WITS
UNIVERSITY

Computational modeling approaches to validate the druggability of the 26- and 28-kDa *Schistosoma* glutathione transferase enzymes using bromosulfophthalein as a benchmark ligand

by

AKEEL VALLI

1016205

Thesis

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, ZA

Supervisor: Ikechukwu A. Achilonu (PhD)

March 2024

I, AKEEL VALLI, am a student registered for the degree of *Doctor of Philosophy* in the 2024 academic year.

I hereby declare:

- I am aware that plagiarism—the use of someone else's work, without their permission and/or without acknowledging the original source—is wrong.
- I confirm that the research proposal submitted for assessment for the above degree, is my own unaided work except where explicitly indicated otherwise and acknowledged. In this context, I understand that the use of editing services is considered aided work and must be declared.
- I have not submitted this work before for any other degree or examination at this, or any other university.
- I have followed the required conventions for referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work, or that I have failed to acknowledge the source of ideas in my writing.

AKEEL VALLI
Johannesburg
March 2024

to Sadeck and Zarina

Acknowledgements

Dr Ike, *you have taught me how to learn*

Kiara, *you are everything I could have ever hoped for*

My parents, *you do not falter*

AKEEL VALLI
Johannesburg
March 2024

Funding information

This research was funded by the South African Research Chairs Initiative (SARChI) and the National Research Foundation (NRF) of South Africa, grant numbers: 64788, 88047 and 134105.

Abstract

PHARMACOPHORE MODELS are 3-D representations of the chemical and spatial features required for interaction with a drug target. These models offer advantages in early-phase drug design by expediting screening experiments and enabling the sampling of highly specific chemical space subsets, such as those containing quality drug-like candidates. The glutathione transferase enzyme of *Schistosoma* spp. (*S*GST) has been identified as an attractive drug target for the novel treatment of human schistosomiasis. We observed selective inhibition of *S*GST by bromosulfophthalein. Bromosulfophthalein was found to complex with *S*GST at a drug binding site in the target dimer interface, providing a suitable benchmark for the design of discriminative *S*GST pharmacophores. The aim of this research is to construct, deploy and evaluate pharmacophore models of the *S*GST drug binding site. The objectives are: to characterise the *S*GST drug binding site, to develop the pharmacophore models and finally to evaluate the drug-resolving ability of the models. We observed significant differences in the drug-binding character of *S*GST, compared to human glutathione transferase (hGST) counterparts, particularly that *S*GST supports binding of phenol and sulfonate moieties. Five- and four-feature pharmacophores were developed for the respective 26- and 28-kDa *S*GST variants. Finally, the models demonstrated remarkable ability to retrieve candidates displaying drug-like qualities. In conclusion, we characterised and developed pharmacophore models of the drug binding domains from two major *S*GST variants. Assessment of drug-resolving power validates the capability of the models to sample drug-like chemicals. Altogether, these accomplishments enable efficient and reliable screening toward novel drug treatment for human schistosomiasis.

List of figures

List of tables

List of abbreviations

BSP	bromosulfophthalein, see chapter 7.4
hGST	human glutathione transferase
GST	glutathione transferase, see chapter 7
kDa	kilodalton
PCA	principal component analysis, see chapter 6.1
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
*S*GST	*Schistosoma* glutathione transferase

Contents

Document arrangement

This document is divided into six sections denoted by Roman numerals I – VI. Each of the six sections are composed of various chapters labelled with Arabic numerals, all of which begin with a short chapter summary.

A description of each section is provided below:

I: Prologue

In this section, the thesis is introduced to the reader by establishing the problem statement and providing the study rationale. The over-arching research goals, as well as the document structure, are presented.

II: Literature review

In this section, we discuss experimental approaches and background topics which constitute fundamental themes of the study. Familiarity with these topics is critical to understanding the thesis.

III: Dimer interface analysis

In this section, we perform a comparative analysis of the dimer interface region of human and *Schistosoma* glutathione transferase enzymes.

This section is based on the material published in [37].

IV: Pharmacophore models

In this section, we demonstrate a molecular-dynamics based approach for pharmacophore model development.

This section is based on the material published in [38].

V: Drug-resolving power

In this section, we present a novel method to assess the drug-resolving capabilities of the *Schistosoma* glutathione transferase pharmacophores.

This section is based on the material in [39] submitted for peer-review

VI: Epilogue

In this closing section, we summarise the research endeavour, document its implications and provide critical recommendations for future work.

I

PROLOGUE

1

Introduction

This chapter expresses the significance of the work by establishing the problem statement. Next, various aspects of the rationale are presented to illustrate the study premise. The research goals are then documented.

1.1 Problem statement

PARASITIC TREMATODE BLOODWORMS of the genus *Schistosoma* cause human schistosomiasis [1], a neglected tropical disease common in Africa, south-east Asia and South America [2]. The reliance of native populations on contaminated water for cooking, cleaning and agriculture in these regions contribute to a global prevalence in excess of 290 million [3]. A further 700 million individuals living in endemic areas are susceptible to schistosome infection, more than half of which are children younger than 12 years of age [4].

Infection occurs when water-borne larval schistosomes penetrate the skin and enter the dermal vasculature, leading to skin rash, fever and nausea [2]. Subsequent maturation and migration of the worms to the systemic vasculature elicit profound host immune responses which culminate in intestinal, liver, kidney [2] and neural conditions [5]. Advanced symptoms are often debilitating, and persist long after the parasite has been eliminated. Various comorbidities have also been identified, including HIV infection in women with urinary schistosomiasis [6], as well as cognitive and growth disorders in affected children [7]. The combined effect of these occurrences implicates this disease in seemingly unrelated

varieties of organ dysfunction. It is therefore recognised that the prevalence of schistosome infections is considerably underestimated.

Praziquantel is an oral pyrazino-isoquinoline antischistosomal released in 1975 by Merck KGaA and Bayer AG [8]. The efficacy, tolerability and low cost of this drug were quickly realised [9], and chemotherapeutic intervention with praziquantel was adopted as the standard population-based control measure for schistosomiasis [10]. Alternative strategies relating to safe water and sanitation provisions declined in the years to follow [10], establishing an inadvertent dependence on praziquantel. The exclusive use of this drug in treatment campaigns spanning almost four decades has been identified to provoke resistance in schistosomes. This was demonstrated in several endemic African regions [11-15], including northern Senegal and the Nile river delta where praziquantel failed to clear human schistosome infection. Furthermore, parasites extracted from these subjects were reported to prevail even after subsequent *in vitro* treatment [11, 13]. These data might suggest an unresolved mechanism of praziquantel drug resistance.

The chronic immune-based pathophysiology of advanced schistosomiasis were determined to originate from the parasite egg antigens [16]. This reveals a critical inadequacy of praziquantel, whose mechanism of action targets only worm-phase schistosomes [17]. The limitations of this drug were described in subjects with high egg yields, where unusually low cure rates were observed [18]. In support of this observation is a high mortality rate in endemic regions stemming from hepatic and renal failure [3, 19, 20], conditions which are associated with chronic schistosome infection. Thus, the ability of praziquantel to provide a comprehensive therapeutic profile against human schistosomiasis is uncertain.

In summary, the problem statement contains two principal elements. First, the scourge of schistosomiasis in the context of the developing world is identified to pose a major health and socioeconomic burden. The demand to address this neglected disease is emphasized by the overlooked morbidity and mortality statistics arising from uncorrelated chronic infections. Second, the limitation of praziquantel for treating

advanced stages of schistosomiasis is described. This issue is accompanied by the impending emergence of drug resistant schistosome strains. Hence, a systematic approach toward novel agents for the treatment of human schistosomiasis is required.

1.2 Rationale

Glutathione transferase (GST) enzymes are central to the maintenance of cell health and viability in schistosomes [21, 22]. These metabolic species are responsible for a major proportion of cellular detoxification by carrying out phase II metabolism of reactive electrophiles [22]. Furthermore, GSTs support the synthesis of host immunosuppressive agents [23] which are required for successful parasite infection. Thus, the GST enzymes of *Schistosoma* have been established as attractive drug targets.

A variety of small-molecule ligands, including praziquantel [24], artemisinin analogues [25, 26], ellagic acid [27] and bromosulfophthalein (BSP) [28], were found to occupy a unique non-substrate site on *Schistosoma* GST. This site exists in the dimer interface of GST within proximity to the substrate binding sites. Certain ligands displayed noteworthy inhibitory effects on the enzyme activity upon binding at this site [27, 28]. Therefore, the dimer interface region is strongly implicated in the inhibition of *Schistosoma* GST.

Despite its intravenous use in medicine, BSP is not established as an orally bioavailable agent, making it unsuitable as a drug candidate. However, a remarkable observation of this small-molecule ligand is that its inhibitory effect is discriminative [28]. In particular, our laboratory has demonstrated the ability of BSP to exclusively inhibit the activity of the 26- and 28-kDa *Schistosoma* GST variants, compared to human GST alpha. Hence, BSP is a compelling archetype for designing novel selective inhibitors to target the *Schistosome* GST non-substrate site.

Modern drug design encompasses a variety of computational tools [29] for rational identification and development of inhibitory molecules. The advent of high-throughput, low-latency integrated processing technolo-

gies, which leverage the respective advantages of CPUs and GPUs [30], have transformed the paradigm of current drug design. These advanced processors provide rapid screening of large data sets [31], combined with efficient molecular modelling simulations [32] to generate reliable protein ligand complexes, binding affinities and ligand poses. Nonetheless, the design of strategic *in silico* protocols must be observed in order to harness these capabilities. In this regard, ligand-based screening with pharmacophore models provides an elegant alternative to conventional docking-based screening. The pharmacophore approach queries a three-dimensional feature map bearing the prerequisite interactions for binding of a ligand [33], effectively circumventing the complex energy calculations [34] required for docking. This provides satisfactory enrichment of inhibitory molecules with minimised computational cost, which demonstrates the value of the pharmacophore in the rational design of drug candidates.

The structural and chemical properties of a drug molecule have significant implications for efficacy and safety [35]. Properties most relevant to the behaviour of a drug are called drug-like properties and include the molecular weight, lipophilicity, membrane transporter affinities and hydrogen bonding nature [35]. Early prioritisation of candidates bearing desirable drug-like properties is imperative to expedite the drug design pipeline. These ideal drug-like properties, however, are not defined by absolute values [36], in which case desirable candidates could conveniently be obtained. Instead, ideal properties are defined by ranges of values which are subject to the mode of delivery of the drug, as well as the internal environment of the target tissue [36]. We therefore require more subtle approaches to assess the drug-like qualities of candidates. One such method involves the use of statistical analyses to determine the overlap in the properties of shortlisted candidates, compared to those from large databases of known drug molecules. In this manner, the degree of overlap represents the quality of the candidates.

In summary, there are various aspects to the rationale of this study. The first is the presence of a non-substrate ligand binding site in the dimer interface of *Schistosoma* GST. This enzyme has been established as a

compelling drug target due to its critical detoxification role in schistosome parasites. Second, the discriminative binding of BSP to GST reveals a valuable archetype for the design of novel selective inhibitors. Third, we demonstrate the advantages of pharmacophore screening methods for rapid retrieval of candidates. Finally, statistical analyses offer a sophisticated approach to estimate the drug-like qualities of candidates.

1.3 Research goals

The non-substrate inhibitor binding site in the *Schistosoma* GST dimer interface, coupled with the selective nature of the BSP small-molecule ligand, present an attractive opportunity for the design of discriminative pharmacophore models. Such models will offer considerable gains in speed and accuracy toward the prioritisation of candidates for selective *Schistosoma* GST inhibition. To ensure a reliable outcome, a rigorous preliminary characterisation of the dimer interface domains from relevant *Schistosoma* GST forms is required. Following this, suitable models can be developed and thereafter assessed in terms of drug-resolving quality.

Hence, this research project consists of three major goals:

1. Perform a comparative structural analysis of the relevant human and *Schistosoma* GST dimer interface domains.
2. Develop and implement pharmacophore models of the *Schistosoma* spp. GST drug targets.
3. Assess the chemical and structural properties of shortlisted candidates to determine the drug-resolving capabilities of the pharmacophore models.

2

Original research publications

---

1. **A. Valli and I. Achilonu**, Comparative structural analysis of the human and *Schistosoma* glutathione transferase dimer interface using selective binding of bromosulfophthalein.

   Proteins: Structure, Function and Bioinformatics
   Vol. 90, no. 9, pp. 1561 – 1569, 2022
   https://doi.org/10.1002/prot.26338

   Impact factor: 3.8

   Author contributions:

   A. Valli: 75% (experimentation, analysis, manuscript preparation)
   I. Achilonu: 25% (study concept, manuscript review)

2. **A. Valli and I. Achilonu**, Molecular dynamics-derived pharmacophores of *Schistosoma* glutathione transferase in complex with bromosulfophthalein: Screening and analysis of potential inhibitors.

Journal of Molecular Graphics and Modelling
Vol. 122, 2023
https://doi.org/10.1016/j.jmgm.2023.108457

Impact factor: 2.9

Author contributions:

A. Valli: 75% (experimentation, analysis, manuscript preparation)
I. Achilonu: 25% (study concept, manuscript review)

3. **A. Valli and I. Achilonu**, Investigating the drug-resolving capabilities of pharmacophore models derived from the 26- and 28-kDa *Schistosoma* spp. glutathione transferase enzymes.

   Journal of Computational Biology
   Under review

   Impact factor: 1.7

   Author contributions:

   A. Valli: 75% (experimentation, analysis, manuscript preparation)
   I. Achilonu: 25% (study concept, manuscript review)

4. **A. Valli and I. Achilonu**, Pharmacophore modelling to accelerate drug discovery.

   Molecular Biosciences Research Thrust (MBRT) Symposium
   University of the Witwatersrand, 2023

II

LITERATURE REVIEW

3

Molecular modelling principles

This chapter discusses fundamental concepts involved in the generation of chemical and molecular models. First, molecular mechanics methods for energy parameterisation are defined. Approaches used to evaluate the effect of solvation are then documented. Finally, a procedure for the systematic minimisation of ligand and protein geometries is described.

3.1 Molecular mechanics

STRUCTURAL DATA CONSISTING OF ATOMIC COORDINATE SETS and bonding arrangements represent the fixed three-dimensional description of a molecular species. The interactive and conformational behaviours of the molecule are apprehended by modelling its sum of potential energy. This is obtained with the molecular mechanics approach [40, 41], which formalises bonded atom pairs as charged spheres connected by springs. Pairwise classical mechanics energy terms are hence enumerated across the molecule, each corresponding to a particular bonded or nonbonded interaction, producing a total potential energy:

$$U_{molecule} = \sum U_{bond\ stretching} + \sum U_{bond\ bending} + \sum U_{bond\ torsion} + \sum U_{van\ der\ Waals} + \sum U_{electrostatic}. \quad (3.1)$$

Various potential energy functions exist to model these terms, including Hooke's law for bond stretching, Coulomb's law for charged interactions and the Lennard-Jones potential for Van der Waals forces [41]. The combination of these functions along with their parameters and empirical constants are collectively called a force field.

3.2 Implicit solvent models

The effects of solvent, which are vital in representing the total potential energy of a molecule, are not captured by the intrinsic molecular energy. Solvation energy contributions are thus resolved independently, either by explicit species whose energies are computed by the force field, or by an implicit solvent model [42].

Implicit models [42, 43] delineate the solvent as a homogenous dielectric continuum occupying all surfaces of the target. Interaction of the target with this dielectric causes polarisation of the bulk solvent, generating an instantaneous reaction potential which is rapidly evaluated. This method offers substantially reduced complexity compared with explicit models, which are encumbered by the tasks of equilibrating the solvent network and accounting for each solvent molecule in the system [42]. These implicit methods are, however, incapable of resolving the solvation effects of buried target groups, since the associated sharing of solvent distorts the dielectric uniformity [44, 45].

Solvent reaction potentials can be approximated by either a distance dependent dielectric, or the generalised Born/surface area method (GBSA) [42]. The distance dependent dielectric is a summation of local Coulomb potentials U based on a sigmoidal distance-dependent dielectric function $\varepsilon(\vec{r})$. Solvation energies are calculated between each species a on the target surface and its corresponding element on a set of regular probes b occupying the solvent interface:

$$U_{solvation} = \sum \frac{q_a q_b}{\varepsilon(\vec{r}_{ab})|\vec{r}_{ab}|}. \qquad (3.2)$$

The GBSA method identifies separate polar and non-polar contributions to the solvation energy. Reaction potentials of polar target groups are approximated with the Poisson-Boltzmann model [46]. This defines the polar solvation effect as a sum of partial and formal potentials, which are parameterised by the Poisson- and Boltzmann equations respectively.

From the linear Poisson equation for electrostatics,

$$\nabla\varepsilon(\vec{r})\nabla\phi(\vec{r}) = -4\pi\rho(\vec{r}), \qquad (3.3)$$

the point magnitude of the electrostatic potential $\nabla\phi(\vec{r})$ at each position $\vec{r}$, is related to the target charge density $\rho(\vec{r})$ and the magnitude of the solvent dielectric field $\nabla\varepsilon(\vec{r})$.

Adding the linear Boltzmann term for the formal potentials produces the Poisson-Boltzmann equation:

$$\nabla\varepsilon(\vec{r})\nabla\phi(\vec{r}) = -4\pi\rho(\vec{r}) + \kappa^2\varepsilon(\vec{r})\phi(\vec{r}). \qquad (3.4)$$

Solving for $\nabla\phi(\vec{r})$ in 3.4 gives the energy due to the polar solvent effects. This equation incorporates the Debye-Hückel [42] screening effect κ of salt ions, where $\kappa^2 = [\text{salt}]$.

Computing solvation energies for large targets in this manner remains a time-intensive task [42, 46]. The generalised Born model (Figure 3.1) addresses inefficiencies by discretising the target charge density to atomic positions i defined by $\vec{r}_i$, effectively forming a finite set of point charges q_i upon which the reaction potentials are calculated [46, 47]. In addition, the solvent polarizability for each point is restricted to that which occupies a sphere of radius α determined by the solvent exposure of the charge.

Nonpolar reaction potentials are calculated with a simple function dependent upon the total solvent-accessible surface area A of the target [42]. This incorporates both hydrophobic interactions and the entropic losses of solvent ordering into a proportionality constant σ, such that

$$U_{nonpolar} = \sigma A. \qquad (3.5)$$

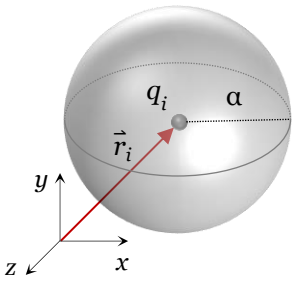


Figure 3.1. Generalised Born solvent model.

3.3 Energy minimisation

Molecular mechanics resolves energies in terms of bonded and nonbonded interactions. The dependence of these interactions on the atomic configuration of a molecule demonstrates the energetic implications of conformational change. Energy minimisation [48] is a technique which explores the structural flexibility of a protein or ligand to generate the lowest energy atomic configuration. We begin by defining the potential energy U of the molecule as a function of its various degrees of freedom, represented by a set of vectors X:

$$U = f(X). \tag{3.6}$$

In the cartesian coordinate system, a given molecule of N atoms contains $3(N\text{-}1)$ degrees of freedom $\vec{a}$ such that

$$X \in \{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_{3(N\text{-}1)}\}. \tag{3.7}$$

Plotting the energy dependence upon any degree of freedom $\vec{a}_x$ yields Figure 3.2. The goal of the minimisation algorithm is to identify which magnitudes of the degree of freedom produce low energy conformers, represented by troughs on the potential energy function. This is achieved by traversing the curve downhill from a given starting position such that the region of zero-gradient is reached.

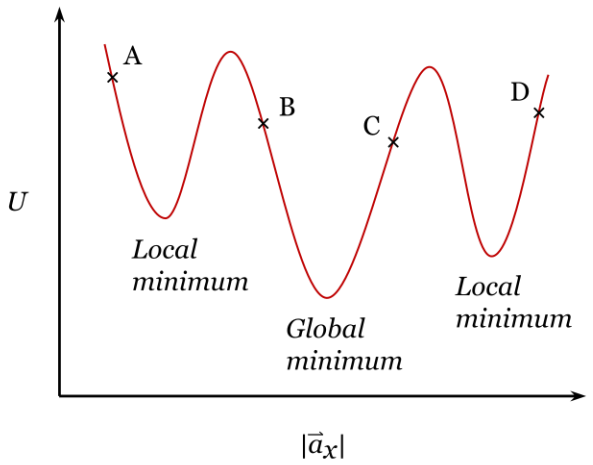


Figure 3.2. Potential energy as a function of atomic motion.

Consider point A, whose slope is expressed as

$$\frac{\partial U}{\partial \vec{a}_x} < 0. \tag{3.8}$$

The optimisation problem

$$\frac{\partial U}{\partial \vec{a}_x} = 0, \tag{3.9}$$

is satisfied for minima where

$$\frac{\partial^2 U}{\partial^2 \vec{a}_x} > 0. \tag{3.10}$$

Applying the same for point B on the energy curve returns an alternate minimum. Optimisation solutions for points C and D, where a positive slope is evaluated in 3.8, are obtained by reversing the inequality in 3.10. The global minimum energy represents the most stable conformation of $\vec{a}_x$. This can be approximated by random sampling to produce a set of starting positions upon which the optimisation principle is applied.

Systematic minimisation of each degree of freedom in 3.7 generates the equilibrium configuration of the molecule, which is vital in ensuring an accurate representation of a ligand or target structure during molecular modelling. Minimisation is also required when adjustments are made to the structure [49], for example when enumerating ligand isomeric states and repairing protein side chains.

3.4 Protein-ligand complexes

Calculating the potential energy sum of a protein in complex with a ligand is a valuable approach to assess the stability of the interaction [32]. This provides a preliminary indication of the empirical binding energy which bears strong implications for binding affinities of enzyme substrates as well as drug candidates.

There are two dynamic aspects of the protein-ligand interaction which are captured by the binding energy term, namely the entropy and enthalpy of binding [37, 40]. The conformational freedom accessible to an unbound ligand is generally reduced upon binding with a protein, prompting the enthalpically-driven formation of non-covalent bonding arrangements at the binding site [37]. Hence, a favourable binding interaction generally requires a gain in enthalpy which exceeds the combined entropic loss of the ligand and protein binding site.

However, the role of solvent [32, 37] must also be considered, as binding sites may accommodate solvent-mediated interactions for a stable interaction to occur. In some cases, the energetic effects of the solvent supersede that of the ligand, where the enthalpy of solvent interactions must necessarily exceed the entropic loss of solvent ordering.

Certain proteins have intrinsically high mobility relating to their function, while others might have low mobility corresponding to low entropy [41]. The same is true for large ligands with many rotatable bonds which tend to exhibit greater entropy. Therefore, evaluating changes in mobility is crucial to resolve the entropic effects of protein-ligand interactions in molecular modelling.

4

Computational approaches

This chapter describes some common tools for representing and evaluating molecular models. Firstly, the techniques, use cases and analyses of docking are illustrated. A description of molecular dynamics is then provided. Next, the implementation of pharmacophores in drug design is presented, followed by an explanation of pharmacophore generation.

