ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Bioinformatics Study of the Structure and Function of *Avicennia marina* NAC-domain Protein

***Uzma Jabeen[1], Arif Zubair[2] and Asmat Salim[3] and Sikandar Khan Sherwani [4]**

1. Department of Biochemistry, Federal Urdu University of Arts Science and Technology Gulshan-e-Iqbal Campus Karachi, Karachi, Pakistan.

2. Department of Environmental Sciences, Federal Urdu University of Arts Science and Technology Gulshan-e-Iqbal Campus Karachi, Karachi, Pakistan.

3. Dr Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University, Karachi, Karachi-75270, Pakistan.

4. Department of Microbiology, Federal Urdu University of Arts Science and Technology Gulshan-e-Iqbal Campus Karachi, Karachi, Pakistan.

Manuscript Info

Manuscript History:

Received: 12 October 2013
Final Accepted: 24 October 2013
Published Online: November 2013

Key words:
Homology Modeling, Structure Prediction, NAC protein, Stress, *Avicennia marina*.

Abstract

The NAC domain protein family is a distinct plant family of transcription factors that consists of DNA binding domains. NAC proteins perform diverse vital functions that not only regulate necessary developmental processes but also play a significant role in regulating defense responses to plant stress as well as in transcriptional control in different plant processes. NAC domain of NAC proteins has shown the ability to bind DNA, while the C-terminal region comprises transcriptional activation domain. Recent work was conducted to predict structure of functionally distinct regions of this protein. Phylogenetic analysis of salt tolerant *Avicennia marina* NAC-domain protein (AMNAC) was done by using sequences of salt sensitive and salt tolerant NAC proteins. Structural comparison of AMNAC with that of salt-sensitive species was performed. Theoretical model of AMNAC closely resembled that of *Arabidopsis* with slight difference in the structure with conserved overall fold having rich basic and some acidic charge distribution. Lys residues are found to be located at positions 98, 109,110 and 124 in β-sheet providing the binding cleft for the DNA. Differences were observed in a small part of the structure where the amino acid residues are involved in a different set of hydrogen bonded interactions. The overall protein sequence was also studied for the presence of possible post-translational modification sites.

Introduction

The NAC domain protein family is a fascinating family of transcription factors distinct to plants [1, 2]. Transcription factors play important role in plant-specific reactions and interestingly, many of them illustrate no noticeable sequence homology to those of other eukaryotes or bacteria. Plant-specific transcription factors are classified according to DNA-binding domains (DBDs) that are distinct from those of prokaryotes or other lineages of eukaryotes [3]. Recently, large families of plant specific transcription factors were identified and classified according to their DNA binding domain as determined by NMR spectroscopy or X-ray crystallography [4]. Out of 1500 probable transcription factors identified in the *Arabidopsis thaliana* genome, approximately 45% are classified into these plant-specific transcription factors [5].

NAC transcription factors (the name derives from NAM, no apical meristem, *Petunia*; ATAF1-2, and CUC2, cup-shaped cotyledon, *Arabidopsis*), consist of same DNA binding domains [1, 6-8]. Since the first NAC gene isolated

from *Petunia* [8] and CUC2 from *Arabidopsis* that are involved in shoot apical meristem development, several NAC members were discovered that appear to coordinate in plant development like NAP (NAC-like activated by AP3/PI) in development of specific flower organs [9], NAC1 in auxin-dependent formation of the lateral root system [10] and NST1 and NST3 in the regulation of secondary wall thickening in woody tissues [11]. Additionally some members of these families play a significant role in the regulation of defense responses against plant stress i.e. high salinity, drought in case of *Arabidopsis thaliana* [12-13], *Oryza sativa* [14-16] and *Brassica napus* [17]. As transcription activators, NAC proteins are implicated to control a variety of plant processes [8, 10, 18].

The NAC domain consists of a twisted antiparallel β-sheet packed against an α-helix on one side and a shorter helix on the other side at the N-terminus surrounded by a few helical elements. The transcriptional activities reside in the C-terminal region, although the C-terminal sequences are quite diverse [13]. The NAC domain is divided into 5 subdomains, A-E. These subdomains also consist of nuclear localization signals (NLS) [2]. The overall structure suggests that the NAC domain mediates dimerization through conserved interactions [19]. It was also determined experimentally that NAC-DNA complex structure and the unbound NAC have a limited flexibility of the NAC dimer arrangement, which involves recognizing suboptimal binding sites [20]. Variation of highly conserved basic residues in a loop region was found to abolish DNA binding [21].

In the present study, 3D structure of NAC-domain protein from salt-tolerant specie, *Avicennia marina* was predicted by homology modeling. Sequence analysis was performed by aligning the amino acid sequences of salt-tolerant and salt-sensitive plant species and conserved and non conserved amino acid residues were analyzed. The 3D model of NAC domain protein from *Avicennia marina* (AMNAC) was compared with that of *Arabidopsis thaliana* (ANAC) (salt-tolerant) and *Oryza sativa* (salt-sensitive) and structurally variable regions were analyzed. Sequence analysis with respect to the possible post-translational modification sites in AMNAC was also performed.

Methods

Sequence analysis

Pair-wise sequence alignment: Primary amino acid sequence of *Avicennia marina* NAC-domain protein (AMNAC) (A7LKM8_AVIMR) was retrieved from UniProt database [22]. The sequence was then submitted to BLAST server [23] for search against PDB [24] to identify template for molecular modeling. The sequence was also submitted to Protein Structure Prediction Server, PSIPRED [25] for prediction of secondary structural elements. BLAST search gave highest homology (70%) with *Arabidopsis thaliana* NAC-domain protein (pdb id: 1ut4) [26]. The N-terminal residues, 1-161 of the target sequence, AMNAC were aligned with residues, 5-163 of 1ut4 by the BLAST server. Some manual adjustments were done where needed.

Multiple sequence alignment and phylogenetic analysis: Primary amino acid sequences of NAC-domain protein from *Avicennia marina* (AMNAC) and that of salt stress and salt tolerant plant species were retrieved from UniProt database [22]. Multiple sequence alignment was performed by CLUSTAL X [27]. In all, 46 sequences of salt sensitive and 44 sequences of salt tolerant plant NAC proteins were used for constructing phylogenetic tree [28-29]. Phylogenetic analysis based on Neighbor-Joining (NJ) method was performed for reconstructing phylogenetic tree from evolutionary distance data by using Drawgram program and Drawtree of PHYLIP tree package version 367.

Prediction of post-translational modification sites: Primary sequence of AMNAC was searched at PROSITE database using ScanProsite [30] for the possible post-translational modification sites.

Homology modeling

Homology model of AMNAC and AMNAC bound with DNA were built using protein structure-modeling program MODELLER 9.10 [31, 32]. The structural coordinates of the DNA-binding NAC domain of *Arabidopsis thaliana*, pdb, id: 1ut4 and pdb, id: 3swm were used as the templates respectively. Structural comparison of AMNAC was performed with that of salt-sensitive *Oryza sativa* (pdb, id: 3ulx) and salt-tolerant *Arabidopsis thaliana* (pdb, id: 1ut4). Protein structures were visualized and analyzed using Swiss PDB Viewer (SPDV) 37 [33] and WebLab viewer 4.0 [34].

Model assessment

Assessment of the predicted homology models was done based on the analysis of geometry, stereochemistry and energy distributions. The consistency of the predicted homology models was assessed using the ENERGY command

of the MODELLER. The stereochemical quality of the best model was further evaluated by the program PROCHECK [35] and ProSa [36, 37]. The variability among the models was compared by superposition of the Cα traces of the model and the template and the RMSD between the equivalent atoms was determined.

Results and Discussion

Multiple sequence alignment and phylogenetic analysis

The complete protein sequences of all NAC subfamilies were aligned using the CLUSTAL X program [27] (Table 1). NAC transcription factor, AMNAC (A7LKM8_AVIMR) consists of 303 amino acid residues. The N terminal region of the AMNAC consists of 1-161 residues. Subdomains, A-E consist of sequence patch comprising residues 6-35, 41-57, 61-99, 104-135 and 146-161, respectively. AMNAC is highly identical to that of *Petunia x hybrida, Bruguiera gymnorhiza, Solanum lycopersicum* and *Solanum tuberosum.* Multiple sequence alignment of NAC family protein has also shown the distribution of charged amino acid residues in each subdomain. AMNAC consists of unequal distribution of charged amino acids as in case of other NAC proteins. Subdomain B consists of a high percentage of acidic amino acid residues. Subdomains, C and D consist of large number of basic amino acid residues as compared to acidic amino acids. These subdomains are involved in nuclear localization signals (NLS) [1, 2]. The subdomains, D and E are involved in DNA binding [1]. In contrast to the conserved N-terminal region, the C-terminal region of NAC protein is highly variable in length and amino acid sequence in all members of the NAC family.

The phylogenetic relationship of AMNAC was done by the construction of unrooted and rooted neighbor-joining trees [28, 29] using 46 sequences of salt sensitive and 44 sequences of salt tolerant plant NAC proteins (Fig. 1). Phylogenetic tree of the salt tolerant proteins was divided into three clades; AMNAC protein is present in the first clade together with other NAC domain proteins belonging to the major subfamilies of *Arabidopsis thaliana*, *Gossypium hirsutum, Hordeum vulgare var distichum*, *Lotus japonicus* and *Saccharum officinarum.* The cladogram shows that AMNAC is phylogenetically related to the identical NAC proteins of Tomato; *Solanum lycopersicum* (SlNAC1), Potato; *Solanum tuberosum* (StNAC) and *Arabidopsis thaliana* (ANAC). The members of these subfamilies are involved in the biological processes of regulation of transcription and response to environmental stress. Proteins consisting of similar domains may have similar biological functions. The member of these proteins does not have a classical helix–turn–helix motif; in its place it reveals a new transcription factor fold consisting of a twisted β-sheet surrounded by a few helical elements [19]. It can therefore, be predicted that members of one subgroup of NAC-domain proteins may also have similar roles in the development and in response to environmental stress.

Molecular model of AMNAC

Analysis of the amino acid sequence deduced for AMNAC was done through BLAST non-redundant database [23]. AMNAC showed 70% sequence identity with that of *Arabidopsis thaliana* (ANAC; pdb id; 1ut4) [19] (Fig. 2). Theoretical model of AMNAC was built by the program MODELLER 9.10 [31, 32]. The Ramachandran plot of PROCHECK [35] and energy plot of ProSa [36, 37] showed that the quality of the models was reasonably good (Fig. 3). Moreover, the Cα superposition of AMNAC model with the respective template gave the root mean square deviations (RMSD) of 0.46Å which indicates that there is a strong structural conservation between the two proteins.

