Available on line www.jocpr.com

Journal of Chemical and Pharmaceutical Research

J. Chem. Pharm. Res., 2010, 2(2): 440-451

ISSN No: 0975-7384

Homology modeling and evaluation of human TEK tyrosine kinase using SWISS-MODEL Workspace

Raj K. Prasad[a*], Rajesh Sharma[a] and G. L. Prajapati[b]

[a*]*School of Pharmacy, Devi Ahilya Vishwavidyalaya, Khandwa Road, Indore, M. P., India*
[b]*Institute of Engineering and Technology, Devi Ahilya Vishwavidyalaya, Indore, M.P. India*

Abstract

TEK is an endothelial cell receptor tyrosine kinase that induces signal transduction pathways involved in cell migration upon angiopoietin-1 stimulation. Homology models of proteins are of great interest for planning and analysis biological experiments. Building homology models requires specialized programs and up to date sequence and structural database. SWISS-MODEL provides several levels of interaction through its World Wide Web interface: in the 'first approach mode' only an amino acid sequence of a protein is submitted to build a 3D model. Template selection, alignment and model building are done completely automated by the server. In the 'alignment mode' the modeling process is based on a user defined target-template alignment. Complex modeling tasks can be handled with the 'project mode' using Deep View (Swiss Pdb-Veiwer), an integrated sequence-to-structure workbench. All human TEK models are sent back via email with a detailed modeling report. The evaluation of model by what if and procheck method from SWISS-MODEL work space accessed freely at http:/swissmodel.expasy.org/workspace/

Keywords: Homology modeling, TEK tyrosine kinase, SWISS-MODEL work space.

Introduction

Three-dimensional (3D) protein structures are of great interest for the rational design of many different types of biological experiments, such as site-directed mutagenesis or structure-based discovery of specific inhibitors. How ever, the number of structurally characterized proteins is small compared with the number of known protein sequences: as of November 2005, more than 33000 experimentally determined protein structures were deposited in the Protein Data Bank [1] While the UniProt protein knowledge database [2,3] held more than 2.3 million sequences. Various computational methods for modeling 3D structure of protein have been

developed to overcome this limitation. Since the number of possible fold in nature appears to be limited and the 3D structure of proteins is better conserved than their sequences. It is often possible to identify a homologous protein with a known structure (template) for a given protein sequence (target).

Homology modeling is routinely used in many applications, such as virtual screening, or rationalizing the effect of sequence variations [4-6]. Although great progress was made in the field of experimental structure solution by X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR), it is still a time-consuming process without guaranteed success. Thus, no structural information is available for the vast majority of protein sequences. Therefore, it is an obvious demand to bridge this 'structure knowledge gap and computational methods for protein structure prediction have gained much interest in recent years.

Angiogenesis [7], the formation of new vessels from the existing vasculature, is a critical process during early development as well as in a number of disease processes. TEK is an endothelium-specific receptor tyrosine kinase involved in both angiogenesis and vasculature maintenance.

The TEK receptor tyrosine kinase is expressed almost exclusively in endothelial cells in mice, rats, and humans. This receptor possesses a unique extra cellular domain containing 2 immunoglobulin-like loops separated by 3 epidermal growth factor-like repeats that are connected to 3 fibronectin type III-like repeats. The ligand for the receptor is angiopoietin-1. Defects in TEK are associated with inherited venous malformations; the TEK signaling pathway appears to be critical for endothelial cell-smooth muscle cell communication in venous morphogenesis.

Materials and Methods

Modeling modes in SWISS-MODEL workspace

Depending on the difficulty of the modelling task, three different types of modelling modes were provided, which differ in the amount of user intervention: automated mode, alignment mode and project mode [8].

Automated mode

Modeling requests were directly computed by the SWISSMODEL server homology modelling pipeline. The 'automated mode' used cases where the target–template similarity was sufficiently high to allow for fully automated modelling. Sequence alignments were sufficiently reliable when target and template share >50% identical residues. 'Automated mode' submissions require only the amino acid sequence or the UniProt accession code of the target protein as input data. The modelling pipeline was automatically selected suitable templates based on a Blast E-value limit, which can be adjusted upon submission. The automated template selection will favors high-resolution template structures with reasonable stereo-chemical properties as assessed by ANOLEA mean force potential and Gromos96 force field energy.

Project mode

The 'project mode' allowed to the submitted a manually optimized modeling request to the SWISS-MODEL server. The starting point for this mode was a Deep View project file. It

contains the superposed template structures, and the alignment between the target and the templates. This mode was giving the user control over a wide range of parameters, e.g. template selection or gap placement in the alignment.

DeepView-Swiss-Pdb Viewer

The program Deep View (Swiss-Pdb-Viewer) was designed to integrate functions for protein structure visualization, analysis and manipulation into a sequence-to-structure workbench with a user-friendly interface. It was publicly available from the ExPASy server (http://www.expasy.org/spdbv). With DeepView one can searched for suitable modeling templates and downloaded the corresponding PDB or ExPDB files directly from the DeepView server. Using the integrated sequence alignment tools and structural superposition algorithms, a target sequence can be mapped onto the modeling templates in one step. Then the initial sequence alignment can be optimized manually while the anticipated changes in the model backbone were reflected in real-time in the displayed structural superposition. The complete project file is then submitted to the SWISS-MODEL server for model building. The resulting protein model can be visualized and analyzed using the integrated tools Fig. 1.

