Title

Scoring protein-protein docked structures based on the balance and tightness of binding

Journal

Journal of Computer-Aided Molecular Design, 18(4)

ISSN

0920-654X

Authors

Yu, Y H

Lu, B Z

Han, J G

et al.

Publication Date

2004-04-01

Peer reviewed

Journal of Computer-Aided Molecular Design **0:** 1–9, 2004.

Scoring protein–protein docked structures based on the balance and tightness of binding

Y.H. Yu[a], B.Z. Lu[b,*], J.G. Han[b] & P.F. Zhang[b]

[a]*Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093-0365, USA;* [b]*The National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei, Anhui 230022, China*

Received 2 June 2003; accepted in revised form 6 June 2004

Key words: balance and tightness of binding, protein–protein docking, scoring function

Summary

One main issue in protein–protein docking is to filter or score the putative docked structures. Unlike many popular scoring functions that are based on geometric and energetic complementarity, we present a set of scoring functions that are based on the consideration of local balance and tightness of binding of the docked structures. These scoring functions include the force and moment acting on one component (ligand) imposed by the other (receptor) and the second order spatial derivatives of protein–protein interaction potential. The scoring functions were applied to the docked structures of 19 test targets including enzyme/inhibitor, antibody/antigen and other classes of protein complexes. The results indicate that these scoring functions are also discriminative for the near-native conformation. For some cases, such as antibody/antigen, they show more discriminative efficiency than some other scoring functions, such as desolvation free energy (ΔG_{des}) based on pairwise atom–atom contact energy (ACE). The correlation analyses between present scoring functions and the energetic functions also show that there is no clear correlation between them; therefore, the present scoring functions are not essentially the same as energy functions.

Introduction

The goal of predictive protein–protein docking is to obtain a near-native structure for the bound complex from the coordinates of the unbound component molecules [1]. Solving the docking problem involves two components: an efficient search procedure and a good scoring function. The development of a scoring function that can reliably distinguish correct docked structures from incorrect ones is a challenging topic of current research.

Presently it is well known that the geometric complementarity and energetic complementarity are two main factors in evaluating the potential solutions from the docked structures of a protein complex [2, 3]. From the early days of docking, it has been postulated and repeatedly reaffirmed that geometric matching plays an important role in determining the structure of a complex [4, 5]. The scoring function of early docking algorithms used practically exclusively geometric complementarity criteria. Current scoring functions frequently use additional criteria in combination with geometric complementarity [6–8]. However, there are cases where the correct solution does not possess the largest contact area, while the incorrect solutions display a better shape complementarity than the correct one [6]. On the other hand, from a thermodynamic point of view, the native protein–protein complex should be the structure with the lowest binding free energy. This leads to the second criterion in evaluating the docked structures: the binding energy or binding free energy of the protein complex. Some scoring functions involve solvation potentials, empirical atom–atom or residue–residue contact energies, and continuum electrostatics [9–15]. However, the empirical free energy and the molecular mechanics

*To whom correspondence should be addressed; present address: Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093-0365, USA. E-mail: blu@mccammon.ucsd.edu

potential alone cannot provide a valid discrimination between native and misdocked structures [16, 17]. One possible reason is that the molecular mechanics potential is just part of the binding free energy, and the entropy is not taken into account. Some elegant free-energy simulations or calculations may be a reliable discrimination to check the solutions [15], such as MMPBSA (or MMGBSA) [18, 19], or the free energy perturbation method [20]. However, the calculation of binding free energy is complicated and is still the subject of research, which makes it impractical to use such an approach in protein–protein docking. Also some other scoring functions, such as that based on pairwise atom–atom potential functions [12], can be effective, but they are limited to the cases of enzyme–inhibitor complexes [11, 14]. In addition, detailed molecular dynamics calculation is also used, but knowledge of the location of the binding site is absolutely essential [3].

