

Setting up and Analysing Biomolecular Simulations: Good Practice and Pitfalls

CCPBioSim Workshop Bristol, 2019

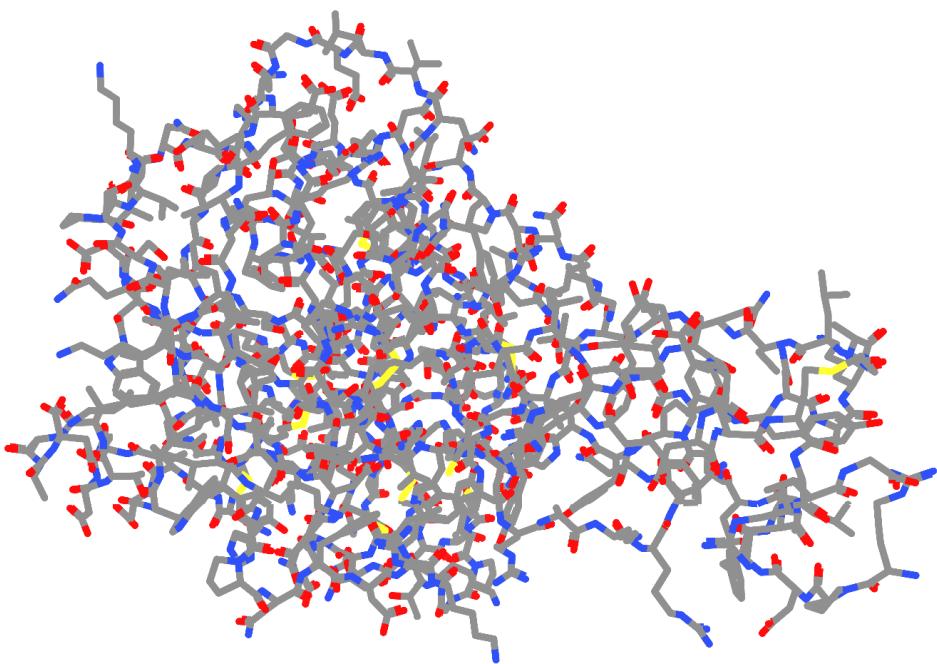
Charlie Laughton

University of Nottingham

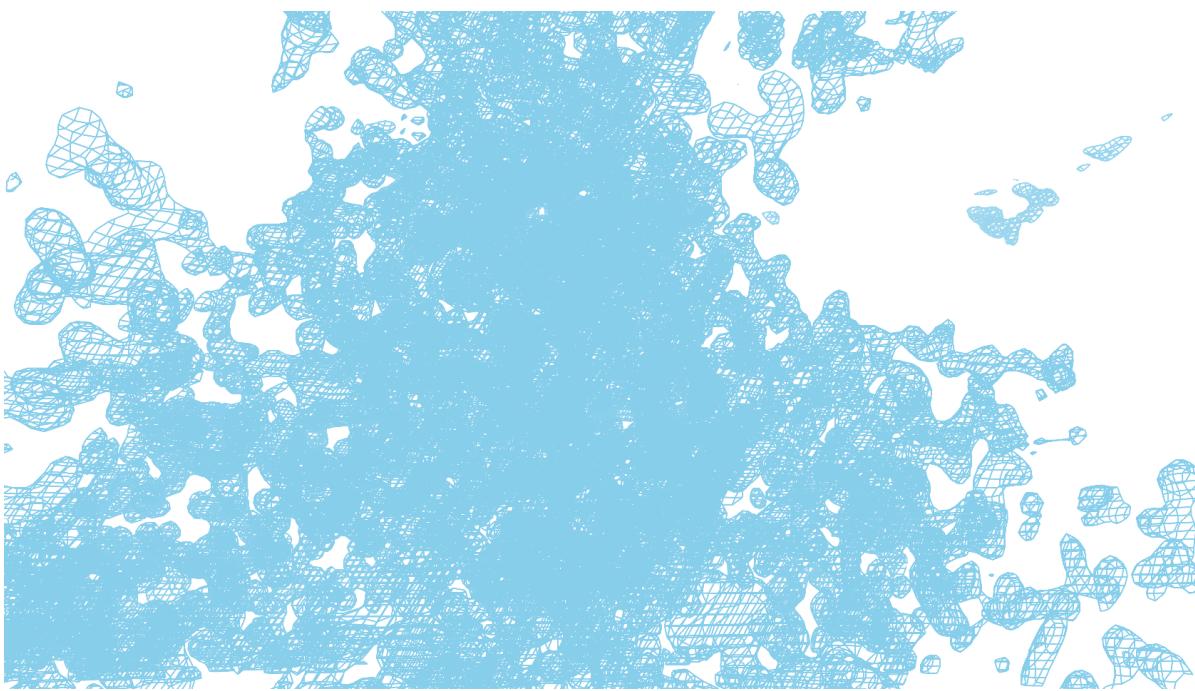
Outline

- Setting up simulations
 - Starting structures: rubbish in, rubbish out.
 - Equilibration – says who?
- Analysing simulations
 - Interactive simulation analysis with Jupyter notebooks:
 - Basic simulation analysis methods using Python.
 - Analysing equilibration and sampling.
 - Dealing with uncertainty – simple statistical methods.

Crystal structures



Crystal structures



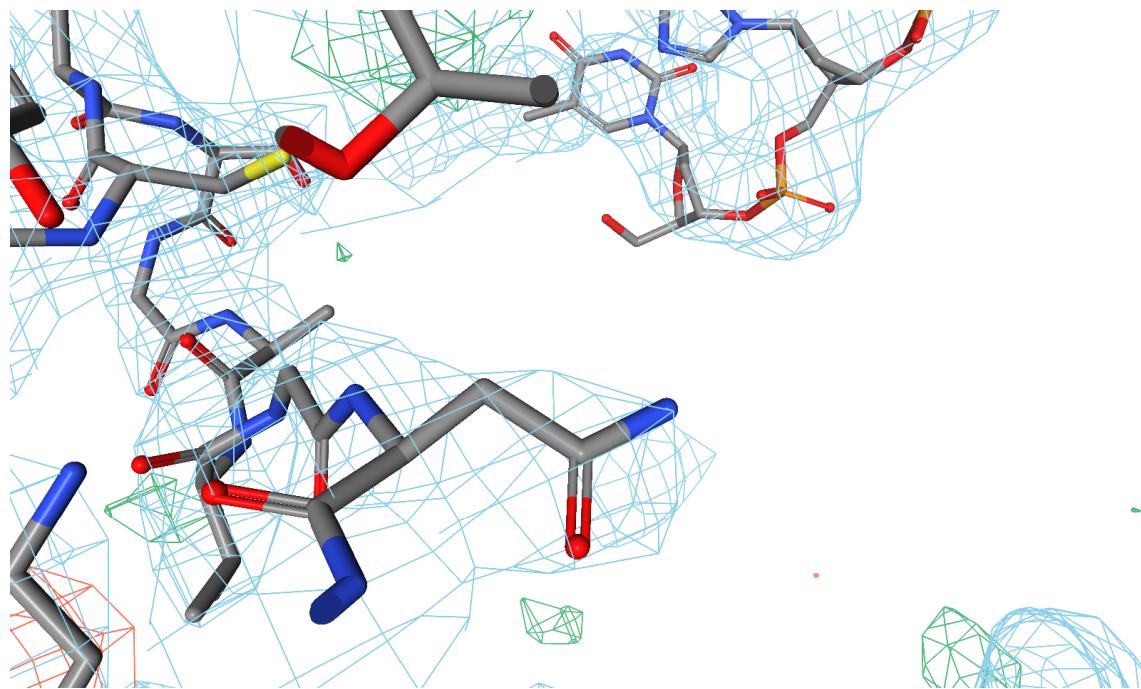
Fitting a model into the electron density

- 1T38 Asn157:



Fitting a model into the electron density

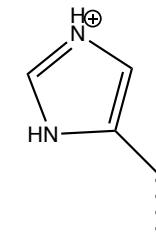
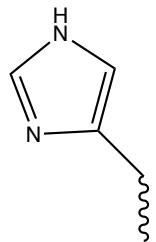
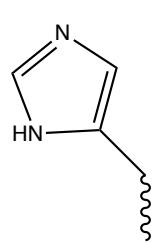
- 1T38 Asn157:



