Journal of Molecular Graphics and Modelling 31 (2011) 5–9

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

HBonanza: A computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis

Jacob D. Durrant[a,*], J. Andrew McCammon[b,c,d]

[a] Department of Chemistry & Biochemistry, University of California San Diego, 9500 Gilman Drive, Mail Code 0365, La Jolla, CA 92093-0365, United States
[b] Department of Chemistry & Biochemistry, NSF Center for Theoretical Biological Physics, National Biomedical Computation Resource, University of California San Diego, La Jolla, CA 92093, United States
[c] Department of Pharmacology, University of California San Diego, La Jolla, CA 92093, United States
[d] Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA, United States

ARTICLE INFO

Article history:
Received 18 June 2011
Received in revised form 27 July 2011
Accepted 31 July 2011
Available online 7 August 2011

Keywords:
Hydrogen bond
Computational chemistry
Molecular dynamics simulations
Molecular recognition
Computer-aided drug design

ABSTRACT

In the current work, we present a hydrogen-bond analysis of 2673 ligand–receptor complexes that suggests the total number of hydrogen bonds formed between a ligand and its receptor is a poor predictor of ligand potency; furthermore, even that poor prediction does not suggest a statistically significant correlation between hydrogen-bond formation and potency. While we are not the first to suggest that hydrogen bonds on average do not generally contribute to ligand binding affinities, this additional evidence is nevertheless interesting. The primary role of hydrogen bonds may instead be to ensure specificity, to correctly position the ligand within the active site, and to hold the protein active site in a ligand-friendly conformation.

We also present a new computer program called HBonanza (hydrogen-bond analyzer) that aids the analysis and visualization of hydrogen-bond networks. HBonanza, which can be used to analyze single structures or the many structures of a molecular dynamics trajectory, is open source and python implemented, making it easily editable, customizable, and platform independent. Unlike many other freely available hydrogen-bond analysis tools, HBonanza provides not only a text-based table describing the hydrogen-bond network, but also a Tcl script to facilitate visualization in VMD, a popular molecular visualization program. Visualization in other programs is also possible. A copy of HBonanza can be obtained free of charge from http://www.nbcr.net/hbonanza.

1. Introduction

Hydrogen bonds, first described in 1912 [1], involve two components: a hydrogen-bond donor, comprised of an electronegative atom covalently bound to a hydrogen-atom partner, and a hydrogen-bond acceptor, also an electronegative heteroatom. The donor heavy atom attracts the electrons normally associated with its hydrogen-atom partner, imparting to that partner a positive partial charge. This positive partial charge can interact with a lone pair of electrons on the hydrogen-bond acceptor, forming a bond that is part electrostatic and part chemical [2]. When binding a ligand to its receptor, a hydrogen bond is traditionally thought to contribute between 0.5 and 4.7 kcal/mol to the binding energy [3].

Despite this traditional wisdom, some have speculated that the hydrogen-bond contribution to ligand–receptor binding energies is generally overstated, certain convincing exceptions aside [4,5]. Ligand hydrogen-bond donors and acceptors are energetically favorable both in solution, where they form hydrogen bonds with the surrounding water molecules, and when bound to an amenable active site, where they form hydrogen bonds with receptor residues. Thus, ligand binding simply substitutes hydrogen bonds with water for those with the receptor. There is arguably little net gain in energy; binding is merely an exchange process [3].

In the current work, we present an analysis of 2673 ligand–receptor complexes of known binding affinity that suggests hydrogen bonds do not generally contribute to ligand binding energies. Their primary role in ligand binding may instead be to ensure specificity; to correctly position the ligand within the active site, thus facilitating catalysis; and to hold the protein active site in a conformation amenable to ligand binding.

One way to study these important roles is through structural/computational biology. A common task is to perform a hydrogen-bond analysis of an entire molecular-dynamics (MD) trajectory, often consisting of thousands of individual protein conformations. Such an analysis can give a sense of the persistence of selected hydrogen bonds across the many MD frames, providing

* Corresponding author. Tel.: +1 858 822 0169; fax: +1 858 534 4974.
E-mail address: jdurrant@ucsd.edu (J.D. Durrant).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.07.008

insight into their relative importance. In our experience, however, performing hydrogen-bond analyses of molecular-dynamics trajectories with existing free software (e.g., [6,7]) can be tedious. Most of the available free programs produce large text-based charts; the important hydrogen-bond interactions can be easily buried in the data.