4.1 Molecular docking

THE OBJECTIVE OF MOLECULAR DOCKING is to predict the orientation of a ligand within a target domain. Obtaining accurate geometries of a ligand-target association demands a thorough approach to account for the various degrees of freedom afforded by the ligand and binding site residues. Docking [34, 50] begins with sampling of ligand conformations and generation of a fixed three-dimensional descriptor of the binding site called a grid. The primary component of the grid is the energy field, calculated by resolving the potential at each of its vertices. In addition, this grid contains the Van der Waals radii and electronic distributions of all atoms existing within the binding site. The preliminary orientation of each ligand conformer is then identified by placing it in the grid such that it minimises spatial collisions with target groups. Next, optimisation of ligand geometry is performed, followed by rigid-body translations and rotations to further reduce target collisions and stabilise the interaction energies. Here, conformers below an energy threshold are eliminated to filter out poor ligand orientations. Those which prevail succeed to the scoring step, where rewards are calculated by computing energies of the

interacting target groups, and penalties by assessing Van der Waals overlap. Some docking protocols incorporate a grid refinement step during scoring, where minimisation of target binding groups is performed [51, 52]. Finally, the best-ranked orientations are presented along with their binding energies, scores and root-mean-square deviation values. Ligand interaction analysis is usually performed on the resulting complexes to identify bonding residues, ligand-target bond properties, as well as the spatial attributes of the ligand [50].

Resolving the binding geometries, energetic parameters and interaction profiles of protein-ligand complexes is useful to estimate key features of potential drugs relating to binding affinity and inhibitory behaviour [50]. These features are used to prioritise candidates from ligand databases during virtual screening. In addition, docking is implemented to create complexes for structure-based pharmacophore generation, as well as to rationalise empirical inhibition studies of ligands [37].

4.2 Molecular dynamics

Classical mechanics force fields report the net scalar potential energy of a system, which fails to capture the structural flexibility of large macromolecules such as proteins. Molecular dynamics offers insight into these spatial fluctuations to reveal the behaviour of the protein during ligand docking and catalytic events. This process resolves the potential energy functions iteratively [53], each time applying displacements to the bonded and nonbonded species under the corresponding forces (Figure 4.1). The result is a spatio-temporal energy representation which displays the atomic trajectories of the protein. Analysis of these trajectories, particularly when simulating the binding sites of protein-inhibitor complexes, allows for interpretation of energetic parameters, as well as assessments of ligand mobility and inhibitory potential [52].

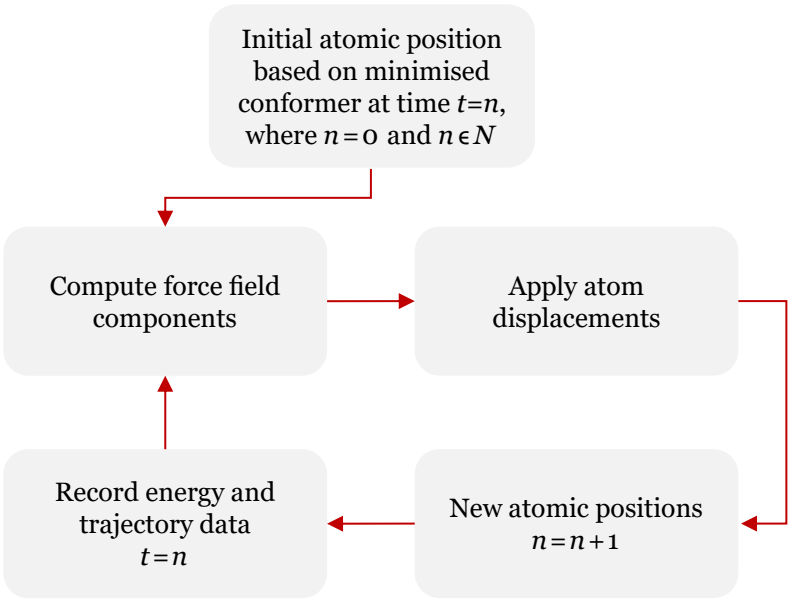


Figure 4.1. Molecular dynamics simulation.

4.3 Pharmacophores

Molecular recognition events [54] between a ligand and its target during enzyme catalysis, receptor activation and inhibitor binding are mediated by geometric complementarity and specific intermolecular forces. The particular spatial and chemical features required for association of a ligand with its target are encoded in an abstract model called a pharmacophore, which may be represented in several ways [55, 56]. A schematic pharmacophore model is a three-dimensional graph whose nodes depict the prerequisite features separated by distance. These features include ligand hydrophobic, aromatic, ionic or hydrogen bonding moieties. Numerical pharmacophore representations [57] are produced by applying hash functions to the graphical models, allowing downstream processing with pattern-matching and clustering algorithms.

Please consult chapter 14.5 for examples of pharmacophore models.

4.3.1 Implementation

Pharmacophore-based virtual screening [56, 58] is a common practice in drug design, where active molecules are identified based upon a query delineated by the set of features. This technique offers rapid screening of large drug libraries compared with molecular docking approaches [50], which are subject to long runtimes due to intensive energy, refinement and scoring calculations. Searching of fragment libraries is a similar approach to identify critical functional groups for *de novo* synthesis of drug leads [55, 59]. Furthermore, lead optimisation workflows commonly use pharmacophores to maintain the essential binding features of promising candidates during scaffold hopping. Outside the realm of drug design, models representing binding domains are used to predict potential metabolites in enzymatic and receptor-mediated pathways [60, 61].

4.3.2 Pharmacophore generation

Methods used to construct pharmacophores are typically distinguished by the input data used. Ligand-based techniques involve training datasets of active molecules which undergo alignment, followed by feature abstraction from overlapping chemical groups [55, 58]. These methods require preliminary energy minimisation steps to optimise ligand geometries and generate approximate bioactive conformations. Performing simultaneous conformer generation of large ligand datasets during the alignment phase is likely to cause some computational inefficiencies. The runtimes are therefore improved by precomputing conformer ensembles prior to pharmacophore generation [55]. Improvements to the efficiency are also achieved by adding inhibition data to the training set, this allows for prioritisation of features from ligands with favourable inhibitory behaviour [59]. Similarly, incorporating inactive ligands in the dataset is used to expedite the protocol, by allowing elimination of certain features.

First, preliminary conformers of elements in the dataset are enumerated with rotamer manipulations. Local minima for each rotamer are then sampled in individual succession with nitrogen pyramidal inversions and ring flips. This forms a heuristic decision tree [58] which is traversed

to approximate the optimal ligand energy. Next, pairwise alignments of all conformers, with a randomly selected reference, are performed using either point- or property-based algorithms [62]. Point-based alignment uses least squares fitting of points arranged on each atom or chemical group. Alternatively, property-based methods employ the superposition of molecular descriptors to perform alignment. These descriptors most notably contain the molecular fields of elements in the conformer library, which are resolved by computing the potential energy at each vertex in a three-dimensional grid surrounding the conformer. Shape, volume and electron density are also included in these molecular descriptors. Finally, the pharmacophore model resulting from a particular library is scored based on the fit of the conformers [55, 58]. Successful alignment is challenging when corresponding points or overlap of molecular descriptors on the aligned conformers cannot be obtained. This is often encountered with diverse datasets and is remedied by grouping similar conformers and constructing separate models of each subset [58].

Structure-based pharmacophore methods [55, 56] generate the feature set from an individual target or ligand-target complex. Abstraction from an apo target is achieved by evaluating the potential energies at the vertices of a regularly spaced grid in the binding domain. Much like in the property-based ligand alignment technique, these potentials allow for the creation of an energy field permeating the target site. Vertices existing within high energy field densities are converted into corresponding features based on predefined potential energy categories. Models developed from target inputs often contain excessive feature counts, this is easily mitigated by implementing user inspection or higher energy field thresholds for feature detection [55]. Apo target methods are used even in instances where reliable ligand complexes are available, providing the advantage of novel binding poses [56]. In such cases, manual configuration of the models may be applied to bias certain features based on pre-established interaction profiles [55, 59].

Pharmacophore construction methods using ligand-inhibitor complexes [55] are advantageous when the preservation of binding modes and inhi-

bition characteristics is necessary. These methods involve independent assessment of the ligand, the target site, as well as the interaction profile for generation of comprehensive feature sets [33, 59]. Approaches incorporating molecular dynamics simulations of the complex [33] are useful to provide better perception of the ligand-target association. In addition to enriching the model with solvent effects, this step enables the use of interaction frequency distributions for feature selection.

5

Rational drug design

This chapter reviews the current paradigm of drug design. The typical procedure for identification of targets is demonstrated. Furthermore, routine approaches used in lead development protocols are described.

5.1 Current approaches

TRADITIONAL DRUG DISCOVERY [63] relied on chance outcomes of high throughput screening experiments. Such methods typically lack the ability to resolve the fundamental molecular basis for the behaviour of a drug. Modern computational tools [31, 52] coupled with the availability of biomolecular structures, metabolic pathways and pathologic mechanisms have ushered in a novel paradigm of drug design [29, 64, 65]. These techniques offer a systematic, multi-faceted approach to identify and refine various properties of a potential drug including its binding affinity, inhibition characteristics and therapeutic action. The drug development pipeline [63, 64] begins with identification of a molecular target, which is most commonly a protein. Next, lead candidates are sought either by direct ligand screening or *de novo* synthesis. Empirical studies follow to identify and validate the lead molecule. Finally, the druglike properties of the lead are optimised, concluding the initial phase of drug design in preparation for animal studies and clinical trials.

5.2 Target identification

Identifying drug targets can be a demanding task [52] due to the complex biochemical pathways in human and pathogen metabolism. Difficulties are especially prevalent in the design of antiparasitics [66], whose target agents observe variable lifecycles with corresponding variations in metabolite expression and accessibility. However, improvements in biological data collection and processing [67] has enabled the resolution of detailed metabolic networks ideal for use in parasite target identification [66, 68]. An empirical analysis [65] is generally used to verify preliminary targets. In helminths such as schistosomes, these range from carbon metabolism species [68, 69] such as triose phosphate isomerase and phosphofructokinase, to detoxification enzymes such as glutathione S-transferase [24].

5.3 Lead development

Several variables influence lead design and optimisation protocols, most notably the availability of accurate structure data of the target enzyme or receptor. This distinguishes structure-, from ligand-based schemes [29], which instead use the structures of inhibitors or binding agents for lead development (Figure 5.1). Whilst docking methods [50, 52] require the target structure file, pharmacophore methods [70] can operate in either structure- or ligand-based protocols. Both these approaches enable prioritisation of lead candidates from ligand datasets using virtual screening.

Molecular docking [50, 52] evaluates the binding energy and ligand root-mean-square deviation (RMSD) of each element in the set to generate docking scores. Increasingly stringent scoring functions might be applied in succession to converge upon the most promising lead candidate. In contrast, pharmacophore screening [59, 70] identifies potential leads by aligning each ligand to a query pharmacophore, derived from either the target structure data, known ligands or inhibitor complexes. Similarities in feature sets, inter-feature distances and shape are used to score these ligand alignments [57]. Such approaches offer drastically faster screening [59] by circumventing the intensive calculations present in docking.

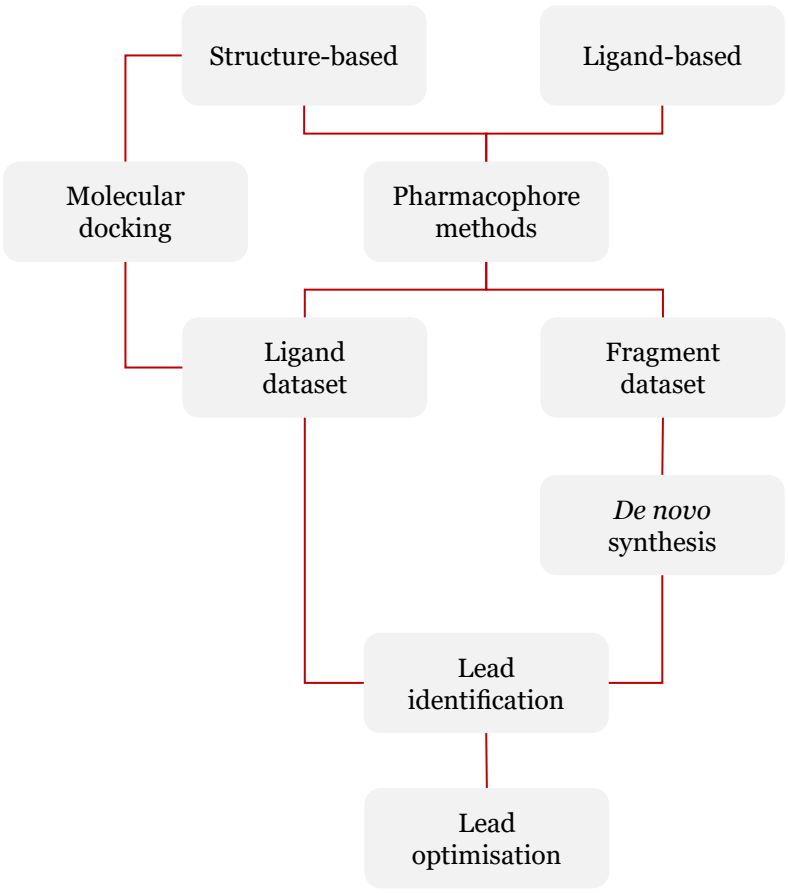


Figure 5.1. Lead development schemes.

However, these quicker turnaround times must be compared against the inherently constrained chemical sampling of pharmacophore screening, which reveals only candidates structurally similar to the query model. In some cases [37], the detriment of limited chemical sampling is leveraged to prioritise candidates possessing certain selectivity or inhibitory properties, based upon pre-established ligand data. Pharmacophore models may also be used in screening of fragment datasets for *de novo* synthesis [55]. Here, the feature set and topology of the model are used to assemble combinations of isolated molecular fragments. These fragments are then joined with linker groups to form novel leads for empirical assessment.

The best lead molecule is identified and validated with laboratory-based experiments [65]. Lead optimisation procedures follow to improve its

druglike properties including target affinity, potency, metabolic stability and bioavailability [29, 65, 71]. This is an iterative process which involves the modification of chemical groups, followed by re-assessment with *in vitro* and animal experimentation to refine the lead molecule for clinical testing. Scaffold-hopping [72] is a lead optimisation technique where the central region of the lead, typically consisting of heterocycles and linker moieties [71], is modified to enhance its druglike qualities. This central region is termed the molecular scaffold and has profound implications for drug permeability and metabolism. Synthesis and empirical analyses are first used in this optimisation scheme to identify the most favourable scaffold. Then, derivatives of the scaffold are enumerated by varying the substituents to generate a shortlist of refined lead molecules. Testing of refined leads is finally performed to establish structure-activity relationships [73], which are used to select the final drug candidate.

6

Statistical techniques

This chapter explains some key statistical techniques used in this study.

6.1 Principal component analysis

Consider a numerical data table of m samples bearing n variables each. Visualising all variables in this table requires n dimensions, which cannot conveniently be achieved when $n > 3$. Principal component analysis is a technique which reduces the n dimensions to a hypothetical lower-dimensional representation. This is obtained by creating a new set of variables called principal components which capture the essential features of the data [74]. Subsequent visualisation along these principal components allows for interpretation of high-dimensional data tables.

Principal component analysis is a convenient solution for comparative molecular modelling, such as determining structural variations across protein super-families and distinguishing protein-ligand interaction profiles. This method also allows us to characterise chemical and structural properties of ligands, such as those contributing to drug-likeness.

We begin the analysis by defining the $n \times n$ covariance matrix of the data, where the major diagonal contains the variance of each variable and the remaining elements contain the covariance values.

For example, if $n = 4$, the covariance matrix is expressed as

$$Cov = \begin{bmatrix} V_1 & C_{1,2} & C_{1,3} & C_{1,4} \\ C_{1,2} & V_2 & C_{2,3} & C_{2,4} \\ C_{1,3} & C_{2,3} & V_3 & C_{3,4} \\ C_{1,4} & C_{2,4} & C_{3,4} & V_4 \end{bmatrix}. \tag{6.1}$$

Next, we identify its eigenvectors

$$\vec{v}_1 , \vec{v}_2 , \vec{v}_3 , \vec{v}_4 \tag{6.2}$$

and eigenvalues

$$\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4. \tag{6.3}$$

The eigenvectors describe the directionality of the principal components in the lower-dimensional subspace, whereas the eigenvalues represent the magnitude of variation captured by each component [74]. Since the covariance matrix is symmetrical, it forms a linear transformation of the data to a subspace defined by a set of orthogonal principal component axes [74, 75]. We finally project the variables in the data table along axes delineated by the first few principal components, effectively reducing the number of dimensions n.

If the data variables have different units of measure, unequal variance is likely to be observed. In such a case, the linear transformation of 6.1 will favour principal components which correspond to variables bearing units of greatest variance. Instead, we ensure consistent treatment of each variable by calculating the standardised score Z_{ij} of each data point as a function of its mean μ_j and standard deviation σ_j:

$$Z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}. \tag{6.4}$$

These standard scores represent scaled data points with zero mean and unit variance. Calculating the covariance matrix of the standard scores forms the correlation matrix, which is used in the linear transformation.

6.2 Clustering analysis

Grouping of samples based on similarity is a useful technique to identify patterns in data. Clustering analysis uses algorithms to group the data along their principal axis pairs [76]. This allows for accurate and reliable clustering which accounts for all variables in the data table. The partitioning around medoids (PAM) [76, 77] algorithm performs clustering by first initialising a user-specified k preliminary clusters defined by central samples S called medoids. This is achieved by calculating a distance-based dissimilarity matrix for each axis pair and designating each non-medoid point U to its closest medoid. Next, the algorithm swaps each S with each U and recomputes the dissimilarity matrix to identify the average dissimilarity coefficient. By identifying the minimal average dissimilarity, the optimal clustering structure is obtained. PAM is less susceptible to the presence of outliers compared to other clustering algorithms, since it defines the central point by samples in the dataset, instead of deriving a hypothetical cluster centre from the mean.

This approach allows us to extract key features of protein atomic data and ligand properties. For example, we can identify common motifs in protein structures as well as similar binding behaviours amongst ligands.

7

Glutathione transferases

This chapter rationalises glutathione transferase (GST) enzymes as the principal drug targets in schistosomes. First, the role of cytosolic GSTs in parasite metabolic detoxification is documented. The classification schemes of the GST superfamily are then explored to identify human GST controls. Next, the shared catalytic mechanism and binding site architecture of the relevant GST isoforms are outlined. Finally, studies on the selective inhibition of the schistosome drug target are presented.

7.1 Enzymatic metabolism

REACTIVE CHEMICAL SPECIES pose a threat to cellular macromolecules such as proteins, DNA and lipid membranes, causing dysfunction associated with oxidative stress [78, 79]. In the *Schistosoma* parasite, cellular detoxification of lipophilic reactive species is fulfilled by phase I and II enzymatic metabolism [80, 81]. These processes maintain parasite survival by solubilising both endogenous and host-derived toxins, allowing for their elimination. Phase I metabolism, which involves hydrolytic or redox reactions [80], is mediated by low quantities of deacetylase and reductase enzymes [81]. This results in the introduction of reactive hydroxyl or epoxide groups to the toxin, increasing its hydrophilicity [79] and preparing it for phase II reactions. In phase II metabolism, either these activated metabolites, or more commonly, unprocessed reactive species, are conjugated with the polar tripeptide glutathione [80, 82]. Glutathione transferase (GST) enzymes catalyse this reaction (Figure 7.1) to produce soluble derivatives, which readily associate with efflux proteins.

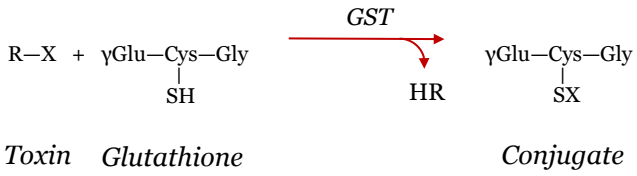


Figure 7.1. Glutathione conjugation reaction.

The predominance of glutathione adducts [80] in the phase II scheme of *Schistosoma*, in lieu of glucuronidation or sulfation products, correlates with the abundant cell concentrations of GST [81, 83] compared to other conjugation enzymes. Hence, when coupled with modest phase I activity [81], the dependence of parasite viability on GST-mediated metabolism is clear. Furthermore, GSTs have been implicated in the synthesis of host immunosuppressive compounds [23] such as prostaglandins E2 and D2, enabling successful infection. Due to these observations, GST enzymes have been established as antischistosomal drug targets.

7.2 Cytosolic glutathione transferases

GSTs belong to a vast superfamily [82, 84] of ubiquitous detoxification agents present in various forms across animal and plant species. These enzymes are most broadly classified [82, 85] based upon their subcellular location, leading to cytosolic, mitochondrial and membrane-bound GST categories. The most prominent GSTs identified in *Schistosoma* spp. are the cytosolic 26- and 28-kDa isoforms [83], which have been associated with the design of novel antischistosomal drugs and vaccines. Both these isoforms fulfil the canonical glutathione conjugation role, with the minor differences between them [27, 86] producing distinct binding specificities for cell toxins. In comparison, human cytosolic GSTs (hGSTs) [87] form a comprehensive set of classes based on sequence identity and substrate specificity: alpha, mu, pi, theta, zeta, omega and sigma.

Human and *Schistosoma* cytosolic GSTs share a dimeric structure [37] possessing two catalytic domains per enzyme, each containing a G- and an H-site where glutathione and the reactive species bind respectively. Subunits of 200 – 240 residues [88] associate at the dimer interface via

hydrophobic interactions [82] to form a pocket in the core of the enzyme [27, 87]. This pocket extends laterally upward and downward in opposite directions from the plane of the dimer, toward the surface of each of the subunits, where it expands to form the two respective catalytic domains. Altogether, this produces a channel spanning the diameter of the enzyme where ligand binding is accommodated. Figure 7.2 shows a longitudinal section of the binding channel, with glutathione molecules present in the G-sites and ellagic acid in the dimer interface pocket [27].

GST catalysis [89] initiates with deprotonation of the bound glutathione sulfhydryl to form a nucleophilic thiolate ion. Residues within the G-site promote formation of this thiolate by interacting with the sulfhydryl and decreasing its effective pK_a. Subsequent attack of the thiolate nucleophile upon the electrophilic reactive species in the nearby H-site completes the enzymatic reaction to produce the glutathione conjugate. In the GSTs of *Schistosoma*, a catalytic tyrosine residue has been implicated in generation of the glutathione nucleophile [89]. Specifically, pi-cation bonding between this tyrosine and an auxiliary arginine allows for displacement of its phenol hydrogen, resulting in a tyrosinate which captures the thiol hydrogen of glutathione.

All cytosolic GSTs demonstrate conserved structural characteristics due to their common detoxification role [78, 82]. The hGST alpha, mu and pi classes bear the greatest resemblance to *Schistosoma* GST isoforms [87,

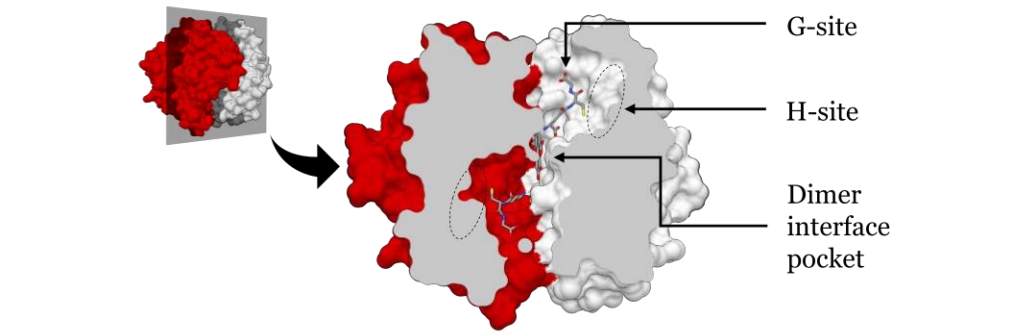


Figure 7.2. Cross-section of the 26-kDa glutathione transferase enzyme from *Schistosoma,* coloured by subunit. PDB ID: 6RWD.

89]. In addition to possessing the same catalytic tyrosine as schistosome GST, these isoforms share a collection of basic residues with the parasite enzyme. This includes arginine and lysine residues in the region linking the dimer interface pocket and G-site of the binding channel [87], which are responsible for stabilising the glutathione. Therefore, the alpha, mu and pi hGSTs are identified to be most relevant to the design of selective inhibitors for schistosome GST.