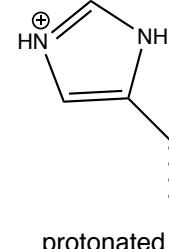
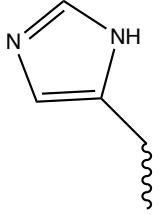
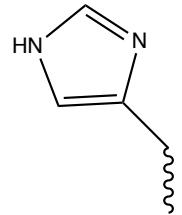
Fitting a model into the electron density

- 1T38 His171:

PDB file:



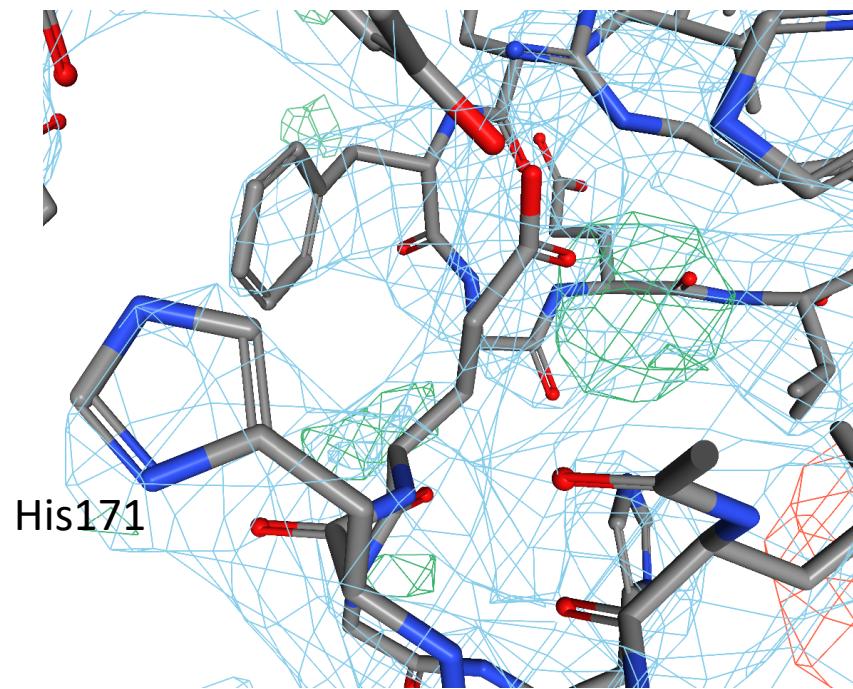
Or flipped?:



δ -tautomer

ϵ -tautomer

protonated

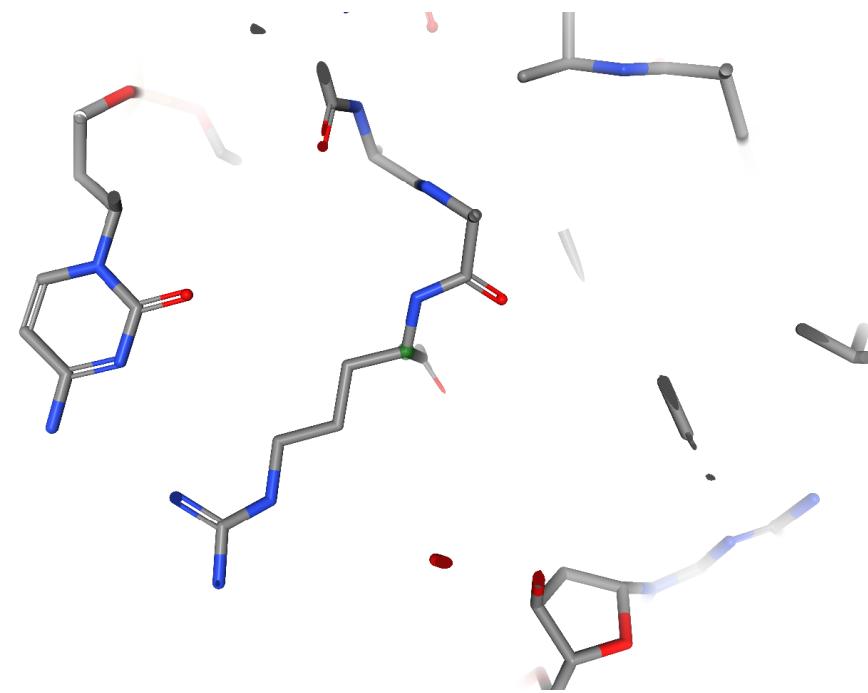
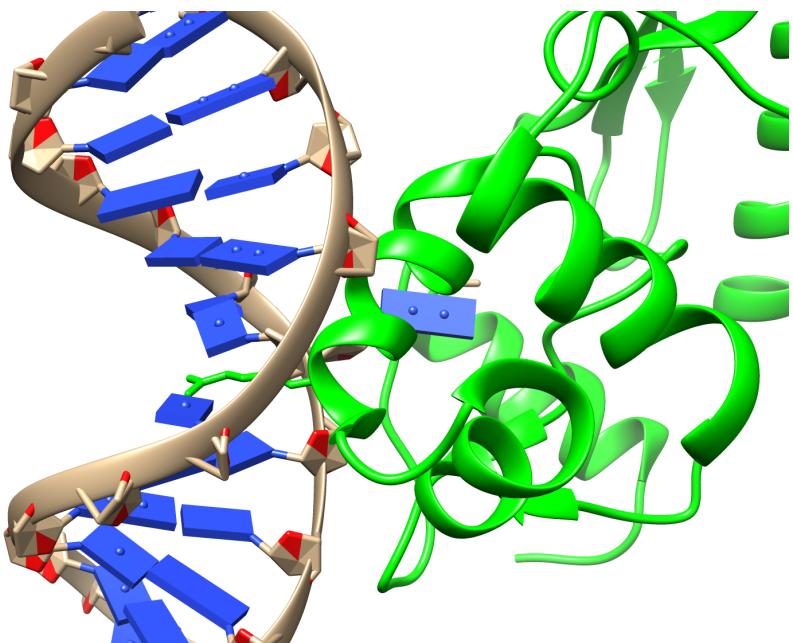


Assigning protonation states/tautomers

- PropKa: Web server currently down but can use PDB2PQR server:
http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/
- PropKa software can be downloaded:
<https://github.com/jensengroup/propka-3.1>
- Note 1: the Propka software only calculates pKas, but the PDB2PQR server will also tell you about predicted tautomers for His.
- Note 2: Other on-line services are available!

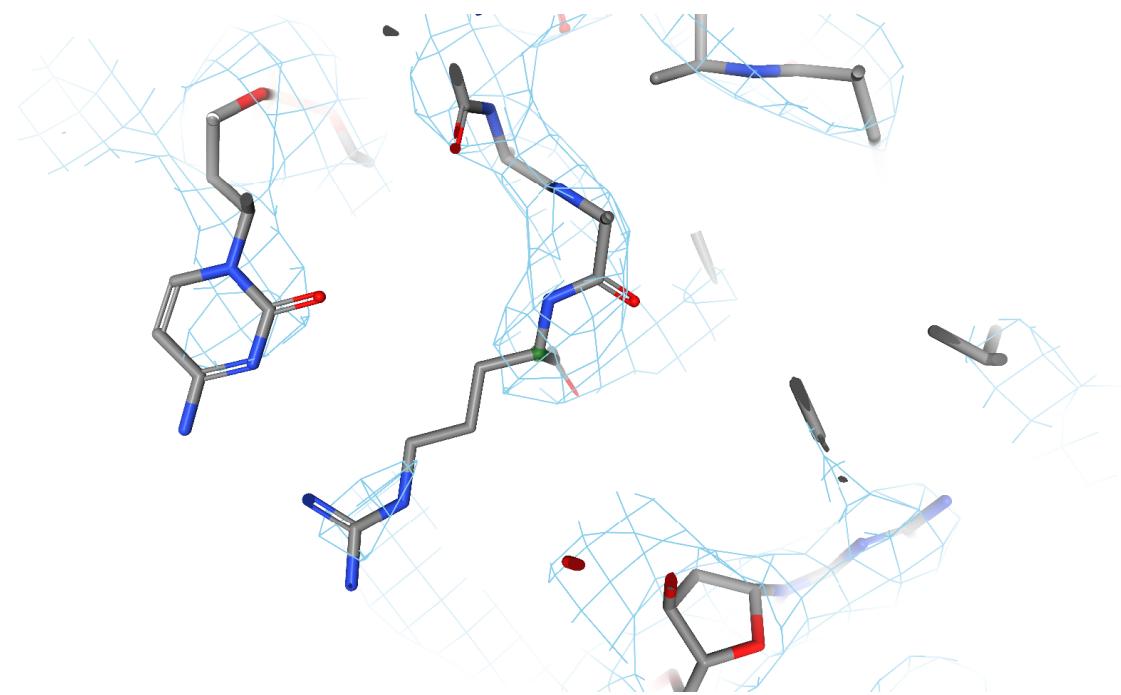
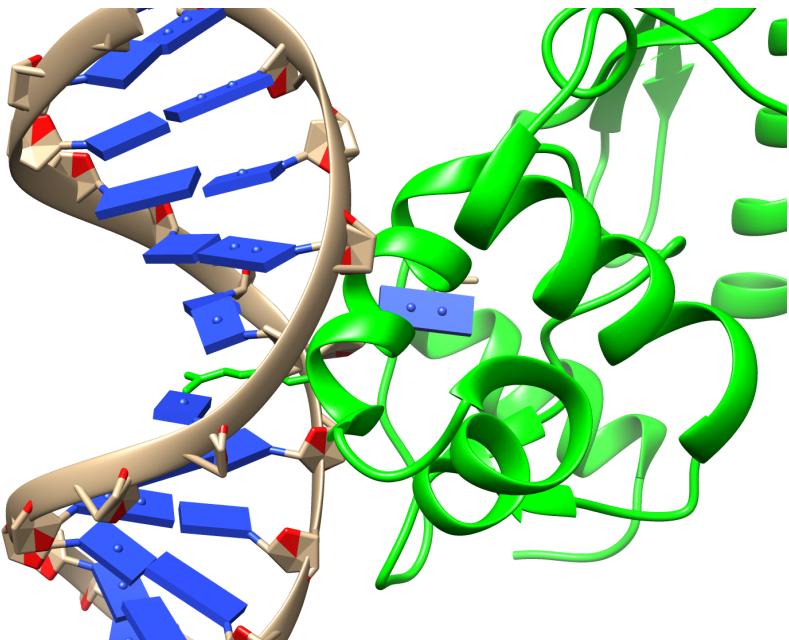
Learning the hard way

