vLife
adding value to life

Performance of Homology Modeling in BioPredicta

By

Dr. Reena Gollapudy and S. Janardan
VLife Sciences Technologies Pvt Ltd.

Aim

The aim of protein homology modeling/comparative modeling is to predict a structure from its sequence with an accuracy that is comparable to experimental results. Homology modeling helps to build a three dimensional structure of a target based on template that helps to identify the putative active sites and binding pockets and the probable ligand-protein interactions. The purpose of this case study is to check performance of homology modeling tool available in BioPredicta.

Methodology

Homology modeling involves taking a known sequence with an unknown structure (target) and mapping it against a known structure (template) of one or several homologous proteins. In this case study human oestrogen receptor alpha (ER-α) (Swiss-Prot: P03372) was taken as target and human oestrogen receptor beta (PDB: 1QKM) was chosen as template. Although, human ER-α is known to have crystal structure (PDB: 1L2I), for the purpose of this case study an assumption was made that such a structure does not exist. The homology modeling was performed by four sequential steps such as template selection, target-template alignment, model construction, and model assessment.

1. The target human ER-α sequence was submitted to BLAST (Basic Local Alignment Search Tool) search to identify suitable templates. From the BLAST search 1QKM was chosen as template based on percentage scores of identities and similarities with respect to human oestrogen receptor alpha.
2. The template was then cleaned with respect to missing residues, improper bond lengths and template residues were mutated to corresponding residues of target human oestrogen receptor alpha.
3. The gaps in target sequence was excised and loops were inserted. For loop building, starting and ending anchoring residues were selected to search the PDB database for loops satisfying the distance criteria within the adjoining secondary structures. The ëhitî loops were selected on the basis of RMSD. Further, the obtained hit fragment was optimized by using MMFF force field upto RMS gradient of 1.0 Kcal/mol/≈.
4. The generated complete homology model was further optimized with MMFF force field upto a RMS gradient of 1.0 Kcal/mol/≈. The superimposition of 1L2I and model protein gives a RMSD of 1.68≈.

Results and analysis

1. The quality of the generated model depends on the confidence in the fold of a model. The correct fold in the model protein depends on high sequence similarity with template and conservation of the key functional or structural residues in the target sequence.
2. The active site of 1L2I consists of (R,R)-5,11-cis-diethyl-5,6,11,12-tetrahydrochrysene-2,8-diol (ETC) as co-crystal and the important active site residues are T347, Q353, R394, G521, H524 and Y525. The superimposition of key active site residues of 1L2I and model protein gives a RMSD of 0.542 ≈. These results show that active site of the modeled protein and the template are conserved. Further, this result could help in predicting ligand binding site of the modeled protein, characterization of active site of target protein and identification of ligand-receptor interactions. Also, homology modeling by VlifeMDS helps not only in identification of active site but also helps in design of site-directed mutant proteins, high-throughput docking and design of highly selective potent compounds.

The sequence identity (59%) and similarity (81%) between target and template sequences were as depicted by their BLAST alignment shown below.

```
Target 309      SLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGF  367
                +L+ +Q+V  LL+AEPP +L S   P+ PF+EASMM  LT LAD+ELVHMI+WAK++PGF
Template 8      ALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGF  65
```

```
Target 368      VDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDM  427
                V+L+L DQV LLE  W+E+LM+GL+WRS++HPGKL+FAP+L+LDR++GKCVEG++EIFDM
Template 66     VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM  125

Target 428      LLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLI  487
                LLAT+SRFR + LQ +E++C+K++ILLNS +Y  +++T  +   +  H +L+ +TD L+

Template 126    LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAH-LLNAVTDALV  184

Target 488      HLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAH  547
                ++AK+G++ QQQ  RLA LL++LSH+RH SNKGMEHL +MKCKNVVP+YDLLLEML+AH
Template 185    WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH  244

Target 548      RLHAPTS  554
                 L    S
Template 245    VLRGCKS  251
```

FIGURE 1: The structure of homology modeled protein, helix (in red) turns/C-α (in green) and sheets (in pink)

