BIOINFORMATICS ORIGINAL PAPER Vol. 26 no. 18 2010, pages 2259–2265
doi:10.1093/bioinformatics/btq404

Structural bioinformatics

Advance Access publication July 11, 2010

iAlign: a method for the structural comparison of protein–protein interfaces

Mu Gao and Jeffrey Skolnick*

Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology, Atlanta, GA, USA

Associate Editor: Burkhard Rost

ABSTRACT

Motivation: Protein–protein interactions play an essential role in many cellular processes. The rapid accumulation of protein–protein complex structures provides an unprecedented opportunity for comparative studies of protein–protein interactions. To facilitate such studies, it is necessary to develop an accurate and efficient computational algorithm for the comparison of protein–protein interaction modes. While there are many structural comparison approaches developed for individual proteins, very few methods are available for protein–protein complexes.

Results: We present a novel ***i***nterface ***align***ment method, iAlign, for the structural alignment of protein–protein interfaces. New scoring schemes for measuring interface similarity are introduced, and an iterative dynamic programming algorithm is implemented. We find that the similarity scores follow extreme value distributions. Using statistical models, we empirically estimate their statistical significance, which is in good agreement with manual classifications by human experts. Large-scale tests of iAlign were conducted on both artificial docking models and experimental structures. In a benchmark test on 1517 dimers, iAlign successfully detects biologically related, structurally similar protein–protein interfaces at a coverage percentage of 90% and an error per query of 0.05. When compared against previously published methods, iAlign is substantially more accurate and efficient.

Availability: The iAlign software package is freely available at http://cssb.biology.gatech.edu/iAlign

Contact: skolnick@gatech.edu

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

Received on May 19, 2010; revised on July 1, 2010; accepted on July 2, 2010

1 INTRODUCTION

Virtually all biological processes are dependent on protein–protein interactions. The role of protein–protein interactions can be structural, e.g. stabilization of homo-oligomers or macromolecular assemblies and/or functional, e.g. inhibition of antigens or enzymes (Goodsell and Olson, 2000; Nooren and Thornton, 2003). Given the essentiality of protein–protein interactions for a cell, one ultimate goal of current research is to identify and elucidate all protein–protein interactions (Russell *et al*., 2004). Toward this goal, one promising avenue and necessary task is the structural determination of all representative protein–protein complexes at high resolution. Over the past two decades, there has been an exponential growth in the number of solved protein complex structures (Tuncbag *et al*., 2008). As structural genomics initiatives select more protein complexes as targets, it is anticipated that the growth rate of protein complex structures will be further accelerated (Bravo and Aloy, 2006; Strong *et al*., 2006).

These rich structural data allow us to characterize protein–protein interactions at the atomic level. By analyzing various properties of a protein–protein interface, such as hydrophobicity, hydrogen bonding, buried surface area, topology, planarity, compositions and so on, one can gain insights into the mechanisms of protein–protein recognition (Janin *et al*., 2008; Jones and Thornton, 1996; Keskin *et al*., 2008). However, these analyses are typically performed for individual protein complexes, and they do not provide detailed structural alignments between protein complexes.

It is well known that the structures of proteins are more likely to be conserved than their sequences (Chothia and Lesk, 1986). Therefore, structural comparison often provides evolutionary insights that are not obvious in sequence comparison. In the era of structural genomics, structural comparison is very relevant because significant structural similarity may guide the study of biological relationships, e.g. whether two proteins are evolutionarily related and/or fulfill similar functions (Redfern *et al*., 2008). Consequently, many computational tools have been developed for structural alignment of individual proteins, such as DALI (Holm and Sander, 1993), CE (Shindyalov and Bourne, 1998) and TM-align (Zhang and Skolnick, 2005). Extensive classifications of protein domains based on their structural, sequence and functional similarity have also become available, such as SCOP (Murzin *et al*., 1995) and CATH (Orengo *et al*., 1997).

Since protein–protein interactions are responsible for the stability and/or function of protein complexes, it makes sense to directly compare the interaction modes of protein complexes and to categorize these complexes according to their interaction modes. One simple strategy is to utilize the alignments of individual protein structures and define interaction modes by the orientation of two complexes and/or the overlap in the interfaces (Aloy *et al*., 2003; Kim *et al*., 2006; Shoemaker *et al*., 2006). While these studies are informative, they have two assumptions: first, the individual corresponding proteins from two complexes are sufficiently similar in structure. Second, the combination of individual protein alignments produces a good interface alignment. However, these assumptions are often not true in biologically interesting cases. One example is shown in Figure 1. Two-headed tomato inhibitor-II (TI-II) simultaneously inhibits two molecules of

*To whom correspondence should be addressed.

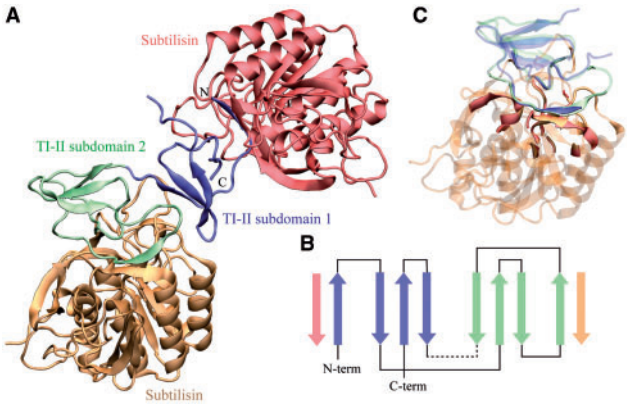

Fig. 1. Complexes of TI-II/subtilisin. (**A**) Cartoon representation. The coordinates were taken from a crystal structure (PDB code 1oyv). Two subdomains (1 and 2) of TI-II are shown in blue and green, and the two subtilisins are shown in red and orange. Molecular visualizations were generated with VMD (Humphrey *et al*., 1996). (**B**) The topology diagram of TI-II and interacting β-strands from subtilisins. The dashed line represents residues (74–85) with missing coordinates. (**C**) Interface alignment of two complexes by iAlign. The alignment has a RMSD of 0.4 Å and a *P*-value of 3×10^{-15}. For clarity, non-interface regions are dimmed.

