

Basics of Molecular Modeling & Computer-Aided Drug design

Prof. Giulio Poli

Dep. of Pharmacy, University of Pisa, Italy

E-mail: giulio.poli@unipi.it

Web site: <http://www.mmvsl.it/wp/group-members/poli/>

Introduction & Graphical Representation

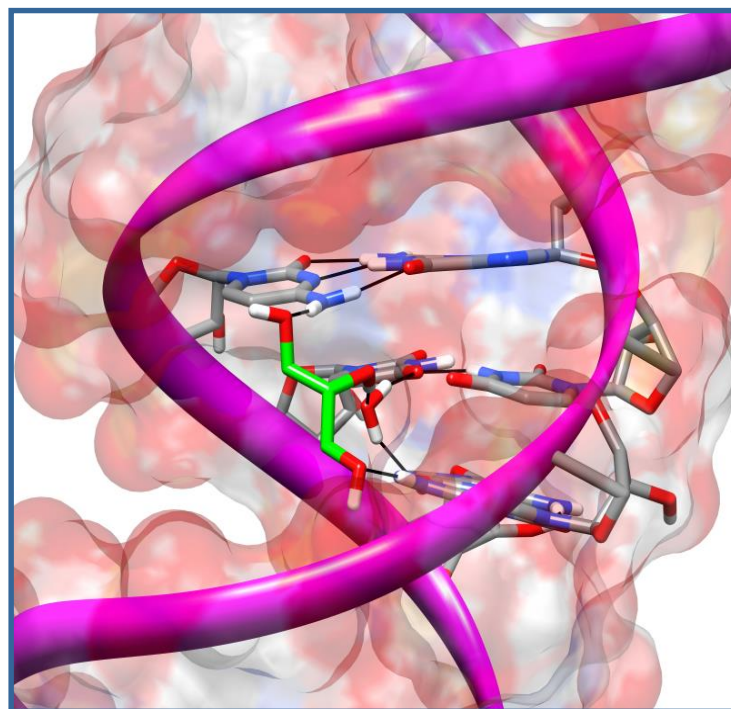
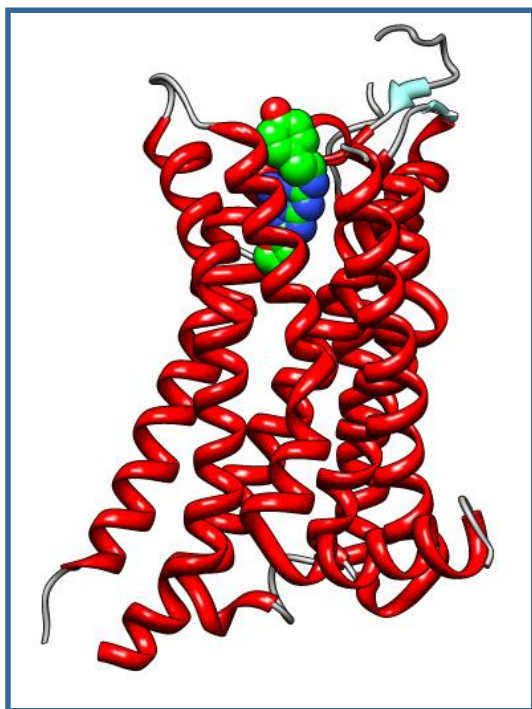
Introduction

Computational chemistry is a branch of chemistry that uses principles of computer science to assist in solving chemical problems. It uses the results of theoretical chemistry, incorporated into efficient computer programs, to calculate the structures and properties of **molecules** and solids. While its results normally complement the information obtained by chemical experiments, it can in some cases predict hitherto unobserved chemical phenomena. It is widely used in the design of new **drugs** and materials.

It covers a range of chemical applications such as quantum chemistry, molecular dynamics, molecular modelling, molecular mechanics and chemoinformatics. In **computer-aided drug design** many of these applications are employed to study and evaluate the (possible) interactions of **small-molecules** with a **biological target (receptor)**.

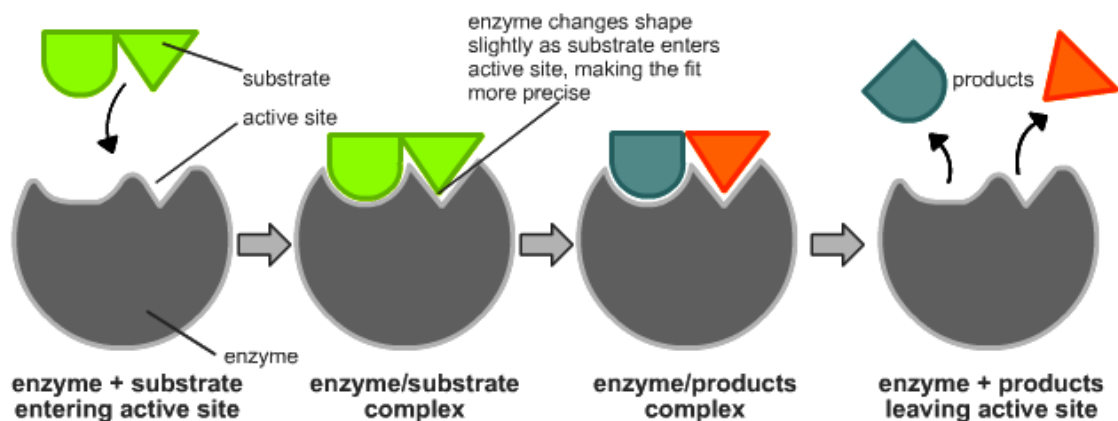
Receptor definition

A **receptor** is a biological structure to which one or more specific kinds of signaling molecules may attach. The main receptors are proteins or nucleic acids. A molecule which binds to a receptor is called a **ligand**, and may be a peptide or other small molecule, such as a neurotransmitter, a hormone, a pharmaceutical drug, or a toxin.



Induced-Fit Model

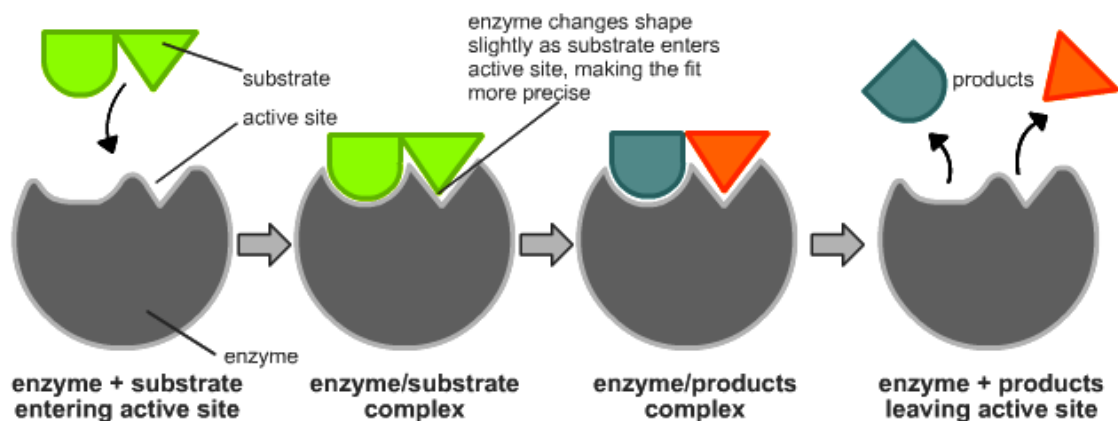
Since enzymes are rather flexible structures, the active site is continually reshaped by interactions with the substrate as the substrate interacts with the enzyme. As a result, the substrate does not simply bind to a rigid active site; the amino acid side chains which make up the active site are modeled into the precise positions that enable the enzyme to perform its catalytic function. In some cases, such as glycosidases, the substrate molecule also changes shape slightly as it enters the active site.



Induced-Fit Model

In the induced-fit model of enzyme action:

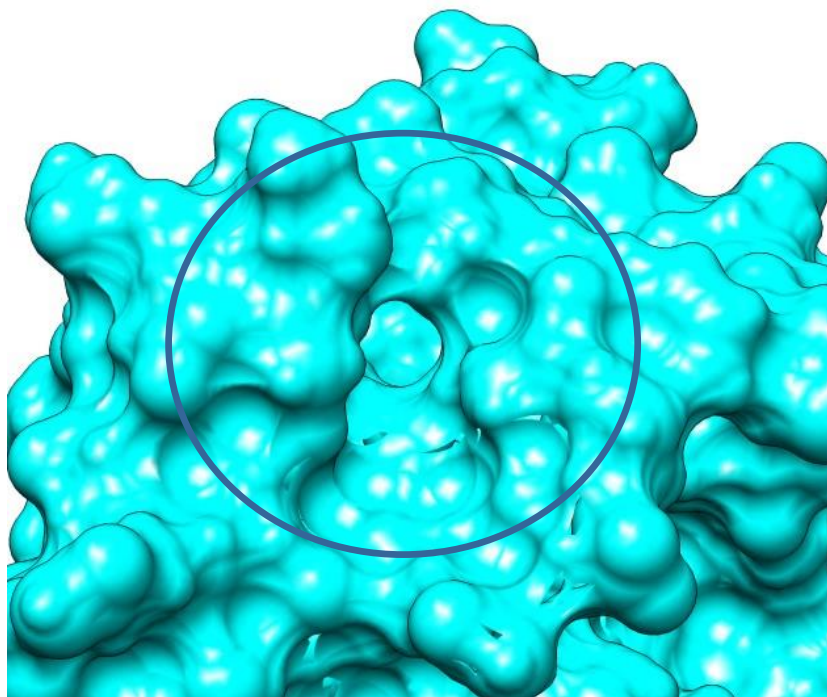
- the active site is flexible, not rigid
- the shapes of the enzyme, active site, and substrate/ligand adjust to maximize the fit, which improves catalysis
- there is a greater range of substrate specificity



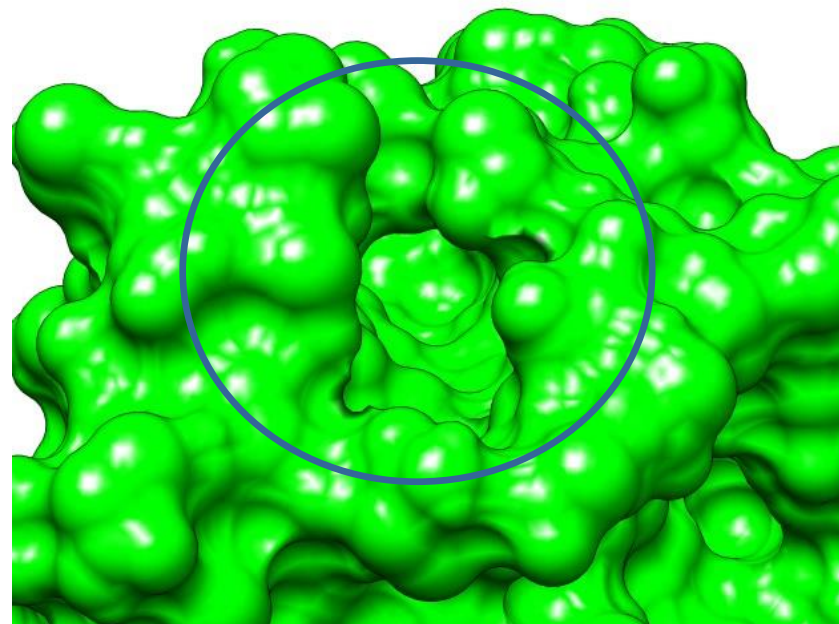
Induced-Fit Model

The active site geometry of a protein complex depends upon conformational changes induced by the bound ligand.