Here, we try to construct some novel scoring functions from the consideration of rigid body mechanics. When the two molecules bind to the determinate active sites and in the 'correct' orientation to form a native or near-native conformation, the complex should locate at the local energy minimum, and then the force and moment opposed on each molecule are zero. We refer to this situation as being in 'balance'. However, the docked structures, even the 'correct' docked structures, are not in the rigorous local minima, even after removing some local clash by performing a short time energy minimization as was done in Ref. 21. We suppose that the near-native docked structures should also be in near balance, which means that the force and moment acting on each molecule should be small. Moreover, enlightened from the picture of a funnel-like energy landscape [22, 23], there exist many local funnel-like energy landscapes in the six-dimensional docking space. The native state (or binding) is trapped into the bottom of the deepest energy funnel that corresponds to the lowest free energy, which means it is more difficult for the complex to overcome the energy barrier to dissociate relative to other binding sites. Here, we refer to this case as 'tight binding'. It is still a concern to define and quantitatively describe the tightness of binding. Generally, it can be assumed that the deeper the funnel, the steeper the energy surface, and consequently, there should be larger second order spatial derivatives of the interaction potential, while the first derivatives are equal to zero at the local energy minimum. Moreover, the energy funnel represents a competition result between entropy and interaction potential. The entropy loss becomes large at the funnel bottom, which is consistent with the above tightness point of view. Because in the case of large second derivatives of the potential, there is only narrow room for the movement (vibration) of the ligand, which results in decrease of entropy. Therefore, the new scoring functions can be constructed using the calculations of force, moment, and the second spatial derivatives, which are used as a measure for the balance and tightness of binding in the present context. However, because the above functions just describe the local properties of the energy surface, the efficiency of these criteria lies on the level of how well the local property can reflect the global property of the energy landscape. On a smoother energy surface, corresponding to a simpler interaction model, the local surface property is more likely to reflect the tendency of the global property. To sum up, we use the second derivative instead of directly looking at the actual energy to find the deep funnel, because a single energy calculation for one docked conformation cannot decide if the energy is at the deep funnel, in other words, the energy value only depends on the unique conformation, while the second derivative reflects the property of the local energy surface and thus contains some energy information of its near conformations. As mentioned above, the binding energy surface may have a lot of local energy funnels; the second derivative is designed to, if possible as an assumption, find the deep funnel. In fact, the docking problem is to find the deepest free energy funnel, with the lowest free energy. Therefore, the energy criterion and its derivatives may be used complementarily. In this work, we used energy criteria as filters at first, and then use the energy derivatives to rescore the docked structures, which is a try to incorporate the above idea. The method was applied to a test set of protein complexes. Comparison was also made among the ranking results using the conventional electrostatic binding energy (ΔE_{ele}), desolvation free energy $\Delta G_{des}(ACE)$ [13], and one of our scoring functions, which showed the good ability of our scoring function to discriminate the near-native structures from other misdocked ones. Also, the analysis of the correlation of different scoring functions showed that these scoring functions were not essentially the completely the same as the energy function.

Materials and methods

Scoring functions for two-stage ranking

In this work, a two-stage ranking strategy was used, in which the traditional scoring was applied to rapidly scan possible docked solutions and obtain initial 'good' candidates, followed by our new scoring functions to further discriminate the limited conformations. The interface area, desolvation free energy $\Delta G_{des}(ACE)$ based on the atomic contact energy (ACE) [13], and electrostatic binding energy ΔE_{ele} were combined as a triple filter in the first stage. The charge parameters were from the CHARMM 19 force field [24]. The electrostatic energy was calculated by the Coulombic potential function with a distance-dependent dielectric $\varepsilon = 4r$, which was similarly treated in other works [21, 27]. The distance-dependent dielectric was normally used to take into account the screening effect of solvent.

In the second stage, we aimed to construct scoring functions to evaluate and rank the candidates. The mechanical force and moment acting on one component molecule in the complex were calculated (the same but with opposite signs), and the second derivatives of the interaction potential were calculated also. From the above section, it was reasonably supposed that the force and moment should be small for the near-native structure, while the second order spatial derivatives should be large. In this work, the interaction potential between the two partner molecules involed Coulombic electrostatic energy and van der Waals energy. With a soft-core Coulombic potential function, the interaction potential of the complex was:

$$\mathrm{V} = \sum_{\mathrm{i}\in\mathrm{m1}}\sum_{\mathrm{j}\in\mathrm{m2}}\left(\frac{\mathrm{Q_i Q_j}}{16\pi\varepsilon_o(\mathrm{r_{ij}}+\mathrm{c})^2} + \frac{\mathrm{C_{12}}}{(\mathrm{r_{ij}}+\mathrm{c})^{12}} - \frac{\mathrm{C_6}}{(\mathrm{r_{ij}}+\mathrm{c})^6}\right) \quad (1)$$

where i and j denote the atoms of the component molecules, r_{ij} is the distance between atoms i and j, Q_i, Q_j are their point charges, C_{12} and C_6 are the van der Waals parameters dependent on atoms i and j, and $m1$ and $m2$ number the two component molecules, respectively. The first term in Equation 1 takes into account the Coulombic potential with a distance-dependent dielectric $\varepsilon = 4r_{ij}$, and the last two terms account for the van der Waals potential. The parameter c is a dampening constant added to the distance separating both nuclei to lower the unrealistically high interaction potentials, which are due to the allowance of limited interpenetration of positions of both molecules in the docking search.

