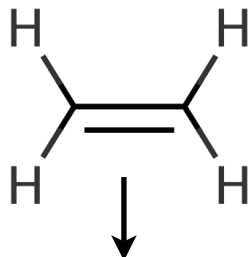


Chemistry on Computers

Aaron Virshup
and
Agostino Migliore
Duke Chemistry

Computational chemistry is about calculating molecular energies from atomic coordinates.

1. Molecular structure



2. Atomic coordinates \mathbf{R}

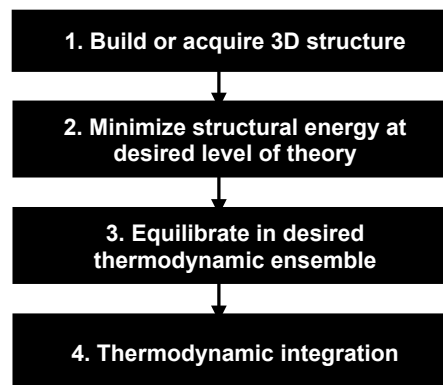
ATOM	X	Y	Z
C	-0.65293	0.0000	2.52293
C	0.65293	0.0000	2.52293
H	-1.22900	-0.9300	2.52293
H	-1.22900	0.9300	2.52293
H	1.22900	-0.9300	2.52293
H	1.22900	0.9300	2.52293

3. Molecular energy

$$E(\mathbf{R}) = -14.2184 \text{ eV}$$

- Most chemistry can be derived from the energy function $E(\mathbf{R})$: bond lengths, rate constants, vibrational modes, ground state electronic structure, time evolution, thermodynamic properties
- Most computational chemistry calculations based on either classical or quantum many-body simulations

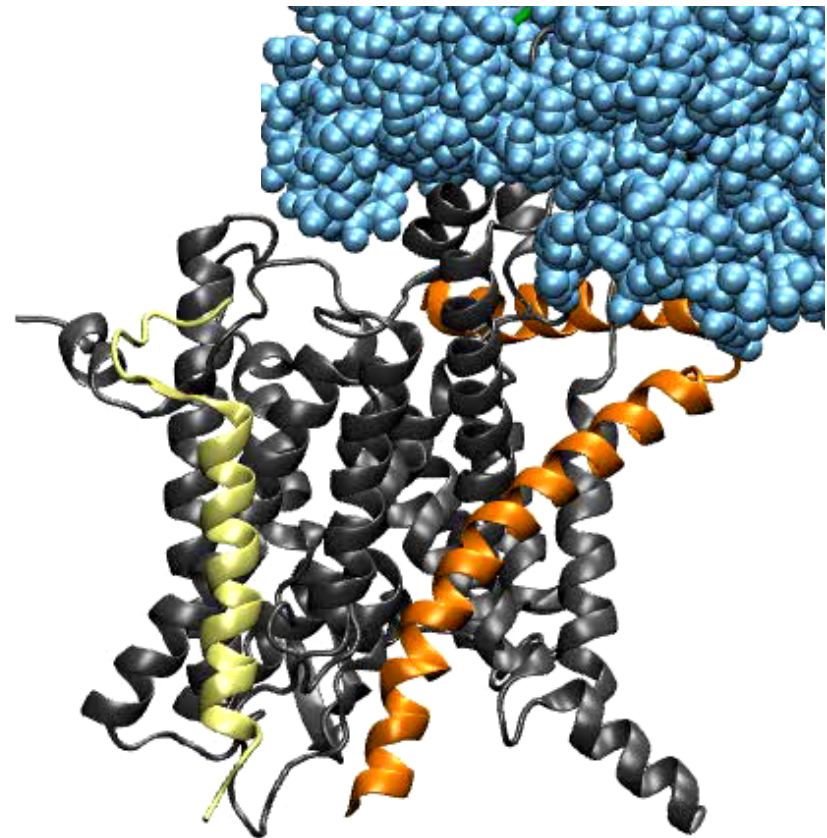
- Workflows are usually single user building, analyzing a chemical system:



Molecular mechanics uses simple, computationally facile energy expressions.

