Chapter 13

Predicting Real-Valued Protein Residue Fluctuation Using FlexPred

Lenna Peterson, Michal Jamroz, Andrzej Kolinski, and Daisuke Kihara

Abstract

The conventional view of a protein structure as static provides only a limited picture. There is increasing evidence that protein dynamics are often vital to protein function including interaction with partners such as other proteins, nucleic acids, and small molecules. Considering flexibility is also important in applications such as computational protein docking and protein design. While residue flexibility is partially indicated by experimental measures such as the B-factor from X-ray crystallography and ensemble fluctuation from nuclear magnetic resonance (NMR) spectroscopy as well as computational molecular dynamics (MD) simulation, these techniques are resource-intensive. In this chapter, we describe the web server and stand-alone version of FlexPred, which rapidly predicts absolute per-residue fluctuation from a three-dimensional protein structure. On a set of 592 nonredundant structures, comparing the fluctuations predicted by FlexPred to the observed fluctuations in MD simulations showed an average correlation coefficient of 0.669 and an average root mean square error of 1.07 Å. FlexPred is available at http://kiharalab.org/flexPred/.

Key words Bioinformatics, Computational biology, Support vector machine, Support vector regression, Protein residue fluctuation, Protein flexibility, Protein conformational flexibility, Protein structure, Protein design, Molecular dynamics

1 Introduction

The function of many proteins is determined not only by their rigid three-dimensional (3D) structure but also by the flexibility of protein chains [1]. Protein flexibility can influence function by, for example, determining catalytic rates [2] and affecting ligand and protein interactions [3]. Knowledge of flexibility is also important for accurate protein design [4, 5] and computational protein/ligand docking [3, 6].

Despite the importance of chain flexibility, it is difficult to glean a full picture of a protein's flexibility via experimental methods. Information about atomic fluctuations is reflected in the B-factor in X-ray crystallography [7]; however, the fluctuation is only one component of this uncertainty convolved with other

Yaoqi Zhou et al. (eds.), *Prediction of Protein Secondary Structure*, Methods in Molecular Biology, vol. 1484,
DOI 10.1007/978-1-4939-6406-2_13,

factors that cause errors in model building. In particular, B-factor tends to underestimate the motion of flexible regions [8]. Nuclear magnetic resonance (NMR) spectroscopy currently provides the most direct experimental evidence of flexibility; nevertheless, the accuracy depends on the experimental setup and the mathematical model used [9–12]. Cryogenic electron microscopy (cryo-EM) can detect heterogeneous conformational states [13] but not small conformational flexibility. Above all, experimental methods are costly in terms of time and resources and thus are not always applicable.

In order to augment experimental methods for determining protein flexibility, many computational approaches have been developed to model protein dynamics. Molecular dynamics (MD) simulations model the motion of all the atoms in a protein on the picosecond to microsecond timescale [14], from which per-residue fluctuation can be extracted. One approach to achieve faster results compared to MD is coarse-grained simulations [15–19]. Using trajectories from MD or coarse-grained simulations, normal mode analysis [20] can depict an overview of the motion that is easy to grasp. Alternatively, Gaussian network model (GNM) [21–24] uses a simplified model of protein structures to simulate protein motion. Many works have used GNM or related approaches to predict B-factor [21, 24–29]. Another work uses a mean-field model to predict fluctuations [30]. Although these physics-based computational methods provide a physical view of atom- or residue-level protein fluctuation, they are generally targeted toward computational biophysicists and may not be easy to use by experimental biologists. In addition to structure-based computational methods, there are methods that use sequence features to predict B-factor [31–33], relative motion [34], and two-state [35] or three-state [36] flexibility. These methods are applicable to a larger number of proteins since no structure is required, but with an inevitable decline in accuracy.

In this chapter, we describe the web server and stand-alone version of FlexPred, which rapidly predicts the absolute fluctuation of each residue in a protein structure. FlexPred is designed to predict the fluctuations exhibited by a protein during a 10 ns MD simulation (fast, comparatively small motions). Full details of the FlexPred method have been previously published [37]. We provide detailed instructions for using FlexPred and present example predictions for X-ray crystal structures of a monomer protein and a protein complex as well as a monomer structure determined using NMR. Both the web server and the stand-alone version can be found at http://kiharalab.org/flexPred.

