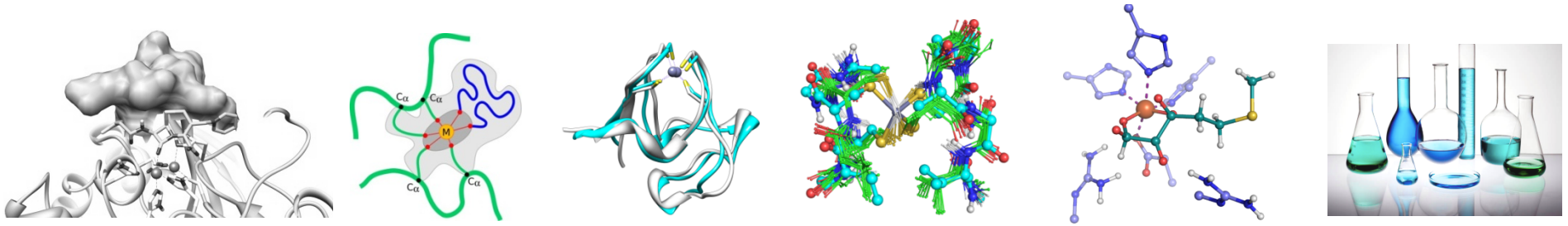
Inside-out design of artificial enzymes led by multi-scale modeling



Anastassia Alexandrova

University of California, Los Angeles & California NanoSystems Institute

Overview

1. Why design?

2. Non-metal enzymes

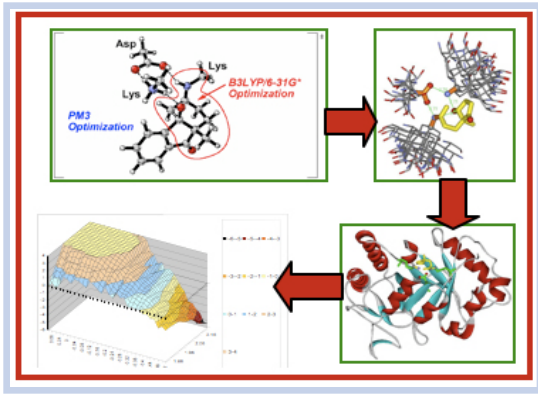
70 % - computation
30 % - experimentation

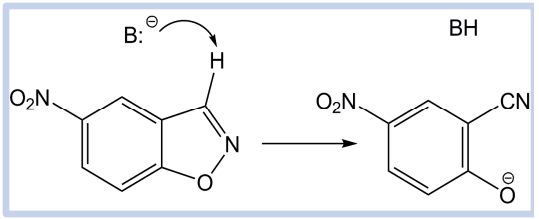
3. Metalloenzymes – harder

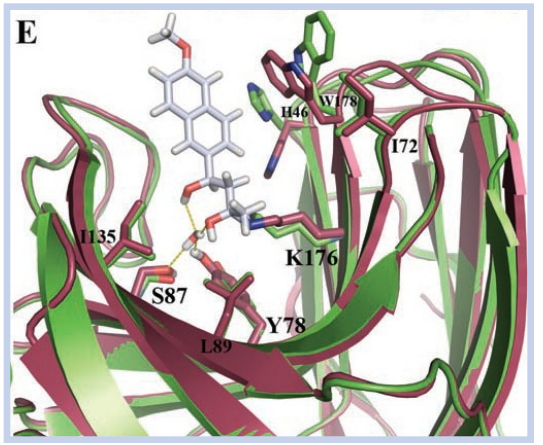
Serious QM, and very serious sampling
Multi-scale most desired

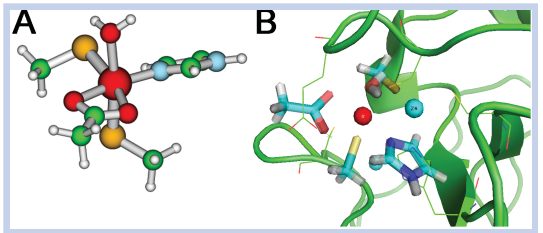
4. Protein-protein interfaces

Pure MM, all-atom, and coarse graining

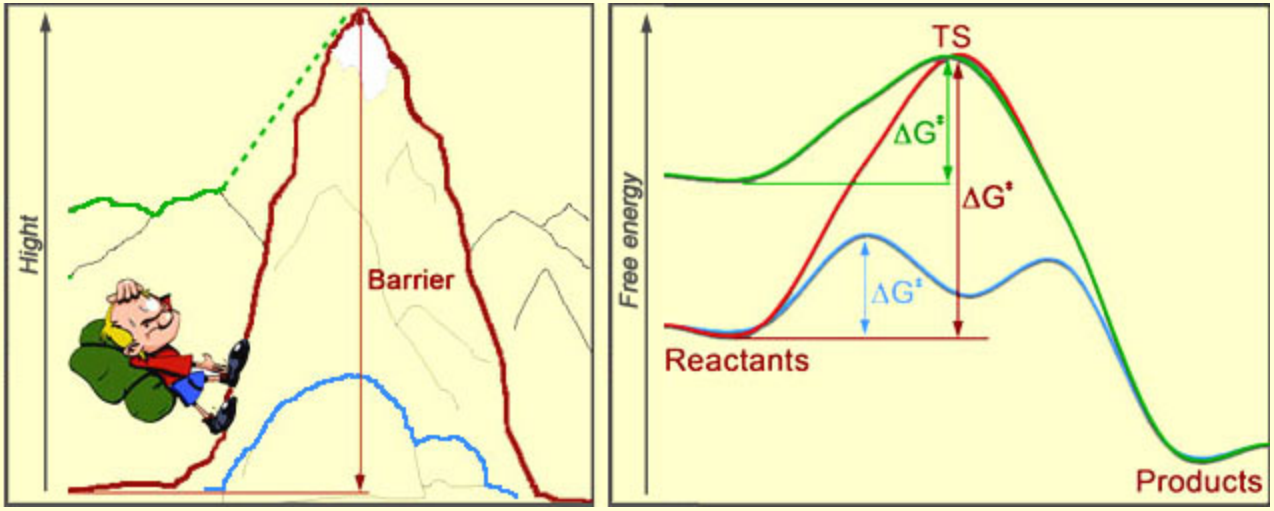




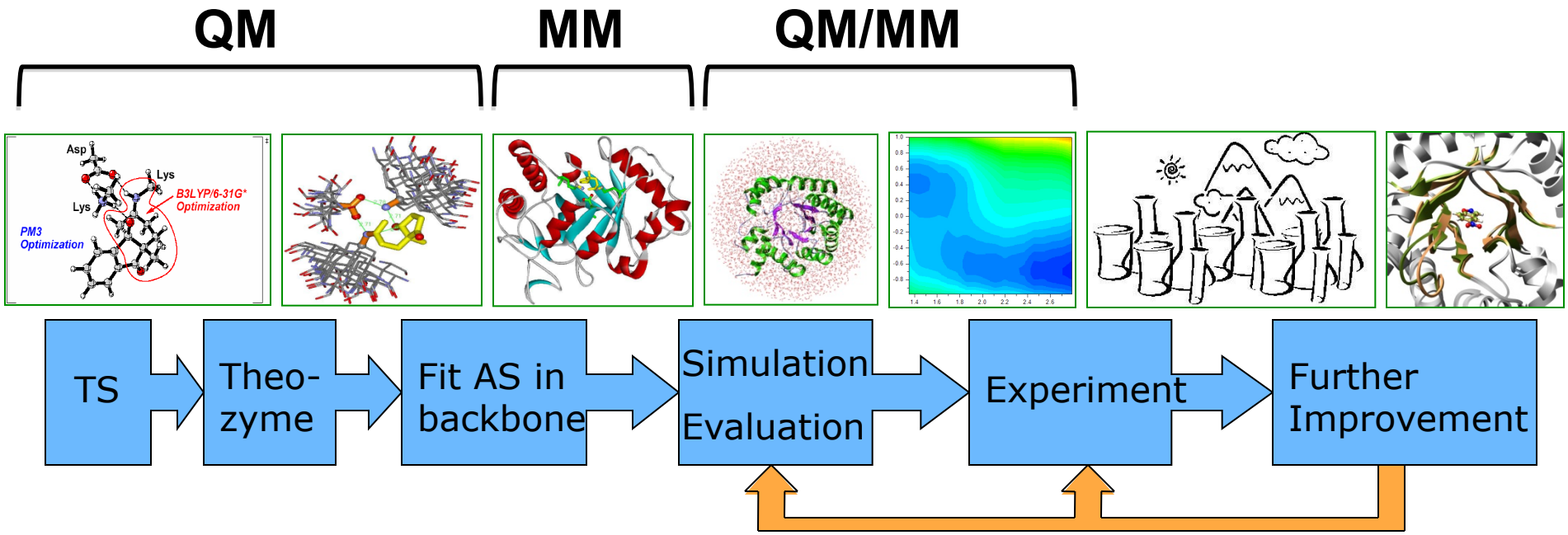




How catalysis works

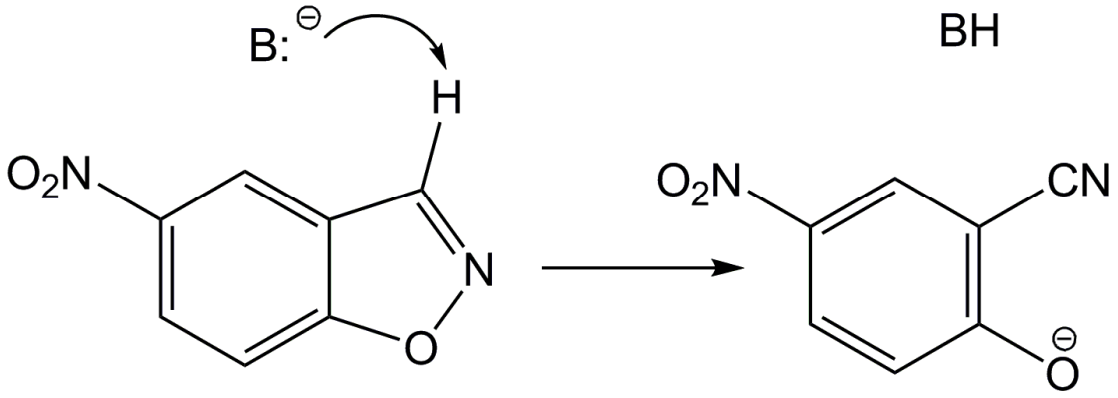


The inside-out design protocol



D. Rothlisberger et al. *Nature* 453 (2008), 109-195
A. Alexandrova et. al. *JACS* 130 (2008), 15907-15915
A. Alexandrova et al. *Mol. Sim.* 37 (2011), 557-571

The "sandbox" reaction: Kemp elimination

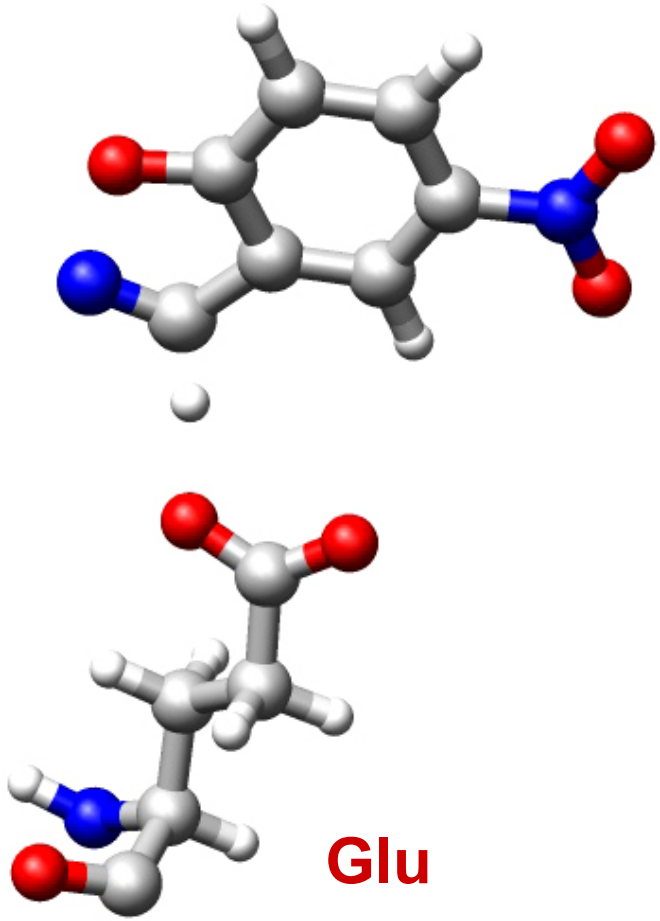


In the protein, we need:

1. General base: Asp, Glu, His, catalytic dyad/triad
2. H^+-source to complete the reaction: e.g. $LysH^+$, Ser
3. Other stabilizer(s) of the negative charge on the O

