

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

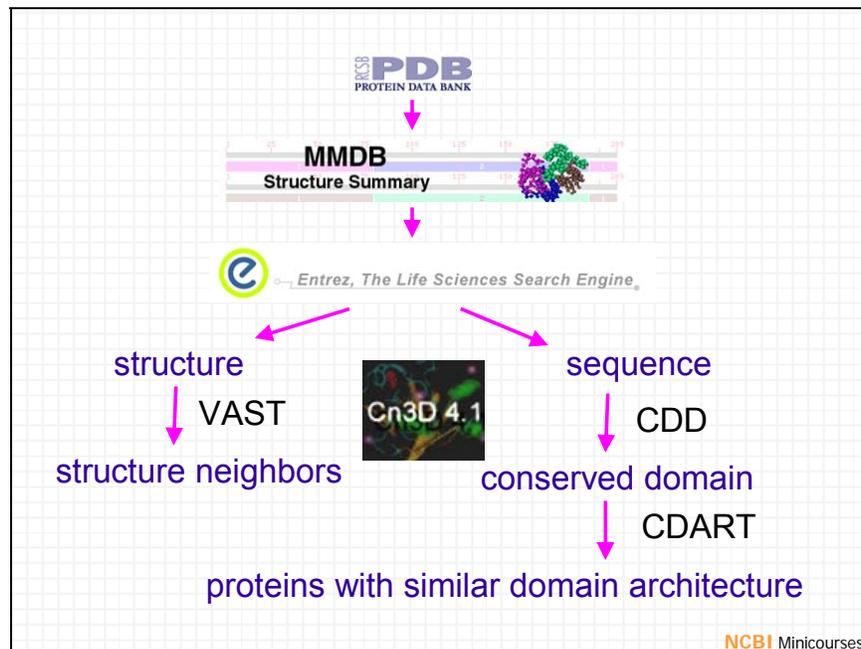
- Identify a conserved domain present in the query protein using **CDD**
- Search for other proteins containing similar domain(s) using **CDART**
- Explore a 3D modeling template for the query sequence using **CDD**
- Find similar structures using **VAST**
- Visualize and annotate the 3D protein structures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

Slides





Conserved Domains

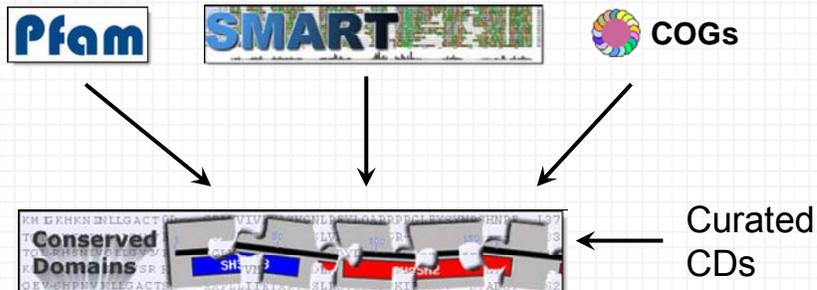
- recurring unit in molecular evolution;
extent can be determined by
sequence and structure analysis;
- performs a defined function;
- represented as a multiple local sequence
alignment of proteins containing the domain
(position specific scoring matrix, pssm).

CDD Search page:

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

NCBI Minicourses

Conserved Domain Database

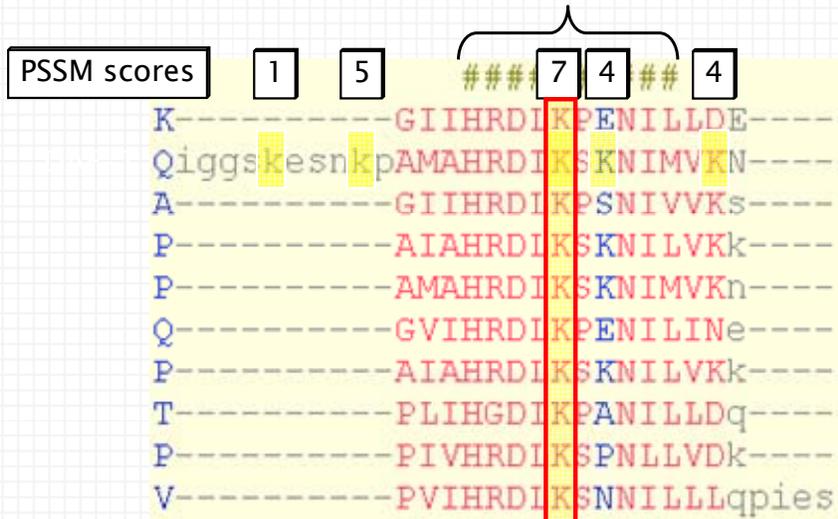


- A database of position-specific scoring matrices (PSSM)
- CD-Search can be used to search against the PSSMs
- Manual curation of CDs is ongoing (cd12345.version)

