Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/dit

Original Article

Comparative evaluation of commercially available homology modelling tools: A structural bioinformatics perspective

Sween Dahiya [a], *Anjum Gahlaut* [b], *Mahesh Kulharia* [a,*]

[a] *Centre for Bioinformatics, Maharishi Dayanand University, Rohtak, Haryana 124001, India*
[b] *Centre for Biotechnology, Maharshi Dayanand University, Rohtak, Haryana 124001, India*

ARTICLE INFO

Article history:
Received 4 March 2013
Accepted 15 April 2013

Keywords:
Coagulation
Homology modelling
PSI-BLAST
Gapped BLAST
SWISS MODEL

ABSTRACT

Background: Structure based drug design has revolutionised the way new drug molecules are being looked for. A very important technique in this process is homology modelling of protein structures. Although a number of protocols are proposed by a number of research groups, yet a comparative assessment is desired to identify the relative merits and demerits of these programs. Comparative assessment of various homology modelling tools was evaluated using prediction of structure of B-domain of factor V.

Methods: The methods employed, here, for comparative assessment were ESyPred3D, SWISS MODEL, PHYRE2 and YASARA. The criteria for selection of these tools were their popularity among users and level of automation. All these are completely automated and require only protein sequence or alignments as input. These tools were fast and the results were obtained within few hours.

Results: To evaluate the various models of the protein structures, we carried out exhaustive evaluation through "WHATif" and "QMEAN". The parameters included the bond angle, bond length, coarse packing quality control, collision symmetry, omega angle, hand check dihedrals etc.

Conclusion: According to our study YASARA emerged as best performer.

1. Introduction

Coagulation is an important part of haemostasis. The coagulation cascade is characterized by the sequential, rapid, and highly localized activation of a series of serine proteases, culminating in the generation of thrombin, with subsequent conversion of fibrinogen into a fibrin clot.[1] When damaged, sub-endothelial cells (after injury) expose their proteins (sub-endothelial proteins or collagen) which bind and activate the circulating platelets and trigger the release of coagulation factors for instance ATP, kallikerin, factor V etc. Aggregated platelets further activate the coagulation factors in the cascade. Platelet aggregation results in the formation of primary plug at the site of vascular injury and stops the bleeding. Fibrin is the end product of the coagulation cascade. Blood clotting factors play significant role in coagulation and form

* *Corresponding author.*
E-mail address: kulharia@gmail.com (M. Kulharia).
0975-7619/$ – see front matter
http://dx.doi.org/10.1016/j.dit.2013.04.003

coagulation cascade. Proteins in the blood plasma are known as coagulating factors.[2] The factors which are involved in blood clotting are named as coagulation factor I to factor XII.[3,4] In this study main focus is on B-domain of factor V and its homology modelling with various homology modelling tools.

Factor V (FV) is a large, multidomain and asymmetric plasma glycoprotein consisting of six domains (A1-A2-B-A3-C1-C2). It is an essential pro-cofactor which is the proteolytically activated by alpha-thrombin via prethrombin2 pathway. This function is taken over by meizothrombin in amplification phase of the coagulation cascade. A and C-domains are 40% identical between factors V and VIII, but the homology between B domains is very limited.[5] B-domain is mainly found on a single large exon. It regulates expression and activity of factor V and VIII.[6,7] The activation and deactivation of factor V is controlled by thrombin and "Activated Protein C" (APC).[8–10] Factor V is cleaved by thrombin and divided into two heterodimers. One is the heavy chain fragment of 94 KD and the light chain fragment is of 74 KD.[11,12] To predict a structure from its sequence with an accuracy that is comparable to the best results achieved experimentally. Experimental structures provide a solid basis for structure-based drug design, analysis of protein function, interactions, antigenic behaviour, and rational design of proteins with increased stability or novel functions. The CASP (Critical Assessment for protein Structure Prediction) shows the development of the different prediction methods in the last decade. There are different software's which can be used for comparative modelling of proteins. The softwares used in the study are SWISS-MODEL, YASARA, phyre2 and ESyPred3D. These are the automated modelling tools used for homology modelling.

SWISS MODEL is used as web-based servers for automated structure prediction requires minimum user input (i.e. Protein sequence only), SWISS-MODEL offers two more advanced user modes in which users can submit their own multiple sequence alignment or manually adjust the modelling parameters.[13,14] The second method used in the study is ESyPred3D which is also an automated homology modelling program which uses the alignment strategy of neural networks. The next method used is YASARA which is also an automated homology modelling program. YASARA features a complete homology modelling module that automatically generates a high-resolution model using a CASP approved protocol.[15] The templates were ranked on the basis of the alignment score and structural quality according to WHAT_CHECK[16] obtained from the PDBFinder2 database.[17] Phyre2 (Protein homology/Analogy Recognition Engine) is a most popular web-based service for protein structure prediction that are free for non-commercial use.[18] QMEAN and WHATif are web based validation servers used in this study for structure validation and assessment.

2. Materials and methods

The first step in predicting a protein structure consists in searching for database information on the target sequence. Four cycles of PSI-BLAST[19] are performed on the Non Redundant (NR) database of protein sequences with an e-value inclusion threshold of 0.005. Only sequences with at least 30% sequence identity, e-value $\leq 10^{-10}$ and alignment of length at least 2/3 of the target sequence were considered. In difficult cases the e-value inclusion threshold in PSI-BLAST was lowered to 0.02 and sequences up to an e-value of 10^{-5} aligned for over half of the target sequence considered. The domain structure of the target was searched in parallel on the PFAM database.[20] The sequences obtained after PSI-BLAST was checked for the B-domain existence using SMART.

The primary objective of protein modelling is to predict a structure from sequence of target protein. The predicted structure can be used in structure based drug designing, analysis of function and interaction of proteins and their antigenic behaviour.[21] Here our main focus was on the comparative assessment of the various homology modelling tools with respect to B-domain of factor V and validating the structures using appropriate tool. Sequences can be retrieved from GenBank or from Swiss Prot. In this study, sequences were retrieved from GenBank. PSI-BLAST was performed with the sequences retrieved from GenBank. The sequences used in the study are given in Table 1.

These sequences of B-domain were used for BLAST and for the further alignment of the given sequences. ClustalW was used for the sequence alignment. Four different softwares were used for the homology modelling which were SWISS MODEL, ESyPred3D, YASARA and Phyre2.

2.1. SWISS MODEL

SWISS MODEL is an automated server which uses the sequence of the target protein and predicts the structure. The sequence of the target protein which is B-domain of factor V was subjected to SWISS MODEL protocol to obtain its structure. Results were collected from the e-mail ID or directly from the server.

2.2. PHYRE2 and ESYPred3D

Both perform in similar way. Request for the homology modelling was submitted by putting the amino acid sequence of the target protein, the job's name and e-mail ID. Results were collected from e-mail ID or directly from the server.

2.3. YASARA

In YASARA based protocol a target sequence was put into the experiment. The sequence was modelled with the hm_build

Table 1 – Sequences used in study, their accession no. And pI values/molecular weight.

S. no.	Organism	Accession no./pdb id	pI and molecular wt.
1.	*Homo sapiens*	CAC82573.1	4.89/104589.69
2.	*Homo sapiens*	3P6Z	8.88/29738.19
3.	*Bos tarus*	NP_776304.1,	5.53/248983.19
4.	*Bos tarus*	Q28107.1	5.53/248983.19
5.	*Mus musculus*	O88783.1	5.68/247230.37
6.	*Mus musculus*	NP_032002.1	5.68/247230.37