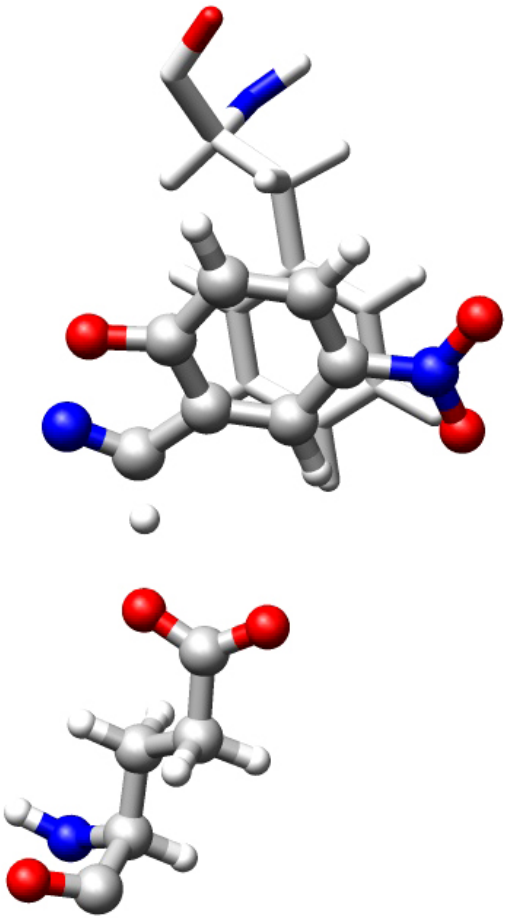
Let's assemble a theo-zyme

1. Pick a catalytic base
2. Find the TS (DFT)

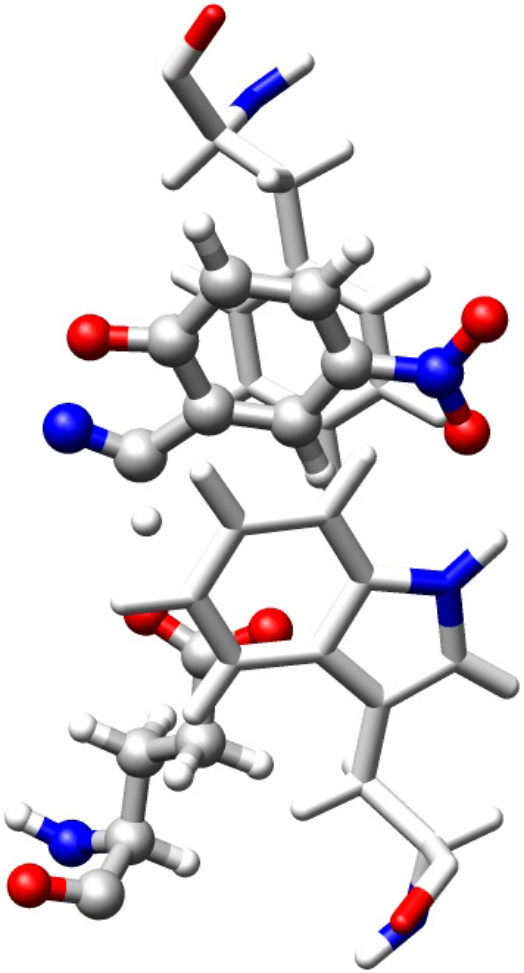


Let's assemble a theo-zyme

3. Stabilize the TS via π-stacking: orientation, and charge stabilization



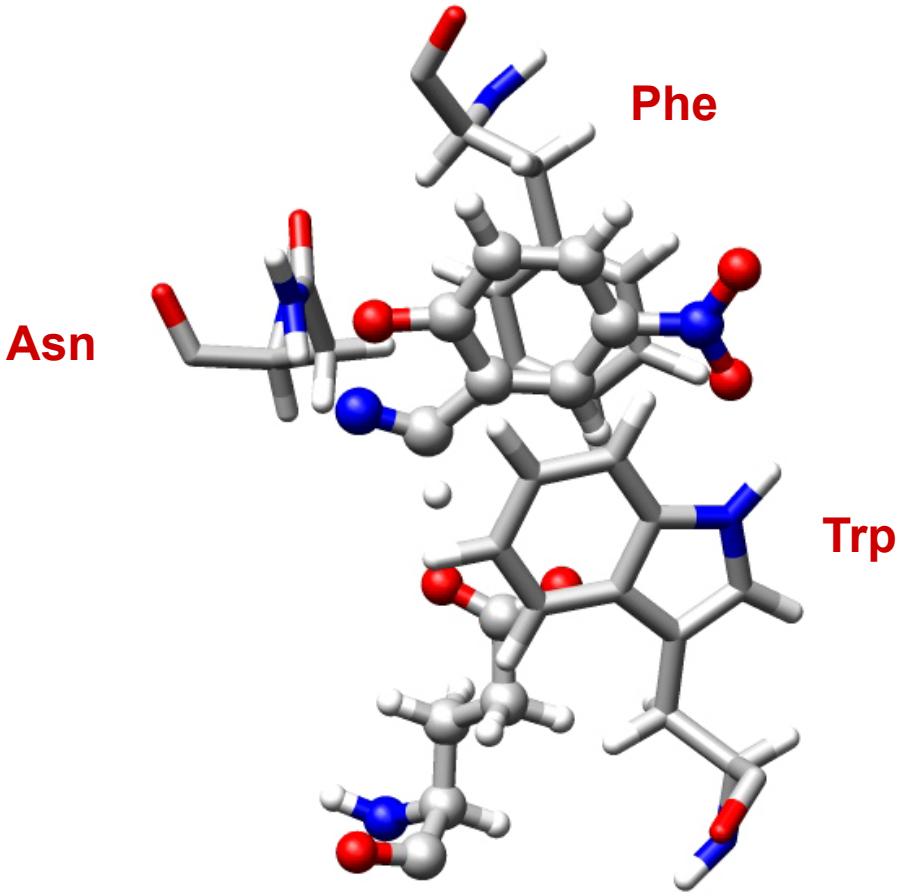
Phe on the back



Trp on the front

Let's assemble a theo-zyme

4. Stabilize the TS via hydrogen bond to O^-



Choosing a host protein

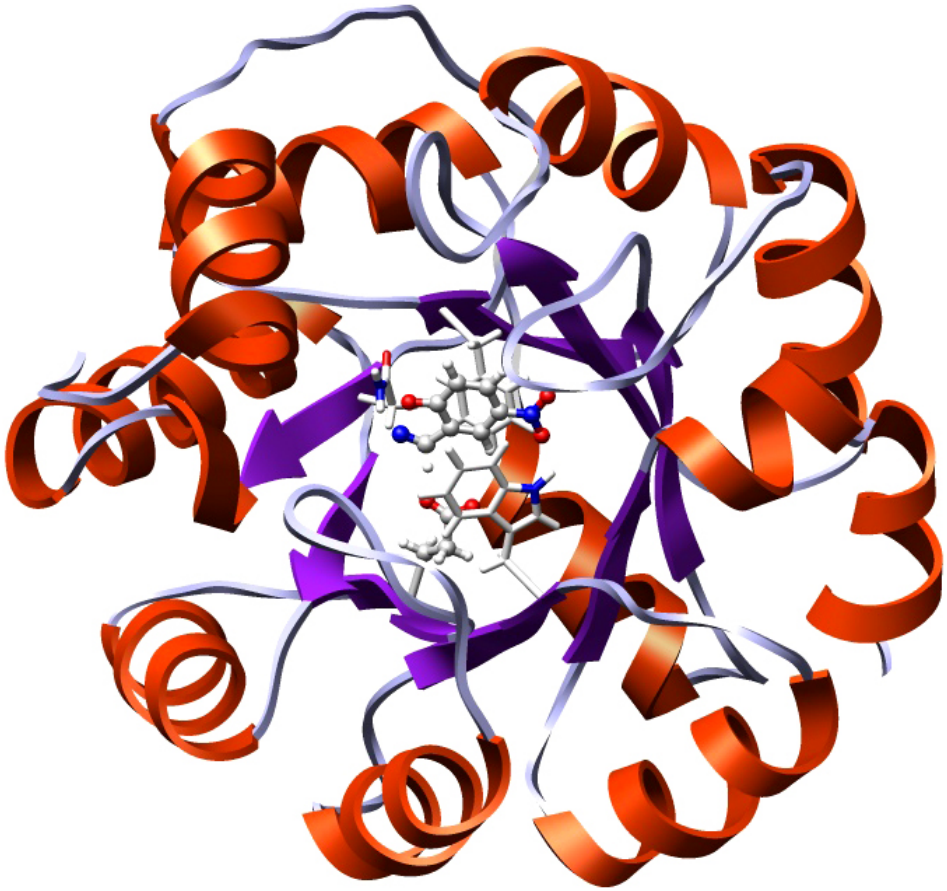
PROTEIN FOLD SHOULD:

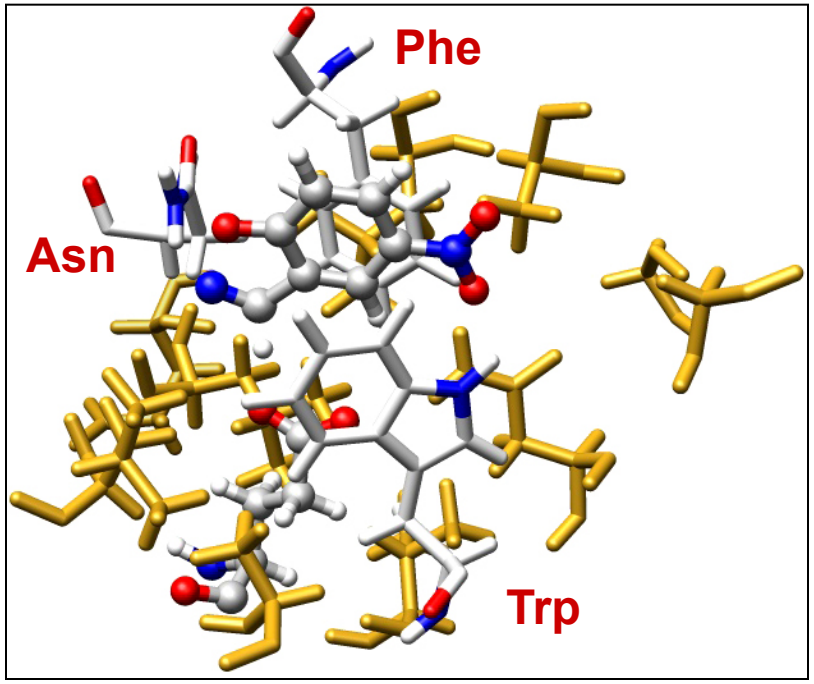
1. Have a cavity
2. Be able to accommodate catalytic residues and the substrate
3. Be easily expressible in *E. Coli*
4. Be stable under a few mutations

PROCEDURE:

1. Catalytic residues are placed at potential catalytic positions
2. Substrate is put in
3. *ROSETTA, Eris*: automated re-building and packing of the remainder of the binding site

Completing the design - MM



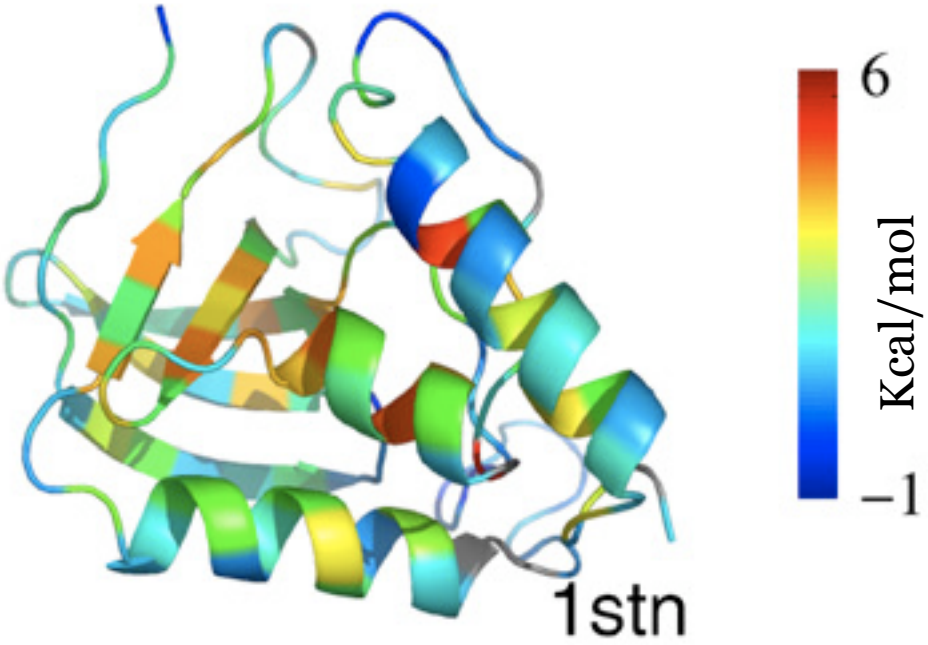


Yellow stuff: Leu, Ile, Val, Ala, Gly

Classical Force Field

Empirical scoring function: bonuses for H-bonding, π-stacking, helixes, dehydration of hydrophobic residues; penalties for exposed hydrophobic residues, etc.=> ΔG for the change in protein stability upon mutation

