Course contents: Review

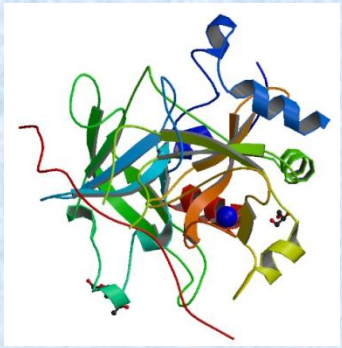
Bioinformatics: introduction	**DONE**
SSH, Unix commands, running programs in unix	**DONE**
Bioinformatics Databases	**DONE**
Sequence alignment: algorithms and online resources	**DONE**
Protein sequence analysis	**DONE**
Protein secondary and tertiary structure	**DONE**
Protein structure prediction	**DONE**
Phylogeny: methods and applications	**DONE**
Online resources in Bioinformatics/programming	**DONE**
Machine learning techniques/Genome analysis	**Ongoing**

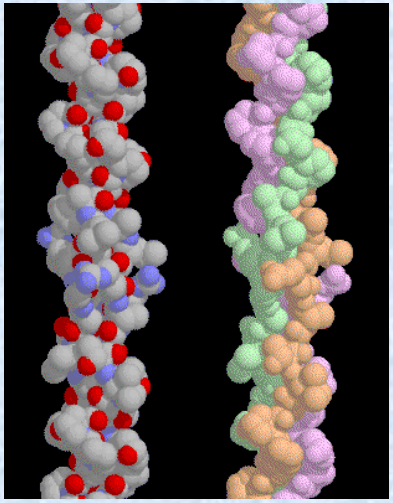
Proteins: secondary databases

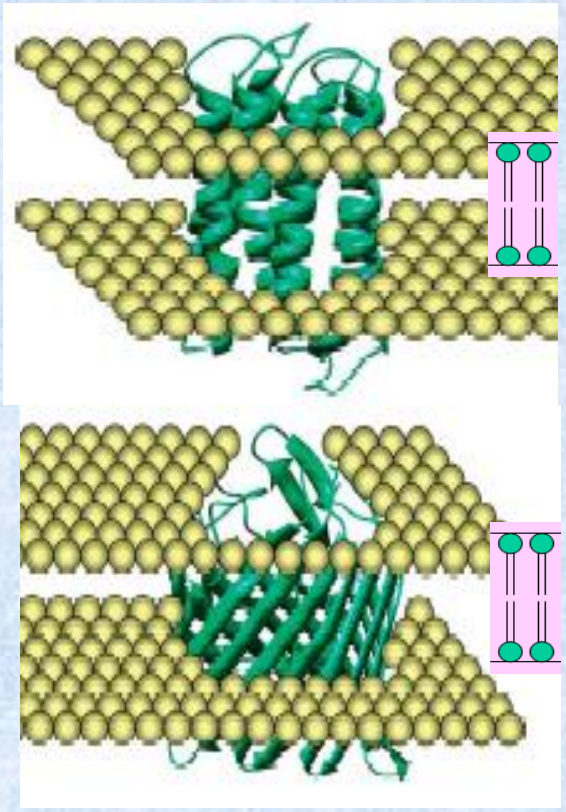
Globular proteins

Fibrous proteins

Membrane proteins







Cytoplasm

Inner membrane

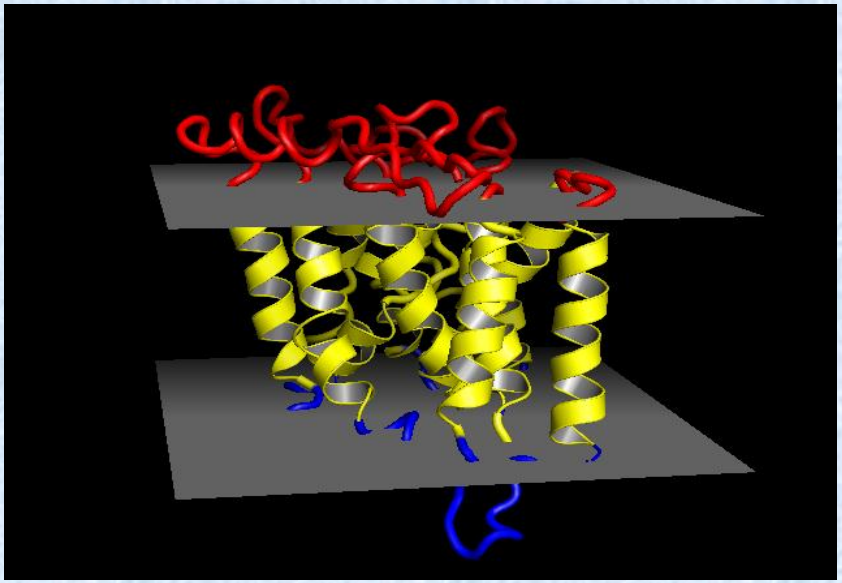
α-helical

Periplasm

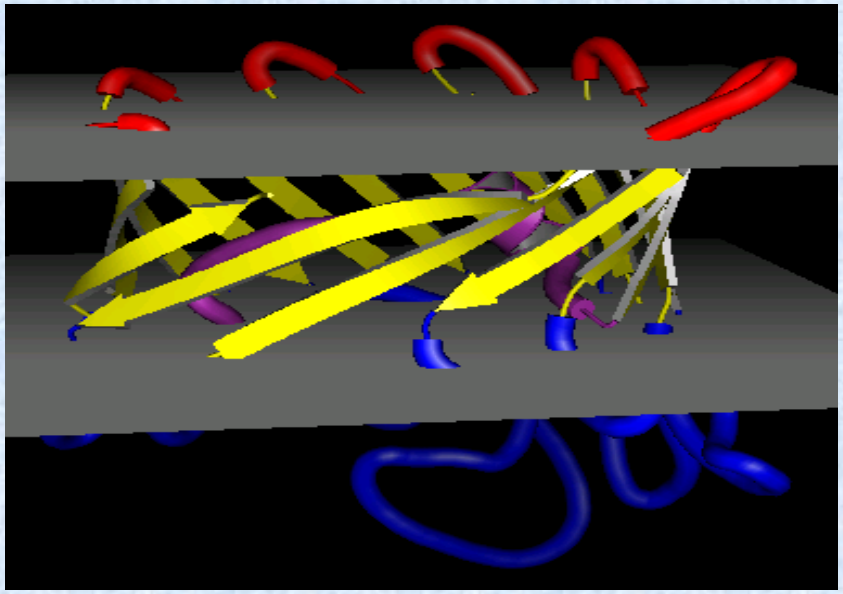
Outer membrane

β-barrel (TMβ)

Outer space



(a)



(b)

Figure 1.9

PDBTM

Database for membrane proteins

Algorithm to (TMDET) to identify the membrane spanning regions from 3D structure

- Protein is cut into 1Å slices and compute the membrane exposed area of hydrophobic (F, G, I, L, M, V, W and Y) and hydrophilic residues
- **Hydrophobic factor** is the hydrophobic area divided by total surface area.
- **Structure factor** is defined as the product of three factors, the straightness, turn and end-chain factor. These factors depend on the projection of Ca atoms of the residues i-3, i and i+3 onto a normal vector of membrane planes.
- The **objective function** (Q-value) is the average of the products of the hydrophobic factor and the structure factor in each slice.

PDBTM: Protein Data Bank of Transmembrane Proteins

PDBTM version: 2017-02-10 Number of transmembrane proteins: 3099 (alpha: 2723 , beta: 364)

all 1a0s

- Home
- Search
- Download
- Statistics
- Documents
- Help

Welcome to the PDBTM home page

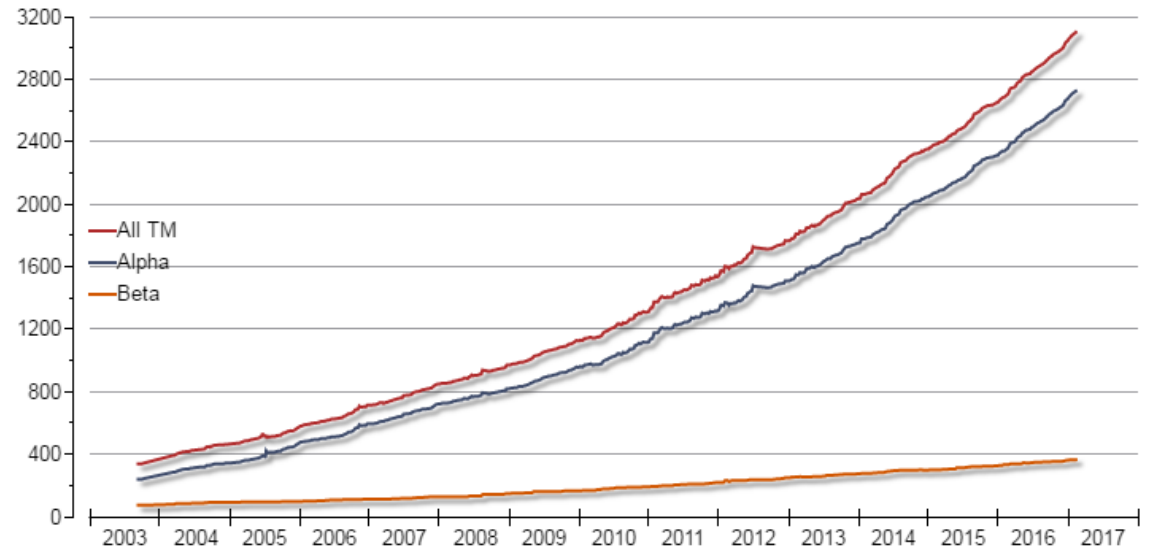
PDBTM is the first comprehensive and up-to-date transmembrane protein selection of the Protein Data Bank (PDB). PDBTM database is maintained at the Institute of Enzymology by the Membrane Protein Bioinformatics Research Group. The PDBTM database was created by scanning all PDB entries with the TMDET algorithm. You can get more information about PDBTM in our articles and in the PDBTM manual. If you find PDBTM useful in your research, please cite our articles (Bioinformatics 20, 2964-2972; Nucleic Acids Research 33 Database Issue, D275-D278; Nucleic Acids Research 41 Database Issue, D524-D529).