7.3 Non-substrate binding

The primary antischistosomal, praziquantel, was found to bind with the dimer interface pocket of the 26-kDa GST from *Schistosoma japonicum* [24]. Initially, the presence of this drug in close proximity to the catalytic domains was theorised to obstruct substrate binding and hinder the GST conjugation. However, no noteworthy inhibition could be identified [90]. Recent studies have confirmed the association of several other molecules with the same dimer interface region of *Schistosoma* spp. GST, including bromosulfophthalein [28], indanyloxyacetic acid, ellagic acid [27], artemether [26], as well as various anthraquinone dyes [91]. Furthermore, these molecules demonstrated empirical inhibition of GST catalytic activity. These observations altogether validate the GST drug target and highlight the significance of the dimer interface in non-substrate ligand binding.

7.4 Selective inhibition

Bromosulfophthalein (Figure 7.3) is a bi-sulfonated polycyclic dye used in the medical industry to assess liver function [92]. Empirical analysis of this molecule has revealed its ability to inhibit the 26- and 28-kDa schistosome GST isoforms with half-maximal inhibitory concentrations of 2.83 and 0.74 μM [28], respectively. Similar effects were absent in human GST (hGST) studies [28], where a minimum average inhibitory concentration of 7.00 μM was recorded for hGST alpha. Hence, bromosulfophthalein displays discriminative inhibition of *Schistosoma* GST upon binding at the non-substrate site, establishing it as a benchmark for designing novel selective inhibitors.

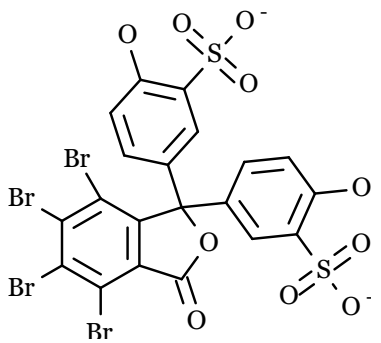


Figure 7.3. Chemical structure of bromosulfophthalein.

III

DIMER INTERFACE ANALYSIS

This section is based on:

VALLI, A. and ACHILONU, I.

Comparative structural analysis of the human and *Schistosoma* glutathione transferase dimer interface using the selective binding of bromosulfophthalein

Proteins: Structure, Function, and Bioinformatics

Volume 90, Issue 8

August 2022

8

Synopsis

In this chapter, we summarise the aims, methods and highlights of the comparative dimer interface analysis of human and schistosome GST.

THE BINDING CHANNEL of the *Schistosoma* glutathione transferase has been shown to possess a non-substrate site associated with catalytic inhibition. This binding channel is formed from the interface of the GST dimer. We produced a comparative characterisation of the *Schistosoma* GST dimer interface with respect to analogous human GST (hGST) forms using the selective binding of bromosulfophthalein (BSP).

Firstly, two *Schistosoma* GST and three hGST structures were selected to be used as search queries in assembling a dataset of 48 empirical GST structures. Sequence alignment to generate a universal residue indexing scheme was then performed, followed by a local structural superposition of the dimer interface site. Principal component analysis (PCA) revealed appreciable variation of the dimer interface, suggesting the potential for selective inhibition of schistosome GST. BSP was found to dock invariably in the dimer interface core pocket, placing it in proximity to the GST catalytic domains, through which it may exert its inhibitory behaviour. Binding poses across the various GST isoforms were distinguished with ligand interaction profiling, where schistosome GST complexes showed stabilization of ligand aromatic- and sulfonate moieties, which anchor the ligand and produce a tight association. In comparison, the missing aromatic stabilisation in the hGST complexes manifest in large bonding distances, causing mobile poses likely to dissociate.

In total, this study demonstrates the potential for selective inhibition of *Schistosoma* GST enzymes, rationalises the selective binding behaviour of the BSP inhibitor, and produces a reliable metric for the development of pharmacophore models of the schistosome GST binding channel.

9

Methods

This chapter documents the approaches used in the comparative dimer interface study of human and schistosome GST, beginning with a brief introduction to the computational utilities used. Next, dataset assembly and analysis of structural variation are described. Finally, selective BSP binding is rationalised with docking and ligand profiling methods.

9.1 Software utilities

BIO3D IS A STRUCTURAL BIOINFORMATICS TOOL [93] offered as a package for the R statistical environment. It provides a wealth of functionalities relevant to the manipulation and comparison of large structural datasets, including sub-setting of atomic structure files, sequence- and structural alignment, principal component analysis (PCA) and graphing of coordinate data. These capabilities were harnessed to achieve the local structural comparison of the GST isoforms, as well as to prune structures for the molecular docking protocol.

UCSF Chimera [94] is a molecular visualisation program which contains features for structure editing, optimisation and docking. In addition to the vast array of three-dimensional structure and sequence visualisation tools, Chimera's integrated docking preparation utility was implemented in tandem with the AutoDock Vina [95, 96] extension to generate BSP complexes of each of the GST isoforms.

The web-based Protein-Ligand interaction Profiler (PLIP) [97] was used to obtain the ligand-receptor interactions of each complex, allowing for

interpretation of the comparative analysis and rationalisation of the BSP binding behaviour. This tool allows for exporting of three-dimensional structure files showing the interactions, making it useful for third-party assessment with PyMOL version 2.3.2 [98].

9.2 Dataset assembly

The dataset was initialised by querying the protein data bank [99] for the most recent and best resolution wildtype structure of each target isoform. Wildtype structures were retrieved to avoid mismatch of mutated residues during alignments, which would compromise enrichment of the dataset. Empirical structures with the identifiers 1PKW, 1XW5 and 8GSS for the human alpha, mu and pi isoforms, as well as 6RWD and 1OE8 for Sj26GST and Sh28GST respectively, were identified as the most eligible. First, these five structures were obtained, and their A chains sub-set and loaded into individual bio3d objects. Residue sequences of the resulting objects were then used to enrich the dataset by employing a basic local alignment search tool (blastp) [100]. A blastp search was invoked through bio3d for each source object, where the top scoring hits beyond a threshold –log (*E*-value) of 240 were retained. Accessing blastp from bio3d does not allow configuration of filters for the search query, resulting in irrelevant hits including homology models, mutants and isoforms from other species. Therefore, analogous blastp protocols were performed via the web-based service [100] where species-specific filters and an exclusion criterion for homology models were incorporated. Based on these outcomes, the bio3d blastp results for each source were pruned to eliminate irrelevant hits. Merging of the pruned datasets containing 22 structures for hGSTA, 6 for hGSTM and hGSTP, 4 for Sj26GST and 5 for Sh28GST, resulted in a total of 48 structures including the sources.

9.3 Alignment and superposition

A multiple sequence alignment was conducted on the GST dataset using multiple sequence comparison with log expectation (MUSCLE), version 3.8.31 [101], which is accessible through bio3d. The resulting alignment

object containing 176 non-gap positions out of a total 247 residues, was employed as a universal residue indexing scheme for the dimer interface superposition. Corresponding indices for this domain were mapped to positions 57 – 131 by selecting the Sj26GST (6RWD) residues in chain A within 10.0 Å of chain B. Local multiple structure superposition was then performed with reference to 6RWD by specifying these dimer interface bounds. Superposed coordinate data was finally saved to a matrix object for analysis and written to a file directory for visualisation.

9.4 Structural variation

Root-mean-square fluctuation (RMSF) graphs revealed a right-skewed distribution. A 10.0% truncated average was calculated across the set of 126 non-outlying elements, to capture a representative measure of the overall fluctuation for comparison. Principal component analysis (PCA) resulted in 528 eigenvalues of which the first three expressed 95.60% of the variance in the superposed dataset. The first three principal components including the per-residue decomposition of each axis were plotted.

9.5 Docking and interaction analysis

The crystal structure of each source isoform was loaded into bio3d where solvent, ions and heterogenous species, including the C chain of hGSTP (8GSS), were eliminated. Global superposition with respect to Sj26GST (6RWD) followed to generate an overlapping region upon which to delineate a consistent docking site. Then, UCSF Chimera was used to prepare the target structures by addressing missing atoms and implementing an energy minimisation protocol. Minimisation was applied to optimise the geometry of the BSP ligand (DB13215) to identify a stable conformation, after retrieval from DrugBank [102]. Docking into each of the five targets was fulfilled by Autodock Vina [95, 96], which was configured to produce 10 poses for each complex, with a maximal inter-pose energy difference of 3.00 kcal/mol. Sampling of poses was restricted to a 25 × 25 × 25 Å region positioned at the centroid of ellagic acid, the cocrystal inhibitor of the Sj26GST isoform (6RWD).

10

Outcomes

This chapter discusses the outcomes of the comparative dimer interface study of human and Schistosoma spp. glutathione transferase enzymes. The principal components of the coordinate data are first presented to report on the structural variation of the isoforms. Second, the binding parameters of BSP are documented to explain its empirical behaviour. We then resolve the major sources of structural variation in the dataset.

10.1 Superposition and principle component analysis

ROOT-MEAN-SQUARE FLUCTUATION (RMSF) of the 176 aligned positions are presented in Figure 10.1. The data show positive outlying regions spanning the dimer interface domain at residues 76 – 78 and 96 – 121, with respective average RMSF values of 2.27 ± 0.00 Å and 4.12 ± 1.22 Å. Upon comparison with the collective truncated average of 1.73 ± 0.12 Å, the magnitude and frequency of these outliers, present in the GST target binding domain, suggest the likelihood of distinct orientations of ligand-stabilising residues. These occurrences underscore the unique inhibitor binding characteristics at the dimer interface of these five GST isoforms.

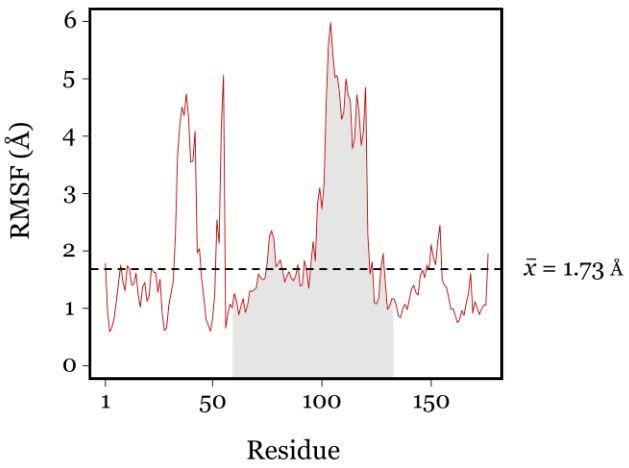


Figure 10.1. Root-mean-square fluctuation of the superposed 48-structure GST dataset.

The first principal component (Figure 10.2 A) clearly distinguishes all but the human mu and pi GSTs. A 73.55% variance captured here highlights the hGSTA isoforms as the most dissimilar from the parasite GSTs, in comparison with the other human isoforms mu and pi. These hGSTs are instead more closely related to, yet still differentiated from the Sj26GSTs. Little resemblance of the Sh28GSTs to any of the other GSTs is presented here. The 16.40% variance of the second principal component (Figure 10.2 A and B) exists only between the 28-kDa isoforms and the remaining GSTs. Distinction between the 26-kDa parasite GSTs, the hGSTM forms and the hGSTPs is captured by the third principal component of 5.65% variance, as shown in Figure 10.2 B. Here, no difference between hGSTA and Sh28GST is observed.

It is therefore understood that the first principal component displays the variance of the human alpha GSTs relative to both parasite forms, as well as hGSTs mu and pi relative to the 28-kDa parasite GSTs. The second principal component demonstrates an almost exclusive 16.40% variability in the Sh28GSTs compared to the hGSTs, confirming their disparity. Hence, all three of the human GST forms are adequately distinguished from the 28-kDa parasite GSTs, whereas it is only the hGST alpha which

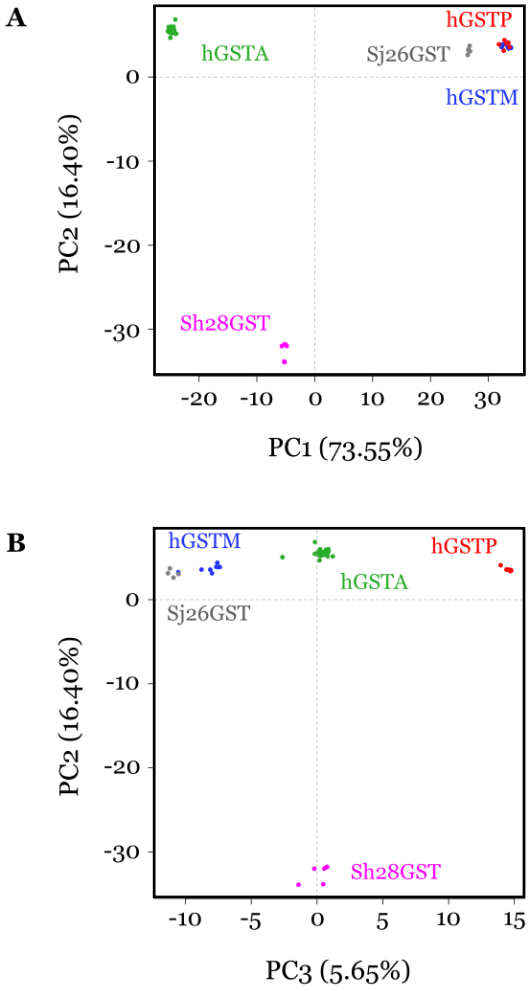


Figure 10.2. Principal component analysis of the coordinate dataset.

shows noteworthy variation from the 26-kDa parasite forms. Although limited contrast exists between the Sj26GST and the human mu and pi forms, the prevalence of their disparity in both principal components two and three reveals an appreciable variance.

10.2 Molecular docking

The best poses of BSP were observed across all the targets to exist in the dimer interface pocket at the core of the GST enzyme binding channel (Figure 10.3). Association with this particular site places BSP in intimate proximity to both catalytic domains of GST, and may therefore explain

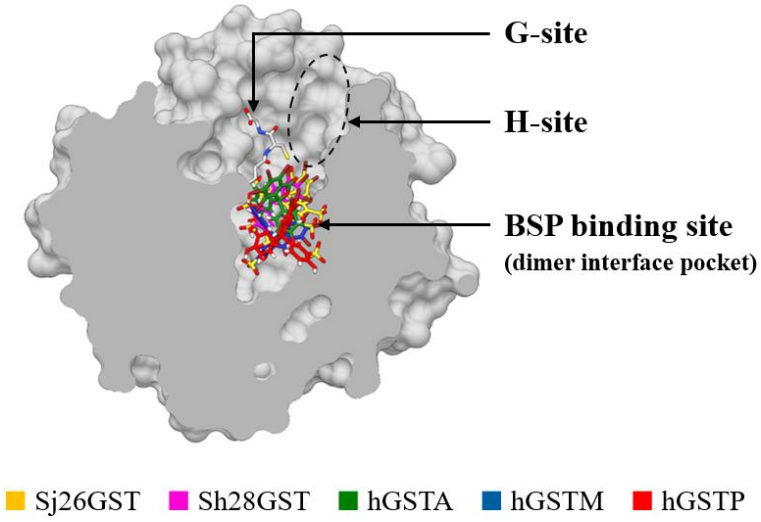


Figure 10.3. Top-scoring docking poses of BSP across the five GST targets.

its inhibitory behaviour. Binding energies (Table 10.1) revealed clear preference for the two parasite isoforms with an average of 8.90 versus the 7.30 kcal/mol across the three hGSTs. Average ligand root-mean-square deviation was calculated on an identical atom-wise basis from 9 out of a total 10 poses. Binding interactions of each complex were pruned to exclude hydrogen bonds with donor-acceptor distances beyond 3.50 Å, and hydrophobic contacts beyond 4.20 Å, owing to their modest energy contribution in the absence of cooperative effects.

The ligand interaction profiles of each of the five complexes are shown in Figure 10.4. Synergistic stabilising effects are observed in the schistosome GSTs (Figure 10.4 A and B), where interactions are made with both the phenol moiety of BSP, and either the adjacent or opposite sulfonate. This interaction pair which is formed with a sulfonated polycyclic ligand such as BSP, has been shown to predominate the stabilisation effects offered to similar ligands such as anthraquinone dye derivatives [91]. The hGST forms, in comparison (Figure 10.4 C, D and E), stabilise the sulfonate whilst neglecting the BSP aromatic groups.

Table 10.1. Molecular docking parameters of BSP in complex with five GST isoforms.

Target	Binding energy (kcal/mol)	Average ligand RMSD (Å)
Sh28GST	-9.5	2.1
Sj26GST	-8.3	2.4
hGSTA	-7.8	12.3
hGSTM	-7.7	16.3
hGSTP	-6.4	8.6

Despite the presence of the short-range pi-interactions which favour the parasite binding modes, similarities in hydrogen and ionic bonding exist amongst the human and parasite complexes. The geometric parameters of these interactions were investigated to assess their contribution to the differential binding energies observed. Student's t-test performed on the donor-acceptor distances of hydrogen bonding interactions revealed a statistically significant difference (p = 0.045) between the schistosome and human complexes, with the former exhibiting 0.28 Å closer binding when averaged across the respective targets. Significant differences did not prevail for ionic bond distances (p = 0.062) and hydrogen bonding angles (p = 0.139), although the parasite complexes displayed shorter average salt bridge lengths.

Hence, larger distance-dependent energy contributions, accompanied with both phenol- and sulfonate-stabilising interactions, corroborate the increased binding energies observed in the parasite targets. In further support is the lower average ligand root-mean-square deviation (RMSD) of the Sj26GST and Sh28GST complexes compared to that of the hGSTs, which suggests decreased mobility, indicative of binding strength. By contrast, long-range bonding in the hGSTs caused by partial stabilisation of the phenol-sulfonate, contribute to somewhat labile ligand-receptor

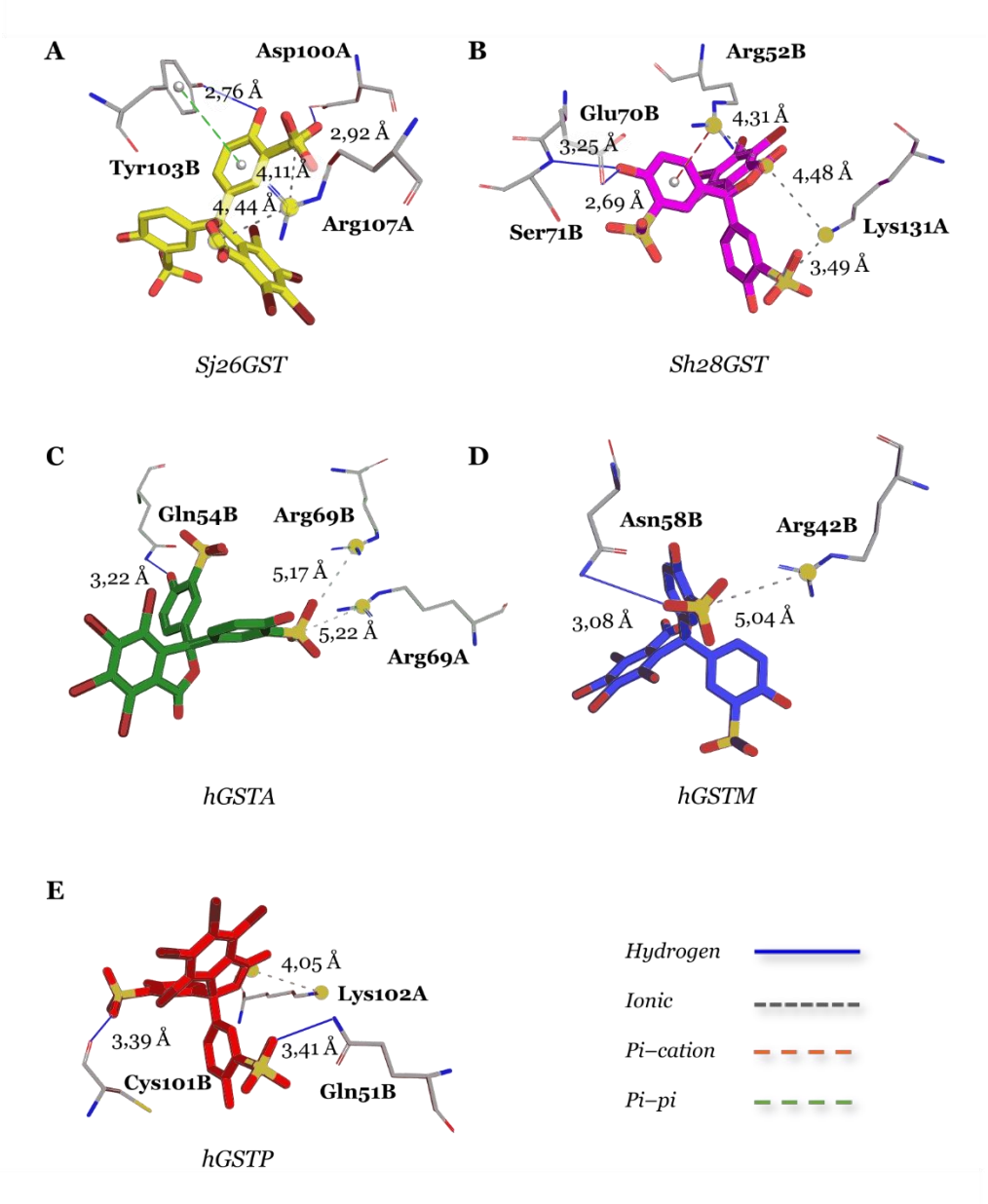


Figure 10.4. Ligand interactions of the top-scoring BSP pose in the dimer interface of five GST structures.

interactions, producing mobile binding poses likely to spontaneously dissociate. It is therefore determined that proximal interactions of BSP with the schistosome GSTs result in stable, tightly bound complexes.

10.3 Parasite complex analysis

Per-residue plots of the PCA present the contribution of each element to the variance displayed across the respective first and second principal components (Figure 10.5). Indices of relevant residues, as revealed by the docking interaction profile data were converted into the universal 1–176 numbering scheme for comparison. The Asp, Tyr and Arg residues which stabilise BSP in the Sj26GST complex, correspond with the respective aligned positions 114, 117 and 121. RMSF and per-residue plots (Figure 10.5 A) correlate the dramatic spatial variation of these residues across the dataset, with an average RMSF of 3.49 Å coupled with a substantial contribution to principal component one. Such observations disparage the involvement of these particular residues in the docking profiles of the remaining BSP complexes. Therefore, the Sj26GST complex possesses a unique interaction profile for the BSP ligand, which includes short-range hydrogen- and ionic bonding.

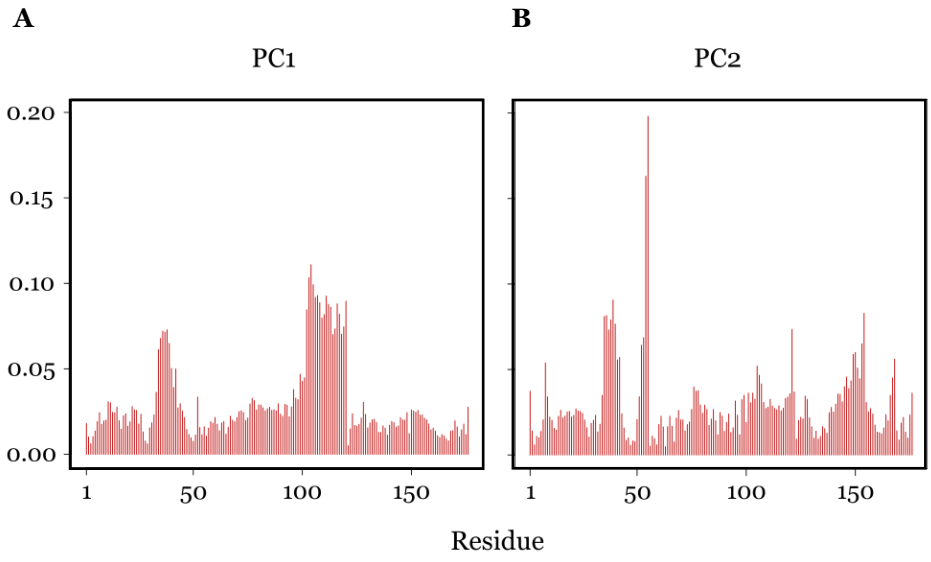


Figure 10.5. Decomposition of the first two principal components of the GST coordinate dataset, showing the contributions of each residue.

