Nucleic Acids Research, 2007, Vol. 35, Web Server issue ***W561–W567***
doi:10.1093/nar/gkm346

3D-partner: a web server to infer interacting partners and binding models

Yung-Chiang Chen[1], Yu-Shu Lo[1], Wen-Chang Hsu[1] and Jinn-Moon Yang[1,2,3,*]

[1]Institute of Bioinformatics, National Chiao Tung University, Hsinchu, 30050, Taiwan, [2]Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, 30050, Taiwan and [3]Core Facility for Structural Bioinformatics, National Chiao Tung University, Hsinchu, Taiwan

Received February 1, 2007; Revised April 3, 2007; Accepted April 22, 2007

ABSTRACT

The 3D-partner is a web tool to predict interacting partners and binding models of a query protein sequence through structure complexes and a new scoring function. 3D-partner first utilizes IMPALA to identify homologous structures (templates) of a query from a heterodimer profile library. The interacting-partner sequence profiles of these templates are then used to search interacting candidates of the query from protein sequence databases (e.g. SwissProt) by PSI-BLAST. We developed a new scoring function, which includes the contact-residue interacting score (e.g. the steric, hydrogen bonds, and electrostatic interactions) and the template consensus score (e.g. couple-conserved residue and the template similarity scores), to evaluate how well the interfaces between the query and interacting candidates. Based on this scoring function, 3D-partner provides the statistic significance, the binding models (e.g. hydrogen bonds and conserved amino acids) and functional annotations of interacting partners. The correlation between experimental energies and predicted binding affinities of our scoring function is 0.91 on 275 mutated residues from the ASEdb. The average precision of the server is 0.72 on 563 queries and the execution time of this server for a query is ~15 s on average. These results suggest that the 3D-partner server can be useful in protein-protein interaction predictions and binding model visualizations. The server is available online at: http://3D-partner.life.nctu.edu.tw.

INTRODUCTION

Protein–protein interactions are involved in most biological processes. Identifying their associated networks comprehensively is the key to understanding cellular mechanisms (1). Some systematic identification of protein–protein interactions have been constructed by high-throughput experimental methods, such as large-scale two-hybrid system (2) and affinity purifications (3). A basic problem with most large-scale experimental methods is the high false-positive rate (4). Many computational methods have been developed to predict protein–protein interactions by using gene expression profiles (5), domain–domain interactions (6–8), phylogenetic profiles (9), known 3D complexes (10,11) and interologs (12,13). These large-scale methods are often unable to respond how a protein interacts with another one.

To identify interacting domains from three-dimensional (3D) structural complexes is able to study domain–domain interactions. A known 3D structure of interacting proteins provides interacting domains and atomic details for thousands of direct physical interactions. In addition, it is usually possible to build an interaction model of two proteins by comparative modeling if a known complex structure comprising homologs of these two sequences is available (10,11,14,15). For a pair sequences, these methods often search a 3D-complex library to find homologous templates and score how well the query protein pair fit the known template structures by using a scoring matrix. In this way, they should evaluate all possible protein pairs (18 000 000) in one species if it has 6000 proteins. Our previous study proposed '3D-domain interologs' which is similar to 'interologs' (12). The 3D-domain interologs is defined as 'Domain *a* (in chain A) interacts with domain *b* (in chain B) in a known 3D complex, their inferring protein pair A′ (containing domain *a*) and B′ (containing domain *b*) in the same species would be likely to interact with each other if both protein pairs are homologous.' Based on this concept, we are able to search protein databases to predict protein–protein interactions for many species by using a 3D-dimer complex (16).

*To whom correspondence should be addressed. Tel: +886-3-571212-56942; Fax: +886-3-579288; Email: moon@faculty.nctu.edu.tw, moon@cc.nctu.edu.tw

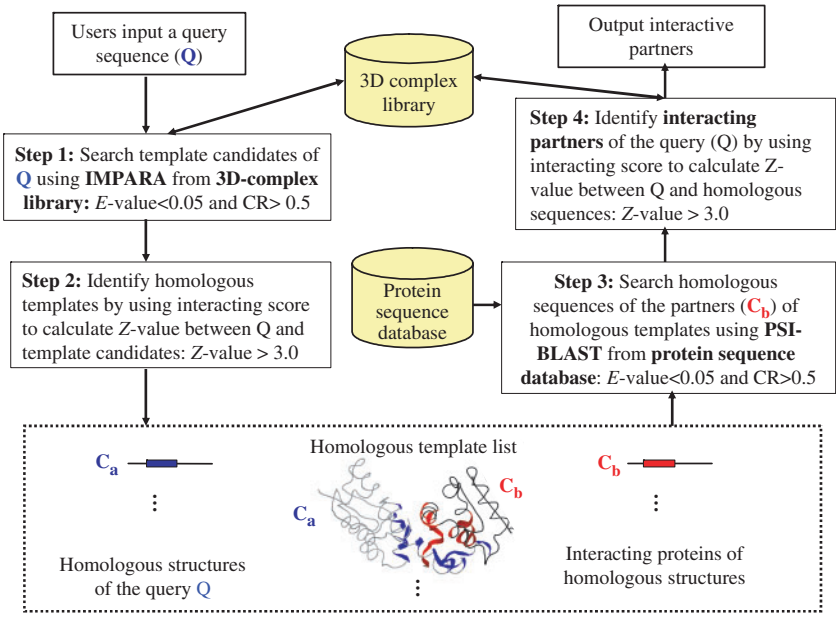

Figure 1. Overview of the 3D-partner server for inferring interaction partners and binding models.