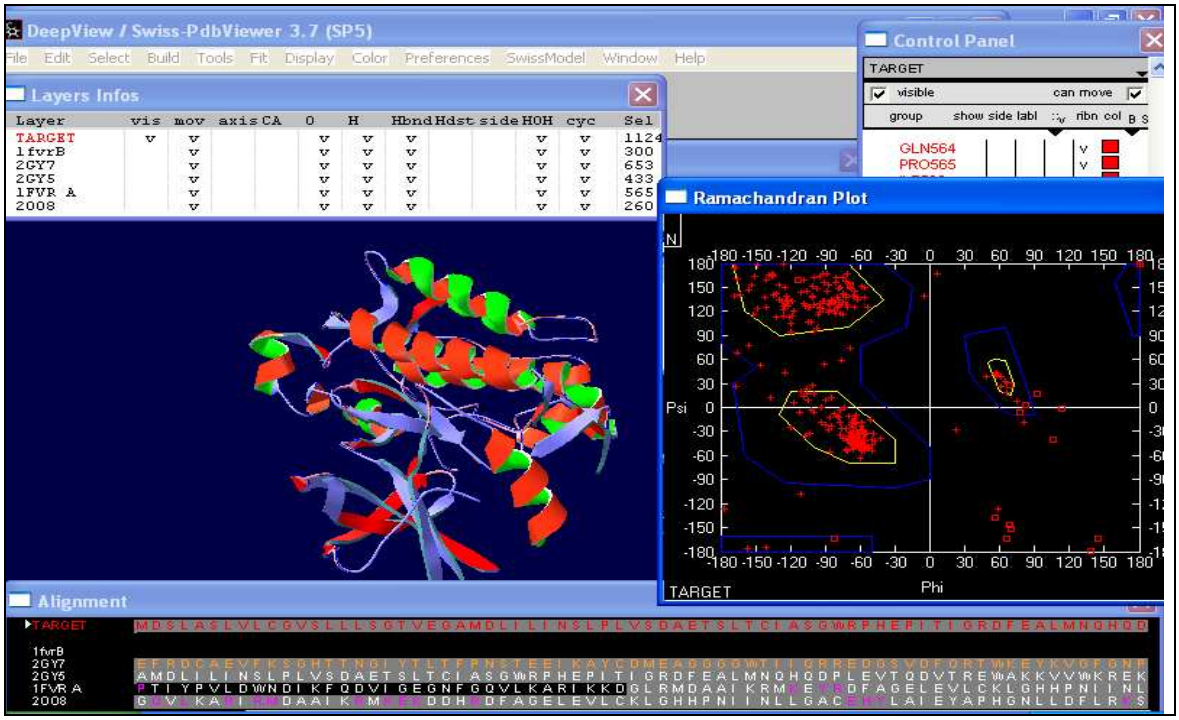


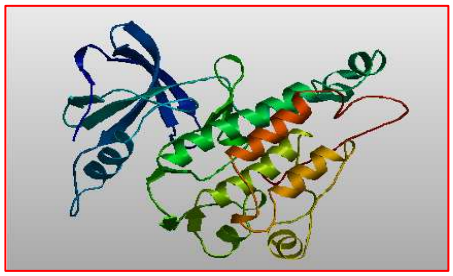
Fig. 1. SWISS-MODEL project displayed in DeepView

Domain assignment, secondary structure and disorder prediction

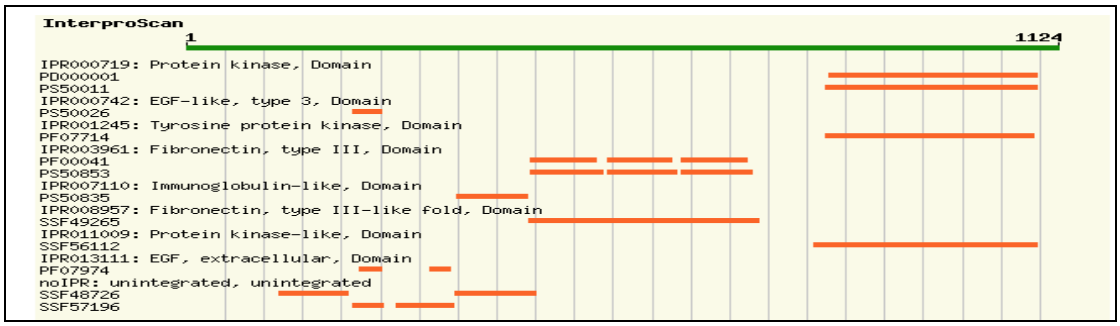
Many proteins are modular and made up of several structurally distinct domains, which often reflect evolutionary relationships and may correspond to units of molecular function. Most soluble proteins adopt well-defined 3D structures. However, some proteins have regions that were natively disordered, unstructured or have flexible regions without permanent regular secondary structure.

