Bioorganic & Medicinal Chemistry 24 (2016) 2159–2189

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Towards understanding the unbound state of drug compounds: Implications for the intramolecular reorganization energy upon binding

Nicolas Foloppe *, I-Jen Chen *

Vernalis (R&D) Ltd, Granta Park, Abington, Cambridge CB21 6GB, UK

ARTICLE INFO

Article history:
Received 14 January 2016
Revised 9 March 2016
Accepted 12 March 2016
Available online 14 March 2016

Keywords:
Conformational dynamics
Drug discovery
Intramolecular energy
Molecular modelling
Molecular recognition
Reorganization energy
Simulation

ABSTRACT

There has been an explosion of structural information for pharmaceutical compounds bound to biological targets, but the conformations and dynamics of compounds free in solution are poorly characterized, if at all. Yet, knowledge of the unbound state is essential to understand the fundamentals of molecular recognition, including the much debated conformational intramolecular reorganization energy of a compound upon binding (ΔE_{Reorg}). Also, dependable observation of the unbound compounds is important for ligand-based drug discovery, e.g. with pharmacophore modelling. Here, these questions are addressed with long ($\geqslant$0.5 μs) state-of-the-art molecular dynamics (MD) simulations of 26 compounds (including 7 approved drugs) unbound in explicit solvent. These compounds were selected to be chemically diverse, with a range of flexibility, and good quality bioactive X-ray structures. The MD-simulated free compounds are compared to their bioactive structure and conformers generated with ad hoc sampling in vacuo or with implicit generalized Born (*GB*) aqueous solvation models. The *GB* conformational models clearly depart from those obtained in explicit solvent, and suffer from conformational collapse almost as severe as in vacuo. Thus, the global energy minima in vacuo or with *GB* are not suitable representations of the unbound state, which can instead be extensively sampled by MD simulations. Many, but not all, MD-simulated compounds displayed some structural similarity to their bioactive structure, supporting the notion of conformational pre-organization for binding. The ligand–protein complexes were also simulated in explicit solvent, to estimate ΔE_{Reorg} as an enthalpic difference ΔH_{Reorg} between the intramolecular energies in the bound and unbound states. This fresh approach yielded ΔH_{Reorg} values $\leqslant$ 6 kcal/mol for 18 out of 26 compounds. For three particularly polar compounds $15 \leqslant \Delta H_{Reorg} \leqslant 20$ kcal/mol, supporting the notion that ΔH_{Reorg} can be substantial. Those large ΔH_{Reorg} values correspond to a redistribution of electrostatic interactions upon binding. Overall, the study illustrates how MD simulations offer a promising avenue to characterize the unbound state of medicinal compounds.

1. Introduction

Using structural information on compounds and their biological targets has become mainstream to elaborate small molecules in drug discovery. Indeed, one can frequently visualize the bioactive structure of medicinal compounds bound to their biological targets, usually via X-ray crystallography.[1] However, much less is known about the structural properties of the compounds in the unbound state in aqueous solution.[2,3] Yet, understanding the physical-chemistry of the recognition of a compound by its biological target(s), its binding free energy and kinetics, requires a characterization of the unbound state in addition to the bound state. Reliable structural information about an unbound ligand could facilitate the design of its conformational preorganization for molecular recognition

Abbreviations: ACE, angiotensin converting enzyme; DHFR, dihydrofolate reductase; *Diel*, distance-dependent dielectric; FF, force field; *GB*, generalized Born solvation model; *GlobMin_Diel*, global energy minimum obtained with a distant-dependent dielectric constant; *GlobMin_GB*, global energy minimum obtained with a GB solvation model; *GlobMin_Vac*, global energy minimum obtained in vacuo; GPU, Graphical Processing Unit; GUI, graphical user interface; LowModeMD, search method combining low-mode moves and molecular dynamics; MD, molecular dynamics; MM, molecular mechanics; MMFF, merck molecular force-field; MT/LMOD, Mixed Torsional/Low-mode; MOE, molecular operating environment; MW, molecular weight; MM, molecular mechanics; NMR, nuclear magnetic resonance; NRot, Oprea number of rotatable bonds; OPLS, optimized potential for liquid simulations force-field; *PB*, Poisson–Boltzmann; PDB, Protein Data Bank; Rgyr, radius of gyration; SBDD, structure-based drug design; VS, virtual screening.

* Tel.: +44 (0) 1223 895 338; fax: +44 (0) 1223 895 556 (N.F.); tel.: +44 (0) 1223 895 354; fax: +44 (0) 1223 895 556 (I.-J.C.).

E-mail addresses: n.foloppe@vernalis.com (N. Foloppe), i.chen@vernalis.com (I-Jen Chen).

http://dx.doi.org/10.1016/j.bmc.2016.03.022
0968-0896/

and selectivity. It would provide a sounder basis for ligand-based inferences,[4–6] and would be relevant to investigations of membrane permeability.[7,8] Furthermore, an accurate representation of the unbound ligands is essential for theoretical insights, including the quantification of a compound intramolecular conformational reorganization energy upon binding to a macromolecule, which remains a strongly debated question.[2,5,9–17] The reorganization energy has also been called the conformational energy penalty upon binding[10] or the strain energy.[13]

A compound in the unbound state is generally thought of as being in dynamic equilibrium between a number of conformations. An accurate representation of such conformational ensemble has been elusive, because it is not only about identifying conformers, but also being able to discern their populations in aqueous solution. It is of course possible to obtain an X-ray structure of a compound on its own,[18] but that would not describe the conformational spread of a flexible molecule in solution. Thus, the main contribution of small molecule crystallography resides arguably in the statistical distributions of conformations for specific functional groups,[19,20] but that does not tackle a molecule as a whole. Alternatively, NMR spectroscopy can yield highly valuable insights on the conformational flexibility of a free compound.[21–25] For example, a recent NMR study of Streptomycin in solution found that its conformers can be grouped in two main families of three-dimensional shapes, one of them was very similar to the bioactive shape.[25] This led to the suggestion that determining the unbound conformations of a ligand may help access its bioactive shape in absence of crystallography, and that NMR is the method of choice for these efforts.[3,25] Yet, there is apparently surprisingly little NMR work performed in this area, maybe because that remains labor intensive, such that only a small number of compounds can be studied in details. The interpretation of NMR data is complicated by the underlying dynamic conformational exchanges, and it usually involves molecular modelling tools to distil the results. Even in absence of experimental restraints, the tools of computational chemistry can help gain insights into the unbound state of small molecules.

