Original Article

3D MODELING AND CHARACTERIZATION OF HDAC9

LALIT R. SAMANT[1*], VIKRANT C. SANGAR[2], MADHURA KHANZODE[3], ABHAY CHOWDHARY[1,2]

[1]Systems Biomedicine Division, Haffkine Institute for Training, Research & Testing, Acharya Donde Marg, Parel, Mumbai 400012, India, [2]Department of Virology & Immunology, Haffkine Institute for Training, Research & Testing, Acharya Donde Marg, Parel, Mumbai 400012, India, [3]Department of Biotechnology and Bioinformatics, Padmashree Dr. D. Y. Patil University, Belapur CBD, Navi Mumbai 400614, India. Email: samantlalit@gmail.com

ABSTRACT

Objective: Histones are the most abundant proteins associated with the eukaryotic DNA. The N-terminal tails of these histones are subjected to modifications primarily by two enzymes namely, Histone acetyl transferases (HATs) and Histone deacetylases (HDACs). HDACs help in the regulation of the acetylation of histones and the condensation of the chromatin in its sTable form. HDACs are considered as one of the promising targets in cancer biology studies. HDAC9 is a class II member of HDAC family and they are associated with many neurological disorders and a variety of cancers. The 3D structure of this HDAC9(Q9UKV0) was not published. Thus, the aim of this study was to develop and validate the model structure of HDAC9 (Q9UKV0) using bioinformatics tools.

Methods: The Physiochemical characterization was carried out using Ex PASy Prot Param tool, the Functional characterization using Cysteine Recognition Server and HMMTOP Server and Molecular Modeling using I-TASSER. Model Refinement, Validation and verification are carried out using SPDBV, RAMPAGE Server and ERRAT Server respectively.

Result and Conclusion: This3D model of HDAC9 now can be further used to target drug discovery studies related to HDAC9 neurological disorders and a variety of cancers.

Keywords: Histone deacetylase, HDAC9, 3D modelling of HDAC9, I-TASSER, RAMPAGE Server, ERRAT Server.

INTRODUCTION

Histones are the most abundant proteins associated with the eukaryotic DNA. Eukaryotic cells contain histones-H1, H2A, H2B, H3 and H4. Histones H2A, H2B, H3 and H4 are the core histones and form the protein core around which the nucleosomal DNA is wrapped around while H1 is the linker histone. When histones are isolated from cells, their N-terminal tails are modified with small molecules. Lysines are acetylated or methylated while serines are phosphorylated. Histone modifications are mediated by specific enzymes namely, Histone acetyl transferases (HATs) and histone deacetylases (HDACs). HATs are responsible for acetylation of the lysines while HDACsremoves these modifications. [1] Due to their critical role in regulation of chromatin structure and gene expression, HDACs are considered as a major candidate for drug targets [2, 3]. HDAC inhibitors are specific toward tumor cells. That is the reason why they are used as anticancer drugs [4].

Histone deacetylases are comprised of a family of 18 genes. These are furthered divided into 4 classes, namely-class I, class II, class III and class IV, based on their homology to their yeast ortho logs [5, 6]. Class I HDACs are closely related to yeast RPD3 and comprise HDAC1, HDAC2, HDAC3 and HDAC8. Class II HDACs are related to yeast HDA1 and are subdivided into subclass IIa (HDAC4, HDAC5, HDAC7 and HDAC9) and subclass IIb (HDAC6 and HDAC10). Class III HDACs consist of seven sirtuins [7], which require the *NAD+*cofactor for activity. Class IV contains only HDAC11. HDACs from the classical family are dependent on Zn[2+]for deacetylase activity. Inhibitors of Zn[2+]-dependent HDACs are inducers of transformed cell growth arrest and cell death and are identified as inhibitors of HDAC activity [8]. The HDAC9 gene is located on chromosome 7p21 [9]. This region is associated with many neurological disorders and a variety of cancers. Due to alternative splicing, HDAC9 encodes a variety of multiple protein isoforms. The most common HDAC9 isoform contains 1011 amino acids. It has a molecular mass of 111.3 kDa and an isoelectric point of 6.41 [10,11]. HDAC9 plays an important role in heart development [12] and also control the fate of regulatory T-cells[13, 14]. Currently there is no 3D structure available for HDAC9 (Q9UKV0). Therefore, the aim of this study was to build a HDAC9 (Q9UKV0) 3D model using bioinformatics tools. This HDAC model then can be further used for docking studies. These studies can provide valuable information regarding the binding sites of receptor which are very crucial elements for ligand binding.

