W388–W392 *Nucleic Acids Research, 2017, Vol. 45, Web Server issue* *Published online 2 May 2017*
doi: 10.1093/nar/gkx352

Amino Acid Interaction (INTAA) web server

Jakub Galgonek[1,*], Jiří Vymětal[1], David Jakubec[1,2] and Jiří Vondrášek[1,*]

[1]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 542/2, Prague CZ-16610, Czech Republic and [2]Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University, Hlavova 8, Prague CZ-12843, Czech Republic

Received March 02, 2017; Revised April 12, 2017; Editorial Decision April 18, 2017; Accepted April 21, 2017

ABSTRACT

Large biomolecules—proteins and nucleic acids—are composed of building blocks which define their identity, properties and binding capabilities. In order to shed light on the energetic side of interactions of amino acids between themselves and with deoxyribonucleotides, we present the Amino Acid Interaction web server (http://bioinfo.uochb.cas.cz/INTAA/). INTAA offers the calculation of the residue Interaction Energy Matrix for any protein structure (deposited in Protein Data Bank or submitted by the user) and a comprehensive analysis of the interfaces in protein–DNA complexes. The Interaction Energy Matrix web application aims to identify key residues within protein structures which contribute significantly to the stability of the protein. The application provides an interactive user interface enhanced by 3D structure viewer for efficient visualization of pairwise and net interaction energies of individual amino acids, side chains and backbones. The protein–DNA interaction analysis part of the web server allows the user to view the relative abundance of various configurations of amino acid–deoxyribonucleotide pairs found at the protein–DNA interface and the interaction energies corresponding to these configurations calculated using a molecular mechanical force field. The effects of the sugar-phosphate moiety and of the dielectric properties of the solvent on the interaction energies can be studied for the various configurations.

INTRODUCTION

All existing biomolecules, such as proteins and DNA, reach their functional repertoire via a synergy of their composing elements (building blocks), which either stabilizes their functional structures or facilitates interactions with other biomolecules through sequential or structural epitopes. Both of the mentioned features are maintained by forces of physical origin, which can be divided into three general categories—electrostatics, dispersion and exchange-repulsion. The description of a structure (a molecule or molecular complex) can be significantly simplified by decomposing any phenomenon into energy contributions that reflect the complex behaviour of the molecule in question and by assigning these contributions to biomolecular building blocks—amino acids in proteins and deoxyribonucleotides in DNA.

The processes of structure formation and biomolecular interactions can be explained based on abundant information from structural databases. Even if the information is incomplete (not covering, for example, the phenomena of transient interactions or disordered proteins), it is comprehensive and the knowledge and the derived principles can be used for the prediction of biomolecular behaviour or functions. The effort of mapping non-covalent interactions for biomolecules in the past few years forms the basis of the proposed web server. The service is meant as a tool for theoreticians as well as bench scientists studying the properties of biomolecules and their interactions, namely protein stability via the Interaction Energy Matrix (IEM) and protein–DNA interactions represented by amino acid–nucleotide interaction analysis tool.

INTERACTION ENERGY MATRIX SERVER—A TOOL FOR ANALYSIS AMINO ACID INTERACTIONS IN PROTEINS

Historically, very soon after several proteins were characterized sequentially and structurally, the folding problem was stated. This highlights the fact that different sequences of amino acids in existing proteins result in very different and unique protein structures. The present view of protein folding emphasizes the role of thermodynamic forces that guide the peptide chain through the funnel-shaped energy landscape. Hydrophobic forces facilitate the formation of compact structures, but the conformational entropy of the chain always opposes the folding process. Consequently, the enthalpy, resulting from the contributions of many individual intra-chain amino–acid interactions as well as amino-

*To whom correspondence should be addressed. Tel: +420 220183294; Fax: +420 220410320; Email: jakub.galgonek@uochb.cas.cz
Correspondence may also be addressed to Jiří Vondrášek. Tel: +420 220183267; Fax: +420 220410320; Email: jiri.vondrasek@uochb.cas.cz

acid–solvent interactions, must be responsible for the major stabilization and specificity of protein folding (the folding code).

A large body of evidence from mutational studies suggests that amino-acid residues at certain positions in the sequence are more important for the stability and correct formation of a protein fold. Such key residues are usually structurally important and evolutionary conserved across the homologous sequences from different organisms. Apart from experimental and alignment-based approaches, an independent, structure-based method was proposed for the identification of the key residues and their effect on the stability of a protein (1). This method utilizes the calculation of physically sound interaction energies and evaluates a complete Interaction Energy Matrix (IEM), involving all pairs of amino acids. The key residues are determined based on the hypothesis that the amino-acid residues with the most stabilizing interactions contribute significantly to the folding enthalpy.

