Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 4 Number 1 (2011) pp. 47—54

http://www.ripublication.com/jcib.htm

Comparative Modeling Study of RNA Polymerase C of *Hypericum perforatum*

Ashish Kumar[1], Prakhar Bisht[1], Ashish Gupta[1], Sharada M. Potukuchi[2] and A.K. Srivastava[1*]

[1]Amity Institute of Biotechnology, Amity University, Noida (U.P.), India
[2]School of Biotechnology, Shri Mata Vaishno Devi University, Katra (J&K), India
[*]Corresponding Author E-mail: aksrivastava@amity.edu

Abstract

Hypericum perforatum, also known as St John's Wort, has received special attention due to its pharmacological characteristics. It is widely known perennial herb for treatment of both mild and severe depression, especially in children and adolescents. RNA polymerase C is a protein which transcribes the information into mRNA, specifically having role in DNA binding or transcription factor recognition. With such prominent significance the three dimensional (3D) structure of RNA Polymerase C is still not reported. Homology modeling, a method to predict the structure of proteins, is utilized to design RNA Polymerase C protein structure. The template protein is obtained by Geno3D server via template protein pdb|3IYD|chain D, which has identity 63%, E value 9e-63 and alignment Score 599. A server for comparative modeling namely SWISS MODEL is used to compare the template protein with the target protein to construct a rough model. The structure of RNA polymerase C protein of *Hypericum perforatum* is found to resemble the structure of peroxiredoxin high resolution structure of Ding Protein of *Pseudomonas fluorescence*. From Ramachandran plot analysis the portion of residues falling into the most favored regions is calculated to be 69.5%. The 3-D model so designed may be further characterized, analyzed and illustrated using other modeling techniques.

Keywords: Homology Modeling, Structure, Blast, mRNA, Ramachandran plot, St. John Wort, Swiss Model, P27SJ, Template, Target.

Introduction

Hypericum perforatum a yellow flowered herb also known as St John's Wort, is used in healing for more than two thousand years. It contains both immune enhancing and

anti-viral components and is being studied by AIDS researchers. St John's Wort has significant concentrations of immune-modulating flavonoids, and also contains hypericin, a substance that is both antiviral and antidepressive [1-2]. Extracts from this plant are utilized in the treatment of mild to moderate depression, anxiety, mania, hypochondriasis, fatigue, sleep disorders, treating wounds for allaying the pain of contusions, all pulmonary complaints, kidney conditions, and treatment of inflammation and bacterial infections [3, 4]. Likewise, P27SJ, a novel protein acquired from the St John's Wort suppresses expression of HIV-1 genome. P27SJ associates with C/EBPβ, a transcription factor that regulates expression of the HIV-1 genome in macrophages and monocytic cells, and the viral transactivator Tat [5].

RNA polymerase is a protein that transcribes the genomic information stored within the DNA into messenger RNA (mRNA) [6-7]. The process of transcription can be divided into initiation, transcript elongation and termination and recycling of RNA Polymerase C (RNAPC) [8].

A very challenging problem in the field of drug discovery and computational biology is Protein modeling. Advances and progress in computational power have helped to solve this problem to a considerable extent; however, predicting accurate three-dimensional structure of proteins has always been and remains a complicated assignment. Two common methods of protein structure prediction are template-based modeling and *ab initio* modeling. Template based modeling is more popular [9]. Comparative modeling also known as Homology modeling (HM) is a method for constructing an atomic-resolution model of a protein from its amino acid sequence [10].

When the target and template are closely related HM can produce high-quality structural models. Thus, inspiring the formation of a structural genomics directed towards the production of representative experimental structures for all classes of protein folds. [11]. In bioinformatics, the utilization of the technique like HM has proved to be of immense significance. The 3D structure of the proteins held no meaning until the correct interpretation of their detailed structure was illustrated contextually [12]. For this pioneer plant protein the primary aim is to predict the 3D structure in order to show the importance of this medicinal plant protein and underline its other applications [13].

Materials and Methods

Target Sequence Retrieval

From the NCBI sequence database the retrieval of the amino acid sequence of the RNAPC [15] is achieved. The Accession No. of the sequence so obtained is ADD47958.1 and the GeneID is 290585426. In Protein Data Bank the 3D structure of the protein is not available. Thus, the designing of the 3D model is carried out considering the novel protein of *Hypericum perforatum.* The protein consists of 178 amino acid residues.

