Technische Universität München

Fakultät für Chemie

Combining Computer-Aided Drug Design, NMR-Fragment based Drug Discovery and Medicinal Chemistry in Early-Stage Drug Development

ROBERTO FINO

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Priv.-Doz. Dr. Gerd Gemmecker

Prüfer der Dissertation: 1. Prof. Dr. Michael Sattler

2. Prof. Dr. Oliver Plettenburg

Die Dissertation wurde am 27.10.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 07.12.2021 angenommen.

Combining Computer-Aided Drug Design, NMR-Fragment based Drug Discovery and Medicinal Chemistry in Early-Stage Drug Development

(Kombination von computergestütztem Wirkstoffdesign, NMR-Fragment-basierter Wirkstoffentdeckung und medizinischer Chemie in der frühen Phase der Wirkstoffentwicklung)

Dissertation

zum Erlangen des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat)

vorgelegt von

Roberto Fino

geboren am 17.09.1984 in Taranto

Technische Universität München

Fakultät für Chemie

Bayerisches NMR-Zentrum

Lehrstuhl für Biomolekulare NMR-Spektroskopie

HelmholtzZentrum münchen

Deutsches Forschungszentrum für Gesundheit und Umwelt

Institut für Strukturbiologie

München, Oktober 2021

Table of Contents

List of publications viii

Abbreviations ix

Abstract x

Zusammenfassung xii

1. Introduction 1

1.1. Computer-aided drug design 1

1.1.1. A brief history of drug discovery 1

1.1.2. Introduction to molecular modeling and CADD 3

1.2. Drug discovery 7

1.2.1. Phenotypic and target-based drug discovery 7

1.2.2. Structure- and ligand-based techniques 9

1.2.3. Computational screening and ligand docking 10

1.2.4. NMR in fragment-based drug discovery 11

1.3. PEX5-PEX14 as a target for drug discovery 14

1.3.1. The PEX5-PEX14 interface 14

1.3.2. PEX14 inhibitors 15

1.3.3. The water envelope around the ligand-PEX14 complex 18

1.3.4. Aim of the thesis 19

2. Materials and methods 20

2.1. Structural biology 20

2.1.1. Protein expression and purification 20

2.1.2. NMR fragment screening 20

2.2. Biophysical assays 21

2.2.1. AlphaScreen™ competition assays 21

2.2.2. NMR spectroscopy ... 21
2.2.3. Trypanolytic assays ... 22
2.3. Computational methods ... 23
2.3.1. Molecular modeling and CADD ... 23
2.3.2. 2D Fingerprint-based HTVS ... 23
2.3.3. Docking and binding affinity predictions ... 24
2.3.4. Molecular dynamics simulations ... 25
3. Publications ... 26
3.1. Computer-Aided Design and Synthesis of a New Class of PEX14 Inhibitors: Substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepines as Potential New Trypanocidal Agents. ... 26
3.2. Introducing the CSP Analyzer: A novel Machine Learning-based application for automated analysis of two-dimensional NMR spectra in NMR fragment-based screening. ... 28
4. Conclusions and outlook ... 31
4.1. Automatic NMR FBS spectra processing ... 31
4.2. PEX14 Ligand optimization ... 32
5. Bibliography ... 33
Acknowledgments ... 45
6. Appendices ... 49
6.1. Introducing the CSP Analyzer: A novel Machine Learning-based application for automated analysis of two-dimensional NMR spectra in NMR fragment-based screening ... 49
6.2. Computer-Aided Design and Synthesis of a New Class of PEX14 Inhibitors: Substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepines as Potential New Trypanocidal Agents. ... 49

6.3. Water envelope has a critical impact on the design of protein-protein interaction inhibitors 49
6.4. Structure-Activity Relationship in Pyrazolo [4, 3-c] pyridines, First Inhibitors of PEX14–PEX5 Protein-Protein Interaction with Trypanocidal Activity 50
6.5. Structure and molecular recognition mechanism of IMP-13 metallo-β-lactamase 50
STATUTORY DECLARATION 51

Zwar ist's mit der Gedankenfabrik
Wie mit einem Weber-Meisterstück,
Wo ein Tritt tausend Fäden regt,
Die Schifflein herüber hinüber schießen,
Die Fäden ungesehen fließen,
Ein Schlag tausend Verbindungen schlägt.
Der Philosoph, der tritt herein
Und beweist Euch, es müßt so sein:
Das Erst wär so, das Zweite so,
Und drum das Dritt und Vierte so;
Und wenn das Erst und Zweit nicht wär,
Das Dritt und Viert wär nimmermehr.

Il fabbricar pensieri, in verità,
Somiglia a quanto avviene sul telaio del tessitore,
Dove mille fili mette in moto un sol premere di piedi.
Scattan su e giù le spole;
Invisibile, corre via ogni stame;
E una percossa sola, intrecci innumerevoli compone.
Il filosofo, qui, entra in scena,
E vi dimostra che essere non doveva, se non così com'è:
Che se il Primo e il Secondo erano tali,
Non potean Terzo e Quarto conseguire a differente specie;
E che se i primi duenon fossero già stati,
Neppure il Terzo e il Quarto sarebbero mai nati.

Johann Wolfgang von Goethe
Faust

List of publications

I hereby declare that this thesis reports results of two manuscripts that are published in international journals. Publication 1 and 2 and a statement on my contributions to these publications are provided in Chapter 3.

1. **Fino, R.***, Byrne, R.*, Softley, C. A., Sattler, M., Schneider, G., & Popowicz, G. M. (2020). Introducing the CSP Analyzer: A novel Machine Learning-based application for automated analysis of two-dimensional NMR spectra in NMR fragment-based screening. *Computational and Structural Biotechnology Journal*.

2. **Fino, R.***, Lenhart*, D., Kalel, V., Softley, C., Napolitano, V., Byrne, R., Schliebs, W., Dawidowski, M., Erdmann, R., Sattler, M., Schneider, G., Plettenburg, O., Popowicz, G. (2021). Computer-Aided Design and Synthesis of a New Class of PEX14 Inhibitors: Substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepines as Potential New Trypanocidal Agents. *Journal of Chemical Information & Modeling.*
   **These authors have contributed equally to this work.*

A complete list of my publications is given in the appendix at the end of this thesis.

Abbreviations

CADD	computer-aided drug design	**TOCSY**	total correlation spectroscopy
PPI	protein-protein interaction	**TROSY**	transverse relaxation optimized spectroscopy
FBS	fragment-based screening	**FPBS**	fingerprint-based screening
FBDD	fragment-based drug design	**DMSO**	dimethyl sulfoxide
ESDD	early-stage drug design	**PEX**	peroxin (peroxisomal protein)
SBDD	structure-based drug design	**T.b.**	*Trypanosoma brucei*
LBDD	ligand-based drug design	**T.c.**	*Trypanosoma cruzi*
HTVS	high-throughput virtual screening	**PCR**	polymerase chain reaction
CATS	chemically-advanced template search	**FAST**	features from accelerated segment test
CSP	chemical shift perturbation	**BRIEF**	Binary robust independent elementary features
GUI	graphical user interface	**ORB**	oriented FAST and BRIEF
ML	machine learning	**HOG**	histogram of oriented gradients
SMOTE	synthetic-minority oversampling technique	**FNR**	false-negative rate
ENN	extended nearest-neighbor	**PCA**	Principal component analysis
MSE	mean square error	**SVC**	support-vector classifier
RMSD	root mean square deviation	**FP**	fluorescence polarization
NMR	nuclear magnetic resonance	**AS**	AlphaScreen™
FID	free induction decay	**ppm**	part per million
HSQC	heteronuclear single quantum coherence/correlation	**kDa**	kilodalton
HMQC	heteronuclear multiple quantum correlation	**IC_{50}**	inhibitory concentration
NOE	nuclear Overhauser effect	**EC_{50}**	effective concentration
NOESY	nuclear Overhauser effect spectroscopy	**K_i**	inhibition constant
PDD	phenotypic drug discovery	**K_d**	dissociation constant

Abstract

This thesis focuses on the synergy between computational drug discovery methods, structural biology, and biophysical methods for the development of new treatments for Sleeping Sickness and Chagas Disease. A high-throughput virtual screening (HTVS) protocol was optimized to assess and consider the dynamics of a given scaffold to improve the selection of the best molecules for synthesis. The integration of these methods led to the discovery of a new potential class of inhibitors for the trypanosomal PEX5-PEX14 protein-protein interaction (PPI), the key target in the quest for the development of new bioactive molecules to treat trypanosomiases.

In summary, this thesis project presents novel methods and the application of structure-based drug discovery by exploiting the druggability of PEX14 as a target that is essential for glycosome biogenesis in *Trypanosoma*. The results highlight how the combination between computational methods and structural biology provides unique opportunities for the development of a new family of lead molecules as trypanocidal agents.

Building on the available structural information from NMR and X-ray crystallography on the PEX5-PEX14 PPI a combination of computational methods such as *in silico* screening, docking, and computer-aided drug design (CADD) and experimental techniques, such as AlphaScreen™, NMR, and immunofluorescence assays, were used to guide a first quantitative structure-activity relationship of the new class of inhibitors based on the 1,4 oxazepine scaffolds. NMR-based screening and computational docking identified three candidate scaffolds, of which the 1,4 oxazepine core was selected and prioritized for medicinal chemistry optimization because of the convenient synthetic accessibility of chemical decorations of the core. The first round of structure-activity relationship evaluation was performed with a library of 1,4 oxazepine-based scaffolds, leading to a library of 19 compounds with IC_{50} against *Trypanosoma brucei* ranging from 4 to 30 μM.

Monitoring fragment binding in protein-detected 2D NMR experiments requires analysis of hundreds of spectra to detect chemical shift perturbations CSPs in the presence of ligands screened. To automate the NMR data analysis, *CSP Analyzer* combines a powerful backend Python-based Machine Learning classifier with an intuitive GUI written in C# to provide the user with a tool that allows quickly identifying interacting fragments and select them for further structural analysis. The software allows rapid evaluation of 2D screening data from a large number of spectra, reducing user-introduced bias in the evaluation and is available on GitHub (https://github.com/rubbs14/CSP-Analyzer/releases/tag/v1.0) under the GPL license 3.0, free to use for academic and commercial users.

Zusammenfassung

Diese Arbeit konzentriert sich auf die Synergie zwischen computergestützten Methoden der Arzneimittelentdeckung, Strukturbiologie und biophysikalischen Methoden zur Entwicklung neuer Behandlungsmethoden für die Schlafkrankheit und die Chagas-Krankheit. Ein virtuelles Hochdurchsatz-Screening-Protokoll (HTVS) wurde optimiert, um die Dynamik eines bestimmten Gerüsts zu bewerten und zu berücksichtigen und so die Auswahl der besten Moleküle für die Synthese zu verbessern. Die Integration dieser Methoden führte zur Entdeckung einer neuen potenziellen Klasse von Inhibitoren für die trypanosomale PEX5-PEX14-Protein-Protein-Interaktion (PPI), dem Schlüsselziel bei der Suche nach der Entwicklung neuer bioaktiver Moleküle zur Behandlung von Trypanosomiasen.

Zusammenfassend stellt dieses Dissertationsprojekt neuartige Methoden und die Anwendung der strukturbasierten Wirkstoffforschung vor, indem es PEX14 als ein für die Glykosom-Biogenese in Trypanosomen wesentliches Ziel für einen Wirkstoff-Ansatz nutzt. Die Ergebnisse zeigen, wie die Kombination von Berechnungsmethoden und Strukturbiologie einzigartige Möglichkeiten für die Entwicklung einer neuen Familie von Leitmolekülen als trypanozide Wirkstoffe bietet.

Aufbauend auf den verfügbaren Strukturinformationen aus der NMR- und Röntgenkristallographie über den PEX5-PEX14-PPI wurde eine Kombination aus rechnergestützten Methoden wie in-silico-Screening, Docking und computergestütztem Wirkstoffdesign (CADD) und experimentellen Techniken wie AlphaScreen, NMR und Immunfluoreszenz-Assays eingesetzt, um eine erste quantitative Struktur-Wirkungs-Beziehung der neuen Klasse von Inhibitoren auf der Grundlage der 1,4-Oxazepin-Gerüste zu erstellen. Durch NMR-basiertes Screening und computergestütztes Docking wurden drei Gerüstkandidaten identifiziert, von denen der 1,4-Oxazepin-Kern ausgewählt und für die medizinisch-chemische Optimierung priorisiert wurde, da die

chemische Dekoration des Grundgerüsts synthetisch leicht zugänglich sind. Eine erste Runde der Bewertung der Struktur-Aktivitäts-Beziehung wurde mit einer Bibliothek von 1,4-Oxazepin-basierten Gerüsten durchgeführt, die zu einer Bibliothek von 19 Verbindungen mit einer IC_{50} gegen *Trypanosoma brucei* zwischen 4 und 30 μM führte.

