Molecular simulations of biomolecules
Remote Lecture #8

thermodynamic cycles
connections to experimental restraints

CCB 550, Spring 2020

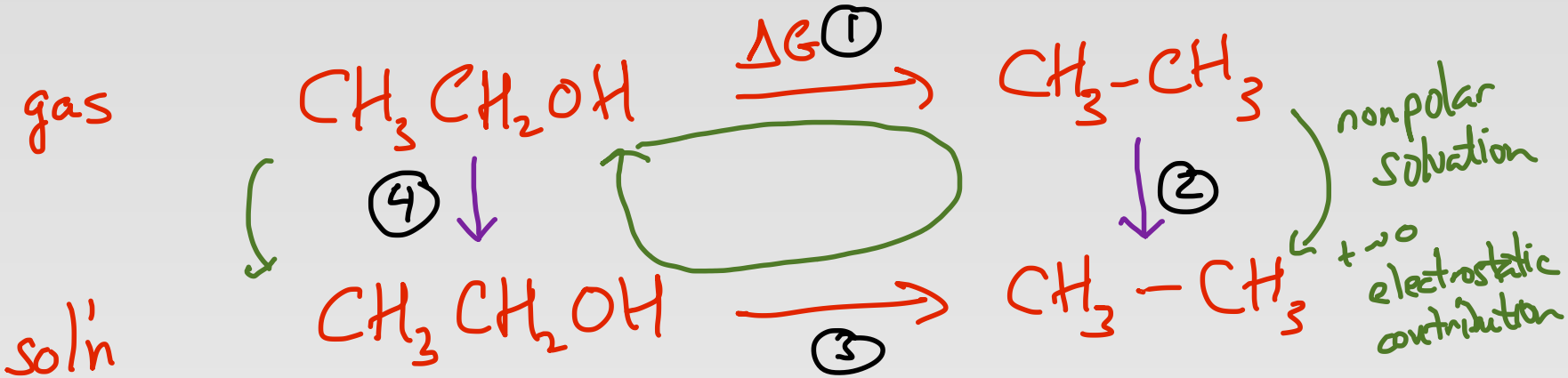
Types of thermodynamic calculations

- "alchemical" changes: $V_0 \rightarrow V_1$, using TI or FEP
  - end states can have different chemistry; or different environments; or just different force fields
- conformational free energy changes: using umbrella sampling or other enhanced sampling techniques
- "end-point" methods:

$$G \simeq < H_{MM} > - TS_{chain} + < \Delta G_{solvation} >$$

Thermodynamic cycles: transfer free energies

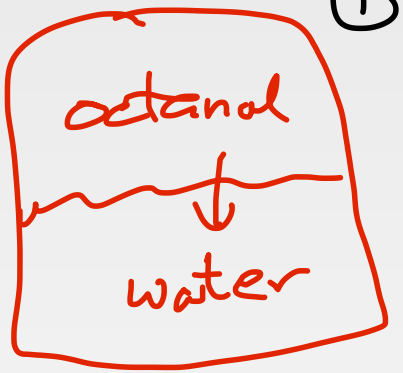
- See Fig. 12.3 in your text



difference in horizontal free energy

= " " vertical "

