



Computational Modelling of Histone Deacetylases 5

Lalit R. Samant^{1*}, Vikrant C. Sangar², Akshay S. Jadhav³, Dr. Abhay S. Chowdhary^{1,2}

¹Systems Biomedicine Division, Haffkine Institute for Training, Research and Testing, Acharya Donde Marg, Parel, Mumbai, India.

²Department of Virology and Immunology, Haffkine Institute for Training, Research and Testing, Acharya Donde Marg, Parel, Mumbai, India.

³Department of Biotechnology and Bioinformatics, Padmashree Dr. D.Y. Patil University, Belapur CBD, Navi Mumbai, India.

*Corresponding author's E-mail: samantlalit@gmail.com

Accepted on: 26-08-2014; Finalized on: 31-12-2014.

ABSTRACT

Histones are the most abundant proteins associated with the eukaryotic DNA. The N-terminal tails of these histones are subjected to modifications by two enzymes – Histone Acetyl Transferases (HATs) and Histone Deacetylases (HDACs). These two enzymes help to regulate the acetylation of histones and condensation of the chromatin in its stable form. HDACs are considered as one of the promising targets in cancer biology studies. HDAC-5 belongs to class II member of histone family. Currently, there is no published 3D structure model available for HDAC-5 (Q9UQL6). As a result, we planned a study where we are going to build a 3D structure for HDAC-5 with the help of suitable bioinformatics tools. ExPASy ProtParam Tool was used in retrieving the physicochemical characterization of the protein. The functional characterization was carried out using Cysteine Recognition Server and HMMTOP Server. I-TASSER was used for modelling the 3D structure for HDAC-5. Model refinement and model validation were carried out using Swiss-Pdb Viewer and RAMPAGE Server respectively. This modelled HDAC-5 can be further used for target based drug discovery studies.

Keywords: Histone deacetylases, HDAC-5, ExPASy ProtParam, I-TASSER, RAMPAGE Server.

INTRODUCTION

Every cell in the body goes through a cell cycle phase where they need to grow, proliferate, differentiate and die. Any alteration in the cell cycle can lead into Cancer. Cancer is a condition where the cells grow at an uncontrolled rate. The cancerous cells can invade and destroy surrounding healthy cells, tissues and also organs. It has been realised that Histones play an important role in cancer development.¹

Histones are the most abundant proteins associated with the eukaryotic DNA. Eukaryotic cells contain histones. These histones are of five types viz. H2A, H2B, H3, H4 and H1, from which the H1 is the linker histone and H2A, H2B, H3 and H4 are core histones. A pair of these core histones assemble to form an octamer. To this octamer, the DNA is wrapped and hence this assembly is known as “nucleosome”. The remaining linker histone is associated with the nucleosome for maintaining the chromosome structure. When histones are isolated from cells, their N-terminal tails are modified with small molecules. Lysines are acetylated or methylated while serines are phosphorylated. These histone modifications are mediated by specific enzymes. Histone acetyl transferases (HATs) are responsible for acetylation of the lysines while histone deacetylases (HDACs) removes these modifications. Due to their critical role in regulation of chromatin structure and gene expression, HDACs have become major drug targets. HDAC inhibitors are specific toward tumour cells. This is the reason why they are used as anticancer drugs.²⁻⁴

Histone deacetylases are class of enzymes which act in removal of acetyl groups from a ϵ -N-acetyl lysine amino

acid on a histone, because of which the DNA is wrapped around the histones tightly. There are in all 4 classes in which all 18 Histone deacetylases are classified. Class I comprises of HDAC-1, HDAC-2, HDAC-3 and HDAC-8. Class II is divided into two sub-classes – Class IIA and Class IIB in which Class IIA comprises of HDAC-4, HDAC-5, HDAC-7 and HDAC-9 and Class IIB comprises of HDAC-6 and HDAC-10. Class III are known as Zn-dependent histone deacetylases, comprising of zinc. Class IV comprises of the remaining HDAC-11.^{5,6}

HDAC-5 is a protein coding gene. It is located on 17q21. Diseases associated with HDAC-5 include intellectual disability, herpes simplex and various types of cancers. The main role of HDAC-5 is deacetylation of lysine residues on the N-terminal part of the core histones (H2A, H2B, H3 and H4). It also plays an important role in transcriptional regulation, histone deacetylation which gives a tag for epigenetic repression, cell cycle progression and developmental events.⁷

Currently, there is no published 3D structure model available for HDAC-5 (Q9UQL6). As a result, we planned a study where we are going to build a 3D structure for HDAC-5 with the help of suitable bioinformatics tools. The modelled HDAC-5 can be further used for docking studies in development of novel HDAC inhibitor molecules.

