#### MOLECULAR MODELING 2021

#### A course about proteins

protein conformation
protein shape
Protein energy
protein interactions
protein dynamics
protein drug interactions

#### Administrative

- Course website: on Blackboard learn: https://lms.rpi.edu/
- Office hour: 4 5 PM Wednesdays, on webex
- Grading policy
  - Assignment and quiz (30 % total)
  - midterm (20%)
  - final (20%)
  - Presentation (20%)
  - Class participation (answering questions in class) and attendance (10%)
- Questions ALWAYS welcome!

## Syllabus

## Installing MOE

Download MOE to your personal computer using file on RPI box

```
https://rpi.box.com/s/
mdjcw69wffo5roe5p7nse4n16wznf97u
```

!! If your RPI box account is not set up, use this link to set it up: <a href="https://itssc.rpi.edu/hc/en-us/articles/">https://itssc.rpi.edu/hc/en-us/articles/</a> 360004829931-Requesting-a-Box-RPI-Account

- Copy the license file (license.dat in the box directory) into the installed MOE local directory
- To run MOE, you need to access the license server located on RPI campus: got to have VPN running!
  - Install "cisco anyconnec"t: <a href="https://itssc.rpi.edu/hc/en-us/">https://itssc.rpi.edu/hc/en-us/</a>
     articles/360008783172-VPN-Installation-and-Connection

#### Instructor

- Chunyu Wang
- Associate Professor of Biological Sciences
- Expert in NMR, protein structure, Alzheimer's disease, Inteins and Hedgehog autoprocessing
- Hobbies: photography, astronomy, astrophotography, tennis, bridge, gardening, swimming etc.



#### NMR structure of Pab PollI intein: Hint fold

Cover of Nov.4th 2011 Issue of *JBC* 

Hint = Hedgehog Intein domain/fold

the journal of biological chemistry ASBMB AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOG

Du et.al. *J. Biol. Chem.*, 2011, 286, 38638-38648.

#### What are A $\beta$ , A $\beta$ 40 and A $\beta$ 42 ?

Aβ42 DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA

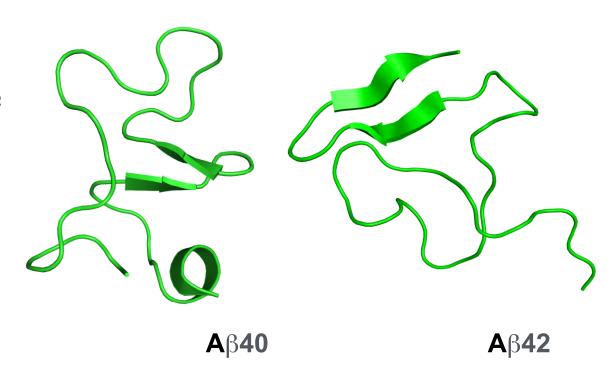
**A**β40 DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV

Major component of the amyloid plaque

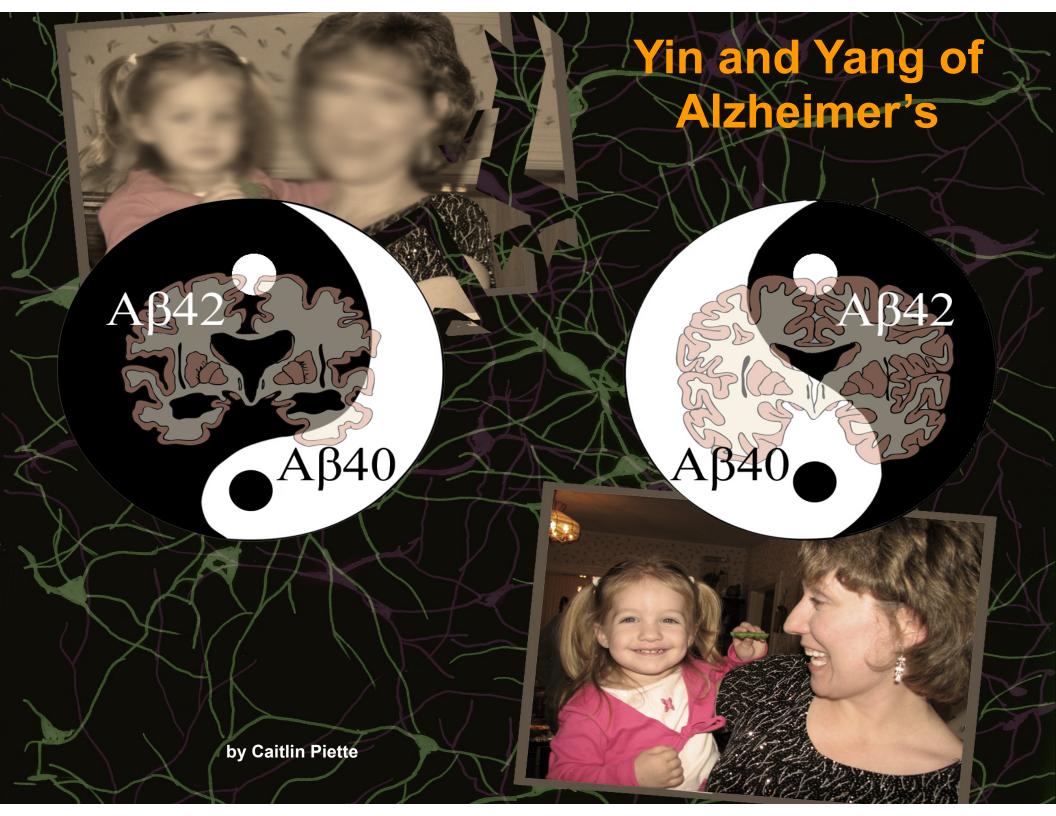
A $\beta$ 40 and A $\beta$ 42 are major species of A $\beta$ 

