

Bioinformatics 2: Molecular Modeling

Lecture 1: the amino acids

Administrative stuff

- Set office hours.
- Read syllabus.
- Install MOE, starting now.
- 3-button mouse.
- Course website:
<http://www.bioinfo.rpi.edu/bystrc/courses/biol4550/biol4550.html>

1.1 Molecular modeling

Molecular modeling

- The use of stereochemical and energetic information to predict conformations, interactions and dynamic behavior of a chemical system.

Are you a biologist or a chemist?

**Q: What determines the structure/
function of a protein?**

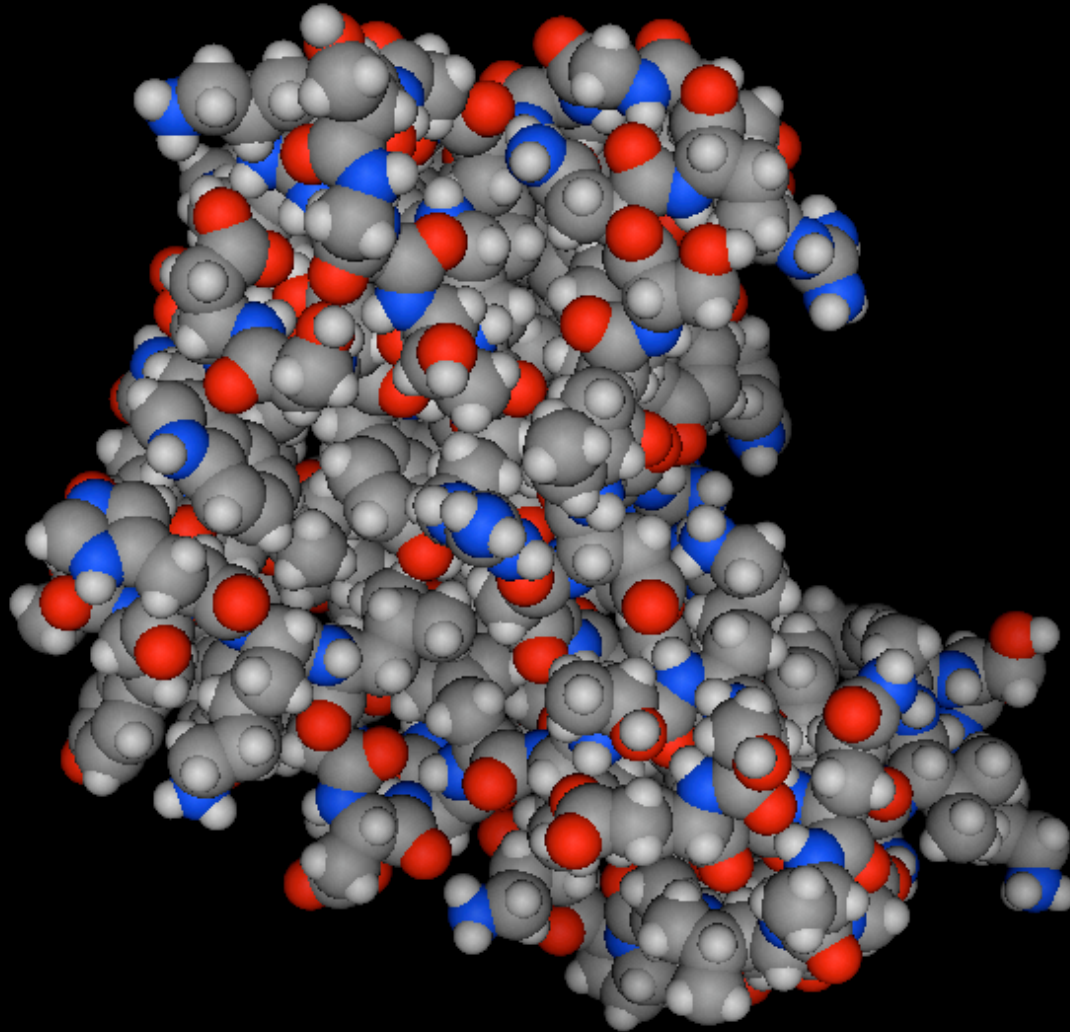
- 1. Evolution** -- *protein sequences are selected for function.*
- 2. Energy** -- *biochemical systems seek the lowest free energy state.*

What a protein looks like... biologist's point of view.

**MSAIQASWPSGTECIAKYNFHGTAEQDLPFC
KGDVLTIVAVTKDPNWKAKNKVGREGIIPA
NYVQKREGV**

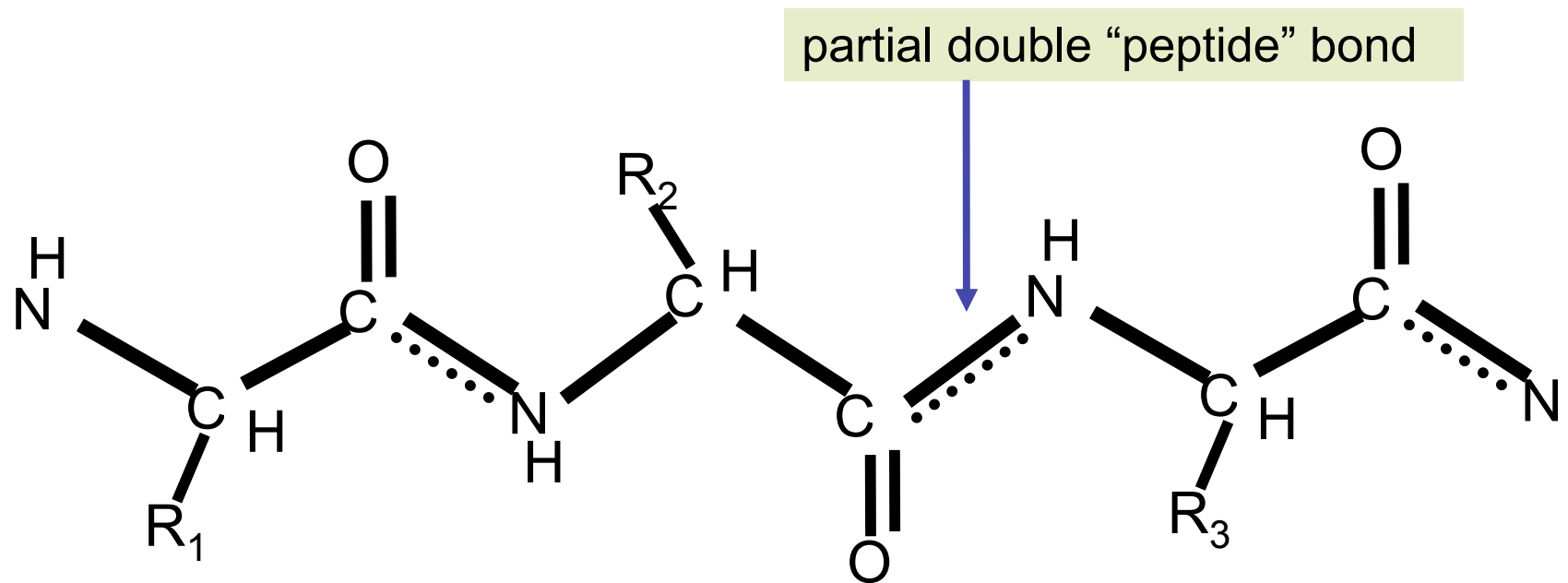
This is all the information the cell needs need to make a 3D protein. Add water and you get a unique 3D structure.

The biochemist's view of a protein



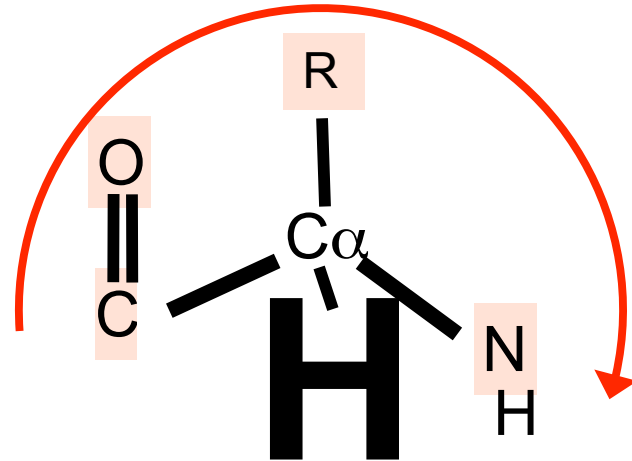
1.2 Drawing amino acids

the backbone atoms



Backbone atom names are "N", "C-alpha" and "C" (or N,CA,C). Oxygen "O" is also considered a backbone atom, though strictly speaking it is a side chain. All atoms in all amino acids have conventional names.

