PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

SHORT COMMUNICATION

Distance-dependent atomic knowledge-based force in protein fold recognition

Mehdi Mirzaie[1,2]* and Mehdi Sadeghi[3]*

[1]Department of Basic Sciences, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[2]Department of Bioinformatics, School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

[3]Department of Bioinformatics, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

ABSTRACT

We have recently introduced a novel model for discriminating the correctly folded proteins from well-designed decoy structures using mechanical interatomic forces. In the model, we considered a protein as a collection of springs and the force imposed to each atom was calculated by using the relation between the potential energy and the force. A mean force potential function is obtained from statistical contact preferences within the known protein structures. In this article, the interatomic forces are calculated by numerical derivation of the potential function. For assessing the knowledge-based force function we consider an optimal structure and define a score function on the 3D structure of a protein. We compare the force imposed to each atom of a protein with the corresponding atom in the optimum structure. Afterwards we assign larger scores to those atoms with the lower forces. The total score is the sum of partial scores of atoms. The optimal structure is assumed to be the one with the highest score in the dataset. Finally, several decoy sets are applied in order to evaluate the performance of our model.

Proteins 2012; 00:000–000.

Key words: knowledge-based potential; interatomic force; score function; 3D structure; decoy set.

INTRODUCTION

Understanding the mechanism of how proteins fold and bind is one of the central problems in molecular biology. The native structure of a protein is generally assumed to be at the lowest free energy.[1] On the basis of this hypothesis an accurate free energy would enable the prediction and assessment of the protein structure. Two different types of energy functions are used in protein fold recognition and threading studies. The first is based on laws of physics and is referred to as physical energy function. In the physical energy function, a molecular mechanics force field is used. Molecular mechanics force fields such as AMBER,[2–4] CHARMM,[5] GROMOS,[6] and OPLS[7] are parameterized from *ab initio* calculation and small molecular structural data. Although atomic level structure refinement can be achieved with molecular dynamic simulation, these potential functions are very time consuming, thus these functions have been undesirable in protein structure prediction. The second type is extracting the potential from the known protein structures and is referred to as knowledge-based potentials. Derivation of these potential functions is based on the statistical averages and distributions. There are various forms of statistical energy functions depending on how statistics are calculated and how proteins are modeled,[8] for example, distance independent contact energies,[9–13] solvent accessible surface potential,[14,15] distance dependent potentials,[16–26] and angular dependence.[27,28] Recently, combination of these statistics such as distance and orientation are widely used.[29–36]

In most cases, one or two points for each residue are considered to represent a protein.[37–39] These points are usually C_α, C_β, or the center of mass of each side chain. The interaction can be either distance dependent or

Grant sponsor: IPM; Grant number: CS 1389-0-01.
*Correspondence to: Mehdi Mirzaie, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, P.O. Box 19716-53313, Tehran, Iran. E-mail: mirzaie@ipm.ir and Mehdi Sadeghi, National Institute of Genetic Engineering and Biotechnology, P.O. Box 14155-6343, Tehran, Iran. E-mail: sadeghi@nigeb.ac.ir.
Received 13 June 2011; Revised 15 November 2011; Accepted 6 December 2011
Published online 10 December 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24011

only dependent on contact. In fact, two atoms within a specific cutoff distance have interaction. Contacts by distance cutoff can potentially include many neighbors which have no significant physical interaction. In other words, two atoms which are close to each other may be shielded from contact by other atoms. The protein potential function based on Delaunay tessellations solves this problem. This protein potential function was originally proposed by Tropsha and coworkers.[40,41] The amino acids in a protein chain are represented by their C_α atoms, C_β atoms, or all atoms. Then the Delaunay tessellation of the resulting point set is computed using Qhull algorithm.[42] It defines the nearest neighbors in an unambiguous and parameter freeway by decomposing the structure into tightly packed tetrahedra which have the atoms as their corners. From the database of residue frequencies for these tetrahedra, a potential function is extracted using the inverse Boltzmann distribution.[16]

We have previously described the calculation of an atomic knowledge-based potential and force for discrimination of native structures from decoys.[26] In fact, by considering an ideal spring between a pair of atoms and the relation between the energy and the force we calculated the harmonic approximated force function.

The value of force on an object is the derivative of the potential function U. Consequently, if we calculate the mean force potential for two atoms *i* and *j* at distance d we would have the value of force at distance d, particularly, the force value at the minimum energy is equal to zero. Additionally, the direction of the aforementioned force is discussed in the material and methods section.

In this study, we obtain a reasonable and accurate knowledge-based force function, which can help us to recognize the native structure from decoys. Through calculating the resultant of the forces imposed to each atom in a native structure, it is expected that the norm of the resultant of the forces on each atom to be close to zero.

