Using Molecular Modelling to Study Interactions Between Molecules with Biological Activity

María J. R. Yunta

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/54007

1. Introduction

To better understand the basis of the activity of any molecule with biological activity, it is important to know how this molecule interacts with its site of action, more specifically its conformational properties in solution and orientation for the interaction. Molecular recognition in biological systems relies on specific attractive and/or repulsive interactions between two partner molecules. This study seeks to identify such interactions between ligands and their host molecules, typically proteins, given their three-dimensional (3D) structures. Therefore, it is important to know about interaction geometries and approximate affinity contributions of attractive interactions. At the same time, it is necessary to be aware of the fact that molecular interactions behave in a highly non-additive fashion. The same interaction may account for different amounts of free energy in different contexts and any change in molecular structure might have multiple effects, so it is only reliable to compare similar structures. In fact, the multiple interactions present in a single two-molecule complex are a compromise between attractive and repulsive interactions. On the other hand, a molecular complex is not characterised by a single structure, as can be seen in crystal structures, but by an ensemble of structures. Furthermore, changes in the degree of freedom of both partners during an interaction have a large impact on binding free energy [Bissantz et al., 2010].

The availability of high-quality molecular graphics tools in the public domain is changing the way macromolecular structure is perceived by researchers, while computer modelling has emerged as a powerful tool for experimental and theoretical investigations. Visualisation of experimental data in a 3D, atomic-scale model can not only help to explain unexpected results but often raises new questions, thereby affecting future research. Models of sufficient quality can be set in motion in molecular dynamic (MD) simulations to move beyond a static picture and provide insight into the dynamics of important biological processes.
Computational methods have become increasingly important in a number of areas such as comparative or homology modelling, functional site location, characterisation of ligand-binding sites in proteins, docking of small molecules into protein binding sites, protein-protein docking, and molecular dynamic simulations [see for example Choe & Chang, 2002]. Current results yield information that is sometimes beyond experimental possibilities and can be used to guide and improve a vast array of experiments.

To apply computational methods in drug design, it is always necessary to remember that to be effective, a designed drug must discriminate successfully between the macromolecular target and alternative structures present in the organism. The last few years have witnessed the emergence of different computational tools aimed at understanding and modelling this process at the molecular level. Although still rudimentary, these methods are shaping a coherent approach to help in the design of molecules with high affinity and specificity, both in lead discovery and in lead optimisation. Moreover, current information on the 3D structure of proteins and their functions provide a possibility to understand the relevant molecular interactions between a ligand and a target macromolecule. As a consequence, a comprehensive study of drug structure–activity relationships can help identify a 3D pharmacophore model as an aid for rational drug design, as a pharmacophore model can be defined as ‘an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response’, and a pharmacophore model can be established either in a ligand-based manner, by superposing a set of active molecules and extracting common chemical features that are essential for their bioactivity, or in a structure-based manner, by probing possible interaction points between the macromolecular target and ligands.

Molecular recognition (MR) is a general term designating non-covalent interactions between two or more compounds belonging to host-guest, enzyme-inhibitor and/or drug-receptor complexes. A rigorous approach to an MR study should involve the adoption of a computational method independent from the chemical intuition of the researcher. Drug design purposes prompt another challenging feature of such an ideal computational method, the ability to make sufficiently accurate thermodynamic predictions about the recognition process.

2. Molecular modelling methods and their usefulness

Molecular recognition is a central phenomenon in biology, for example, with enzymes and their substrates, receptors and their signal inducing ligands, antibodies and antigens, among others. Given two molecules with 3D conformations in atomic detail, it is important to know if the molecules bind to each other and, if it is so, what does the formed complex look like (“docking”) and how strong is the binding affinity (that can be related to the “scoring” functions).

Molecules are not rigid. The motional energy at room temperature is large enough to let all atoms in a molecule move permanently. That means that the absolute positions of atoms in a molecule, and of a molecule as a whole, are by no means fixed, and that the relative location
of substituents on a single bond may vary with time. Therefore, any compound containing one or several single bonds exists at every moment in many different conformers, but generally only low energy conformers are found to a large extent [Kund, 1997, as cited in Tóth et al., 2005].

The biological activity of a drug molecule is supposed to depend on one single unique conformation amongst all the low energy conformations, the search for this so-called bioactive conformation for compound sets being one of the major tasks in Medicinal Chemistry. Searching for all low energy conformations is possible with molecular modelling studies, since molecular modelling is concerned with the description of the atomic and molecular interactions that govern microscopic and macroscopic behaviours of physical systems. These molecular interactions are classified as: (a) bonded (stretching, bending and torsion), (b) non-bonded (electrostatic (including interactions with metals), van der Waals and $\pi$-stacking), and (c) derived, as they result from the previous ones (hydrogen bonds and hydrophobic effect).

