BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 20 2011, pages 2913–2914
doi:10.1093/bioinformatics/btr482

Structural bioinformatics

Advance Access publication August 27, 2011

Fast and accurate prediction of protein side-chain conformations

Shide Liang[1], Dandan Zheng[2], Chi Zhang[3] and Daron M. Standley[1,*]

[1]Systems Immunology Lab, Immunology Frontier Research Center, Osaka University, Suita, Osaka 565-0871, Japan, [2]Department of Radiation Oncology, Massey Cancer Center, Virginia Commonwealth University, Richmond, VA 23298 and [3]School of Biological Sciences, Center for Plant Science and Innovation, University of Nebraska, Lincoln, NE 68588, USA

Associate Editor: Anna Tramontano

ABSTRACT

Summary: We developed a fast and accurate side-chain modeling program [Optimized Side Chain Atomic eneRgy (OSCAR)-star] based on orientation-dependent energy functions and a rigid rotamer model. The average computing time was 18 s per protein for 218 test proteins with higher prediction accuracy (1.1% increase for χ_1 and 0.8% increase for χ_{1+2}) than the best performing program developed by other groups. We show that the energy functions, which were calibrated to tolerate the discrete errors of rigid rotamers, are appropriate for protein loop selection, especially for decoys without extensive structural refinement.

Availability: OSCAR-star and the 218 test proteins are available for download at http://sysimm.ifrec.osaka-u.ac.jp/OSCAR

Contact: standley@ifrec.osaka-u.ac.jp

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

Received on June 23, 2011; revised on July 29, 2011; accepted on August 14, 2011

1 INTRODUCTION

Over the past two decades, much effort has been spent improving the accuracy or speed of side-chain modeling methods. Most methods exploit a limited number of representative conformations, called rotamers, at each residue position and use efficient search algorithms to find a low-energy rotamer combination for the whole protein. In spite of their efficiency, rigid rotamers are inherently accompanied by a discrete error and not suited for physics-based force fields, which are sensitive to small atomic clashes: the calculated energies can be quite different for the native conformation and near-native rotamers. Force fields thus have to be modified, by either scaling the atomic radii (Dahiyat and Mayo, 1997), or using softer Lennard–Jones repulsive terms (Yanover *et al.*, 2008) to reduce the influence of the steric clashes. Alternatively, knowledge-based, coarse-grained energy functions have been developed that can tolerate rigid rotamers while achieving high accuracy (Liang and Grishin, 2002). A third approach is to use extremely detailed rotamer libraries, or flexible rotamer models, in combination with accurate energy functions at the cost of speed (Peterson *et al*., 2004).

Recently, we developed a side-chain modeling program combining accurate, orientation-dependent, Optimized Side Chain Atomic eneRgy (OSCAR-o) with a flexible rotamer model (Liang *et al*., 2011a). The prediction accuracy was significantly higher (2.2% for χ_1 and 4.0% for χ_{1+2}) than that of the next-best method, but the run time was as long as 28 min for a single protein. In this study, we adopted OSCAR to a rigid rotamer model by modifying the distance-dependent component for fast side-chain modeling. The parameters of the orientation-dependent functions were optimized so that decoy proteins with low RMSD (root mean square deviation) from native structures could be distinguished from a pool of decoys (obtained by perturbing the energy functions and then modeling the entire protein). The proposed methodology (OSCAR-star) is very fast while maintaining high accuracy.