Die Überwachung der Fragmentbindung in 2D-NMR-Experimenten mit Proteindetektion erfordert die Analyse von Hunderten von Spektren, um Störungen der chemischen Verschiebung von CSPs in Gegenwart der untersuchten Liganden zu erkennen. Um die NMR-Datenanalyse zu automatisieren, kombiniert CSP Analyzer einen leistungsstarken, auf Python basierenden Machine-Learning-Klassifikator mit einer intuitiven, in C# geschriebenen Benutzeroberfläche, um dem Benutzer ein Werkzeug an die Hand zu geben, mit dem interagierende Fragmente schnell identifiziert und für die weitere Strukturanalyse ausgewählt werden können. Die Software ermöglicht die schnelle Auswertung von 2D-Screening-Daten aus einer großen Anzahl von Spektren und reduziert das vom Benutzer verursachte Bias bei der Auswertung. Sie ist auf GitHub (https://github.com/rubbs14/CSP-Analyzer/releases/tag/v1.0) unter der GPL-Lizenz 3.0 verfügbar und kann von akademischen und kommerziellen Nutzern kostenlos verwendet werden.

1. Introduction

This chapter will introduce the main fields covered throughout the doctoral research work. First, a general overview of the role of CADD will be discussed, with a particular focus on the exploration of the available chemical space using HTVS 2D- and 3D-based methods. In section 1.2, the focus will be put on the structural biology techniques involved in target recognition and how this can be used for drug design. Finally, in section 1.3, the reader will have more insights into the structural biology of the PEX importomer system and how the proteins involved represent an excellent target for drug discovery.

This work was co-supervised by Dr. Grzegorz Popowicz (HMGU-STB/TUM-BNMRZ) and Prof. Dr. Michael Sattler (HMGU-STB/TUM-BNMRZ) and financed by the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Skłodowska-Curie Grant Agreement No. 675555 and Accelerated Early-staGe drug dIScovery (AEGIS) and supported by the Helmholtz Association Initiative and Networking Funds.

1.1. Computer-aided drug design

1.1.1. A brief history of drug discovery

The origins of drug discovery can be dated back to the early dawn of mankind. Since the early human civilizations, the occasional discovery of medicinal plants and pigments has always accompanied and boasted the development of new settlements and improved the quality of life of the first societies. A remarkable finding was the discovery of pieces of the fungus birch polypore (fomitopsis betulina) in a sachet found on Ötzi, 5.300-year-old mummy found by hikers on the Alps at the Italian-Austrian border in 1991.

The birch polypore is known in medicine because is the source of a cocktail of alkaloids and secondary metabolites that have a range of benefits, from pain relief

to anthelmintic effects. The latter is most likely to be the reason why Ötzi was taking pieces of the polypore with him, given the infestation of intestinal worms that was plaguing him [1].

With the advent of the Age of Enlightenment and the rise of the scientific method as the principal form of investigation, the first attempts to correlate the activity of chemicals to clinical effects were conducted (e.g., the discovery of the anesthetic effects of ethyl ether and the early trial-and-fail attempts to optimize its effects).

It was not before the advent of modern chemistry in early 1900, though, that appreciable progress was made. Before the Fischer family and the school of German organic chemistry pioneers set the path for modern organic chemistry, the discovery of new treatments was mostly serendipitous.

With the improvement of separation and purification techniques, a plethora of natural compounds was identified and studied; amongst these compounds, salicylic acid and morphine are maybe the most notable ones. The rudimentary set of chemical modifications available to the pioneers of medicinal chemistry, though it was limited, led to the discovery of derivatives of natural compounds and the seeds of modern combinatorial library design. The results of these efforts afforded the first 'blockbuster' drug known to mankind: the acetylated derivative of salicylic acid, more known under its commercial name: Aspirin (Bayer). The same acetylation reaction that led to the discovery of Aspirin led also to a more notorious derivative of morphine, heroin, which was extensively used as a powerful anesthetic. The commercial name 'Heroin' was chosen by Bayer in 1903 because of the incredibly powerful effects in the suppression of pain if compared to morphine; what was not completely elucidated at that time, was that the better affinity to the opioid receptor mu-opioid receptor (MOR) was also followed by an increased dependency after the treatment and, in the worst cases, by the death of the patients treated upon respiratory depression (the most common side effect of morphine and its derivatives). The casualties and the social effects of the abuse of heroin, pushed the governments and the scientific community to increase the controls and the regulation of the studies required for the validation

of the clinical effects of drugs. The scientific community also set the standards and protocols for modern pharmacological screening.

1.1.2. Introduction to molecular modeling and CADD

In recent years, molecular modeling and CADD became progressively more important tools in early-stage drug design. Recent developments in target identification techniques provided medicinal chemists with more information on biological targets. Li et al. define molecular modeling as *the science of representing molecular structures numerically and simulating their behavior with the equations of quantum and classical physics* [2, 3].
. A CADD software package should ideally fulfill two main purposes: 1) molecule 3D visualization and 2) interactive molecular editing.
Several packages are available for academic use. Amongst the most used there is PyMol [4], which provides visualization, some basic editing features, and even support tools for binding site identification and molecular docking protocols [5]; a commercial version of PyMol with more advanced features is licensed by Schrödinger LLC. One of the main advantages of PyMol is the python-based scripting, which opens up a wide range of possible customizations using both user-written scripts and more sophisticated extensions via custom plugins.
Besides PyMol, the University of Illinois' Theoretical and Computational Biophysics group lead by Klaus Schulten designed another blockbuster for molecular modeling: VMD [6] (aka visual molecular dynamics). VMD is a full-stack and extensible suite for molecular visualization, editing, and simulations written in C and interfaced with other software via Tcl/Tk command-line console. It features a large library of plugins for several tasks and extensions. Of the latter, one of the most used is the NAMD scalable molecular dynamics package. [7]
Other notable mention software packages available for free for academic and commercial uses are UCSF's Chimera [8] which provides both visualization and editing features, Avogadro [9] that features plugins for the generation of topology

files and other useful features for the customization of input files for other packages (i.e. ORCA QM calculations input files), and the Java-based Jmol 3D visualizer [10], which is also available as a web-based plugin.

One of the state-of-the-art applications which fulfill most of the purposes of computational drug design is the Schrödinger Maestro [11] molecular modeling package, featuring in the Linux-based installation only – the D.E. Shaw Desmond molecular dynamics package [12], and Glide [13] for HTVS and molecular docking.

A very popular drug discovery platform is developed by the Chemical Computing Group and released under the name of Molecular Operating Environment (MOE) [14], and it stands as a full-stack software package for small-molecules drug discovery, with applications for library preparation, molecular design, HTVS, and molecular dynamics.

A full suite for molecular modeling is available from YASARA Biosciences is Yasara Structure [15], which features also molecular visualization based on OpenCL, molecular editing, and a command-line console with its built-in programming language (Yanaconda, with python-based syntax) to automate tasks for high-throughput applications.

Last but not least, is worth mentioning Acellera Labs gpGPU-powered software for biophysical simulations ACEMD [16] which features in the HTMD [17] package. Also released from Acellera Labs, PlayMolecule is a web-based GUI and workflow handler available as a free service for testing at http://www.playmolecule.org. [18-20]

In recent times, the development of molecular modeling applications has been also focusing on the new opportunities offered by augmented and virtual reality systems such as Oculus Rift, HTC Vive, and Microsoft's Hololens. In this context, it is worth mentioning the effort made by Norrby et al. that lead to the development of Molecular Rift [21], a C#-based software for drug designers that runs a virtual molecular modeling environment for the Oculus Rift system. Besides, the main

packages mentioned several other software are available for molecular modeling and editing.[1]

Molecular modeling is a useful tool in the quest of drug design and ligand optimization and it often represents one of the nodes of more complex virtual workflows in drug discovery. *In silico* drug discovery is an iterative process that

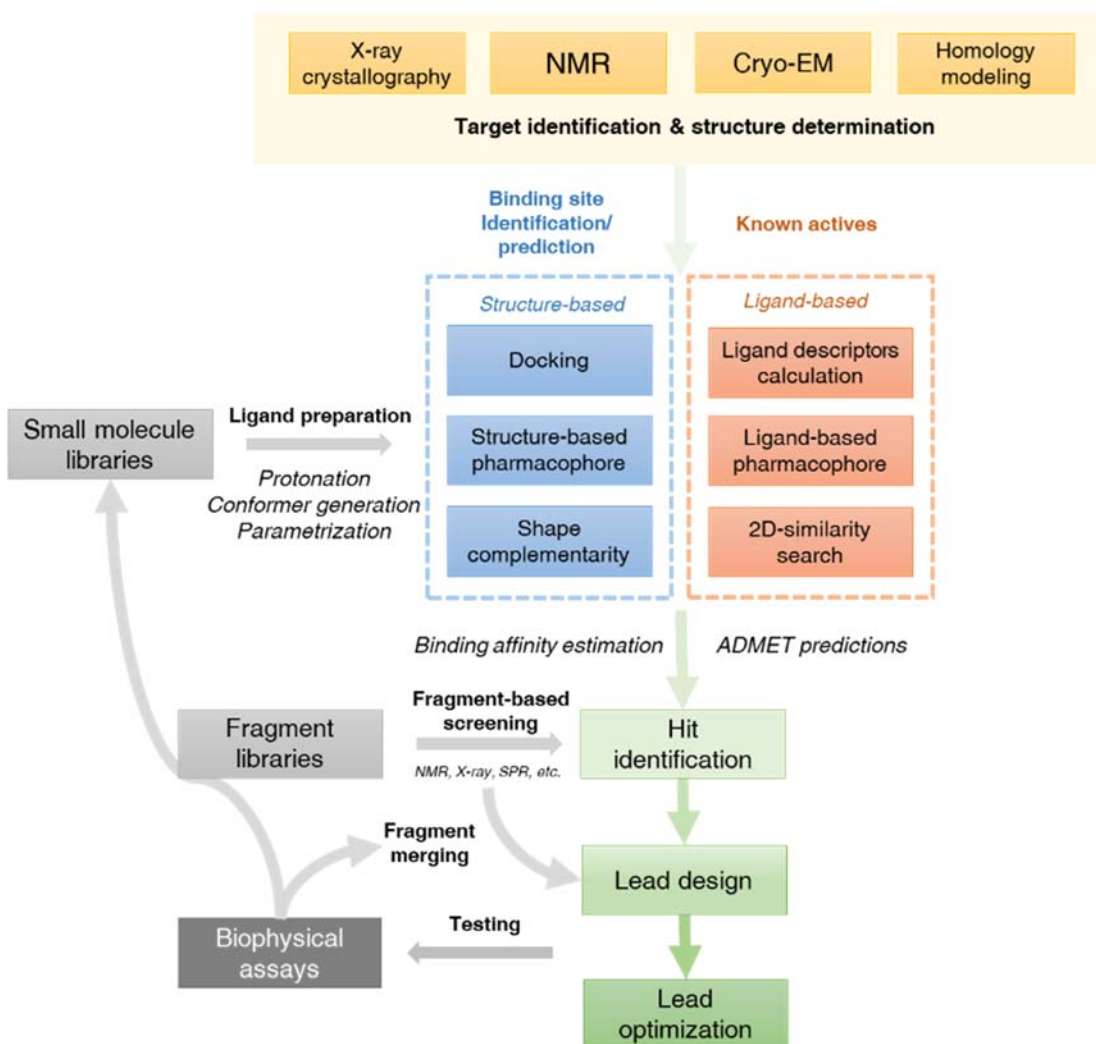


Figure 1. A typical workflow for in silico drug discovery.

consists of several interconnected stages (Figure 1). Besides molecular modeling, other methods are widely used for ligand design and optimization.

[1] For a more comprehensive list of the available packages please refer to: http://www.click2drug.org/directory_MolecularModeling.html

A crucial step in every drug design workflow is the correct structure preparation of the selected biological target and the ligand libraries. In general, the target structure is available for download from the online database Protein data bank (PDB). Depending on the method used to determine the structure (e. g. NMR or X-ray crystallography), the protein may require some fixing and checkups before performing any further analysis. Typically, hydrogens are missing from X-ray determined structure, because the electron density of hydrogen atoms is too weak to be detected even at higher resolutions.
Since the PDB file format is only containing the information on the atom coordinates, bond typing is required to assign generate the right bond connectivity for the atoms in the molecule, though, valence is formally reported in the CONECT field of the PDB file. Target preparation is often performed by adding hydrogens according to an appropriate pH, adding any missing side chains, fixing the N- and C-termini, and assigning the right bond orders to the aromatic residues.
Biological targets of interest, usually protein structures but more recently also DNA- or RNA-based receptors [22-24], often require structural refinement before being used for HTVS purposes; routine protein preparation is performed adding the eventually missing sidechains of flexible residues (e.g. lysine or disordered loops), adding missing hydrogens, and estimate the protonation states of chargeable residues (i.e. aspartate, glutamate, arginine, etc.)
For the latter purpose, some methods are available, with PROPKA [25] being one of the standards for the estimation of the stability of protomers of the sidechains at a given pH.
From the perspective of the ligand libraries, due to the nature of the chemical properties of small organic molecules such as tautomerization and isomerization, the prediction of protonation states becomes a more complicated task. A first step in the ligand preparation is that of correctly assigning the bond orders to the structures. While this may seem a trivial task, it is heavily dependent on the nature of the chemical file in input: PDB files, for example, do not bear any information

about the bond orders, even though the CONECT field may refer to the surrounding chemical connectivity of a given set of atoms. For this purpose, SMILES (for 2D structures), SDF, and MOL2 file formats are better suited to describe the chemical structure of small organic molecules, because they keep the information of the atom typing and thus the bond order connectivity between them.