MATERIALS AND METHODS

The 3D structure of HDAC-5 (Q9UQL6) is unavailable in Uniprot under Protein Knowledgebase (UniProtKB). Its FASTA Sequence of HDAC-5 was retrieved from UniProtKB Database and then subjected to physicochemical and functional characterization. I-TASSER was used to build



the model and the best model was selected on the basis of confidence score. The refinement of the model was done by using Swiss-Pdb Viewer and later on validation of the model was done by RAMPAGE Server. The FASTA Sequence which was used is mentioned below-

```
>sp|Q9UQL6|HDAC5_HUMAN Histone deacetylase 5 OS=Homo sapiens GN=HDAC5 PE=1 SV=2
MNSPNSDGMGSRGPELEILPRTSLHSIPVTVKVFVLPFRMPSSMGGGGSSPSFVELR
GALVGSVDFTLREQQLQELIALKQQQQQLKQQLLFAEFQKQKHDLTRQHEVQLQKHLKQQ
QEMLAQQQQQEMLAQRQQLLEQQQRQEQEQEELKQRLEQLLLLRNKEKSKESALIAS
TEVKRLRQEFLLSKSKEPTPGGLNHSILPQHFKCKGAAHSLDQSSPPQSGPPGTPPSYKL
PLBGPYDSRDDFLRKTASEPNLKVRSRLKQKVAERSSPFLRRKDGTVISTFKRAVEI
TGAGPGASVNCNAPFSGSPSSPNSHSTIAENGFTGSVNPITMPLQHRALPLDSSPNQ
FSLYTSFSLNPLSLGLQATVTVTNSHLTASPKLSTQQAERQALQSLRQGGTLTGKFMST
SSIPGCLLVALEGGDSSPHGASLLQHVILLQARQQSTLIAVFLHSGSPPLVTEGRVATS
MRIVKGLPARRHDFLRKTASEPNLKVRSRLKQKVAERSSPFLRRKDGTVISTFKRAVEI
LPRQPTTPEETEELTEQQEVLLGEGALTMPREGSTSESTQEDLEEEDEEDDGEEDD
CIQVKDEEGESGAEQDLLEEFEGAGYKLFSDAQPLQPLQVYQAPLSLATVPHALGRTO
SSPAAPGGMKSPDPQVVKHLFTTGVYDITFMLKHQCMGNTHVHEHAGRIQSWRSLQE
TGLLSKCERIRGRKATLDEIQTVHSEYHTLLYGTSLNBRKLDKSKLLGPISQMYAVLP
CGGIGVDSDTVWEMHSSAVRMAVGCLELAFKVAAGELKNGFAIIRPPGHAEESTAM
GFCFNSVAITAKLLQKLNKLVKVLIVDWDIHHNGTQQAIFYNDPVLVYISLHRYDNGNF
FPGSGAPEEYVGGGPGVGNVNVAVTQGVDPPIGDVEYLTAFTVVMPIAHEFSDDVVLVS
AGFDVAEHLPLSGGYSVTAFCFGLTRQLMLAGRVLVLALEGGHDLTAICDASEACVS
ALLSVELQPLDEAVLQQKPNINAVATLEKRVIEIQSHWVCVQKFAAGLGRSLREAQAGET
EEAETVSAMALLSVGAEQAQAAAAREHSRPAEAEPEMQEPAL
```

Physiochemical Characterization

The information of HDAC-5 regarding its molecular weight, theoretical isoelectric point, instability index, aliphatic index, GRAVY, atomic composition, amino acid composition and extinction coefficient are all included in physiochemical characterization. For retrieving information of physiochemical characterization, ExPasy ProtParam Server was used. The result is presented in Table 1. Table 3 highlights the Amino Acid Composition of HDAC-5.⁸