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

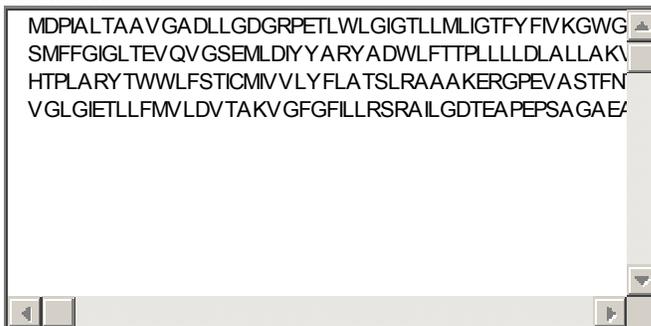
In this problem, we will follow these steps:

- 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, 1E12_A, on the CD-Browser page. Then click Links → Structure on the top right, then on 1E12 again in the Entrez Structure page, and finally on the chain A graphic. To view neighbors with 1E12_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, a 3D ribbon diagram of a protein structure is shown, with two conserved domains, SH3 and SH2, highlighted in green and blue respectively. Below the diagram, the NCBI logo and navigation tabs (HOME, SEARCH, SITE MAP, PubMed, Entrez, CDD, Structure, Protein, Taxonomy, BLAST, Help) are visible. The main content area is titled "A Conserved Domain Database and Search Service, v2.1.0". It includes a search bar with "GO" and "CLEAR" buttons, a "Submit Query" button, and a search database dropdown set to "CDD v2.10 - 12589 PSSMs". The search results display a protein sequence in FASTA format: "SMFFGIGLTEVQVGSSEMLDIYYARYADWLF...". A pink arrow points to the "Submit Query" button.

NCBI Conserved Domains

Query sequence: [(local sequence)\cl|Undefined_sequence]

Concise Result Full Result Show Search Information

Descriptions

+	Title	Pssmid	Multi-Dom	E-value
Hpfam01036	Bac_rhodopsin, Bacteriorhodopsin..	41106	No	1e-47

[Search for similar domain architectures](#)

CD Search Reference:
 Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly", *Nucleic Acids Res.*32(W)327-331.
[Help](#) | [Disclaimer](#) | [Write to the Help Desk](#)
 NCBI | NLM | NIH

NCBI National Center for Biotechnology Information

Search: for

What do we do?
 In 1988 as a national resource for biology information, NCBI creates databases, conducts research in molecular biology, develops software tools for genome data, and disseminates information - all for the better understanding of molecular processes and human health and disease. [More...](#)

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

NCBI Bookshelf

Search: for [Save Search](#)

Display: Show: 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 α Helices

Although [Figure 3-33](#) depicts glycophorin as a monomer with a single α helix spanning the bilayer, this protein is present in erythrocyte membranes as a dimer of two identical polypeptide chains. The two membrane-spanning α 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 α 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 α helices (see [Figure 16-40](#)).

A large and important family of integral proteins is defined by the presence of seven membrane-spanning α 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 α 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.

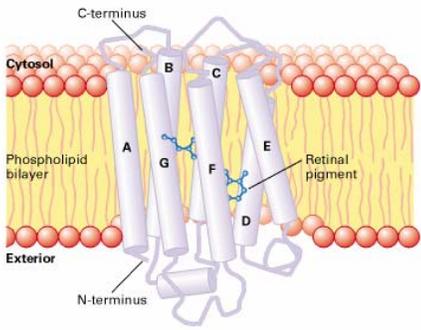
MOLECULAR CELL BIOLOGY

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 α 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

Query sequence: [(local sequence)|cl|Undefined_sequence]

Concise Result Full Result Show Search Information

Descriptions

Title	PssmId	Multi-Dom	E-value
Hjpfam01036, Bac_rhodopsin, Bacteriorhodopsin...	41106	No	1e-47

[Search for similar domain architectures](#)

NCBI CDART: Conserved Domain Architecture Retrieval Tool

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

[About CDART](#)