Protein-ligand or, in general, molecule-molecule binding free energy differences can \textit{a priori} be computed from first principles using free energy perturbation techniques and a full atomic detailed model with explicit solvent molecules using molecular dynamics simulations. However, these are computationally demanding. More affordable approaches use end-point molecular dynamic simulations and compute free energies accounting for solvent effects with continuum methods, such as MM-PBSA (molecular mechanics Poisson-Boltzman surface area) or MM-GBSA (generalized Born surface area) [Kollman et al., 2000; Wang et al., 2005]. One of the first approaches was comparative molecular field analysis (CoMFA) [Cramer et al., 1988], which enabled interpretation and understanding of enzyme active sites when the crystal structure was absent. However, this type of analysis was not possible until \textit{in vitro} drug-drug interaction studies were widely used (through the 1990s).

2.1. Molecular mechanics, molecular dynamics and docking

Molecular mechanics (MM) is often the only feasible means with which to model very large and non-symmetrical chemical systems such as proteins and polymers. Molecular mechanics is a purely empirical method that neglects explicit treatment of electrons, relying instead on the laws of classical physics to predict the chemical properties of molecules. As a result, MM calculations cannot deal with problems such as bond breakage or formation, where electronic or quantum effects dominate. Furthermore, MM models are wholly system-dependent. MM energy predictions tend to be meaningless as absolute quantities, as the zero or reference value depends on the number and types of atoms and their connectivity, and so they are generally useful only for comparative studies. A force field is an empirical approximation for expressing structure-energy relationships in molecules and is usually a compromise between speed and accuracy.

Molecular mechanics have been shown to produce more realistic geometry values for the majority of organic molecules, owing to the fact that they are highly parameterised. Parameterisation of structures should be performed with care and non-“standard”
molecules will need to have new parameters. This is usually done by analogy for bonded
terms and assigning charges by a procedure consistent with the used force field.

There are many levels of theory in which computational models of 3D structures can be
constructed. The overall aim of modelling methods is often to try to relate biological activity
to structure. An important step towards this goal is to be able to compute the potential
energy of the molecule as a function of the position of the constituent atoms. Once a method
for evaluating the molecular potential energy is available, it is natural to search for an
optimum molecular geometry by minimising the energy of the system. In a biological
macromolecule, the potential energy surface is a complicated one, in which there are many
local energy minima as well as a single overall energy minimum. All the energy
minimisation algorithms commonly used have a marked tendency to locate only a local
energy minimum that is close to the starting conformation. For a biological macromolecule,
the number of conformations that have to be searched rises exponentially with the size of
the molecule; hence, systematic searching is not a practical method for large molecules.

Molecular dynamics (MD) is a conformation space search procedure in which the atoms of a
biological macromolecule are given an initial velocity and are then allowed to evolve in time
according to the laws of Newtonian mechanics [van Gunsteren & Berendsen, 1977].
Depending on the simulated temperature of the system, the macromolecule can then
overcome barriers at the potential energy surface in a way that is not possible with a
minimisation procedure. One useful combination of molecular dynamics and minimisation
schemes is a method known as simulated annealing [Kirkpatrick et al, 1983, Černý, 1985].
This method uses a molecular dynamics calculation in which the system temperature is
raised to a high value to allow for a widespread exploration of the available conformational
space. The system temperature is then gradually decreased as further dynamics are
performed. Finally, a minimisation phase may be used to select a minimum energy
molecular conformation.

One of the most important applications of molecular modelling techniques in structural
biology is the simulation of the docking of a ligand molecule onto a receptor. These methods
often search to identify the location of the ligand binding site and the geometry of the ligand
in the active site, to get the correct ranking when considering a series of related ligands in
terms of their affinity, or to evaluate the absolute binding free energy as accurately as
possible. To select a force field and the adequate modelling methodology for a given task, it
is important to appreciate the range of molecular systems to which it is applicable and the
types of simulations that can be performed.

2.2. Most used existing force fields

AMBER (Assisted Model Building with Energy Refinement) developed by Kollman et al.
[http://ambermd.org/] was originally parameterised specifically for proteins and nucleic
acids [Weiner et al., 1984, 1986; Cornell et al., 1995], using 5 bonding and non-bonding terms
along with a sophisticated electrostatic treatment and with no cross terms included. The
results obtained with this method can be very good for proteins and nucleic acids, but less
so for other systems, although parameters that enable the simulation of other systems have been published [for examples see Doshi & Hamelberg, 2009; Zgarbová et al., 2011].

CHARMM (Chemistry at HARward Macromolecular Mechanics) developed by Karplus et al. [http://www.charmm.org] was originally devised for proteins and nucleic acids [Brooks et al., 1983], and is now used for a range of macromolecules, molecular dynamics, solvation, crystal packing, vibrational analysis and QM/MM (quantum mechanics/molecular mechanics) studies. It uses five valence terms, one of which is electrostatic and is a basis for other force fields (e.g., MOIL [Elber et al., 1995]).

GROMOS (Groningen Molecular Simulation) developed at the University of Groningen and the ETH (Eidgenössische Technische Hochschule) of Zurich [http://www.igc.ethz.ch/GROMOS/index] is quite popular for predicting the dynamical motion of molecules and bulk liquids, also being used for modelling biomolecules. It uses five valence terms, one of which is electrostatic [van Gunsteren and Berendsen., 1977]. Its parameters are currently being updated [Horta et al., 2011].

