UNIVERZITA KARLOVA V PRAZE

Přírodovědecká fakulta

DIPLOMOVÁ PRÁCE

Vladimír Palivec

Počítačové modelování interakcí iontů s proteiny: Allosterický efekt iontů a fenolických ligandů na strukturu insulinového hexameru

Katedra fyzikální a makromolekulární chemie PřF UK

Ústav organické chemie a biochemie, AV ČR, v.v.i.

Vedoucí diplomové práce: prof. RNDr. Pavel Jungwirth, DSc.

Studijní program: Modelování chemických vlastností nano- a biostruktur

2016

CHARLES UNIVERSITY IN PRAGUE

Faculty of Science

DIPLOMA THESIS

Vladimír Palivec

Computer modeling of ion protein interactions: Allosteric effects of phenolic ligands and ions on insulin hexamer structure

Department of Physical and Macromolecular Chemistry

Institute of Organic Chemistry and Biochemistry, AS CR, v.v.i

Thesis advisor: prof. RNDr. Pavel Jungwirth, DSc.

Specialization: Modeling of Chemical Properties of Nano- and Biostructures

2016

Science is a wonderful thing providing us with many opportunities. It is truly miraculous and I could go on to fully describe its full potential. However, I wouldn't have achieved anything without happiness, love and support I get from my family and friends. They are the reason why I finished this thesis as they ignite my passion for science and, more importantly, for life. Thank you all.

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 4. 5. 2016

Vladimír Palivec

Název práce: Počítačové modelování interakcí iontů s proteiny: Allosterický efekt iontů a fenolických ligandů na strukturu insulinového hexameru

Autor: Vladimír Palivec

Katedra: Katedra fyzikální a makromolekulární chemie PřF UK

Vedoucí diplomové práce: prof. RNDr. Pavel Jungwirth, DSc., UOCHB AV ČR, v.v.i.

Email vedoucího: pavel.jungwirth@uochb.cas.cz

Abstrakt: Insulinový hexamer je alosterický protein, který můžeme nalézt ve třech různých konformacích (T6, T3R3, R6). Tvorbu a konformaci inzulinového hexameru můžeme ovlivnit například vazbou iontů, či takzvaných fenolických ligandů. V této práci jsem zkoumal inzulinové hexamery v různých konformacích za pomoci molekulové dynamiky. Studie je rozdělena na dvě části. V první části zkoumám efekt vazby kationtů (Zn^{2+}, Ca^{2+}, K^{+} and Na^{+}) na T6 a T3R3 inzulinové hexamery. V druhé části se zaměřím na neurotransmitery serotonin a dopamin, které by mohly sloužit jako fenolické ligandy v in vivo podmínkách. Výsledky první části výzkumu ukazují, že ionty s vysokou nábojovou hustotou (Zn^{2+}, Ca^{2+}) jsou mnohem více lokalizované v kavitě, která se nachází uprostřed hexameru. Toto vede ke zpomalení difuze vodních molekul, což se projeví také tím, že uvnitř se nachází vždy pouze jeden kation. Monovalentní kationty tento efekt nemají. V druhé části práce ukáži, že z obou uvažovaných neurotransmiterů je pouze serotonin slibným fenolickým ligandem. Dále jsem pro neurotransmitery našel nová, dosud neznámá vazebná místa. Dopamin se na tyto vazebná místa váže nejsilněji. Nakonec porovnám všechny teoretické výsledky s experimentální prací našich spolupracovníků J. Jiráčka et al. a M. Brzozowskeho et al.

Klíčová slova: molekulová dynamika, empirické potenciály, insulinový hexamer, fenolický ligand, fenol, neurotransmitery, dopamin, serotonin, ionty, zinek, vápník

Title: Computer modeling of ion protein interactions: Allosteric effects of phenolic ligands and ions on insulin hexamer structure

Author: Vladimír Palivec

Department: Department of Physical and Macromolecular Chemistry Faculty of Science UK

Advisor: prof. RNDr. Pavel Jungwirth, DSc., IOCB AS CR, v.v.i.

Advisor's email address: pavel.jungwirth@uochb.cas.cz

Abstract: Insulin hexamer is an allosteric protein capable of undergoing conformational changes between three states: T6, T3R3, and R6. Transitions between them, as well as the formation of insulin hexamers, are mediated through binding of phenolic ligands or ions. This thesis presents a molecular dynamics study of allosteric behavior of insulin using empirical force fields. Two effects are closely inspected – cation (Zn^{2+}, Ca^{2+}, K^{+}, and Na^{+}) binding to the insulin hexamers and a possible binding of two neurotransmitters – dopamine and serotonin to the phenolic pocket. The results show that high charge density cations (Zn^{2+} and Ca^{2+}) are mostly localized in the B13 glutamate cavity, slow-down diffusion, while preventing other cations from entering. In contrast, low charge density cations (Na^{+} and K^{+}) do not have this effect. Concerning neurotransmitters, dopamine does not bind to the phenolic pocket whereas serotonin binds in a similar way like phenol. During an investigation of the phenolic pocket, previously unknown binding pockets for neurotransmitters were found on the surface of hexamer. These pockets were closely inspected and it was found that among the investigated species, dopamine binds the strongest. The present computational results are supported by experimental evidence based on protein crystallography and biochemical essays (collaborators J. Jiracek et al. and M. Brzozowski et al.).

Keywords: molecular dynamics, empirical potentials, insulin hexamer, phenolic ligand, phenol, neurotransmitters, dopamine, serotonin, ions, zinc, calcium

CONTENT

Chapter 1

Introduction

This chapter introduces the topic of insulin and its connection to ions and neurotransmitters. I am going to make an attempt to go through the story of this remarkable hormone while introducing all important parts of an insulin molecule. There are still many unknowns which I am going to emphasize during this rather brief introduction. They will be summarized at the end of this chapter.

Human insulin is a signaling peptide consisting of 51 amino acids. It is composed of two chains linked together by one intramolecular and two intermolecular disulfide bridges as shown in Figure 1.1.

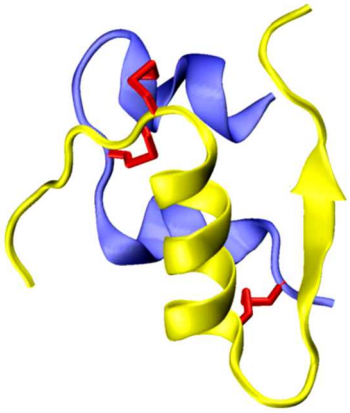

Figure 1.1. Insulin monomer composed of a chain A (blue, 21 amino acids) and a chain B (yellow, 30 amino acids). Three disulfide bridges are shown as red color.

Its function is to regulate the metabolism of saccharides and fats. Insulin release into the bloodstream leads, for instance, to an absorption of glucose into the cells, synthesis of glycogen or increased lipid synthesis. Disruption of insulin secretion leads to diabetes mellitus. Treatment for this disease often includes daily intravenous injection of insulin. Since the hormone's discovery in 1922[1], it has been known that there is a

connection between zinc and insulin[2–6]. The effect of zinc was used in pharmaceutical preparations, although the exact effect was not known. Later, the connection between the additives in insulin medicaments, like antimicrobial preservatives (phenol, m-cresol, chloride, or acetate) and physiochemical stability was discovered. During testing of different substances, it was shown that these additives in insulin medicaments can prolong its desired effect. Since then, it has been shown in many studies that the insulin – zinc complex is an allosteric protein. Due to many experimental methods such as NMR, X-ray, crystallography and other in order to investigate this problem, we now know that insulin forms dimers which form insulin hexamers[7–10] in the presence of zinc. These hexamers essentially include three conformational states abbreviated as T6, T3R3, and R6 as shown in Figure 1.2.

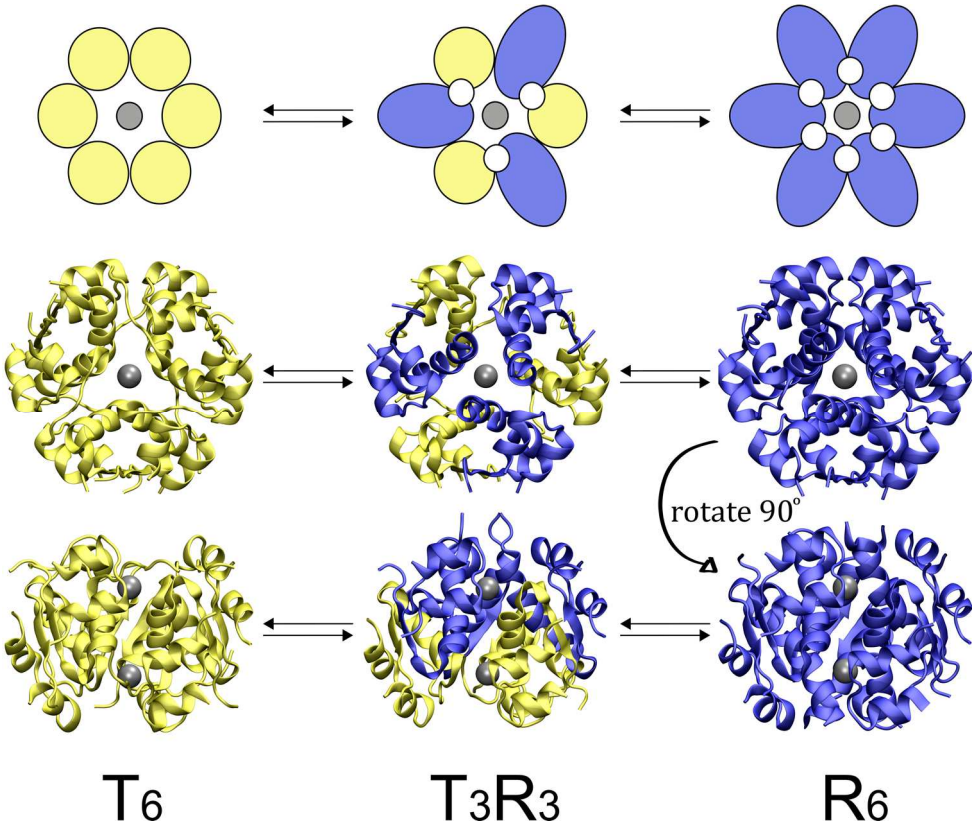

Figure 1.2. The scheme illustrates insulin hexamer conversions between T6 – T3R3 – R6 conformations. The conversion of insulin monomer from the T (yellow) to the R (blue) conformation is known to form one phenolic pocket (white circle). Grey dots/atoms depict zinc cations.

The conversion between these three states is mediated through binding of phenolic ligands and ions. In insulin hexamer, two zinc cations lie on a three-fold symmetry axis coordinated by three B10 histidines from each trimeric unit (this will be discussed later). Upon the change of the conformation from a T to an R state, the first 9 residues of the B chain undergo a transition from an extended to an alpha-helix state (Figure 1.3.). This conformational change from the T to the R state opens a phenolic pocket. The details of the phenolic pocket will be also discussed later.

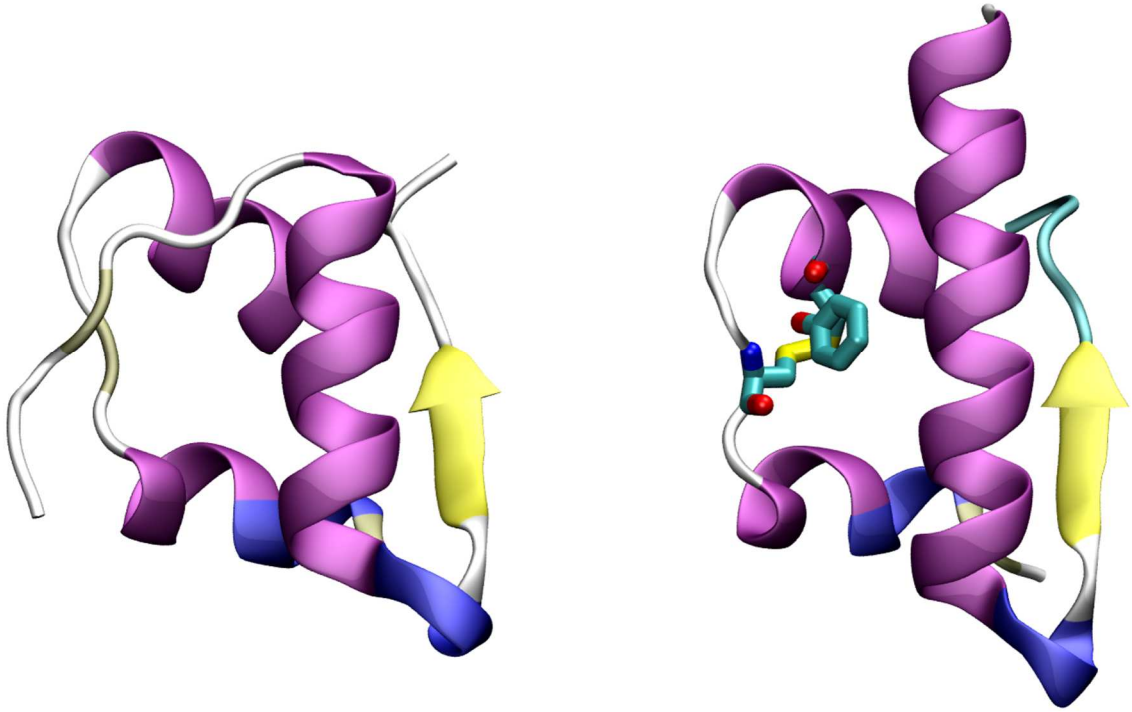

Figure 1.3. The pictures depict a comparison of the T conformation of insulin monomer (left) and the R conformation (right). In the case of R conformation, a phenol and amino acids crucial for binding of phenol (A6 and A11) are shown. Note that both monomers were extracted from hexamer structure and therefore, an artificial beta sheet structure can be seen.

The T6 abbreviation corresponds to a state where all of the monomeric units are in the T state whereas the R6 state corresponds to a state where all of the monomeric units are in the R state while the T3R3 state is an intermediate state between the two. Many in vitro studies have been performed on insulin hexamers. Crystallization of insulin in the presence of various cations lead, for instance, to Ni^{2+}, Co^{2+}, Cu^{2+}, and Cd^{2+} hexamers[11–15]. Apart from cations found not only in the B10 site, an additional binding site for cations

was discovered as well. This site is located in the middle of the hexamer between two B10 zincs where six B13 glutamates can be found. The glutamates are relatively close to each other while being positioned in a circle around the center of a ~10 Å diameter cavity, hence essentially forming a perfect trap for cations. A detailed picture of the middle part of hexamer is shown in Figure 1.4.

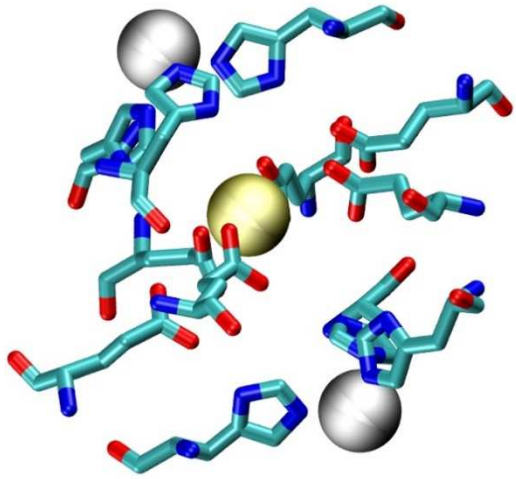

Figure 1.4. The B10 histidines and B13 glutamates of insulin hexamer are shown. These amino acids play a critical role in the insulin hexamer structure. B10 histidines are binding zinc cations (grey van der Walls balls) while B13 glutamates can bind various cations[16–19] which is depicted by a yellow van der Waals ball.

Using crystallography and NMR studies, it has been shown that Cd^{2+} binds to two B13 glutamates in three different conformations each with 0.33 occupancy. Same binding is predicted for the Ca^{2+} cation[16–19]. Binding of divalent cations to the B13 site has been shown to be contributory to insulin hexamer formation[20].

There are also studies which were focused on the effect of phenolic ligands. Many different substances such as phenol, m-cresol, hexanol, acetate, resorcinol, and other[7,21–23] have been considered. From these, the phenol and resorcinol were discovered to be the most effective at directing the T conformation toward the R state. It was found that not only phenolic ligands can drive insulin hexamer toward the R state, but also that anions do by binding to the B10 zinc cations. Many different anions were investigated, ranging

from inorganic Cl^-, SCN^- to organic acids like p-aminobenzoate[21,24]. Among these, the thiocyanate (SCN^-) was found to be the most effective. Crystallization of insulin in the presence of SCN^- and zinc lead to T3R3 insulin hexamer. It has been shown that R6 hexamer is far more stable than T3R3 hexamer and that T3R3 hexamer is more stable than T6 hexamer[25,26]. However, every experiment which was mentioned was done in vitro and little is known about the actual conformation/composition of insulin hexamers inside the secretory granules of beta-cells.

The story of insulin molecule starts in pancreas, or more specifically in the beta-cells of islets of Langerhans (Figure 1.5.) on ribozymes where a preproinsulin is synthesized.

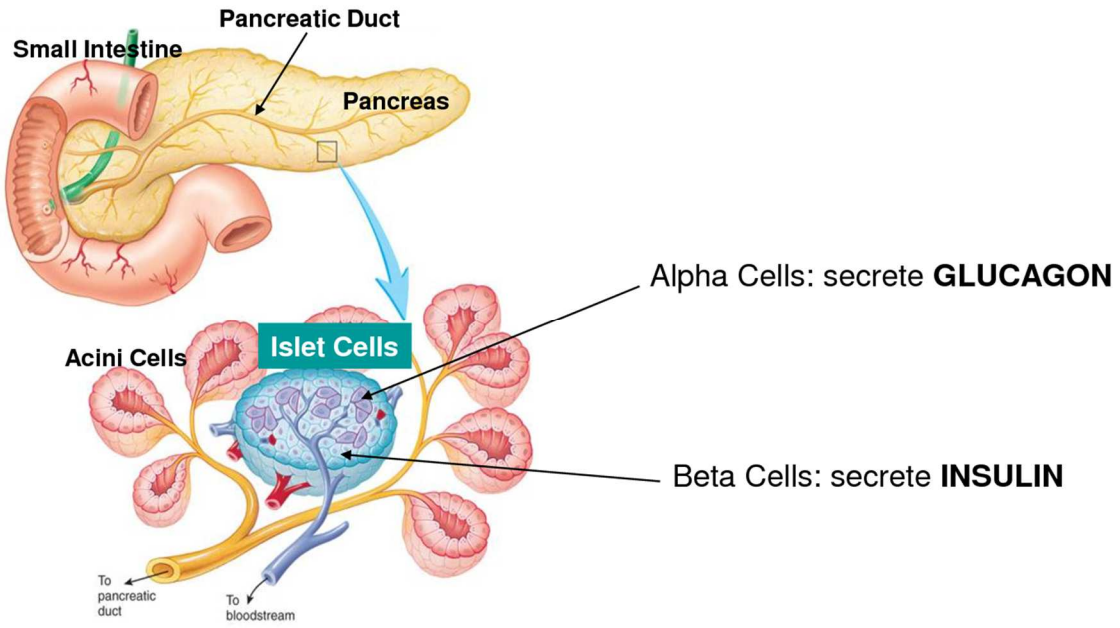

Figure 1.5. Location of islets of Langerhans in human body. Source: Review Nursing

The preproinsulin consists of insulin, a C peptide, and a signaling peptide at the N terminus. The preproinsulin is then transported to endoplasmic reticulum, where the signaling peptide is cleaved, yielding a proinsulin. Subsequently, the proinsulin is transported to the secretory granule which is the final storage place for insulin. The granules are loaded with zinc[27–29] and the environment is mildly acidic (pH ~ 5.5)[30,31]. There, in the granules, the proinsulin forms proinsulin hexamers in the presence of zinc cations. After the formation of proinsulin hexamers, the C peptide is cleaved, yielding

insulin hexamers. Since the pH is ~5.5, which is also an isoelectric point of insulin, the insulin hexamers start to precipitate forming a solid crystal[32], which reduces the susceptibility of insulin to enzymes[33]. The secretory granules are densely packed with insulin content of about 50-60%[34] (Figure 1.6.). However, no one knows the exact stored form of insulin hexamers.