The first calculations and application of the IEM employed the precise but time-consuming *ab initio* methods. Therefore, they were limited to rather small proteins such as Trp-cage or rubredoxin (1,2). However, it was later demonstrated that the common force fields for biomolecules (OPLS-AA (3), AMBER parm03 (4)) provide an acceptable precision for the description of interaction energies (5,6). It has made it possible to scale up the computations dramatically and make the calculations feasible for any protein size. The faster calculations thus enabled further application of the IEM, such as an alternative energy-based definition of an amino-acid residue contact (7) or the investigation of protein–protein binding interfaces (8).

Methods

We propose the calculation of the IEM as a web service in order to make it easily available for other researchers. The IEM evaluates pairwise interaction energy (comprising only Lennard-Jones potential and point-charge electrostatics) between well-defined molecular fragments, such as protein and nucleic-acid residues. Although the initial concept of the IEM involved only the mutual amino-acid interactions, we generalized it towards protein–DNA interactions by including support for DNA and RNA residues in our web service.

This service offers the evaluation of the IEM by four common biomolecular force fields, namely OPLS-AA (3), AMBER parm03 (4) and parm99 (9), and charmm36 (10). The supported force fields are commonly used in molecular simulations and represent different parametrization approaches and strategies. These particular force fields were selected to match the community standards and to reflect our previous work on this topic (11–13). Concerning amino-acid residues, all of them provide sounded interaction energies. For calculations of complexes containing nucleic acids we recommend to use AMBER parm99 or charmm36 force field, of which nucleic-acid parameters are supported in the current version of IEM service.

The only mandatory input required from a user is a PDB identifier or a protein structure file compliant with the official PDB format and atom nomenclature. Due to the heuristic algorithm for the assignment of atomic types, the atom names may also follow the conventions of any force field supported for the calculation of the IEM. This makes the usage more straightforward if the structures to be analysed originate from molecular dynamics or modelling software.

Since X-ray protein structures deposited in the PDB are usually resolved without hydrogen atoms, we incorporated optional preprocessing by Reduce (14) to reconstruct them in full-atom resolution, which is necessary for further calculations. Afterwards, the PDB file is parsed removing the residues, which are not supported or excluded from IEM calculation (such as organic ligands, ions and water molecules). If multiple rotamers are present, only the first one is used for evaluation of interaction energies. In case of missing atoms, the assignment of the force field parameters for given molecular fragment fails and the particular interaction energies are reported in IEM as not available (NA).

The calculations of the IEM are performed immediately after submission and take from several seconds to a couple of minutes depending on the service load and the size of the protein. The results and analysis are afterwards presented in an interactive user interface (UI). The UI contains interactive tables and a structure viewer for the effective visualization of the chosen residues in the structure. Additionally, the whole IEM can be shown in a single table and exported in the CSV format (see Figure 1).

The identification of key residues relies on the net interaction energies, which are listed for all residues in the first interactive panel. To guide the eye, the strength of the net interactions is also visualized intuitively by bars next to particular numeric values. For a specific residue selected in the first panel, the decomposition of the net interaction energy is presented in the second panel. Simultaneously, the chosen residues are highlighted in the structure viewer.

The structure viewer works in two modes. By default, it colours the amino acids based on their net interaction energies, helping the user in finding the key residues. In the alternative mode, the colouring follows the pairwise interaction energies between the chosen reference residue and the others. In this regime, the colours refer directly to a particular row (or column) of the IEM.

Furthermore, the web application also offers a decomposition of interaction energies into side-chain–side-chain, backbone–backbone and backbone–side-chain contributions. All analyses and visualizations can be presented for any component of interaction energy.

For medium or bigger proteins, the interaction energy matrices can be very large. To save the user's internet bandwidth, the application obtains only summary information (net interaction energies) and residue parameters from the server. If necessary, these residue parameters can then be used to calculate a requested subset of pairwise interaction energies on the client's side.

