International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Functional Characterization and Structure Prediction of HDAC-10: A Bioinformatics Approach

Sabeena M [1], Kaiser Jamil[2]

[1]Research Associate, Centre for Biotechnology and Bioinformatics, Jawaharlal Nehru Institute of Advanced Studies (JNIAS), Buddha Bhawan, 6th Floor, M.G.Road, Hyderabad-500003, Telangana, India

[2]Dean, School of Life Sciences, & Director, Centre for Biotechnology and Bioinformatics, Jawaharlal Nehru Institute of Advanced Studies (JNIAS), Buddha Bhawan, 6th Floor, M.G.Road, Hyderabad-500003, Telengana, India

Abstract: ***In eukaryotic cells, nuclear DNA is packaged around histone proteins. There are two enzymes namely Histone Acyl Transferases (HATs) and Histone Deacetylases (HDACs) that regulate the acetylation of histones and helps to condense the chromatin in its stable form. HDACs are considered as one of the promising targets in cancer biology studies. There are three major classes of mammalian HDACs which comprise of 18 genes. Since the crystal structure of HDAC-10 is not available yet, this molecular modeling study can be of great importance in structural biology especially in drug discovery researches. The aim of this study was to develop and validate model structure and the active site determination of HDAC-10 which is a class II member using Bioinformatics applications. The design of the model is based on a thorough evaluation of the HDAC-10 query sequence, Q969S8 by its primary sequence analysis using Expasy Protparam, Secondary analysis using Expasy SOPMA, Sequence similarity and alignment recognition using NCBI Protein BLAST, template alignment using LALIGN, and molecular modeling and dynamics studies using Swiss model and Deepview. Based on the individual sequence alignment scores obtained for query sequence, Q969S8 from LALIGN, four templates (3C0Y_A, 2VQO_A, 2VQJ_A and 2VQW_G) were selected for modeling. Each of these four templates was used for the modeling with the help of Swiss model and generated four corresponding models. The constructed models underwent minimization process in Deep View Swisspdb viewer. Validation protocols assessed the structure, fold and stereo chemical quality of the model. Rampage program showed model1 based on 3C0Y_A template and model 2 on 2VQO_A both were reliable but model 1 was considered as the best one, which showed 95.5% residues in the most favorable region. Verify 3D acquired best value with 89.39% of the residues had an average 3D-1D score of > 0.2 and Prosaweb recorded a Z score of -7.65. The predicted HDAC-10 model1 and its active site residues can be used for target based drug discovery process and it could be a promising receptor for docking and scoring studies.***

Keywords: Template alignment**,** HDAC-10 model building and evaluation, Rampage program, active site prediction

1. Introduction

Histones are the basic protein and its acetylation can be considered as an important factor governing gene expression by its effects on chromatin structure and assembly [1]. There are many studies that reveal the mechanism of histone acetylation and deacetylation and its role in transcription. Histone deacetylase (HDACs) regulates various cellular processes through enzymatic deacetylation of both histone and nonhistone proteins [2]. Also several recent studies have implicated that, individual HDAC enzymes as potential therapeutic targets in the development of cancer. [7], [8]. Histone deacetylases superfamily consists of 18 genes which are divided into two families and four classes - I, II, and IV, in eukaryotic cells. All classes consist of 11 family members, which are referred to as “classical” HDACs, whereas 7 of class III members are called “sirtuins” [3]. There are two major classes of this histone deacetylase family- class I and class II, both share a highly conserved catalytic domain which is considered as receptor for many drugs especially in cancer studies [4]. HDAC10 is included in class II Histone Deacetylase family. The homology comparison studies indicate that HDAC10 is most similar to HDAC6. Both HDAC-10 and HDAC-6 contains a unique catalytic domain which is not found in other HDACs. However this domain does not seem to be functional in HDAC-10 [5]. It is believed that the leucine-rich domain of HDAC10 is responsible for its cytoplasmic enrichment [6]. In recent studies, histone deacetylases (HDACs) were proved to be novel molecular targets for the treatment of cancer [7], [8], [9], [10]. The role of HDAC in cancer was studied for the first time 1998 [11]. It has been found that the knockdown of HDAC-10 significantly increased the mRNA expression levels of thioredoxin interacting protein (TXNIP) in human gastric cancer cells [12]. Histone deacetylase inhibitors (HDACi) have extensively demonstrated the antitumor efficacy in vitro and in vivo [13]. Therefore, the determination of HDACi has become one of the most important research fields of the antitcancer drugs.

Since at present there is no crystal structure data available for HDAC-10 our aim was to construct this model. Molecular modeling techniques can build and refine the model for the target which can subsequently be used for docking studies for determining the possible lead molecules as reported by earlier researches [14], [15]. Also these studies can provide valuable information regarding binding sites of receptor which are crucial elements for ligand binding. Similar kind of studies has been carried out in case of other classes of HDAC superfamily proteins. Hence our objective was to construct and validate a model of HDAC-10 and determine the active site residues by computational tools. Computational Biology application efforts have now provided the biologists with the information of large number of Databases development and management [16]. These data repositories help us to understand the entire mechanism of biology through IT applications.

Our hypothesis is based on the fact that bioinformatics and drug discovery studies play an important role in determining and predicting the efficiency of novel molecules which can turn to medicine. Technological advances have greatly increased the real understanding of the molecular basis of tumor progression and treatment response which facilitates the development of effective medicines [17]. Recently, a study was published that found the tunnel behavior of HDAC active sites and it suggested the isoform based drug discovery process [18]. We propose a method to construct a suitable model for query sequence, Q969S8 of HDAC-10 by finding related sequences as template.

2. Materials and Methods

2.1. Sequence retrieval and primary analysis

At the outset, the query sequence of HDAC-10 with the accession id Q969S8 was retrieved from UniprotKB resource [19]. The complete sequence consists of 669 residues. The physicochemical properties of Q969S8 were determined using Expasy ProtParam tool. ProtParam in Expasy Proteomics Server computes various physicochemical properties that can be deduced from a protein sequence [20].We computed properties such as theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues in the query, its value extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY).