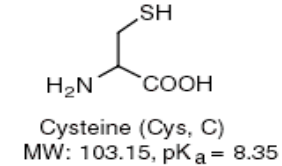
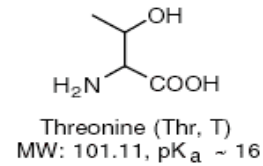
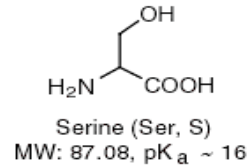
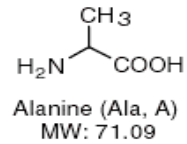
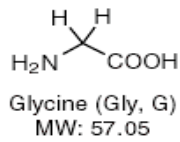
How to check chirality.



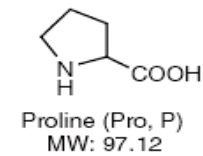
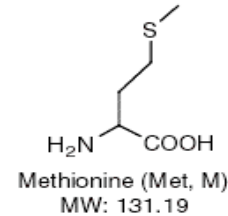
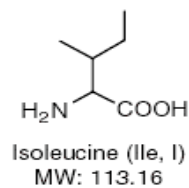
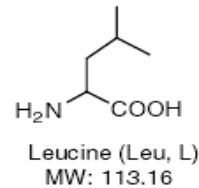
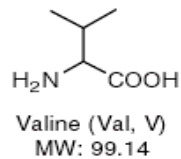
When an L-amino acid is drawn with the alpha-H forward and the R-group in the back, the letters read clockwise spell “CORN”. The “Corn Crib” is a good way to remember which side the R-group (i.e. sidechain) goes on.

the sidechains

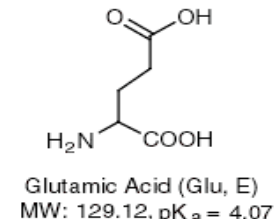
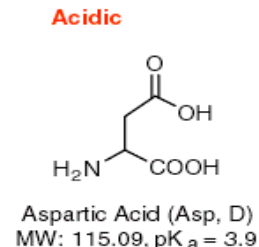
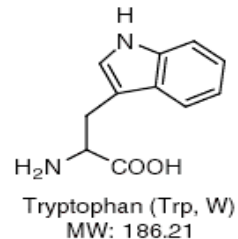
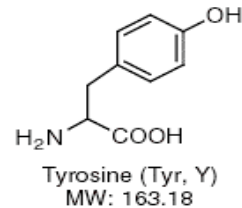
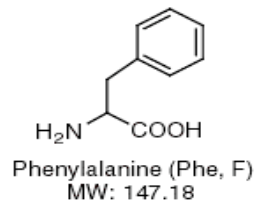
Small



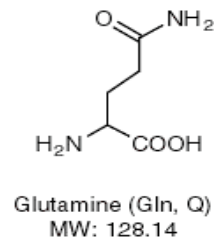
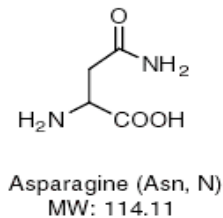
Hydrophobic



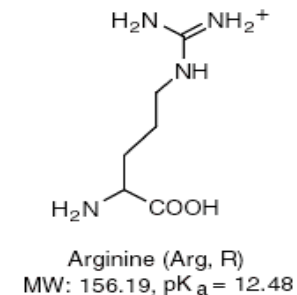
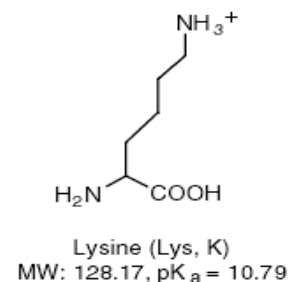
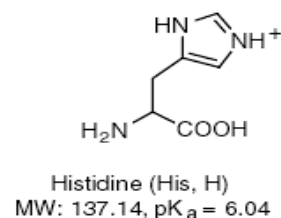
Aromatic



Amide

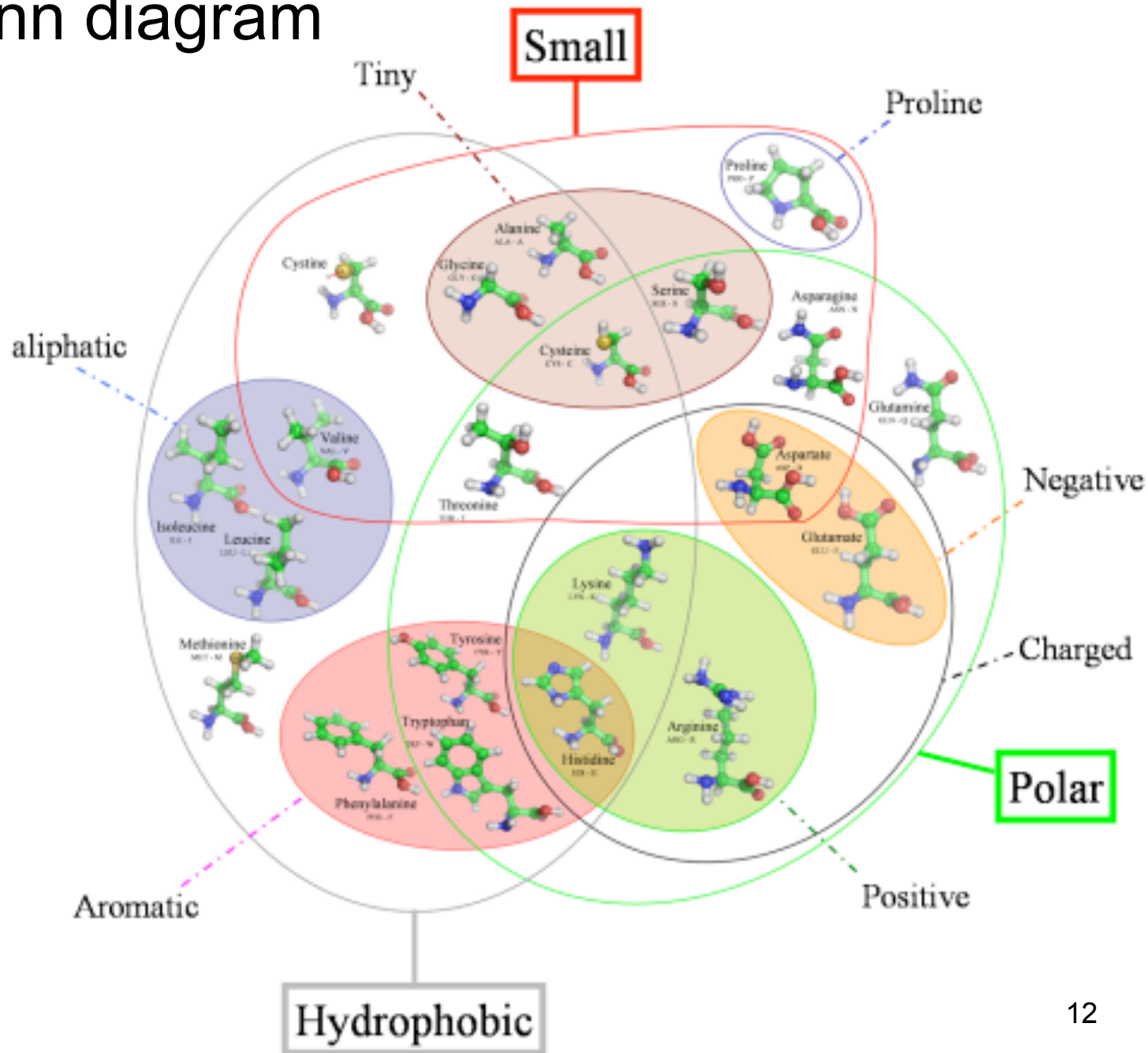


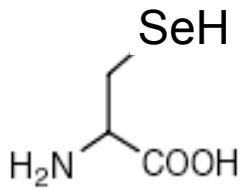
Basic



These chemical nature of these 20 side chains account for the folding and function of proteins.