We begin by calculating a mean force potential from a sample of the native structures. Next, we derive the distance dependent force function and use it to calculate the resultant of the forces imposed to each atom of a protein structure and then we introduce a scoring function. To test the performance of our approach we apply it on several decoy sets to measure its ability to discriminate native structure from decoys. Several decoy sets, which contain hundreds of decoy proteins and are generated in different ways, have been used. In most cases, this approach has been able to distinguish native structure from decoys. The calculated *Z*-score, as a useful quantitative measure of the validity of the calculated potential, shows high value for all protein data sets. At the end, we compare our theory to the previously published harmonic approximated force function with the aid of six multiple target decoy sets.

A detailed description of the training and decoy sets, the scoring functions and the evaluation criteria are provided in the results and discussion section. At the end, we compare our results to the previous and three well known other scoring functions with the aid of eight multiple target decoy sets.

MATERIALS AND METHODS

Calculation of distance-dependent potential

We extract knowledge-based mean force potential for assessment of contact of three atom types within TOP500H records.[43] This is a nonredundant set of 500 proteins resolved by X-ray crystallography with at least 1.8 Å resolutions. The three atom types are containing C (carbon) and S (sulfur) in one group, N (nitrogen) and O (oxygen). To quantify contact preferences, we use Delaunay tessellation.[26] In this procedure two atoms are in contact if they do not shielded from contact by other atoms.

The distances between any two atoms are divided into 30 distance shells, starting from 0.75 Å with distance shell 0.5 Å in width. All pairwise occurrences out of this range are excluded. Another parameter that we consider is the sequence separation. Usually, sequence separation longer than about 10 amino acids consider as nonlocal.[44] So we consider six distinct values for sequence separation (k = 1, 2, 3, 4,), ($5 \leq k \leq 8$), and ($k \geq 9$). Only all pairwise atoms of N and C with a sequence separation $k = 1$ are omitted. We count each pairwise contact asymmetrically, which means that a contact between atom A and atom B (when A is closer to the N-terminal side than B) is counted separately from a contact between atom B and atom A (when B is closer to the N-terminal side than A). For that reason, we have constructed 54 different atom pairwise empirical distributions. The pairwise pseudo-energy terms for the atomic pair i and j at sequence separation k and at distance shell l has been calculated by the following equation as described by Sippl.[16]

$$\Delta E_k^{ij}(l) = RT \ln[1 + M_{ijk}\sigma] - RT \ln\left[1 + M_{ijk}\sigma \frac{f_k^{ij}(l)}{f_k^{xx}(l)}\right] \quad (1)$$

Where M_{ijk} is the number of observations for the atomic pair i and j at sequence separation k, f_k^{ij} is the relative frequency of occurrence for the atomic pair i and j at sequence separation k in the class of distance l and $f_k^{xx}(l)$ is the relative frequency of occurrence for all the atomic pairs at sequence separation k in the distance shell l.

The temperature was set to 293 K, so that RT is equivalent to 0.582 kcal/mol and $\sigma = 1/50$.

Distance-dependent force function

In the previous study,[26] it is assumed that the energy function around the minimum value follows Hook's law; hence the distance dependent force function is approximated by the relation between the energy and the force function. Figure 1 shows the curve for some pairs of

