Research Article

Structural Insights of Induced pluripotent stem cell regulatory factors Oct4 and its Interaction with Sox2 and Fgf4 Gene

Vimal Kishor Singh[1*], Neeraj Kumar[2], Ramesh Chandra[2]

[1]Department of Biotechnology, Stem Cell Research Laboratory, Delhi Technological University, Delhi, India

[2]Department of Chemistry, University of Delhi, Delhi, India

*Corresponding author 1: Vimal Kishor Singh, Department of Biotechnology, Stem Cell Research Laboratory, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi-110042 E-mail: vim_kissor@yahoo.co.in; vimalkishorsingh@gmail.com; vimalksingh@dce.ac.in

Corresponding author 2: Ramesh Chandra, Department of Chemistry, University of Delhi, Delhi, India, Email: acbrdu@hotmail.com

Abstract

Pluripotency of stem cells is governed by various factors, and Octamer-binding transcription factor 4 (Oct4) has been shown to be most essential regulator of Embryonic Stem Cells (ESCs) pluripotency. It is a key molecule for the reprogramming process and vastly used for IPSCs generation in research laboratories. Oct4 interacts directly with another important molecule Sox2 and elicitvarious downstream signals during the reprogramming. Sox2 is sex determining region-Y protein and involved in reprogramming through interaction with Oct4 and other pluripotency factors. Both of these transcription factors are known to bind enhancer element of specific genes regulating pluripotency such as fibroblast growth factor 4 (Fgf4) gene. Fgf4 has been proposed to promote self-renewal in human ESCs and support proliferation of mouse inner cell mass. The molecular mechanism for Oct4 interactions with Sox2 and their target gene(s) remain cryptic. The present study focuses on these aspects, and a comprehensive in silico analysis of Oct4-Sox2 and Fgf4 structure and their interactions are reported here. Briefly, the Three-Dimensional (3D) models of Oct4 and Sox2 were generated and analyzed using de novo structure prediction approach. Further, these molecules were used to define the mechanism of their interactions to each other (Oct4 and Sox2); and their complex interaction with Fgf4 gene by molecular docking. The interaction of proteins and enhancer element of Fgf4 gene and resulting complexes were further evaluated through Dim plot analysis. This work reports that Oct4 binds to Sox2 with significant stability as indicated by their binding energy score (-618.95 kJ/mol) and two hydrogen bonds in between Gly307-Ser228 of 2.97 Å and Glu219-His260 of 3.02 Å bond lengths with hydrophobic interfaces. More importantly, we found that Oct4 in complex formation with Sox2 bind moreefficientlywithscore-615.86kJ/molthenOct4andSox2bindsindividuallywiththescoreof-585.22 and -395.99 kJ/mol respectively.

We suggest Oct4-Sox2 complex play a crucial role by synergistically increasing their binding to enhancer element of Fgf4 gene. Insights into the molecular mechanism of interaction of Oct4 and Sox2 would help to better understand the reprogramming regulatory network.

Keywords: Molecular Docking; Protein-Protein Interactions; Reprogramming; Pluripotency Transcription Factor.

Introduction

Oct4 is a master transcription factor for regulating early mammalian development and pluripotent cell self-renewal. Oct4 expression starts from the oocyte stage, becomes confined to Inner Cell Mass (ICM) of the blastocyst and ultimately remains only in primordial germ cells and maintains the pluripotency. Oct4 known as POU5F1, a homeodomain transcription factor of the

Pit1-Oct1/4-Unc-86 (POU) family consists a bipartite DNA binding domain [1]. Oct4 in combination with othertranscription factors mainly Sox2, Klf4, and C-Myc are responsible for the generation of Induced Pluripotent Stem Cells (iPSCs) from somatic cells [2].

iPSCs are pluripotent stem cells and possess the important property of self-renewability, and they can give rise to all three germ layers endoderm, mesoderm, and ectoderm. This reversion of adult cells to a reprogrammed state which resembles ESCs offers a non-diminishing potential source for the pluripotent cells with enormous applications such as generation of patient-specific tissues to facilitate disease modeling. Further, it is highly applicable in screening methods of drugs to develop new therapies [3].

In 2006, Yamanaka first time reported the generation of iP-SCs. Initially, they have used 24 transcription factors for reprogramming, and in the consequent studies, they reported the four transcription factors for the generation of iPSCs which are Oct4, Sox2, Klf4, C-myc. Further, in the subsequent studies on iPSCs generation, many reports suggest the Oct4 as a master regulator of source cells for reprogramming. Oct4 in cooperation to a Sox2 transcription factor is involved in somatic cell reprogramming. Sox2 is Sex determining region-Y box-2 related proteins which are activated in response to Leukemia Inhibitory Factor (LIF), which regulates the JAK-STAT signaling pathway. Downstream expression of signaling pathway results in Sox2 activation [4]. Both Oct4-Sox2 alone have been reported to generate iPSCs. Sox2 is a key factor required for iPSCs also thought to regulate Oct4 expression [5]. During early embryonic development, Members of POU family and Sox family exemplify functional cooperation and Sox proteins interact specifically with each other and bind to DNA [6]. Transcription regulation by Oct4 and Sox2 transcription factors reflects many of regulations showing the combinatorial control. Previous studies have demonstrated that specific promoters are selectively responsive to members of Oct family. Oct4 bind to the promoter element of Fgf4 gene adjacent of a Sox2 transcription factor. FGF4 gene promotes the auto self-renewable of the human ESCs and both Oct4 and Sox2 transcription factors are known to regulate the Fgf4 gene [7,8]. Fgf4 enhancer element was the first of the DNA elements found to contain composite DNA binding element of Oct4 and Sox2 [9].

