

Preliminary analysis of the vascular-related
NAC-domain 1 (VND 1) transcription factor

参与维管形成的转录因子VND1的初步分析



G02 : 孟 强
解振诚
石 磊
黄鹏辉



中国农业科学院
Chinese Academy of Agricultural Sciences

2017年12月13号

CONTENT

01 Background



02 Analyzed by Uniport and NCBI

- Sequence
- Publications
- Subcellular location and expression
- MEGA



03 Structure



- Secondary Structure
- 3D Structure



04 Acknowledgement



➤ Background

VND1是植物中与维管形成相关的转录因子，属于NAC转录因子家族中的一员。NAC (NAM/ATAF/CUC) 转录因子家族是植物特有的一类转录因子，在水稻、番茄、拟南芥中均有存在，也是目前发现的最大的转录因子家族之一。第一个NAC基因 (NAM) 是1996年Soure等在矮牵牛中克隆得到的，1997年Aida在拟南芥中发现具有类似功能的CUC2基因，NAM、ATAF、CUC编码的蛋白的N端都有一段保守的氨基酸序列，被命名为NAC结构域，包含该结构域的蛋白组成了NAC转录因子家族。该家族在植物的生长发育调控中发挥着重要作用，例如：调控顶端分生组织发育、细胞分化、子叶形成、侧根发生、花发育、激素信号传导、植物器官衰老、胚胎发育以及次生细胞壁形成和纤维的形成。此外，NAC转录因子还具有相应病原菌侵染和逆境胁迫应答的作用，当植物受到生物和非生物胁迫时，会诱导大量NAC转录因子成员的产生，从而调控植物进行胁迫响应



中国农业科学院

Chinese Academy of Agricultural Sciences

➤ Analyzed by NCBI

The screenshot shows the NCBI Nucleotide search interface. At the top, the search term 'vnd1' is entered in the search box, with a 'Search' button to its right. Below the search bar, there are options for 'Create alert' and 'Advanced'. The main content area displays search results for 'vnd1'. A summary box at the top of the results area states: 'See [VND1 vascular related NAC-domain protein 1](#) in the Gene database' and 'vnd1 reference sequences [Transcript \(3\)](#) [Protein \(3\)](#)'. Below this, there are three items listed:

- [Arabidopsis thaliana ecotype Landsberg erecta chromosome 2, whole genome shotgun sequence](#)
1. 19,037,554 bp linear DNA
Accession: CM004360.1 GI: 1032966057
[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- [Arabidopsis thaliana chromosome 2, whole genome shotgun sequence](#)
2. 19,037,554 bp linear DNA
Accession: LUHQ01000002.1 GI: 1032292779
[BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- [Arabidopsis thaliana vascular related NAC-domain protein 1 \(VND1\), mRNA](#)
3. 1,671 bp linear mRNA
Accession: NM_001335579.1 GI: 1063700651
[BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

On the left side of the page, there are various filters and options, including 'Species' (Animals (1), Plants (9)), 'Molecule types' (genomic DNA/RNA (5), mRNA (5)), 'Source databases' (INSDC (GenBank) (6), RefSeq (4)), 'Sequence length', 'Release date', 'Revision date', and 'Clear all'. On the right side, there are sections for 'Results by taxon' (listing Arabidopsis thaliana (7), Helianthus annuus (1), Musa AAB Group (1), Acropora millepora (1)), 'Analyze these sequences' (Run BLAST), 'Find related data' (Database: Select), and 'Search details' (vnd1 [All Fields]).



➤ Analyzed by BLAST

```
/db_xref="TAIR:AT2G18060"  
/translation="MEPMESCSVPPGFRFHPTDEELVGYLRKKIASQKIDLDVIRDI  
DLYRIEPLDQEQCRIGYEEQNEWYFFSHKDKKYPTGTRINRATMAGFWKATGRDKAV  
YDKTKLIGMRKTLVIFYKGRAPNGKSDWIMHEYRLESNDENAPPQEEGWVVCRAFVKRA  
TGQAKNTEIWSYFYDEVAPNGVNSVMDPIDYISKQQHNI FGKGLMCKQELEGMVDG  
INYIQSNQFIQLPQLQSPSLPLMKRPSSSMSITSMDDNNYKPLADEESFESFIRGE  
DRRKKKQVMMTGNWRELDKFKVASQLMSQEDNGTSSFAGHHIVNEDKNNNDVEMDSSM  
FLSEREEENRFVSEFLSTNSDYDIGICVFDN"
```

ORIGIN

```
1  tcttcaccac  ttctctcttc  ttcttaatta  agttgttctc  accagtcacc  accattgatc  
61  ttcttaggtt  ttctttttgt  ttccgtcaca  tactogatca  attataatta  tatattoggc  
121  tgcgaagatt  tacaagaaat  aatggagcca  atggaatctt  gtagcgttoc  tccaggattt  
181  aggttccatc  cgacggacga  agagcttgtc  gggtactatc  taaggaagaa  aatcgcatcg  
241  caaaagattg  atctcgacgt  catcagagac  atcgatctct  acagaataga  accatgggat  
301  ctacaagaac  aatgtogaat  cggttatgag  gaacaaaatg  aatggatttt  ttttagtcac  
361  aaggacaaga  aatatccaac  ggggacaaga  actaatagag  cgaccatggc  tggattttgg  
421  aaagccacgg  gaagagacaa  agctgtttac  gacaaaacaa  aactaattgg  tatgaggaaa  
481  acacttgtgt  tctacaaaag  acgtgcaact  aatggcaaga  aatccogattg  gatcatgcat  
541  gagtaccggc  tcgagtcaga  tgagaatgca  cggccccagg  aagaaggatg  ggtggtttgt  
601  agagcattca  aaaaaagagc  tacagggcaa  gccaaagaac  cggaaacttg  gagctcaagt  
661  tacttttacg  atgaagttgc  accgaatgga  gtttaactcg  ttatggaccc  cattgattac  
721  atatctaagc  agcaacataa  catttttggg  aaaggtttga  tgtgtaagca  agaactagaa  
781  ggaatggttg  atggtataaa  ctatatacaa  togaatcaat  tcattcagct  cccacaactc  
841  caaagccctt  ctctcccgct  gatgaaaaga  ccttcaagct  cgatgtccat  aacatcaatg  
901  gataacaatt  acaactataa  actcccatta  ggggatgaag  aaagcttoga  gtcattcata  
961  agaggagagg  atagaaggaa  gaagaaaaag  caagttaatg  tgacgggaaa  ttggagagag  
1021  ttagacaagt  ttgttgcttc  acaacttatg  agccaagaag  acaatggaac  ttcaagtttc  
1081  gcaggtcac  atatagttaa  tgaagataaa  aacaacaatg  atgtggagat  ggattcgtca  
1141  atgtttttga  gcgaaagaga  agaagaaaac  aggttcgtca  gtgaattctt  gagtacaaac  
1201  tcggattatg  atattgggat  ttgcgtatct  gataattgaa  tgaactaca  tagatggcat  
1261  gtgagatttg  ttggtacgtt  tacttacact  tatatgaaa  tattaacaat  ataactaata  
1321  aagaaaatta  aacatattgt  ttagccatgt  ctagttttat  gaaaagaatg  tgttaggggt  
1381  gtgaatgagt  gaggatataa  atctatatta  cgttacgtac  tatttatatt  cgagtttgag  
1441  tgactaggtc  agagttattg  ttaagttatg  agggatgaa  ttgtgaggtt  ttctaattt  
1501  tttaatata  gtttagttta  tagagatgat  ttaaggggoc  attaattaag  tatagttaga  
1561  aatcagaata  agaaatcatt  gttgtggggc  aaggagatgt  cgagcgtgog  gttgtcacaa  
1621  cggggctccg  ttgggcatgc  tttgtctgct  acaataccaa  cgagactttg  g
```

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query sequence.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) [Query subrange](#)

NM_001335579

Or, upload file 未选择文件.