Venn diagram



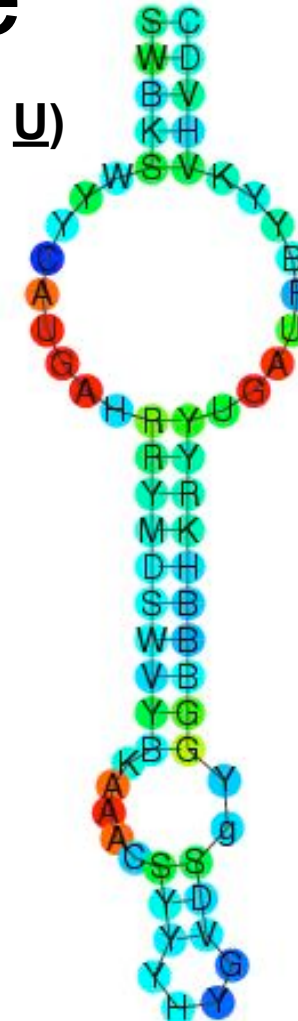


Cysteine (Cys, C)
MW: 103.15, pK_a = 8.35

Selenocysteine

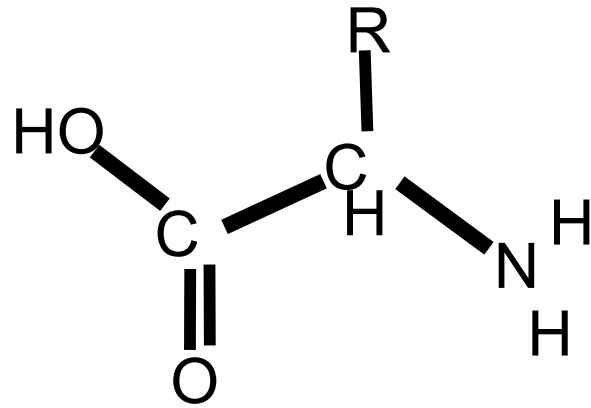
(3-letter code Sec, one-letter code U)

The presence of a selenocysteine insertion sequence, **SECIS**, signals the cell to translate UGA Stop codons with a special, modified tRNA. Initially charged with serine, Ser-tRNA(Sec) is then converted to Sec-tRNA(Sec) by selenocysteine synthase. Selenocysteine is not formed if no selenium (Se) is present, leading to translation stop. SECIS is located within the coding sequence (bacteria) or in the 3'-UTR (eukaryotes, archea).

