Chapter 2

Computational Modeling of Small Molecule Ligand Binding Interactions and Affinities

Marino Convertino and Nikolay V. Dokholyan

Abstract

Understanding and controlling biological phenomena via structure-based drug screening efforts often critically rely on accurate description of protein–ligand interactions. However, most of the currently available computational techniques are affected by severe deficiencies in both protein and ligand conformational sampling as well as in the scoring of the obtained docking solutions. To overcome these limitations, we have recently developed MedusaDock, a novel docking methodology, which simultaneously models ligand and receptor flexibility. Coupled with MedusaScore, a physical force field-based scoring function that accounts for the protein–ligand interaction energy, MedusaDock, has reported the highest success rate in the CSAR 2011 exercise. Here, we present a standard computational protocol to evaluate the binding properties of the two enantiomers of the non-selective β-blocker propanolol in the β2 adrenergic receptor's binding site. We describe details of our protocol, which have been successfully applied to several other targets.

Key words Flexible docking, MedusaDock, MedusaScore, Induced Fit, Gaia, Chiron, Protein–ligand interactions, Protein structure refinement

1 Introduction

The interactions between small molecules or small peptides and protein targets are at the basis of many biological processes; therefore, the scientific community has been very prolific in developing algorithms, protocols, and methodologies to describe, understand, and control the process of recognition and formation of protein–ligand and protein–peptide complexes [1–5]. The ability to elucidate the pharmacodynamical properties of low molecular weight compounds or small peptides, along with the possibility of rationally designing novel drugs, relies on the accurate prediction of atomic interactions between ligands and target proteins. However, the ligands' large number of degrees of freedom and proteins' backbone and side chains flexibility present a critical challenge for an effective computational description of the ligand–receptor

interaction (i.e., docking calculations) [6–8]. Modeling the induced fit phenomenon, whereby both the target and the ligand undergo mutually adaptive conformational changes upon binding, is particularly demanding due to significant conformational sampling required for computational optimization of such interactions [8–10]. In order to properly account for this effect, experimentally (via X-ray crystallography or NMR spectroscopy) and/or computationally (via molecular dynamics or normal mode analysis) determined protein conformations have been included in current docking calculations [11–15]. However, multiple conformations of the protein may not be available, or be biased toward the protein–ligand complex conformations, and, thus not able to capture new rearrangements of protein binding sites upon binding of novel compounds.

To overcome these limitations, we have recently developed a new docking algorithm, namely MedusaDock [16], which accounts for ligand and receptor flexibility at the same time. In MedusaDock, we build a stochastic rotamer library for each ligand, and simultaneously model the protein sidechain conformation using a rotamer library for all natural amino acids. The efficient sampling of our docking is associated with the use of MedusaScore [17], a physical force field-based scoring function accounting for the protein–ligand interaction energy. The adoption of MedusaScore circumvents the problem of low transferability among different targets and ligands, which is typical of empirical scoring functions classically used in docking calculations [18, 19]. MedusaDock and MedusaScore have been successfully adopted in the evaluation of the binding properties of both peptides [5] and small molecules [16, 20, 21].

Our docking approach has successfully predicted the native conformations of 28 out of the 35 study cases proposed in the recent CSAR-2011 competition [20], more than any other group in the exercise (H. Carlson, personal communications). In this chapter, we present a standard protocol to perform the docking of the propanolol enantiomers in the binding site of the β2 adrenergic receptor (β2AR). We (1) assess the structural quality of this G protein-coupled receptor's structure using our in-house developed software Gaia, which compares the intrinsic properties of protein structural models to high-resolution crystal structures (http://chiron.dokhlab.org [22]); (2) generate the optimized starting structures of ligands using widely used molecular modeling tools; and finally (3) calibrate and run docking calculations using MedusaDock [16], which will eliminate any possible bias originated from the starting conformations of the amino acids in β2AR binding pockets.

2   Materials

To implement the reported docking calculation procedure, it is necessary to have access to an internet-connected computer running a Linux operative system and mount a licensed copy of the Schröedinger Suite (Schröedinger, LLC), as well as a licensed copy of the MedusaDock software (Molecules in Action, LLC).