Aβ42 aggregates faster and is more toxic than Aβ40

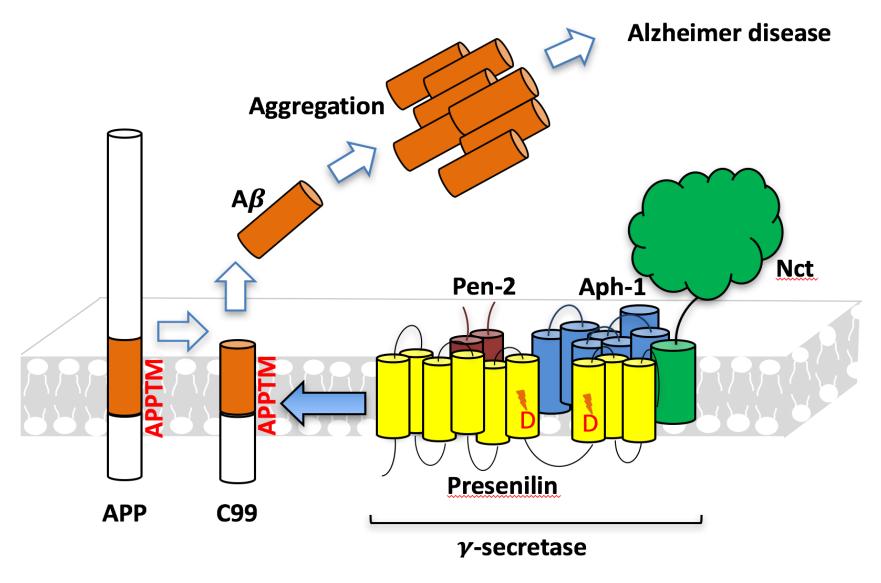
In familial Alzheimer's disease (FAD), Aβ42/Aβ40 ↑



Sgourakis et al. (2007) *JMB* 368:1448-1457

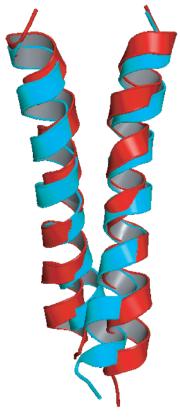


## Aβ generation and AD

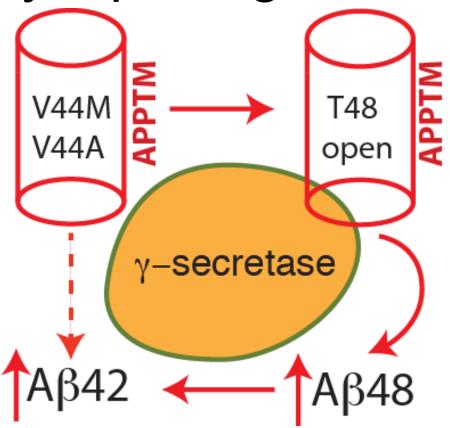


Presenilin homolog

# AD-Causing Mutations in APPTM Increase Aβ42 by Opening T48



NMR structures of WT APPTM (red) and V44M (blue), an AD-causing mutant.



Wen Chen, Eric Gamache, David Rosenman, Jian Xie, Maria Lopez, Yueming Li, and Chunyu Wang (2014). "Familial Alzheimer's Mutations within APPTM Increase Abeta42 Production by Enhancing the Accessibility of the Epsilon-Cleavage Site". Nature Communications 4:3037 doi: 10.1038/ncomms4037.