In response to this challenge, we created HBonanza (hydrogen-bond analyzer), an open-source, python-implemented computer program that, given an atomistic molecular structure, can greatly facilitate the identification and analysis of hydrogen-bond networks. HBonanza can be used to analyze single protein structures or entire molecular-dynamics trajectories. While all hydrogen bonds can certainly be identified, the program also facilitates the identification of only those hydrogen-bond networks connected to a selected residue or residues of interest (e.g., the ligands). Importantly, HBonanza outputs a Tcl file that creates an intuitive visualization when loaded into VMD [8], a popular molecular visualization program. Visualization using other programs is also possible.

HBonanza has been tested on Ubuntu 10.04.1 LTS, Mac OS X 10.6.6, and Windows XP using Python versions 2.6.5, 2.6.1, and 2.7.1, respectively. A copy can be obtained free of charge from http://www.nbcr.net/hbonanza.

2. Materials and methods

2.1. A database of experimentally characterized ligand–receptor complexes

The database of ligand–receptor complexes with experimentally measured binding affinities used in the current work has been described previously [9,10]. In brief, structures with K_d values listed in the PDBbind-CN [11,12] and MOAD [13] databases were downloaded from the Protein Data Bank [14]. Hydrogen atoms were added to all ligands and receptors using the Schrodinger Maestro (Schrodinger) and AutoDockTools 1.5.1 [15] programs. An in-house script was used to optimize the geometry of the hydrogen bonds between ligand and receptor atoms, and the computer program BINANA [9] was used to identify the number of hydrogen bonds formed between the receptor and ligand for each complex. Default BINANA hydrogen-bond parameters were used.

2.2. A molecular dynamics simulation of RNA editing ligase 1

A molecular model of *Trypanosoma brucei* RNA editing ligase 1 bound to **V2**, a recently discovered low-micromolar inhibitor [16], was generated from the 1XDN structure [17] deposited in the Protein Data Bank [14]. The ligand, prepared using Schrodinger Maestro (Schrodinger) and AutoDockTools [15], was docked into the 1XDN active site *sans* the co-crystallized ATP ligand using AutoDock Vina [18]. The docked pose with the lowest predicted binding energy was accepted.

Hydrogen atoms were added to the protein receptor using the PDB2PQR server [19,20], and the receptor was parameterized according to the AMBER-99SB force field [21]. Ligand partial charges were assigned using the Gaussian computer program [22] and the RESP module of the Amber molecular-dynamics package [6]. The molecular electrostatic potential was calculated at the HF/6-31G* level. Ligand atom types were assigned according to the GAFF force field [23]. GAFF does not have parameters for the angle between atom types nh-c3-ca; the angle parameters associated with nh-c3-c3 were used instead.

A box of TIP3P water molecules [24] was built around the protein/ligand model, extending 10 Å beyond the model in all directions, using the xleap module of the Amber molecular-dynamics package. Sufficient Na^+ cations were added to bring the system to electrical neutrality; additional Na^+ and Cl^- ions were added to approximate 20 mM NaCl solution.

NAMD 2.7b1 [25] was used to minimize the system in four stages. First, the hydrogen atoms of the system were minimized for 5000 steps. Second, the hydrogen atoms, water molecules, and ions were minimized for 5000 steps. Third, the hydrogen atoms, water molecules, ions, and protein side chains were minimized for 10,000 steps. Finally, all the atoms of the system were minimized for an additional 25,000 steps. The system was next equilibrated by gradually heating in four stages, beginning at 50 °K and ending at 310 °K. The simulation was run at each temperature for 250,000 steps with a time step of 1.0 fs. Bonds to hydrogen atoms were held rigid, and Langevin dynamics were performed to maintain constant pressure and temperature.

Following minimization and equilibration, the isothermal-isobaric simulation was continued for an additional 10,000,000 steps using a 2.0 fs time step. The system was saved every 5000 steps, yielding 2000 distinct protein conformations that were used for subsequent analysis.

