Bioinformatics

Protein structure modelling

Bioinformatics - lectures

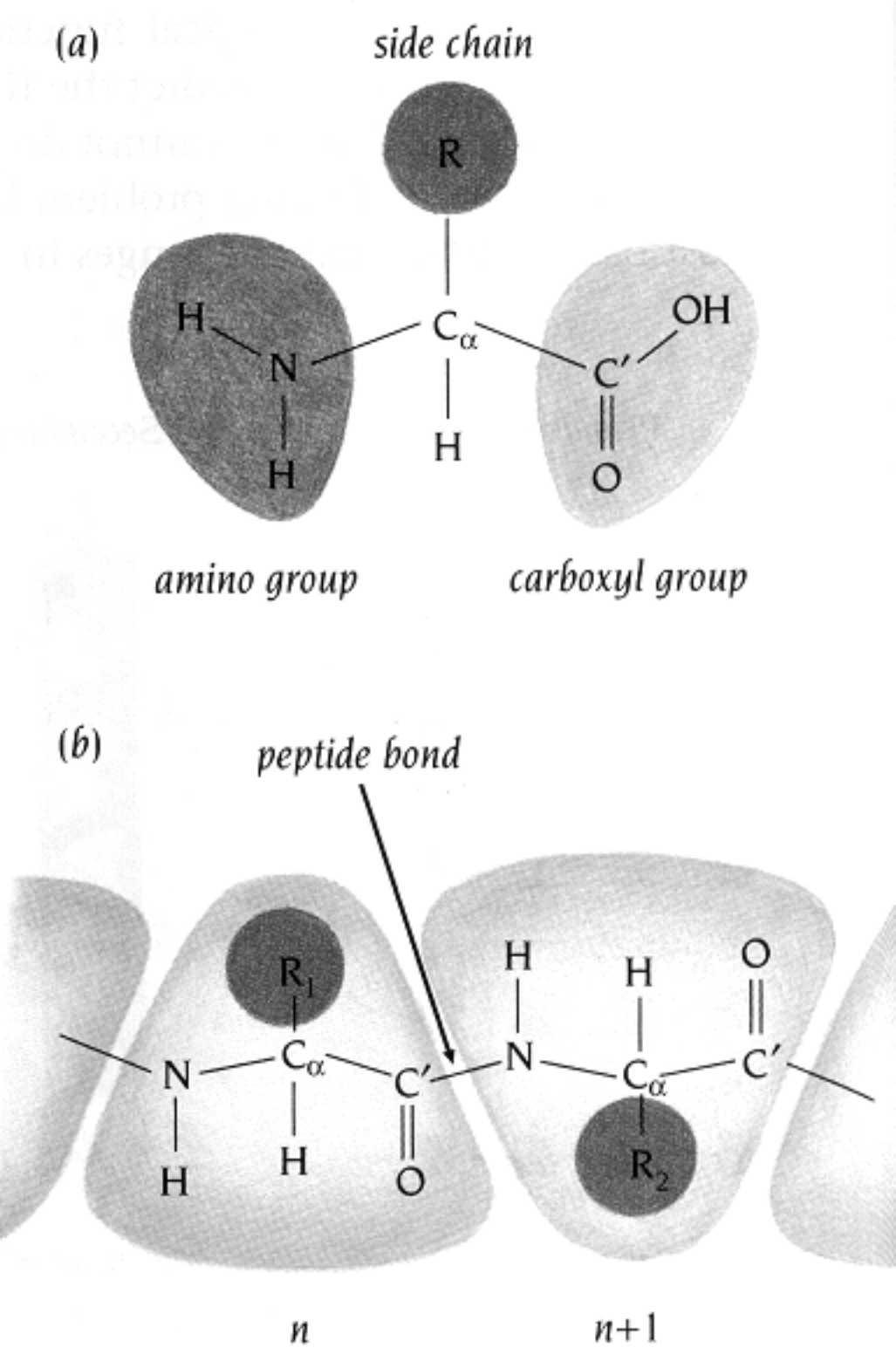
- Introduction
- Information networks
- Protein information resources
- Genome information resources
- DNA sequence analysis
- Pairwise sequence alignment
- Multiple sequence alignment
- Secondary database searching
- Analysis packages
- Protein structure modelling

Protein structure modelling

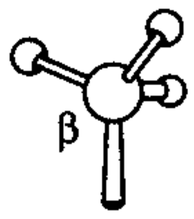
- protein structure
- protein structure databases
- prediction of secondary structure
- prediction of protein fold
- prediction of tertiary structure
- modelling of protein-ligand complexes

Protein structure

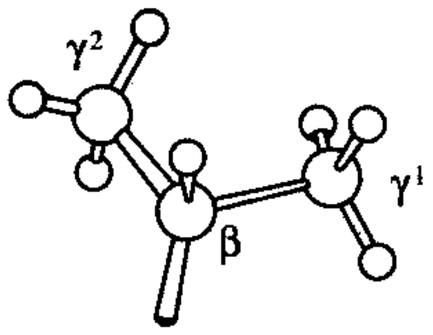
- Proteins are build up by amino acids that are linked by peptide bonds. The 20 different amino acids occur naturally in proteins.
- Protein structure can be experimentally determined by X-ray crystallography, nuclear magnetic resonance (NMR) or by electron crystallography.
- Levels of protein structure:
  - primary structure
  - secondary structure
  - supersecondary structure
  - tertiary structure
  - quaternary structure


(a)
side chain
R
H
N
H
C_α
H
C′
O
OH
amino group
carboxyl group
(b)
peptide bond
R_1
N
H
C_α
H
C′
O
H
N
H
C_α
H
R_2
O
C′
n
n+1