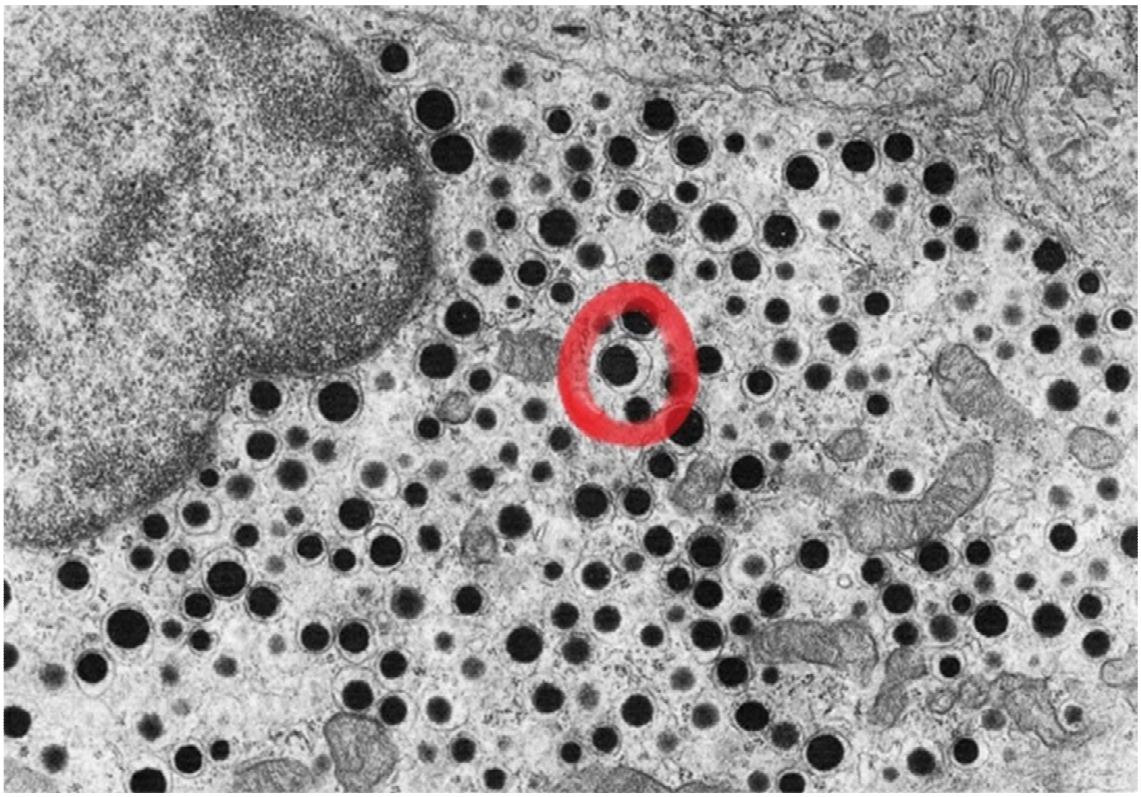

Figure 1.6. A beta cell containing high amount of secretory granules. One of the granules is marked by a red circle. Source: http://images.nigms.nih.gov/

While insulin is stored as hexamer, the active species is a monomer. When a secretory granule merges with membrane and the insulin gets released to the extracellular milieu, the insulin hexamers starts to break down into monomers[35]. It has been shown that the insulin crystal dissolves at different rates[36], which directly leads to an idea that the difference may be caused by a mixture of T6/T3R3/R6 hexamers. It has been known for many years that the beta-cells contained high amount of Zn, Ca and S elements[37–39] and that there is also a high content of phosphates, ATP. and neurotransmitters[40–43] (serotonin, dopamine, L-DOPA). There is strong evidence that the beta-cells are connected with the neurotransmitters[44–48]. For instance, there are transporters VMAT1

and VMAT2 which cumulate biogenic amines inside these cells[49]. Dopamine has been shown to be stored and co-secreted with insulin[50,51]. The beta-cells have also been shown to accumulate the previously mentioned neurotransmitters[52–55]. We shall take a closer look at the structure of phenolic pocket with a phenol molecule bound (Figure 1.7.)

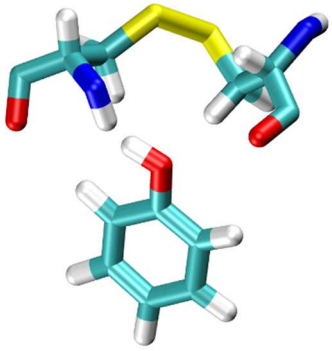

Figure 1.7. Phenol molecule inside the phenolic pocket. Strong binding is mediated through hydroxyl group of the phenol molecule. Only A6 and A11 amino acids are shown.

and compare the chemical structure of dopamine and serotonin with phenols (Figure 1.8.),

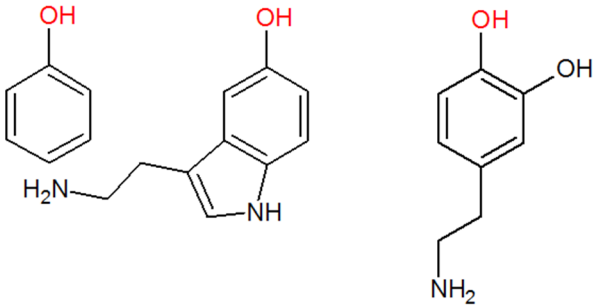

Figure 1.8. Phenol (left) and two neurotransmitters serotonin (middle) and dopamine (right) which could serve as phenolic ligands instead of phenol. The red hydroxyl group (OH) is crucial for binding.

we can acknowledge that all these substances have a crucial hydroxyl group required for binding to the phenolic pocket. In the light of these facts, might it be possible that the neurotransmitters could substitute phenol as a phenolic ligand in in vivo conditions?

During the introduction, we have stumbled upon several interesting questions which remained unanswered up to this point. In order to begin, there are a few publications which examined the inner B13 cavity, however, only using cadmium. Of course, the experimental investigation of the cavity is rather difficult and it has not been done yet to my knowledge. Molecular dynamics is a suitable method for tackling this experimentally elusive problem. Therefore, the first problem I investigated was binding of calcium and other cations to the insulin hexamers and their respective effects. Moreover, I have proposed an idea that dopamine and serotonin neurotransmitters work in secretory granules as phenolic ligands. This is thus the second question I will try to answer in this thesis.

Chapter 2

Computational techniques

1. Theory and methods

As mentioned at the end of the last chapter, the research was done with the usage of classical molecular dynamics simulations. This chapter therefore discusses all general concepts and methods of molecular dynamics as well as the quantum mechanical methods used throughout this work.

1.1. Molecular dynamics

This section covers basic theories and technical tricks from classical molecular mechanics which have been applied throughout this thesis. At first a concept of computer experiment is introduced which is then discussed more thoroughly in a form of molecular dynamics simulation. The notations and theoretical framework were taken from Frenkel, D, Smit, B.[56], Jungwirth, P.[57], Allen, M. P., Tildesley, D. J.[58].

1.1.1. General concepts

There is only one real world, however, there are many ways to model it. If we choose to model our system with every detail, a quantum mechanical description has to be used. If we need to study bigger systems, we have to make some approximations. One way of doing this is to describe all particles just by classical mechanics. Computer simulations use theories of statistical mechanics which were been developed long time ago, but computers allowed us to use them thoroughly later. In classical computer experiment, e.g. molecular dynamics simulation, one obtains positions and velocities of

every particle. This information, of course, is not accessible by any present experiment. What gets usually measured is an average property of the studied system, e.g. its density. This real experiment is done in a finite amount of time and the studied property is averaged over many atoms/molecules. Similar thing can be done during a computer experiment. If system is ergodic then the average property of the system is just the time average over the trajectory of our molecular dynamics simulation. We can achieve certain condition such as constant volume during a computer experiment which is also similar to the real world. According to the met condition, we distinguish for example following ensembles: NpT, NVT, NVE, or μVT. In molecular dynamics, the pressure (NpT) and constant volume (NVT) ensembles are most commonly used constants.

1.1.2. Molecular dynamics – basic ideas

In a molecular dynamics simulation, all atoms move according to the laws of classical mechanics. This is usually an excellent approximation for a wide range of conditions. All atoms are described as balls of certain size with a partial charge. Any microstate is fully described by positions and momenta of all atoms and the system evolves according to Newton's equations of motion. Second Newton's law reads

$$\vec{F} = m\vec{a}, \tag{2.1}$$

which connects force $\vec{F}$ with mass m and acceleration $\vec{a}$. When the forces acting on all particles are known, we can integrate this equation to obtain future position of all atoms. Many algorithms have been proposed, although one usually uses a Verlet algorithm

$$r(t + \Delta t) \approx 2r(t) - r(t - \Delta t) + \frac{f(t)}{m}\Delta t^2, \tag{2.2}$$

where $r(t)$ stands for a position, Δt is the time step in our molecular dynamics simulation and $f(t)$ is a force. The force is computed in a following way

$$f(t) = \left(\frac{\partial V(r)}{\partial r}\right)_t, \tag{2.3}$$

Where we introduced an interaction potential energy $V(r)$. This potential energy is divided into several terms including intermolecular (non–bonded interactions) terms and intramolecular terms (bonded interactions). The overall potential can be, and usually is, constructed as follows

$$\begin{aligned} V(r^N) = & \sum_{bonds} k_b(r - r_{eq})^2 + \sum_{angles} k_a(\theta - \theta_{eq})^2 \\ & + \sum_{dihedral\ angles} \sum_n \frac{1}{2} k_n(1 + \cos(n\varphi - \omega) \\ & + \sum_{i<j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon r_{ij}}\right], \end{aligned} \tag{2.4}$$

where the first three terms describe bonded interactions while the last term describes non-bonded interactions composed of van der Waals and electrostatic energies. The first term involves summation of all bonds and it takes form of a simple harmonic spring described by a bond force constant k_b. Similarly to the first term, the second term describes an angular dependence of energy. A simple harmonic description is usually sufficient. The third term represents energy of twisting bonds and it is described by Fourier series. The fourth term describes all non-bonded interactions composed of van der Waals and electrostatic terms. The van der Waals term usually takes form of the Lennard-Jones potential and the electrostatic term of the simple Coulombs law. A set of these parameters which describe all interactions between atoms is called a force field. It is derived from the best fit to experimental or ab initio data.

1.1.3. Temperature and pressure coupling

One is usually interested in constant temperature or/and constant pressure simulation. The constant temperature ensemble can be achieved by using so-called *thermostats*. Many thermostats have been introduced and they can be divided into three bundles. The first of them simply rescales the velocities (e.g. velocity rescale or the Berendsen thermostat). Furthermore, we can regulate the temperature by introducing stochastic forces (e.g. the Andersen or Langevin thermostat). The last one uses extended Lagrangian formalisms and the Nosé-Hoover thermostat could serve as an example. The pressure coupling can be conducted in a similar way to the temperature. In analogy to the Berendsen thermostat, we can rescale not temperature but dimensions. Another method how to control the pressure is to use the Parinello-Rahman barostat which is an analog to the Nosé-Hoover thermostat.

1.1.4. Periodic boundary conditions

Another widely used trick to suppress the fact that our system in simulation is very small in comparison with the real system is to use periodic boundary conditions (PBC). The method is used for simulating an ‘infinite’ system with limited computational time. The idea is as follows. We construct a unit cell and when an object passes through one side, it reappears on the opposite side of our unit cell. In molecular dynamics simulations, we usually use this trick in order to calculate bulk properties of a system, e.g. a protein solvated in a bath of explicit solvent. Another advantage of this method is that it saves computer time. For instance we can make simulation box bigger in order to simulate bulk water environment. However, this is not feasible even with the best computers which are available nowadays. The PBC are thus an acceptable way to do it.

1.1.5. Ewald summation

As mentioned in the last paragraph, computer time is always limited. Therefore, next of many computer tricks we introduce is a so-called *cutoff*. The non-bonded interactions take a long computation time to be calculated. We can take advantage of the knowledge that these interactions go to zero in direct relation to an increasing distance between the two considered atoms. Therefore, if we neglect all interaction after certain distance value, we can save a lot of computer time. Truncation of the van der Waals term usually does not cause any problem, although truncating electrostatic interaction leads (due to slow convergence to zero in relation to the distance) to huge errors and some correction has to be introduced. Under PBC, one can, for example, use the Particle Mesh Ewald method[59]. The trick is as follows. Under PBC, the electrostatic term has a form of an infinite sum. This sum is divided into a short-range part which is solved in real space. The rest (long-range part), which originates mostly from periodic images, is then solved in reciprocal space. A cutoff of 0.9 –1.2 nm is usually sufficient after this correction.

1.1.6. Bond constraining

In molecular dynamics simulation, one has to decide how big the time step in integrating Newton's equations should be. We would like to have the time step as big as possible. However, we still need to capture all possible motions in our system. This covers diffusion but also vibrations or rotations, from which the vibrational motions of bonds, formed by hydrogen atoms, are the fastest. In order to capture these motions, a time step of ~0.5–1 fs has to be used. Nevertheless, if we manage to constrain these bonds, we can use a time step of ~2 fs. There are several algorithms which are used in molecular dynamics. There is a SETTLE[60] algorithm which is used for water molecules. There are SHAKE and LINCS[61] algorithms which are used for all other bonds.

1.2. Quantum mechanics

As the quantum mechanical calculations are not the main goal of this work but solely a ligand parameterization, I will briefly go through this section. Firstly, the basic ideas of quantum chemistry calculations will be introduced. Then we proceed to the core of all wave function methods – the Hartree-Fock approximation. The Hartree-Fock approximation is important as this is the common method to parameterize small molecules. As we speak about parameterization, we need to obtain a reasonable geometry of our parameterized molecule. For this purpose, the density functional theory is an excellent quantum chemistry method. At the end of this section, we are finally going to gather our desired results – partial atomic charges which are part of parameterization of a small molecule in molecular dynamics simulation. Notations and theoretical descriptions were mainly taken from Szabo, A., Ostlund, N. S.[62] and Cramer, C.[63].

1.2.1. Electronic structure calculations of the ground state

The key relationship that is being solved in all quantum mechanical calculations is the non-relativistic time-independent Schrödinger equation

$$\widehat{H}|\Phi\rangle = \mathcal{E}|\Phi\rangle, \tag{2.5}$$

with $\widehat{H}$ being the Hamiltonian operator for the system of M nuclei and N electrons described by positions of vectors R_M and r_N. Φ is our unknown wave function and $\mathcal{E}$ is its corresponding energy. The Hamiltonian for N electrons and M nuclei is defined as follows

$$\widehat{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}, \tag{2.6}$$

with M_A being the mass of a nucleus, A, Z_A being the atomic number of the nucleus, A, i, j being the electron indices, and A, B being the nuclei indices. r_{ij} denotes the distance between the ith and jth electron. Similarly, r_{iA} stands for the distance between the ith electron and Ath nucleus. The distance between Ath and Bth nucleus is denoted R_{AB} in the same way. The Laplacian operators ∇_i^2 and ∇_A^2 involve differentiation with respect to the coordinates of the ith electron and Ath nucleus. The first two terms in equation 2 describe the kinetic energy of electrons and nuclei. The third term defines the attraction between electrons and nuclei. The last two terms describe repulsion between two nuclei or between two electrons.

As we said, the ultimate goal is to solve the Schrödinger equation but this is not feasible even for a system of two electrons. Therefore, approximations have to be made. One of the first and central of them is the Born-Oppenheimer approximation. Since nuclei are much heavier than electrons, we can separate the motion of electrons from nuclei. With this in mind, we arrive to the so-called electronic Hamiltonian which reads as follows

$$\hat{H}_{elec} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}}. \tag{2.7}$$

All indices here have the same meaning as in the Hamiltonian defined in equation 2.6. Nevertheless, terms which depend only on nuclei coordinates are missing. With the Born-Oppenheimer approximation, the wave function becomes parametrically dependent on nuclei coordinates. The electrons are now moving in an external field of fixed nuclei and the nuclear repulsion becomes just an additive constant to the overall energy. Following equations summarize the thoughts stated above

$$\hat{H}_{elec}|\Phi_{elec}\rangle = \mathcal{E}_{elec}|\Phi_{elec}\rangle, \tag{2.8}$$

$$\mathcal{E}_{tot} = \mathcal{E}_{elec} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}. \tag{2.9}$$

By solving time-independent electronic Schrödinger equation (Equation 2.8), we obtain the electronic wave function $\Phi_{elec}(r_i, R_A)$ with the corresponding electronic energy $\mathcal{E}_{elec}$. The fundamental problem of all ab initio methods is solving the time-independent electronic Schrödinger equation but it cannot be done analytically. Therefore approximations have to be made again, which leads us to the Hartree-Fock approximation. The key concept of the Hartree-Fock method is that we neglect instantaneous motion of electrons. By this we can treat each electron as if it was moving under an average field created by all other electrons and nuclei.

1.2.2. The Hartree-Fock approximation

To proceed any further in solving this incredibly difficult problem, one has to make a few assumptions. First of them is a wave function form. What does this beast look like? The following wave function form has been proposed

$$\Phi_{SD}(x_1, x_2, \dots, x_N) = (N!)^{-\frac{1}{2}} \begin{vmatrix} \chi_i(x_1) & \chi_j(x_1) & \dots & \chi_k(x_1) \\ \chi_i(x_2) & \chi_j(x_2) & \dots & \chi_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(x_N) & \chi_j(x_N) & \dots & \chi_k(x_N) \end{vmatrix}. \tag{2.10}$$

It has a form of a single Slater determinant Φ_{SD} of N electrons. The prefactor is just the normalization constant and χ_i are so-called molecular spin orbitals. One advantage is that this assumption ensures that the wave function will be antisymmetric. The molecular spin orbitals are further expanded into a linear combination of K known atomic orbitals ϕ_k with expansion coefficients c_{ki}.

$$\chi_i = \sum_{k=1}^{K} c_{ki}\phi_k. \tag{2.11}$$

A set of known atomic orbitals is called a basis set. Infinite amount of different basis sets[64–66] can be found in the literature and to decide which one is the best, that depends largely on the experience of the user.

Secondly, we have to apply so-called mean field approximation, otherwise we would end up in an unsolvable problem. The electron-electron repulsion is replaced by an interaction of electron with an averaged external field $v_i^{HF}\{j\}$ created by all of the other electrons *j* as follows

$$\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \rightarrow v_i^{HF}\{j\}. \tag{2.12}$$

Using this approximation, we have come to an effective one-electron operator which is called the Fock operator. It reads

$$\hat{f}_i = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + v_i^{HF}\{j\}, \tag{2.13}$$

in which the first two terms were already described and the last term is the Hartree-Fock potential $v_i^{HF}\{j\}$. With this approximation, we arrive to a set of one-electron equations

$$\hat{f}_i|\chi_i\rangle = \varepsilon_i|\chi_i\rangle, \tag{2.14}$$

however, each equation depends on all the other electrons through the average external potential $v_i^{HF}\{j\}$. Because of this, one has to solve this set iteratively until the self-consistency has been achieved.

1.2.3. Density functional theory

The wave function is rather complicated 3N dimensional function which is hard to obtain. Is there not any easier way to deal with the ab initio problems? In 20th century, M. Born interpreted squared wave function as a probability to find an electron, i.e. electron density. Do we need this complicated 3N dimensional wave function or does the electron density provide us with all the required information? These questions were raised in the last century and thanks to P. Hohenberg, L.J. Sham, and W. Kohn, we know that the electron density itself is enough for describing the ground state of a quantum system. The simplification in the wave function is amazing. From 3N electron problem, we come down to a problem with only 3 dimensions. An overall energy functional $E[\rho]$ has a form of

$$E[\rho] = T_{ni}[\rho] + V_{eN}[\rho] + V_{ee}[\rho] + \Delta\, T[\rho] + \Delta\, V_{ee}[\rho], \tag{2.15}$$

where $T_{ni}[\rho]$ stands for functional kinetic energy of non-interacting electrons, $V_{eN}[\rho]$ represents interaction of the nuclei with electron density, and $V_{ee}[\rho]$ is the electron-electron repulsion term. The last two terms $\Delta\, T[\rho]$ and $\Delta\, V_{ee}[\rho]$ are the corrections which are together referred to as the exchange-correlation term $E_{xc}[\rho]$

$$E_{xc}[\rho] = \Delta\, T[\rho] + \Delta\, V_{ee}[\rho], \tag{2.16}$$

but unfortunately the form of this functional is unknown. We know certain properties of this functional such as what it should look like and so on. However, it still has to be approximated. The simplest form of this exchange functional depends only on the electron density and it is thus called a local density approximation (LDA functionals). Of

course, there are far more advanced methods which also include first derivatives. These methods are called generalized gradient approximations (GGA functionals). There is another option and that is to combine the exact exchange energy from the Hartree-Fock method with some empirical approximation like LDA or GGA in certain ratio. Functionals created in this way are called hybrid functionals. The ratio is usually determined by fitting onto experimental values.