Eris – predicting protein stability upon mutagenesis



$$\Delta\Delta G = \Delta G^{mutant} - \Delta G^{wildtype}$$

Positive number = disfavorable mutation
Negative number = favorable mutation

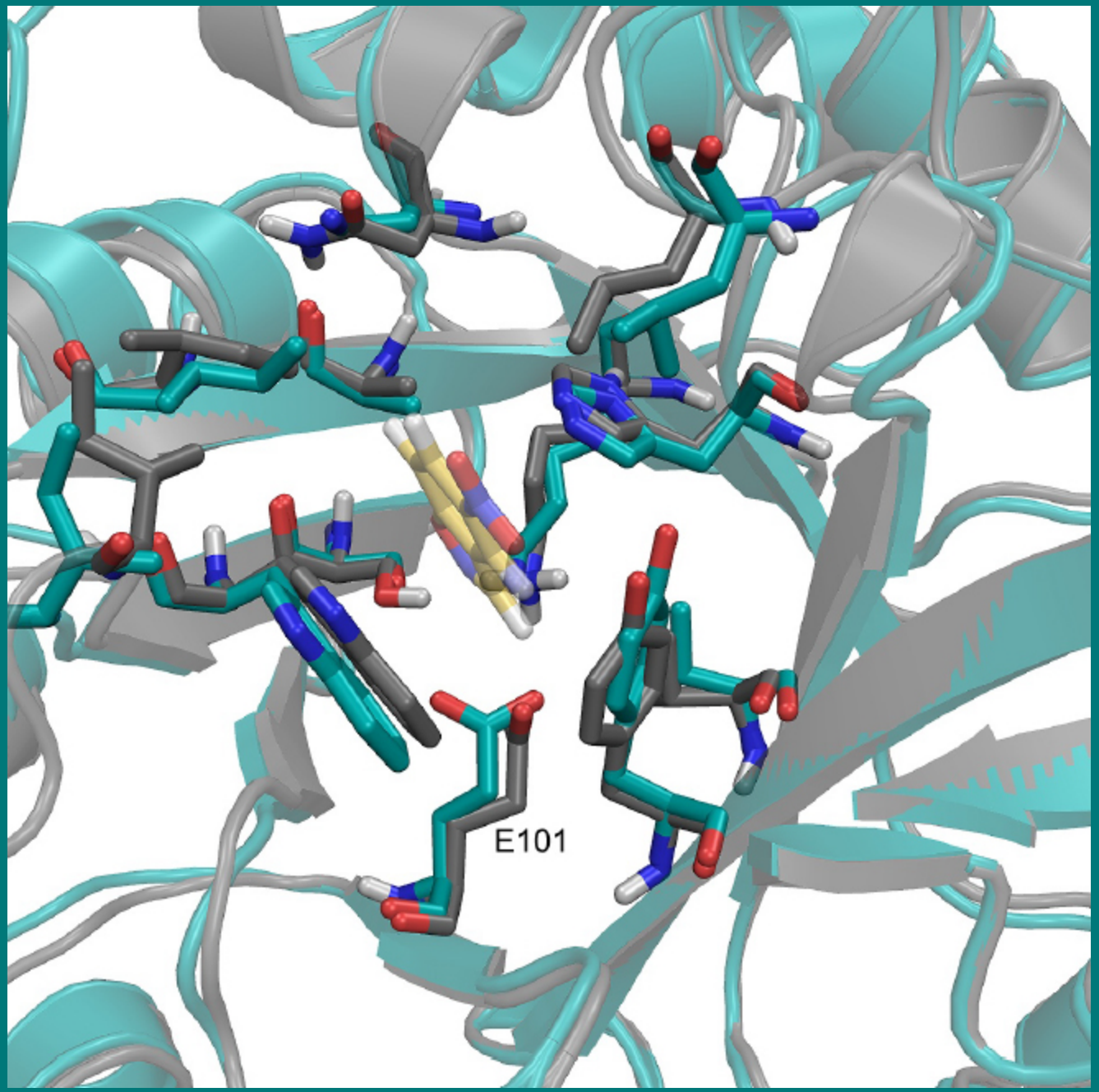
1. Yin, S., Ding, F., and Dokholyan, N. V., "Modeling backbone flexibility improves protein stability estimation", Structure, 15:1567-1576, (2007).

Eris – predicting protein stability upon mutagenesis

$$\Delta E = W_{vdw_attr} E_{vdw_attr} + W_{vdw_rep} E_{vdw_rep} + W_{solv} E_{solv}$$
$$+ W_{bb_hbond} E_{bb_hbond} + W_{sc_hbond} E_{sc_hbond} + W_{bb_sc_hbond} E_{bb_sc_hbond}$$
$$+ W_{aa|\phi,\psi} E_{aa|\phi,\psi} + W_{rot|\phi,\psi,aa} E_{rot|\phi,\psi,aa} - E_{ref}$$

1. Yin, S., Ding, F., and Dokholyan, N. V., "Modeling backbone flexibility improves protein stability estimation", Structure, 15:1567-1576, (2007).

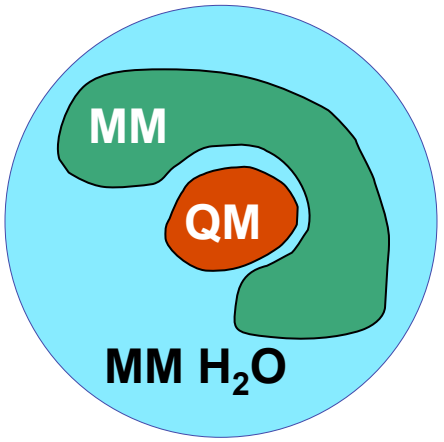
KE07: crystal structure (cyan) vs. design (grey). $RMSD_{backbone}$ = 0.32Å, $RMSD_{overall}$ = 0.95Å



Testing methodology: QM/MM Monte Carlo

System:

chopped protein (ca. 300 residues)
22 Å TIP4P water cap: (400-500 water molecules)
QM-region: the substrate and the base
MM-region: the rest of the protein and solvent



QM/MM Metropolis Monte Carlo at 25˚C and 1 atm:

5 M solvent equilibration, 10 M full eq., 25 M sampling

Degrees of freedom:

QM-region - bond lengths, angles, dihedrals
MM-protein - angles and dihedrals
Solvent - translations and rotations

QM: PDDG-PM3 - Pair-wise Distance Directed Gaussian

(A single small Gaussian function is added to pair-wise core repulsion function)

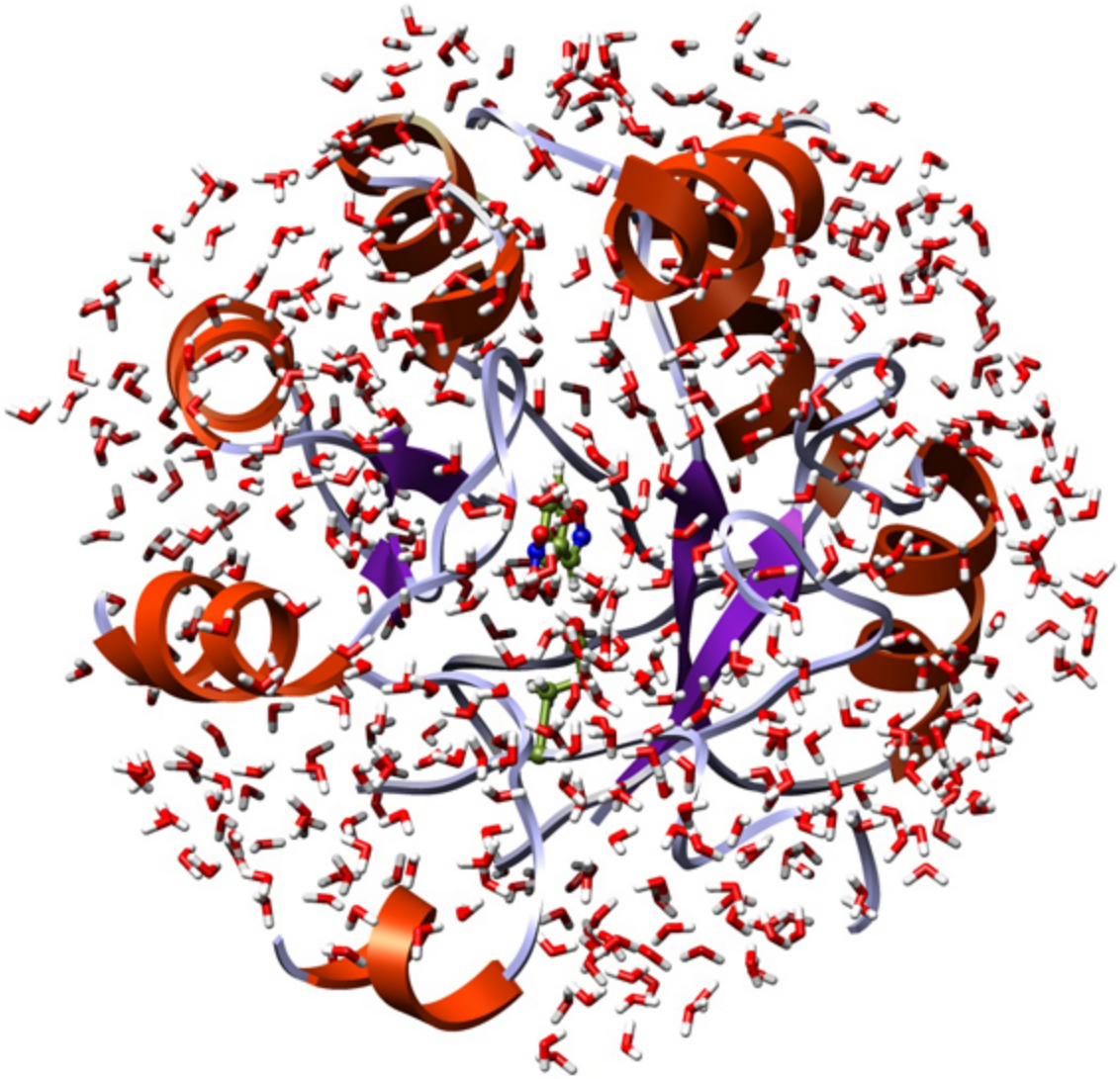
$$U_{ss}, U_{pp}, \beta_s, \beta_p, \zeta_s, \zeta_p, \alpha, a_1, b_1, c_1, a_2, b_2, c_2, P_1, P_2, D_1, D_2$$

MNDO **PM3** **PDDG**

MM: OPLS-force field

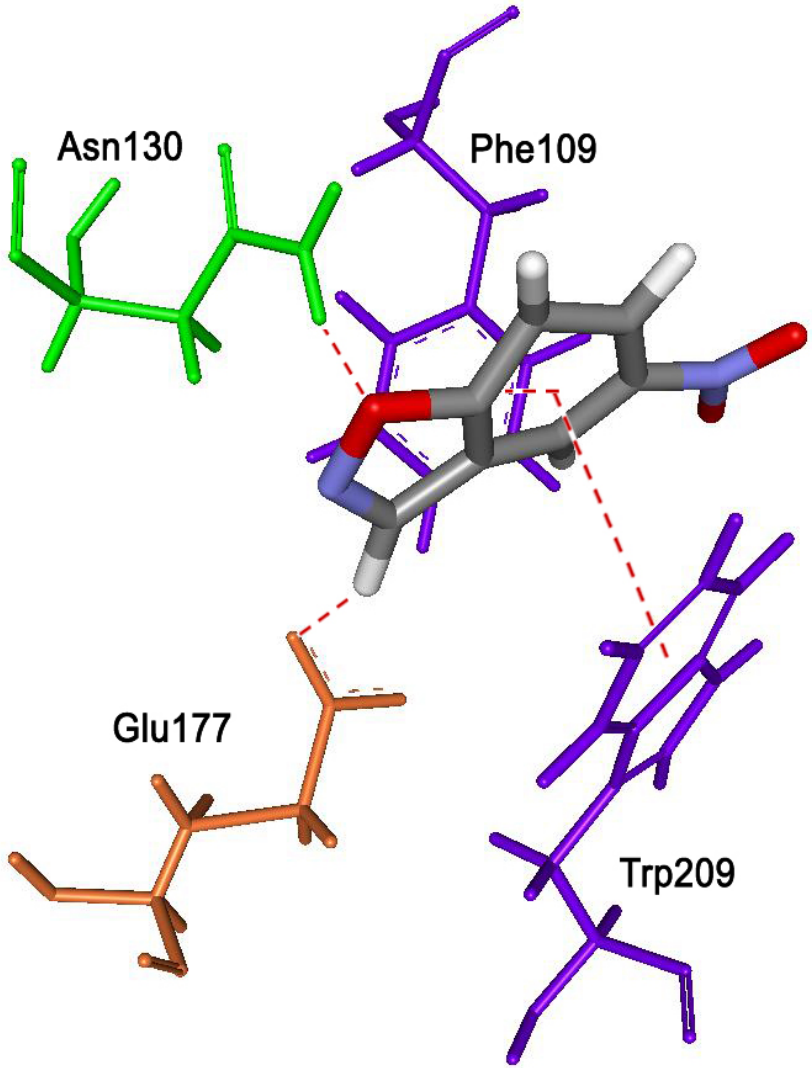
QM-MM interface: link atom approach, CM charges for Coulomb, and LJ