MATERIALS AND METHODS

The 3D structure of HDAC9 (Q9UKV0) was unavailable in Uni Prot KB Databank. Its FASTA sequence was retrieved from Uni Prot KB. This sequence was then subjected to various bioinformatics tools to predict its physiochemical and functional characterization. The best model was selected based on the confidence score. The refined and validated model can further be used for various drug discovery studies.

Sequence retrieval and physiochemical characterization

The query sequence of HDAC9 with the accession id Q9UKV0 was retrieved from UniProtKB. Table1 shows the HDAC9 query sequence having 1011 residues. The Physiochemical characterization of Q9UKV0 was determined using ExPASyProtParam tool [15]. Table 2 shows the results of Physiochemical characterization of HDAC9 using ExPASy's Prot Paramtool.

Functional characterization

The Functional Characterization of Q9UKV0 was determined using Cysteine Recognition Server [16]. Table 3 shows the result of Functional characterization of HDAC9 using CYS_REC Server. Table 4 shows the Amino acid Composition of HDAC9. Transmembrane region prediction was carried out using HMMTOP Server [17]. Table 5 shows the Trans Membrane Region of HDAC9 as predicted by HMMTOP Server.

Model building and model refinement

The three dimensional structure of HDAC9 was modeled using I-TASSER model workspace [18-20]. Table 6 shows the Top 10 templates used by I-TASSER to build the model. These templates were selected using a meta-threading approach. Table 7 shows the result of I-TASSER modeling score. The best among the resultant modeled structures was then selected depending on confidence score. Fig. 1 shows the I-TASSER 3D Modeled structure of HDAC9 using UCSF Chimera [21].

I-TASSER server is an on-line platform for protein structure and function predictions. 3D models are built based on multiple-threading alignments by LOMETS and Iterative template fragment assembly simulations; function insights are derived by matching the 3D models with BioLiP protein function database. Fig. 2 shows the energy minimization of the modeled structure using Swiss-Pdb Viewer [22].

Table 1: It shows the HDAC9 query sequence having 1011 residues.

>sp|Q9UKV0|HDAC9_HUMAN Histone deacetylase 9 OS=Homo sapiens GN=HDAC9 PE=1 SV=2
MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMMPVVDPVVREKQLQQELLLIQQQQQIQKQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKL
EQQRQEQEVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLWYTAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKD
DFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPTPHAEQMVSQQRILIHEDSMNL
LSLYTSPSLPNITLGLPAVPSQLNASNSLKEKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQHLLLKEQMRQQKLLVAGGVPLHPQSPLAT
KERISPGIRGTHKLPRHRPLNRTQSAPLPQSTLAQLVIQQQHQQFLEKQKQYQQQIHMNKLLSKSIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSA
CVDDTLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQPFLEPTHTRALSVRQAPLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIAYD
PLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGVDSDTIWNEL
HSSGAARMAVGCVIELASKVASGELKNGFAVVRPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAFYADPSILYISLHRYDEGNFFPG
SGAPNEVGTGLGEGYNINIAWTGGLDPPMGDVEYLEAFRTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHLTKQLMTLADGRVVLALEGGHDL
TAICDASEACVNALLGNELEPLAEDILHQSPNMNAVISLQKIIEIQSMSLKFS

Table 2: It shows the results of Physiochemical Characterization of HDAC9 using Ex PASy's ProtParamtool