Once the bond orders are assigned correctly, it is necessary to generate protomers according to a given pH and to add the missing hydrogens to the ligands. The affinity of a ligand for a given binding site is strictly dependent on the protonation state of the ligand and thus on its total charge. [26] Several tools are available for ligand protonation and tautomer generation [27], with some of them capable of predicting the pKa of chargeable sites and others protonating the compounds in a deterministic manner according to a specific set of rules. Amongst the free tools available for ligand preparation, a notable mention is *dimorphite_dl* [28] which is available under the Apache License v2 and freely available both for commercial and academic purposes. While dimorphite is capable of enumerating all the possible protomers in a range of pH values, it is worth noticing that it accepts only 2D structures as input (SMILES). On the other hand, another valuable tool for ligand preparation (available under the less permissive GPLv2 License) is OpenBabel [29], which can protonate a ligand at a given pH.

1.2. Drug discovery

1.2.1. Phenotypic and target-based drug discovery

The discovery of compounds in a mechanism-agnostic way by the screening of large libraries against a given target is known as phenotypic drug discovery (PDD).

The pioneer of this technique, and the first to discover his so-called 'golden bullet', was the German scientist Paul Ehrlich and his studies on trypanocidal drug salvarsan.

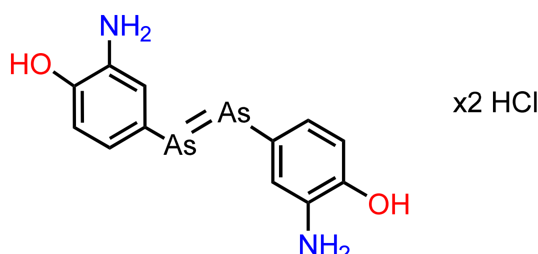


Figure 2. The structure of Salvarsan.

Phenotypic drug discovery is one of the key methods used in drug discovery to identify first-in-class drugs [30]. While PDD is a powerful technique, a relatively hard and complex deconvolution is often required to elucidate the causes and mechanism of the pharmacological effects induced by the compounds tested.
On the other end, this technique will not require any structural information in the early stages of drug discovery, thus allowing a higher speed of the overall process in the first phases of the project.
The advances in protein engineering and recombinant gene technologies, the skyrocketing evolution of structure identification techniques such as X-ray protein crystallography, Cryo-EM, and NMR, had led the pharmaceutical industries to focus more on a target-first approach, switching to more sophisticated alternatives to PDD. Recently, structure prediction through machine-learning powered algorithms with DeepMind AlphaFold has proven as a promising method to gain insights on the yet structurally unexplored proteasome. [31]
In this context, target-based drug design, either structure- or ligand-based has progressively gained momentum, leading to the development of new, and more optimized drugs.

1.2.2. Structure- and ligand-based techniques

On a general basis, the drug design workflow starts with the identification of the biological target of interest. In the vast majority of the cases, the target is identified as a particular protein.

The very first step right after the identification and the expression of the biological target is that to determine its 3D structure. For this purpose, X-ray crystallography, NMR, or Cryo-EM techniques may be employed according to the physicochemical properties of the target. The selected target is usually addressed as a *receptor*.

In this context, structure-based drug design is performed to retrieve key features on the receptor surface. In the vast majority of cases, the receptor has some natural binding partner, which can be another protein, a small peptide, a nucleic-based (either DNA or RNA) sequence, or an endogenous organic small molecule (such as vitamins or co-enzymes as NAD or FAD). In this case, the drug designer can retrieve the key features in the interaction between the receptor and its natural ligand to generate a pharmacophore binding hypothesis. It can then be tested via experimental methods, for example, through FBDD.

The IUPAC defines a pharmacophore as "the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response" [32]. It is worth noticing that a pharmacophore is not a defined unique chemical component, but more a defined arrangement in space of particular features and moieties. The pharmacophoric hypothesis can be considered as the synopsis of the receptor-ligand binding hypothesis, and it is the basis of SAR and QSAR assessment of virtual screening results.

On the other hand, problems may arise if only the target's *apo* structure is known; in this case, the identification of the right binding site and pharmacophoric features on the surface of the target is required.

1.2.3. Computational screening and ligand docking

Thanks to recent advances in computer technologies and with the spreading of multi-threading processors, high-throughput virtual screening (HTVS) software has become a valuable asset in early-stage drug development.

High-throughput virtual screening techniques allow fast access to the exploitable chemical space. Compounds suitable for screening are often available from different vendors' databases or common comprehensive databases such as ZINC [33, 34], or ChEMBL [35].

Databases can be either screened by 2D pharmacophore matching (e.g. substructure matching or maximum common scaffold – MCS – matching) [36-38] or 3D properties (for example, shape complementarity to the binding site) and 3D pharmacophore matching [39].

Virtual screening from 2D structures is mostly performed converting the molecules to fingerprints. Fingerprints can be described as vectors that store the information about the atom patterns in the molecule as binary arrays of 0 and 1. Once the fingerprint of a molecule is calculated, a metric to test the difference (namely the "distance") between the fingerprint of a reference molecule and the tested library can be chosen to quantify the molecular similarity score.

Fingerprint-based virtual screening methods are a fast and efficient way to screen large libraries of compounds in the early stages of rational drug design. Several fingerprints are available to represent the different molecular properties of compounds into arrays [40-45] which can then be compared using different distance metrics; in this context, Tanimoto distance calculated between binary arrays is one of the most common measures of similarity used in chemoinformatics [46]. From a ligand-based screening approach perspective, one of the main advantages of using 2D fingerprints is that, given the structural distribution in space of a known ligand, it is possible to compare up to millions of structures to the reference compound and retrieve those that better resemble the same arrangement of the pharmacophoric features required by the binding hypothesis

in a very short amount of time. While common topological fingerprints were designed initially for similarity searching, other circular fingerprints like ECPF and MACCS were specifically created for structure-activity modeling [47], which represents a useful tool in hit-to-lead optimization and lead optimization QSAR studies. Other topological fingerprints such as the CATS (chemically advanced template search) [48] allowed medicinal chemists to facilitate core-hopping and scaffold-hopping design, thus unlock new routes toward the discovery of new compounds against known targets that cover a new, IP novel, part of the chemical space against a selected target. Besides 2D fingerprinting, 3D fingerprints such as the rapid overlay of chemical scaffolds (ROCS) and shape complementarity are commonly used to characterize and screen compounds [49]. A (relatively) novel approach is also that based on the ligand-receptor interactions fingerprinting, a technique that unlocked the capability of assessing the goodness of fitting of a binding hypothesis in a dynamic way (such as in a ligand-protein binding simulation)[50, 51].

1.2.4. NMR in fragment-based drug discovery

In recent times, fragment-based drug discovery (FBDD) has become progressively more important in early-stage drug design. Fragment-based screening (FBS) offers an efficient, rational, way to find small molecule inhibitors. By testing for the binding of small fragments, a large chemical space can be tested with fewer molecules than with other approaches. This is due to the higher probability of a suitable binding pocket or position being present with lower complexity molecules [52], leading to higher efficiency (both screening efficiency and ligand efficiency), and elucidating possible starting points for further drug discovery. The binding sites of these fragments can then be located on the surface of the protein and from there, fragments are connected or grown to

maximize interactions with the surrounding area, eventually leading to hit and lead molecules.

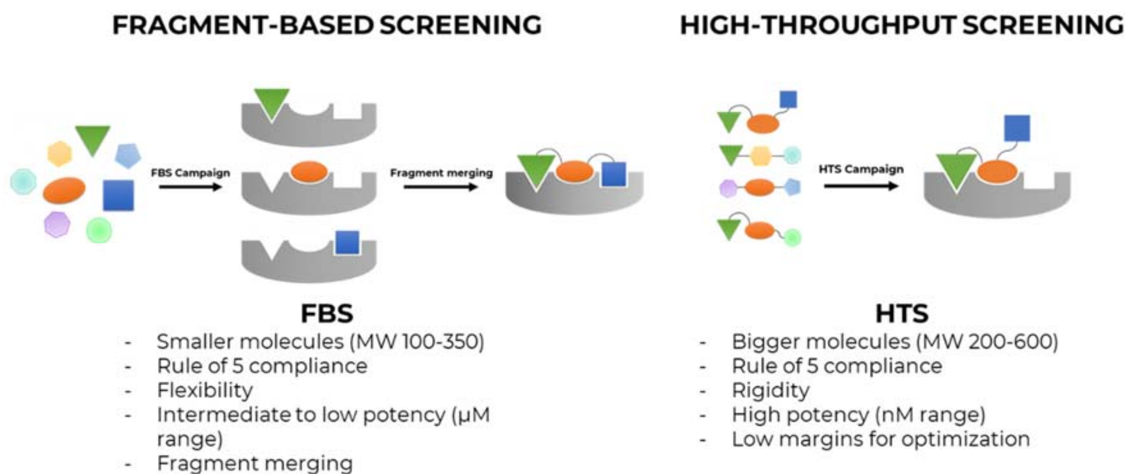


Figure 3. FBS vs HTS. FBS is more versatile than HTS in terms of hit optimization.

X-ray crystallography and nuclear magnetic resonance (NMR) can be used to both identify hits and get structural information about the binding. NMR, working in the solution state, and requiring lower concentrations, is well placed to detect weakly-binding fragments, in an environment more representative of *in vivo* conditions. Since the turn of the last century [53], FBS has grown in use, in both academic and industrial settings, owing to its speed and reliability at testing large chemical space with standard and cost-effective biophysical methods. As such, thousands-strong libraries can be quickly measured against a huge variety of biological targets.

With NMR, both target- and ligand-observed screening approaches are possible [54-59]. From a technical point of view, NMR FBS allows medium to large libraries of fragments (1000 – 5000 fragments) to be efficiently screened. Concerning binding affinity ranges, FBS by NMR allows detecting binding of ligands with affinities as weak as low millimolar [53]. Furthermore, certain experiments may provide helpful insights into the kinetics and dynamics of fragment binding [60].

Ligand-observed NMR screening utilizes simple and sensitive 1D experiments (i.e. STD [59], WaterLOGSY [61], SLAPSTIC [62], T2 and T1rho [63]) and requires no isotopic-labeling of the target protein, simplifying the experimental preparation. In

STD and other ligand-based screening techniques, usually, the polarization is transferred from the solvent (WaterLOGSY) or from the protein (STD) to the ligand, and, depending on the proximity of the binding site the signal of the ligand in the ^{1}H dimension will be affected (the closer the ligand, the higher the intensity of the ligand peaks will be). These ligand-observed NMR experiments are applicable to medium to very high molecular-weight targets. Depending on the ligand affinity, direct or competitive binding is monitored [59, 60, 64].

For smaller proteins (< 30 kDa), protein-observed 2D heteronuclear NMR experiments [65], such as ^{1}H, ^{15}N correlation (HSQC [66], SOFAST-HMQC [67]) are also used [68]. This requires isotope labeling of the target protein - often ^{15}N or ^{13}C alone for small proteins or combined with ^{2}H-labeling [69] for larger proteins [70]. As STD and T1rho, for example, work better with larger proteins as a result of slower tumbling, 2D experiments are especially preferred with small proteins as a reliable procedure that measures the direct effect on protein chemical environment, rather than transferred magnetization or change in ligand properties. It has the additional advantage, in cases where the amino acid shifts are assigned, that it gives an indication of possible binding regions utilizing the peaks' chemical shift perturbations (CSP) [71-73].

Automated analysis tools are available for 1D NMR experiments [74-76] and 2D NMR experiments [77-79]. These tools allow the analysis of 2D NMR titration experiments by tracking CSPs. Following titration assignments to determine rate constants as in NvMap [77], however, is quite different from the needs of fragment screening, which requires a more global view of whether the overall spectrum is significantly altered. In practice, this means that a significant number of peaks alter position or intensity or broaden upon the addition of the fragment or ligand in question. Each spectrum measured with a fragment is overlaid with the reference spectrum and compared directly. This is easily done by eye, but depending on the number of NMR spectra recorded in an FBS campaign, manual analysis can be time-consuming. The other significant issue with manual analysis is the addition of human bias. Whether as a result of tiredness, splitting the analysis over multiple

sessions, or simply as a result of comparison with those already considered, the same spectrum can be classified differently depending on its position within the dataset.

Automation or partial automation of this process would improve efficiency and accuracy, as well as reducing opportunities for the introduction of human bias. It would enable direct comparisons of achieved hit rates with different proteins, as currently these campaigns are often analyzed by different people, leading to differences in subjective evaluation of the spectra. As far as we are aware, fast and reliable automated analysis software for the simultaneous analysis of large numbers of 2D NMR spectra that provides useful binning for hit identification has not been reported in the literature.

