Secondary and tertiary structure of proteins

Josef Houser

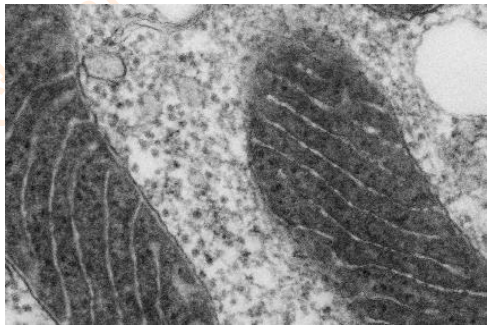
Autumn 2023

S1004 Methods for structural characterization of biomolecules

Structure

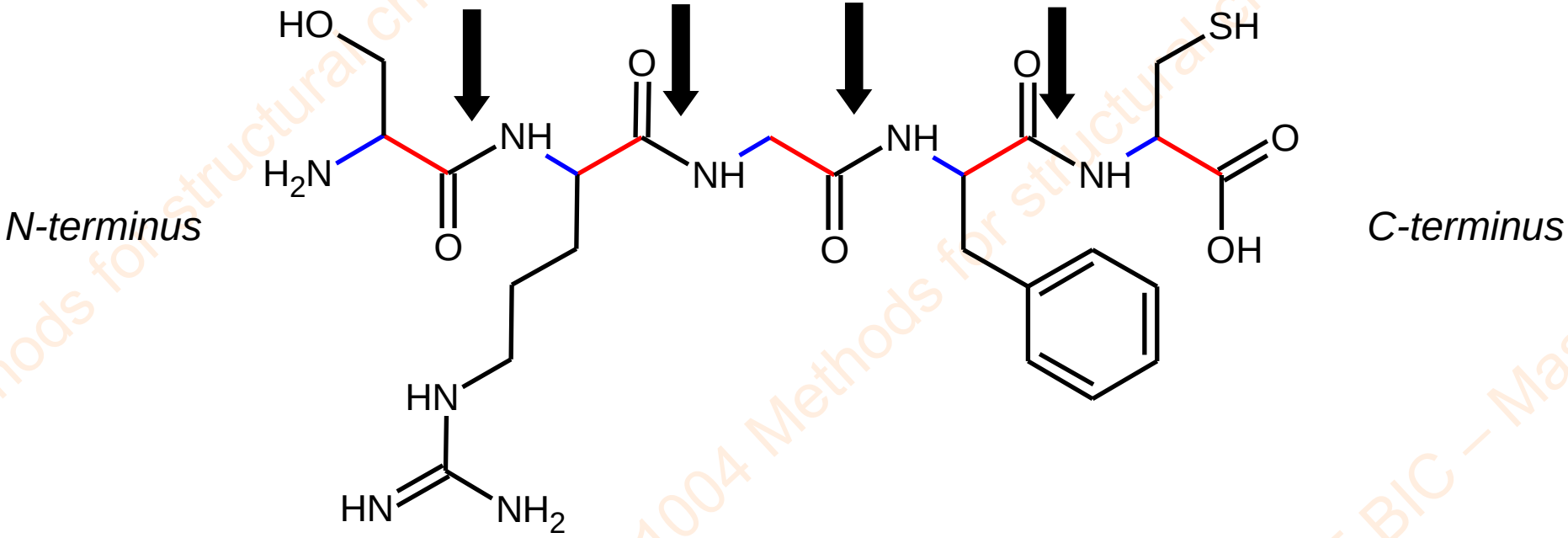
- “Relative” **position of individual building blocks** within the sample

- Block definition reflects structural details:
  - Whole molecules
  - Domains
  - Secondary structures
  - Residues
  - Atoms
  - Orbitals
  - Elementary particles



Protein

Linear polymer of amino acids linked by peptide bond

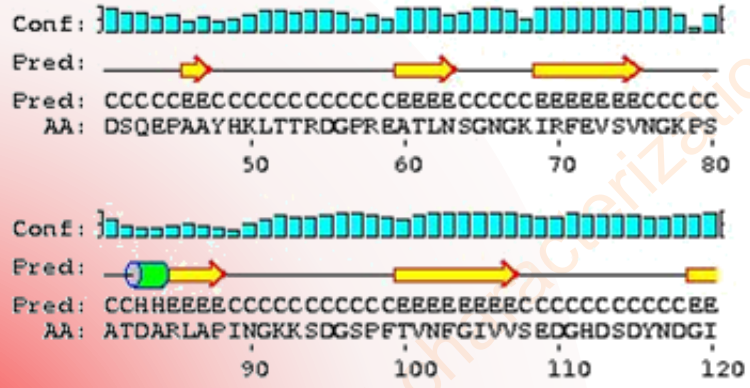


Protein structure

1D

ADSQTSSNRAGEFSIPPNTDFRAIFFANAAE
QQHIKLFIGDSQEPAAYHKLTTRDGPREATL
NSGNGKIRFEVSVNGKPSATDARLAPINGK
KSDGSPFTVNFGIVVSEDGHDSDYNDGIVV
LQWPIG

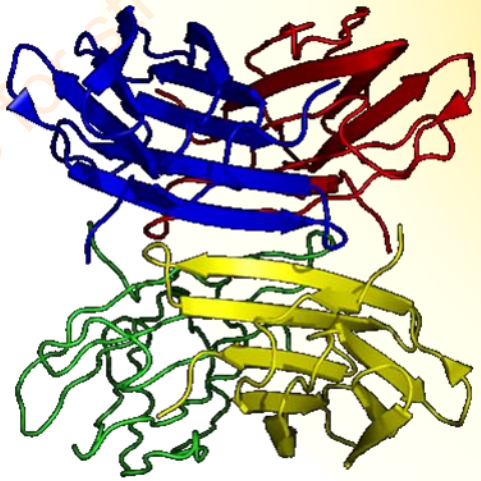
primary (sequence)



2D

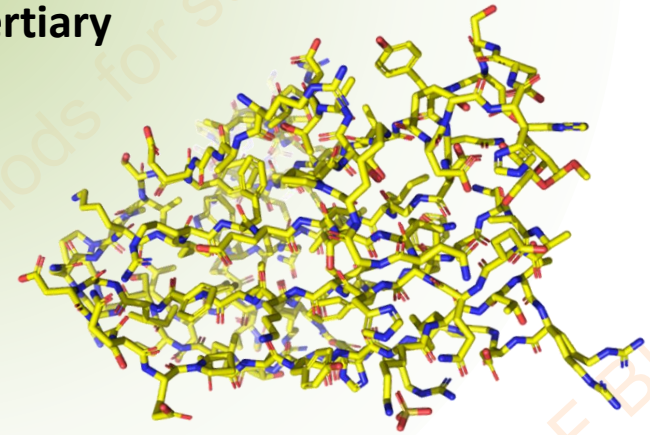
secondary

4D



quaternary

tertiary

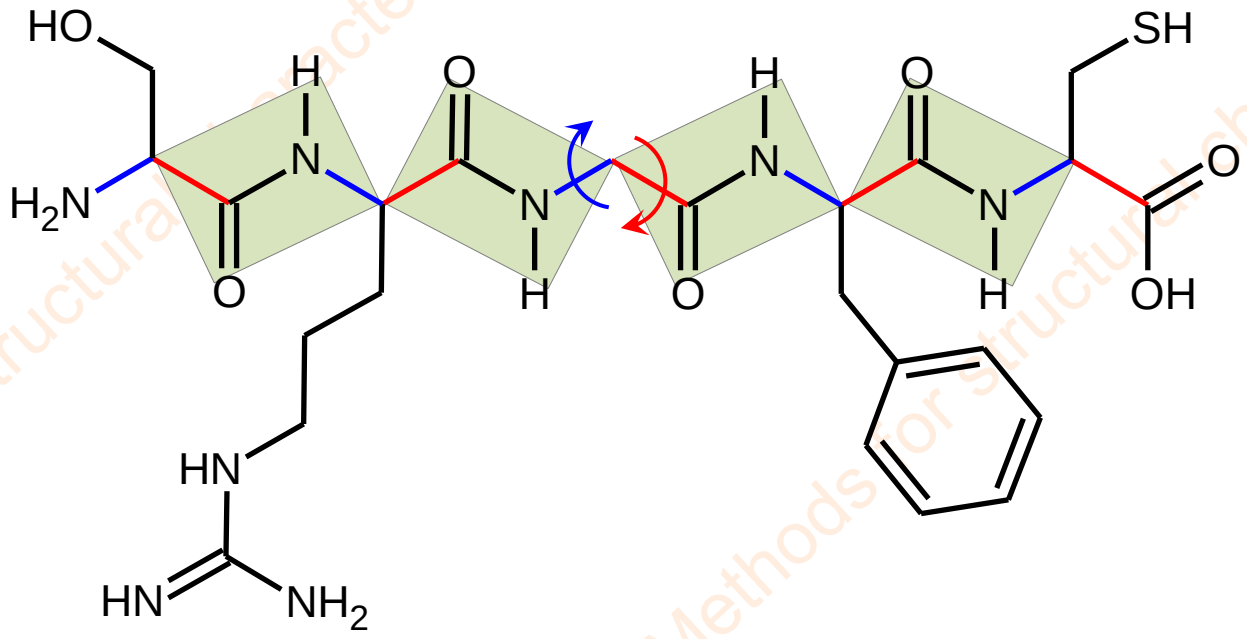


3D

Protein secondary structure

2D

Peptide bond – planar (angle ω is ± 180°)

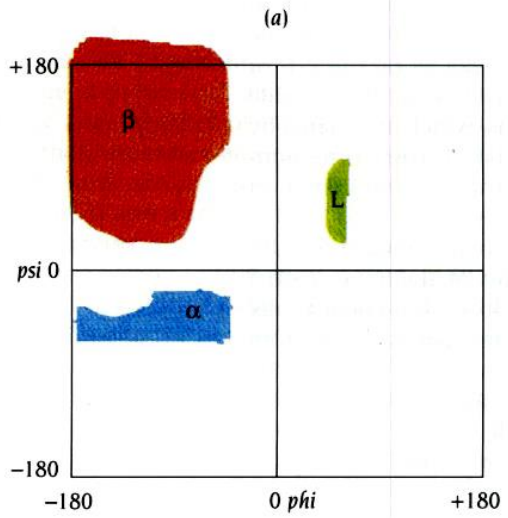


Backbone conformation defined by two torsion angles ϕ and ψ

Ramachandran diagram

2D

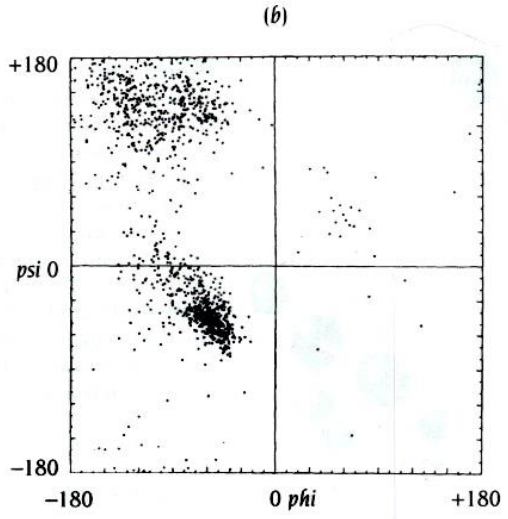
- Combination of ϕ and ψ angles for individual amino acids in protein
- Populated in several areas (combination of angles)
- Main areas labeled as α and β