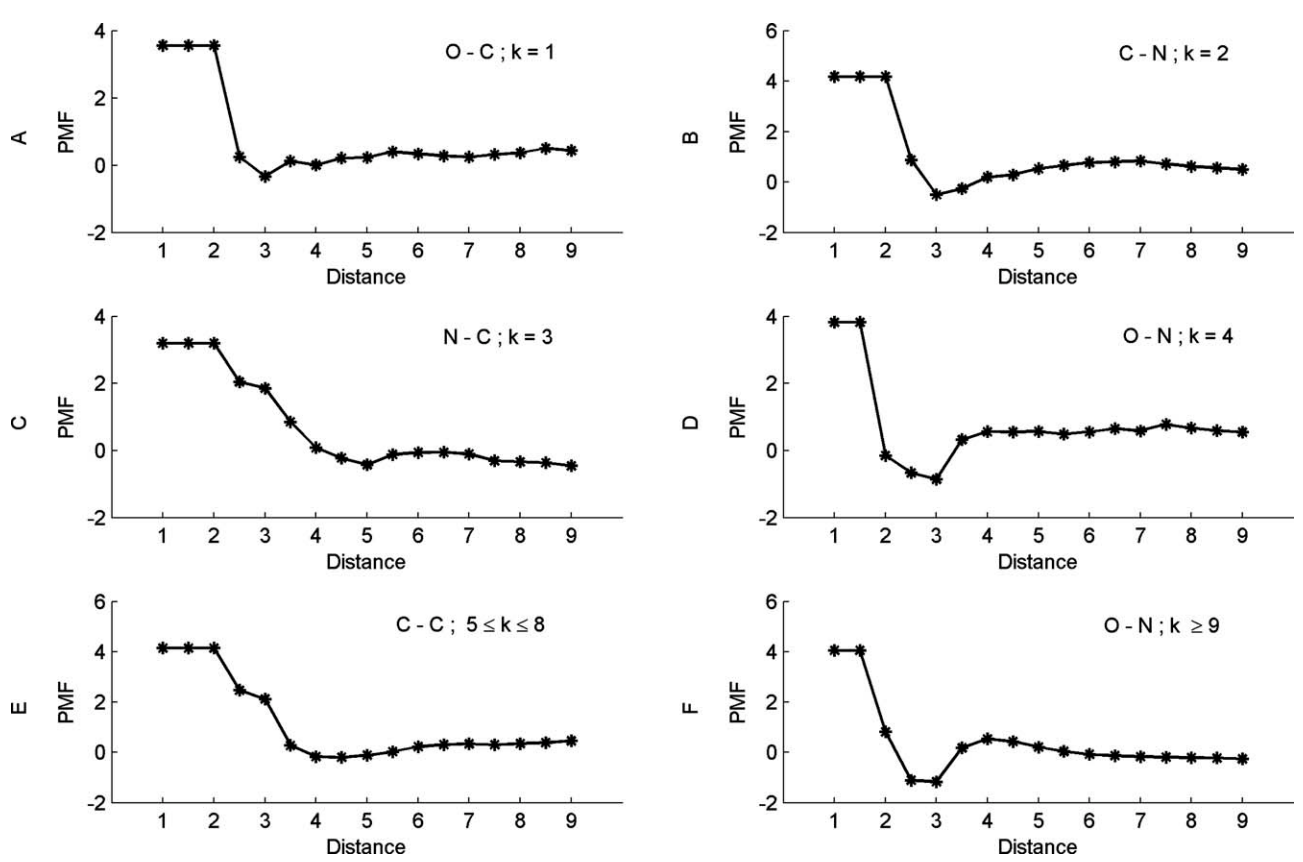

Figure 1

The potential of mean force (PMF): (**A**) for the atomic pair O—C at sequence separation of $k = 1$; (**B**) for the atomic pair C—N at sequence separation of $k = 2$; (**C**) for the atomic pair of N—C at sequence separation of $k = 3$; (**D**) for the atomic pair O—N at sequence separation of $k = 4$; (**E**) for the atomic pair C—C at sequence separation of $5 \leq k \leq 8$; (**F**) for the atomic pair O—N at sequence separation of $k \geq 9$.

atoms at some particular sequence separation. As shown by the Figure 1, the energy function increases slowly toward zero after the minimum point and the rate of change of energy in the repulsive manner is different from the attractive one. In the repulsive manner, the energy function decreases considerably, but in the attractive manner increases very gently, therefore a simple method to estimate the derivative of a function is used in order to calculate the force function.

Numerical differentiation is a technique of numerical analysis to produce an estimate of the derivative of a mathematical function using values from the corresponding function. The simplest method is to use finite difference approximations. A simple estimation is to compute the slope of a nearby secant line through the points $(x - h, f(x - h))$ and $(x + h, f(x + h))$ as shown in Figure 2. The slope of this line is

$$m = f(x + h) - f(x - h)/(2h) \qquad (2)$$

The force on an object is the negative of the derivative of the potential function E. This means that it is the negative of the slope of the potential energy curve. So we can estimate the value of the force imposed to each atom by using the values of the energy function with respect to the types of atoms and distances among them.

Let E be $\Delta E_k^{AiAj}(d)$ that obtained from (1). So by (2), the magnitude of the force that atom A_i imposed to atom A_j at distance d is

$$E(d + 0.5) - E(d - 0.5) \qquad (3)$$

Specially, if d be the bin of the distance in which E is at the minimum value, then the force is zero.

Vector of force imposed to an atom

Force is a vector quantity which has both magnitude and direction. The magnitude of the force that atom A_i imposes to atom A_j is dependent on the distance d which is obtained in the previous section. A positive value implies the repulsiveness, while a negative value implies attractiveness of the force. In the repulsive manner, the direction of the force vector is from A_i to A_j, where in the attractive manner, the direction of the force vector is from A_j to A_i.

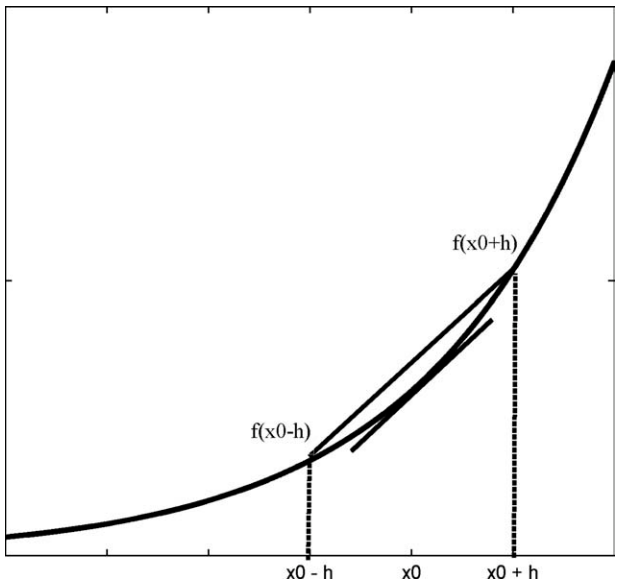

Figure 2

Derivative of $f(x)$ at x_0 approximated by secant line through the points $(x_0 - h, f(x_0 - h))$ and $(x_0 + h, f(x_0 + h))$.