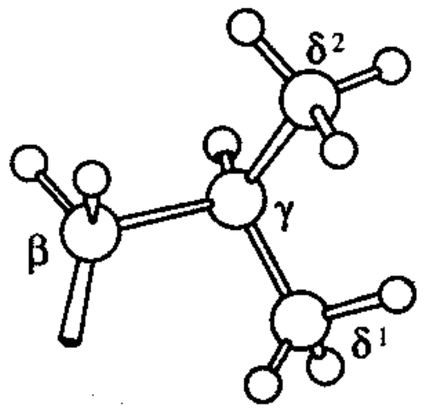
Glycine
Gly
G



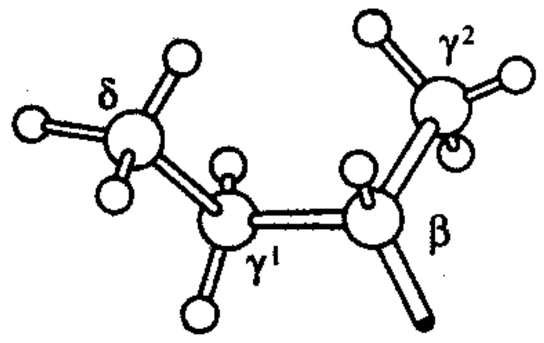
Alanine
Ala
A



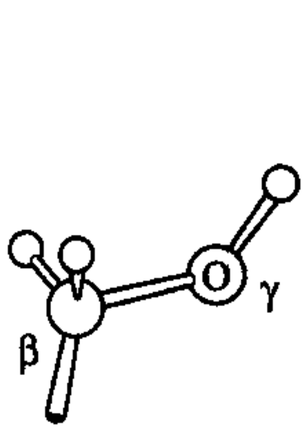
Valine
Val
V



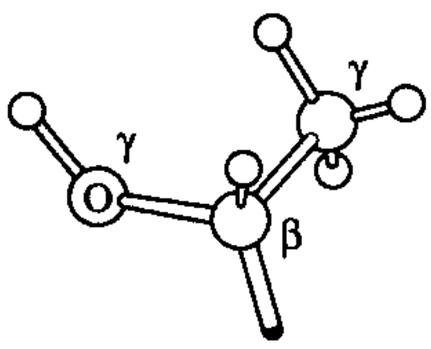
Leucine
Leu
L



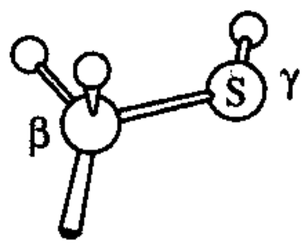
Isoleucine
Ile
I



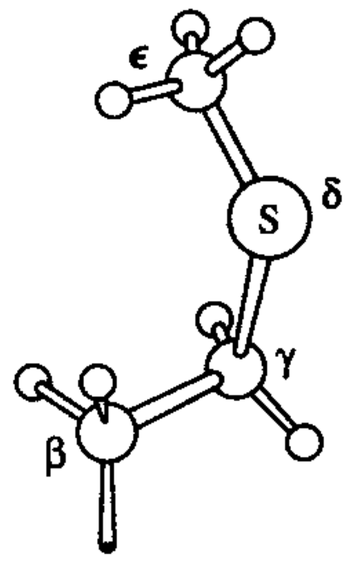
Serine
Ser
S



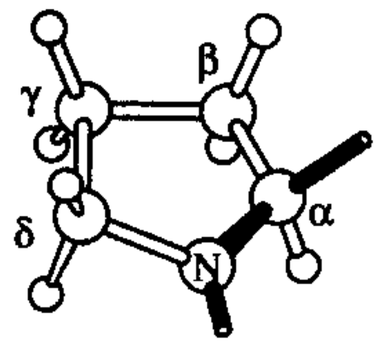
Threonine
Thr
T



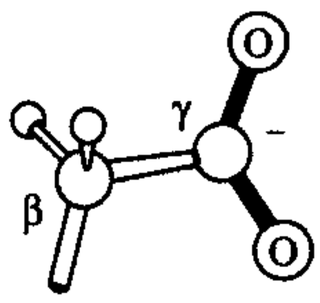
Cysteine
Cys
C



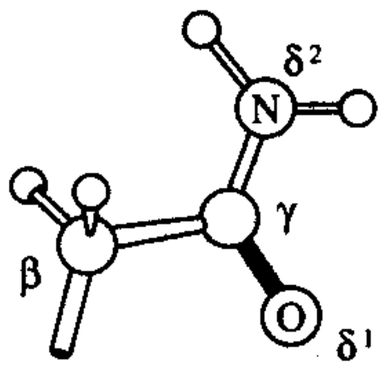
Methionine
Met
M



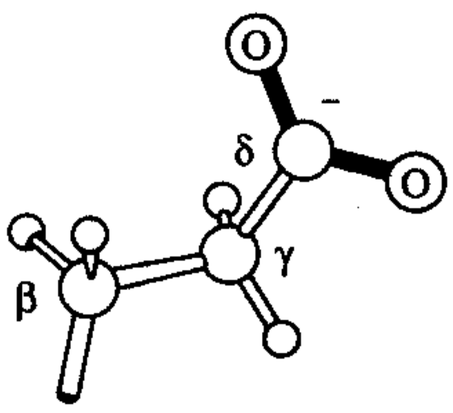
Proline
Pro
P



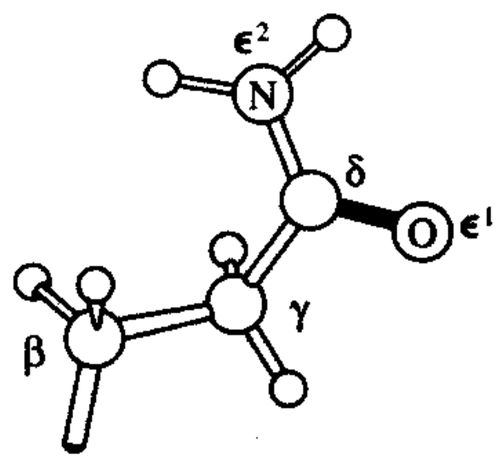
Aspartic acid
Asp
D



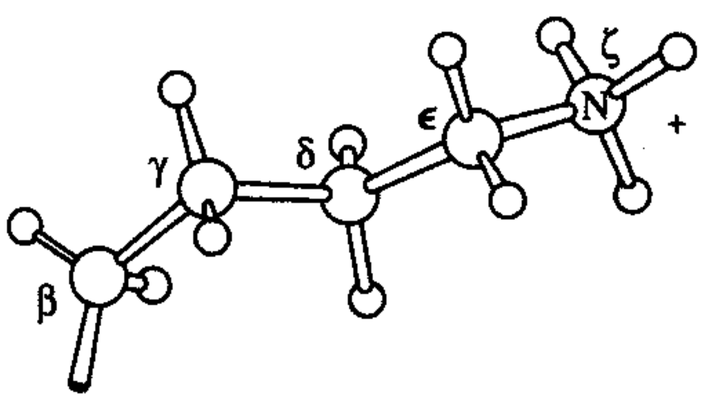
Asparagine
Asn
N



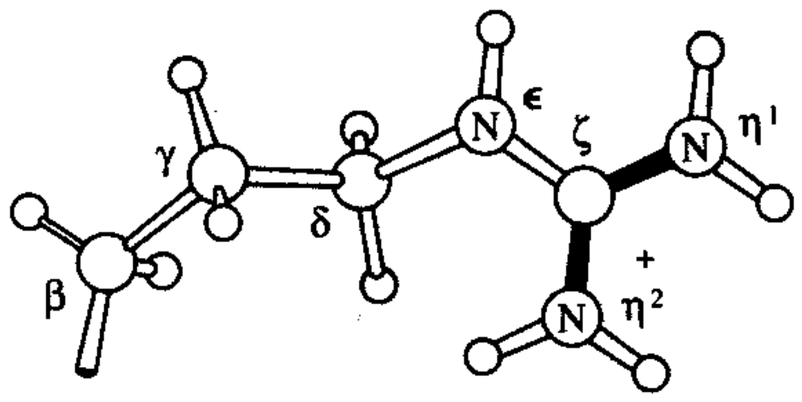
Glutamic acid
Glu
E



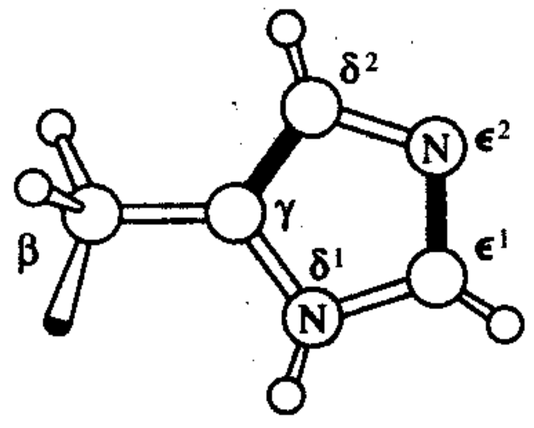
Glutamine
Gln
Q



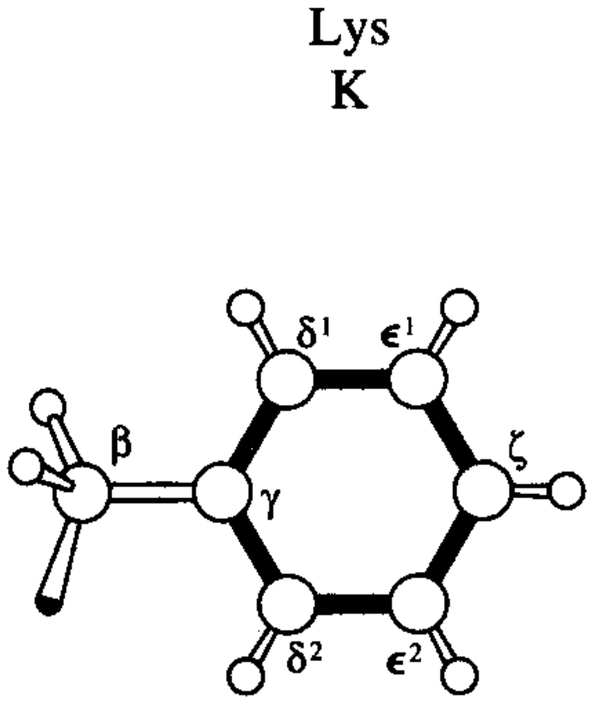
Lysine
Lys
K



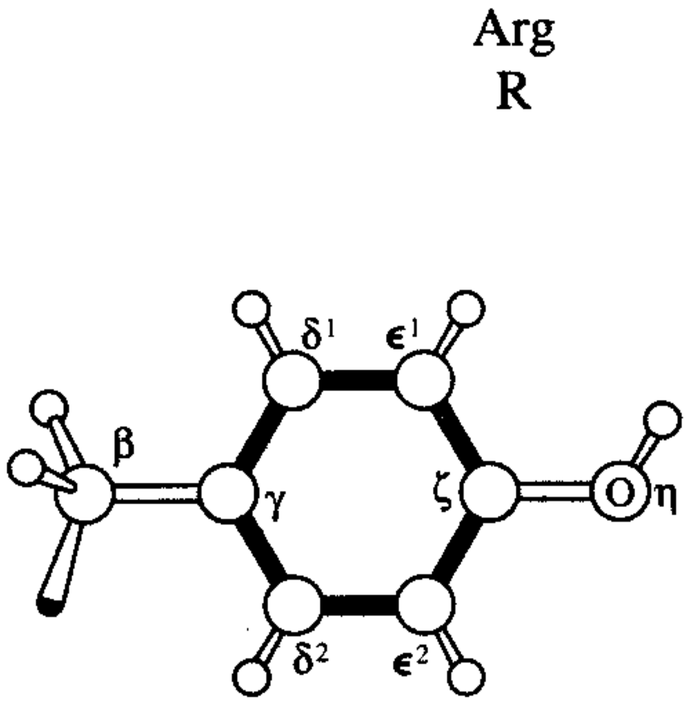
Arginine
Arg
R



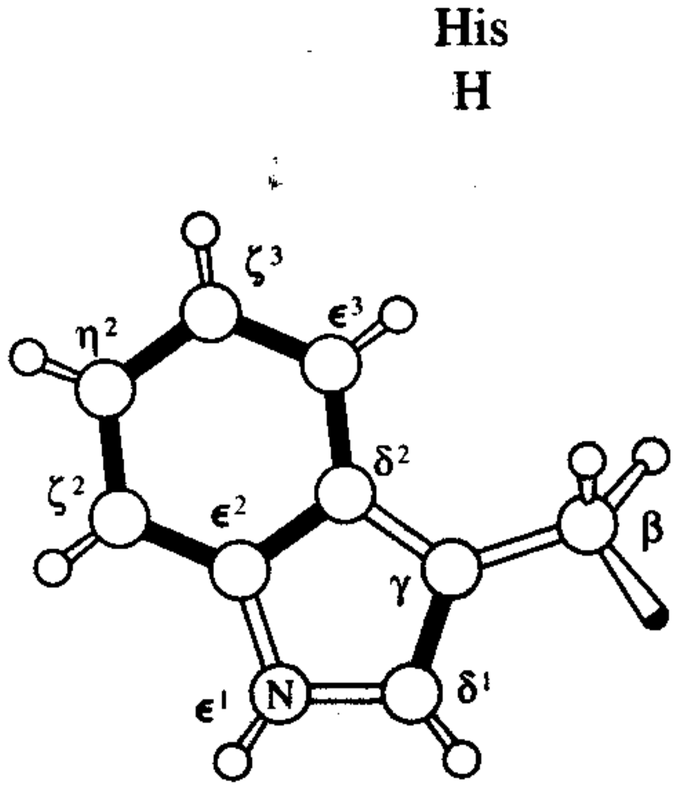
Histidine
His
H



Phenylalanine
Phe
F

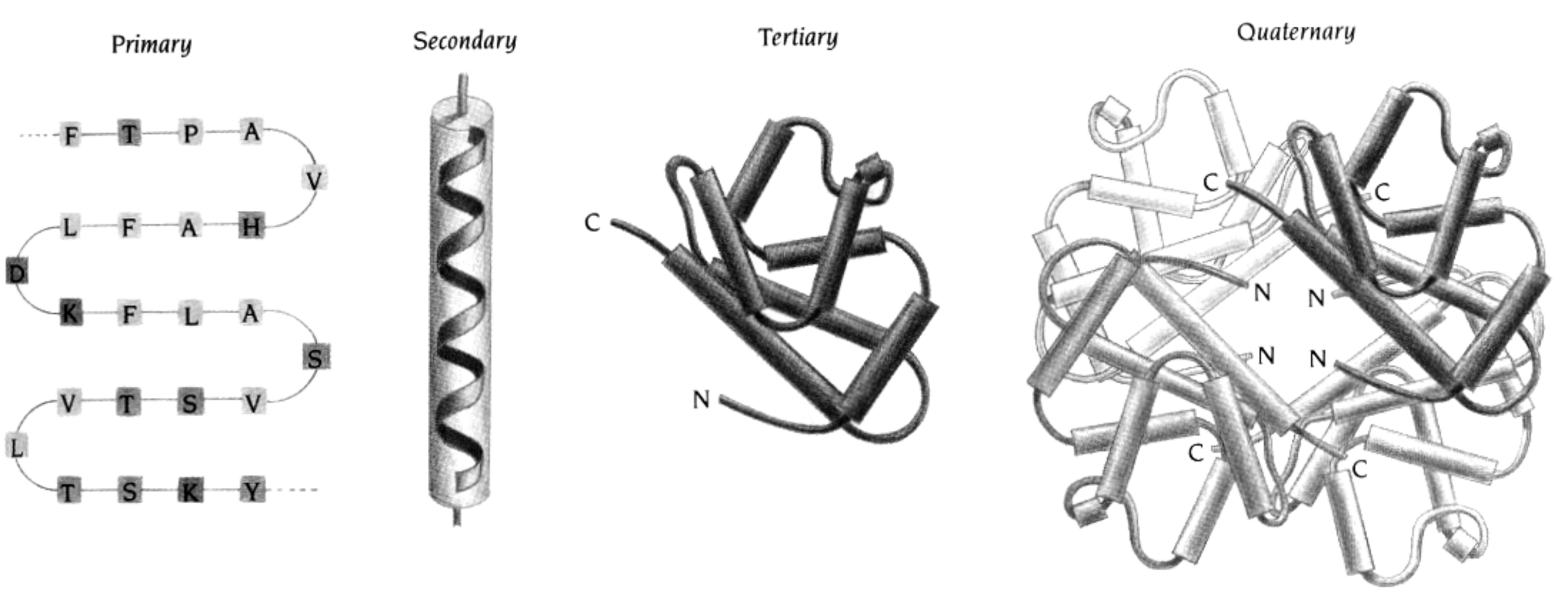


Tyrosine
Tyr
Y



Tryptophan
Trp
W

Primary structure:	the linear sequence of amino acids in a protein molecule
Secondary structure:	regions of local regularity within a protein fold (e.g., α-helices, β-turns, β-strands)
Super-secondary structure:	the arrangement of α-helices and/or β-strands into discrete folding units (e.g., β-barrels, βαβ-units, Greek keys, etc.)
Tertiary structure:	the overall fold of a protein sequence, formed by the packing of its secondary and/or super-secondary structure elements
Quaternary structure:	the arrangement of separate protein chains in a protein molecule with more than one subunit
Quinternary structure:	the arrangement of separate molecules, such as in protein–protein or protein–nucleic acid interactions


Primary
F T P A V L F A H D K F L A S V T S V L T S K Y
Secondary