Here, we report the development of an automatic server, 3D-partner, for interacting partners and binding models prediction by using 3D-domain interologs through structure complexes and a knowledge-based scoring function which is the key novelty in this article. The 3D-partner utilizes IMPALA and PSI-BLAST to identify homologous structures (templates) and interacting partners of a query protein sequence from a 3D-dimer template library and protein sequence databases [i.e. SwissProt (17)], respectively. These homologous structures and interacting partners were evaluated by a scoring function which considered steric and special-bond matrices (i.e. hydrogen bonds, electrostatic interactions and disulfide bonds) but also the template consensus scores (couple-conserved residue score and template similarity). After interacting partners were identified, the 3D-partner provides 3D interacting domains and contact residues for visualizing molecular details of any protein pairs between the query and interacting partners. The 3D-partner server was tested on 275 mutated residues selected from the Alanine Scanning Energetics database (ASEdb) (18) to predict the binding affinities. The correlation between experimental energies and predicted energies is 0.91. In addition, the average precision of this server for interacting partner prediction is 0.72 by using a non-redundant set.

METHOD AND IMPLEMENTATION

Figure 1 presents the details of the 3D-partner server for inferring interacting partners and binding models of a query sequence through structure complexes and a new scoring function by the following steps. First, the server uses IMPALA to search template candidates of a query (Q) from 3D-complex profile library (1894 heterodimers). IMPALA, widely used for local sequence alignments, searches the query sequence against each of the template profiles, which constitute a database of PSI-BLAST-generated position-specific score matrices (PSSMs). A template is considered as a candidate if the E-value is <0.05 and the aligned contact residue ratio (CR) between the Q and candidate is >0.5. The aligned procedure of IMPALA is a sequence (Q) to profile (template) alignment. Second, our scoring function is applied to calculate the interacting score and Z-value for each candidate, which is selected as a homologous template (C_a in Figure 1) of Q if its Z-value > 3.0, according to the aligned contact pairs on the template.

After homologous templates are identified, the 3D-partner identified interacting partner candidates of the query. For each homologous template (C_a), this server applies PSI-BLAST to scan the interacting-partner sequence profile (C_b in Figure 1) of C_a against each of protein sequences in the SwissProt version 51.3 (containing 250 296 protein sequences). The sequence profile, built by using the same procedure for template sequence profiles, is the initial PSSM of PSI-BLAST and the number of iteration is set to one. Therefore, this search procedure can be considered as a profile-to-sequence alignment. The sequences whose E-value < 0.05 and $CR > 0.5$ are selected as homologous sequences of C_b. Finally, for each homologous sequence, our scoring function is applied to calculate the interacting score and to evaluate the Z-value between the query Q and the homologous sequence according to the aligned contact pairs on the hit template. A homologous sequence is considered as an interacting partner of the query if the Z-value >3.0. The server reports interacting partners of the query ordered by Z-values which represent the statistical significances of hit interacting partners.

3D-dimer library and interacting domains

The 3D-partner uses IMPALA to identify the 3D-dimer templates of a query sequence. Here, the 3D-dimer template library, which consists of 1894 heterodimers (i.e. 3788 sequences), was extracted from the Protein Data Bank (PDB) (19) released in 24 February 2006. Any sequence in the library has <98% sequence identity to each other to eliminate the duplicated complexes. We excluded the dimer whose chains are shorter than 30 residues (15,20). For each complex in the 3D-dimer library, we identified interacting domains and contact residues of two chains. Contact residues, whose any heavy atoms should be within a threshold (distance ≤ 4.5 Å) to any heavy atoms of another chain, were considered as the core parts of the 3D-interacting domains in a complex. Each domain must have > contact residues and the number of interacting contact-residue pairs >25 to make sure that the contact between the domains was reasonably extensive (21). After interacting domain were determined, we identified its SCOP domain (22). Finally, each template profile in the IMPALA profile library was constructed using PSI-BLAST by searching SCOP domain sequence against the UniRef90 database (23) in which the sequence identity is <90% of each other.