Theoretical 3D model of AMNAC closely resembled the DNA-binding NAC domain of *Arabidopsis* (ANAC), with slight difference in the structure. The overall fold is maintained which consists of twisted anti-parallel β-sheet packed against an N-terminal α-helix and loops [19] (Fig. 4). A small part of the structure in AMNAC did not exactly superpose with the template because of the difference in amino acid residues. Fig. 6 shows part of the structure in AMNAC where these residues are involved in different sets of hydrogen bonded interactions. These interactions include Val100 (O)-Leu106 (N), Val100 (N)-Gly107 (O) and Lys102-Lys104. Furthermore, Thr106 formed four hydrogen bonds; one side chain-side chain interaction with Glu25 and three main chain-main chain interactions with Gly108 and Gln109 in ANAC (Fig. 4). These interactions were not observed in AMNAC. Homology model of AMNAC has the surface representations of rich basic and some acidic charge distribution (Fig. 5). The conserved basic (Arg lys and His), acidic (Glu and Asp) residues are involved in almost similar hydrogen bonds and ionic interactions as in case of NAC domains of *Arabidopsis* (ANAC) and *Saccharum* sp (SsNACs). Arg15 of AMNAC formed ionic interaction with Glu63 and hydrogen bond with His17 as in case of ANAC, although Tyr is present at position 17 in the latter. It may be possible that the conserved basic residues in the terminal two β strands present in most of the plant NAC proteins mediate DNA binding [4].

Structural comparison of AMNAC with that of salt-sensitive and salt-tolerant plants

Superposition studies of NAC domain protein from *Avicennia marina* (AMNAC) with that of *Arabidopsis thaliana* (ANAC) (salt-tolerant) and *Oryza sativa* (salt-sensitive) showed RMSD of 0.3 Å and 1.1 Å respectively (Fig. 6). These plant species have revealed that there are four regions of structural deviation (Fig. 6; Table 2). These regions were analyzed with respect to their functions and the role of significant residues was highlighted.

Region 1 (residues 29-42): The first major region comprising residues 29-42 occupies large portion of AMNAC structure. It did not exactly superpose with the templates ANAC (pdb id: 1ut4) and *Oryza sativa* (pdb id: 3ulx) because of the difference in amino acid residues and having structurally variable RMSD of 0.13 Å and 1.59 Å respectively. Homology model of AMNAC closely resembles salt tolerant plant protein (Fig. 6). These residues are involved in different sets of hydrogen bonded interactions (Fig 6; Table 2). These interactions include Cys29-Cys32 and Arg30-Val-25 observed in both salt tolerant NAC proteins and not in salt sensitive *Oryza sativa* NAC protein. Similarly, hydrogen bonding interaction between Arg27-His23 was found in salt sensitive *Oryza sativa* NAC protein while not in case of both salt tolerant species, AMNAC and ANAC.

Region 2 (residues 73-83): In this region, conserved residues are present but these residues did not exactly superpose due the missing residues 79-85 in 1ut4 and 79-83 in 3ulx in the structure of these two proteins.

Region 3 (residues 98-108 residues): In this region, Lys98 formed ionic interaction with Asp20 in salt tolerant AMNAC and ANAC proteins while this interaction was not found in salt sensitive *Oryza sativa* NAC protein.

Region 4 (residues 133-149): This part of the structure did not exactly superpose due to the presence of different structurally variable amino acid residues. A hydrogen bond was observed between Lys145-Thr144 in both salt tolerant species AMNAC and ANAC while in case of salt sensitive specie *Oryza sativa*, Lys145 formed interaction with Asn143.

Homology model of AMNAC bound with DNA

The homology model of AMNAC domain bound with DNA mainly consists of β-fold structure closely resembling the structure of ANAC [20] (Fig. 7). Plant specific NAC proteins, WRKY transcription factors and the mammalian GCM (Glial cell missing) transcription factors use a β-strand motif for DNA-binding. Similarly the homology model of AMNAC has shown that the NAC domain is placed at the edge of its core β-sheet containing Lys residues at position 98, 109,110 and 124. Trp115 and Thr125 formed major channel for DNA binding (Fig. 7). It was documented that the structure of the NAC-DNA complex has shown limited flexibility of the NAC dimer arrangement, which could be important to discriminate suboptimal binding sites [20].

Predicted post-translational modification sites

Transcription factors are present in their inactive forms in the cytoplasm, and upon stimulation, they are activated and localized into the nucleus. These transcription factors can be activated by protein phosphorylation [38]. The activity of some NAC proteins can also be modulated post-translationally by binding with co-factors, glucosamines or by proteolytic processing [39]. The Scan Prosite [30] result predicted that AMNAC has several potential glycosylation, phosphorylation and N-myristoylation sites (Table 3).

Glycosylation site: An N-glycosylation site, Asn78 in the sequence patch, ^{78}Asn-Gly-Ser-Arg^{81} was predicted in AMNAC protein by PROSITE (Table 3). It has been documented that the information for the N-glycosylation of particular Asn residues is encoded in the primary structure of the polypeptide chain itself and a triplet sequence has been recognized as Asn-X-Thr/Ser [40, 41]. In the AMNAC model, Asn78 was found on the surface having surface accessibility of 41.98Å^2 and is part of the loop structure. We compared this sequence patch with the proposed criteria for such a site in glycosylated proteins [40, 42]. As Asn78 is present in the loop region, there is a strong probability that glycosylation can take place at this residue. However, the position of proline residue is also critical. Its presence either between Asn and Ser/Thr or at the next position C-terminal to Ser/Thr can prevent glycosylation [40]. In AMNAC, there is no Pro within the proposed sequence while Pro82 is present C-terminal to Arg81. This Pro might prevent glycosylation. In the absence of any experimental data regarding glycosylation in NAC proteins, further evidence is needed to confirm it.

Phosphorylation sites: Table 3 shows the potential sites of phosphorylation in AMNAC. PROSITE search has identified some sequence patches consistent with the phosphorylation sites in the known proteins. These patches

consist of phosphorylation sites recognized by many phosphokinases such as casein kinase 2 and protein kinase C. The potential residues for phosphorylation were analyzed with respect to their positions in the homology model. A few of these sites cannot be the predicted as phosphorylation sites because of their extremely low surface accessibilities, whereas others with their exposed positions on the surface could be considered as possible phosphorylation sites. Sequences of plant NAC proteins were also aligned by CLUSTAL X and analyzed for the presence of conserved phosphorylation sites.

(a) Casein kinase 2 phosphorylation sites: Casein kinase 2 (CK2) is a well-conserved protein-kinase present in different organisms, including plants, mammals and yeast [43-45]. In several organisms including *Arabidopsis,* it functions as a circadian clock component. One of the CK2 targets in *Arabidopsis* is the circadian clock associated protein-1 (CCA1) where it influences photoperiodic flowering [44]. In addition, CK2 is involved in various biological phenomena which include modulation of DNA-binding ability, protein stability, intracellular localization etc. [46-49]. PROSITE predicted various potential CK2 sites in AMNAC; three N-terminal sites were ^{19}Thr, ^{70}Ser and ^{94}Thr (Table 3). Analysis of the homology model showed that Thr19, Ser70 and Thr94 have surface accessibilities of 22.97Å^2, 0.317Å^2 and 16.52Å^2 respectively. The sequence patch containing Thr19 is highly conserved in the NAC family. Thr19 is also part of loop which makes it a highly probable target for CK2 phosphorylation site as compared to the other possible Thr/Ser phosphorylation sites. Ser70 has a low surface accessibility while Thr94 is part of the β-sheet structure which makes these residues less likely to be the targets for phosphorylation.

(b) Protein kinase C phosphorylation site: Protein kinase C (PKC) is an enzyme family of at least 10 distinct isoforms with unique tissue distributions and dynamic subcellular localization. It induces both short term alterations in cellular activities and long term effects such as differentiation, proliferation, and apoptosis [50]. Multiple signal transduction pathways can be mobilized to mediate PKC activation during oxidative stress. PROSITE has predicted one potential PKC phosphorylation site at Ser70 in the N-terminal region (Table 3). Homology model of AMNAC have shown that Ser70 is mainly buried (accessibility = 0.3171Å^2) and is part of the β-sheet. Although, it is highly conserved in the NAC protein family, Ser70 cannot be considered as a phosphorylation site due to its location and low accessibility.

N-myristoylation site: The PROSITE pattern search results has identified one N-myristoylation site in the amino-acid sequence of AMNAC at Gly34 (Table 3). This region was further confirmed by the prediction of potential myristoylation site by using the NMT website (http://mendelimpunivieacat/myristate/). Homology model of AMNAC has shown that Gly34 is present on the surface and it is a part of the loop. Gly34 has a relatively higher surface accessibility of 13.78Å^2. This sequence patch is 30% conserved in NAC family. Therefore, Gly34 can be considered a potential myristoylation site but experimental evidence is needed to confirm this site.

Table 1: CLUSTAL X multiple sequence alignment showing the sequence of members of the salt tolerant NAC-domain protein

