University of Hradec Králové

Faculty of Science

Department of Chemistry

Computational Analysis of Human Glycogen Synthase Kinase 3β Structure and Binding Sites

Bachelor thesis

Author: Michaela Melíková
Study programme: B1407 Chemistry
Field of study: 1407R005 Chemistry

Supervisor: Mgr. et Mgr. Rafael Doležal, Ph.D.

Hradec Králové May 2016

Declaration

I hereby declare that I have writen this thesis on my own and that I have duly referenced all the sources which I had used.

In Pardubice, May 2016

Michaela Melíková

Acknowledgement

In the first place, I would like to express my deep thanks to my supervisor, Mgr. et Mgr. Rafael Doležal, Ph.D., for all his helpfulness, patience and the kind approach.

I also thank Assoc. Prof. Ing. Ondřej Krejcar, Ph.D. for having granted me entrance to the study room J16 at the Faculty of Informatics and Management where I could carry out some of the computational experiments.

The development of computational methods was done at Meta VO. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

The scale-up and finalization of computations for two targets and 12 ligands with 10 iterations were performed on the supercomputers Anselm and Salomon within the scope of the research project “Computerized drug design of novel glycogen synthase kinase 3 beta inhibitors for Alzheimer’s disease treatment enhanced by structure-based virtual screening and QM/MM calculations” at Biomedical Research Center in Faculty Hospital Hradec Králové. This work was supported by The Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project „IT4Innovations National Supercomputing Center – LM2015070“.

Anotace

MELÍKOVÁ, M. *Počítačová analýza struktury a vazebných míst lidské glykogen synthasy kinasy 3β*. Hradec Králové 2016. Bakalářská práce na Přírodovědecké fakultě Univerzity Hradec Králové. Vedoucí bakalářské práce Rafael Doležal. 47 s.

Bakalářská práce se zaměřuje na počítačovou analýzu struktury lidské glykogen synthasy kinasy 3β (GSK-3β) s cílem lokalizovat a charakterizovat potenciální vazebná místa pro nové modulátory její aktivity. Tento enzym je kromě biosyntézy glykogenu zahrnut v řadě patologických procesů, jako je rakovina, diabetes mellitus II a Alzheimerova nemoc. Rozbor topochemických vlastností vazebných míst GSK-3β pomocí specializovaného softwaru poskytne poznatky pro racionální návrh nových selektivních léčiv na rakovinu či Alzheimerovu nemoc metodami strukturního virtuálního screeningu.

Klíčová slova

glykogen synthasa kinasa 3β, počítačová analýza, vazebná místa, dokování

Annotation

MELÍKOVÁ, M. *Computational Analysis of Human Glycogen Synthase Kinase 3β Structure and Binding Sites*. Hradec Králové, 2016. Bachelor thesis at Faculty of Science University of Hradec Králové. Thesis supervisor Rafael Doležal. 47 p.

The bachelor thesis is focused on computational analysis of the structure of human glycogen synthase kinase 3β (GSK-3β) with the aim to localize and characterize potential binding sites for novel modulators of its activity. Apart from glycogen biosynthesis, the enzyme is involved in an array of pathological processes, like cancer, diabetes mellitus type II and Alzheimer's disease. The study of topochemical characteristics of GSK-3β binding sites using specialized software will yield findings for a rational design of the new selective molecules for cancer or Alzheimer's disease by the means of virtual screening.

Keywords

glycogen synthase kinase 3β, computational analysis, binding sites, docking

Table of Contents

List of Abbreviations

2D, 3D, *n*D	two-dimensional, three-dimensional, *n*-dimensional
ASA	accessible surface area
CADD	computer-aided drug design
CPU	central processor unit
EC_{50}	half-maximal effective concentration
GSK-3β	glycogen synthase kinase-3β
GYS	glycogen synthase
IC_{50}	half-maximal inhibitory concentration
MD	molecular dynamics
MM	molecular mechanics
NMR	nuclear magnetic resonance
PDB	Protein Data Bank (free online database of 3D protein structures)
RMSD	root mean square deviation
RNA	ribonucleic acid
SASA	solvent accessible surface area
SES	solvent excluded surface

List of Figures

Introduction

Drug development is a lengthy and costly process with a low chance of stumbling upon relevant substances. Until recently, it had relied on assiduity and, in great part, on chance. Thanks to the introduction of computers as tools able to quickly handle immense amounts of data, it was possible to transfer some of the tasks in the early stages of drug design to the newly arisen field of computational chemistry, substantially reducing the time spent on screening while increasing hit rate by discarding entities with predicted poor properties. Although *in silico* simulations cannot replace experimental steps in drug development, they prove helpful in finding and focusing on substances with higher likelihood of exerting biological activity.

Human glycogen synthase kinase-3β (GSK-3β) is a serine-threonine specific protein kinase discovered in 1980 [1]. Although 35 years have passed, the many mechanisms it participates in haven't been fully understood. The enzyme has been shown to be virtually omnipresent in human cells and to interact with a number of other proteins. Beside its role in glucose metabolism, it is also involved in several signaling transduction pathways where its excessive activity appears to be related to neurodegenerative and inflammatory diseases and certain types of cancer. GSK-3β is responsible for hyperphosphorylation of tau proteins which, together with amyloid beta plaques, accumulates in the brain tissue of patients with Alzheimer's disease [2]. This having been said, it comes naturally that finding pharmaceutically usable inhibitors of its activity arouses major interest in pharmaceutical companies. To date, more than 40,000 compounds have been reported to inhibit GSK-3β but there are insufficient data concerning their accurate IC_{50} values, their binding sites and binding modes, their selectivity and drug-likeness [3].

The objective of this thesis was to study the surface of GSK-3β in order to find pockets which could further be investigated as potential allosteric binding sites for selective down-regulation of GSK-3β's activity with pharmaceuticals. In order to cover the range of already determined structures of GSK-3β and a library of molecules to be tested for affinity to potential cavities, *in silico* approach was chosen.

The search was conducted employing a tool called fpocket, optimized for finding putative binding pockets protein structures. The structures used in the study had been determined by the methods of X-ray analysis of GSK-3β co-crystallized (a) with an inhibitor and (b) with a substrate. Fpocket implements Voronoi tessellation and α-spheres for quick search of pockets in the protein surface. Descriptors of found pockets were determined using a bundled algorithm named dpocket.

A secondary aim of the thesis was to validate any conserved pockets reported by fpocket. This was done by the means of a virtual screening in which several already known GSK-3β inhibitors were docked into the pockets.

1 Theoretical Part

1.1 Rational Design of Protein-Targeting Drugs

1.1.1 Proteins Make Desirable Targets

Throughout the history of mankind, drugs used to be discovered by observing the effects of substances of natural origin, mainly ones derived from medicinal plants, such as extracts or brews. However, little was known about the underlying principles of their action. It was as late as in the 19th and 20th centuries that little by little, scientific methods started seeping into pharmacology. Thanks to advances in all branches of chemistry, structures and functions of biomacromolecules as well as mechanisms of their interactions with other molecules had become partly elucidated. In the 1960s, functions of enzymes, receptors and ion channels were described, marking a turning point in drug design [4].

Various theories elucidating physico-chemical mechanisms of interaction between biomacromolecules and drugs have been proposed, providing a scaffold for a more complex estimation of therapeutic action of pharmaceuticals. The current widely accepted hypothesis states that by modulating the activity of a molecule involved in the pathways of a disease, therapeutic effects can be achieved. Endogenous molecules whose activity can be regulated by mediators, including drugs, are called biological targets. They are usually present in the biosystem only in low concentrations but play a vital role within the organism. Biological targets can be whole bacteria or viruses, cytoplasmic membranes, but in most cases, they are either nucleic acids (genes) or proteins (enzymes, receptors, ligand or voltage-gated ion channels, but also structural proteins). Those biological targets which are known or predicted to bind to a drug are called druggable, and as different proteins have different structures and chemical properties, they can potentially be targeted with high selectivity [5].

When on the lookout for a drug, it is advisable that, if possible, one first get acquainted with the structure, localization and function of its intended target. The structure of proteins can be described at four distinct levels of organization. Primary structure is the sequence of a protein's building blocks – residues of amino acids. There are twenty common proteinogenic amino acids and several rare, non-standard ones. Primary structure is essential for bioinformatics because it allows for comparison of chemical composition, similarity searches etc. Secondary structure is the local spatial orientation of the amino acid backbone and its stabilization by non-covalent bonds. The positioning of backbone atoms often shows a degree of regularity, with the two prominent motives being α-helices and β-pleated sheets. Tertiary structure, or fold, is the overall three-dimensional (3D) arrangement of the whole protein molecule, including side-chains. In its natural environment in the biological system, the molecule is held together by non-covalent bonds and may further be cross-linked by covalent disulfide bonds. The process of protein folding does not depend solely on its primary structure; it requires proper conditions and the involvement of auxiliary elements present within the cell environment. Quaternary level only applies to proteins consisting of more than one subunit, each represented by a single molecule with its own tertiary structure. It describes the mutual connection and orientation of the subunits [6].

Proteins play a vast number of roles in the organism. They may, among others, have structural, transport, immune, signaling, regulatory, or catalytic functions. The latter group is the most numerous and its members are called enzymes. Each cell contains literally thousands of enzymes catalyzing its chemical reactions, allowing for higher reaction rates, milder reaction conditions, greater specificity and fine-tuned control [7]. This is why proteins have increasingly been attracting the interest of pharmaceutical researchers, especially in the case of so-called molecular diseases, *i.e.* diseases with a simple etiology which can be traced back to point mutations or protein misfolding [8].

1.1.2 A Primer to Enzymes

Enzyme is a biomolecule with catalytic function highly specific for the molecules it processes, called substrates. A vast majority of enzymes are proteins, although RNA molecules demonstrating catalytic activity are also known. Enzymes are divided into classes according to the reactions they catalyze. Main classes recognized are oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases. Enzymes with absolute specificity are capable of catalyzing the reaction of one substrate only. Most enzymes, however, are group-specific, meaning that they catalyze reactions of a whole group of related substrates, even though for each, they reach different reaction speeds. In order to function properly, enzymes may require the presence of non-protein structures –cofactors. A covalently bound cofactor is called a prosthetic group while non-covalently bound ones are known as coenzymes. The protein part of an enzyme is called apoenzyme, the complete enzyme a holoenzyme [7].

A hypothesis has been put forward which claims that the sequence of proteins, among them enzymes, unambiguously predetermines their 3D structure which, in turn, determines their function. This so-called sequence-structure-function paradigm would mean that each protein is a deterministic static, rigid structure with a predictable function [9]. Although there certainly are tight relations of sequence-to-structure and of structure-to-function, the subject turns out to be more complicated. In the last decades, the concept has been challenged by suggestions that also dynamics are vital for protein function. These claims are supported, for example, by observing conformational changes in proteins and by resolving structures of proteins lacking persistent structure, or intrinsically disordered proteins (IDPs) [10]. How does this pertain to the functioning of enzymes?

The pioneers of enzyme research formulated first theories at the close of 19th century. It had been known since 1850s that enzymes act as biological catalysts and that they somehow bind to their substrates. Scientists scrutinized enzyme action and wondered how come that they always so minutely maintain specificity. By the study of invertase (an enzyme catalyzing hydrolysis of sucrose to fructose and glucose), Emil Fischer came up with the idea that to preserve chirality, enzymes must themselves be chiral [11]. According to him, the reaction would run within a relatively small restricted space of a given shape within the protein, into which the substrate molecule fits just like a key fits in the lock. Nowadays we call the aforementioned space the enzyme's active site. The lock-and-key complementarity model was improved by the notion of an enzyme-substrate complex [12], which would be a transitional state occurring between the binding of the substrate to the active site and its conversion into the product.

The theory was further worked upon by Koshland who stated that the enzyme and the substrate aren't perfectly complementary (Figure 1). Rather, the bond between the ligand and the active center is weak in the beginning; it is the proximity of ligand that induces a gradual conformational change of the protein, strengthening the bond [13]. The induced-fit model was the first to acknowledge certain dynamics of protein function and provided grounds for later models. The fluctuation fit theory by Brunó Ferenc Straub suggests that the enzyme molecule fluctuates in time – some of its forms are able to bind a ligand whereas others aren't but may bind different ligands [14]. The experimental techniques at the time were not sufficient for determining which of the theories was correct, though, so it wasn't until recently that further progress has been made.

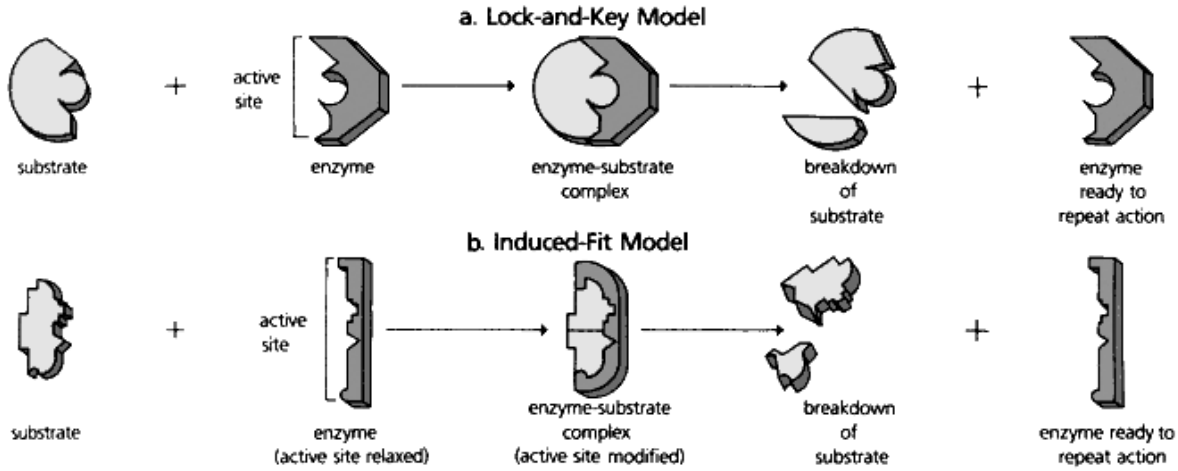


Figure 1: Comparison of lock and key model and induced-fit model.
Taken from http://ibbiology.wikifoundry.com/page/Describe+the+induced-fit+model.

It is now generally accepted that each enzyme has at least one domain – a set of dozens to hundreds independently folded conserved amino acid residues performing a specific function. Domains may contain an active site formed by the side chains of amino acid residues with specific physico-chemical properties; the residues can be located far from each other in the terms of their position in the sequence but thanks to the enzyme's 3D conformation they get topologically close. Active site is the region where the binding process and the catalytic reaction take place. Domains typically have the characteristics of small independent globular protein units, and in larger, multi-domain proteins they may form lobes, although at times these are so intertwined that it is difficult to tell them apart. During conformational changes, the domains themselves don't undergo substantial changes in shape, rather they are hinged by less rigid parts that move relatively freely [6].

Since the emergence of new simulation and experimental methods, a synthesis of both Koshland's and Straub's theories has been accepted: it regards each of the two concepts as extremes while usually the two co-exist and contribute with different ratios. Conformation ensemble model [15] further generalizes the idea to say that the enzyme occupies a delicately balanced set of possible states in the conformational space among which it fluctuates. The

changes in states are related to shifts in probability of population distribution caused by a multitude of stabilizing and destabilizing interactions [16].

The latest views border closely on the idea of enzyme regulation. Roughly put, there are three ways of regulating enzymes' activity: first, localization of individual enzymes within the cell and their competition for available substrate. Second, controlling the synthesis and degradation of enzyme molecules. And third, modulation of their activity through changes in their structure or by the action of effectors. From pharmaceutical point of view, the last principle is the most enticing, as by the use of different agents, a range of effects can be achieved.

