*International Journal on Recent Advancement in Biotechnology & Nanotechnology*
*Vol. 1, Issue 1 - 2018*

# PROTEOMIC ANALYSIS OF OCT4 PROTEIN, A MASTER REGULATOR OF REPROGRAMMING FACTOR OF INDUCED PLUROPOTENT CELLS (IPSCS)

**PURNIMA[*], SAURABH MISHRA[**]**

## ABSTRACT

OCT4 is a component of a network of transcription factors, including Sox2, Klf4, and c-myc selected by Takahashi and Yamanaka to induce somatic cells into pluripotent stem cells (iPSCs). Several questions regarding reprogramming factors of stem cells are remaining unanswerable due to limited experimental availability and ehilical issues. Proteomic analysis of OCT 4 is still remaining unpredicted as protein structure is not available in PDB. Three dimensional structure of OCT4 has been modeled using homology modeling approach through MODELLER program. The modeled structure uploaded to PDBsum and obtained a set of password protected analysis for OCT4. This paper shows annotated plots of each protein chain's secondary structure, detailed structural analyses generated by the PROMOTIF program, protein-ligand interactions, PROCHECK analyses of structural quality, and many others. Stem cell research is used in treatment of a number of diseases including genetic disorders.

**KEYWORDS:** Proteomic analysis, PDBsum, OCT4, Reprogramming factors, iPSCs.

## INTRODUCTION

Oct4 is considered a master regulator among Yamanaka's six transcription factor, selected to reprogram fibroblast cell to iPSC, **Nicholset el., 1998**. In human POU5F1 gene encodes OCT4 transcription factor and shows three isoforms, namely OCT 4A, OCT4B, OCT4B1.OCT 4A is found in nucleus and play role in pluropotency of embryonic stem cells **Takeda, Seino and Bell, 1992**]. Human POU5F1 gene is located at chromosome 6 with four exons and three introns .Various publications discussed wealth of information but Oct4 protein-protein interaction networks have not been defined. Three dimensional structure of OCT4 protein is not available in RCSB PDB to study detailed mechanism of interaction. PDB database have detailed structural information of various proteins of different organisms based on X-ray crystallography or nuclear magnetic resonance (NMR) techniques **Laskowski, 2009**. Previously OCT4 protein is modeled using homology modeling method and published. In this paper we will discuss interaction mechanism of OCT4 through proteomic analysis.

[*]Department of Biotechnology, Mewar University, Gangrar, Chittorgarh, Rajasthan, India.
[**]Department of Biotechnology, Mewar University, Chittorgarh, Rajasthan, India.

 **ISSN: 2582-1571**

## METHODOLOGY

PDBsum is a web-server providing a largely pictorial structural information of the deposited macromolecular structures at the Protein Data Bank (PDB).RCSB and EMBL-EBI servers offer a catalog of the structural information of all PDB entries through PDBsum. An addition feature of PDBsum is PDBsumProprietary, that is a in-house version for companies to process own structure. On submitting our predicted 3D structure of OCT4, we got a PDB code and password to access structural information Magyar et. 2005. The PyMOL molecular graphical tool is used to visualize three dimensional structure of the modeled protein. Predicted model of OCT 4 was accessed in PDB format and analyzed.

## RESULTS AND DISCUSSION

We received a corresponding summary web page of PDBsum, accessible by the four-character PDB identifier (n697). Proteomic analysis are performed through retrieving structural information of modeled OCT4 protein.

## PICTORIAL OVERVIEWS

The header details start with a thumbnail image of the molecule that can be rotated and zoomed. Thumbnnail image shows the protein chain only while DNA, ligands, metals and other molecules become transparent.

**Figure 1.Thumbnail image of the structure, as shown using 3Dmol.js on the entry's PDBsum page**

To study interactions of any protein, shape and chemical properties of its surface are the key factors. The surface contains binding sites in the form of clefts of different sizes. A large cleft provides an increased surface area to interact with other molecules or ligands. It has been suggested **Kuntz et al., 1982; DesJarlais et al., 1988** that the active site usually lies in the largest of all the protein's clefts or cavities.