AMINO ACID–NUCLEOTIDE INTERACTION ANALYSIS TOOL

The purpose of this part of the proposed web server is to provide the user with the knowledge of how a specific geometrical configuration of some amino acid–DNA residue contact relates to the background of all contacts of that

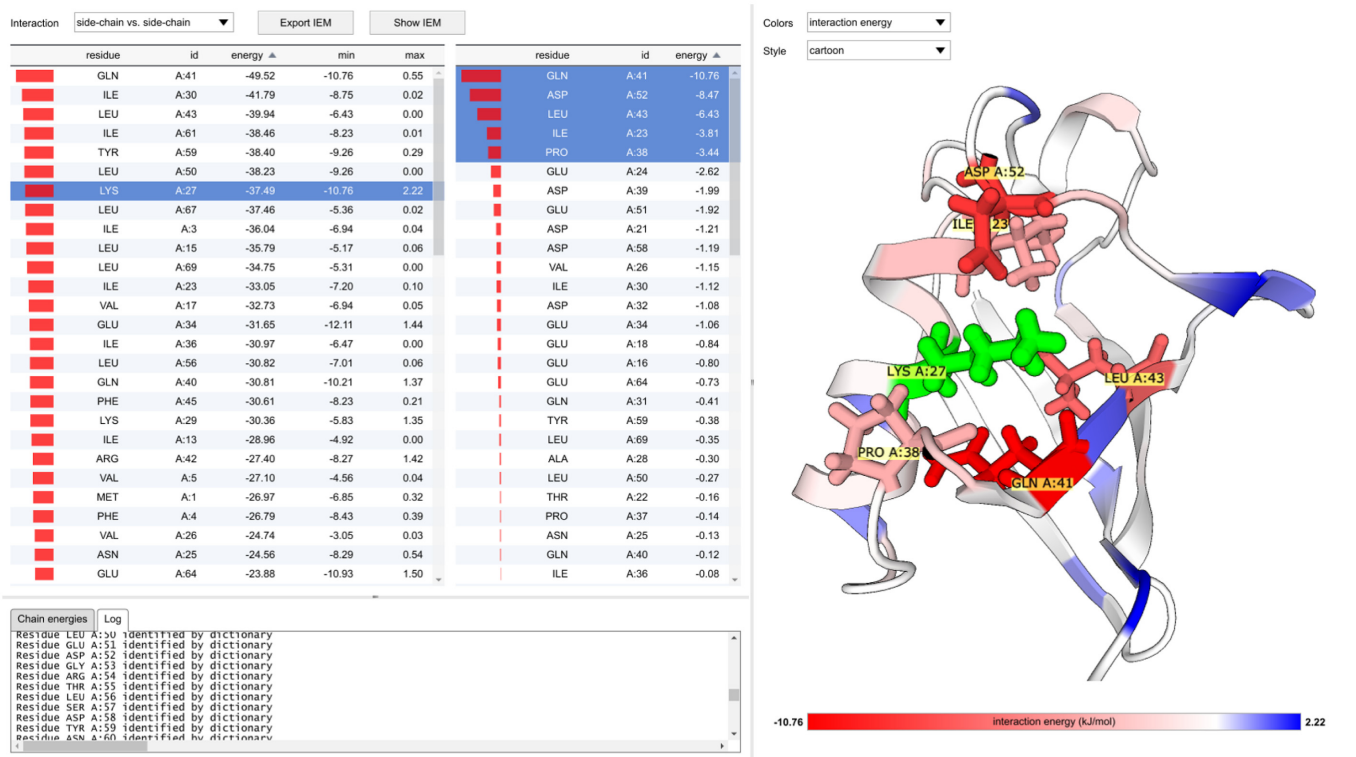

Figure 1. The user interface of the Interaction Energy Matrix Application. The UI provides two interactive tables and an interactive structure viewer. This screenshot captures an analysis of the stabilization role of LYS27 (in 1UBQ). This particular amino-acid residue provides one of the top net interaction energies as found on the left panel, where all net interactions are listed and optionally sorted. The right panel shows the decomposition of the net energy for the selected amino acid. Sorting by energy reveals the strongest interaction partners. The rightmost structure viewer reacts instantly on the actual selection in both panels. In the 'interaction energy' mode, the reference residue is coloured green and the others by corresponding interaction energies (the stabilizing interactions in red, the destabilizing and the repulsions in blue). The selected residues with the most stabilizing interactions (ILE23, PRO38, GLN41, LEU43 and ASP52) are additionally highlighted using full-atom representation.

type. The user can view how abundant each particular configuration is in the structures of protein–DNA complexes, what its interaction energy (IE) is, and how large that IE is when compared to other IEs in that particular distribution. Using the options provided by the web server interface (Figure 2), the user can also assess the relative importance of the DNA base and sugar–phosphate moieties for the process of specific recognition, or reflect on the role of the dielectric properties of the environment.

