Mini Courses NCBI Resources one course at a time

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' 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

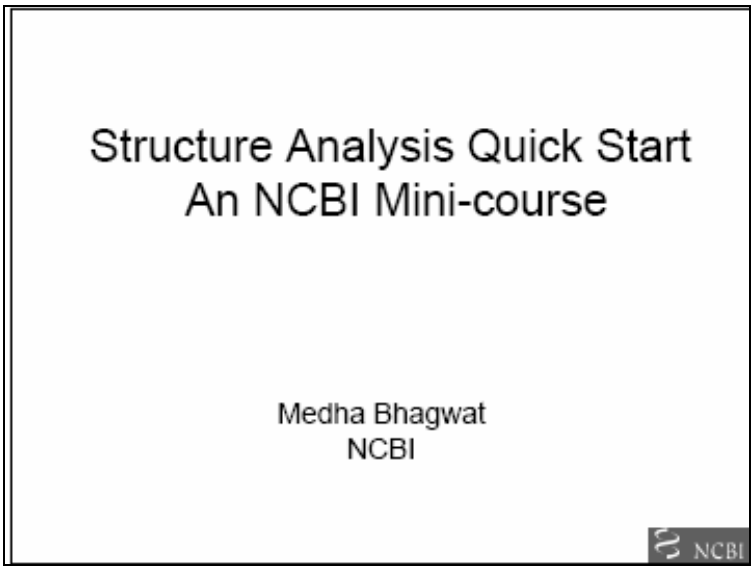
- Identify a conserved domain present in the query protein using **CDD**
- Search for other proteins containing similar domain(s) using **CDART**
- Explore a 3D modeling template for the query sequence using **CDD**
- Find similar structures using **VAST**
- Visualize and annotate the 3D protein structures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

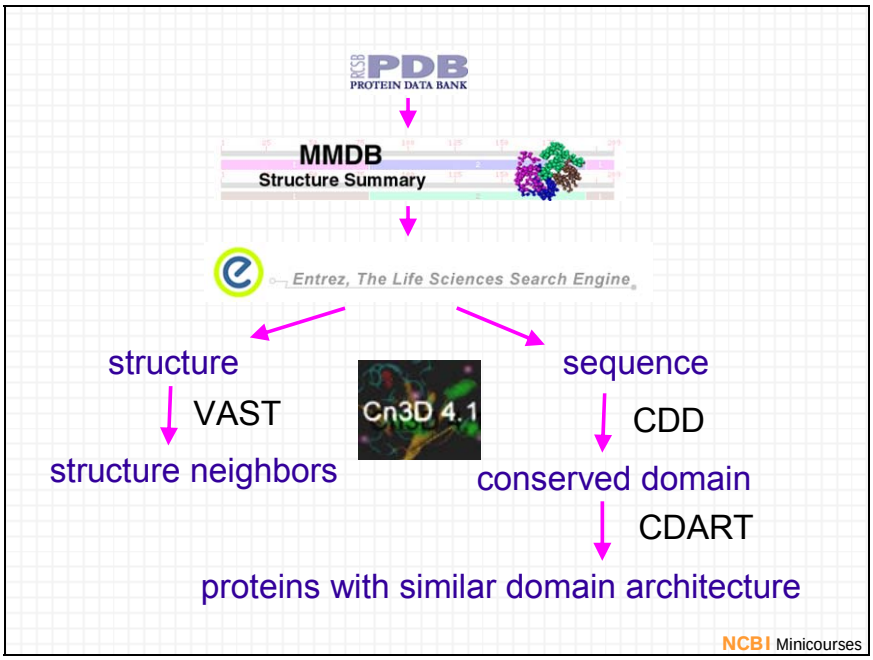
The remainder of the handout includes the introductory slides and the screen shots of the exercise demonstrated in Problem 1.

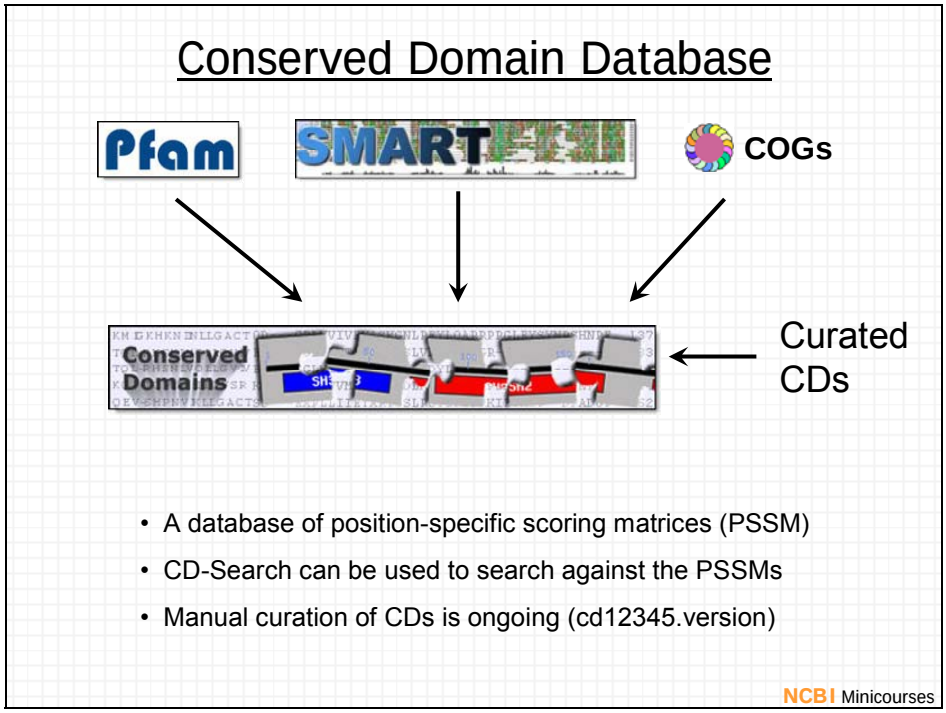
URL: http://www.ncbi.nlm.nih.gov/Class/minicourses/quickstructure.html

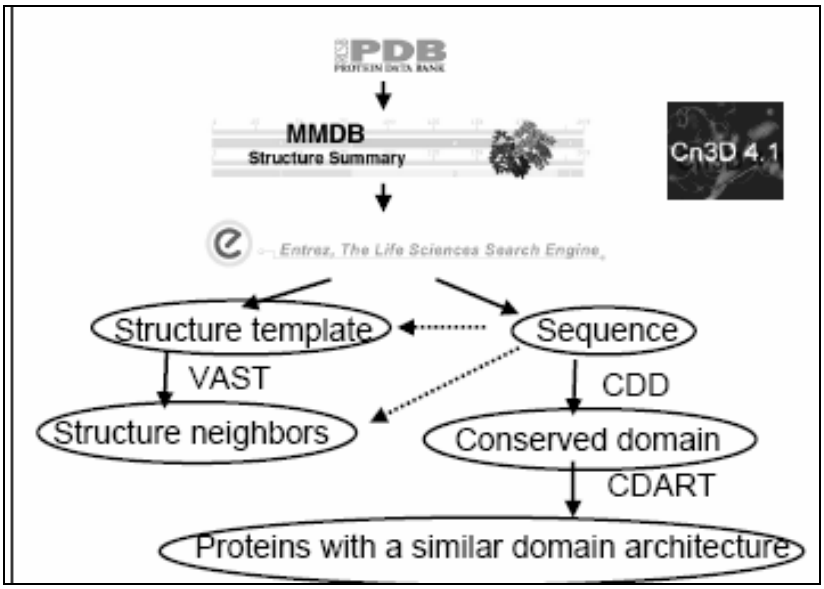
Course developed by: Dr. Medha Bhagwat (bhagwat@ncbi.nlm.nih.gov)

Slides




PDB
PROTEIN DATA BANK
MMDB
Structure Summary
Entrez, The Life Sciences Search Engine
structure
VAST
structure neighbors
Cn3D 4.1
sequence
CDD
conserved domain
CDART
proteins with similar domain architecture
NCBI Minicourses


Conserved Domain Database
Pfam
SMART
COGs
Conserved Domains
Curated CDs
• 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



Outline

For a query protein:

1. Identify the conserved domain(s) present in it.
2. Search for other proteins containing similar domain(s).
3. Explore a 3D modeling template.
4. Find distant sequence homologs.

Problem 1

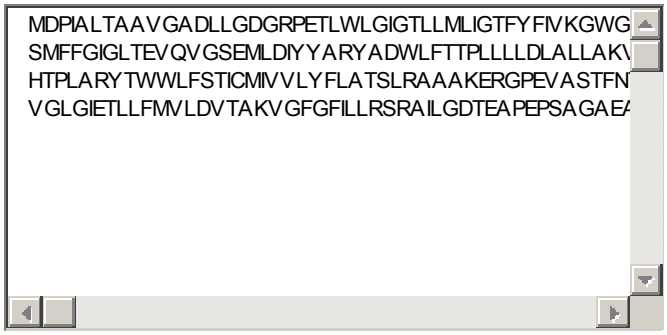
In this problem, we will follow these steps:

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”.