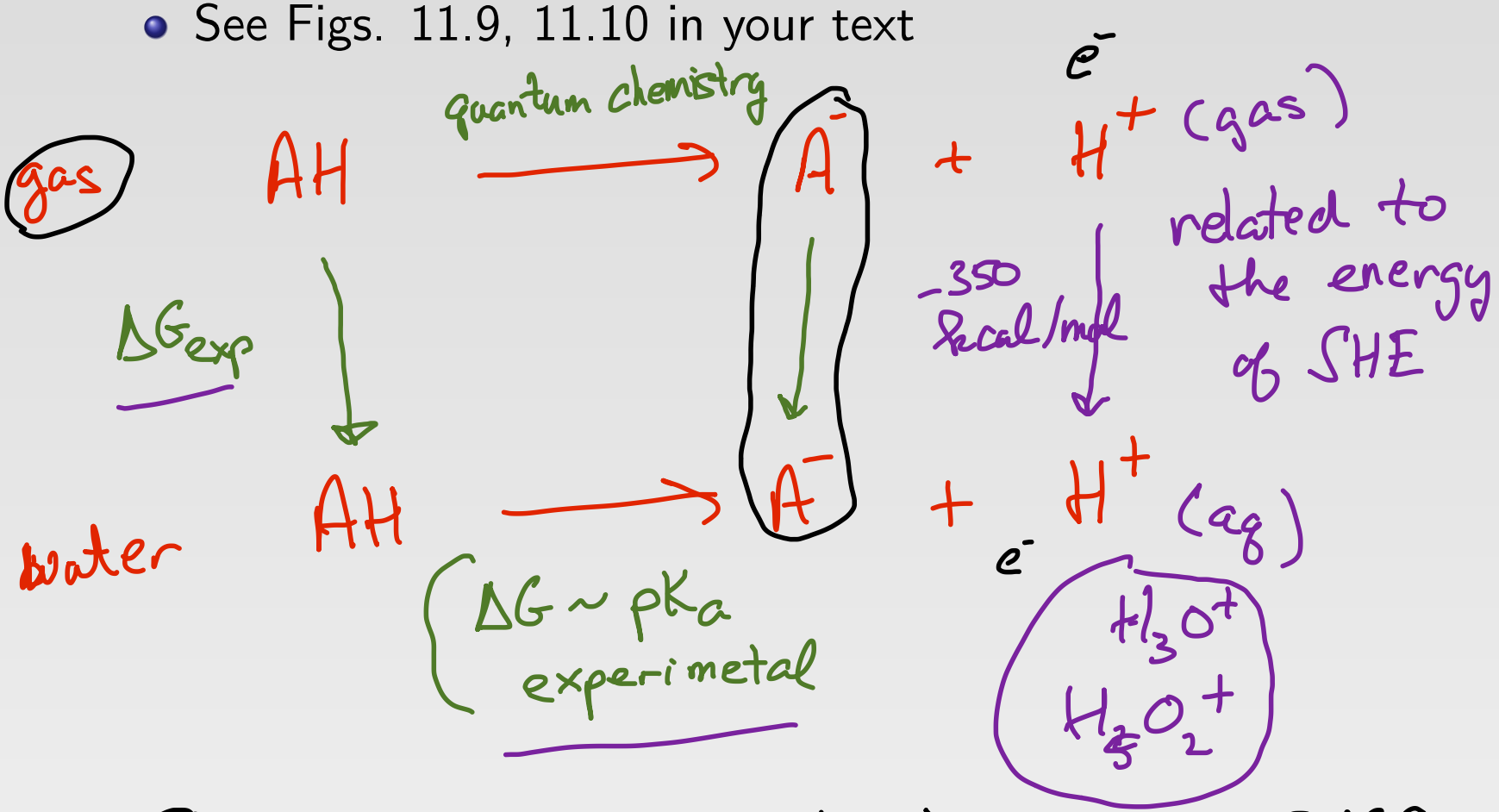
① + ② + ③ = ④



partition coefficient

Thermodynamic cycles:

- See Figs. 11.9, 11.10 in your text



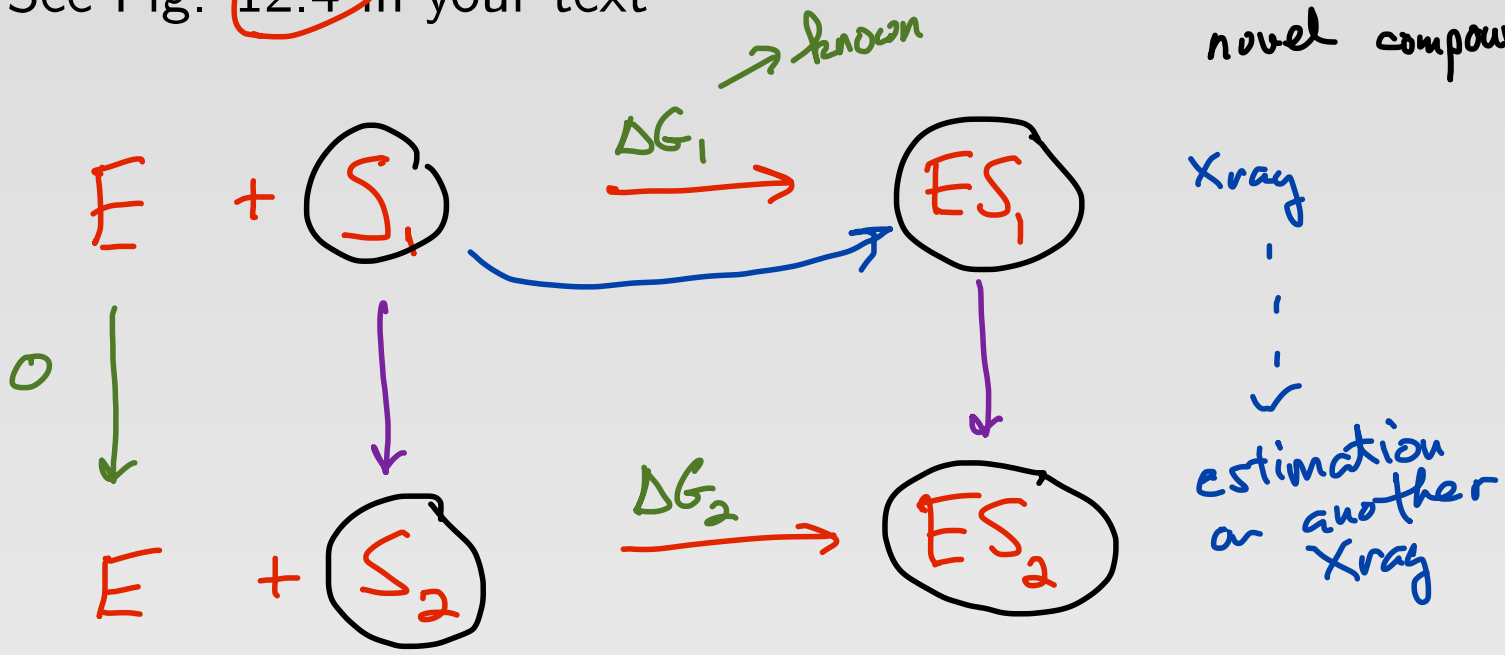
① want simulations to give right answer because we want to simulate polar → non-polar

Thermodynamic cycles: pH and redox behavior

inhibitor design

lead compound vs novel compound

- See Fig. 12.4 in your text



(1) alchemy : two purple arrows

(2) umbrella sampling : two horizontal lines

more recently gaining attention

Bayes theorem in probability theory

conditional probability

$$p(A \cap B) = p(A|B)p(B) = p(B \cap A) = p(B|A)p(A) \qquad (1)$$

prob. that both A and B are true

$$\boxed{p(A|B) = \frac{p(B|A)p(A)}{p(B)}}$$

posterior likelihood

prior

- Nomenclature (jargon): $p(A)$ is the "prior" estimate of the likelihood that A is true; $p(A|B)$ is the "posterior" estimate of likelihood, given evidence B
- philosophy of probabilities: frequentist interpretation vs. Bayesian ("degree of belief") interpretation

B represents additional information

Using force fields instead of conventional geometric restraints

contruct a model that — maximize ℓ this

$$p(model|data) = \frac{p(data|model)\,p(model)}{p(data)}$$

$p(data)$ → fixed constant

$$-\ln[p(data|model)] = \text{ML target}; \quad -\ln[p(model)] \approx E_{MM}/k_B T_{eff}$$