$$E(\mathbf{R}) = \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \delta)] \\ + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u (u - u_0)^2 \\ + \sum_{\text{nonbonded}} \epsilon \left[\left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon r_{ij}}$$

- “Atomistic” simulations: $\sim 10^3$ - 10^5 atoms, as many as 10 million
- Cheap enough for long-time dynamical simulation (ps- μ s)
- Developed for modeling biomolecules
- Straightforward parallelization-split energy function evaluation over multiple nodes
- Scales as $O(\text{Number of atoms})^2$

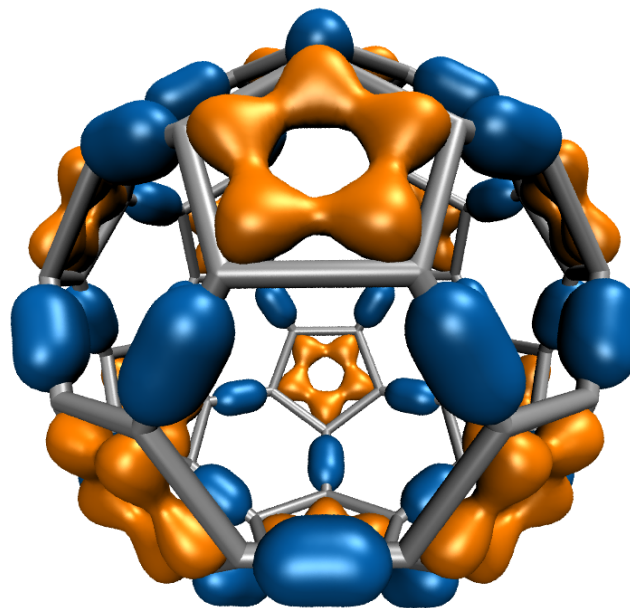


Quantum chemistry models molecules at the electronic level.

- Used to determine structure, electronic properties of molecules <100 atoms
- Exact solution of electronic Schrodinger equation scales as $O(e^N)$
- Many approximate methods exist, with polynomial scaling, usually at least $O(N^4)$
- Parallelization is not straightforward
- Big systems need fast network communications, fast and plentiful storage, lots of memory

$$H\Psi = E(\mathbf{R})\Psi$$

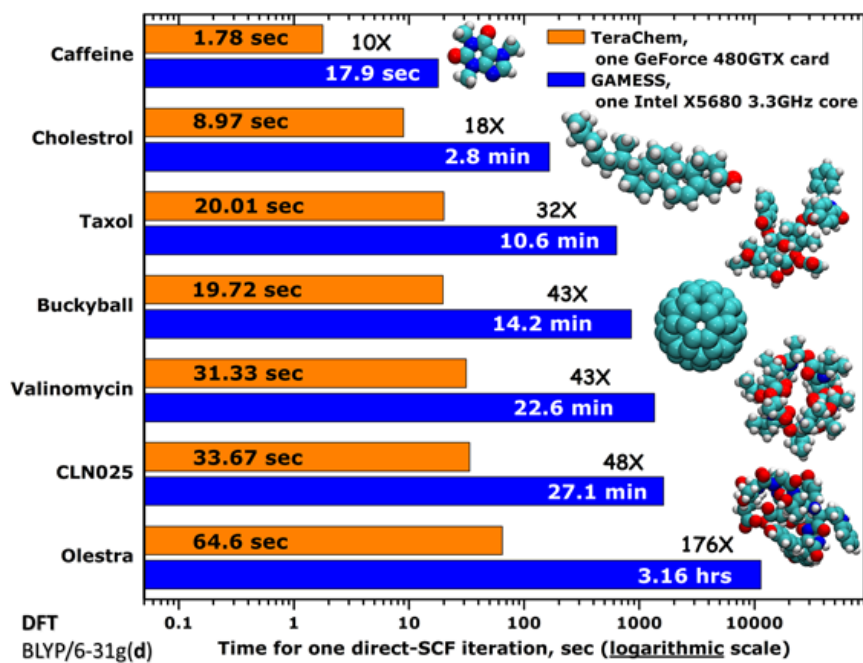
$$H = T_{electron} + V_{Nuclear-Nuclear} + V_{electron-electron} + V_{Nuclear-electron}$$



GPUs show great potential for both classical and quantum computation.