Subdomain-A

```
A7LKM8_AVIMR      ---------------------------------------------MKG---GDQQLNLPA 12
NAC68_ARATH       -----------------------------------------------------MMKGLI  6
NAC69_ARATH       -----------------------------------------------------MVKDLV  6
NAC3_ARATH        -----------------------------------------------------METPV  5
NAC4_ARATH        -----------------------------------------------------MMNPV  5
NAC5_ARATH        -----------------------------------------------------MANPV  5
Q948Z2_SOLTU      ---------------------------------------------MNKGATGNQQLELPA 15
Q6RH27_SOLLC      ---------------------------------------------MNKGANGNQQLELPA 15
C0J1R1_GOSHI      ---------------------------------------------MKA------ELELPP  9
B3IX39_LOTJA      ---------------------------------------------MKG------ELELPP  9
C0J1R4_GOSHI      ---------------------------------------------MTA---TELQLRLPA 12
C0J1R2_GOSHI      ---------------------------------------------MTA---SELQL--PP 10
NAC2_ARATH        ---------------------------------------------MS---ELLQL--PP  9
A4HRC1_HORVD      ---------------------------------------------MSGG----QELNLPP 11
Q4QWQ6_SACOF      ---------------------------------------------MSGGG---QDLQLPP 12
NA102_ARATH       --------MDFALFSSISIFEINHKDPPIRRFIKTQNRILSTRKQQGTFPKMKAELNLPA 52
A4HRC0_HORVD      -------------------------------MVKAEAMTAEAEGSSGRRDAEAELNLPP 28
NAC19_ARATH       ---------------------------------------------MGIQETDPLTQLSLPP 16
NAC55_ARATH       ---------------------------------------------MGLQELDPLAQLSLPP 16
NAC72_ARATH       ---------------------------------------------MGVREKDPLAQLSLPP 16
C0J1R5_GOSHI      ---------------------------------------------MGVPETDPLAQLSLPP 16
NAC29_ARATH       ------------------------------------------MEVTS--------QSTLPP 11
```

```
C0J1R6_GOSHI      ----------------------------------------MNMKHP-------QSSLPP 12
NAC18_ARATH       ----------------------------------------MESTDSSGGPPPPQPNLPP 19
NAC42_ARATH       ---------------------------------------MSGEGNLGKDHEEENEAPLP 20
NAC9_ARATH        ----------------------------------------MGDRNNDGDQKMED-VLLP 18
NAC54_ARATH       ---------------------------------------MDVDVFNGWGRPRFEDESLMPP 22
NAC98_ARATH       ---------------------------------------MDIPYYH---YDHGGDSQYLPP 19
C0J1R3_GOSHI      ---------------------------------------MEDAIVVN----QRQQELMELPP 19
C0KLH1_PONTR      ----------------------------------------MEK----YDDQEQIELPP 14
NAC22_ARATH       ---------------------------------------METEEEMKE---SSISMVEAKLPP 21
A4HRC2_HORVD      ----------------------------------------MSM---SFLSMVETELPP 15
NAC78_ARATH       ----------------------------------------------MGRGSVTSLAP 11
Q8LAH6_ARATH      ----------------------------------------------MGRGSVTSLAP 11
NAC43_ARATH       ---------------------------------------MMSKSMSISVNGQSQVPP 18
NAC66_ARATH       ----------------------------------------------MNISVNGQSQVPP 13
NAC12_ARATH       ---------------------------------------MADNKVNLSINGQSKVPP 18
NAC7_ARATH        ----------------------------------------------MNSFSHVPP 9
NAC61_ARATH       ------------------------MG-----------------------------EELSV 7
Q5Y5S4_HORVD      ---------------MTRSRPTTTMGGSVPDQHHQQQQHDGEVDGGQLQHGGEHVETVMP 45
NACA1_ARATH       ----------------------------------------------------MTTE--- 4
NACA2_ARATH       ---------------------------------------------MSPPPAVVTESAD 13
NACA5_ARATH       ----------------------------------------------MPGAIVEE--- 8
NAC1_ARATH        ------------------------------------------------------MEDQV 5
NAC8_ARATH        MAGRSWLIDSNRIATKIMSASASSDPRQVVWKSNPSRHCPKCQHVIDNSDVVDDWPGLPR 60
NAC6_ARATH        ------------------------------------------------------MKILPV 6

                            Subdomain-A                    Subdomain-B
A7LKM8_AVIMR      GFRFHPTDEELVVHYLCRKCAGQQ--------IGVPVIAEIDLYKFDP-WELPDLALYGE 63
NAC68_ARATH       GYRFSPTGEEVINHYLKNKLLG-K------YWLVDEAISEINILSHKPSKDLPKLARIQS 59
NAC69_ARATH       GYRFYPTGEELINHYLKNKILG-K------TWLVDEAISEINICSYDP-IYLPSLSKIKS 58
NAC3_ARATH        GLRFCPTDEEIVVDYLWPKNSDRD------TSHVDRFINTVPVCRLDP-WELPCQSRIKL 58
NAC4_ARATH        GFRFRPNDEEIVDHYLRPKNLDSD------TSHVDEVISTVDICSFEP-WDLPSKSMIKS 58
NAC5_ARATH        GFRFRPTDGEIVDIYLRPKNLESN------TSHVDEVISTVDICSFDP-WDLPSHSRMKT 58
Q948Z2_SOLTU      GFRFHPTDDELVQHYLCRKCAGQP--------IAVSIITEIDLYKFDP-WQLPEKALYGE 66
Q6RH27_SOLLC      GFRFHPTDDELVQHYLCRKCAGQS--------IAVSIIAEIDLYKFDP-WQLPEKALYGE 66
C0J1R1_GOSHI      GFRFHPTDDELVNHYLCRKCASQP--------ISVPIIAEIDLYKFDP-WQLPDMALYGE 60
B3IX39_LOTJA      GFRFHPTDEELVNHYLCTKCAGQS--------FNYSVIKEIDLYKFDP-WQLPEMGFDGE 60
C0J1R4_GOSHI      GFRFHPTDEELVMHYLCRKCASQS--------IAVPIIAELDLYKYDP-WDLPDLALYGE 63
C0J1R2_GOSHI      GFRFHPTDEELVMHYLCRKCASQS--------IAVPIIAEIDLYKYDP-WDLPDLALYGE 61
NAC2_ARATH        GFRFHPTDEELVMHYLCRKCASQS--------IAVPIIAEIDLYKYDP-WELPGLALYGE 60
A4HRC1_HORVD      GFRFHPTDEELVMHYLCRRCAGAP--------IAVPIITEIDLYKLDP-WQLPKMAMYGE 62
Q4QWQ6_SACOF      GFRFHPTDEELVMHYLCRRCASLP--------IAVPIIAEIDLYKFDP-WQLPRMALYGE 63
NA102_ARATH       GFRFHPTDEELVKFYLCRRCASEP--------INVPVIAEIDLYKFNP-WELPEMALYGE 103
A4HRC0_HORVD      GFRFHPTDDELVVHYLCRKVAGQP--------QPVPIIAEVDLYKFNP-WDLPERALFGS 79
NAC19_ARATH       GFRFYPTDEELMVQYLCRKAAGYD--------FSLQLIAEIDLYKFDP-WVLPNKALFGE 67
NAC55_ARATH       GFRFYPTDEELMVEYLCRKAAGHD--------FSLQLIAEIDLYKFDP-WVLPSKALFGE 67
NAC72_ARATH       GFRFYPTDEELLVQYLCRKVAGYH--------FSLQVIGDIDLYKFDP-WDLPSKALFGE 67
C0J1R5_GOSHI      GFRFYPTDEELLVQYLCRKVAGHH--------FSLQIIGEIDLYKFNP-WDLPSKALFGE 67
NAC29_ARATH       GFRFHPTDEELIVYYLRNQTMSKP--------CPVSIIPEVDIYKFDP-WQLPEKTEFGE 62
C0J1R6_GOSHI      GFRFHPTDEELILHYLKKKITSSP--------FPVSIIADVDIYKFDP-WDLPDKAAFGE 63
NAC18_ARATH       GFRFHPTDEELVIHYLKRKADSVP--------LPVAIIADVDLYKFDP-WELPAKASFGE 70
NAC42_ARATH       GFRFHPTDEELLGYYLRRKVENKT--------IKLELIKQIDIYKYDP-WDLPRVSSVGE 71
NAC9_ARATH        GFRFHPTDEELVSFYLKRKVQHNP--------LSIELIRQLDIYKYDP-WDLPKFAMTGE 69
NAC54_ARATH       GFRFHPTDEELITYYLLKKVLDSN--------FSCAAISQVDLNKSEP-WELPEKAKMGE 73
NAC98_ARATH       GFRFHPTDEELITHYLLRKVLDGC--------FSSRAIAEVDLNKCEP-WQLPGRAKMGE 70
C0J1R3_GOSHI      GFRFHPTDEEIITHYLLEKVMNSN--------FSAAAIGEADLNKSEP-WDLPDKAKMGE 70
C0KLH1_PONTR      GFRFHPTDEELITHYLTPKVFDGC--------FSARAIGEVDLNKCEP-WDLPRRAKMGE 65
NAC22_ARATH       GFRFHPKDDELVCDYLMRRSLHNN-------HRPPLVLIQVDLNKCEP-WDIPKMACVGG 73
A4HRC2_HORVD      GFRFHPRDDELICDYLAPKVTGKVGFS---GRRPP--MVDVDLNKVEP-WDLPVTASVGG 69
NAC78_ARATH       GFRFHPTDEELVRYYLKRKVCNKP--------FKFDAISVTDIYKSEP-WDLPDKSKLKS 62
Q8LAH6_ARATH      GFRFHPTDEELVRYYLKRKICNKP--------FKFDAISVTDVYKSEP-WDLPDKSRLKS 62
NAC43_ARATH       GFRFHPTEEELLQYYLRKKVNSIE--------IDLDVIRDVDLNKLEP-WDIQEMCKIGT 69
NAC66_ARATH       GFRFHPTEEELLKYYLRKKISNIK--------IDLDVIPDIDLNKLEP-WDIQEMCKIGT 64
NAC12_ARATH       GFRFHPTEEELLHYYLRKKVNSQK--------IDLDVIREVDLNKLEP-WDIQEECRIGS 69
NAC7_ARATH        GFRFHPTDEELVDYYLRKKVASKR--------IEIDFIKDIDLYKIEP-WDLQELCKIGH 60
NAC61_ARATH       GFRFYPTEVELLTYYLRIQLGGGN-------ATIHSLIPILDVFSVEP-TQLPNLAGERC 59
Q5Y5S4_HORVD      GFRFHPTEEELIEFYLRRKVDGKR-------FNID-LIASVDLYRYDP-WDLPALASIGD 96
NACA1_ARATH       ---EKEILAAKLEE--QKIDLDKP--------EVEDDDDNEDDDSDDDDKDDDEADGLDG 51
NACA2_ARATH       GQPEQPPVTAIAEELEKKLQTDEP--------IVEDVKDDEDDDDDDEEEEDDDAQGVS- 64
NACA5_ARATH       ---EKSQIESIKEQL-KLEKEDDV--------VVEDVKDGEEED-DDEDDEDVEVEGEGG 55
NAC1_ARATH        GFGFRPNDEELVGHYLRNKIEGNTSR------DVEVAISEVNICSYDP-WNLRFQSKYKS 58
NAC8_ARATH        GVKFDPSDPEIIWHLLAKSGLSGLSSHPFIDEFIPTVNQDDGICYTHP-KNLPGVKSDGT 119
```

```
NAC6_ARATH        GSRFCPTDLGLVRLYLRNKVERNQ----------SSFITTMDIHQDYP-WLLPHVNNPLF 55

                                    Subdomain-C
A7LKM8_AVIMR      K----EWYFFSPRDRKYPNGSRPNRAAGTG------YWKATGADKPVGKP---------- 103
NAC68_ARATH       EDL--EWYFFSPIEYTNPNKMKMKRTTGSG------FWKPTGVDREIRDKRG------NG 105
NAC69_ARATH       DDP--VWYFFCPKEYTSAKKKVTKRTTSSG------YWKATGVDRKIKDKRG------NR 104
NAC3_ARATH        KDV--AWCFFRPKENKYGRGDQQMRKTKSG------FWKSTGRPKPIMR---------NR 101
NAC4_ARATH        RDG--VWYFFSVKEMKYNRGDQQRRRTNSG------FWKKTGKTMTVMRKRG------NR 104
NAC5_ARATH        RDQ--VWYFFGRKENKYGKGDRQIRKTKSG------FWKKTGVTMDIMRKTG------DR 104
Q948Z2_SOLTU      K----EWYFFSPRDRKYPNGSRPNRAAGTG------YWKATGADKPVGKP---------- 106
Q6RH27_SOLLC      K----EWYFFSPRDRKYPNGSRPNRAAGTG------YWKATGADKPVGKP---------- 106
C0J1R1_GOSHI      K----EWYFFSPRDRKYPNGCRPNRAAGTG------YWKATGADKPIGGP---------- 100
B3IX39_LOTJA      K----EWYFFSPRDRKYPNGSRPNRAAGSG------YWKATGADKPIGKP---------- 100
C0J1R4_GOSHI      K----EWYFFSPRDRKYPNGSRPNRAAGSG------YWKATGADKPIGQP---------- 103
C0J1R2_GOSHI      E----EWYFFSPRDRKYPNGSRPNRAAGSG------YWKATGADKPIGQP---------- 101
NAC2_ARATH        K----EWYFFSPRDRKYPNGSRPNRSAGSG------YWKATGADKPIGLP---------- 100
A4HRC1_HORVD      K----EWYFFSPRDRKYPNGSRPNRAAGSG------YWKATGADKPVGTP---------- 102
Q4QWQ6_SACOF      K----EWYFFSPRDRKYPNGSRPNRAAGSG------YWKATGADKPVGTP---------- 103
NA102_ARATH       K----EWYFFSHRDRKYPNGSRPNRAAGTG------YWKATGADKPIGKP---------- 143
A4HRC0_HORVD      R----EWYFFTPRDRKYPNGSRPNRSAGTG------YWKATGADKPVAPRESG------G 123
NAC19_ARATH       K----EWYFFSPRDRKYPNGSRPNRVAGSG------YWKATGTDKIISTE---------G 108
NAC55_ARATH       K----EWYFFSPRDRKYPNGSRPNRVAGSG------YWKATGTDKVISTE---------G 108
NAC72_ARATH       K----EWYFFSPRDRKYPNGSRPNRVAGSG------YWKATGTDKIITAD---------G 108
C0J1R5_GOSHI      K----EWYFFSPRDRKYPNGSRPNRVAGSG------YWKATGTDKIITTE---------G 108
NAC29_ARATH       N----EWYFFSPRERKYPNGVRPNRAAVSG------YWKATGTDKAIHSGSS--------- 104
C0J1R6_GOSHI      K----EWYFFSPRDRKYPNGARPNRAAASG------YWKATGTDKVIVTSSMADRRGEVQ 113
NAC18_ARATH       Q----EWYFFSPRDRKYPNGARPNRAATSG------YWKATGTDKPVISTGGGGSK---- 116