The density functional theory has proved itself to be a valuable method in ab initio calculations. It outmatches the wave function theories many times with the same resources. Therefore, it is a feasible and quite accurate method for determining equilibrium geometries and many other chemical properties.

1.2.4. Population analysis and parametrization of small molecules

Finally we have reached the purpose of the whole quantum chemical section. One of the non-bonded interactions included in every classical force field is the electrostatic term. This electrostatic part takes form of a simple Coulomb's law and all atoms interact according to this law. Therefore, every atom has a defined partial atomic charge but the way to obtain such charge is not completely clear. The partial charges are not physically observable quantity and hence there are many ways to obtain them. A common way is to perform an ab initio calculation and assign the partial atomic charges in accordance with some method. There are many methods which have been proposed such as the Natural Population Analysis (NPA)[67] which assigns partial atoms charges according to the degree of a contribution of an atomic basis function to the overall wave function. The next type of methods uses the electrostatic potential. From the wave function, we can calculate the electron density and subsequently the electrostatic potential. This is already an observable property and we can therefore compare the computed electrostatic potential to the one experimentally observed. Partial atomic charges can be then fitted to reproduce this computed/experimentally measured electrostatic potential. As an example of this method can serve a CHELPG method[68]. A modification to this procedure, which ensures that rotationally degenerate atoms have equivalent partial atoms charges, is called the Restrained Electrostatic Potential method (RESP)[69]. This method is recommended to

parameterize a small molecule for the AMBER force fields in a condensed phase. The electrostatic potential should be calculated by a HF/6–31G* level of theory in a gas phase. In spite of using such low level of description, the error in this calculation is close to the difference in charge distribution between the gas phase and what we observe in the solution.

2. Advanced sampling methods

After going through general methods and theories, we will now focus on more advanced methods of computer modeling which are used throughout this thesis. The common theoretical framework was taken from D.A. McQuarrie[70], C. Chipot, A. Pohorille[71], D. Frenkel, B. Smith[56]. More advanced approaches, which have not been yet, present in books are cited on the fly.

2.1. Free energy calculations

When any reaction, e.g. sn2 chemical reaction, occurs, we can say that there is a certain driving force which pushes the reaction along its reaction coordinate. This driving force is often expressed in a magical term - free energy difference.

Statistical mechanics dictate that if we know canonical partition function, then we know everything. The canonical partition function Q of a system can be calculated if we integrate over the whole phase space

$$Q = \frac{1}{\lambda^{3N} N!} \int e^{-\frac{U(\vec{q})}{k_b T}} d\vec{q}, \tag{2.17}$$

where U expresses the potential energy of a system, $\vec{q}$ being the set of 3N coordinates, k_b being the Boltzmann's constant, T being the absolute thermodynamic temperature, N

being number of particles, and λ being the temperature dependent de Broglie wavelength. The canonical partition function can be related to Helmholtz free energy A

$$A = -k_b T ln(Q). \tag{2.18}$$

Helmholtz free energy is relevant if a process is performed under constant number of particles, constant volume and constant temperature. However, Gibbs free energy G usually has greater importance. It takes place under constant pressure instead of constant volume. Under small pressures and in a condensed phase, in which the volume changes are small, the difference between Helmholtz free energy A and Gibbs free energy G is negligible.

The difference between two states, e.g. bound and unbound state of a protein and a ligand is what people are usually interested in. This difference is commonly expressed by a free energy difference ΔA. When we now express this free energy difference in terms of partition functions, we get

$$\Delta A = -k_b T ln\left(\frac{Q_2}{Q_1}\right) = -k_b T ln\left(\frac{\int e^{-\frac{U_2(\vec{q})}{k_b T}} d\vec{q}}{\int e^{-\frac{U_1(\vec{q})}{k_b T}} d\vec{q}}\right). \tag{2.19}$$

All free energy methods used in computer simulations can be derived from this equation.

As the partition function defines the effective number of available states, the free energy difference could be also evaluated by a so called probability method

$$\Delta A = -k_b T ln\left(\frac{P_2}{P_1}\right), \tag{2.20}$$

where P_2 is a probability of being in the state 2, while P_1 is a probability of being in the state 1. This method is convenient when the energetic barrier between these two states is not high and when the free energy difference is low. If these requirements are not met then we need to use more advanced methods like thermodynamic integration or umbrella sampling methods.

In computer modeling, we are often interested in a free energy difference between two states along certain pathway. It is convenient to have some reaction coordinate λ which makes a distinction between our two thermodynamic states. The reaction coordinate can be of any kind, ranging from the simplest like distance between two atoms to more complex like root mean square deviation between two structures.

2.1.1. Thermodynamic integration

Thermodynamic integration is one of the most common methods for calculating the free energy difference. Starting with the formula for a free energy

$$A = -k_b T ln(Q) = -k_b T ln\left(\int e^{-\frac{U(\vec{q})}{k_b T}} d\vec{q}\right), \tag{2.21}$$

defining a reaction coordinate λ and taking a derivate with respect to the reaction coordinate λ yields

$$\frac{dA}{d\lambda} = -k_b T \frac{d}{d\lambda} ln(Q) = -k_b T \frac{d}{d\lambda} ln\left(\int e^{-\frac{U(\lambda,\vec{q})}{k_b T}} d\vec{q}\right), \tag{2.22}$$

which can be rearranged into

$$\frac{dA}{d\lambda} = \frac{\int \frac{dU(\lambda,\vec{q})}{d\lambda} e^{-\frac{U(\lambda,\vec{q})}{k_b T}} d\vec{q}}{\int e^{-\frac{U(\lambda,\vec{q})}{k_b T}} d\vec{q}} = \langle \frac{dU(\lambda,\vec{q})}{d\lambda} \rangle_\lambda. \tag{2.23}$$

The brackets denote an average of the quantity which is inside. By integration over the reaction coordinate λ, we get the free energy difference ΔA

$$\Delta A = \int_0^1 \langle \frac{dU(\lambda,\vec{q})}{d\lambda} \rangle_\lambda \, d\lambda, \tag{2.24}$$

where $\lambda = 0$ defines an initial state with a free energy $A_1$and $\lambda = 1$ defines a final state with a free energy A_2.

One of the simplest application of the above derived formula is to evaluate $\langle \frac{dU(\lambda,\vec{q})}{d\lambda} \rangle_\lambda$ for a number of intermediate values of λ and then perform numerical integration. There are various methods for numerically evaluating the integral using, for instance, the trapezoidal or the Simpson's rule.

At first the potential $U(\lambda,\vec{q})$ has to be defined. It is done by dividing overall potential to common part $U_{common}(\vec{q})$ and a part that is being perturbed $V(\lambda,\vec{q})$ as follows

$$U(\lambda,\vec{q}) = U_{common}(\vec{q}) + V(\lambda,\vec{q}). \tag{2.25}$$

The way we define the coupling between initial state $\lambda = 0$ and the final state $\lambda = 1$ is also important. The simplest approach uses linear mixing between an initial state 0 and a final state 1

$$V(\lambda) = (1-\lambda)V_0 - \lambda V_1, \tag{2.26}$$

with V_0 being the potential of the initial state and V_1being the potential of the final state. This form of coupling is often sufficient. In this case however,we make some atoms appear or disappear and the integrand in Equation 2.24 diverges at $\lambda = 0$ or $\lambda = 1$ (depends on whether atoms are appearing or disappearing). An elegant way of solving this problem is to use softcore potentials. A modified version of the Lennard-Jones potential then gets a form of

$$V_{0,dissapearing} = 4\varepsilon(1-\lambda)\left[\frac{1}{\left[\alpha\lambda + \left(\frac{r_{ij}}{\sigma}\right)^6\right]^2} - \frac{1}{\alpha\lambda + \left(\frac{r_{ij}}{\sigma}\right)^6}\right], \tag{2.27}$$

$$V_{1,apearing} = 4\varepsilon\lambda\left[\frac{1}{\left[\alpha(1-\lambda) + \left(\frac{r_{ij}}{\sigma}\right)^6\right]^2} - \frac{1}{\alpha(1-\lambda) + \left(\frac{r_{ij}}{\sigma}\right)^6}\right], \tag{2.28}$$

where α is an empirical parameter. With this modification, the linear coupling between potentials can, and should be, used for the van der Waals transformations.

The statistical error of thermodynamic integration is usually evaluated as follows. $\langle\frac{dU(\lambda,\vec{q})}{d\lambda}\rangle_\lambda$ is evaluated in every microstate from which variance is calculated. The total variance $var(\Delta A)$ is then weighted sum of the intermediate variances $\left(\frac{dU}{d\lambda}\right)_i$

$$var(\Delta A) = \sum_{i=1}^{N} w_i var\left(\frac{dU}{d\lambda}\right)_i, \tag{2.29}$$

where w_i are the weights which are determined by the integration method. To get a correct statistical error, each individual average needs to be further corrected for an autocorrelation time τ by multiplying each of the intermediate variances by a factor of $\sqrt{2\tau}$.

In order to access free energy difference between two states, one is often forced to use a thermodynamic cycle. An example of such cycle is shown in Figure 2.1.

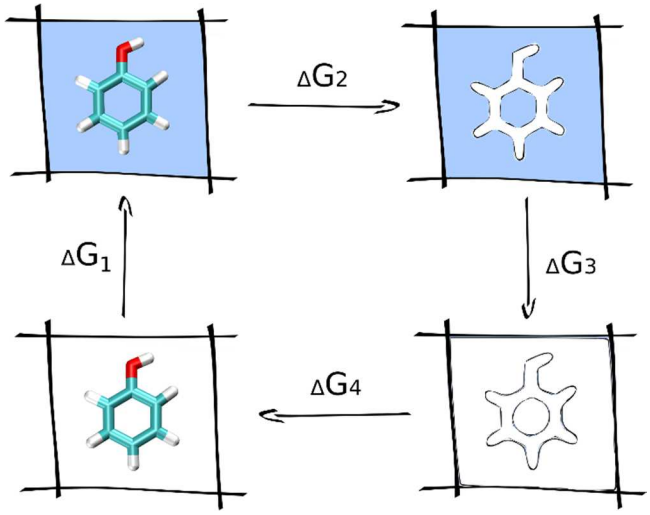

Figure 2.1. An illustration of a complete thermodynamic cycle to calculate the solvation energy of phenol.

We start here with a phenol molecule in vacuum (left, bottom). ΔG_1 represents free energy difference of the phenol molecule solvation. ΔG_2 represents free energy difference of turning off all interactions of the phenol molecule while being present in a water solution. ΔG_3 represents free energy difference of transferring non-interacting phenol molecule from water environment into vacuum. As the phenol molecule gets completely decoupled, the free energy difference equals zero. ΔG_4 represents free energy difference of turning on all interactions in vacuum. As free energy is state variable we can state that

$$\sum_{i=1}^{i=4} \Delta G_i = 0. \tag{2.30}$$

For instance, if we want to evaluate the solvation free energy of phenol molecule ΔG_1, we know from this thermodynamic cycle that

$$\Delta G_1 = -\Delta G_2 - \Delta G_4 \tag{2.31}$$

2.1.2. Umbrella sampling

Another elegant method for calculating free energy difference between two states is called the umbrella sampling method. We can overcome the sampling issues by using this method if the brute force fails (the probability method). Graphical representation of the whole idea is shown in Figure 2.2.

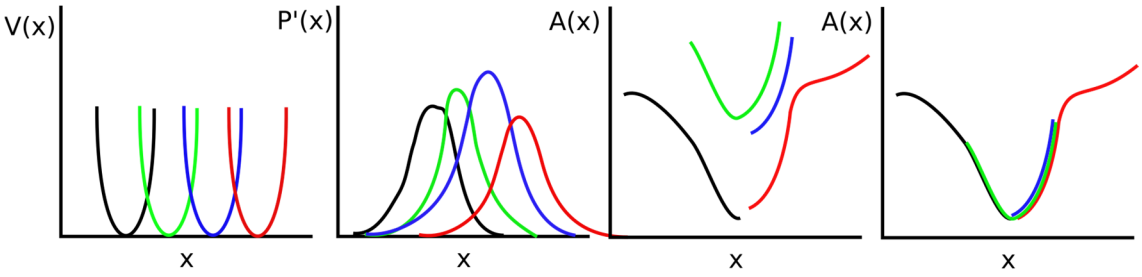

Figure 2.2. Graphical representation of the umbrella sampling method. From left – at first, we sample the reaction coordinate x with help of umbrella potentials $V(x)$. Then we calculate the biased probability $P'(x)$ from which we obtain unbiased free energy profiles $A(x)$. Nevertheless, these profiles are shifted by a constant and need to be combined together as shown in the last picture.

We would like to sample phase space along some reaction coordinate, where the barriers and the free energy differences are too big. Therefore, we sample it with help of umbrella potentials $V(x)$.

Then the biased probability $P'(x)$ is calculated from which we estimate unbiased free energy using the following formula

$$A(x) = -k_b T ln\big(P'(x)\big) - V(x) + F, \tag{2.32}$$

with k_b being the Boltzmann's constant, T being the absolute thermodynamic temperature, and F being the constant. The value of F is undetermined but completely irrelevant. We obtain corresponding free energy profiles after doing many simulations along the reaction coordinate. However, each of them is shifted by a different constant F. The important thing is that the neighboring free energy profiles have to overlap. If this is achieved then we can combine all free energy profiles together.

It is not completely clear how to combine all free energy profiles together. A weighted histogram analysis method (WHAM) was developed[72,73] for this reason. This method utilizes all histograms at once in order to determine the optimal F values so that the free energy profiles are appropriately overlapped. The whole method consists of two following equations

$$P(x) = \frac{\sum_{i=1}^{N_{sim}} n_i(x)}{\sum_{i=1}^{N_{sim}} n_i(x) e^{\frac{F_i - V_i(x)}{k_b T}}}, \tag{2.33}$$

$$F_i = -k_b T ln\left[\sum_{x_{bin}} P(x) e^{-\frac{V_i(x)}{k_b T}}\right], \tag{2.34}$$

where N_{sim} stands for the number of simulations, $n_i(x)$ stands for the number of counts in histogram bin associated with x, x_{bin} is the number of bins, $V_i(x)$ is the biasing potential, F_i is the free energy shift from simulation i, and $P(x)$ is the best estimate of unbiased probability distribution. $P(x)$ and F_i are the unknowns and the equations need

to be solved iteratively to self-consistency. $P(x)$ is then directly proportional to the free energy profile as stated above.

In order to obtain correct estimate of error, one usually divides simulation data into several blocks and then performs the WHAM. The data sets have to be reasonably long (> 10 x autocorrelation time). We can evaluate the variance from the obtained free energy profiles.

A free energy profile calculated by the umbrella sampling method is sometimes called potential of mean force. Potential of mean force binding is connected to a binding constant K_b in the context of protein-ligand by the following relation

$$K_b = \int_V e^{-\frac{W(r)}{RT}} dr, \tag{2.35}$$

where the integration is restricted to the region where the ligand is defined to be bound. If the ligand can be found in one distinct binding site, this can be approximated with a good degree of accuracy by a square well potential. Then

$$K_b = V_b e^{-\frac{W_0}{RT}}, \tag{2.36}$$

where V_b stands for the volume occupied by the ligand when bound to the protein, and W_0 represents the depth of the potential well which in our case is

$$W_0 = W\left(r \to r_{flat}\right) - W(r \to r_{min}), \tag{2.37}$$

where r_{min} denotes minimum in the potential of the mean force calculated by the

umbrella sampling, and r_{flat} represents the distance where the potential gets flat already (bulk solution). Binding constant K_b, in the context of simple protein-ligand binding, is then related to the standard free energy of binding ΔG_b° by

$$\Delta G_b^\circ = -RTln(C^\circ K_b). \tag{2.38}$$

C° stands here for a standard concentration (usually 1M). The inverse of the standard concentration can be interpreted as volume V° occupied by a single molecule at standard concentration 1M. In my case, I chose $C^\circ = 1M$ and then the $V^\circ = 1661$ Å^3. There is also an additional contribution to free energy of binding. From the statistical mechanics point of view, the free energy binding is determined by the following equation

$$\Delta G_b^0 = -RTln\left(\frac{C^0}{8\pi^2}\frac{\sigma^P\sigma^L}{\sigma^{PL}}\frac{Z^{PL}}{Z^P Z^L}\right) + p^0\Delta V_{PL}, \tag{2.39}$$

where C° represents standard state, σ stands for a symmetry number, and Z denotes a configurational integral. Superscripts P, L, and PL refer to protein, ligand, and protein-ligand complex respectively. The term $p^0\Delta V_{PL}$ represents standard pressure times the change of the volume on complex formation correction, which is negligible at standard pressures. In molecular dynamics simulation, it is difficult to capture this symmetry problem. To account for this, one has to adjust final free energy of binding by a factor of

$$\Delta G_{SYM} = -RTln\left(\frac{\sigma^P\sigma^L}{\sigma^{PL}}\right). \tag{2.40}$$

To summarize, the final standard free energy of binding ΔG_b°, which is calculated by the umbrella sampling method and corrected for symmetry, is as follows

$$\Delta G_b^\circ = -RTln\left(\frac{V_b}{V^\circ}e^{-\frac{W_0}{RT}}\right) + \Delta G_{SYM}. \quad (2.41)$$

2.2. Spatial distribution functions

Spatial binding functions are convenient for presenting data in a clear and simple way. An illustration of such function is presented in Figure 2.3.

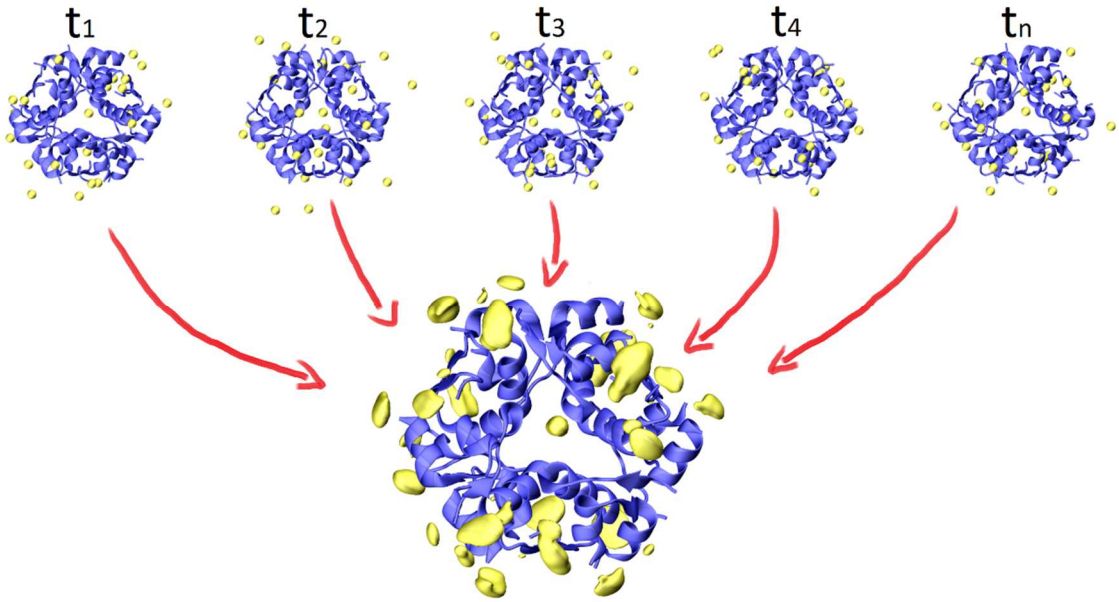

Figure 2.3. An illustration of a spatial distribution function of yellow atoms around a solute of interest (blue protein).

