Structural Analysis & Protein-Protein Interaction of Sox2 & Sox6 to Study the Regulatory Mechanism of Pluripotency & Differentiation: A Bioinformatics Approach

[1]Abhishek Saini, [2] Jai Gopal Sharma, [3]Vimal Kishor Singh

[1]Research Scholar, [2]Professor, and Head, [3]Associate Professor
[1&2]Stem Cell Research Laboratory, Department of Biotechnology,
[1&2]Delhi Technological University, Delhi, India
Corresponding Author-[3] Tissue Engineering and Regenerative Medicine Laboratory, Department of Biomedical Engineering,
Amity School of Engineering and Technology (ASET), Amity University Haryana,
Amity Education Valley, Gurgaon (Manesar) – 122 413, Haryana

Abstract: In 2006, Takahashi and Yamanaka introduced a breakthrough by generating pluripotent stem cells having properties related to Embryonic Stem Cells (ESCs), by the introduction of four factors (Oct4, Sox2, Klf4, and c-Myc) into mouse fibroblasts and termed as induced pluripotent stem cells (iPSCs). Since these cells are pluripotent, they can serve as a potential source for the generation of differentiated cells of a specific type from somatic cells of any other kind by employing right set of cytokines along with the appropriate media which make them a precious asset in the fields of regenerative medicine and disease modeling. Many researchers have reported Sox2 to be having essential roles in the regulation of development of an embryo and cell fate determination. The primary function of Sox2 is the maintenance of pluripotency or stemness. Researchers have also found that a transcription factor from same protein family, Sox6 plays a role in the regulation of the differentiation of pluripotent cells to an erythroid lineage which is opposite to the role of Sox2 to maintain stemness. So, an exciting line of work can be whether Sox6 inhibits Sox2 at later stages which decrease the expression level of Sox2 and hence, the pluripotent cells differentiate to erythroid lineage. Our work focuses on studying the interaction of Sox2 with Sox6 to explore the underlying mechanism for the regulation of pluripotency and differentiation. The 3D structures of the proteins Sox2 and Sox6 were predicted by iTasser resulting in C-scores of -2.26 and -0.76 respectively. The Ramachandran plots of both structures were predicted by RAMPAGE. The protein-protein docking of Sox2 and Sox6 was carried out by ClusPro which revealed the interactions between both proteins. The lowest energy score of the docked complex was -1460.4 kcal/mol while being -1218.6 kcal/mol at the center of the complex. The interaction between both the proteins as revealed by docking defines the complex to be having a high number of the residues involved in hydrogen bonding and hydrophobic interactions which can further be utilized to explore the underlying mechanism which allows these proteins to regulate the cell fate.

***IndexTerms* – Sox2, Sox6, pluripotency, protein-protein interaction.**

I. INTRODUCTION

Induced pluripotent stem cells (iPSCs) were first generated by the introduction of primary four reprogramming factors (Oct4, Sox2, Klf4, and c-Myc) into the fibroblasts required to be reprogrammed (Takahashi et al. 2006). These different factors were selected by their roles in reprogramming and differentiation. These factors work together in cooperation to bring back the primitive stage of somatic cells where those cells were still pluripotent and hence were able to differentiate into different cell types (Singh et al. 2015a). All of these factors play a distinct and vital role in the regulation of reprogramming, but our primary work focuses on the study of mechanistic insight into the role of Sox2 which is SRY (Sex-determining Region Y)-box 2, a transcription factor essential for the maintenance of self-renewal and pluripotency of embryonic stem cells. Sox2 is also required for epiblast development. Sox6 is another protein belonging to the SRY-related High-Mobility-Group box family of transcription factors, which plays a primary role in the regulation of cell-fate specification for many types of cells. The primary function of Sox6 lies in the differentiation of Neural Progenitor cells to neurons. But recently, Sox2 has been found to enhance the induction of erythroid differentiation in primary erythroid cultures generated from human CD34+ cells and erythroleukemic cell line K562 (Cantu et al. 2011). The roles that Sox2 and Sox6 play in the self-renewal & pluripotency and differentiation for determination of cell fate are apparently opposed to each other. Hence we hypothesize that Sox2 and Sox6 interact with each other to regulate the process of differentiation. It may occur by the inhibition of Sox2 by Sox6 which results in the inhibition of the maintenance of pluripotency and hence the expression of Sox6 can induce the differentiation of stem cells.