2 Algorithm

FlexPred predicts the residue fluctuation observed during a 10 ns molecular dynamics (MD) simulation. MD fluctuation, the theoretical range of motion of the atoms in a protein structure, was computed as the root mean square distance between the Cα atom in the MD simulation and the Cα atom in the reference PDB file averaged across all time steps of the MD simulation [37]. The values produced by FlexPred can be used to estimate whether a specific portion of a protein chain is flexible and how flexible that region is.

FlexPred uses static features of a protein structure to predict MD residue fluctuation. The features tested were B-factor [7], residue distance from the protein center of mass [38, 39], residue contact number [40, 41], hydrophobic/hydrophilic [42] residue contact number, residue solvent accessible surface area [43, 44], residue depth [45], residue lower/upper half-sphere exposure [46], and secondary structure [43]. Details of each feature have been previously described [37]. FlexPred combines these features using the framework of support vector regression (SVR) implemented by LIBSVM [47]. It was trained on a non-redundant set of 592 molecular dynamics (MD) simulations from the Molecular Dynamics Extended Library (MoDEL) [48]. Almost all (96.11%) of the simulations were 10 ns in length and the rest were shorter. This is in the timescale of "fast" motions [14]; thus, FlexPred is not appropriate for predicting large movements such as domain-domain motion.

FlexPred predictions were evaluated using Pearson's correlation coefficient and the root mean square error (RMS) of the difference between the predicted fluctuation and the MD fluctuation. The highest single static feature correlation to real fluctuation was for residue contact number with a cutoff of 15 or 16 Å. This term had higher correlation to MD fluctuation than did B-factor. GNM prediction was also tested as a feature. While GNM alone had a higher correlation to MD fluctuation than did any static feature, including GNM in the feature set led to a consistent decrease in correlation coefficient [37]. Multiple combinations of features were tested and the highest correlation of 0.669 was observed with B-factor combined with residue contact number with cutoffs of 6, 8, 12, 16, 18, 20, and 22 Å (Feature set 15) [37]. The RMS of this feature combination was 1.07 Å [37]. This feature set effectively encodes information about 3D structures in a lower dimensional feature space.

3 Web Server

The web server takes a 3D protein structure in PDB format as input and predicts a fluctuation value for each residue of the protein. Figure 1 shows a screenshot of the web server input page.

3.1 Web Server Input

1. Go to the web server homepage: http://kiharalab.org/flexPred/.
2. Choose a PDB structure (*see* **Notes 1** and **2**):
   (a) To upload your own PDB file, click "Browse…" (Fig. 1 (1)).
   (b) To use a published PDB structure, enter the 4-character PDB code (e.g. 1bfg) into the PDB code box. (Fig. 1 (2)).
3. Choose the feature set (Fig. 1 (3); *see* **Note 3**):
   (a) "With B-factor": use this for X-ray structures.
   (b) "Without B-factor": use this for NMR structures and computational models.
4. Click "Predict Fluctuation" (Fig. 1 (4)).
5. The server will generally complete in 2–20 s.

3.2 Web Server Output

Figure 2 shows a screenshot of an example web server output page. The top of the page shows an image of the structure with high fluctuation residues colored in red and low fluctuation residues colored in blue (Fig. 2 (1)). The middle of the page shows a graph of the predicted fluctuation for each residue number (Fig. 2 (2)). Below the graph are links to download a comma separated value (CSV) file