From a collection of microstates, e.g. frames of molecular dynamics simulation (t_1, t_2, …, t_n), we can construct a spatial binding map of an atom/molecule (e.g. the yellow atom) around another molecule (e.g. the protein depicted as blue) we are interested in. Final spatial distribution function then presents overall mean distribution of a molecule of our interest around the chosen molecule. Spatial distribution function is

constructed in the following way. At first all of the microstates are collected and our molecule of interest (protein) is overlapped in every frame according to the lowest root mean square displacement. Then the space is divided into a grid and studied molecules (the yellow atoms) are mapped onto this grid. In order to smoothen the spatial distribution function, the studied molecule does not only contribute to one point on this grid as its contribution is unequally divided into all surrounding grid points. This final spatial distribution function can be then visualized as depicted.

3. Simulated systems

The last section covers specific preparation of all parameters, systems, and calculations which are used throughout this work. At first general things are described which include the protein structure, ligand parameters, and general simulation parameters. Moreover, the composition, simulation settings and specific things for every particular calculation are listed.

3.1. General preparation and parameterization

Insulin hexamer preparation

Insulin is stored at mildly acidic conditions (pH ~ 5.5)[31], which corresponds to protonated B5 histidine. However, the B10 histidine is deprotonated[30,74] due to zinc coordination. The six B13 glutamates in the middle region of the insulin hexamer were considered as deprotonated. One cation was always present in the middle region to compensate for the negative charge of glutamates. In case of R conformation, there was also one chloride anion binding to the B10 zinc. The initial protein coordinates were taken from the following pdb structures: 1AIY[75] (R6 insulin hexamer), 1MSO[3] (T6), and 1TRZ[8] (T3R3). These structures were used and modified as described below.

Neurotransmitters preparation

Phenol, serotonin, as well as dopamin were the assigned parameters from ff03 AMBER force field using the ANTECHAMBER package[76]. Partial atomic charges were derived by the standard HF/6–31G* method in the gas phase using the RESP method. These calculations were performed in the Gaussian 09 package[77]. Note that both neurotransmitters are charged at pH 5.5 (charge +1).

The B10 zinc binding site

The B10–Zn interaction potential had to be re-parameterized in order to at least partially account for electronic polarization and charge transfer effects and to reproduce experimental data. The partial atomic charges on the B10 zincs and B10 histidines were changed according to results from ab initio calculations which lead to the charge of B10 Zn cation of +1.5 and each B10 histidine residues having a net charge of +0.1677. Zinc non-bonded van der Waals parameters r_0 and ε were set to 2.44 Å and 0.25 kcal/mol. The presence of another cation (in addition to the two B10 coordinated zincs) in the highly negatively charged middle region of protein (B13 glutamates site) is crucial for achieving stability of the B10–Zn site.

Ion parameterization

In order to model ions properly, I used recently published force fields for NaCl and $CaCl_2$[78,79]. These new force fields have, so that we can account for the polarizability of ions, rescaled charges by a factor of 0.75[80–82]. The van der Waals parameters were also rescaled to match neutron scattering data reference. By using these rescaled charges, I was sometimes forced to have a chloride ion of charge -1.0 to achieve systems' electroneutrality. In a similar way, all other ionic charges were rescaled by a factor of

0.75 in order to account for electronic polarization effects at least in a mean-field way[80–82]. However, the van der Waals parameters were not changed.

General simulations parameters

All molecular dynamics simulations were prepared and performed using the AMBER 14 program with GPU acceleration whenever possible[83,84]. Throughout all simulations, the AMBER ff03 non-polarizable protein force field and the SPC/E water model were used[85]. Production simulations were performed in isothermal-isobaric ensemble at ambient conditions T = 300 K and p = 1 atm using the Berendsen barostat and thermostat[86]. The only exception were thermodynamic integration calculations, where the temperature was controlled by Langevin dynamics with a reference temperature of T = 300 K and a collision frequency 5 ps^{-1} in order to avoid problem of ergodicity when the ligand is fully decoupled from its environment. 3D periodic boundary conditions were applied in with a general cutoff of 9 Å. The long range electrostatic interactions were accounted for by using the particle mesh Ewald method[59]. Bonds containing hydrogen atoms were constrained using the SHAKE algorithm[61](analytical version SETTLE for water molecules[60]). The time step of 2 fs was used for dynamics.

3.2. Specific simulation settings

Cation binding to T6 and T3R3 insulin hexamers – simple MD simulations

The simulations of cation binding to insulin hexamers were done in 0.5 M $ZnCl_2$, $CaCl_2$, NaCl or KCl. The unit cell consisted of a single insulin hexamer, 27 000 SPC/E water molecules, 2 B10 Zn cations, 1 central B13 glutamate cation of the same type as in the salt used, 242 cations, 239 chloride anions (or 482 chloride anions in case of divalent cations), and one chloride anion with -1.0 charge to ensure electroneutrality. After the

preparation, the systems were minimized using 5 000 steps of the steepest descent method, where the protein was restrained with a harmonic potential. To equilibrate, at first the velocities for T = 10 K were assigned from the Maxwell-Boltzmann distribution. The systems were then subjected to 200 ps of isothermal-isochoric molecular dynamics while the temperature was slowly raised to T = 300 K. This was followed by 1.2 ns of isothermal-isobaric equilibration which lead to a cell of approximately 95 × 95 × 95 $Å^3$. The total simulation time for each ion binding simulation was 1.5 µs, with preceding 10 ns of equilibration.

Phenolic pocket – initial direct MD simulations

I investigated substitution of phenol in its known binding place by serotonin or dopamine neurotransmitters. As a reference, I used an insulin structure crystalized in the presence of phenol[75]. This resulted in the R6 insulin hexamer with 6 phenol molecules in the phenolic pockets. In order to obtain an initial picture, the phenol molecules were exchanged either for dopamine or serotonin. For both neurotransmitters, two initial starting orientations in the phenolic pocket were used. All starting geometries of phenolic ligands inside the phenolic pockets are shown in Figure 2.4.

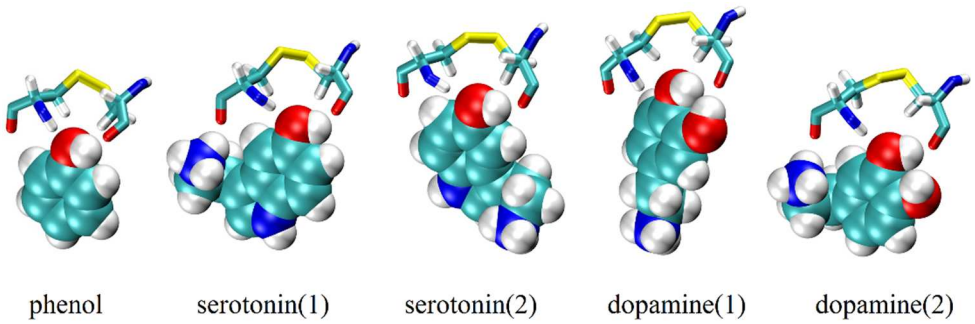

Figure 2.4. Phenol together with serotonin and dopamine neurotransmitters are depicted in the phenolic pocket. Two initial orientations of serotonin and dopamine were used.

Furthermore, every insulin R6 hexamer was immersed into a unit cell containing 9 000 SPC/E water molecules with chloride/sodium ions added to ensure overall electroneutrality. After the preparation, the systems were minimized using 5 000 steps of the steepest descent method, where the protein and phenolic ligands were restrained with a harmonic potential. For equilibrating, at first the velocities for T = 10 K were assigned from the Maxwell-Boltzmann distribution. The systems were then subjected to 200 ps of isothermal-isochoric molecular dynamics, where the temperature was slowly raised to T = 300 K. This was followed by 1.2 ns of isothermal-isobaric equilibration which lead to a cell of approximately 69 × 68 × 66 $Å^3$. After the equilibration, the production simulations were run for 600 ns under the conditions stated above in the general simulation parameters.

The phenolic pocket – free energy calculations

Thermodynamic integration is an excellent way to compare different phenolic ligands inside the phenolic pocket. Forward and backward mutations between phenol, dopamine, and serotonin were performed. Each mutation consisted of several smaller steps. Using this approach, one can effectivelly calculate the so-called binding free energy difference ΔΔG. However, it is also possible to calculate the absolute free energies of binding for small molecules by a double annihilation method[87]. The double annihilation method is the most precise when the perturbation is small, i.e. when the number of mutated atoms is the smallest. Therefore, a phenol molecule was chosen as this calculation should be the most accurate. Only one ligand was being mutated at that time. By combining these methods, one gets absolute free energies of all three binding neurotransmitters. Illustration of all calculations is shown in Figure 2.5. Details of all calculations will be described below.

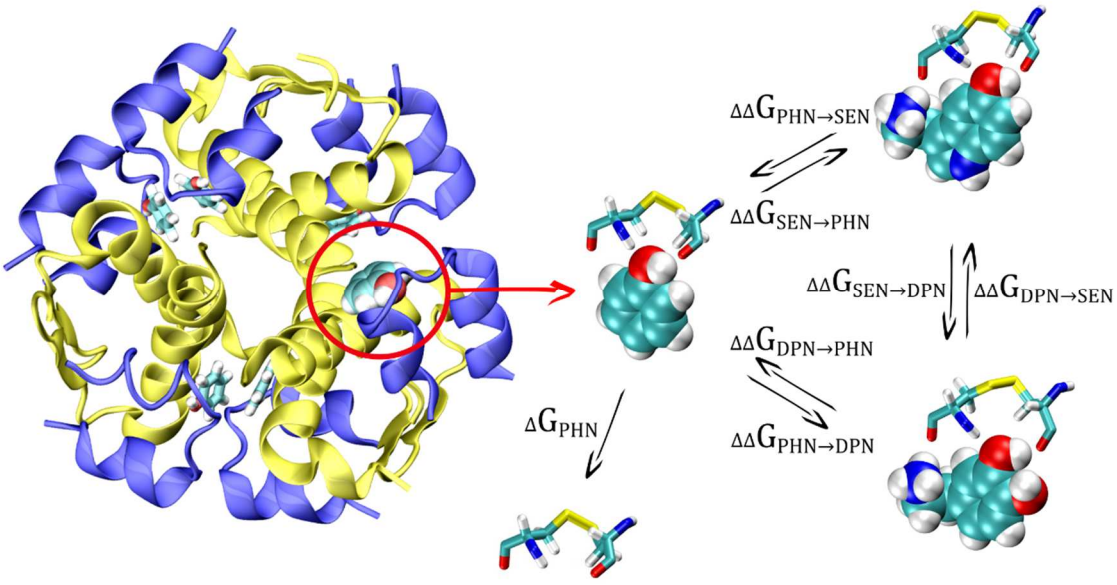

Figure 2.5. An illustration of free energy calculations, which are the aim of this work, (relative free energies of binding ($\Delta\Delta G$), and absolute free energies of binding (ΔG).

Composition of prepared systems was as following: insulin R6 hexamer, 9 000 SPC/E water molecules with chloride/sodium ions added to ensure overall electroneutrality, 5 phenol molecules present in the phenolic pockets, and 1 phenolic ligand (phenol, dopamine, or serotonin). In thermodynamic cycles, which will be introduced below. there are also calculations of mutating/decoupling phenolic ligand in bulk water. One phenolic ligand was immersed here into a box of 2611 SPC/E water molecules.

After the preparation, the systems were minimized using the steepest descent method. In order to equilibrate, at first the velocities for T = 10 K were assigned from the Maxwell-Boltzmann distribution. The systems were then subjected to 200 ps of isothermal-isochoric molecular dynamics, where the temperature was slowly raised to T = 300 K. This was followed by 1.2 ns of isothermal-isobaric equilibration which lead to a cell of approximately 69 × 68 × 66 Å^3, or 42 × 42 × 42 Å^3.

At first the differences in free energy of binding ($\Delta\Delta G_{1-2}$) between phenol (PHN), dopamine (DPN), and serotonin (SEN) were calculated as follows. For these types of calculations, a complete thermodynamic cycle shown in Figure 2.6. was used (example calculation $\Delta\Delta G_{PHN-SEN}$).

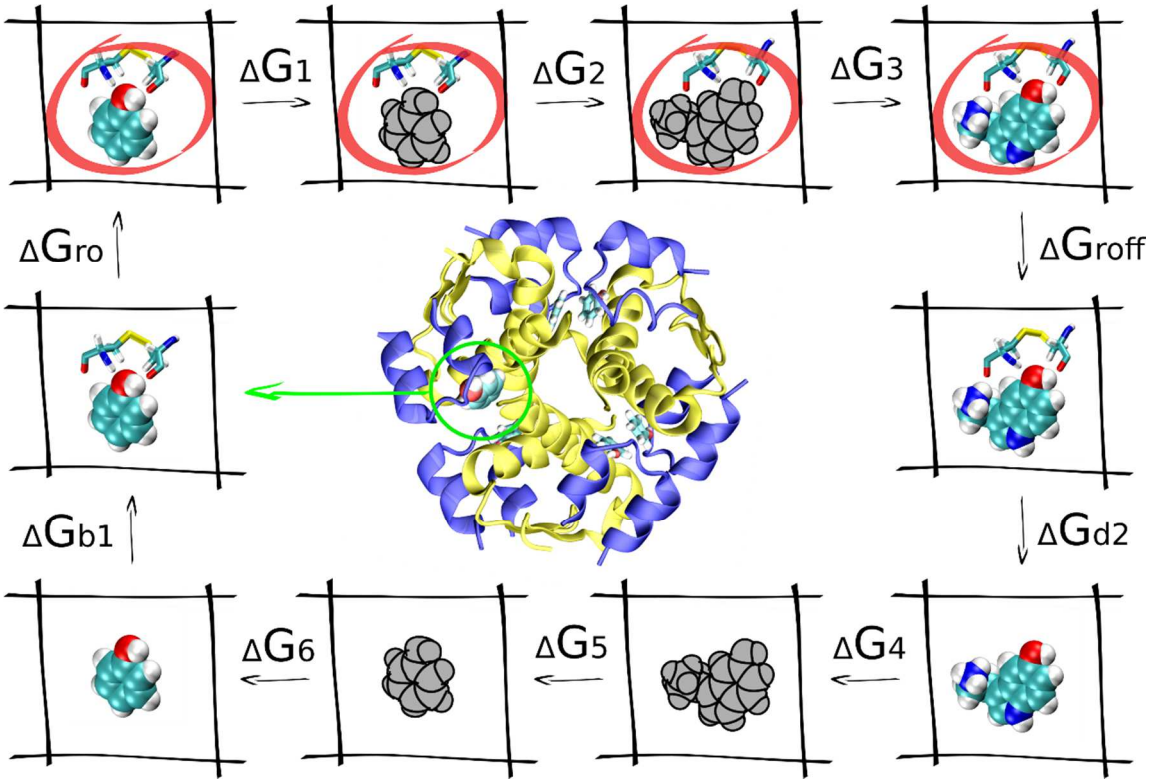

Figure 2.6. A practical example of a complete thermodynamic cycle used to calculate the differences in the free energies of binding ($\Delta\Delta G_{1-2}$) between different phenolic ligands to phenolic pocket of insulin R6 hexamer. A thermodynamic cycle to calculate $\Delta\Delta G_{PHN-DPN}$ is shown here. The restraints are indicated by a red circle; the grey ligand indicates that the electrostatic interactions of a ligand are turned off. ΔG_{b1} represents free energy of binding of phenol to the phenolic pocket whereas ΔG_{d2} represents free energy of dissociation of serotonin from the phenolic pocket. ΔG_{ro} represents free energy of restraining phenol to a certain position inside the phenolic pocket. ΔG_1 represents gradual turning off electrical charges of phenol inside the phenolic pocket while keeping proposed restraints on. ΔG_2 represents mutation of a restrained phenol to a restrained serotonin inside the phenolic pocket while all charges are turned off. ΔG_3 represents free energy of turning on the electrical charges of the serotonin while the restrains are on. ΔG_{roff} represents free energy of releasing restraints on the serotonin inside the phenolic pocket. ΔG_4 stands for free energy of turning off electrical charges of the serotonin in the bulk solution. ΔG_5 represents free energy of mutating serotonin to phenol while all electrical charges are off in the bulk solution. ΔG_6 represents free energy of turning on electrical charges of phenol in the bulk solution.

ΔG_{b1} represents free energy of binding of a ligand 1 to the insulin while ΔG_{d2} represents free energy of dissociation of a ligand 2 from the insulin. $\Delta\Delta G_{1-2}$ reflects the difference in free energies of binding between these two ligands. As this is a complete thermodynamic cycle, this energy equals to

$$\Delta\Delta G_{1-2} = -\Delta G_{b1} - \Delta G_{d2}, \tag{2.42}$$

$$\Delta\Delta G_{1-2} = \Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_4 + \Delta G_5 + \Delta G_6 + \Delta G_{ron} + \Delta G_{roff}. \tag{2.43}$$

Explanation of these terms can be found in Figure 2.6. One of the corrections that has been proposed is the long range dispersion correction ΔG_{LRC} [88]. However, this contribution turned out to introduce significant errors and thus it was omitted. Each subsequent simulation was performed using linear scaling between potential V0 and V1 with lambda windows 0, 0.1, up to 1.0, resulting in 11 windows. The only exception were simulations where the van der Waals parameters were changed. The softcore potentials were used here while the lambda step stayed shorter – 0.0, 0.05, up to 1.0, resulting in 21 windows. Integration was always carried out by the trapezoidal rule. All simulations were performed with a simulation step of 2 fs for a total simulation time 5 ns (at first) with preceding 1.2 ns equilibration. In certain cases, an additional set of 20 ns calculations was performed to check for a convergence. Altogether, a single calculation of binding free energy difference $\Delta\Delta G_{1-2}$ consisted of 108 subsequent simulations. To restrain a ligand in the phenolic pocket, a simple distance restraint between center of mass of A6 backbone nitrogen atom, and A11 backbone oxygen atom to phenolic ligands crucial oxygen atom was used with a force constant of $0.5\,\frac{kcal}{mol\cdot Å^2}$. and equilibrium value of 1.5 Å. With this force constant, it was proved that the steps of applying (ΔG_{ro}) and releasing restraints (ΔG_{roff}) are negligible. In order to be able to reasonably estimate an error in these calculations, every calculation was performed in both directions (forward and backwards mutations) and multiple times as depicted in Figure 2.5. This lead to 6 separate mutations with the following differences in binding

free energies $\Delta\Delta G_{PHN\rightarrow DPN}$, $\Delta\Delta G_{DPN\rightarrow PHN}$, $\Delta\Delta G_{PHN\rightarrow SEN}$, $\Delta\Delta G_{SEN\rightarrow PHN}$, $\Delta\Delta G_{DPN\rightarrow SEN}$, and $\Delta\Delta G_{SEN\rightarrow DPN}$.

Furthermore, one can take the advantage of the fact, that forward and backward calculations like $\Delta\Delta G_{PHN\rightarrow DPN}$ and $\Delta\Delta G_{DPN\rightarrow PHN}$ should be of the same value, differing only in a sign. With this in mind, one can combine all 6 simulations to obtain a reasonable error estimate. Moreover, free energy of mutating dopamine to serotonin with all electrical charges off ($\Delta G_{DPN\rightarrow SEN}$) was calculated in accordance with the following thermodynamic relation

$$\Delta G_{PHN\rightarrow DPN} + \Delta G_{DPN\rightarrow SEN} + \Delta G_{SEN\rightarrow PHN} = 0. \quad (2.44)$$

This is due to the fact that the step of mutating atoms is the most accurate when one is mutating only a few atoms (perturbation is not large). Other general parameters of calculations, which are not listed here, are stated in the General simulation parameters section.

To obtain the absolute free energy of a phenol molecule binding to a phenolic pocket (ΔG_{PHN}), a thermodynamic cycle shown in Figure 2.7. was used.