2 RESULTS

2.1 Parameter optimization

The parameters of the distance-dependent energy functions (OSCAR-dstar) were initialized to the corresponding values (OSCAR-d) previously optimized, by maximizing the energy gap between the native conformation and rotamers at each modeled position (Liang *et al.*, 2011a). To model a side chain at a given position, OSCAR-dstar exploited a limited number of rigid rotamers to find the rotamer that had the lowest energy. The original OSCAR-d parameters were sensitive to discrete errors of rigid rotamers and the mean RMSD of the lowest energy rotamers was as large as 0.785 Å for a training set of 40 000 side chains per residue type (the rotamer interior energy was calculated the same as OSCAR-d). We then optimized the parameters to improve the accuracy by Monte Carlo (MC) simulation. Consequently, the RMSD was dropped to 0.734 Å and the accuracy (90.9% for χ_1 and 80.8% for χ_{1+2}) for single residues in 30 test proteins was similar to that of OSCAR-d with a flexible rotamer model. In the next step, the optimized OSCAR-dstar potential was multiplied by an orientation-dependent function to yield OSCAR-star. The parameters of the orientation-dependent function were optimized by simultaneously minimizing the RMSD of the lowest energy rotamer at each modeled position and the RMSD of the lowest energy decoy obtained by perturbing the energy functions and then modeling the entire protein. As a result, the prediction accuracy of OSCAR-star increased by 0.6 and 0.7% for χ_1 and χ_{1+2}, respectively, compared with OSCAR-dstar when modeling all residues in each of the 218 test proteins.

2.2 Comparison with other methods

We compared the performance of OSCAR-star with other top-ranked side-chain modeling programs (Table 1) such as CISRR (Cao *et al*., 2011), SCWRL4 (Krivov *et al*., 2009), LGA

*To whom correspondence should be addressed.

Table 1. Comparison of side-chain modeling programs in prediction accuracy and running time for 218 independent test proteins

Program[a]	All residues			Core residues			
	χ_1 (%)	χ_{1+2} (%)	RMSD (Å)	χ_1 (%)	χ_{1+2} (%)	RMSD (Å)	CPU time/ protein[b]
CISRR	84.7	73.1	1.49	92.6	85.9	0.95	23 s
SCWRL4[c]	85.1	74	1.48	93	86.9	0.96	7 s
LGA[c]	86.1	72.3	1.42	93.9	85.9	0.91	5 m 53 s
NCN[c]	86.3	74.3	1.48	93.8	87.9	0.87	20 m 50 s
OSCAR-d[c]	86.6	75.3	1.41	95.5	90.4	0.7	9 m 26 s
OPUS_Rota[c]	86.6	75.7	1.4	94.3	87.6	0.86	7 s
OSCAR-dstar	87.1	75.7	1.37	93.9	86.3	0.87	14 s
OSCAR-star	87.7	76.4	1.35	94.4	87.3	0.85	18 s
OSCAR-o[c]	88.8	79.7	1.24	95.9	91.9	0.62	27 m 49 s

[a]The list of programs are sorted according to χ_1 accuracy. Default parameters/arguments were used in the calculations.
[b]OPUS_Rota was run on one Intel Xeon 3.0 GHz processor and other programs were run on one AMD Opteron 2.7 GHz processor.
[c]The prediction accuracies of SCWRL4, LGA, NCN, OPUS_Rota, OSCAR-d and OSCAR-o were obtained from our previous work (Liang *et al*., 2011a).

(Liang and Grishin, 2002), NCN (Peterson *et al*., 2004), OPUS_Rota (Lu *et al*., 2008), OSCAR-d and OSCAR-o. OSCAR-star had better prediction accuracies than other programs except OSCAR-o and was faster than all but three programs: SCWRL4, OPUS_Rota and OSCAR-dstar. In other words, OSCAR-star was more accurate than all of the faster side-chain modeling programs. According to a paired *t*-test, the χ_1 accuracy difference between OSCAR-star and the three programs was statistically significant ($P < 0.0001$).

The performance of a side-chain modeling is affected by the energy function, structural representation and search algorithm. Efficient search algorithms save time but help little to improve the prediction accuracy. For rigid-rotamer-based side-chain modeling programs such as OSCAR-star, the MC simulation time is less than that used to calculate rotamer–backbone and rotamer–rotamer interaction energies in the initial stage (see Methods in Supplementary Material). Orientation-dependent energy functions are essential for high accuracy. For example, OPUS_Rota, the most accurate side-chain modeling program after OSCAR methods (Table 1), uses orientation-dependent statistical energy functions. On the other hand, flexible rotamer models, which are time consuming, cannot achieve accurate predictions without high-quality energy functions. In fact, the three programs using flexible rotamers, CISRR, SCWRL4 and NCN, have lower accuracies than the rigid-rotamer-based OPUS_Rota and OSCAR-star. OSCAR-o, which uses both orientation-dependent energy functions and a flexible roamer model, is the most accurate and also slower than the other methods. With a state-of-the-art search algorithm, SCWRL4 is the fastest, even though a flexible rotamer model is used.

