
IN SILICO ANALYSIS OF LIN28, AN IMPORTANT REPROGRAMMING FACTOR OF INDUCED PLURIPOTENT STEM CELLS (iPSCs)

POORNIMA CHAUHAN
Head of Department Biology
Dewan Public School International
Meerut

ABSTRACT:

Recent research of Takahashi and Yamanaka shows that fibroblasts could be reprogrammed to generate induced pluripotent stem cells (iPSCs). It was concluded that six common reprogramming factors (OCT4, SOX2, KLF4, C-MYC, NANOG, and LIN28) are widely used for the generating the iPS cells. For proteomics analyses, conservation patterns and residues, binding grooves, binding pockets, evaluation of amino acid composition, high scoring hydrophobic segments and tandem repeats have been investigated.

The molecular dynamics and interaction property of LIN28 is studied by using Bioinformatics tools. We used 4A4I.pdb file for the crystal structure of human LIN28. Two identical chains (A & B) was observed which have GLY32, LYS35, PHE37, ASN38, GLY42, VAL62, PHE63, VAL64, HIS65, GLN66, SER76, VAL99, GLY109, SER110 conserved residues. Beta strands are dominating than helix in forming secondary structure. Approximately 40 % of chains residues were involved in the stabilization centers. In our paper various bioinformatics tools are used on 3D structural data that may be helpful to the future researchers to understand stem cell biology, identify new therapeutic targets and docking studies for drug discovery.

INTRODUCTION

Pluripotent stem cells (PSCs) are defined by their potential to generate all cell types of an organism [1]. The cells can be generated by reprogramming readily available somatic cells, such as fibroblasts, into induced pluripotent stem cells (iPSCs), which can replicate indefinitely and give rise to any somatic cell type [2]. Induced pluripotent stem cells (iPSCs) are generated through fibroblasts those have similar pluripotent quality and embryonic stem (ES) cells like property [3,4]. The generation of validated "naïve" human ESCs will allow the molecular dissection of a previously undefined pluripotent state in humans and may open up new opportunities for patient-specific, disease-relevant research [5]

Initially such experiment was performed on mouse; Takahashi et al. (2007) isolated human pluripotent stem cells that resemble human embryonic stem cells by all measured criteria. Transcription factor-based cellular reprogramming has opened the way to converting somatic cells to a pluripotent state, but has faced limitations resulting from the requirement for transcription factors and the relative inefficiency of the process [6]. 24 transcription factor encoded genes were selected that expressed in ES of Mouse and considered as candidates of inducing pluripotency in somatic cells [7]. After screening it was concluded that six common reprogramming factors (OCT4, SOX2, KLF4, C-MYC, NANOG, and LIN28) which are widely used for the generating the iPS cells. The reprogramming technology established by Yamanaka and coworkers have shed a new light [8].

It was show that four factors (OCT4, SOX2, NANOG, and LIN28) are sufficient to reprogram human somatic cells to pluripotent stem cells with the essential characteristics of embryonic stem (ES) cells [9]. This landmark discovery by Takahashi and Yamanaka has opened platform to develop disease specific

iPSCs. Induced pluripotent stem cells (iPSCs), reprogrammed from somatic cells with defined factors, hold great promise for regenerative medicine as the renewable source of autologous cells [10].

This aspect can be used to identify new therapeutic targets and cell replacement therapy for various degenerative disorders [11]. Peripheral blood was collected from a clinically diagnosed 64-year old male multiple schwannoma patient. Peripheral blood mononuclear cells (PBMCs) were reprogrammed with the Yamanaka KMOS reprogramming factors using the Sendai-virus reprogramming system [12]. Peripheral blood was collected from a clinically diagnosed 79-year old male and 72-year old male with sporadic Parkinson's disease and Alzheimer's disease patients respectively. Peripheral blood mononuclear cells (PBMCs) were reprogrammed with the Yamanaka KMOS reprogramming factors using the Sendai-virus reprogramming system [13, 14].

LIN28 gene encodes 'LIN-28 homolog A protein' [15]. Human LIN28A and LIN28B are RNA-binding proteins (RBPs) conserved in animals that play important roles in stem cell reprogramming [16,17]. Lin-28 has two conserved patterns cold shock domain and CCHC zinc fingers domain [18]. Presently the six common reprogramming factors (OCT4, SOX2, KLF4, C-MYC, NANOG, and LIN28) are the central point of attraction for researchers, working on stem cells. In this paper, we have focused on structural analysis of human LIN-28 homolog B with the help of bioinformatics tools.

First of all, physiochemical parameters of LIN28 sequence are calculated that provide valuable information about stability and functionality [19, 20]. Secondary structural elements analysis, hydrophobic segment determination, Conserved residues determination, tandem repeats calculations, post translational modifications analysis and other related parameters prediction depicted structure-function relationship [21]. Identification of conserve part is the most informative link that may codes functional or structural unit of a protein. [22] Several other factors are associated with the molecular dynamics like hydrogen bonds, interaction between protein subunits, hydrophobic segments, charge distribution, stabilization centers, stabilizing residues and solvent accessibility. Stabilization center of a protein and Solvent accessibility are associated with the protein stability [23]. Protein functionality depends on its proper folding that is maintained by a network of interactions between its amino acid residues [24].

