Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 5, Issue1, 2014, pp51-55
http://www.bipublication.com

IN- SILICO STRUCTURAL CHARACTERIZATION OF P300 INVOLVED IN EPITHELIAL CANCER

Ashish A. Kulkarni * and Sana A. Shaikh

Dept. of PG Studies and Research in Bioinformatics
Walchand Centre for Biotechnology, Solapur, Maharashtra, India.
Corresponding author: Email: *kulkarni.ashu.a@gmail.com,sanbio7@gmail.com, Tel: 09503814931.

[Received-22/01/2014, Accepted-24/02/2014]

ABSTRACT:

EP300 gene codes for protein Histone acetyltransferase (p300), serve as tumor suppressor protein and transcription factor. The p300 plays important role in cell proliferation, differentiation, and apoptosis. Mutations of p300 associated with certain types of cancer mainly epithelial cancer (carcinoma). Cancers of the epithelial cells make up about 85 out of every 100 cancers (85%). Chromosomal aberrations involving EP300 may be a cause of acute myeloid leukemias (AML). The mechanisms by which the inactivation of p300 contributes to carcinogenesis have not been fully elucidated. The paper deals with understanding of p300 function in tumor suppression by in-silico approach. The physicochemical properties of the selected human p300 protein was analyzed by using ExPASy's ProtParam tool and it was found that the molecular weight is 264160.6 Da, its isoelectric point (pI) was found to be basic in nature. The instability index infers that p300 is unstable, its GRAVY value (-0.725) indicates better interaction with water. The SOPMA secondary structure prediction showed that random coil dominated all the other conformations. The 3D structure of p300 was obtained by Swiss Model server. The data provide a structural basis for understanding the known biochemical properties of p300 turning into an oncogene and indicate additional regions in the molecule that may possibly participate in other cellular functions. In vivo studies have demonstrated that tumor growth, therefore the gene is a clear target for the development of small molecule inhibitors.

Keywords: *p300, ProtParam, SOPMA, Swiss model, oncogene, mutation.*

[I] INTRODUCTION

The integration of wet experiments and the use of bioinformatics analyses have become an indispensable part of the biological and clinical research of this century. A typical scenario of cancer research using bioinformatics tools is an analysis of global profiles of gene expression in cancer. EP300 or p300 gene codes for protein Histone acetyltransferase (p300), serve as tumor suppressor protein and transcription factor. The EP300 gene is located on the long (q) arm of the human chromosome 22 at position 13.2. P300 and CBP (CREB binding protein) are

highly related transcriptional coactivators. P300 Protein regulates the activity of many genes in tissues throughout the body. It plays an essential role in controlling cell growth and division also prompting cells to mature and assume specialized functions (differentiate). p300 connects transcription factors, which are proteins that start the transcription process, with the complex of proteins that carries out transcription. The ability of p300 to acetylate many transcription factors, including p53, E2F, TFIIE, and TFIIF etc. demonstrates a novel mechanism of targeted p300 regulation of gene expression [1, 2, 3]. Defects in P300 may play a role in epithelial cancer. Chromosomal aberrations involving EP300 may be a cause of acute myeloid leukemias. Translocation t (8;22) (p11;q13) with KAT6A. Several mutations in the EP300 gene have been identified in people with Rubinstein-Taybi syndrome. Rubinstein-Taybi syndrome 2 (RSTS2) [MIM:613684] is a disorder characterized by craniofacial abnormalities, postnatal growth deficiency, broad thumbs, broad big toes, mental retardation and a propensity for development of malignancies.

The raw sequence information of proteins and nucleic acid can convert to analytical and relative information with the help of soft computing tools. Protein sequence was retrieved from UniprotKB and were subjected to ProtParam to analyze various physicochemical properties, secondary structure was predicted by SOPMA, the protein 3D model and its characteristics were predicted by Swiss model server and were visualized using RasMol, molecular graphics visualization program.

[II] MATERIALS AND METHODS

2.1. Sequence retrieval

The sequence of the protein p300 was retrieved From UniprotKb database. Sequence retrieved belongs to the *Homo sapiens.* UniprotKb is a comprehensive, high-quality and freely accessible database of protein sequence and functional information [4].

2.2 Analysis of physicochemical parameters

The different physicochemical properties of p300 were computed using ExPASy's ProtParam tool [5] and these properties can be deduced from a protein sequence. The ProtParam includes the following computed parameters: Molecular weight (M.Wt.), theoretical pI, instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY).

2.3 Secondary structure prediction

The secondary structure was predicted by self-optimized prediction method with alignment (SOPMA) [6]. SOPMA was employed for calculating the secondary structural features of the selected protein sequence considered in this study. This method calculates the content of α-helix, β-sheets, turns, random coils and extended strands. SOPMA is a neural network based methods; global sequence prediction may be done by this sequence method [7].

