

# Structural Analysis Quick Start

## An NCBI Mini-Course

A protein domain is considered to be a distinct functional and/or structural unit. A domain in a structural context refers to a segment of a polypeptide chain that can fold into an independent three dimensional structure. It may interact with other domains of the protein or may simply be joined to other domains by a polypeptide chain. A domain in a sequence context refers to a long sequence pattern that is shared by other proteins having a common evolutionary origin. A domain may include all of the protein sequence or a part of it. A conserved domain is a recurring unit in molecular evolution whose extents can be determined by sequence and structure analysis.

**The Conserved Domain Database (CDD)** contains domains derived from the Smart, Pfam and Clusters of Orthologous Groups (COGs) databases. Conserved domains can be represented as multiple sequence alignments. Source alignments are processed by NCBI as follows:

- Sequences in the alignment for which a link can not be provided to a protein in Entrez are removed.
- If possible, a closely related sequence with a known structure is substituted.
- A representative sequence, preferably with a structure link, is chosen from among those in the alignment.
- A consensus sequence is made.
- A position-specific scoring matrix (PSSM) is constructed.

The Conserved Domain search (CD-search) compares a protein sequence to the PSSMs in the CDD database to identify conserved domains within it and to identify a 3-D modeling template. Since the PSSMs are the "subject", instead of the query as in PSI-Blast, the CD-search is a form of Reverse Position-Specific Blast (RPS-Blast).

**The Conserved Domain Architecture Retrieval Tool (CDART)** can be used to identify proteins containing the domain(s) present in the query sequence. Conserved domain(s) present in all sequences within Entrez proteins are identified using CD-search during routine NCBI processing. These pre-computed results are accessed through CDART.

**The Vector Alignment Search Tool (VAST)** is a computer algorithm developed at NCBI to detect similar protein 3-dimensional structures. The "structure neighbors" for every structure in NCBI's Molecular Modeling DataBase (MMDB)

are pre-computed. These neighbors can be used to identify distant homologs that cannot be recognized by sequence comparison alone. A VAST-search can be used for determining the structure neighbors for recently solved structures not yet in MMDB.

**Cn3D** is a helper application for web browsers to view 3-dimensional structures from NCBI's Entrez retrieval service. Cn3D runs on Windows, Macintosh, and Unix. Cn3D simultaneously displays structure, sequence, and alignment, and now has powerful annotation and alignment editing features.

## In this course, we will learn to

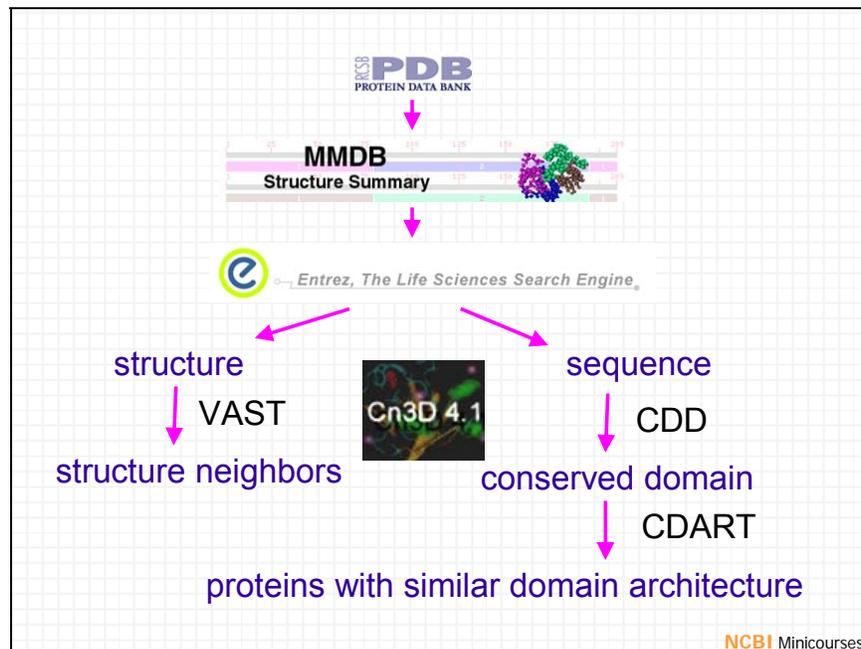
- Identify a conserved domain present in the query protein using **CDD**
- Search for other proteins containing similar domain(s) using **CDART**
- Explore a 3D modeling template for the query sequence using **CDD**
- Find similar structures using **VAST**
- Visualize and annotate the 3D protein structures using **Cn3D**

The remainder of the handout includes the introductory slides and the screen shots of the exercise demonstrated in Problem 1.

URL: <http://www.ncbi.nlm.nih.gov/Class/minicourses/quickstructure.html>

Course developed by: Dr. Medha Bhagwat ([bhagwat@ncbi.nlm.nih.gov](mailto:bhagwat@ncbi.nlm.nih.gov))

## Slides





## Conserved Domains

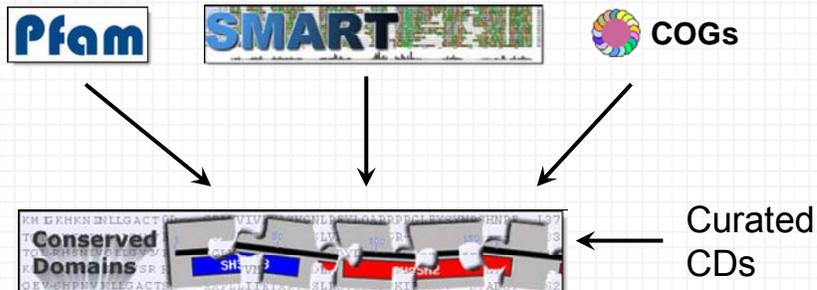
- recurring unit in molecular evolution;  
extent can be determined by  
sequence and structure analysis;
- performs a defined function;
- represented as a multiple local sequence  
alignment of proteins containing the domain  
(position specific scoring matrix, pssm).

CDD Search page:

<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>

NCBI Minicourses

## Conserved Domain Database

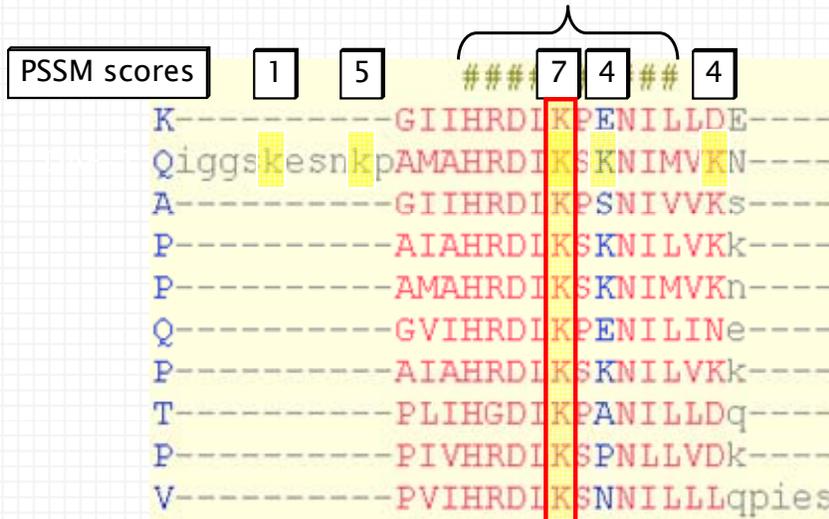


- A database of position-specific scoring matrices (PSSM)
- CD-Search can be used to search against the PSSMs
- Manual curation of CDs is ongoing (cd12345.version)