Methods

The IE calculations were performed using parameters derived from AMBER parm94 (DNA) (15) and parm99 (protein) (9,16) classical molecular mechanical force fields and, where applicable, a GB/SA implicit solvent and their technical details were described in detail elsewhere (17,18). The amino acid–deoxyribonucleotide interaction analysis part of the web server allows the user to view the interaction energies of the amino acid–DNA base or amino acid–deoxyribonucleoside monophosphate (dNMP) pairs. These pairs were extracted from the structures of 1584 protein–DNA complexes solved by X-ray crystallography to a resolution better than 2.5 Å deposited in the RCSB PDB (17,19). As such, they cover the configurational variability of amino acid–deoxyribonucleotide contacts under the constraining requirements of the protein–DNA interface.

The 3D transformation of all amino acid–deoxyribonucleotide pairs of a certain type (e.g. all contacts of asparagine with dAMP) extracted from the structures in order to minimize the RMSD of the atomic positions of the DNA base non-hydrogen atoms leads to the 3D distribution of that amino acid around that DNA base. Due to energy constraints, some areas of the amino acid–DNA residue configurational space are more populated than others. This results in the occurrence of spatially-defined clusters in these distributions. Up to six such clusters were identified in each distribution as described in our previous works (17,18). In each amino acid cluster within each distribution, a single amino acid–DNA residue pair, called cluster representative, can be defined as the pair containing the amino acid which has the lowest RMSD of the atomic positions from all the other members of that cluster (17,18,20). The amino acid–deoxyribonucleotide interaction analysis web server interface is shown in Figure 2.

IMPLEMENTATION DETAILS

The web applications are implemented in the Java language by using the Google Web Toolkit. Molecule and protein vi-

