Article

3D structure prediction of histone acetyltransferase (HAC) proteins of the p300/CBP family and their interactome in *Arabidopsis thaliana*

Amar Ćemanović[1], **Jasmin Šutković**[1], **Rabab Elamawi**[2], **Waleed Elkhoby**[2], **Mohamed Ragab Abdel Gawwad**[1]

[1]Genetics and Bioengineering department, International University of Sarajevo, Ilidza, 71220 Bosnia and Herzegovina

[2]Rice Research and Training Center, Sakha, Egypt

E-mail: mragab@ius.edu.ba,mrgawad@hotmail.com

Received 29 October 2013; *Accepted 5 December 2013*; *Published online 1 September 2014*

Abstract

Histone acetylation is an important posttranslational modification correlated with gene activation. In *Arabidopsis thaliana* the histone acetyltransferase (HAC) proteins of the CBP family are homologous to animal p300/CREB (cAMP-responsive element-binding proteins, which are important histone acetyltransferases participating in many physiological processes, including proliferation, differentiation, and apoptosis. In this study the 3-D structure of all HAC protein subunits in *Arabidopsis thaliana:* HAC1, HAC2, HAC4, HAC5 and HAC12 is predicted by homology modeling and confirmed by Ramachandran plot analysis. The amino acid sequences HAC family members are highly similar to the sequences of the homologous human p300/CREB protein. Conservation of p300/CBP domains among the HAC proteins was examined further by sequence alignment and pattern search. The domains of p300/CBP required for the HAC function, such as PHD, TAZ and ZZ domains, are conserved in all HAC proteins. Interactome analysis revealed that HAC1, HAC5 and HAC12 proteins interact with S-adenosylmethionine-dependent methyltransferase domain-containing protein that shows methyltransferase activity, suggesting an additional function of the HAC proteins. Additionally, HAC5 has a strong interaction value for the putative c-myb-like transcription factor MYB3R-4, which suggests that it also may have a function in regulation of DNA replication.

Keywords *Arabidopsis thaliana*; histone acetyltransferase; 3D structure; interactome; functional annotation.

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

In order for the whole genome to be packed into the nucleus of a eukaryotic cell, the DNA is associated with special proteins, called histones, to form chromatin fibers. Histones are small proteins that function in DNA packaging and gene regulation (Hardin et al., 2012). The region of DNA that is part of the core particle is inaccessible to transcription factors and therefore transcriptionally inactive. Various studies demonstrate the

direct association between modification of chromatin structure and levels of gene expression (Allfrey et al., 1964; Katsani et al., 2003; Martinowich et al., 2003; McGraw et al., 2007; Kishimoto et al., 2006; Hublitz et al., 2009). These modifications can be in the form of methylation, ubiquitination, phosphorylation, ADP-ribosylation and acetylation. Of those mechanisms, the best known is acetylation – modification by the action of histone acetyltransferase (HAT) enzymes (Sterner and Berger, 2000).

HAT proteins function by transferring acetyl groups from acetyl-CoA to lysine residues on the N-terminal tails of histones. Thereby, they lower the affinity of histones for DNA and, accordingly, make the DNA more accessible to the transcription machinery. The process is reversible, with the deacetylation being performed by deacetylase enzymes. Based on sequence homology, histone acetyltransferases have been organized into four families: GNAT, MYST, p300/CBP and TAFII250. The p300/CBP family is especially important since it acetylates not only histones, but also various other proteins, such as the cAMP response element-binding (CREB) protein, p53, HIV-1 Tat protein and Stat3. Thus, the proteins of this family are also termed factor acetyltransferases (Loidl, 1994).

The p300 / CBP family encompasses the p300 and CBP proteins in animals and homologous proteins in other eukaryotes. Due to their high homology and similarity in structure and function, p300 and CBP have until recently been treated as one protein. Since it suits the purpose, this paper does likewise. p300 / CBP is a large, universally expressed, regulatory protein found in many multicellular organisms. It is involved in a wide variety of cellular processes, including differentiation, cell cycle control and apoptosis (Loidl, 1994). In addition, it is shown to participate in tumor-suppressor pathways, cell transformation and action of cellular oncogenes, associated with viral oncoproteins, contribute to the differentiation of specific cell lineages (processes of hematopoiesis and myogenesis), play a role in development and participate in transcriptional repression pathways (Goodman and Smolik, 2000). It is also suggested that p300/CBP is brought in contact with specific promoters by physical interactions with sequence-specific transcription factors (Vandel and Trouche, 2001).

On the molecular level, besides CREB, p300/CBP interacts with numerous promoter-binding trascription factors including nuclear hormone receptors and activators such as c-Jun, c-Fos, c-Myb etc (Sterner and Berger, 2000). Defects in its expression have been linked to a large number of developmental abnormalities, such as Rubinstein-Taybi syndrome (RTS) in humans (Goodman and Smolik, 2000). Additionaly, Vandel and Trouche (2001) have shown that p300/CBP physically associates with a histone methyltransferase enzyme.

Many conserved domains are found in the p300/CBP protein, including the bromodomain, CREB-binding domain (KIX), N-terminal nuclear receptor-interacting domain (RID),glutamine/proline-rich domain (QP), three cysteine/histidine-rich regions (CH1, CH2 and CH3) and a bipartite nulear localization signal (NLS-BP). The CH regions contain four zinc finger motifs: TAZ1, TAZ2, ZZ and PHD (Yuan and Giordano, 2002). The domains RID, KIX, TAZ1 and TAZ2 function in binding to transcriptional activators and regulators (Goodman and Smolik, 2000). The HAT function is performed by a large conserved region that spans from the PHD domain to the ZZ motif,with a putative CoA-binding domain being located in the middle (Ogryzko et al., 1996).