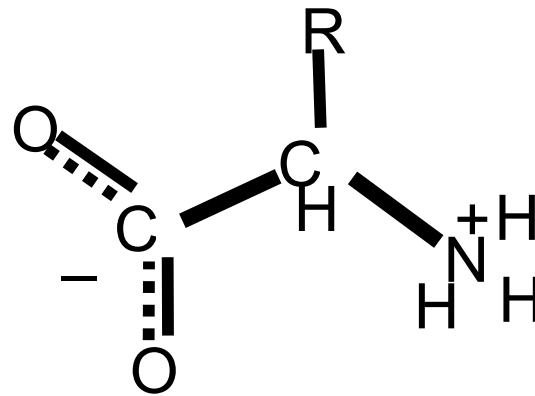


0 1
Sequence conservation

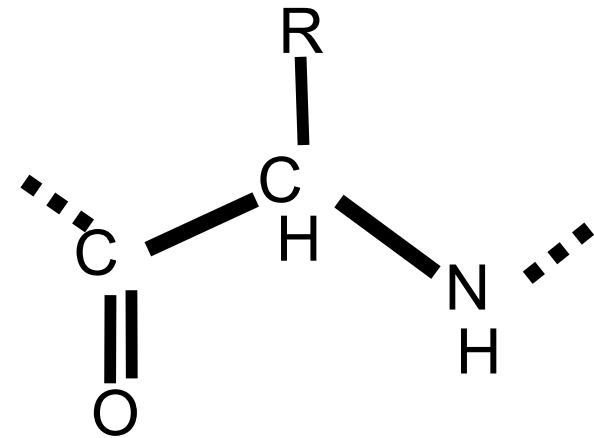
Free AA vs polypeptides



Free amino acid

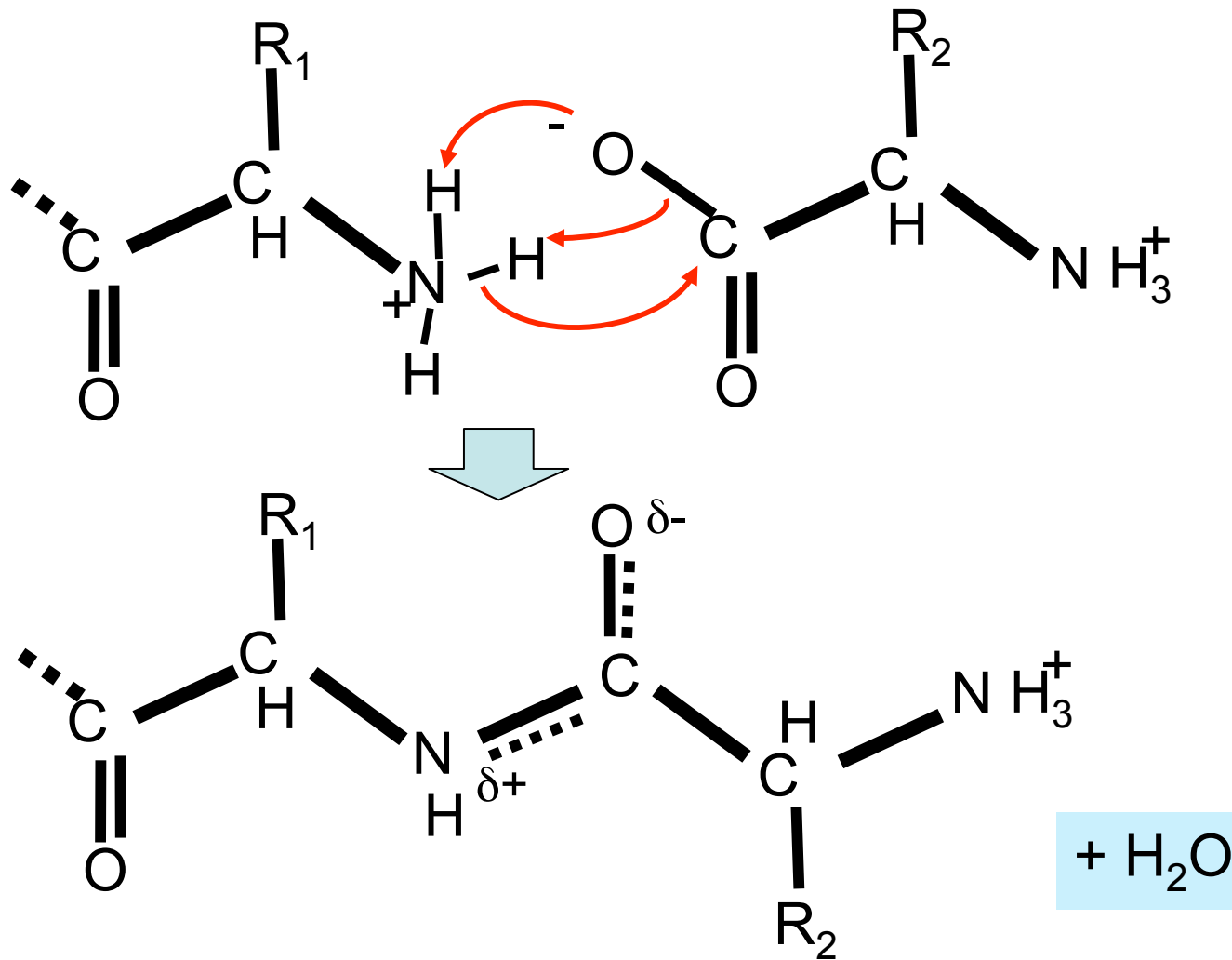


Zwitterionic
form, pH7

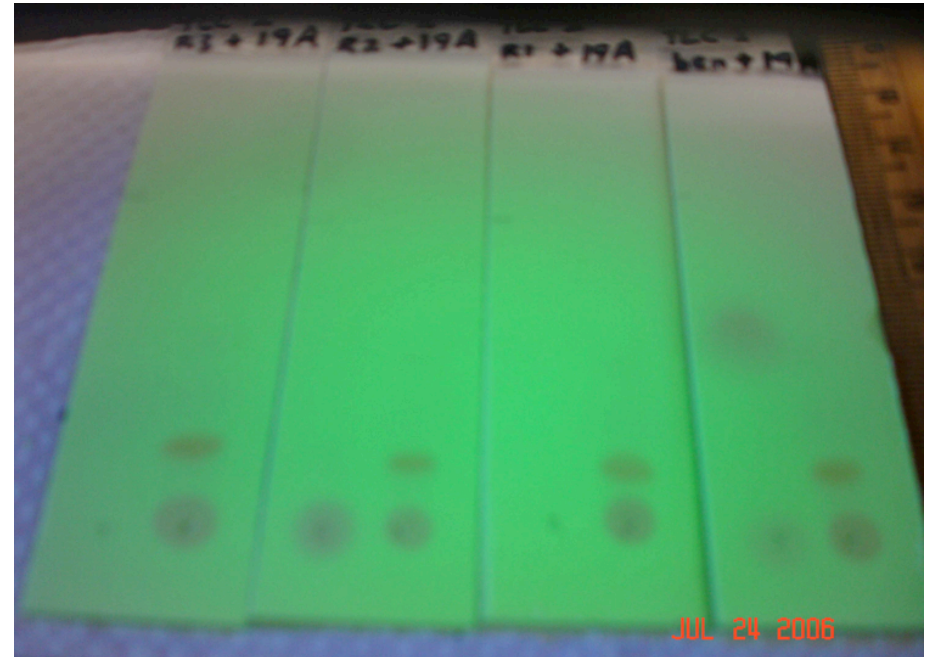
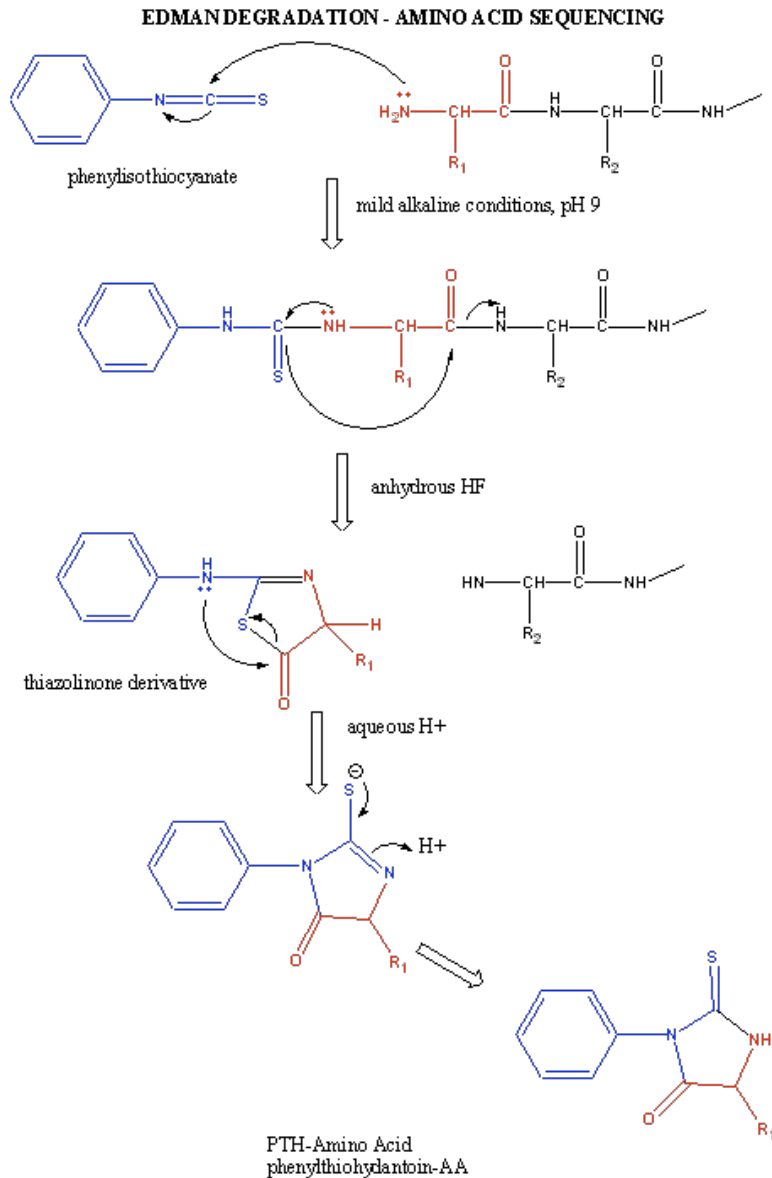


Polymer
repeating unit
("residue")

peptide bond formation is dehydrating



Why is it called a “residue”?



Thin layer chromatography used to identify the PTH-amino acid, product of the Edman degradation, a stain or “residue” on the plate.

The Protein Data Bank (PDB)

Go to www.pdb.org

Download "1CA2"

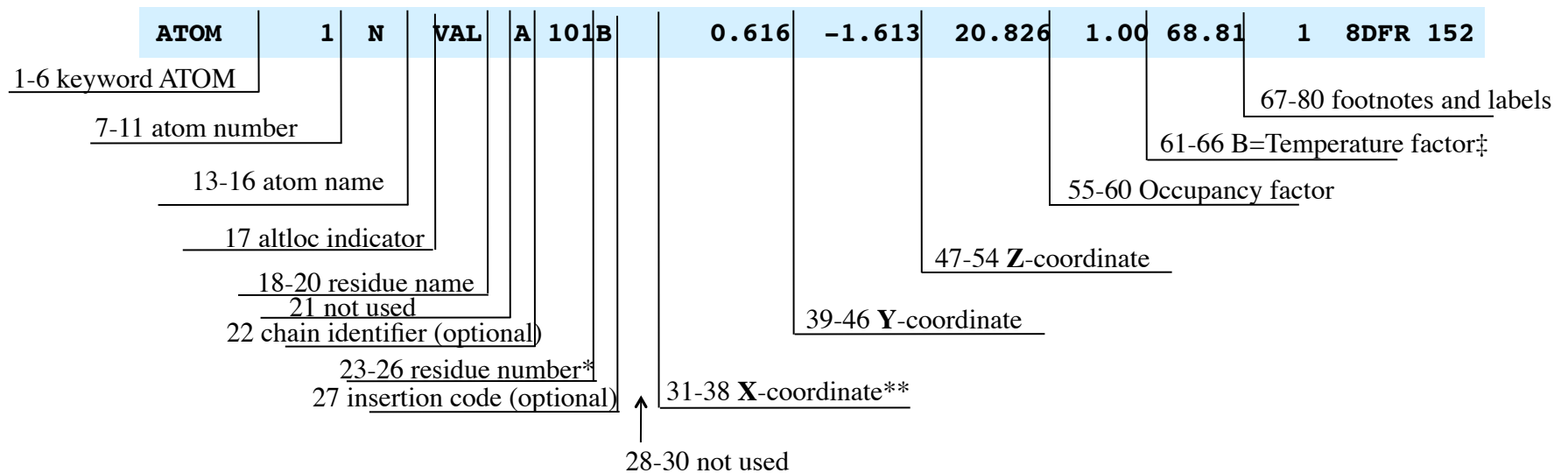
Open in text editor, or browser window

PDB file format. Keywords.

- HEADER, CMPND, REMARK : reference information.
- HET, FORMUL, HETNAM : ligands, non-standard groups.
- HELIX, SHEET, TURN : secondary structure elements.
- ATOM : coordinates, names, numbering.
- HETATM : coordinates, names, numbering, for HET groups.

- There is no *explicit* information about what atoms are bonded to what. (This is determined by distances and atom names.)
- No direct information about the formal or partial charges on atoms.

Keyword 'ATOM'



* Usually, but not always, residues are numbered sequentially 1,2,3 etc. Often the numbering starts from a number other than 1.

** Coordinates are in orthogonal angstroms by convention. May be converted to crystallographic coordinates using CRYST lines.

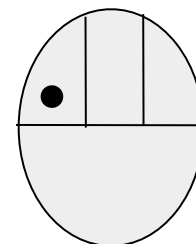
‡ Mean square displacement $\langle u^2 \rangle$ is proportional to B: $\langle u^2 \rangle = B/(8\pi^2)$

Download, install Jmol

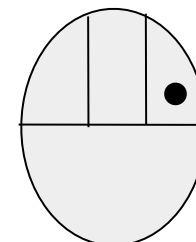
<http://jmol.sourceforge.net/>

Manual (very useful)