Tertiary
C
N
Quaternary
C
C
N
N
N
N
C
C

Synchrotron radiation facility

European Synchrotron Radiation Facility at Grenoble, France

Protein structure databases

- PDB
- PDBsum

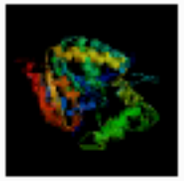
Protein structure classification databases

- SCOP
- CATCH

Protein structure databases

- PDB - Protein Data Bank
  - developed at Brookhaven National Laboratory
  - currently maintained by Research Collaboratory for Structural Bioinformatics (RSCB)
  - world repository of three-dimensional protein structures
  - entries from crystallographic analysis (80%), nuclear magnetic resonance (16%) and modelling (2%)
  - entries stored as flat files composed of section for information records and section for co-ordinates
  - entries identified by unique PDB-ID code (e.g., 1EDE)
  - searchable by keywords
  - interactive visualization of structures

Structure Explorer – 1CV2



Summary Information

Summary Information
View Structure
Download/Display File
Structural Neighbors
Geometry
Other Sources
Sequence Details

Explore
SearchLite SearchFields

Title: **Hydrolytic Haloalkane Dehalogenase Linb From Sphingomonas Paucimobilis Ut26 At 1.6 A Resolution**

Compound: **Mol_Id: 1; Molecule: Haloalkane Dehalogenase; Chain: A; Synonym: Linb, 1,3,4,6–Tetrachloro–1,4–Cyclohexadiene Hydrolase; Ec: 3.8.1.5; Engineered: Yes**

Authors: **J. Marek, J. Vevodova, J. Damborsky, I. Smatanova, L. A. Svensson, J. Newman, Y. Nagata, M. Takagi**

Exp. Method: **X–ray Diffraction**

Classification: **Hydrolase**

EC Number: **3.8.1.5**

Source: **Sphingomonas Paucimobilis**

Primary Citation: **Marek, J., Vevodova, J., Smatanova, I., Nagata, Y., Svensson, L. A., Newman, J., Takagi, M., Damborsky, J.: Crystal Structure of the Haloalkane Dehalogenase from Sphingomonas Paucimobilis Ut26 *Biochemistry* 39 *pp.* 14082 (2000)**
[Medline]

Deposition Date: **22–Aug–1999** *Release Date:* **11–Sep–2000**

