Università degli Studi di Cagliari

DOTTORATO DI RICERCA

SCIENZE E TECNOLOGIE PER L'INNOVAZIONE

Ciclo XXX

Computational Modeling and Simulations of Protein-Drug and Protein-Protein Complexes: as potential target for therapeutics development

Settore scientifico disciplinari di afferenza

CHIM/07 FONDAMENTI CHIMICI DELLE TECNOLOGIE

Presentata da: **Dr. Amit Kumar**

Coordinatore Dottorato: Prof. Roberto Orru

Tutor: Prof. Francesco Delogu

Esame finale anno accademico 2016– 2017
Tesi discussa nella sessione d'esame Marzo 2018

ABSTRACT

The main objective of my thesis is to illustrate the potential of computational modeling techniques in determining decisive protein-protein interactions and protein-ligand interactions of two relevant macromolecular biological systems associated to human diseases. Computational tools such as homology modeling, molecular docking, molecular dynamics simulations and the developed protocols implemented for the preparation, simulation and analysis of each biological system are presented. The first contribution is the simulation of modeling of protein-peptide-protein complexes related to adaptive immune system and multiple sclerosis disease. Investigation of molecular similarity between self-peptide and two microbial peptides for the complexes with respect to molecular recognition mechanism is presented.

The second contribution is the investigation of protein-ligand interactions of biological systems associated to Alzheimer's disease. Computational results are compared with experiments to evidence the origin and degree of selective inhibition displayed by 2-Phenylbenzofurans ligands against butyrylcholinesterase (BChE) protein. The final contribution is on the application of a *priori* knowledge gathered on protein-ligand interactions in designing ligands with specific structural modifications that display an improved inhibitory activity against BChE protein. In conclusion, therapeutical perspectives and application of hybrid computational approaches to design and develop of potential drugs are discussed.

ACKNOWLEDGEMENTS

LIST OF PUBLICATIONS

1. Delogu GL, Matos MJ, Fanti M, Era B, Medda R, Pieroni E, Fais A, **Kumar A***, Pintus F. (2016) 2-Phenylbenzofuran derivatives as butyrylcholinesterase inhibitors: Synthesis, biological activity and molecular modeling. *Bioorganic & Medicinal Chemistry Letters* **26**, 2308-2313.

   Contribution: Idea conceived, MD simulations performed, analyzed and discussed computational results, and wrote the manuscript. I am also joint corresponding author of the paper.

2. **Kumar A***, Delogu F. (2017) Dynamical footprint of cross-reactivity in a human autoimmune T-cell receptor. *Sci Rep* **7**, 42496

   Contribution: Designed the study, performed MD simulations, analysed and discussed the results and wrote the manuscript. I am also the corresponding author of the paper.

3. **Kumar A*,** Pintus F, Petrillo AD, Medda R, Caria P, Matos MJ, Vina D, Pieroni E, Delogu F, Era B, Delogu GL, Fais A. (2018) Novel 2-pheynlbenzofuran derivatives as selective butyrylcholinesterase inhibitors for Alzheimer's disease. *Sci Rep* (Under Review)

   Contribution: Idea conceived, MD simulations performed, analyzed MD results and discussion, and wrote the manuscript. I am the corresponding author of the paper.

Table of Contents

1. Introduction

Proteins are long molecular chains that our cells need to function properly. The term "**Protein**" was derived from the Greek word "*Prota*", meaning "of primary importance", which was introduced by **Jons Jakob Berzelius** in the year 1838. Proteins are biological macromolecules[1,2] made from building blocks of twenty amino acids and make up about 42% of the dry weight of our bodies. For instance, the protein collagen holds our muscles, skin, and bones together, that accounts up about a quarter of total protein in our body, and all of our cells and even blood are packed with protein molecules. Amino acids are organic compounds formed by amine (-NH2) and carboxyl (-COOH) functional groups, along with a side chain (R group), which is specific to each amino acid[1] (Fig. 1).
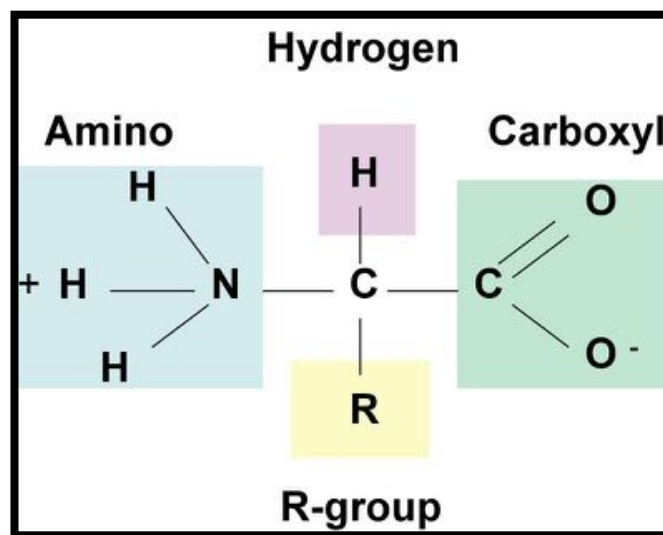
Figure 1. Two-dimensional structure representation of an amino acid.

Protein structure can be classified into four different aspects based on different of covalent structure and folding patterns (Fig. 2), namely primary, secondary, tertiary, and quaternary structure. A linear chain of amino acid residues is called a polypeptide

and sequence of amino acids that make up a polypeptide chain is the primary structure, while regular and repeated patterns of folding (two most common folding patterns are the alpha helix and the beta sheet) of the protein backbone form the secondary structure. The overall folding of the entire polypeptide chain into a specific three-dimensional shape (compact, globular shape) is referred as protein tertiary structure. Finally, the quaternary structure is the way in which the different subunits (more than one polypeptide chain) are packed together to form the overall structure of the protein[3].
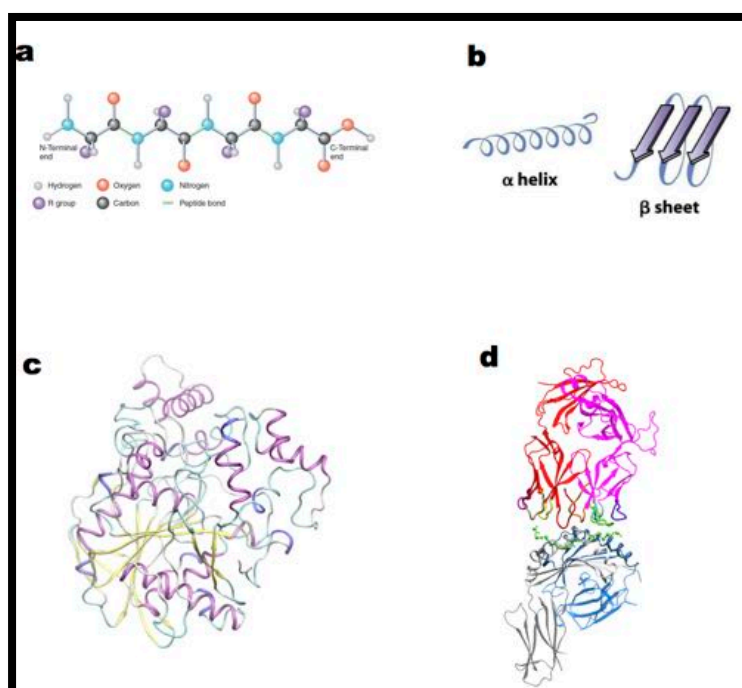
Figure. 2 Four aspects of protein structure. (a) Primary (b) Secondary (c) Tertiary (d) Quaternary.

Proteins need to physically interact with each other, in order to accomplish their biological functions. Protein folding into one or more specific spatial conformations is driven by a number of non-covalent interactions such hydrogen bonding, ionic interactions, Van der Waals forces, and hydrophobic packing[2]. To understand the functions of proteins at a molecular level, it is often necessary to determine their three-dimensional (3D) structure[4]. Experimental techniques such as X-ray crystallography[5], nuclear magnetic resonance (NMR) spectroscopy[6] and dual

polarisation interferometry[7] have been employed to determine the structure of proteins.

1.1 Protein–Protein interactions (PPIs)

PPIs are of pivotal importance in the molecular recognition, signaling and regulation in biological systems[8]. The relevance of these complex networks of interactions has been widely recognized in many subjects as biology, immunology and medicine and drawn considerable attention for designing drugs of the future[9]. Dysfunction in protein interactions can cause diseases, for instance aberrant interactions can lead to the accumulation of protein aggregates, which result in a number of neurodegenerative diseases[10,11]. The nature of interactions between the proteins depends not just on their shape but also on their chemical properties, positive and negative charged amino acids are attracted to each other[12]. These physical properties allow proteins to interact in specific ways. The stability and specificity of protein interactions are highly dependent on the presence of electrostatic interactions, salt bridges, hydrogen bonds and hydrophobic attractions. PPIs are dynamic in nature and can been classified into diverse types ranging from transient or permanent obligate interactions to obligate interactions[13-15]. The interaction between the proteins is facilitated through their respective interfaces and the structural; physicochemical properties vary across different protein-protein interfaces[13,16]. The residues that contribute significantly to the free energy of binding are called "hot spots"[17,18](Fig. 3).
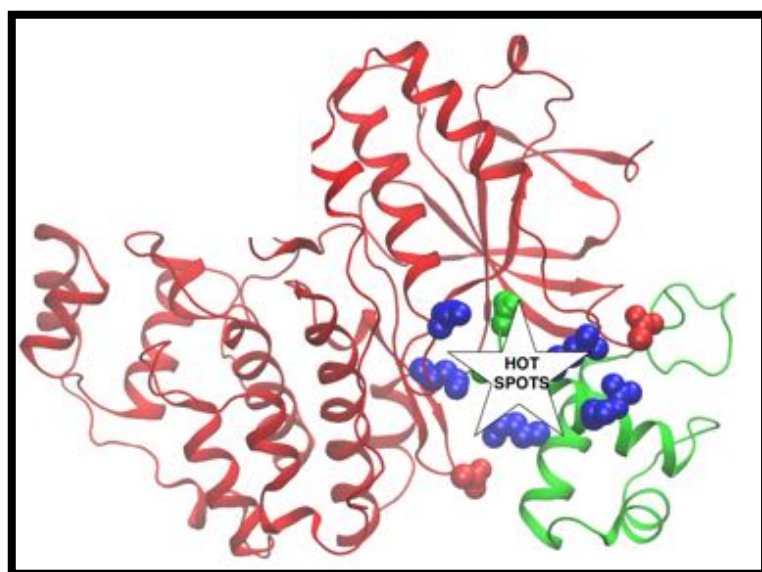
Figure 3. Interaction at protein-protein interface. Hot spot residues at the interface are shown as blue (positive charged) and in red (negative charged) spheres.

These hot spots are also considered as potential drug targets, to interfere with the biological binding process at protein-protein interface[19,20]. Identification of hot spot in the protein interfaces is possible, for instance using alanine scanning mutagenesis experiments[18], wherein a variation in binding energy by at least 2 kcal/mol upon mutation is used as a criterion to label the residue as hot spot. However, scarce experimental data available in literature have opened way for computational techniques to predict PPIs[21]. The majority of computational techniques can be classified as simulation-based and machine-learning based ones. In my thesis, I will present molecular docking[22] and molecular dynamics (MD) simulations[23] approaches to model the forces governing the interaction of the proteins and estimate the strength of these interactions.

1.2 Protein-ligand interaction (PLIs)

PLI is a molecular recognition process of biological macromolecules (proteins) interacting with various small molecules (biological ligands, drugs, activators, inhibitors), typically with a high specificity and affinity to form a specific complex[24]. PLI thus constitutes the basis of almost all the processes in living organisms. PLIs also play a vital role in inhibition of enzymes, which are related to diseases such as Alzheimer's disease[25,26]. A detailed understanding of the protein-ligand interactions (Fig. 4) is therefore central to understanding biology at the molecular level. The driving forces of various interactions and energy exchanges among the protein and ligand determine their association and binding characteristics[27]. Gibbs binding free energy (G) can be parsed into the enthalpy and entropy contributions with the following equation[28,29]:

$$\Delta G = \Delta H - T\Delta S \ (Equation\ 1)$$

where ΔH and ΔS are change in enthalpy and entropy, respectively, of the system upon ligand binding, and T is the temperature in Kelvin.
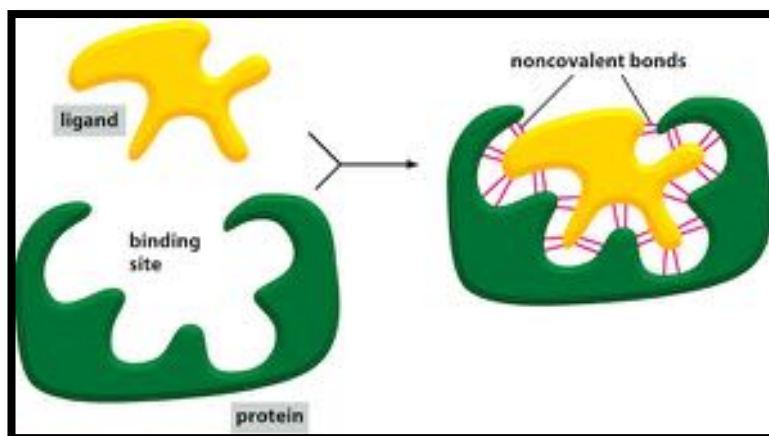
Figure 4. Cartoon representation of Protein-Ligand Interaction.

Enthalpy is a measure of the total energy of a thermodynamic system. In a binding process, ΔH, or the binding enthalpy, reflects the change in total energy of the system upon ligand binding to the protein. The binding enthalpy in a non-strict sense is generally treated as the changes in energy resulting from the formations of non-covalent interactions (Fig. 4) at the binding interface. Entropy is viewed as a measure of the disorder or randomness in atoms and molecules in a system. The protein–ligand binding can occur spontaneously only when the change of the system free energy is negative, which is determined jointly by two thermodynamic quantities, enthalpy and entropy. An interesting phenomenon is "enthalpy-entropy compensation",[30] wherein the free energy of binding is maintained constant or modulated by complementary changes of enthalpy and entropy. This phenomenon has been confirmed from analysis of calorimetric data for protein–ligand binding in many biological systems[31]. Among three major models, lock and key, induced-fit and conformational selection have been proposed to explain the protein-ligand binding mechanisms[27,32]. Depending on the site of inhibition in the enzymes the ligands (molecules) are classified as competitive, mixed-type or non-competitive ones.

Experimental techniques such as X-ray crystallography, NMR and cryo-electron microscopy (Noble prize in Chemistry 2017) provide atomic-resolution or near-atomic-resolution structures of the unbound proteins and the protein–ligand complexes. Experiments such as ligand binding assay techniques[33], isothermal titration calorimetry[28,34] and surface plasmon resonance[35] have been employed to measure protein–ligand binding affinity. Even tough, experimental techniques allows us to investigate thermodynamic profiles for protein-ligand complex, the procedures

for determination of binding affinity are laborious, time-consuming, and expensive. Moreover, modern rational drug design strategy usually involves screening of databases containing millions of compounds to find the lead molecules[36,37]. In this scenario, computational approaches have enormous potential in providing insights in facilitating the interpretation of the existing experimental data and also direct the design of new experiments. Thus, providing insights into underlying biological processes and functions. Indeed, structure-based computational approaches are valuable tool in all aspects of investigating protein-ligand binding interactions.

In this thesis, the suitability of employing computational approaches to investigate protein-protein interactions and protein-ligand interactions in biological complexes important to human diseases will be presented. The reliability of computational methods for investigating protein-protein binding characteristics, protein-ligand binding affinity, and validation with available experimental approaches and their advantages, disadvantages, and challenges will be discussed and examined.

The thesis is organized in the following manner:

- ♣ Chapter 2 — Computational Methodology
- ♣ Chapter 3 — Molecular insight into protein-peptide-protein complexes found in a patient suffering from Multiple Sclerosis disease
- ♣ Chapter 4 — 2-Phenylbenzofuran ligands as selective butyrylcholinesterase protein inhibitors
- ♣ Chapter 5 —New 2-Phenylbenzofuran ligands with an improved inhibitory selectivity against BChE protein
- ♣ Chapter 6— concluding remarks and future perspective

2. Methods

2.1 Homology Modeling

Technological incrementation in DNA sequencing methodology has resulted in an enormous, and ever- growing, number of protein sequences. At the same time, the number of experimentally determined protein structures has lagged increasingly behind, owing to the inherently slower, more resource intensive, and less-predictable nature of these experiments. Determination of three-dimensional protein structure by experimental methods such as X-ray crystallography or NMR spectroscopy is time consuming, laborious and an expensive process. Currently, there are 138,840 (access date 24 October 2017) experimental protein structures deposited in the Protein Data Bank (www.wwpdb.org)[38], while there are 500,000 (access date 24 October 2017) protein sequences freely accessible at Uniprot databsase (www.unitprot.org)[39]. The "structure knowledge gap" between the huge number of protein sequences and small number of known structures has hampered the widespread use of structure-based approaches in life science research[21].

Template based homology modeling techniques[40] is the method of choice to generate a reliable 3D model of a protein from its amino acid sequence as notably shown in several meetings of the bi-annual critical assessment of techniques for protein structure prediction (CASP). Homology modeling involves searching the conformation space by minimally disturbing those existing experimentally solved protein structures. The method is based on the fact that structural conformation of a protein is more highly conserved than its amino acid sequence, and that small or medium changes in sequence normally result in little variation in the 3D structure.
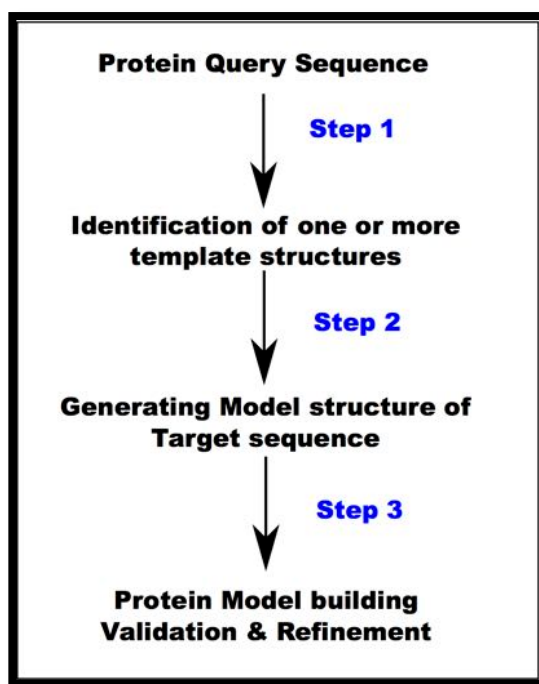
Figure 5. Homology Modeling Protocol

The process of homology modeling steps has been depicted in Fig. 5. These steps are repeated until suitable models were built. The quality of models generated depends on sequence similarity with the template structure and is directly linked with the identity between template and target sequences. A general accepted rule is that models built over 50% sequence similarities are accurate enough for drug discovery applications, while those between 25 and 50% identities can be helpful in designing of mutagenesis experiments and those in between 10 and 25% are tentative at superlative.

In the present work, Modeller software[41], Swiss-model webserver[42] were used as tools to build reasonable 3D structures of protein.

2.2 Molecular Docking

Molecular docking is a multidimensional optimization problem that has become an increasingly important tool for drug discovery[22]. Molecular docking programs perform a search algorithm in which the conformation of the ligand is evaluated recursively until the convergence to the minimum energy is reached. Finally, an affinity scoring function, ΔG (in kcal/mol), is employed to rank the candidate poses as

8

the sum of the electrostatic and Van der Waals energies. The driving forces for these specific interactions in biological systems aim toward complementarities between the shape and electrostatics of the binding site surfaces and the ligand or substrate.

Knowledge of potential ligand binding site before docking processes significantly increases the docking efficiency. It is also possible to obtain information about the sites by comparison of the target protein with a family of proteins sharing a similar function or with proteins co-crystallized with other ligands. However, in the absence of knowledge about the binding sites, cavity detection programs can be utilized to identify putative active sites within proteins. In general, protein-ligand docking strategies can be performed either considering the protein as "rigid" or "flexible". To date 60 different docking tools and programs have been developed for both academic and commercial needs, such as AUTODOCK[43], GOLD[44], and GLIDE[45], which allow rigid and partial flexibility of the receptor and full flexibility for the ligand. Flexibility of the ligand and receptor is computationally cumbersome to include in the docking experiments. A successful docking run is considered if the binding of a ligand into its active site results closer to the known protein-ligand complex structure.
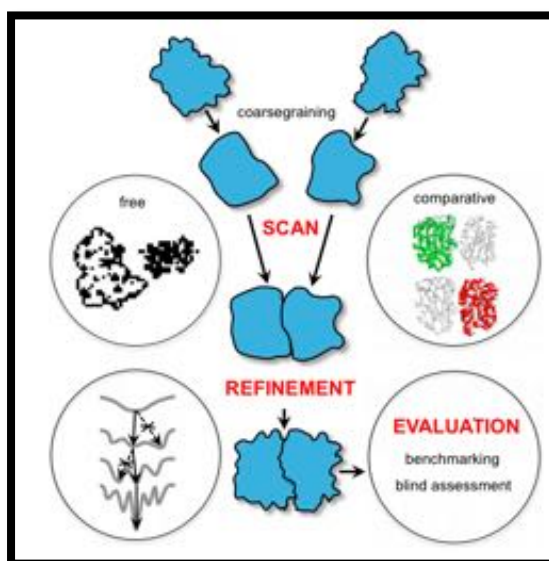
Figure 6. Protein-protein docking methodology development (reproduced from reference[46])

Protein-Protein Docking. 3D structure determination of a protein-protein complex, generally, is challenging and more difficult to determine experimentally than the structure of an individual protein. In Fig. 6, docking protocol for generating protein-protein complex from the known individual protein structure is shown. Proper ways to

accommodate flexibility in the docking simulations without increasing the computational effort too much is one of the most important issues in the development of new methodologies at the moment.

2.3 Molecular Dynamics Simulations

Molecular Dynamics (MD) simulation is a technique founded upon the basic principles of classical mechanics that provide a dynamical picture of the individual particles of the system at a microscopic level[47]. Temporal configurations of the molecular system can be generated using this technique by integrating Newton's law of motion. The result is a trajectory, which contains the microscopic time evolution of the system in the phase space. From the trajectory generated, one can compute the dynamical properties such as absorption spectra, rate constants and transport properties. Further, on combining MD with statistical mechanics as a mean of sampling, one can compute equilibrium properties such as average thermodynamics quantities, structure, and free energies along the reaction path seen as a union of all possible states of the system.