Although the human p300/CBP is well characterized and has its crystal structureexperimentally determined (Liu et al., 2008), little is known about its orthologs in Arabidopsis thaliana. The p300/CBP family in Arabidopsis includes the proteins: HAC1, HAC2, HAC4, HAC5 and HAC12 14, all displaying a significant degree of homology to the p300/CBP protein (Deng et al., 2007). The members of this family have previously been called PCAT proteins (p300/CBP acetyltransferase-related proteins) (Bordoli, 2001).

All five members of the family function in the nucleus, with HAC4 additionally being found in the chloroplast. Out of them, only HAC2 has been shown to lack direct HAT activity (Bordoli, 2001). HAC1,

HAC2 and HAC12 also display transcription cofactor activity, while all members of the family exhibit zinc ion binding activity. Furthermore, HAC1, HAC5 and HAC12 have been shown to be involved in the process of flower development (The UniProt Consortium, 2012).

All HAC proteins contain ZZ and TAZ zinc finger domains, implicated in interactions with transcription factors, and a C-terminal HAT domain that provides HAT activity in vitro. As compared to the p300/CBP protein, the HAC proteins differ in four main aspects: 1. they lack a bromo-domain; 2. they lack a KIX domain; 3. they have two ZZ zinc finger domains, whereas p300/CBP has only one and; 4. they lack the QP region near the C-terminus (Deng et al., 2007).

This in silico research is aimed to attribute additional functions to the HAC proteins of the p300/CBP protein family in *Arabidopsis thaliana* by interactome analysis. In addition, the phylogenetic relationship of the proteins will be assessed using multiple sequence alignment and phylogenetic tree construction, and the subcellular localization of each HAC protein determined. Finally, the three-dimensional (3D) structures will be predicted and confirmed with Ramachandran plot analysis and several validation tools available at the Swiss Model server.

2 Materials and Methods

The sequences for the HAC proteins of the p300/CBP family (HAC1, HAC2, HAC4, HAC5 and HAC12) were obtained from National Center for Biotechnology Information (NCBI) database (Sayers, 2009). Additionally, TAIR (The Arabidopsis Information Resource) ID numbers were obtained (Lamesch et al., 2012). The NCBI and TAIR accession numbers are shown in Table 1.

Table 1 HAC proteins of the p300/CBP family and their Ids.

HAC protein	Gene ID code	TAIR IDs
HAC1	NP_565197.3	At1g79000
HAC2	NP_564891.4	At1g67220
HAC4	NP_564706.1	At1g55970
HAC5	NP_187904.1	At3g12980
HAC12	NP_173115.1	At1g16710

Multiple sequence alignment (MSA) has been performed using the ClustalW software (version 2) located on the website of the European Bioinformatics Institute (EBI), using default options (Larkin et al., 2002; Goujon, 2010). MSA is an invaluable bioinformatics tool used to measure the similarity between sequences, examine patterns of conservation and variability and derive evolutionary relationships (Lesk, 2002). ClustalW uses a progressive method of alignment, meaning that it aligns the sequences one by one, instead of aligning all at once (Claverie and Notredame, 2007).

In order to infer the evolutionary relationship between the HAC proteins, a phylogenetic tree was constructed using Phylogeny.fr, a web service for phylogenetic analysis of molecular sequences (Dereeper et al., 2008). The service was run on default settings, and the steps that it performed to construct the phylogenetic tree involved multiple sequence alignment, alignment organization and construction and visualization of the phylogenetic tree using different integrated tools (Castresana, 2000; Chevenet et al., 2003; Guindon and Gascuel, 2003; Edgar, 2004; Anisimova and Gascuel, 2006; Dereeper, 2010).

As there exists no experimental published data on the 3D structure of the HAC proteins in Arabidopsis in

the Protein Data Bank (Bernstein, 1977), the structures were hereby predicted with the help of the Phyre2 protein homology modeling server (Kelley and Sternberg, 2009). Phyre2 is a web-based service for protein structure prediction that is free for non-commercial use. Phyre2 has been designed and funded by the Biotechnology and Biological Sciences Research Council (BBSRC) from United Kingdom. Recently, a comparative structural study for the Arabidopsis HAC1 protein was performed to predict the most accurate structure of HAC1, where the HAC1 structural modeling by the Phyre2 server proved to be the most accurate, along with the Swiss Model homology modeling server (Cemanovic et al., 2014). A practical and widely cited molecular visualization tool, PyMOL, was used for structure visualization and representation (PyMOL).

After the generation of the 3D structures, structural evaluation and stereochemical analysis was performed using different evaluation and validation tools. Backbone conformation of all models and their stereo-chemical quality was evaluated by analysis of Ramachandran plots using the PROCHECK software (Laskowski, 1993).

Furthermore, 3D structure validation for all protein members was performed by the Swiss Model server with several incorporated validation tools (Schwede, 2003). The Swiss Model server includes a Swiss model workspace, an automated comparative protein modeling environment that incorporates several structural validation tools for model quality estimation. One of these tools is QMEAN6, a composite scoring function that is used to estimate the quality of a 3D protein model and to compare and rank alternative models of a target. The scores are given in values between 0 and 1, with higher values for better models (Benkert et al., 2008).

Another tool integrated into the Swiss Model Server is Verify3D, which assesses protein structures using three-dimensional profiles. It analyzes the compatibility of a 3D model with its own amino acid sequence (1D). The scores range from -1 (bad score) to +1 (good score) (Eisenberg et al., 1997).

The identification of domains in the five HAC proteins was performed using the online tool SMART (Simple Modular Architecture Research Tool) located on the website of the European Molecular Biology Laboratory (EMBL) (Schultz, 1998; Letunic et al., 2011).