Scoring function and matrices

We have developed a scoring function to measure the reliability of a protein–protein interaction. This scoring function includes the contact-residue interacting score, which consists of the steric (i.e. shape complementary) and specific energies (e.g. hydrogen-bond energy), and the template consensus scores which contain couple-conserved residue and the template similarity scores. Based on this scoring function, the 3D-partner server is able to evaluate how well the interfaces of pairing proteins and provides the statistic significance (Z-value), the binding models and functional annotations of interacting partners. The scoring function is defined as

$$E_{\text{tot}} = E_{\text{vdw}} + E_{\text{SF}} + wE_{\text{cons}} \quad \mathbf{1}$$

where E_{vdw} is the van der Waal's energy; E_{SF} is the special energy (i.e. hydrogen-bond energy, electrostatic energy and disulfide-bond energy); w is constant weight; and E_{cons} is the template consensus score. Here, w is set to 0.8. The E_{vdw} and E_{SF} are given as

$$E_{\text{vdw}} = \sum_{i,j}^{CP} \left(Vss_{ij} + Vsb_{ij} + Vsb_{ji}\right) \quad \mathbf{2}$$

$$E_{\text{SF}} = \sum_{i,j}^{CP} \left(Tss_{ij} + Tsb_{ij} + Tsb_{ji}\right) \quad \mathbf{3}$$

where CP denotes the number of the aligned-contact residues of proteins A and B aligned to a hit template; Vss_{ij} and Vsb_{ij} (Vsb_{ji}) are the sidechain–sidechain and sidechain–backbone van der Waals energies between residues i (in protein A) and j (in protein B), respectively. Tss_{ij} and Tsb_{ij} (Tsb_{ji}) are the sidechain–sidechain and sidechain-backbone special interacting energies between i and j, respectively, if the pair residues i and j form the special bonds (i.e. hydrogen bond, salt bridge or disulfide bond) in the template structure. The van der Waals energies (Vss_{ij}, Vsb_{ij} and Vsb_{ji}) and special interacting energies (Tss_{ij}, Tsb_{ij} and Tsb_{ji}) can be obtained from our four knowledge-based scoring matrices (Figure S1 in Supplementary Data), including sidechain–sidechain (Figure S1-A) and sidechain–backbone van-der Waals scoring matrices (Figure S1-B); and sidechain–sidechain (Figure S1-C) and sidechain–backbone special-bond scoring matrices (Figure S1-D). The sidechain–sidechain scoring matrices are symmetric and sidechain–backbone-scoring matrices are non-symmetric. The interaction scores from these matrices provide protein–protein interaction binding affinity estimates and preferences of two contacted residues in the interfaces.

For the sidechain-sidechain van-der Waals scoring matrix (Figure S1-A), the scores are high (yellow blocks) if large-aliphatic residues (i.e. Val, Leu, Ile and Met) interact to large-aliphatic residues or aromatic residues (i.e. Phe, Tyr and Trp) interact to aromatic residues. In contrast, the scores are low (orange blocks) when non-polar residues interact to polar residues. The top two highest scores are 3.0 (Met. to Met) and 2.9 (Trp to Trp). For the sidechain–sidechain special-bond scoring matrix (Figure S1-C), the score is high when the pair-interacting residue (i.e. Cys to Cys) forms a disulfide bond or basic residues (i.e. Arg, Lys and His) interact to acidic residues (Asp and Glu). The scoring values are zero if non-polar residues interact to other residues.

These four knowledge-based matrices are the key components of the 3D-partner for predicting protein–protein interactions. Here, a general mathematical structure (24) is used to construct these matrices. The entry (S_{ij}), which is the interacting score for a contact residue i, j pair ($1 \leq i, j \leq 20$), of a scoring matrix is defined as $S_{ij} = \ln q_{ij}/e_{ij}$, where q_{ij} and e_{ij} are the observed probability and the expected probability, respectively, of the occurrence of each i, j pair. These values of q_{ij} and e_{ij} are derived from a non-redundant set which consists of 621 3D-dimer complexes proposed by Glaser *et al.* (25) This dataset consists of 217 heterodimers and 404 homodimers and the sequence identity is <30% to each other.

The consensus score is defined as $E_{\text{cons}} = \sum_{i,j}^{CP} (C_{ij} + 0.5I_{ij})$, where C_{ij} and I_{ij} are the conserved score and template similar score for a contact residue i and j pair, respectively. They are given as

$$\begin{aligned} C_{ij} &= \max\left(0, \left(M_{ip} - K_{ii}\right) + \left(M_{jp'} - K_{jj}\right)\right) \\ I_{ij} &= \max\left(0, \left(M_{ip} - K_{ii} + 2\right) + \left(M_{jp'} - K_{jj} + 2\right)\right) \end{aligned} \quad \mathbf{4}$$

where M_{ip} is the score in the PSSM for residue type i at position p in Protein A; $M_{jp'}$ is the score in the PSSM for residue type j at position p' in Protein B; K_{ii} and K_{jj} are the diagonal scores of BLOSUM62 for residue type i and j.

