Article

Comparative structural analysis of HAC1 in *Arabidopsis thaliana*

Amar Ćemanović, **Jasmin Šutković**, **Mohamed Ragab Abdel Gawwad**
Genetics and Bioengineering department, International University of Sarajevo, Ilidza, 71220 Bosnia and Herzegovina
E-mail: mragab@ius.edu.ba,mrgawad@hotmail.com

Received 30 July 2013; *Accepted 4 September 2013*; *Published online 1 June 2014*

Abstract
Histone acetylation is an important posttranslational modification correlated with gene activation. In *Arabidopsis thaliana*, the histone acetyltransferase 1 (AtHAC1) is homologous to animal p300/CREB (cAMP-responsive element-binding protein)-binding proteins, which are the main histone acetyltransferases participating in many physiological processes, including proliferation, differentiation, and apoptosis. In this study the 3-D structure of the HAC1 protein in *Arabidopsis thaliana* was predictedusing 4 homology-based prediction servers: ESyPred3D, 3D-JIGSAW, SWISS-MODEL and PHYRE2. The homology modeled structureswere evaluated and stereochemical analysis done by Ramachadran plot analysis. The amino acid sequences of *Arabidopsis thaliana* HAC1protein are highly similar to the sequence of the homologous human p300/CREB. SWISS MODEL and Phyre2 servers computed the identical 3D structures. Validation and verification methods, using Z-score and 3D-1D score, showed that these 3D models are of good and acceptable quality.

Keywords *Arabidopsis thaliana*; HAC1 protein; 3-D structure; structure prediction, homology.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Arabidopsis thaliana, also known as thale-cress or mouse-ear cress, or simply Arabidopsis,is a small flowering plant belonging to the *Brassicaceae* family, a plant family comprising species such as radish and cabbage. Although not of agronomic importance, it is a widely known plant due to its use as a model organism in various genetic and plant biology studies. In addition, it is the first plant that had its genome sequenced. The prominent features of Arabidopsis are its small genome (135 Mb), rapid life cycle and ease of cultivation in laboratory conditions (The Arabidopsis Genome Innovative, 2000).

In cells of eukaryotic organisms, including plants, the DNA is associated with histone proteins to form structures called nucleosomes. In nucleosomes DNA is wound around a group of eight histones, which makes the region inaccessible to transcription factors and therefore transcriptionally inactive. Histone acetyltransferases (HATs) are a group of enzymes which function in transcriptional regulation by means of histone modification. Specifically, they transfer acetyl groups from acetyl-CoA to specific residues in histone

tails, thereby lowering the affinity of histones for DNA and, accordingly, making the DNA more accessible to the transcription machinery. The process of acetylation is reversible, with the deacetylation being performed by deacetylase enzymes (Sterner and Berger, 2000).

Based on sequence similarity, histone acetyltransferases have been organized into four families: GNAT, MYST, p300/CBP and $TAF_{II}250$. The p300/CBP family is especially important since it acetylates not only histones, but also various other proteins, including CREB, p53, Stat3 and HIV-1 Tat protein (Sterner and Berger, 2000).This protein family in *Arabidopsis thaliana*includes the following members: HAC1, HAC2, HAC4, HAC5 and HAC12 (Pandey et al., 2002).

From the members of the p300/CBP family, HAC1 has the highest homology with the human and animal p300/CREB protein (Deng et al., 2007). Although the human p300/CREB is well characterized and has an experimentally determined crystal structure, little is known about its homolog in *Arabidopsisthaliana*. It has been found that it is located in the nucleus and exhibits, besides its acetyltransferase activity, also transcription cofactor activity. Furthermore, it is found to acetylate primarily histones H3 and H4 (Earley et al., 2007).

The aim of this paper is to predict the best 3D model for the HAC1 protein in *Arabidopsis thaliana*, using homology modeling techniques by model comparison and evaluation of the predicted structures.