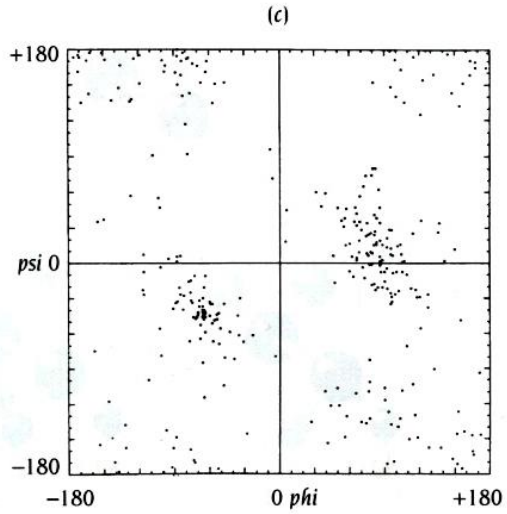


a) Colored areas show sterically allowed combinations of the ϕ and ψ angles.

b) Observed values for all residue types except for glycine. Each point represents ϕ and ψ values for an amino acid residue in a well-refined x-ray structure.

c) Observed values for glycine

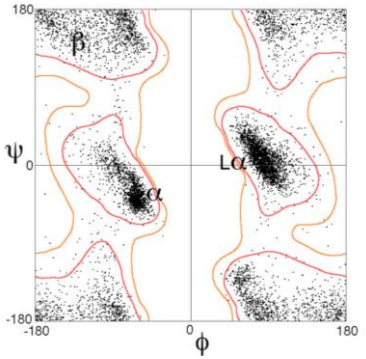




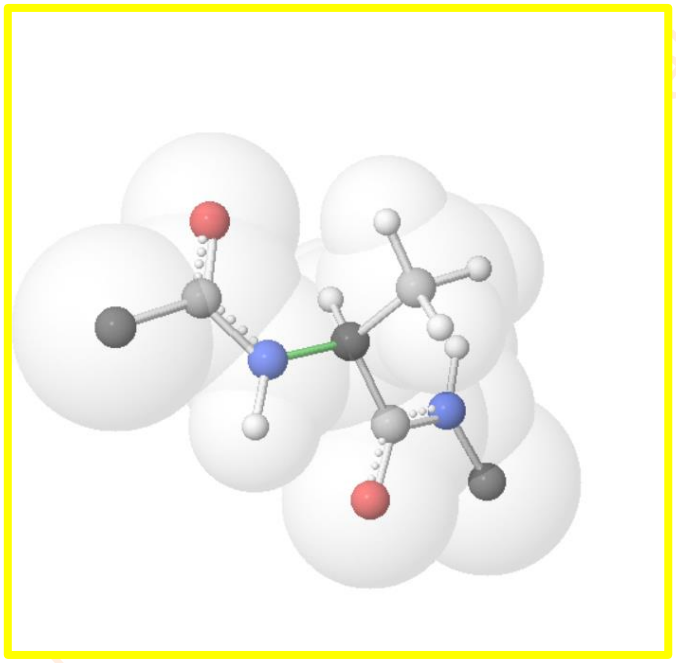
Sterical hindrance

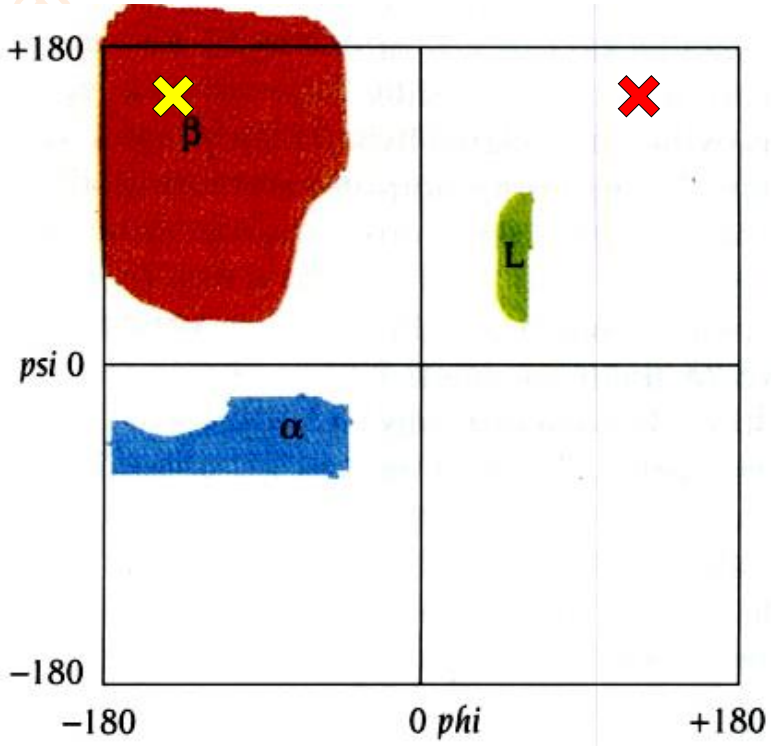
2D

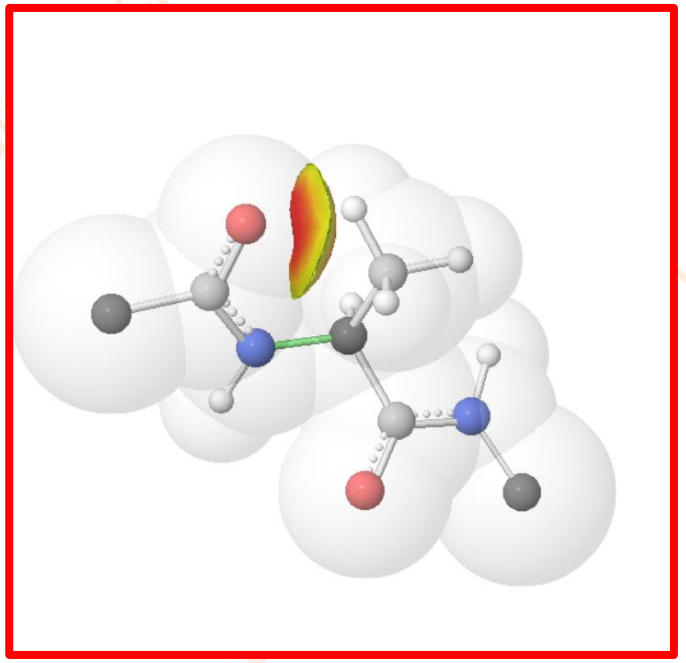
- Non-represented combinations suffer from **sterical clashes**
- Much smaller problem for **glycine** – Cα atom missing



wikipedia.org







http://bioinformatics.org/molvis/phipsi/

Bonds involved in structure stbilization

2D

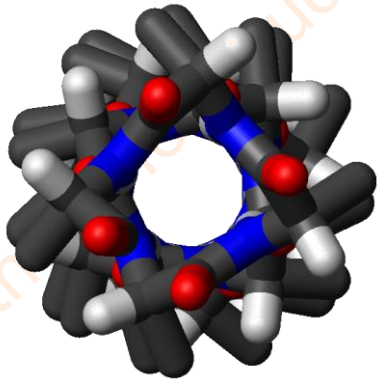
- **Hydrogen bond** (H-bridge)

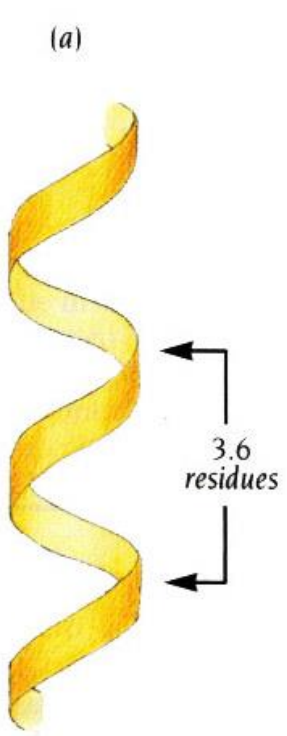
- Charge-charge
- Polar AA contacts
- Non-polar / hydrophobic AAs
- Stacking – aromatic AAs
- Cysteine / cystine – sulfur-sulfur bond
- Metal ions