Transcription factors regulations involve the protein-protein interaction for formation of regulatory complex and further their interaction with DNA binding element of a gene that can interact with and modulate the downstream genes responsible for maintaining the pluripotency. Their functional partnership has been characterized by regulatory elements in various species, including human, mouse, and the fruit fly. During development, POU and Sox proteins are expressed differentially, and their interactions may lead to a differential expression of the genes which are im-

portant for the determination of cell-fate [10]. The genes encoding the transcription factors Oct4 and Sox2 are critically regulated during embryonic development and somatic cell reprogramming. Their combinatorial role is critical, asit functions to specify the three germ layers of the mammalianembryo [11]. Although Sox2 and Oct4are considered to have a combinatorial role in vivo, the binding of POU factors by Sox2 isindiscriminate in vitro [12]. Biological data indicates that DNA binding region of two proteins synergizes the interaction with target genes, however not elucidate the combinatorial control at the molecular level and also the mechanism of interactions. Therefore, we have chosen the Oct4-Sox2-Fgf4 element ternary complex tostudy.

These transcription factors play a pivotal role in the development of embryonic stem cells and iPSCs generation, however; mechanism of their interactions for iPSCs generation is not well understood. Hence, detailed study is required to demonstrate the interaction of Oct4 and Sox2 proteins. It is also unclear how these four transcriptions factors Oct4, Sox2, Klf4 and Nanog lead to changes in the cellstate and systematic analysis of these transcription factors are lacking. Unearthing the interactions of Oct4, would provide insight to how this transcription factor interaction acts importantly to the decision of cell fate and maintenance of stem cellpluripotency.

Present study elucidates the molecular mechanistic interactions involved among Oct4/Sox2 and target gene Fgf4 to demonstrate important regulatory element in the overall process of pluripotency. Interactions of Oct4 protein and Sox2 protein have been done by using Molecular docking methods. Subsequently, Oct4 and Sox2 protein complex were evaluated to define its ability to interact with their direct target Fgf4 gene. These studies revealed the structural information of Oct4 and Sox2 transcriptionfactors and may exemplify their downstream interactions during the reprogramming. Moreover, this studies would help in better understand the mechanistic interaction between Oct4 and Sox2 proteinswith their regulatory Fgf4gene.
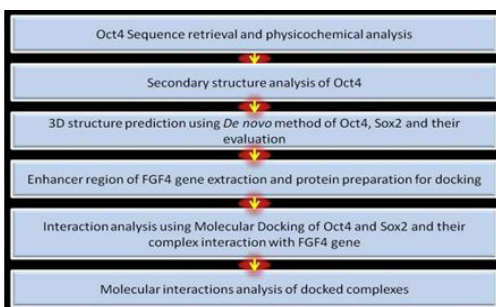
Methodology

Primary Structural Analysis of Oct4 and Sox2

The amino acid sequences of Oct4 and Sox2 proteins were retrieved from the NCBI protein sequence database (http://www.ncbi.nlm.nih.gov/protein) in Fasta format. Sequence analyses of Oct4 and Sox2 were done by local alignment and multiple sequence alignment using Basic Local Alignment Search Tool (BLAST)(http://blast.ncbi.nlm.nih.gov/Blast.cgi[13]andT-Coffee (http://www.ebi.ac.uk/Tools/msa/tcoffee/) [14] respectively.

Physicochemical properties of proteins were studied using the Expasy's Protparam server (http://web.expasy.org/protparam/).

Physicochemical properties including molecular weight, theoretical Isoelectric Point (pI), molecular formula, the number of negative and positive residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY) were determined [15]. The models quality was checked by energetic and geometricmeans (figure 1).

Figure 1: Flow chart diagram of the methodology used: Oct4 and Sox2 transcription factors sequence analysis and 3D structure prediction. Molecular interaction analysis of Oct4 and Sox2 and their heterodimer complex interaction with Fgf4 gene through docking.

3D Structure Prediction

The amino acid sequence of Oct4 and Sox2 were subjected to BLAST to find out the template for structure prediction through homology modeling. Lack of suitable structural template for Oct4 and Sox2 resorted to de novo structure modeling of Oct4 and Sox2 using iterative threading assembly refinement (I-TASSER) (http://zhanglab.ccmb.med.umich.edu/I-TASSER/).It is a web-based protein 3D structure prediction tool. It involves the simulation of protein structures as well as domains of the proteins andTm score and active sites. Tm score in which smaller distance between the structures is weighted high. The tm-score range is a value of more than 0.5 ensures a model of right topology and a value of 0.17 refer to random similarity or structure would be discarded [16]. With these parameters, the complete models wereobtained.

