

# 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.

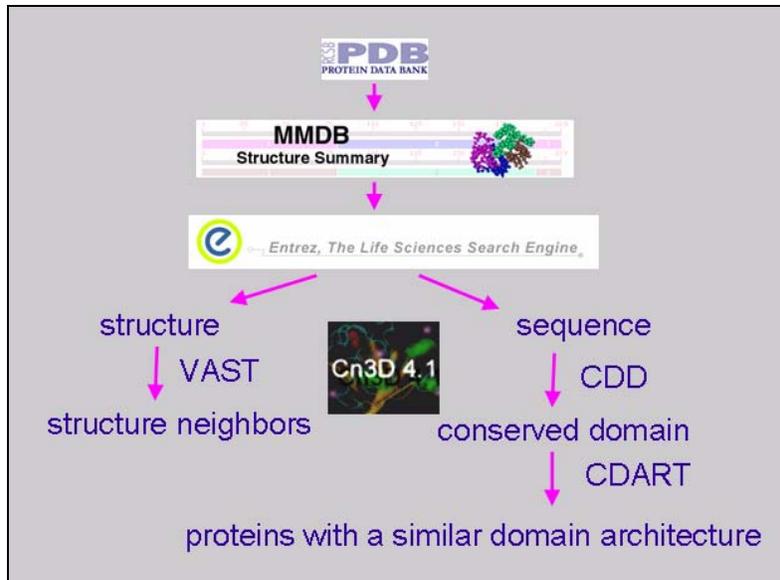
### **In this course, we will learn to**

- 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 following handout includes the screen shots of the exercise demonstrated in the mini-course.

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))



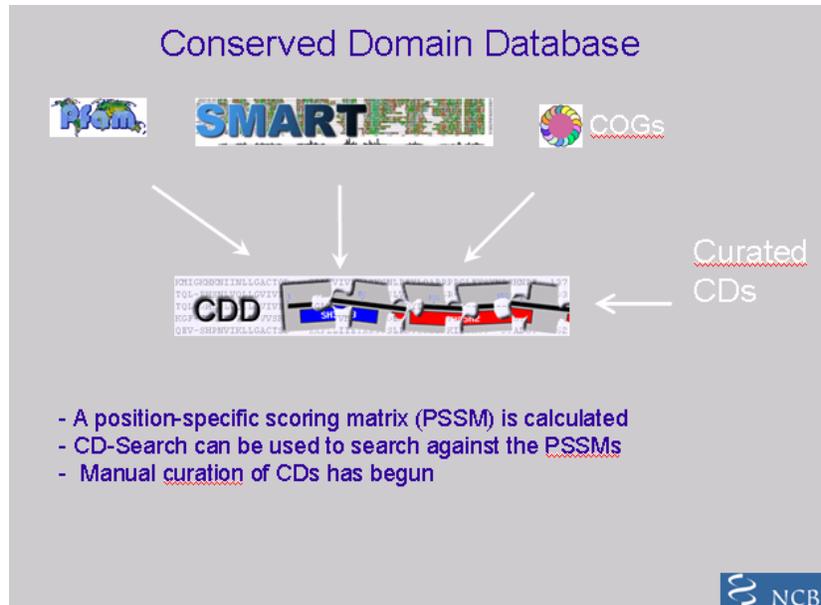
**CDD**

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

**Conserved Domain**

- recurring unit in molecular evolution, whose extents can be determined by sequence and structure analysis
- performs a particular function
- represented as a multiple local sequence alignment of proteins containing the domain

NCBI



## 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.

```

MDPIALTAAVGADLLGDGRPETLWLGIGTLLMLIGTFYFIVKGWG
SMFFGIGLTEVQVGSEMLDIYARYADWLFTTPLLLLDLALLAKV
HTPLARYTWWLFSTICMIVVLYFLATSLRAAAKERGPEVASTFN
VGLGIETLLFMVLDVTAKVGFIFILLRSRAILGDTEAPEPSAGAE/
  
```

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 a similar domain architecture by clicking on the "Show" 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 "Aligned rows" list box pull down menu to specify "up to 5" sequences. 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 of the similar protein from the CD-Browser page. Then click on the structure link on the top right side, then on 1E12, and finally on the chain A 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.