Functional Characterization

For prediction of transmembrane helices and topology of proteins, HMMTOP Server has been used and Cys_Rec is used for Predicting SS-bonding States of Cysteines and disulphide bridges in Protein Sequences. The result of Cys_Rec is presented in Table 2.^{9,10}

Model Building & Refinement

The 3D structure of HDAC-5 was modelled using I-TASSER. The top 10 templates which are shown in Table 4 were used to build the model by I-TASSER. The I-TASSER Results are presented in Table 5. Confidence Score (C-score) should be in the range of [-5 to 2], where a C-score of higher value signifies a model with a high confidence and vice versa. With the help of the C-score, the best among the resultant modelled structure was selected. The Energy Minimization of the modelled HDAC-5 was done using Swiss-Pdb Viewer. Figure 1 shows the Energy Minimised Model of HDAC-5 using UCSF Chimera.¹¹

Model Validation

Model Validation was done by RAMPAGE Server. It plotted a Ramachandran Plot whose result is displayed in Table 6. Figure 2 is the screenshot of the resulted Ramachandran Plot.¹²

RESULTS AND DISCUSSION

Physiochemical Characterization

ExPasy ProtParam tool gives a detailed information about the molecular weight, theoretical isoelectric point, instability index, aliphatic index, GRAVY, atomic composition, amino acid composition and extinction coefficient of the protein. Isoelectric Point is the "pH at which a particular molecule carries no net electrical charge". The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. The instability index provides an estimate of the stability of your protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. The aliphatic index of a protein is defined as "the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine)". Higher value of Aliphatic Index shows that protein is stable for wide range of temperature.

Grand average of hydropathicity index indicates the solubility of the proteins, positive GRAVY indicates hydrophobic nature and negative GRAVY indicates hydrophilic nature.

Functional Characterization

HMMTOP predicts membrane topology of integral membrane proteins using hidden Markov model and it was found out that there were no transmembrane helices present. Cys_Rec predicts SS-bonding States of Cysteines. Also, Cys_Rec results shows, the presence of two disulphide bonds which indicates, the stability of protein which might be increased due to this, along with non-covalent interactions.

There are in all 14 Cysteines present in the query sequence, whose positions are – 213, 311, 426, 601, 696, 698, 727, 781, 807, 843, 982, 1012, 1018 and 1060.

Table 2: Cys_Rec Result by HMMTOP.

No. of Cysteines	Positions	Score
1	213	-28.1
2	311	4.4
3	426	-26.6
4	601	-27.0
5	696	-27.2
6	698	-27.5
7	727	-21.0
8	781	-31.1
9	807	-29.9
10	843	-27.3
11	982	-3.2
12	1012	-20.7
13	1018	-20.9
14	1060	-25.5

Table 1: Physicochemical Characterization predicted by Expsy Protparam Tool.

Sequence Length	Molecular Weight	Theoretical Isoelectric Point	Instability Index	Aliphatic Index	Grand Average of Hydropathicity (GRAVY)
1122	121977.5	5.83	64.25	80.07	-0.493

Table 3: Amino Acid Composition.

Amino Acids	Symbol	Number of Residues	Number of Residues (In Percentage)
Alanine	A	84	7.5%
Arginine	R	51	4.5%
Asparagine	N	26	2.3%
Aspartic Acid	D	35	3.1%
Cysteine	C	14	1.2%
Glutamine	Q	97	8.6%
Glutamic Acid	E	95	8.5%
Glycine	G	92	8.2%
Histidine	H	39	3.5%
Isoleucine	I	31	2.8%
Leucine	L	128	11.4%
Lysine	K	52	4.6%
Methionine	M	22	2.0%
Phenylalanine	F	25	2.2%
Proline	P	83	7.4%
Serine	S	98	8.7%
Threonine	T	62	5.5%
Tryptophan	W	6	0.5%
Tyrosine	Y	15	1.3%
Valine	V	67	6.0%
L-Pyrrolysine	O	0	0.0%
L-Selenocysteine	U	0	0.0%

Model Building

The 3D structure of HDAC-5 was modelled using I-TASSER by ab initio method using a meta-threading approach. The best among the resultant modelled structures was selected depending on the C-score. The top ten templates which were used for building the model using I-TASSER are highlighted in the Table 4.