2 Materials and Methods

The protein sequence and information (Name and Origin, Protein attributes, General annotation and Entry information) of HAC1 was retrieved from the National centre of Biotechnology information (NCBI), in FASTA format, with the accession number NP_565197. Additional information were taken from the Arabidopsis Information Resource (TAIR), with the identification number AT1G79000. Secondary Structure Prediction of HAC1 protein was done by GOR IV server prediction server, based on the information theory (Garnier et al., 1996). There is no defined decision constant. GOR IV uses all possible pair frequencies within the window of 17 amino acid residues.

The protein sequence was subjected for comparative homology modelling via SWIS-MODEL Server (Schwede et al., 2003) and PHYRE2 server (Kelley LA and Sternberg MJE, 2009). SWISMODEL software assured a better control of the homology modelling process and promoted a deeper understanding of various features of the protein structure meant to be modeled.Phyre2 is a major update to the original Phyre server with a range of new features, accuracy is improved, using the alignment of hidden Markov models via HH search to significantly improve accuracy of alignment and detection rate (Jefferys et al., 2010).For additional structural and comparative analysis, ESyPred3D (Cristophe et al., 2002) and 3DJIGSAW (Bates et al., 2001) online servers predicted the 3D structure of HAC1.

Finally, once the 3D structures were generated, structural evaluation and stereo-chemical analysis was performed using different evaluation and validation tools. Backbone conformation of all models was evaluated by analysis of Psi/Phi Ramachandran plot using RAMPAGE program (Lovel et al., 2002). Energy calculations on a protein chain were performed by atomic empirical mean force potential ANOLEA (Melo et al., 1997). The program performs energy calculations on a protein chain, evaluating the "Non- Local Environment" (NLE) of each heavy atom in the molecule. A plot obtained from the Verify3D (Eisenberg, D., 1997) structure Evaluation Server that represented the average3D-1D profile score. Verify3D analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D).

3 Results and Discussion

The most successful general approach for predicting the structure of proteins involves the detection of homolog's of known three-dimensional (3D) structure—the so-called template-based homology modeling or

fold-recognition. Protein 3D structure is very important in understanding the protein functions, interactions and localizations (Denise Chen, 2009). In this study, for *Arabidopsis thaliana* histone acetyltransferase 1 protein , with 1697 amino acids, the best homolog found in all modeling programs, is the human p300 histone acetyltransferase (HAT), with protein data bank ID '3BIY'. The best homolog found in both modelers, Phyre2and SWISS modeler, with 45.8% sequence similarity, is a human p300 histone acetyltransferase (HAT). HAT1 regulates gene expression by acetylating histones and other transcription factors (Xu et al., 2001).

The secondary structural analysis of the protein was done and random coil was found to be most frequent (54.63%), followed by alpha helix (28.87%). Extended strand (Ee) was found to be least frequent (16.50%) (Data not shown).The dominance of the coiled regions indicates the high level of conservation and stability of the protein structure (Neelamathi et al., 2009).The predicted models were visualized by PyMol software, in the mode of cartoon view, with the lines removed.