The interaction between the atoms in a complex system can be described using an empirical force-field, which is a mathematical expression describing the dependence of the energy of a system on the coordinates of its particles. The potential function V(r) from which the forces used in MD are derived depends on the atomic coordinates V(r) has the following expression:

$$
\begin{aligned}
V(r) = \sum_{bonds} \frac{1}{2} k_d (d - d_0)^2 &+ \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\
&+ \sum_{dihedrals} k_\varphi \left[1 + \cos(n\varphi - \delta) \right] \\
&+ \sum_{atom\ i} \sum_{j \neq i} 4\epsilon_{i,j} \left\{ \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right\} + \sum_{atom\ i} \sum_{j \neq i} \frac{q_i q_j}{\epsilon_0 r_{i,j}} \qquad (Eq.2)
\end{aligned}
$$

In equation 2, the first three represents the bonded interactions (bonds, angles, dihedrals), while the fourth term describes Leonard-Jones (LJ) potential and the last term is the columbic (electrostatic) interaction[48].

These interactions are modeled using the two-body approximation, which does not explicitly account for the polarization effects. The parameters used in equation 2 are typically obtained from quantum chemical calculations and experimental data (e.g. crystallographic data, spectroscopic data, etc). The popular sets of force fields used for MD simulations[49] of proteins are AMBER[50], GROMOS[51], CHARMM[52] and OPLS[53].
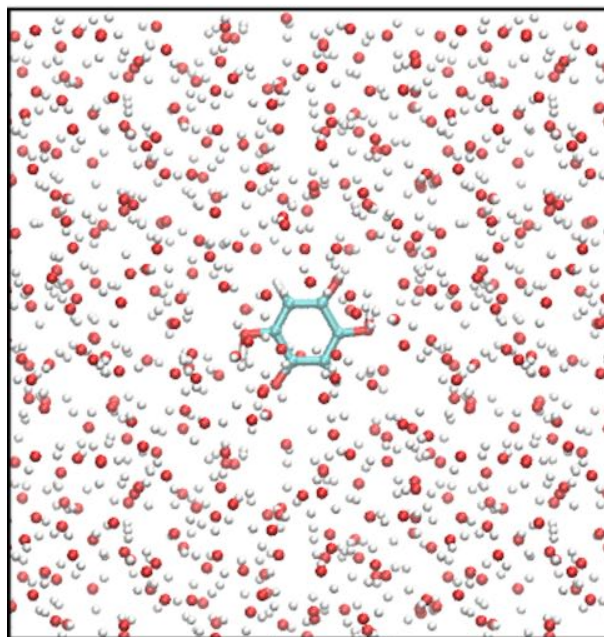
Figure 7. MD Simulation of a molecule in solvent

To mimic the real biological systems, MD simulations of the protein are always performed in presence of the solvent. The description of the solvent (water for most of the biologically interesting systems) can be explicit or implicit. In the first case solvent molecules with a full atomistic force field description are added in the simulation box at the experimental density. In the implicit solvent description the solvent is treated as a dielectric medium in which the system is embedded. This is clearly a more approximated description but it is also computationally much more efficient since in many practical cases the solvent constitutes the majority of the atoms. In the thesis the solvent is treated in an explicit manner.

To avoid artifacts near the border of the simulation box (Fig. 7) periodic boundary conditions (PBC) are employed in MD simulations. In this scheme, short-range non-bonded interactions are calculated using the minimal image convention (only the nearest replica is considered). Typically a cut-off radius (R_c) is used for LJ

interactions of the order of 10 Å. To avoid interactions between a particle and its periodic image each box side must be larger than $2R_c$.

Molecular dynamics can be performed in different statistical ensembles. In the micro-canonical (NVE) ensemble, total number of atoms N, the volume V (of the unit cell) are kept constant. Generally, if the simulation system is sufficiently large, the small part of it may be considered as a canonical system. For large NVE systems the fluctuations in temperature are small, and it may be considered approximately constant. However, there are situations in which temperature must be kept constant, and therefore for these classes of problems MD must reproduce an isothermal ensemble, such as canonical NVT ensemble, in which the number of atoms N, volume, and temperature are constant. The temperature T, in contrast to the number of particles N and volume V, is an intensive parameter. In my thesis, I used NPT ensemble, which is an extension of NVT ensemble, where pressure and temperature are kept constant. The NPT ensemble is extensively used for comparison of MD simulations with experiments. As most experimental measurements are usually made under conditions, which include a fixed pressure P, temperature T, and number of atoms N (constant-NPT ensemble).

3. Dynamical footprint of cross-reactivity in a human autoimmune T-cell receptor

Multiple Sclerosis (MS) is a chronic inflammatory and degenerative disease of the central nervous system affecting more than 2.5 million people worldwide[54,55]. MS involves an abnormal response of the human body's **immune system** directed against brain and spinal cord, triggered, in genetically susceptible individuals, by a combination of one or more environmental factors. Pathogen derived and self-antigen presentation by **major histocompatibility complex** (MHC) is a critical step for T lymphocyte triggering and subsequent immune response, whose failure could lead to initiate specific autoimmune diseases.

Peripheral **T-cell receptors** (TCR's) are commonly educated to **recognize** a maximum of pathogen-derived epitopes while ignoring self-antigens. However, there are also cases in which some TCRs are able to recognize self-antigens, thus initiating an autoimmune response. The term **TCR cross-reactivity** is associated exactly to **the recognition of many different peptide** antigens presented by the MHC of an individual.

In a study published in Nature Communications[56], the authors isolated an autoimmune **Hy.1.B11 TCR** from a relapsing-remitting MS patient. In their work, the **structural**

basis of cross-reactivity displayed by human Hy.1.B11 TCR between MBP and two microbial peptides was investigated. Their study did reveal a common tilted TCR binding topology onto the peptide-MHC surface and pointed out a dominant involvement of complimentary determining region (CDR) in TCR cross-reactivity and thus providing strong indication about the role of local molecular configurations in the activation of cross-reactivity. However, an exclusively static conceptual framework can be hardly thought to enable a satisfactory understanding of the complex mechanisms underlying cross-reactivity.

In this scenario, our **computational molecular dynamics** (MD) simulations performed on these structures can provide the necessary atomistic level description on dynamic changes of TCR in the presence of peptide-MHC complex, which could be useful to understand better engagement of TCR.

In this chapter, we present first report of involvement of a small number of structurally and energetically important **hot spots** that provides new insights into the **dynamical basis** of Hy.1.B11 **TCR cross-reactivity**. This approach can be useful is designing truly **personalized** immunotherapeutic peptides. This work has been published[57] in Scientific Reports of Nature Publishing group.

3.1 Abstract

The present work focuses on the dynamical aspects of cross-reactivity between myelin based protein (MBP) self-peptide and two microbial peptides (UL15, PMM) for Hy.1B11 T-cell receptor (TCR). This same TCR was isolated from a patient suffering from multiple sclerosis (MS). The study aims at highlighting the chemical interactions underlying recognition mechanisms between TCR and the peptides presented by Major Histocompatibility Complex (MHC) proteins, which form a crucial component in adaptive immune response against foreign antigens. Since the ability of a TCR to recognize different peptide antigens presented by MHC depends on its cross-reactivity, we used molecular dynamics methods to obtain atomistic detail on TCR-peptide-MHC complexes. Our results show how the dynamical basis of Hy.1B11 TCR's cross-reactivity is rooted in a similar bridging interaction pattern across the TCR-peptide-MHC interface. Our simulations confirm the importance of TCR CDR3α E98 residue interaction with MHC and a predominant role of P6 peptide residue in MHC binding affinity. Altogether, our study provides energetic and dynamical insights into factors governing peptide recognition by the cross-reactive Hy.1B11 TCR, found in MS patient.

3.2 Introduction

Multiple sclerosis (MS) is a chronic inflammatory and degenerative disease of the central nervous system[58], affecting more than 2.5 million people worldwide[59]. MS exhibits a heterogeneous geographical pattern affecting populations across the globe. In particular, it is more common far from the equator and shows latitude gradient[60].

MS involves an abnormal response of the human body's immune system directed against brain and spinal cord. In particular, the immune system attacks myelin, i.e. the protective substance covering and insulating nerve fibers. The disease owes its name exactly to the sclerosis formed by damaged myelin[61]. Damage, or destruction, of any part of the myelin sheath or nerve fibers cause the distortion, or interruption, of the nerve impulses that travel to and from brain and spinal cord. Eventually, nerve fibers themselves can deteriorate or suffer from permanent damage. A wide variety of symptoms determined by the location of lesions within the central nervous system can arise[62], ranging from loss of sensitivity and changes in sensation to pain, muscle

weakness, difficulty in moving and sight and speech problems[63]. Overall, the exact cause of MS disorder still remains elusive[64], with the disease linked to both genetic[65,66] and environmental[67,68] factors. Experimental and clinical studies have provided evidence reinforcing the hypothesis that immune mechanisms are involved in the pathogenesis of inflammatory demyelination in MS[64,69]. In fact, the exact antigen that immune system cells are sensitized to attack has not been recognized yet, and this leads many experts to consider MS as an "immune-mediated" process, rather than an "autoimmune" disease. Nevertheless, multiple findings identified the human leukocyte antigen (HLA) class II system as the main genetic determinant region related to MS[65,70,71]. An important role of autoreactive T lymphocytes in the initiation and perpetuation of disease has also been suggested[59,72].

T-cells form a subset of lymphocytes, critical for providing an adaptive immune response against invading pathogens[73]. In particular, the T-cell receptor (TCR) at the surface of T lymphocytes is a complex of integral membrane proteins that participates in the activation of T-cells in response to an antigen[74]. Stimulation of TCR is triggered after recognition of antigenic peptides presented by the major histocompatibility complex (MHC), corresponding to HLA in humans, located on the surface of antigen-presenting cells[75]. A successful TCR engagement initiates positive and negative cascades leading to T-cell activation, differentiation, proliferation and, finally, to a specific immune response to the invading pathogen[76,77].

Peripheral T-cells are commonly trained to recognize a widest set of pathogen-derived epitopes while ignoring self-antigens[75]. However, there are also cases in which some TCRs escape this selection and are able to recognize self-antigens, thus initiating an autoimmune response and becoming self-reactive[78]. The term TCR cross-reactivity is associated exactly to the recognition of many different peptide antigens presented by the HLA of an individual[79-81]. The three-complementarity determining region (CDR) loops of the α and β chains present in TCR facilitate the recognition of peptide-HLA-II complex (Figure 1)[56]. Majority of contacts with the bound peptide involve CDR3 rather than CDR1 and CDR2[82].
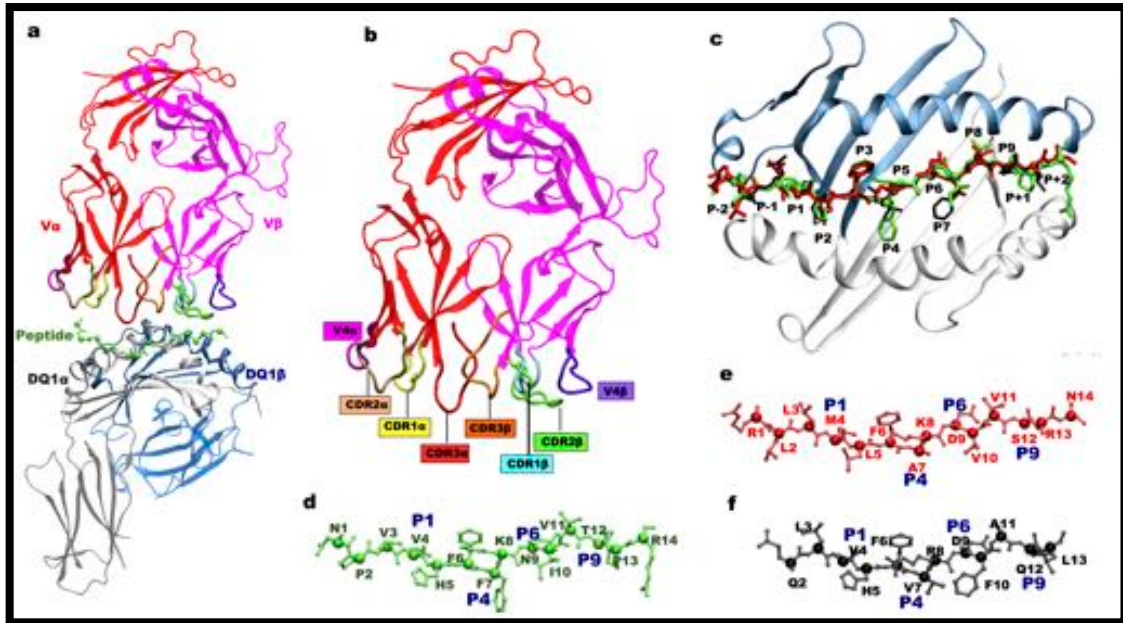
Figure 8. Peptide cross-reactivity of Hy.1B11 TCR. (a) Structure of TCR-MBP-HLA-DQ1-with HLA-DQ1 complex: in red TCR Vα; in pink TCR Vβ; MBP peptide in green; HLA-DQ1α in grey; HLA-DQ1β in light blue. (b) Complementary determining regions (CDR's) of TCR. (c) HLA-DQ1 peptide binding groove with the three peptides and their pockets. (d,e,f) Peptides structures with residues named using single letter nomenclature: MBP in green; PMM in red and UL15 in black, and the pockets P1, P4, P6 and P9 are indicated in blue.

Many experimental studies have shown the importance of TCR cross-reactivity in initiating adaptive immune response[83-85]. However, a direct correlation between TCR binding affinity[86-88] and potency of T-cell activation has not been proved, and the overall process is still poorly understood[89]. Multiple mechanisms of T-cell receptor cross-reactivity have been proposed on the basis of the solved three-dimensional structures for the tri-molecular complex TCR-peptide-HLA[90]. Specifically, induced fit[91], differential TCR docking[79], structural degeneracy[92], molecular mimicry[83,93], and antigen-dependent tuning of peptide-HLA flexibility[84], were proposed. Around twenty three-dimensional structures for the TCR-peptide-MHC-II complex structures have been determined providing structural insights into MHC restriction[75]. In a previous study, investigators isolated an autoimmune Hy.1.B11 TCR from a MS patient[94]. This TCR was initially found to be specific for myelin basic protein (MBP) peptide bound

to the HLA-DQ1 (DQA1*0102-DQB1*0502) class II protein. However, it has been observed that the same Human Hy.1.B11 T-cell clone is not only activated by MBP, but also by other distinct microbial peptides[95]. The structural basis of cross-reactivity displayed by human Hy.1.B11 TCR between MBP and microbial peptides from UL15 terminase protein of *Herpes simplex virus* and phosphomannomustase protein of *Pseudomonas aeruginosa* has been recently investigated[56]. The crystallographic structures of these complexes revealed a common tilted TCR binding topology onto the peptide-MHC surface, pointing out a dominant involvement of CDR3α residues towards both self and microbial peptide. This aspect clarified the docking geometry and static interaction picture between TCR and peptide-MHC molecule, providing strong indication about the role of local molecular configurations in cross-reactivity. However, an exclusively static conceptual framework can be hardly thought to allow a satisfactory understanding of the complex mechanisms underlying cross-reactivity. Accurate dynamical information is needed as well to unveil the role of atomic and molecular motion in the involvement of CDR residues.

Computational molecular dynamics (MD) simulations can provide the necessary finite temperature atomistic level description dynamic changes of TCR in the presence of peptide-MHC complex, which could be useful to understand better engagement of TCR by the immune system. MD simulations have been already employed to investigate, for instance, the dynamics of tri-molecular TCR-peptide-MHC-I complexes, with either altered/different peptides[96-98], or different MHC's[99] or different TCR's[100]. Yet, until date only one computational study has been performed on the TCR-peptide-MHC-II complex[101], wherein the authors investigated the energetic and flexibility properties of the complex with a native peptide as well as for twelve mutations introduced in the peptide.

In this chapter, we present MD simulations of the three tri-molecular complexes relevant to observed TCR cross-reactivity in a MS patient. To further probe the role of peptide-MHC complex dynamics in TCR cross-reactivity, we performed additional simulation of three peptide-MHC complexes in the absence of TCR structure.

Our simulations highlight the important energetic role of CDR3α TCR loop in binding to HLA-DQ1-peptide complexes; in particular the key contribution to peptide recognition by CDR3α E98, a residue that is conserved in the three tri-molecular complexes. Furthermore, we found a new interaction between another TCR loop

CDR2β D55 residue and HLA-DQ1α K39 residue, located outside the peptide-binding groove, to constitute a conserved anchor point for docking TCR on to MHC class II protein. Altogether, these results explain the dynamical basis for cross-reactivity between the MBP self-peptide and the two microbial peptides for Hy.1B11 TCR.

3.3 Results

For the three peptides (MBP, UL15, PMM), we performed TCR-pMHC MD simulations and pMHC MD simulations, each 110 ns in length. The convergence of MD simulations was estimated using a novel Good-Turing statistical approach[102].

Molecular interactions of TCR with HLA-DQ1 bound peptides. The persistent H-bond interaction evaluated between the HLA-DQ1 bound peptide residues and TCR residues survived for than 60% of the MD simulations. H-bond interactions of TCR with MBP and the two microbial peptides (UL15, PMM) were characterized by a unique pattern of pair-wise contact between CDR3 Vα E98 and pocket 5 (P5) K8/R8 of the peptide complexes (Fig. 9). A marked difference in H-bond interaction pattern between the self and microbial peptide complexes was observed. In the two microbial peptide complexes, the same peptide residue K8/R8 (Fig. 8e-f) shared a common interaction pattern with CDR3β D97 residue. While, a specific pair-wise contact between fourth hyper variable loop V4β residue E69 and R14 of MBP peptide complex was observed (Fig. 9b).
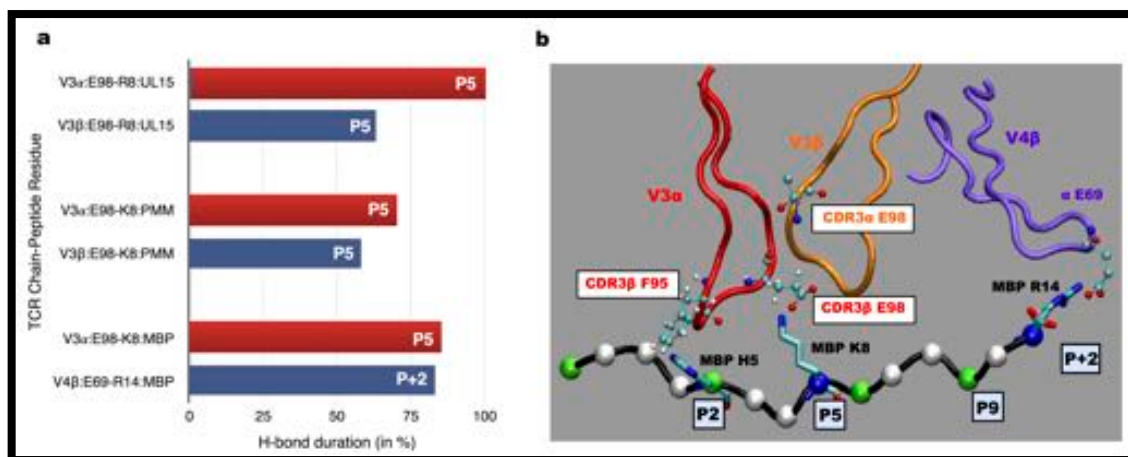
Figure 9. Hydrogen bond interaction TCR and peptide residues. (a) H-bond duration between TCR-peptide residues in %. (b) Interaction picture between TCR and MBP peptide.

Persistent stacking interactions considered between the TCR and peptide residues, survived for more than 40% of MD simulations. CDR3α F95 residue formed a persistent stacking interaction with the P3 (Phe6) PMM peptide residue and with the P2 (His5) MBP peptide residue, respectively. However, in the UL15 peptide complex no persistent stacking interaction involving CDR3α F95 was noted.

Interactions of TCR with HLA-DQ1. The residues of TCR/HLA-DQ1 complex (Fig. 10a) were selected to perform dihedral angle principal component analysis[103] (dPCA) for the three complexes on the 5500 snapshots extracted from MD simulation trajectory. The dihedral angles of TCR-MHC residues were projected onto the first two principal components (PC) for each trajectory snapshot from MD simulations for the three complexes (Fig. 10b). Each point in the plot (Fig. 10b) represents a specific configuration explored by the TCR-MHC complex during MD simulations. Projection of dihedral angle fluctuations along the first two principal components in three peptide complexes (UL15, PMM, MBP, Fig. 10b), suggested a more limited phase space exploration in the microbial peptide complexes with respect to MBP complex.
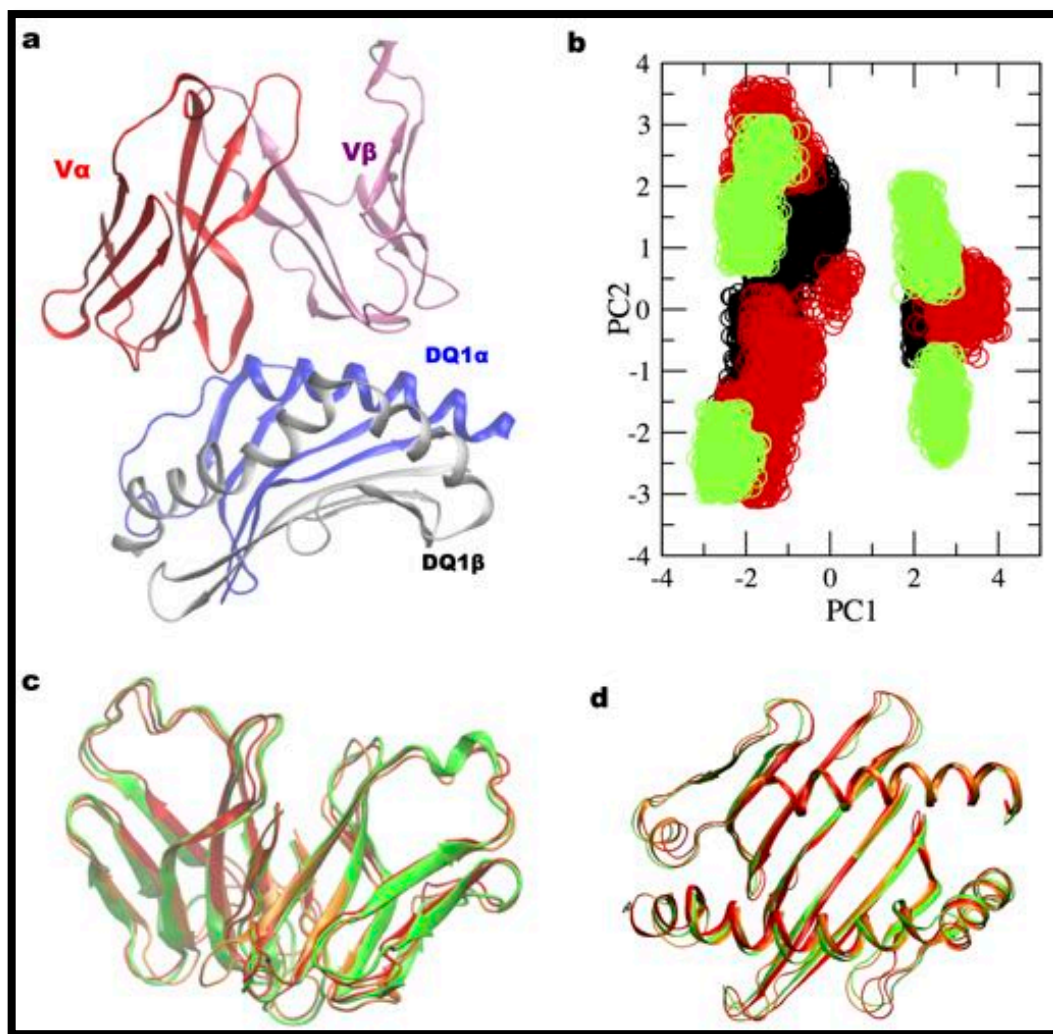
Figure 10. dPCA on the TCR-MHC class II for the three molecular complex systems. (a) TCR residues α-chain (24:104) in red and β-chain (24:103) in magenta; HLA-DQ1 α-chain (5:76) in blue and β-chain (7:90) in grey. (b) The two dimensional point maps correspond to projection of dihedral angles (phi, psi) fluctuations (from MD simulation trajectory) on the plane defined by first two principal components: (i) black: UL15, (ii) red: PMM and (iii) Green: MBP peptide complexes. Superposition of (c) TCR Vα and Vβ domains and (d) HLA-DQ1 α1 and β1 domains, for the three-peptide complexes.