5jjj

5jjj

PDBTM type: Tm_Alpha
Chain(s): A[7], B[2], C[7], D[2]

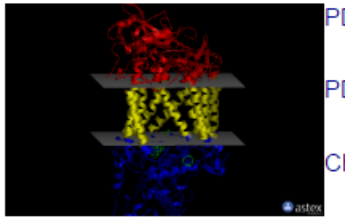
PDBTM growth statistics



search pdb_id: 1prc Submit

number of match: 1 Download XML files as zip Get it!

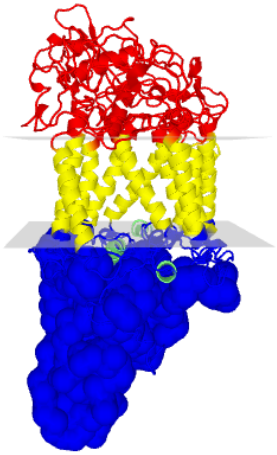
1. 1prc



PDB classification: PHOTOSYNTHETIC REACTION CENTER 04-FEB-88 1PRC 1PRC 3

PDBTM type: Tm_Alpha

Chain(s): C[0], L[5], M[5], H[1]



Sequence

Chain: L

C
H
L
M

ALLSFE GGTLIGGDLFDFWVGPYFVGF
GVSAIF SLIGYAASQGPTWDPFAISINP
PDLKYG AQAITVCALGAFISWML
RE FCVPIFMFCVLQVFRP
LLLGSWGHAFPYGIL SHLDWVNNFGYQYLNWHYN
PGHMSSVSFLFVNAMALGLHGG LILSVANPGDGD
KVKTAEHENQYFRDVVGYSIGALSIHRL GLFLAS
NIFLTGAFGTIA SGPFWTRGWPEWWGWWLDIPFW
S

ytosolic

Downloads

Cross references

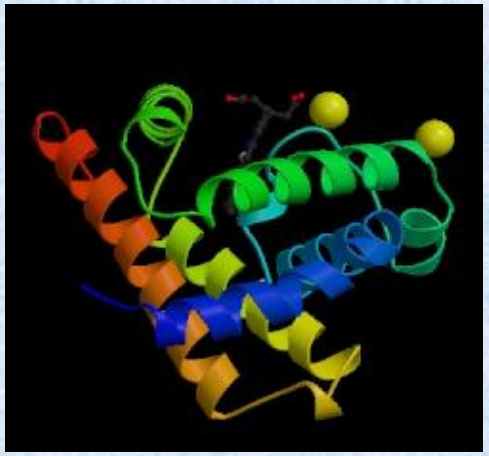
```
    <TMTYPE>Tm_Alpha</TMTYPE>
    <SPRES>Tm_Alpha</SPRES>
    <PDBKWRES>yes</PDBKWRES>
    <PDBKWORD>Membrane</PDBKWORD>
  </RAWRES>
▼<MEMBRANE>
    <NORMAL X="0.00000000" Y="0.00000000" Z="17.75000000"/>
  ▼<TMATRIX>
      <ROWX X="0.83854312" Y="-0.09631874" Z="-0.53625381" T="-52.36920929"/>
      <ROWY X="0.00000000" Y="0.98424947" Z="-0.17678508" T="-14.93281200"/>
      <ROWZ X="0.54483527" Y="0.14824191" Z="0.82533562" T="-33.65978378"/>
    </TMATRIX>
  </MEMBRANE>
▼<CHAIN CHAINID="A" NUM_TM="7" TYPE="alpha">
  ▼<SEQ>
      MESPIQIFRG EPGPTCAPSA CLPPNSSAWF PGWAEPDSNG SAGSEDAQLE PAHISPAIPV IITAVYSVVF
      MNSWPFGDVL CKIVISIDYY NMFTSIFTLT MMSVDRYIAV CHPVKALDFR TPLKAKIINI CIWLLSSSVG
      IIIVCYTLMI LRLKSVRLLS GSREKDRNLR RITRLVLVVV AVFVVCWTPI HIFILVEALG STSHSTAALS
      RNTVQDPAYL RDIDGMNKPV
    </SEQ>
    <REGION seq_beg="1" pdb_beg="1" seq_end="59" pdb_end="59" type="1"/>
    <REGION seq_beg="60" pdb_beg="60" seq_end="87" pdb_end="87" type="H"/>
    <REGION seq_beg="88" pdb_beg="88" seq_end="90" pdb_end="90" type="2"/>
    <REGION seq_beg="91" pdb_beg="91" seq_end="118" pdb_end="118" type="H"/>
    <REGION seq_beg="119" pdb_beg="119" seq_end="132" pdb_end="132" type="1"/>
    <REGION seq_beg="133" pdb_beg="133" seq_end="158" pdb_end="158" type="H"/>
    <REGION seq_beg="159" pdb_beg="159" seq_end="172" pdb_end="172" type="2"/>
    <REGION seq_beg="173" pdb_beg="173" seq_end="195" pdb_end="195" type="H"/>
    <REGION seq_beg="196" pdb_beg="196" seq_end="225" pdb_end="225" type="1"/>
    <REGION seq_beg="226" pdb_beg="226" seq_end="250" pdb_end="250" type="H"/>
    <REGION seq_beg="251" pdb_beg="251" seq_end="272" pdb_end="272" type="2"/>
    <REGION seq_beg="273" pdb_beg="273" seq_end="294" pdb_end="294" type="H"/>
    <REGION seq_beg="295" pdb_beg="295" seq_end="307" pdb_end="307" type="1"/>
    <REGION seq_beg="308" pdb_beg="308" seq_end="334" pdb_end="334" type="H"/>
    <REGION seq_beg="335" pdb_beg="335" seq_end="380" pdb_end="380" type="2"/>
  </CHAIN>
</pdbtm>
```

Different structural classes

4MBN

all-α

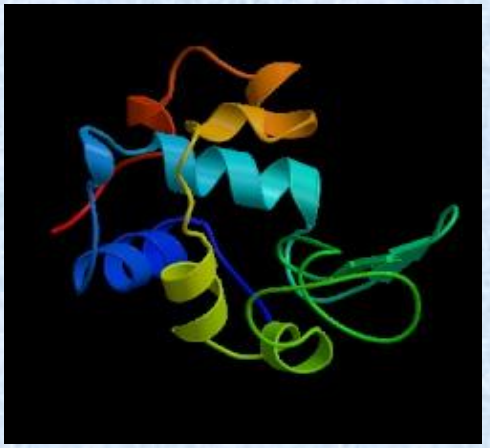
Dominated by α-helices
α>40; β<5



4LYZ

α+β

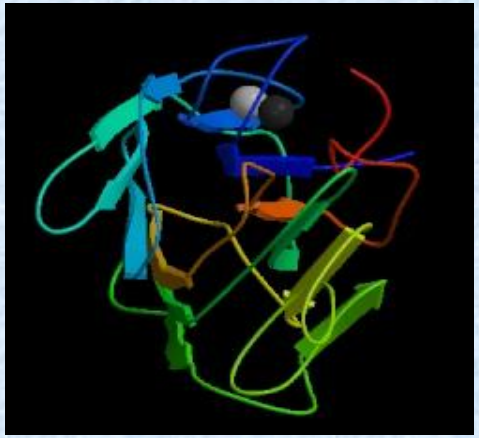
Helices and strands tend to Segregate
α>15; β>10



3CNA

all-β

Dominated by β-strands
β>40; α<20



1TIM

α/β

Helices and strands mix each other
α>15; β>10

SCOP
Structural Classification of Proteins

Provides a detailed and comprehensive description of the structural and evolutionary relationships of the proteins of known structure.

For each protein, the classification has the hierarchical levels, family, superfamily, fold and structural class.