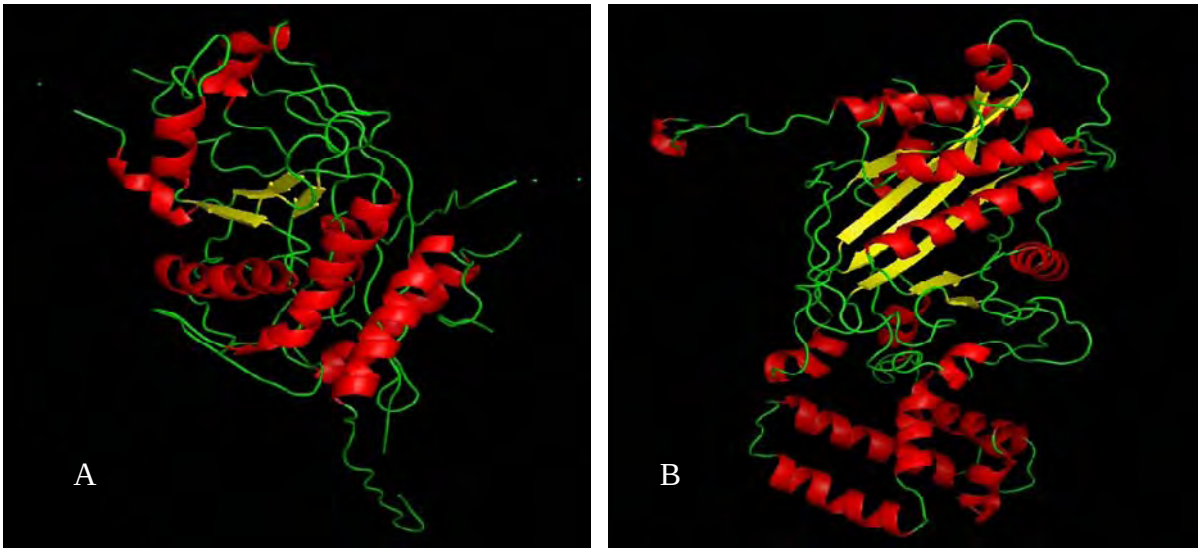

Fig. 1 Predicted models of HAC1 protein by (A)ESyPred3D and (B)3D-JIGSAW servers.

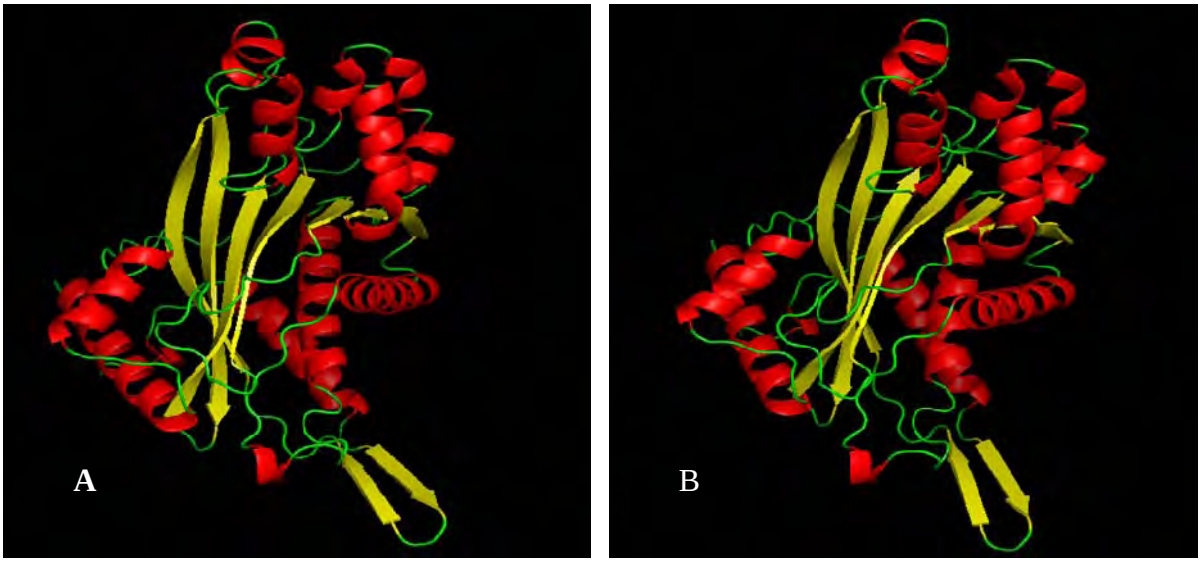

Fig. 2 Predicted models of HAC1 protein by (A) Phyre2and (B) SWISS-MODEL servers.

Benchmark of 4 different homology modeling programs: 3DJIGSAW, Swiss-Model, EasyPred3D and Phyre2, has been performed used to transform the alignment to a 3D model. The difference between the programs is how the information contained in the alignment is used to build a 3D model.

One of the most frequently used modelling programs; SWISS-MODEL generates the 3D models from a small number of rigid bodies, obtained from the core of the aligned regions (Blundell et al., 1987). The final three-dimensional (3D) structure was then confirmed by the Phyre2 server, an automatic fold recognition server for predicting the structure and the function of proteins based on homology modeling (Lawrence, 2011).

Errors of the model are usually estimated either from the energy of the model or from theresemblance of a given characteristic of the model to real structures (Sippl, 1995). Predicted models of HAC1 (Figs. 1-4) are evaluated by ANOLEA and Verify3D tools (Table 1).