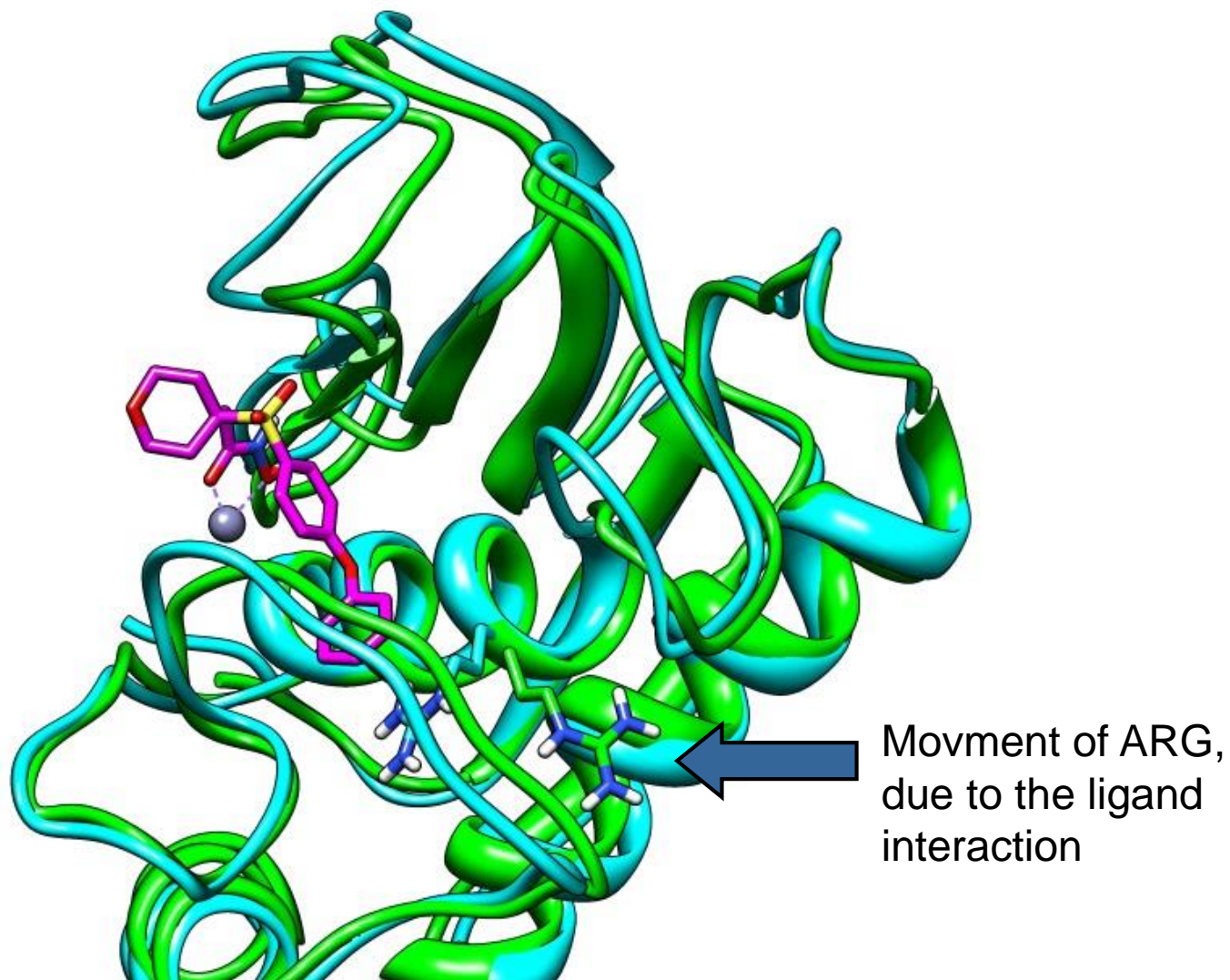
Binding site A



Binding site B



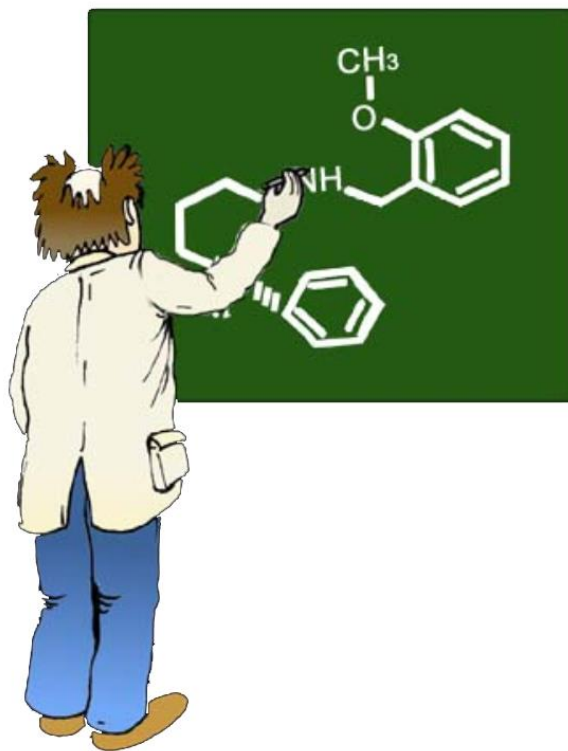
Induced-Fit Model



Molecular Modeling

Molecular Geometry

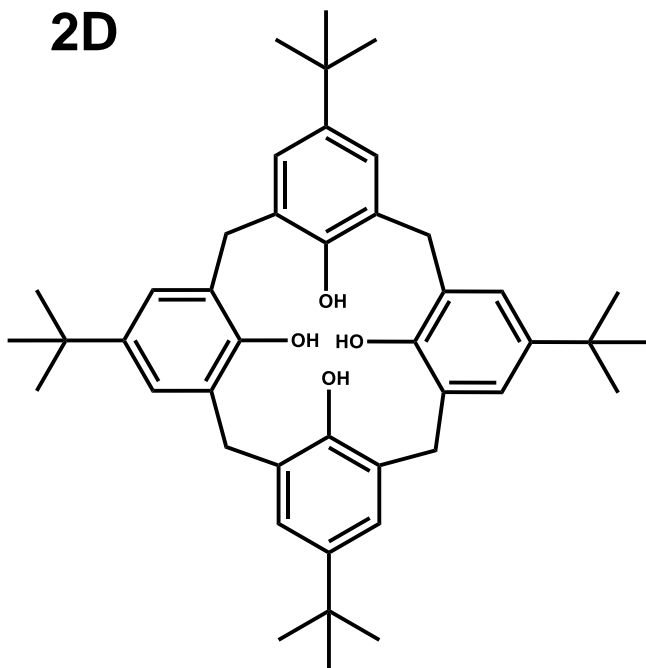
The way molecules have been perceived and defined has changed over the years. In the early 1970s, medicinal chemists considered molecules as mere topological two-dimensional (2D) entities with associated chemical and physicochemical properties.



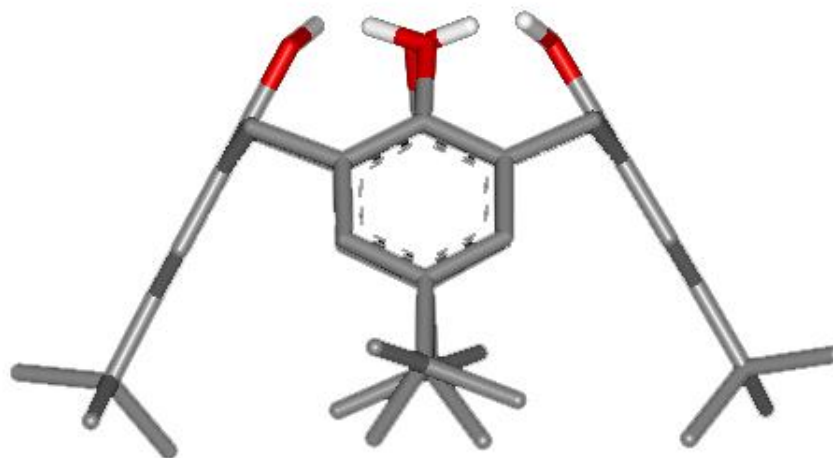
Molecular Geometry

The formula of a molecule can be drawn in two-dimensions, however it really exists in 3D space with precise geometrical features. For example the calix[4]arene showed here has a cone-like structure, a geometry that is not carried by its 2D formula.

