In silico Study of tertiary structure and binding interaction of Profilin – *Schizosaccharomyces pombe*

Gangawane A.K.[1,*], Gomase V.S.[1], Anjali R.[1] and Patil C.S.[2]

[*1]Department of Biotechnology and Bioinformatics, Padmashree Dr. D. Y. Patil University, Sector 15, CBD Belapur, Navi Mumbai-400614, Email- ajganga@yahoo.com

[2]B V Bhoomraddi College of Arts, Sci and Com, Bidar, Dist- Bidar, Karnataka, India-585403

Abstract-Profilin plays a crucial role by directly linking actin cytoskeleton dynamics to various signal transduction pathways. Profilin has been well conserved during the evolution and can be found throughout the animal and plant kingdom. All known profilin are defined by common structural and biochemical properties and are known to interact with at least three types of ligands: they are complex with the G actin and actin – related proteins; interact with PLP presented either as a peptide or as a sequence motif within specific proteins (with the exception of vaccinia profilin), and bind to polyphosphoinositides. Profilin has different functional roles: Actin binding and membrane refluxes, regulation of phospholipase C-Y1, focal contacts, nuclear transport of proteins, allergen, and tumor suppression. A protein function is tightly linked to its tertiary structure and as a residue located far apart in the sequence can be very close in space, and only few residues are responsible for a proteins function, insights into the tertiary structure of a protein can represent a key component of the functional analysis process. The homology or comparative protein structure of profilin was obtained by optimization of molecular probability density function (pdf) using the software program MODELLER. A stable structure was obtained by fixing the atomic charges by CHARMM 22 force field, and minimizing the energy by conjugate gradient method (1000 iterations) using NAMD software. Protein- protein docking was carried out with Vegazz for profilin dimer formation and profilin –ligand interactions were studied with Auto dock software programs. The *in silico* studies help in understanding the profilin monomeric interaction and its effect on PLP binding and polyphosphoinositide interactions, the two important processes which regulate number of signaling event.

Keywords: Profilin, Tertiary structure, docking studies, *Schizosaccharomyces pombe*

List of Abbreviations: Poly (L-proline) - PLP, Probability density function- pdf, Actin binding proteins –ABPs, Vasodilator stimulated phosphoprotein- VASP, Structural classification of protein- SCOP

Introduction:

Profilin has been well conserved during the evolution and can be found throughout the animal and plant kingdom[1]. They are found in protozoa [2, 3, 4] echinodermata, insect [5] and mammalian [6, 7]. The plant profilin was first identified as an allergen from Birch pollen [8, 9]. Profilin also plays a crucial role by directly linking actin cytoskeleton dynamics to various signal transduction pathways. Though profilin a small protein it has amazingly diverse functions in the cells that continue to elude the understanding of researchers even today, after two decades of its discovery. Profilin was among the first actin binding protein (ABPs) which is involved in regulation, organization and orchestration of the dynamic system[6]. It was found to interact with the membrane phosphoinositides and poly (L-proline) – PLP domain containing proteins such as VASP family protein, N-WASP and diaphanous[10]. Ena (Drosophila enabled), mena (Mammalian Ena), aczonin, drebrin I [11] p^{85} a subunit of P-13 kinase [12] etc. Many other proline rich have been identified as profiling ligands, such as a forming homology proteins, the actin related protein - 2/3 complex 9 Arp 2/3complex). Recently profiling is also shows to play a role in tumor suppression [13].All the known profiling is defined by common structural and biochemical properties and is known to interact with at least 3 types of ligands: the complex with G-actin [14, 15, 16 &17] and actin related protein [18] interact with PLP presented either as a peptide or as a sequence motifs within the specific proteins and binds to phosphoinositides [19, 20]. Profilin comprise 124 to 153 amino acid sequences although amino acid sequences of the common isoforms in distantly related species might show less than 25 % identify. Studies on profilin from diverse sources show that they possess closely related tertiary structure [21]. The sequence alignment of plant profilin show that they are linearly ranging from 131 to 137 amino acids. The X ray structure determination of human and bovine profiling I showed that it forms two major contacts with actin in the crystal. The binding of profiling to ligands might provide a means of linking different pathways, by mechanism that remains unclear, to cytoskeleton dynamics (Figure 1). The first proline rich protein to be identified as profiling ligands is vasodilator stimulated phosphoprotein (VASP) [22]. It is an *in vitro* substrate of protein kinase C and P14m 5P2 is the effective activator of profiling phosphorylation by PKC [23].The profilin molecule is unique in having positive and negative effects on polymerization of actin. It also helps in polymerization and depolymerization of actin [23,24,25,26]. Profilin is shown to interact with the scaffolding proteins in neurons [27]. It is an important plant allergen [8, 9]. PFN1 and PNF2 are present in