Agents binding to an enzyme may do so reversibly by creating non-covalent bonds, or irreversibly *via* covalent bonds. Effectors increasing the activity of an enzyme are called activators, those decreasing it, inhibitors; it is interesting to note that virtually all drugs act as inhibitors. Four types of reversible inhibition can occur: competitive inhibitors are molecules which resemble the enzyme's substrate so closely that they bind to the active site. However, no product is created after the formation of the enzyme-substrate complex and the active site becomes blocked. Because the effect occurs directly in the active site, it is sometimes called orthosteric inhibition; it can be overcome by increasing substrate concentration. Uncompetitive inhibitors are those which react with the enzyme-substrate complex; by binding to it, they decrease the maximum speed of the reaction. Mixed inhibition is the combination of competitive and uncompetitive inhibition – the inhibitor can bind to the enzyme whether or not the substrate has already bound to it. The binding of the substrate and the binding of the inhibitor affect each other. In non-competitive inhibition, the inhibitor binds to the enzyme regardless of whether the substrate has been bound (Figure 2). Non-competitive inhibitors always bind at locations other than the active (orthosteric) site, called allosteric sites [7]. Inhibition type can be determined by the properties of the line resulting from plotting the ratio of reciprocal values of reaction speed (y-axis) as a function of time (x-axis). This method is based on Michaelis-Menten equation in a reciprocal form (Equation 1) [17].

$$\frac{1}{V} = \frac{K_M}{V_{\max}} \cdot \frac{1}{[S]} + \frac{1}{V_{\max}}$$

Equation 1: Reciprocal Michaelis-Menten equation. V – speed of reaction; V_{max} – maximum speed of reaction; KM – Michaelis-Menten constant; [S] – concentration of substrate.

Molecules bound at the allosteric sites usually provoke a conformational change in the protein; they can act as allosteric activators as well as allosteric inhibitors (inhibitors binding at the allosteric site don't necessarily act as non-competitive ones but may well fall into either category of competitive, non-competitive, or uncompetitive inhibitors). Compared to orthosteric sites, allosteric sites offer a greater spatial diversity because there is no need for them to bind to a specific substrate which would necessitate their high evolutionary conservation. Therefore, allosteric sites may differ among an enzyme's subtypes sharing the same substrate, allowing for a higher selectivity of drugs targeting them. It has been suggested that all proteins are potentially allosteric and that their allosteric behavior only needs to be awoken by the use of a properly chosen effector [15].

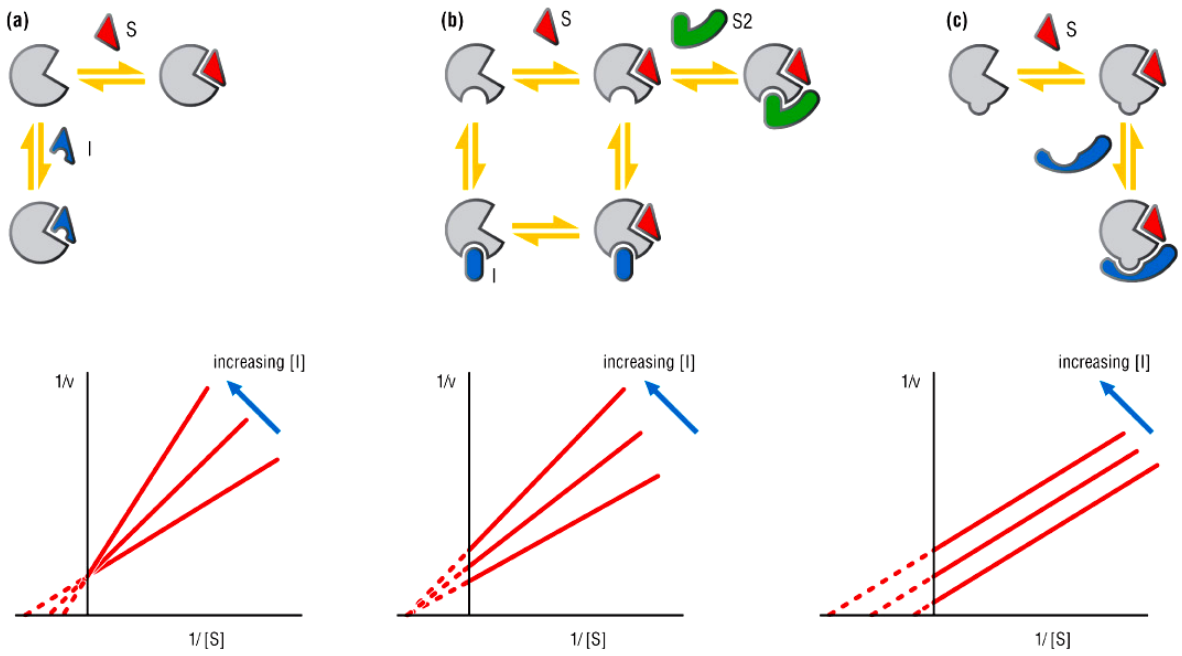


Figure 2: Illustration of (a) competitive, (b) non-competitive and (c) uncompetitive inhibition mechanisms with their corresponding and Lineweaver-Burk plots. *v* – reaction speed; [I] – concentration of inhibitor; [S] – concentration of substrate.
Taken from https://biochemshariestar.wordpress.com/2013/04/14/enzymes.

1.1.3 Evolution of Drug Design Methods

Natural medicines have accompanied man since the dawn of civilization. The findings of pollen from healing herbs at burial sites dating to Paleolithic, 60,000 B.C., suggest that beneficial effects of plants were known to Neanderthals. What led ancient humans to first use natural substances as crude drugs can now hardly be ascertained but what we do know is that the millennia of accumulated experience have bequeathed to future generations vast amounts of medicinal lore. Although the teachings of ancient medicine were often coupled with mysticism, spiritualism and philosophy, they dealt, sometimes in great detail, with procuring, treating and storage of crude drugs, complemented with instructions on indications and dosage. These copious data gave rise to traditional medicines, notably Chinese traditional medicine.

After the advent of script, much of extant information interspersed across cultures was written down, constituting a body of historical literature collectively related to as *māteria medica* (Latin for "medicinal substance"). As more and more information on medicinal substances was being gathered in Europe, the scope of pharmacology started to expand beyond the mere knowledge of how to use medicinal plants, incorporating domains of other, then emerging, natural sciences such as chemistry, biology and physics. In the 18th century, the extended medical information containing additional data on purification, potency, safety etc. became known by the name of pharmacognosy, from the Greek words φάρμακον (farmakon) – drug and γνῶσις (gnosis) – knowledge. Yet for that whole period, the mechanisms of drug action remained entirely unknown. Among concepts attempting to elucidate them, the doctrine of signatures is worth mentioning. It was based on the assertion that plants affected those body parts to which they were morphologically similar. Pharmacognosy along with its

botany-centered point of view was regarded as one of the pillars of pharmacy and persisted as a teaching subject of medical doctors and pharmacists until the beginning of 20th century. In conjunction with other research branches, it is making a return today [18].

Thanks to the influence it exercised on modern natural sciences, pharmacognosy contributed to the emergence of classical pharmacology – a science which relies on the assessment of activity of substances on models of disease, represented usually by cell cultures or animals. After an effect is confirmed during primary screening, classical pharmacology takes up a study of the substance's mechanism of action. Detection of any positive hits is followed by the search for affected molecular structures and may result in the discovery of relevant biological targets. Through an iterative process, methods of combinatorial chemistry are used to synthesize new ligands based on the structure of the hit compounds with the aim to improve their properties. Resulting derivatives are then again subjected to high-throughput screening and sifted to find suitable leads. Lead compounds are further chemically modified in the attempt to meet strict criteria of low toxicity, high selectivity, good bioavailability etc. imposed by European Medicines Agency (Europe), Food and Drug Administration (USA) or their counterparts elsewhere. [4]. The duration of a typical preclinical phase is 5.5 years on average; its costs amount to over 180 million USD. If the meticulous process of lead optimization happens to yield any promising drug candidates, clinical phases ensue. The three stages of clinical trials take about 7-10 years and bring expenditures of another 200 million USD. In case of final market approval of the drug, a long-term fourth clinical testing phase may be required which may last up to decades and require even more funds [19]. Upon summing up the above numbers, we arrive to long 10 to 15 years and a whopping 400 million USD needed for the process between finding a biologically active compound and introducing it to the market. That naturally only holds if everything goes well. As this way of designing drugs relies heavily on chance in a long trial-and-error marathon, it is quite obvious that high failure rates combined with nine to ten-figure costs of development render classical drug design a risky business.

Serendipity has played a noteworthy role in drug discovery. In the world of science, the word "serendipity" does not refer to a pure stroke of luck, it also has a connotation of making a discovery thanks to a keen mind. In drug discovery, this was the case of chloral hydrate (1869), paracetamol (1887), penicillin (1928), LSD (1943), warfarin (1948), chlorpromazine (1950), cisplatin (1965) and sildenafil (1996), to name a few famous ones [20, 21]. Unfortunately, crazes for new drugs sometimes led to sinister consequences which could have been avoided had due caution been taken. An example is the Elixir Sulfanilamide scandal from 1937, during which 105 patients died because of a poisoning with ethylene glycol solution of sulfanilamide. The consequent public outcry caused an elevated interest in drug regulation and resulted in the prompt implementation of Food, Drug and Cosmetic Act in 1938 [22]. Another wave of updates in drug regulations began in 1970s in the wake of the thalidomide affair which caused tens of thousands babies to be born with severe malformations [23]. The impact of the rigorous regulations is, of course, vastly beneficial in that it brings an increase in patients' safety but as a trade-off it ties, so to say, Fortuna's hands.

An approach envisioning a suppression of the dependence of drug discovery on happenstance is rational drug design, also called reverse pharmacology. As opposed to classical pharmacology, it starts from a genomic or a proteomic study and, usually with the help of a

bioinformatic search, estimates which biological structure is likely to be involved in the disease. After inferring the relevant structure, it designs molecules with the aim to find ones which will have a biological effect on the target. The rest of the process, *i.e.* clinical phases, is analogous to traditional pharmacology. Main difference between rational design and classical pharmacology lies in the reduction of time (about 2 years for rational design vs. 5 years for classical pharmacology) and costs of the pre-clinical phase [4].

To-date, rational design has brought numerous drugs to the market. First attempts to systematically design a drug began on the verge of 19th and 20th century, when purposeful chemical modifications were introduced into salicylic acid to reduce its irritating properties, yielding acetylsalicylic acid (1897). Shortly, salvarsan (1910), the first chemotherapeutic agent, and several drugs developed in Germany by Bayer, notably sulfonamides (1930s), followed [24, 25]. The most productive Czech scientist in the field of biomedicine and rational drug design was Antonín Holý who designed three potent antivirotics used against hepatitis B and HIV infections [26].

1.2 Computer-Aided Drug Design

1.2.1 Structure-Based Drug Design

In the 1980s, considerable progress was made in the techniques used for experimental determination of protein 3D structure – X-ray crystallography and NMR spectroscopy. The two approaches provide an insight at a near-atomic resolution and have been providing the majority of now available protein structures. Many resolved structures, often with already co-crystallized drug candidates, are uploaded by scientists to the free online worldwide database PDB (Protein Data Bank) [27]. As of now, PDB contains over 118,000 structures of biomacromolecules and their complexes. Such a variety makes it possible to design drugs based on the knowledge of their corresponding target's structure in a process called structure-based design. The central point of structure-based design is that structure-specific drugs react selectively with their biological target based on their mutual structural and chemical complementarity [7]. Their advantage over non-specific drugs lies in high selectivity and thus lower probability to cause side effects.

In the search for a drug, the first step is determining whether we need an activator or an inhibitor. Inhibitor design is always easier than that of an activator, among other reasons because inhibitors can be obtained by suitably modifying the substrate. Second step should focus on finding whether there are any experimental leads yet. As for the target, it should be essential, unique or conditionally expressed for the disease of interest. When a lead is found, it is tested for drug-likeness (solubility, lipophilicity, non-toxicity, stability, molar mass) and pharmacokinetics (absorption, distribution, metabolism, excretion, and toxicity – collectively denoted as ADMET) [28].

The binding of a ligand into an enzyme is mediated by short to mid-range non-covalent interactions. The contributions of interactions in order of decreasing strength are as follows: hydrophobic forces, coulombic charge–charge, charge–dipole, dipole–dipole, dipole-induced dipole and induced dipole–induced dipole. The abovementioned interactions include ion-pair formation, π-π interactions, hydrogen bonds and van der Waals forces. A significant contribu-

tion to protein-ligand interaction can be attributed to the hydrophobic effect. It is an entropy-driven force causing water molecules to orient in such a way along a non-polar surface that the disruption of their 3D H-bond network is minimized. The total change of free energy in the reaction of ligand binding at given temperature and pressure can be calculated from its enthalpic (released heat) and entropic (probability of states) contributions (Equation 2).

$$\Delta G^{\circ}(p, T) = \Delta H^{\circ} - T\Delta S^{\circ}$$

Equation 2: Free energy change of ligand-binding reaction. G° – Gibbs free energy; H° – enthalpy; S° – entropy; p – pressure; T – thermodynamic temperature.

Gibbs free energy itself is related to the value of the dissociation constant by Equation 3 [29]:

$$\Delta G^{\circ} = RT \ln K_D$$

Equation 3: Relationship between Gibbs free energy and enzyme-ligand dissociation constant. G° – Gibbs free energy; R – molar gas constant; T – thermodynamic temperature; K_D – dissociation constant.

The measure of how strongly a ligand binds to its target is affinity. The higher the affinity, the lower dose of the drug is needed to provoke biological response. Affinity is expressed in the terms of energy needed to break the bond between the ligand and the target. It is characterized by dissociation constant of the enzyme-ligand complex. For the reaction

$$\text{Enzyme} + \text{Ligand} \rightleftharpoons \text{Complex}$$

dissociation constant K_D can be calculated using Equation 4.

$$K_D = \frac{[Enzyme][Ligand]}{[Complex]}$$

Equation 4: Dissociation constant of enzyme-ligand complex. K_D – dissociation constant; [Enzyme] – equilibrium concentration of enzyme; [Ligand] – equilibrium concentration of ligand; [Complex] – equilibrium concentration of enzyme-ligand complex.

In biological systems, K_D can span many orders, from 10 mmol L^{-1} (weak binding strength) to 10^{-16} mol L^{-1} (very strong binding).

It is important to note that although their concepts are related, affinity is not the same as biological potency. Whereas affinity merely denotes the strength of a bond, potency is a measure of the efficacy of a substance in producing biological response upon binding to its target. It is often expressed as half-maximal inhibitory concentration IC_{50} at which the substance inhibits its target's function by 50 %, or, analogically, half-maximal effective concentration EC_{50} for activators (Figure 3).

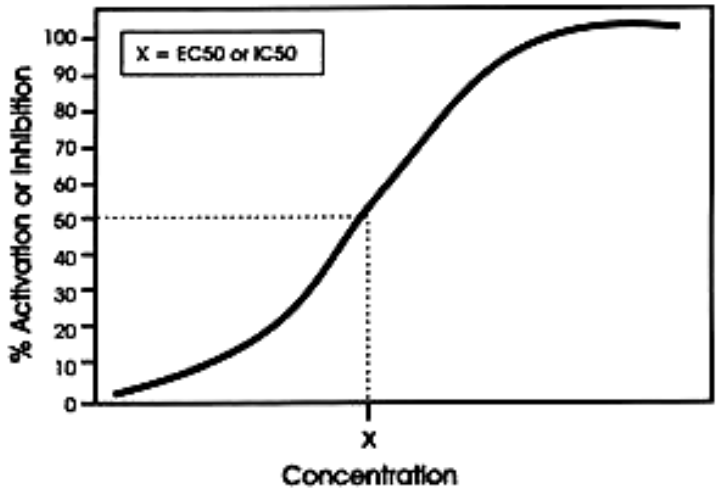


Figure 3: Illustration of IC_{50} and EC_{50}. The sigmoid curve represents dependence of % response (activation or inhibition) on drug concentration.
Taken from http://www.preskorn.com/books/ssri_s3.html.