Arg , Glu and Lys amino acids in the Sh28GST complex, corresponding to respective positions 62, 80 and 94, exhibit moderate variation with neither principle component one, nor two demonstrating any fluctuation

(Figure 10.5). Conservation of these residues across the hGST complexes are therefore observed, most prominently of the Arg residue at position 62, which stabilises the ligand sulfonic acid in the hGSTA and hGSTM forms. In comparison, the orientation of BSP in the Sh28GST complex allows for intimate stabilisation of the phenol by the same Arg residue. The Sh28GST Lys present at position 94 completes the characteristic phenol-sulfonate binding pair by securing the opposing sulfonate group of BSP. While this residue remains conserved in the hGSTP complex, the necessary stabilising effect does not prevail, with the Lys instead forming a bond with the smaller hydroxyl moiety of BSP. Finally, the hydrogen bond acceptor Glu at position 80 offers a close-proximity interaction with BSP. This provides a greater stabilisation effect than the analogous hGST Gln and Asn residues, which possess weaker amide electrophiles. In short, the structural deviation and residue substitutions in the hGST complexes manifest per-residue interaction profiles which are subsets of that of the Sh28GST complex.

11

Conclusion

In this chapter, we gather the principal findings from the comparative dimer interface analysis, and report on a few of its major implications.

The dimer interface of the *Schistosoma* GST enzyme is the prime inhibition target for the novel treatment of human schistosomiasis. In this study, we produced a comparative structural assessment of an enriched GST dataset bearing 48 structures sourced from the cytosolic schistosome isoforms Sj26GST and Sh28GST, as well as analogous hGST forms alpha, mu and pi. The resulting superposed coordinate data were interpreted in the context of the interaction profiles of BSP, a molecule shown to bind preferentially with schistosome GST.

Root-mean-square fluctuation and principal component analysis data have illustrated significant variation among the dimer interface residues of the human and target enzymes, which indicates their unique binding capabilities. This highlights the potential for selective inhibition of the schistosome GST forms. Docking of BSP was shown to occur within the dimer interface pocket. This site exists within the GST binding channel, proximal to the catalytic domains, and may therefore explain the inhibitory behaviour of the BSP ligand. Binding energies of the schistosome complexes were consistently more favourable compared to their human counterparts, supporting the empirical observations of this molecule. Ligand profiling revealed a repertoire of interactions in the schistosome GST complexes which stabilise the phenol-sulfonate group of BSP. This occurrence is shared across various similar polycyclic sulfonates such as

anthraquinone dyes, allowing anchoring of the ligand core for intimate binding. BSP ligand interaction profiling of hGSTs revealed subsets of interactions from the Sh28GST complex. Despite the stabilisation of the BSP sulfonate in these hGST complexes, missing aromatic interactions impart excessive bond distances, leading to transient ligand poses.

The accomplishments of this work demonstrate a reliable metric for the generation and validation of pharmacophore models of a non-substrate site in the schistosome GST binding channel. These approaches bear the potential to expedite screening and optimisation protocols in the drug development schemes for novel antischistosomals, such that the disease burden of human schistosomiasis may readily be addressed.

IV

PHARMACOPHORE MODELS

This section is based on:

VALLI, A. and ACHILONU, I.

Molecular dynamics-derived pharmacophores of *Schistosoma* glutathione transferase in complex with bromosulfophthalein: screening and analysis of potential inhibitors

Journal of Molecular Graphics and Modelling

Volume 122

July 2023

12

Synopsis

In this chapter, we summarise the aims, methodologies and highlights of Schistosoma spp. GST pharmacophore generation, using molecular dynamics trajectories of bromosulfophthalein as an archetypal model.

We used *in silico* methods to validate the discriminative inhibitory properties of bromosulfophthalein (BSP) against the 26-kDa GST from *S. japonicum* (Sj26GST), and the 28-kDa GST from *S. haematobium* (Sh28GST), versus human GST (hGST) isoforms alpha (hGSTA), mu (hGSTM) and pi (hGSTP).

The use of BSP as an archetypal selective agent was harnessed to produce molecular dynamics-derived pharmacophores of the two GST targets. Pharmacophore-based screening using a large dataset of experimental and approved drug compounds was performed to produce a shortlist of candidates. The top candidate for each target was prioritised via molecular docking, yielding guanosine-3'-monophosphate-5'-diphosphate (G3D) for Sj26GST, and quercetin-3'-*O*-phosphate (Q3P) for Sh28GST. Comparative molecular dynamics studies of both candidates compared to BSP showed similar characteristics of binding stability and strength, suggesting their potential to emulate the inhibitory effects of BSP.

13

Methods

In this chapter, we describe the various approaches used to generate pharmacophore models of the Schistosoma GST drug targets. We first document the protocol used for generation and preliminary analysis of molecular dynamics trajectories. Next, procedures used to determine the pharmacophore features are outlined. Finally, we explain the methods employed during screening and stability assessment of the candidates.

13.1 Generation of complexes

THE NEWEST AND BEST RESOLUTION wild-type structure of each target GST enzyme was selected from the Protein Data Bank (PDB) [99] (www.rcsb.org). Structures bearing the PDB identifiers 6RWD and 1OE8 were retrieved to represent the experimental Sj26GST and Sh28GST respectively. Similarly, the respective hGSTA, hGSTM and hGSTP controls were represented by 1PKW, 1XW5 and 8GSS. The Schrödinger, Inc. software suite release 2019-3 [103], was used to generate the complexes. The structures were first processed using the Protein Preparation Wizard [104], by removing solvent molecules and heterogenous groups, as well as by addressing missing atoms and bonding data. Next, ionisable side chain states were resolved at pH 7.0, followed by optimisation of the hydrogen bonding network. Finally, restrained minimisation of the target geometry was performed using the OPLS2005 force field, with heavy-atom deviation restricted to 0.30 Å.

The BSP ligand, retrieved from PubChem [105] with compound identifier 102371197, was prepared using the LigPrep utility [106]. First, neutralising sodium ions were removed. Next, the preferred ligand ionisation state at empirical inhibition conditions of pH 7.0, was identified as the BSP (II) anion. Thereafter, conjugate gradient energy optimisation with the OPLS2005 force field was implemented to produce a stable three-dimensional structure in preparation for docking.

Rigid-receptor docking was fulfilled by Glide XP [107] using a 25 × 25 × 25 Å bounding box centred at the dimer interface pocket. This central position was identified based on the centroid coordinates of ellagic acid, a symmetrical cocrystal ligand present in the dimer interface core of 6RWD. Superposition of all structures, performed with Quick Align [103], ensured consistent positioning of the GST targets before defining the bounding box. Flexible ligand sampling was implemented, including conjugate gradient energy minimisation with a maximum of 100 steps, and scaling of van der Waals radii to 80% magnitude.

Docking scores of the single best BSP pose averaged -7.24 kcal/mol with the schistosome GST targets, compared to -4.16 kcal/mol with human GST. The relative scoring correlates with our previous molecular docking experiments of BSP bound to these GST targets, where schistosome GST complexes recorded an average -8.90 kcal/mol versus -7.30 kcal/mol of the human GST forms [37]. The weaker overall docking scores obtained here are due to increased penalties which are implemented in Glide XP, such as those related to the de-solvation of polar and charged groups.

13.2 Molecular dynamics (MD)

Desmond 5.9 [108] (Schrödinger, Inc.) was used to perform the MD simulations. The MD systems were initialised by positioning each protein-ligand complex in a periodic orthorhombic bounding box containing TIP3P (transferable intermolecular potential with three-points) explicit solvent, whose edges were positioned at a minimum 10 Å buffer distance from the solute surface. An appropriate number of sodium or chloride ions were added to neutralise each system.

The systems were first processed by executing a hybrid energy minimisation protocol, which incorporates quasi-Newton and steepest descent optimisation algorithms. Additional minimisation was performed via a series of short simulations using a combination of the canonical NVT (constant atom number, volume and temperature) and NPT (constant atom number, pressure and temperature) statistical ensembles to relax the geometry of the systems. Production simulations were performed for 300 ns using the NPT ensemble, with temperature and pressure of 300 K and 1.0 atm respectively. The OPLS2005 force field was used to resolve the energy of the system, which was captured along with each trajectory frame at 150 ps intervals. The same MD approaches were used to simulate the apo experimental and control targets.

13.3 Trajectory analysis

The Simulation Interactions Diagram module of Maestro 12.0 [103] (Schrödinger, Inc.) was used to analyse the trajectories relative to the input systems. Binding stability of schistosome and hGST complexes, as well as that of their respective apo forms, was assessed with protein and ligand root-mean-square deviation (RMSD). Per-residue evaluation of target structural deviation in the complexes was provided by root-mean-square fluctuation (RMSF) data. Next, frequency-based interaction profiles from the stable trajectory portion of each complex were used to determine relative BSP binding strength.

Clustering of stable schistosome protein-ligand trajectories followed. This was achieved by first sub-setting the stable trajectory portions of each system, and thereafter using Trajectory Clustering from the Schrödinger software suite release 2019-3, to sample every 5 frames and merge those within a maximum heavy-atom RMSD of 2.0 Å, to form a single representative structure. The resulting structures were used to display stable binding site conformations of each of the schistosome GST complexes, which were then compared to the respective pre-stable trajectory clusters generated in the same manner.

13.4 Pharmacophore generation and screening

Phase [109] (Schrödinger, Inc.) was used to generate the pharmacophores from stable clustered representations of each schistosome GST complex. Manual selection of features was performed using the corresponding frequency-based interaction profiles.

Each ligand from the DrugBank [110] version 5.1.9 structure file dataset was processed using LigPrep [106], by first enumerating ionisation and tautomeric states at pH 7.0, followed by an energy minimisation protocol with OPLS2005. The resulting 23 928 ligands were screened against the respective models using a minimum of four matching pharmacophore sites. Successful candidates were identified according to a fitness score measured on a scale of -1.00 to 3.00.

13.5 Comparative trajectory analysis

The top five candidates arising from each virtual screening experiment were docked into their respective targets using the same Glide XP [107] docking procedure as with generation of the BSP complexes.

Next, the best-scoring complex of each target was subjected to a 300 ns MD simulation as previously described. Finally, radius of gyration was measured with reference to the longest diameter of each GST dimer. This provided an assessment of target compactness in response to binding of the ligand, effectively resolving the structural effects induced by BSP.

14

Outcomes

In this chapter, we show and discuss the results gathered from various phases in the design of molecular dynamics-derived pharmacophores. We first analyse the trajectories of the molecular dynamics simulations to determine the stability and mobility characteristics of each system. Frequency-based interaction profiles of the target complexes are then presented, followed by a description of the respective ligand equilibration events. Next, the features of each pharmacophore are displayed, and the top candidates are identified via screening of ligand databases. Finally, we provide a comparative trajectory analysis of the candidates.

14.1 Stability analysis

BOTH THE PROTEIN BACKBONE and ligand RMSD of the Sj26GST-BSP complex (Figure 14.1 A) exhibit rapid equilibration observed within 60 ns, relative to the apo target which has a 50 ns equilibration phase. After equilibration, the 240 ns working trajectory of the protein-ligand system reflects strong ligand RMSD convergence to 1.66 ± 0.13 Å, and protein backbone RMSD convergence to values of 1.84 ± 0.11 Å and 2.07 ± 0.14 Å for the bound and apo forms of Sj26GST respectively.

Similarly, the Sh28GST-BSP complex shown in Figure 14.1 B demonstrates relatively short equilibration phases of 75 ns for the protein-ligand system, to the 25 ns of the apo target. Strong ligand RMSD convergence to 1.15 ± 0.15 Å was also recorded in the 225 ns working

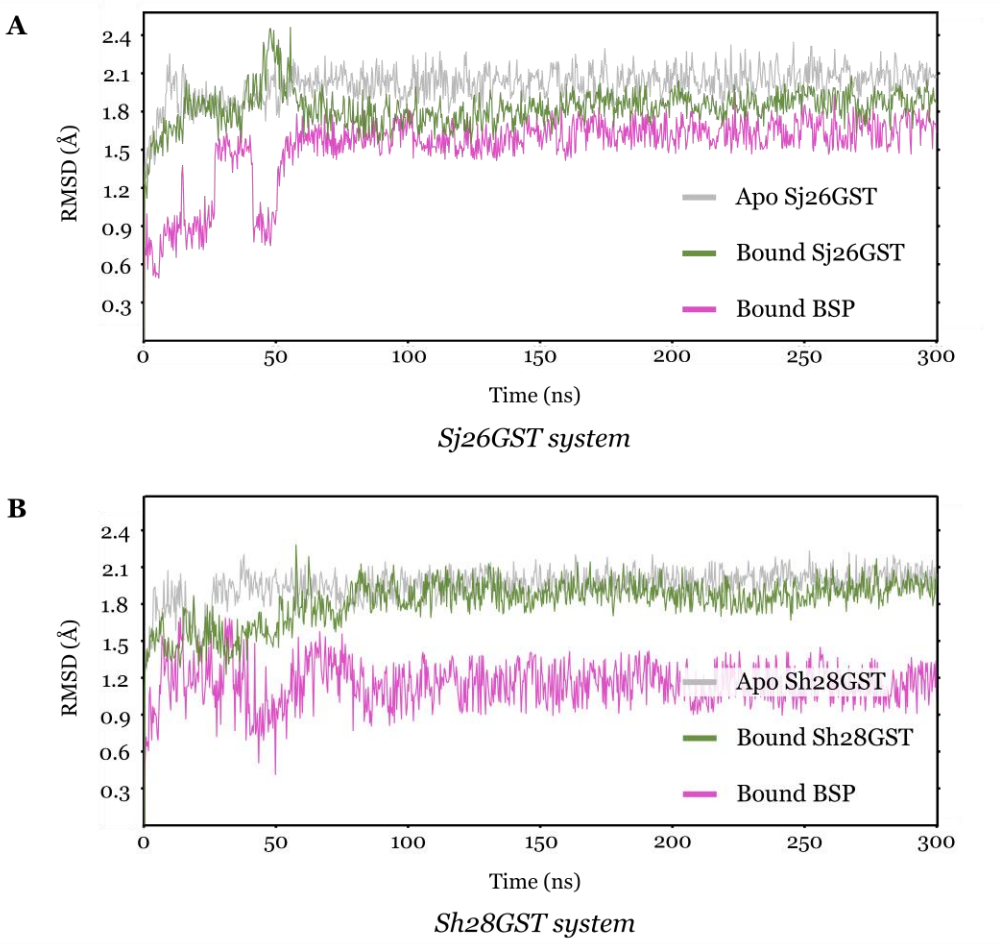


Figure 14.1. Root-mean-square deviation of the schistosome GST systems.

trajectory, together with convergence of the bound protein backbone to 1.87 ± 0.22 Å, and the apo protein backbone to 2.01 ± 0.08 Å.

In total, the large working trajectory proportions of 80% and 75% of the respective Sj26GST and Sh28GST protein-ligand systems, coupled with the strong convergence of RMSD data, highlight the stability of BSP in complex with the schistosome GST targets.

The apo systems of hGSTA, hGSTM and hGSTP share quick equilibration events with their schistosome counterparts as shown in Figure 14.2, with respective time phases of 10, 55 and 50 ns followed by strong RMSD convergence. However, the stability characteristics of the three hGST protein-ligand systems differ drastically from the schistosome complex-

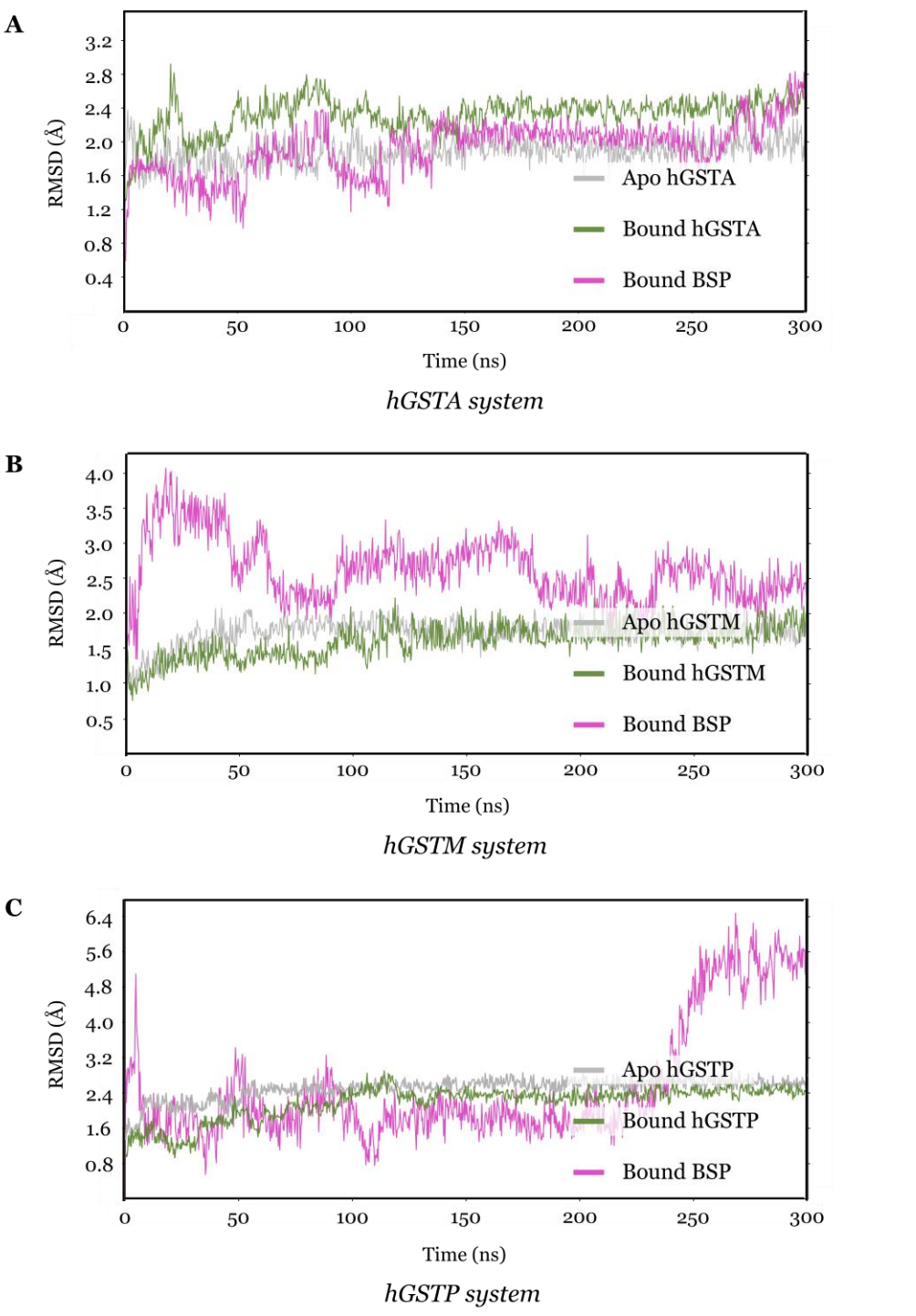


Figure 14.2. Root-mean-square deviation of the human GST systems

es. The hGSTA-BSP system (Figure 14.2 A) equilibrates after a prolonged 140 ns, lasting only until 240 ns, where dramatic deviation of the ligand RMSD is observed. This relatively short pseudo-equilibrated working trajectory, which is succeeded by regions of large RMSD fluctuations, demonstrates erratic ligand geometries consistent with instability of the complex. Nevertheless, convergence of the ligand RMSD in the 100 ns pseudo-stable phase of the system was measured to 2.11 ± 0.15 Å, and that of the bound protein backbone to 2.38 ± 0.18 Å. The backbone RMSD of the apo hGSTA converged to a value of 1.93 ± 0.23 Å within its 290 ns working trajectory.

Despite stable protein backbone RMSD data after 125 ns in the hGSTM-BSP system (Figure 14.2 B), no equilibration is recognised due to lack of ligand RMSD convergence. Here, the respective apo and bound protein backbone RMSD converge to similar values of 1.85 ± 0.17 Å and 1.87 ± 0.22 Å, calculated across the working trajectories of 175 ns and 245 ns respectively. The inconsistent ligand RMSD data share the same characteristic of binding instability with the hGSTA-BSP complex.

The hGSTP-BSP system (Figure 14.2 C) likewise demonstrates a pseudo-stabilised trajectory spanning 115 ns to 215 ns, with the ligand RMSD data converging to 1.84 ± 0.34 Å, and bound protein backbone RMSD converging to 2.42 ± 0.24 Å, compared to 2.56 ± 0.18 Å of the apo hGSTP backbone measured for its 250 ns working trajectory. The time phase beyond 215 ns of the protein-ligand system is burdened by a dramatic surge in ligand RMSD, which confirms the instability of the complex.

In total, the pseudo-stable working trajectories of 100 ns exhibited by the hGSTA- and hGSTP-BSP systems suggest weak, transient interactions which do not provide sufficient binding stability to the complexes. This outcome manifests as profound fluctuations of ligand geometries spanning the majority of each simulation. Similar occurrences are likely to be observed in the side chain RMSD profiles of the bound proteins, as the systems attempt to stabilise each complex by adjusting protein side chain geometries. The weakest protein-ligand interaction is observed in the hGSTM-BSP complex, to the extent that even transiently stable lig-

and geometries are absent throughout the entire 300 ns simulation. In all the hGST protein-ligand systems, comparison of RMSD convergence values is inconsequential to the assessment of binding stability, due to either exceedingly short or non-existent working trajectories. Finally, all apo systems demonstrate adequate backbone RMSD convergence, which signifies that the unbound proteins are stable. Hence, it is altogether determined that the binding stability of BSP with schistosome GST far exceeds that with human GST.

14.2 Side chain fluctuations

The average side chain root-mean-square fluctuation (RMSF) of the BSP complexes was evaluated to assess local deviation in the GST structures across the 300 ns simulation. As shown in Figure 14.3, the Sj26GST- and Sh28GST-BSP complexes displayed significantly lower RMSF values (p-value < 0.001), with less drastic swings compared to all the human GST counterparts. This observation is consistent with superior stability of the schistosome GST complexes. Furthermore, the most pronounced swings in the RMSF data of the schistosome complexes appear to be localised to the dimer interface, compared to the hGST targets where they appear consistently throughout most of the protein structure. A likely explanation for this occurrence in the hGST complexes is the steric effect of highly variable, destabilised side chain geometries in the dimer interface, which lead to disruption of neighbouring side chain orientations.

It is therefore identified that the source of structural variation in the BSP complexes of Sj26GST and Sh28GST is largely constrained to the target dimer interface, where side chain orientations are sampled to achieve stability of the ligand complex. In comparison, the hGST-BSP complexes demonstrate significantly greater RMSF swings in the dimer interface and neighbouring regions, consistent with the lack of adequately stable side chain geometries. These data correlate with the stability observed in the schistosome GST complexes of BSP, and the lack thereof in the corresponding hGST complexes.