3   Methods

3.1   Protein Preparation

1. Navigate through the Protein Data Bank (PDB) website [23] to download the crystallographic coordinates of the human β2AR at 2.8 Å resolution (PDB-ID: 3NY8 [24]). From the downloaded file, remove the coordinates of (1) the co-crystallized inverse agonist ICI 118,551; (2) water molecules not mediating the binding of ICI 118,551 to β2AR; and (3) molecules used for technical purposes and present in the final crystal structure.

2. In order to estimate the quality of the resulting β2AR protein structure, run the in-house developed software Gaia [22]. Navigate to the following address http://chiron.dokhlab.org. Click on the Submit Task button in the starting page (Fig. 1a). In the **step 1** section, enter a Job Title in the dedicated window, and upload the file containing the β2AR crystallographic structure in pdb format. You can choose to receive an e-mail notification when the submitted job is completed. In the **step 2** section, choose the task Gaia to validate the submitted protein structure. The status of the calculation can be monitored via the panel Gaia, which is accessible by clicking the Home/Overview button in the starting page (Fig. 1a). Upon completion of the job (indicated by a green mark in the Status), a short report of some protein features will be presented on the web page (Fig. 1b). The user can download a detailed report on the structural features of the protein clicking on the eye icon in the table (Fig. 1b, *see* **Note 1**).

3.2   Ligand Preparation

1. Several applications can be used to prepare the structure of ligands to be used in docking calculations. In this specific case, we will use a number of applications available via the Schrödinger Suite. Starting from the Maestro interface (v. 9.3.5), use the 2D Sketcher tool to draw the chemical structures of the inverse agonist ICI 118,551, co-crystallized with the β2AR protein, as well as the two propanolol enantiomers, whose binding modes will be investigated through docking.

a

Χειρων | Protein Structure Refinement   Dokholyan Group

Logged in as **mconvert** [Logout]

Home/Overview	**Submit a task**
Submit Task	Dear guests,
User Profile	We have recently migrated to a newer, faster server to ensure shorter turn-around times for submitted jobs. If you are a new user, you may not notice a difference. If you are returning user, you may notice that some things have changed or have become better. If you face problems with any of the features, please let us know and we will be glad to work with you.
Documentation	Chiron team
Contact Us	

Step 1 : Enter task parameters

	Task Parameters
Job Title ? :	
Input Type :	⦿ PDB ID ○ File
PDB ID :	
Constrain Sidechain ? :	☐
E-mail Notification :	☐

Step 2 : Choose relevant task

	Choose Task
▸ Chiron	
▸ Gaia	

If you are using Chiron for the first time, we recommend that you read the documentation before submitting your job.

b

Χειρων | Protein Structure Refinement   Dokholyan Group

Logged in as **mconvert** [Logout]

Home/Overview	**Overview**
Submit Task	
User Profile	**Chiron** **Gaia**
Documentation	**You have no recent submissions.**
Contact Us	**Your recent results.**

⏳ – Waiting   ↻ – Running   ✓ – Done   ⬤ – Failed

HBS – Hydrogen bonds in the shell   HBC – Hydrogen bonds in the core
MSA – Molecular Surface Area   Voids – Internal cavities only

					Scores				
Job ID	Title	Protein	Clashes	HBS	HBC	MSA	Voids	Status	Action
11910		Custom	0.0172	7.6	0.5	152.208	0.647	✓	👁 🗑
Job ID	Job Title	PDB ID		Submitted at			Status		Action

c

Table 1: Summary of scores for the input structure

Criterion		Observed	Target
Steric clashes		0.02	0.02
Hydrogen bonds	%Unsatisfied in shell	7.60	9.56
	%Unsatisfied in core	0.50	1.45
Solvent accessible surface area		243.36	221.64
Void volume		0.65	0.097

Fig. 1 (**a**) Home page of Chiron/Gaia server for protein structure refinement, which is available at the following link: http://chiron.dokhlab.org. (**b**) Short report of protein's structural features from the Chiron/Gaia server.

2. The ligand structures need to be further optimized using the LigPrep application. The user can choose the appropriate force field (in this case MMFFs [25]) for the optimization of atom distances, angles, and dihedral angles, along with the most appropriate pH for the determination of the formal charges of titratable groups (*see* **Note 2**). Several options are available for the determination of the ligands' stereochemistry. Since we have manually drawn the ligand structures, we determine the appropriate chiralities from the generated 3D structures without constructing any tautomers. The optimized structures of ligands are saved in mol2 format for docking calculations, and in Structure Data Format (i.e., SDF format by MDL Information Systems) for storage.