1.2.2 Computer-Aided Structure-Based Drug Design

It is estimated that the number of molecules synthesizable with currently available synthesis methods amounts to somewhere between 10^{20} and 10^{24} [30]. Although possible in theory, it would be an unthinkable mission to perform a "wet" screening in search for a new drug trying all of them. Luckily, the onset of computer era has brought about a rapid growth of computational power, memory and data storage capacity, enabling scientists to transfer a great deal of tasks of the early stages of drug development to the virtual plane – namely looking for a hit, hit-to-lead optimization and lead optimization. The displacement alleviates many a burden linked to wet screenings, such as having to purchase chemicals, handle and store unstable or toxic substances, perform time-consuming repetitive tests etc. Experimental screening can then be performed only on compounds which are predicted with some level of confidence to possess biological activity. Computer-aided drug design (CADD) usually focuses on the prediction of physico-chemical properties of chemicals followed by clustering them into libraries based on similarity in some of those properties [31]. Two main approaches are structure-based design (based on detailed knowledge of target biomolecule) and ligand-based design (stemming from knowing which molecules are active and which are not while target's structure remains a mystery) [32]. Another use of computers lies in estimating the structure of biomolecules with a known sequence by modeling them *ab initio* ("from scratch") or by searching them against entries in specialized sequence databases. Similarly, their biological function can be inferred from phylogenetic evidence by searching their structures against structural databases of molecules of known function [33].

The invaluable services offered by computers consist in graphical visualization of molecules and in powerful computational tools built upon the knowledge brought by theoretical chemistry, computer science and information science. Cutting-edge supercomputers achieve performance in the order of peta (10^{15}) FLOPS (floating-point operations per second). They are massively-parallel machines with tens or hundreds of thousands individual processors (CPU) [34]. There are basically two approaches in parallel computing: either a centralized

cluster (server farm) of processors located physically close to one another performs the computations, or there are many discrete heterogeneous stations connected *via* a network and coordinated by a central server (distributed computing). The asset of chemical problems is that often, many subtasks can be computed in parallel. Determination of chemical properties can be carried out in a pleasingly parallel workload, *i.e.* such where the separation of the problem into partial tasks doesn't pose a problem [33].

Typical theoretical apparatus underlying CADD algorithms employs quantum mechanics (solving Schrödinger equation is feasible only for systems not exceeding hundreds of atoms), molecular mechanics (classical mechanics applied to atoms represented by rigid balls and connected into molecules by springs), molecular dynamics (physical movement of particles in time) or hybrid methods.

To assess types and strengths of enzyme-ligand interactions, it is useful to describe the surface of the molecules. In computer modeling, this can be done in several ways. For the purpose of computation and visualization, surface area can be represented in different modes (Figure 4). Van der Waals surface is given by the union of the set of all spheres of atoms' van der Waals radii. Van der Waals radius is half of the distance between two free (unbound) atoms in the crystal. Accessible surface area (ASA) or solvent accessible surface area (SASA) is the surface obtained by tracing van der Waals surface of the molecule with the center of a ball-shaped probe of a given radius, typically approximated to water molecule (r = 1.4 Å). Solvent-excluded surface (SES, a.k.a. Connolly surface) is the surface eroded by the same probe [8].

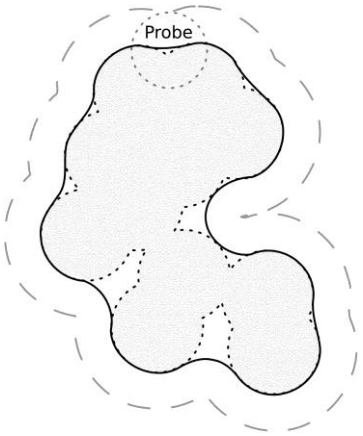


Figure 4: Representations of molecular surfaces. Dotted line – van der Waals surface; dashed line – ASA; solid line – SES (Connolly surface).
Taken from http://link.springer.com/article/10.1186/1471-2105-10-276.

Additionally, surface-exposed atoms or moieties can be assigned descriptors such as elementary charges, polarity, acidity or basicity, hydrophobicity and hydrophilicity. The descriptors can be qualitative (e.g. positive/negative) or quantitative (e.g. hydropathy index); they can typically by visualized as color-coded.

Curvature of surfaces is of great interest in molecular modeling as concave surfaces fail to fully satisfy solvent hydrogen bonding requirements (Figure 5). Their relative inaccessibility means fewer H-bonds can form, thus small non-polar ligands binding into concave areas result in a greater enthalpy-driven desolvation change [35]. Hence, clefts between domains are

often the binding sites of small molecules because the flexible connection between otherwise rather rigid conserved domains enables more than one domain to participate in the interactions [6].

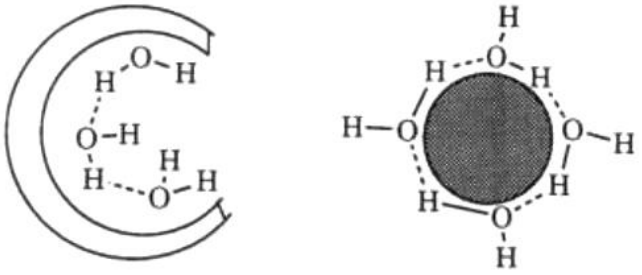


Figure 5: Hydration of concave (left) and convex non-polar surface (right). Taken from http://ncbr.muni.cz/~martinp/C3210/StructBioinf6.pdf.

As witnessed by approved drugs captopril (1981), dorzolamide (1995), saquinavir (1995), ritonavir, indinavir (1996), tirofiban (1998) and others [32], structure-based CADD is a powerful servant. Nevertheless, it is essential to remember that it is still but a tool the results of which must be taken with a grain of salt. When it comes to reliability, it is unlikely that in near future it will be able to hold a candle to experimental methods. As much as CADD takes into account steric effects, affinity, lipophilicity etc., thorough *in vitro* and *in vivo* testing still need to take place to confirm the drug's actual selectivity, pharmacokinetics, stability, solubility etc. A good starting point in this regard is sticking to Lipinski's "rule of five": for the molecule to be bioavailable, it should have at most 5 H-bond donors, 10 H-bond acceptors, a molecular mass of 500 Da and a partition coefficient (octanol : water) of 5 [36]. This is a necessary but not sufficient condition, for many unexpected factors can occur.

1.2.3 Detecting Pockets in Proteins

As mentioned earlier, concave areas in proteins are often ligand-binding sites. There is no naming convention as to their differentiation by name but generally they can be divided to cavities (void spaces buried in the protein), invaginations (cavities connected with the protein surface *via* a tunnel), and pockets, clefts or grooves (shallow depressions on the surface). There are several dozen pocket-finding computational tools at hand, many of which come with a front-end for visualization and description of pocket metrics, surface properties, listing of residues involved in binding etc. They can be divided into several groups by the general approach they exploit for recognizing concave areas in 3D structures of biomolecules.

Geometric algorithms usually employ splitting the molecule into a number of smaller sectors using some type of 3D grid. Subsequently a ball-shaped probe is rolled over the grid points located near the surface while its clashes with protein's atoms are registered. Sufficiently overlapping segments containing no clashes are then aggregated into larger clusters. The tools SURFNET [37], LIGSITE [38], CAST [39], fpocket [40] and VICE [41] fall into this category. Early methods would e.g. measure angles between centers of spheres located on the protein surface [42], while VOIDOO algorithm used iterative mapping of the surface with a probe followed by increasing van der Waals radii to close the cavities off [43]. Others rely on various shape descriptors or filling the cavities with spheres which are then clustered or sub-

stituted with bigger spheres. The outcome of analysis with geometric algorithms is influenced by structure resolution, orientation of coordinates, cavity size and position. Their drawbacks consist in neglecting protein dynamics and in their tendency to miss less common binding sites outside concave areas but on the other hand, they work fast and require few computational resources [32].

Energy-based approaches apply a grid to a static structure as well but instead of relying purely on geometry, they calculate binding energies of a small probe docked into the structure. They take into account electrostatic, van der Waals, H-bond, hydrophobic and solvent terms. GRID [44], CS-Map [45], Q-SiteFinder [46] and PocketPicker [47] employ this type of algorithm. Simple energy-based computations are almost as fast as geometric methods and offer a higher sensitivity. However, there are many energy minima on proteins' potential energy surfaces, resulting in a high chance of false-positive hits. Additionally, which pockets are detected depends on the properties of selected probe [32].

Unlike the previously mentioned methods, pocket detection based on molecular dynamics (MD) perceives proteins as structures whose shape fluctuates in time. These methods often require a set of several conformations as input; such data can be obtained by running an MD simulation carried out from a single structure. MD simulations numerically solve Newtonian equations of motion for all atoms placed in a force field defined by molecular mechanics (MM). The structure of the molecule is calculated for all states along a trajectory fractionated into many short time steps of e.g. 1-10 fs. MD-based tools include MDpocket [48], dxTuber [49] and Caver 3.0 [50]. Drawbacks of MD methods include the tendency of the algorithms to get stuck in local energy minima and high time complexity [32].

Finally, some tools use not-so-standard methods such as homology searches, e.g. SiteEngines [51], introduction of mutations [52], or consensus approaches, like MetaPocket [53].

Each of the algorithms must employ ranking of some kind to estimate which pockets are most likely to bind small molecules. Nearly planar surfaces happen to be the sites of protein-protein interactions more often than small ligand binding sites, therefore exclusion of shallow pockets may take place. Besides, descriptors suitable for detecting binding site of one type of molecule may overlook sites for other types of ligands. Another issue is posed by the fact that pocket-finding algorithms neglect induced fit mechanics and thus may fail to detect cavities without a bound ligand.

Fpocket is a free open-source geometric tool employing Voronoi tessellation and α-spheres. It is able to detect over 90 % of top-three ranked pockets, which is an above-average feat among currently available algorithms, and is complemented with adequately high speed [40].

In a metric space R^n with a defined distance function d, Voronoi cell V_i of an atom a_i is the set of all points x in R^n whose distance from a_i is lesser than or equal to their distance to each atom a_j while $i \neq j$ (Equation 5).

$$V_i = \left\{ x \in R^n \;\middle|\; d(a_i, x) \leq d(a_j, x) \; \forall \; i \neq j \right\}$$

Equation 5: Definition of a Voronoi cell. V_i – Voronoi cell of an atom a_i; R^n – metric space; $d(a, b)$ – distance of points a and b (for Euclidean space, d is their Euclidean distance).

Each Voronoi cell contains exactly one atom. In layman's terms, a 2D Voronoi tessellation is a polygonal partition of the plane in which every point of an edge is equidistant from exactly two nearest atoms and each vertex is equidistant from three or more [54]. Each convex polygon is the intersection of all half-planes delimited by the perpendicular bisectors between a_i and each a_j. When extended to 3D space, Voronoi edges form planar 2D faces and the cells, convex 3-polytopes. Delaunay triangulation diagrams convey equivalent information as Voronoi tessellation diagrams because one can be mapped directly from the other. Delaunay diagram can be obtained from a Voronoi diagram by connecting every tuple of atoms separated with a Voronoi edge with a straight line.

For each Voronoi vertex in 2D, a circle can be constructed with the center in the vertex such that it touches at least three atoms but contains none inside (Figure 6). Analogously, α-spheres touching at least 4 atoms can be obtained in 3D space. α-sphere radii correspond with the curvature of the protein's surface; highly curved surfaces yield spheres of small radii whereas four atoms located on a plane would produce an infinitely great radius. Thus, by filtering α-spheres by their radii gives information about protein surface; discarding too big and too small spheres will leave only spheres located in concave areas.

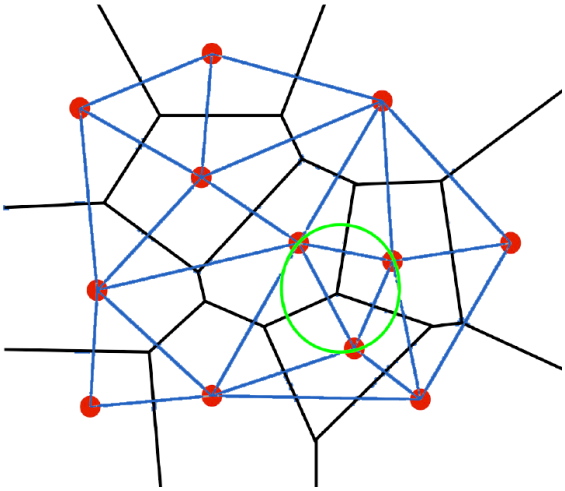


Figure 6: 2D Voronoi diagram (black lines) and its relation to Delaunay triangulation (blue lines) for a set of atoms (red). A Delaunay triangle is circumscribed by a green circle (2D counterpart of α-sphere) with the center located in a Voronoi vertex.
Adapted from [55].

In the next step, fpocket clusters related α-spheres with proximate centers of mass. Polarity of found pockets is also evaluated – α-spheres containing at least two polar atoms are marked as polar while those containing at least 3 atoms with low electronegativity are marked as non-polar. Finally, a simple scoring function based on Partial Least Squares is applied to determine the propensity of each pocket to bind small molecules. The output of the software presents information about detected pockets sorted by rank [40].

1.2.4 Molecular Docking

After binding sites in a protein are identified, it is useful to find out what molecules satisfy spatial and physico-chemical requirements which would enable them to non-covalently bind in those sites and thus perhaps produce a biological response in the biomolecule. This can be done by the means of molecular docking – a computational procedure which determines the best mutual alignment of two molecules and the affinity of the ligand. Besides scanning libraries of small compounds in search for suitable ligands for a protein, it can as well be used to infer the conformation of the complex of two molecules of known structure but unknown binding mode (*i.e.* position and conformation) [56].

The molecules can be represented in three ways which influence how the docking will be done. Atomic representations are computationally expensive as they have to tackle the interactions between all tuples of close atoms. It is used with potential energy functions used for scoring. Surface representation considers only shapes of the molecules and looks for complementarity. It considers mainly surface atoms in the computation but the interior atoms cannot be neglected either for they contribute to the properties of the surface. The surface is considered as a set of geometric features like convex, concave and saddle-shaped areas modeled according to their accessibility to a probe. Grid representation replaces the protein with a grid of points with given properties, then assigns them "interior", "exterior" and "surface" markers according to their position in the protein. A similar procedure is performed with the ligand, after which the method looks for maximization of the contact of surface areas of the protein and the ligand [32].

The simplest case of docking considers two rigid molecules. The protein remains fixed in space whereas the ligand is subject to a search for best binding mode. The ligand molecule has 6 degrees of freedom – three translational and three rotational ones. The docking algorithm separates the space with a grid, then positions the ligand molecule (usually its center of mass) into different points in it, calculating binding energies for possible rotations. Decreasing the angle of rotation steps and lowering grid granularity lead to longer computational times, because the number of ligand poses (set of specific translational and rotational variables) grows rapidly and easily reaches tens of thousands poses for even simple ligands in small, coarse grids. By introducing flexible (stretchable, rotatable) bonds, the number of degrees of freedom rockets and so does the computational complexity. Therefore exhaustive or brute force algorithms (systemically going through all possible poses) would be highly inefficient and are replaced with optimized methods [57].

One of the simplest solutions is manual, *i.e.* user-driven docking. However, this requires that the user have a good idea of how the ligand will bind in the pocket, for example thanks to knowing a crystal structure with a very similar ligand. The search space is very narrow in this case. The drawback is that even very similar ligands can adopt quite different poses, as demonstrated by crystallographic studies. One of the first docking algorithms, DOCK, was based on complementarity of ligand and pocket shapes represented by overlapping spheres. Fragment docking is a method where a rigid ligand moiety (such as an aromatic ring) is docked into a favorable position and anchored there; subsequently the rest of the molecule is searched for acceptable conformations. This approach is used by FLEXX, FLOG, SURFLEX and SEED. Stochastic methods are, so to say, descendants of serendipity as they cannot guarantee

that they will find the optimal solution; they rely on generators of pseudorandom numbers based on which they place and orient the ligand. Already explored poses can be discarded to ensure that broader conformational space is explored (so called tabu search). By repeating the steps many times, different statistically distributed poses are covered. Monte Carlo methods are an example of stochastic approach. They methods can be implemented with Metropolis criterion which directs the search to converge to lower energy states. Stochastic methods are an alternative for exhaustive methods when the search space is too large. An example of Monte Carlo-Metropolis approach is MDOCK. MD methods employ classical Newtonian laws. They are not very suitable for docking because of high dependence on starting conformation and inefficiency in overcoming high energy barriers on potential energy surfaces; this holds especially for small, very flexible ligands [58]. This inconvenience can partly be suppressed by strategies like simulated annealing. Finally, genetic algorithms can also be used. They simulate evolution – the parameters of "parent" generation undergo minor random mutations to give rise to "child" generations. New conformations are kept if their parameters score better. An example of a genetic algorithm in docking is GOLD [32]. All of the algorithms make simplifying assumptions, for example that the bond lengths and angle do not change; that charges remain the same in the course of the docking; that the protonation states do not change, etc. [56].

