

## Homology modeling and consensus protein disorder prediction of human filamin

Suresh Kumar\*

Department of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Coimbatore - 641114, Tamil Nadu, India; Suresh Kumar - Email: sureshbio@gmail.com; \*Corresponding author

Received July 12, 2011; Accepted July 14, 2011; Published August 02, 2011

**Abstract:**

Filamins are dimeric actin-binding proteins participating in the organization of the actin-based cytoskeleton. Their modular domain organization is made up of an N-terminal actin-binding domain composed of two CH domains followed by flexible rod regions that consist of 24 Ig-like domains. Homology modeling was used to model human filamin using Modeller 9v5. The resulting model assessed by Verify 3D and PROCHECK showed that the final model is reliable. The conformational disorder prediction of human filamin residues were also mapped on the validated structure of human filamin. Prediction of protein disorder in filamin structures will help understand its role in function.

**Keywords:** human filamin, Calponin CH domain, Ig-like domain, homology modeling, protein disorder

**Background:**

Filamins are large cytoplasmic proteins that crosslink cortical actin into three-dimensional structures and give mechanical force to cells by binding to actin filaments forming bundles or gel networks [1]. Filamins have been reported to interact directly with more than 30 cellular proteins of great functional diversity [2]. In the human filamin there are three isoforms (FLN - A, B and C), in which FLNA and FLNB are expressed ubiquitously and FLNC is expressed in skeletal and cardiac muscles. They are structurally very similar to each other. Each isoform has a relative molecular mass of 280 kDa, consists of an amino-terminal actin binding domain (ABD) and 24 repeated immunoglobulin Ig-like rod domains [3]. The actin-binding domain of filamin is composed of two calponin homology (CH) domains: the amino- and carboxy-terminal CH1 and CH2 domains respectively. All human filamins have two unique long hinges positioned between repeats 15-16 (27 residues) and 23-24 (35 residues) that are postulated to be flexible. But filamin C contains 81 amino acid insertions in repeat 20, not present in filamin A or filamin B [4]. Each CH domain consists of four main  $\alpha$ -helices connected by long loops, and two or three shorter, less regular  $\alpha$ -helices. Three dominant alpha helices form a triple helical bundle, against which the amino-terminal  $\alpha$ -helix packs in a perpendicular orientation [5]. The human filamin rod domains consist of a  $\beta$  sandwich, which resembles the subtype C1 fold of the immunoglobulin family [6]. It is commonly assumed that a protein must attain a stable, folded conformation in order to carry out its specific biological function. However, it was recently shown that several proteins do not assume a well defined and stable three-dimensional structure but are natively unfolded. A disordered protein can be either completely unfolded or comprise of both folded and unfolded segments [7]. Disordered protein regions often lead to difficulties in purification and crystallization of proteins, and become a bottleneck in high throughput structural determination [8]. Therefore, it is quite necessary to identify the disordered regions of target proteins from their amino acid sequences. The prediction of disordered regions is also important for the functional annotation of proteins. The techniques for

predicting conformational disorder are extremely important in structural biology, where they are becoming routine filters in the pipeline of finding suitable targets to be analyzed. In order to visualize predicted disorder features in the context of a filament structure, homology modeled 3D structures of individual domains of human filamin isoforms A, B and C were generated. Consensus conformational disorder residues were mapped onto filamin models to observe the occurrence and nature of disorder.

**Materials and Methodology:**

**Sequence retrieval:**

The FASTA sequence of individual domains of human filamin isoforms A, B and C was obtained from the uniprot database [9]. For each filamin isoforms, 26 sets were constructed for CH1, CH2 and the Ig-like domains 1-24. Domain boundaries were taken directly from the uniprot annotations.

**Sequence Alignment:**

Comparative modeling usually starts by searching the PDB of known protein structures using the 26 individual filamin A, B and C isoforms sequences (target sequence) as the query [10]. This search is generally done by comparing the target sequence with the sequence of each of the structures in the database. The target sequence was searched for similar sequence using the FASTA search tool against Protein Database (PDB) with default parameter. Maximum sequence identity between target and template was selected. Ig-like domains 6 and 8 were ignored since their sequence identity was below 30%. (**Table 1, see supplementary material**).

**Comparative modeling and structure verification:**

Filamin CH and Ig-like domains were modeled using templates composed of homologous Ig-like domains of human and Dictyostelium discoideum filamin (ddFilamin). A template for each target domain was chosen so that both sequence and insertion pattern were similar to the query. Homology models of

filamin domains generally have high sequence-structure compatibility scores. The template and target sequence were carefully aligned to remove potential alignment. The homology modeling of 3D structure was performed by Modeller software version 9v5 [11]. The resulting individual filamin domain model was evaluated by using PROCHECK [12] and energy minimization using Verify3D [13]. The overall stereo chemical quality of the protein was assessed by Ramchandran plot analysis [14]. The structures were visualized through Pymol program.