It had been suggested that disordered regions may possess biological functions, and could be involved in signaling and regulation processes. The protein disorder prediction program Disopred estimated the propensity of protein sequences to be disordered. The result of

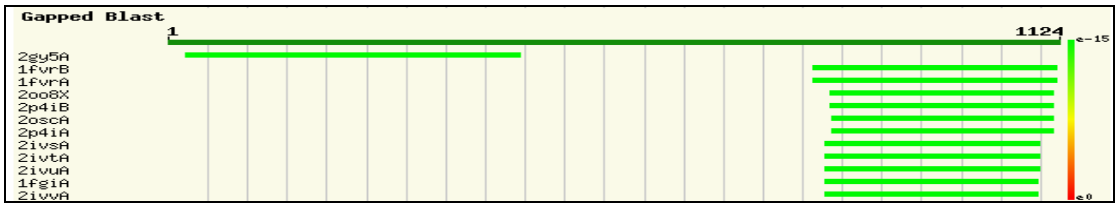
secondary structure, disorder and domain boundary prediction Fig. 2, can aid the selection of modelling templates for specific regions of the target protein.



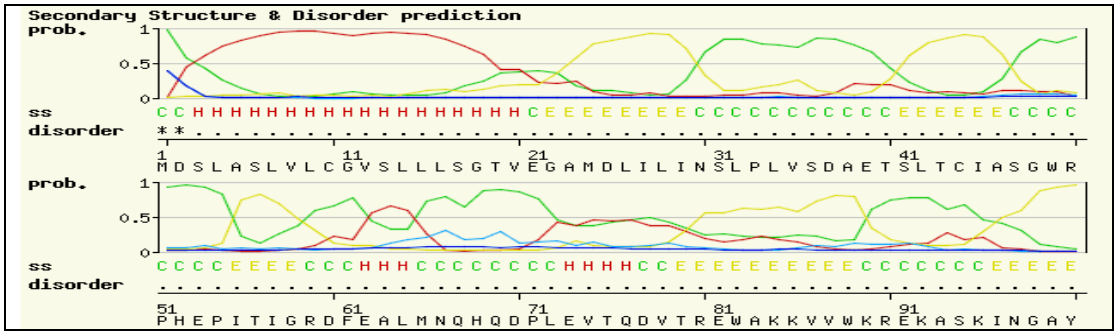
(a)



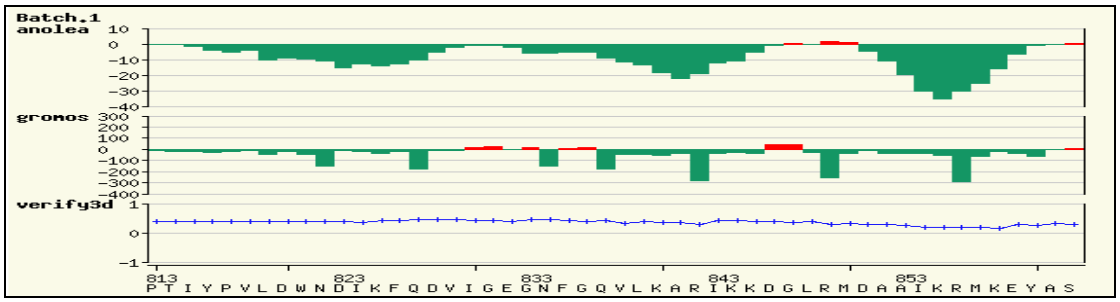
(b)



(c)



(d)



(e)

Fig. 2.Graphical output of SWISS-MODEL workspace representing typical steps of a modeling experiment

(a) Ribbon structure of ten domains modelled for the target protein human TEK tyrosine kinase. (b) IprScan of the target sequence detected ten domains in the target protein. (c) Sequence-based searches of the template library identified twelve segments with suitable template structures. (d) Secondary structure and disorder prediction of the target protein. (e) Anolea mean force potential plot allows for quality assessment of the final models.

Homology Modeling Method

All homology-modeling methods consist of the following four steps: (i) template selection; (ii) target template alignment; (iii) model building; and (iv) evaluation. These steps can be iteratively repeated, until a satisfying model structure is achieved. Several different techniques for model building have been developed [9-11].

Template selection

It is tools for searching the SWISS-MODEL template library for suitable template structures. I was retrieve the human TEK sequence from the NCBI [12] and submitted.

The SWISS-MODEL server template library ExPDB was extracted from the PDB. In order to allow a stable and automated workflow of the server, the PDB coordinate files were splited into individual protein chains and unreliable entries. Additional information useful for template selection is gathered and added to the file header, e.g. probable quaternary structure; quality indicators like empirical force field energy [13] or ANOLEA mean force potential scores [14]. To selected templates for a given protein (human TEK), the sequences of the template structure library are searched [15, 16]. If these templates cover distinct regions of the target sequence, the modeling process will be split into separate independent batches.