NCBI Conserved Domains

Search across Entrez databases

Submit Query

Search Database: CDD v2.09 - 12422 PSSMs

Enter a Protein query as Accession, GI, or Sequence in FASTA format:

NDPIALTAAVGADLLGDRPETLWLGIGTLLMIGTFYFIVKNGVTDKAEYYSITLVPGIASAAYL  
SMFFGLGTEVQVGSMLDIYYARYADWLFPTPLLLDLALLAKVDRVSICTLVQVDALMIVTGLVGLS  
HTPLARYTWLFTICMIVLVYFLATSLRAAAKERGPEVASTFNLTALVLVLTWYFILWIGTEGAGV

Find CDs in Entrez

Computational biologists define conserved domains based on recurring sequence patterns or motifs. The un-curated section of CDD contains domains imported from SMART, Pfam and COGs. The source databases also provide descriptions and links to citations. Because conserved domains correspond to compact structural units, CDs are linked to 3D structure when possible. The NCBI-curated section of CDD attempts to group ancient domains related by common descent into family hierarchies.

NCBI Conserved Domains

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

Concise Result Full Result Show Search Information

Bac\_rhodopsin

Descriptions

Title	Pssmid	Multi-Dom	E-value
Hjpfam01036, 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

National Library of Medicine National Institutes of Health

Search: All Databases for bacteriorhodopsin Go

What NCBI do?

Hot Spots

Assembly Archive

Clusters of orthologous groups

Coffee Break, Genes & Disease, NCBI Handbook

Electronic PCR

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

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.

About Entrez  
 Books  
 Overview  
 Using the books  
 Information for authors and publishers  
 Contact us  
 Mailing list

### 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.]

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

Concise Result  
  Full Result  
  Show Search Information  

Title	PssmId	Multi-Dom	E-value
[H]pfam01036, Bac_rhodopsin, Bacteriorhodopsin...	41106	No	1e-47

Search for similar domain architectures

### CDART: Conserved Domain Architecture Retrieval Tool

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

[About CDART](#)

**Similar domain architectures:**

AAQ16277 Chlamydomonas rein chlamyopsin-5  
 523 Sequences cellular proteins hypothetical prote



NCBI **Entrez** Protein

My NCBI (Sign In) (Register)

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 523 page 1 of 27 Previous Next

1: [XP\\_001217277](#), Reports conserved hypothe... [gi:115402401] BLink, Conserved Domains, Links

```
>gi|115402401|ref|XP_001217277.1| conserved hypothetical protein [Aspergillus terreus NIH2624]
MALSSLVYVILAAAPLSKRVFHSLSVIATTVSFIVVLALATQGLTWKHDLIREHKKHVPNVTEYYRQ
VFWRYLNNFLTEPLLANLALLSGLPGAHLLSAVVADYVMLGSGLLGTFAGHTARRVANFAISAIGYLT
TVYHIGINGSAAVNRKDVQIKRFFGTISAVTLFVKALYFVAIAAGPLAKMNVNAETVIFAIDYDFTQGI
IGVLLIAHDSPPGLTSLVDGFWSSGIGNEGSIREEEGA
```

2: [ZP\\_01449284](#), Reports bacteriorhodopsin... [gi:114771895] BLink, Conserved Domains, Links

```
>gi|114771895|ref|ZP_01449284.1| bacteriorhodopsin [alpha proteobacterium HTCC2255]
MKDVMLNPTDVGMSFWIISAAMLAATFFFFMERDRAVGKWKTSLSVAALVTGIAALHYFYHREVIMTG
ESPTTFRYVDWLLTVPQLIEFYLILTAIAVVRALFVKLLVASLVMLIFGYGFEIGTMNVULAFVISMAG
WLYIIEYVFVGEASKISAANGTAAASKKAFNALRLIVTIGWSIYPIGVVLGYTAEGTDDALNLVYMLADF
VNKIAFGVVIWAAATSGSEKA
```