The interactome of the HAC proteins was obtained with help of the Arabidopsis Interaction Viewer, which uses a database of 70944 predicted and 28556 confirmed Arabidopsis interacting proteins (Geisler-Lee, 2007).

The subcellular localization of each HAC protein was determined using WoLF PSORT, an online tool which functions by detecting known sorting signals and specific amino acid content in proteins. Experiments showed that the overall accuracy of WoLF PSORT is over 80% (Horton et al., 2007).

3 Results

3.1 Predicted 3D structure models and Ramachandran plot confirmations

For additional 3D structure validation and confirmation, statistical analysis with QMEAN6 and Verify3D software was performed. The results are shown in Table 2 and include also the Ramachandran plot PROCHECK results (Fig. 1-2).

3.2 Multiple sequence alignment and phylogenetic tree construction

The amino acid sequences of all HAC proteins is obtained from the NCBI database, analyzed in the ClustalW online tool for multiple sequence alignment. The obtained results are summarized in Table 3. The scores presented in the table are a measure of the similarity between the respective sequences, as determined by their pairwise alignment. A higher score represents a higher similarity.

The phylogenetic tree created for the HAC proteins shows that HAC1 and HAC12 as well as HAC4 and HAC5 are evolutionary close relatives. HAC2 is phylogenetically the most distant member of the family, diverging very early in the evolution of the HAC proteins (Fig. 3).

Table 2 QMEAN6, Verify3D and PROCHECK structure validation results.

HAC PROTEIN MEMBERS	VALIDATION TOOLS		
	QMEAN6 (0-1)	Verify3D	PROCHECK– favored region in %
HAC1	0.653	0.72	91.2
HAC2	0.630	0.69	91.4
HAC4	0.608	0.77	92.1
HAC5	0.649	0.82	91.2
HAC12	0.642	0.69	90.6

Table 3 ClustalW multiple sequence alignment scores for HAC proteins.

Seq. 1	Name	Length	Seq. 2	Name	Length	Score
1	HAC1	1697	2	HAC2	1367	26
1	HAC1	1697	3	HAC4	1456	44
1	HAC1	1697	4	HAC5	1670	50
1	HAC1	1697	5	HAC12	1706	75
2	HAC2	1367	3	HAC4	1456	24
2	HAC2	1367	4	HAC5	1670	24
2	HAC2	1367	5	HAC12	1706	26
3	HAC4	1456	4	HAC5	1670	53
3	HAC4	1456	5	HAC12	1706	44
4	HAC5	1670	5	HAC12	1706	48

3.3 Domain search

The domain search using SMART showed a high degree of similarity in domain presence and location among the HAC proteins, similar to what has been shown by recent research.

All the HAC protein members contain the following domains: zinc finger TAZ (transcriptional adaptor zinc-binding) domain, consisting of a distinct fold and is characteristic for proteins of the CPB/P300 family. In addition, HAC proteins in Arabidopsis have zinc finger ZZ domains, named so because they are able to bind two zinc ions, and a PHD (plant homeodomain) domain. All the domain locations are shown in Table 4.

3.4 Subcellular localization

The analysis of the subcellular localization using the WoLF PSORT tool showed that all HAC proteins are localized in the nucleus, which is also confirmed by TAIR. HAC2, HAC4, HAC5 and HAC12 were predicted to be localized in the chloroplast as well, whereas HAC1 was predicted also in the cytoplasm. Additionally, HAC2 and HAC4 were predicted in the mitochondria and plastid, respectively, although with low accuracy. The results are summarized in Table 5.

3.5 Interactome analysis

The results obtained using the Arabidopsis Interactions Viewer showed that HAC1, HAC5 and HAC12 interact with various proteins which are localized mainly in the nucleus. However, it provided no interactome for HAC2 and HAC4.

Table 4 Position of domains in HAC proteins.

PROTEINS	DOMAINS									
	TAZ 1		TAZ2		ZZ		ZZ 2		PHD	
	Start	End	Start	End	Start	End	Start	End	Start	End
HAC1	631	709	1582	1660	1398	1442	1518	1564	991	1064
HAC2	-	-	1275	1357	1093	1139	121	1265	704	763
HAC4	417	495	1345	1420	1160	1204	1280	1325	792	839
HAC5	612	690	1557	1632	1378	1421	1498	1542	972	1044
HAC12	638	716	1591	1669	1407	1451	1407	1451	1000	1073

Table 5 Results of localization analysis for HAC proteins.

Protein	Subcellular localization	Prediction accuracy
HAC1	Nucleus	12.5
	Cytoplasm	7.0
HAC2	Nucleus	9.0
	Chloroplast	3.0
	Mitochondria	1.0
HAC4	Chloroplast	6.0
	Nucleus	6.0
	Plastid	1.0
HAC5	Nucleus	7.0
	Chloroplast	6.0
HAC12	Nucleus	10.0
	Chloroplast	4.0

Common interactome partners for HAC1, HAC5 and HAC12 are ‘smad nuclear-interacting protein 1’ (At3g20550), ‘S-adenosylmethionine-dependent methyltransferase domain-containing protein’ (At1g45231) and ‘transcription regulator/ zinc ion binding protein’ (At3g47610). Interactome partners common just to HAC1 and HAC5 include ‘N-terminal asparagine amidohydrolase family protein’ (At2g44420) and ‘DNA repair protein RAD51-like 1’ (At5g20850). HAC5 and HAC12 show interaction with the ‘DNA-directed RNA polymerase II subunit RPB1’ (At4g35800), whereas HAC1 and HAC12 interacts with the ‘abscisic acid-insensitive 5-like protein 5 (At1g45249). In addition, HAC5 shows strong interaction with both ‘histone acetyltransferase GCN5’ (At3g54610) and ‘putative c-myb-like transcription factor 3r-4’ (At5g11510). The results are summarized in Fig. 4 and Table 5. Proteins that interact with more than one HAC protein are written in the same color for easier recognition.