Query: Bac_rhodop

Similar domain architectures

578 Sequences

cellular organisms			
hypothetical prote			
NP_346428		HATPase_c	more>
Streptococcus pneu			
sensor histidine k	Icc		more>
EAY23562			
Trichomonas vaginalis			
hypothetical prote			

NCBI CDART: Conserved Domain Architecture Retrieval Tool

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

[About CDART](#)

Query: Bac_rhodop

Similar domain architectures

CAK41066		more>
Aspergillus niger		
unnamed protein pr		
CAK48429		more>
Aspergillus niger		
unnamed protein pr		
ZP_01737880		more>
Moribacter sp. E		
hypothetical prote		
ZP_01734914		more>
Flavobacterium bact		
bacteriorhodopsin		
YP_001384623		more>
Pichia stipitis OB		
plasma membrane he		
ZJAF8		more>
Moribacterium selli		
unnamed protein pr		
XP_001272404		more>
Aspergillus clavus		
opsin, putative		
ABL97827		more>
uncultured marine		
pop proteorhodopsi		
ABL97764		more>
uncultured marine		
pop proteorhodopsi		

[Look Up Sequences in Entrez](#)

NCBI Entrez Protein

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort by Relevance Send to

Summary ASN.1

FASTA

XML

GenPept

GI List

Graphics

TinySeq XML

INSDSeq XML

LinkOut

Related Sequences

Conserved Domain Links

3D Domain Links

Gene Links

Genome Links

Genome Project 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

PubMed Links

PubMed (RefSeq) Links

SNP Links

2JAF A

unnamed protein product [Halobacterium salinarum]

gi|122920878|pdb|2JAF|A|122920878

Page 1

NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Display FASTA Show 20 Send to

Item 1 - 20 of 578

page 1 of 29 Previous Next

1: CAK41066 Reports ...[gi|134080019]

>gi|134080019|emb|CAK41066.1| unnamed protein product (Aspergillus niger)

MSLFMRNDALNPNFIAGVDEALSVHGSDDLWLAITAVIVSFLGLTMSFGKESERVFHYLFTFLLVVG

AITYYQASDLGWSAVEQVNNLDNGAIRQMFNAKYINWVAFPSLALGLVSGVSWITMCIHALACYW

VSYLVSAITPSTYKMGFFAFGLTSLWVILAMSTINESREAAQRLGVDRDYIILSGWNLWLLYPVAVGL

SDGGNVIGITEATIFFGILDILMVLPLTFMFLFFARNDYQKLSIAFSDARFSPREAGSKQQQVAESSPS

A

2: CAK48429 Reports ...[gi|134079945]

>gi|134079945|emb|CAK48429.1| unnamed protein product (Aspergillus niger)

MIESFKAFITFLWEPSPSTSVAPITPTVPGSDPIFQVLDHTGKRLNWNVIMGSSLVFVYVLRARVE

ITKRVFHLVALASTISFTIYALATGGQMDKLDIIRKHNKHWNTQEVYRQLMLRVYVWFLFPLG

LISLSLGLPGAMHLLAIAADFMILGSGILGTAGHSSRRVWVTISAIYGLTVVYVHSGSRAASNK

DVQTKRFFGTLGVMFLVYVLAAGALALKINVDVETILFAIYDIFSQGILGWLLIAHDGAPGIT

LNVDFWVSHGLNGEAGIRISEEDGA

3: ZP_01737880 Reports hypothetical prot...[gi|126666904]

>gi|126666904|ref|ZP_01737880.1| hypothetical protein ME1B17_06684 (Marinobacter sp. ELB17)

MELEESLVGQYGFISNVFSLANASAILFHWLMSVAPAVRMAIISGLVTAIAVHYVLIQMLSWH

GAASVITGEVLTSSEGFNRARYVDMVLLVPLLVLLVLLVLMKLSNSETVRSVYVLSGAAALMIIIGYFGE

VSGVGTALFWLSMIFPLYIVRELVIGLKESSISQQPDEVKGLISAATLVVASNAFPIVYVLELLGI

TGSAIVAVETGYTIADIVSKAVFGLLIFTIAVRKSKADERHVEVPVETATQAEQKTFVVKKA

4: ZP_01734914 Reports bacteriorhodopsin...[gi|126663919]

>gi|126663919|ref|ZP_01734914.1| bacteriorhodopsin (Flavobacteria bacterium BAL38)

NCBI Conserved Domains

Query sequence: [(local sequence)|cl|Undefined_sequence]

Concise Result Full Result Show Search Information

Descriptions

Title	PssmId	Multi-Dom	E-value
Hpfam01036_Bac_rhodopsin, Bacteriorhodopsin	81106	No	1e-47