Input, output and options

The 3D-partner is an easy-to-use web server. Users input the query protein sequence in FASTA format. Users are also able to assign a specific species for the query

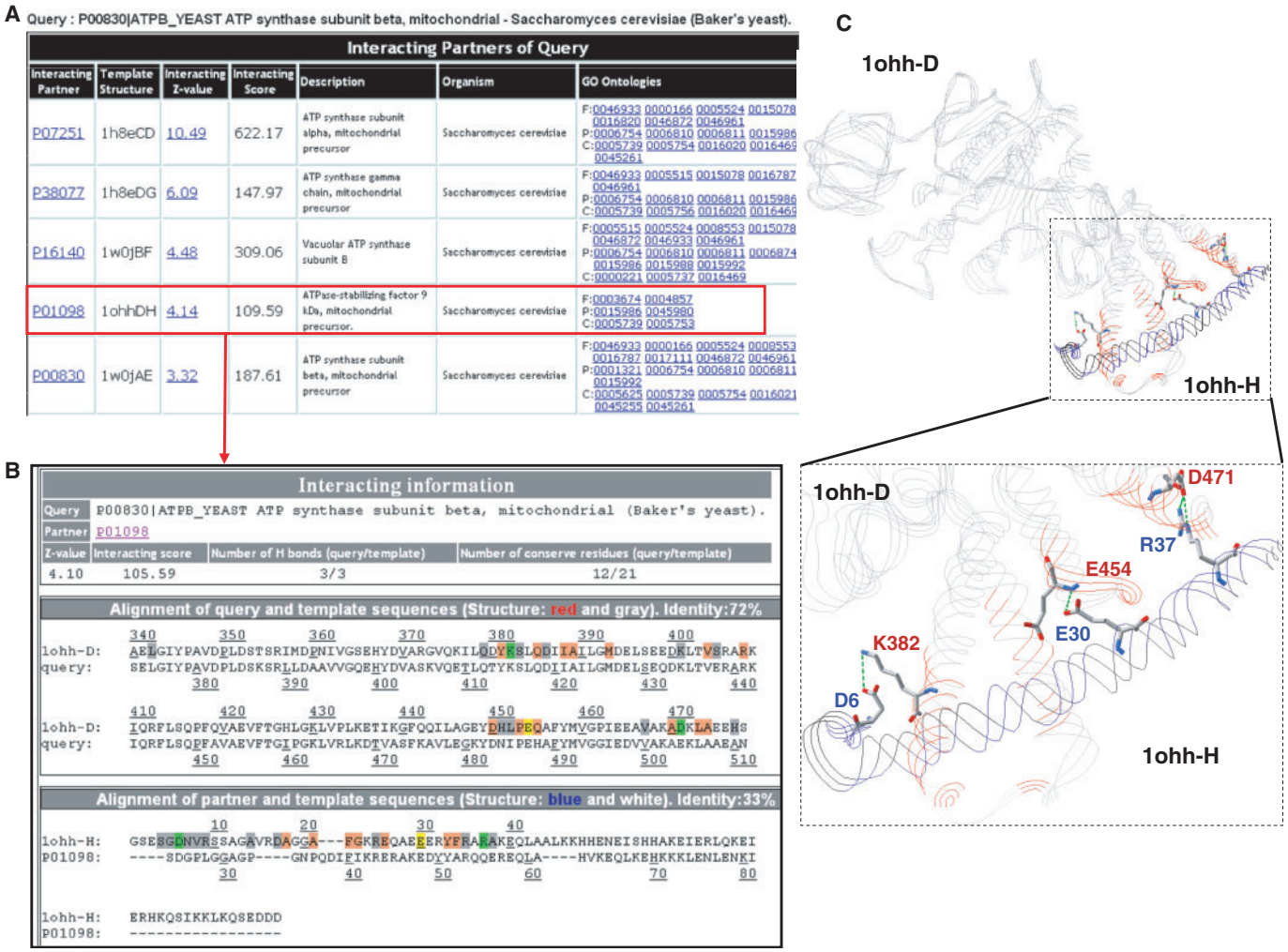

Figure 2. The 3D-partner server search results of using yeast ATP2 (SwissProt entry P00830) as query. (**A**) The 3D-partner server predicts five interacting partners of ATP2. For each interacting partner, this server provides the SwissProt entry, template structure with PDB entry, interacting *Z*-value and score, description, organism and Gene Ontology annotations. (**B**) Detailed interactions between the query and its interacting partner (SwissProt entry P01098). The server first presents summary interacting binding models, such as numbers of hydrogen bonds and conserved residue pairs. The alignments of both the query and its partner aligned to template sequences, respectively, are also indicated. The contacted residues are marked in template sequence based on their interacting characteristics, including hydrogen-bond residues (green); conserved residues (orange); both (yellow), and others (gray). In this example, D6, E30 and R37 of ATPase inhibitor (PDB entry 1ohh-H) form hydrogen bonds to K382, E454 and D471 of ATP synthase subunit beta (PDB entry 1ohh-D), respectively. (**C**) The template structure consists of ATPase inhibitor (black) and ATP synthase subunit beta (gray). The backbones are indicated in the ribbon model and the contact residues of 1ohh-D are colored by red and those of 1ohh-H are colored by blue. The residues forming hydrogen bonds (E454-E30 and D471-R37) and electrostatic interaction (K382-D6) are indicated.