Alignment

Up to five template structures per batch were superposed. A structural alignment was generated after removing incompatible templates, i.e. omitting structures with high Ca root mean square deviations to the first template. A local pair-wise alignment of the target sequence to the main template structures was calculated [17], followed by a heuristic step to improve the alignment for modeling purposes. The placement of insertions and deletions is optimized considering the template structure context. In particular, isolated residues in the alignment were moved to the flanks to facilitate the loop building process.

Model building

To generate the core of the model, the backbone atom positions of the template structure were averaged. The templates were thereby weighted by their sequence similarity to the target sequence, while significantly deviating atom positions were excluded. The template coordinates cannot be used to model regions of insertions or deletions in the target-template alignment. To generate those parts, an ensemble of fragments compatible with the neighboring stems was constructed using constraint space programming (CSP). The best loop was selected using a scoring scheme, which accounts for force field energy, steric hindrance and favorable interactions like hydrogen bond formation. There were some no suitable loop can be identified, the flanking residues were included to the rebuilt fragment to allow for more flexibility. In cases where CSP did not give a satisfying solution and for loops above 10 residues, a loop library derived from experimental structures is searched to find compatible loop fragments.

Tools for protein structure and model assessment

The quality of individual models can vary significantly from the average accuracy expected for a given target–template similarity or modelling method. SWISS-MODEL workspace provided the graphical plots of Anolea mean force potential, GROMOS empirical force field energy, Verify3D profile evaluation and Whatcheck and Procheck reports were generated to estimate the quality of protein models and template structures.

Table 1: Sequence Alignment between Target and Template domain

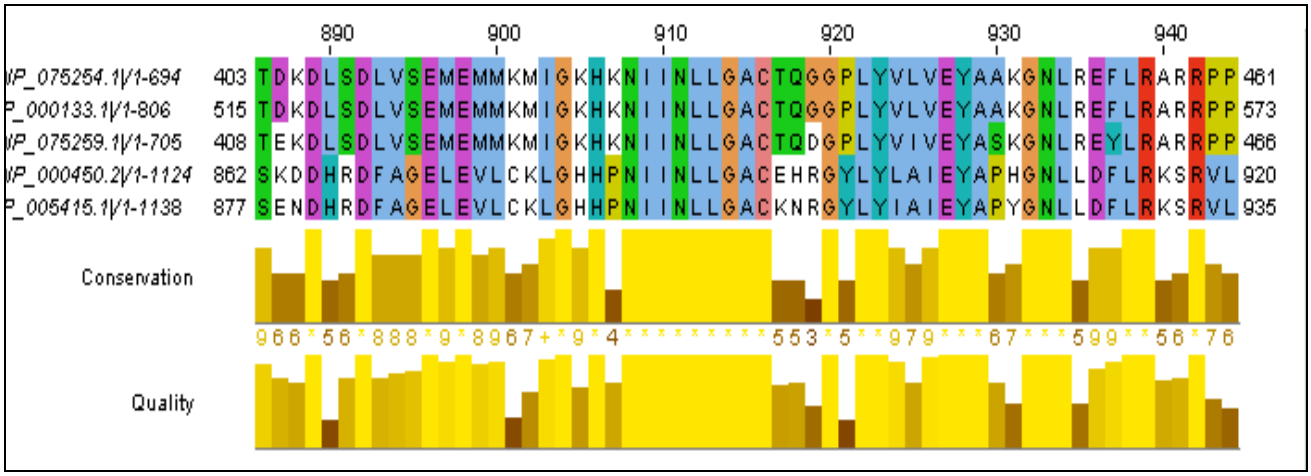
```
TARGET      813  PTIYPVLDWN DIKFQDVIGE GNFGQVLKAR IKKDGLRMDA AIKRMKEYAS
1fvrB            813 ptiypvldwn dikfqdvige gnfgqvlkar ikkdglrmda aikrmkey--

TARGET                                  ssssssss  sssssss  sss   sssss  sssssss
1fvrB                                   ssssssss  sssssss  sss   sssss  sssssss


TARGET      863  KDDHRDFAGE LEVLCKLGHH PNIINLLGAC EHRGYLYLAI EYAPHGNLLD
1fvrB            867 ----rdfage levlcklghh pniinllgac ehrgylylai eyaphgnlld

TARGET                        hhhhh hhhhhh          ssssss ss sssssss s    ssshh
1fvrB                         hhhh hhhhhh           ssssss ss sssssss s    ssshh


TARGET      913  FLRKSRVLET DPAFAIANST ASTLSSQQLL HFAADVARGM DYLSQKQFIH
1fvrB            913 flrksrvlet dpafaianst astlssqqll hfaadvargm dylsqkqfih

TARGET            hhhh   hh     hhhhh ss s    hhhhh hhhhhhhhhh hhhh   sss
1fvrB             hhhh   hh     hhhhhhss s    hhhhh hhhhhhhhhh hhhh   sss


TARGET      963  RDLAARNILV GENYVAKIAD FGLSRGQEVY VKKTMGRLPV RWMAIESLNY
1fvrB            963 rdlaarnilv genyvakiad fglsrgqevy vkk---rlpv rwmaieslny

TARGET                  ssss s   sssss     sss  ss s                  hhhhh
1fvrB                   ssss s   sssss     sss  ss s                  hhhhh


TARGET     1013  SVYTTNSDVW SYGVLLWEIV SLGGTPYCGM TCAELYEKLP QGYRLEKPLN
1fvrB            1013 svyttnsdvw sygvllweiv slggtpycgm tcaelyeklp qgyrlekpln

TARGET            sss hhhhhh hhhhhhhhhh h          hhhhhhhhh          s
1fvrB             sss hhhhhh hhhhhhhhhh h          hhhhhhhhh          s


TARGET     1063  CDDEVYDLMR QCWREKPYER PSFAQILVSL NRMLEERKTY VNTTLYEKFT
1fvrB            1063 cddevydlmr qcwrekpyer psfaqilvsl nrmleerkty vnttlyekft

TARGET            sshhhhhhhh hh          hhhhhhhh hhhh    s sssss
1fvrB             sshhhhhhhh h           hhhhhhhh hhhh    s sssss


TARGET                     1113                                  YAGIDCSAE
1fvrB                      1113                                  yagidcsae-

TARGET
1fvrB
```