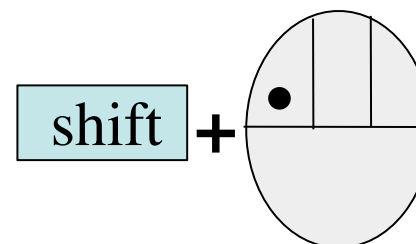
jmol.sourceforge.net/docs



rotate



translate



scale

Display an amino acid in Jmol

Download *1CA2.pdb* from rcsb.org

Open in Jmol

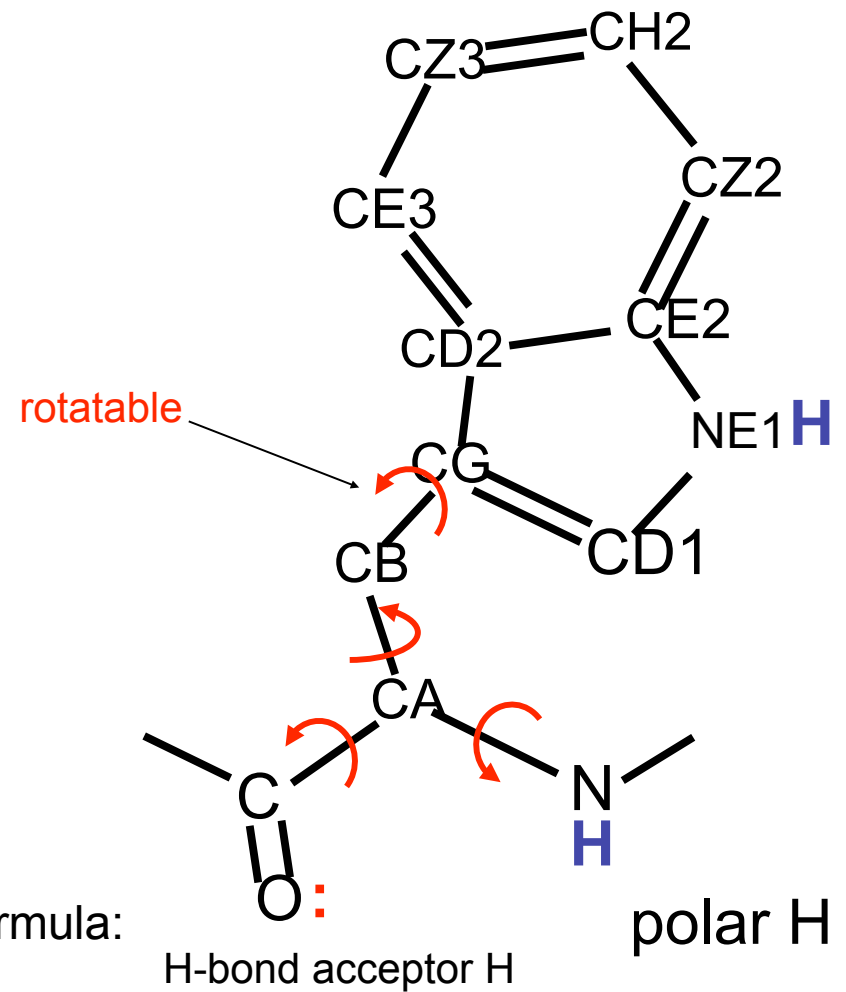
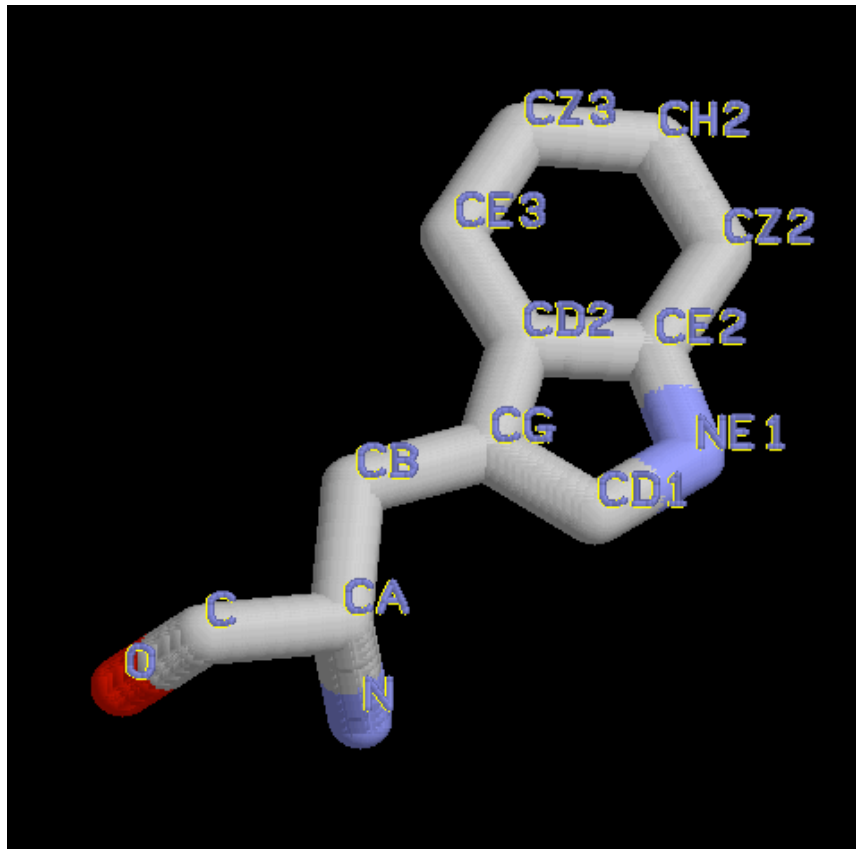
Click on one of the displayed residues to get its residue number

- `restrict nnn` (where *nnn* is the residue number)
- `wireframe 100`
- `label %a` (%a shows atoms names)

To return to whole protein:

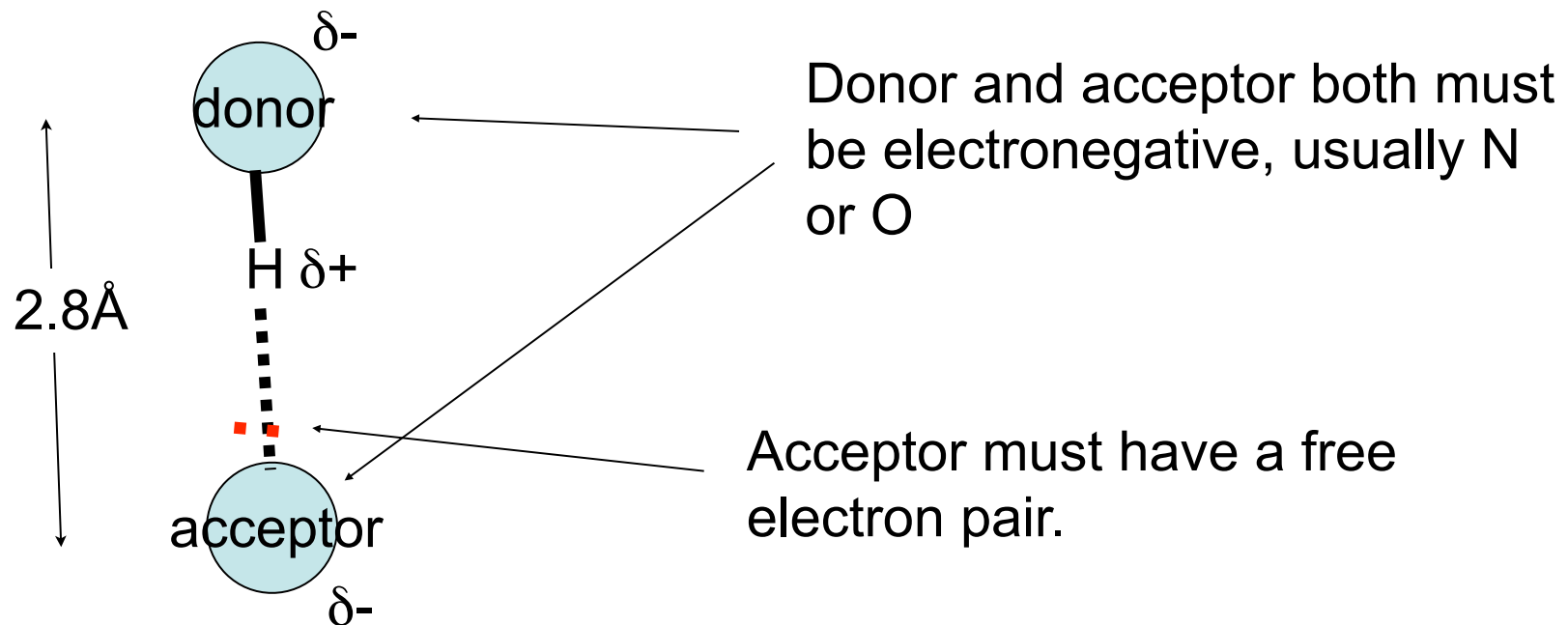
- `select protein`
- `wireframe 100`

tryptophan



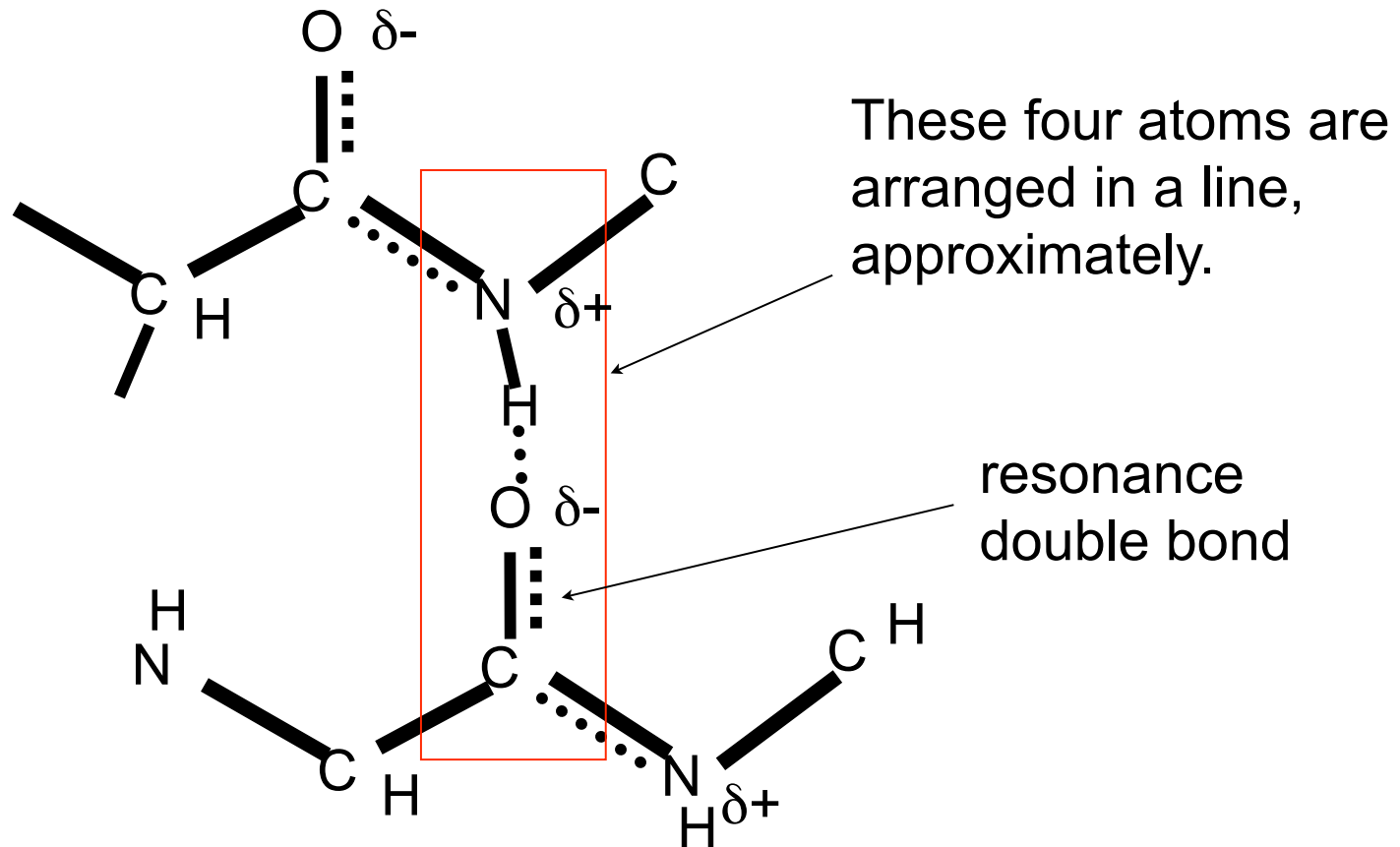
PDB convention atom names follow the formula:
<element><greek letter><alt posit>

Hydrogen bonds



Hydrogen bonds are a **linear** arrangement of three atoms, two electronegative (O or N) and an electropositive hydrogen in the middle. The atoms are closer together than expected for a “non-bonded” interaction, but not close enough for a covalent interaction. In that sense, they exist in the “limbo” world between covalent and non-covalent interactions.

Typical backbone H-bond

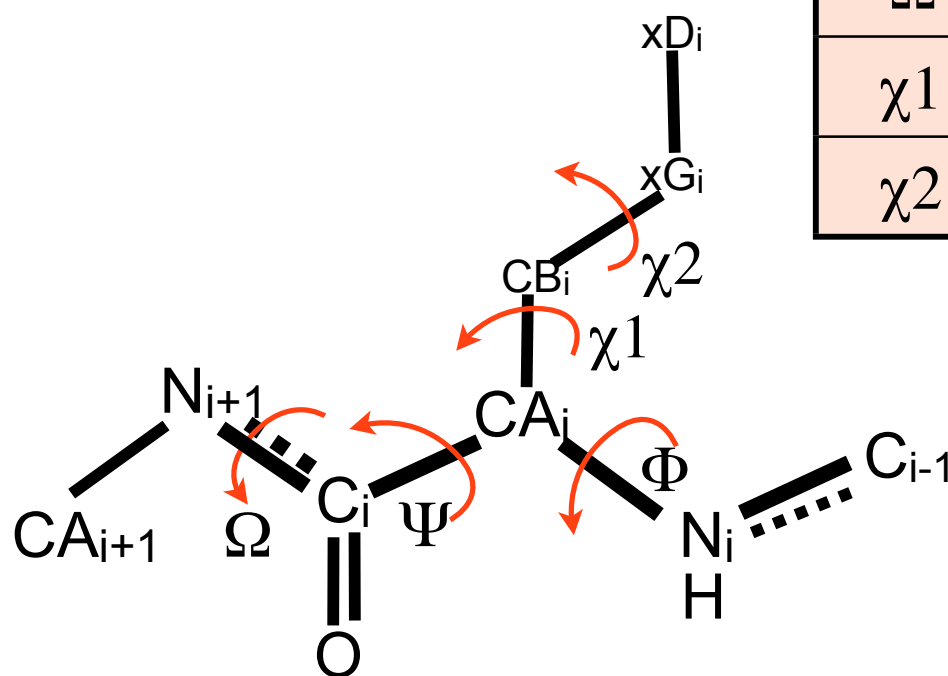


H-bonds between backbone atoms define secondary structure like buttons on a shirt, in the sense that **donors** match **acceptors** like buttons to button-holes. Only sometimes the shirt has buttons on alternating sides rather than all on one side.

1.3 Torsion angles

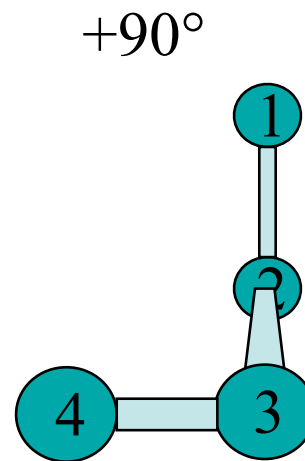
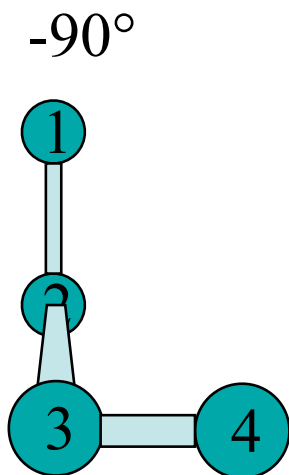
Side chain and backbone Torsion angles

Angle	atom1	atom2	atom3	atom4
Φ	C_{i-1}	N_i	CA_i	C_i
Ψ	N_i	CA_i	C_i	N_{i+1}
Ω	CA_i	C_i	N_{i+1}	CA_{i+1}
χ_1	C_i	CA_i	CB_i	xG_i
χ_2	CA_i	CB_i <td xG_i	xD_i	