2.4 Tertiary structure prediction

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server or from the program DeepView (Swiss Pdb-Viewer) [8]. The purpose of this server is to make protein modeling accessible to all biochemists and molecular biologists worldwide. The 3D structure of p300 protein was generated by homology modeling using SWISS-MODEL automated modeling server. The 3D structure template was searched using automated program present in the server.

2.5 Structure Validation

The PROCHECK suite of programs provides a detailed check on the stereochemistry of a protein structure. Its output comprise a number of plots in postScript format and a comprehensive residue-by-residue listing. The PROCHECK programs are useful for assessing the quality not only of protein structures in the process of being solved but also of existing

structures and of those being modeled on known structures[9]. PROCHECK from PDBSum of EBI is used for structure validation of p300.

2.6 Structure visualization

The predicted three dimensional structure of the protein p300 (Homo sapiens) was visualized using molecular graphics visualization program RasMol used primarily for the depiction and exploration of biological macromolecule structures [10]. RasMol works with graphical user interface and command line interface.

[III] RESULTS AND DISCUSSION

3.1 Retrieved sequences

The sequence of p300 protein of Homo sapiens was retrieved from protein database UniprotKb having accession number Q09472. The sequence length of p300 protein in Homo sapiens is 2414 amino acids [Figure 1]. P300 protein of Homo sapiens consists of major domains like ZnF_TAZ, Bromo, KAT11 and ZnF_ZZ.

```
>sp|Q09472|EP300_HUMAN Histone acetyltransferase p300 OS=Homo sap
MAENVVEPGPPSAKRPKLSSPALSASASDGTDFGSLFDLEHDLPDELINSTELGLTNGGD
INQLQTSLGMVQDAASKHKQLSELLRSGSSPNLNMGVGGPGQVMASQAQQSSPGLGLINS
MVKSPMTQAGLTSPNMGMGTSGPNQGPTQSTGMMNSPVNQPAMGMNTGMNAGMNPGMLAA
GNGQGIMPNQVMNGSIGAGRGRQNMQYPNPGMGSAGNLLTEPLQQGSPQMGGQTGLRGPQ
PLKMGMMNNPNPYGSPYTQNPGQQIGASGLGLQIQTKTVLSNNLSPFAMDKKAVPGGGMP
NMGQQPAPQVQQPGLVTPVAQGMGSGAHTADPEKRKLIQQQLVLLLHAHKCQRREQANGE
VRQCNLPHCRTMKNVLNHMTHCQSGKSCQVAHCASSRQIISHWKNCTRHDCPVCLPLKNA
GDKRNQQPILTGAPVGLGNPSSLGVGQQSAPNLSTVSQIDPSSIERAYAALGLPYQVNQM
PTQPQVQAKNQQNQQPGQSPQGMRPMSNMSASPMGVNGGVGVQTPSLLSDSMLHSAINSQ
NPMMSENASVPSLGPMPTAAQPSTTGIRKQWHEDITQDLRNHLVHKLVQAIFPTPDPAAL
KDRRMENLVAYARKVEGDMYESANNRAEYYHLLAEKIYKIQKELEEKRRTRLQKQNMLPN
AAGMVPVSMNPGPNMGQPQPGMTSNGPLPDPSMIRGSVPNQMMPRITPQSGLNQFGQMSM
AQPPIVPRQTPPLQHHGQLAQPGALNPPMGYGPRMQQPSNQGQFLPQTQFPSQGMNVTNI
PLAPSSGQAPVSQAQMSSSSCPVNSPIMPPGSQGSHIHCPQLPQPALHQNSPSPVPSRTP
TPHHTPPSIGAQQPPATTIPAPVPTPPAMPPGPQSQALHPPPRQTPTPPTTQLPQQVQPS
LPAAPSADQPQQQPRSQQSTAASVPTPTAPLLPPQPATPLSQPAVSIEGQVSNPPSTSST
EVNSQAIAEKQPSQEVKMEAKMEVDQPEPADTQPEDISESKVEDCKMESTETEERSTELK
TEIKEEEDQPSTSATQSSPAPGQSKKKIFKPEELRQALMPTLEALYRQDPESLPFRQPVD
PQLLGIPDYFDIVKSPMDLSTIKRKLDTGQYQEPWQYVDDIWLMFNNAWLYNRKTSRVYK
YCSKLSEVFEQEIDPVMQSLGYCCGRKLEFSPQTLCCYGKQLCTIPRDATYYSYQNRYHF
CEKCFNEIQGESVSLGDDPSQPQTTINKEQFSKRKNDTLDPELFVECTECGRKMHQICVL
HHEIIWPAGFVCDGCLKKSARTRKENKFSAKRLPSTRLGTFLENRVNDFLRRQNHPESGE
VTVRVVHASDKTVEVKPGMKARFVDSGEMAESFPYRTKALFAFEEIDGVDLCFFGMHVQE
YGSDCPPPNQRRVYISYLDSVHFFRPKCLRTAVYHEILIGYLEYVKKLGYTTGHIWACPP
SEGDDYIFHCHPPDQKIPKPKRLQEWYKKMLDKAVSERIVHDYKDIFKQATEDRLTSAKE
LPYFEGDFWPNVLEESIKELEQEEEERKREENTSNESTDVTKGDSKNAKKKNNKKTSKNK
SSLSRGNKKKPGMPNVSNDLSQKLYATMEKHKEVFFVIRLIAGPAANSLPPIVDPPDLIP
CDLMDGRDAFLTLARDKHLEFSSLRRAQWSTMCMLVELHTQSQDRFVYTCNECKHHVETR
WHCTVCEDYDLCITCYNTKNHDHKMEKLGLGLDDESNNQQAAATQSPGDSRRLSIQRCIQ
SLVHACQCRNANCSLPSCQKMKRVVQHTKGCKRKTNGGCPICKQLIALCCYHAKHCQENK
CPVPFCLNIKQKLRQQQLQHRLQQAQMLRRRMASMQRTGVVGQQQGLPSPTPATPTTPTG
QQPTTPQTPQPTSQPQPTPPNSMPPYLPRTQAAGPVSQGKAAGQVTPPTPPQTAQPPLPG
PPPAAVEMAMQIQRAAETQRQMAHVQIFQRPIQHQMPPMTPMAPMGMNPPPMTRGPSGHL
EPGMGPTGMQQQPPWSQGGLPQPQQLQSGMPRPAMMSVAQHGQPLNMAPQPGLGQVGISP
LKPGTVSQQALQNLLRTLRSPSSPLQQQQVLSILHANPQLLAAFIKQRAAKYANSNPQPI
PGQPGMPQGQPGLQPPTMPGQQGVHSNPAMQNMNPMQAGVQRAGLPQQQPQQQLQPPMGG
MSPQAQQMNMNHNTMPSQFRDILRRQQMMQQQQQQGAGPGIGPGMANHNQFQQPQQVGYP
PQQQQRMQHHMQQMQQGNMGQIGQLPQALGAEAGASLQAYQQRLLQQQMGSPVQPNPMSP
QQHMLPNQAQSPHLQGQQIPNSLSNQVRSPQPVPSPRPQSQPPHSSPSPRMQPQPSPHHV
SPQTSSPHPGLVAAQANPMEQGHFASPDQNSMLSQLASNPGMANLHGASATDLGLSTDNS
```

