Xjenza Online - Journal of The Malta Chamber of Scientists
www.xjenza.org
DOI: 10.7423/XJENZA.2016.2.03

Review Article

Role of Protein Structure in Drug Discovery

Rosalin Bonetta[1], Jean-Paul Ebejer[1], Brandon Seychell[2], Marita Vella[2], Thérèse Hunter[2] and Gary J. Hunter[*2]
[1] *Centre of Molecular Medicine and Biobanking, University of Malta, Msida, MSD 2080, Malta*
[2] *Department of Physiology and Biochemistry, University of Malta, Msida, MSD 2080, Malta*

Abstract. Many pharmaceuticals currently available were discovered either during the screening of natural of synthetic product libraries or by serendipitous observation. Such a "random" approach entails testing numerous compounds and developing countless high-throughput screening assays. On the other hand, a "rational" approach involves the structure-based route to drug discovery, where the structure of a target protein is determined. Hypothetical ligands may be predicted by molecular modelling, while movement of a molecule may be predicted by Molecular Dynamics Simulations prior to synthetic chemical synthesis of a particular molecule. Here, we will be discussing protein structure-based approaches to drug discovery.

Keywords: Protein Structure, X-ray crystallography, Molecular Dynamics Simulations, Drug design

1 Introduction

Proteins are complex molecules composed of long strings of twenty different types of amino acid. The length of the string and the order of amino acids are vitally important for the protein to function properly in its biological role. This part of the process of protein function, the gene encoding the protein determines these factors. A single mistake (mutation) in the gene may cause the wrong amino acid to be incorporated into the sequence or a nonsense mutation may cause the protein to be truncated. However, protein function is more directly determined by the protein's three dimensional shape, the protein structure, and the availability of non-protein cofactors.

2 Protein Structure - why is it important?

As can be seen in Fig. 1, a complex protein such as xanthine oxidoreductase (XOR) forms a highly convoluted structure, but one which accommodates cofactors and substrates perfectly. Protein structure is typically determined by one of three methods today; X-Ray crystallography, Nuclear Magnetic Resonance spectroscopy (NMR) or cryoelectron microscopy. The former is the oldest and most commonly used technique, while the latter is only just becoming available for the analysis of proteins at atomic resolution. X-Ray crystallography relies on the ability of the protein to form a regular molecular array and crystallise; a completely biologically unnatural condition for any protein. Even so, it is possible and there are now over one hundred thousand entries in the biological structures databank, RCSB (Deshpande et al., 2005). The advantage of NMR over X-Ray crystallography is that it can be performed in solution (no crystals required) but the major problem is size; NMR cannot be used to determine the structure of large proteins. Electron microscopy will soon be capable of providing structural information about protein as good as X-Ray crystallography, and is performed in solution. Today it can yield protein structures to 2.2 Å (X-Rays typically give resolutions as high as 0.6 to 1.3 Å). In the laboratory of Biochemistry and Protein Science, at the University of Malta, we use X-Ray crystallography to determine protein structure, with crystallisation conditions determined in our laboratory applied and subjected to X-Ray diffraction at the University of Leeds, UK in collaboration with Dr Chi Trinh. We have determined the structures of several superoxide dismutase enzymes and mutants (to a minimum of 1.7 Å) and are currently working to solve the structures of others, in-

Correspondence to: Gary J. Hunter (gary.hunter@um.edu.mt)

