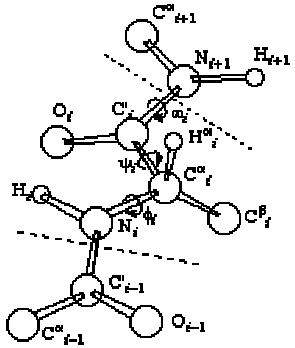
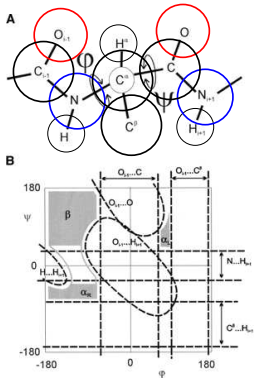
Polypeptide backbone (or the main chain)



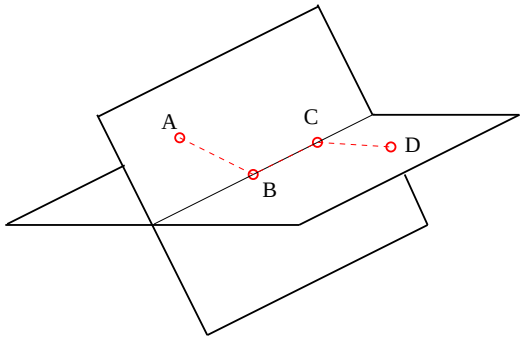
[IUPAC-IUB Commission on Biochemical Nomenclature, Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. Eur. J. Biochem., 1969, 17, 193-201]

Ramachandran steric map



[Ho, K.H., Thomas, A. and Brasseur, R., Protein Science, 2003, 12:2508-2522]

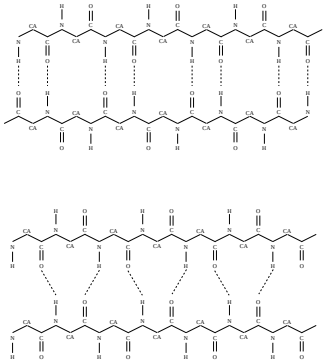
Dihedral angle



Protein Data Bank entry (extract)

```
COMPND    TRIOSE PHOSPHATE ISOMERASE (E.C.5.3.1.1)
SOURCE    CHICKEN (GALLUS GALLUS) BREAST MUSCLE
AUTHOR    D.W.BANNER,A.C.BLOOMER,G.A.PETSKO,D.C.PHILLIPS,
AUTHOR   2 I.A.WILSON
  :
JRNL        AUTH   D.W.BANNER,A.C.BLOOMER,G.A.PETSKO,D.C.PHILLIPS,
JRNL        AUTH 2 I.A.WILSON
JRNL        TITL   ATOMIC COORDINATES FOR TRIOSE PHOSPHATE ISOMERASE
JRNL        TITL 2 FROM CHICKEN MUSCLE
JRNL        REF    BIOCHEM.BIOPHYS.RES.COMM.     V.  72   146 1976
JRNL        REFN   ASTM BBRCA9  US ISSN 0006-291X                  146
  :
REMARK   2 RESOLUTION. 2.5 ANGSTROMS.
  :
SEQRES   1 A  247  ALA PRO ARG LYS PHE PHE VAL GLY GLY ASN TRP LYS MET
SEQRES   2 A  247  ASN GLY LYS ARG LYS SER LEU GLY GLU LEU ILE HIS THR
  :
ATOM      1  N   ALA A   1      43.240  11.990  -6.915  1.00  0.00
ATOM      2  CA  ALA A   1      43.888  10.862  -6.231  1.00  0.00
ATOM      3  C   ALA A   1      44.791  11.378  -5.094  1.00  0.00
ATOM      4  O   ALA A   1      44.633  10.992  -3.937  1.00  0.00
ATOM      5  CB  ALA A   1      44.722  10.051  -7.240  1.00  0.00
ATOM      6  N   PRO A   2      45.714  12.244  -5.497  1.00  0.00
ATOM      7  CA  PRO A   2      46.689  12.815  -4.561  1.00  0.00
ATOM      8  C   PRO A   2      46.042  13.601  -3.411  1.00  0.00
ATOM      9  O   PRO A   2      46.030  13.141  -2.267  1.00  0.00
  :
```

Hydrogen bonds in β-sheets



DSSP summary codes

H 4-helix (α-helix)
B residue in isolated β-bridge
E extended strand, participates in β-ladder
G 3-helix
I 5-helix
T H-bonded turn
S bend