NAC42_ARATH       K----EWYFFCMRGRKYRNSVRPNRVTGSG------FWKATGIDKPVYSN---------L 112
NAC9_ARATH        K----EWYFYCPRDRKYRNSSRPNRVTGAG------FWKATGTDRPIYSSEG-------N 112
NAC54_ARATH       K----EWYFFTLRDRKYPTGLRTNRATEAG------YWKATGKDREIKSSKTK------- 116
NAC98_ARATH       K----EWYFFSLRDRKYPTGLRTNRATEAG------YWKATGKDREIFSSKTC------- 113
C0J1R3_GOSHI      K----EWYFFCQRDRKYPTGMRTNRATQAG------YWKATGKDKEIFKGKG-------- 112
C0KLH1_PONTR      K----EWYFFCVRDRKYPTGLRTNRATEAG------YWKATGKDKEIYKAK--------- 106
NAC22_ARATH       K----DWYFYSQRDRKYATGLRTNRATATG------YWKATGKDRTILRKG--------- 114
A4HRC2_HORVD      K----EWYFYSLKDRKYATGQRTNRATVSG------YWKATGKDRVVARRG--------- 110
NAC78_ARATH       RDL--EWYFFSMLDKKYSNGSKTNRATEKG------YWKTTGKDREIRNGS--------- 105
Q8LAH6_ARATH      RDL--EWYFFSMLNKKYRNGSKTNRATEMG------YWKTTGKDREILNGS--------- 105
NAC43_ARATH       TPQN-DWYFFSHKDKKYPTGTRTNRATAAG------FWKATGRDKIIYSNG--------- 113
NAC66_ARATH       TPQN-DWYFYSHKDKKYPTGTRTNRATTVG------FWKATGRDKTIYTNG--------- 108
NAC12_ARATH       TPQN-DWYFFSHKDKKYPTGTRTNRATVAG------FWKATGRDKIICSCV--------- 113
NAC7_ARATH        EEQS-DWYFFSHKDKKYPTGTRTNRATKAG------FWKATGRDKAIYLRH--------- 104
NAC61_ARATH       RGDAEQWIFFVPRQERGARGGRPSRTTGSG------YWKATGSPGPVFSPD---------N 105
Q5Y5S4_HORVD      K----EWFFYVPRDRKYRNGDRPNRVTPSG------YWKATGADRMVKVVEG-------N 139
NACA1_ARATH       EAG-------GKSKQSRSEKKSRKAMLKLG------MKPITGVSRVTVKKSK-------N 91
NACA2_ARATH       ----------GSSKQSRSEKKSRKAMLKLG------MKPVTGVSRVTIKRTK-------N 101
NACA5_ARATH       N---------ENAKQSRSEKKSRKAVLKLG------MKPVSDVSRVTIKRAK-------N 93
NAC1_ARATH        RD--AMWYFFSRRENN--KGNRQSRTTVSG------KWKLTGESVEVKDQWGFCSEG-FR 107
NAC8_ARATH        VS-----HFFHKAIKAYSTGTRKRRKIHDDDFGDV-RWHKTGRTKPVVLDG--------- 164
NAC6_ARATH        NNN--EWYYFVPLTERGGKILSVHRKVAARGGSEGGTWRSNDGKKEIKDGHMQ------- 106


                          Subdomain-D                          Subdomain-E
A7LKM8_AVIMR      KTLGIKK-ALVFYAGKAPRGVKTNWIMHEYRLANVDRSAGKTK---------N-LRLDDW 152
NAC68_ARATH       VVIGIKK-TLVYHEGKSPHGVRTPWVMHEYHITCLP-HHKRK-----------------Y 146
NAC69_ARATH       GEIGIKK-TLVYYEGRVPKGVWTPWVMHEYHITCLP-QDQRN-----------------Y 145
NAC3_ARATH        QQIGEKK-ILMFYTSK---ESKSDWVIHEYHGFSHN-QMMMT-----------------Y 139
NAC4_ARATH        EKIGEKR-VLVFKNRD---GSKTDWVMHEYHATSLFPNQMMT-----------------Y 143
NAC5_ARATH        EKIGEKR-VLVFKNHG---GSKSDWAMHEYHATFSSPNQIMT-----------------Y 143
Q948Z2_SOLTU      KTLGIKK-ALVFYAGKAPRGIKTNWIMHEYRLANVDRSAGKNN---------N-LRLDDW 155
Q6RH27_SOLLC      KTLGIKK-ALVFYAGKAPRGIKTNWIMHEYRLANVDRSAGKNN---------N-LRLDDW 155
C0J1R1_GOSHI      KALGIKK-ALVFYAGKAPRGVKTNWIMHEYRLANVDRSAGKRS---------NNLRLDDW 150
B3IX39_LOTJA      KALGIKK-ALVFYVGKAPKGIKTNWIMHEYRLANVDRSAGKKS---------N-LRLDDW 149
C0J1R4_GOSHI      KPVGIKK-ALVFYSGKAPKGEKTNWIMHEYRLADVDRSVRKK----------NSLRLDDW 152
C0J1R2_GOSHI      KPVGIKK-ALVFYAGKAPKGEKTNWIMHEYRLADVDRSARKK----------NSLRLDDW 150
NAC2_ARATH        KPVGIKK-ALVFYAGKAPKGEKTNWIMHEYRLADVDRSVRKKK---------NSLRLDDW 150
A4HRC1_HORVD      KPLAIKK-ALVFYAGKAPKGEKTNWIMHEYRLADVDRSARKK----------NSLRLDDW 151
Q4QWQ6_SACOF      KPLAIKK-ALVFYAGKAPKGEKTNWIMHEYRLADVDRSARKK----------NSLRLDDW 152
NA102_ARATH       KTLGIKK-ALVFYAGKAPKGIKTNWIMHEYRLANVDRSASTNKK--------NNLRLDDW 194
A4HRC0_HORVD      RTVGIKK-ALVFYSGRAPRGVKTDWIMHEYRIAQADRTPGKKG----------SLKLDEW 172
NAC19_ARATH       QRVGIKK-ALVFYIGKAPKGTKTNWIMHEYRLIEPSRRNGST-------------KLDDW 154
```

```
NAC55_ARATH       RRVGIKK-ALVFYIGKAPKGTKTNWIMHEYRLIEPSRRNGST-------------KLDDW 154
NAC72_ARATH       RRVGIKK-ALVFYAGKAPKGTKTNWIMHEYRLIEHSRSHGSS-------------KLDDW 154
C0J1R5_GOSHI      RKVGIKK-ALVFYVGKAPKGTKTNWIMHEYRLIETSRKSGSS-------------KLDDW 154
NAC29_ARATH       -NVGVKK-ALVFYKGRPPKGIKTDWIMHEYRLHDSRKASTK---------RNGSMRLDEW 153
C0J1R6_GOSHI      ENIGVKK-ALVFYKGRPPKGMKTNWIMHEYRLADNPNSNFNNRPLK---SKDSSMRLDDW 169
NAC18_ARATH       -KVGVKK-ALVFYSGKPPKGVKSDWIMHEYRLTDNKPTHICDFG-----NKKNSLRLDDW 169
NAC42_ARATH       DCVGLKK-SLVYYLGSAGKGTKTDWMMHEFRLPSTTKT------------DSPAQQAEVW 159
NAC9_ARATH        KCIGLKK-SLVFYKGRAAKGVKTDWMMHEFRLPSLSEPSPPSKRF----FDSPVSPNDSW 167
NAC54_ARATH       SLLGMKK-TLVFYKGRAPKGEKSCWVMHEYRLDGKFSYHYIS-----------SSAKDEW 164
NAC98_ARATH       ALVGMKK-TLVFYKGRAPKGEKSNWVMHEYRLEGKFSYHFIS-----------RSSKDEW 161
C0J1R3_GOSHI      CLVGMKK-TLVFYKGRAPRGEKTNWVMHEYRLEGKFSYYNLP-----------KAAKDEW 160
C0KLH1_PONTR      ALVGMKK-TXXFYRGRAPKGQKTNWVMHEYRLEGKYSAYDHP-----------KTAKNEW 154
NAC22_ARATH       KLVGMRK-TLVFYQGRAPRGRKTDWVMHEFRLQGSHHPPNHSL----------SSPKEDW 163
A4HRC2_HORVD      ALVGMRK-TLVFYQGRAPKGRKTEWVMHEYRLEGAHEQAS----------------KEDW 153
NAC78_ARATH       RVVGMKK-TLVYHKGRAPRGERTNWVMHEYRLS---DEDLKKA----------GVPQEAY 151
Q8LAH6_ARATH      KVVGMKK-TLVYHKGRAPRGERTNWVMHEYRLV---DQDLDKT----------GVHQDAF 151
NAC43_ARATH       RRIGMRK-TLVFYKGRAPHGQKSDWIMHEYRLDDNIISPEDVTVHEVVSIIGEASQDEGW 172
NAC66_ARATH       DRIGMRK-TLVFYKGRAPHGQKSDWIMHEYRLDESVLISSCGDHDVNVETCDVIGSDEGW 167
NAC12_ARATH       RRIGLRK-TLVFYKGRAPHGQKSDWIMHEYRLDDTPMSNG---YADVVTEDPMSYNEEGW 169
NAC7_ARATH        SLIGMRK-TLVFYKGRAPNGQKSDWIMHEYRLETDENGTP----------------QEEGW 148
NAC61_ARATH       RVIGVKK-TMVFYTGKAPTGRKTKWKMNEYKAVETASVSTIP----------------KS 148
Q5Y5S4_HORVD      RSIGLKK-TLVFYVGKAPKGLRSSWIMNEYRLPHGETERYQK----------------EI 182
NACA1_ARATH       ILFVISK-PDVFKSPA----SDTYVIFGEAKIEDLSSQIQSQ-------------AAEQF 133
NACA2_ARATH       VLFFISK-PDVFKSPH----SETYVIFGEAKIEDLSSQLQTQ-------------AAQQF 143
NACA5_ARATH       VLFVISK-PDVYKSPN----AETYVIFGEAKVDDLSSQLQTQ-------------AAQRF 135
NAC1_ARATH        GKIGHKR-VLVFLDGRYPDKTKSDWVIHEFHYDLLPEHQRTYVICRLEYKGDDADILSAY 166
NAC8_ARATH        VQRGCKK-IMVLYGGK---AVKTNWVMHQYHLGIEEDEKEGD----------------Y 203
NAC6_ARATH        KGDGLRASDDLQKVVLCRIRYKKEANVNEFGLVNHQAHQTQD-------------ALTGF 153
                                            :

                  Subdomain-E
A7LKM8_AVIMR      VLCRIYNKKGTLEKHLSVD------------------------------------------ 171
NAC68_ARATH       VVCQVKYKGEAAEIS------YEPSPSLVSDSHTVIAITGEP------------------- 182
NAC69_ARATH       VICQVMYKGEDGDVPSGGNNSSEPSQSLVSDSNTVRATS----------------------- 184
NAC3_ARATH        TLCKVMFNGGMREKSSSSPSSSGVSGIEQSRRDSLIPQLVNN------------------- 181
NAC4_ARATH        TVCKVEFKGEETEISSSS------TGSEIEQIHSLIP-LVNS------------------- 178
NAC5_ARATH        TLCKVKFKGERREFSVAT-------GSGIKHTHSLIPPTNNSG------------------ 179
Q948Z2_SOLTU      VLCRIYNKKGTLEKHYNVD------------------------------------------ 174
Q6RH27_SOLLC      VLCRIYNKKGTLEKHYNVD------------------------------------------ 174
C0J1R1_GOSHI      VLCRIYNKKGSIEKHFPSE------------------------------------------ 169
B3IX39_LOTJA      VLCRIYNKKGKIEKYNTPT------------------------------------------ 168
C0J1R4_GOSHI      VLCRIYNKKGCIEKQPPRGS----------------------------------------- 172
C0J1R2_GOSHI      VLCRIYNKKGAIEKQSPQGS----------------------------------------- 170
NAC2_ARATH        VLCRIYNKKGATERRGPPPP----------------------------------------- 170
A4HRC1_HORVD      VLCRIYNKKGGMEKPASVDRKPVTM------------------------------------ 176
Q4QWQ6_SACOF      VLCRIYNKKGGLEKPAAASSGDHKP------------------------------------ 177
NA102_ARATH       VLCRIYNKKGTMEKYLPAA------------------------------------------ 213
A4HRC0_HORVD      VLCRLYNKKNNWDKVKVEQ------------------------------------------ 191
NAC19_ARATH       VLCRIYKKQSSAQKQVYDNG---------------------------------------- 174
NAC55_ARATH       VLCRIYKKQTSAQKQAYNNL---------------------------------------- 174
NAC72_ARATH       VLCRIYKKTSGSQRQAVTPV---------------------------------------- 174
C0J1R5_GOSHI      VLCRIYKKNSSGQK-PLSSV---------------------------------------- 173
NAC29_ARATH       VLCRIYKKRGASKLLNEQ------------------------------------------ 171
C0J1R6_GOSHI      VLCRIYKKSNALSSTTAIG----------------------------------------- 188
NAC18_ARATH       VLCRIYKKNNSTASRHHHH----------------------------------------- 188
NAC42_ARATH       TLCRIFKR--VTSQRNPTILPP-------------------------------------- 179
NAC9_ARATH        AICRIFKKTNTTTLRALSHSFV-------------------------------------- 189
NAC54_ARATH       VLCKVCLKS-----GVVS------------------------------------------ 177
NAC98_ARATH       VISRVFQKTTLASTGAVSE----------------------------------------- 180
C0J1R3_GOSHI      VVCRVFHKN------TAG------------------------------------------ 172
C0KLH1_PONTR      VICRVFQKS------GGG------------------------------------------ 166
NAC22_ARATH       VLCRVFHKNT----EG------------------------------------------- 175
A4HRC2_HORVD      VLCRVICKKK----SG------------------------------------------- 165
NAC78_ARATH       VLCRIFQKSGTGPKNGEQYGAPYLEEEWEE-DGMTYVPAQDAFSEGLALNDDVYVDIDDI 210
Q8LAH6_ARATH      VLCRIFQKSGSGPKNGEQYGAPFVEEEWEEEDDMTFVPDQ----EDLGSEDHVYVHMDDI 207
NAC43_ARATH       VVCRIFKKKNLHKTLNSPV----------------------------------------- 191
NAC66_ARATH       VVCRVFKKNNLCKNMIS------------------------------------------- 184
NAC12_ARATH       VVCRVFRKKNYQKIDDCP------------------------------------------ 187
```

```
NAC7_ARATH        VVCRVFKKR-LAAVRRMGD---------------------------------------- 166
NAC61_ARATH       GSSRAFDR--------------------------------------------------- 156
Q5Y5S4_HORVD      SLCRVYKRPGIDDNLHLTGTTTRSSGSRAA----------------------------- 212
NACA1_ARATH       KAPDLSNVISKGESSS------------------------------------------- 149
NACA2_ARATH       RMPEIGATSQRAEASTATVE--------------------------------------- 163
NACA5_ARATH       KMPDVTSMLPNAGSEATMAP--------------------------------------- 155
NAC1_ARATH        AIDPTPAFVPNMTSSAGSVVNQSRQ---------------------------------- 191
NAC8_ARATH        VVSKIFYQQPQQLVVKRGDK--------------------------------------- 223
NAC6_ARATH        ADQLEMMLEGQEDREQKEE---------------------------------------- 172
```