MATERIALS AND METHODS

COLLECTION OF DATA

We have collected data of the human LIN28 homolog B from the National Center for Biotechnology Information [25]. Amino acid sequence of LIN28 was retrieved in FASTA format through ENTREZ tool of NCBI with the AAZ38897.1 accession number. The PDB file of LIN28 (4A4I.pdb) was extracted from the protein data bank (www.rcsb.org) for further analysis [26].

Amino Acid Composition and other physiochemical parameters

We used PROTPARAM to calculate physiochemical parameters like Molecular weight, Theoretical pI, Total number of atoms, Total number of negatively and positively charged residues, Atomic composition, Instability index and Grand average of hydropathicity (GRAVY). Statistical analysis of protein sequences (SAPS) tool was used to analyze the amino acid composition, high-scoring hydrophobic segments and tandem and periodic repeats [27].

Charge Distribution Prediction

SAPS tool uses algorithms for SAPS (Statistical Analysis of Protein Sequences) to predict the distribution of charges, high-scoring positive charge segment, high-scoring negative charge segment, high-scoring mixed charge segment and high-scoring uncharged segments.

Secondary Structure Topology Mapping and Disulphide Bond Topology

PDBsum server used to view the topology of secondary structure elements [28, 29]. PDBsum is a brief view of proteins from the Protein Data Bank (PDB) [30]. Using this server we found arrangements of alpha helices and Beta strands. We can predict the interacting chains in three dimensional model and intra H bonding within secondary structural elements. The PDBsum server uses Gail Hutchinson's HERA algorithm to generate hydrogen bond plots to identify the hydrogen bonds in the secondary structure [31]. To explore the secondary structure aspects, such as disulphide bond topology prediction, we used "SCRATCH protein predictor" which has different tools for structural proteomics computational analysis [32].

Interaction studies

PDBsum server also predicted interaction between identical chains of LIN28. the server predicts the various types of interactions between subunit and also finds the residues those actually interacted between interfaces.

Hydrophobic Segment and Conserved Residue Prediction

For hydrophobic segment prediction, we used ProtScale server. This server creates hydrophobicity plots using a sliding window algorithm [33].

Stabilization Centers and Stabilizing Residues Prediction

Stabilization centers were identified and analyzed by using SCide server [34], and the stabilizing residues were identified by using SRide server [35]. We also calculated the stabilization center elements in the sequence. Stabilizing residues were calculated by using the surrounding hydrophobicity, long-range order, stabilization center and conservation score as described previously [36]. We used a conservation score of 6 as the cutoff value to identify the stabilizing residues.

Calculation of Highly Conserved Amino Acids

The conservation patterns of COMT using ConSurf server [37,38] has been developed. The conservation scores at each amino acid position were calculated using the same web server. Highly conserved amino acids from proteins were used for further analysis. This server can calculate the evolutionary conservation of amino acid positions in proteins using an empirical Bayesian inference, starting from protein structure and sequence.

Generation of Surface Cavity

The PyMOL [39] has been used to for generation of surface cavity as well as identification of binding grooves of LIN28. ".pdb" files were used to generate the surface structure and the cavities of the given protein.

Active sites prediction

The Q-Site Finder server was used to identify the potential possible active sites where ligand can bind. This online tool uses energy-based method for generating the protein-ligand binding sites [40].

RESULTS AND DISCUSSION

The prediction results of the amino acid composition of LIN28 are shown in Fig. 1. As shown, the composition of Gly is highest (11.6 %) among all other amino acids. The lowest composition was observed for Tyr (0.8 %) and Trp (0.4 %). Other parameters calculated at the primary level structure of LIN28B are given Table 1. In general, disulphide bond perform various functions in proteins, from maintaining the folding and stability of proteins to preserving bioactive structure essential to exact protein function [41-43]. Disulphide bonds stabilize the native conformation of a protein by destabilizing the unfolded form [44,45]. The software, SCRATCH protein classified the sequence as DOES NOT HAVE disulfide bonds. The

predicted Cysteine pairs and the probable disulphide bonds are given in Table 2. Atomic composition of LIN28B protein is represented in Fig 2

Cysteine predicted at respective positions	Disulphide bonds by decreasing order of probability	
	Cysteine position 1	Cysteine position 2
Total number of cysteines:	187	201
Predicted number of bonds:	154	164
	129	142
Cysteines at the following positions are predicted to form the disulfide bond:	34	107
34,107,129,142,154,164,187,201		

Table 1- Prediction of the cysteine position and the probable disulphide bonds of LIN28B.