For further investigation, we performed cluster analysis (see Methods) and identified the most populated configurations in the three complexes and subsequently obtained their corresponding representative structures. The difference between the representative TCR-MHC configurations in the microbial peptide complexes to that in

the MBP peptide complex case was evaluated by calculating the RMSD values (Table 1).

Table 1. RMSD and buried surface area (BSA) calculations. RMSD values for Vα, Vβ chains of TCR (column 2, column 3) and α1, β1 chains (column 4, column 5) of HLA-DQ1. In column 6, we report accessible surface area values evaluated on MD simulation trajectory. The RMSD fluctuation values are reported with respect to MBP peptide complex as reference (REF).

RMSD (backbone)	TCR			HLA-DQ1			Accessible surface area ($Å^2$)
	Vα-chain	Vβ-chain	Total	α1-chain	β1-chain	Total	
UL15	1.3 Å	1.3 Å	1.6 Å	1.6 Å	1.4 Å	1.8 Å	1655±63
PMM	1.1 Å	1.4 Å	1.9 Å	1.2 Å	1.4 Å	1.6 Å	1704±64
MBP	REF	REF	REF	REF	REF	REF	1702±63

We found the TCR-MHC class II structures in microbial peptide complexes and MBP complex to overlap quite nicely (Fig. 10c-d) with an RMSD difference less than 2 Å. This observation is consistent with comparable buried surface area values in the three complexes (Table 1).

Persistent hydrogen bonded interactions between the TCR and HLA-DQ1 residues are reported in Table 2. We found both conserved and non-conserved interactions between the MBP peptide and two microbial peptide complexes. Overall, we found a higher number of interactions between the TCR residues and DQ1 α1 helix residues, while peptide specific TCR interactions with DQ1 β1 helix residue was noted only for the microbial peptide complexes.

Table 2. Hydrogen bonded interactions between TCR and HLA-DQ1. Conserved interacting pairs are highlighted in bold, with the total number of interactions in ().

TCR – MHC	UL15	PMM	MBP
Vα-DQ1 α1	**E98-R61** (3)	**E98-R61** (4)	**E98-R61** (2)
	-	-	E98-Q57 (1)
	-	-	K99-Q57 (1)
Vβ-DQ1 α1	**D55-K39** (2)	**D55-K39** (1)	**D55-K39** (2)
	E69-K75 (2)	Q48-H68 (1)	D54-K39 (3)
Vα-DQ1 β1	-	R51-E66 (4)	-
	-	R51-E69 (2)	-
Vβ -DQ1 β1	G26-Y60 (1)	-	-

Remarkably, the titled orientation of TCR limits its interactions with DQ1 β1 helix residues. Moreover, CDR2β loop overlays the central portion of DQ1 α1 helix and Vβ D55 residue involved in H-bond interaction with DQ1 α1 residue K39 located outside the peptide-binding groove (Table 2, Fig. 11).
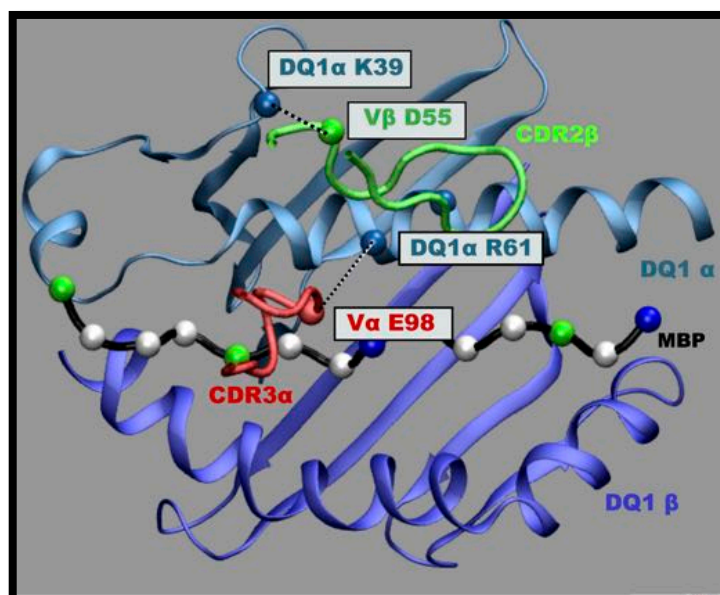
Figure 11. Conserved TCR-MHC contacts. H-bond interactions and placement of CDR2β loop over DQ1 α-helix. The participating residues are boxed, and H-bond is denoted by dashed line.

23

HLA-DQ1 interactions with bound peptides in the TCR complex. We analysed MD simulation trajectory for the three TCR-pMHC complexes and report only the persistent H-bond interactions observed between the peptide-MHC pairs (Fig. 12). In general, we found DQ1 α-chain residues (Fig. 12a) to display a predominant involvement in H-bond interactions with the peptide residues. This observation derives higher number of interacting pairs with respect to that of β-chain (Fig. 12b). Even tough the two microbial peptides are quite different in their sequence from MBP peptide, the majority of the H-bond interactions between MHC residues and specific positioned peptide residues was conserved.
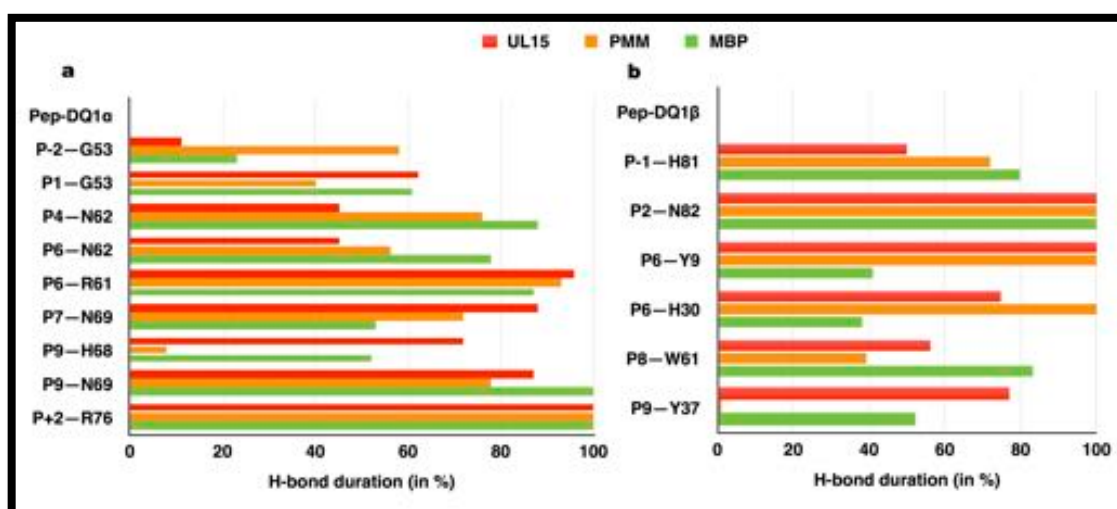
Figure 12. **Hydrogen bonded interactions between peptide and HLA-DQ1 residues.** In (a) peptide-DQ1α interacting pairs, and in (b) peptide-DQ1β interacting pairs.

Subsequently, we also analysed persistent stacking[104] interactions between the peptide residues and the two chains of MHC protein. A conserved pair-wise stacking contact between DQ1 α-chain residue R61 and pocket 3 of the peptide cases was observed. On the other hand, we also found peptide-specific stacking interaction pattern involving DQ1 β-chain residues.

Table 3. **Stacking interaction (pep-MHC) evaluation and configuration entropy calculations.**

	Stacking interaction		Configurational Entropy (J/molK)	
	Peptide - DQ α	Peptide-DQ β	HLA-DQ1	TCR-MHC
UL15	F6(CD1)-R61(NE)	F10(CD1):Y47(CD1)	5672	8687
PMM	F6(CD1)-R61(NE)	R13(NE)-P56(CD) R13(NE)-Y60(CD1)	5965	8890
MBP	F6(CD1)-R61(NE)	H5(CG)-R77(NE) H5(CG)-H81(CG) P13(CD)-Y60(CD1)	5839	8877

Entropy and interaction energy estimation. Configurational entropy[105] values calculated for HLA-DQ1 peptide binding groove residues and for the TCR-MHC components respectively, showed highest value for the microbial peptide PMM complex, while the self-peptide MBP complex displayed an intermediate value. The calculated interaction energy of TCR with peptide-DQ1 complexes showed highest value for DQ1/MBP than DQ1/PMM and DQ1/UL5 cases (Fig. 13a). A similar trend was also noted by better binding affinity (higher interaction energy value) of MBP peptide for TCR (-206 kcal/mol), compared with the two microbial peptide cases (-125 kcal/mol). However, observing binding affinity of peptides for DQ1 alone, we found the lowest interaction energy value for MBP complex (Fig. 13b).

HLA-DQ1 peptide binding groove width fluctuation in TCR simulations. To probe the importance of binding groove flexibility at a local level description, we dissected the groove into four regions[106] (D1-D4, Fig. 13c). The center of mass distance variation between the DQ1 α-chain and β-chain residues was selected as a parameter to monitor the distance fluctuations in these four regions.
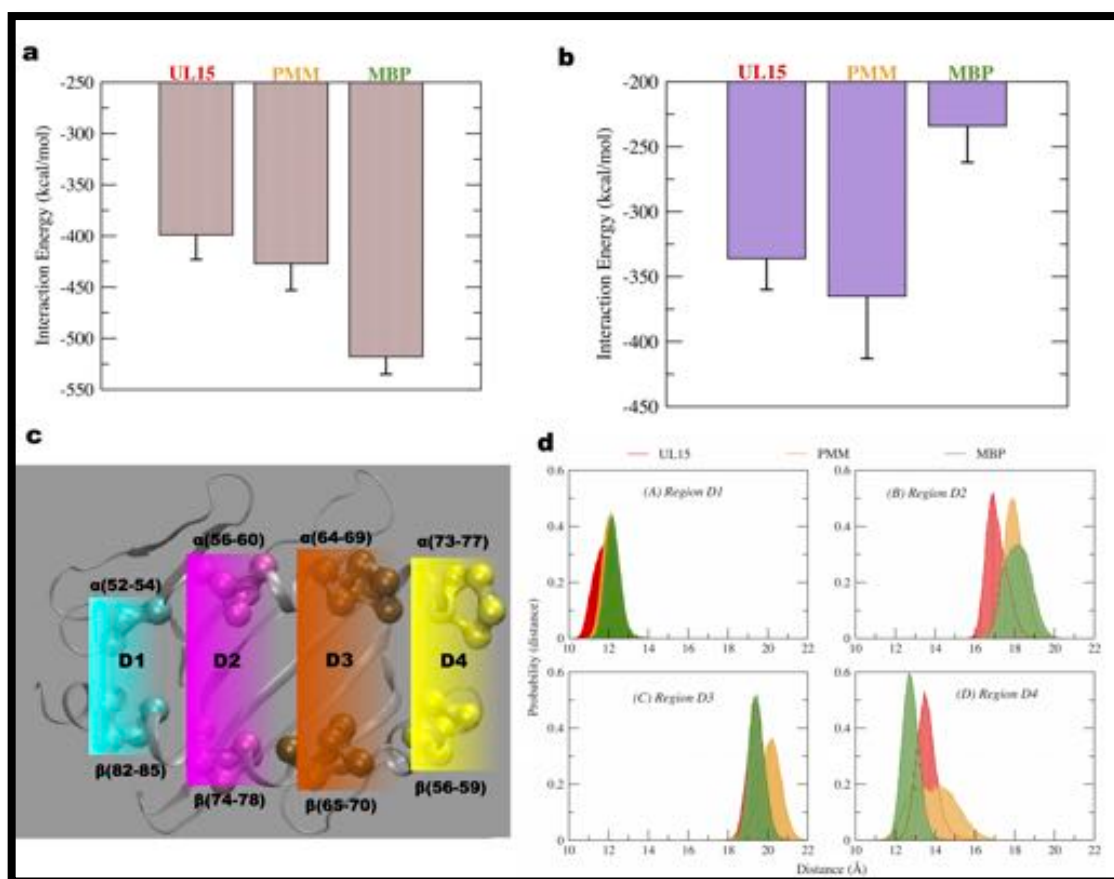
Figure 13. **Interaction energy plots and HLA-DQ1 groove width analysis.** (a-b) Interaction energy corresponds to non-bonded energy values comprising of Van der Waals and electrostatic energy between: (a) TCR and pep-MHC and in (b) peptide and MHC. (c) Dissection into four regions D1 to D4, (d) center of mass distance variation in the four different regions, MBP complex in green, UL15 complex in red and PMM complex in orange.

In region D1, the distance profile distribution is quite similar for the MBP and PMM peptide bound cases, while it is slightly left shifted in the UL15 peptide bound case (Fig. 13d). In region D2, we note a broader distribution for the MBP case; while a narrow distribution in the microbial peptide cases and left shifted peak distribution was noted in the UL15 peptide case. In region D3, we note a perfect overlap this time between MBP and UL15 peptide cases, while a broader and right shifted distribution in the PMM peptide case. In region D4, we note a left shifted distance peak distribution for the MBP case, with respect to the two microbial peptide cases.

Moreover, we observed region D4 to be very flexible for the PMM bound case, with fluctuation in the range 12-16 Å. Overall, we found MHC binding groove displayed higher flexibility in PMM bound case, in particular in the regions D3 and D4. This finding is consistent with a higher value of configurational entropy observed (Table 3) for the MHC binding groove in the PMM bound case.

Comparative study for p-MHC complex simulations without TCR. To address the role of peptide/HLA-DQ1 complex dynamics in T-cell receptor crossreactivity we performed additional simulations without TCR and compared the results with simulations performed with the TCR. The average interaction energy values calculated from MD simulations suggested a better peptide-MHC binding in simulations performed without the TCR (Fig. 14a).
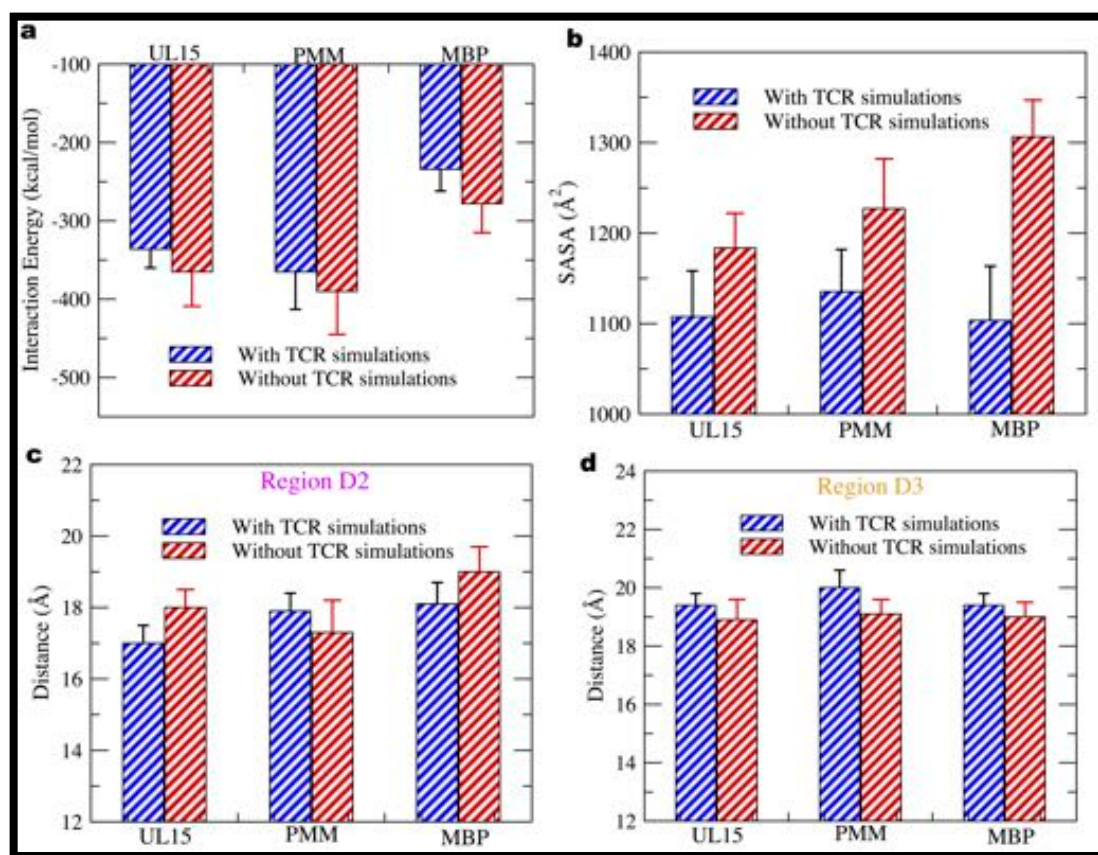
Figure 14. HLA-DQ1 binding groove analysis in the presence and absence of TCR. In (a) peptide-HLA-DQ1 interaction energy corresponds to non-bonded energy values comprising of Van der Waals and electrostatic energy (b) HLA-DQ1 peptide binding groove SASA. In (c) and (d) average distance values in region D2 and D3 of the HLA-DQ1 binding groove.

Furthermore, the average solvent accessible surface area (SASA)[107] values of HLA-DQ1 binding groove showed an increased value in all three peptide cases for TCR free simulation cases (Fig. 14b). The maximum increase in average value of SASA (~15%) was noted in MBP peptide complex. Binding groove width analysis on a local scale was performed in the absence and presence of TCR on the MD simulations trajectories. We found significant variation in average distance values in the regions D2 (Fig. 14c) and D3 (Fig. 14d) in all the three-peptide complexes. In region D2 (Fig. 14c) we found that groove width increases when not bound to TCR, in the UL15 and MBP peptide complexes. However, for all the peptide cases, in region D3 we note the groove to be slightly narrow in the absence of TCR (Fig. 14d). Only in the MBP peptide case, we observed an opening (~1.5 Å) of HLA-DQ1 region D4 when not bound to TCR (Fig. 15).
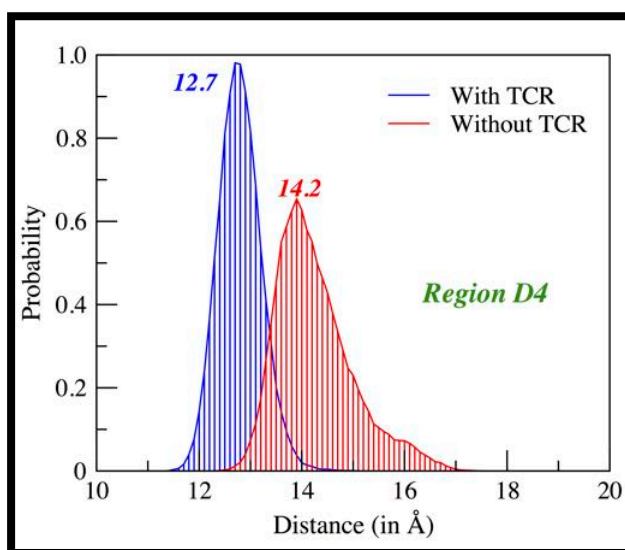
Figure 15. Region D4 distance probability plot for MBP bound MHC binding groove.

To understand better the differences noted in the binding groove dynamics between the TCR bound and unbound cases, we examined the peptide-MHC interaction network (Table 4). In all the peptide complex cases, we note a striking absence of H-bond interaction between DQ1 αR61 and P6 peptide residue in the simulations performed without TCR. On the other hand, we observed a novel H-bond interaction

for the unbound TCR simulations between DQ1 βE66 and P5 peptide residue, in all the peptide complexes. Overall, we found an increase (~6-8%) in the HLA-DQ1 binding groove entropy values in the TCR unbound cases, suggesting a much more flexible binding groove in the absence of TCR.

3.4 Discussion and Conclusions

The objective of our study has been to provide dynamical insight into Hy.1B11 TCR crossreactivity between MBP self-peptide and two microbial peptides while bound to HLA-DQ1 complex. To address this issue, we performed molecular dynamics simulations on available experimental atomistic model of Hy.1B11 TCR from a patient with relapsing-remitting MS disease. The central role of TCRs is to recognize the peptide presented by MHC molecules and provide an immune protection against foreign peptides[108].

Table 4. **Peptide-MHC hydrogen bonded interactions and entropy evaluation in the presence and absence of TCR.** (a) H-bond interactions absent in TCR-unbound simulations with respect to TCR–bound. (b) New interactions in unbound-TCR simulations. (c) MHC binding groove entropy analysis; an increase is reported in % with respect to TCR simulations.

	UL15	PMM	MBP
a. Absent H-bond interactions in simulations without TCR			
DQ1 α-residues	αR61—D9 (P6) αC8—V7 (P4) αN11—D9 (P6)	αR61—D9 (P6)	αR61—N9 (P6) αH68—T12 (P9) αN62—F7 (P4)
DQ1 β-residues	βY37—Q12 (P9) βS74—R8 (P5)	-	βY9—N9 (P6) βH30—N9 (P6)
b. New H-bond interactions in simulations without TCR			
DQ1 α-residues	αY77—Q12 (P9)	-	αC8—F7 (P4)
DQ1 β-residues	βE66—R8 (P5)	βE66—R8 (P5)	βE66—K8 (P5)
c. Configurational entropy (J/molK)			
HLA-DQ1	6068 (7%↑)	6308 (6%↑)	6339 (8%↑)

It is interesting to note that the total number of possible peptides of 14-mer length that can be constructed from the 20 amino acids are of order $\sim 10^{18}$. Without entering in the details of the peptide and complex structural constraints, even assuming only a very low percentage of this peptide repertoire to bind to MHC class II molecules, the possible number of peptides is still many order of magnitude greater than the theoretical number of possible TCRs in humans ($\sim 10^{8}$). Thus, the bandwidth of TCR cross-reactivity is inevitable to compensate this disparity and to provide an immune cover for vast number of peptide-MHC complexes[109]. TCR cross-reactivity can have both positive and negative implications. On one hand, a positive implication can be providing polyclonal recognition of peptide-MHC molecules, thus providing immune cover against pathogens that escapes recognition. While, on other hand, a negative consequence can be for causing autoimmune diseases[109].