Resolution [Å]: **1.58** *R–Value:* **0.149**

Space Group: **P 21 21 2**

Unit Cell: *dim [Å]:* *a* **50.26** *b* **71.67** *c* **72.70**
angles [°]: *alpha* **90.00** *beta* **90.00** *gamma* **90.00**

Polymer Chains: **A** *Residues:* **296**

Atoms: **2750**

HET groups: HOH

Entry from the PDB database (header)

```
HEADER    HYDROLASE                               22-AUG-99   1CV2
TITLE     HYDROLYTIC HALOALKANE DEHALOGENASE LINB FROM SPHINGOMONAS
TITLE    2 PAUCIMOBILIS UT26 AT 1.6 A RESOLUTION
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: HALOALKANE DEHALOGENASE;
COMPND   3 CHAIN: A;
COMPND   4 SYNONYM: LINB, 1,3,4,6-TETRACHLORO-1,4-CYCLOHEXADIENE
COMPND   5 HYDROLASE;
COMPND   6 EC: 3.8.1.5;
COMPND   7 ENGINEERED: YES;
COMPND   8 BIOLOGICAL_UNIT: MONOMER
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: SPHINGOMONAS PAUCIMOBILIS;
SOURCE   3 STRAIN: UT26;
SOURCE   4 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE   5 EXPRESSION_SYSTEM_STRAIN: HB101;
SOURCE   6 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE   7 EXPRESSION_SYSTEM_PLASMID: PMYLB1
KEYWDS    DEHALOGENASE, LINDANE, BIODEGRADATION, ALPHA/BETA-HYDROLASE
EXPDTA    X-RAY DIFFRACTION
AUTHOR    J.MAREK,J.VEVODOVA,J.DAMBORSKY,I.SMATANOVA,L.A.SVENSSON,
AUTHOR   2 J.NEWMAN,Y.NAGATA,M.TAKAGI
REMARK   1 REFERENCE 1
REMARK   1  AUTH   I.SMATANOVA,Y.NAGATA,L.A.SVENSSON,M.TAKAGI,J.MAREK
REMARK   1  TITL   CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION
REMARK   1  TITL 2 ANALYSIS OF HALOALKANE DEHALOGENASE LINB FROM
REMARK   1  TITL 3 SPHINGOMONAS PAUCIMOBILIS UT26
REMARK   1  REF    ACTA CRYST. D                 V. D53  1231 1999
REMARK   1  REFN                   DK ISSN 0907-4449
REMARK   2
REMARK   2 RESOLUTION. 1.58 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : SHELXL-97
REMARK   3   AUTHORS     : G.M.SHELDRICK
REMARK   3
```

