγ	0.5, 0.8
His2 Cα	1.4	Arg7 Cα, Cγ	0.5, 0.9
Thr3 Cα	1.2	Glu8 Cα, Cγ	0.6, 1.3
His4 Cα, Cβ	1.3, 1.7	Lys9 Cα	0.4
Ser5 Cα, Cβ	1.0, 1.2		
Glu6 Cα, Cγ	0.7, 1.3	Molecule F	Distance Å
Arg7 Cα, Cγ	0.7, 0.9	Glu6 Cα	0.5
Glu8 Cα, Cγ	0.6, 1.3	Arg7 Cα	0.5
Lys9 Cα, Cγ	0.5, 2.8	Glu8 Cα	0.5
		Lys9 Cα	0.6
Molecule C	Distance Å		
Met0 Cα, Cγ, Sδ	2.5, 6.7, 8.1	Molecule G	Distance Å
Trp1 Cα	2.4	Arg7 Cα	0.1
His2 Cα	1.4	Glu8 Cα	0.6
Thr3 Cα	1.2	Lys9 Cα	0.2
His4 Cα, Cβ	1.2, 1.6		
Ser5 Cα, Cβ	0.8, 1.0	Molecule H	Distance Å
Glu6 Cα, Cγ	0.6, 0.4	Trp1 Cα	2.0
Arg7 Cα, Cγ	0.6, 0.9	His2 Cα	1.0
Glu8 Cα, Cγ	0.6, 1.2	Thr3 Cα	0.7
Lys9 Cα	0.5	His4 Cα, Cβ	0.4, 0.5
		Ser5 Cα, Cβ	0.3, 0.5
Molecule D	Distance Å	Glu6 Cα, Cγ	0.2, 0.2
Ala-1 Cα	5.4	Arg7 Cα, Cγ, Nε	0.3, 0.5, 2.5
Met0 C α , CE, S δ	3.3, 8.1, 8.7	Glu8 Cα, Cγ	0.4, 1.1
Trp1 Cα	2.7	Lys9 Cα	0.4
His2 Cα	1.5		
Thr3 Cα A, B	1.4, 1.4		

As the fragment F6 contains a thiophene moiety, it has anomalous scattering attributes due to the sulfur atom. This made it feasible to run an anomalous map on the collected dataset after molecular replacement which confirmed the presence of the fragment from the anomalous signal, as seen in Figure 3.13.

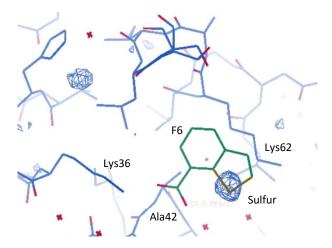


Figure 3.13: Superimposition of the crystal structure of E1W PEX14 in complex with F6 (green, PDB: 6S7M) and the generated anomalous map shows a clear anomalous signal in the same position where the sulfur atom of F6 is. The anomalous map is shown at a sigma level of 3σ .

F6 Interactions

All eight copies of the protein in the asymmetric unit have either one or both binding pockets free (Figure 3.14). Overall, six copies of the fragment can be found in the crystal structure sharing the same binding modes.

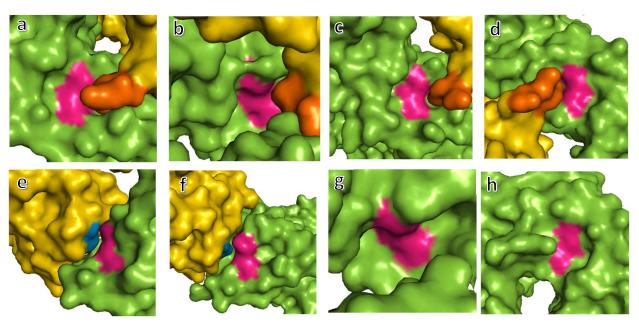


Figure 3.14: Eight molecules make up the asymmetric unit of the protein crystal. The asymmetric unit of each molecule is colored in green, whereas the generated neighboring crystal mates within 4 Å of the asymmetric unit are shown in yellow. The residues at the N terminus until the mutated tryptophan residue Trp1 are highlighted in orange, the Thr50 is highlighted in blue, and the two phenylalanine residues separating the binding pockets are highlighted in pink. Four protein molecules (a,b,c,d) have the binding pocket where the Phe107 of PEX5 binds free, while the other pocket where the Trp103 of PEX5 binds is blocked by the Trp1 of the neighboring crystal mate. Two molecules (e,f) have the Trp103 pocket free, and the Phe107 pocket is blocked by a Thr50 residue of the neighboring crystal mate. Two molecules (g,h) have both binding pockets unoccupied.

Chapter 3 PEX14

The electron density in the binding pockets of PEX14 seen in protein molecules A, B, D, F, and G suggests that the fragments could be binding in the pockets (Figure 3.15). Additionally, their planar structure with the aromatic system would fit into the hydrophobic pocket and gives a negatively polarized π system capable of forming an interaction with the positive Arg25 seen in some cases beneath the electron density of the bound fragment. However, when the fragments were built into the electron density, they had very high B factors and their exact orientation could not be clearly resolved even after consecutive rounds of structural refinements. Therefore, they were not added into the final structure.

F6 Molecule 1

The first molecule of F6 (Figure 3.16a) binds in protein molecule A at an occupancy of 98%, where it forms a hydrogen bond between its carboxylic acid moiety and the nitrogen atom of Lys36/A (2.9 Å), and the backbone nitrogen of Ala42/A (2.7 Å). F6 forms another hydrogen bond to Lys36'/C of the neighboring crystal mate (2.8 Å), and a π - π stacking with the F6 from the neighboring crystal mate (4.2 Å). It also stacks against Lys62/A. The bound fragments are shown with their 2m|Fo|-|Fc| density map in green which is contoured at a standard sigma level of 1 σ , hydrogen bonds are shown as yellow dotted lines, and distances are in Å.

F6 Molecule 2

The second molecule of F6 (Figure 3.16b) binds in protein molecule B at an occupancy of 100%, where it forms a hydrogen bond between its carboxylic acid moiety and the nitrogen atom of Lys36/B (2.8 Å), and the backbone nitrogen of Ala42/B (3.0 Å). F6 forms another hydrogen bond to Lys36'/D of the neighboring crystal mate (2.8 Å), and a π - π stacking with the F6 from the neighboring crystal mate (4.0 Å). It also stacks against Lys62/B. The bound fragments are shown with their 2m|Fo|-|Fc| density map in green which is contoured at a standard sigma level of 1 σ , hydrogen bonds are shown as yellow dotted lines, and distances are in Å.

F6 Molecule 3

The third molecule of F6 (Figure 3.16c) binds in protein molecule C at an occupancy of 100%, where it forms a hydrogen bond between its carboxylic acid moiety and the nitrogen atom of Lys36/C (2.6 Å), and the backbone nitrogen of Ala42/C (3.0 Å). F6 forms another hydrogen bond to Lys36'/A of the neighboring crystal mate (2.8 Å), and a π - π stacking with the F6 from the neighboring crystal mate (4.1 Å). It also stacks against Lys62/C. The bound fragments are shown with their 2m|Fo|-|Fc| density map in green which is contoured at a standard sigma level of 1 σ , hydrogen bonds are shown as yellow dotted lines, and distances are in Å.

F6 Molecule 4

The fourth molecule of F6 (Figure 3.16d) binds in protein molecule D at an occupancy of 100%, where it forms a hydrogen bond between its carboxylic acid moiety and the nitrogen atom of Lys36/D (2.6 Å), and the backbone nitrogen of Ala42/D (3.0 Å). F6 forms another hydrogen bond to Lys36'/B of the neighboring crystal mate (2.9 Å), and a π - π stacking with the F6 from the neighboring crystal mate (4.0 Å). It also stacks against Lys62/D. The bound fragments are shown with their 2m|Fo|-|Fc| density map in green which is contoured at a standard sigma level of 1 σ , hydrogen bonds are shown as yellow dotted lines, and distances are in Å.

F6 Molecule 5 and 6

The fifth molecule of F6 (Figure 3.16e) binds in protein molecule E at an occupancy of 100% and the sixth molecule of F6 binds in protein molecule H at an occupancy of 100%. Both fragment molecules form a π - π stacking to each other (3.9 Å). F6 molecule 5 forms a hydrogen bond between its carboxylic acid moiety and the nitrogen atom of Lys36/E (2.5 Å), and the backbone nitrogen of Ala42/E (3.0 Å). It also forms another hydrogen bond to Lys36/H (2.8 Å). F6 molecule 6 forms a hydrogen bond between its carboxylic acid moiety and the nitrogen atom of Lys36/H (2.7 Å), and the backbone nitrogen of Ala42/H (3.6 Å). It also forms another hydrogen bond to Lys36/E (2.6 Å). The bound fragments are shown with their 2m|Fo|-|Fc| density map in green which is contoured at a standard sigma level of 1σ , hydrogen bonds are shown as yellow dotted lines, and distances are in Å.

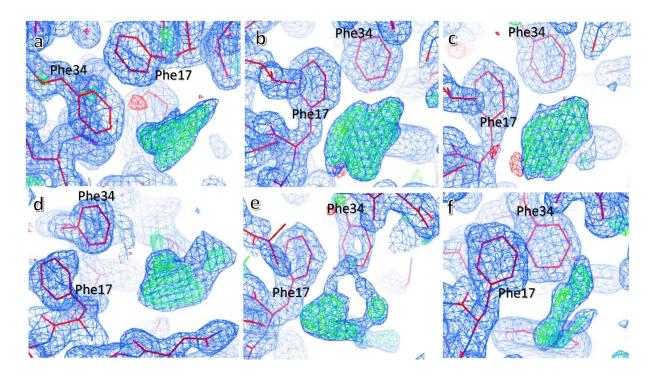


Figure 3.15: The eight independent protein molecules A (a), B (b), D (c), F (d), G (e), and H (f) have electron densities near the Phe17 and Phe34 residues that separate the hydrophobic binding pocket, suggesting that F6 could also be binding. The fragments were left out of the final structure due to high B factors after refinement and the electron density being unclear for concluding the fragment orientation. The 2m|Fo|-|Fc| density map is shown in blue contoured at a standard sigma level of 1σ and the m|Fo|-|Fc| density map is shown in green contoured at a standard sigma level of 3σ .

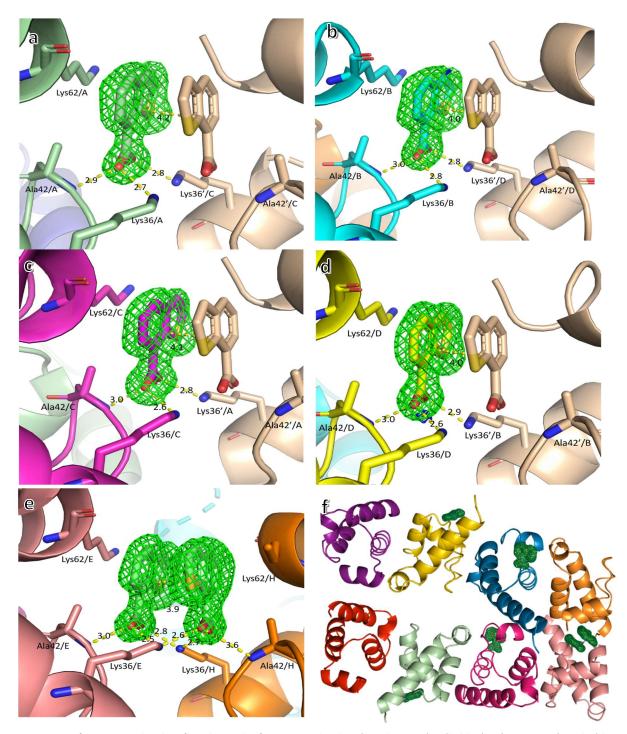


Figure 3.16: One fragment molecule is found in each of protein molecule A (a, pale green), B (b, blue), C (c, magenta), and D(d, yellow). Two fragment molecules are found stacking against one another in protein molecules E and H (e, E in salmon, H in orange). The interactions found for all fragment molecules are very similar where they form hydrogen bonds between their carboxylic acid moieties and the nitrogen of Lys36, and backbone nitrogen of Ala42. Additionally, the fragments stack against Lys62 and form a second hydrogen bond to Lys36 of the symmetry equivalent crystal mate, or in the case of molecule E and H to the neighboring protein molecule. With the exception of the two fragment copies that stack against each other in protein molecules E and H (e), all other four fragment copies stack against the fragment of the symmetry equivalent mate (images a – d, colored in wheat). An overview of the fragment binding positions is shown in (f). The bound fragments are shown with their 2m|Fo|-|Fc| density map in green which is contoured at a standard sigma level of 1σ , hydrogen bonds are shown as yellow dotted lines, and distances are in Å. Residues Leu65 and Ser66 are omitted for image clarity.

All crystal forms discussed above can be transformed into the other by applying a specific matrix, which includes a shift of the crystal origin. The matrices for transformation are listed in Table 3.7, and an image depicting the origins of the crystal forms is also shown.

Table 3.7: A list of the transformation matrices that can be applied to transform one crystal form of PEX14 to the other.

Transformation ^a	Original Matrix	Matrix Applied to
		Transform
crystal form 5 to crystal form 1	(1.0, 0.0, 0.0, 0.0	(-1.0, 0.0, 0.0, 20.9
	0.0, 1.0, 0.0, 0.0	-0.0, -1.0, 0.0, 18.3
	0.0, 0.0, 1.0, 0.0	0.0, 0.0, 1.0, -1.1
crystal form 5 crystal form 1	0.0, 0.0, 0.0, 1.0)	0.0, 0.0, 0.0, 1.0)
crystal form 6 to crystal form 5	(1.0, 0.0, 0.0, 0.0	(0.1, -1, -0.0, 3.1
	0.0, 1.0, 0.0, 0.0	-1.0, -0.1, -0.0, 92.5
	0.0, 0.0, 1.0, 0.0	0.0, 0.0, -1.0, 191.7
	0.0, 0.0, 0.0, 1.0)	0.0, 0.0, 0.0, 1.0)
crystal form 6 crystal form 5		

a) The transformations can be achieved by applying the listed matrix to the second crystal form listed.

3.2.7 Discussion

The mutated E1W residue at the N-terminal proved to be an obstacle in crystallization, despite selecting a residue far from the binding pocket. It appears that the small size of the construct yields crystals with crystal mates in proximity, and because of the hydrophobicity of tryptophan, it binds into the binding pockets of its neighboring crystal mates. These crystals are easily influenced by the fragment used to soak them, as seen by the fragment hit F6 where the fragment has lowered the symmetry of the crystals and suggested a setting within an alternative space group (Figure 3.11).

As fragment soaking with the E1W PEX14 variant proved to be difficult, the more probable construct to work with for fragment soaking appeared to be the wild-type construct that bears a glutamate instead of the mutated tryptophan. This will be discussed in detail in section 3.4.

3.3 Virtual Screening

3.3.1 Definition

In drug discovery, virtual screening (VS) is a computational approach used to search databases of compounds to find those that are most likely to bind to a drug target with a known structure either from X-ray crystallography or NMR, or high confidence homology models. ^{134,135} The databases to be searched may contain up to millions of available compounds, and can be obtained from commercial libraries, public libraries, or commercial vendors. In VS, there are two broad categories; ligand-based (LBVS) and structure-based (SBVS) approaches. In LBVS, a set of known active compounds are used as a reference to define and subsequently identify candidate compounds. This approach includes similarity searching, quantitative structure-activity relationships (QSAR), pharmacophores, as well as 3D shape matching. ¹³⁶ In SBVS, the 3D structure of the biological target is used as the reference to dock the candidate molecules and give them a rank based on how well they fit into the binding site. ¹³⁷

3.3.2 Screening Methodology

In the VS performed here, we used a combination of LBVS and SBVS utilizing several tools which include *Scorpion*¹³⁸ (DesertSci), *Anchor Query*¹³⁹, *PocketPicker*¹⁴⁰, *LIQUID*, *KNIME* (The Konstanz Information Miner)¹⁴¹, as well as *CATS light* 2^{142,143} (Chemically Advanced Template Search) for molecular similarity search and *DOGS*¹⁴⁴ (Design of Genuine Structures).

The first part of the performed screen included the following steps:

- I. Scorpion.⁵ This online server was used to rank the protein-ligand interactions of a published PEX14 ligand, compound **2** (PDB:5L87, Figure 3.17a) and the protein-fragment interactions of the fragment hit (PDB:6S6R) to assess which moieties, in terms of anchors, are important for binding.
- II. Anchor Query. 139 These anchors were then used individually to screen a database of MCR reactions that synthesize compounds possessing moieties with similar properties (pharmacophore search). The outcome of this screen with the fragment anchor was 250 compounds, while the ligand anchor gave 500 compounds.
- III. PocketPicker¹⁴⁰ & LIQUID (structure-based drug design). Pocket Picker is a grid-based detection that detects pockets on the surface of the protein. The PEX14 receptor was used here after removing all solvent molecules and ligands. "VirtualLigand", an accompanying program, was then used to project pharmacophoric properties of the detected pocket residues into this receptor pocket, creating a "LIQUID" model of the protein pocket or a pseudo pharmacophore. This "LIQUID" is a set of descriptors that was then used to screen the ACC library, which is a library compiled from several available databases being Asinex Elite, Asinex Fragments, Asinex Gold + Platinum Collections, ChemBridge SC Collection, Enamine Advanced, Enamine HTS Collection, Specs Natural Products, and Specs Screening Compounds. This resulted in 2,000 molecules matching these descriptors. To ensure that all our compound data sets were defined using the same descriptors, the compounds from Anchor Query were also converted into LIQUID descriptors.
- IV. KNIME. 141 KNIME was used to compare the distances (RMSD) between compounds from the screenings, and the original LIQUID model to assess which compounds were most similar to the pseudo pharmacophore.

The second part of the screen:

- V. CATS^{142,143} (Ligand based drug design). CATS takes an input ligand and describes it using a set of descriptors called CATS. The published ligand, compound **3** (PDB: 5N8V, Figure 3.17b) with a K_i of 0.207 μ M¹²⁹ was used. These CATS descriptors were also used to define the compounds from the first part of the screen.
- VI. *KNIME*.¹⁴¹ KNIME was used to compare the distances (RMSD) between the compounds and the published ligand compound **3** (PDB: 5N8V) ligand, to assess which of these compounds are most similar to the ligand. This was done as a cross validation of the Anchor Query results and PocketPicker.
- VII. DOGS. 144 DOGS uses an input ligand (in our case we used the same ligand, compound 3) to suggest de novo design methods for synthesizing similar compounds. This screen gave 432 compounds.

The third part of the screen:

VIII. Docking in MOE. 145 All compounds were docked into the PEX14 receptor pocket after removing duplicates (10 poses each, rigid docking). The compounds were then filtered by visual inspection of the best binding poses.

52 compounds were selected, and after filtering based on Lipinski RO5¹⁴⁶, **28** compounds were selected for purchase based on the number of carbon-carbon double bonds, compound flexibility, and conformation (Table 3.8).

Additionally, we also ran *Haddock2.2*¹⁴⁷ from Utrecht University to study whether docking the E1W PEX14 variant monomer against itself would predict the energetics of the mutated tryptophan upon binding into the pocket of the other monomer. The result showed that there are some energy penalties to some of the other monomer residues, but the tryptophan residue docked into the binding pocket of the other monomer, which is what we observe in our crystal structures.

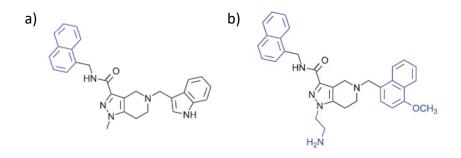


Figure 3.17: a) Chemical structures of compound 2 (PDB:5L87) and b) compound 3 (PDB: 5N8V).

Table 3.8: Chemical structures of the selected compounds for PEX14 binders found through a combination of SBVS and LBVS.

Compound No.	Chemical Structure	Compound No.	Chemical Structure
VS 1		VS 15	N H HN
VS 2		VS 16	o N H N
VS 3		VS 17	
VS 4		VS 18	C PART OF THE PART
VS 5		VS 19	C NH NH
VS 6	O N N N N N N N N N N N N N N N N N N N	VS 20	

VS 7	N H HR	VS 21	G N N N N N N N N N N N N N N N N N N N
VS 8		VS 22	
VS 9	N. H. IIN.	VS 23	
VS 10	C NH HN	VS 24	O NH
VS 11	O IN	VS 25	NH NH
VS 12	N N N N N N N N N N N N N N N N N N N	VS 26	OH OH

VS 13	VS 27	NH NH
VS 14	VS 28	HN HN HN H

3.3.3 Determining apparent K_D and IC₅₀ using 2D NMR and AlphaScreen

3.3.3.1 SOFAST HMQC 2D NMR Screening

The 28 purchased compounds suggested by the virtual screening campaign, in addition to lead compound $\bf 1$ (Scheme 3.1), were screened in a SOFAST HMQC 2D NMR measurement to record their binding. The compounds leading to significant chemical shift perturbations were then measured with a SOFAST HMQC titration (Figure 3.18) to determine their apparent K_D (Table 3.9). SOFAST HMQC (selective optimized flip angle short transient heteronuclear multiple quantum correlation)¹⁴⁸ allows recording of 2D 1 H- 1 SN correlation spectra of our protein. The protein used was the E1W PEX14 variant expressed and purified to have 1 SN labelling (further details can be found in the Materials and Methods of this chapter). The ligands were added to the protein and spectra were recorded.

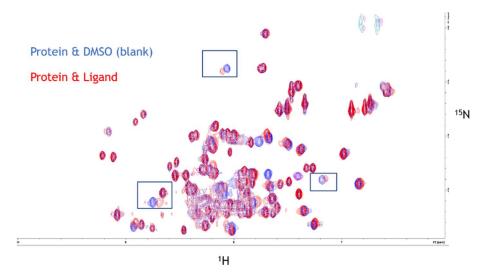


Figure 3.18: SOFAST HMQC spectra of labelled PEX14 with the VS compounds. The ¹H-¹⁵N correlation spectra that show pronounced shifts between protein alone (blue) and protein with the ligand (red) indicate ligand binding to the protein, as highlighted in square boxes.

AlphaScreen-based Competition Assay

The AlphaScreen[™] (Amplified Luminescent Proximity Homogenous Assay Screen) is a method to measure biological interactions and estimate affinities. It relies on the use of "Donor" and "Acceptor" beads that are bioconjugated with the protein and another molecule. In our experiment, the Histagged PEX14 was associated with a Nickel chelate Acceptor bead, while a biotinylated high affinity PEX5-derived peptide ALS (ALSENWAQEFLA)¹²⁹ was associated with the Streptavidin-coated Donor bead. If an interaction occurs between the protein and peptide, it will bring the two beads close together which starts a series of chemical reactions and produces a signal. Upon laser excitation at 680 nm, the oxygen from ambient air is converted to a singlet state by a photosensitizer in the Donor bead. This excited singlet oxygen reacts with a chemiluminescer in the Acceptor bead which activates fluorophores that emit light at 520–620 nm (Figure 3.19a).¹⁴⁹ If the compounds being tested bind to PEX14-associated Acceptor beads, the high-affinity peptide on the Donor beads will no longer be close to the Acceptor beads and the recorded signal will drop. The level of the recorded signal is then plotted against the log arithmic concentration of the ligand to calculate the IC₅₀ value (Figure 3.19b).

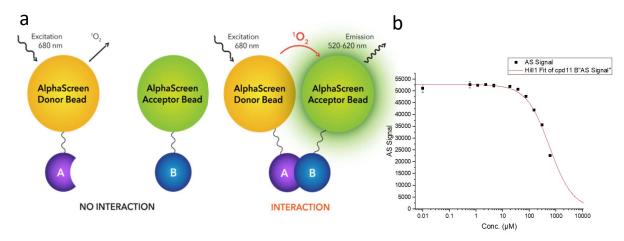


Figure 3.19: a) Scheme showing how the AlphaScreen functions (source of image from procomcure.com). A singlet oxygen released from the Donor bead diffuses across to the Acceptor bead where it is responsible for setting off a cascade of reactions that result in a light signal emitted, indicating an interaction due to close proximity of the two beads. b) Plot of the recorded signal versus the log concentration of the ligand used, where the IC₅₀ is calculated.

We measured the NMR hits from the VS campaign in this orthogonal AlphaScreen-based assay described in the literature¹²⁹ to obtain IC_{50} values for the discovered compounds. IC_{50} values are given in Table 3.2. VS 25 gave the highest IC_{50} value in our AlphaScreen and so was also measured in an *in vitro* trypanocoidal cell-based assay to give an IC_{50} of 18.5 μ M, a promising value for a VS compound.

SOFAST HMQC NMR Titrations

The strength of a protein-ligand complex can be defined by a single quantity that sums its stability or its reciprocal – the dissociation constant (K_D) : $K_D = [P][L]/[PL]$. This dissociation rate depends on the concentration of the bound ligand as well as the strength of the binding interaction.¹⁵⁰ Protein-ligand complexes (except high-affinity ligands) coexist in solution in equilibrium with significant amounts of free protein and/or free ligand¹⁵¹. In titration NMR, the concentration of the unbound protein will gradually decrease upon increasing concentrations of the ligand, while the concentration of the bound

protein increases. This is reflected in the gradual movement of the NMR peak, which is the average of the bound and unbound states, towards the fully bound state. This is the concept we utilized in our experiments, assuming our ligands are in fast-exchange and form a 1:1 complex.

Five out of the 28 VS ligands produced pronounced chemical shifts (in addition to lead compound **1** (Scheme 3.1). Their SOFAST HMQC spectra were recorded at eight different concentrations and a fixed concentration of the protein. The chemical shifts of both, the ¹H and ¹⁵N of our protein residues (relative to chemical shifts of blank measurements) at each ligand concentration were noted. An example of such a chemical shift in a 2D correlation plot is seen in Figure 3.20.

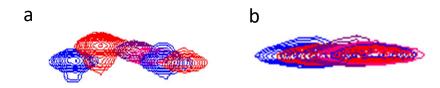


Figure 3.20: a) A strong chemical shift can be observed in the 2D correlation plot of lead compound **1** at consequently increasing concentrations from 0 mM until 1.5 mM (only 0 mM, 0.125mM, 0.375 mM, 0.75mM, 1mM, 1.5mM are displayed for clarity), where the y-axis representing the ¹⁵N and the x-axis representing the ¹H chemical shifs are not displayed. b) A chemical shift can also be observed in the 2D correlation plot of VS 7 at the same concentrations, but less prominent relative to lead compound **1**.

The weighted chemical shifts ($\Delta\delta$) of each residue were then calculated by the following equation, accounting for the different weights of ¹⁵N and ¹H chemical shifts (Equation 3.1). ¹⁵²

$$\Delta \delta = \sqrt{(\Delta H^2 + (\Delta N/10)^2)}$$
 (Equation 3.1)

 $\Delta\delta/[L]_0$ is then calculated by dividing the $\Delta\delta$ at the ligand concentration.

Using Scatchard Plots, the $\Delta\delta/[L]_0$ is then plotted against the $\Delta\delta$ as seen in Figure 3.21 to give a linear plot where the gradient = -1/ K_D and x-intercept = $\Delta\delta$ max, based on (Equation 3.2). 150

$$\Delta\delta/[L]_0 = -\Delta\delta/K_D + \Delta\delta \max/K_D$$
 (Equation 3.2)

The apparent K_D values calculated from the Scatchard plots after removing anomalies, are given in Table 3.9.

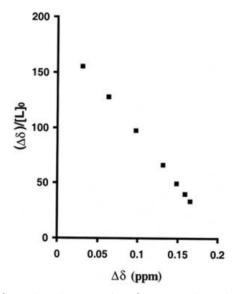
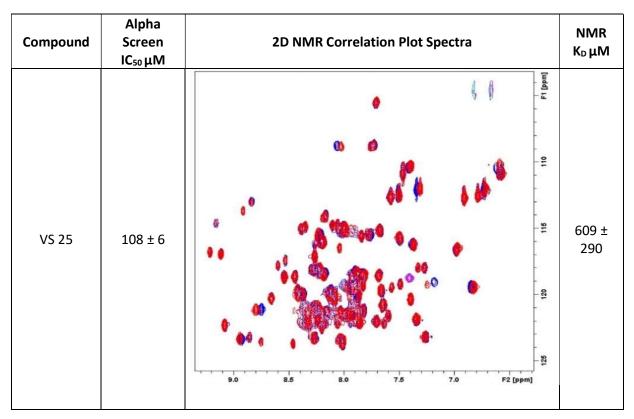
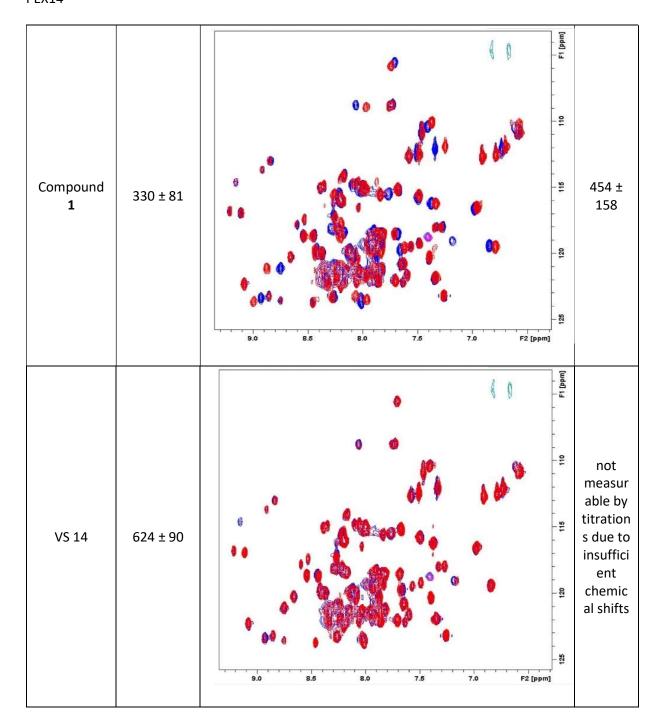


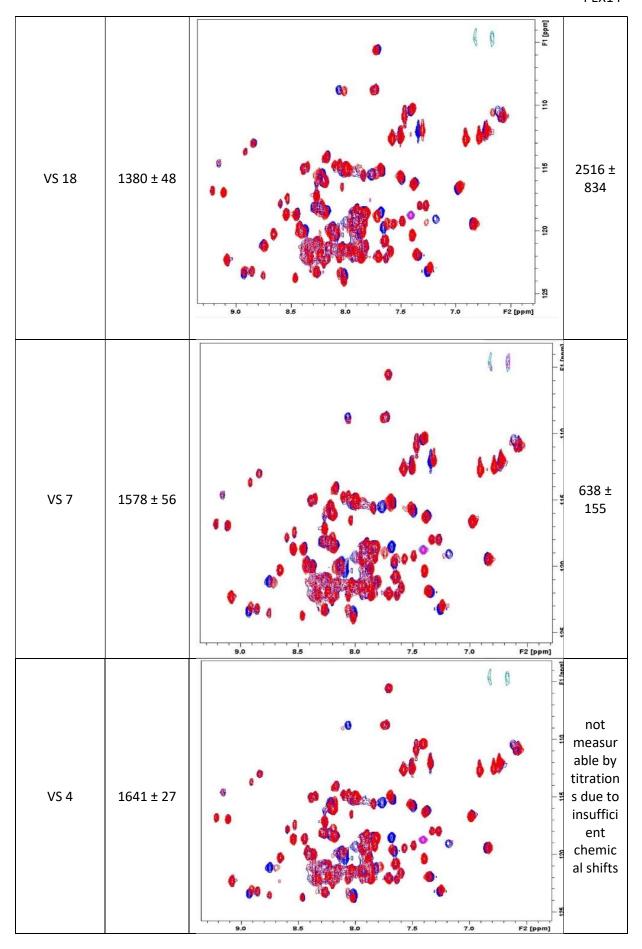
Figure 3.21: Scatchard plot where $\Delta\delta/[L]0$ is plotted against the $\Delta\delta$ to give a linear line where the slope equals -K; the association constant of ligand binding. The reciprocal value gives the dissociation constant K_D^{16} .

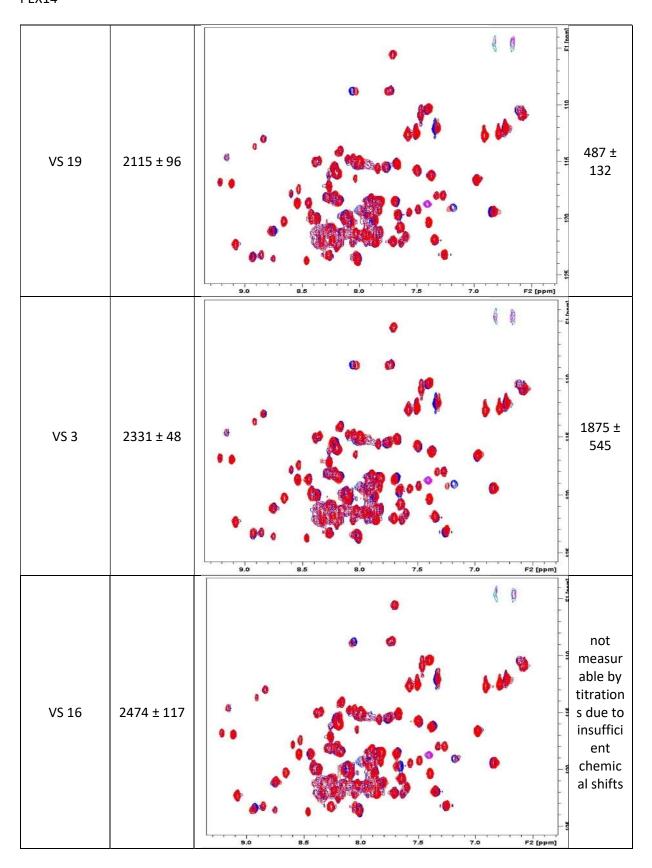
Table 3.9: IC_{50} values obtained from the AlphaScreen and apparent K_D values obtained from 2D NMR titrations.^a

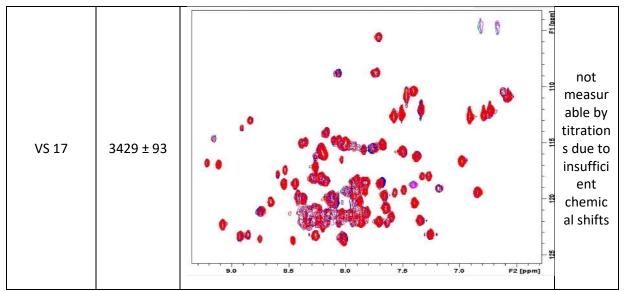


Chapter 3 PEX14









a) For our VS compounds are plotted below, where VS 25 gave the highest IC_{50} value and was measured in an in-vitro trypanocoidal assay to give an IC_{50} of 18.5 μ M. The SOFAST HMQC correlation plot spectra for each compound are also shown.

3.3.4 Co-crystallization of Ligands

The most potent VS compounds from the NMR titrations VS 19 and compound **1** were screened against almost 1500 conditions to find possible co-crystallization conditions. The co-crystallization screening of both compounds led to the common condition listed in Table 3.10. This condition was used to co-crystallize also the other VS compounds from Table 3.2. High resolution data sets (1.25 - 1.90 Å) were collected for all the compounds, with the exception of VS 14, VS 16, and VS 17, as upon addition of these ligands to the protein during co-crystallization, the protein did not crystallize. However, even the X-ray structures collected for the VS compounds that did crystallize, did not show any difference electron density indicating the bound compounds.

Table 3.10: Co-crystallization screen of the compound 1 and VS 19.^a

Compound	Temp.	p. Salt Buffer Pi		Precipitant	Protein: Ligand	Co-crystal Image	
Compound 1	18°C	0.17 M Ammonium Acetate	0.085 M Sodium Citrate pH 5.6	25.5% PEG 4000, 15% Glycerol	1 mM: 20 mM	1.78Å	
VS 19	18°C	0.17 M Ammonium Acetate	0.085 M Sodium Citrate pH 5.6	25.5% PEG 4000, 15% Glycerol	1 mM: 20 mM	1.72Å	

a)The common crystallization condition was used to crystallize the other compounds of the VS that were found to bind in 2D NMR.

It seems that the ligands are not potent enough to support co-crystallization and to assemble in formed crystals. In reported cases where ligands were co-crystallized with PEX14, the least potent compound which was a result of an optimized VS hit, had a K_i of 12.3 μ M¹²⁹, supposedly having a higher potency than the compounds studied here, where the most potent compound had a K_D above 450 μ M.

Nevertheless, the VS ligands somehow aid crystallization of the E1W PEX14 variant, as crystallization trials using the same crystallization conditions in the absence of any of the VS ligands does not yield any crystal formation. The obtained "VS ligand influenced" **crystal form 5** has slightly different unit cell dimensions than the **crystal form 1** we previously used for soaking, and thus were used for further fragment screening, as discussed in section 3.2.4.

3.4 Crystallizing the Wild-Type

3.4.1 Crystallization Screening

A crystallization screen was set up for the wild-type construct, which resulted in only one reproducible hit out of almost 1500 conditions. The reproduced crystals differed in their morphology to the screening hit, where the latter gave a 'shower' of crystals meaning many crystals grew in the well resulting in very thin needles (Figure 3.22). As suggested in literature and applied earlier, 5% glycerol was added to the reservoir solution to reduce nucleation and crystal growth, which resulted in larger crystals thick enough for fishing and testing at a microfocus beamline.