2.2 Secondary Structure analysis

Using Expasy SOPMA tool the secondary structure of our query Q969S8 was predicted. We imported the query protein sequences in FASTA format and submitted to SOPMA server. SOPMA is Self Optimized Prediction Method of Alignment, which is useful for predicting secondary structure of proteins [21]. Self optimized prediction method is based on the homologue method. Using SOPMA we obtained details of the secondary structure components such as percentage of α-helix, β-sheets, turns, random coils and extended strands present in the query Q969S8.

2.3 Template recognition/ Query – Template Alignment studies

The HDAC-10 query (Q969S8) was analyzed in NCBI BLASTP against Protein Data Bank (PDB) and visually inspected to determine the possible Hits. We considered the identity cutoff range between 35-45% and the finding of the most identical sequences has resulted 8 Hits. NCBI Blast and LALIGN programs based sequence analysis and comparison was adopted to mark all the conserved amino acid residues among the most four similar templates to query protein Q969S8. NCBI Blast Identity and LALIGN identity were recorded for each template. The possible identical residues occurred in each of the sequences were denoted as ':' and '.' represents the conservative replacements among them. Each of these shortlisted template sequences were aligned with query sequences by using LALIGN program, which required two files such as a file containing target query in FASTA format and a file containing template in FASTA format. We feel that this step is essential to detect the consensus residues or binding site residues present in both the query and the template sequences. The calculated scores from BLAST were reevaluated by using LALIGN score and 4 templates were selected as suitable templates for model generation. 3C0Y_A, 2VQO_A, 2VQJ_A and 2VQW_G are the templates selected for model building, which represents the PDB accession id and the corresponding chain information. Local alignment between the query and the template was applied to this selection.

2.4 Molecular modeling and validation studies

An initial phase of the molecular modeling was carried out using Swiss model. 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) [22], [23]. Four models were generated based on the selected templates (3C0Y_A:T1, 2VQO_A :T2, 2VQJ_A :T3 and 2VQW_G :T4). Each of these four generated theoretical models (model-1, model-2, model-3 and model-4) were subjected to Swiss-PDB Viewer for energy minimization using the steepest descent and conjugate gradient technique to verify the stereochemistry feature. For each model developed, interactive analysis of the effects of individual Ramachandran angles on the protein fold were carried out using Swiss PDB viewer. Also Geometry optimization and energy minimization were carried out using GROMOS module.

The generated model was subjected to a series of analysis to determine its stability and reliability. Backbone conformation of the refined model was assessed by the Rampage web server which explains the feature of Psi and Phi angle orientation [24]. Verify 3D Structure Evaluation Server was used to determine the 3D-profiling of the residue in the model [25]. The overall quality score of the model was calculated by ProSA which displayed as a plot. Prosaweb is a tool widely used to check 3D models of protein structures for potential errors [26]. Errat was used to understand the statistics of various bond interactions between different atom types by plotting the graph.

2.5 Active Site analysis and residues recognition

In order to proceed further with the ligand binding studies by docking, active site analysis was performed by submitting the model-1 to Castp server (http://cast.engr.uic.edu). [27].

3. Results/Discussion

3.1 Sequence retrieval and Primary Sequence analysis using Expasy Protparam

Our first step of analysis was to identify the query sequence of HDAC-10 (Q969S8) with 669 residues of amino acids which was retrieved from UniprotKB (Table 1)

Table 1: HDAC-10 query sequence (Q969S8) with 669 residues

>sp|Q969S8|HDA10_HUMAN Histone deacetylase 10 OS=Homo sapiens
MGTALVYHEDMTATRLLWDDPECEIERPERLTAA
LDRLRQRGLEQRCLRLSAREASEEELGLVHSPEYV
SLVRETQVLGKEELQALSGQFDAIYFHPSTFHCAR
LAAGAGLQLVDAVLTGAVQNGLALVRPPGHHGQ
RAAANGFCVFNNVAIAAAHAKQKHGLHRILVVD
WDVHHGQGIQYLFEDDPSVLYFSWHRYEHGRFW
PFLRESDADAVGRGQGLGFTVNLPWNQVGMGNA
DYVAAFLHLLLPLAFEFDPELVLVSAGFDSAIGDP
EGQMQATPECFAHLTQLLQVLAGGRVCAVLEGG
YHLESLAESVCMTVQTLLGDPAPPLSGPMAPCQS
ALESIQSARAAQAPHWKSLQQQDVTAVPMSPSSH
SPEGRPPPLLPGGPVCKAAASAPSSLLDQPCLCPAP
SVRTAVALTTPDITLVLPPDVIQQEASALREETEA
WARPHESLAREEALTALGKLLYLLDGMLDGQVN
SGIAATPASAAAATLDVAVRRGLSHGAQRLLCVA
LGQLDRPPDLAHDGRSLWLNIRGKEAAALSMFH
VSTPLPVMTGGFLSCILGLVLPLAYGFQPDLVLVA
LGPGHGLQGPHAALLAAMLRGLAGGRVLALLEE
NSTPQLAGILARVLNGEAPPSLGPSSVASPEDVQA
LMYLRGQLEPQWKMLQCHPHLVA

The advantage of using this UniprotKB database is to find a centralized repository of protein sequences with comprehensive coverage and a systematic approach to protein annotation, incorporating, interpreting, integrating and standardizing data from large and disparate sources and this was the most comprehensive catalog of protein sequence and functional annotation [19].

The primary and secondary features of the query were studied using Expasy proteomics server. By using ProtParam tool, the physicochemical characters were analyzed. In Protparam, no additional information was required about the query protein. The query sequence can either be specified as a Swiss-Prot/TrEMBL accession number or ID, or in the form of a raw sequence. White space and numbers were ignored.

The molecular weight of HDAC-10 obtained was 71444.8, Number of residues in the query were 669 aa, Theoretical pI was 5.44, Extinction coefficient in 280 nm solution was 66765. The half-life was estimated as 30 hours (unit measured based on mammalian reticulocyte in vitro studies) and the instability index was found to be 45.46. (Table 2 a).