protease subtilisin (Barrette-Ng *et al*., 2003). The two interaction sites, located in the two subdomains of TI-II, are well separated (Fig. 1A). Obviously, these two interaction sites cannot be aligned in the comparison of the inhibitor against itself. Moreover, due to the different topologies exhibited in the two TI-II subdomains (Fig. 1B), most structural alignment programs only detect weak similarity between them, and cannot properly align the two interaction sites even when one attempts to align the two subdomains. In contrast, it is straightforward for iAlign to identify highly significant similarity between the two protein–protein interfaces (Fig. 1C), suggesting that the same inhibition mechanism is employed.

In a second strategy, one can extend a structural alignment algorithm for individual proteins to the entire protein complex, e.g. MM-align (Mukherjee and Zhang, 2009). An intrinsic limitation of such an approach is that it does not differentiate between interface regions from non-interface regions. As a result, a significant similarity score does not necessarily mean that two complexes have similar interaction modes. An illustration of this issue is the following: Suppose that a protein A forms complexes with two different partners B and C. Assuming no significant structural changes of A, the comparison of two complexes A/B and A/C will lead to a significant score because of the alignment of A to itself, yet A may associate with B and C in totally different ways. This is particularly problematic when the size of protein A is much bigger than the sizes of B and C. Thus, the similarity between these two A proteins dominates the structural comparison, regardless of the actual interaction modes, which could in fact be quite different. In addition, the alignment of full complexes does not necessarily provide the best alignment of interfaces. In the TI-II/subtilisin example mentioned above, the two interaction sites on the inhibitor still cannot be properly aligned by MM-align.

Thus, the development of a dedicated method for comparing protein–protein interfaces is necessary for studying protein–protein interaction modes. An early approach employed a geometric hashing algorithm (Tsai *et al*., 1996). More recently, methods that compare physical chemical interactions, non-covalent interactions or contact maps have been proposed in I2I-SiteEngine (Shulman-Peleg *et al*., 2004), Galinter (Zhu *et al*., 2008) and CMAPi (Pulim *et al*., 2008), respectively. None of these studies, however, provides an assessment of statistical significance of the interface similarity. Moreover, these methods were tested on small datasets, and it is not clear how well they perform in large-scale benchmarks. In particular, it has not been established that interface comparison is useful for the detection of biological relationships.

To address these issues, we present a novel method, iAlign, for the structural comparison of protein–protein interfaces. Below, we first introduce scoring functions for measuring similarity between protein–protein interfaces. We then describe the alignment algorithm and statistical models for the estimation of statistical significance. Large-scale benchmark tests were performed on both docking models and experimental structures. In addition, iAlign is compared with MM-align and I2I-SiteEngine. Finally, we discuss both the advantages and limitations of our approach.

2 METHODS

We adopt a common definition of a protein–protein interface (Janin *et al*., 2008): an interfacial contact is defined if the two residues from two separate proteins have at least one pair of their respective heavy-atoms within 4.5 Å. A contact is, therefore, defined at the residue level. A protein–protein interface is the collection of all residues that have at least one interfacial contact. The length (or size) of an interface is the number of amino acids constituting the interface.

2.1 Similarity measure

In a typical scenario, one compares a query protein–protein interface against a template interface from a library of interfaces. Suppose a query of length L_Q is aligned to a template of length L_T. We consider two scoring functions for measuring interface similarity in iAlign. The first is the Template Modeling score (TM-score; Zhang and Skolnick, 2004),

$$\text{TM-score} = \frac{1}{L_Q} \max \left[\sum_{i=1}^{N_a} 1/(1+d_i^2/d_0^2) \right] \quad (1)$$

where N_a is the number of aligned residue pairs, d_i is the distance in Å between the C_α atoms from the i-th aligned residue pair and the empirical scaling factor $d_0 \equiv 1.24(L_Q - 15)^{1/3} - 1.8$. The constants in d_0 were obtained through fitting the distribution of C_α distances in random alignments (Zhang and Skolnick, 2004). In order to calculate the distance d_i, aligned residues are superimposed with the Kabsch algorithm (Kabsch, 1976). The notation max denotes that the TM-score is the maximum of all possible superpositions. A heuristic iterative extension algorithm is employed to calculate the TM-score (Zhang and Skolnick, 2004), similar to the one used for calculating the GDT-score (Zemla, 2003) and MaxSub (Siew *et al*., 2000). The definition of the TM-score is exactly the same as used in measuring an alignment of individual proteins by TM-align (Zhang and Skolnick, 2005) or of complexes by MM-align (Mukherjee and Zhang, 2009). To avoid confusion, we denote iTM-score for the TM-score of two interfaces compared by iAlign, mTM-score for TM-score of two monomeric proteins compared by TM-align, and dTM-score for TM-score of two dimeric complexes compared by MM-align. Note that these TM-scores have different levels of statistical significance at the same numerical value (see below).