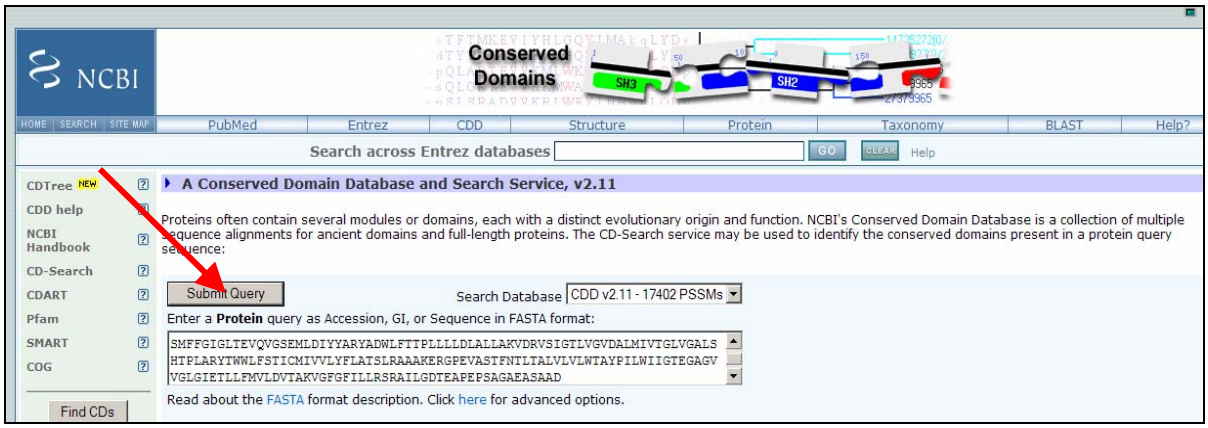
C. Go back to the CD-Search results page. Generate a multiple sequence alignment for the top 10 sequences representative of the conserved domain hit by clicking on the graphic of the domain. Use the "Row Display" list box pull down menu to specify "up to 5" sequences and reformat sequence alignment. 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.

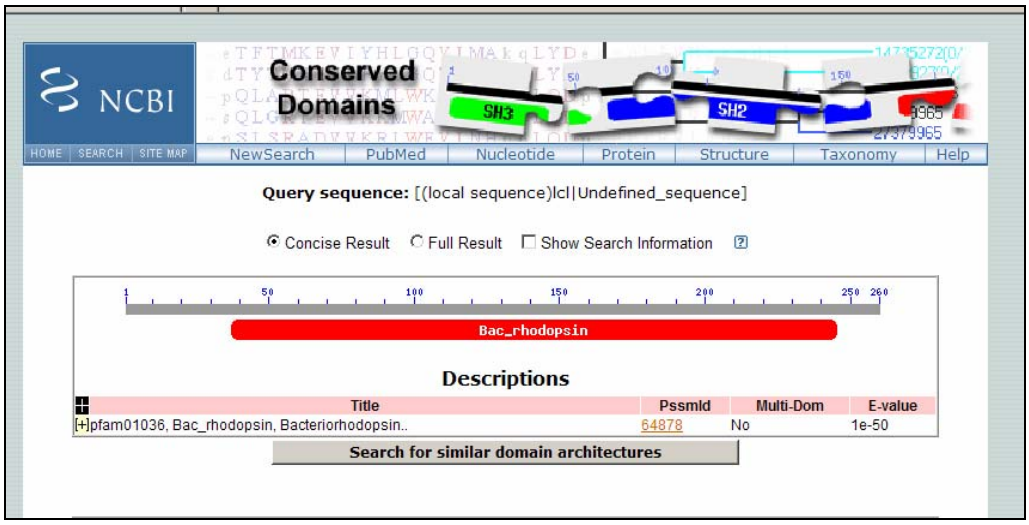
The structure of the *Halobacterium salinarum* mutant halorhodopsin protein and its sequence alignment with our query protein are displayed. For a better view of the backbone, remove the side chains globally (Style--Edit global style--Protein side chains). The query protein contains a bacterial rhodopsin 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.

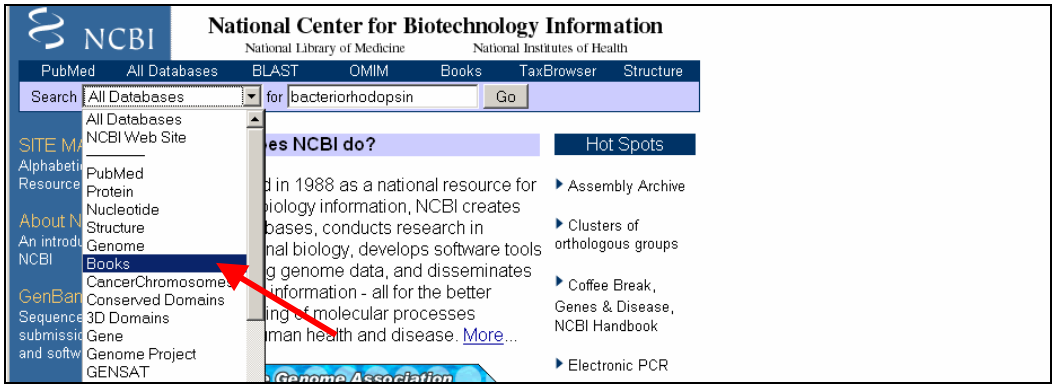
Display the side chains of this residue (Use Style--Annotate--New--Edit Style. Change the protein backbone Rendering to Tubes, Color Scheme to User Selection and User Color to choose the color for the highlighted residue, for example yellow. Repeat these steps for the Protein Side chains row and click 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.

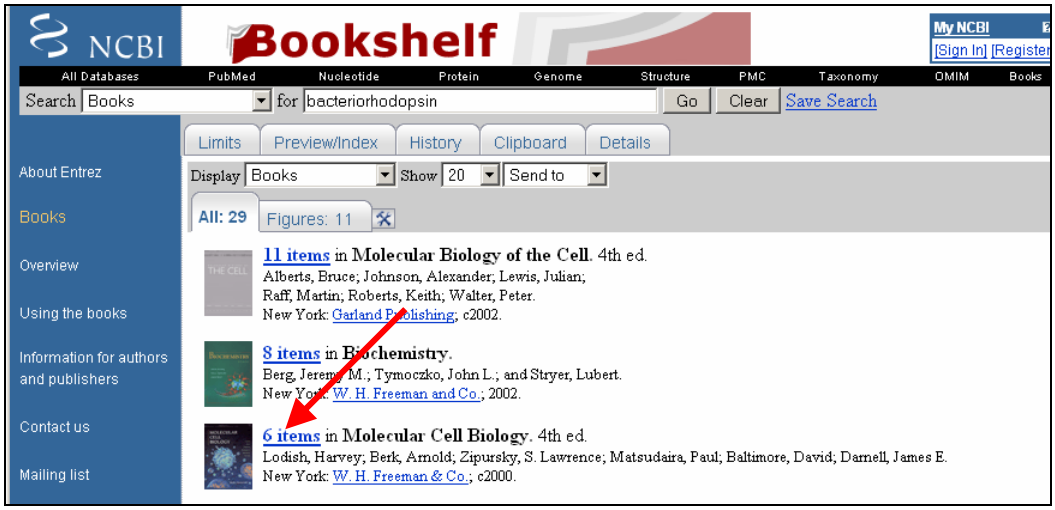
D. To obtain the structural neighbors for the halorhodopsin protein, first click on the structure entry link, 1TNO_A, on the CD-Browser page. Then click Links → Structure on the top right, then on 1TNO again in the Entrez Structure page, and finally on the chain A graphic. To view neighbors with 1TNO_A,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

Screenshots




NCBI
Conserved Domains
SH3
SH2
HOME
SEARCH
SITE MAP
NewSearch
PubMed
Nucleotide
Protein
Structure
Taxonomy
Help
Query sequence: [(local sequence)lcl|Undefined_sequence]
Concise Result
Full Result
Show Search Information
Bac_rhodopsin
Descriptions
Title
PssmId
Multi-Dom
E-value
[+]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
All Databases
NCBI Web Site
PubMed
Protein
Nucleotide
Structure
Genome
Books
CancerChromosomes
Conserved Domains
3D Domains
Gene
Genome Project
GENSAT
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
About Entrez
Books
Overview
Using the books
Information for authors and publishers
Contact us
Mailing list
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