Length	Mol. wt.	-pI	+R	-R	Extinction coefficient	Instability Index	Aliphatic Index	GRAVY
1011	111297.0	6.40	103	115	44975	55.41	83.06	-0.524

Table 3: It shows the result of Functional Characterization of HDAC9 using CYS_REC

No of Cys residues	Position	Score
1	353	-5.4
2	534	-9.0
3	646	-19.5
4	648	-23.0
5	677	-23.4
6	731	-17.6
7	757	-26.2
8	793	-23.1
9	932	-25.1
10	962	-21.3
11	968	-7.4

11 cysteins are found in positions	The most probable pattern of pairs
353 534 646 648 677 731 757 793 932 962 968	534-968, 646-731

Table 4: It shows the Amino acid composition of HDAC9

Name of amino acid	No. of amino acid	Percentage of amino acid
Ala (A)	67	6.6%
Arg (R)	47	4.6%
Asn (N)	32	3.2%
Asp (D)	43	4.3%
Cys (C)	11	1.1%
Gln (Q)	85	8.4%
Glu (E)	72	7.1%
Gly (G)	68	6.7%
His (H)	41	4.1%
Ile (I)	40	4.0%
Leu (L)	115	11.4%
Lys (K)	56	5.5%
Met (M)	23	2.3%
Phe (F)	22	2.2%
Pro (P)	70	6.9%
Ser (S)	94	9.3%
Thr (T)	48	4.7%
Trp (W)	4	0.4%
Tyr (Y)	15	1.5%
Val (V)	58	5.7%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

Table 5: It shows the Trans membrane Region of HDAC9 as predictedby HMMTOP

Protein	Length	N-terminus	Number of transmembrane helices	Transmembrane helices
HDAC9	1011	IN	1	789-805

Model validation and verification

The refined model was validated by RAMPAGE Server by verifying the parameter of Ramachandran plot quality [23]. Fig. 3 shows the Ramachandran plot of the modeled HDAC9. The summary of the model building and model quality assessment are as shown in Table 8. Verification of the refined model was done using ERRAT [24]. ERRAT is a protein structure verification algorithm that is especially well-suited for evaluating the progress of crystallographic model building and refinement. The program works by analyzing the statistics of non-bonded interactions between different atom types. A single output plot is produced that gives the value of the error function vs. Position of a 9-residue sliding window. By comparison with statistics from highly refined structures, the error values have been calibrated to give confidence limits. Fig. 4 shows the results of ERRAT showing the error value.

RESULTS AND DISCUSSION

Physiochemical characterization

The Physiochemical Characterization was carried out using ExPA Sy Prot Paramtool. The computed isoelectric pointcan be useful for developing the buffer system which can be used for the purification using the isoelectric focusing method. Extinction coefficient values for protein at 280 nm was found to be44975 M^{-1} cm^{-1}, this indicates the presence of higher concentration of Tyr and Trp. Instability index was found to be 55.41 which is beyond 40. This indicates that the protein is slightly un table. Higher value of aliphatic index shows that the protein is Table for wide range of temperature indicating greater amount of aliphatic to aromatic residues. The very low GRAVY index of protein infers that this protein could result in a better interaction with water.

Functional characterization

The CYS_REC result shows the presence of two disulphide bond. This indicates that the stability of protein might be increased due to this along with non-covalent interactions. Table 4 shows the Amino acid composition of HDAC9. Higher values of Lys, Leu, Serine indicates that the amino acids have a high chance of forming the helix and alpha helixes are dominant in these proteins. Trans membrane region predation was carried out using HMMTOP Server.

Model building

The three dimensional structure of HDAC9 was modeled using I-TASSER model workspace using a meta-threading approach. The best among the resultant modeled structures was selected depending on the confidence score.

The top ten templates used for building the model using I-TASSER were 2vqjA, 2vqjA, 3c10A, 2pqpA, 2pqpA, 2vqjA, 2pqpA, 3c10A, 2vqjA and 2nvrA (Table 6).