sequence. The server typically returns a list of predicted interacting partners of the query ordered by *Z*-values, which represent the statistical significance of a hit interacting partner, within 20 s. For each predicted interacting partner, 3D-partner provides the visualization of the binding model between the query protein and its partner by aligning them to respective template sequences and structures (Figure 2). The important contact residues in the interface are indicated in the following formats: hydrogen-bond residues (green); conserved residues (orange), both (yellow), and other (gray). The structure is visualized in PNG format generated by MolScript and Raster3D packages. If the Java software is installed in a browser, the output will display the structures and users are allowed to dynamically view the binding model, interfaces, and the important residues in the browser.

Example analysis

The mitochondrial ATP synthase couples energy of the proton gradient across the mitochondrial membrane, derived by respiration, to the phosphorylation of ADP to ATP. The F_1 catalytic domain of ATP synthase has 3 catalytic sites formed by three pairs of α/β subunits, which are arranged as a sphere forming the core of the enzyme. The central stalk is in the center of the core of F_1 and is physically coupled to F_0 (26). If users want to known how assembly of yeast ATP synthase, the sequence of β subunits (ATP2) of yeast ATP synthase could be used

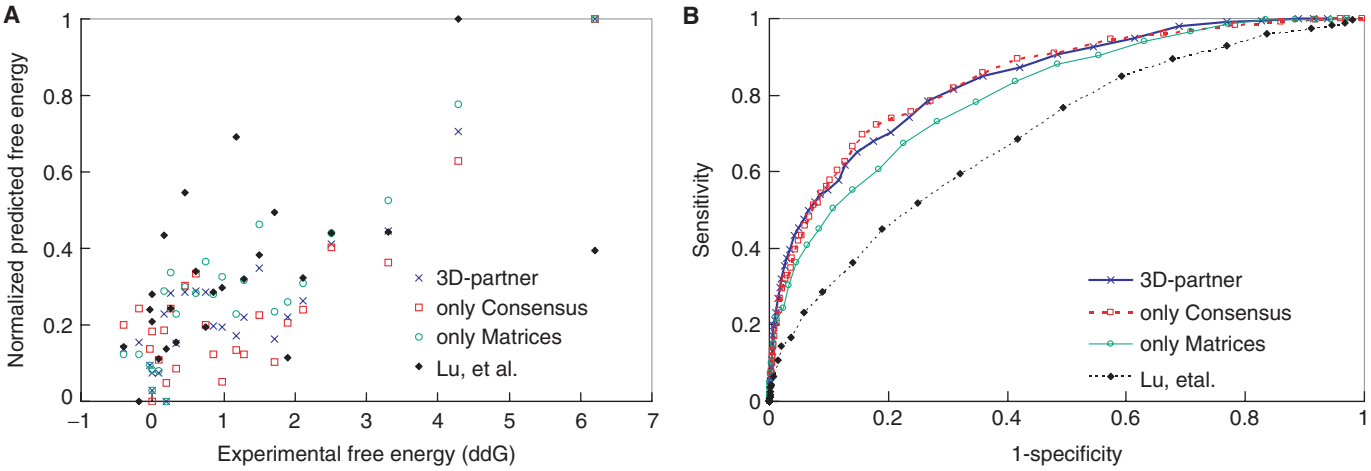

Figure 3. Evaluation of the 3D-partner server in (**A**) binding affinities prediction and (**B**) interacting partners prediction. (A) The correlations between experimental free energies (ddG) and the predicted values of the 3D-partner using four scoring functions, including E_{tot} (3D-partner, blue), E_{cons} (only Consensus, red), $E_{vdw}+E_{SF}$ (only Matrices, green) and one matrix (black) proposed by Lu *et al.*, on 275 mutated residues in Alanine Scanning Energetics database. (B) The ROC curves of the 3D-partner applying these four scoring functions on the data set NR-563. The predicted accuracy of 3D-partner was superior when using both knowledge-based scoring matrices and consensus scores. The performance of using one matrix is the worst among these four scoring functions.

to query 3D-partner server. Five proteins, including P07251 (ATP1), P38077 (ATP3), P01098 (STF1), P16140 (VMA2) and P00830 (ATP2) (Figure 2A) in yeast, are predicted to interact with the query protein. The interactions between the query and ATP1, ATP3 and ATP2 are recorded in the core subset in DIP database, but no structural data is available in PDB. ATP1 and ATP3, which are α and γ subunits of ATP synthase, respectively, bind to the query to form F_1 catalytic domain of ATP synthase (27). ATP2 is the same protein of the query, the aggregation of ATP2 occurred when the β-barrel domain of ATP1 is not expressed (28).