2D

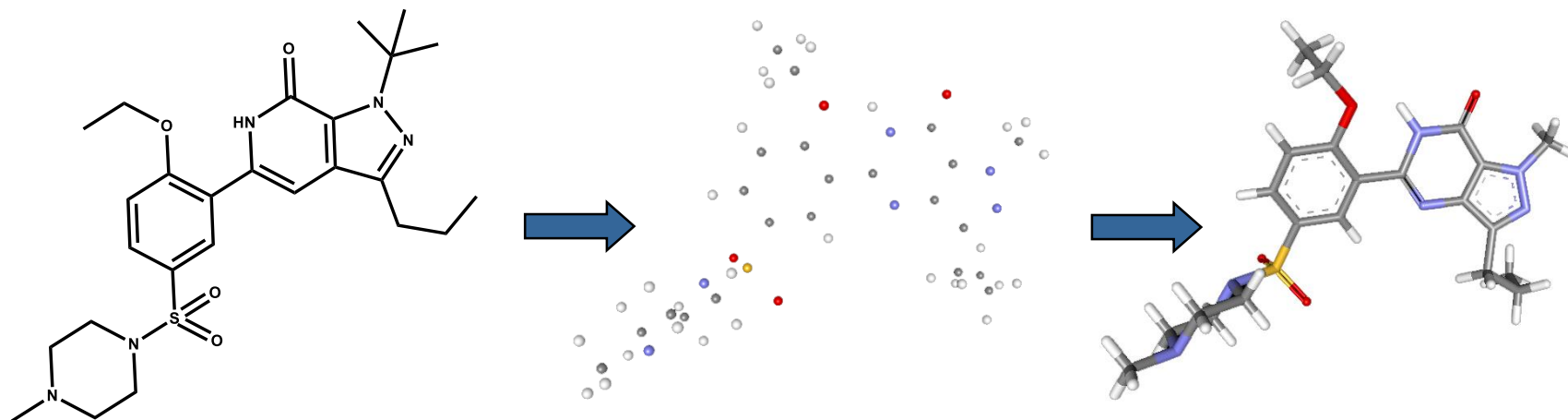


3D



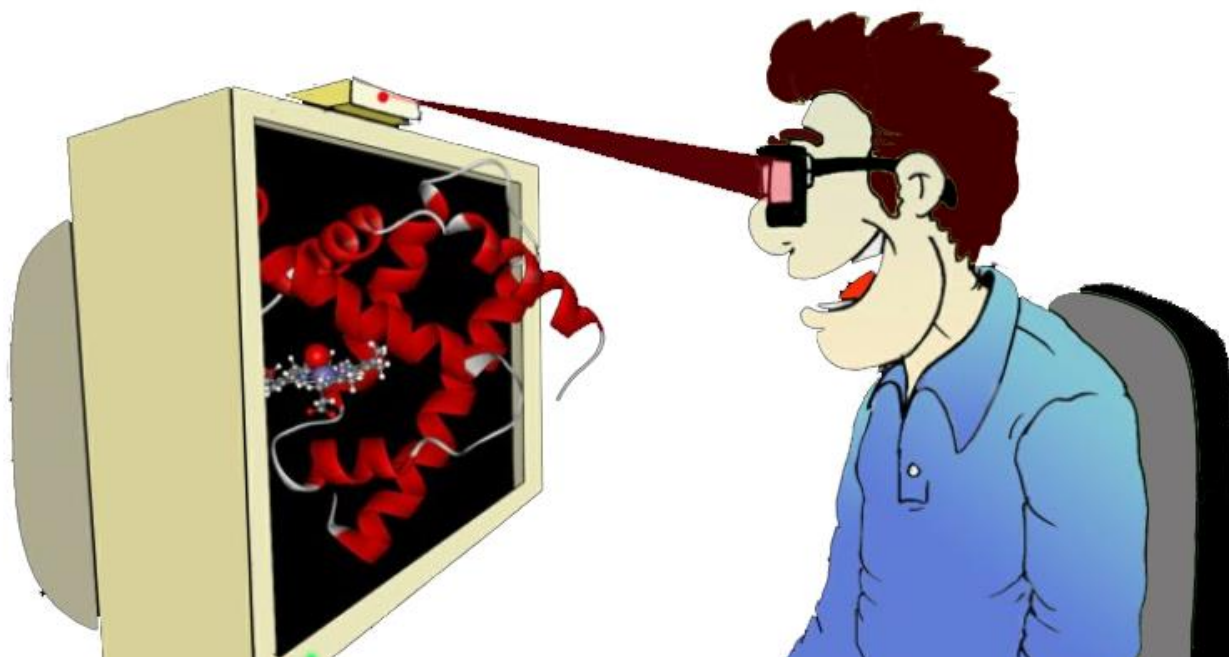
Molecular Geometry

A molecule is an assembly of atoms in 3D. The 2D structure defines the atoms and the connections between them; it becomes 3D when the location of the atoms is considered. The following molecule consists of 63 atoms, 66 bonds and 4 rings.



3D Stereo

Hardware stereo is a trick incorporated in graphics systems. The monitor runs at double frequency so that the screen presents alternate eye views one after another. The user wears a pair of goggles containing liquid crystal shutters and an infrared emitter on the workstation synchronizes the visibility of the screen to each eye.



Department of Pharmacy

Molecular Modeling & Virtual Screening Laboratory

3D Representation

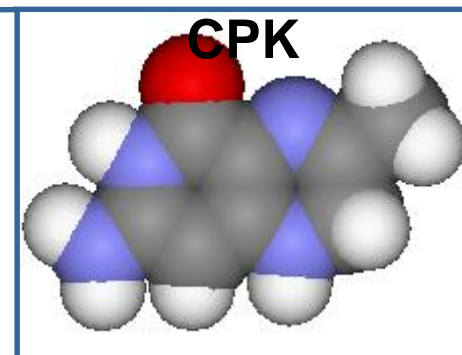
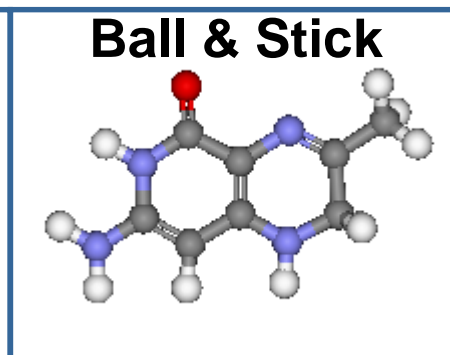
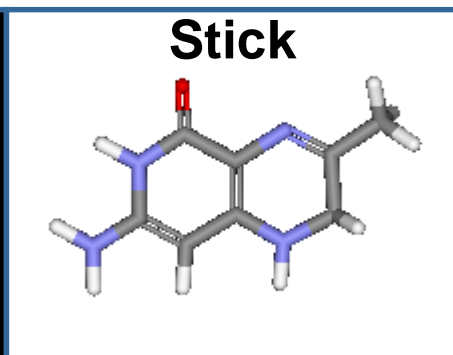
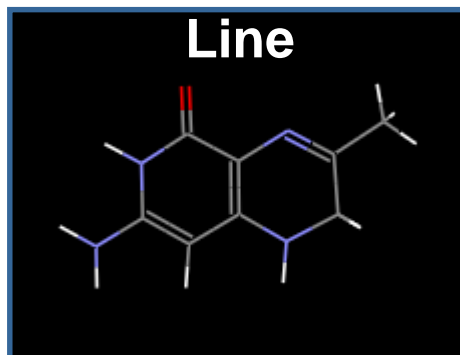
Molecules can be displayed using different rendering techniques.

Line. This is the simplest and most common way to visualize molecules where the bonding arrangement is represented in 3D. This type of representation is also called “wireframe”.

Stick. The bonds are represented as tubes.

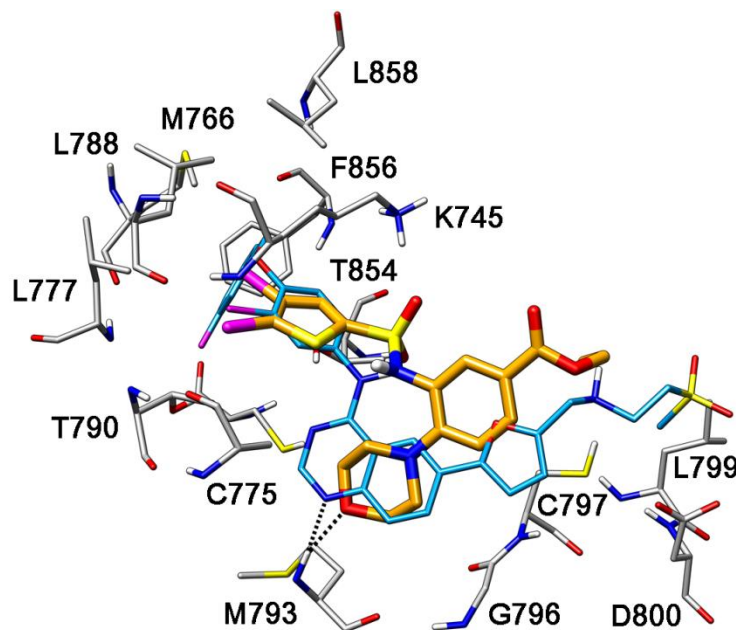
Ball & Stick. The molecule is displayed as the assembly of atoms and bonds. Atoms are represented as small spheres and bonds as tubes.

CPK. The molecule is defined as a set of spheres of van der Waals radii of the individual atoms.



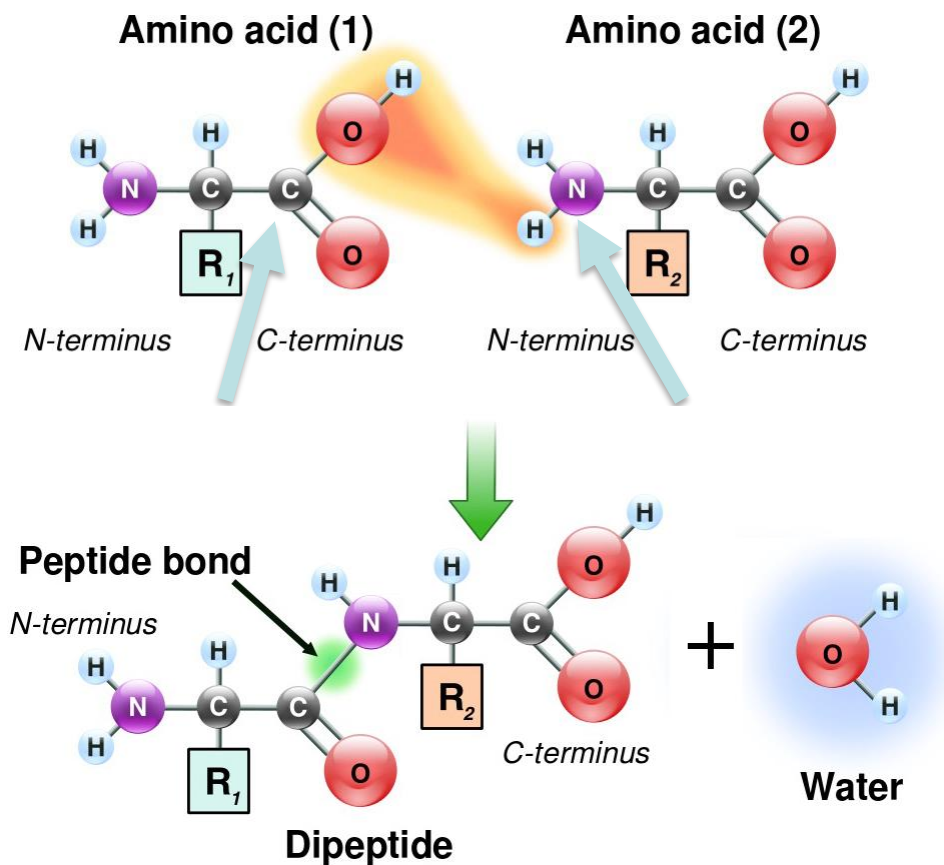
Proteins Representation

Macromolecules are complex entities. They can be displayed as small molecules using various techniques (line, ball and stick, etc.). Other representations are specific to macromolecules and give an overview of the overall molecular architecture of the protein.