For a docked structure, we then can calculate the absolute values of mechanical quantities: $F = |\nabla V|$, $M = \left|\sum_{j\in m2}(r_j - r_c)\times\nabla_j V\right|$, where F and M are the magnitudes of force and moment acting on $m2$ (molecule 2) by $m1$ (molecule 1), $\boldsymbol{r}_j$ is the displacement of the j-th atom of $m2$, and $\boldsymbol{r}_c$ is the geometric center of m2. The gradient operator ∇j is relative to the displacement $\boldsymbol{r}_j$ in Cartesian space. From the formula, the moment M acting on molecule 2 was relative to its center. For the tightness of binding, there were six degrees of freedom to be considered, since the ligand could both translate and rotate. The second order partial derivatives relative to three translational directions constructed a 3×3 symmetric matrix $[\frac{\partial^2 V}{\partial x_i \partial x_j}]$ (x_i or x_j denote any one of the x, y, z directions), and the second derivatives in the three rotational directions constructed another 3×3 matrix $[\frac{\partial^2 V}{\partial\theta_i\partial\theta_j}]$ (θ_i or θ_j denote any of the rotational angles θ, ϕ, ψ). Since the elements in these two matrices had different units, they needed to be factorized in order to form a unified measure. If, expanding the potential at the local minimum, the first derivatives of the potential equaled zero then, the equations of translational and rotational motions became

$$[M_{ij}]\ddot{\vec{r}} = -[\frac{\partial^2 V}{\partial x_i \partial x_j}],$$

and

$$[I_{ij}]\ddot{\vec{\theta}} = -[\frac{\partial^2 V}{\partial\theta_i\partial\theta_j}],$$

respectively, where the square brackets denote a 3×3 matrix, $[M_{ij}]$ is the diagonal mass matrix, $[I_{ij}]$ is the matrix of the moment of inertia, and $\vec{\Delta r}$ and $\vec{\Delta\theta}$ are the translational and rotational displacement vector relative to the local minimum. Therefore, $[\frac{\partial^2 V}{\partial x_i\partial x_j}]$ was multiplied by the inverse of $[M_{ij}]$, and $[\frac{\partial^2 V}{\partial\theta_i\partial\theta_j}]$ was multiplied by the inverse of $[I_{ij}]$. After this, the two second derivative matrices can be compared. To obtain a unique value from the two matrices to measure the binding tightness of the docked structure, the two second order derivative matrices after factorization were then diagonalized. We selected two functions: the maximum marked as *VDMAX* of the 6 diagonal elements in the two matrices after diagonalizing, and the sum of the arithmetic mean and geometric

mean of the 6 diagonal elements, which is marked as *VDMEAN*.

When the docked structure was near the native, the F,M should be the smallest, and *VDMAX*, *VDMEAN* should be large enough, which were assumed to be the largest among the docked structures in this work. Therefore, F, M, *VDMAX*, and *VDMEAN* could be used as four new scoring functions. The essence of this set of scoring functions is to find the docked structures with the smallest moment and force, and the largest second derivatives of the interaction potential. However, it should be noted that these criteria are just the necessary or supposedly necessary conditions of the near-native conformation, they are not sufficient to identify the near-native structures. These criteria may be satisfied at the minimum of the interaction potential. Therefore, these new scoring functions can only be expected to find a list of potential candidates, instead of just one, for the near-native docked structures, which was also the normal expectation as in a usual docking procedure. In fact, this can be considered as an avoidance of the open problem to search the global energy minimum.

A note should also be made here that the solvation energy term would indeed affect the position of the local minimum on the potential of the energy surface of the complexes. In the presently constructed scoring functions, however, except for the electrostatic screening effect represented by the distance-dependent dielectric in Equation 1, we have not performed a full solvation calculation (such as PB or GB) to account simultaneously for electrostatic and solvation effects. There are two reasons; one is that the electrostatic interaction is the dominant interaction between molecules and is more sensitive to the relative position of the two molecules than, for example, a SASA related solvation term; the other one is that those calculations will introduce mathematical complexity for the calculation of the spatial derivatives we considered in this work. Our goal is to find some simple and fast criteria for a large amount of docked structures.