- 1T38 (DNA repair protein MGMT)
 - Arginine – cytosine interaction



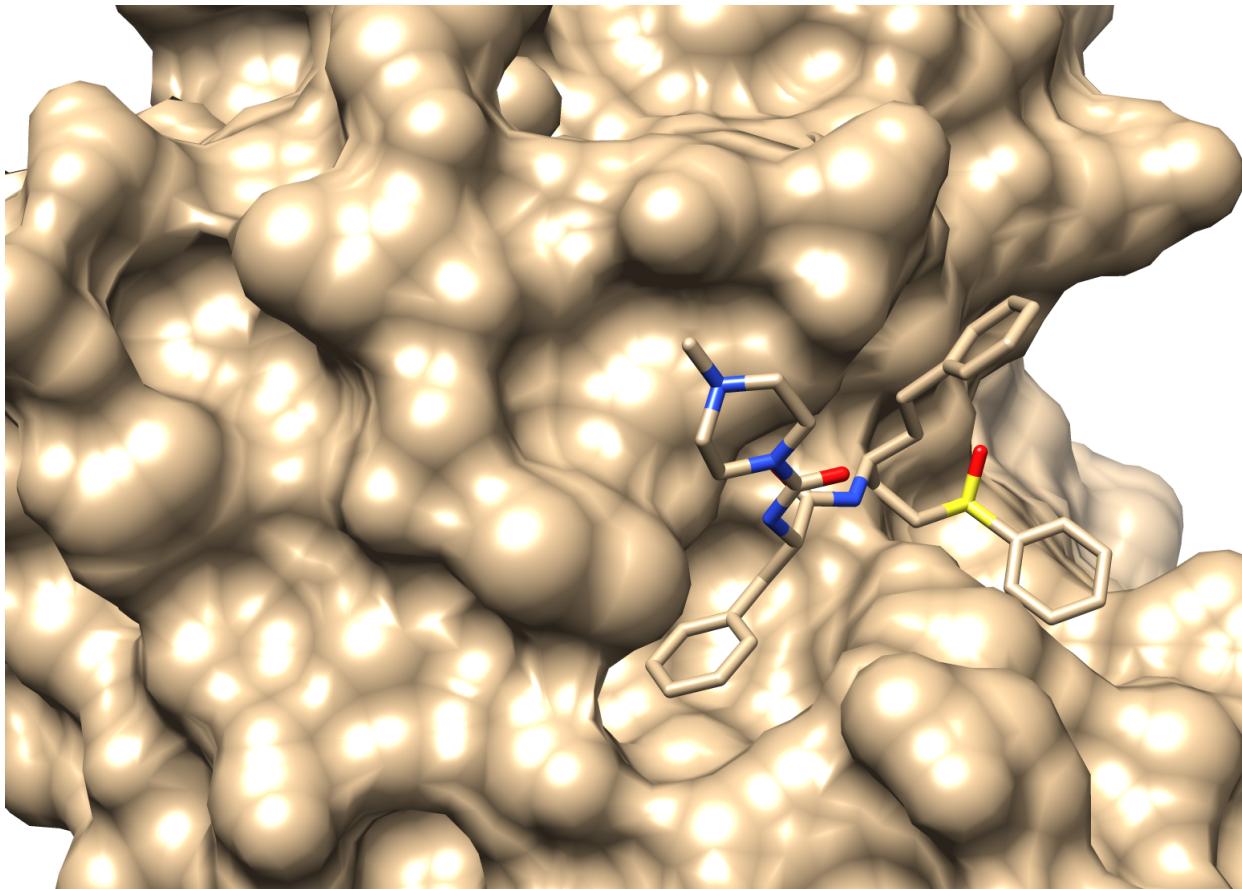
Learning the hard way

- 1T38 (DNA repair protein MGMT)
 - Arginine – cytosine interaction



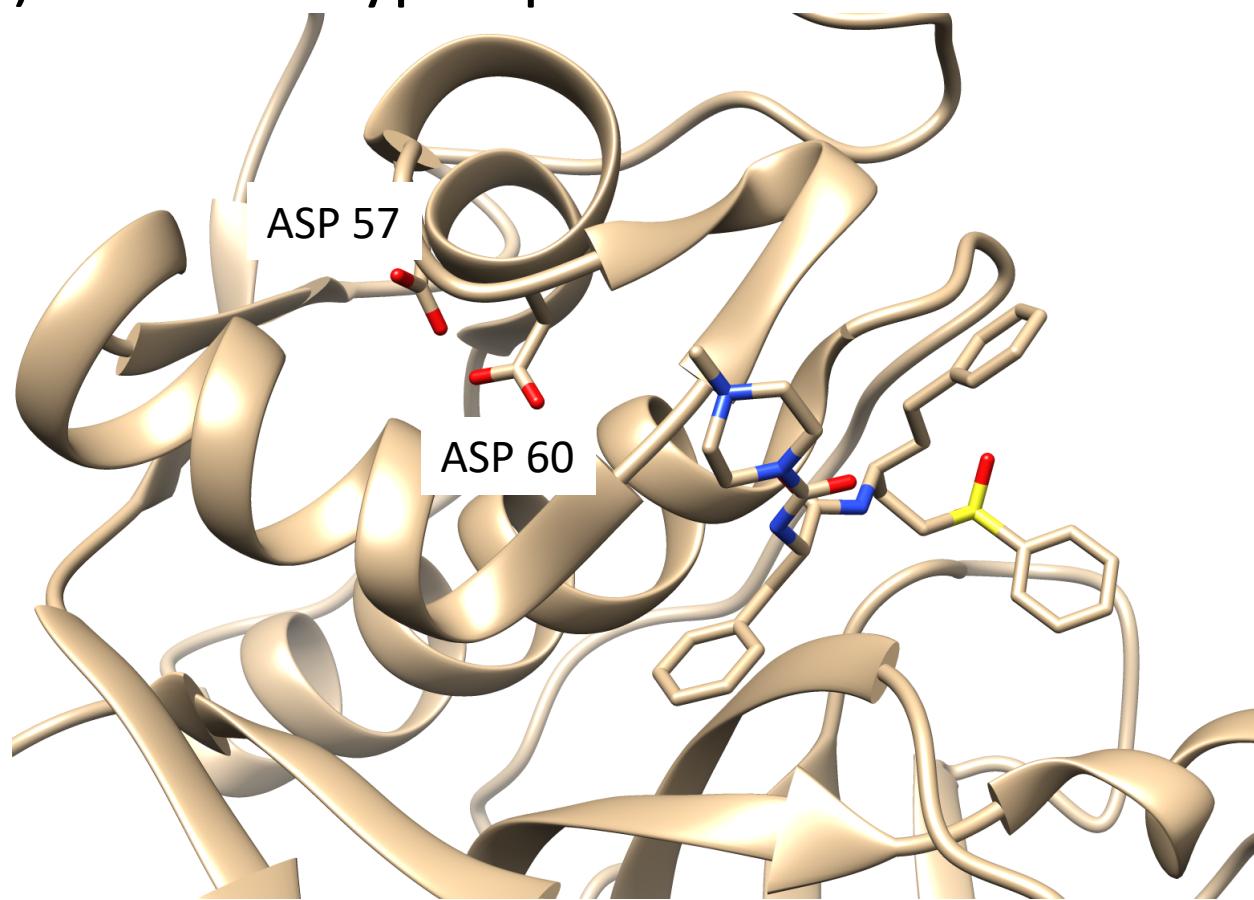
Learning the hard way

- Cruzain (2oz2) and the cryptic pocket



Learning the hard way

- Cruzain (2oz2) and the cryptic pocket



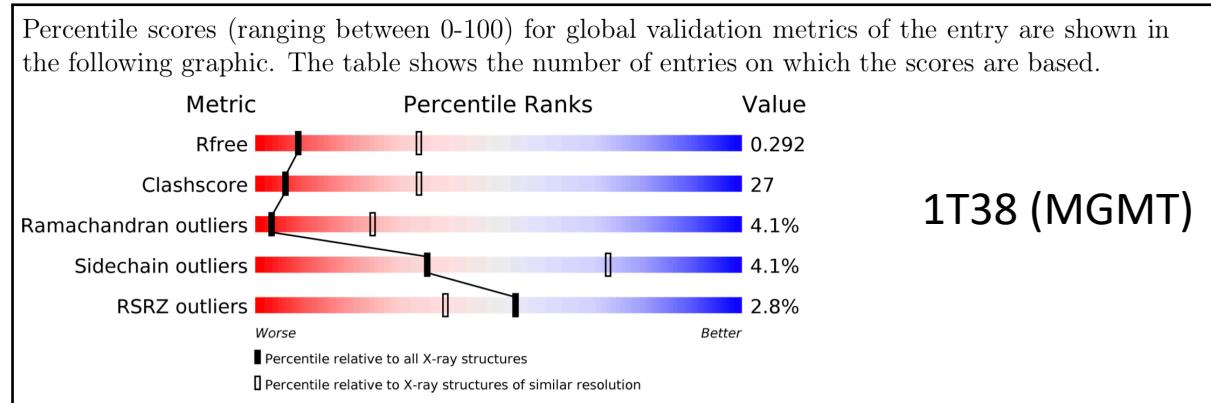
PDB-bad

- <http://swift.cmbi.ru.nl/teach/pdbad/>
- from Gert Vriend (author of WHAT_CHECK).
- Examples of errors/oddities in PDB structures.



Tools to help

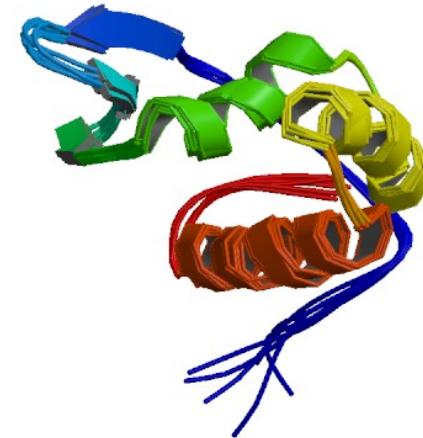
- Most are available as web services rather than as downloadable tools:
- MolProbity: <http://molprobity.biochem.duke.edu>
- SAVES: multiple tools “meta-server”:
<http://services.mbi.ucla.edu/SAVES/>
- Tools from the PDB: http://www.rcsb.org/#Subcategory-analyze_quality



Exercise 1 – using MolProbity

** USE YOUR OWN LAPTOP/WORKSTATION FOR THIS – NOT THE WORKSHOP JUPYTER HUB **

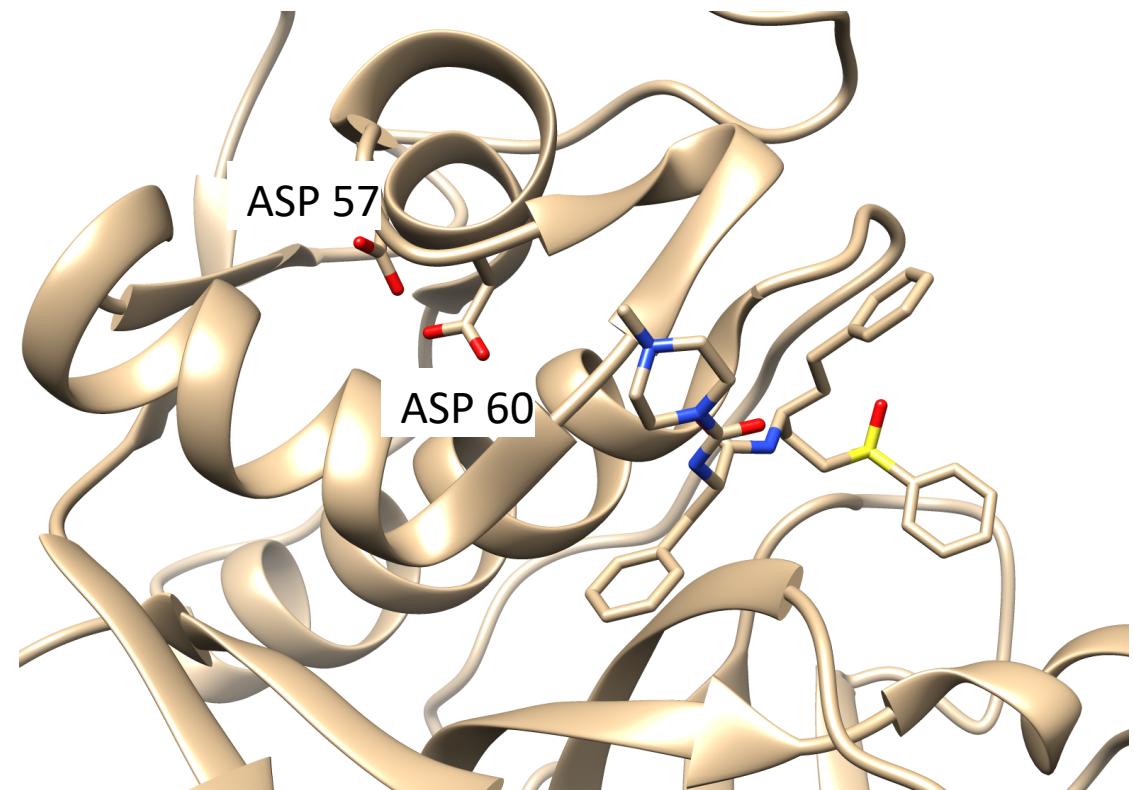
- 5SXY:
 - Solved by NMR, 20 structures deposited in PDB.
 - Are they all of equivalent quality?
1. Go the RCSB website and download the pdb file.
 2. Edit the pdb file 5sxy.pdb to leave just one model, save under a new name (e.g. 5sxy_1.pdb).
 3. Submit to MolProbity at <http://molprobity.biochem.duke.edu>
 4. You will see a message in red telling you that the hydrogen atoms in the model have been removed. When you click on “continue” at the bottom of the page, reselect the original file you uploaded from the “Currently working on” drop-down list on the next page.
 5. Click on “Analyze all-atom contacts and geometry”.
 6. On the next page leave the list of suggested analyses to run at the default values, and click “Run programs to perform these analyses” at the bottom of the page.
 7. Look through the results, and make a note in particular of:
 - a) The MolProbity score.
 - b) Any poor geometry metrics shown in red.
 8. Add your data to the workshop etherpad.