NCBI Minicourses

## Position-Specific Score Matrix

Serine/Threonine protein kinases  
catalytic loop



NCBI Minicourses

## Position-Specific Score Matrix

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
435 K	-1	0	0	-1	-2	3	0	3	0	-2	-2	1	-1	-1	-1	-1	-1	-1	-1	-2
436 E	0	1	0	2	-1	0	2	-1	0	-1	-1	0	0	0	-1	0	0	-1	-1	-1
437 S	0	0	-1	0	1	1	0	1	1	0	-1	0	0	0	2	0	-1	-1	0	-1
438 N	-1	0	-1	-1	1	0	-1	3	3	-1	-1	1	-1	0	0	-1	-1	1	1	-1
439 K	-2	1	1	-1	-2	0	-1	-2	-2	-1	-2	5	1	-2	-2	-1	-1	-2	-2	-1
440 P	-2	-2	-2	-2	-3	-2	-2	-2	-2	-1	-2	-1	0	-3	7	-1	-2	-3	-1	-1
441 A	3	-2	1	-2	0	-1	0	1	-2	-2	-2	0	-1	-2	3	1	0	-3	-3	0
442 M	-3	-4	-4	-4	-3	-4	-4	-5	-4	7	0	-4	1	0	-4	-4	-2	-4	-1	2
443 A	4	-4	-4	-4	0	-4	-4	-3	-4	4	-1	-4	-2	-3	-4	-1	-2	-4	-3	4
444 H	-4	-2	-1	-3	-5	-2	-2	-4	10	-6	-5	-3	-4	-3	-2	-3	-4	-5	0	-5
445 R	-4	8	-3	-4	0	-1	-2	-3	-2	-5	-4	0	-3	-2	-4	-3	-3	0	-4	-5
446 D	-4	-4	-1	8	-6	-2	0	-3	-3	-5	-6	-3	-5	-6	-4	-2	-3	-7	-5	-5
447 I	-4	-5	-6	-6	-3	-4	-5	-6	-5	3	5	-5	1	1	-5	-5	-3	-4	-3	1
448 K	0	0	1	-3	-5	-1	-1	-3	-3	-5	-5	7	-4	-5	-3	-1	-2	-5	-4	-4
449 S	0	-3	-2	-3	0	-2	-2	-3	-3	-4	-4	-2	-4	-5	2	6	2	-5	-4	-4
450 K	0	3	0	1	-5	0	0	-4	-1	-4	-3	4	-3	-2	2	1	-1	-5	-4	-4
451 N	4	-3	8	-1	-5	-2	-2	-3	-1	-6	-6	-2	-4	-5	-4	-1	-2	-6	-4	-5
452 I	3	-5	-5	-6	0	-5	-5	-6	-5	6	2	-5	2	-2	-5	-4	-3	-5	-3	3
453 M	-4	-4	-6	-6	-3	-4	-5	-6	-5	0	6	-5	1	0	-5	-4	-3	-4	-3	0
454 V	-3	-3	-5	-6	-3	-4	-5	-6	-5	3	3	-4	2	-2	-5	-4	-3	-5	-3	5
455 K	-2	1	1	4	-5	0	-1	-2	1	-4	-2	4	-3	-2	-3	0	-1	-5	-2	-3
456 N	1	1	3	0	-4	-1	1	0	-3	-4	-4	3	-2	-5	-2	2	-2	-5	-4	-4
457 D	-3	-2	5	5	-1	-1	1	-1	0	-5	-4	0	-2	-5	-1	0	-2	-6	-4	-5
458 L	-3	-1	0	-3	0	-3	-2	3	-4	-2	3	0	1	1	-2	-2	-3	5	-1	-3

catalytic loop

NCBI Minicourses

## Problem 1

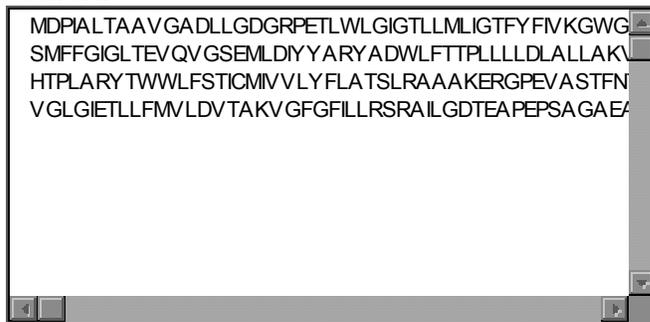
In this problem, we will follow these steps:

- A. Identify conserved domain(s) present in a protein.
- B. Search for other proteins containing similar domain(s).
- C. Explore a 3D modeling template for the query sequence.
- D. Find distant sequence homologs that may not be identified by BLAST.

NCBI's Conserved Domain Search allows you to match your protein sequence to a library of conserved protein domains, generate a multiple sequence alignment based on this match, and explore 3D modeling templates for your sequence. Click on the CDD link provided below,

### CDD

Paste the following protein sequence in the CD-Search query box and run the search.



- A. What is the domain present in this protein?  
Obtain more information about the domain by searching in [NCBI's Bookshelf](#)
- B. Go back to the CD-Search results page. Obtain a list of proteins with similar domain architecture by clicking on the "Search for similar domain architectures" button. To display the records, click on the link to the sequences and from there on the "Look up Sequences in Entrez". Change the display from "Summary" to "FASTA".
- C. Go back to the CD-Search results page. Generate a multiple sequence alignment for the top 10 sequences representative of the conserved domain hit by clicking on the graphic of the domain. Use the "Row Display" list box pull down menu to specify "up to 5" sequences and reformat sequence alignment. Extend the "Structure" display and invoke Cn3D with a display of a 3D modeling template and a multiple sequence alignment including your query sequence by pressing the "Show Structure" button.

The structure of the *Halobacterium salinarum* halorhodopsin protein and its sequence alignment with our query protein are displayed. For a better view of the backbone, remove the side chains globally (Style--Edit global style--Protein side chains). The query protein contains a bacterial rhodopsin signature (FMVLDVTAKVGF) where K is the retinal binding site. Identify these residues in the query protein and highlight the corresponding lysine residue in the halorhodopsin protein sequence.

Display the side chains of this residue (Use Style--Annotate--New--Edit Style. Change the protein backbone Rendering to Tubes, Color Scheme to User Selection and User Color to choose the color for the highlighted residue, for example yellow. Repeat these steps for the Protein Side chains row and click the Protein Side chains on. Click on the "Done" button. To zoom in, press z on the keyboard. Identify the cofactor near the lysine residue.

D. To obtain the structural neighbors for the halorhodopsin protein, first click on the structure entry link, 1TNO\_A, on the CD-Browser page. Then click Links → Structure on the top right, then on 1TNO again in the Entrez Structure page, and finally on the chain A graphic. To view neighbors with 1TNO\_A, select one or more of the check boxes next to the structure neighbors and view by clicking on the "View 3D Structure" button.