Complications concerning docking arise from the fact that even when an algorithm finds a well-scoring binding mode, it may still not correspond to reality. The algorithms can't fully take into account all factors influencing the binding, e.g. solvent model, low Gibbs free energy value of the interaction, multiple binding sites, induced fit [8]. Another problem is that crystallographic structures are usually the ones used for docking. Their ligand is removed and the freed binding site stays in the conformation adapted to the original ligand's shape, therefore docking has a tendency to identify ligands of a very similar structure while in fact, the protein could accommodate a much wider range of shapes.

Docking software typically immediately filters out poses with substantial steric clashes. The rest are evaluated with scoring functions. When a single ligand is docked, only its fit with the protein needs to be assessed but when scanning large libraries, the requirements on the scoring function include not only scoring affinities of individual ligands quickly but also being able to compare the binding of ligands relatively to each other. For this purpose, maintaining a good comparative ranking is more important than the absolute scale of scores. A handful of scoring functions can be used; each of the available approaches has its pros and cons. Most scoring functions approximate the binding free energy by adding up the terms describing energetic contributions of solvent effects, conformational changes of both the ligand and the protein, specific protein-ligand interactions, free energy loss due to freezing rotatable bonds, the loss of free rotational and translational free energy caused by binding of the two molecules into a single complex, changes in vibrational modes and possibly other terms (Equation 6). The inclusion or omission of different parameters are a hot topic to-day. The parameters in consideration may further be assigned a weighing coefficient extrapolated from experimental data or quantum mechanical calculations.

$$\Delta G_{binding} = a_1 \Delta G_{solvent} + a_2 \Delta G_{conf} + a_3 \Delta G_{\text{int}} + a_4 \Delta G_{rot} + a_5 \Delta G_{t/r} + a_6 \Delta G_{vib} + \ldots$$

Equation 6: Gibbs free energy of binding between the protein and the ligand. ΔG – Gibbs free energies of individual contributions; a_n – experimentally determined weighing factors.

Molecular mechanics scoring functions are more simplistic as they calculate only some of the free energy terms but as a tradeoff, they are more likely to yield inaccurate results. Empirically derived functions observe the dependences of total binding energy on individual parameters, then sum the contributions up: ionic interactions, hydrogen bonding, lipophilic interactions, loss of internal degrees of freedom of the ligand etc. [58]. An example of an atomic interaction-based function neglecting solvation/desolvation and entropy changes could look like Equation 7.

$$S = \sum_{protein} \sum_{ligand} \left(\frac{R_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{r_{ij}} \right) + \sum_{h-bonds} \cos\Theta_{ij} \left(\frac{\overline{R}_{ij}}{r_{ij}^{12}} - \frac{\overline{A}_{ij}}{r_{ij}^{10}} \right)$$

Equation 7: Example of an interaction-based scoring function with atomistic resolution. The first term is empirical Pauli repulsion; the second, van der Waals attraction; the third, electrostatic potential. Fourth and fifth term denote angular-dependent H-bond potential [28].

Knowledge-based scoring functions are built upon knowledge contained in experimentally determined structures. Specific interatomic distances that occur often are considered favored ones while rarer are not. This scoring is applied by DRUGSCORE. Because as of now, no single function achieves perfect accuracy, consensus scoring is sometimes used to get the best results. The point of all these functions is to perform at high speed and throughput; candidate poses with promising scores can be refined in successive steps to calculate more accurate values of binding energies [28].

An important decision in docking for a given purpose is whether or not one should consider flexibility of the molecules and to what extent. Each flexible element in a structure increases the computational complexity immensely because not only relative orientation of molecules must be calculated but also their conformations. In fact, this task has been proved to be NP-hard, which means that increasing the size of an input parameter leads to a polynomial increase in time required to solve the problem [59]. That means that fully flexible docking is not feasible in practice. Flexibility is thus allowed only for a subset of the variables. There are rigid receptor – flexible ligand docking; flexible receptor – rigid ligand docking; and flexible receptor – flexible ligand docking. They are a compromise between rigid-body docking and fully-flexible docking, typically used to refine binding modes found by other methods [60].

It may seem that docking requires huge computational resources and lots of time in return for questionable output. It has, however, proved to be a serviceable tool in drug design. To name at least a few instances of a successful outcome, it has been helping in elucidating the function of acetylcholinesterase reactivators and to design novel antidotes for organophosphorus compound and nerve agent poisonings [61]. Virtual screening has provided the most hits for G-protein coupled receptors and for kinases [62]. And, to conclude, MD docking has led to the discovery of the first approved HIV-integrase inhibitor [32].

1.3 Human GSK-3β as a Biological Target

1.3.1 GSK-3β Characterization and Function

Kinases are enzymes catalyzing the transfer of an activated γ-phosphate moiety (PO_4^{3-}) from ATP to a substrate molecule. This leads to the priming of the substrate or to a modulation of its activity, resulting in transmission of signals and regulation of metabolic pathways. Phosphorylation is one of the possible post-translational modifications of proteins and as such, it plays an important role in metabolism regulation. It follows naturally that dysfunctioning kinases can have a variety of adverse effects on health. Human proteome sports over 500 kinases sharing a highly conserved catalytic site and an ATP-binding pocket. One kinase can phosphorylate a variety of molecules and, conversely, one substrate can be subject to the regulation by a handful of different kinases [63].

GSK exists in isoforms α and β in mammals, encoded by two different genes (GSK3A and GSK3B) and having distinct functions. Insufficiency in one type cannot be compensated for by the abundance of the other. Although they possess 97 % similarity in the active center, there are major differences between them in the remainder of the molecule [3]. GSK-3β (EC 2.7.11.26) is expressed ubiquitously in animals, plants and microorganisms; it is located mainly in the cytosol, but to a lesser extent also in the nucleus and mitochondria. It is a proline-directed serine-threonine protein kinase, *i.e.* its substrate residues can be both serine and threonine side chains which are immediately followed by a proline residue. The general reaction it catalyzes is

$$\text{ATP} + \text{protein} \rightarrow \text{ADP} + \text{phosphoprotein}.$$

It was first isolated from rabbit skeletal muscles and reported to regulate glycogen synthase (GYS1, GYS2) by the means of phosphorylation [1], hence its trivial name. Under physiological conditions, phosphorylated GYS is the inactive form *in vivo*. It easily becomes allosterically inhibited with cytosolic ATP, ADP and P_i; the active form doesn't show such inhibition. Phosphorylation of GYS by GSK-3β inactivates the former. Glycogen synthase, as its name suggests, is responsible for the synthesis of glycogen from an oligosaccharide primer provided by glycogenin. Glycogen is a glucose polysaccharide serving as energy depository, therefore the regulation of this pathway has impact on glucose uptake and storage [6].

Since its discovery, GSK-3β has been proved to partake in many metabolic pathways beyond that reported by early studies. To name a few, the protein tyrosine phosphatase 1B (PTP1B) insulin signaling pathway modulates insulin-dependent signaling and thus blood levels of glucose; in Wingless integration gene (WNT) signaling pathway, GSK-3β forms a degradation complex with axin, adenomatous polyposis protein (APC), protein phosphatase 2A (PP2A) and casein kinase 1α (CK1α) which degrades β-catenin, thus activating gene transcription [64]; through phosphorylation of Tau-protein it interferes with the stabilization of microtubules in the nervous system; *via* interaction with Microtubule-actin cross-linking factor 1 (MACF1), it disrupts cell migration; and there seem to be more pathways, many of which still awaiting elucidation [3]; more details can be seen in Figure 7.

GSK-3β is implicated in metabolic disorders – diabetes mellitus type II (which actually was the first disease of interest after GSK-3β discovery) by regulating glucose levels, and obesity

by promoting adipogenesis. Besides that, it regulates many neuronal signaling pathways, wherefore its overexpression or aberrant activity are connected with neuroinflammation and several neurodegenerative diseases. It takes part in the formation of neurofibrillary tangles (NFT) and, *via* hyperphosphorylation of tau-protein, amyloid-β plaques resulting in neuronal apoptosis: these conditions are the staple degenerative processes in Alzheimer's disease. [3]. It has been demonstrated that certain GSK-3β genotypes are associated with an increased risk for Parkinson's disease; furthermore, it is strongly suspect from accounting for Huntington's disease, bipolar disorder, depression, schizophrenia and stroke [65]. Last but not least, since GSK-3β acts as a negative regulator of cell growth, its dysregulation can result in development of colorectal, breast, prostate, lung, ovarian and stomach cancer. Some studies also indicate its involvement in the formation of melanoma and glioblastoma [3].

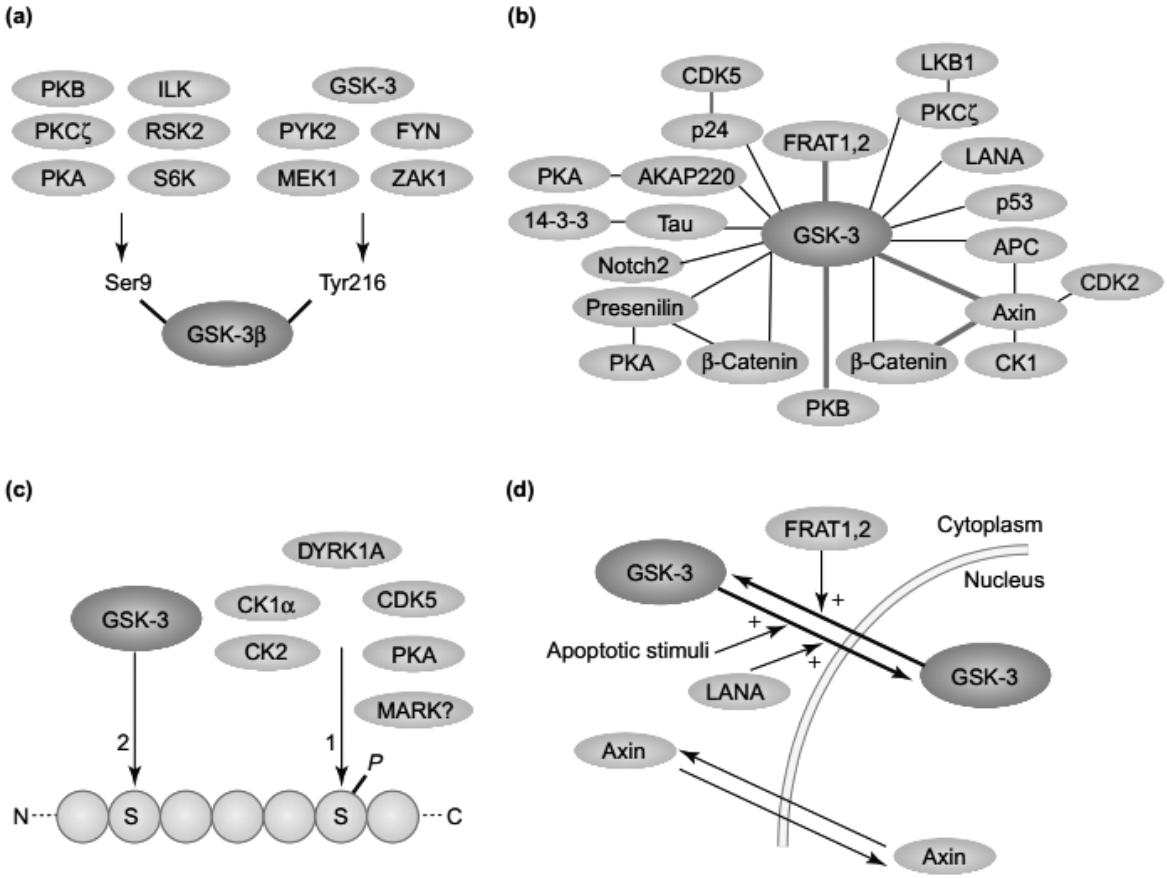


Figure 7: Diagram showing the involvement of GSK-3β in metabolism [66]. (a) post-translational modifications of GSK-3β; (b) association into multi-protein complexes; (c) primed GSK-3β substrates; (d) distribution within the organism.

From the aforementioned information it is quite obvious why GSK-3β has become a tempting target in the quest for drugs treating so far unmet diseases. Unfortunately, as it is involved in so many pathways, it will be difficult to pinpoint and exploit suitable regulatory sites without disrupting the enzyme's function in the pathways which need to stay intact.

1.3.2 GSK-3β Structure and Modulation

PDB contains several dozen GSK-3β structures, hinting at how rich its conformational space is. GSK-3β in crystals occurs as a slightly asymmetric, low-affinity homodimer. Each subunit has 420 residues but experimentally resolved structures only show the residues from Lys35 to Ser386 as the terminal regions lack rigid structure in the crystals and thus don't show clearly localized electron densities (Figure 8). The molar mass of monomer is about 47 kDa. The N-terminal domain (residues 35-134) forms a seven-stranded orthogonal β-barrel whose 5th and 6th strands are connected with a short, two-turn α-helix. The barrel serves as an activation segment dissimilar from other Ser/Thr kinases, reminding instead of traits of tyrosine kinases. The C-terminal α-helical domain (152-342) is connected to the N-terminal domain by an α-helix (138-149). It is *via* the α-helical domain that the dimer subunits bind. The activation segment link is stabilized by the presence of an exogenous anion, a sulfate group (SO_4^{2-}) in the crystal but possibly a phosphate *in vivo*. The core α-helical domain shows topological resemblance to mitogen-activated protein kinases (MAPK) and cyclin-dependent kinases (CDK) [67]. The closest conformation to GSK-3β among other kinases has been found in ERK2 (a MAPK). While ERK2 shows a strong preference for proline residues at the n+1 position, GSK-3β is more tolerant [68].

Activation segment is present in many kinases; it contains residues which can be phosphorylated in order to increase the kinase's activity manifold (over 1000-fold for MAPK) by altering the conformation of catalytic and substrate-binding sites. In GKS-3β, this residue is Tyr216 but the resulting modulation is subtler. As phosphorylated kinases are active and dephosphorylated ones, inactive, it is apparent that different kinases can regulate one another, participating in an intricate network of delicately balanced signal transduction pathways.

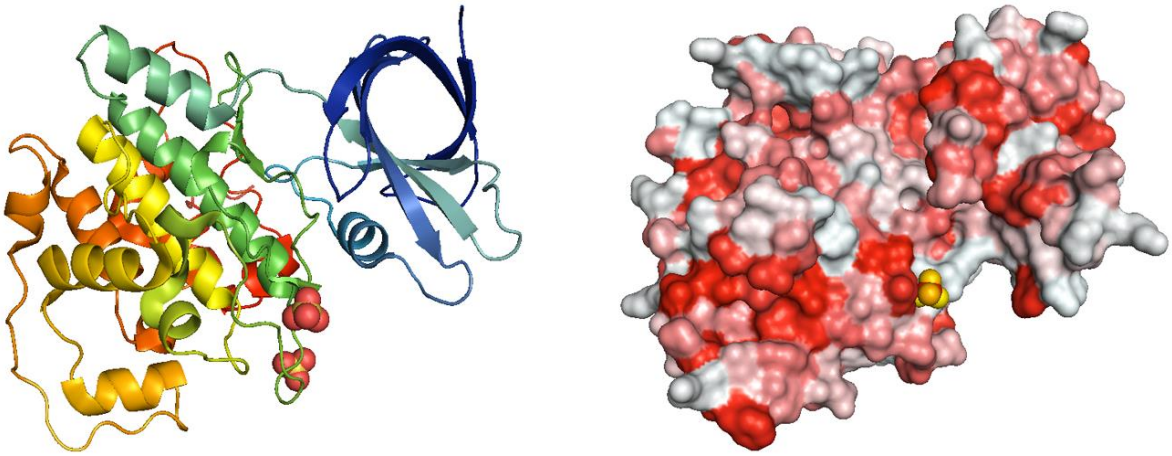


Figure 8: Structure of human GSK-3β (PDB 1GNG) monomer. Left: cartoon demonstrating the position of N-terminal β-barrel with its auxiliary α-helix (blue), and the α-helical domain (rainbow). Two co-crystallized sulfate anions can be seen near the joint between both domains. Right: SES (Connolly surface) for a water probe. Red regions are hydrophobic.