Table 2 a: Physical and Chemical parameters for HDAC-10, Q969S8 protein sequence were analyzed by Expasy Protparam

Parameters	*Value*
Total number of negatively charged residues (Asp + Glu)	68
Total number of positively charged residues (Arg + Lys)	43
Ext. coefficient	66765
The estimated half-life is	30 hours (value measured based on mammalian reticulocytes, in vitro)
The instability index (II)	45.46
Aliphatic index	99.10
Grand average of hydropathicity (GRAVY)	0.070

3.2 Secondary Structure analysis using Expasy SOPMA

The self-optimized prediction method (SOPM) has been developed to improve the success rate in the prediction of the secondary structure of proteins. Even though, with the absence of Bend region, 3^{10} helices and other states structural information, the SOPMA results showed that the query Q969S8 protein had a high level of Alpha helix secondary structure of more than 40% and highly variable random coils of nearly 40%. SOPMA results infers that the protein consists of abundant helices with 277 residues is 41.41%, Random coil (Cc) is of 247 is 36.92%, Extended strand (Ee) 102 is 15.25% etc. It also suggests that the structure have less number of Beta turn (Tt) with 43 residues only which is 6.43% of the entire sequence. (Table 2 b).

It seems likely that the structural elements suggested by SOPMA are the ensemble of diverse and stable nature of the protein conformation of HDAC-10. Also it is possible that the presence of more helix can make the structure more stable. SOPMA program determined the role of individual amino acid for building the secondary structure with their positions. SOPMA is neural network based method; global sequence prediction may be done by this sequence method [30].

Table 2b: Secondary structure analysis for HDAC-10, Q969S8 protein sequence were analyzed by SOPMA of ExPASy's server

Secondary structure	**Percentage**
Alpha helix (Hh)	277 is 41.41%
3_{10} helix (Gg)	0 is 0.00%
Pi helix (Ii)	0 is 0.00%
Beta bridge (Bb)	0 is 0.00%
Extended strand (Ee)	102 is 15.25%
Beta turn (Tt)	43 is 6.43%
Bend region (Ss)	0 is 0.00%
Random coil (Cc)	247 is 36.92%
Other states	0 is 0.00%

3.3Template determination and query-template matching studies

Although, initially thought to be a HDAC superfamily member, but HDAC-10 showed a relatively low sequence similarity with other database sequences (average range of

35-45%) from NCBI Blast search (Table 3). We have chosen minimum 35% as identity cutoff value for the selection of template. The BLAST algorithm calculated similarity scores for local alignments (i.e., the most similar regions between 2 sequences) between the query sequence and subject sequences using specific scoring matrices, and returns a table of the best matches or "hits" from the database sequences [28].

Table 3: The finding of the most identical sequences for The Query sequence of HDAC-10, Q969S8 in NCBI Protein BLAST has resulted 8 Hits.

Accession	Identity (%)	Query cover (%)	e-Value	Total Score
2VQO_A	38	63	4e-72	278
2VQW_G	38	63	6e-72	278
2VQQ_A	38	63	3e-72	240
2VQJ_A	38	63	4e-72	239
3C0Y_A	38	61	2e-72	233
1ZZ0_A	41	30	9e-39	148
2VCG_A	41	30	1e-38	148
3MEN_A	42	27	4e-25	108

First Column shows the accession id of the resultant entry (eg: 2vQO shows the PDB id and A shows the corresponding chain information, same representation for other 7 more entries). Identity represents extent to which the query sequence Q969S8 shows the occurrence of same residues at the same positions for each template alignment and it shows in percentage value. Query cover denotes the percent of the query sequence Q969S8 that overlaps the subject sequence and e-Value (Expect Value) is the parameter which describes the number of hits expected while searching with database entries. Total score is the entire score obtained after the alignment of query with each of the template sequence. LALIGN program were considered for the selection for four template with the identity scores such as 40.6%, 32.9%, 32.9% and 32.9% and the corresponding templates T1, T2, T3, and T4 respectively) (Figure 1).