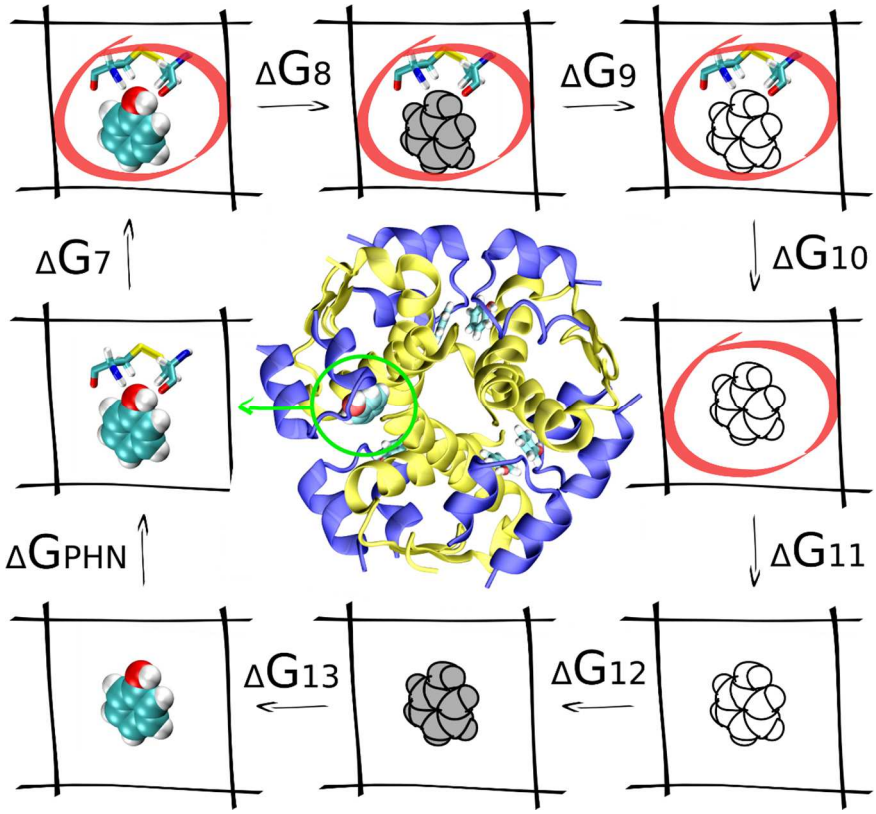

Figure 2.7. A complete thermodynamic cycle used to calculate the absolute free energy of binding of a phenol molecule to a phenolic pocket of insulin R6 hexamer (ΔG_{PHN}). The restraints are depicted by a red circle; grey ligand means that the electrostatic interactions are turned off; fully transparent ligand means that the ligand is fully decoupled from its environment. ΔG_{PHN} represents free energy of binding of phenol to the insulin. ΔG_7 represents free energy of restraining phenol to a certain position inside the phenolic pocket. ΔG_8 represents free energy of turning off electrical charges of phenol inside the phenolic pocket while keeping the proposed restraints on. ΔG_9 stands for the free energy of decoupling restrained phenol from the phenolic pocket while all charges stay turned off. ΔG_{10} stands for the free energy of transferring decoupled phenol from phenolic pocked to the bulk water environment. As the phenol does not interact with the protein at all, the energy of bound and unbound phenol is the same. Hence the free energy difference between these two states is zero. ΔG_{11} represents free energy of releasing the proposed restrains from phenol which is now situated in the bulk solution. ΔG_{12} represents the free energy of turning on the van der Waals interactions of phenol in the bulk solution. ΔG_{13} stands for the free energy of turning on electrical charges of phenol in the bulk solution.

The computational protocol was as follows. Each subsequent simulation was performed using linear scaling between potential V0 and V1. The only exception were the simulations where the phenol molecule was completely decoupled from its environment (ΔG_9). The softcore potentials were used here for achieving better stability and smoother shape of free energy derivations' curve. Integration was always done using the trapezoidal rule. The first step of the whole thermodynamic cycle was proposing restrains on the phenol molecule (ΔG_7). This was done by using same restrains as Boresch et al.[87,89], where I restrain the phenol molecule by 3 dihedral angles ($\chi(P1,P2,P3,L4)$; $\psi(P2,P3,L4,L5)$; $\omega(P3,L4,L5,L6)$, 2 angles ($\varphi(P2,P3,L4)$; $\phi(\mathrm{P3,L4,L5})$), and one distance restraint ($\mathrm{r(P3,L4)}$). While using these restraints, we need to choose 3 protein atoms (labeled P), and 3 phenol atoms (labeled L). In my calculations, I chose P1 – A6 cysteine backbone oxygen, P2 – A11 cysteine backbone H (NH), P3 – A11 cysteine side change Sulphur, L4 – phenols C4, L5 – phenols oxygen, and L6 – phenols C2 (Figure 2.8.).

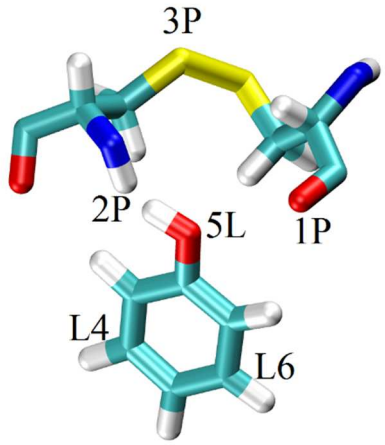

Figure 2.8. Numbering of the atoms used for the restraining the orientation of the phenol in the phenolic pocket. P1 – A6 cysteine backbone oxygen, P2 – A11 cysteine backbone H (NH), P3 – A11 cysteine side change Sulphur, L4 – phenols C4, L5 – phenols oxygen, and L6 – phenols C2.

The phenol is thus effectively restrained the phenolic pocket. The equilibrium values were taken from a non-restrained molecular dynamics simulation. The values were $\chi =$

$80.0°, \psi = -46.35°, \omega = -137.75°, \varphi = 37.25°, \phi = 30.6°, r = 7.15$ Å. The force constant values for restraining a ligand in the binding pocket should not generally matter. However, it was proved that the distance force constant $K_r = 10 \frac{kcal}{mol \cdot Å^2}$, and the dihedral/angle restraint force constant $K_i = 200 \frac{kcal}{mol \cdot rad^2}, i = \varphi, \phi, \chi, \psi, \omega$ work well in the case when the ligand has a well-defined position in the binding pocket. These restraints were used during the whole process of decoupling the phenol from its phenolic binding pocket in insulin R6 hexamer. Process of calculating the free energy difference of restraining the phenol was broken into 12 windows with a lambda value of 0.000, 0.200, …, 0.800, 0.850, 0.875, …, and 1.000. Each window was equilibrated for 2 ns and then the data were gathered for 5 ns. The second step was turning the electrical charges of the phenol in the binding pocket off (ΔG_8), while the restraints are on. This was done in 11 windows (lambda values: 0.000, 0.200, …, 0.600, 0.700, 0.750, …, 1.000), each equilibrated for 2 ns, followed by 30 ns data acquisition. The next step involved full decoupling of the phenol while the proposed restrains stayed on and the electrical charged turned off (ΔG_9). As this was the most problematic step, the calculation was divided into 33 windows (lambda values: 0.000, 0.025, …, 0.250, 0.300, …, 0.600, 0.625, …, 1.000), each equilibrated for 2 ns, followed by 50 ns of data collection. The next step, the transition from the bound to the unbound state (bulk solution) has a zero free energy difference $\Delta G_{10} = 0$ as the ligand gets fully decoupled from its environment. The following step is releasing the restraints from the phenol (ΔG_{11}). As the phenol is completely decoupled, the releasing of the proposed restrains can be done anallytically[87] by the following formulae

$$\Delta G_{10} = -RTln\left(\frac{8\pi^2 V_0}{r^2 \sin(\varphi) \sin(\phi)} \cdot \frac{(K_r K_\varphi K_\phi K_\chi K_\psi K_\omega)^{0.5}}{(2\pi RT)^3}\right), \quad (2.45)$$

where T stands for the thermodynamic temperature, R is the molar gas constant, r is the equilibrium distance of our chosen distance restraint, φ and ϕ are the equilibrium angles of our chosen angle restraints, V_0 is factor which corrects the energy to a 1M standard state and its value is 1661 Å^3, K_r is the force constant for distance restraint, K_φ and K_ϕ

are force constants for angle restraints; K_χ, K_ψ, and K_ω are the force constants for dihedral angle restraints. Now that the restraints got released, the only remaining steps are to turn on the van der Waals interactions (ΔG_{12}) and electrical charges of the phenol (ΔG_{13}) in the bulk solution. The van der Waals interactions were turned on in 21 windows (lambda values: 0.00, 0.05, …, 1.00), each equilibrated for 2 ns, and followed by 10 ns of data collection. The electrostatic interactions were turned on in 11 windows (lambda values: 0.0, 0.1, …, 1.0), each equilibrated for 2 ns, and followed by 10 ns of data collection.

Summing all the terms, we obtain the standard free energy difference of decoupling the phenol from the phenolic pocket ΔG_{PHN}

$$\Delta G_{PHN} = -(\Delta G_7 + \Delta G_8 + \Delta G_9 + \Delta G_{11} + \Delta G_{12} + \Delta G_{13}). \quad (2.46)$$

However, due to the symmetry of insulin R6 hexamer, there is also an additional contribution toward the free energy of binding. In my calculations, I did not sample the whole configurational space (due to the restraints). There are 6 equivalent binding sites for a phenol molecule. To account for this, the final free energy of binding has to be adjusted by a factor of ΔG_{symm}

$$\Delta G_{symm\_i} = -RTln(i), \quad (2.47)$$

where i is dependent on remaining free binding sites for the phenol molecule. For R6 insulin hexamer without any phenol bound, i equals 6 hence the entropy contribution is the highest. This number goes to zero as the phenolic pockets get occupied by phenol molecules. Table 2.1. summarizes the entropy effects $\Delta G_{symm\_}i$ on binding of phenol while there are $i - 6$ phenol molecules bound.

Table 2.1. Entropy contribution to the free energy of binding of a phenol to the phenolic pockets according to how many phenolic pockets are unoccupied (ΔG_{symm_i}). If $i = 6$, then there is no phenol molecule bound and the contribution is the highest.

i	6	5	4	3	2	1	$\frac{\langle \Delta G_{symm} \rangle}{kcal/mol}$
$\frac{\Delta G_{symm\_i}}{kcal/mol}$	-1.07	-0.96	-0.83	-0.65	-0.41	0.00	-0.65

I will report entropy contribution as an average of these values $\langle \Delta G_{symm} \rangle$. By combining both approaches, one gets the absolute standard free energies of binding of phenol, serotonin, and dopamine to a phenolic pocket of the insulin R6 hexamer as follows

$$\Delta G^{\circ}_{PHN} = \Delta G_{PHN} + \langle \Delta G_{symm} \rangle, \tag{2.48}$$

$$\Delta G^{\circ}_{DPN} = \Delta G^{\circ}_{PHN} + \Delta\Delta G_{PHN \to DPN}, \tag{2.49}$$

$$\Delta G^{\circ}_{SEN} = \Delta G^{\circ}_{PHN} + \Delta\Delta G_{PHN \to SEN}. \tag{2.50}$$

The standard free energy of binding is also often expressed in terms of a disociation constant K_D

$$K_D = e^{\frac{\Delta G^{\circ}_b}{RT}}. \tag{2.51}$$

The A14/A17 binding pocket – direct MD simulations

During the direct simulations a hitherto unknown binding sites for neurotransmitters on the surface of insulin R6 hexamer were discovered. To investigate

these new, previously unknown binding sites, four additional simulations were performed.Insulin R6 hexamer with six phenol molecules present in the phenolic pockets was propagated here in time in 74 mmol solution of one of the neurotransmitters. This was done either in pure water or in a 0.5 M $CaCl_2$ solution. The computational setup of insulin R6 hexamer in 74 mmol solution of neurotransmitters was the following: insulin R6 hexamer with 6 phenol molecules present in the phenolic pockets, 9 000 SPC/E water molecules, 12 dopamine, or 12 serotonin molecules, and counterions to ensure overall electroneutrality. Computational setup for insulin hexamer in solution of neurotransmitters and 0.5 M $CaCl_2$ was similar. However, due to the ions, the box had to be larger. The simulation box contained the following: insulin R6 hexamer with 6 phenol molecules present in the phenolic pockets, 27 000 SPC/E water molecules, 36 dopamine, or 36 serotonin molecules, 243 calcium cations with a charge +1.5, 529 chloride anions with a charge -0.75, and 1 chloride anion with a charge -1.0. After the preparation, the systems were minimized using 5 000 steps of steepest descent method, where the protein and phenol molecules were restrained with a harmonic potential. In order to equilibrate the system, at first the velocities for T = 10 K were assigned from the Maxwell-Boltzmann distribution. The systems were then subjected to 200 ps of isothermal-isochoric molecular dynamics, where the temperature was slowly raised to T = 300 K. This was followed by 1.2 ns of isothermal-isobaric equilibration which lead to a cell of approximately 70 × 68 × 66 $Å^3$ in the case of a smaller system and to 96 × 95 × 95 $Å^3$ in the case of the larger one. After the equilibration, we ran production simulations for 1000 ns (pure water) or 800 ns (salt solution).

The A14/A17 binding pocket – free energy calculations

As the A14/A17 binding pocket is located on the surface of the insulin hexamer, umbrella sampling is a suitable option for obtaining energetics of binding of neurotransmitters to the A14/A17 binding site. As a reaction coordinate, the distance from the center of mass of the insulin hexamer to center of mass of heavy atoms of a phenolic ligand was used. In order to restrain the phenolic ligand, a simple harmonic potential with

value of a force constant $6\frac{kcal}{mol\cdot Å^2}$ was used. The calculations were performed using 21 evenly spaced windows from 20 Å to 35 Å. Slightly longer equilibration run (5 ns) was used to ensure right position of a ligand at the reaction coordinate. Each of the production windows were simulated for 50 ns (phenol), 80ns (dopamine), and 100 ns (serotonin). The initial computational setup was following: insulin R6 hexamer with 6 phenol molecules present in the phenolic pockets, 9 000 SPC/E water molecules, 1 dopamine, phenol, or serotonin molecule in the A14/A17 binding pocket, and counterions to ensure overall electroneutrality. Finally, the results were analyzed by the WHAM procedure[90] for obtaining the potential of mean force W. As our reaction coordinate is expressed in spherical coordinates, one has to correct the results by a factor of $RTln(4\pi \boldsymbol{r}^2)$, where R is a molar gas constant, T stands for the thermodynamic temperature, and $\boldsymbol{r}$ is the distance used in each of the umbrella window. This correction is already accounted for in potentials of mean force presented later.

Because of the symmetry, there are 3 equivalent binding sites again. Therefore, the free energy of binding has to be adjusted similarly as in the case of phenolic pocket. Table 2.2. shows entropy correction to a free energy of binding in the case of the A14/A17 binding pocket.

Table 2. Entropy contribution to the free energy of binding of a phenolic ligand to the A14/A17 binding pockets according to how many pockets are unoccupied(ΔG_{symm_i}). If $i = 3$ then there is no phenolic ligand bound and the contribution is the highest. $\langle \Delta G_{symm} \rangle$ represents an average value.

i	3	2	1	$\frac{\langle \Delta G_{symm} \rangle}{kcal/mol}$
$\frac{\Delta G_{symm\ i}}{kcal/mol}$	-0.65	-0.41	0.00	-0.36

Chapter 3

Results

This chapter summarizes behavior of different conformations of insulin hexamers under various conditions. The first part covers the influence of different cations on T6 and T3R3 insulin hexamer structures. This is truly essential since the B13 glutamates cavity is located in the middle of insulin hexamer regardless of insulin hexamer conformation. The second part is dedicated to the famous phenolic pocket of insulin hexamer. In this section, I am going to focus on the question whether dopamine and serotonin neurotransmitters could work as in vivo ligands of insulin hexamers. In the last part of Chapter 3, a new, and hitherto unknown, binding pocket for studied neurotransmitters dopamine and serotonin will be introduced and described.

1. Cations and the middle B13 glutamate cavity

Ions have a profound effect on insulin hexamer structures. Insulin is one of the few known proteins in which zinc cations are used to store the hormone rather than modulating its release into the bloodstream. Here we discuss effects and binding sites of various other cations present in beta-cells, namely effects of Na, K, Zn, and Ca on the T3R3, and T6 insulin hexamer structures. R6 conformation was omitted to keep system as simple as possible because phenol molecules are needed for stabilizing the R6 structure.

T3R3, and T6 insulin hexamers were simulated in 0.5 M solutions of the following salts: NaCl, KCl, $CaCl_2$, and $ZnCl_2$ for 1.5 μs from which the ion spatial distributions were obtained for each of the cation. The results are summarized in Figure 3.1.1.

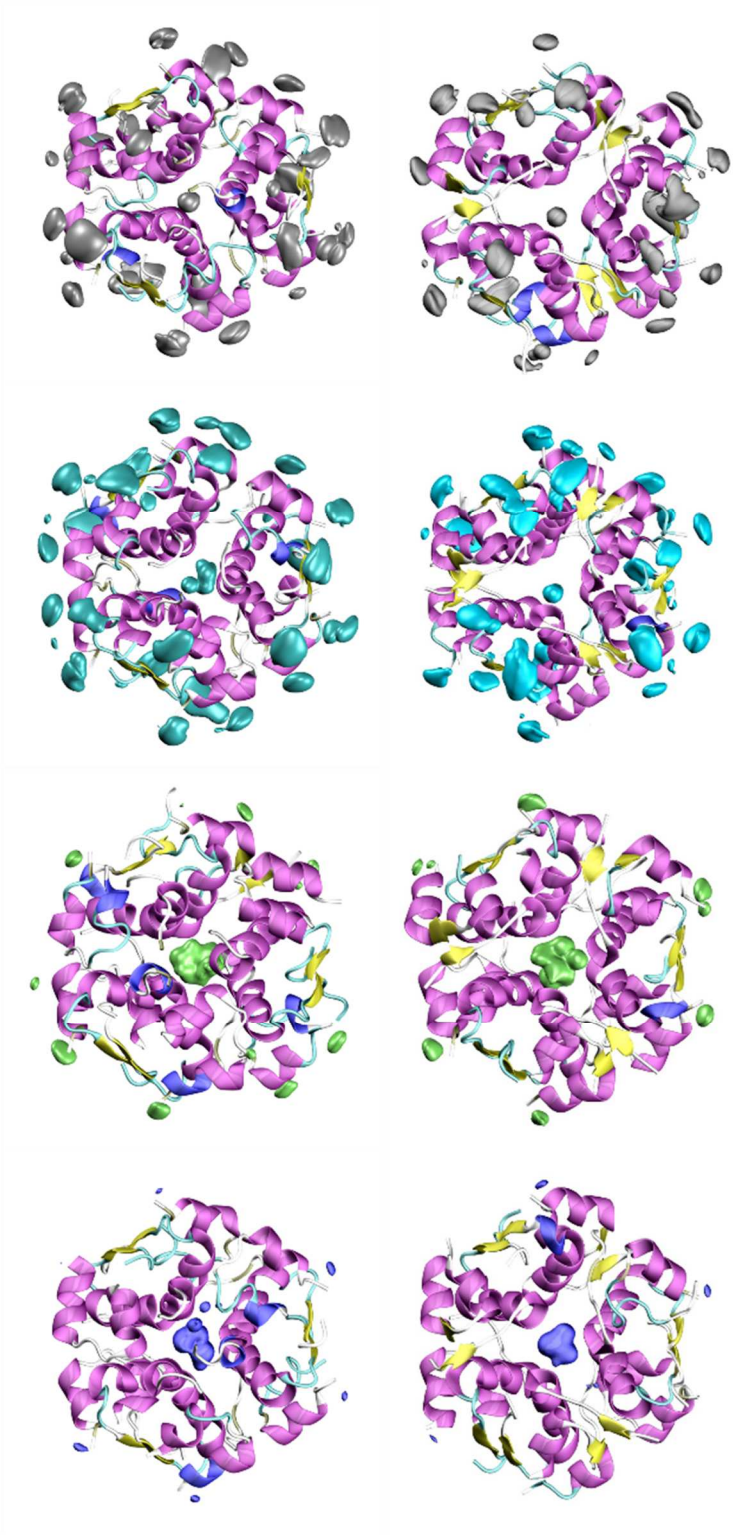

Figure 3.1.1. The cation spatial distribution functions of the T6 (top) or T3R3 (bottom) insulin hexamer in different 0.5 M salt solutions (silver – $Zn^{1.5+}$, cyan – $Ca^{1.5+}$, green – $Na^{0.75+}$, blue – $K^{0.75+}$) are shown with an isovalue of 10 (10 x higher concentration than what is present in the solution).