## Screenshots

The screenshot displays the NCBI Conserved Domain Database (CDD) search interface. At the top, there is a navigation bar with links for HOME, SEARCH, and SITE MAP. Below this, there are tabs for PubMed, Entrez, CDD, Structure, Protein, Taxonomy, BLAST, and Help?. The main content area is titled "A Conserved Domain Database and Search Service, v2.11". A search bar is present with a "GO" button and a "CLEAR" button. Below the search bar, there is a "Submit Query" button and a dropdown menu for "Search Database" set to "CDD v2.11 - 17402 PSSMs". The search results show a list of conserved domains, with the first one being "SH3". A red arrow points to the "Submit Query" button. The search results also include a "Find CDs" button and a link to "Read about the FASTA format description. Click here for advanced options."

NCBI

Conserved Domains

Query sequence: [(local sequence)|cl|Undefined\_sequence]

Concise Result
  Full Result
  Show Search Information

1 50 100 150 200 250 260

**Bac\_rhodopsin**

**Descriptions**

Title	Pssmid	Multi-Dom	E-value
[H]pfam01036_Bac_rhodopsin_Bacteriorhodopsin..	64878	No	1e-50

Search for similar domain architectures

NCBI National Center for Biotechnology Information  
National Library of Medicine National Institutes of Health

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search All Databases for bacteriorhodopsin Go

What can NCBI do? Hot Spots

d in 1988 as a national resource for  
 ology information, NCBI creates  
 ases, conducts research in  
 al biology, develops software tools  
 g genome data, and disseminates  
 nformation - all for the better  
 ing of molecular processes  
 man health and disease. [More...](#)

- ▶ Assembly Archive
- ▶ Clusters of orthologous groups
- ▶ Coffee Break, Genes & Disease, NCBI Handbook
- ▶ Electronic PCR

[Genome Association](#)

NCBI **Bookshelf**

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Books for bacteriorhodopsin Go Clear Save Search

Limits Preview/Index History Clipboard Details

About Entrez

Books

Overview

Using the books

Information for authors and publishers

Contact us

Mailing list

Display Books Show 20 Send to

All: 29 Figures: 11

[11 items](#) in **Molecular Biology of the Cell**. 4th ed.  
 Alberts, Bruce; Johnson, Alexander, Lewis, Julian; Raff, Martin; Roberts, Keith, Walter, Peter.  
 New York: [Garland Publishing](#); c2002.

[8 items](#) in **Biochemistry**.  
 Berg, Jeremy M.; Tymoczko, John L.; and Stryer, Lubert.  
 New York: [W. H. Freeman and Co.](#); 2002.

[6 items](#) in **Molecular Cell Biology**. 4th ed.  
 Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul, Baltimore, David; Darnell, James E.  
 New York: [W. H. Freeman & Co.](#); c2000.

### Many Integral Proteins Contain Multiple Transmembrane $\alpha$ Helices

Although [Figure 3-33](#) depicts glycophorin as a monomer with a single  $\alpha$  helix spanning the bilayer, this protein is present in erythrocyte membranes as a dimer of two identical polypeptide chains. The two membrane-spanning  $\alpha$  helices of glycophorin are thought to form a coiled-coil structure (see [Figure 3-9a](#)) stabilized by specific interactions between the amino acid side chains at the interface of the two helices. It is now known that many other transmembrane proteins contain two or more membrane-spanning  $\alpha$  helices. For instance, the *bacterial photosynthetic reaction center (PRC)* comprises four subunits and several prosthetic groups, including four chlorophyll molecules. In this complex protein, three of the four subunits span the membrane; two of these subunits (L and M) each contain five membrane-spanning  $\alpha$  helices (see [Figure 16-40](#)).

A large and important family of integral proteins is defined by the presence of seven membrane-spanning  $\alpha$  helices. More than 150 such “seven-spanning” membrane proteins have been identified. This class of integral proteins is typified by *bacteriorhodopsin*, a protein found in a photosynthetic bacterium ([Figure 3-34](#)). Absorption of light by the retinal group attached to *bacteriorhodopsin* causes a conformational change in the protein that results in pumping of protons from the cytosol across the bacterial membrane to the extracellular space. The proton concentration gradient thus generated across the membrane is used to synthesize ATP, as discussed in [Chapter 16](#). Both the overall arrangement of the seven  $\alpha$  helices in *bacteriorhodopsin* and the identity of most of the amino acids can be resolved by computer analysis of micrographs of two-dimensional crystals of the membrane-embedded protein taken at various angles to the electron beam.

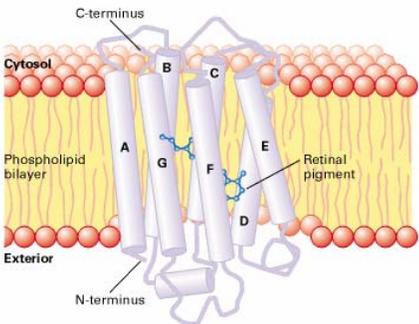
# MOLECULAR CELL BIOLOGY

Lodish Berk Zipursky Matsudaira Baltimore Darnell

W H FREEMAN AND COMPANY

[Short Contents](#) [Full Contents](#) [Other books @ NCBI](#)

*Molecular Cell Biology* → **3. Protein Structure and Function** → 3.4. Membrane Proteins



**Navigation**

- [About this book](#)
- 3. Protein Structure and Function**
- [3.1. Hierarchical Structure of Proteins](#)
- [3.2. Folding, Modification, and Degradation of Proteins](#)
- [3.3. Functional Design of Proteins](#)
- ↪ 3.4. Membrane Proteins**
- [3.5. Purifying, Detecting, and Characterizing Proteins](#)
- [PERSPECTIVES for the Future](#)
- [PERSPECTIVES in the Literature](#)
- [Testing Yourself on the Concepts](#)
- [MCAT/GRE-Style Questions](#)
- [References](#)

**Figure 3-34.** Overall structure of bacteriorhodopsin as deduced from electron diffraction analyses of two-dimensional crystals of the protein in the bacterial membrane. The seven membrane-spanning  $\alpha$  helices are labeled A–G. The retinal pigment is covalently attached to lysine 216 in helix G. The approximate position of the protein in the phospholipid bilayer is indicated. [Adapted from R. Henderson et al., 1990, *J. Mol. Biol.* 213:899.]