3D Structure Assessment

Predicted structures of Oct4 and Sox2 proteins were evaluated. I-TASSER server resulted in the generation of five structures for the each protein. Predicted structures were evaluated using the Rampage for Ramachandran plot, Verify3D, Errat and Prosa web interface. Ramachandran plot is to visualize dihedral angles phi (Φ) and psi (ψ)of amino acid residues in the proteins structures. It shows the possible allowed conformation of Oct4 and Sox2 polypeptides. Rampage server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) was used to analyzes the stereochemical properties to assess the quality of the structures, the planarity of the peptide bond, the main chain hydrogen bond energy; Cαchiralities, nonbonded interactions. The Errat (http://services.mbi.ucla.edu/ERRAT/) algorithm based on the statistical parameters of non-bonded interactions between different types of atoms and subsequently provides the accuracy of the protein models [17]. Verify3D (http://services.mbi.ucla.edu/Verify_3D/) analyze the compatibility of the predicted 3D structure with its primary amino acid sequence structural properties. A high Verify3D profile score indicates a better quality of model [18]. Prosa web interface (https://prosa.services.came.sbg.ac.at/prosa.php)calculatestheoverallquality ofthestructures;scoreshould be in the range of native experimentally determined protein structures. Prosa score (Z-score) determines the overall model quality and fold reliability of the predicted structure [19].

Oct4 and Sox2 Interaction Analysis through Molecular Docking and their Interaction Sites Identification

Interaction study of Oct4 and Sox2 was done by molecular docking. There are various tools available for protein-protein interactions docking. In the present study, we have used Hex 8.0 cuda program forthe molecular docking simulations. Hex determines the steric shape, electrostatic potential and the charge density of each protein. Molecular docking analysis was accomplished to obtain the best native conformation of protein-protein docked complex [20]. In the case of Hex, the input parameters were the Oct4 transcription factor PDB coordinate files for receptor protein and Sox2/Fgf4 as ligand files with set parameters. The total energy of interactions was calculated based on shape and electrostatics, and the final search was set to 25 (N= 25), and the angular search range of interaction was confined by selecting interface residues for Oct4 with angle 45° range. Other parameters were set to default values. The best model of docking complex, which showed the largest binding affinity were subjected to further molecular interactionstudies.

Predicated protein structure Oct4 and Sox2 were prepared before molecular docking since, these PDBs may contain some unwanted ligands or repetition of chains. Proteins were prepared using build/check/repair application of WhatIF server (http://swift.cmbi.ru.nl/servers/html/index.html) [21]. We studied the interactions of Oct4 and Sox2 proteins with Fgf4 gene enhancer element. First, we have studied the interaction of Oct4 and Sox2 proteins, and then Oct4 protein interaction with the enhancer element of Fgf4 gene and similarly Sox2 protein was docked with the target Fgf4 gene through molecular docking. Furthermore, we studied the interaction of Oct4-Sox2 heterodimer complex with the enhancer element of Fgf4 gene by molecular docking. Fgf4 enhancer element was extracted from the crystal structure of POU/HMG/DNA ternary complex available at protein data bank (PDB ID 1GT0), using Swiss PDBviewer.

Molecular Interactions Analysis

The molecular interaction plots between proteins were gen-

erated using Dimplot application of Ligplot software (v. 4.5.3). The Dimplot program gives a plot of the interactions across a protein-protein domain interface. Interactions plot includes hydrogen bonds, hydrophobic interactions and non-bonded contacts [22]. Molecular interactions (hydrogen bonds and hydrophobic interactions) of docked complex of Oct4-Sox2 and Oct4/Sox2-Fgf4 gene were determined using the Ligplot.
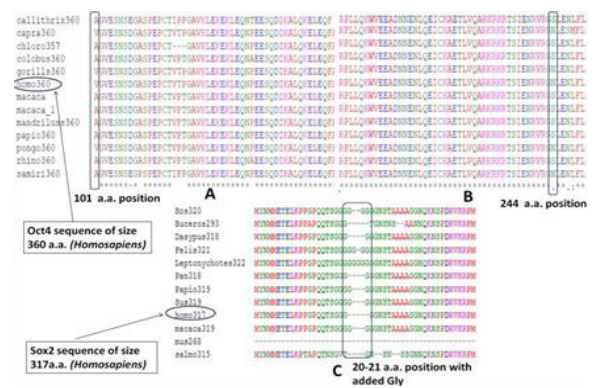
Results