Template Searching and Sequence Alignment

The modeling of the target protein is obtained by finding a suitable template protein.

An attempt is made to search the template protein through Geno3D server; an online tool to search similar sequences on the basis of structure and sequence-wise similarity [16]. From the homology searching an intact activator-dependent transcription initiation complex (3IYD) of [3IYD: D] Chain D [17] is selected as template protein. The Swiss-Pdb Viewer package is used to derive the amino acid sequence alignment of the target and template proteins [18]. The minimization of the number of gaps and insertions are inspected and adjusted manually by aligning the sequences using default parameters.

Rough Model

Rough 3-D model is constructed from the sequence alignment between RNAPC and the template proteins using SWISS MODEL [19-20].

Refinement and Evaluation of Refined Model

With the aid of steepest descent and conjugate gradient techniques the rough model, so constructed, is solvated and subjected to constraint energy minimization with harmonic constraint of 100 kJ/ mol/Å2 applied for all protein atoms. Due to this the bad contacts between protein atoms and structural water molecules are eliminated. Computations are carried out in *vacuo* with the GROMOS96 43B1 [21] parameters set, applying Swiss –Pdb Viewer.

In the last step of HM, a series of tests for testing the internal consistency and reliability of the refined structure model are carried out. As a result, the backbone conformation is evaluated by the inspection of the Psi/Phi Ramachandran plot obtained from PROCHECK [22] analysis. The Swiss-Pdb Viewer energy minimization test is applied to check for energy criteria in comparison with the potential of mean force derived from a large set of known protein structures. Packing quality of the refined structure is investigated by the WHATIF [23].

Pocket Identification

The submitted model (PM0077218) is further used for finding pocket to identify residues which are involved in binding of substrate or transcription factor. A Probe radius 1.4 Angstrom is used to do this study.

Result and Discussion

While generating useful model only 30 % of the sequence homology is constructive. In this case, the sequence alignment score i.e. 599, was determined by the aid of Clustal W2 [24]. During our work, based on results calculated from mGenTHREADER [25] program and geno3D, the X-ray structure of the Chain A - pdb |3IYD| D, a high resolution structure of intact activator –dependent transcription initiation complex is selected as template. By using BLAST [26], for template protein 3IYD E value 9e-63, identity 63% and alignment score 599 is obtained. SWISS MODEL is used for building the model and global energy minimization [27].

With the help of SWISS MODEL technique, model is generated. The stereo-chemical parameters of the proteins like main and side chains data of RNAPC are

considered, to determine the quality of the model which is generated using PROCHECK 3.0. The main chain parameters like Ramachandran plot quality; peptide bond planarity, C-alpha Chirality, over-all G factor and the bad contacts per 100 residues are found to be within the limits for the model. The side chain parameters are in better range and within the limits for RNAPC. After comparing all parameters best model is obtained. Best model is subjected to refinement by using energy minimization. Refined model was analyzed by different protein analysis programs including RAMPAGE Software [14] for the evaluation of the Ramachandran plot quality, and WHATIF for the calculation of packing quality.