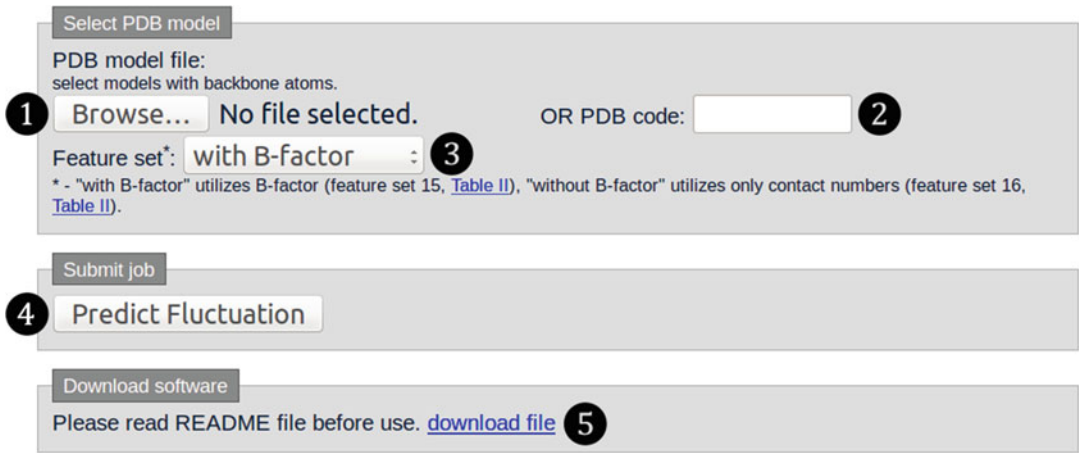

Fig. 1 The FlexPred web server query submission page. (1) Upload a PDB file or (2) choose a published PDB structure by ID. (3) Choose the feature set with B-factor for X-ray structures or without B-factor for NMR structures. (4) Finally, click "Predict Fluctuation." (5) The software may also be downloaded