# Intro to Molecular Modeling and Proteins

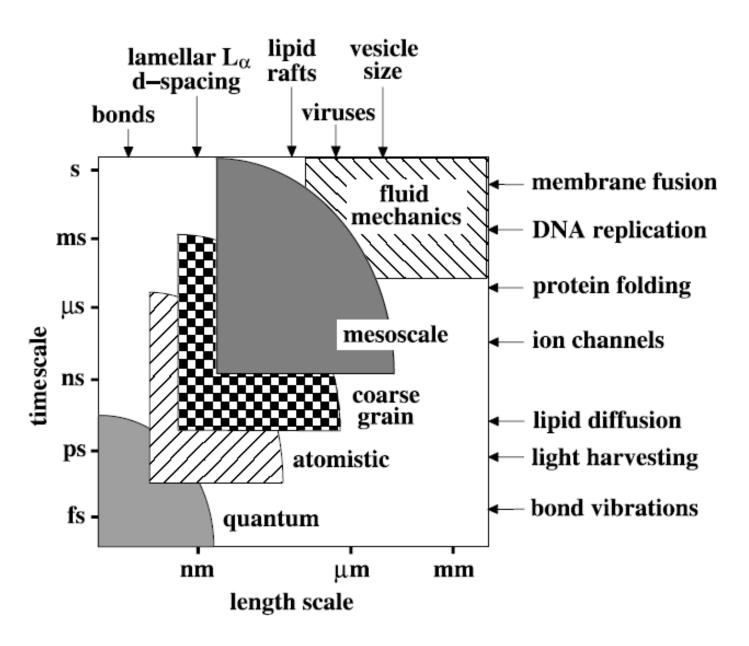
#### What is molecular modeling?

- Experiments performed on a computer
- Application of computational methods to study the structure, dynamics, reaction, and thermodynamics of molecules, and to understand molecular processes
- The models and methods must be tailored to the question at hand.
- This implies that molecular modeling must be both multidisciplinary and multiscale

## Multidiscipline

- At the interface of chemistry, physics, mathematics, computer science, molecular biology, material science, nanotechnology
- Numerous names: computational chemistry, molecular modeling, computational biochemistry, computational biophysics, computational nanoscience, computational structural biology, computational material science ....
- Drawing increasing interest: academia, industry (drug discovery)
- The field is developing with dazzling speed.

#### Multiscale (temporal and spatial)



# Connection to Wet Lab Experiments: Crucial

- Confirm, extrapolate and predict wet lab experimental results --- representing one more experimental tool, such as NMR
- Require information from wet-lab experiments
- Can help understand experimental results
- Can guide experiments (design)
- Can provide details not available in typical experiments

## Some Important Applications

- Refine the X-ray and NMR structures
- Understand structure (dynamics) function relationship
- Understand chemistry of the biological processes, including enzyme catalysis, DNA damage...
- Rational design of drugs, novel inhibitors, new protein and materials
- Predict the biomolecular structure, property, function

•

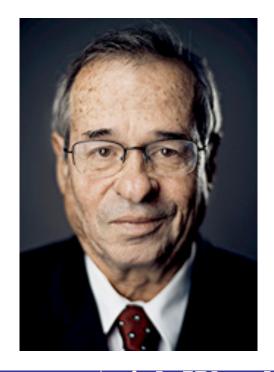
#### **NOBEL PRIZE IN CHEMISTRY 2013**

#### Taking the experiment to cyberspace

Chemical reactions occur at lightning speed; electrons jump between atoms hidden from the prying eyes of scientists. The Nobel Laureates in Chemistry 2013 have made it possible to map the mysterious ways of chemistry by using computers. Detailed knowledge of chemical processes makes it possible to optimize catalysts, drugs and solar cells.