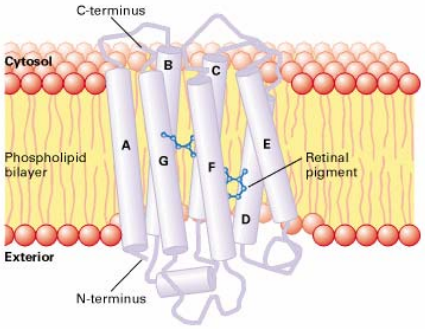
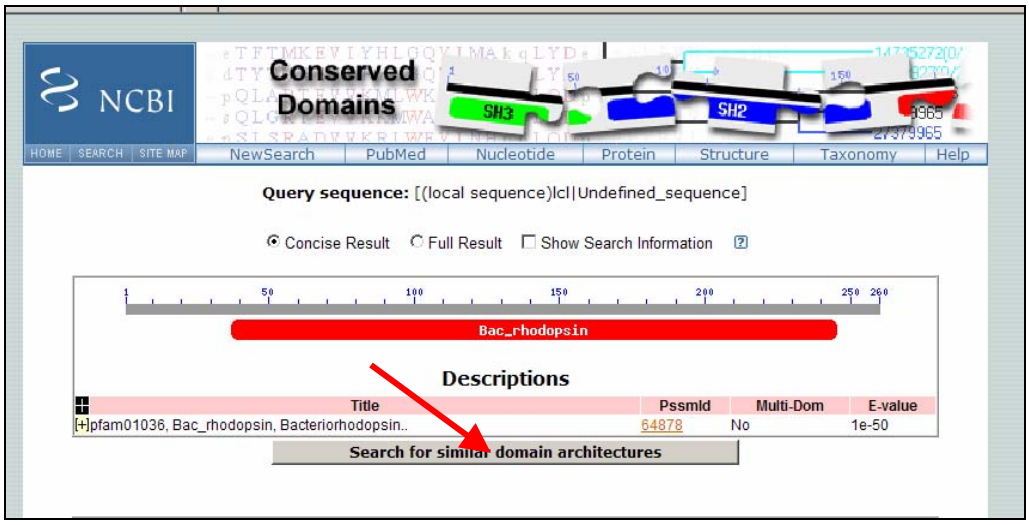
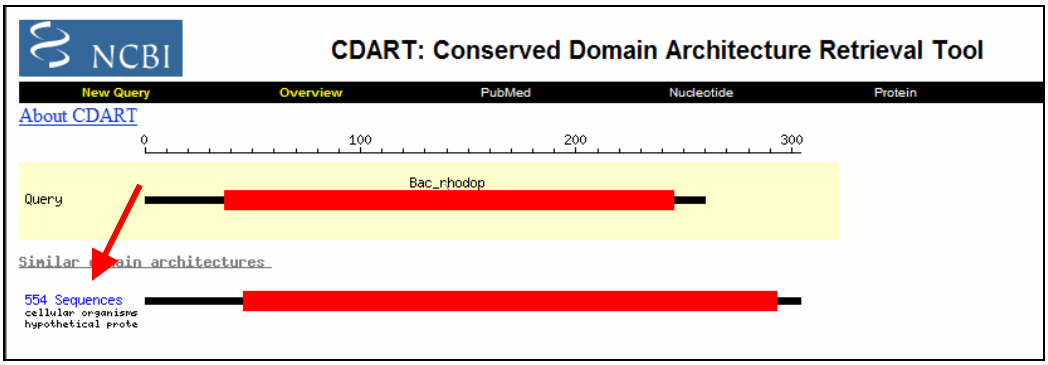


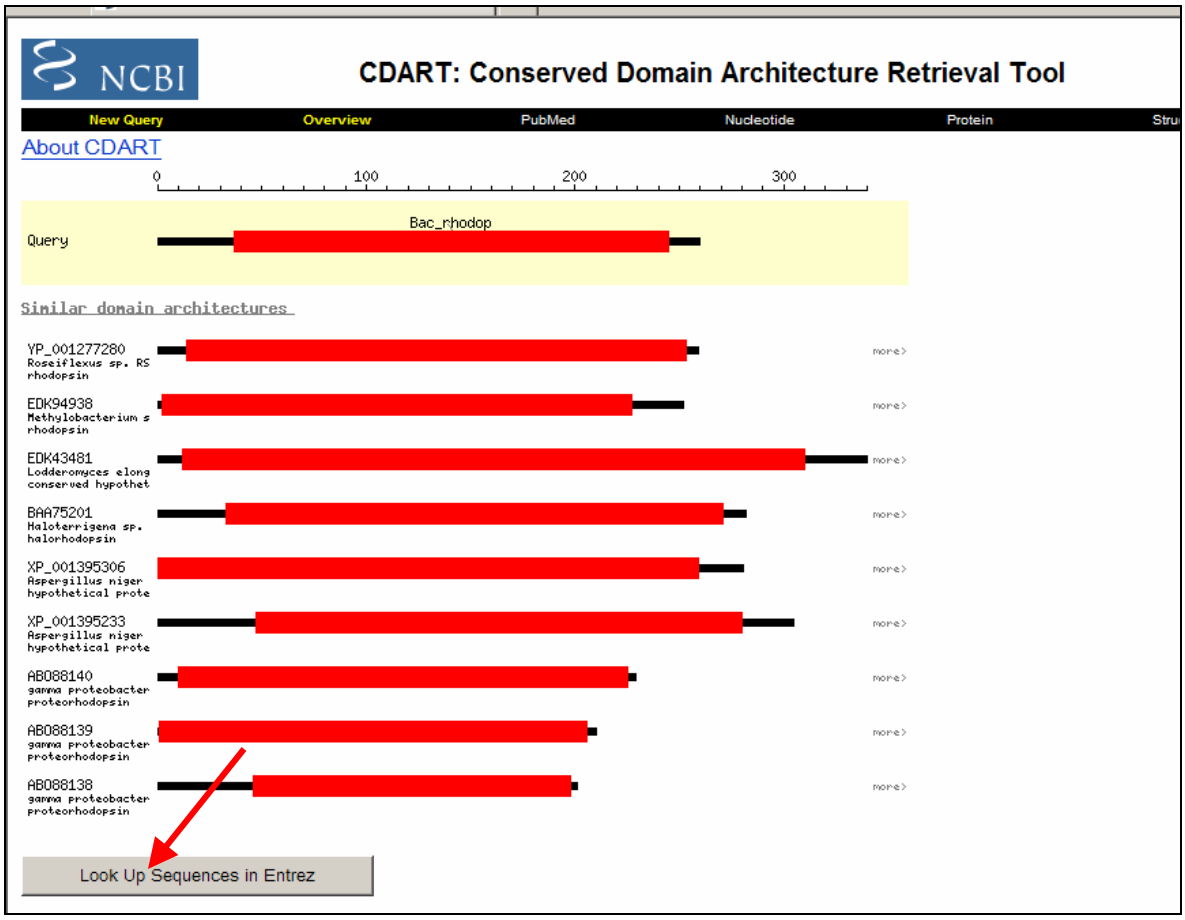
Figure 3-34. Overall structure of bacteriorhodopsin as deduced from electron diffraction analyses of two-dimensional crystals of the protein in the bacterial membrane. The seven membrane-spanning α helices are labeled A–G. The retinal pigment is covalently attached to lysine 216 in helix G. The approximate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

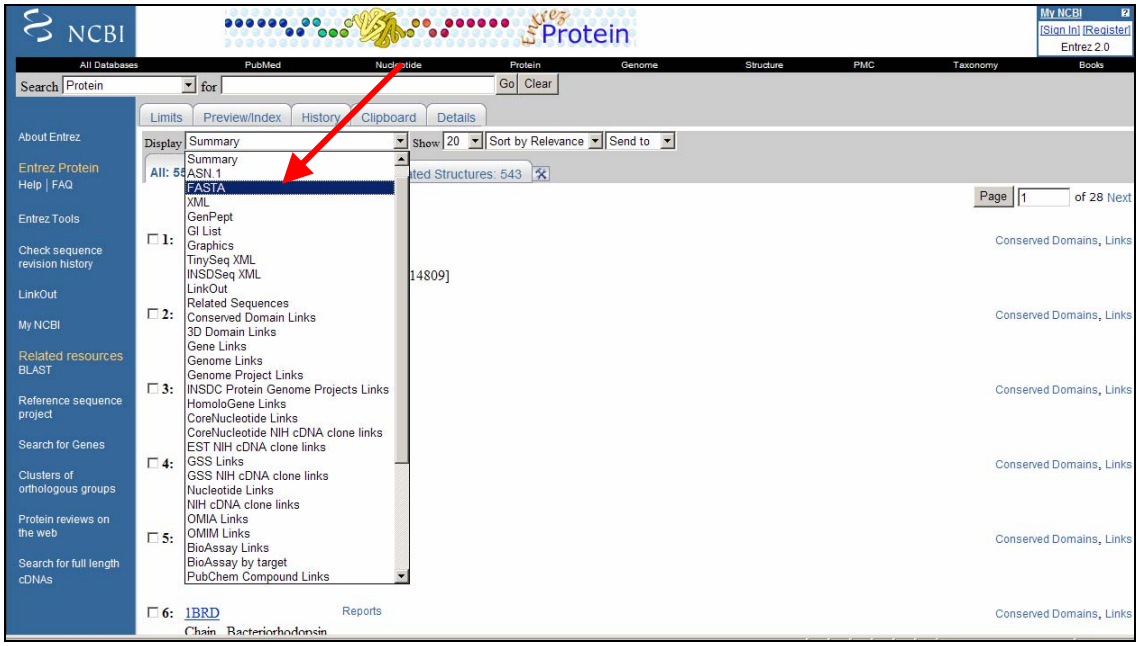
Navigation

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


NCBI
Conserved Domains
SH3
SH2
HOME
SEARCH
SITE MAP
NewSearch
PubMed
Nucleotide
Protein
Structure
Taxonomy
Help
Query sequence: [(local sequence)lcl|Undefined_sequence]
Concise Result
Full Result
Show Search Information
Bac_rhodopsin
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
About CDART
Bac_rhodop
Query
554 Sequences
cellular organisms
hypothetical prote