Family: C-type lysozyme

Fold: Lysozyme-like *common alpha+beta motif for the active site region*

Structural class: α+β

The structure can be identified with a six letter code (1lz1__; 5th Chain; 6th Domain

Protein: Lysozyme from Human (*Homo sapiens*)

Lineage:

1. Root: scop
2. Class: Alpha and beta proteins (a+b) [53931]
   Mainly antiparallel beta sheets (segregated alpha and beta regions)
3. Fold: Lysozyme-like [53954]
   common alpha+beta motif for the active site region
4. Superfamily: Lysozyme-like [53955]
5. Family: C-type lysozyme [53960]
6. Protein: Lysozyme [53961]
   ubiquitous in a variety of tissues and secretions
7. Species: Human (*Homo sapiens*) [53969]

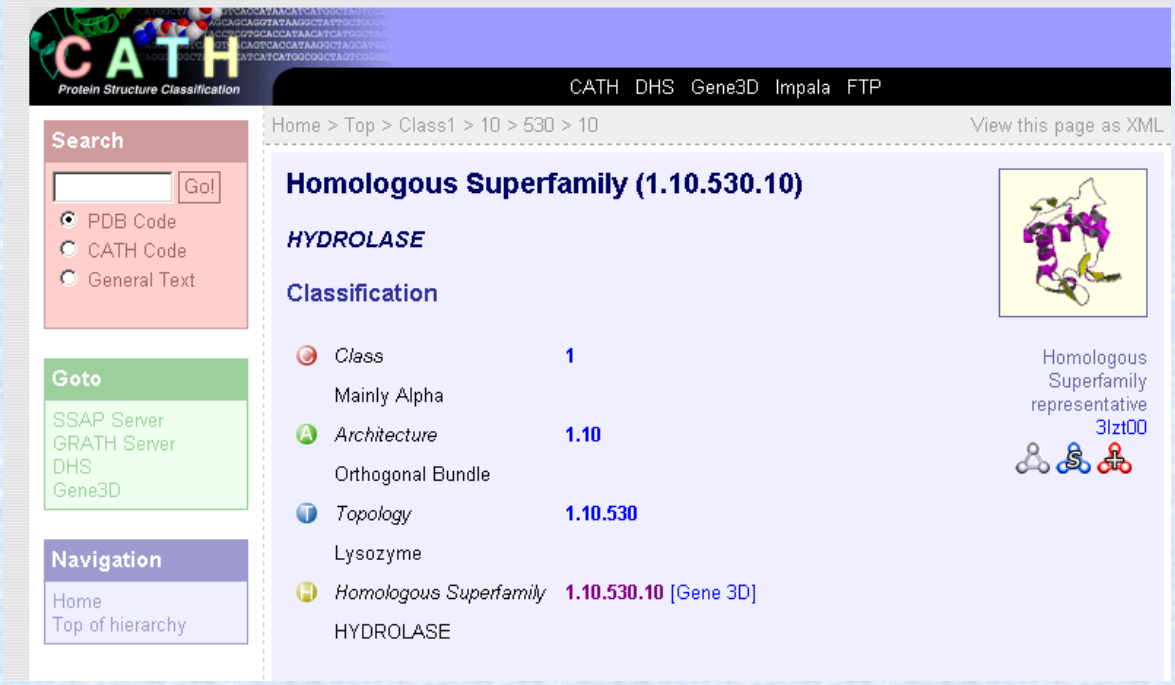
PDB Entry Domains:

1. 1jsf [36410]
   complexed with no3
2. 1iwu
   complexed with cl
   1. chain a [76891]
3. 1iwt
   complexed with cl
   1. chain a [76890]
4. 1iwv
   complexed with cl
   1. chain a [76892]
5. 1iww
   complexed with cl
   1. chain a [76893]
6. 1jwr
   1. chain a [63311]

CATH: Hierarchical classification of protein domain structures

Provides class information for all the structures in PDB.

The four main levels of CATH classification are protein class (C), architecture (A), topology (T) and homologous superfamily (H).



CATH: Hierarchical classification of protein domain structures

Class is the simplest level, and it essentially describes the secondary structure composition of each domain.

Architecture summarizes the shape revealed by the orientations of the secondary structure units, such as barrels and sandwiches.

At the topology level, sequential connectivity is considered, such that members of the same architecture might have quite different topologies.

The homologous superfamilies cluster proteins with highly similar structures and functions.

The CATH classification for human lysozyme: Architecture of orthogonal bundle and it is assigned as all-α class proteins in CATH.

SCOP and CATH: Comparison

Usually CATH and SCOP have similar classification for structural class.

It may be possible to have different assignments in SCOP and CATH.

Proteins: percentage of helical (or strand) content is high and strand (helix) content is less or vice versa.

In human lysozyme, the content of α-helical structures is 31% and that of β-strands is 8%.

Based on the high content of α-helix it was classified as all-α proteins in CATH

Due to the presence of α-helices and β-strands SCOP classified it as α+β protein.

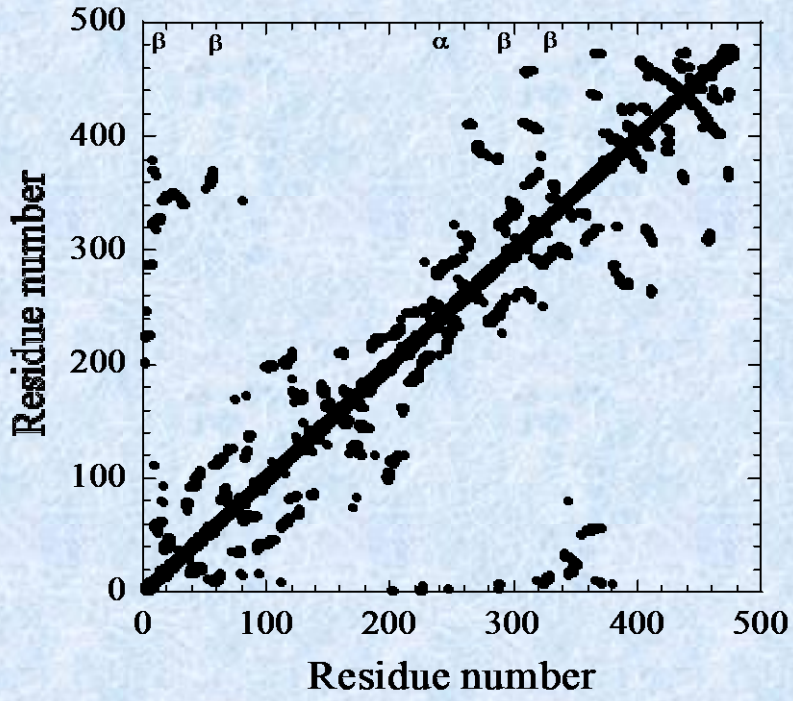
Protein structure analysis

Contact maps

A protein **contact map** represents the **distance** between all possible residue pairs of a three-dimensional protein structure using a **binary two-dimensional matrix**.

For two residues *i* and *j*, the ***ij* element of the matrix is 1** if the **two residues are closer than a predetermined threshold**, and 0 otherwise.

Various contact definitions have been proposed: The distance between the **C_α-C_α atom with threshold 6-12 Å**; distance between C_β-C_β atoms with threshold 6-12 Å (C_α is used for Glycine); distance between the side-chain centers of mass; or consideration of any atom and lower threshold 4.5-6 Å.



Construction of contact maps

Cα atoms (limit of 8Å)

ATOM	1	N	MET	A	1	36.644	-24.949	8.853	1.00 29.12	N
ATOM	2	CA	MET	A	1	36.942	-23.581	8.984	1.00 19.55	C
ATOM	3	C	MET	A	1	35.712	-22.887	9.526	1.00 22.27	C
ATOM	4	O	MET	A	1	34.626	-23.375	9.258	1.00 18.31	O
ATOM	5	CB	MET	A	1	37.365	-23.090	7.599	1.00 8.40	C
ATOM	6	CG	MET	A	1	37.639	-21.603	7.644	1.00 30.36	C
ATOM	7	SD	MET	A	1	39.309	-21.106	7.226	1.00 39.80	S
ATOM	8	CE	MET	A	1	40.241	-22.126	8.356	1.00 44.83	C
ATOM	9	N	ASN	A	2	35.890	-21.796	10.310	1.00 17.74	N
ATOM	10	CA	ASN	A	2	34.809	-21.015	10.918	1.00 5.91	C
ATOM	11	C	ASN	A	2	35.236	-19.557	10.931	1.00 11.34	C
ATOM	12	O	ASN	A	2	36.390	-19.244	10.620	1.00 9.58	O
ATOM	13	CB	ASN	A	2	34.487	-21.602	12.355	1.00 9.68	C
ATOM	14	CG	ASN	A	2	35.645	-21.566	13.309	1.00 11.82	C
ATOM	15	OD1	ASN	A	2	36.176	-20.496	13.515	1.00 18.42	O
ATOM	16	ND2	ASN	A	2	36.013	-22.685	13.919	1.00 9.43	N
ATOM	17	N	ILE	A	3	34.315	-18.689	11.287	1.00 6.65	N
ATOM	18	CA	ILE	A	3	34.543	-17.262	11.353	1.00 10.61	C
ATOM	19	C	ILE	A	3	35.794	-16.849	12.198	1.00 14.73	C
ATOM	20	O	ILE	A	3	36.530	-15.891	11.865	1.00 13.47	O
ATOM	21	CB	ILE	A	3	33.235	-16.515	11.748	1.00 6.98	C
ATOM	22	CG1	ILE	A	3	33.352	-14.965	11.537	1.00 7.31	C
ATOM	23	CG2	ILE	A	3	32.955	-16.849	13.220	1.00 12.46	C
ATOM	24	CD1	ILE	A	3	33.729	-14.545	10.094	1.00 7.03	C
ATOM	25	N	PHE	A	4	36.027	-17.510	13.335	1.00 6.09	N
ATOM	26	CA	PHE	A	4	37.178	-17.109	14.110	1.00 6.45	C
ATOM	27	C	PHE	A	4	38.472	-17.460	13.428	1.00 7.23	C
ATOM	28	O	PHE	A	4	39.418	-16.667	13.450	1.00 11.23	O
ATOM	29	CB	PHE	A	4	37.105	-17.747	15.496	1.00 13.74	C
ATOM	30	CG	PHE	A	4	35.931	-17.216	16.290	1.00 22.31	C
ATOM	31	CD1	PHE	A	4	34.698	-17.879	16.239	1.00 14.28	C
ATOM	32	CD2	PHE	A	4	36.093	-16.097	17.122	1.00 12.02	C
ATOM	33	CE1	PHE	A	4	33.635	-17.415	17.016	1.00 13.79	C
ATOM	34	CE2	PHE	A	4	35.038	-15.603	17.909	1.00 15.06	C
ATOM	35	CZ	PHE	A	4	33.840	-16.322	17.873	1.00 15.21	C
ATOM	36	N	GLU	A	5	38.539	-18.674	12.852	1.00 9.52	N
ATOM	37	CA	GLU	A	5	39.746	-19.063	12.152	1.00 16.61	C
ATOM	38	C	GLU	A	5	39.972	-18.177	10.940	1.00 14.11	C
ATOM	39	O	GLU	A	5	41.077	-17.723	10.662	1.00 10.00	O