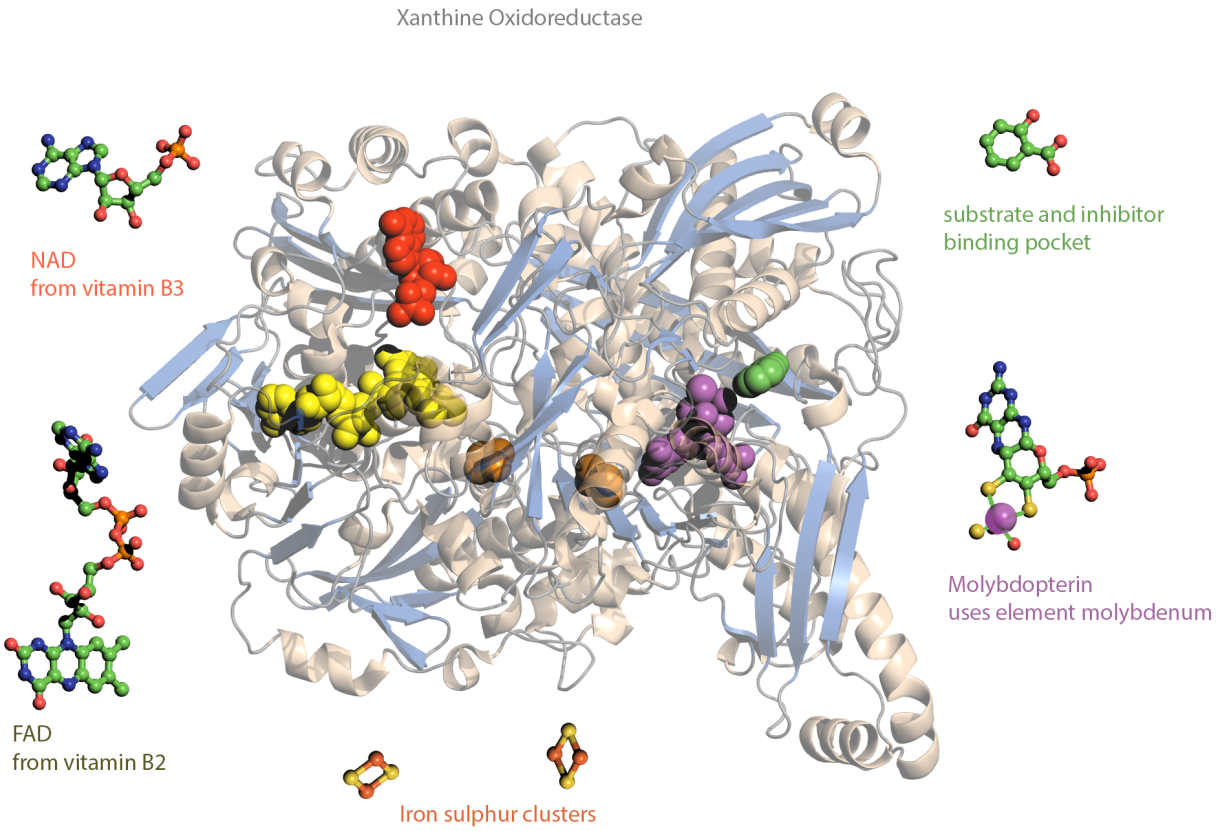

Figure 1: Bovine Xanthine Oxidoreductase. Only one of two identical protein subunits is shown, in cartoon representation with blue beta sheets (arrows) and cream alpha helices (spirals). Cofactors and substrates are shown surrounding the protein in ball-and-stick representation and their corresponding positions of binding within the protein as atomic spheres. NAD is in orange, FAD is in yellow, iron-sulphur clusters are brown, the molybdopterin is purple and xanthine (substrate) in green. Together, the protein forms a scaffold for the cofactors which form an electron transport chain from one side of the protein (substrate) to the other (FAD/NAD). The figure was created using the PyMOL molecular Graphics System (Schrödinger LLC, 2010).

cluding human XOR. Structures such as these help us to understand how the protein functions, and will help to design chemicals to be used pharmaceutically as modifiers of enzyme activity. X-Ray structures usually provide us with a quite static picture of the protein, and it is best combined with other techniques in order to obtain a detailed idea of how the protein functions.

3 Molecular Dynamics Simulations of Biomolecules

Molecular Dynamics Simulations are applied in the investigation of numerous dynamic properties and processes by scientists in a variety of fields that include structural biochemistry, enzymology, biophysics, molecular biology, biotechnology and pharmaceutical chemistry. Molecular Dynamics Simulations allow the researcher to study the thermodynamic and time-dependent (kinetic) properties of biomolecules such as proteins. This provides an understanding of numerous dynamic aspects of biomolecular structure, recognition, and function (Adcock & McCammon, 2006) The techniques involving Molecular Dynamics Simulations involve Langevin's or Newton's equations of motion, as well as a particular molecular bond structure, parametrized force fields, and an initial conformation of atomic positions, together with the velocities that are necessary to generate the atomic dynamics in a molecular system. Molecular Dynamics Simulations have a limited function when used in isolation. The trajectory of Molecular Dynamics (i.e., the progress of a simulated structure correlated to time) usually generates data related only to the level of atomic positions, velocities and single-point energies. Researchers are usually interested in obtaining macroscopic properties. The latter requires the application of statistical mechanics, which combines microscopic simulations together with macroscopic observables. Statistical mechanics provide the mathematical expressions associating the distributions

