International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

„Dinosaur Collagen" Computational Structure Modelling and Docking Analysis with α1 (I) Integrin Domain

B. Sai Shankar[1], Dr. C. Joyce Priyakumari[2]

[1]Department of Zoology, Madras Christian College, hennai-600059, Tamil Nadu, India

[2*]Corresponding Author, Assistant Professor, Department of Zoology, Co-coordinator, Bioinformatics Infrastructure Facility (BIF), BTIS Net, Madras Christian College, Chennai-600059, Tamil Nadu, India

Abstract: ***Three peptides from two unique collagen sequences (α1 and α2 chains) of two dinosaur species i.e. Tyrannosaurus rex and Brachylophosaurus canadensis had previously been found to overlap with the integrin binding site on the rat collagen microfibril structure and was reported to be a significant non-random distribution suggesting mechanisms of protein survival over long geologic periods. The sequences of these peptides featuring certain conserved residues as was revealed by analyzing multiple alignments from different organisms, functional importance of recognition motif and the nature of collagen itself favoring conservation, were used to computationally model the triple helix structure and the resultant structure was docked with the crystal structure of α1 (I) Integrin domain (PDB Code-2M32). Similar patterns of interaction were observed between the receptor i.e. α1 (I) Integrin domain and each of the different ligands i.e. modelled collagen structure of the two dinosaur species and the synthetic collagen mimetic peptide (present in the original complex of 2M32).***

Keywords: Dinosaur, collagen, docking, integrin, modelling

1. Introduction

The retrieval of bio-molecular information of evolutionary value from fossils is a coveted yet often elusive goal. Although both proteins and DNA are components of the tissues of all organisms, they undergo degradation over geological time. The time range over which meaningful protein and DNA sequence information is preserved depends on a variety of variables such as the environmental temperature, humidity and the type of matrix in which the molecules are contained [1]. It is widely accepted that proteins have the potential to survive significantly longer periods of time than DNA [2]. Asara *et al.* (2007) [3] reported the sequencing of proteins from 68-million-year-old *Tyrannosaurus rex* (Museum of the Rockies-MOR 1125)fossil and established similarities between dinosaur and chicken genomes. The authors generated seven *T. rex* peptides by matching mass spectra against collagen proteins. Collagen peptide sequences were subsequently derived from a second dinosaur, *Brachylophosauraus canadensis* (MOR 2598) [4], and included many of the earlier lines of supporting evidence as well as independent replication of data in multiple labs.

The dinosaur collagen peptide sequences that have been determined by Mass spectrometry were mapped onto the 3 dimensional structure of rat collagen micro-fibril and a non-random localization to specific regions within the structure was reported [5]. Out of the many features of peptide distribution reported by them two features were concluded by the workers as statistically significant. One of the feature was a non-random co-localization of three peptides (from two unique sequences of α1 and α2 chains) belonging to the two species *i.e., T. rex* and *B. canadensis,* to a region on the collagen monomer identified to be the integrin binding site that promotes cell- collagen interactions, angiogenesis and osteoblast differentiation.

In the present study the structures for the dinosaur collagen peptides have been computationally modelled and docked with an Integrin structure available from PDB. The primary structure of the collagen itself with its repeating GXY triplets along its entire sequence favors computational structure prediction and analysis possible. An analysis of the docked protein-protein complex is carried out for identifying any similarities in the patterns of interacting residues.

2. Literature Survey

Eleven fossil derived peptide sequences from the two dinosaurs, *T. rex* and *B. canadensis*[3],[4] were mapped on molecular models of extant human and rat collagens [5] and it was found that they were non-randomly distributed in several respects. In particular, fossil sequences mapped to regions of the protein partly shielded by tight molecular packing [6], which may physically stabilize and protect them from enzymatic degradation, thus contributing to their preservation. In fact, the only position where alpha 1 chain peptides [*i.e.,* Peptide „GVQGPPGPQGP" of both *T. rex* and *B. canadensis*] co-localize with an alpha 2 chain peptide [*i.e.,* Peptide GLPGESGAVGPAGPP(I*)GSR of *B. canadensis*/*T.rex**] mapped to the integrin binding site that promotes cell-collagen interactions, angiogenesis, and osteoblast differentiation; its fibril location and association with severe mutations also suggest its crucial nature [7] and hence strong selective pressure for conservation of sequence.