Search for similar domain architectures

NCBI Conserved Domains

pfam01036.12 Bac_rhodopsin, with user query added

[+] Links: Bacteriorhodopsin.
 [+] Statistics:
 [+] Structure:

Other Related Conserved Domains: C005524

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

```

1E12_A      8 . [16] .LVFVYM. [1].RTIRPGRPLINGATLMIPLVSISSYLGLLSGLTVGMIEMP. [11].SQWGRYLTWALSTPMI 98
query      21 . [16] .FYFIVK. [1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG. [ 5].IYYARYADWLFPTPLL 105
1UAZ_A     15 . [16] .FYFIVK. [1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG. [ 5].IYYARYADWLFPTPLL 99
1MOK_A     22 . [16] .LYFLVK. [1].WGVSDPDAKFFYAITTLVPAIAFTMYLSMLLGYGLTMVPPG. [ 5].IYARYADWLFPTPLL 106
qi 114809   34 . [16] .LLFVFM. [1].RGLDDPRAKLIQVSTILVFPVVSISYTLGLASGLTISVLEMP. [21].TMWGRYLTWALSTPMI 134
qi 461609   34 . [16] .LLFVFM. [1].RNVEDPRAQLIFVATLMVPLVSISSYTLGLVSGLTVGFLEMP. [11].TPWGRYLTWALSTPMI 124
qi 2499383  29 . [16] .LLFVFM. [1].RDIESPRAKLIQVATLMVPLVSISSYAGLASGLTVGFLEMP. [11].SPWGRYLTWTFSTPMI 119
qi 1168614   4 . [16] .AVLAYG. [1].YTLVPEETRRKRVLLIIAIPGIAIVAYALMALGFGSIQSEGH. [ 1].YVVVRYVDULLTTPLN 83
qi 2499387  14 . [16] .LYFIAR. [1].WVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG. [ 5].IYARYADWLFPTPLL 98
qi 2499386   7 . [16] .LYFIAR. [1].WGETSRRQKFYIATILITAFVNYLAMALGFGLTIVEFA. [ 5].IYARYSDWLFPTPLL 91
  
```

NCBI Conserved Domains

pfam01036.12 Bac_rhodopsin, with user query added

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

Show Structure

Program: Cn3D
 Drawing: AllAtoms
 Aligned Rows: up to 5
 [Download Cn3D]

Other Related Conserved Domains: C005524

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

```

1E12_A      8 . [16] .LVFVYMGRTIRPGRPLINGATLMIPLVSISSYLGLLSGLTVGMIEMPAGH. [ 8].SQWGRYLTWALSTPMILL 100
query      21 . [16] .FYFIVKGVGVTDKAREYYSITILVPGIASAAYLSMFFGIGLTEVQVGSEM. [ 2].IYYARYADWLFPTPLLLL 107
1UAZ_A     15 . [16] .FYFIVKGVGVTDKAREYYSITILVPGIASAAYLSMFFGIGLTEVQVGSEM. [ 2].IYYARYADWLFPTPLLLL 101
1MOK_A     22 . [16] .LYFLVKGVSDDPDAKFFYAITTLVPAIAFTMYLSMLLGYGLTMVPPGGEQ. [ 2].IYARYADWLFPTPLLLL 108
qi 2499387  14 . [16] .LYFIARGWVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELGEE. [ 2].IYARYADWLFPTPLLLL 100
  
```

CDD Descriptive Items [X]