Score = 640 bits (1650), Expect = 0.0; Identities = 300/309 (97%), Positives = 300/309 (97%), Gaps = 9/309 (2%)

Table 2: CLUSTAL W (1.83) multiple sequence alignments

```
gi|4503711|ref|NP_000133.1|   GCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMK 530
gi|13186255|ref|NP_075259.1   GCFGQVVMAEAVGIDKDKPKEAVTVAVKMLKDDATEKDLSDLVSEMEMMK 423
gi|88758596|ref|NP_000450.2   DVIGEGNFGQVLKARIKKDGLRMDAAIKRMKEYASKDDHRDFAGELEVLC 877
gi|4885631|ref|NP_005415.1    DLIGEGNFGQVIRAMIKKDGLKMNAAIKMLKEYASENDHRDFAGELEVLC 892
                               . :*:  :.:.:    .:    : .*:* :*: *:..* *:..*:*::

gi|4503711|ref|NP_000133.1|   MIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFD 580
gi|13186255|ref|NP_075259.1|  MIGKHKNIINLLGACTQDGPLYVIVEYASKGNLREYLRARRPPGMEYSYD 473
gi|88758596|ref|NP_000450.2|  KLGHHPNIINLLGACEHRGYLYLAIEYAPHGNLLDFLRKSRVLETDPAFA 927
gi|4885631|ref|NP_005415.1|   KLGHHPNIINLLGACKNRGYLYIAIEYAPYGNLLDFLRKSRVLETDPAFA 942
                               :*:* ********* : * **: :***. *** ::**  *    : ::

gi|4503711|ref|NP_000133.1|   MKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSF 680
gi|13186255|ref|NP_075259.1|  MKIADFGLARDINNIDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSF 573
gi|88758596|ref|NP_000450.2|  AKIADFGLSRGQE---VYVKKTMGRLPVRWMAIESLNYSVYTTNSDVWSY 1024
gi|4885631|ref|NP_005415.1|   SKIADFGLSRGEE---VYVKKTMGRLPVRWMAIESLNYSVYTTKSDVWSF 1039
                               *******:*. .    * *.* *****:*** *:*   *** :*****:

gi|4503711|ref|NP_000133.1|   GVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMREC 730
gi|13186255|ref|NP_075259.1|  GVLMWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTNELYMMMRDC 623
gi|88758596|ref|NP_000450.2|  GVLLWEIVSLGGTPYCGMTCAELYEKLPQGYRLEKPLNCDDEVYDLMRQC 1074
gi|4885631|ref|NP_005415.1|   GVLLWEIVSLGGTPYCGMTCAELYEKLPQGYRMEQPRNCDDEVYELMRQC 1089
                              ***:***.:***:** *:.  **:: * :*:*:::* ** .::* :**:*

gi|4503711|ref|NP_000133.1|   WHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSS 780
gi|13186255|ref|NP_075259.1|  WHAVPSQRPTFKQLVEDLDRILTLTTNEEYLDLSQPLEQYSPSYPDTRSS 673
gi|88758596|ref|NP_000450.2|  WREKPYERPSFAQILVSLNRMLEERKTYVNTTL---YEKFTYAGIDCSAE 1121
gi|4885631|ref|NP_005415.1|   WPDPPYERPPFAQIALQLGRMLEARKAYVNMSL---FENFTYAGIDATAE 1136
                              *:  * :**.* *:  .*.*:*   .       *    *::: .  *  :.

gi|4503711|ref|NP_000133.1|   SSSGDDSVFAHD------LLPPAPPSSGGSRT 806
gi|13186255|ref|NP_075259.1|  CSSGDDSVFSPDPMPYEPCLPQYPHINGSVKT 705
gi|88758596|ref|NP_000450.2|  EAA---------------------------- 1124
gi|4885631|ref|NP_005415.1|   EA---------------------------- 1138
```

