Computer simulation approaches in structural biology

Mónika Fuxreiter
Institute of Enzymology
Budapest

Possible applications

- when the problem cannot be examined experimentally (transition states)
- to complement experimental data (eg. structure determinations)
- to improve resolution of experimental data (eg. build a model)
- to interpret experimental data (e.g. kinetic measurements on mutants)
- to study enzyme families (e.g. to study evolutionary relationships)

- **prediction** (structure, mutants, mechanism ...)

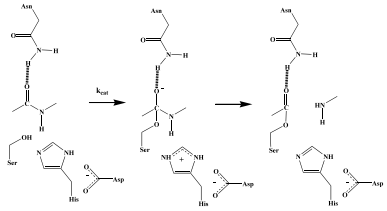
Problems in structural biology

1. Enzymatic mechanisms

	k_{cat} [s^{-1}]	
	enzyme	water
carbonic anhydrase	$6x10^5$	10^{-9}
acetylcholine esterase	$2x10^4$	$8x10^{-10}$
staphylococcal nuclease	10^2	$2x10^{-14}$

1. Enzymatic mechanisms

Serine proteases

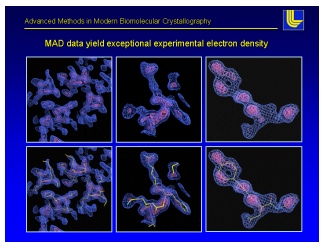


Enzymatic mechanisms
Problems

- Mechanism of action
  - description of the reaction steps
  - catalytic residues
  - rate limiting step
- Origin of catalytic power
  - proximity effect
  - steric stress
  - general acid/base catalysis
  - contribution of far-lying residues
  - gas phase like environment
  - Low Barrier Hydrogen Bonds
- Mutation effects

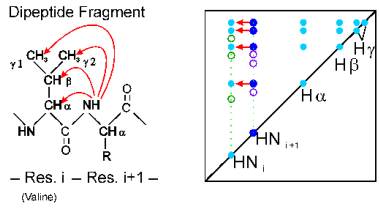
2. Structure determination

X-ray crystallography



Structure determination

NMR

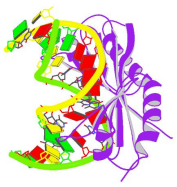


Structure determination

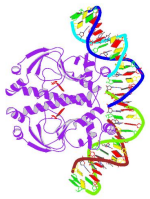
- bad parameter/data ratio
  - model building
  - refinement using force-fields
  - multiple conformers (low energy?)
- identification of missing parts
  - unsructured
  - loops
- conformational changes
  - substrate binding
  - inhibitor binding
  - cofactor binding

Examples

DNA binding proteins



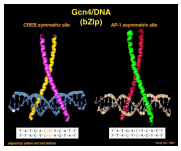
TATA box binding protein (TBP)

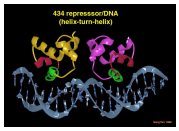


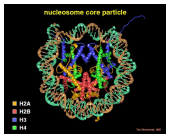
catabolic gene activator protein (CAP)

Examples

DNA binding proteins







Quantitative applications

- dynamical analysis of protein (or complex) structures
- refinement of crystallographic or NMR data
- enviromental effects (solution studies)
- kinetics of folding
- homology modelling (> 30%)
- vibrational spectra (low resolution)
- structural stability (differences)

- calculation of k_{cat} of enzymatic reactions
- ΔG of ligand binding (differences)
- mutational effects
- interpretation of catalytic effects

Validation

Comparison with experimental results

Structure determination

X-ray crystallography: comparison of computed and measured structure factors (R factor)

Comparison with homolog proteins

likely mechanism of action

Interpretation of catalytic effects

reproducing k_{cat}

reproducing k_{cat}(muténs)/ k_{cat}(natív) data

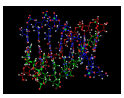
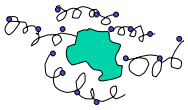
explaining pH dependence

explaining dependence on ion strenght

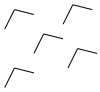
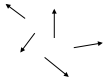
prediction