Docked structures

To produce the docked structures, we used the docking algorithm of Wodak and Janin [25] implemented in the program DOCK [26], in which some side chain flexibility was allowed [27]. The use of a simplified model of a protein and the 'softness allowance' in the docking procedure compensates for certain conformational changes that may take place upon binding. The six parameters that defined the position and orientation of one molecule relative to the other were five Euler rotation angles (θ_1, ϕ_1, θ_2, ϕ_2 and χ) and an intermolecular distance ρ. Angles θ_1 and ϕ_1 locate the center of the ligand relative to the receptor; θ_2 and ϕ_2 locate the center of the receptor relative to the ligand; χ is a spin angle about the center line. The five angles were systematically searched in steps of 7.5°. We explored the full range of θ_2 ($\pm 90°$), ϕ_2 and χ ($\pm 180°$), that is to say, the full surface of the ligand. For the receptor, we restricted the search range of θ_1 and ϕ_1 to $\pm 30°$. With a 7.5° step, about 4.86×10^6 different binding modes were generated for each complex. One target T06 (1KXQ) of CAPRI Round 2 (http://capri.ebi.ac.uk) and 18 other protein–protein complexes (from http://zlab.bu.edu/~rong/dock/index.shtml) were selected as the test set (see Table 1). They were chosen from complexes of different types, including enzyme/inhibitors, antibody/antigens and other complexes. In Table 1, XX added after the complex PDB code refers to the cases of bound docking, in which the structures of receptor and ligand were from the co-crystallized structure. If the complex was reconstructed from two component structures, one of which was from the complexed protein and the other from the free form, FX or XF is added after the PDB code, where F and X denote the free form and co-crystallized form, respectively. If the complex is reconstructed from both proteins of the free form, FF is added to the PDB code. The third column of Table 1 lists the source PDB files from which the structures of the receptor and ligand were taken.

Filtering

During the search, all docked complexes with an interface area less than 600 Å^2 were eliminated. In the following filtering, a number (300–500) of binding modes with the largest interface areas were retained, then a number (300–500) of those with the lowest electrostatic energy were added, and finally followed by the addition of hundreds of conformations with the lowest $\Delta G_{des}(ACE)$. Then the retained candidate solutions were clustered by the method of Cherfils et al. [26], keeping the average position in each cluster. The reason to use separate filters instead of a combined filter such as electrostatic + solvation energy is due to the consideration that different types of complexes may favor different types of protein–protein interaction or complementarities, e.g. electrostatic complementarity or geometric complementarity. For different

Table 1. The selected 19 protein–protein complexes used to test the scoring functions.

Dock name	Description	PDB Code
Enzyme/inhibitor		
1CHOFF	Alpha-chymotrypsin/Ovomucoid 3rd domain	5cha, 2ovo
2SICXX	Subtilisin BPN/Subtilisin inhibitor	2sic
1TGSXX	Trypsinogen/Pancreatic secretory trypsin inhibitor	1tgs
1BRSFF	Barnase/Barstar	1a2p, 1a19
2PTCXX	Beta-trypsin/Pancreatic trypsin inhibitor	2ptc
2SNIFF	Subtilisin Novo/Chymotrypsin inhibitor 2	1sup, 2ci2
Antibody/antigen		
1AHWFF	Antibody Fab 5G9/Tissue factor	1fgn, 1boy
1BVKFF	Antibody Hulys11 Fv/Lysozyme	1bvl, 3lzt
1BQLXF	Hyhel-5 Fab/Lysozyme	1bql, 1dkj
1JHLXF	IgG1 Fv Fragment/Lysozyme	1jhl, 1ghl
2VIRXF	Igg1-lamda Fab/Hemagglutinin	2vir, 2viu
1WEJFF	IgG1 E8 Fab fragment/Cytochrome C	1qbl, 1hrc
1NMBXF	Fab NC10/Neuraminidase	1nmb, 7nn9
1KXQXF	Antibody Vhh Fragment /Amylase	1kxq, 1pif
1QFUXF	Igg1-k Fab/Influenza/Hemagglutinin	1qfu, 2viu
Others		
1A0OFX	Che A/Che Y	1chn, 1a0o
1MDAFF	Methylamine dehydrogenase/Amicyanin	2bbk, 1aan
4INSXX	Pig insulin dimer	4ins
1IGCXF	IgG1 Fab fragment/Protein G	1igc, 1igb

complexes in our test, the number of clusters ranged from 200 to 1600. These solutions were used as the docked structures for the test of our scoring functions, to increase the chance of finding more native-like conformations.