ATOM 2 CA MET A 1 36.942 -23.581 8.984 1.00 19.55 C

ATOM 10 CA ASN A 2 34.809 -21.015 10.918 1.00 5.91 C

ATOM 18 CA ILE A 3 34.543 -17.262 11.353 1.00 10.61 C

ATOM 26 CA PHE A 4 37.178 -17.109 14.110 1.00 6.45 C

ATOM 37 CA GLU A 5 39.746 -19.063 12.152 1.00 16.61 C

ATOM 46 CA MET A 6 39.005 -17.022 9.081 1.00 8.98 C

ATOM 54 CA LEU A 7 39.394 -13.637 10.850 1.00 9.64 C

ATOM 62 CA ARG A 8 42.437 -14.708 12.525 1.00 19.95 C

ATOM 73 CA ILE A 9 44.096 -15.463 9.190 1.00 11.21 C

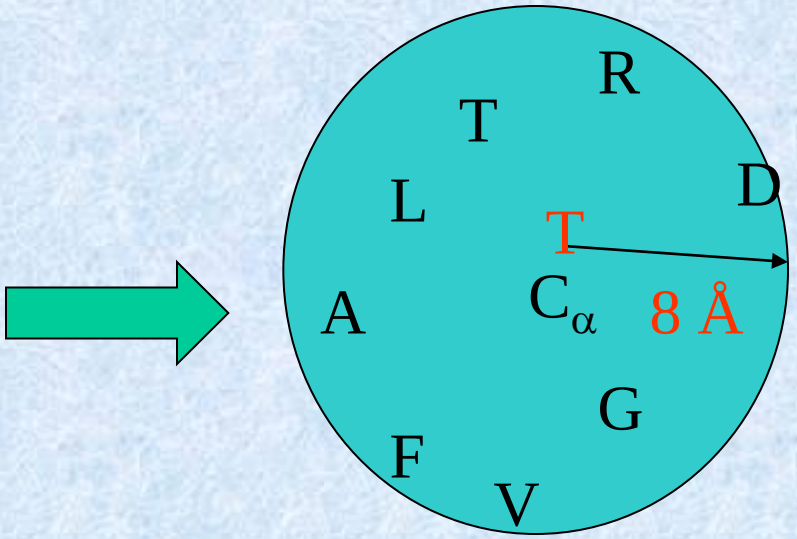
ATOM 81 CA ASP A 10 43.052 -12.038 7.999 1.00 14.11 C

Construct contact map?

Inter-residue contacts: Definition

Basic concepts:

- The amino acid residues, which are in contact with each other in the *native structure* (8Å)
- Their respective locations in the *amino acid sequence*



DFA…LTRTDAH…STA

Classification

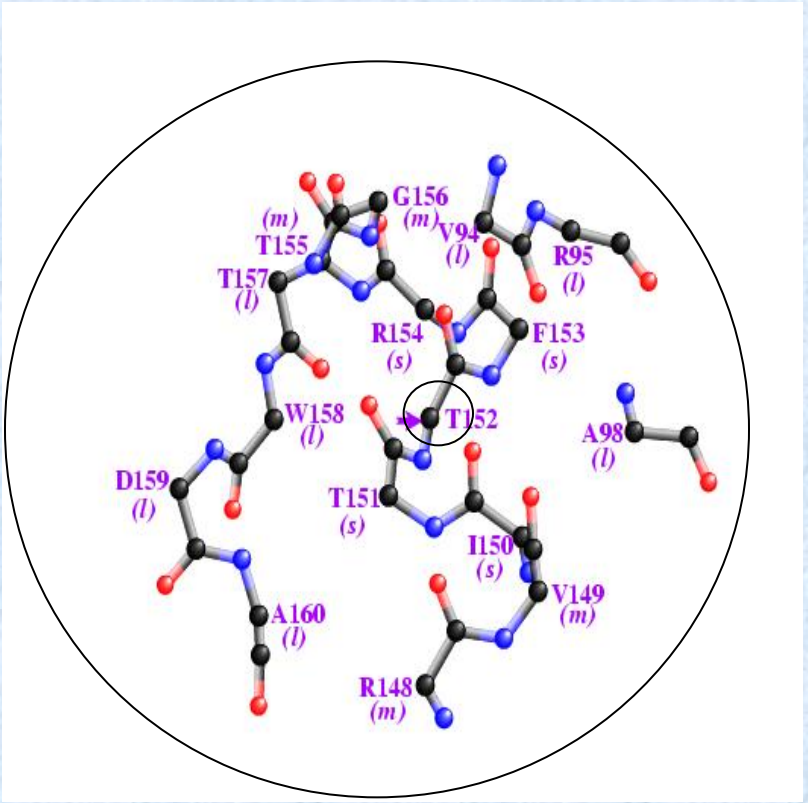
- Distance of separation between amino acid residues in a polypeptide chain

s: short (< ± 3 residues)

m: medium (± 3 or ± 4 residues)

l: long (> ± 4 residues)

(further divided into small bins of 10 residues)



Gromiha, MM and Selvaraj, S. (2004) Prog. Biophys. Mol. Biol. 86, 235-277.

Representation

Inter-residue contacts can be pictorially represented by contact maps.

Short-range contacts:

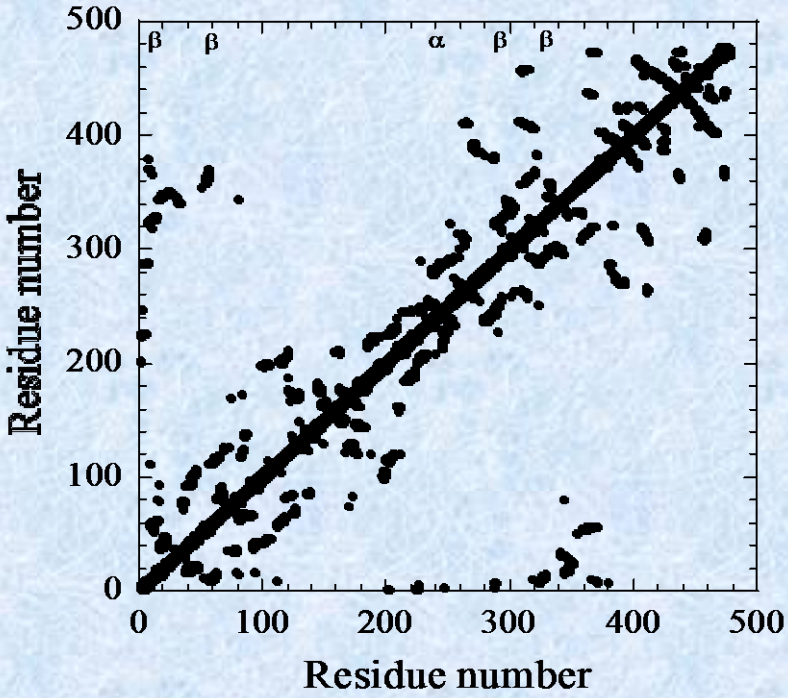
diagonal residues

Medium-range contacts:

close to diagonal residues

Long-range contacts:

far away from the diagonal.