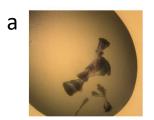




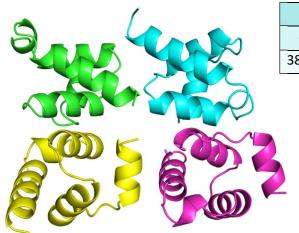


Figure 3.22: Images of the a) original crystallization hit of the wild-type PEX14 and the b) reproduced crystals which grow in a shower of very thin crystals. c) Addition of 5% glycerol helped slow down crystal growth and resulted in crystals thicker in size.

Table 3.11: Crystallization conditions of the wild-type PEX14, requiring very high concentrations of the protein.

Protein	Concentration	Buffer	Salt	Precipitant	Temperature
Wild-Type	95mg/ml	0.1M Trisodium	none	40% PEG 600	4°C
PEX14		Citrate pH 5.5		(+5% Glycerol)	

The wild-type PEX14 crystal was soaked in a soaking solution consisting of the crystallization solution, 100 mM of compound 1, and 25% glycerol for 19 hours. The diffraction data of the crystal suggested a setting of the crystal symmetry in the monoclinic $P2_1$ space group with four molecules in the asymmetric unit (PDB: 6S9Y). The data was refined to a resolution of 1.64 Å, but no electron density could be seen for compound 1 (crystal form 7, Figure 3.23).



Unit Cell Dimensions									
а	b	С	α	β	γ				
38.9	81.3	39.3	90.0	92.8	90.0				

4 molecules/ asymmetric unit monoclinic P2₁

Figure 3.23: Crystal of the wild-type PEX14 soaked in compound $\bf 1$ yields a structure in P2₁ space group with four independent molecules in the asymmetric unit (A, green, B, blue, C, magenta, D, yellow), where compound $\bf 1$ does not appear to bind. No electron density could be seen for compound $\bf 1$ even at a resolution of 1.64 Å. (PDB: 6S9Y)

Another dataset for the wild-type PEX14 was collected from a crystal that was not exposed to a ligand to be soaked. The diffraction data suggested a setting of the crystal symmetry in the orthorhombic space group C222₁ with two molecules in the asymmetric unit. After structural refinement of the dataset, we noticed that both the R_{work} (27.9%) and the R_{free} (32.1%) values were in the high range for the collected resolution. Additionally, the average B-factor of the protein molecule A was 23.1 Ų, whereas the average B-factor of the protein molecule B was 33.3 Ų, which is almost twice. This made us question whether the data was processed in the correct space group. To validate the chosen space group, the dataset was processed in the lowest symmetry space group P1 and the XDS_ASCII.HKL file from processing was run in XPREP¹⁵³ to search for higher symmetry space groups. The practical processing statistics of all space groups, the dataset was processed in, are shown in Table 3.12. XPREP is a program used to analyze the symmetry of the data and to merge the reflections according to the space group found. The output from XPREP is shown in Figure 3.24. The monoclinic P lattice with the calculated lowest R-sym value was selected, and the dataset was processed in space group P2₁.

Table 3.12: The data processing statistics after processing in three different space groups.^a

Space group	C222 ₁	P1	P2 ₁
R-sym (%)	6.5	4.1	4.8
R-sym (%) highest resolution shell	34.1	31.4	36.8
R-sym (%) lowest resolution shell	2.8	1.8	2.1
Completeness	96.7 (93.6)	82.5 (73.0)	93.6 (85.9)
Multiplicity	6.1	1.9	3.3
Resolution	1.93 Å	1.90 Å	1.90 Å

a) Data in parenthesis indicates the highest resolution shell.

```
Determination of reduced (Niggli) cell
Transformation from original cell (HKLF-matrix):
  0.0000 1.0000 0.0000
                          1.0000 0.0000 0.0000
                                                  0.0000 0.0000 -1.0000
Unitcell:
             39.116 39.200 80.828 90.01 90.00
                                                     92.81
Niggli form:
                     1530.06
                                 b.b = 1536.64
              a.a =
                                                     C.C =
                                                            6533.17
              b.c =
                       -0.39
                                 a.c =
                                          -0.11
                                                     a.b =
Search for higher METRIC symmetry
Identical indices and Friedel opposites combined before calculating R(sym)
Option A: FOM = 0.123 deg. ORTHORHOMBIC C-lattice R(sym) = 0.086 [ 21723]
       54.003 56.719 80.828 90.00
                                     90.01
                                            89.88
                                                     Volume:
                                                               247576.75
Matrix: 1.0000 1.0000 0.0000 1.0000 -1.0000 0.0000 0.0000 0.0000 -1.0000
                         MONOCLINIC P-lattice R(sym) = 0.039 [ 13488]
Option B: FOM = 0.007 deg.
      39.116 80.828 39.200 89.99
                                    92.81
                                            90.00
                                                              123788.38
                                                     Volume:
Matrix: 0.0000 1.0000 0.0000 0.0000 1.0000 1.0000 0.0000 0.0000
                  ..........
Option C: FOM = 0.123 \text{ deg.}
                        MONOCLINIC C-lattice R(sym) = 0.081 [ 13653]
       54.003 56.719 80.828 90.00
                                     90.01
                                            89.88
                                                     Volume:
Matrix: 1.0000 1.0000 0.0000 1.0000 -1.0000 0.0000 0.0000 0.0000 -1.0000
Option D: FOM = 0.123 deg. MONOCLINIC C-lattice R(sym) = 0.083 [
       56.719 54.003 80.828 89.99 90.00 90.12
                                                     Volume:
Matrix: 1.0000 -1.0000 0.0000 -1.0000 -1.0000 0.0000 0.0000 0.0000 -1.0000
```

Figure 3.24: Screenshot from the output of XPREP, showing the space groups of higher symmetry suggested for the dataset based on the calculated R-sym value. The monoclinic p-lattice space group has the lowest R-sym value at 3.9%.

The processing in P2₁ gave the most promising statistics, and Matthew's coefficient analysis suggested that four molecules in the asymmetric unit and 40.2% solvent content in the crystal, as well as three molecules in the asymmetric unit and 55.17% solvent content were the most likely states. Molecular replacement (MR) was performed using the MTZ file from the P2₁ processing with three molecules in the asymmetric unit. The solution from MR gave three molecules; A, B, and C. For molecules A and B, the average B factors were comparable (23.3 Ų and 23.9 Ų, respectively) but for molecule C it had slightly higher values (32.7 Ų). Additionally, the built molecules had several overlapping residues which suggested an incorrect MR solution. To check the correctness of this MR solution, the PDB file from MR was run in the program ARP/wARP ¹⁵⁴ which is used to build macromolecular models in the electron density map. Only 56% of molecule A could be built into the electron density (residues Val23 until Glu60), and 20% of molecule B (residues Leu40 until Gly53). None of the residues of molecule C could be built.

Despite better R-sym values of the dataset after processing in the P2₁ solution, the MR solution was not correct based on the results from ARP/wARP. Therefore, we proceeded with the original data processed in orthorhombic space group. This higher symmetry space group cannot be excluded for PEX14, as C222₁ is in fact a supergroup of P2₁. The MR solution of two molecules in the asymmetric 108

unit was tested in ARP/wARP ¹⁵⁴, which built the entire protein molecule A in the electron density but was only able to build half of protein molecule B. Due to the poor dataset, the resulting electron density shows disorder for this second protein molecule. The obtained crystal structure (**crystal form 8**) can be seen in Figure 3.25, where the protein molecule B has also been built in for visualization purposes of the entire asymmetric unit.

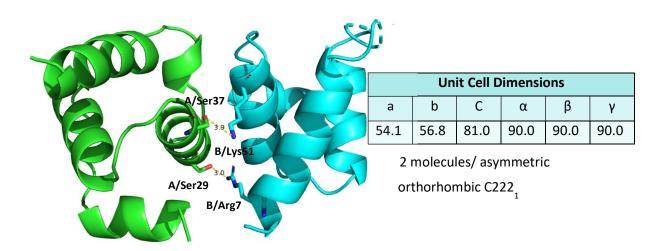


Figure 3.25: Crystal structure of the wild-type PEX14 at a resolution of 1.93 Å revealed two molecules in the asymmetric unit corresponding to the orthorhombic space group C222₁. Protein molecule A was fully resolved in the electron density whereas protein molecule B was only partially resolved. The first resolved residue is Ser5 which indicates that even Glu1 from the wild-type sequence is not visible in the crystal packing. The unit cell dimensions of the crystal structure are indicated.

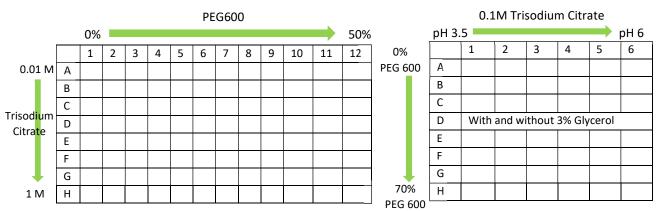
This crystal form was soaked again in a higher concentration of the lead compound **1** at 143 mM, for a longer time of 23 hours. This yielded a crystal structure of the wild-type PEX14 in complex with compound **1**, which will be described in section 3.5.1.

3.4.2 Wild-Type Crystals Optimization

3.4.2.1 Screening Around Buffer Components

To improve the stability of the wild-type crystals when soaked in fragment solutions, systematic attempts to optimize the crystallization conditions were performed included screening around buffer pH, buffer concentration, precipitant concentration as well as addition of glycerol, as indicated in Table 3.13.

Table 3.13: Screening around the crystallization conditions of wild-type PEX14. a



a) This included screening around the buffer concentrations, buffer pH, precipitant concentration as well as addition of glycerol, to control crystal growth rate.²

Another optimization screen included adding varying concentrations of salt to the crystallization conditions as seen in Table 3.14 .

Table 3.14: Varying concentrations of salt in addition to varying concentrations of the components in our crystallization conditions.

%PEG600		0.1M Trisodium Citrate											
		1	2	3	4	5	6	7	8	9	10	11	12
		pH 4.5		pH 5.5		pH 4.5		pH 5.5					
Α	45% PEG 600												
В	B 50% PEG 600		25mM NoCl		25mM NaCl		75mM NaCl		75mM NaCl				
С	55% PEG 600	25mM NaCl											
D	60% PEG 600												
Ε	45% PEG 600	FOm MANACI											
F	50% PEG 600			FOrm MANACI		100mM NaCl		100mM NaCl		laCl			
G	55% PEG 600	50mM NaCl			50	50mM NaCl		100mivi NaCi		1001111VI NaCi			
Н	60% PEG 600												

The optimization screening yielded crystals in some cases but did not seem to improve stability of the wild-type crystals.

3.4.2.2 Thermal Shift Analysis (TSA)

To determine whether wild-type PEX14 was stable under the different buffer conditions, we applied TSA to screen for stabilizing effects on the studied protein. The concept of TSA is monitoring the thermal denaturation of protein in different conditions. This is done through incubating the folded protein with SYPRO Orange and systematically increasing the temperature. In case of stabilization this can give rise to a 'thermal shift' towards higher unfolding temperature and hence quantifies the stabilization of the protein in different buffers. The observed thermal shift is recorded by the change in fluorescence resulting from SYPRO Orange. Upon unfolding, the hydrophobic core of the protein becomes exposed to the solvent and the hydrophobic SYPRO Orange dye binds to the protein. A shift in the fluorescence peak and hence 'thermal shift' indicates whether the unfolding temperature of the protein indicates increasing stabilization in the studied buffer condition. ¹⁵⁵ We ran the experiments with the wild-type protein but could only observe a very noisy plot. When repeating the experiments with the E1W PEX14 variant, the plots were not noisy but they lacked any fluorescent peaks, and thus no differences in the thermal stability of the protein could be observed (Figure 3.26). As our experimental set up worked with other proteins like endothiapepsin, we attributed the inconclusive plots of PEX14 to the small size of the protein (only three alpha helices of 66 amino acids) and its hydrophobic nature, which would not produce any visible differences in the observed fluorescence upon unfolding, as has been reported with other small proteins that unfold in a two-state transition

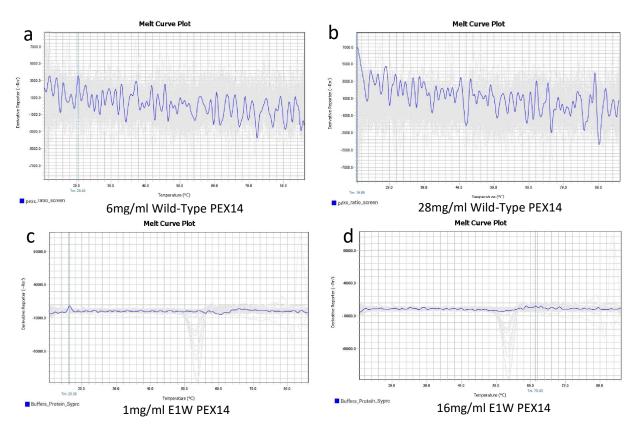


Figure 3.26: Melting plots produced from the thermal shift analysis of wild-type PEX14 at different concentrations (a) and (b) show only a noisy baseline, whereas the melting plots c) and d) of the E1W PEX14 variant show no signal, which indicates no observable differences with PEX14. This may be attributed to the small size of the PEX14 protein, with only three alpha helices and its hydrophobic nature which may not produce any visible differences upon unfolding of the protein from its quaternary

3.4.2.3 Dynamic Light Scattering (DLS) for Stable Buffers

DLS is a technique that is used to determine the size distribution profile of small particles in a suspension, as well as polymers in a solution. We used DLS to screen for another buffer that the wild-type protein would be more stable in, thus improving chances of finding promising crystallization hits. Thanks to Patrick Walter from the Pasteur Institute, the results of DLS revealed that the protein was already quite stable in its storage buffer (5 mM Tris pH 8, 150 mM NaCl) as seen in Figure 3.27a. Another buffer that gave a slightly more stable protein was 20 mM MES pH 6 and 300 mM NaCl as seen in Figure 3.27b. After performing a buffer exchange, a new crystallization screen of over 800 conditions for the wild-type protein resulted in a single hit (60 mg/mL of protein, 0.1 M Tris pH 8, 0.2 M Calcium Chloride, 20% PEG 6000 at 4°C). The conditions were irreproducible, likely owing to the crystal being a salt crystal.

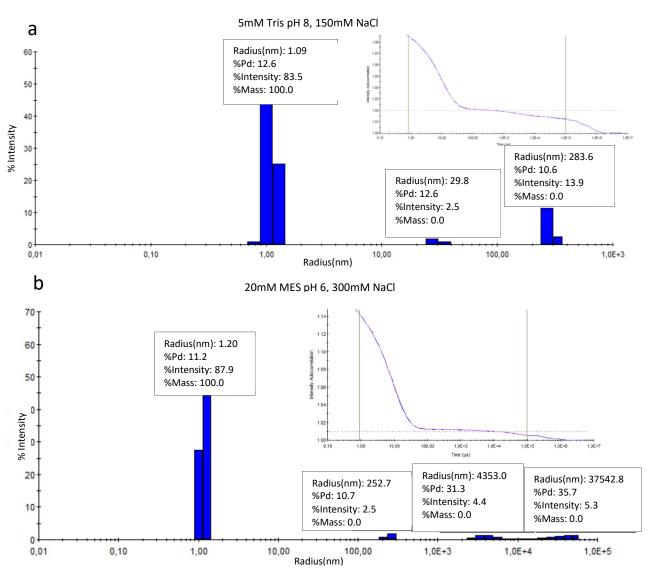


Figure 3.27: DLS measurements of the wild-type protein in its original storage buffer (top) showed that the protein was already monodispersed and stable. Upon screening the protein in different buffers, a second buffer was found to also keep the protein particles monodispersed and stable (bottom).

3.4.2.4 Additives

Our final attempt at optimizing the wild-type crystals for soaking, was using additives. We ran a screen using TEW $[TeW_6O_{24}]^{6-}$ (Tellurium-Centered Anderson-Evans Polyoxotungstate) (Figure 3.29), a technique reported in literature to facilitate crystal lattice formation by mediating and stabilizing crystal contacts by electrostatically cross-linking protein monomers.¹⁵⁸

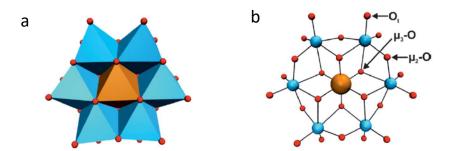
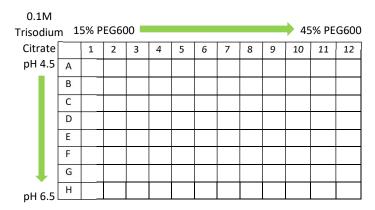
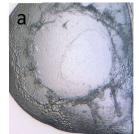


Figure 3. 28: Polyhedral (a) and ball and stick (b) representation of $[TeW_6O_{24}]^{6-}$. Different coordination modes of the oxygen atoms are assigned in panel B. Color code: tungsten, cyan; tellurium, ochre; oxygen, red.' ⁷

We set up two crystallization plates, one with TEW added at a concentration of 2 mM to the protein, and the other plate with only protein. In addition to the TEW, the crystallization drops were set up in a way to allow screening around the crystallization conditions (Table 3.15). However, TEW did not enhance our crystallization. In fact, small crystals could be observed in the plate without TEW whilst the plate with TEW gave no crystals. In the case of the wild-type PEX14 it appears that TEW somehow hinders the crystallization (Figure 3.29).

Table 3.15: Crystallization screening conditions of wild-type PEX14 in the absence and presence of 2 mM of TEW.





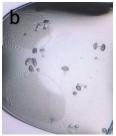


Figure 3.29: Image of the same crystallization drops in the presence (a) and absence of TEW (b).

3.5 Wild Type with a Bound Ligand

3.5.1 Soaking of Wild Type Crystals

As discussed above, trials to optimize the wild type crystals were not successful, and several attempts to yield crystals of the mutated PEX14 construct with unblocked binding pockets lead us to slightly different crystal forms, all of which did not have completely accessible pockets. The crystal structure of the wild type PEX14 proved to be more suited for fragment screening as it has unblocked binding pockets, which allowed it to accommodate compound 1 through soaking.

3.5.2 Wild Type PEX14 in Complex with a Lead Compound

We succeeded in resolving a structure of the wild-type PEX14 with a ligand, somewhat larger than fragments, which proves the suitability of the studied wild-type crystal form for soaking as it harbors solvent channels large enough for a ligand to diffuse into the crystal. To keep wild-type crystals as stable as possible during soaking and avoid mechanical damage that could occur upon transferring the crystals into soaking solutions, the DMSO stock of our lead compound 1 was added directly into the well with the wild-type crystals at a final concentration of 143 mM, for 23 hours.

A dataset of 1.83 Å of the wild-type crystal in complex with our synthesized lead compound $\bf 1$ (Scheme 3.1) was collected. The compound $\bf 1$ results in a crystal with two molecules in the asymmetric unit for which the diffraction data suggested a setting of the crystal symmetry in the orthorhombic space group C222₁. After structural refinement of the dataset, we noticed that both the R_{work} (27.9%) and the R_{free} (33.0%) values were in the high range for the collected resolution, with a large gap of 5.1% between the values. Additionally, the average B-factor of the protein molecule A was 21.3 Ų, whereas the average B-factor of the protein molecule B was 37.7 Ų, which is almost twice as large. This made us question whether the data was processed in the correct space group. To validate the chosen space group, the dataset was processed in the lowest symmetry space group P1 and the XDS_ASCII.HKL file from processing was run in XPREP 153 to search for higher symmetry space groups. The practical processing statistics of all space groups the dataset was processed in are shown in Table 3.16. XPREP is a program used to analyze the symmetry of the data and to merge the reflections according to the space group found. The output from XPREP is shown in Figure 3. 30. The monoclinic P lattice with the calculated lowest R-sym value was selected, and the dataset was processed in space group P2₁.

Table 3.16: The data processing statistics of wild-type PEX14 in complex with compound 1 after processing in three different space groups.^a

Space group	C222 ₁	P1	P2 ₁
R-sym (%)	10.8	4.9	6.8
R-sym (%) highest resolution shell	49.6	39.3	45.1
R-sym (%) lowest resolution shell	5.4	2.4	3.4
Completeness	98.9 (97.8)	76.3 (77.7)	91.0 (93.2)
Multiplicity	4.5	1.6	2.6
Resolution	1.83 Å	1.8 Å	1.8 Å

a) Data in parenthesis indicates the higher resolution shell.

```
Determination of reduced (Niggli) cell
Transformation from original cell (HKLF-matrix):
  0.0000 -1.0000 0.0000 1.0000 0.0000 0.0000
                                                     0.0000 0.0000 1.0000
Unitcell:
              38.892
                       39.210 80.881 90.03 90.00
                                                       92.76
Niggli form:
                      1512.59
                                   b.b =
                                          1537.42
                                                                6541.74
               a.a =
               b.c =
                        -1.88
                                   a.c =
                                           -0.11
                                                       a.b =
Search for higher METRIC symmetry
Identical indices and Friedel opposites combined before calculating R(sym)
Option A: FOM = 0.468 deg. ORTHORHOMBIC C-lattice R(sym) = 0.177 [ 22203]
Cell: 53.881 56.541 80.881 90.02 90.03 89.53
                                                        Volume:
Matrix: 1.0000 -1.0000 0.0000 1.0000 1.0000 0.0000 0.0000 1.0000
Option B: FOM = 0.034 deg.
                           MONOCLINIC P-lattice
10 89.97 92.76 90.0
                                                   R(sym) = 0.081 [
                                                                     133351
       38.892 80.881 39.210
                                               90.00
                                                        Volume:
                                                                  123196.79
Matrix: 0.0000 -1.0000 0.0000 0.0000 -1.0000 1.0000 0.0000 0.0000
                           MONOCLINIC C-lattice R(sym) = 0.185 [ 12786]
B1 90.02 90.03 89.53 Volume: 246393.56
Option C: FOM = 0.468 deg.
       53.881 56.541 80.881
Matrix: 1.0000 -1.0000 0.0000 1.0000 1.0000 0.0000 0.0000 0.0000 1.0000
Option D: FOM = 0.468 deg. MONOCLINIC C-lattice R(sym) = 0.200 [ 12206]
       56.541 53.881 80.881 89.97
                                              90.47
                                                        Volume:
                                       90.02
                                                                  246393.56
Cell:
Matrix: 1.0000 1.0000 0.0000 -1.0000 1.0000 0.0000 0.0000 0.0000 1.0000
Option E retains original cell
```

Figure 3. 30: The space groups of higher symmetry suggested by XPREP for the dataset, based on the calculated R-sym value. The monoclinic P-lattice space group has the lowest R-sym value at 8.1%.

The processing in $P2_1$ gave the most promising statistics, and Matthew's coefficient suggested four molecules in the asymmetric unit and 39% solvent content in the crystal. MR was performed using the MTZ file from the $P2_1$ processing with four molecules in the asymmetric unit. The solution from MR gave four molecules; A, B, C, And D. For molecules A, B, and C the average B factors were comparable with molecule C having slightly higher values. For molecule D however, the B factors were significantly higher, suggesting an incorrect solution. MR was performed again but three molecules instead were searched for in the asymmetric unit. The solution gave three molecules in the asymmetric unit; A, B, and C. The B factors were comparable for all three molecules with molecule C having slightly higher B factors (molecule A 14.5 Ų, molecule B 17.9 Ų, molecule C 26.2 Ų). Despite better R-sym values of the dataset after processing in the $P2_1$ solution, refinement in PHENIX (simulated annealing: Cartesian, and XYZ coordinates) of the MR solution did not give a clear electron density for the binding ligand, so

we proceeded with the original data processed in orthorhombic space group. This is similar to the apo structure of the wild-type PEX14 described earlier in section 3.3.4, which was also processed in the higher symmetry space group. To check the correctness of the MR solution which gave two protein molecules – A and B - in the asymmetric unit, the PDB file from MR was run in the program ARP/wARP which is used to build macromolecular models in the electron density map. ARP/wARP built the entire molecule A in the electron density but due to the poor dataset quality, it was only able to build half of molecule B, due to disorder in molecule B. This disorder was not seen in the refined structure due to model bias. For molecule A however, the electron density shows a completely resolved protein molecule in addition to the electron density seen for compound 1 binding to molecule A. Figure 3.31 depicts the binding mode of the lead compound 1 to the binding pockets of molecule A (**crystal form 9**).

Superimposition of the structure of the wild-type in complex with compound 1 and the original fragment hit obtained from soaking the E1W PEX14 crystal form 1 as seen in Figure 3.32 shows the agreement between the benzimidazole moiety of compound 1 and the indole moiety of F5. The synthesized compound 1 is flipped in comparison to the fragment hit so that the phenyl ring of F5 overlays with the imidazole of compound 1, and the phenyl ring of compound 1 overlays with the pyrrole of F5 (Figure 3.31). The overlay of the bicyclic systems however is preserved which fixes the orientation of the compound 1 to fit into the binding pocket.

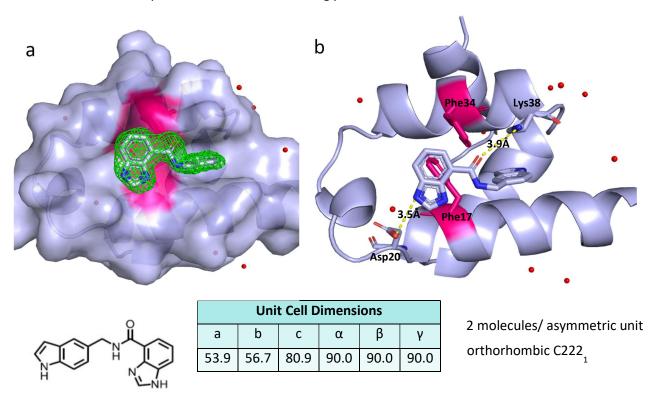


Figure 3.31: First crystal structure of the wild-type PEX14 in complex with a ligand – compound 1. a) The ligand binds in the binding pocket that usually becomes occupied by Trp103 of PEX5. b) The benzimidazole moiety of the ligand forms a hydrogen bond (3.5 Å) between one of its nitrogens and the carboxylate of Asp20. The carboxamide of the ligand also interacts with the terminal nitrogen of Lys38 (3.9 Å). The ligand is shown with its 2m|Fo|-|Fc| density map in green which is contoured at a standard sigma level of 1, hydrogen bonds are shown as yellow dotted lines, and distances shown are in Å. The phenylalanine residues separating the hydrophobic pockets are shown in pink. The chemical structure of lead compound 1, unit cell dimensions, space group, and number of molecules per asymmetric unit are also shown.

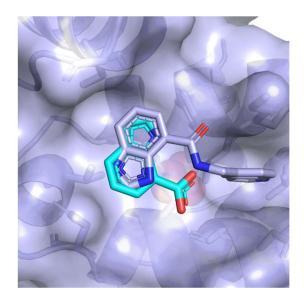


Figure 3.32: Superimposition of the wild-type compound 1 complex (purple) and the original fragment hit F5 obtained from soaking the E1W PEX14 crystal form 1 (cyan). The benzimidazole moiety of compound 1 overlays with the indole moiety of F5 but rather than the phenyl rings of both compounds overlaying, the superimposition is flipped so that the phenyl ring of F5 overlays with the imidazole of compound 1, and the phenyl ring of compound 1 overlays with the pyrrole of F5

3.6 Materials and Methods

3.6.1 Protein Expression and Purification

3.6.1.1 Wild-Type and Mutated T. brucei N-His-PEX14 (19-84)

The E. coli cells BL21-CodonPlus (DE3)-RIPL (Cam r) were transformed with the plasmid vector pETM-11 (EMBL) of interest (either the wild-type or E1W T. brucei N-His-PEX14 variant (19-84) provided by the group of Prof. Dr. Michael Sattler, HMGU Munich and incubated in a pre-culture of 125 mL LB medium containing 30 mg · L⁻¹ kanamycin for 17 h at 37 °C and 220 rpm. In addition to kanamycin resistance, the plasmid contained a sequence encoding for an N-terminal 6xHis-tag® separated from the PEX14 start codon by a spacer sequence and a sequence encoding a TEV protease cleavage site. The pre-culture was added to 6×1 L main culture (ZY auto-induction medium including 100 mg . L ¹ kanamycin, 1 mg. L⁻¹ thiamin, and 1 mg. L⁻¹ biotin) which is incubated at 37 °C and 220 rpm until an OD₆₀₀ of 0.6-0.8. and moved to a 20°C shaker overnight. The cell pellets were harvested by centrifugation (9,000 rpm at 4 °C for 8 min.) and re-suspended in 250 mL lysis/TRIS buffer (50 mM TRIS pH 8, 20 mM imidazole, 300 mM NaCl, 0.5 mM TCEP and 2 cOmplete™-Protease Inhibitor Cocktail Tablets (Roche) per 1.5 L of bacterial culture). Cell disruption was achieved via an EmulsiFlex-C5™ highpressure homogenizer (Avestin Europe GmbH). The lysate was clarified by centrifugation (centrifugation speed 19,000 rpm, 45 min, 4 °C). Purification of the protein was achieved at room temperature by a two-step purification including affinity chromatography and size-exclusion chromatography (SEC). The protein was loaded onto an affinity chromatography prepacked NiNTA column (HisTrap HP 5 mL column from GE Healthcare) that had been equilibrated with 5 column volumes of elution buffer (50 mM TRIS pH 8, 250 mM imidazole, 300 mM NaCl, 0.5 mM TCEP) followed by 10 column volumes of TRIS buffer. After loading the protein onto the column, the column was washed with 5% of elution buffer, and the protein was eluted with 100% elution buffer. The collected protein was checked for purity in an SDS gel, diluted with TRIS buffer in a 1:1 ratio, and the buffer was exchanged to remove imidazole using room temperature dialysis overnight (50 mM TRIS pH 8, 1 mM EDTA, 150 mM NaCl, 0.5 mM TCEP). The following day the protein was spun down to remove any precipitation and the protein concentration was measured. In a fresh dialysis buffer, 1 mg TEV protease was added per 10 mg protein and left to cleave for 3 hrs at room temperature. Separation of the cleaved protein and His tag was achieved using a reverse NiNTA column equilibrated with 5 column volumes of TRIS buffer. The cleaved protein was eluted with 100% TRIS buffer and purified into crystallization buffer (5 mM TRIS pH 8, 150 mM NaCl, 0.5 mM TCEP) using a Superdex75 SEC column for the final purification step. The fractions containing PEX14 protein were then concentrated in a VIVASPIN®20 centrifugal concentrator (Sartorius, MWCO = 3,000) until a final concentration of approximately 16 mg \cdot mL $^{-1}$. Finally, the protein was flash frozen into aliquots of 100 μ L and stored at 80 °C.

Pierce BCA Protein Assay

For measuring the concentration of the wild-type PEX14, as it lacks any intrinsic fluorescence, the Pierce™ BCA (bicinchoninic acid) Protein Assay was used as indicated by Thermo Scientific™. This assay is based on the colorimetric detection and quantitation of total protein. It combines the reduction of Cu²+ to Cu⁺ by protein in an alkaline medium (the Biuret reaction) with the selective colorimetric detection of the cuprous cation (Cu⁺) using bicinchoninic acid. The chelation of two molecules of BCA with one cuprous ion produces a complex which gives the purple-colored reaction product of the assay. This complex exhibits a strong absorbance at 562nm nearly linear with increasing protein concentrations over a broad working range (20-2000 µg · mL⁻¹). Protein concentrations are determined with reference to standards of a common protein bovine serum albumin (BSA).¹60

3.6.1.2 ¹⁵N labelled E1W *T. brucei* N-His-PEX14 (19-84)

The isotopically labelled protein produced for NMR and also used in the AlphaScreen[™] was prepared in a similar manner but expressed in a P-5052 minimal media¹⁵⁹ and with an additional 3 mg. L⁻¹ DNAse, 3 mg. L⁻¹ AEBSF.HCL, and lysozyme powder in the lysis buffer. Affinity chromatography was performed using a gravity-flow NiNTA column using the same buffers as the non-labelled PEX14, and the cleavage step was omitted as the His-tag is needed in AlphaScreen. The final purification was done using a Superdex75 SEC column using the NMR buffer (20 mM sodium phosphate pH 6.5 and 50 mM NaCl).

3.6.2 Crystallization

T. brucei E1W PEX14 crystal form 1 crystals were crystallized at 4°C using the sitting-drop vapor diffusion methods. 1.5 µL of protein were mixed with 1.5 µL of reservoir solution in the wells of a crystallization plate containing 550 μL of reservoir solution, where the protein concentration and the reservoir solution used varied, as has been previously discussed in detail in section 3.2. T. brucei E1W PEX14 crystal form 5 crystals were crystallized at 18°C using the sitting-drop vapor diffusion methods in presence of the VS ligands. 1.5 µL of the incubated protein and ligand solution (1 mM protein: 20 mM ligand) were mixed with 1.5 μL of reservoir solution in the wells of a crystallization plate containing 550 µL of reservoir solution (0.17 M ammonium acetate, 0.085 M sodium citrate pH 5.6, 25.5% PEG 4000, 15% glycerol). Crystals and cocrystals were soaked for a time ranging between 4 hours and 16 hours into a solution containing fragment or ligand solution (dissolved in 100% DMSO) in a soaking solution consisting of the reservoir solution with an additional 25% glycerol for cryoprotection. Crystals were directly flash-frozen in liquid nitrogen and stored for synchrotron data collection. Data collection and refinement statistics are described in the Appendix (Table 6 - Table 8). The diffraction data were indexed, scaled, and merged using XDS⁹⁰ and XDSAPP.⁹¹ Molecular replacement from the program PHASER MR⁹² from the CCP4 suite⁹³ was used to determine all crystal structures. Arcimboldo¹⁶¹ was used to create the model file using model helices from the PDB structure 5AON. In the refinement, a 5% subset of all reflections was omitted during refinement to be used for Rfree calculation. Model building was achieved in COOT⁴⁴ and refinement using PHENIX.refine version 1.10.1-2155. 94 Cartesian simulated annealing with default parameters was used as a first refinement step for all the structures. This was followed by refinement of XYZ coordinates and occupancies of protein residues and fragments (with the exception of water molecules whose occupancies were fixed). In the case of protein residues that gave additional density, they were refined in double confirmation and kept if their refined occupancy was ≥ 20%. %. The structure of E1W PEX14 in complex with F5 (PDB: 6S6R) was refined anisotropically except for water molecules. The structure of E1W PEX14 in complex with F6 (PDB: 6S7M) was refined isotropically with 30 TLS groups. The structure of wild-type PEX14 (PDB: 6S9Y) was refined isotropically with 6 TLS groups. The crystal structures of the un-soaked wild-type PEX14 and the structure of the wild-type PEX14 in complex with compound 1 contained unresolved motifs in the second protein molecule. Because they do not meet the criteria for PDB depositions, they were not deposited. Chemicalize⁹⁶ developed by ChemAxon⁹⁷ was used for name-to-structure generation and SMILES code notation. The ligand PDB and restraint files were generated with the Grade Web Server.⁹⁸

3.6.3 NMR Conditions

All NMR 1 H- 15 N correlation spectrum experiments were performed at 298K on a Bruker Avance 600 MHz NMR spectrometer equipped with a cryogenically cooled probe head. Automatic sample changing was accomplished with a SampleJet system. NMR samples were prepared for screening in a buffer consisting of 20mM sodium phosphate pH 6.5 and 50 mM NaCl, 10% D $_{2}$ O, 100 μ M of 15 N labelled PEX14, and 1 mM of VS ligand (50 mM stock in d $_{6}$ -DMSO). A blank sample was prepared with the equivalent volume of DMSO instead of the ligand. For the titration experiments, samples were prepared in the same way but at different ligand concentrations of 0.125 mM, 0.25 mM, 0.375 mM, 0.5 mM, 0.75 mM, 1 mM, 1.25 mM, and 1.5 mM. A liquid handling pipetting robot Gilson was used to fill 3mm NMR tubes in a SampleJet rack. A SOFAST HMQC pulse sequence with Watergate W5 water suppression (SFhmqcf3gpph) was used, with an interscan delay of 0.3 s and 16 scans. Shift perturbations between DMSO blanks and ligand samples were calculated with Pythagorus' theorem. Since the measurement range (7–9 ppm) of observations of 1 H is approximately one-tenth of that (105–124 ppm) of 15 N chemical shifts were scaled one-tenth relative to 1 H chemical shifts.

3.6.4 AlphaScreen-based Competition Assay Conditions

Experiments were carried out in a 96-well plate where 20 μ L of DMSO were added to the first column while the rest of the wells were filled with 12.5 μ L. In the first well of each row, 5 μ L of 50 mM DMSO ligand stock solution were added. Serial dilutions were carried out by transferring 12.5 μ L from each well to the next in every row, and discarding the volume from volume 11, leaving the final column with no compound. To every well, 200 μ L of assay mixture (prepared by mixing 40 mL of AS Buffer (consisting of 0.250 mg of BSA, 50 mL of 1XPBS buffer, and 50 μ L 10% Tween20), 48 μ L AS donor beads, 48 μ L AS acceptor beads, 1.2 μ L ALS (0.1mM) biotinylated PEX5-derived peptide (ALSENWAQEFLA), and 12 μ L 100 μ M 15 N labelled *T. Brucei* PEX14) were added and incubated for 1 hour at room temperature in dark conditions. Each well reaction was divided into 4 to fill a 384-well white Optiplate and measured as quadruplicates in an EnVision plate reader. The resulting curves were analyzed with Origin¹⁶⁴ software and are provided in the Appendix.