1.3. PEX5-PEX14 as a target for drug discovery

1.3.1. The PEX5-PEX14 interface

Peroxisomes are organelles present in eukaryotic cells that acquit different functions, from the metabolism of fatty acids to the scavenging of oxidative species. The proteins involved in the functions of the peroxisomes are named peroxins (PEX) [80]. PEX14 is a transmembrane protein that interacts with PEX5 which is a shuttle peroxin that recognizes the PTS-1 (peroxisome targeting signal) and participates in the active transport inside of the peroxisome of the proteins tagged with the PTS-1 sequence that are subsequently degraded.

PEX5 and PEX14 interface consists mainly of hydrophobic interactions, with a conserved WxxxF(Y) motif that has the tryptophan and the phenylalanine as the 2 main anchors for the PEX5 motif. The binding site of PEX14 presents two hydrophobic pockets that recognize specifically the tryptophan and the phenylalanine [81].

Previous works by Dawidowski et al., have shown that the PEX14 is a suitable target for drug design [82, 83]. Figure 4 summarizes the key interactions of the WxxxF motif and the binding site of PEX14.

The inhibitors reported by Dawidowski et al. have been designed to connect the two hydrophobic aromatic anchors with a pyrazolo[4,3-c]pyridine core. The core provides both the functions of giving rigidity to the compounds and interacting with the two phenylalanines in the center of PEX14 binding pockets.

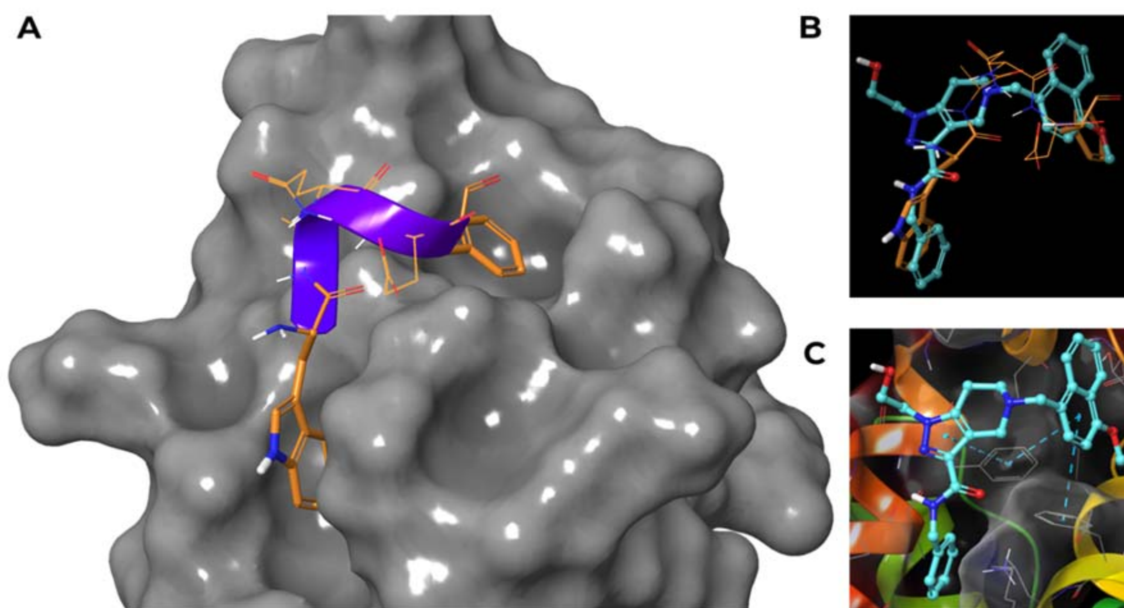


Figure 4. The PEX14 (grey) interface with the WxxxF motif of PEX5 (A). (B) One of the inhibitors reported by Dawidowski et al. overlaid to the WxxxF motif. (C) X-ray structure of the inhibitor bound to PEX14.

1.3.2. PEX14 inhibitors

Building on the work previously done by Emmanoulidis et al. [82], a new series of compounds were synthesized to elucidate the SAR activity of the pyrazolo[4, 3-c] pyridine derivatives.

A first attempt to optimize the lead scaffolds identified as novel binders to PEX14 is here reported in the work by Dawidowski et al. [83]; amongst around 65 compounds listed in the work, some of them shown sub-micromolar IC_{50}'s (down

to 230 nM) against *T.b.* in vitro. Besides the optimization of the compounds, this work reported an exhaustive validation of the binding mode of the pyrazolo[4,3-c]pyridines and elucidate the key interactions of the compounds with the binding site.

In this context, molecular dynamics studies conducted on the new crystal structures obtained during the hit-to-lead optimization phase, shown that pi-stackings between Phe17 and Phe34 and the compounds are crucial in stabilizing the interactions between ligands and PEX14.

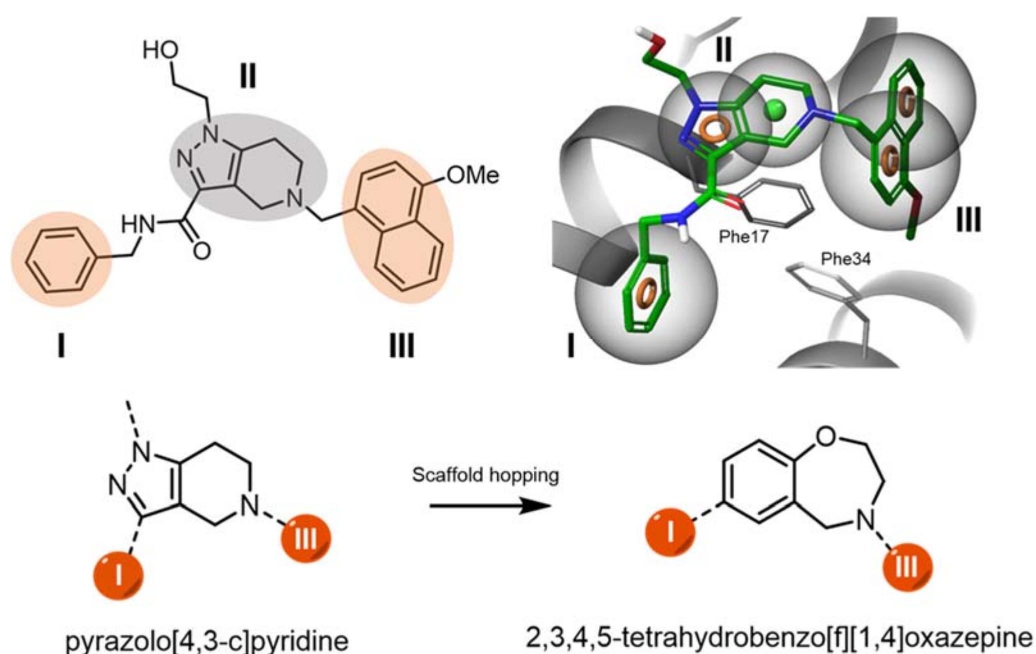


Figure 5. Distribution in space of the pharmacophoric features of Pex14 inhibitors.

In particular, a multi-directional, T-shaped pi-stacking is formed between the pyrazolo[4,3-c]pyridine core (marked as feature II of the pharmacophore), the decoration I and III of the ligands, and the phenylalanines that form the two shallow cavities of the binding site (Figure 5).

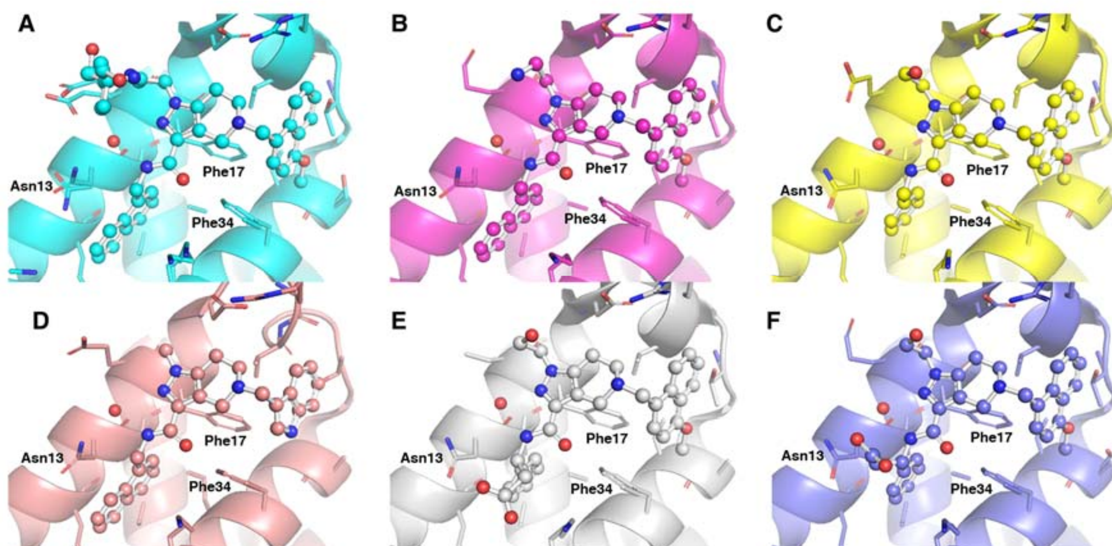


Figure 6. Water coordination found in the X-ray structures of Pex14-inhibitor complexes, PDB codes: 6SPT (A), 5N8V (B), 5L8A (C), 5L87 (D), 5OML (E), 6RT2 (F). The lone pair of the unsubstituted pyrazolo nitrogen is coordinating a stable water molecule in synergy with Asn13.

Besides, the lone pair of the unsubstituted nitrogen in the pyrazolo part of the core is important in the coordination of a water molecule with the nearby Asn13 (Figure 6).

In this work, in addition to NMR, AlphaScreen™ assay (PerkinElmer) was used to evaluate the binding affinity of the compounds toward Tb and Tc PEX14. AlphaScreen™ is a technique based on the production of short-lifetime singlet-state oxygen molecules by beads attached to one interaction partner. Another bead, attached to the second interaction partner, emits light upon reacting with singlet oxygen. The proximity of donor and acceptor beads is used to distinguish between binding and non-binding events [84]. In our case, the assay consists of acceptor beads bound to PEX14 protein and the donor beads to PEX5 WxxxF motif peptide (the natural binder). In the case of a non-binding event, the laser excitation at 680 nm will induce the donor to produce singlet oxygen, which is an extremely oxidative specimen. The singlet oxygen will then react with the thioxene present in the acceptor bead, inducing chemiluminescence at 370 nm. This will subsequently cause the emission of fluorescence in the other fluorophore compounds present in the acceptor beads, which is then recorded by

a detector placed at 90° of the plate. In the case of a binding event, the donor bead is displaced from the protein, thus the emission of fluorescence of the fluorophores will be impeded and the resultant signal will have a lower intensity.
Building on the structural information collected after the extensive optimization of pyrazolo[4,3c]pyridine compounds, scaffold-hopping was performed to identify new compounds as potential binders of PEX14. This effort resulted in the second paper described in chapter 3, leading to the discovery of new series of compounds based on oxazepines aimed at covering yet another portion of the chemical space of PEX14 ligands.

1.3.3. The water envelope around the ligand-PEX14 complex

Water and solvent envelope are crucial but often underestimated players in protein-ligand interactions. Due to the spatial distribution of the features of the constituents of water molecules and the electric dipole induced by the difference in the electronegativity between the oxygen and the hydrogen atoms, water molecules can act as an “extension” in space of directional interactions such as hydrogen bonding. Hydrogen bonding networks are of particular importance in profiling specific interactions and thus fine-tuning the ligand specification to better fulfill the required pharmacophoric features of the binding site.
While, in general, these interactions lead to a negative contribution to the enthalpy component of the total ΔG binding free energy, the reorganization of water molecules in the solvation sheet around the ligand is followed by a decrease in entropy, which is deleterious for binding free energy. Fine-tuning the ligand physicochemical properties to stabilize the water bridges at the interface of the protein-ligand complex can lead to compensation on the loss in entropy due to an extended hydrogen-bonding network.
As shown by Ratkova et al. [85], building on the structural information retrieved by identifying the structural waters found in the x-ray structures of PEX14 bound to pyrazolo[4,3-c] pyridine derivatives, the ligands may be optimized to stabilize the

water solvation sheet around the compounds. Solvation model studies were conducted using a combination of structural biology techniques (X-ray crystallography), binding assays (AlphaScreen™), and molecular simulations based on classical molecular dynamics simulations and 3D Reference Interacting Site Model (3D-RISM) shown that it is possible to improve the binding affinity of the ligands by stabilizing the coordination of nearby water molecules in the solvation sheet of the binding interface.

1.3.4. Aim of the thesis

This work aimed at elucidating the binding dynamics of ligands to the binding site of PEX14, and improving the ligand design by combining structural information collected using a plethora of biophysical methods, from NMR FBS drug discovery to binding assays.
In this context, CADD, molecular modeling, docking, and molecular simulations were extensively used in order to collect the most exhaustive information on the nature of the ligand binding.
To improve the performances of the NMR-based FBS campaign, a new software package that leverages state-of-the-art image-recognition machine-learning-based algorithms was developed and released as a free resource for the other experimentalist in the field.