The first thing to be noticed is that the surface binding of cations follow the series $Zn^{1.5+} \approx Ca^{1.5+} > Na^{0.75+} > K^{0.75+}$, which is in agreement with the normal Hoffmeister ordering. However, the only significantly different place is the middle part of the insulin hexamers. The ions with the biggest charge densities ($Ca^{1.5+}$, $Zn^{1.5+}$) are localized, i.e., occupy a very specific position in the cavity. On the other hand, ions with the small charge density ($Na^{0.75+}$, $K^{0.75+}$) are more delocalized. The ion localization further affects the water structure inside the central cavity. Upon entering of a cation with high charge density in the middle of the cavity, the water molecules form a rigid structure and the diffusion is slowed down. As an example of this behavior, the water molecule map of a T6 hexamer in $ZnCl_2$ solution is presented in Figure 3.1.2.

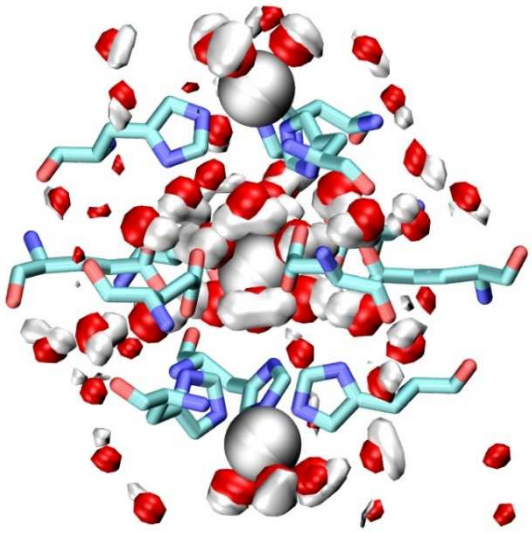

Figure 3.1.2. The spatial distribution function of water molecules in the central cavity of the insulin T6 hexamer in 0.5 M $ZnCl_2$. The density isovalue of 2 for hydrogen atoms and 3 for oxygen atoms is used. The top and the bottom B10 histidines coordinating zinc cations are shown and also the six B13 glutamates with a central cation (zinc in this case). Zinc cations are depicted in grey.

It is necessary to state that such slow diffusing water molecules cannot be found anywhere else in the insulin hexamer. The water structuring is observed up to the B10 histidines, essentially closing all 6 possible paths for another cation entering the central cavity. I also tried to quantify this phenomenon in Figure 3.1.3.

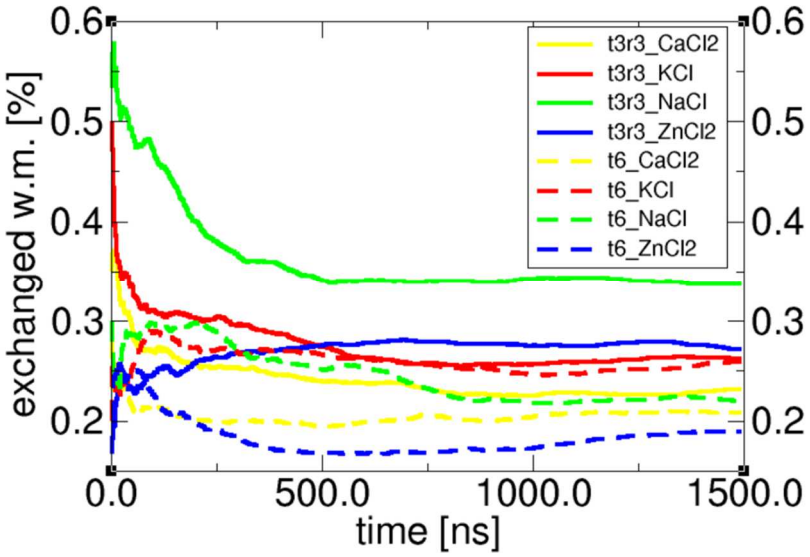

Figure 3.1.3. Cumulative mean of the exchanged water molecules in the B13 region after 2 ns. The end value on Y axis essentially shows mean percentage of water molecules which is exchanged with a bulk solution in 2 ns. T3R3 insulin hexamers are depicted by a full curve while T6 hexamers are shown by a dashed line.

The closest 35 water molecules in the B13 glutamate region were assigned as cavity water molecules. Figure 3.1.3 shows cumulative mean of a fraction of exchanged water molecules in the B13 region after 2 ns. In other words, for example the T3R3 insulin hexamer in NaCl solution (full green) – in average, 3.4 out of the 10 water molecules are exchanged with the bulk solution within 2 ns. The first thing that strikes us is that water exchange is slower in T6 insulin hexamers than in T3R3 hexamers. Another, but maybe not surprising thing, is that doubly charged cations of zinc and calcium generally lay below sodium and potassium. The only exceptions are T3R3_$ZnCl_2$ and T3R3_KCl which otherwise lay close together. Another interesting fact is the number of ions that enter the B13 cavity which is shown in a Figure 3.1.4.

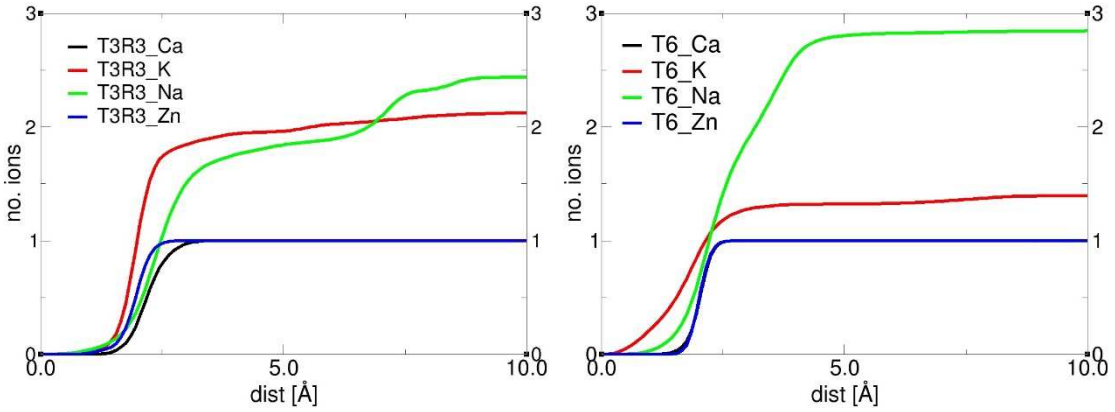

Figure 3.1.4. The integrated radial distribution function for cations from the middle of the cavity essentially showing averaged number of cations inside the B13 glutamates cavity during the 1.5 μs simulations. Results are divided into T3R3 insulin hexamer (left) and T6 insulin hexamer (right). In the case of T6 insulin hexamer, the curves of T6_Zn and T6_Ca overlap.

This figure shows an integrated radial distribution function for cavity cations, i.e., an averaged number of cations inside the cavity during the simulation. Every simulation started with one cation already inside and it is observed that none of them left (average number of cations is always higher than one). In the case of the high charge density cations ($Zn^{1.5+}$, $Ca^{1.5+}$), the number of cations remained exactly one, and in case of the low charge density cations ($Na^{0.75+}$, $K^{0.75+}$), additional cations diffused inside the cavity (value 1.4 – 2.9).

2. The phenolic pocket

In this part, I thoroughly explore the possibility of dopamine and serotonin neurotransmitters substituting phenol inside the phenolic pocket of insulin R6 hexamer. For this purpose, I firstly substituted phenols in the phenolic pockets by serotonin or dopamine molecules while assuming two possible initial orientations of the neurotransmitters as shown in Figure 3.2.1.

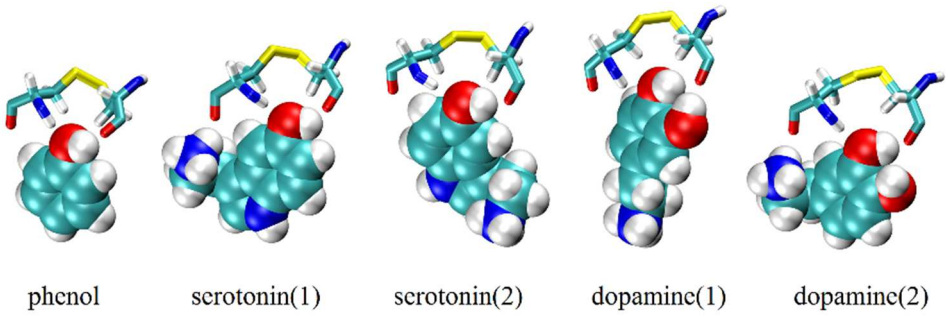

Figure 3.2.1. Phenol together with serotonin and dopamine neurotransmitters are depicted in the phenolic pocket. Two initial orientations of serotonin, and dopamine were considered.

These systems were then propagated in time without any constraints. The resulting root mean square deviations (RMSD) of the backbone of the insulin R6 hexamers with different phenolic ligands are presented in Figure 3.2.2.

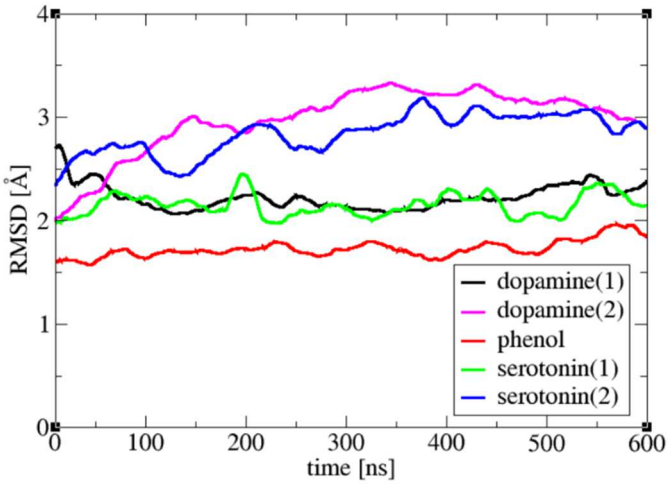

Figure 3.2.2. Root mean square deviations of a protein backbone from the insulin R6 hexamer crystal structure with different phenolic ligands (phenol, dopamine, and serotonin). Both for serotonin and dopamine, there were two starting orientations. The crystal structure with phenols was the reference geometry for the RMSD evaluations[75].

As expected, the insulin R6 hexamer containing phenol molecules has the lowest RMSD of ~1.6–1.8 Å indicating that the system does not significantly deviate from the crystal structure during the simulation. Starting from the first orientation, both serotonin and

dopamine behave similarly to phenol with backbone RMSD just above 2 Å while, starting from the second orientation, leading to somewhat higher RMSDs (~2.8 Å). Closer picture can be obtained upon inspecting RMSDs pertinent directly to the binding pocket. Figure 3.2.3. thus shows RMSDs with respect to the phenolic ligand and the A6 cysteine backbone oxygen.

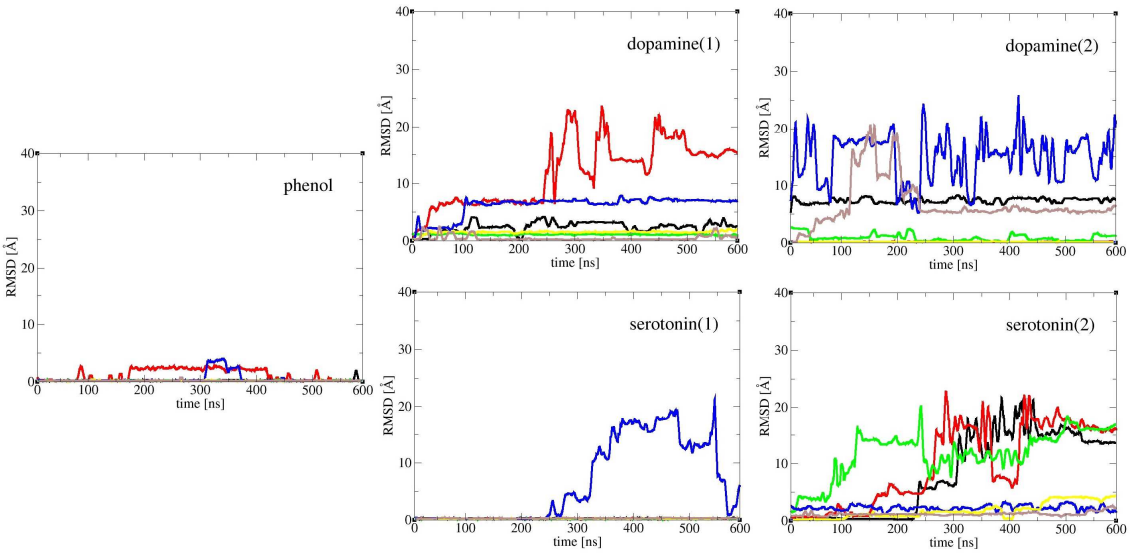

Figure 3.2.3. Root mean square deviations of all 6 phenolic ligands from the phenol binding pockets in the insulin R6 hexamers. Each line corresponds to RMSD of one phenolic ligand from its starting position with respect to the phenolic ligand and A6 cysteine backbone oxygen. The zero value of RMSD shows direct binding on phenolic ligand to insulin R6 hexamer through hydrogen bond while any significant increase indicates breaking of this hydrogen bond.

As expected, phenol essentially stays in its binding pocket again. Two in six phenols transiently break the hydrogen bonds with protein A6 and A11 amino acids. However, they remain in the binding pocket and eventually reestablish the original geometry. Starting from the first geometry, serotonins also stay in the binding pocket with the exception of one molecule which temporarily leaves it (but eventually diffuses back). This demonstrates the reversibility of the binding/dissociation process. In the other cases however, the neurotransmitter molecules mostly leave the original binding site. This suggests that the phenolic pocket is rather tight and unfit for strong binding of dopamine

in either orientation and serotonin in the second one. In Figure 3.2.4., I compare binding of phenol and serotonin (in the first geometry) in the phenolic pocket.

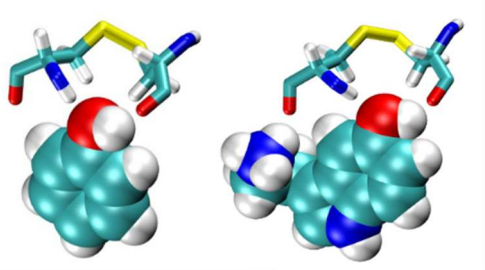

Figure 3.2.4. Phenol (left) and substituted serotonin(1) (right) in the phenolic pocket. The binding pocket is mainly composed of hydrophobic amino acids (not shown) and A6 and A11 amino acids which provide crucial binding through hydroxyl group.

Based on the figure, it is clear that the charged amino group of serotonin provides another contribution toward the binding energy by binding to the A11 backbone oxygen atom. These results suggest that serotonin may be a viable candidate for a substitution of phenol in the in vivo conditions. In order to explore this possibility further, I performed free energy calculations of binding of phenol, serotonin or dopamine to the phenolic pocket using the thermodynamic integration method. The final results are presented in Table 3.2.1. From Table 3.2.1., we see that while phenol is the strongest binder among the three molecules, the strength of serotonin binding is comparable. In contrast, dopamine, which does not fit into the phenolic pocket, is not bound at all. Free energies of successive transformations steps are summarized in Table 3.2.2.

Table 3.2.1. Standard free energies of binding of dopamine, serotonin, and phenol molecules to the phenol binding pocket in the insulin R6 hexamer with corresponding dissociation constants. ΔG_b° – standard free energy of binding, K_D – dissociation constant.

	ΔG°_b [kcal/mol]	K_D [M]
phenol	-4.49 ± 1.55	5.4E-04
dopamine	1.10 ± 1.73	6.3E-00
serotonin	-4.24 ± 1.87	8.1E-04

Table 3.2.2. Free energies of binding of dopamine, serotonin, and phenol molecules to the phenolic binding pocket are divided into several steps.

$\frac{\Delta G}{[kcal/mol]}$	phenol→serotonin	phenol→dopamine	$\frac{\Delta G}{[kcal/mol]}$	phenol
ΔG_1	23.21 ± 0.08	23.21 ± 0.08	ΔG_7	4.01 ± 0.25
ΔG_2	-3.28 ± 0.45	-3.28 ± 0.32	ΔG_8	25.11 ± 0.40
ΔG_3	-61.63 ± 0.91	-29.16 ± 0.68	ΔG_9	7.15 ± 1.46
ΔG_4	60.28 ± 0.07	32.04 ± 0.03	ΔG_{11}	-11.60
ΔG_5	0.69 ± 0.21	1.80 ± 0.06	ΔG_{12}	-19.03 ± 0.07
ΔG_6	-19.02 ± 0.02	-19.02 ± 0.02	ΔG_{13}	-1.80 ± 0.21
$\Delta\Delta G_{disp}$	-2.58	-1.47	ΔG_{PHN}	-3.84 ± 1.55
$\Delta\Delta G_{ele}$	2.83	7.06	$<\Delta G_{symm}>$	-0.65
$\Delta\Delta G_{1\text{-}2}$	0.25 ± 1.04	5.59 ± 0.76	$\Delta G°_{PHN}$	-4.49 ± 1.55

ΔG_1 – free energy difference of charge decoupling of phenol molecule inside the phenolic pocket, ΔG_2 – free energy difference of mutation of phenol into dopamine/serotonin inside the phenolic pocket, ΔG_3 – free energy difference of introducing charges on dopamine/serotonin inside the phenolic pocket, ΔG_4 – free energy difference of charge decoupling of dopamine/serotonin in the bulk solution, ΔG_5 – free energy difference of van der Waals transformation of dopamin/serotonin into phenol in the bulk solution, ΔG_6 – free energy difference of introducing charges on phenol in the bulk solution, $\Delta\Delta G_{disp}$ – dispersion contribution to the free energy difference between phenol/neurotransmitter, $\Delta\Delta G_{ele}$ – electrostatic contribution to the free energy difference between phenol/neurotransmitter, $\Delta\Delta G_{1-2}$ – overall free energy difference between phenol/neurotransmitter.

ΔG_7 – free energy difference of introducing the restraints to the phenol molecule inside the phenolic pocket, ΔG_8 – free energy difference of charge decoupling of phenol inside the phenolic pocket, ΔG_9 – free energy difference of van der Waals decoupling of phenol inside the phenolic pocket, ΔG_{11} – free energy difference of releasing restrains (analytically, correction to the standard state), ΔG_{12} – free energy difference of charge decoupling of phenol in the bulk solution, ΔG_{13} – free energy difference of van der Waals decoupling of phenol in the bulk solution, ΔG_{PHN} – overall standard free energy of binding without the entropy contribution originating from the symmetry , $\langle\Delta G_{symm}\rangle$ – average symmetry contribution to the free energy of binding, $\Delta G°_{PHN}$ – overall standard free energy of binding of phenol to the phenolic pocket.

When we compare electrostatic ($\Delta\Delta G_{ele}$) and van der Waals contributions ($\Delta\Delta G_{disp}$) with the overall free energy of binding of serotonin over phenol, we see that serotonin does not gain as much from electrostatics as phenol. Nevertheless, the dispersion term does

compensate for this loss. In the case of dopamine, the electrostatic term is largely positive and the dispersion term is insufficient to compensate.

3. The A14/A17 binding pocket

While dopamine does not bind into the phenolic pocket, my simulations indicate an existence of new binding sites for this molecule on the surface of the R6 hexamer. The resulting spatial distributions of dopamine corresponding to these three equivalent binding sites are depicted in Figure 3.3.1. For comparison, I also show spatial distributions of serotonin which occupies similar sites but seems to be more loosely distributed on the surface of the R6 hexamer.