NCBI

**Conserved Domains**

HOME SEARCH SITE MAP NewSearch PubMed Nucleotide Protein Structure Taxonomy Help

Query sequence: [(local sequence)|cl|Undefined\_sequence]

Concise Result  Full Result  Show Search Information

**Descriptions**

Title	Pssmid	Multi-Dom	E-value
[+]pfam01036, Bac_rhodopsin, Bacteriorhodopsin...	64878	No	1e-50

[Search for similar domain architectures](#)

NCBI

**CDART: Conserved Domain Architecture Retrieval Tool**

[New Query](#) [Overview](#) [PubMed](#) [Nucleotide](#) [Protein](#) [Structure](#)

[About CDART](#)

Query

Bac\_rhodop

Similar domain architectures

551 Sequences  
cellular organisms  
hypothetical prote

**NCBI** **CDART: Conserved Domain Architecture Retrieval Tool**

New Query Overview PubMed Nucleotide Protein

[About CDART](#)

0 100 200 300

Query **Bac\_rhodop**

Similar domain architectures

XP_001395306 Aspergillus niger hypothetical prote		more>
XP_001395233 Aspergillus niger hypothetical prote		more>
AB088140 gamma proteobacter proteorhodopsin		more>
AB088139 gamma proteobacter proteorhodopsin		more>
AB088138 gamma proteobacter proteorhodopsin		more>
AB088137 gamma proteobacter proteorhodopsin		more>
AB064306 synthetic construc channel rhodopsin		more>
ZP_01737880 Marinobacter sp. E hypothetical prote		more>
ZP_01734914 Flavobacteria bact bacteriorhodopsin		more>

[Look Up Sequences in Entrez](#)

**NCBI** **Protein** My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search Protein for [ ] Go Clear

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort by Relevance Send to

All: 544 related Structures: 542

1: [FASTA] XML GenPept GI List Graphics TinySeq XML INSDSeq XML LinkOut

2: Related Sequences Conserved Domain Links 3D Domain Links Gene Links Genome Links

3: Genome Project Links INSDC Protein Genome Projects Links HomoloGene Links Nucleotide Links NIH cDNA clone links OMIA Links OMIM Links BioAssay Links BioAssay by target PubChem Compound Links PubChem Substance Links PMC Links PopSet Links Protein Cluster Links PubMed Links

4: [FASTA] XML GenPept GI List Graphics TinySeq XML INSDSeq XML LinkOut

5: Related Sequences Conserved Domain Links 3D Domain Links Gene Links Genome Links

6: [FASTA] XML GenPept GI List Graphics TinySeq XML INSDSeq XML LinkOut

proteorhodopsin [gamma proteobacterium HTCC6245] gi|141534095|gb|ABO88137.1|[141534095]

NCBI Logo

My NCBI (Sign In) (Register)

Search Protein for [ ] Go Clear

Limits Preview/Index History Clipboard Details

Display FASTA Show 20 Send to

Item 1 - 20 of 551 page 1 of 28 Previous Next

Item 1: [XP\\_001395306](#), Reports hypothetical prot...[gi:145246114] Links

>gi|145246114|ref|XP\_001395306.1| hypothetical protein An12g02200 [Aspergillus niger] [Next sequence](#)

```
MSLFMRANDALNVNFIAGVDEALSVHGSDDLWNAITAIYIVSFLGLLMSFSGRESRVFHYLFTFLLVG
AITVYQASDLGWSAVEQVNNLDNGAIRQMFNWKYINNVVAFPSLALALGLVSGVSWTICNIALACYW
VTSYLVSAITFSTYKMGFFAFGLTSLWVILAMSTINESREAAQLGVDYDIILSGWCNLLWLLVPAWGL
SDGGNVGITEATIFFGLDILMVLPLTFMFLFFARNWYQKLSIAFSDARPSREAQSSKQQVAESSPS
A
```

Item 2: [XP\\_001395233](#), Reports hypothetical prot...[gi:145245968] Links

>gi|145245968|ref|XP\_001395233.1| hypothetical protein An12g01460 [Aspergillus niger] [Previous sequence](#) [Next sequence](#)

```
MIESFKATPTLEWPEFSPSTSVAPIPTVVPQSDPIFQVLHDTGKRLWVVTMLGSSLVFVYLGARVP
ITKRVFHVVALASTISFIYALATGQGMWDKLDIKEHNKHPVNTTQYRQLNLRVYVFLTEFLG
LISLSLGLPGAHLIAAADFIMLGGILGTVAGHSSRRVWFTISAIGYLTIVYHISVHGSRAASNK
DVQTKRFFGLSGVMLFVKLLVVALAAGALALINVDVETILFAIYDIFSQGLGYWLLIAHDGAPGIT
LNVDFWVSHGLNEGAIREEDEGA
```

Item 3: [ABO88140](#), Reports proteorhodopsin [...][gi:141534101] BLINK, Conserved Domains Links

>gi|141534101|gb|ABO88140.1| proteorhodopsin [gamma proteobacterium HTCC2207] [Previous sequence](#) [Next sequence](#)

```
MINNLSASDPVGISFWLISMAVATAFFVDRDRVAGKWKTSITVSGLVTLIAAVHYFMRDWWVATGT
SPTVFRYIDWLLTVPLLMIEFYLLSAVTKVPAVGFWRLLIGTIAMLGFGYAGEAGWMDAKMAFIPMAAT
WFYLIWEIFRGEASQINAGLANANVQKAYKMTLLVTVGWSIYPLGFFGFMGSDQFVMLNVIYNVADF
WNKIAFGVVIWAAAVADSE
```

Item 4: [ABO88139](#), Reports proteorhodopsin [...][gi:141534099] BLINK, Conserved Domains Links

>gi|141534099|gb|ABO88139.1| proteorhodopsin [gamma proteobacterium HTCC6216] [Previous sequence](#) [Next sequence](#)

```
MALVVAATAFFVERDRVAGKWKTSITVSGLVTLIAAVHYFMRDWWVATGTSTVFRYIDWLLTVPLLMH
EYLLSAVTKVPAVGFWRLLIGTIAMLGFGYAGEAGWMDAKMAFIPMAAFYIWEIFRGEASQINAG
LANANVQKAYKMTLLVTVGWSIYPLGFFGFMGSDQFVMLNVIYNVADFVNKIAFGVVIWAAAVADSEK
```

Item 5: [ABO88138](#), Reports proteorhodopsin [...][gi:141534097] BLINK, Conserved Domains Links

>gi|141534097|gb|ABO88138.1| proteorhodopsin [gamma proteobacterium HTCC6124] [Previous sequence](#) [Next sequence](#)

```
YFFIERDRVAGKWKTSITVSGLVTLIAAVHYFMRDWWVSTGTSPTVFRYIDWLLTVPLLMIEFYLLSA
VTKVPAVGFWRLLIGTIAMLGFGYAGEAGWMDAKMAFIPMAAFYIWEIFRGEASQINAGLANANVQK
AYKMTLLVTVGWSIYPLGFFGFMGSDQFVMLNVIYNVADFVNKIAFGVVIWAAAVADSE
```

Item 6: [ABO88137](#), Reports proteorhodopsin [...][gi:141534095] BLINK, Conserved Domains Links

>gi|141534095|gb|ABO88137.1| proteorhodopsin [gamma proteobacterium HTCC6245] [Previous sequence](#) [Next sequence](#)

```
TAFFFIERDRVAGKWKTSITVSGLVTLIAAVHYFMRDWWVSTGTSPTVFRYIDWLLTVPLLMIEFYLL
SAVTKVPAVGFWRLLIGTIAMLGFGYAGEAGWMDAKMAFIPMAAFYIWEIFRGEASQINAGLANANV
```

NCBI Logo

Conserved Domains

SH3 SH2

Query sequence: [(local sequence)|Undefined\_sequence]

Concise Result  Full Result  Show Search Information

1 50 100 150 200 250 260

Bac\_rhodopsin

Descriptions

Title	Pssmid	Multi-Dom	E-value
[+]pfam01036, Bac_rhodopsin, Bacteriorhodopsin...	64878	No	1e-50

Search for similar domain architectures

NCBI Conserved Domains

pfam01036.13 Bac\_rhodopsin, with user query added

Bacteriorhodopsin.