3.6.5 Trypanocidal Assay

Anti-trypanosomal activity of compound VS 25 was tested against the bloodstream form (BSF) of \mathcal{T} . brucei brucei using resazurin-based 96-well plate assay. Two-fold serial dilutions of VS 25 were prepared in 96-well plates in HMI-11 or SDM-79 medium on the wells per row. As controls, one well was left without addition of any compound and one well contained only media. Into all the wells (except the well with only media) 100 μ L of parasite cultures (4x10 $^3 \cdot m$ L BSF) were inoculated and the final concentration of parasites was 2x10 $^3 \cdot m$ L he plates were then incubated for 66 h followed by addition of 25 μ L resazurin (0.1 mg· mL hanks Balanced Salt Solution) and the plates were further incubated to a total of 72 h. Reduction of resazurin by the living cells was quantified by measuring the fluorescence with Synergy H1 microplate reader (excitation 530 nm, emission 585 nm). The background fluorescence of the well with only media was subtracted and the percent survival values were calculated by setting the fluorescence of the well without any compound to "100% survival". GraphPad Prism (version 6.04) was used to plot non-linear regression graphs and the from the resulting sigmoidal dose-response curve the half-maximal inhibitory concentration (IC50) value was determined.

Chapter 4 Endothiapepsin

Introductory Remarks

The following chapter is the preparation of a manuscript to be submitted.

The first STD NMR experiments were performed by Boeringer Ingelheim in a previous collaboration (Schiebel. J. et al., 2015).

The second STD NMR experiments were performed in collaboration with Charlotte Softley from the group of Prof. Dr. Michael Sattler during an academic scientific secondment at the Helmholtz Zentrum in Munich, Germany. The analysis of the NMR spectra was done by Charlotte Softley and the author of this thesis.

The WaterLOGSY NMR experiments were performed by Charlotte Softley. The analysis of the NMR spectra was done by the author of this thesis.

Buffer exchange of endothiapepsin and its crystallization was done by the author of this thesis, in addition to the soaking experiments, structural refinements, and analysis of results.

The SPR screen was done by the author of this thesis in collaboration with Dr. Stefan Geschwindner, and the deuterated versus non-deuterated STD NMR experiments and the images of their spectra used in this thesis were done by Dr. Per-Olof Eriksson during a scientific academic secondment at Astrazeneca R&D in Gothenburg, Sweden.

As part of a previous project, the X-ray crystallography cocktail experiments were done by Dr. Johannes Schiebel and Helene Koester, structure refinements and detailed analysis for cocktails A, B, C, and D were done by Dr. Nedyalka Radeva, and autorefinement for the remaining cocktail sets was done by Dr. Johannes Schiebel. Refinement of cocktail L was done by Dr. Johannes Schiebel. The NMR cocktail experiments for cocktail sets A, B, C, D, and E, were done by the author of this thesis and Charlotte Softley during the academic scientific secondment. The NMR cocktail experiments for cocktail sets F, G, H, I, J, K, L, M, and N were done by Charlotte Softley. The comparative analysis for all cocktail sets and the structural refinements of cocktails E, F, G, H, I, J, K, M, and N were done by the author of this thesis

ITC measurements to determine binding affinities of the X-ray fragment hits were done by Dr. Johannes Schiebel as part of a previous project (Schiebel. J. et al., 2016).

The comparative analysis of all methods was done by the author of this thesis.

Chapter 4 EP

4. 1 Introduction

4.1.1 Endothiapepsin

4.1.1.1 Family

Endothiapepsin belongs to the family of pepsin-like aspartic proteases. This class of enzymes is responsible for the hydrolytic cleavage of peptide substrates in vertebrates, fungi, plants, and viruses. Aspartic proteases gain their name from their conserved aspartic acid residues where the cleaving takes place by utilizing a catalytically activated water molecule. Endothiapepsin is an A1 subfamily aspartic protease from the chestnut blight fungus (*Cryphonectria* or *Endothia* parasitica), and has an acid pH optimum. It is often used in cheese production as a fungal rennet due to its lactic coagulation properties. Owing to the high degree of similarity between endothiapepsin and other pharmacologically relevant aspartic proteases, it has served as the model enzyme for studying the mechanism of aspartic proteases 171,172 as well as for the development of antihypertensives (inhibitors of renin) and treatment of Alzheimer's disease (inhibitors of β -secretase). Additionally, endothiapepsin is pursued as a surrogate for drug targets in malaria (plasmepsins), cancer (cathepsin D), and AIDS (HIV1 protease (human immunodeficiency virus type 1) protease).

4.1.1.2 Structure of Endothiapepsin

Endothiapepsin is a monomer of 330 amino acids and has a molecular mass of 33.8 kDa. It crystallizes in space group P2₁. Its active site is rather large and covered by the so-called flap region. This region must open to give access for a substrate or ligand to enter the binding pocket and in most cases closes again upon their binding. Endothiapepsin has nine binding pockets in its active site (Figure 4.1) as listed below.

- 1- S1' pocket: surrounded by Ile300, Ile302, and Ile304.
- 2- S2' pocket: surrounded by Ile77, Leu133, and Phe194.
- 3- Adjacent S3' pocket: exhibiting Ser78, Tyr79, and Gly80.
- 4- S2 pocket: surrounded by Thr222, Thr223, and Tyr226.
- 5- S3 pocket: surrounded by Ile10, Ala16 and Ile122 and partially covered by the flap region.
- 6- Adjacent S1 pocket: surrounded by Tyr79, Phe116 and Leu125.
- 7- Solvent exposed S4 pocket.
- 8- Solvent exposed S5 pocket.
- 9- Solvent exposed S6 pocket: surrounded by Leu13, Phe280, Ile283 and Phe291.

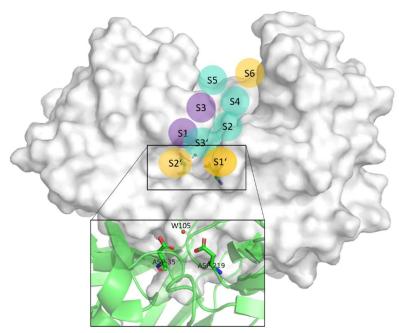


Figure 4.1: Surface representation of endothiapepsin with its binding pockets represented as spheres. Hydrophobic pockets are colored in orange, hydrophilic ones in blue, and amphipathic ones in purple. The catalytic site consisting of the conserved aspartate residues Asp35 and Asp219 as well as the catalytic water W105 (red sphere) are shown in the magnified rectangle.

4.1.2 Characterization of Binding Using Orthogonal Biophysical Methods

In previous work done by other members from our group to identify and characterize endothiapepsin binders, X-ray crystallography was consulted as a primary screening method and its hit identification potential or power was compared to several biochemical and biophysical screening methods which included high-concentration biochemical screens (HCS), microscale thermophoresis (MST), thermal shift assay (TSA), saturation transfer difference nuclear magnetic resonance (STD-NMR), reporter displacement assay (RDA), and native electrospray ionization mass spectrometry (ESI-MS) assay.¹⁷⁵ The fragment library screened was designed for general purposes and contained 361 entries. Further details of this fragment library design have been described in Chapter 1. The entire library was first screened against endothiapepsin via a fluorescence-based assay (HCS), resulting in a hit rate of 15%. The obtained hits from HCS had then been analyzed crystallographically using cocktails of two, which resulted in a hit rate of 20% of the original HCS hits. 10 The question that was raised at this point was whether an alternative screening method would have indicated similar promising hits to be followed up in a subsequent crystallographic study. 176 This led to the consequent study of the hit rates using the aforementioned screening methods. The fragment library was also screened as individual soaks on separate crystals of endothiapepsin, resulting in a hit rate of 20%, much higher to that obtained by two-entry cocktail screenings done as a follow up to the HCS screen, which gave only 3% of the entire library. The analysis and comparison of these cocktails with the individual soaking experiments demonstrated that the use of cocktails in crystallography to accelerate hit identification often leads to loss of hits, not to mention the detrimental impact of using a prescreening assay like the HCS which had missed out on 70% of the individual soaked X-ray hits. 175

To understand the reason behind the low overlap of hit rates found in the aforementioned screening techniques, we decided to prioritize two of these screening techniques to analyze in closer detail the underlining differences between them and the obtained fragment hit rates. We did this by selecting the 71 fragment hits detected by individually soaked crystals for X-ray crystallography¹⁷⁵ and rescreened them again with STD NMR under slightly different buffer conditions, in addition to

WaterLOGSY NMR experiments. It is worth mentioning that 44% of these 71 X-ray hits were not detected by any of the previously mentioned biophysical screening approaches. ¹⁷⁵ The outcomes of these additional screens will be discussed in the coming sections.

In addition to this comparative analysis, we added a new biophysical screen that was not analyzed in the previous studies, which was SPR. Thus, in this work we also screened the entire 361 fragment library with SPR, adding another biophysical method for our comparative hit rate analysis to give us further insight and understanding of which conditions are crucial to maintain while transferring across different techniques. Further details of this SPR screen are given in section 4.3.

4.2 Fragment Screening using Water LOGSY and STD NMR

The sequential NMR screening that we describe here was performed on only those confirmed 71 X-ray hits. These 71 fragments were screened with both Water LOGSY and a second slightly different STD NMR. The second STD NMR was a single fragment screen, rather than a cocktail screen, and a different NMR buffer was used. This was to test whether different experimental settings would yield similar results. In a previous STD NMR screen, 206 out of 361 fragments in our fragment library were initially screened¹⁷⁶, thus leaving out 155 fragments. These 155 fragments were deemed either impure (HPLC MS purity \geq 80%), poorly soluble (500 μ M), or prone to aggregation (self STD intensity <2% at 500 μ M), which is why they were left out of the initial STD NMR screen.¹⁷⁶ When the 206 fragments were screened, they were studied as cocktails of two compounds and 71 hits were detected. 22 of these STD NMR hits were confirmed as active-site binders in a competitive displacement assay with a potent active-site ligand namely, ritonavir. Interestingly, an additional 8 hits were found only in the presence of ritonavir (Figure 4.2).

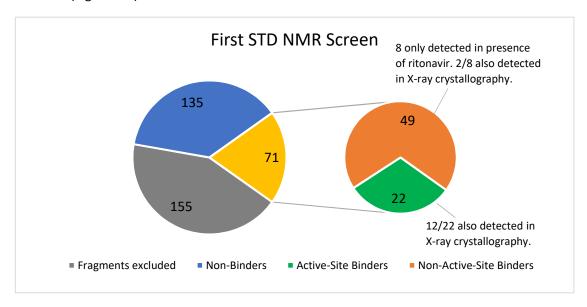


Figure 4.2: A pie of pie chart indicating the total number of fragments screened in the first STD NMR. 206 fragments were screened, 135 of which were non-binders (blue) and 71 were binders (yellow). 155 fragments were excluded from the screen (gray). From the 71 binding fragments (yellow), 22 were active-site binders (green) and 49 were non-active-site binders (orange). Amongst the non-active-site binders, 8 were detected only in the presence of the potent active-site ligand ritonavir, 2 of which were also detected by X-ray crystallography. From the active-site binders, 12 fragments were also detected by X-ray crystallography.

The entire 361 fragment library was also screened with X-ray crystallography. This resulted in 71 hits. The number of hits is purely coincidental, as the hits detected in both methods are not the same (Figure 4.3).

Upon comparison of the hits, it was found that 12 of the 22 STD NMR hits were also detected as hits by crystallography. As for the additional 8 STD NMR hits found only in the presence of ritonavir, only 2 of these were also detected as hits by crystallography. It is worth mentioning that 33 of the excluded fragments from the initial STD NMR screen were detected as hits by crystallography. Soothingly, 19 of these were detected as hits in the second STD NMR discussed here.

In the coming sections we will discuss the sequential NMR screening that was performed on only the 71 X-ray hits. This includes both a second STD NMR screen as well as a WaterLOGSY screen.

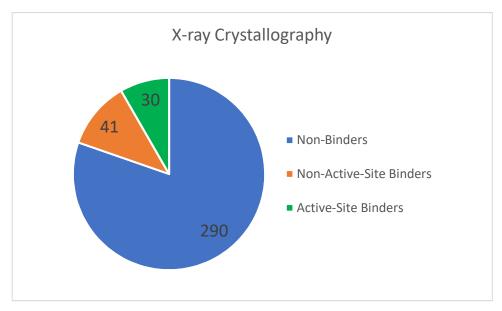


Figure 4.3: A pie chart indicating that from the 361 fragments screened using X-ray crystallography, 290 were non-binders (blue), 41 were non-active-site binders (orange) and 30 were active-site binders.

4.2.1 Fragment Hit Rate with WaterLOGSY

The 71 fragment hits detected by X-ray crystallography were screened by WaterLOGSY NMR. The concept of WaterLOGSY (Water-Ligand Observed via Gradient Spectroscopy) relies on the partial transfer of the magnetization of bulk water to the free ligands in solution via the protein-ligand complex. The resonances of free non-binding ligands appear in the spectra with an opposite sign to the resonances of the interacting ligands. The signs of the ligands with and without protein addition, the interacting ligands and thus hits can be identified. The number of hits in WaterLOGSY was 50 fragments, which corresponds to a 70% overlap in the fragments detected by X-ray. The fragments that were not detected as hits in WaterLOGSY (30%) were also not detected by either of the two STD NMR screens. The exception to this is five fragments shown in Table 4.1.

Table 4.1: Chemical structures of fragments that were detected as hits in STD NMR but not in Water LOGSY.

Fragment	Chemical Structure	Corresponding STD NMR that detected fragment as hit	Binding Mode in the crystal
158	N N N O-N	Second STD NMR screen (absence of salt)	S1 pocket (Asp81)
216	F NH	Both initial and second STD NMR screens	Catalytic Dyad
306	H ₂ N S	Both initial and second STD NMR screens	Catalytic Dyad
311	S O	Second STD NMR screen (absence of salt)	S6 pocket (Phe291)
333	s N N N	Second STD NMR screen (absence of salt)	S6 pocket (Phe291)

4.2.2 Comparison of Fragment Hits and Hit Rate with and without Salt in STD NMR

The initial STD NMR screen was performed at a buffer concentration of 50mM CD_3COOD in D_2O at a pH of 4.3, in the presence of 100mM NaCl and 4mM NaN₃ which is added as an antibacterial agent.¹⁷⁸ The resulting hit rate was 27% (when considering only the 71 X-ray hits).

The second STD NMR screen was performed at a buffer concentration of 3mM CH₃COOH in H₂O at a pH of 4.6, in the absence of any salt. The resulting hit rate was 59% of the tested 71 fragments.

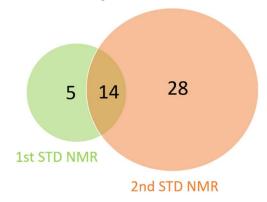


Figure 4.4: Venn Diagram depicting the number of fragment hits overlapping between both STD NMR screens, as well as the additional fragment hits of each screen independently. In the first STD NMR screen in green, a total of 19 fragments were detected, 14 of which were also detected in the second STD NMR screen in orange. An additional 28 fragments were detected in the second STD NMR screen but not in the first, 19 of which were not screened in the first STD NMR due to being deemed either insoluble, impure, or prone to aggregation.

Role of Buffer in NMR Sensitivity

We must keep in mind that the presence of salt in a buffer increases the ionic strength of the solution, which can affect the binding interactions between the protein and the fragments. The strength of

hydrophobic interactions in the protein will likely increase with increasing ionic strength. The addition of salt can also weaken the electrostatic interactions experienced between the ligand and the protein. Not only that, but the salt in an NMR buffer can affect the sensitivity of the NMR probes. A buffer with salt produces a conductive sample, where the conductivity is a function of both the salt concentration and the mobility of the ions in solution. ¹⁷⁹This electrical conductivity adds a resistance to the coil which causes significant reduction in the signal-to-noise ratio.¹⁸⁰ Of course solutions buffered with salts are required in NMR measurements to keep the pH constant and the macromolecules in a defined protonation state, as well as to increase the solubility. But one must also keep in mind the additional ions of strong bases or acids that are added to the solution whilst adjusting pH values. When adjusting pH values, this is achieved through titration of weak acid with a strong base (HCl) or weak base with a strong acid (NaOH).¹⁸¹ These additional ions contribute greatly to the detrimental effect on the sensitivity because they are of a high mobility. 181 Additionally, buffers with multiple charges reduce sensitivity more than buffers with single charges, especially if this translates to an increase in the counterion concentration. It has been shown by several researchers that the typical salt concentrations of 100-150 mM do decrease the sensitivity advantage of a cryogenic probe head to about a factor of two compared to that of a conventional probe with the same sample 182, and they can lead to longer pulses and decreased sensitivity. 183 That been said, other researchers have reported that this effect is seen only at higher concentrations of salt between 200-300 mM. Additionally, as the first STD NMR screen in our work was performed in 2.5 mm NMR tubes, the effect is much less significant than if performed in a 5 mm tube as the total amount of salt in the sample is less. 180 Finally, the STD signal was interpreted as % STD_AV where it was compared to a regular 1H spectrum, and because both spectrums are affected just as much by salt, when they are divided to calculated the percentage the salt effect cancels out.

To study this possible effect of salt on fragment binding, we examined fragments that were only detectable as binders in the presence of higher salt concentration (higher concentrations of acetate buffer, sodium azide, and sodium chloride which contribute more ions of CH₃COO⁻, H⁺, Na⁺, N⁻³, and Cl⁻) in the screening buffer, and the fragments that appear to only bind in the absence of this salt. We analyzed their predicted charged state at a pH of 4.3 and 4.6, as seen in Table 4.2, as well as the charge of the protein. The predicted pKa values of the fragments were calculated using the webserver chemicalize.96 Endothiapepsin has a theoretical pl of 4.14 (calculated using Protparam184), which suggests that at a pH of 4.3 and 4.6, the protein is likely negatively charged. As seen from Table 4.2, most of the fragments are also charged at a pH of 4.3 and 4.6. This means that the interactions they make could either be stabilized or destabilized by the change in salt concentration and thus ionic strength in the buffer solution. Additionally, the fragments that were not detected in the first STD NMR screen could have been missed due to the loss of sensitivity in the NMR probe as discussed above, although unlikely. 180 Considering the four fragment hits that could only be detected in the initial STD NMR, for one of these fragments – F039 - the reason for its discovery is the use of deuterated buffer rather than non-deuterated buffer. The effect of using deuterated or non-deuterated buffers in STD NMR will be described in detail in section 4.2.3. As for the other four fragments we believe the reason could be due to the described effect of salt on the stability of the interaction between protein and fragment. The salt may be stabilizing the fragment binding to endothiapepsin and improving its binding efficiency. The presence of salt may be important for fragment solubility in NMR screens but not in Xray crystallography where the fragment is soaked at such high concentrations that absence of salt supposedly does not affect binding. To trace such dependencies, F236 and F114 were soaked into endothiapepsin crystals in a buffer of 100 mM NaCl to compare their binding to the protein in presence and absence of high salt concentrations. If the presence of salt in the solution enhances binding to the crystal even further, this may be possible to see in terms of achieved fragment occupancy in the crystal structure. For F236, the presence of salt did not appear to enhance binding and surprisingly, we could not see any density for the F114, even at a resolution of 1.15 Å for the collected data set. In addition to the effect of NaCl, we wanted to see whether the NaN₃ in the first STD NMR buffer was interacting with endothiapepsin in a way that would hinder fragment binding. Apo crystals of endothiapepsin were soaked in a solution of 125 mM NaN₃ and a data set was collected at a resolution of 1.16 Å. No electron density could be seen for NaN₃ in the resulting crystal structure. Based on the crystal structure at least, we can eliminate hinderance of fragment binding by NaN₃.

Table 4.2: Chemical structures and predicted pKa values at pH 4.3 and 4.6 of fragment hits detected in only the first STD NMR screen and only the second STD NMR screen. Most likely present microspecies are also depicted below. ^a

First STD NMR	Predicted pk _a	Corresponding acidic form	Second STD NMR	Predicted pk _a	Microspecies
14	Br CH ₃ CH ₃	Br CH ₃	31	2.82 _{H2} N CH ₃	H ₂ N CH ₃
39	9 H N	H++ C	52	7.30 H ₂ N Br	H ₃ N Br
114	N NH ₂ N-NH 14.04	70% charged species, 30% neutral species	158	CH ₃ CH ₃ CH ₃	H,C NH NH
	H ₃ C N N N N N N N N N N N N N N N N N N N	NH ₂ +		H ₂ N 5,22 CH ₃	74% top species , 13% bottom species.
236			171		H ₂ N
291	5.14 N 4.29	Top to bottom species; 46%, 36%, 11% 7% neutral	209	4.82 N N N CH	76% charged species, 23%

	NH ₂			
		211	9.12 _{H2} N CH ₃	H ₃ M CH ₃
		311	0 0	
		333	H ₃ C CH ₃	S CH ₃
		285	H ₃ C NH ₂ ^{8.91}	H ₃ C NH ₃ *

a) pKa values and microspecies percentage were calculated using the server chemicalize.com

Table 4.3: Chemical structures, binding affinities, and endothiapepsin binding positions of all seven X-ray fragment hits that were measured in both STD NMR screens but could not be detected in either of them.

Fragment	Structure	Binding affinities ^a (K _d in mM)	Binding Position in Endothiapepsin
4	N, NH ₂	>10	Catalytic dyad (S1 and S1')
56	CIOH	n. d	Surface
227	Вг	n. d	Surface
266	о о о о н	>10	Catalytic dyad (water mediated)
51	N N N N N N N N N N N N N N N N N N N	1	S3/S5 pocket (Asp119)
73	NH NH	>10	S1 pocket (Asp81)
207	HN N	4.4	Catalytic dyad, S1 pocket (Asp81) S6 pocket (Phe291)

a) Binding affinities were determined via displacement ITC ¹⁷⁵

There are fragments that did not show up as hits in either of the STD NMR screens, as seen in Table 4.3. There are several aspects that could explain this, most apparently is the low binding affinity. As seen in Table 4.3, the majority of these fragments - with the exception of fragments 51 and 207 - have binding affinities either too low to be quantified or weaker than 10 mM.

In case of fragments of higher potency like F051 which has a K_d of 1 mM, the crystal structure reveals a direct interaction with Asp119 through the pyridine-type nitrogen atom (H-bond distance 2.7 Å). This suggests that either the Asp119 or the pyridine ring must be in a protonated state for the binding to occur. The calculated pK_a of the pyridine-type nitrogen atom is predicted to be 4.09, meaning at a pH of 4.3 or 4.6 in equilibrium both the protonated and neutral species are present at a ratio of 35% and 62%, respectively. With the higher probability of the neutral species being present, it is more likely that Asp119 is protonated. Additionally, a glycerol molecule is found at a van der Waals distance from the fragment which stabilizes F051 in position (Figure 4.5). In NMR, Asp119 may be deprotonated which could explain why F051 does not bind in NMR, in addition to the absence of glycerol in the NMR buffer to support fragment stabilization. Another factor that could explain why X-ray fragment hits do not show up in NMR is fragment solubility. In X-ray crystallography soaking a crystal in the presence of solid material of the fragment is still feasible, whereas in NMR clear solutions with no precipitation are preferred. This could be a possible explanation to why F207, with a binding affinity of 4.4 mM, could not be detected by the STD NMR.

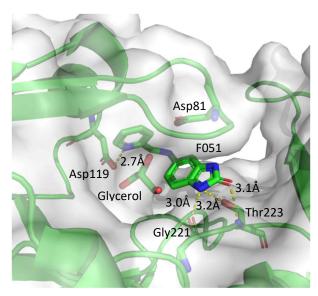


Figure 4.5: The binding mode of F051 (PDB: 4Y4X) in the protein crystal shows a hydrogen bond (2.7Å) between the nitrogen of the pyrimidine moiety of F051 and the carboxylate of Asp119, suggesting that either the pyrimidine or the Asp119 must be protonated. Additionally, the fragment is further stabilized in its binding position by the presence of a glycerol molecule which is found in the soaking conditions of the protein crystal.

4.2.3 Displacement of Fragments by Reporter Ligands and Vice Versa.

To test whether fragment hits were active-site binders or non-active-site binders, a displacing compound with high affinity was used. In our experiments, this compound was an active-site specific inhibitor, ritonavir, with a K_i value of 15 nM for endothiapepsin (Figure 4.6). When ritonavir is bound to endothiapepsin, only non-active-site binders will continue to bind while active-site specific binders will be competing with the high affinity ligand ritonavir for the binding site.

In the initial STD NMR screen, a total of 19 fragment hits were selected as binders. When tested in the presence of ritonavir, this number dropped to 14. However, we noticed that two of the fragment hits in the presence of ritonavir did not bind in the absence of ritonavir, which indicated that the presence of ritonavir somehow aided the binding of these additional hits. We also noticed this phenomena when

analyzing the results of the second STD NMR screen where a total of 39 fragment hits were selected as binders but when screened in the presence of ritonavir, the number of fragment hits dropped to 29 and one of which was only binding in the presence of ritonavir. Additionally, some fragments that bind to the catalytic dyad in the crystal structures obtained by the individual soaks continued to bind even in the presence of ritonavir. As ritonavir is an inhibitor smaller than the inhibitor pepstatin¹⁸⁹, it still leaves some unoccupied sub-pockets in the binding cleft of endothiapepsin. This provides the opportunity for additional interaction sites for putative fragments, which can recruit additional fragment hits as we observed. To check whether these fragment hits were binding in the unoccupied clefts or perhaps to the ritonavir itself, we used crystallography to map the details of the interactions. Structures of these fragments in addition to their binding affinities are shown in Table 4.4. In the first set of experiments, endothiapepsin crystals, presoaked in ritonavir, were transferred into single soaking solutions of the fragments, however the crystal structures revealed only an electron density for the fragments with no sign of ritonavir. This made us consider whether the ritonavir was simply washed out of the crystals upon transfer into the new solution without ritonavir, rather than being displaced by the fragment. To avoid this potential limitation, the presoaked crystals were transferred into soaking solutions containing both the fragment to be studied and ritonavir, so that washing out of ritonavir is limited. The observed electron densities can be seen in the Figure 4.7– Figure 4.10 below, the 2m|Fo|-|Fc| density contoured at a σ level of 1 is colored in green, while the m|Fo|-|Fc| density contoured at a σ level of 3 is colored in blue.

For some fragments, soaking into a solution of both, ritonavir and the fragment, resulted in an electron density of only ritonavir. Although it did not explain the result of the STD NMR where the fragment was seen to bind in the presence of ritonavir, it was indeed the expected outcome of soaking a fragment with such a high affinity reporter ligand like ritonavir. This holds true for F109, F236, F267, and F268, as seen in Figure 4.7. For other cases, the fragment manages to bind in the presence of ritonavir, and we can clearly fit both fragment and ritonavir into the electron density, which explains the results of the STD NMR. This holds true for F216 and F218 where they change their binding mode from the catalytic dyad to a new location where usually a PEG molecule can be found in the crystal structure obtained from endothiapepsin crystals soaked with the fragment alone. This can be seen in Figure 4.8. Such an active-site fragment displacement of potent inhibitors was also observed by Dr. Alexander Metz and Dr. Jan Wollenhaupt with the potent endothiapepsin inhibitor SAP114, (Ki 560 nM) (personal communication). Simultaneous soaking of SAP114 with a fragment-sized follow-up compound resulted in a crystal structure where the initial active-site fragment found in single soaking is displaced, but a second copy of the fragment can be seen that binds remotely. This second fragment copy binds in the same position where a PEG molecule is often seen in soaked crystal structures of endothiapepsin (PDBs 4YCT, 4YCY), and is conserved in soaking with and without SAP114 (Figure 4.9). To our surprise, we even found instances where the fragment managed to displace the ritonavir and bind instead of it. It is particularly astonishing as a strong binding displacer such as ritonavir with a binding affinity of 15 nM can still be displaced by weak binding fragments. This could be due to the exposure of the crystals to very high concentrations of the fragments up to 100 mM. But more importantly, ritonavir might have a different affinity in the crystal structure than in solution, due to packing contacts. This has been studied before with other protein and ligands. A prominent example is one where aldose reductase was used to study the extent to which the crystalline state of protein affects the binding and affinity of the ligand. 190 This was done through a competition experiment where two ligands, with different potencies but targeting the same binding site of aldose reductase, where added simultaneously to the protein at different concentrations. The ratio of the ligand occupancies was measured by X-ray crystallography and by mass spectrometry. The results in solution showed a higher occupancy for the ligand with the higher affinity, whereas the results in crystal showed a 4.6 times lower occupancy for the ligand with the higher affinity. This implies that the ligand potency is influenced by crystal contacts. Displacement of ritonavir by the fragments is seen with F290, F306, and F063 (Figure 4.10). When considering the measured binding affinities of the fragments, there appears to be a relationship between the ability of a fragment to displace ritonavir, and its binding affinity. We see this displacement for the most potent fragments which are F290, F306, and F063 whose binding affinities are 0.1 mM, 0.1 mM, and 0.5 mM, respectively. On the other hand, the binding affinities of fragments that could not displace ritonavir being F109, F236, F267, and F268, are 1 mM, 0.6 mM, 0.7 mM, and 3.7 mM, respectively, thus close to the same range of the successful displacers. It can be speculated that this results from differences in the affinities experienced in solution and in the crystal, or that kinetic effects of the diffusion process in the crystal channels are responsible for this deviation of behavior in some fragments.

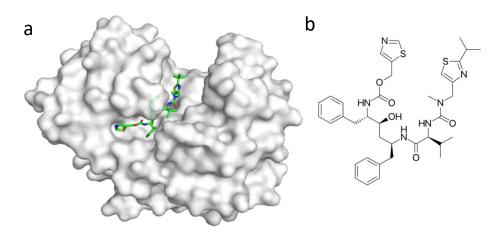


Figure 4.6: a) Binding position of the potent inhibitor ritonavir (PDB: 3PRS) in endothiapepsin and b) chemical structure. Ritonavir is used in competitive binding experiments to check for active-site binders. It has a K_I of 15nM with endothiapepsin.

Table 4.4: Chemical structures of catalytic dyad fragments that were found to bind in the competitive binding experiments in the presence of ritonavir.^a

Fragment	Structure	Binding Affinity (K _d mM)	Observed Electron Density Ritonavir or Fragment?
109	H ₂ N-NH N	1	Ritonavir
216	NH NH ₂	0.6	Fragment and Ritonavir
218	N NH ₂	0.1	Fragment and Ritonavir
236	N O N O N O N O N O N O N O N O N O N O	0.6	Ritonavir
267	TZ O	0.7	Ritonavir
268		3.7	Ritonavir

290	H ₂ N S CI	0.1	Fragment
306	H ₂ N S	0.1	Fragment
63	N O O HN	0.5	Fragment

(a) Their binding affinities determined via displacement ITC are also listed, in addition to the resulting electron density observed after soaking the endothiapepsin crystals into soaking solutions containing both ritonavir and the fragment of interest.

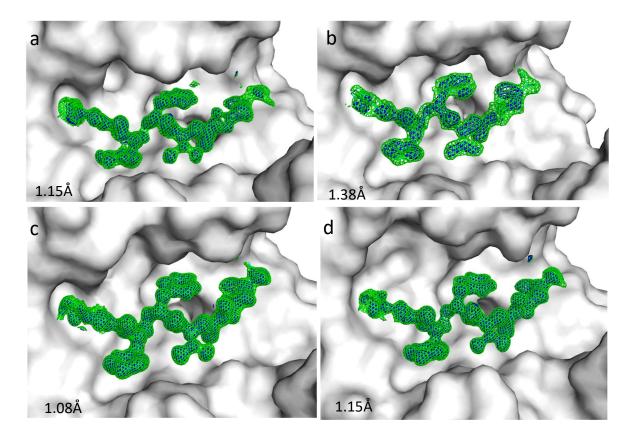


Figure 4.7: Electron densities (2|Fo|-|Fc| density contoured at a σ level of 1 in green, |Fo|-|Fc| density contoured at a σ level of 3 in blue) of ritonavir seen in crystals soaked in a) F109, b) F236, c) F267, and d) F268. No electron densities for the fragments could be observed, showing that these fragments could not compete with the ritonavir in binding to the active site.

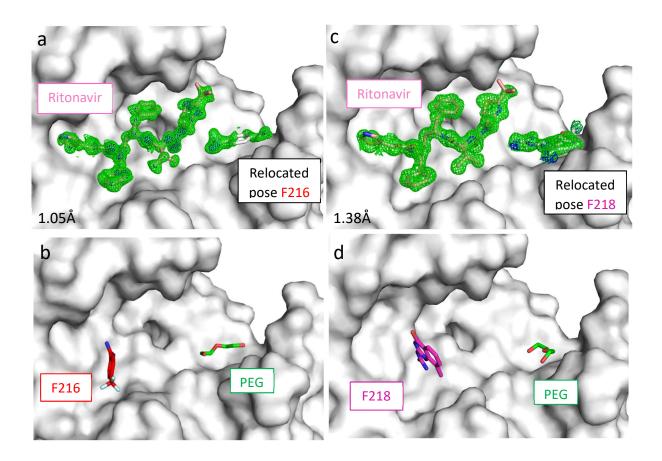


Figure 4.8: Electron densities (2|Fo|-|Fc| density contoured at a σ level of 1 in green, |Fo|-|Fc| density contoured at a σ level of 3 in blue) seen in structures of a) F216 where both fragment and ritonavir are binding, and the closest interacting distance is 3.16 Å between one molecule of the *para*-trifluoromethyl benzamidine and the nitrogen of the thiazole in ritonavir. The previously reported binding pose of F216 (PDB: 4YCT) can be seen in b) where F216 is in red, binding at the position of ritonavir. Additionally, the PEG molecule, also in red, in the original structure can also be seen binding in the position as found for the relocated fragment. The structure c) of F218 also shows simultaneous binding of reporter ligand and the fragment, and the previously reported binding site of F218 (PDB: 4YCY) shown in d) in magenta binding in the same position as ritonavir in the other complex, in addition to the PEG molecule in magenta seen in this structure to bind in the position where the

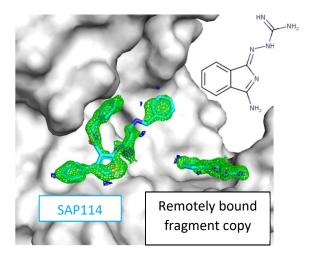


Figure 4.9: Crystal structure of endothiapepsin soaked in a solution of an active site fragment (chemical structure shown) and potent inhibitor SAP114. The binding of SAP114 in the active site displaces the fragment, but a second copy of the fragment that binds remotely can still be seen to bind. This remote binding site is where a PEG molecule is seen in many published endothiapepsin structures obtained by soaking. Residues 79 – 83, and 113 – 118 have been omitted for clarity.

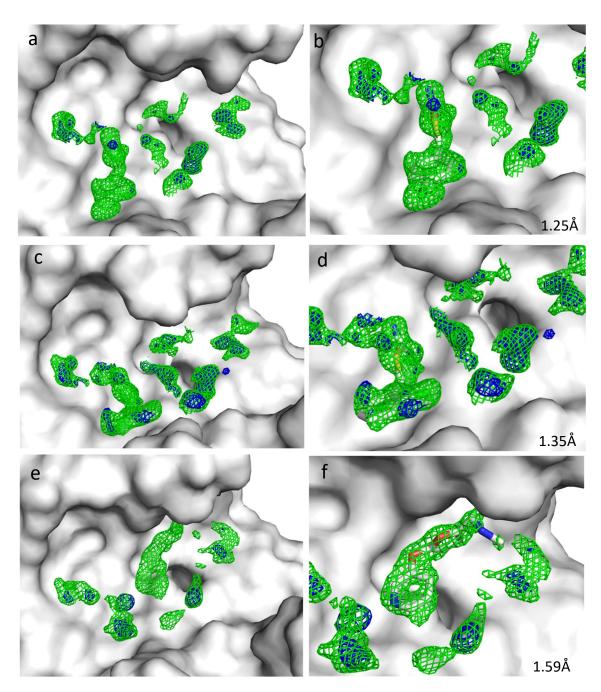


Figure 4.10: Electron densities (2|Fo|-|Fc|) density contoured at a σ level of 1 in green, |Fo|-|Fc| density contoured at a σ level of 3 in blue) observed in the structures of a), b) F290 occupancy 77%, c), d) F306 occupancy 85%, and e), f) F063 occupancy 81%. There is dispersed density seen that is not resolved well enough to fit ritonavir and it overlaps with the electron density of the fragments. The occupancies are assigned after one round of refinement.

4.2.4 Comparison of Light Water STD NMR and Heavy Water STD NMR

The second STD NMR experiments were run in non-deuterated water (H_2O). It is not uncommon to run STD NMR in non-deuterated solvents ¹⁹¹ considering costs of deuterated water, and the fact that shorter interscan delays can be used due to faster T_1 relaxation in light water. In deuterated water, much longer interscan delays are required to allow for full relaxation of small molecules. Otherwise artifacts may arise from ligands that are re-excited before being fully relaxed.

