INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print) 2222-5234 (Online)
http://www.innspub.net
Vol. 5, No. 12, p. 374-389, 2014

RESEARCH PAPER OPEN ACCESS

Fragment based homology modeling and simulation based study of endoglin (CD-105) from *Homo sapiens*

Bijal Shah[1], Pramodkumar P. Gupta[2*]

[1]*Veer Narmad South Gujarat University, Surat, Gujarat, India*

[2*]*Department of Biotechnology and Bioinformatics, Padmashree Dr D Y Patil University, Navi Mumba, Maharashtra, India*

Key words: Endoglin, CD105, Homology - modeling, 3D-Structure, CHARMM, NAMD, VMD.

http://dx.doi.org/10.12692/ijb/5.12.374-389 Article published on December 27, 2014

Abstract

The structural data elucidates more information and detailed study in structural informatics, Endoglin (CD105), own a multi dimensional activity results in cure and prophylaxis in many disease and disorder as angiogenesis and a types of cancer. As till date no structural domain or information is available in any of the biological database. Several non-relevant structure are modeled using a homology and automated system, as the identity were relatively below a relevant percentage among the sequence and templates the results has to be discarded. Finally, here we implemented the fragment based homology modeling to model a 3D structure of Endoglin, resulting in a more appropriate fold the protein was studied in dynamic condition by simulating in a solvent body system at 310K using a CHARMM package in VMD-NAMD. The final model was checked by using PROSA and Procheck.

* **Corresponding Author:** Pramodkumar P. Gupta ✉ pramod.gupta@dypatil.edu

Introduction

A numerous experiments and large amount of experimental data has shown that Endoglin/CD105 is expressed on endothelial cells,(Fonsatti *et al.*, 2008; Wong *et al.*, 2000; Wikstrom *et al.*, 2002) and that it is over expressed in vascular endothelial cells of tissues undergoing angiogenesis such as regeneration and inflamed tissues or tumors (Wikstrom *et al.*, 2002; Burrows *et al.*, 1995; Fonsatti *et al.*, 2000; Miller *et al.*, 1999). Research suggests that, levels of endoglin/CD105 positively correlate with the extent of proliferation of endothelial cells (Burrows *et al.*, 1995; Fonsatti *et al.*, 2000; Miller *et al.*, 1999) and has been suggested to be the most suitable marker available to quantify tumor angiogenesis (Fonsatti *et al.*, 2008; Wikstrom *et al.*, 2002; Fonsatti *et al.*, 2000; Wang *et al.*, 1993; Wang *et al.*, 1994; Kumar *et al.*, 1999; Shariat *et al.*, 2008). Altogether, these valuable research findings support the role of endoglin/CD105 as an optimal marker of proliferation of endothelial cells and identifies a promising clinical potential target for clinical investigation such as prognostic, diagnostic, and therapeutic vascular target in human cancer (Fonsatti *et al.*, 2008; Burrows *et al.*, 1995; Ten *et al.*, 2008; Bernabeu *et al.*, 2008). Since over a decade a biochemical and biological functions of endoglin/CD105 are under active investigation, and new data continue to emerge the knowledge on its functional role within the TGF-β receptor complex. The modulation of TGF- β in cellular responses and its involvement in vascular physiology and angiogenesis (Bernabeu *et al.*, 2008; Haruta and Seon, 1986; Bernabeu *et al.*, 2007). Since no reported 3D structure of Endoglin/CD105 in biological databases, here we have considered an *in-silico* methodology to model the 3D structure of Endoglin/CD105 and optimized it using simulation methods.

Materials and methods

Template Identification and Protein modeling

Searching the RCSB Protein Data Bank (http://www.rcsb.org/) confirmed that the tertiary structure of Endoglin (CD-105) was not publicly available (Berman 2008; Fonsatti *et al.*, 2008; Wong *et al.*, 2000; Wikstrom *et al.*, 2002). The complete Endoglin (CD-105) protein sequence, that consists of 658 amino acids and has a calculated molecular weight of 70578 kDa, was retrieved from the UniProtKB database (http://www.uniprot.org/) (accession number P17813) (McAllister *et al.*, 1994). BLASTP (Altschul *et al.*, 1990) was used to identify homologs in the RCSB Protein Data bank (Berman, 2008) but revealed a poor homology with 2WOK (Mackenzie *et al.*, 2010).

Secondary Structure Prediction

Secondary structure for Endoglin (CD-105) sequence was predicted using CFSSP- Chou & Fasman Secondary Structure Prediction Server (Chou and Fasman., 1974; Chou and Fasman., 1974), GOR (Garnier *et al.*, 1996) and SOPMA (Geourjon and Deleage 1995).

Tertiary Structure Prediction

Tertiary structure for Endoglin sequence (CD-105) was predicted and modeled using SWISS-MODEL (Arnold *et al.*, 2006; Kiefer *et al.*, 2009; Peitsch, 1995), CPHmodels (Nielsen *et al.*, 2010), ESyPres3D (Lambert *et al.*, 2002), Geno3d (Combet *et al.*, 2002), Phyre (Successor of 3D-PSSM) (Kelley and Sternberg 2009), HHpred (Remmert *et al.*, 2011; Soding, 2005; Soding *et al.*, 2005), SAM-T08 (Karplus, 2009) and Bhageerath (Jayaram *et al.*, 2012).

Fragment based homology Modeling

As the initial results from BLASTP with poor homolog templates and subsequently low optimal structure from prediction servers leads us to fragment the sequence and search the optimum identical structural template from PDB (Altschul *et al.*, 1990; Berman 2008). The iterative fragmentation to the Endoglin (CD-105) sequence showed some promising identical structural templates from PDB (McAllister *et al.*, 1994; Berman, 2008) and modeled using Modeler tool (Eswar *et al.*, 2006; Martin-renom *et al.*, 2000; Sali and Blundell, 1993; Fiser *et al.*, 2000).

Model Building and validation

The modeled fragmented structure of Endoglin (CD-105) were manually ligated using SPDBV tool (Guex and Peitsch, 1997) program and subjected to molecular dynamics simulation at 500 pico seconds, 01, 05 and 10 nano seconds using CHARMM algorithm under the water solvent system using NAMD (Brooks *et al.*, 2009; Philips *et al.*, 2005). The simulated structures were finally passed through validation process using PROSA (Wiederstein and Sippl 2007; Sippl 1993), Procheck . (Laskowski *et al.*, 1993) and Ramachandran analysis (Ramachandran *et al.*, 1963).

Results and discussions

The output from BLAST-P for Human Endoglin (CD 105) was extremely poor and Delta BLAST as follows: 15% identity with 3NK3_A (Han L *et al.*, 2010), 15% identity with 4AJV_A (Diestel *et al.*, 2013), 16% identity with 3QW9_A (Lin *et al.*, 2011) and 11% identity with 3D4C_A (Monne *et al.*, 2008), belongs to Zona Pellucida domain, ranging among query amino acid 359 - 559.

Table 1. Secondary structure prediction result.

Algorithm	Helix	Extend Strand	Random Coil
Choufasmann	423.00	277.00	73.00
	64.30%	42.10%	11.10%
Gor	163.00	135.00	360.00
	24.77%	20.52%	54.71%
Sopma	130.00	171.00	357.00
	19.60%	25.99%	54.26%

Table 2. Template homology with Endoglin sequence in whole and fragment (1st and 2nd iteration).