2.3. Using HBonanza to analyze the molecular dynamics trajectory

Default HBonanza values were used for trajectory analysis. Hydrogen bonds were considered to be those that satisfied the following criteria: (1) The heavy-atom participants in the candidate bond were oxygen, nitrogen, fluorine, or sulfur atoms. (2) These heavy atoms were no farther than 3.5 Å apart. (3) The angle between the hydrogen atom, the donor heavy atom, and the acceptor heavy atom was less than 30°. (4) The hydrogen bond as defined above was present in at least 75% of the frames of the trajectory.

3. Results and discussion

The purpose of the current work is two fold. First, we present a small analysis of 2673 ligand–receptor complexes with known binding affinities in order to elucidate the general contribution of hydrogen bonds to ligand binding energies. Second, we describe in detail a novel computer program called HBonanza (hydrogen-bond analyzer) that has been developed to facilitate the hydrogen-bond analyses of single structures as well as molecular-dynamics trajectories.

3.1. The hydrogen-bond contribution to ligand binding energies

We recently described a large structural database of 2673 ligand–receptor complexes with known binding affinities [9,10]. As the affinities and structures of these complexes have been well characterized experimentally, it is possible to compare their measured K_d values to various binding characteristics. As an example, we used a computer program called BINANA to see how the number of ligand–receptor hydrophobic contacts varied with ligand potency [9].

We here present a similar analysis of ligand–receptor hydrogen bonds. First, the experimentally measured pK_d value of each of the 2673 complexes was plotted against the BINANA-determined number of ligand–receptor hydrogen bonds (Fig. 1). In order to determine if these two variables were correlated, linear regression was used to identify a single line that best fit the data. The R^2 value of the regression was 2.7×10^{-4}, suggesting that the number of hydrogen bonds is a poor predictor of ligand potency if a linear relationship is assumed.

The slope, F statistic, and number of degrees of freedom for the regression were -2.9×10^{-2}, 0.72, and 2671, respectively. If there were no linear correlation between ligand potency and the

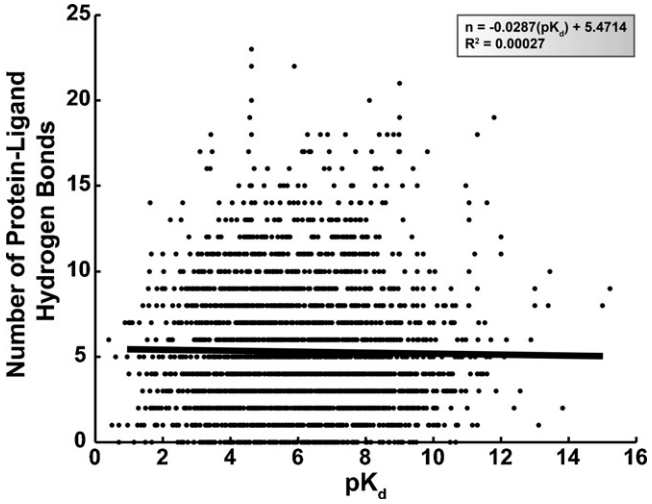

Fig. 1. The experimentally measured pK_d values of 2673 ligand–receptor complexes plotted against the corresponding number of ligand–receptor hydrogen bonds. A linear regression line is shown in black; the equation of this line and the R^2 value of the fit are given in the legend.

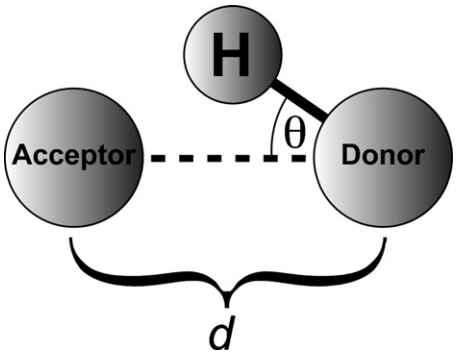

Fig. 2. A schematic of a hydrogen bond. For the purposes of the current work, d is defined as the distance between the donor and acceptor heavy atoms, and θ is defined as the angle formed by the hydrogen atom, the donor heavy atom, and the acceptor heavy atom.

number of hydrogen bonds, one would expect a regression with a slope of 0, a model with only 1 degree of freedom. We accepted as the null hypothesis that the regression model generated from the 2673 data points does not differ significantly from the linear model of slope 0. This null hypothesis was tested using an F probability distribution; the corresponding p-value was 0.40, suggesting the null hypothesis should not be rejected. Thus, not only is the hydrogen-bond count a poor predictor of ligand potency (based on the R^2 value of the linear regression), but even that poor prediction does not suggest a statistically significant correlation between these two variables. We are not the first to suggest that hydrogen bonds do not generally contribute to binding energies, but this additional evidence is nevertheless interesting.