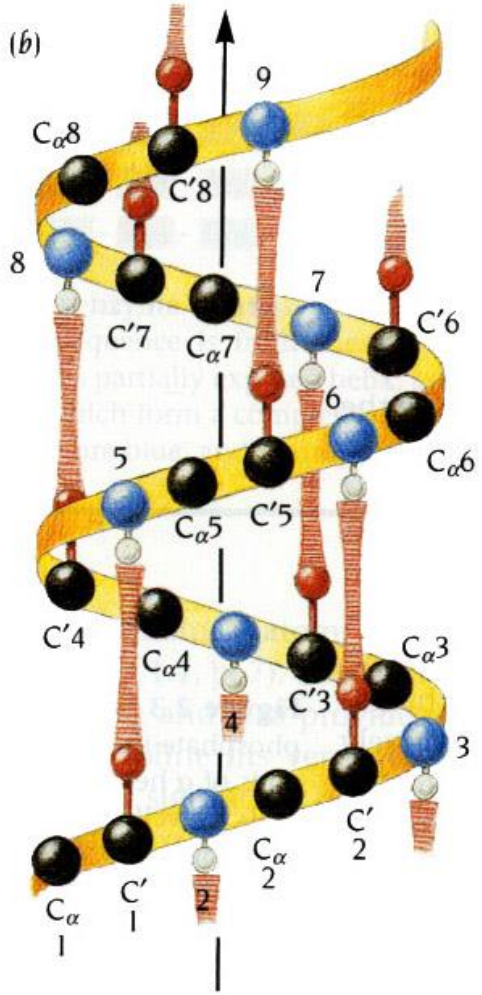
α-helix

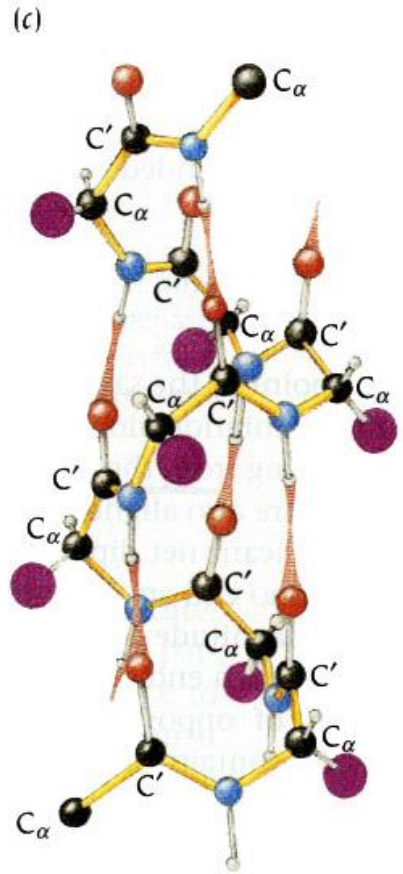
2D

- Most frequent
- Stabilized by **intra-main chain** hydrogen bonds



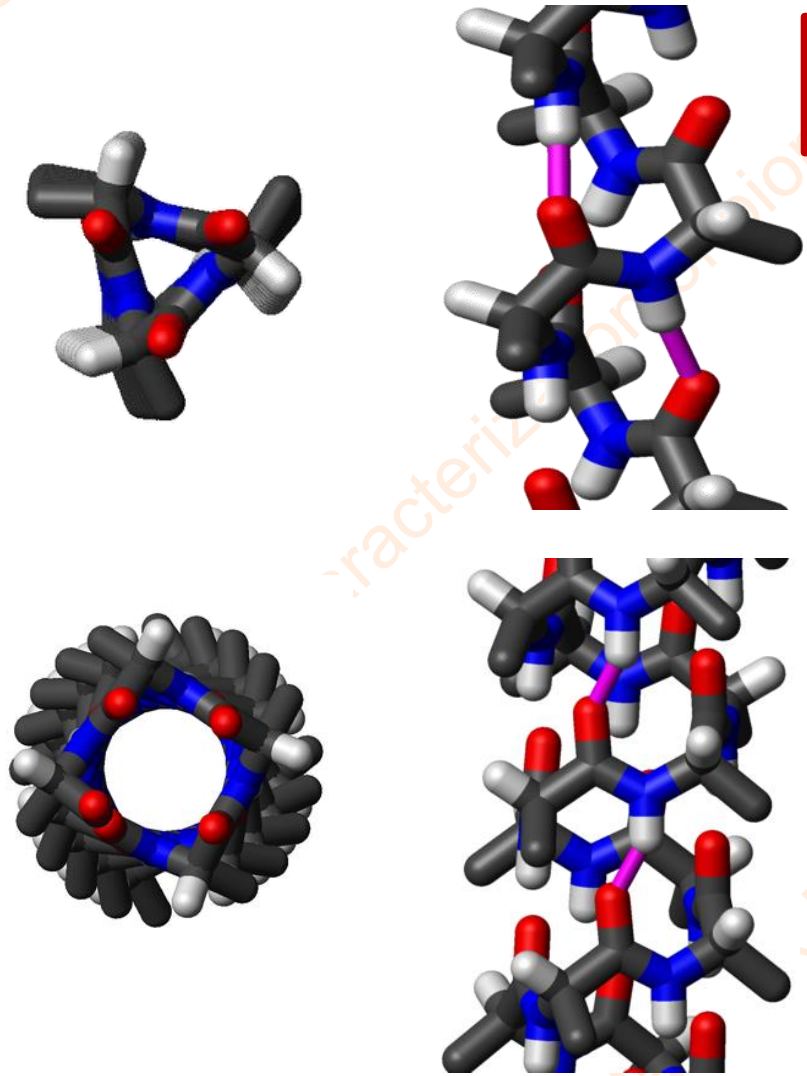


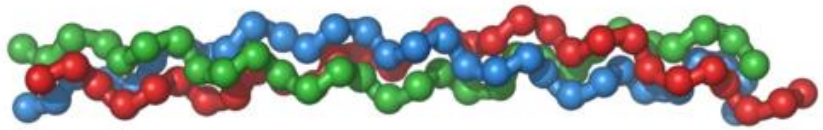




Other helix structures

- **3_{10} helix**
  - “more tight”
  - ends of α helix, turns
- **π-helix**
  - “more loose”
  - ends of helixes, very rare
- **left-hand helix**
  - sequence dependent – proline/glycine rich
  - colagen





β-sheet (β-strand)

2D

- Second main 2D structure type
- Stabilized by **inter-main chain** hydrogen bonds
- Two types based on mutual orientation of neighboring chains
- **Antiparallel** more stable than **parallel**

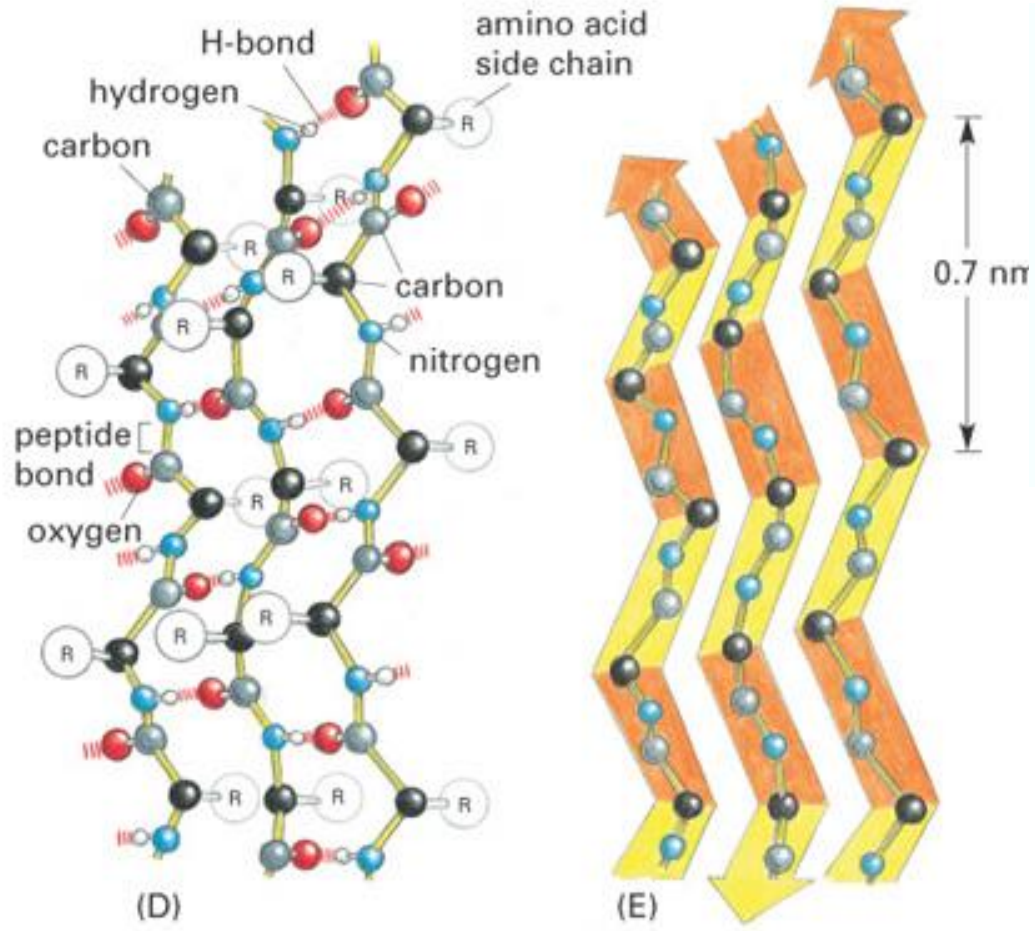
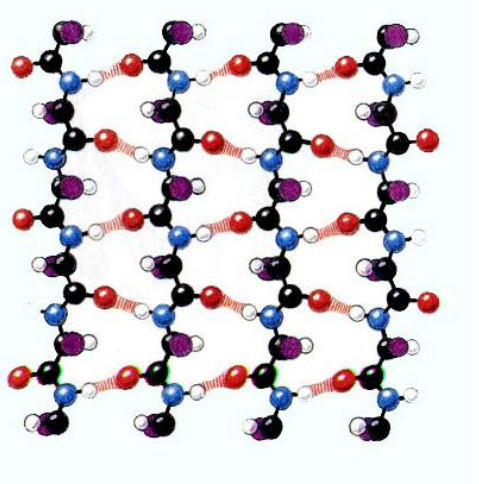
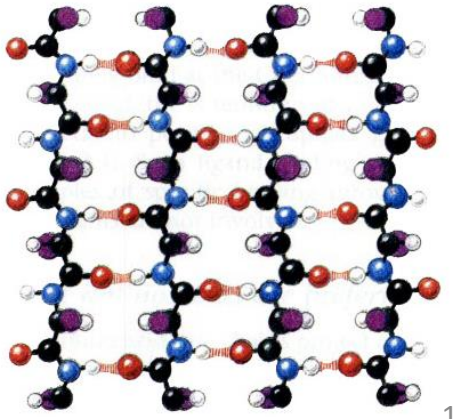


Figure 4-10 part 2 of 2 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Parallel