Integrins α1β1 and α2β1 are the major integrin collagen receptors [8]. The α1, α2, and α10 subunits each contain an "inserted" (I) domain near the N terminus. The „I" domains

have been shown to contain a ligand-binding site and a MIDAS motif, which needs to be occupied by an appropriate cation for effective ligand binding by the integrin. Recombinant proteins duplicating these small (approximately 200 amino acids) I domain polypeptide segments effectively bind collagen; presumably it is these regions that are responsible for the integrin‟s binding to collagens [9].

A specific collagen motif recognized by the α2-I domain has recently been identified as the hexapeptide GFOGER (O- 5 hydroxyproline) within the context of a collagen triple helix [10]. Analysis of published data suggests a critical role for the central integrin binding site, GFPGER $^{502–507}$ according to four lines of evidence [7].

3. Materials and Methods

Dinosaur collagen sequences were retrieved from UniProtKB in FASTA format. Three peptides representing four collagen sequences (α_1 and α_2 chains of collagen from each dinosaur species) were identified based on the observation that these peptides overlap with the integrin binding site of the rat collagen micro-fibril comprising the residues 502-510 as reported [5]. The identified peptides were: Peptide 1- „GVQGPPGPQGPR‟- from α_1 type 1 collagen chain of both *Tyrannosaurus rex* and *Brachylophosaurus canadensis;* Peptide 2: „GLPGESGAVGPAGPPGSR‟- from α_2 type 1 collagen chain of *Brachylophosaurus canadensis* and Peptide 3: „GLPGESGAVGPAGPIGSR‟- from α_2 type 1 collagen chain of *Tyrannosaurus rex.*

Before computationally modelling the structure for the selected peptides, a multiple sequence alignment was performed using COBALT and the results were submitted to ConSurf web server [11]-[13] to determine the extent of evolutionary conservation of the residues in the peptides. Also from the literature [7], it was found that a critical region was necessary for interaction of collagen with ECM proteins including integrins. Thus the residues GFPGER, forming the critical recognition motif for the integrins on collagen present in a majority of species as evident from the alignment was included upstream of the selected peptide 1 in BioEdit. Then the peptides after analysis were ascertained for computational structure modelling using THeBuScr ver. 1.07 [14]. The 3 dimensional structure of collagen (type 1) is a triple helix consisting of two α_1 and one α_2 chains. Accordingly the peptide sequences were fed in FASTA format to the THeBuScr along with setting of parameters to build the structure of collagen and the output was obtained in pdb format. The modelled collagen structures were assessed using RAMPAGE [15] and RaptorX [16].

The structure of α_1 (I) integrin domain was retrieved from PDB (Code-2M32). The structure was then analyzed using Consurf to determine the functional importance of the ligand binding site (site where the collagen binds) based on evolutionary conservation of residues at the site. Docking of the prepared dinosaur collagen structures (ligand) with the integrin receptor was then carried out using PatchDock [17]. Top 20 solutions were refined using Firedock [18], [19] (refinement based on energy minimization) and from the top 10 refined solutions, the solution with the highest rank (solution no. 1) was taken up for further analysis. Analysis of the top ranking solution was done using SPPIDER [20] and CCP4MG [21]-[23].