Previous crystallographic studies of two human TCR's[92,110] from MS patients, displayed between them a different binding geometry to the peptide-MHC complex, but a common CDR footprint displaced towards the N-terminal of the bound MBP peptide. In our study, the Hy.1B11 TCR structure not only displays a different binding geometry with respect to these human TCR structures, but also a different CDR footprint. In particular, with CDR2β loop to overlay onto the central portion of HLA-DQ1 α-helix (Fig. 11) and CDR3α, CDR3β chains positioned towards the center portion of peptide binding groove.
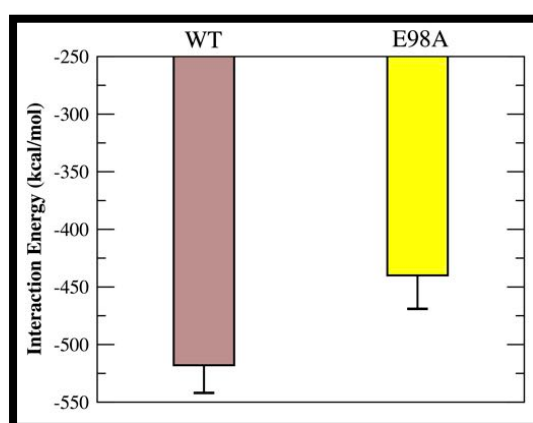
Figure 16. Interaction Energy plot. Comparison of interaction energy values between TCR and MBP-MHC residues for WT and CDR3α E98A mutant TCR simulations.

Our MD simulations confirmed the energetic role of CDR3α E98 residue[94] in interaction with the P5 (Lys/Arg) peptide segment (Fig. 9). Mutation of CDR3α residue E98 to alanine (A98) resulted in a significant reduction in interaction energy value (Fig. 16) between TCR and MBP-MHC residues with respect to wild-type case. These observations are consistent with previous experimental data, wherein alanine substitution of P5 resulted in a complete loss of activity and CDR3α E98A mutation was particularly severe to HLA-DQ1 peptide binding[56]. Previous structural investigation of Hy.1B11 TCR for HLA-DQ1-peptide (MBP, UL15, PMM) complexes[56] suggested a dominant role of CDR3α F95 residue for recognition of both self-peptide (MBP) and microbial peptides (UL15, PMM). No such predominant role is observed from our MD simulations. However, we do observe persistent stacking interactions between CDR3α F95 and with peptide pockets P2 and P3 in the MBP and PMM peptide cases, respectively.

Recent experimental study[85] investigated the TCR-peptide-MHC cross-reactivity for nine peptides with limited sequence similarity and noted a consistent TCR-MHC interaction mode in all peptide complexes. With a similar strategy, we investigated how the TCR-MHC docking configurations changed during MD simulations by dPCA analysis and also evaluated the interaction mode for the three cross-reactive peptide complexes. Notably, the three different peptide complexes superimposed neatly (Fig. 10c-d, within a RMSD difference of 2.0 Å) and displayed similar values of total buried surface area, obtained from MD simulations (Table 1). Specifically, about MHC class II system one must consider that the peptide-binding groove is quite well characterized in literature and one can expect only local rearrangements in bound peptide conformations. Therefore, no drastic difference in the global conformation between the three-peptide complexes is expected with respect to the starting crystal structures.

Moreover, a common interaction picture (Fig. 11) was noted in the peptide complexes, mediated through TCR Vα domain, in which CDR3α E98 formed H-bonds to R61 of HLA-DQ1α and through Vβ domain, in which CDR2β D55 formed H-bonds to K39 of HLA-DQ1α (Table 2). Indeed, surface plasmon resonance (SPR) experiments showed CDR3α E98-DQ1α R61 interaction to be energetically the most important interaction between the Hy.1B11 TCR and HLA-DQ1 residues[56].

31

An energetic picture of cross-reactive Hy.1B11 TCR was obtained by performing interaction energy calculations to estimate the binding affinity of TCR for three DQ1-peptide complexes (Fig. 13a). MD simulations provided best interaction energy value of Hy.1B11 TCR for DQ1/MBP complex, than DQ1/PMM and DQ1/UL15 complexes. This trend in TCR binding affinity for DQ1-peptide complexes is in perfect agreement with SPR experiments[56], which showed Hy.1B11 TCR higher binding affinity for DQ1-MBP complex. Furthermore, the same experiments showed the lower TCR affinity for the DQ1-microbial peptide complexes to be partially compensated by higher value of binding affinity of PMM and UL15 for DQ1, compared with the MBP peptide. To address the same, we evaluated the interaction energy values of DQ1 for MBP, UL15 and PMM peptides (Fig. 13b). Our data correlated well with the experimental data. Thus, supporting the importance of an energetic balance between both TCR affinity for DQ1-peptide and DQ1 affinity for peptide in Hy.1B11 TCR cross-reactivity. With the unique exception of P5, all other key anchor (pockets) peptide residues, contributed to the DQ1-peptide binding affinity (Fig. 12, Table 3). In spite of different anchor residues between the microbial peptides and MBP self-peptide, we note conserved interaction with DQ1 residues. Interestingly, the conserved peptide residue (Phe) at P3 position formed persistent stacking interaction with DQ1 residue αR61 (Table 3). The predominant role of peptide residue at P6 position in H-bond interaction with both α-chain (Fig. 12a) and β-chain residues (Fig. 12b) of DQ1 binding groove was observed. The two microbial peptides shared a common peptide residue (aspartic acid) at P6 position, different from asparagine residue in MBP. This difference at P6 position was reflected by a lesser persistent interaction between asparagine residue in MBP and residues Y9 and H30 of DQ1 β-chain (Fig. 12b), consistent with lower binding DQ1 affinity of MBP. The local fluctuations of the binding grove in the four regions highlighted similarity in width fluctuation between self-peptide MBP and the microbial peptides: in region D1 involving UL15, while in region D2 with PMM (Fig. 13b). A wider and flexible regions D3 and D4 in the PMM complex resulted in an overall higher value of entropy (Table 3).

To understand the role of peptide-MHC dynamics in Hy.1B11 TCR cross-reactivity, we performed additional simulations of the peptide-MHC complexes without the TCR. In the simulations without TCR, we observed an overall increase (6-8%) in

HLA-DQ1 binding groove flexibility (Table 4), and the trend in local MHC width fluctuations to be inverted between regions D2 and D3. A slight but a significant difference in peptide-DQ1 binding affinity (Fig. 14a) and an increase in solvent accessible surface area (Fig. 14b) was observed. Interestingly, in the simulations without TCR, we note absence of a conserved H-bond interaction between peptide residue at position P6 and residue R61 of DQ1 α-chain, and a new interaction between peptide residue at position P5 and residue E66 of DQ1 β-chain, in the three peptide-MHC complexes (Table 4). These observations confirm the role of TCR in bridging interaction between peptide position P6 and residue R61 of DQ1 α-chain. It is important to mention that in the TCR simulations, peptide residue at position P5 was the one involved in interactions with TCR residues. In summary, we found the presence of TCR to have an important impact on both local and global level description of peptide-MHC interactions.

In conclusions, using MD simulations, we identified a bridging interaction involving CDR3α (E98) – DQ1α (R61) – peptide (P6) as an energetic hot spot on the TCR-peptide-MHC interface that contributes to Hy.1B11 TCR cross-reactivity (Fig. 17). We further identified a structurally relevant new H-bond interaction between CDR2β D55 and DQ1α K39 to constitute a key anchor point for interaction of TCR on to the MHC class II. Our findings confirm the energetic role of CDR3α residue E98 in Hy.1B11 TCR cross-reactivity.
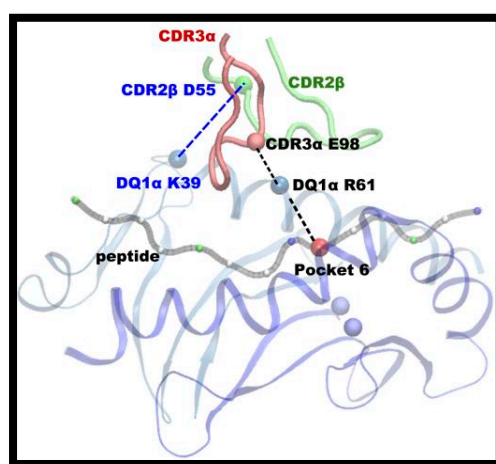
Figure 17. Bridging interaction across TCR-MHC-peptide interface determining Hy.1B11 TCR cross-reactivity. H-bond interactions are represented with dashed lines.

Altogether, using MD simulations we were able to identify involvement of a small number of structurally and energetically important hot spots that provide dynamical insights into Hy.1B11 TCR cross-reactivity.

3.5 Methods

Model preparation. In the current study, we chose the available X-ray structures of the tri-molecular TCR-peptide-MHC-II complexes for the two microbial peptides UL15 (PDB id: 4MAY), PMM (PDB id: 4GRL) and one self-peptide MBP (PDB id: 3PL6) complex. The tri-molecular structures consisted of same Hy.1B11 TCR, same HLA-DQ1 protein complexed alternatively with the three peptides under investigation (Fig. 9). HLA-DQ1 protein is a heterodimer composed of two chains: α (DQA1*0102) and β (DQB1*0502); and the peptide-binding groove is formed from two non-covalently linked subunits of α1 (5-90 residues) and β1 (5-90 residues) chains. The missing hydrogen atoms in the X-ray structures of the three TCR-peptide-MHC-II complexes were built using psfgen package of VMD software[111]. Each of the trimolecular complex system was then immersed in a water box, and subsequently counter-ions were added in order to have a neutral system. Details about the simulation box size and the total number of atoms for each of the systems are presented in Table 5.

Table 5. MD simulation of peptide-MHC complexes with and without Hy.1B11 TCR. We report total number of atoms and simulation box size in each of the system investigated.

Peptide – MHC complex	With Hy.1B11 TCR		Without Hy.1B11 TCR	
	Total atoms	Box-size	Total atoms	Box-size
MBP-HLA-DQ1	184164	[113 111 162]	48465	[79 67 107]
PMM-HLA-DQ1	184122	[113 111 162]	48450	[79 67 107]
UL15-HLA-DQ1	184151	[113 111 162]	48479	[79 67 107]

TIP3P[112] parameters for water molecules and Charmm-27 force-field parameters for the proteins and peptides were used. Correct protonation state was assigned to all the protein and peptide residues using Propka software[113]. For TCR-unbound peptide-MHC complexes, we removed the TCR structure from the tri-molecular complex and performed MD simulations of only the peptide-MHC complexes. For simulations of peptide-MHC complex, we chose the chains corresponding to peptide and MHC molecules from the solved crystal structure. We performed simulations both with/without TCR for the three systems. Each system was energy minimized and slowly heated to 310 K in steps of 30 K with initial positional constraints of 50 kcal/(mol $Å^2$) on carbon alpha atoms. Subsequently, positional constrains was slowly released in steps of 10 kcal/(mol $Å^2$). Molecular dynamic simulation of 110 ns was performed in NPT ensemble with T=310 K, and 1 atm pressure. Further simulation protocol details have been described in our previous works[114,115]. All-atom molecular dynamics (MD) simulations were performed employing NAMD[116] software package on 64 processors cluster.

Simulation analysis. MD trajectory of a total simulation time 110 ns, for each complex under investigation, was used for analysis. The stability of protein-peptide-protein complexes and peptide-protein complexes was evaluated by calculating the root mean square deviation (RMSD) values for the C-alpha atoms of residues during MD simulations (Fig. 18).
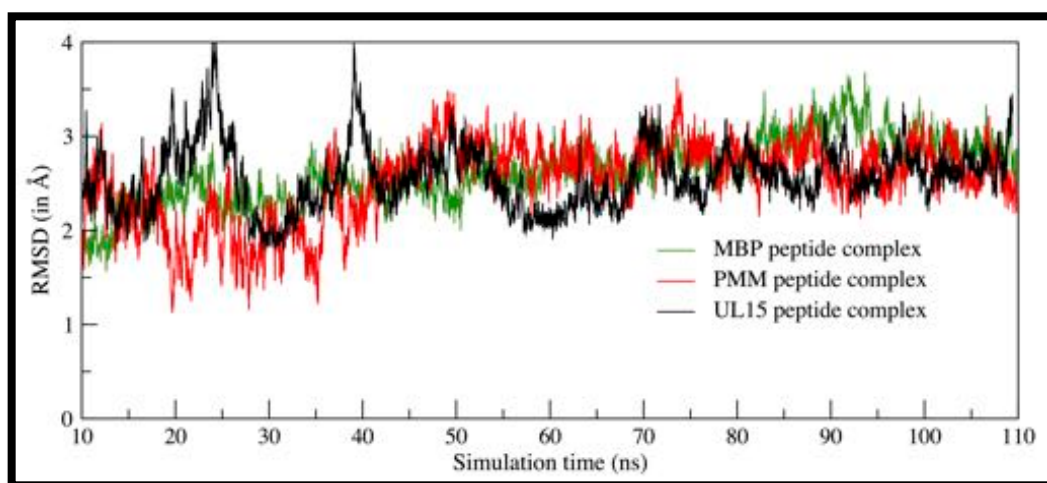
Figure 18. RMSD plot of C-alpha atoms for TCR-pep-MHC complexes.

To estimate convergence of our MD simulations, we employed a novel Good-Turing statistical approach proposed recently by Koukos and Glykos[102]. This method allows estimating the probability distribution of unobserved configurations, $p_{unobserved}$(RMSD), as a function of the RMSD distance between unobserved and observed molecular configurations in MD simulations (Fig. 19).
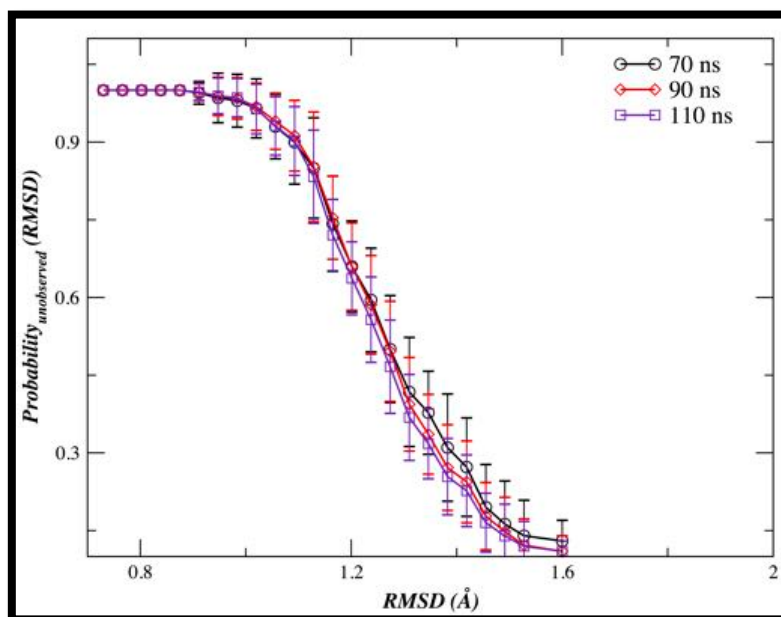
Figure 19. Good-Turing convergence test. Probability$_{unobserved}$ (RMSD) as a function of RMSD distance. For TCR bound MBP-MHC complex at different simulation time lengths.

The hydrogen bonded (H-bond) interaction between peptide-protein or protein-protein residues pairs was calculated using a geometrical criterion, with a donor-acceptor cutoff distance of 3.1 Å and donor-hydrogen-acceptor cut-off angle 130 degree. H-bonds present for at least 20% of trajectory time length are reported. The aromatic stacking interaction between the residues pairs was calculated using EUCB software[117] with following geometrical criteria: – maximum dihedral angle cut-off parameters between the planar/ring side chains of 30° – centroid distance cut-off between side chains 5.0 Å – persistence simulation time 20%.

The interaction energy between the two selected groups of atoms (for instance, between the peptide residues and HLA-DQ1 residues) was calculated by evaluating

the non-bonded energy values comprising of Van der Waals and electrostatic energy, using the energy plugin of Namd software[116]. A cutoff distance of 12 Å was used for non-bonded interactions and for the electrostatic interaction we also adopted the particle mesh Ewald scheme[118]. The interaction energy scheme adopted in our calculations provides only a rough estimate in terms of enthalpic contributions to binding, as solvent effects are not included. Thus, interaction energy values obtained can be used only for the ranking the different systems based on their energy values.

Configurational entropy calculations using the quasi-harmonic approximation scheme[105] were performed to investigate differences in protein flexibility and stability between the different protein-peptide complexes. From 110 ns MD simulation trajectory of each system, we extracted 5500 extracted structures, and configurational entropy estimate was done by evaluating the covariance matrices of atomic fluctuations of selected residues, within a routine of CARMA software[119].

Dihedral angles principal component analysis (dPCA) was performed on selected TCR and MHC class II binding site residues on MD simulation trajectory using CARMA software[119]. This resulted is a matrix containing the values of the top three principal components for each and every structure recorded in the trajectory. To perform cluster analysis, the method incorporated within the CARMA software uses a peak-picking algorithm that is applied to three-dimensional density distributions of the principal components derived from the MD trajectory. The classification of different clusters is done automatically using the density distribution threshold that can explain at least 80% of the original principal component map's variance. Each classified clusters, represents a different number of structures (from the trajectory) that have values for their principal components corresponding to the specific point of the principal component plane. Cluster 1 corresponds to the most populated cluster with highest number of structures (from MD trajectory). A major limitation of this methodology it that it does not comprehensively assign each frame of a trajectory to a cluster. Indeed, the algorithm aims at efficiently identifying the most prominent molecular conformations.

4. 2-Phenylbenzofuran ligands as selective butyrylcholinesterase protein inhibitors

Alzheimer's Disease (AD) is an irreversible progressive neurodegenerative brain disorder affecting more than 44 million people with partly understood pathophysiology[120]. An accepted treatment strategy is to restore the levels of acetylcholine by inhibiting cholinesterase enzymes, such as butyrylcholinesterase (BChE) and acetylcholinesterase (AChE).Benzofuran scaffold has drawn considerable attention over the last few years, with many studies investigating their inhibitory activity towards the above-mentioned enzymes.

In this chapter, inhibition activity of a series of 16 2-phenylbezonfurans compounds towards these enzymes is presented. The inhibitory activity (IC_{50}) of the 16 compounds was evaluated against BChE enzyme was compared with the reference compound galantamine. To better understand the enzyme inhibition mechanisms, in relation to the substituents and their positions in our compounds, molecular modeling studies were also performed. The best inhibitory activity for the compound with a hydroxyl substituent in the 2-pheynl ring and chlorine atom at position 7 of benzofuran moiety.

Overall, biological assays and MD simulations results were integrated to highlight the molecular basis of the selective BChE inhibition by the benzofuran scaffold. This work has been published in Bioorganic Medicinal Chemistry Letters[121].

4.1 Abstract

A series of 2-phenylbenzofurans compounds were designed, synthetized and evaluated as cholinesterase inhibitors against Alzheimer's disease (AD). Methoxy and hydroxy substituents were introduced in the 2-phenyl ring, whereas the position 5 and 7 in the benzofuran moiety was not substituted or substituted with methyl, bromine and chlorine atoms. The activities of each compound were evaluated towards both acetylcholinesterase (AChE) and butyrylcholinesterase (BChE) enzymes. The biological assay experiments showed that most of the compounds displayed a clearly selective inhibition for BChE, while a weak or no effect toward AChE was detected. Among these benzofuran derivatives, compound **16** exhibited the highest BChE inhibition with an IC_{50} value of 30.3 µM. Lineweaver-Burk plot and molecular modeling studies revealed that compound **16** was a mixed-type inhibitor, binding to both catalytic active (CAS) and peripheral anionic sites (PAS) of BChE. Moreover, the same compound exerts antioxidant activity and resulted to be non-cytotoxic towards motor-neuron cells. Furthermore, we employed molecular dynamic simulations to provide information on enzyme dynamics and interaction energies between the aforementioned ligands and both cholinesterase enzymes. In particular, the simulations revealed that compound **16** binds to both the CAS and the PAS of BChE and displayed the best interaction energy value, in agreement with our experimental data.

4.2 Introduction

Alzheimer's disease (AD) is an irreversible and progressive brain disorder which is characterized by progressive memory loss and a wide range of cognitive impairments[122]. Although the precise cause of AD is not completely known, there are some factors that seem to play a significant role in the pathogenesis of AD, such as: deficit of acetylcholine (ACh), presence of amyloid-β deposits, τ-protein

aggregation, oxidative stress and metal ions imbalance. Among these distinct research approaches, the cholinergic hypothesis has been examined more extensively. In fact, low levels of ACh appear to be a critical element in the development of cognitive and neurodegenerative disorders in AD patients[123].

Accordingly, one strategy in AD treatment is to restore the levels of ACh by inhibiting acetylcholinesterase (AChE, EC 3.1.1.7) and butyrylcholinesterase (BChE, EC 3.1.1.8) which are mainly responsible for ACh hydrolysis. These enzymes belong to the superfamily of α/β-hydrolase fold proteins and are able to hydrolyse ACh with different efficiencies[124]. They are encoded by two distinct human genes and display 65% homology in their amino acid sequences. These two proteins also show a great similarity in both their tertiary structure and their overall architecture of active sites[125,126]. Both AChE and BChE have indeed a primarily hydrophobic active gorge into which ACh diffuses and is cleaved.

Ligand binding specificity between the two enzymes, have been related to differences in the residue's structural arrangement which lead to the active site located near the bottom of a deep and narrow gorge (Figure 20, Table 6). The gorge is characterised by (i) a peripheral site at the entrance, (ii) an oxyanion hole, (iii) a choline-binding site located within the entrance, and (iv) the active site constituted by an acyl pocket buried near the catalytic triad.
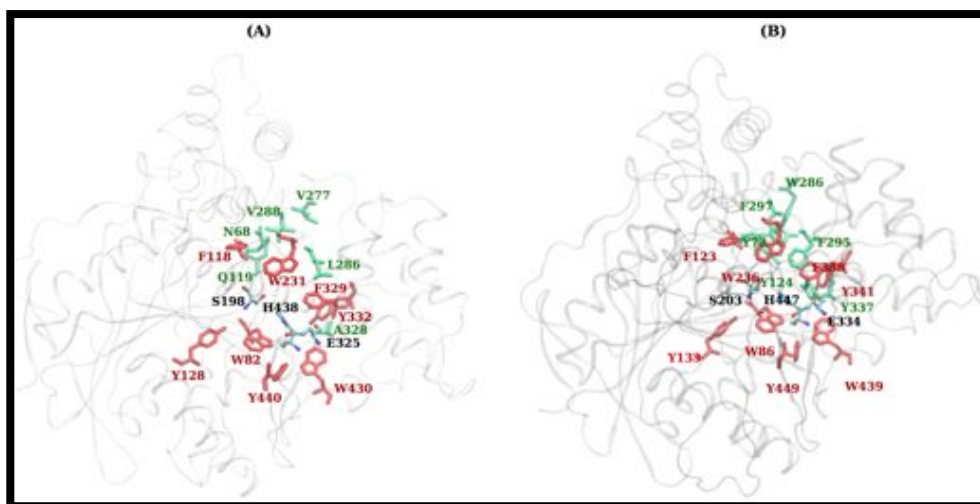
Figure 20. Enzymes under investigation. (A) *Equine serum* BChE (B) *Electrophorus electricus* AChE. The residues lining the gorge of the two enzymes are shown. The conserved residues between the two are shown in red (licorice representation) and non-conserved in green. The catalytic triad residues are shown in ball-stick representation.