Exercise 2 – using Propka

** USE THE WORKSHOP JUPYTER HUB FOR THIS**

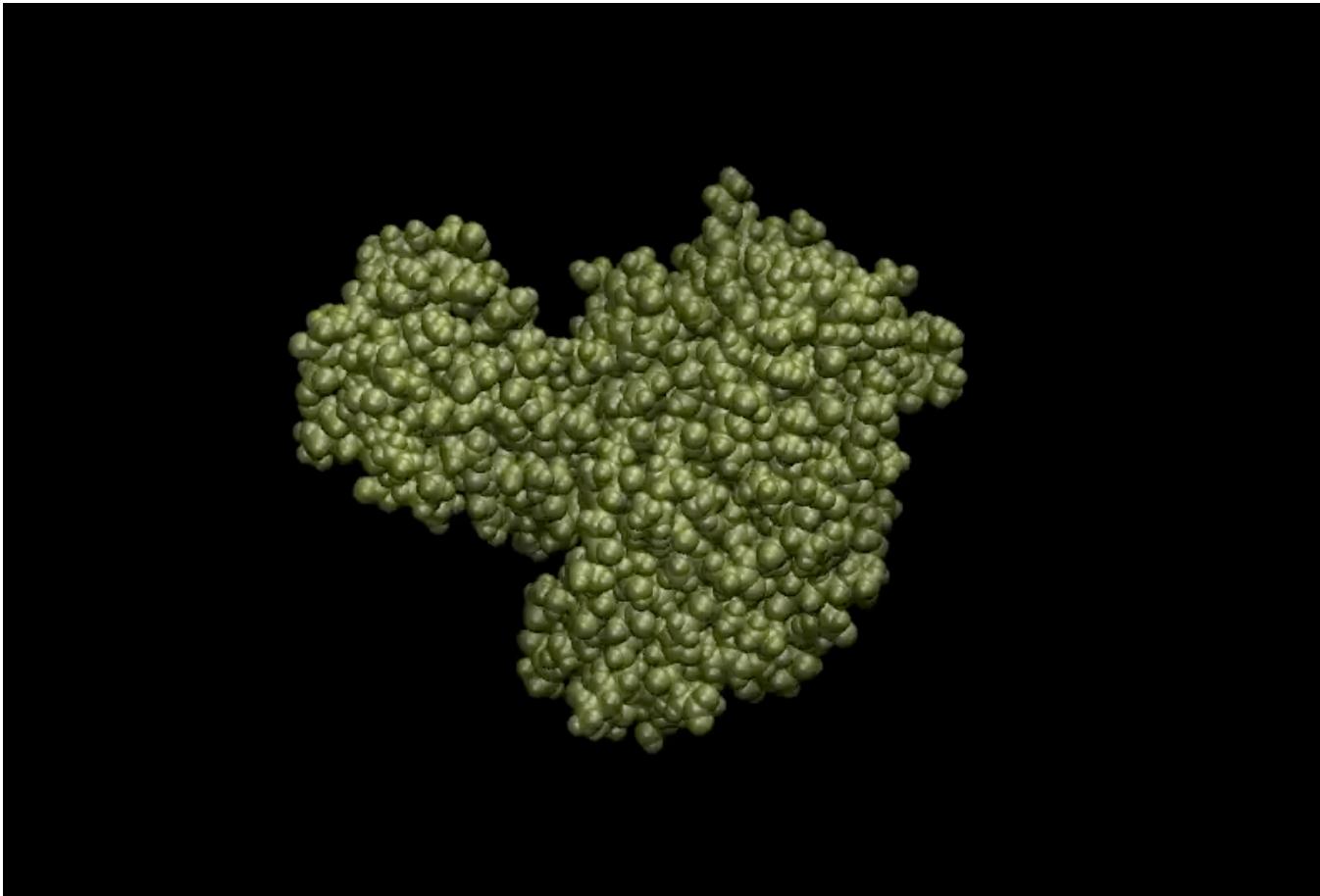
- 2oz2:
 - That cruzain structure we talked about earlier.
 - Which of those two aspartates is the protonated one?
 - Are there other “tricky” residues?



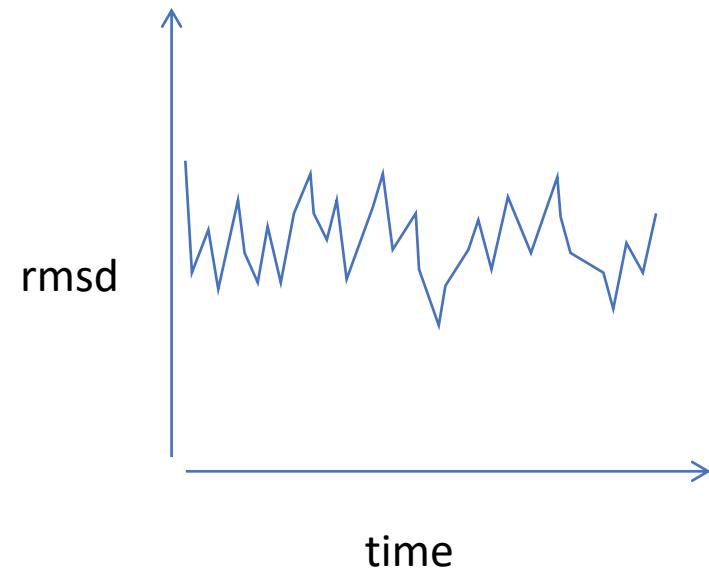
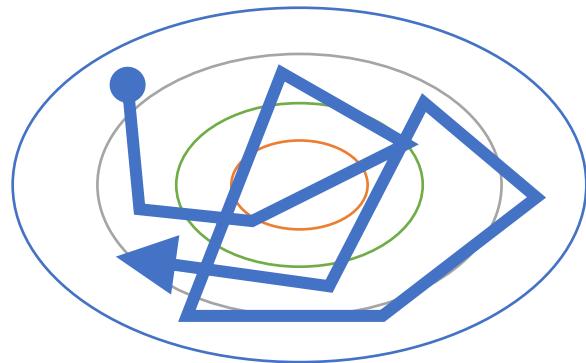
Kerr et al. (2009) J. Biol. Chem. **284**: 25697-25703

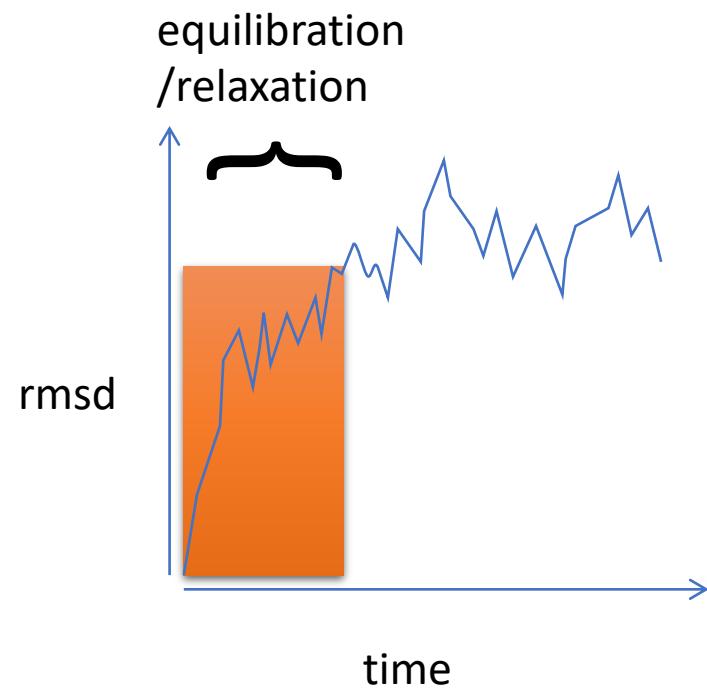
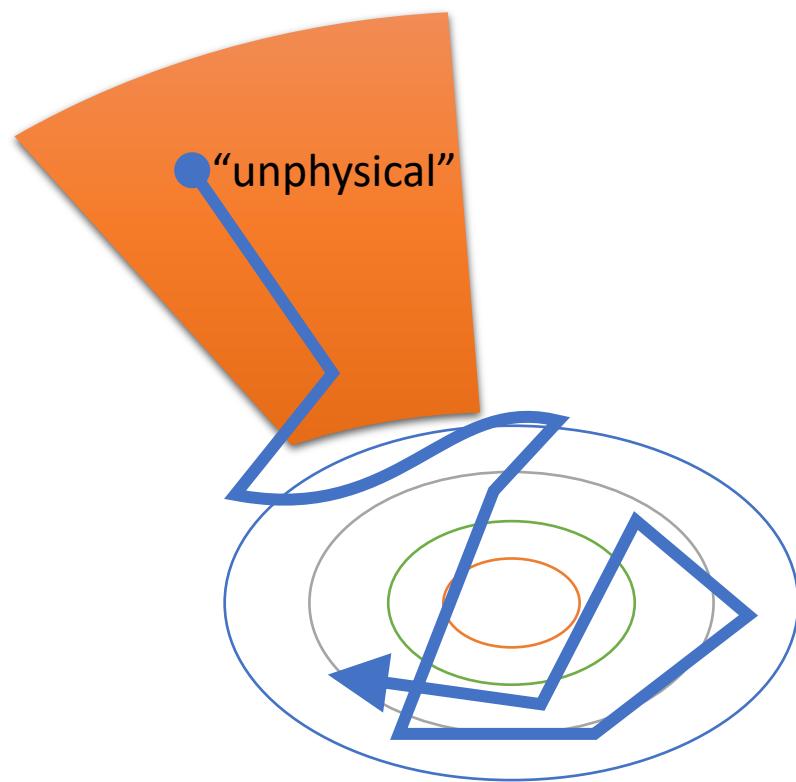
Equilibration