NCBI
CDART: Conserved Domain Architecture Retrieval Tool
New Query
Overview
PubMed
Nucleotide
Protein
About CDART
0
100
200
300
Bac_rhodop
Query
Similar domain architectures
YP_001277280
Roseiflexus sp. RS
rhodopsin
EDK94938
Methylobacterium s
rhodopsin
EDK43481
Lodderomyces elong
conserved hypothet
BAA75201
Haloterrigena sp.
halorhodopsin
XP_001395306
Aspergillus niger
hypothetical prote
XP_001395233
Aspergillus niger
hypothetical prote
AB088140
gamma proteobacter
proteorhodopsin
AB088139
gamma proteobacter
proteorhodopsin
AB088138
gamma proteobacter
proteorhodopsin
more>
Look Up Sequences in Entrez


NCBI
Entrez Protein
My NCBI
[Sign In] [Register]
Entrez 2.0
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
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
INSDC Protein Genome Projects Links
HomoloGene Links
CoreNucleotide Links
CoreNucleotide NIH cDNA clone links
EST NIH cDNA clone links
GSS Links
GSS NIH cDNA clone links
Nucleotide Links
NIH cDNA clone links
OMIA Links
OMIM Links
BioAssay Links
BioAssay by target
PubChem Compound Links
Page 1 of 28 Next
Conserved Domains, Links
1BRD
Reports
About Entrez
Entrez Protein
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI
Related resources
BLAST
Reference sequence project
Search for Genes
Clusters of orthologous groups
Protein reviews on the web
Search for full length cDNAs

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 5 Send to

Item 1 - 5 of 554 page 1 of 111 Previous Next

1: P15647. Reports Halorhodopsin (HR...[gi:114809] BLink, Conserved Domains Links

Next sequence

```
>gi|114809|sp|P15647|BACH_NATPH Halorhodopsin (HR)
MTETLPPVTESAVALQAEVTQRELFEFVLNDPLLASSLYINIALAGLSILLFVFMTRGLDDPRAKLIAVS
TILVPVVSIASYTGLASGLTISVLEMPAGHFAEGSSVMLGGEEVDGVVTMWGRYLTWALSTPMILLALGL
LAGSNATKLFTAITFDIAMCVTGLAAALTTSSHLMRWFWYAISCACFLVVLYILLVEWAQDAKAAGTADM
FNTLKLLTVVMWLGYPIVWALGVEGIAVLPVGVTSWGYSFLDIVAKYIFAFLLLNYLTSNESVVSGSILD
VPSASGTPADD
```

2: AAA72184. Reports bacteriorhodopsin...[gi:208055] BLink, Conserved Domains Links

Previous sequence Next sequence

```
>gi|208055|gb|AAA72184.1| bacteriorhodopsin
MQAQITGRPEWIWLALGTALMGLGTLYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMV
PFGGEQNPIYWARYADWLFTTPLLLLDLALLVDADQGTILALVGADGIMIGTGLVGALTKVYSYRFVWWA
ISTAAMLYILYVLFFGFTSKAESMRPEVASTFKVLRNVTVVLWSAYPVVWLIGSEGAGIVPLNIETLLFM
VLDVSAKVGFGLILLRSRAIFGEAEAPEPSAGDGAAATS
```

3: 0501217A. Reports bacteriorhodopsin...[gi:223063] BLink, Conserved Domains Links

Previous sequence Next sequence

```
>gi|223063|prf||0501217A bacteriorhodopsin
EAQITGRPEWIWLALGTALMGLGTLYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVP
FGGEQNPIYWARYADWLFTTPLLLLDLALLVDADEGTILAIVGADGLMIGTGLVGALTKVYSYRFVWAIS
TAAMSYILYVLFFGFTSKAESMRPEVASTFKVLRNVTVVLWSAYPVVWLIGSEGAGIVPLNIETALFMVL
DVSAKVGFGLILLRSRAIFGEAEAPEPSAGDGAAATS
```

4: 0502635B. Reports bacteriorhodopsin...[gi:223077] BLink, Conserved Domains Links

Previous sequence Next sequence

```
>gi|223077|prf||0502635B bacteriorhodopsin N term fragment
EAQITGRPEWIWLALGTALMGLGTLYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVP
FGGEQNPIYWARYADWLFTTPLLLLDLALLVDADEGTILAIVGADGLMIGTGLVGALTKVYSYRFVWAIS
TAAM
```

5: 0712242A. Reports bacteriorhodopsin...[gi:223370] BLink, Conserved Domains Links

Previous sequence

```
>gi|223370|prf||0712242A bacteriorhodopsin precursor
MLELLPTAVEGVSQAQITGRPEWIWLALGTALMGLGTLYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMY
LSMLLGYGLTMVPFGGEQNPIYWARYADWLFTTPLLLLDLALLVDADQGTILALVGADGIMIGTGLVGAL
TKVYSYRFVWWAISTAAMLYILYVLFFGFTSKAESMRPEVASTFKVLRNVTVVLWSAYPVVWLIGSEGAG
IVPLNIETLLFMVLDVSAKVGKGLILLRSRAIFGEAEAPEPSAGDGAAATSD
```