3: [ZP\\_01447408](#), Reports bacteriorhodopsin... [gi:114769798] BLink, Conserved Domains, Links

```
>gi|114769798|ref|ZP_01447408.1| bacteriorhodopsin [alpha proteobacterium HTCC2255]
NTILSKKGFVALATVAAGILPASARHAATLAPNDYVGSFVLSMALIAATAFFFIETTRVEGKWKTSLT
VSGLVTLVAAYHYFYMRDVVSTGESPTVYRIDWLLITVPLLVVEFYLLRAITKVSIGIFWRLMIGTVV
MLVGGYAGEIGYISATLGLIIGMLGWATLYEIFAGEASKVAADAPASVQSASFSTRMIVITIGWAIYPL
GYFNGYNGAVSDEALNIIYVNFADVNNKIAFGVVIINVAVTESEK
```

NCBI **Conserved Domains**

HOME SEARCH SITE MAP NewSearch PubMed Nucleotide Protein Structure CDD Taxonomy Help

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
pfam01036, Bac_rhodopsin, Bacteriorhodopsin	41106	No	1e-47

Search for similar domain architectures

NCBI Conserved Domains

pfam01036.12 Bac\_rhodopsin, with user query added

Bacteriorhodopsin.

Links: Statistics: Structure:

Show Structure

Program: Cn3D

Drawing: All Atoms

Aligned Rows: up to 10

[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 members

```

1E12_A      6  .[16].LVFVYM.[1].RTIRPGRPLINGATLMPLVVISVYLGLLSGLTVGMIEMP.[11].SQMGRYLTWALSTPMI 98
query      21  .[16].FYFIVK.[1].WGVTDKEAREYYSITILVPGIASAAVLSMFFGIGLTVQVQG.[ 5].IYYARYADULFTTPLL 105
1U02_A     15  .[16].FYFIVK.[1].WGVTDKEAREYYSITILVPGIASAAVLSMFFGIGLTVQVQG.[ 5].IYYARYADULFTTPLL 99
1MOR_A     22  .[16].LVPVVK.[1].MGVSDPDAKKFYAITLVPVIAFTMYSMLLGGGLTVVPPG.[ 5].IYYARYADULFTTPLL 106
qi 114809   34  .[16].LLFVYM.[1].RGLDDPRAKLIASVITLVPVVISVYTGASGLTISVLEMP.[21].TWGGRYLTWALSTPMI 134
qi 461809   34  .[16].LLFVYM.[1].RNVEDPRAGLIVATLWVPLVISVYTGASGLTVSFLVEMP.[11].TPWGRYLTWALSTPMI 124
qi 2499383  29  .[16].LLFVAM.[1].RDIESPRAKLIWATMLVPLVISVYTGASGLTVGFLQMP.[11].SPWGRYLTWTFSTPMI 119
qi 1168614   4  .[16].AVLAYG  YTLVPEETREKRYVLLIAPGIAIAYALMALGFGSIQSEGH.[ 1].VYVURYVDULLTTPLEN 83
qi 2499387  14  .[16].LVFIAR.[1].WVSDQDRQKQFYIATLITAIAPVNYLALMALGFGVITIELG.[ 5].IYYARYADULFTTPLL 98
qi 2499386   7  .[16].LVFIAR.[1].WGETDSRQKQFYIATLITAIAPVNYLALMALGFLTIVEFA.[ 5].IYYARYADULFTTPLL 91