many organisms, of which the ubiquitously expressed PFN1 is considered for normal cellular functions [28, 13].The many neurospora has provided insights into biological processes have been enumerated briefly [29, 30]. The entire genome of the species contains 38, 69,769 base pairs of DNA encode: 10,082 proteins (9,200 of them longer ect than 100 amino acid) 424 tRNAs, 74 rRNAs, 175-200 copies of the 25S/17S/ 5.8S gene cluster [31].

Materials and Methods

Tools for *in silico* analysis

Sequence analysis

In this protocol we analyzed the sequence of profiling from *Neurospora crassa*.

Multiple sequence alignment

We used dynamic programming for analysis of conserved region from set of gene which is similar in diversified species.

MSWQAYVDSSLVGTGHIDKAAIISAAGDSTWAATP
GFTLSPDEMKFISAVLGDNGKGSNVDKVFAEGLHV
AGQRYVAFNIEGRHVYRQGKTGVIIVKTTQAILVAHY
GENAVAGNSTQTVEALADYLVKAGY (GenBank 85106701)

Structure analysis

The 3D SM server takes the sequence of the protein which is interesting and attempts to predict its 3D – structure and also its function. The probe is scanned by 3D-PSSM library using the global dynamic programming algorithm. The amino acid sequence of the protein of interest is submitted in FASTA format with a one line description about the protein. An email ID is indispensable as the result are mailed

MSWQAYVDSSLVGTGHIDKAAIISAAGDSTWAATP
GFTLSPDEMKFISAVLGDNGKGSNVDKVFAEGLHV
AGQRYVAFNIEGRHVYRQGKTGVIIVKTTQAILVAHY
GENAVAGNSTQTVEALADYLVKAGY (GenBank 85106701)

The server generates the results with an in depth analysis. The browser comes up with proteins that are very similar to the query sequence after searching the non redundant protein databanks. The secondary structure is predicted by using PSI-Blast. The matching of the sequences the server comes up with the list of proteins with similar domain and also classifies the protein in to the super family based on its properties using SCOP- (Structural classification of protein) by scanning the library.

Spatial restraints

MODELLER [32] is a computer program that models three dimensional structures of proteins and their assemblies by satisfaction of spatial restraints. MODELLER is most frequently used for homology or comparative proteins structure modeling. The user provides an alignment of a sequence to be modeled with known related structure and MODELLER will automatically calculate a model with all non hydrogen atoms.

There are three kinds of input files protein data Bank atom files with co-ordinates for the template structure, the alignment file with the alignment of the template structures with the target sequence and MODELLER commands in script files that instruct MODELLER what to do.

Each atom file is named code.atm where code is a short protein code –PDB code while 1ypr.atm the code must be used as that proteins identifier throughout the modeling. The alignment is done in PIR format. MODELLER is command line only tool and has no graphical user interface; MODELLER commands

Stereo chemical analysis

PROCHECK suite of programs [33] for assessing the stereo chemical quality of a given protein structure. The aim of PROCHECK is to assess how normal or conversely how unusual, the geometry of the residues in a given protein structure is as compared with stereo chemical parameters derived from well refined high resolution structures. The parameters are to be described in details [34].