Results and discussion

First, we make a comparison of the discriminative ability of the commonly used scoring functions and one of our mechanical scoring functions. Table 2 gives the ranking results for the 19 complexes using four scoring functions ΔE_{ele}, $\Delta G_{des}(ACE)$, the sum of these two energies, and the maximum of second derivative *VDMAX*. For each complex, the best rank of the near-native docked structure and the corresponding root mean square deviation (RMSD) from the X-ray crystallographic complex are listed. Here, the near-native structures of the listed complexes are those with RMSD $<$ 4.0 Å. The calculations of the interaction potential use 0.5 Å as the damping constant c in Equation 1. For 5 out of 6 complex cases of enzyme/inhibitor, the scoring function of *VDMAX* gives better and higher ranking places for the near-native docked structures than the scoring function ΔE_{ele} does. The best results for the enzyme/inhibitor complexes appear to come from the desolvation $\Delta G_{des}(ACE)$. The exception is the barnase/barstar complex (1BRS), which is strongly charged and the desolvation in the bound structure is known to be repulsive [28]. Since the native conformation of this complex is determined almost exclusively by electrostatic interactions, it is not surprising that the ΔE_{ele} is able to give a rank within the top 100 ranking places, and the desolvation $\Delta G_{des}(ACE)$ is unable to identify the near-native structures. However, for this class of enzyme/inhibitors, $\Delta G_{des}(ACE)$ discriminates very well the near-native structures from the docked structures for the other 5 complexes. Hu et al. [29] have shown that in enzyme/inhibitor complexes, residues are more conserved at the interfaces than at other locations. This may explain why a pairwise potential function (such as ACE) derived from enzyme/inhibitors is more successful when applied to the same complex class. In contrast, for anti-

Table 2. The best rank of the near-native docked structure and the corresponding RMSD (Å) using scoring functions ΔE_{ele}, ΔG_{ACE}, ΔE_{tot} (sum of ΔE_{ele} and ΔG_{ACE}), and *VDMAX* for the 19 protein–protein complexes (dashes indicate that no near-native docked structure was ranked within the top 100).

Dock name	Scoring							
	ΔE_{ele}	RMSD	ΔG_{ACE}	RMSD	ΔE_{tot}	RMSD	*VDMAX*	RMSD
Enzyme/inhibitor								
1CHOFF	33	3.54	1	0.06	33	3.54	2	3.54
2SICXX	27	3.30	1	1.80	4	1.80	8	3.63
1TGSXX	25	0.37	1	0.37	2	0.37	14	2.98
1BRSFF	33	1.96	–	–	33	1.96	32	1.96
2PTCXX	80	0.66	28	0.66	80	0.66	4	1.31
2SNIFF	80	3.75	31	3.98	80	3.75	–	–
Antibody/antigen								
1AHWFF	1	2.66	2	2.66	2	2.66	59	2.66
1BVKFF	–	–	91	3.21	–	–	–	–
1BQLXF	–	–	–	–	–	–	–	–
1JHLXF	54	3.75	–	–	–	–	89	1.54
2VIRXF	65	2.25	–	–	47	3.75	–	–
1WEJFF	94	3.91	–	–	–	–	1	3.91
1NMBXF	–	–	–	–	–	–	25	3.44
1KXQXF	4	3.46	–	–	–	–	8	2.37
1QFUXF	–	–	6	2.19	–	–	44	2.19
Others								
1A0OFX	12	3.99	100	3.65	12	3.99	67	3.70
1MDAFF	–	–	5	2.46	72	3.90	5	3.81
4INSXX	18	2.76	12	0.44	4	2.76	42	2.76
1IGCXF	76	2.81	–	–	76	2.87	–	–

body/antigen complexes, the $\Delta G_{des}(ACE)$ and the sum of $\Delta G_{des}(ACE)$ and ΔE_{ele} fail to identify the near-native docked structures in most cases, while our *VDMAX* succeeds in ranking the near-native solutions for 6 out of 9 complexes, with 2 cases ranked within the top 10. It is noted here that the complex 1KXQ (Antibody Vhh Fragment/Amylase) was a target of CAPRI Round 2. Our *VDMAX* also gets the near-native conformation with a low RMSD of 1KXQ within the top 10. For the other class of complexes, the ranking results are comparable using the above four scoring functions. Because the third column is from the sum of ΔE_{ele} and $\Delta G_{des}(ACE)$, it is not surprising that the scoring results for most cases seem to be the average of the two individual results. As a whole, for our test set, the second derivative *VDMAX* identifies most complexes within the top 100 ranking places.