Modelling models

Quantum Mechanics | Molecular Mechanics | QM/MM

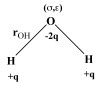
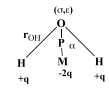
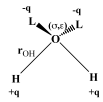




All Atom | Dipoles | Continuum Models

Models Water

3 point model | 4 point model | 5 point model

SPC, SPC/E,TIP3P | TIP4P polarizable: BSV, Chialvo-Cummings, Dang-Chang, PPC | ST2, TIP5P

Reproduce:
SPC, TIP4P: structure, thermodynamics, dynamics
SPC/E: average polarizability of molecules
BSV: dielectric constant, structure
Density maximum: BSV, PPC (SPC, SPC/E no max., TIP4P –30 °C max.)

Strategies

Principles

- The model has to match with the chemical nature of the problem (QM is not always the best)
- Consider the global problem. (define the questions)
- All important parts have to be represented. (environmental effect)
- profit/cost optima
- **The error in different parts of the model has to be similar for the given problem.**

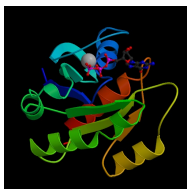
Model building Errors

- wrong iterpretation of the catalytic effect
- wrong environment (e.g. missing ions)
- small environment (e.g. *ab initio* calculations)
- gas phase (neutrality)
- incorrect water environment (crystallographic waters)
- wrong treatment of electrostatic effects (ε_{in}, pK_a)
- ignoring dynamic effects

Model building Examples

Ras p21

$$GTP + H_2O \rightarrow GDP + HPO_4^{2-}$$



- What is the mechanism?
- Proton transfer pathways?
- Role of Lys-16?
- Role of Gln-61?
- Role of substrate in catalysis?
- Why cofactor is required?

Ras p21 Possible mechanism

1. Activation of attacking nucleophile (water)?

A deprotonation by Gln-61
B deprotonation by another water
C not deprotonated (neutral water attacks)
D concerted mechanism
E γ-phosphate is the general base
F proximal Asp or Glu (Asp-57) deprotonates

Ras p21
Mechanisms

Experimental results

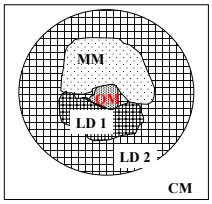
- Mutational data: Gln-61, Gly-12, Gly-13 drop in activity (10^{-3})
- Structure: all are in proximity to γ phosphate
- NMR: pKa of the γ phosphate is affected by the mutations
- Kinetics: k_{cat} const. if pH > 3, and dependenton pKa of the γ phosphate

⇩

Hypothesis: catalysis occurs with substrate assistance

Ras p21
Studies on mechanism

Warshel et al. Nat. Struct. Biol. (1994) 1 pp. 476 – 484.



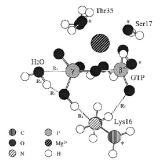
Method: EVB/FEP

Activation barrier calculation for
all possible mechanisms including
all sidechains and the whole environment

Conclusion: it is the γ phosphate that
deprotonates the attacking water (**E**)

Ras p21
Studies on mechanism

Futatsugi et.al. *Biophys. J.* (1999) **77** 3287-3292.



Method:
- structure: optimized enzyme
- *ab initio* calculations on ground and transition state

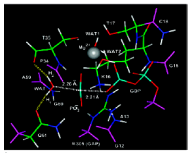
Conclusion:
- ″proton ping-pong″ with the assistance of Lys
- Gln helps to position of the catalytic water

Error:
- assume the protonation state of Lys and GTP
- missing enzymatic environment

Ras p21
Studies on mechanism

Cavalli et al. J. Am.Chem.Soc. (2002) 124, pp.3763-3768.

DFT-MD (Carr-Parinello)



- AlF_3 TS analog

- 85% of the electric field of the protein is represented
- 2 ps MD 60000 CPU hours
- O_w-Pγ 1.8 Å distant constraints
- when released proton jumps to O of Gln61

Error:

irreal pKa values,
no comparison with solution reaction
Gln→Glu mutation cannot be explained