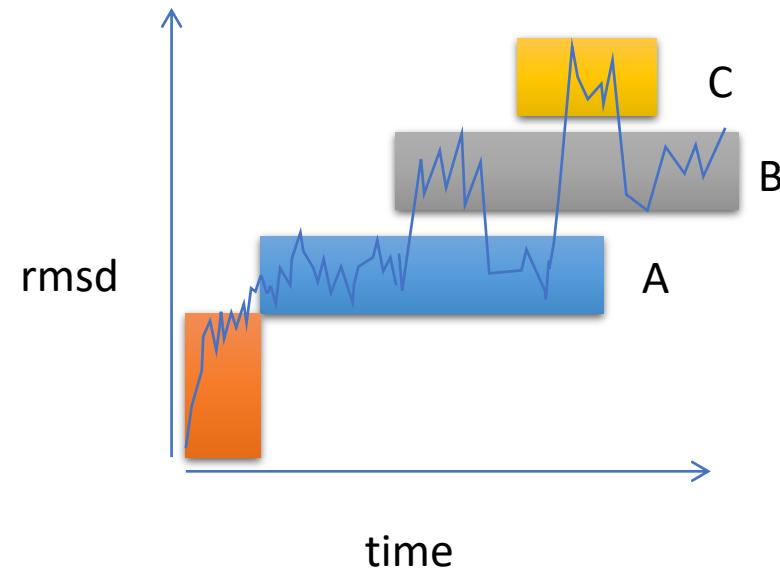
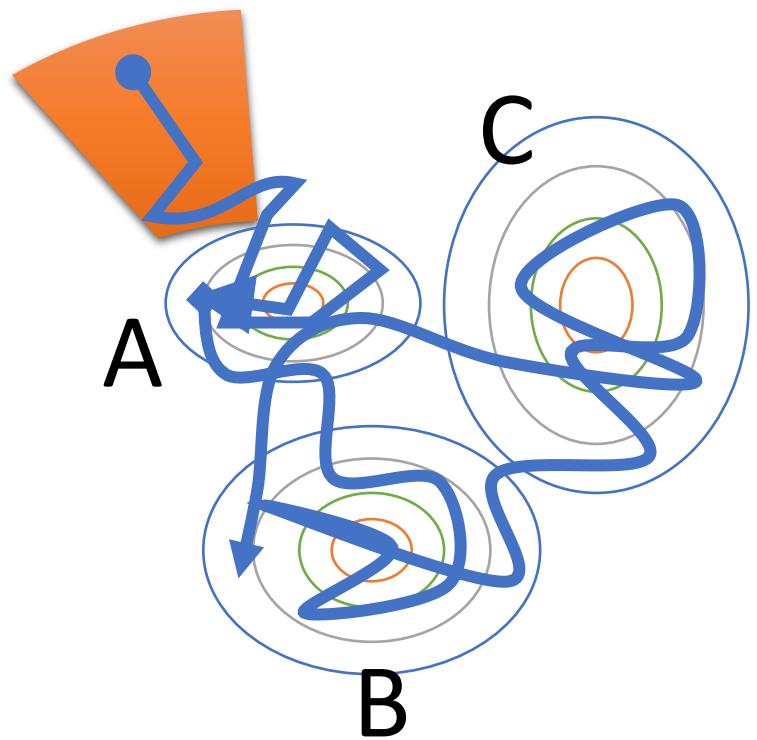
- What do we mean by equilibration?



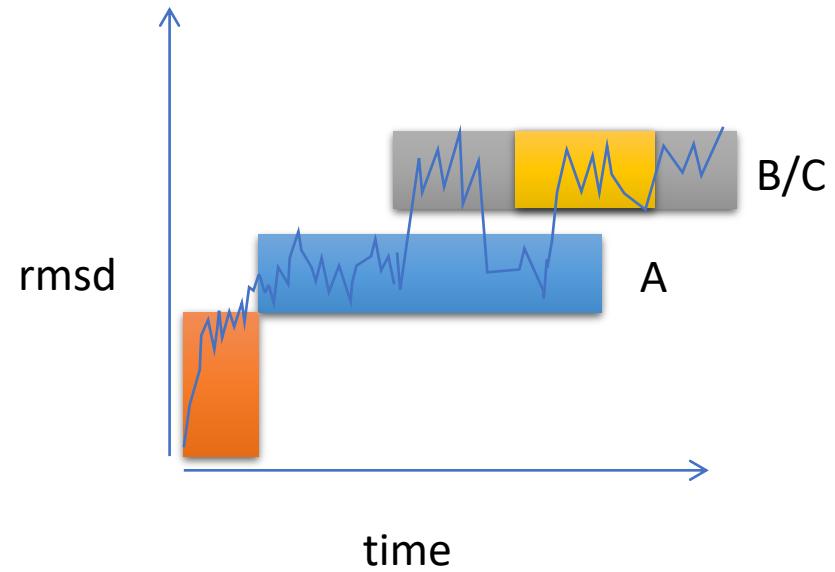
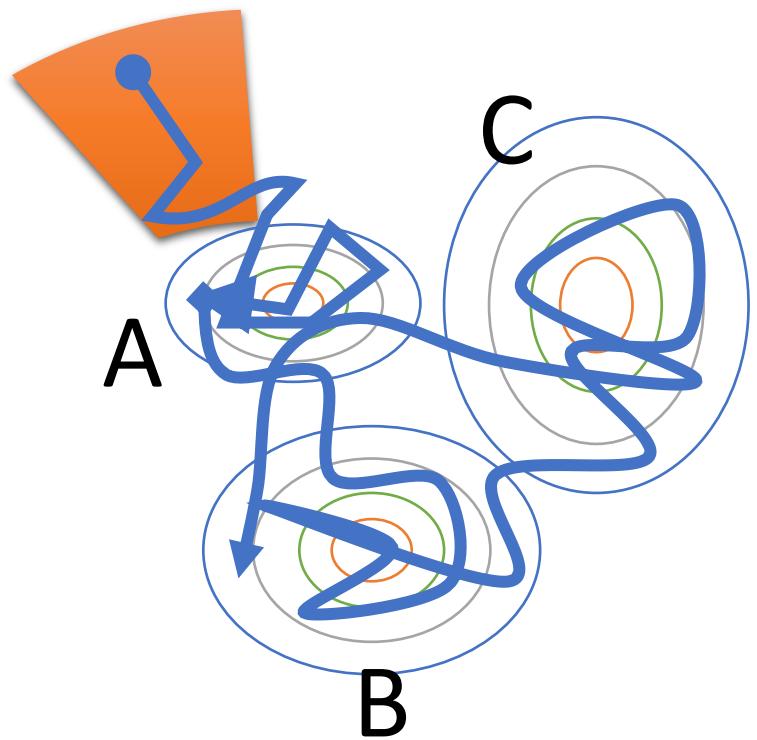
MD: A walk over a potential energy surface