Table 3 summarizes the ranking results of near-native docked structures of the 19 complexes using our scoring functions: the moment M, the force F, the second derivative *VDMAX*, and a mean value of the second derivative *VDMEAN* (see details of the definitions in the Method section). The first two ranking columns in Table 3 show that for 6 cases of 1TGS, 1AHW, 1BVK, 1JHL, 1WEJ, and 1NMB complexes, the two scorings using M and F both fail to select the near-native conformations. And interestingly, for most of the other 13 cases, the two scoring functions also get ranking results on the same level and identify the same docked structures. This fact indicates that, in the calculation under the present frame, the force and moment seem to approach a similar level of balance for the native or near-native docked structures, which may be due to the close physical foundations of force and moment. Therefore, the two scoring criteria based on M and F may be substituted for each other. The last

Table 3. The best rank of the near-native docked structure and the corresponding RMSD (Å) using scoring functions M, F, *VDMAX*, and *VDMEAN* for the 19 protein–protein complexes (dashes indicate that no near-native docked structure was ranked within the top 100).

Dock name	Scoring							
	M	RMSD	F	RMSD	*VDMAX*	RMSD	*VDMEAN*	RMSD
Enzyme/inhibitor								
1CHOFF	1	0.06	1	0.06	2	3.54	24	2.06
2SICXX	3	3.44	12	1.80	8	3.63	6	3.63
1TGSXX	–	–	–	–	14	2.98	79	3.62
1BRSFF	–	–	95	2.29	32	1.96	–	–
2PTCXX	16	0.66	5	0.66	4	1.31	5	1.31
2SNIFF	54	2.53	45	2.53	–	–	47	3.64
Antibody/antigen								
1AHWFF	–	–	–	–	59	2.66	97	3.10
1BVKFF	–	–	–	–	–	–	–	–
1BQLXF	94	0.72	80	0.72	–	–	–	–
1JHLXF	–	–	–	–	89	1.54	45	1.54
2VIRXF	19	1.68	36	1.68	–	–	–	–
1WEJFF	–	–	–	–	1	3.91	21	3.91
1NMBXF	–	–	–	–	25	3.44	–	–
1KXQXF	12	3.77	17	3.77	8	2.37	5	2.37
1QFUXF	7	2.19	7	2.19	44	2.19	–	–
Others								
1A0OFX	99	3.74	74	3.74	67	3.70	33	3.70
1MDAFF	11	1.98	13	1.98	5	3.81	–	–
4INSXX	3	0.44	7	0.44	42	2.76	45	1.98
1IGCXF	14	3.17	14	3.17	–	–	6	3.17

two ranking columns using scoring functions *VDMAX* and *VDMEAN* are based on the tightness criteria of the binding of docked structures. They show different ranking pictures from the scoring functions M and F by comparing the highest rank and the corresponding RMSD of near-native conformations of the 19 protein complexes. For 1TGS, 1AHW, 1JHL, and 1WEJ, the scoring functions *VDMAX* and *VDMEAN* can identify the near-native docked conformations, while M and F fail. In the 19 complex set, there are 14 cases where *VDMAX* gets the highest ranking places of the near-native conformations within the top 100, and 12 cases where *VDMEAN* succeeds.

Considering that there are only a few near-native structures in the filtered docked structures by our procedure for some antibody/antigen complexes (e.g. there is only one for 1BQL, 1NMB), it is found in Table 3 that for each scoring function, M, F, *VDMAX* or *VDMEAN*, there are no obvious differences in the discrimination ability between different classes of protein complexes. Therefore, differing from $\Delta G_{des}(ACE)$, each of them may be used as a uniform criterion for ranking the docked conformations of all classes of protein complexes. However, as mentioned above, M and F often pick out the same near-native conformations (with the same RMSDs), while these conformations are often not the same as those identified by using the last two scoring functions *VDMAX* and *VDMEAN* (with different RMSDs). Also the conformations picked out by *VDMAX* and *VDMEAN* are often not the same one (see Table 3). First, these phenomena may be due to the different physics of the four quantities; M and F are related to the balance of the binding of two molecules, and *VDMAX* and *VDMEAN* are related to the binding tightness as supposed. Second, the criteria of M and F can sufficiently describe the condition for balance of the binding of two molecules; however, *VDMAX* and *VDMEAN* defined in this work are not sufficient to decide the 'binding tightness'.