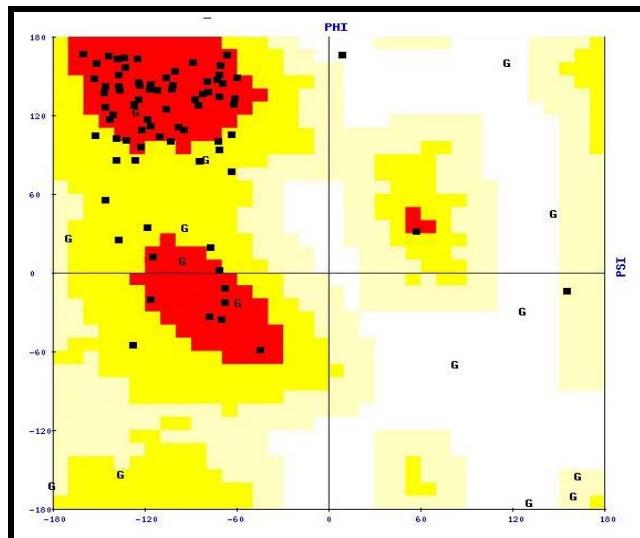
#### Conformational disorder prediction:

For each filamin isoforms A, B and C, protein sequences were obtained from the Uniprot database and the protein disorder were predicted using conformational disorder prediction method developed. The detailed methodology on how the consensus prediction of protein conformational disorder from amino acid sequence was calculated is given elsewhere [15].

#### Results and Discussion:

##### Comparative modeling of human filamin:

The FASTA sequences of human filamin isoforms A, B and C were obtained from Uniprot. This sequence was queried against the PDB using FASTA tool. Comparative modeling predicts the 3-D structure of filamin A, B and C isoforms model of given protein sequence (target), based primarily on its alignment to the template (structure determined experimentally). The model was also checked for  $\phi$  and  $\psi$  torsion angles using the Ramchandran plots. A comparison of the results shows that all the models expect Ig-like domain 6 and 8 models generated by Modeller program are more acceptable. An example of Ramachandran plot for filamin A of Ig-like domain 3 is show in Figure 1.



**Figure 1:** Ramachandran Plot of filamin A of Ig-like domain 3. The plot calculation was done with PROCHECK program

##### Conformational protein disorder prediction:

The prediction of conformational disorder for individual domains of filamin was done with the consensus method developed [15]. The predicted disordered residues were mapped on the homology models of filamin domains as discussed below.

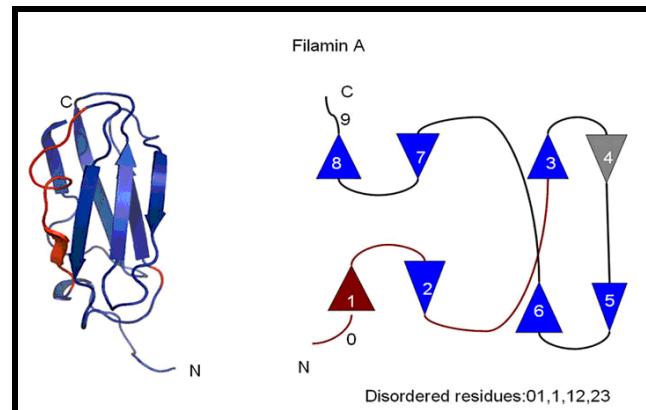
##### (i) Mapping disorder residues on CH domain structures:

The topology diagram of filamin isoforms of CH domain consists of four main  $\alpha$ -helices (1, 2, 3, 4) connected by loop regions (0-1, 1-2, 2-3, 3-4) that might contain two or three shorter helices. Three dominant helices form a parallel bundle against which N-terminal helix packs at a right angle. It can be seen that disordered residues were predicted for CH1 in the loop region 2-3 and for CH2 domain between the helix 2 and loop 3-4.