GSK-3β's active site is formed by the residues Lys85, Glu97, Asp181 and Asp200. The ATP-binding pocket is found in the cleft between β and α-lobes. In the vicinity of this site are the binding sites of two Mg^{2+} ionic cofactors. The enzyme exhibits a preference for substrates con-

taining the sequence S/T-X-X-X-Sp/Tp, *i.e.* such where the serine or threonine residue at the position n+4 (n denotes the substrate side chain) has already been phosphorylated (Figure 7). Substrates lacking the primed site are phosphorylated less readily. Binding pocket for the primed substrate is found in the activation loop (Arg96, Arg180 and Lys205) [68]. Inhibition is achieved by the phosphorylation of Ser9, signaled by insulin. It works as a competitive auto-inhibition mediated by binding of the phosphorylated N-terminal region, similar to a primed pseudosubstrate, into the enzyme's own active site [67].

There are over 40,000 molecules reported to inhibit GSK-3β to a certain extent but IC_{50} values have only been determined for about 7,500 of them [3]. Lithium has been known as a treatment for bipolar disorders since 1950s; it was later found that its main target is GSK-3β and that it competes with its magnesium cofactors. Lithium ions are, alas, not sufficiently selective so high concentrations are required. Most to-day known inhibitors compete for the ATP-active site but due to relatively high structure similarity in the kinase family, selectivity still remains an issue. It is unlikely that selectivity for GSK isoforms could be achieved by an ATP competitor; higher selectivity could, however, be attained by targeting β-isoform-specific pockets. So far the most specific inhibitors are paullones, substituted indirubins and substituted maleimides. Their shared traits are low molecular mass (<600 Da) and flat molecules with hydrophobic heterocycles [66].

2 Practical Part

2.1 Finding Pockets in Human GSK-3β

2.1.1 Experimental Setup

The study is a follow-up of the work carried out in 2011 by Palomo et al. [63] who undertook the study of binding pockets in then available 25 GSK-3β structures in sufficient resolution. They found 7 pockets conserved across structures, three of which were already-known binding sites of ATP, substrate, and axin and frattide peptides.

PDB currently stores 65 GSK-3β structures, all of which have been determined by X-ray crystallography. 48 of the structures are co-crystals with inhibitors; vast majority of the inhibitors binds in the ATP site and the enzyme in remaining crystals is either inhibited by phosphorylated peptides or autoinhibited. The enzyme is co-crystallized with different ligands and the resolutions range from 1.80 to 3.20 Å. Two different structures were selected based on their novelty compared to Palomo's research, their reasonable resolution and suitable ligands: GSK-3β bound to an inhibitor (N-[4-(isoquinolin-7-yl)pyridin-2-yl]cyclopropanecarboxamide) as a representative of inhibited enzyme conformation, PDB access code 4PTE (released in 2015); and uninhibited GSK-3β complexed with a primed axin peptide bound at its substrate binding site, PDB ID 4NM0 (released in 2014).

One of the measures of structure quality in PDB is Ramachandran plot. The protein backbone is created by repeating regular segments of C_α and planar amide bonds stabilized by delocalized electrons. Due to rigidity caused both by amide bonds and by steric effects of residues' side chains, the backbone dihedral angles φ and ψ (see Figure 9) are limited to certain value ranges.

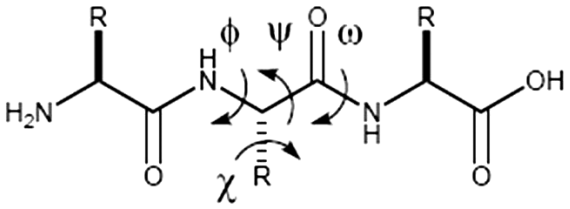


Figure 9: Dihedral angles φ, ψ and ω in protein backbone.
Taken from http://www.keyword-suggestions.com/ZGloZWRyYWwgYW5nbGUgcHJvdGVpbnM.

When φ and ψ of individual amino acid residues are plotted on x and y axis, densely populated areas of naturally occurring conformations form a characteristic map showing favored and allowed conformations, called Ramachandran plot or Ramachandran diagram [6]. Different amino acids have different rigidity and different side chain volumes so their accessible areas in the plot can vary. In X-ray crystallography, outliers in Ramachandran plot can indicate that something had gone wrong during crystallization or during the experiment. The lower the number of outliers, the more representative the obtained structure. Therefore Ramachandran plots are important indicators of structure validity. As can be seen from Figure 10, the selected structures have a good degree of reliability: there is but one outlier in 4PTE and none in 4NM0.

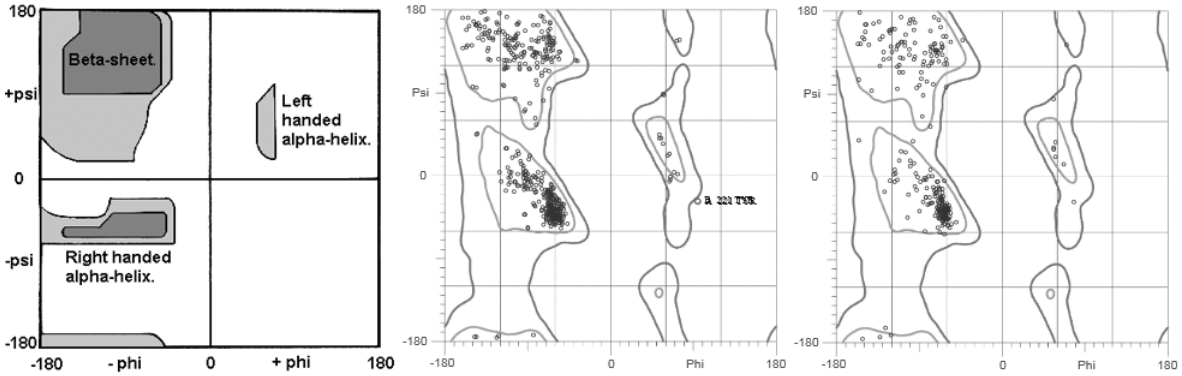


Figure 10: Ramachandran plots. Left: generic plot indicating favored (dark grey) and allowed (light grey) conformations; center: 4PTE; right: 4NM0.
Taken from http://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/rama.html (left) and from corresponding PDB entries (center and right).

Fpocket was chosen for detecting pockets because of its reported decent performance, free availability for distribution, and platform independence. Fpocket accepts adjusted PDB structures as input. The structure cleanup consisted in removing all solvent (water and buffer) molecules, ions, ligands and non-polar hydrogen atoms. Only one monomer subunit was left in place. Resulting structures were randomly rotated to reduce bias, as the pocket finding algorithm may perform slightly differently for different coordinates.

Fpocket parameters were set as follows: minimal radius of αsphere to 3 Å; maximum radius of α-sphere to 6 Å; minimum apolar neighbors to 3; minimum number of α-spheres to 35; maximum clustering distance to 1.73 Å; maximum distance between pockets to 4.5 Å; maximum distance for single linkage to 2.5 Å; and minimum proportion of apolar spheres to 0.0.

Minimal radius of α-spheres sorts out too small spheres which arise mainly in small internal cavities. Maximum radius, contrariwise, discards too big spheres which are located on more or less planar surfaces. Apolar neighbors parameter sets the criterion for polarity: only spheres containing at least the indicated number of apolar atoms on their surface will count. Minimum number of α-spheres parameter also prunes too small cavities; only pockets with sufficient number of spheres are left. The last three parameters govern clustering. Maximum clustering distance influences the first clustering step – it clusters spheres whose distance is lesser or equal to the threshold and whose centers are connected by a Voronoi edge, forming small pockets. Maximum distance between pockets determines how small primary pockets will be clustered based on the distance of their centers of mass. Linkage distance checks how many α-spheres belonging to different pockets coincide and how much. It is determines the results of the last clustering step. Minimum proportion of apolar spheres spans from 0.0 to 1.0 and has the final word about whether or not a pocket will be discarded due to overly high polarity. The value is defined as the ratio of apolar spheres to the total amount of spheres in the pocket; 0 means that all pockets are kept regardless of polarity [69].

Volume of the pockets was calculated in 2,500 iterations. The calculation uses a stochastic approach based on pseudo-random numbers. The implemented Monte Carlo method generates random coordinates and checks if the point lies within the pocket. This step is repeated

many times and the volume is eventually obtained from the ratio of hits to number of iterations.

2.1.2 Results

Fpocket detected 15 pockets in 4PTE (Figure 11, Figure 12). The top-scoring pocket corresponds to the site where the ATP-competitive inhibitor was bound in the original structure. The 4th-scoring pocket lies on the opposite side of the joint between α and β-domains. The remaining sites might act as allosteric sites. The 6th-scoring pocket forms a gap in the otherwise more or less planar area where the two subunits of the dimer contact. It is the same site where axin binds. The 2nd-scoring pocket is the largest and more polar than the ATP-binding site. It is located close to the 5th pocket; they might form a single binding site. Detailed data about the pockets can be found in Table 1.

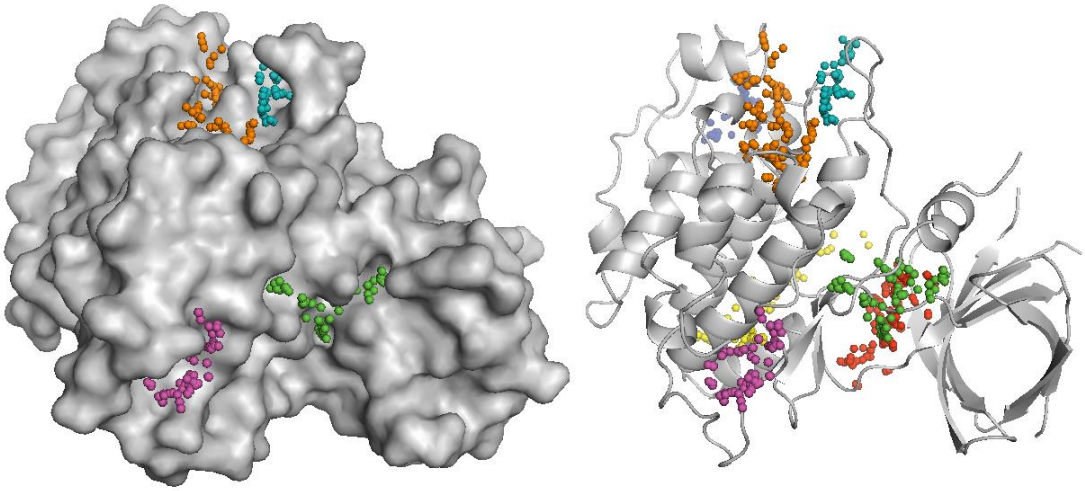


Figure 11: Top 7 scoring pockets in 4PTE. 1st – red; 2nd – orange; 3rd – yellow; 4th – green; 5th – teal; 6th – blue; 7th – magenta.

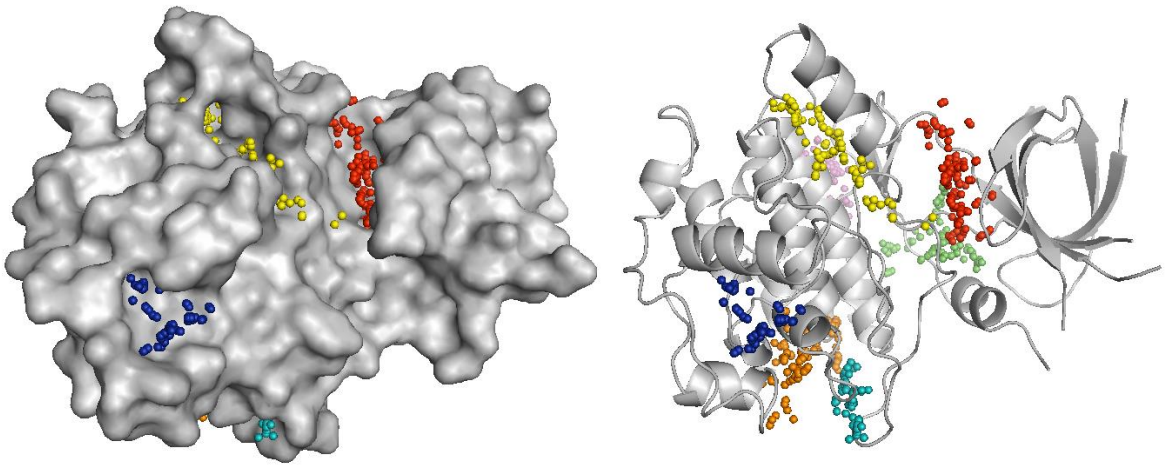


Figure 12: Top 7 scoring pockets in 4PTE – rear view. 1st – red; 2nd – orange; 3rd – yellow; 4th – green; 5th – teal; 6th – blue; 7th – magenta.

The search of 4NM0 structure returned 20 pockets (Figure 13, Figure 14). Again, the top-scoring pocket, along with pocket no. 6, are located in the crevice between the two domains. It is the binding site of ATP along with two Mg^{2+} cofactor ions. The 4th-scoring pocket is part of the binding site of axin peptide. Interestingly, just like in the inhibited protein 4PTE, the 2nd and the 5th pockets are located in the same lobe, their ranking is the same and they are so adjacent that they might form a single site. Pocket no. 7 at the opening of the β-barrel accommodates a glycerol molecule in the original structure. More details can be seen in Table 2.

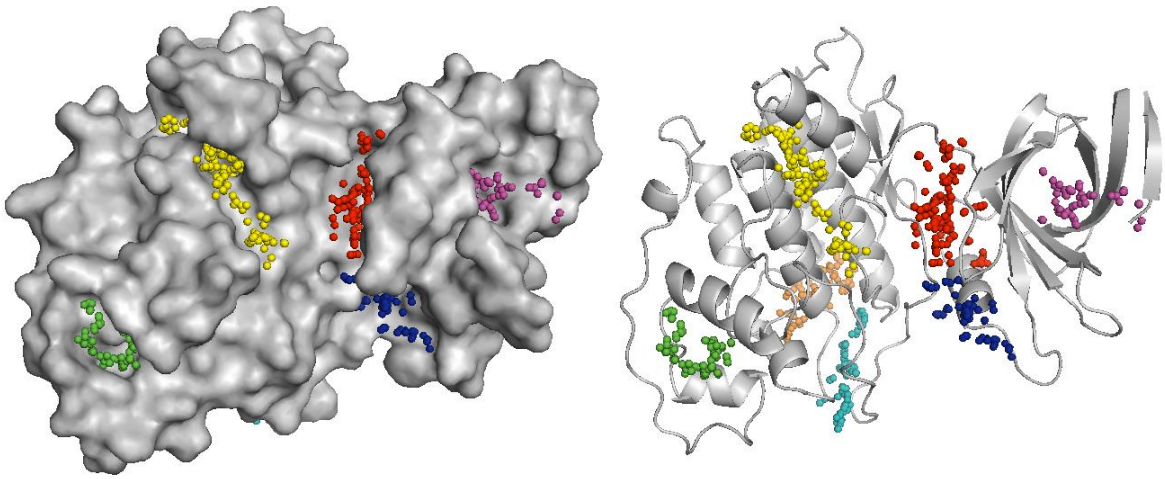


Figure 13: Top 7 scoring pockets in 4NM0. 1st – red; 2nd – orange; 3rd – yellow; 4th – green; 5th – teal; 6th – blue; 7th – magenta.