Other Related Conserved Domains: C069524

Reformat Sequence Alignment Format: Compact Hypertext Row Display: Up to 5 Color Bits: 2.0 bits Type Selection: the most

```

1TN0_A      9  .[16].LYFLVK.[1].MGVSDPDAKKFYAITTLVPPIAFTMYLSMLLGYGLTMVFFG.[ 5].IYWARYADWLFTFTPLL 93
query      21  .[16].FYFIVK.[1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLIEVQVG.[ 5].IYYARYADWLFTFTPLL 105
gi 114809   34  .[16].LLFVFM.[1].RGLDDPRAKLIIVSTILVPGIASAAYLSMFFGIGLIEVQVG.[21].TMNGRYLTWALSTPMI 134
gi 60391839 21  .[16].FYFIVK.[1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLIEVQVG.[ 5].IYYARYADWLFTFTPLL 105
gi 2499387  14  .[16].LYFIAR.[1].WVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVITIELG.[ 5].IYWARYADWLFTFTPLL 98
gi 2499386  7  .[16].LYFIAR.[1].WGETDSRRQKFYIATILITAIIVNYLSMALGFGVITIEFA.[ 5].IYWARYADWLFTFTPLL 91
gi 1168614  4  .[16].AVLAYG  YTLVPEETRRKRYLLIIPGIAIVVAYALMALGFGVITIESEG.[ 1].VYVRYVDWLLTFTPLN 83
gi 2499383  29  .[16].LLFVAM.[1].RDIESPRAKLIIVVATMLVPLVSISSYAGLAGLTVGFLQMP.[11].SPWGRYLTWTFSTPMI 119
1E12_A     8  .[16].LVFVYM.[1].RTIRPGRPLIIVGATMIPVLSISSYAGLAGLTVGMIEMP.[11].SQWGRYLTWALSTPMI 98
1UAZ_A    15  .[16].FYFIVK.[1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLIEVQVG.[ 5].IYYARYADWLFTFTPLL 99

```

NCBI Conserved Domains

pfam01036.13 Bac\_rhodopsin, with user query added

Bacteriorhodopsin.

Show Structure

Program: Cn3D Drawing: All Atoms Aligned Rows: up to 5

[Download Cn3D]

Other Related Conserved Domains: C069524

Reformat Sequence Alignment Format: Compact Hypertext Row Display: up to 5 Color Bits: 2.0 bits

```

1TN0_A      9  .[16].LYFLVKGWVSDPDAKKFYAITTLVPPIAFTMYLSMLLGYGLTMVFFG.[ 4].PIYWARYADWLFTFTPLLLDL 97
query      21  .[16].FYFIVKGVGVTDKAREYYSITILVPGIASAAYLSMFFGIGLIEVQVG.[ 4].DIYYARYADWLFTFTPLLLDL 109
gi 60391839 21  .[16].FYFIVKGVGVTDKAREYYSITILVPGIASAAYLSMFFGIGLIEVQVG.[ 4].DIYYARYADWLFTFTPLLLDL 109
gi 2499387  14  .[16].LYFIARGWVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVITIELG.[ 4].AIYWARYDWLFTFTPLLLYL 102
1UAZ_A    15  .[16].FYFIVKGVGVTDKAREYYSITILVPGIASAAYLSMFFGIGLIEVQVG.[ 4].DIYYARYADWLFTFTPLLLDL 103

```

**CDD Descriptive Items**

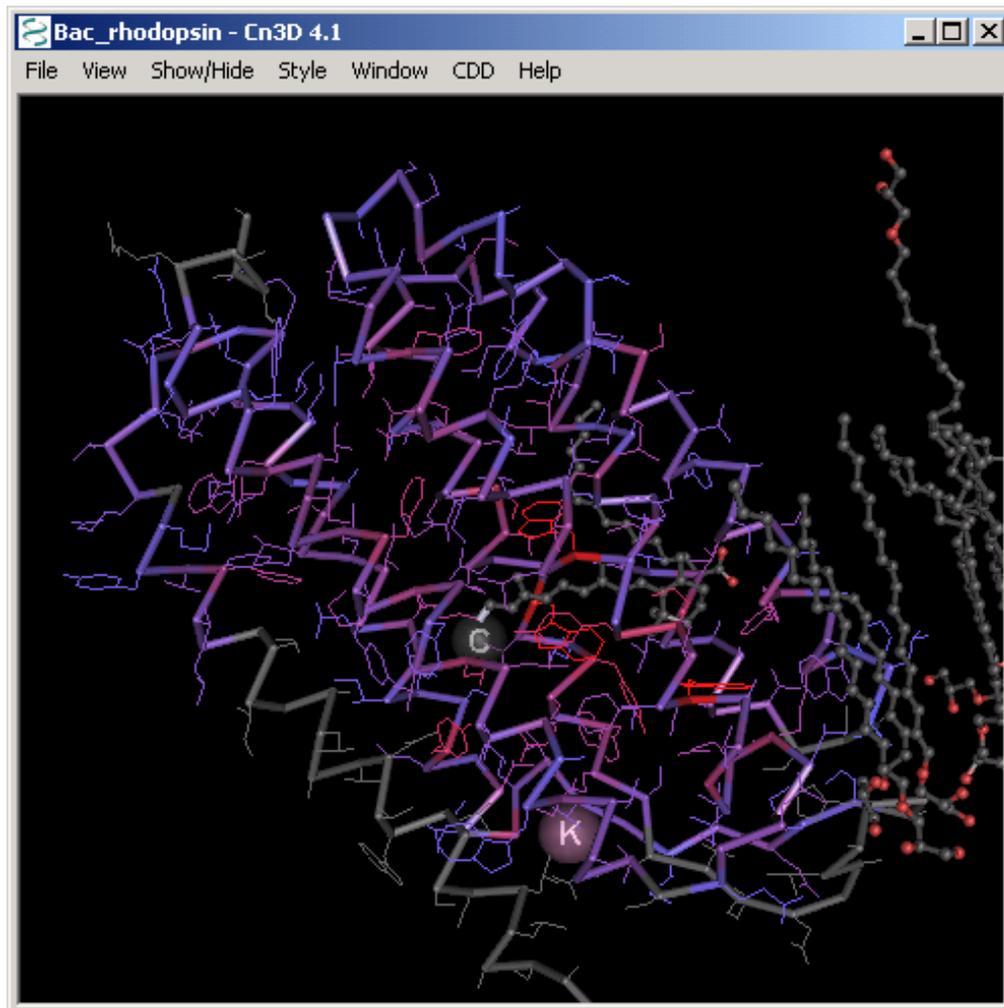
Name: Bac\_rhodopsin

Bacteriorhodopsin.