NCBI Conserved Domains

HOME | SEARCH | SITE MAP

NewSearch | PubMed | Nucleotide | Protein | Structure | Taxonomy | Help

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

Concise Result Full Result Show Search Information

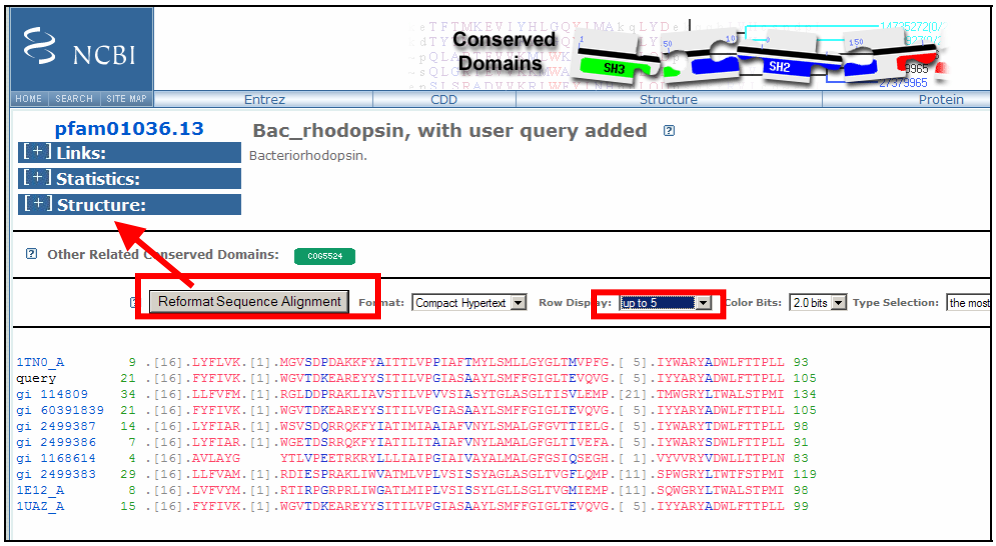
1 50 100 150 200 250 260

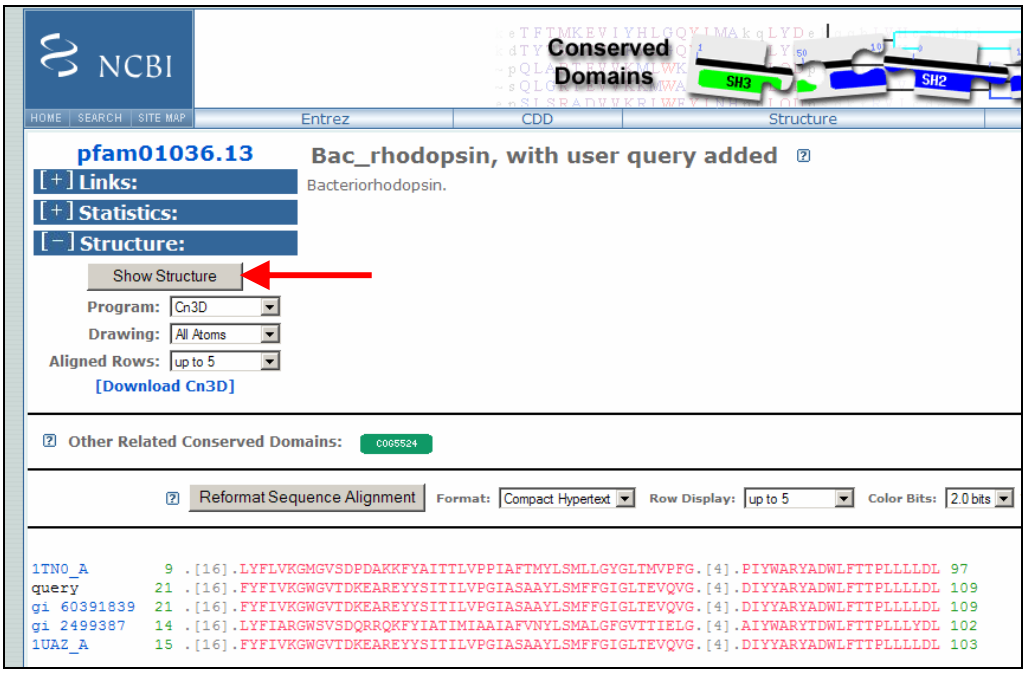
Bac_rhodopsin

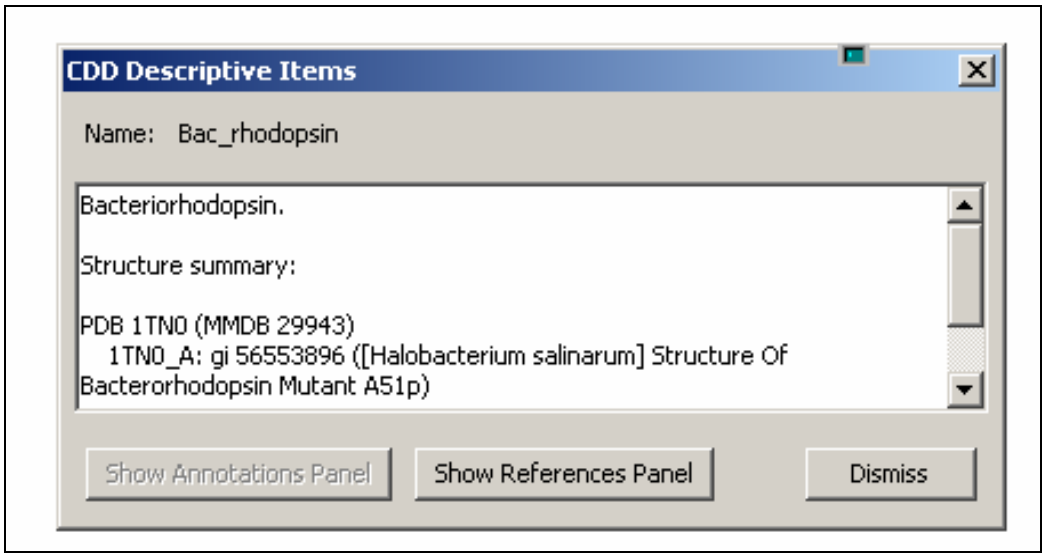
Descriptions

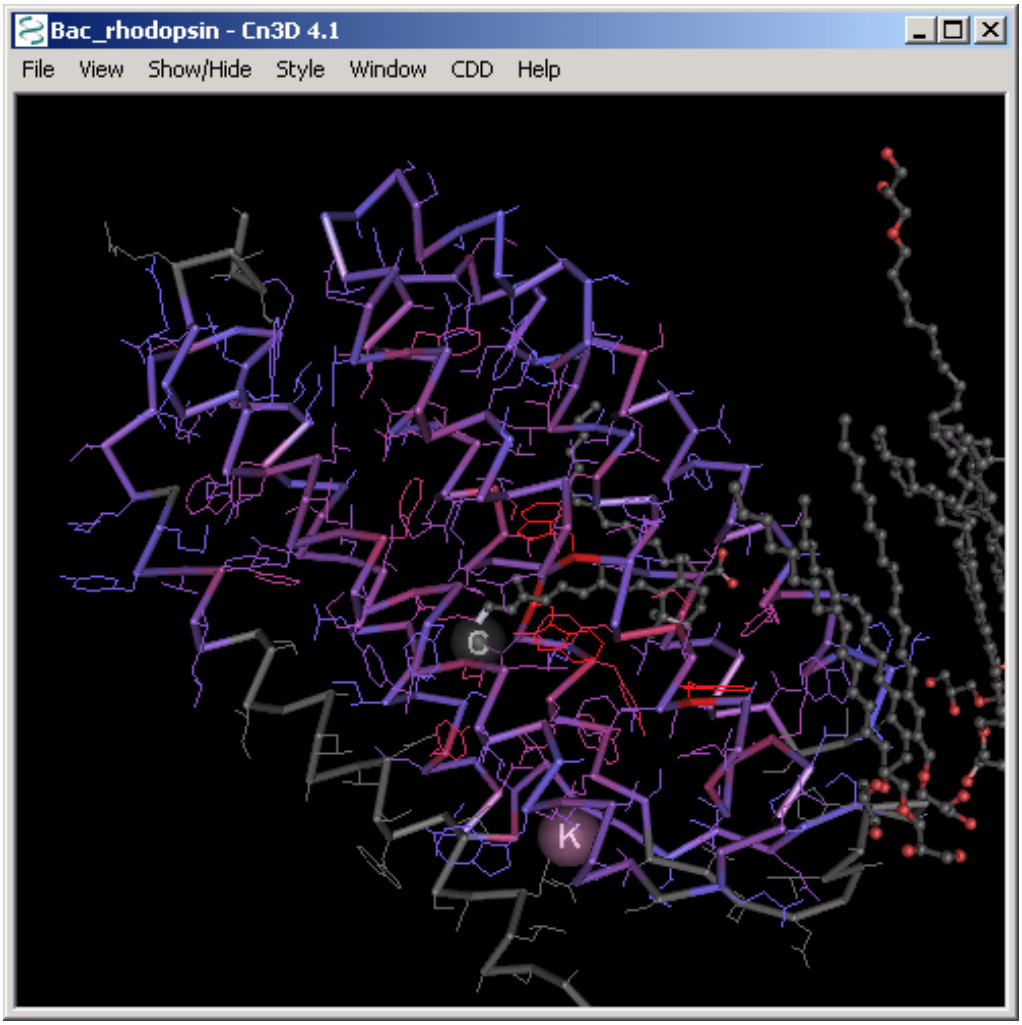
Title	PssmId	Multi-Dom	E-value
[+]pfam01036, Bac_rhodopsin, Bacteriorhodopsin..	64878	No	1e-50