Entry from the PDB database (crystallographic info)

```
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 1.58
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 20.0
REMARK   3   DATA CUTOFF            (SIGMA(F)) : 0.000
REMARK   3   COMPLETENESS FOR RANGE        (%) : 94.2
REMARK   3   CROSS-VALIDATION METHOD           : THROUGHOUT
REMARK   3   FREE R VALUE TEST SET SELECTION   : RANDOM
REMARK
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: P 21 21 2
REMARK 290
REMARK 290      SYMOP   SYMMETRY
REMARK 290     NNNMMM   OPERATOR
REMARK 290       1555   X,Y,Z
REMARK 290       2555   -X,-Y,Z
REMARK 290       3555   1/2-X,1/2+Y,-Z
REMARK 290       4555   1/2+X,1/2-Y,-Z
REMARK 290
REMARK 290     WHERE NNN -> OPERATOR NUMBER
REMARK 290           MMM -> TRANSLATION VECTOR
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS
REMARK 290 THE FOLLOWING TRANSFORMATIONS OPERATE ON THE ATOM/HETATM
REMARK 290 RECORDS IN THIS ENTRY TO PRODUCE CRYSTALLOGRAPHICALLY
REMARK 290 RELATED MOLECULES.
REMARK 290   SMTRY1   1  1.000000  0.000000  0.000000        0.00000
REMARK 290   SMTRY2   1  0.000000  1.000000  0.000000        0.00000
REMARK 290   SMTRY3   1  0.000000  0.000000  1.000000        0.00000
REMARK 290   SMTRY1   2 -1.000000  0.000000  0.000000        0.00000
REMARK 290   SMTRY2   2  0.000000 -1.000000  0.000000        0.00000
REMARK 290   SMTRY3   2  0.000000  0.000000  1.000000        0.00000
```

Entry from the PDB database (sequence, sec. elements)

```
DBREF  1CV2 A    1   296  DBJ    BAA03443 BAA03443          1    296
SEQRES   1 A  296  MET SER LEU GLY ALA LYS PRO PHE GLY GLU LYS LYS PHE
SEQRES   2 A  296  ILE GLU ILE LYS GLY ARG ARG MET ALA TYR ILE ASP GLU
SEQRES   3 A  296  GLY THR GLY ASP PRO ILE LEU PHE GLN HIS GLY ASN PRO
SEQRES   4 A  296  THR SER SER TYR LEU TRP ARG ASN ILE MET PRO HIS CYS
SEQRES   5 A  296  ALA GLY LEU GLY ARG LEU ILE ALA CYS ASP LEU ILE GLY
SEQRES   6 A  296  MET GLY ASP SER ASP LYS LEU ASP PRO SER GLY PRO GLU
SEQRES
HELIX    1   1 SER A   42  ALA A   53
HELIX    2   2 TYR A   82  LEU A   96
HELIX    3   3 TRP A  109  ARG A  120
HELIX    4   4 GLU A  145  ARG A  155
HELIX    5   5 GLY A  159  LEU A  164
HELIX    6   6 VAL A  168  LEU A  177
HELIX    7   7 GLU A  184  GLU A  192
HELIX    8   8 ARG A  202  ILE A  211
HELIX    9   9 ALA A  218  SER A  234
HELIX   10  10 THR A  250  ARG A  258
HELIX   11  11 ILE A  274  ASP A  277
HELIX   12  12 SER A  278  LEU A  293
SHEET    1  S1 8 LYS A  12  ILE A  14  0
SHEET    2  S1 8 MET A  21  GLU A  26 -1  N  MET A  21   O  ILE A  14
SHEET    3  S1 8 ARG A  57  ASP A  62 -1  N  ALA A  60   O  ILE A  24
SHEET    4  S1 8 ASP A  30  HIS A  36  1  N  ILE A  32   O  ARG A  57
SHEET    5  S1 8 VAL A 102  HIS A 107  1  N  VAL A 103   O  PRO A  31
SHEET    6  S1 8 VAL A 125  MET A 131  1  N  ALA A 129   O  LEU A 104
SHEET    7  S1 8 LYS A 238  PRO A 245  1  N  ILE A 241   O  TYR A 130
SHEET    8  S1 8 GLN A 263  GLY A 270  1  N  THR A 264   O  LYS A 238
CISPEP   1 ASN A   38    PRO A   39          0        -2.50
CISPEP   2 ASP A   73    PRO A   74          0        -2.40
CISPEP   3 THR A  216    PRO A  217          0        -3.84
CISPEP   4 GLU A  244    PRO A  245          0         3.01
CISPEP   5 PRO A  295    ALA A  296          0        20.14
```

Entry from the PDB database (co-ordinates)

```
ATOM      1  N   GLY A   4       7.096   3.531   6.684  1.00 20.14           N
ATOM      2  CA  GLY A   4       7.885   3.530   5.461  1.00 31.48           C
ATOM      3  C   GLY A   4       9.300   4.043   5.633  1.00 20.62           C
ATOM      4  O   GLY A   4       9.599   4.865   6.501  1.00 12.02           O
ATOM      5  N   ALA A   5      10.240   3.571   4.814  1.00 15.21           N
ATOM      6  CA  ALA A   5      11.609   4.057   4.935  1.00 10.23           C
ATOM      7  C   ALA A   5      11.883   5.182   3.955  1.00 11.96           C
ATOM      8  O   ALA A   5      12.950   5.809   3.978  1.00 13.69           O
ATOM      9  CB  ALA A   5      12.621   2.943   4.674  1.00 10.47           C
ATOM     10  N   LYS A   6      10.929   5.437   3.056  1.00 12.31           N
ATOM     11  CA  LYS A   6      11.251   6.452   2.053  1.00 17.87           C
ATOM     12  C   LYS A   6      11.223   7.850   2.660  1.00  9.09           C
ATOM     13  O   LYS A   6      10.310   8.161   3.422  1.00 10.53           O
ATOM     14  CB  LYS A   6      10.274   6.419   0.870  1.00 20.08           C
ATOM     15  CG  LYS A   6      10.901   6.898  -0.436  1.00 47.79           C
ATOM     16  CD  LYS A   6      10.695   8.377  -0.703  1.00 63.02           C
ATOM     17  CE  LYS A   6      11.654   8.950  -1.734  1.00 62.33           C
ATOM     18  NZ  LYS A   6      11.574  10.435  -1.832  1.00 50.12           N
ATOM     19  N   PRO A   7      12.171   8.696   2.307  1.00 12.48           N
ATOM     20  CA  PRO A   7      12.108  10.087   2.748  1.00 15.50           C
ATOM     21  C   PRO A   7      10.895  10.808   2.144  1.00 16.27           C
ATOM     22  O   PRO A   7      10.244  10.396   1.170  1.00 15.29           O
ATOM     23  CB  PRO A   7      13.394  10.717   2.217  1.00 12.40           C
ATOM     24  CG  PRO A   7      13.877   9.803   1.151  1.00 23.02           C
ATOM     25  CD  PRO A   7      13.347   8.427   1.456  1.00 21.52           C
ATOM     26  N   PHE A   8      10.608  11.942   2.751  1.00 10.86           N
ATOM     27  CA  PHE A   8       9.557  12.848   2.302  1.00  6.69           C
ATOM     28  C   PHE A   8      10.134  13.914   1.384  1.00 17.38           C
ATOM     29  O   PHE A   8      11.121  14.590   1.716  1.00 17.05           O
ATOM     30  CB  PHE A   8       8.912  13.490   3.531  1.00  6.78           C
ATOM     31  CG  PHE A   8       7.776  14.444   3.183  1.00 12.53           C
ATOM     32  CD1 PHE A   8       6.526  13.921   2.874  1.00 18.59           C
ATOM     33  CD2 PHE A   8       7.984  15.811   3.166  1.00 15.33           C
ATOM     34  CE1 PHE A   8       5.494  14.797   2.547  1.00 19.62           C
ATOM     35  CE2 PHE A   8       6.961  16.701   2.851  1.00 15.73           C
ATOM     36  CZ  PHE A   8       5.718  16.160   2.537  1.00 22.66           C
ATOM     37  N   GLY A   9       9.544  14.110   0.215  1.00 18.38           N
```