Structure summary:

PDB 1TN0 (MMDB 29943)  
1TN0\_A: gi 56553896 ([Halobacterium salinarum] Structure Of Bacteriorhodopsin Mutant A51p)

Show Annotations Panel    Show References Panel    Dismiss



Bac\_rhodopsin - Sequence/Alignment Viewer

View Edit Mouse Mode Unaligned Justification Imports

```

ITN0_A      m g l g t L Y F L V K G M G V S D P D A K K F Y A I T T L V P P I A F T M Y L S M L L G Y G L T M V P F G g e q n P I Y V
query       m l i g t F Y F I V K G W G V T D K E A R E Y Y S I T I L V P G I A S A A Y L S M F F G I G L T E V Q V G s e m 1 D I Y
gi 60391839 m l i g t F Y F I V K G W G V T D K E A R E Y Y S I T I L V P G I A S A A Y L S M F F G I G L T E V Q V G s e m 1 D I Y
gi 2499387  m f l g m L Y F I A R G W S V S D Q R R Q K F Y I A T I M I A A I A F V N Y L S M A L G F G V T T I E L G g e e r A I Y V
IUAZ_A     m l i g t F Y F I V K G W G V T D K E A R E Y Y S I T I L V P G I A S A A Y L S M F F G I G L T E V Q V G s e m 1 D I Y

```

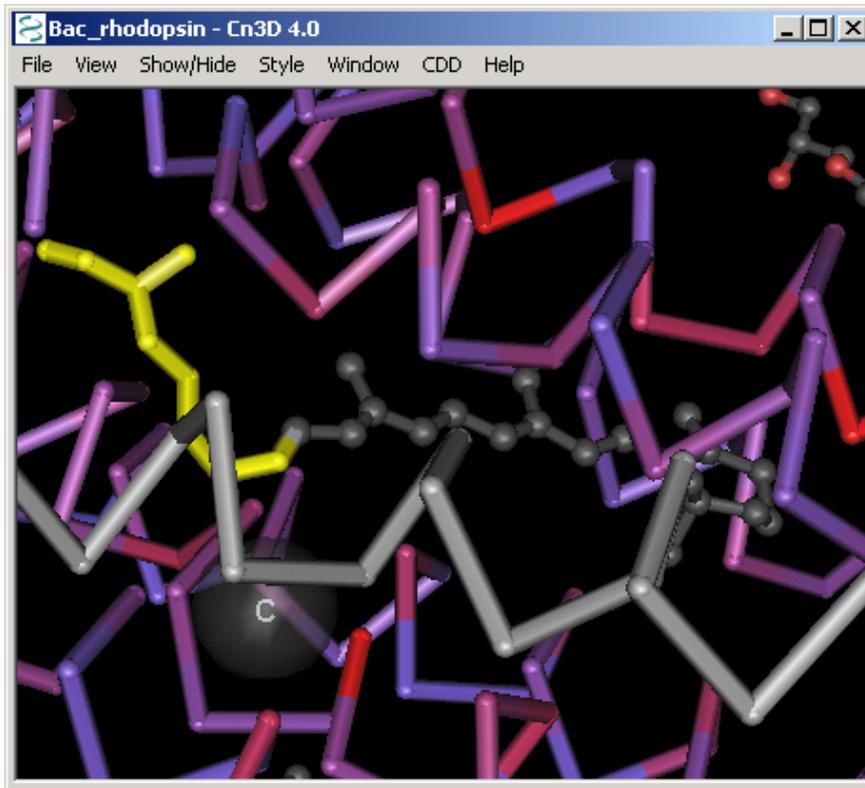
Style Options

Settings Labels Details

Rendering Settings

Group	Show	Rendering	Color Scheme	User Color
Protein backbone:	<input type="button" value="Complete"/>	<input type="button" value="Tubes"/>	<input type="button" value="User Selection"/>	<input type="color" value="#FFFF00"/>
Protein sidechains:	<input checked="" type="checkbox"/>	<input type="button" value="Tubes"/>	<input type="button" value="User Selection"/>	<input type="color" value="#FFFF00"/>
Nucleotide backbone:	<input type="button" value="Trace"/>	<input type="button" value="Tubes"/>	<input type="button" value="Molecule"/>	<input type="color" value="#808080"/>
Nucleotide sidechains:	<input checked="" type="checkbox"/>	<input type="button" value="Wire"/>	<input type="button" value="Molecule"/>	<input type="color" value="#808080"/>
Heterogens:	<input checked="" type="checkbox"/>	<input type="button" value="Ball and Stick"/>	<input type="button" value="Element"/>	<input type="color" value="#808080"/>
Solvents:	<input type="checkbox"/>	<input type="button" value="Ball and Stick"/>	<input type="button" value="Element"/>	<input type="color" value="#808080"/>
Connections:	<input checked="" type="checkbox"/>	<input type="button" value="Tubes"/>	<input type="button" value="User Selection"/>	<input type="color" value="#CCCCFF"/>
Helix objects:	<input type="checkbox"/>	<input type="button" value="With Arrows"/>	<input type="button" value="Object"/>	<input type="color" value="#808080"/>
Strand objects:	<input type="checkbox"/>	<input type="button" value="With Arrows"/>	<input type="button" value="Object"/>	<input type="color" value="#808080"/>
Virtual disulfides:	<input checked="" type="checkbox"/>			<input type="color" value="#FF8C00"/>
Hydrogens:	<input type="checkbox"/>		Background:	<input type="color" value="#000000"/>

Done Cancel Apply after each change?  Apply



NCBI

Conserved Domains

SH3 SH2

HOME SEARCH SITE MAP Entrez CDD Structure

**pfam01036.13** Bac\_rhodopsin, with user query added

Bacteriorhodopsin.

[+] Links:  
 [+] Statistics:  
 [-] Structure:

Show Structure

Program: Cn3D  
 Drawing: All Atoms  
 Aligned Rows: up to 5

[Download Cn3D]

Other related Conserved Domains: C005624

Reformat Sequence Alignment Format: Compact Hypertext Row Display: up to 5 Color Bits: 2.0 bits

1TNO_A	9	[16]	.LYFLVKGWGSVDPDAKKFYAITTLVPPIAFTMYLSMLLGYGLTMVPPG	[4]	.PIYWARYADWLFITPLLLLDL	97
query	21	[16]	.FYFIVKGSVTDKEAREYYSITILVPGIASAAIYLSMFFGIGLIEVQVG	[4]	.DIYYARYADWLFITPLLLLDL	109
g1_60391839	21	[16]	.FYFIVKGSVTDKEAREYYSITILVPGIASAAIYLSMFFGIGLIEVQVG	[4]	.DIYYARYADWLFITPLLLLDL	109
g1_2499387	14	[16]	.LYFIARGWSVSDQRKQKFYIATIMIAAIAFVNYLSMALGFGVITIELG	[4]	.AIYWARYTDWLFITPLLLYDL	102
1UAZ_A	15	[16]	.FYFIVKGSVTDKEAREYYSITILVPGIASAAIYLSMFFGIGLIEVQVG	[4]	.DIYYARYADWLFITPLLLLDL	103

NCBI **Links Protein**

Search: Protein for 1TN0A [Go] [Clear]

Display: GenPept Show: 20 Send to: [v]

Range: from begin to end Features:  CDD + Refresh

1: 1TN0A, Reports Chain A, Structur... [gi:56553896]

[Comment](#) [Features](#) [Sequence](#)

LOCUS 1TN0\_A 249 aa linear BCT 11-JUN-2004  
 DEFINITION Chain A, Structure Of Bacterorhodopsin Mutant A51p.  
 ACCESSION 1TN0\_A  
 VERSION 1TN0\_A GI:56553896  
 DBSOURCE pdb: molecule 1TN0, chain 6S, release Jun 11, 2004;  
 deposition: Jun 11, 2004;  
 class: Membrane Protein;  
 source: Mol\_id: 1; Organism\_scientific: Halobacterium Salinarium;  
 Organism\_common: Halobacterium Halobium; Gene: Bop, Vng1467g;  
 Expression\_system: Escherichia Coli; Expression\_system\_common:  
 Bacteria;  
 Exp. method: X-Ray Diffraction.