Table 6 Interactome of HAC proteins showing only significant interactions.

HAC proteins	Interactome	Interolog Confidence Value	Interolog Confidence	Interactome Location	Interactome function
HAC1	At1g10390	4	Medium	nuclear membrane, nucleus	Nucleoporin autopeptidase
	At1g45231	4	Medium	nucleus, cytoplasm	S-adenosylmethionine-dependent methyltransferase domain-containing protein
	At1g45249	4	Medium	nucleus	Abscisic acid-insensitive 5-like protein 5
	At3g20550	4	Medium	chloroplast, cytoplasm, nucleus	Smad nuclear-interacting protein 1
	At3g47610	9	Medium	nucleus	Transcription regulator/ zinc ion binding protein
	At2g44420	2	Medium	chloroplast, nucleus, extracellular	N-terminal asparagine amidohydrolase family protein
	At5g20850	2	Medium	nucleus	DNA repair protein RAD51-like 1
	At5g26680	4	Medium	nucleus, cytoplasm	Flap endonuclease-1
HAC5	At1g45231	8	Medium	nucleus, cytoplasm	S-adenosylmethionine-dependent methyltransferase domain-containing protein
	At3g20550	1	Low	chloroplast, cytoplasm, nucleus	Smad nuclear-interacting protein 1
	At3g47610	1	Low	nucleus	Transcription regulator/ zinc ion binding protein
	At4g35800	4	Medium	chloroplast, nucleus, plasmodesma, vacuole	DNA-directed RNA polymerase II subunit RPB1
	At2g44420	1	Low	chloroplast, nucleus, extracellular	N-terminal asparagine amidohydrolase family protein
	At5g20850	1	Low	nucleus	DNA repair protein RAD51-like 1
	At3g54610	12	High	nucleus	Histone acetyltransferase GCN5
	At5g11510	96	High	nucleus	AtMYB3R4_putative c-myb-like transcription factor 3r-4
HAC12	At1g10390	4	Medium	nuclear membrane, nucleus	Nucleoporin autopeptidase
	At1g45231	4	Medium	nucleus, cytoplasm	S-adenosylmethionine-dependent methyltransferase domain-containing protein
	At1g45249	1	Low	nucleus	Abscisic acid-insensitive 5-like protein 5
	At3g20550	4	Medium	chloroplast, cytoplasm, nucleus	Smad nuclear-interacting protein 1
	At3g47610	4	Medium	nucleus	Transcription regulator/ zinc ion binding protein
	At4g35800	4	Medium	chloroplast, nucleus, plasmodesma, vacuole	DNA-directed RNA polymerase II subunit RPB1