MM1-4 (Molecular Mechanics) developed by Allinger [1976] are general purpose force fields for monofunctional organic molecules. The first version of this method was the MM1 [Allinger, 1976]. MM2 was parameterised for a lot of functional groups while MM3 [Allinger & Durkin, 2000; Allinger & Yan, 1993] is probably one of the most accurate ways of modelling hydrocarbons. MM4 is the latest version with several improvements [Allinger et al., 1996].

MMFF (Merck Molecular Force Field) developed by Halgren [1996] is also a general purpose force field mainly for organic molecules. MMFF94 [Halgren, 1996] was originally designed for molecular dynamics simulations, but has also been widely used for geometrical optimisation. It uses five valence terms, one of which is electrostatic and another is a cross term. MMFF was parameterised based on high level ab initio calculations. MMFF94 contains parameters for a wide variety of functional groups that arise in Organic and Medicinal Chemistry.

OPLS (Optimized Potential for Liquid Simulations) developed by Jorgensen at Yale [http://zarbi.chem.yale.edu] was designed for modelling bulk liquids [Jorgensen & Tirado-Rives, 1996] and has been extensively used for modelling the molecular dynamics of biomolecules. It uses five valence terms, one of which is an electrostatic term and none of them is a cross term.

TRIPOS (Sybil force field) is a commercial method designed for modelling organics and biomolecules. It is often used for CoMFA analysis and uses five valence terms, one of which is an electrostatic term.

CVFF (Consistent Valence Force Field) developed by Dauber-Osguthorpe is a method parameterised for small organic (amides and carboxylic acids, among others) crystals and gas phase structures [Dauber-Osguthorpe et al., 2004]. It handles peptides, proteins and a wide range of organic systems. It was primarily intended for studies of structures and binding energies, although it predicts vibrational frequencies and conformational energies reasonably well.
2.3. Popular docking programs

One of the most important and useful areas of application of molecular modelling is the approach of docking a protein onto a second molecule, typically a small ligand. This is of interest because it models the possible interactions between the protein and the ligand in the formation of a biologically important protein-ligand complex. To perform a computational docking, experimental or model 3D structures of both the protein and ligand molecules are required together with the charge distribution for each molecule.

There are several software programs that are available for carrying out docking calculations, only some of them will be considered here. The DOCK program suite [Kuntz, 1992] is one of the best known. First of all, a set of overlapping spheres are used in the program to construct a negative image of a specified site on the protein or another macromolecule, and the negative image is then matched against structures of potential ligands. Matches can be scored in this program by the quality of the geometric fit, as well as by the molecular mechanics interaction energy [Meng et al., 1992] and can lead to protein-binding ligands that have micromolecular levels of binding affinity [Kuntz et al., 1994]. It has also been used for modelling protein-protein docking [Shoichet & Kuntz, 1996].

The program GRID [Goodford, 1985] identifies likely protein binding sites for ligands [Reynolds et al., 1989; Cruciani & Goodford, 1994] using a 3D grid around the protein.

The program AutoDock developed by Morris et al.; [http://www.scripps.edu/pub/olson-web/doc/autodock/] uses a grid-based scheme for energies of individual atoms, allowing a quick computation of the interaction energy of the protein-ligand complex as the interaction between the ligand and the grid.

GLIDE software [Friesner et al., 2004, 2006; Halgren et al., 2004] also uses a grid-based scheme to represent the shape and properties of the receptor and then uses a systematic search algorithm to produce a set of initial conformations, using a OPLS-AA force field for ligand minimisation in the field of the receptor.

SURFLEX [Jain, 2003, 2007] is a fully automatic flexible molecular docking algorithm that presents results evaluated for reliability and accuracy in comparison with crystallographic experimental results on 81 protein/ligand pairs of substantial structural diversity.

In a recent study, comparison of seven popular docking programs [Plewczynski et al., 2011] clearly showed that the ligand binding conformation could be identified in most cases by using the existing software. Yet, there is still the lack of universal scoring function for all types of molecules and protein families. One can always hope that incremental improvements in current techniques will gradually lead to major advances in this field.

3. The solvent and how to model it

Solvation plays an important role in ligand-protein association and has a strong impact on comparisons of binding energies for dissimilar molecules. The binding affinity of a ligand
for a receptor ($\Delta G_{\text{bind}}$) depends on the interaction free energy of the two molecules relative to their free energy in solution:

$$
\Delta G_{\text{bind}} = \Delta G_{\text{interact}} - \Delta G_{\text{solv,L}} - \Delta G_{\text{solv,R}}
$$

where $\Delta G_{\text{interact}}$ is the interaction free energy of the complex, $\Delta G_{\text{solv,L}}$ is the free energy of desolvating the ligand, and $\Delta G_{\text{solv,R}}$ is the free energy of occluding the receptor site from the solvent. Various methods have been proposed to evaluate or estimate these terms. The problem is difficult because the energy of each component on the right hand side of Equation 1 is large while the difference between them is small.