3.3 Docking Calibration

1. Docking calculations are executed via our Monte Carlo-based algorithm MedusaDock [16], which simultaneously accounts for ligands' and receptors' (side chains) flexibility. We calibrate docking calculations to the target protein by performing a self-docking of any co-crystallized binder as retrieved from the PDB to assess both the convergence of docking calculations, and the ability of reproducing the native pose of the co-crystallized ligand (i.e., ICI 118,551) in the β2AR binding site.

2. In order to test the convergence of docking results, submit several independent docking calculations of ICI 118,551 in the β2AR binding site (e.g., 100, 200, 500) using MedusaDock [16] (*see* **Note 3**), and plot the distributions of the binding energies as estimated by MedusaScore [17] (Fig. 2a). The number of calculations by which there is no more variation of the poses' binding energy distributions will be the minimal number of docking runs normally submitted to explore the binding modes of compounds (with similar molecular weight and rotatable bonds to ICI 118,551) in the β2AR binding site.

3. The estimated binding energies for all of the docking poses of ICI 118,551 (as for any docked compound) show a normal distribution (Fig. 2b). Therefore, according to the central limit theorem [26], it is possible to retrieve as statistical significant solutions from only those docking poses for which the

---

Fig. 1 (continued) The green mark below the Status column indicates the completion of the job; the eye icon in the table gives access to a detailed report, which can be downloaded in pdf format. (**c**) Initial summary about protein's structural features as downloaded from the Chiron/Gaia server. Values highlighted in red usually need the user attention in order to further refine the submitted protein structure (*see* **Note 1**). A detailed report about steric clashes, hydrogen bonds in the shell and in the core of the protein, solvent accessible surface area, and void volume is also available to the user
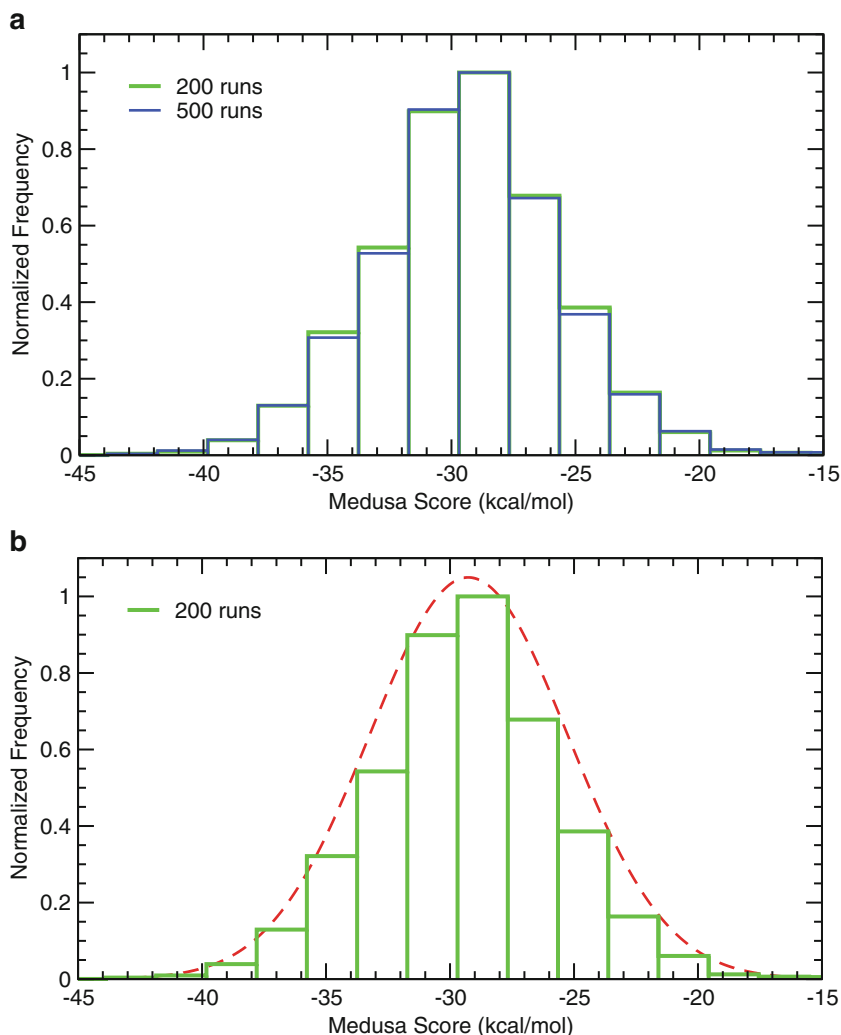
Fig. 2 (**a**) Convergence of the distributions of docking pose's binding energies extracted from 200 and 500 independent MedusaDock calculations are reported in *green* and *blue*, respectively. (**b**) Normal distribution (*red dashed curve*) of docking pose's binding energies extracted from 200 independent MedusaDock calculations (*green bars*)