41

In fact, structural analysis had revealed that these enzymes have two major substrate-binding sites. One is a peripheral anionic site (PAS) at the entrance of the gorge, acting as a regulatory site; the other is the catalytic anionic site (CAS), which is located in the bottom of the gorge and is assigned to a Ser-His-Glu catalytic triad[127].

Table 6. **Key residues lining the gorge site of the two Cholinesterase's (ChE)**. The non-conserved residues of BChE with respect to AChE are highlighted in bold. Non-conserved ChE's residues in human are shown in red.

	Equine serum BChE	*Electrophorus electricus AChE*
Catalytically active site	S198, E325, H438	S203, E334, H447
Peripheral Site	**N68, Q119, V277** (A277), D70, Y332	Y72, Y124, W286 D74, Y341
Acyl Pocket	**L286, V288,** F329	F295, F297, F338
Wall of Gorge	Y114, Y128, W231 W430, Y440	Y119, Y133, W236 W439, Y449
Oxyanion hole	G116, G117, A199	G120, G121, A204
Choline binding site	W82, **A328**	W86, Y337

AChE and BChE appear to be simultaneously active in the synaptic hydrolysis of ACh, terminating its neurotransmitter action, and co-regulating levels of ACh[128]. AChE has a very high catalytic efficiency for ACh hydrolysis and is mainly found in cholinergic synapses, while BChE has lower efficiency and is widely distributed in tissues and plasma. In the healthy brain, AChE predominates and BChE is considered to play a minor role in regulating ACh levels. On the

contrary, BChE activity increases in the temporal cortex and hippocampus during the development of AD, while at the same time AChE activity decreases[129].

Since AD is characterized by a forebrain cholinergic neuron loss and a progressive decline in ACh, a possible therapeutic strategy involves the use of cholinesterases inhibitors to restore the neurotransmitter level and thus alleviate AD symptoms[130-132]. These inhibitory molecules may act by binding the CAS site (competitive mechanism) or PAS (non-competitive mechanism) or may exert a dual binding enzyme inhibition – acting as mixed-type inhibitors[133]. Moreover, since the oxidative stress may be a risk factor for the initiation and progression of AD, drugs with both antioxidant and inhibitory actions, might be useful for either the prevention or the treatment of AD.

The benzofuran (oxygen heterocycle) is a common moiety found in many biologically active natural and therapeutic products and thus represents a very important pharmacophore[134]. It is present in many medicinally important compounds that show biological activity, including anticancer and anti-inflammatory properties[135]. Benzofuran scaffold has drawn considerable attention over the last few years due to its profound physiological and chemotherapeutic properties[136]. Some benzofuran derivatives are also known as MAO and 5-lipoxygenase inhibitors, antagonists of the angiotensin II receptor, blood coagulation factor Xa inhibitors, ligands of adenosine A1 receptor and so forth[137]. Recent studies have also investigated their inhibitory activity towards AChE[138].
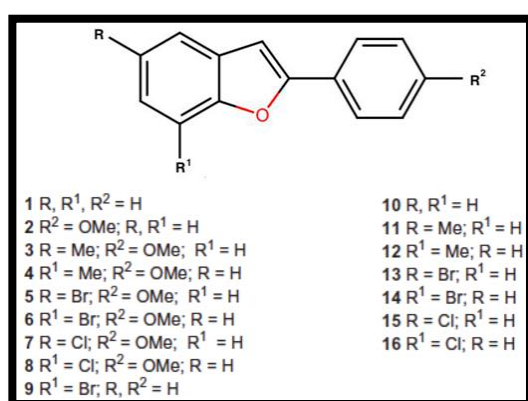
Figure 21. 2-Phenylbenzofuran derivatives with functional groups R, R^1 and R^2.

43

In Fig. 21, the complete list of 16 2-pheynl compounds which were synthesized and differ with each other in either of their functional groups at position R, R^1 and R^2. The activity of all the synthetized compounds was first evaluated at compound concentration of 100 µM (Table 7). As observed, only compounds **9** and **11** exert a very weak inhibitory activity towards AChE, while all the compounds, except for **2** and **10**, inhibit BChE activity with a varying efficiency. In particular, compounds **12**, **14** and **16** show the highest inhibition percentages and the lowest IC_{50} values.

Table 7. Cholinesterase inhibitory activity of compound 1, 2, 8-16.

Compound	% Inhibition at 100 µM		IC_{50} (µM)
	AChE	BChE	BChE
1	-	6	> 100
2	-	-	> 100
8	-	5	> 100
9	4	12	> 100
10	-	-	> 100
11	4.8	19	> 100
12	-	58	77 ± 6.7
13	-	16.6	> 100
14	-	54	82.5 ± 7.1
15	-	15	> 100
16	-	77	30.3 ± 1.9
Galantamine			28.29 ± 2.12

4.3 Results and Discussion

The series of 2-phenylbezonfurans compounds (Fig. 21) used in this study are composed by: (i) benzofuran moiety with substituents at position 5, 7 (R, R^1) and (ii) 2-phenyl ring with substituent at position *para* (R^2). The biological assays showed that compounds **12**, **14** and **16** displayed high BChE inhibition (> 50%). As already observed, these compounds share the same substituents (Fig. 21) R (H) and R^2 (OH), but differ in R^1 (Met, Br, Cl respectively). Both chlorinated benzofurans **15** (R = Cl; R^2 = OH) and **8** (R^1 = Cl; R^2 = OMe), resulted in low values of BChE inhibition (Table 2). To understand the impact of substitutions (benzofuran moiety and 2-phenyl ring) on enzyme inhibition ability, compounds **8**, **15** and **16** were selected for molecular dynamics simulations. Compound **16** showed the best inhibition activity with compound **15** differing at chlorine position and compound **8** differing at *para* position with respect to compound **16** (see Fig. 22).
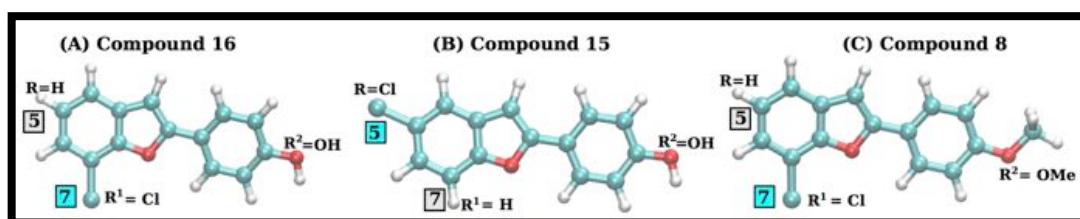
Figure 22. Three compounds selected for Molecular modeling investigation.

Molecular docking results showed that the three compounds were able to interact with CAS residues in BChE, and with PAS in AChE. The three compounds displayed very similar docking energies (~6-7 kcal/mol) in both cases. To investigate the structural and dynamical aspects upon ligand binding it is necessary to perform molecular dynamics simulations. For this purpose, we simulated the two enzymes for 50 ns in the free and ligand bound configurations. The systems stability was monitored during MD simulations by evaluating the root mean square deviation (RMSD) values for the backbone atoms of the two enzymes (Fig. 23). After a careful analysis of the MD trajectories, we found the interaction sites of the two enzymes in complex with the ligands to be stable, also noted from the low RMSD ligand fluctuations (< 0.5 Å).
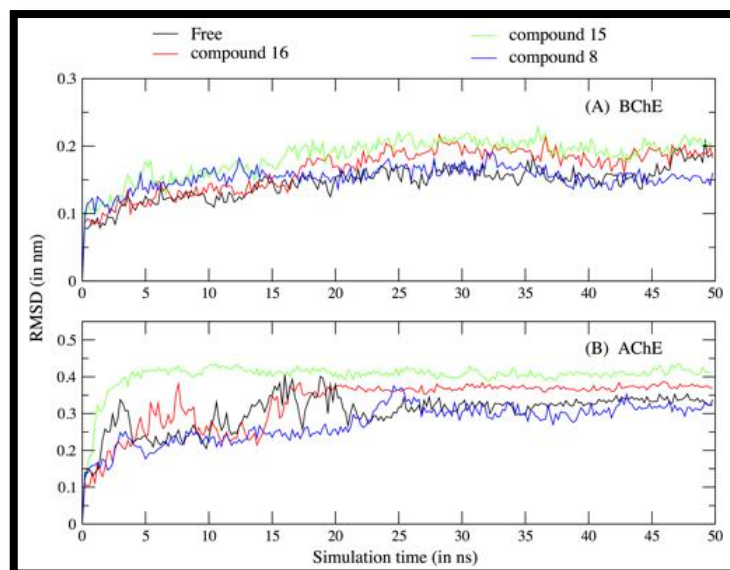
Figure 23. Root mean square deviation of complex systems under investigation.

Enzyme dynamics in presence of the three compounds (**8**, **15**, **16**) was investigated by evaluating dynamic cross-correlation map for carbon-alpha atoms, calculated on 500 snapshots extracted from 50 ns MD trajectories using Prody software[139]. A clear difference is observed in the enzyme ligand complex dynamics between compound **16** and compound **15** (Fig. 24).
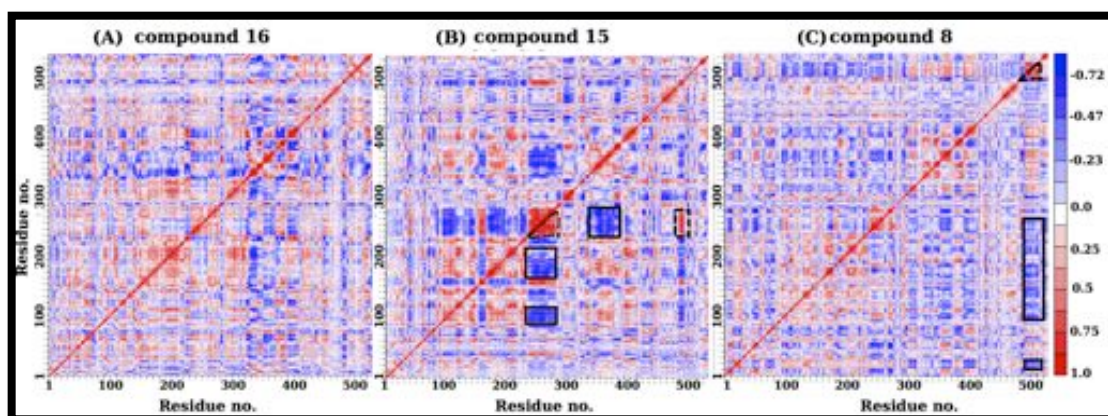
Figure 24. Cross-Correlated motions of *equine serum* BChE bound to the three compounds. In (A) compound **16** (B) compound **15** and (C) compound **8**. Positive correlations are indicated in red and negative correlations in blue. In (B) and (C); with boxed regions represent those displaying better positive correlation (red) and better negative correlation (blue) with respect to compound **16**.

Changing chlorine from position 7 (R^1) to position 5 (R) (switching compound from **16** to **15**) has a significant effect on enzyme dynamics (Fig. 24) in at least three regions, as shown by the increased negative correlation. On the other hand, only changing the substituent R^2 (compound **8** and **16**) results in a very little difference in enzyme dynamics. Overall, these results suggest that substitution in positions 7 for benzofurans compounds play an important role in enzyme dynamics.

We also estimated the binding strength by evaluating the interaction energy between the enzyme residues and the compound (Table 8).

Table 8. Interaction energy and buried surface area (BSA) evaluation.

BChE with	VDW* (kcal/mol)	Elec.** (kcal/mol)	Total energy (kcal/mol)	BSA (in %)
compound **16**	-22.4	-26.5	-48.9	75
compound **15**	-10.1	-27.5	-37.6	64
compound **8**	-8.1	-29.8	-37.9	58

* Van-der-Waal (VDW);**Electrostatic (Elec.)

In the case of BChE, compound **16** displayed the best interaction energy, with respect to the other two compounds, that displayed very similar energies (Table 8). Interestingly, the same compound **16** also possessed the highest percentage of its total surface area buried in the gorge (75%) with respect to the other two compounds. To understand the origin of these differences, we carefully inspected the binding mode of the ligands in complex with BChE (Fig. 25), and with AChE (Fig. 26, shown for comparison). The interaction picture was plotted using Ligplot[140].
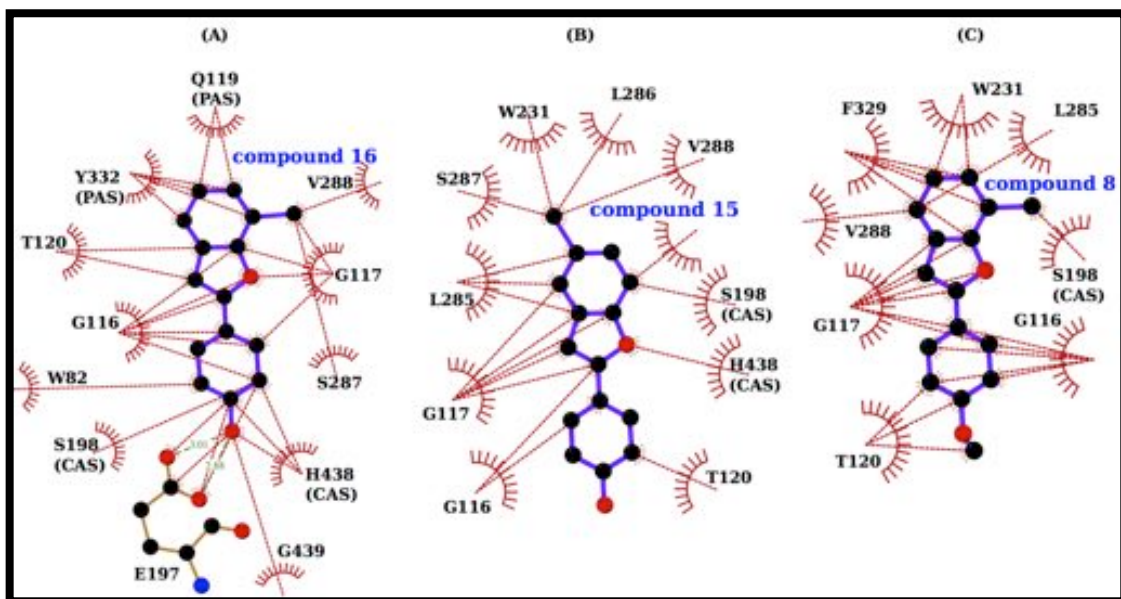
Figure 25. Molecular interaction picture of *equine serum* BChE bound to the three compounds. In (A) compound **16** (B) compound **15** and (C) compound **8**.
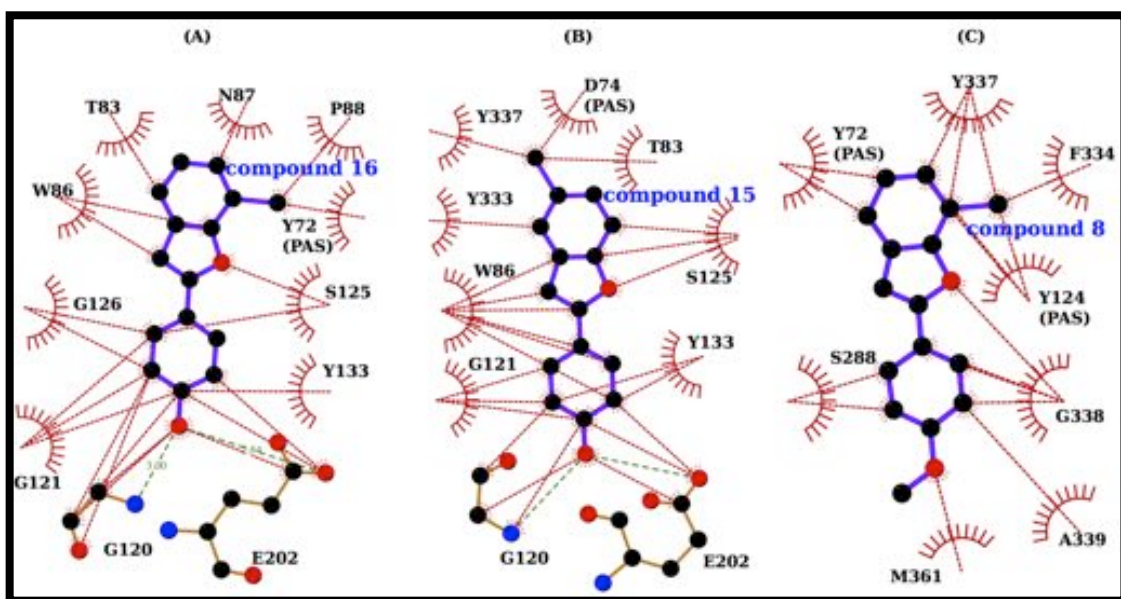
Figure 26. **Molecular interaction picture of *Electrophorus electricus* AChE bound to the three compounds.** In (A) compound **16** (B) compound **15** and (C) compound **8**.

In the case of AChE the three compounds (**8, 15** and **16**) do not interact with any CAS residues, thus, no enzyme inhibition was observed (Fig. 26). On the other hand, in BChE the three compounds interacted with at least one of the three CAS residues, thus displaying enzyme inhibition. Therefore, the simulations of our models suggest interaction of 2-phenylbenzofurans with CAS residues to be crucial for enzyme inhibition. Concerning *Equine serum* BChE, in the three compounds under investigation we observed a mixture of conserved and non-conserved interacting residue partners (Fig. 25).

Previous biochemical and molecular studies[141] found the difference in inhibitory property of E2020 towards AChE and BChE to be related to simultaneous participation of CAS and PAS binding residues (present only for AChE). Our modeling results confirm a similar effect for 2-phenylbenzofurans (compound **16**). The participation of both CAS and PAS residues also resulted in better interaction energy for the compound **16** complex. Interestingly, in all three complexes, the compounds globally interacted with CAS residues (S198, H438). However, only in compound **16** we find the 2-phenyl ring moiety involved in interaction with CAS residues (Fig. 25). Moreover, only in compound **16**, do we observe the benzofuran moiety to interact with PAS residues (Q119, Y332) and the 2-phenyl ring moiety to interact with W82 and residues E197 and G439. Thus explaining the high interaction energy and the related high inhibition (Figure 25). In a recent study, residue Q119 has been shown to be important towards selective inhibition of mouse BChE by two biscarbamates compounds[142]. Residue W82 (W86 in AChE) has also been suggested as a crucial component of the anionic site[127] and as a controller for CAS opening and closing of CAS[143]. Moreover, previous studies analyzed the relevance of residue E197 (E202 in AChE) in substrate inhibition[144] and of residue Y332 (Y341 in AChE) in substrate binding[145]. Overall, jointly with all these previous proposals and findings, our current experimental and modeling results confirm and explain the highest BChE inhibition characteristic noted for compound **16**.

4.4 Conclusions

In conclusion, in this study a series of 2-phenylbenzofurans compounds have been designed, synthetized and evaluated for cholinesterase inhibitory activity. These compounds showed no inhibition toward AChE while inhibiting BChE with varying efficiencies. Compounds **12**, **14** and **16** were the most potent inhibitors. These compounds displayed also the highest antioxidant activity. According to our results, the contemporary presence of a hydroxyl group in the *para* position of the phenyl ring and a substitution at the position 7 of the benzofuran scaffold improved the inhibitory activity, with respect to the other synthesized compounds. In particular, compound **16** exhibited the highest BChE inhibition with an IC_{50} value (30.3 μM). Molecular modeling demonstrated that the interaction of 2-phenylbenzofurans with CAS residues is crucial for enzyme inhibition. All the compounds analyzed here interact with CAS in BChE, while no interactions involving CAS in AChE were detected. Our simulation also revealed that compound **16** binds both CAS and PAS sites in BChE, in accordance to the experimental data, which showed that this compound acts as a mixed-type inhibitor. We can therefore conclude that the combination of biological assays and molecular dynamics simulations, allowed highlighting the molecular basis of the selective BChE inhibition by the benzofuran scaffold.

4.5 Methods

Computational modeling of enzyme-ligand complexes. Preparation of three-dimensional structures for proteins and ligands: the biological assays have been performed using AChE from *Electrophorus electricus* and BChE from *Equine serum*. As no good resolution X-ray or NMR three dimensional structures are available for these proteins, homology modeling was performed using the Swiss-model interface[42]. After BLAST sequence alignment, we selected the following templates: the human BChE (PDB id: 4TPK)[146], for the *Equine serum* BChE enzyme isoform, and the *Torpedo californica* AChE (PDB id: 2W6c)[147], for *Electrophorus electricus* AChE. Two-dimensional atomic coordinates of the ligands were drawn using ChemWriter software and subsequently three-dimensional coordinates were generated using Open Babel software. All the ligands were then subjected to geometry optimization using

the Hartree-Fock basis set 6-31G* within Gaussian03 software package[148]. The charges and the force field parameters for the ligands were evaluated following the standard AMBER protocol[149].

Molecular Docking. Docking was performed using SwissDock web server, which is based on the docking software EADock DSS[150]. The docking mode was accurately chosen, with a blind docking procedure that considers the entire protein surface as a potential target. Using this procedure, a large number of ligand binding modes (~15.000) were generated, with the simultaneous rough interaction energy estimation. The binding modes possessing favourable energies were then ranked and classified into different clusters, this time based on the full fitness scoring function. The most consistent and favourable ligand conformation was chosen from 10 independent docking runs for each ligand and was further considered for molecular dynamics (MD) simulation runs.

MD Simulations. The starting structures of protein-ligand complexes were built using leap module of Amber11. Each complex was inserted in an explicit TIP3P water-box with a minimum distance of 1.8 nm between the solute and box boundary. We used Amber force-field parameters for the enzymes and the compounds were assigned using AMBER protocol. Energy minimization was performed for each system, followed by gradual heating to 300 K in steps of 30 K using positional restraints.[55] The positional restraints were gradually removed during 0.3 ns of simulation time and then an equilibration run of 3 ns was performed. The time step used in MD simulation was of 2 fs using SHAKE. Simulations were performed in NPT ensemble using periodic boundary conditions. Long-range electrostatic interactions were treated with particle mesh Ewald method with 12 Å cutoff for non-bonded interactions. For each complex we performed a MD simulation of 50 ns, using NAMD software package on the CRS4 High Performance Computing facility.