Table 4: Top 10 Templates used by I-TASSER

Rank	PDB Hit	Iden1	Iden2	Cov.	Normal Z-score
1	2vqjA	0.76	0.27	0.36	2.36
2	2vqjA	0.75	0.27	0.36	4.34
3	3c10A	0.70	0.24	0.34	2.89
4	2pqpA	0.70	0.24	0.34	4.99
5	2pqpA	0.70	0.24	0.34	3.79
6	2vqjA	0.75	0.27	0.36	4.02
7	2pqpA	0.70	0.24	0.34	5.41
8	2vqmA	0.74	0.00	0.35	7.68
9	2vqjA	0.76	0.27	0.36	2.72
10	2nvrA	0.70	0.24	0.34	3.71

Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence. Ident2 is the percentage sequence identity of the whole template chains with query sequence. Cov. represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein. Normal Z-score is the normalized Z-score of the threading alignments. Alignment with a normalized Z-score >1 mean a good alignment and vice versa. The result of the different models generated by I-TASSER is shown in Table 5.

Table 5: I-TASSER Modelling Score

Name	C-score	Exp. TM-Score	Exp. RMSD	No. of decoys	Cluster Density
Model1	-1.79	0.50+-0.15	13.7+-4.0	76	0.0323
Model2	-2.07			58	0.0246
Model3	-2.14			50	0.0229
Model4	-2.11			48	0.0236
Model5	-2.23			42	0.0208

C-score is a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated on the basis of the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5 to 2] where, a C-score of higher value signifies a model with a high confidence and vice-versa. TM-score and RMSD are known standards for measuring structural similarity between two structures which, are usually used to measure the accuracy of structure modelling when the native structure is known. A TM-score >0.5 indicates a model of correct topology and a TM-score <0.17 means a random similarity. This cut-off does not depend on the protein length. The first model is the best out of the others modelled by I-TASSER. It has a C-score of -1.79 which is the highest among the rest. A higher C-score value signifies a model with a high confidence. The TM-score is 0.35 - 0.65 which indicates a correct topology.

Software other than I-TASSER like Swiss-Pdb Viewer, PHYRE 2, M4T Server, ModWeb, RaptorX, etc were not used for building a 3D model for HDAC-5. This is because the coverage for these Servers & Software's was very low (< 50%), because of which the authentication of the modelled structure was low. Due to this limitation, I-TASSER was used to model the 3D structure of HDAC-5. The total coverage was only 41% in case on RaptorX. This means that only 469 residues of the 1122 residues were modelled. In case of PHYRE 2, the total coverage was only 32% which means only 361 residues of the 1122 residues were modelled. While in the case of I-TASSER the C-score was -1.70 which was in the range of [-5 to 2].^{13, 14}

Model Refinement & Validation

Energy Minimization of the 3D structure was done using Swiss-Pdb Viewer. HDAC-5 has Torsion of 7540.591, Electrostatic Energy -25198.29 KJ/mol and total energy -15077.916 KJ/mol. Figure 1 shows the refined structure of HDAC-5 using UCSF Chimera.

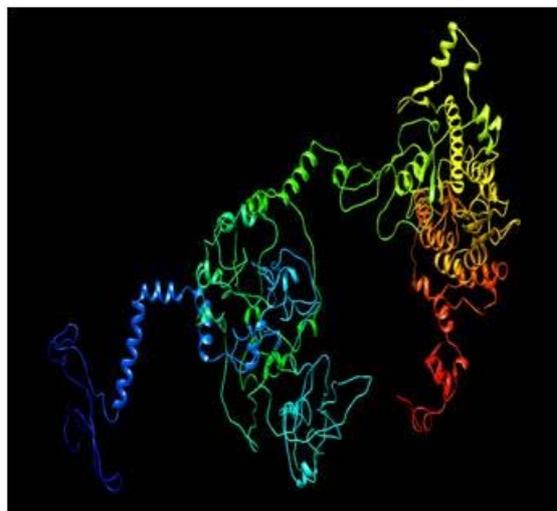


Figure 1: Energy Minimised Model is displayed using UCSF Chimera.

The 3D model was validated by RAMPAGE Server by verifying the parameter of Ramachandran Plot quality (Figure 2). The Plot Analysis are presented in Table 6.