Table 3: Predict the conservation and quality in multiple Sequence Alignment of deferent sequence of family TEK tyrosine kinase by Jalveiw program



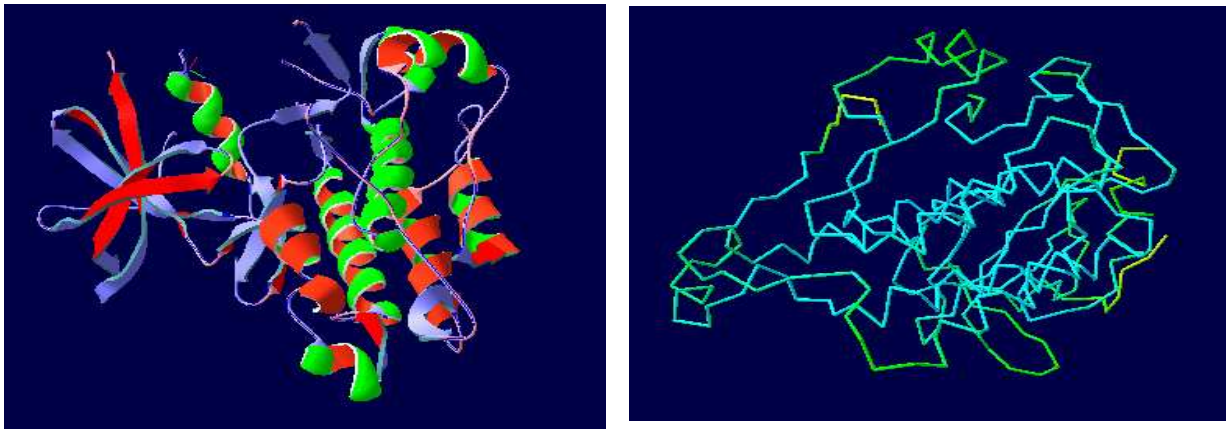


Fig. 3. Secondary structure of human TIE-2 kinase Domain (template)

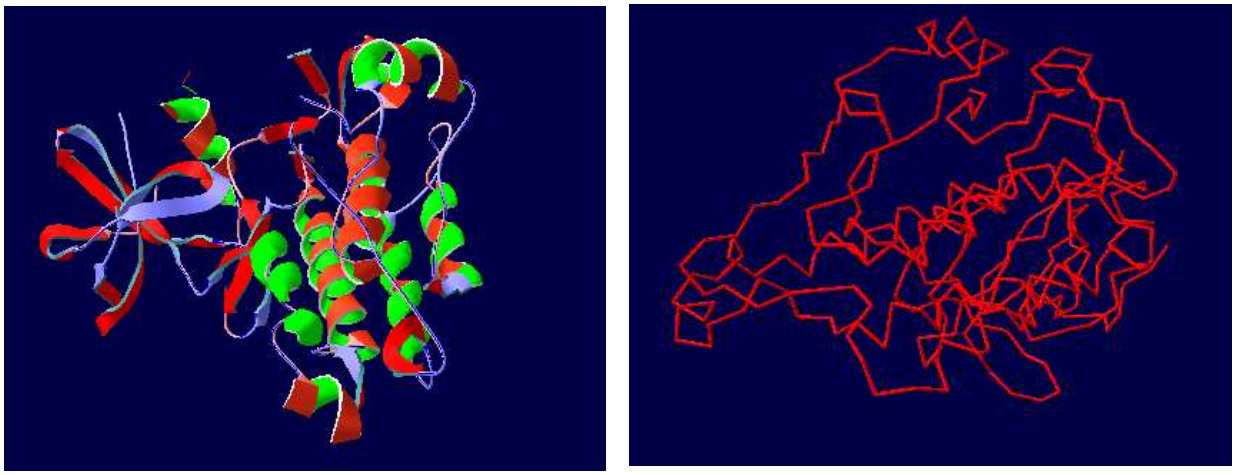


Fig. 4. Secondary structure of homology modeled Protein (target)

