Deciphering Molecular Interactions of viral oncogene E-7 in human genital papillomavirus (HPV) for cancer therapy

Sabahat Sohail[1*], Naureen Aslam Khattak[2], Syeda Naqsh Zahra[2,] Raja Adnan Tahir[2]

1. Sabahat Sohail (Corresponding Author)

Department of Basic Sciences. PMAS-Arid Agriculture University, Muree Road, Rawalpindi, Pakistan.

Sabahat.sohail25@gmail.com

2. Naureen Aslam Khattak[2], Syeda Naqsh Zahra[2]Raja Adnan Tahir[2]

Department of Bioinformatics and Biotechnology, International Islamic University Islamabad, H-10 Sector Islamabad, Pakistan

naureen.aslam@uaar.edu.pk

naqsh.z@gmail.com

Abstract

Human papillomaviruses (HPV) cause almost 90% of cervical cancers in associated with E6 and E7 proteins. Tumour supressor protein p53 and pRB becomes inactivated by E6 and E7 viral oncogenes respectively. E6/E7 oncoproteins are essential for the neoplastic growth maintenance and their repression stop the cell proliferation and causes the cells to undergo senescence. MODELLER (9v10) was utilized to generate three dimensional structures of E-7 candidate gene. Protparam server were applied for physiochemical analysis. AutoDock utilized for protein ligand docking and PatchDock for protein-protein docking. Discovery studio, chimera , ligplot were explored for visualization of receprot-ligand interactions and Pymol for protein protein functional analysis. In current study, 3-dimensional structure of E7 protein was proposed which showed 75% sequence homology with template 2EWL with e-value of 5e-24. Detailed molecular studies of receptor-ligand and protein-protein docking were employed to explore binding interactions of E7. Ligand (KUC104527N) binding with E7 substrate revealed binding interactions with polar (K73) and non- polar (I72, L74, V76, F86, F90, L94 and F96) amino acid residues with estimated free energy of binding -4.69 kcal/mol , Inhibition contact 366.88 μM, and torsion free energy 1.37 kcal/mol respectively. While in protein-protein docking E7 showed six hydrogen bonds between E7 and its functional partner KRT14 involving Q57,S79, V 75, S95 and E77 as most interacting residues. Current research may facilitate 3D structural insights of E7 oncoprotein, to efficiently reveal the functional aspect of candidate gene . Furthermore inhibitory mechanism of ligand was analyzed for designing novel drug against HPV.

Keywords: HPV, E7 protein, Cervical cancer, Molecular docking.

Introduction

Human genital papillomaviruses (HPVs) are frequently involved in causing cervical cancers. It has several types among which HPV types 16, 18, 31 and 33 are more frequently found in high grade lesions and carcinomas, whereas types 6 and 11 are involved in condylomata and low grade cervical intraepithelial lesions. In the majority of cervical carcinomas, DNA of HPV integrate into the genome, which usually occurs within the E1 or E2 open reading frames (ORF) and most oftenly cause deletions in this region [23].

E7 oncoprotein of HPV plays an important role in carcinogenic transformation and life cycle of the HPV virus by binding to pRb tumor suppressor protein and E2F transcription factors leading to pRb dissociation from E2F transcription factors and the premature progression into the cell cycle [10]. It disrupts the two key events, cellular differentiation and proliferation in normal epithelium, causing the virus to replicate itself in cells that are no longer in the dividing state [14].