The iTM-score only considers geometric distances. We further introduce the Interface Similarity score (IS-score), which not only measures geometric

distance, but also the conservation of interfacial contact patterns. The IS-score is derived from the iTM-score as follows:

$$\text{IS-score} = \frac{S+s_0}{1+s_0} \tag{2}$$

$$S = \frac{1}{L_Q}\max\left[\sum_{i=1}^{N_a}\frac{f_i}{(1+d_i^2/d_0^2)}\right]. \tag{3}$$

Here, the contact overlap factor $f_i \equiv (c_i/a_i + c_i/b_i)/2$, where a_i and b_i are the numbers of interfacial contacts of the template and of the query interface at the i-th position of the alignment, respectively, and c_i is the number of pairs of overlapped interfacial contacts at the same position (see Supplementary Fig. S1). A pair of interfacial contacts overlaps if the residues forming these contacts are aligned in the two pairs of chains. The scaling factor $s_0 \equiv 0.18 - 0.35/L_Q^{0.3}$ is introduced to make the IS-score length independent (see below). The constants in s_0 were obtained by fitting the distribution of raw scores of unrelated interfaces. Both the iTM/IS-score range from 0 to 1 and identical structures give the maximum score of one.

2.2 Alignment algorithm

The algorithm of iAlign is a further development of the original algorithm implemented in TM-align (Zhang and Skolnick, 2005). Although the algorithm is heuristic, as shown below, the algorithm is sufficiently accurate and highly efficient for practical use. Briefly, the algorithm has two major phases: in the first phase, several guessed solutions are generated through gapless alignments or secondary structure comparison. In the second phase, starting from a guessed alignment, dynamic programming is iteratively applied. The best alignment according to either the iTM-score or IS-score (whichever is specified by the user) is retained.

Four initial alignments are generated during the first phase: (i) the first initial alignment is the gapless alignment that gives the best iTM/IS-score of two interfaces. (ii) the second initial alignment is the best secondary structure match. The match is obtained through dynamic programming with a scoring matrix whose elements are 1 for residues with identical secondary structure and 0 otherwise. A gap penalty of -1 is used. (iii) The third initial alignment is obtained by superimposing fragments of interfaces, similar to ideas suggested in Fr-TM-align (Pandit and Skolnick, 2008) and MM-align (Mukherjee and Zhang, 2009). Let $L_{\min} \equiv \min(L_T, L_Q)$. The interfaces are partitioned into fragments with a length of $\min(L_{\min}/20, 5)$. Superposition is performed for all fragment pairs with at least one pair of residues having identical secondary structure. Corresponding to each fragment superposition, a scoring matrix [Equation (4)] for dynamic programming is calculated and applied to align the two full-length interfaces. The global alignment with the highest iTM-score is the third initial alignment. (iv) After the first three initializations and dynamic programming iterations, the best alignment gives a distance matrix with elements $1/[1+(d_{ij}/d_0)^2]$. The elements of the distance matrix and of the secondary structure matching matrix are summed with weight 0.5, leading to a new scoring matrix for dynamic programming, which generates the fourth initial alignment.

In the second phase, the above four initial alignments are subjected to dynamic programming iterations for which the scoring matrix is defined as

$$S_{ij} = \begin{cases} 1/(1+d_{ij}^2/d_0^2) & \text{for TM-score} \\ (f_{ij}+\delta)/(1+d_{ij}^2/d_0^2) & \text{for IS-score} \end{cases} \tag{4}$$

Here, d_{ij} is the distance between the i-th residue of one structure and the j-th residue of the other structure, and $f_{ij} \equiv (c_{ij}/a_i + c_{ij}/b_j)/2$, where a_i and b_j are the numbers of interfacial contacts of the i-th and j-th residues, respectively, and c_{ij} is the number of pairs of overlappable contacts. A contact between residues i and m is defined overlappable to a contact between residues j and n, if the distances between both i and j, and between m and n are less than an empirical distance $d^* \equiv 1.5[\min(L_T, L_Q)]^{0.3} + 3.5$ and < 8 Å. The small constant δ, set at 0.01, is introduced to prevent a score of zero.

Application of dynamic programming with the scoring matrix defined in Equation (4) yields an alignment. The iTM/IS-score is subsequently calculated for the alignment, and the superposition corresponding to the iTM/IS-score is used to obtain a new scoring matrix, which in turn generates a new alignment. The procedure is repeated until the alignment converges, typically within 10 rounds, or reaches an upper limit of 30 iterations. A gap opening penalty score of -0.6 without any gap extension penalty is used.

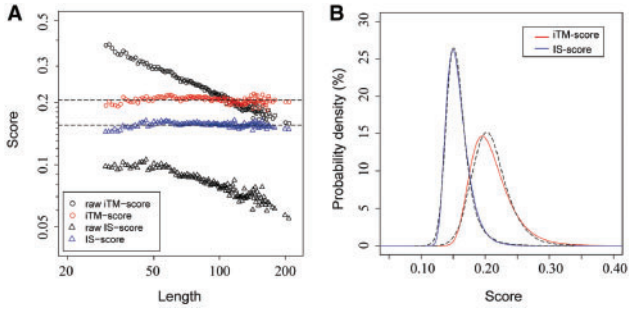

Fig. 2. Distributions of iTM/IS-scores of randomly selected, similar length interface pairs. (**A**) Mean of scores versus the length of interfaces. Horizontal dashed lines are located at 0.156 and 0.206, the means of iTM/IS-scores across all lengths. (**B**) Distributions of iTM/IS-scores. Dashed lines are the observed probability densities, and solid lines are modeled values according to the Gumbel distributions.