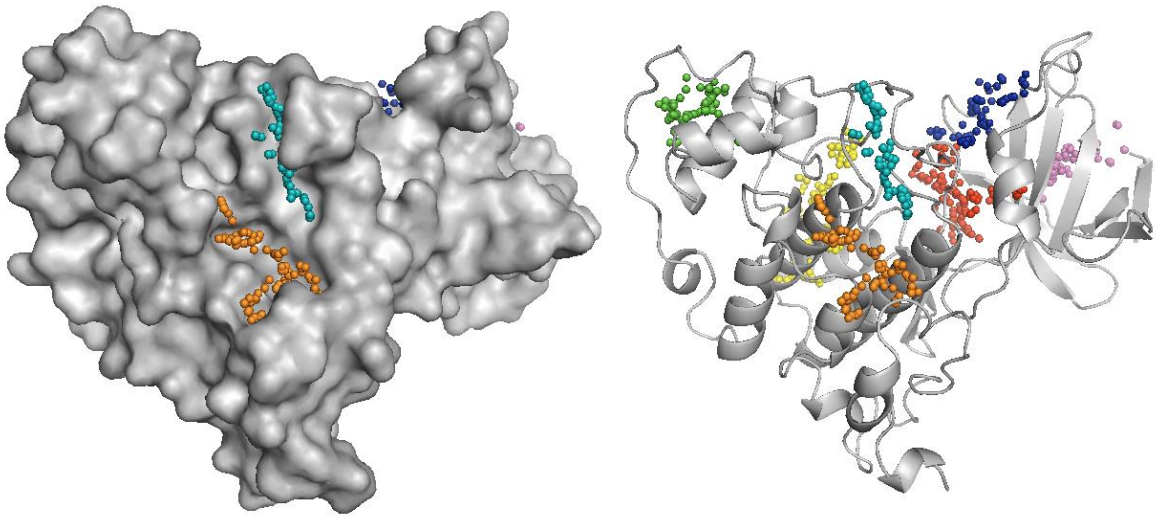


Figure 14: Top 7 scoring pockets in 4NM0 – different view. 1st – red; 2nd – orange; 3rd – yellow; 4th – green; 5th – teal; 6th – blue; 7th – magenta.

Table 1: Pocket descriptors for 4PTE.

No.	Score	Hydrophobicity score	Polarity score	Volume Å^3	Charge score
1	38.6565	32.5200	13	2372.5518	2
2	32.3441	13.9615	15	3555.5063	-4
3	23.1745	51.2174	10	2718.9121	0
4	18.8411	17.9500	11	2012.3595	1
5	13.6342	15.0000	9	1205.1886	0
6	11.4581	49.0000	4	1548.2566	1
7	9.8622	34.0714	6	1778.8041	0
8	9.5371	9.0000	9	1578.1012	1
9	8.5365	33.8333	7	1959.4733	0
10	8.4940	30.4000	6	1328.7073	2
11	8.4226	39.7000	6	1443.9266	0
12	5.8861	21.6000	7	1501.4696	0
13	5.7936	59.2727	4	820.7696	-1
14	5.3918	32.7000	3	1185.0250	-1
15	1.4518	35.8462	6	2548.1270	2

Table 2: Pocket descriptors for 4NM0.

No.	Score	Hydrophobicity score	Polarity score	Volume Å^3	Charge score
1	32.1009	36.3929	14	2727.7402	3
2	26.5142	19.3500	13	2004.8190	-2
3	25.5595	51.5417	8	3434.9426	0
4	17.7633	53.8000	4	1777.3320	1
5	17.0870	9.3750	11	1442.2821	0
6	12.1386	17.6471	10	2054.9795	2
7	11.5500	39.2857	7	1687.4370	3
8	11.4333	33.2308	4	1716.0238	-2
9	10.8063	20.5000	6	1763.3873	3
10	10.6944	16.7273	7	839.7458	-2
11	9.1007	34.0714	6	1662.4253	0
12	6.8876	28.8333	4	1454.2434	3
13	6.1350	51.0000	4	854.5778	0
14	6.1038	-14.5000	8	1640.0389	1
15	6.0868	42.8182	6	1476.6116	0
16	5.2157	21.9000	6	1654.5199	0
17	4.3280	43.0000	3	598.7144	-1
18	4.2593	6.0000	5	1781.4888	-1
19	4.0036	27.3000	4	1256.8938	1
20	2.2033	43.3750	1	984.2368	0

2.2 Docking into Human GSK-3β Selected Pocket(s)

2.2.1 Experimental Setup

After pockets had been identified, the putative binding sites were tested by docking against a library of 12 small molecules. Eleven molecules were proclaimed inhibitors of GSK-3β in the literature, the last one was ADP, the enzyme's natural substrate. The selected ligands and their IUPAC names are listed in Table 3.

Ligand ID	Ligand name
30	4-benzyl-4,5-dihydro-3H-1λ³,2λ³,4-dithiazole-3,5-dione
44	4-benzyl-2-ethyl-1,2,4-thiadiazolidine-3,5-dione
4NM0-ADP	adenosine 5′-diphosphate (substrate bound to 4NM0 in its original structure)
4PTE_ligand	N-[4-(isoquinolin-7-yl)pyridin-2-yl]cyclopropanecarboxamide (inhibitor bound to 4PTE in its original structure)
alsterpaullone	9-nitro-7,12-dihydroindolo-[3,2-d][1]benzazepin-6(5)-one
cimetidine	1-cyano-2-methyl-3-[2-[(5-methyl-1H-imidazol-4-yl)methylsulfanyl]ethyl]guanidine
famotidine	3-[({2-[(diaminomethylidene)amino]-1,3-thiazol-4-yl}methyl)sulfanyl]-*N*′-sulfamoylpropanimidamide
indirubin-oxime	indirubin-3′-oxime
ligand_3du8_h	(7S)-2-(2-aminopyrimidin-4-yl)-7-(2-fluoroethyl)-1H,4H,5H,6H,7H-pyrrolo[3,2-c]pyridin-4-one
maleimide	3-(2,4-Dichlorophenyl)-4-(1-methyl-1H-indol-3-yl)-1H-pyrrole-2,5-dione
manzamine A	(1R,2R,4S,5Z,12R,13S,16Z,22R)-25-{9H-pyrido[3,4-b]indol-1-yl}-11,22-diazapentacyclo[11.11.2.1²,²².0².,¹².0⁴,¹¹]heptacosa-5,16,25-trien-13-ol
ZINC5932684	3-(3-{[(2S)-2,3-dihydroxypropyl]amino}phenyl)-4-(5-fluoro-1-methyl-1H-indol-3-yl)-2,5-dihydro-1H-pyrrole-2,5-dione

Table 3: Overview of ligands selected for docking.

Ligands 30 and 44 are reported in literature as potent GSK-3β inhibitors [70]. 4NM0-ADP is the natural substrate which was co-crystallized in 4NM0 structure. 4PTE_ligand is the inhibitor bound in the original structure of 4PTE. The two latter ligands served for validation of the docking results against their original PDB files. Remaining compounds represent several families of GSK-3β inhibitors discovered up to now [66].

The docking was carried out using free, open-source software AutoDock Vina 1.1.2. This software was selected because of its free license, high reliability, accuracy, robustness, exhaustiveness, speed, and easy parallelization. The scoring function of Vina makes a summation of symmetric interaction functions representing intramolecular and intermolecular contributions (Equation 8).

$$c = \sum_{i<j} f_{t_i t_j}\left(r_{ij}\right)$$

Equation 8: Scoring function of Vina. *t* – type of atom; *f(a, b)* – set of interaction functions assigned to an atom pair *a*, *b*; r_{ij} – interatomic distance between atoms *i* and *j*.

Vina employs a stochastic global optimization approach with a genetic algorithm. It alternates mutation steps and local optimization by Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. BFGS is a quasi-Newtonian method which takes into account not only the value of the scoring function at a given point but also its local gradient with respect to its arguments. Several runs are performed from random conformation seeds; they can be pleasingly parallelized, achieving near-ideal speed-up. For input, Vina takes .pdbqt files, an extension of .pdb format. Their difference lies in the fact that .pdbqt provides additional information about atom types and partial charges. [56].

For the screening, rigid-body docking was selected. The ligands were prepared in HyperChem 8 *via* a semi-empiric PM3 (quantum calculation of molecular electronic structure) geometric optimization method. Afterwards, they were converted from .pdb to .pdbqt format in MGL AutoDock Tools. Gasteiger-Marsili empirical atomic partial charges were added and non-polar hydrogens of protonated forms of the ligands calculated for physiological pH 7.4 were merged into heavy atoms. Finally, torsion angles were detected automatically.

Preparation of protein molecules was also carried out in MGL AutoDock Tools. The procedure included the removal of redundant enzyme subunits, cofactors, ions, ligands and water molecules. Although water takes part in the formation of enzyme-ligand complex, Vina is not suited for multi-particle simulations, therefore solvent effects were disregarded. Polar hydrogens were added and non-polar hydrogens were collapsed into associated carbon atoms. Alternative conformations of residues were detected and removed. The resulting files were converted to .pdbqt format. Center of the docking gridbox and its dimensions were set by an in-house C++ program.

The following computation parameters for the targets were used:

4MN0

```
cpu = 16
center_x = -7.259
center_y = 34.229
center_z = -15.976

size_x = 60
size_y = 63
size_z = 75

exhaustiveness = 124

num_modes = 20
```

4PTE

```
cpu = 16
center_x = -13.959
center_y = 1.950
center_z = -44.282

size_x = 69
size_y = 57
size_z = 55

exhaustiveness = 124

num_modes = 20
```

CPU setting denotes the number of processor units which are to be used for calculation. The x, y and z are Cartesian coordinates of the center of the gridbox, followed by the dimensions of the gridbox. The size of the gridbox was chosen to accommodate the bulk of the protein plus some additional volume so as to leave enough space for the ligand to probe the whole of the enzyme's surface. Increasing the exhaustiveness parameter exponentially decreases the probability that the global minimum will not be found. This is achieved at the cost

of linear increase of the time spent on the calculation. Num_modes parameter defines the maximum number of binding modes to be generated.

Setting the CPU parameter to 16 means that multi-threading was used. Distribution of calculations was achieved by the use of an in-house shell script using PBS Torque scheduler for high-performance computing (HPC) systems. The preliminary calculations and refinement of computational methods was carried out at MetaCentrum VO (Virtual Organization). MetaCentrum VO provides infrastructure for distributed computing and storage resources free of charge for students and researchers in the Czech Republic. Subsequent scale-up and finalization of the calculations were performed at IT4Innovations Supercomputing Center research institution, at HPC Cluster with nodes of the following configuration: 2x Intel Xeon E5-2680v3, 2.5GHz, 24 CPU, 128GB RAM, x86-64, CentOS 6.6 Linux, InfiniBand FDR56 / 7D Enhanced hypercube, 0.5PB /home NFS disk storage and DDN Lustre shared storage.

2.2.2 Results

Each of the 12 ligands was docked sequentially into both receptor molecules. Each pair of molecules underwent 10 iterations of docking, each yielding 20 top-scoring binding modes, providing a total of 200 binding modes. An example of one output file can be seen below:

```
Using random seed: 1457334310
mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1         -5.8      0.000      0.000
   2         -5.7      2.735      3.125
   3         -5.7      0.278      2.143
   4         -5.6      1.536      2.284
   5         -5.5      1.132      1.681
   6         -5.5     15.541     16.373
   7         -5.5     15.813     16.578
   8         -5.4      1.869      1.900
   9         -5.4     16.161     16.793
  10         -5.4     16.028     16.634
  11         -5.4     15.968     16.581
  12         -5.4     16.126     16.791
  13         -5.4     15.504     16.409
  14         -5.4      2.257      3.276
  15         -5.4     16.029     16.663
  16         -5.3     15.659     16.668
  17         -5.3     38.914     40.076
  18         -5.3     15.525     16.500
  19         -5.3      2.491      3.621
  20         -5.2     38.855     40.083
```

The "random seed" item is the pseudorandom number generated for the stochastic part of Vina's algorithm. The rows in the file correspond the top-20 binding modes. The first column is the ranking of the mode; the second, estimated affinity (binding energy). Negative number means that energy is released upon creation of the bond, therefore the binding is favorable – the lower the energy, the greater the affinity. RMSD columns show the root mean square deviations between actual binding mode and upper bound (ub) or lower bound (lb) mode. RMSD is a measure of structural similarity – the lower its value, the better.

An overview of calculated binding sites for each protein and ligand can be seen in Table 4.

Ligand ID	Avg 4PTE affinity kcal/mol	Binding sites 4PTE	Avg 4NM0 affinity kcal/mol	Binding sites 4NM0
30	-5.60	ATP barrel opening 2&5	-5.80	ATP barrel opening 2&5
44	-6.30	ATP 2&5 Axin site undetected pocket	-6.40	ATP 2&5 barrel opening undetected pocket
4NM0-ADP	-7.44	ATP 2&5	-7.53	ATP 2&5
4PTE_ligand	-8.89	ATP 2&5	-8.40	ATP undetected pocket
alsterpaullone	-8.70	ATP	-8.37	ATP 2&5 Axin site
cimetidine	-5.23	ATP 2&5 3	-5.41	ATP 2&5 Axin site
famotidine	-5.67	ATP – whole cleft 2&5 undetected pocket	-5.67	ATP 2&5 Axin site
indirubin-oxime	-9.00	ATP	-8.60	ATP
ligand_3du8_h	-8.70	ATP 2&5	-8.10	ATP 2&5
maleimide	-9.28	ATP	-8.31	ATP
manzamine A	-11.40	ATP – whole cleft 2&5 4PTE – pocket 7	-11.10	ATP 2&5 Axin site
ZINC5932684	-8.79	ATP	-8.69	ATP 2&5

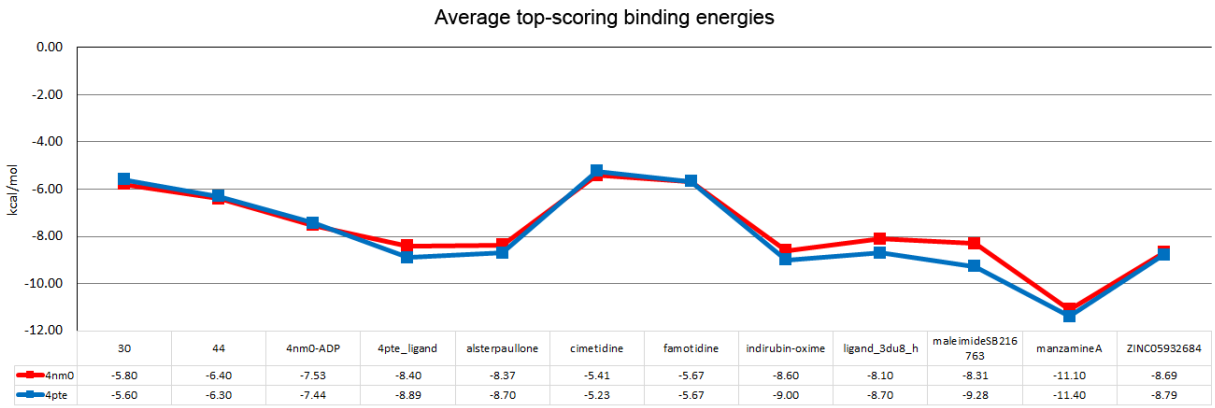
Table 4: Binding energies and binding sites for 4PTE, 4NM0 and 12 ligands. ATP – ATP-binding site, in both proteins detected as no. 1 by fpocket. 2&5 – adjacent pockets scoring second and fifth by fpocket in both proteins. Barrel opening – ATP-cleft-side mouth of the β-barrel, detected as pocket no. 7 in 4NM0, undetected in 4PTE. Axin site – detected as no. 6 in 4PTE and as no. 4 in 4NM0. 3 – 4PTE 3rd ranking pocket (next to ATP-site).