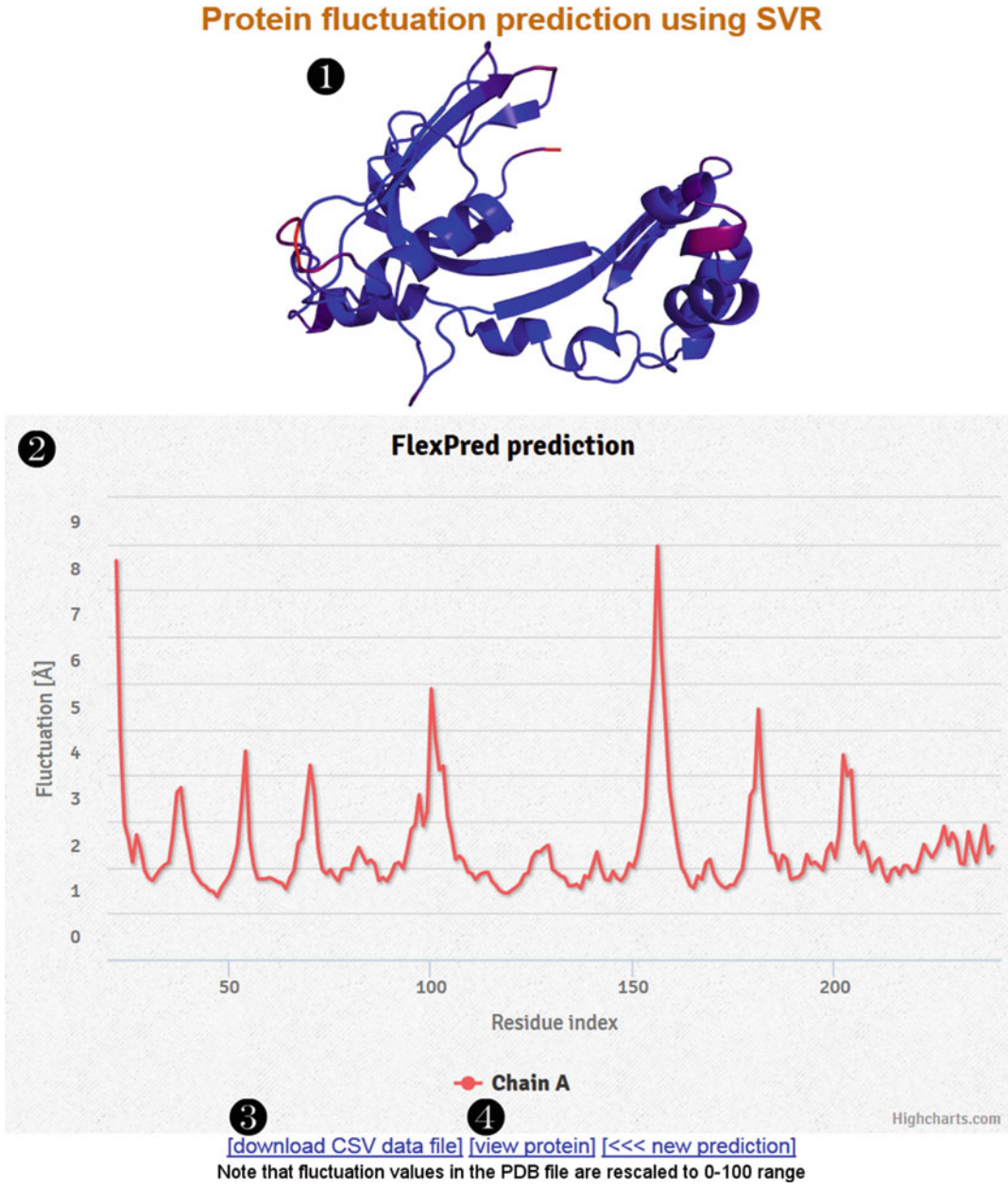

Fig. 2 The FlexPred output page. (1) The structure with high fluctuation shown in *red*. (2) The fluctuation of each residue. (3) Download link for the raw data file. (4) Download link for the PDB file with predicted fluctuation in the B-factor column

containing the predicted fluctuations for each residue (Fig. 2 (3)) and a PDB file with normalized predicted fluctuations in the B-factor column (Fig. 2 (4)). The CSV file can be conveniently viewed using spreadsheet software.

4 Stand-Alone Software

The stand-alone software requires Python and has only been tested on Linux. It takes a 3D protein structure in PDB format as input and predicts a fluctuation value for each residue of the protein.

4.1 Stand-Alone Input

1. Go to the web server homepage (http://kiharalab.org/flexPred/).
2. Under the heading "Download software," "click" "download file" (Fig. 1 (5)).
3. Expand the file using the command "tar -xvf flexPred.tar.xz."
4. Download libsvm from http://www.csie.ntu.edu.tw/~cjlin/libsvm/ and compile.
5. Edit "predictFluct.py" to set paths to libsvm and FlexPred.
6. The script has two positional arguments:
   (a) The first argument is the path to the PDB file (see Notes 1 and 2).
   (b) The second argument is the feature set. "NMR" uses the feature set without B-factor and "XRAY" uses the feature set with B-factor (*see* **Note 3**).
7. Try the example protein: "cd example; python ../predictFluct.py 1BFG.pdb XRAY" (*see* **Note 4**).
8. To run FlexPred on other proteins: "python predictFluct.py model_file.pdb [NMR|XRAY]."

4.2 Stand-Alone Output

The stand-alone version of FlexPred should complete within seconds. The stand-alone software produces a text file containing the predicted fluctuations for each residue.

5 Case Studies

We present example FlexPred predictions using three types of proteins: a monomer X-ray structure, a dimer X-ray structure, and a monomer NMR structure. The predictions for the X-ray structures use the "With B-factor" feature set while the prediction for the NMR structure uses the "Without B-factor" feature set. Figure 3 and Table 1 show that FlexPred predictions have moderate to high correlation to real fluctuation from MD simulations. On the test set in the previous paper, the average correlation coefficient was 0.669 and the average RMS was 1.07 Å [37]. The MD fluctuation is the RMS of the difference between each snapshot in the trajectory and the mean position of the trajectory.

The first example is the ssDNA binding protein gp32 (PDB ID 1gpc) (Fig. 3a, b). The highest core fluctuation is in a loop around