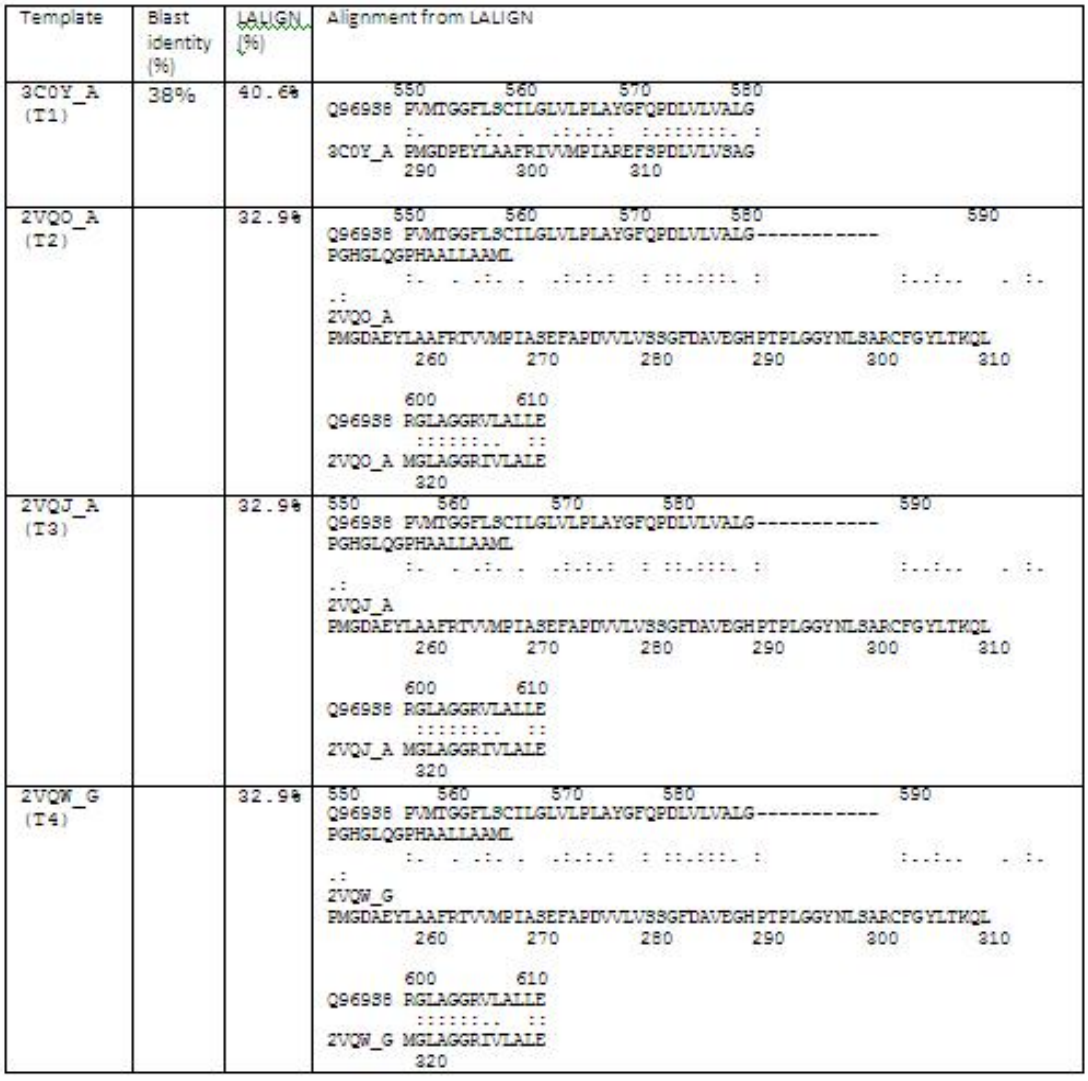


Figure 1: Templates recognition for HDAC-10 model building

We propose that, this study may facilitate direct assessment of the effectiveness of the model building procedure based on sequence identity between the query and template. Thus two different criteria for ranking were suggested to evaluate the sequence similarity between the query protein and template for model. One is based on the NCBI Blast

alignment and other one is the further verification of the individual similarity between the query and template using LALIGN [29]. LALIGN results shows that out of four templates studied, T1 and T2 have the high score and identity compared to T3 and T4. Also the fact that the four templates selected belongs to the same HDAC superfamily which may be viable choice as the template for a proper homology modeling studies.

3.4 Molecular modeling studies

Homology/comparative modeling of protein structures usually requires extensive expertise in structural biology and the use of highly specialized computer programs for each of the individual steps of the modeling process [31]. The concept of automated molecular modeling facility with integrated expert knowledge was first implemented in 1993 [32].

Based on the eight sequence alignments using NCBI protein Blast and four templates (Table3) obtained from LALIGN program, Homology models were generated by using Swissmodel automated program. Out of the four models generated two of them were selected for further energy minimization and validation studies which are represented as model1 and model2 and were visualized using Accelrys DS visualiser (Figure 2 a and b).

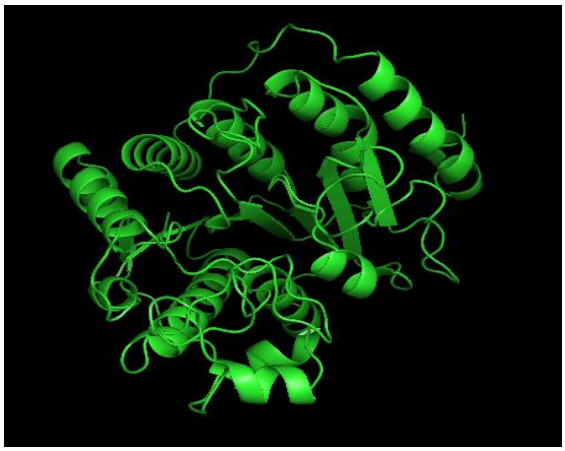


Figure 2 a: The proposed tridimensional structure of HDAC-10 was obtained by comparative modeling based on the template 1 (3C0Y_A: T1) using Swiss model and visualized with PyMOL visualization software

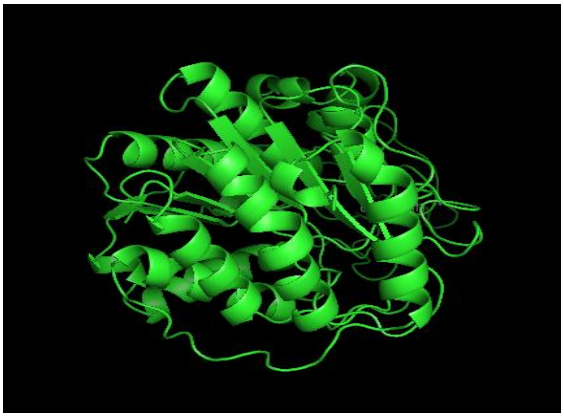


Figure 2 b: The proposed tridimensional structure of HDAC-10 was obtained by comparative modeling based on the template 2 (2VQO_A (T2) using Swiss model and visualized with PyMOL visualization software

The energy minimization for these two models (model1 and model2) were carried out on the steepest descent energy minimization using the GROMOS96 force field in Deep View Swiss PDB viewer based on the concept that, empirical force fields are useful to detect parts of the model with conformational errors [33].