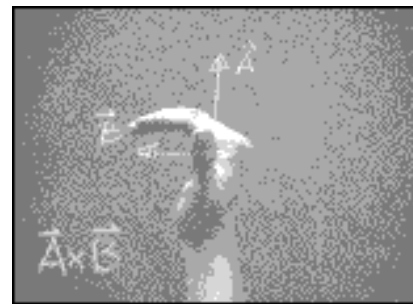
$\chi = \text{chi}$

Measuring a torsion angle by eye

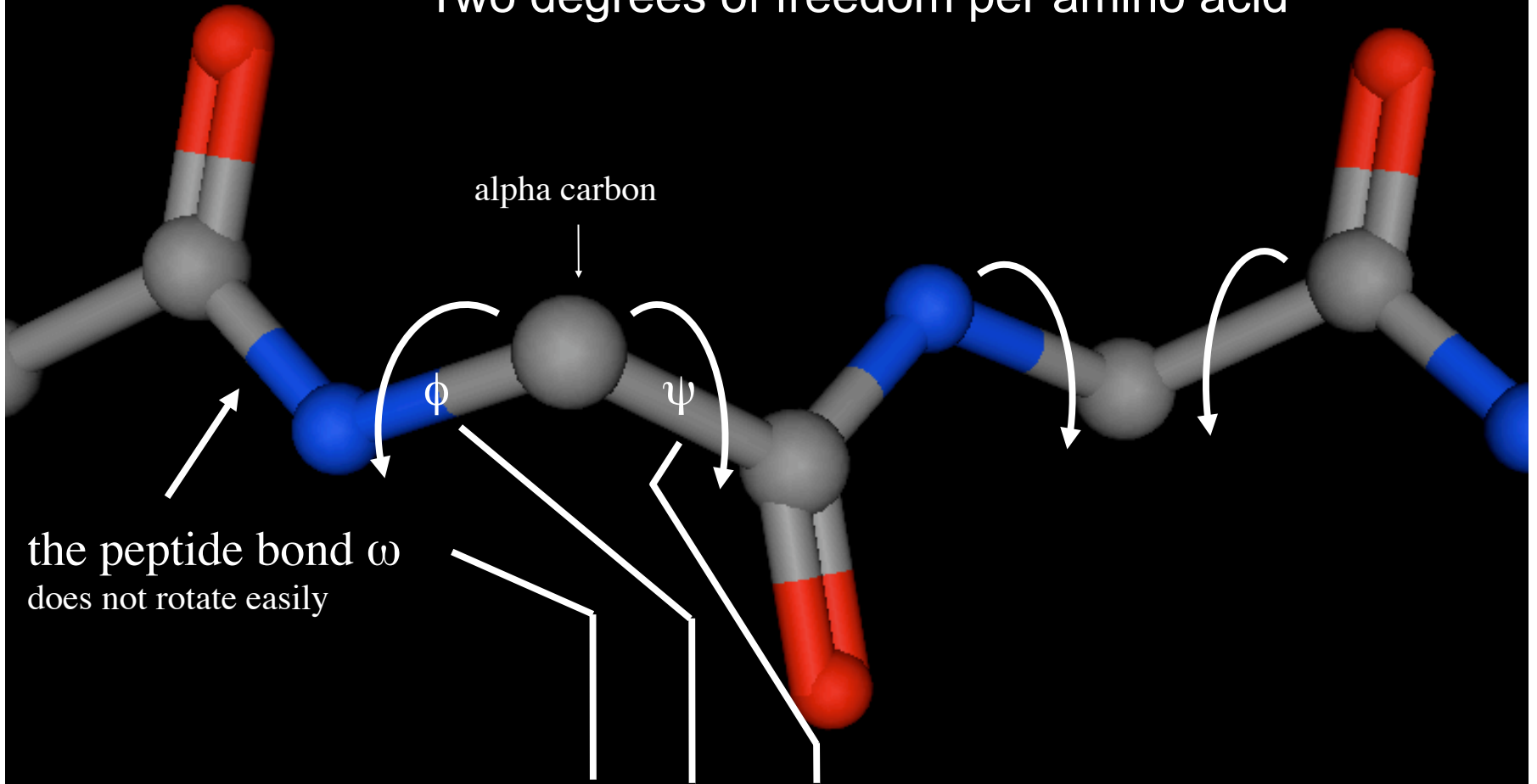


Use the right-hand rule

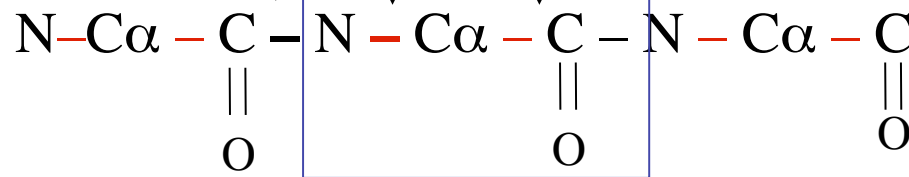
Positive torsion is in the direction of the fingers on the thumb-side of the rotation axis.



Two degrees of freedom per amino acid



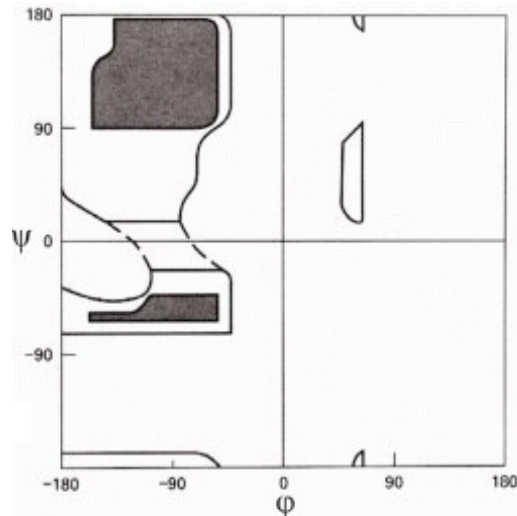
One amino acid



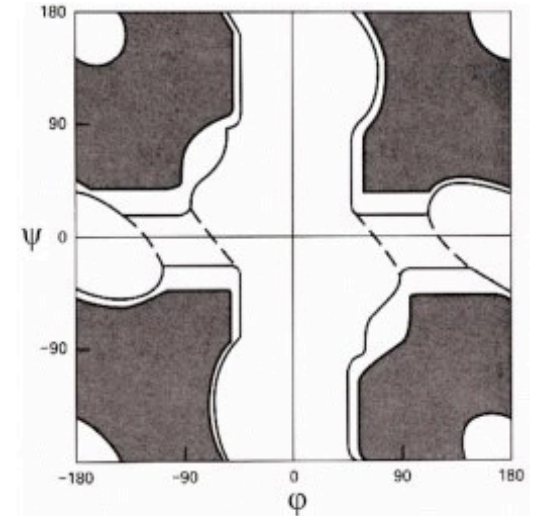
Ramachandran Plot maps allowable phi, psi regions

Ramachandran used a physical model of dipeptides to determine the allowed (dark) and disallowed (white) combinations of phi and psi backbone angles. The observed frequencies roughly agree with R's allowed regions.

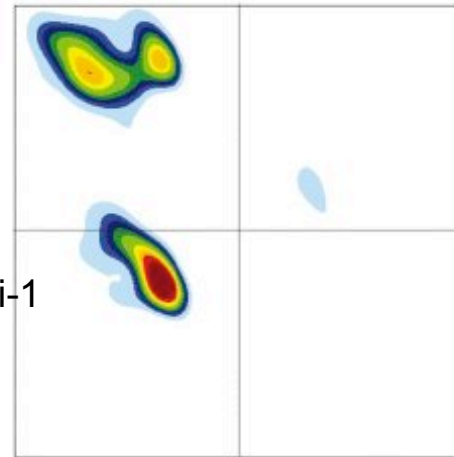
Ramachandran & Sasisekharan (1968)



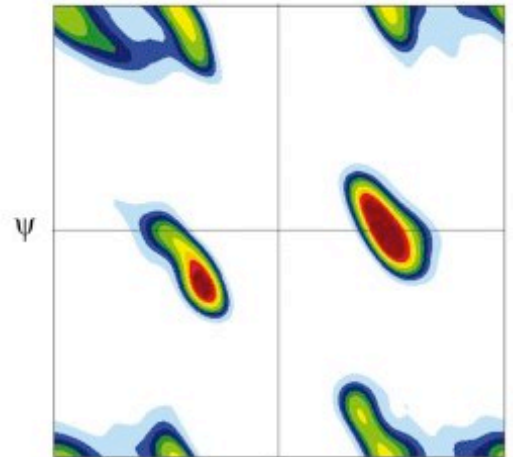
(a) non-glycine, non-proline



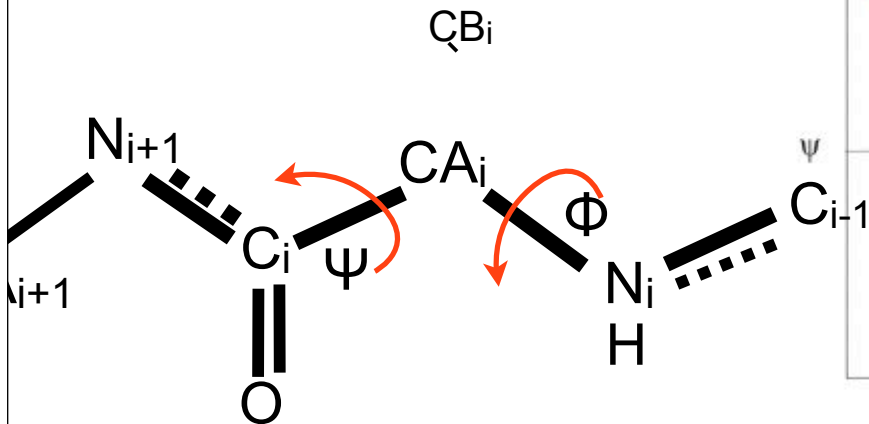
(b) glycine



(c)