Job Title [Enter a descriptive title for your BLAST search](#)

Align two or more sequences

Choose Search Set

Database Human genomic + transcript Mouse genomic + transcript Others (nr etc.):
Nucleotide collection (nr/nt)

Organism [Optional](#) Exclude [+](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude [Optional](#) Models (XM/XP) Uncultured/environmental sample sequences

Limit to [Optional](#) Sequences from type material

Entrez Query [Optional](#) [YouTube](#) [Create custom database](#)

Program Selection

Optimize for Highly similar sequences (megablast) More dissimilar sequences (discontiguous megablast)



中国农业科学院

Chinese Academy of Agricultural Sciences

➤ Analyzed by BLAST

Mouse over to see the title, click to show alignments



Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Arabidopsis thaliana vascular related NAC-domain protein 1 (VND1), mRNA	3086	3086	100%	0.0	100%	NM_001335579.1
<input type="checkbox"/> Arabidopsis thaliana vascular related NAC-domain protein 1 (VND1), partial mRNA	2174	2174	70%	0.0	100%	NM_127362.2
<input type="checkbox"/> Arabidopsis thaliana vascular related NAC-domain protein 1 (VND1), mRNA	2111	2679	86%	0.0	99%	NM_001335580.1
<input type="checkbox"/> Arabidopsis thaliana chromosome 2 sequence	2021	3103	100%	0.0	100%	CP002685.1
<input type="checkbox"/> Arabidopsis thaliana chromosome 2 clone T27K22 map c245, complete sequence	2021	3103	100%	0.0	100%	AC006201.4
<input type="checkbox"/> PREDICTED: Arabidopsis lyrata subsp. lyrata NAC domain-containing protein 37 (LOC9320174), mRNA	1810	1810	77%	0.0	92%	XM_002886099.2
<input type="checkbox"/> Eutrema salsugineum hypothetical protein (EUTSA_v10022756mq), mRNA, complete cds	1304	1304	71%	0.0	87%	XM_006409138.1
<input type="checkbox"/> PREDICTED: Brassica oleracea var. oleracea NAC domain-containing protein 37 (LOC106304367), mRNA	1225	1225	68%	0.0	86%	XM_013740775.1
<input type="checkbox"/> PREDICTED: Brassica rapa NAC domain-containing protein 37 (LOC103828433), mRNA	1203	1203	68%	0.0	86%	XM_009104034.2
<input type="checkbox"/> PREDICTED: Brassica napus NAC domain-containing protein 37-like (LOC106356176), mRNA	1201	1201	68%	0.0	86%	XM_013795985.2
<input type="checkbox"/> PREDICTED: Raphanus sativus NAC domain-containing protein 37 (LOC108809570), transcript variant X1, mRNA	1098	1098	68%	0.0	84%	XM_018581709.1
<input type="checkbox"/> PREDICTED: Raphanus sativus NAC domain-containing protein 37 (LOC108809570), transcript variant X2, mRNA	1086	1086	68%	0.0	84%	XM_018581710.1
<input type="checkbox"/> PREDICTED: Camelina sativa NAC domain-containing protein 76-like (LOC104729563), mRNA	747	747	57%	0.0	81%	XM_010448515.1
<input type="checkbox"/> Capsella rubella hypothetical protein (CARUB_v10004997mq), mRNA, partial cds	736	736	57%	0.0	81%	XM_006283818.1
<input type="checkbox"/> PREDICTED: Camelina sativa NAC domain-containing protein 76 (LOC104721153), mRNA	728	728	55%	0.0	81%	XM_010439067.1
<input type="checkbox"/> PREDICTED: Arabidopsis lyrata subsp. lyrata NAC domain-containing protein 76 (LOC9303104), mRNA	695	695	56%	0.0	80%	XM_021020091.1
<input type="checkbox"/> Arabidopsis thaliana NAC domain containing protein 76 (NAC076), mRNA	693	693	57%	0.0	80%	NM_119783.3
<input type="checkbox"/> Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSLTFB61ZG08 of Flowers and buds of strain col-0 of Arabidopsis thaliana (thale cress)	693	693	57%	0.0	80%	BX826769.1
<input type="checkbox"/> Arabidopsis thaliana NAC domain containing protein 76 (NAC076), mRNA	691	691	57%	0.0	80%	NM_001342407.1
<input type="checkbox"/> Arabidopsis thaliana NAC domain containing protein 76 (NAC076), mRNA	691	691	56%	0.0	80%	NM_001342406.1
<input type="checkbox"/> Arabis alpina genome assembly, chromosome:3	688	1334	90%	0.0	79%	LT669790.1



➤ Analyzed by Clustal Omega

Input form | Web services | Help & Documentation

Wise2DBA and Promoterwise are scheduled for retirement on 15th April 2018. Alternatives please contact us via [support](#)

Results for job clustalo-I20171212-081443-0325-32291771-p

Alignments | Result Summary | **Phylogenetic Tree** | Submission Details

Download Alignment File | Send to Simple_Phylogeny

CLUSTAL O (1.2.4) multiple sequence alignment

```
Capsella_rubella
Camelina_sativa
C_sativa
Arabidopsis_thaliana
Arabidopsis_lyrata
Raphanus_sativus
Brassica_oleracea
Brassica_rapa
Brassica_napus
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
```

```
Capsella_rubella
Camelina_sativa
C_sativa
Arabidopsis_thaliana
Arabidopsis_lyrata
Raphanus_sativus
Brassica_oleracea
Brassica_rapa
Brassica_napus
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
```

```
Capsella_rubella
Camelina_sativa
C_sativa
Arabidopsis_thaliana
Arabidopsis_lyrata
Raphanus_sativus
Brassica_oleracea
Brassica_rapa
Brassica_napus
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
```

```
Capsella_rubella
Camelina_sativa
C_sativa
Arabidopsis_thaliana
Arabidopsis_lyrata
Raphanus_sativus
Brassica_oleracea
Brassica_rapa
Brassica_napus
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
```

```
Capsella_rubella
Camelina_sativa
C_sativa
Arabidopsis_thaliana
Arabidopsis_lyrata
Raphanus_sativus
Brassica_oleracea
Brassica_rapa
Brassica_napus
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
```

Input form | Web services | Help & Documentation

please contact us via [support](#).

Results for job clustalo-I20171212-081443-0325-32291771-p2m

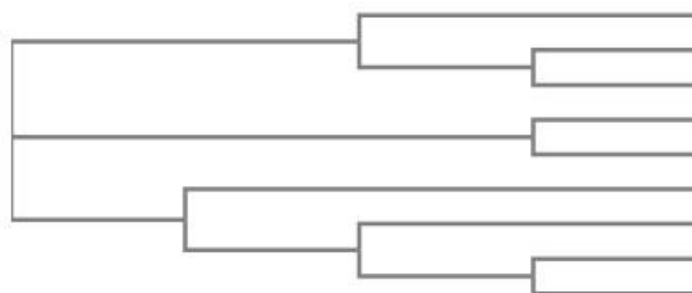
Alignments | Result Summary | **Phylogenetic Tree** | Submission Details

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Download Phylogenetic Tree Data

Branch length: Cladogram Real

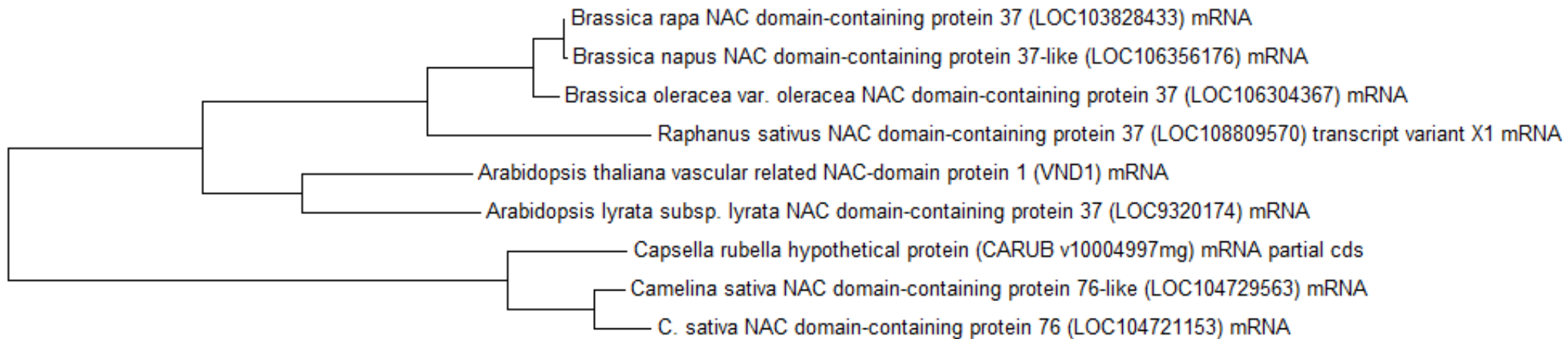


Capsella_rubella 0.03722
Camelina_sativa 0.01065
C_sativa 0.01417
Arabidopsis_thaliana 0.09066
Arabidopsis_lyrata 0.0923
Raphanus_sativus 0.06916
Brassica_oleracea 0.00865
Brassica_rapa 0.00248
Brassica_napus -0.00173