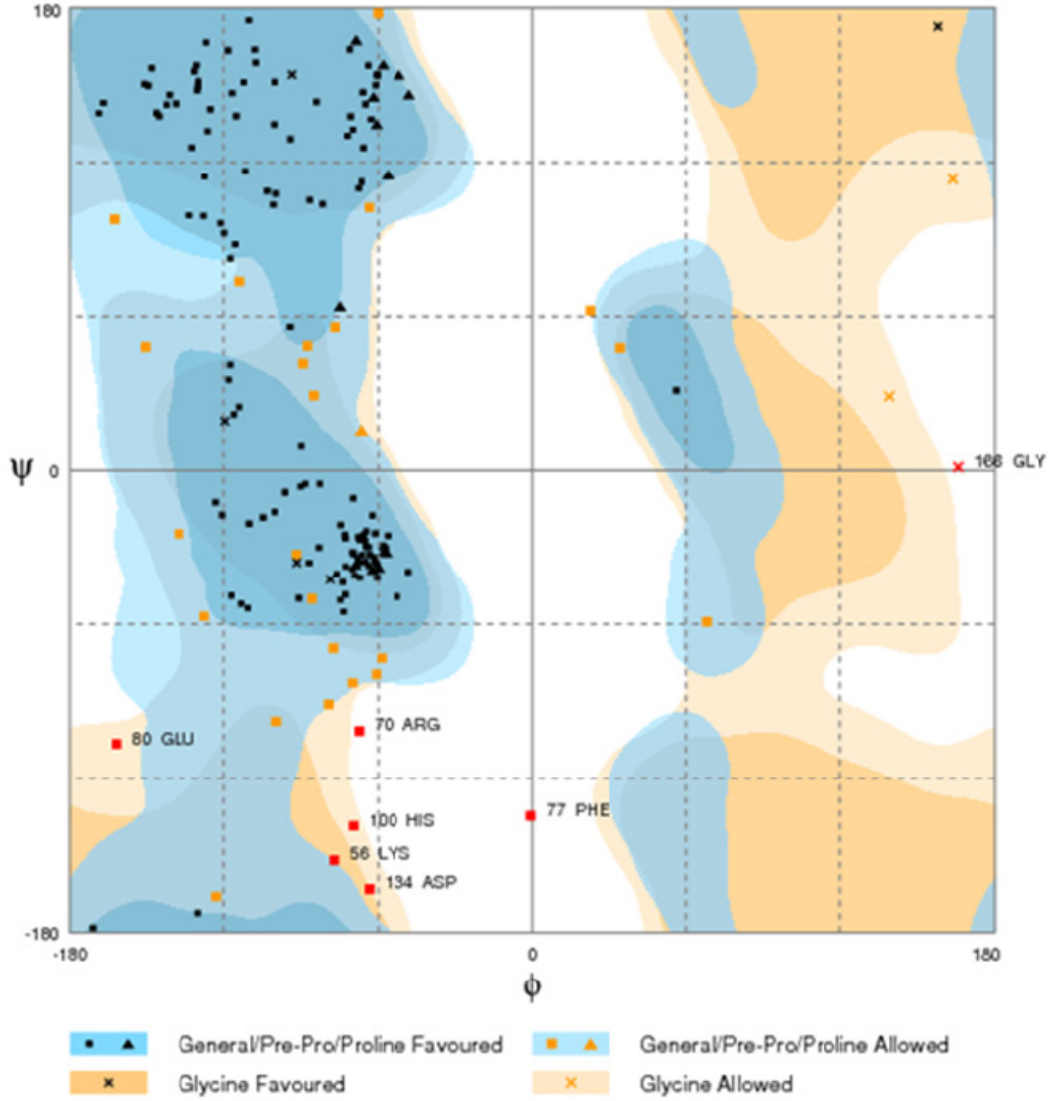

Figure 1: Ramachandran plot analysis of model using RAMPAGE [14] server. It shows the various residues falling in favoured, allowed and outlier region and the Glycine residues (143 residues are in favoured region, 28 in allowed region and 7 in disallowed region) so > 90% residues have allowed conformations.

The Ramachandran plot for RNAPC using RAMPAGE software revealed that among the 178 residues, 143 residues are in favored region, 28 in allowed region and 7 in disallowed region proving that the predicted model is acceptable (Fig. 1). Ramachandran plot for general, glycine, pre-proline and proline is also done and found that the glycine, pre-Pro and proline of RNAPC all fall under allowed regions. Packing quality for this model is found in normal range with the help of WHATIF server. From the tools VERIFY_3D [28] & ERRAT [29] which is available at Structural Analysis and Verification Server (SAVES) demonstrated that 50.57% of the residues had an averaged 3D-1D score > 0.2 and Overall quality factor 43.871.