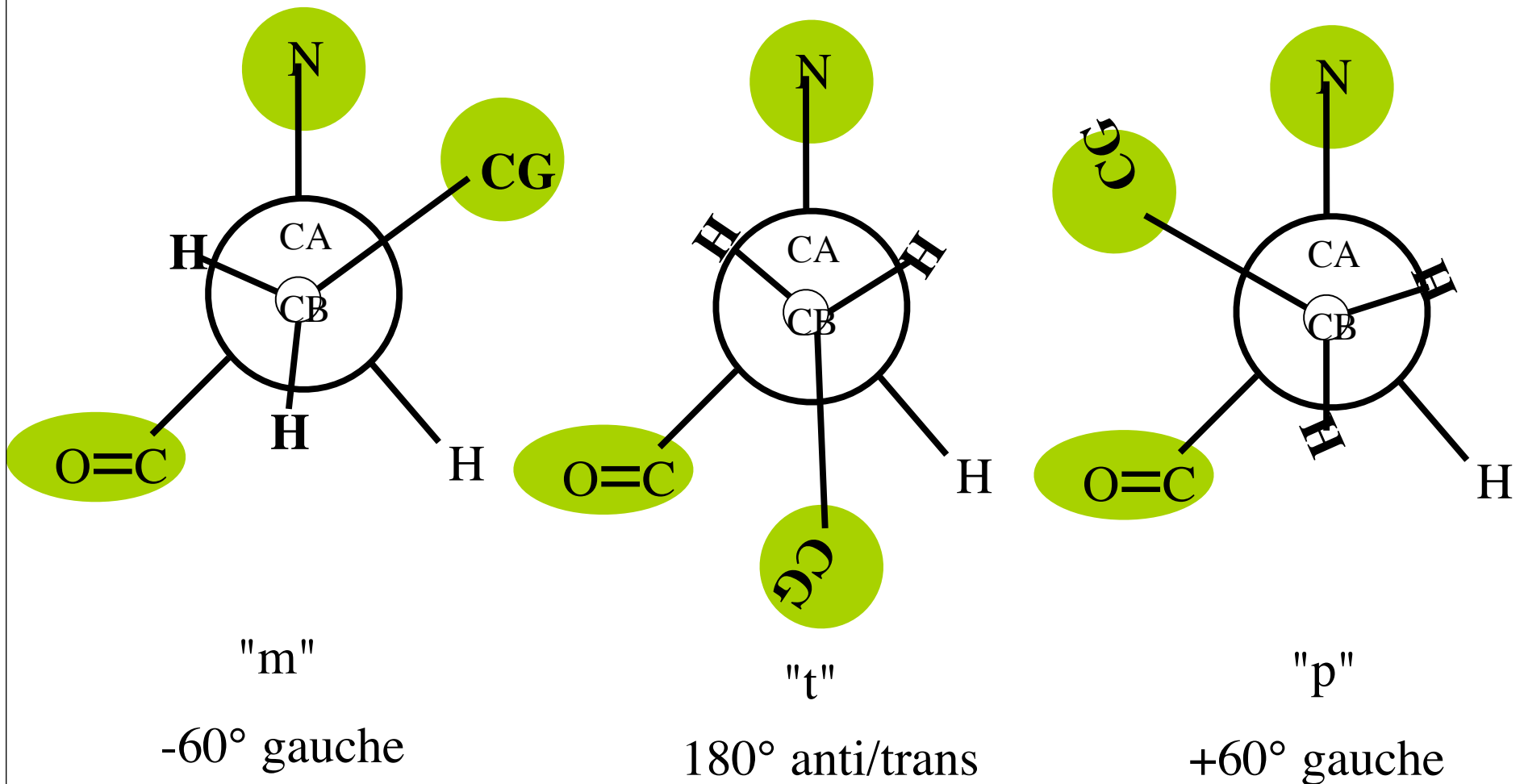


(d)



Sidechain rotamers

1-4 interactions differ greatly in energy depending on the moieties involved.

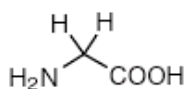


Exercise 1.1: Draw all 20 amino acids

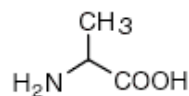
A C D E F G H I K L M N P Q R S T V W Y

- Draw using approximately correct bond angles
- Draw only polar hydrogens, H-bond donors.
- Draw a lone-pair “:” at each H-bond acceptor
- Circle chiral centers in Thr, Ile.
- Draw a curved arrow through each rotatable bond if it rotates at least one heavy atom (not H).
- **Turn this page in** at the end of class.

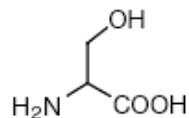
Small



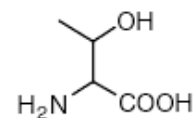
Glycine (Gly, G)
MW: 57.05



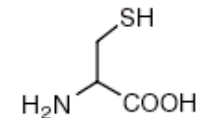
Alanine (Ala, A)
MW: 71.09



Serine (Ser, S)
MW: 87.08, pK_a ~ 16

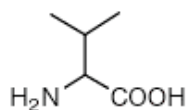


Threonine (Thr, T)
MW: 101.11, pK_a ~ 16

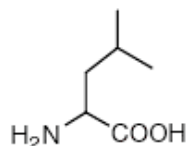


Cysteine (Cys, C)
MW: 103.15, pK_a = 8.35

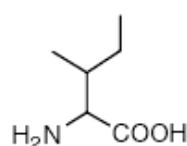
Hydrophobic



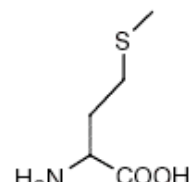
Valine (Val, V)
MW: 99.14



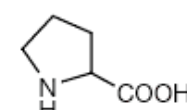
Leucine (Leu, L)
MW: 113.16



Isoleucine (Ile, I)
MW: 113.16

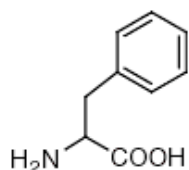


Methionine (Met, M)
MW: 131.19

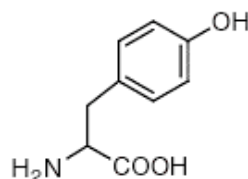


Proline (Pro, P)
MW: 97.12

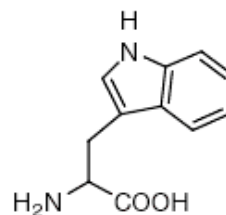
Aromatic



Phenylalanine (Phe, F)
MW: 147.18

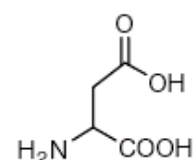


Tyrosine (Tyr, Y)
MW: 163.18

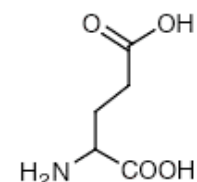


Tryptophan (Trp, W)
MW: 186.21

Acidic

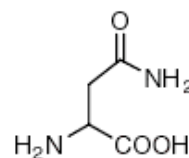


Aspartic Acid (Asp, D)
MW: 115.09, pK_a = 3.9

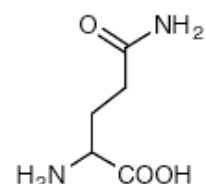


Glutamic Acid (Glu, E)
MW: 129.12, pK_a = 4.07

Amide

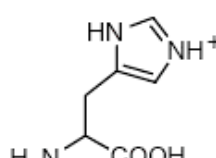


Asparagine (Asn, N)
MW: 114.11

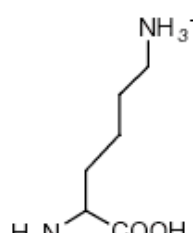


Glutamine (Gln, Q)
MW: 128.14

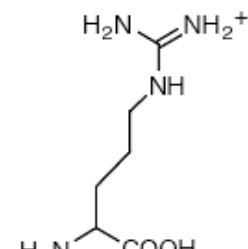
Basic



Histidine (His, H)
MW: 137.14, pK_a = 6.04



Lysine (Lys, K)
MW: 128.17, pK_a = 10.79



Arginine (Arg, R)
MW: 156.19, pK_a = 12.48

Further reading

- <http://www.johnkyrk.com/aminoacid.html>
- http://en.wikipedia.org/wiki/Secondary_structure
- http://en.wikipedia.org/wiki/Protein_structure
-



**CHEMICAL
COMPUTING
GROUP INC.**

INSTALL MOE 2011.10

This week

Get a 3-button mouse: Read Help->Mouse...

Familiarize yourself with the graphical user interface (GUI):

Help-->Tutorials-->Getting Started..., click on **GUI**

Run first part of **MoeTour** (Help-->Tutorials-->Getting Started)