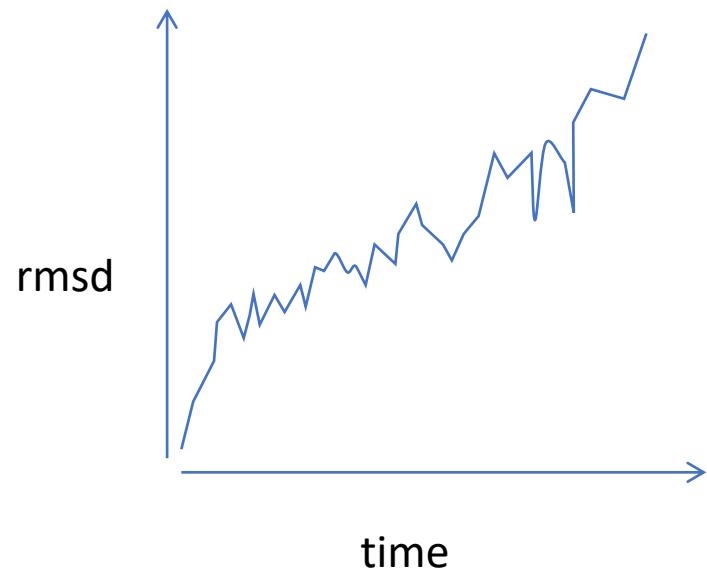


“Jumping Among Minima”