2. Materials and methods

2.1. Structural biology

2.1.1. Protein expression and purification

Proteins were expressed and prepared using the recombinant expression in E. coli cells as reported in Softley. C. A. work [86]. Plasmids were transformed into DH5α cells for plasmid amplification and the genetic material isolated. For expression, BL21 cells were used. For the use in crystallography and ligand-observed NMR methods, protein without isotope labeling was expressed using an autoinduction medium [87]. For protein-observed NMR experiments, the ^{15}N-labelled protein was produced with a ^{15}N autoinduction medium or, where required, ^{15}N ^{13}C-labelled protein was produced using M9 minimal medium, supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose. After expression, cells were lysed with sonication, and affinity chromatography with Ni or Zn NTA beads was used for initial purification. Size-exclusion chromatography was used for a further purification step. Purity and construct size were checked using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and the concentration was measured using a Thermofisher Nanodrop 2000. All proteins were stored frozen in NMR buffer (HEPES for IMP-13 and phosphate NMR buffer for PEX14) at -80 °C. Fragments were stored at -20 °C at 100 mM in d6-DMSO, while ligands were stored at 50 mM in d6-DMSO, also at -20 °C.

2.1.2. NMR fragment screening

The in-house libraries of fragments were purchased from Maybridge. Aqueous protein solutions of approximately 100 µM concentration with 10% D_2O were added to 8 µL of fragment cocktails dissolved in deuterated DMSO in a 96-well plate with a total volume of 180 µL. Samples of 160 µL were then transferred to

3mm NMR tubes using a Gilson liquid handling robot and SOFAST HMQC [88] or HSQC [89] spectra were recorded at 298K using a Bruker Avance 600 MHz spectrometer equipped with a 5 mm QCI cryo-cooled probe head. A reference sample was made with d6-DMSO instead of fragment solution.

2.2. Biophysical assays

2.2.1. AlphaScreen™ competition assays

Compounds were evaluated for their ability to inhibit PEX5-PEX14 using an optimized protocol for AlphaScreen™ competition-based assay as described in Dawidowski et al. [82]. N-terminal his-tagged PEX14 at the final concentration of 3nM was mixed with 10nM biotinylated PEX5-derived peptide (ALSENWAQEFLA) in a PBS buffer supplemented with 5 mg/mL of BSA and 0.01 % (v/v) Tween-20. 5 μg/mL of streptavidin donor beads and 5 μg/mL of nickel chelate acceptor beads (PerkinElmer) were added to the mixture. Serial dilutions of inhibitors were prepared in DMSO (12 points) and mixed with the above mixture and incubate for 1h in the dark. Doing so the concentration of DMSO was kept constant to 5% (which was shown to not affect the assay readout) and the only variable was the ligand concentration. Data were analyzed using Origin Pro 9.0. Experimental points were interpolate using Hill sigmoidal fitting fixing the asymptotes at the maximal assay signal (no inhibitor added) and 0, respectively, and half-maximal effective concentration (EC_{50}) was calculated for each compound.

2.2.2. NMR spectroscopy

TcPEX14 N-terminal domain backbone assignment was carried out using ^{1}H,^{15}N 2D correlation experiments, 3D HNCA, HNCACB, and CBCACONH experiments [90] recorded on a uniformly ^{15}N,^{13}C-labelled sample in NMR phosphate buffer. The

sample concentration was 600µM and the solvent 90% water, supplemented with 10% D_2O. Experiments were carried out on a Bruker Avance III 800 MHz spectrometer (^1H frequency 800 MHz) equipped with a 5 mm TCI cryoprobe. Data processing was carried out with NMRpipe [91] and visualized with nmrDraw and analysis and assignment were carried out with CCPN Analysis [79].
Compounds were tested using ^1H, ^{15}N 2D correlation spectra on a Bruker Avance III 600MHz spectrometer (^1H frequency 600 MHz) with a QCI cryoprobe). Samples were made up of 200µM uniformly ^{15}N-labelled TcPEX14 protein in phosphate NMR buffer (pH 6.5, 20mM NaCl, 5mM Na_2PO_4) in water, supplemented with 10% D_2O. Compounds, dissolved in d_6-DMSO were added to the test samples at 3:1 and the equivalent volume of d_6-DMSO to the reference sample.

2.2.3. Trypanolytic assays

Trypanocidal activity against bloodstream-form *T.b.* (mammalian stage) was performed by our collaborator Dr. Vishal Khalel at RUB University Bochum and procedures are described in Dawidowski et al. [82]
Logarithmically growing *T.b.* parasites were treated with DMSO alone as a negative control or with 5 µM of the compound tested for 24 hours. Parasites were fixed with formaldehyde and immunofluorescence microscopic analysis was performed essentially as described in Dawidowski et al.[82] with the exception that the primary antibodies against glycosomal GAPDH were used at 1:1500 dilution and images were analyzed using Zen 2.3 Blue edition software (Carl Zeiss).

2.3. Computational methods

2.3.1. Molecular modeling and CADD

Proteins and protein-ligand complexes were inspected using PyMol [92], Schrödinger Maestro 2018.3 [93], and Yasara Structure [15] software. Molecules were designed and modified using either Maestro or ChemDraw18 [94] software.
Protein structures were downloaded from the RCSB Protein Data Bank website and prepared using Schrödinger's Maestro built-in module Protein Preparation, using PROPKA [25, 95] to predict protonation states for sidechains at pH 7.4, add hydrogens and optimize the hydrogen bonding network. In the case of missing sidechains, proteins were prepared first by using the "Clean" command in YASARA Structure.
Ligand libraries were obtained from ZINC12 [33] using ZINCPharmer [96] or AnchorQuery™ [97], and commercial vendors such as ChemDiv, eMolecules, Enamine, and ChemBridge either as SMILES strings or as SDF files. Ligands were prepared using YASARA Structure (based on OpenBabel v2.4.1 [29]), protomers were generated at pH 7.4, and hydrogens were added automatically according to the assigned bond orders.

2.3.2. 2D Fingerprint-based HTVS

Fingerprints and 2D-based similarity searches were performed using the RDKit [98] KNIME [99] nodes release 2017.03, using RDKit AtomPair fingerprinting starting from 2D SMILES [100] strings generated by RDKit reading nodes.
Structures in SDF files were loaded using the KNIME SDF reader node and passed through the "MoleculeToRDKit" node to filter out broken molecules. SMILES strings were parsed using the "Read File" KNIME node and then send to the "MoleculeToRDKit" node to generate 2D structures and filter broken molecules.

The AtomPair fingerprint was selected as the molecular pathway descriptor and Tanimoto distance [101] was the metric selected as a similarity index.

2.3.3. Docking and binding affinity predictions

Docking calculations were performed on a standard workstation (Intel® Core™ i7-5960X, 8 cores/16 threads). All molecular dynamics simulations were performed on NVIDIA GTX 1070. Docking experiments were performed using VINA [102] and the setup was carried out with the YASARA molecular modeling program. Docking was performed against the available X-ray structure of TbPEX14 (PDB accession code: 5L87). The protein was prepared for docking using Yasara Structure, adding missing sidechains, missing hydrogens, and generating protonation states using Yasara's built-in "Clean" command. The docking box was generated around the center of mass of the existing ligand, keeping a distance of 5Å from the center of mass of the ligand atoms. The cell generated had a total volume of 9.625$Å^3$ (X-axis = 23.0Å, Y-axis = 27.0Å, Z-axis = 15.5Å; $\alpha = \beta = \gamma = 90°$), to accommodate larger ligands than the one in the original structure, and to avoid artifacts in pose generation due to steric constraints.

Compounds' best poses were selected according to VINA binding energy estimation, and then rescored using VINA after short (500 ps for each simulation followed by simulated annealing) in vacuo [103] and solvated complex using AMBER14 force field [104] molecular dynamics simulations to assess both the ligand pose stability and the quality of fit of the pharmacophore model.

A comprehensive example macro script for YASARA structure combining Docking and MD simulations is available in the supporting information of the article *Computer-Aided Design and Synthesis of a New Class of PEX14 Inhibitors: Substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepines as Potential New Trypanocidal Agents*.

2.3.4. Molecular dynamics simulations

The PDB files of the different ligand-protein complexes were prepared by adding missing sidechains and hydrogens using YASARA Biosciences YASARA Structure's clean built-in command. Structures were then imported into Schrödinger Maestro version 2018.3 and further refined using the Maestro v11.1 "Protein Preparation Wizard". Protonation states were calculated using PROPKA at pH 7.0 ± 2.0 and minimization of hydrogen positions with restrained backbone was performed using OPLS_2005 FF [105] to optimize the hydrogen bonding network. All the systems were then prepared for simulation using Maestro v11.1 "System Builder" GUI using a TIP4P solvent model [105, 106] (crystallographic water molecules were deleted) in an automatically generated cubic cell with periodic boundary conditions. In addition to the solvated complex, Na+ and Cl– ions corresponding to a 150 mM buffer were placed in the cell to set the total net charge to zero. MDs simulations were run using Schrödinger Maestro Desmond Molecular Dynamics Package version 2018.3 [12] on an NVIDIA 1070 graphics processing unit using the Desmond graphical user interface for a total simulation time of 50 ns to ensure system convergence, with XYZ coordinates recording interval every 50 ps (1000 snapshots in total) and 1.2 ps for potential energy calculations of the ensemble. Ensemble class was set to NPT at a temperature of 300 K and pressure of 1.01 bar, force cutoff radius was set to 9.0 Å and each solvated model was relaxed with the Desmond default relaxation protocol before starting the actual simulation [12].

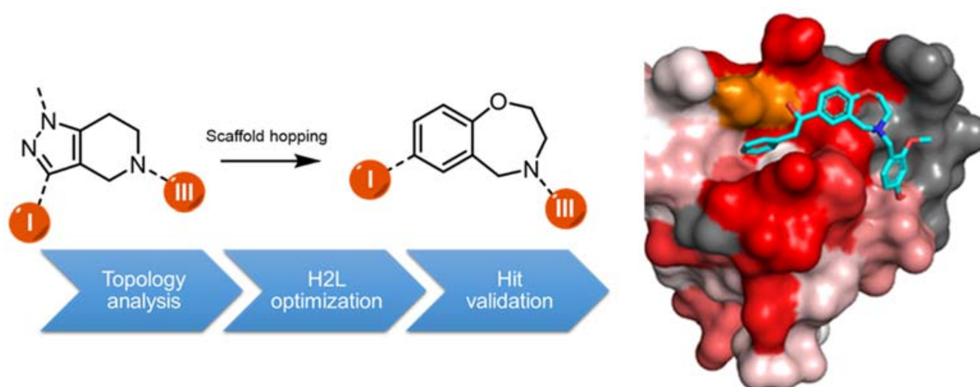
3. Publications

This is a publication-based thesis. The works published in international peer-reviewed journals are summarized in this section of the thesis. Some of the works cited in these papers are partially reported in the Bibliography section and referenced when appropriate.

3.1. Computer-Aided Design and Synthesis of a New Class of PEX14 Inhibitors: Substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepines as Potential New Trypanocidal Agents.

The article Computer-Aided Design and Synthesis of a New Class of PEX14 Inhibitors: Substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepines as Potential New Trypanocidal Agents *has been published by the Journal of Computational Information and Modeling (JCIM). Roberto Fino shares the first authorship of this paper with Dr. Dominik Lenhart; Roberto Fino implemented the 2D CATS HTVS of libraries available at ETH Modlab, selected the virtual hits, designed and performed the docking/MD screening of the HTVS output, inspected and selected the core-hopping results, prepared the chemoinformatic framework for analysis results, prepared, organized, and edited the manuscript. Dr. Dominik Lenhart designed, synthesized, and characterized the oxazepine derivatives of the virtual hits found.*

With the progressive raising of the temperatures and the menace of global warming, neglected tropical diseases (NTDs) are supposed to become endemic and widespread even in the EU and USA. In this perspective, African- and American trypanosomiases are amongst some of the most dangerous diseases which urge the development of new treatments. Building on the previous work by Popowicz and Sattler AGs on validating the proof-of-concept of the druggability of PEX14 [80], this work aimed to perform 2D scaffold-hopping using the CATS algorithm [48] to explore the chemical space in search of new compounds that could be designated as new binders for PEX14.