##### (ii) Mapping disorder residues on Ig-like domain structures:

The topology diagram of Ig-like domain presents an immunoglobulin-like fold made up of seven  $\beta$ -strands organized in two beta sheets giving a  $\beta$ -sandwich. The first  $\beta$ -sheet consisted of strands 1, 2, 5 and 6. The second  $\beta$ -sheet consisted of strands 3, 4, 7 and 8. Only in some of filamin isoforms strand 4 was present. The loops connecting different betastrands were 0-1, 1-2, 2-3, 3-4,

4-5, 5-6, 6-7 and 8-7 (Figure 2). The predicted disordered residues were marked on homology models of Ig-like domain 3. From the topology diagram, we can infer that disordered residues tend to be in the loop region (between loops 1 and 2) and at the N-terminus. Among filamin protein, Ig-like domain 15 of filamin A and B were predicted to be ordered, whereas Ig-like domain 15 of filamin C was predicted to be disordered. Also Ig-like domain 24 of filamin A and C were predicted to be completely ordered, whereas Ig-like domain 24 of filamin B was predicted to be partially disordered. For e.g. as shown in Figure 2, predicted disordered residues were marked on homology model filamin A of Ig-like domain 3. From the topology diagram, we can infer that disordered residues tend to be in the loop region (between loop 1 and loop 2) and at the N-terminus (Table 2, see supplementary material).



**Figure 2:** Disordered residues marked on the homology model of filamin A of Ig-like domain 3. Ig-like domain-  $\beta$ -sheets organized in two beta sheets giving a  $\beta$ -sandwich. First  $\beta$  -sheet consists of strands 1, 2, 5 and 6. The second  $\beta$  -sheet consists of strands 3, 4, 7 and 8. Strand 4 is present only in some filamin isoforms. The loop connecting different beta-strand is 0-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6-7 and 8-7.

##### Conformational disorder in the experimental filamin PDB entries:

The conformation disorder status in experimentally solved structures deposited in the PDB was studied. The conformational disordered status was inferred from the missing residues in the REMARK 465 entries. In filamin A, Ig-like domain 17 along with GPIB alpha cytoplasmic domain complex (PDB ID: 2BP3) and the disordered residues were in the loop 0-1 and in the loop 8-9. In filamin A, Ig-like domain 21 complexed with MIG FILIN peptide (PDB ID: 2WOP) and the conformation disordered were found to be in the loop 0-1, loop 5-6 and loop 8-9. In filamin C, Ig-like domain 23 (PDB ID: 2NOC) and the conformational disordered residues were found to be in the loop 0-1. The conformational disorder residues were only observed either in the N-terminus (loop 0-1) or at the C-terminus (loop 8-9) of Ig-like domain of human filamin. Despite the paucity of the data, they agreed quite well with the predictions described above.

#### Conclusion:

The consensus conformational disorder prediction approach was performed on the modeled filamin proteins to know the occurrence of disorder residues. The results suggested that there were more occurrences of disorder residues in N-terminus and occurrence of order residues on C-terminus of Ig-like domain. It also suggested that more disorder residues can be found between loop 1-2 and loop 3-4 of Ig-like domain. In CH domain of filamin isoforms, the disorder residues were predicted between the loop 2 and 3. The conformational disorder status studied from the experimentally determined filamin structures agreed quite well with the prediction. Mapping of Predicted disorder residues in homology model filamin protein will give valuable insights for target selection and the design of constructs, particularly in structural biology and structural genomics projects.

#### Acknowledgements:

This work was supported by the Austrian GEN-AU project BIN-II.

#### References:

- [1] van der Flier A & Sonnenberg A. *Biochim Biophys Acta*. 2001 **1538**: 99 [PMID: 11336782]
- [2] Feng Y & Walsh CA. *Nat Cell Biol*. 2004 **6**: 1034 [PMID: 15516996]

- [3] Popowicz GM *et al.* *Trends Biochem Sci.* 2006 **31**: 411 [PMID: 16781869]  
[4] Xie Z *et al.* *Biochem Biophys Res Commun.* 1998 **251**: 914 [PMID: 9791010]  
[5] Stossel TP *et al.* *Nat Rev Mol Cell Biol.* 2001 **2**: 138 [PMID: 11252955]  
[6] Bork P *et al.* *J Mol Biol.* 1994 **242**: 309 [PMID: 7932691]  
[7] Fink AL. *Curr Opin Struct Biol.* 2005 **15**: 35 [PMID: 15718131]  
[8] Oldfield CJ *et al.* *Proteins.* 2005 **59**: 444 [PMID: 15789434]  
[9] Apweiler R *et al.* *Nucleic Acids Res.* 2004 **32**: D115 [PMID: 14681372]  
[10] Westbrook J *et al.* *Nucleic Acids Res.* 2002 **30**: 245 [PMID: 11752306]  
[11] Sali A *et al.* *Proteins.* 1995 **23**: 318 [PMID: 8710825]  
[12] Laskowski RA *et al.* *J Biomol NMR.* 1996 **8**: 477 [PMID: 9008363]  
[13] Bowie JU *et al.* *Science.* 1991 **253**: 164 [PMID: 1853201]  
[14] Ramachandran GN *et al.* *J Mol Biol.* 1963 **7**: 95 [PMID: 13990617]  
[15] Kumar S & Carugo O. *Open Biochem J.* 2008 **2**: 1 [PMID: 18949069]

Edited by P Kangueane

Citation: Kumar. Bioinformation 6(10): 366-369 (2011)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

## Supplementary material:

**Table 1:** Homology Models of human isoforms. Table shows template taken from PDB and sequence identity (%) between target and the template.