Tree Data

➤ Analyzed by MEGA

Species/Abbrv	Group Name	*****	***	*	**	*****	*****	**	**	**	*****	*	*****	*****	*****	*****	**	**	*****	*****
1. Arabidopsis thaliana vascular related NAC-domain protein 1		GG	AA	G	AAA	TCGCATCGCAAAAGATGATCTCGACGTCTCAGAGACATCGATCTCTTACAGAAATGAAACCTTGGGATCTTCAAGAGCAGT														
2. Arabidopsis lyrata subsp. lyrata NAC domain-containing prot		GG	AA	G	AAA	TCGCATCGCAAAAGATGATCTCGATGTTGATCAGAGACATTTGATCTCTTACAGAAATGAAACCTTGGGATCTTCAAGAGCAGT														
3. Brassica oleracea var. oleracea NAC domain-containing prote		GG	AA	G	AAA	GTCGCATCGCAAAAGATCGATCTCGATGTTCAATTAGAGAGCATCGATCTCTTACAAAAATTGAAACCTTGGGATCTTCAAGAGCAGT														
4. Brassica rapa NAC domain-containing protein 37 (LOC10382843		GG	AA	G	AAA	GTCGCATCGCAAAAGATCGATCTCGATGTTCAATTAGAGAGCATCGATCTCTTACAAAAATTGAAACCTTGGGATCTTCAAGAGCAGT														
5. Brassica napus NAC domain-containing protein 37-like (LOC10		GG	AA	G	AAA	GTCGCATCGCAAAAGATCGATCTCGATGTTCAATTAGAGAGCATCGATCTCTTACAAAAATTGAAACCTTGGGATCTTCAAGAGCAGT														
6. Raphanus sativus NAC domain-containing protein 37 (LOC10880		GG	AA	G	AAA	GTCGCTCGCAAAAGATCGATCTCGATGTTCAATTAGAGAGCATCGATCTCTTACAAAAATTGAAACCTTGGGATCTTCAAGAGCAGT														
7. Camelina sativa NAC domain-containing protein 76-like (LOC1		GG	AA	G	AAA	GTCGCATCGCAAAAGATCGATCTTGTATGTTCAATAGAGAGCATCGATCTCTTACAGAAATGAAACCTTGGGATCTTCAAGAGAGAT														
8. Capsella rubella hypothetical protein (CARUB_v10004997mg)_m		GG	AA	G	AAA	GTITGCAATCGCAAAAGATCGATCTTGTATGTTCAATAGAGAGCATCGATCTCTTACAGAAATGAAACCTTGGGATCTTCAAGAGAGAT														
9. C. sativa NAC domain-containing protein 76 (LOC104721153)_m		GG	AA	G	AAA	GTCGCATCGCAAAAGATCGATCTTGTATGTTCAATAGAGAGCATCGATCTCTTACAGAAATGAAACCTTGGGATCTTCAAGAGAGAT														



0.05



➤ Analyzed by NEBcutter



NEBcutter V2.0

[Program Guide](#)[Help](#)[Comments](#)

This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E. coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. **The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.**

[What's new in V2.0](#)[Citing NEBcutter](#)

Local sequence file: Vnd1.fasta

GenBank number: [\[Browse GenBank\]](#)

or paste in your DNA sequence: *(plain or FASTA format)*

Standard sequences:
Plasmid vectors ▾
Viral + phage ▾

The sequence is: Linear Circular

Enzymes to use:

- NEB enzymes
- All commercially available specificities
- All specificities
- All + defined oligonucleotide sequences
- Only defined oligonucleotide sequences [\[define oligos\]](#)

Minimum ORF length to display: a. a.

Name of sequence: *(optional)*



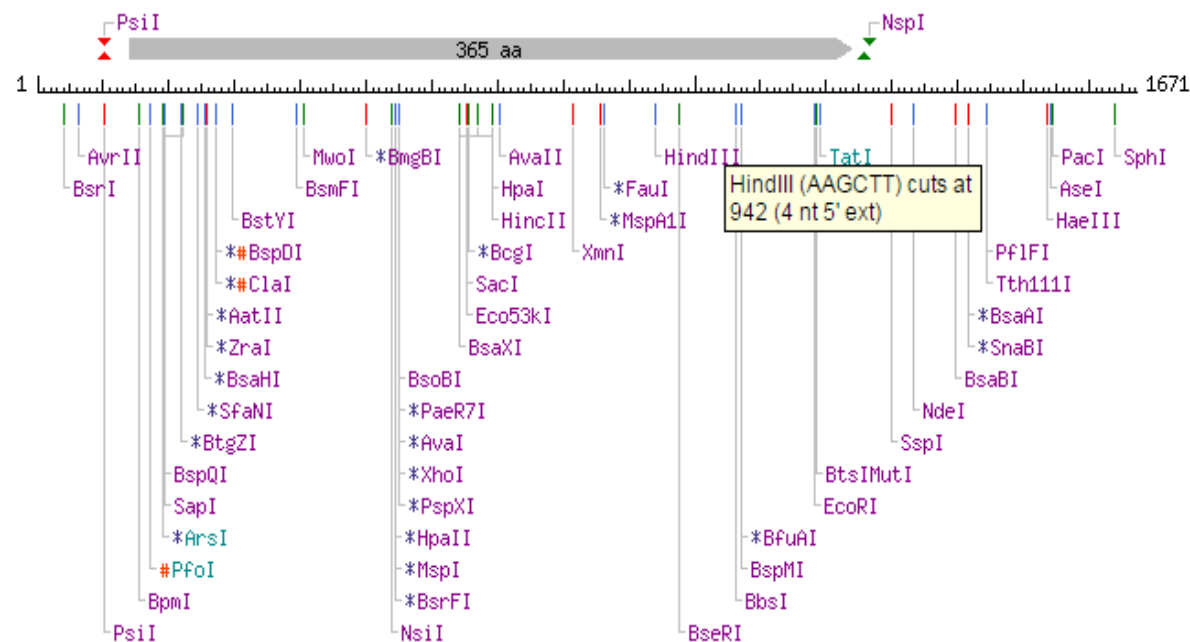
Linear Sequence: VndI

[Help](#)
[Comments](#)

Display: - All commercial single cutter restriction enzymes
 - Main non-overlapping, min. 100 aa ORFs

GC=39%, AT=61%

Cleavage code	Enzyme name code
✂ blunt end cut	Available from NEB
⏏ 5' extension	Has other supplier
⏏ 3' extension	Not commercially available
⏏ cuts 1 strand	*: cleavage affected by CpG meth.
	#: cleavage affected by other meth.
	(enz.name): ambiguous site



Main options	Availability	Display	Zoom	List
New DNA	NEB	2 cutters	Zoom in	0 cutters
Custom digest	All	3 cutters	More...	1 cutters
View sequence				All sites
ORF summary				Save all sites




[\[Back to main display\]](#)

Available from NEB, Catalog # R0104

[View product page](#)

HindIII-HF (High-Fidelity version) NEB Catalog # R3104

[View product page](#)

 Time-Saver Supplied with CutSmart buffer.