and motions of atoms to macroscopic observables including free energy, pressure and heat capacity (Callen, 1985; McQuarrie, Salvaterra, De Blas, Routes & Mahler, 1976). Molecular Dynamics Simulation programs include AMBER, CHARMM, NAMD and POLY-MD.

Kinetic rate constants of ligand-receptor interactions are essential in enzymology (Bar-Even et al., 2011) and drug discovery (Copeland, Pompliano & Meek, 2006), as they provide a good indication of drug efficacy (Copeland et al., 2006). Thus, the prediction and optimisation of these parameters is an important challenge in medicinal chemistry (Copeland, 2016). Even though these values may be measured experimentally, an accurate computational prediction would result in a useful alternative in cases where the experiment is either expensive or difficult to perform. Additionally, advances in computational power, have allowed simulations to be carried out in significantly less time. This provides a great potential for methods that require vast amounts of computational power.

Predicting the interaction between an enzyme and its substrate and other ligands via Molecular Dynamics Simulations is essential to fully understand the mechanism of the enzyme. Predicting hydrogen bonding in an enzyme is crucial for analysing the structure and function of this type of biological molecule, especially in terms of enzyme catalysis. Molecular Dynamics Simulations provide information on the molecule that is not observable in the data obtained via X-ray crystallography experiments alone. With this knowledge it is then possible to design new chemicals based upon the binding requirements discovered to inhibit or enhance the biological activity of the protein. In many cases it may be adventitious to modify the structure of an existing, known effector molecule (enhancer or inhibitor) to increase or decrease its activity. With computer aided rational design, a new or modified pharmaceutical may be created with better effectiveness and reduced side effects. Molecular simulations give us the power to suggest or reject such modifications prior to chemical synthesis of the compound. This saves time, effort and money.

4 Protein Structure, Molecular Dynamics, Drug Discovery - tying the knot with computational approaches

Structure-based virtual screening is a computational method employed to find small, bioactive molecules which sterically fit and interact with a protein. A library of small molecules (ligands) is "docked" to the protein's binding site in a typical "lock-and-key" fashion (Meng, Zhang, Mezei & Cui, 2011). Three things are required for this computational approach. Firstly, a protein structure is either determined experimentally (as described earlier) or modelled computationally, typically using homology modelling. In homology modelling, we use one or more known protein structures with close sequence similarity as a template to model our protein of interest. The binding site on the protein needs to be identified. Secondly, a library of small molecules must be prepared and provided to the docking algorithm. This preparation may imply many steps such as sanitisation, setting the appropriate ionization state, removing salts, etc. Thousands to millions of molecules form part of the digital library, only a fraction of which could possibly be tested physically in a laboratory. Thirdly, a docking protocol is required which defines the parameters used in the docking experiment. This includes, but is not limited to, ligand flexibility, protein side-chain flexibility, role of water molecule in the binding site, and which scoring function to use. The scoring function is of critical importance as it assesses the goodness of the fit, producing a quantitative score which can be used to rank each individual ligand. Many aspects are taken into consideration when evaluating the interaction of the protein with each ligand including steric fit, electrostatics, polar interactions and hydrogen bonding. The problem is compounded by the many possible conformations the ligand (or protein) takes on. The scoring function must evaluate each of these binding poses. Some of the major critiques of docking are the inability to calculate the free binding energy correctly (possibly because of the additive nature of most scoring functions), protein main-chain flexibility, the correct prediction of water in binding and the intensive computational resources required. In order to alleviate some of these issues, docking is sometimes used as a filtering first step before a more rigorous and computationally intensive molecular dynamics simulation. The top hits of the docking experiment are then rescored using MD. In a typical workflow, large virtual screening databases are first filtered using fast and inexpensive docking protocols. This rescoring is based on more physically realistic techniques for binding free energy estimations such as thermodynamic integration, free energy perturbation, linear interaction energy and molecular mechanics/Poisson-Boltzmann and surface area (MM/PB-SA). Overall, this provides a more accurate prediction of the binding affinity between the protein and the ligand (compared to the scoring function in docking tools). Computer-aided drug design is an active field of research, which has gained a lot of momentum in recent years - mostly driven by the decreasing productivity of the pharmaceutical industry to find new drugs.