Table 1 Model evaluation of proteins by using ANOLEA and Verfy3D tools.

Protein	Modeling program	Evaluation program	High energyzone (%)	Z-score	3D-1D profile
AtHAC1	ESyPred3D	ANOLEA	88.27%	18.04	-
		Verify3D	-	-	0.68
	3DJIGSAW	ANOLEA	51.67%	6.15	-
		Verify3D	-	-	0.72
	SWISS MODEL	ANOLEA	36.95%	3.51	-
		Verify3D	-	-	0.71
	PHYRE2	ANOLEA	36.95%	3.51	-
		Verfy3D	-	-	0.71

Verify3D method assesses protein structures using three-dimensional profiles. This program analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D). Each residue is assigned a structural class based on its location and environment (alpha, beta, loop, polar, apolar etc).The scores ranges from -1 (bad score) to +1 (good score) (Eisenberg, 1997). ANOLEA program performs energy calculations on a protein chain, evaluating the "Non- Local Environment" (NLE) of each heavy atom in the molecule (Melo et al., 1997).ANOLEA results (Tab.1) are shown as High energy (%) and pseudo Z-score. High energy zone (HEz) in the protein profile correlates with errors in the structure. Z-score is represents as pseudo energies of target protein sequences, where lower Z-score means higher reliability and vice versa.

Energy calculations performed by ANOLEA, gave the Z-score of 3.51 in 3D models generated by Phyre 2 and SWISS-MODEL server, indicating same quality of the models. While the HAC1 models obtained from EasyPred3D and 3DJIGSAW gave 18.04 and 6.15 Z-scores, two to six fold higher scores compared to SWISS Models and Phyre2 servers. The high Z-score obtained from EasyPred3D server is caused by the specific alignment method, obtained by combining, weighting and screening the results of several multiple alignment programs. Therefore the final three-dimensional structures is built using the modeling package MODELLER, which is not done in this study. 3D-1D profile scores, obtained from Verify3D, of HAC1 models is 0.71 obtained from Phyre 2 and SWISS-MODEL servers. EasyPred3D and 3DJIGSAW homology modeler resulted in 0.72 and 0.68 respectively.

The 3D structures generated in SWISS-MODEL and Phyre2 servers are identical, having the same number

of α helixes and β sheets. Z-score of SWISS-MODEL and Phyre2 modeler is much smaller compared to EasyPred3D and 3DJIGSAW server. Furthermore, the 3D-1D score of 0.72 (maximum is 1) indicates a good quality 3D model, compared to the human homolog model. Ramachandran plots analysis, generated by the RAMPAGE program (Lovel et al., 2002) for all four 3D models, shows all residues in the most favorable regions and disallowed regions (Fig. 3).