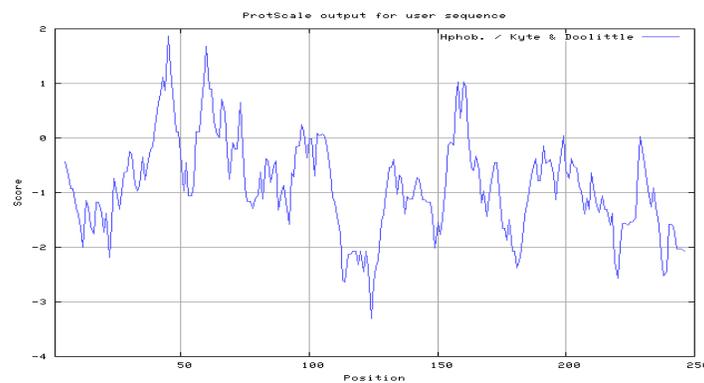


Fig1 Hydrophobicity plot of the amino acid sequence

Physiochemical parameters	Compositional analysis of amino acids	Atomic composition:	Tandem, and periodic repeats
Number of amino acids: 250 Molecular weight: 27083.6 Theoretical pI: 9.15 Formula: $C_{1161}H_{1859}N_{353}O_{363}S_{16}$ Total number of atoms: 3752 Total number of negatively charged residues (Asp + Glu): 28 Total number of positively charged residues (Arg + Lys): 38 The instability index (II): 79.67 Aliphatic index: 45.64 Grand average of hydropathicity (GRAVY): -0.890	A : 12(4.8%); C : 10(4.0%); D : 4(1.6%); E : 24(9.6%); F : 9(3.6%); G : 29(11.6%); H : 8(3.2%); I : 7(2.8%); K : 23(9.2%); L : 11(4.4%); M : 6(2.4%); N : 5(2.0%); P : 24(9.6%); Q : 13(5.2%); R : 15(6.0%); S : 27(10.8%); T : 9(3.6%); V : 11(4.4%); W : 1(0.4%); Y: 2(0.8%);	Carbon C Hydrogen H Nitrogen N Oxygen O Sulfur S	Aligned matching blocks: [120- 123] QKRK [245- 248] QKRK [207- 210] PQEA [222- 225] PQEA

Table 2: Calculated Physiochemical parameters of LIN28B

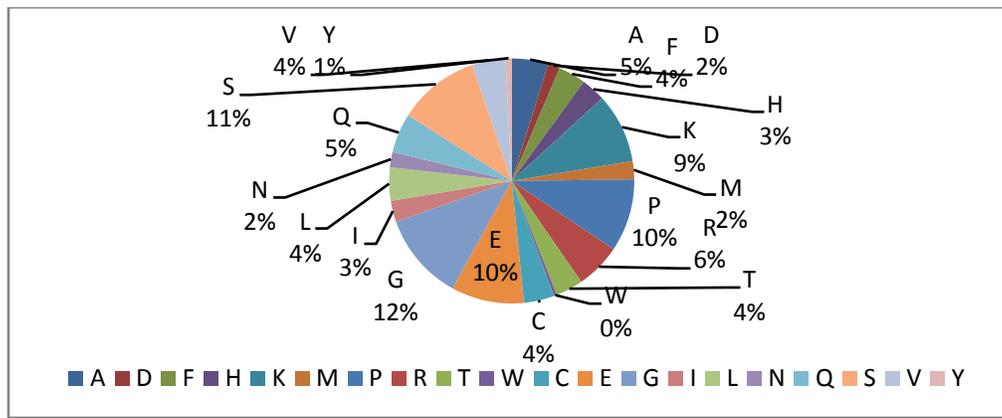


Figure 2 Compositional amino acid analysis of LIN28B

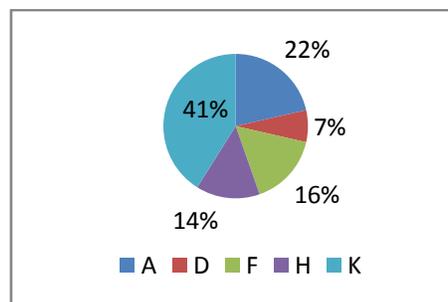


Figure 3 Atomic composition of LIN28B

For the secondary structure topology elements, such as α helices, β sheets, β hairpins, β bulges, β strands, β turns and turns of lin28B, The structure was analyzed through PDBsum. It was found the structure have two identical chains of 82 amino acid long as shown in Fig. 3. Complete structure of LIN28B was not available in PDB. Both chains encoded by **cold-Shock domain positioned from 27th to 110th amino acid.**

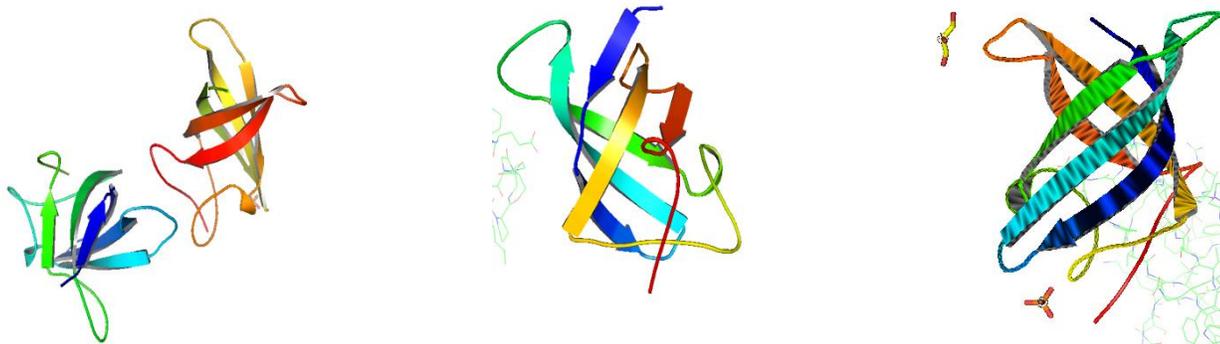


Fig 4 LIN28 structure retrieved from PDB (a) Both chain A and chain B (b) Chain a with the ligands GOI (c) Chain b with SO4 ligand

Secondary structure topology, conserved residues, β sheet topology and hydrogen bond pattern of LIN28.