Buffer name: NEBuffer 2.1

Salt: 50 mM NaCl

Main: 10 mM Tris-HCl

pH: 7.9

Mg: 10 mM MgCl₂

BSA: 100

Reaction temperature: 37 °C

Neoschizomers:

NONE

Sites in sequence: 1

End produced at 942:

HindIII

5'... A^vA G C T T ... 3'
3'... T T C G A_A ... 5'

[Help](#)

[Comments](#)

[REBASE enzyme page](#)

[Methylation Sensitivity](#)

Overlapping methylation:

NONE

Isoschizomers:

NONE

Enzymes producing compatible ends:

Those cutting VndI:

➤ Analyzed by BioEdit

BioEdit Sequence Alignment Editor

File Edit **Sequence** Alignment View Accessory Application RNA World Wide Web Options Window Help

D:\Bac

Mode: Sele

Arabid

- New Sequence
- Edit Sequence
- Edit all selected
- Select Positions
- Open at cursor position
- Extract Positions
- Go to Pubmed references
- Overwrite/Retrieve sequences by gi number in title by HTTP to GenBank
- Overwrite/Retrieve Genbank data by gi number in title by HTTP to GenBank (ignore sequence)
- True positions from alignment positions
- Phylogeny / Taxonomy
- Filter out sequences containing certain characters
- Rename
- Sort
- PCR Primers / oligos
- Pairwise alignment
- Similarity Matrix (for pairwise alignments and shading)
- Features
- Sequence groups (or families)
- Edit Mode
- Mask
- Toggle Color
- Gaps
- Manipulations
- Nucleic Acid**
- Protein
- Translate or Reverse-Translate (permanent)
- Translate in selected frame (permanent)

fast

70 80 90 100 110 120 130

tttctttttgtttccggtcacatactcgatcaattataaattatatttcggctgccaagattttacaa

- Nucleotide Composition
- Base composition and mass export (monoisotopic)
- Base composition and mass export with average masses
- Complement
- Reverse Complement Shift+Ctrl+R
- DNA->RNA
- RNA->DNA
- Translate
- Find next ORF
- Find ORFs from a list of positions
- Create Plasmid from Sequence
- Gap beginning to minimize stop codons in reading frame 1 codon search
- Restriction Map**
- Sorted Six-Frame Translation
- Unsorted Six-Frame Translation



Create Restriction Map

Display:

<input checked="" type="checkbox"/> Display Map	<input checked="" type="checkbox"/> Sites that cut five or fewer times	<input checked="" type="checkbox"/> large recognition sites (>6)
<input checked="" type="checkbox"/> Alphabetical by Name	<input type="checkbox"/> Summary table of frequencies	<input type="checkbox"/> All Isoschizomers (not recommended)
<input checked="" type="checkbox"/> Numeric by position	<input checked="" type="checkbox"/> Summary of enzymes that do not cut	Manufacturer:
<input type="checkbox"/> List of unique sites	<input type="checkbox"/> 4-base cutters	All Enzymes <input type="button" value="→"/>
<input type="checkbox"/> Fasta list of fragments	<input type="checkbox"/> 5-base cutters	<input type="button" value="View by Manufacturer"/> <input type="button" value="Select from list"/>
<input type="checkbox"/> Create feature map	<input checked="" type="checkbox"/> 6-base cutters	
	<input checked="" type="checkbox"/> Enzymes with degenerate recognition	

Title: Arabidopsis thaliana vascular related NAC-domain protein 1 (VND1), mRNA Restriction Map

Display Translations:	Top Strand	Bottom Strand	<input type="checkbox"/> Circular DNA (ends joined)
	<input type="checkbox"/> Frame 1	<input type="checkbox"/> Frame -1	<input type="button" value="Generate Map"/> <input type="button" value="Cancel"/>
	<input type="checkbox"/> Frame 2	<input type="checkbox"/> Frame -2	<input type="button" value="View All Currently Selected Enzymes"/>
	<input type="checkbox"/> Frame 3	<input type="checkbox"/> Frame -3	

tcttcaccacttctctcttctttctaattaagttggttctcaccagtcaccaccattgatcttcctagggtttctttttggttccgtcacatact

1671 base pairs

Translations: none

Restriction Enzyme Map:

```
1      TCTTCACCACTTCTCTCTTCTTTCTAATTAAGTTGTTCTCACCAGTCACCACCATTGATCTTCCTAGGTTTTCTTTTTGT      80
1      AGAAGTGGTGAAGAGAGAAGAAAGATTAATTCAACAAGAGTGGTCAGTGGTGGTAACTAGAAGGATCCAAAAGAAAAACA      80
           MboII      EarI      HphI      HphI      MboII      AvrII      TspGWI
                               BsrI
                                   BsaJI
                                       StyI

81     TTCGGTCACATACTCGATCAATTATAATTATATATTTCGGCTGCGAAGATTTACAAGAAATAATGGAGCCAATGGAATCTT      160
81     AAGGCAGTGTATGAGCTAGTTAATATTAATATATAAGCCGACGCTTCTAAATGTTCTTTATTACCTCGGTTACCTTAGAA      160
           PsiI      MboII      NlaIV      BpmI
           BbvI
                                   Eco57MI

161    GTAGCGTTCCTCCAGGATTTAGGTTCCATCCGACGGACGAAGAGCTTGTTCGGGTACTATCTAAGGAAGAAAATCGCATCG      240
161    CATCGCAAGGAGGTCCATAAATCCAAGGTAGGCTGCCTGCTTCTCGAACAGCCCATGATAGATTCTTCTTTTAGCGTAGC      240
           FokI      MnlI      BstF5I      EarI      TspGWI      MboII
           NlaIV      BslI      MboII
           SapI      MmeI

241    CAAAAGATTGATCTCGACGTCATCAGAGACATCGATCTCTACAGAATAGAACCATGGGATCTACAAGAACAATGTCGAAT      320
241    GTTTTCTAACTAGAGCTGCAGTAGTCTCTGTAGCTAGAGATGTTCTTATCTTGGTACCCTAGATGTTCTTGTTACAGCTTA      320
           SfaNI      Hpy188III      ClaI      SfcI      BsaJI      AlwI
           BsaHI      BtgI
           ZraI      NcoI
           AatII      StyI
           BsmAI      BstYI
```

Restriction table:

Enzyme	Recognition	frequency	Positions
AatII	G_ACGT'C	1	261
AlwI	GGATCnnnn'n_	2	306, 539
ApaLI	G'TGCA_C	2	504, 1614
ApoI	r'AATT_y	2	1184, 1496
AseI	AT'IA_AI	1	1543
AvaI	C'yCGr_G	1	551
AvrII	C'CTAG_G	1	64
BanII	G_rGCy'C	2	656, 1628
BbsI	GAAGACnn'nnnn_	1	1065
BbvI	GCAGCnnnnnnnn'n'nnnn_	2	107, 742
BcgI	CGAnnnnnnnTGCnnnnnnnnnn'nn'	1	693
BcgI	GCAnnnnnnnTCGnnnnnnnnnn'nn'	1	659
BfrBI	ATG'CAT	1	538
Bme1580I	G_kGCm'C	2	508, 1618
BmgBI	CAC'GIC	1	503
BpmI	CTGGAGnnnnnnnnnnnnnn'nn'	1	156
BpuEI	CTTGAGnnnnnnnnnnnnnn'nn'	2	640, 1210
BsaAI	yAC'GTr	1	1417
BsaBI	GATnn'nnATC	1	1399
BsaHI	Gr'CG_yC	1	258
BsaJI	C'CnnG_G	5	64, 293, 405, 426, 576
BsaXI	ACnnnnnnCTCCnnnnnnnn'nnn'	1	643
BsaXI	GGAGnnnnnnGTnnnnnnnnnn'nnn'	1	673
BseRI	GAGGAGnnnnnnnn'nn'	1	978
BsiHKAI	G_wGCw'C	3	508, 656, 1618
BslI	CCnn'nnn'nnGG	2	193, 931
BsmI	GAATG_Cn'	2	571, 605
BspMI	CTCTCn'nnnn'	2	261, 429, 1657