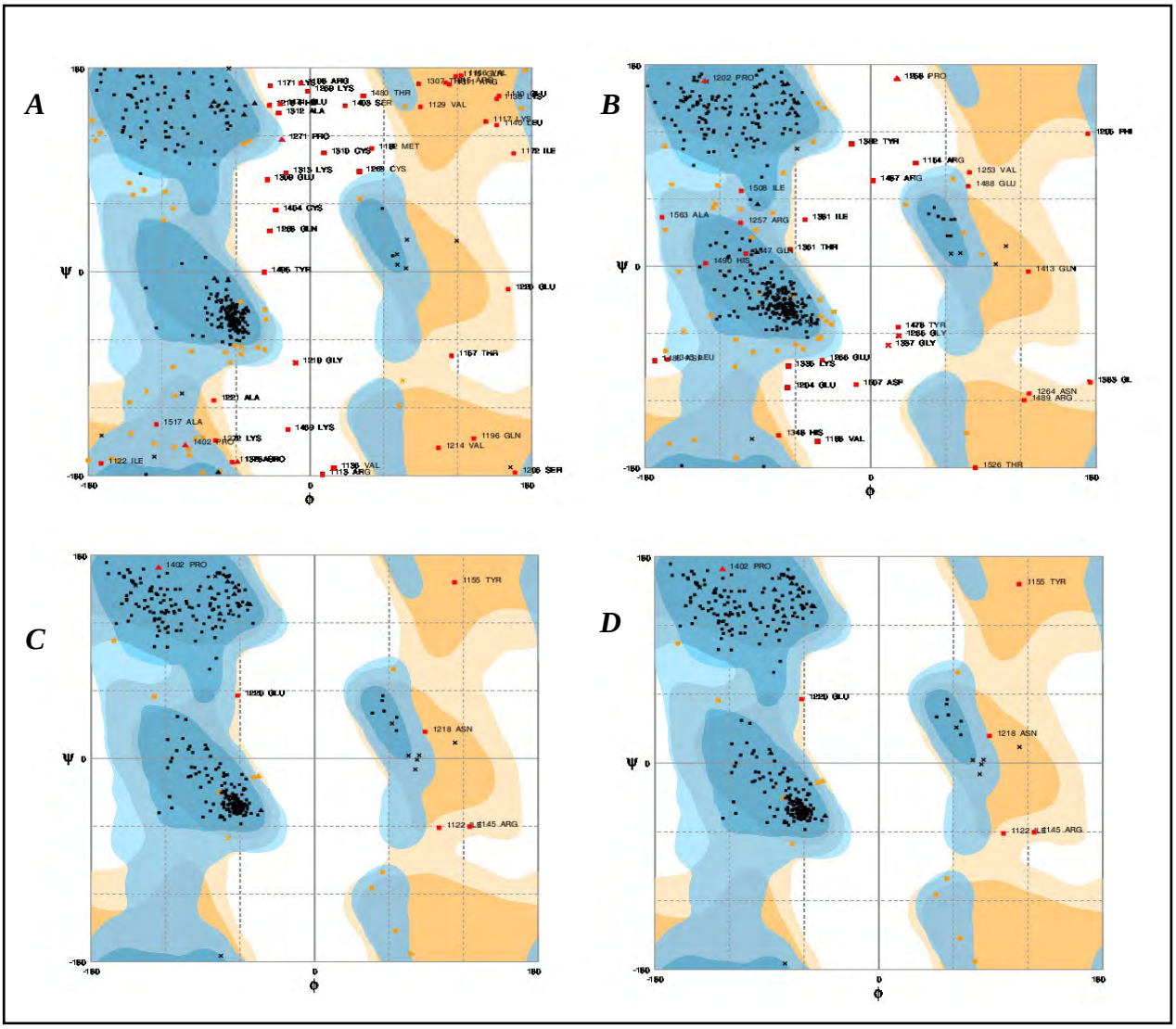

Fig. 3 ***A.*** Ramachandran plot of the ESyPred3D HAC1 model; ***B.*** Ramachandran plot obtained from the3D-JIGSAWserver; ***C.*** Ramachandran plot from the SWISS-MODEL server; ***D.*** Ramachadran plot of the PHYRE2HAC1 model.

The number of residues in favored region in the ESyPred3D 3D model of HAC1 is 73.0% (Fig. 3A). Ramachandran plot analysis for 3DJIGSAW HAC1 model (Fig. 3B), resulted in 83.9% of all residues placed in favored regions. These two results confirm the model check analysis of Verify3D and ANOLEA, indicating that 3DJIGSAW and ESyPred3D servers predicted the models with insufficient residues in the expected favored region. However, SWISS-MODEL and PHYRE2 servers were capable to produce a good 3D structure of *Arabidopsis thaliana* HAC1 protein, having 95% of all residues located in expected regions in Ramachandran plot analysis (Fig. 3 C and D).

Although the high similarity of 45.8 with thep300/CBP HAT homolog from humans , with a well predicted X-ray crystallography structure and known function, it is difficult to attribute some the functional characteristics to our predicted AtHAC1(Arabidopsis histone acetyltransferase of the CBP family) just according the structural similarities, that are stated in this paper.

The p300/CBP HAT protein from humans, with the sequence length of 2414 amino acids, is a bigger protein than HAC1 in Arabidopsis thaliana, with 1697 amino acids.