STF1 is the ATPase-stabilizing factor which involves in ATP synthase regulation. Currently, no study has demonstrated that the STF1 binds ATP2 directly. Hashimoto *et al.* (29) proved that STF1 binds to F_1 domain of ATP synthase and inhibits ATP synthase activity. In addition, a known structure of ATP synthase with inhibitor protein (PDB entry 1ohh) from *Bos taurus* (30) is able to be used as the template of the interacting model between STF1 and ATP2. STF1 and 1ohh-H chain share 33% sequence identity; and the important contact residues (D6, E30 and R37), which forms special bonds (i.e. hydrogen bonds and electrostatic interactions), are conserved in these two proteins (Figures 2B and 2C). These results suggest that the predicted interaction between STF1 and ATP2 is reasonably reliable and should be a novel interaction in yeast.

VMA2 may not interact to ATP2 based on the subcellular locations and functions. VMA2 is non-catalytic subunit of the peripheral V1 complex of vacuolar ATPase and the subcellular locations of VMA2 and ATP2 are vacuolar (GO:0016469) and mitochondria (GO:0005739), respectively, according to the annotations of Gene Ontology (31). In the future, our method will integrate functional annotations (e.g. Gene Ontology) to reduce the false-positive rate.

RESULTS

First, we evaluated our scoring function with different combinations on 275 mutated residues selected from the ASEdb database (18) to predict the binding affinities (Figure 3A). In addition, a non-redundant set (563 complexes) was used to evaluate the performance of the 3D-partner server using various scoring methods for interacting partner predictions (Figure 3B).

Binding affinity prediction

To determine the contribution of a residue to the binding affinity, the alanine-scanning mutagenesis is frequently used as an experimental probe. We selected 275 mutated residues from the ASEdb (18) with 16 heterodimers whose 3D structures were known. Those mutated residues should position at protein–protein interfaces and be the contact residues which were shown in the 3D-partner web server. ASEdb gives the corresponding delta G value representing the change in free energy of binding upon mutation to alanine for each experimentally mutated residue. Residues that contribute a large amount of binding energy are often labeled as hot spots of binding energy.

Based on the interacting characteristics, these 275 mutated residues can be divided into three types, including the special-force residues forming hydrogen bonds or electrostatic interactions; conserved residues if the conserved score [i.e. C_{ij} defined in Equation (4)] exceeds zero, and the other residues. The average and standard deviation of experimental ddG values are 1.92 and 1.97 for 99 special-force residues, respectively. For 176 non-special-force residues, the average and standard deviation of ddG values are 0.8 and 1.06, respectively. Standard two-sample *t*-test shows that the mean of ddG values for special-force residues is significantly higher (P-value $< 10^{-6}$) than that of non-special-force residues. At the same time, for 71 conserved residues, the average

and standard deviation of experimental ddG values are 1.77 and 2.14, respectively, and these two values are 1.0 (average) and 1.23 (standard deviation) for 204 non-conserved residues. The *P*-value of standard two-sample *t*-test is 0.005 and shows that the mean of ddG values for conserved residues is significantly higher than that of non-conserved residues. These results suggest that special-force and conserved residues should be more important than the other residues in the interacting surface, and the scoring matrix could be divided into van der Waal's energy (E_{vdw}) and special energy (E_{SF}).

Figure 3A illustrates the correlations between ddG values and predicted energies of the 3D-partner server applying four different scoring functions, including E_{tot} (3D-partner using both consensus and matrices), E_{cons} (only consensus), $E_{vdw}+E_{SF}$ (only matrices), and one matrix proposed by Lu, *et al.* (15), on 275 mutated residues, where E_{tot}, E_{cons}, E_{vdw} and E_{SF} are defined in Equation (1). Among these four scoring functions, the 3D-partner server applying both consensus and matrices is the best (0.91) and one matrix is the worst (0.54). The correlations are 0.91 and 0.84 for using only matrices and consensus, respectively.

Interacting partner prediction

Several metrics were utilized to assess the predicted quality of the 3D-partner server. Precision is defined as A_h/T_h and recall is defined as A_h/A, where A_h is the number of true hits in the hit list, T_h is the total number of hits in the hit list, and A is total number of true hits in the database. The ROC curve plots the sensitivity (i.e. recall) against the '1.0-specificity' (i.e. false-positive rate). The average precision is defined as $(\sum_{i=1}^{A} i/T_h^i)/A$, where T_h^i is the number of compounds in a hit list containing i correct hits.