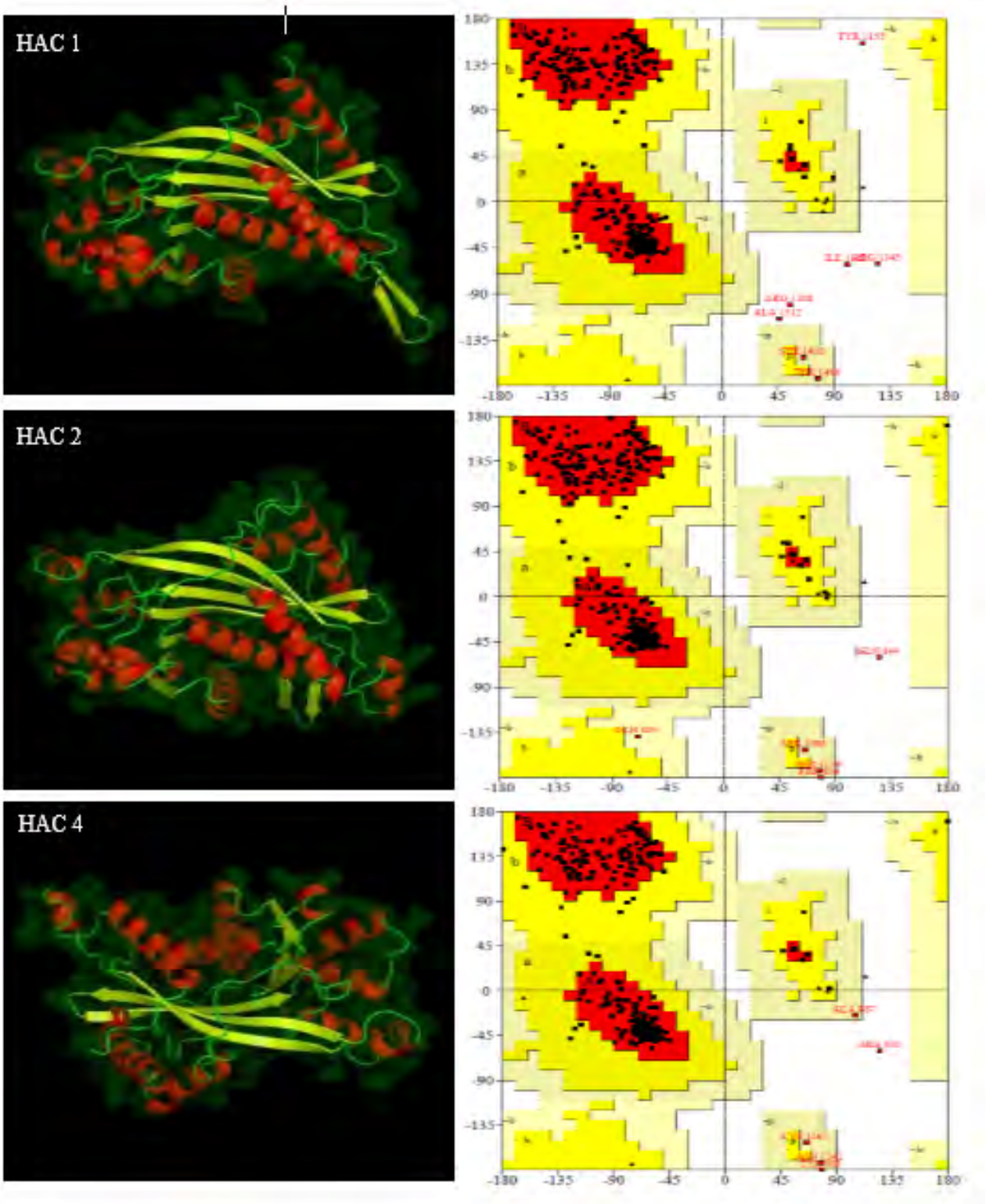

Fig. 1 Predicted 3D structures of HAC1, HAC2 and HAC4 proteins and the respective Ramachandran plots.

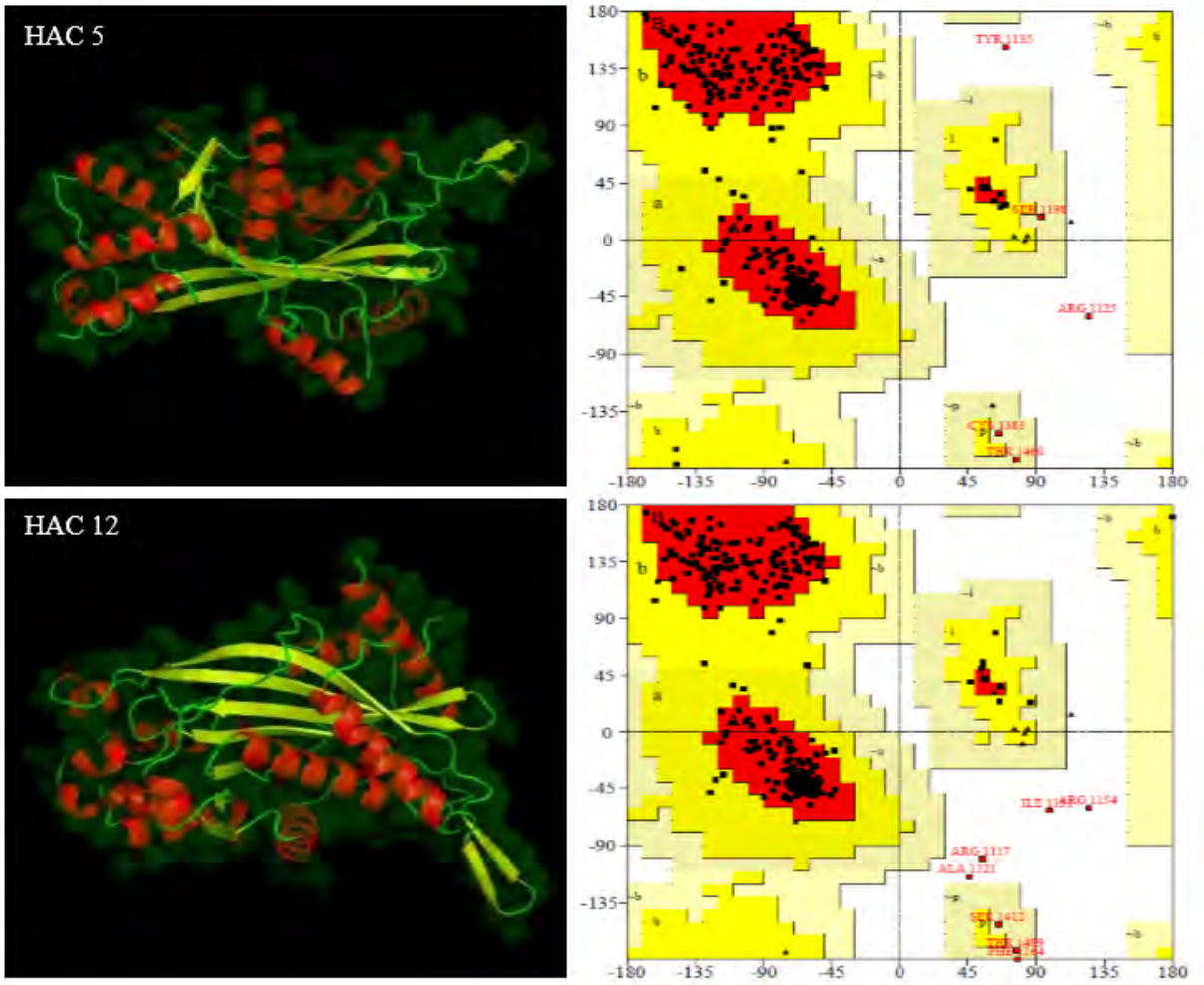

Fig. 2 Predicted 3D structures of HAC5 and HAC12 proteins and the respective Ramachandran plots.

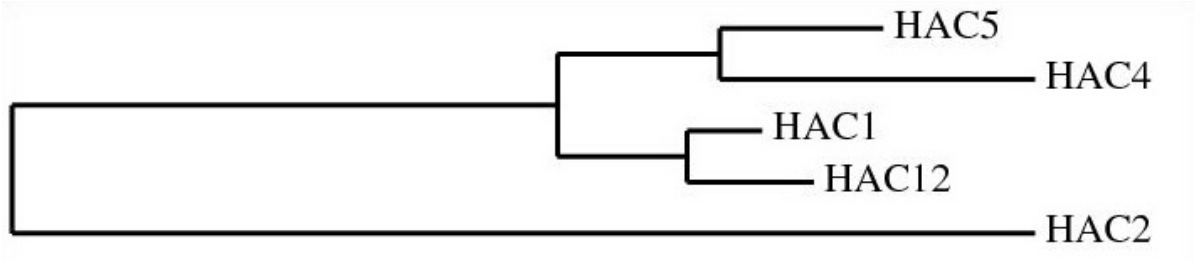

Fig. 3 Phylogenetic tree of the HAC proteins.