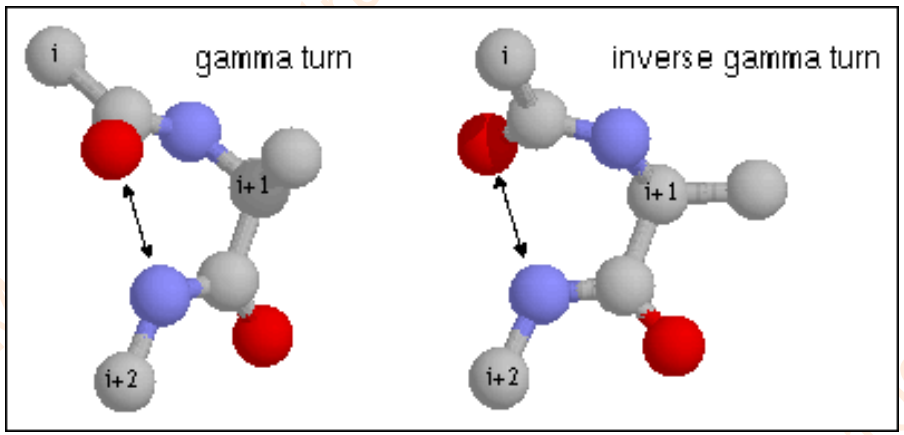


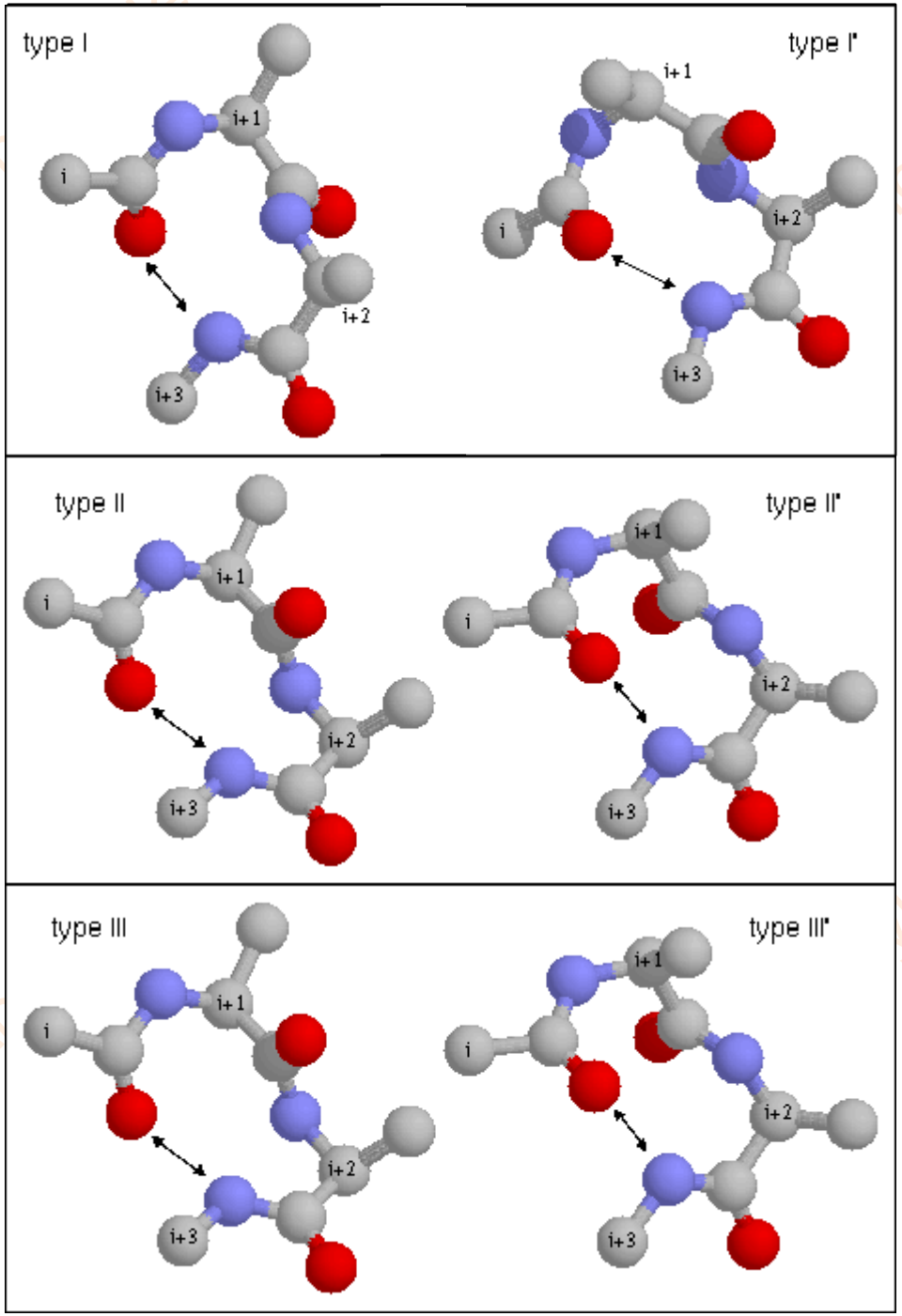
Antiparallel



Turns

- Several types
- Various AAs number – 3 - 5
- Examples: β-turn, γ-turn





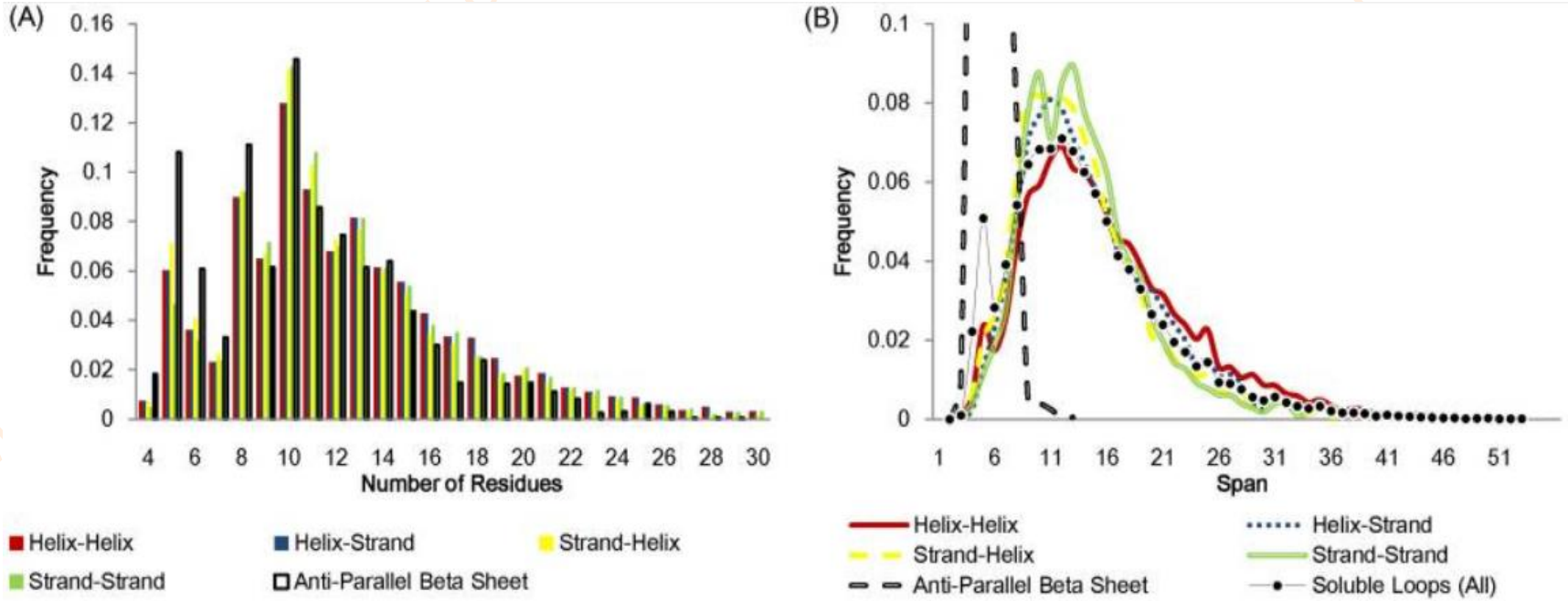
2D

www.cryst.bbk.ac.uk

Loops

2D

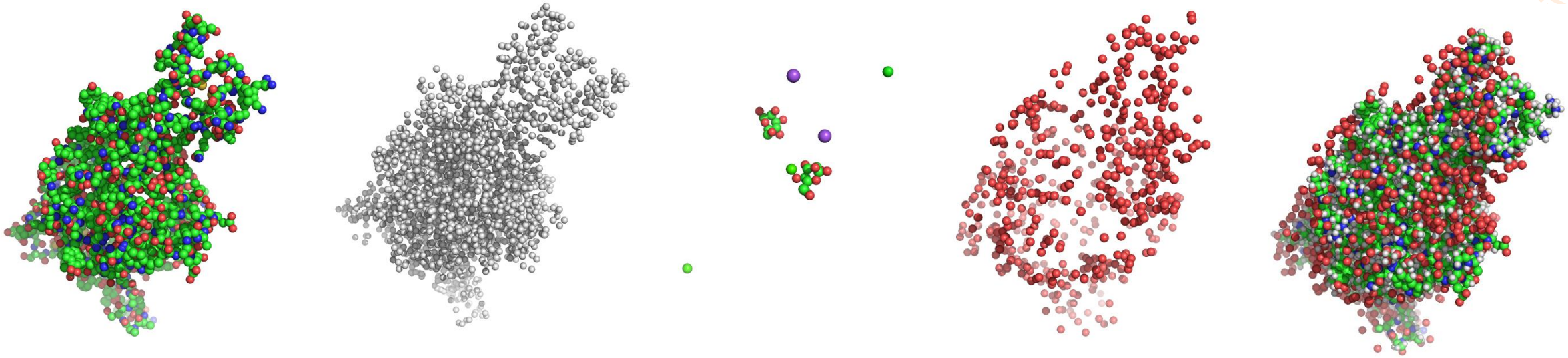
- Connecting two elements of 2D structure
- **Partially organized** structures – between turns and random coil
- Typically 5-16 AAs with dominantly polar residues



Y. Choi *et al* (2013) Peerj.

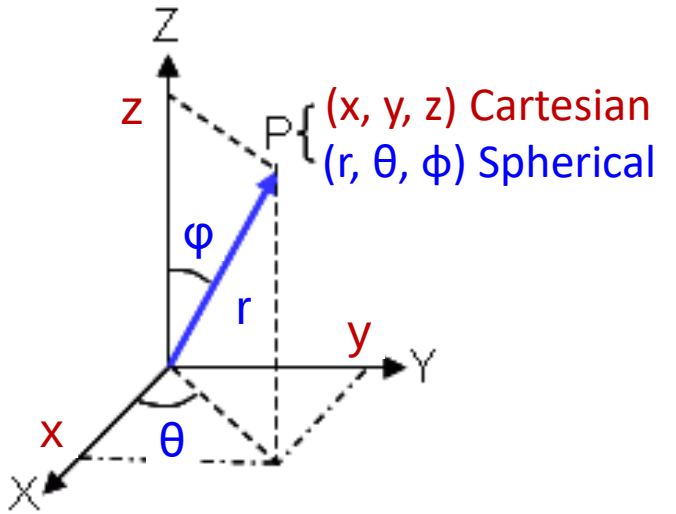
Protein terciary structure

- Location of individual atoms in space
  - **“heavy atoms”** – C, N, O, S, (P)
  - hydrogens
  - bound molecules
  - hydration shell



Absolute coordinates

- Related to the defined origin of coordinate system [0, 0, 0]
- **Cartesian coordinates** – x, y, z
- Spherical coordinates – r, θ, ϕ or ρ, θ, ϕ



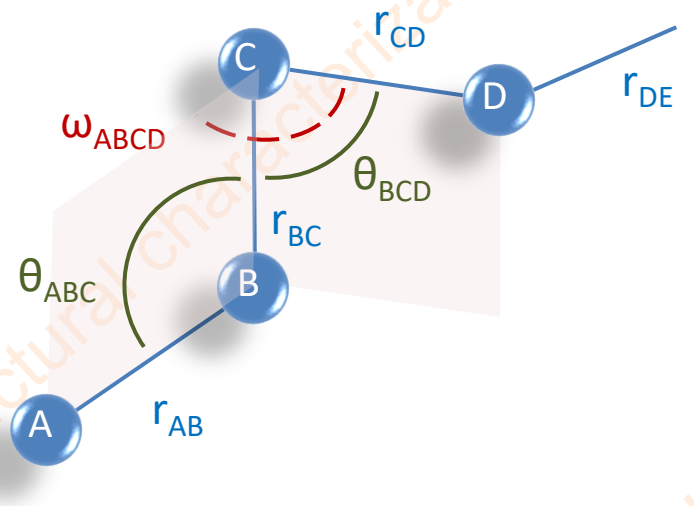
- N atoms → 3N coordinates

Relative coordinates

- Related to previous defined point (atom)
- **Distance** to previous atom
- **Angle** between three atoms
- **Torsion angle** between four atoms

- For N atoms → 3N – 6 coordinates

etc.



A	-	-	-
B	r_{AB}	-	-
C	r_{BC}	θ_{ABC}	-
D	r_{CD}	θ_{BCD}	ω_{ABCD}
E	r_{DE}	θ_{CDE}	ω_{BCDE}
...			

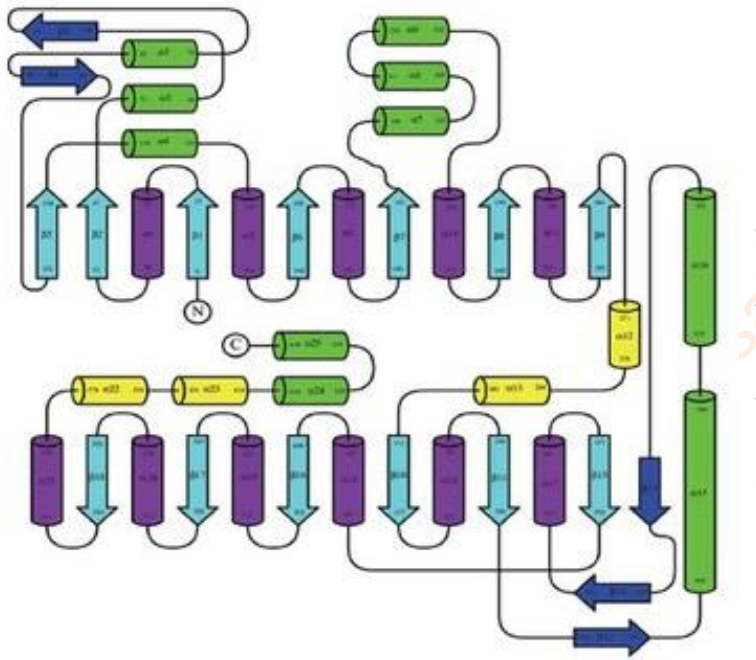
From 2D to 3D structure

- Complicated **hierarchy**:

  Secondary – (Supersecondary) – Terciary

- **Topology** (in structural biology) – mutual orientation of 2D structure elements

- Motives – Folds – Domains



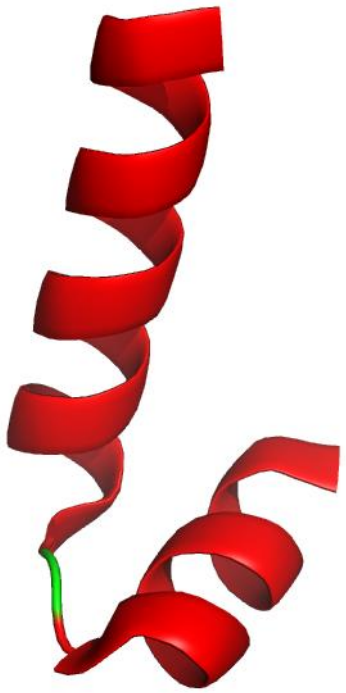
S. Baskaran (2010)