Sr no	Query length	Template id	Identity
1	1-658	2WOK_A	30.00%
2	1-329	2YOF_A	34.00%
3	330-658	1HT6_A	29.30%
4	1 – 200	No significant result	---------
5	200-400	3DUZ_A	24.00%
6	400-658	2QPS_A	23.00%

Secondary structure

The output of secondary structure prediction were not clear with percentage determination among the 3 used algorithm, Helix region were greatly supported by Choufasmann and lesser supported by Gor method and least followed by Sopma, on the another hand the Beta sheet / strands were supported by Choufasmann and lesser by Sopma followed by the Gor, Where as the coiled region was merely 1/5th as compare to the output from Gor and Sopma (Table 1.0), supplementary data.

Tertiary Structure

As initially in the search of homolog to Endoglin protein resulted in an extremely poor structure templates, the sequence is subjected to online server for 3D structure prediction.

CPHmodels

The total modeled **r**esidues among the 658 is only 86 range: 252 -338 with a high loop region (Figure 1a).

EsyPred 3D

The total modeled residues among the 658 is only 230 range: 357 -587 with a high loop region (Figure 1b).

Table 3. Template homology with Endoglin sequence in fragment (3rd iteration).

No	Query sequence length	Template pdb_id	Identity	Best model	Ramachandran Analysis
1	1 – 53	1CNS (Song and Suh 1996).	39.00%	Model_1	69.80%
				Model_2	65.10%
				Model_3	79.10%
				Model_4	81.40%
				Model_5	72.10%
2	54 – 200	2EF4_A (kumravel *et al*).	35.00%	Model_1	83.40%
				Model_2	81.60%
				Model_3	79.20%
				Model_4	78.40%
				Model_5	76.00%
3	201 – 300	3P8S_A (Yang and Czapla , 1993; sharma *et al*).	39.00%	Model_1	70.00%
				Model_2	81.20%
				Model_3	75.00%
				Model_4	78.80%
				Model_5	78.30%
4	301 – 400	1HZO_A (Nukaga *et al*., 2002).	34.00%	Model_1	81.40%
				Model_2	79.50%
				Model_3	87.50%
				Model_4	83.00%
				Model_5	87.30%
5	401 – 500	3DUI_A (Lopez-Lucendo *et al*., 2009).	37.00%	Model_1	83.10%
				Model_2	82.00%
				Model_3	84.30%
				Model_4	83.10%
				Model_5	80.90%
6	501 - 658	3CN5_A (Frick *et al*., 2009).	39.00%	Model_1	81.10%
				Model_2	80.30%
				Model_3	82.60%
				Model_4	84.80%
				Model_5	84.10%

Table 4. Optimum selected 3D model for ligation on the basis of Ramachandran analysis.

No	Query sequence length	Template pdb_id	Identity	Best model	Ramachandran Analysis
1	1 – 53	1CNS	39.00%	Model_4	81.40%
2	54 – 200	2EF4_A	35.00%	Model_1	83.40%
3	201 – 300	3P8S_A	39.00%	Model_2	81.20%
4	301 – 400	1HZO_A	34.00%	Model_3	87.50%
5	401 – 500	3DUI_A	37.00%	Model_3	84.30%
6	501 - 658	3CN5_A	39.00%	Model_4	84.80%

Table 5. Ramachandran plot distribution.

	500ps		1ns		5ns		10ns	
Residues in most favoured regions	362	64.3%	363	64.5%	363	64.5%	359	63.8%
Residues in additional allowed regions	151	26.8%	151	26.8%	151	26.8%	151	26.8%
Residues in generously allowed regions	29	5.2%	26	4.6%	26	4.6%	32	5.7%
Residues in disallowed regions	21	3.7%	23	4.1%	23	4.1%	21	3.7%
Number of non-Glycine and non-Proline residues	563	100.0%	563	100.0%	563	100%	563	100.0%
Number of end-residues (excl. Gly and Pro)	2		2		2		2	
Number of Glycine residues (shown as triangles)	44		44		44		44	
Number of Proline residues	49		49		49		49	
Total	658		658		658		658	

Phyre2 (Successor of 3D-PSSM)

The generated 3-D structure with 100% confidence but among 658 residues only 197 residues was modeled (Figure 1c).

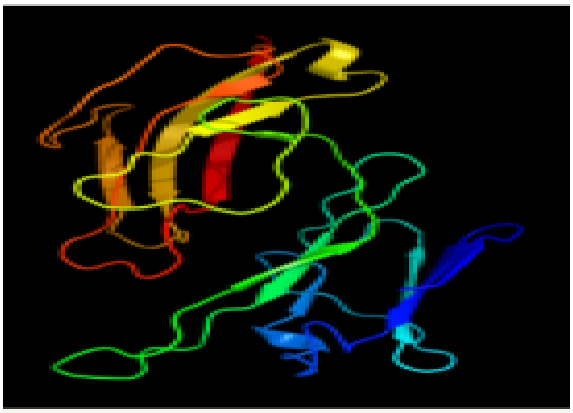

Fig. 1a. 3D model from CPH.

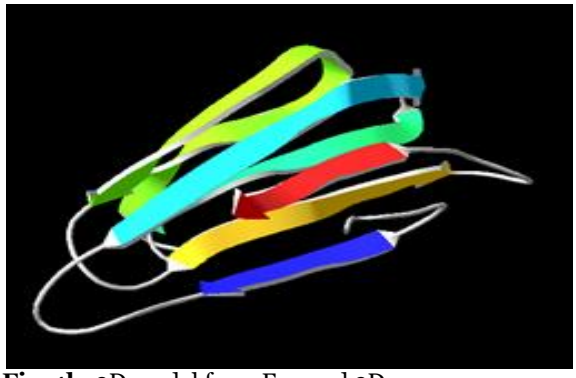

Fig. 1b. 3D model from Esypred 3D.

HHpred

The 3D model have maximum loop region. Only short lengths of residues were modeled (Figure 1d).

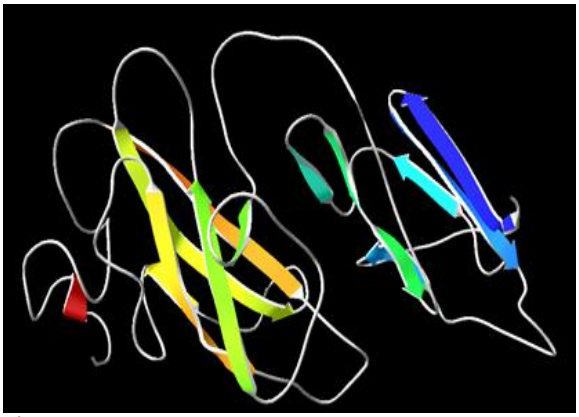

Fig. 1c. 3D model from Phyre2.

SAMTo8

The 3-D model have several cuts present between residues (backbone was not complete) (Figure 1e).

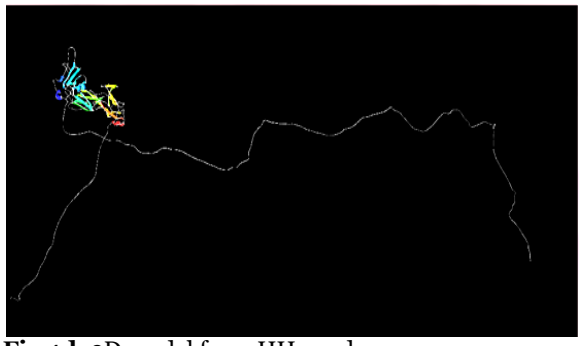

Fig.1d. 3D model from HH-pred.