An accurate way to calculate relative binding energies is with free-energy perturbation techniques, although they are usually restricted to calculating the differential binding of similar compounds and require extensive computation, making it impractical as an initial screen, but quite useful sometimes [Buch et al., 2011; Reddy & Erion, 2007]. Several authors have described force fields that consider the bound and solvated states [see for example Chen et al., 2008; Moon & Howe, 1991], successfully predicting new ligands and also the structures of ligand-receptor complexes [Wilson et al., 1991].

When calculating interactions in congeneric series, the cost in electrostatic free energy of desolvating both the enzyme binding site and the burial part of the ligand ($\Delta G_{\text{desolv}}$) is roughly constant within the series. This is particularly true when the calculation is done partitioning the electrostatic free energy contributions into a van der Waals term from the molecular mechanics force field, and an electrostatic contribution computed using a continuum method [Checa et al., 1997]. For that reason, it has been proposed to neglect $\Delta G_{\text{desolv}}$ in earlier studies.

The binding energy between ligand and receptor is approximated to the interaction enthalpy calculated by means of empirical energy functions that represent van der Waals repulsion, dispersion interactions by a Lennard-Jones term, and electrostatic interactions in the form of a Coulomb term that uses atom-centred point charges [Ajay & Murcko, 1995]. In most cases, these calculations of molecular mechanics are performed on a structure that is taken to represent the ensemble average of each complex. Entropy contributions are usually ignored although solvation terms are sometimes added to the scoring function by calculating changes in buried nonpolar surface area [Viswanadhan et al., 1999] or differences in the ease of desolvation of both the ligand and the binding site upon complex formation [Checa et al., 1997]. Molecular mechanics-based QSAR studies on ligand-receptor complexes can benefit greatly from proper incorporation of solvation effects into a COMBINE framework based on residue-based interaction energy decomposition [Pérez et al., 1998].

The relevance of solvation in modulating the biological activity of drugs is well known [Orozco & Luque, 2000]. In the last years, theoretical methods have been developed to calculate fragment contributions to the solvation free energy, particularly in the framework of quantum mechanical (QM) continuum solvation methods [Klamt et al., 2009]. Thus, fractional methods based on GB/SA methods have been developed [Cramer & Truhlar,
2008], as well as those based on the MST(Miertus-Scrocco-Tomasi) solvation method model [Soteras et al., 2004].

An explicit solvent model includes individual solvent molecules and calculates the free energy of solvation by simulating solute-solvent interactions. It requires an empirical interaction potential between the solvent and the solute, and between the solvent molecules, usually involving Monte Carlo (MC) calculations and/or molecular dynamics. MC calculations can be used to compute free energy differences and radial distribution functions, among others, and cannot be used to compute time-dependent properties such as diffusion coefficients or viscosity. MD simulations, on the other hand, can be used to compute free energies and time-dependent properties, transport properties, correlation functions, and others.

An implicit solvent model treats solvent as a polarisable continuum with a dielectric constant, $\varepsilon$, instead of explicit solvent molecules. The charge distribution of the solute polarises the solvent, producing a reaction potential that alters the solute. This interaction is represented by a solvent reaction potential introduced into the Hamiltonian. As interactions should be self consistently computed, they are also known as self-consistent reaction field (SCRF) methods [Onsager, 1936]. These models are significantly easier than explicit solvent models, but cannot model specific interactions such as hydrogen bonds.

Changes in hydration free energy during complex formation are a crucial element of binding free energies [Gilson & Zhou, 2007]. With the use of methods to predict binding free energies becoming common-place in the field of drug design, there is still a need for solvation methods that are both quick and accurate [Mancera, 2007], although much research has been carried out on the improvement of existing methods and development of new solvation models at many levels of theory [Chambers et al., 1996; Gallicchio et al., 2002; Palmer et al., 2011].

Explicit solvation models such as free energy perturbation (FEP), thermodynamic integration (TI) [Gilson & Zhou, 2007; Khavertsiki & Wallquist, 2010] and the faster linear interaction energy (LIE) [Aqvist et al., 1994; Carlson & Jorgensen, 1995] offer detail on the distinct nature of water around the solute and are transferable across a wide range of data sets, although there is a lack of throughput in the field of drug design.

Implicit solvation models offer a faster alternative to explicit models by replacing the individual water molecules with a continuous medium [Baker, 2005; Chen et al., 2008], combining the hydration free energy density and group contribution [Jäger & Kast, 2001], or, more recently, calculating solvation free energy directly from the molecular structure [Delgado & Jaña, 2009]. For small organic molecules, the loss of molecular detail of the solvent results in relatively small differences between hydration free energy prediction accuracies calculated with explicit solvent models relative to the explicit treatment [Mobley et al., 2009; Nicholls et al., 2009]. To cope with some of these pitfalls, the variational implicit solvent model (VISM) has been proposed for calculating the solute/water interface where established models fail [Dzubiella et al., 2006a, 2006b].
It is sometimes possible to get quite accurate results with very simple models, such as the case of the molecular modellisation of phenethylamine carriers conducted in our lab. Calculations were carried out using chloride anion to mimic the picrate anion used in experimental measurements and with no explicit solvent molecules. The chloroform environment was simulated by a constant dielectric factor, as this solvent has a low dielectric constant and thus, interactions should not end quickly with the distance [Campayo et al., 2005]. When complexation takes place in water as the solvent, the environment is simulated by a distance-dependent dielectric factor, as it takes into account the fact that the intermolecular electrostatic interactions should vanish with distance faster than in the gas phase. This assumption proves to work as it gives theoretical results in good agreement with experimental transportation values [Miranda et al., 2004; Reviriego et al., 2008]. Results for theoretical interactions have been supported by NMR experiments.