```

CDD Descriptive Items

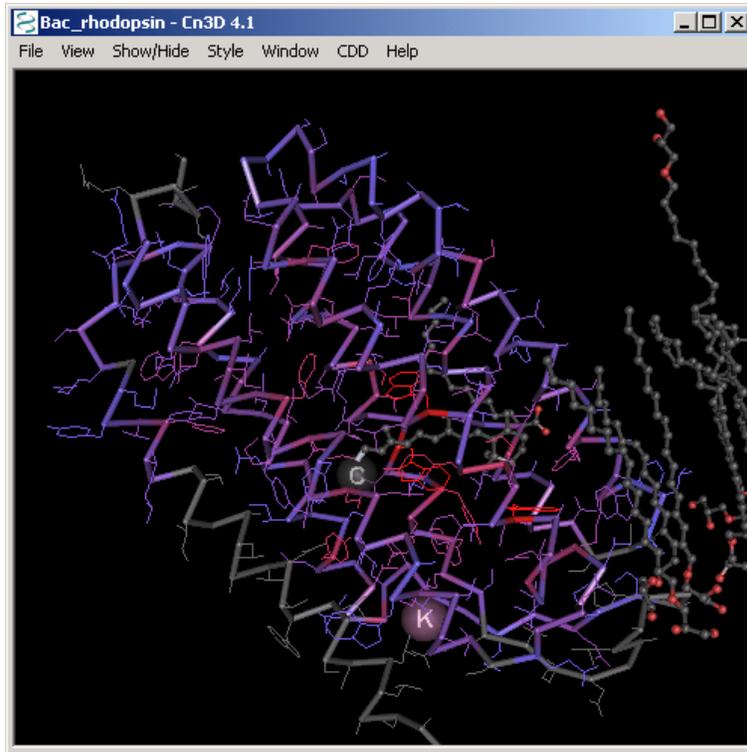
Name: Bac\_rhodopsin

Bacteriorhodopsin.

Structure summary:

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

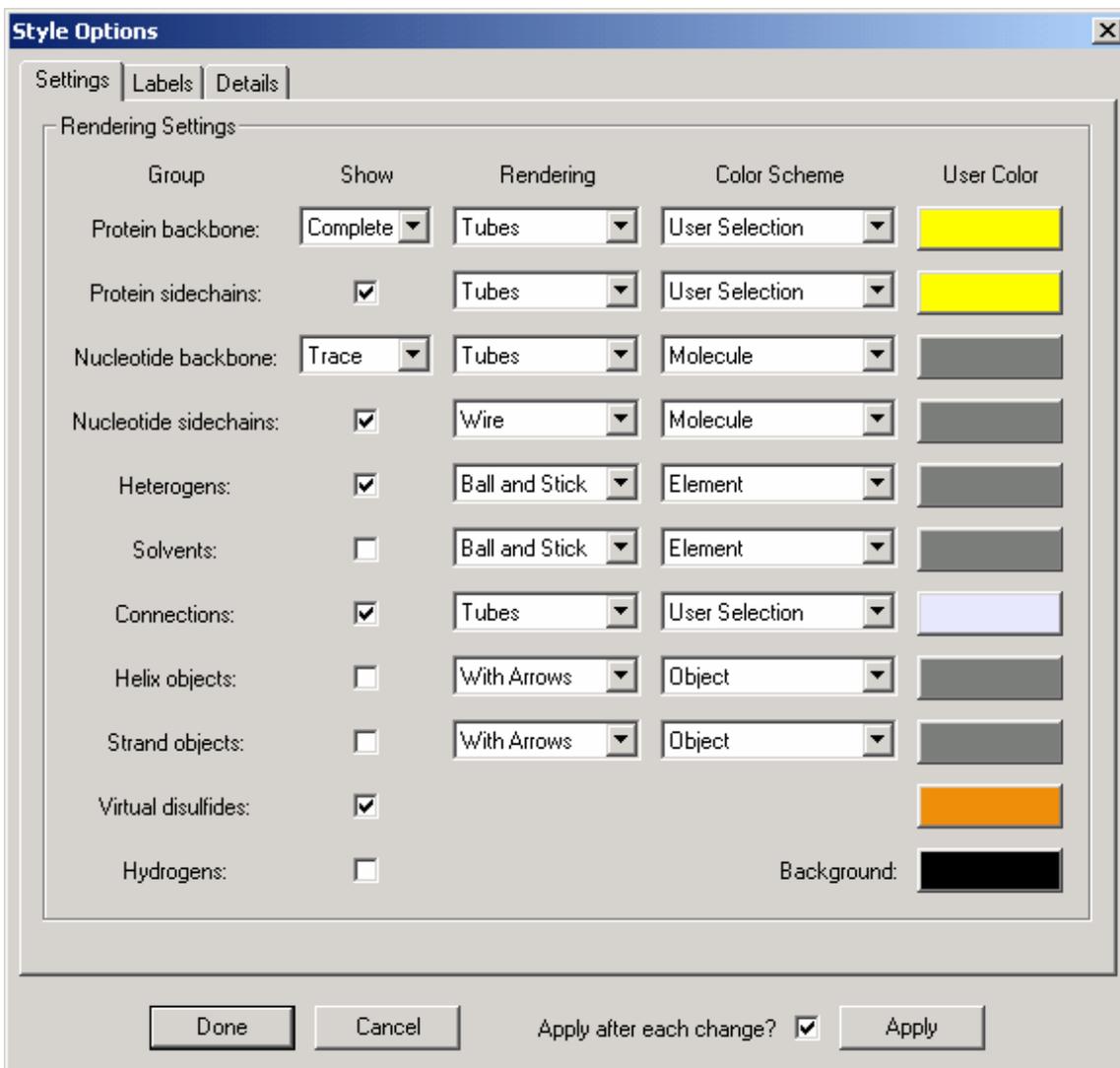
Show Annotations Panel Show References Panel Dismiss