4. Results and Discussion

From the results that were obtained by submitting the multiple sequence alignment from COBALT to „ConSurf‟ it was seen that certain residues of the peptides were identical in different organisms included in the alignment, and coloured according to ConSurf‟s conservation score (Fig 1). As can be seen from the ConSurf coloured structure of 2M32 (Fig 2a) the region where the collagen triple helix (displayed as a grey backbone) binds features highly conserved residues which are involved in ligand binding and thus implies their functional importance. The same can be attributed to the ConSurf coloured structures of modelled dinosaur collagen docked with the α1 integrin I domain in which the region where the collagen binds i.e. the trench shaped groove contains residues that have been conserved (Figs 2b and 2c).

The structures of dinosaur collagen peptides modelled using THeBuScr ver. 1.07 are presented (Figs 3a and 3b). The collagen peptides are of type I, consisting of two alpha 1 and one alpha 2 chains. Evaluation of the modelled structures of dinosaur collagen peptides by „RAMPAGE‟ (Ramachandran Plot Assessment) indicated that 100.0% of the residues lied in the favoured region which is greater than the expected value (~98.0%) suggesting the reliability and the quality of structure prediction. Also the result obtained from performing a pair-wise structure alignment of the dinosaur collagen micro-fibril with rat collagen micro-fibril (PDB code 3HQV) using „RaptorX‟ showed how the modelled structures exhibited a similar fold as that of the rat micro-fibril structure. This is reflected in the computed TM scores of the top solutions i.e., 0.4420 (>0.3) for the modeled collagen structures of both the dinosaur species. Visual inspection of the top ranking refined solution structures showed that the dinosaur collagen peptides binds to a trench shaped groove in the α1 (I) Integrin domain (Figs 4a and 4b) which is comparable to the binding of collagen mimetic peptide to the α1 (I) Integrin domain in the PDB structure 2M32.

A comparative analysis of the α1 (I) Integrin domain bound with collagen mimetic peptide complex (PDB code-2M32) and the refined solution structures of dinosaur collagen docked with the α1 (I) Integrin domain to identify the residues involved in the protein-protein interaction was done using SPPIDER. The SPPIDER analysis revealed that the following residues i.e., N^{14}, S^{15}, Y^{17}, G^{77}, Q^{80}, D^{114}, G^{115}, E^{116}, S^{117}, H^{118}, S^{145}, R^{148}, F^{156}, E^{159} of α1 (I) Integrin domain i.e., of the receptor molecule were involved in the protein-protein interaction and was common to all the complexes analyzed- 2M32, *T. rex* and *B. canadensis* collagen docked with α1 (I) Integrin domain.

Visualization and analysis in CCP4MG showed that the residues Lys^{155}, Glu^{159} and Thr^{81} , $Glu^{116}$of the receptor formed H-bond interaction with Arg^{6} of leading strand and

Gly[115] of the receptor formed H-bond interaction with Gln[9] and Arg[6] of the middle strand in the refined model of *T. rex* collagen docked with α1 (I) Integrin domain. Similarly the residues Ile[16] , Ser[13] and Arg[76] of the receptor formed H-bonds with Arg[6] of middle strand of collagen and Ser[15] of the receptor formed H-bond interaction with Gln[9] of middle strand of collagen ligand in the refined model of *B. canadensis* collagen docked with α1 (I) Integrin domain. The different residues involved in interaction between the docked dinosaur collagen peptides and the α1 (I) Integrin domain is presented in Fig 5.