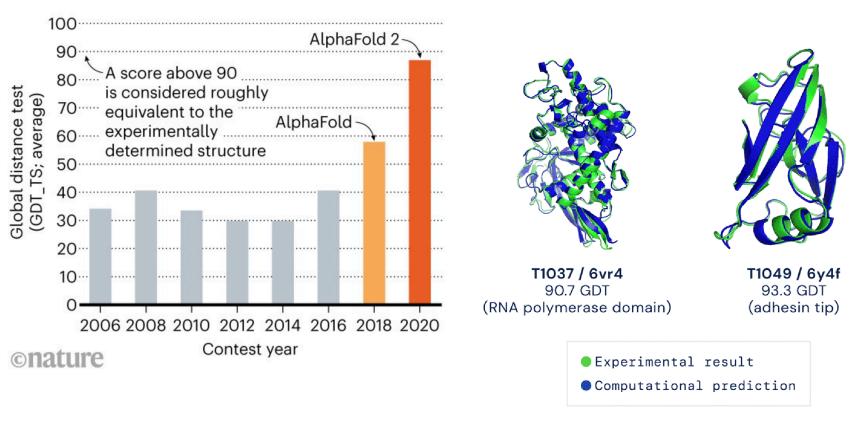


**Martin Karplus** 

**Michael Levitt** 

Arieh Warshel

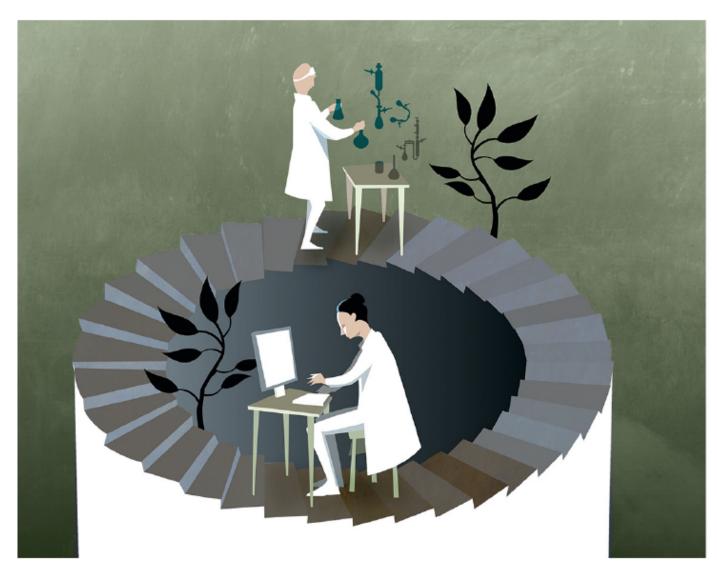
# Alpha Fold 2: predicting protein structure w deep learning



https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

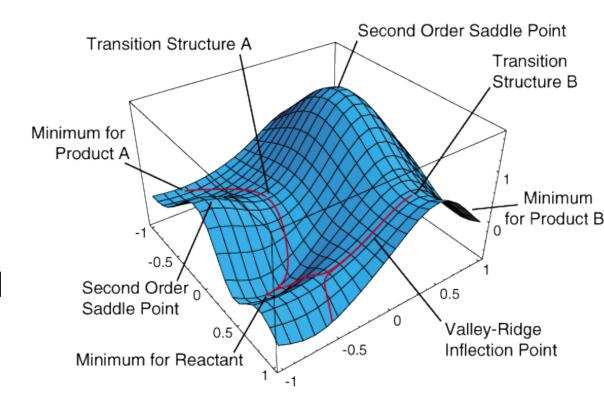
https://www.nature.com/articles/d41586-020-03348-4

Today chemists experiment just as much on their computers as they do in their labs. Theoretical results from computers are confirmed by real experiments that yield new clues to how the world of atoms works. Theory and practice crossfertilize each other.



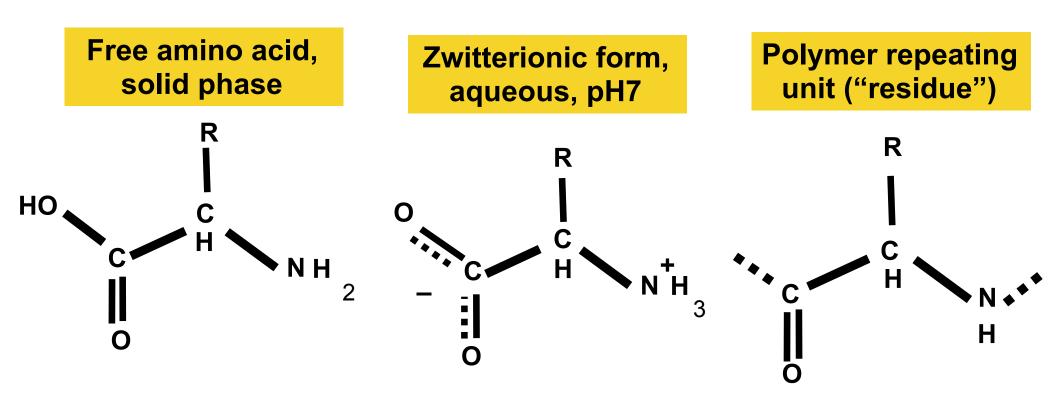
#### Important Elements for Molecular Modeling

- Structure of the target molecule
- Potential Energy Surface
- Simulation. (sampling, how to explore the PES)
- analyze the results and calculate the desired properties



#### Amino acids

### Amino acid states

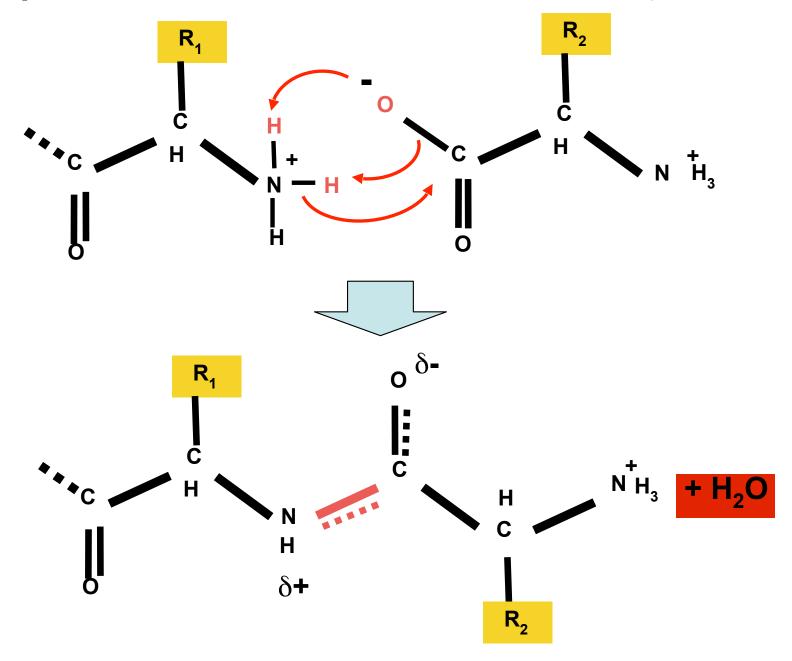


The chemical nature of the AA backbone depends on the context.