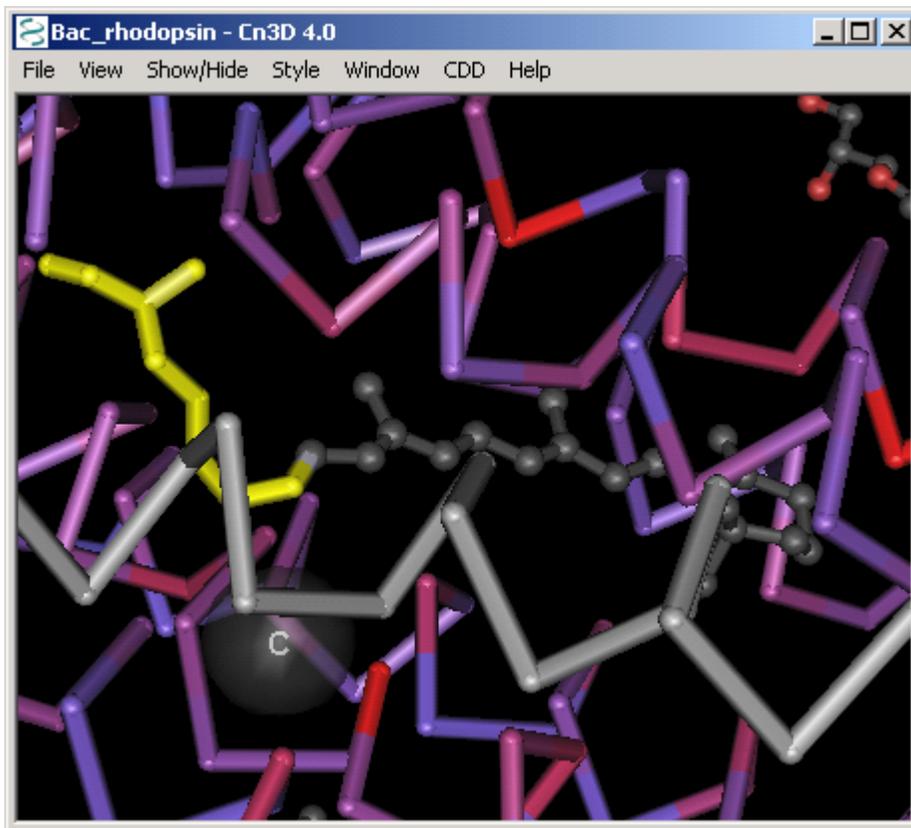


Bac\_rhodopsin - Sequence/Alignment Viewer

View Edit Mouse Mode Unaligned Justification Imports

<i>IB12_A</i>	LIWGATLMIPLVSISSYLGLLSGLTVGMIEMPAGH a l a g e m v r S Q W G R Y L T W A L S T P M I L L A L G L L A D V D L G S L F T V I A A D I G
query	EYYSITILVPGIASAAYLSMFFGIGLTEVQVGSEMI ~~~~~ d I Y Y A R Y A D W L F T T P L L L L D L A L L A K V D R V S I G T L V G V D A L
<i>IUAZ_A</i>	EYYSITILVPGIASAAYLSMFFGIGLTEVQVGSEMI ~~~~~ d I Y Y A R Y A D W L F T T P L L L L D L A L L A K V D R V S I G T L V G V D A L
<i>IMOK_A</i>	KFYA I T T L V P A 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 W A R Y A D W L F T T P L L L L D L A L L V D A D Q G T I L A L V G A D G I
gi 2499387	KFYIATIMIAAIAFVNYLSMALGFVTTIELGGEE r ~~~~~ a I Y W A R Y T D W L F T T P L L L Y D L A L L A G A D R N T I Y S L V G L D V L





NCBI Conserved Domains

pfam01036.12 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: C095524

Reformat Sequence Alignment

Format: Compact Hypertext

Row Display: Up to 5

Color Bits: 2.0 bits

Type Selection: the most similar members

```

1E12_A      8  .[16].LVFYMGRTIRPGRPLIIGATLMIPLVSISSYLGLLSGLTVGHMNPAG.[8].SQWRYLTWALSTPNILL 100
query      21  .[16].FFYIVKGVGVDKEAREYYSITILVPGIASAAYLSMFFGIGLTVQVQSEM.[2].IYARYADWLFPTPLLL 107
1UA2_A     15  .[16].FFYIVKGVGVDKEAREYYSITILVPGIASAAYLSMFFGIGLTVQVQSEM.[2].IYARYADWLFPTPLLL 101
1MOK_A     22  .[16].LVFLVKGWVSDPPDAKIFYAITTLVPAIATNYLSMLLGYGLTHVFPQGEQ.[2].IYARYADWLFPTPLLL 108