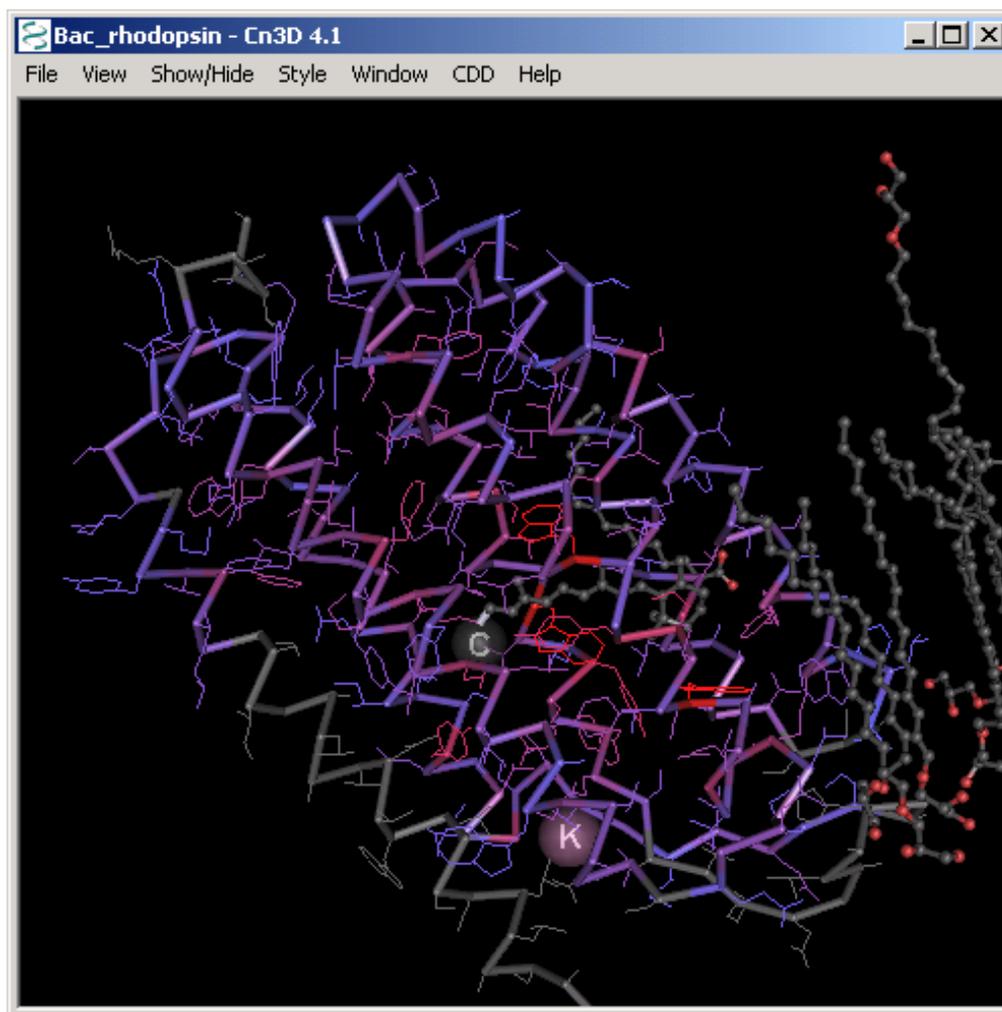
Name: Bac_rhodopsin

Bacteriorhodopsin.

Structure summary:

PDB 1E12 (MMDB 13348)
1E12_A: gi 8569313 ([Halobacterium salinarum] Chain A, Halorhodopsin, A
Light-Driven Chloride Pump)

Show Annotations Panel Show References Panel Dismiss



Bac_rhodopsin - Sequence/Alignment Viewer

View Edit Mouse Mode Unaligned Justification Imports

```

iE12_A agia iLVFVYMGRTIRPGRPLIWGATLMIPLVSISSYLGLLSGLTVGMIEMP aghala-----gem
query  mli g tFYFIVK gWGVTDKEAREYYSITILVPGIASAAAYLSMFFGIGLTEVQVG sem-----
iUA2_A mli g tFYFIVK gWGVTDKEAREYYSITILVPGIASAAAYLSMFFGIGLTEVQVG sem-----
iMOK_A mli g tLYFIVK gMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVFPF g e q-----
gi 114809 agl s iLLFVFM tRGLDDPRAKLI AVSTILVPPVSIASVTGLASGLTISVLEMP aghfae g s s v m l g g e e v d g
gi 461609 agl s iLLFVYMG RNVEDPRAQLIFVATLMVPLVSISSYTG LVSGLTVSFLEMP aghala-----g q e
gi 2499383 agv v iLLFVAM gRDI ESPRAKLIWVATMLVPLVSISSYAGLASGLTVGFLQMP pghala-----g q e
gi 1168614 e l l g tAVLAYG ~YTLVPEETRKRYLLIIAIPGIAIVAYALMALGFGSIOSEGH a-----
gi 2499387