Table 6: It shows the Top 10 templates used by I-TASSER to build the model

Rank	PDB Hit	Iden1	Iden2	Cov.	Norm. Z-score
1	2vqjA	0.74	0.28	0.38	2.35
2	2vqjA	0.73	0.28	0.38	4.82
3	3c10A	0.69	0.26	0.38	2.99
4	2pqpA	0.69	0.26	0.38	4.96
5	2pqpA	0.69	0.26	0.38	3.71
6	2vqjA	0.74	0.28	0.38	3.95
7	2pqpA	0.69	0.26	0.38	5.31
8	3c10A	0.67	0.00	0.37	7.68
9	2vqjA	0.74	0.28	0.38	2.79
10	2nvrA	0.69	0.26	0.38	3.89

Rank of templates represents the top ten threading templates used by I-TASSER. Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence. Ident2 is the percentage sequence identity of the whole template chains with query sequence. Cov. represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein. Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa. Table 7 shows the result of I-TASSER modeling score.

Table 7: It shows the result of I-TASSER modeling score

Name	C-score	Exp. TM-Score	Exp. RMSD	No. of decoys	Cluster density
Model1:	-2.12	0.46+-0.15	14.3+-3.8	94	0.0223
Model2:	-2.16			80	0.0215
Model3:	-2.27			78	0.0193
Model4:	-2.33			63	0.0181
Model5:	-2.49			60	0.0155

C-score is a confidence score for estimating the quality of the predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of (-5, 2), where a C-score of higher value signifies a model with a high confidence and vice-versa.

TM-score and RMSD are known standards for measuring the structural similarity between two structures which are usually used to measure the accuracy of structure modeling when the native structure is known. A TM-score >0.5 indicates a model of the correct topology and a TM-score<0.17 means a random similarity. This cutoff does not depend on the protein length.

Here we only report the quality prediction (TM-score and RMSD) for the first model, because it was found that the correlation between C-score and TM-score is weak for lower rank models. However, the C-score of all models is listed just for a reference. The first model was

found to be the best out of the others modelled by I-TASSER. It has a C-score of -2.12 which is the highest among the rest. A higher C-score value signifies a model with a high confidence. The TM-score is 0.31- 0.61 which indicates a correct topology. Fig. 1shows the modeled structure of HDAC9 using UCSF Chimera.
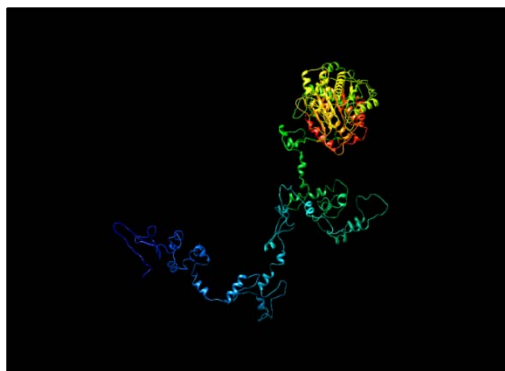

Fig. 1: It shows the I-TASSER 3-D Modeled structure of HDAC9 using UCSF Chimera

Other softwares which are available for model building include Swiss-Pdb Viewer, PHYRE2, M4T server, ModWeb, HMM Modellor, RaptorX etc. These softwares were not used to build the model for HDAC9. This was because the total coverage for these servers and softwares was very low (<50%) because of which the authentication of the modelled structure was low. Due to this limitation, I-TASSER was used to model the structure of HDAC9 with a higher confidence score. The total coverage was only 34% in the case of PHYRE2. This means that only 342 residues of the 1011 residues were modeled. While in the case of I-TASSER the confidence score was -2.12 which is in the range of (-5, 2).