Interface alignment is complicated by the fact that the order of protein chains may be important. Given two interfaces with four chains A/B and A′/B′, there are four ways to pair the chains: (i) A/B versus A′/B′, (ii) B/A versus B′/A′, (iii) A/B versus B′/A′ and (iv) B/A versus A′/B′. When a highly significant alignment exists, the first two pairings (i–ii) lead to the same alignment as does pairs (iii–iv). However, for interfaces with low/no similarity, each of the respective pairings might yield different alignments. This is because the algorithm for the iTM/IS-score calculation is dependent on chain order. To guarantee an identical alignment regardless of chain order, iAlign considers all four ways of chain pairings. We then pick the pairing whose score is the best. In addition, cross-chain alignment [e.g. A to B′ or B to A′ in (i)] is prohibited by assigning large penalties in the scoring matrix [Equation (4)].

Above, we have described the procedure for sequential structural alignment, i.e. the alignment of interfacial residues follows their sequential order. An algorithm for non-sequential alignment has also been developed for iAlign (details will be published elsewhere). Below, we only present results from sequential alignments.

2.3 Statistical significance

The statistical significance of iTM/IS-scores is estimated through comparing about 1.8 million random interface pairs (Section 2.4). Figure 2A shows the means of both iTM and IS-scores of random interfaces of similar lengths. For a given length, we consider all random pairs whose lengths are between 95% and 105% of the length. The raw iTM-score is calculated using a fixed value of d_0 at 4 Å, and the raw IS-score is calculated using Equation (3). Without applying proper scaling factors, the raw iTM/IS-scores decrease exponentially as the length of interfaces increase. In contrast, proper scaling yields approximately length-independent iTM/IS-scores. The means of iTM/IS-scores for random interfaces of similar lengths are 0.206/0.156, respectively.

As shown in Figure 2B, the scores from the random background (RB) follow Gumbel distributions [Equation (5)], also known as type I extreme value distributions. These distributions are over maximum values and, therefore, are suitable to our cases since the iTM/IS-score are the maxima of many structural alignments. The statistical models allow us to calculate the P-values of scores, as proposed previously (Levitt and Gerstein, 1998). A list of P-values and corresponding iTM/IS-scores is given in Table 1. For example, an iTM/IS-score of 0.31/0.21 indicates a similarity at a significant

Table 1. Statistical significance of the scores for interfaces alignments

P-value	5×10^{-2}	1×10^{-2}	1×10^{-3}	1×10^{-4}	1×10^{-5}	1×10^{-6}	1×10^{-10}
iTM-score	0.270	0.311	0.368	0.426	0.484	0.542	0.773
IS-score	0.191	0.214	0.247	0.279	0.311	0.343	0.473

P-value of 0.01. One may use these scores to quickly estimate statistical significance.

A more accurate estimation of statistical significance is achieved through modeling the distributions of scores at specific lengths. Since the scores are asymmetric for interfaces of different lengths, due to score normalization, we calculate the *P*-value that corresponds to the higher score, namely, the score normalized by the smaller interface. This gives a single *P*-value for a pair of interfaces. Supplementary Figure S2 shows the observed and modeled distributions of scores at various lengths. Each distribution is modeled by the Gumbel distribution,

$$P(z)=\exp[z-\exp(z)] \tag{5}$$

where $z=(s-\mu)/\sigma$. The variable s denotes the iTM/IS-score; μ and σ are the location and the scale parameters, respectively. These parameters are estimated through linear regression fits

$$\begin{aligned}\mu&=a+b\ln(L_Q)+c\ln(L_T)\\ \sigma&=c+d\ln(L_Q)+f\ln(L_T).\end{aligned} \tag{6}$$

The parameters a to f, given in Supplementary Table S1, were obtained by linear fitting to the location and scale parameters, which were obtained through maximum likelihood estimates with the EVD package in R (http://www.r-project.org/). Finally, the *P*-value is calculated using the formula

$$P\text{-value}=1-\exp[-\exp(-z)]. \tag{7}$$

2.4 Datasets

The three datasets used in the study were derived from the M-TASSER template library (Chen and Skolnick, 2008). The library consists of 1838 dimeric protein–protein complexes, non-redundant at 35% sequence identity. Since coiled–coil complexes are trivially similar, we removed the 48 such complexes from the library, resulting in 1790 complexes.

2.4.1 Random background RB is a set of ~1.77 million pairs of interfaces curated from all-against-all comparison of 1790 complexes. For each pair of complexes, the structural similarity of individual proteins that form the complexes was assessed with TM-align, which reports mTM-score between individual proteins. There are four combinations of individual proteins from two dimers. We discarded the pair of complexes if the maximum mTM-score among these four combinations is higher than 0.35. Note that for mTM-score a value of 0.35 suggests that two protein structures are dissimilar (Zhang *et al.*, 2006). The protein–protein interfaces of the remaining complex pairs are the RB chosen for estimating the statistical significance described above.

