(PS)²: protein structure prediction server version 3.0

Tsun-Tsao Huang[1,2], Jenn-Kang Hwang[1,2,3,4], Chu-Huang Chen[5,6,7,8], Chih-Sheng Chu[6,9], Chi-Wen Lee[2,3] and Chih-Chieh Chen[6,10,11,*]

[1]Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu 30068, Taiwan, [2]Center for Bioinformatics Research, National Chiao Tung University, Hsinchu 30068, Taiwan, [3]Department of Biological Science and Technology, National Chiao Tung University, Hsinchu 30068, Taiwan, [4]Department of Bioinformatics and Medical Engineering, Asia University, Taichung 41354, Taiwan, [5]Department of Vascular and Medicinal Research, Texas Heart Institute, Houston, TX 77030, USA, [6]Center for Lipid Biosciences, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 80708, Taiwan, [7]L5 Research Center, China Medical University Hospital, China Medical University, Taichung 40402, Taiwan, [8]Section of Cardiovascular Research, Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA, [9]Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 80708, Taiwan, [10]Center for Lipid and Glycomedicine Research, Kaohsiung Medical University, Kaohsiung 80708, Taiwan and [11]Institute of Medical Science and Technology, National Sun Yat-sen University, Kaohsiung 80424, Taiwan

ABSTRACT

Protein complexes are involved in many biological processes. Examining coupling between subunits of a complex would be useful to understand the molecular basis of protein function. Here, our updated (PS)² web server predicts the three-dimensional structures of protein complexes based on comparative modeling; furthermore, this server examines the coupling between subunits of the predicted complex by combining structural and evolutionary considerations. The predicted complex structure could be indicated and visualized by Java-based 3D graphics viewers and the structural and evolutionary profiles are shown and compared chain-by-chain. For each subunit, considerations with or without the packing contribution of other subunits cause the differences in similarities between structural and evolutionary profiles, and these differences imply which form, complex or monomeric, is preferred in the biological condition for the subunit. We believe that the (PS)² server would be a useful tool for biologists who are interested not only in the structures of protein complexes but also in the coupling between subunits of the complexes. The (PS)² is freely available at http://ps2v3.life.nctu.edu.tw/.

INTRODUCTION

Proteins are important molecules involved in almost all biological processes. The analysis of protein three-dimensional (3D) structures plays important role in understanding the molecular basis of their functions. Comparative modeling is one of the most important computational approaches to predict structures from amino acid sequences. Our previous comparative modeling web tools, (PS)² (1,2), have been widely used by many studies for proteomic research and have been cited in various journals such as *The American Journal of Human Genetics* (3), *Human mutation* (4,5), *BBA Reviews on Cancer* (6), *Journal of Clinical Microbiology* (7). Due to many fundamental cellular processes are mediated by protein–protein interactions, the prediction of protein complex structure is more and more important in proteomic studies. Indeed, many comparative modeling tools have been published to build complex structures (8–15). However, the predicted complex structures emerge a relevant question: whether individual subunit itself is enough to perform biological function or the whole complex structure is necessary in biological condition. Recently, Chang *et al.* addressed this issue only through comparing packing density and sequence conservation profiles, where only structure coordinates and sequence homologs are required to obtain these two properties (16).

Previous studies have shown that packing density is correlated with the sequence conservation (17–20). Furthermore, Shih *et al.* showed that the reciprocal value of packing density is proportional to the sequence conservation at residue level (20) and this linear relationship could be explained by a mechanistic model based on statistical physics (21). Using various methods to calculate packing density and sequence conservation, Yeh *et al.* found that the highest correlation between these two properties occurs when the packing density is estimated by weighted contact num-

ber (WCN) (22) as well as the sequence conservation is estimated by ConSurf (23,24). In order to obtain positive correlation between packing density and ConSurf profiles, the reciprocal WCN (rWCN) is used in these studies (16,18,20,25). For the given 3D structures of protein complexes, there are two ways to compute rWCN profile for each protein subunit: one ignores the packing contribution of the other subunits of the complex (denoted by rWCN I) and the other considers packing contribution of the other subunits (denoted by rWCN II). Chang *et al.* showed that rWCN I better agrees with sequence conservation for the set of enzymes whose active sites are located in individual subunits; in contrast, rWCN II better agrees with sequence conservation for the set of enzymes whose active sites comprise catalytic residues from multiple subunits (16). These findings imply that an individual subunit might be enough to perform biological function when its rWCN I profile better agrees with sequence conservation profile; on the contrary, complex form is preferred in biological condition when rWCN II profile better agrees with sequence conservation profile. As a consequence, Chang *et al.* suggested that comparing packing density and conservation profiles is a novel way of looking at couplings between subunits of a complex by comparing structural and sequence conservation profiles (16).