Amino acid sequence of *Avicennia marina* NAC domain protein (A7LKM8_AVIMR): NAC domain protein consists of subdomains, A-E; subdomain A (in blue), subdomain B (in green), subdomain C (in brown), subdomain D (in magenta) and subdomain E (in orange) Each subdomain of NAC family has shown the variable distribution of charged amino acid residues (in red)

Table 2: Sequence patches that show structural deviation between *Avicennia marina* NAC protein (AMNAC) and that of *Oryza sativa* (salt-sensitive) and *Arabidopsis thaliana* (salt-tolerant) NAC proteins

Avicennia marina	***Arabidopsis thaliana***	***Oryza sativa***
Region 1 (residues 29-42)		
29Cys (N)-24 Val(O)	29 Cys (N) -24 Val (O)	26Cys (N) -21Val (O)
29Cys (O) - 33 Ala(N)	29 Cys(O) - 33 Ala (N)	27Arg (O) -30Ala (N)
29Cys(O) -32 Cys (N)	29 Cys (O) -32 Cys (N)	-
30Arg (N) - 25 Val (O)	30Arg(N) -25 Val(O)	-
-	-	27Arg(NE) -23His(O)
30Arg (O) - 34 Gly(N)	30Arg (O) - 34 Gly(N)	27Arg (O) -31 Gly (N)
30Arg (O) - 35Gln (N)	30Arg (O) - 35Gln (N)	27Arg (O) -32 Gln (N)
31Lys(NZ) – 44Glu (OE1)	31Lys (NZ) - 44Glu(OE1)	28Lys(NZ) -41Glu(OE1)
Region 2 (residues 73-83)		
-	-	-
Region 3 (residues 98-108)		
98Lys(NZ) –20Asp(OD2)	98 Lys(NZ) –20Asp(OD2)	-
Region 4 (residues 133-149)		
145Lys (N)-144Thr (OG1)	145Lys (N)-144Thr (OG1)	145Lys (N) - 143Asn (OD1)

Table 3: Sequence patches that contain the predicted post translational modification sites in *Avicennia marina* NAC protein as predicted by PROSITE. The predicted residue in the patch is denoted by asterisk (*). In case of glycosylation, phosphorylation and myristoylation, Asn, Ser (S) / Thr (T) and Gly (G) are post translationally modified respectively.