Given the coordinates of A_i and A_j, the magnitude of the force and the direction of it, we can calculate vector of the force. For instance, let the direction of the force which is imposed to A_j be from A_i to A_j and the magnitude of it be f_{ij}. The unit vector from A_i to A_j, u, is

$$u = \frac{((x_j - x_i), (y_j - y_i), (z_j - z_i))}{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}}$$

So, the vector of force is $f_{ij}.u$

Now, for the calculation of the vector of force imposed to atom A_i, we first determine neighborhood of A_i by using Delaunay algorithm and then calculate the vectors of forces which imposed to A_i from each of the neighbors and consider the resultant of these as the vector of force imposed to A_i.

Energy and force model

In the energy model, for a given 3D structure of a protein, the total potential of mean force, G, is

$$G = \frac{1}{2}\sum_{i,j} \Delta E_k^{ij}(l),$$

where the summation is over all pairs of atoms and in the force model, the norm of the resultant of forces imposed to each atom is calculated. For the calculation of the resultant of the forces imposed to each atom we determine the neighboring atoms of any atom by Delaunay tessellation procedure. For every pair of atoms at any sequence separation we consider an interval with radius 6 Å around the minimum value as a valid region and then we exclude the neighbor atoms with distances out of the valid region. Finally, we calculate the force imposed to each atom by the remaining atoms. Therefore, in the force model an n x l matrix denoted by F is obtained for decoy set containing n structures which each structure consisting of l atoms.

We know from mechanics that a system S which is composed of n particles is at equilibrium, if the force imposed on each particle be zero. Thus if we consider a protein as a system of atoms and calculate the force on each atom, it is expected in a strict physical law that the force imposed to each atom in the native structure should be zero. Since the protein is not rigid, this assumption is unrealistic. But we can assume that the structure S is a good approximation of the native structure, if the force on each atom be close to zero. In order to identify the native structure from decoys based on this assumption we consider the following procedure.

1. The aforementioned F matrix is calculated.
2. We find the minimum force imposed to each atom, for example, i'th atom, among the n structures. Let m_i be this minimum (m_i is the minimum of the values in column i) and k_i be the number of structures that the force imposed to their ith atom is equal to m_i. In order to assign larger scores to those atoms with minimal force, we add $s_i = (1/k_i)$ to their corresponding score. Therefore, s_i for larger number of structures getting minimum force is lower than s_i for fewer numbers of structures.
3. The total score for each structure is the summation of s_i over all atoms.
4. We calculate the median of scores and omit the structures with scores lower than median. Now, we have a new set of structures.
5. We repeat Steps 2, 3, and 4 twice.
6. The highest score among the remaining structures is expected to be the native structure.

In the energy model, for each structure in the decoy sets, including the native state and decoy structures, the total potential G is calculate. It is expected that the structure with the lowest energy among decoys be the native structure.

RESULTS AND DISCUSSION

In the previous study,[26] we considered a protein as a collection of springs and assumed that the force function

Table 1

The Averages and the Variances of the Forces on the Atoms in Three Secondary Structures; Alpha, Beta, and Coil

Secondary structure	Average	Variance
Alpha	2.0818	3.0874
Beta	1.9931	2.6882
Coil	2.1257	3.433

Table 2
The Performance of the Force and the Energy Models on Decoy Sets

	Force model			Energy model			
Decoy sets	Top 1 (*Z*-score[a])	Top 2	Top 5	Top 1 (*Z*-score[a])	Top 2	Top 5	#Targets
Lmds	10 (−34.68)	10	10	1 (−1.98)	3	3	10
lattice-ssfit	7 (−18.6)	8	8	7 (3.71)	8	8	8
Fisa	3 (−9.67)	4	4	2 (0.87)	2	2	4
fisa-casp3	3 (−6.3)	3	3	5 (3.75)	5	5	6
4state-reduced	4 (−4.18)	5	5	4 (2.81)	4	5	7
ig-structal	53 (−4.65)	59	61	27 (1.27)	29	34	61
ig-strudtal-hire	19 (−7.67)	20	20	15 (2.59)	15	16	20
hg-structal	26 (−6.12)	27	−29	16 (1.69)	17	21	29
Moulder	18 (−11.98)	19	20	20 (4.69)	20	20	20
ROSETTA	26 (−3.38)	33	37	0 (−2.9)	0	0	59
I−TASSER	39 (−9.43)	44	50	38 (1.8)	41	42	56
#Total	208 (−10.6)	232	247	135 (1.66)	144	156	280

[a]The numbers in parentheses are the average *Z*-scores.

follows Hook's law. As shown in Figure 1 the potential function is asymmetric with different rate of change of energy around the minimum energy. A simple method to overcome this problem is using numerical differentiation to estimate the force function from the values of potential function. In the force model, we expect a structure with the greatest number of atoms at relative minimum force to be the native structure. In the Step 2 of the algorithm comparing the corresponding force imposed to atom i of each structure results in finding the minimum value of the force. It is well known that long loops tend to be more flexible than regular secondary structures such as helices and strands. Therefore, it is expected that the force imposed to atoms in the secondary structures elements, helices and strands, should be different from atoms in coil. Even the force imposed to atoms in helices must be different from atoms in strands. To examine the hypothesis of variation among these three groups we use the Kruskal Wallis one way analysis of variance. The results of Kruskal Wallis test show the significant differences among groups (P-value is < 0.01). The averages and the variances of the forces on three structures alpha; beta and coil are shown in Table I.