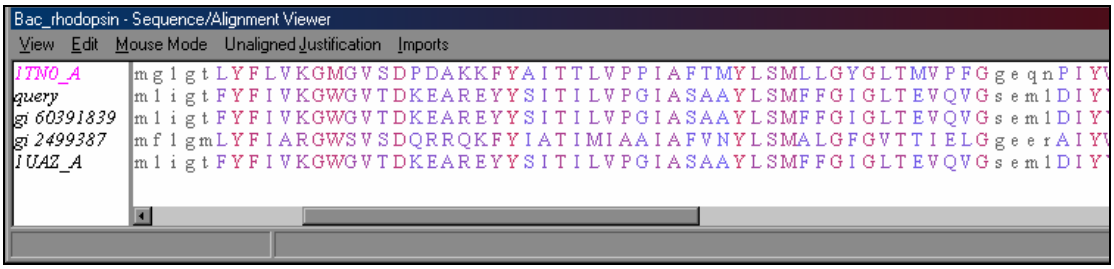
Search for similar domain architectures

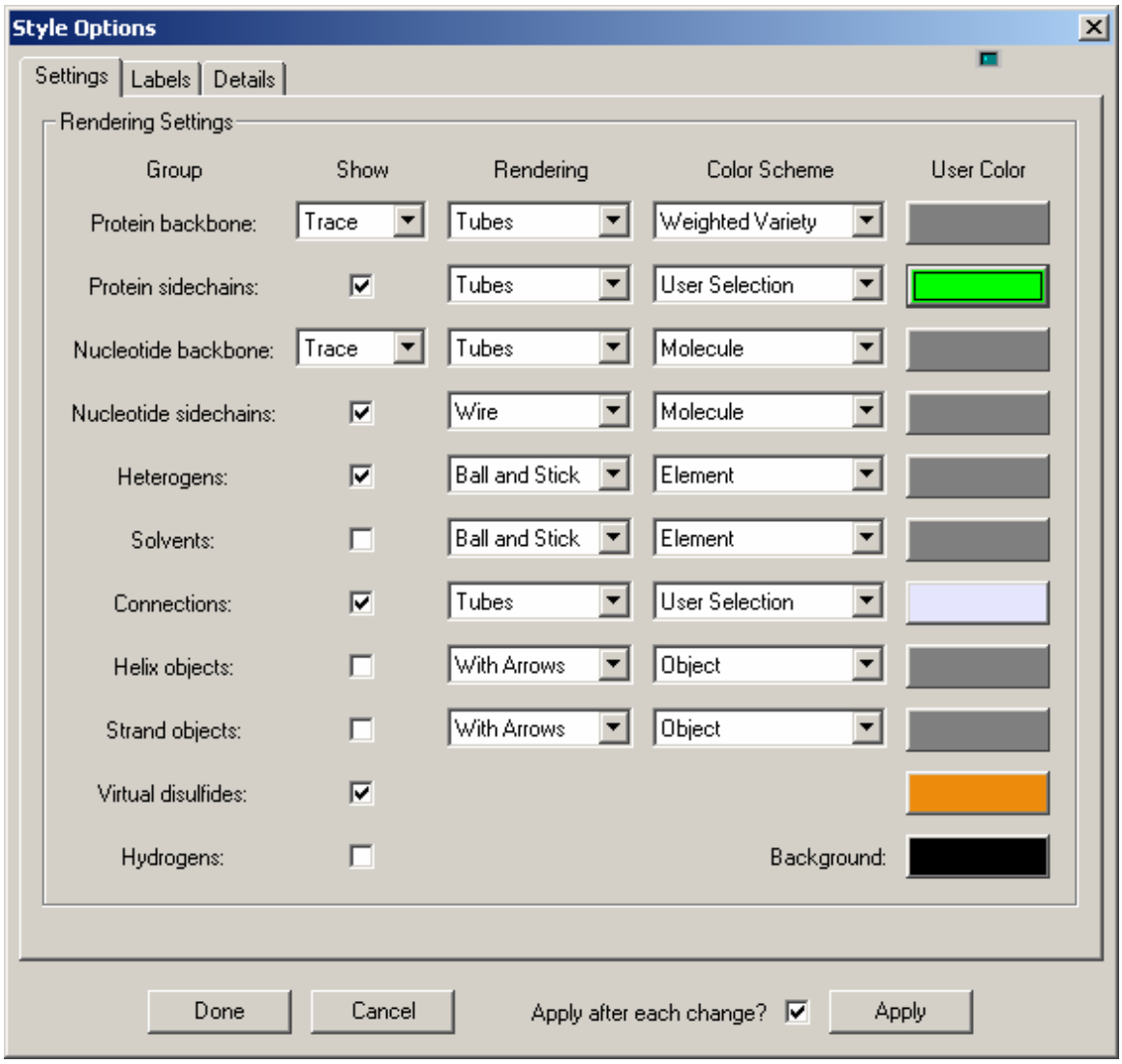

NCBI
Conserved Domains
SH3
SH2
HOME | SEARCH | SITE MAP
Entrez
CDD
Structure
Protein
pfam01036.13
Bac_rhodopsin, with user query added
Bacteriorhodopsin.
[+] Links:
[+] Statistics:
[+] Structure:
Other Related Conserved Domains:
Reformat Sequence Alignment
Format: Compact Hypertext
Row Display: up to 5
Color Bits: 2.0 bits
Type Selection: the most
1TN0_A 9 .[16].LYFLVK.[1].MGVSDPDAKKFYAITTLVPPIAFTMYLSMLLGYGLTMVPFG.[5].IYWARYADWLFTTPLL 93
query 21 .[16].FYFIVK.[1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[5].IYYARYADWLFTTPLL 105
gi 114809 34 .[16].LLFVFM.[1].RGLDDPRAKLIAVSTILVPVVSIASYTGLASGLTISVLEMP.[21].TMWGRYLTWALSTPMI 134
gi 60391839 21 .[16].FYFIVK.[1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[5].IYYARYADWLFTTPLL 105
gi 2499387 14 .[16].LYFIAR.[1].WSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG.[5].IYWARYTDWLFTTPLL 98
gi 2499386 7 .[16].LYFIAR.[1].WGETDSRRQKFYIATILITAIAFVNYLAMALGFGLTIVEFA.[5].IYWARYSDWLFTTPLL 91
gi 1168614 4 .[16].AVLAYG YTLVPEETRKRYLLLIAIPGIAIVAYALMALGFGSIQSEGH.[1].VYVVRYVDWLLTTPLN 83
gi 2499383 29 .[16].LLFVAM.[1].RDIESPRAKLIWVATMLVPLVSISSYAGLASGLTVGFLQMP.[11].SPWGRYLTWTFSTPMI 119
1E12_A 8 .[16].LVFVYM.[1].RTIRPGRPRLIWGATLMIPLVSISSYLGLLSGLTVGMIEMP.[11].SQWGRYLTWALSTPMI 98
1UAZ_A 15 .[16].FYFIVK.[1].WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[5].IYYARYADWLFTTPLL 99

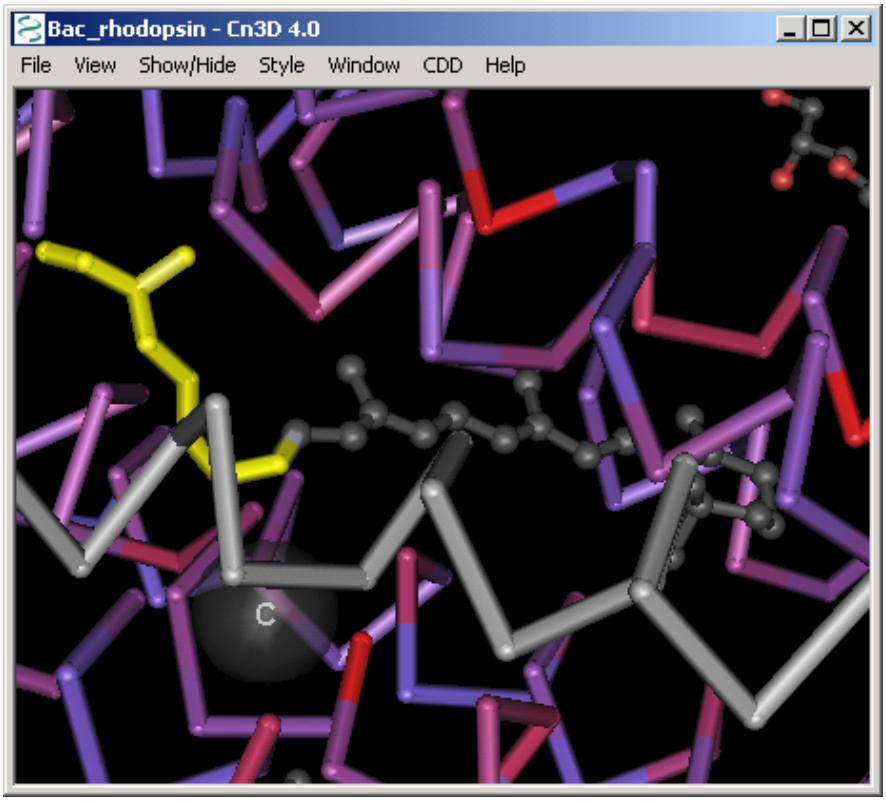

NCBI
Conserved Domains
SH3
SH2
HOME | SEARCH | SITE MAP
Entrez
CDD
Structure
pfam01036.13
Bac_rhodopsin, with user query added
Bacteriorhodopsin.
[+] Links:
[+] Statistics:
[−] Structure:
Show Structure
Program: Cn3D
Drawing: All Atoms
Aligned Rows: up to 5
[Download Cn3D]
Other Related Conserved Domains:
Reformat Sequence Alignment
Format: Compact Hypertext
Row Display: up to 5
Color Bits: 2.0 bits
1TN0_A 9 .[16].LYFLVKGMGVSDPDAKKFYAITTLVPPIAFTMYLSMLLGYGLTMVPFG.[4].PIYWARYADWLFTTPLLLLDL 97
query 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLDL 109
gi 60391839 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLDL 109
gi 2499387 14 .[16].LYFIARGWSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG.[4].AIYWARYTDWLFTTPLLLYDL 102
1UAZ_A 15 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLDL 103

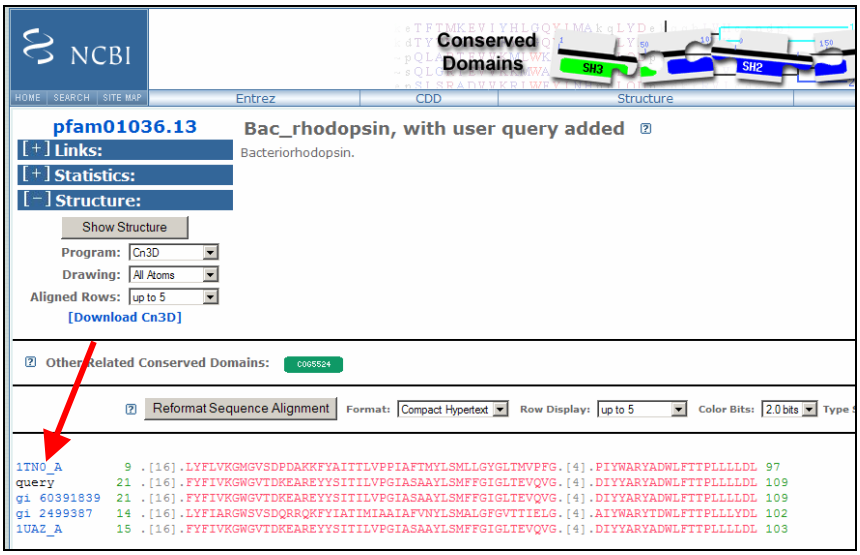

CDD Descriptive Items
Name: Bac_rhodopsin
Bacteriorhodopsin.
Structure summary:
PDB 1TN0 (MMDB 29943)
1TN0_A: gi 56553896 ([Halobacterium salinarum] Structure Of Bacterorhodopsin Mutant A51p)
Show Annotations Panel
Show References Panel
Dismiss