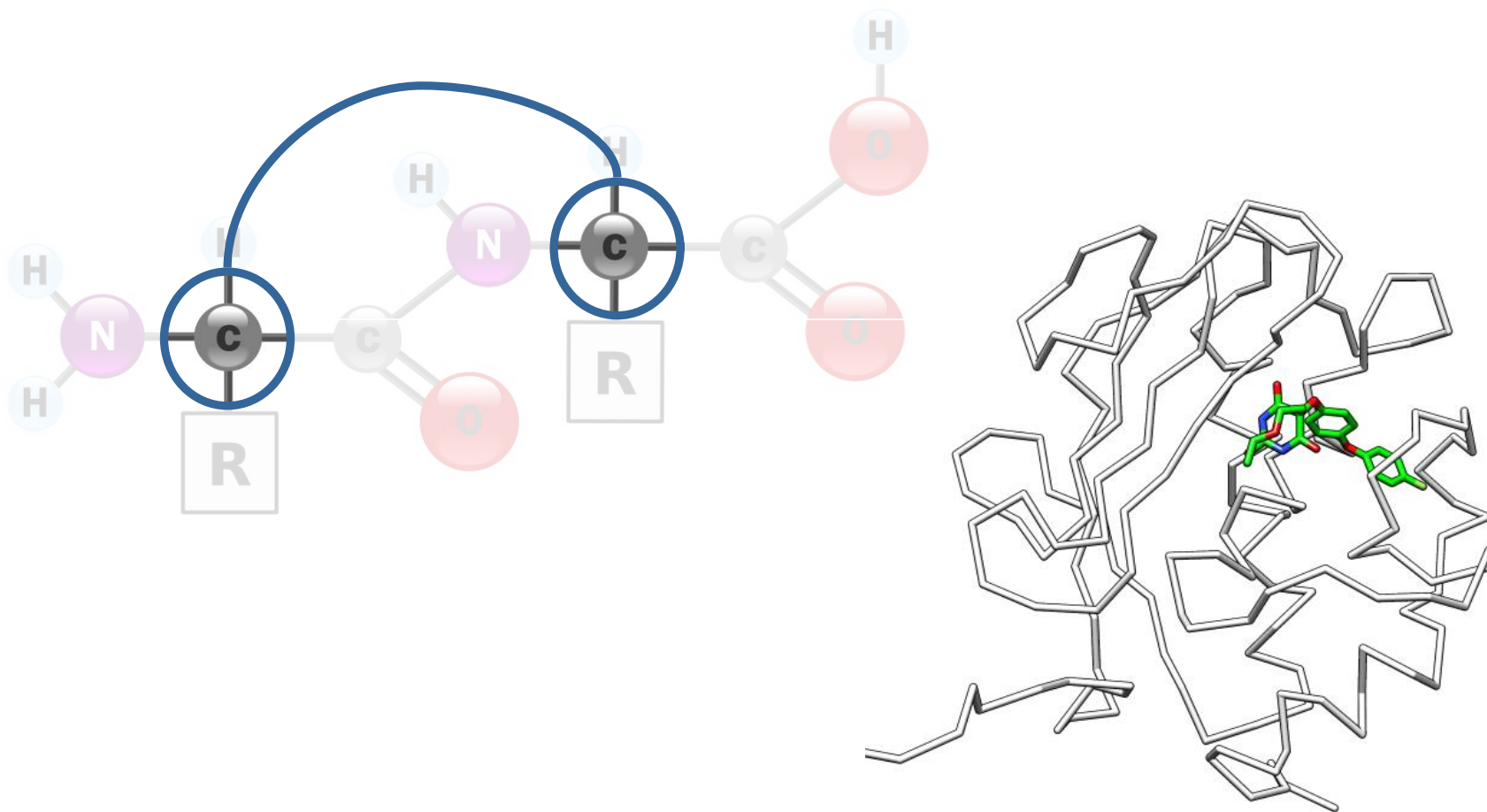
From Amino Acids to Proteins

Peptide Bonds



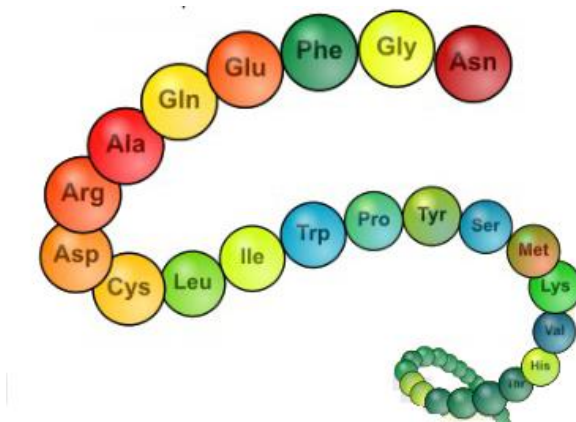
Proteins Representation: C α trace

This representation is useful for editing and aligning different proteins.



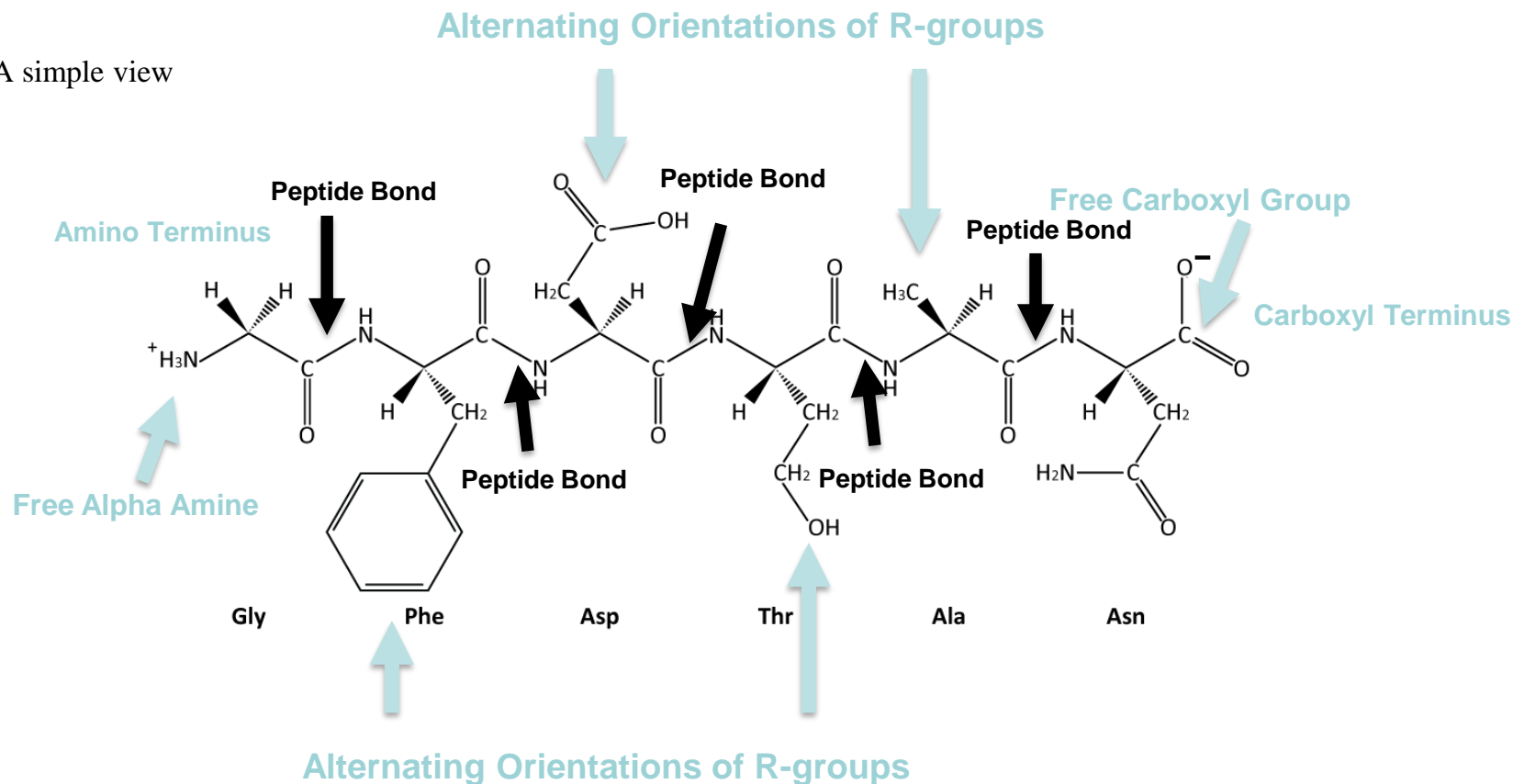
Primary Protein Structure

- Linear sequence of amino acids
- Joined by Peptide Bonds
- Translated from mRNA using Genetic Code
- Synthesis begins at amino-end, and terminates at carboxyl-end
- **Ultimately determines all properties of a protein**