Prediction of protein-protein interactions with the help of bioinformatics tools is an advantageous technique for the exploration of interactions between proteins which helps in decreasing the burden of prophecy by conducting wet lab experiments. The interactions predicted can be further validated with the help of tests conducted in the laboratories. Hence, in our study, we use this technique to find the insights of the interactions between the proteins Sox2 and Sox6

We first carried out the 3D structure prediction using the denovo method and further evaluated for stability of the structures. Sox2 was docked with Sox6 to analyze the possible interactions between them. Resulting docked complex was analyzed though DIMPLOT to define the mechanistic interaction and various residues involved in the interaction.

II. MATERIAL & METHODS

(i) Protein interaction pathway analysis using STRING

The protein interaction pathways of Sox2 and Sox6 were predicted using STRING. It is a tool which predicts possible interactions of a protein by data from experiments, text mining, homology, microarray, etc. The interaction pathways help to predict the protein which can interact with specific proteins of interest which in our case are Sox2 and Sox6 (Szklarczyk et al., 2015).

(ii) Sequence retrieval of Sox2 and Sox6 and their analysis

The sequence of the protein Sox2 and Sox6 were retrieved from NCBI from Accession numbers NP_003097 and AAH47064 respectively. The sequences were aligned to protein sequences in the database of PDB using BLAST only to know that there were no suitable templates available to carry homology modeling to predict the 3D structures of both the proteins. So, to predict the tertiary structures of Sox2 and Sox6, modeling was carried out using iTasser (Iterative Threading ASSEmbly Refinement) (Yang et.al., 2015)

(iii) Prediction of 3D structures of Sox2 and Sox6 and their validation

iTasser was used to predict tertiary structures of these proteins; It is an online server to predict the structure and functions of proteins with high quality by using the amino acid sequence. iTasser gives top 5 predicted models from which the top models are the best-predicted model, and hence it was used further. The top models of Sox2 and Sox6 were selected for further study, and their validation was done. Ramachandran plots of both the protein were predicted by RAMPAGE. An overall structure quality of the models was predicted by ERRAT (Lovell et. Al.,2003 and Colovos et. al., 1993).

(iv) Protein-protein docking

To study the mechanistic interaction of Sox2 and Sox6, the protein-protein interaction of these proteins were carried out using ClusPro. It gives 100 complexes. As a result, the top model of which was further analyzed in our study. The results include the lowest energy of the whole complex and the energy at the center (Comeau et. al., 2004).

(v) Interaction analysis

The interaction of the docked complex retrieved from ClusPro was plotted by DIMPLOT server available in Ligplus (LigPlot+) which is a graphical front-end to LIGPLOT as well. DIMPLOT is used to plots interactions between two proteins across protein-protein or domain-domain interface (Wallace et. al., 1995).

III. RESULTS AND DISCUSSION

(i) Protein interaction pathways analysis using STRING

The protein interaction pathways of Sox2 and Sox6 were predicted by using STRING. In the interaction pathway of Sox6, the list of interactants included Sox2 which has been shown in Fig 1.