Due to recent advances in sequencing of genomes of various organisms, researchers have to deal with a huge amount of raw data which is too much laborious and time consuming. In order to help the researchers, field of Bioinformatics emerge which is application of computational techniques to understand and organize the information associated with biological macromolecules [25]. It organizes and analyses data for access to existing information and to submit new data [3]. For raw DNA sequences, this field helps in identification of exons, introns and promoter regions and separating coding and non-coding regions for annotating genomic DNA [28]. Investigation for protein sequences include development of sequence comparison algorithms [4] techniques for multiple sequence alignment, and searching functional domains from conserved sequence motifs in such alignments. Structural data analysis involve the secondary and tertiary protein structures prediction, tools for 3D structural alignment [18][19] and protein interactions analysis with DNA, RNA and small molecules. Then structural data is used for understanding of protein's function and different protein fold [15][9] energy calculations of macromolecular structures, simulating movements and compute energies involved in molecular docking. Other subject areas include linkage analysis and metabolic pathway simulations [12].

Here in this study, we use Bioinformatics approaches like comparative modeling, determining physiochemical properties and molecular docking for computational analysis of HPV E7 protein.

Materials and Methods

Comparative Modeling

HPV genome encoded six early and 2 late proteins. Early proteins E6 and E7 are directly involved in celluar cell cycle and inhibit the p53 and pRB functionality respectively leading to carcinogen. 3D structure of E6 protein is reported in PDB but the 3D structure of E7 protein is not known yet. The amino acid sequence (105aa) of E7 protein was retrieved from Uniprot Knowledgebase (UniprotKB) [12], (Accession No. P06788) in fasta format. To get the best template of target protein, NCBI Basic Local Alignment Search Tool (Psi-BLAST) (Altschul et al., 1990) was applied against Protein Databank (PDB). Template (PDB ID: 2EWL) having 75% identity score and E value 5e-24 was selected for homology modeling. Alignment and model was built by using homology modeling program Modeller [22].

Predicted 3D structure of E7 protein was evaluated by Rampage [11] and by ERRAT [6] tools. ERRAT is a protein structure verification algorithm for evaluating the progress of model building and refinement. It works by statistical analysis of non-bonded interactions between various atom types.

Determination of Physiochemical Properties

In order to predict the physiochemical properties, ProtParam server was used which computes various physical and chemical parameters i.e., molecular weight, theoretical pI, amino acid composition, atomic composition, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) for a particular protein (Gasteiger et al., 2005)

Molecular Docking

Protein ligand and protein-protein docking of E7 protein was carried out through AuotoDock Tools and Hex server [13] respectively.

Protein-Ligand Docking

Ligand (KUC104527N) for E7 protein retrieved from PubChem database was drawn using Chem Draw Ultra Version 8.0[16] and saved in PDB format. Ligand file was then read in AutoDockTools and automatic calculation of the best root was done using tools.

Flexible docking was performed by Autodock[26] AutoDock was used for docking calculations using the Lamarckian Genetic Algorithm. Grid and docking parameters are presented in Table 1. To get optimal docking conformation we undertook 100 runs of docking. 3-Dimensional structure predicted for E7 in

this study was used as receptor in molecular docking and results are analyzed by Chimera[20] Ligplot[27] and Discovery Studio.

Table 1: Detailed grid and docking parameters were used in docking are presented

Grid Parameters	Values	Docking Parameters	Values
Spacing	1.000	Rotateable bonds	05
Grid Center	8.593X	Torsional degrees of freedom	05
	-6.221Y	Rate of Gene Mutation	0.02
	9.935Z	Rate of Crossover	0.8

Protein-Protein Docking

Protein to be used as a ligand in protein-protein docking was retrieved from STRING databas [24] and its 3D structure was predicted using ab-initio approach through I-TASSER server[21] docking server was used for protein-protein docking which generated a docked complex with energy of -7.777990e+02 and Root Mean Square of -1.00. Post docking analysis was carried out using PyMol software.

Results

Comparative Modeling

3-Dimensional structure of the E7 protein was predicted shown in Figure 1 using template 2EWL with 75% identity and E-value of 5e-24.