Wall time for QM calculation,

GPU vs. **CPU**



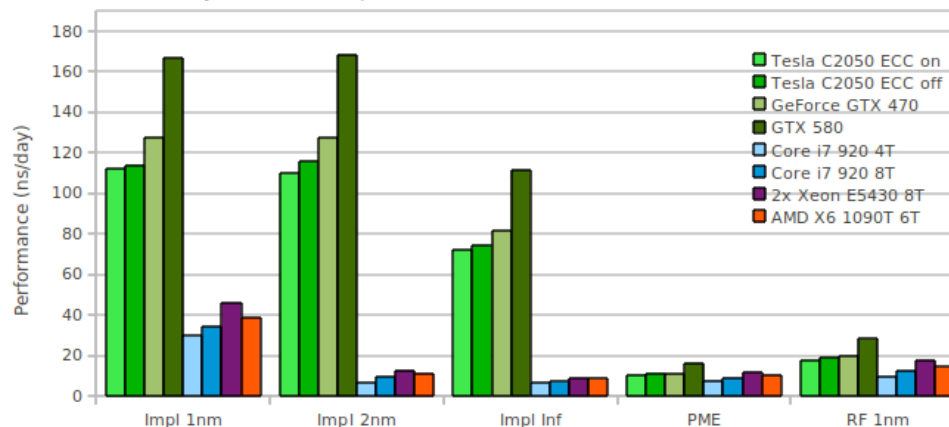
Source: TeraChem

Simulation time per wall day,

GPU vs. **CPU**

GROMACS 4.5 performance comparison

system: DHFR implicit (2489 atoms), solvated (23569 atoms)



Source: GROMACS

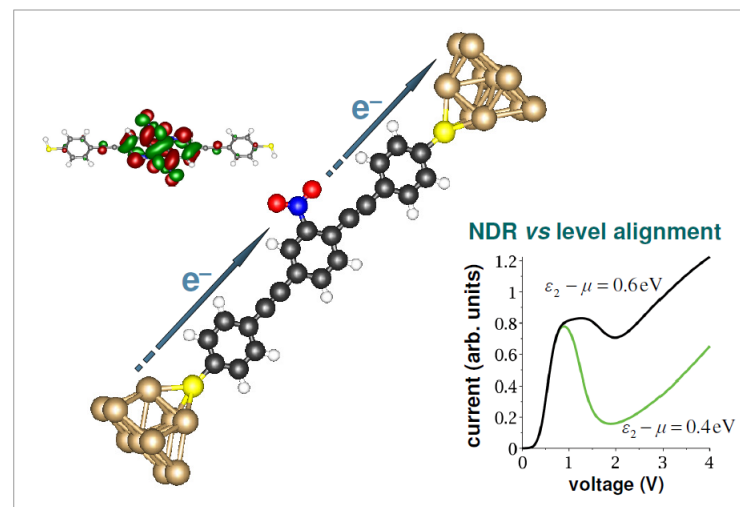
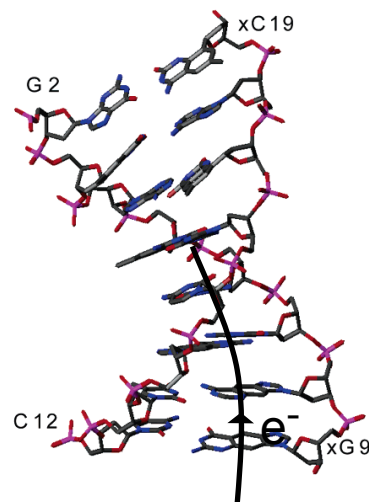
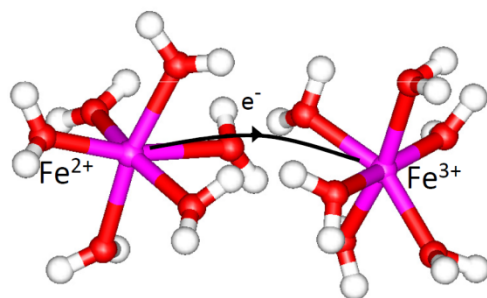
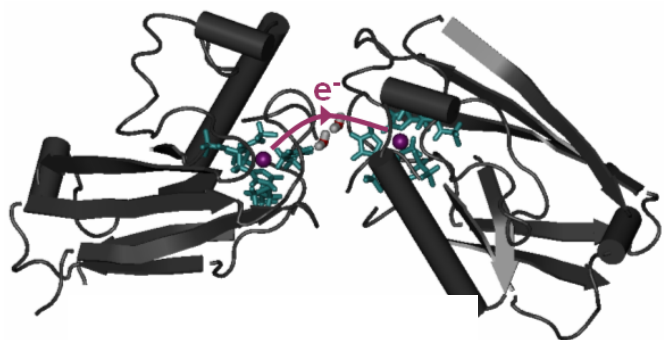
**The future – more compounds, bigger systems,
better sampling, more accuracy.**