Polypeptides

A simple view

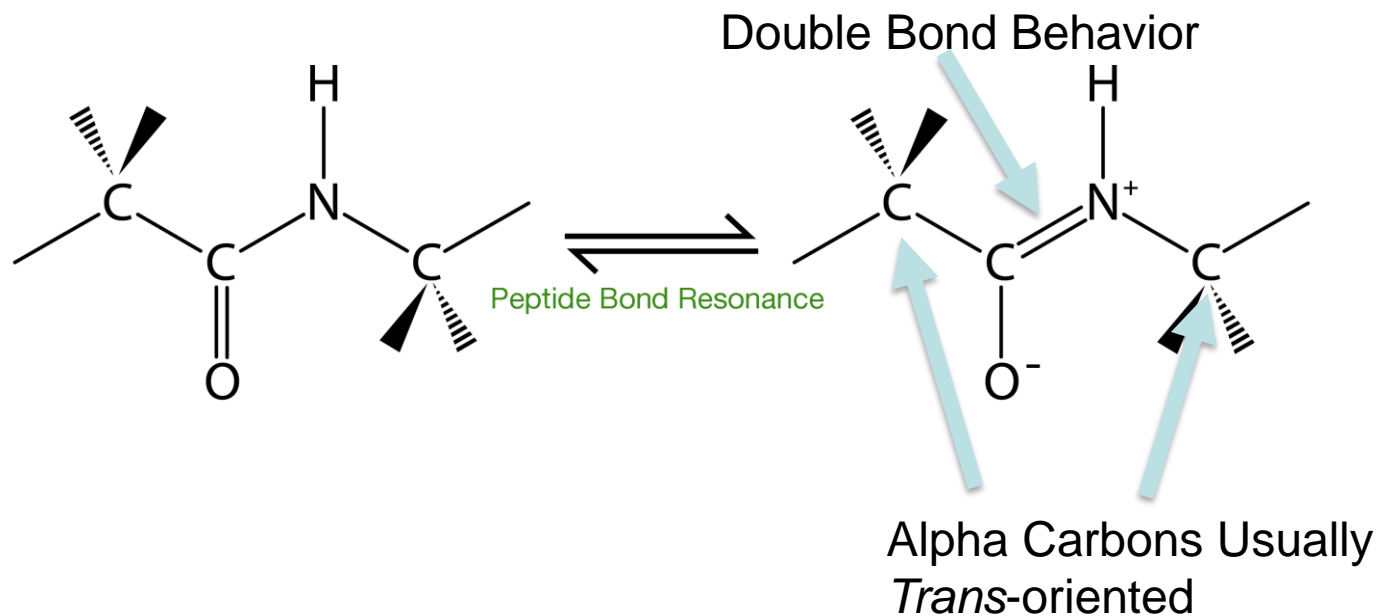


Department of Pharmacy

Molecular Modeling & Virtual Screening Laboratory

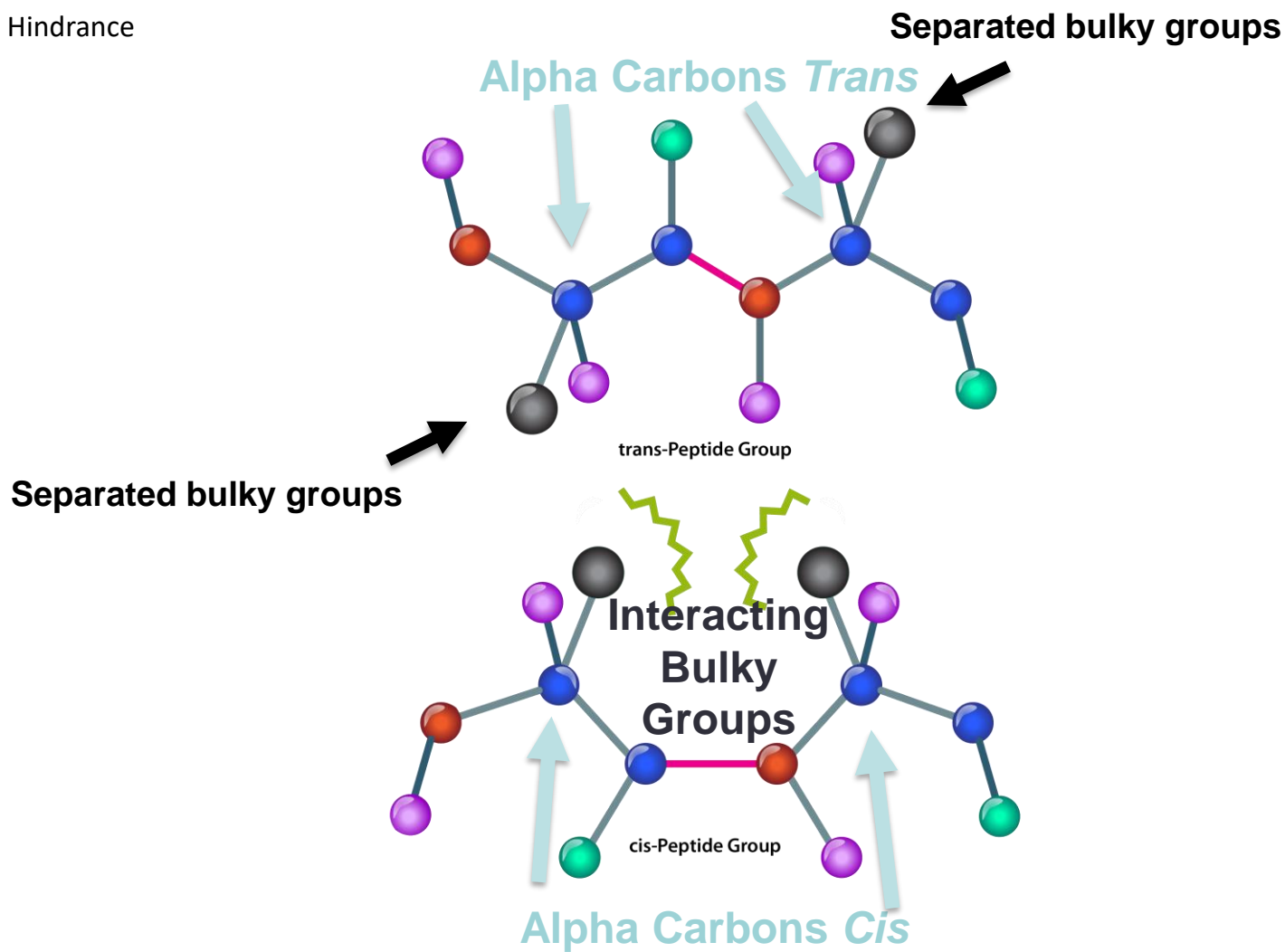
Peptide Bonds

Chemical Character



Polypeptides

Steric Hindrance



Department of Pharmacy

Molecular Modeling & Virtual Screening Laboratory

Polypeptides

Multiple Peptide Bond Planes

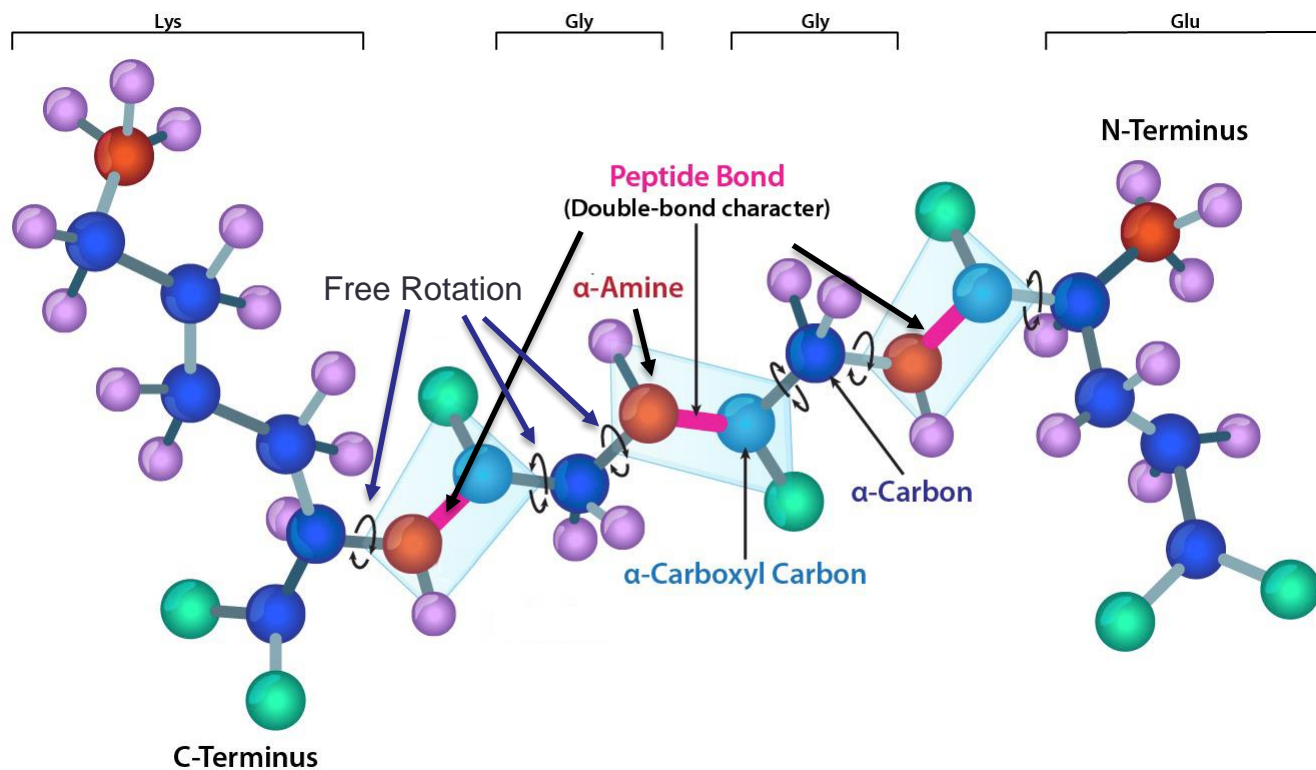


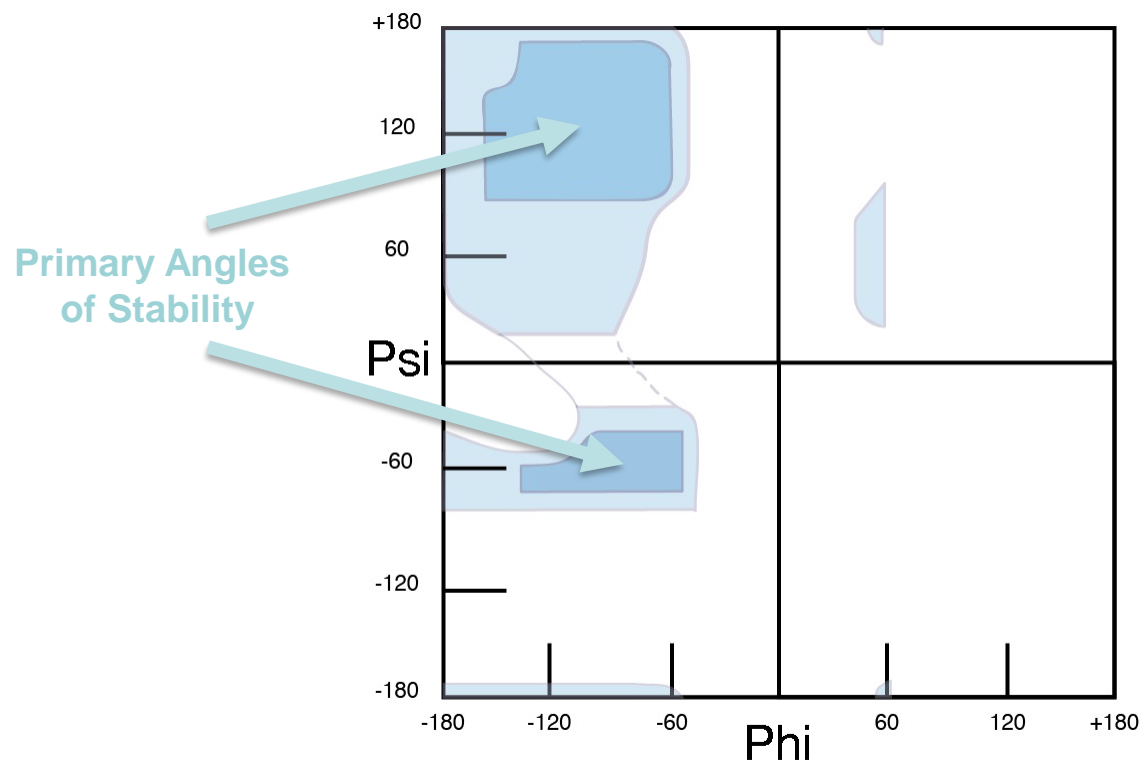
Diagram illustrating the three torsion angles that define the conformation of a protein backbone:

- Phi Angle (ϕ):** Rotation around the $C_{\alpha}-C$ bond.
- Psi Angle (ψ):** Rotation around the $C-C$ bond.
- Omega Angle (ω):** Rotation around the $C-N$ bond.