Here, we report the updated (PS)^2 server that automatically predicts protein complex structures from query sequences and shows the coupling between subunits of predicted complex structures based on the findings of Chang *et al.* The updated (PS)^2 accepts protein sequence in FASTA format and tries to build homologous 3D structures using (PS)^2 methodology (1,2), which was previously developed by one of the authors and is based on effective consensus strategy in both template selection and target-template alignment to build homologous structures; afterward, the updated (PS)^2 provides the packing density profiles calculated by rWCN I and rWCN II (16) for the predicted structure, as well as provides the sequence conservation profile calculated by ConSurf (23,24). The correlation coefficients between these structure-derived and sequence-derived profiles give a clue about the degree of coupling among the subunits of the predicted complex structure. Finally, the predicted tertiary structure and the structure-sequence profile comparisons can be visualized and indicated by Java-based 3D graphics viewers. This web server is freely available at http://ps2v3.life.nctu.edu.tw.

MATERIALS AND METHODS

The workflow of (PS)^2 is schematically shown in Figure 1. One of the new features of the updated (PS)^2 server is 3D structure prediction for protein complexes. Users can enter two query protein sequences, then the server models 3D complex structure through the (PS)^2 homology modeling strategies (1,2). The complex template dataset, consisting of 56547 3D structures, contains all available protein complexes in Protein Data Bank. With the predicted 3D structure and the query sequences, the packing density and the sequence conservation of each residue are respectively calculated by WCN (22) and ConSurf (24).
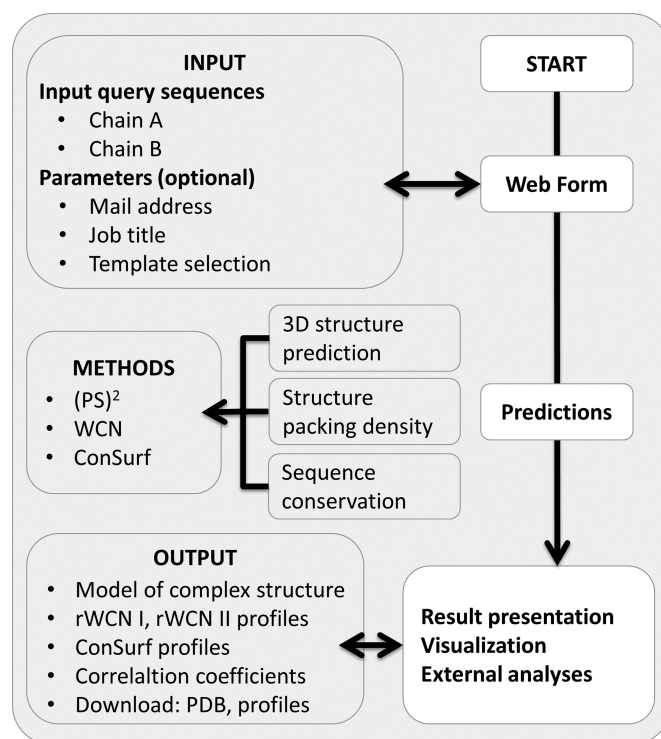

**Figure 1.** The schematic workflow of (PS)^2 version 3.0.

For a monomeric structure, the WCN of the ith residue is $WCN_i = \sum_{i \neq j}^{N} \frac{1}{d_{ij}^2}$, the summation of the square of inversed Cα–Cα distances to the other residues (22). In order to compare with the conservation score calculated by ConSurf (24), the reciprocal value of WCN (rWCN) is used. For a dimeric structure, rWCN could be calculated by two ways (16): rWCN I only considers residues in the same subunit (chain), while rWCN II considers residues in the whole complex. In the result page, profiles of rWCN I, rWCN II and ConSurf are z-score normalized by $z = \frac{x - \bar{x}}{\sigma_x}$, where $\bar{x}$ and σ_x are the mean and the standard deviation of x, respectively.