Some may wonder if our database of ligand–receptor complexes is biased by the presence of drug-like molecules. Among the complexes with high potency, are we capturing the characteristics of good drugs rather than the general characteristics of good binders? After all, many drugs conform to Lipinski's rule [26], which includes constraints on the number of ligand hydrogen-bond donors and acceptors. Additionally, drugs may also have reduced numbers of hydrogen-bond-forming hydroxyl and amino groups in order to avoid phase-II conjugation and to improve cell-membrane and blood–brain barrier permeability.

In order to address this concern, we searched DrugBank [27] for each of the ligands in our database and determined that only 5.4% (144 ligands) are currently approved drugs. A t-test revealed that the approved drugs are significantly more likely ($p = 3.5 \times 10^{-3}$) to form fewer hydrogen bonds with their respective receptors than ligands that are not approved drugs, but the difference is not great, averaging 4.6 ± 2.9 (standard deviation) vs. 5.3 ± 3.8. When the pK_d values of the approved drugs are compared to the pK_d values of the remaining ligands in the database, the two groups do not differ significantly ($p = 0.46$), suggesting that the approved drugs do not tend to be more or less potent than the remaining ligands.

3.2. HBonanza: a novel computer program for characterizing hydrogen bonds

In the analysis above, the computer program BINANA [9] was used to identify ligand–receptor hydrogen bonds. BINANA is well suited for counting the number of hydrogen bonds in a single structure and producing a visualization of those bonds in VMD [8], but it is not well suited for identifying and tallying the hydrogen bonds across an entire molecular-dynamics trajectory, which contains multiple frames with subtly different hydrogen-bond networks. In order to analyze and visualize the hydrogen-bond networks of both trajectory and single-structure PDB files, we have created a new computer program called HBonanza.

HBonanza considers a number of factors when identifying hydrogen bonds. First, the heavy-atom participants in the potential hydrogen bond must be oxygen, nitrogen, fluorine, or sulfur atoms. Second, the distance between these heavy atoms (d) must be no greater than a user-defined cutoff distance, 3.5 Å by default (Fig. 2). Third, the angle between the hydrogen atom, the donor heavy atom, and the acceptor heavy atom (θ) must be less than a user-specified angle, 30° by default (Fig. 2). Finally, if an entire MD trajectory is analyzed, the hydrogen bonds as defined above must be well represented among the many frames of the simulation. If a hydrogen bond appears only infrequently, it is less likely to play an important role in ligand binding or protein structure. The required frequency cutoff is user defined; by default, it is set to 75%.

Though it is certainly possible to analyze all of the receptor/ligand hydrogen bonds of a given trajectory, HBonanza may be most useful when one wishes to examine the specific hydrogen-bond network relevant to a residue or residues of interest (i.e., a "seed" residue, often the ligand). HBonanza identifies relevant networks recursively. First, all residues that are connected to the seed residue(s) *via* hydrogen bonds are identified. Next, all residues that are connected to those that are connected to the seed residues are identified, then all residues that are connected to those, etc., thereby delineating the entire hydrogen-bond network that supports the seed residue(s) of interest.

Perhaps most helpfully, HBonanza not only prints out a text-based chart describing these hydrogen bonds but also builds a visualizing Tcl script that can be loaded directly into VMD [8], a popular molecular visualization program. The identified hydrogen bonds are colored according to their frequency, facilitating the identification of the most important bonds (see, for example, Fig. 3). For those who do not use VMD, the generated Tcl script contains helpful comments that should make visualization in other software packages easy as well.