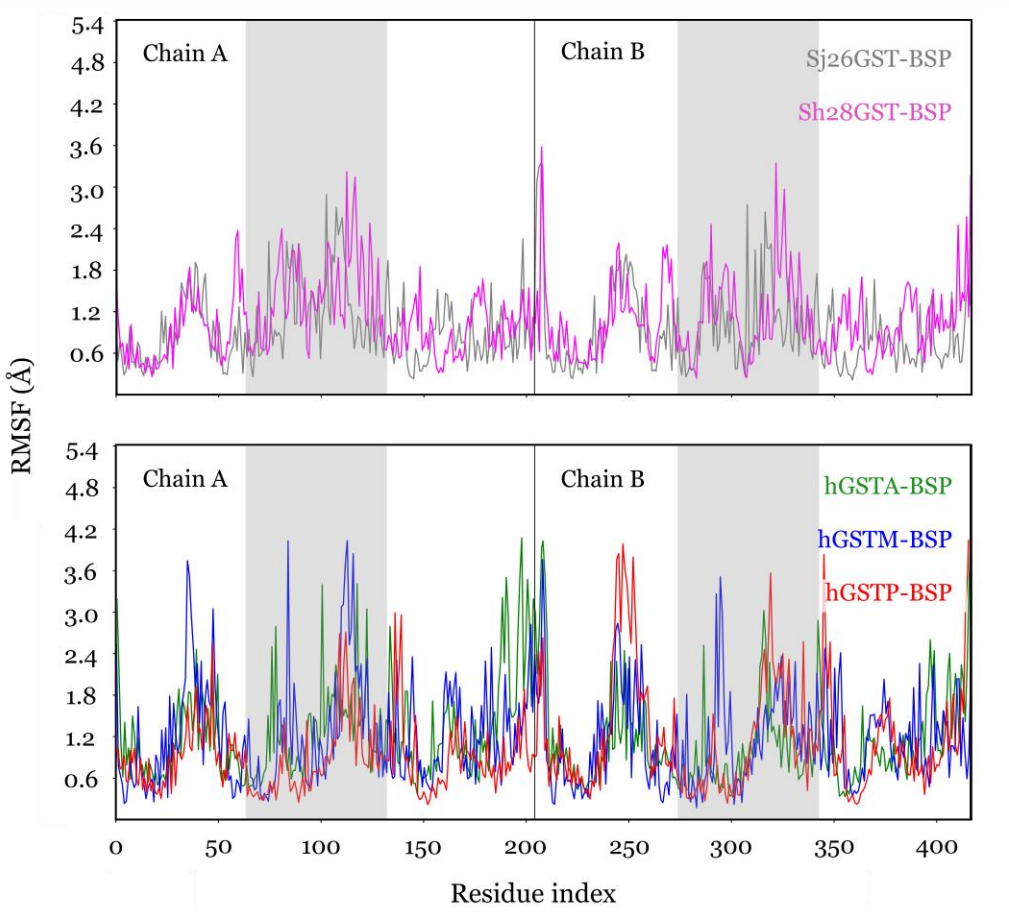


Figure 14.3. Root-mean-square fluctuation of the GST targets averaged across the 300 ns simulation.

14.3 Interaction profiles

The most common protein-ligand interactions present in the respective 240 and 225 ns working trajectories of the Sj26GST- and Sh28GST-BSP systems were displayed with a 50% frequency threshold (Figure 14.4). Note the presence of three water bridges in the Sj26GST complex, as well as a water molecule responsible for coordinating Asp100 (Figure 14.4 A). The presence of solvent, including one water molecule in the Sh28GST-BSP complex (Figure 14.4 B), indicates the considerable water exposure of the schistosome GST binding channel. This occurrence is appropriate, considering the requirement for binding variably-sized substrates at this region, particularly in the absence of any known conformational adapta-

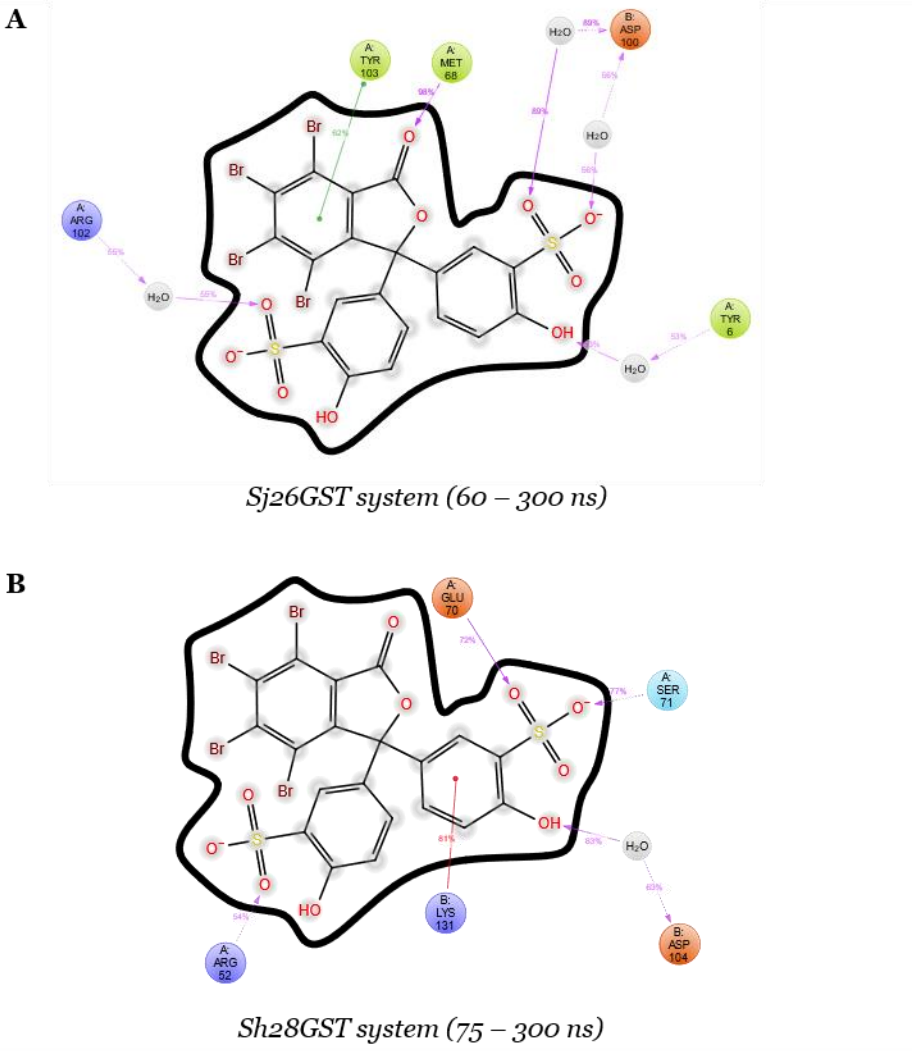
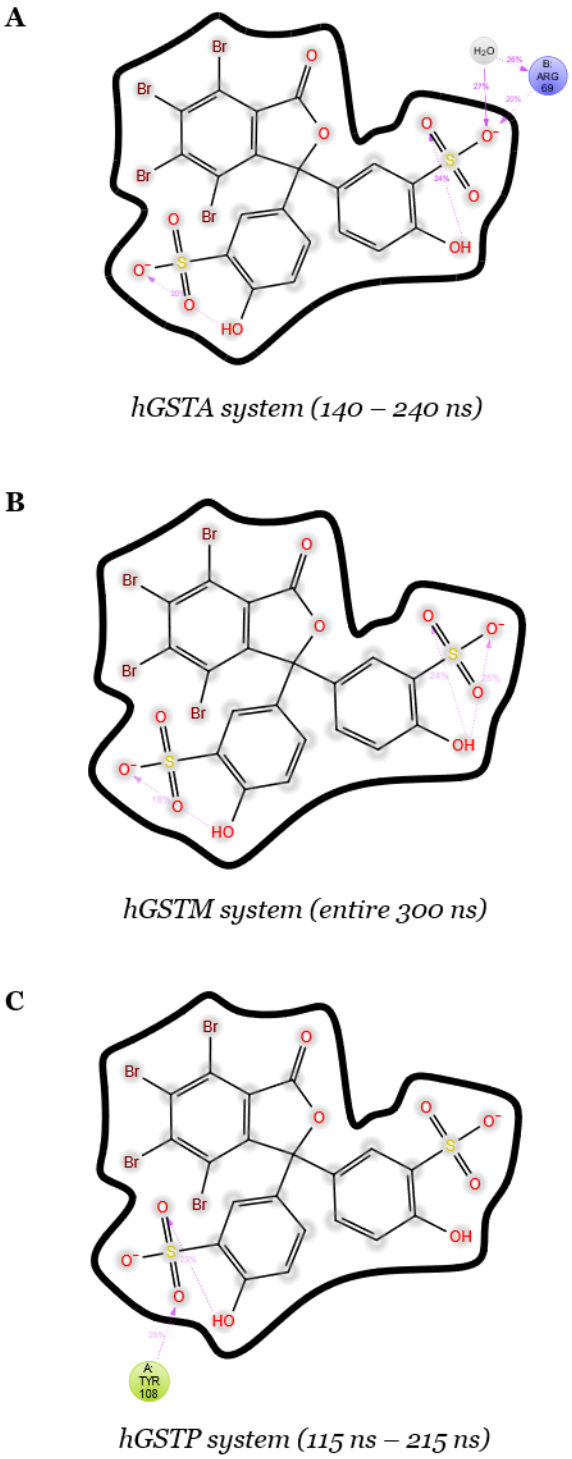


Figure 14.4. Interaction profiles from the working trajectory portions of the schistosome GST complex simulations, based upon a 50% frequency threshold.

tions. Furthermore, comprehensive stabilisation of the ligand sulfonate and aromatic groups appears in the schistosome binding profiles, which concur with our previous study of BSP in complex with the Sj26GST and Sh28GST targets [37]. Finally, the large working trajectories observed in the schistosome protein-ligand simulations offer substantial confidence in the interaction profiles, which makes them suitable for the purpose of pharmacophore generation.

Binding profiles of hGST-BSP systems were evaluated across the pseudo-stable working trajectories of 100 ns each for the hGSTA and hGSTP targets, and across the entire 300 ns trajectory for the hGSTM target (Figure 14.5). In comparison with the schistosome complexes, human GST interaction profiles of 50% frequency threshold showed no prominent stabilising residues. Reduction of the frequency threshold to 30% revealed poorly populated residue interactions in the hGSTA- and hGSTP-BSP systems (Figure 14.5 A and C). Here, interaction with a single ligand sulfonate group is inadequate to provide stable binding. Furthermore, the hGSTM-BSP system (Figure 14.5 B) failed to produce any notable interactions with the reduced 30% frequency threshold. The lacking protein-ligand interaction in these hGST-BSP complexes, compared to that of schistosome GST, is consistent with our previous study [37] which demonstrates the disparate binding channel architecture of these enzymes. This renders the hGST forms incapable of forming the appropriate interactions with BSP. In addition, the scarcity of solvent in the hGST complexes is likely due to the presence of intra-ligand bonding, which displaces water molecules and fulfils the stabilisation energy with entropic effects, instead of the enthalpy afforded by water bridges and coordination complexes.

Hence, the compatibility of the binding channel, coupled with the presence of bridging and coordinating water molecules in the Sj26GST- and Sh28GST-BSP complexes, produce a richer interaction profile to that of hGSTA, hGSTM and hGSTP. The ligand interaction data therefore support the relatively greater binding strength of BSP for schistosome GST.



hGSTA system (140 – 240 ns)

hGSTM system (entire 300 ns)

hGSTP system (115 ns – 215 ns)

Figure 14.5. Interaction profiles from the pseudo-stable trajectory portions of the human GST complex simulations, based upon a reduced 30% frequency threshold.

14.4 Ligand equilibration

Clustering of the respective 240 and 225 ns working trajectories of the Sj26GST- and Sh28GST-BSP complexes was performed, followed by clustering of each pre-equilibrated trajectory portion. This provided representative equilibrated and pre-equilibrated binding site conformations of the schistosome protein-ligand systems, as shown in Figure 14.6. In this way, the principal motions responsible for achieving stable ligand states were resolved.

In addition to the translation of ligand coordinates in the Sj26GST-BSP system (Figure 14.6 A), the ligand exhibits a 135° rotation about the indicated C_{17} – C_{20} bond. This brings the BSP sulfonate in sufficient proximity for the water-mediated Asp100 interaction, and the adjacent hydroxyl in proximity to Tyr6. The predominant ligand motion in the Sh28GST-BSP system is a 180° rotation about C_{36} – O_{10} which promotes the water-mediated hydrogen bond with Asp104 (Figure 14.6 B).

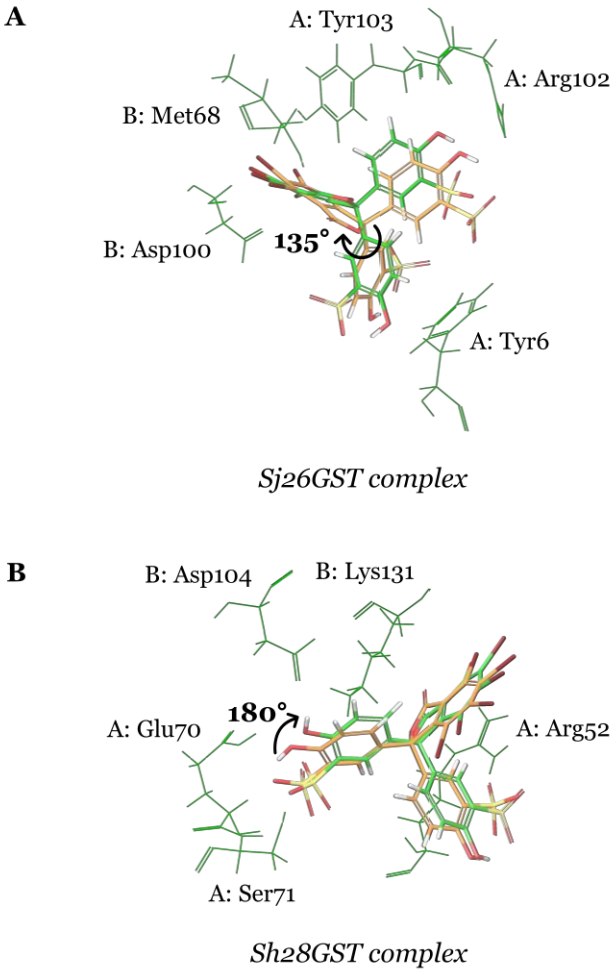


Figure 14.6. Representative equilibrated (green) and pre-equilibrated (orange) ligand orientations displaying the principal motions required for stabilisation of the schistosome GST complexes.

14.5 Pharmacophore generation

The respective five and four-feature pharmacophores models, derived from molecular dynamics simulations of the Sj26GST and Sh28GST targets in complex with BSP, are shown in Figure 14.7. Note the common features, including the characteristic aromatic-sulfonate stabilisation pair [37], which anchors the ligand in the binding site.

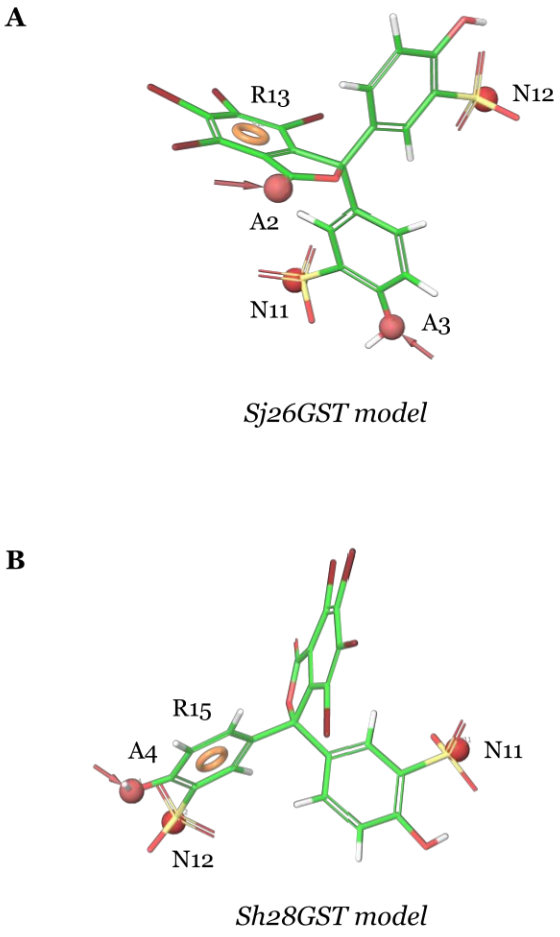


Figure 14.7. Pharmacophore models derived from the stable clustered representations of each schistosome GST form in complex with BSP. Features are shown superposed onto BSP.

14.6 Pharmacophore screening and docking

Screening of the 23 928 DrugBank ligands yielded a total of 39 ligands which exhibited fitness scores beyond 50% of the maximal score of 3.00. The top five ligand candidates arising from the screening study are displayed in Table 14.1, ranked according to their fitness score. Subsequent docking revealed guanosine-3'-monophosphate-5'-diphosphate (G3D) and quercetin-3'-*O*-phosphate (Q3P) as candidates possessing the most favourable binding energy for the Sj26GST and Sh28GST targets respectively. Both these ligands bear the full complement of features for each

Table 14.1. Fitness scores (-1.00 to 3.00) of the top five candidates determined via pharmacophore-based screening, including their subsequent docking scores.

Sj26GST	Fitness score	Glide XP score (kcal/mol)
BSP	3.00	-7.20
Guanosine-3'-monophosphate-5'-diphosphate (G3D)	1.92	-8.64
Calcein	1.90	-5.75
Phytic acid	1.88	-6.58
Fluorescein	1.86	-3.86
Phenolphthalein	1.82	-3.88
Sh28GST		
BSP	3.00	-7.28
Olsalazine	2.23	-6.67
Clinofibrate	2.14	-6.80
2-(4-chlorophenyl)-1-hydroxyguanidine	2.13	-7.03
Quercetin-3'-*O*-phosphate (Q3P)	2.07	-9.74
Mycophenolic acid	1.94	-7.35

model and were found to bind in the same dimer interface pocket as BSP. G3D belongs to a family of nucleoside polyphosphate compounds, which have not previously been investigated for antischistosomal properties. The parent compound of Q3P, a flavonoid called quercetin, is present in plant-based foods such as capers, onions and garlic, all of which have been associated with some degree of antischistosomal activity [111-113]. Consequently, both G3D and Q3P, in complex with their respective schistosome GST targets, were selected for comparative MD analysis.

14.7 Comparative molecular dynamics

The Sj26GST and Sh28GST systems, containing G3D and Q3P respectively, demonstrate similar stability properties as the corresponding BSP systems of each target, which are characterised by rapid equilibration and strong RMSD convergence (Figure 14.8). The Sj26GST-G3D system (Figure 14.8 A) equilibrates after a 100 ns period, with respective protein and ligand RMSD convergence measured to 2.52 ± 0.26 Å and 2.29 ± 0.15 Å along the remaining 200 ns working trajectory. An equilibration phase of 65 ns in the Sh28GST-Q3P system (Figure 14.8 B) is followed by 235 ns of ligand RMSD convergence to 1.95 ± 0.24 Å, and protein RMSD convergence to 3.72 ± 0.22 Å.

Marginal variation in the RMSD convergence values of the Sj26GST-G3D system, compared to the corresponding BSP system suggest largely si-

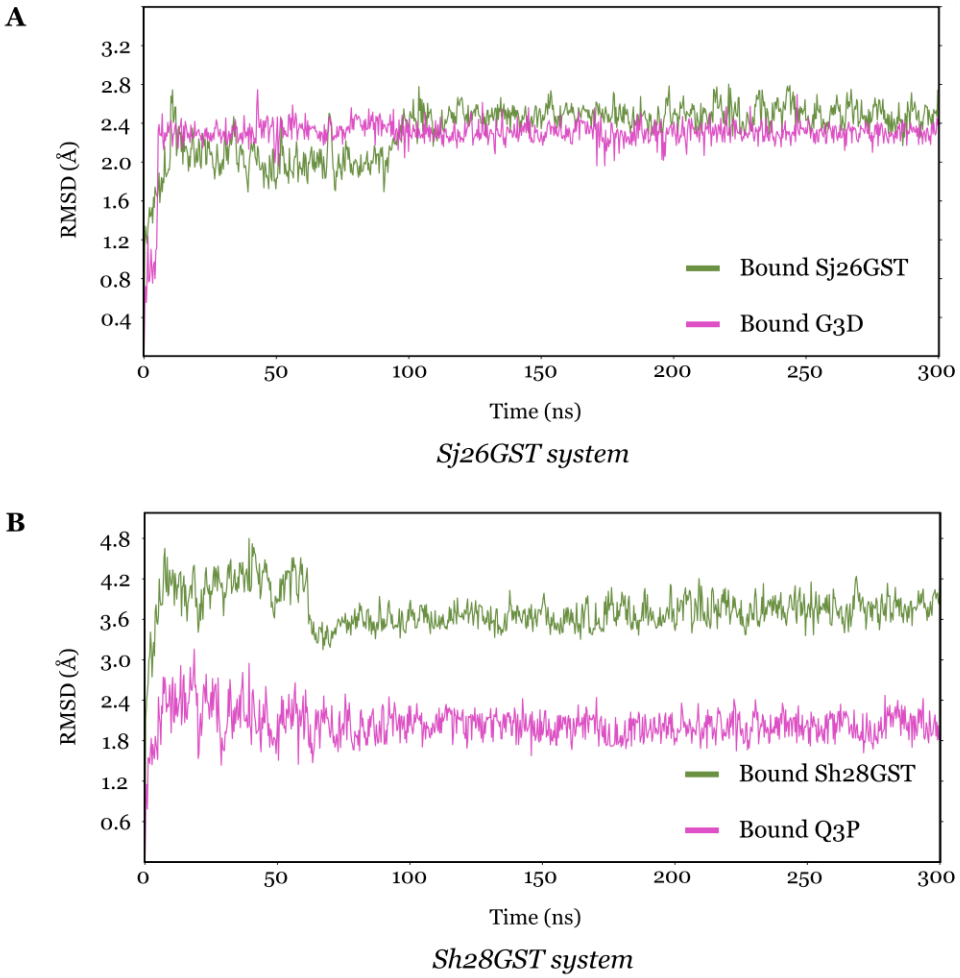


Figure 14.8. Root-mean-square deviation of top-scoring schistosome GST complexes.

milar binding characteristics. Therefore, much like BSP, the ability of G3D to exert selective inhibitory effects on the 26-kDa GST target is plausible. Average protein RMSD of 3.72 Å in the Sh28GST-Q3P system exceeds that of the corresponding BSP complex, indicating augmented mobility of the target in response to binding Q3P compared to BSP. Due to the matching ligand RMSD behaviour in this complex compared to Sh28GST-BSP, it is determined that the source of structural variation

exists outside of the ligand binding site. This observation bears similar implications for the empirical activity of Q3P, as with BSP and G3D.

Frequency-based interaction profiles generated with a 50% threshold from the respective 200 ns and 235 ns working trajectories of the Sj26GST-G3D (Figure 14.9 A) and Sh28GST-Q3P (Figure 14.9 B) systems, exhibit great similarity with their corresponding BSP counterparts. Conserved residues include Asp100, Arg102 and Tyr103 in the Sj26GST systems, as well as Arg52 and Ser71 in the Sh28GST systems. Additionally, glutamate residues appear in both 28-kDa schistosome GST systems.

Hence, the binding mechanics of the Sj26GST and Sh28GST targets in complex with the respective candidates G3D and Q3P bear strong resemblance to the BSP inhibitor. These ligands are therefore liable to produce selective empirical inhibition of the GST targets.

The radius of gyration plot of the Sj26GST target (Figure 14.10 A) shows no noteworthy change in global protein compactness, either in response to BSP or G3D, compared to the apo form. Specifically, binding of BSP is associated with a negligible increase in target compactness, whereas G3D binding results in a correspondingly minor decrease. These observations are insufficient to produce any inferences toward the potential empirical effects of G3D.