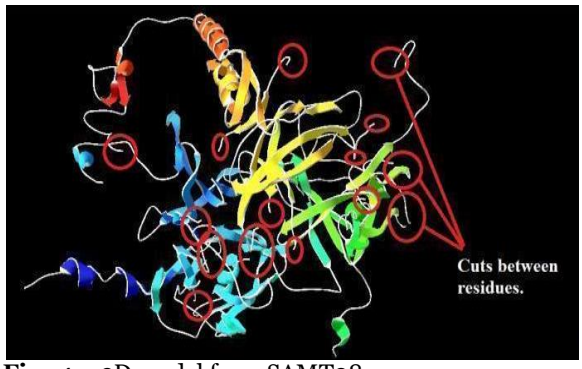

Fig. 1e. 3D model from SAMTo8.

Bhageerath H

The 3-D model have multiple number of 40 cuts i.e. missing residual information (backbone was not complete) (Figure 1f).

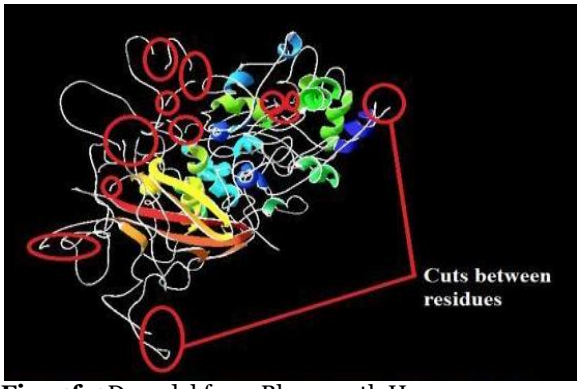

Fig. 1f. 3D model from Bhageerath H.

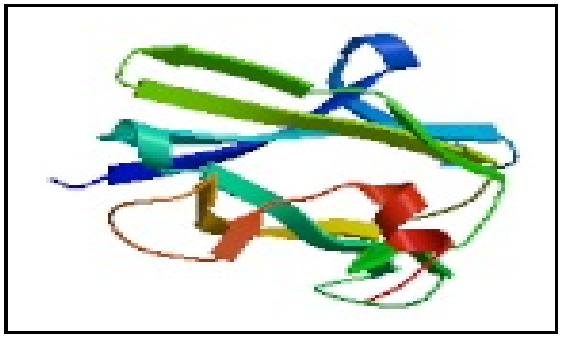

Fig. 2a. 3D model from Swiss model server

Homology modelling

Using the homology Swiss model server, based on template 3QW9_A (Lin *et al.*, 2011) with 16% identities with a short stretch of structure ranging sequence 445 – 578 (Figure 2a). The first output for the entire sequence length with 2WOK_A (Fiser *et al.*, 2000) shows a very poor homology with template and modeled structure had a high loop region (Figure 2b).

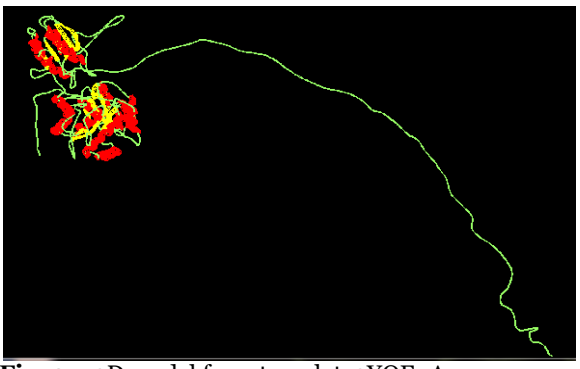

Fig. 2c. 3D model from template2YOF_A.

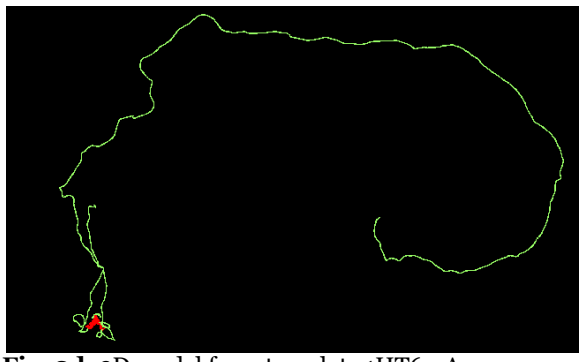

Fig. 2d. 3D model from template 1HT6 _A.

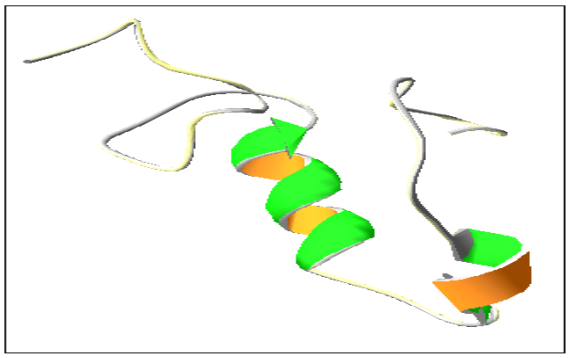

Fig. 3a. 3D model of 1-53Amino acids. Amino acids.

Fragment Based homology modeling

The first fragmentation process resulted into two segments, the first shows 34.00% identity with 2YOF_A (Cui *et al.*, 2012) having a maximum loop region (Figure 2c). The second segment shows a 29.30% identity with 1HT6_A (Robert *et al.*, 2003) (Figure 2d). Whereas in second iteration the entire sequence is segmented in three fragments, the first segment from 1-200 show no significant result. The second segment 200-400 shows a 24.00% identity wit 3DUZ_A (kadlec *et al.*, 2008) and third segment shows a 23.00% identity with 2QPS_A (Bozonnet *et al.*, 2007). All models predicted by using different matrix were non relevant structures. All models have maximum loop region in their 3-D structure. In the third iteration the protein was fragmented down in six parts and homology models were generated (Figure 3a – 3.f).

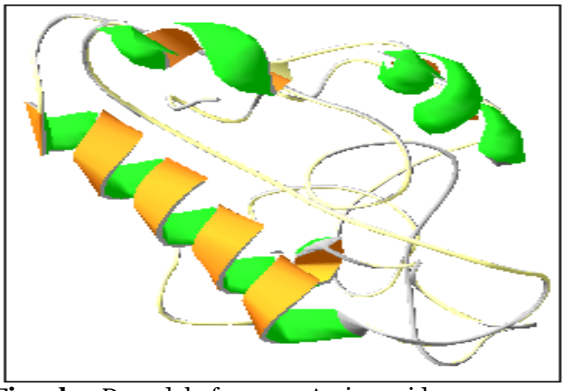

Fig. 3b. 3D model of 54-200 Amino acids.

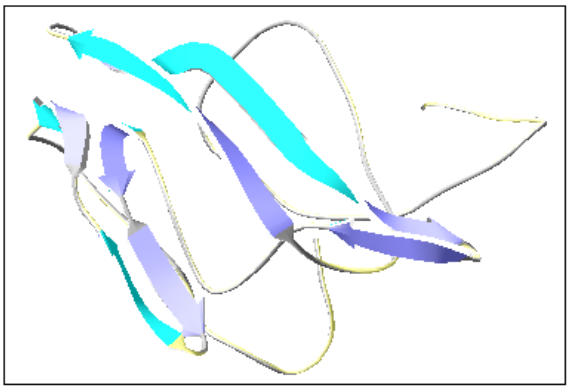

Fig. 3c. 3D model of 201-300 Amino acids.