When applied to complex biomolecular systems, this loss of detail may become problematic in locations where water does not behave as a continuum medium, for example, the individual water molecules occurring in concave pockets at the surfaces of proteins [Li & Lazaridis, 2007]. The ELSCA (Energy by Linear Superposition of Corrections Approximation) method [Cerutti et al., 2005] has also been proposed for the rapid estimation of solvation energies. This procedure calculates the electrostatic and apolar solvation energy of bringing two proteins into close proximity or into contact compatible with the AMBER ff99 parameter set. The method is most useful in macromolecular docking and protein association simulations.

Solvent treatment is also of considerable interest in MD simulations as the solvent molecules (usually water, sometimes co-solvent and counterions/buffer or salt for electrolyte solutions) enter pockets and inner cavities of the proteins through their conformational changes. This is a very slow process and nearly as difficult to model as protein solving. One solution to this problem is using an efficient coupling of molecular dynamics simulation with the 3D molecular theory of solvation (3D-RISM-KH), contracting the solvent degrees of freedom [Luchko et al., 2010] or using free energy perturbation and OPLS force field together with molecular dynamics [Shivakumer et al., 2010].

4. Molecule-molecule or ion-molecule interactions in active molecule design

Non-covalent interactions are central to biological structure and function. In considering potential interactions of molecules and/or ions and their receptor, the focus has been on hydrophobic interactions, hydrogen bonding and ion pairing. Although hydrogen bonds are by far the most important interactions in biological recognition processes, the cation-π interaction is a general, strong, non-covalent binding force that occurs throughout nature, being energetically comparable or stronger than a typical hydrogen bond.

Cooperativity in multiple weak bonds (hydrogen bond and ion-π interactions among others) has been considered and studied at the MP2/6-311++ G(d,p) computational level.
[Alkorta et al., 2010]. Due to the presence of a great number of aromatic rings containing heteroatoms in biological systems, this effect might be important and help to understand some biological processes where the interplay between both interactions may exist.

Computer-assisted drug design (CADD) has contributed to the successful discovery of numerous novel enzyme inhibitors, having been used to predict the binding affinity of an inhibitor designed from a lead compound prior to synthesis [Reddy & Erion, 2005]. A free energy simulation technique known as the thermodynamic cycle perturbation (TCP) approach [Reddy et al., 2007], used together with calculations of molecular dynamics, offers a theoretically precise method to determine the binding free energy differences of related inhibitors.

Many small molecules are transported across cell membranes by large integral membrane proteins, which are referred generically as transporters. Selection among competing alternatives is always interesting and cation-π interactions are strongly involved in substrate recognition by many transporters. Drug transporters are able to carry small molecules or ions across membranes, being an important target for pharmaceutical development [Zacharias & Dougherty, 2002].

The regulation of metal ions plays a major role in enzymes, allowing to catalyse a range of biological reactions. Identification and characterisation of the metal ion binding sites and their selectivity have received immense attention over the past few decades [Ma & Dougherty, 1997]. It is evident from earlier studies that metal ions can bind to aromatic groups in a covalent as well as non-covalent fashion. Non-covalent interactions between metal ions and an aromatic ring, which are considered strong cation-aromatic interactions, are increasingly being recognised as an important binding force relevant to structural biology [Meyar et al., 2003; Elguero et al., 2009]. However, in many cases, the cation is the side chain protonated nitrogen of a basic amino acid. Reddy et al. have made available a web-based cation-aromatic database (CAD) including metal ions and basic amino acids [Reddy et al., 2007b].

Macrocyclic entities that act as ion receptors and carriers exhibit a large number of conformations in crystals and solutions, depending on the nature of their environments and of the complexed ion. To ensure the formation of the most favourable cavity for a given ion, as well as to enhance the binding and release of the ion during transport at interfaces, flexibility in the ligand structure is of utmost importance.

Complexation studies of ions with macrocycles are well documented in the literature. Some representative trends in these studies would include the following: taking into account the existence of hydrogen-bonded water molecules [Hill & Feller, 2000; Durand et al., 2000; Fantoni, 2003] and sometimes using molecular dynamics and free energy perturbation studies [Varnek et al., 1999]. The complexation phenomena have also been studied in cases where the ligand can exist as different conformers able to complex the cation [Hashimoto & Ikuta, 1999].
One of the most important aspects of ion complexation is ion selectivity, which is considered in terms of the more or less favourable binding energies. The binding energy (BE) is defined as the difference between the energy of the complex and the energy of the free ligand and ion:

\[
BE = E_{\text{complex}} - (E_{\text{ion}} + E_{\text{ligand}})
\]

Metal ion affinity is enhanced if the host molecule has a unique conformation that is optimal for complexation, that is, with all the binding sites positioned to structurally complement the metal ion [Lumetta et al., 2002].