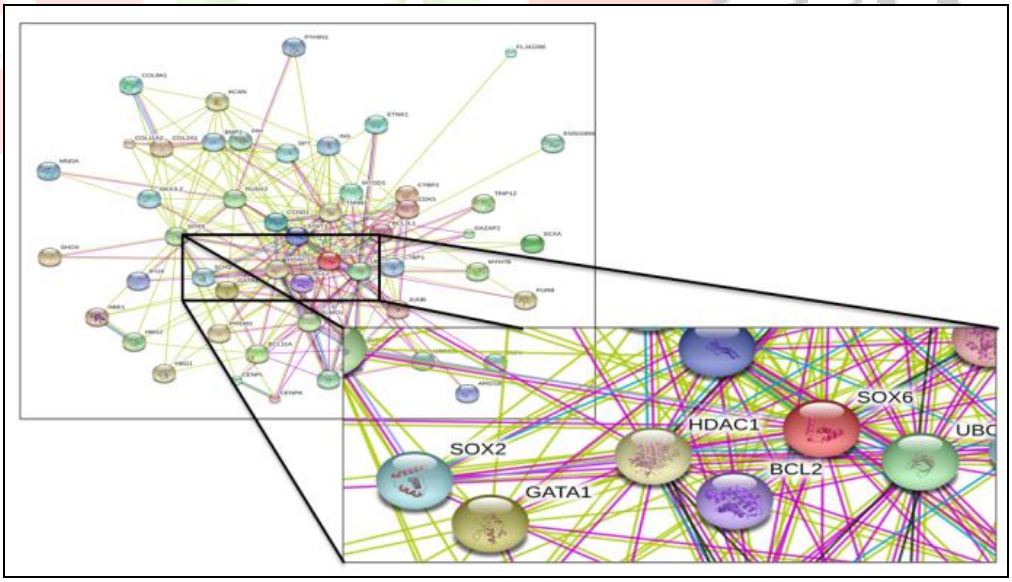

Figure 1: Interaction pathway as obtained from STRING has been zoomed to highlight the interaction between Sox2 and Sox6.

(ii) Structure prediction and validation

The top models from the structures predicted by iTasser (Fig 2) were further evaluated with the tools RAMPAGE and ERRAT. RAMPAGE predicts the Ramachandran Plot of the protein whereas ERRAT gives an overall quality of the predicted structure. For Sox2 structure, RAMPAGE predicted 231 (73.3%) residues, 59 (18.7%) residues and 25 (7.9%) residues in favored, allowed and outlier region respectively in the Ramachandran Plot of Sox2 as shown in Fig 3A. An ERRAT score of 93.793 indicates Sox2 structure with very high quality. 3D structure of Sox6 was predicted to be having

523 (63.3%) residues, 204 (24.7%) residues, and 99 (12.0%) residues in favored, allowed and outlier regions respectively as shown in Fig 3B. An ERRAT score of 77.065 indicates a good quality structured Sox6 model.

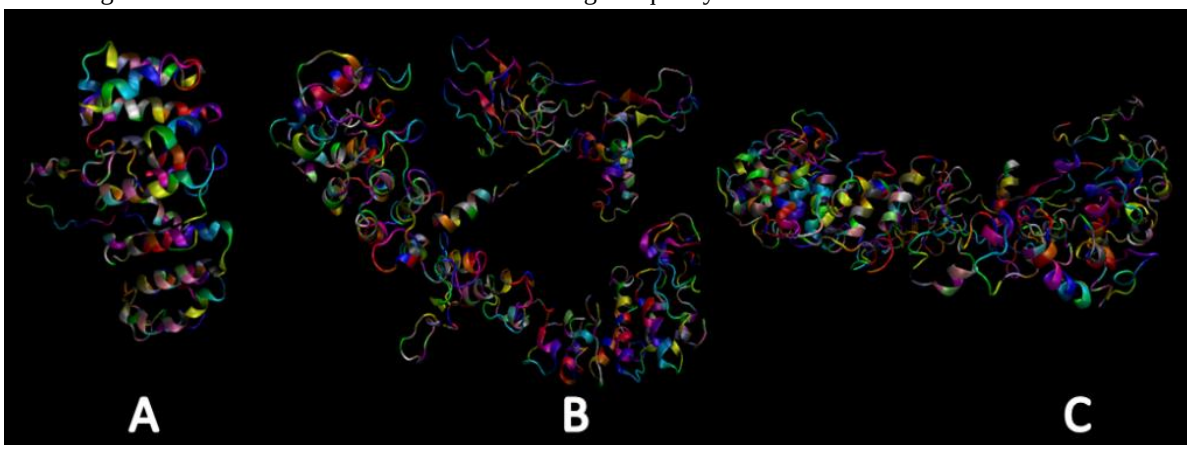

Figure 2: A) Predicted structure of Sox2; B) Top view of the predicted structure of Sox6 and C) Side view of Sox6