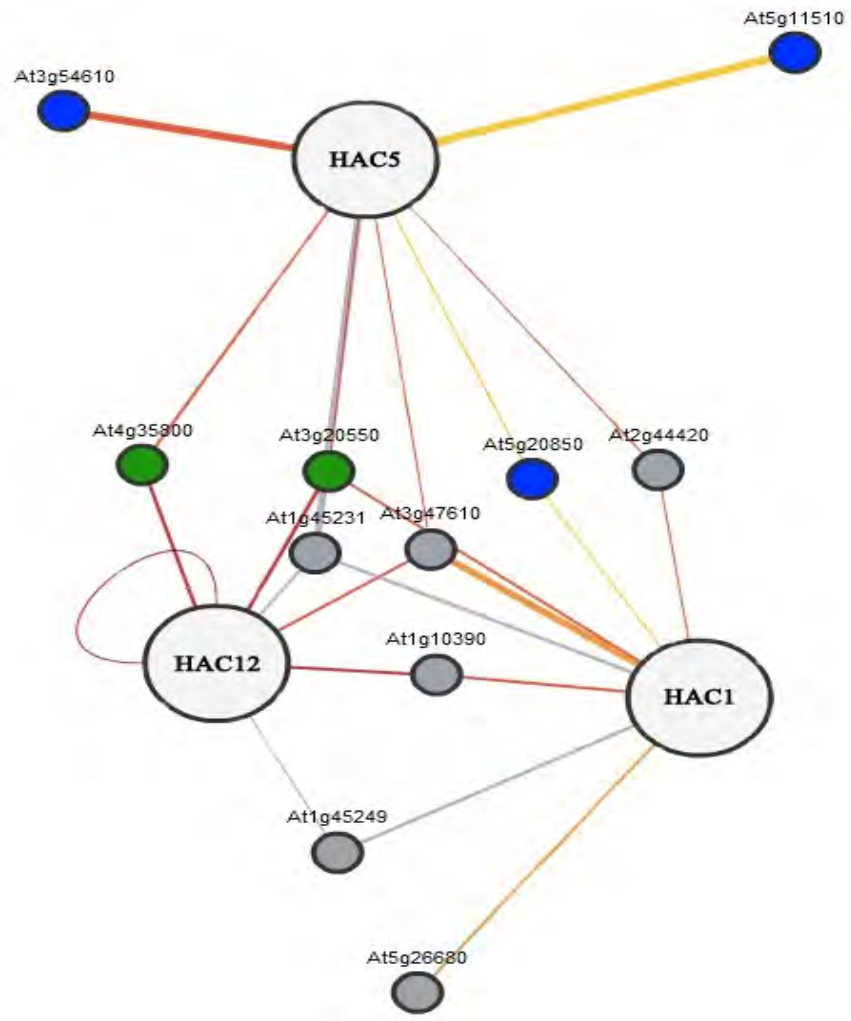

Fig. 4 Interactome of HAC proteins (only significant interaction are shown).

4 Discussion and Conclusion

The five members of the histone CBP acetyltransferase (HAC) protein family in *Arabidopsis thaliana* (HAC1, HAC2, HAC4, HAC5 and HAC12) are plant homologs of the human p300/CBP, involved in transcription regulation, cell-cycle control and differentiation . Its primary function, histone acetylation, is an important posttranslational modification correlated with gene activation. Histone acetylation modifies the protein and lowers the affinity of the respective histone for the DNA molecule, causing the DNA to render and to be accessible to polymerase and transcription factors (Sterner and Berger, 2000). The involvement of p300/CBP and other histone acetyltransferases in processes other than histone acetylation is mostly associated with their interaction with other proteins in the cell. This is important since it allows the assignment of new functions to proteins by analyzing their interactome.

The multiple sequence alignment scores and the phylogenetic tree revealed the evolutionary relationship among HAC homologs. The closest ancestors are HAC1 and HAC12, with an alignment score of 75. This implies that HAC1 and HAC12 have the most recent common ancestor of all HAC proteins. Furthermore, it is noticeable that HAC2 diverged earliest, also reflected in the alignment scores, being HAC2 as the least similar with other HAC homologs. In addition, it can be seen that HAC4 and HAC5 have a relatively high similarity and that they have diverged from HAC1 and HAC12.

By the analysis of the three-dimensional structures of the HAC proteins by aforementioned validation methods, we can conclude that the Phyre2 server created 3D structure models of good quality, confirming the previous in silico structural analysis of Arabidopsis HAC1 (Cemanovic et al., 2014). Phyre2 is based on identifying and aligning remote homologous sequences and relies on profiles or hidden Markov models (HMMs). These profiles/HMMs contain information about the tendency for mutation of each amino acid in a sequence and are unique for each protein. They are created for a set of known 3D structures as well as for the user sequence, and then scanned to find a match (Baum, 1972). Further confirmation and verification of the modeled structure was tested by three validation methods. All the results related to these methods lie in the range of acceptance.

The interactome analysis has linked the HAC1 protein with flap endonuclease-1, which has DNA binding, catalytic and nuclease activity (Tabata, 2000). This may suggest an involvement of HAC1 in these processes, possibly by making the target region of DNA available for flap endonuclease-1 by acetylation of the associated histone proteins.