Therefore, to compare the forces, the standard force imposed to each atom is calculated using:

$$f_s(a,s) = \frac{f(a,s) - \mu_s}{\sigma_s},$$

where $f(a, s)$ is the force impose to an atom a in secondary structure s, and μ_s and σ_s are the average and the standard deviation of forces in the secondary structure s, respectively.

Validation of the energy and the force model

Eleven decoy sets, including the *4state_reduce*,[45] *fisa_casp*,[46] *fisa*,[47] *lattice_ssfit*,[48] *lmds*,[49] *hg_structal, ig_structal_hires, ig_structal, moulder*,[50] *ROSETTA*,[51] and *I-TASSER*[52] decoy sets are used to evaluate the performance of the model. These decoys contain different number of structures and are made with different methods and are very appropriate for assessment of the model. The

Table 3
CC and RMSD of Structure Selected in Absence of Native Structure

Decoy Set	#Decoy	RMSD	CC	Decoy Set	#Decoy	RMSD	CC	Decoy Set	#Decoy	RMSD	CC
Lmds				ITasser				ITasser			
1b0nB	498	3.39	−0.62	1cewI	454	4.55	−0.26	1sro	517	4.23	−0.3
4pti	344	6.60	−0.49	1csp	317	3.39	−0.33	1ten	296	2.12	−0.26
1ctf	496	7.10	−0.31	1dtjA	287	2.93	−0.2	256bA	508	3.44	−0.4
1bba	501	5.33	−0.34	1egxA	354	2.65	−0.32	1ah9	512	3.27	−0.21
1dtk	216	8.44	−0.65	1fadA	516	3.61	−0.54	1jnuA	271	3.49	−0.23
1fc2	501	4.56	−0.5	1g1cA	309	2.57	−0.13	1kviA	552	2.09	−0.34
1shf-A	437	9.43	−0.5	1gpt	471	4.42	−0.29	1mkyA	287	4.22	−0.21
2cro	501	10.3	−0.44	1gyvA	339	3.31	−0.26	1mla	337	2.54	−0.18
1igd	501	7.81	−0.38	1orgA	444	2.78	−0.32	1npsA	471	2.29	−0.2
2ovo	348	9.07	−0.56	1sfp	310	5.52	−0.65	1b72A	536	5.34	−0.17

Table 4
Assessments of Different Scoring Functions by the Rank Native Structure in Eight Decoy Sets

Decoy sets	DFIRE	DOPE	RWplus	Force	#Target
4state-reduced	6	7	6	4	7
Fisa	3	3	3	3	4
Fisa-casp3	3	3	4	3	5
Lmds	7	7	7	10	10
Lattice-ssfit	8	8	8	7	8
Moulder	19	19	19	18	20
ROSETTA	22	21	20	26	58
I-TASSER	47	30	56	39	56
#Total	115	98	123	110	168

performances of the energy and the force model are compared in Table I. The results are presented in terms of the number of the first ranked (Top 1), the second ranked (Top 2), and the fifth ranked (Top 5) native structures within the decoy sets and the average *Z*-scores. The *Z*-score of the native structure in the decoys set is equal to

$$Z - \text{score} = \frac{\langle \text{score}_{\text{decoys}} \rangle - \text{score}_{\text{native}}}{\sigma_{\text{decoys}}},$$

in which $\text{score}_{\text{native}}$ is the score calculated for native structure and $\langle \text{score}_{\text{decoys}} \rangle$ and σ_{decoys} are the average and the standard deviation of scores distributions of decoys proteins, respectively. It is noticeable that *Z*-score in the force model is calculated based on the average and standard deviation of scores for remaining decoy structures mentioned in Step 6 of the algorithm.

Detection of the native structure using the energy model is mediocre. As shown in Table II, the energy model works well on decoy sets *lattice-ssfit, fisa-casp3*, and *moulder*, but the performance of the energy model on the other decoy sets is not as efficient as the force model. For decoy sets *lattice-ssfit* and *4state-reduced* the performance of the energy model in decoy discrimination is the same as the force model whereas for decoy sets *fisa-casp3* and *moulder* the performance in decoy discrimination of the energy model is better than the force model, but corresponding *Z*-scores in the force model are very high. This indicates that there is a high gap between the score of the native structure and the average scores of decoys for the force model.