5 Limitations of MD-based methods

The main force fields that are currently being employed for biomolecular simulations include AMBER (Asensio & Jimenez-Barbero, 1995), CHARMM (MacKerell et

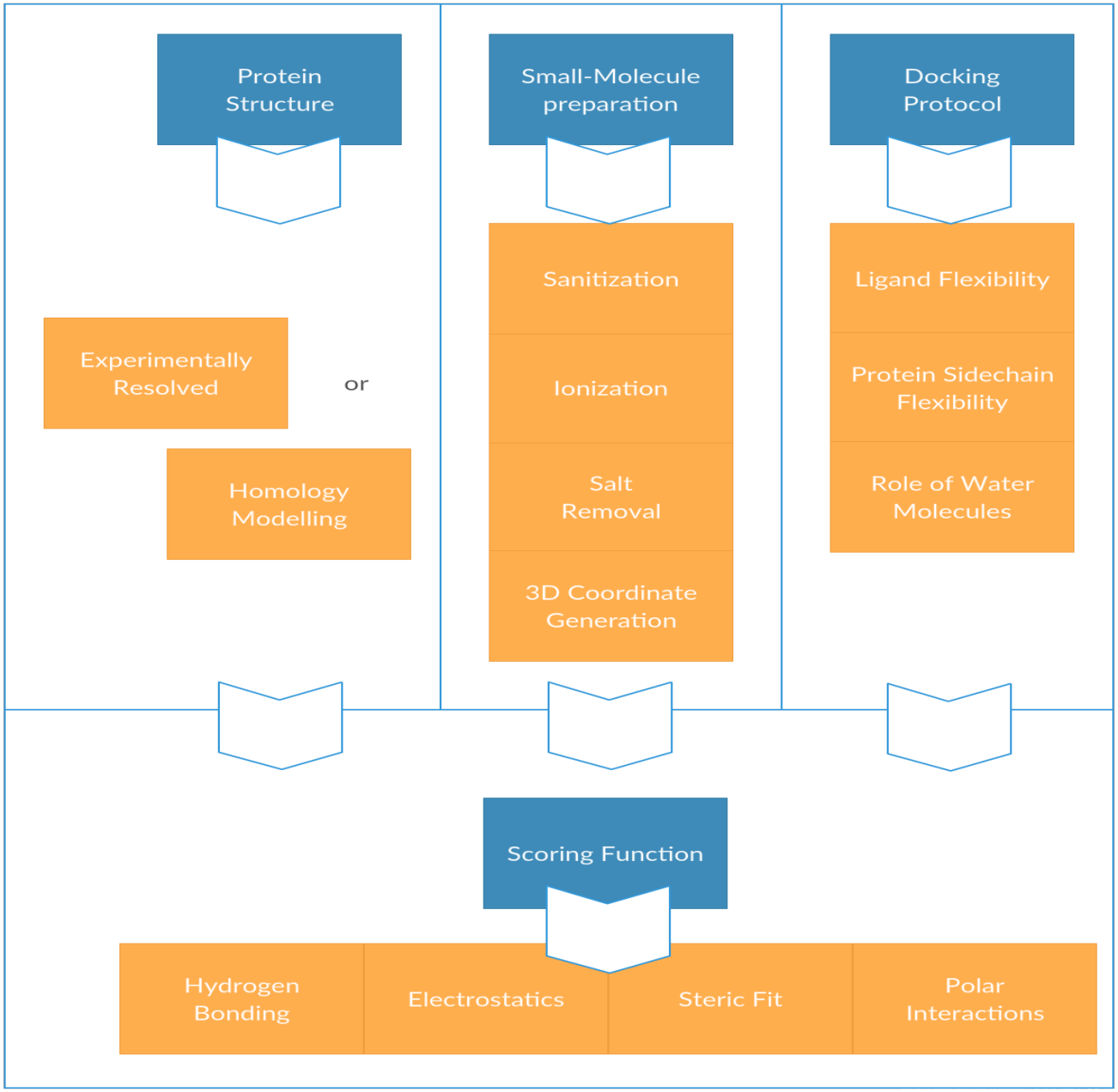

Figure 2: Components of a computational protein-ligand docking experiment. The goodness-of-fit of different small-molecules in a protein's pocket is assessed by means of a scoring function. The top-ranked results may serve as input to more computationally exhaustive techniques, such as Molecular Dynamics.