```

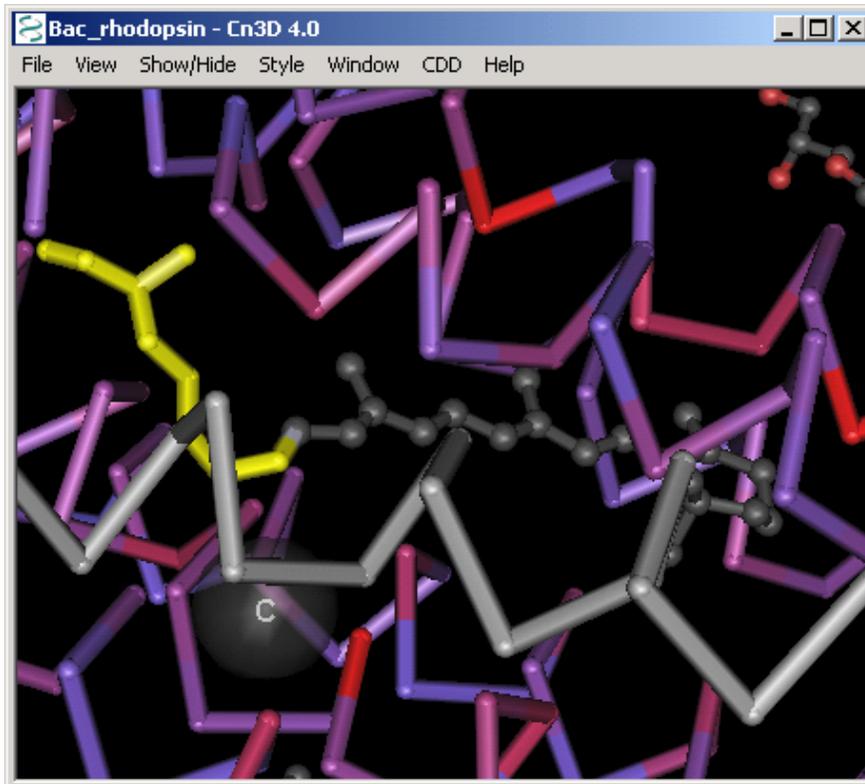
Style Options

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"/>	
Protein sidechains:	<input checked="" type="checkbox"/>	<input type="button" value="Tubes"/>	<input type="button" value="User Selection"/>	
Nucleotide backbone:	<input type="button" value="Trace"/>	<input type="button" value="Tubes"/>	<input type="button" value="Molecule"/>	
Nucleotide sidechains:	<input checked="" type="checkbox"/>	<input type="button" value="Wire"/>	<input type="button" value="Molecule"/>	
Heterogens:	<input checked="" type="checkbox"/>	<input type="button" value="Ball and Stick"/>	<input type="button" value="Element"/>	
Solvents:	<input type="checkbox"/>	<input type="button" value="Ball and Stick"/>	<input type="button" value="Element"/>	
Connections:	<input checked="" type="checkbox"/>	<input type="button" value="Tubes"/>	<input type="button" value="User Selection"/>	
Helix objects:	<input type="checkbox"/>	<input type="button" value="With Arrows"/>	<input type="button" value="Object"/>	
Strand objects:	<input type="checkbox"/>	<input type="button" value="With Arrows"/>	<input type="button" value="Object"/>	
Virtual disulfides:	<input checked="" type="checkbox"/>			
Hydrogens:	<input type="checkbox"/>			
			Background:	

Done Cancel Apply after each change? Apply



NCBI

Conserved Domains

Bac_rhodopsin, with user query added

pfam01036.12 Bacterorhodopsin.

Structure:

Other Related Conserved Domains: [466924](#)

Retformat Sequence Alignment

Accession	Length	Score	Start	End	Model	EV
1E12_A	0	(-16)	.LVFVYRGRTPFPQPFPLWGA	TLNIPLVLSIDSYLGLLSGLTVQKIERFAGH	(8)	.00000174
query	21	(-16)	.FTFIVRGRGQVTDKAREYYS	ITILVFGIASAAYLSMFFGIGLTVQVQGER	(2)	.1174747
1UAZ_A	15	(-16)	.FTFIVRGRGQVTDKAREYYS	ITILVFGIASAAYLSMFFGIGLTVQVQGER	(2)	.1174747
1MOK_A	22	(-16)	.LVFLVGRGQVTDKAREYYS	ITILVFGIASAAYLSMFFGIGLTVQVQGER	(2)	.1174747
gi_2499387	14	(-16)	.LVFLVGRGQVTDKAREYYS	ITILVFGIASAAYLSMFFGIGLTVQVQGER	(2)	.1174747

NCBI Protein

Search Protein for 1E12A[ACCN] Go Clear

Limits Preview/Index History Clipboard Details

Display GenPep Show 20 Send to

Range: from begin to end Features: CDD + Refresh

1: 1E12A Reports Chain A, Halorhod. [gi:8569313] BLink, Conserved

Comment Features Sequence

LOCUS 1E12_A 253 aa linear BCT 06-APR-2000

DEFINITION Chain A, Halorhodopsin, A Light-Driven Chloride Pump.