HAC1 and HAC12 are predicted to interact with nucleoporin autopeptidase, a transporter protein localized in the nuclear membrane (Tamura, 2010). It suggests that this protein transports the newly synthesized proteins into the nucleus where they perform their further functions.

HAC1, HAC5 and HAC12 were all shown to interact with S-adenosylmethionine-dependent methyltransferase domain-containing protein, a methyltransferase enzyme localized mainly in the nucleus. They were also shown to have another common interactome partner - Smad nuclear-interacting protein 1, which functions in RNA and protein binding, and is involved in the processes of DNA methylation, chromatin modification and transcription regulation (Machida and Yuan, 2013). This is in concordance with the findings of Vandel and Trouche (2001), who showed that immunoprecipitation of the human HAC homolog CBP/p300 leads to co-immunoprecipitation of histone methytransferase activity (HMT), and the study of Yang et al. (2009) who proved that CBP/p300 interacts with a methyltransferase enzyme (Protein arginine N-methyltransferase 5). This suggests a link between the processes of acetylation and methylation of histones and an involvement of histone methylation in transcription activation.

HAC1, HAC5 and HAC12 also interact with transcription regulator/ zinc ion binding protein, a transcriptional regulator found in the nucleus. This is in concordance with the fact that the HAC proteins are all involved in the process of transcription regulation. The common interactome protein partners of HAC1, HAC5 and HAC12 can be associated with their high similarity as shown in their alignment score, phylogenetic relationship analysis and their 3D structure similarities.

HAC1 and HAC5 were predicted to interact with N-terminal asparagine amidohydrolase family protein, a protein which also functions in N-terminal protein modification, but by means of myristoylation, i.e. attachment of myristoyl groups to N-termini of proteins (Lin et al., 1999). The two aforementioned HAC proteins also potentially interact with DNA repair protein RAD51-like 1, which is involved in processes of DNA repair, DNA methylation, chromatin modification, transcription- and cell-cycle regulation (Doutriaux et al., 1998). This may also be brought in connection with the findings of Vandel and Trouche (2001) about the potential interaction of acetyltransferases and methyltransferases. Apart from the involvement of RAD51, it is shown that radiation sensitive protein 1 (RAD1) and Replication factor C subunits (RFCs) play an important role in DNA repair and DNA replication processes in *Arabidopsis thaliana* (Abdel Gawwad et al., 2013).

HAC1 and HAC12 have also a possible common interactome partner - abscisic acid-insensitive 5-like protein 5. However, the confidence value for HAC12 is very low in this case, as compared to that of HAC1 (not shown). The interactome partner has DNA binding and protein binding function, and is involved in positive regulation of transcription, similar to the two HAC proteins (Theologis, 2000).

HAC5 and HAC12 were predicted to interact with DNA-directed RNA polymerase II subunit RPB1, the largest subunit of the polymerase II core complex. Besides its involvement in transcription, DNA-directed RNA polymerase II subunit RPB1 is also involved in the process of desumoylation (cleaving of a SUMO – small ubiquitin-related modifier – protein from its target protein) (Nawrath, 1990).

Finally, the strongest confidence values were obtained for the HAC5 interaction with histone acetyltransferase GCN5 and putative c-myb-like transcription factor 3r-4 (MYB3R-4). The former is an acetyltransferase enzyme involved in positive regulation of transcription as well as flower development, like the HAC proteins. In addition, it is also involved in protein ubiquitination and deubiquitination (Servet et al., 2010). This suggests an interaction between HAC5 and GCN5 as DNA modifying enzymes of different protein families at the nuclear level and in the same physiological processes. The latter interactome partner of HAC5, MYB3R-4, is a DNA binding protein involved in transcription regulation, but also in regulation of DNA replication (Haga et al., 2011). A similar interaction is found in humans as well, where the human proto-oncogene c-myb interacts with the CBP/p300 protein (Giordano and Avantaggiati, 1999; Sterner and Berger, 2000). Having a very high confidence value for interacting with MYB3R-4, we can annotate an additional function to HAC5 - regulation of DNA replication.

Conclusively, the analysis of the HAC proteins confirmed their homology and common evolutionary descent. The predicted 3D structures were tested and characterized as being of good quality, and the interactome analysis hinted a possible new functions for some of the HAC proteins.

References

Abdel Gawwad MR, Sutkovic J, Zahirovic E. 2013. 3D structure prediction of replication factor C subunits (RFC) and their interactome in *Arabidopsis thaliana*. Network Biology, 3(2): 74-86

Allfrey VG, Faulkner R, Mirsky AE. 1964. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proceedings of the National Academy of Sciences USA, 51: 786-794

Anisimova M, Gascuel O. 2006. Approximate likelihood ratio test for branchs: A fast, accurate and powerful alternative. Systems Biology, 55(4): 539-552

Baum LE. 1972. An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process. Inequalities, 3: 1-8

Benkert P, Tosatto SCE, Schomburg D. 2008. QMEAN: A comprehensive scoring function for model quality assessment. Proteins: Structure, Function, and Bioinformatics, 71(1): 261-277

Bernstein FC, Koetzle TF, Williams GJ. 1977. The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures. Journal of Molecular Biology, 112:535

Bordoli L, Netsch M, Lüthi U, 2001. Plant orthologs of p300/CBP: conservation of a core domain in metazoan p300/CBP acetyltransferase-related proteins. Nucleic Acids Research, 29: 589-597

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution, 17(4): 540-552

Cemanovic A, Sutkovic J, Abdel Gawwad MR. 2014. Comparative structural analysis of HAC1 in *Arabidopsis thaliana*. Network Biology, 4(2): 67-73