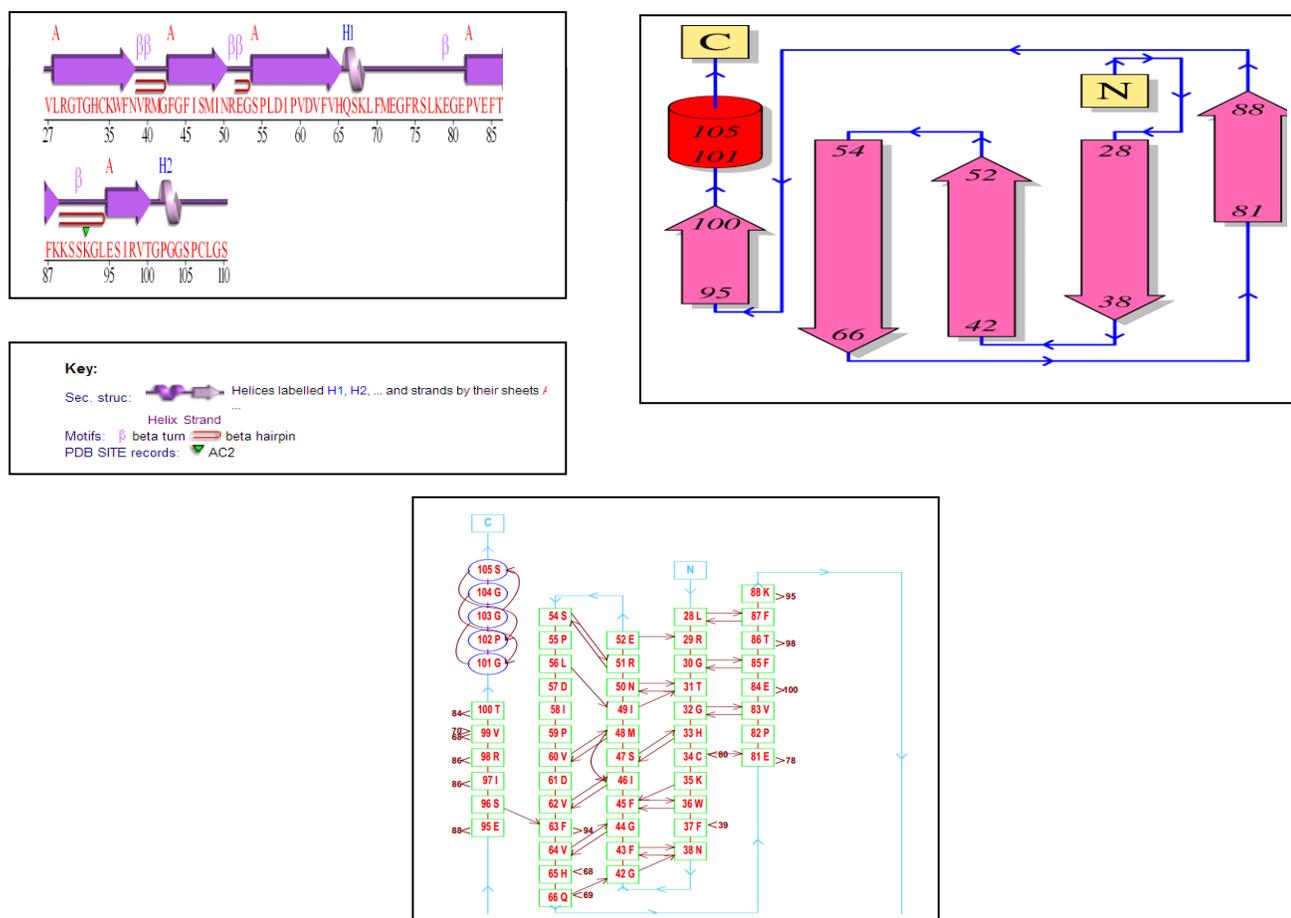


Fig. 5 (a) Secondary structural elements (β sheets, β hairpins, β bulges, β turns, strands, helices) (b) Topology diagram illustrating the β strands, represented by the large arrows, join up, side-by-side, to shape the domain's central β sheet. The figure also illustrates the relative locations of the alpha helices (the red cylinders). The small arrows specify the directionality of the protein chain, from the N- to the C-terminus.

(C) Intra- and inter-hydrogen bond pattern between residues in β sheet topology (Colour figure online)

We calculated α helices, β sheets and β hairpins from the graphical figure and represented in the figure 5. We used PDBsum server to predict the hydrogen bond pattern both in α helices and β sheet within the **cold-Shock domain** domain. Most of the hydrogen bonds were inter hydrogen bond between the beta sheets; although, four intra-hydrogen bonds were observed among residues of α helix. Hydrogen bonds (H-bond)

play a significant role in the formation of three dimensional structures and stabilization [46].It was concluded the structure has dominated by beta sheets, although two helices are also present [fig 5].