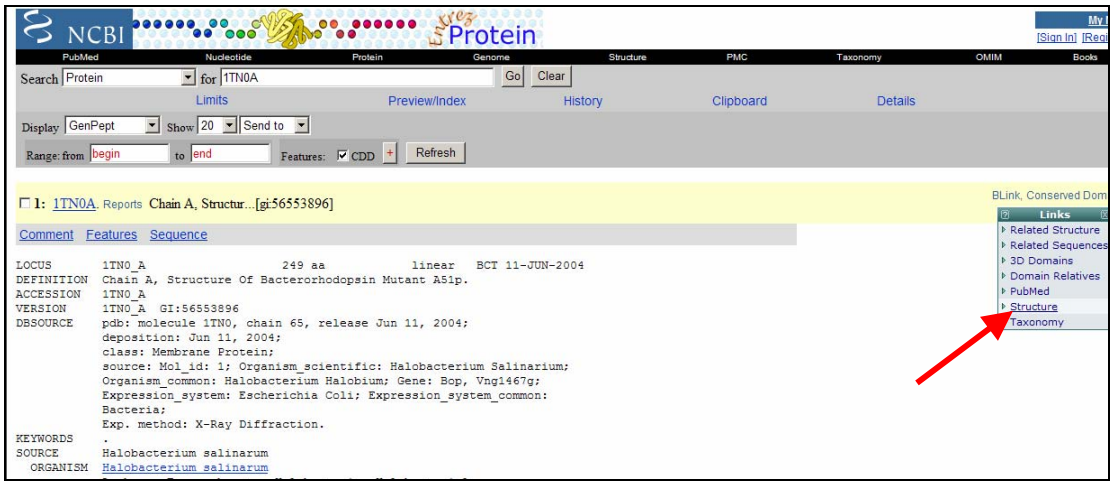

Bac_rhodopsin - Cn3D 4.1
File View Show/Hide Style Window CDD Help
C
K

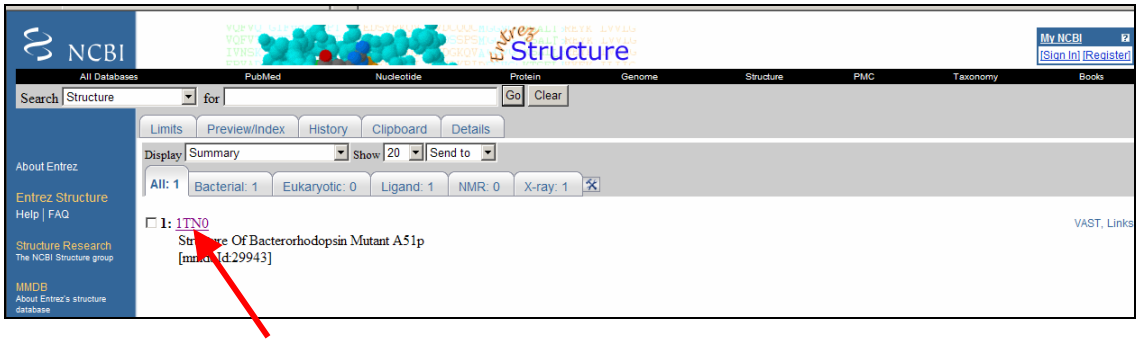

Bac_rhodopsin - Sequence/Alignment Viewer
View Edit Mouse Mode Unaligned Justification Imports
1TN0_A
query
gi 60391839
gi 2499387
1UAZ_A

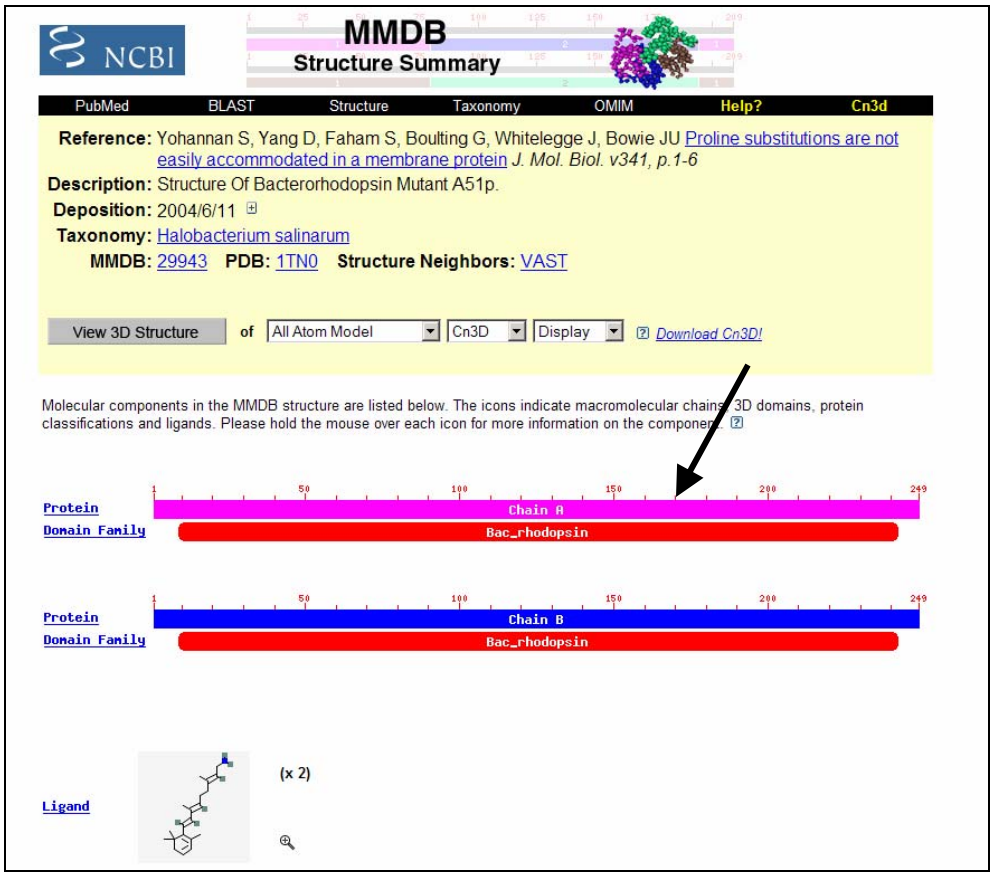

Style Options
Settings
Labels
Details
Rendering Settings
Group
Show
Rendering
Color Scheme
User Color
Protein backbone:
Trace
Tubes
Weighted Variety
Protein sidechains:
Tubes
User Selection
Nucleotide backbone:
Trace
Tubes
Molecule
Nucleotide sidechains:
Wire
Molecule
Heterogens:
Ball and Stick
Element
Solvents:
Ball and Stick
Element
Connections:
Tubes
User Selection
Helix objects:
With Arrows
Object
Strand objects:
With Arrows
Object
Virtual disulfides:
Hydrogens:
Background:
Done
Cancel
Apply after each change?
Apply