To rationalize the differences between hits seen by STD NMR experiments and X-ray crystallography, we considered the role of protons in non-deuterated water. The concept of STD NMR is based on irradiating the protein, NMR transitions of which get saturated, this saturation is transferred from the protein to its bound ligand. The effect of the transferred saturation is recorded by directly observing NMR signals of the (large) fraction of free, unbound ligand, in typical applications of weak (micro to millimolar) interactions. The experiment is performed by applying a low-power irradiation in the region of a ¹H NMR spectra that covers protein signals but no ligand signals. Through spin diffusion, the magnetization spreads quickly throughout the protein and saturates all the protein ¹H NMR signals. If a ligand is temporarily bound to the protein, its ¹H NMR signals also become saturated and upon its dissociation this leads to a decrease in the intensity of ¹H NMR signals measured from the sample pool of screened free ligand in solution. Another experiment is then recorded where the irradiation pulse is directed to a spectral region where no protein and ligand atoms exhibit signals, to avoid saturation of protein and saturation transfer to the ligand. The first and second spectra are subtracted to yield the difference spectrum STD (Saturation Transfer Difference).¹⁹²

When magnetization is distributed through the protein by spin diffusion, it does not only saturate the ¹H NMR signals of bound ligands, but also those of water molecules associated with the protein. Once exchanged with bulk water molecules these waters can then transfer magnetization to the free ligand molecules in solution, which can affect the intensity of the ¹H NMR signals measured from the pool of free ligands. In consequence, not only the bound ligand is irradiated via the protein, but also to some degree the unbound ligand is irradiated in solution.⁷² The magnetized bound ligand and magnetized free ligand molecules will have opposite signs in the ¹H NMR signals. There will be a positive sign from protein bound ligand, and negative sign from ligand in solution that receives magnetization from the water. These signals will either cause attenuation of the NMR signal of interest making it difficult to observe it in the ¹H NMR spectra, or even cancel each other out leading to false negatives. We can also observe false positives from the same phenomena, only in that case there is only a signal coming from the free ligand molecule in solution purely obtaining magnetization from protein released water. This will be explained in more detail in the discussion section.

By using heavy water, this should theoretically reduce the aforementioned phenomena, as only a negligible fraction of protein spins will be in contact with protonated water molecules. To test this, we selected eight fragments that were hits in crystallography, WaterLOGSY, and SPR but gave either ambiguous, weak, or no signals in STD NMR and screened them in light (90% $H_2O/10\%$ D_2O , referred to here as third NMR screen) and heavy (100% D_2O , referred to here as fourth STD NMR screen) water NMR. We also selected a fragment observed to bind in all assays, and therefore remained unaffected by light water, as a control. Only one of these fragments, F039, was screened in the initial STD NMR where CD_3COOD in D_2O was used and it showed up as a hit, however the other fragments were excluded from the screening due to inefficient solubility, impurity, or being prone to aggregation. Structures of these fragments can be seen in Table 4.5.

Table 4.5: Fragments selected for STD NMR in heavy water and light water comparison.^a

Fragment	Structure	Hit in Crystallography	Hit in SPR	Hit in Water LOGSY	Hit in third (90% H ₂ O/10% D ₂ O) STD NMR	Hit in fourth (D₂O) STD NMR
39	The Co	Yes	Yes	Yes	Ambiguous	Yes
54	H N N N N N N N N N N N N N N N N N N N	Yes	Yes	Yes	Weak	Yes
63	N O O HN	Yes	Yes	Yes	Weak	Yes
66	N NH ₂	Yes	Yes	Yes	Ambiguous	Yes
162	S H N N	Yes	Yes	Yes	Weak	Yes
189	NH ₂	Yes	Yes	Yes	No	Yes
206	F NH ₂	Yes	No	Yes	No	No
286	NH ₂	Yes	No	Yes	Ambiguous	Yes
41		Yes	Yes	Yes	Yes	Yes

a) All fragments were detected as hits in x-ray crystallography and SPR, with the exception of F206 and F286 which are not detected in SPR. All fragments bind when screened in heavy water STD NMR.

The collected 1H NMR spectra for the fragments is shown below in addition to the STDD (Saturation Transfer Double Difference) 193 spectra of the fragments in 90% $H_2O/10\%$ D_2O and 100% D_2O buffers. The buffer used for the 90% $H_2O/10\%$ D_2O STD NMR was 3 mM CH_3COOH at a pH of 4.6, whereas the buffer used for the 100% D_2O STD NMR was 3 mM CD_3COOD at a pD of 5. The fragments were tested at a concentration of 310 μ M and endothiapepsin was added at a concentration of 9.6 μ M.

The STDD NMR spectra of F039 in CH₃COOH (Figure 4.11 red) shows only one fragment signal at 6.9 ppm, whereas the spectra in CD₃COOD (Figure 4.11 blue) shows two fragment signals at both 6.9 ppm and 7 ppm.

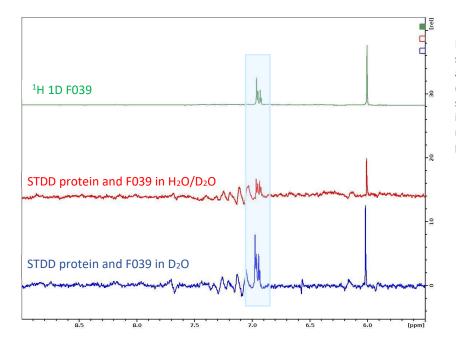


Figure 4.11: ¹H 1D NMR reference spectra of F039 and endothiapepsin and STDD NMR spectra in CH₃COOH (red) and CD₃COOD (blue). More signals can be seen for fragment 039 in the deuterated solvent (spectra region of interest is highlighted in pale blue).

Fragment 54

The STDD NMR spectra of F054 in CH₃COOH (Figure 4.12 red) shows weak signals for the fragment. However, the spectra in CD₃COOD (Figure 4.12 blue) show three strong signals of the fragment in the aromatic region.

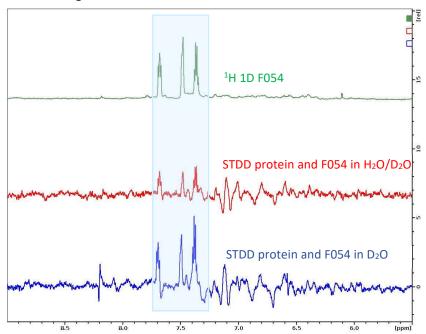


Figure 4.12: ¹H 1D NMR reference spectra of F054 and endothiapepsin and STDD NMR spectra in CH₃COOH (red) and CD₃COOD (blue). All three signals of the F054 in the aromatic region can be much more clearly seen in the deuterated solvent (spectra region of interest is highlighted in pale blue).

The STDD NMR spectra of F063 in CH₃COOH (Figure 4.13 red) shows fragment binding, however the fragment signals have a low intensity. The spectra for the same fragment in CD₃COOD (Figure 4.13 blue) shows signals with higher intensity of the fragment binding.

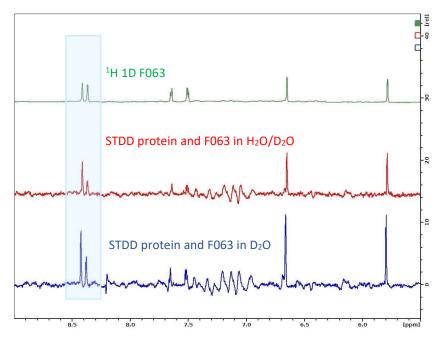


Figure 4.13: ¹H 1D NMR reference spectra of F063 and endothiapepsin and STDD NMR spectra in CH₃COOH (red) and CD₃COOD (blue). Signals for F063 can be detected in both the nondeuterated and deuterated solvent (spectra region of interest is highlighted in pale blue) however the signals are more prominent in the spectra with the deuterated solvent.

Fragment 66

The STDD NMR spectra of F066 in CH_3COOH (Figure 4.14 red) shows some fragment binding, however not all the fragment signals in the region 8 to 7.7 ppm can be seen to bind. The spectra for the same fragment in CD_3COOD (Figure 4.14 blue) shows higher intensity for the fragment as well as all the signals in the region 8 to 7.7 ppm specifically at 8 ppm.

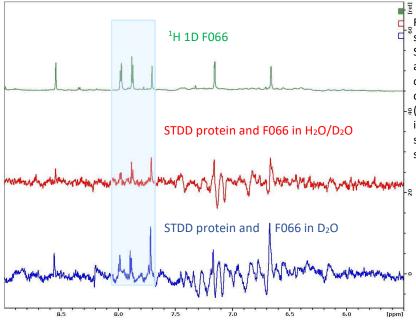


Figure 4.14: ¹H 1D NMR reference spectra of F066 and endothiapepsin and STDD NMR spectra in CH₃COOH (red) and CD₃COOD (blue). Fragment signals can be seen for F066 in both the non-deuterated and deuterated solvent (spectra region of interest is highlighted in pale blue) however all the fragment signals can only be observed in the STDD spectra in deuterated solvent.

The STDD NMR spectra of F162 in CH₃COOH (Figure 4.15 red) shows fragment binding, however the intensity of the fragment signal at 8.7 ppm is very weak, making it unclear whether it is a signal or noise. The intensity of the bound fragment in the STDD NMR in CD₃COOD (Figure 4.15 blue) is stronger.

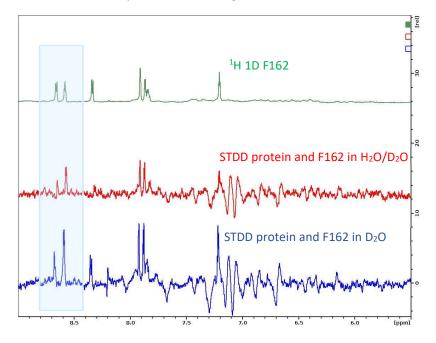


Figure 4.15: ¹H 1D NMR reference spectra of F162 and endothiapepsin and STDD NMR spectra in CH₃COOH (red) and CD₃COOD (blue). Both fragment signals are clearly observed in the spectra of the deuterated solvent, whereas the non-deuterated solvent spectra show one clear fragment peak, and a weaker signal difficult to observe within the noise of the spectra.

Fragment 189

The STDD NMR spectra of F189 in CH_3COOH (Figure 4.16 red) shows no fragment peaks, indicating that the fragment is a non-binder. The spectra for the same fragment in CD_3COOD (Figure 4.16 blue) shows fragment binding in the region between 8.6 and 7.7 ppm. This could be considered a false negative in H_2O STD NMR.

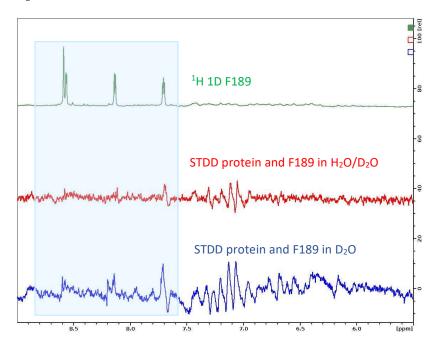


Figure 4.16: ¹H 1D NMR reference spectra of F189 and endothiapepsin and STDD NMR spectra in CH₃COOH (red) and CD₃COOD (blue). Fragment binding is only seen in deuterated solvent.

The STDD NMR spectra of F206 in CH₃COOH (Figure 4.17 red) shows no fragment signals, indicating that the fragment is a non-binder. The spectra for the same fragment in CD₃COOD (Figure 4.17 blue) also shows no fragment binding. It is worth mentioning that F206 was also not detected as a hit in SPR.

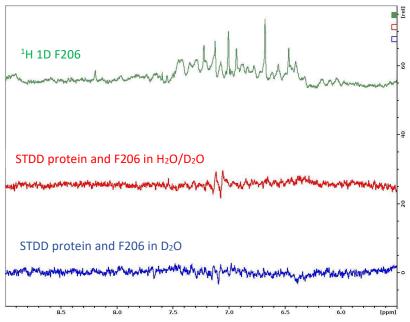


Figure 4.17: ¹H 1D NMR reference spectra of F206 and endothiapepsin and STDD NMR spectra in CH₃COOH (red) and CD₃COOD (blue). Fragment binding cannot be seen in neither deuterated nor non-deuterated solvent.

Fragment 286

The STDD NMR spectra of F286 in CH_3COOH (Figure 4.18 red) shows no fragment signals, indicating that the fragment is a non-binder. The spectra for the same fragment in CD_3COOD (Figure 4.18 blue) shows a fragment signal at 7 ppm, indicating binding. This is another clear case of a false negative in H_2O STD NMR.

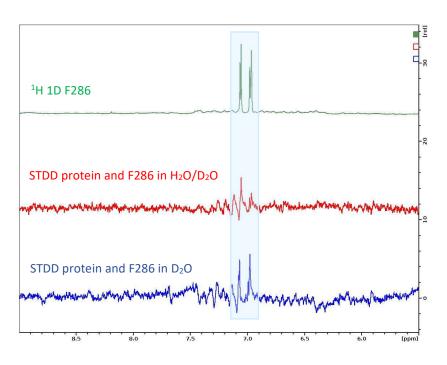


Figure 4.18: ¹H 1D NMR reference spectra of F286 and endothiapepsin and STDD NMR spectra in CH₃COOH (red) and CD₃COOD (blue). Fragment binding can only be seen in non-deuterated solvent.

4.2.5 Discussion of Comparative Analysis

In addition to STD NMR, the 71 crystallographically discovered fragment hits were also screened by WaterLOGSY, which showed the highest correlation with the X-ray hits with a 70% overlap (Figure 4.19). The sensitivity of WaterLOGSY has been reported to be higher than that of STD NMR¹⁹⁴, which our studies also confirm. We can also attribute this higher sensitivity to the concept of WaterLOGSY, which eliminates false positives and negatives arising from non-deuterated solvents as previously discussed.

In STD NMR, we have a negative NOE (*nuclear Overhauser effect*) between protein and bound ligand since the protein-fragment complex is big and tumbles slowly. On the other hand, we have a positive NOE between the water and unbound ligand in solution, since the motion here is much faster. The non-deuterated water in the solution that is associated with the protein has received polarization when it came in contact with the protein, where it received a negative NOE. When this water is released again into the bulk phase, it receives the positive NOE that dominates the magnetization transfer between water and unbound ligand in solution. These two processes are responsible for cross relaxation in opposite directions and accordingly they may cancel each other out. When the experiments are run in deuterated water, the cross relaxation between water and unbound ligand is negligible since most protons in the solution are replaced by deuterium, and thus the magnetization transfer between the protein and bound ligand is the dominant effect observed (positive NOE) and this leads to a stronger STD effect being observed, and thus more intense ¹H NMR signals can be seen in the spectra.

The disadvantages of STD NMR have been described elsewhere, and they include false positives resulting from direct irradiation of ligand as discussed in the previous section, as well as false negatives resulting from incomplete saturation of protein.⁷² The variables leading to the difference in hit rates between the initial STD NMR and the second STD NMR screen have been discussed in detail, which included contributions of using a buffer with high ionic strength, as well as using non-deuterated solvents. Additionally, the fact that the initial STD NMR was run using cocktails of two fragments is also responsible that some hits were missed.¹⁷⁶ Efficiency of screening cocktails in NMR will be discussed in section 4.5.

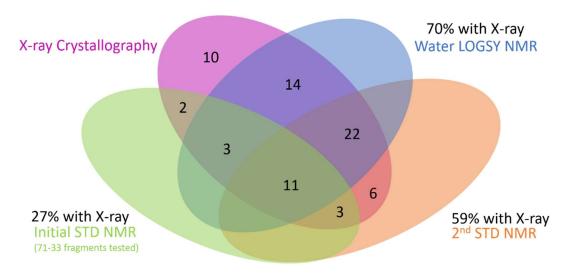


Figure 4.19: Venn diagram showing a comparison of fragment hit rates of water LOGSY NMR in cyan, the initial STD NMR in green, and the second STD NMR screen in orange relative to the 71 fragment hits detected by crystallography in purple.

4.3 Fragment Screening Using Surface Plasmon Resonance

In our previous comparative analysis study¹⁷⁵, Surface Plasmon Resonance (SPR) as a prominent biophysical method in fragment screening was not included. The reason for this was the difficulty to immobilize endothiapepsin to the dextran chip. The theoretical pl of endothiapepsin is 4.14 (calculated using protparam¹⁸⁴). To immobilize a protein successfully, the pH at which this process takes place should be at least one order of magnitude below the pl of the protein (so it can adopt an overall positive charge) but above a pH of 3.5 as the carboxylates of dextran on the chip surface have to remain negatively charged. For successful immobilization this opens a very small pH window - between 3.5 and 4.1 - to work with and likely the required charged states are impossible to achieve. To overcome this challenge, we decided to use a reversed inhibition assay – the inhibition in solution assay (ISA) rather than the classic direct binding assay (DBA). Experimental details are discussed in Materials and Methods.

4.3.1 ISA Methodology

In ISA, a target definition compound (TDC) which is a known active-site binder of the target protein is immobilized onto the sensor chip instead of immobilizing the target protein as is done in DBA (Figure 4.20a). In our experiments we used a potent inhibitor of endothiapepsin, the peptide H-142 ¹⁹⁵ (sequence: PHPFHL_RVIHK, where R depicts the reduced isostere of the scissile peptide bond between residue Leu10 and Val11 in human angiotensinogen) (Figure 4.20b).¹⁷⁴ The target (endothiapepsin) and the fragments are then incubated together in a 96-well plate and the mixture is injected over the TDC immobilized surface. The initial binding rate is used to determine the percentage of free protein in solution, in other words the unbound endothiapepsin available in the mixture to bind to the immobilized H-142 peptide, which will change depending on the concentration of the competing fragment. When the fragment is binding to the endothiapepsin in the mixture, it competes with the immobilized TDC for the endothiapepsin active site and the recorded 'free protein' signal is decreased. This is then recorded as a reduction of the percentage of bound endothiapepsin to the H-142 peptide and hence a higher percentage of active-site binding of the fragment to endothiapepsin. This allows a qualitative affinity ranking of the fragments since the experiments are all conducted at similar concentrations.¹⁹⁶

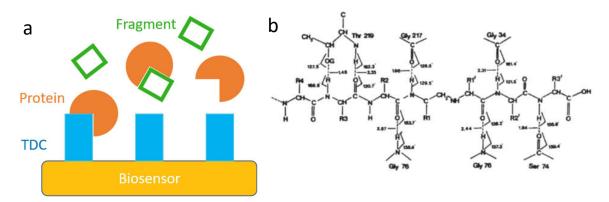


Figure 4.20: a) Concept of ISA derived from Geschwindner, S. et al. (2013) where a TDC is immobilized onto the surface of a sensor chip, and a solution of fragment and protein is passed over the surface so that the fragment and TDC are in competition for the binding site of the protein. b) Structure of peptide H-142 potent endothiapepsin inhibitor with reported K_i of 160 nM 4 . Image taken from Geschwindner, S. et al. J. Med. Chem. (2013).

4.3.2 Fragment Hits

The entire 361-fragment library was screened, resulting in a hit rate of 34%. Hits were considered as such giving a percentage inhibition above 3%. This value was determined based on the calculated standard deviation of the percentage inhibition measured for the buffer control. The overlapping hits between crystallography and SPR were 11% (38 fragments), the highest match in comparison with the other biophysical methods.¹⁷⁵

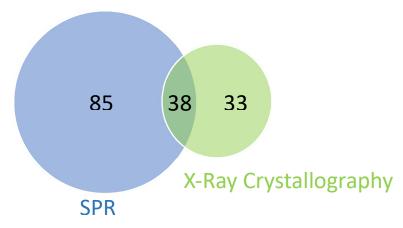


Figure 4.21: Venn Diagram depicting the number of fragment hits overlapping between the SPR and X-ray crystallography screens, as well as the additional fragment hits of each screen independently. In the SPR screen in blue, a total of 123 fragments were detected, 38 of which were also detected in X-ray crystallography in green. An additional 33 fragments were detected as hits by X-ray crystallography but not by SPR.

For the SPR fragment hits that gave a percentage inhibition of at least 50%, the IC_{50} values were determined by running a dose-response curve and plotting the logarithmic fragment concentration against the percentage of inhibition (Figure 4.22). The strongest fragment is F177 which was previously reported to also be the most potent fragment in other five out of six assays.¹⁹⁷ The reason behind this has been described in detail in literature¹⁹⁷ where it was proven that the reactive nature of this fragment causes its oxidization and oligomerization into another compound, making it essentially a false positive binder. The Table 4.6 below lists the IC_{50} values and the previously calculated K_d values from ITC^{175} and IC_{50} from the reporter displacement assay (RDA)⁷⁰ where available. To rule out non-specific binding of these fragment hits to the TDC, they were injected over the H-142 immobilized surface without addition of endothiapepsin and the resulting percentage binding analyzed. The values were comparable to those from the buffer controls, so we could confirm that the percentage inhibition calculated in the presence of endothiapepsin is due to specific binding to the protein.

Table 4.6: Overview of the most potent fragment hits by SPR and their calculated $IC_{50}s$ from SPR, as well as previously calculated K_d from ITC, and $IC_{50}s$ from the RDA.

Fragment	IC ₅₀ μM SPR	K _d μM	IC ₅₀ μM RDA
177	135		3
248	526		
306	597	100	40
79	628		
284	781		260
290	834	100	40
255	846	200	1,160
236	915	600	
149	974		150
167	1136		

218	1250	100	
47	1304		
64	1581		
337	1551		
162	1614		
178	1672		230
148	2088		> 1,300
159	2294		
65	2544		
334	2958		
342	3388		

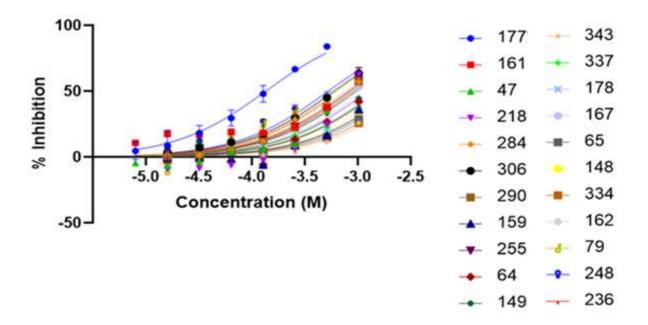


Figure 4.22: Plot of log concentration of fragments against their % inhibition. The resulting IC₅₀s were calculated for fragments that inhibit endothiapepsin in the ISA by at least 50%.

4.4 Comparative Analysis of Fragment Binding using X-ray Crystallography, SPR, and STD NMR

4.4.1 Discussion

SPR was the method that showed the highest overlapping hit rate with crystallography of 11%, even higher than the RDA from our previous study that had a correlation of 8% with the crystallographic hits. The SPR detected 84 additional fragments as hits that were not detected by X-ray crystallography, giving SPR a total hit rate of 34% which is quite high. This hit rate, however, does not account for unspecific binding that may have occurred between the TDC on the chip surface and the fragments themselves, which would also change the recorded signal. Also, a more stringent cut off set to a higher threshold than 3% inhibition would reduce the number of hits. For example, setting the threshold to 10% inhibition would lead to 61 hits, cutting the number of hits by half. The 21 most potent hits, however, were tested for unspecific binding, but their specific binding could be confirmed.

If we analyze the 33 crystallographic fragment hits that were not detected as hits by SPR, we find that they all share low binding affinities in common (Table 4.7). There are two fragments that had measurable binding affinities below 1.5 mM but were still not detected as hits in SPR. In fact, one of them, F051 with binding affinity of 1mM, was not detected by any of the STD NMR screens either due to reasons discussed previously in section 4.2.

Table 4.7: Chemical structures of crystallographic fragment hits that were not detected as hits in SPR.a

Fragment	Chemical Structure	Binding Affinity (Kd in mM)	Fragment	Chemical Structure	Binding Affinity (K _d in mM)
4	N NH ₂	>10	206	F NH ₂	1.9
14	Br CI s N N H	5	211	H ₂ N	1.9
31	H ₂ N O	5.2	227	О	NA
35	H_2N	3.2	231	N OH	2.5

48	CI N N N N N N N N N N N N N N N N N N N	3.1	240	2	>10
52	H ₂ N Br	3.3	266	O O O H	>10
56	CI OH	NA	203	N HN S S H ₂ N	NA
58	N-N NH ₂	8.8	268	O NH	3.7
73	HO NH	>10	274	F NH ₂	3.8
78	H_2N	>10	278	OH N N	>10
81	N NH ₂	>10	286	NH ₂	3.3
103	F O NH2	4.3	305	O-B O-B	NA
125	NH NH ₂	1.7	311	S N O	NA

131	O N F F	2.8	330	O N N CI	3.4
205	SH NH2	1.5	51	TZ ZI	1.0
224	HN N	0.4	207	HN N	4.4
164	Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z	n.d			

a) Their binding affinities determined by ITC are all above 1.5 mM.

There are 16 fragments in total that were measured and detected as hits in crystallography, the first NMR screen, and SPR. Their chemical structures, binding affinities, and binding positions which deviate are shown below (Figure 4.23). They are color-coded based on their binding positions. The diverse nature of these fragments suggests that there are no cardinal rules that we can use to claim a certain moiety or property as transferable across different biophysical methods. We can, however, keep in mind the aspects discussed in this chapter to keep false positives and false negatives at a minimum.

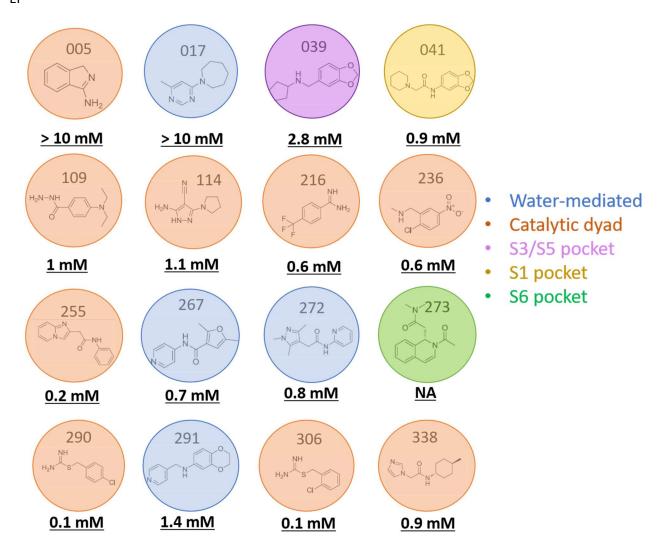


Figure 4.23: Chemical structures, binding affinities, and color-coded binding positions of all 16 fragments that were detected as hits in crystallography, the first STD NMR, and SPR. The diverse nature of the fragments makes it difficult to attribute a single characteristic for making these fragment hits recursively detected by the different biophysical methods

Highest Overlapping Methods

As mentioned earlier, the overlap between SPR and X-ray hits was the highest reported, at 11%. The second highest overlap between the methods was seen between SPR and RDA hits, at 9%, followed by an 8% overlap between RDA and X-ray hits. These three methods are also from the ones that detected the highest number of hits, where SPR detected 123 fragments, X-ray detected 71, and RDA detected 50. The first STD NMR detected 58 hits, which brings the overlap of the first STD NMR and SPR at 7%. Finally, the first STD NMR and X-ray had an overlap of 6%.

4.5 Comparison of Cocktail Screening in X-Ray Crystallography vs NMR

4.5.1 Definition of Cocktail Screening

Cocktail screening, whether in the realm of crystallography or NMR, was designed as a technique to allow accelerated examination of larger libraries in a reduced time and to enable rapid hit identification. There are several aspects that must be considered in designing a cocktail experiment in crystallography, among these is the structural shape diversity of fragments and the number of fragments in every cocktail set. However, it has been proven in previous fragment screening studies that cocktail soaking can miss a great deal of hits due to blurred and fuzzy electron density in the final structure, as well as competitive behaviors between the fragments; both which could be avoided by single soak experiments. The structure is the structure of t

While previous studies have analyzed differences in experimental results between cocktail soaking and single soaking crystallography, there is still a need to study differences between fragment detection and cocktailing in crystallography in comparison to fragment cocktailing in NMR.

4.5.2 Results & Discussion

In this study, we performed a retrospective STD NMR screen using the same sets of fragment cocktails that were used before in a cocktail fragment screening in crystallography to determine whether STD NMR could pick up hits in the fragment cocktails with comparable or improved efficiency to crystallography.

All the cocktails tested are composed of fragments previously detected as hits in crystallography with the exception of F308, F040, and F142. Additionally, all fragments were detected as hits when screened separately by the second STD NMR, with the exception of, F236, F103, F058, F207, F227, F274, F051, F236. Additionally, F040 and F142 were excluded from the single fragment screens in the second STD NMR but were not found to bind in the first STD NMR either.

In the figures below, the ${}^1\text{H}$ 1D NMR spectra of the cocktails were compared to the ${}^1\text{H}$ 1D NMR spectra of the single fragments, to ensure that the peaks in the cocktail set correspond to the single fragments. The STD NMR spectra of the endothiapepsin protein with the cocktail sets were then compared to the ${}^1\text{H}$ 1D NMR spectra of the cocktail set alone and the ${}^1\text{H}$ 1D NMR spectra of the single fragments thus identifying the bound fragments in the STD NMR. An overview of the results is listed in Table 4.8, in addition to the chemical formulas of the fragments. For the electron densities of the crystal structures in the figures below, the 2m|Fo|-|Fc| density contoured at a σ level of 1 is colored in blue while the m|Fo|-|Fc| density contoured at a σ level of 3 is colored in green. For all of the cocktail sets processing with the Gaussian multiplication (Gaussian transformation)¹⁹⁹, which is done to improve resolution at the expense of the signal-to-noise ratio, lead to a spectrum with a noisy baseline and the binding signals could not be clearly seen. Therefore, the spectra is displayed only after FT processing. The only exception to this is cocktail set D. Additional information regarding fragment binding occupancies in the crystal structures can be found in Table 4.8.

Cocktail A - F063, F267, F291

The STD NMR spectra (blue) of endothiapepsin and the cocktail set shows binding for F063 (green), F267 (violet), and F291 (yellow). In the crystal structure (PDB: 5MB0) all three fragments could be clearly assigned to the electron densities. In the individually soaked crystal structures, F063 (PDB 4Y57) binds in a different position than that where it binds as part of the cocktail set (Figure 4.38). F291 binds as two copies when soaked individually (PDB 4Y45) but when part of a cocktail set, one of these copies binds with an RMSD of 0.057 to its original binding position and the second copy binds stacked to F267, partially assigned with benzodioxanyl in the S1 pocket instead of pyridine. In the cocktail set, the new binding position of F063 would clash with the binding position of the second copy of F291, thus they seem to be in competition and F063 binds instead. F267 binds in the same position when individually soaked and when part of a cocktail set.

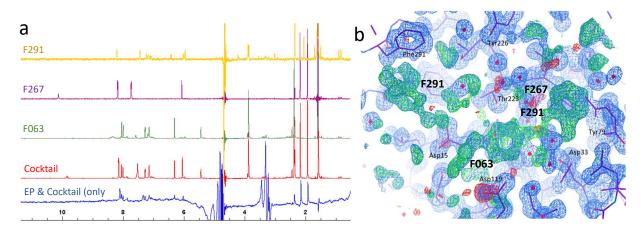


Figure 4.24: a) STD NMR spectra of endothiapepsin and cocktail set shows binding signals for all three fragments in the cocktail set. b) The crystal structure shows clear electron densities that can be assigned to F063, F267, and F291.

Cocktail B - F224 and F236

The STD NMR spectra (blue) of endothiapepsin and the cocktail set shows binding for both F224 (green) and F236 (violet). In the crystal structure (PDB: 5MB7), only F236 could be assigned to the electron density as F224 seems to be displaced by F236.

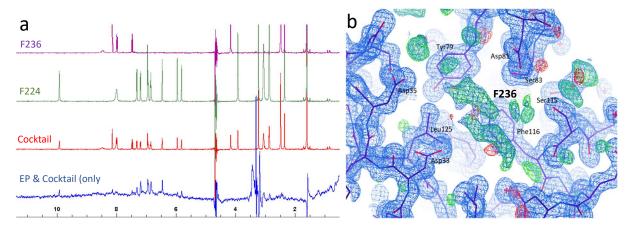


Figure 4.25: a) STD NMR spectra of endothiapepsin and cocktail set shows binding signals for both fragments in the cocktail set. b) The x-ray structure shows the electron density for F236 only and no visible electron density for F224.

Cocktail C - F103 and F171

The STD NMR spectra (blue) of endothiapepsin and the cocktail set shows no binding for any of the fragments. It is worth mentioning that F103 also did not show binding in the STD NMR when tested individually with endothiapepsin. In the crystal structure (PDB: 5MB5) only F171 could be assigned to the electron density where it binds similarly to its original binding position in the individual fragment soaking (PDB 4Y3X).

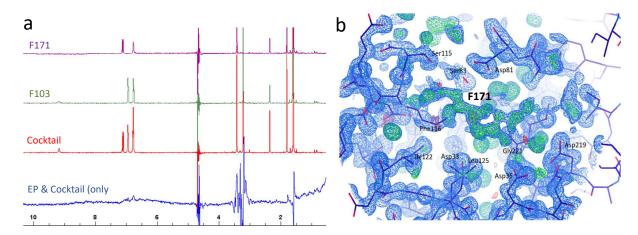


Figure 4.26: a) STD NMR spectra of endothiapepsin and cocktail set shows no clear binding for any of the fragments. b) The crystal structure shows electron density only for F171, and no visible electron density for F103.

Cocktail D – F308 and F333

The STD NMR spectra (blue) of endothiapepsin and the cocktail set shows binding for only F333 (violet). F308 was also not a crystallographic hit. In the crystal structure (PDB: 5MB6) F333 could not be assigned to the electron density.

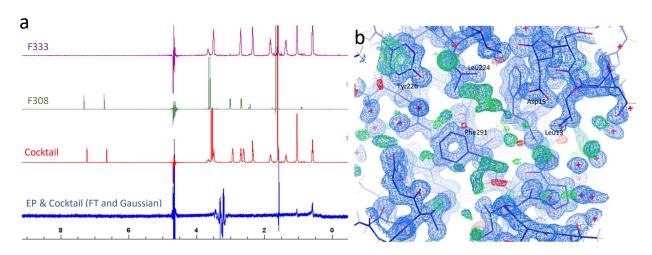


Figure 4.27: a) STD NMR spectra of endothiapepsin and cocktail set shows binding only for F333. b) The crystal structure shows no clear electron density for either fragment to be built in.

Cocktail E - F052, F058, and F063

The STD NMR spectra (blue) of endothiapepsin and the cocktail set show binding for both F063 (yellow) and F052 (green). However, there is no clear binding for F058 (violet). F058 did not show binding in the STD NMR when tested individually with endothiapepsin. In the crystal structure only F063 could be clearly assigned to the electron density, as it seems to win the competition with F052 and F058 for the same binding position.

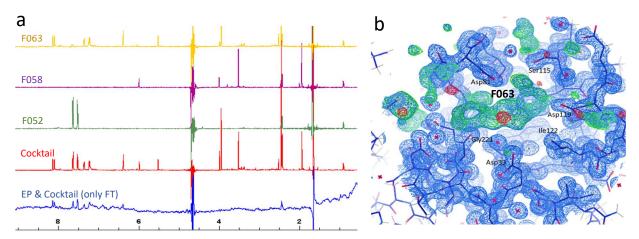


Figure 4.28: a) STD NMR spectra of endothiapepsin and cocktail set shows binding for F063 and F052, but no sign of binding for F058. b) The crystal structure shows the electron density for F063 but no clear electron density for F052 and F058 to be built in.

Cocktail F - F207 and F261

The STD NMR spectra (blue) of endothiapepsin and the cocktail set show binding for F261 (violet) and a signal at 2.5 ppm which may correspond to either F261 or F207 (green). The spectrum of the cocktail set however, does not correspond to the ¹H 1D spectra of F207 and may reflect insolubility of F207 in the cocktail set, and additional signals around 3 ppm may reflect an impurity in the cocktail set. In the crystal structure F207 could be clearly assigned to the electron density, as well as fragments of F261. One copy of F261 binds in its original binding position in the Asp81 flap region as seen in the structure of the single soaked fragment (PDB: 4Y5B) but the electron density is not sufficient to build the entire molecule. F207 and F261 are in competition as F207 binds in the same position as F261 when soaked as a single fragment.

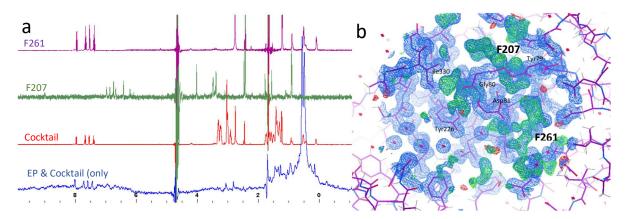


Figure 4.29: a) STD NMR spectra of endothiapepsin and cocktail set shows binding for F261 and a peak that may correspond to either F261 or F207. b) The crystal structure showing the electron densities for F207 and fragments of F261.

