OXFORD

Structural bioinformatics

PREFMD: a web server for protein structure refinement via molecular dynamics simulations

Lim Heo and Michael Feig*

Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

*To whom correspondence should be addressed.

Abstract

Summary: Refinement of protein structure models is a long-standing problem in structural bioinformatics. Molecular dynamics-based methods have emerged as an avenue to achieve consistent refinement. The PREFMD web server implements an optimized protocol based on the method successfully tested in CASP11. Validation with recent CASP refinement targets shows consistent and more significant improvement in global structure accuracy over other state-of-the-art servers.

Availability and implementation: PREFMD is freely available as a web server at http://feiglab.org/prefmd. Scripts for running PREFMD as a stand-alone package are available at https://github.com/feiglab/prefmd.git.

Contact: feig@msu.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Protein structure prediction has become an essential tool in structural biology (Cavasotto and Phatak, 2009). Although predicted model structures can be very useful they often have deficiencies such as incorrect secondary structure orientations, loop structures, side-chain packings and/or poor local stereochemical properties (Modi and Dunbrack, 2016; Nugent *et al.*, 2014). This necessitates further improvements of the model qualities via refinement methods (Feig, 2017) that complement the homology modeling techniques commonly used to generated the initial models (Zhang, 2008). Since CASP10, refinement methods have started to achieve consistent refinement (Nugent *et al.*, 2014). Refinement methods based on molecular dynamics (MD) were most successful as a result of improved force fields and ensemble averaging approaches (Feig and Mirjalili, 2016; Mirjalili and Feig, 2013).

Several refinement methods tested in CASP have been implemented as web servers (Heo *et al.*, 2013; Khoury *et al.*, 2014; Rodrigues *et al.*, 2012; Zhang *et al.*, 2011), including one refinement server, PRINCETON_TIGRESS, that is based on MD-based refinement using implicit solvent (Khoury *et al.*, 2014). However, because of short simulations and strong restraints, significant model improvements are not achieved (Khoury *et al.*, 2014). There remains a need for a web service that offers more extensive MD-based structure refinement.

Here, the PREFMD (Protein REFinement via Molecular Dynamics) web server is presented that implements a more extensive MD-based refinement protocol based on the best-performing refinement method in CASP11 by the FEIG group (Feig and Mirjalili, 2016). In the original method, 100 ns-scale MD simulations were employed followed by structure selection and ensemble averaging to obtain a refined model. To adopt this method for use in a web service, the computational costs were optimized while still performing close to the original method. The web service relies on GPU-acceleration to provide fast turn-around times and delivers significant and consistent refinement for most structures based on tests with refinement targets from the last rounds of CASP.

2 Results

2.1 The PREFMD method

The PREFMD web server is based on the protein structure refinement method by the FEIG group in CASP11 (Feig and Mirjalili, 2016). The method has been updated to improve the refinement accuracy and the computational efficiency while the main idea of the refinement protocol is preserved. Briefly, this method relies on initial rounds of explicit solvent MD simulations with weak positional restraints to prevent large structural deviations. From the structures

Table 1. PREFMD benchmark results for CASP refinement targets

CASP (#)[a]	ΔGDT-HA[b]	ΔGDC-SC[b]	−ΔRMSD[b]	ΔSG[b,c]	Mol Probity[d]
9 (14)	2.11 (86%)	0.64 (50%)	−0.01 (57%)	0.42 (50%)	0.69 (100%)
10 (27)	1.21 (74%)	2.89 (78%)	0.02 (70%)	−0.24 (37%)	0.75 (89%)
11 (37)	1.96 (81%)	3.18 (86%)	0.04 (78%)	1.02 (76%)	0.76 (100%)
12 (28[e])	0.85 (64%)	0.82 (68%)	−0.01 (46%)	0.02 (46%)	0.77 (93%)
All (106)	1.50 (75%)	2.15 (75%)	0.01 (65%)	0.35 (55%)	0.75 (95%)

[a]Number of targets.
[b]average improvements.
[c]Sphere-Grinder.
[d]Average Molprobity scores; percentage of improved cases is given in parentheses.
[e]14 targets from the CASP12 were excluded from the benchmark because they exceeded 300 residues (TR520, TR905, TR909, TR910, TR912, TR913, TR917, TR928, TR942, TR945) or the unavailability of the native structure (TR874, TR875, TR876, TR887, TR910).

sampled during MD, a subset is selected using empirical scoring and the distance from the initial model and subsequently averaged to obtain a refined model (Mirjalili and Feig, 2013). Different from the original protocol, a reduced set of simulations of five runs, each over 30 ns (150 ns in total) is carried out, and the latest CHARMM force field, CHARMM36m (Huang et al., 2017) is used. MD simulations are run on GPUs using OpenMM (Eastman et al., 2013) via CHARMM (Brooks et al., 2009). A local refinement protocol [locPREFMD (Feig, 2016)] was also incorporated to improve stereochemical properties once before starting MD simulations and again as the final step after averaging.