Table 6: Plot statistics of the modelled HDAC-5

Number of residues in favoured region (-98.0% expected)	890 (79.5%)
Number of residues in allowed region (-2.0% expected)	143 (12.8%)
Number of residues in outlier region	87 (7.8%)

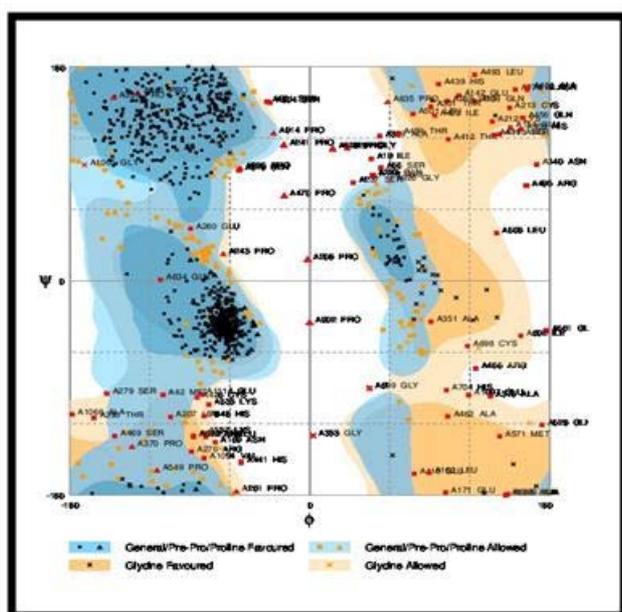


Figure 2: Ramachandran plot of the modelled HDAC-5.

CONCLUSION

This study presents the methodology for generating 3D structure models for proteins whose crystal structures are not available. A computational model of HDAC-5 was successfully designed by combining the primary sequence studies, secondary structure analysis, molecular modelling and validation approaches. This 3D model of HDAC-5 can further be used for various drug discovery studies.

REFERENCES

1. Glozak MA, Seto E, Histone Deacetylases and Cancer, *Oncogene*, 26, 2007, 5420–5432.
2. Barneda-Zahonero B, Parra M, Histone Deacetylases and Cancer, *Molecular Oncology*, 6, 2012, 579-589.
3. Hagelkruys A, Sawicka A, Rennmayr M, Seiser C, The biology of HDAC in Cancer: The Nuclear and Epigenetic Components, *Handbook of Experimental Pharmacology*, 206, 2011, 13-37.
4. Marks PA, Richon VM, Rifkind RA, Histone Deacetylase Inhibitors: Inducers of Differentiation or Apoptosis of Transformed Cells, *Journal of the National Cancer Institute*, 92, 2000, 1210–1216.
5. De Ruijter AJM, Van Gennip AH, Caron HN, Kemp S, Van Kuilenburg ABP, Histone Deacetylases (HDACs): Characterization of the Classical HDAC family, *Biochem. J.*, 370, 2003, 737-749.
6. Sabeena M, Jamil K, Functional Characterization and Structure Prediction of HDAC-10: A Bioinformatics Approach, *International Journal of Science and Research*, 3, 2014, 1792–1800.
7. http://uswest.ensembl.org/Homo_sapiens/Location/View?db=core:g=ENSG00000108840;r=17:44076746-44123702 [Last Accessed on 2nd August, 2014 on 14:42].
8. <http://web.expasy.org/cgi-bin/protparam/protparam1?Q9UQL6@noft> [Last Accessed on 3rd August, 2014 on 18:51].
9. http://www.softberry.com/cgi-bin/programs/propt/cys_rec.pl [Last Accessed on 3rd August, 2014 on 19:22].
10. <http://www.enzim.hu/hmmtop/server/hmmtop.cgi> [Last Accessed on 3rd August, 2014 on 19:43].
11. <http://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S188421> [Last Accessed on 13th August, 2014 on 17:09].
12. <http://mordred.bioc.cam.ac.uk/~rapper/rampage2.php> [Last Accessed on 15th August, 2014 on 12:32].
13. <http://raptorx.uchicago.edu/StructurePrediction/status/93722718/> [Last Accessed on 5th August, 2014 on 15:37].
14. http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/97719ea8bc10e222/summary.html [Last Accessed on 5th August, 2014 on 16:08].

Source of Support: Nil, Conflict of Interest: None.