For the remaining decoy sets (which the performance of the force model is indicated by boldface numbers in

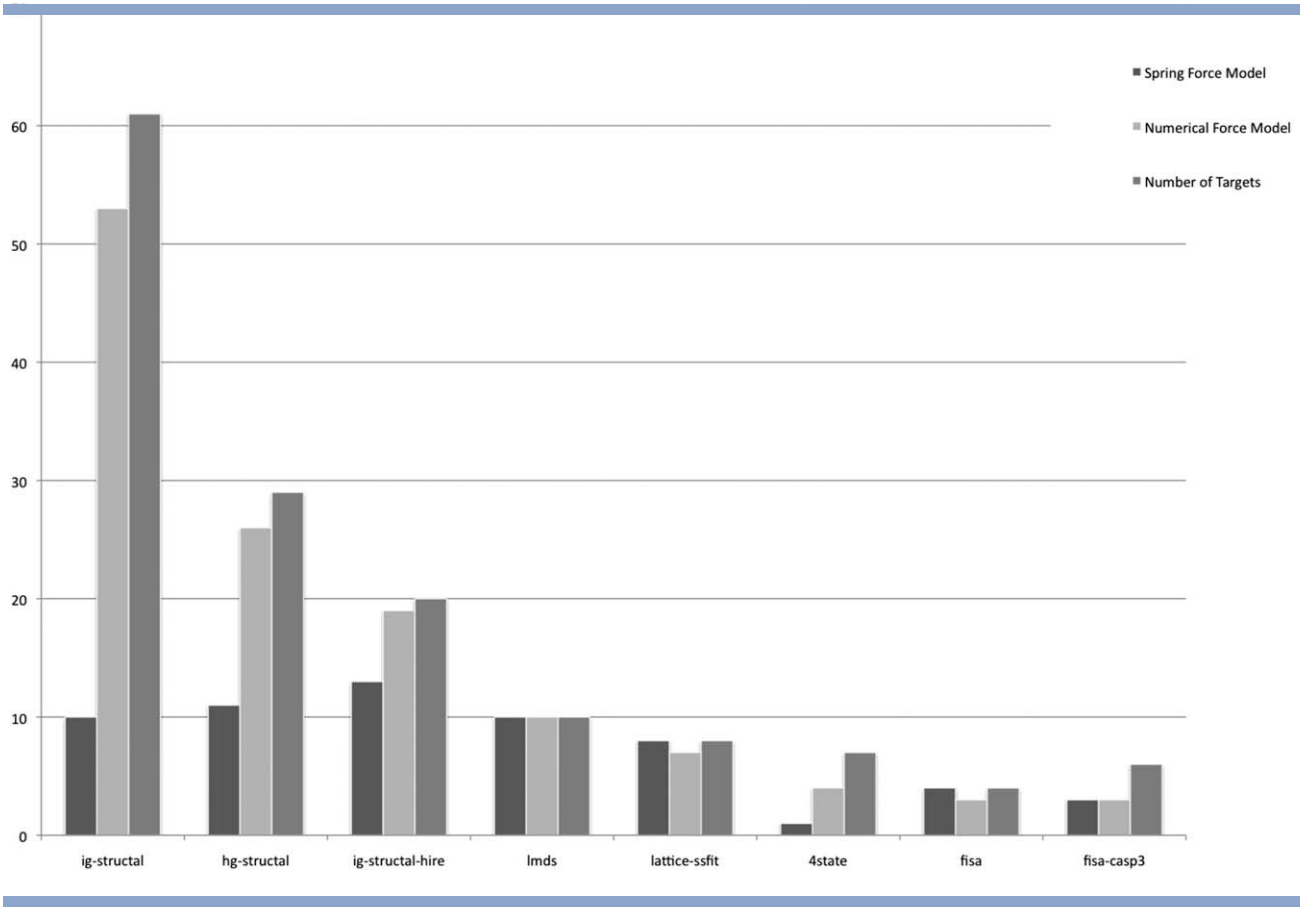

Figure 3
Assessment of numerical force model versus spring force model.

Table II) the results of the force model is significantly better than the energy model, especially for decoy sets *lmds*, *ig-structal*, *hg-structal*, and *ROSETTA*. Furthermore, if the success is defined by ranking the native structure as one of the two or five highest score (TOP2 and TOP5), the performance of the force model is remarkably better than the energy model. The force model correctly identified 208 native targets in 11 independent decoy sets with a success rate of 74%. This success increases up to 82 and 88% for TOP2 and TOP5, respectively while the energy model correctly identified 135 native targets with a success rate of 48% and the results of TOP2 and TOP5 increase only up to 51 and 56%, respectively.

For some decoy sets including ROSETTA and I-TASSER, detection of the native structure is mediocre, but the average *Z*-scores for all decoy sets are high. There are some reasons for failure in discrimination of native structure using knowledge-based potential functions, as discussed by Shen and Sali.[24] In the present case, we believe that the force model can be modified by using an appropriate reference state and atom type definition.