Z-score is lower than –2 (i.e., less than 5 % probability that the specific docking pose is extracted by chance). In this case, *Z* is defined as:

$$Z = \frac{x - \mu}{\sigma}$$

where x is the estimated binding energy of a specific docking poses, and μ and σ are the mean and the standard deviation of the binding energies in the population of binding poses, respectively.

4. On the subset of extracted docking poses (i.e., poses with
   Z-score lower than −2), perform a cluster analysis to retrieve
   the most representative docking pose (i.e., centroid of the most
   populated cluster of poses). Cluster the ensemble of docking
   solutions according to the root mean square deviation (RMSD)
   computed over the ligand's heavy atoms. The optimal number
   of highly populated clusters can be identified by applying the
   average linkage method [27] and the Kelley penalty index [28]
   in order to minimize the number of clusters and the spread of
   internal values in each cluster. The clustering level with the low-
   est Kelley penalty represents a condition where the clusters are
   highly populated and concurrently maintain the smallest inter-
   nal spread of RMSD values (*see* **Note 4**). The centroid of the
   most populated cluster is chosen as the representative confor-
   mation of the ICI 118,551 bound to β2AR.

5. Calculate the RMSD of the extracted solution of ICI 118,551
   with respect to the original co-crystallized conformation of the
   ligand in β2AR. The RMSD computed over the ligand's heavy
   atoms (1.4 Å) is below the X-ray resolution (2.8 Å). Therefore,
   the applied strategy is successful in reproducing the native pose
   of ICI 118,551 as also demonstrated by the consistency with
   the electron-density map of the crystal as downloaded from
   the Uppsala Electron Density Server [29] (Fig. 3a).

3.4 Docking Calculations for Propanolol Enantiomers

1. Using MedusaDock submit the number of independent dock-
   ing calculations determined in the **step 2** of docking calibra-
   tion (*see* **Note 5**).

2. Isolate, cluster, and retrieve the obtained docking poses of pro-
   panolol enantiomers (Fig. 3b) as described in the **steps 3–5** of
   docking calibration.

4 Notes

1. Starting from Gaia panel in the Home/Overview page (Fig.
   1b), the user can download a detailed report of the structural
   properties of the submitted protein in comparison with what
   observed in high-resolution crystal structures. The initial sum-
   mary is reported in Fig. 1c. Values highlighted in red usually
   need the user attention in order to further refine the submitted
   protein structure. Such operation can be performed using the
   software Chiron [30], which minimizes the number of non-
   physical atom interactions (clashes) in the given protein
   structure.

2. The user can choose several options for the ligands' optimiza-
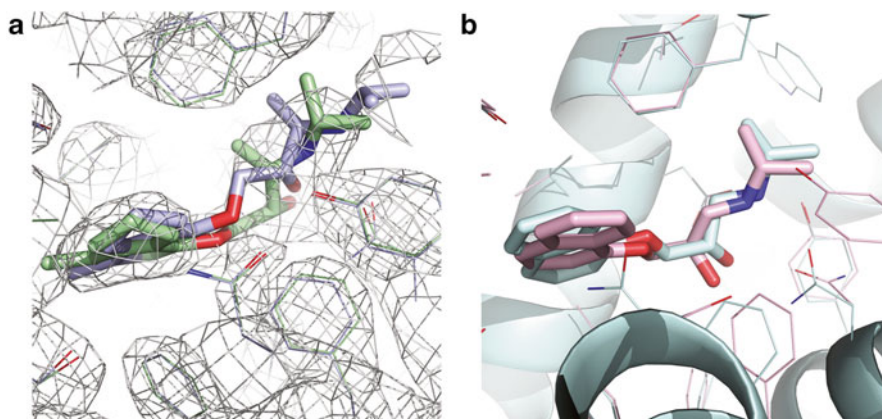   tion. Available force fields are MMFFs [25] or OPLS_2005