The new (PS)^2 server is built on the basis of the WCN, ConSurf and the original (PS)^2 systems, where the MODELLER is upgraded to version 9v8 (26) and Perl scripts are used to integrate the prediction pipeline. The web page is constructed by HTML and PHP. This web server runs on a Linux system with 2.40 GHz Intel Xeon processors consisting of 24 cores. The OpenAstexViewer (http://openastexviewer.net) was used for visualization of the predicted models.

WEB SERVER

Input format

The (PS)^2 web server accepts two types of input (Figure 2A): users could paste protein sequence or upload a sequence file in FASTA format. Furthermore, there are three ways for template selection: (i) our web server automatically selects templates based on the *E*-values from sequence alignments, (ii) users could assign a PDB structure and choose which chains would be used as template, (iii) users could up-
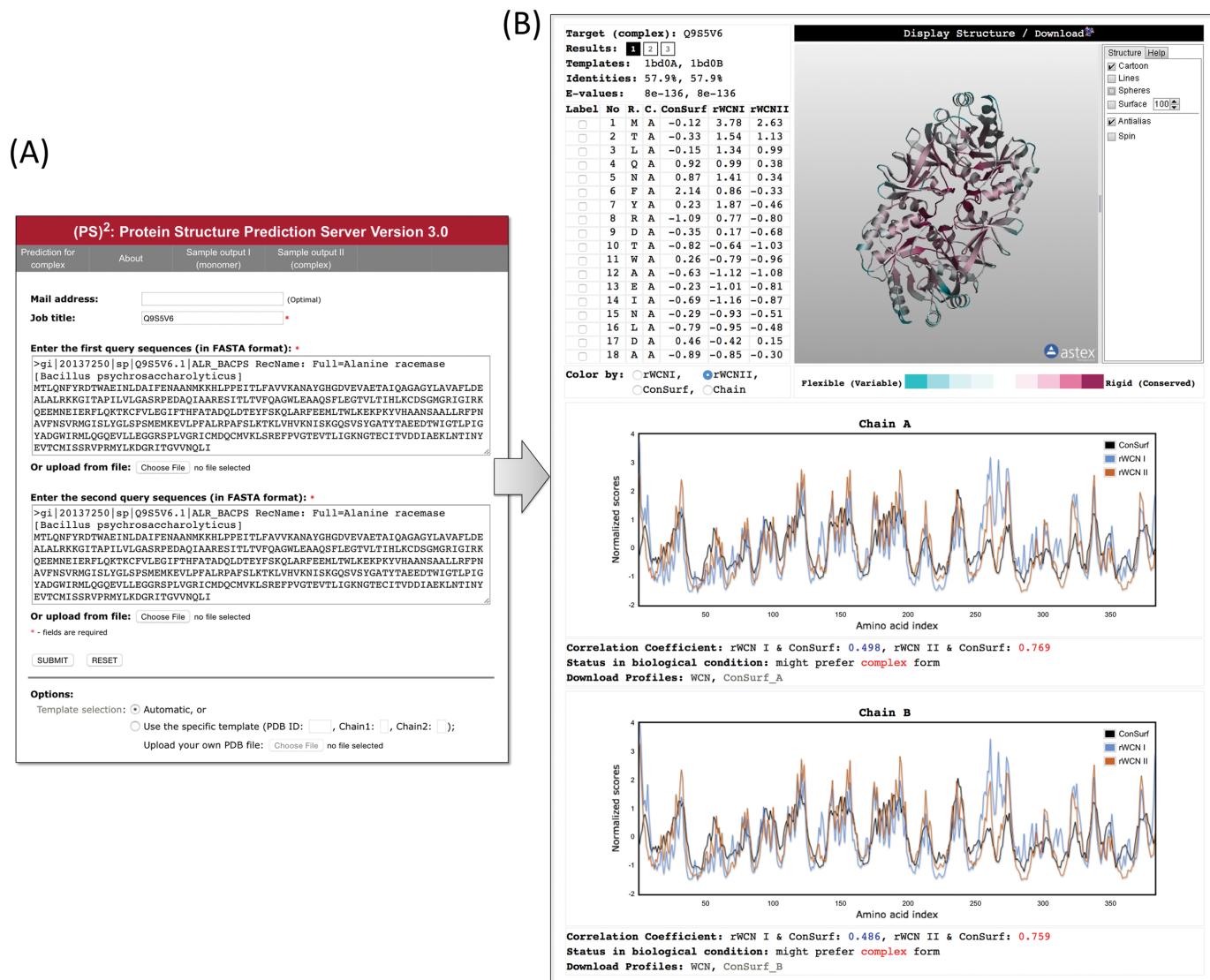
Figure 2. The features of the (PS)^2 web server: (**A**) Input: users can enter or upload protein sequences in FASTA format. (**B**) Output: users can view the predicted structure in 3D graphics viewer (top-right panel) as well as get the structural and sequence properties for each residue (top-left panel) and can compare the sequence conservation and packing density profiles chain-by-chain (bottom panel).

