Comparative study of three benchmark protein structure prediction techniques for the prediction of epidermal growth factor receptor (EGFR) family protein structures

Apoorv Tiwari and Saurav B Saha*
Department of Computational Biology & Bioinformatics, SHIATS, Allahabad-211007, India
*Corresponding Author: saurav.saha@shiats.edu.in

Background

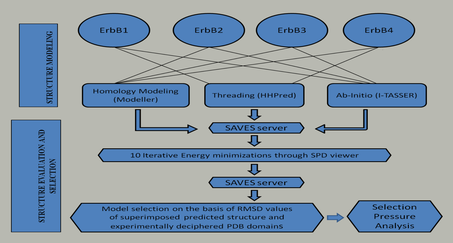
Epidermal growth factor receptor (EGFR) is among the most exploited and important transmembrane receptor subfamily involved in fate decision of very canonical biological pathways viz. RAS, JAK/STAT, PI3 etc. EGFR family consist of four members; EGFR (ErbB-1), HER2/c-neu (ErbB-2), Her 3 (ErbB-3) and Her 4 (ErbB-4) with an amino acid length of 1210, 1255, 1342 and 1308 residues. Reports have shown that abnormalities like mutation in check mechanism of these receptors are associated to tumor development and henceforth cancer which makes it one of the most lucrative cancer targets with many monoclonal antibodies (mAB) like Cetuximab, ABX-EGF , Nimotuzumab and Tyrosine kinase inhibitors (TKIs) like Gefitinib, Erlotinib, Canertinib targeting it available in market. However due to lack of complete protein structure, few remarkable question do exists which prevents optimal exploitation of this magic target such as:

1.How mutant EGFR kinase activation occurs?

2. How do mutations affect binding of inhibitors?

The present work was carried out to address these questions by determining complete protein structure of Human EGFR family members through conventional computational approaches viz. Homology modeling, threading and Ab-initio.

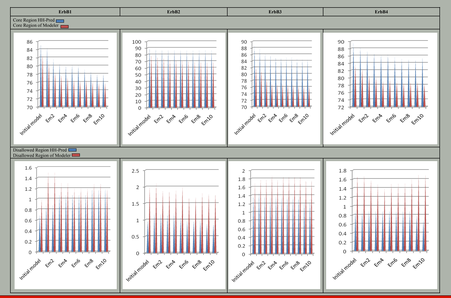
Methodology



Results

Past recent decades has witnessed advent of computational biology in protein structural prediction. The goal of current work is to analyze conventional *in silico* prediction approaches viz. the most exhaustive homology modeling, where all the combinations of templates with e-value 0 to the target were taken into consideration, threading and ab-initio to decipher complete protein structure of Human EGFR family members. e is Despite the fact that homology modeling with its added advantage of longer template knowledge is considered the best, surprisingly threading showed a better result in our case (Table 1). Repeated energy minimization showed a similar pattern being threading to be best followed by homology modeling and ab-initio (Table 2).

Table 1



Reference

1. **Sali A. and Blundell T. L. (1993).** Comparative protein modeling by satisfaction of spatial restraints. *Journal of Molecular Biology,* **234:**779-815

2. **Jorissen R. N. et al., (2003).** Epidermal growth factor receptor: mechanisms of activation and signaling, *Experimental Cell Research* **284**:31–53.

3. **Soding J. et al 2005.** The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research,* **33**:244-248.

4. **Zhang Y. 2008**. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**:40

Table 2

S. No.	Receptor	Template	Method	RMSD	Structure
1	ERBB1	3NJP	Homology Modeling	42.838	
2	ERBB1	3NJP	Threading	19.590	
3	ERBB1	3NJP	Ab-initio	16.902	
4	ERBB2	3N85	Homology Modeling	29.651	
5	ERBB2	3N85	Threading	1.108	
6	ERBB2	3N85	Ab-initio	1.543	
7	ERBB3	3P11	Homology Modeling	116.706	
8	ERBB3	3P11	Threading	2.085	
9	ERBB3	3P11	Ab-initio	3.975	
10	ERBB4	3P11	Homology Modeling	53.416	
11	ERBB4	3P11	Threading	2.284	
12	ERBB4	3P11	Ab-initio	1.452	

Furthermore, after prediction and selection pressure analysis of active sites of ERBB1, residues 720S, 745K, 766M, 775C, 776R, 788L, 791Q, 792L, 793M, 796G, 997F, in ErbB2, residue 726P,731V, 734P, 751E, 753L, 774C, 799R, 801R, 802L, 804S, 849F, 850G, 852A, 863H, in ErbB3 all 12 and in ErbB4 all 20 residues were found to be negatively selected and can act as potential drug target residues.

Conclustion

The present work shows supremacy of Threading over Homology modeling. The results of threading might have resulted from methodological principle advantage of small template folds matches. Therefore, we propose threading to be a better approach for multi domain large proteins. If not, consensus of algorithm should always be taken into consideration during any similar case.

Acknowledgment

The authors acknowledge the support of Department of Computational Biology and Bioinformatics, Sam Higginbottom Institute of Agricultural, Technology and Sciences, Allahabad, U.P., India.