Fig. 3 (**a**) Superimposition of MedusaDock docking solution of ICI 118,551 to its crystallographic conformation in the β2AR binding site (PDB-ID: 3NY8). The described docking procedure demonstrates high reliability as it reproduces the binding pose of the original co-crystallized molecule with a RMSD computed over the ligand's heavy atoms of 1.4 Å, which is below the X-ray resolution (2.8 Å). The binding energy as estimated by MedusaDock is −39.4 kcal/mol and −37.9 kcal/mol for ICI 118,551 in its docked and crystallized conformation, respectively. Carbon atoms are represented in blue and green for ICI 118,551 in its docked and crystallized conformation, respectively. β2AR electron density map available from the Electron Density Server is reported as white mesh. (**b**) R/S propanolol bound conformations obtained by combining the MedusaScore values with a hierarchical cluster analysis of statistically significant docking solutions (i.e., poses with *Z*-score lower than −2, main text). The binding energy as estimated by MedusaScore is −38.1 kcal/mol and −38.8 kcal/mol for R- and S-propanolol, respectively. The reported solutions represent the centroids of the most populated clusters of statistically significant docking poses of R- and S-propanolol (i.e., 61.5 % and 57.7 % of the conformational ensembles, respectively). Carbon atoms are represented in pink and cyan for R- and S-enantiomers, respectively. The same color code is adopted to indicate the sidechains of β2AR amino acids when in complex with the two enantiomers

[31, 32]. The ionization state of titratable groups can be refined at the appropriate pH (the user should retrieve any available information about the pH value at the protein binding site) using either the Epik or the Ionizer application. The user can also decide to generate tautomers or all possible combinations of stereoisomers for each optimized ligand.

3. MedusaDock command can be submitted in a machine running a Linux operating system using the following command:

```
$>  ./medusaDock.linux  -i  TARGET_PROTEIN  -m
MOLECULE_TO_DOCK -o DOCKING_SOLUTION -p ./ MEDUSADOCK_
PARAMETERS/ -M BINDING_SITE_CENTER -r BINDING_SITE_
RADIUS -S SEED_NUMBER -R
```

In this specific case TARGET_PROTEIN is β2AR; MOLECULE_TO_DOCK is ICI 118,551; DOCKING_SOLUTION is the output name for the calculation; MEDUSADOCK_PARAMETERS is the directory where parameters for docking calculations are stored; BINDING_SITE_CENTER is the centroid of the ICI 118,551's crystallographic coordinates as retrieved from the PDB (PDB ID: 3NY8),

which has been chosen as center of the β2AR binding site; `BINDING_SITE_RADIUS` is 8 Å; `SEED_NUMBER` is a random number to be used to define a new independent Monte Carlo cycle; and `-R` is the flag which specify the initialization of a docking calculation in MedusaDock. The command is customizable for running multiple independent docking calculations as in the following *bash* script:

```
$> for i in $(seq -w 1 200 )
$> do
$>    rng = \$RANDOM    #random number generation
$>    ./medusaDock.linux -i TARGET_PROTEIN -m
      MOLECULE_TO_DOCK -o DOCKING_SOLUTION -p ./
      MEDUSADOCK_PARAMETERS/ -M BINDING_SITE_
      CENTER -r BINDING_SITE_RADIUS -S ${rng} -R
$> done
```

In this case, we perform 200 independent docking calculations of ICI 118,551 in β2AR. Even though MedusaDock can perform on a single 8-core CPU, each docking calculation requires on average 8 min to be completed, therefore the user should consider the use of supercomputer for the docking of small libraries of compounds.

4. We perform the cluster analysis using an *ad hoc* developed program. The less experienced user is advised to refer to the Conformer Cluster script available in the Resources of the Schrödinger Suite.