From 2D to 3D structure

- **Motifs**
  - 2-3 elements of secondary structure combined
- **Folds**
  - Combination of simple motifs
- **Domains**
  - Consist of motifs/folds

Simple motifs

Helix-turn-helix

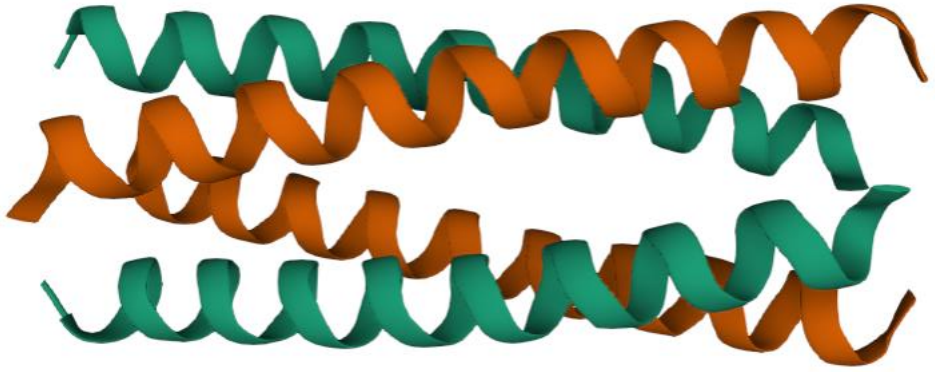


β-hairpin

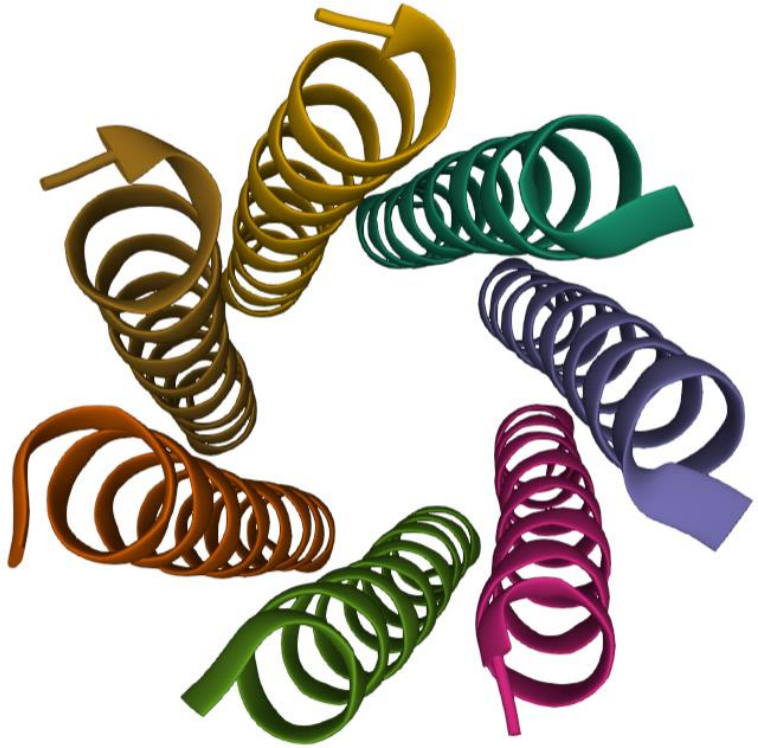
β-α-β

Complex α-motifs/folds

4-helix bundle

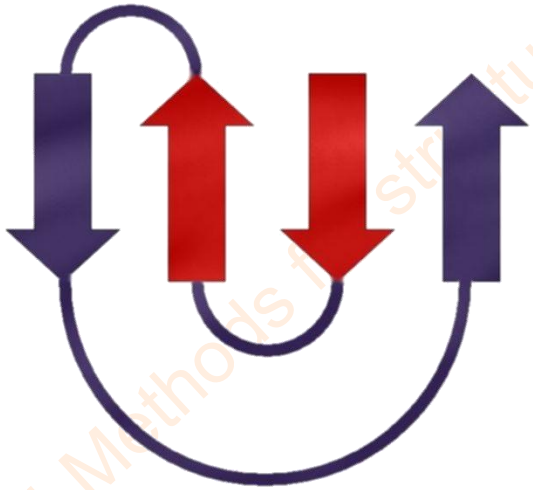


7-helix barrel



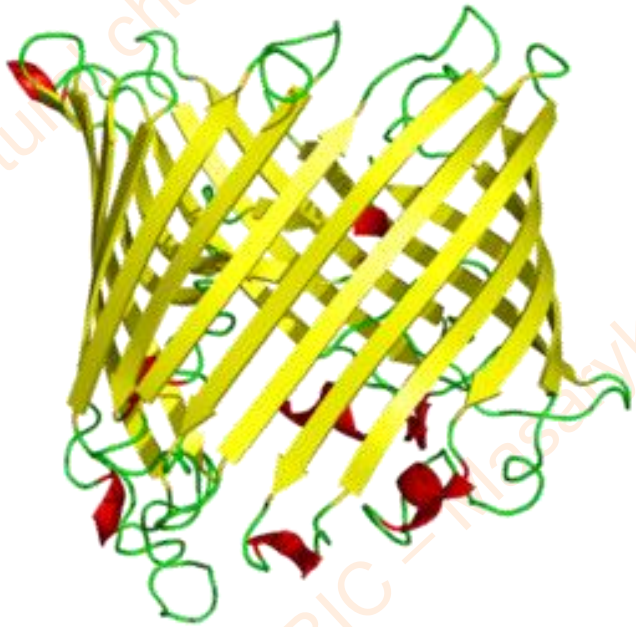
Complex β-motifs/folds

Greek key



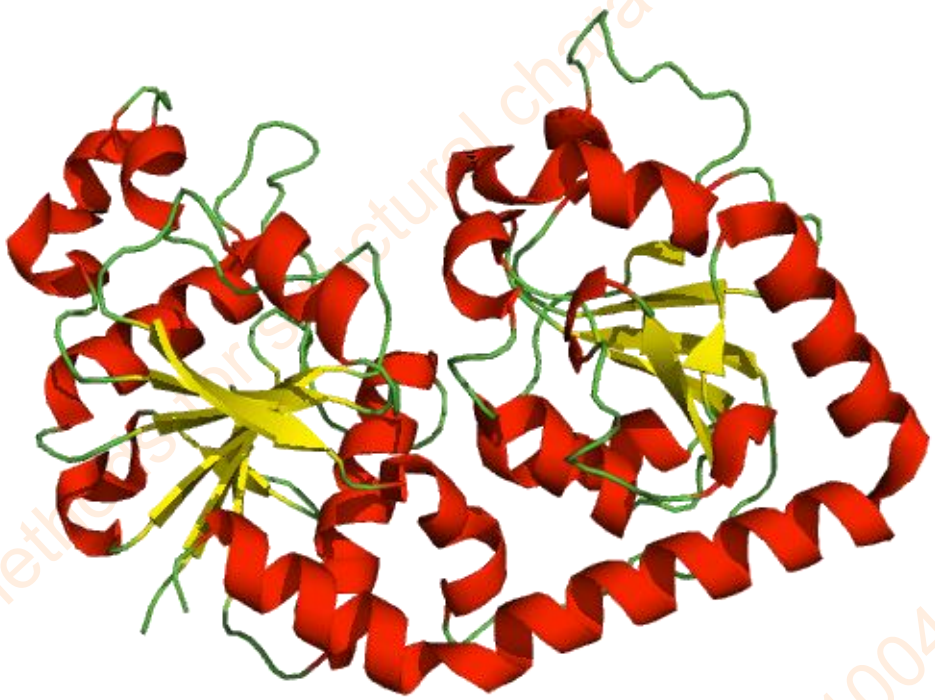
β-meander

β-barrel

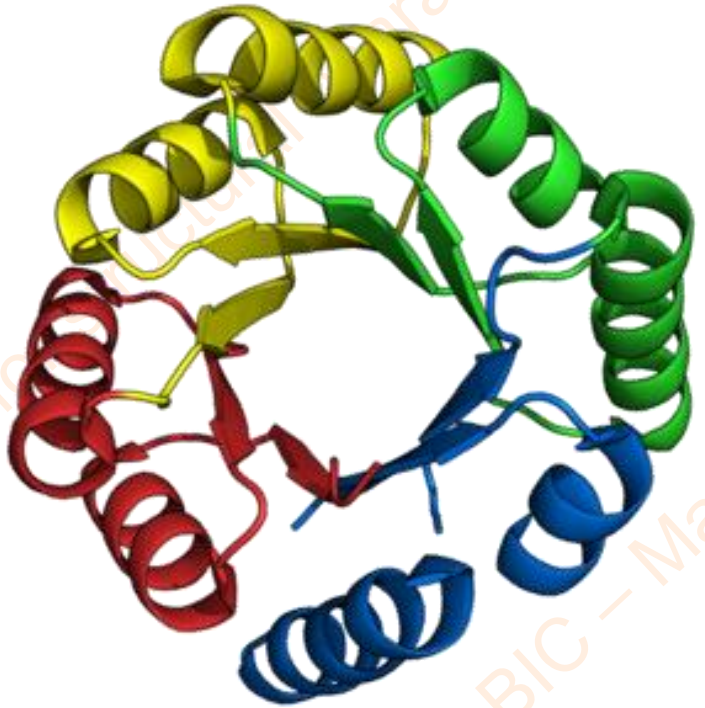


Complex α/β-motifs/folds

Rossman fold

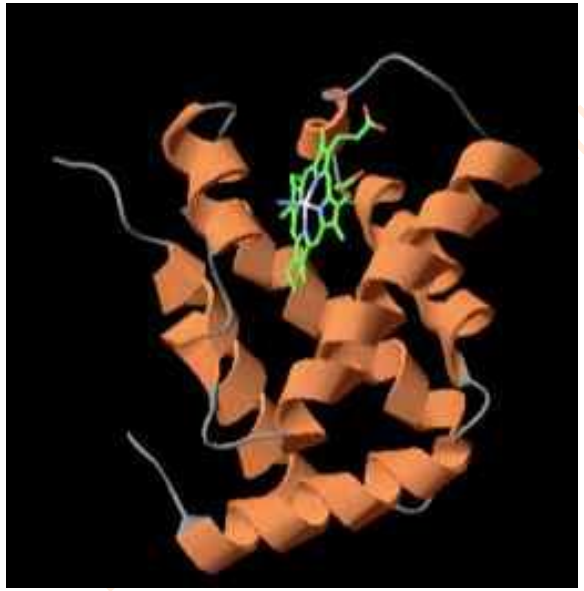


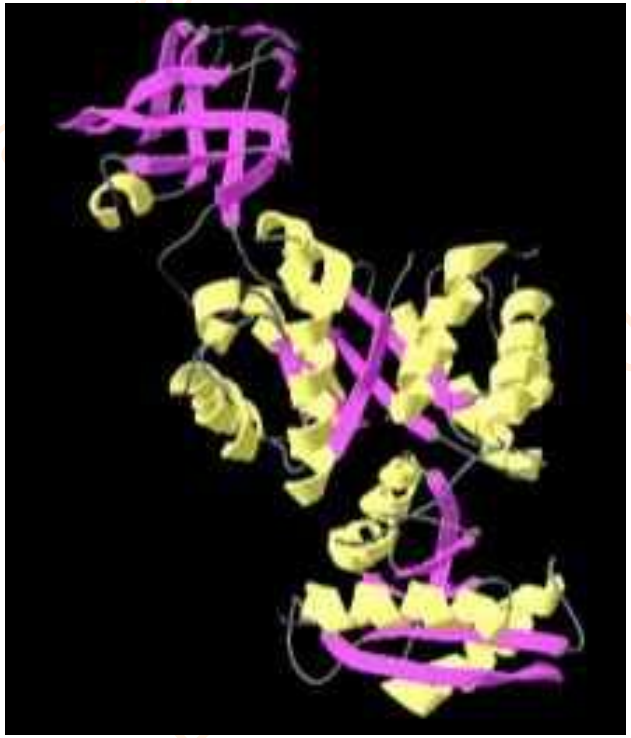
TIM-barrel



Protein domains

- Part of the structure with **defined function** (smallest functional unit)
- Independent unit (at least partially)
- **Single-domain** vs. **multi-domain** proteins