- File
 - Edit
 - View
 - Function
 - Translate
 - Window
 - Help
- New
 - DNA Sequence
 - Protein Sequence
 - Open
 - Insert Sequence
 - Save (Ctrl+S)
 - Save As
 - Preferences (Ctrl+R)
 - Degenerate Bases (Ctrl+B)
 - Print
 - Quit
- 1 D:\Program Files (x86)\BioSoft\Primer Premier 5\data\DemoTNF.ALN
2 C:\Program Files\Primer Premier 5\data\DemoTNF.ALN

Translation:

Active Sequence:

Translations:

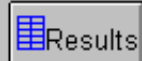
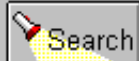
DNA Protein

5' Seq No Header 3 10 Find Find Next S A dsDNA

1

Pos: 00001

Primer:



Direct Select:

3' AGAAGTGGTGAAGAGAGAAGAAAGA 5'
 |||||
 5' TCITCACCACCTCTCTCTTCTTCTAATTAAGTTGTTCTCACCAGTCACCACCATTGATCTTCCTAGGTTTTCTTTTTGTTT 3'

S S P L L S S F - L S C S H Q S P P L I

	Rating	Seq No	Length	Tm [°C]	GC%	Δ G [kcal/mol]	Activi [μg/OD]
Sense	83	1	25	58.1	40.0	-41.6	37.0
Anti-sense	83	25	25	58.1	40.0	-41.6	27.9
Product	28	--	25	65.5	40.0	--	--

	Hairpin	Dimer	False Priming	Cross Dimer	No Hairpins Found
Sense	None	None	Found	Found	
Anti-sense	None	None	Found		

Search Criteria

Search For:

 PCR Primers
 Sequencing Primers
 Hybridization Probes

Search Type:

 Sense Primer
 Compatible with Sense Primer
 Anti-sense Primer
 Compatible with Anti-sense Primer
 Both
 Pairs

Search Ranges:

Sense Primer:

1 to 1671

Anti-sense Primer:

1 to 1671

PCR Product Size:

100 bp to 500 bp

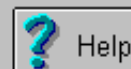
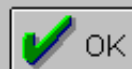
Primer Length:

25 bp ? 1 bp

Search Mode:

 Automatic
 Manual

Search Parameters



1321
 1381
 1441
 1501
 1561
 1621

Primer:

Direct Select:

3' CAGGTGTTATGGTTGCTCTGAAAC 5'

5' GATGTCGAGCGTGCGGTGTGCACAACGGGGCTCCGTTGGGCATGCTTTGTGCTCCACAATACCAACGAGACTTTGG 3'

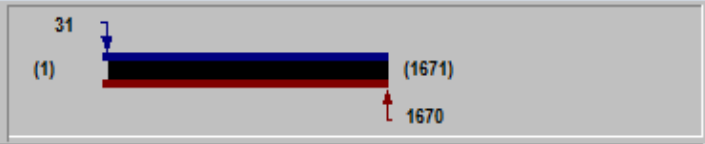
1600 1610 1620 1630 1640 1650 1660 1670

M S S V R C A Q R G S V G H A L S S T I P T R L W

	Rating	Seq No	Length	Tm [°C]	GC%	Δ G [kcal/mol]	Activity [μg/OD]	Degeneracy	Ta Opt [°C]
Sense	100	31	24	63.2	50.0	-42.4	32.5	1	--
Anti-sense	93	1670	24	60.6	45.8	-42.4	32.4	1	--
Product	82	--	1640	84.9	38.6	--	--	--	52.6

	Hairpin	Dimer	False Priming	Cross Dimer
Sense	None	None	None	Found
Anti-sense	Found	Found	None	

No Hairpins Found



NewSequence

Original DNA

DNA Protein

nd Next dsDNA

GAAAAC AGGTTGTCGA GTGAATTCTT GAGTACAAAC
 GTATTT GATAATTGAA TGAACTACA TAGATGGCAT
 TACACT TATATGAAAA TATTAACAAT ATAACTAATA
 1321 AAGAAAATTA AACATATGGT TTAGCCATGT CTAGTTTTAT GAAAAGAATG TGTTAGGGTT
 1381 GTGAATGAGT GAGGATATAA ATCTATATTA CGTTACGTAC TATTTATATT CGAGTTTGAG
 1441 TGA CTAGGTC AGAGTTATTG TTACGTTATG AGGGTATGAA TTGTGAGGTT TTCTAAATTT
 1501 TTTAATTATA GTTAGTTTTA TAGAGATGAT TTAAGGGGCC ATTAATTAAG TATAGTTAGA
 1561 AATCAGAATA AGAAATCATT GTTGTGGGGC AAGGAGATGT CGAGCGTGCG GTGTGCACAA
 1621 CGGGGCTCCG TTGGGCATGC TTTGTCGTCC ACAATACCAA CGAGACTTTG G

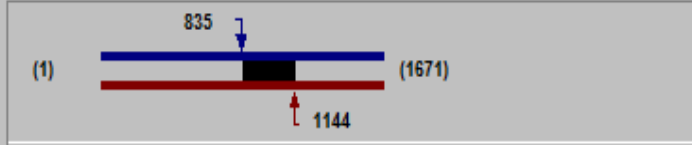
Pos: 01672

Sense Anti-sense Pairs

824 anti-sense primers found.

#	Rating	Seq No	Length	Tm [°C]	GC%	Mark
1	100	58	24	63.5	50.0	<input type="checkbox"/>
2	100	770	24	63.6	41.7	<input type="checkbox"/>
3	100	858	24	69.3	58.3	<input type="checkbox"/>
4	100	974	24	54.2	41.7	<input type="checkbox"/>
5	100	1144	24	62.3	41.7	<input type="checkbox"/>
6	100	1146	24	62.1	41.7	<input type="checkbox"/>
7	93	1226	24	61.6	37.5	<input type="checkbox"/>

Primer:



Direct Select:

3' TACACCTCTACCTAAGCAGTTACA 5'
 5' TGAAGATAAAAAACAACAATGATGTGGAGATGGATTGTCGAATGTT
 E D K N N N D V E M D S S M F

	Rating	Seq No	Length	Tm [°C]	GC%
Sense	100	835	24	69.3	58.3
Anti-sense	100	1144	24	62.3	41.7
Product	88	--	310	84.1	39.0

	Hairpin	Dimer	False Priming	Cross Dimer
Sense	None	None	None	None
Anti-sense	None	None	None	

Sense Anti-sense Pairs
 824 anti-sense primers found.

#	Rating	Seq No	Length	Tm [°C]	GC%	Mark
	100	58	24	63.5	50.0	<input type="checkbox"/>
	100	770	24	63.6	41.7	<input type="checkbox"/>
	100	858	24	69.3	58.3	<input type="checkbox"/>
	100	974	24	54.2	41.7	<input type="checkbox"/>
	100	1144	24	62.3	41.7	<input type="checkbox"/>
	100	1146	24	62.1	41.7	<input type="checkbox"/>
	93	1226	24	61.6	37.5	<input type="checkbox"/>

Edit Primer

HinFI

V E M D S S M

3' TACACCTCTACCTAAGCAGTTACA 5'

5' ATGTGGAGATGGATTGTCGAATGTTTTGAGCGAAAGAGAAGAAACAGGTTGTCAGTGAATTCCTTGGTACAAACT 3'

V E M D S S M F L S E R E E E N R F V S E F L S T N S

HinFI EcoRI

Rating	Seq No	Length	Tm [°C]	GC%	ΔG [kcal/mol]	Activity [μg/OD]	Degeneracy
100	1144	24	62.3	41.7	-42.8	31.7	1

Hairpin Dimer False Priming

None None None

No Hairpins Found

Analyze Enzyme Prime OK Cancel Help

➤ Structure

Primary Structure

>NP_179397.1 vascular related NAC-domain protein 1 [Arabidopsis thaliana]

MEPMESCSVPPGFRFHPTDEELVGYYLRKKIASQKID
LDVIRDIDLYRIEPWDLQEQRIGYEEQNEWYFFSHK
DKKYPTGTRTNRATMAGFWKATGRDKAVYDKTKLI
GMRKTLVIFYKGRAPNGKKSDWIMHEYRLESDENAP
PQEEGWVVCRAFKKRATGQAKNTETWSSSYFYDEV
APNGVNSVMDPIDYISKQQHNIFGKGLMCKQELEG
MVDGINYIQSNQFIQLPQLQSPSLPLMKRPSSSMISI
TSMDNNYNYKLPLADEESFESFIRGEDRRKKKKQVM
MTGNWRELDKFVASQLMSQEDNGTSSFAGHHIVN
EDKNNNDVEMDSSMFLSEREEENRFVSEFLSTNSDY
DIGICVFDN