- Molecular mechanics calculations need: bigger, better clusters
- Quantum methods: More shared memory with more processors, more methods need to be adapted for GPUs
- Better ways to track, store, and visualize enormous amounts of data
- High-throughput methods to perform standard calculations on large databases of molecules: more nodes = bigger, better properties

Computational Investigation of Charge Transfer Processes in Enzymes, DNA, Interfaces.

Agostino Migliore (*David Beratan's group*)

Department of Chemistry, Duke University, Durham, NC, USA

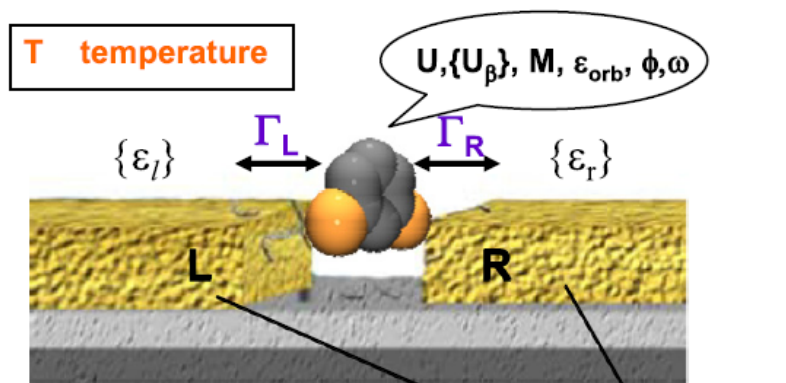
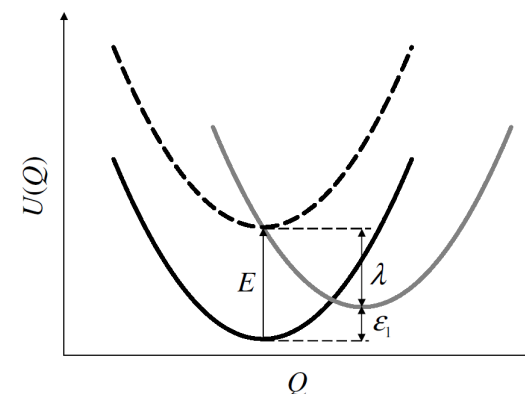
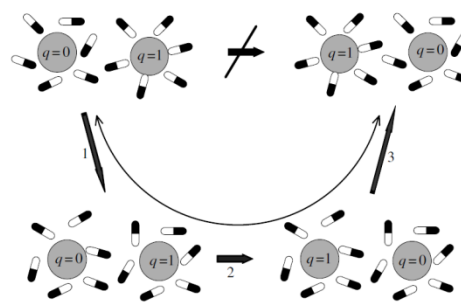