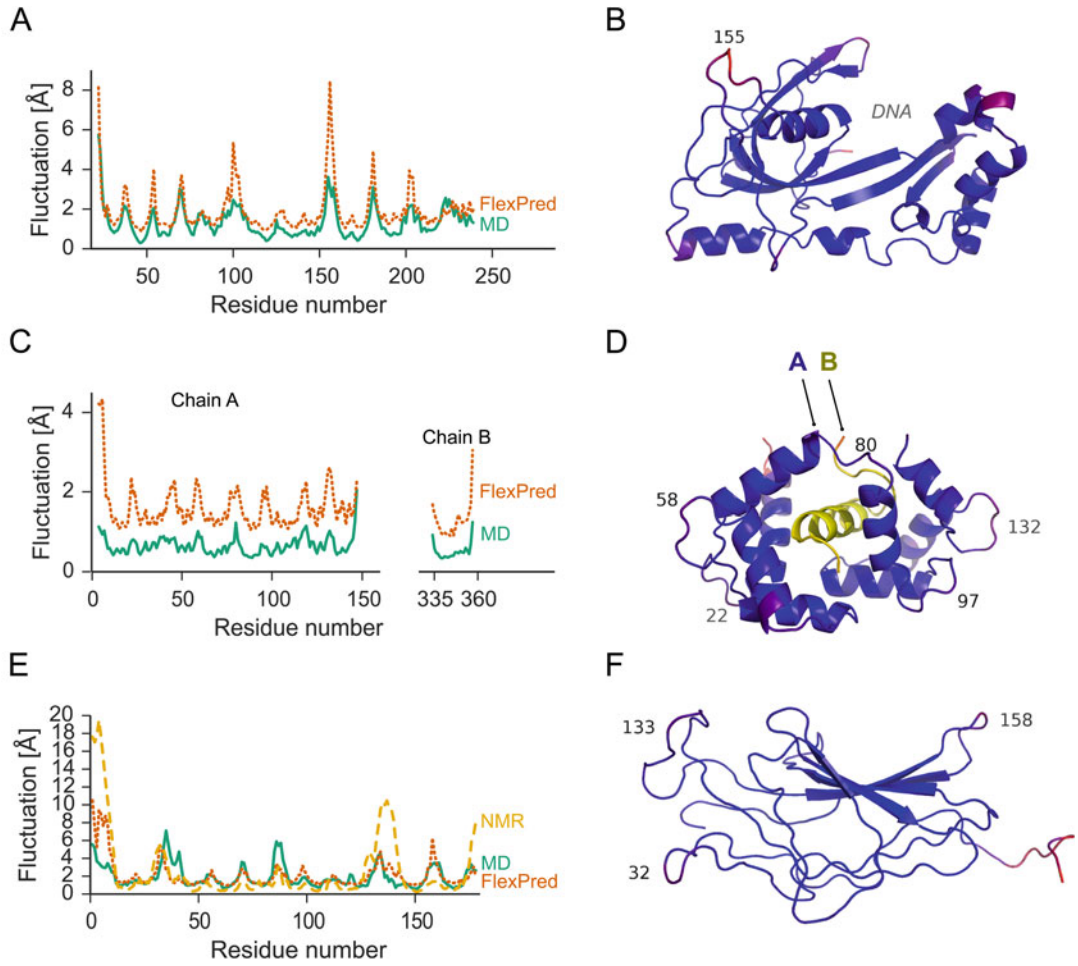

Fig. 3 Examples of FlexPred predictions. (**a**, **c**, **e**) fluctuation of each residue. *Green solid line*: MD fluctuation. *Orange dotted line*: Predicted fluctuation by FlexPred. *Yellow dashed line*: NMR fluctuation. (**b**, **d**, **f**) structures with high fluctuation shown in *red* and notable high fluctuation residues indicated with numbers. Proteins used: (**a**, **b**) single-stranded DNA binding protein gp32 from bacteriophage T4, residues 22–239 (PDB ID: 1gpc). (**c**, **d**) Calmodulin (chain A) in complex with a fragment of Ca(2+)/calmodulin-dependent kinase kinase (chain B) (PDB ID: 1iq5). Calmodulin, Chain A, is residues 4–147 and shown in *blue*. CaMKK, chain B, is residues 334–357 and shown in *yellow*. (**e**, **f**) human transcription factor NFATc, DNA-binding domain, residues 3–178 (PDB ID: 1nfa)

residue 155. ssDNA fits into a cleft sized to exclude dsDNA [49], where minimal flexibility was observed. High flexibility could allow dsDNA to bind. Overall, the predicted fluctuation agrees well with those in MD simulation. The prediction had a correlation coefficient of 0.839 and RMS of 0.83 Å to the MD simulation, better than the average correlation and RMS observed in the benchmark in the original paper [37] (Table 1).