Amino acid composition: [CSV format](#)

Ala (A)	14	3.8%
Arg (R)	20	5.5%
Asn (N)	22	6.0%
Asp (D)	26	7.1%
Cys (C)	5	1.4%
Gln (Q)	17	4.7%
Glu (E)	31	8.5%
Gly (G)	21	5.8%
His (H)	6	1.6%
Ile (I)	19	5.2%
Leu (L)	20	5.5%
Lys (K)	28	7.7%
Met (M)	16	4.4%
Phe (F)	18	4.9%
Pro (P)	16	4.4%
Ser (S)	31	8.5%
Thr (T)	15	4.1%
Trp (W)	7	1.9%
Tyr (Y)	16	4.4%
Val (V)	17	4.7%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Number of amino acids: 365

Molecular weight: 42476.67

Theoretical pI: 5.40

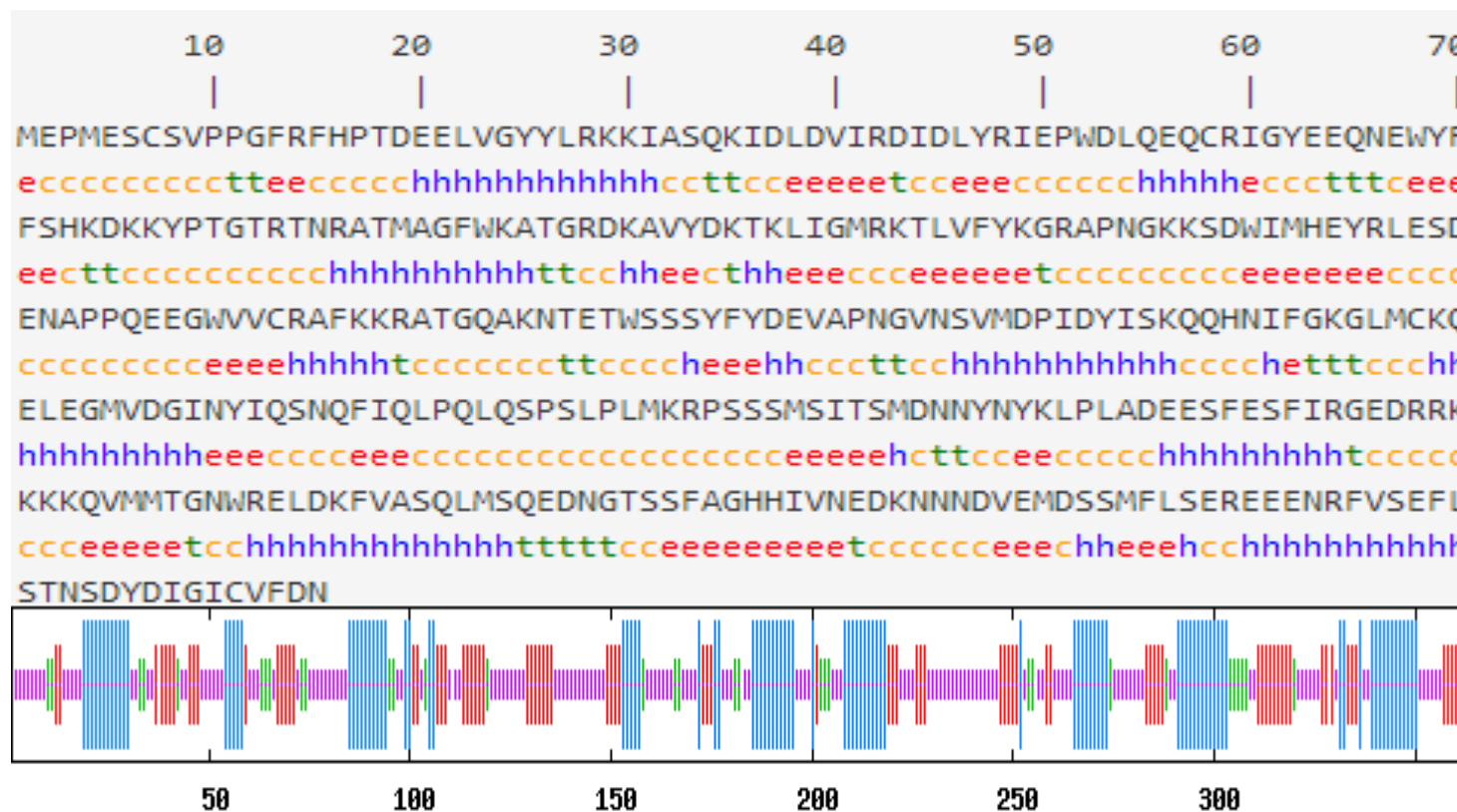
Total number of negatively charged residues (Asp + Glu): 57

Total number of positively charged residues (Arg + Lys): 48



➤ Structure

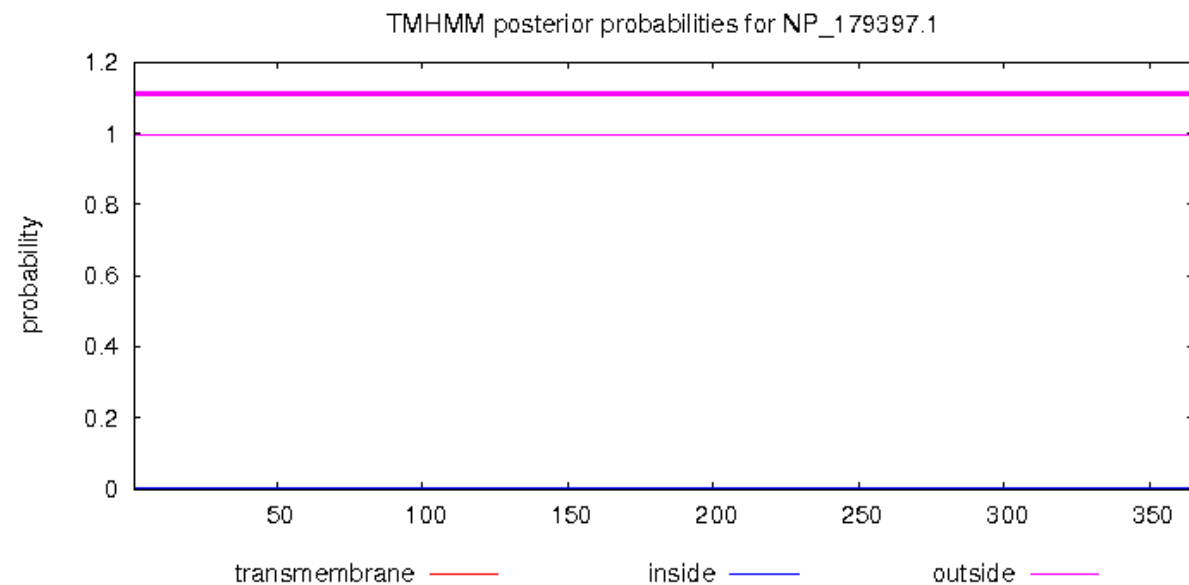
Secondary Structure



➤ Structure

Secondary Structure

```
# NP_179397.1 Length: 365
# NP_179397.1 Number of predicted TMHs: 0
# NP_179397.1 Exp number of AAs in TMHs: 0
# NP_179397.1 Exp number, first 60 AAs: 0
# NP_179397.1 Total prob of N-in: 0.00392
NP_179397.1 TMHMM2.0 outside 1 365
```