Model refinement

Energy Minimization of the structure was done using Swiss-Pdb Viewer (Fig.2). Computations were done in vacuo with the GROMOS96 43B1 parameters set, without reaction field. For more information about GROMO96, refer to: W. F. van Gunsteren et al (1996) in Biomolecular simulation: the GROMOS96 manual and user guide; (http: //iqc. ethz. ch/gromos). HDAC9 has torsion of 6111.165, electrostatic energy -25124.33 KJ/mol and total energy -20914.396 KJ/mol.

Model validation and verification

The refined model was validated by RAMPAGE Server by verifying the parameter of Ramachandran plot quality (Fig.3). Verification of the refined model was carried out by ERRAT (Fig.4).

Fig. 2: It shows the Energy minimization of the modeled structure using Swiss-PdbViewer

Table 8: It shows the Plot statistics of the modeled HDAC9

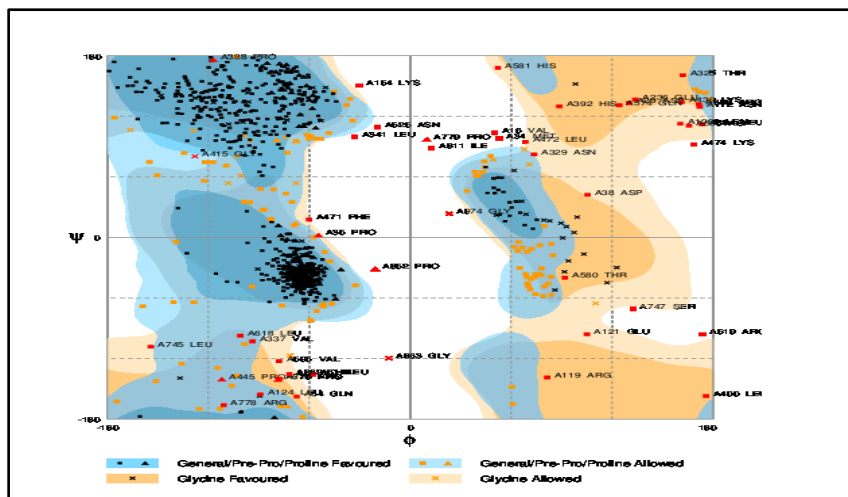| Plot Analysis	
Number of residues in favored region (~98.0% expected)	841 (83.3%)
Number of residues in allowed region (~2.0% expected)	120 (11.9%)
Number of residues in outlier region	48 (4.8%)

Fig. 3: It shows the Ramchandran plot of the modeled HDAC9

ERRAT is a program for verifying protein structures determined by crystallography. The error values are plotted as a function of the position of a sliding 9-residue window. The error function is based on the statistics of non-bonded atom-atom interactions in the reported structure as compared to that of a database of reliable high-resolution structures. The overall quality factor was found to be52.840. This model evaluation method produces a model with good resolution.
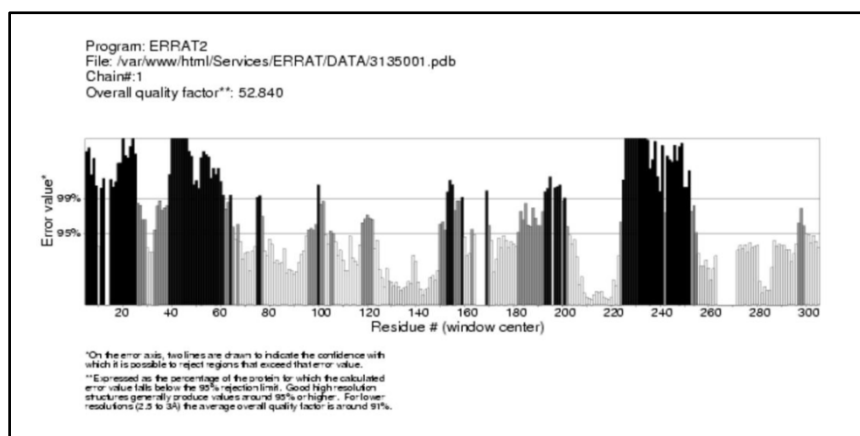

Fig. 4: It shows the result of ERRAT showing the error value

CONCLUSION

HDACs play a major role in regulation of chromatin structure and gene expression. Till now, no 3D model structure was available for HDAC9 but this study has successfully generated a 3D structure for the query sequence using various bioinformatics tools. This 3D model in the future can be used to carry out in vitro and in vivo study of these deacetylases. Various HDAC inhibitors can be developed to combat cancer and other various deadly diseases.