System prepared for simulation

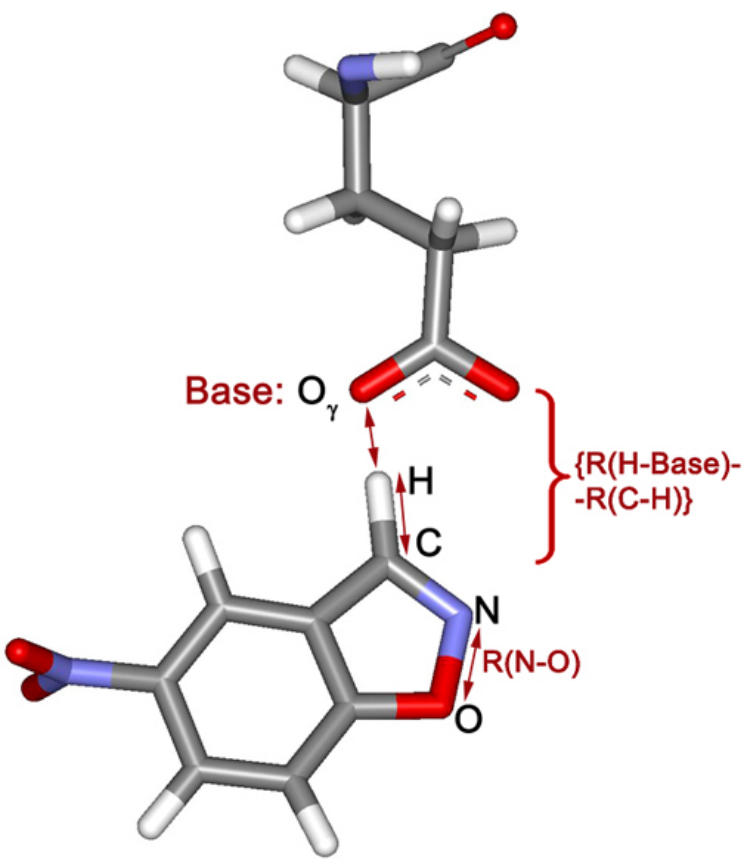


QM/MM MC relaxation with the substrate

1) π-stacking is present
2) Glu177- correct protonation state
3) Asn130 interacts with O
4) No water at the reaction site



QM/MM MC + Free Energy Perturbation



1. Reaction coordinates:

R(N-O), {R(C-H)-(R(O-H)}

2. Drive the reaction along the reaction coordinates: increment of 0.02-0.04 Å

3. compute ΔG per each step:

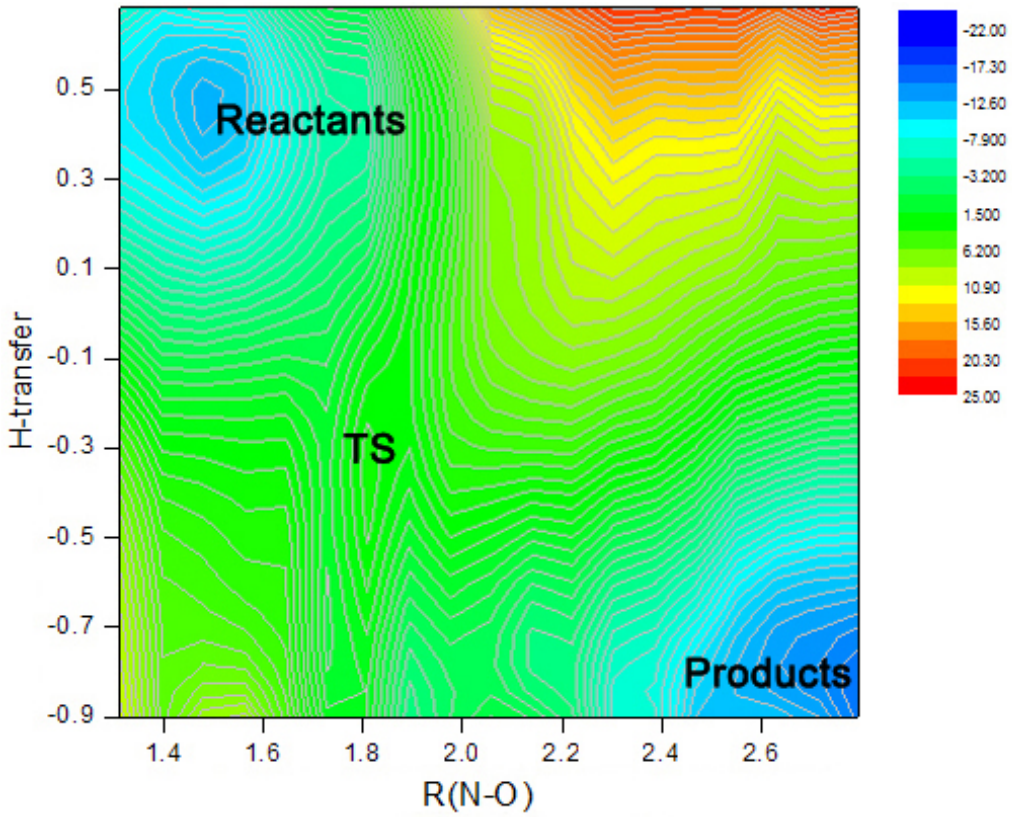
$$\Delta G(AtoB) = G_B - G_A = -k_b T \ln \left\langle \exp\left(-\frac{E_B - E_A}{k_b T} \right) \right\rangle_A$$

50 x 10^6 conf. QM/MM Monte Carlo at each step.

4. construct 1-D or 2-D free energy maps
(Potential of Mean Force maps)
→ $\Delta G^{\ddagger}$

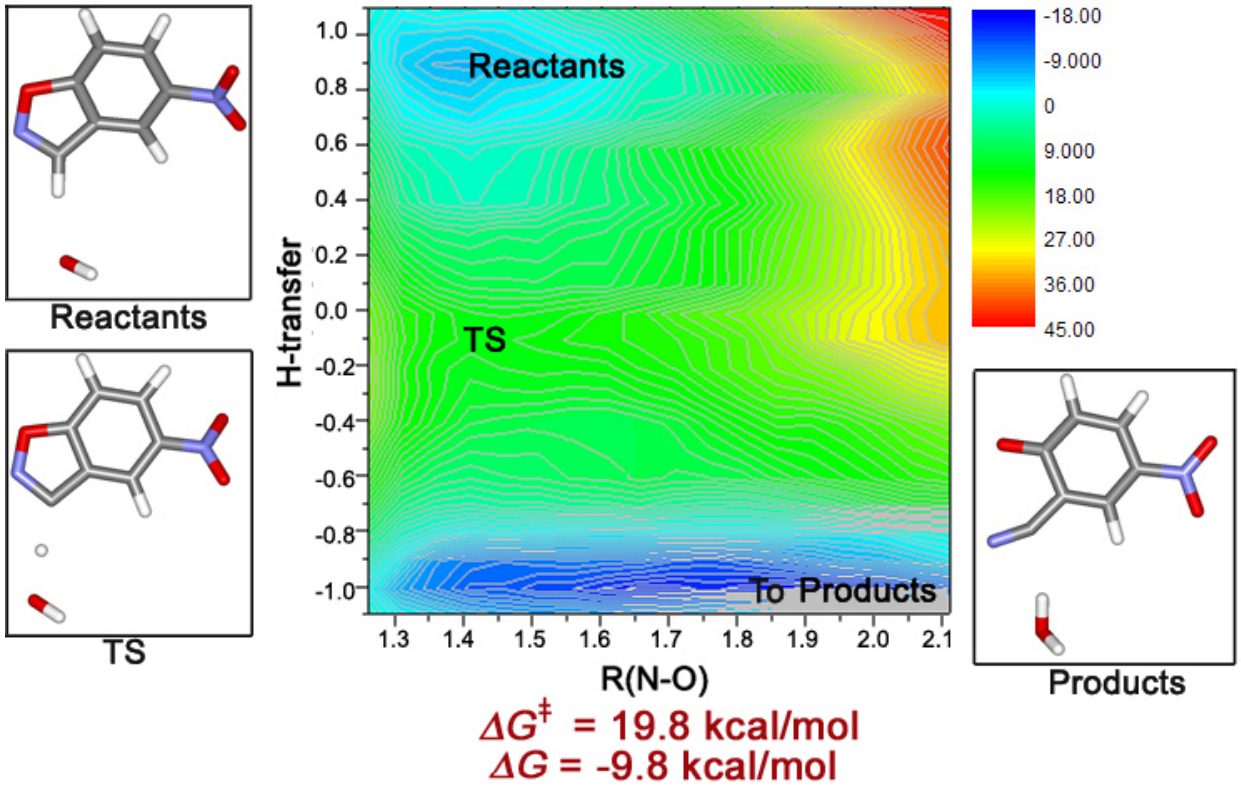
Free energy map for Kemp elimination in 1a53

Computationally intense: takes ca. 1 week to generate such a map (50 3GHz processors)



Concerted, single-step mechanism.
Asymmetric TS. H-transfer driven process.
$\Delta G^{\ddagger}$ = 13.5 kcal/mol ΔG = -4.6 kcal/mol

Uncatalyzed reaction in water

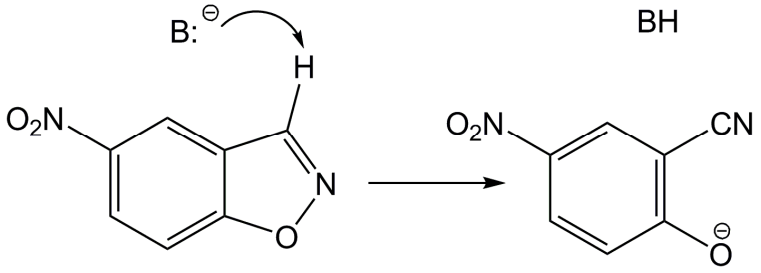


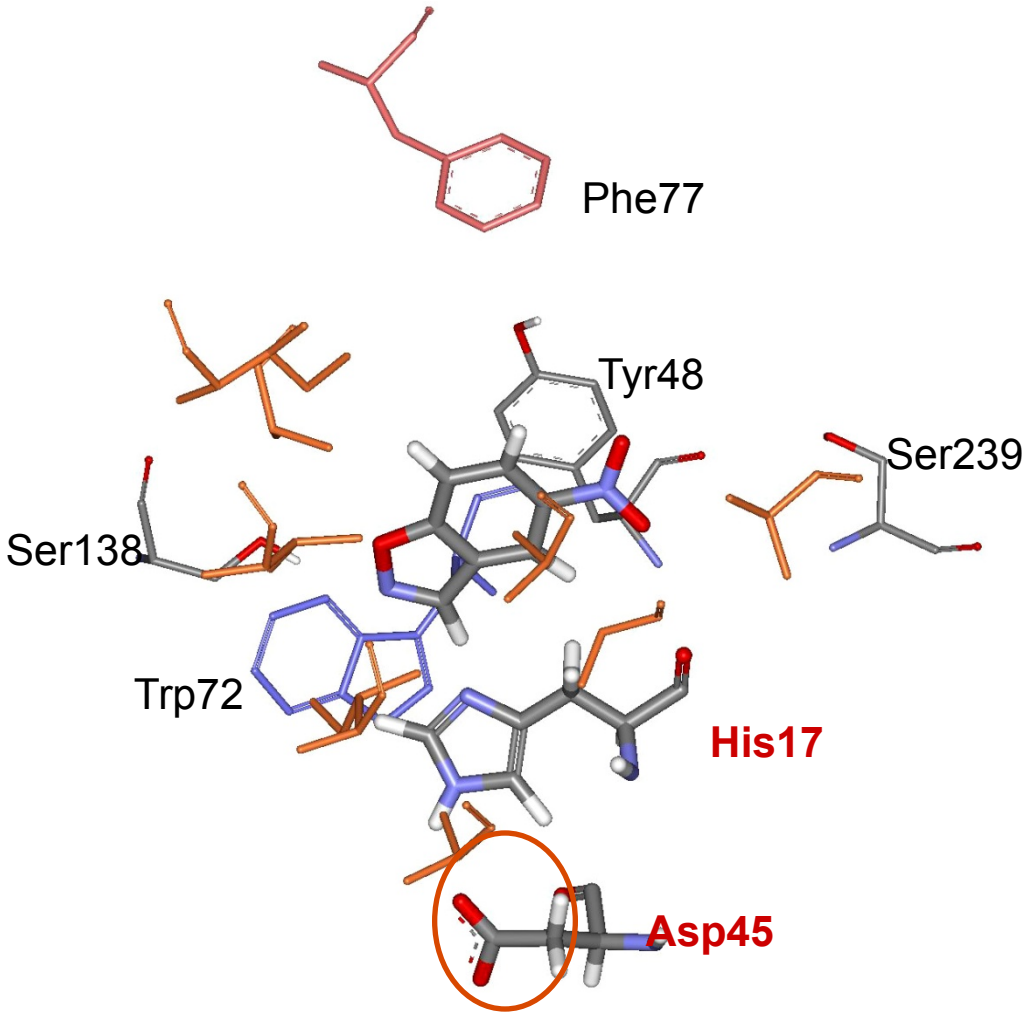
compare to ΔG‡ = 13.5 kcal/mol => 1a53 is an active enzyme

1a53 was expressed and purified. Activity was confirmed experimentally.
Experimental k_{cat}/k_{uncat} is 2.5×10^5

The best designed Kemp eliminase

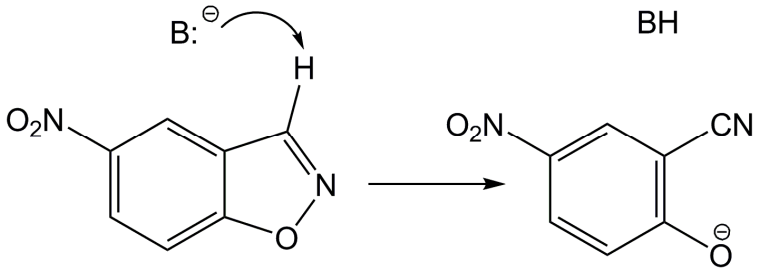
D. Rothlisberger et al. *Nature* 453 (2008), 109-195





$k_{cat}/k_{uncat} = 1.2 \times 10^6$

Kemp eliminases: what's important?