Crambin (1CRN)

```
TTCCPSIVARSNFNVCRLPGTPEAICATYTGCIIIPGATCPGDYAN
 EE SSHHHHHHHHHHHHTT  HHHHHHHHS EE SSS   TTS
```


DSSP

Hydrogen bond energy

$$E = q_1 q_2 \left(\frac{1}{d(ON)} + \frac{1}{d(CH)} - \frac{1}{d(OH)} - \frac{1}{d(CN)} \right) \times f$$

Antiparallel bridge:

[hbond(i,j) and hbond(j,i)]
or
[hbond(i-1,j+1) and hbond(j-1,i+1)]

Parallel bridge:

[hbond(i-1,j) and hbond(j,i+1)]
or
hbond(j-1,i) and hbond(i,j+1)]

Protein stability

- good stereochemistry; no steric clashes;
- buried charged atoms must be paired;
- enough hydrophobic surface must be buried, and the interior must be sufficiently densely packed, to provide thermodynamic stability.

Modular proteins

- multi-domain proteins, often with many copies of related domains;
- domains recur in many proteins in different structural contexts.

3D Transformations

Translation

$$T(d_x, d_y, d_z) = \begin{bmatrix} 1 & 0 & 0 & d_x \\ 0 & 1 & 0 & d_y \\ 0 & 0 & 1 & d_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Scaling

$$S(s_x, s_y, s_z) = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Rotation

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_z(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

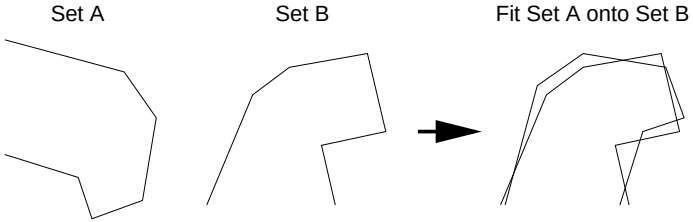
Domain definitions

"The definition of protein domains varies widely across the discipline of biology. Domains are defined simultaneously as:

(1) regions that display a significant level of sequence homology;
(2) a minimal part of the gene that is capable of performing a function;
(3) a region of the protein with an experimentally assigned function;
(4) parts of structures that have significant structural similarity; and
(5) compact spatially distinct units of protein structure."

[Veretnik S, Bourne PE, Alexandrov NN, Shindyalov IN.
Toward consistent assignment of structural domains in proteins.
J Mol Biol. 2004 Jun 4;339(3):647-678.]

Comparing molecular fragments



- 3-D transformation to map Set A onto Set B
- Root Mean Square (RMS) distance

DOMAK (DOmain MAKer)

[Siddiqui AS, Barton GJ. Continuous and discontinuous domains: an algorithm for the automatic generation of reliable protein domain definitions. Protein Sci. 1995 May;4(5):872-884.]

Based on the principle that the residues comprising a domain make more contacts between themselves (internal contacts) than they do with the rest of the protein (external contacts).

Two residues make contact if a heavy atom of one is within 5*angstrom* of a heavy atom of the other.

$$\frac{int_A}{ext_{AB}} \times \frac{int_B}{ext_{AB}}$$

Can deal with domains consisting of two segments.

STRUDL (STRUctural Domain Limits)

[Wernisch L, Hunting M, Wodak SJ. Identification of structural domains in proteins by a graph heuristic. Proteins. 1999 May 15;35(3):338-352.]

Algorithm designed to identify domains with any number of non-contiguous chain segments.

Uses the Kernighan-Lin graph heuristic to partition the protein into residue sets which display minimum interactions between them.

Interactions are deduced from contact areas between atoms in the weighted Voronoi diagram.

The radius of the "accessible sphere" around each atom is the van der Waals radius of the atom increased by 1.4*angstrom*.

Conflicting domain assignments

"The major factors responsible for conflicting domain assignments between methods, both experts and automatic, are:

(1) the definition of very small domains;
(2) splitting secondary structures between domains;
(3) the size and number of discontinuous domains;
(4) closely packed or convoluted domain-domain interfaces;
(5) structures with large and complex architectures; and
(6) the level of significance placed upon structural, functional and evolutionary concepts in considering structural domain definitions."

[Veretnik S, Bourne PE, Alexandrov NN, Shindyalov IN.
Toward consistent assignment of structural domains in proteins.
J Mol Biol. 2004 Jun 4;339(3):647-678.]