Secondary structure

DSSP: Dictionary of Secondary Structure of Proteins

Algorithm:

Hydrogen bonding pattern **ftp://ftp.cmbi.kun.nl/**

Download the program dssp-2

Command:

./dssp-2 input output

```
==== Secondary Structure Definition by the program DSSP, updated CMBI version by ElmK / April 1,2000 ==== DATE=2-APR-2004      .
REFERENCE W. KABSCH AND C.SANDER, BIOPOLYMERS 22 (1983) 2577-2637                                                            .
HEADER    HYDROLASE (O-GLYCOSYL)                  12-OCT-84   1LZ1                                                       .
COMPND    LYSOZYME (E.C.3.2.1.17)                                                                                        .
SOURCE    HUMAN (HOMO $SAPIENS)                                                                                          .
AUTHOR    P.J.ARTYMIUK,C.C.F.BLAKE                                                                                       .
  130   1   4   4   0 TOTAL NUMBER OF RESIDUES, NUMBER OF CHAINS, NUMBER OF SS-BRIDGES(TOTAL,INTRACHAIN,INTERCHAIN)      .
  6787.0    ACCESSIBLE SURFACE OF PROTEIN (ANGSTROM**2)                                                                  .
   89 68.5    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(J)  , SAME NUMBER PER 100 RESIDUES                       .
    2  1.5    TOTAL NUMBER OF HYDROGEN BONDS IN     PARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES                       .
   12  9.2    TOTAL NUMBER OF HYDROGEN BONDS IN ANTIPARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES                       .
    2  1.5    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-5), SAME NUMBER PER 100 RESIDUES                       .
    1  0.8    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-4), SAME NUMBER PER 100 RESIDUES                       .
    1  0.8    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-3), SAME NUMBER PER 100 RESIDUES                       .
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-2), SAME NUMBER PER 100 RESIDUES                       .
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-1), SAME NUMBER PER 100 RESIDUES                       .
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+0), SAME NUMBER PER 100 RESIDUES                       .
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+1), SAME NUMBER PER 100 RESIDUES                       .
    9  6.9    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+2), SAME NUMBER PER 100 RESIDUES                       .
   25 19.2    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+3), SAME NUMBER PER 100 RESIDUES                       .
   34 26.2    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+4), SAME NUMBER PER 100 RESIDUES                       .
    5  3.8    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+5), SAME NUMBER PER 100 RESIDUES                       .
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30     *** HISTOGRAMS OF ***     .
 0  0  0  0  0  1  0  0  0  1  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    RESIDUES PER ALPHA HELIX   .
 1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    PARALLEL BRIDGES PER LADDER .
 2  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    ANTIPARALLEL BRIDGES PER LADDER .
 0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    LADDERS PER SHEET          .
  #  RESIDUE AA STRUCTURE BP1 BP2  ACC     N-H-->O    O-->H-N    N-H-->O    O-->H-N     TCO  KAPPA ALPHA  PHI   PSI    X-CA   Y-CA   Z-CA
    1    1   K              0   0   63     0, 0.0    39,-2.6     0, 0.0     2,-0.7    0.000 360.0 360.0 360.0 146.7    1.9   20.6   21.1
    2    2   V  B     -A   39   0A  99    37,-0.2    37,-0.2    38,-0.1     2,-0.1   -0.949 360.0-145.6-104.2 112.9    4.6   19.8   18.5
    3    3   F              0   0   10    35,-2.4     2,-0.2    -2,-0.7     3, 0.0   -0.453  10.1-122.2 -70.2 150.5    7.4   22.4   18.9
    4    4   E    >  -      0   0  146    -2,-0.1     4,-2.6     1,-0.1     5,-0.3   -0.657  34.8-107.1 -82.3 155.8    9.4   23.7   16.0
    5    5   R  H  > S+     0   0   91    -2,-0.2     4,-2.1     1,-0.2     5,-0.1    0.886 115.0  37.2 -56.3 -50.8   13.1   23.1   16.5
    6    6   a  H  > S+     0   0   27     1,-0.2     4,-2.5     2,-0.2    -1,-0.2    0.826 112.9  58.3 -78.4 -24.6   14.3   26.7   17.2
    7    7   E  H  > S+     0   0   58     1,-0.2     4,-2.3     2,-0.2    -1,-0.2    0.907 109.6  45.4 -60.1 -45.4   11.1   27.5   19.2
    8    8   L  H  X S+     0   0    1    -4,-2.6     4,-3.2     2,-0.2     5,-0.3    0.903 110.4  54.1 -62.5 -42.1   12.0   24.6   21.5
    9    9   A  H  X S+     0   0    0    -4,-2.1     4,-2.1    -5,-0.3    -2,-0.2    0.937 111.4  43.9 -61.1 -44.3   15.7   25.6   21.7
   10   10   R  H  X S+     0   0  114    -4,-2.5     4,-2.5     2,-0.2    -1,-0.2    0.906 113.6  51.9 -67.4 -34.7   14.8   29.2   22.8
   11   11   T  H  X S+     0   0   29    -4,-2.3     4,-1.3     1,-0.2    -2,-0.2    0.910 111.0  46.7 -66.8 -47.6   12.2   27.8   25.3
   12   12   L  H  <>S+     0   0    0    -4,-3.2     5,-2.5     2,-0.2     6,-0.3    0.845 111.4  52.9 -61.5 -37.6   14.7   25.4   26.8
   13   13   K  H  ><5S+     0   0   89    -4,-2.1     3,-1.8    -5,-0.3     5,-0.3    0.970 109.3  47.8 -61.0 -44.3   17.3   28.3   27.0
   14   14   R  H  3<5S+     0   0  1[illegible]    -4,-2.5    -1,-0.2     1,-0.3    -2,-0.2    0.809 106.9  58.1 -67.8 -23.4   14.9   30.5   28.9
   15   15   L  T  3<5S-     0   0   49    -4,-1.3    -1,-0.3    -5,-0.2    -2,-0.2    0.399 121.3-107.0 -91.0    6.3   14.0   27.6   31.3
   16   16   G  T  < 5S+     0   0   40    -3,-1.8    -3,-0.2    -4,-0.3    -2,-0.1    0.719  79.9 128.7  86.8   24.2   17.7   27.4   32.3
   17   17   M     > < +      0   0    0    -5,-2.5     3,-1.7     2,-0.1     2,-0.6    0.625  35.2 105.3 -87.8 -19.2   18.7   24.2   30.5
   18   18   D  T 3  S-      0   0   78    -6,-0.3     6,-0.2    -5,-0.3    4,-0.1    -0.578 104.8  -7.7 -68.4 107.8   21.8   25.4   28.6
   19   19   G  T 3  S+      0   0   44     4,-2.4     3,-2.1    -2,-0.6     2,-0.4    0.604  88.0 164.0  88.8   6.3   24.8   23.9   30.6
   20   20   Y  B  X  S-B   23   0B  54    -3,-1.7     3,-1.6     3,-0.8    -1,-0.3   -0.584  80.4 -10.2 -63.8 126.7   22.6   22.6   33.4
   21   21   R  T 3  S-      0   0  165    -2,-0.4    -1,-0.3     1,-0.2     3,-0.1    0.784 134.6 -56.6  46.1  39.0   24.8   20.0   35.2
   22   22   G  T <  S+      0   0   75    -3,-2.1     2,-0.7     1,-0.2    -1,-0.2    0.507 104.9 132.1  83.1   6.6   27.2   20.4   32.3
   23   23   I  B  X   -B   20   0B  15    -3,-1.6    -4,-2.4    -6,-0.1    -3,-0.8   -0.753  50.3-136.2-103.2 121.5   24.8   19.5   29.6
```

Secondary structure

Solvent accessibility

Hydrogen bonding partners/ Electrostatic energy

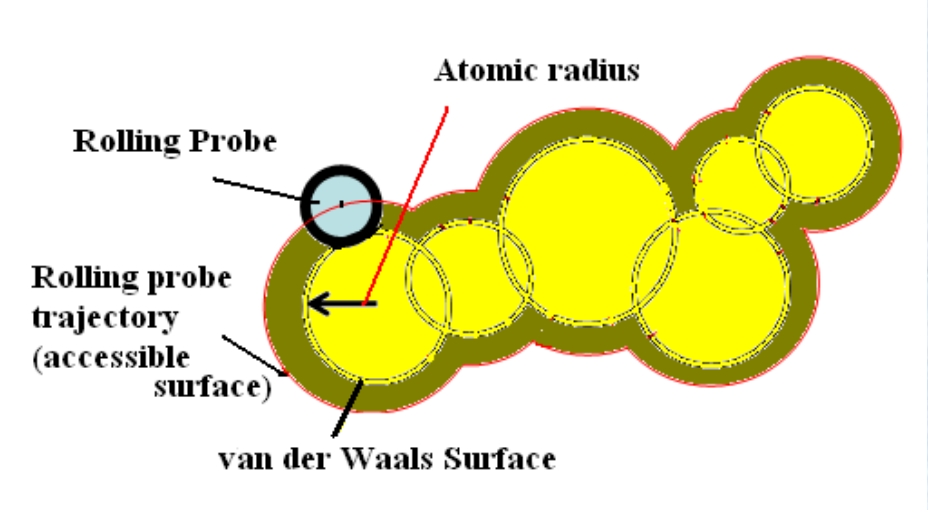
Solvent accessibility

The solvent accessible surface area is defined as the locus of the centre of the solvent molecule as it rolls over the van der Waals surface of the protein.