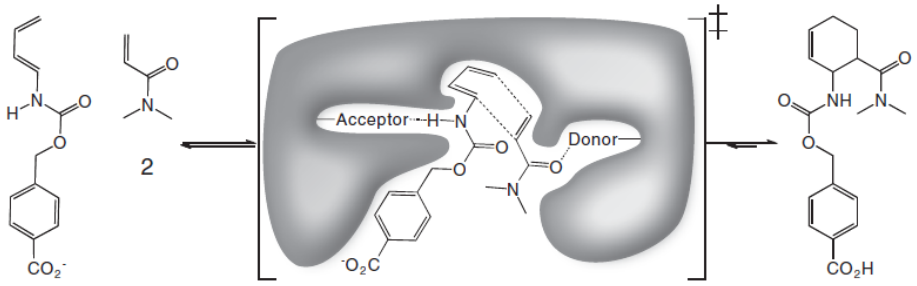


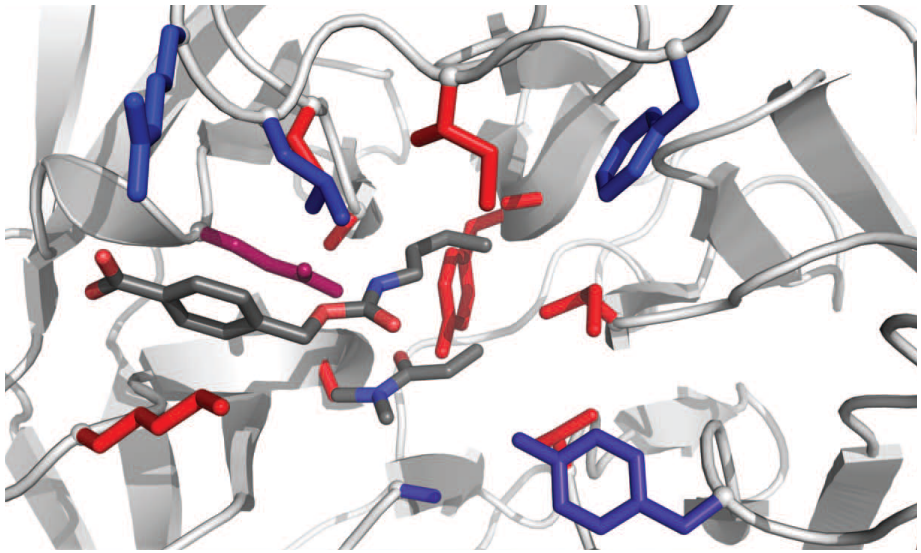
1. Strength of the catalytic base
2. Tightness of the contact with the base
3. Dynamic stability/immobility of the contact
4. Hydrophobic environment

Baker, Houk, Jorgensen, Alexandrova, Mayo, Hilvert, etc.

Artificial enzymes beyond sandbox

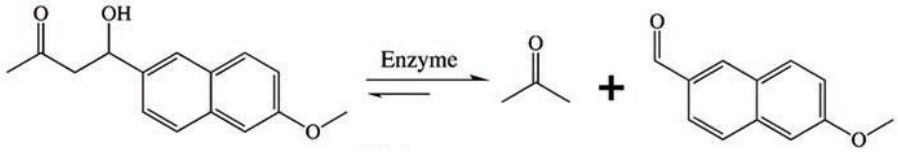
Diels-alderases

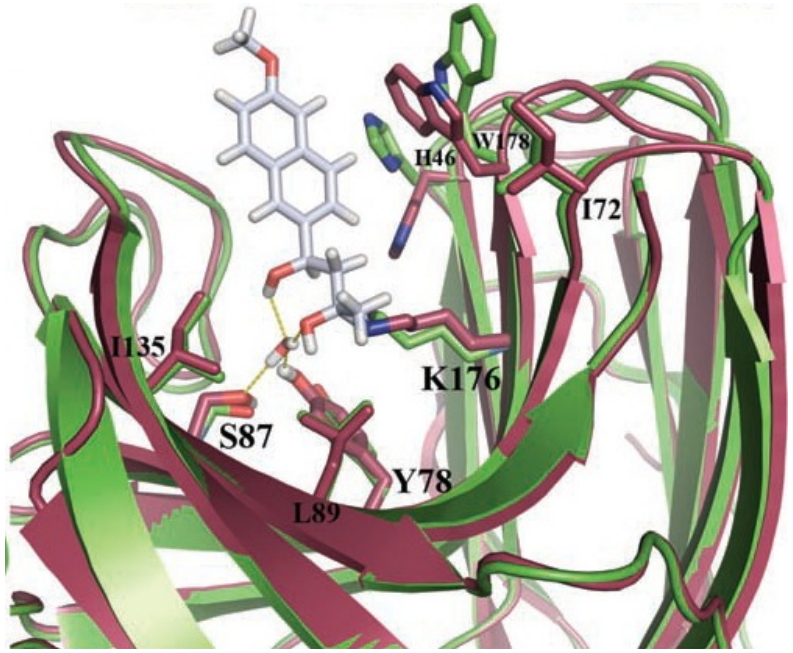




16 JULY 2010 VOL 329 SCIENCE

Retro-aldolases



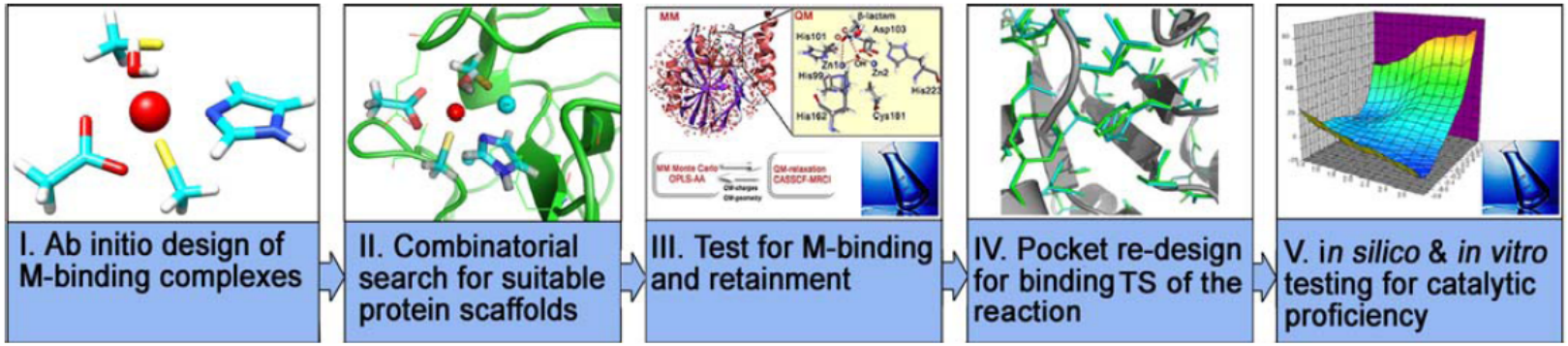


SCIENCE VOL 319 7 MARCH 2008

Used methods: MM, QM, QM/MM

Metalloenzymes are harder

Heavy role of QM
Extensive sampling!



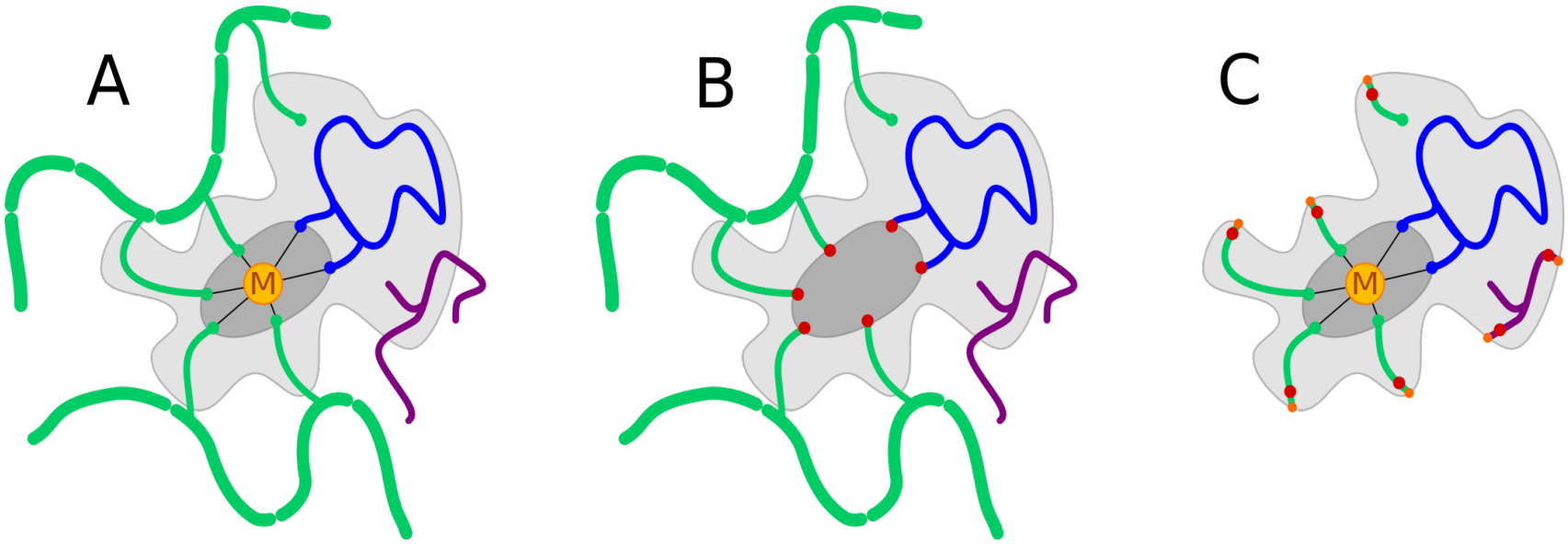
Additional dimension:
Metal and its electronic structure

Original inside-out protocol
Combined with QM on-the-fly

Severe need for new methods

QM/DMD: new hybrid dynamics method

Partitioning of the protein



The "breathing" QM-DMD boundary

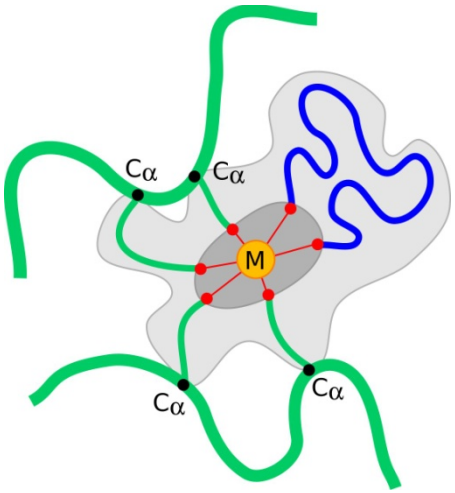
DMD management area

QM management area

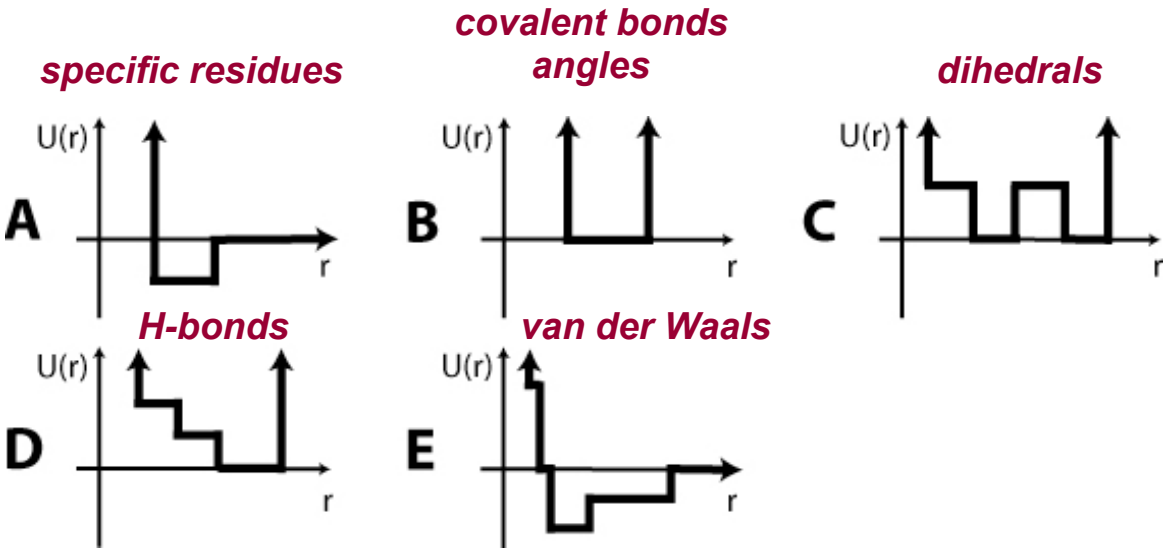
Sparta, Shirvanyants, Ding, Dokholyan, Alexandrova *Biophys. J.* **2012**, *103*, 767-776.