5. New Phenylbenzofuran ligands with an improved selectivity against BChE protein

BChE activity progressively increases in patients with AD, while AChE activity remains unchanged or declines, with changes that become more and more pronounced during the disease course. Therefore, it is important to design compounds that not only selectively interact with BChE but also display an improved inhibitory activity with respect to reference compounds, which might have a relevant role in treatment of patients with advanced AD. In the previous chapter, we reported specific structural characterization of benzofuran moiety able to improve the selective inhibitory activity towards BChE.

In this chapter we investigate the importance of hydroxyl group substitution in the benzofuran phenyl-ring, in the novel synthesized 14 2-phenylbezonfurans compounds. Most of our compounds displayed selective BChE inhibitory activity with varying efficiencies. In particular, we found two compounds to display 4- and 8- times better BChE inhibitory activity than the reference compound, galantamine. The results from this work is submitted and under review.

5.1 Abstract

Alzheimer's disease (AD) is a neurodegenerative disorder representing the leading cause of dementia and is affecting nearly 44 million people worldwide. AD is characterized by a progressive decline in acetylcholine levels in the cholinergic systems, which results in severe memory loss and cognitive impairments. Expression levels and activity of butyrylcholinesterase (BChE) enzyme has been noted to increases significantly in the late stages of AD, thus making it a viable drug target. In the present study, a series of hydroxylated 2-phenylbenzofurans compounds were designed, synthesized and their inhibitory activities toward acetylcholinesterase (AChE) and BChE enzymes were evaluated. Two compounds (**15 and 17)** displayed significantly higher inhibitory activity towards BChE with IC_{50} values of 6.23 μM and 3.57 μM, respectively. Interestingly, the same compounds further exhibited selective inhibitory activity against BChE over AChE. Computational studies were used to compare protein-binding pockets and evaluate the interaction fingerprints of the compound. MD simulations showed a conserved protein residue interaction network between the compounds, resulting in similar interaction energy values. Thus, combination of biochemical and computational approaches could represent rational guidelines for further structural modification of these hydroxy-benzofuran derivatives as future drugs for treatment of AD.

5.2 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative brain disorder, named after German psychiatrist Alois Alzheimer. AD is the most common cause of dementia, accounting for up to 80 % of all dementia cases, as well as being a major cause of death worldwide[120,151,152]. It is common in elderly people over 65 years old and exhibits heterogeneous distribution across the globe, being most prevalent in Western Europe and North America, while less prevalent in Sub-Saharan Africa region[153].

Being a multifactorial neurodegenerative brain disorder, the exact pathophysiology of AD is not yet entirely known[154]. However, several pathogenesis of AD have been suggested: deficits in the cholinergic system[155,156], accumulation and deposits of beta-amyloid outside the neurons in the brain[157], oxidative stress[158] and inflammation[159]. Early studies performed on patients suffering from AD[156] found an altered cholinergic activity, which resulted in cognitive and functional symptoms. In the present study, we focus our attention on the cholinergic system, which is the most cited potential mechanism[160,161]. The cholinergic system directly contributes to regulation and memory process, thus represents a suitable target for the AD drug design[120,162,163]. In the cholinergic system disruption in the levels of acetylcholine (ACh) is caused by hydrolytic action of cholinesterases (ChEs)[164]. ACh is a neurotransmitter that plays a role in the modulation of memory function in normal and neurodegenerative conditions[165].

Butyrylcholinesterase (BChE) and Acetylcholinesterase (AChE) belong to ChEs family of enzymes and play a role in ACh regulation and in the cholinergic signalling[166]. The two enzymes are extraordinarily efficient and are able to cleave more than 10000 ACh molecules per second[167]. AChE is substrate specific in nature and is found in high concentrations in the brain, while BChE is non-specific and is distributed throughout the body[163]. In particular, it is primarily found in the liver, pancreas and associated with glial and endothelial cells in the brain[166,168]. In a healthy brain, the AChE enzyme dominantly degrades ACh while BChE plays only a supportive role. The two enzymes display diverse kinetic characteristics depending on ACh concentrations. At low ACh concentrations, AChE's activity becomes highly dominant, while BChE is more efficient in the hydrolysis at high ACh concentrations[163]. Initial studies underestimated the importance of BChE in human brain owing to its low expression[169]. However, other studies have shown the importance of BChE within the nervous system to be pivotal in the late stages of AD[155,170]. Indeed in patients with AD, BChE activity progressively increases, while AChE activity remains unchanged. Moreover, AChE knockouts experiments performed on mouse models demonstrated the role of BChE to maintain the cholinesterasic function even in the absence of AChE[171].

Despite being encoded by different genes on human chromosomes 7 and 3[172], at molecular level the two enzymes AChE and BChE share nearly 65% sequence homology. The availability of several X-ray crystallographic structures for the two enzymes[173,174] further revealed the similarity of the tertiary structure and particularly the architecture of the active site. The active site consists of a catalytic triad (Ser, His, Glu) and a choline-binding pocket buried nearly 20 Å deep into the surface of the enzymes[127]. The main difference between the two enzymes is located in the acyl-binding pocket, which accommodates the acyl moiety. In detail, two bulky amino acids (Phe) in AChE are replaced with two smaller amino acids; Val and Leu, thus allowing BChE accommodate large and chemically different molecules.
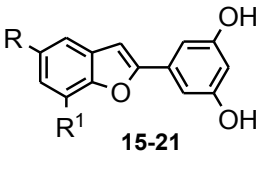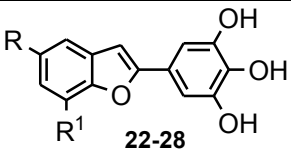
A well-documented strategy towards an effective management of AD is by developing inhibitors that suppress the ChEs enzymes from breaking down ACh and therefore increasing both the level and duration of the neurotransmitter action[162]. Current Food and Drug Administration (FDA) approved cholinesterase inhibitors namely: donepezil, rivastigmine and galantamine, help only in controlling the symptoms of AD and do not treat the underlying disease or delay its progression. In this scenario, a continuous research related to development of more potent and highly efficacious cholinesterase inhibitors becomes even more essential.

Heterocyclic ring compounds are known to display broad biological, medicinal and pharmacological characteristics and thus form an important moiety to construct inhibitors against ChEs enzymes. Among them, benzofuran derivatives, since synthesized for the first time by Perkin[175] in 1870, has been constantly explored in the treatment of various diseases, including AD[176]. Initially most of the research studies were focused on development of AChE inhibitors towards treatment of AD. However, molecules displaying very high selectivity for BChE over AChE have also been developed[177,178]. Recent studies designed and synthesized benzofuran derivatives that displayed a selective inhibitory profile against AChE enzyme[138,179,180]. In this context, we recently developed a series of 2-phenylbenzofuran derivatives[121], which exhibited selective inhibitory property for BChE enzyme and with an inhibition IC_{50} value similar to that of galantamine (~30 μM). It was noted that the contemporary presence of a single hydroxyl group

in the para position of the phenyl ring and a halogen substitution at position 7 of the benzofuran scaffold improved the inhibitory activity towards BChE[121].

In this work, we elucidate the importance of hydroxyl group substitution in the phenyl-ring in the new series of 2-phenylbenzofuran derivatives (Table 9) with a contemporary presence of either chlorine or bromine at position 7 (R^1) of the benzofuran scaffold. MD simulations were employed to identify key structural and dynamical aspects that influence the inhibitory activity of the potent compounds against hBChE enzyme.

Table 9. Inhibition of EeAChE and eqBChE enzymes by Compounds **15-28**.

Compound	R	R^1	IC$_{50}$ (µM)* EeAChE	eqBChE	Selectivity to eqBChE**
15	H	Cl	>100	6.23 ± 0.43	>16.0
16	Cl	H	80 ± 7.3	36.6 ± 2.90	2.2
17	H	Br	100 ± 6.1	3.57 ± 0.25	28.0
18	Br	H	66 ± 4.2	30 ± 2.70	2.2
19	H	CH$_3$	>100	10.03 ± 0.96	>10.0
20	CH$_3$	H	>100	12.51 ± 1.29	>8.0
21	H	H	>100	25.18 ± 1.30	>4.0
22	H	Cl	50 ± 3.3	25.7 ± 1.60	1.9
23	Cl	H	30 ± 2.8	38.2 ± 2.40	0.78
24	H	Br	37 ± 2.6	18.41 ± 0.93	2.0
25	Br	H	25 ± 1.9	27.6 ± 1.90	0.90

26	H	CH$_3$	66 ± 5.4	19.8 ± 1.20	3.3
27	CH$_3$	H	>100	16.05 ± 1.05	>6.2
28	H	H	>100	>100	>1.0
Galantamine			0.95 ± 0.02	28.3 ± 2.1	0.033

*EeAChE and eqBChE inhibition is expressed as the mean ± SD (n = 3 experiments). **Selectivity to BChE: IC$_{50}$ for AChE)/IC$_{50}$ for BChE

5.3 Results

Inhibitory activity of 2-phenylbenzofuran derivatives against AChE and BChE. To investigate the importance of hydroxyl substituents in the synthesized 2-phenylbenzofuran derivatives, the inhibitory effect of these compounds were (**15-28**) on EeAChE and eqBChE activity by determining their inhibitory potency IC$_{50}$ which is concentration of inhibitor needed to reduce the enzyme activity by half. For the initial screening of the compounds, we used enzymes of non-human origin namely EeAChE and eqBChE due to their lower cost and high degree of similarity with their respective human enzymes.

The inhibition results of the compounds against the two enzymes are summarized in Table 9. We noted that compound **28**, with three hydroxyl substituents in phenyl-ring and hydrogen atom in position 5 (**R**) and 7 (**R^1**) of benzofuran scaffold did not exert any cholinesterase inhibitory activity. In general, except compounds **23** and **25**, all other compounds displayed better activity against eqBChE enzyme. In detail, only six compounds (**15**, **17**, **19-21** and **27**) displayed inhibitory activity against eqBChE and with IC$_{50}$ values for EeAChE being equal or greater to 100 μM. While, on the other hand the remaining compounds inhibited both the enzymes with varying efficiency. Among these derivatives, maximum inhibitory activity against eqBChE enzyme were displayed by compound **15** (IC$_{50}$ = 6.23 μM) and **17** (IC$_{50}$ = 3.57 μM), with two hydroxyl substituents in phenyl-ring and with presence of chlorine and bromine atoms respectively at position 7 (R^1) of benzofuran scaffold. Interestingly, eqBChE inhibitory activity displayed by compounds **15** and **17** was about 4- and 8- times

more active than the reference compound, galantamine (IC_{50} = 28.3 µM). We therefore focused our attention on the compounds **15** and **17**, which exhibited maximum inhibitory action against eqBChE enzyme. Kinetic analysis of steady state inhibition data revealed that compound **15** acts as a mixed-type inhibitor and compound **17** as a non-competitive inhibitor. The inhibition constants for the compound **17** were determined to be 4.3 µM and 4.7 µM.

The inhibitory activity of the most potent inhibitors (compounds **15**, **17**) was further investigated on hBChE enzyme; the results are presented in Table 10. We note that both these compounds inhibit hBChE enzyme with IC_{50} values in the µM range and display similar IC_{50} values.

Table 10. Inhibition of hBChE by Compounds **15** and **17**.

Compound	IC_{50} (µM)*
15	27.51 ± 1.82
17	27.46 ± 1.53
Galantamine	56.8 ± 4.11

*hBChE inhibition is expressed as the mean ± SD (n = 3 experiments)
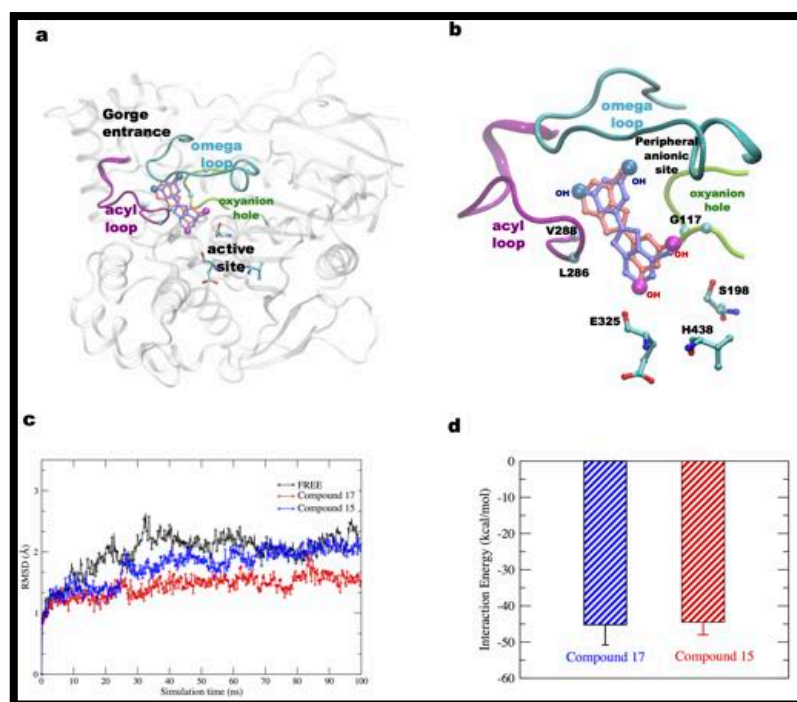
Figure 27. Molecular Modelling. (a) Superimposition of best-docked positions of compounds 17 (blue) and 15 (red) into binding site of hBChE protein. The protein is represented in cartoon representation, the active site residues in licorice, and loops leading to hBChE active site are shown. (b) Zoomed representation of hBChE interaction site for the two compounds, and key residues are shown. (c) RMSD plots for the free and compound-bound hBChE simulations. (d) Interaction energy plots between the compound and hBChE residues.

Molecular modelling studies. To predict how the compounds **15** and **17** bind to hBChE and to understand the molecular origin of their high inhibitory activity and selectivity, we performed molecular docking experiments. Docking results suggested similar interaction sites (Fig. 27a, 27b) and similar binding energy values (~7.5 kcal/mol), for the two compounds. The stability of the docking poses of the two compounds was investigated using MD simulations, which is a standard technique used to study the dynamical properties of biomolecules[57,71,114]. The stability of the systems during the MD simulations was evaluated by calculating the root mean square deviation (RMSD) of C-alpha atoms of protein residues (Fig. 27c) from the starting structure. The average RMSD values of protein bound compound simulations were lower than in free protein simulations,

59

with lowest value noted for compound **17** complex simulations. Subsequently, the interaction energy between the hBChE residues and the two compounds was calculated by evaluating the non-bonded energy values comprising of Van der Waals and electrostatic energy in the two simulations. Both the complexes exhibited similar interaction energy values (Fig. 27d).
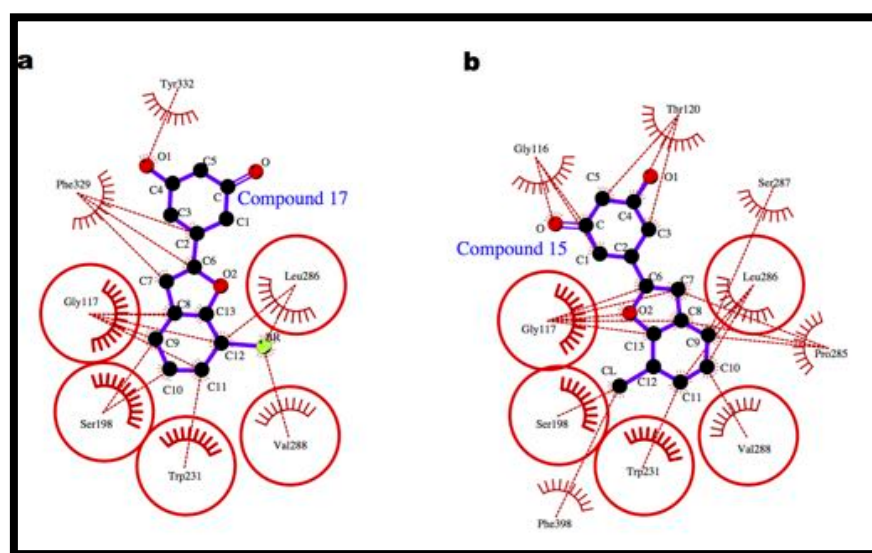
Figure 28. Molecular interaction picture of hBChE protein bound to (a) compound **17** and (b) compound **15**. The conserved interactions between the two complexes are represented as red circles.

To understand the origin of this similarity, we carefully inspected the binding mode of the compounds in complex with hBChE using Ligplot[140]. The compounds (**15** and **17**) were stably bound to hBChE active site (Fig. 28) encompassing the region between peripheral anionic site (PAS) and the catalytic triad site (CAS). Figure 28 depicts five overlapping hBChE residues interacting with the two compounds. In detail, these residues are located in catalytic triad (S198), oxyanion hole (G117), acyl-pocket (L286, V288) and wall of BChE active site. The hydroxyl substituents in compound **17** interact with peripheral anionic site residue (Y332), while compound **15** interacts with oxyanion hole residue (G116) and residue T120.
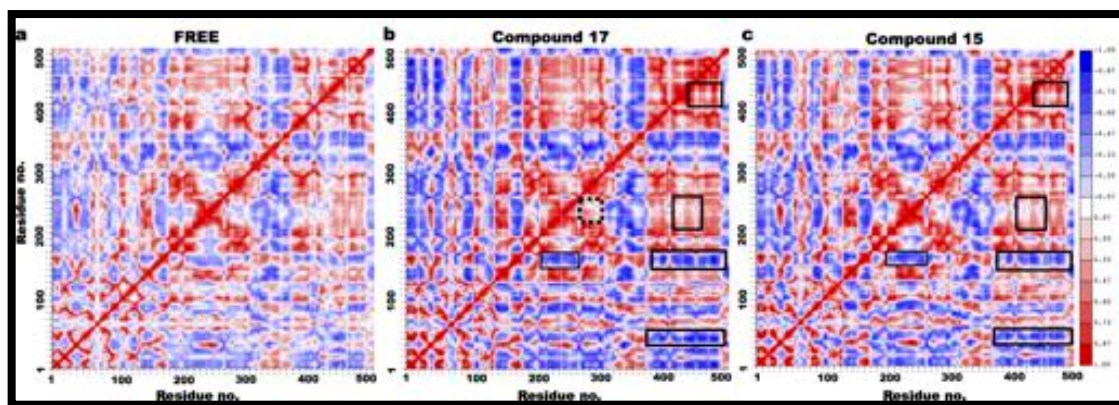
Figure 29. Dynamical cross-correlation map for C-alpha atoms. (a) Free hBChE, (b) compound **17** and (c) compound **15** complexes. Positive correlations are indicated in red and negative or anti-correlations in blue, while no correlation in white. In (b) and (c) boxed regions represent those regions different with respect to free hBChE protein. While in (b), the dashed box for compound **17** represents the region different with respect to both the free and compound **15** complexes.

To examine the effects of compound **15** and **17** on the protein structural dynamics, comparative analysis of a series of snapshots of the protein coordinates from MD simulations trajectories between the complex (bound to the compounds) and free protein was done. Calculation of all inter-residue cross-correlations fluctuations (see Methods) of C-alpha atoms resulted in a matrix of cross-correlation coefficient (C_{ij}) elements, which are displayed in a graphical representation as a dynamical cross-correlation map, shown in Fig. 29.

As expected, we note strong fluctuations occur along the diagonal occur (between the same residue), wherein C_{ij} is always equal to 1. A clear difference in the cross-correlations maps between the free and complex simulations was observed (Fig. 29). With respect to free protein simulations (Fig. 29a), we observed between few domains, an increase in either a positive or a negative correlation dynamics for the complex simulations (Fig. 29b, 29c). In detail, the regions involved in higher negative correlated dynamics included residues 40-60, 170-190 and 380-500, while residues 230-280 displayed lower negative correlated dynamics. On the other hand, residues 430-470 exhibited higher positive correlated dynamics in the compound complexes.

As expected, most of these regions are in close vicinity to the hBChE active site gorge. Interestingly, only for compound **17** complex (Fig. 29b), positive correlation dynamics was noted between the domains surrounding the BChE active site gorge, i.e. residues 240-280 and 300-330, respectively.

5.4 Discussion

There is increasing clinical evidence suggesting an important role of BChE in the regulation of ACh levels and in particular in the development and progression of AD. Particularly, in progressed or late stage of AD, BChE mostly dominates hydrolysis of ACh[129]. Moreover, alongside its involvement in AD progression, an emerging role of BChE as a prognostic marker (which determines the progress of the disease) in liver and non-liver diseases, as well as in protein-energy malnutrition and obesity, has been reported[164,181]. Design and development of compounds with the ability to selectively inhibit BChE would not only improve understanding of the aetiology of AD but also assist in developing wider variety of new treatments. Therefore, the objective of our study has been to design and develop 2-phenylbenzofuran compounds that display selective BChE inhibitory activity employing biochemical, kinetics and computational techniques.

In our recent study[121], we reported that the contemporary presence of a hydroxyl group in the para position of the 2-phenyl ring and a halogen substitution at position 7 (R^1) of the benzofuran scaffold resulted in a good and selective BChE inhibition, with best inhibitor displaying an IC$_{50}$ of 30 µM. Following the results of our previous findings, in this present work we decided to explore the importance of the number and position of hydroxyl groups located in the 2-phenyl ring of the benzofuran moiety. We therefore synthesized new 2-phenylbenzofurans compounds with two hydroxyl substituents (compounds **15-21**) and with three hydroxyl substituents (compounds **22-28**). Galantamine was used as our reference compound. The inhibitory action of the newly synthesized compounds presented in Table 1 demonstrate that, regardless the type of substituent at position 7 of benzofuran scaffold, the 2-phenylbenzofuran derivatives with two hydroxyl substituents (compounds **15-21**) in meta position of the 2-phenyl ring displayed rather high inhibitory activity toward eqBChE and very low

activity against EeAChE. In particular, compounds **15** and **17** displayed eqBChE inhibitory activity 4- and 8- times more effective than the reference compound, respectively. However, in the compounds with three hydroxyl substituents (instead of two) in the 2-phenyl ring (compounds **22, 24**), we found lower inhibitory activity against eqBChE. This fact suggest that contemporary presence of three hydroxyl groups in the 2-phenyl ring of the compounds could decrease the inhibitory activity of the compounds against eqBChE. It has been shown previously[182] that the position and number of hydroxyl group in the ligand can influence the magnitude of hydrogen bond interactions with the protein. The BChE active site is located at the bottom of a 20 Å deep gorge that is lined mostly with hydrophobic residues. Thus, binding of an additional hydroxyl substituent (a polar group) within the gorge could result in a thermodynamic penalty of additional 4.3−5.3 kcal/mol[183], due to energetic cost of desolvation. Hence, this could be one possible hypothesis to explain the low BChE inhibitory activity detected for the compounds with three hydroxyl groups in the 2-phenyl ring.