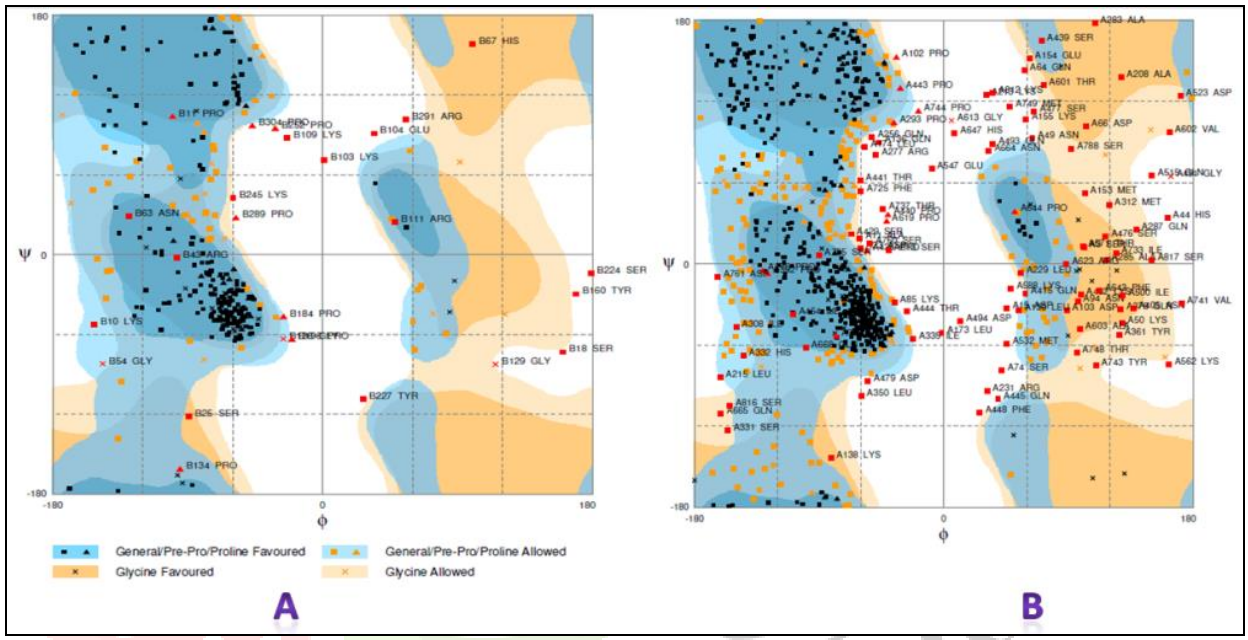

Figure 3: A) Ramachandran Plot of Sox2 and B) Ramachandran Plot of Sox6 as predicted from the tool RAMPAGE.

(iii) Protein-Protein Docking

The top result of the docking from ClusPro, which is the best-docked confirmation between two proteins, was selected. After docking of Sox2 and Sox6, energy at center and lowest energy of the whole complex was found to be -1218.6 and -1460.4 respectively which represent a docking score with a high binding affinity.

(iv) Interaction analysis

The interaction between Sox2 and Sox6 was plotted using DIMPLOT. The data including specific residues forming H-bonds and the residues involved in hydrophobic interactions have been mentioned in the table. The residues involved in the formations of H-bonds have also been labeled in the docked complex and shown as a figure of the interaction site between both the proteins in Fig 4.