Density functional theory based on electronic structure calculations is computationally affordable. It has very good predictability power for various structural and thermodynamic properties of a molecular system, and has therefore been used to model M⁺-crown ether complexes [Ali et al., 2008] and collarenes acting as ionophores and receptors [Choi et al., 1998]. In both cases, the most stable equilibrium structure for complexes are estimated based on PM3 semi-empirical calculations followed by B3LYP calculations using the G-311++G(d,p) basis set of functions. Currently, the Protein Data Bank (PDB) contains over 25,000 structures that contain a metal ion. Thus, methodologies to incorporate metal ions into the AMBER force field have been developed there [Hoops et al., 1991; Reichert et al., 2001; Peters et al., 2010].

Molecular modellisation of Cu(II) and Zn(II) coordination complexes has been studied by our research group, among many others. The parameters used by us were checked against a known X-ray structure and the data obtained agreed quite well with similar deviations published for other theoretical results [Miranda et al., 2005]. The models obtained were useful in explaining the differences observed among the complexes obtained in different environments. Our cation metal parameters have also been used to help in the data elucidation of coordination metal complex structures [Rodríguez-Ciria, 2000; Rodríguez-Ciria et al., 2002].

Coordination and complexation of ions by aromatic moieties have been studied, taking into account the different characteristics of the electronic charge distribution on the aromatic frame as an addition of coulombic potentials [Alberti et al., 2010]. This type of interaction is quite common in biology as signalling in the nervous system is generally mediated by the binding of small molecules (neurotransmitters) to the appropriate receptors, which usually contain a cationic group at physiological pH.

An organic ammonium ion never exists as a sole cation; an anion is always associated with it. Depending on the polarity and hydrogen donor/acceptor abilities of the solvent, the association strength is different [Marcus & Hefter, 2006]. Strongly coordinating counter ions such as chloride generally lead to weaker binding constants upon recognition of the associated cation, when compared to weakly coordinating counterions such as iodide or perchlorate [Gevorkyan et al., 2001].
5. Selective formation of complexes

There are several examples of molecular modelling studies on complexes between cyclic receptors and ammonium ions, calixarenes [Choe & Chang, 2002] and crown ethers being the most used. As an example, it is noteworthy to mention the theoretical studies on calix[4]crown-5 and a series of alkyl ammonium ions [Park et al., 2007], having shown that the energy of complex formation depends on the number of amine groups in the alkyl chain as well as on the number of methylene groups between the primary and secondary amine groups, results that agree with experimental measurements. Although the calculations are performed under quite different conditions of vacuum compared with the experimental conditions of the phase system of chloroform-water, the binding properties of calixarene-type compounds towards alkyl ammonium ions have been successfully simulated, providing general and useful explanations for the molecular recognition behaviour.

Complex formation of compounds containing benzene rings with ammonium cations has also been theoretically studied using many computational techniques, including ab initio calculations [Kim et al., 2000]. It has been shown that two types of NH-aromatic π and CH-aromatic π interactions, which are important in biological systems, are responsible for binding, and that charged hydrogen bonds versus cation-π interaction is the origin of the high affinity and selectivity of novel receptors for NH₄⁺ over K⁺ ions [Oh et al., 2000]. Organic molecules complexed with metal cations have also been studied by MM2 molecular modelling [Mishra, 2010]. The search for metal ion selectivity is of interest in the field of biomimetic models of metalloenzymes and molecular modelling helps in the design of new ligands with this purpose [Kaye, 2011].

Molecular modelling has been used to suggest possible contributions of carrier effectivity and selectivity to complex formation in accordance with experimental results [Chipot et al., 1996; Ilioudis et al., 2005]. Our research group has evaluated the possible cation-receptor interactions involved in the complexes with ammonium and metal cations of selective carriers using the Amber force field with appropriate parameters developed by us. The complexation energies obtained are in reasonable agreement with experimental values, taking into account that complexation/decomplexation processes have a great influence on transport rates and are not equally favoured in cyclic and acyclic carriers [Campayo et al., 2004].

Both binding and selectivity in binding can be understood through the combined efforts of several non-covalent interactions, such as hydrogen bonding, electrostatic interactions, hydrophobic interactions, cation-π interactions, π-π stacking interactions and steric complementarity [Späth & König, 2010]. Formation of complexes is also possible in the case of neutral ligands. For example, the interactions between cholesterol and cyclodextrins have been theoretically studied to investigate their 1:1 and 1:2 complexes [Castagne et al., 2010], while the formation of stable complexes between trehalose and benzene compounds have been investigated by the general Amber force field (GAFF) and Gaussian 03 for MP2/6G-31G** calculation of atomic charges [Sakakura et al., 2011].
Docking of a ligand into a receptor may occur via an automated procedure [Subramanian et al., 2000] or manually [Filizola et al., 1999]. In both cases, docking is a combination of two components: a search strategy and a scoring function [Taylor et al., 2002]. The computational method MOLINE (Molecular Interaction Evaluation) was created to study complexes in an unbiased fashion [Alcaro et al., 2000]. It is based on a systematic, automatic and quasi-flexible docking approach that prevents the influence of the chemist’s intuition on generating the configuration. This method has been used with acceptable results in studying inclusion complexes [Alcaro et al., 2004].