Protein Sequence and Physicochemical Analysis of Oct4 and Sox2

The amino acid sequence of the Oct4 transcription factor of *Homosapiens* was obtained from NCBI protein database with accession number GenBank: AAI17436.1. On analyzing Oct4 sequence, we found 371 variants for *Homosapiens* and 109 variants for *Mus musculus* of different length present on NCBI, among which most sequences were of 360 amino acids. BLAST analysis of Oct4 (100 hits) showedmore than 90% sequence identity for *Homosapiens* variants and more than 86% identity with variantsin other organisms. Multiple sequence alignment of Oct4 in different organisms was done by T-Coffee, which showed that major regions where amino acids differ in organisms are 101 and 244 amino acids (Figure2(a)). The position101of Oct4 has Valto Alaaminoacid change in *Callithrixjacchus,Chlorocebussabaeus, Colobus angolensispalliatus, Saimiriboliviensis, Papioanubis, Mandrillusleucophaeus, Macacafascicularis, Macacamulatta, Rhinopithecusroxellana* and at 224 position Asn to Ser amino acid change in *Chlorocebussabaeus, Colobus angolensispalliatus, Macacafascicularis, Macacamulatta, Papioanubis, Mandrillusleucophaeusand Rhinopithecusroxellana. Capra hircus* has been found to have different amino acids at 7 different positions (247, 248, 293, 306, 327, 332, 347,352),*Saimiriboliviensis* differedat5 positions(27,101,108,110,351),*Callithrixjacchus* at 5 positions (27,55,108,133,351), *Mandrillusleucophaeus* at 4 positions (101, 244, 352, 354),*Papioanubis* at 3 positions (39,101, 244) and Gorilla at only position 51. These results showed Oct4 is conserved protein, present in various organisms and putative conserved domains are POU domain and DNA-binding homeodomain.

Similarly, the amino acid sequences of the Sox2 transcription factor of *Homosapiens* was obtained from NCBI protein database with sequence accession number: NP_003097.1. Through Sox2 sequence analysis, we found Sox2 has 108 variants of different amino acid length for *Homosapiens* but mostly of size 317 amino acids. Sox2 has 90% and more identity with other variants of *Homosapiens* and other organisms. Multiple sequence alignment showed that between 20-21 amino acid position of Oct4, Gly residues are added in various other organisms (Figure 2(b)). *Leptonychotesweddellii* of size 322 amino acids consisted of four

added Gly at the same position. Moreover, Salmo salar with 100% coverage of Sox2 and 87% identity found to have different amino acids at 90,143, 146, 152, 156, 172, 175, 177, 181,195, 196, 207, 217, 236, and 253 positions and *Buceros rhinoceros silvestris* with 96% coverage and90% identity at 143, 146, 293, 172, 175, 177, 181, 195, 251 and 253 positions. *Mus musculus* found to be different at only 301 position (Ser to Ala) and *Papioanubis* at 246 amino acid position (Gly to Ser). This suggests Sox2 is a conserved protein and its putative conserved domains are DNA-binding HMG- box region and Sox transcription factor domain of size ~80 aminoacids.

Figure 2: Multiple sequence alignment results of Oct4 and Sox2 in different organisms

(A)   Oct4 protein sequence (Homosapiens) at 101 amino acid position shows changes from Val to Ala

(B)   244 amino acid position of Oct4 of Homosapiens shows changes from Asn to Ser

(C)   Sox2 sequence at 20-21 amino acid position shows the different number of added Gly in the protein sequences of other organisms.

The primary structural features of Oct4 and Sox2 were studied using Protparam Tool and are described in (Table 1). As indicated, Oct4 protein (Human) consists 360 amino acids and molecular weight of 38570.6 daltons and molecular formula $C_{1718}H_{2657}N_{469}O_{517}S_{13}$. The calculated Isoelectric Point (pI) for Oct4 was 5.69, suggesting more negatively charged residues. Oct4 possess the higher extinction coefficient value which was 36940 M-1cm-1and indicating the presence of Cys, Trp, and Tyr in abundance. A Higher number of these residues aid to the quantitative study of protein-protein interactions. The aliphatic index value of Oct4 found to be 66.61 indicating its stability at a wide range of temperature. However, instability index value 53.24 (more then 40) indicates it may be thermally unstable. The Grand average hydropathy (GRAVY) value of Oct4 was -0.435, suggesting a hydrophilic pattern with high interaction with watermolecules.

Properties	Oct4 Protein	Sox2 Protein
Accession number	NCBI-GenBank: AAI17436.1	NCBI Reference Sequence: NP_003097.1
Molecular formula	C1718H2657 N469O517S13	C1467H2321 N443O457S26
Molecular weight	38570.6	34309.8
Amino acids	360	317
pI (Isoelectric Point)	5.69	9.74
Negative Residues	38	21
Positive Residues	33	34
Extinction Coefficient	36940	37360
Instability Index	53.24	58.73
Aliphatic Index	66.61	48.71
GRAVY	-0.435	-0.742

Table 1:Physico-chemical properties of Oct4 and Sox2 protein.

Whereas, Sox2 consists of 317 amino acids and molecular weight 34309.8 daltons, molecular formula C1467H2321N443O457S26 and pI 9.74 representing the basic nature of the protein. Extinction coefficient value was 37360 M-1-cm-1suggesting the presence of aromatic amino acids. Sox2 has aliphatic indexand instability index value 48.71 and 58.73 respectively, indicating the stability of Sox2. GRAVY value of Sox2 was -0.742, shows the hydrophilic pattern of the protein with more interaction with water.