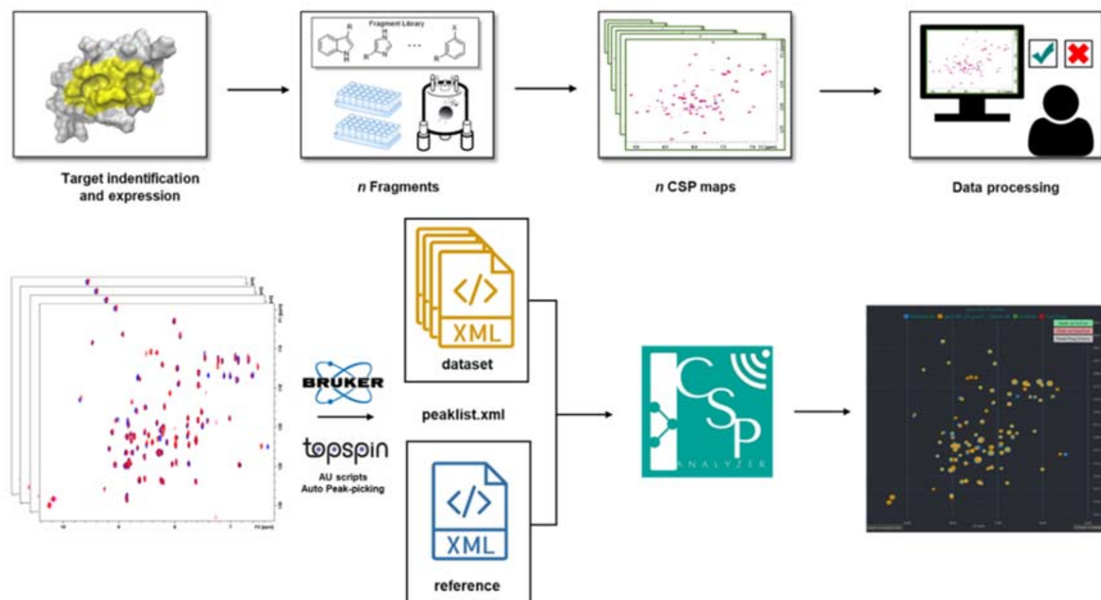


In this work, we used a synergy of structural biology, computer-aided drug design, and medicinal chemistry to design, optimize, and assay a library of substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepines against T.b. PEX14. A general library of commercially available compounds from different sources (Asinex, ChemBridge, Enamine, etc.) was screened using the CATS fingerprint similarity using as a reference the best pyrazolo [4, 3-c] pyridine-based compound available from previous hit-to-lead optimization efforts [83]. The screening led to a shortlist of 37 unique core scaffolds from which were subsequently selected 3 main compounds. Amongst these, it was identified as the first substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepine. This compound was screened then for similar compounds on the ChemBridge database to find analogs. After manual inspection and CADD, a total of 23 molecules were selected for the first QSAR evaluation. A synthesis protocol to obtain the oxazepine-based compounds was established and the compounds were subsequently assayed using AlphaScreen™ and trypanocidal immunofluorescence assays. The best compound in this series showed an IC_{50} of 4 μM against tbPEX14. Besides, the first assessment of ADME properties was conducted to evaluate the physicochemical properties of the best oxazepine candidates. While the compounds are still in the hit-to-lead optimization and showed some metabolic liabilities, this work stands as an excellent example of how an early-stage drug design project may benefit from a workflow designed combining different techniques, ranging from classical structural biology to chemoinformatic and advanced HTVS protocols.

3.2. Introducing the CSP Analyzer: A novel Machine Learning-based application for automated analysis of two-dimensional NMR spectra in NMR fragment-based screening.

The article Introducing the CSP Analyzer: A novel Machine Learning-based application for automated analysis of two-dimensional NMR spectra in NMR fragment-based screening *has been published in the Computational and Structural Biotechnology Journal and is available open access at the address:* https://doi.org/10.1016/j.csbj.2020.02.015. *Roberto Fino, the author of this thesis, shares the first authorship of this paper with Dr. Ryan Byrne; Roberto Fino developed the idea original idea of Dr. Grzegorz Popowicz. He designed and coded the UI of the front-end of the CSP-Analyzer in C#, expressed some of the proteins cited in the article, performed NMR FBDD, collected, refined, processed, and manually analyzed the NMR spectra reported in the study, tested and validated the machine-learning framework deployed in the python backend, prepared and edited part of the manuscript. Citations of this paper are included in paragraph 1.2.1.*

The analysis capabilities of the results of NMR FBS techniques have not coped with the amount of data generated during FBDD campaigns. This work is aimed at providing NMR scientists with fast, reliable, and automated software for analyzing hundreds of 2D NMR spectra resulting from FBS campaigns. The CSP-Analyzer is a software program written in C# and python that deploys advanced machine-learning-driven statistical discrimination to identify as fast as possible interacting fragments with a given target of interest.



The CSP Analyzer features a C# frontend that is interfaced with a Python machine learning classifier and is available free-of-charge on GitHub for both commercial and academic uses: https://github.com/rubbs14/CSP-Analyzer/releases/tag/v1.0 under the GPL license v3.0.

According to the principles of target-based 2D NMR FBS, the software is designed to identify the “active” fragments by comparing each spectrum (protein with fragment or fragment pool) in a screening set to its reference spectrum (protein only). This operation does not require the assignment of the peaks because the statistics are based on the fingerprinting of each spectrum's properties (peaks position, peaks scattering, intensities, etc.). In other words, the algorithms included in the package are capable of indicating, with good recall, the “most different” experiments to the reference spectrum. The software is designed to enable either partial or fully automated hit determination from measured 2D spectra, depending on the user’s requirements, thus being a versatile tool for FBS campaigns by NMR.

Featuring a rich and user-friendly interface, the CSP-Analyzer uses XML peak lists generated with Bruker TopSpin and converts them to scatterplots mapped

by ^{1}H and ^{15}N ppm and intensities. It fast compares hundreds of spectra of screened fragments against a given reference using a synergy of image-processing and CNN algorithms to identify the spectra that are most dissimilar to the reference. The analysis of the peak lists can be conducted in a fully automated mode but the user can manually assign a flag according to its choices. After the analysis is completed, the results can be exported to PDF or Excel files for further processing or data storage. This work represents the first attempt at deploying machine-learning image recognition techniques to 2D NMR FBS data analysis.

4. Conclusions and outlook

4.1. Automatic NMR FBS spectra processing

The reported method for spectra analysis is capable of discriminating and recalling with good reliability the spectra which show sufficient change, promoting greater efficiency in screening analysis, and reduction of human bias induced by the repetitive nature of the task. While the program currently relies on third-party software for peak picking, future versions might directly utilize raw NMR time-domain data to create a comprehensive NMR analysis tool for FBDD. Allowing the user to adjust the intensity levels to their needs and set the thresholds for peak picking, would be a logical extension of the GUI. We also considered implementing functionalities to bridge the visualization module with available NMR data processing and visualization tools currently available, such as NMRPipe [91], PINT [107], NMRView [108], Sparky [109], and CCPNMR [79]. While the CSP Analyzer was designed to screen HSQC ^{1}H-^{15}N spectra, it could also be used for ^{1}H-^{13}C or any other bidimensional experiment. To facilitate this, a module for automatic recognition of the F1 and F2 dimensions could be added. We also considered the development of a KNIME [99] node for the CSP Analyzer which could then be interfaced with a general workflow for batch NMR data processing. The backend utilizes several established computer vision approaches, coupled with machine learning, to achieve solid results in the recall of active-labeled spectra. Future work might replace this image-based approach with a model that can directly analyze waveform data [110], perhaps combined with some more direct means of determining which noise level can be safely ignored [111].

Overall, however, the relatively simple framework set out here sufficed to achieve good recall on a laborious task. It is uncertain how much of the gap from perfect recall to that achieved is owing to ambiguity and variation between experts in the analysis of these spectra, but, in general, we could demonstrate that the speed

of NMR FBDD data analysis can be greatly improved using our implementation of machine learning methods without a substantial loss in terms of accuracy.

4.2. PEX14 Ligand optimization

Thanks to the efforts reported in the previous chapter, it has been shown further evidence that PEX14 is a suitable drug target for the development of new lead compounds against trypanosomiasis. The results show that this can be approached by exploring the chemical space available using different scaffolds with well-known, reliable, and simple chemistry.

It has been demonstrated that scaffold hopping of pyrazolo[4,3-c]pyridine derivatives using the CATS algorithm for the HTVS campaign was successful in highlighting new strategies in the early-stage drug development of new ligands for the PEX5-PEX14 PPI. Established molecular modeling workflows such as docking and molecular dynamics guided the selection of the best poses that were then selected for the next steps in compound optimization.

The reported best hits further characterized the best hits using a combination of biophysical assays (NMR, AlphaScreen™, and trypanocidal assays) and run the first campaign of chemical synthesis to produce optimized compounds for the SAR analysis, leading to a trypanosomal cellular IC_{50} of 4 μM against the bloodstream form of *T.b.* for the best compound in the library (**7a**).

Our substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepine compounds are a promising starting point for further lead optimization. While the oxazepine family of compounds has shown promising activity, they still are in the hit optimization phase. Metabolism liabilities, such as the potential hydrolysis of the C-O bond in the oxazepine ring, still need to be fully addressed and resolved, but the alternatives provided by the scaffold-hopping protocols stand as an excellent starting point for the discovery of new leads in the development of new trypanocidal agents.

5. Bibliography

1. Kirsch, D. R.; Ogas, O., *The Drug Hunters: The Improbable Quest to Discover New Medicines*. Simon and Schuster: 2016.
2. Comba, P.; Hambley, T. W.; Martin, B., *Molecular Modeling of Inorganic Compounds*. John Wiley & Sons: 2009.
3. Richon, A. B., An Introduction to Molecular Modeling. *Mathematech* **1994,** *1*, 83.
4. DeLano, W. L., Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsletter on protein crystallography* **2002,** *40*, 82-92.
5. Seeliger, D.; de Groot, B. L., Ligand Docking and Binding Site Analysis with Pymol and Autodock/Vina. *J. Comput. Aided Mol. Des.* **2010,** *24*, 417-422.
6. Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14*, 33-38.
7. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with Namd. *J. Comput. Chem.* **2005,** *26*, 1781-1802.
8. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., Ucsf Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004,** *25*, 1605-1612.
9. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminform.* **2012,** *4*, 17.
10. Jmol, J., An Open-Source Java Viewer for Chemical Structures in 3d. *Jmol web page: http://www. jmol. org/, last accessed* **2013,** *15*.
11. Release, S., 1: Maestro. *Schrödinger, LLC, New York, NY* **2017,** *2017*.
12. Release, S., 4: Desmond Molecular Dynamics System. *DE Shaw Research: New York, NY* **2017**.
13. Release, S., 4: Glide. *Schrödinger, LLC, New York, NY* **2016**.

14. Inc., C. C. G., Molecular Operating Environment (Moe). Chemical Computing Group Inc 1010 Sherbooke St. West, Suite# 910, Montreal …: 2016.

15. Krieger, E.; Vriend, G.; Spronk, C., Yasara–yet Another Scientific Artificial Reality Application. *YASARA. org* **2013,** *993*.

16. Harvey, M. J.; Giupponi, G.; Fabritiis, G. D., Acemd: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009,** *5*, 1632-1639.

17. Doerr, S.; Harvey, M.; Noé, F.; De Fabritiis, G., Htmd: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* **2016,** *12*, 1845-1852.

18. Martinez-Rosell, G.; Lovera, S.; Sands, Z. A.; De Fabritiis, G., Playmolecule Crypticscout: Predicting Protein Cryptic Sites Using Mixed-Solvent Molecular Simulations. *J. Chem. Inf. Model.* **2020,** *60*, 2314-2324.

19. Skalic, M.; Martínez-Rosell, G.; Jiménez, J.; De Fabritiis, G., Playmolecule Bindscope: Large Scale Cnn-Based Virtual Screening on the Web. *Bioinformatics* **2019,** *35*, 1237-1238.

20. Martínez-Rosell, G.; Giorgino, T.; De Fabritiis, G., Playmolecule Proteinprepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2017,** *57*, 1511-1516.

21. Norrby, M.; Grebner, C.; Eriksson, J.; Bostrom, J., Molecular Rift: Virtual Reality for Drug Designers. *J. Chem. Inf. Model.* **2015,** *55*, 2475-2484.

22. Juru, A. U.; Hargrove, A. E., Frameworks for Targeting Rna with Small Molecules. *J. Biol. Chem.* **2021**, 100191.

23. Stefaniak, F.; Bujnicki, J. M., Annapurna: A Scoring Function for Predicting Rna-Small Molecule Binding Poses. *PLoS Comput. Biol.* **2021,** *17*, e1008309.

24. Warner, K. D.; Hajdin, C. E.; Weeks, K. M., Principles for Targeting Rna with Drug-Like Small Molecules. *Nature Reviews Drug Discovery* **2018,** *17*, 547-558.

25. Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., Propka3: Consistent Treatment of Internal and Surface Residues in Empirical P K a Predictions. *J. Chem. Theory Comput.* **2011,** *7*, 525-537.

26. Petukh, M.; Stefl, S.; Alexov, E., The Role of Protonation States in Ligand-Receptor Recognition and Binding. *Curr. Pharm. Des.* **2013,** *19*, 4182-4190.

27. Liao, C.; Nicklaus, M. C., Comparison of Nine Programs Predicting P K a Values of Pharmaceutical Substances. *J. Chem. Inf. Model.* **2009,** *49*, 2801-2812.

28. Ropp, P. J.; Kaminsky, J. C.; Yablonski, S.; Durrant, J. D., Dimorphite-Dl: An Open-Source Program for Enumerating the Ionization States of Drug-Like Small Molecules. *J. Cheminform.* **2019,** *11*, 1-8.

29. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011,** *3*, 33.

30. Moffat, J. G.; Vincent, F.; Lee, J. A.; Eder, J.; Prunotto, M., Opportunities and Challenges in Phenotypic Drug Discovery: An Industry Perspective. *Nature reviews Drug discovery* **2017,** *16*, 531-543.

31. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A., Highly Accurate Protein Structure Prediction with Alphafold. *Nature* **2021,** *596*, 583-589.