It would be adequate at this point to remember that testing the ‘drug-receptor complexation’ for a receptor model against available experimental data usually involves the use of site-directed mutagenesis experiments. This fact provides information on the amino acids involved in ligand binding and receptor activation. However, it should be noted that the results of mutagenesis studies are not necessarily related to receptor-ligand interactions. In fact, mutations can also alter the 3D structure of a receptor and therefore, modify the binding profile of a ligand by this mechanism. Besides that, efficient binding to a receptor does not guarantee that a ligand will produce a pharmacological action, given that the ligand may act as an agonist or antagonist.

6. Interaction of molecules with DNA

Anthrycycline antibiotics such as doxorubicin and its analogues have been in common use as anticancer drugs for around half a century. There has been intense interest in the DNA-binding sequence specificity of these compounds in recent years, with the hope of identifying a compound that can modulate gene expression or exhibit reduced toxicity. Cashman and Kellog have studied models of binding for doxorubicin and derivatives [Cashman & Kellog, 2004], looking for sequence specificity and the effects of adding aromatic or aliphatic ring substituents or additional amino or hydroxy groups. They performed a hydropathic interaction analysis using the HINT program (a Sybyl program module, Tripos Inc.) and four double base pair combinations. Interaction of some intercalators with two double DNA base pairs have also been studied with the density functional based tight binding (DFTB) method [Riahi et al., 2010], despite DFT methods being known to be inherently deficient in calculating stacking interactions, and the Amber force field and then AM1 to dock the intercalator between DNA base pairs [Miri et al., 2004].

Studies on sequence-selectivity of DNA minor groove binding ligands have shown that the most reliable results for AT-rich DNA sequences are obtained when MD simulations are performed in explicit solvent, when the data are processed using the MM-PB/SA approach, and when normal mode analysis is used to estimate configurational entropy changes [Shaikh et al., 2004; Wang & Laughton, 2009]. Use of the GB/SE model with a suitable choice of parameters adequately reproduces the structural and dynamic characteristics in explicitly solvated simulations in approximately a quarter of the computational time, although limitations become apparent when the thermodynamic properties are evaluated [Sands & Laughton, 2004]. Water molecules taking part in the complexation have been studied using
the MMX force field and then PMB for gas phase optimisation, followed by re-optimisation in aqueous phase with the PM3 method using the AMSOL package [Silva & Jayasundera, 2002]. Optimisation geometries with AM1 and the use of implicit solvent have been taken into account when considering intercalation versus insertion into the minor or major groove [Bendic & Volanschi, 2006].

Calculating the curvature radius of molecular DNA structures has been reported [Slickers et al., 1998] as a new method for understanding the dependence of binding affinity on ligand structure, assuming that strong binders should have a shape complementary to the DNA minor groove. A method for predicting sequence selectivity and minor groove binding, based on MD simulations on DNA sequences with and without the bound ligand, to obtain an approximate free energy of binding has been proposed [Wang & Laughton, 2010].

Amber force field, developing the necessary parameters, has also been used together with electrostatic potential-derived (ESP) charges and explicit solvent molecules to study bisintercalation into DNA. The targeted molecular dynamics (tMD) approach has been considered for comparing the relative energetic cost involved in creating the intercalation sites and also studying the mechanisms of action [Braña et al., 2004]. It has been found that the electrostatic contribution is a critical characteristic of binding selectivity [Marco et al., 2005]. Reports on duplex and triplex formation of oligonucleotides by stacking aromatic moieties in the major groove, using Amber force field and the GB/SA solvation model in molecular dynamic simulations, can be found in the literature [Andersen et al., 2011]. Studies on docking using GOLD [Kiselev et al., 2010] to optimise the starting structures with the MMFF94 force field have also been performed.

Most of the published molecular modelling studies use two double base pairs or more than eight double base pairs to represent DNA. In our opinion, molecular modelling of DNA intercalation complexes should be done using at least the two base pairs of the intercalation site and an additional base pair at the two strand ends to maintain DNA shape and avoid distortion leading to inaccurate results. That means four base pairs for monointercalation studies and five or six base pairs for bisintercalation ones should be used. Using these DNA models, our studies on the mono and bisintercalation of benzo[g]phthalazine derivatives strongly suggest the possibility of bisintercalation and the important role played by an N-methyl group in stabilising the DNA complex of one of the compounds, throwing some light over the experimental results obtained [Rodríguez-Ciria et al., 2003]. The possibility of bisintercalation for a 1,4-disubstituted piperazine has been studied on duplexes of five and six base pairs, obtaining much better results in the case of five base pairs, in accordance with the theoretical calculations of binding mode not conforming to the neighbouring exclusion principle proposed by different authors [Veal et al., 1990].