The two most active compounds (**15, 17**) differ in halogen atom at position 7 of the benzofuran moiety (chlorine, bromine atoms), respectively. It is interesting to note that this little difference is reflected in the protein interaction network characterizing these compounds (Fig. 28). The chlorine atom in compound **15** interacts with the CAS residue (S198) and F398, while bromine atom in compound **17** interacts with the acyl pocket residues (L286, V288). The ChE inhibition can occur either via a competitive interaction with CAS, or a non-competitive binding with PAS, or via mixed-type mechanisms, by exerting a dual binding ChE inhibition[133]. Enzyme kinetic analysis demonstrated only compound **15** as mixed-type inhibitor, while compound **17** as non-competitive inhibitor of eqBChE activity. The results from kinetic experiments are confirmed from MD simulations, which provide molecular-level insights into how ligand binding at an allosteric site can affect protein structure and, consequently, enzymatic activity. Indeed, the difference observed in the nature of correlated dynamics between the domains of residues surrounding BChE active site gorge provide dynamical information about the protein structure, which could explain the different BChE inhibition mechanisms between the compounds.

Previous clinical studies evidence that oxidative stress is a crucial factor in AD and plays an important role in inducing and activating multiple cell signalling pathways, contributing to the development of AD[184,185]. Indeed, development of new avenues to

reduce oxidative damages can provide therapeutic efficacy in the treatment of AD[186]. We therefore investigated the antioxidant properties of the new synthesized compounds. Comparing the results with the antioxidant property of benzofurans derivatives analyzed in our previous study[121], compounds **15** and **17** showed a higher antioxidant activity. Thus, substitution and positioning the groups within the 2-phenyl ring of the compounds, led to an improvement in terms of both BChE inhibitory activity and antioxidant property.

5.5 Conclusions

In this study, a series of hydroxylated 2-phenylbenzofurans compounds were designed, synthesized and their selective inhibitory activity BChE was evaluated. Combining biochemical analysis and computational approaches, we identified two potent BChE inhibitors as compound **17** (IC_{50}=3.5 μM) and compound **15** (IC_{50}=6.25 μM), with the presence of two hydroxyl substituents in meta position of the 2-phenyl ring and bromine or chlorine at position 7 of benzofuran moiety. The BChE selective inhibition property decreased with the introduction of a third hydroxyl group in the 2-phenyl ring of the compounds. Detailed kinetic experiments revealed compound **15** as a mixed-type inhibitor, while **17** as non-competitive inhibitor of BChE activity. Experimental results were confirmed by MD simulations, which revealed a conserved interaction pattern resulting in similar interaction energy values. Finally, compounds **15**, and **17** examined on hBChE revealed 2-times more active inhibitory action than the reference compound. In conclusion, gathering the information obtained in this study, compounds **15** and **17** could be considered as promising candidates for the design and development of drugs against AD.

5.6 Methods

Molecular Modeling. High-resolution three-dimensional protein structure of hBChE was obtained from protein data bank (PDB id: 4TPK). For the compounds (**15** and **17**), the three-dimensional coordinates were generated using Open Babel software[187].

The geometry of the compounds were optimized using the Hartree-Fock basis set 6-31G* within Gaussian03 software package[148]. The charges and the force field parameters of the compounds were evaluated following the standard protocol within AMBER software tools[149,188].

Molecular docking of the compounds into hBChE protein was performed using SwissDock web server, which is based on the docking software EADock DSS[150]. The docking poses of the compounds were accurately chosen with a blind docking procedure that considers the entire protein surface as a potential target. Using this procedure, a large number of ligand binding modes (~15000) were generated, with the simultaneous rough interaction energy estimation. The binding modes possessing favorable energies were then ranked and classified into different clusters, this time based on the full fitness scoring function. The most consistent and favorable conformation chosen from 10 independent docking runs for each compound was further considered for MD simulations.

The hBChE-compound complexes were built using leap module of Amber11. Each complex was inserted separately in an explicit water-box with a minimum distance of 1.8 nm between the solute and box boundary. We used amber force-field parameters[189] for hBChE protein and TIP3P[112] parameters for water molecules. The simulation box and number of atoms for the complexes investigated are presented in Table 11. Energy minimization, followed by heating of the complexes to temperature 300 K, was done with positional restraints on C-alpha atoms. The positional restraints were gradually removed during the simulation time and an equilibration run of 10 ns was performed. The time step used in MD simulation was of 2 fs using SHAKE algorithm. Simulations were performed in NPT ensemble using periodic boundary conditions. All-atom MD simulations of free protein and protein-compound complexes were performed for a simulation time of 100 ns employing NAMD[116] software package.

Table 11. Details of total number of atoms and starting simulation box size [x,y,z] in the hBChE protein with and with compound complexes.

hBChE	System Details	
	No. of Atoms	Box Size (Å)
Free complex	120545	[110 122 108]
Compound **15 complex**	119827	[108 120 105]
Compound **17 complex**	119854	[108 120 106]

The stability of systems was evaluated by calculating the RMSD values for the C-alpha atoms of residues during MD simulations, using VMD[111]. The interaction energy between the compound and protein residues was calculated by evaluating the non-bonded energy values comprising of Van der Waals and electrostatic energy, using the energy plugin of NAMD software. A cut-off distance of 12 Å was used for non-bonded interactions and for the electrostatic interaction we also adopted the particle mesh Ewald[118] scheme. The dynamic cross-correlation[190] coefficients for C-alpha atoms was calculated on 1000 snapshots extracted from 100 ns MD trajectories using Prody[139] software. The matrix of all inter-atomic cross-correlations of atomic fluctuations C_{ij} where i and j are C-alpha atoms, can be represented as a dynamical cross-correlation map. If the fluctuations of two C-alpha atoms are completely correlated then $C_{ij} = 1$ (red), if anticorrelated then $C_{ij} = -1$ (blue), and if $C_{ij} = 0$ (white) then the fluctuations of i and j are not correlated.

6. Conclusions and Future Perspective

The main goal of my research activity is to understand the molecular recognition mechanism of biologically relevant protein-protein and protein-ligand complexes associated to human diseases. For my research activity, I adopted classical molecular dynamics simulations for the investigation of bio-molecular complexes.

With the recent advances in technology and techniques used in experiments, more and more PPI data have become available, while in parallel, computational methods emerge to validate and complete the missing interactions. In this thesis, my aim is to introduce the importance of protein−protein interactions, protein-ligand interactions and to provide a broad and informative methodology for predicting such interactions, and furthermore present tools to analyse the data.

Computational approaches as every method in Science have their own advantages and limitations. Indeed, the outcome of the thesis supports the importance to integrate experimental and computational techniques. In this light it is encouraging to learn that the scientific community is already pursuing strategies to integrate and standardize differently annotated data in an organized way defining a common data format to exchange PPI information[8]. In detail, three case studies are presented.

6.1 Application of Molecular Modeling to biological complexes associated to MS disease

Three-dimensional X-ray structures of the protein-peptide-protein complexes provided a good starting point to model and investigate dynamical aspects of cross-

reactivity between myelin based protein (MBP) self-peptide and two microbial peptides (UL15, PMM) for Hy.1B11 T-cell receptor (TCR). Employing MD simulations, estimates of binding energy that were consistent with the trend noted in experiment is provided in our work. In addition, we identified a set of molecular interactions involving "hot spot" residues in the interface region, possibly responsible for TCR cross-reactivity.

6.2 Multidisciplinary approach to investigate protein-ligand complexes related to Alzheimer's disease.

A series of 2-pheynlbenzofurans compounds were synthesized and experimental data of their inhibitory activity against butylcholinesterases (BChE) and acetylcholinesterases (AChE) protein were provided by our experimental partners. These compounds showed different degree of selectivity against BChE protein. Employing, homology modeling of the 3D protein structures, molecular docking and MD simulations for the protein-ligand complexes, computational analysis revealed the molecular mechanism of BChE selectivity displayed by the compounds. The interaction of ligand with both the catalytic anionic site (CAS) and peripheral anionic site (PAS) was essential to selectively inhibit BChE enzyme. The importance of hydroxyl substituent in the 2-pheynl ring emerged for our multidisciplinary approach.

6.3 Design of ligands with improved BChE inhibitory activity.

Gathering information from our multidisciplinary approach, new series of 2-pheynlbenzofurans compounds were designed and synthesized with two and three hydroxyl substituents in the 2-pheynl ring. Experimental data predicted compounds with two hydroxyl substituents to displayed higher BChE inhibitory activity even with respect to our reference compound galanthamine, which was confirmed using computational modeling techniques.

As a whole in my thesis, a sincere effort has been made to demonstrate the role and application of computational approaches to provide deep understanding of macromolecular biological complexes, thus facilitating discovery, design, and development of potential drugs.

6.4 Perspectives and work in progress.

The simulation of protein–protein association is challenged by the fact that dissociation times is not directly accessible to standard MD simulation lengths by far, as long MD simulations can get trapped after a single binding minima, which is not necessarily the true minima of the complex and required simulation timescales are not yet easily reachable. Future direction of my work will be to use hybrid computational approaches such as the combination of an extensive and adaptively generated ensemble of unbiased molecular dynamics simulations with hidden Markov modeling to protein–protein association kinetics.

The ultimate goal is to put all these findings and data in the cellular environment, consider the interactions stability, affinity, and dynamics to gain further insight into cellular mechanisms. In the near future, I believe a well-defined proteome-scale map of protein interactions will be obtained by the integrative approaches, thus helping our understanding of the human "interactome".

7. References

1   Whitford, D. Proteins: Structure and Function. **Wiley** (2005).

2   Misra, G. Introduction to Biomolecular Structure and Biophysics. **Springer** (2017).

3   Ouzounis, C. A., Coulson, R. M. R., Enright, A. J., Kunin, V. & Pereira-Leal, J. B. Classification schemes for protein structure and function. *Nature Reviews Genetics* **4**, 508-519, doi:10.1038/nrg1113 (2003).

4   Slabinski, L. *et al.* The challenge of protein structure determination-lessons from structural genomics. *Protein Sci.* **16**, 2472-2482, doi:10.1110/ps.073037907 (2007).

5   Chruszcz, M., Borek, D., Domagalski, M., Otwinowski, Z. & Minor, W. X-Ray Diffraction Experiment. **77**, 23-40, doi:10.1016/s1876-1623(09)77002-6 (2009).

6   Opella, S. J. Structure Determination of Membrane Proteins by Nuclear Magnetic Resonance Spectroscopy. *Annual Review of Analytical Chemistry* **6**, 305-328, doi:10.1146/annurev-anchem-062012-092631 (2013).

7   Swann, M. J., Peel, L. L., Carrington, S. & Freeman, N. J. Dual-polarization interferometry: an analytical technique to measure changes in protein structure in real time, to determine the stoichiometry of binding events, and to differentiate between specific and nonspecific interactions. *Anal. Biochem.* **329**, 190-198, doi:10.1016/j.ab.2004.02.019 (2004).

8   Keskin, O., Tuncbag, N. & Gursoy, A. Predicting Protein–Protein Interactions from the Molecular to the Proteome Level. *Chem. Rev.* **116**, 4884-4909, doi:10.1021/acs.chemrev.5b00683 (2016).

9   Scott, D. E., Bayly, A. R., Abell, C. & Skidmore, J. Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. *Nature Reviews Drug Discovery* **15**, 533-550, doi:10.1038/nrd.2016.29 (2016).

10   Forman, M. S., Trojanowski, J. Q. & Lee, V. M. Y. Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs. *Nat. Med.* **10**, 1055-1063, doi:10.1038/nm1113 (2004).

11   Kuzmanov, U. & Emili, A. Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome Med.* **5**, 37, doi:10.1186/gm441 (2013).

12   Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13-20 (1996).

13   Nooren, I. M. & Thornton, J. M. Diversity of protein-protein interactions. *EMBO J.* **22**, 3486-3492, doi:10.1093/emboj/cdg359 (2003).

14   La, D., Kong, M., Hoffman, W., Choi, Y. I. & Kihara, D. Predicting permanent and transient protein-protein interfaces. *Proteins: Structure, Function, and Bioinformatics* **81**, 805-818, doi:10.1002/prot.24235 (2013).

15   Mintseris, J. & Weng, Z. Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10930-10935, doi:10.1073/pnas.0502667102 (2005).

16   Prasad Bahadur, R., Chakrabarti, P., Rodier, F. & Janin, J. A Dissection of Specific and Non-specific Protein–Protein Interfaces. *J. Mol. Biol.* **336**, 943-955, doi:10.1016/j.jmb.2003.12.073 (2004).

17   Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1-9, doi:10.1006/jmbi.1998.1843 (1998).

18   Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383-386 (1995).

19   Keskin, O., Ma, B. & Nussinov, R. Hot Regions in Protein–Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues. *J. Mol. Biol.* **345**, 1281-1294, doi:10.1016/j.jmb.2004.10.077 (2005).

20   Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* **450**, 1001-1009, doi:10.1038/nature06526 (2007).

21   Marks, D. S., Hopf, T. A. & Sander, C. Protein structure prediction from sequence variation. *Nat. Biotechnol.* **30**, 1072-1080, doi:10.1038/nbt.2419 (2012).

22   Pagadala, N. S., Syed, K. & Tuszynski, J. Software for molecular docking: a review. *Biophysical Reviews* **9**, 91-102, doi:10.1007/s12551-016-0247-1 (2017).

23   Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646-652, doi:10.1038/nsb0902-646 (2002).

24   Yura, K. Preface of Special Issue "Protein-Ligand Interactions". *Biophysics and Physicobiology* **13**, 85-86, doi:10.2142/biophysico.13.0_85 (2016).

25   Huang, H.-J., Lee, C.-C. & Chen, C. Y.-C. Lead Discovery for Alzheimer's Disease Related Target Protein RbAp48 from Traditional Chinese Medicine. *BioMed Research International* **2014**, 1-14, doi:10.1155/2014/764946 (2014).

26   Liao, K. H. *et al.* Ligand-Based and Structure-Based Investigation for Alzheimer's Disease from Traditional Chinese Medicine. *Evid. Based Complement. Alternat. Med.* **2014**, 1-16, doi:10.1155/2014/364819 (2014).

27   Du, X. *et al.* Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *International Journal of Molecular Sciences* **17**, 144, doi:10.3390/ijms17020144 (2016).

28   Perozzo, R., Folkers, G. & Scapozza, L. Thermodynamics of Protein–Ligand Interactions: History, Presence, and Future Aspects. *J. Recept. Signal Transduct.* **24**, 1-52, doi:10.1081/rrs-120037896 (2009).

29   Alberty, R. A. Standard Gibbs free energy, enthalpy, and entropy changes as a function of pH and pMg for several reactions involving adenosine phosphates. *J. Biol. Chem.* **244**, 3290-3302 (1969).

30   Ferrante, A. & Gorski, J. Enthalpy-entropy compensation and cooperativity as thermodynamic epiphenomena of structural flexibility in ligand-receptor interactions. *J. Mol. Biol.* **417**, 454-467, doi:10.1016/j.jmb.2012.01.057 (2012).

31   Hinz, H. J. Thermodynamics of Protein-Ligand Interactions: Calorimetric Approaches. *Annu. Rev. Biophys. Bioeng.* **12**, 285-317, doi:10.1146/annurev.bb.12.060183.001441 (1983).

32   Pang, X. & Zhou, H.-X. Rate Constants and Mechanisms of Protein–Ligand Binding. *Annual Review of Biophysics* **46**, 105-130, doi:10.1146/annurev-biophys-070816-033639 (2017).

33   Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **6**, 365-387, doi:10.1038/nprot.2011.305 (2011).

34   Ladbury, J. E. & Chowdhry, B. Z. Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem. Biol.* **3**, 791-801, doi:10.1016/s1074-5521(96)90063-0 (1996).

35   Jonsson, U. *et al.* Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *BioTechniques* **11**, 620-627 (1991).

36   Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432**, 862-865, doi:10.1038/nature03197 (2004).

37   Irwin, J. J. & Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177-182, doi:10.1021/ci049714+ (2005).

38   Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980-980, doi:10.1038/nsb1203-980 (2003).

39   UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158-D169, doi:10.1093/nar/gkw1099 (2017).

40   Fiser, A. Template-Based Protein Structure Modeling. **673**, 73-94, doi:10.1007/978-1-60761-842-3_6 (2010).

41   Eswar, N. *et al.* Comparative Protein Structure Modeling Using Modeller. 5.6.1-5.6.30, doi:10.1002/0471250953.bi0506s15 (2006).

42   Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201, doi:10.1093/bioinformatics/bti770 (2006).

43   Goodsell, D. S., Morris, G. M. & Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **9**, 1-5, doi:10.1002/(SICI)1099-1352(199601)9:1<1::AID-JMR241>3.0.CO;2-6 (1996).

44   Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. *J. Mol. Biol.* **267**, 727-748, doi:10.1006/jmbi.1996.0897 (1997).

45   Friesner, R. A. *et al.* Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **49**, 6177-6196, doi:10.1021/jm051256o (2006).

46   Vakser, Ilya A. Protein-Protein Docking: From Interaction to Interactome. *Biophys. J.* **107**, 1785-1793, doi:10.1016/j.bpj.2014.08.033 (2014).

47   Leach, A. *Molecular Modelling: Principles and Applications.*, (Prentice Hall, 2001).

48   Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179-5197, doi:10.1021/ja00124a002 (1995).

49   Lopes, P. E. M., Guvench, O. & MacKerell, A. D. Current Status of Protein Force Fields for Molecular Dynamics Simulations. **1215**, 47-71, doi:10.1007/978-1-4939-1465-4_3 (2015).

50   Ponder, J. W. & Case, D. A. Force Fields for Protein Simulations. **66**, 27-85, doi:10.1016/s0065-3233(03)66002-x (2003).

51   Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656-1676, doi:10.1002/jcc.20090 (2004).

52   MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586-3616, doi:10.1021/jp973084f (1998).

53   Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for

crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657-1666, doi:10.1021/ja00214a001 (1988).

54   Ascherio, A. & Munger, K. L. Epidemiology of Multiple Sclerosis: From Risk Factors to Prevention-An Update. *Semin. Neurol.* **36**, 103-114, doi:10.1055/s-0036-1579693 (2016).

55   Zwibel, H. L. & Smrtka, J. Improving quality of life in multiple sclerosis: an unmet need. *Am. J. Manag. Care* **17 Suppl 5 Improving**, S139-145 (2011).

56   Sethi, D. K., Gordo, S., Schubert, D. A. & Wucherpfennig, K. W. Crossreactivity of a human autoimmune TCR is dominated by a single TCR loop. *Nat Commun* **4**, 2623, doi:10.1038/ncomms3623 (2013).

57   Kumar, A. & Delogu, F. Dynamical footprint of cross-reactivity in a human autoimmune T-cell receptor. *Sci. Rep.* **7**, 42496, doi:10.1038/srep42496 (2017).

58   Steinman, L. *Nat. Immunol.* **2**, 762-764, doi:10.1038/ni0901-762 (2001).

59   Dendrou, C. A., Fugger, L. & Friese, M. A. Immunopathology of multiple sclerosis. *Nat. Rev. Immunol.* **15**, 545-558, doi:10.1038/nri3871 (2015).

60   Arnaiz-Villena, A. *et al.* The origin of Palestinians and their genetic relatedness with other Mediterranean populations. *Hum. Immunol.* **62**, 889-900 (2001).

61   McDonald, W. I. The dynamics of multiple sclerosis. *J. Neurol.* **240**, 28-36, doi:10.1007/bf00838443 (1993).

62   Steinman, M. D. L. Multiple Sclerosis: A Coordinated Immunological Attack against Myelin in the Central Nervous System. *Cell* **85**, 299-302, doi:10.1016/s0092-8674(00)81107-1 (1996).

63   Thompson, A. J. Symptomatic management and rehabilitation in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* **71 Suppl 2**, ii22-27 (2001).

64   Steinman, L., Martin, R., Bernard, C., Conlon, P. & Oksenberg, J. R. Multiple sclerosis: deeper understanding of its pathogenesis reveals new targets for therapy. *Annu. Rev. Neurosci.* **25**, 491-505, doi:10.1146/annurev.neuro.25.112701.142913 (2002).

65   International Multiple Sclerosis Genetics, C. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214-219, doi:10.1038/nature10251 (2011).

66   Fugger, L., Friese, M. A. & Bell, J. I. From genes to function: the next challenge to understanding multiple sclerosis. *Nat. Rev. Immunol.* **9**, 408-417, doi:10.1038/nri2554 (2009).

67   Kakalacheva, K. & Lunemann, J. D. Environmental triggers of multiple sclerosis. *FEBS Lett.* **585**, 3724-3729, doi:10.1016/j.febslet2011.04.006 (2011).

68   Ascherio, A., Munger, K. L. & Lunemann, J. D. The initiation and prevention of multiple sclerosis. *Nat. Rev. Neurol.* **8**, 602-612, doi:10.1038/nrneurol.2012.198 (2012).

69   Martin, R., McFarland, H. F. & McFarlin, D. E. Immunological aspects of demyelinating diseases. *Annu. Rev. Immunol.* **10**, 153-187, doi:10.1146/annurev.iy.10.040192.001101 (1992).

70   Cocco, E. *et al.* Interaction between HLA-DRB1-DQB1 Haplotypes in Sardinian Multiple Sclerosis Population. *PLoS One* **8**, e59790, doi:10.1371/journal.pone.0059790 (2013).

71   Kumar, A. *et al.* Dynamical insights into the differential characteristics of Mycobacterium avium subsp. paratuberculosis peptide binding to HLA-DRB1 proteins associated with multiple sclerosis. *New. J. Chem.* **39**, 1355-1366, doi:10.1039/c4nj01903b (2015).

72   Markovic-Plese, S., Pinilla, C. & Martin, R. The initiation of the autoimmune response in multiple sclerosis. *Clin. Neurol. Neurosurg.* **106**, 218-222, doi:10.1016/j.clineuro.2004.02.018 (2004).

73   Chaplin, D. D. Overview of the immune response. *J. Allergy Clin. Immunol.* **125**, S3-S23, doi:10.1016/j.jaci.2009.12.980 (2010).

74   Miles, J. J., McCluskey, J., Rossjohn, J. & Gras, S. Understanding the complexity and malleability of T-cell recognition. *Immunol. Cell Biol.* **93**, 433-441, doi:10.1038/icb.2014.112 (2015).

75   Rossjohn, J. *et al.* T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.* **33**, 169-200, doi:10.1146/annurev-immunol-032414-112334 (2015).

76   Rothbard, J. B. & Gefter, M. L. Interactions between immunogenic peptides and MHC proteins. *Annu. Rev. Immunol.* **9**, 527-565, doi:10.1146/annurev.iy.09.040191.002523 (1991).

77   Sprent, J., Zhang, X., Sun, S. & Tough, D. T-cell proliferation in vivo and the role of cytokines. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**, 317-322, doi:10.1098/rstb.2000.0568 (2000).

78   Regner, M. Cross-reactivity in T-cell antigen recognition. *Immunol. Cell Biol.* **79**, 91-100, doi:10.1046/j.1440-1711.2001.00994.x (2001).

79   Colf, L. A. *et al.* How a single T cell receptor recognizes both self and foreign MHC. *Cell* **129**, 135-146, doi:10.1016/j.cell.2007.01.048 (2007).

80   Birnbaum, M. E. *et al.* Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073-1087, doi:10.1016/j.cell.2014.03.047 (2014).

81   Wooldridge, L. *et al.* A single autoimmune T cell receptor recognizes more than a million different peptides. *J. Biol. Chem.* **287**, 1168-1177, doi:10.1074/jbc.M111.289488 (2012).