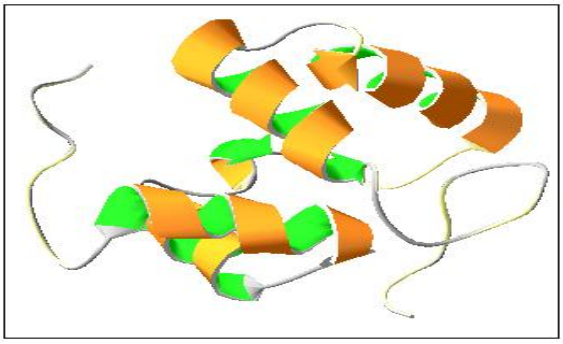

Fig. 3d. 3D model of 301-400 Amino acids.

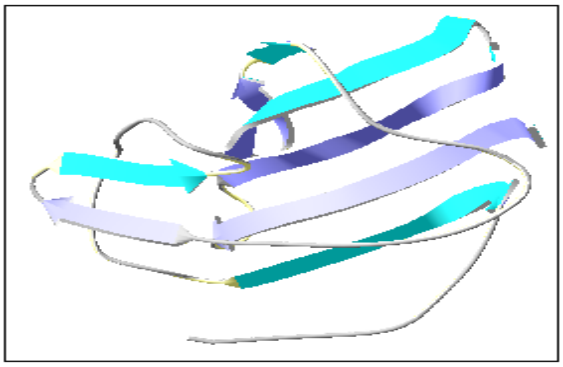

Fig. 3e. 3D model of 401-500 Amino acids.

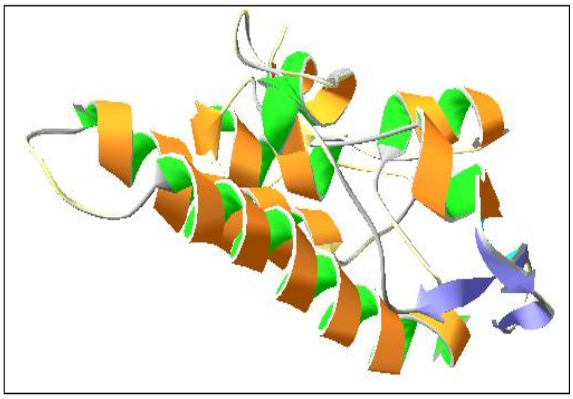

Fig. 3f. 3D model of 501-658 Amino acids.

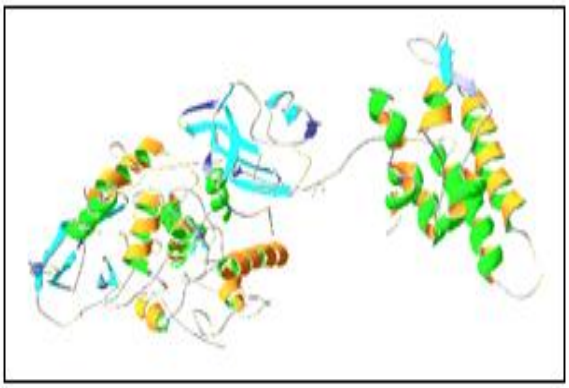

Fig. 3g. Final 3D structure of Endoglin.

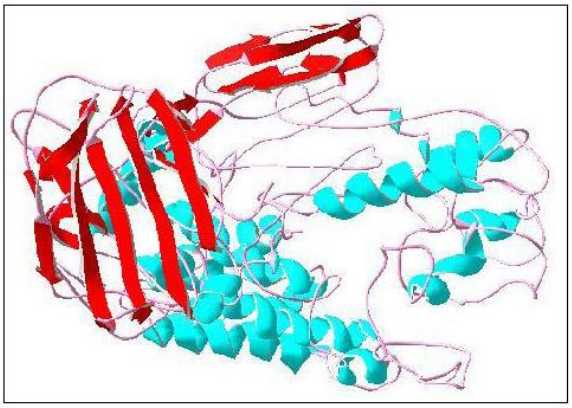

Fig. 4a. 3D structure of Endoglin system.

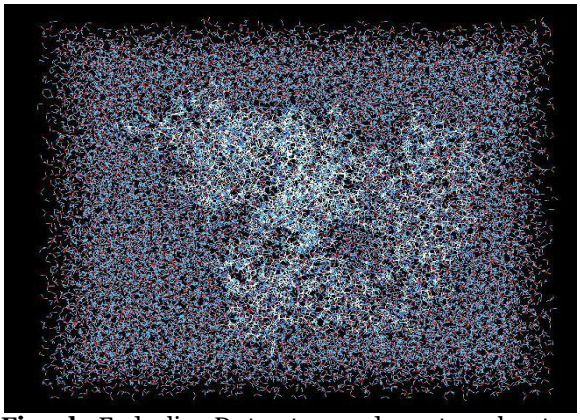

Fig. 4b. Endoglin 3D structure under water solvent.

For ligating the 3D structure, we used the best result of modelled fragments by validating it through Ramachandran plot analysis considering the most favoured regions of different models (Figure 3g). The structure was initially optimized using Gromacs and energy was calculated: 3677522.703884 k/j (Figure 4a) and simulated under water body condition using CHARMM in NAMD with a time scale 500 pico seconds, 1, 5, and 10 nano seconds (Figure 4b). The following result in table no: 5.0 are giving the residual organizing in Ramachandran plot after the 500 pico-seconds, 1 nano-second, 5 nano-seconds and 10 nano-seconds.

Structure validation

The distribution of φ, ψ angles showed a significant residues in the most favored, additional allowed and generously allowed regions comprising more than 90% (Table: 5.0). The model quality was tested under PROSA and Procheck and resulted a Z-score as foolows 500 ps: -3.34, 01 ns: -3.41, 05 ns: -3.41 and 10 ns: -3.44 (Figure 5a – 5l).

Discussion

The 3D modeled structure of Endoglin was travelled through a multi process of simulation and optimization resulted into a refined structure, but even though the 3D modeled structure is not at optimum level. In future the identification and submission of newer protein structure may increase in the percentage homology between the template and query sequence of Endoglin may reveal an optimum model.

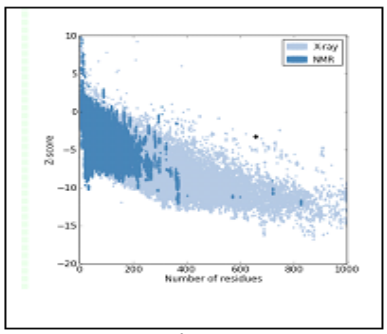

Fig. 5a.

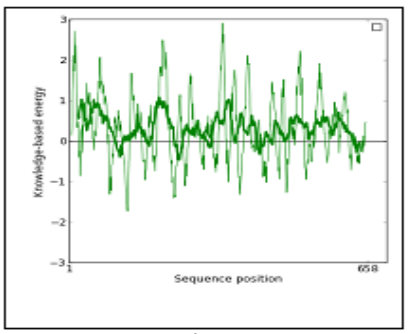

Fig. 5b.

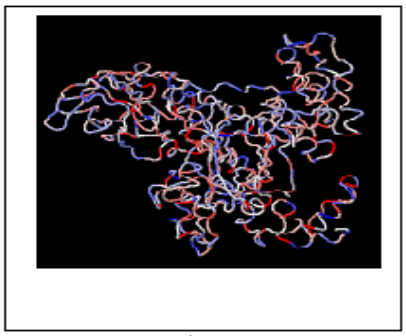

Fig. 5c.