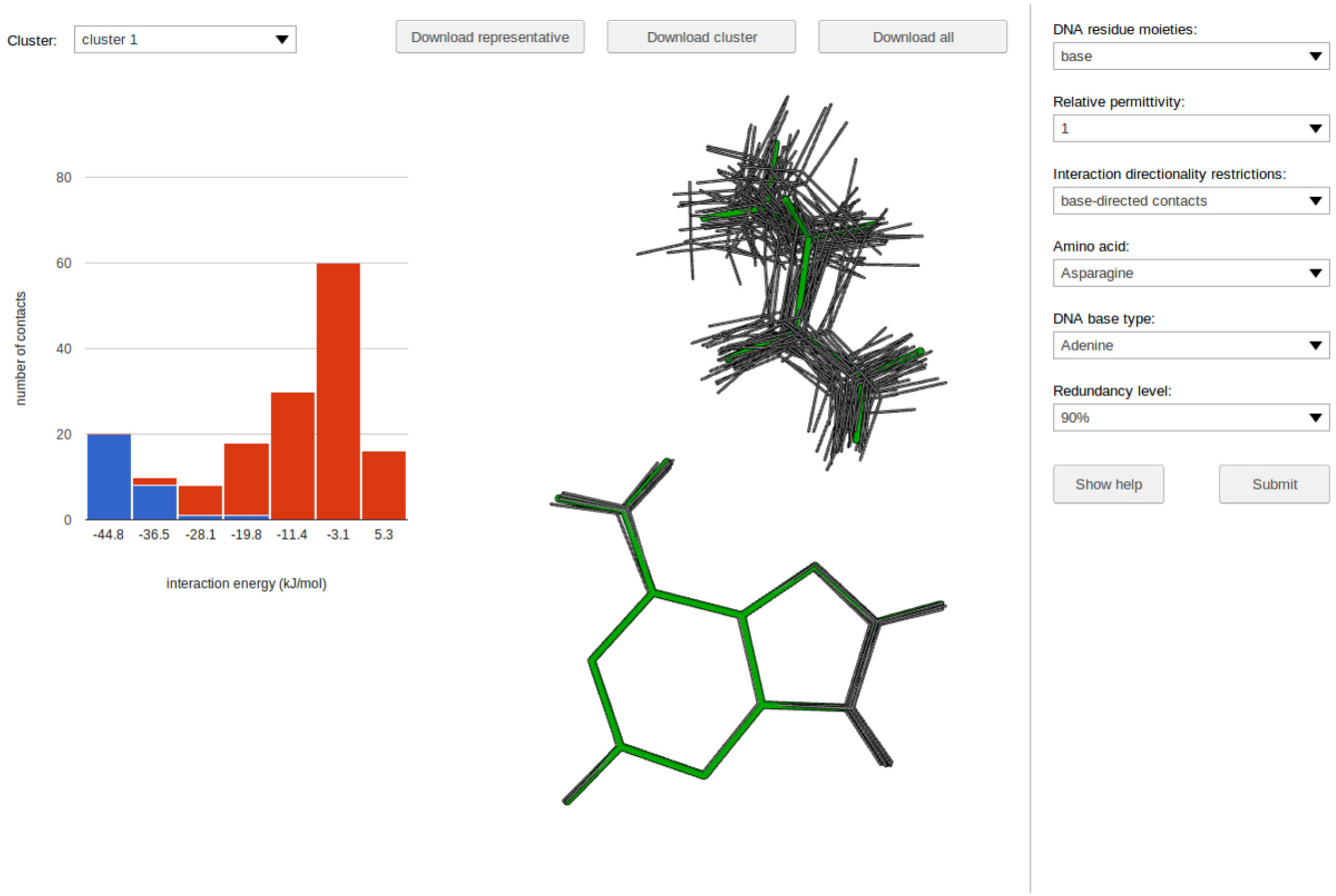

Figure 2. In the menu on the right, the user can select the parameters of the distribution, with details available in the 'Help' section. In the example provided, the distribution of asparagine side chains around the adenine base was chosen. Only the DNA base atoms were considered in the IE calculation, which was performed *in vacuo* (relative permittivity = 1). The only amino acid side chains considered in the IE calculations were those that directly contact the DNA base moiety. The protein chains from which these amino acids were extracted from a set in which the sequence identity of any pair of chains is lower than 90%. After you click 'Submit', a histogram of the IEs of the amino acid–DNA residue pairs from the selected distribution (the IE profile) is drawn on the left side of the screen. The numbers along the x-axis show the average IE of the contacts represented by that column. The height of each column represents the number of contacts falling into the IE range of that column. If at least one amino acid side-chain cluster exists in the distribution with the chosen parameters, a part of one or more columns is coloured blue. This corresponds to the contacts from that cluster having IEs within the range represented by that column. If some cluster is available and selected from the menu on the top left of the screen, the WebGL visualization tool displays the set of amino acid–DNA residue pairs from that cluster using the parameters chosen from the menu on the right side of the screen. The number (1–6) associated with each cluster in the menu indicates the rank of each cluster if one was to sort them from the one containing the most (cluster 1) to the one containing the fewest (cluster 6) amino acid–deoxyribonucleotide pairs in the original data set. It is possible that fewer than six clusters are listed, particularly when more restrictive sequence redundancy criteria are selected. The pair containing the cluster representative is drawn using green sticks and its 3D coordinates can be downloaded in PDB format by clicking the 'Download representative' button. Clicking the 'Download cluster' makes it possible to download the 3D coordinates of all members of the selected cluster. Clicking the 'Download all' button (available even when no cluster is selected) makes it possible to obtain the 3D coordinates of all amino acid–DNA residue pairs in the distribution.

sualization is supported by the WebGL-based viewer PV (21), and graphs are generated by Google Charts. The applications are deployed on the Apache Tomcat server, which runs under CentOS Linux on a virtual machine hosted by the CERIT Scientific Cloud. The applications require a modern web browser supporting HTML5 technologies (e.g. WebGL and File API). The applications have been tested on all major browsers, especially on Chromium and Firefox.

CONCLUSION

We provide the calculation of protein residue Interaction Energy Matrix as a web service with an interactive user interface and helpful visualizations which simplify the identification of residues important for the stability of the examined protein. Our service helps the users identify key residues with highly favourable interaction energies, which are supposed to energetically stabilize the particular protein fold. Alternatively, residues with unfavourable or suboptimal energetic contributions can be revealed as candidates for site-directed mutagenesis or *in silico* design with the aim of improving protein stability. Thus, the presented web service can be utilized as a valuable tool in the field of protein engineering and protein science in general.