PROCHECK is the PDB file[35] holding the co-ordinates of the structure of interest. PROCHECK generates a number of output files in the default directory which have the same name as the original PBD file but with different extensions.

Molecular optimization

The program Auto Dock was developed to provide an automated procedure for predicting the interaction of ligands with biomolecular targets. The motivation for this work arises from problems in the design of bioactive compounds and in particular the field of computer aided drug design. Bio macromolecule: - *Neurospora crassa* profilin and Ligands: - Poly-L – proline

Results & Interpretation

The template selection was carried out on basis of sequence and secondary structure similarity, (Table 1) shows the results of the search programs mentioned above and their respective statistical scores that were considered in selecting the best template. A systematic representation of the multiple sequence alignment of target *Neurospora crassa*

profilin sequence with their respective template proteins and their secondary structure details are shown respectively. *Neurospora crassa* profilin share 43% similarity with *Saccharomyces cerevisiae* (Table 1 & Figure 2). The 3D structure model of *Neurospora crassa* profilin was built by using MODELLER 9v1 tool. The few amino acids which were in disallowed region of Ramchandran plot were also modeled (Table 2 &3). The final model shows an improved Ramchandran plot and energy profiles with respect to the initial models (Table 4). This cycle of loop building minimization and evaluation was done till no further improvement in the structure can be made (Figure 3 and 4). The improved model was evaluated using programs PROCHECK and Verify 3D (Figure 5). The dimer was obtained by docking two monomer of *Neurospora crassa*. Dimers were screened based on the binding affinities and the top ranked dimer was selected with good steric fit. It was further subjected to energy minimization with the steepest descent algorithm for optimization in Swiss PDB viewer. The electrostatic energy of the complex was found to be -6394 KJ/mol. The binding free energy was found to be -9.924380Kcal/mol. Docked complex ranked 6 (based on electrostatic term) was found to be more close to the biological information available for the dimer formation (Figure 6). The crystal structure of human profilin tetramer is available. The tetramer was obtained by docking two dimer of *Neurospora crassa* (Figure 7). A model of *Neurospora crassa* profilin PLP complex was obtained by docking *Neurospora crassa* profilin monomer with PLP by using AUTO DOCK. PLP structure was obtained from crystal structure of human profilin I-PLP complex. The complex structure obtained after energy minimization was used for further analysis (Figure 8).

Conclusion

Profilin has been of great interest to biophysicists owing to its abundance in the cell and lot of structural data is available on profilin monomer. The structural similarity between *Neurospora crassa* profilin and human profilin is forms was between 20-32%. *In silico* model of *Neurospora crassa* and *Saccharomyces cerevisiae* profilin showed 43% identical. The model was evaluated using programs Procheck and verify3D. Solvent pockets were generated by using POPS server. Oligomers of *Schizosaccharomyces pombe* profilin were also obtained, Monomer-Monomer docking; to form Dimer and again two dimer were docked together to obtain tetramer (using Vega ZZ dock). *In-silico* analysis was compared with human profilin data *Schizosaccharomyces pombe* Profilin also contain conserved residue also known as “Profilin motif” TRP3 ASN4 ALA5 TYR6. The Schizosaccharomyces pombe Profilin dimer obtained with docking shows during dimer formation “N” and “C” terminal are in accessible with PLP same as previously obtained result for human and bovine profilin [36]. Slime profilin exists in three forms monomer, dimer and tetramer similar to human profilin. It also exhibits in vitro oligomerization and only tetramer/ monomer profilin binds with PLP. Dimer did not shows binding to PLP due to inaccessible PLP Binding region.

Acknowledgement

We are thankful to the Department of Biotechnology and Bioinformatics, Padmashree Dr. D. Y. Patil University, Navi Mumbai for providing the financial and technical support.