Cocktail G - F216 and F338

The STD NMR spectra (blue) of endothiapepsin and the cocktail set show binding for both F216 (green) and F338 (violet). In the crystal structure, F216 could be assigned to the electron density but only a fragment of F338, indicating that the F338 may have been in competition with F216 which hinders F338 interacting fully with its binding position

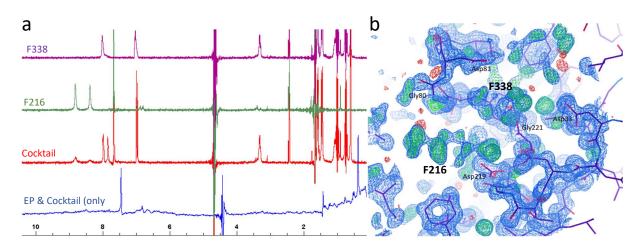


Figure 4.30: a) STD NMR spectra of endothiapepsin and cocktail set shows binding for both F216 and F338. b) The crystal structure shows a clear electron density for F216 and partial electron density for F338.

Cocktail H – F218 and F224

The STD NMR spectra (blue) of endothiapepsin and the cocktail set show binding for only F218 (green), albeit only a single peak at 1.5 ppm is clearly from F218. In the crystal structure only F218 could be clearly assigned to the electron densities. Several copies bind in different places, one of which pushes residues Asp81 to Tyr79 out of their original positions (PDB: 4Y5L). A total of four copies of F218 could be built into the electron densities, three of which stack against one another.

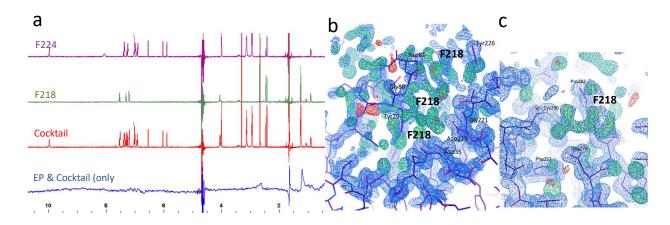


Figure 4.31: a) STD NMR spectra of endothiapepsin and cocktail set shows binding only for F218. b) The crystal structure shows the electron density for stacked copies of F218 and c) another copy of F218 in a remote binding location next to the S6 pocket.

Cocktail I - F227 and F274

The STD NMR spectra (blue) of endothiapepsin and the cocktail set show no binding for either fragments. It is worth mentioning that both fragments did not show binding in the STD NMR when tested individually with endothiapepsin. In the crystal structure, both F227 and F274 could be clearly assigned to the electron density.

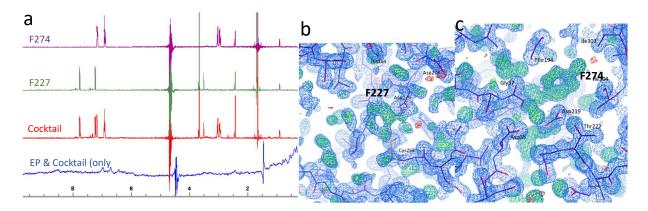


Figure 4.32: a) STD NMR spectra of endothiapepsin and cocktail set shows no binding for either fragment. b) The x-ray structure shows clear electron densities for F227 and c) F274 which bind to a different part of the protein crystal next to the S1' S2' pocket.

Cocktail J - F041 and F051

The STD NMR spectra (blue) of endothiapepsin and the cocktail set show only weak binding for F051 (violet). It is worth mentioning that both fragments did not show binding in the STD NMR when tested individually with endothiapepsin. In the crystal structure both fragments could be clearly assigned to the electron density.

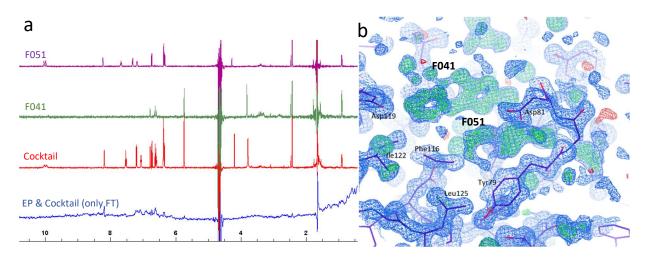


Figure 4.33: a) STD NMR spectra of endothiapepsin and cocktail set shows only binding for F051. b) The crystal structure shows clear electron densities for F041 and F051.

Cocktail K - F162, F236, and F338

The STD NMR spectra (blue) of endothiapepsin and the cocktail set show only binding for F338 (yellow) where a signal can be seen at 0.5 ppm (orange spectrum). It is worth mentioning that F236 (violet) did not show binding in the STD NMR when tested individually with endothiapepsin. In the crystal structure none of the fragments could be clearly assigned to the weak electron density.

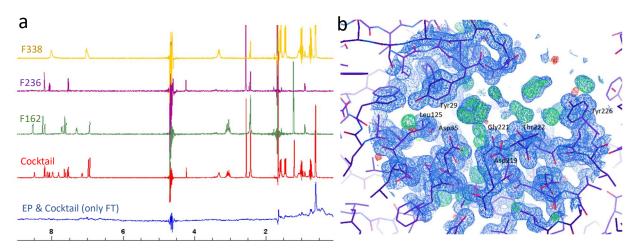


Figure 4.34: a) STD NMR spectra of endothiapepsin and cocktail set shows binding only for F338. b) The crystal structure shows dispersed electron densities unclear for assigning any of the three fragments.

Cocktail L - F063 and F272

The STD NMR spectra (blue) of endothiapepsin and the cocktail set show only binding for F063 (green). As F272 has several peaks that overlap with F063, it could also be that F272 is binding (e.g. peak at 2.5ppm). In the crystal structure only F063 could be clearly assigned to the electron density.

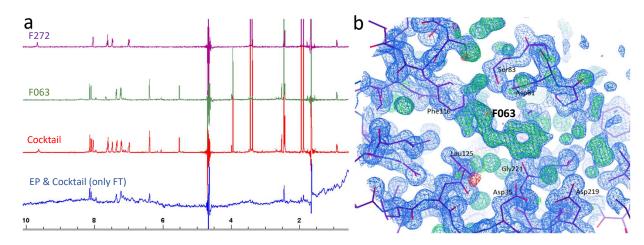


Figure 4.35: a) STD NMR spectra of endothiapepsin and cocktail set shows binding for F063 and overlapping peaks of F063 and F272 may also indicate binding for F272. b) The crystal structure shows clear electron density for F063 only.

Cocktail M - F063 and F268

The STD NMR spectra (blue) of endothiapepsin and the cocktail set shows binding for both F063 (green) and F268 (violet). In the crystal structure F063 could be clearly assigned to the electron density while F268 had a weak and dispersed electron density so could not be built in.

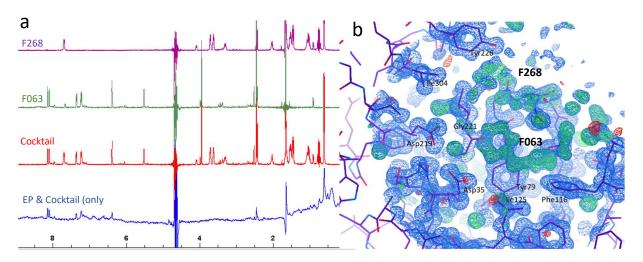


Figure 4.36: a) STD NMR spectra of endothiapepsin and cocktail set shows binding for F063. b) The crystal structure shows a clear electron density for F063 and a less clear electron density for F268.

Cocktail N - F207 and F240

The STD NMR spectra (blue) of endothiapepsin and the cocktail set shows no binding for either fragment. It is worth mentioning that F207 did not show binding in the STD NMR when tested as individual fragments with endothiapepsin. Similar to cocktail N, the spectrum of the cocktail set does not correspond to the ¹H 1D spectra of F207. This may reflect insolubility of F207 in the cocktail set and additional signals at 3 ppm may reflect an impurity in the cocktail set. In the crystal structure both fragments could be clearly assigned to the electron density where two copies of F207 bind clearly, one of each stereoisomer, in addition to one copy of F240.

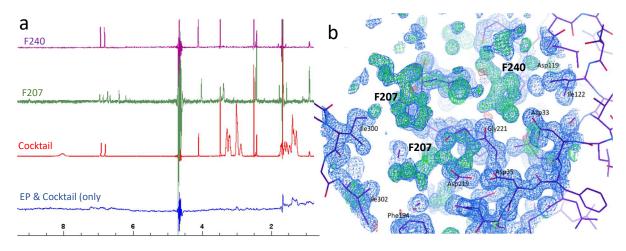


Figure 4.37: a) STD NMR spectra of endothiapepsin and cocktail set shows no binding for either fragment. b) The x-ray structure shows the electron densities for two copies of F207 and one copy of F240.

STD NMR records the binding properties of a fragment to the protein as a dynamic event and not only the final result of the binding event as is the case in crystallography. The temporary binding and unbinding of a fragment in NMR make it possible that several fragments can bind consecutively at different time points, allowing all of them to experience the magnetization transfer. This eliminates drawbacks faced with cocktailing in crystallography as competitive binding between fragments can result in poorly defined electron density, making it difficult to assign it to fragments during structure refinement. 175 Here fragments penetrate simultaneously the solvent channels in the crystal packing and assemble at available binding positions on the protein. The efficiency with which the population at a site builds up depends on the affinity, but also on the diffusion rate through the channels. These can be different and there is possibly a gradient in the fragment occupancy from the crystal's rim to the center. An example of this is F338, for which the electron density in cocktail soaking was weakened due to possible expelling by the competitive F216. Therefore, fragment cocktailing in NMR can allow detection of fragments binding, even in the presence of direct competition. On the other hand, fragments in NMR may have signals at the same value on the ppm scale (if not assembled in a way to avoid this) and therefore overlapping information in the resulting STD NMR spectra make interpretation difficult to reliably distinguish which fragments are binding. Some fragments were hits in crystallography when soaked as individual fragments but could not be assigned to the electron density in the structures of cocktail soaking. These fragments could be easily detected in NMR cocktails, for example F224, F333, F052, and F338.

In contrast, some fragments that could be resolved by crystallography in cocktail soaking could not be seen as binding events in STD NMR, for example F240. Fragments that did not bind when tested as single fragments in STD NMR such as F207, F227, and F274 also showed no binding when tested in a cocktail set, as expected.

For the tested cocktail sets, STD NMR was just as likely to detect fragment binding of the fragments in the cocktail set as crystallography was. We must conclude that despite having cases where NMR cocktailing out-performed crystallography in detecting the fragments in cocktails, NMR did not detect binding for all of the fragments in the tested cocktail sets. They were, however, detected as hits when tested as individual fragments which is definitely a case of a false negative in NMR cocktailing. Cocktailing in crystallography can also lead to false negatives due to competitive fragment behavior. This fact still holds the advantage of elucidating how the fragments in the cocktail are binding and whether they adopt different binding poses compared to single soaks or not. It must be considered, however, that in some cases cocktailing in crystallography can reveal a different binding mode for a given fragment compared to the adopted geometry found when soaking them individually. An example of this can be seen for cocktail set A (fragments 63, 267, and 291). F063 in the cocktail soak binds in a remote location to its original binding position in the individual soak (PDB: 4Y57). This can provide incorrect information regarding the binding locations of fragments in cocktail soaking (Figure 4.38).

An overview of our comparative analysis can be seen in Table 4.8 below.

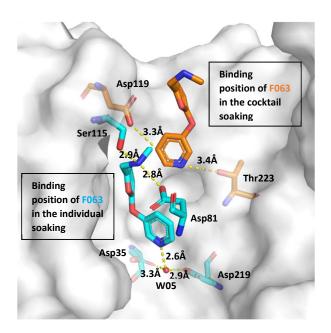


Figure 4.38: Binding position of F063 (cyan) when soaked as an individual fragment (PDB: 4Y57) binds in the catalytic dyad where it forms a water-mediated hydrogen bond (W05) (2.6Å) with its pyridine to the catalytic residues Asp35 (3.3Å) and Asp219 (2.9Å), in addition to two hydrogen bonds formed by its terminal nitrogen to the carboxylate of Asp81 (2.8Å) and carbonyl of Ser115 (2.9Å). The binding position of F063 (orange) when soaked as part of a cocktail set however is markedly different. It no longer forms a water-mediated interaction to the catalytic dyad but instead binds near the S1 and S3 pockets, where it interacts with Asp119 (3.3Å) and Thr223 (3.4Å).

Table 4.8: Results of fragment hit detection in fragment cocktails screened by STD NMR in comparison to cocktailing in X-ray crystallography.

Cocktail	Fragment	Structures	Cocktail STD NMR Hit	Cocktail X-ray Hit (occupancy)	Individual STD NMR Hit (second STD NMR)	Individual X-ray Hit
A (pyridine moieties)	63	N O O HN	yes	yes (73 %) In competition with F267, and thus binds in a different position in the cocktail soak than when soaked as an individual fragment.	yes	yes
	267		yes	yes (73%)	yes	yes

	291	N N N N N N N N N N N N N N N N N N N	yes	yes (73%, 25%)	yes	yes
Cocktail	Fragment	Structures	Cocktail STD NMR Hit	Cocktail X-ray Hit (occupancy)	Individual STD NMR Hit (second STD NMR)	Individual X-ray Hit
B (aliphatic	224	HN	yes	no Seems to be displaced by F236.	yes	yes
secondary amine as warhead)	236	H CI N O	yes	yes (80%)	yes	yes
C (structural diversity)	103	F O NH ₂	no	no	no	yes
	171	H ₂ N—N	yes	yes (72%)	yes	yes
D	308	٥	no	no	yes	no
(structural diversity)	333	-S-N	yes	no Electron density unclear to assign to F333.	yes	yes
E	52	H ₂ N Br	yes	no It is in competition with F063.	yes	yes
	58	N-N NH ₂	no	no It is in competition with F063.	no	yes
	63	N O O HN-	yes	yes (82%) ^a	yes	yes

	1	Ι	I			<u> </u>
F	207	HN N	no Possible signal overlap of F207 and F261.	yes (100%)	no	yes
	261	N N N N	yes	yes (83%) Only three out of the two fragment copies seen in individual soaks are seen in cocktail soaking, due to competition with F207 which binds in one of the positions of the three copies of F261.	yes	yes
Cocktail	Fragment	Structures	Cocktail STD NMR Hit	Cocktail X-ray Hit	Individual STD NMR Hit (second STD NMR)	Individual X-ray Hit
	216	NH NH ₂	yes	Yes (80%)	yes	yes
G	338	N N N N N N N N N N N N N N N N N N N	yes	weak (67%) Seems to be in competition with F216 which slightly expels it out of its binding position.	yes	Yes
Н	218	N N NH ₂	yes	yes (82%, 74%, 76%, 78%) Binds in three copies in the catalytic dyad, as well as another copy in a remote binding location near the S6 pocket.	yes	yes
	224	HN	no	no	yes	yes

	227	Вг	no	Yes (69%)	no	yes
I	274	F NH ₂	no	Yes (69%)	no	yes
Cocktail	Fragment	Structures	Cocktail STD NMR	Cocktail X-ray Hit	Individual STD NMR Hit (second STD	Individual X-ray Hit
			Hit		NMR)	
J	41		no	Yes (73%) ^a	yes	yes
	51	N N N N N N N N N N N N N N N N N N N	weak	Yes (76%) ^a	no	yes
				no		
K	162	S T T N	no	Electron density unclear for assignment.	yes	yes
	236	H C N', O.	no	no Electron density unclear for assignment.	yes	yes
	338	N N N N N N N N N N N N N N N N N N N	yes	no Electron density unclear for assignment.	yes	yes
	63	N O O HN-	yes	Yes (100%)	yes	yes
			no			
L	272	N N N N N N N N N N N N N N N N N N N	Possible overlap in signals of F207 and F261.	no	yes	yes
	63	N O O HN	yes	Yes (92%) ^a	yes	yes
М	268	Sin C	yes	Weak, cannot be built in	yes	yes

N	207	H C H	no	Yes (99%, 92%)	no	yes
	240	Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z	no	Yes (100%)	no	yes

a) Value after one round of structural refinement

Overall, it remains dubious whether the presumed speed improvement of screening fragments in cocktails is advisable, particularly in crystallography, in light of the huge number of created artifacts, e.g. arbitrarily modulated binding modes, or false positively and negatively detected hits. The danger exists that false information is followed up or promising fragments are lost in a screen. To recover such information would involve complex deconvolution of the cocktails which at the end will be elaborate, time consuming, and most likely avoidable with individual soaks. As for NMR, despite the ability to detect competitive binding of fragments due to the temporary binding and unbinding events, the parallel binding and thus detection of fragments is not always guaranteed. As seen in our experiments, 20% of the fragments that were detected as hits when screened individually in NMR were not detected as hits when screened as cocktails. Additionally, there are two cases were fragments were detected or not detected in the cocktail NMR depending on the cocktail set they were a part of. These fragments are F224 and F236. When screened as pairs in a cocktail set, both F224 and F236 were detected as hits. On the other hand, when F224 was screened as part of a cocktail set with F218 it was not detected, and F236 when screened as part of a cocktail set with F162 and F338 was also not detected.

4.6 Materials and Methods

4.6.1 Purification of EP

Endothiapepsin aliquots were obtained from Suparen (provided by DSM Food Specialties) and the storage buffer was exchanged with 0.1 M sodium acetate (pH 4.6) using a Vivaspin 20 with a 10 kDa molecular weight cutoff. Protein concentration was determined by absorbance at λ = 280 nm, assuming an extinction coefficient of 1.15 for 1 mg·mL⁻¹ solutions. For NMR experiments, the buffer was further exchanged with 3 mM acetic acid (pH 4.6).

4.6.2 Crystallization and Structure Refinement of EP

Endothiapepsin crystals were grown at 18°C using the sitting-drop vapor diffusion method.

 $2~\mu L$ of 5 mg \cdot mL $^{-1}$ were mixed with $2~\mu L$ of reservoir solution (0.1 M sodium acetate pH 4.6, 0.1 M ammonium acetate, 28% - 30% (w/v) PEG4000) in the wells of a crystallization plate containing 1 mL reservoir solution. Crystals could be obtained within three days.

The crystallography screening of the 361 fragment library was done by Dr. Nedyalka Radeva and experimental details can be found in her dissertation. ²⁰⁰

Displacer ritonavir was purchased from Sigma Aldrich. For the reporter ligand experiments, initially the endothiapepsin crystals were soaked with a solution containing 2 mM of ritonavir (dissolved in 100% DMSO), 0.1 M sodium acetate pH 4.6, 0.1 M ammonium acetate, 28% - 30% (w/v) PEG4000, 25% glycerol, for 6 hours. They were then transferred to a well containing 100 mM of fragment solution

(dissolved in 100% DMSO), 0.1 M sodium acetate pH 4.6, 0.1 M ammonium acetate, 28% - 30% (w/v) PEG4000, 25% glycerol, for 18 hours. In the second part of the reporter ligand experiments, the endothiapepsin crystals were soaked with 2 mM of ritonavir in the same way as described above but for 5 hours, after which they were transferred to a well containing both 2 mM of the reporter ligand and 100 mM of the fragment solution (same solution described as above) for 24 hours.

For the cocktail soaking experiments performed by Helene Köster and Johannes Schiebel, *apo* endothiapepsin crystals were soaked into cocktails of fragments each at 90 mM concentration, with the exception of cocktails B and D where the fragment concentration was 50 mM, and cocktails L, M, and N where the fragment concentration was 100 mM.

Data collection and refinement statistics can be found in the Appendix (Table 9 – Table 15). The diffraction data were indexed, scaled, and merged using XDS^{90} and XDSAPP. Molecular replacement from the program PHASER MR⁹² from the CCP4 suite⁹³ was used to determine all crystal structures. The structure 5DQ4 was used as a search model. In the refinement, a 5% subset of all reflections was omitted during refinement to be used for R_{free} calculation. Model building was achieved in COOT⁴⁴ and refinement using PHENIX.refine version 1.10.1-2155. Cartesian simulated annealing with default parameters was used as a first refinement step for all the structures. This was followed by refinement of XYZ coordinates and occupancies of protein residues and fragments (with the exception of water molecules whose occupancies were fixed). In the case of protein residues that gave additional density, they were refined in double confirmation and kept if their refined occupancy was \geq 20%. When appropriate, hydrogen atoms were added to the refined structures in the last refinement step in PHENIX.refine. For published structures, details of their refinement can be found in the dissertation of Dr. Nedyalka Radeva. Chemicalize developed by ChemAxon was used for name-to-structure generation and SMILES code notation. The ligand PDB and restraint files were generated with the Grade Web Server.

4.6.3 SPR Conditions

SPR screens were performed at 25°C using a BIAcore 3000 instrument.

Immobilizing H-142 Peptide

The functional groups of a CM7 GE sensor chip surface were first activated by injecting for 10 min at 10 μ L/min with a 1:1 mixture of 0.5 M 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and 0.5 M *N*-hydroxysuccinimide (NHS), immediately followed by injecting the H-142 target compound for 7 min at a flow rate of 10 μ L/min to 865RU (where RU is the resonance units). The relation between RU and ng of material bound varies depending on the refractive index of the analyte. The H-142 was prepared by diluting a 25 mg · mL ⁻¹ DMSO stock solution of H-142 to 0.5 mg · mL ⁻¹ in a 100 mM Na-acetate buffer, pH 4. Remaining activated carboxyl groups on the surface were blocked with 4x1.5 min pulses of 0.5 M ethanolamine.

<u>Calibration Free Concentration Analysis</u>

A 96-well plate of various endothiapepsin concentrations (0 to 100 mM) was prepared by diluting the protein into the running buffer composed of 25 mM Na-acetate pH 4.5, 125 mM NaCl, 0.005% Tween 20, 1% DMSO. These protein solutions were passed over the immobilized H-142 and the concentration was plotted against the corresponding response values (RU) to create a dose-response curve. The response in the SPR assay, expressed as the initial binding rate, was linear with the free protein concentration in solution, which ensures that the measurements are done under conditions of mass transport limitation, as can be seen in Figure 4.39. The endothiapepsin concentration used in subsequent screening was 50 nM.

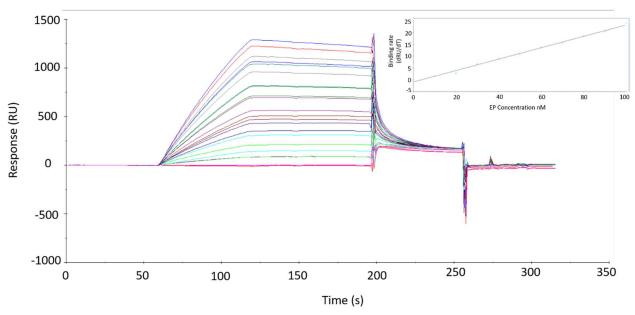


Figure 4. 39: Calibration of BIAcore ISA experiments. Concentration series (0 to 100nM) of EP were run in duplicates at 20 μ L/min at 25°C over a CM7 surface (sensor chip CM7 - GE Healthcare) with an amine-coupled peptide as TDC. The running buffer was composed of 25 mM sodium acetate, 200 mM NaCl, 0.005% Tween 20, pH 4.5. After each cycle the surface was regenerated with a pulse of 100 mM Tris at pH 8.5. The insert depicts the concentration dependency of the initial binding rate (dRU/dt) with the protein concentration.

Inhibition in Solution Assay

Endothiapepsin was diluted to a concentration of 100 nM in the running buffer. Fragment plates were prepared by diluting 100 mM fragment stocks in running buffer to a final concentration of 2 mM. The plates were centrifuged to remove any precipitation and the supernatant was used to prepare the screening plates containing a final concentration of 50 nM endothiapepsin and 1 mM fragment. The fragment and endothiapepsin mixture were injected onto the H-142 immobilized surface at 20 μL/min, and the surface was regenerated after each cycle using 15 s pulses of 100 mM Tris at pH 8.5. The initial binding rate (dRU/dt) determined within the first 15 s after injection was used as a measure of the free protein concentration (using the subtracted sensorgrams). By dividing the free protein concentration measured for each fragment by that of the control measurement without fragment, we can get the percentage bound to protein, and therefore percentage inhibited by the fragment. Anything above 3% inhibition was considered a hit. Fragments that gave an inhibition of at least 50% were then run again using varying concentrations from 0 to 1024 µM. Using the nonlinear regression in GraphPad Prism, the logarithmic fragment concentration was plotted against the percentage inhibition to give the IC₅₀. In GraphPad Prism, all negative percentage inhibition values were changed to zero, the constraint parameters used included setting the bottom and top constants to 0 and 100, respectively, and the Hill slope to shared value for all data sets. Only the IC₅₀ values with R-squared above 80% were taken into consideration. In some cases, the fragments were run in duplicates when possible.

4.6.4 NMR conditions

STD and Water LOGSY fragment screening experiments were performed at 283K on a Bruker Avance 600 MHz NMR spectrometer equipped with a cryogenically cooled probe head. Automatic sample changing was accomplished with a SampleJet system. Three separate NMR samples with each fragment were prepared for screening in a buffer consisting of 3 mM acetate pH 4.6. The first sample contained 310 μ M fragment, the second contained 310 μ M fragment in addition to 9.6 μ M of endothiapepsin, and the third contained 310 μ M fragment in addition to 9.6 μ M of endothiapepsin

and 16 μ M of ritonavir. A liquid handling Gilson pipetting robot was used to fill 3 mm NMR tubes in a SampleJet rack. Final DMSO concentration in solutions with fragments was 0.20%, and 0.36% in solutions with fragment and ritonavir. For STD spectra (pulse program $kf_zggpw5std$) 320 scans were collected and the interscan delay was 2.5 s. In WaterLOGSY experiments (pulse program $s_sM.ephogsy.noe_sO$) 16 scans were taken and the interscan delay was 2 s. Spectra were processed and analyzed with $TopSpin^{\circ}$. The water suppression used in STD NMR was WATERGATE w5 with gradients. ¹⁶² The water suppression used in WaterLOGSY was excitation sculpting and water flip back. The same methods were applied to the cocktail screening, with the exception of the samples preparation where fragments were added in pairs of two or three, at a fragment concentration of 310 μ M. Final DMSO concentration in solutions with two fragments was 0.41%, and 0.62% in solutions with three fragments. ¹H 1D NMR spectra of the single fragments were generated using the Water Suppression by Excitation Sculpting (pulse sequence zggpw5¹⁶²) where 64 scans were accumulated, with interscan delay of 1 s.

The STD NMR comparative analysis experiments between fragments in deuterated and non-deuterated buffers were performed at at 25°C using a Bruker Avance III-HD 800 MHz NMR spectrometer equipped with a cryogenically cooled CPTXI probe head. Sample preparation was done off-line using a TECAN EVO Robot running Gemini 4.0. Sample changing was done via a Bruker SampleRail system and data acquisition was done using ICON NMR. Water suppression was done using excitation sculpting with gradients. For the STD (pulse program: *std2.poe*) number of scans collected was 1024, and the interscan delay was 3 s, with the exception of the data set for fragment 39 which had 512 scans. For the ¹H 1D spectrum (pulse program: *zgesgppr*) 32 scans were accumulated with a relaxation delay of 2 s.

NMR samples were prepared in buffers of heavy and light water. For the deuterated heavy water buffer the sample was prepared in 20 mM CD $_3$ COONa at pH 4.6 with 300 μ M fragment in 100% D $_2$ O. Endothiapepsin was then added at a concentration of 9.6 μ M, as well as ritonavir in case of the competitive binding experiments at a concentration of 16 μ M. For the non-deuterated light buffer the sample was prepared in 20 mM NaAc at a of pH 4.6, 10% D $_2$ O (added to allow locking using the deuterium resonance), 90% H $_2$ O, and 300 μ M of fragment. Endothiapepsin was then added at a concentration of 9.6 μ M, as well as ritonavir for the competitive binding experiments at a concentration of 16 μ M. Final DMSO concentration in solutions with fragments was 0.20%, and 0.36% in solutions with fragment and ritonavir.

Bibliography

- Bosch, F., Rosich, L. The Contributions of Paul Ehrlich to Pharmacology: A Tribute on the Occasion of the Centenary of His Nobel Prize. *Pharmacology* 82, 171–179 (2008).
- Gund, P. Three-Dimensional Pharmacophoric Pattern Searching. *Prog. Mol. Subcell. Biol.* 5, 117-143 (1977).
- 3. Whittaker, M., Law, R. J., Ichihara, O., Hesterkamp, T., Hallett, D. Fragments: past, present and future. *Drug Discov. Today Technol.* **7**, e163–e171 (2010).
- 4. Shuker, S. B., Hajduk, P. J., Meadows, R. P., Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **274**, 1531–1534 (1996).
- 5. Mattos, C., Dagmar, R. Locating and characterizing binding spots on proteins. *Nat. Biotechnol.* **14**, 595–599 (1996).
- Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857 (1985).
- Joseph-McCarthy, D., Tsang, S. K., Filman, D. J., Hogle, J. M., Karplus, M. Use of MCSS to Design Small Targeted Libraries: Application to Picornavirus Ligands. *J. Am. Chem. Soc* 123, 12758–12769 (2001).
- 8. Patel, D., Bauman, J. D., Arnold, E. Advantages of crystallographic fragment screening: Functional and mechanistic insights from a powerful platform for efficient drug discovery. *Prog. Biophys. Mol. Bio.* **116**, 92–100 (2014).
- 9. Congreve, M., Carr, R., Murray, S., Jhoti, H., A 'Rule of Three' for fragment-based lead discovery?

 Drug Discov. Today 8, 876-877 (2003).
- Koester, H., Craan, T., Brass, S., Herhaus, C., Zentgraf, M., Neumann, L., Heine, A., Klebe, G. A
 Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal
 Structures with Various Fragment Chemotypes. J. Med. Chem. 54, 7784–7796 (2011).
- 11. Scott, D. E., Coyne, A. G., Hudson, S. A., Abell, C. Fragment-Based Approaches in Drug Discovery and Chemical Biology. *Biochemistry* **51**, 4990–5003 (2012).

- 12. Hajduk, P. J., Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Disc.* **6**, 211–219 (2007).
- 13. Erlanson, D. A., McDowell, R. S., O'Brien, T. Fragment-Based Drug Discovery. *J. Med. Chem.* 47, 3463–3482 (2004).
- 14. Hopkins, A. L., Groom, C. R., Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **9**, 430–431 (2004).
- Dimitropoulos, D., Ionides, J., Henrick, K. Using MSDchem to Search the PDB Ligand Dictionary.
 In Curr. Protoc. Bioinformatics (John Wiley & Sons, Inc., 2006).
- 16. Huschmann, F. U., Linnik, J., Sparta, K., Uehlein, M., Wang, X., Metz, A., Schiebel, J., Heine, A., Klebe, G., Weiss, M. S., Mueller, U. Structures of endothiapepsin–fragment complexes from crystallographic fragment screening using a novel, diverse and affordable 96-compound fragment library. *Acta Cryst., F* 72, 346–355 (2016).
- 17. Bakhtiar, R. Surface Plasmon Resonance Spectroscopy: A Versatile Technique in a Biochemist's Toolbox. *J. Chem. Edu.* **90**, 203–209 (2013).
- 18. SPRpages, accessed June 2019 < https://www.sprpages.nl/spr-overview/spr-theory >
- Thillaivinayagalingam, P., Gommeaux, J., McLoughlin, M., Collins, D., Newcombe, A. R.
 Biopharmaceutical production: Applications of surface plasmon resonance biosensors. *J. Chromatogr., B* 878, 149–153 (2010).
- 20. GE Healthcare Life Sciences, accessed June 2019
 - < https://www.gelifesciences.com/en/us/solutions/protein-research/knowledgecenter/surface-plasmon-resonance/surface-plasmon-resonance >
- 21. Biacore, accessed May 2019 < https://www.biacore.com >
- 22. GE Health Care Life Sciences. Biacore Assay Handbook 29-0194-00 Edition AA.
- 23. Kuroki, K., Maenaka, K. Analysis of Receptor–Ligand Interactions by Surface Plasmon Resonance.

 In *Immune Receptors* 748, 83–106 (Humana Press, 2011).

- 24. Holdgate, G. A., Anderson, M., Edfeldt, F., Geschwindner, S. Affinity-based, biophysical methods to detect and analyze ligand binding to recombinant proteins: Matching high information content with high throughput. *J. Struct. Biol.* **172**, 142–157 (2010).
- 25. Radiopaedia, accessed June 2019 < https://radiopaedia.org >
- 26. Questions and Answers in MRI, accessed June 2019 < http://mriquestions.com >
- 27. Chemistry LibreTexts, accessed June 2019 < https://chem.libretexts.org >
- 28. James, T. L. Fundamentals of NMR. University of California, 1- 31 (1998):
 https://qudev.phys.ethz.ch/phys4/studentspresentations/nmr>
- 29. Harner, M. J., Frank, A. O., Fesik, S. W. Fragment-Based Drug Discovery Using NMR Spectroscopy. *J. Biomol. NMR* **56**, 65–75 (2013).
- 30. Institute of Chemistry, NMR Lab, accessed June 2019 < http://chem.ch.huji.ac.il >
- 31. The resonance, accessed June 2019 < https://www.theresonance.com >
- 32. Hahn, T. International tables for crystallography. Vol. A: Space-group symmetry. (Wiley, 1983)
- 33. Radeva, N. From Active-site Mapping to Lead Discovery using Fragment-based Approaches on the Aspartic protease Endothiapepsin (Philipps-Universitaet Marburg, 2016).
- 34. Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369**, 253–287 (1921).
- 35. McPherson, A., Gavira, J. A. Introduction to protein crystallization. *Acta Crysta., F* **70**, 2–20 (2014).
- 36. Smyth, M. S., Martin J. H. X Ray Crystallography. *Mol Pathol.* **53**, 8–14 (2000).
- 37. MIT Department of Chemistry X-Ray Diffraction Facility, accessed June 2019
 < https://web.mit.edu/x-ray/cystallize.html >
- 38. Petsko, G. A. Protein crystallography at sub-zero temperatures: Cryo-protective mother liquors for protein crystals. *J. Mol. Biol.* **96**, 381–392 (1975).
- 39. Spectral Instruments, accessed June 2019 < http://www.specinst.com/What Is A CCD.html >
- 40. X-Ray Detectors, accessed June 2019 < https://www.rigaku.com >
- 41. Klebe, G. *Drug Design*. (Springer Reference, 2013).

- 42. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., Leslie, A. G. W. *iMOSFLM*: a new graphical interface for diffraction-image processing with *MOSFLM*. *Acta Crysta.*, *D* **67**, 271–281 (2011).
- 43. Electron Density Maps, accessed June 2019< https://openwetware.org/wiki/BMCB625:pdb_analysis >
- 44. Emsley, P., Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crysta., D* **60**, 2126–2132 (2004).
- 45. Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter III, H. L., Deng, S-J. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R., Shewchuk, L. M. Crystallization of protein–ligand complexes. *Acta Crysta., D* 63, 72–79 (2007).
- 46. Ehrmann, F. R., Stojko, J., Metz, A., Debaene, F., Barandun, L. J., Heine, A., Diederich, F., Cianferani, S., Reuter, K., Klebe, G. Soaking suggests "alternative facts": Only co-crystallization discloses major ligand-induced interface rearrangements of a homodimeric tRNA-binding protein indicating a novel mode-of-inhibition. *PLOS ONE* **12**, e0175723 (2017).
- 47. Christopher, P. R., David, K. V., John, S. M., Sankarapandian, V. Antibiotic therapy for Shigella dysentery. *Cochrane Database of Syst. Rev.* **8**, 1-102 (2010).
- 48. Heta, S., Robo, I. The Side Effects of the Most Commonly Used Group of Antibiotics in Periodontal Treatments. *Med. Sci.* **6**, 6 (2018).
- 49. S. K. Shigellosis. *J. Microbiol.* **43**, 133–143 (2005).
- 50. Romier, C., Reuter, K., Suck, D., Ficner, R. Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange. *EMBO. J* **15**, 2850–2857 (1996).
- 51. Graedler, U. H.-D. Gerber, D. M. Goodenough-Lashua, G. A. Garcia, R. Ficner, K. Reuter, M. T. Stubbs, Klebe, G. A new target for shigellosis: rational design and crystallographic studies of inhibitors of tRNA-guanine transglycosylase11Edited by R. Huber. J. Mol. Biol. 306, 455–467 (2001).

- 52. Hurt, J. K., Olgen, S., Garcia, G. A. Site-specific modification of Shigella flexneri virF mRNA by tRNA-guanine transglycosylase in vitro. *Nucleic Acids Res.* **35**, 4905–4913 (2007).
- 53. Stengl, B., Reuter, K., Klebe, G. Mechanism and Substrate Specificity of tRNA-Guanine

 Transglycosylases (TGTs): tRNA-Modifying Enzymes from the Three Different Kingdoms of Life

 Share a Common Catalytic Mechanism. *ChemBioChem* **6**, 1926–1939 (2005).
- 54. Sievers, F. Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez., R., McWilliam, H., Remmert, M., Soeding, J., Thompson, J. D., Higgins D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2014).
- 55. Meyer, E. A., Donati, N., Guillot, M., Schweizer, W. B., Diederich, F. Synthesis, Biological Evaluation, and Crystallographic Studies of Extended Guanine-Based (lin-Benzoguanine) Inhibitors for tRNA-Guanine Transglycosylase (TGT). Helv. Chim. Acta 89, 573–597 (2006).
- 56. Barandun, L. J., Ehrmann, F. R., Zimmerli, D., Immekus, F., Giroud, M., Gruenenfelder, C., Schweizer W. B., Bernet, B., Betz, M., Heine, A., Klebe, G., Diederich, F. Replacement of Water Molecules in a Phosphate Binding Site by Furanoside-Appended *lin* -Benzoguanine Ligands of tRNA-Guanine Transglycosylase (TGT). *Chem. Eur. J.* 21, 126–135 (2015).
- 57. Naomi Tidten. *Structural and mutational characterisation of Shigella Pathogenicity Factors*. (Philipps-Universitaet Marburg, 2007).
- 58. Koester, H., Craan, T., Brass, S., Herhaus, C., Zentgraf, M., Neumann, L., Heine, A., Klebe, G. A Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes. *J. Med. Chem.* **54**, 7784–7796 (2011).
- 59. Silvestre, H. L., Blundell, T. L., Abell, C., Ciulli, A. Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery. *PNAS* **110**, 12984–12989 (2013).
- 60. Mashalidis, E. H., Śledź, P., Lang, S., Abell, C. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat. Protoc.* **8**, 2309–2324 (2013).
- 61. Myszka, D. G. Improving biosensor analysis. J. Mol. Recognit. 12, 279–284 (1999).