82   Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. How Tcrs Bind Mhcs, Peptides, and Coreceptors. *Ann Rev Immunol* **24**, 419-466, doi:10.1146/annurev.immunol.23.021704.115658 (2006).

83   Macdonald, W. A. *et al.* T cell allorecognition via molecular mimicry. *Immunity* **31**, 897-908, doi:10.1016/j.immuni.2009.09.025 (2009).

84   Borbulevych, O. Y. *et al.* T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility. *Immunity* **31**, 885-896, doi:10.1016/j.immuni.2009.11.003 (2009).

85   Adams, J. J. *et al.* Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nat. Immunol.* **17**, 87-94, doi:10.1038/ni.3310 (2016).

86   Holler, P. D. & Kranz, D. M. Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. *Immunity* **18**, 255-264 (2003).

87   McMahan, R. H. *et al.* Relating TCR-peptide-MHC affinity to immunogenicity for the design of tumor vaccines. *J. Clin. Invest.* **116**, 2543-2551, doi:10.1172/JCI26936 (2006).

88   Tian, S., Maile, R., Collins, E. J. & Frelinger, J. A. CD8+ T cell activation is governed by TCR-peptide/MHC affinity, not dissociation rate. *J. Immunol.* **179**, 2952-2960 (2007).

89   Stone, J. D., Chervin, A. S. & Kranz, D. M. T-cell receptor binding affinities and kinetics: impact on T-cell activity and specificity. *Immunology* **126**, 165-176, doi:10.1111/j.1365-2567.2008.03015.x (2009).

90   Yin, Y. & Mariuzza, R. A. The Multiple Mechanisms of T Cell Receptor Cross-reactivity. *Immunity* **31**, 849-851, doi:10.1016/j.immuni.2009.12.002 (2009).

91   Mazza, C. *et al.* How much can a T-cell antigen receptor adapt to structurally distinct antigenic peptides? *EMBO J.* **26**, 1972-1983, doi:10.1038/sj.emboj.7601605 (2007).

92   Li, Y. *et al.* Structure of a human autoimmune TCR bound to a myelin basic protein self-peptide and a multiple sclerosis-associated MHC class II molecule. *EMBO J.* **24**, 2968-2979, doi:10.1038/sj.emboj.7600771 (2005).

93   Harkiolaki, M. *et al.* T cell-mediated autoimmune disease due to low-affinity crossreactivity to common microbial peptides. *Immunity* **30**, 348-357, doi:10.1016/j.immuni.2009.01.009 (2009).

94   Sethi, D. K. *et al.* A highly tilted binding mode by a self-reactive T cell receptor results in altered engagement of peptide and MHC. *J. Exp. Med.* **208**, 91-102, doi:10.1084/jem.20100725 (2011).

95   Wucherpfennig, K. W. & Strominger, J. L. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell* **80**, 695-705, doi:0092-8674(95)90348-8 [pii] (1995).

96   Knapp, B., Dunbar, J. & Deane, C. M. Large scale characterization of the LC13 TCR and HLA-B8 structural landscape in reaction to 172 altered peptide ligands: a molecular dynamics simulation study. *PLoS Comput. Biol.* **10**, e1003748, doi:10.1371/journal.pcbi.1003748 (2014).

97   Smith, S. N. *et al.* Changing the peptide specificity of a human T-cell receptor by directed evolution. *Nat Commun* **5**, 5223, doi:10.1038/ncomms6223 (2014).

98   Stavrakoudis, A. Insights into the structure of the LC13 TCR/HLA-B8-EBV peptide complex with molecular dynamics simulations. *Cell Biochem. Biophys.* **60**, 283-295, doi:10.1007/s12013-011-9151-2 (2011).

99   Reboul, C. F., Meyer, G. R., Porebski, B. T., Borg, N. A. & Buckle, A. M. Epitope flexibility and dynamic footprint revealed by molecular dynamics of a pMHC-TCR complex. *PLoS Comput. Biol.* **8**, e1002404 (2012).

100   Wolfson, M. Y., Nam, K. & Chakraborty, A. K. The effect of mutations on the alloreactive T cell receptor/peptide-MHC interface structure: a molecular dynamics study. *J. Phys. Chem. B* **115**, 8317-8327, doi:10.1021/jp202471d (2011).

101   Bello, M. & Correa-Basurto, J. Energetic and flexibility properties captured by long molecular dynamics simulations of a membrane-embedded pMHCII–TCR complex. *Mol. BioSyst.* **12**, 1350-1366, doi:10.1039/c6mb00058d (2016).

102   Koukos, P. I. & Glykos, N. M. On the application of Good-Turing statistics to quantify convergence of biomolecular simulations. *J. Chem. Inf. Model.* **54**, 209-217, doi:10.1021/ci4005817 (2014).

103   Altis, A., Nguyen, P. H., Hegger, R. & Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **126**, 244111, doi:10.1063/1.2746330 (2007).

104   Balaraju, T. *et al.* Aromatic interaction profile to understand the molecular basis of raltegravir resistance. *Struct. Chem.* **24**, 1499-1512, doi:10.1007/s11224-012-0181-1 (2013).

105   Andricioaei, I. & Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **115**, 6289-6292, doi:10.1063/1.1401821 (2001).

106   Yaneva, R., Schneeweiss, C., Zacharias, M. & Springer, S. Peptide binding to MHC class I and II proteins: new avenues from new methods. *Mol. Immunol.* **47**, 649-657, doi:10.1016/j.molimm.2009.10.008 (2010).

107   Marsh, J. A. Buried and accessible surface area control intrinsic protein flexibility. *J. Mol. Biol.* **425**, 3250-3263, doi:10.1016/j.jmb.2013.06.019 (2013).

108   Mason, D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today* **19**, 395-404 (1998).

109   Sewell, A. K. Why must T cells be cross-reactive? *Nat. Rev. Immunol.* **12**, 669-677, doi:10.1038/nri3279 (2012).

110   Hahn, M., Nicholson, M. J., Pyrdol, J. & Wucherpfennig, K. W. Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. *Nat. Immunol.* **6**, 490-496, doi:10.1038/ni1187 (2005).

111   Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38, 27-38, doi:0263785596000185 [pii] (1996).

112   Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935, doi:10.1063/1.445869 (1983).

113   Rostkowski, M., Olsson, M. H., Sondergaard, C. R. & Jensen, J. H. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct. Biol.* **11**, 1-6, doi:10.1186/1472-6807-11-6 (2011).

114   Kumar, A. *et al.* Antigenic peptide molecular recognition by the DRB1-DQB1 haplotype modulates multiple sclerosis susceptibility. *Mol. Biosyst.* **10**, 2043-2054, doi:10.1039/c4mb00203b (2014).

115   Kumar, A., Cocco, E., Atzori, L., Marrosu, M. G. & Pieroni, E. Structural and Dynamical Insights on HLA-DR2 Complexes That Confer Susceptibility to Multiple Sclerosis in Sardinia: A Molecular Dynamics Simulation Study. *PLoS One* **8**, e59711, doi:10.1371/journal.pone.0059711 (2013).

116   Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-1802, doi:10.1002/jcc.20289 (2005).

117   Tsoulos, I. G. & Stavrakoudis, A. Eucb: A C++ program for molecular dynamics trajectory analysis. *Comput. Phys. Commun.* **182**, 834-841, doi:10.1016/j.cpc.2010.11.032 (2011).

118   Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577, doi:10.1063/1.470117 (1995).

119   Glykos, N. M. Software news and updates. Carma: a molecular dynamics analysis program. *J. Comput. Chem.* **27**, 1765-1768, doi:10.1002/jcc.20482 (2006).

120   Alzheimer's Association. 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **13**, 325-373, doi:10.1016/j.jalz.2017.02.001 (2017).

121   Delogu, G. L. *et al.* 2-Phenylbenzofuran derivatives as butyrylcholinesterase inhibitors: Synthesis, biological activity and molecular modeling. *Biorg. Med. Chem. Lett.* **26**, 2308-2313, doi:10.1016/j.bmcl.2016.03.039 (2016).

122   Schuster, D. *et al.* Morphinans and isoquinolines: Acetylcholinesterase inhibition, pharmacophore modeling, and interaction with opioid receptors. *Bioorg. Med. Chem.* **18**, 5071-5080, doi:10.1016/j.bmc.2010.05.071 (2010).

123   Lane, R. M., Potkin, S. G. & Enz, A. Targeting acetylcholinesterase and butyrylcholinesterase in dementia. *The Int J Neuropsychopharmacol* **9**, 101, doi:10.1017/s1461145705005833 (2005).

124   Valle, A. M. *et al.* The cholinesterases: Analysis by pharmacogenomics in man. *Chem Biol Interac* **175**, 343-345, doi:10.1016/j.cbi.2008.04.042 (2008).

125   Allderdice, P. W. *et al.* The cloned butyrylcholinesterase (BCHE) gene maps to a single chromosome site, 3q26. *Genomics* **11**, 452-454, doi:10.1016/0888-7543(91)90154-7 (1991).

126   Getman, D. K., Eubanks, J. H., Camp, S., Evans, G. A. & Taylor, P. The human gene encoding acetylcholinesterase is located on the long arm of chromosome 7. *Am. J. Hum. Genet.* **51**, 170-177 (1992).

127   Sussman, J. L. *et al.* Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. *Science* **253**, 872-879 (1991).

128   Mesulam, M. Butyrylcholinesterase in the normal and Alzheimer brain. *Butyrylcholinesterase: Its Function and Inhibitors*, 29-37 (2003).

129   Greig, N. H. *et al.* A new therapeutic target in Alzheimer's disease treatment: attention to butyrylcholinesterase. *Curr. Med. Res. Opin.* **17**, 159-165, doi:10.1185/0300799039117057 (2001).

130   Jann, M. W., Shirley, K. L. & Small, G. W. Clinical pharmacokinetics and pharmacodynamics of cholinesterase inhibitors. *Clin. Pharmacokinet.* **41**, 719-739, doi:10.2165/00003088-200241100-00003 (2002).

131   Zemek, F. *et al.* Outcomes of Alzheimer's disease therapy with acetylcholinesterase inhibitors and memantine. *Expert Opin. Drug Saf.* **13**, 759-774, doi:10.1517/14740338.2014.914168 (2014).

132   Khan, M. A. *et al.* Regioselective synthesis of novel 2,3,4,4a-tetrahydro-1H-carbazoles and their cholinesterase inhibitory activities. *RSC Adv.* **5**, 59240-59250, doi:10.1039/c5ra10461k (2015).

133   Semenov, V. E. *et al.* 6-Methyluracil Derivatives as Bifunctional Acetylcholinesterase Inhibitors for the Treatment of Alzheimer's Disease. *Chem. Med. Chem.* **10**, 1863–1874, doi:10.1002/cmdc.201500334 (2015).

134   Jiang, Y. *et al.* Simple, Convenient, and Efficient Synthesis of 2-Aryl-substituted Benzo[b]furans. *Synth. Commun.* **39**, 197-204, doi:10.1080/00397910701860323 (2008).

135   Galal, S. A., Abd El-All, A. S., Abdallah, M. M. & El-Diwani, H. I. Synthesis of potent antitumor and antiviral benzofuran derivatives. *Bioorg. Med. Chem. Lett.* **19**, 2420-2428, doi:10.1016/j.bmcl.2009.03.069 (2009).

136   Erber, S., Ringshandl, R. & von Angerer, E. 2-Phenylbenzo[b]furans: relationship between structure, estrogen receptor affinity and cytostatic activity against mammary tumor cells. *Anticancer Drug Des.* **6**, 417-426 (1991).

137   Delogu, G. L. *et al.* Monoamine Oxidase (MAO) Inhibitory Activity: 3-Phenylcoumarins versus 4-Hydroxy-3-phenylcoumarins. *Chem Med Chem* **9**, 1672-1676, doi:10.1002/cmdc.201402010 (2014).

138   Baharloo, F. *et al.* Benzofuran-derived benzylpyridinium bromides as potent acetylcholinesterase inhibitors. *Euro. J. Med. Chem.* **93**, 196-201, doi:10.1016/j.ejmech.2015.02.009 (2015).

139   Bakan, A., Meireles, L. M. & Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **27**, 1575-1577, doi:10.1093/bioinformatics/btr168 (2011).

140   Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778-2786, doi:10.1021/ci200227u (2011).

141   Saxena, A. *et al.* Aromatic amino-acid residues at the active and peripheral anionic sites control the binding of E2020 (Aricept) to cholinesterases. *Eur. J. Biochem.* **270**, 4447-4458 (2003).

142   Bosak, A. *et al.* Peripheral site and acyl pocket define selective inhibition of mouse butyrylcholinesterase by two biscarbamates. *Arch. Biochem Biophys* **529**, 140-145, doi:10.1016/j.abb.2012.11.012 (2013).

143   Greenblatt, H. M. *et al.* The Complex of a Bivalent Derivative of Galanthamine withTorpedoAcetylcholinesterase Displays Drastic Deformation of the Active-Site Gorge: Implications for Structure-Based Drug Design. *J. Am. Chem. Soc.* **126**, 15405-15411, doi:10.1021/ja0466154 (2004).

144   Radic, Z. *et al.* Expression of recombinant acetylcholinesterase in a baculovirus system: kinetic properties of glutamate 199 mutants. *Biochemistry* **31**, 9760-9767 (1992).

145   Shafferman, A. *et al.* Mutagenesis of human acetylcholinesterase. Identification of residues involved in catalytic activity and in polypeptide folding. *J. Biol. Chem.* **267**, 17640-17648 (1992).

146   Brus, B. *et al.* Discovery, Biological Evaluation, and Crystal Structure of a Novel Nanomolar Selective Butyrylcholinesterase Inhibitor. *J. Med. Chem.* **57**, 8167-8179, doi:10.1021/jm501195e (2014).

147   Paz, A. *et al.* The Crystal Structure of a Complex of Acetylcholinesterase with a Bis-(−)-nor-meptazinol Derivative Reveals Disruption of the Catalytic Triad. *J. Med. Chem.* **52**, 2543-2549, doi:10.1021/jm801657v (2009).

148   Frisch, M. J. *et al.* Gaussian 03, Revision C.02.   **Gaussian 03**, Revision C.02 Wallingford CT (2004).

149   Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157-1174, doi:10.1002/jcc.20035 (2004).

150   Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **39**, W270-W277, doi:10.1093/nar/gkr366 (2011).

151   Wilson, R. S. *et al.* The natural history of cognitive decline in Alzheimer's disease. *Psychol. Aging* **27**, 1008-1017, doi:10.1037/a0029857 (2012).

152   Barker, W. W. *et al.* Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer Dis. Assoc. Disord.* **16**, 203-212 (2002).

153   Qiu, C., Kivipelto, M. & von Strauss, E. Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. *Dialogues Clin. Neurosci.* **11**, 111-128 (2009).

154   Kumar, A. & Dogra, S. Neuropathology and therapeutic management of Alzheimer's disease - An update. *Drugs Future* **33**, 433, doi:10.1358/dof.2008.033.05.1192677 (2008).

155   Perry, E. K. The cholinergic system in old age and Alzheimer's disease. *Age Ageing* **9**, 1-8 (1980).

156   Bartus, R. T., Dean, R. L., 3rd, Beer, B. & Lippa, A. S. The cholinergic hypothesis of geriatric memory dysfunction. *Science* **217**, 408-414 (1982).

157   Sinha, S. & Lieberburg, I. Cellular mechanisms of beta -amyloid production and secretion. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11049-11053, doi:10.1073/pnas.96.20.11049 (1999).

158   Gella, A. & Durany, N. Oxidative stress in Alzheimer disease. *Cell Adh. Migr.* **3**, 88-93 (2009).

159   Muszyński, P. *et al.* The Relationship between Markers of Inflammation and Degeneration in the Central Nervous System and the Blood-Brain Barrier Impairment   in Alzheimer's   Disease.   *J.   Alzheimer's   Dis.*,   1-10, doi:10.3233/jad-170220 (2017).

160   Bachurin, S. O. *et al.* Novel conjugates of aminoadamantanes with carbazole derivatives as potential multitarget agents for AD treatment. *Sci. Rep.* **7**, 45627, doi:10.1038/srep45627 (2017).

161   Mufson, E. J., Counts, S. E., Perez, S. E. & Ginsberg, S. D. Cholinergic system during the progression of Alzheimer's disease: therapeutic implications. *Expert Rev. Neurother.* **8**, 1703-1718, doi:10.1586/14737175.8.11.1703 (2014).

162   Anand, P. & Singh, B. A review on cholinesterase inhibitors for Alzheimer's disease. *Arch. Pharmacal Res.* **36**, 375-399, doi:10.1007/s12272-013-0036-3 (2013).

163   Mushtaq, G., Greig, N. H., Khan, J. A. & Kamal, M. A. Status of acetylcholinesterase and butyrylcholinesterase in Alzheimer's disease and type 2 diabetes mellitus. *CNS Neurol. Disord. Drug Targets* **13**, 1432-1439 (2014).

164   Li, Q., Yang, H., Chen, Y. & Sun, H. Recent progress in the identification of selective butyrylcholinesterase inhibitors for Alzheimer's disease. *Eur. J. Med. Chem.* **132**, 294-309, doi:10.1016/j.ejmech.2017.03.062 (2017).

165   Micheau, J. & Marighetto, A. Acetylcholine and memory: A long, complex and chaotic but still living relationship. *Behav. Brain Res.* **221**, 424-429, doi:10.1016/j.bbr.2010.11.052 (2011).

166   Mesulam, M., Guillozet, A., Shaw, P. & Quinn, B. Widely spread butyrylcholinesterase can hydrolyze acetylcholine in the normal and Alzheimer brain. *Neurobiol. Dis.* **9**, 88-93, doi:10.1006/nbdi.2001.0462 (2002).

167   Bazelyansky, M., Robey, E. & Kirsch, J. F. Fractional diffusion-limited component of reactions catalyzed by acetylcholinesterase. *Biochemistry* **25**, 125-130 (1986).

168   Colovic, M. B., Krstic, D. Z., Lazarevic-Pasti, T. D., Bondzic, A. M. & Vasic, V. M. Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. *Curr. Neuropharmacol.* **11**, 315-335, doi:10.2174/1570159x11311030006 (2013).

169   Li, B. *et al.* Abundant tissue butyrylcholinesterase and its possible function in the acetylcholinesterase knockout mouse. *J. Neurochem.* **75**, 1320-1331 (2000).

170   Darvesh, S., Hopkins, D. A. & Geula, C. Neurobiology of butyrylcholinesterase. *Nat. Rev. Neurosci.* **4**, 131-138, doi:10.1038/nrn1035 (2003).

171   Mesulam, M. M. *et al.* Acetylcholinesterase knockouts establish central cholinergic pathways and can use butyrylcholinesterase to hydrolyze acetylcholine. *Neuroscience* **110**, 627-639 (2002).

172   Ehrlich, G. *et al.* Mapping the human acetylcholinesterase gene to chromosome 7q22 by fluorescent in situ hybridization coupled with selective

PCR amplification from a somatic hybrid cell panel and chromosome-sorted DNA libraries. *Genomics* **13**, 1192-1197 (1992).

173   Nicolet, Y., Lockridge, O., Masson, P., Fontecilla-Camps, J. C. & Nachon, F. Crystal Structure of Human Butyrylcholinesterase and of Its Complexes with Substrate and Products. *J. Biol. Chem.* **278**, 41141-41147, doi:10.1074/jbc.M210241200 (2003).

174   Cheung, J. *et al.* Structures of Human Acetylcholinesterase in Complex with Pharmacologically Important Ligands. *J. Med. Chem.* **55**, 10282-10286, doi:10.1021/jm300871x (2012).

175   Perkin, W. H. XVI.—On artificial alizarin. *J. Chem. Soc.* **23**, 133-143, doi:10.1039/js8702300133 (1870).

176   Ono, M., Kung, M.-P., Hou, C. & Kung, H. F. Benzofuran derivatives as Aβ-aggregate-specific imaging agents for Alzheimer's disease. *Nucl. Med. Biol.* **29**, 633-642, doi:10.1016/s0969-8051(02)00326-8 (2002).

177   Greig, N. H., Lahiri, D. K. & Sambamurti, K. Butyrylcholinesterase: An Important New Target in Alzheimers Disease Therapy. *Int. Psychogeriatr.* **14**, 77-91, doi:10.1017/s1041610203008676 (2002).

178   Makhaeva, G. F. *et al.* Conjugates of γ-Carbolines and Phenothiazine as new selective inhibitors of butyrylcholinesterase and blockers of NMDA receptors for Alzheimer Disease. *Sci. Rep.* **5**, doi:10.1038/srep13164 (2015).

179   Zha, X. *et al.* Novel Tacrine–Benzofuran Hybrids as Potent Multitarget-Directed Ligands for the Treatment of Alzheimer's Disease: Design, Synthesis, Biological Evaluation, and X-ray Crystallography. *J. Med. Chem.* **59**, 114-131, doi:10.1021/acs.jmedchem.5b01119 (2016).

180   Wu, W.-Y. *et al.* Novel multitarget-directed tacrine derivatives as potential candidates for the treatment of Alzheimer's disease. *J. Enzyme Inhib. Med. Chem.* **32**, 572-587, doi:10.1080/14756366.2016.1210139 (2017).

181   Santarpia, L., Grandone, I., Contaldo, F. & Pasanisi, F. Butyrylcholinesterase as a prognostic marker: a review of the literature. *Journal of Cachexia, Sarcopenia and Muscle* **4**, 31-39, doi:10.1007/s13539-012-0083-5 (2013).

182   Roccatano, D. *et al.* Influence of Hydroxyl Group Position and Temperature on Thermophysical Properties of Tetraalkylammonium Hydroxide Ionic Liquids with Alcohols. *PLoS One* **9**, e86530, doi:10.1371/journal.pone.0086530 (2014).

183   Barratt, E. *et al.* Thermodynamic Penalty Arising from Burial of a Ligand Polar Group Within a Hydrophobic Pocket of a Protein Receptor. *J. Mol. Biol.* **362**, 994-1003, doi:10.1016/j.jmb.2006.07.067 (2006).

184   Feng, Y. & Wang, X. Antioxidant Therapies for Alzheimer's Disease. *Oxid Med Cell Long* **2012**, 1-17, doi:10.1155/2012/472932 (2012).

185   Mecocci, P. & Polidori, M. C. Antioxidant clinical trials in mild cognitive impairment and Alzheimer's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1822**, 631-638, doi:10.1016/j.bbadis.2011.10.006 (2012).

186   Adalier, N. & Parker, H. Vitamin E, Turmeric and Saffron in Treatment of Alzheimer's Disease. *Antioxidants* **5**, 40, doi:10.3390/antiox5040040 (2016).

187   O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminform.* **3**, 33, doi:10.1186/1758-2946-3-33 (2011).

188   Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668-1688, doi:10.1002/jcc.20290 (2005).

189   Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999-2012, doi:10.1002/jcc.10349 (2003).

190   Brooks CL, I., Karplus M, Pettitt BM. . *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*.   (John Wiley and Sons; New York, 1988).