ACCESSION 1E12_A

VERSION 1E12_A GI:8569313

DBSOURCE pdb: molecule 1E12, chain 65, release Apr 6, 2000; deposition: Apr 6, 2000; class: Ion Pump; source: Mol_id: 1; Organism_scientific: Halobacterium Salinarum; Strain: D2; Cellular_location: Membrane; Gene: Hop; Other_details: H. Sal. Strain D2 Was Constructed For Homologous Overexpression Of Hr. See Also Heymann Et Al., Mol. Microbiol., Vo. 7, 623-630 (1993).; Exp. method: X-Ray Diffraction.

KEYWORDS

SOURCE Halobacterium salinarum

ORGANISM Halobacterium salinarum
Archaea; Euryarchaeota; Halobacteria; Halobacteriales; Halobacteriaceae; Halobacterium.

REFERENCE 1 (residues 1 to 253)

AUTHORS Havelka, W.A., Henderson, R. and Oesterheit, D.

TITLE Three-dimensional structure of halorhodopsin at 7 Å resolution

JOURNAL J. Mol. Biol. 247 (4), 726-738 (1995)

PUBMED 7723027

REFERENCE 2 (residues 1 to 253)

Links
 Related Structure
 Related Sequences
 3D Domains
 Domain Relatives
 PubMed
 Structure
 Taxonomy

NCBI Structure

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search Structure for Go Clear

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Send to

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

1: 1E12 Halorhodopsin, A Light-Driven Chloride Pump [mmdbId:13348] VAST, Links

NCBI **MMDB** Structure Summary

PubMed BLAST Structure Taxonomy OMIM **Help?** Cn3d

Reference: Kolbe M, Besir H, Essen LO, Oesterhelt D [Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution](#) *Science* v288, p. 1390-1396
[All References](#)

Description: Halorhodopsin, A Light-Driven Chloride Pump.
Deposition: 2000/4/6
Taxonomy: [Halobacterium salinarum](#)
MMDB: [13348](#) **PDB:** [1E12](#) **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

Ligand (x 10)

NCBI **VAST** Structure Neighbors

PubMed BLAST Structure Taxonomy OMIM **Help?** Cn3d

VAST neighbors for: MMDDB [13348](#), 1E12 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

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: 140; 20 representatives from the [Medium redundancy](#) subset displayed.

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

[1E12 A](#) Protein Family Bac_rhodopsin Ali_Len

[2JAF A](#) 239

[1H2S A](#) 219

Graphics
 Graphics
 Table
 Download Asn1
 Download Xml
 Entrez



VAST Structure Neighbors



PubMed BLAST Structure Taxonomy OMIM **Help?** **Cn3D**

VAST neighbors for: **MMDB 13348, 1E12 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