References

1. Markey, F., Lindberg, U., and Erikson.L. (1978) *FEBES letter* 88:75-79
2. Reichstein, E., and Korn, E.D.(1979)*J Biol Chem*. 254: 6174-6179
3. Ozaki, K., Sugino, H., Hasegawa, T., Takahashi, S., Hatano, S. (1983) *J Biochem*. 93: 295-298
4. Haugwitz, M., Noegel, A., Reider, R d., Lotispeich, P., and Schleicher, M.(1991) *J.Cell Science* 100: 481-489
5. Cooley, L., Verheyen, E., and Ayers,K.(1992) *Cell* 69: 173-184, *Trends Cell Biol* 13: 435-446
6. Carlsson,L., Nystrom, L.E., Lindberg, U., Kannan K K., Cid- Dresdner, H., Lovergren, S., and Jornvall, H (1976) *J.Mol Biol* 105: 353-366
7. Kwiatkowski, N., and Bruns, G.A. (1988) *J Biol Chem* 263: 5910-5915
8. Valenta, R., m., Breitenbach, M., Pettennburger, K., Koller, L., Rumpold, H., Scheiner, O., and Kraft, D. (1991) *Int Arch Allergy Appl Immunol* 94: 368-370
9. Valenta, R., Duchene, M., Eber, C., Valent, P., Sillaber, C., Deviller,P., Ferreira, F., Tajkl,M., Edelmann, H., Kraft, D., and Scheiner, O. (1992) *J Exp Med* 175: 377-385
10. Holt M,.R., Koffer, A (2001) *Trends Cell Biology* 11: 38-46
11. Witke, W., Podtelejnikov, A, V., Di Nardo, A., Sutherland, J.D., Gurniak, C.B., Dotti, C & Mann, M (1998) *EMBO J* 17: 967-976
12. Surya S. Singh, Abha Chauhan, Noriiko Murakami, James Styles, Marshall Elzinga

and Ved P.S., Chauhan (1996) *Receptors and Signal transduction* 6(2) 77-86

13. Janke, J., Schluter, K., Jandrig, B., Theile, M., Kolble, K., Arnold, W., Grinstein E., Schwartz, A., Estevez-Schwarz, L., Schlag, P M (2000) *J Exp Med* 191: 1675-1686

14. Carlsson,L., Nystrom, L.E., Sundkvist, I., Markey, F., Lindberg, U. (1977) *J Mol.Biol* 115 : 465-483

15. Tseng, P.C., and Pollard, T.D (1982) *J Cell Biol* 94 : 213-218

16. Tobacman, L.S., Brenner, S.L., Korn, E.D (1983) *J Biol Chem* 258: 8806-8812

17. Schutt, C.E., Lindberg, U., Myslik, J., Strauss, N., (1989) *J Mol Biol* 209: 735-746

18. Machesky, L., Cole, N., Moss, B., and Pollard, T.,(1994) *Biochemistry* 33: 10815-10824

19. Lasssing, I., and Lindberg, U. (1985) *Nature* 314: 472-474

20. Lasssing, I., and Lindberg, U. (1985) *J Cell Biochem* 37: 255-267

21. Fedorov, A.A., Pollard, T.D., and Almo, S.C (1994) *J Mol Biol* 241: 480-482

22. Reinhard, M., Giehl, K., Abel. K., Haffner, C., Jarchau, T., Hoppe, V., Jockusch, B.M., and Walter U (1995) *EMBO J* 14: 1583-1589

23. Hansson, A., Goran Skoogland, Ingrid Lassing, UnoLindberg and Magnus Ingelman Sundberg (1988) *Biochemical and Biophysical Research Communication* 150(2) : 526-531

24. Pantaloni, D., Carlier, M F (1993) *Cell* 75: 1007-1014

25. Pollard, T.D (1984) *J Cell Biol* 99: 769-777

26. Pring , M., Weber, A., and Bubb, M,R (1992) *Biochemistry* 31: 1827-1836

27. Mammoto, a., Sasaki, T., Asakura, T., Hotta, I., Imamura, H., Takahashi. K., Matsuura, Y.,Sirao, T., and Takai, Y., (1998) *Biochemical and Biophysical Research Communication* 24: 86-89