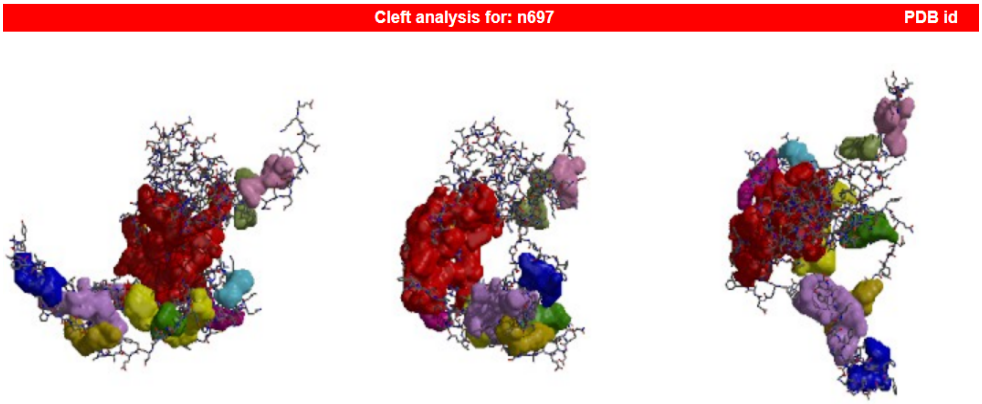


**Figure 2.Cleft analysis of OCT4 protein**

 **ISSN: 2582-1571**

The SURFNET program is employed to depict surface clefts in the predicted model of OCT4.The program shows cleft as solid surfaced coloured according to volume. The largest cleft is colored red, and often turns out to be the protein's binding site. This automatically generated image only gives a rough idea of the sizes and locations of the clefts. To identify, the most probable binding sites of our predicted model, R1 and volume are important functional requirements as ligands tend to preferentially bind to the largest cleft **Laskowski, 1996**. Our analysis showed existence of four major as shows in Table:

**Table 1 Parameters to analyze clefts in the predicted model**

| Clefts | Volume | R1 ratio |
|---|---|---|
| 1 (Red) | 7613.16 | 3.98 |
| 2(violet ) | 1911.52 | 0.00 |
| 3(Yellow) | 1504.41 | 0.00 |
| 4 (Blue) | 770.34 | 0.00 |

## PROTEIN DOMAINS

Human OCT4 protein has 360 amino acids and consists of three domains namely as an N-transactivation domain (NTD, 137-amino acids), a POU domain (152-aminoacids), and a C-transactivation domain (CTD,71-amino acids). The POU domain comprises of two DNA binding patterns -N-terminal 138-213 long amino acids POUS (POU specific and a C-terminal 230-290 amino acid long POUH (POU homeo) domain) that are connected by a linker of 213-230 amino acids (Fig. 1b).

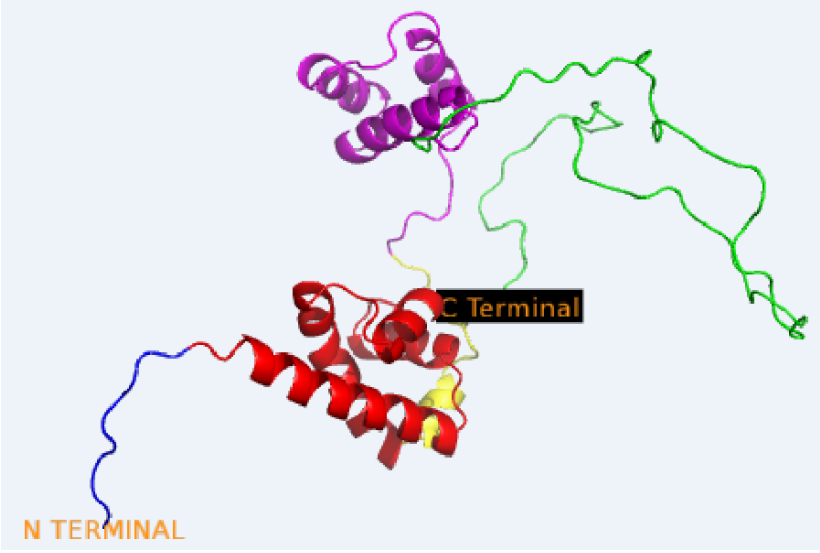


**Figure 3.Three dimensional structure showing domains: I domain (blue), II domain (red), III(yellow), IV domain(magenta), V domain(green)**

## INTERACTIONS

Figure show a protein-protein interaction diagram, generated using the HBPLUS29 and LIGPLOT30 programs, . The protein-DNA plot in Figure 4 is created with NUCPLOT. PDBsum predicted the hydrogen bond and nonbonded pattern both in α helices and β sheet within the domain.