Protein structure databases

- PDBsum
  - developed at University College London
  - summaries and analyses of protein structures (secondary database derived from PDB)
  - summary of PDB entries: resolution, R-factor, # protein chains, topology, ligands, metal ions, etc.
  - analysis of PDB entries: protein-metal and protein-ligand interactions, protein validation
  - provides links to many related databases

RasMol

VRML v.1.0

Structure viewers HELP!

PDB id: 1cv2

Hydrolase

Title: *Hydrolytic haloalkane dehalogenase linb from sphingomonas paucimobilis ut26 at 1.6 a resolution*

Structure: *Haloalkane dehalogenase. Chain: a. Synonym: linb, 1,3,4,6–tetrachloro–1,4–cyclohexadiene hydrolase. Engineered: yes*
Source: *Sphingomonas paucimobilis. Strain: ut26. Expressed in: escherichia coli.*

Resolution: 1.58Å. **R–factor:** 0.152. **R–free:** 0.211.

Authors: J.Marek, J.Vevodova, J.Damborsky, I.Smatanova, L.A.Svensson, J.Newman, Y.Nagata, M.Takagi

Date: 22–Aug–99

PDB header records

Enzyme class from PDB file: E.C.3.8.1.5 E.C.->PDB

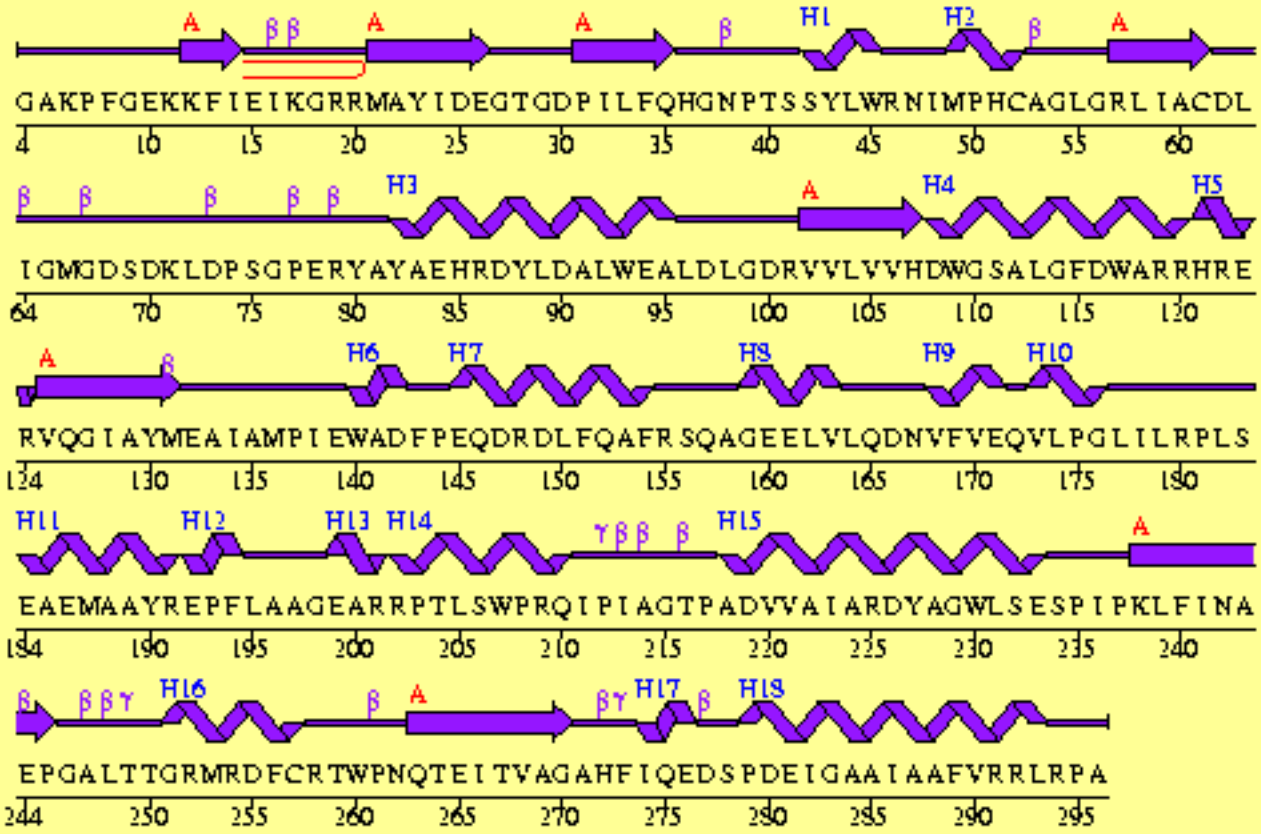
Links:

PDB
OCA
MMDB
IMB Jena
STING
GRASS
PQS
CATH
scop
FSSP
PROCHECK
WHATCHECK
PROMOTIF

Chain A (**293** residues)

- **CATH** structural classification (1 domain) :

Links	CATH no.		Class	Architecture
CATH DHS	**3.40.50.950**	–>	*Alpha Beta*	*3–Layer(aba) Sandwich*



Postscript version

View chain **A** alone.