5. Perform MedusaDock calculations for propanolol enantiomers by adapting the command reported in **Note 3** to the new compounds.

Acknowledgments

References

1. Guedes IA, de Magalhães CS, Dardenne LE (2014) Receptor–ligand molecular docking. Biophys Rev 6:75–87

2. Grinter S, Zou X (2014) Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. Molecules 19:10150–10176

3. Audie J, Swanson J (2012) Recent work in the development and application of protein-peptide docking. Future Med Chem 4:1619–1644

4. Bhattacherjee A, Wallin S (2013) Exploring protein-peptide binding specificity through computational peptide screening. PLoS Comput Biol 9:e1003277

5. Dagliyan O, Proctor EA, D'Auria KM et al (2011) Structural and dynamic determinants of protein-peptide recognition. Structure 19:1837–1845

6. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. J Med Chem 49:5851–5855

7. Sousa SF, Fernandes PA, Ramos MJ (2006) Protein-ligand docking: current status and future challenges. Proteins 65:15–26

8. Teague SJ (2003) Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2:527–541

9. Carlson HA, McCammon JA (2000) Accommodating protein flexibility in computational drug design. Mol Pharmacol 57:213–218

10. Teodoro ML, Kavraki LE (2003) Conformational flexibility models for the receptor in structure based drug design. Curr Pharm Des 9:1635–1648

11. Barril X, Morley SD (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J Med Chem 48:4432–4443

12. Damm KL, Carlson HA (2007) Exploring experimental sources of multiple protein conformations in structure-based drug design. J Am Chem Soc 129:8225–8235

13. Karplus M (2003) Molecular dynamics of biological macromolecules: a brief history and perspective. Biopolymers 68:350–358

14. Karplus M, Kuriyan J (2005) Molecular dynamics and protein function. Proc Natl Acad Sci U S A 102:6679–6685

15. Rueda M, Bottegoni G, Abagyan R (2009) Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. J Chem Inf Model 49:716–725

16. Ding F, Yin SY, Dokholyan NV (2010) Rapid flexible docking using a stochastic rotamer library of ligands. J Chem Inf Model 50:1623–1632

17. Yin S, Biedermannova L, Vondrasek J, Dokholyan NV (2008) MedusaScore: an accurate force field-based scoring function for virtual drug screening. J Chem Inf Model 48:1656–1662

18. Gohlke H, Klebe G (2001) Statistical potentials and scoring functions applied to protein-ligand binding. Curr Opin Struct Biol 11:231–235

19. Golbraikh A, Tropsha A (2002) Beware of q(2)! J Mol Graph Model 20:269–276

20. Ding F, Dokholyan NV (2012) Incorporating backbone flexibility in medusadock improves ligand-binding pose prediction in the csar2011 docking benchmark. J Chem Inf Model 53:1871–1879

21. Serohijos AWR, Yin SY, Ding F et al (2011) Structural basis for mu-opioid receptor binding and activation. Structure 19:1683–1690

22. Kota P, Ding F, Ramachandran S, Dokholyan NV (2011) Gaia: automated quality assessment of protein structure models. Bioinformatics 27:2209–2215

23. Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. Nucleic Acids Res 28:235–242

24. Wacker D, Fenalti G, Brown MA et al (2010) Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. J Am Chem Soc 132:11443–11445

25. Halgren TA (1995) The Merck molecular force field. I. basis, form, scope, parameterization, and performance of MMFF94. J Comp Chem 17:490–519

26. Central limit theorem. Encyclopedia of Mathematics. http://www.encyclopediaof-math.org/index.php?title=Central_limit_theorem&oldid=18508

27. Legendre P, Legendre L (1998) Numerical Ecology. Second English Edition. Developments in Environmental Modelling 20:302–305. Elsevier, Amsterdam

28. Kelley LA, Gardner SP, Sutcliffe MJ (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng 9:1063–1065

29. Kleywegt GJ, Harris MR, Zou J et al (2004) The Uppsala electron-density server. Acta Crystallogr D Biol Crystallogr 60:2240–2249

30. Ramachandran S, Kota P, Ding F, Dokholyan NV (2011) Automated minimization of steric clashes in protein structures. Proteins 79:261–270

31. Jorgensen WL, Tirado-Rives J (1988) The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110(6):1657–1666

32. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236