Main Tasks and Computations

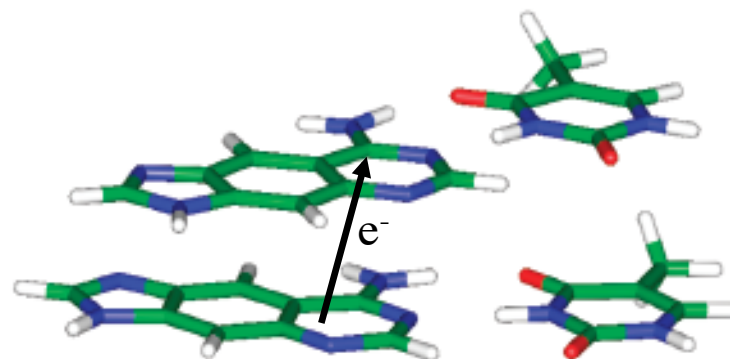
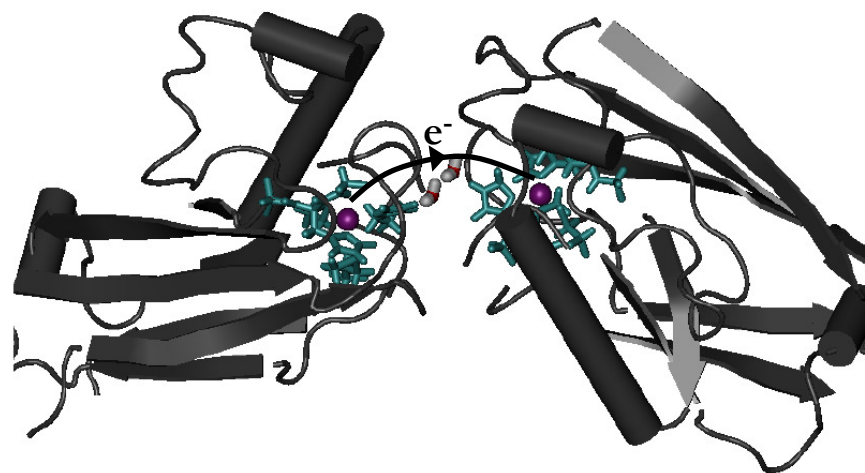
- **MD and QM/MM to explore protein, DNA, electrode-molecule interface dynamics: instant properties and their fluctuations.**
- **Accurate quantum computation of electronic properties on MD snapshots.**
- **Semi-empirical methods on wider snapshot selections.**

ET, PT, PCET: ubiquitous processes

A. Nitzan, *Chemical dynamics in condensed phases* (Oxford University Press, Oxford, 2007)



- ϵ_k band states in electrode
 - U electron repulsion on molecule
 - $\{U_\beta\}$ coupling to external bath modes
 - M vibronic coupling on molecule
 - ϵ_{orb} molecular orbital energies
 - ϕ molecular orbitals
 - ω molecular vibrational frequency
 - Γ spectral density (electron-lead coupling)
 - E_{FK} (K=L,R) Fermi energies
 - Φ bias potential
- $E_{FL} = E_{FR} + |e|\Phi$