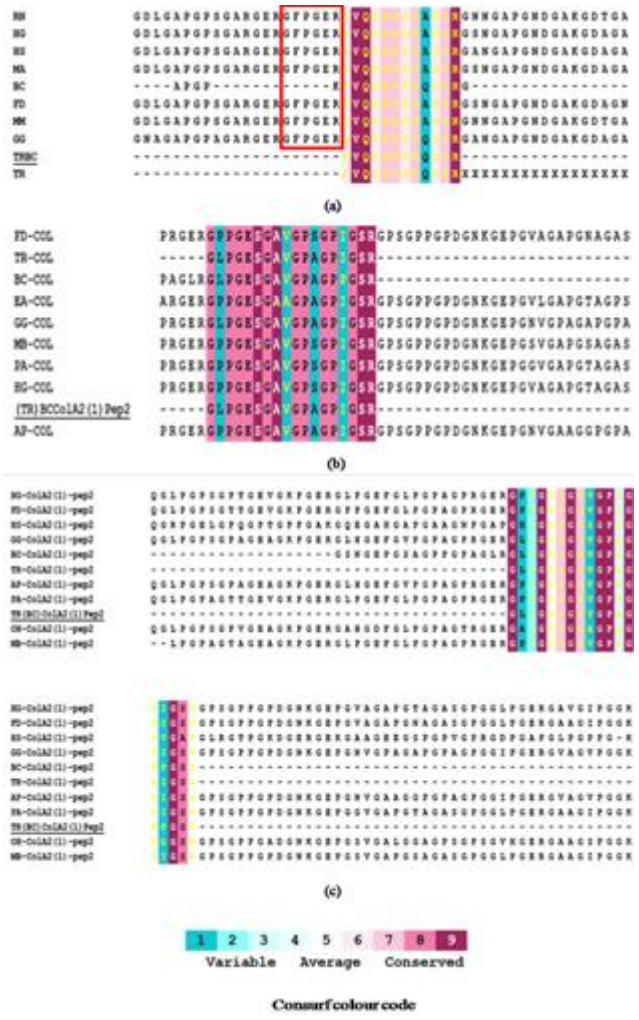


Figure 1: Multiple sequence alignment of peptides (a) GVQGPPGPQGPR, (b) GLPGESGAVGPIGSR and (c) GLPGESGAVGPPGSR color coded by ConSurf. [TRBC- collagen peptide from the alpha 1 (I) chain of both *Tyrannosaurus rex* and *Brachylophosaurus canadensis*; (TR)BC Col A2(1) - collagen peptide from the alpha 2 (1) chain of *Tyrannosaurus rex*; TR(BC) Col A2(1) - collagen peptide from the alpha 2 type 1 chain of *Brachylophosaurus canadensis* ;BC-*Brachylophosaurus canadensis*; TR-*Tyrannosaurus rex*; GG- *Gallus gallus*; MA- *Mammut americanum*; HS- *Homo sapiens*; MM-*Mus musculus*; AP-*Anas platyrhynchos*; OH- *Ophiophagus Hannah*; FD-*Fukomys damarensis*; RN- *Rattus norvegicus*; MB- *Myotis brandtii*; HG- *Heterocephalus glaber*; PA- *Pteropus alecto*] „GFPGER" critical motif highlighted.

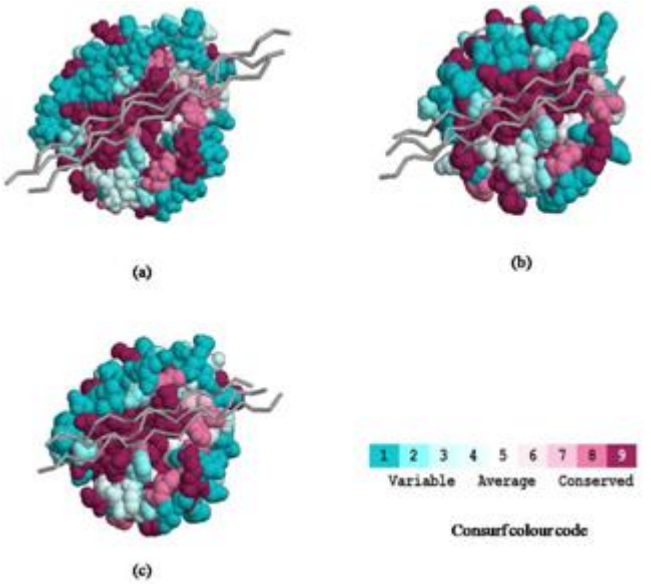


Figure 2: Visualization of evolutionarily conserved residues (colored spheres) of 2M32 an α 1 (I) integrin bound with (a) collagen present in the original structure (b) *T. rex* collagen and (c) *B. canadensis* collagen (shown as grey backbone in foreground). The residues in the region where the collagen triple helix binds, are conserved showing its functional significance (ConSurf output- coloration according to conservation score and viewed in Rasmol)