QM/DMD: the DMD part

Extensive nanosecond-scale sampling of the protein, incl. the backbone



Interaction potentials:



QM-DMD boundary around dark-grey

Dokholyan, N.V. et al. Fold. Des. 3, 577 (1998). Ding F. et al. J. Mol. Biol 350, 1035 (2005).

QM/DMD: the DMD part

United atom model (heavy atoms and polar hydrogens)
Backbone-dependent rotamer library

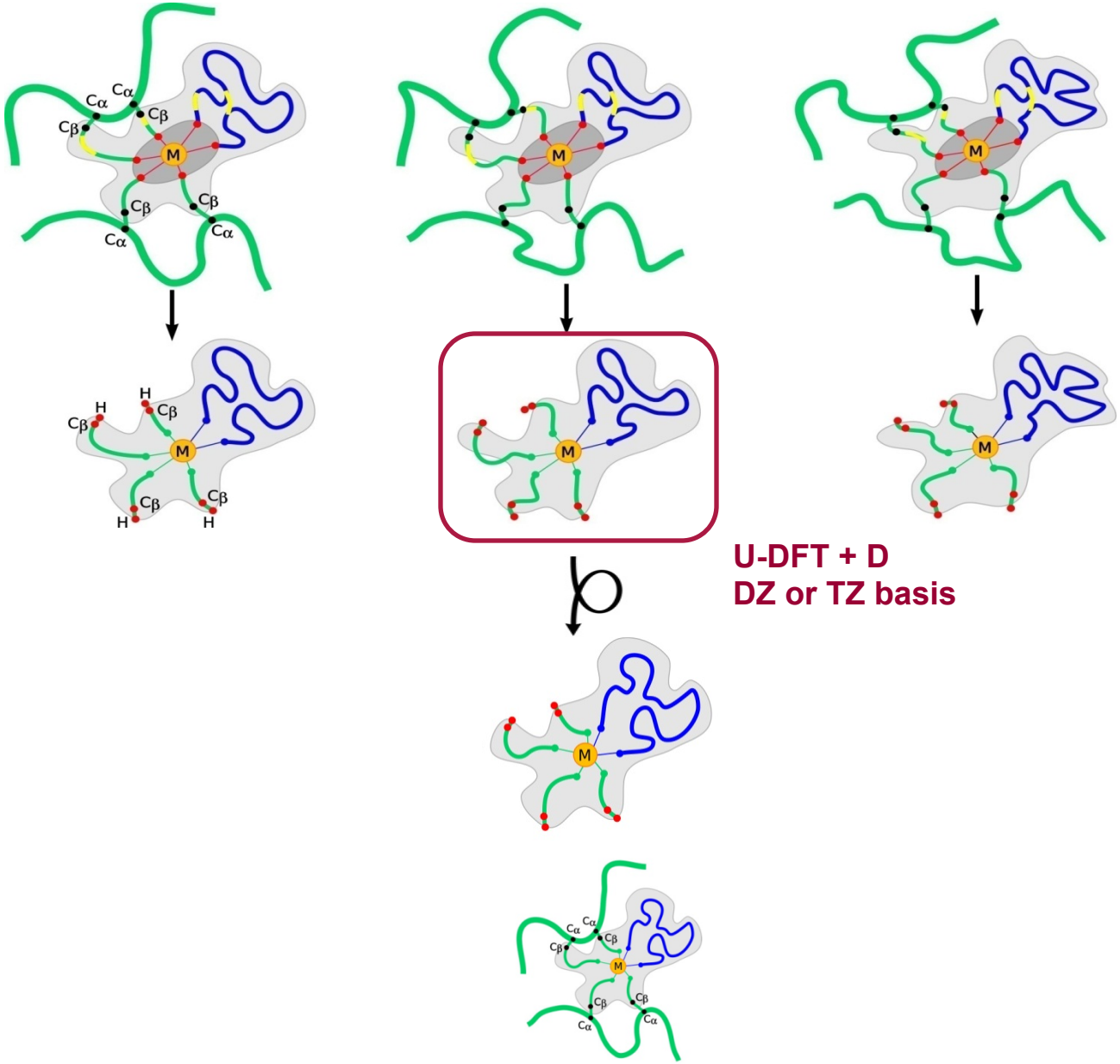
$$\Delta E = W_{vdw_attr} E_{vdw_attr} + W_{vdw_rep} E_{vdw_rep} + W_{solv} E_{solv}$$
$$+ W_{bb_hbond} E_{bb_hbond} + W_{sc_hbond} E_{sc_hbond} + W_{bb_sc_hbond} E_{bb_sc_hbond}$$
$$+ W_{aa|\phi,\psi} E_{aa|\phi,\psi} + W_{rot|\phi,\psi,aa} E_{rot|\phi,\psi,aa} - E_{ref}$$

E_{vdw_attr}, E_{vdw_rep} → Standard Lennard-Jones (LJ), parameters from CHARMM19
E_{solv} → Implict solvation, EEF1 model
E_{bb_hbond}, E_{sc_hbond}, $E_{bb_sc_hbond}$ → Orientation-dependant hbond model
$E_{aa|\varphi,\psi}$, $E_{rot|\varphi,\psi,aa}$ → Backbone dependent internal energies
E_{ref} → Empirical energy for each amino acid to model unfolded state

Dokholyan, N.V. et al. Fold. Des. 3, 577 (1998). Ding F. et al. J. Mol. Biol 350, 1035 (2005).

QM/DMD: the QM part

QM-DMD boundary around light-grey



Clustering
by geometric similarity

Single point QM on centroids

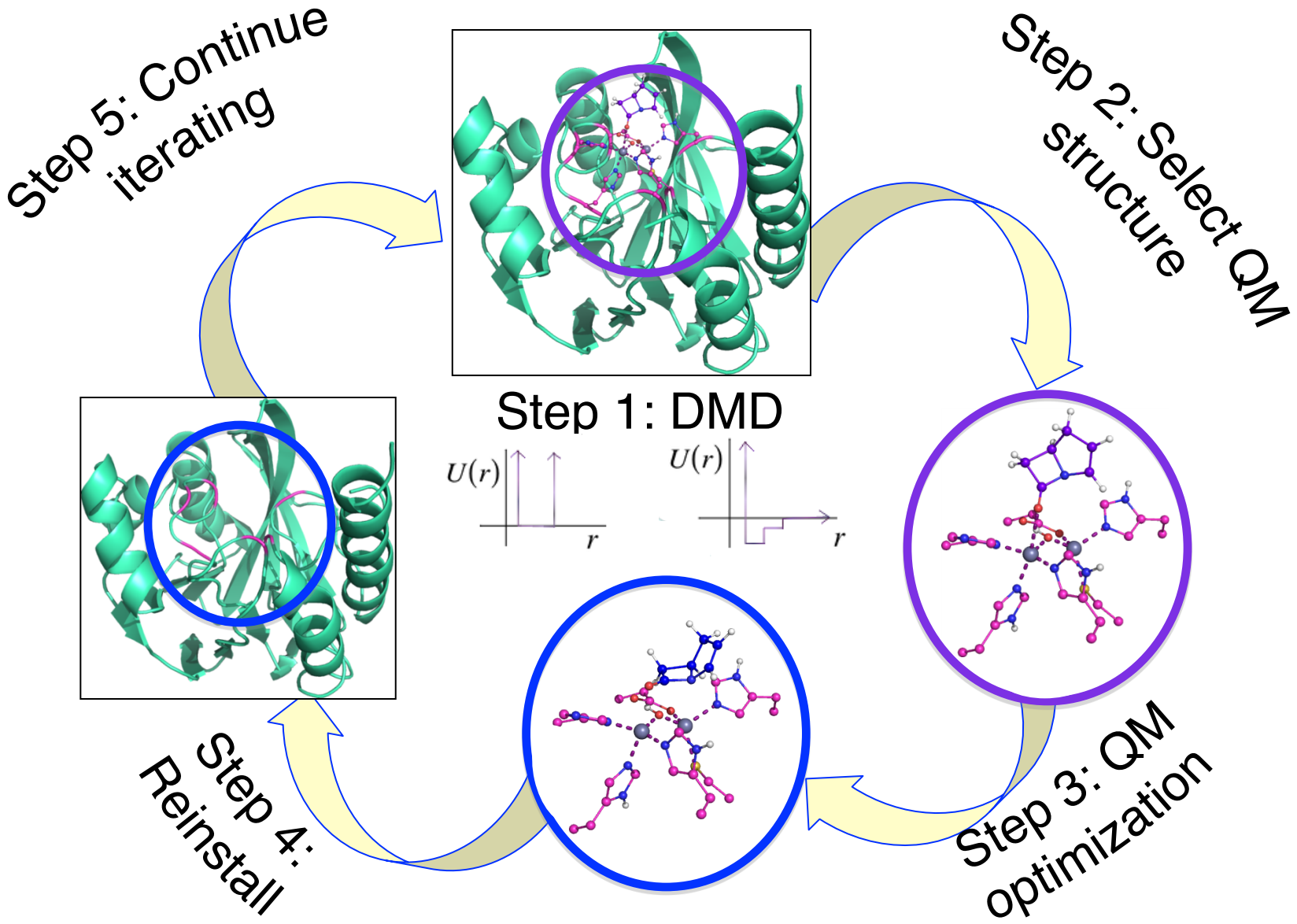
Hybrid scoring and selection

QM optimization

Reinstallation of the QM region into the protein

Continue with DMD

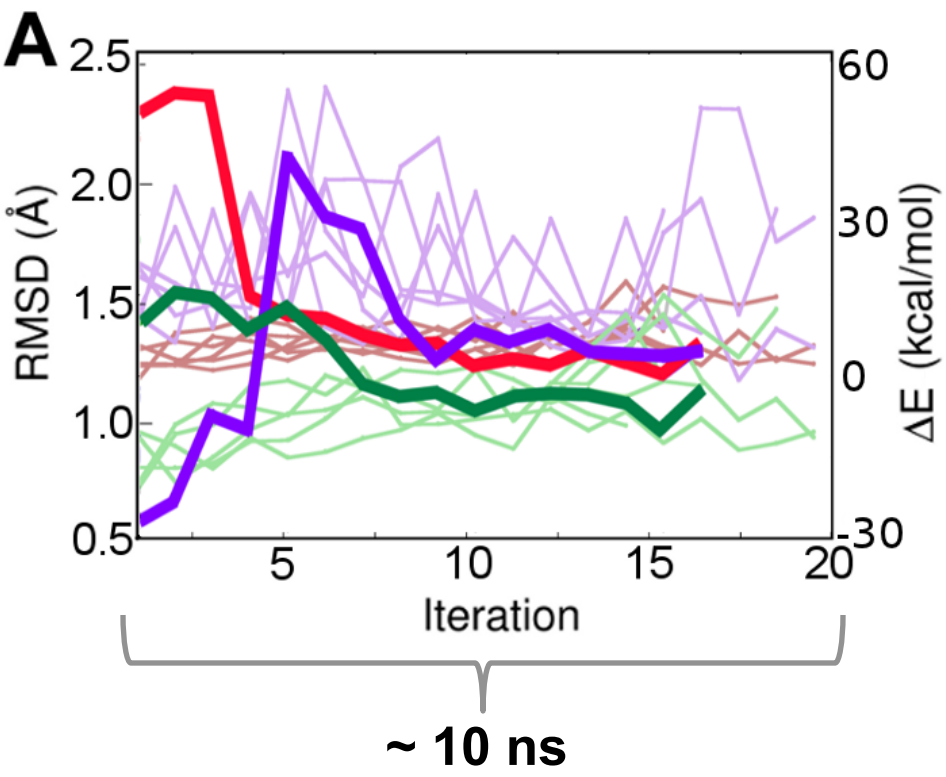
QM/DMD: new hybrid dynamics method

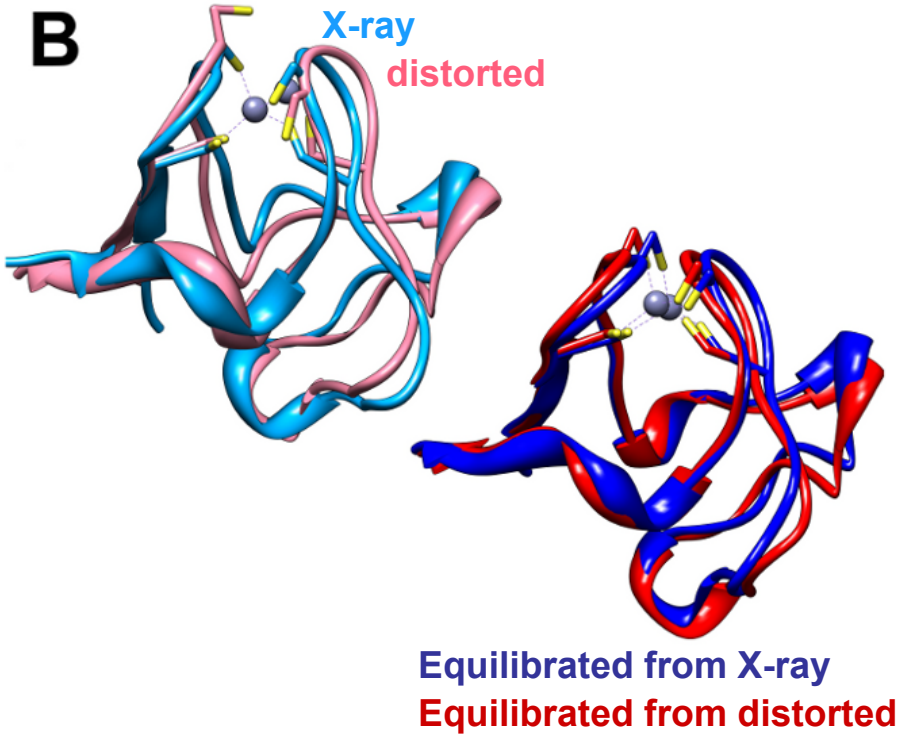


Sparta, Shirvanyants, Ding, Dokholyan, Alexandrova *Biophys. J.* **2012**, *103*, 767-776.