There are many computational methods to explore the conformations of small molecules.[2,26] All conformational search algorithms use a scoring or energy model to evaluate the likelihood of a conformer; when many conformers must be generated and evaluated, intramolecular energies are usually estimated with a molecular mechanics force-field.[2,26–30] Errors on such energies can be on the order of a few kcal/mol for diverse chemistries.[14,30–34] Conformers are sometimes generated in vacuo,[8,9,17,35,36] but force-field intramolecular energies may be combined with implicit (continuum) aqueous solvation models which estimate the influence of water on the stability of a conformer.[37] Implicit solvation models are popular since they are much faster to compute than with explicit solvent, however the underlying mean-field approximations neglect the details of solvent structure around the solute.[37] Dampening electrostatic interactions with a distance-dependent dielectric constant (*Diel*) provides a rudimentary implicit solvation model still in use.[5,38] More sophisticated continuum solvation frameworks include the Poisson–Boltzmann (*PB*)[11,37,39,40] and generalized Born (*GB*)[37,40–42] treatments. The modern *GB* models perform well when compared to the more rigorous *PB* formalism,[40] and they have been popular with small molecules,[5,10,13,38,41,43,44] with various implementations in different molecular modelling packages.

Since *GB* models can be computed efficiently, they are frequently used in conjunction with ad hoc conformational sampling methods which aim to quickly generate diverse conformers of a compound.[2] These methods use various expedient schemes to traverse physical energy barriers,[2] hence their ad hoc character. Such ad hoc sampling currently provides the main avenue to generate conformers of free compounds, e.g. for ligand-based modelling.[5,8,35,38,45–50] Indeed, ad hoc methods were used to identify the global energy minimum conformer, which typically represented the compound unbound state in studies of the intramolecular reorganization energy upon binding.[5,9,10,12,13,17] Since ad hoc sampling methods operate independently of thermodynamically controlled conditions of temperature and pressure, they do not inform directly about the populations of the conformers. This is a major difference with the conformational sampling obtained with physics-based simulation techniques such as molecular dynamics (MD) simulations, which can be performed under controlled temperature and pressure in explicit solvent.[51,52] Assuming enough sampling performed with an adequate force-field, MD simulations directly generate physically relevant conformational populations, i.e. the most stable conformers are the most populated. Of course, the ability to access the populations offers important insights.[2,43,53–56] Also, physics-based simulations are now routinely performed in explicit solvent, more reliable than the simplified continuum models. The main drawback of MD simulations in explicit solvent is their computational cost, which has been intimidating. That may be why little work has been done to characterize the unbound state of pharmaceutical compounds with MDs,[2] although MDs have proved informative for peptides,[53] nucleosides[56] and cofactors[54,55] in solution.

The recent ability to run MDs much faster in parallel on commodity Graphical Processing Units (GPUs)[52] delivers longer simulations with significantly improved sampling. This work is taking advantage of this breakthrough to simulate the unbound state of 26 diverse drug-like compounds in aqueous solution. Those simulations afforded extensive sampling of the free compounds, providing a wealth of details on their conformational dynamics in solution. This provides a context to examine other common ways to represent the unbound state of compounds, such as the global energy minimum obtained in vacuo (*GlobMin_Vac*), or with *GB* (*GlobMin_GB*). As anticipated, the *GlobMin_Vac* conformers suffer from extensive intramolecular collapse driven by electrostatic and van der Waals interactions. More surprisingly, the *GlobMin_GBs* obtained with the widely used softwares MacroModel[57,58] or MOE[59] are also dominated by collapsed conformers with spurious intramolecular interactions, which artificially lower their intramolecular conformation energy. It follows that those *GlobMin_GBs* are not an adequate representation of the unbound state of the compounds, in particular when investigating their reorganization energy upon binding to a macromolecule. Also, the full set of conformers generated with *GB* can clearly deviate from the MD-simulated ensemble.

The studied compounds were selected to meet a number of criteria, including the availability of a good quality[60] bound bioactive X-ray structure. This allows to compare the MD-simulated unbound state to the bioactive X-ray structures, and examine how conformationally close they may be. We find that some preorganization is not uncommon but varies across compounds, such that general statements on this matter are dangerous.

The compounds were also simulated in the bound state, to estimate their intramolecular reorganization energy from unbound to bound state. To our knowledge, this approach for estimating the reorganization energy has not been attempted before and provides a fresh perspective. The reorganization energy is a component of the binding free energy[10] which cannot be accessed experimentally, and has proved notoriously difficult to determine computationally.[2] This is partly because energies cannot be computed directly on the X-ray coordinates of a bound ligand.[5,9,10,13,17] Also, a robust representation of the reference unbound state has been elusive, since previous work had to reduce the free ligand to its global energy minimum obtained in vacuo,[9] with *GB*,[5,10,13] or approximations in the same vein.[11,17] These technical difficulties have