2.4.2 Dimer1517 From the template library, we collected 1517 dimers that have SCOP assignments (version 1.75). The set consists of 327 heterodimers and 1190 homodimers. We further examine all-against-all pairs among Dimer1517. According to SCOP, protein domains within the same superfamily are biologically related (Murzin *et al.*, 1995). Two complexes are considered to have related protein–protein interfaces, if (i) at least one pair of their interacting domains has the same SCOP superfamily assignments, and (ii) the two protein–protein interfaces share strong structural similarity. To avoid self-testing iAlign, we designed an alternative procedure to assess interface similarity for complexes that have one pair of domains from the same SCOP superfamily. In this procedure, alignments of each of the monomers in one dimer to the monomers in the other dimer were obtained with TM-align. From these alignments, one can count the number of overlapped contacts and calculate the contact overlap ratio, which is the count divided by the smaller number of total interfacial contacts between the two complexes. A minimum contact overlap ratio of 0.3 is required for assigning related interfaces. In these related pairs, the best mTM-score is usually higher than 0.5 between individual proteins from these complexes. On the other hand, two protein–protein complexes have a pair of unrelated protein–protein interfaces, if (i) none of their interacting domains has the same SCOP superfamily assignments, or (ii) the contact overlap ratio is zero and the fraction of aligned interface residues over the shorter length of the two interfaces is <0.15. This cutoff was conservatively chosen to tolerate a scenario where a small fraction of interfacial residues are aligned by chance. The second condition incorporates complexes with similar global structures but dissimilar interaction modes. In total, 1128 pairs of biologically related and ~1.15 million pairs of unrelated interfaces were classified. The remaining 3167 interface pairs are ambiguous cases that cannot be confidently classified by analyzing the global structural alignments of individual proteins, and they were therefore discarded for this analysis, although many have significant iTM/IS-scores according to iAlign.

2.4.3 Dimer597 A subset (597) of Dimer1230 is curated by limiting the lengths of individual proteins within 200 amino acids. The subset contains 373 related pairs and 176 875 unrelated pairs.

3 RESULTS

We first tested iAlign on docking models and obtained encouraging results (Supplementary Material). Below, we describe the result of detecting evolutionarily related, structurally similar protein–protein interfaces, and comparison with MM-align and I2I-SiteEngine.

3.1 Predicting biologically related interfaces

We benchmark the performance of iAlign in detecting biologically related protein–protein interfaces from experimental structures. The SCOP superfamily classification and structural similarity are used to decide whether two complexes share biologically related interfaces (Section 2.4.2).

Using either the iTM-score or the IS-score as the similarity measure, iAlign estimates a significant $P<0.01$ for all related protein–protein interfaces, except for one case (Fig. 3A). On the other hand, about 0.8/1.8% of unrelated pairs have a significant iTM/IS-score with an estimated $P<0.01$. The cumulative fraction of unrelated interface pairs according to their SCOP classification is the observed, or 'ideal', *P*-value. Ideally, the observed *P*-value should match the *P*-value estimated by iAlign from Equation (5). In the regime above 1×10^{-4}, the *P*-values according to our statistical model are in good agreement with the observed *P*-values. In particular, corresponding to estimated *P*-values of 0.01, 0.001 and 0.0001 for the iTM-score, the observed *P*-values are 0.009, 0.0009 and 0.0001, respectively. This excellent agreement is a bit surprising, because we did not use the SCOP classification as the RB for deriving our statistical models. In the regime below 1×10^{-5}, however, there is some separation between the estimated and the observed *P*-values. This is largely attributed to the existence of structurally similar yet evolutionarily unrelated pairs, such as four-helix bundles. Overall, the result shows that iAlign identifies significant similarity between related interfaces, and that the estimates of statistical significance are reasonable. The results are consistent between homodimers and

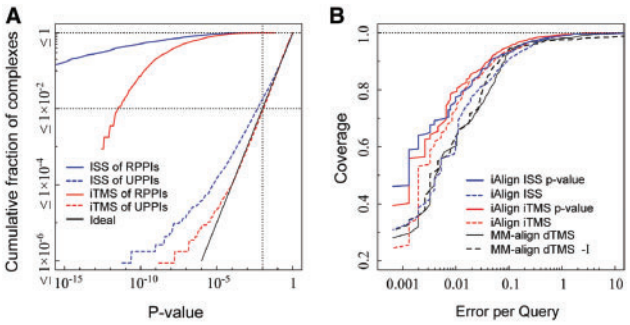

Fig. 3. Detection of biologically related protein-protein interfaces (RPPIs) from unrelated protein-protein interfaces (UPPIs) among 1517 dimers. (**A**) Cumulative fraction of complexes versus the estimated *P*-values for iTM/IS-scores (iTMS/ISS) by iAlign. The cumulative factions of unrelated pairs are the observed *P*-values. The black line represents the ideal match between the estimated *P*-values and the observed *P*-values. (**B**) Coverage versus EPQ curves using different criteria for predicting RPPIs (see text).

heterodimers (Supplementary Fig. S5), though the estimated IS-score *P*-values for heterodimers are closer to the ideal *P*-values than the estimates for homodimers.

Figure 3B shows receiver operating characteristic (ROC) type curves, which analyze the performance of various criteria (e.g. IS-score) for predicting related protein–protein interfaces. Given a threshold for a criterion, the positives/negatives are those interface pairs satisfying/dissatisfying the criterion. True positives and true negatives are related and unrelated interface pairs (Section 2.4.2) that are correctly predicted as positives and negatives, respectively. Based on these binary predictions, coverage is the fraction of true positives correctly predicted as positives, and error per query (EPQ) is the number of false positives divided by the total number of query interfaces (Brenner *et al*., 1998; Levitt and Gerstein, 1998). Compared with the false positive rate adopted in a typical ROC curve, EPQ provides a better measure in the high accuracy regime, which is relevant to practical applications because the amount of unrelated pairs is enormous. In our benchmark tests, we assessed five criteria: the iTM/IS-score and their *P*-value, and dTM-score, the TM-score from comparing global structures of complexes by MM-align (Mukherjee and Zhang, 2009). All three types of scores are all asymmetric, in the sense that two different values are possible for a pair of structures of different lengths. In benchmarks, the higher score was chosen. In the case of MM-align, two sets of runs were conducted with the default setting and with the –I option, respectively. The latter adds a heavy weight to interface regions in an alignment.