QM/DMD: Structure recapitulation

Rubredoxin: Fe(II)/Fe(III) protein, 53 residues, X-Ray 15 Å



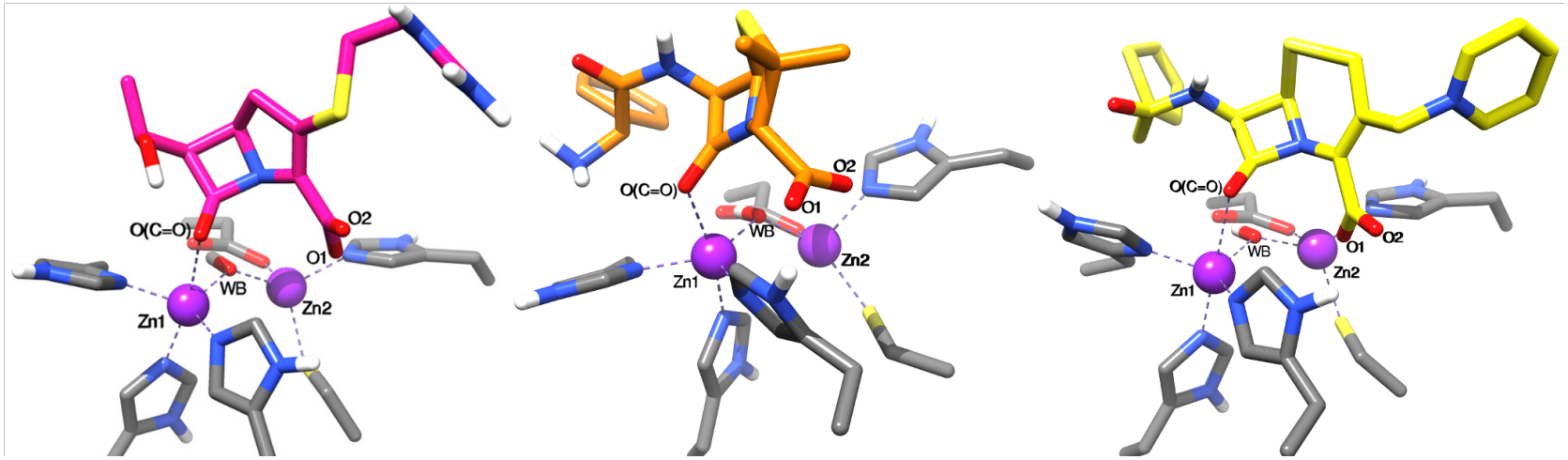


Sparta, Shirvanyants, Ding, Dokholyan, Alexandrova *Biophys. J.* **2012**, *103*, 767-776.

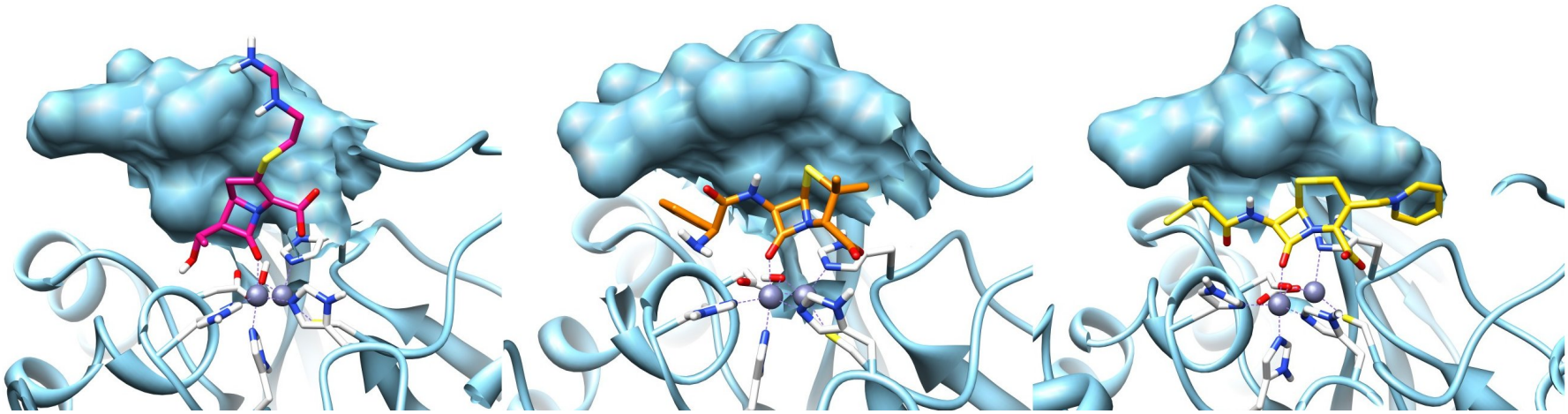
Large scale motions. Flexible docking.

di-Zn metallo-β-lactamase (CcrA): flexible loop above the active site

CG



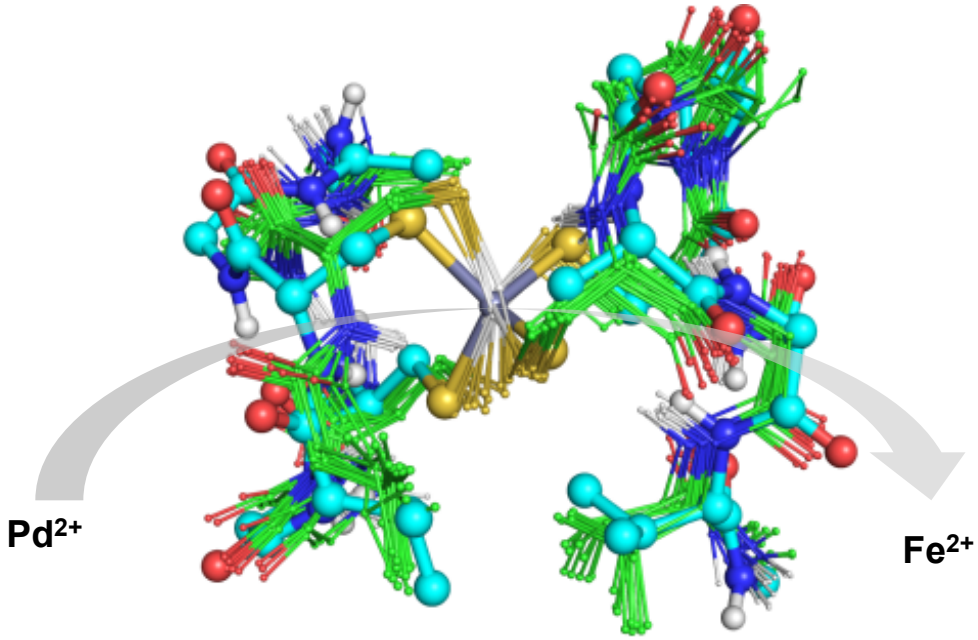
Imipenem **Ampicillin** **Cephalosporine**



Valdez, AA **2013** *JCTC, 9*, 730

Structure prediction: M replacement

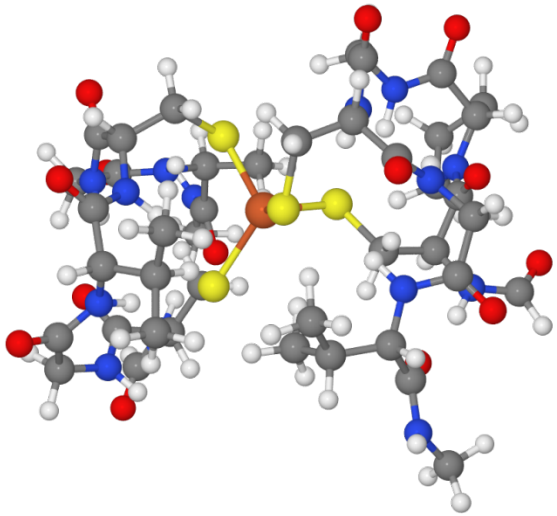
Pd-substituted Rubredoxin: compromising

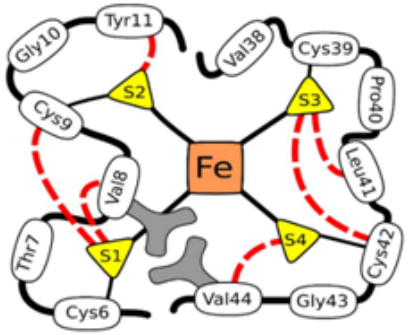


Going toward square planar Pd, but not quite
Slightly away from the optimum for the protein

Sparta, Shirvanyants, Ding, Dokholyan, Alexandrova *Biophys. J.* **2012**, *103*, 767-776.

Finer structural details

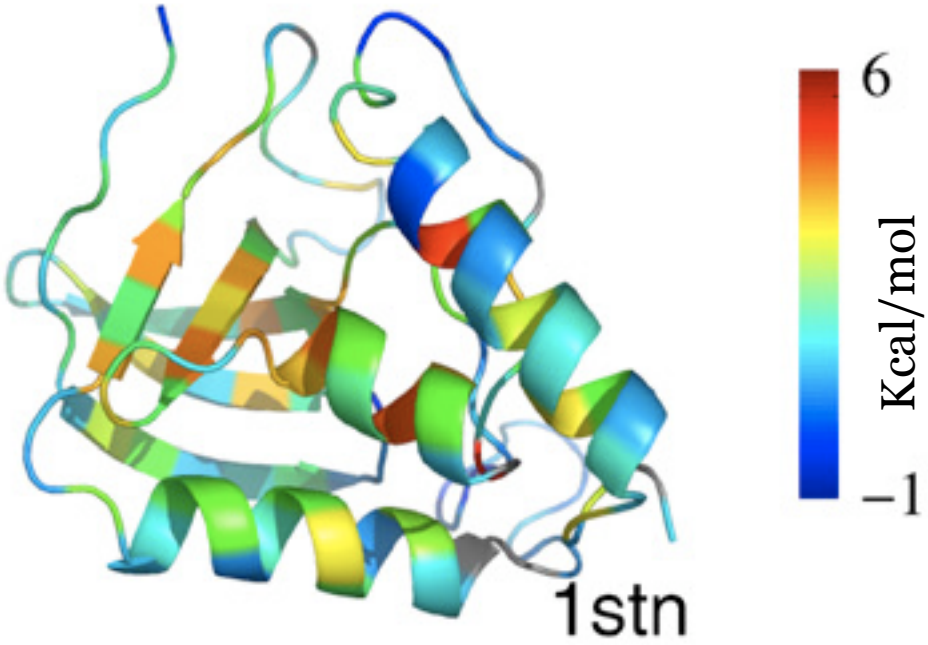




	Fe(II) exper	Fe(II), calc	Fe(III), exper	Fe(III), calc
$Fe\text{-}S_{Cys6}$	2.38	2.35 (0.02)	2.25	2.26 (0.01)
$Fe\text{-}S_{Cys9}$	2.37	2.28 (0.01)	2.30	2.17 (0.01)
$Fe\text{-}S_{Cys39}$	2.39	2.33 (0.01)	2.27	2.22 (0.02)
$Fe\text{-}S_{Cys42}$	2.31	2.31 (0.01)	2.25	2.19 (0.02)

Sparta, Shirvanyants, Ding, Dokholyan, Alexandrova *Biophys. J.* **2012**, *103*, 767-776.