The second example is the heterodimer of calmodulin (CaM) in complex with the CaM binding domain of CaM-dependent

Table 1
FlexPred prediction on two example X-ray protein structures

Structure	PCC	RMS (Å)
1gpcA	0.839	0.83
1iq5A	0.620	1.03
1iq5B	0.933	0.81
Prev. avg.	0.669	1.04

Pearson's correlation coefficient (PCC) (perfect correlation is 1, no correlation is 0, and perfect negative correlation is −1) and root mean square deviation (no deviation is 0) between FlexPred and MD residue fluctuations. Average values are from the X-ray dataset in the original paper [37]

kinase kinase (CaMKK) (PDB ID 1iq5) (Fig. 3c, d). The prediction overestimated fluctuations in comparison with the MD simulation, although the average correlation and RMS of the two chains were better than the average (Table 1). The high predicted flexibility of the calcium binding loops (at residues 22, 58, 97, and 132) may be because the features only consider the position of protein atoms, not the bound calcium ion. In the MD fluctuation, the linker between the two domains (residue 80) shows the highest core fluctuation. It was observed that binding of the CaMKK peptide causes a shift in the relative angle of the two domains [50]. However, while FlexPred does predict flexibility at the domain linker (residue 80), it is of similar magnitude to multiple other loops. The lower magnitude of the predicted linker flexibility could be due to the timescale of the domain-domain motion, which is likely longer than the 10 ns timescale of the training data. The accuracy for this complex is comparable to the accuracy for monomers (Table 1), indicating that while FlexPred was not trained using multimers, it is applicable to predict the flexibility of a protein-protein complex.

The third example is the NMR structure of the DNA binding domain of the transcription factor NFATc (PDB ID 1nfa) (Fig. 3e, f). Because NMR structures lack B-factor, the prediction used the "Without B-factor" setting. The predictions are slightly worse than the average computed on the previous test of NMR structures [37]. For this example, we also computed NMR ensemble fluctuation by computing the RMS of each model to the first model (Table 2). It is interesting that FlexPred shows higher correlation to NMR (0.85) than to MD (0.58). In the NMR ensemble, the highest fluctuation in the core region (residues 13–172 [51]) is in a loop around residue 133, where both MD and FlexPred match the location of the increased fluctuation but with much lower magnitude.

Table 2
FlexPred prediction on an example NMR protein structure

Pair	Structure	PCC	RMS (Å)
FlexPred-MD	1nfaA	0.584	1.37
	Prev. avg.	0.686	2.16
FlexPred-NMR	1nfaA	0.846	2.52
	Prev. avg.	0.739	1.81
MD-NMR	1nfaA	0.471	3.37
	Prev. avg.	0.651	2.42

Pearson's correlation coefficient (PCC) (perfect correlation is 1, no correlation is 0, and perfect negative correlation is −1) and root mean square deviation (no deviation is 0) of fluctuations between FlexPred and MD, between FlexPred and NMR, and between MD and NMR. Average values are from the NMR dataset in the original paper [37]

For NMR, the second highest core fluctuation is in the DNA binding loop around residue 32 [51], where FlexPred matches NMR almost perfectly while the peak from MD simulation is shifted. Both FlexPred and MD show much higher fluctuation than NMR for the loop around residue 158.

6 Conclusions

We outline the web server and stand-alone software for FlexPred, which predicts a real-valued absolute fluctuation for each residue of an input 3D protein structure. The web server is easy to use and quickly provides accurate prediction with intuitive visualization. It is useful for analyzing function of a protein from its structure and for artificial design of proteins.

7 Notes

1. If the PDB file contains multiple models (generally only in NMR structures and marked by lines such as "MODEL 1" and "MODEL 2"), only the first model will be considered.
2. If the PDB file contains multiple chains (indicated with different chain IDs, e.g., A, B), all chains will be used to compute contact maps and a separate prediction will be made for each chain. If the protein is a biological monomer but the PDB file contains crystal contacts, additional chains should be removed from the file before prediction for the most accurate results. Protein oligomeric state is often annotated in the "REMARK 350" section of a PDB file and can be predicted using the PISA server [52].