Obtained data show a satisfactory consistency with literary sources. The proteins' corresponding ligands were in most occasions correctly docked in the places where they had originally been found in experimentally determined crystallographic structure (Figure 16). Uniformly, all of the inhibitors and the substrate show a strong preference for the ATP-binding pocket. The docking experiment shows that most of the ligands may be able to bind to other sites as well, only indirubin-oxime and maleimide derivative seem to bind solely in the ATP-pocket. 5 of the ligands show some degree of affinity to the axin-binding site. Ligands 30 and 44 (thiadiazolidinone derivatives) show a considerable affinity toward the mouth of the β-barrel, occluding its side turned to the ATP-binding pocket. This shallow β-barrel pocket has been detected by fpocket but only in 4NM0 and with a low ranking (7^{th} rank). The reported binding modes support the claim that compounds 30 and 44 act as non-ATP-competitive inhibitors. According to Vina, they do bind in the ATP site but much less frequently than all the other tested ligands (Figure 17).

Ten ligands seem to be inclined to adopt positions in the area detected by fpocket as two pockets in close vicinity of each other, ranking identically 2 and 5 in both protein structures. These depressions might together form one larger pocket able to contain a rather large ligand. Seven of the ten ligands have been reported by Vina to bind into this pocket in both of the protein structures. This information is of interest because the place might function as an allosteric site.

The binding energies for corresponding pairs of protein-ligand complexes were virtually the same in both proteins. The chart displaying average energies for each binding couple is shown in Figure 15.

Manzamine was found to have the lowest negative binding energy (lower than -11), *i.e.* it has the greatest affinity to GSK-3β from among the ligands selected for the study. It also tied for having the richest portfolio of binding modes – apart from being able to bind along the cleft between α and β domains, which is the binding site of ATP, it can also occupy three more pockets (Figure 18, Figure 19).



	30	44	4nm0-ADP	4pte_ligand	alsterpaullone	cimetidine	famotidine	indirubin-oxime	ligand_3du8_h	maleimideSB216763	manzamineA	ZINC05932684
4nm0	-5.80	-6.40	-7.53	-8.40	-8.37	-5.41	-5.67	-8.60	-8.10	-8.31	-11.10	-8.69
4pte	-5.60	-6.30	-7.44	-8.89	-8.70	-5.23	-5.67	-9.00	-8.70	-9.28	-11.40	-8.79

Figure 15: Average top-scoring binding energies with different ligands. Red –NM0; blue- 4PTE.

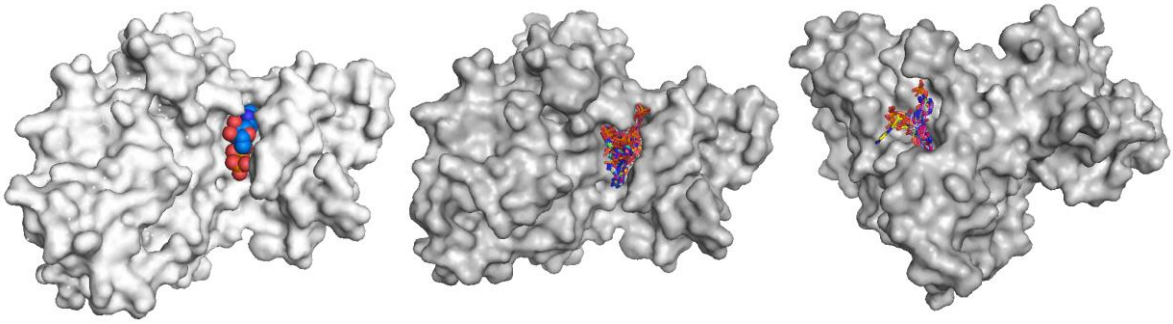


Figure 16: 4NM0 bound to ADP. Left: original crystallographic structure; center: docked structure; right: binding modes with ADP docked in the large, potentially allosteric pocket, identified by fpocket as 2nd and 5th ranking pocket (rear view).

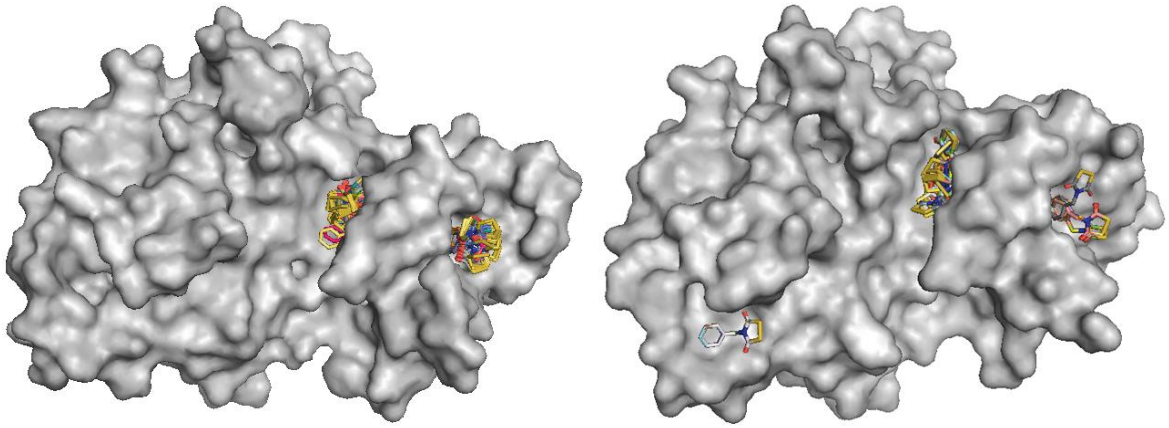


Figure 17: 4NM0 (left) and 4PTE (right) with ligand 30 docked in. The cluster of binding modes close to the center is ATP-binding site. The cluster to the right is bound in the mouth of the β-barrel and may account for non-competitive inhibition.

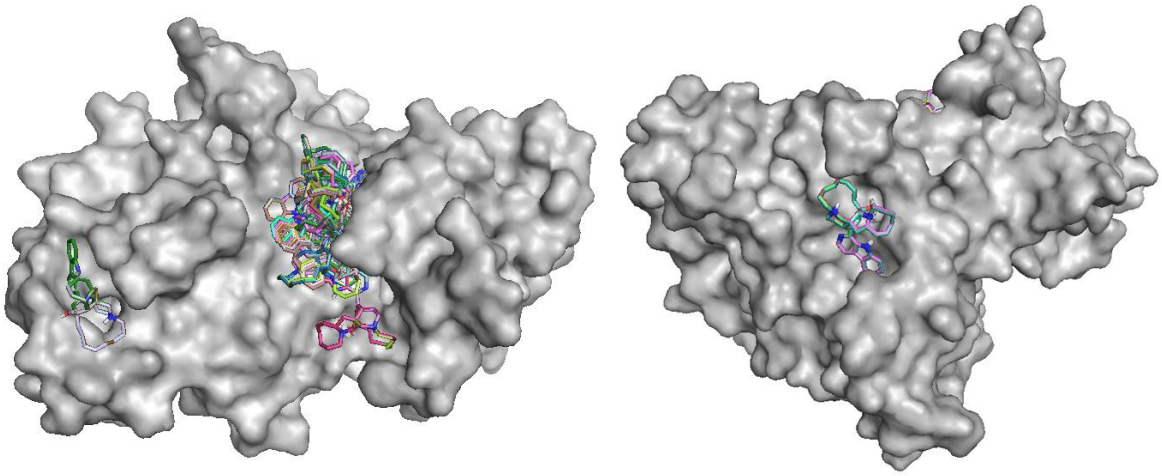


Figure 18: 4NM0 bound to manzamine A. Left: manzamine bound to the ATP-binding site (huge cleft in the middle). Another binding mode can be seen to the left, in the area corresponding to rank 7 pocket found in 4NM0. Right: rear view of the same structure. Manzamine A is bound to the potentially allosteric site (pockets 2 and 5 as ranked by fpocket).

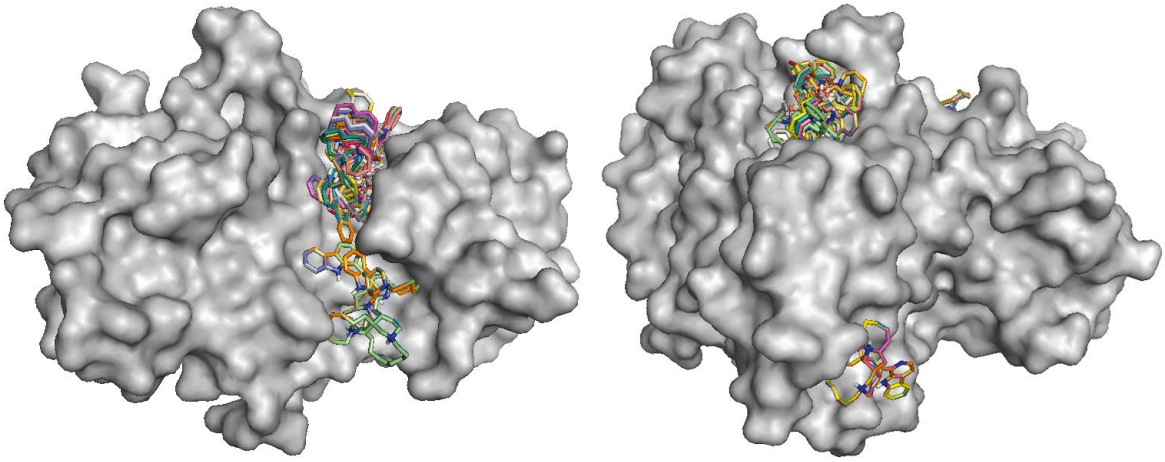


Figure 19: 4PTE bound to manzamine A. Left: manzamine bound to the ATP-binding site (huge cleft in the middle). Right: rear view of the same structure. Near the top of the picture, manzamine A is bound to the potentially allosteric site (pockets 2 and 5 as ranked by fpocket). The bottom position of the ligand corresponds to pocket ranked as no. 7 in 4PTE.

Conclusion

The results of pocket finding are in accordance with available literature on the topic. The pocket which scored by far the best is the ATP-binding site found in the cleft between α and β-domains of GSK-3β. It is followed by the other two already known binding sites, that of protein substrate and that of axin which is known to associate with GSK-3β and other proteins into a gene transcription regulatory unit. No pockets were detected on the planar side where two GSK-3βsubunits contact in crystals except for a small depression forming a part of axin-binding site.

The docking experiment confirmed relevance of top-scoring pockets found by fpocket. First-ranking pocket had by far the highest score; it corresponds to the ATP-binding site. All docked molecules had a decent fit and a high affinity for it and can be considered ATP-competitive inhibitors (except for ADP which is the natural substrate). The cofactor and protein substrate sites were also well characterized but hardly any binding interactions in these regions were found for the ligands. Instead, the ligands showed a tendency to bind to a large pocket which had been confirmed by fpocket and ranked as 2nd and 5th in both proteins. Each ranking corresponds to what could be considered one half of a larger pocket. This might prove to be a potential allosteric site but this hypothesis would require further essays.

It is possible that there are other possibly allosteric sites but they haven't been detected with the setup used in the experiment. This may have been caused by the fact that all "probes" used for rigid-body scanning were ATP competitors and thus had similar properties. It might be worthwhile to try docking a probe with different characteristics in future experiments.

Part of this study's results was submitted to 8th International Conference on Computational Collective Intelligence (ICCCI 2016) which will take place in Halkidiki, Greece, 28th – 30th September 2016.

References

[1] EMBI, N., D. B. RYLATT and P. COHEN. Glycogen synthase kinase-3 from rabbit skeletal muscle. Separation from cyclic-AMP-dependent protein kinase and phosphorylase kinase. *European journal of biochemistry / FEBS*. 1980, 107, no. 2, p. 519–527. ISSN 0014-2956.

[2] PETIT-PAITEL, Agnès. GSK-3beta: a central kinase for neurodegenerative diseases? *Médecine Sciences: M/S* [online]. 2010, 26, no. 5, p. 516–521. ISSN 0767-0974. Available at: doi:10.1051/medsci/2010265516

[3] KANNOJI, Akanksha, Samiron PHUKAN, V. SUDHER BABU and Vitukudi N. BALAJI. GSK3beta: a master switch and a promising target. *Expert Opinion on Therapeutic Targets* [online]. 2008, 12, no. 11, p. 1443–1455. ISSN 1744-7631. Available at: doi:10.1517/14728222.12.11.1443

[4] TAKENAKA, T. Classical vs reverse pharmacology in drug discovery. *BJU international*. 2001, 88 Suppl 2, p. 7-10-50. ISSN 1464-4096.

[5] LANDRY, Yves and Jean-Pierre GIES. Drugs and their molecular targets: an updated overview. *Fundamental & Clinical Pharmacology* [online]. 2008, 22, no. 1, p. 1–18. ISSN 1472-8206. Available at: doi:10.1111/j.1472-8206.2007.00548.x

[6] VOET, Donald and Judith G. VOET. *Biochemistry*. 4th ed. Hoboken, NJ: John Wiley & Sons, 2011. ISBN 978-0-470-57095-1.

[7] VODRÁŽKA, Zdeněk. *Biochemie*. Praha: Academia, 1996. ISBN 978-80-200-0600-4.

[8] LESZCZYNSKI, Jerzy, ed. *Handbook of computational chemistry*. Dordrecht ; New York: Springer, 2012. Springer reference. ISBN 978-94-007-0710-8.

[9] BAKER, D. and A. SALI. Protein structure prediction and structural genomics. *Science (New York, N.Y.)* [online]. 2001, 294, no. 5540, p. 93–96. ISSN 0036-8075. Available at: doi:10.1126/science.1065659

[10] FORMAN-KAY, Julie D. and Tanja MITTAG. From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins. *Structure* [online]. 2013, 21, no. 9, p. 1492–1499. ISSN 0969-2126. Available at: doi:10.1016/j.str.2013.08.001

[11] FISCHER, Emil. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* [online]. 1894, 27, no. 3, p. 2985–2993. ISSN 1099-0682. Available at: doi:10.1002/cber.18940270364

[12] HENRI, Victor. Théorie générale de l'action de quelques diastases. *Comptes rendus biologies* [online]. 1902, 329, no. 1, p. 47–50. ISSN 1631-0691, 1768-3238. Available at: doi:10.1016/j.crvi.2005.10.007

[13] KOSHLAND, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*. 1958, 44, no. 2, p. 98–104. ISSN 0027-8424.

[14] БРАУНШТЕЙН, АЕ, ed. *Молекулярная биология. Проблемы и перспективы*. Moscow: Наука, 1964.

[15] GUNASEKARAN, K., Buyong MA and Ruth NUSSINOV. Is allostery an intrinsic property of all dynamic proteins? *Proteins* [online]. 2004, 57, no. 3, p. 433–443. ISSN 1097-0134. Available at: doi:10.1002/prot.20232

[16] ZÁVODSZKY, Péter and István HAJDÚ. Evolution of the concept of conformational dynamics of enzyme functions over half of a century: A personal view. *Biopolymers* [online]. 2013, 99, no. 4, p. 263–269. ISSN 1097-0282. Available at: doi:10.1002/bip.22159

[17] LINEWEAVER, Hans and Dean BURK. The Determination of Enzyme Dissociation Constants. *Journal of the American Chemical Society* [online]. 1934, 56, no. 3, p. 658–666. ISSN 0002-7863. Available at: doi:10.1021/ja01318a036

[18] DHAMI, Namraj. Trends in Pharmacognosy: A modern science of natural medicines. *Journal of Herbal Medicine* [online]. 2013, 3, no. 4, p. 123–131. ISSN 2210-8033. Available at: doi:10.1016/j.hermed.2013.06.001

[19] CIOCIOLA, Arthur A., Lawrence B. COHEN, Prasad KULKARNI, Costas KEFALAS, Alan BUCHMAN, Carol BURKE, Tedd CAIN, Jason CONNOR, Eli D. EHRENPREIS, John FANG, Ronnie FASS, Robyn KARLSTADT, Dan PAMBIANCO, Joseph PHILLIPS, Mark POCHAPIN, Paul POCKROS, Philip SCHOENFELD and Raj VUPPALANCHI. How Drugs are Developed and Approved by the FDA: Current Process and Future Directions. *The American Journal of Gastroenterology* [online]. 2014, 109, no. 5, p. 620–623. ISSN 0002-9270. Available at: doi:10.1038/ajg.2013.407

[20] BAN, Thomas A. The role of serendipity in drug discovery. *Dialogues in Clinical Neuroscience*. 2006, 8, no. 3, p. 335–344. ISSN 1294-8322.