Eris – predicting protein stability upon mutagenesis



Need in addition:

QM for the metal

More backbone sampling

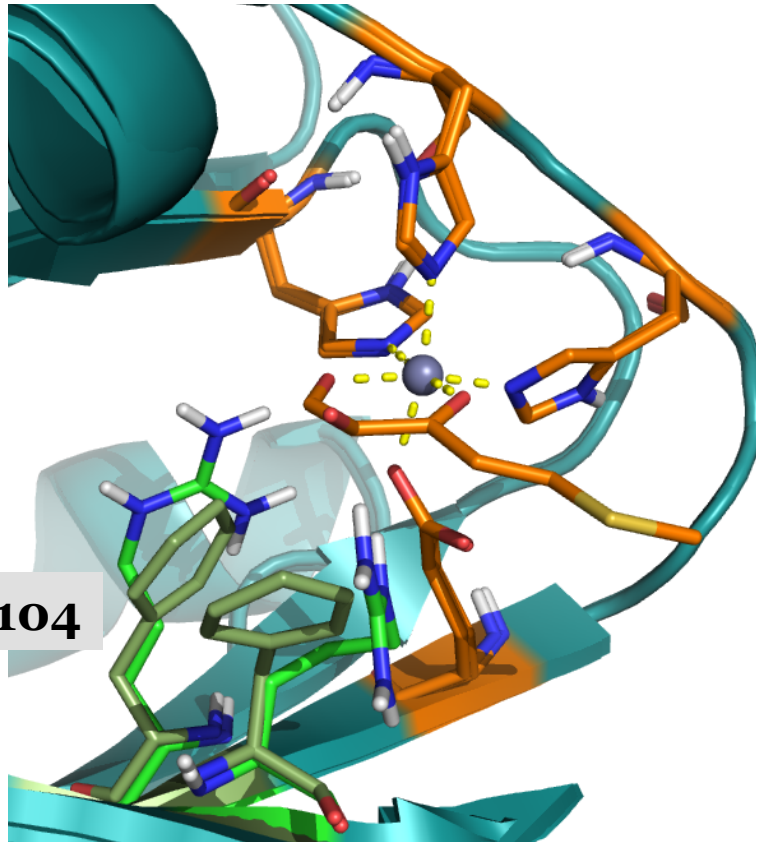
$$\Delta\Delta G = \Delta G^{mutant} - \Delta G^{wildtype}$$

Positive number = disfavorable mutation

Negative number = favorable mutation

1. Yin, S., Ding, F., and Dokholyan, N. V., "Modeling backbone flexibility improves protein stability estimation", Structure, 15:1567-1576, (2007).

Eris – test and exercise

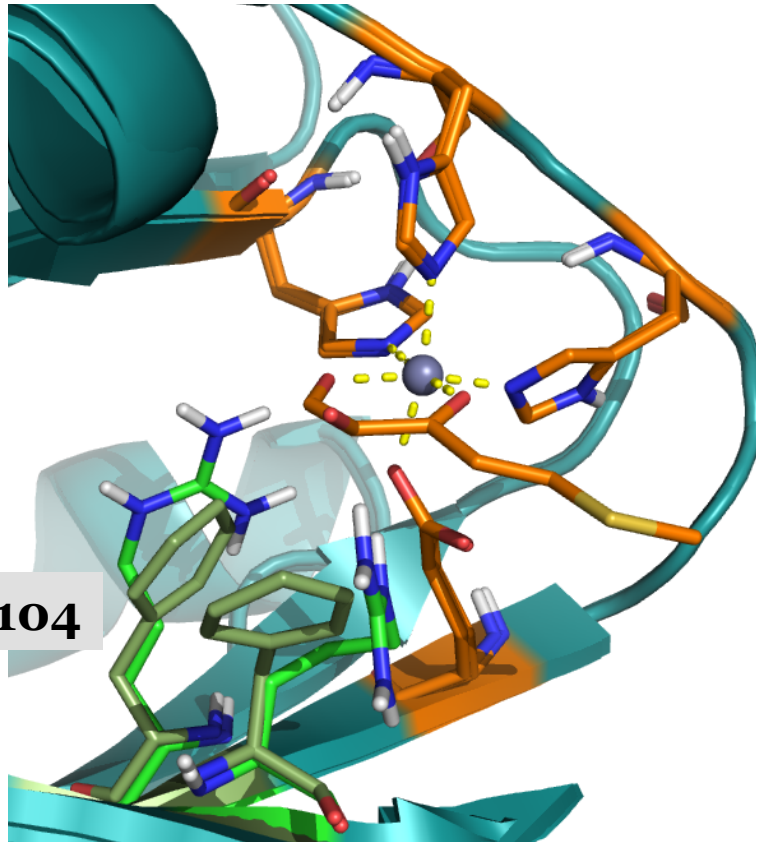


R104	→ R104F
R154	→ R154F

1. Flex Backbone *Eris*
***2. Eris*+QM/DMD for sampling and QM treatment of the metal**

$$\Delta\Delta G = \Delta G^{mutant} - \Delta G^{wildtype}$$

Eris – test and exercise



R104	→ R104F
R154	→ R154F

	ΔΔG Phe→ Arg (kcal/mol)
Eris (Flex Backbone)	13.5
Eris + QM/DMD	-12.9

Eris **is not providing enough backbone movement for reintroducing native ARG.**

Eris **+ QM/DMD is the recipe.**

First time QM effects are included in the design process.

Multi-scale methods for Protein Design

Methods used so far:

QM, MM, QM/MM

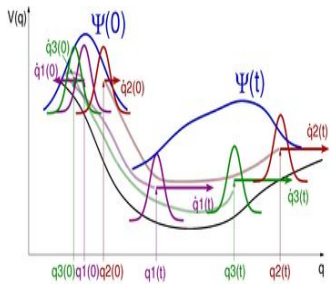
Methods that could be added to the toolbox:

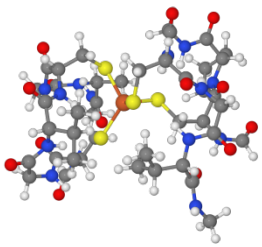
Quantum dynamics (H^+ or e^- transfer, PCET)

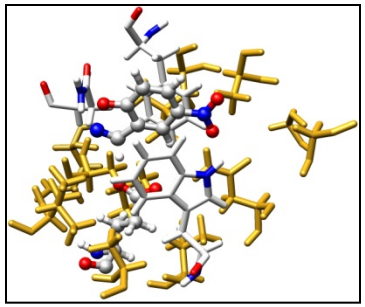
Coarse graining (more aggressive protein motifs design, loops)

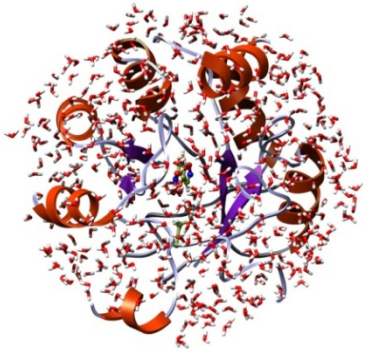
Hybrid methods combining several components

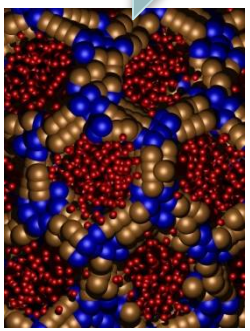
Quantum dynamics → **QM** → **MM** → **QM/MM** → **CG**







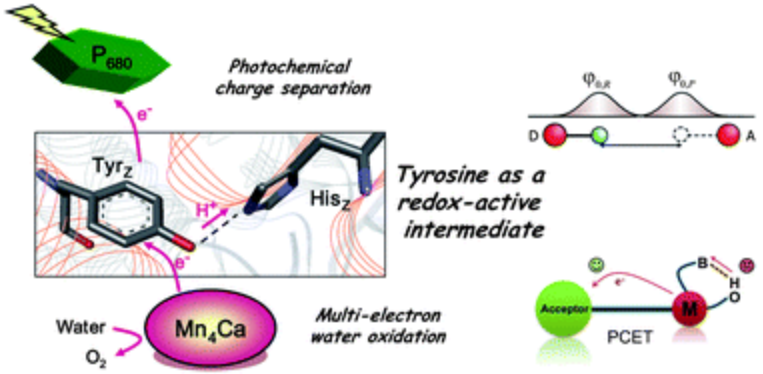




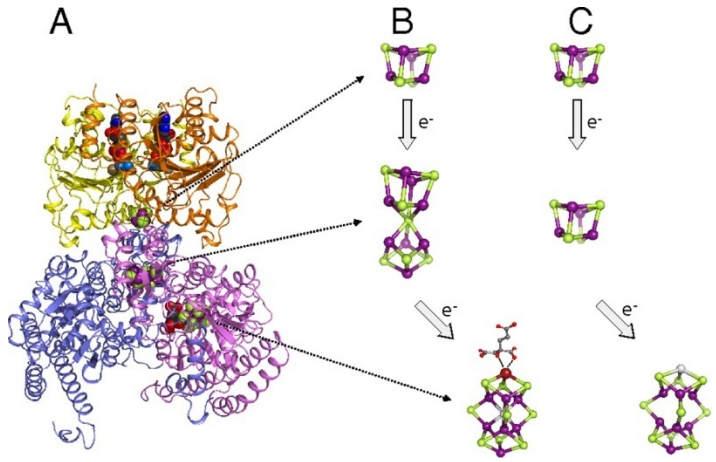
Multi-scale methods for Enzyme Design

Coupling to quantum dynamics

PCET. Photosystem II



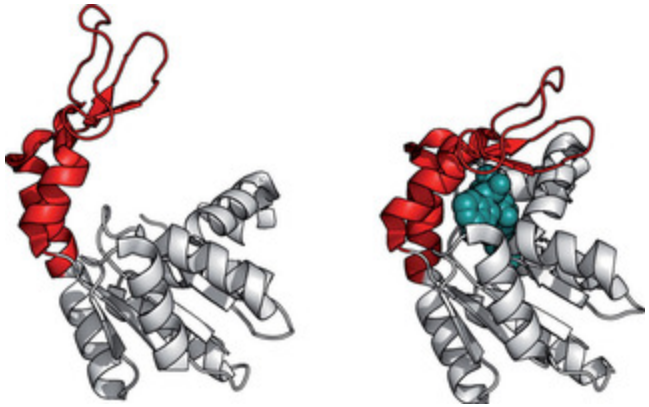
Multiple e- transfers. Nitrogenase



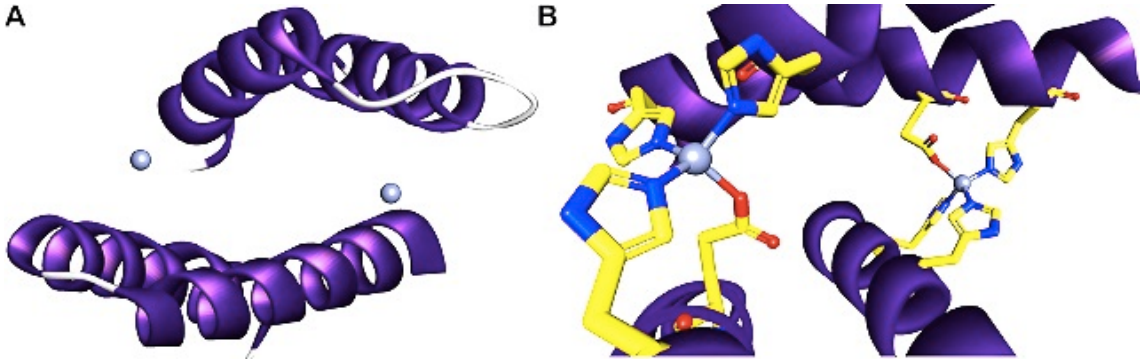
Multi-scale methods for Enzyme Design

Coupling to CG

Large-scale motions can be expected



Zn-fingers for protein-protein interfaces



Multi-scale methods for Protein Design

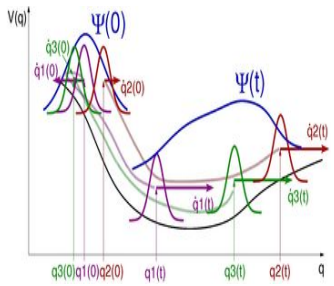
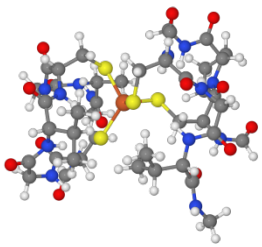
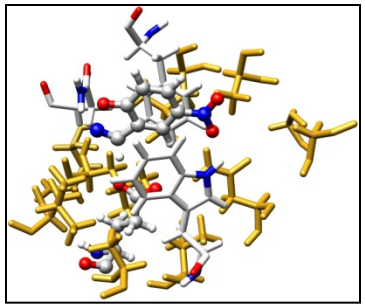
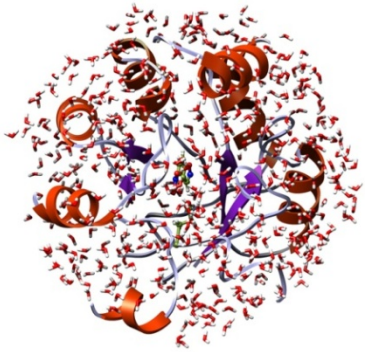
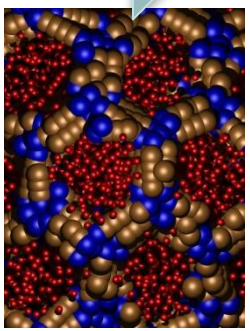
Methods used so far:

QM, MM, QM/MM

Methods that could be added to the toolbox:

Quantum dynamics (H^+ or e^- transfer, PCET)

Coarse graining (more aggressive protein motifs design, loops)

Hybrid methods combining several components

Quantum dynamics → **QM** → **MM** → **QM/MM** → **CG**