3D structure databases

- **wwPDB** (http://www.wwpdb.org)

  RCSB PDB – Research Collaboratory for Structural Bioinformatics Protein Data Bank

  PDBe – Protein Data Bank Europe

  PDBj – Protein Data Bank Japan

  BMRB – Biological Magnetic Resonance Data Bank

- **SCOP** (http://scop.mrc-lmb.cam.ac.uk/scop/)

  structural classification of proteins

- **CATH** (http://www.cathdb.info/)

  classification of protein-domains

- **EMDataBank** (http://www.emdatabank.org/)

  electron microscopy structures

Formats for 3D structure files

PDB (Protein Data Bank)

- PDB File Format (http://www.wwpdb.org/documentation/file-format)
- mmCIF File Format and PDB Exchange Dictionary
- PDBML - XML File Format

PDB format

- Created 1976
- Fixed column position and width, capacity limitation
- Still very frequent but **outdated**

```
HEADER    HYDROLASE                               14-MAR-03   1HL8
TITLE     CRYSTAL STRUCTURE OF THERMOTOGA MARITIMA ALPHA-FUCOSIDASE
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: PUTATIVE ALPHA-L-FUCOSIDASE;
COMPND   3 CHAIN: A, B;
COMPND   4 EC: 3.2.1.51;
COMPND   5 ENGINEERED: YES;
COMPND   6 OTHER_DETAILS: ORF TM0306
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: THERMOTOGA MARITIMA;
SOURCE   3 ORGANISM_TAXID: 243274;
SOURCE   4 STRAIN: MSB8;
SOURCE   5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE   6 EXPRESSION_SYSTEM_TAXID: 511693;
SOURCE   7 EXPRESSION_SYSTEM_STRAIN: BL21;
SOURCE   8 EXPRESSION_SYSTEM_VECTOR: PDEST17
KEYWDS    HYDROLASE, GLYCOSIDE HYDROLASE, ALPHA-L-FUCOSIDASE, THERMOSTABLE
EXPDTA    X-RAY DIFFRACTION
AUTHOR    G.SULZENBACHER,C.BIGNON,Y.BOURNE,B.HENRISSAT
REVDAT   5   13-JUL-11 1HL8    1       VERSN
REVDAT   4   24-FEB-09 1HL8    1       VERSN
REVDAT   3   25-MAR-04 1HL8    1       JRNL
REVDAT   2   13-FEB-04 1HL8    1       REMARK
REVDAT   1   15-JAN-04 1HL8    0
JRNL        AUTH   G.SULZENBACHER,C.BIGNON,T.NISHIMURA,C.A.TARLING,S.G.WITHERS,
JRNL        AUTH 2 B.HENRISSAT,Y.BOURNE
JRNL        TITL   CRYSTAL STRUCTURE OF THERMOTOGA MARITIMA ALPHA-L-
JRNL        TITL 2 FUCOSIDASE. INSIGHTS INTO THE CATALYTIC MECHANISM AND THE
JRNL        TITL 3 MOLECULAR BASIS FOR FUCOSIDOSIS.
JRNL        REF    J.BIOL.CHEM.                  V. 279 13119 2004
JRNL        REFN                   ISSN 0021-9258
JRNL        PMID   14715651
JRNL        DOI    10.1074/JBC.M313783200
REMARK   2
REMARK   2 RESOLUTION.    2.4  ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : REFMAC 5.1.24
REMARK   3   AUTHORS     : MURSHUDOV,VAGIN,DODSON
REMARK   3
REMARK   3    REFINEMENT TARGET : MAXIMUM LIKELIHOOD
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.40
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 37.27
```

Residue number

Chain

Amino acid type

Coordinates

ATOM	1	N	ARG	A	7	-26.699	-11.392	48.842	1.00	56.84	N
ATOM	2	CA	ARG	A	7	-25.554	-10.912	49.663	1.00	55.29	C
ATOM	3	C	ARG	A	7	-24.623	-9.995	48.864	1.00	53.39	C
ATOM	4	O	ARG	A	7	-24.414	-10.191	47.661	1.00	54.25	O
ATOM	5	CB	ARG	A	7	-24.761	-12.105	50.193	1.00	56.00	C
ATOM	6	CG	ARG	A	7	-25.374	-12.749	51.426	1.00	58.45	C
ATOM	7	CD	ARG	A	7	-24.396	-12.945	52.578	1.00	59.72	C
ATOM	8	NE	ARG	A	7	-25.048	-12.736	53.869	1.00	61.30	N
ATOM	9	CZ	ARG	A	7	-24.413	-12.499	55.014	1.00	61.72	C
ATOM	10	NH1	ARG	A	7	-23.087	-12.440	55.065	1.00	61.05	N
ATOM	11	NH2	ARG	A	7	-25.115	-12.320	56.126	1.00	63.61	N
ATOM	12	N	TYR	A	8	-24.055	-9.007	49.545	1.00	50.83	N
ATOM	13	CA	TYR	A	8	-23.096	-8.100	48.940	1.00	48.87	C
ATOM	14	C	TYR	A	8	-21.680	-8.609	49.201	1.00	47.84	C
ATOM	15	O	TYR	A	8	-21.378	-9.123	50.279	1.00	47.98	O
ATOM	16	CB	TYR	A	8	-23.287	-6.680	49.481	1.00	47.56	C
ATOM	17	CG	TYR	A	8	-24.700	-6.147	49.294	1.00	48.37	C
ATOM	18	CD1	TYR	A	8	-25.123	-5.630	48.067	1.00	49.00	C
ATOM	19	CD2	TYR	A	8	-25.619	-6.180	50.332	1.00	48.91	C
ATOM	20	CE1	TYR	A	8	-26.419	-5.156	47.889	1.00	48.83	C
ATOM	21	CE2	TYR	A	8	-26.918	-5.707	50.160	1.00	50.24	C
ATOM	22	CZ	TYR	A	8	-27.306	-5.192	48.936	1.00	49.98	C
ATOM	23	OH	TYR	A	8	-28.589	-4.719	48.773	1.00	51.15	O
ATOM	24	N	LYS	A	9	-20.837	-8.493	48.178	1.00	46.89	N

mmCIF format

- Created 1990, preferred since 2014, since 2019 the only accepted for deposition
- Non-context grammar, more flexible, possibility to add further information

```
data_4I1S
#
_entry.id   4I1S
#
_audit_conform.dict_name       mmcif_pdbx.dic
_audit_conform.dict_version    4.027
_audit_conform.dict_location   http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
loop_
_database_2.database_id
_database_2.database_code
PDB  4I1S
RCSB RCSB076195
#
_database_PDB_rev.num            1
_database_PDB_rev.date           2013-01-30
_database_PDB_rev.date_original  2012-11-21
_database_PDB_rev.status         ?
_database_PDB_rev.replaces       4I1S
_database_PDB_rev.mod_type       0
#
_pdbx_database_status.status_code                     REL
_pdbx_database_status.entry_id                        4I1S
_pdbx_database_status.deposit_site                    RCSB
_pdbx_database_status.process_site                    RCSB
_pdbx_database_status.status_code_sf                  REL
_pdbx_database_status.status_code_mr                  ?
_pdbx_database_status.SG_entry                        ?
_pdbx_database_status.status_code_cs                  ?
_pdbx_database_status.methods_development_category    ?
#
loop_
_audit_author.name
_audit_author.pdbx_ordinal
'Motz, C.'      1
'Witte, G.'     2
'Hopfner, K.P.' 3
#
_citation.id                        primary
_citation.title
'Paramyxovirus V Proteins Disrupt the Fold of the RNA Sensor MDA5 to Inhibit Antiviral
_citation.journal_abbrev            Science
```