N-glycosylation site	**Protein kinase C phosphorylation site**	**Casein kinase II phosphorylation site**	**N-myristoylation site**
78-81 *NGSR	70-72 *SPR	19-22 *TDEE	34-39 *GQQIGV
		70-73 *SPRD	79-84 *GSRPNR
		94-97 *TGAD	107-112 *GIKKAL
			122-127 *GVKTNW

Fig. 1: Phylogenetic analysis of *Avicennia marina* NAC domain protein showing relatedness with NAC proteins belonging to (a) salt sensitive and (b) salt tolerant plants. Evolutionary tree was constructed from distance data by

using online Neighbor joining method and represented using the Drawgram program of the PHYLIP tree package version 367

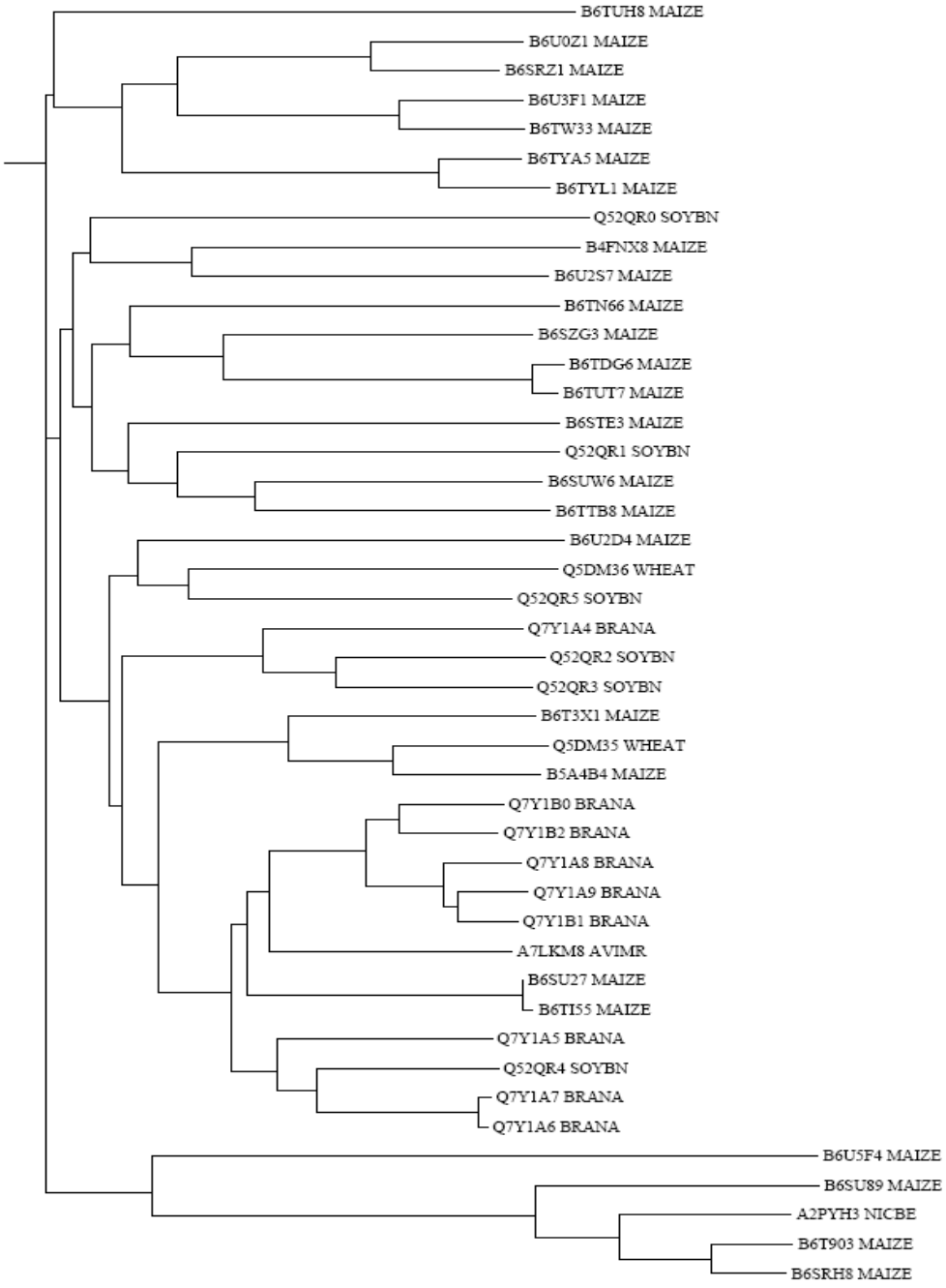


Fig. 1b

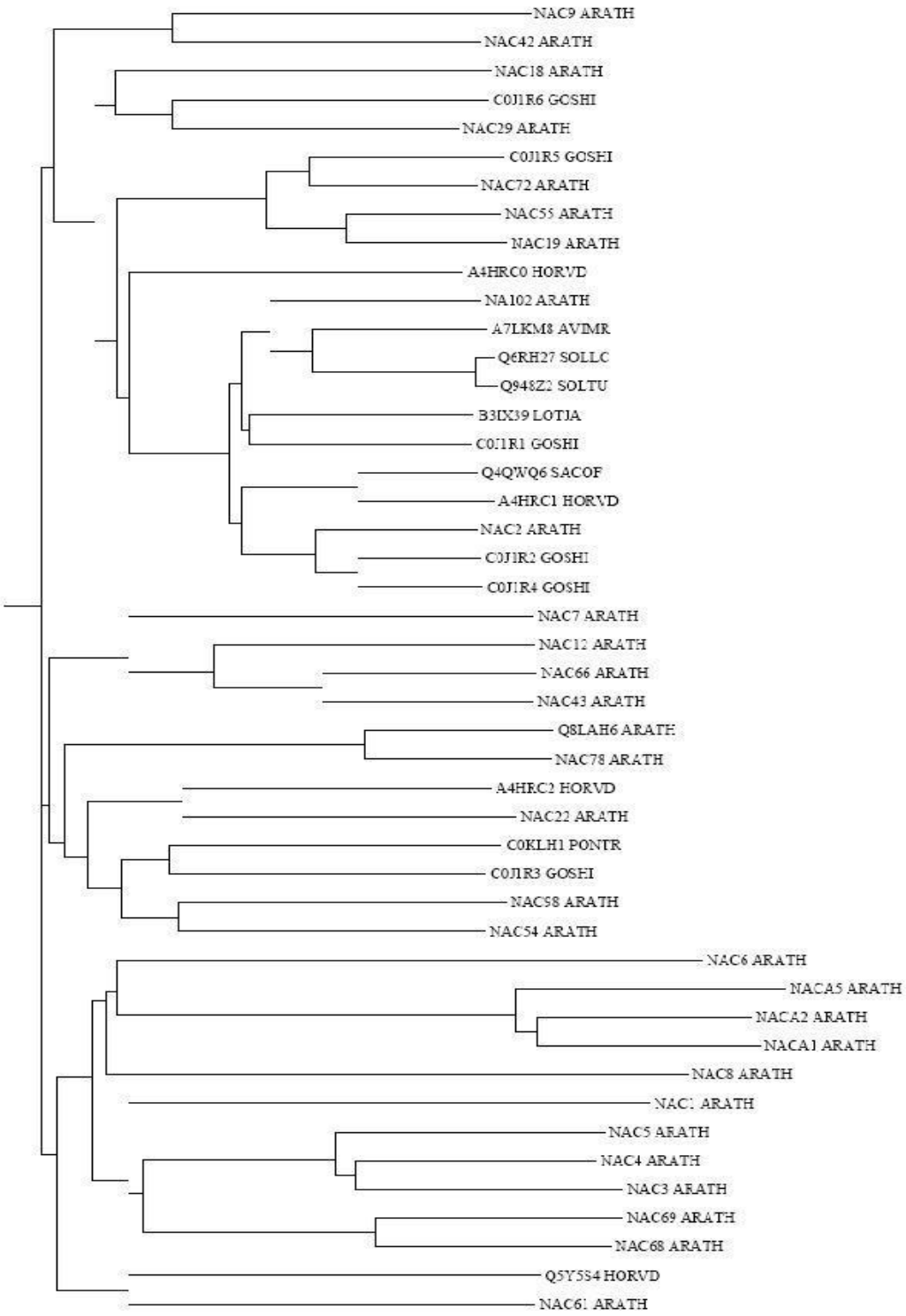

NAC9 ARATH
NAC42 ARATH
NAC18 ARATH
C0J1R6 GOSHI
NAC29 ARATH
C0J1R5 GOSHI
NAC72 ARATH
NAC55 ARATH
NAC19 ARATH
A4HRC0 HORVD
NA102 ARATH
A7LKM8 AVIMR
Q6RH27 SOLLC
Q948Z2 SOLTU
B3IX39 LOTJA
C0J1R1 GOSHI
Q4QWQ6 SACOF
A4HRC1 HORVD
NAC2 ARATH
C0J1R2 GOSHI
C0J1R4 GOSHI
NAC7 ARATH
NAC12 ARATH
NAC66 ARATH
NAC43 ARATH
Q8LAH6 ARATH
NAC78 ARATH
A4HRC2 HORVD
NAC22 ARATH
C0KLH1 PONTR
C0J1R3 GOSHI
NAC58 ARATH
NAC54 ARATH
NAC6 ARATH
NACA5 ARATH
NACA2 ARATH
NACA1 ARATH
NAC8 ARATH
NAC1 ARATH
NAC5 ARATH
NAC4 ARATH
NAC3 ARATH
NAC69 ARATH
NAC68 ARATH
Q5Y5S4 HORVD
NAC61 ARATH

Fig. 2: Pairwise sequence alignment of the target sequence, *Avicennia marina* NAC domain protein with 1ut4 using BLAST server.

```
pdb|1UT4|A Chain A, Structure Of The Conserved Domain Of Anac, A Member
Of The Nac Family Of Transcription Factors

Score =  237 bits (604),  Expect = 5e-63, Method: Compositional matrix adjust.
Identities = 113/161 (70%), Positives = 129/161 (80%), Gaps = 4/161 (2%)

Query  7    QLNLPAGFRFHPTDEELVVHYLCRKCAGQQIGVPVIAEIDLYKFDPWELPDLALYGEKEW  66
Sbjct  14   QLSLPPGFRFYPTDEELMVQYLCRKAAGYDFSLQLIAEIDLYKFDPWVLPNKALFGEKEW  73

Query  67   YFFSPRDRKYPNGSRPNRAAGTGYWKATGADKPVG-KPKTLGIKKALVFYAGKAPRGVKT  125
Sbjct  74   YFFSPRDRKYPNGSRPNRVAGSGYWKATGTDKIISTEGQRVGIKKALVFYIGKAPKGTKT  133

Query  126  NWIMHEYRLANVDRSAGKTKNLRLDDWVLCRIYNKKGTLEK  166
Sbjct  134  NWIMHEYRLIEPSRRNGSTK---LDDWVLCRIYKKQSSAQK  171
```

Fig. 3: Evaluation of homology model of *Avicennia marina* NAC domain (AMNAC) protein by (a) PROCHECK showing no residue in the disallowed region and (b) Protein Structure Analysis (ProSA) showing with z score of -5.62. ProSA plot shows local model quality by plotting energies as a function of amino acid sequence position. The thick line shows the average energy over each 40-residue fragment while the thin line shows a smaller window size of 10 residues.