al., 1998), and OPLS (Jorgensen & Tirado-Rives, 1988). Although, extended parametrisation for amino acids, nucleic acids, lipids, carbohydrates, and several ionic species has been included in the parent force fields in recent years, the variability of small molecules (i.e., ligands) still poses a challenge to condensed-phase force fields. Thus, the user must carry out specific parametrisation. The latter is a time-consuming and an error-prone procedure, and has lead to the development of some general force field sets such as GAFF57 for AMBER, and CGenFF58 for CHARMM, together with specific parametrisation toolkits. Several challenges must be overcome to further increase the importance of MD-based methods on drug design. The molecular mechanics force fields that are presently available partially or fully neglect charge transfer and polarisation effects, as well as many electronic-based interactions. The current limits of force field and MD-based methods allow certain

target families, such as metalloproteins, to be studied with limited accuracy (De Vivo, Masetti, Bottegoni & Cavalli, 2016).

6 Conclusion

It is the combination of computational approaches that encompass techniques such as molecular dynamics simulations and docking, together with the interpretation of related experimental structural data, which is essential to provide a comprehensive understanding of the motions in proteins and their assemblies. Information on the latter is crucial when synthesising improved biomolecules and designing new drugs.

Acknowledgements

We would like to thank Dr Rebeca Garcia Fandinho from University of Santiago de Compostela, Spain, who is collaborating with us on Molecular Dynamics Simulations of different proteins. This collaboration was made possible due to our participation in COST Action CM1306 "Understanding Movement and Mechanism in Molecular Machines". We would also like to thank our collaborators Prof. Dr Arwen Pearson (University of Hamburg, Germany) and Dr Chi Trinh (University of Leeds, UK) for our ongoing collaborations in the fields of X-ray crystallography.

References

Adcock, S. A. & McCammon, J. A. (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev*, *106*(5), 1589–1615.

Asensio, J. L. & Jimenez-Barbero, J. (1995). The use of the AMBER force field in conformational analysis of carbohydrate molecules: determination of the solution conformation of methyl alpha-lactoside by NMR spectroscopy, assisted by molecular mechanics and dynamics calculations. *Biopolymers*, *35*(1), 55–73.

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S. & Milo, R. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, *50*(21), 4402–4410.

Callen, H. B. (1985). *Thermodynamics and an Introduction to Thermostatistics*. Student Edition. Wiley India Pvt. Limited.

Copeland, R. A. (2016). The drug-target residence time model: a 10-year retrospective. *Nat Rev Drug Discov*, *15*(2), 87–95.

Copeland, R. A., Pompliano, D. L. & Meek, T. D. (2006). Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov*, *5*(9), 730–739.

De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. (2016). Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem*, *59*(9), 4035–4061.

Deshpande, N., Addess, K. J., Bluhm, W. F., Merino-Ott, J. C., Townsend-Merino, W., Zhang, Q., ... Bourne, P. E. (2005). The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res*, *33*(Database Issue), D233–237.

Jorgensen, W. L. & Tirado-Rives, J. (1988). 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*(6), 1657–1666.

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., ... Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, *102*(18), 3586–3616.

McQuarrie, C., Salvaterra, P. M., De Blas, A., Routes, J. & Mahler, H. R. (1976). Studies on nicotinic acetylcholine receptors in mammalian brain. Preliminary characterization of membrane-bound alpha-bungarotoxin receptors in rat cerebral cortex. *J Biol Chem*, *251*(20), 6335–6339.

Meng, X. Y., Zhang, H. X., Mezei, M. & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des*, *7*(2), 146–157.

Schrödinger LLC. (2010). The PyMOL Molecular Graphics System (Version 1.3r1) [Computer Program].