load their own 3D structure as template. It should be noted that, for the 'prediction for complex', the entered specific template must be a complex structure and the input chain IDs, 'Chain1' and 'Chain2', must be different. For a protein with about 400 amino acids, it costs about 10 min for the template-based structure modeling, packing density calculation with and without the consideration of other subunits, but it costs about 30 min for sequence conservation calculation. For a longer sequence, it may take more than one hour in run time because of the time-consuming calculation for sequence conservation. Therefore, users are encouraged to enter their e-mail addresses so that notification will be sent by e-mail when the submitted job is finished.

Output format

After the prediction, the updated (PS)^2 server will return two types of results (Figure 2B): (i) if the 3D structure being

successfully built, the visualization of its predicted complex structure (Display Structure) will be shown in the right region of the top panel. (ii) The sequence conservation and structure packing density of each residue will be shown in the left region of the top panel, including rWCN I, rWCN II and ConSurf. The server integrates these two types of results for users to easily analyze and view 3D structures and manipulate their orientations in space. If clicked on the check box of the (Label) column of the top-left panel, the selected residues will be labeled in 3D graphics viewer of the top-right panel. The different structure display modes (e.g. Cartoon, Lines, Spheres and Surface) can be visualized together or individually in the 3D graphics viewer for easy analysis. The (PS)^2 server also allows user to download the predicted structure coordinates in the PDB format.
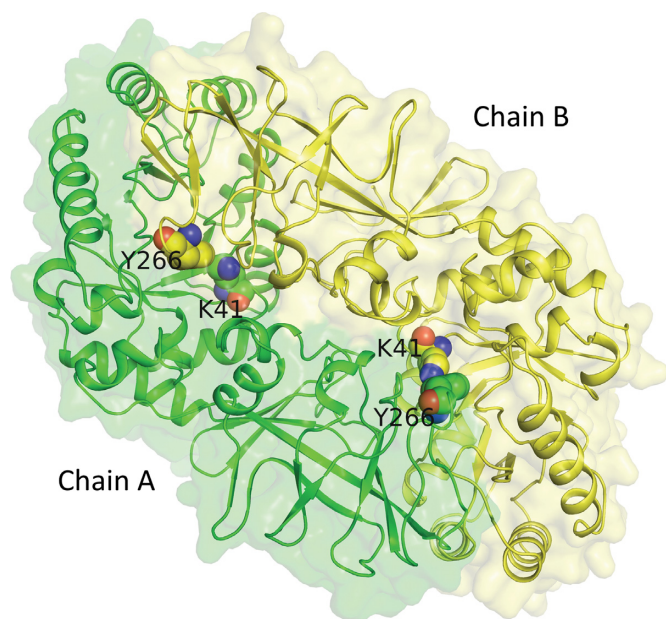
Figure 3. The predicted complex structure of alanine racemase (ALR_BACPS) by (PS)^2 server. Structure models of ALR_BACPS complex are built using sequence homolog (PDB ID: 1BD0, chains A and B) as their templates. The active-site residues Lys41 and Tyr266 are shown in spheres mode.