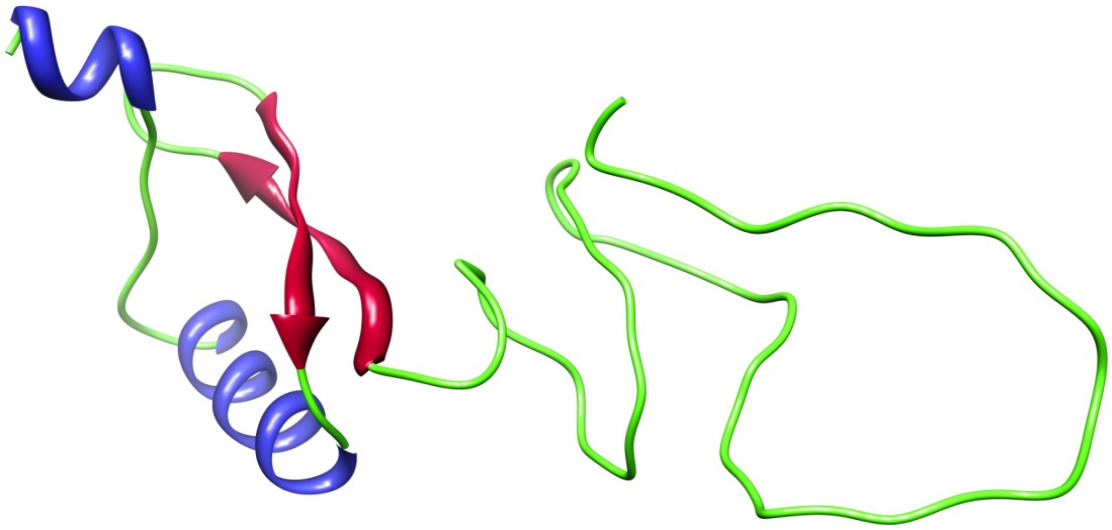


Figure 1: Predicted Structure of E7 protein using 2EWL as a template. Helixes are displayed in blue colour, strands in red colour and coils are in green colour.

To evaluate the predicted structure, Ramachandran plot and overall quality factor were accessed through Rampage and ERRAT respectively (Table 2). Ramachandran plot showed the distribution of amino acids in favoured, allowed and outlier regions. As mentioned in Table 2, most of the residues lie in favoured regions so it gives indication that model is valid. Physiochemical properties were calculated using PROTPARAM program shown in Table 3.

Table 2: Overall Quality of Protein by RAMPAGE and ERRAT

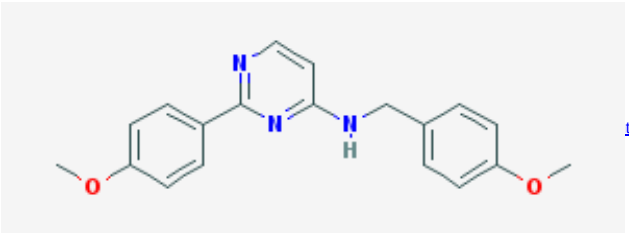
RAMPAGE			ERRAT
Favoured Regions	**Allowed Regions**	**Outlier Regions**	**Quality Factor**
93.2%	3.9%	2.9%	49 %

Table 3: Physiochemical properties of E-7 protein

Sr. No.	Parameters	Output
1	Molecular weight	11994.6
2	Theoretical isoelectric point	4.91
3	Total number of positively charged residues	8
4	Total number of negatively charged residues	17
5	Instability index	67.13
6	Aliphatic index	86.38
7	GRAVY	-0.379

Protein-ligand Docking

Ligand retrieved for E7 is shown in Figure 2. Docked complex retrieved from AutoDock Vina was analysed by Chimera, LigPlot and Discovery Studio to determine binding interactions. Figure 3 showing the binding residues of receptor protein with ligand.