Fig. Shows the z-score of -3.34 after 500 pico-seconds of simulation.

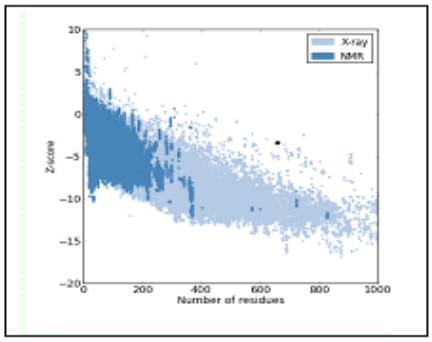

Fig. 5d.

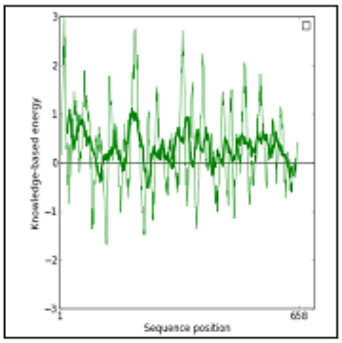

Fig. 5e.

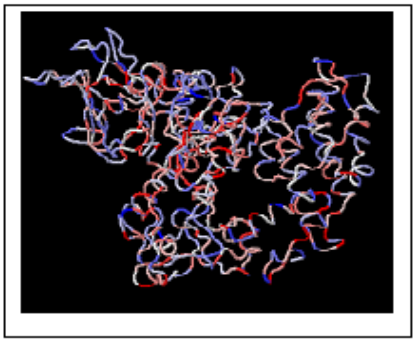

Fig. 5f.

Fig. Shows the z-score of -3.41 after 01 nano-seconds of simulation.

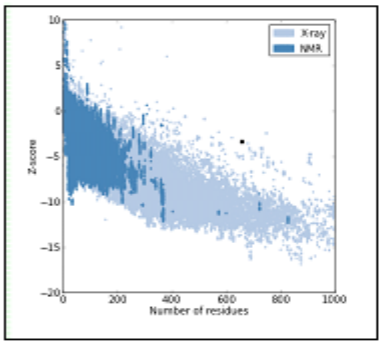

Fig. 5g.

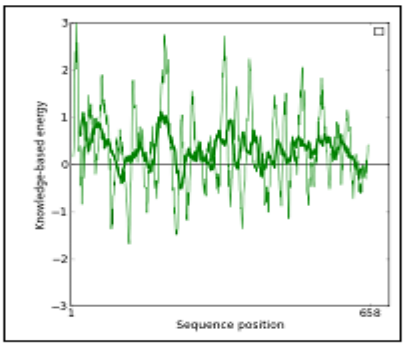

Fig. 5h.

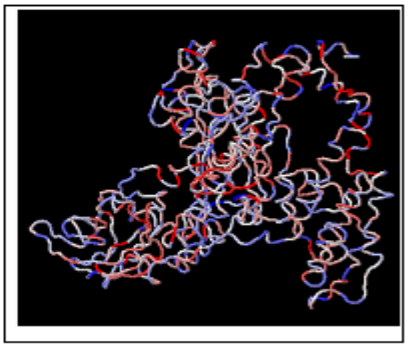

Fig. 5i.

Fig. Shows the z-score of -3.41 after 05 nano-seconds of simulation.

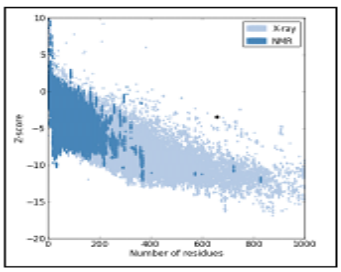

Fig. 5j.

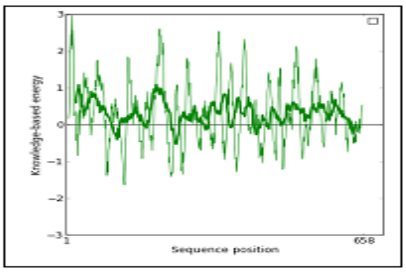

Fig. 5k.

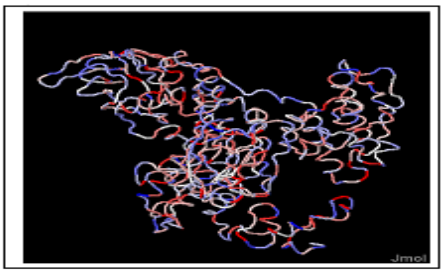

Fig. 5l.

Fig. Shows the z-score of -3.44 after 10 nano-seconds of simulation.

Disclaimers

An author declares that there is no conflict of interests.

Supplementary data (1)

