Bioorganic & Medicinal Chemistry 21 (2013) 7898–7920

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

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

Tackling the conformational sampling of larger flexible compounds and macrocycles in pharmacology and drug discovery

I-Jen Chen *, Nicolas Foloppe *

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

ARTICLE INFO

Article history:
Received 18 August 2013
Revised 29 September 2013
Accepted 4 October 2013
Available online 16 October 2013

Keywords:
Bioactive structures
Conformational sampling
Computational chemistry
Drug discovery
Flexible compounds
Global energy minimum
Low-mode
Macrocycles

ABSTRACT

Computational conformational sampling underpins much of molecular modeling and design in pharmaceutical work. The sampling of smaller drug-like compounds has been an active area of research. However, few studies have tested in details the sampling of larger more flexible compounds, which are also relevant to drug discovery, including therapeutic peptides, macrocycles, and inhibitors of protein–protein interactions. Here, we investigate extensively mainstream conformational sampling methods on three carefully curated compound sets, namely the 'Drug-like', larger 'Flexible', and 'Macrocycle' compounds. These test molecules are chemically diverse with reliable X-ray protein-bound bioactive structures. The compared sampling methods include Stochastic Search and the recent LowModeMD from MOE, all the low-mode based approaches from MacroModel, and MD/LLMOD recently developed for macrocycles. In addition to default settings, key parameters of the sampling protocols were explored. The performance of the computational protocols was assessed via (i) the reproduction of the X-ray bioactive structures, (ii) the size, coverage and diversity of the output conformational ensembles, (iii) the compactness/extendedness of the conformers, and (iv) the ability to locate the global energy minimum. The influence of the stochastic nature of the searches on the results was also examined. Much better results were obtained by adopting search parameters enhanced over the default settings, while maintaining computational tractability. In MOE, the recent LowModeMD emerged as the method of choice. Mixed torsional/low-mode from MacroModel performed as well as LowModeMD, and MD/LLMOD performed well for macrocycles. The low-mode based approaches yielded very encouraging results with the flexible and macrocycle sets. Thus, one can productively tackle the computational conformational search of larger flexible compounds for drug discovery, including macrocycles.

1. Introduction

Conformational sampling underpins much of the computational modeling of the three-dimensional (3D) conformations and properties of organic compounds. In medicinal chemistry and pharmacology, the compound conformations are a critical input for many analyses and molecular design efforts, regarding conformational analysis,[1–3] examination of intramolecular contacts,[4] the effect of conformation on reactivity,[5] the influence of crystal lattices on conformations,[6] fitting of molecular models to the X-ray electron density maps,[7,8] interpretation of NMR data,[9–12] compound overlays,[13,14] pharmacophore elucidation,[15] pharmacophore-based[16] and shape-based[17] virtual screening, docking to a receptor,[18] exploitation of molecular fields,[19] and estimates of the conformational energy of bound ligands.[20–23] These tasks require an ensemble of conformations for every compound, since flexible compounds interconvert between low-energy conformations. Therefore, the conformational sampling of small molecules is a very active area of research.[22,24–39]

Abbreviations: 3D, three-dimensional; *GB*, generalized born; *Diel*, distance-dependent dielectric; GUI, graphical user interface; LMOD, low-mode; LLMOD, large-scale low-mode; LowModeMD, search method combining low-mode moves and molecular dynamics; Max-Iteration, maximum number of search iterations per compound; MD/LLMOD, MD-based simulated annealing followed by large-scale low-mode; MMFF, Merck molecular force field; MOE, molecular operating environment; MT/LMOD, mixed torsional/low-mode; MT/LLMOD, mixed torsional/large-scale low-mode; MW, molecular weight; NbConfs, number of conformers; NRot, number of rotatable bonds; OPLS, optimized potentials for liquid simulations force field; opr_nrot, Oprea number of rotatable bonds; PDB, protein data bank; Rgyr, radius of gyration; RotSteps, number of moves per rotatable bond; *%BioConf_Rep*, percentage of compounds for which the bioactive structure was reproduced; *%GlobMin_found*, percentage of compounds for which the global energy minimum was found.

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

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

0968-0896/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bmc.2013.10.003

To generate conformers, computational methods rely on an energy model and a sampling algorithm which acts on the conformational degrees of freedom. The many methods devised to sample the conformational space and the underlying energy models have been reviewed.[38,39] Search methods typically involve a stochastic element, which may be torsional Monte Carlo moves,[40] random pulses in Cartesian coordinates,[41] seed inter-atomic distances in distance geometry calculations,[24] initial velocities of a molecular dynamics simulation,[42] or the chosen low-modes and seed conformer in low-mode searches.[34,43] Thanks to improved computational resources, algorithmic innovations,[9,34,44,45] and optimization of the search parameters that control a sampling protocol,[30] the conformational sampling of relatively small drug-like compounds with only moderate flexibility has become very powerful.[46]

Thus, it is timely to address the conformational sampling of compounds larger and more flexible than conventionally smaller 'drug-like' compounds. Indeed, a good number of approved drugs are larger and more flexible than required by mainstream drug-likeness criteria;[24,27,47] examples include lipitor, eribulin,[48] and a number of antibiotics.[49] In addition, there is a resurgence of interest in chemotypes which do not fit narrow drug-like prescriptions, such as therapeutic peptides[50] and macrocycles.[27,34,51] Inhibitors of protein–protein interactions also tend to be larger and more flexible since, to gain binding affinity for rather flat and open binding sites, they have to make numerous contacts with the protein.[52] In addition, pharmacology is not only concerned with drug candidates, but also with tool compounds which can be quite flexible. Also, modelling of larger compounds may reveal a key binding fragment core, from which new smaller and more ligand-efficient compounds could be designed. Therefore, conformational sampling of large flexible compounds is of general interest.