Usually we are talking about AAs in the polymer (polypeptide)

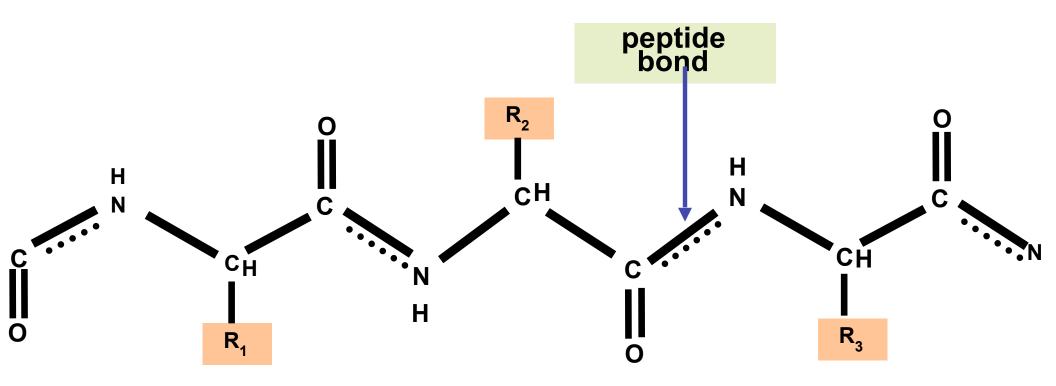
context.

#### peptide bond formation is dehydrating



Rx is catalyzed by the ribosome.

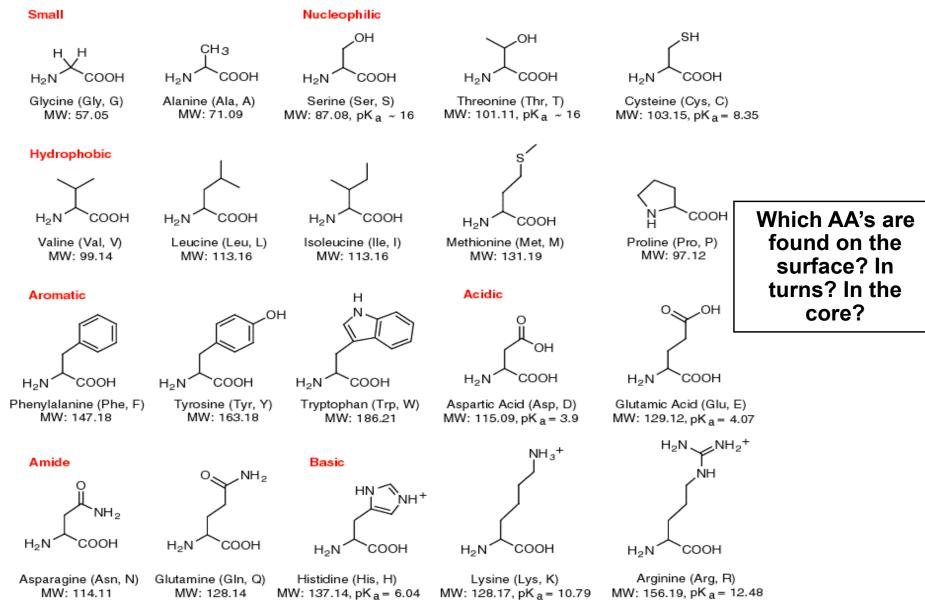
### the polypeptide backbone



Backbone atom names are "N, "C-alpha" and "C" (or N,CA,C).

Oxygen "O" is also considered a backbone atom. All atoms in all amino acids have conventional atom names.

#### the side chains



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

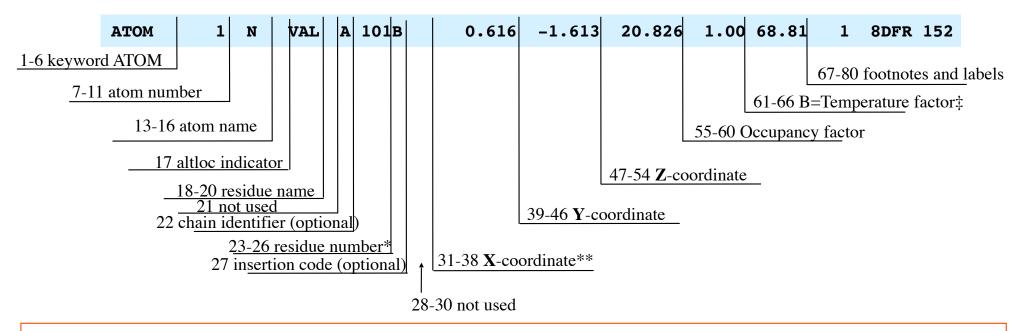
## How and where to obtain a suitable initial structure for your molecule?