Chain

Residue number

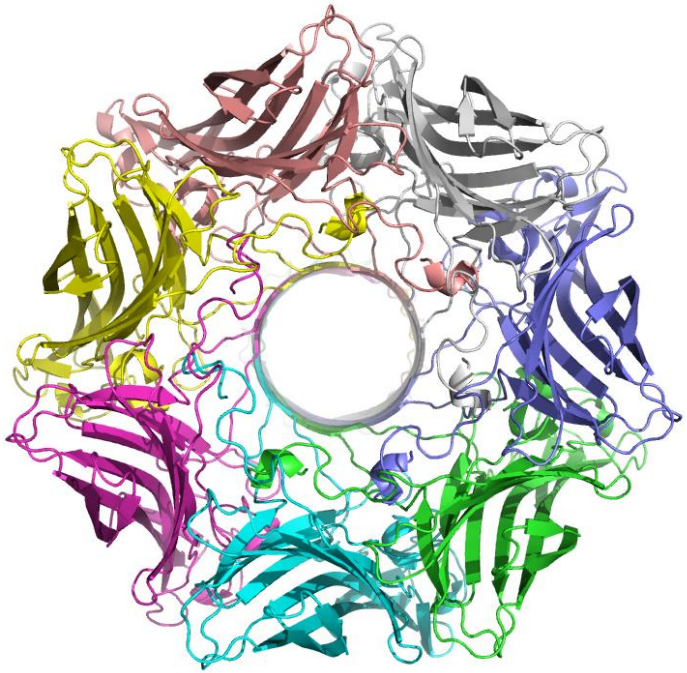
Amino acid type

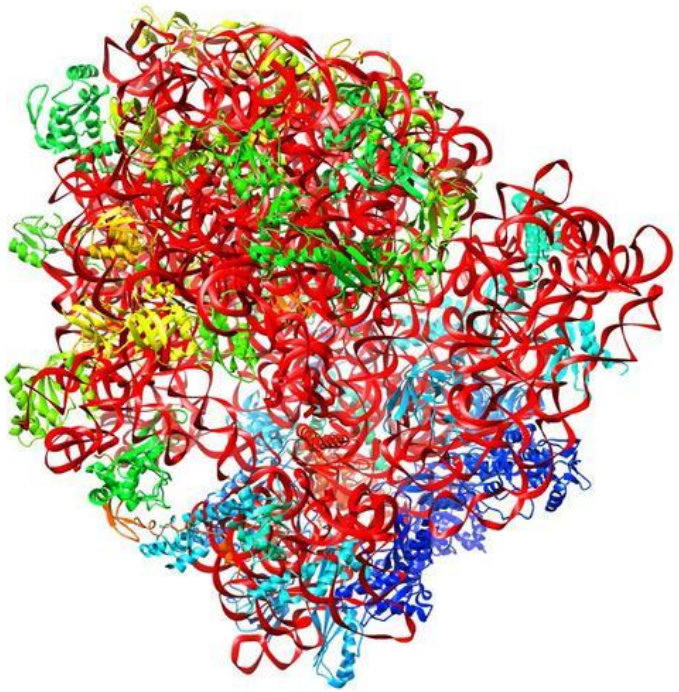
Coordinates

```
_atom_site.pdbx_PDB_model_num
ATOM   1     N  N    . GLU A 1 1   ? 7.254   11.020 4.888  1.00 61.38  ? ? ? ? ? ? 546 GLU A N    1
ATOM   2     C  CA   . GLU A 1 1   ? 6.404   12.200 5.071  1.00 67.04  ? ? ? ? ? ? 546 GLU A CA   1
ATOM   3     C  C    . GLU A 1 1   ? 7.111   13.526 4.729  1.00 59.60  ? ? ? ? ? ? 546 GLU A C    1
ATOM   4     O  O    . GLU A 1 1   ? 6.576   14.360 3.999  1.00 64.05  ? ? ? ? ? ? 546 GLU A O    1
ATOM   5     C  CB   . GLU A 1 1   ? 5.842   12.232 6.500  1.00 74.02  ? ? ? ? ? ? 546 GLU A CB   1
ATOM   6     C  CG   . GLU A 1 1   ? 5.625   13.627 7.094  1.00 74.52  ? ? ? ? ? ? 546 GLU A CG   1
ATOM   7     C  CD   . GLU A 1 1   ? 4.448   14.369 6.495  1.00 78.40  ? ? ? ? ? ? 546 GLU A CD   1
ATOM   8     O  OE1  . GLU A 1 1   ? 3.968   13.977 5.409  1.00 81.00  ? ? ? ? ? ? 546 GLU A OE1  1
ATOM   9     O  OE2  . GLU A 1 1   ? 3.997   15.354 7.118  1.00 79.97  ? ? ? ? ? ? 546 GLU A OE2  1
ATOM   10    N  N    . ASP A 1 2   ? 8.299   13.714 5.287  1.00 44.26  ? ? ? ? ? ? 547 ASP A N    1
ATOM   11    C  CA   . ASP A 1 2   ? 9.213   14.768 4.873  1.00 34.80  ? ? ? ? ? ? 547 ASP A CA   1
ATOM   12    C  C    . ASP A 1 2   ? 10.508  14.039 4.527  1.00 30.06  ? ? ? ? ? ? 547 ASP A C    1
ATOM   13    O  O    . ASP A 1 2   ? 11.245  13.650 5.424  1.00 29.92  ? ? ? ? ? ? 547 ASP A O    1
ATOM   14    C  CB   . ASP A 1 2   ? 9.460   15.735 6.039  1.00 34.15  ? ? ? ? ? ? 547 ASP A CB   1
ATOM   15    C  CG   . ASP A 1 2   ? 10.399  16.909 5.672  1.00 36.09  ? ? ? ? ? ? 547 ASP A CG   1
ATOM   16    O  OD1  . ASP A 1 2   ? 11.138  16.835 4.665  1.00 33.05  ? ? ? ? ? ? 547 ASP A OD1  1
ATOM   17    O  OD2  . ASP A 1 2   ? 10.397  17.917 6.418  1.00 36.96  ? ? ? ? ? ? 547 ASP A OD2  1
ATOM   18    N  N    . LEU A 1 3   ? 10.778  13.854 3.239  1.00 32.19  ? ? ? ? ? ? 548 LEU A N    1
ATOM   19    C  CA   . LEU A 1 3   ? 11.922  13.061 2.787  1.00 30.81  ? ? ? ? ? ? 548 LEU A CA   1
ATOM   20    C  C    . LEU A 1 3   ? 13.253  13.688 3.155  1.00 27.21  ? ? ? ? ? ? 548 LEU A C    1
```

Quaternary structure

- **Association of individual** (protein) **chains**
- Consisting of identical chains (**homooligomers**) or different chains (**heterooligomers**), including non-protein molecules, e.g. nucleic acids





Structure dynamics

- **Molecular structure is not static picture**
- Various degrees of dynamics
  - Atom thermal motion
  - Flexibility of side chains
  - Flexibility of backbone
  - Association/dissociation of subunits

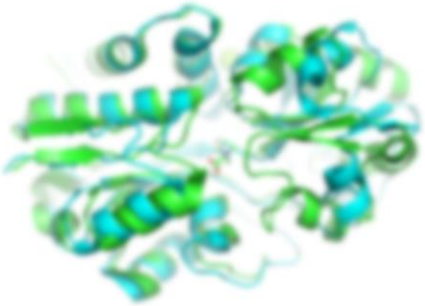
Structure in time

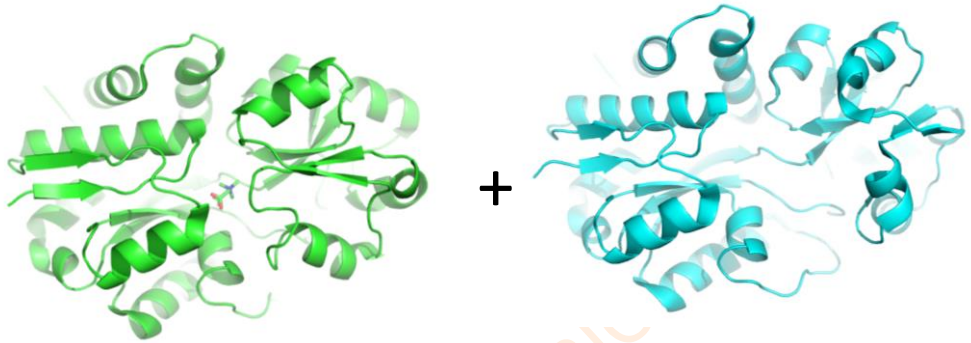
- **Static**
  - Thermal motion – low temperature
  - Short time scale
  - Averaging

- **Dynamic**
  - Multiple structures comparison
  - Precision of structure determination
  - Dedicated methods

Speed of change

- **Fast change**
  - **Average** signal
  - **Continuum** of states
- **Slow change**
  - Sample **heterogeneity**
  - **Separation** of species or signals





2D structure dynamics

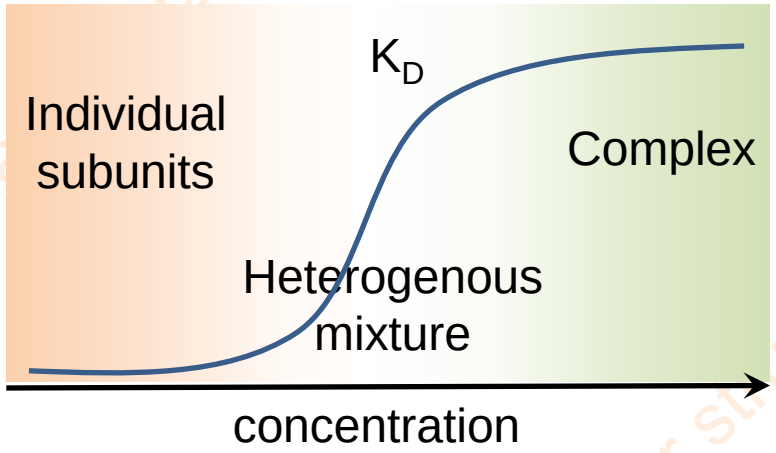
- 2D structure enables formation of **higher structures** (3D, 4D)
- Experimentaly (CD, IR) determined **average**
- Environment/interaction-**induced** 2D structure change

3D structure dynamics

- Flexible **backbone – intrinsically disordered proteins** (IDPs)
- Flexible parts – **loops**, N- and C- **termini**
  - Random movement
  - Stabilization by **ligand binding**
- Side chains
  - Preferred **conformations**
  - Stabilization by additional bonds (H, polar, hydrophobic)

4D structure dynamics

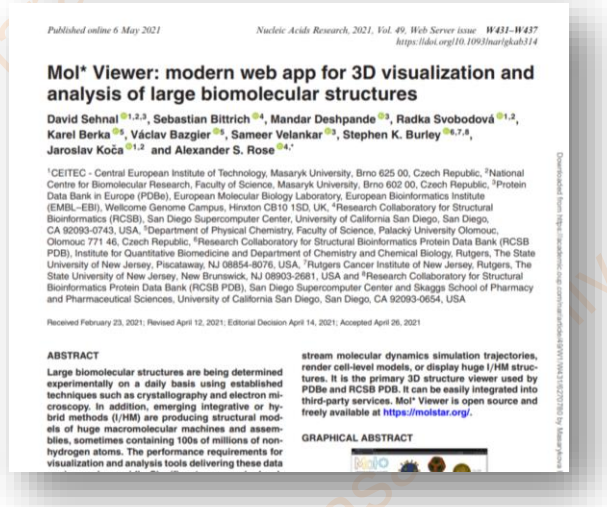
- **Association/dissociation equilibrium**



- Multi-component complex – various dissociation constants

Visualization of 3D structures

- Many SW tools: **Mol***, PyMol, Jmol, LiteMol, RasMol, VMD, Chimera, Cn3D,…
- Various applications (web-based, high-resolution images, platform-specific)
- Several styles for different purposes



Published online 6 May 2021 — Nucleic Acids Research, 2021, Vol. 49, Web Server issue W431–W437 — https://doi.org/10.1093/nar/gkab314

Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures

David Sehnal[1,2,3], Sebastian Bittrich[4], Mandar Deshpande[3], Radka Svobodová[1,2], Karel Berka[5], Václav Bazgier[5], Sameer Velankar[3], Stephen K. Burley[6,7,8], Jaroslav Koča[1,2] and Alexander S. Rose[4,*]

[1]CEITEC - Central European Institute of Technology, Masaryk University, Brno 625 00, Czech Republic, [2]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno 602 00, Czech Republic, [3]Protein Data Bank in Europe (PDBe), European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL–EBI), Wellcome Genome Campus, Hinxton CB10 1SD, UK, [4]Research Collaboratory for Structural Bioinformatics (RCSB), San Diego Supercomputer Center, University of California San Diego, San Diego, CA 92093-0743, USA, [5]Department of Physical Chemistry, Faculty of Science, Palacký University Olomouc, Olomouc 771 46, Czech Republic, [6]Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854-8076, USA, [7]Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ 08903-2681, USA and [8]Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), San Diego Supercomputer Center and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, San Diego, CA 92093-0654, USA

Received February 23, 2021; Revised April 12, 2021; Editorial Decision April 14, 2021; Accepted April 26, 2021