- Barandun, L. J., Immekus, F., Kohler, P. C., Ritschel, T., Heine, A., Orlando, P., Klebe, G., Diederich,
 F. High-affinity inhibitors of *Zymomonas mobilis* tRNA–guanine transglycosylase through convergent optimization. *Acta Crysta., D* 69, 1798–1807 (2013).
- 63. Meiboom, S., Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **29**, 688–691 (1958).
- 64. Tidten, N., Stengl, B., Heine, A., Garcia, G. A., Klebe, G., Reuter, K. Glutamate versus Glutamine Exchange Swaps Substrate Selectivity in tRNA-Guanine Transglycosylase: Insight into the Regulation of Substrate Selectivity by Kinetic and Crystallographic Studies. *J. Mol. Biol.* **374**, 764–776 (2007).
- 65. Biela, I., Tidten-Luksch, N., Immekus, F., Glinca, S., Nguyen, P. T. X., Gerber, H-D., Heine, A., Klebe, G., Reuter, K. Investigation of Specificity Determinants in Bacterial tRNA-Guanine Transglycosylase Reveals Queuine, the Substrate of Its Eucaryotic Counterpart, as Inhibitor. *PLoS ONE* 8, e64240 (2013).
- 66. Ritschel, T., Atmanene, C., Reuter, K., Van Dorsselaer, A., Sanglier-Cianferani, S., Klebe, G. An Integrative Approach Combining Noncovalent Mass Spectrometry, Enzyme Kinetics and X-ray Crystallography to Decipher Tgt Protein-Protein and Protein-RNA Interaction. *J. Mol. Biol.* 393, 833–847 (2009).
- 67. Jakobi, S., Nguyen, P. T. X., Debaene, F., Cianferani, S., Reuter, K., Klebe, G. What Glues a Homodimer Together: Systematic Analysis of the Stabilizing Effect of an Aromatic Hot Spot in the Protein–Protein Interface of the tRNA-Modifying Enzyme Tgt. *ACS Chem. Biol.* **10**, 1897–1907 (2015).
- 68. Pearce, N. M., Krojer, von Delft, F. Proper modelling of ligand binding requires an ensemble of bound and unbound states. *Acta Crysta., D* **73**, 256–266 (2017).
- 69. Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., Marsden, B. D., Kelm, S., Shi, J., Deane, C. M., von Delft, F. A multi-crystal method for extracting obscured

- crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **8**, 15123 (2017).
- Schiebel, J., Radeva, N., Koester, H., Metz, A., Krotzky, T., Kuhnert, M., Diederich, W. E., Heine, A., Neumann, L., Atmanene, C., Roecklin, D., Vivat-Hannah, V., Renaud, J-P., Meinecke, R., Schlinck, N., Sitte, A., Popp, F., Zeeb, M., Klebe, G. One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* 10, 1511–1521 (2015).
- 71. Ruth Brenk. Virtuelles Screening, strukturbasiertes Design und Kristallstrukturanalyse von Inhibitoren der tRNA-Guanin Transglykosylase, ein Target der Bakterienruhr. (Philipps-Universitaet Marburg, 2002).
- 72. Davis, B. J., Erlanson, D. A. Learning from our mistakes: The 'unknown knowns' in fragment screening. *Bioorg. Med. Chem. Lett.* **23**, 2844–2852 (2013).
- 73. Murray, C. W., Verdonk, M. L., Rees, D. C. Experiences in fragment-based drug discovery. *Trends Pharmacol. Sci.* **33**, 224–232 (2012).
- 74. Neeb, M., Czodrowski, P., Heine, A., Barandun, L. J., Hohn, C., Diederich, F., Klebe, G. Chasing Protons: How Isothermal Titration Calorimetry, Mutagenesis, and p K a Calculations Trace the Locus of Charge in Ligand Binding to a tRNA-Binding Enzyme. *J. Med. Chem.* **57**, 5554–5565 (2014).
- 75. Kohler, P. C., Ritschel, T., Schweizer, W. B., Klebe, G., Diederich, F. High-Affinity Inhibitors of tRNA-Guanine Transglycosylase Replacing the Function of a Structural Water Cluster. *Chem. Eur. J.* **15**, 10809–10817 (2009).
- 76. Hoertner, S. R., Ritschel, T., Stengl, B., Kramer, C., Schweizer, W. B., Wagner, B., Kansy, M., Klebe, G., Diederich, F. Potent Inhibitors of tRNA-Guanine Transglycosylase, an Enzyme Linked to the Pathogenicity of theShigella Bacterium: Charge-Assisted Hydrogen Bonding. *Angew. Chem. Int. Ed.* 46, 8266–8269 (2007).

- 77. Barandun, L. J., Immekus, F., Kohler, P. C., Tonazzi, S., Wagner, B., Wendelspiess, S., Ritschel, T., Heine, A., Kansy, M., Klebe, G. From lin-Benzoguanines to lin-Benzohypoxanthines as Ligands for Zymomonas mobilis tRNA-Guanine Transglycosylase: Replacement of Protein-Ligand Hydrogen Bonding by Importing Water Clusters. *Chem. Eur. J.* **18**, 9246–9257 (2012).
- 78. Christoph Hohn. PhD Thesis DISS. ETH No. 23162 (ETH Zürich, 2015).
- 79. Lecoutey, C., Fossey, C., Rault, S., Fabis, F. Efficient Room-Temperature One-Pot Synthesis of 2-Amino-3-alkyl(3-aryl)quinazolin-4(3H)-ones. *Eur. J. of Org. Chem.* **2011**, 2785–2788 (2011).
- 80. Ehrmann, F. R., Kalim, J., Pfaffeneder, T., Bernet, B., Hohn, C., Schaefer, E., Botzanowski, T., Cianferani, S., Heine, A., Reuter, K., Diederich, F., Klebe, G. Swapping Interface Contacts in the Homodimeric tRNA-Guanine Transglycosylase: An Option for Functional Regulation. *Angew. Chem. Int. Ed.* 57, 10085–10090 (2018).
- 81. Nasielski-Hinkens, R., Levêque, P., Castelet, D., Nasielski, J. The Four 6-Halo-7-nitroquinoxalines.

 Heterocycles **26**, 2433–2442 (1987).
- 82. Neeb, M., Betz, M., Heine, A., Barandun L. J., Hohn, C., Diederich, F., Klebe, G. Beyond Affinity: Enthalpy–Entropy Factorization Unravels Complexity of a Flat Structure–Activity Relationship for Inhibition of a tRNA-Modifying Enzyme. *J. Med. Chem.* **57**, 5566–5578 (2014).
- 83. Ritschel, T., Hoertner, S., Heine, A., Diederich, F., Klebe, G. Crystal Structure Analysis and in Silico pKa Calculations Suggest Strong pKa Shifts of Ligands as Driving Force for High-Affinity Binding to TGT. *ChemBioChem* **10**, 716–727 (2009).
- 84. Brenk, R., Stubbs, M. T., Heine, A., Reuter, K., Klebe, G. Flexible Adaptations in the Structure of the tRNA-Modifying Enzyme tRNA-Guanine Transglycosylase and Their Implications for Substrate Selectivity, Reaction Mechanism and Structure-Based Drug Design. *ChemBioChem* **4**, 1066–1077 (2003).
- 85. Ruehmann, E. H., Rupp, M., Betz, M., Heine, A., Klebe, G. Boosting Affinity by Correct Ligand Preorganization for the S2 Pocket of Thrombin: A Study by Isothermal Titration Calorimetry,

- Molecular Dynamics, and High-Resolution Crystal Structures. *ChemMedChem* **11**, 309–319 (2016).
- 86. Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures: Protein crystallography for non-crystallographers. *FEBS J.* **275**, 1–21 (2008).
- 87. Haider, K., Cruz, A., Ramsey, S., Gilson, M. K., Kurtzman, T. Solvation Structure and Thermodynamic Mapping (SSTMap): An Open-Source, Flexible Package for the Analysis of Water in Molecular Dynamics Trajectories. *J. Chem. Theory Comput.* **14**, 418–425 (2018).
- 88. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig H., Shindyalov, I. N., Bourne, P. E. Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- 89. Romier, C., Meyer, J. E. W., Suck, D. Slight sequence variations of a common fold explain the substrate specificities of tRNA-guanine transglycosylases from the three kingdoms. *FEBS Lett.*416, 93–98 (1997).
- 90. Kabsch, W. XDS. Acta Crysta., D 66, 125–132 (2010).
- 91. Krug, M., Weiss, M. S., Heinemann, U., Mueller, U. *XDSAPP*: a graphical user interface for the convenient processing of diffraction data using *XDS*. *J. Appl. Crystallogr.* **45**, 568–572 (2012).
- 92. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., Read, R. J. *Phaser* crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- 93. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crysta., D* **50**, 760–763 (1994).
- 94. Adams, P. D., Afonine P.V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., Zwart P. H. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crysta., D Biological Crystallography* 66, 213–221 (2010).

- 95. Frederik Rainer Ehrmann. *Strukturbasierte Entwicklung und Charakterisierung von Inhibitoren der TGT, ein mögliches Ziel zur Therapie der Bakterienruhr.* (Philipps-Universitaet Marburg, 2016).
- 96. https://chemicalize.com
- 97. http://www.chemaxon.com
- 98. Smart, O. S., Womack, T. O., Sharff, A., Flensburg, C., Keller, P., Paciorek, W., Vonrhein, C., BricogneG (2011) grade, version v1.103. http://www.globalphasing.com.
- 99. Reducing Non-Specific Binding in Surface Plasmon Resonance Experiments, accessed March 2019
 < https://www.nicoyalife.com/wp-content/uploads/2015/11/ >
- 100. Staahlberg, J., Joensson, B., Horvath, C. Theory for electrostatic interaction chromatography of proteins. *Anal. Chem.* **63**, 1867–1874 (1991).
- 101. Feng, B. Y., Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors.

 Nat. Protoc. 1, 550–553 (2006).
- 102. Roberts, S. M., Davies, G. J. The Crystallization and Structural Analysis of Cellulases (and Other Glycoside Hydrolases). in *Methods Enzymol.* **510**, 141–168 (Elsevier, 2012).
- 103. Radeva, N., Schiebel, J., Wang, X., Krimmer, S. G., Fu, K., Stieler, M., Ehrmann, F. R., Metz, A., Rickmeyer, T., Betz, M., Winquist, J., Park, A. Y., Huschmann, F. U., Weiss, M. S., Mueller, U., Heine, A., Klebe, G. Experimental Active-Site Mapping by Fragments: Hot Spots Remote from the Catalytic Center of Endothiapepsin. *J. Med. Chem.* **59**, 7561–7575 (2016).
- 104. Bertini, I., Turano, P., Vila, A. J. Nuclear magnetic resonance of paramagnetic metalloproteins. *Chem. Rev.* **93**, 2833–2932 (1993).
- 105. Hwang, T-L., Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitary Waveforms and Pulsed Field Gradients. *J. Magn. Reson.* **112**, 275–279 (1995).
- 106. Molecular Operating Environment (MOE), Version 2015.10 (2016) Chemical Computing Group Inc., Montreal.
- 107. Case, D.A.; Cerutti, D.S.; Cheatham, T.E.; Darden, T.A.; Duke, R.E.; Giese, T.J.; Gohlke, H.; Goetz, A.W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Li, P.; Lin, C.; Liu J.;

- Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.R.; Roitberg, A.; Sagui, C.; Simmerling, C.L.; Botello-Smith, W.M.; Swails, J.; Walker, R.C.; Wang, J.; Wolf, R.M.; Wu, X.; Xiao, L.; York, D.M.; Kollman, P.A. AMBER 2017 *University of California, San Francisco*.
- 108. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters.

 *Proteins 65, 712–725 (2006).
- 109. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **120**, 9665–9678 (2004).
- 110. Le Grand, S.; Goetz, A. W.; Walker, R. C. SPFP: Speed without Compromise A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* 184, 374–380 (2013).
- 111. Goetz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. J. Chem. Theory Comput. 8, 1542–1555 (2012).
- 112. Salomon-Ferrer, R.; Goetz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald.
 J. Chem. Theory Comput. 9, 3878–3888 (2013).
- 113. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* 23, 327–341 (1977).
- 114. https://github.com/KurtzmanLab/SSTMap.
- 115. Hotez, P. J., Kamath, A. Neglected Tropical Diseases in Sub-Saharan Africa: Review of Their Prevalence, Distribution, and Disease Burden. *PLoS Negl. Trop. Dis.* **3**, e412 (2009).

- 116. Bruce, D. *Preliminary report on the tsetse fly disease or nagana, in Zululand*. (Bennett & Davis Durban, 1895).
- 117. Castellani, A. Some Observations on the Morphology of the Trypanosoma Found in Sleeping Sickness. *Br. Med. J.* **1**, 1431-1432 (1903).
- 118. Brun, R., Blum, J., Chappuis, F., Burri, C. Human African trypanosomiasis. Lancet 375, 12 (2010).
- 119. World Health Organization (April 2019) African trypanosomiasis (sleeping & sickness). Fact sheet no. 259. Available: http://www.who.int/mediacentre/factsheets/fs259/en/print.htmlAccessed 27 July 2009.
- 120. Pentreath, V. W. Trypanosomiasis and the nervous system: Pathology and immunology. *Trans. Royal Soc. Trop. Med. Hyg.* **89**, 9–15 (1995).
- 121. Parsons, M. Glycosomes: parasites and the divergence of peroxisomal purpose: Glycosome function. *Mol. Microbiol.* **53**, 717–724 (2004).
- 122. Haanstra, J. R., González-Marcano, E. B., Gualdrón-López, M., Michels, P. A. M. Biogenesis, maintenance and dynamics of glycosomes in trypanosomatid parasites. *Biochim. Biophys. Acta* 1863, 1038–1048 (2016).
- 123. Galland, N., Demeure, F., Hannaert, V., Verplaetse, E., Vertommen, D., Van Der Smissen, P., Courtoy, P. J., Michels, P. A. M. Characterization of the role of the receptors PEX5 and PEX7 in the import of proteins into glycosomes of Trypanosoma brucei. *Biochim. Biophys. Acta* 1773, 521–535 (2007).
- 124. Gatto, G. J., Geisbrecht, B. V., Gould, S. J., Berg, J. M. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nat. Struct. Biol.* **7**, (2000).
- 125. Emmanouilidis, L., Gopalswamy, M., Passon, D. M., Wilmanns, M., Sattler, M. Structural biology of the import pathways of peroxisomal matrix proteins. *Biochim. Biophys. Acta* **1863**, 804–813 (2016).

- 126. Neufeld, C., Filipp, F. V., Simon, B., Neuhaus, A., Schueller, N., David, C., Kooshapur, H., Madl, T., Erdmann, R., Schliebs, W., Wilmanns, M., Sattler, M. Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19. *EMBO J.* **28**, 745–754 (2009).
- 127. Haanstra, J. R., van Tuijl, A., Kessler, P., Reijnders, W., Michels, P. A. M., Westerhoff, H. V., Parsons, M., Bakker, B. M. Compartmentation prevents a lethal turbo-explosion of glycolysis in trypanosomes. *PNAS* **105**, 17718–17723 (2008).
- 128. Kessler, P. S., Parsons, M. Probing the Role of Compartmentation of Glycolysis in Procyclic Form

 Trypanosoma brucei: RNA INTERFERENCE STUDIES OF PEX14, HEXOKINASE, AND

 PHOSPHOFRUCTOKINASE. J. Biol. Chem. 280, 9030–9036 (2005).
- 129. Dawidowski, M., Emmanouilidis, L., Kalel, V. C., Tripsianes, K., Schorpp, K., Hadian, K., Kaiser, M., Maeser, P., Kolonko, M., Tanghe, S., Rodriguez, A., Schliebs, W., Erdmann, R., Sattler, M., Popowicz, G. M. Inhibitors of PEX14 disrupt protein import into glycosomes and kill *Trypanosoma* parasites. *Science* **355**, 1416–1420 (2017).
- 130. Emmanouilidis Leonidas. *Structural biology in peroxisome biogenesis.* (Technische Universitaet Muenchen, 2016).
- 131. Carpenter, E. P., Beis, K., Cameron, A. D., Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struc. Biol.* **18**, 581–586 (2008).
- 132. Galkin, O., Vekilov, P. G. Control of protein crystal nucleation around the metastable liquid-liquid phase boundary. *PNAS* **97**, 6277–6281 (2000).
- 133. Mueller, U., Foerster, R., Hellmig, M., Huschmann, F. U., Kastner, A., Malecki, P., Puehringer, S., Roewer, M., Sparta, K., Steffien, M., Uehlein, M., Wilk, P., Weiss, M. S. The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. *Eur. Phys. J. Plus* **130**, (2015).
- 134. Kontoyianni, M. Docking and Virtual Screening in Drug Discovery **1647**, 255–266. In *Proteomics for Drug Discovery* (Springer New York, 2017).

- 135. Lavecchia, A., Giovanni, C. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Curr. Med. Chem.* **20**, 2839–2860 (2013).
- 136. Jahn, A., Hinselmann, G., Fechner, N., Zell, A. Optimal assignment methods for ligand-based virtual screening. *J. Cheminformatics* **1**, (2009).
- 137. Villoutreix, B. O., Renault, N., Lagorce, D., Montes, M., Miteva, M.A. Free Resources to Assist Structure-Based Virtual Ligand Screening Experiments. *Curr. Protein Pept. Sc.* **8**, 381–411 (2007).
- 138. Kuhn, B., Fuchs, J. E., Reutlinger, M., Stahl, M., Taylor, N. R. Rationalizing Tight Ligand Binding through Cooperative Interaction Networks. *J. Chem. Inf. Model.* **51**, 3180–3198 (2011).
- 139. Meireles, L. M. C., Domling, A. S., Camacho, C. J. ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. *Nucleic Acids Res.* **38**, W407–W411 (2010).
- 140. Weisel, M., Proschak, E., Schneider, G. PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem. Cent. J.* **1**, 7 (2007).
- 141. Data analysis, machine learning and applications: proceedings of the 31st Annual Conference of the Gesellschaft fuer Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7-9, 2007. (Springer, 2008).
- 142. Reutlinger, M., Koch, C. P., Reker, D., Todoroff, N., Schneider, P., Rodrigues, T., Schneider, G. Chemically Advanced Template Search (CATS) for Scaffold-Hopping and Prospective Target Prediction for 'Orphan' Molecules. *Mol. Inform.* 32, 133–138 (2013).
- 143. Schneider, G., Neidhart, W., Giller, T., Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem. Int. Ed.* **38**, 2894–2896 (1999).
- 144. Hartenfeller, M., Zettl, H., Walter, M., Rupp, M., Reisen, F., Proschak, E., Weggen, S., Stark, H., Schneider, G. DOGS: Reaction-Driven de novo Design of Bioactive Compounds. *PLoS Comput. Biol.* **8**, e1002380 (2012).

- 145. Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group ULC, 1010

 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2018.
- 146. Lipinski, C. A., Lombardo, F., Dominy, B. W., Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Deliv. Rev.* **23**, 3–25 (1997).
- 147. van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Melquiond, A. S. J., van Dijk, M., de Vries, S. J., Bonvin A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol. Biol.* **428**, 720–725 (2016).
- 148. Schanda, P., Brutscher, B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. J. Am. Chem. Soc. 127, 8014–8015 (2005).
- 149. AlphaScreen™: Sensitive homogeneous assay technology handbook, PerkinElmer
- 150. Fielding, L. NMR Methods for the Determination of Protein-Ligand Dissociation Constants. *Curr. Top. Med. Chem.* **3,** 39-53 (2003).
- 151. Gemmecker, G. NMR as a Tool in Drug Research. In *NMR Spectroscopy in Drug Development and Analysis* (Wiley-VCH Verlag GmbH, 2007).
- 152. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
- 153. Sheldrick, G.M. XPREP. (Bruker AXS Inc., Madison, 2008).
- 154. Langer, G., Cohen, S. X., Lamzin, V. S., Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
- 155. Huynh, K., Partch, C. L. Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay: Protein Stability by Thermal Shift Assay. In *Current Protocols in Protein Science* (John Wiley & Sons, Inc., 2015).

- 156. Knapp, S., Mattson, P. T., Christova, P., Berndt, K. D., Karshikoff, A., Vihinen, M., Smith, C. I. E., Ladenstein, R. Thermal unfolding of small proteins with SH3 domain folding pattern. *Proteins* **31**, 309–319 (1998).
- 157. Berne, B.J., Pecora, R. Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics. (John Wiley, New York, 1976)
- 158. Bijelic, A., Rompel, A. Ten Good Reasons for the Use of the Tellurium-Centered Anderson–Evans Polyoxotungstate in Protein Crystallography. *Acc. Chem. Res.* **50**, 1441–1448 (2017).
- 159. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
- 160. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., Klenk, D. C. Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76–85 (1985).
- 161. Rodríguez, D., Sammito, M., Meindl, K., de llarduya, I. M., Protratz, M., Sheldrick, G. M., Uson, I. Practical structure solution with ARCIMBOLDO. *Acta Crysta.*, *D* **68**, 336–343 (2012).
- 162. Liu, M., Mao, X-a., Ye, C., Huang, H., Nicholson, J. K., Lindon, J. C. Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy. *J. Magn. Reson.* **132**, 125–129 (1998).
- 163. Murakami, K., Yoshioka, T., Horii, S., Hanaki, M., Midorikawa, S., Taniwaki, S., Gunji, H., Akagi, K-i., Kawase, T., Hirose, K., Irie, K. Role of the carboxy groups of triterpenoids in their inhibition of the nucleation of amyloid β42 required for forming toxic oligomers. *Chem. Comm.* 54, 6272–6275 (2018).
- 164. Origin (Pro), OriginLab Corporation, Northampton, MA, USA.
- 165. Ali-Vehmas, T., Louhi, M., Sandholm, M. Automation of the Resazurin Reduction Test using Fluorometry of Microtitration Trays. *J. Vet. Med., Series B* **38**, 358–372 (1991).
- 166. Hirumi, H., Hirumi, K. Continuous Cultivation of Trypanosoma brucei Blood Stream Forms in a Medium Containing a Low Concentration of Serum Protein without Feeder Cell Layers. *J. Parasitol.* **75**, 985 (1989).

- 167. Brun R., Schoenenberger, C. Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. *Acta Trop.* **36**, 289–92 (1979).
- 168. Hartman, A., Mondal, M., Radeva, N., Klebe, G., Hirsch, A. Structure-Based Optimization of Inhibitors of the Aspartic Protease Endothiapepsin. *Int. J. Mol. Sci.* **16**, 19184–19194 (2015).
- 169. Cooper, J. B. Endothiapepsin. In *Handbook of Proteolytic Enzymes* 147–150 (Elsevier, 2013).
- 170. Barrett, A. J., Rawlings, N. D., Woessner, J. F. *Handbook of Proteolytic Enzymes*. (Academic Press, 2012).
- 171. Coates, L., Tuan, H-F., Tomanicek, S., Kovalevsky, A., Mustyakimov, M., Erskine, P., Cooper, J. The Catalytic Mechanism of an Aspartic Proteinase Explored with Neutron and X-ray Diffraction. *J. Am. Chem. Soc.* **130**, 7235–7237 (2008).
- 172. Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A. A., Cooper, J. B. A Neutron Laue Diffraction Study of Endothiapepsin: Implications for the Aspartic Proteinase Mechanism. *Biochemistry* **40**, 13149–13157 (2001).
- 173. Cooper, J., Quail, W., Frazao, C., Foundling, S. I., Blundell, T. L. X-ray crystallographic analysis of inhibition of endothiapepsin by cyclohexyl renin inhibitors. *Biochemistry* **31**, 8142 8150.
- 174. Geschwindner, S., Olsson, L-L., Albert, J. S., Deinum, J., Edwards, P. D., de Beer, T., Folmer, R. H.

 A. Discovery of a Novel Warhead against β-Secretase through Fragment-Based Lead Generation

 § J. Med. Chem. **50**, 5903–5911 (2007).
- 175. Schiebel, J., Radeva, N., Krimmer, S. G., Wang, X., Stieler, M., Ehrmann, F. R., Fu, K., Metz, A., Huschmann, F. U., Weiss, M. S., Mueller, U., Heine, A., Klebe, G.Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chem. Biol.* 11, 1693–1701 (2016).
- 176. Schiebel, J., Radeva, N., Koester, H., Metz, A., Krotzky, T., Kuhnert, M., Diederich, W. E., Heine, A., Neumann, L., Atmanene, C., Roecklin, D., Vivat-Hannah, V., Renaud, J-P., Meinecke, R., Schlinck, N., Sitte, A., Popp. F., Zeeb, M., Klebe, G. One Question, Multiple Answers: Biochemical

- and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* **10**, 1511–1521 (2015).
- 177. Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomol. NMR* **21**, 349 359. (2001).
- 178. Lu-Yun Lian, Gordon Roberts. *Protein NMR Spectroscopy: Practical Techniques and Applications.*(John Wiley & Sons, 2011).
- 179. Barry, P. H., Lynch, J. W. Liquid junction potentials and small cell effects in patch-clamp analysis. *J. Membr. Biol.* 121, 101–117 (1991).
- 180. Voehler, M. W., Collier, G., Young, J. K., Stone, M. P., Germann, M. W. Performance of cryogenic probes as a function of ionic strength and sample tube geometry. *J. Magn. Reson.* **183**, 102–109 (2006).
- 181. Kelly, A. E., Ou, H. D., Withers, R., Doetsch, V. Low-Conductivity Buffers for High-Sensitivity NMR Measurements. *J. Am. Chem. Soc.* **124**, 12013–12019 (2002).
- 182. Flynn, P. F., Mattiello, D. L., Hill, H. D. W., Wand, A. J. Optimal Use of Cryogenic Probe Technology in NMR Studies of Proteins. *J. Am. Chem. Soc.* **122**, 4823–4824 (2000).
- 183. Roberts, G. C. K. NMR of Macromolecules. A practical approach. (Oxford University Press, 1993).
- 184. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server, 571–607. In *The Proteomics Protocols Handbook* (Humana Press, 2005)
- 185. Dalvit, C., Fasolini, M., Flocco, M. Knapp, S., Pevarello, P., Veronesi, M. NMR-Based Screening with Competition Water–Ligand Observed via Gradient Spectroscopy Experiments: Detection of High-Affinity Ligands. *J. Med. Chem.* **45**, 2610–2614 (2002).
- 186. Siriwardena, A. H., Tian, F., Noble, S., Prestegard, J. H. A Straightforward NMR-Spectroscopy-Based Method for Rapid Library Screening. *Angew. Chem. Int. Ed.* **41**, 3454–3457 (2002).

- Dalvit, C., Flocco, M. Knapp, S., Mostardini, M., Perego, R., Stockman, B. J., Veronesi, M., Varasi,
 M. High-Throughput NMR-Based Screening with Competition Binding Experiments. *J. Am. Chem.* Soc. 124, 7702–7709 (2002).
- 188. Behnen, J., Koester, H., Neudert, G., Craan, T., Heine, A., Klebe, G. Experimental and Computational Active Site Mapping as a Starting Point to Fragment-Based Lead Discovery.

 ChemMedChem 7, 248–261 (2012).
- 189. Bailey, D., Cooper, J. B., Veerapandian, B., Blundell, T. L., Atrash, B., Jones, D. M., Szelke, M. X-ray-crystallographic studies of complexes of pepstatin A and a statine-containing human renin inhibitor with endothiapepsin. *Biochem. J.* **289**, 363–371 (1993).
- 190. Cousido-Siah, A., Petrova, T., Hazemann, I., Mitschler, A., Ruiz, F. X., Howard, E., Ginell, S., Atmanene, C., Van Dorsselaer, A., Sanglier- Cianferani, S., Joachimiak, A., Podjarny, A. Crystal packing modifies ligand binding affinity: The case of aldose reductase. *Proteins* **80**, 2552–2561 (2012).
- 191. Hoye, T. R., Eklov, B. M., Ryba, T. D., Voloshin, M., Yao, L. J. No-D NMR (No-Deuterium Proton NMR) Spectroscopy: A Simple Yet Powerful Method for Analyzing Reaction and Reagent Solutions. Org. Lett. 6, 4 (2004).
- 192. Viegas, A., Manso, J., Nobrega, F. L., Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Ed.* **88**, 990–994 (2011).
- 193. Claasen, B., Axmann, M., Meinecke, R., Meyer, B. Direct Observation of Ligand Binding to Membrane Proteins in Living Cells by a Saturation Transfer Double Difference (STDD) NMR Spectroscopy Method Shows a Significantly Higher Affinity of Integrin α_{IIb} β_{II} in Native Platelets than in Liposomes. *J. Am. Chem. Soc.* **127**, 916–919 (2005).
- 194. Antanasijevic, A., Ramirez, B., Caffrey, M. Comparison of the sensitivities of WaterLOGSY and saturation transfer difference NMR experiments. *J. Biomol. NMR* **60**, 37–44 (2014).

- 195. Szelke, M., Leckie, B., Hallett, A., Jones, D. M., Sueiras, J., Atrash, B., Lever, F. Potent new inhibitors of human renin. *Nature* **299**, 555–557 (1982).
- 196. Holdgate, G. A., Anderson, M., Edfeldt, F., Geschwindner, S. Affinity-based, biophysical methods to detect and analyze ligand binding to recombinant proteins: Matching high information content with high throughput. *J. Struct. Biol.* **172**, 142–157 (2010).
- 197. Cramer, J., Schiebel, J., Wulsdorf, T., Grohe, K., Najbauer, E. E., Ehrmann, F. R., Radeva, N., Zitzer, N., Linne, U., Linser, R., Heine, A., Klebe, G. A False-Positive Screening Hit in Fragment-Based Lead Discovery: Watch out for the Red Herring. *Angew. Chem. Int. Ed.* **56**, 1908–1913 (2017).
- 198. Patel, D., Bauman, J. D., Arnold, E. Advantages of Crystallographic Fragment Screening:

 Functional and Mechanistic Insights from a Powerful Platform for Efficient Drug Discovery. *Prog Biophys Mol Biol.* **116**, 92–100 (2014).
- 199. Lindon, J. C., Ferrige, A.G. Digitisation and data processing in Fourier transform NMR. *Prog. Nucl. Mag. Res. Sp.* **14**, 27–66 (1980).
- 200. Nedyalka Radeva. From Active-Site Mapping to Lead Discovery using Fragment-based Approaches on the Aspartic protease Endothiapepsin (Philipps-Universitaet Marburg, 2016).
- 201. Laskowski, R. A., MacArthur, M. W., Moss D. S., Thornton J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
- 202. Kleywegt G. J., Zou, J. Y., Kjeldgaard, M., Jones, T. A. *International Tables for Crystallography, Volume F. Crystallography of Biological Macromolecules*. (Rossmann, M.G. and Arnold, E., 2001).
- 203. Matthews, B. W. Solvent Content of Protein Crystals. J. Mol. Biol. 33, 491-497 (1968).

Table 1: Data collection and refinement statistics for TGT structures.

PDB Code	6FSO	5SW3	5N6F
Fragment Jena ID	J14	J41	J64
A) Data Collection and Processing ^a			
Wavelength	0.9184	1.000	0.9184
Beamline	BESSY BL 14.1	ELETTRA BL 5.2R	BESSY BL 14.1
Detector	PILATUS 6M	DECTRIS PILATUS 2M	PILATUS 6M
Resolution range (Å)	43.29 - 1.45 (1.54 - 1.45)	44.46 - 1.38 (1.46 - 1.38)	44.28 - 1.12 (1.19 - 1.12)
Space group	C121	C 1 2 1	C 1 2 1
Unit cell parameters (a, b, c Å)	90.9, 64.9, 70.8	89.0, 64.1, 70.5	88.7, 64.9, 70.7
Unit cell parameters (α, β, γ°)	90.0, 95.7, 90.0	90.0, 92.8, 90.0	90.0, 93.4, 90.0
Matthews coefficient b Å ³ /Da)	2.5	2.4	2.5
Solvent content ^b (%)	51.1	49.4	50.8
Total reflections	183142 (27068)	303178 (47561)	537838 (59077)
Unique reflections	68249 (10718)	81197 (12961)	149008 (20548)
Multiplicity	2.7	3.7	3.6
Completeness (%)	94.8 (92.3)	99.7 (98.9)	97.1 (83.2)
Mean I/sigma(I)	9.8 (2.2)	12.4 (2.5)	10.3 (2.0)
Wilson B-factor (Ų)	11.5	13.6	10.9
R-sym ^c (%)	7.3 (48.6)	6.0 (49.3)	6.1 (48.7)
R-meas (%)	9.0 (60.8)	6.9 (57.7)	7.1 (59.2)
CC1/2	99.6 (76.3)	99.8 (79.7)	99.7 (72.0)
B) Refinement			
Resolution range (Å)	43.29 - 1.45	44.46 - 1.38	44.28- 1.12
Total reflections used in refinement	68178	81188	148997
Reflections used for R-work	64769	77129	141547
Reflections used for R-free	3409	4059	7450
Final R value for R-work (%)	16.6	13.3	13.6
Final R value for R-free (%)	18.6	16.2	15.2
No. of Protein residues	368	371	364
No. of Water molecules	252	223	280
No. of Ligand molecules	1	2	1
No. of Other ligand molecules	8	6	4
RMSD bond lengths (Å)	0.008	0.007	0.007
RMSD bond angles (°)	1.0	0.9	1.0
Ramachandran plot ^d			
Residues in most favored regions (%)	93.9	95	96.1
Residues in additionally allowed regions	5.4	4.7	3.6
(%)	3.4	7.7	3.0
Residues in generously allowed regions	0.3	0.3	0.3
(%)	0.5	0.5	0.5
Residues in disallowed regions (%)	none	none	None
Average B-factor all atoms ^e (Å ²)	16.2	17.9	16.1
Protein main chain	13.4	15.7	13.7
Protein side chain	16.6	18.2	16.1
Protein all atoms	15.1	17.0	15.0
Ligand	21.6	30.1	12.2
Water molecules	25.5	26.4	27.4
Other ligands ^f	36.3	36.3	33.5
Number of TLS groups	6	N/A	N/A

Table 2: Data collection and refinement statistics for TGT structures.

PDB Code	5UTI	5UTJ	5V3C
Fragment Jena ID	J72	J79	J86
A) Data Collection and Processing ^a			
Wavelength	1.000	0.9184	0.9184
Beamline	ELETTRA BL 5.2R	BESSY BL 14.1	BESSY BL 14.1
Detector	DECTRIS PILATUS 2M	PILATUS 6M	PILATUS 6M
Resolution range (Å)	44.51 - 1.36 (1.44 - 1.36)	44.50 - 1.55 (1.64 - 1.55)	44.33 - 1.42 (1.51 - 1.42)
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell parameters (a, b, c Å)	89.2, 64.3, 70.9	89.1, 64.8, 70.5	88.8, 63.7, 70.4
Unit cell parameters (α, β, γ°)	90.0, 93.1, 90.0	90.0, 93.2, 90.0	90.0, 92.6, 90.0
Matthews coefficient ^b (Å ³ /Da)	2.5	2.5	2.4
Solvent Content b (%)	50.1	51.0	48.8
Total reflections	239209 (38188)	219498 (34906)	276359 (43372)
Unique reflections	81480 (12780)	58079 (9314)	73712 (11753)
Multiplicity	2.9	3.8	3.8
Completeness (%)	95.0 (92.5)	99.7 (99.4)	99.4 (98.4)
Mean I/sigma(I)	11.6 (2.3)	15.4 (2.0)	15.6 (2.0)
Wilson B-factor (Å ²)	15.4	18.0	15.7
R-sym ^c (%)	5.0 (44.5)	5.2 (55.6)	4.6 (55.1)
R-meas (%)	6.1 (54.2)	6.0 (64.7)	5.4 (64.3)
CC1/2	99.8 (81.6)	99.9 (77.4)	99.9 (74.1)
B) Refinement			
Resolution range (Å)	44.51 - 1.36	44.50 - 1.55	44.33 - 1.42
Total reflections used in refinement	81466	58069	73705
Reflections used for R-work	77393	55166	70020
Reflections used for R-free	4073	2903	3685
Final R value for R-work (%)	13.5	14.7	13.4
Final R value for R-free (%)	15.8	17.7	15.5
No. of Protein residues	371	364	373
No. of Water molecules	269	322	359
No. of Ligand molecules	1	1	1
No. of Other ligand molecules	4	4	6
RMSD bond lengths (Å)	0.007	0.008	0.007
RMSD bond angles (°)	0.9	0.9	0.9
Ramachandran plot ^d			
Residues in most favored regions (%)	94.6	94.2	94
Residues in additionally allowed regions	5	5.5	5.6
(%)	5	5.5	5.0
Residues in generously allowed regions (%)	0.3	0.3	0.3
Residues in disallowed regions (%)	none	none	none
Average B-factor all atoms ^e (Å ²)	20.5	24.6	22.1
Protein main chain	17.9	22.1	19.7
Protein side chain	20.9	25.9	22.5
Protein all atoms	19.5	24.1	21.1
Ligand	27.1	57.5	26.8
Water molecules	29.8	30.4	31.2
Other ligands ^f	44	36.4	46
Number of TLS groups	N/A	N/A	N/A

Table 3: Data collection and refinement statistics for TGT structures.