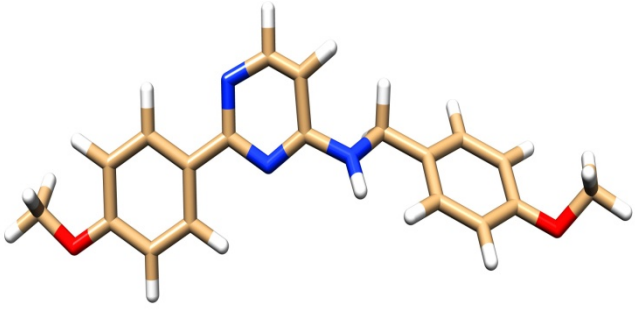


Figure 2: Chemical and PDB structure of Ligand (NSC673925) for E7 protein

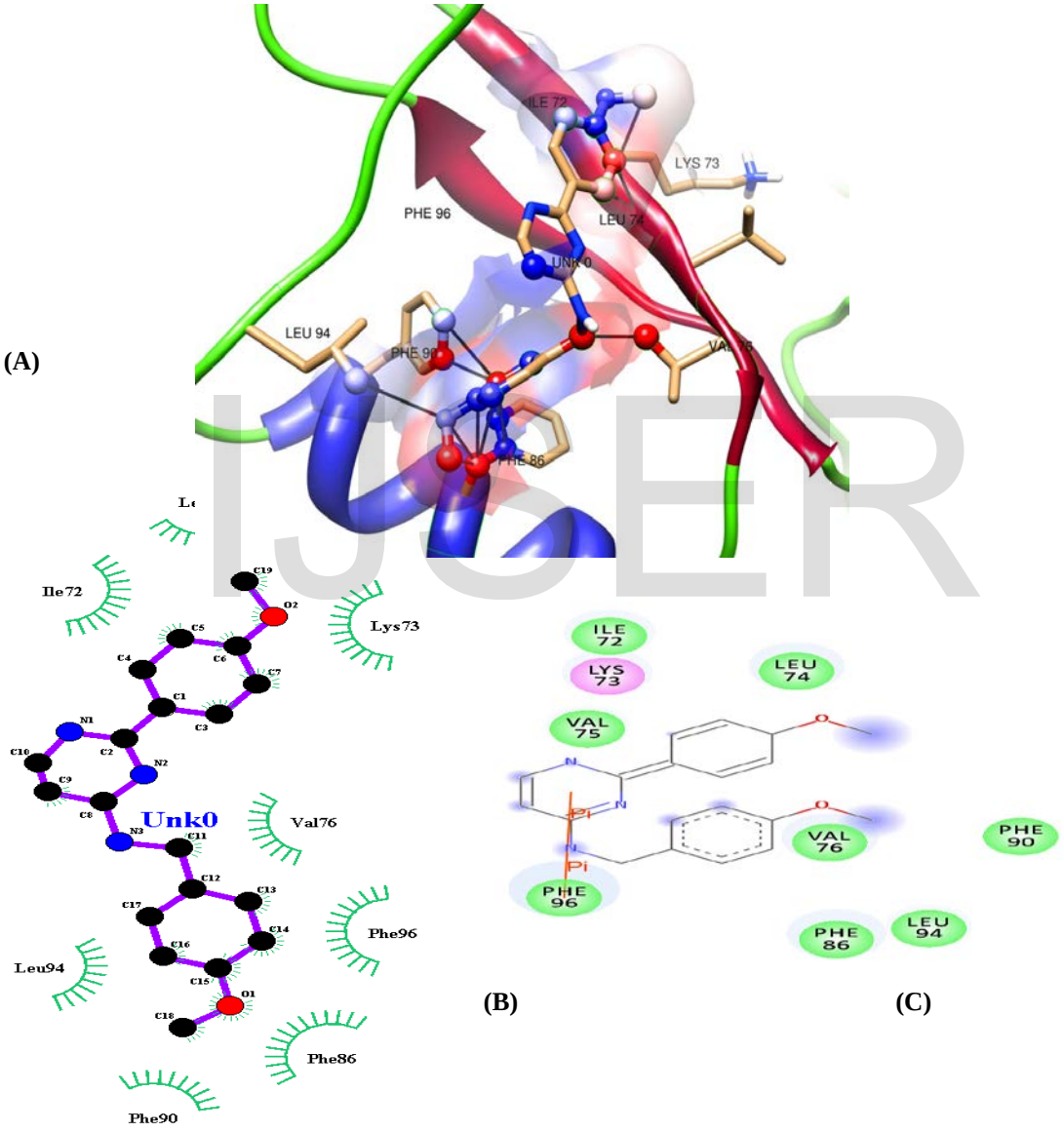


Figure 3: The interactions of ligand with E7 receptor protein **(A)** Binding residues were determined by Chimera 1.6v. Ligand surface is shown and depicted in sticks format. **(B)** interactions of docked complex were explored by LigPlot. **(C)** Docked complex was analyzed and binding residues were determined by Discovery Studio 3.5.

Protein-Protein Docking

Docked complex obtained after protein-protein docking was visualized through PyMol software for post docking analysis. Figure 4 shows the interactions between receptor (E7) and ligand protein (KRT14). Table 4 displays the interactions showed by PyMol.

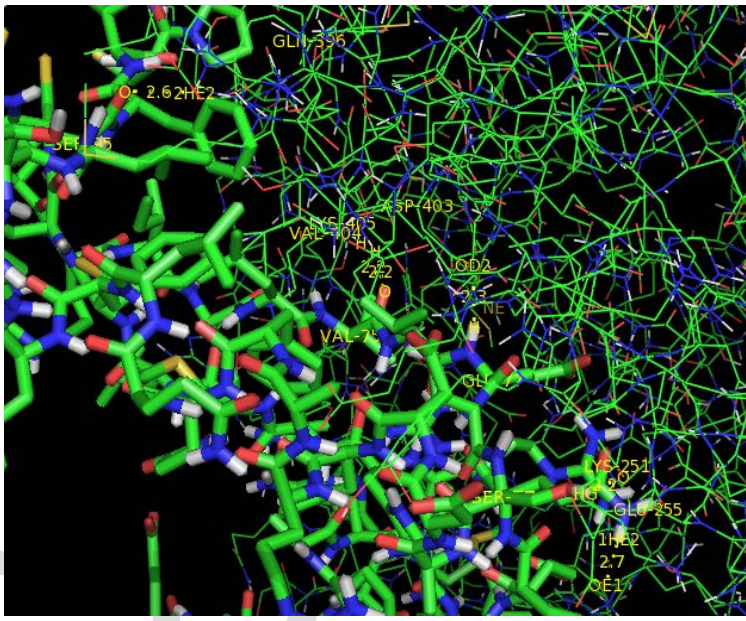


Figure 4: Interactions between E7 and KRT14

Table 4: Hydrogen Bonding as a result of protein-protein docking of E7 and KRT14

Receptor (E7) atoms	Ligand (KRT14) atoms	Bond Distance
Gln-57:1HE2	Glu-255:OE1	2.7
Ser-79:HG	Lys-251:O	2.2
Glu-77:H	Asp-403:OD2	2.3
Val-75:O	Val-404:H	2.2
Val-75:O	Lys-405:H	2.2
Ser-95:O	Gln-396:2HE2	2.6

Discussion

Cytogenetic changes after series of somatic mutations in particular genes lead to uncontrolled cellular proliferation and in turn cause cancer. In almost all types of cancers, carcinogenic events occur in a single cell i.e. these events are monocolonal in their origin which distinguish neoplasms from hyperplasias having polyclonal origin.

Role of Human papillomavirus (HPV) in the bio-pathological processes of carcinogenesis of the anogenital region has been widely researched and documented and is of considerable importance.In current research, computational analysis of HPV oncoprotein E7 was done. Different bioinformatics techniques were utilized for analysis of E7 protein.