 **ISSN: 2582-1571**

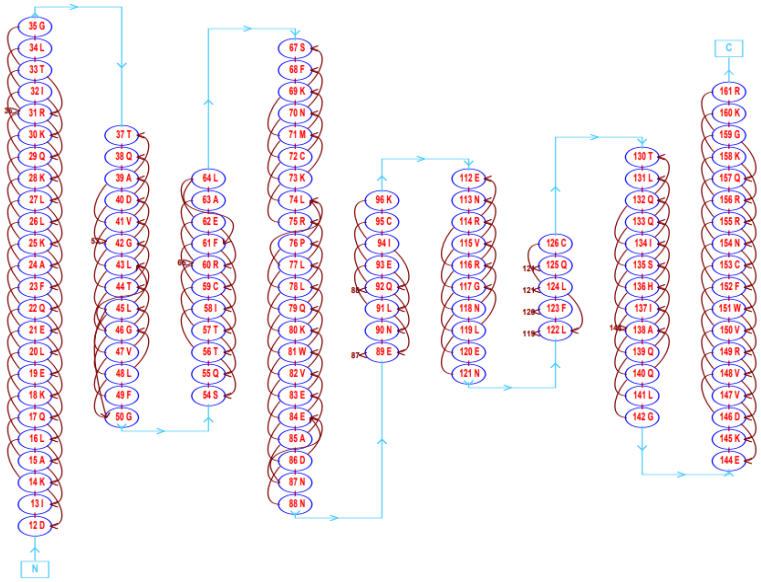

**Figure 4.Intra- and inter-hydrogen bond pattern between residues in b sheet topology of OCT4 predicted model**

Most of the hydrogen bonds were inter hydrogen bond between the beta sheets. Hydrogen bonds (H-bond) play a significant role in the formation of three dimensional structures and stabilization.

## SECONDARY STRUCTURE

The protein's secondary structure assignments and topology diagram show the proteins' secondary structure elements (alpha helices and beta sheets) together with various structural motifs such as beta and gamma turns, beta hairpins. These structural elements are computed ROMOTIF32 using The single-letter amino acid codes showing the protein's sequence are coloured according to their CATH structural domain **Hutchinson & Thornton, 1990**.

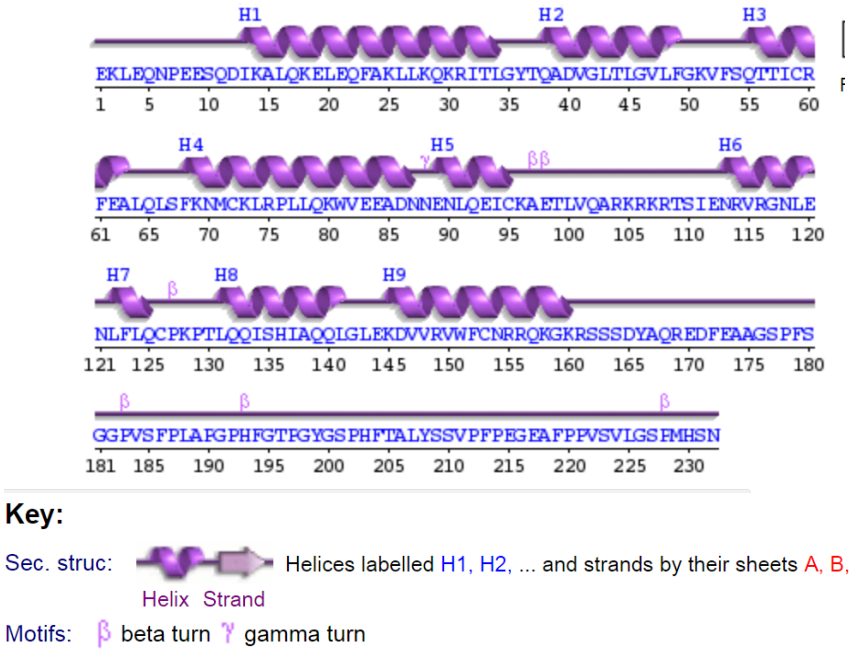


**Figure 5.Secondary structural elements (b sheets, b hairpins, b bulges, b turns, strands, helices)**

 **ISSN: 2582-1571**

Schematic "wiring diagram" of the protein's secondary structure-including strands (pink arrows), helices (purple springs), and other motifs in red (e.g., β-hairpins, γ-turns, etc). Helices are leballed H1, H2, etc while strands are labelled A, B, C, etc according to the beta sheet to which they belong.

Number of helices in chain : 9

Number of helix interactions involving helices of chain: 15

Number of beta turns in chain : 6

Number of gamma turns in chain 1

## TOPOLOGY DIAGRAM

The diagram illustrates how the beta strands, represented by large pink arrows, join up, side-by-side, to form beta-sheets and how the alpha-helices (the large cylinders) are located relative to these. The small arrows indicate the directionality of the protein chain, from the N- to the C-terminus. The numbers within the secondary structural elements correspond to the residue numbering given in the PDB file.