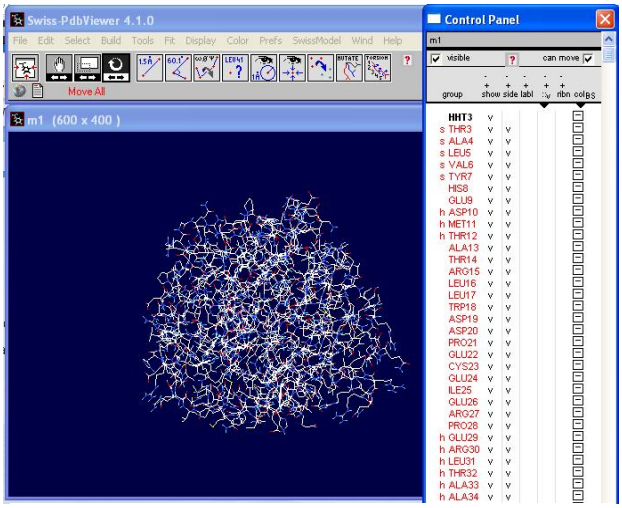


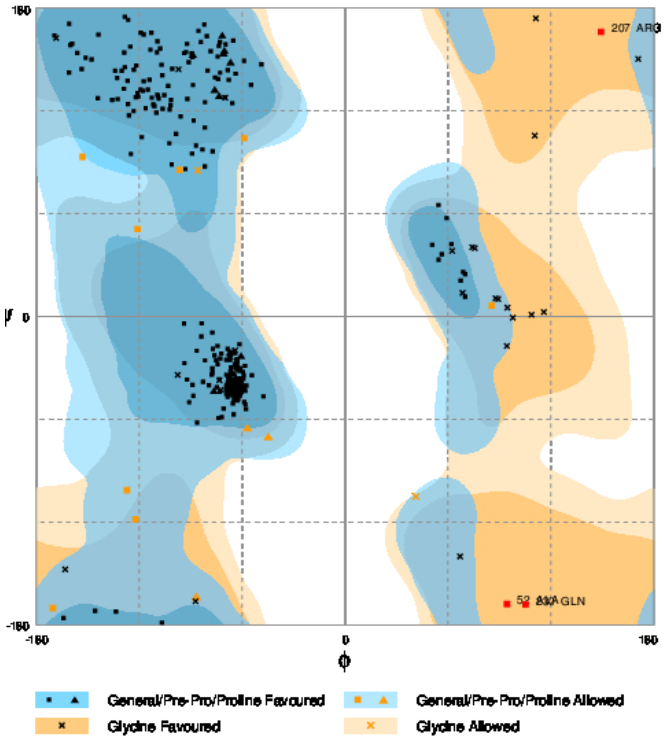
Figure 3: Model 1 in GROMOS96 of SwissPdb Viewer

Also we feel that, in general energy minimization or molecular dynamics methods only are not able to improve the efficacy of the predicted models. The accuracy of the predicted model topology is an essential step, since stereochemistry play a major role on how the newly produced structure behaves in molecular dynamics environment. Thus we cross validated model1 and model2 structure by using various model evaluation tools. The local geometry of the predicted models were determined by Rampage server, which evaluated the steriochemical nature of the psi and phi angle and its spatial arrangement. (Table 4) [34].

Table 4: Ramachandran plot calculations on 3D models of HDAC-10 computed with the Rampage program

Template Name	Residues in favoured region (overall in %)	Residues in disallowed region (overall in %)
3C0Y_A (T1)	14 :THR, 76 :GLN,84 :GLN,176 :GLY,204 :PRO, 205 :PHE, 226 :LEU, 271 :PRO,272 :GLU, 273 :GLY, 276 :GLN,302 :GLU, 328 :PRO (95.5%)	52 :ALA, 207 :ARG, 230 :GLN (0.8%)
2VQO_A (T2)	16 :LEU,20 :ASP,28 :PRO,29 :GLU,79 :GLY,176 :GLY,209 :SER, 213 :ALA, 231 :VAL, 232 :GLY, 246 :LEU, 254 :ASP, 271 :PRO, 302 :GLU,305 :TYR (92.8%)	18 :TRP,27 :ARG,75 :THR,89 :GLN,207 :ARG226 :LEU,227 :PRO,233 :MET,274 :GLN,332 :PRO,333 :MET (3.0%)

As seen from the figure, in the model1 based on the template 3C0Y_A (T1) in Rampage Ramachandran plots identified the probable number of residues in most favoured region were 339 (95.5%), number of residues in allowed region were 13 (3.7%) and number of residues in outlier region were only 3 (0.8%). In the case of model2 based on template 2VQO_A (T2), Rampage Ramachandran plots identified the probable number of residues in most favoured region were 335 (92.8%), number of residues in allowed region were recorded as 15 (4.2%) and the number of residues in the outlier region were 11(3.0%) only. (Figure 4 a and b).



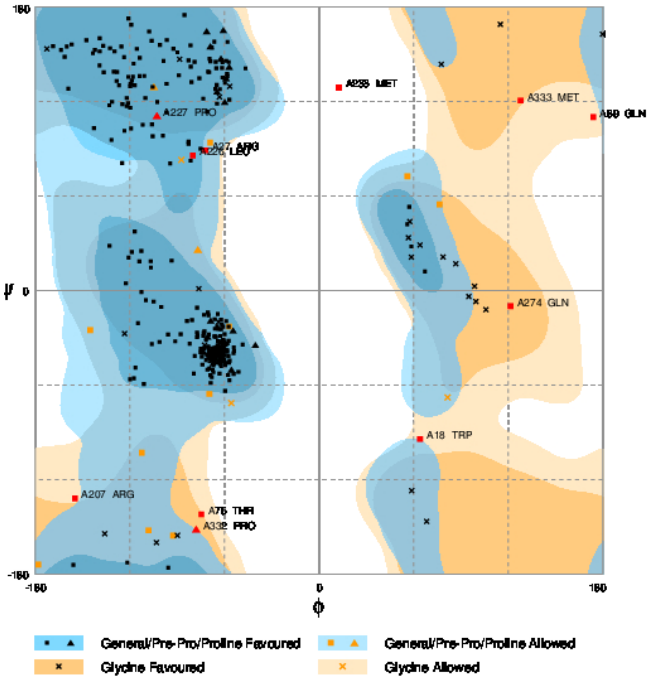


Figure 4 a and b shows Assessing the steriochemical stability of generated models model1 and model2 by Rampage Ramachandran Plot.