It is evident that *P*-values of the iTM/IS-score always achieve better performance than the scores themselves (Fig. 3B). At 0.005 EPQ, for example, the *P*-value of the iTM/IS-score gives 0.68/0.70 coverage, compared with 0.65/0.56 coverage yielded by the iTM/IS-score. At an extremely low EPQ of <0.005, the *P*-value of IS-score produces the best performance among all prediction schemes. In the EPQ regime from 0.005 to 1, the *P*-value of the iTM-score performs the best, but the difference is very small between the two *P*-values. At the same EPQ of 0.05, for example, *P*-values of the iTM-score and IS-score give coverage values of 0.92 and 0.91, respectively. Moreover, regardless of the similarity measure employed, interface alignments by iAlign always provide better discrimination than global alignments of complex structures by MM-align, with the difference being most notable in the highly confident regime relevant to practical applications. At 0.01 EPQ, for example, dTM-score produces 0.60 coverage, which is considerably lower than 0.79/0.76 coverage by *P*-value of iTM/IS-score calculated with iAlign. Heavily weighting the interface region gives MM-align slightly better performance, but the improvement is small, for instance, ∼4% increase in coverage at 0.01 EPQ.

The difference between interface alignment and global alignment of complexes is not surprising because interface similarity is not a priori equivalent to global similarity, though the two are correlated. As shown in Supplementary Figure S6A, more than half of complex pairs with significant interface similarity at a $P<0.01$ have an insignificant dTM-score <0.4. Nevertheless, highly significant interface similarity often leads to highly significant global similarity. The number of complexes with dissimilar global structures drops dramatically when a significant *P*-value threshold $<1\times10^{-5}$ is employed, though there are exceptions with one example provided in Supplementary Figure S7A. On the other hand, as we mentioned in Section 1, a high dTM-score does not guarantee interface similarity. For example, among dimer pairs with dTM-score >0.6, ∼15% of them have dissimilar interfaces with an insignificant iAlign $P<0.01$ (Supplementary Fig. S6B). These are complexes sharing very similar global folds but dissimilar interaction modes. Such false positives are significantly reduced through interface similarity evaluation, which is the main reason for the better performance of iAlign than MM-align. One example is shown in Supplementary Figure S7B.

3.2 Comparison with I2I-SiteEngine

The performance of iAlign is compared with a previously published interface alignment method in I2I-SiteEngine (Shulman-Peleg *et al*., 2004). The I2I-SiteEngine demands relatively large computing resources. To reduce computing costs, we selected a subset (Dimer597) of Dimer1517 by limiting the length of the individual protein chains to 200 residues. I2I-SiteEngine reports two scores for each of top 10 alignments, the Match-score (M-score) and the Total-score (T-score), the former normalized by the Total-score of query compared with itself. For a pair of interfaces, we conducted two runs with each interface as the query, and then selected the best M-score or T-score among all top alignments.

As shown in Figure 4A, the coverage-precision (also known as recall-precision) curve evaluates the accuracy of two methods. The coverage is the same as above, and the precision is defined as the fraction of true positives (both pairs of complexes have evolutionarily related, structurally similar interfaces) among all positives predicted. Clearly, iAlign has substantially higher accuracy than the I2I-SiteEngine. At 80% precision, iAlign can identify 85% of true positives according to the *P*-value of the IS-score, whereas I2I-SiteEngine can only identify 21/47% of true positives with the M/T-score, respectively. At a higher precision of 90%, iAlign has a coverage value of 75%, about five/two times the coverage of 15/39% by the M/T-score.

Regarding the requisite computer time, iAlign is about two orders of magnitude faster than I2I-SiteEngine (Fig. 4B). We collected the total computing time statistics for iAlign, and statistics of total and essential computing time for I2I-SiteEngine. Note that the total computing time does not include the time for generating input files, which one needs to construct only once. The essential computing

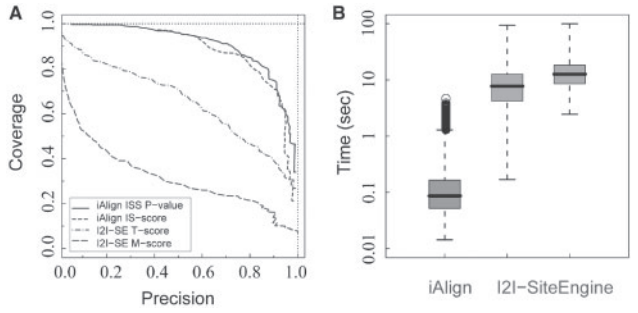

Fig. 4. Comparison of iAlign with I2I-SiteEngine. (**A**) Coverage versus precision curves from tests on the set Dimer597. (**B**) Box plots of runtime costs of iAlign (blue) and I2I-SiteEngine (green). The left and right box plots represent the overall runtime of iAlign and I2I-SiteEngine, and the middle plot represents the runtime of essential computing as reported by I2I-SiteEngine. Medians are represented by thick horizontal black lines.

time of I2I-SiteEngine includes only costs for hashing, matching and scoring. All runs were conducted on the same cluster composed of 2.4 GHz AMD Opteron 8431 processors. The median total time of iAlign is 0.087 s, about 146/89 times shorter than 12.66/7.75 s, the median total/essential computing time reported by I2I-SiteEngine.