3D Structure Prediction and Evaluation

Due to unavailability of suitable templates with appropriate sequence identity with proteins which is essential for protein homology modeling, we have designed the 3D structure for both the Oct4 and Sox2 (Human) proteins using de novo method through I-TASSER. I-TASSER is the hierarchical approach for protein 3D structure prediction, in this approach templates are identified by multiple threading approach and models are designed by adjoining the multiple fragment assembly simulations. I-TASSER generatedthe five models of acceptable high physicochemical properties, and out of them, high score model with best properties was selected as measured by TM-score. The Oct4 protein model-1 has TM-score forOct4 0.33 ± 0.11, which represents the absolute high quality of predicted structures. The Sox2 protein model-1 has TM-score0.45 ± 0.14, signifying the overall good quality of predicted structure (figure3).
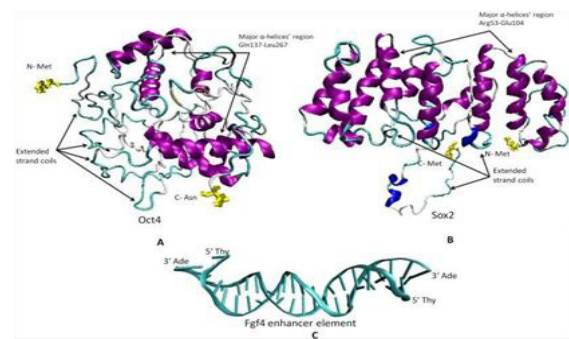

Figure 3: Diagram depicting 3D de novomodels

(A) Predicted structure of Oct4 protein, showing major region of alpha helices Gln137-Leu267 (purple) and extended strand coiled region (sky blue)

(B) Predicted structure of Sox2 protein showing the major region of alpha helices Arg53-Glu104 (purple) and extended strand coiled region (sky blue)

(C) Enhancer element (24 nucleotides) of Fgf4 gene extracted from the crystal structure of POU/HMG/DNA ternary complex

3D Structures Quality Assessment

The physicochemical quality of predicted structures of Oct4 and Sox2 were further assessed by Ramachandran plot, Prosa–web server, Verify-3D, and Errat. The stereochemical quality and accuracy of predicted models were evaluated by Ramachandran plot using Rampage server. For Oct4, the number of residues falling in the most favored regions of Ramachandran plot obtained by Rampage amounted to 72.1% and 17.9% in the allowed region and 10.1% in the outlier region. For Sox2, the number of residues which fall in the most favored region is 73.3% and 18.7% in the allowed region and 7.9% in the outlier region (figure 4).
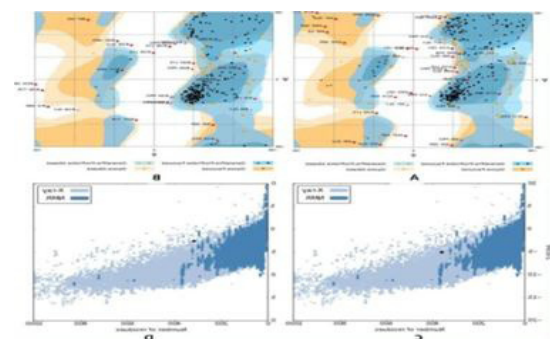

Figure 4: Diagram depicting the

(A) Ramachandran plot of predicted Oct4 structure, shows the 72.1% and 17.9% in the most favored and allowed regions

(B) Ramachandran plot of predicted Sox2 protein structure, shows 73.3% and 18.7% regions in most favored and allowed region respectively. Ramachandran plot shows the phi (φ)-psi (ψ) torsion angles for all amino acid residues and mostfavored and additional allowed regions of the structures.

(C) NMR/X-ray plot of the Oct4 structure through Prosa-web interface

(D) NMR/X-ray plot of Sox2 protein structure using Prosa-web

The Z-score represents the overall physio-chemical quality of the predicted protein 3D structure. The Z- score plot for Oct4 using Prosa-web showed the Z score -5.1, which was within the range of template proteins used for structure prediction (figure 4). Most identical template proteins PDBs were 3I1P,1GT0 and 2XSD and they have Z-score values of -3.88, -4.87 and -5.18 respectively. Similarly, Z-score for the Sox2 structure was obtained which was -2.62, shows the overall quality of thepredicted structure. Z-score of Sox2 also found to be within range with the templates used for structure prediction using de novo methods. Identical templates were 4N16, 1J46 and their Z-scores were -1.78 and-3.34.

For analyzing the statistics of non-bonded interactions and crystallographic model building Errat server was used. The overall quality score for Oct4 found to be 76.657 (as evaluated by Errat), representing the crystallographic model building and its refinement. Whereas, Sox2 predicted structure overall quality score was 93.793 suggest better crystallographic properties and non-bonded interactions. Both predicted structures were further evaluated through Verify 3D tool, which assesses the three-dimensional profile and entails at least 80% of amino acids residues should show absolute quality interactions (alpha, beta, loop, polar, non-polar, etc) and location in the predicted structure. Verify 3D resultedin81.39% residues of Oct4 shows the absolute the 3D crystallographic properties. On the hand, only 56.47% residues of Sox2 protein possess the 3D absolute qualities and indicates the moderate 3D structure qualities.