As seen in the Figure 4 a and b, model1 generated based on the template was found to be more reliable and as a best one as per the psi-phi orientation. According to Ramachandran Plot, a good quality model should have >90% residues in favored region, thus the final model was validated as good quality model and the structure was viewed via Accelrys DS visualiser and it depicted an alpha helix rich structure.

The steriochemical quality of both the models (model1 and model2) was further verified by Verify 3D, which is a Structure Evaluation Server designed to help in the refinement of crystallographic structures. The problem of finding which amino acid sequences fold into a known three-dimensional (3D) structure, can be effectively attacked by finding sequences that are most compatible with the environments of the residues in the 3D structure [25].Verify3D usually analyse the crystal structure in PDB format which provides the plot with a visual analysis of the quality of a putative crystal structure for a protein. In the Verify3D plot, the vertical axis represents the average 3D-1D profile score for residues in a 21-residue sliding window, the center of which is at the sequence position indicated by the horizontal axis [35]. In the model1 based on the template 3C0Y_A (T1) Verify3D plots recorded 89.39% of the residues had an average 3D-1D score of > 0.2 (Figure 3a) whereas in the case of the model2 based on the template 2VQO_A (T2), showed Verify3D plots of 78.30% of the residues had an averaged 3D-1D score > 0.2 (**Figure 5a and b)**

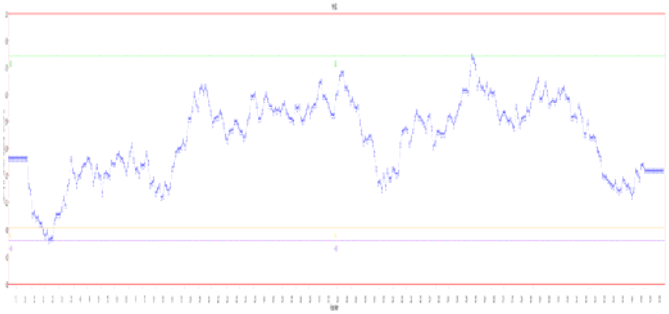


Figure 5a: Verify3D Plots obtained for model1 based on the template 3C0Y_A (T1) which recorded 89.39% of the residues had an averaged 3D-1D score of > 0.2.

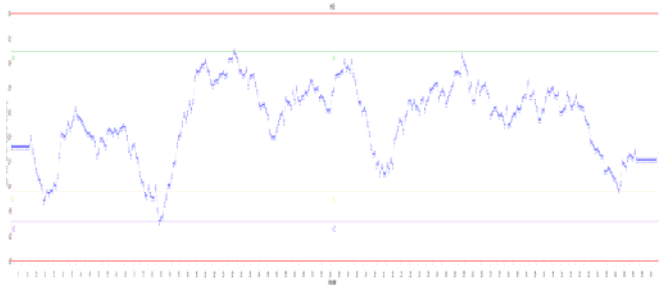
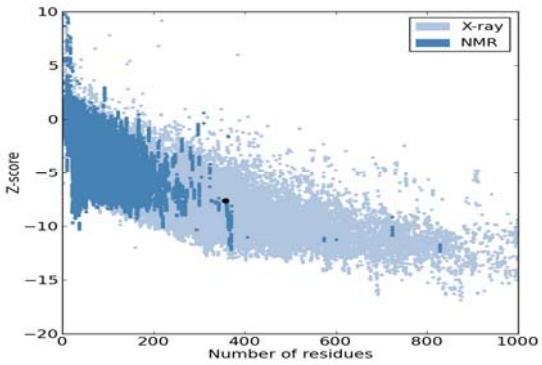


Figure 5b: Verify3D Plots obtained for model2 based on the template 2VQO_A (T2) which recorded 78.30% of the residues had an averaged 3D-1D score of > 0.2.

The profile score above zero in the Verify3D graph corresponded to the acceptable environment of the model. ProSA (3), a tool widely used to check 3D models of protein structures for potential errors. Prosa web results of model1 based on the template 3C0Y_A (T1) shows that the overall good model quality of the full-length protein lies within the crystal structure defined limit for protein with 357 amino acid residues with Z-Score value of -7.65 (Figure 6a) and results of model2 based on the template 2VQO_A (T2), shows the protein with 363 amino acid residues and Z score value of -6.72. (Figure 6b). The Z-score is indicative of overall model quality which is used to check whether the input structure is within the range of scores typically found in native proteins of similar size.



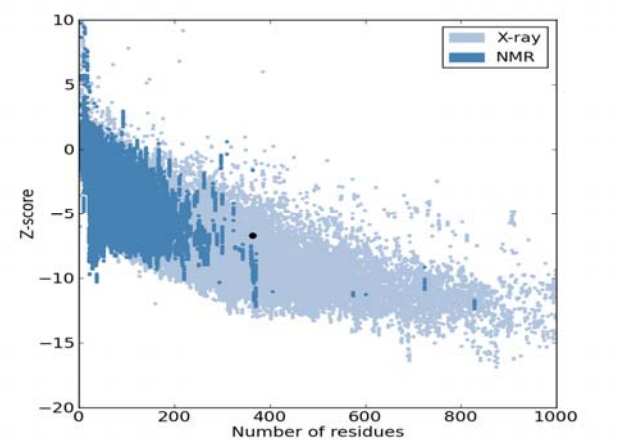


Figure 6a and b shows Investigation of two theoretical models of HDAC-10 based on the templates 3C0Y_A and 2VQO_A ProSA-web z-scores for 3C0Y_A shows 357 amino acid residues with Z-Score value of -7.65 and ProSA-web z-scores for 2VQO_A shows that 363 amino acid residues and Z score value of -6.72. In each of the figures, residue energies averaged over a sliding window are plotted as a function of the central residue in the window. A window size of 80 is used due to the large size of the protein chain (default: 40).