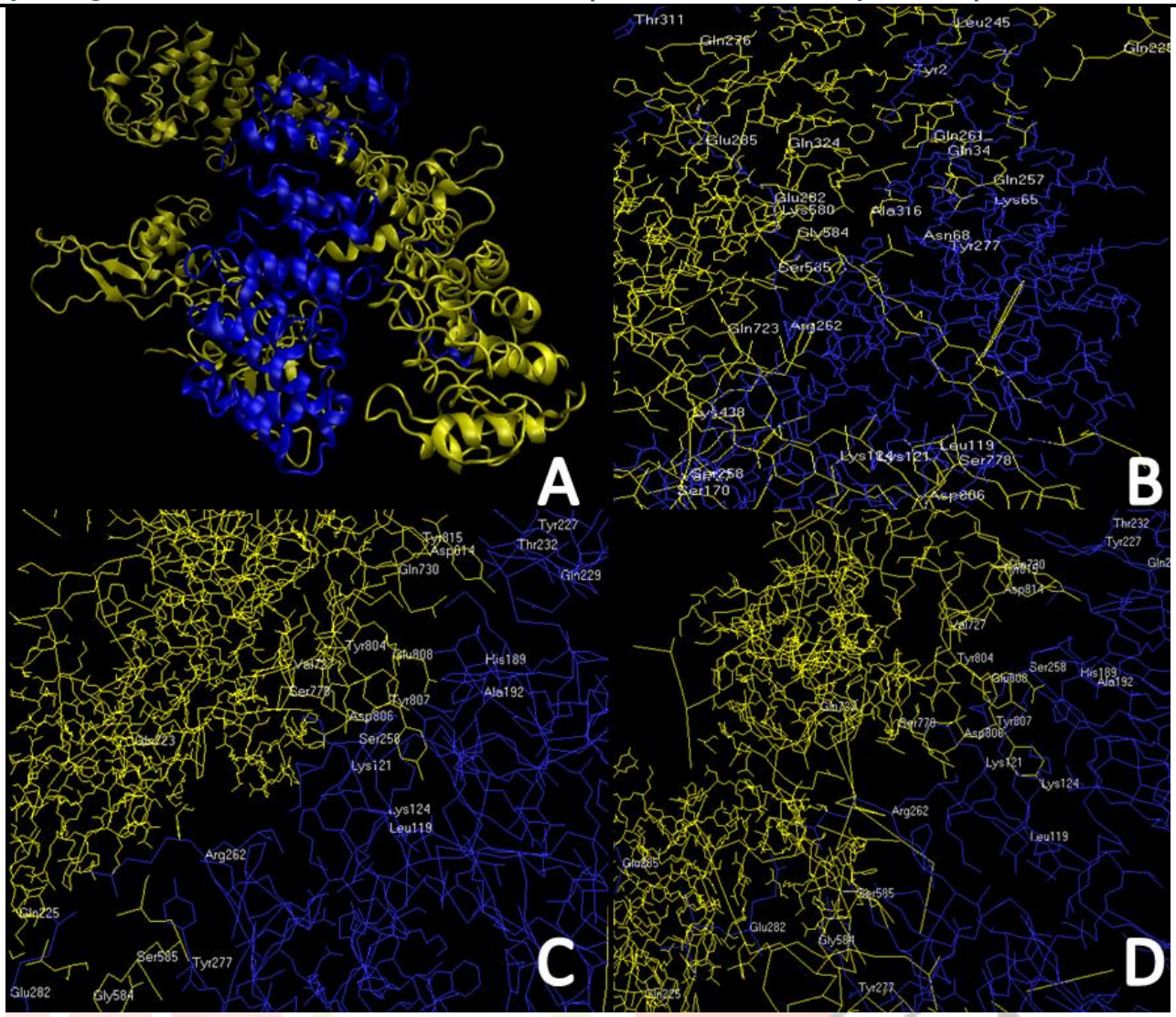

Figure 4: The docked complex of Sox2 and Sox6 as given from ClusPro; B), C), and D) the residues from the interface of the complex have been labeled and shown from different views. The Sox2 protein has been shown as blue and Sox6 as yellow.

In the Sox2-Sox6 complex, 28 H-bonds were detected. A total of 21 residues from Sox2 & 22 residues from Sox6 were found to interact with each other via hydrogen bonding, and 29 residues from Sox2 & 33 residues from Sox6 were found to be involved in hydrophobic interactions which have been listed in Table 1. Analysis of the interaction between both the proteins indicates a perfect binding between both them by which we can say that Sox2 and Sox6 may also interact with each other in real life scenario.

Table 1: The specific residues which play a role in the interaction between both the protein have been mentioned in the above table. These specific residues include the ones involved in H-bonding and hydrophobic interactions.