The diagram also shows the **Peptide Bond** (highlighted in blue) and the atoms involved: C_{α} , C , N , C_{β} , H , O , and $H(+1)$.

Ramachandran Plot

Bond Angles

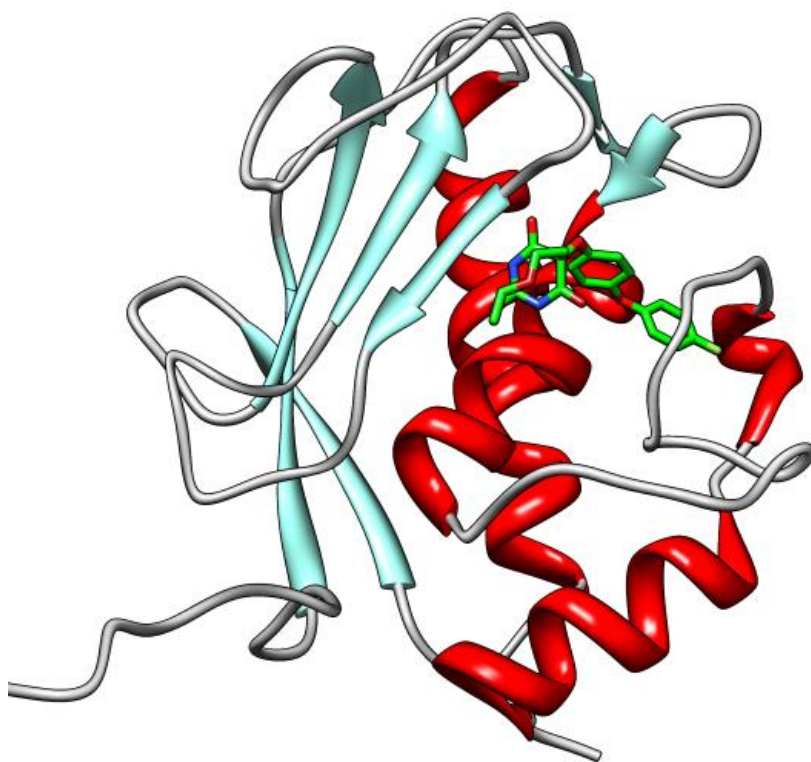


Department of Pharmacy

Molecular Modeling & Virtual Screening Laboratory

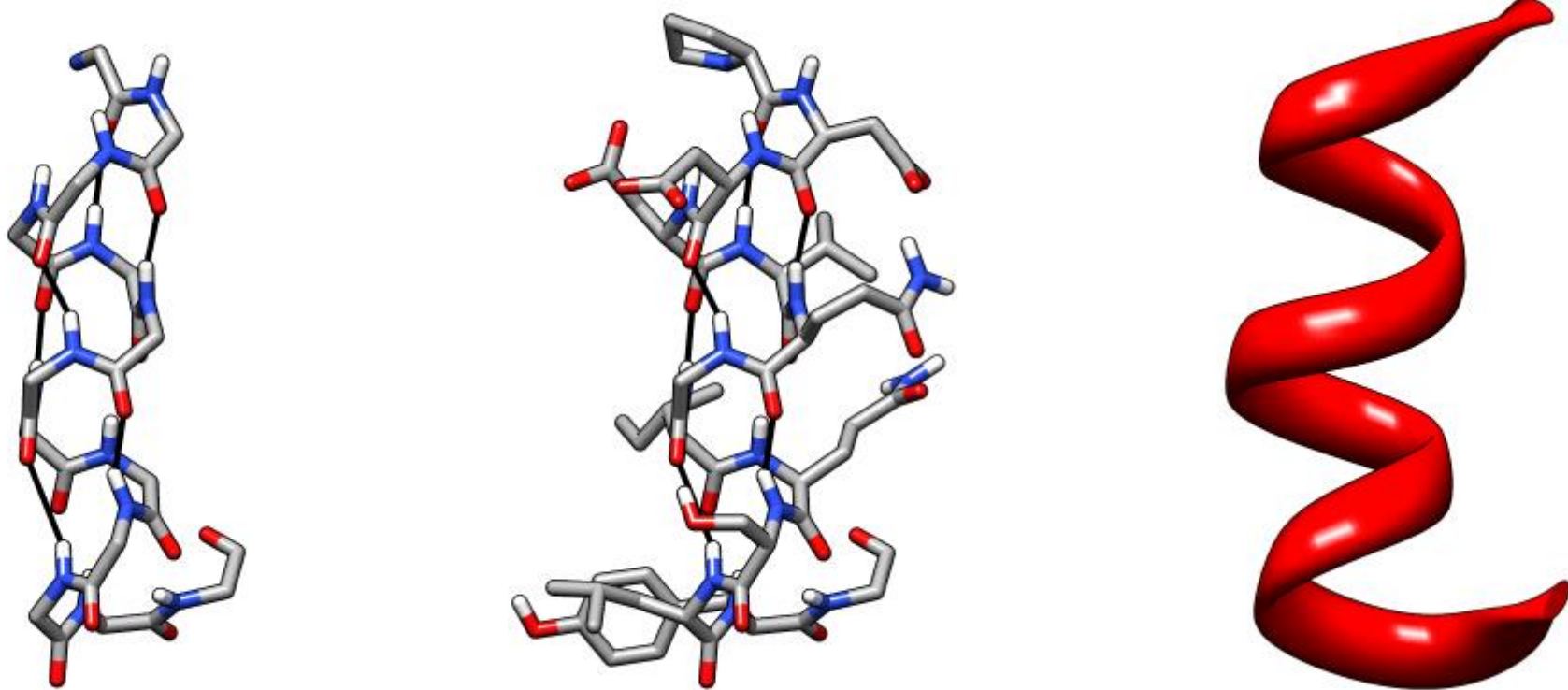
Proteins Representation: Ribbon

Ribbon representations are specific to proteins and provide an overview of the overall molecular architecture (secondary structure) of the protein. α -helices, β -sheets and β -turns are easily recognized.



Proteins Representation: α -helices

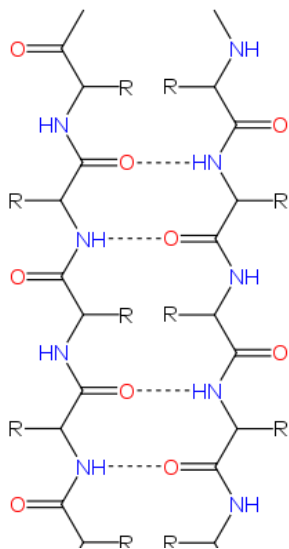
The **alpha helix** is a right-handed coiled or spiral conformation, in which every backbone N-H group donates a hydrogen bond to the backbone C=O group of the amino acid four residues earlier ($i+4 \rightarrow t$).



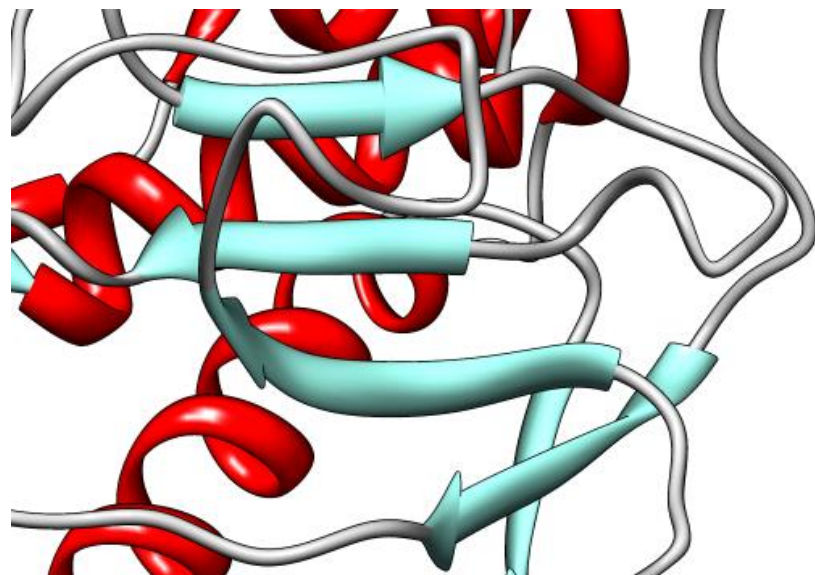
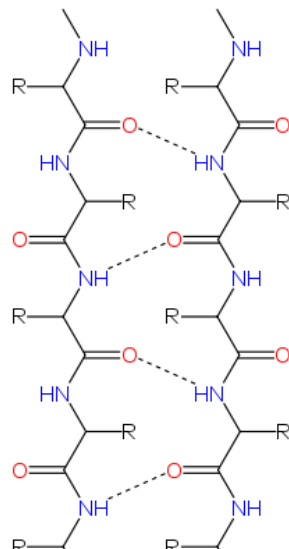
Proteins Representation: β -sheets

Beta sheet consists of **beta strands** connected laterally by at least two or three backbone hydrogen bonds, forming a generally twisted, pleated sheet. The majority of **beta strands** (typically 3 to 10 amino acids long) are arranged adjacent to other strands and form an extensive hydrogen bond network with their neighbors in which the N-H groups in the backbone of one strand establish hydrogen bonds with the C=O groups in the backbone of the adjacent strands.