The amino acid–deoxyribonucleotide interaction analysis tool complements the protein stability assessment service by providing the means to examine the relative abundance and interaction energies in various binding configurations

of these biomolecular building blocks. The web interface of the tool allows the user to visualize these binding motifs and provides means to study the effects of the dielectric properties of the solvent and of the sugar-phosphate moiety on the process of amino acid–deoxyribonucleotide interaction specificity.

Together, these services strive to provide the user with a robust toolset for an interaction energy-based analysis of the stability of protein and protein–DNA complex structures.

FUNDING

ELIXIR CZ (MEYS) [LM2015047]; CESNET [LM2015042]; CERIT Scientific Cloud [LM2015085]. Funding for open access charge: Institute of Organic Chemistry and Biochemistry of the CAS (RVO) [61388963].
Conflict of interest statement. None declared.

REFERENCES

1. Bendova-Biedermannova,L., Hobza,P. and Vondrasek,J. (2008) Identifying stabilizing key residues in proteins using interresidue interaction energy matrix. *Proteins*, **72**, 402–413.
2. Vondrasek,J., Bendova,L., Klusak,V. and Hobza,P. (2005) Unexpectedly strong energy stabilization inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: correlated ab initio quantum chemical calculations. *J. Am. Chem. Soc.*, **127**, 2615–2619.
3. Kaminski,G.A., Friesner,R.A., Tirado-Rives,J. and Jorgensen,W.L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, **105**, 6474–6487.
4. Duan,Y., Wu,C., Chowdhury,S., Lee,M.C., Xiong,G., Zhang,W., Yang,R., Cieplak,P., Luo,R., Lee,T. *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, **24**, 1999–2012.
5. Berka,K., Laskowski,R.A., Hobza,P. and Vondrasek,J. (2010) Energy matrix of structurally important side-chain/side-chain interactions in proteins. *J. Chem. Theory Comput.*, **6**, 2191–2203.
6. Berka,K., Laskowski,R., Riley,K.E., Hobza,P. and Vondrasek,J. (2009) Representative amino acid side chain interactions in proteins. A comparison of highly accurate correlated ab initio quantum chemical and empirical potential procedures. *J. Chem. Theory Comput.*, **5**, 982–992.
7. Fackovec,B. and Vondrasek,J. (2012) Optimal definition of inter-residual contact in globular proteins based on pairwise interaction energy calculations, its robustness, and applications. *J. Phys. Chem. B*, **116**, 12651–12660.
8. Kysilka,J. and Vondrasek,J. (2012) Towards a better understanding of the specificity of protein-protein interaction. *J. Mol. Recognit.*, **25**, 604–615.
9. Wang,J., Cieplak,P. and Kollman,P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049–1074.
10. Best,R.B., Zhu,X., Shim,J., Lopes,P.E., Mittal,J., Feig,M. and Mackerell,A.D. Jr (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.*, **8**, 3257–3273.
11. Vymetal,J. and Vondrasek,J. (2014) Parametrization of 2,2,2-trifluoroethanol based on the generalized AMBER force field provides realistic agreement between experimental and calculated properties of pure liquid as well as water-mixed solutions. *J. Phys. Chem. B*, **118**, 10390–10404.
12. Vymetal,J. and Vondrasek,J. (2013) Critical assessment of current force fields. Short peptide test case. *J. Chem. Theory Comput.*, **9**, 441–451.
13. Vymetal,J. and Vondrasek,J. (2010) Metadynamics as a tool for mapping the conformational and free-energy space of peptides—the alanine dipeptide case study. *J. Phys. Chem. B*, **114**, 5632–5642.
14. Word,J.M., Lovell,S.C., Richardson,J.S. and Richardson,D.C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.*, **285**, 1735–1747.
15. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
16. Lindorff-Larsen,K., Piana,S., Palmo,K., Maragakis,P., Klepeis,J.L., Dror,R.O. and Shaw,D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950–1958.
17. Jakubec,D., Laskowski,R.A. and Vondrasek,J. (2016) Sequence-specific recognition of DNA by proteins: Binding motifs discovered using a novel statistical/computational analysis. *PLoS One*, **11**, e0158704.
18. Jakubec,D., Hostas,J., Laskowski,R.A., Hobza,P. and Vondrasek,J. (2015) Large-scale quantitative assessment of binding preferences in protein-nucleic acid complexes. *J. Chem. Theory Comput.*, **11**, 1939–1948.
19. Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.F. Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem.*, **80**, 319–324.
20. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
21. Biasini,M. (2015) pv: v1.8.1. doi:10.5281/zenodo.20980.