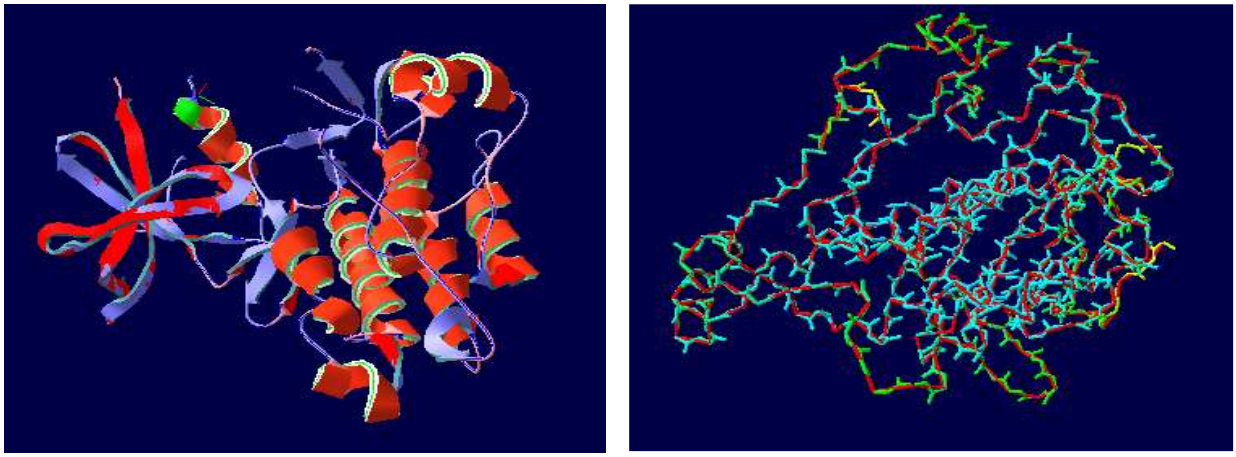
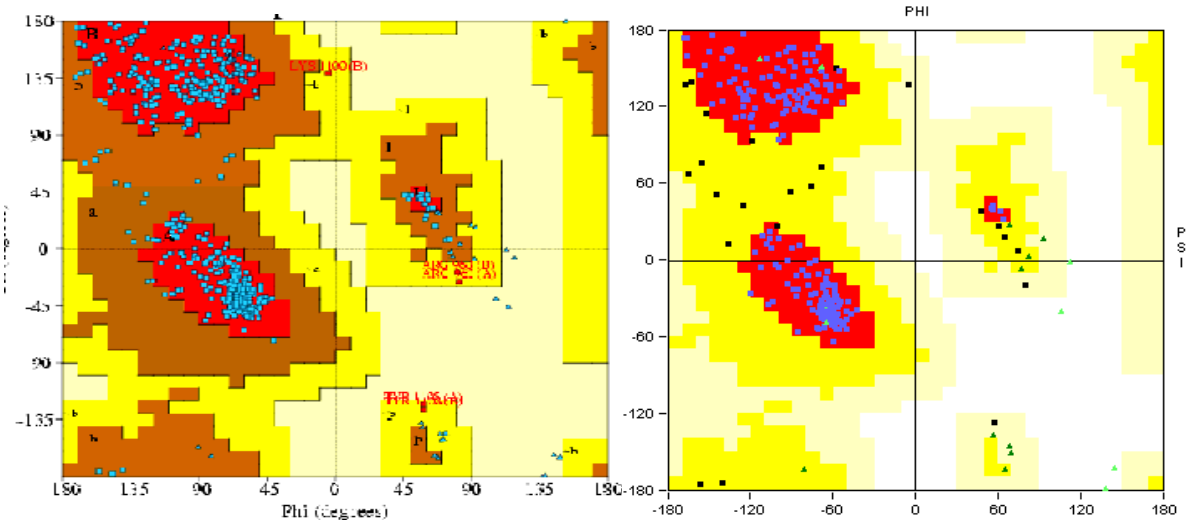


Fig. 5. Superposition structure of target and template protein



(a) **(b)**

Fig.6. Comparison between the Ramchandran plot of TIE-2 protein (a) and TEK protein (b)

Table 4: Summary of WHAT IF and PROCHECK analysis of TEK

Report was created by WHAT IF version 19970813-1517
This is a summary of the quality of the structure as compared with current reliable structures. Further it is useful for modeling calculations. The second part of the table mostly gives an impression of how well the model conforms to common refinement constraint values. The first part of the table shows a number of constraint-independent quality indicators.

WHAT IF	
Structure Z-scores (positive is better than average)	
1st generation packing quality	: -1.407
2nd generation packing quality	: -1.289
Ramchandran plot appearance	: 0.234
Chi-1/chi-2 rotamer normality	: 1.364
Backbone conformation	: 0.719
RMS Z-scores (should be close to 1.0)	
Bond lengths	: 0.779
Bond angles	: 1.040
Omega angle restraints	: 0.754
Side chain planarity	: 1.423
Improper dihedral distribution	: 1.451
Inside/Outside distribution	: 1.028

Table 6 continues…

PROCHECK statistics

Ramchandran plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	484	91.5%
Additional allowed regions	[a,b,l,p]	40	7.6%
Generously allowed regions	[~a,~b,~l,~p]	5	0.9%
Disallowed regions	[XX]	0	0.0%
Non-glycine and non-proline residues		529	100.0%
End-residues (excl. Gly and Pro)		10	
Glycine residues		38	
Proline residues		22	
Total number of residues		599	

Based on an analysis the resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A,B,L].