EXAMPLE ANALYSIS

The alanine racemase (gene name: ALR_BACPS, UniProt ID: Q9S5V6) from *Bacillus psychrosaccharolyticus* catalyzes the alanine racemization between the L and D forms. Based on the phenylhydrazine method of Wada and Snell (27), the study of Inagaki *et al.* indicated that alanine racemases are dimeric complexes (28). In order to build ALR_BACPS complex model, the (PS)^2 server automatically selects the alanine racemase from *Geobacillus stearothermophilus* (PDB ID: 1BD0, chains A and B) (29) as the template whose sequences are most similar to the query sequence (about 58%) for the whole PDB database. The complex structure predicted by (PS)^2 server is shown in the top-right panel of Figure 2B, while the packing density (both rWCN I and rWCN II) as well as sequence conservation (ConSurf) of each residue are shown in the top-left panel. In order to compare structural and sequence profiles for each subunit, the profiles of rWCN I, rWCN II and ConSurf are shown in the bottom panel chain-by-chain. For each individual chain, the Pearson's correlation coefficients for rWCN I-ConSurf and rWCN II-ConSurf are shown in the bottom of the profile comparison figure. Take the profiles of chain A as example, it is clear that rWCN II have better agreement with ConSurf (0.769) than rWCN I (0.498) and this is the same for chain B (0.759 > 0.486). Based on the findings of Chang *et al.* (16) and the profile comparisons results, we infer that alanine racemases are dimers in biological condition. This inference is consistent with the experimental results from Inagaki *et al.* (28).

Figure 3 shows the predicted complex structure of ALR_BACPS by the (PS)^2 server. As a result, the left active site consists of Lys41 from chain A and Tyr266 from chain

B and so the right active site did it too. Since each active site of alanine racemase consisting of residues from different chains, dimer form is necessary to perform the function. This is consistent with the inference that alanine racemase is dimer based on the method proposed by Chang *et al.* (16). In summary, this example shows that our (PS)^2 server not only correctly predicts the protein complex structure but also has the potential to predict biological functional unit.

CONCLUSION

Here we present an updated (PS)^2 web server for predicting complex structures and analyzing these structures by integrating ConSurf, WCN and (PS)^2 systems. One of the unique features of (PS)^2 is the integration of packing density and sequence conservation analysis into complex structure prediction, and this integration predicts structures of protein complexes as well as points out that the biological functional unit should be monomeric or complex form. The example demonstrates that the updated (PS)^2 server is effective for complex structure prediction and provides a new way to look at the coupling between subunits. We believe that the (PS)^2 server will be useful to general biologists.

REFERENCES

1. Chen,C.C., Hwang,J.K. and Yang,J.M. (2006) (PS)2: protein structure prediction server. *Nucleic Acids Res.*, **34**, W152–W157.
2. Chen,C.C., Hwang,J.K. and Yang,J.M. (2009) (PS)2-v2: template-based protein structure prediction server. *BMC Bioinformatics*, **10**, 366.
3. Weber,S., Thiele,H., Mir,S., Toliat,M.R., Sozeri,B., Reutter,H., Draaken,M., Ludwig,M., Altmuller,J., Frommolt,P. *et al.* (2011) Muscarinic acetylcholine receptor M3 mutation causes urinary bladder disease and a prune-belly-like syndrome. *Am. J. Hum. Genet.*, **89**, 668–674.
4. Soltys,D.T., Rocha,C.R., Lerner,L.K., de Souza,T.A., Munford,V., Cabral,F., Nardo,T., Stefanini,M., Sarasin,A., Cabral-Neto,J.B. *et al.* (2013) Novel XPG (ERCC5) mutations affect DNA repair and cell survival after ultraviolet but not oxidative stress. *Hum. Mutat.*, **34**, 481–489.

5. Kuo,Y.C., Lin,Y.H., Chen,H.I., Wang,Y.Y., Chiou,Y.W., Lin,H.H., Pan,H.A., Wu,C.M., Su,S.M., Hsu,C.C. *et al.* (2012) SEPT12 mutations cause male infertility with defective sperm annulus. *Hum. Mutat.*, **33**, 710–719.

6. Friedman,R., Boye,K. and Flatmark,K. (2013) Molecular modelling and simulations in cancer research. *Biochim. Biophys. Acta*, **1836**, 1–14.

7. Huang,S.W., Hsu,Y.W., Smith,D.J., Kiang,D., Tsai,H.P., Lin,K.H., Wang,S.M., Liu,C.C., Su,I.J. and Wang,J.R. (2009) Reemergence of enterovirus 71 in 2008 in taiwan: dynamics of genetic and antigenic evolution from 1998 to 2008. *J. Clin. Microbiol.*, **47**, 3653–3662.

8. Mosca,R., Ceol,A. and Aloy,P. (2013) Interactome3D: adding structural details to protein networks. *Nat. Methods*, **10**, 47–53.