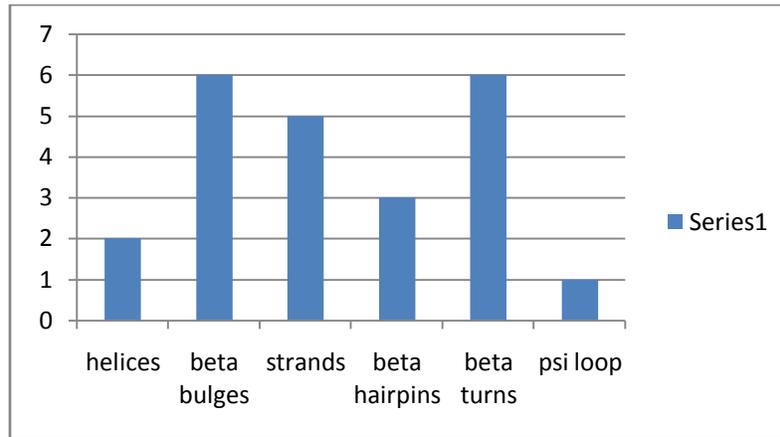
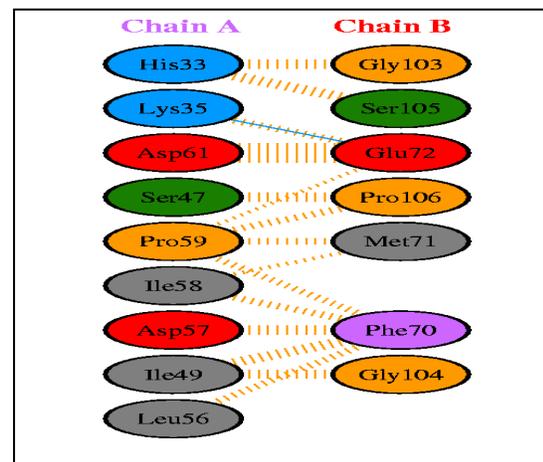
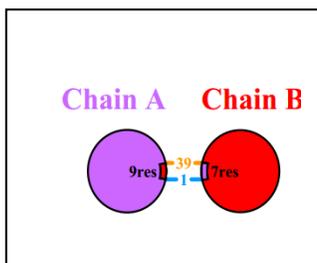


Fig 6: Graphical presentation of secondary structural elements

INTERACTIONS BETWEEN AMINO ACIDS

Figure 6a shows a model of the interaction between the subunits. The model shows that the hydrogen bond and 39 non-bonded contacts are responsible for the interaction. Our model is showing the number of interactions across two interfaces. Figure 6b shows details of the individual residue-residue interactions between the interfaces. The number of residues involved in residue-residue interactions between A-B chain were 33,35,61,47,59,58,57,49,56 and 103,105,72 ,106,71,70,104 respectively. These protein-protein interactions stabilize the native structure of proteins appear to be a novel way for transforming the activities of proteins [47].



Key: Salt bridges (red line), Disulphide bonds (yellow line), Hydrogen bonds (blue line), Non-bonded contacts (orange dashed line)

Fig. 7 a Schematic diagram showing the interactions between the subunits b. A model showing the individual residue–residue interactions

STABILIZATION CENTERS AND STABILIZING RESIDUES

The stabilization centers in LIN28B are shown in Fig. 6. We found that 40 % of domain residues were involved with the stabilization centers. We inferred that these residues might contribute additional stability to CRP the protein. From the prediction of stabilizing residues, we found that seven stabilizing residues were available in LIN28 (Table). Among the stabilizing residues, highest conservation score (9) was observed in CYS34, GLY44 and PHE45 and their surrounding hydrophobicity (Hp) were 25.33, 26.16 and 23.26, respectively. Highest surrounding hydrophobicity (Hp) was observed in GLY44 (26.16). Computer prediction and modeling of protein structures play a dominant role in stabilizing the native structures [48-50]. This information will also be important for protein engineering and to unfold the mechanism of the structure–function relationships [51].