[21] HARGRAVE-THOMAS, Emily, Bo YU and Jóhannes REYNISSON. Serendipity in anticancer drug discovery. *World Journal of Clinical Oncology* [online]. 2012, 3, no. 1, p. 1–6. ISSN 2218-4333. Available at: doi:10.5306/wjco.v3.i1.1

[22] SHAW, Leslie M. *The Clinical Toxicology Laboratory: Contemporary Practice of Poisoning Evaluation*. B.m.: Amer. Assoc. for Clinical Chemistry, 2001. ISBN 978-1-890883-53-9.

[23] VARGESSON, Neil. Thalidomide-induced teratogenesis: history and mechanisms. *Birth Defects Research. Part C, Embryo Today: Reviews* [online]. 2015, 105, no. 2, p. 140–156. ISSN 1542-9768. Available at: doi:10.1002/bdrc.21096

[24] WILLIAMS, KJ. The introduction of 'chemotherapy' using arsphenamine – the first magic bullet. *Journal of the Royal Society of Medicine* [online]. 2009, 102, no. 8, p. 343–348. ISSN 0141-0768. Available at: doi:10.1258/jrsm.2009.09k036

[25] OTTEN, H. Domagk and the development of the sulphonamides. *Journal of Antimicrobial Chemotherapy* [online]. 1986, 17, no. 6, p. 689–690. ISSN 0305-7453, 1460-2091. Available at: doi:10.1093/jac/17.6.689

[26] DE CLERCQ, Erik. The acyclic nucleoside phosphonates from inception to clinical use: historical perspective. *Antiviral Research* [online]. 2007, 75, no. 1, p. 1–13. ISSN 0166-3542. Available at: doi:10.1016/j.antiviral.2006.10.006

[27] BERMAN, Helen M., John WESTBROOK, Zukang FENG, Gary GILLILAND, T. N. BHAT, Helge WEISSIG, Ilya N. SHINDYALOV and Philip E. BOURNE. The Protein Data Bank. *Nucleic Acids Research* [online]. 2000, 28, no. 1, p. 235–242. ISSN 0305-1048, 1362-4962. Available at: doi:10.1093/nar/28.1.235

[28] KUKOL, Andreas, ed. *Molecular modeling of proteins*. Totowa, NJ: Humana Press, 2008. Methods in molecular biology, 443. ISBN 978-1-58829-864-5.

[29] DUNN, Michael F. Protein–Ligand Interactions: General Description. In: *eLS* [online]. B.m.: John Wiley & Sons, Ltd, 2001 [vid. 2016-05-12]. ISBN 978-0-470-01590-2. Avai-

lable
at: http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0001340.pub2/abstract

[30] ERTL, Peter. Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *Journal of Chemical Information and Computer Sciences* [online]. 2003, 43, no. 2, p. 374–380. ISSN 0095-2338. Available at: doi:10.1021/ci0255782

[31] DUFFY, Bryan C., Lei ZHU, Hélène DECORNEZ and Douglas B. KITCHEN. Early phase drug discovery: cheminformatics and computational techniques in identifying lead series. *Bioorganic & Medicinal Chemistry* [online]. 2012, 20, no. 18, p. 5324–5342. ISSN 1464-3391. Available at: doi:10.1016/j.bmc.2012.04.062

[32] SLIWOSKI, Gregory, Sandeepkumar KOTHIWALE, Jens MEILER and Edward W. LOWE. Computational methods in drug discovery. *Pharmacological Reviews* [online]. 2014, 66, no. 1, p. 334–395. ISSN 1521-0081. Available at: doi:10.1124/pr.112.007336

[33] RAMACHANDRAN, K. I., G. DEEPA and K. NAMBOORI. *Computational chemistry and molecular modeling: principles and applications*. Berlin: Springer, 2008. ISBN 978-3-540-77302-3.

[34] *November 2015 | TOP500 Supercomputer Sites* [online]. 2015 [vid. 2016-05-13]. Available at: http://www.top500.org/lists/2015/11/

[35] SNYDER, Phillip W., Matthew R. LOCKETT, Demetri T. MOUSTAKAS and George M. WHITESIDES. Is it the shape of the cavity, or the shape of the water in the cavity? *The European Physical Journal Special Topics* [online]. 2014, 223, no. 5, p. 853–891. ISSN 1951-6355, 1951-6401. Available at: doi:10.1140/epjst/e2013-01818-y

[36] LIPINSKI, Christopher A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technologies* [online]. 2004, 1, no. 4, p. 337–341. ISSN 1740-6749. Available at: doi:10.1016/j.ddtec.2004.11.007

[37] LASKOWSKI, R. A. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *Journal of Molecular Graphics*. 1995, 13, no. 5, p. 323–330, 307–308. ISSN 0263-7855.

[38] HENDLICH, M., F. RIPPMANN and G. BARNICKEL. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *Journal of Molecular Graphics & Modelling*. 1997, 15, no. 6, p. 359–363, 389. ISSN 1093-3263.

[39] LIANG, J., H. EDELSBRUNNER and C. WOODWARD. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Science: A Publication of the Protein Society* [online]. 1998, 7, no. 9, p. 1884–1897. ISSN 0961-8368. Available at: doi:10.1002/pro.5560070905

[40] LE GUILLOUX, Vincent, Peter SCHMIDTKE and Pierre TUFFERY. Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics* [online]. 2009, 10, p. 168. ISSN 1471-2105. Available at: doi:10.1186/1471-2105-10-168

[41] TRIPATHI, Ashutosh and Glen E. KELLOGG. A novel and efficient tool for locating and characterizing protein cavities and binding sites. *Proteins* [online]. 2010, 78, no. 4, p. 825–842. ISSN 1097-0134. Available at: doi:10.1002/prot.22608

[42] CONNOLLY, M. L. Shape distributions of protein topography. *Biopolymers* [online]. 1992, 32, no. 9, p. 1215–1236. ISSN 0006-3525. Available at: doi:10.1002/bip.360320911

[43] KLEYWEGT, G. J. and T. A. JONES. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallographica. Section D, Biological Crystallography* [online]. 1994, 50, no. Pt 2, p. 178–185. ISSN 0907-4449. Available at: doi:10.1107/S0907444993011333

[44] REYNOLDS, C. A., R. C. WADE and P. J. GOODFORD. Identifying targets for bioreductive agents: using GRID to predict selective binding regions of proteins. *Journal of Molecular Graphics*. 1989, 7, no. 2, p. 103–108, 100. ISSN 0263-7855.

[45] KORTVELYESI, Tamas, Sheldon DENNIS, Michael SILBERSTEIN, Lawrence BROWN and Sandor VAJDA. Algorithms for computational solvent mapping of proteins. *Proteins* [online]. 2003, 51, no. 3, p. 340–351. ISSN 1097-0134. Available at: doi:10.1002/prot.10287

[46] LAURIE, Alasdair T. R. and Richard M. JACKSON. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics (Oxford, England)* [online]. 2005, 21, no. 9, p. 1908–1916. ISSN 1367-4803. Available at: doi:10.1093/bioinformatics/bti315

[47] WEISEL, Martin, Ewgenij PROSCHAK and Gisbert SCHNEIDER. PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chemistry Central Journal* [online]. 2007, 1, p. 7. ISSN 1752-153X. Available at: doi:10.1186/1752-153X-1-7

[48] SCHMIDTKE, Peter, Axel BIDON-CHANAL, F. Javier LUQUE and Xavier BARRIL. MDpocket : Open Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* [online]. 2011, p. btr550. ISSN 1367-4803, 1460-2059. Available at: doi:10.1093/bioinformatics/btr550

[49] RAUNEST, Martin and Christian KANDT. dxTuber: detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *Journal of Molecular Graphics & Modelling* [online]. 2011, 29, no. 7, p. 895–905. ISSN 1873-4243. Available at: doi:10.1016/j.jmgm.2011.02.003

[50] CHOVANCOVA, Eva, Antonin PAVELKA, Petr BENES, Ondrej STRNAD, Jan BREZOVSKY, Barbora KOZLIKOVA, Artur GORA, Vilem SUSTR, Martin KLVANA, Petr MEDEK, Lada BIEDERMANNOVA, Jiri SOCHOR and Jiri DAMBORSKY. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS computational biology* [online]. 2012, 8, no. 10, p. e1002708. ISSN 1553-7358. Available at: doi:10.1371/journal.pcbi.1002708

[51] SHULMAN-PELEG, Alexandra, Ruth NUSSINOV and Haim J. WOLFSON. Recognition of functional sites in protein structures. *Journal of Molecular Biology* [online]. 2004, 339, no. 3, p. 607–633. ISSN 0022-2836. Available at: doi:10.1016/j.jmb.2004.04.012

[52] DELANO, Warren L. Unraveling hot spots in binding interfaces: progress and challenges. *Current Opinion in Structural Biology*. 2002, 12, no. 1, p. 14–20. ISSN 0959-440X.

[53] HUANG, Bingding. MetaPocket: a meta approach to improve protein ligand binding site prediction. *Omics: A Journal of Integrative Biology* [online]. 2009, 13, no. 4, p. 325–330. ISSN 1557-8100. Available at: doi:10.1089/omi.2009.0045

[54] AURENHAMMER, Franz. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*. 1991, 23, no. 3, p. 345–405.

[55] LUKAT, Gunther, Jens KRÜGER and Björn SOMMER. APL@Voro: a Voronoi-based membrane analysis tool for GROMACS trajectories. *Journal of Chemical Information and Modeling* [online]. 2013, 53, no. 11, p. 2908–2925. ISSN 1549-960X. Available at: doi:10.1021/ci400172g

[56] TROTT, Oleg and Arthur J. OLSON. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* [online]. 2010, 31, no. 2, p. 455–461. ISSN 1096-987X. Available at: doi:10.1002/jcc.21334

[57] JENSEN, Frank. *Introduction to computational chemistry*. 2nd ed. Chichester, England ; Hoboken, NJ: John Wiley & Sons, 2007. ISBN 978-0-470-01186-7.

[58] LEACH, Andrew R. *Molecular modelling: principles and applications*. 2nd ed. Harlow, England ; New York: Prentice Hall, 2001. ISBN 978-0-582-38210-7.

[59] ANDRUSIER, Nelly, Efrat MASHIACH, Ruth NUSSINOV and Haim J. WOLFSON. Principles of Flexible Protein-Protein Docking. *Proteins* [online]. 2008, 73, no. 2, p. 271–289. ISSN 0887-3585. Available at: doi:10.1002/prot.22170

[60] ALVAREZ-GARCIA, Daniel and Xavier BARRIL. Relationship between Protein Flexibility and Binding: Lessons for Structure-Based Drug Design. *Journal of Chemical Theory and Computation* [online]. 2014, 10, no. 6, p. 2608–2614. ISSN 1549-9618. Available at: doi:10.1021/ct500182z

[61] DOLEZAL, Rafael, Jan KORABECNY, David MALINAK, Jan HONEGR, Kamil MUSILEK and Kamil KUCA. Ligand-based 3D QSAR analysis of reactivation potency of mono- and bis-pyridinium aldoximes toward VX-inhibited rat acetylcholinesterase. *Journal of Molecular Graphics & Modelling* [online]. 2015, 56, p. 113–129. ISSN 1873-4243. Available at: doi:10.1016/j.jmgm.2014.11.010

[62] RIPPHAUSEN, Peter, Britta NISIUS, Lisa PELTASON and Jürgen BAJORATH. Quo vadis, virtual screening? A comprehensive survey of prospective applications. *Journal of Medicinal Chemistry* [online]. 2010, 53, no. 24, p. 8461–8467. ISSN 1520-4804. Available at: doi:10.1021/jm101020z

[63] PALOMO, Valle, Ignacio SOTERAS, Daniel I. PEREZ, Concepción PEREZ, Carmen GIL, Nuria Eugenia CAMPILLO and Ana MARTINEZ. Exploring the binding sites of glycogen synthase kinase 3. Identification and characterization of allosteric modulation cavities. *Journal of Medicinal Chemistry* [online]. 2011, 54, no. 24, p. 8461–8470. ISSN 1520-4804. Available at: doi:10.1021/jm200996g

[64] NUSSE, Roel. Wnt signaling in disease and in development. *Cell Research* [online]. 2005, 15, no. 1, p. 28–32. ISSN 1001-0602. Available at: doi:10.1038/sj.cr.7290260

[65] GOLPICH, Mojtaba, Elham AMINI, Fatemeh HEMMATI, Norlinah Mohamed IBRAHIM, Behrouz RAHMANI, Zahurin MOHAMED, Azman Ali RAYMOND, Leila DARGAHI, Rasoul GHASEMI and Abolhassan AHMADIANI. Glycogen synthase kinase-3 beta (GSK-3β) signaling: Implications for Parkinson's disease. *Pharmacological Research* [online]. 2015, 97, p. 16–26. ISSN 1096-1186. Available at: doi:10.1016/j.phrs.2015.03.010

[66] MEIJER, Laurent, Marc FLAJOLET and Paul GREENGARD. Pharmacological inhibitors of glycogen synthase kinase 3. *Trends in Pharmacological Sciences* [online]. 2004, 25, no. 9, p. 471–480. ISSN 0165-6147. Available at: doi:10.1016/j.tips.2004.07.006

[67] DAJANI, R., E. FRASER, p. M. ROE, N. YOUNG, V. GOOD, T. C. DALE and L. H. PEARL. Crystal structure of glycogen synthase kinase 3 beta: structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell*. 2001, 105, no. 6, p. 721–732. ISSN 0092-8674.

[68] BAX, Benjamin, Paul p. CARTER, Ceri LEWIS, Angela R. GUY, Angela BRIDGES, Robert TANNER, Gary PETTMAN, Chris MANNIX, Ainsley A. CULBERT, Murray J. B. BROWN, David G. SMITH and Alastair D. REITH. The Structure of Phosphorylated GSK-3β Complexed with a Peptide, FRATtide, that Inhibits β-Catenin Phosphorylation. *Structure* [online]. 2001, 9, no. 12, p. 1143–1152. ISSN 0969-2126. Available at: doi:10.1016/S0969-2126(01)00679-7

[69] LE GUILLOUX, Vincent, Peter SCHMIDTKE and Pierre TUFFERY. *Fpocket Users' manual* [online]. 2010 [vid. 2016-05-17]. Available at: http://fpocket.sourceforge.net/manual_fpocket2.pdf

[70] MARTINEZ, Ana, Mercedes ALONSO, Ana CASTRO, Isabel DORRONSORO, J. Luis GELPÍ, F. Javier LUQUE, Concepción PÉREZ and Francisco J. MORENO. SAR and 3D-QSAR Studies on Thiadiazolidinone Derivatives: Exploration of Structural Requirements for Glycogen Synthase Kinase 3 Inhibitors. *Journal of Medicinal Chemistry* [online]. 2005, 48, no. 23, p. 7103–7112. ISSN 0022-2623, 1520-4804. Available at: doi:10.1021/jm040895g

Appendix

Contents of enclosed DVD-R

`Docking`	contains data relevant to docking
`input structures`	contains structures used for docking experimental part
`4NM0`	structures of 4NM0 prepared for docking
`4PTE`	structures of 4PTE prepared for docking
`Ligands`	structures of 12 ligands prepared for docking
`results_4nm0`	output from docking 4NM0 with 12 ligands
`results_4pte`	output from docking 4PTE with 12 ligands
`Images`	contains images created for the thesis
`Pockets`	contains data for and output from pocket calculation
`4NM0_monomer`	4NM0 input and output for pocket search
`4PTE_monomer`	4PTE input and output for pocket search
`Melikova BP text.pdf`	electronic version of the thesis