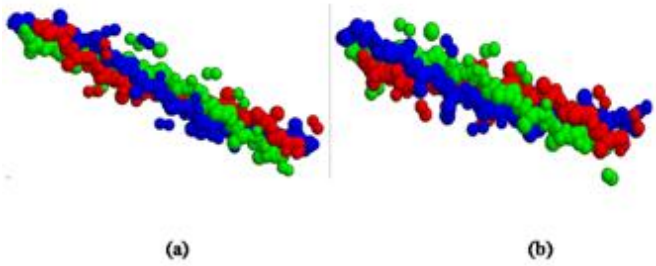


Figure 3: *T. rex*(a) and *B. canadensis*(b) collagen modelled using THeBuScr and visualized in Rasmol [Chain A-Blue; Chain B- Green; Chain C- Red] (Display type- Space fill; Color- by chain type)

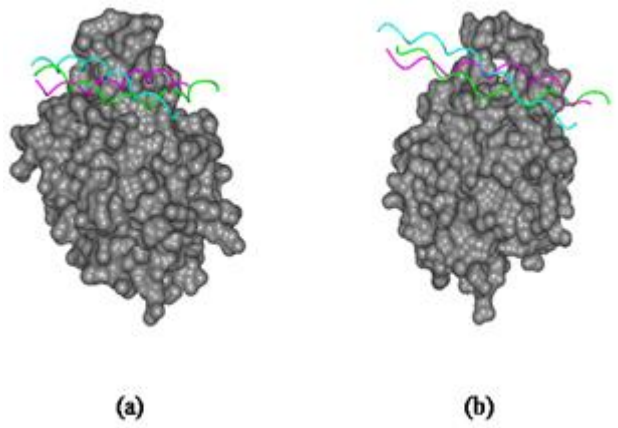
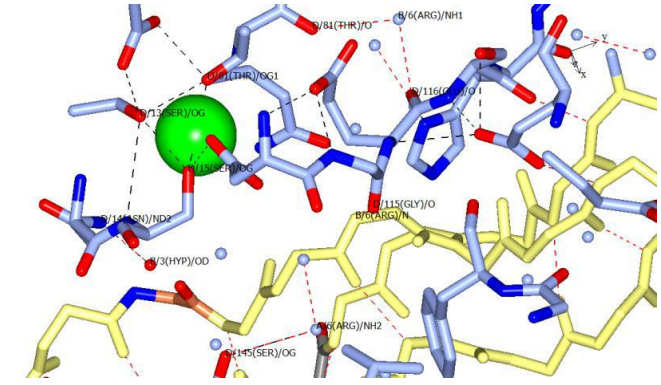
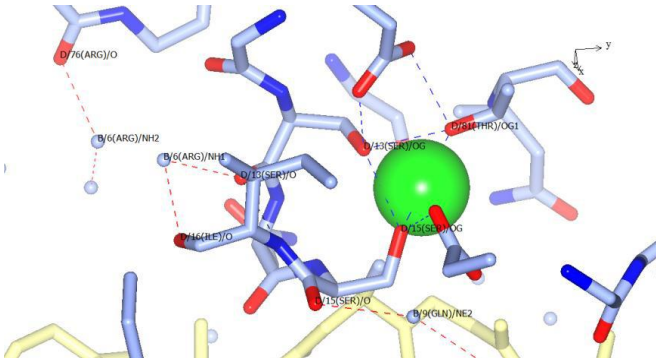


Figure 4: Interface representation of modeled *T. rex*(a) and *B. canadensis*(b)collagen docked with α1 (I) Integrin domain. The collagen triple helix binds across a surface trench in the α1 (I) Integrin domain (displayed as a grey molecular surface).The leading, middle and trailing strands of the collagen triple helix shown here in cyan, magenta and green respectively.