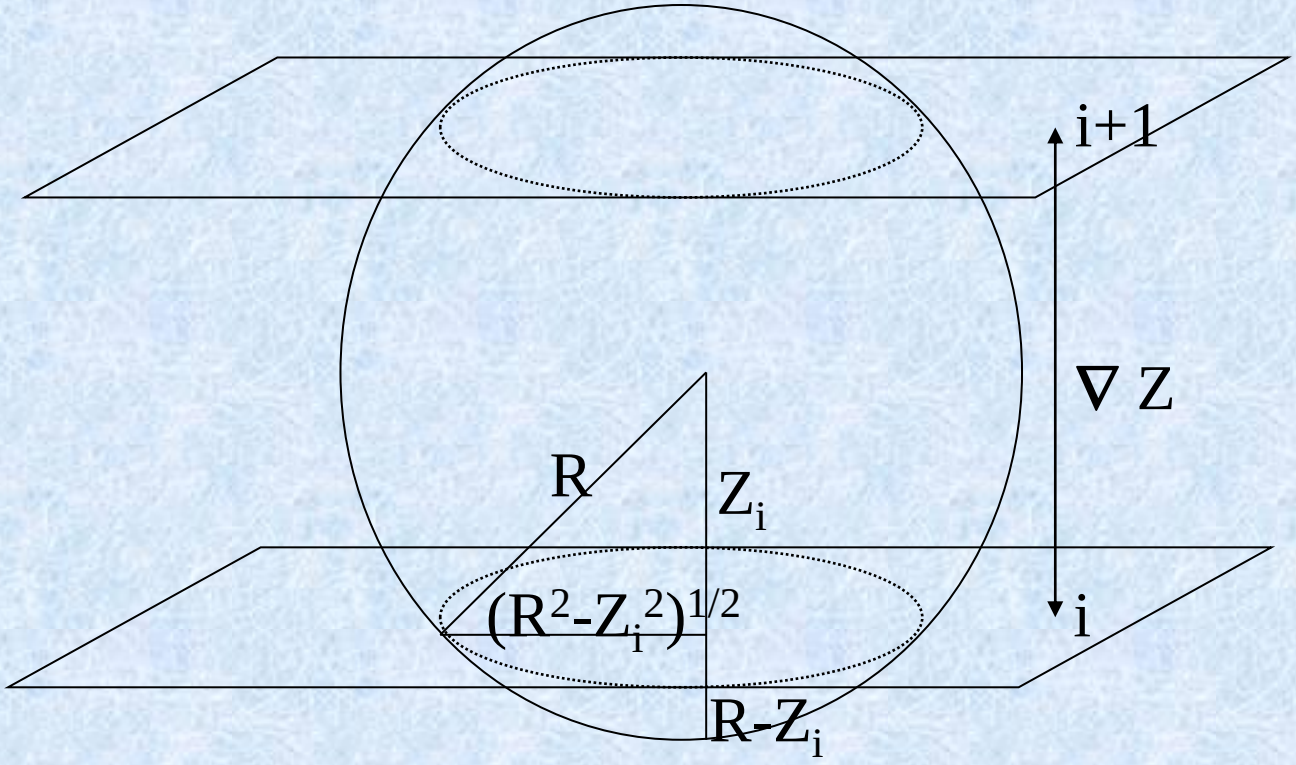
Generally, a sphere of water is assumed to be the solvent molecule with a radius 1.4 Å.

The solute molecule is represented by a set of interlocking spheres of appropriate van der Waals radii assigned to each atom

The solvent molecule is rolled along the envelope of the van der Waals surface at planes conveniently sectioned.



The accessible surface area of an atom of radius *r* is then the area on the surface of sphere of radius $R = r + r_{solv}$ on each point of which the center of the solvent molecule can be placed in contact with this atom without penetrating any other atoms of the solute molecule.



- **The solvent accessible surface area (ASA) is calculated using the formula (Lee and Richards, 1971):**

- **$ASA = \sum[R/(R^2 - Z_i^2)^{1/2}] L_i . D; D = \nabla Z/2 + \nabla' Z$**

- **where, L_i is the length of the arc computed on a given section *i*, Z_i is the perpendicular distance from the center of the sphere to the section *i*, ∇Z is the spacing between the sections and $\nabla' Z$ is $\nabla Z/2$ or $R-Z_i$, whichever is smaller. Summation is over all of the arcs drawn for the given atom.**

Methods

ACCESS: Lee and Richards (1978)

NACCESS: Hubbard and Thornton (1993)

ASC: Eisenhaber and Argos (1993)

DSSP: Kabsch and Sandor (1983)

GETAREA: Fraczkiewicz and Braun (1998)

Comparison index	ACCESS	DSSP	NACCESS	ASC	GETAREA
Standalone executable availability	Yes	Yes	Licensed	Yes	No
Online calculations/ database	No	Yes	No	Yes	Yes
Polar and nonpolar area	No	No	Yes	No	Yes
Atom-wise surface area	Yes	No	Yes	Yes	Yes
Source code availability	No	Yes	No	Yes	No
Choice of probe radius	Yes	No	Yes	Yes	No
Choice of van der Waals and other parameters	Yes	No	Yes	Yes	By Manual editing
Secondary structure	No	Yes	No	No	No
Reference	and Richards, 1978	Kabsch and Sander, 1983	Hubbard and Thornton, 1993	Eisenhaber and , 1993	Fraczkiewicz and Braun, 1998

```
==== Secondary Structure Definition by the program DSSP, updated CMBI version by ElmK / April 1,2000 ==== DATE=2-APR-2004      .
REFERENCE W. KABSCH AND C.SANDER, BIOPOLYMERS 22 (1983) 2577-2637                                                            .
HEADER    HYDROLASE (O-GLYCOSYL)                  12-OCT-84   1LZ1                                                       .
COMPND    LYSOZYME (E.C.3.2.1.17)                                                                                        .
SOURCE    HUMAN (HOMO $SAPIENS)                                                                                          .
AUTHOR    P.J.ARTYMIUK,C.C.F.BLAKE                                                                                       .
  130   1   4   4   0 TOTAL NUMBER OF RESIDUES, NUMBER OF CHAINS, NUMBER OF SS-BRIDGES(TOTAL,INTRACHAIN,INTERCHAIN)      .
  6787.0    ACCESSIBLE SURFACE OF PROTEIN (ANGSTROM**2)                                                                  .
   89 68.5    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(J)  , SAME NUMBER PER 100 RESIDUES                       .
    2  1.5    TOTAL NUMBER OF HYDROGEN BONDS IN     PARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES                       .
   12  9.2    TOTAL NUMBER OF HYDROGEN BONDS IN ANTIPARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES                       .
    2  1.5    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-5), SAME NUMBER PER 100 RESIDUES                       .
    1  0.8    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-4), SAME NUMBER PER 100 RESIDUES                       .
    1  0.8    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-3), SAME NUMBER PER 100 RESIDUES                       .
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-2), SAME NUMBER PER 100 RESIDUES                       .
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-1), SAME NUMBER PER 100 RESIDUES                       .
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+0), SAME NUMBER PER 100 RESIDUES                       .
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+1), SAME NUMBER PER 100 RESIDUES                       .
    9  6.9    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+2), SAME NUMBER PER 100 RESIDUES                       .
   25 19.2    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+3), SAME NUMBER PER 100 RESIDUES                       .
   34 26.2    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+4), SAME NUMBER PER 100 RESIDUES                       .
    5  3.8    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+5), SAME NUMBER PER 100 RESIDUES                       .
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30     *** HISTOGRAMS OF ***     .
 0  0  0  0  0  1  0  0  0  1  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    RESIDUES PER ALPHA HELIX   .
 1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    PARALLEL BRIDGES PER LADDER .
 2  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    ANTIPARALLEL BRIDGES PER LADDER .
 0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    LADDERS PER SHEET          .
  #  RESIDUE AA STRUCTURE BP1 BP2  ACC     N-H-->O    O-->H-N    N-H-->O    O-->H-N     TCO  KAPPA ALPHA  PHI   PSI    X-CA   Y-CA   Z-CA
    1    1   K              0   0   63     0, 0.0    39,-2.6     0, 0.0     2,-0.7    0.000 360.0 360.0 360.0 146.7    1.9   20.6   21.1
    2    2   V  B     -A   39   0A  99    37,-0.2    37,-0.2    38,-0.1     2,-0.1   -0.949 360.0-145.6-104.2 112.9    4.6   19.8   18.5
    3    3   F              0   0   10    35,-2.4     2,-0.2    -2,-0.7     3, 0.0   -0.453  10.1-122.2 -70.2 150.5    7.4   22.4   18.9
    4    4   E    >  -      0   0  146    -2,-0.1     4,-2.6     1,-0.1     5,-0.3   -0.657  34.8-107.1 -82.3 155.8    9.4   23.7   16.0
    5    5   R  H  > S+     0   0   91    -2,-0.2     4,-2.1     1,-0.2     5,-0.1    0.886 115.0  37.2 -56.3 -50.8   13.1   23.1   16.5
    6    6   a  H  > S+     0   0   27     1,-0.2     4,-2.5     2,-0.2    -1,-0.2    0.826 112.9  58.3 -78.4 -24.6   14.3   26.7   17.2
    7    7   E  H  > S+     0   0   58     1,-0.2     4,-2.3     2,-0.2    -1,-0.2    0.907 109.6  45.4 -60.1 -45.4   11.1   27.5   19.2
    8    8   L  H  X S+     0   0    1    -4,-2.6     4,-3.2     2,-0.2     5,-0.3    0.903 110.4  54.1 -62.5 -42.1   12.0   24.6   21.5
    9    9   A  H  X S+     0   0    0    -4,-2.1     4,-2.1    -5,-0.3    -2,-0.2    0.937 111.4  43.9 -61.1 -44.3   15.7   25.6   21.7
   10   10   R  H  X S+     0   0  114    -4,-2.5     4,-2.5     2,-0.2    -1,-0.2    0.906 113.6  51.9 -67.4 -34.7   14.8   29.2   22.8
   11   11   T  H  X S+     0   0   29    -4,-2.3     4,-1.3     1,-0.2    -2,-0.2    0.910 111.0  46.7 -66.8 -47.6   12.2   27.8   25.3
   12   12   L  H  <>S+     0   0    0    -4,-3.2     5,-2.5     2,-0.2     6,-0.3    0.845 111.4  52.9 -61.5 -37.6   14.7   25.4   26.8
   13   13   K  H  ><5S+     0   0   89    -4,-2.1     3,-1.8    -5,-0.3     5,-0.3    0.970 109.3  47.8 -61.0 -44.3   17.3   28.3   27.0
   14   14   R  H  3<5S+     0   0  1[illegible]    -4,-2.5    -1,-0.2     1,-0.3    -2,-0.2    0.809 106.9  58.1 -67.8 -23.4   14.9   30.5   28.9
   15   15   L  T  3<5S-     0   0   49    -4,-1.3    -1,-0.3    -5,-0.2    -2,-0.2    0.399 121.3-107.0 -91.0    6.3   14.0   27.6   31.3
   16   16   G  T  < 5S+     0   0   40    -3,-1.8    -3,-0.2    -4,-0.3    -2,-0.1    0.719  79.9 128.7  86.8   24.2   17.7   27.4   32.3
   17   17   M     > < +      0   0    0    -5,-2.5     3,-1.7     2,-0.1     2,-0.6    0.625  35.2 105.3 -87.8 -19.2   18.7   24.2   30.5
   18   18   D  T 3  S-      0   0   78    -6,-0.3     6,-0.2    -5,-0.3    4,-0.1    -0.578 104.8  -7.7 -68.4 107.8   21.8   25.4   28.6
   19   19   G  T 3  S+      0   0   44     4,-2.4     3,-2.1    -2,-0.6     2,-0.4    0.604  88.0 164.0  88.8   6.3   24.8   23.9   30.6
   20   20   Y  B  X  S-B   23   0B  54    -3,-1.7     3,-1.6     3,-0.8    -1,-0.3   -0.584  80.4 -10.2 -63.8 126.7   22.6   22.6   33.4
   21   21   R  T 3  S-      0   0  165    -2,-0.4    -1,-0.3     1,-0.2     3,-0.1    0.784 134.6 -56.6  46.1  39.0   24.8   20.0   35.2
   22   22   G  T <  S+      0   0   75    -3,-2.1     2,-0.7     1,-0.2    -1,-0.2    0.507 104.9 132.1  83.1   6.6   27.2   20.4   32.3
   23   23   I  B  X   -B   20   0B  15    -3,-1.6    -4,-2.4    -6,-0.1    -3,-0.8   -0.753  50.3-136.2-103.2 121.5   24.8   19.5   29.6
```