Figure 1: Protein sequence of p300 Homo sapiens.

3.2 Primary protein sequence analysis

The physicochemical properties of p300 protein were predicted by using ProtParam tool as shown in the Table 1. The protein has a molecular weight of about 264160.6 Dalton. The instability index showed that p300 (67.04) is unstable. The computed pI value of p300 (8.81) indicates that it is basic in nature. The very high aliphatic index of p300 infers that it may be stable for a wide range of temperature. The very low GRAVY index of p300 (Homo sapiens) could result in a better interaction with water.

Alpha Helix(Hh)	Extended strand (Ee)	Beta turn (Tt)	Random coil (Cc)
28.33%	9.82%	6.55%	55.30%

Table 1: ProtParam Analysis.

3.3 Secondary structure analysis

The detailed secondary structure analysis is shown in the Table 2, which illustrates that random coil predominates the other secondary structure elements (55.30%) and extended strand being the least conformational structure (9.82%). In p300, alpha helix and beta turn showed 28.33% and 6.55% respectively. The secondary structure indicates whether a given amino acid lies in a helix, strand, turn or coil and other states [11]. P300 protein showed zero percentage of Beta Bridge, 3_{10} helix, Pi helix and other states.

Molecular weight (Da)	Instability index	pI value	Aliphatic index	(GRAVY)
264160.6	67.04	8.81	59.74	-0.725

Table 2: Sopma Analysis

3.4 Tertiary structure analyses

The 3D structure analyses of p300 protein were done by using SWISS-MODEL automated modeling server [Figure 2]. Predicting the protein 3D structures by this method implements the four steps of the homology modeling approach like Template searching to identify the structure homolog, Target-template alignment to get the maximum identity, Model building & optimization and Model evaluation. Template selection, alignment and model building are

automatically done by the server. 1jspB was the template selected for p300 protein from Homo sapiens and was aligned with the target protein which showed 95.73% identity according to Schwede *et al.,*. 1JSP is the NMR structure of CBP bromodomain in complex with p53 peptide which is having two chains (chainA, chainB).