3-Dimensional structure of E7 protein was predicted using template 2EWL with 75% identity and E-value of 5e-24 and evaluated with ramachandran plot and accessing overall quality factor which ensures that the predicted structure of E7 is a valid structure for further analysis. Physiochemical properties, predicted through PROTPARAM, indicates that E7 protein has molecular weight of 11994.6, Theoretical isoelectric point of 4.91, 8 positively charged residues, 17 negatively charged residues, Instability index of 67.13, Aliphatic index of 86.38 and GRAVYof -0.379. Molecular docking of E7 was divided into two parts, i.e. Protein-ligand docking and Protein-protein docking. In protein-ligand docking, ligand KUC104527N was docked with E7 through Autodock Vina and docked complex was analysed through Chimera, LigPlot and Discovery Studio and interactions having bond distances less than 4 are selected. Eight interactions were identified in post docking analysis. Detailed interactions of complex were determined and showed that amino acid residues ILE-72, LYS-73, LEU-74, VAL-76, PHE-86, PHE-90, LEU-94 and PHE-96 involved in interaction with ligand.

For Protein-Protein docking, interacting protein (KRT14) to be used as a ligand was retrieved through STRING database and docked using HEX docking server.Docking results were analyzed through PyMol software which revealed six hydrogen bondings between the two proteins. Hydrogen atoms of receptor protein E7 glutamine, serine and glutamic acid residues showed interactions with Oxygen atoms of interacting protein (KRT14) glutamic acid, lysine and aspartic acid residues with bond distances of 2.7, 2.2 and 2.3 respectively. Oxygen atoms of receptor protein E7 valine and serine residues showed interactions with Hydrogen atoms of interacting protein (KRT14) valine, lysine and glutamine residues with bond distances of 2.2, 2.2 and 2.6 respectively.

Conclusion

Computer aided drug designing strategy was implemented to find out the most important interacting residues of receptor and ligand to inhibit the activity of mutated protein *Insilico* receptor-ligand docking analysis suggests strong ionic and hydrophobic interactions between HPV E7 oncoprotein and ligand whereas protein-protein docking revealed extensive hydrogen bonding. These ligand-receptor interactions might inhibit aggressive carcinogenic transformation associated with HPV E7 oncoprotein. Current research may facilitate 3D structural insights of E7 oncoprotein, to eficiently reveal the functional aspect of candidate gene . For future perspective, this study could play a significant role in finding the possible involvement of E7 protein of HPV in carcinogenesis. Furthermore inhibitory mechanism of ligand was analyzed for designing novel drug against HPV.

List of Abbreviations:

HPV = Human papillomavirus

ORF = Open Reading Frame

PDB =. Protein Databank

E-Value = Expected Value

KRT1 = Keratin 14

Competing interest

The authors declare there is no competing interest among them.

Authors' contributions

NAK defined the research theme and designed methods analyzed the data, interpreted the results

and wrote the paper. RAT, SAS and NZ carried out all the work and analysis of results under the guidance of NAK and RGK. NAK also provided suggestions to the interpretation of results. All authors have contributed to, seen, read and approved the manuscript.

Acknowledgements

We are thankful to PMAS-Arid

Agriculture university and International Islamic university, Islamabad to carry out this research work.

References

[1] Altschul, "Basic local alignment search tool," Journal of Molecular Biology, 215, pp. 403-410, 1990.

[2] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, "The Protein Data Bank," Nucleic Acid Residues, 28, pp. 235-242, 2000.

[3] Bernstein FC, Koetzle TF, Williams GJ, MeyerEF JR, Brice MD, Rodgers JR, "The Protein Data Bank. A computer-based archival file for macromolecular structures," European Journal of Biochemistry, 80, pp. 319-324, 1977.

[4] Boguski MS, "Biosequence exegesis," Science, 286, pp. 453-455, 1999.

[5] Cole ST, Danos O, "Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products," Journal of Molecular Biology, 193, pp. 599-608, 1987.