Oct4 and Sox2 Interaction Analysis

The protein interactions of Oct4 and Sox2 with Fgf4 gene was analyzed in two different manners.Firstly, the Oct4 protein and Sox2 were interacted to each other to form their heterodimer protein complex as depicted in (Table-2) (Figure 5). In the second approach, the Oct4 protein was used to interact with enhancer element of Fgf4 gene, and then Sox2 protein interacted with the Fgf4 gene. Then Oct4/Sox2 heterodimer complex was docked with enhancer element of Fgf4 gene. For the molecular docking PDB files of proteins were prepared and docked though Hex cuda 8.0. Hex docking module withsettingtheparameterswasrun.Hexshowedsignificantbindingenergytotalscore(E-totalscore)-618.95 kJ/mol for Oct4 and Sox2 protein. Fgf4 enhancer element PDB file was extracted from the1GT0 PDB and prepared for the molecular docking. Extracted Fgf4 enhancer element consists of two chains A and B of length 24 nucleotides each. Hex was run with receptor protein Oct4 and DNA PDB file with protein-DNA docking module selection. For Oct4 and enhancer element of Fgf4 gene, Hex resulted in binding energy total score -615.86 kJ/mol. Next, Sox2 was docked with the enhancer element ofFgf4gene,whichshowsenergytotalscore-395.99kJ/mol(Table-2).Oct4-Sox2proteincomplexalso docked with the enhancer element of Fgf4 gene, which resulted in higher energy total score of -615.86 kJ/mol, their interacting complex shown in figure 6. Higher energy total score of heterodimer complex of Oct4 and Sox2 then the individually interact with Fgf4 gene, indicates their synergistic enhancement in binding to the target gene.

Receptor	Ligand	E-total (kJ/mol)	Bonds	Hydrophobic interfaces
4-Oct	Sox2	-618.95	Two hydrogen bonds: Gly307-Ser228 of 2.97Å and Glu219-His260 of 3.02 Å bond lenghts	Oct4 residues: Pro346, Pro347, Phe345, Val348, Pro340, Pro338, Pro309, Leu226, Phe305, Ser306, Gly308, Ser335, Ser336, Gly342, Cys221, Phe312, Thr225, Gln224, Lys222 and Ala223. Sox2 residues: Lys122, Lys121, Met120, Thr118, Lys117, Gly190, Ala191, Tyr277, Tyr227, Gln229, Met194, His189, Thr85, Ser259, Ser261, Ser258, Gln282 and Arg262.
4-Oct	Fgf4(Enhancer element)	-585.22	One external bond between : Pro119- guanine15 of Fgf4	Oct4 residues: Lys206, Pro95, Glu96, Ala121, Val122, Lys123, Glu98 and Gly100.
Sox2	Fgf4(Enhancer element)	-395.99	No bond	Sox2 residues: Asn33, Gln34, Met1, Tyr2, Met276, Pro279 and Met4.
Oct4-Sox2 Comple x	Fgf4(Enhancer element)	-615.8	Three hydrogenbonds: 1. Gly120-guanine 7 of 3.2 Å, 2. Val117-thymine 5 of 2.59 Å, 3. Glu210-adenine13 of 2.54 Å bond lengths	Oct4 residues: Glu209, Lys206, Gln205, Thr116, Val117, Thr118, Ala121, Val122, and Gly83.

Table 2: Table is enlisting the energy total scores, hydrogen bonds and hydrophobic interfaces residues of docked complexes of Oct4/Sox2/Fgf4, through molecular docking interaction analysis using Hex cuda8.0.
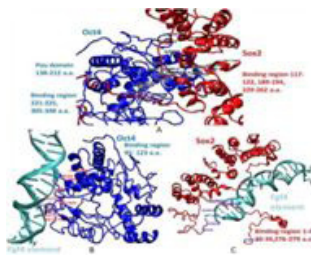
Figure 5: Diagram depicting the

(A)   Oct4-Sox2 protein complex resulting from molecular docking, where blue color ribbon coiled structure is Oct4 protein, and red color ribbon coiled structure is Sox2 protein.

(B)   Oct4-enhancer element Fgf4 gene (DNA strands in blue color) complex and shows the binding domain of Oct4 protein to DNA element

(C)   Sox2 protein docking complex with the enhancer element of Fgf4 gene through molecular docking
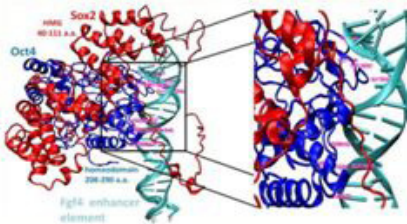

Figure 6: Diagram depicting the molecular interactions of docked complex of an Oct4-Sox2 heterodimer with enhancer element of Fgf4 gene and enlarged view of binding domains of Oct4 and Sox2 transcription factors to the enhancer element DNA strands of Fgf4 gene.