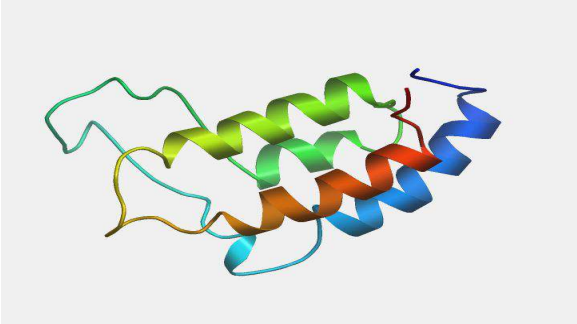

Figure 2: p300 three dimensional structure.

3.5 3D Model Validation

The built up 3D model of p300 protein was validated using Procheck of PDBSum at EBI. The procheck server showed 97.1% (100 residues) and 0.0% (0 residues) in most favored regions and disallowed regions, respectively [Figure 3]. As per Laskowski *et a*l., the model developed is a good quality model hence can be used for further study.

PROCHECK statistics

1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	100	97.1%
Additional allowed regions	[a,b,l,p]	3	2.9%
Generously allowed regions	[~a,~b,~l,~p]	0	0.0%
Disallowed regions	[XX]	0	0.0%
Non-glycine and non-proline residues		103	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		3	
Proline residues		10	
Total number of residues		118	

Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A,B,L].

Figure 2: Procheck (Ramchandran Plot) Analysis.

3.5 Structure visualization using Rasmol program

The predicted 3d structure of Homo sapiens p300 was visualized using molecular graphics visualization program RasMol, showing different display features and labels. [Figure .4]

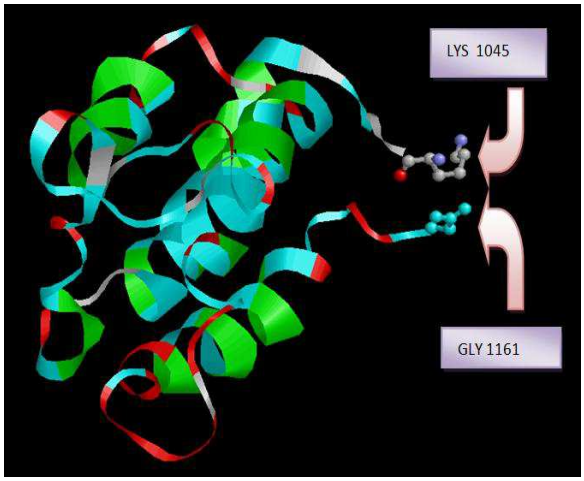

Figure 4: p300 structure visualization (RasMol).

[IV] CONCLUSION

p300 has a direct role in the control of cell growth and differentiation in primary epithelial cells, and p21Waf1/CIP1 is an important mediator of these p300 functions. In the present study we determine only the impact of p300 and its mechanism in epithelial cancer. For developing any realistic model one need to incorporate other proteins which influence p300 protein simultaneously and then study the impact collectively. Our study is just one step forward towards understanding p300 regulatory network. In this paper, bioinformatics tools were used to probe basic questions in tumor suppression and developing a 3D model for a tumor suppressor p300 in its normal state. This analysis is important because it allows us to begin to understand how epithelial cells produce and receive signals in small intestine as well as finding out the mutations in p300.

ACKNOWLEDGEMENT

The special thank goes to my Principal Dr. A. H. Manikshete and my guide Miss. Sana Shaikh, Asst. Professor, Dept of Bioinformatics, WCBT, Solapur for their great appreciation, I would like to thank to all departmental staff for valuable

guidance and their kind support and encouragement during the research work.

REFERENCES

1. Gu, W. et al., (1997) Cell 90, 595-606.
2. Eckner, R. et al., (1994) Genes & Dev. 8, 869-884.
3. Chrivia, JC. et al., (1993) Nature 365, 855-859.
4. http://www.uniprot.org/help/uniprotkb.
5. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch, A. (2005). ; Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607
6. Ashokan K V., Mundaganur D S and Mundaganur Y D. (2010). Catalase: Phylogenetic Characterization to Explore Protein Cluster, Journal of Research in Bioinformatics: 1, 001-008.
7. Prashant VT., Uddhav SC., Madura SM., Vishal PD and Renuka RK. (2010). Secondary Structure Prediction and Phylogenetic Analysis of Salt Tolerant
8. Schwede T., Kopp J., Guex N and Peitsch M.C. (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 31:3381-3385.
9. Roman A. Laskowski, Malcolm W. Macarthur, David S. Moss, Janet M. Thornton, PROCHECK: a program to check the stereo chemical quality of protein structures.
10. Sayle R.A and Milner White EJ. (1995). RASMOL: biomolecular graphics for all. Trends Biochem. Sci. 20:374.
11. Jyotsna C., Ashish P., Shailendra G., Verma M K. (2010). Homology Modeling and Binding Site.