28. Wittenmayer N., Burkhard Jandrig, martin Rothkegel, Karthrin Schluter, Wolfgang A, Wolfgang H, Siegfried Scherneck and Brigitte M. Jockusch (2004) *Molecular biology of the Cell* 15: 1600-1608

29. Horowitz N H.(1991) *Genetics* 127: 631-636

30. Perkins.D.D (1992) *Genetics* : 130: 687-701

31. Giraldo R (2003) *Nat Struct Biol* 10: 565-571

32. M.A. MartiRenom, A. Stuart, A. Fiser, R, Sanchez, F. Melo, A. Sali (2000) *Annu Rev. Biophys Biomol* 29: 291-325

33. Laskowski R A., Mac Arthur M W., Moss D S & Thornton J M (1993) *J Appl Cryst* 26: 283-291

34. Morris A L., Mac Arthur M W., Hutchinson E G & Thornton J M (1992) *Proteins* 12: 345-364

35. Breitwieser , G E (2004) *Circ Res* 94: 17-27

36. Mahoney, NM., Janmey, PA., and Almo, S.C.(1997) *Nat Struct Biol* 4: 953-960

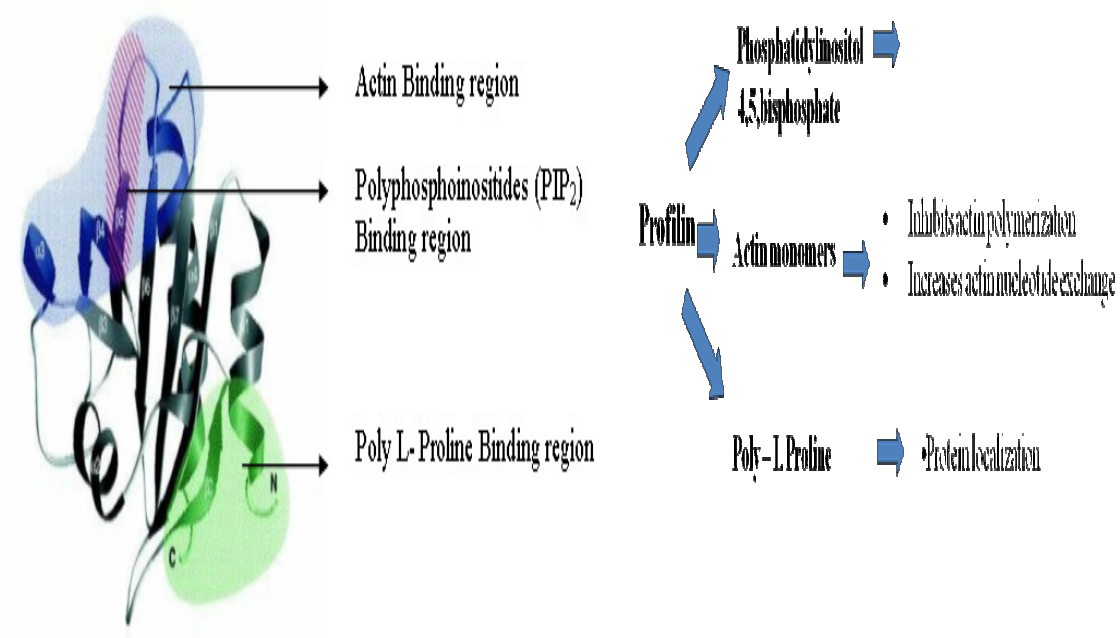


Figure 1: - Structure and Function of profilin

Table 1: - Sequence of profilin from various sources and their amino acid residues

Sequence	Organism	Amino acid residues
Sequence 1	*Schizosaccharomyces pombe*	127
Sequence 2	*Saccharomyces cerevisiae*	126
Sequence 3	*Neurospora crassa*	133
Sequence 4	*Rattus norvegicus*	140
Sequence 5	*Homo sapiens* Profilin 1	140
Sequence 6	*Homo sapiens* Profilin 2	140
Sequence 7	*Homo sapiens* Profilin 3	129
Sequence 8	*Homo sapiens* Profilin 4	137
Sequence 9	Bovine Profilin 1	140
Sequence 10	*Arabidopsis thaliana*	131
Sequence 11	*Nicotiana tabacum*	134
Sequence 12	*Aspergillus niger*	131
Sequence 13	*Variola virus*	133