3.5 Active Site analysis and residues recognition

Cavities on a receptor as well as its specific amino acid positioning within it create the physicochemical properties needed for a protein to perform its function. Pockets are empty concavities on a protein surface into which solvent (probe sphere 1.4 A) can gain access. CASTp is an online tool that locates and measures pockets and voids on 3D protein structures. This new version of CASTp includes annotated functional information of specific residues on the protein structure. The annotations are derived from the Protein Data Bank (PDB), Swiss-Prot, as well as Online Mendelian Inheritance in Man (OMIM) [27].

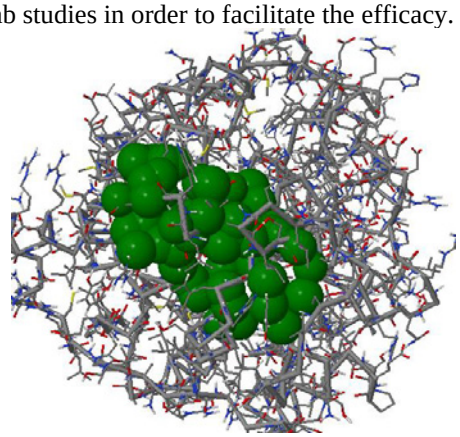
CASTp identifies all pockets and voids on the receptor structure and provided detailed delineation of all atoms participating in their formation. It also measures the volume and area of each pocket and void analytically, using both the solvent accessible surface model (Richards' surface) and molecular surface model (Connolly's surface). The default value used by CASTp server was 1.4 Angstroms. Only values between 0.0 and 10.0 will be accepted. CASTp is based on recent theoretical and algorithmic results of Computational Geometry. It has many advantages: 1) pockets and cavities are identified analytically, 2) the boundary between the bulk solvent and the pocket is defined precisely, 3) all calculated parameters are rotationally invariant, and do not involve discretization and they make no use of dot surface or grid points [36]. Active site residue determination of the predicted and validated model (model1) through CASTp is shown (Figure 7).

The default probe radius value used by CASTp server is 1.4 Angstroms. Only values between 0.0 and 10.0 will be accepted. The residues included in active site were 14THR,16LEU,19ASP,21PRO,24GLU,26LUC,27ARG,28P

RO,67LEU,68GLU,69ASP,70ALA,76ILE,77VAL,121GLU, 130ARG,131PRO,144PHE and146VAL

Figure 7: active site determination of model 1 using CASTP server. Only values between 0.0 and 10.0 will be accepted. Green color residues represent the possible active site residues.

Active site determination through CASTp suggests that this predicted and validated model (model1) can be utilized as a potential drug target. However, these will further be tested by wet lab studies in order to facilitate the efficacy.



As per the results from Rampage server, in model1 there were three residues (Residue [52:ALA] (94.52,-167.66) in Outlier region, Residue [207 :ARG] (149.30, 165.96) in Outlier region, Residue [230 :GLN] (105.31,-168.01) in Outlier region), which were in the disallowed region in the predicted model. From Castp Server results, with careful visual inspection we could verify that these three residues were found to be located far from the binding site region. This region never occurs to affect the ligand binding process by any means. The overall results of our study strongly suggest that the model 1 was more reliable for further analysis. (Table 5)

Table 5: Model evaluation of HDAC-10 using various validation softwares

Template Name	A. Rampage Residues in favoured region (overall in %)	B. Verify3D plot (3D-1D score)	C. Prosa web (Z score)
3C0Y_A (T1)	95.5%	89.39% (> 0.2)	357 (-7.65)
2VQO_A (T2)	92.8%	78.30%(> 0.2)	363 (-6.72)

4. Conclusion

This study presents the methodology for generating 3D structure models for proteins whose crystal structures are not available. We have successfully constructed two computationally viably designed models of HDAC-10. Further this elucidation could aid in studies on the molecular mechanism of inhibiting HDAC-10. A meaningful and more realistic study can be derived by selecting this model as target for performing the molecular docking studies to determine the possible lead molecules which can inhibit the activity of HDAC-10. By combining primary sequence studies, secondary structure analysis, sequence analysis and alignment, molecular modeling and validation approaches together; we could determine and validate this 3 dimensional model for the query protein of HDAC-10.

5. Future Scope

This study proposes the fact that a fast and reliable homology model was possible by considering the sequences with profound similarity at sequence level. The present study also records a set of active site residues which is expected to greatly help in the development of novel HDAC inhibitor molecules by docking and scoring studies. Our models of HDAC-10 are only predictive, and needs to be confirmed experimentally. The active site residue prediction criteria used in this study can be applied to model other types of functional residues such as binding pocket specificity determinants and interaction interfaces between the members in the superfamily of HDAC's.

6. Acknowledgments

The authors gratefully acknowledge the support from JNIAS for the successful completion of this project.

References

[1] Min-Hao Kuo, C. David Allis (1998) Roles of histone acetyltransferases and deacetylases in gene regulation. BioEssays 20: 615–626.
[2] Geneviève P Delcuve, Dilshad H Khan, James R Davie (2012) Roles of histone deacetylases in epigenetic regulation: emerging paradigms from studies with inhibitors.Clinical Epigenetics Biomed Central 4:5,
[3] Witt O, Deubzer HE, Milde T, Oehme I (2009) HDAC family: What are the cancer relevant targets?. Cancer Letters 277: 8–21.
[4] Paul A. Marks, Victoria M. Richon, Richard A. Rifkind, (2000) Histone Deacetylase Inhibitors: Inducers of Differentiation or Apoptosis of Transformed Cells. Journal of the National Cancer Institute 92: 1210–1216.
[5] Amaris R. Guardiola, Tso-Pang Yao, (2002) Molecular Cloning and Characterization of a Novel Histone Deacetylase HDAC10. The Journal of Biological Chemistry 277: 3350 –3356.
[6] Jenny J. Tong, Jianhong Liu, Nicholas R. Bertos, Xiang-Jiao Yang (2002) Identification of HDAC10, a novel class II human histone deacetylase containing a leucine-rich domain. Nucleic Acids Research 30: 1114–1123.
[7] Hagelkruys A, Sawicka A, Rennmayr M, Seiser C (2011) The biology of HDAC in cancer: the nuclear and epigenetic components, Handbook of Experimental Pharmacology 206: 13-37.
[8] Ren J, Zhang J, Cai H, Li Y, Zhang Y, et al. (2013) HDAC as a therapeutic target for treatment of endometrial cancers. Current Pharmaceutical Design
[9] Bertrand P (2010) Inside HDAC with HDAC inhibitors. European Journal of Medicinal Chemistry 45: 2095–2116