As the interactions in *Saccharomyces cerevisiae* are the most extensive, reliable and well studied, we measured the quality of our predicted interactions in *S. cerevisiae*. In order to evaluate performance of the 3D-partner server using various scoring methods, we selected a non-redundant set, called NR-563. This set consists of 563 dimer complexes from the 3D-dimer library according to their SCOP interacting-domain pairs. At least one chain of these complexes has different SCOP family. Protein–protein interactions to the tune of 5882 recorded as the core subset in the DIP database were used as the positive cases and 2 708 746 non-interacting protein pairs defined by Jansen *et al.* (5) were selected as the negative cases. The 3D-domain sequence profiles (i.e. 1026 sequences) of 563 complexes were used as the queries to search on the yeast proteome (6714 sequences) selected from SGD (32) by using PSI-BLAST. Based on this set, the 3D-partner can yield 4206 protein–protein interaction candidates by setting three criteria, including the sequence identity is >15%, *CR* exceeds 50%, and *E*-value is <0.05. Among these 4206 candidates, 226 (CR > 80% and sequence identity >25%) and 3980 protein–protein candidates were recoded in the positive and negative sets, respectively (Figure S2 in Supplementary Data).

The 3D-partner server tested these four different scoring functions (i.e. E_{tot}, only E_{cons}, only $E_{vdw}+E_{SF}$ and one matrix) on these 4206 candidates. The average precisions of these four methods are 0.49 (one matrix), 0.65 (only $E_{vdw}+E_{SF}$), 0.70 (only E_{cons}) and 0.72 (E_{tot}). The ROC curve (Figure 3B) provides an estimation of the likely number of true-positive and false-positive predictions. The 3D-partner server using both scoring matrices ($E_{vdw}+E_{SF}$) and consensus scores (E_{cons}) yields much better predictions than one matrix (black). The performance of using only scoring matrices and only consensus scores are also much better than that of one matrix.

The 3D-partner provides a threshold *Z*-score to identify interacting partners with the query. The Z-score reveals that the proportion of true positives rises when a higher *Z*-score is utilized (Figure S2 in Supplementary Data). If the sequence identity is restricted to over 25%, the sensitivity is 0.29 and precision is 0.79 (*Z*-score > 5); the sensitivity is 0.37 and precision is 0.70 (*Z*-score > 4); and the sensitivity and precision are 0.41 and 0.61, respectively (*Z*-score > 3). If the sequence identity is restricted to over 30%, the sensitivity is 0.20 and precision is 0.83 (*Z*-score > 5); the sensitivity is 0.25 and precision is 0.81 (*Z*-score > 4); and the sensitivity and precision are 0.27 and 0.76, respectively (*Z*-score > 3).

Although the query and partner proteins can be potential interactors based on reasoning that they are homologous to interacting domains of the template, they might be not structurally similar to the template structure. The 3D-partner server is able to reduce the ill-effect if the sequence identity and the Z-score are restricted to over 30% and 3.5, respectively (Figure S2 in Supplementary Data). This result is consistent with the previous results, pairs of interacting proteins can be considered structurally similar if their sequence identity is >30%, proposed by Aloy *et al.* (33).

CONCLUSION

This study demonstrates the robustness and feasibility of the 3D-partner server to infer interacting partners and binding models. The key novelty of the present work is the cooperative integration of the 3D-domain interologs and a new scoring function; the former uses interacting-domain sequence profile to search candidates for many species efficiently and the latter evaluates candidates reliably. Our scoring function achieves good agreement for the binding affinity in protein–protein interactions and provides the statistic significance (*Z*-value) for predicting protein–protein interactions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

J.-M.Y. was supported by National Science Council and partial support of the ATU plan by MOE. Authors are grateful to both the hardware and software supports of the Structural Bioinformatics Core Facility at National

Chiao Tung University. Funding to pay the Open Access publication charges for this article was provided by National Science Council.

Conflict of interest statement. None declared.