Results for 2D prediction:

```
          10      20      30      40      50      60
          |       |       |       |       |       |
          *       *       *       *       *
Query 1   MDRGTLPLAVALLLASCSLSPTSLAETVHCDLQPVGPERGEVTYTTSQVSKGCVAQAPNA
Helix 1     <----------->     <------->           <-----------
Sheet 1     EEEEEEEEEE        EEEEEEEE      EEEEEEE
Turns 1     TT            T          T T     T  T     T
Choufasman


GOR       MDRGTLPLAVALLLASCSLSPTSLAETVHCDLQPVGPERGEVTYTTSQVSKGCVAQAPNA
          ccccccchhhhhhhccccccccccceeeccccccccccceeeeeeeccccccccccc

Sopma     MDRGTLPLAVALLLASCSLSPTSLAETVHCDLQPVGPERGEVTYTTSQVSKGCVAQAPNA
          ccccccchhhhhhhhhcccccchhhhhhhhhccccccccceeeeeccccccccehccccc

          70      80      90      100     110     120
          |       |       |       |       |       |
          *       *       *       *       *
Query 61  ILEVHVLFLEFPTGPSQLELTLQASKQNGTWPREVLLVLSVNSSVFLHLQALGIPLHLAY 120
Helix 61  --------->     <--------> <--------> <------------> 120
Sheet 61  EEEEEEEEEE     EEEEEE        EEEEEEEEEEEEEEEEEEEEEEEE   120
Turns 61          T TT       T TT  T       T            120
Choufasman

GOR       ILEVHVLFLEFPTGPSQLELTLQASKQNGTWPREVLLVLSVNSSVFLHLQALGIPLHLAY
          Hhhhheeeeccccccchhhhhhhhhcccccccceeeeeeccccchhhhhhhccchhhhcc

Sopma     ILEVHVLFLEFPTGPSQLELTLQASKQNGTWPREVLLVLSVNSSVFLHLQALGIPLHLAY
          hhheeeeeeccccccchheeeeeccccccccccceeeeeeccchheeeeccccccceeeee

          130     140     150     160     170     180
          |       |       |       |       |       |
          *       *       *       *       *
Query 121 NSSLVTFQEPPGVNTTELPSFPKTQILEWAAERGPITSAAELNDPQSILLRLGQAQGSLS 180
Helix 121   <>          <---------------------------------------- 180
Sheet 121   EEEE    EEE    EEEEEEEE            EEEEEE      EE 180
Turns 121 T     T T        T       TT       T T       T   180
Choufasman

GOR       NSSLVTFQEPPGVNTTELPSFPKTQILEWAAERGPITSAAELNDPQSILLRLGQAQGSLS
          ccceeeccccccccccccccchhhhhhhhhccccchhhhccchhhhhhhhccccccc
```

```
Sopma    NSSLVTFQEPPGVNTTELPSFPKTQILEWAAERGPITSAAELNDPQSILLRLGQAQGSLS
             Ccceeeeccccccccccccccchhhhhhhhcccceeeeeeccccceeeeeeccccccc


          190     200     210     220     230     240
          |       |       |       |       |       |
          *       *       *       *       *
Query 181 FCMLEASQDMGRTLEWRPRTPALVRGCHLEGVAGHKEAHILRVLPGHSAGPRTVTVKVEL 240
Helix 181 -------->           <------------------>        <------ 240
Sheet 181 EEE                       EEEE       EEEEEE   240
Turns 181       T  T    T                    T        240
Choufasman

GOR   FCMLEASQDMGRTLEWRPRTPALVRGCHLEGVAGHKEAHILRVLPGHSAGPRTVTVKVEL
      Hhhhhhhhhhhhhhcccccceeeccccccccccchhheeeeeccccccccceeeeeeee

Sopma    FCMLEASQDMGRTLEWRPRTPALVRGCHLEGVAGHKEAHILRVLPGHSAGPRTVTVKVEL
      ccccccccccchhhhhcccccccccccccchhccchheeeeeccccccccceeeeeeee


          250     260     270     280     290     300
          |       |       |       |       |       |
          *       *       *       *       *
Query 241 SCAPGDLDAVLILQGPPYVSWLIDANHNMQIWTTGEYSFKIFPEKNIRGFKLPDTPQGLL 300
Helix 241 ->  <------>    <------------------------------>    <--- 300
Sheet 241     EEEEEEE   EEEEE    EEEEEEEEEEEE          EEEE   300
Turns 241   T       TT           T      TT      T T   300
Choufasman

GOR   SCAPGDLDAVLILQGPPYVSWLIDANHNMQIWTTGEYSFKIFPEKNIRGFKLPDTPQGLL
      Ccccceeeeeeeccccceeeeeeccccccceeeeccccccccccccccccccccchh

Sopma        SCAPGDLDAVLILQGPPYVSWLIDANHNMQIWTTGEYSFKIFPEKNIRGFKLPDTPQGLL
      Ccccccceeeeeccccceeeeeeccccceeeeeccccceeeeccccccccccccchhhhh


          310     320     330     340     350     360
          |       |       |       |       |       |
          *       *       *       *       *
Query 301 GEARMLNASIVASFVELPLASIVSLHASSCGGRLQTSPAPIQTTPPKDTCSPELLMSLIQ 360
Helix 301 -------------------------->               <--------- 360
Sheet 301     EEEEEEEEEE                EEE       EEEEEEE 360
Turns 301     T               T         TT   T      360
Choufasman
```

```
GOR   GEARMLNASIVASFVELPLASIVSLHASSCGGRLQTSPAPIQTTPPKDTCSPELLMSLIQ
      Hhhhhhhhhhhhhhhhccccceeeeeccccccccccccccccccccccccchhhhhhhh
Sopma GEARMLNASIVASFVELPLASIVSLHASSCGGRLQTSPAPIQTTPPKDTCSPELLMSLIQ
      hhhhhhhhhheeeeeeccccceeeeeeccccccccccccccccccccccchhheeehhc

                    370     380     390     400     410     420
             |      |       |       |       |       |
             *      *       *       *       *
Query 361 TKCADDAMTLVLKKELVAHLKCTITGLTFWDPSCEAEDRGDKFVLRSAYSSCGMQVSASM 420
Helix 361 ---------------------->          <--------->     <------ 420
Sheet 361 EE    EEEE    EEEEEEEEEEEE              EEEEE    420
Turns 361 T       T          T    T T    T         420
Choufasman

GOR   TKCADDAMTLVLKKELVAHLKCTITGLTFWDPSCEAEDRGDKFVLRSAYSSCGMQVSASM
      Hcccchhhhhhhhhhhhhhhceeeeeeeeeccccccccccccceeeeeccccccceecchh

Sopma TKCADDAMTLVLKKELVAHLKCTITGLTFWDPSCEAEDRGDKFVLRSAYSSCGMQVSASM
      ccccccchheeeehhhhhhhhhhhheeeeeeccccccccccccheeeeccccccccehcccc

             430     440     450     460     470     480
             |       |       |       |       |       |
             *       *       *       *       *
Query 421 ISNEAVVNILSSSSPQRKKVHCLNMDSLSFQLGLYLSPHFLQASNTIEPGQQSFVQVRVS 480
Helix 421 ---------------------> <--------> <------------------>  480
Sheet 421     EEEE       EEE    EEEEEEEE          EEEEEEEE  480
Turns 421  T      T    T           T    T  T T      480
Choufasman

GOR   ISNEAVVNILSSSSPQRKKVHCLNMDSLSFQLGLYLSPHFLQASNTIEPGQQSFVQVRVS
      hhhhhheeeeccccccceeeeeeccccchhhhhhhccccccccccccccccceeeeecc


Sopma ISNEAVVNILSSSSPQRKKVHCLNMDSLSFQLGLYLSPHFLQASNTIEPGQQSFVQVRVS
      Cccheeeeeccccccchheeeeccccchhheeeeccchhhhcchhhhccccceeeeecc

             490     500     510     520     530     540
             |       |       |       |       |       |
             *       *       *       *       *
Query 481 PSVSEFLLQLDSCHLDLGPEGGTVELIQGRAAKGNCVSLLSPSPEGDPRFSFLLHFYTVP 540
Helix 481  <------>          <------->           <---->     540
Sheet 481  EEEEEEE          EEEEE      EEEE       EEEEEEEEEEE 540
Turns 481 T           T T     T   T     T   T         540
Choufasman
GOR   PSVSEFLLQLDSCHLDLGPEGGTVELIQGRAAKGNCVSLLSPSPEGDPRFSFLLHFYTVP
```

Ccchhhhhhhhcccceeccccccchhhhhhhhhccccceeeeccccccccccceeeeccceec

Sopma PSVSEFLLQLDSCHLDLGPEGGTVELIQGRAAKGNCVSLLSPSPEGDPRFSFLLHFYTVP
Ccchhhhhhhhcceeccccccceeeeeccccccccceeeeeccccccccceeeeeeeeecc

550 560 570 580 590 600
| | | | | |
* * * * *

Query 541 IPKTGTLSCTVALRPKTGSQDQEVHRTVFMRLNIISPDLSGCTSKGLVLPAVLGITFGAF 600
Helix 541 <----> <----------> <---------- 600
Sheet 541 EEEEEEEEEEEE EEEEEEEEEEE EEEEEEEEEEEEEE 600
Turns 541 T T T T T T 600
Choufasman

GOR IPKTGTLSCTVALRPKTGSQDQEVHRTVFMRLNIISPDLSGCTSKGLVLPAVLGITFGAF
Ccccceeceeeeeccccccchhhhhhhhhheeeeccccccccccceeceeccccchh

Sopma IPKTGTLSCTVALRPKTGSQDQEVHRTVFMRLNIISPDLSGCTSKGLVLPAVLGITFGAF
ccccccceeeeeeccccccchhhhhhhhhhhhcccccchhhccceehheeeeccchh

610 620 630 640 650 658
| | | | | |
* * * * *

Query 601 LIGALLTAALWYIYSHTRSPSKREPVVAVAAPASSESSSTNHSIGSTQSTPCSTSSMA 658
Helix 601 ----------> <----------> 658
Sheet 601 EEEEEEEEEEEEEE EEE 658
Turns 601 T T T T T 658
Choufasman

GOR LIGALLTAALWYIYSHTRSPSKREPVVAVAAPASSESSSTNHSIGSTQSTPCSTSSMA
Hhhhhhhhhhheeeccccccccccceeeeeccccccccccceeeeccccccceeeeec

Sopma LIGALLTAALWYIYSHTRSPSKREPVVAVAAPASSESSSTNHSIGSTQSTPCSTSSMA
Hhhhhhhhhhhheeccccccccccceeeeeccccccccccccccccccccccceh

Abbreviation

ps: Pico second
ns: Nano second.