Indeed, recent studies have started to investigate the conformational sampling of more flexible compounds, for example with Monte-Carlo torsional sampling,[53] distance geometry[24,27] and the LowModeMD method which combines low-mode and MD sampling.[34] Distance geometry is an interesting approach since it can incorporate heuristics which bias the search towards more extended or more compact conformers, to direct the search across a range of molecular compactness values.[24] It can handle macrocycles.[27] Low-mode based methods are also of special interest for flexible and/or cyclic compounds.[9,34,43,54,55] The principle of low-mode search methods has been explained.[9] Briefly, the low-frequency vibrational mode eigenvectors can be seen as pointing along the low-energy paths connecting energy minima via saddle points on the conformational energy surface. So, moving the coordinates along the low-frequency modes is an efficient way to cross energy barriers between energy minima; once a move following a low-mode eigenvector has located a new energy well, energy minimization is performed and another search cycle is initiated. It has the advantage to be performed in the space of reduced dimensionality of the low-frequency modes, and is well-adapted to cyclic topologies. Low-mode searches can be tuned by controlling how frequently the eigenvectors are re-calculated along the search. To apply low-mode based searches to large systems, at least two approaches have been devised, the Large-scale low mode method (LLMOD),[43] and the recently developed LowModeMD approach[34] implemented in the software MOE[56]. LLMOD generates the eigenvectors without explicitly diagonalizing the entire Hessian.[43] LowModeMD does not calculate the low-frequency modes explicitly, but instead efficiently channels and amplifies atomic motions along directions of low curvature of the potential energy surface,[34] that is along directions similar to those followed by the low-frequency modes. The vibrational motions are imparted via a short molecular dynamics (MD) run at the beginning of every iteration. Initial tests of LowModeMD with macrocycles and protein loops have been encouraging,[34] but it is important to assess further the performance of this new method. Moreover, various low-mode based search schemes have long been implemented in the widely used MacroModel package[57] distributed by Schrödinger.[58] MacroModel offers a plain low-mode search termed Low-mode (LMOD), and the variant Large-scale low-mode (LLMOD) which obtains the eigenvectors in a more approximate manner that is more efficient for larger systems.[43] Two other variants add random torsional moves to the low-mode sampling, and are called Mixed torsional/Low-mode (MT/LMOD) and Mixed torsional/Large-scale low-mode (MT/LLMOD). Schrödinger has also recently proposed a specialized method to explore the conformational space of macrocycles, which starts with a high-temperature MD-based simulated annealing followed by LLMOD sampling; we refer to this method as MD/LLMOD. Thus, low-mode based sampling methods are conceptually well-adapted to the conformational sampling of larger flexible compounds, and are widely available from mainstream software. Also, it is of interest to assess if the recent developments implemented in LowModeMD[34] offer advantages over the earlier incarnations of low-mode based searches,[9,43,54] or over the MOE Stochastic Search.

Low-mode search approaches have been investiagted,[9,24,27,33,34,43,54,59,60] but early studies usually could only handle a small number of compounds, and most studies were essentially performed at the default settings of the program. Yet, conformational sampling protocols depend on many tunable settings which can be adjusted and strongly influence their performance. Such settings include the number of search cycles, the energy window within which the conformers are accepted, how similar to each other the retained conformers can be, and variants on the energy model. Consequently, a more systematic assessment of low-mode search methods would provide guidance regarding best-practices for medicinal chemistry applications and data to compare search algorithms, while clarifying what can be expected of these methods for particularly flexible compounds. Such results may give hints for protein loop modeling as well.

The present work addresses the conformational sampling of particularly flexible compounds, including macrocycles, in comparison to smaller drug-like compounds. This draws on three carefully curated compound sets, called the 'Drug-like set', the 'Flexible set' and the 'Macrocycle set'. The Drug-like set has been presented before,[22,30] but the Flexible and Macrocycle sets were compiled for the present work. Here, the term 'Flexible compound' refers specifically to nonmacrocyclic molecules with at least 12 rotatable bonds, represented by the Flexible set. For every test compound there is at least one good-quality publicly available X-ray structure in complex with a biological macromolecule; the X-ray structure of the bound test compound will be referred to as the bioactive structure. Each set contains a sizable number of compounds (Drug-like: 253, Flexible: 50, Macrocycle: 30), but the emphasis was on selecting the test compounds carefully rather than collating the largest possible sets. Such balance allowed detailed tests of computationally demanding protocols with the Flexible and Macrocycle sets. The investigated sampling methods are comprised of LowModeMD and Stochastic Search implemented in MOE, and LMOD, LLMOD, MT/LMOD, MT/LLMOD and MD/LLMOD in MacroModel. Where applicable, these were explored with three force fields, MMFF,[61] OPLS2005[62] and the recent OPLS2.0.[63] The topic is not high-throughput library generation, but thorough conformational searches for compounds of special interest.

First, the study addresses the ability to 'reproduce' the bioactive X-ray structures, since the performance of computational protocol in that respect with Flexible and Macrocycle compounds was highly uncertain at the outset. Second, we address the conformational coverage, via the number of generated conformers (NbConfs), their compactness/extendedness, and the number of 3D pharmacophores they visit. Since considerable sampling was