- The first and essential task for a molecular modeling project.
- Building with softwares, such as MOE
- Crystal molecules of small molecules (Cambridge structural database <a href="http://www.ccdc.cam.ac.uk/products/csd">http://www.ccdc.cam.ac.uk/products/csd</a>
   )
- Large biomolecules: searching the protein data bank. http://www.rcsb.org. The files can be downloaded from the web

## What is a protein structure?

- A collection of 3D coordinates (x, y, z) of all atoms in the protein
- a pdb file in protein databank (pdb)
- Visualized by many viewer programs, such as VMD, pymol etc.

#### PDB ATOM lines



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

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

<sup>\*\*</sup> Coordinates are in orthogonal angstroms by convention. May be converted to crystallographic coordinates using CRYST lines.

## Sample pdb without header: specific atoms, x, y, z coordinates

ATOM	31	N	PRO	Α	3	2.202	-4.696	-6.182	1.00	0.00	A
ATOM	32	CA	PRO	Α	3	2.702	-4.606	-7.534	1.00	0.00	Α
ATOM	33	HA	PRO	Α	3	3.628	-4.050	-7.573	1.00	0.00	Α
ATOM	34	CB	PRO	Α	3	2.943	-6.070	-7.923	1.00	0.00	Α
ATOM	35	HB1	PRO	Α	3	3.965	-6.345	-7.706	1.00	0.00	A
ATOM	36	HB2	PRO	Α	3	2.737	-6.207	-8.974	1.00	0.00	Α
ATOM	37	CG	PRO	Α	3	1.990	-6.810	-7.073	1.00	0.00	Α
ATOM	38	HG1	PRO	Α	3	2.313	-7.833	-6.950	1.00	0.00	Α
ATOM	39	HG2	PRO	Α	3	1.000	-6.769	-7.505	1.00	0.00	Α
ATOM	40	CD	PRO	Α	3	2.031	-6.089	-5.770	1.00	0.00	Α
ATOM	41	HD1	PRO	Α	3	2.890	-6.425	-5.242	1.00	0.00	Α
ATOM	42	HD2	PRO	Α	3	1.133	-6.232	-5.193	1.00	0.00	Α
ATOM	43	С	PRO	Α	3	1.623	-4.031	-8.427	1.00	0.00	Α
ATOM	44	0	PRO	Α	3	0.442	-4.040	-8.056	1.00	0.00	Α
ATOM	45	N	GLY	Α	4	1.989	-3.539	-9.582	1.00	0.00	A
ATOM	46	HN	GLY	Α	4	2.938	-3.475	-9.817	1.00	0.00	Α
ATOM	47	CA	GLY	Α	4	0.981	-3.069	-10.497	1.00	0.00	Α
ATOM	48	HA1	GLY	Α	4	1.451	-2.673	-11.379	1.00	0.00	Α
ATOM	49	HA2	GLY	Α	4	0.415	-2.284	-10.014	1.00	0.00	A
ATOM	50	C	GLY	Α	4	0.037	-4.197	-10.883	1.00	0.00	A
ATOM	51	0	GLY	Α	4	-1.116	-3.958	-11.237	1.00	0.00	A
ATOM	52	N	ASP	Α	5	0.554	-5.441	-10.809	1.00	0.00	Α
ATOM	53	HN	ASP	Α	5	1.522	-5.538	-10.647	1.00	0.00	A
ATOM	54	CA	ASP	Α	5	-0.228	-6.669	-11.087	1.00	0.00	A
ATOM	55	HA	ASP	Α	5	-0.519	-6.652	-12.126	1.00	0.00	A
ATOM	56	CB	ASP	Α	5	0.655	-7.903	-10.846	1.00	0.00	A
ATOM	57	HB1	ASP	Α	5	1.112	-7.818	-9.870	1.00	0.00	A
ATOM	58	HB2	ASP	Α	5	1.430	-7.933	-11.597	1.00	0.00	Α
ATOM	59	CG	ASP	Α	5	-0.115	-9.217	-10.905	1.00	0.00	Α
ATOM	60	OD1			5	-0.117	-9.864	-11.972	1.00	0.00	Α
ATOM	61		ASP		5	-0.705	-9.614	-9.882	1.00	0.00	Α
ATOM	62	C	ASP		5	-1.483		-10.220	1.00	0.00	Α
ATOM	63	0	ASP	Α	5	-2.479	-7.374	-10.610	1.00	0.00	Α
ATOM	64	N	THR	Α	6	-1.436	-6.138	-9.075	1.00	0.00	A