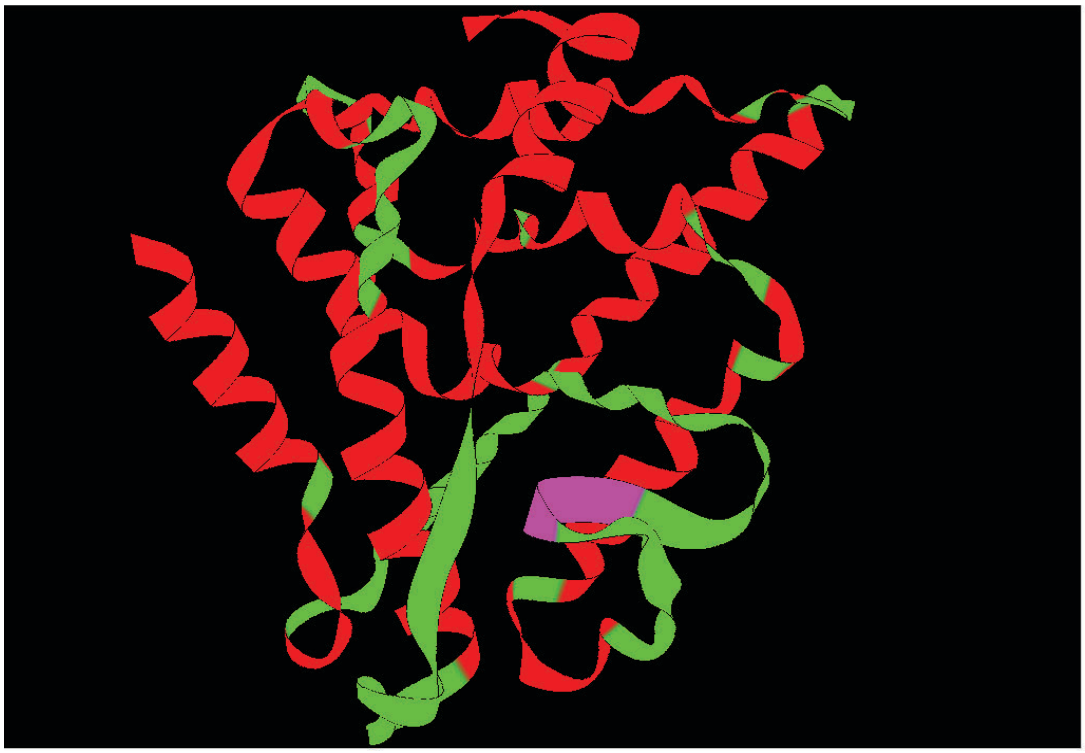

FIGURE 2: Superimposition of generated model with 1L2I

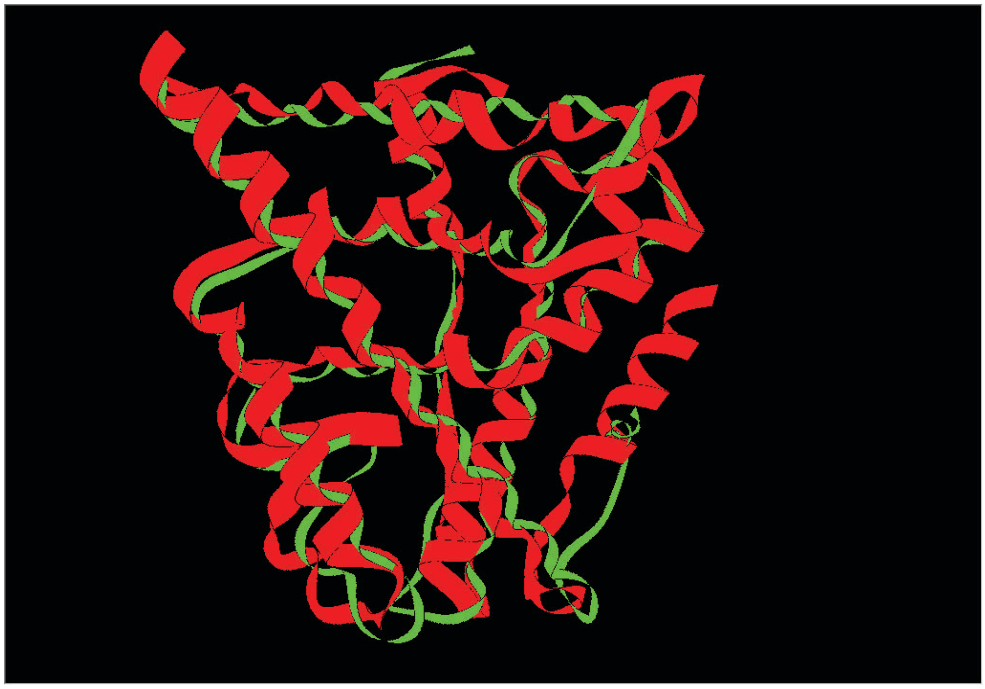

FIGURE 3: Superimposition of active site residues of generated model and 1L2I

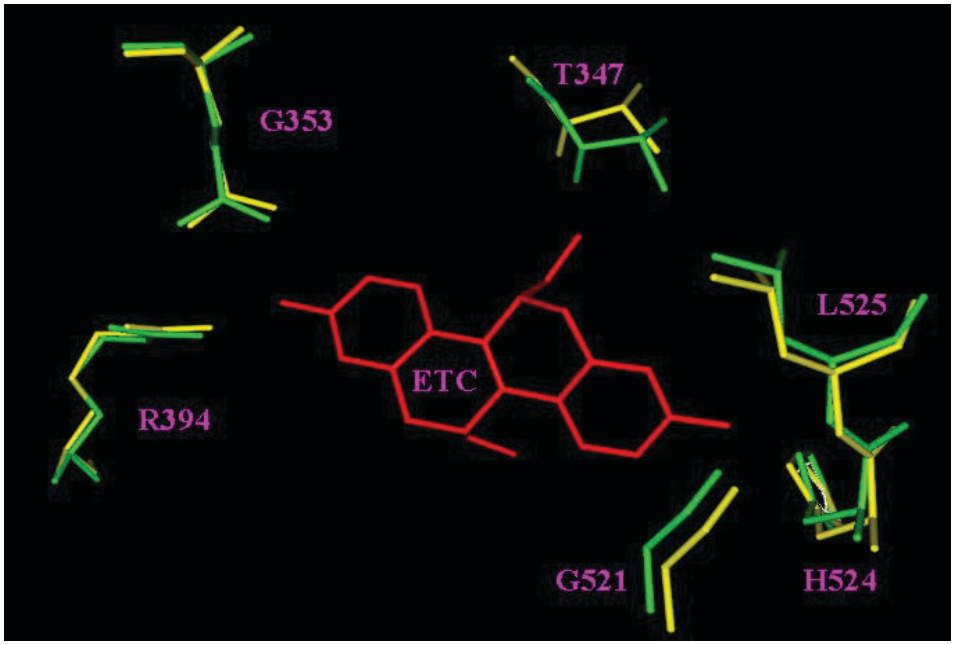

FIGURE 4: Back bone superimposition of generated model with the template (1QKM)