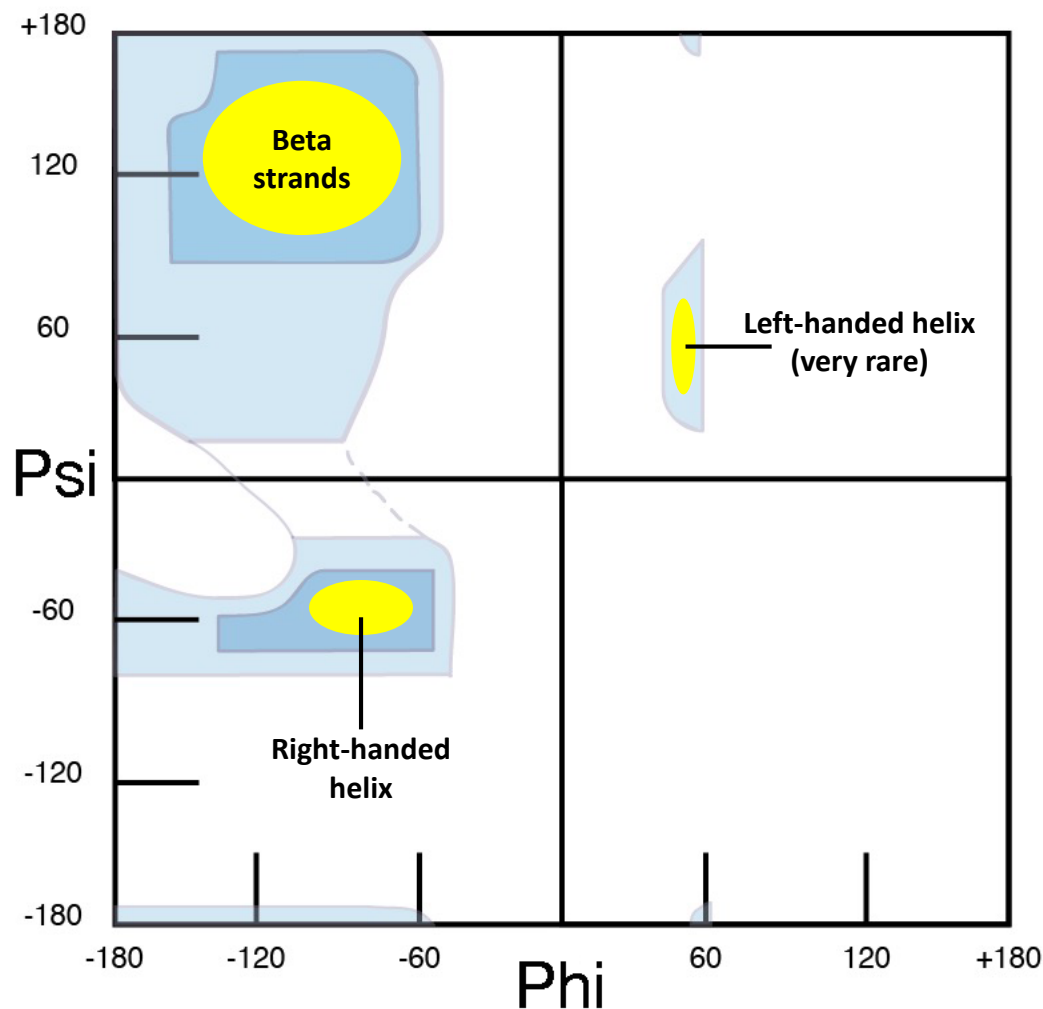
Antiparallel Beta sheet



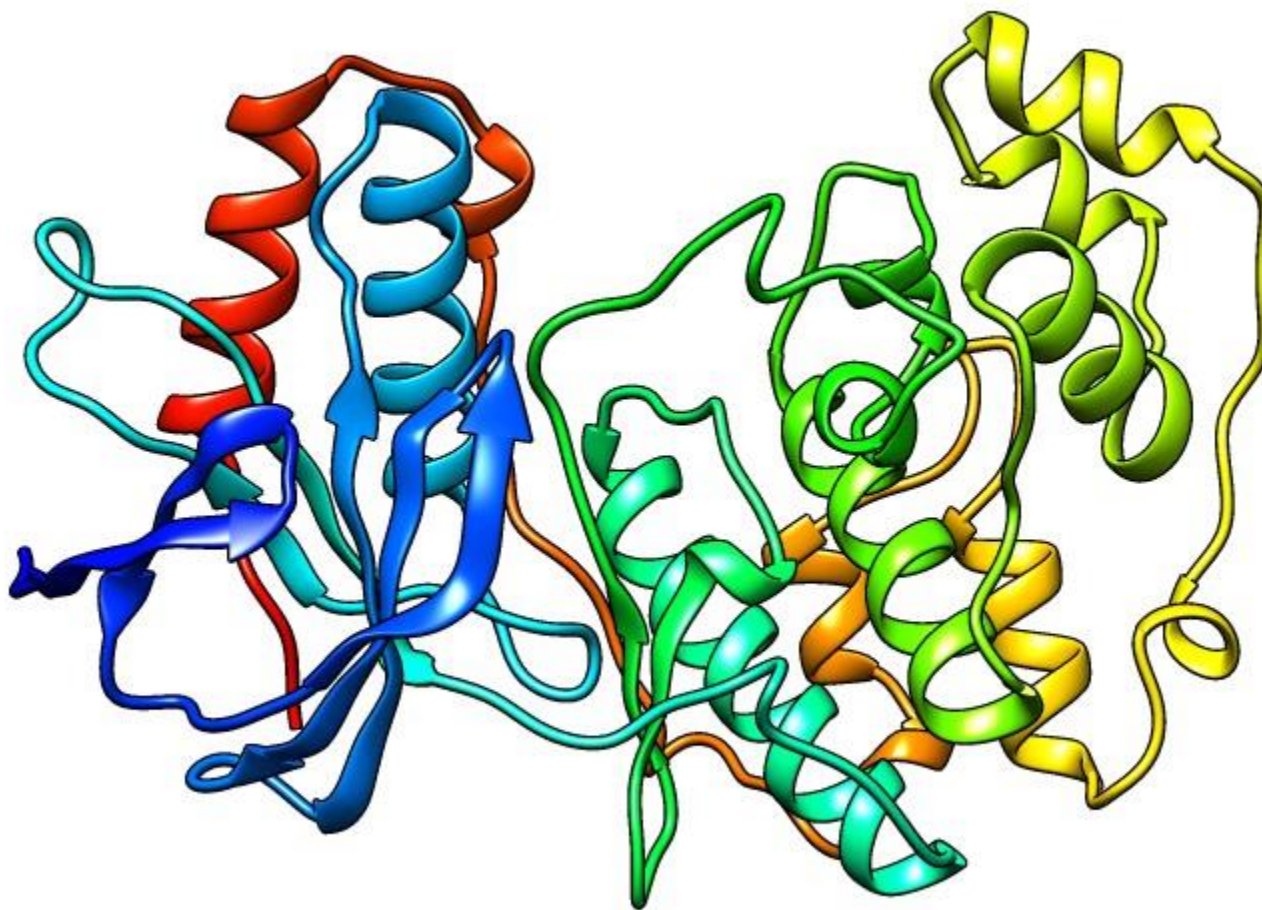
Parallel Beta sheet



Ramachandran Plot Labeled



Molecule format: PDB

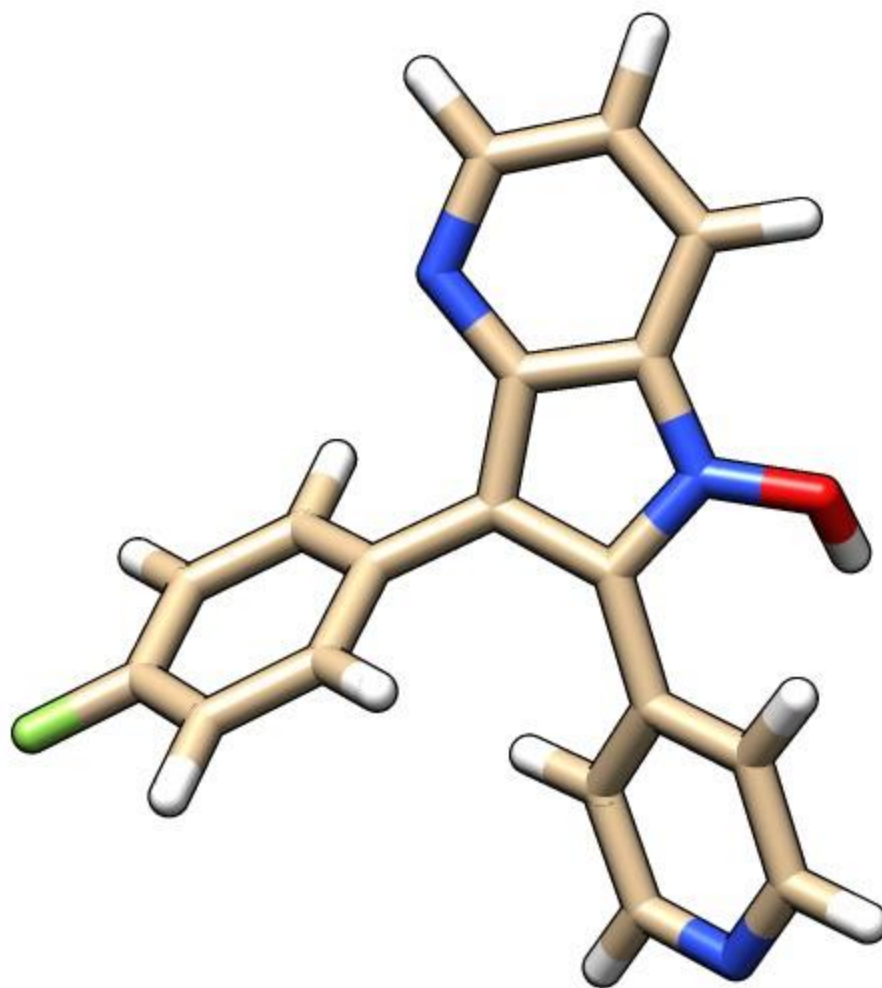


Molecule format: PDB

ATOM	1	N	ARG	A	5	-2.440	2.802	39.839	1.00	34.45	N
ATOM	2	CA	ARG	A	5	-1.393	3.612	40.550	1.00	34.02	C
ATOM	3	C	ARG	A	5	-0.474	2.718	41.392	1.00	33.04	C
ATOM	4	O	ARG	A	5	-0.945	1.957	42.215	1.00	34.00	O
ATOM	5	CB	ARG	A	5	-2.020	4.713	41.416	1.00	34.82	C
ATOM	6	CG	ARG	A	5	-1.029	5.810	41.901	1.00	35.95	C
ATOM	7	CD	ARG	A	5	-1.575	6.691	43.038	1.00	38.02	C
ATOM	8	NE	ARG	A	5	-0.657	7.728	43.548	1.00	37.74	N
ATOM	9	CZ	ARG	A	5	0.091	7.610	44.656	1.00	38.04	C
ATOM	10	NH1	ARG	A	5	0.065	6.479	45.378	1.00	38.47	N
ATOM	11	NH2	ARG	A	5	0.878	8.619	45.048	1.00	35.72	N
ATOM	12	N	PRO	A	6	0.840	2.782	41.166	1.00	31.98	N
ATOM	13	CA	PRO	A	6	1.790	2.041	42.004	1.00	30.33	C
ATOM	14	C	PRO	A	6	1.926	2.648	43.394	1.00	29.20	C
ATOM	15	O	PRO	A	6	1.477	3.784	43.647	1.00	27.67	O
ATOM	16	CB	PRO	A	6	3.111	2.203	41.260	1.00	30.53	C
ATOM	17	CG	PRO	A	6	2.955	3.522	40.524	1.00	31.77	C
ATOM	18	CD	PRO	A	6	1.520	3.510	40.073	1.00	31.67	C
ATOM	19	N	THR	A	7	2.537	1.874	44.291	1.00	27.45	N
ATOM	20	CA	THR	A	7	2.900	2.386	45.608	1.00	27.35	C
ATOM	21	C	THR	A	7	4.261	3.097	45.524	1.00	26.19	C
ATOM	22	O	THR	A	7	5.198	2.609	44.897	1.00	26.31	O
ATOM	23	CB	THR	A	7	2.884	1.237	46.679	1.00	27.29	C
ATOM	24	OG1	THR	A	7	1.551	0.736	46.776	1.00	27.51	O
ATOM	25	CG2	THR	A	7	3.169	1.794	48.102	1.00	27.42	C
ATOM	26	N	PHE	A	8	4.341	4.269	46.129	1.00	26.06	N
ATOM	27	CA	PHE	A	8	5.581	5.017	46.237	1.00	26.66	C
ATOM	28	C	PHE	A	8	6.152	4.860	47.639	1.00	26.92	C
ATOM	29	O	PHE	A	8	5.405	4.740	48.590	1.00	27.23	O
ATOM	30	CB	PHE	A	8	5.333	6.519	46.018	1.00	26.78	C
ATOM	31	CG	PHE	A	8	4.880	6.891	44.630	1.00	26.11	C
ATOM	32	CD1	PHE	A	8	3.622	6.506	44.145	1.00	27.62	C
ATOM	33	CD2	PHE	A	8	5.691	7.691	43.830	1.00	24.50	C
ATOM	34	CE1	PHE	A	8	3.210	6.870	42.857	1.00	26.15	C
ATOM	35	CE2	PHE	A	8	5.302	8.062	42.562	1.00	25.15	C
ATOM	36	CZ	PHE	A	8	4.056	7.658	42.058	1.00	25.39	C