G-Factor Parameter	Score	Average Score
Dihedral angles:-		
Phi-psi distribution	-0.10	
Chi1-chi2 distribution	0.18	
Chi1 only	0.21	
Chi3 & chi4	0.48	
Omega	0.58	
	0.25	
Main Chain covalent forces:-		
Main-chain bond lengths		0.67
Main-chain bond angles		0.47
		0.55
OVERALL AVERAGE		0.37

G-factors provide a measure of how **unusual**, or out-of-the-ordinary, a property is.
Values below -0.5* - unusual , Values below **-1.0**** - highly unusual

Results and Discussion

We described how the different modelling modes and tools available from the SWISS-MODEL workspace were applied to identify suitable templates and to build homology models for biologically relevant human proteins: TEK tyrosine kinase.

Modeling of human TEK tyrosine kinase

The human TEK tyrosine kinase was modeled by using various sequence Alignments such as BLAST and cluster W program searches, to identify the suitable templates, leads to several highly significant matches. these matches span the second half of the protein (Starting at about residues 813), which correspond to the TIE-2 kinase domain as assigned by scanning the InterPro database of Protein families and domains.

All detected templates belong to the kinase family, in particular the TIE-2 kinase domain share the highest sequence similarity with TEK Tyrosine kinase .For kinase several Experimental structures are known. The Blast Alignment between the sequence of TEK tyrosine kinase and TIE-2 kinase. The two sequences share ~97% identify and contains only 9 Gaps (2%).Table 1 and multiple sequence alignment by cluster W and Jalview program is reported in Table 2 and 3.

Based on these observations, the TEK tyrosine kinase can be easily modeled using the automated mode of SWISS-MODEL pipeline. the modeling pipeline select the structure of the human TIE-2 kinase [PDB:1fvrB] as a template , although other structures of the same protein have been deposited in the PDB, 1fvrB selected by modeling pipeline because it has the highest resolution (**2.2Å**) among the suitable template.

Exhaustive information about the individual templates is available directly from the template selection output as link to the SWISS-MODEL template library and external resource, MSD, PDB, PDBsum, SCOP and CATH. As expected from the high target-template similarity the quality of models (based on the structures 1fvrB) is very good overall ANOLEA, GROMOS and verify 3D quality assessment do not detect major problems in the models. Fig. 2(e).

Visual inspection and Evaluation of the model and template structure using Deep-view show good sequence conservation. Secondary structure of both TEK and TIE-2 were predicted and shown in Fig. 3 & 4. Superposition structure of target and template protein reported at Fig. 5. The modeled TEK tyrosine kinase was validated by Ramchandran plot and What If Server. Ramchandran plot of target and template protein reported at Fig. 6. The What If and PROCHECK result is shown in Table 4.

Acknowledgements

We are grateful to Head, School of Pharmacy for providing facilities to carry out proposed work. We are deeply indebted to Manuel C. Peitsch (Novartis AG, Basel) and to Nicolas Guex (GSK, Raleigh, NC, USA) for their pioneering work on large-scale protein structure modelling. One of the authors Mr. Raj K. Prasad is grateful to All India council of Technical education (AICTE) for providing junior research fellowship.

References

[1] J. Westbrook, Z. Feng, L. Chen, H. Yang, H. M. Berman, *Nucleic Acids Res.* **2003**, 31, 489–491
[2] A. Bairoch, *Nucleic acid Res. (Database Issu)* **2005**, 33, D154-D159
[3] A.Tramontano, R. Leplae, V. Morea, *Proteins*, **2001**, 5, 22–38.
[4] M. A. Marti-Renom, *Biophysics. Biomol.strucure*, **2000**, *29, 291*-325.
[5] J. Kopp, T. Schwede, *Pharmacogenomic*, **2004**, *5*, 405-416.
[6] A. Hillisch, *drug discovery today*, **2004**, *9*. 659-669.
[7] L. M. Shewchuk , A. M. Hassell , B. Ellis, W. D. Holmes , R. Davis , E. L. Horne , S. H. Kadwell , D. D. McKee , J.T. Moore, *Structure of the Tie2 RTK domain*, **2000**, 8,1105-13.
[8] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, *Bioinformatics*, **2006**, 22 (2), 195-201
[9] A Sali, T.L. Blundell, *J. Mol. Biology*. **1993**, 234, 779–815.
[10] M. J. Sutcliffe, I. Haneef, D. Carney, T. L. Blundell, *Protein Eng.* **1987**, 1, 377–384.
[11] http://www.ncbi.nlm.nih.gov

[12] W. F. V. Gunsteren, S. R. Billeter, A. Eising, P. H. Hu¨nenberger, P. Kru¨ger, A. E. Mark, W. R. P. Scott, I. G.Tironi, *Biomolecular Simulations: The GROMOS96 Manual and User Guide.* **1996.**
[13] F. Melo, E. Feytmans. *J. Mol. Biol.* **1998**, 277, 1141–1152.
[14] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, 215, 403–410.
[15] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997**, 25, 3389–3402.
[16] X. Huang, W. Miller, *Adv. Appl. Math,***1991**, 12, 337–357.
[17] S.C. Lovell, J. M. Word, J. S. Richardson, D. C. Richardson, *Proteins*, **2000**, 40, 389–408.