[6] Colovos VC, Yeates T, "Verification of protein structures: Patterns of non bonded atomic interactions," Protein Sci, 2, pp. 1511–1519, 1993.

[7] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, "Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, pp. 571-607, 2005.

[8] Gonnet GH, Korostensky C, Benner S, "Evaluation measures of multiple sequence alignments," J Comput Biol, 7(1-2), pp. 261-276, 2000.

[9] Hegyi H, Gerstein M, "The relationship between protein structure and function: a comprehensive survey with application to the yeast genome,"J Mol Biol, 288, pp. 147-164, 1999.

[10] Liu X, Clements A, Zhao K, Marmorstein R, "Structure of the Human Papillomavirus E7 Oncoprotein and Its Mechanism for Inactivation of the Retinoblastoma Tumor Suppressor," Journal of Biological Chemistry, 281, pp. 578–586, 2005.

[11] Lovell SC, Davis IW, Arendall WB, de-Bakker PI., Word JM, Prisant MG, "Structure validation by Calpha geometry: phi,psi and C beta deviation," Proteins: Structure, Function & Genetics, , pp. 437-450, 2002.

[12] Luscombe NM, Greenbaum D, Gerstein M, "Review: What is bioinformatics? An introduction and overview," Year book of Medical Informatics, pp. 83-100, 2001.

[13] Macindoe G, Mavridis L, Venkatraman V, Devignes MD, David W, Ritchie, "HexServer: an FFT-based protein docking server powered by graphics processors," Nucleic Acids Research, 38, pp. 445–449, 2010.

[14] Margaret E, McLaughlin-Drubin, Karl-Münger, "The Human Papillomavirus E7 Oncoprotein," Virology, 384,]]. 335–344, 2009.

[15] Martin AC, Orengo CA, Hutchinson EG, Jones S, Karmirantzou M, Laskowski RA, "Protein folds and functions," Structure, 6, pp. 875-884, 1998.

[16] Mendelsohn LD, "ChemDraw 8 Ultra: Windows and Macintosh Versions," J Chem Inf Comput Sci, *pp*. 44, 2225–2226, 2004.

[17] Miller C, Gurd J, Brass A, ”A RAPID algorithm for sequence database comparisons: application to the identification of vector contamination in the EMBL databases,” Bioinformatics, 15, pp. 111-121, 1999.

[18] Orengo CA, Taylor WR, “SSAP: sequential structure alignment program for protein structure comparison,” Methods Enzymolb, 266, pp. 617-635, 1996.

[19] Orengo CA, “CORA—topological fingerprints for protein structural families,” Protein Sci, 8, pp. 699-715, 1999.

[20] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, “Ferrin TE: UCSF Chimera--a visualization system for exploratory research and analysis,” J Comput Chem, 25, pp. 1605-1612, 2004.

[21] Roy A, Kucukural A, Zhang Y, “I-TASSER: a unified platform for automated protein structure and function prediction,” Nature protocols, 5, pp. 725-738, 2010.

[22] Sali A, Blundell TL, “Comparative protein modeling by satisfaction of spatial restraints,” J Mol Biol, 234, pp. 779-815, 1993.

[23] Storey A, Osborn K, Crawforo L, “Co-transformation by human papillomavirus types 6 and 11,” Journal of General Virology, 71, pp. 165-171, 1990.

[24] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, “The STRING database in 2011: functional interaction networks of proteins globally integrated and scored,” Nucleic Acids Research, 39, pp. 561–568, 2001.

[25] The Economist, “Drowning in data,” The Economist, pp. 97-98, 2009.

[26] Trott O, Olson AJ, “AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading,” J Comput Chem, 31, pp. 455-61, 2010.

[27] Wallace AC, Laskowski RA, Thornton. “JM: LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions,” Protein Eng, 8, pp. 127-134, 1996.

[28] Zhang MQ, “Promoter analysis of co-regulated genes in the yeast genome,” Comput Chem, 23, pp. 233-250, 1990.