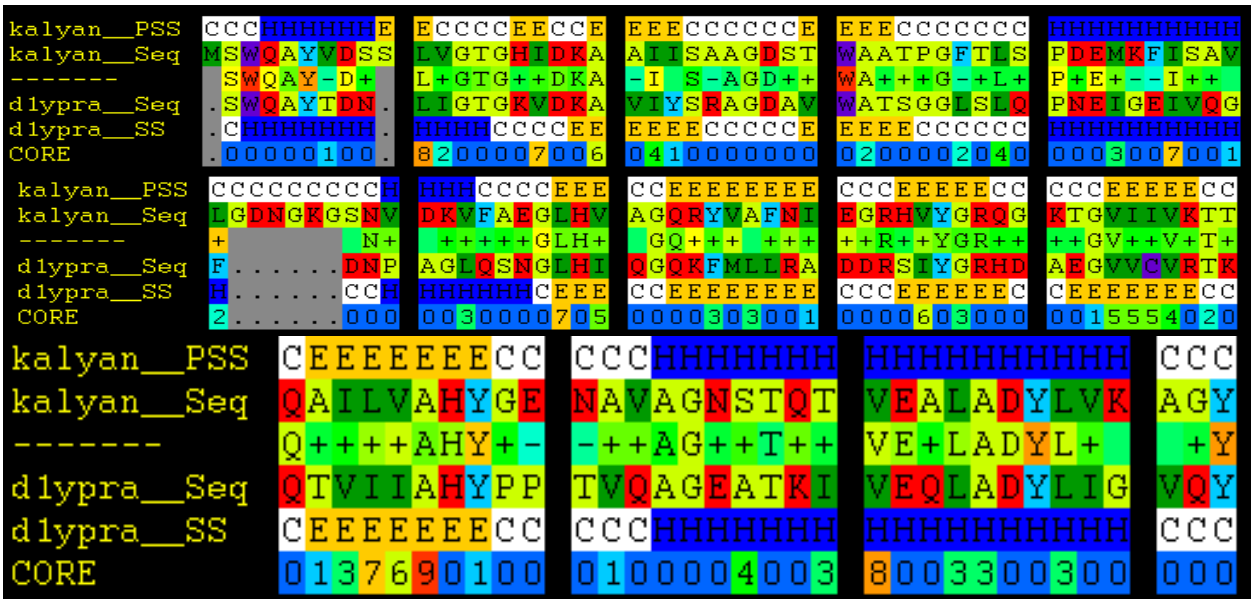


Figure 2: - 2D Confirmation generated from 3D PSSM Server

Figure 3: - *Schizosaccharomyces pombe* profilin was built by using Modeller 9v1 tool.

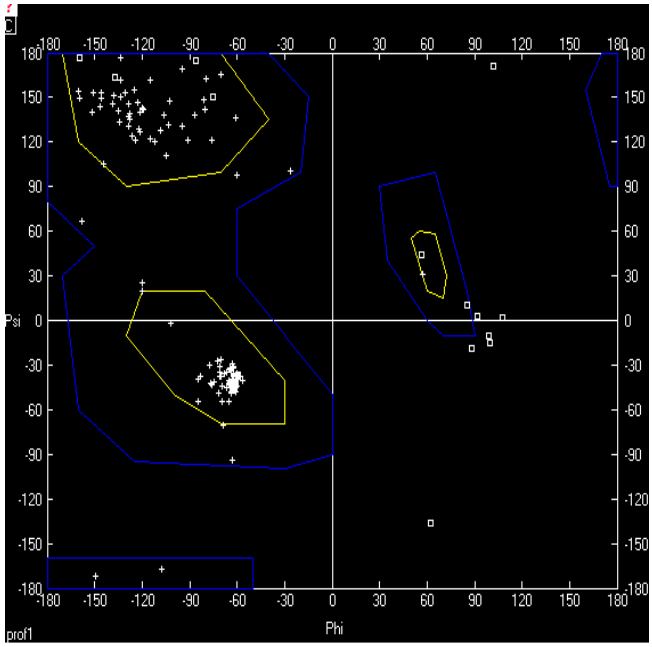


Figure 4: - Validation of protein structure using Ramchandran plot

Table 2: - Ramachandran Plot Statistics

Ramachandran Plot statistics

Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A,B,L].