Molecular InteractionAnalysis

The molecular interactions (hydrogen bonds and hydrophobic interactions) of Oct4/Sox2 docked complex were identified by using Dimplot. We found, two intermolecular hydrogen bonds in the Oct4- Sox2 protein complex. Hydrogen bonds were in between amino acid Gly307-Ser228 of2.97 Åand Glu219-His260 of 3.02 Å bond lengths with various hydrophobic interactions. Hydrophobic residues of the Oct4 protein responsible for interactions were Pro346, Pro347, Phe345, Val348, Pro340, Pro338, Pro309, Leu226, Phe305, Ser306, Gly308, Ser335, Ser336, Gly342, Cys221, Phe312, Thr225, Gln224, Lys222, and Ala223. Hydrophobic residues of Sox2 were Lys122, Lys121, Met120, Thr118, Lys117, Gly190, Ala191, Tyr277, Tyr227, Gln229, Met194, His189, Thr85, Ser259, Ser261, Ser258, Gln282 andArg262. Hydrophobic interactions and hydrogen bonds between Oct4 and Sox2 indicate, the strong interactions between them.

Furthermore, the molecular interactions of Oct4 and Sox2 with enhancer element of Fgf4 gene were studied. Oct4-Fgf4 gene

complex showed one external bond between Pro119-Guanine15 of Fgf4 gene and amino acids Lys206, Pro95, Glu96, Ala21, Val122, Lys123, Glu98, and Gly100 were involved in the hydrophobic interactions. Whereas, Sox2-Fgf4 gene complex showed the hydrophobic interactions only with amino acids Asn33, Gln34, Met1, Tyr2, Met276, Pro279, and Met4. Moreover, molecular interactions of Oct4-Sox2 heterodimer complex with Fgf4 gene were analyzed. Docked complex of Oct4-Sox2 with enhancer element Fgf4 gene showed the three hydrogen bonds between amino acids and various hydrophobic interactions (figure 7). Hydrogen bonds were in between Gly120- Guanine7, Val117-Thymine 5 and Glu210-Adenine13 of 3.2 Å, 2.59 Å and 2.54 Å bond lengths respectively and Glu209, Lys206, Gln205, Thr116, Val117, Thr118, Ala121, Val122 and Gly83 residues of Oct4 were involved in the hydrophobic interactions with the Fgf4gene. This molecular interaction analysis showed the cooperative enhancement in interaction of Oct4-Sox2 complex with Fgf4 gene with more number of hydrogen bonds and hydrophobic interactions.
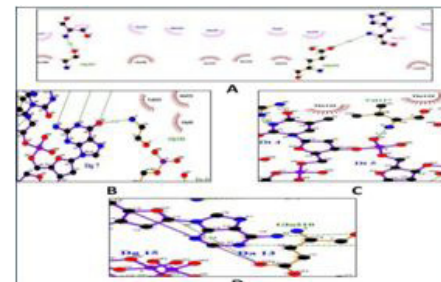
Figure 7: Diagram showing the hydrogen bonds and hydrophobic interactions of docked PDB complexes between (a) Oct4-Sox2 complex with enhancer element of Fgf4 gene, showing the first H-bond. (B) Second H-bond of Oct4/Sox2/Fgf4 ternary complex (C) Third H- bond of Oct4/Sox2/Fgf4 ternary complex. Hydrogen bonds are shown by dashed lines (green), and hydrophobic interactions are shown by spoked arcs (red) between residues.

Discussion

Oct4 is an important transcription factor, employed in the reprogramming of somatic cells in combination with other transcription factors Sox2, Klf4, and c-Myc. Oct4 interacts with Sox2 in pluripotency regulation networks and is useful for iPSCs generation.The Oct4-Sox2 complex has been known to regulate Fgf4 gene which helps in maintaining pluripotency of embryonic stem cells. However mechanism of interaction of pluripotency transcription factors and with downstream target Fgf4 gene is not known hence molecular interactions analyses have been performed.

Firstly, we have studied the physicochemical properties of Oct4 and Sox2 proteins. Oct4 has isoelectric point 5.69 which indicates it has more negatively charged residues and Sox2 has 9.74 shows the basic nature of more positive charged residues. The ali-

phatic index value of Oct4 and Sox2 were found tobeand 48.71 respectively, show the hydrophilic patterns of proteins and indicates Oct4 and Sox2 being stable at a wide range oftemperature.

We have generated the 3D structures of Oct4 and Sox2 transcription factors through de novo method using I-TASSER. The qualities of generated structures were ensured by TM-score. TM-score signifies the absolute quality of both the structures with scores of 0.33 ± 0.11 and 0.45 ± 0.14 for Oct4 and Sox2 respectively. Furthermore, modeled structures were evaluated by the Psi/Phi angle based backbone conformation, bond lengths using Ramachandran plot and found most of the residues of Oct4 andSox2-are in the favored region and allowed region. Both models were analyzed using Prosa and Errat. Z- Scores using the Prosa-web for Oct4 and Sox2 shows over all residues energies value within the rangeof templates used for structure prediction and Errat scores of Oct4 and Sox2 were found to be 76.657 and 93.793 respectively. These results represent a quality indicator for non-bonded atomic interactions, and its high score implies the better quality of structures. The above mentioned validating parameters assured the quality and reliability of models. To get insight into mechanistic interactions and their binding to the target gene, molecular docking method has beenemployed.