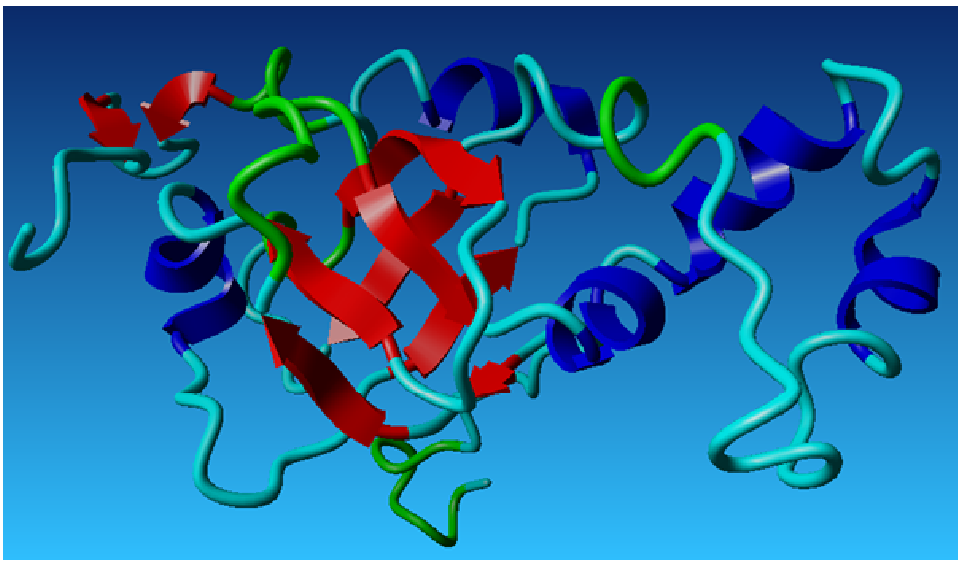

Figure 2: Ray diagram of 3D modeled structure of RNAPC. In this model residues which are responsible for alpha helix and extended strand are 74 and 25 respectively while 79 residues involve for random coil region.

The overall results provided the evidences that the predicted 3D structure of RNAPC is acceptable and of good quality. This structure (Fig 2; see PMDB ID - PM0077218 [30] for the corresponding coordinates in pdb format) is found to be satisfactory based on the above results. To identify the pocket in this model CASTp server was used. The volume and area of the best pocket which identified by using CASTp server was 1117.7 and 738.9 respectively (figure 3).

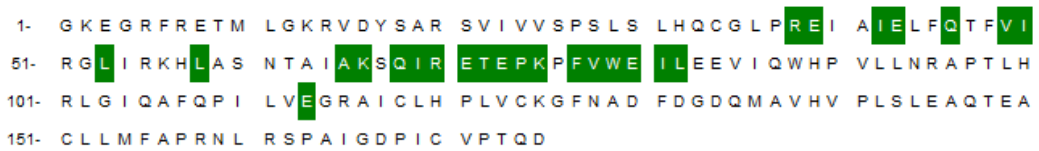

Figure 3: Residues involved in pocket formation are shown in green color. (CASTp server [31])

Drug designing and molecular modeling is the base of dry-laboratory research work. It highlights the *in silico* study of protein and nucleic acid which serves as a source of raw information and knowledge to the field of bioinformatics. The designing and structural prediction of these particular proteins encompasses vast details which serve multifaceted implications. This model of RNAPC, submitted successfully in Protein Model Database (PMDB) with the PMDB ID – PM0077218. (Refer Fig.2), characterizes the proteins wet laboratory applications and illustrates the mechanism of exact functioning of the protein.

Acknowledgement

Authors appreciate the assistance of Amity Institute of Biotechnology, Amity University, Noida and express our gratitude to the faculty members of School of Biotechnology, Shri Mata Vaishno Devi University, Katra for their support. Protein homology modeling software tools offline or online are also acknowledged.