Secondary structure

Solvent accessibility

Hydrogen bonding partners/ Electrostatic energy

GETAREA

Calculation of Solvent Accessible Surface Areas, Atomic Solvation Energies and Their Gradients for Macromolecules

Robert Fraczkiewicz and Werner Braun

Sealy Center for Structural Biology
University of Texas Medical Branch
Galveston, TX 77555, USA

Quite often a biomolecular researcher wants to quickly calculate solvent accessible surface area or solvation energy of, for example, a protein molecule but does not have time/resources/skills/will to find and install an appropriate software. GETAREA, our efficient method of calculating the solvent accesible surface area [1,2] implemented in program FANTOM, can be directly accessed through this form. An on-line manual is available. Atomic coordinates should be supplied in PDB format. Please cite reference [1] in publications that use our service. Comments are welcome, please mail them to webraun@utmb.edu. Sample Getarea output: getarea.pdf

References:

1. Fraczkiewicz, R. and Braun, W. (1998) "Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules" *J. Comp. Chem.*, **19**, 319-333.
2. Fraczkiewicz, R. and Braun, W. "A New Efficient Algorithm for Calculating Solvent Accessible Surface Areas of Macromolecules" presented at the *Third Electronic Computational Chemistry Conference*; Northern Illinois University; November 1996; World Wide Web.

Please select your PDB File: Choose File 2PTL.pdb

Please Enter radius of the water probe(A): 1.4

Do you want gradient in calculations: n y or n, Default is n

Please enter your Email Address: gromiha@iitm.ac.in For information purpose only

Select desired level of output: 2. Area/energy per residue

Submit for Analysis | Reset to Default values

GLN	32	98.49	34.65	6.46	9
THR	33	76.88	62.60	12.26	6
ALA	34	6.64	6.64	6.24	
GLU	35	119.15	44.05	17.76	10
PHE	36	20.53	20.53	3.05	1
LYS	37	109.82	52.86	13.38	9

33	CA	LYS	3	9.
34	C	LYS	3	0.
35	O	LYS	3	3.
36	CB	LYS	3	25.
37	CG	LYS	3	13.
38	CD	LYS	3	35.

```
APOLAR area/energy          =          5001.36
UNKNOW area/energy          =             0.00
-------------------------------------------------
Total  area/energy          =          8496.85
-------------------------------------------------
Number of surface atoms     =  723
Number of buried atoms      =  586
```

Pictorial representation: ASAView

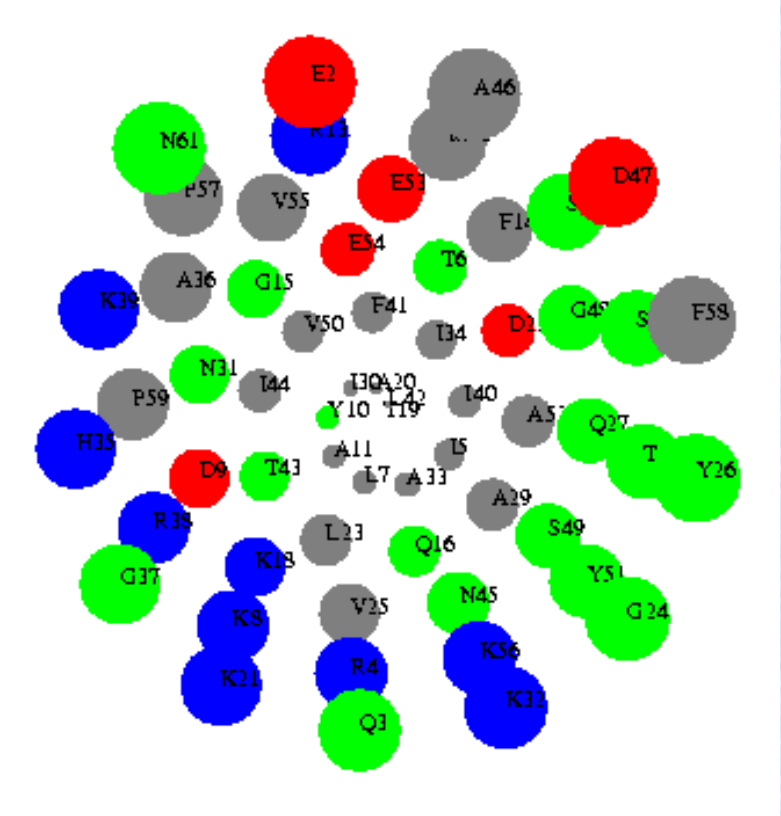
ASAview is an algorithm and a database of schematic representations of solvent accessibility of residues in a protein.

In this program, a characteristic two-dimensional spiral plot of solvent accessibility has been implemented for providing a convenient graphical view of residues in terms of their exposed surface areas.

Further, the sequential plots are also provided in the form of bar charts.

Online plots of the proteins included in the entire Protein Data Bank (PDB), are provided for the whole protein as well as their chains separately.

ASAview

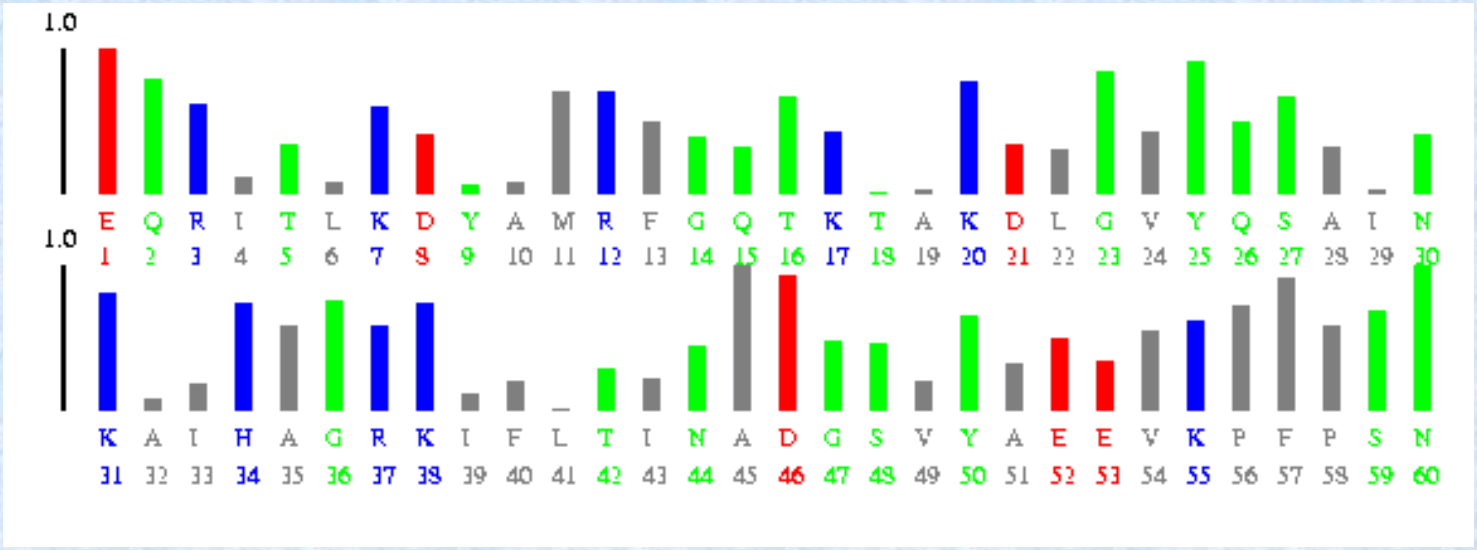


Spiral view

blue, red, green, gray and yellow colors indicates the positively charged, negatively charged, polar, non-polar and Cys residues, respectively. The size of the sphere shows the relative ASA