## pdb header (e.g. 2in0)

- A wealth of information
- Title: what protein
- What PSD methods; Resolution for Xray
- Authors, citation
- Sequence (SEQRES)
- 2nd structure (HELIX, SHEET)
- Mutations (SEQADV)

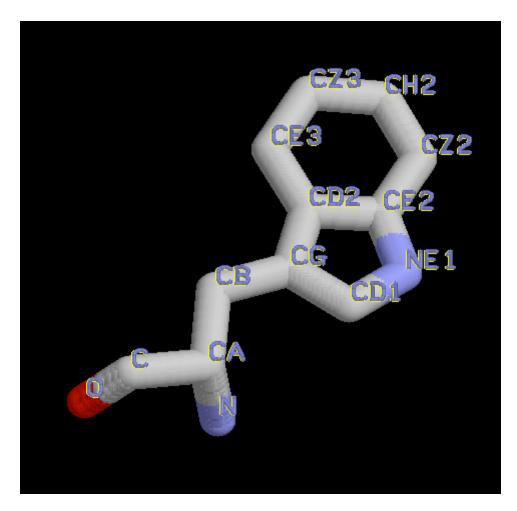
#### Protein Data Bank (www.rcsb.org)

 The pdb archive is a repository for the processing and distribution of 3-D biological macromolecular structure data

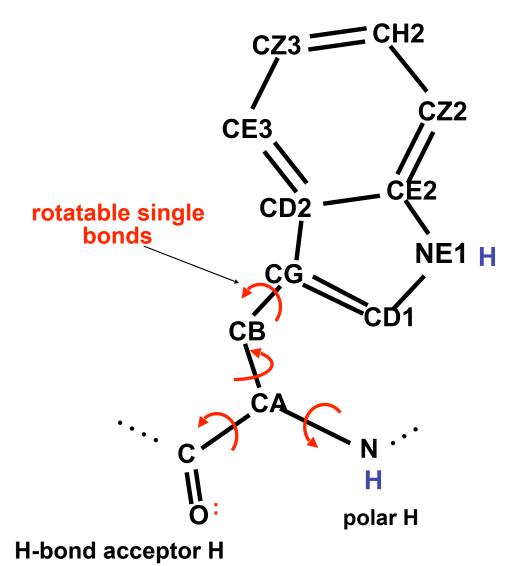
 Tons of information, many useful tools but maybe challenging to explore ...

Understanding PDB Data: Looking at Structures

## atom names: tryptophan

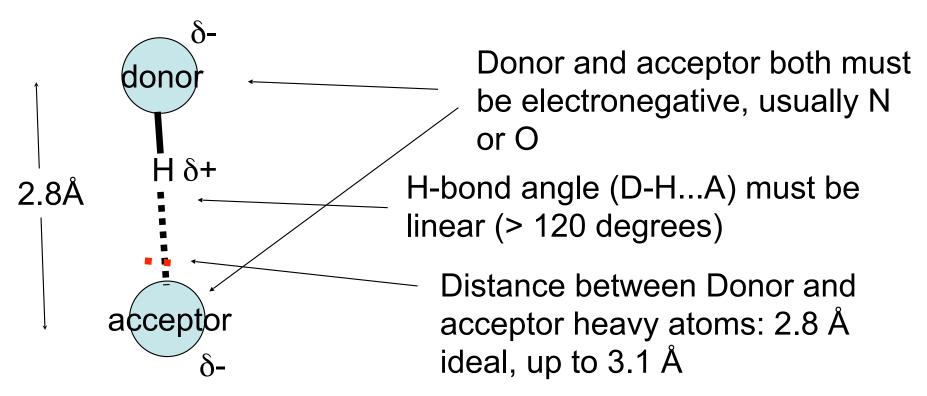


PDB convention atom names follow the formula: <element><greek letter><alt posit>



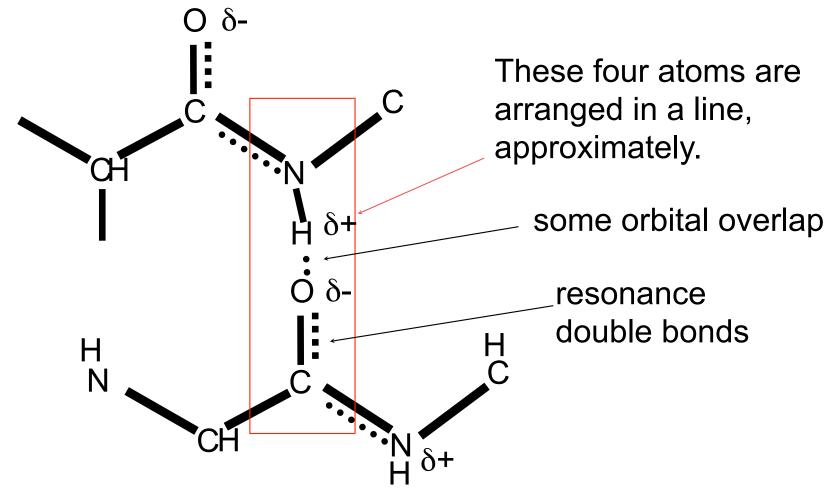
CA = Alpha Carbon, CB = Beta Carbon, OG = Gamma Oxygen, Delta.., Epsilon.., Zeta..,

#### Hydrogen bonding holds it all together



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.