(a)

```
 +----------<<<  P R O C H E C K    S U M M A R Y  >>>----------+
 |                                                               |
 | amnac   2.5                                      161 residues |
 |                                                               |
 | Ramachandran plot:   94.7% core    5.3% allow     .0% gener     .0% disall |
 |                                                               |
 | Gly & Pro Ramach:     0 labelled residues (out of 26)                      |
 | Chi1-chi2 plots:      0 labelled residues (out of 101)                     |
 |                                                               |
 | Main-chain params:    6 better     0 inside      0 worse                   |
 | Side-chain params:    5 better     0 inside      0 worse                   |
 |                                                               |
*| Residue properties: Max.deviation:     2.3             Bad contacts:    0 |
*|                     Bond len/angle:    6.5    Morris et al class:  1  1  2 |
+|     2 cis-peptides                                            |
 | G-factors           Dihedrals:    .00 Covalent:   -.24    Overall:   -.08 |
 |                                                               |
 | M/c bond lengths: 92.3% within limits  7.7% highlighted                    |
 | M/c bond angles:  82.3% within limits  17.7% highlighted                   |
 | Planar groups:   100.0% within limits    .0% highlighted                   |
 |                                                               |
 +---------------------------------------------------------------------------+
      + May be worth investigating further.  * Worth investigating further.
```

Fig. 3(b)

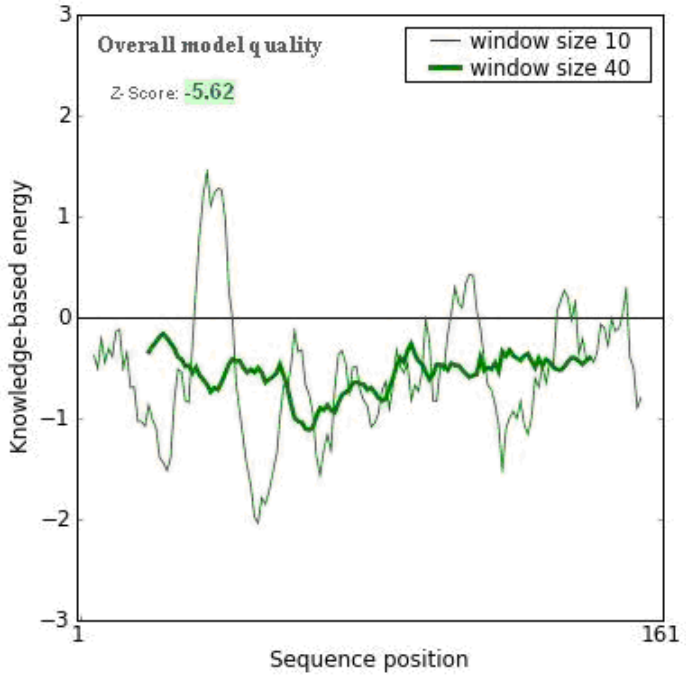


Fig. 4: (a) Superposition of the homology model of *Avicennia marina* NAC domain protein (AMNAC) with the respective template having root mean square deviation (RMSD) of 0.46Å; (b) Part of the structure of AMNAC that did not exactly superpose with the template with.

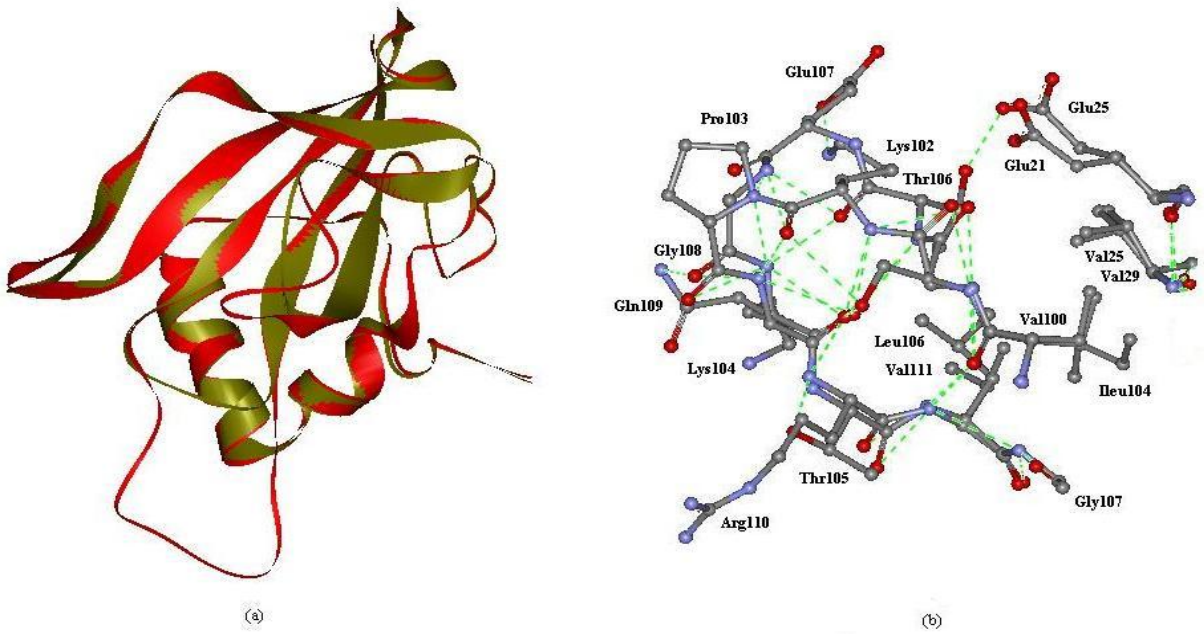


Fig. 4(c): different pattern of hydrogen bonding.

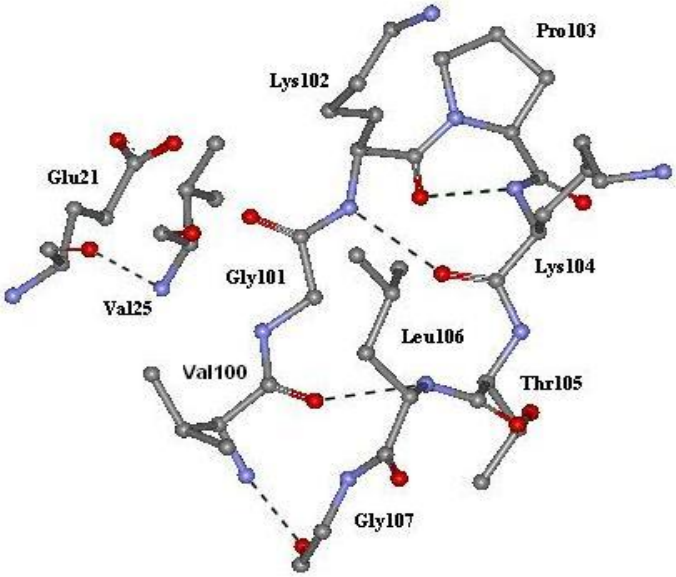


Fig. 5: Homology model *Avicennia marina* NAC protein (AMNAC) showing the surface representations of rich basic and less acidic charge distribution.

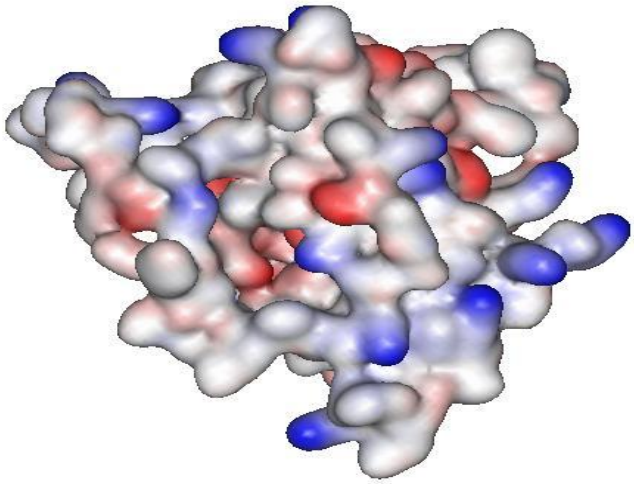


Fig. 6: Superposed models of *Avicennia marina* NAC protein (AMNAC) (blue) with NAC proteins belonging to salt sensitive *Oryza sativa* (red) and salt tolerant Arabidopsis thaliana (green) species. Regions labeled 1-4 did not superpose well due to the difference in amino acid residues.

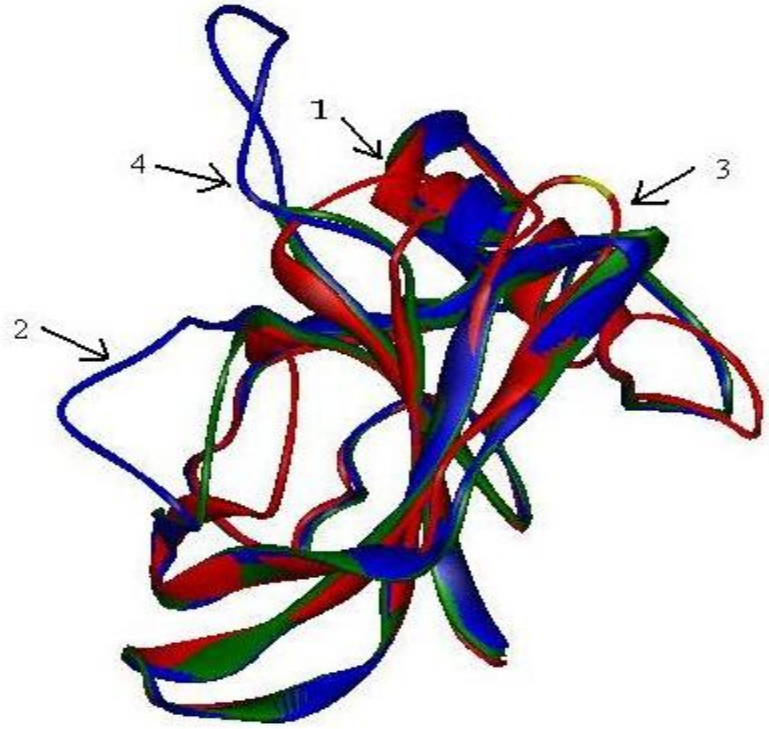
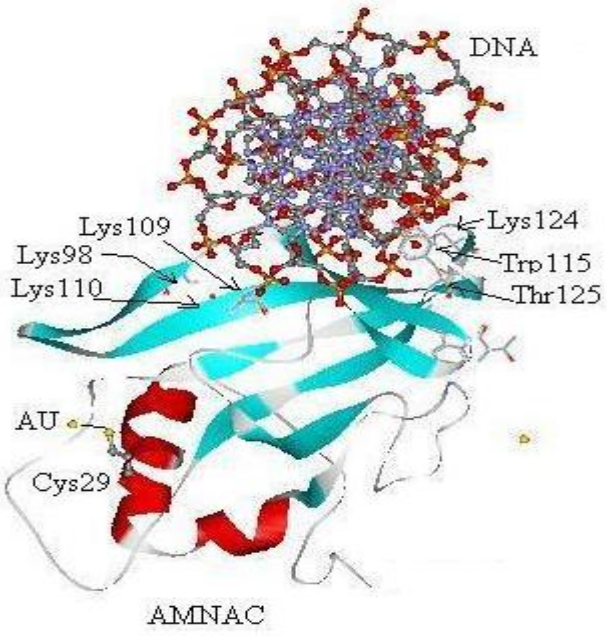


Fig. 7: Homology model *Avicennia marina* NAC protein (AMNAC) bound with DNA.