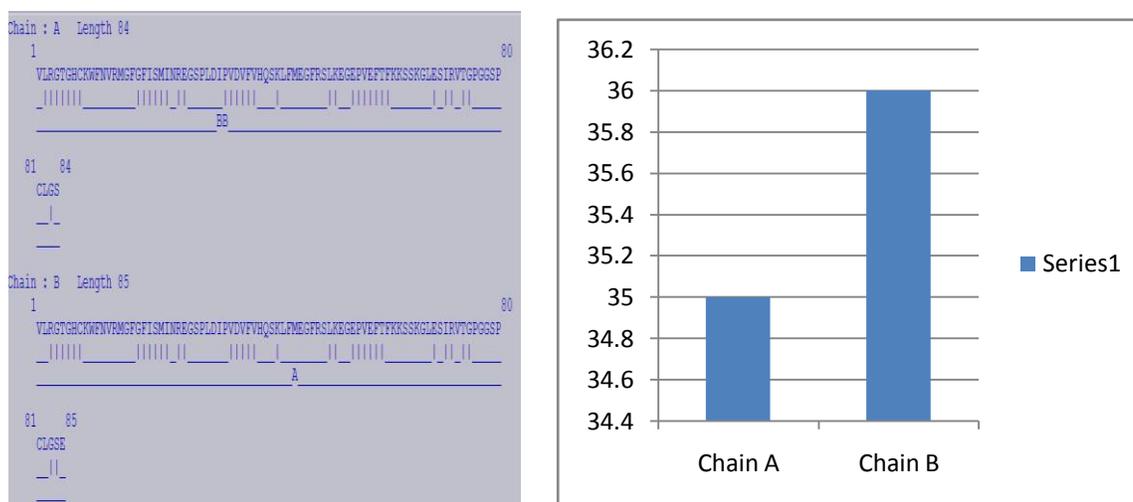


Fig. 8 Stabilization center prediction of LIN28B. a Stabilization center in chains A and B. Location of stabilization center elements in the sequence is marked with vertical line, while other positions are marked with horizontal line. b Total number of residues associated with stabilization center

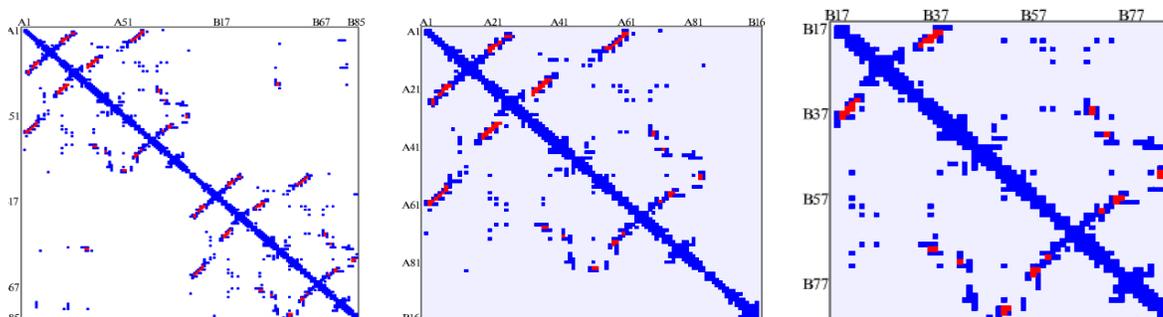


Fig 9 Graphical presentation of interaction between chain A & B in LIN 28B a.blue colour shows interaction and red colour shows residues in stabilization center b.Chain A c. Chain B

Residue	Cons score	H _p
CYS34	9	25.33
GLY44	9	26.16
PHE45	9	23.26
ILE46	8	21.75
VAL62	8	20.91
VAL64	8	24.17
THR100	8	22.27

Table 2 Stabilizing residues in LIN28B

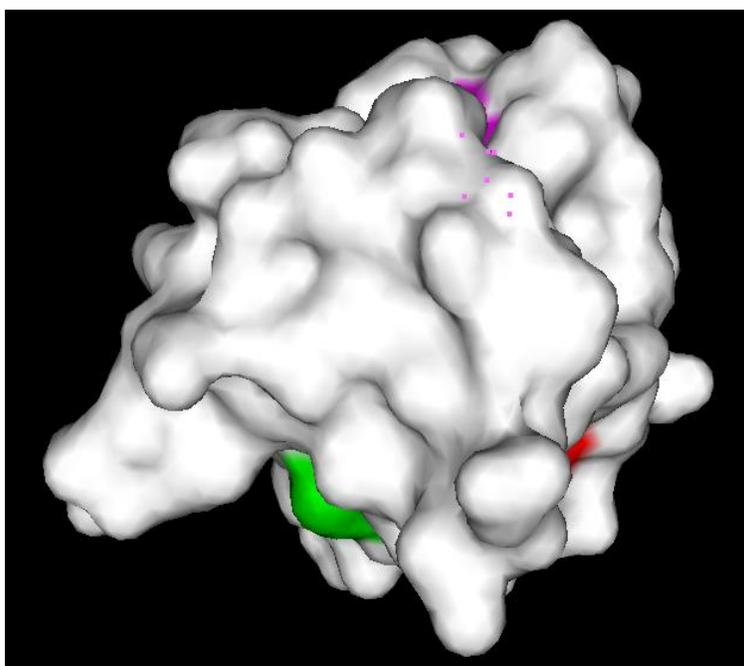


Fig 10 Surface cavities of LIN28B chain A generated through PyMol. C34residue shown as red, GLY44, PHE45, ILE46 as green, VAL62, VAL64 as yellow, THR as purple

HIGHLY CONSERVED AMINO ACIDS

ConSurf server predicts conserve residues through Bayesian algorithm. MAFFT was used for Multiple Sequence Alignment in ConSurf server. Homologues were searched through CSI- BLAST (E-value: 0.0001) with 3 Iterations. There were 480 CSI-BLAST hits, 341 of them are unique sequences fig[9].The calculation was performed on the 150 sequences with the lowest E-value and the most conserved residues with score 9 were highlighted with purple color (table 3).