References

Fonsatti E, Del Vecchio L, Altomonte M, Sigalotti L, Nicotra MR, Coral S, Natali PG, Maio M. 2008. Endoglin: an accessory component of the TGF-beta-binding receptor-complex with diagnostic, prognostic, and bio immune therapeutic potential in human malignancies J Cell Physiol. **188,** 1-7. http://dx.doi.org/10.1002/jcp.1095.

Wong SH, Hamel L, Chevalier S, Philip A. 2000. Endoglin expression on human microvascular endothelial cells association with betaglycan and formation of higher order complexes with TGF-beta

signalling receptors. Eur J Biochem **267,** 5550-5560. http://dx.doi.org/10.1046/j.1432-1327.2000.01621.x.

Wikstrom P, Lissbrant IF, Stattin P, Egevad L, Bergh A. 2002. Endoglin CD105 is expressed on immature blood vessels and is a marker for survival in prostate cancer. Prostate **51,** 268-275. http://dx.doi.org/10.1002/pros.10083.

Burrows FJ, Derbyshire EJ, Tazzari PL, Amlot P, Gazdar AF, King SW, Letarte M, Vitetta ES, Thorpe PE. 1995. Up-regulation of endoglin on vascular endothelial cells in human solid tumors: implications for diagnosis and therapy. Clin Cancer Res. **1,** 1623-1634.

Fonsatti E, Jekunen AP, Kairemo KJ, Coral S, Snellman M, Nicotra MR, Natali PG, Altomonte M, Maio M. 2000. Endoglin is a suitable target for efficient imaging of solid tumors: in vivo evidence in a canine mammary carcinoma model. Clin Cancer Res. **6,** 2037-2043.

Miller DW, Graulich W, Karges B, Stahl S, Ernst M, Ramaswamy A, Sedlacek HH, Müller R, Adamkiewicz J. 1999. Elevated expression of endoglin, a component of the TGF-beta-receptor complex, correlates with proliferation of tumor endothelial cells. Int J Cancer **81,** 568-572. http://dx.doi.org/10.1002/(SICI)10970215(19990517)81:4<568::AID-IJC11>3.0.CO;2-X.

Wang JM, Kumar S, Pye D, van Agthoven AJ, Krupinski J, Hunter RD. 1993. A monoclonal antibody detects heterogeneity in vascular endothelium of tumours and normal tissues. Int J Cancer **54,** 363-370. http://dx.doi.org/10.1002/ijc.2910540303.

Wang JM, Kumar S, Pye D, Haboubi N, al-Nakib L. 1994. Breast carcinoma: comparative study of tumor vasculature using two endothelial cell markers. J Natl cancer inst **86,** 386-388.

Kumar S, Ghellal A, Li C, Byrne G, Haboubi N, Wang JM, Bundred N. 1999. Breast carcinoma: vascular density determined using CD105 antibody correlates with tumor prognosis. Cancer Res **59,** 856-861.

Shariat SF, Karam JA, Walz J, Roehrborn CG, Montorsi F, Margulis V, Saad F, Slawin KM, Karakiewicz PI. 2008. Improved prediction of disease relapse after radical prostatectomy through a panel of preoperative blood- based biomarkers. Clin Cancer Res. **14,** 3785-3791. http://dx.doi.org/10.1158/1078-0432.CCR-07-4969.

Ten Dijke P, Goumans MJ, Pardali E. 2008. Endoglin in angiogenesis and vascular diseases. Angiogenesis **11,** 79-89. http://dx.doi.org/10.1007/s10456-008-9101-9.

Bernabeu C, Lopez-Novoa JM, Quintanilla M. 2009. The emerging role of TGF-beta superfamily coreceptors in cancer. Biochim Biophys Acta. **1792,** 954–973. http://dx.doi.org/10.1016/j.bbadis.2009.07.003.

Haruta Y, Seon BK. 1986. Distinct human leukemia-associated cell surface glycoprotein GP160 defined by monoclonal antibody SN6. Proc Natl Acad Sci USA, **83,** 7898-7902.

Bernabeu C, Conley BA, Vary CPH. 2007. Novel biochemical pathways of endoglin in vascular cell physiology. J Cell Biochem. **102,** 1375-1388. http://dx.doi.org/10.1002/jcb.21594.

Berman HM. 2008. "The Protein Data Bank: a historical perspective". Acta Crystallographica Section A: Foundations of Crystallography A **64,** 88–95.

McAllister KA, Grogg KM, Johnson DW, Gallione CJ, Baldwin MA, Jackson CE, Helmbold EA, Markel DS, McKinnon WC, Murrell J, McCormick5 MK, Pericak-Vance6 MA, Heutink7 P, Oostra7 BA, Haitjema8 T, Westerman8 CJJ, Porteous9 ME, Guttmacher4 AE, Letarte10 M, Marchuk DA.

1994. Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1.Nat Genet **8,** 345-51.
http://dx.doi.org/10.1038/ng1294-345.

Altschul, SF, Gish W, Miller W, Myers EW, Lipman, DJ. 1990. "Basic local alignment search tool." J. Mol. Biol **215,** 403-410.

Mackenzie AK, Valegård K, Iqbal A, Caines ME, Kershaw NJ, Jensen SE, Schofield CJ Andersson I. 2010. Crystal structures of an oligopeptide-binding protein from the biosynthetic pathway of the beta-lactamase inhibitor clavulanic acid. J Mol Biol **396,** 332-44.
http://dx.doi.org/10.1016/j.jmb.2009.11.045.

Chou PY, Fasman GD. 1974. Prediction of protein conformation. Biochemistry **13,** 222–245.
http://dx.doi.org/10.1021/bi00699a002.

Chou PY, Fasman GD. 1974. Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. Biochemistry **13,** 211–222.
http://dx.doi.org/10.1021/bi00699a001

Garnier J, Gibrat J-F, Robson. 1996. GOR secondary structure prediction method version IV. Methods in Enzymology R.F. Doolittle Ed **266,** 540-553.

Geourjon C, Deleage G. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci **11,** 681-684.
http://dx.doi.org/10.1093/bioinformatics/11.6.681.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics **22,** 195-201.
http://dx.doi.org/10.1093/bioinformatics/bti770.

Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. 2009. The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. **37,** D387-D392.
http://dx.doi.org/10.1093/nar/gkn750.

Peitsch MC. 1995. Protein modeling by E-mail. Nature Biotechnology **13,** 658 - 660.
http://dx.doi.org/10.1038/nbt0795-658.

Nielsen M, Lundegaard C, Lund O, Petersen TN. 2010. CPHmodels-3.0 - Remote homology modeling using structure guided sequence profiles Nucleic Acids Research **38,** 576-81.
http://dx.doi.org/10.1093/nar/gkq535.

Lambert C, Leonard N, De Bolle X, Depiereux E. 2002. ESyPred3D: Prediction of proteins 3D structures. Bioinformatics **18,** 1250-1256.
http://dx.doi.org/10.1093/bioinformatics/18.9.1250.