32. Wermuth, C.-G.; Ganellin, C.; Lindberg, P.; Mitscher, L., Glossary of Terms Used in Medicinal Chemistry (Iupac Recommendations 1998). *Pure and Applied Chemistry. Chimie Pure et Appliquee* **1998,** *70*, 1129-1143.

33. Irwin, J. J.; Shoichet, B. K., Zinc− a Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005,** *45*, 177-182.

34. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G., Zinc: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012,** *52*, 1757-1768.

35. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E., The Chembl Database in 2017. *Nucleic Acids Res.* **2017,** *45*, D945-D954.

36. Hattori, M.; Okuno, Y.; Goto, S.; Kanehisa, M., Development of a Chemical Structure Comparison Method for Integrated Analysis of Chemical and Genomic Information in the Metabolic Pathways. *J. Am. Chem. Soc.* **2003,** *125*, 11853-11865.

37. Muchmore, S. W.; Debe, D. A.; Metz, J. T.; Brown, S. P.; Martin, Y. C.; Hajduk, P. J., Application of Belief Theory to Similarity Data Fusion for Use in Analog Searching and Lead Hopping. *J. Chem. Inf. Model.* **2008,** *48*, 941-948.

38. Van Berlo, R. J.; Winterbach, W.; De Groot, M. J.; Bender, A.; Verheijen, P. J.; Reinders, M. J.; De Ridder, D., Efficient Calculation of Compound Similarity Based on Maximum Common Subgraphs and Its Application to Prediction of Gene Transcript Levels. *Int. J. Bioinform. Res. Appl.* **2013,** *9*, 407-432.

39. Gabb, H. A.; Jackson, R. M.; Sternberg, M. J., Modelling Protein Docking Using Shape Complementarity, Electrostatics and Biochemical Information. *J. Mol. Biol.* **1997,** *272*, 106-120.

40. Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A., New Methods for Ligand-Based Virtual Screening: Use of Data Fusion and Machine Learning to Enhance the Effectiveness of Similarity Searching. *J. Chem. Inf. Model.* **2006,** *46*, 462-470.

41. Willett, P., Similarity-Based Virtual Screening Using 2d Fingerprints. *Drug Discovery Today* **2006,** *11*, 1046-1053.

42. Bender, A.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C. K.; Glick, M.; Davies, J. W., How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space. *J. Chem. Inf. Model.* **2009,** *49*, 108-119.

43. Koutsoukas, A.; Paricharak, S.; Galloway, W. R.; Spring, D. R.; IJzerman, A. P.; Glen, R. C.; Marcus, D.; Bender, A., How Diverse Are Diversity Assessment Methods? A Comparative Analysis and Benchmarking of Molecular Descriptor Space. *J. Chem. Inf. Model.* **2014,** *54*, 230-242.

44. Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G., Molecular Fingerprint Similarity Search in Virtual Screening. *Methods* **2015,** *71*, 58-63.

45. Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A., Comparison of Fingerprint-Based Methods for Virtual Screening Using Multiple Bioactive Reference Structures. *J. Chem. Inf. Comput. Sci.* **2004,** *44*, 1177-1185.

46. Bajusz, D.; Rácz, A.; Héberger, K., Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminform.* **2015,** *7*, 1-13.

47. Rogers, D.; Hahn, M., Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010,** *50*, 742-754.

48. 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.* **2013,** *32*, 133.

49. Fontaine, F.; Bolton, E.; Borodina, Y.; Bryant, S. H., Fast 3d Shape Screening of Large Chemical Databases through Alignment-Recycling. *Chem. Cent. J.* **2007,** *1*, 1-14.

50. Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M., Plip: Fully Automated Protein–Ligand Interaction Profiler. *Nucleic Acids Res.* **2015,** *43*, W443-W447.

51. Wójcikowski, M.; Kukiełka, M.; Stepniewska-Dziubinska, M. M.; Siedlecki, P., Development of a Protein–Ligand Extended Connectivity (Plec) Fingerprint and Its Application for Binding Affinity Predictions. *Bioinformatics* **2019,** *35*, 1334-1341.

52. Hann, M. M.; Leach, A. R.; Harper, G., Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2001,** *41*, 856-864.

53. Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R., Fragment-Based Lead Discovery. *Nature Reviews Drug Discovery* **2004,** *3*, 660.

54. Hajduk, P. J.; Augeri, D. J.; Mack, J.; Mendoza, R.; Yang, J.; Betz, S. F.; Fesik, S. W., Nmr-Based Screening of Proteins Containing 13c-Labeled Methyl Groups. *J. Am. Chem. Soc.* **2000,** *122*, 7898-7904.

55. Gossert, A. D.; Jahnke, W., Nmr in Drug Discovery: A Practical Guide to Identification and Validation of Ligands Interacting with Biological Macromolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016,** *97*, 82-125.

56. Jhoti, H.; Cleasby, A.; Verdonk, M.; Williams, G., Fragment-Based Screening Using X-Ray Crystallography and Nmr Spectroscopy. *Curr. Opin. Chem. Biol.* **2007,** *11*, 485-493.

57. Ludwig, C.; Guenther, U. L., Ligand Based Nmr Methods for Drug Discovery. *Front. Biosci.* **2009,** *14*, 24.

58. Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference Nmr Spectroscopy. *Angew. Chem. Int. Ed.* **1999,** *38*, 1784-1788.

59. 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. Educ.* **2011,** *88*, 990-994.

60. Angulo, J.; Enríquez‐Navas, P. M.; Nieto, P. M., Ligand–Receptor Binding Affinities from Saturation Transfer Difference (Std) Nmr Spectroscopy: The Binding Isotherm of Std Initial Growth Rates. *Chemistry–A European Journal* **2010,** *16*, 7803-7812.

61. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B., Waterlogsy as a Method for Primary Nmr Screening: Practical Aspects and Range of Applicability. *J. Biomol. NMR* **2001,** *21*, 349-359.

62. Jahnke, W.; Rüdisser, S.; Zurini, M., Spin Label Enhanced Nmr Screening. *J. Am. Chem. Soc.* **2001,** *123*, 3149-3150.

63. Hajduk, P. J.; Olejniczak, E. T.; Fesik, S. W., One-Dimensional Relaxation- and Diffusion-Edited Nmr Methods for Screening Compounds That Bind to Macromolecules. *J. Am. Chem. Soc.* **1997,** *119*, 12257-12261.

64. Wang, Y. S.; Liu, D.; Wyss, D. F., Competition Std Nmr for the Detection of High - Affinity Ligands and Nmr - Based Screening. *Magn. Reson. Chem.* **2004,** *42*, 485-489.

65. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W., Discovering High-Affinity Ligands for Proteins: Sar by Nmr. *Science* **1996,** *274*, 1531-1534.

66. Bodenhausen, G.; Ruben, D. J., Natural Abundance Nitrogen-15 Nmr by Enhanced Heteronuclear Spectroscopy. *Chem. Phys. Lett.* **1980,** *69*, 185-189.

67. 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.* **2005,** *127*, 8014-8015.

68. Amero, C.; Schanda, P.; Durá, M. A.; Ayala, I.; Marion, D.; Franzetti, B.; Brutscher, B.; Boisbouvier, J., Fast Two-Dimensional Nmr Spectroscopy of High Molecular Weight Protein Assemblies. *J. Am. Chem. Soc.* **2009,** *131*, 3448-3449.

69. Sattler, M.; Fesik, S. W., Use of Deuterium Labeling in Nmr: Overcoming a Sizeable Problem. *Structure* **1996,** *4*, 1245-1249.

70. Gardner, K. H.; Kay, L. E., The Use of 2h, 13c, 15n Multidimensional Nmr Gto Study the Structure and Dynamics of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **1998,** *27*, 357-406.

71. Chuang, S.; Velkov, T.; Horne, J.; Porter, C. J.; Scanlon, M. J., Characterization of the Drug Binding Specificity of Rat Liver Fatty Acid Binding Protein. *J. Med. Chem.* **2008,** *51*, 3755-3764.

72. Patil, R.; Laguerre, A.; Wielens, J.; Headey, S. J.; Williams, M. L.; Hughes, M. L.; Mohanty, B.; Porter, C. J.; Scanlon, M. J., Characterization of Two Distinct Modes of Drug Binding to Human Intestinal Fatty Acid Binding Protein. *ACS Chem. Biol.* **2014,** *9*, 2526-2534.

73. Sundell, G. N.; Vögeli, B.; Ivarsson, Y.; Chi, C. N., The Sign of Nuclear Magnetic Resonance Chemical Shift Difference as a Determinant of the Origin of Binding Selectivity: Elucidation of the Position Dependence of Phosphorylation in Ligands Binding to Scribble Pdz1. *Biochemistry* **2017,** *57*, 66-71.

74. East, K. W.; Leith, A.; Ragavendran, A.; Delaglio, F.; Lisi, G. P., Nmrdock: Lightweight and Modular Nmr Processing. *bioRxiv* **2019**, 679688.

75. Peng, C.; Frommlet, A.; Perez, M.; Cobas, C.; Blechschmidt, A.; Dominguez, S.; Lingel, A., Fast and Efficient Fragment-Based Lead Generation by Fully Automated Processing and Analysis of Ligand-Observed Nmr Binding Data. *J. Med. Chem.* **2016,** *59*, 3303-3310.

76. Willcott, M. R., Mestre Nova. ACS Publications: 2009.

77. Fukui, L.; Chen, Y., Nvmap: Automated Analysis of Nmr Chemical Shift Perturbation Data. *Bioinformatics* **2006,** *23*, 378-380.

78. Peng, C.; Unger, S. W.; Filipp, F. V.; Sattler, M.; Szalma, S., Automated Evaluation of Chemical Shift Perturbation Spectra: New Approaches to Quantitative Analysis of Receptor-Ligand Interaction Nmr Spectra. *J. Biomol. NMR* **2004,** *29*, 491-504.

79. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D., The Ccpn Data Model for Nmr Spectroscopy: Development of a Software Pipeline. *Proteins: Structure, Function, and Bioinformatics* **2005,** *59*, 687-696.

80. Emmanouilidis, L.; Gopalswamy, M.; Passon, D. M.; Wilmanns, M.; Sattler, M., Structural Biology of the Import Pathways of Peroxisomal Matrix Proteins. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **2016,** *1863*, 804-813.

81. Neuhaus, A.; Kooshapur, H.; Wolf, J.; Meyer, N. H.; Madl, T.; Saidowsky, J.; Hambruch, E.; Lazam, A.; Jung, M.; Sattler, M., A Novel Pex14 Protein-Interacting Site of Human Pex5 Is Critical for Matrix Protein Import into Peroxisomes. *J. Biol. Chem.* **2014,** *289*, 437-448.

82. Dawidowski, M.; Emmanouilidis, L.; Kalel, V. C.; Tripsianes, K.; Schorpp, K.; Hadian, K.; Kaiser, M.; Maser, 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* **2017,** *355*, 1416-1420.

83. Dawidowski, M.; Kalel, V. C.; Napolitano, V.; Fino, R.; Schorpp, K.; Emmanouilidis, L.; Lenhart, D.; Ostertag, M.; Kaiser, M.; Kolonko, M., Structure-Activity Relationship in Pyrazolo [4, 3-C] Pyridines, First Inhibitors of Pex14-Pex5 Protein-Protein Interaction (Ppi) with Trypanocidal Activity. *J. Med. Chem.* **2019**.

84. Eglen, R. M.; Reisine, T.; Roby, P.; Rouleau, N.; Illy, C.; Bossé, R.; Bielefeld, M., The Use of Alphascreen Technology in Hts: Current Status. *Curr. Chem. Genomics* **2008,** *1*, 2.

85. Ratkova, E. L.; Dawidowski, M.; Napolitano, V.; Dubin, G.; Fino, R.; Ostertag, M. S.; Sattler, M.; Popowicz, G.; Tetko, I. V., Water Envelope Has a Critical Impact on the Design of Protein–Protein Interaction Inhibitors. *Chem. Commun.* **2020,** *56*, 4360-4363.

86. Softley, C. Development and Application of Nmr-Spectroscopy and X-Ray Crystallography in Early-Stage Drug Discovery. Technische Universität München, 2020.

87. Studier, F. W., Protein Production by Auto-Induction in High-Density Shaking Cultures. *Protein Expr. Purif.* **2005,** *41*, 207-234.

88. Schanda, P.; Kupče, Ē.; Brutscher, B., Sofast-Hmqc Experiments for Recording Two-Dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds. *J. Biomol. NMR* **2005,** *33*, 199-211.

89. Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V., Gradient-Tailored Water Suppression for 1h-15n Hsqc Experiments Optimized to Retain Full Sensitivity. *Journal of Magnetic Resonance, Series A* **1993,** *102*, 241-245.

90. Sattler, M.; Schleucher, J.; Griesinger, C., Heteronuclear Multidimensional Nmr Experiments for the Structure Determination of Proteins in Solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999,** *34*, 93-158.

91. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., Nmrpipe: A Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR* **1995,** *6*, 277-293.

92. Dong, E.; Du, H.; Gardner, L., An Interactive Web-Based Dashboard to Track Covid-19 in Real Time. *The Lancet infectious diseases* **2020**.