PDB Code	Unpublished	Unpublished	6RKT	6RKQ
Ligand	J19	J33	4.1	4.2
A) Data Collection and				
Processing ^a				
Wavelength	0.895	1.033200	0.9184	0.9184
Beamline	BESSY BL 14.3	ESRF BM30A	BESSY BL 14.2	BESSY BL 14.2
Detector	Rayonix MX225 CCD	ADSC Q315r CCD	PILATUS3S 2M	PILATUS3S 2M
Resolution range (Å)	44.63 -1.37 (1.46-1.37)	44.38 -1.25 (1.33-1.25)	44.74 -1.75 (1.85-1.75)	45.19 -1.67 (1.77-1.66
Space group	C 1 2 1	C 1 2 1	C121	C121
Unit cell parameters (a, b, c Å)	89.7, 64.8, 70.5	90.0, 65.0, 70.0	89.6, 64.6, 71.2	90.5, 64.8, 71.5
Unit cell parameters $(\alpha, \beta, \gamma^{\circ})$	90.0, 95.6, 90.0	90.0, 93.0, 90.0	90.0, 93.3, 90.0	90.0, 93.1, 90.0
Matthews coefficient ^b (Å ³ /Da)	2.5	2.5	2.5	2.6
Solvent Content ^b (%)	50.3	50.1	51.4	51.9
Total reflections	248324 (34930)	361277 (52153)	137815 (19404)	180914 (28178)
Unique reflections	82394 (12435)	108083 (16922)	40353 (6118)	47621 (7421)
Multiplicity	3.0	3.3	3.42	3.78
Completeness (%)	98.1 (91.7)	98.5 (95.8)	97.7 (92.2)	98.3 (95.7)
Mean I/sigma(I)	13.2 (2.6)	10.1 (2.3)	17.6 (2.0)	14.1 (2.1)
Wilson B-factor (Å ²)	11.0	14.2	29.6	37.2
R-sym ^c (%)	6.0 (47.7)	6.4 (49.0)	3.5 (50.4)	4.7 (46.3)
R-meas (%)	7.2 (57.8)	7.6 (59.3)	4.2 (60.4)	5.5 (53.8)
CC1/2	99.8 (79.0)	99.5 (75.4)	99.9 (86.7)	99.8 (88.4)
B) Refinement		(,	,	(,
Resolution range (Å)	44.63 – 1.37	44.38 – 1.25	38.90 - 1.75	43.03 - 1.66
Total reflections used in	1.1100 2.07	1.1100 2.120	30.30 1.73	10100 2100
refinement	82364	108075	40307	47606
Reflections used for R-work	78246	102671	38292	45226
Reflections used for R-free	4118	5404	2015	2380
Final R value for R-work (%)	13.6	14.5	16.9	16.3
Final R value for R-free (%)	15.5	16.5	20.5	19.4
No. of Protein residues	371	368	366	369
No. of Water molecules	280	264	193	186
No. of Ligand molecules	1	1	1	1
No. of Other ligand molecules	2	7	3	3
				-
RMSD bond lengths (Å)	0.007	0.005	0.009	0.009
RMSD bond angles (°)	1.0	0.8	1	1
Ramachandran plot ^d				
Residues in most favored regions (%)	94.6	94.6	95.5	95.2
Residues in additionally allowed regions (%)	5.0	5.1	4.2	4.5
Residues in generously allowed regions (%)	0.3	0.3	0.3	0.3
Residues in disallowed regions (%)	none	none	none	none
Average B-factor all atoms ^e (Å ²)			42.3	37.1
Protein main chain	12.5	19.2	39.6	35.1
Protein side chain	14.5	21.8	44.6	38.5
Protein all atoms	13.6	20.5	42.2	36.9
Ligand	25.8	24.7	37.2	32.6
Water molecules	25.8	32.1	44.5	41.2
Other ligands ^f	26.1	42.1	55.1	53.7
Number of TLS groups	N/A	N/A	6	6

a) Statistics for the highest-resolution shell are shown in parenthesis, b) Calculated with $Matthews_coef$ program from CCP4 suite version 7.0.047 93 , c) Calculated by the equation R(I)sym=[$\Sigma h\Sigma i | Ii(h)-(I(h))|/\Sigma h\Sigma Iii(h)] \cdot 100$, where (I(h)) is the mean of the I(h) observation of reflection h. d) Calculated with $PROCHECK^{201}$, e) Calculated with $PROCHECK^{201$

Table 4: Data collection and refinement statistics for TGT structures. Copied from the dissertation of Dr. Frederik Ehrmann, 2016.⁹⁵

PDB Code	5J9M 5JT5		5J9N
Ligand	2.1 2.2		2.5
A) Data Collection and Processing ^a			
Wavelength	0.8944	0.91841	0.91841
Beamline	BESSY BL 14.3	BESSY BL 14.1	BESSY BL 14.1
Resolution range (Å)	45.21 - 1.33(1.41 - 1.33)	43.53 - 1.21 (1.28 - 1.21)	45.12 - 1.64 (1.74 - 1.64)
Space group	C 1 2 1	C121	C 1 2 1
Unit cell parameters (a, b, c Å)	91.0, 64.9, 70.7	90.3, 65.3, 70.6	90.4, 64.5, 71.3
Unit cell parameters (α, β, γ°)	90.0, 96.3, 90.0	90.0, 96.1, 90.0	90.0, 92.8, 90.0
Matthews coefficient ^b (Å ³ /Da)	2.4	2.4	2.4
Solvent Content b (%)	49.0	48.9	49.1
Unique reflections	93288 (14902)	122593 (19268)	49307 (7770)
Multiplicity	2.9	3.7	3.0
Completeness (%)	99.2 (98.4)	98.7 (96.4)	97.5 (95.2)
Mean I/sigma(I)	14.3 (2.3)	15.7 (3.3)	9.28 (2.0)
R-sym ^c (%)	4.5 (48.9)	4.1 (38.4)	6.3 (34.2)
B) Refinement			
Resolution range (Å)	45.21 - 1.33	39.80 - 1.21	41.68 - 1.64
Total reflections used in refinement	93249	122586	49285
Final R value for R-work (%)	12.9	13.5	21.3
Final R value for R-free (%)	15.7	15.7	23.7
No. of Protein atoms	3002	2988	2823
No. of Water molecules	400	430	208
No. of Ligand atoms	14	15	15
RMSD bond lengths (Å)	0.005	0.005	0.006
RMSD bond angles (°)	0.8	0.8	0.8
Ramachandran plot ^d			
Residues in most favored regions (%)	94.9	94.9	95.5
Residues in additionally allowed regions (%)	4.8	4.8	4.2
Residues in generously allowed regions (%)	0.3	0.3	0.3
Average B-factor ^e (Ų)			
Protein atoms	17.2	17.1	30.5
Water molecules	34.7	33.6	37.6
Ligand atoms	25.3	28.8	42.7

Table 5: Data collection and refinement statistics for TGT structures. Copied from the dissertation of Dr. Frederik Ehrmann, 2016 95

PDB Code	5JT6		5JT7
Ligand	2.6	3.1	3.2
A) Data Collection and Processing ^a			
Wavelength	0.91841	0.91841	0.91841
Beamline	BESSY BL 14.1	BESSY BL 14.1	BESSY BL 14.1
Resolution range (Å)	45.11 - 1.54 (1.63 - 1.54)	19.45 - 1.41 (1.49 - 1.41)	42.80 - 1.70 (1.80 - 1.70)
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell parameters (a, b, c Å)	90.3, 64.2, 71.0	90.1, 63.8, 71.0	90.3, 64.2, 71.2
Unit cell parameters (α, β, γ°)	90.0, 92.8, 90.0	90.0, 92.9, 90.0	90.0, 93.2, 90.0
Matthews coefficient ^b (Å ³ /Da)	2.4	2.4	2.4
Solvent Content b (%)	48.6	48.1	48.7
Unique reflections	59224	76131	44054
Multiplicity	2.9	3.8	2.5
Completeness (%)	98.4 (97.3)	97.9 (95.6)	98.1 (98.4)
Mean I/sigma(I)	18.8 (2.1)	16.9 (2.3)	15.8 (2.0)
R-sym ^c (%)	2.8 (49.3)	3.5 (49.3)	3.1 (40.6)
B) Refinement			
Resolution range (Å)	41.11 - 1.54	19.45 - 1.41	42.80 - 1.70
Total reflections used in refinement	59215	76103	44050
Final R value for R-work (%)	16.3	14.0	17.2
Final R value for R-free (%)	18.1	16.9	20.5
No. of Protein atoms	2889	2978	2853
No. of Water molecules	207	294	209
No. of Ligand atoms	15	15	15
RMSD bond lengths (Å)	0.006	0.006	0.006
RMSD bond angles (°)	0.8	0.8	0.8
Ramachandran plot ^d			
Residues in most favored regions (%)	95.9	94.3	95.8
Residues in additionally allowed regions (%)	3.8	5.4	3.9
Residues in generously allowed regions (%)	0.3	0.3	0.3
Average B-factor ^e (Å ²)			
Protein atoms	31.6	26.8	35.9
Water molecules	37.3	40.3	40.6
Ligand atoms	41.2	30.9	41.6

a) Values in parentheses are statistics for the highest resolution shell. b) Calculated with MATTPROB 203 c) R(I)sym=[Σ h Σ i|Ii(h)-(I(h))-(Σ h Σ IIi(h)] · 100, where (I(h)) is the mean of the I(h) observation of reflection h. d) Determined with PROCHECK 201 . e) Calculated with MOLEMAN 202

Table 6: Data collection and refinement statistics for E1W PEX14 structures.

PDB Code	6S6R	6S7M	Not published
Ligand	F5	F6	Apo - Crystal form 1
A) Data Collection and Processing ^a			
Wavelength	0.9184	0.9184	0.9184
Beamline	BESSY BL 14.1	BESSY BL 14.1	BESSY BL 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M
Resolution range (Å)	43.31 - 1.58 (1.68 - 1.58)	39.89 - 1.76 (1.86 –1.76)	43.18 - 1.56 (1.65 - 1.56)
Space group	P4(1)2(1)2	P2(1)	P4(1)2(1)2
Unit cell parameters (a, b, c Å)	43.3, 43.3, 80.9	79.8, 39.7, 81.1	43.18, 43.18, 80.96
Unit cell parameters (α , β , γ °)	90.0, 90.0, 90.0	90.0, 90.1, 90.0	90.0, 90.0, 90.0
Matthews coefficient ^b Å ³ /Da)	2.4	2.1	2.4
Solvent Content b (%)	49.2	42.5	48.9
Total reflections	90149 (14571)	185815 (30160)	95671 (15433)
Unique reflections	10965 (1705)	49925 (7928)	11386 (1761)
Multiplicity	8.2	3.7	8.4
Completeness (%)	98.4 (97.4)	97.6 (96.7)	98.9 (97.9)
Mean I/sigma(I)	16.9 (4.2)	14.7 (2.5)	25.9 (3.9)
Wilson B-factor (Ų)	18.9	23.1	20.6
R-sym ^c (%)	6.9 (49.7)	5.2 (50.2)	4.3 (55.4)
R-meas (%)	7.4 (52.9)	6.1 (58.4)	4.6 (58.8)
CC1/2	99.8 (93.9)	99.9 (88.0)	100 (96.8)
B) Refinement	, ,		• ,
Resolution range (Å)	38.18 – 1.58	39.89 – 1.76	38.10 – 1.56
Total reflections used in refinement	10960	49847	11365
Reflections used for R-work	10412	47355	10797
Reflections used for R-free	548	2492	568
Final R value for R-work (%)	18.7	23.1	20.5
Final R value for R-free (%)	23.5	26.7	26.5
No. of Protein residues	60	67/67/67/61	60
(Protein molecules A/B/C/D/E/F/G/H)	69	60/66/68/68	69
No. of Water molecules	42	151	42
No. of Ligand molecules	1	6	none
No. of Other ligand molecules	5	2	2
RMSD bond lengths (Å)	0.006	0.007	0.009
RMSD bond angles (°)	0.8	0.7	0.9
Ramachandran plot ^d			
(Protein molecules A/B/C/D/E/F/G/H)			
Desidence in secret forward regions (00)	00.0	96.7/96.8/96.7/96.8	00.0
Residues in most favored regions (%)	96.8	96.7/98.2/96.3/95.0	96.8
Residues in additionally allowed regions	3.2	3.3/3.2/3.3/3.2	2.7
(%)	3.2	3.3/1.8/3.7/5.0	3.2
Residues in generously allowed regions (%)	none	none	none
Residues in disallowed regions (%)	none	none	none
Average B-factor all atoms ^e (Å ²)			
(Protein molecules A/B/C/D/E/F/G/H)			
	25.5	23.8/17.9/19.1/20.1	25.1
Protein main chain	20.9	28.3/37.5/36.5/24.1	25.1
Protein side chain	23.8	25.7/21.8/22.1/22.8 30.6/37.6/37.5/27.0	30.6
		24.7/19.8/20.5/21.4	
Protein all atoms	22.3	29.4/37.5/37.0/25.4	28.0
Ligand	26.1	21.3	N/A
Water molecules	31.0	28.3	33.2
Other ligands ^f	42.4	39.8	36.8
Number of TLS groups	N/A	30	6

Table 7: Data collection and refinement statistics for E1W PEX14 structures.

PDB Code	Not published ^g	Not published ^g	Not published ^g
Ligand	Apo – Crystal form 3	Apo – Crystal form 4	Apo - Crystal form 5
A) Data Collection and Processing ^a			
Wavelength	1.541790	0.9184	0.873127
Beamline	Inhouse X-ray Source	BESSY BL 14.1	ESRF ID23-2
Detector	MAR345	PILATUS 6M	DECTRIS PILATUS3 X 2M
Resolution range (Å)	42.61 -2.59 (2.75-2.60)	40.96 - 2.30 (2.43–2.29)	41.66 - 1.60 (1.70 - 1.60)
Space group	P4(3)2(1)2	P2(1)	P4(1)2(1)2
Unit cell parameters (a, b, c Å)	42.6, 42.6, 81.1	39.0, 82.1, 39.4	41.7, 41.7, 81.6
Unit cell parameters (α, β, γ°)	90.0, 90.0, 90.0	90.0, 92.9, 90.0	90.0, 90.0, 90.0
Matthews coefficient ^b Å ³ /Da)	2.4	2.0	2.3
Solvent Content ^b (%)	47.6	38.2	45.6
Total reflections	22462 (3529)	37760 (6091)	121577 (19346)
Unique reflections	2575 (399)	10961 (1755)	10054 (1573)
Multiplicity	8.7	3.4	12.1
Completeness (%)	99.7 (99.8)	98.6 (98.2)	99.9 (99.9)
Mean I/sigma(I)	13.7 (3.5)	7.9 (2.5)	25.1 (4.9)
Wilson B-factor (Ų)	41.2	25.5	23.1
R-sym ^c (%)	11.8 (66.2)	12.3 (50.9)	4.9 (44.0)
R-meas (%)	12.6 (70.3)	14.6 (60.2)	5.2 (45.9)
CC1/2	99.8 (99.0)	99.5 (82.2)	100 (98.9)
B) Refinement			
Resolution range (Å)	37.72 – 2.60	39.35 – 2.23	37.11 – 1.60
Total reflections used in refinement	2551	11989	10022
Reflections used for R-work	2423	599	501
Reflections used for R-free	128	11390	9521
Final R value for R-work (%)	24.2	21.6	27.0
Final R value for R-free (%)	34.1	29.7	30.8
No. of Protein residues	69	69/69/69/69	69
(Protein molecules A/B/C/D)			
No. of Water molecules	none added	none added	none added
No. of Ligand molecules	none	none	none
No. of Other ligand molecules	none added	none added	none added
RMSD bond lengths (Å)	0.010	0.008	0.006
RMSD bond angles (°)	1.2	1.0	0.9
Ramachandran plot ^d			
(Protein molecules A/B/C/D)			
Residues in most favored regions (%)	90.5	90.5/92.1/90.5/93.7	96.8
Residues in additionally allowed regions (%)	9.5	7.9/4.8/7.9/6.3	3.2
Residues in generously allowed regions (%)	none	none/3.2/1.6/0	none
Residues in disallowed regions (%)	none	1.6/0/0/0	none
Average B-factor all atoms ^e (Å ²)			
(Protein molecules A/B/C/D)			
Protein main chain	54.1	34.3/37.3/31.5/37.7	30.2
Protein side chain	56.1	38.1/41.8/36.0/41.9	35.8
Protein all atoms	55.1	36.2/39.5/33.7/39.7	32.9
Ligand	N/A	N/A	N/A
Water molecules	N/A	N/A	N/A
Other ligands ^f	N/A	N/A	N/A
Number of TLS groups	6	6	6
realiser of the groups	1		1

Table 8: Data collection and refinement statistics for wild-type PEX14 structures.

PDB Code	6S9Y	Not published	Not published
Ligand	Apo – Crystal form 7	Apo – Crystal form 8	Compound 1
A) Data Collection and Processing ^a			
Wavelength	0.9184	0.9184	0.9184
Beamline	BESSY BL 14.1	BESSY BL 14.1	BESSY BL 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M
Resolution range (Å)	40.63 - 1.64 (1.74 - 1.64)	39.20 - 1.93 (2.04 -1.93)	40.43 - 1.83 (1.94 – 1.83)
Space group	P2(1)	C222(1)	C222(1)
Unit cell parameters (a, b, c Å)	38.9, 81.3, 39.3	54.1, 56.8, 81.0	53.9, 56.7, 80.9
Unit cell parameters (α , β , γ °)	90.0, 92.8, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Matthews coefficient b (Å3/Da)	2.2	2.3	2.2
Solvent Content b (%)	43.8	45.8	44.1
Total reflections	95230 (10495)	58085 (9425)	50057 (7949)
Unique reflections	28782 (3967)	9461 (1463)	11110 (1753)
Multiplicity	3.3	6.1	4.5
Completeness (%)	96.2 (82.6)	96.7 (93.6)	98.9 (97.8)
Mean I/sigma(I)	12.4 (2.3)	19.0 (4.5)	9.2 (2.4)
Wilson B-factor (Ų)	16.6 ^h	23.7	21.4
R-sym ^c (%)	6.1 (37.4)	6.5 (34.1)	10.8 (49.6)
R-meas (%)	7.3 (46.8)	7.1 (37.1)	12.2 (56.0)
CC1/2	99.8 (86.2)	99.9 (95.3)	99.6 (84.5)
B) Refinement			
Resolution range (Å)	39.24 -1.64	39.20 - 1.93	40.43 – 1.83
Total reflections used in refinement	28758	9445	11103
Reflections used for R-work	27320	8982	10548
Reflections used for R-free	1438	463	555
Final R value for R-work (%)	19.7	25.8	27.9
Final R value for R-free (%)	22.9	31.0	33.0
No. of Protein residues	62/62/62/62	c2/c2	62/62
(Protein molecules A/B/C/D)	62/63/62/63	62/62	62/62
No. of Water molecules	88	15	17
No. of Ligand molecules	none	none	2
No. of Other ligand molecules	2	none	none
RMSD bond lengths (Å)	0.005	0.007	0.007
RMSD bond angles (°)	0.7	0.8	0.8
Ramachandran plot ^d			
(Protein molecules A/B/C/D)			
Residues in most favored regions (%)	98.2	98.2/98.0	98.2
Residues in additionally allowed regions (%)	1.8	1.8/2.0	1.8
Residues in generously allowed regions (%)	none	none	none
Residues in disallowed regions (%)	none	none	none
Average B-factor all atoms ^e (Å ²)	Hone	Hone	Hone
(Protein molecules A/B/C/D)			
Protein main chain	16.5/17.1/21.5 /26.0	23.8/39.9	20.9/38.2
Protein side chain	20.0/19.3/24.1/27.9	25.6/39.6	21.7/37.1
Protein all atoms	18.2/18.1/22.7/26.9	24.6/39.8	21.3/37.7
Ligand	N/A	N/A	23.2
Water molecules	31.1	35.1	29.2
Other ligands ^f			
-	43.7	N/A	N/A
Number of TLS groups a) Statistics for the highest-resolution shell are sh	19	9	6

a) Statistics for the highest-resolution shell are shown in parenthesis, b) calculated with $Matthews_coef$ program from CCP4 suite version 7.0.047⁹³, c) calculated by the equation: R(I)sym=[$\Sigma h\Sigma i|Ii(h)-(I(h))|/\Sigma h\Sigma i|Ii(h)]$ · 100, where $\langle I(h)\rangle$ is the mean of the I(h) observation of reflection h. d) calculated with $PROCHECK^{201}$, e) calculated with $MOLEMAN^{202}$, f) other ligands include DMSO, PEG, Glycerol, SO₄, g) statistics after one cycle of TLS refinement as part of an auto-refinement pipeline. No waters or ligand molecules have been added. h) calculated by Xtriage from the Protein Data Bank ⁸⁸

Table 9: Data collection and refinement statistics for EP structures.

PDB Code	Not published ^g Not published ^g		Not published ^g
Ligand	F268 and ritonavir F109 and ritonavir		F236 and ritonavir
A) Data Collection and Processing ^a			
Wavelength	0.9184	0.9184	0.9184
Beamline	BESSY BL 14.1	BESSY BL 14.1	BESSY BL 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M
Resolution range (Å)	42.87 - 1.15 (1.22– 1.15)	42.87 - 1.15 (1.22 – 1.15)	49.75 - 1.38 (1.46 – 1.38)
Space group	P2 ₁	P2 ₁	P2 ₁
Unit cell parameters (a, b, c Å)	45.3, 72.9, 52.4	45.4, 73.0, 52.5	45.4, 73.0, 52.7
Unit cell parameters (α, β, γ°)	90.0, 108.9, 90.0	90.0, 109.1, 90.0	90.0, 109.3, 90.0
Matthews coefficient ^b (Å ³ /Da)	2.4	2.4	2.4
Solvent Content ^b (%)	49.4	49.6	49.7
Total reflections	419839 (64346)	420086 (64230)	235137 (34332)
Unique reflections	110654 (17454)	111322 (17547)	64037 (9764)
Multiplicity	3.8	3.8	3.7
Completeness (%)	96.8 (94.6)	97.1 (94.9)	95.1 (89.9)
Mean I/sigma(I)	17.8 (4.2)	12.0 (2.1)	14.0 (2.6)
Wilson B-factor (Ų)	10.2	10.7	11.9
R-sym ^c (%)	3.8 (27.4)	6.0 (53.5)	9.5 (53.5)
R-meas (%)	4.4 (32.0)	6.9 (62.6)	11.1 (63.1)
CC1/2	99.9 (93.2)	99.9 (84.1)	99.6 (85.9)
B) Refinement			
Resolution range (Å)	42.87 – 1.15	42.87 – 1.15	49.75 – 1.38
Total reflections used in refinement	110651	111298	64016
Reflections used for R-work	105118	105733	60815
Reflections used for R-free	5533	5565	3201
Final R value for R-work (%)	14.5	15.1	19.1
Final R value for R-free (%)	16.1	16.6	21.4
No. of Protein residues	330	330	330
No. of Water molecules	233	232	173
No. of Ligand molecules	none added	none added	none added
No. of Other ligand molecules	none added	none added	none added
RMSD bond lengths (Å)	0.006	0.007	0.008
RMSD bond angles (°)	1.2	1.2	1.2
Ramachandran plot ^d			
Residues in most favored regions (%)	93.5	94.2	94.2
Residues in additionally allowed regions	6.5	5.8	5.8
(%)	0.5	5.6	5.6
Residues in generously allowed regions	none	none	none
(%)			
Residues in disallowed regions (%)	none	none	none
Average B-factor all atoms ^e (Å ²)			_
Protein main chain	10.6	11.7	14.3
Protein side chain	12.5	13.6	16.0
Protein all atoms	11.6 12.7		15.2
Ligand	N/A N/A		N/A
Water molecules	23.5	24.2	22.4
Other ligands f	N/A	N/A	N/A
Number of TLS groups	N/A	N/A	N/A

Table 10: Data Collection and Refinement Statistics for EP Structures.

PDB Code	Not published ^g	Not published h	Not published h
Ligand	F267 and ritonavir	F267 and ritonavir F216 and ritonavir	
A) Data Collection and Processing ^a			
Wavelength	0.9184	0.9184	0.9184
Beamline	BESSY BL 14.1	BESSY BL 14.1	BESSY BL 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M
Resolution range (Å)	42.86 -1.08 (1.14-1.08)	42.78 - 1.05 (1.11 – 1.05)	42.92 - 1.38 (1.46 – 1.38)
Space group	P2 ₁	P2 ₁	P2 ₁
Unit cell parameters (a, b, c Å)	45.4, 73.0, 52.7	45.3, 72.9, 52.6	45.3, 73.1, 52.6
Unit cell parameters (α, β, γ°)	90.0, 109.2, 90.0	90.0, 109.2, 90.0	90.0, 108.7, 90.0
Matthews coefficient b (Å3/Da)	2.4	2.4	2.4
Solvent Content b (%)	49.7	49.4	49.7
Total reflections	451438 (71810)	548960 (82883)	248840 (39066)
Unique reflections	133364 (21059)	146027 (22766)	65817 (10466)
Multiplicity	3.4	3.8	3.8
Completeness (%)	96.0 (94.0)	96.9 (93.7)	98.3 (97.0)
Mean I/sigma(I)	17.0 (2.3)	14.6 (2.8)	10.9 (2.2)
Wilson B-factor (Ų)	10.6	10.2	12.2
R-sym ^c (%)	3.5 (49.2)	4.5 (43.1)	7.3 (52.4)
R-meas (%)	4.2 (58.2)	5.2 (50.3)	8.5 (61.1)
CC1/2	100.0 (84.7)	99.9 (87.3)	99.7 (80.2)
B) Refinement	, ,	, ,	,
Resolution range (Å)	41.13 – 1.08	41.03 – 1.05	42.92 – 1.38
Total reflections used in refinement	133483	146015	65809
Reflections used for R-work	126809	138714	62519
Reflections used for R-free	6674	7301	3290
Final R value for R-work (%)	15.2	21.5	22.0
Final R value for R-free (%)	16.9	21.8	24.6
No. of Protein residues	330	330	330
No. of Water molecules	244	none added	none added
No. of Ligand molecules	none added	2	2
No. of Other ligand molecules	none added	none added	none added
RMSD bond lengths (Å)	0.006	0.003	0.005
RMSD bond angles (°)	1.2	0.8	0.8
Ramachandran plot ^d			
Residues in most favored regions (%)	93.5	93.5	93.9
Residues in additionally allowed regions			
(%)	6.5	6.5	6.1
Residues in generously allowed regions (%)	none	none	none
Residues in disallowed regions (%)	none	none	none
Average B-factor all atoms e (Ų)			
Protein main chain	11.4	11.1	13.1
Protein side chain	13.2	12.7	14.6
Protein all atoms	12.3	11.8	13.8
Ligand	N/A	15.7	19.1
Water molecules	23.9	N/A	N/A
Other ligands ^f	N/A	N/A	N/A
Number of TLS groups	N/A	6	6
Trainiber of TEO Broaps	1 11/11	1	ı

Table 11: Data Collection and Refinement Statistics for EP Structures.

PDB Code	Not published h	Not published h	Not published h
Ligand	F063 and ritonavir	F290 and ritonavir	F306 and ritonavir
A) Data Collection and Processing ^a			
Wavelength	0.9184	0.9184	0.9184
Beamline	BESSY BL 14.1	BESSY BL 14.1	BESSY BL 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M
Resolution range (Å)	42.72 -1.59 (1.68-1.59)	42.74 - 1.25 (1.32 - 1.25)	42.74 - 1.35 (1.43 – 1.35)
Space group	P2 ₁	P2 ₁	P2 ₁
Unit cell parameters (a, b, c Å)	45.4, 73.2, 52.7	45.3, 73.2, 52.6	45.4, 73.0, 52.8
Unit cell parameters (α, β, γ)	90.0, 109.6, 90.0	90.0, 109.4, 90.0	90.0, 109.5, 90.0
Matthews coefficient b (Å3/Da)	2.4	2.4	2.4
Solvent Content b (%)	49.6	49.6	49.6
Total reflections	159269 (24928)	326127 (51548)	265100 (40471)
Unique reflections	43134 (6881)	88133 (14142)	68512 (10690)
Multiplicity	3.7	3.7	3.9
Completeness (%)	98.2 (97.0)	98.1 (97.7)	96.0 (92.8)
Mean I/sigma(I)	16.3 (2.5)	17.4 (2.5)	13.5 (2.7)
Wilson B-factor (Å ²)	-	-	12.0
R-sym ^c (%)	6.0 (52.1)	4.4 (55.2)	6.0 (45.9)
R-meas (%)	7.0 (61.1)	5.2 (64.7)	7.0 (53.3)
CC1/2	99.9 (76.5)	99.9 (76.4)	99.8 (84.3)
B) Refinement			
Resolution range (Å)	42.72 – 1.59	39.50 - 1.25	42.74 – 1.35
Total reflections used in refinement	43133	88127	68503
Reflections used for R-work	40976	83721	65078
Reflections used for R-free	2157	4406	3425
Final R value for R-work (%)	22.3	22.0	22.6
Final R value for R-free (%)	23.7	23.3	23.2
No. of Protein residues	330	330	330
No. of Water molecules	none added	none added	none added
No. of Ligand molecules	1	1	1
No. of Other ligand molecules	none added	none added	none added
RMSD bond lengths (Å)	0.005	0.004	0.004
RMSD bond angles (°)	0.8	0.8	0.8
Ramachandran plot ^d			
Residues in most favored regions (%)	93.5	93.9	93.9
Residues in additionally allowed regions (%)	6.5	6.1	6.1
Residues in generously allowed regions (%)	none	none	none
Residues in disallowed regions (%)	none	none	none
Average B-factor all atoms ^e (Å ²)			
Protein main chain	15.5	13.2	13.0
Protein side chain	16.9	14.9	14.5
Protein all atoms	16.1	14.0	13.7
Ligand	27.4	19.8	20.1
Water molecules	N/A	N/A	N/A
Other ligands ^f	N/A	N/A	N/A
Number of TLS groups	6	6	6

a) Statistics for the highest-resolution shell are shown in parenthesis, b) calculated with *Matthews_coef* program from *CCP4* suite version 7.0.047⁹³, c) calculated by the equation: (SUM(ABS(I(h,i)-I(h))))/(SUM(I(h,i))) d) calculated with *PROCHECK*²⁰¹, e) calculated with *MOLEMAN*²⁰², f) other ligands include DMSO, PEG, Glycerol, SO₄, ACT. g) statistics after 11 cycles of refinement as part of an auto-refinement pipeline including anisotropic refinement of protein and water molecules, and addition of hydrogen. No ligand molecules have been added. h) only the ligand of interest was added after TLS refinement. No water molecules or other ligand molecules have been added.

Table 12: Data collection and refinement statistics for EP structures. Copied from the dissertation of Dr. Nedyalka Radeva, 2016^{33}

Fragment number	Cocktail A:	Cocktail B:	Cocktail C:	Cocktail D:
PDB code	F63, F267, F291 5MB0	F224 and F236 5MB7	F171 and F103 5MB5	F308 and F333 5MB6
Data collection and	PINIRO	5IVIB/	21/182	PINIRE
processing Wavelength (Å)	0.01941	0.01941	0.01941	0.91841
Beamline	0.91841 BESSY BL14.1	0.91841 BESSY BL14.2	0.91841 BESSY BL14.1	BESSY BL14.2
Detector	PILATUS 6M	MX-225 CCD	PILATUS 6M	MX-225 CCD
	P2 ₁	P2 ₁	P1LATO3 61VI	P2 ₁
Space group Unit-cell parameters a	PZ ₁	PZ ₁	PZ ₁	PZ ₁
	45 4 72 1 52 5	45 4 72 0 52 7	45.4, 72.9, 52.6	45.4, 73.5, 53.2
a, b, c (Å)	45.4, 73.1, 52.5 90.0, 109.6, 90.0	45.4, 72.9, 52.7 90.0, 109.5, 90.0		
α, β, γ (°)	, ,	, , , , , , , , , , , , , , , , , , ,	90.0, 109.7, 90.0	90.0, 110.5, 90.0
Resolution range (Å)	42.8-1.15 (1.22-1.15) 9.5	30.0-1.30 (1.32-1.30)	42.7-0.98 (1.04-0.98)	30.0-1.20 (1.22-1.20)
Wilson B factor (Ų)		10.9	8.4	8.3
No. of unique reflections	113549 (17691)	78870 (3975)	179855 (27996)	95743 (3095)
Multiplicity	3.6 (3.2)	3.2 (3.1)	3.8 (3.5)	2.8 (1.8)
R _{sym} (%)	7.4 (46.6)	5.2 (52.0)	6.6 (45.0)	5.4 (23.8)
Completeness (%)	98.8 (95.2)	99.9 (99.3)	97.4 (94.0)	94.2 (61.2)
<i σ(i)=""></i>	9.5 (2.8)	22.3 (2.0)	10.2 (2.2)	21.7 (3.8)
Refinement	42.0.4.45	10.0 1.20	26.5.000	27.0 4.20
Resolution range	42.8 - 1.15	18.9 - 1.30	36.5 – 0.98	27.9 – 1.20
No. of reflections used for	113540 (5677)	78840 (3850)	179793 (8989)	95693 (4629)
refinement (work (free))	45.0	12.5	44.7	12.2
R _{cryst} (%)	15.2	13.5	14.7	13.3
R _{free} (%)	17.0	15.8	16.2	14.9
No. of refined residues	330	330	330	330
No. of fragment atoms	59	13	13	-
No. of other ligand atoms	30	6	30	11
No. of water molecules	387	337	408	464
RMSD, bond length (Å)	0.006	0.006	0.005	0.006
RMSD, bond angles (°)	0.9	0.9	0.9	0.9
Ramachandran plot (%) b				
Most favoured	94.2	93.5	93.5	94.2
Additionally allowed	5.8	6.5	6.5	5.8
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values (Ų) c				
All protein atoms	11.2	12.5	9.6	9.6
Main chain	10.3	11.5	8.8	8.7
Side chain	12.1	13.4	10.3	10.4
Fragment atoms	13.4	37.6	10.1	-
Other ligand atoms ^d	21.6	17.5	19.8	12.6
Waters	28.7	27.2	27.2	26.9

a) Statistics for the highest-resolution shell are shown in parenthesis, b) Calculated by PROCHECK²⁰¹, c) Calculated by MOLEMAN²⁰², d) Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate.

Table 13: Data Collection and Refinement Statistics for EP Structures.

PDB Code	Not published	Not published	Not published
Ligand	Cocktail E	Cocktail F	Cocktail G
Ligand	F063, F052, F058	F207, F261	F216, F338
A) Data Collection and Processing ^a			
Wavelength	0.9184	0.9184	0.9184
Beamline	BESSY BL 14.1	BESSY BL 14.1	BESSY BL 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M
Resolution range (Å)	41.03 - 1.15 (1.22 – 1.15)	39.64 - 1.23 (1.30 – 1.23)	42.70 - 1.08 (1.14 – 1.08)
Space group	P2 ₁	P2 ₁	P2 ₁
Unit cell parameters (a, b, c Å)	45.4, 73.2, 52.7	45.5, 72.9, 52.7	45.4, 73.2, 52.7
Unit cell parameters (α, β, γ°)	90.0, 109.8, 90.0	90.0, 109.5, 90.0	90.0, 109.7, 90.0
Matthews coefficient ^b (Å ³ /Da)	2.4	2.4	2.4
Solvent Content ^b (%)	49.6	49.6	49.7
Total reflections	424924 (66924)	350510 (55884)	511677 (79502)
Unique reflections	114661 (18351)	93643 (15033)	138339 (22228)
Multiplicity	3.7	3.7	3.7
Completeness (%)	99.5 (98.5)	99.6 (99.1)	99.6 (99.2)
Mean I/sigma(I)	13.9 (2.3)	16.4 (2.5)	10.6 (2.2)
R-sym ^c (%)	5.0 (51.5)	4.4 (49.4)	6.4 (50.5)
R-meas (%)	5.9 (60.4)	5.1 (57.6)	7.4 (59.3)
CC1/2	99.9 (76.8)	99.9 (79.7)	99.8 (77.5)
B) Refinement			
Resolution range (Å)	41.03 – 1.15	39.64 – 1.23	
Total reflections used in refinement	114655	93641	138328
Reflections used for R-work	108922	88959	131412
Reflections used for R-free	5733	4682	6916
Final R value for R-work (%)	15.5	15.7	16.1
Final R value for R-free (%)	17.3	17.8	17.7
No. of Protein residues	330	330	330
No. of Water molecules	213	211	243
No. of Ligand molecules	1	2	2
No. of Other ligand molecules	none added	none added	none added
RMSD bond lengths (Å)	0.004	0.004	0.004
RMSD bond angles (°)	0.8	0.8	0.8
Ramachandran plot d			
Residues in most favored regions (%)	93.9	94.6	94.2
Residues in additionally allowed regions	6.4	F 4	F 0
(%)	6.1	5.4	5.8
Residues in generously allowed regions			
(%)	none	none	none
Residues in disallowed regions (%)	none	none	none
Average B-factor all atoms ^e (Å ²)			
Protein main chain	11.7	11.9	10.2
Protein side chain	13.0	13.1	11.4
Protein all atoms	12.3	12.4	10.7
Ligand	18.0	21.0	16.7
Water molecules	23.0	20.6	22.2
Other ligands ^f	N/A	N/A	N/A
Number of TLS groups	N/A	N/A	N/A

Table 14: Data Collection and Refinement Statistics for EP Structures.