➤ Structure

Secondary Structure

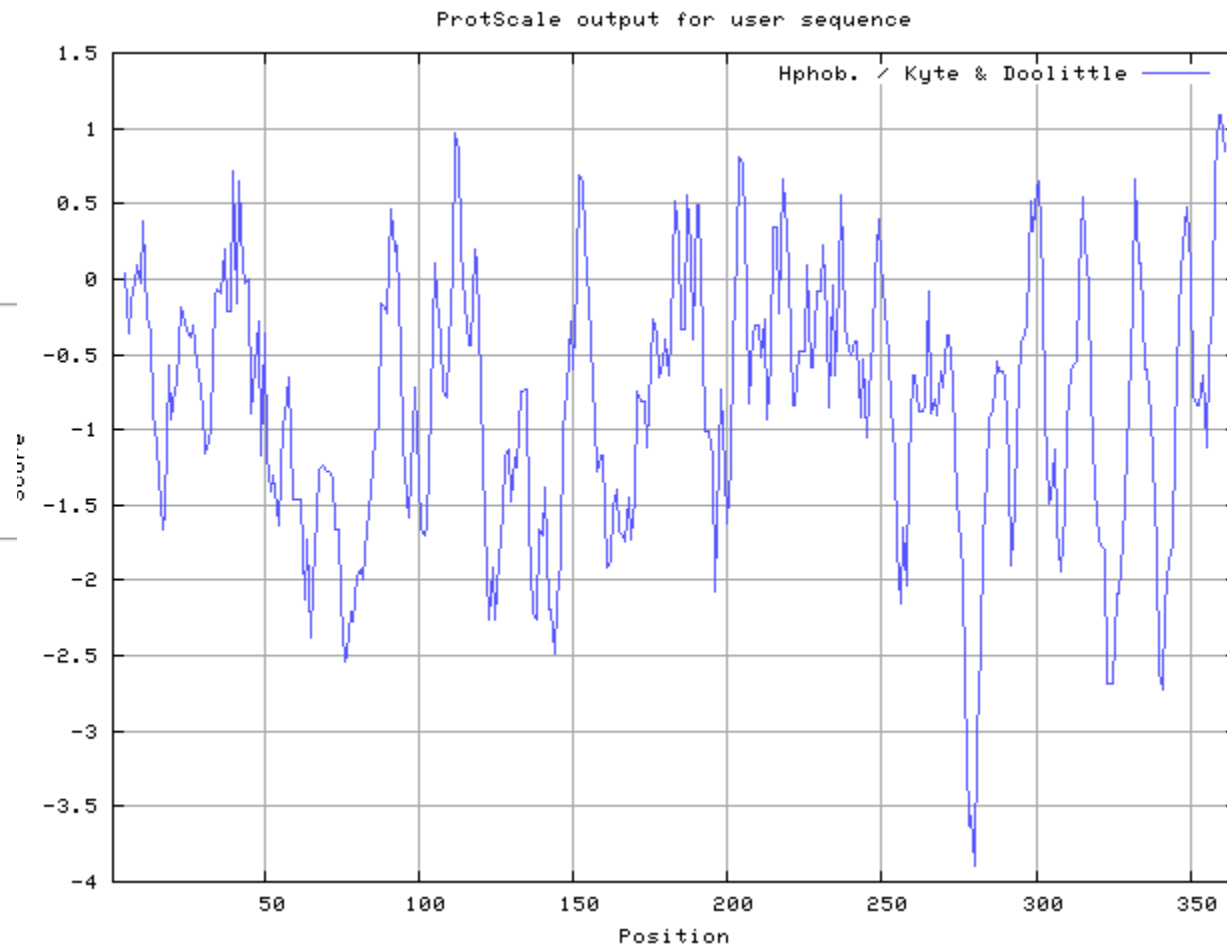
SEQUENCE LENGTH: 365

Using the scale **Hphob. / Kyte & Doolittle**, the individual values for the 20 amino acids are:

Ala: 1.800	Arg: -4.500	Asn: -3.500	Asp: -3.500	Cys: 2.500	Gln: -3.500
Glu: -3.500	Gly: -0.400	His: -3.200	Ile: 4.500	Leu: 3.800	Lys: -3.900
Met: 1.900	Phe: 2.800	Pro: -1.600	Ser: -0.800	Thr: -0.700	Trp: -0.900
Tyr: -1.300	Val: 4.200	: -3.500	: -3.500	: -0.490	

Weights for window positions 1,..,9, using linear weight variation model:

1	2	3	4	5	6	7	8	9
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
edge				center				edge



中国农业科学院
Chinese Academy of Agricultural Sciences

➤ Structure

Secondary Structure



Predicted localization for the Eukarya domain: Nucleus (GO term ID: [GO:0005634](#)) Prediction confidence 64

$k = 9/23$

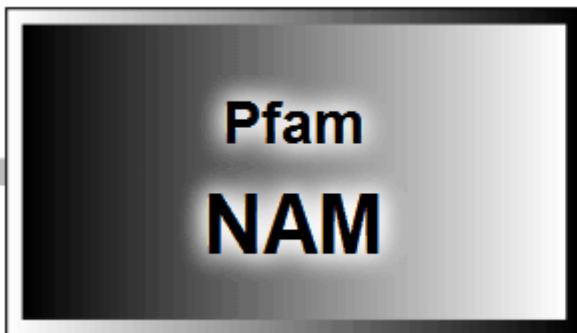
91.3 %: nuclear
4.3 %: cytoplasmic
4.3 %: cytoskeletal

>> prediction for 151289405818780 is nuc (k=23)



➤ Structure

Secondary Structure



DNA-binding domain (DBD) and dimerization domain

This is a *Pfam* domain. Please see the [NAM](#) entry in *Pfam* for full annotation.

Pfam domain sequence (128 aa):

[Submit to BLAST](#)

[Copy to clipboard](#)

```
PPGFRFHPTDEELVGYLRRKIASQKIDLDVIRDIDLYRIEPWDLQEQCRIGYEEQNEWY  
FFSHKDKKYPTGTRTNRATMAGFWKATGRDKAVYDKTKLIGMRKTLVIFYKGRAPNGKKS  
D  
WIMHEYRL
```

Position: 10 to 137
E-value: 1.1e-33 (*HMMER3*)

Accession: [PF02365](#)

Description:

Interpro abstract (IPR003441): *The NAC domain (for [Petunia hybrida](#) (*Petunia*) NAM and for *Arabidopsis ATAF1*, *ATAF2*, and ...[\(full abstract\)](#))*

GO process: *regulation of transcription, DNA-templated ([GO:0006355](#))*

GO function: *DNA binding ([GO:0003677](#))*

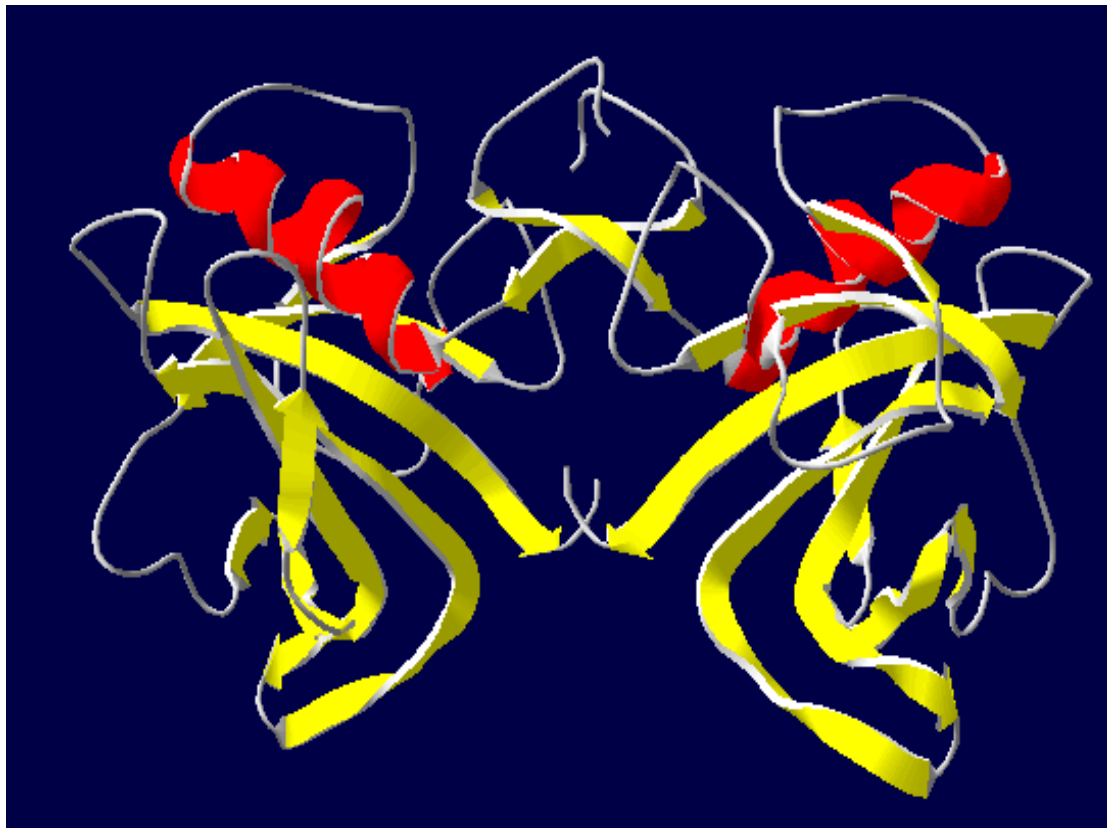
Pfam domain sequence (128 aa):