References

[1] Diwu Z. (1995). Novel therapeutic and diagnostic applications of hypocrellins and hypericins. Invited review. Photochem Photobiol. 61:529–539.
[2] Wagner H, Bladt S. (1994). Pharmaceutical quality of Hypericum extracts. J Geriatr Psych Neurol. 7:65–68.
[3] Laudahn D. (1992). High dosed St John's Wort extract for treatment of mild and moderate depressions. Acta Horticult. 332:265.
[4] Okpanyi S N, Weischer M L. (1987). Animal experiments on the psychotropic action of Hypericum extract. Arzneimittel-Forsch. (Drug Res.) 37:10-13.
[5] Darbinian-Sarkissian N, Darbinyan A, Otte J, Radhakrishnan S, Sawaya BE, Arzumanyan A, Chipitsyna G, Popov Y, Rappaport J, Amini S, Khalili K. (2006). p27SJ, a novel protein in St John's Wort, suppresses expression of HIV-1 genome. Gene Ther. 13(4):288-295.
[6] Prelich G. (2002). RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. Eukaryot Cell 1(2): 153-62.
[7] Svejstrup J Q. (2004). "The RNA polymerase II transcription cycle: cycling through chromatin." Biochim Biophys Acta 1677(1-3): 64-73.
[8] Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. (2005). A structural perspective of CTD function. Genes Dev 19(12): 1401-15.
[9] Daga P R, Patel R Y, Doerksen R J. (2010). Template-Based Protein Modeling: Recent Methodological Advances. Current Topics in Medicinal Chemistry, 10, 84-94
[10] Martí-Renom M A, Stuart A C, 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.
[11] Williamson A R. (2000). Creating a structural genomics consortium. Nat Struct Biol. 7(11s):953.

[12] Gupta A, Pal S K, Narayan J, Potukuchi S M. (2009). InSilico structure prediction of p27SJ, a novel protein in St John's Wort, that suppresses expression of HIV-1 genome. Biofrontiers. 1(1): 32-35.

[13] Gupta A, Malaviya A, Potukuchi S M, Narayan J. (2010). Three Dimensional Structure Modelling and Analysis of Ribulose 1, 5 Bisphosphate Carboxylase of Hypericum perforatum. Conference Proceeding BIOCOMP 2010, Las Vegas, Nevada, USA. (2010). 807-810.

[14] Lovell S C, Davis I W, Arendall III W B, de Bakker P I W, Word J M, Prisant M G, Richardson J S, Richardson D C. (2002) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function & Genetics. 50: 437-450.

[15] Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One. 5(1): e8613.

[16] Combet C, Jambon M, Deléage G, Geourjon C. (2002). Geno3D: Automatic comparative molecular modelling of protein. Bioinformatics.18:213-214.

[17] Hudson B P, Quispe J, Lara-González S, Kim Y, Berman H M, Arnold E, Ebright R H, Lawson C L. (2009). Three-dimensional EM structure of an intact activator-dependent transcription initiation complex. Proc Natl Acad Sci U S A. 2009 Nov 24;106(47):19830-5.

[18] Guex N, Peitsch M C. (1997). Swiss model and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18:2714-2723.

[19] Sali A, Blundell T L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 234(3):779-815.

[20] 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.

[21] Scott W R P, Hünenberger P H, Tironi I G, Mark A E, Billeter S R, Fennen J, Torda A E, Huber T, Krüger P, Wilfred F. van Gunsteren. (1999). The GROMOS Biomolecular Simulation Program Package. J. Phys. Chem. A. 103(19):3596–3607.

[22] Laskowski R A, MacArthur M W, Moss D S, Thornton J M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283-291.

[23] Vriend G. (1990). WHAT IF: A molecular modeling and drug design program. J. Mol. Graph. 8:52-56.

[24] Larkin M A, Blackshields G, Brown N P, Chenna R, McGettigan P A, McWilliam H, Valentin F, Wallace I M, Wilm A, Lopez R, Thompson J D, Gibson T J, Higgins D.G.(2007). ClustalW and ClustalX version 2. Bioinformatics. 23(21):2947-2948.

[25] McGuffin L J, Jones DT. (2003). Improvement of the GenTHREADER method for genomic fold recognition. Bioinformatics 19:874-881.

[26] Altschul S F, Gish W, Miller W, Myers E W, Lipman D J. (1990). Basic local alignment search tool. J Mol Biol. 215(3):403–410.

[27] Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37, D387-D392.
[28] Lüthy R, Bowie JU, Eisenberg D. (1992). Assessment of protein models with three-dimensional profiles. Nature.356 (6364):83-85.
[29] Colovos C, Yeates T O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2(9):1511-1519.
[30] Castrignano` T, Onorio De Meo P D, Cozzetto D, Talamo I G, Tramontano A. (2006). The PMDB Protein Model Database. Nucleic Acids Research. 34: D306–D309.
[31] Liang J, Edelsbrunner H, Fu P, Sudhakar P V, Subramaniam S. (1998). Analytical shape computing of macromolecules II: identification and computation of inaccessible cavities inside proteins. Proteins. 33:18-29.