For data sets where the whole range of RMSD is reported, the correlation coefficient (CC) between RMSD and score should be negative and significantly different from 0. RMSD of selected structure in the absence of the native structure and also correlation coefficient between RMSDs and scores in some decoy sets are listed in Table III. The structure selected in absence of the native structure in some cases is not near native, but it is noticeable that the CC between RMSD and score is negative and significantly different from 0 in almost all decoy sets.

We compared the force model to the three previously published scoring functions, including DFIRE,[23] DOPE,[24] and RWplus.[53] Table IV shows the performance of different methods together with the force model in recognizing the native structures from decoys in several decoy data sets. In comparison to the other scoring functions, the force model does not work well with the *4-state reduced* and *I-TASSER*, but does well with *lmds*. None of the other previously published scoring functions based on energy in Table IV acquire such performance on *lmds*, while the force model can discriminate all of the 10 native structures. The overall ability of recognizing the native structure for the force model on decoy sets including *fisa, fisa-casp3, lattice-ssfit, moulder*, and *ROSETTA* does not differ considerably from other scoring functions; however, the performance of the force model on seven decoy sets, ignoring the results of *I-TASSER* is more efficient from the other scoring functions. In fact the force model identifies 71 native structures while other scoring functions identify at most 68 native structures. However, the force model correctly identifies 110 native structures from 168 targets in all eight decoy sets.

CONCLUSION

In this study, we have extracted the force function from knowledge-based potential function. This force is calculated through the numerical differentiation of mean force potential and finally the force imposed to each atom from its neighboring atoms is calculated. Considering a protein as a system of particles we assessed that in native fold, more residues should be at equilibrium or near equilibrium. Introducing a scoring function, we evaluate it on several decoy sets to argue its ability to discriminate the native fold from decoys. In the previous study,[26] we considered an ideal spring between pair of atoms, but in this study we introduced a more accurate and reasonable force function by numerical derivation of the potential function. Results in Figure 3 show that the force model based on numerical differentiation performs more efficient than the spring force model in detecting the native structure. This model can be applied in *ab initio* protein structure prediction, fold assignment, sequence structure alignment and template selection and in molecular dynamics and normal mode analysis additionally.

REFERENCES

1. Anfisen CB. Principles that govern folding of protein chains. Science 1973;181:223–230.
2. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106: 765–784.
3. Weiner SJ, Kollman PA, Nguyen DT, Case DA. An all atom forcefield for simulations of proteins and nucleic-acids. J Comput Chem 1986;7:230–252.
4. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins 2006;65:712–725.
5. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. J Comp Chem 1983;4: 187–217.
6. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, van Gunsteren WF. The GROMOS bio molecular simulation program package. Phys Chem 1999;103:3596–3607.
7. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 1996;118:11225–11236.
8. Mirzaie M, Sadeghi M. Knowledge-based potentials in protein fold recognition. J Paramed Sci 2010;1:65–75.
9. Miyazawa S, Jernigan RL. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromole 1985;18:534–552.
10. McConkey BJ, Sobolev V, Edelman M. Discrimination of native protein structures using atom-atom contact scoring. Proc Natl Acad Sci USA 2003;100:3215.
11. Pokarowski P, Kloczkowski A, Jernigan RL, Kothari NS, Pokarowska M, Kolinski A. Inferring ideal amino acid interaction forms from statistical protein contact potentials. Proteins 2005;59:49.
12. Zhang C, Kim SH. Environment-dependent residue contact energies for proteins. Proc Natl Acad Sci USA 2000;97:2550.
13. Solis AD, Rackovsky S. Information and discrimination in pairwise contact potentials. Proteins 2008;71:1071–1087.