2.3 Protein loop selection with OSCAR-star

We have previously demonstrated that OSCAR-o has higher accuracy than other energy functions in selecting near native conformations from loop decoys (Liang *et al*., 2011b). Here, we compared the performance of OSCAR-star with OSCAR-o for the RAPPER decoy set (de Bakker *et al*., 2003), in which every loop target contained 1000 decoys optimized by side-chain modeling and 50 top scored decoys further optimized by energy minimization. We modeled side-chain conformations of loop residues with OSCAR-o/OSCAR-star before each energy calculation. For the decoys without energy minimization, OSCAR-star demonstrated better performance than OSCAR-o in 7 out of 11 loop lengths from 2 to 12 and equal accuracy for five-residue loops. For the energy-minimized decoys, OSCAR-star was effective for long loops but poor for short loops compared with the more accurate OSCAR-o. The relatively coarse-grained OSCAR-star was superior to OSCAR-o, which was sensitive to incomplete sampling and atomic clashes, for decoys without energy minimization. Moreover, it took 5 min for OSCAR-star to model side-chain conformations of 1000 decoys for an eight-residue loop target compared with 5 h for OSCAR-o. OSCAR-star is thus appropriate for the initial stage of loop modeling. Side-chain conformations can be modeled very fast at candidate loop backbones, which makes it possible to sample loop conformations extensively (>1000 decoys). The top ranked decoys can be then energy minimized and evaluated by more accurate force fields such as OSCAR-o.

Funding: DMS was supported by the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST), Japan Science for the Promotion of Science (JSPS).

Conflict of Interest: none declared.

REFERENCES

Cao,Y. *et al*. (2011) Improved side-chain modeling by coupling clash-detection guided iterative search with rotamer relaxation. *Bioinformatics*, **27**, 785–790.

Dahiyat,B.I. and Mayo,S.L. (1997) Probing the role of packing specificity in protein design. *Proc. Natl Acad. Sci. USA*, **94**, 10172–10177.

de Bakker,P.I. *et al*. (2003) Ab initio construction of polypeptide fragments: accuracy of loop decoy discrimination by an all-atom statistical potential and the AMBER force field with the Generalized Born solvation model. *Proteins Struct. Funct. Bioinform.*, **51**, 21–40.

Krivov,G.G. *et al.* (2009) Improved prediction of protein side-chain conformations with SCWRL4. *Proteins Struct. Funct. Bioinform.*, **77**, 778–795.

Liang,S. and Grishin,N.V. (2002) Side-chain modeling with an optimized scoring function. *Protein Sci.*, **11**, 322–331.

Liang,S.D. *et al*. (2011a) Protein side chain modeling with orientation dependent atomic force fields derived by series expansions. *J. Comput. Chem.*, **32**, 1680–1686.

Liang,S.D. *et al.* (2011b) Protein loop selection using orientation dependent force fields derived by parameter optimization. *Proteins Struct. Funct. Bioinform.*, **79**, 2260–2267.

Lu,M.Y. *et al.* (2008) OPUS-Rota: a fast and accurate method for side-chain modeling. *Protein Sci.*, **17**, 1576–1585.

Peterson,R.W. *et al.* (2004) Improved side-chain prediction accuracy using an ab initio potential energy function and a very large rotamer library. *Protein Sci.*, **13**, 735–751.

Yanover,C. *et al.* (2008) Minimizing and learning energy functions for side-chain prediction. *J. Comput. Biol.*, **15**, 899–911.