3. "With B-factor" uses feature set 15 from the original paper [37] (mean correlation coefficient 0.669, mean RMS 1.07 Å) and "Without B-factor" uses feature set 16 (mean correlation coefficient 0.660, mean RMS 1.09 Å). If "With B-factor" is selected but the B-factors in the file are all zero (0.0), "Without B-factor" will be substituted automatically.
4. The stand-alone version requires write access to the directory where the input PDB file is located.

Acknowledgements

This work was partly supported by the National Institute of General Medical Sciences of the National Institutes of Health (R01GM097528) and the National Science Foundation (IIS1319551, DBI1262189, IOS1127027).

References

1. Teilum K, Olsen JG, Kragelund BB (2009) Functional aspects of protein flexibility. Cell Mol Life Sci 66:2231–2247
2. Hammes GG, Benkovic SJ, Hammes-Schiffer S (2011) Flexibility, diversity, and cooperativity: pillars of enzyme catalysis. Biochemistry 50:10422–10430
3. Zacharias M (2010) Accounting for conformational changes during protein–protein docking. Curr Opin Struct Biol 20:180–186
4. Mandell DJ, Kortemme T (2009) Backbone flexibility in computational protein design. Curr Opin Biotechnol 20:420–428
5. Lassila JK (2010) Conformational diversity and computational enzyme design. Curr Opin Chem Biol 14:676–682
6. Lill MA (2011) Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. Biochemistry 50:6157–6169
7. Debye P (1913) Interferenz von Röntgenstrahlen und Wärmebewegung. Ann Phys 348:49–92
8. Eastman P, Pellegrini M, Doniach S (1999) Protein flexibility in solution and in crystals. J Chem Phys 110:10141
9. Ishima R, Torchia DA (2000) Protein dynamics from NMR. Nat Struct Biol 7:740–743
10. Baldwin AJ, Kay LE (2009) NMR spectroscopy brings invisible protein states into focus. Nat Chem Biol 5:808–814
11. Nilges M, Habeck M, O'Donoghue SI, Rieping W (2006) Error distribution derived NOE distance restraints. Proteins 64:652–664
12. Chalaoux F-R, O'Donoghue SI, Nilges M (1999) Molecular dynamics and accuracy of NMR structures: effects of error bounds and data removal. Proteins 34:453–463
13. Wang Q, Matsui T, Domitrovic T, Zheng Y, Doerschuk PC, Johnson JE (2013) Dynamics in cryo EM reconstructions visualized with maximum-likelihood derived variance maps. J Struct Biol 181:195–206
14. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE (2009) Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol 19:120–127
15. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA (1997) A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. J Comput Chem 18:849–873
16. Kolinski A (2004) Protein modeling and structure prediction with a reduced representation. Acta Biochim Pol 51:349–371
17. Jamroz M, Kolinski A, Kmiecik S (2013) CABS-flex: server for fast simulation of protein structure fluctuations. Nucleic Acids Res 41:W427–W431
18. Jamroz M, Orozco M, Kolinski A, Kmiecik S (2013) Consistent view of protein fluctuations from all-atom molecular dynamics and coarse-grained dynamics with knowledge-based force-field. J Chem Theory Comput 9:119–125