Author, please supply new artwork

Figure 1a–f. Correlations between different scorings for the docked structures of 1KXQ. The binding potential V (including Coulombic and van der Waals interactions), G_{ACE}, and M are in kcal/mol, F is in kcal/mol/Å, *VDMAX* and *VDMEAN* are in kcal/mol/Å^2.

The following question now arises: are there any correlations in the scorings or rankings using above energetic functions and the proposed scoring functions, in other words, do they choose the same structures? Figures 1a–f give some correlation plots of between different pairs of scorings (rankings) for the case of complex 1KXQ. In this figure all the very large data are cut off, since the corresponding docked conformations have atomic overlaps. Figure 1a clearly shows that ranks by M and F have a close correlation, which just explains the phenomenon in Table 3 that M and F often choose the same docked structures. Figure 1b shows that the scorings using the interaction potential V and$M$ have a correlation to some extent, but they do have some difference in the identifications of near-native structures. However, the above cases do not occur in the scorings by *VDMEAN* and V (Figure 1c), or in the scorings by *VDMAX* and V (Figure 1d), where there are no clear correlations. In fact, these results from Figure 1 are consistent with the observations and analysis in Table 3. Moreover, our study also shows that there is no correlation between the scorings by only electrostatics and that by any of the above quantities, which means they are not essentially the same scoring functions. In addition, we have also checked the correlation between V and the interface area related free energy G_{ACE} used during filtering (Figure 1e), and the correlation between V and G_{ACE} (Figure 1f). They are not closely correlated also as shown.

Conclusion

Based on the mechanical balance and tightness of binding in both translational and rotational space of two proteins, the force and moment acting on one molecule by the other, and the second order derivatives of the binding potential including electrostatic and van der Waals interaction are calculated as scoring functions to rank the docked structures. For our test set including 19 protein complexes, each of our mechanical criteria (scoring functions) is able to identify the near-native conformations within the top 100 ranking places for most docking cases. The performance of these scoring functions does not obviously depend on the type of complex. Moreover, there are no clear correlations between the scoring using interaction energy and the scoring using mechanical criteria, except for that between V and M. This indicates that these new scoring functions can give some different information about the docked structures and thus can identify some near-native structures differing in nature from those identified by the energy functions.

An advantage of present scoring functions is that they only depend on a self-consistent force field, and do not need any other set of empirical parameters such as atom contact energy (ACE). Moreover, the hurdle and complexity of the sophisticated calculation of free energy is avoided, and only a more accurate calculation of the binding potentials of protein complexes is needed. It is worth noting that the local overlap or pack on the interface of a docked structure can heavily affect the scoring and ranking results, especially when the van der Waals interaction is involved in the scoring functions, as in the present work. Because the shape complementarity is not included in the three existing scoring functions – an electrostatic interaction, a desolvation term, and the sum of these two terms – they are incomplete, while the shape complementarity can be reflected through inclusion of van der Waals interaction, as in our proposed scoring functions (see Equation 1). In particular, the second derivative terms may provide information on how snugly the ligand is fitted into the binding site, information that is missing from E(elec), etc. This may be one reason for the good performance of present scoring functions. Despite of this, in our procedure, the structural overlap has two negative effects on the ranking results. One is the abnormally large interaction potential; the other is that the unreasonable interaction may result in negative eigenvalue(s) of the second derivative matrices that should be positive near the local

minimum, which therefore affects the calculations of *VDMAX* and *VDMEAN*. We do observe many very large scores (not listed here) ranked at the top position. This is a possible factor for protecting our scoring from getting a better rank for the near-native docked structures. Using energy minimization to remove the local atomic clash in docked structures as suggested by Carlos et al. [14], our test also shows (not listed here) that molecular mechanics minimization can improve some ranking results for the near-native conformations of the docked structures. Another possible reason for the improvement on ranking results is that the energy minimization may also help to select some complexes with poor tightness but with good tightness after minimization. Because the energy surface of the binding interaction is rugged and very complicated, it is very sensitive to the binding position and direction, even for small changes after a short time of minimization, and sometimes, these changes can significantly affect the rank of the docked structures. Generally, through adjusting the local atomic position to approach a local energy minimum, the energy minimization can increase the tightness of binding for a protein–protein complex and thus affect the ranking of docked solutions.