gi 2499387 14  .[16].LVFIARGVSDQRQRKFYIATINIAAIAFVNYLSMALGFGVTTIELGEE.[2].IYARYADWLFPTPLLL 100

1E12_A     101  ALGLLADVLDGSLFTVIAADIGMCTVGLAAAMT.[1].SALLFRWAFYAISCAFFVVVLSALVTDWAASASSA  GT 171
query      108  DLALLARVDRVSIOTLVGVVDALMIVTGLVGLS  HTPLARYTWLFTSTICIVLVLYFLATSLRAAAKER.[2].EV 179
1UA2_A     102  DLALLARVDRVSIOTLVGVVDALMIVTGLVGLS  HTPLARYTWLFTSTICIVLVLYFLATSLRAAAKER.[2].EV 173
1MOK_A     109  DLALLVDADQGITLALVGADGIMIGTGLVGLT  KVSYRPFVWAISTAANLYLYLVLFPGTSAESH.[2].EV 180
gi 2499387 101  DLALLAGADNTFVSLVGLVLMIGTGLATLS.[6].PAGAEPLVWGISTGFLVLLTFLFSLNLTDASEL.[2].DL 178

```



NCBI Sequence Viewer v2.0 - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&doptcmd=GenPept&db=Prote

Getting Started Latest Headlines

NCBI Protein

My NCBI Sign In Register

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for 1E12A Go Clear

Limits Preview/Index History Clipboard Details

Display GenPept Show 20 Send to

Range: from begin to end Features:  SNP  CDD  MGC  HPRD  STS  tRNA Refresh

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

BLink, Conserved

Links

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

Comment Features Sequence

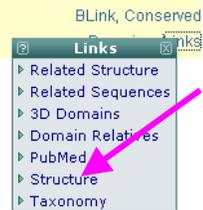
LOCUS 1E12\_A 253 aa linear BCT 06-APR-2000

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



NCBI Entrez Structure

Entrez 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 Text

1: 1E12  
Halorhodopsin, A Light-Driven Chloride Pump  
[mmdbId:13348] [MMDB, Links](#)