Combet C, Jambon M, Deléage G, Geourjon C. 2002. Geno3D: Automatic comparative molecular modelling of protein. Bioinformatics **18,** 213-214.
http://dx.doi.org/10.1093/bioinformatics/18.1.213.

Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the web: a case study using the Phyre server, Nature Protocols **4,** 363-371.
http://dx.doi.org/10.1038/nprot.2009.2.

Remmert M, Biegert A, Hauser A, Söding J. 2011. HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods **9,** 173-5.
http://dx.doi.org/10.1038/nmeth.1818.

Söding J. 2005. Protein homology detection by HMM-HMM comparison. Bioinformatics **21,** 951-960.
http://dx.doi.org/10.1093/bioinformatics/bti125.

Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research **33,** 244 - 248.

http://dx.doi.org/10.1093/nar/gki408.

Karplus K. 2009. SAM-T08, HMM-based protein structure prediction. Nucleic Acids Res. **37,** 492-7. http://dx.doi.org/10.1093/nar/gkp403.

Jayaram B, Dhingra P, Lakhani B, Shekhar S. 2012. Bhageerath - Targeting the Near Impossible: Pushing the Frontiers of Atomic Models for Protein Tertiary Structure Prediction. Journal of Chemical Sciences **124,** 83-91. http://dx.doi.org/10.1007/s12039-011-0189-x.

Eswar N, Marti-Renom MA, Webb B, Madhusudhan M S, Eramian D, Shen M, Pieper U, Sali A. 2006. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics **15(5),** 6.1-5.6.30. http://dx.doi.org/10.1002/0471250953.bi0506s15.

Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, Sali A. 2000. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct **29,** 291-325. http://dx.doi.org/10.1146/annurev.biophys.29.1.291.

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol **234,** 779-815. http://dx.doi.org/10.1006/jmbi.1993.1626.

Fiser A, Do RK, Sali A. 2000. Modeling of loops in protein structures, Protein Science **9,** 1753-1773. http://dx.doi.org/10.1110/ps.9.9.1753.

Guex N, Peitsch MC. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis **18,** 2714-2723. http://dx.doi.org/10.1002/elps.1150181505.

Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. 2009. CHARMM: The Biomolecular Simulation Program J Comput Chem **30,** 1545–1614. http://dx.doi.org/10.1002/jcc.21287.

Phillips J C, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R D, Kale L, Schulten K. 2005. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry **26,** 1781-1802. http://dx.doi.org/10.1002/jcc.20289.

Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research **35,** 407-410. 43.

Sippl MJ. 1993. Recognition of Errors in Three-Dimensional Structures of Proteins. Proteins **17,** 355-362. http://dx.doi.org/10.1002/prot.340170404.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst **26,** 283-291. http://dx.doi.org/10.1107/S0021889892009944.

Ramachandran GN, Ramakrishnan C, Sasisekharan V. 1963. "Stereochemistry of polypeptide chain conFigureurations". Journal of Molecular Biology **7,** 95 - 9.

Han L, Monné M, Okumura H, Schwend T, Cherry AL, Flot D, Matsuda T, Jovine L. 2010. Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. Cell **143,** 404-15. http://dx.doi.org/10.1016/j.cell.2010.09.041.

Diestel U, Resch M, Meinhardt K, Weiler S,

Hellmann TV, Mueller TD, Nickel J, Eichler J, Muller YA. 2013. Identification of a Novel TGF-β-Binding Site in the Zona Pellucida C-terminal ZP-C Domain of TGF-β-Receptor-3 TGFR-3. PLoS One. **8,** 67214. http://dx.doi.org/10.1371/journal.pone.0067214.

Lin SJ, Hu Y, Zhu J, Woodruff TK, Jardetzky TS. 2011. Structure of betaglycan zona pellucida ZP-C domain provides insights into ZP-mediated protein polymerization and TGF-beta binding. Proc Natl Acad Sci U.S.A **108,** 5232-6. http://dx.doi.org/10.1073/pnas.1010689108.

Monné M, Han L, Schwend T, Burendahl S, Jovine L. 2008. Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. Nature **456,** 653-7. http://dx.doi.org/10.1038/nature07599.

Cui H, Carrero-Lérida J, Silva AP, Whittingham JL, Brannigan JA, Ruiz-Pérez LM, Read KD, Wilson KS, González-Pacanowska D, Gilbert IH. 2012. Synthesis and evaluation of α-thymidine analogues as novel antimalarials. J Med Chem **55,** 10948-57. http://dx.doi.org/10.1021/jm301328h.

Robert X, Haser R, Gottschalk TE, Ratajczak F, Driguez H, Svensson B, Aghajari N. 2003. The structure of barley alpha-amylase isozyme 1 reveals a novel role of domain C in substrate recognition and binding: a pair of sugar tongs. Structure **11,** 973-84. http://dx.doi.org/10.1016/S0969-2126(03)00151-5.

Kadlec J, Loureiro S, Abrescia NG, Stuart DI, Jones IM. 2008. The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. Nat Struct Mol Biol **15,** 1024-30. http://dx.doi.org/10.1038/nsmb.1484.

Bozonnet S, Jensen MT, Nielsen MM, Aghajari N, Jensen MH, Kramhøft B, Willemoës M, Tranier S, Haser R, Svensson B. 2007. The 'pair of sugar tongs' site on the non-catalytic domain C of barley alpha-amylase participates in substrate binding and activity. Epub **274,** 5055-67. http://dx.doi.org/10.1111/j.1742-4658.2007.06024.x.

Song HK, Suh SW. 1996. Refined structure of the chitinase from barley seeds at 2.0 a resolution. Acta Crystallogr D Biol Crystallogr **52,** 289-98. http://dx.doi.org/10.1107/S0907444995009061.

Kumarevel TS, Karthe P, Kuramitsu S, Yokoyama S. Crystal structure of the arginase from thermus thermophilus, Journal To be Published; http://www.rcsb.org/pdb/explore/explore.do?structureId=2EF4.

Yang H, Czapla TH. 1993. Isolation and characterization of cDNA clones encoding jacalin isolectins. J Biol Chem **268,** 5905-10.

Sharma U, Ahmed N, Krishnasastry MV, Suresh CG. Crystal Structure of Single Chain Recombinant Jacalin Showin Highly Dynamic Posttranslational Excission Loop that Reduce Binding Affinity; Journal: To Be Published; http://www.rcsb.org/pdb/explore/explore.do?structureId=3P8S

Nukaga M, Mayama K, Crichlow GV, Knox JR. 2002. Structure of an extended-spectrum class A beta-lactamase from Proteus vulgaris K1. J Mol Biol **317,** 109-17. http://dx.doi.org/10.1006/jmbi.2002.5420.

López-Lucendo MF, Solís D, Sáiz JL, Kaltner H, Russwurm R, André S, Gabius HJ, Romero A. 2009. Homodimeric chicken galectin CG-1B (C-14): Crystal structure and detection of unique redox-dependent shape changes involving inter- and intrasubunit disulfide bridges by gel filtration, ultracentrifugation, site-directed mutagenesis, and peptide mass fingerprinting. J Mol Biol **386,** 366-78. http://dx.doi.org/10.1016/j.jmb.2008.09.054.

Frick A, Wang Y, Ekvall M, Hallgren K, Hedfalk K, Neutze R, Tajkhorshid E, Törnroth-Horsefield S. 2009. Structural and functional analysis of SoPIP2;1 mutants adds insight into plant aquaporin gating.Nyblom M1. J Mol Biol. **387,** 653-68. http://dx.doi.org/10.1016/j.jmb.2009.01.065.