Residues involved in H-bonding		**Residues involved in Hydrophobic Interactions**	
Sox2	**Sox6**	**Sox2**	**Sox6**
Gln287, Tyr2, Gln34, Ala30, Lys65, Ser258, Ser170, Thr311, Tyr277, Asn68, Leu119, Lys124, Lys121, His189, Ala192, Tyr227, Gln229, Thr232, Ala281, Glu282, Arg262,	Leu245, Gln261, Gln257, Gln225, Val727, Lys438, Gln276, Gln259, Ala316, Tyr807, Asp806, Ser778, Tyr804, Glu808, Asp814, Tyr815, Gln730, Lys580, Gln324, Gly584, Ser585, Gln723	Met4, Met1, Asn33, Tyr171, Gly32, Ala29, Gly31, His 67, Pro64, Pro134, His101, Met102, Arg96, Ala99, Gly310, Leu205 , Ala248, Gln206, Ser249, Pro 313, Lys35, Asn36, Ser167, Gly136, Met157, Arg156, Gln155, Met120, Tyr160, Ala191, His316, Ala133, Gly190, Ser228, Met194, His198, Val283, Gly280, Ala263	Leu279, Leu244, His249, Met268, Phe307, Ser439, Ile258, Gln256, Ser360, Phe725, Thr726, Tyr361, Met312, Met582, Leu320, Asn582, Asp275, Thr278, His274, Ala313, Ala314, Ala315, Pro440, Met803, Gln240, His359, Pro731, Ala280, Asp575, Ala587, Leu589, Lys588, Ala586

IV. CONCLUSION

The 3D structures of Sox2 and Sox6 after prediction via ab initio modeling and validation were docked with each other. Their docking results indicated a very good binding score and a high number of residues are involved in hydrogen bonding and hydrophobic interactions. This analysis proves our hypothesis of a protein-protein interaction between Sox2 and Sox6 to be true to a certain level, but for further validation of this point, more approaches including wet lab protein-protein interaction studying experiments need to be conducted.

V. ACKNOWLEDGMENT

We are thankful to the Honorable Chairman and the Honorable Vice- Chancellor of the Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi-42, for support. Corresponding author particularly thanks the Department of Science and Technology and Indian National Science Academy (INSA), INDIA, for the research grant.

REFERENCES

[1] Takahashi, K. and Yamanaka, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126 (4), 663–76, doi:10.1016/j.cell.2006.07.024.
[2] Singh, V.K., Kumar, N., Kalsan, M., Saini, A., Chandra, R. 2015. Mechanism of induction: Induced Pluripotent Stem Cells (iPSCs). Journal of Stem Cells. Vol10. Issue 1, 43-62.
[3] Cantù C, Ierardi R, Alborelli I, Fugazza C, Cassinelli L, Piconese S, et.al. 2011. Sox6 enhances erythroid differentiation in human erythroid progenitors. Blood: 117 (13).
[4] Damian Szklarczyk, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, Alexander Roth, Alberto Santos, Kalliopi P. Tsafou, Michael Kuhn, Peer Bork, Lars J. Jensen, and Christian von Mering. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. Jan 28; 43(Database issue): D447–D452. doi: 10.1093/nar/gku1003
[5] Jianyi Yang, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, Yang Zhang. 2015. The I-TASSER Suite: Protein structure and function prediction. Nature Methods, 12: 7-8.
[6] S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson 2002 Structure validation by Calpha geometry: phi, psi, and Cbeta deviation. Proteins: Structure, Function & Genetics. 50: 437-450.
[7] Colovos C, Yeates TO. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. Sep; 2(9):1511-9.
[8] Stephen R. Comeau, David W. Gatchell, Sandor Vajda, Carlos J. Camacho. 2004. ClusPro: an automated docking and discrimination method for the prediction of protein complexes, Bioinformatics, Volume 20, Issue 1, 1, Pages 45–50, https://doi.org/10.1093/bioinformatics/btg371
[9] Wallace, A.C., Laskowski, R. A., Thornton, J.M. 1996. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng., 8, 127-134.