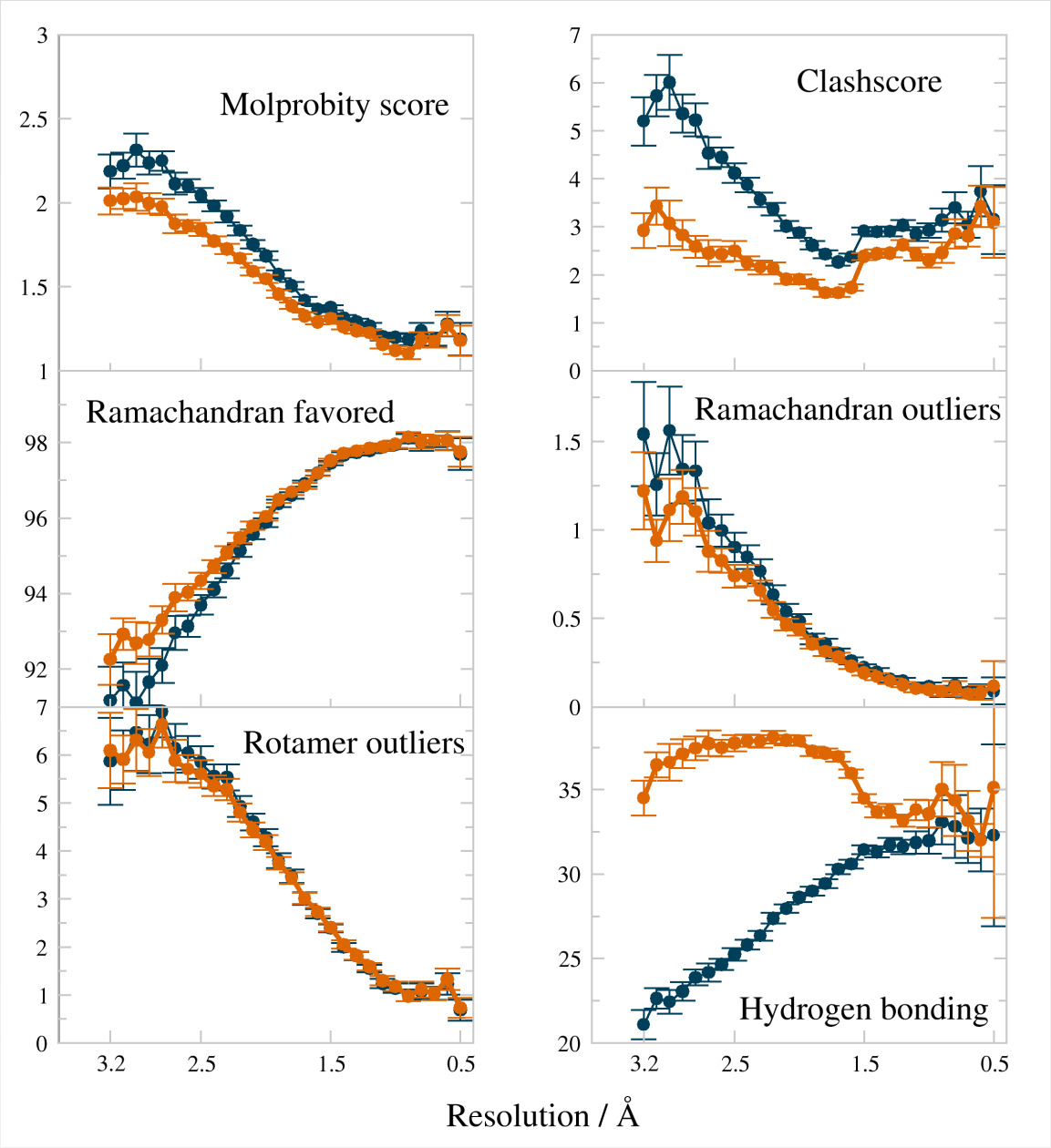
- Why would one want to do this? Or think it might work?
  - provides an approach to correlations among restraints; suggests a rational way to decide how much weight bonds vs. angles vs. clashes vs. torsions....
  - useful when restraints are hard to obtain via small molecule databases
  - a force field energy (eventually?) should optimally encode our prior knowledge about structure
  - especially at low resolution, energies should provide more restrictive restraints than simple geometric restraints

get best model: minimize $\sum_i (F_i^{calc} - F_i^{obs})^2$ + E_{MM}/kT (natural) + ML target (artificial term)