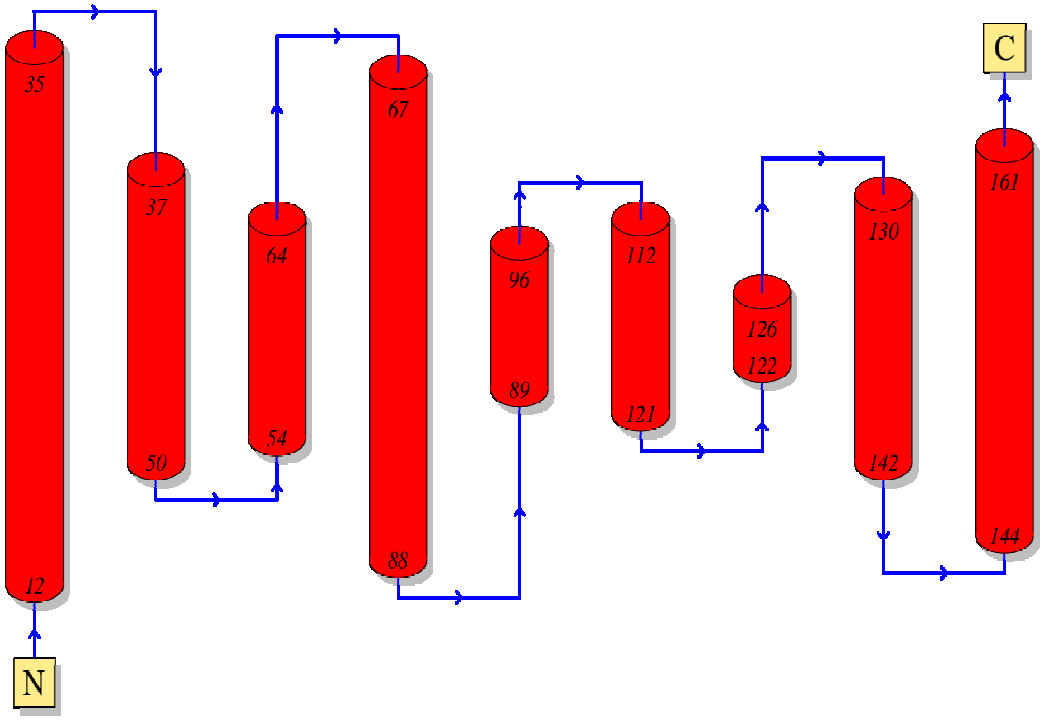


**Figure 6.Topology diagram illustrating the b strands**

The figure also illustrates the relative locations of the alpha helices (the red cylinders). The small arrows specify the directionality of the protein chain, from the N- to the C-terminus schematic diagram illustrating the protein's topology in terms of how the β-strands (pink arrows) are arranged into β-sheets, and the relative disposition of the α-helices (red cylinders

## PROCHECK ANALYSIS

Procheck analysis check the quality of all the protein chains in the predicted model **Sippl, 1993**. Ramachanran plot analyse the stereochemical quality by observing the residues lie in allowed regions.