4 DISCUSSION

Comparative studies of protein structures provide valuable insights into the biological relationship between proteins, especially those with distant evolutionary relationship undetectable from sequence information alone. During the past two decades, many computational methods have been developed for protein structural comparison; some well-known examples include DALI (Holm and Sander, 1993), CE (Shindyalov and Bourne, 1998) and TM-align (Zhang and Skolnick, 2005). However, these methods are designed for comparing the global tertiary structures of individual proteins. Considering that protein–protein interactions play an essential role in a cell, structural comparison of protein–protein interfaces may provide additional insights that are not obvious in standard structural comparisons of individual proteins. The main purpose of this work is to introduce iAlign, an accurate and efficient computational method for protein interface comparison, and to demonstrate that such a comparison is indeed helpful for deciphering biological relationships between proteins.

Like structure comparisons of individual proteins, it is important to define a measure that effectively characterizes the level of structural similarity. One advantage of the TM-score, first introduced for measuring the similarity between monomeric protein structures, is that the TM-score is approximately length independent for globular structures (Zhang and Skolnick, 2004). This property allows one to estimate statistical significance according to the values of the score (an mTM-score >0.4 is considered significant). However, it is not clear whether the TM-score can be applied to protein–protein interfaces, which are often flat and discontinuous. We show that the interfacial TM-score (iTM-score) can be applied to measure the similarity between protein–protein interfaces, and that the property of length independence is still maintained. In addition, we designed a new measure, the IS-score, which compares not only geometric distances, but also contact patterns not considered by the iTM-score. As a result, the IS-score is more specific and significantly better than the iTM-score in differentiating closely related protein–protein complexes, e.g. near-native docking models (Supplementary Results).

Moreover, we employed statistical models to estimate the significance of the iTM/IS-scores. In large-scale benchmark tests on 1517 dimers, our estimates agree well with the SCOP classification, despite the fact that SCOP was not used for deriving the statistical models. In the predictions of biological relationships, the *P*-values of iTM/IS-score yield significantly more accurate results than the scores. In general, accurate predictions were obtained by iAlign with the *P*-values of both scores. For example, about 90% coverage is achieved at 0.05 EPQ.

There is unlikely to be a consensus on the definition of the optimal similarity metric for protein–protein interface comparison. While geometry-based similarity measures are used in iAlign, measures based on physical chemical properties were proposed in I2I-SiteEngine (Shulman-Peleg *et al.*, 2004). Benchmark results show that iAlign is significantly more accurate and two orders of magnitude more efficient than I2I-SiteEngine. However, the difference in accuracy does not mean that physical chemical properties are not important for interface comparison. The better accuracy of iAlign is mainly due to the length independence of the iTM/IS-score and a sound assessment of statistical significance.

It is known that the space of tertiary structures of individual protein domains is very limited, largely because only certain ways of secondary structure packing are physically viable (Finkelstein and Ptitsyn, 1987). As a consequence, two evolutionarily unrelated proteins may converge to a similar structural fold (Kihara and Skolnick, 2003; Kolodny *et al.*, 2006; Zhang *et al.*, 2006). This interesting phenomenon unfortunately creates a challenge for predicting biological relationships from protein structure. In this study, we demonstrate that comparison of protein–protein interfaces may be utilized for differentiating biological relationships. This works because protein–protein interactions are important for their stability and/or function, leading to the conservation of a specific interaction mode during the course of evolution.

However, we emphasize that interface comparison is not a replacement for, but rather is complementary to, structural comparison of individual proteins. This point can be illustrated with one example. Suppose protein A and its homolog A′ interacts with two different proteins B and C, respectively, and the protein–protein interfaces of complexes A/B and A′/C are dissimilar. An interface alignment of A/B and A′/C gives an insignificant score, whereas an alignment between A and A′ yields a significant score. While the score of interface similarity is still informative, indicating that A/B and A′/C exhibit different interaction modes, the homologous relationship between A and A′ is detected only through the alignment of individual protein structures. Therefore, one expects that a combination of both types of structural comparisons should provide a more comprehensive description of protein–protein relationships than using either comparison metric alone. Toward this direction, considerable efforts have been invested recently in classifying protein–protein interfaces (Kim *et al.*, 2006; Mintz *et al.*, 2005; Shoemaker *et al.*, 2006; Tuncbag *et al.*, 2008).

Since two biologically unrelated protein complexes may display similar interfaces (Tsai *et al.*, 1996), the question arises in how similar two protein–protein interfaces have to be for a reliable prediction on their biological relationship. With the *P*-value estimation provided by iAlign, one can quickly set a suitable *P*-value

threshold. When scanning a large library of 10 000 interfaces, which is the same order of magnitude as the size of the current PDB, one expects to see one hit from unrelated interfaces at a *P*-value of 1×10^{-4}. In order to reduce the number of false positives to <0.1, one needs to set a highly significant *P*-value threshold of $< 1 \times 10^{-5}$ for predicting biologically related interfaces from a library the size of the PDB.

In principle, protein–protein interfaces can be geometrically similar without following a particular sequence order. While iAlign provides an option for permitting non-sequential alignment, we found that the sequential alignments yield much fewer hits across SCOP superfamilies than non-sequential alignments (Gao and Skolnick, unpublished data). Non-sequential alignments detect both biologically related and unrelated pairs of complexes that have structurally similar interfaces. Restriction to sequential alignments significantly increases the likelihood that an evolutionary relationship (e.g. as assessed by SCOP) is detected.

Despite many efforts toward understanding the nature of protein–protein interactions, numerous issues remain unresolved. Some important questions include: What is the repertoire of protein–protein interactions modes that Nature employs? How complete is the structural space of protein–protein interaction modes in the current PDB? How can we use structural information to accurately predict protein-protein interactions? How can we design a protein–protein interface for a desired function? We expect that iAlign will be a useful tool for addressing these outstanding questions and for developing applications involving structural comparison of protein–protein interfaces.