3.3. An example: the hydrogen-bond networks of TbREL1

In order to demonstrate the utility of HBonanza, we analyzed the hydrogen-bond networks of RNA editing ligase 1 from *T. brucei* (*Tb*REL1), a recently discovered therapeutic target useful in the fight against African sleeping sickness. First, a 20-ns molecular

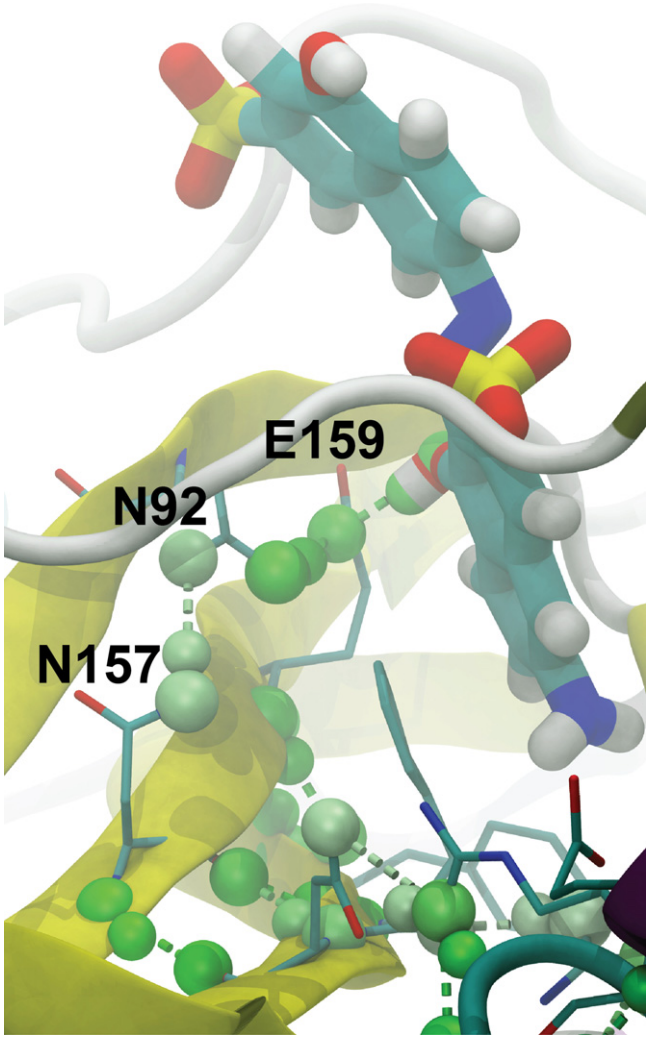

Fig. 3. Predicted hydrogen-bond interactions with **V2**, a recently identified low-micromolar inhibitor of *Tb*REL1. The visualization shown is representative of the default HBonanza-generated VMD visualization script, rendered with Tachyon. The ligand is shown in thick licorice, protein residues thought to participate in hydrogen-bond networks involving the ligand are shown in thin licorice, and the entire protein is shown in cartoon, colored according to secondary structure. Hydrogen bonds are colored in shades of green according to the frequency with which they appear in the many frames of the MD trajectory. The coloring ranges from white, those bonds present in 75% of the frames, to green, those bonds present in 100% of the frames.

dynamics simulation was performed of the protein with a bound, recently discovered inhibitor (**V2**) [16] whose predicted binding pose was determined using docking. The hydrogen-bond networks that interfaced with the ligand were subsequently identified using HBonanza. The HBonanza output file, visualized in VMD and rendered using Tachyon, is shown in Fig. 3.

The analysis revealed that only one hydrogen bond forms between the protein and ligand in more than 75% of the frames of the trajectory. This bond connects a key ligand hydroxyl group to the E159 side-chain carboxylate, which in turn is held in place by a hydrogen bond with the N92 side-chain amine, which is in turn stabilized by a hydrogen bond with the N157 side-chain amine, which may ultimately tie the region of the active-site that interacts with the ligand hydroxyl group to the stability of the underlying beta sheet. Of course it is possible to take this kind of analysis too far; residues that are separated from the ligand by many degrees are not likely to be critical to binding. Nevertheless, HBonanza analyses offer interesting insights into the role distant residues play in stabilizing residues of special interest.

4. Conclusion

The current work is two fold. First, we presented an analysis of 2673 ligand–receptor complexes with known binding affinities. We found that the total number of ligand–receptor hydrogen bonds is a poor predictor of ligand potency; furthermore, even that poor prediction does not suggest a statistically significant correlation between these two variables. While we are not the first to suggest that hydrogen bonds do not generally contribute to ligand binding energies, the additional evidence presented in the current work is nevertheless interesting.