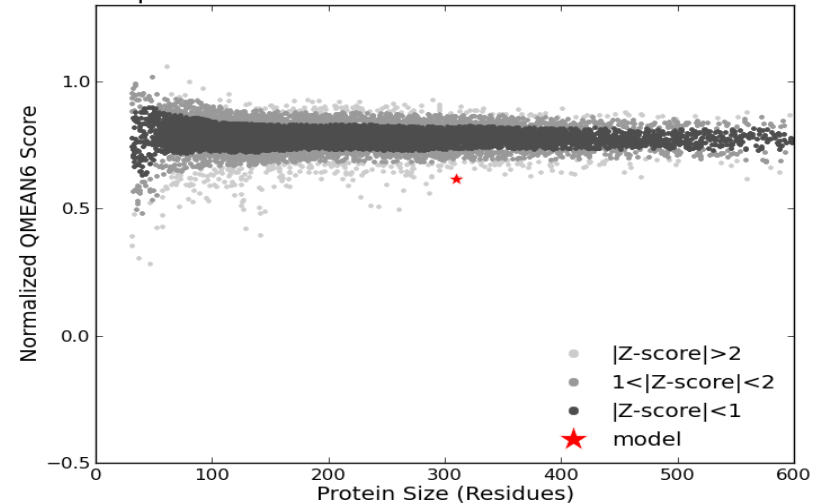
中国农业科学院
Chinese Academy of Agricultural Sciences

➤ Structure

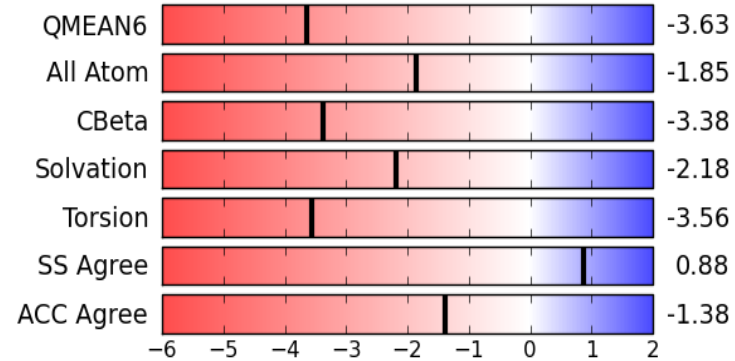
3D structure (Swiss Model)



Comparison with Non-redundant Set of PDB Structures

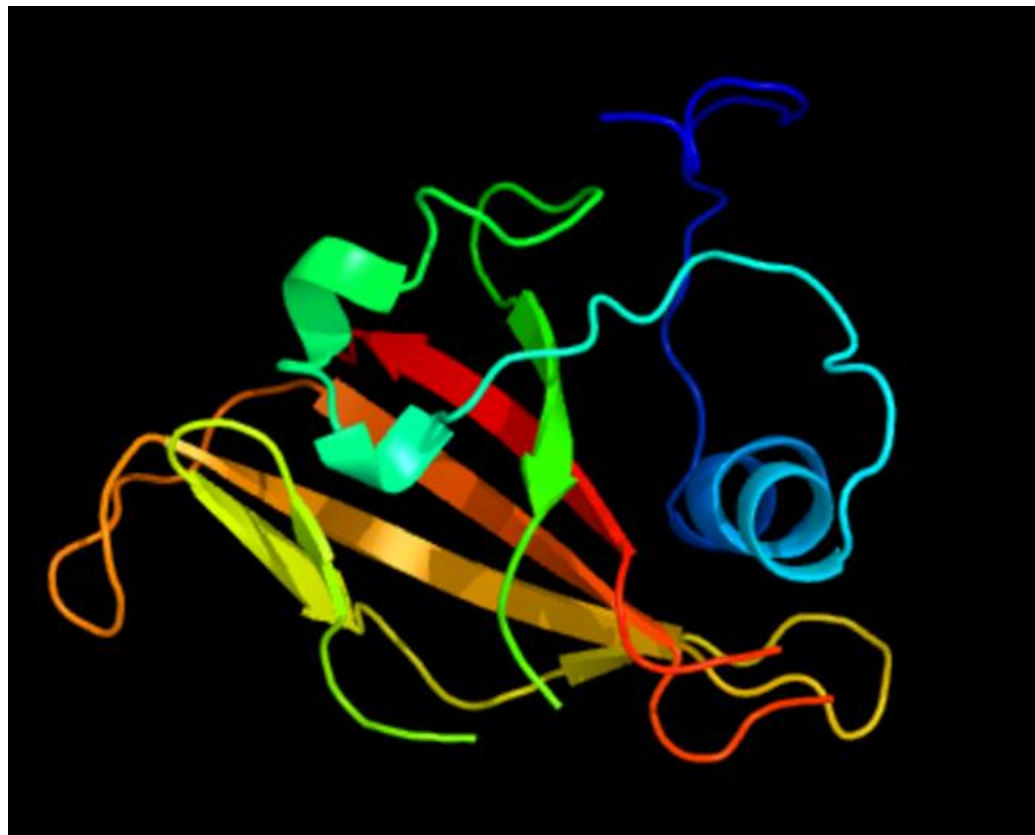


Z-score



➤ Structure

3D structure (Phary2)



Model (left) based on template [c3ulxA](#)

Top template information

PDB header: dna binding protein

Chain: A: **PDB Molecule:** stress-induced transcription factor nac1;

PDB Title: crystal structural of the conserved domain of rice stress-responsive2 nac1

Confidence and coverage

Confidence: **100.0%** Coverage: **40%**

146 residues (40% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.



➤ Summary

- VND1是植物中参与维管形成的转录因子，具 有一段保守的DNA结合结构域（NAC-domain），属于NAC转录因子家族
- VND1总共包含365个氨基酸、相对分子量为42476.67、等电点为5.4
- VND1没有跨膜结构，所以其不是跨膜蛋白
- VND1属于亲水蛋白，并且具有核定位信号， 将其位于细胞核中

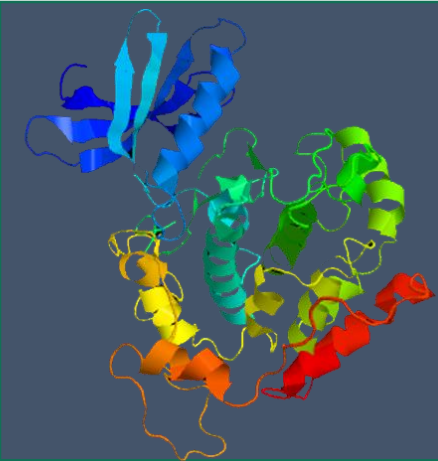


➤ Acknowledgement

衷心感谢罗老师一学期以来的耐心指导！
祝愿罗老师身体健康，工作愉快！

感谢我组所有成员对学习、讨论和实习的
配合，感谢班级同学对我组的帮助和
指导！





请老师批评指正，谢谢！



中国农业科学院
Chinese Academy of Agricultural Sciences