PDB Code	Not published	Not published	Not published
Ligand	Cocktail H	Cocktail I	Cocktail J
Ligaliu	F218, F224	F227, F274	F041, F051
A) Data Collection and Processing ^a			
Wavelength	0.9184	0.9184	0.9184
Beamline	BESSY BL 14.1	BESSY BL 14.1	BESSY BL 14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M
Resolution range (Å)	42.97- 1.34 (1.42 - 1.34)	36.87 - 1.27 (1.35 - 1.27)	41.14 – 1.12 (1.19 - 1.12)
Space group	P2 ₁	P2 ₁	P2 ₁
Unit cell parameters (a, b, c Å)	45.5, 73.2, 52.9	45.3, 73.3, 52.7	73.2, 52.8, 90.0
Unit cell parameters (α, β, γ°)	45.5, 73.2, 52.9	90.0, 109.7, 90.0	90.0, 109.6, 90.0
Matthews coefficient ^b (Å ³ /Da)	2.5	2.4	2.5
Solvent Content ^b (%)	50.2	49.7	49.8
Total reflections	275538 (40917)	319269 (50025)	450342 (71003)
Unique reflections	73346 (11686)	85259 (13719)	123825 (19873)
Multiplicity	3.8	3.7	3.6
Completeness (%)	99.5 (98.5)	99.7 (99.5)	99.1 (98.7)
Mean I/sigma(I)	13.1 (2.3)	14.1 (2.4)	12.4 (2.4)
R-sym ^c (%)	5.5 (50.4)	6.0 (51.3)	5.9 (49.4)
R-meas (%)	6.4 (59.5)	7.0 (60.1)	6.9 (58.0)
CC1/2	99.8 (78.2)	99.9 (79.8)	99.9 (78.0)
B) Refinement	, ,	, ,	
Resolution range (Å)	22.38 – 1.34	28.03 – 1.27	24.12 – 1.12
Total reflections used in refinement	73334	85254	123724
Reflections used for R-work	69667	80991	117538
Reflections used for R-free	3667	4263	6186
Final R value for R-work (%)	17.7	16.1	18.2
Final R value for R-free (%)	19.2	18.1	19.5
No. of Protein residues	330	330	330
No. of Water molecules	166	226	219
No. of Ligand molecules	4	2	2
No. of Other ligand molecules	none added	none added	none added
RMSD bond lengths (Å)	0.004	0.004	0.004
RMSD bond angles (°)	0.8	0.8	0.8
Ramachandran plot d			
Residues in most favored regions (%)	94.2	94.2	93.9
Residues in additionally allowed regions			
(%)	5.8	5.8	6.1
Residues in generously allowed regions (%)	none	none	none
Residues in disallowed regions (%)	none	none	none
Average B-factor all atoms e (Ų)	none	none	none
Protein main chain	15.3	10.5	10.8
Protein side chain	17.0	11.6	11.8
Protein all atoms	16.0	11.0	11.8
	23.7		11.3
Ligand Water molecules	23.7	15.9 19.7	17.5
Water molecules Other ligands ^f			
	N/A	N/A	N/A
Number of TLS groups	6	N/A	N/A

Table 15: Data Collection and Refinement Statistics for EP Structures.

PDB Code	Not published	Not published	Not published	Not published
Ligand	Cocktail K	Cocktail L	Cocktail M	Cocktail N
Ligand	F162, F236, F338	F063, F272	F063, F268	F207, F240
A) Data Collection and				
Processing ^a				
Wavelength	0.9184	0.9184	0.9184	0.9184
Beamline	BESSY BL 14.1	BESSY BL 14.2	BESSY BL 14.2	BESSY BL 14.2
Detector	PILATUS 6M	MX-225 CCD	MX-225 CCD	MX-225 CCD
Resolution range (Å)	42.65 - 1.37 (1.45-	42.68 – 1.18 (1.25-	42.72 -1.06 (1.06 -	42.77 – 1.28 (1.36 –
	1.37)	1.18)	1.12)	1.28)
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Unit cell parameters (a, b, c Å)	45.2. 73.2, 52.6	45.3, 73.0, 52.5	45.4, 72.8, 52.4	45.4, 73.0, 52.7
Unit cell parameters (α, β, γ°)	90.0, 109.4, 90.0	90.0, 109.8, 90.0	90.0, 109.8, 90.0	90.0, 109.8, 90.0
Matthews coefficient ^b (ų/Da)	2.4	2.4	2.4	2.5
Solvent Content b (%)	49.5	49.4	49.3	49.8
Total reflections	252016 (37139)	373081 (26284)	548543 (55318)	347394 (54718)
Unique reflections	68024 (10653)	98039 (10843)	138620 (17410)	83375 (13410)
Multiplicity	3.7	3.8	4.0	4.2
Completeness (%)	99.2 (96.9)	92.7 (63.6)	95.5 (74.5)	99.7 (99.4)
Mean I/sigma(I)	14.3 (2.6)	20.4 (3.0)	19.3 (3.7)	18.7 (3.3)
R-sym ^c (%)	5.3 (46.9)	3.8 (33.2)	4.0 (29.8)	5.1 (48.8)
R-meas (%)	6.2 (55.2)	4.4 (43.2)	4.7 (35.9)	5.9 (56.2)
CC1/2	99.9 (81.7)	99.9 (83.5)	99.9 (90.9)	99.9 (85.3)
B) Refinement				
Resolution range (Å)	42.65 – 1.37	25.46 – 1.18	40.90 – 1.06	42.77 – 1.28
Total reflections used in refinement	68019	98008	138586	83347
Reflections used for R- work	64618	93108	131657	79180
Reflections used for R- free	3401	4900	6929	4167
Final R value for R-work (%)	17.2	12.5	15.2	17.6
Final R value for R-free (%)	18.4	14.5	16.6	18.7
No. of Protein residues	330	330	330	330
No. of Water molecules	181	332	2593	240
No. of Ligand molecules	none	1	1	2
No. of Other ligand molecules	none added	6	none added	none added
RMSD bond lengths (Å)	0.006	0.008	0.004	0.005
RMSD bond angles (°)	1.1	1.3	0.8	0.8
Ramachandran plot ^d				
Residues in most favored regions (%)	94.2	94.2	93.5	94.6
Residues in additionally allowed regions (%)	5.8	5.8	6.5	5.4
Residues in generously allowed regions (%)	none	none	none	none
Residues in disallowed regions (%)	none	none	none	none
Average B-factor all atoms ^e (Å ²)				

Protein main chain	13.6	10.2	9.3	10.3
Protein side chain	15.3	12.0	10.4	11.9
Protein all atoms	14.3	11.1	9.8	11.0
Ligand	N/A	17.3	11.5	14.9
Water molecules	22.0	24.4	18.8	19.0
Other ligands ^f	N/A	19.0	N/A	N/A
Number of TLS groups	2	N/A	N/A	6

a) Statistics for the highest-resolution shell are shown in parenthesis, b) calculated with *Matthews_coef* program from *CCP4* suite version 7.0.047⁹³, c) calculated by the equation: (SUM(ABS(I(h,i)-I(h))))/(SUM(I(h,i))) d) calculated with *PROCHECK*²⁰¹, e) calculated with *MOLEMAN*²⁰², f) other ligands include DMSO, PEG, Glycerol, SO₄, ACT.

Table 16: Chemical Structures of the 96 Jena fragments. Marvin was used for displaying chemical structures, Marvin 6.3.1, 2014, ChemAxon (http://www.chemaxon.com) and the web server chemicalize.com.

Jena Plate ID	Fragment	Jena Plate ID	Fragment	Jena Plate ID	Fragment
1	H ₂ N N O	8	HN HN	15	NH NH
2	N H ₂ N	9	O HN N H	16	Br NH NH
3	NH NH CI	10	NH ₂	17	H ₂ N N
4		11	F N N N N N N N N N N N N N N N N N N N	18	TZ ZZ
5	NH ₂	12	ОН	19	NH O
6	HN	13	H ₂ N N	20	NH ₂
7	O NH	14	NH	21	HN NH ₂

22	NH F	30	NH ₂	38	HIN N N
23	N NH	31	O N N	39	NH NH
24		32	NH ₂	40	N HN N
25	N N N N N N N N N N N N N N N N N N N	33	H ₂ N NH	41	ОН
26	H ₂ N S N	34	NH ₂ N	42	NH NH
27	N NH ₂	35	NH NH	43	NH NH
28	NH ₂	36	HN CI	44	
29	H ₂ N NH	37	HN	45	NH ₂
46	HN O	54	N O O	62	O OCH ₃ NH ₂ · HCI

47	HO NH ₂	55	H ₃ C H ₂ N _{III} ,	63	O CH ₃ HN N N
48	HN O	56	O	64	HN N N N H
49	H ₂ N NH	57		65	O OH
50	N N N N N N N N N N N N N N N N N N N	₊ 58	N CH ₃	66	CI HN O
51	N NH	59	HCI NH	67	H_2N OH
52	NH NH	60	OH Cbz O	68	TZ ZI TZ ZI
53	N I I I I I I I I I I I I I I I I I I I	61	NH ₂	69	O N N O N O O O O O O O O O O O O O O O
70	N NH ₂ OH	78	H ₂ N → N → OH	86	OH H ₂ N
71	H ₃ C N N N N CH ₃		HN H ₂ O	87	H ₃ C ₂ NH ₂

72	H ₂ N H O OH NH ₂ OH	80	H ₂ N O CH ₃	88	HO N CH ₃
73	NH HN NH HH OO OH	81	CH ₃ OH O H ₃ C-N+ CH ₃	89	OH NH ₂
74	HO H. OH HO-CCH ₂ NH ₂	82	ОН ОН ОН	90	NO ₂
75	O ₂ N CF ₃ CH ₃	83	NH ₂	91	М О Н ₃ С О
76	HZ OH OH	84	HN	92	OH OH NO ₂
77	H ₂ N H NH ₂ OH OH	85	NH ₂	93	O ₂ N H O N O
94	0= s=0	95	H ₂ N O	96	NHOH

Table 17: Binding level in RU (response units) of Jena fragments to TGT screened by SPR. a

Jena Fragment ID	RU	Jena Fragment ID	RU
J50	45.6	J66	7.6
J36	21.3	J93	7.5
J92	21.1	J14	7.4
19	17.3	J25	7.4
J51	16.7	189	7.4
J55	15.3	J56	7.3
J24	14.7	J91	7.3
J7	14.3	J18	7.2
J29	13.7	J49	7.2
J79	13.5	J67	6.8
J32	11.7	J61	6.7
J47	11.6	J17	6.6
J84	11.6	J15	6.5
J62	10.6	J63	6.3
J19	10.3	180	6.3
J37	10.3	J58	6.2
J12	9.9	J64	6.2
J5	9.8	J88	6.2
J59	9.8	J52	6.1
J60	9.5	J85	6.1
J38	9.3	J87	6
J76	9.3	J30	5.7
J20	9.2	J69	5.7
J27	9.2	J96	5.6
J23	9.1	J71	5.4
J13	9	J72	5.4
J45	9	J73	5.2
J8	8.8	J68	5

J6	8.6	J77	4.9
J22	8.5	J78	4.9
J10	8.4	J95	4.8
J41	8.4	J39	4.5
J46	8.4	J86	4.3
J43	8.3	J82	4.2
J4	8.1	J70	4.1
J28	7.7	J81	4.1
J44	7.7	J42	1.8

a) Sorted from highest binding level to lowest.

 ${\it Table~18: Structures~of~fragments~not~included~in~SPR~binding~level~screen~to~TGT}.$

Jena Plate ID	Fragment	Jena Plate ID	Fragment
J1	H ₂ N N	J48	HN
J11	F F	J74	HO H. OH HO-CCH ₂ NH ₂
J16	Br O NH	J75	O_2N CF_3 CGH_3 CH_3
J21	HN NH ₂	J83	NH ₂
J26	H ₂ N S N	J90	OH OH
J33	H ₂ N NH F F	J35	THE
J53	O Links	J54	
J57			

Table 19: Structures of fragments not included in NMR fragment screen of TGT

Jena Plate ID	Fragment	Jena Plate ID	Fragment
J2	N H ₂ N	J64	H ₂ N N H
J4		J67	H ₂ N OH
J6	° I I	J69	H ₂ N NH HO NH
J13	H ₂ N N	J73	NH HN NH H OH
J14	NH NH	J77	H ₂ N H NH ₂ OH
J17	H ₂ N N	J79	HN H2O
J59	HCI ON NH	J82	OH OH
J35	N N N N N N N N N N N N N N N N N N N	J53	O I I I I I I I I I I I I I I I I I I I
J54	N O	J55	H ₂ N _{III} ,
J57			

Table 20: Potent inhibitors of TGT used in displacement and competitive binding experiments in NMR and SPR screens.

Compound No.	Chemical Structure	Binding Affinity (K _D nM)
Compound 1	HN HN NH2	68 ± 5
Compound 2	HIN NH NH ₂	7.7

Table 21: Synthesized ligands used to study the $preQ_1$ transient sub-pocket of TGT.

Ligand	Parent Scaffold	R	K _i (μM)	PDB code
1.1	R NH NH 16 NH2	Me	0.058 ± 0.036 ⁷⁶	4PUK ⁷⁴
1.2	R N N NH ₂	O N	0.006 ± 0.006 ⁷⁶	4PUJ ⁷⁴
2.1	H 4	Me	300 ± 37	5J9M
2.2	R NH ₂ NH ₂ NH ₂	$\langle \rangle$	544 ± 43	5JT5
2.3	H 4 N	Me	270 ± 50	
2.4	R N 6	0	264 ± 40	
2.5	U 4 N/NH ₂	Me	283 ± 40	5J9N
2.6	R N 6 H	0 N	282 ± 18	5JT6
3.1	H 4 5 NH	Me	19 ± 2	5J90
3.2	R N N NH ₂	0 N	58 ± 2	5JT7
4.1 ^a	R3 NH NH,		29 ± 21 (K _D)	6RKT
4.2 ^a	4.1 R3= H, 4.2 R3= ———————————————————————————————————	Me	8.9 ± 1.3 (K _D)	6RKQ
5.1	H 2 NH	Me	6.5 ± 2.9 ⁷⁷	3S1G ⁷⁷
5.2	R N 4 5	$\langle \rangle$	4.1 ⁷⁷	4Q4R ⁸²

Table 22: The various crystal forms E1W PEX14 variant was found to crystallize in.

Crystal Form	Unit Cell Dimensions						Space Group	Molecules per asymmetric unit	Ligand
1	Unit Cell Dimensions a b c α β γ 43.2 43.2 81.0 90.0 90.0 90.0						tetragonal P4 ₁ 2 ₁ 2	1	аро
2	a 43.3	Uni b 43.3	t Cell D c 80.9	imensi α 90.0	ons β 90.0	γ 90.0	tetragonal P4 ₁ 2 ₁ 2	1	F5
3	a 42.6	Uni b 42.6	t Cell D c 81.1	imensi α 90.0	ons β 90.0	γ 90.0	tetragonal P4 ₃ 2 ₁ 2	1	apo
4	a 38.9	Uni b 82.1	t Cell D c 39.4	imensi α 90.0	ons β 92.9	γ 90.0	Monoclinic P2 ₁	4	аро
5	A 41.7	Uni b 41.7	t Cell D c 81.6	imensi α 90.0	ons β 90.0	γ 90.0	tetragonal P4 ₁ 2 ₁ 2	1	VS influenced
6	a 79.7	Uni b 39.7	c 81.1	imensi α 90.0	ons β 90.0	γ 90.0	monoclinic P2 ₁	8	F6

Table 23: The crystal forms wild-type PEX14 was processed as.

Crystal Form		Uni	t Cell D	imensi	ons		Space Group	Molecules per asymmetric unit	Ligand
7	a 38.9	Un b 81.3	t Cell C c 39.3	Pimensi α 90.0	ons β 92.8	γ 90.0	Monoclinic P2 ₁	4	Soaked compound 1 but no electron density seen in crystal structure corresponding to compound 1
8	a 54.1	Uni b 56.8	t Cell D c 81.0	imensi α 90.0	ons β 90.0	γ 90.0	Orthorhombic C222 ₁	2	apo
9	a 53.9	Un b 56.7	c 80.9	α 90.0	ons β 90.0	γ 90.0	Orthorhombic C222 ₁	2	Compound 1

Table 24: Fragment hits from the library used to screen T. brucei PEX14 by NMR.

Fragment Number	Chemical Name	Chemical Structure
F1	1-benzothiophene-3-carboxylic acid	HO
F2	1-benzothiophene-5-carboxylic acid	но
F3	2'-Hydroxy-1'-acetonaphthone	O CH ₃
F4	1,8-Diaminonaphthalene	NH ₂ NH ₂
F5	1H-indole-7-carboxylic acid	HOO
F6	benzo[b]thiophene-7-carboxylic acid	HOO
F8	3-benzofuranacetic acid	ОН
F9a	alpha, alpha-dimethyl-6-methoxy-2- napthalenemethanol	OH
F10	4H-Thieno[3,2-b]pyrrole-5-carboxylic acid	S HOH
F11	Ethyl 3-indoleacetate	O_CH ₃
F12	3-Indoleacetic acid	OH OH

 $Table\ 25: Chemical\ structures\ of\ PEX14\ compounds\ and\ PDB\ codes\ of\ their\ crystal\ structures\ in\ complex\ with\ PEX14.$

Compound No.	Chemical Structure	PDB Code
Compound 1	NH NN NH	Not published.
Compound 2	HN N N N N N N N N N N N N N N N N N N	5L87
Compound 3	HN O OCH3	5N8V

Table 26: Chemical structures of the selected PEX14 binders found through a combination of SBVS and LBVS.

Compound No.	Chemical Structure	Compound No.	Chemical Structure
VS 1	N C C C C C C C C C C C C C C C C C C C	VS 15	N N N N N N N N N N N N N N N N N N N
VS 2		VS 16	o N H N
VS 3		VS 17	
VS 4		VS 18	
VS 5		VS 19	N H NH
VS 6	O N N N N N N N N N N N N N N N N N N N	VS 20	o N N N N N N N N N N N N N N N N N N N

VS 7	NH HN	VS 21	G N N N N N N N N N N N N N N N N N N N
VS 8		VS 22	
VS 9		VS 23	
VS 10	C NH IN	VS 24	C NH
VS 11	N H	VS 25	NH NH
VS 12		VS 26	NH OH

VS 13	VS 27	NH NH
VS 14	VS 28	HN HN N N N N N N N N N N N N N N N N N

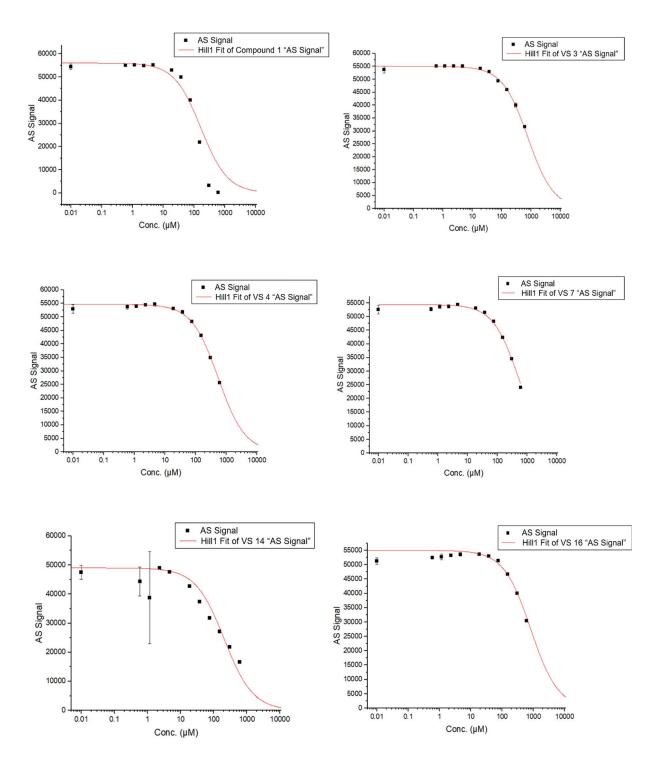


Figure 1a: The IC₅₀ plots generated from the AlphaScreen for the measured VS compounds and compound 1.

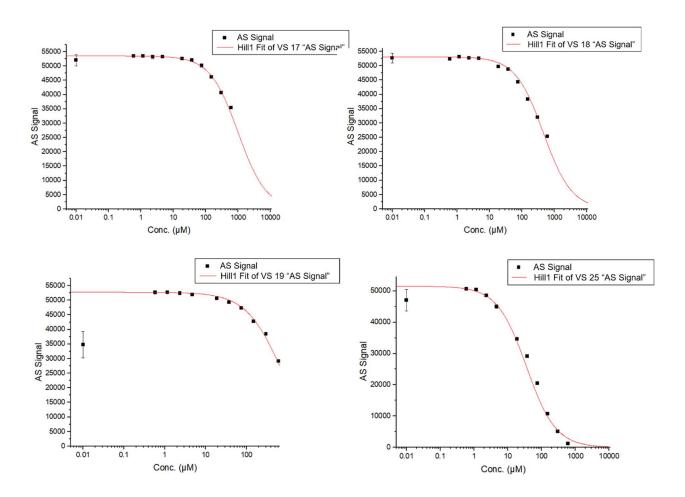


Figure 2b: The IC₅₀ plots generated from the AlphaScreen for the measured VS compounds and compound 1.

 $\label{thm:continuous} \textbf{Table 27: Values of \% inhibition of the fragment hits of ISA SPR screen against endothiapeps in.}$

Fragment ID	% Binding ^a	% Inhibition	Fragment ID	% Binding ^a	% Inhibition
177	11.4	88.6	54	83.8	16.2
218	19.3	80.7	338	83.9	16.1
284	23.0	77.0	356	84.6	15.4
306	23.9	76.1	242	85.4	14.6
290	27.9	72.1	150	85.4	14.6
159	32.1	67.9	76	85.5	14.5
255	35.3	64.7	318	85.8	14.2
64	40.6	59.4	16	86.2	13.8
149	45.4	54.6	243	86.3	13.7
343	46.2	53.8	319	86.4	13.6
337	48.1	51.9	85	87.1	12.9
178	48.9	51.1	128	87.1	12.9
161	52.4	47.6	37	87.6	12.4
167	52.7	47.3	291	88.5	11.5
65	53.7	46.3	109	88.6	11.4
148	57.5	42.5	323	88.7	11.3
47	59.7	40.3	181	89.1	10.9
334	61.3	38.7	8	89.7	10.3
162	62.5	37.5	252	89.9	10.1
79	63.3	36.7	57	89.9	10.1
248	64.7	35.3	251	90.1	9.9
236	69.4	30.6	208	90.1	9.9
63	70.5	29.5	234	90.2	9.8
41	71.7	28.3	5	90.3	9.7
261	73.5	26.5	44	90.4	9.6
88	74.1	25.9	133	90.5	9.5
272	75.7	24.3	114	90.5	9.5
301	76.1	23.9	13	90.5	9.5

175	76.3	23.7	209	90.5	9.5
Fragment ID	% Binding ^a	% Inhibition	Fragment ID	% Binding ^a	% Inhibition
216	76.5	23.5	262	90.9	9.1
101	77.6	22.4	196	90.9	9.1
66	77.9	22.1	267	91.0	9.0
142	79.3	20.7	62	91.1	8.9
40	80.3	19.7	137	91.2	8.8
75	80.7	19.3	332	91.5	8.5
222	80.7	19.3	357	92.0	8.0
293	80.8	19.2	18	92.0	8.0
247	81.4	18.6	158	92.1	7.9
112	82.0	18.0	320	92.1	7.9
17	82.3	17.7	304	92.1	7.9
316	82.7	17.3	333	92.2	7.8
273	92.3	7.7	122	94.5	5.5
321	92.6	7.4	307	94.8	5.2
152	92.6	7.4	173	94.9	5.1
188	92.7	7.3	144	95.0	5.0
33	93.1	6.9	264	95.0	5.0
287	93.2	6.8	260	95.0	5.0
285	93.4	6.6	263	95.1	4.9
328	93.6	6.4	299	95.1	4.9
326	93.7	6.3	254	95.3	4.7
60	93.8	6.2	361	95.3	4.7
335	93.8	6.2	39	95.5	4.5
223	94.0	6.0	126	95.6	4.4
192	94.0	6.0	191	95.7	4.3
180	94.0	6.0	244	95.9	4.1
3	94.0	6.0	258	96.4	3.6
151	94.1	5.9	213	96.5	3.5

189	94.2	5.8	226	96.5	3.5
Fragment ID	% Binding ^a	% Inhibition	Fragment ID	% Binding ^a	% Inhibition
34	94.3	5.7	171	96.6	3.4
277	94.4	5.6	235	96.6	3.4
351	94.4	5.6	42	96.7	3.3
160	94.5	5.5	347	96.7	3.3
346	94.5	5.5			

a) Reflects % of free protein, calculated as (slope produced with frag/control slope)

Table 28: Buffers used in purification, crystallization, and screening of the various proteins.

Protein	TGT	PEX14	EP
Lysis Buffer	20 mM TRIS pH 7.8, 10 mM EDTA, 1mM DTT and 2 cOmplete™-Protease Inhibitor Cocktail Tablets	50 mM TRIS pH 8, 20 mM imidazole, 300 mM NaCl, 0.5 mM TCEP and 2 cOmplete™-Protease Inhibitor Cocktail Tablets	<u></u>
Purification buffer	Q-Sepharose anion exchange column: Buffer A: 10 mM TRIS pH 7.8, 1 mM EDTA, 1mM DTT Elution buffer B: 10 mM TRIS pH 7.8, 1 mM EDTA, 1mM DTT, 1M NaCl Streptactin column: Buffer W: 100 mM TRIS pH 7.8, 1M NaCl, 1 mM EDTA Elution buffer E: 100 mM TRIS pH 7.8, 1 M NaCL, 1 mM EDTA, 2.5 mM D-desthiobiotin	Ni-column: TRIS buffer: 50 mM TRIS pH 8, 20 mM imidazole, 300 mM NaCl, 0.5 mM TCEP Elution buffer: 50 mM TRIS pH 8, 250 mM imidazole, 300 mM NaCl, 0.5 mM TCEP Dialysis buffer: 50 mM TRIS pH 8, 1 mM EDTA, 150 mM NaCl, 0.5 mM TCEP Superdex column: 5 mM TRIS pH 8, 150 mM NaCl, 0.5 mM TCEP	
Storage buffer	10 mM TRIS pH 7.8, 2M NaCl, 1mM EDTA	5 mM TRIS pH 8, 150 mM NaCl, 0.5 mM TCEP	0.1 M sodium acetate pH 4.6

Crystallization	100 mM MES pH 5.5, 1 mM DTT, 13% (w/v) PEG8000, 10% (v/v) DMSO, 0.3 mM protein Co-crystallization: 0.23 mM protein, 6.6 mM ligand	E1W variant: 0.2 M Ammonium Sulphate, 30 %(w/v) PEG 8000, 2 mM protein, 4° C 0.1 M Trisodium Citrate pH 5.6, 0.1 M Lithium Sulphate, 30% PEG 400, 3.7 mM protein, 18° C 0.05 M Trisodium Citrate pH 4.5, 0.07 M Sodium Chloride, 22% PEG 400, 3.7 mM protein, 18° C Co-crystallization: 0.17 M ammonium acetate, 0.085 M sodium citrate pH 5.6, 25.5% PEG 4000, 15% glycerol, 1 mM protein, 20 mM ligand, 18° C Wild-type: 0.1 M Trisodium Citrate pH 5.5, 40% PEG 600, 12 mM protein, 4° C	0.1 M sodium acetate pH 4.6, 0.1 M ammonium acetate, 28% - 30% (w/v) PEG4000, 0.13 mM protein
Soaking	17.5% PEG3350, 25% PEG400, 200 mM NaCl, 100 mM Fragment	E1W Variant: 0.2 M Ammonium Sulphate, 30 %(w/v) PEG 8000, 25% glycerol 0.17 M ammonium acetate, 0.085 M sodium citrate pH 5.6, 25.5% PEG 4000, 15% glycerol 100 mM Fragment Wild-Type: 0.1 M Trisodium Citrate pH 5.5, 40% PEG 600, 25% glycerol, 4° C 100 mM compound 143 mM compound	0.1 M sodium acetate pH 4.6, 0.1 M ammonium acetate, 28% - 30% (w/v) PEG4000, 25% glycerol. 90 mM Fragment

SPR screening	10 mM Hepes buffer pH 7.4, 0.05% Tween 20, 150 mM NaCl, 1 mM TCEP, 1% DMSO, 1.2 μM protein, 1 mM fragment		25 mM Na-acetate pH 4.5, 125 mM NaCl, 0.005% Tween 20, 1% DMSO, 50 nM protein, 1 mM fragment
NMR screening	20 mM deuterated TRIS at pH 7.4, 1 mM EDTA, 150 mM NaCl, 10% (vol/vol) D ₂ O, 3 mM TCEP and 10 μM TMSPd11, 3 μM protein, 200 μM fragment	20mM sodium phosphate pH 6.5 and 50 mM NaCl, 10% D₂O, 100 μM of protein, 1 mM ligand	3 mM acetic acid pH 4.6, 356 μM H-142 peptide, 9.6 μM protein, 310 μM fragment

Acknowledgments

First and foremost, I thank God who in His infinite mercy gave me courage, patience, and strength to complete this work.

My endless appreciation goes to my supervisor Prof. Dr. Gerhard Klebe for believing in me and giving me the opportunity to join his research group. His continuous support, constructive discussions, and immense knowledge have helped me during these years to grow not only as a scientist but also as a person, for which I am forever grateful.

I am also grateful to Prof. Dr. Andreas Heine for his patience and constant devotion in teaching me about the world of X-ray crystallography. I will always appreciate our fruitful discussions.

Thanks are due to all my internal and external collaborators who made this work possible. Prof. Dr. Michael Sattler, Dr. Grzegorz Popowicz, Astrid Lauxen, Charlotte Softley, and Roberto Fino from the Helmholtz Zentrum in Munich; Prof. Dr. François Diedrich, Dr. Christoph Hohn, Dr. Levon Movsisyan, Wieland Goetzke, Prof. Dr. Gisbert Schneider, and Ryan Byrne from ETH Zurich; Dr. Stefan Geschwindner, Dr. Per-Olof Eriksson, Dr. Göran Dahl, Dr. Anna Aagarad, and Ave Kuusk from Astrazeneca Mölndal in Gothenburg; Patrick Walter from Institute Pasteur in Paris, Tobias Wulsdorf from AG Klebe, and Ralf Pöschke from the MarXtal facility of the Philipps University in Marburg.

Huge thanks to all my colleagues from AG Klebe and AG Kolb at the Philipps University in Marburg for the incredible working environment and always lending a helpful hand. Special thanks to Dr. Alexander Metz, the best postdoc of all time for the countless times I asked for his help; Francesca Magari, my Mediterranean sister with whom I have once shared not only an apartment and office, but four years of friendship and countless memories; Lorena Zara, Jessica Desogus, Stefania Monteleone, and Maria Giovanna Papadopoulos for making work fun, and Dr. Frederik Ehrmann and Dr. Nedyalka Radeva for their expertise and support during their time in AG Klebe. I also thank all members of AG Klebe who experienced the sleep deprivation of our synchrotron trips but pulled through to collect datasets of our precious crystals.

A heartfelt thank you goes to Dr. Nicole Bertoletti for always being there for me at the beginning of my PhD. Her presence and encouragement got me through a lot of hard days (and nights!) in the S1 lab. I will always appreciate her help and friendship.

I would like to extend my thanks to Lydia Hartleben for her administrative support and care even before I joined AG Klebe. Thank you for always ensuring things are running as they should! Thanks to Christian Sohn and Steffi Dörr for their technical help, and very importantly Felix Terwesten, Phong Nguyen, Tobias Wulsdorf, and Corey Taylor for managing our computer networks.

Acknowledgment is due to the European Union for funding my work through Horizon 2020 under the Marie Skłodowska-Curie fellowship. Being part of the Accelerated Early staGe drug dlScovery (AEGIS) program has been a huge privilege. To my dear AEGIS fellows, thank you for making these past three years one of the best of my life. I will always cherish this journey that we took together and the memories we share. It has been great! I would also like to thank our project coordinator Dr. Eva Schlosser from the Helmholtz Zentrum in Munich for her organization of this program, and all the AEGIS fellows who worked so hard in organizing all our academic schools.

Acknowledgments

I also give thanks to Prof. Dr. Gerhard Klebe, Prof. Dr. Andreas Heine, Prof. Dr. Michael Sattler, Prof. Dr. Klaus Reuter, Dr. Wolfgang Jahnke, Dr. Stefan Geschwindner, Dr. Per-Olof Eriksson, Dr. Doru Vornicescu, Dr. Grzegorz Popowicz, Charlotte Softley, and Johanna Senst for their contributions in revising various sections of this thesis.

In the end, I deeply thank my dear family for pushing me to pursue my goals. Their support, even thousands of kilometers away, has driven me to see this work through.

Last but not least, an endless thank you to my lovely husband Mostafa Nofal for taking this journey with me. Thank you for your patience during all those long months away, thank you for your understanding, and thank you for your persistent encouragement and confidence in me. I couldn't have done it without you.

ENGI HASSAAN

Conferences & Publications

CONFERENCES

ORAL PRESENTATIONS

POSTER PRESENTATIONS & AWARDS

PUBLICATIONS

- Flash talk at "New Frontiers in Structure-based Drug Discovery"; 2019 September 23-25; Florence, Italy. Talk Title "Do All Roads Really Lead to Rome? Learnings from Comparative Analysis using SPR, NMR, & X-Ray Crystallography to Optimize Fragment Screening in Drug Discovery".
- Speaker at 14th Annual Drug Discovery Chemistry Conference; 2019 April 9-10; San Diego, USA. Talk Title: "Busted! Recognizing False Positives and False Negatives Learnings from Comparative Analysis of Fragment Binding using X-Ray Crystallography and NMR" (30 minutes).
- Best Poster Award in "New Frontiers in Structure-based Drug Discovery Conference"; 2019 September 23-25; Florence, Germany. Poster Title: "Do All Roads Really Lead to Rome? Learnings from Comparative Analysis using SPR, NMR, & X-Ray Crystallography to Optimize Fragment Screening in Drug Discovery".
- Annual BESSY User Meeting; 2018 December 5-7; Berlin, Germany. Poster Title: "Primary Screen Awards".
- Annual DPhG conference; 2018 October 3-5; Hamburg, Germany. Poster Title: "Comparative Analysis of Fragment Binding Using Orthogonal Biophysical Methods Yield Surprising Results".
- Annual Drug Discovery Summit; 2017 June 7-8; Berlin, Germany. Poster Title: "Fragment-based Approach to Develop Ligands to Fight Shigellosis by Inhibiting tRNA Guanine Transglycosylase (TGT)"
- Hassaan, E. et al. Comparative Analysis of Fragment Binding to a tRNA-modifying Enzyme using SPR, NMR, and X-ray Crystallography, submitted to ChemMedChem Journal (June 2019).
- Hassaan, E. et al. Fragment Screening Hit Draws Attention to a Novel Transient Pocket Adjacent to the Recognition Site of the tRNA-Modifying Enzyme TGT, ready for submission.
- Hassaan, E. et al. Do All Roads Really Lead to Rome? Learnings from Comparative Analysis using SPR, NMR, & X-Ray Crystallography to Optimize Fragment Screening in Drug Discovery, manuscript in preparation
- Hassaan, E. et al. Overcoming Pitfalls from the Crystallographic Fragment Screening of the Trypanosomal Peroxisomal Membrane Protein PEX14, manuscript in preparation.
- Hassaan, E. et al. Combining Virtual Screening, SOFAST HMCQ NMR, and AlphaScreen Assay to Find Novel Inhibitors of the Trypanosoma Brucei PEX14, manuscript in preparation.
- Magari, F., Hassaan, E., Gloeckner, S., Scanlan, J. H. W., Bertoletti, N., Metz, A., Huschmann, F. U., Siefker, C., Abazi, N., Heine, A., Weiss, M. S., Klebe, G. Exhaustive X-ray Crystallographic Fragment Screening of a Hit-Enriched 96 Fragment Library Against Diverse Targets, manuscript in preparation.

Erklärung

Ich versichere, dass ich meine Dissertation
"Do All Roads Really Lead to Rome? Learnings from Comparative Analysis using SPR, NMR, & X-Ray Crystallography to Optimize Fragment Screening in Drug Discovery
selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.
Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.
Marburg, den27.05.2019
(Unterschrift mit Vor- und Zuname)