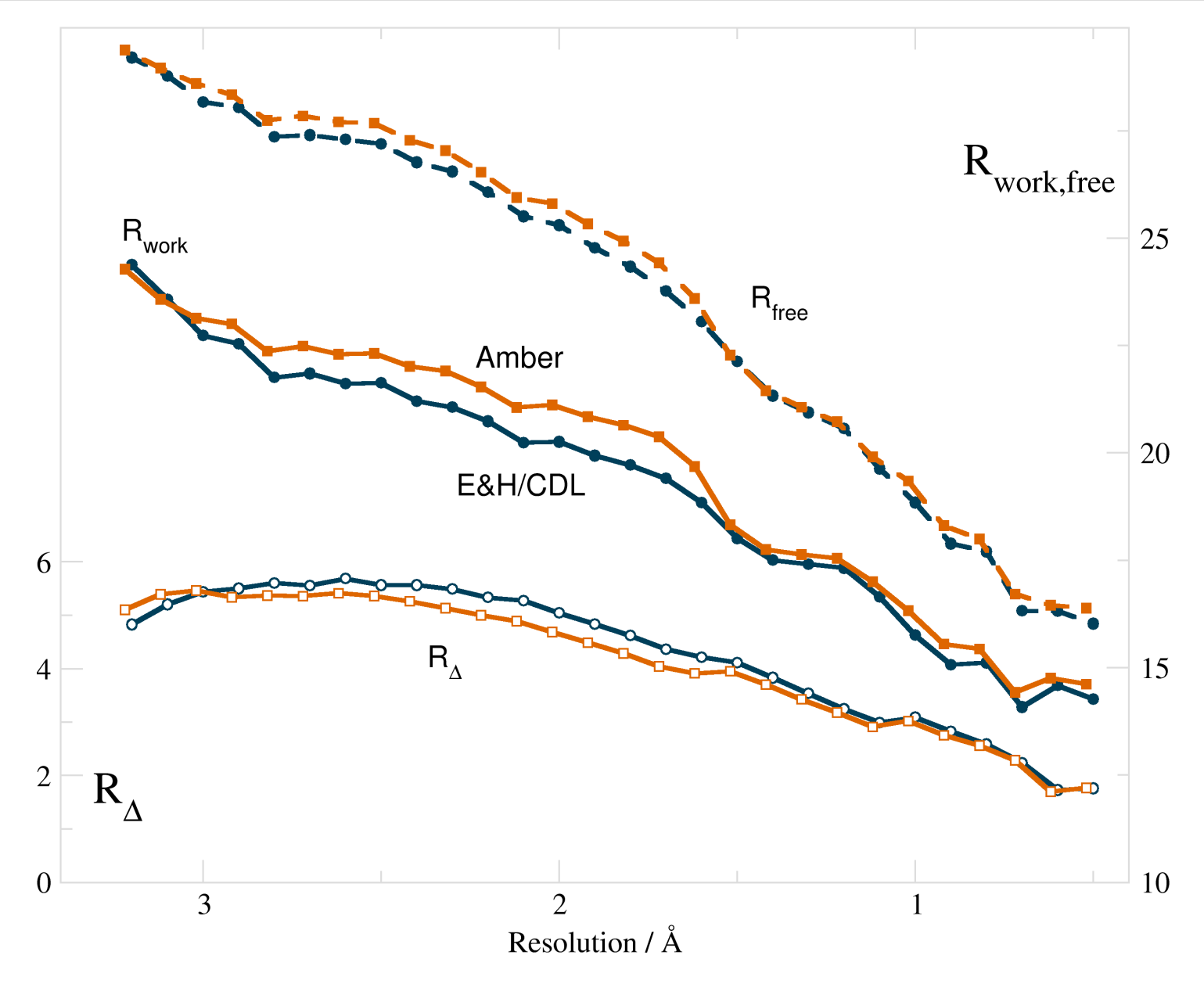
Wait: hasn't this all been done before?

- XPLOR (e.g. Karplus & Brünger, Acc. Chem. Res. 24:54, 1991); CNS; FFX (Schnieders, Fenn, Pande, Brünger, Acta Cryst D 65:952, 2009); PrimeX, QuantumBio, OpenEye, others....
- What can be done now, by someone with no Amber experience?
- `phenix.AmberPrep 9xyz.pdb <options>`
- `phenix.refine 4phenix_9xyz.pdb use_amber=True topology_file_name=4amber_9xyz.prmtop amber.coordinate_file_name=4amber_9xyz.rst7 amber.order_file_name=4amber_9xyz.order ....`