		No. of residues	%-tage
Most favoured regions	[A,B,L]	105	92.9%
Additional allowed regions	[a,b,l,p]	8	7.1%
Generously allowed regions	[~a,~b,~l,~p]	0	.0%
Disallowed regions	[XX]	0	.0%
Non-glycine and non-proline residues		113	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		16	
Proline residues		2	
Total number of residues		133	

Table 3:- G-Factors

G-Factors

Parameter	Score	Average Score
Dihedral angles:-		
Phi-psi distribution	.27	
Chi1-chi2 distribution	-.04	
Chi1 only	.11	
Chi3 & chi4	.10	
Omega	-.13	
		.06
Main-chain covalent forces:-		
Main-chain bond lengths	-.24	
Main-chain bond angles	-.29	
		-.27
OVERALL AVERAGE		-.07

G-factors provide a measure of how **unusual**, or out-of-the-ordinary, a property is.

Values below -0.5* - unusual

Values below **-1.0**** - highly unusual

Important note: The main-chain bond-lengths and bond angles are compared with the Engh & Huber (1991) ideal values derived from small-molecule data. Therefore, structures refined using different restraints may show apparently large deviations from normality.

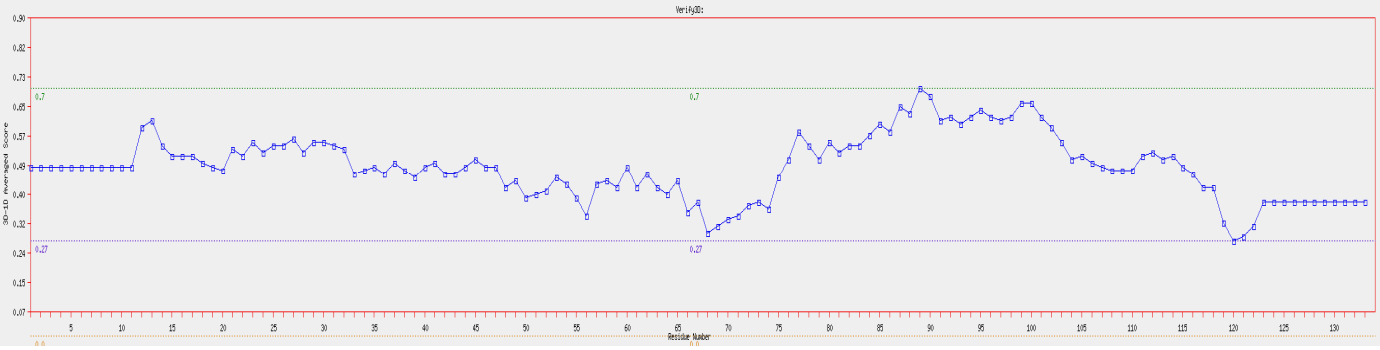


Figure 5: - Verify 3D

Table 4: - Solvent Pockets Generated by POPS Server

Three solvent pockets identified for the given protein PROF.pdb

First solvent pocket 4Ang

LYS19 ALA20 PHE37 LEU39 GLU43 PHE46 ILE47 VAL63 GLU66 GLY67 LEU68 HIS69 VAL70 TYR75 GLY93 VAL94 ILE96 VAL105 ALA106 HIS107

Second solvent pocket 4Ang

ALA21 ILE22 ILE23 THR30 PHE37 LEU39 ILE47 ILE96 ILE103 VAL105

Third solvent pocket 4Ang

TRP3 ILE22 TRP31 VAL97 THR99 GLN101 ALA102 LEU104 VAL121 LEU124 ALA125 LEU128 TYR133

```
            Total SASAs of Molecule :
prof1.pdb

      Total Hydrophobic SASA =      4043.7    Angs^2
      Total Hydrophilic SASA =      2837.9    Angs^2
      Total SASA             =      6881.6    Angs^2
```

Figure 6: - Dimer structure from two monomers of *Schizosaccharomyces pombe*.