93. Release, S., 1: Maestro. *Schrödinger, LLC: New York, NY, USA* **2018**.

94. Cousins, K. R., Computer Review of Chemdraw Ultra 12.0. ACS Publications: 2011.

95. Bas, D. C.; Rogers, D. M.; Jensen, J. H., Very Fast Prediction and Rationalization of Pka Values for Protein–Ligand Complexes. *Proteins: Structure, Function, and Bioinformatics* **2008,** *73*, 765-783.

96. Koes, D. R.; Camacho, C. J., Zincpharmer: Pharmacophore Search of the Zinc Database. *Nucleic Acids Res.* **2012,** *40*, W409-W414.

97. Koes, D. R.; Dömling, A.; Camacho, C. J., A Nchor Q Uery: R Apid Online Virtual Screening for Small‐Molecule Protein–Protein Interaction Inhibitors. *Protein Sci.* **2018,** *27*, 229-232.

98. Landrum, G., Rdkit: A Software Suite for Cheminformatics, Computational Chemistry, and Predictive Modeling. Academic Press: 2013.

99. Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B., Knime-the Konstanz Information Miner: Version 2.0 and Beyond. *AcM SIGKDD explorations Newsletter* **2009,** *11*, 26-31.

100. Weininger, D., Smiles, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988,** *28*, 31-36.

101. Bajusz, D.; Rácz, A.; Héberger, K., Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminform.* **2015,** *7*, 20.

102. Trott, O.; Olson, A. J., Autodock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010,** *31*, 455-61.

103. Krieger, E.; Koraimann, G.; Vriend, G., Increasing the Precision of Comparative Models with Yasara Nova--a Self-Parameterizing Force Field. *Proteins* **2002,** *47*, 393-402.

104. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., Ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99sb. *J. Chem. Theory Comput.* **2015,** *11*, 3696-713.

105. Jorgensen, W. L.; Tirado-Rives, J., The Opls [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988,** *110*, 1657-1666.

106. Jorgensen, W. L.; Madura, J. D., Temperature and Size Dependence for Monte Carlo Simulations of Tip4p Water. *Mol. Phys.* **1985,** *56*, 1381-1392.

107. Ahlner, A.; Carlsson, M.; Jonsson, B.-H.; Lundström, P., Pint: A Software for Integration of Peak Volumes and Extraction of Relaxation Rates. *J. Biomol. NMR* **2013,** *56*, 191-202.

108. Johnson, B. A., Using Nmrview to Visualize and Analyze the Nmr Spectra of Macromolecules. In *Protein Nmr Techniques*, Springer: 2004; pp 313-352.

109. Lee, W.; Tonelli, M.; Markley, J. L., Nmrfam-Sparky: Enhanced Software for Biomolecular Nmr Spectroscopy. *Bioinformatics* **2014,** *31*, 1325-1327.

110. Oord, A. v. d.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K., Wavenet: A Generative Model for Raw Audio. *arXiv preprint arXiv:1609.03499* **2016**.

111. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.-A., Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of machine learning research* **2010,** *11*, 3371-3408.

Acknowledgments

One field in which Italians are extremely good at, besides cooking, is greeting people. We spend most of our time greeting each other, and that is not only for elevating the gentle art of procrastination but also because we value the people we met. So, beware: this part of my thesis may need some time to read through, as it can be considered a chapter on its own.

I am incredibly grateful to my supervisors Dr. Grzegorz Popowicz and Prof. Dr. Michael Sattler for granting me the opportunity to work on this project. I could have not desired any better supervision in all these years of my Ph.D. work. They provided me with a fantastic chance to express my creativity and strive toward being a better scientist and a better person. In particular, I would like to say that working with you was an honor for me and, most importantly, it was great fun. I cannot describe by words how grateful I am to you.

So long, and thanks for all the fish!

I want to say thank you to my family and my sisters for without them I would not be here today.

Thanks to Antonella Rella, my one and only.

These scars long have yearned for your tender caress
To bind our fortunes, damn what the stars own
Rend my heart open, then your love profess
A winding, weaving fate to which we both atone

You flee my dream come the morning
Your scent, berries tart, lilac sweet
To dream of raven locks entwisted, stormy
Of violet eyes, glistening as you weep

Thank you very much also to Franco, Angela, and all the Rellas family for all the support. At this point, I should thank also the University of Bari for training me and

for they made me a scientist, but I will not. Well, thanks for nothing, UniBa. I hope you fail miserably and burn in hell as soon as possible and, hopefully, in the worst way conceivable. I still hate that place from the deepest of my heart, and I always will hate it with all of my might.
A huge thank you is for Dr. Eva Schlosser.
She took care of all of us, AEGIS fellows, assisted us with any problems we encountered during our studies and did an excellent job with the project management of the AEGIS consortium. She was more than an added supervisor to us; she was more like a mother. Thank you very much for everything you did for me and us, it will never be forgotten!
Thank you very much to Dr. Oliver Plettenburg and for all the evening spent talking about chemistry, life lessons, and other amenities. I will surely miss the beers and the pizzas together, and all the chemistry homework that came afterward.
A big thank you is for Prof. Michael Nilges, who supervised me and provided precious suggestions during my research.
A huge thank you to Prof. Alex Dömling, the man who taught me how to apply the Miles Davis “less is more” rule to medicinal chemistry.
Thanks to Prof. Gisbert Schneider, for hosting me in Zurich and for teaching me the Zen of computer-aided drug design.
A big thank you goes to Prof. Stefano Piotto, Prof. Simona Concilio, Carmine Cappetta, Luigi De Biasi, Lucia Sessa, and the UniSa Piotto Lab because without you I could have not been here. A thank you also to Prof. Lucia D’Accolti, my organic chemistry equivalent of Gunnery Sergeant Hartman.
A huge thank you to all my AEGIS fellows, Atilio, Ave, Joy, Laura, Giulia, Francesca, Engi, Markella, it has been incredibly fun to spend time with you in all the schools we had together!
A big thank you to Charlotte Softley, my fellow in the AEGIS ITN and NMR specialist at HMGU. A big thank you also to Michael Ostertag and Miriam Sonntag.

A German, a French, an Irishman, and an Italian step into a bar...

Well, that could easily be the starting line for a joke, but it was actually what Patrick, Maxime, Ryan, and I did all the time we met for the AEGIS schools. I will miss our wild nights out, and I will always love you as my brothers.

A special thanks to my other Portuguese brother, João 'Ahmed Al-' Encarnação for all the nights spent talking, playing PES, and the laughs during Munich's coldest of the winters.

A big thank you to all the people in the Popowicz/Sattler Lab and all my partners in crime in the German-Polish-Italian drug cartel: Krzystof Zak (the Polish White Duke of crystallography), Mark Bostock (aka Marek Bowsztowszki - the Polish-naturalized NMR artist), Maciek Dawidowski (the Godfather of Pex14 ligand design and synthesis) and my AEGIS fellow Valeria Napolitano (the golden lady of crystallography). Thank you very much guys, it has been a lot of fun to work with you!

Many thanks to Dominik Lenhart and Julia Dürner, it was great to spend time with you, and thank you for the Bayern Küche handbook! Love you guys!

Thank you to Sarah Rioton, Tony Fröhlich, Patrick O'Connor for all the nice time spent in the lab and all the fun during lunch breaks, I could have not asked for better lab mates!

A big thank you to all the people in the Sattler Group who I met in these four years, in particular, many thanks to Leonidas Emmanoulidis, the mastermind of the Pex14 project. A huge thank you to Waltraud Wolfson, our light in the darkness of German universities' bureaucracy. A special thank you to Gerd Gemmecker for all the deuterated methanol he provided us with!

Special thanks are to Arie Geerlof, Astrid Lauxen, Ana Messias, Stephanie, André, and all the people at HMGU for the wonderful working environment you built: I will miss you all guys.

A modern-day warrior
Mean, mean stride
Today's Tom Sawyer
Mean, mean pride
Though his mind is not for rent
Don't put him down as arrogant
His reserve, a quiet defense
Riding out the day's events
The river

And what you say about his company
Is what you say about society
Catch the mist, catch the myth
Catch the mystery, catch the drift

The world is, the world is
Love and life are deep
Maybe as his skies are wide

Today's Tom Sawyer
He gets high on you
And the space he invades
He gets by on you

No, his mind is not for rent
To any god or government
Always hopeful, yet discontent
He knows changes aren't permanent
But change is

And what you say about his company
Is what you say about society
Catch the witness, catch the wit
Catch the spirit, catch the spit

The world is, the world is
Love and life are deep
Maybe as his eyes are wide

Exit the warrior
Today's Tom Sawyer
He gets high on you
And the energy you trade
He gets right on to the friction of the day

6. Appendices

6.1. Introducing the CSP Analyzer: A novel Machine Learning-based application for automated analysis of two-dimensional NMR spectra in NMR fragment-based screening

Publication (Open Access)

Fino, R.*, Byrne, R.*, Softley, C. A., Sattler, M., Schneider, G., & Popowicz, G. M.

Computational and Structural Biotechnology Journal (2020)

Volume 18, Pages 603-611

https://doi.org/10.1016/j.csbj.2020.02.015

6.2. Computer-Aided Design and Synthesis of a New Class of PEX14 Inhibitors: Substituted 2,3,4,5-tetrahydrobenzo[f][1,4]oxazepines as Potential New Trypanocidal Agents.

Publication (Open Access)

Fino, R.*, Lenhart, D.*, Kalel, V. C., Softley, C. A., Napolitano, V., Byrne, R., Schliebs, W., Dawidowski, M., Erdmann, R., Sattler, M., Schneider, G., Plettenburg, O., Popowicz, G. M.

Journal of Chemical Information and Modeling (2021), *in press*

https://pubs.acs.org/doi/full/10.1021/acs.jcim.1c00472

6.3. Water envelope has a critical impact on the design of protein-protein interaction inhibitors

Communication (Open Access)

Ratkova, E. L., Dawidowski, M., Napolitano, V., Dubin, G., **Fino, R.**, Ostertag, M., Sattler, M., Popowicz, G. M., Tetko, I. V.

Chemical Communications (2020)

Volume 56, Number 31, Pages 4360-4363

https://doi.org/10.1039/C9CC07714F

6.4. Structure-Activity Relationship in Pyrazolo [4, 3-c] pyridines, First Inhibitors of PEX14–PEX5 Protein-Protein Interaction with Trypanocidal Activity

Publication

Dawidowski, M., Kalel, V. C., Napolitano, V., **Fino, R.**, Schorpp, K., Emmanouilidis, L., Lenhart, D., Ostertag, M., Kaiser, M., Kolonko, M., Tippler, B., Schliebs, W., Dubin, G., Mäser, P., Tetko, I. V., Hadian, K., Plettenburg, O., Erdmann, R., Sattler, M., Popowicz, G. M.

Journal of Medicinal Chemistry (2019)

Volume 63, Number 2, Pages 847-879

https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.9b01876

6.5. Structure and molecular recognition mechanism of IMP-13 metallo-β-lactamase

Publication (Open Access)

Softley, C. A., Zak, K., M., Bostock, M., J., **Fino, R.**, Zhou, R., X., Kolonko, M., Mejdi-Nitiu, R., Meyer, H., Sattler, M., Popowicz, G. M.

Antimicrobial agents and chemoterapy (2019)

Volume 64, Number 6, Pages e00123-20

https://doi.org/10.1128/AAC.00123-20

STATUTORY DECLARATION

Anhang I

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die bei der promotionsführenden Einrichtung

Technischen Universität München

der TUM zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:

Combining Computer-Aided Drug Design, NMR-Fragment based Drug Discovery and Medicinal Chemistry in Early-Stage Drug Development

in

Fakultät für Chemie - Lehrstuhl für Biomolekulare NMR-Spektroskopie

Fakultät, Institut, Lehrstuhl, Klinik, Krankenhaus, Abteilung

unter der Anleitung und Betreuung durch: Prof. Dr. Michael Sattler ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Ab. 6 und 7 Satz 2 angebotenen Hilfsmittel benutzt habe.

■ Ich habe keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht, oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

■ Ich habe die Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.

☐ Die vollständige Dissertation wurde in englischer sprache veröffentlicht. Die promotionsführende Einrichtung

Technische Universität München - Fakultät für Chemie

hat der Veröffentlichung zugestimmt.

■ Ich habe den angestrebten Doktorgrad noch nicht erworben und bin nicht in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.

☐ Ich habe bereits am _______ bei der Fakultät für ______________________________

der Hochschule______________________________

unter Vorlage einer Dissertation mit dem Thema ______________________________

____

die Zulassung zur Promotion beantragt mit dem Ergebnis: ______________________________

Die öffentlich zugängliche Promotionsordnung der TUM ist mir bekannt, insbesondere habe ich die Bedeutung von § 28 (Nichtigkeit der Promotion) und § 29 (Entzug des Doktorgrades) zur Kenntnis genommen. Ich bin mir der Konsequenzen einer falschen Eidesstattlichen Erklärung bewusst.

Mit der Aufnahme meiner personenbezogenen Daten in die Alumni-Datei bei der TUM bin ich

■ einverstanden, ☐ nicht einverstanden.

____

Ort, Datum, Unterschrift