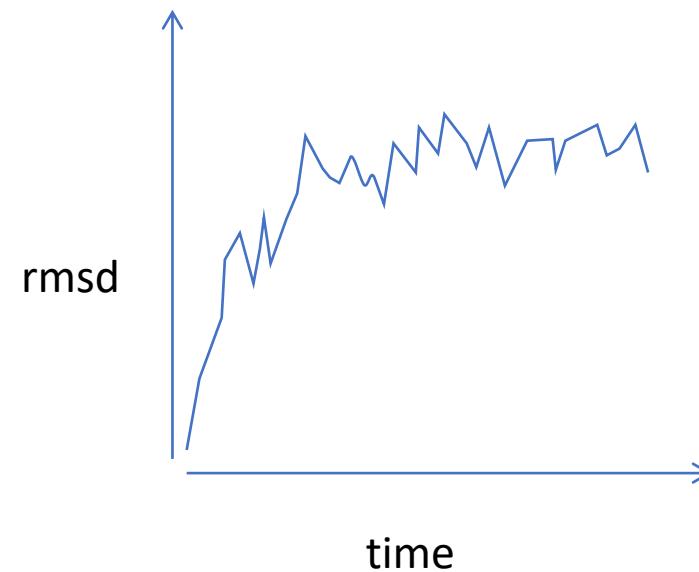


The Analysis of MD Data

- RMSD plots are of limited value

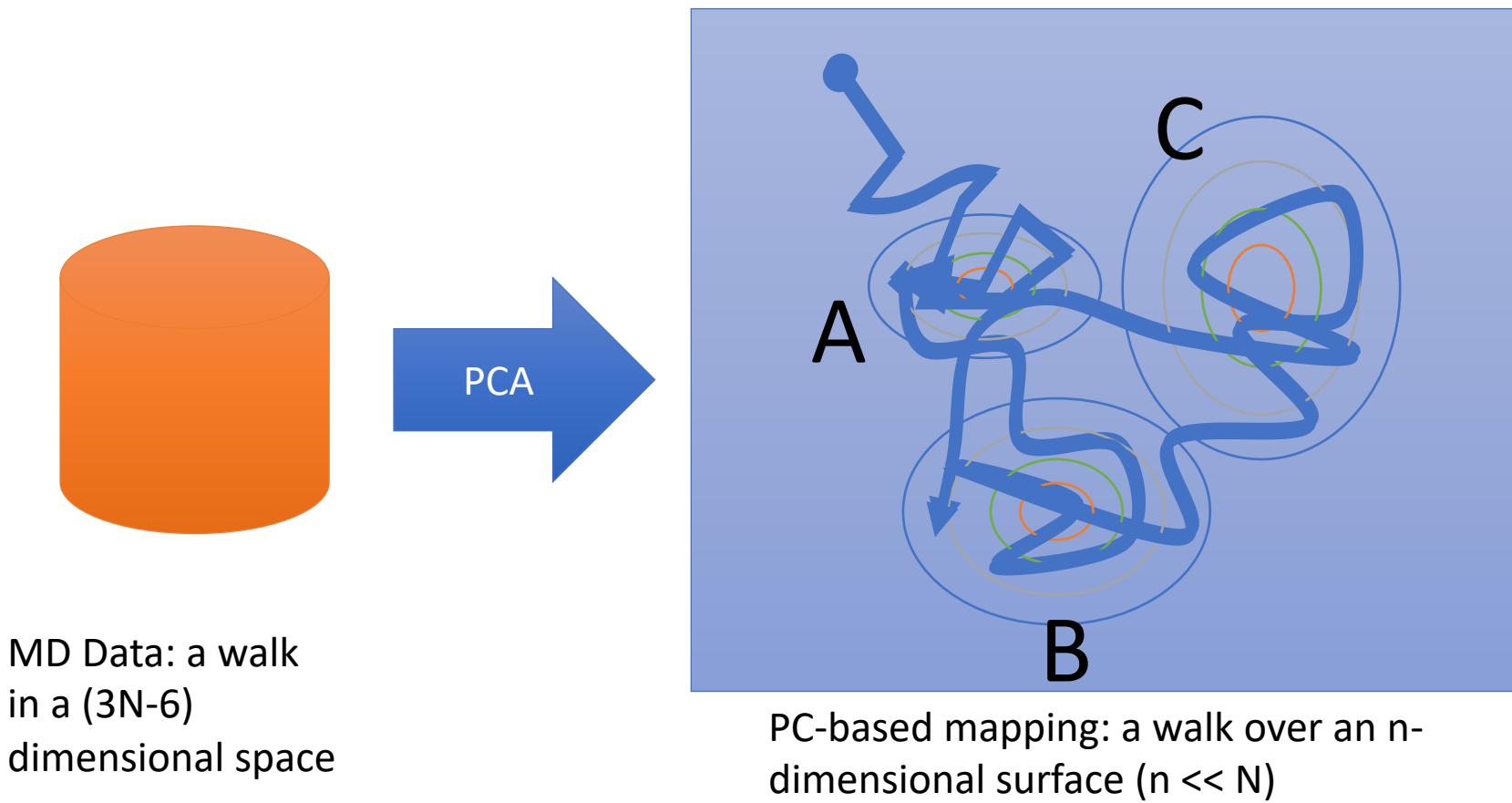


This is definitely bad

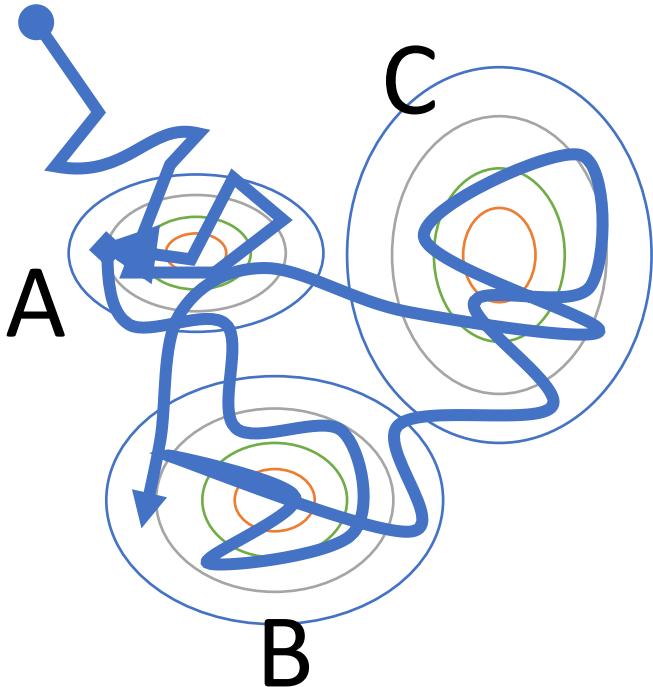


This is not necessarily good

Principal Component Analysis

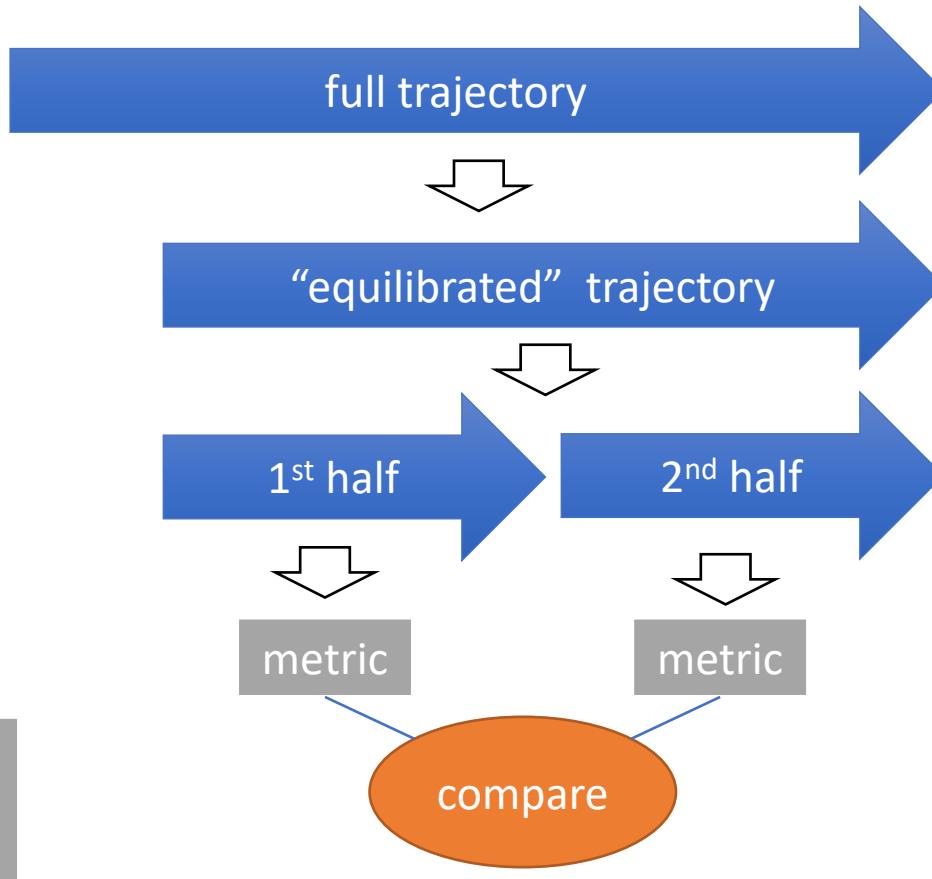


Assessing Sampling and Convergence

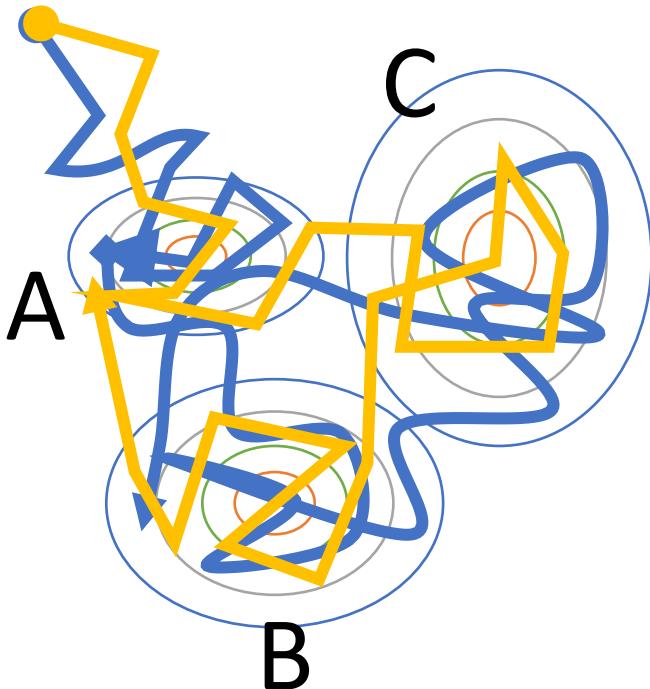


Metrics:

- % time in A, B, C
- average distance between two atoms
- radius of gyration
- (etc.)

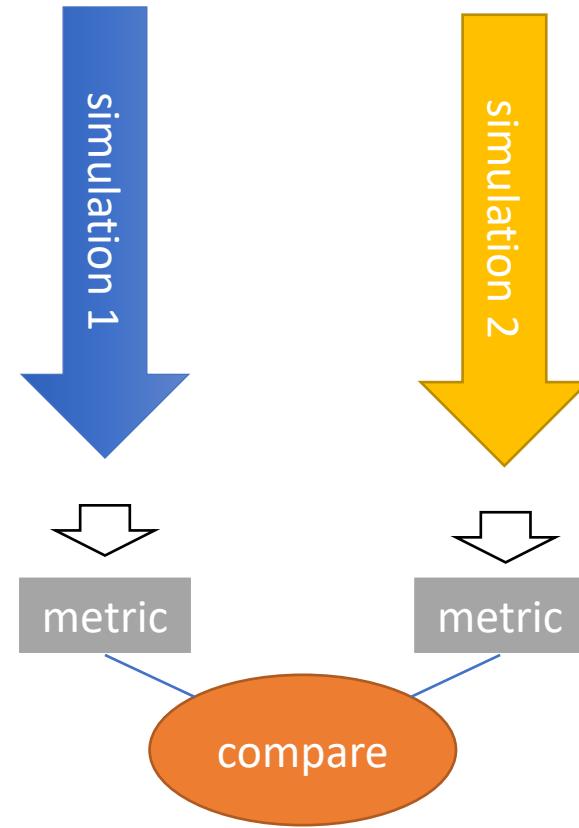


Assessing Sampling and Convergence



Metrics:

- % time in A, B, C
- average distance between two atoms
- radius of gyration
- (etc.)



Example: Alanine pentapeptide

- Sampling of local minima by 100 independent 100ns simulations:

		Local minimum								
		c1	c2	c3	c4	c5	c6	c7	c8	c9
replicate	R0	32	16	6	8	11	9	5	5	6
	R1	33	17	6	8	8	8	9	4	7
	R2	52	16	4	9	3	10	1	2	4
	R3	28	31	8	4	1	5	4	16	5
	R4	20	11	10	21	9	5	10	7	7
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

% occupancy of each local minimum