9. Mukherjee,S. and Zhang,Y. (2011) Protein-protein complex structure predictions by multimeric threading and template recombination. *Structure*, **19**, 955–966.

10. Guerler,A., Govindarajoo,B. and Zhang,Y. (2013) Mapping monomeric threading to protein-protein structure prediction. *J. Chem. Inf. Model.*, **53**, 717–725.

11. Szilagyi,A. and Zhang,Y. (2014) Template-based structure modeling of protein-protein interactions. *Curr. Opin. Struct. Biol.*, **24**, 10–23.

12. 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.

13. Baspinar,A., Cukuroglu,E., Nussinov,R., Keskin,O. and Gursoy,A. (2014) PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. *Nucleic Acids Res.*, **42**, W285–W289.

14. Kundrotas,P.J., Lensink,M.F. and Alexov,E. (2008) Homology-based modeling of 3D structures of protein-protein complexes using alignments of modified sequence profiles. *Int. J. Biol. Macromol.*, **43**, 198–208.

15. Fukuhara,N. and Kawabata,T. (2008) HOMCOS: a server to predict interacting protein pairs and interacting sites by homology modeling of complex structures. *Nucleic Acids Res.*, **36**, W185–W189.

16. Chang,C.M., Huang,Y.W., Shih,C.H. and Hwang,J.K. (2013) On the relationship between the sequence conservation and the packing density profiles of the protein complexes. *Proteins*, **81**, 1192–1199.

17. Franzosa,E.A. and Xia,Y. (2009) Structural determinants of protein evolution are context-sensitive at the residue level. *Mol. Biol. Evol.*, **26**, 2387–2395.

18. Yeh,S.W., Liu,J.W., Yu,S.H., Shih,C.H., Hwang,J.K. and Echave,J. (2014) Site-specific structural constraints on protein sequence evolutionary divergence: local packing density versus solvent exposure. *Mol. Biol. Evol.*, **31**, 135–139.

19. Liao,H., Yeh,W., Chiang,D., Jernigan,R.L. and Lustig,B. (2005) Protein sequence entropy is closely related to packing density and hydrophobicity. *Protein Eng. Des. Sel.*, **18**, 59–64.

20. Shih,C.H., Chang,C.M., Lin,Y.S., Lo,W.C. and Hwang,J.K. (2012) Evolutionary information hidden in a single protein structure. *Proteins*, **80**, 1647–1657.

21. Huang,T.T., del Valle Marcos,M.L., Hwang,J.K. and Echave,J. (2014) A mechanistic stress model of protein evolution accounts for site-specific evolutionary rates and their relationship with packing density and flexibility. *BMC Evol. Biol.*, **14**, 78.

22. Lin,C.P., Huang,S.W., Lai,Y.L., Yen,S.C., Shih,C.H., Lu,C.H., Huang,C.C. and Hwang,J.K. (2008) Deriving protein dynamical properties from weighted protein contact number. *Proteins*, **72**, 929–935.

23. Goldenberg,O., Erez,E., Nimrod,G. and Ben-Tal,N. (2009) The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res.*, **37**, D323–D327.

24. Ashkenazy,H., Erez,E., Martz,E., Pupko,T. and Ben-Tal,N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38**, W529–W533.

25. Yeh,S.W., Huang,T.T., Liu,J.W., Yu,S.H., Shih,C.H., Hwang,J.K. and Echave,J. (2014) Local packing density is the main structural determinant of the rate of protein sequence evolution at site level. *Biomed. Res. Int.*, **2014**, 572409.

26. Webb,B. and Sali,A. (2014) Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics*, **47**, 5.6.1–5.6.32.

27. Wada,H. and Snell,E.E. (1961) The enzymatic oxidation of pyridoxine and pyridoxamine phosphates. *J. Biol. Chem.*, **236**, 2089–2095.

28. Inagaki,K., Tanizawa,K., Badet,B., Walsh,C.T., Tanaka,H. and Soda,K. (1986) Thermostable alanine racemase from Bacillus stearothermophilus: molecular cloning of the gene, enzyme purification, and characterization. *Biochemistry*, **25**, 3268–3274.

29. Stamper,G.F., Morollo,A.A. and Ringe,D. (1998) Reaction of alanine racemase with 1-aminoethylphosphonic acid forms a stable external aldimine. *Biochemistry*, **37**, 10438–10445.