In fact, as revealed in the Results and discussion section, moment M and force F as scoring functions have abilities on nearly the same level to identify the near-native conformations among docked structures. The similar case appears for *VDMAX* and *VDMEAN*. Therefore, only two criteria are necessary, one (M or F) is for the balance criterion and the other (*VDMAX* or *VDMEAN*) is for the binding tightness criterion. A better scoring scheme may be achieved by adopting a combination form of these scoring functions. However, it is still a great challenge to decide the 'correct' docked site by study of the properties of the energy surface, which has much more complexities and details.

Acknowledgements

This work was supported by the Chinese Natural Science Foundation (No. 20173055), Foundation of National Key Program for Basic Research of China (No. 2001CCB01000), and Natural Science Foundation of Anhui Province (No. 01046202).

References

1. Sternberg, M.J.E., Gabb, H.A. and Jackson, R.M., Curr. Opin. Struct. Biol., 8 (1998) 250.
2. Smith, G.R. and Sternberg, M.E.J., Curr. Opin. Struct. Biol., 12 (2002) 28.
3. Halperin, I., Ma, B., Wolfson, H. and Nussinov, R., Proteins, 47 (2002) 409.
4. Kuntz, I., Blaney, J., Oatley, S., Langridge, R. and Ferrin, T.A., J. Mol. Biol., 161 (1982) 269.
5. Connoly, M., Biopolymers, 25 (1986) 1229.
6. Norel, R., Petrtrey, D., Wolfson, H. and Nussinov, R., Proteins, 35 (1999) 403.
7. Gardina, E.J., Willett, P. and Artimiuk, P.J., Proteins, 44 (2001) 44.
8. Palma, P.N., Krippahl, L., Wampler, J.E. and Moura, J.J.G., Proteins, 39 (2000) 178.
9. Jackson, R.M., Gabb, H.A. and Sternberg, M.J.E., J. Mol. Biol., 276 (1998) 265.
10. Jackson, R.M. and Sternberg, M.J.E., J. Mol. Biol., 250 (1995) 258.
11. Weng, Z., Vajda, S. and DeLisi, C., Protein Sci., 5 (1996) 614.
12. Wallquist, A. and Covell, D.G., Proteins, 35 (1996) 403.
13. Zhang, C., Vasmatzis, G., Cornette, J.L. and DeLisi., C., J. Mol. Biol., 267 (1997) 707.
14. Camacho, C.J., Gatchell, D.W., Kimura, S.R. and Vajda, S., Proteins, 40 (2000) 525.
15. Pearlman, D.A. and Charifson, P.S., J. Med. Chem., 44 (2001) 3417.
16. Gatchell, D.W., Dennis, S. and Vajda, S., Proteins, 41 (2000) 518.
17. Vajda, S., Sippl, M. and Novotny, J., Curr. Opin. Struct. Biol., 7 (1997) 222.
18. Srinivasan, J., Cheatham III, T.E., Cieplak, P., Kollman, P.A. and Case, D.A., J. Am. Chem. Soc., 120 (1998) 9401.
19. Wang, W. and Kollman, P.A., J. Mol. Biol., 303 (2000) 567.
20. Straatsma, T.P. and McCammon, J.A., Ann. Rev. Phys. Chem., 43 (1992) 407.
21. Li, L., Chen, R. and Weng, Z.P., Proteins, 53 (2003) 693.
22. Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z. and Socci, N.D., Proc. Natl. Acad. Sci. USA, 92 (1995) 3626.
23. Tsai, C.J., Kumar, S., Ma, B. and Nussinov, R., Protein Sci.,8 (1999) 1181.
24. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., J. Comput. Chem., 4 (1983) 187.
25. Wodak, S.J. and Janin, J., J. Mol. Biol., 124 (1978) 323.
26. Cherfils, J., Duquerroy, S. and Janin, J., Proteins, 11 (1991) 271.
27. Li, C.H., Ma, X.H., Chen, W.Z. and Wang, C.X., Protein Eng., 16 (2003) 265.
28. Camacho, C.J., Weng, Z., Vajda, S. and DeLisi, C., Biophys J., 76 (1999) 1166.
29. Hu, Z., Ma, M., Wolfson, H. and Nussinov, R., Proteins, 39 (2000) 331.