- **PROMOTIF** summary :
  1 sheet, 8 strands, 18 helices, 19 beta turns, 3 gamma turns, 1 beta bulge, 1 beta hairpin, 4 beta alpha beta units, 1 psi–loop.

Protein structure classification databases

- Classification attempts to capture the structural similarities among proteins.
- The structural similarities relate to the evolution.
- The structural similarities may imply function.
- The classification scheme is dependent on the underlying philosophy.

Protein structure classification databases

- SCOP - Structural Classification of Proteins
  - developed at MRC Laboratory of Molecular Biology
  - construction: combination of manual and automatic methods (complicated by multidomain proteins)
  - fold = same secondary elements in same arrangement, independently of common evolutionary origin
  - superfamily = low identity but common evolutionary origin implied from common structure and function
  - family = sequence identity >30%

Protein: Haloalkane dehalogenase from *Sphingomonas paucimobilis*, UT26, LinB

Lineage:

1. Root: scop
2. Class: Alpha and beta proteins (a/b)
   Mainly parallel beta sheets (beta–alpha–beta units)
3. Fold: alpha/beta–Hydrolases
   core: 3 layers, a/b/a; mixed beta–sheet of 8 strands, order 12435678, strand 2 is antiparallel to the rest
4. Superfamily: alpha/beta–Hydrolases
   many members have left–handed crossover connection between strand 8 and additional strand 9
5. Family: Haloalkane dehalogenase
6. Protein: Haloalkane dehalogenase
7. Species: *Sphingomonas paucimobilis*, UT26, LinB

PDB Entry Domains:

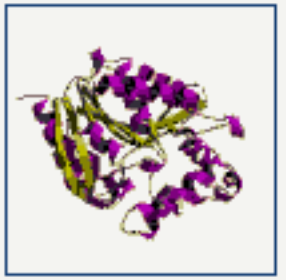
1. 1cv2
   1. chain a
2. 1d07
   complexed with br, gol
   1. chain a

Protein structure classification databases

- CATCH - Class, Architecture, Topology, Homology
  - developed at University College London
  - construction: mostly automatic
  - unique numbering scheme analogous to Enzyme Classification (E.C.) scheme
  - class = gross secondary structure content
  - architecture = gross secondary structure arrangement
  - topology = shape and connectivity of secondary structures (60% of larger protein matches smaller one)
  - homology = sequence identity >35%, common ancestry
  - sequence = clustering based on sequence identity

View as XML

Domain 1cv2A0



1cv2A0

View Rasmol

- Alpha Beta
- 3–Layer(aba) Sandwich
- Rossmann fold
- Alpha/beta hydrolase
- HYDROLASE
- HYDROLASE
- HYDROLASE

Fold relatives

There are either no other non–identical relatives within this fold group or the structural comparisons for this domain have not yet been calculated.

Search

Go!

PDB code
CATH code
General text

Goto...

SSAP
DHS
Gene3D
PDBsum

Navigation

Top of heirarchy
Up one level

Help

Select a topic

Prediction of secondary structure

- Algorithms assign probability for occurrence of α-helix, β-strand, turn and random coil at particular position in the sequence.

- Methods: statistical, stereochemical and homology/neural networks based. All methods rely on information derived from known 3D structures. Most recent methods use the information from multiple alignments.

- Reliability of the best current methods is >70%.

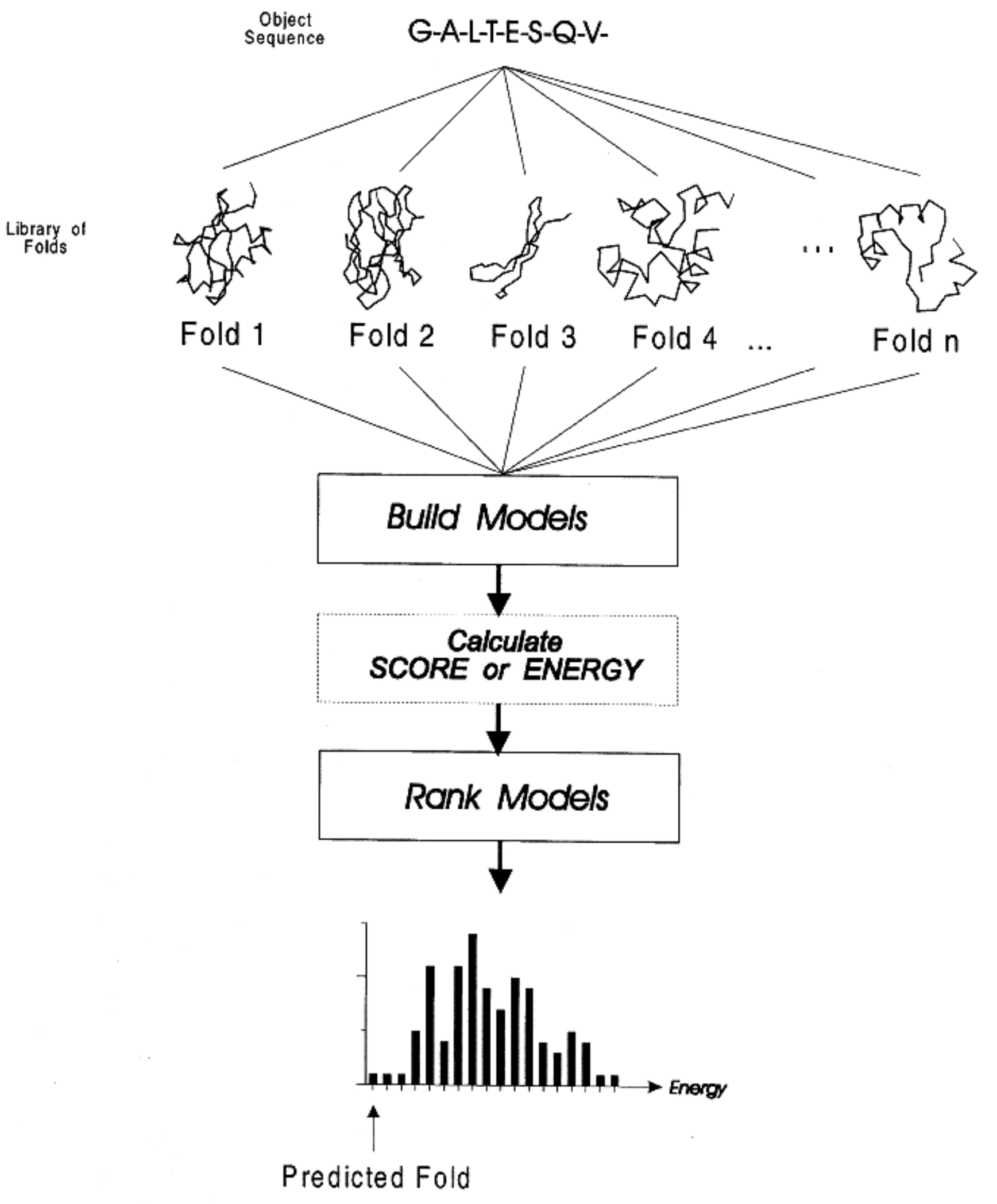
Prediction of secondary structure

- Chou-Fasman and GOR
  - statistical - amino acids show preference for particular secondary structure elements
- PHD and NNPredict
  - neural networks - the rules for prediction are not defined in advance, they are created by training
- NNSSP and PREDATOR
  - nearest neighbour approach
- JPRED
  - consensus approach - utilises multiple alignments and state-of-art method - makes consensus