For future research, structural prediction and conformations of other *Arabidopsis thaliana* members p300/CBP histone acetyltransferase proteins, this HAC1 structural analysis could be used to find out the other proteins that not yet have been identified in this pathway. Furthermore, interactome analysis of all HAC member of the CBP family is recommended in order to assign new functions to this family.

4 Conclusion

An attempt was made to predict the three-dimensional structure of Arabidopsis histone acetyltransferase 1 of the CBP family. The structural analysis approach was based on homology modeling, using SWISS-MODEL, Phyre2, ESyPred3D and 3DJIGSAW model servers. For 3D model validations, Verify3D and ANOLEA tools have been used. Additionally, Ramachandran plot analysis was performed in order to verify whether all residues of our predicted models lye in the most favorable regions. SWISS MODEL and Phyre2 servers computed the identical 3D structure. Validation and verification methods, using Z-score and 3D-1D score, showed that these 3D models are of good and acceptable quality.

References

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, 408(6814): 796-815

Arnold K, Bordoli L, Kopp J, et al. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22: 195-201

Bates PA, Kelley LA, MacCallum RM, et al. 2001. Enhancement of Protein Modelling by Human Intervention in Applying the Automatic Programs 3D-JIGSAW and 3D-PSSM. Proteins: Structure, Function and Genetics, Suppl(5): 39-46

Blundell TL, Sibanda BL, Sternberg MJ, et al. 1987. Knowledge-based prediction of protein structures and the design of novel molecules. Nature, 326: 347-352

Chen D. 2009. Structural GENOMICS: Exploring the 3D Protein Landscape. Simbios, the NIH National Center for Physics-Based Simulation of Biological Structures,11-18

Deng W, Liu C, Pei Y, et al. 2007. Involvement of the histone acetyltransferase AtHAC1 in the regulation of flowering time via repression of FLOWERING LOCUS C in Arabidopsis. Plant Physiology, 143: 1660-1668

Earley KW, Shook MS, Brower-Toland B, et al. 2007. In vitro specificities of Arabidopsis co-activator histone acetyltransferases: implications for histone hyperacetylation in gene activation. The Plant Journal, 52: 615-626

Eisenberg D, Luthy R, et al. 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods in Enzymology, 277: 396-404

Garnier J, Gibrat JF, Robson B. 1996. GOR secondary structure prediction method version IV. Methods in Enzymology, 266: 540-553

Greer J. 1990. Comparative Modeling Methods: Application to the Family of the Mammalian Serine Proteases.

Proteins, 7: 317-334

Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the web: a case study using the Phyre server. Nature Protocols, 4: 363-371

Kelley LA, Bennett-Lovsey R, Herbert A, et al. 2011. Phyre: Protein Homology/analogY Recognition Engine. Structural Bioinformatics Group, Imperial College, London, UK

Lovell SC, Davis IW, Arendall III WB, et al. 2002. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function and Genetics, 50: 437-450

Melo F, Feytmans E. 1998. Assessing protein structures with a non-local atomic interaction energy. Journal of Molecular Biology, 2775: 1141-1152

Neelamathi E, Vasumathi E, Bagyalakshmi S, et al. 2009. Insilico prediction of structure and functional aspects of a hypothetical protein of Neurosporacrassa. Journal of Cell and Tissue Research, 93: 1889-1894

Pandey R, Müller A, Napoli CA. 2002. Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. Nucleic Acids Research, 30(23): 5036-5055

Raghava GPS. 2002. APSSP2 : A combination method for protein secondary structure prediction based on neural network and example based learning. CASP5. A-132

Sippl MJ. 1995. Knowledge based potentials for proteins. Current Opinion in Structural Biology, 5: 229-235

Söding J, Biegert, A, Lupas AN. 2005 TheHHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research, 33(Web Server issue): W244-8

Sterner DE, Berger SL. 2000. Acetylation of Histones and Transcription-Related Factors. Microbiology and Molecular Biology Reviews, 64(2): 435

The PyMOL Molecular Graphics System, Version 1.3. Schrödinger, LLC

Xu W, Chen H, Du K, et al. 2001. A transcriptional switch mediated by cofactor methylation. Science, 294: 2507-2511