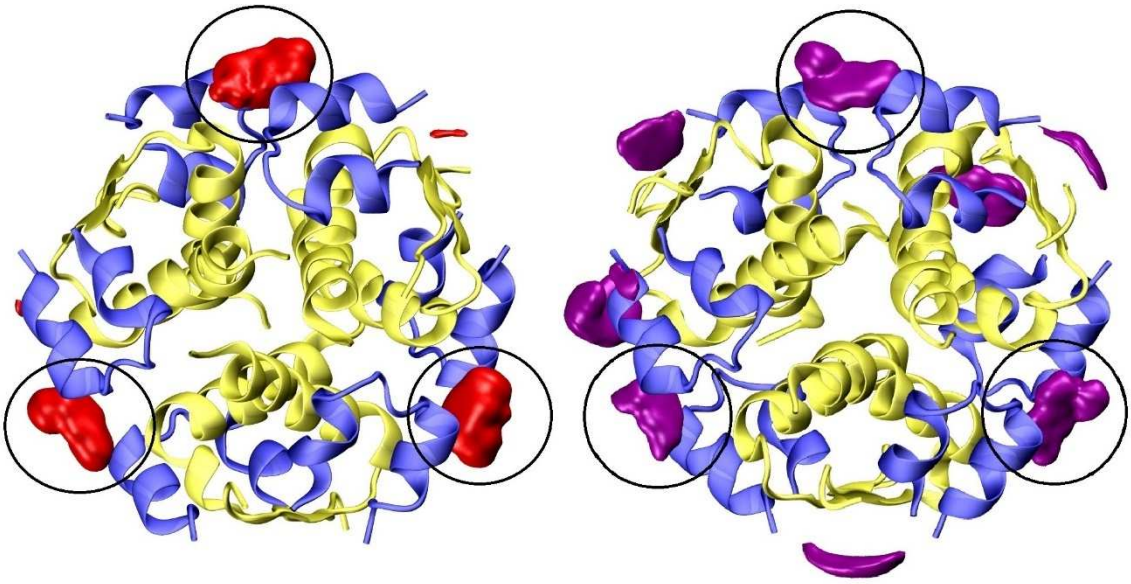

Figure 3.3.1. Dopamine (left, red) and serotonin (right, magenta) spatial distributions around the insulin R6 hexamer using the same isodensity value (~50x). Chain A is shown in blue color while a chain B is shown in yellow color. Black circles depict the A14/A17 binding pockets.

These three binding sites for dopamine, which are identical due to the symmetry of insulin hexamer, are also present in the T3R3 or the T6 conformation of the insulin hexamer. A detailed view on one of these dopamine binding sites is shown in Figure 3.3.2. The binding pocket is formed between two adjacent insulin monomers. The strong binding to dopamine is mediated by a hydrogen bond, a salt bridge, and hydrophobic interactions, including the crucial A14 tyrosine and A17 glutamate amino acids. As a result, I denote this new site as the A14/A17 binding pocket.

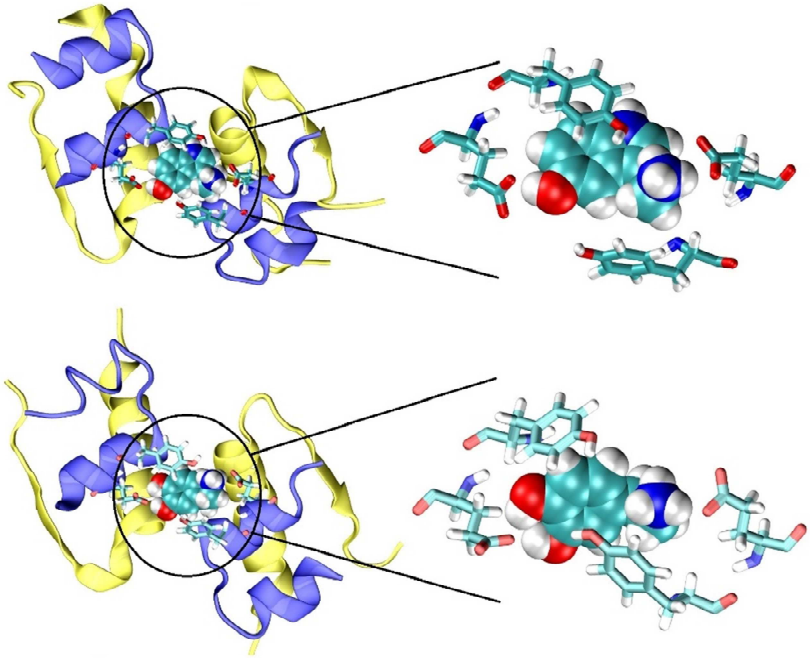

Figure 3.3.2 Dopamine (bottom) and serotonin (top) in a new A14/A17 binding pocket. It is formed by two adjacent insulin monomers, where chain A is depicted in blue color while the chain B is depicted in yellow color.

These results suggest that dopamine in particular may bind to these sites with high affinity. To this end, free energies of binding dopamine as well as serotonin (and phenol for comparison) to the A14/A17 binding site were evaluated using the umbrella sampling method. The resulting binding free energy curves are presented in Figure 3.3.3. with the binding constants summarized in Table 3.3.1. Upon applying symmetry and volume entropy corrections, we see that it is dopamine distributions which are attached the strongest to the new binding site while the corresponding binding of serotonin (as well as phenol) is rather weak.

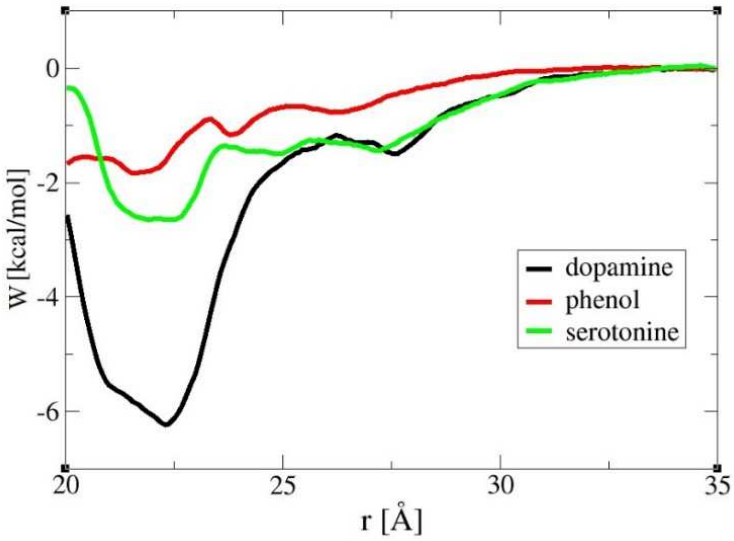

Figure 3.3.3. Free energy profiles of phenolic ligands entering the A14/A17 binding pocket in a bulk solution. The A14/A17 binding pocket is located at ~21-23 Å at the reaction coordinate while ~35 Å represents the bulk solution.

Table 3.3.1. Standard free energies of binding of dopamine, serotonin, and phenol molecules to the A14/A17 binding pocket with corresponding dissociation constants. W_0 – free energy difference form umbrella sampling calculations, $\langle \Delta G_{symm} \rangle$ – symmetry contribution, ΔG_{vol} – volume entropy contribution (correction to standard state (1M)), ΔG_b° – overall standard free energy of binding, K_D – corresponding dissociation constant.

$A14/A17\ pocket$	$\frac{W_0}{kcal/mol}$	$\frac{\langle \Delta G_{symm} \rangle}{kcal/mol}$	$\frac{\Delta G_{vol}}{kcal/mol}$	$\frac{\Delta G_b^\circ}{kcal/mol}$	$\frac{K_D}{[M]}$
phenol	-1.83 ± 0.28	-0.36	1.38 ± 0.22	-0.81	2.59E-01
dopamine	-6.23 ± 0.67	-0.36	3.2 ± 0.2	-3.24	4.38E-03
serotonin	-2.65 ± 0.55	-0.36	2.0 ± 0.1	-0.92	2.13E-01

It is also important to address the question whether calcium cations, which are abundantly present in beta cells, could affect the interactions with the binding sites above. This is very topical since an important contribution to binding comes from the A17 glutamates. Figure 3.3.4. shows the resulting spatial of neurotransmitters and calcium cations on the insulin R6 hexamer in 0.5 M $CaCl_2$ aqueous solution.

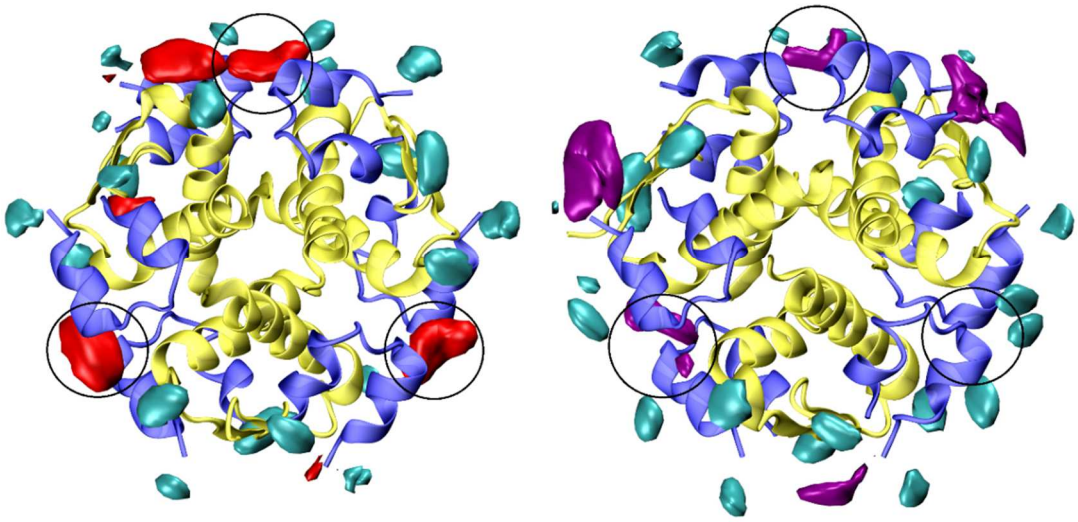

Figure 3.3.4. Dopamine (left, red) and serotonin (right, purple) spatial distribution functions around the insulin R6 hexamer in a 0.5 M $CaCl_2$ solution using the same isodensity value (~50x). Spatial distribution functions of calcium cations are depicted in cyan color (isodensity value 15x). Chain A is shown in blue color; chain B is shown in yellow color. Black circles depict the A14/A17 binding pockets.

Even though the neurotransmitter distributions are quantitatively altered in the calcium chloride solution compared with pure water (the comparison in Figure 3.3.1.), the same binding patterns to the A14/A17 pocket can still be seen. However, in the case of serotonin in particular, the resulting binding is weaker. This weakening is likely due to the relatively strong binding of Ca^{2+} cations to each of the A17 glutamates. In other words, binding of dopamine to the new A14/A17 biding pocket effectively inhibits binding of calcium to the A17 glutamates whereas binding of calcium to the A17 glutamates inhibits binding of serotonin.

Chapter 4

Discussion

1. Ions and the inner cavity

The experimental results proved that cation concentrations inside the secretory granules are dramatically different from anywhere else. As the secretory granules contain high amount of insulin, a question regarding specific binding of cations to the insulin hexamers arose. I investigated this phenomenon by means of molecular dynamics simulations.

Binding of cations to the surface of hexamers follows classical Hoffmeister series. However, behavior of cations in the middle of the B13 glutamates cavity is completely different. The high charge density cations (zinc and calcium) stabilize the water structure inside the cavity and consequently prevent other cations from entering. The low charge density cations (sodium and potassium) do not stabilize the structure strongly enough to hinder all possible paths which eventually leads to the presence of ~2–3 cations at the same time. It seems that either the repulsion of two divalent cations inside the inner cavity is too high or that the effective radius of divalent cations is too big for entering the B13 cavity. Sadly, the results obtained are purely theoretical and no comparative experiment has been performed so far.

2. Neurotransmitters

An idea that neurotransmitters could somehow be connected with insulin hexamer conformation is very thrilling. To this day, they are no substances serving as stabilizing agents which are naturally present in our bodies. In order to shed more light on this idea,

a possible substitution of phenol in the phenolic pocket by dopamine or serotonin was investigated.

At first I substituted phenol molecules for dopamine or serotonin and then performed simple classical MD simulations to get the initial idea. The first results showed that only binding strength of serotonin was similar to the phenols'. These results were further supported by the free energy calculations. One phenolic ligand was mutated between phenol, serotonin and dopamine to obtain the difference in free energies of binding between these considered phenolic ligands. Results clearly show that serotonin is similar to phenol in terms of thermodynamics. However, the strength of dopamine binding is by ~5.6 kcal/mol lower. Furthermore, the absolute free energies of binding of considered phenolic ligands were calculated which lead to standard free energies of binding of phenol (-4.5 ± 1.6 kcal/mol), serotonin (-4.2 ± 1.9 kcal/mol), and dopamine (1.1 ± 1.7 kcal/mol).

During the research, a new, and previously unknown, binding pocket for dopamine and serotonin was found. The new binding pocket was named after the interactions involved in binding – a A14/A17 binding pocket. I have described the A14/A17 binding pocket both in structural and energetics terms. From the structural point of view, the A14/A17 binding pocket is located on the surface of hexamer regardless its conformation (T6/T3R3/R6). Moreover, due to the three-fold symmetry of insulin hexamer there are three identical binding pockets. From the thermodynamics point of view, I have shown that dopamine binds the strongest while serotonin and phenol not so much. Standard free energies of binding are as follows: dopamine (-3.2 kcal/mol), serotonin (-0.9 kcal/mol), and phenol (-0.8 kcal/mol).

As previously stated, the research was carried out in collaboration with J. Jiracek et al, and M. Brzozowski et al. M. Brzozowski who succeeded in crystalizing insulin hexamer in the presence of serotonin. The neutron scattering analysis revealed the T3R3 conformation with 6 serotonin molecules present. Three of them were in the phenolic pocket and three of them were in the new binding pockets which had been predicted by molecular dynamics simulation before the experiment. The structural arrangement of serotonin inside the phenolic pocket and comparison to my molecular dynamics simulation is shown in a Figure 4.1.

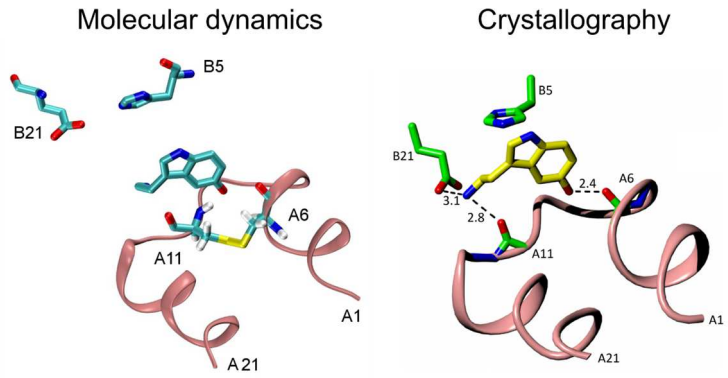

Figure 4.1. Comparison of structural arrangement of serotonin inside the phenolic pocket predicted by molecular dynamics simulations with the experimental results (crystallography).

In the case of the A14/A17 binding pocket, I predicted two degenerate low energy laying structures. Comparison with the experimental result is shown in Figure 4.2.

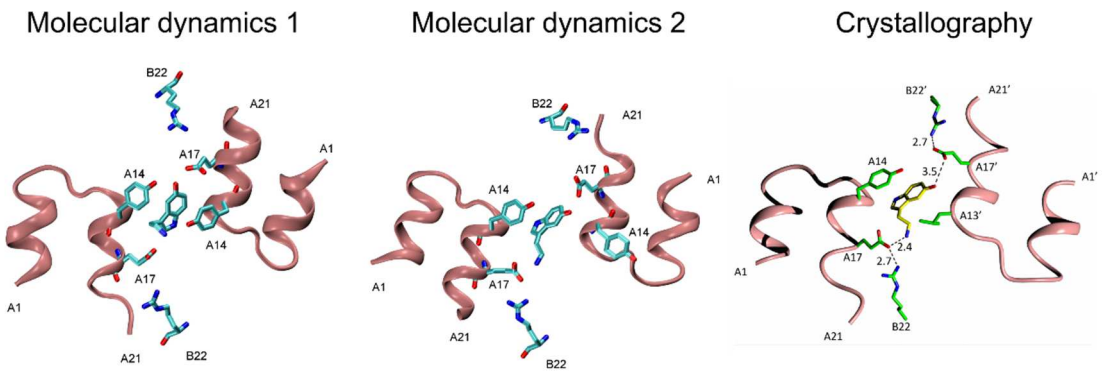

Figure 4.2. The newly discovered A14/A17 binding pocket for neurotransmitters and two possible orientations predicted by the molecular dynamics simulation in comparison with the experimental result.

From the structural point of view, the agreement is very good. From the energetics point of view however, the MD simulations predicted that serotonin binds only non-specifically. Clearly more theoretical investigation has to be done in this topic. Moreover,

my simulations predict that not serotonin, but dopamine is the ultimate binder to this A14/A17 binding pocket. Unfortunately, this has not been proved yet experimentally as there are problems with dopamine's molecule stability.

Another piece of information which support this work is experiment done by J. Jiracek et al. There are three possible conformations (T6, T3R3, R6) of insulin hexamer and all hexamers contain two B10 coordinated zinc cations. Various ligands can bind to these zinc cations. However, only if that part of the hexamer is present in R3 form. Therefore, the following experiment was performed. A phenolic ligand was titrated to a solution of insulin T6 hexamers in the presence of a ligand which binds to the B10 zinc cation. If the hexamers are in T conformation, there is no binding. Nevertheless, if they are in R conformation, the binding occurs. This binding can be monitored by absorption spectroscopy. In this experiment, a 4-hydroxy-3-nitrobenzoic acid (4H3N) was used. This molecule change the absorption near 441 nm as it binds. And so if titrated ligand changes conformation of insulin hexamer, we will see it as a change in the absorption. Figure 4.3. shows the ideas of the experiment and results. The first step is to measure a titration curve (against blank titration) which can be fitted using the Hill equation. This equation will give us three parameters – a strength of response ΔA_{max}, an apparent dissociation constant Kd, and the Hill coefficient. What we see from this experiment is that only phenol and dopamine display change in the absorption and therefore, only these two ligands bind to the phenolic pocket and change the conformation of insulin hexamer. This is in agreement with the molecular dynamics simulations. Moreover, MD simulations predict that phenol binds stronger than serotonin, which is also in agreement with the experiment. However, one should be careful about interpreting these results as the binding of a phenolic ligand to the phenolic pocket is observed only indirectly.