About Entrez  
Entrez Structure Help | FAQ  
Structure Research The NCBI Structure group  
MMDB

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**  
**Domain Family**

Chain A  
Bac\_rhodopsin

**Ligand**

NCBI **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 **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: 125; 1 - 60 of 101 representatives from the **Medium redundancy** subset displayed. Page: 1

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

[1E12 A](#)  
Protein Family

**Bac\_rhodopsin**

Accession	Length
<input type="checkbox"/> <a href="#">1VGO B</a>	229
<input type="checkbox"/> <a href="#">1BRK C</a>	227
<input type="checkbox"/> <a href="#">1VGO B</a>	227
<input type="checkbox"/> <a href="#">1QHJ B</a>	226
<input type="checkbox"/> <a href="#">1QKQ B</a>	226
<input type="checkbox"/> <a href="#">1QKP B</a>	226
<input type="checkbox"/> <a href="#">1BRK B</a>	226
<input type="checkbox"/> <a href="#">1BRK B</a>	226

NCBI **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 **Graphics**

Advanced neighbor search

*(A pink arrow points to the 'List' button, and a pink box highlights the dropdown menu options: Graphics, Graphics, 1E12 A, Download Asn1, Download Xml, Entrez.)*

NCBI **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  with  Display  [Download Cn3D!](#)