Conclusion

Plant cells have to survive constantly in the presence of various metabolites that are produced in response to various stress conditions. Mangroves that are salt-tolerant plants are better protected from oxidative damage under salt stress. They have evolved a complex series of enzymatic and non-enzymatic antioxidant protective mechanisms. Various genes responsible for providing salt tolerance have been identified including the genes involved in synthesis and uptake of various compatible solutes. *Avicennia marina* NAC-domain protein belongs to the NAC protein family, which is the largest family of plant transcription factors. Some members of these families play a significant role in the regulation of defense responses against plant stress i.e. high salinity, drought etc. It has also been demonstrated that transgenic *Arabidopsis* and *Oryza sativa* plant species over-expressing stress-responsive NAC genes have exhibited improved stress tolerance via biotechnological approaches. Members of these subfamilies are involved in the biological processes of regulation which is essential for plant development, transcription regulation and regulatory pathways involving protein-protein interactions. It can therefore be predicted that members of one subgroup of NAC-domain proteins may also have similar roles in the development and in the response of plant to environmental stress. The present study regarding the sequence and structural analysis of *Avicennia marina* NAC protein outlines their structural and functional aspects which may further assist in understanding their roles in resistance to salt stress**.**

References

1. Duval M, Hsieh T-F, Kim SY Thomas TL. Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. Plant Mol Biol 2008; 50:237-248

2. Kikuchi K, Ueguchi-Tanaka M, Yoshida KT, Nagato Y, Matsusoka M, Hirano H-Y. Molecular analysis of the NAC gene family in rice. Mol Gen Genet 2000; 262: 1047-1051

3. Riechemann JL Ratcliffe OJ. A genomic perspective on plant transcription factors. Curr Opin Plant Biol 2000; 3: 423-434

4. Yamasaki K, Kigawa T, Inoue M, Watanabe S, Tateno M, Seki M, Shinozaki K, Yokoyama S. Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. Plant Physiology and Biochemistry 2008; 46:394-401

5. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C-Z, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G-L. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 2000; 290:2105-2110

6. Olsen AN, Ernst HA, Lo Leggio L, Skriver K. NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci 2005; 10:79-87

7. Ooka H, Satoh K, Doi K, Nagata T, Otomo Y Murakami K, Matsubara K, Osato N, Kawai J, Carninci P, Hayashizaki Y, Suzuki K, Kojima K, Takahara Y, Yamamoto K, Kikuchi S. Comprehensive analysis of NAC family genes in Oryza sativa and Arabidopsis thaliana. DNA Res 2003; 10**:** 239-247

8. Souer E, Van AH, Kloos D, Mol J, Koes R. The No Apical Meristem gene of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell, 1996; 85: 159-170

9. Sablowski RWM, Meyerowitz EM. A homolog of no apical meristem is an immediate target of the floral homeotic genes. APETAL3. Cell 1998; 92: 93-103

10. Xie Q, Frugis G, Colgan D, Chua NH. Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev 2000; 14:3024-3036

11. Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme Takagi M. NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis. Plant Cell 2007; 19:270-280

12. Bu QY, Jiang HL, Li CB, Zhai QZ, Zhang J, Wu XY, Sun JQ, Xie Q, Li CY, Role of the Arabidopsis thaliana NAC transcription factors AN AC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. Cell Res 2008; 18: 756-767

13. Fujita M, Fujita Y, Maruyama K, Seki M, Hiratsu K, Ohme-Takagi M, Tran LS, Yamaguchi-Shinozaki K, Shinozakil K. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. Plant J 2004; 39: 863-876

14. Hu H, Dai MQ, Yao JL, Xiao BZ, Li XH, Zhang QF, Xiong LZ. Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resis tance and salt tolerance in rice. Proc Natl Acad Sci USA 2006; 103:12987-12992

15. Hu H, You J, Fang YJ, Zhu XY, Qi ZY, Xiong LZ. Characterization of transcription factor gene SNAC2 conferring cold and salt tolerance in rice. Plant Mol Biol 2008; 67: 169-181

16. Zheng XN, Chen B, Lu GJ, Han B. Overexpression of a NAC transcription factor enhances rice drought and salt tolerance. Biochem Biophys Res Commun 2009; 379:985-989

17. Hegedus D, Yu M, Baldwin D, Gruber M, Sharpe A, Parkin I, Whitwill S, Lydiate D. Molecular characterization of *Brassica napus* NAC domain transcriptional activators induced in response to biotic and abiotic stress. Plant Mol Biol 2003; 53: 383-397.

18. Ren T, Qu F, Morris TJ. HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus. Plant Cell 2000; 12:1917-1925

19. Ernst HA, Olsen AN, Skriver K, Larsen S, Lo Leggio L. Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors, EMBO Rep 2004; 5: 297-303

20. Welner DH, Lindemose S, Grossmann JG, Møllegaard NE, Olsen AN, Helgstrand C, Skriver K, Lo Leggio L. DNA binding by the plant-specific NAC transcription factors in crystal and solution: a firm link to WRKY and GCM transcription factors. Biochem J 2012; 444:395-404

21. Addie NO, Heidi AE, Leila LL, Karen S. DNA-binding specificity and molecular functions of NAC transcription factors. Plant Science 2005; 169:785-797

22. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. The Universal Protein Resource (UniProt). Nucleic Acids Res 2005; 33: D154-159

23. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BTLAS and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25: 3389-3402

24. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2002; 58:899-907

25. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000; 16:404-405

26. Olsen AN, Ernst HA, Lo Leggio L, Johansson E, Larsen S, Skriver K. Preliminary crystallographic analysis of the NAC domain of ANAC, a member of the plant-specific NAC transcription factor family. Acta Crystallogr D Biol Crystallogr 2004; 60 (PT 1):112-115

27. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997; 25: 4876-4882

28. Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 367). Cladistics 1889; 5: 164-166

29. Howe K, Bateman A, Durbin R. QuickTree: building huge Neighbour-Joining trees of protein sequences. Bioinformatics 2002; 18: 1546-1547

30. Obenauer JC, Cantley LC, Yaffe MB. Scansite 20: proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res 2003; 31:3635-3641

31. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993; 234: 779-815.

32. http://salilab.org/modeller/

33. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 1997; 18: 2714-2723 ; http://wwwexpasyorg/spdbv/;)

34. http://accelrys.com/products/discovery-studio/visualization-download.php).

35. Laskowski RA, McAurthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Cryst 1993; 26: 283-291

36. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1003; 17: 355-362

37. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007; 35:W407–410

38. Hoppe T, Rape M, Jentsch S. Membrane-bound transcription factors: regulated release by RIP or RUP. Curr Opin Cell Biol 2001; 13:344-348

39. Kim HS, Park BO, Yoo JH… Jung MS, Lee SM, Han HJ, Kim KE, Kim SH, Lim CO, Yun DJ Chung WS. Identification of a calmodulin-binding NAC protein (CBNAC) as a transcriptional repressor in Arabidopsis. J Biol Chem 2007; 282: 36292-362302

40. Bause E, Hettkamp H, Legler G. Conformational aspects of N-glycosylation of proteins Studies with linear and cyclic peptides as probes. Biochem J 1982; 203:761-768

41. Marshall RD.Glycoproteins. Ann Rev Biochem 1972; 41: 673-702

42. Aubert JP, Helbecque N, Loucheux-Lefebvre MH. Circular dichroism studies of synthetic Asn-X-Ser/Thr-containing peptides: structure- glycosylation relationship. Arch Biochem Biophys 1981; 208: 20-29.

43. Meggio F, Pinna LA. One-thousand-and-one substrates of protein kinase CK2? FASEB J 2003; 17:349-368

44. Ogiso E**,** Takahashi Y**,** Sasaki T**,** Yano M**,** Izawa T**.** The Role of Casein Kinase II in Flowering Time Regulation Has Diversified during Evolution. Plant Physiology 2010; 152: 808-820

45. Olsten ME, Litchfield DW. Order or chaos? An evaluation of the regulation of protein kinase CK2. Biochem Cell Biol 2004; 82:681-693

46. Krick R, Aschrafi A, Hasgun D, Arnemann J. CK2-dependent C-terminal phosphorylation at T300 directs the nuclear transport of TSPY protein. Biochem Biophys Res Commun 2006; 341:343-350

47. Miyata Y, Nishida E. CK2 controls multiple protein kinases by phosphorylating a kinase-targeting molecular chaperone. Mol Cell Biol 2004; 24: 4065-4074

48. Riera M, Figueras M, Lopez C, Goday A, Pages M. Protein kinase CK2 modulates developmental functions of the abscisic acid responsive protein Rab17 from maize. Proc Natl Acad Sci USA 2004; 101: 9879-9884

49. Taghli-Lamallem O, Hsia C, Ronshaugen M, McGinnis W. Context dependent regulation of Hox protein functions by CK2 phosphorylation sites. Dev Genes Evol 2008; 218:321-332

50. Dempsey E C, Newton A C, Mochly-Rosen D, Fields A P, Reyland M E, Insel P A, Messing R O. Protein kinase C isozymes and the regulation of diverse cell responses. Am J Physiol 2000; 279: L429-L438