ATOM/HETATM

Molecule format: MOL2



Molecule format: MOL2

@<TRIPOS>MOLECULE

loz1

35 38 1 0 0

PROTEIN

AMBER ff14SB

@<TRIPOS>ATOM

1	O12	22.9580	13.8160	33.2690	O.3	1	FPH	-0.4462
2	N07	21.6960	14.0160	33.0600	N.ar	1	FPH	0.0275
3	C01	21.1480	15.2740	33.0560	C.ar	1	FPH	-0.1626
4	C06	21.6920	16.5690	33.2740	C.ar	1	FPH	-0.0710
5	C05	20.7940	17.6450	33.2060	C.ar	1	FPH	-0.2353
6	C04	19.4490	17.3950	32.9130	C.ar	1	FPH	0.3872
7	N03	18.9110	16.1640	32.6780	N.ar	1	FPH	-0.6260
8	C02	19.7840	15.0960	32.7640	C.ar	1	FPH	0.3193
9	C08	19.5260	13.6820	32.5980	C.ar	1	FPH	-0.0162
10	C24	18.2190	13.0650	32.3100	C.ar	1	FPH	-0.0548
11	C25	17.7200	12.0650	33.1830	C.ar	1	FPH	-0.0780
12	C26	16.4790	11.4420	32.9480	C.ar	1	FPH	-0.1720
13	C27	15.7220	11.8500	31.8500	C.ar	1	FPH	0.1329
14	F32	14.5350	11.2560	31.6140	F	1	FPH	-0.1399
15	C28	16.1900	12.8430	30.9700	C.ar	1	FPH	-0.1720
16	C29	17.4320	13.4490	31.2020	C.ar	1	FPH	-0.0780
17	C13	20.7700	13.0000	32.7920	C.ar	1	FPH	0.0008
18	C14	21.1970	11.5670	32.7220	C.ar	1	FPH	-0.0118
19	C19	22.2090	11.0690	33.5820	C.ar	1	FPH	-0.2228
20	C18	22.6200	9.7380	33.4590	C.ar	1	FPH	0.3882
21	N17	22.0850	8.8740	32.5420	N.ar	1	FPH	-0.6550
22	C16	21.1230	9.3570	31.7230	C.ar	1	FPH	0.3882
23	C15	20.6600	10.6740	31.7710	C.ar	1	FPH	-0.2228
24	H06	22.7416	16.7170	33.4811	H	1	FPH	0.1500
25	H05	21.1371	18.6545	33.3779	H	1	FPH	0.1470
26	H04	18.7837	18.2446	32.8694	H	1	FPH	0.0261
27	H25	18.3030	11.7756	34.0448	H	1	FPH	0.1480
28	H26	16.1205	10.6635	33.6052	H	1	FPH	0.1510
29	H28	15.5941	13.1372	30.1187	H	1	FPH	0.1510
30	H29	17.7896	14.2145	30.5293	H	1	FPH	0.1480
31	H19	22.6579	11.7117	34.3248	H	1	FPH	0.1575
32	H18	23.3951	9.3751	34.1178	H	1	FPH	0.0261
33	H16	20.6899	8.6890	30.9932	H	1	FPH	0.0261
34	H15	19.8956	11.0066	31.0843	H	1	FPH	0.1575
35	H12	23.2990	13.2073	32.6096	H	1	FPH	0.4300

@<TRIPOS>BOND

@<TRIPOS>BOND

1	1	2	1
2	2	3	ar
3	2	17	ar
4	3	4	ar
5	3	8	ar
6	4	5	ar
7	5	6	ar
8	6	7	ar
9	7	8	ar
10	8	9	ar
11	9	10	1
12	9	17	ar
13	10	11	ar
14	10	16	ar
15	11	12	ar
16	12	13	ar
17	13	14	1
18	13	15	ar
19	15	16	ar
20	17	18	1
21	18	19	ar
22	18	23	ar
23	19	20	ar
24	20	21	ar
25	21	22	ar
26	22	23	ar
27	24	4	1
28	25	5	1
29	26	6	1
30	27	11	1
31	28	12	1
32	29	15	1
33	30	16	1
34	31	19	1
35	32	20	1
36	33	22	1
37	34	23	1
38	35	1	1

@<TRIPOS>SUBSTRUCTURE

1 FPH 1 RESIDUE

4 A

FPH

0 ROOT

Protein Data Bank

www.rcsb.org

RCSB PDB Deposit ▾ Search ▾ Visualize ▾ Analyze ▾ Download ▾ Learn ▾ About ▾ Documentation ▾ Careers COVID-19 [MyPDB ▾](#) [Contact us](#)

RCSB PDB
PROTEIN DATA BANK

205,133 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

▾ 3D Structures ? ☐ Include CSM ?

[Advanced Search](#) | [Browse Annotations](#) [Help](#)

New: More Computed Structure Models (CSM) available [Learn more](#)

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

- Experimentally-determined 3D structures from the **Protein Data Bank (PDB)** archive
- Computed Structure Models (CSM)** from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

Explore NEW Features

PDB-101 Training Resources

May Molecule of the Month

Cellulases and Bioenergy

IN SUPREME DIGNITAS
1343

Department of Pharmacy
Molecular Modeling & Virtual Screening Laboratory

MMVSL
Molecular Modeling & Virtual Screening Laboratory

Protein Data Bank

If you insert for example the 1YOU PDB code you will obtain the following results:

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment Literature

Biological Assembly 1

1YOU

Crystal structure of the catalytic domain of MMP-13 complexed with a potent pyrimidinetrione inhibitor

DOI: 10.2210/pdb1you/pdb

Classification: [HYDROLASE](#)

Deposited: 2005-01-28 Released: 2005-03-15

Deposition author(s): [Pandit, J.](#)

Organism: [Homo sapiens](#)

Expression System: Escherichia coli

Structural Biology Knowledgebase: 1YOU (4 models >24 annotations) [SbKB.org](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.3 Å

R-Value Free: 0.295

R-Value Work: 0.208

wwPDB Validation

Full Report

Metric	Percentile Ranks	Value
Rfree		0.293
Clashscore		8
Ramachandran outliers		1.8%
Sidechain outliers		7.2%
RSRZ outliers		6.3%

Legend:
■ Percentile relative to all X-ray structures
□ Percentile relative to X-ray structures of similar resolution

Literature

Download Primary Citation

Potent pyrimidinetrione-based inhibitors of MMP-13 with enhanced selectivity over MMP-14.

[Blagg, J.A.](#), [Hoe, M.C.](#), [Wolf-Gouveia, L.A.](#), [Reiter, L.A.](#), [Laird, E.R.](#), [Chang, S.P.](#), [Danley, D.E.](#), [Downs, J.T.](#), [Elliott, N.C.](#), [Eskra, J.D.](#), [Griffiths, R.J.](#), [Hardink, J.R.](#), [Haugeto, A.J.](#), [Jones, C.S.](#), [Liras, J.L.](#), [Lopresti-Morrow, L.L.](#), [Mitchell, P.G.](#), [Pandit, J.](#), [Robinson, R.P.](#), [Subramanyam, C.](#), [Vaughn-Bowser, M.L.](#), [Yocum, S.A.](#)

(2005) *Bioorg. Med. Chem. Lett.* 15: 1807-1810

PubMed: 15780611 [Search on PubMed](#)

DOI: 10.1016/j.bmcl.2005.02.038

PubMed Abstract:

Through the use of computational modeling, a series of pyrimidinetrione-based inhibitors of MMP-13 was designed based on a lead inhibitor identified through file screening. Incorporation of a biaryl ether moiety at the C-5 position of the pyrimidinetrione ring resulted in

X-ray Resolution



Protein Data Bank

Macromolecules

Find similar proteins by: Sequence ▾ (by identity cutoff) | [Structure](#)

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details
Collagenase 3	A, B	168	Homo sapiens	Mutation(s): 0 Gene Names: MMP13 EC: 3.4.24

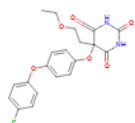
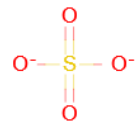


Find proteins for [P45452](#) (*Homo sapiens*)

Explore [P45452](#)

Protein Data Bank

Small Molecules

Ligands 4 Unique

ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
PFD Query on PFD Download CCD File	A, B	5-(2-ETHOXYETHYL)-5-[4-(4-FLUOROPHENOXY)PHENOXY]PYRIMIDINE-2,4,6(1H,3H,5H)-TRIONE <chem>C20 H19 F N2 O6</chem> XRSYNYGEEYTXJV-UHFFFAOYSA-N		Ligand Interaction
SO4 Query on SO4 Download CCD File	A, B	SULFATE ION <chem>O4 S</chem> QAOWNCQODCNURD-UHFFFAOYSA-L		Ligand Interaction
ZN Query on ZN Download CCD File	A, B	ZINC ION <chem>Zn</chem> PTFCDOFLOPIGGS-UHFFFAOYSA-N		Ligand Interaction
CA Query on CA Download CCD File	A, B	CALCIUM ION <chem>Ca</chem> BHPQYMZQTOCNFJ-UHFFFAOYSA-N		Ligand Interaction

External Ligand Annotations

ID	Binding Affinity (Sequence Identity %)
PFD	IC50: 0.8700000047683716 nM Binding MOAD
PFD	IC50 : 0.8700000047683716 nM PDBBind
PFD	IC50: 0.6000000238418579 nM BindingDB

Department of Pharmacy

Molecular Modeling & Virtual Screening Laboratory