2.2 Web-server implementation

On the submission page at http://feiglab.org/prefmd (see Supplementary Fig. S1), the input model in PDB format is expected. An optional e-mail address is used for notifications and to deliver results. To limit computational costs protein sizes are limited to 300 residues. Typical runs take 1–3 days (see Supplementary Fig. S2). Results are then available for download. The PREFMD method uses CHARMM with OpenMM and the MMTSB Tool Set (Feig et al., 2004). The web server is implemented using perl and bash scripts with a MySQL database backend. The server hardware consists of shared resources with multi-core Intel Xeon CPUs and NVIDIA K40, GTX-980/1080 GPUs.

3 Results and discussion

The PREFMD web server was tested on the recent CASP refinement category targets. The results are summarized in Table 1, and the performance as a result of initial model quality is described in the Supplementary Figure S3. Although the simulation time is dramatically reduced compared to the original protocol, significant refinement close to the performance reported during CASP is still possible. Both global and local structure quality were improved as measured by GDT-HA (global distance test-high accuracy), GDC-SC (global distance cutoff-side chain) and MolProbity scores, respectively. Especially, the improvement in global structure accuracy by 1.5 GDT-HA units on average is significantly better than any

other refinement web service (Heo et al., 2013; Khoury et al., 2014; Rodrigues et al., 2012; Zhang et al., 2011). Moreover, global refinement consistently succeeded in 75% of the cases whereas the local stereochemistry was improved for 95% of the targets. CASP12 targets were refined less, presumably because MD-based refinement has already become part of some prediction pipelines.

We further analyzed the performance of PRFMD as a function of the prediction group, and thereby the method that generated the initial model. (see Supplementary Table S1) Interestingly, models from the Zhang group (using I-TASSER and QUARK) (Yang et al., 2016), the Baker group (using Rosetta), RaptorX, MULTICOM, and Pcons using Modeller could be refined significantly whereas models from the LEE group were not refined at all on average. Presumably this is, again, because MD-based refinement was already part of the LEE group protocol (Joo et al., 2016). However, local accuracy could be improved for almost all models, except a subset of models generated by Pcons.

4 Conclusions

The new PREFMD protein structure refinement web server is described. The method is based on a top-performing MD-based refinement protocol that was optimized to perform well at reduced computational costs to operate under the resource constraints of a community web service. The service provides consistent refinement with better performance than other state-of-the-art refinement servers. We expect that the server will be useful as the final stage in protein structure prediction.

References

Brooks,B.R. et al. (2009) CHARMM: the biomolecular simulation program. J. Comput. Chem., 30, 1545–1614.

Cavasotto,C.N., and Phatak,S.S. (2009) Homology modeling in drug discovery: current trends and applications. Drug Discov. Today, 14, 676–683.

Eastman,P. et al. (2013) OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. J. Chem. Theory Comput., 9, 461–469.

Feig,M. (2016) Local protein structure refinement via molecular dynamics simulations with locPREFMD. J. Chem. Inf. Model., 56, 1304–1312.

Feig,M. (2017) Computational protein structure refinement: almost there, yet still so far to go. WIREs Comput. Mol. Sci., 7, e1307.

Feig,M. et al. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J. Mol. Graph. Model., 22, 377–395.

Feig,M., and Mirjalili,V. (2016) Protein structure refinement via molecular-dynamics simulations: what works and what does not?. Proteins, 84, 282–292.

Heo,L. et al. (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res., 41, W384–W388.

Huang,J. et al. (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods, 14, 71–73.

Joo,K. et al. (2016) Template based protein structure modeling by global optimization in CASP11. Proteins, 84, 221–232.

Khoury,G.A. et al. (2014) Princeton_TIGRESS: protein geometry refinement using simulations and support vector machines. Proteins, 82, 794–814.

Mirjalili,V., and Feig,M. (2013) Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. *J. Chem. Theory Comput.*, **9**, 1294–1303.

Modi,V., and Dunbrack,R.L. Jr. (2016) Assessment of refinement of template-based models in CASP11. *Proteins*, **84** (Suppl. 1), 260–281.

Nugent,T. *et al*. (2014) Evaluation of predictions in the CASP10 model refinement category. *Proteins*, **82** (Suppl. 2), 98–111.

Rodrigues,J.P. *et al*. (2012) KoBaMIN: a knowledge-based minimization web server for protein structure refinement. *Nucleic Acids Res.*, **40**, W323–W328.

Yang,J. *et al*. (2016) Template-based protein structure prediction in CASP11 and retrospect of I-TASSER in the last decade. *Proteins*, **84** (Suppl. 1), 233–246.

Zhang,J. *et al*. (2011) Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure*, **19**, 1784–1795.

Zhang,Y. (2008) Progress and challenges in protein structure prediction. *Curr. Opin. Struct. Biol.*, **18**, 342–348.