Domain (target)	Template			TEMPLATE -PDB ID			Sequence Identity (%)		
	A	B	C	A	B	C	A	B	C
CH1	CH	CH	CH	2eyi	2eyi	2eyi	43.9	43.9	43.9
CH2	CH	CH	CH	2eyi	2eyi	1wku	26.7	24.7	27.7
Ig-like domain 1	F14_C	F14_C	F14_C	2d7m	2d7m	2d7m	31.3	32.3	36.3
Ig-like domain 2	F14_C	F14_C	F14_C	2d7m	2d7m	2d7m	38.3	38.7	37.3
Ig-like domain 3	F12_B	F12_B	F15_B	2dic	2dic	2dmb	38.5	36.0	37.1
Ig-like domain 4	F9_B	F9_B	F9_B	2di9	2di9	2di9	51.6	45.1	50.5
Ig-like domain 5	F14_C	F14_C	F14_C	2d7m	2d7m	2d7m	41.2	38.1	41.2
Ig-like domain 6	F13_B	F13_C	F14_C	2dj4	2dj4	2d7m	26.2	27.1	33.0
Ig-like domain 7	F14_B	F13_B	F13_B	2e9j	2dj4	2dj4	41.4	43.4	39.3
Ig-like domain 8	F11_B	F22_B	F12_B	2dib	2eeb	2dic	27.0	27.0	28.1
Ig-like domain 9	F9_B	F9_B	F9_B	2di9	2di9	2di9	75.2	100	82.7
Ig-like domain 10	F10_B	F10_B	F10_B	2dia	2dia	2dia	51.5	96.8	63.1
Ig-like domain 11	F11_B	F11_B	F11_B	2dib	2dib	2dib	65.0	100	60.0
Ig-like domain 12	F12_B	F12_B	F12_B	2dic	2dic	2dic	65.6	97.8	69.8
Ig-like domain 13	F13_B	F13_B	F13_B	2dj4	2dj4	2dj4	68.0	96.8	71.8
Ig-like domain 14	F14_C	F14_B	F14_C	2d7m	2e9j	2d7m	76.2	100	95.8
Ig-like domain 15	F15_B	F15_B	F15_B	2dmb	2dmb	2dmb	66.6	100	59.5
Ig-like domain 16	F16_C	F16_B	F16_C	2d7n	2ee9	2d7o	64.3	91.9	81.1
Ig-like domain 17	F17_A	F17_B	F17_C	2aav	2eea	2dmc	97.7	100	97.8
Ig-like domain 18	F18_B	F18_B	F18_B	2dmc	2dmc	2dmc	61.7	100	64.3
Ig-like domain 19	F19_B	F19_B	F19_B	2di8	2di8	2di8	71.1	98.9	68.8
Ig-like domain 20	F20_B	F20_B	F20_B	2dlg	2e9i	2dlg	60.6	97.8	69.8
Ig-like domain 21	F21_B	F21_B	F21_B	2ee6	2ee6	2ee6	82.7	98.9	84.9
Ig-like domain 22	F22_B	F22_B	F22_C	2eeb	2eeb	2d7p	79.7	97.8	98.9
Ig-like domain 23	F23_B	F23_B	F23_C	2eec	2eec	2nqe	75.2	100	97.8
Ig-like domain 24	F24_B	F24_B	F24_C	2eed	2eed	1vo5	68.4	97.8	96.8

**Table 2:** Frequency of occurrence of conformational disorder in the 24 segments of the Ig-like domains of human filamin (percentage)

Segment	Filamin Isoforms			
	All	A	B	C
N-terminus	7	8	4	8
Strand 1	6	5	5	10
Loop 1-2	45	35	43	57
Strand 2	6	4	8	4
Loop 2-3	21	21	33	8
Strand 3	18	17	29	8
Loop 3-4	35	33	50	21
Strand 4	15	0	40	0
Loop 4-5	4	4	4	4
Strand 5	11	9	14	9
Loop 5-6	25	29	21	25
Strand 6	17	21	12	17
Loop 6-7	17	17	17	17
Strand 7	14	12	12	17
Loop 7-8	14	12	12	17
Strand 8	0	0	0	0
C-terminus	0	0	0	0