Fig 11 ConSurf Color Coded Multiple Sequence Alignment

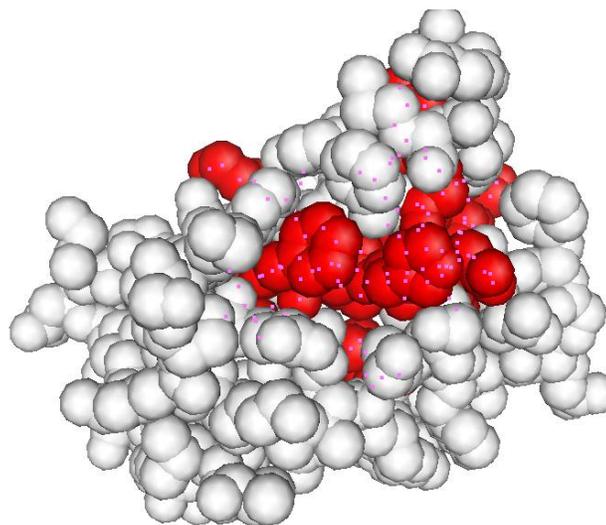


Fig 12 Lin28B three dimensional structure of chain A, Conserved residues shown in red color

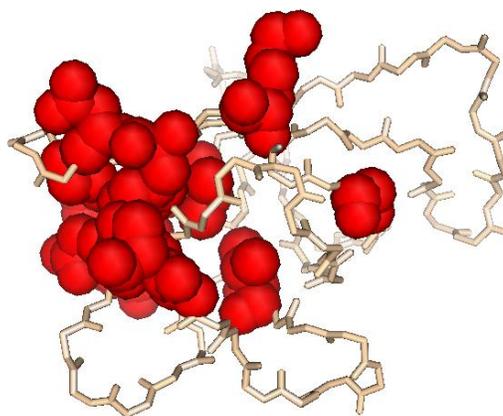


Fig 13b Backbone structure of Lin28B chain A with highly conserved amino acids

REFERENCES

1. Müller, F.J., Schuldt, B.M., et al (2011). *A bioinformatic assay for pluripotency in human cells*
2. Xie, M., Tang, S., Li, K., Ding, S. (2017). Pharmacological Reprogramming of Somatic Cells for Regenerative Medicine. *Acc Chem Res*(Vol 50(5))pp.1202-1211).
3. Takahashi K, Tanabe K, Ohnuki M, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007; 131(5):861–72.
4. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006;126(4):663–76.
5. Hanna, J. et al. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci U S A* (Vol. 107(20), pp.9222-7).
6. Anokye-Danso, F. et al (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* (Vol 8(4),pp.376-88).
7. Zaehres H, Scholer HR. Induction of pluripotency: from mouse to human. *Cell* 2007;131(5):834–5.
8. Matsuda, Y., Semi, K., Yamada, Y. (2014). Application of iPSC cell technology to cancer epigenome study: Uncovering the mechanism of cell status conversion for drug resistance in tumor.
9. Tada, M, Takahama Y, Abe K, Namatsuji N, tada T (2001). Nuclear reprogramming of somatics cells by in vitro hybridization with ES cells, *Curr boil* 11;1553-1558
10. Zhao T¹, Zhang ZN, Rong Z, Xu Y. Immunogenicity of induced pluripotent stem cells.(2011). *Nature* (Vol 474(7350))pp.212-5)
11. Yu, J., Vodyanik, M.A., Smuga-Otto K, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007;318(5858):1917–20.
12. Zhang, S., Lv, Z., Hu, Y., Liu, L., Gong, W., Li, Q., Wu, H.(2017). Generation of a human induced pluripotent stem cell (iPSC) line from a 64year old male patient with multiple schwannoma. *Stem Cell Res* (Vol 19,pp.34-36).
13. Zhang, S., Lv, Z., Zhang, S., Liu, L., Li, Q., Gong, W., Sha, H., Wu, H.(2017). Characterization of human induced pluripotent stem cell (iPSC) line from a 72year old male patient with later onset Alzheimer's disease. *Stem Cell Res* (Vol 19,pp.34-36).
14. Zhang S, Liu L, Hu Y, Lv Z, Li Q, Gong W, Sha H, Wu H.(2017). Derivation of human induced pluripotent stem cell (iPSC) line from a 79 year old sporadic male Parkinson's disease patient. *Stem Cell Res* (Vol 19,pp.43-45).
15. Das A K, Pal R.(2010). Induced pluripotent stem cells (iPSCs): the emergence of a new champion in stem cell technology-driven biomedical applications. *J Tissue Eng Regen Med.* (Vol 4(6),pp.413-21).
16. Moss EG, Tang L.(2003). Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev Bio.*(Vol 258(2).pp.432–42).
17. Hafner M, Max E.A, Bandaru P, Morozov P, Gerstberger S, Brown M, Molina H, Tuschl T. Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. *Advance March 12, 2013*
18. Moss, E.G., Lee, R.C., Ambros, V. the cold shock domain protein LIN-28 controls developmental timing in *C.elegans* and is regulated by the *lin-4* RNA. *cell* 2007; 88:637-646.
19. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; *Protein Identification and Analysis Tools on the ExPASy Server*; (In) [John M. Walker \(ed\): The Proteomics Protocols Handbook, Humana Press \(2005\).](#) pp. 571-607
20. Walsh CT, Garneau-Tsodikova S, Gatto JR (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* 44: 7342–7372
21. Kirnarsky L, Nomoto M, Ikematsu Y, Hassan H, Bennett EP, et al. (1998) Structural analysis of peptide substrates for mucin-type O-glycosylation. *Biochem* 37: 12811–7.v
22. Tokuriki, N., Stricher, F., Serrano, L., & Tawfik, D. S. (2008). How protein stability and new functions trade off. *PLoS Computational Biology*, 4, e1000002.
23. Manning, J. R., Jefferson, E. R., & Barton, G. J. (2008). The contrasting properties of conservation and correlated phylogeny in protein functional residue prediction. *BMC Bioinformatics*, 9, 51.
24. Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., et al. (2011). Database resources of the national center for bio- technology information. *Nucleic Acids Research*, 39, D38–D51.
25. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., et al. (2000). The protein data bank. *Nucleic Acids Research*, 28, 235–242.
26. Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B. E., & Karlin, S. (1992). Methods and algorithms for statistical analysis of protein sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 2002–2006
27. Laskowski, R. A. (2001). PDBsum, summaries and analyses of PDB structures. *Nucleic Acids Research*, 29, 221–222.
28. 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.
29. Laskowski, R. A. (2009). PDBsum new things. *Nucleic Acids Research*, 37, D355–D359.