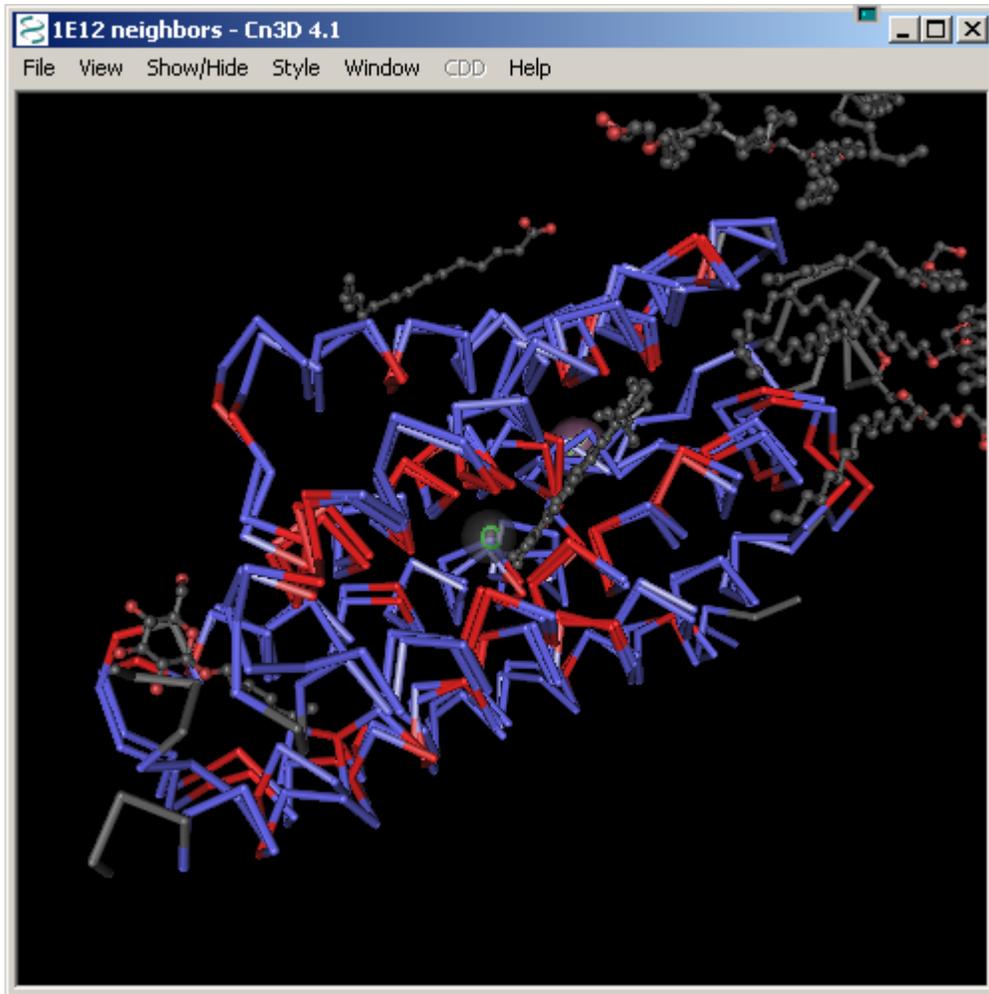
List **Medium redundancy** subset, sorted by **Aligned Length** in **Table**

Advanced neighbor search

Total neighbors: 140; 1 - 18 of 20 representatives from the **Medium redundancy** subset with 2 selected neighbors displayed.

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

	PDB	C	D	Ali. Len	Score	E_Val	Rmsd	%Id	MMDB Date	LHM	GSP	Description
<input type="checkbox"/>	2AF	A		239	16.0	10e-17.3	0.9	99.6	01/2007	0.0	0.4	Ground State Of Halorhodopsin T203v
<input checked="" type="checkbox"/>	1H2S	A		219	15.6	10e-15.9	1.4	27.4	11/2002	3.3	0.7	Molecular Basis Of Transmembrane Signalling By Sensory Rhodopsin li-Transducer Complex
<input type="checkbox"/>	1C3W	A		217	15.2	10e-15.1	1.6	33.6	03/2001	2.2	0.8	BacteriorhodopsinLIPID COMPLEX AT 1.55 A RESOLUTION
<input type="checkbox"/>	2F93	A		216	16.0	10e-17.0	1.3	27.8	05/2006	3.3	0.7	K Intermediate Structure Of Sensory Rhodopsin liTRANSDUCER Complex In Combination With The Ground State Structure
<input type="checkbox"/>	1XIO	A		209	11.9	10e-11.1	1.7	26.3	11/2004	4.5	0.8	Anabaena Sensory Rhodopsin



1E12 neighbors - Sequence/Alignment Viewer

View Edit Mouse Mode Unaligned Justification Imports

```

1E12_A avreNALLSSSLWVNVALAGIAILVFVYMGRTIRPG:PRLIWGATLMIPLVSISSYLGLLSGLTVGMIEmpaghala g eMVRSQW
1H2S_A ~~~~MVGLTTLFWLGAIGMLVGT LAFAWAGR DAGSG~ERRYYVTLVGI SGIAAVAYVVMALGVGWV PVA~~~~~ERTVFA
  
```

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.

```
>gij2851597|sp|P25848|PHY1_CERPU Light-sensor Protein kina:
MSATKKTY SSTTSAKSKHSVRVAQTTADAALAEVYEMSGDSC
QREGLIQNFGCMVAVEEPNFCVIA YSENA SEFLDLIPQAVPSMGE
AATQDISLLNPTVHCRRSGKPLYAJAHRIDIGVIDFEAVKMIDVPV
LPGGDIELLCDTVEEVRELTGYDRVMAFKFHEDEHGEVVAEIRR
KNRVRLIADCYASPVKLIQDPDIRQPVSLAGSTLRAPHGCHAQYI
IQGRKLVGLVV CQHTSPRTV PFPLRSVCEFLMQVFGMQLNLH
PIGIVSQTPNIMDLVKCDGAALYYGKR 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.