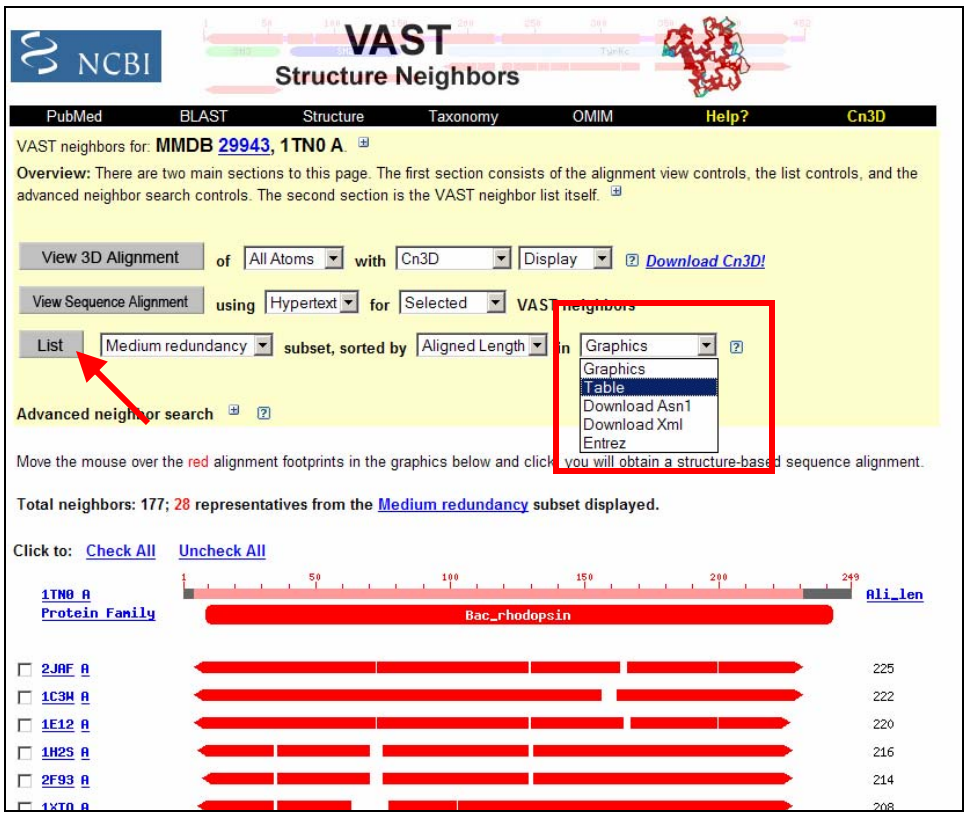

Bac_rhodopsin - Cn3D 4.0
File View Show/Hide Style Window CDD Help
C


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:
Reformat Sequence Alignment
Format: Compact Hypertext
Row Display: up to 5
Color Bits: 2.0 bits
1TN0_A 9 .[16].LYFLVKGMGVSDPDAKKFYAITTLVPPIAFTMYLSMLLGYGLTMVPFG.[4].PIYWARYADWLFTTPLLLLDL 97
query 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLDL 109
gi 60391839 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLDL 109
gi 2499387 14 .[16].LYFIARGWSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG.[4].AIYWARYTDWLFTTPLLLYDL 102
1UAZ_A 15 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLDL 103


NCBI
Entrez Protein
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Search Protein for 1TN0A Go Clear
Limits Preview/Index History Clipboard Details
Display GenPept Show 20 Send to
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 65, 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
BLink, Conserved Dom
Links
Related Structure
Related Sequences
3D Domains
Domain Relatives
PubMed
Structure
Taxonomy


NCBI
Entrez Structure
My NCBI
[Sign In] [Register]
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: 1TN0
Structure Of Bacterorhodopsin Mutant A51p
[mmdb Id:29943]
VAST, Links
About Entrez
Entrez Structure
Help | FAQ
Structure Research
The NCBI Structure group
MMDB
About Entrez's structure database


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
Domain Family
Chain A
Bac_rhodopsin
Protein
Domain Family
Chain B
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
Graphics
Table
Download Asn1
Download Xml
Entrez
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
Bac_rhodopsin
Ali_len
2JAF A 225
1C3W A 222
1E12 A 220
1H2S A 216
2F93 A 214

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

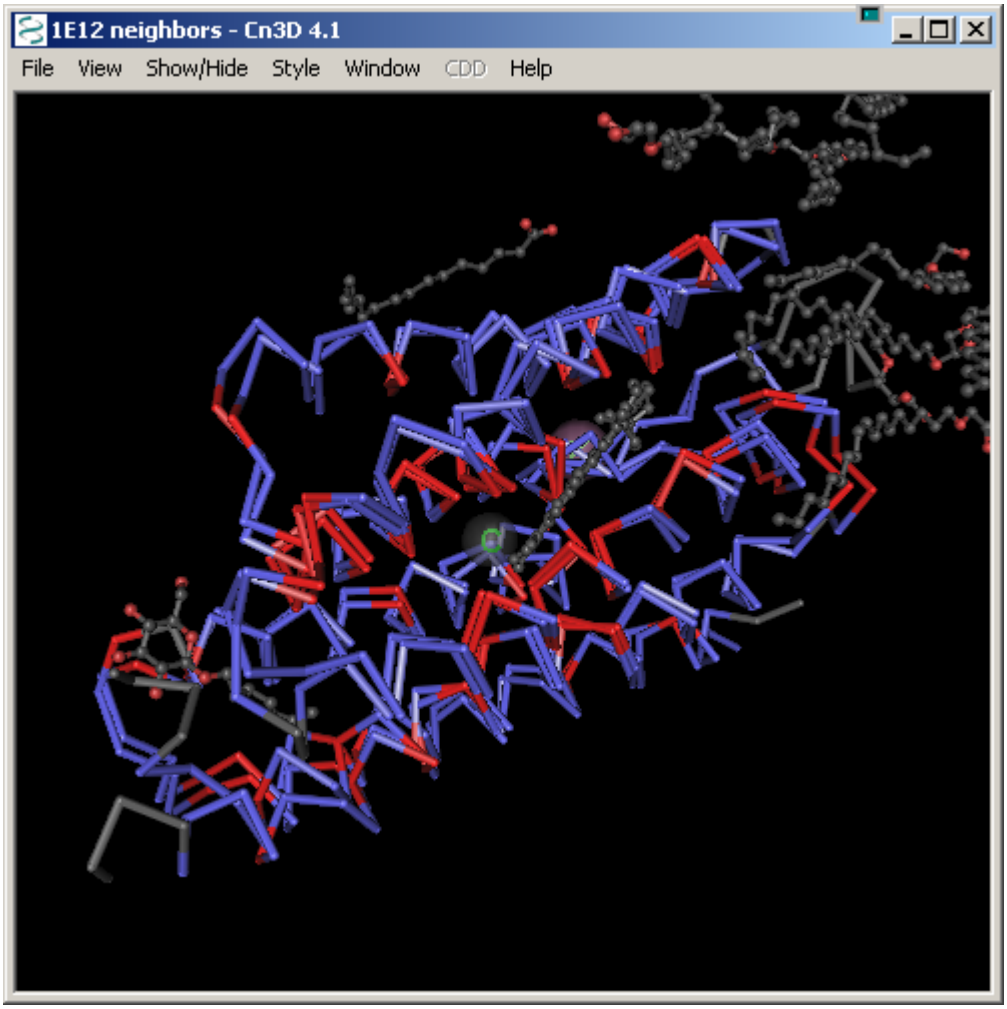
List Medium redundancy subset, sorted by Aligned Length in Table

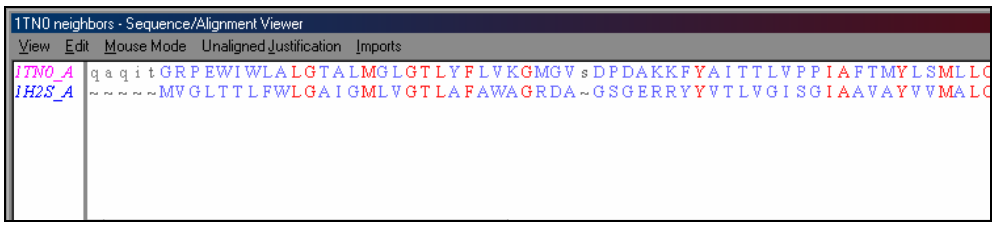
Advanced neighbor search

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

Click to: Check All Uncheck All

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


1E12 neighbors - Cn3D 4.1
File View Show/Hide Style Window CDD Help


1TN0 neighbors - Sequence/Alignment Viewer
View Edit Mouse Mode Unaligned Justification Imports
1TN0_A qaqitGRPEWIWLALGTALMGLGTLYFLVKGMGVsDPDAKKFYAITTLVPPIAFTMYLSMLLG
1H2S_A ~~~~~MVGLTTLFWLGAIGMLVGTLAFAWAGRDA~GSGERRYYVTLVGISGIAAVAYVVMALG

Problem 2

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.

```
>gi|2851597|sp|P25848|PHY1_CERPU Light-sensor Protein kina
MSATKKTYSSTTSAKSKHSVRVAQTTADAALEAVYEMSGDSG
QREGLIQNFGCMVAVEEPNFCVIAYSENASEFLDLIPQAVPSMGE
AATQDISLLNPITVHCRRSGKPLYAIAHRIDIGIVIDFEAVKMIDVPV
LPGGDIELLCDTIVEEVRELTGYDRVMAFKFHEDEHGEVVAEIRRM
KNRVRLIADCYASPVKLIQDPDIRQPVSLAGSTLRAPHGCHAQYI
IQRGRKLWGLVVCQHTSPRTVPFPLRSVCEFLMQVFGMQLNLH
PIGIVSQTPNIMDLVKCDGAALYYGKRVWLLGTTPTENQIKEIADW
HLLGDAVCGMAAAKITAKDFLFWFRSHTATEVKWGGAKHDPDE
EDVEMDAIHSLQLILRGSFRDIADSDTKTMIHARLNDLKLQGVEER
```

A. 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

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 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”.

C. Go back to the CD-Search results page. Click on the “Full Report” radio button. 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 "Aligned Rows" list box pull down menu to specify "up to 5" sequences

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.

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 kinase 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 conserved 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 Alignment" button.