30. Hutchinson, E. G., & Thornton, J. M. (1990). HERA, a program to draw schematic diagrams of protein secondary structures. *Proteins*, 8, 203–212.
31. Cheng, J., Randall, A. Z., Sweredoski, M. J., & Baldi, P. (2005). SCRATCH, a protein structure and structural feature prediction server. *Nucleic Acids Research*, 33, W72–W76.
32. Roseman, M. A. (1988). Hydrophilicity of polar amino acid side- chains is markedly reduced by flanking peptide bonds. *Journal of Molecular Biology*, 200, 513–522
33. Doszta ́nyi, Z. S., Magyar, C. S., Tusna ́dy, G. E., & Simon, I. (2003). SCide, identification of stabilization centers in proteins. *Bioinformatics*, 19, 899–900.
34. Magyar, C., Gromiha, M. M., Pujadas, G., Tusna ́dy, G. E., & Simon, I. (2005). SRide, a server for identifying stabilizing res- idues in proteins. *Nucleic Acids Research*, 33, W303–W305.
35. Gromiha, M. M., Pujadas, G., Magyar, C., Selvaraj, S., & Simon, I. (2004). Locating the stabilizing residues in (alpha/beta)₈ barrel proteins based on hydrophobicity, long-range interactions, and sequence conservation. *Proteins*, 55, 316–329.
36. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., & Ben-Tal, N. (2010). *Nucleic Acids Research*, 38, W529– W533.
37. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., & Ben-Tal, N. (2003). *Bioinformatics*, 19, 163–164.
38. DeLano, W. L. (2002). The PyMOL molecular graphics System.
39. Peitsch MC. Protein modeling by E-mail. *Biol Technol*. 1995; 13:658-60.
40. Fahnoe, D. C., Johnson, G. D., Herman, S. B., & Ahn, K. (2000). Disulfide bonds in big ET-1 are essential for the specific cleavage at the Trp21-Val22 bond by soluble endothelin converting enzyme-1 from baculovirus/insect cells. *Arch Biochem Biophys*, 373, 385–393.
41. Van Vlijmen, H. W. T., Gupta, A., Narasimhan, L. S., & Singh, J. (2004). A Novel Database of Disulfide Patterns and its Application to the Discovery of Distantly Related Homologs. *J Mol Biol*, 335, 1083–1092.
42. Betz, S. F. (1993). Disulfide bonds and the stability of globular proteins. *Protein Sci*, 2, 1551–1558.
43. Perlman, J. H., Wang, W., Nussenzveig, D. R., & Gershengorn, M. C. (1995). A disulfide bond between conserved extracellular cysteines in the thyrotropin-releasing hormone receptor is critical for binding. *J Biol Chem*, 270, 24682–24685
44. Thornton, J.M. (1981) Disulphide bridges in globular proteins. *J. Mol.Biol.* 151, 261–287
45. Wetzel, R. (1987) Harnessing disulfide-bonds using protein engineering. *Trends Biochem. Sci.* 12, 478–482
46. Politi, R., Harries, D. (2010). Enthalpically driven peptide stabilization by protective osmolytes. *Chem Commun (Camb)* 46, 6449-6451.
47. Bourgeas, R., Basse, M.-J., Morelli, X., & Roche, P. (2010). Atomic analysis of protein–protein interfaces with known inhibitors, the 2P2I database. *PLoS ONE*, 5, e9598.
48. Abkevich, V. I., Gutin, A. M., & Shakhnovich, E. I. (1995). Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. *The Journal of Biological Chemistry*, 270, 460–471.
49. Mirny, L. A., & Shakhnovich, E. (1996). How to drive a protein folding potential? A new approach to an old problem. *The Journal of Biological Chemistry*, 271, 1164–1179.
50. Bahar, I., & Jernigan, R. L. (1997). Inter-residue potentials in globular proteins and the dominance of highly specific hydro- philic interactions at close separation. *The Journal of Biological Chemistry*, 272, 195–214.
51. Doszta ́nyi, Z., Fiser, A., & Simon, I. (1997). Stabilization centers in proteins, identification, characterization and predictions. *The Journal of Biological Chemistry*, 272, 597–612.