(a)



(b)

Figure 5: CCP4MG analysis of the refined solution structure of modelled (a) *T. rex* collagen and (b) *B. canadensis* collagen docked with α1 (I) Integrin domain displaying the residues (labeled) involved in the interaction. [Receptor chain residues Ser (13), Ser (15) and Thr (81) coordinating with the Mg2+ ion (green sphere) displayed in blue dashed lines; H-bonds between the residues of receptor and ligand displayed in red dashed lines]

5. Acknowledgements

The authors thank the Biotechnology Information System Network (BTISNET), Department of Biotechnology (DBT), Government of India, as part of the work was carried out utilizing the resources provided under the scheme of Bioinformatics Infrastructure Facility (BIF) at Madras Christian College.

References

[1] J. L. Bada, S. W. Xueyun and H. Healy, "Preservation of key biomolecules in the fossil record: current knowledge and future challenges", Philosophical Transactions of the Royal Society of London, 354, pp. 77-87,1999.

[2] C.Nielsen-Marsh, "Biomolecules in fossil remains-Multidisciplinary approach to endurance", The Biochemist, 24 (3), pp. 12–14, 2002.

[3] J.M.Asara, M. H. Schweitzer, L. M. Freimark, M. Phillips,L. C. Cantley, "Protein Sequences from Mastodon and Tyrannosaurus Rex Revealed by Mass Spectrometry", Science, 316 (5822), pp. 280-285, 2007.

[4] M. H. Schweitzer, W. Zheng, C. L. Organ, R. Avci, Z. Suo, L. M. Freimark, V. S. Lebleu, M. B. Duncan, G. Matthew, V. Heiden, J. M. Neveu, W. S. Lane, J. S. Cottrell, J. R. Horner, L. C. Cantley, R. Kalluri, J. M. Asara, "Biomolecular characterization and protein sequences of the CampanianhadrosaurBrachylophosauruscanadensis", Science, 324 (5927), pp. 626–631, 2009.

[5] J. D. San Antonio, M. H. Schweitzer, S. T. Jensen, R. Kalluri, M. Buckley, *et al*., "Dinosaur Peptides Suggest Mechanisms of Protein Survival", PLoS ONE, 6(6), e20381, 2011.doi:10.1371/journal.pone.0020381.

[6] S.Perumal, O.Antipova, J.P.Orgel, "Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis", Proceedings of the National Academy of Sciences USA, 105 (8), pp.2824–2829, 2008.

[7] S.M.Sweeney, J.P.Orgel, A.Fertala, J.D.McAuliffe, K.R.Turner, *et al*., "Candidate Cell and Matrix Interaction Domains on the Collagen Fibril, the Predominant Protein of Vertebrates", Journal of Biological Chemistry, 283 (30), pp. 21187-21197, 2008.doi: 10.1074/jbc.M709319200 originally published online May 15, 2008.

[8] K.Kühn and J. Eble, "The structural bases of integrin-ligand interactions", Trends in Cell Biology, 4 (7), pp. 256-261, 1994.

[9] R.L.Rich, C.C.Deivanayagam, R.T.Owens, M.Carson, A.Höök, D.Moore, J.Symersky, V.W.Yang, S. V.Narayana and M.Höök, "Trench-shaped binding sites promote multiple classes of interactions between collagen and the adherence receptors, alpha(1)beta(1) integrin and Staphylococcus aureuscna MSCRAMM", Journal of Biological Chemistry, 274 (35), pp. 24906-13, 1999.