ACKNOWLEDGEMENTS

We thank Dr Shashi B. Pandit for stimulating discussions and Dr Bartosz Ilkowski for computing support.

Funding: National Institutes of Health (grant GM-48835).

Conflict of Interest: none declared.

REFERENCES

Aloy,P. *et al*. (2003) The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.*, **332**, 989–998.

Barrette-Ng,I.H. *et al*. (2003) Structural basis of inhibition revealed by a 1 : 2 complex of the two-headed tomato inhibitor-II and subtilisin carlsberg. *J. Biol. Chem.*, **278**, 24062–24071.

Bravo,J. and Aloy,P. (2006) Target selection for complex structural genomics, *Curr. Opin. Struct. Biol.*, **16**, 385–392.

Brenner,S.E. *et al*. (1998) Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships. *Proc. Natl Acad. Sci. USA*, **95**, 6073–6078.

Chen,H.L. and Skolnick,J. (2008) M-TASSER: an algorithm for protein quaternary structure prediction. *Biophys. J.*, **94**, 918–928.

Chothia,C. and Lesk,A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.*, **5**, 823–826.

Finkelstein,A.V. and Ptitsyn,O.B. (1987) Why do globular-proteins fit the limited set of folding patterns. *Prog. Biophys. Mol. Biol.*, **50**, 171–190.

Goodsell,D.S. and Olson,A.J. (2000) Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 105–153.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Humphrey,W. *et al*. (1996) VMD: visual molecular dynamics. *J. Mol. Graphics*, **14**, 33–38.

Janin,J. *et al*. (2008) Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.*, **41**, 133–180.

Jones,S. and Thornton,J.M. (1996) Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.

Kabsch,W. (1976) Solution for best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A*, **32**, 922–923.

Keskin,Z. *et al*. (2008) Principles of protein-protein interactions: What are the preferred ways for proteins to interact? *Chem. Rev.*, **108**, 1225–1244.

Kihara,D. and Skolnick,J. (2003) The PDB is a covering set of small protein structures. *J. Mol. Biol.*, **334**, 793–802.

Kim,W.K. *et al*. (2006) The many faces of protein-protein interactions: a compendium of interface geometry. *PLoS Comp. Biol.*, **2**, 1151–1164.

Kolodny,R. *et al*. (2006) Protein structure comparison: implications for the nature of 'fold space', and structure and function prediction. *Curr. Opin. Struct. Biol.*, **16**, 393–398.

Levitt,M. and Gerstein,M. (1998) A unified statistical framework for sequence comparison and structure comparison. *Proc. Natl Acad. Sci. USA*, **95**, 5913–5920.

Mintz,S. *et al*. (2005) Generation and analysis of a protein-protein interface data set with similar chemical and spatial patterns of interactions. *Proteins Struct. Funct. Bioinform.*, **61**, 6–20.

Mukherjee,S. and Zhang,Y. (2009) MM-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. *Nucleic Acids Res.*, **37**, e83.

Murzin,A.G. *et al*. (1995) SCOP - a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.

Nooren,I.M.A. and Thornton,J.M. (2003) Diversity of protein-protein interactions. *EMBO J.*, **22**, 3486–3492.

Orengo,C.A. *et al*. (1997) CATH - a hierarchic classification of protein domain structures. *Structure*, **5**, 1093–1108.

Pandit,S.B. and Skolnick,J. (2008) Fr-TM-align: a new protein structural alignment method based on fragment alignments and the TM-score. *BMC Bioinformatics*, **9**, Article 531.

Pulim,V. *et al*. (2008) Optimal contact map alignment of protein-protein interfaces. *Bioinformatics*, **24**, 2324–2328.

Redfern,O.C. *et al*. (2008) Exploring the structure and function paradigm. *Curr. Opin. Struct. Biol.*, **18**, 394–402.

Russell,R.B. *et al*. (2004) A structural perspective on protein-protein interactions. *Curr. Opin. Struct. Biol.*, **14**, 313–324.

Shindyalov,I.N. and Bourne,P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, **11**, 739–747.

Shoemaker,B.A. *et al*. (2006) Finding biologically relevant protein domain interactions: conserved binding mode analysis. *Protein Sci.*, **15**, 352–361.

Shulman-Peleg,A. *et al*. (2004) Protein-protein interfaces: recognition of similar spatial and chemical organizations. In Jonassen,I.K.J. (ed.), *Algorithms in Bioinformatics*, Springer, Berlin, pp. 194–205.

Siew,N. *et al*. (2000) MaxSub: an automated measure for the assessment of protein structure prediction quality. *Bioinformatics*, **16**, 776–785.

Strong,M. *et al*. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl Acad. Sci. USA*, **103**, 8060–8065.

Tsai,C.J. *et al*. (1996) A dataset of protein-protein interfaces generated with a sequence-order-independent comparison technique. *J. Mol. Biol.*, **260**, 604–620.

Tuncbag,N. *et al*. (2008) Architectures and functional coverage of protein-protein interfaces. *J. Mol. Biol.*, **381**, 785–802.

Zemla,A. (2003) LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Res.*, **31**, 3370–3374.

Zhang,Y. *et al*. (2006) On the origin and highly likely completeness of single-domain protein structures. *Proc. Natl Acad. Sci. USA*, **103**, 2605–2610.

Zhang,Y. and Skolnick,J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins Struct. Funct. Bioinform.*, **57**, 702–710.

Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.*, **33**, 2302–2309.

Zhu,H. *et al*. (2008) Alignment of non-covalent interactions at protein-protein interfaces. *PLoS One*, **3**, e1926.