ACKNOWLEDGEMENT

CONFLICT OF INTEREST

None

REFERENCES

1. Cress WD, Seto E. Histone deacetylases, transcriptional control, and cancer. J Cell Physiol 2000;184:1-16.

2. Hagelkruys A, Sawicka A, Rennmayr M, Seiser C. The biology of HDAC in cancer: the nuclear and epigenetic components. Handb Exp Pharmacol 2011;206:13-37.

3. Ren J, Zhang J, Cai H, Li Y, Zhang Y, Zhang X. HDAC as a therapeutic target for treatment of endometrial cancers. Curr Pharm Des 2013;20:1847-56.

4. Marks PA, Richon VM, Rifkind RA. Histone deacetylase inhibitors: inducers of differentiation or apoptosis of transformed cells. J Natl Cancer Inst 2000;92:1210-6.

5. deRuijter AJM, van Gennip AH, Caron HN, Kemp S, van Kuilenburg AB. Histone deacetylases (HDACs): characterization of the classical HDAC family. Biochem J 2003;370:737-49.

6. Gregoretti IV, Lee YM, Goodson HV. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. J Mol Bio 2004;338:17-31.

7. Witt O, Deubzer HE, Milde T, Oehme I. HDAC family: what are the cancer relevant targets? Cancer Lett 2009;277:8-21.

8. Smith KT, Workman JL. Histone deacetylase inhibitors: anticancer compounds. Int J Biochem Cell Biol 2009;41:21-5.

9. Mahlknecht U, Schnittger S, Will J, Cicek N, Hoelzer D. Chromosomal organization and localization of the human histone deacetylase 9 gene (HDAC9). Biochem Biophys Res Commun 2002;293:182-91.

10. Zhou X, Marks PA, Rifkind RA, Richon VM. Cloning and characterization of a histone deacetylase, HDAC9. Proc Natl Acad Sci USA 2001;98:10572-7.

11. Petrie K, Guidez F, Howell L, Healy L, Waxman S, Greaves M, *et al*. The histone deacetylase 9 gene encodes multiple protein isoforms. J Biol Chem 2003;278:16059-72.

12. Zhang CL, McKinsey TA, Chang S, Antos CL, Hill JA, Olson EN. Class II Histone Deacetylases act as signal-responsive repressors of cardiac hypertrophy. Cell 2002;110:479-88.

13. Tao R, de Zoeten EF, Ozkaynak E, Chen C, Wang L, Porrett PM. Deacetylase inhibition promotes the generation and function of regulatory T cells. Nat Med 2007;13:1299-307.

14. de Zoeten EF, Wang L, Sai H, Dillmann WH, Hancock WW. Inhibition of HDAC9 increases T regulatory cell function and prevents colitis in mice. Gastroenterol 2010;138:583-94.

15. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD. Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook; 2005. p. 571-607.

16. http://www.softberry.com/berry.phtml?topic=cys_rec&group =programs&subgroup=propt. (Last accessed on 16th July 2014, 21:25 hrs).

17. Tusnady GE, Simon I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J Mol Biol 1998;283:489-506.

18. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinf 2008;9:40.

19. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 2010;5:725-38.

20. Roy A, Yang J, Zhang Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res 2012;40:471-7.

21. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC. UCSF Chimera-a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605-12.

22. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophor 1997;18:2714-23.

23. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, *et al*. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 2003;50:437-50.

24. Colovos C, Yeates TO. Verification of protein structures: patterns of non-bonded atomic interactions. Protein Sci 1993;2:1511-9.