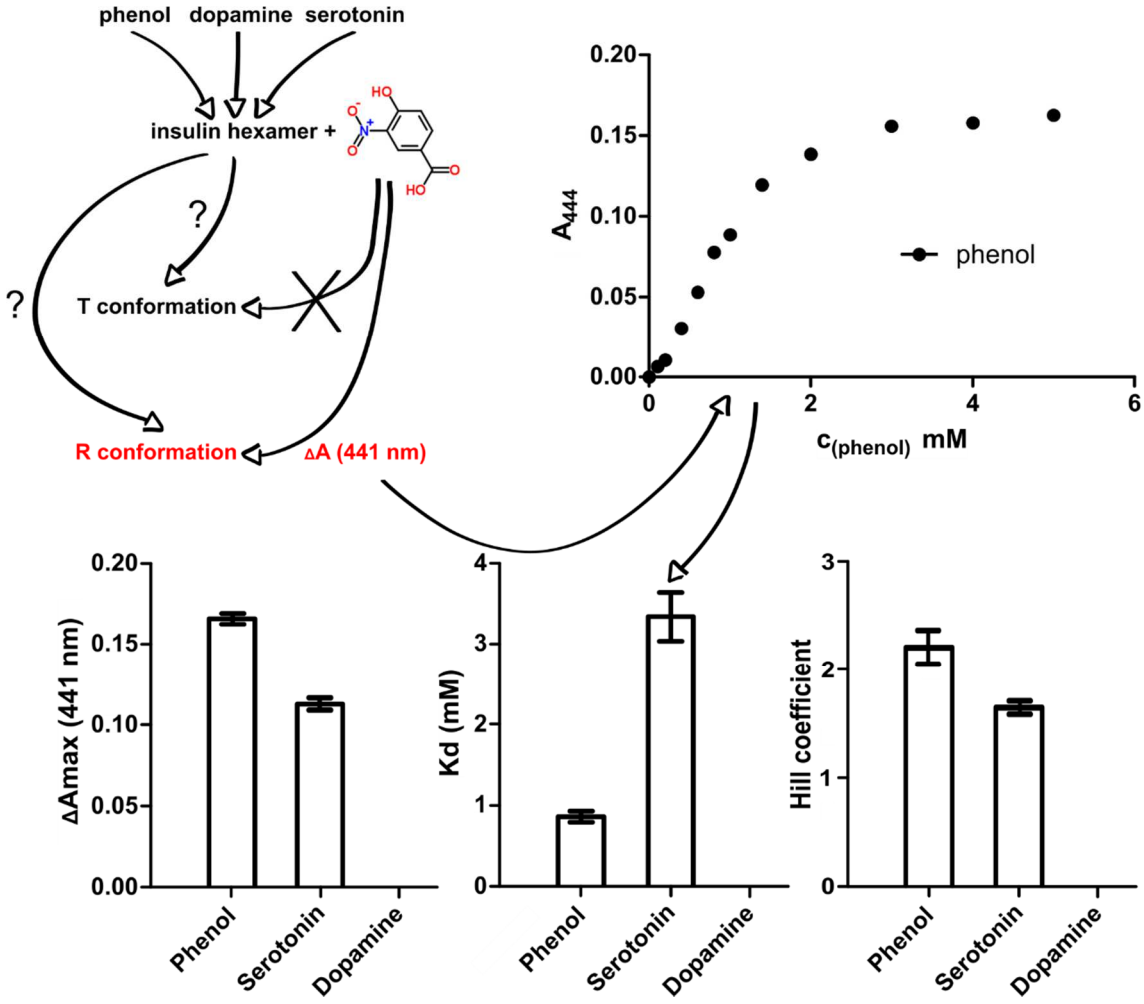

Figure 4.3. Results of an experiment conducted by J. Jiracek et al. Left top: Illustration of the experiment arrangement – a phenolic ligand (phenol/serotonin/dopamine) is titrated to insulin hexamer solution with 4-hydroxy-3-nitrobenzoic acid (4H3N). 4H3N binds to the proximity of B10 zinc cation but only to the R conformer. This binding changes absorption of 4H3N at 441 nm. A titration curve is measured (right top) and fitted using the Hill equation. From this technique, the strength of response ΔAmax, dissociation constant Kd, and the Hill coefficient was obtained. Note that no response was acquired in the case of dopamine titration.

Chapter 5

Conclusions

The results of my computer experiments show clear difference between doubly charged cations (zinc and calcium) compared with singly charged cations (sodium and potassium). However, there are still many unanswered questions like whether a divalent cation present in the B13 cavity prevents other monovalent cations from entering and furthermore, whether it is possible for a divalent cation to enter the cavity while monovalent cation/cations is/are present. The next fact is that only T6 and T3R3 hexamers were investigated and therefore, the R6 hexamer should also be probed. Another interesting question is why zinc is so important to insulin hexamer although this is probably a question concerning the QM/MM methods. All these questions were asked for now since proper parametrization of all ions is essential. A group of P. Jungwirth is now developing a force field for zinc cation and we have decided to wait until we have proper description of this crucial ion regarding this problem.

Concerning neurotransmitters overall in both MD experimental simulations, the results suggest that serotonin might work as in vivo phenolic ligand while dopamine does not. Nevertheless, all the presented results are indirect and the final answer about the actual conformation of insulin hexamer in in vivo conditions still cannot be answered. However, this is the first progress in this topic after several decades. Another questions, which must be raised now, include: How does the transition from T to R conformation happen? Why is zinc cation so unique to insulin hexamer? Moreover, we know that arginine inhibits hexamerization (according to recent M. Brzozowski's experimental results) while an addition of serotonin overcomes this inhibition. Because arginine is a residual product of insulin hexamer formation, it is quite abundant in secretory granules. What is thus the overall interplay between cations – neurotransmitters – arginine? We hope that these questions will be answered soon.

Chapter 6

List of abbreviations

HF – Hartree–Fock

DFT – density functional theory

NPA – natural population analysis

RESP – restrained electrostatic potential fit

WHAM – weighted histogram analysis method

NMR – nuclear magnetic resonance

LINCS – A linear constraint solver for molecular simulations

L-DOPA – L-3, 4-dihydroxyphenylalanine

DPN – dopamine

SEN – serotonin

PHN – phenol

MD – molecular dynamics

T6 – insulin hexamer with all monomers in T conformation

T3R3 – intermediate form of insulin hexamer between T6 and R6 conformation

R6 – insulin hexamer with all monomers in R conformation

Literature

1. Bliss, M. The discovery of insulin: the inside story. *Am. Inst. Hist. Pharm.* **16,** 93–99 (1997).

2. Smith, G. D., Swenson, D. C., Dodson, E. J., Dodson, G. G. & Reynolds, C. D. Structural stability in the 4-zinc human insulin hexamer. *Proc. Natl. Acad. Sci.* **81,** 7093–7097 (1984).

3. Smith, G. D., Pangborn, W. A. & Blessing, R. H. The structure of T6 human insulin at 1.0 A resolution. *Acta Crystallogr. D. Biol. Crystallogr.* **59,** 474–482 (2003).

4. Scott, D. A. & Fisher, A. M. The Insulin and the Zinc Content of Normal and Diabetic Pancreas. *J. Clin. Invest.* **17,** 725–728 (1938).

5. Goldman, J. & Carpenter, F. H. Zinc binding, circular dichroism, and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives. *Biochemistry* **13,** 4566–4574 (1974).

6. Dunn, M. F., Pattison, S. E., Storm, M. C. & Quiel, E. Comparison of the zinc binding domains in the 7S nerve growth factor and the zinc-insulin hexamer. *Biochemistry* **19,** 718–725 (1980).

7. Derewenda, U. *et al.* Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer. *Nature* **338,** 594–6 (1989).

8. Ciszak, E. & Smith, G. D. Crystallographic Evidence for Dual Coordination Around Zinc in the T3R3 Human Insulin Hexamer. *Biochemistry* **33,** 1512–1517 (1994).

9. Brader, M. L. & Dunn, M. F. Insulin Hexamers - New Conformations and Applications. *Trends Biochem. Sci.* **16,** 341–345 (1991).

10. Baker, E. N. *et al.* The Structure of 2Zn Pig Insulin Crystals at 1.5 A overset Resolution. *Philos. Trans. R. Soc. B Biol. Sci.* **319,** 369–456 (1988).

11. Sreekanth, R., Pattabhi, V. & Rajan, S. S. Metal induced structural changes observed in hexameric insulin. *Int. J. Biol. Macromol.* **44,** 29–36 (2009).

12. Nicholson, J., Perkins, L. & Korber, F. 1M5A: Crystal Structure of 2-Co(2+)-Insulin at 1.2A Resolution. *Protein Data Bank* doi:http://dx.doi.org/10.2210/pdb1m5a/pdb

13. Kadima, W. Role of Metal Ions in the T- To R-Allosteric Transition in the Insulin Hexamer. *Biochemistry* **38,** 13443–13452 (1999).

14. Gursky, O., Li, Y. L., Badger, J. & Caspar, D. L. D. Monovalent Cation Binding to Cubic Insulin Crystals. *Biophys. J.* **61,** 604–611 (1992).

15. Brader, M. L., Borchardt, D. & Dunn, M. F. The T to R transition in the copper(II)-substituted insulin hexamer. Anion complexes of the R-state species exhibiting type 1 and type 2 spectral characteristics. *Biochemistry* **31,** 4691–4696 (1992).

16. Sudmeier, J., Bell, S., Storm, M. & Dunn, M. Cadmium-113 nuclear magnetic resonance studies of bovine insulin: two-zinc insulin hexamer specifically binds calcium. *Science (80-.).* **212,** 560–562 (1981).

17. Hill, C. P., Dauter, Z., Dodson, E. J., Dodson, G. G. & Dunn, M. F. X-ray structure of an unusual calcium site and the roles of zinc and calcium in the assembly, stability, and storage of the insulin hexamer. *Biochemistry* **30,** 917–924 (1991).

18. Storm, M. C. & Dunn, M. F. The Glu(B13) carboxylates of the insulin hexamer form a cage for cadmium and calcium ions. *Biochemistry* **24,** 1749–1756 (1985).

19. Palmieri, R., Lee, R. W. K. & Dunn, M. F. Proton fourier transform NMR studies of insulin: coordination of calcium to the Glu(B13) site drives hexamer assembly and induces a conformation change. *Biochemistry* **27,** 3387–3397 (1988).

20. Coffman, F. D. & Dunn, M. F. Insulin-metal ion interactions: the binding of divalent cations to insulin hexamers and tetramers and the assembly of insulin hexamers. *Biochemistry* **27,** 6179–6187 (1988).

21. Ferrari, D., Diers, J. R., Bocian, D. F., Kaarsholm, N. C. & Dunn, M. F. Raman signatures of ligand binding and allosteric conformation change in hexameric insulin. *Biopolymers* **62,** 249–60 (2001).

22. Choi, W. E., Brader, M. L., Aguilar, V., Kaarsholm, N. C. & Dunn, M. F. Allosteric transition of the insulin hexamer is modulated by homotropic and heterotropic interactions. *Biochemistry* **32,** 11638–11645 (1993).

23. Brader, M. L., Kaarsholm, N. C., Lee, R. W. K. & Dunn, M. F. Characterization of the r-state insulin hexamer and its derivatives. The hexamer is stabilized by heterotropic ligand binding interactions. *Biochemistry* **30,** 6636–6645 (1991).

24. Bonaccio, M., Ghaderi, N., Borchardt, D. & Dunn, M. F. Insulin allosteric behavior: detection, identification, and quantification of allosteric states via 19F NMR. *Biochemistry* **44,** 7656–68 (2005).

25. Hassiepen, U., Federwisch, M., Mülders, T. & Wollmer, A. The lifetime of insulin hexamers. *Biophys. J.* **77,** 1638–54 (1999).

26. Rahuel-Clermont, S., French, C. A., Kaarsholm, N. C., Dunn, M. F. & Chou, C. I. Mechanisms of stabilization of the insulin hexamer through allosteric ligand interactions. *Biochemistry* **36,** 5837–45 (1997).

27. Chimienti, F., Favier, A. & Seve, M. ZnT-8, a pancreatic beta-cell-specific zinc transporter. *Biometals* **18,** 313–7 (2005).

28. Nicolson, T. J. *et al.* Insulin storage and glucose homeostasis in mice null for the granule zinc transporter

ZnT8 and studies of the type 2 diabetes-associated variants. *Diabetes* **58,** 2070–83 (2009).

29. Chimienti, F., Devergnas, S., Favier, A. & Seve, M. Identification and Cloning of a -Cell-Specific Zinc Transporter, ZnT-8, Localized Into Insulin Secretory Granules. *Diabetes* **53,** 2330–2337 (2004).

30. Bryant, C. *et al.* Acid stabilization of insulin. *Biochemistry* **32,** 8075–8082 (1993).

31. Hutton, J. C. The Internal pH and Membrane Potential of the Insulin Secretory Granule. *Biochem. J.* **204,** 171–178 (1982).

32. Greider, M. H. Isolation and Properties of Secretory Granules from Rat Islets of Langerhans: II. Ultrastructure of the Beta Granule. *J. Cell Biol.* **41,** 162–166 (1969).

33. Steiner, D. F., Park, S.-Y., Støy, J., Philipson, L. H. & Bell, G. I. A brief perspective on insulin production. *Diabetes. Obes. Metab.* **11 Suppl 4,** 189–96 (2009).

34. Hutton, J. C., Penn, E. J. & Peshavaria, M. Isolation and Characterizataion of Insulin Secretory Granules from a Rat Islet Cell Tumor. *Diabetologia* **23,** 365–373 (1982).

35. Gold, G. & Grodsky, G. M. Kinetic aspects of compartmental storage and secretion of insulin and zinc. *Experientia* **40,** 1105–1114 (1984).

36. Michael, D. J. Pancreatic -Cells Secrete Insulin in Fast- and Slow-Release Forms. *Diabetes* **55,** 600–607 (2006).

37. Foster, M. C., Leapman, R. D., Li, M. X. & Atwater, I. Elemental Composition of Secretory Granules in Pancreatic Beta-cells Islets of Langerhans. *Biopfysical J.* **64,** 525–532 (1993).

38. Hutton, J. C., Penn, E. J. & Peshavaria, M. Low Molecular Weight Constituents of Isolated Insulin Secretory Granules - Bivalent Cations, Adenine Nucleotides and Inorganic Phosphate. *Biochem. J.* **210,** 297–305 (1983).

39. Vinkenborg, J. L. *et al.* Genetically encoded FRET sensors to monitor intracellular Zn2+ homeostasis. *Nat. Methods* **6,** 737–40 (2009).

40. Falck, B. & Hellman, B. Evidence for the presence of biogenic amines in pancreatic islets. *Experientia* **19,** 139–140 (1963).

41. Hansen, S. E. & Hedeskov, C. J. Simulataneous Determination of Content of Serotonin, Dopamine, Noradrenaline and Adrenaline in Pancreatic Islets Isolated from Fed and Starved Mice. *Acta Endocrinol. (Copenh).* **86,** 820–832 (1977).

42. Gylfe, E. Serotonin as marker for the secretory granules in the pancreatic beta-cell. *Acta Physiol. Scand. Suppl.* **452,** 125–128 (1977).

43. Braun, M. *et al.* Corelease and differential exit via the fusion pore of GABA, serotonin, and ATP from LDCV

in rat pancreatic beta cells. *J. Gen. Physiol.* **129,** 221–31 (2007).

44. Blum, I. *et al.* The influence of meal composition on plasma serotonin and norepinephrine concentrations. *Metabolism* **41,** 137–140 (1992).

45. Goldstein, D. S., Eisenhofer, G. & Kopin, I. J. Sources and significance of plasma levels of catechols and their metabolites in humans. *J. Pharmacol. Exp. Ther.* **305,** 800–11 (2003).

46. Lundquist, I., Panagiotidis, G. & Stenstrom, A. Effect of L-DOPA Administration on Islet Monoamine Oxidase Activity and Glucose-Induced Insulin Release in the Mouse. *Pancreas* **6,** 522–527 (1991).

47. Lundquist, I., Ahéh, B., Hansson, C. & Håkanson, R. Monoamines in Pancreatic Islets of Guinea Pig, Hamster, Rat, and Mouse Determined by High Performance Liquid Chromatography. *Pancreas* **4,** 662–667 (1989).

48. Lernamark, A. Significance of 5-Hydroxytryptamine for Insulin Secretion in Mouse. *Horm. Metab. Res.* **3,** 305 (1971).

49. Saisho, Y. *et al.* Relationship between pancreatic vesicular monoamine transporter 2 (VMAT2) and insulin expression in human pancreas. *J. Mol. Histol.* **39,** 543–51 (2008).

50. Simpson, N. *et al.* Dopamine-mediated autocrine inhibitory circuit regulating human insulin secretion in vitro. *Mol. Endocrinol.* **26,** 1757–72 (2012).

51. Ustione, A. & Piston, D. W. Dopamine synthesis and D3 receptor activation in pancreatic β-cells regulates insulin secretion and intracellular [Ca(2+)] oscillations. *Mol. Endocrinol.* **26,** 1928–40 (2012).

52. Lindstrom, P. Further Studies on 5- Hydroxytryptamine Transport in Pancreatic Islets and Isolated Beta-Cells. *Br. J. Pharmacol.* **73,** 385–391 (1981).

53. Lindstrom, P., Sehlin, J. & Taljedal, I. B. Characteristics of 5-Hydroxytryptamine Transport in Pancreatic Islets. *Br. J. Pharmacol.* **68,** 773–778 (1980).

54. Sweet, I. R. *et al.* Systematic screening of potential beta-cell imaging agents. *Biochem. Biophys. Res. Commun.* **314,** 976–983 (2004).

55. Ericson, L. E., Hakanson, R. & Lundquist, I. Accumulation of dopamine in mouse pancreatic B-cells following injection of L-DOPA. Localization to secretory granules and inhibition of insulin secretion. *Diabetologia* **13,** 117–124 (1977).

56. Frenkel, D. & Smit, B. *Understanding Molecular Simulation*. *Understanding Molecular Simulation* (Elsevier, 2002).

57. Jungwirth, P. *Classical and Quantum Molecular Dynamics*.

58. Allen, M. P. & Tildesley, D. J. *Computer simulations of liquids*. (Clarendon Press, 1987).

59. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577 (1995).

60. Miyamoto, S. & Kollman, P. A. SETTLE - An Analytical Version Of The SHAKE and RATTLE Algorithm For Rigid Water models. *J. Comput. Chem.* **13,** 952–962 (1992).

61. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).

62. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. (Dover editions, inc., 1996).

63. Cramer, C. J. *Essentials of Computational Chemistry*. (Willey & Sons, 2002).

64. Hehre, W. J., Stewart, R. F. & Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *J. Chem. Phys.* **51,** 2657 (1969).

65. Binkley, J. S., Pople, J. A. & Hehre, W. J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **102,** 939–947 (1980).

66. Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90,** 1007 (1989).

67. Reed, A. E., Weinstock, R. B. & Weinhold, F. Natural population analysis. *J. Chem. Phys.* **83,** 735 (1985).

68. Breneman, C. M. & Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **11,** 361–373 (1990).

69. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97,** 10269–10280 (1993).

70. McQuarrie, D. A. *Statistical Mechanics*. (2000).

71. Pohorille, A. & Chipot, C. *Free Energy Calculations - Theory and Applications in Chemistry and Biology*. (Springer).

72. Ferrenberg, A. & Swendsen, R. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **63,** 1195–1198 (1989).

73. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13,** 1011–1021 (1992).

74. Ishikawa, T. *et al.* An abnormal pK(a) value of internal histidine of the insulin molecule revealed by neutron crystallographic analysis. *Biochem. Biophys. Res. Commun.* **376,** 32–5 (2008).

75. Chang, X., Jorgensen, A. M., Bardrum, P. & Led, J. J. Solution structures of the R6 human insulin hexamer,. *Biochemistry* **36,** 9409–9422 (1997).

76. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25,** 247–60 (2006).

77. Frisch, M. J. *et al.* Gaussian 09, Revision A.02. (2009).

78. Kohagen, M., Mason, P. E. & Jungwirth, P. Accurate description of calcium solvation in concentrated aqueous solutions. *J. Phys. Chem. B* **118,** 7902–9 (2014).

79. Kohagen, M., Mason, P. E. & Jungwirth, P. Accounting for Electronic Polarization Effects in Aqueous Sodium Chloride via Molecular Dynamics Aided by Neutron Scattering. *J. Phys. Chem. B* (2015).

80. Leontyev, I. & Stuchebrukhov, A. Accounting for electronic polarization in non-polarizable force fields. 2613–2626 (2011).

81. Leontyev, I. V & Stuchebrukhov, A. A. Polarizable molecular interactions in condensed phase and their equivalent nonpolarizable models. 1–45

82. Leontyev, I. V & Stuchebrukhov, A. A. Polarizable Mean-Field Model of Water for Biological Simulations with AMBER and CHARMM Force Fields. (2012).

83. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **9,** 3878–3888 (2013).

84. D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B., X. W. and P. A. K. Amber 14. (2014).

85. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91,** 6269–6271 (1987).

86. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684 (1984).

87. Boresch, S., Tettinger, F., Leitgeb, M. & Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **107,** 9535–9551 (2003).

88. Shirts, M. R., Mobley, D. L., Chodera, J. D. & Pande, V. S. Accurate and Efficient Corrections for Missing Dispersion Interactions in Molecular Simulations. *J. Phys. Chem. B* **111,** 13052–13063 (2007).

89. Mobley, D. L., Chodera, J. D. & Dill, K. A. On the use of orientational restraints and symmetry corrections in

alchemical free energy calculations. *J. Chem. Phys.* **125,** 84902 (2006).

90. Grossfield, A. WHAM: the weighted histogram analysis method, 2.0.9. (2016).

91. Pal, S. K., Peon, J. & Zewail, A. H. Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1763–8 (2002).

92. Pal, S. K., Peon, J., Bagchi, B. & Zewail, A. H. Biological Water: Femtosecond Dynamics of Macromolecular Hydration. *J. Phys. Chem. B* **106,** 12376–12395 (2002).