Results from 13,237 paired protein refinements



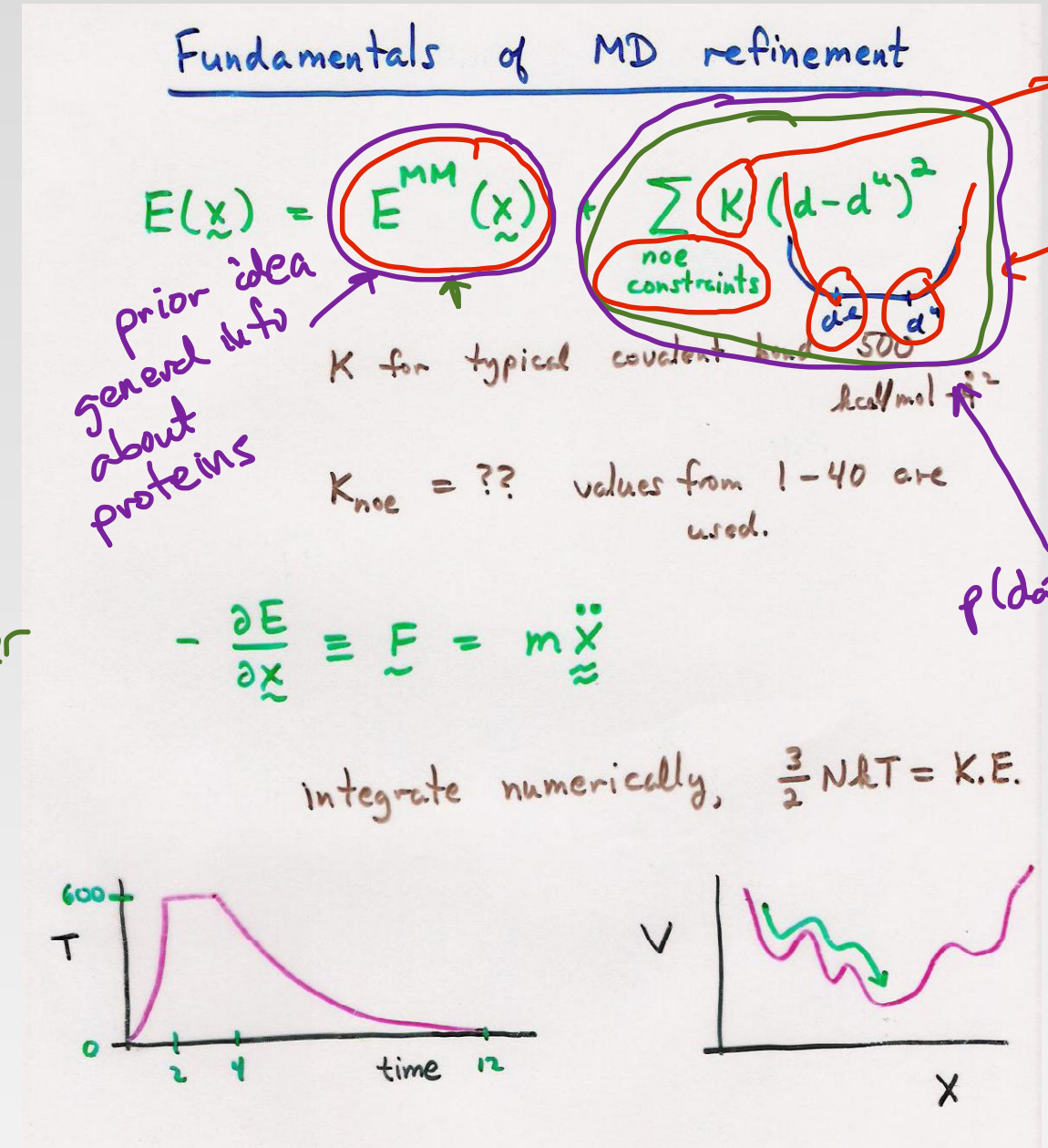
Results from 13,237 paired protein refinements



Molecular dynamics-based structure refinement

estimates distances between nuclei

there is a range of K that give almost the same answer



weighting between E^{MM} and data-driven constraints

$p(data|model)$

$\Downarrow$

$p(model|data)$

outcome of a refinement