Binding of BSP and Q3P to Sh28GST both elicit a decrease in target compactness (Figure 14.10 B), more so with Q3P, where the average change is measured at 0.58 Å with respect to the apo target. This behaviour correlates with the increase in side chain RMSD of the Sh28GST-Q3P complex. Although the associated deviations are minimal, they may

indicate disordering of the global structure and therefore hold the potential to influence enzyme activity.

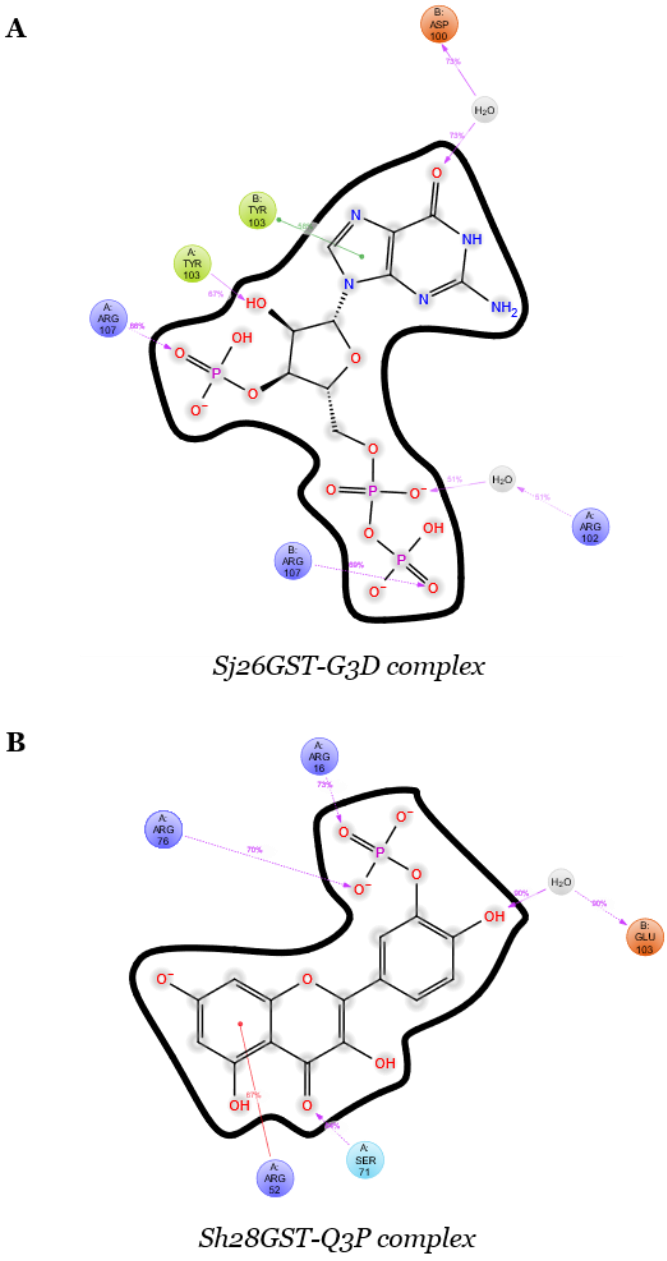


Figure 14.9. Interaction profiles from the working trajectory portions of each schistosome GST system in complex with its top-scoring candidate.

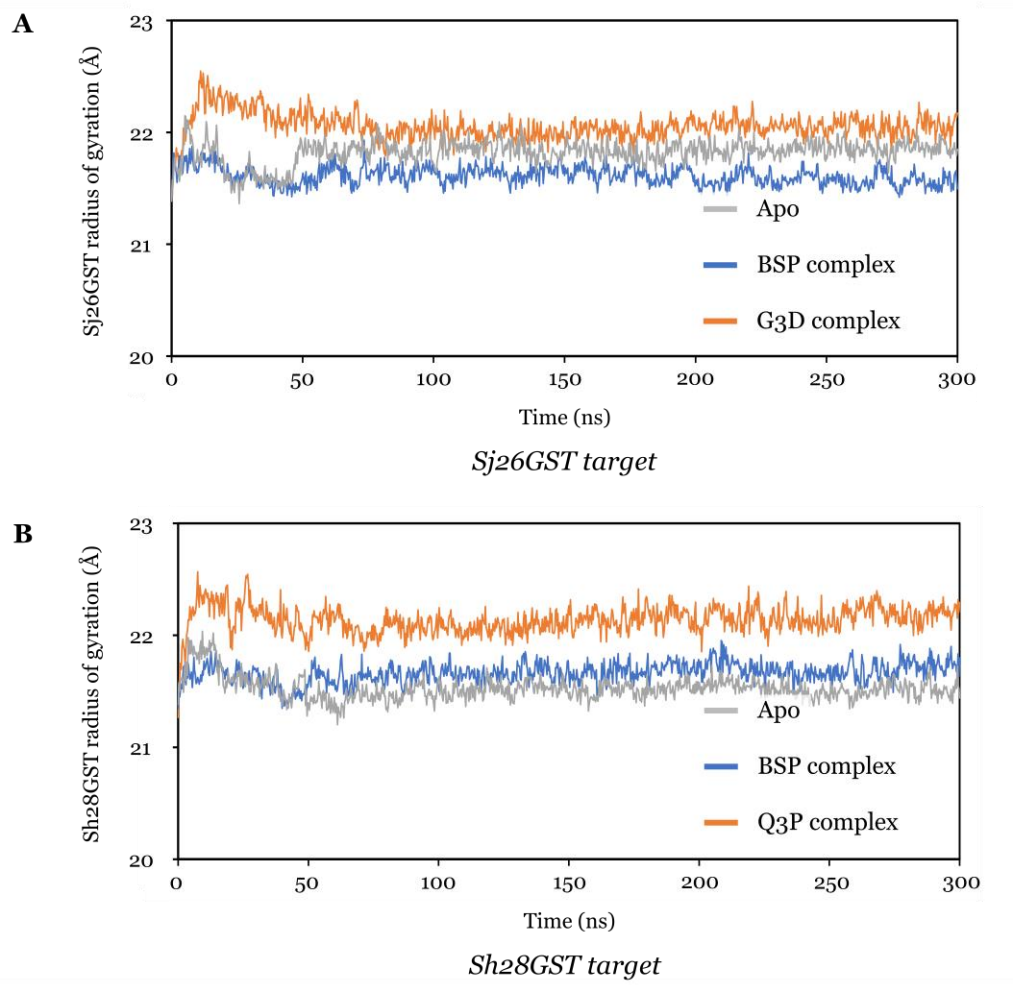


Figure 14.10. Radius of gyration showing the compactness of each target in response to binding with BSP and the respective top-scoring candidates.

15

Conclusion

We summarise the central findings and provide some perspectives on further research emanating from the pharmacophore-based screening.

MOLECULAR DYNAMICS ANALYSIS has revealed remarkably stable binding of BSP to the Sj26GST and Sh28GST drug targets, compared with human GST alpha (hGSTA), mu (hGSTM) and pi (hGSTP) controls. This validates the selective inhibitory behaviour of BSP upon the targets and establishes BSP as a benchmark ligand for the design of novel selective inhibitors. Pharmacophore screening with MD-derived models of the BSP complexes, combined with molecular docking has yielded G3D and Q3P as ligands bearing the greatest candidacy for targeting Sj26GST and Sh28GST. Comparative MD showed invariable binding characteristics of both candidates to BSP, suggesting their potential to exert similar empirical effects on the schistosome GST targets.

V

DRUG-RESOLVING POWER

This section is based on:

VALLI, A. and ACHILONU, I.

Investigating the drug-resolving capabilities of pharmacophore models derived from the 26- and 28-kDa *Schistosoma* spp. glutathione transferase enzymes

Journal of Computational Biology

Under review

16

Synopsis

In this chapter, we summarise the methodology used to assess the drug-resolving power of the two Schistosoma GST pharmacophore models. We also briefly report on the major findings observed during the study.

The structural and biochemical properties of a drug candidate influence drug-likeness. Hence, early prioritisation of candidates with favourable drug-like characteristics is important to mitigate failure rates during drug design. We have previously developed pharmacophore models of the 26- and 28-kDa *Schistosoma* spp. glutathione transferase (GST) enzymes, which are key targets in novel antischistosomal design. In this study, we evaluate the ability of these pharmacophores to retrieve candidates with desirable drug-likeness. Principal component analysis of both models revealed inappreciable variation between the drug-like properties of candidates from a generalised ligand database and known drug molecules. Subsequent clustering analysis indicated remarkable congruence in the data, suggesting the drug-like nature of candidates from the generalised database. Finally, candidates from both models displayed average clustered drug-likeness values within ideal ranges for orally bioavailable agents, altogether supporting the drug-resolving capabilities of the 26- and 28-kDa *Schistosoma* GST pharmacophores.

17

Methods

In this chapter, we detail various approaches involved in assessing the drug-resolving ability of the Schistosoma GST pharmacophore models. We begin by outlining the process used to construct screening libraries. Next, the pharmacophore screening protocol is documented. Methods used to compute the drug-like properties of shortlisted ligands are then described. Finally, we explain the statistical techniques used in the assessment, which includes principal component analysis and clustering.

17.1 Screening library construction

THREE DATABASES CONTAINING LIGAND STRUCTURE DATA FILES (.sdf) were used to construct separate screening libraries. First, we retrieved two-dimensional structure files of all approved, experimental and nutraceutical drugs present in DrugBank [110] version 5.1.9 (go.drugbank.com). These 8 946 ligands were processed using LigPrep [106] version 2019-3 (Schrödinger, Inc.) by computing ionisation and tautomeric states at pH 7.0, as well as by performing energy optimisation using the OPLS2005 force field to determine three-dimensional ligand geometries. The resulting 11 394 ligand structures were stored in a screening library. Then, 2 946 two-dimensional ligand structures present in the FDA-approved Drug Library Plus of MedChemExpress (www.medchemexpress.com) were retrieved and processed in the same manner, resulting in 6 561 ligands for screening. A subset of ZINC20 [114] (zinc20.docking.org) was finally retrieved, containing 7 112 310 lead-like ligands bearing LogP (lipophilicity) values less than 3.51, though certain experimental mole-

cules in violation of this criterion were included. Energy optimised three-dimensional structures of these ligands were retrieved using a batch file (.bat) script and stored in a third screening library.

17.2 Pharmacophore screening

Screening of the three ligand libraries was performed against two pre-developed pharmacophore models [38] using Phase [103] version 2019-3 (Schrödinger, Inc.). The algorithm was configured to match a minimum of four features for the five- and four-feature models of Sj26GST and Sh28GST respectively. Each library was screened independently for each of the two models. The vast size of the ZINC lead-like library necessitated the use of two screening experiments per model, each involving separate halves of the ligand library.

After collating the pharmacophore screening results of the ZINC library, each of the six output tables were arranged in order of Phase fitness score, measured on a scale of 0.00 to 3.00. The tables were pruned by removing duplicate entries and, save for the ZINC tables, eliminating ligands beyond 200 SMILES string characters. Removal of ligands with exceedingly large SMILES was necessary due to input restrictions of downstream analytical techniques.

Sampling of top-scoring ligands from the DrugBank and MedChemExpress screening experiments was constrained to the total number of entries present in the smallest output table between the two pharmacophore models. For the large outputs returned by the ZINC screening experiments, the top 10% of entries from the smallest table was chosen.

17.3 Computing drug-like properties

SMILES strings of top-scoring ligands sampled from the DrugBank and MedChemExpress data tables were submitted to the SwissADME [115] web server (www.swissadme.ch/index.php) for enumeration of physico- and biochemical parameters. This enables prediction of the absorption,

distribution, metabolism, excretion and toxicity (ADMET) profile of each ligand, which provides an indication of drug-likeness.

Entries in the ZINC data tables of both pharmacophore models lacked SMILES strings, therefore the SMILES property for the sampled ZINC ligands was regenerated prior to SwissADME submission. This was achieved by submitting the relevant ZINC identifiers to a search query on the ZINC web server and incorporating the corresponding SMILES data into the output tables. The SwissADME results were stored as individual comma-separated values (.csv) files.

17.4 Data table pre-processing

Each SwissADME output file consists of a table bearing numerical and categorical properties relating to the drug-likeness of ligands retrieved via pharmacophore screening. Pre-processing of the tables was performed to convert categorical data into numbers by assigning integer values corresponding to the relative degrees of a particular property. For example, the GI absorption property represented as either low, medium or high was converted to respective integers 1 to 3. Finally, the resulting numerical data tables were consolidated for each of the two pharmacophores, while maintaining a property designating the library from which each ligand originates.

17.5 Principal component analysis

The two consolidated numerical tables, each containing top-scoring entries from all three ligand libraries for a particular pharmacophore, were imported into RStudio [116] version 2023.03.0 for analysis with R. [117] version 4.2.0. Here, principal component analysis (PCA) was performed using a correlation matrix, followed by visualisation of the first three components with the ggfortify [118] package version 0.4.15.

17.6 Clustering analysis

R [117] version 4.2.0 was used to cluster the data points along the first two principal components as described below for each model. The NbClust [119] package version 3.0.1 was first used to establish the optimal number of clusters k, using the silhouette method. This approach uses a silhouette plot which displays the average silhouette width observed for various values of k, where the optimal k coincides with the maximal average silhouette width. Then, the partitioning around medoids algorithm from the cluster [120] package version 2.1.4 was used to perform grouping of the points into k clusters, followed by visualisation of the clusters along components one and two with the factoextra [121] package version 1.0.7. The silhouette coefficient for each clustered data point was thereafter plotted using the same package. This indicates the fit of a point within a particular cluster, providing a means to validate the outcome of the clustering analysis.

17.7 Central tendency and dispersion of properties

The numerical table containing the drug-like properties of candidates for each model was updated to include the cluster number assigned to each element. Means and standard deviations of a subset of properties crucial for oral bioavailability were then computed in R [117] version 4.2.0 for each cluster. Central tendencies of the clusters were compared to known drug-likeness standards, whereas standard deviation values were compared between the two pharmacophore models to determine dispersion of the drug-like data.

17.8 Structures and properties of top candidates

Two-dimensional structures of the top five ZINC candidates for each model were generated by submitting the relevant SMILES strings to the Marvin JS (Chemaxon) [122] web-based interface. The structures of these molecules were presented, along with the subset of drug-like properties for oral bioavailability.

18

Outcomes

In this chapter, we present and analyse the results garnered during the assessment of the drug-resolving ability of the Schistosoma spp. GST pharmacophore models. We first display the number of hits retrieved via the pharmacophore screening experiments. Next, we plot the drug-like property data of candidates from each ligand library along the first three principal components. Optimal clustering of the data points is then presented, followed by validation of the clustered data. Finally, we report the mean and deviation of the clustered data points and showcase the two-dimensional structures and properties of top candidates.

18.1 Pharmacophore screening

THE NUMBER OF ELEMENTS IN EACH SCREENING LIBRARY, as well as the number of candidates retrieved with pharmacophore screening and sampled for further analysis, are all displayed below in Table 18.1. ZINC represents the generalised ligand library, whereas the drug libraries are represented by DrugBank and MedChemExpress.

Table 18.1. Number of ligands in each screening library, along with the number of ligands shortlisted and selected for analysis.

Screening libraries			
Total number of elements	DrugBank	11 394	
	MedChemExpress	6 561	
	ZINC	7 112 310	
Pharmacophore screening			
		26-kDa model	28-kDa model
Unique elements retrieved	DrugBank	150	28
	MedChemExpress	307	31
	ZINC	63 146	1 982
Elements sampled	DrugBank	28	28
	MedChemExpress	31	31
	ZINC	198	198

18.2 Principal component analysis

The first, second and third principal components capture a respective 51.33%, 14.52% and 11.63% variance in the drug-like properties of hits retrieved with the 26-kDa model (Figure 18.1). The low spread of most points along principal component one, (Figure 18.1 A) in addition to moderate variance suggest minor disparity in the data. Considerable overlap of points from the MedChemExpress and DrugBank libraries is observed along this principal component, where fewer than ten points are distinguishable from the remaining elements. In comparison, the ZINC library contains 26 elements distal to this overlapping region, equating to 13.13% of mildly distinct candidates. Variation across principal component two (Figure 18.1 A and B) is inappreciable, except for a single outlier which also features in component one. The third principal component (Figure 18.1 B) of 11.63% variance shares characteristics with principal component one, with fewer distally located ZINC data points. In total,

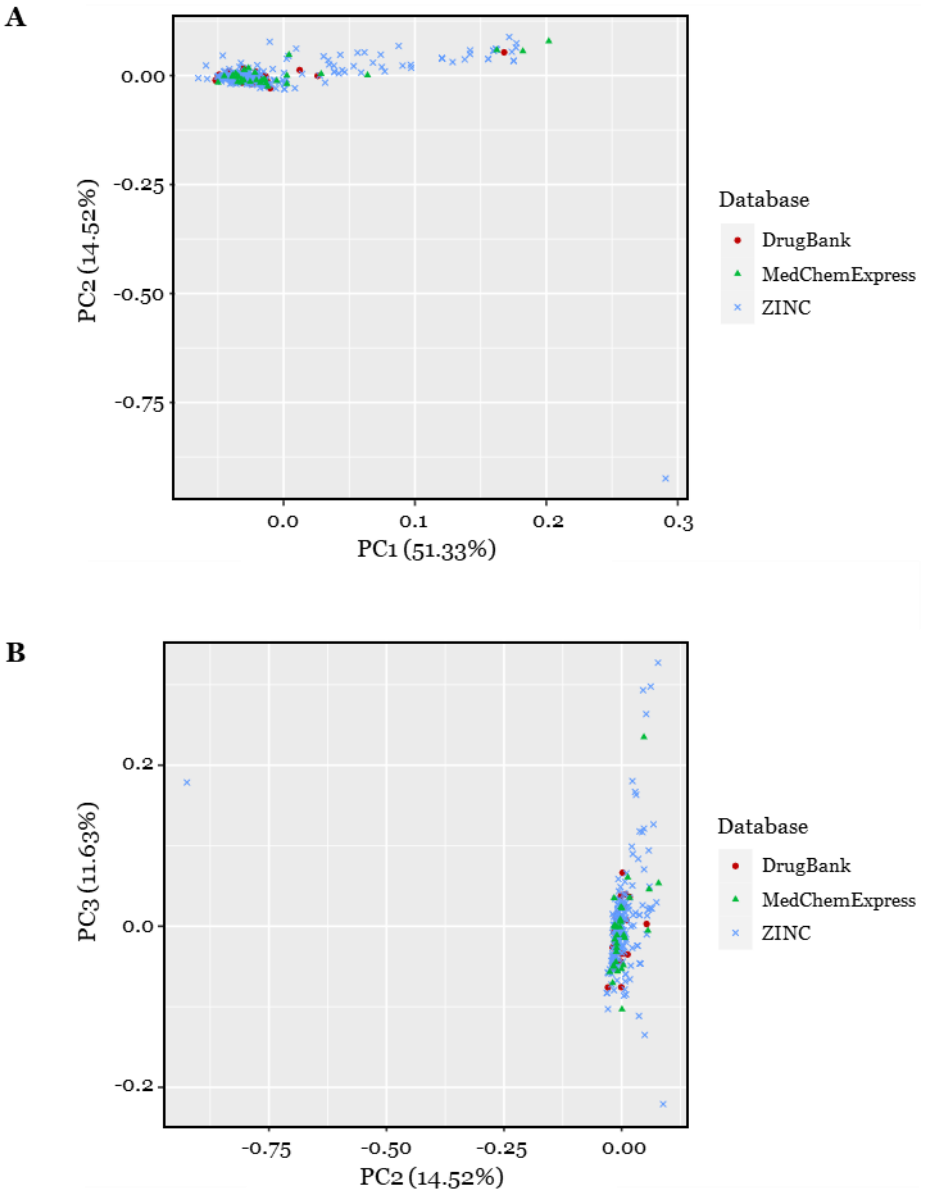


Figure 18.1. Principal component analysis of the drug-like properties from candidates retrieved via screening with the 26-kDa GST model.

the low spread and marginal variance which are observed across principal components one and three, combined with negligible variation across principal component two, indicate uniformity in the drug-like data of the 26-kDa model candidates.

Data belonging to candidates from the 28-kDa model demonstrate comparable variance values of 47.82%, 15.82% and 11.44% for each of the three principal components (Figure 18.2). The greater spread observed here correlates with the less stringent 28-kDa pharmacophore model containing four features compared to five of the 26-kDa model. Similarity in the data elements are evident, with overlapping of the three libraries along principal component one (Figure 18.2 A), where the ZINC library contains fewer than 25 distal elements. Furthermore, a small number of distinctly varied points belonging to DrugBank and MedChemExpress elements are observed along this component. In contrast with the 26-kDa model, the second principal component (Figure 18.2 A and B) also captures variation among the 25 distal elements of the ZINC library, with the remaining points demonstrating similar overlap to principal component one. Principal component three (Figure 18.2 B) exhibits the least variation among the data, with a couple of DrugBank elements being the only notable exceptions. Altogether, the marginal ±12.6% proportion of distal ZINC candidates, as well as a few distinct DrugBank and MedChemExpress elements, do not amount to any significant dissimilarity in the data. This is especially true when considering the modest variance captured by the principal components. Therefore, the drug-like data are sufficiently uniform amongst the candidates retrieved from the 28-kDa model.

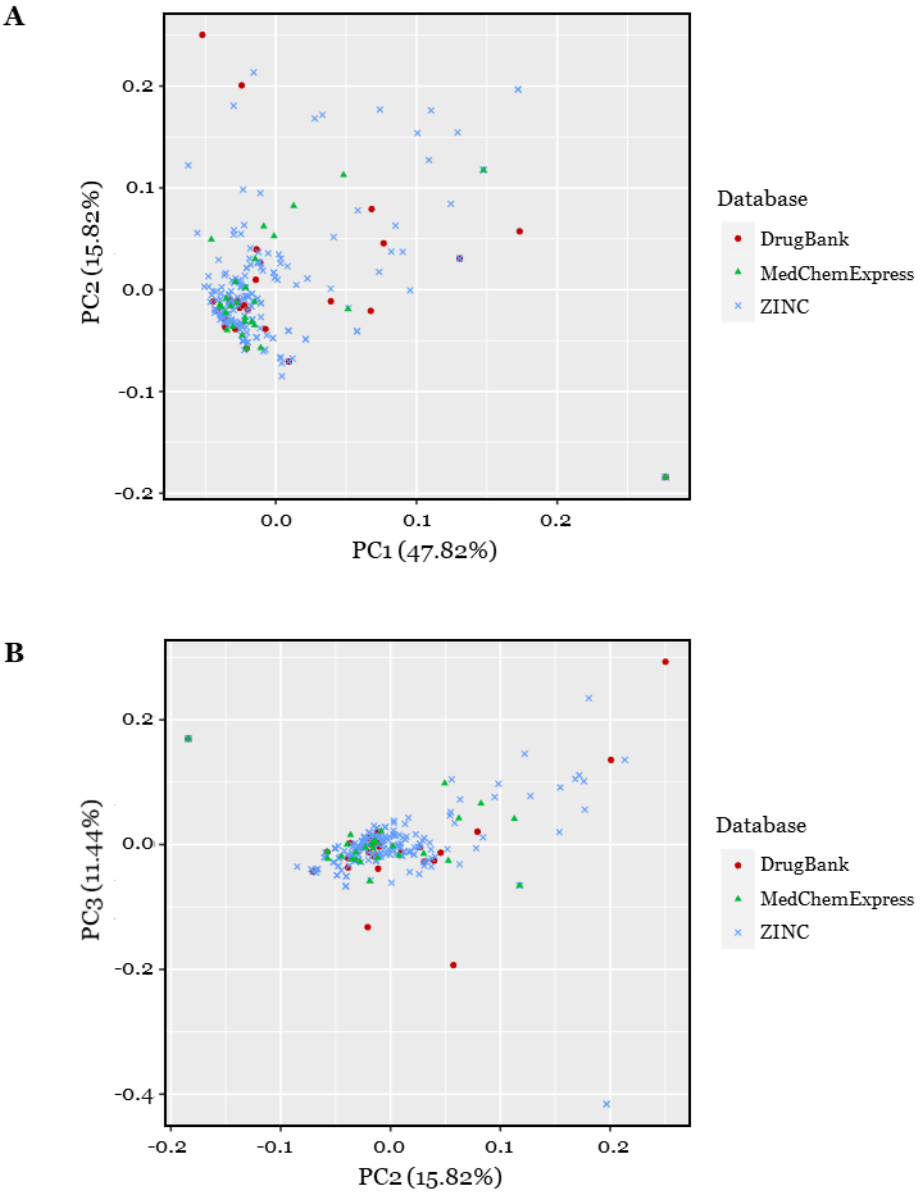


Figure 18.2. Principal component analysis of the drug-like properties from candidates retrieved via screening with the 28-kDa GST model.