View Sequence Alignment using  for  VAST neighbors

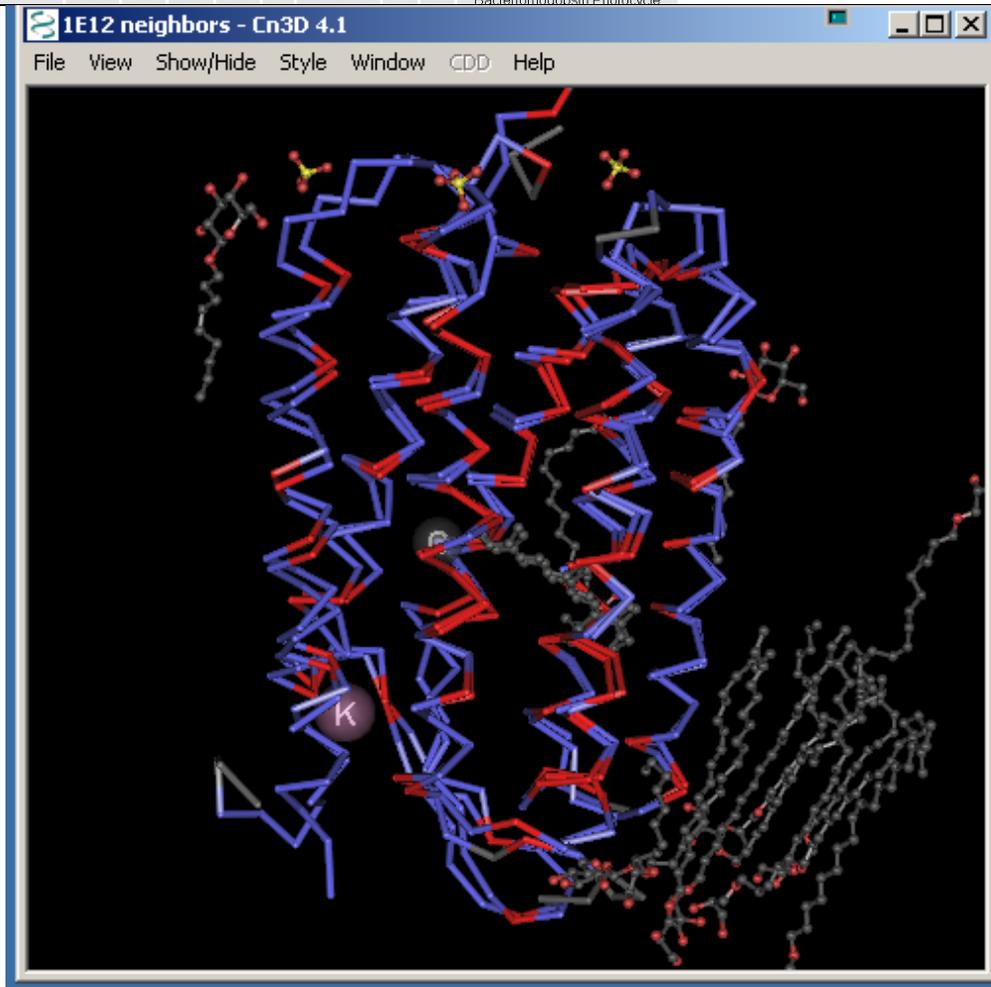
List  subset, sorted by  in

Advanced neighbor search

Total neighbors: 125; 1 - 60 of 101 representatives from the [Medium redundancy](#) subset displayed. Page:

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

	PDB	C	D	Ali. Len	Score	E_Val	Rmsd	%ld	MMDB Date	LHM	GSP	Description
<input checked="" type="checkbox"/>	<a href="#">1VGO</a>	<a href="#">B</a>		229	15.7	10e-16.1	2.0	33.2	10/2005	2.5	0.9	Crystal Structure Of Archaeorhodopsin-2y
<input type="checkbox"/>	<a href="#">1BRR</a>	<a href="#">C</a>		227	14.5	10e-12.3	1.9	33.5	03/2001	2.1	0.9	X-Ray Structure Of The Bacteriorhodopsin TrimerLIPID Complex
<input checked="" type="checkbox"/>	<a href="#">1VGO</a>	<a href="#">A</a>		227	15.2	10e-15.1	1.9	33.0	10/2005	2.4	0.9	Crystal Structure Of Archaeorhodopsin-2y
<input type="checkbox"/>	<a href="#">1QHJ</a>	<a href="#">A</a>		226	15.0	10e-13.8	1.7	33.6	03/2001	2.1	0.8	X-Ray Structure Of Bacteriorhodopsin Grown In Lipidic Cubic Phases
<input checked="" type="checkbox"/>	<a href="#">1QKO</a>	<a href="#">A</a>		226	15.2	10e-14.2	1.7	33.6	03/2001	2.1	0.8	High Resolution X-Ray Structure Of An Early Intermediate In The Bacteriorhodopsin Photocycle



1E12 neighbors - Sequence/Alignment Viewer  
View Edit Mouse Mode Unaligned Justification Imports

```
1E12_A ~ ~ ~ a v R E N A L L S S L W W N V A L A G I A I L V F V Y M G R T I R P G R P R L I W G A T L M I P L V S I S S Y L G L L S G L T V G M I E M P a g h a I A G E M V  
1VGO_B g f d I I N D G R P E T L W L G I G T L L M L I G T F Y F I A R G W G V T D K E A R E Y Y A I T I L V P G I A S A A Y L A M F F G I G V T E V E L A ~ ~ ~ ~ S G T V L
```

## 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 SSTTSA KSKHSV RVA QTTADA ALEA VY EMSGD SG
QREGLIQNF GCMV A V EEPNFCV IA Y SENA SEFLDLIPQA V PSMGE
AA TQDISLLN PTTV HCRRSGK PLYA IA HRIDIGIV IDFEA V KMIDV PV
LPGGDIELLCD TIV EEV RELTGY DRV MAFK FHEDEHGEV V A EIRR
KNRVRLIADCY A SPVKLIQDPDIRQPV SLA GSTLRA PHGCHA QYI
IQRGRKLWGLV V CQHTSPRTV PFPLRSV CEFLMQV FGMQLNLH
PIGIVSQTPNIMDLV KCDGAA LYY GKRVWLLGTTPTENQIKEIA DV
HLLGDA VCGMAA AKITAKDFLFWFRSHTA TEV KWGGAKHDPDE
EDVEMDA IHSLQLILRGSFRDIA DSDTKTMIHARLNDLKLQGV EER
```

- What are the domains present in this protein?

-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 a similar domain architecture by clicking on the "Show" 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".

- 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|29142). Use the "Aligned rows" list box pull down menu to specify "up to 5" sequences. 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.