REFERENCES

1. Cary,M.P., Bader,G.D. and Sander,C. (2005) Pathway information for systems biology. *FEBS Lett.*, **579**, 1815–1820.
2. Ito,T., Chiba,T., Ozawa,R., Yoshida,M., Hattori,M. and Sakaki,Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA*, **98**, 4569–4574.
3. Pandey,A. and Mann,M. (2000) Proteomics to study genes and genomes. *Nature*, **405**, 837–846.
4. von Mering,C., Krause,R., Snel,B., Cornell,M., Oliver,S.G., Fields,S. and Bork,P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, **417**, 399–403.
5. Jansen,R., Yu,H., Greenbaum,D., Kluger,Y., Krogan,N.J., Chung,S., Emili,A., Snyder,M., Greenblatt,J.F. *et al.* (2003) A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, **302**, 449–453.
6. Ng,S.K., Zhang,Z. and Tan,S.H. (2003) Integrative approach for computationally inferring protein domain interactions. *Bioinformatics*, **19**, 923–929.
7. Itzhaki,Z., Akiva,E., Altuvia,Y. and Margalit,H. (2006) Evolutionary conservation of domain-domain interactions. *Genome Biol.*, **7**, R125.
8. Deng,M., Mehta,S., Sun,F. and Chen,T. (2002) Inferring domain-domain interactions from protein-protein interactions. *Genome Res.*, **12**, 1540–1548.
9. Pellegrini,M., Marcotte,E.M., Thompson,M.J., Eisenberg,D. and Yeates,T.O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad. Sci. USA*, **96**, 4285–4288.
10. Aloy,P. and Russell,R.B. (2002) Interrogating protein interaction networks through structural biology. *Proc. Natl Acad. Sci. USA*, **99**, 5896–5901.
11. Lu,L., Arakaki,A.K., Lu,H. and Skolnick,J. (2003) Multimeric threading-based prediction of protein-protein interactions on a genomic scale: application to the Saccharomyces cerevisiae proteome. *Genome Res.*, **13**, 1146–1154.
12. Matthews,L.R., Vaglio,P., Reboul,J., Ge,H., Davis,B.P., Garrels,J., Vincent,S. and Vidal,M. (2001) Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Genome Res.*, **11**, 2120–2126.
13. Jonsson,P.F., Cavanna,T., Zicha,D. and Bates,P.A. (2006) Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis. *BMC Bioinformatics*, **7**, 2.
14. Aloy,P., Bottcher,B., Ceulemans,H., Leutwein,C., Mellwig,C., Fischer,S., Gavin,A.C., Bork,P., Superti-Furga,G. *et al.* (2004) Structure-based assembly of protein complexes in yeast. *Science*, **303**, 2026–2029.
15. Lu,L., Lu,H. and Skolnick,J. (2002) MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. *Proteins*, **49**, 350–364.
16. Chen,Y.-C., Chen,H.-C. and Yang,J.-M. (2006) DAPID: a 3D-domain annotated protein-protein interaction database. *Genome Informatics*, **17**, 206–215.
17. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.C., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.
18. Thorn,K.S. and Bogan,A.A. (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*, **17**, 284–285.
19. Deshpande,N., Addess,K.J., Bluhm,W.F., Merino-Ott,J.C., Townsend-Merino,W., Zhang,Q., Knezevich,C., Xie,L., Chen,L. *et al.* (2005) The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.*, **33**, D233–D237.
20. Ofran,Y. and Rost,B. (2003) Analysing six types of protein-protein interfaces. *J. Mol. Biol.*, **325**, 377–387.
21. Park,J., Lappe,M. and Teichmann,S.A. (2001) Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast. *J. Mol. Biol.*, **307**, 929–938.
22. Andreeva,A., Howorth,D., Brenner,S.E., Hubbard,T.J., Chothia,C. and Murzin,A.G. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res.*, **32**, D226–D229.
23. Wu,C.H., Apweiler,R., Bairoch,A., Natale,D.A., Barker,W.C., Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H. *et al.* (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, **34**, D187–D191.
24. Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA*, **89**, 10915–10919.
25. Glaser,F., Steinberg,D.M., Vakser,I.A. and Ben-Tal,N. (2001) Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins: Struct. Funct. Bioinformatics*, **43**, 89–102.
26. Gibbons,C., Montgomery,M.G., Leslie,A.G. and Walker,J.E. (2000) The structure of the central stalk in bovine F(1)-ATPase at 2.4 A resolution. *Nat. Struct. Biol.*, **7**, 1055–1061.
27. Arnold,I., Pfeiffer,K., Neupert,W., Stuart,R.A. and Schagger,H. (1999) ATP synthase of yeast mitochondria. Isolation of subunit j and disruption of the ATP18 gene. *J. Biol. Chem.*, **274**, 36–40.
28. Yao,B. and Mueller,D.M. (1999) The role of the amino-terminal beta-barrel domain of the alpha and beta subunits in the yeast F1-ATPase. *J. Bioenerg. Biomembr.*, **31**, 95–104.
29. Hashimoto,T., Yoshida,Y. and Tagawa,K. (1990) Regulatory proteins of F1F0-ATPase: role of ATPase inhibitor. *J. Bioenerg. Biomembr.*, **22**, 27–38.
30. Cabezon,E., Montgomery,M.G., Leslie,A.G. and Walker,J.E. (2003) The structure of bovine F1-ATPase in complex with its regulatory protein IF1. *Nat. Struct. Biol.*, **10**, 744–750.
31. Harris,M.A., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**, D258–D261.
32. Cherry,J.M., Ball,C., Weng,S., Juvik,G., Schmidt,R., Adler,C., Dunn,B., Dwight,S., Riles,L. *et al.* (1997) Genetic and physical maps of Saccharomyces cerevisiae. *Nature*, **387**, 67–73.
33. Aloy,P., Pichaud,M. and Russell,R.B. (2005) Protein complexes: structure prediction challenges for the 21st century. *Curr. Opin. Struct. Biol.*, **15**, 15–22.