KEYWORDS .  
 SOURCE Halobacterium salinarum  
 ORGANISM [Halobacterium salinarum](#)

ELink, Conserved Dom  
 Links  
 Related Structure  
 Related Sequences  
 3D Domains  
 Domain Relatives  
 PubMed  
 Structure  
 Taxonomy



NCBI **Structure**

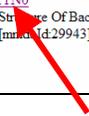
All Databases Search: Structure for [Go] [Clear]

Display: Summary Show: 20 Send to: [v]

All: 1 Bacterial: 1 Eukaryotic: 0 Ligand: 1 NMR: 0 X-ray: 1

1: 1TN0  
 Structure Of Bacterorhodopsin Mutant A51p  
 [mmdbId:29943]

VAST, Links



NCBI **MMDB** Structure Summary

PubMed BLAST Structure Taxonomy OMIM Help? Cn3d

**Reference:** Yohannan S, Yang D, Faham S, Boulting G, Whitelegge J, Bowie JU [Proline substitutions are not easily accommodated in a membrane protein](#) *J. Mol. Biol.* v341, p.1-6

**Description:** Structure Of Bacterorhodopsin Mutant A51p.

**Deposition:** 2004/6/11

**Taxonomy:** [Halobacterium salinarum](#)

**MMDB:** [29943](#) **PDB:** [1TN0](#) **Structure Neighbors:** [VAST](#)

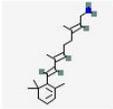
View 3D Structure of All Atom Model Cn3D Display Download Cn3D!

Molecular components in the MMDB structure are listed below. The icons indicate macromolecular chains, 3D domains, protein classifications and ligands. Please hold the mouse over each icon for more information on the component.

**Protein** Chain A  
**Domain Family** Bac\_rhodopsin

**Protein** Chain B  
**Domain Family** Bac\_rhodopsin

**Ligand** (x 2)



NCBI **VAST** Structure Neighbors

PubMed BLAST Structure Taxonomy OMIM Help? Cn3d

VAST neighbors for **MMDB 29943, 1TN0 A**

**Overview:** There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced neighbor search controls. The second section is the VAST neighbor list itself.

View 3D Alignment of All Atoms with Cn3D Display Download Cn3D!

View Sequence Alignment using Hypertext for Selected VAST neighbors

List Medium redundancy subset, sorted by Aligned Length in Graphics

Advanced neighbor search

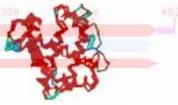
Move the mouse over the red alignment footprints in the graphics below and click you will obtain a structure-based sequence alignment.

Total neighbors: 177; 28 representatives from the Medium redundancy subset displayed.

Click to: [Check All](#) [Uncheck All](#)

1TN0 A Protein Family	Ali_len
<input type="checkbox"/> <a href="#">2JRF A</a>	225
<input type="checkbox"/> <a href="#">1C3M A</a>	222
<input type="checkbox"/> <a href="#">1E12 A</a>	220
<input type="checkbox"/> <a href="#">1H2S A</a>	216
<input type="checkbox"/> <a href="#">2F93 A</a>	214
<input type="checkbox"/> <a href="#">1XT0 A</a>	208





[PubMed](#)   [BLAST](#)   [Structure](#)   [Taxonomy](#)   [OMIM](#)   [Help?](#)   [Cn3D](#)

VAST neighbors for: [MMDB 29943](#), [1TN0 A](#) 

**Overview:** There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced neighbor search controls. The second section is the VAST neighbor list itself. 

of  with    [Download Cn3D!](#)

using  for  VAST neighbors

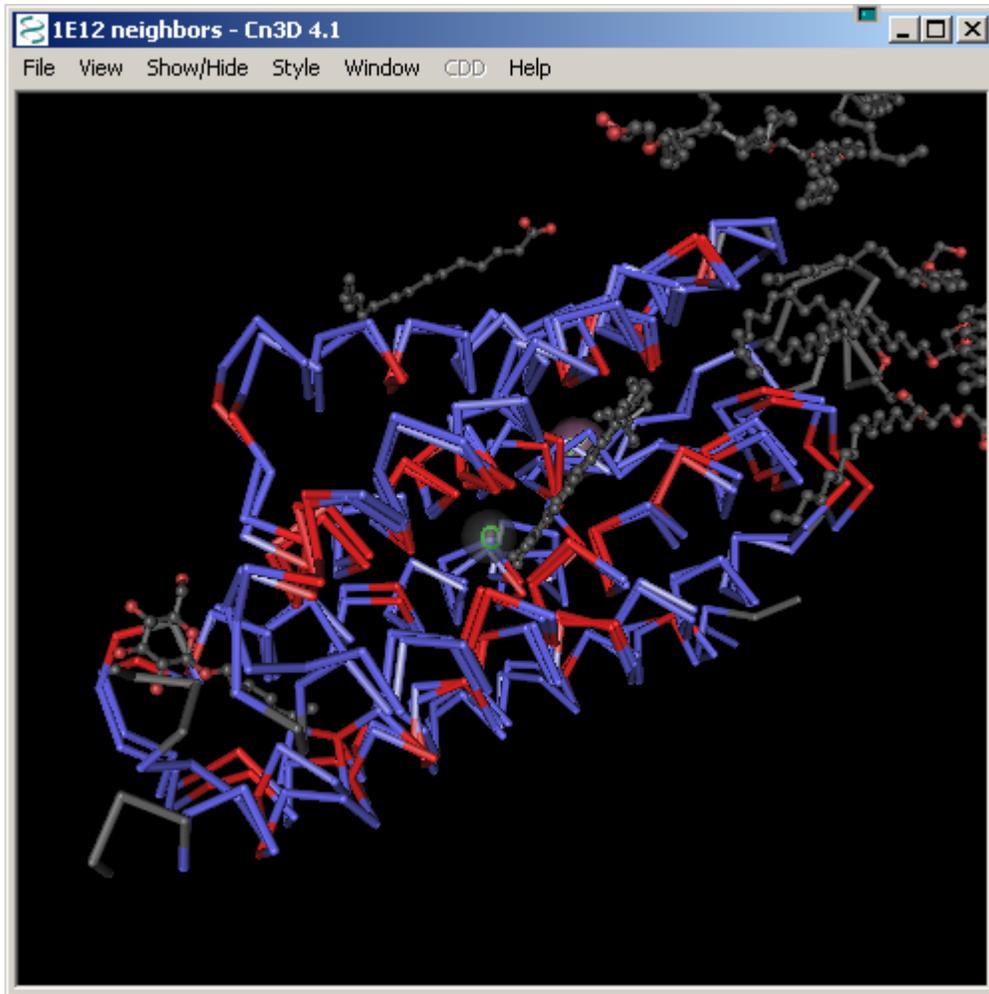
subset, sorted by  in  

[Advanced neighbor search](#)  

Total neighbors: 177; 28 representatives from the [Medium redundancy](#) subset displayed.