1
10
20
30
40
50
60
Flavodoxin
A K I G L F Y G T Q T G V T Q T I A E S I Q Q E F G G E S I V D L N D I A N A D A S D L N A Y D Y L I I G C P T W N V G
nnpredict
PredictProtein
SSPRED
GOR
Levin
DPM
SOPMA
CNRS Consensus
1OFV
Beta 1
Alpha 1
Beta 2
Alpha 2
Beta 3
70
80
90
100
110
120
Flavodoxin
E L Q S D W E G I Y D D L D S V N F Q G K K V A Y F G A G D Q V G Y S D N F Q D A M G I L E E K I S S L G S Q T V G Y W
nnpredict
PredictProtein
SSPRED
GOR
Levin
DPM
SOPMA
CNRS Consensus
1OFV
Alpha 3
Alpha 4
Beta 4
Alpha 5
130
140
150
160
Flavodoxin
P I E G Y D F N E S K A V R N N Q F V G L A I D E D N Q P D L T K N R I K T W V S Q L K S E F G L
nnpredict
PredictProtein
SSPRED
GOR
Levin
DPM
SOPMA
CNRS Consensus
1OFV
Beta 5
Alpha 6

Prediction of protein fold

- Threading
  - treading = protein fold assignment or fold recognition
  - target sequence is searched against database of folds (3D profiles) and threaded models are constructed
  - 3D profile - each residue in 3D structure is assigned environmental variables (buried area, fraction of side chain covered polar atoms, secondary structure, etc.). Assumption - environment of the residue should be more conserved that the residue itself.
  - residue can be also described by its interactions
  - match of target sequence with 3D profile (quality of threaded models) is quantified by Z-score or energy
  - limitation: can not handle multi-domain proteins


Object Sequence
G-A-L-T-E-S-Q-V-
Library of Folds
Fold 1
Fold 2
Fold 3
Fold 4
...
Fold n
Build Models
Calculate SCORE or ENERGY
Rank Models
Energy
Predicted Fold

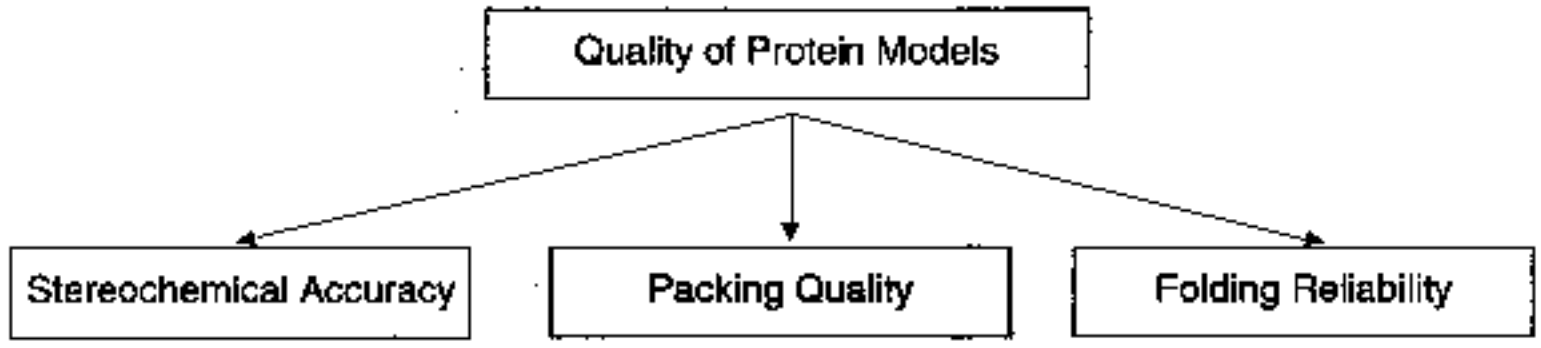
Prediction of protein fold

- Bioinbgu
  - consensus method utilising predictions from five different algorithms
- 3D-PSSM
  - scoring functions: 1D-PSSMs (sequence profiles built from relatively close homologues), 3D-PSSMs (more general profiles containing more remote homologues), matching of secondary structure elements, and propensities of the residues for solvent accessibility
- GenThreader
  - hybrid method: profile-based alignment, evaluation of alignments by threading, evaluation of threaded models by neural network

Prediction of tertiary structure

- *Ab initio*
  - 3D structure of a protein is predicted from first principles (search for global minimum structure)
  - current algorithms are not very reliable
- Homology modelling
  - 1. alignment of modelled sequence against sequences of structurally similar proteins (templates)
  - 2. "extraction" of the backbone from template structure and positioning of side-chain
  - 3. modelling of loops
  - 4. structure refinement and validation

Validation of protein models



- Torsion angles
  - → Mainchain torsion angle distribution, (Ramachandran plots)
  - → Sidechain torsion angle distribution (χ_1 - χ_2 plots)
- Planarity of peptide bonds
  - → ω angle distribution
- Chirality of C^{α}-atoms
  - → ζ angle distribution
- Bond lengths
- Bond angles
- Planarity
  - → Aromatic ring systems and sp^2-hybridized end groups

- Interatomic distances
  - → 'Bump check'
  - → 'Atomic contact quality'
- Secondary structural elements
  - › Location and geometry of secondary structural elements
- Hydrophobicity
  - → Distribution of polar and nonpolar amino acids
- Solvent accessible surface of amino acids
- Unsatisfied buried H-bond donors/acceptors

- 3D-comparison model/ template structure
  - → RMS deviations between backbone atoms
- 3D-1D-profiles
  - → Comparison of environment strings with amino acid sequences
- Knowledge-based potentials
  - → Energy-based comparison

Prediction of tertiary structure

- SWISS-MODEL
  - fully automated modelling server
  - input = protein sequence; output = PDB file
  - 1. search of ExNRL-3D using BLASTP for potential templates; 2. select all templates with sequence identities above 25%; 3. Generate structures of 3D models; 4. energy minimise models using GROMOS 96
  - first approach and optimise mode (Swiss-PDBViewer)

- MODELLER
  - most widely used academic program for homology modelling (satisfaction of spatial restrains)

Modelling of protein ligand-complexes

- Docking
  - positioning of small organic molecules (ligands) inside the protein active site
  - different orientations and conformations of the ligand are evaluated using geometric or energetic scoring functions
  - Protein-ligand interaction energy = van der Waals term + electrostatic term + H-bond term + entropic term
  - flexible docking - considers different conformation of ligand; different rotamers of protein side chains
- Software: DOCK, AUTODOCK, FLEX