## H-bond stabilized 2<sup>nd</sup> Structure: helices, strand and turns.



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.

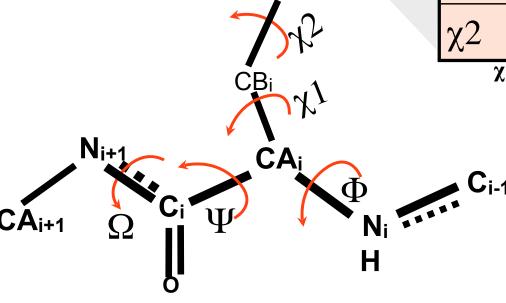
#### Torsion angles

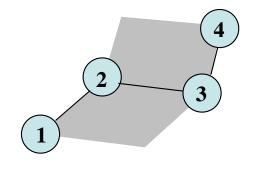
Protein flexibility is due to rotations around single bonds, backbone and side chain.

4 atoms define two planes

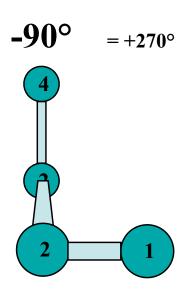
Angle	atom1	atom2	atom3	atom4	
Φ	C <sub>i-1</sub>	Ni	CAi	Ci	
Ψ	Ni	CAi	Ci	N <sub>i+1</sub>	
Ω	CAi	Ci	N <sub>i+1</sub>	CA <sub>i+1</sub>	
χ1	Ci	CAi	CBi	xGi	
χ2	CAi	CBi	xGi	хDі	

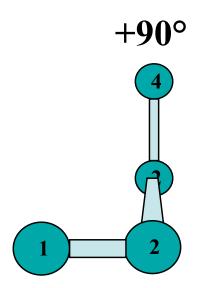
 $\chi = chi$ 





#### Measuring a torsion angle by eye





Positive direction: clockwise

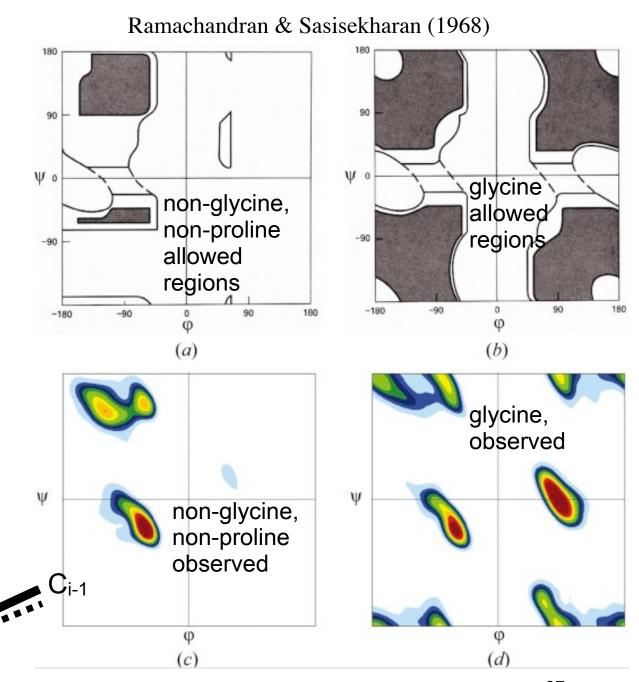


#### Ramachandran Plot maps allowable phi, psi regions

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.

 $N_{i+1}$ 

CBi



#### MOE demo

- Load 4Y6K, a membrane protease bound to an inhibitor (L679), rotate, translate, zoom in and out
- open Sequence editor (SE)
- delete unwanted chains
- Synchronize selection in SE/Select
- Select the drug: center, try different atom presentation
- Select the catalytic aspartates D162, D200
- Measure distance to ligand: is there a hydrogen bond between the ligand and the catalytic aspartates? (add hydrogen, measure hbond length and angle)?



#### Assignments:

- 1. Complete MOE tutorial (help/tutorials/getting started)
- MOE Tour from "Building a Small Molecule" to "Introducing the Sequence Editor"
- 3. Review amino acid side chain structure: quiz on Tuesday

### Review questions

- What is a "residue"?
- What atoms make up the polypeptide backbone?
- What do we find in a PDB file?
- How are amino acid side chain atoms named?
- What is a Ramanchandran plot?
- What is a hydrogen bond made of?
- How do I measure a torsion angle?
- Is there any difference between a torsion angle and a bond angle?