[10] Hidemi Rikiishi (2011) Autophagic and apoptotic effects of HDAC inhibitors on cancer cells. Journal of Biomedicine and Biotechnology,
[11] Warrell RP Jr, He LZ, Richon V, Calleja E, Pandolfi PP (1998) Therapeutic targeting of transcription in acute promyelocytic leukemia by use of an inhibitor of histone deacetylase. Journal of the National Cancer Institute 90:1621–1625.
[12] Ju-Hee Lee, Eun-Goo Jeong, Moon-Chang Choi, Sung-Hak Kim, Jung-Hyun Park, et al, (2010) Inhibition of histone deacetylase 10 induces thioredoxin-interacting protein and causes accumulation of reactive oxygen species in SNU-620 human gastric cancer cells. Molecules and Cells 30: 107-112.
[13] Praveen Rajendran, Emily Ho, David E Williams, Roderick H Dashwood (2011) Dietary phytochemicals, HDAC inhibition, and DNA damage/repair defects in cancer cells. Clinical Epigenetics Biomed Central: 3:4
[14] Sabita Kotra, Madala KK, Kaiser Jamil (2008) Homology models of the mutated EGFR and their response towards quinazolin analogues. Journal of Molecular Graphics and Modeling 27(3): 244-254
[15] Kaiser Jamil, Sabeena Muhammed Mustafa (2012) Thioredoxin System: A Model for Determining Novel Lead Molecules for Breast Cancer Chemotherapy. Avicenna Journal of Medical Biotechnology 4:3: 121-130.
[16] Cynthia Gibas & Per Jambeck (2001) Developing Bioinformatics Computer Skills
[17] Sunali Mehta, Andrew Shelling, Anita Muthukaruppan, Annette Lasham, Cherie Blenkiron, et al, (2010) Predictive and prognostic molecular markers for cancer medicine. Therapeutic Advances in Medical Oncology 2(2): 125–148.
[18] Sundarapandian Thangapandian, Shalini John, Yuno Lee, Venkatesh Arulalapperumal, Keun Woo Lee (2012) Molecular Modeling Study on Tunnel Behavior in Different Histone Deacetylase Isoforms. Plos One 7:11.
[19] Apweiler, Rolf ; Martin, Maria Jesus ; O'Donovan, ,(2010), Claire (UniProt Consortium), The Universal Protein Resource (UniProt) in 2010, Nucleic Acids Research, 38D142-D148.
[20] Elisabeth Gasteiger, Alexandre Gattiker, Christine Hoogland, Ivan Ivanyi, Ron D. Appel and Amos Bairoch, ,(2003), ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Research, 31:133784–3788.
[21] C. Geourjon, G. Deléage, (1995), SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Computer Applications in The Biosciences, Oxford Journals, 11:6681-684.
[22] Arnold K., Bordoli L., Kopp J., and Schwede T, (2006), The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling, Bioinformatics, 22, 195-201.
[23] Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T, (2009), The SWISS-MODEL Repository and associated resources. Nucleic Acids Research, 37, D387-D392.
[24] Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, ,(2003):Word JM, Prisant MG, Richardson JS, Richardson DC, Structure validation by Calpha geometry: phi,psi and Cbeta deviation,50:3437-50.
[25] Bowie JU, Lüthy R, Eisenberg D, 1991, A method to identify protein sequences that fold into a known three-dimensional structure, Science. , 253(5016), 1991, 164-70.
[26] Sippl,M.J, (1993) Recognition of errors in three-dimensional structures of proteins. Proteins, 17, 355–362.
[27] Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. (2006),CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 34
[28] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, (1990), Basic local alignment search tool, Journal of Molecular Biology, 5;215(3), 403-10.
[29] X. Huang and W. Miller, (1991), lalign the Huang and Miller SIM algorithm, Adv. Appl. Math, 12 337-357
[30] Combet C., Blanchet C., Geourjon C. and Deléage G, (2000), NPS@: Network Protein Sequence Analysis, 25:3, 147-150.
[31] Tramontano, Leplae, Morea, (2001), Analysis and assessment of comparative modeling predictions in CASP4. Proteins, 22–38.
[32] Peitsch, M.C. and Jongeneel, C.V, (1993), A 3-D model for the CD40 ligand predicts that it is a compact trimer similar to the tumor necrosis factors. International Immunology, 5, 233–238
[33] Van Gunsteren W.F., Billeter,S.R., Eising,A., Hünenberger,P.H., Krüger,P., Mark,A.E., Scott,W.R.P. and Tironi,I.G. (1996) Biomolecular Simulations: The GROMOS96 Manual and User Guide,
[34] S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, (2002), Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function & Bioinformatics, 50, 437-450
[35] Lüthy R, Bowie JU, Eisenberg D. (1992), Assessment of protein models with three-dimensional profiles. Nature, 35683-85.
[36] T. Andrew Binkowski, Shapor Naghibzadeh, Jie Liang, (2003), CASTp: Computed Atlas of Surface Topography of proteins, Nucleic Acids Research, 31:13, 3352–3355

Author Profile

Sabeena M, is presently doing her PhD in Bioinformatics in Jawaharlal Nehru Technological University, Anantapur, A.P. She had obtained her masters in Bioinformatics from Mahatma Gandhi University, Kerala. She have been working as Research Associate with Centre for Biotechnology and Bioinformatics, Jawaharlal Nehru Institute of Advanced Studies (JNIAS), Hyderabad, A.P. She is mainly focusing on Cancer Biology aspects for her research. She published few papers in International Journals as part of her studies.