18.3 Clustering analysis

A silhouette plot is used to determine the number of clusters k required for performing clustering analysis upon a dataset. The plots for both models (Figure 18.3) indicate an optimal value of $k = 2$ based upon the maximal average silhouette width, where the average silhouette width signifies the overall fit of all clustered data points from $0 - 1$. Hence, the partitioning around medoids clustering algorithm was configured to produce two clusters for each model. Due to the uniformity in the data, this suggests grouping of at least two out of the three screening libraries. Furthermore, upon considering the profound overlap in elements of MedChemExpress and DrugBank for both models, the likeness of the ZINC candidates to these drug libraries seems apparent.

Figure 18.4 shows cluster plots generated along the first two principal axes. The 26-kDa model (Figure 18.4 A) exhibits one large red cluster spanning all but one outlying ZINC data element, which forms a second cluster of little significance. The presence of a single cluster of low spread, along with only one vastly outlying point forming the second cluster, supports the remarkable congruence in the data. This reveals unequivocal similarity of the ZINC candidates to those from the drug libraries.

In comparison, the cluster plot belonging to the 28-kDa pharmacophore model (Figure 18.4 B) contains two discernible clusters, the primary red cluster of moderate spread displaying the majority of data elements, in addition to a blue cluster displaying 15 elements with relatively high spread. Assuming the worst case where all elements of the blue cluster belong to the ZINC library, the proportion of 15 out of 198 non-drug-like candidates amounts to a mere 7.58%. Thus, a retrieval rate of 92.42% for drug-like ZINC candidates prevails for the model.

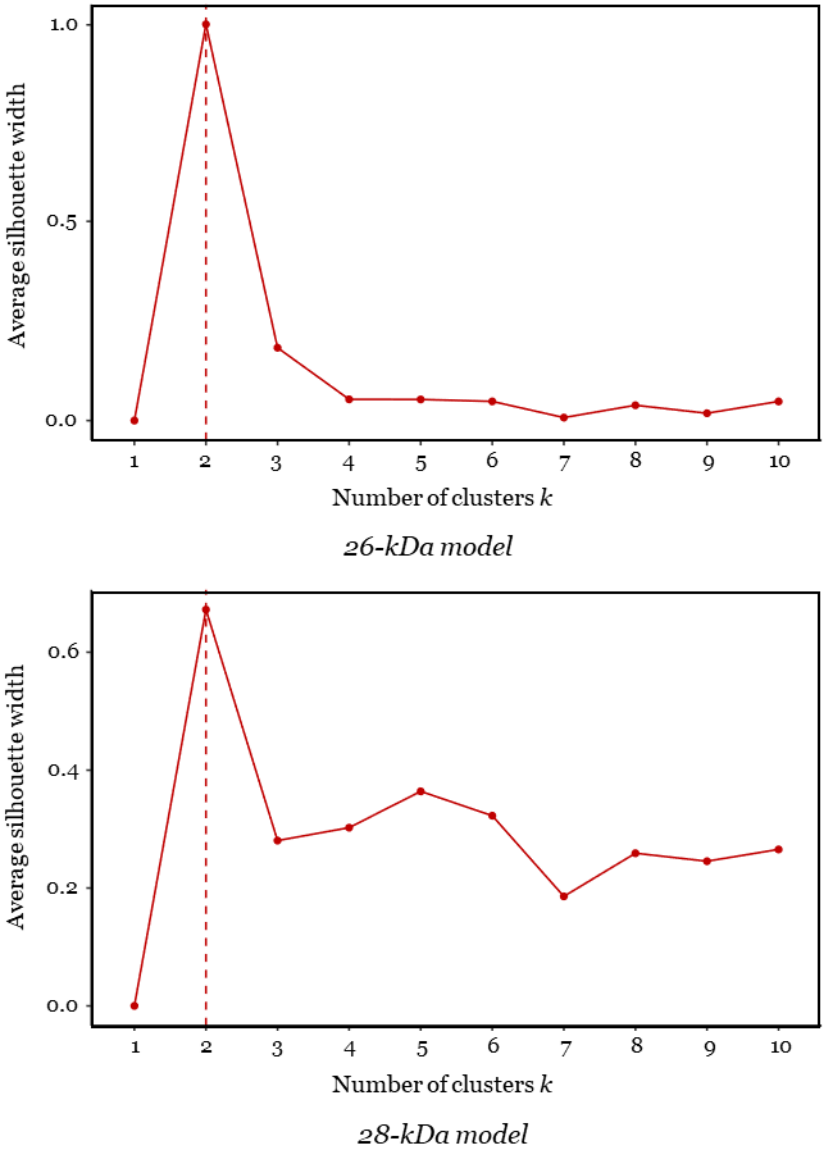


Figure 18.3. Silhouette plot showing the optimal number of clusters *k* in which to group candidates retrieved with each *Schistosoma* GST pharmacophore model

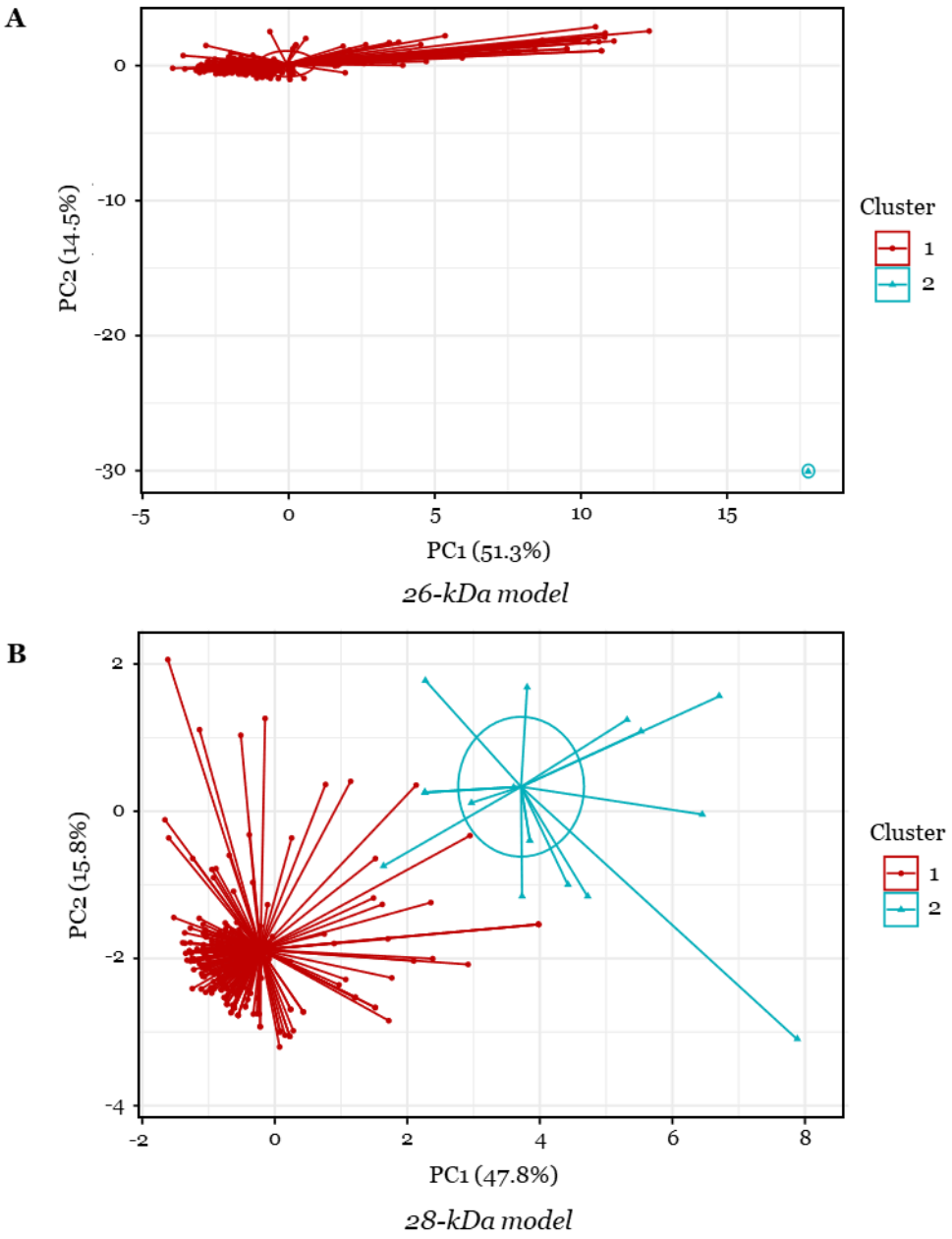


Figure 18.4. Cluster plot of $k = 2$ clusters showing the drug-like properties of candidates retrieved using two *Schistosoma* GST pharmacophore models.

18.4 Validation of clusters

The validity of the clustering analysis for each model was assessed by computing individual silhouette widths for all grouped data elements. Plotting these widths enables visualisation of the fit of each point compared to the average silhouette width, where negative values indicate placement in the wrong cluster. The exceptionally low spread of points in the primary cluster of the 26-kDa model yields individual silhouette widths which are invariably greater than 0.98. In addition, the silhouette width of the single point lying in the vastly outlying second cluster defaults to 1, due to no additional points in the same cluster for comparison. Thus, cluster validation of the 26-kDa model is not required due to the trivial solutions of 1 for all silhouette widths. The validation plot of the 28-kDa model shown in Figure 18.5 displays a favourable average silhouette width of 0.67 among all the points. Moreover, assessment of individual points shows that most elements in each cluster exceed the average fit, which suggests satisfactory clustering of the data elements.

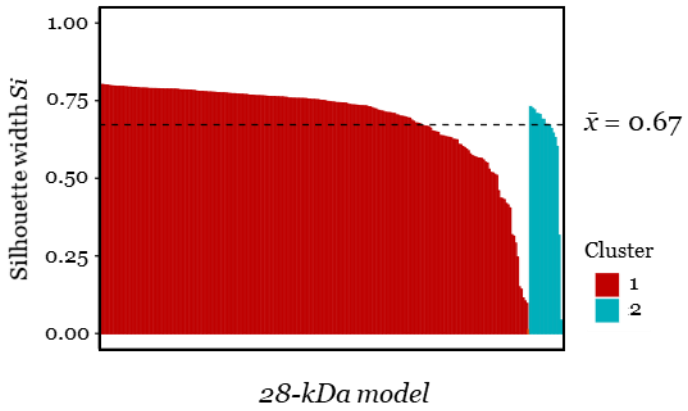


Figure 18.5. Cluster validation plot of data from the 28-kDa pharmacophore model of *Schistosoma* GST.

18.5 Central tendency and dispersion of properties

Table 18.2 displays the mean and standard deviation values of seven drug-like properties relating to oral bioavailability, computed across each cluster for the two models. The two primary clusters of each model, which contain a respective 99.49% and 92.52% of the candidates from the generalised ZINC library, demonstrate desirable values of all seven properties. Lower relative standard deviations of the 26-kDa model candidates are observed compared to those from the 28-kDa model. This corroborates the visual discrepancy in the spread of points along the principal components, confirming that the increased number of features of the 26-kDa model confers specificity to the screening experiments. When contrasted with the more compliant 28-kDa pharmacophore, the two models offer a strategic combination of sensitivity and specificity, presenting the potential to minimise false negative and false positive errors.

Secondary clusters of both models present candidates which violate the ranges for orally bioavailable drugs (Table 18.2). In particular, excessively large molecular weights are liable to compromise cell permeability, [123, 124] hence disrupting absorption and transport. Similarly, too many rotatable bonds results in high entropic penalties during membrane diffusion and target binding events, [123] which diminishes bioavailability. Despite the adequate solubility (LogS) metrics, the great number of H-bonding groups causes unsuitably high polarity of the candidates, which is signified by the outlying topological polar surface area (TPSA). This results in suboptimal lipophilicity values (LogP), which also impair membrane transport and distribution. [123-125]

Table 18.2. Mean and standard deviation of seven drug-like properties of candidates computed within each data cluster belonging to the two models.

		26-kDa model		28-kDa model	
Property	Oral bioavailability range	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Molecular weight (Da)	150 – 500 [58, 59]	330.23 ± 54.38	970.09	391.14 ± 92.21	678.29 ± 246.95
Rotatable bonds	< 9 [57]	4.42 ± 2.91	27.00	5.01 ± 4.37	16.50 ± 10.79
H-bond acceptors	< 10 [58]	5.33 ± 1.86	19.00	6.52 ± 3.58	12.63 ± 4.77
H-bond donors	< 5 [58]	1.67 ± 0.70	9.00	2.65 ± 1.84	5.75 ± 3.38
TPSA [1] (Å^2)	20 – 130 [60]	106.84 ± 23.46	351.68	110.21 ± 33.90	287.87 ± 111.87
LogP [2]	- 0.7 – 6 [61]	1.59 ± 0.72	-6.05	0.29 ± 2.01	-1.07 ± 2.79
LogS (ESOL method)	< 6: very soluble 6 – 10: soluble [49, 62]	-3.08 ± 0.80	3.81	-3.30 ± 1.61	-1.65 ± 3.06

[1] Topological Polar Surface Area

[2] LogP is determined using the consensus value arising from a total of five indices generated with various approximation methods.

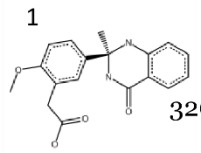
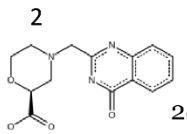
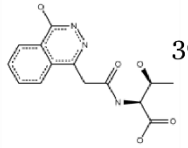
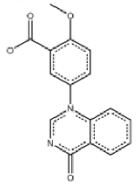
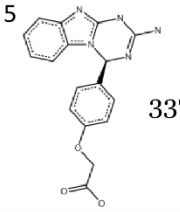
18.6 Structures and properties of top candidates

Table 18.3 and Table 18.4 show the two-dimensional chemical structures of the top five candidates retrieved using the 26- and 28-kDa models respectively, along with the seven properties which are critical determinants of oral bioavailability. From the top five candidates of the 26-kDa model (Table 18.3), a single violation of the rules outlined in Table 18.2 is observed, where the TPSA of candidate 3 marginally exceeds the upper limit of 130 Å^2 with a value of 132.64 Å^2. Even though this presents ramifications for the absorption and distribution of candidate 3 across

membranes, [126] the LogP appears to exist within the predefined limits for oral bioavailability. The remaining candidates from the 26-kDa model all exhibit satisfactory drug-likeness.

As shown in Table 18.4, three of the top five candidates of the 28-kDa model also violate the TPSA upper limit of 130 Å^2. Candidates 1 and 3 display minor violations with respective TPSA values of 132.13 and 138.38 Å^2, whereas candidate 2 is more severe with 157.51 Å^2. Here, desirable LogP values still prevail for all candidates. Substitution of polar moieties is, however, advised in order to improve permeability of these candidates for optimal absorption and transport. For instance, certain hydrogen bonding donor groups which are insignificant to the models can be eliminated. Ultimately, the presence of at most one violation of these properties among the top candidates of both models, all of which reside only marginally outside ideal oral bioavailability ranges, exemplifies the drug-resolving capabilities of the two pharmacophores.

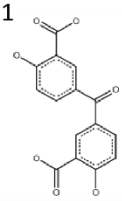
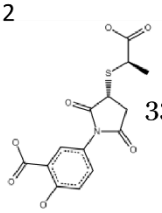
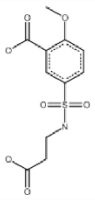
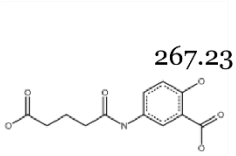
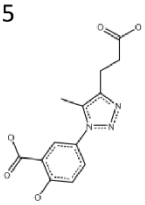
Table 18.3. Chemical structures and relevant drug-like properties of top candidates retrieved with the 26-kDa pharmacophore model.

Candidate	MW (Da)	Rot. bonds	H-bond acc.	H-bond don.	TPSA (Å²)	LogP	LogS	Oral bio-avail. viol.
1 	326.35	4	4	3	87.66	1.98	-3.36	None
2 	289.29	3	6	2	95.52	0.14	-0.36	None
3 	305.29	6	7	4	132.64	0.15	-1.48	TPSA
4 	296.28	3	5	1	81.42	1.94	-3.24	None
5 	337.33	4	5	3	114.76	1.16	-3.07	None

TPSA: Topological Polar Surface Area

LogP is determined using the consensus value arising from a total of five indices generated with various approximation methods.

Table 18.4. Chemical structures and relevant drug-like properties of top candidates retrieved with the 28-kDa pharmacophore model.

Candidate	MW (Da)	Rot. bonds	H-bond acc.	H-bond don.	TPSA (Å²)	LogP	LogS	Oral bio-avail. viol.
1 	302.24	4	7	4	132.13	1.56	-3.56	TPSA
2 	339.32	5	7	3	157.51	0.67	-2.56	TPSA
3 	303.29	7	8	3	138.38	0.3	-1.40	TPSA
4 	267.23	7	6	4	123.93	0.89	-2.16	None
5 	291.26	5	7	3	125.54	0.77	-2.50	None

TPSA: Topological Polar Surface Area

LogP is determined using the consensus value arising from a total of five indices generated with various approximation methods.

19

Conclusion

In this chapter, we summarise the central observations gathered while assessing the drug-resolving capabilities of the pharmacophore models.

In this study, we determine the drug-resolving capabilities of two *Schistosoma* GST pharmacophore models by evaluating the relative drug-likeness of candidates retrieved from the generalised ZINC library, compared to candidates from the drug libraries MedChemExpress and Drugbank. Principal component analysis does not reveal any appreciable variation in the data of either model, highlighting the relatedness of the ZINC candidates to those from the drug libraries. Clustering analysis performed on both models, followed by validation, shows robust grouping of over 92% of ZINC candidates with candidates from the drug libraries. Finally, mean values of drug-like properties computed across each primary cluster exist within the ranges for oral bioavailability in both models, demonstrating satisfactory ability of these *Schistosoma* GST pharmacophores to retrieve candidates with desirable drug-likeness.

VI

EPILOGUE

20

Research summary

We collate the premise, approach and major highlights of this research.

THE NON-SUBSTRATE BINDING SITE occupying the dimer interface domain of *Schistosoma* GST is a compelling target for novel antischistosomal drugs. In order to develop selective pharmacophore models, we leverage the discriminative binding capabilities of BSP to this target. It is theorised that the use of these pharmacophores will expedite screening protocols and enable the retrieval of quality drug-like candidates.

Comprehensive characterisation of the GST dimer interface indicates substantial differences in binding character between human and *Schistosoma* GST variants. Particularly, the binding of chemicals bearing phenol and sulfonate moieties to schistosome GST is afforded by the presence of appropriate stabilising residues which are absent in human GST. In total, we have established a suitable chemical space from which potential drug candidates may be identified.

Ligand interaction profiles derived from the dynamic behaviour of BSP in the target binding site were employed to develop five- and four-feature pharmacophores of the respective 26- and 28-kDa *Schistosoma* GST variants. These models both contained the characteristic aromatic- and sulfonate-stabilising residues previously observed. Candidates retrieved with these pharmacophores displayed *in silico* behaviour typical of the BSP benchmark, suggesting their potential to emulate the empirical effects of BSP upon the schistosome GST drug target.

Finally, rigorous statistical approaches have revealed the drug-likeness of shortlisted candidates retrieved using both models, demonstrating their ability to sample highly specific chemical space subsets. This ensures a satisfactory reliability index for the models, which is a crucial measure in identifying candidates likely to endure the drug development pipeline.

21

Future perspectives

In this final chapter, we provide some implications of this work, as well as document a few recommendations regarding future research efforts.

The use of benchmark molecules in the construction, refinement and evaluation of screening models such as pharmacophores, has proven to be a beneficial endeavour. Our body of research has moreover validated the capabilities of such models to strategically bias chemical sampling of small-molecule species in order to identify quality, selective ligands for drug targets. Hence, the routine use of pharmacophores is recommended, particularly when rapid screening of large datasets is required, or when protocols demand highly specialised sampling, such as when there is increased propensity for cross-reactivity, or otherwise general safety concerns about non-target binding.

Implementation of the statistics workflow performed in this study offers an accurate and robust method to assess various aspects of candidates retrieved with screening models. These assessments provide indications of the nature of the models, such as the survival of shortlisted candidates in the scope of drug development, or the likelihood of them interacting with a particular drug target.

In vitro screening assays will provide the necessary empirical evidence regarding the behaviours of shortlisted candidates retrieved with these pharmacophores. These assays should ideally be combined with iterative directed chemical refinement in order to improve upon any observed

empirical activity. If desirable results are obtained, the candidates can finally be admitted into subsequent steps of the drug design process.

References

[1] B. Gryseels, K. Polman, J. Clerinx, and L. Kestens, Human schistosomiasis. *The Lancet,* vol. 368, no. 9541, pp. 1106-1118, 2006.

[2] D. G. Colley, A. L. Bustinduy, W. E. Secor, and C. H. King, Human schistosomiasis. *The Lancet,* vol. 383, no. 9936, pp. 2253-2264, 2014.

[3] A. K. Deol, F. M. Fleming, B. Calvo-Urbano, M. Walker, V. Bucumi, I. Gnandou, E. M. Tukahebwa, S. Jemu, U. J. Mwingira, and A. Alkohlani, Schistosomiasis—assessing progress toward the 2020 and 2025 global goals. *New England Journal of Medicine,* vol. 381, no. 26, pp. 2519-2528, 2019.

[4] M. L. Nelwan, Schistosomiasis: life cycle, diagnosis, and control. *Current Therapeutic Research,* vol. 91, pp. 5-9, 2019.

[5] A. Akhaddar and A. Akhaddar, Other parasitic infections of the central nervous system. *Atlas of infections in neurosurgery and spinal surgery,* pp. 311-316, 2017.

[6] B. D. Furch, J. R. Koethe, V. Kayamba, D. C. Heimburger, and P. Kelly, Interactions of schistosoma and HIV in sub-Saharan Africa: a systematic review. *The American journal of tropical medicine and hygiene,* vol. 102, no. 4, p. 711, 2020.

[7] S. S. W. Sakari, A. K. Mbugua, and G. M. Mkoji, Prevalence of soil-transmitted helminthiases and schistosomiasis in preschool age children in Mwea division, Kirinyaga south district, Kirinyaga county, and their potential effect on physical growth. *Journal of Tropical Medicine,* vol. 2017, 2017.

[8] R. Gönnert and P. Andrews, Praziquantel, a new broad-spectrum antischistosomal agent. *Zeitschrift für Parasitenkunde,* vol. 52, pp. 129-150, 1977.

[9] C. H. King and A. A. F. Mahmoud, Drugs five years later: praziquantel. *Annals of Internal Medicine,* vol. 110, no. 4, pp. 290-296, 1989.

[10] L. S. Iarotski and A. Davis, The schistosomiasis problem in the world: results of a WHO questionnaire survey. *Bulletin of the World Health Organization,* vol. 59, no. 1, p. 115, 1981.