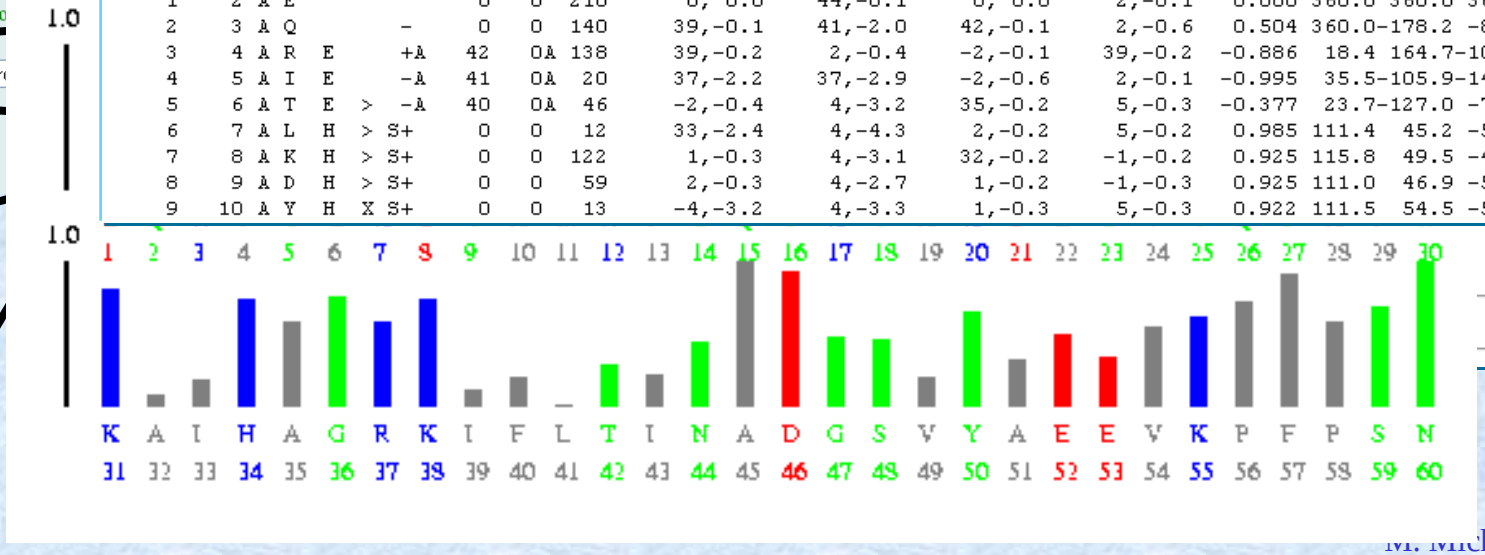
S. Ahmad, M.M. Gromiha and A. Sarai, (2004) BMC Bioinformatics

ASAview


1.0
E Q R I T L K D Y A M R F G Q T K T A K D L G V Y Q S A I N
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
1.0
K A I H A G R K I F L T I N A D G S V Y A E E V K P F P S N
31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

```
==== Secondary Structure Definition by the program DSSP, updated CMBI version by ElmK / April 1,2000 ==== DATE=2
REFERENCE W. KABSCH AND C.SANDER, BIOPOLYMERS 22 (1983) 2577-2637
HEADER    GENE REGULATION/DNA                     22-APR-98   6CRO
COMPND    2 MOLECULE: DNA (5'-
SOURCE    2 SYNTHETIC: YES;
AUTHOR    R.A.ALBRIGHT,B.W.MATTHEWS
   60   1   0   0   0 TOTAL NUMBER OF RESIDUES, NUMBER OF CHAINS, NUMBER OF SS-BRIDGES(TOTAL,INTRACHAIN,INTERCHAIN)
 4773.0    ACCESSIBLE SURFACE OF PROTEIN (ANGSTROM**2)
   40 66.7    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(J)  , SAME NUMBER PER 100 RESIDUES
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(J)  , ANTIPARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES
   11 18.3    TOTAL NUMBER OF HYDROGEN BONDS IN     PARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS IN ANTIPARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-5), SAME NUMBER PER 100 RESIDUES
    1  1.7    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-4), SAME NUMBER PER 100 RESIDUES
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-3), SAME NUMBER PER 100 RESIDUES
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-2), SAME NUMBER PER 100 RESIDUES
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-1), SAME NUMBER PER 100 RESIDUES
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+0), SAME NUMBER PER 100 RESIDUES
    0  0.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+1), SAME NUMBER PER 100 RESIDUES
    3  5.0    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+2), SAME NUMBER PER 100 RESIDUES
    4  6.7    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+3), SAME NUMBER PER 100 RESIDUES
   20 33.3    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+4), SAME NUMBER PER 100 RESIDUES
    2  3.3    TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+5), SAME NUMBER PER 100 RESIDUES
  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30      *** HISTOGRAMS OF
  0  0  0  0  0  0  0  2  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0     RESIDUES PER ALPHA
  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0     PARALLEL BRIDGES P
  0  0  1  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0     ANTIPARALLEL BRIDG
  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0     LADDERS PER SHEET
  #  RESIDUE AA STRUCTURE BP1 BP2  ACC     N-H-->O    O-->H-N    N-H-->O    O-->H-N     TCO  KAPPA ALPHA  PHI   P
    1    2 A E              0   0  210      0, 0.0    44,-0.1     0, 0.0     2,-0.1   0.000 360.0 360.0 360.0   8
    2    3 A Q        -     0   0  140     39,-0.1    41,-2.0    42,-0.1     2,-0.6   0.504 360.0-178.2 -87.1 14
    3    4 A R  E     +A   42   0A 138     39,-0.2     2,-0.4    -2,-0.1    39,-0.2  -0.886  18.4 164.7-100.9 12
    4    5 A I  E     -A   41   0A  20     37,-2.2    37,-2.9    -2,-0.6     2,-0.1  -0.995  35.5-105.9-146.6 15
    5    6 A T  E  >  -A   40   0A  46     -2,-0.4     4,-3.2    35,-0.2     5,-0.3  -0.377  23.7-127.0 -71.5 14
    6    7 A L  H  > S+     0   0   12     33,-2.4     4,-4.3     2,-0.2     5,-0.2   0.985 111.4  45.2 -58.0 -5
    7    8 A K  H  > S+     0   0  122      1,-0.3     4,-3.1    32,-0.2    -1,-0.2   0.925 115.8  49.5 -45.6 -5
    8    9 A D  H  > S+     0   0   59      2,-0.3     4,-2.7     1,-0.2    -1,-0.3   0.925 111.0  46.9 -50.4 -5
    9   10 A Y  H  X S+     0   0   13     -4,-3.2     4,-3.3     1,-0.3     5,-0.3   0.922 111.5  54.5 -57.6 -4
```

```
E2 100.0
Q3 78.4
R4 60.3
I5 10.8
T6 33.2
L7 6.6
K8 59.3
D9 40.9
Y10 6.1
A11 6.4
M12 70.0
R13 70.3
F14 48.3
G15 38.1
Q16 30.8
T17 65.6
K18 41.3
T19 0.0
A20 1.8
K21 75.8
D22 32.5
L23 30.6
G24 83.9
V25 41.6
Y26 90.8
Q27 48.2
S28 65.7
A29 31.8
I30 2.7
N31 39.6
K32 80.7
A33 7.3
I34 18.4
H35 74.2
A36 58.1
G37 75.0
R38 58.5
K39 73.9
```



Percentage Accessibility

Percentage accessibility

Ratio between

Accessible surface area computed with 3D structure and

Accessible surface area in extended state

Gly-X-Gly or Ala-X-Ala

The values are Ala-110.2; Asp-144.1; Cys-140.4; Glu-174.7; Phe-200.7; Gly-78.7; His-181.9; Ile-185.0; Lys-205.7; Leu-183.1; Met-200.1; Asn-146.4; Pro-141.9; Gln-178.6; Arg-229.0; Ser-117.2; Thr-138.7; Val-153.7; Trp-240.5; Tyr-213.7 (the units are in A**2).

ASA less than 5%: Buried

ASA between 5% and 20%: Partially buried

ASA between 20% and 50%: Partially exposed

ASA more than 50%: Exposed.