Example: Alanine pentapeptide

- Convergence through amalgamation of data:

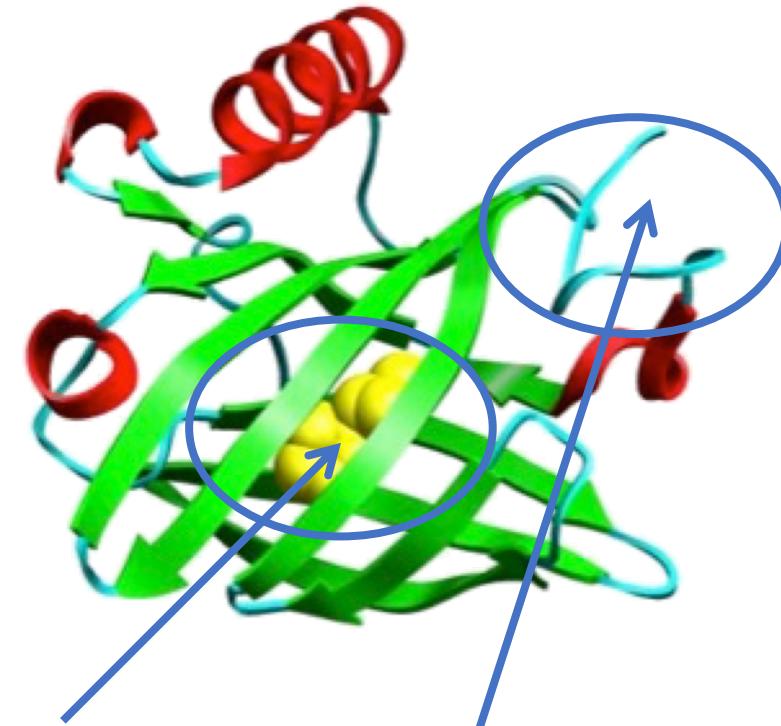
		Local minimum								
		c1	c2	c3	c4	c5	c6	c7	c8	c9
replicate	R0-19	28	16	11	10	9	7	6	7	6
	R20-39	27	15	11	11	10	9	6	6	5
	R40-59	26	14	11	11	12	7	7	7	5
	R60-79	29	16	11	10	10	7	6	5	6
	R80-99	27	16	11	10	10	8	6	5	5

% occupancy of each local minimum

Not all parts of a protein relax at the same rate

Typically:

- The core relaxes faster than the surface.
- The mainchain relaxes faster than sidechains.
- Helices and sheets relax faster than unstructured elements.



If you are interested in this

You *may* not need to worry so much about this

Part 2: The analysis of simulation data

- Three Jupyter notebooks for you to work through at your own pace:
 1. An introduction to trajectory analysis using Jupyter notebooks.
 2. An introduction to the use of PCA for data analysis.
 3. An introduction to the statistical analysis of simulation data.

Other notebooks contain “extension” exercises.