[11] A. Danso-Appiah and S. J. De Vlas, Interpreting low praziquantel cure rates of Schistosoma mansoni infections in Senegal. *Trends in parasitology,* vol. 18, no. 3, pp. 125-129, 2002.

[12] B. Gryseels, A. Mbaye, S. J. De Vlas, F. F. Stelma, F. Guisse, L. Van Lieshout, D. Faye, M. Diop, A. Ly, and L. A. Tchuem-Tchuenté, Are poor responses to praziquantel for the treatment of Schistosoma mansoni infections in Senegal due to resistance? An overview of the evidence. *Tropical Medicine & International Health,* vol. 6, no. 11, pp. 864-873, 2001.

[13] M. Ismail, S. Botros, A. Metwally, S. William, A. Farghally, L.-F. Tao, T. A. Day, and J. L. Bennett, Resistance to praziquantel: direct evidence from Schistosoma mansoni isolated from Egyptian villagers. *The American journal of tropical medicine and hygiene,* vol. 60, no. 6, pp. 932-935, 1999.

[14] M. J. Doenhoff, J. R. Kusel, G. C. Coles, and D. Cioli, Resistance of Schistosoma mansoni to praziquantel: is there a problem? *Transactions of the Royal Society of Tropical Medicine and Hygiene,* vol. 96, no. 5, pp. 465-469, 2002.

[15] P. G. Fallon, Schistosome resistance to praziquantel. *Drug Resistance Updates,* vol. 1, no. 4, pp. 236-241, 1998.

[16] M. S. Wilson, M. M. Mentink-Kane, J. T. Pesce, T. R. Ramalingam, R. Thompson, and T. A. Wynn, Immunopathology of schistosomiasis. *Immunology and cell biology,* vol. 85, no. 2, pp. 148-154, 2007.

[17] M. J. Doenhoff, D. Cioli, and J. Utzinger, Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Current Opinion in Infectious Diseases,* vol. 21, no. 6, pp. 659-67, 2008.

[18] D. G. Colley and W. E. Secor, Immunology of human schistosomiasis. *Parasite Immunology,* vol. 36, no. 8, pp. 347-357, 2014.

[19] R. S. Barsoum, Urinary schistosomiasis. *Journal of advanced research,* vol. 4, no. 5, pp. 453-459, 2013.

[20] X.-N. Zhou, L.-Y. Wang, M.-G. Chen, X.-H. Wu, Q.-W. Jiang, X.-Y. Chen, J. Zheng, and U. Jürg, The public health significance and control of schistosomiasis in China—then and now. *Acta tropica,* vol. 96, no. 2-3, pp. 97-105, 2005.

[21] C.-l. Tang, H.-h. Zhou, Y.-w. Zhu, J. Huang, and G.-b. Wang, Glutathione S-transferase influences the fecundity of Schistosoma japonicum. *Acta Tropica,* vol. 191, pp. 8-12, 2019.

[22] W. U. Tiu, K. M. Davern, M. D. Wright, P. G. Board, and G. F. Mitchell, Molecular and serological characteristics of the glutathione S-transferases of Schistosoma japonicum and Schistosoma mansoni. *Parasite Immunology,* vol. 10, no. 6, pp. 693-706, 1988.

[23] B. K. Kubata, M. Duszenko, K. S. Martin, and Y. Urade, Molecular basis for prostaglandin production in hosts and parasites. *Trends in parasitology,* vol. 23, no. 7, pp. 325-331, 2007.

[24] M. A. McTigue, D. R. Williams, and J. A. Tainer, Crystal Structures of a Schistosomal Drug and Vaccine Target: Glutathione S-Transferase from Schistosoma japonica and its Complex with the Leading Antischistomal Drug Praziquantel. ed: Elsevier, 1995.

[25] M. E. M. Saeed, S. Krishna, H. J. Greten, P. G. Kremsner, and T. Efferth, Antischistosomal activity of artemisinin derivatives in vivo and in patients. *Pharmacological Research,* vol. 110, pp. 216-226, 2016.

[26] X. Shu-Hua, Y. Ji-Qing, G. Hui-Fang, M. Jin-Yan, J. Pei-Ying, J. Chollet, M. Tanner, and J. Utzinger, Schistosoma japonicum: effect of artemether on glutathione S-transferase and superoxide dismutase. *Experimental parasitology,* vol. 102, no. 1, pp. 38-45, 2002.

[27] B. O. Akumadu, R. Pandian, J. Olfsen, R. Worth, M. Thulo, T. Mentor, S. Fanucchi, Y. Sayed, H. W. Dirr, and I. Achilonu, Molecular basis of inhibition of Schistosoma japonicum glutathione transferase by ellagic acid: Insights into biophysical and structural studies. *Molecular and Biochemical Parasitology,* vol. 240, p. 111319, 2020.

[28] K. Pooe, R. Worth, E. A. Iwuchukwu, H. W. Dirr, and I. Achilonu, An empirical and theoretical description of Schistosoma japonicum glutathione transferase inhibition by bromosulfophthalein and indanyloxyacetic acid 94. *Journal of Molecular Structure,* vol. 1223, p. 128892, 2021.

[29] D. C. Young, *Computational drug design: a guide for computational and medicinal chemists*. John Wiley & Sons, 2009.

[30] S. Wang, Y. Zhong, and E. Wang, An integrated GIS platform architecture for spatiotemporal big data. *Future Generation Computer Systems,* vol. 94, pp. 160-172, 2019.

[31] S. McIntosh-Smith, J. Price, R. B. Sessions, and A. A. Ibarra, High performance in silico virtual drug screening on many-core processors. *The international journal of high performance computing applications,* vol. 29, no. 2, pp. 119-134, 2015.

[32] Z. Fan, W. Chen, V. Vierimaa, and A. Harju, Efficient molecular dynamics simulations with many-body potentials on graphics processing units. *Computer Physics Communications,* vol. 218, pp. 10-16, 2017.

[33] D. Schaller, D. Šribar, T. Noonan, L. Deng, T. N. Nguyen, S. Pach, D. Machalz, M. Bermudez, and G. Wolber, Next generation 3D pharmacophore modeling. *Wiley Interdisciplinary Reviews: Computational Molecular Science,* vol. 10, no. 4, p. e1468, 2020.

[34] N. S. Pagadala, K. Syed, and J. Tuszynski, Software for molecular docking: a review. *Biophysical Reviews,* vol. 9, no. 2, pp. 91-102, 2017.

[35] L. Di, E. H. Kerns, and G. T. Carter, Drug-like property concepts in pharmaceutical design. *Current Pharmaceutical Design,* vol. 15, no. 19, pp. 2184-94, 2009.

[36] I. Yusof and M. D. Segall, Considering the impact drug-like properties have on the chance of success. *Drug Discovery Today,* vol. 18, no. 13, pp. 659-666, 2013.

[37] A. Valli and I. Achilonu, Comparative structural analysis of the human and Schistosoma glutathione transferase dimer interface using selective binding of bromosulfophthalein. *Proteins: Structure, Function, and Bioinformatics,* 2022.

[38] A. Valli and I. Achilonu, Molecular dynamics-derived pharmacophores of Schistosoma glutathione transferase in complex with bromosulfophthalein: Screening and analysis of potential inhibitors. *Journal of Molecular Graphics and Modelling,* vol. 122, p. 108457, 2023.

[39] A. Valli and I. Achilonu, Investigating the drug-resolving capabilities of pharmacophore models derived from the 26- and 28-kDa Schistosoma spp. glutathione transferase enzymes. *Molecular Informatics,* 2023.

[40] K. Hinsen, The molecular modeling toolkit: a new approach to molecular simulations. *Journal of Computational Chemistry,* vol. 21, no. 2, pp. 79-85, 2000.

[41] N. L. Allinger, X. Zhou, and J. Bergsma, Molecular mechanics parameters. *Journal of Molecular Structure: THEOCHEM,* vol. 312, no. 1, pp. 69-83, 1994.

[42] A. Onufriev, Implicit solvent models in molecular dynamics simulations: A brief overview. *Annual Reports in Computational Chemistry,* vol. 4, pp. 125-137, 2008.

[43] C. Tan, Y.-H. Tan, and R. Luo, Implicit nonpolar solvent models. *The Journal of Physical Chemistry B,* vol. 111, no. 42, pp. 12263-12274, 2007.

[44] J. Zhang, H. Zhang, T. Wu, Q. Wang, and D. van der Spoel, Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents.

Journal of Chemical Theory and Computation, vol. 13, no. 3, pp. 1034-1043, 2017.

[45] J. Kleinjung and F. Fraternali, Design and application of implicit solvent models in biomolecular simulations. *Current Opinion in Structural Biology,* vol. 25, pp. 126-134, 2014.

[46] N. A. Baker, Biomolecular applications of Poisson–Boltzmann methods. *Reviews in computational chemistry,* vol. 21, pp. 349-379, 2005.

[47] V. Tsui and D. A. Case, Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers: Original Research on Biomolecules,* vol. 56, no. 4, pp. 275-291, 2000.

[48] B. Gautam, Energy Minimization. in *Homology Molecular Modeling-Perspectives and Applications*: IntechOpen, 2020.

[49] K. Roy, S. Kar, and R. N. Das, *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*. Academic press, 2015.

[50] A. Sethi, K. Joshi, K. Sasikala, and M. Alvala, Molecular docking in modern drug discovery: Principles and recent applications. *Drug discovery and development-new advances,* vol. 2, pp. 1-21, 2019.

[51] Schrödinger, *Induced Fit Docking*. New York: Schrödinger Press, 2015.

[52] P. Śledź and A. Caflisch, Protein structure-based drug design: from docking to molecular dynamics. *Current Opinion in Structural Biology,* vol. 48, pp. 93-102, 2018.

[53] T. Hansson, C. Oostenbrink, and W. van Gunsteren, Molecular dynamics simulations. *Current Opinion in Structural Biology,* vol. 12, no. 2, pp. 190-196, 2002.

[54] E. Persch, O. Dumele, and F. Diederich, Molecular recognition in chemical and biological systems. *Angewandte Chemie International Edition,* vol. 54, no. 11, pp. 3290-3327, 2015.

[55] S.-Y. Yang, Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today,* vol. 15, no. 11-12, pp. 444-450, 2010.

[56] X. Qing, X. Y. Lee, J. De Raeymaecker, J. R. H. Tame, K. Y. J. Zhang, M. De Maeyer, and A. Voet, Pharmacophore modeling: advances, limitations, and current utility in drug discovery. *Journal of Receptor, Ligand and Channel Research,* vol. 7, pp. 81-92, 2014.

[57] A. Kutlushina, A. Khakimova, T. Madzhidov, and P. Polishchuk, Ligand-based pharmacophore modeling using novel 3D pharmacophore signatures. *Molecules,* vol. 23, no. 12, p. 3094, 2018.

[58] G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, Computational methods in drug discovery. *Pharmacological Reviews,* vol. 66, no. 1, pp. 334-395, 2014.

[59] S. A. Khedkar, A. K. Malde, E. C. Coutinho, and S. Srivastava, Pharmacophore modeling in drug discovery and development: an overview. *Medicinal Chemistry,* vol. 3, no. 2, pp. 187-197, 2007.

[60] M. J. de Groot, A. A. Alex, and B. C. Jones, Development of a combined protein and pharmacophore model for cytochrome P450 2C9. *Journal of Medicinal Chemistry,* vol. 45, no. 10, pp. 1983-1993, 2002.

[61] R. E. Salmas, M. Stein, M. Yurtsever, P. Seeman, I. Erol, M. Mestanoglu, and S. Durdagi, The signaling pathway of dopamine D2 receptor (D2R) activation using normal mode analysis (NMA) and the construction of pharmacophore models for D2R ligands. *Journal of Biomolecular Structure and Dynamics,* vol. 35, no. 9, pp. 2040-2048, 2017.

[62] K. Poptodorov, T. Luu, and R. D. Hoffmann, Pharmacophore model generation software tools. *Pharmacophores and Pharmacophore Searches,* vol. 32, pp. 15-47, 2006.

[63] T. Mavromoustakos, S. Durdagi, C. Koukoulitsa, M. Simcic, M. G Papadopoulos, M. Hodoscek, and S. Golic Grdadolnik, Strategies in the rational drug design. *Current Medicinal Chemistry,* vol. 18, no. 17, pp. 2517-2530, 2011.

[64] A. Sethi, K. Joshi, K. Sasikala, and M. Alvala, Drug Discovery and Development—New Advances. ed: IntechOpen London, UK:, 2019.

[65] K. Strømgaard, P. Krogsgaard-Larsen, and U. Madsen, *Textbook of drug design and discovery*. CRC press, 2017.

[66] J. A. Frearson, P. G. Wyatt, I. H. Gilbert, and A. H. Fairlamb, Target assessment for antiparasitic drug discovery. *Trends in parasitology,* vol. 23, no. 12, pp. 589-595, 2007.

[67] J. H. McKerrow, Update on drug development targeting parasite cysteine proteases. *PLoS Neglected Tropical Diseases,* vol. 12, no. 8, p. e0005850, 2018.

[68] D. J Timson, Metabolic enzymes of helminth parasites: potential as drug targets. *Current Protein and Peptide Science,* vol. 17, no. 3, pp. 280-295, 2016.

[69] F. Ferraro, I. Corvo, L. Bergalli, A. Ilarraz, M. Cabrera, J. Gil, B. M. Susuki, C. R. Caffrey, D. J. Timson, X. Robert, C. Guillon, T. Freire, and G. Álvarez, Novel and selective inactivators of Triosephosphate isomerase with anti-trematode activity. *Scientific Reports,* vol. 10, no. 1, p. 2587, 2020.

[70] X. Qing, X. Y. Lee, J. De Raeymaecker, J. R. H. Tame, K. Y. J. Zhang, M. De Maeyer, and A. R. D. Voet, Pharmacophore modeling: advances, limitations, and current utility in drug discovery. *Journal of Receptor, Ligand and Channel Research,* vol. 7, pp. 81-92, 2014.

[71] H. Sun, G. Tawa, and A. Wallqvist, Classification of scaffold-hopping approaches. *Drug Discovery Today,* vol. 17, no. 7-8, pp. 310-324, 2012.

[72] H. Zhao, Scaffold selection and scaffold hopping in lead generation: a medicinal chemistry perspective. *Drug Discovery Today,* vol. 12, no. 3-4, pp. 149-155, 2007.

[73] J. Boström, K. Berggren, T. Elebring, P. J. Greasley, and M. Wilstermann, Scaffold hopping, synthesis and structure–activity relationships of 5, 6-diaryl-pyrazine-2-amide derivatives: a novel series of CB1 receptor antagonists. *Bioorganic and Medicinal Chemistry,* vol. 15, no. 12, pp. 4077-4084, 2007.

[74] I. T. Jolliffe, *Principal component analysis for special types of data*. Springer, 2002.

[75] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, An overview of principal component analysis. *Journal of Signal and Information Processing,* vol. 4, no. 3B, p. 173, 2013.

[76] R. Garcia-Dias, S. Vieira, W. H. L. Pinaya, and A. Mechelli, Clustering analysis. in *machine learning*: Elsevier, pp. 227-247, 2020.

[77] J. Swarndeep Saket and S. Pandya, An overview of partitioning algorithms in clustering techniques. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),* vol. 5, no. 6, pp. 1943-1946, 2016.

[78] X. Li, Glutathione and Glutathione-S-Transferase in Detoxification Mechanisms. *General, Applied and Systems Toxicology,* 2009.

[79] E. Croom, Metabolism of xenobiotics of human environments. *Progress in Molecular Biology and Translational Science,* vol. 112, pp. 31-88, 2012.

[80] P. Matoušková, I. Vokřál, J. Lamka, and L. Skálová, The role of xenobiotic-metabolizing enzymes in anthelmintic deactivation and resistance in helminths. *Trends in parasitology,* vol. 32, no. 6, pp. 481-491, 2016.

[81] J. Barrett, Helminth detoxification mechanisms. *Journal of Helminthology,* vol. 71, no. 2, pp. 85-90, 1997.

[82] D. Sheehan, G. Meade, V. M. Foley, and C. A. Dowd, Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical journal,* vol. 360, no. 1, pp. 1-16, 2001.

[83] P. M. Brophy and J. Barrett, Glutathione transferase in helminths. *Parasitology,* vol. 100, no. 2, pp. 345-349, 1990.

[84] D. W. Nebert and V. Vasiliou, Analysis of the glutathione S-transferase (GST) gene family. *Human genomics,* vol. 1, no. 6, pp. 1-5, 2004.

[85] T. M. Buetler and D. L. Eaton, Glutathione S-transferases: Amino acid sequence comparison, classification and phylogenetic relationship. *Journal of Environmental Science & Health Part C,* vol. 10, no. 2, pp. 181-203, 1992.

[86] K. A. Johnson, F. Angelucci, A. Bellelli, M. Hervé, J. Fontaine, D. Tsernoglou, A. Capron, F. Trottein, and M. Brunori, Crystal structure of the 28 kDa glutathione S-transferase from Schistosoma haematobium. *Biochemistry,* vol. 42, no. 34, pp. 10084-10094, 2003.

[87] B. Wu and D. Dong, Human cytosolic glutathione transferases: structure, function, and drug discovery. *Trends in Pharmacological Sciences,* vol. 33, no. 12, pp. 656-668, 2012.

[88] A. E. Salinas and M. G. Wong, Glutathione S-transferases-a review. *Current Medicinal Chemistry,* vol. 6, no. 4, pp. 279-310, 1999.

[89] F. Angelucci, P. Baiocco, M. Brunori, L. Gourlay, V. Morea, and A. Bellelli, Insights into the catalytic mechanism of glutathione S-transferase: the lesson from Schistosoma haematobium. *Structure,* vol. 13, no. 9, pp. 1241-1246, 2005.

[90] J. L. Milhon, R. L. Thiboldeaux, K. Glowac, and J. W. Tracy, Schistosoma japonicum GSH S-transferase Sj26 is not the

molecular target of praziquantel action. *Exp Parasitol,* vol. 87, no. 3, pp. 268-74, 1997.

[91] M. Platis, D. Vlachakis, A. I. Foudah, M. M. Muharram, M. H. Alqarni, A. C. Papageorgiou, and N. E. Labrou, The Interaction of Schistosoma Japonicum Glutathione Transferase with Cibacron Blue 3GA and its Fragments. *Medicinal Chemistry (Shariqah, United Arab Emirates),* vol. 17, no. 4, pp. 332-343, 2021.

[92] S. M. Rosenthal and E. C. White, Clinical application of the bromsulphalein test for hepatic function. *Journal of the American Medical Association,* vol. 84, no. 15, pp. 1112-1114, 1925.

[93] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. D. Caves, Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics,* vol. 22, no. 21, pp. 2695-2696, 2006.

[94] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry,* vol. 25, no. 13, pp. 1605-1612, 2004.

[95] J. Eberhardt, D. Santos-Martins, A. Tillack, and S. Forli, AutoDock Vina 1.2. 0: new docking methods, expanded force field, and Python bindings. 2021.

[96] O. Trott and A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry,* vol. 31, no. 2, pp. 455-461, 2010.

[97] M. F. Adasme, K. L. Linnemann, S. N. Bolz, F. Kaiser, S. Salentin, V. J. Haupt, and M. Schroeder, PLIP 2021: Expanding the scope of the protein-ligand interaction profiler to DNA and RNA. *Nucleic Acids Research,* 2021.

[98] L. Schrödinger, "The PyMOL Molecular Graphics System, Version 2.3.2," November, 2015.

[99] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, The protein data bank. *Nucleic acids research,* vol. 28, no. 1, pp. 235-242, 2000.

[100] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden, NCBI BLAST: a better web interface. *Nucleic acids research,* vol. 36, no. suppl_2, pp. W5-W9, 2008.

[101] R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research,* vol. 32, no. 5, pp. 1792-1797, 2004.

[102] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research,* vol. 34, no. suppl_1, pp. D668-D672, 2006.

[103] *Maestro, Schrödinger Inc.*, New York, 2021.

[104] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design,* vol. 27, pp. 221-234, 2013.

[105] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research,* vol. 49, no. D1, pp. D1388-D1395, 2020.

[106] *LigPrep, Schrödinger Inc.*, New York, 2021.

[107] R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, and D. T. Mainz, Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *Journal of Medicinal Chemistry,* vol. 49, no. 21, pp. 6177-6196, 2006.

[108] E. C. Kevin J. Bowers, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, "Proceedings of the 2006 ACM/IEEE conference on Supercomputing," Tampa, Florida, 2006: Association for Computing Machinery.

[109] S. L. Dixon, A. M. Smondyrev, E. H. Knoll, S. N. Rao, D. E. Shaw, and R. A. Friesner, PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *Journal of Computer-Aided Molecular Design,* vol. 20, no. 10, pp. 647-671, 2006.

[110] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, and M. Wilson, DrugBank

5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research,* vol. 46, no. D1, pp. D1074-d1082, 2018.

[111] D. M. Metwally, E. M. Al-Olayan, M. Alanazi, S. B. Alzahrany, and A. Semlali, Antischistosomal and anti-inflammatory activity of garlic and allicin compared with that of praziquantel in vivo. *BMC Complementary and Alternative Medicine,* vol. 18, no. 1, pp. 1-11, 2018.

[112] S. S. El-Hawary, K. F. Taha, F. N. Kirillos, A. A. Dahab, A. A. El-Mahis, and S. H. El-Sayed, Complementary effect of Capparis spinosa L. and silymarin with/without praziquantel on mice experimentally infected with Schistosoma mansoni. *Helminthologia,* vol. 55, no. 1, p. 21, 2018.

[113] J. M. Muema, M. A. Obonyo, S. N. Njeru, and J. K. Mwatha, Antischistosomal effects of selected methanolic plant extracts in swiss albino mice infected with Schistosoma mansoni. *European Journal of Medicinal Plants,* vol. 9, no. 1, pp. 1-11, 2015.

[114] J. J. Irwin and B. K. Shoichet, ZINC– a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling,* vol. 45, no. 1, pp. 177-182, 2005.

[115] A. Daina, O. Michielin, and V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports,* vol. 7, no. 1, p. 42717, 2017.

[116] J. Allaire, RStudio: integrated development environment for R. *Boston, MA,* vol. 770, no. 394, pp. 165-171, 2012.

[117] R. Ihaka and R. Gentleman, R: a language for data analysis and graphics. *Journal of computational and graphical statistics,* vol. 5, no. 3, pp. 299-314, 1996.

[118] Y. Tang, M. Horikoshi, and W. Li, ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages. *R J.,* vol. 8, p. 474, 2016.

[119] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software,* vol. 61, pp. 1-36, 2014.

[120] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, Cluster: cluster analysis basics and extensions. 2012.

[121] A. Kassambara and F. Mundt, Package 'factoextra'. *Extract and visualize the results of multivariate data analyses,* vol. 76, no. 2, 2017.

[122] B. Cherinka, B. H. Andrews, J. Sánchez-Gallego, J. Brownstein, M. Argudo-Fernández, M. Blanton, K. Bundy, A. Jones, K. Masters, and D. R. Law, Marvin: A tool kit for streamlined access and visualization of the SDSS-IV MaNGA data set. *The Astronomical Journal,* vol. 158, no. 2, p. 74, 2019.

[123] D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *Journal of Medicinal Chemistry,* vol. 45, no. 12, pp. 2615-2623, 2002.

[124] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews,* vol. 64, pp. 4-17, 2012.

[125] A. K. Ghose, V. N. Viswanadhan, and J. J. Wendoloski, A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *Journal of Combinatorial Chemistry,* vol. 1, no. 1, pp. 55-68, 1999.

[126] P. Ertl, B. Rohde, and P. Selzer, Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *Journal of Medicinal Chemistry,* vol. 43, no. 20, pp. 3714-3717, 2000.