Chevenet F, Brun C, Banuls A. 2003. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics, 7: 439

Claverie JM, Notredame C. 2007. Bioinformatics for Dummies (Second Edition). Wiley Publishing Inc., Indianapolis, USA

Deng W, Liu C, Pei Y. 2007. Involvement of the Histone Acetyltransferase AtHAC1 inthe Regulation of

Flowering Time via Repression of FLOWERING LOCUS C in *Arabidopsis*. Plant Physiology, 143: 1660-1668

Dereeper A, Audic S, Claverie JM, Blanc G. 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evolutionary Biology, 12: 10-18

Dereeper A, Guignon V, Blanc G. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research, 36: 465-469

Doutriaux MP, Couteau F, Bergounioux C, White C. 1998. Isolation and characterisation of the RAD51 and DMC1 homologs from *Arabidopsis thaliana*. Molecular Genetics and Genomics, 257: 283-291

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5): 1792-1797

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

Geisler-Lee J, O'Toole N, Ammar R. 2007. A Predicted Interactome for *Arabidopsis*. Plant Physiology, 145: 317-329

Giordano A, Avantaggiati ML. 1999. p300 and CBP: partners for life and death. Journal of Cellular Physiology, 181(2): 218-230

Goodman RH, Smolik S. 2000. CBP/p300 in cell growth, transformation, and development. Genes Dev., 14: 1553-1577.

Goujon M, McWilliam H, Li W. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Research, 38: 695-699

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systems Biology, 52(5): 696-704

Haga N, Kobayashi K, Suzuki T. 2011. Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in *Arabidopsis*. Plant Physiology, 157(2): 706-17.

Hardin J, Bertoni G, Kleinsmith LJ. 2012. Becker's world of the cell, Eight edition. Benjamin Cummings, San Francisco, USA

Horton P, Park KJ, Obayashi T. 2007. WoLF PSORT: Protein Localization Predictor. Nucleic Acids Research, 1-3 (DOI 10.1093/nar/gkm259)

Hublitz P, Albert M, Peters A. 2009. Mechanisms of Transcriptional Repression by Histone Lysine Methylation. The International Journal of Developmental Biology, 10(1387): 335-354

Katsani KR, Mahmoudi T, Verrijzer CP. 2003. Selective gene regulation by SWI/SNF-related chromatin remodeling factors. Current Topics in Microbiology and Immunology, 274: 113-141

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

Kishimoto M, Fujiki R, Takezawa S. 2006. Nuclear receptor mediated gene regulation through chromatin remodeling and histone modifications. Endocrine Journal, 53(2): 157-172

Lamesch P, Berardini TZ , Li D. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research, 40: 1202-1210

Larkin MA, Blackshields G, Brown NP. 2002. ClustalW and ClustalX version 2. Bioinformatics, 21: 2947-2948

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK - a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26: 283-291

Lesk AM. 2002. Introduction to Bioinformatics. Oxford University Press Inc., New York, USA

Letunic I, Doerks T, Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Research,40: D302-D305

Lin X, Kaul S, Rounsley S. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature, 402 (6763): 761-768

Liu X, Wang L, Zhao K. et al. 2008. The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. Nature, 451: 846-850

Loidl P. 1994. Histone acetylation: facts and questions. Chromosoma, 103(7): 441-449

Machida S, Yuan YA. 2013. Crystal Structure of *Arabidopsis thaliana* Dawdle Forkhead-Associated Domain Reveals a Conserved Phospho-Threonine Recognition Cleft for Dicer-Like 1 Binding. Molecular Plant, [Epub ahead of print].

Martinowich K, Hattori D, Wu H. 2003. DNA methylation-related chromatin remodeling in activity - dependent BDNF gene regulation. Science, 302(5646): 890-893

McGraw S, Vigneault C, Sirard MA. 2007. Temporal expression of factors involved in chromatin remodeling and in gene regulation during early bovine in vitro embryo development. Reproduction, 133(3): 597-608

Nawrath C, Schell J, Koncz C. 1990. Homologous domains of the largest subunit of eucaryotic RNA polymerase II are conserved in plants. Molecular Genetics and Genomics, 223: 65-75

Ogryzko VV, Schiltz RL, Russanova V. 1996. The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases. Cell, 87: 953-959

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

Sayers EW, Barrett T, Benson SH. 2009. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, 37: 5-15

Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. Proceedings of the National Academy of Sciences USA, 11: 5857-5864

Schwede T, Kopp J, Guex N., Peitsch MC. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research, 31: 3381-3385

Servet C, Conde e Silva N, Zhou DX. 2010. Histone acetyltransferase AtGCN5/HAG1 is a versatile regulator of developmental and inducible gene expression in *Arabidopsis*. Molecular Plant, 3(4): 670-677

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

Tabata S, Kaneko T, Nakamura Y. 2000. Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. Nature, 408: 823-826

Tamura K, FukaoY, Iwamoto M. 2010. Identification and Characterization of Nuclear Pore Complex Components in *Arabidopsis thaliana*. Plant Cell, 22: 4084-4097

The UniProt Consortium. 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Research, 40: D71-D75

Theologis A, Ecker JR, Palm CJ et al. 2000. Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. Nature, 408(6814): 816-820

Vandel L, Trouche D. 2001. Physical association between the histone acetyl transferase CBP and a histone methyl transferase. EMBO Reports, 2: 21-26

Yang M, Sun J, Sun X et al. 2009. Caenorhabditis elegans Protein Arginine Methyltransferase PRMT-5 Negatively Regulates DNA Damage-Induced Apoptosis. PLoS Genetics, 5(6): e1000514

Yuan LW, Giordano A. 2002. Acetyltransferase machinery conserved in p300/CBP family proteins. Oncogene, 21: 2253-2260