7. Interaction of small molecules with enzymes

The potential of molecular simulations to enhance our understanding of drug behaviour and resistance relies ultimately on their ability to achieve an accurate ranking of drug binding affinities at clinically relevant time scales. Several computational approaches exist to
Using Molecular Modelling to Study Interactions Between Molecules with Biological Activity

estimate ligand binding affinities and selectivities, with various levels of accuracy and computational expense: free energy perturbation (FEP), thermodynamic integration (TI), lineal response (LR), and molecular mechanics Poisson-Boltzman surface area (MM/PBSA). Identification of conformational preferences and binding site residues, as well as structural and energetic characterisation, is possible using MD simulations [Anzini et al., 2011; Dastidor et al., 2008; Stoika et al., 2008]. It is also possible to estimate conformational energy penalties for adopting the bioactive conformation identified by using a pharmacophore model [Frølund et al., 2005].

A model based on van der Waals intermolecular contribution from Amber and electrostatic interactions derived from the Poisson-Boltzman equation has been used to predict the change in the apparent dissociation constant for a series of six enzyme-substrate complexes during COMBINE analysis [Kmunicek et al., 2001]. In COMBINE analysis, binding energies are calculated for the set of enzyme-substrate complexes using the molecular mechanics force field. The total binding energy, $\Delta U$, may be assumed to be the sum of five terms: the intermolecular interaction energies between the substrate and each enzyme residue, $E_{interES}$, the change in the intramolecular energy of the substrate upon binding to the enzyme, $\Delta E^S$, the change in the intramolecular energy of the enzyme upon binding, $\Delta E^E$, the desolvation energy of the substrate, $E_{desolvS}$, and the desolvation energy of the enzyme, $E_{desolvE}$.

$$\Delta U = E_{interES} + \Delta E^S + \Delta E^E + E_{desolvS} + E_{desolvE}$$ (3)

When the substrate is a rather small molecule, there is no evidence for large differences in the structure of the enzyme when different substrates are bound and so the second and third are neglected. This method identifies the amino acid residues responsible for modulating enzyme activity [Kmunicek et al., 2005].

Molecular modelling of proteins is sometimes directed towards homology modelling, enabling progress in understanding the mechanisms of action despite the lack of detailed information on the 3D structure of a protein. Molecular dynamic simulations are usually used to test the stability of the complete structure derived from homology modelling [Srinivas et al., 2006].

Molecular docking examples can be used to compare relative stabilities of the complexes, but not calculate binding affinities, since changes in entropy and solvation effects are not taken into account [Pastorin et al., 2006; Tscharmer et al., 2011]. In any case, docking calculations are common studies on novel drugs, Autodock being one of the most used docking programs [see for example Venskutonyte et al., 2011]. Docking programs treat enzymes and substrates as rigid entities, but flexible docking is also possible, if several different protein conformations extracted from molecular dynamic simulations are used [Roumen et al., 2010].

In our laboratory, molecular modelling has been tentatively used to study the trypanosomicidal activity of some phthalazine derivatives. Results obtained with Amber force field implemented in HyperChem 8.0 plus our own necessary parameters, and with AutoDock 4.2 using the PDB structure for T. cruzi Fe-SOD enzyme, were in accordance with
experimental data, helping to explain the experimental results obtained. However, if there is no PDB structure for the desired enzyme and only a model of the active site, as for *Leishmania* Fe-SOD enzyme, results obtained with our calculations do not agree with the experimental ones when compared to the *T. cruzi* ones. This indicates that the interaction with the external part of the enzyme plays an important role as it might collaborate in, or make access to the active site difficult, since the enzyme shape and conformation plays a crucial role in its activity [Sanchez-Moreno et al., 2011; Yunta, unpublished results].

8. Conclusion

Modern molecular modelling techniques are remarkable tools in the search for potentially novel active agents by helping to understand and predict the behaviour of molecular systems, having assumed an important role in the development and optimisation of leading compounds. Moreover, current information on the 3D structure of proteins and their functions provide a possibility of understanding the relevant molecular interactions between a ligand and a target macromolecule. Although improvements are still needed in the techniques used, they have been shown to be invaluable in structure–activity relationship research.

On the basis of the current improved level of understanding of molecular recognition and the widespread availability of target structures, it is reasonable to assume that computational methods will continue to aid not only the design and interpretation of hypothesis-driven experiments in disease research, but also the fast generation of new hypotheses.

Author details

María J. R. Yunta
*Universidad Complutense, Madrid, Spain*

Acknowledgement

Financial support from the Spanish MEC project (CGL2008-0367-E/BOS) and the MCINN projects (CTQ2009-14288-C04-01 and CONSOLIDER INGENIO 2010 CSD2010-00065) are gratefully acknowledged.

9. References


Using Molecular Modelling to Study Interactions Between Molecules with Biological Activity 203


Rodríguez-Ciria, M. (2000). Síntesis de 1-amino y 1,4-diamino derivados de benzo[g]ftalazina, evaluación de sus propiedades complejantes frente a cationes metálicos y catecolaminas involucrados en mecanismos de neurotransmisión. *Ph.D. Thesis*, Universidad Complutense, Madrid