Second, we also presented a new computer program called HBonanza (hydrogen-bond analyzer) that, given either a single structure or an entire molecular dynamics trajectory in PDB format, facilitates the identification, analysis, and visualization of hydrogen-bond networks. HBonanza is open source and python implemented, making it easily editable, customizable, and platform independent. Unlike many other freely available hydrogen-bond analysis tools, HBonanza generates not only a text-based table describing the hydrogen-bond network, but also a Tcl script to facilitate visualization in VMD [8], a popular molecular visualization program. Visualization in other programs is also possible. A copy of HBonanza can be obtained free of charge from http://www.nbcr.net/hbonanza.

Acknowledgements

This work was carried out with funding from NIH GM31749, NSF NSF MCB-1020765, and MCA93S013 to JAM. The funding sources had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. Additional support from the Howard Hughes Medical Institute, the National Center for Supercomputing Applications, the San Diego Supercomputer Center, the W.M. Keck Foundation, the National Biomedical Computational Resource, and the Center for Theoretical Biological Physics is gratefully acknowledged. We would also like to thank Aaron Friedman for helpful discussions.

References

[1] T.S. Moore, T.F. Winmill, The state of amines in aqueous solution, J. Chem. Soc. Trans. 101 (1912) 1635.
[2] E. Arunan, G.R. Desiraju, R.A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, et al., Defining the Hydrogen Bond: An Account, IUPAC, 2011.
[3] A.M. Davis, S.J. Teague, Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis, Angew. Chem. Int. Ed. 38 (1999) 736.
[4] J. DeChancie, K.N. Houk, The origins of femtomolar protein–ligand binding: hydrogen-bond cooperativity and desolvation energetics in the biotin-(strept)avidin binding site, J. Am. Chem. Soc. 129 (2007) 5419–5429.
[5] K.A. Armstrong, B. Tidor, A.C. Cheng, Optimal charges in lead progression: a structure-based neuraminidase case study, J. Med. Chem. 49 (2006) 2470–2477.
[6] D.A. Case, T.E. Cheatham 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., et al., The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.
[7] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.
[8] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[9] J.D. Durrant, J.A. McCammon, BINANA: a novel algorithm for ligand-binding characterization, J. Mol. Graph. Model. 29 (2011) 888–893.
[10] J.D. Durrant, J.A. McCammon, NNScore: a neural-network-based scoring function for the characterization of protein–ligand complexes, J. Chem. Inf. Model. 50 (2010) 1865–1871.
[11] R. Wang, X. Fang, Y. Lu, S. Wang, The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures, J. Med. Chem. 47 (2004) 2977–2980.
[12] R. Wang, X. Fang, Y. Lu, C.Y. Yang, S. Wang, The PDBbind database: methodologies and updates, J. Med. Chem. 48 (2005) 4111–4119.
[13] L. Hu, M.L. Benson, R.D. Smith, M.G. Lerner, H.A. Carlson, Binding MOAD (mother of all databases), Proteins 60 (2005) 333–340.
[14] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, et al., The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.

[15] M.F. Sanner, Python: a programming language for software integration and development, J. Mol. Graph. Model. 17 (1999) 57–61.
[16] J.D. Durrant, L. Hall, R.V. Swift, M. Landon, A. Schnaufer, R.E. Amaro, Novel naphthalene-based inhibitors of *Trypanosoma brucei* RNA editing ligase 1, PLoS Negl. Trop. Dis. 4 (2010) e803.
[17] J. Deng, A. Schnaufer, R. Salavati, K.D. Stuart, W.G. Hol, High resolution crystal structure of a key editosome enzyme from *Trypanosoma brucei*: RNA editing ligase 1, J. Mol. Biol. 343 (2004) 601–613.
[18] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2009) 455–461.
[19] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, et al., PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, Nucleic Acids Res. 35 (2007) W522–W525.
[20] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, Nucleic Acids Res. 32 (2004) W665–W667.
[21] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins 65 (2006) 712–725.
[22] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, et al., Gaussian 03, Gaussian Inc., Wallingford, CT, 2004.
[23] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174.
[24] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926.
[25] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, et al., Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.
[26] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 46 (2001) 3–26.
[27] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, et al., DrugBank: a comprehensive resource for in silico drug discovery and exploration, Nucleic Acids Res. 34 (2006) D668–D672.