19. Jamroz M, Kolinski A, Kmiecik S (2014) CABS-flex predictions of protein flexibility compared with NMR ensembles. Bioinformatics 30:2150–2154
20. Brooks B, Karplus M (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A 80:6571–6575
21. Haliloglu T, Bahar I, Erman B (1997) Gaussian dynamics of folded proteins. Phys Rev Lett 79:3090–3093
22. Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 77:1905–1908
23. Bahar I, Erman B, Haliloglu T, Jernigan RL (1997) Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. Biochemistry 36:13512–13523
24. Yang L, Song G, Jernigan RL (2009) Protein elastic network models and the ranges of cooperativity. Proc Natl Acad Sci U S A 106:12347–12352
25. Kondrashov DA, Cui Q, Phillips GN Jr (2006) Optimization and evaluation of a coarse-grained model of protein motion using X-ray crystal data. Biophys J 91:2760–2767
26. Lin T-L, Song G (2010) Generalized spring tensor models for protein fluctuation dynamics and conformation changes. BMC Struct Biol 10:S3
27. Micheletti C, Carloni P, Maritan A (2004) Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models. Proteins 55:635–645
28. Canino LS, Shen T, McCammon JA (2002) Changes in flexibility upon binding: application of the self-consistent pair contact probability method to protein-protein interactions. J Chem Phys 117:9927
29. Opron K, Xia K, Wei G-W (2015) Communication: capturing protein multiscale thermal fluctuations. J Chem Phys 142:211101
30. Pandey BP, Zhang C, Yuan XZ, Zi J, Zhou YQ (2005) Protein flexibility prediction by an all-atom mean-field statistical theory. Protein Sci 14:1772–1777
31. Zhang H, Kurgan L (2014) Improved prediction of residue flexibility by embedding optimized amino acid grouping into RSA-based linear models. Amino Acids 46:2665–2680
32. Chen P, Wang B, Wong H-S, Huang D-S (2007) Prediction of protein B-factors using multi-class bounded SVM. Protein Pept Lett 14:185–190
33. Schlessinger A, Yachdav G, Rost B (2006) PROFbval: predict flexible and rigid residues in proteins. Bioinformatics 22:891–893
34. Hirose S, Yokota K, Kuroda Y, Wako H, Endo S, Kanai S, Noguchi T (2010) Prediction of protein motions from amino acid sequence and its application to protein-protein interaction. BMC Struct Biol 10:20
35. Gu J, Gribskov M, Bourne PE (2006) Wiggle—predicting functionally flexible regions from primary sequence. PLoS Comput Biol 2:e90
36. de Brevern AG, Bornot A, Craveur P, Etchebest C, Gelly J-C (2012) PredyFlexy: flexibility and local structure prediction from sequence. Nucleic Acids Res 40:W317–W322
37. Jamroz M, Kolinski A, Kihara D (2012) Structural features that predict real-value fluctuations of globular proteins. Proteins 80:1425–1435
38. Shih C-H, Huang S-W, Yen S-C, Lai Y-L, Yu S-H, Hwang J-K (2007) A simple way to compute protein dynamics without a mechanical model. Proteins 68:34–38
39. Kloczkowski A, Jernigan RL, Wu Z, Song G, Yang L, Kolinski A, Pokarowski P (2009) Distance matrix-based approach to protein structure prediction. J Struct Funct Genomics 10:67–81
40. Lin C-P, Huang S-W, Lai Y-L, Yen S-C, Shih C-H, Lu C-H, Huang C-C, Hwang J-K (2008) Deriving protein dynamical properties from weighted protein contact number. Proteins 72:929–935
41. Halle B (2002) Flexibility and packing in proteins. Proc Natl Acad Sci U S A 99:1274–1279
42. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132
43. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637
44. Miller S, Janin J, Lesk AM, Chothia C (1987) Interior and surface of monomeric proteins. J Mol Biol 196:641–656
45. Chakravarty S, Varadarajan R (1999) Residue depth: a novel parameter for the analysis of protein structure and stability. Structure 7:723–732
46. Hamelryck T (2005) An amino acid has two sides: a new 2D measure provides a different view of solvent exposure. Proteins 59:38–48
47. Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:27:1–27:27

48. Meyer T, D'Abramo M, Hospital A et al (2010) MoDEL (molecular dynamics extended library): a database of atomistic molecular dynamics trajectories. Structure 18:1399–1409
49. Shamoo Y, Friedman AM, Parsons MR, Konigsberg WH, Steitz TA (1995) Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. Nature 376:362–366
50. Kurokawa H, Osawa M, Kurihara H, Katayama N, Tokumitsu H, Swindells MB, Kainosho M, Ikura M (2001) Target-induced conformational adaptation of calmodulin revealed by the crystal structure of a complex with nematode Ca2+/calmodulin-dependent kinase kinase peptide. J Mol Biol 312:59–68
51. Wolfe SA, Zhou P, Dötsch V, Chen L, You A, Ho SN, Crabtree GR, Wagner G, Verdine GL (1997) Unusual Rel-like architecture in the DNA-binding domain of the transcription factor NFATc. Nature 385:172–176
52. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774–797