14. Sippl MJ. Boltzmann's principle, knowledge based mean force and protein folding. An approach to the computational determination of protein structures. J Comput Aided Mol Des 1993;7:473–501.
15. Melo F, Feytmans E. Assessing protein structures with nonlocal atomic interaction energy. J Mol Biol 1998;277:1141–1152.
16. Sippl MJ. Calculation of conformational ensembles from potentials of mean force: an approach to the knowledge-based prediction of local structures in globular proteins. J Mol Biol 1990;213:859–883.
17. Tobi D, Elber R. Distance dependent, pair potential for protein folding: results from linear optimization. Proteins 2000;41:40–56.
18. Jones DT, Taylor WR, Thornton JM. A new approach to protein fold recognition. Nature 1992;358:86–89.
19. Samudrala R, Moult J. An all atom distance dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol 1998;275:895–916.
20. Lu H, Skolnick J. A distance dependent atomic knowledge based potential for improved protein structure selection. Proteins 2001;44:223–232.
21. Melo F, Feytmans E. Novel knowledge-based mean force potential at atomic level. J Mol Biol 1997;267:207.
22. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure derived potentials of mean force for structure selection and stability prediction. Protein Sci 2002;11:2714–2726.
23. Chi Zhang, Song Liu, Hongyi Zhou, Yaoqi Zhou. An accurate, residue-level, pair potential of mean force for folding and binding based on the distance–scaled, ideal–gas reference state. Protein Sci 2004;13:400–411.
24. Shen M, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci 2006;15:2407–2524.
25. Ferrada E, Vergara IA, Melo F. A knowledge-based potential with an accurate description of local interactions improves discrimination between native and near-native protein conformations. Cell Biochem Biophys 2007;49:111–124.
26. Mirzaie M, Eslahchi C, Pezeshk H, Sadeghi M. A distance dependent atomic knowledge based potential and force for discrimination of native structures from decoys. Proteins 2009;77:454–463.
27. Rooman MJ, Kocher JPA, Wodak SJ. Prediction of protein backbone conformation based on seven structure assignments. Influence of local interactions. J Mol Biol 1991;211:961–979.
28. Fang Q, Shortle D. A consistent set of statistical potentials for quantifying local side-chain and backbone interactions. Proteins 2005;60:90.
29. Pohl FM. Empirical protein energy maps. Nat New Biol 1971;234:277.
30. Bagci Z, Kloczkowski A, Jernigan RL, Bahar I. The origin and extent of coarse-grained regularities in protein internal packing. Proteins 2003;53:56–67.
31. Buchete NV, Straub JE, Thirumalai D. Orientational potentials extracted from protein structures improves native fold recognition. Protein Sci 2004;13:862.
32. Buchete NV, Straub JE, Thirumalai D. Continuous anisotropic representation of coarse-grained potentials for proteins by spherical harmonics synthesis. J Mol Graph Model 2004;22:441.
33. Mukherjee A, Bhimalapuram P, Bagchi B. Orientation-dependent potential of mean force for protein folding. J Chem Phys 2005;123: 014901.
34. Misura KM, Morozov AV, Baker D. Analysis of anisotropic side-chain packing in proteins and application to high-resolution structure prediction. J Mol Biol 2004;342:651–664.
35. Miyazawa S, Jernigan RL. How effective for fold recognition is a potential of mean force that includes relative orientations between contacting residues in proteins? J Chem Phys 2005;122:024901.
36. Wu Y, Lu M, Chen M, Li J, Ma J. OPUS-Ca: a knowledge based potential function requiring only Calpha positions. Protein Sci 2007;16:1449–1463.
37. Sippl MJ. Knowledge-based potentials for proteins. Curr Opin Struct Biol 1995;5:229–235.
38. Covell DG. Folding protein alpha-carbon chains into compact forms by Monte Carlo methods. Proteins 1992;14:409–420.
39. Sun S. Reduced representation model of protein structure prediction: statistical potential and genetic algorithms. Protein Sci 1993;2:762–785.
40. Singh RK, Tropsha A, Vaisman II. Delaunay tessellation of proteins:Four body nearest neighbor propensities of amino acid residues. J Comput Biol 1996;3:213–221.
41. Munson PJ, Singh RK. Statistical significance of hierarchical multi body potentials based on Delaunay tessellation and their application insequence structure alignment. Protein Sci 1997;6:1467–1481.
42. Barber CB, Dobkin DP, Huhdanpaa H. The quickhull algorithm for convex hulls. ACM Trans Math Software 1996;22:469–483.
43. Lovell S, Davis I, Arnedall W, de Baker P, Word J, Prisant M, Richardson J, Richardson D. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins 2003;50:437–450.
44. Ferrada E, Melo F. Nonbonded terms extrapolated from nonlocal knowledge-based energy functions improve error detection in near-native protein structure models. Protein Sci 2007;16:1410–1421.
45. Park B, Levitt M. Energy functions that discriminate X-ray and near native folds from well-constructed decoys. J Mol Biol 1996; 258:367–392.
46. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 1997;268:209–225.
47. Simons KT, Ruczinski I, Kooperberg C, Fox B A, Bystroff C, Baker D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 1999;34:82–95.
48. Xia Y, Huang ES, Levitt M, Samudrala R. Ab initio construction of protein tertiary structures using a hierarchical approach. J Mol Biol 2000;300:171–185.
49. Keasar C, Levitt M. A novel approach to decoy set generation: designing a physical energy function having local minima with native structure characteristics. J Mol Biol 2003;329:159–174.
50. John B, Sali A. Comparative protein structure mideling by iterative alighnment, model building and model assessment. Nucleic Acid Res 2003;31:3982–3992.
51. Das R, Qian B, Raman S, Vernon R, Thompson J, Bradley P, Khare S, Tyka MD, Bhat D, Chivian D, Kim DE, Sheffler WH, Malmström L, Wollacott AM, Wang C, Andre I, Baker D. Structure prediction for CABP7 targets using extensive all-atom refinement with Rosetta@home. Proteins 2007;69:118–128.
52. Wu S, Skolnick J, Zhang Y. Ab initio modeling of small proteins by iterative TASSER simulations. BMC Biol 2007;5:17.
53. Zhang J, Zhang Y. A novel side chain orientation dependent potential derived from random walk reference state for protein fold selection and structure prediction. PloS ONE 2010;5:e15386.