[10] C.G.Knight, L.F.Morton, A.R.Peachey, D.S.Tuckwell, R.W.Farndale andM. J.Barnes, "The Collagen-binding A-domains of Integrins a1b1 and a2b1Recognize the Same Specific Amino Acid Sequence, GFOGER, in

Native (Triple-helical) Collagens", The Journal of Biological Chemistry, 275 (1), pp. 35-40, 2000.doi: 10.1074/jbc.275.1.35.
[11] M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko and N. Ben-Tal, "ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures",Nucleic Acids Research, 33 (suppl 2),W299-W302, 2005.
[12] F. Glaser, T. Pupko, I. Paz, R.E. Bell, D. Bechor, E. Martz and N. Ben-Tal, "ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information", Bioinformatics, 19 (1), pp.163-164. 2003.
[13] C. Berezin, F. Glaser, J. Rosenberg, I. Paz, T. Pupko, P. Fariselli, R. Casadio and N. Ben-Tal, "ConSeq: The Identification of Functionally and Structurally Important Residues in Protein Sequences", Bioinformatics, 20 (8), pp.1322-1324, 2004.
[14] J.K. Rainey, and M.C. Goh, "An interactive triple-helical collagen builder", Bioinformatics, 20(15), pp.2458-2459, 2004.
[15] S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson, "Structure validation by Calpha geometry: phi,psi and Cbetadeviation",Proteins: Structure, Function & Genetics, 50 (3), pp. 437-450, 2002.
[16] Morten Källberg, Haipeng Wang, Sheng Wang, JianPeng, Zhiyong Wang, Hui Lu and JinboXu, "Template-based protein structure modeling using the RaptorX web server", Nature Protocols, 7 (8), pp. 1511–1522, 2012.
[17] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson, "PatchDock and SymmDock: servers for rigid and symmetric docking", Nucleic Acids Research, 33(suppl 2),W363-367, 2005.
[18] N. Andrusier, R. Nussinov and H. J. Wolfson, "FireDock: Fast Interaction Refinement in Molecular Docking", Proteins, 69(1), pp. 139-159, 2007.
[19] E. Mashiach, D. Schneidman-Duhovny, N. Andrusier, R. Nussinov, H. J. Wolfson, "FireDock: a web server for fast interaction refinement in molecular docking", Nucleic Acids Research, 36(Web Server issue):W229-32, 2008.
[20] A. Porollo, and J. Meller, "Prediction-based fingerprints of protein–protein interactions", Proteins:structure, function, and bioinformatics, 66(3), pp.630-645, 2007.
[21] S. McNicholas, E. Potterton, K. S. Wilson and M. E. M. Noble, "Presenting your structures: the CCP4mg molecular-graphics software", ActaCrystallographica, D67 (4), pp. 386-394, 2011.
[22] L. Potterton, S. McNicholas, E. Krissinel, J. Gruber, K. Cowtan, P. Emsley, G. N. Murshudov, S. Cohen, A. Perrakis and M. Noble, "Developments in the CCP4 molecular-graphics project", ActaCrystallographica, D60 (12), pp. 2288-2294, 2004.
[23] E. Potterton, S. McNicholas, E. Krissinel, K. Cowtan and M. Noble, "The CCP4 molecular-graphics project", ActaCrystallographica, D58 (11), 1955-1957, 2002.

Author Profile

B. Sai Shankar completed his B. Sc. and M. Sc. in Zoology from Madras Christian College, Chennai in 2013 and 2015 respectively. He has completed a project under the „Star college scheme" of DBT, India in the area of vector control during the year 2013. He has also received „Studentship" from Bioinformatics Infrastructure Facility (BIF), BTIS Net at Madras Christian College.

Dr. C. Joyce Priyakumari is the co-coordinator of Bioinformatics Infrastructure Facility (BIF) of BTIS Net centre at Madras Christian College. She holds a PhD from Madras University and currently teaches Molecular biology, Biotechnology and Bioinformatics at Madras Christian College, Chennai.