 **ISSN: 2582-1571**

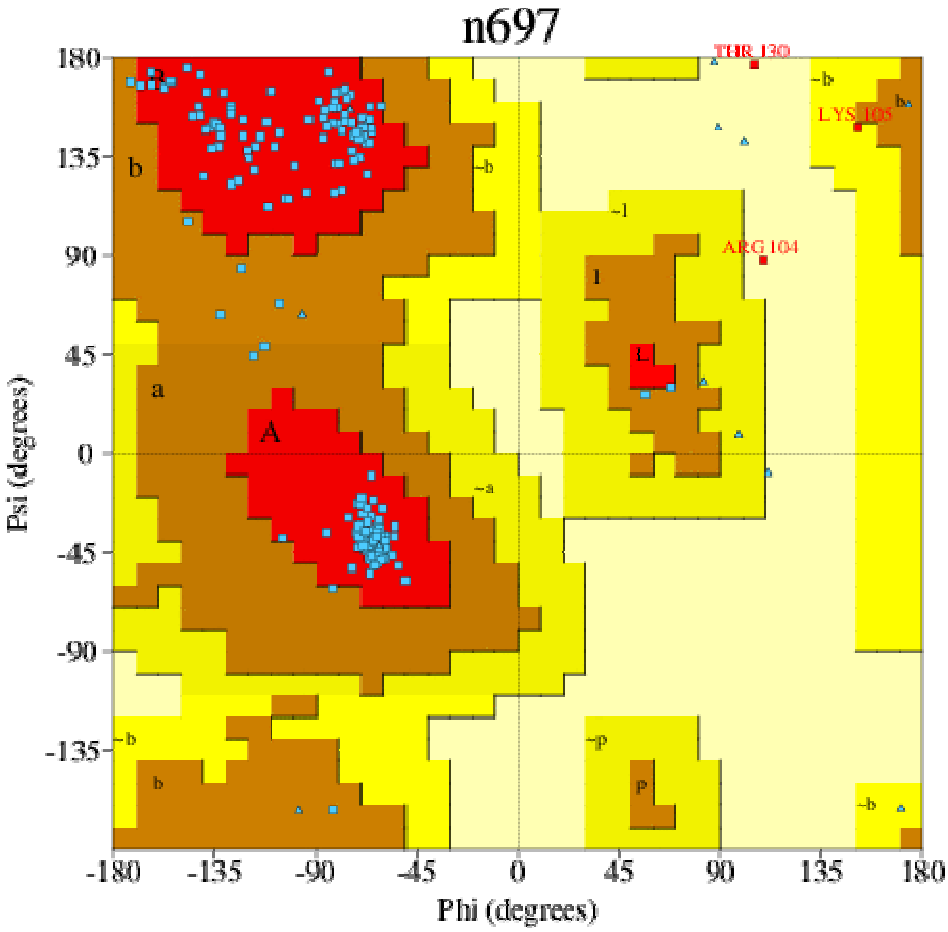


**Figure 7.Ramachandran Plot of predicted 3D structure of OCT 4 protein structure**

## ENZYME REACTION

Summary PROCHECK analyses (including a Ramachandran plot) giving an indication of the stereochemical 'quality' of all the protein chains in the structure, and links to related databases. In the list of molecules that follows, each protein chain is shown schematically by a 'wiring diagram' depicting its secondary structural motifs, primary sequence, structural domains and highlighting active site residues and residues that interact with ligands, metals or DNA/RNA molecules. The secondary structural motifs are computed by the PROMOTIF, **Wallace et al. 1995**, program, whose detailed outputs are available via hyperlinks, while the domain definitions come from the CATH protein structural classification database . For each ligand molecule a LIGPLOT diagram gives a schematic depiction of the hydrogen bonds and non-bonded interactions between it and the residues of the protein with which it interacts.

In PDBsum the matching residues are coloured according to their conservation from red for highly conserved, to blue for highly variable. Not all matching PROSITE patterns are shown; only those that appear to be true positives are included, **Hofmann et al., 1999**.

## REFERENCES

[1]. DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. J. Med. Chem. 31(4): 722-729, 1988.

[2]. Hofmann K., Bucher, P., Falquet, L. and Bairoch, A. (1999) The PROSITE database,

 **ISSN: 2582-1571**

its status in 1999. Nucleic Acids Res., 27, 215-219.

[3]. Hutchinson, E. G., & Thornton, J. M. (1990). HERA, a program to draw schematic diagrams of protein secondary structures. Proteins, 8, 203-212.

[4]. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161: 269-288, 1982.

[5]. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM. Protein clefts in molecular recognition and function. Protein Sci. 1996; 5:2438-2452.

[6]. Laskowski, R. A. (2009). PDBsum new things. Nucleic Acids Research, 37, D355-D359.

[7]. Laskowski, R. A., Chistyakov, V. V., & Thornton, J. M. (2005). PDBsum more, new summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Research, 33, D266-D268.

[8]. Magyar, C., Gromiha, M. M., Pujadas, G., Tusna´dy, G. E., & Simon, I. (2005). SRide, a server for identifying stabilizing residues in proteins. Nucleic Acids Research, 33, W303-W305.

[9]. Milburn D., Laskowski, R.A. and Thornton, J.M. (1998) Sequences annotated by structure: a tool to facilitate the use of structural information in sequence analysis. Protein Eng., 11, 855-859.

[10]. Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell, 95, 379-91.

[11]. Sippl, M.J. (1993). Recognition of Errors in Three-Dimensional Structures of Proteins. Proteins, 17, 355-362.

[12]. Takeda, J., Seino, S., and Bell, G.I. (1992). Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues. Nucleic Acids Res, 20, 4613-4620.

[13]. Wallce A. C., Laskowsi R. A., Thornton. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. 8(2):127-34.

 **ISSN: 2582-1571**