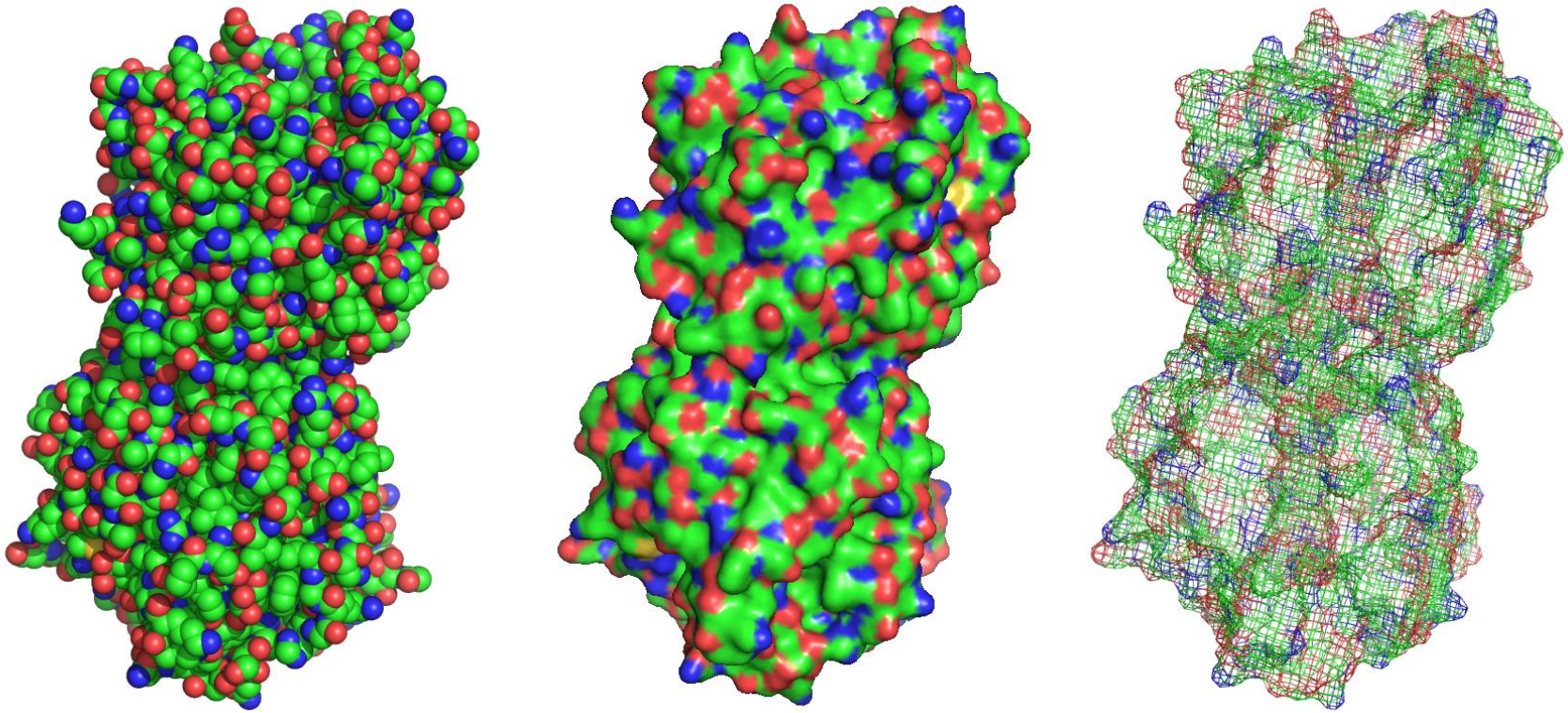
ABSTRACT

Large biomolecular structures are being determined experimentally on a daily basis using established techniques such as crystallography and electron microscopy. In addition, emerging integrative or hybrid methods (I/HM) are producing structural models of huge macromolecular machines and assemblies, sometimes containing 100s of millions of non-hydrogen atoms. The performance requirements for visualization and analysis tools delivering these data stream molecular dynamics simulation trajectories, render cell-level models, or display huge I/HM structures. It is the primary 3D structure viewer used by PDBe and RCSB PDB. It can be easily integrated into third-party services. Mol* Viewer is open source and freely available at https://molstar.org/.

GRAPHICAL ABSTRACT

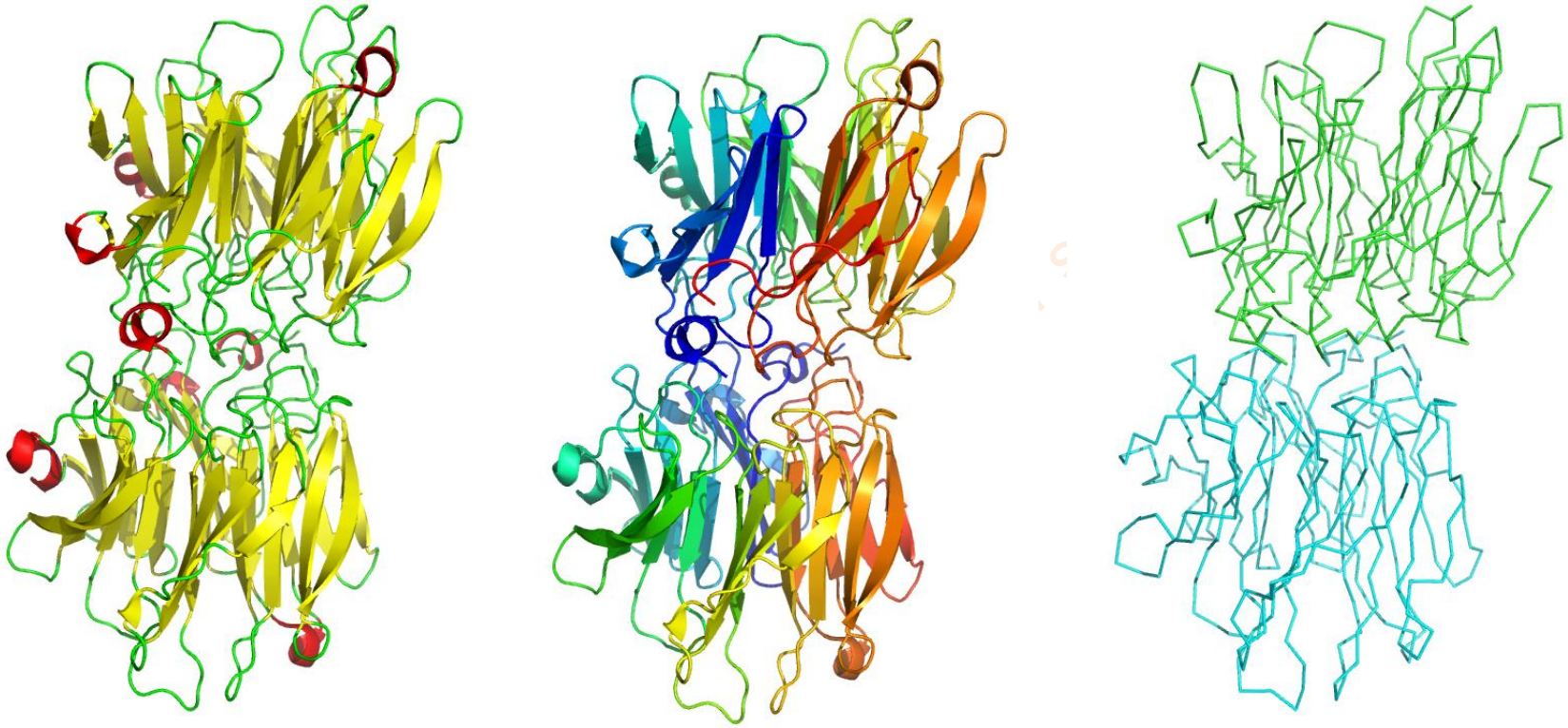
Visualization of 3D structures

- Spheres / Surface / Mesh
- Space filling, interaction surfaces, overall shape



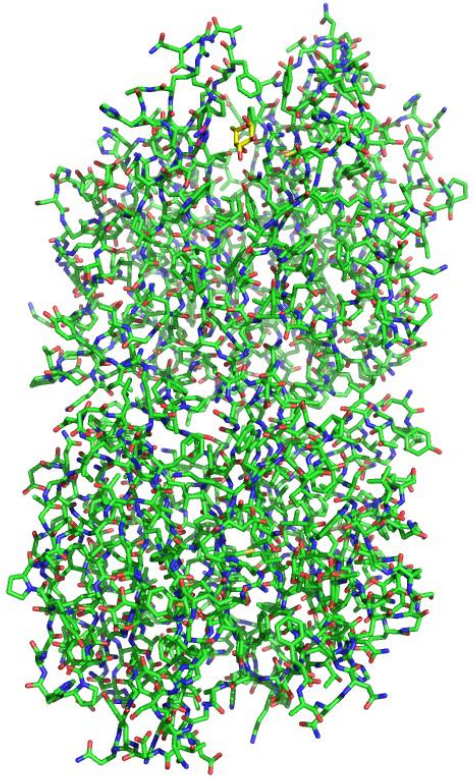
Visualization of 3D structures

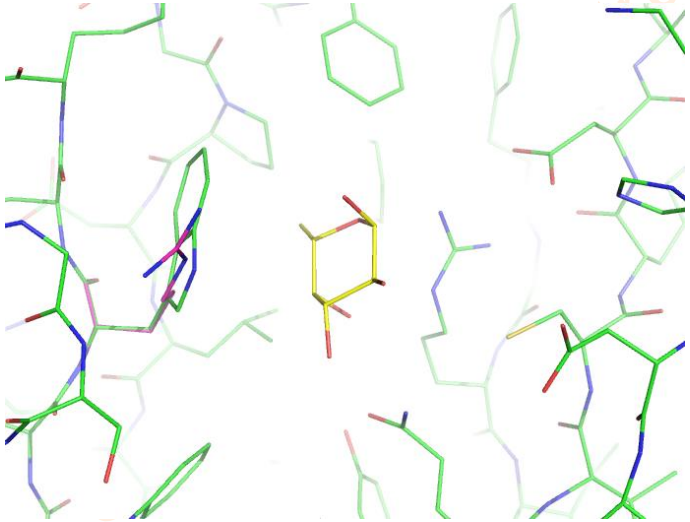
- Cartoon / Ribbon
- Secondary structures, domain assignment, connectivity, main chain orientation

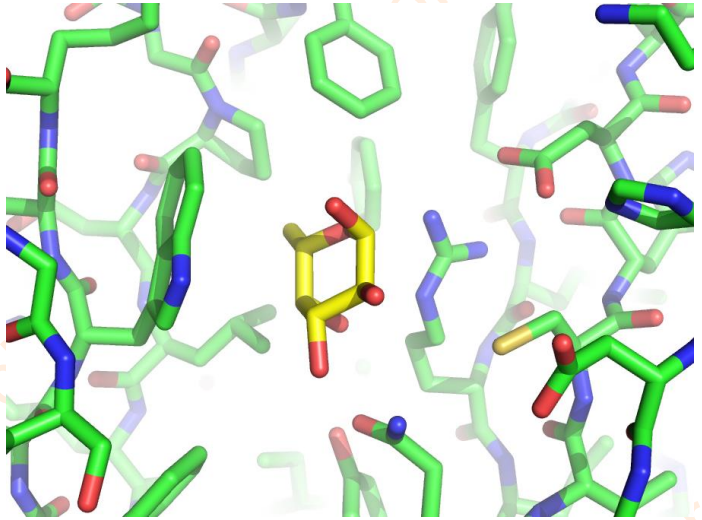


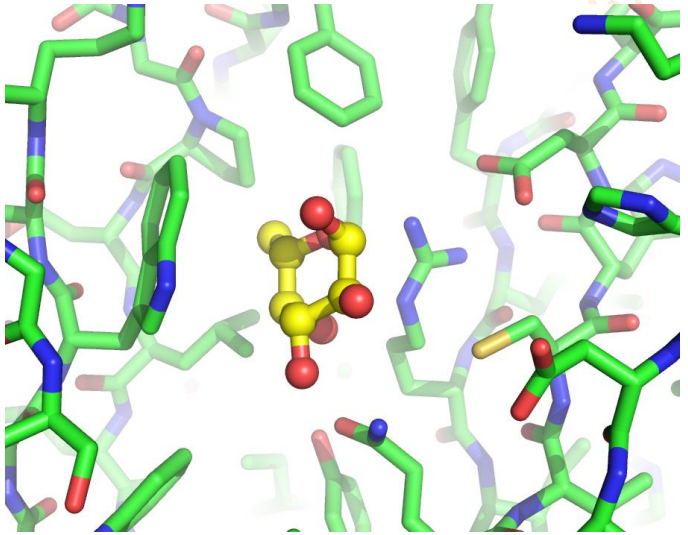
Visualization of 3D structures

- Sticks / Balls & sticks / Lines
- Detailed view, side chain orientation, mutations









Visualization of 3D structures

- Combination of representations – for publication purposes
- Additional graphics – hydrogen bonds, distances, clashes, labeling, electron density, ...

Good figure should be both **nice** and **clear** (!)

Questions?

Biomolecular Interactions and Crystallography Core Facility

bic@ceitec.cz

bic.ceitec.cz

MUNI | CEITEC

CF Head

Josef Houser

- +420 549 492 527
- josef.houser@ceitec.cz