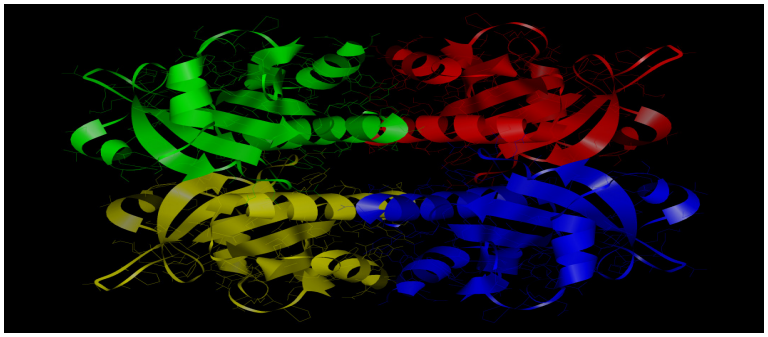
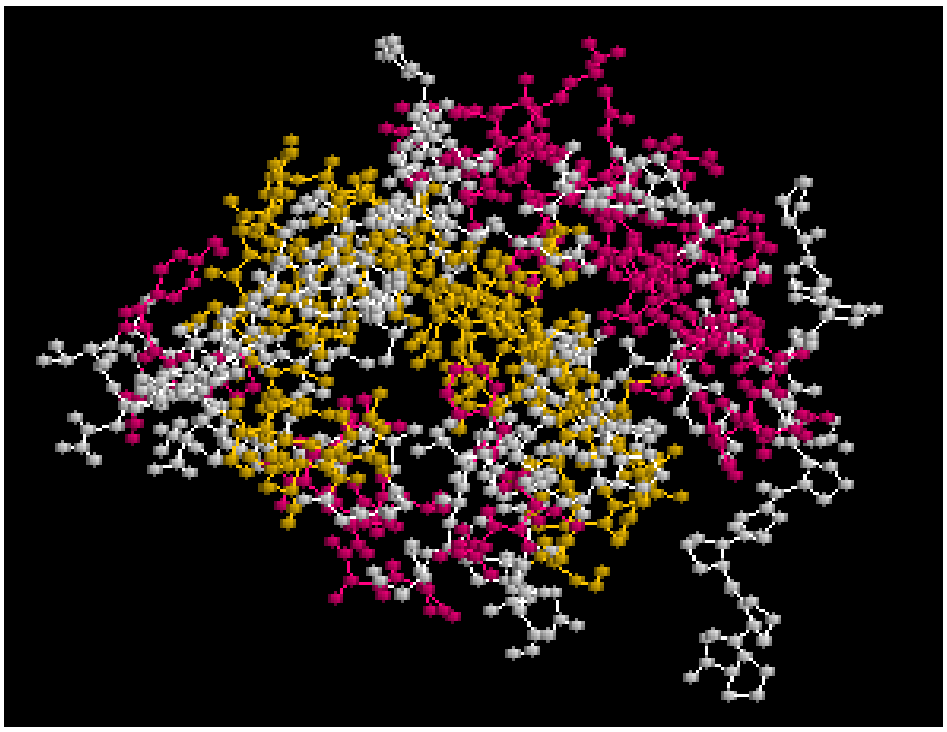


Figure 7: - The tetramer structure from two dimers of *Schizosaccharomyces pombe*



Poly –L – Proline (PLP) Binding

Figure 8: - Complex of *Neurospora crassa* profilin monomer with PLP by Using AUTO DOCK.