Oct4 protein is known to interact with Sox2 directly and regulates the various genes including the Fgf4 gene by binding to its transcription domain or enhancer element. Our results showed that Oct4 binds potently with Sox2 with a significant binding energy total score -618.95 kJ/mol with two hydrogen bonds and various hydrophobic interactions, shown in the table-2. For analyzing the binding specificity of Oct4 and Sox2 with Fgf4 gene, both proteins were docked individually with enhancer element ofFgf4 gene, and then their heterodimer complex was docked with Fgf4 gene element. Oct4-Fgf4 complex resulting in binding energy score -585.22 kJ/mol with eight hydrophobic interfaces and Sox2-Fgf4 complex had a score of -395.55 kJ/mol with seven hydrophobic residues only which shows Oct4 binds more potently than Sox2 with Fgf4 gene to regulate its expression. Next, molecular docking of Oct4-Sox2 protein heterodimer complex with enhancer element of Fgf4 gene showed a higher binding energy total score of -615.86 kJ/mol than Oct4 and Sox2 individual binding scores with Fgf4 gene. Oct4-Sox2 heterodimer complex through docking with Fgf4 gene showed three hydrogen bonds and strong hydrophobic interactions of Oct4 protein with promoter element, which suggest Oct4 is major protein responsible for binding and subsequent regulation of Fgf4 gene and binding of Sox2 to Oct4 introduces some conformational changes that resulted in the synergistic enhancement in binding of Oct4-Sox2 heterodimer complex with Fgf4 gene and possible reason for higher reprogramming efficiency when Oct4-Sox2 together induced in source cells for iPSCs derivation than lone Oct4 or Sox2 induced to the sourcecells.

Conclusion

We have generated the 3D structures of Oct4 and Sox2 transcription factors and found to be stable. We found Oct4 protein has DNA binding motif in homeodomain region. We found Oct4 binds strongly to a Sox2 protein with the formation of 2 hydrogen bonds and hydrophobic interactions. We also depict Oct4 binds more efficiently with Fgf4 gene in association with Sox2, forming 3 hydrogen bonds and strong hydrophobic interactions and suggest the possible cause of enhancement of binding of Oct4 with Fgf4 gene in the presence of Sox2 may be attributed to the possibility that Sox2 creates some conformational change in Oct4 and/or Fgf4 gene. Oct4-Sox2 complex results in better binding to Fgf4 gene than which is attained in individual binding and resulting in higher reprogramming efficiency of iPSCs generation with both Oct4 and Sox2 transcription factors together.

Acknowledgment

References

1. Guang JP, Chang ZY, Scholer HR, Duanqing PEI (2002) Stem cell pluripotency and transcription factor Oct4. Cell Research 12: 321-329.

2. Takahashi K, Yamanaka S (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663-676.

3. Singh VK, Kalsan M, Kumar N, Saini A, Chandra R (2015) Induced Pluripotent Stem Cells: Applications in regenerative medicine, disease modelling and drug discovery. Front. Cell Dev Bio l3: 2.

4. Singh VK, Kumar N, Kalsan M, Saini A, Chandra R (2015) Mechanism of induction: Induced PluripotentStem Cells (iPSCs). Journal of Stem Cells 10: 43-62.

5. Zhao R, Daley GQ (2008) From fibroblasts to iPS cells: induced pluripotency by defined factors. J Cell Biochem 10: 949-955.

6. Herr W, Cleary MA (1995) The POU domain: Versatility in transcriptional regulation by a flexible two-in-one DNAbinding domain. Genes & Dev 9: 1679-1693.

7. Dailey L, Yuan H, Basilico C (1994) Interaction between a novel F9-specific factor and octamer-binding proteins is required for cell-type-restricted activity of the fibroblast growth factor 4 enhancer. Mol Cell Biol 14: 7758-7769.

8. Ambrosetti DC, Basilico C, Dailey L (1997) Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol Cell Biol 17: 6321-6329.

9. Yuan H, Corbi N, Basilico C (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev 9: 2635-2645.

10. Dailey L, Basilico C (2001) Coevolution of HMG domains and home-odomains and the generation of transcriptional regulation by Sox/POU complexes. J Cell Physiol1 86: 315-328.

11. Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24: 372-376.

12. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N,et al. (2003) Mul-tipotent cell lineages in early mouse developmentdepend on SOX2 function. Genes & Dev 17: 126-140.

13. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. (2008) "BLAST+: architecture and applications." BMC Bioinformatics 10: 421.

14. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecu-lar biology 302: 205-217.

15. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Hu-mana Press, 571-607.

16. Yang J, Yan R, Roy A, Xu D, Poisson J, et al. (2015) The I-TASSER Suite: Protein structure and function prediction. Nature Methods 12: 7-8.

17. Colovos C, Yeates TO (1993) Verification of protein structures: pat-terns of nonbonded atomic interactions. Protein Sci 2: 1511-1519.

18. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83-85.

19. Wiederstein M, Sippl M. J (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res35: 407-410.

20. Ghoorah AW, Smail-Tabbone M, Devignes MD, Ritchie DW (2013) Protein Docking Using Case- Based Reasoning. Proteins: Structure, Function, Bioinformatics 81: 2150-2158.