Click to: [Check All](#)   [Uncheck All](#)

	<a href="#">PDB</a>	<a href="#">C</a>	<a href="#">D</a>	<a href="#">Ali. Len</a>	<a href="#">Score</a>	<a href="#">E_Val</a>	<a href="#">Rmsd</a>	<a href="#">%Id</a>	<a href="#">MMDB Date</a>	<a href="#">LHM</a>	<a href="#">GSP</a>	<a href="#">Description</a>
<input type="checkbox"/>	<a href="#">2JAF</a>	<a href="#">A</a>		225	15.3	10e-14.9	1.7	33.3	01/2007	2.2	0.8	Ground State Of Halorhodopsin T203v
<input type="checkbox"/>	<a href="#">1C3W</a>	<a href="#">A</a>		222	16.0	10e-16.8	0.8	99.5	03/2001	0.8	0.4	BacteriorhodopsinLIPID COMPLEX AT 1.55 A RESOLUTION
<input type="checkbox"/>	<a href="#">1E12</a>	<a href="#">A</a>		220	15.5	10e-15.3	1.5	34.1	03/2001	2.1	0.7	Halorhodopsin, A Light-Driven Chloride Pump
<input checked="" type="checkbox"/>	<a href="#">1H2S</a>	<a href="#">A</a>		216	15.3	10e-14.8	1.1	29.2	11/2002	1.5	0.5	Molecular Basis Of Transmembrane Signalling By Sensory Rhodopsin li-Transducer Complex
<input type="checkbox"/>	<a href="#">2F93</a>	<a href="#">A</a>		214	15.5	10e-15.2	1.0	29.4	05/2006	1.4	0.5	K Intermediate Structure Of Sensory Rhodopsin liTRANSDUCER Complex In Combination With The Ground State Structure
<input type="checkbox"/>	<a href="#">1XIO</a>	<a href="#">A</a>		208	11.4	10e-8.9	1.6	29.3	11/2004	2.5	0.8	Anabaena Sensory Rhodopsin



1TNO neighbors - Sequence/Alignment Viewer

View Edit Mouse Mode Unaligned Justification Imports

1TNO_A	q a q i t G R P E W I W L A L G T A L M G L G T L Y F L V K G M G V s D P D A K K F Y A I T T L V P P I A F T M Y L S M L L C
1H2S_A	~ ~ ~ ~ M V G L T T L F W L G A I G M L V G T L A F A W A G R D A ~ G S G E R R Y Y V T L V G I S G I A A V A Y V V M A L C

## Problem 2

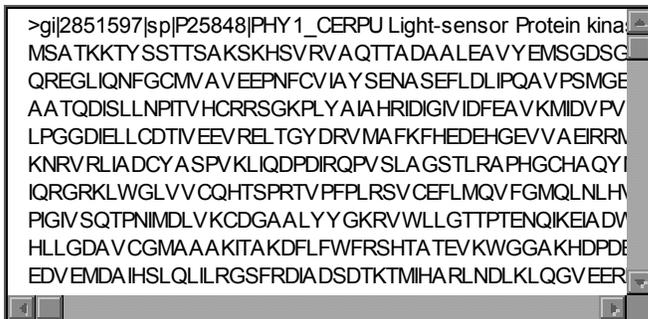
In this problem, we will follow these steps:

- Identify conserved domain(s) present in a protein.
- Search for other proteins containing similar domain(s).
- Explore a 3D modeling template for the query sequence.
- Find distant sequence homologs that may not be identified by BLAST.

NCBI's Conserved Domain Search allows you to match your protein sequence to a library of conserved protein domains, generate a multiple sequence alignment based on this match, and explore 3D modeling templates for your sequence. Click on the CDD link provided below,

### [CDD](#)

paste the following protein sequence in the CD-Search query box and run the search.



```
>gj|2851597|sp|P25848|PHY1_CERPU Light-sensor Protein kina:
MSATKKTY SSTTSAKSKHSVRVAQTTADAALAEVYEMSGDSC
QREGLIQNFGCMVAVEEPNFCVIA YSENA SEFLDLIPQAVPSMGE
AATQDISLLNPTVHCRRSGKPLYAJAHRIDIGVIDFEAVK MIDVPV
LPGGDIELLCDTVEEVRELTGYDRVMAFKFHEDEHGEVVAEIRR
KNRVRLIADCYASPVKLIQDPDIRQPVSLAGSTLRAPHGCHAQYI
IQRGRKLWGLVV CQHTSPRTV PFPLRSVCEFLMQVFGMQLNLH
PIGV SQT FNIMDLVKCDGAALYYGKR V WLLGTTPTENQIKEIADV
HLLGDAVCGMAAAKITAKDFLWFRSHTATEV/KWGGAKHDPDE
EDVEMDAIHSLQLILRGSFRDIADSDTKTMIHARLNDLKLQGV EER
```

- What are the domains present in this protein?  
(Select the "Full Result" radio button to display all of the domains.)

-Suppose, we are interested in the serine/threonine protein kinase domain.  
Obtain more information about it by searching in [NCBI's Bookshelf](#)

- Go back to the CD-Search results page. Obtain a list of proteins with similar domain architecture by clicking on the "Search for similar domains architectures" button. To display the records, click on the links to the subsets of sequences and from there on the "Look up Sequences in Entrez". Change the display from "Summary" to "FASTA".

- Go back to the CD-Search results page. Generate a multiple sequence alignment for the top 10 sequences representative of the conserved domain hit by clicking on the graphic representation of the serine/threonine kinase domain from CDD (CDD|00180). Use the "Row Display" list box pull down menu to specify "up to 5" sequences and reformat sequence alignment. Invoke Cn3D with a display of a 3D modeling template and a multiple sequence alignment

including your query sequence by pressing the "Show Structure" button.

To show only one top structure, click on the down arrow key. For better view of the backbone, remove the side chains globally (Style--Edit global style--Protein side chains). The query protein contains a serine/threonine protein kinases active-site signature (IIHRDLKSMNILV) where K is the ATP binding site. Identify these residues in the query protein and highlight the corresponding lysine residue in the first protein sequence.

Display the side chains of this residue (Use Style--Annotate--New--Edit Style. Change the protein backbone Rendering to Tubes, Color Scheme to User Selection and User Color to choose the color for the highlighted residue, for example yellow. Repeat these steps for the Protein Side chains row and click the Protein Side chains on. Click on the "Done" button. To zoom in, press z on the keyboard. Note the heterogen near the lysine residue.

D. To obtain the structural neighbors for the serine/threonine protein kinase protein, first click on the structure entry link 1JNK of the similar protein from the CD-Browser page. Then click on the structure link on the top right side, then on 1JNK, and finally on the chain graphic. Select one or more of the check boxes next to the structure neighbors and download the structures by clicking on the "View 3D Structure" button.