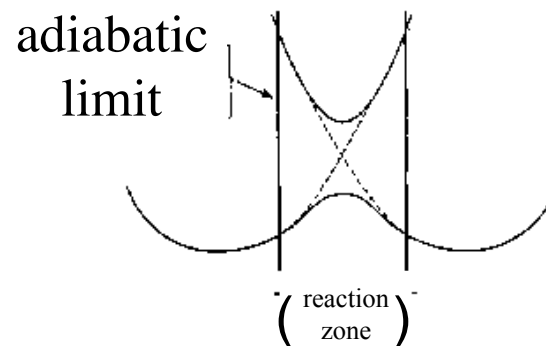
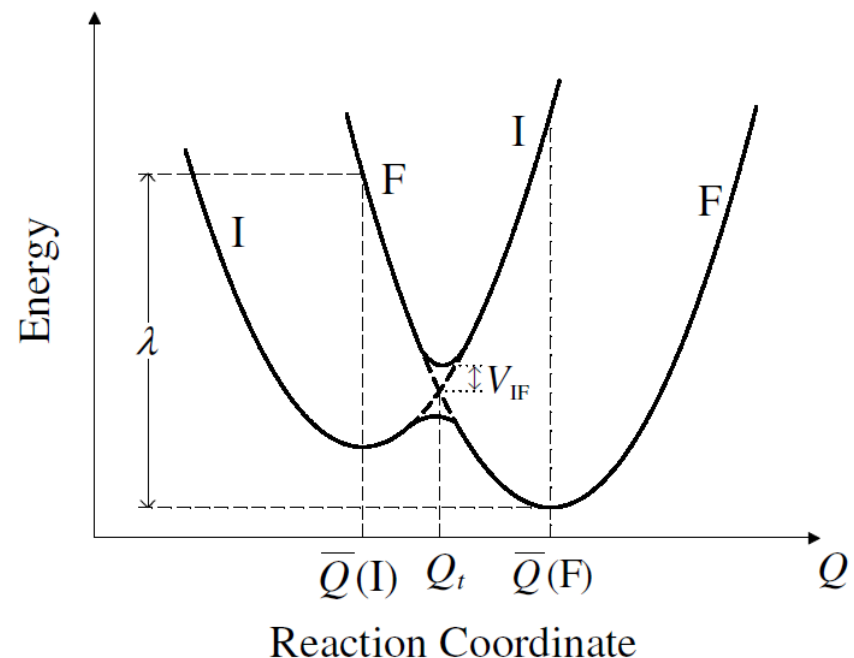
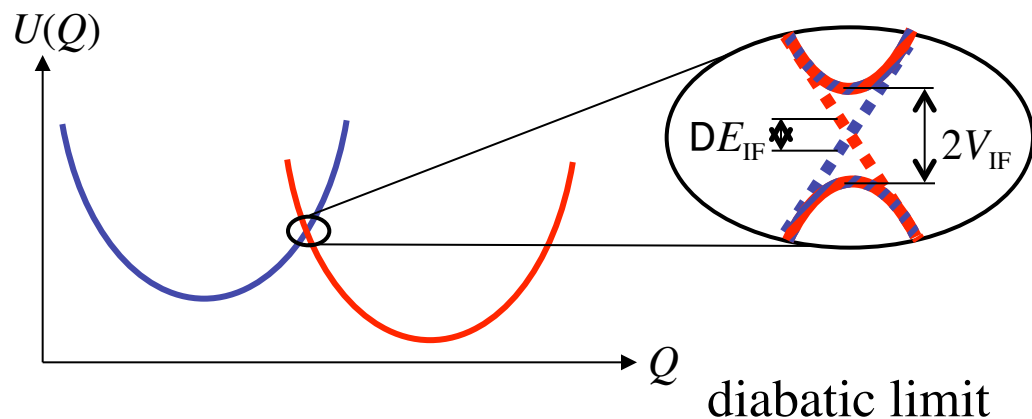
M. Galperin, M. A. Ratner, A. Nitzan, J. Phys.: Condens. Matter 19, 103201 (2007).

Marcus' equation for ET rate (linear response of the polarization of a thermal bath with Gaussian statistics, and equal force constants for reactant and product modes):

$$k_{\text{ET}} = \kappa(V_{\text{IF}}) \nu \exp\left[-\frac{(\Delta G^0 + \lambda)^2}{4\lambda k_{\text{B}}T}\right]$$

Proximity of transition state coordinate:

$$Q \sim Q_t \rightarrow \Delta E_{\text{IF}} \ll 2V_{\text{IF}}$$



$$q_{\text{D}} = q_{\text{A}} \text{ at } Q = Q_t$$

Marcus-Hush-Levich ET theory

Homogeneous ET

$$k_{\text{ET}} = \kappa(V_{\text{IF}}) \nu \exp\left[-\frac{(\Delta G^0 + \lambda)^2}{4\lambda k_B T}\right]$$

k = electronic transmission coefficient

ν_n = effective frequency for the nuclear motion along Q

λ = nuclear reorganization energy

V_{IF} = effective electronic coupling

ΔG_0 = free energy of reaction

R. A. Marcus and N. Sutin,
Biochim. Biophys. Acta 811, 265

R. A. Marcus, Annu. Rev. Phys. Chem. 15, 155

Heterogeneous ET

$$R_{\text{mol-met}} = \int_{-\infty}^{\infty} dE \gamma(E, V_{\text{IF}}, \rho_E) f(E; \Delta V) \\ \times \exp\left[-\frac{(E - \varepsilon - \lambda_{\text{mol}})^2}{4\lambda_{\text{mol}} k_B T}\right]$$

g = coupling to the metal contact

E = metal level energy

ρ_E = density of metal electron states

V_{IF} = molecule-metal coupling

ΔV = interface voltage

ε = molecular level

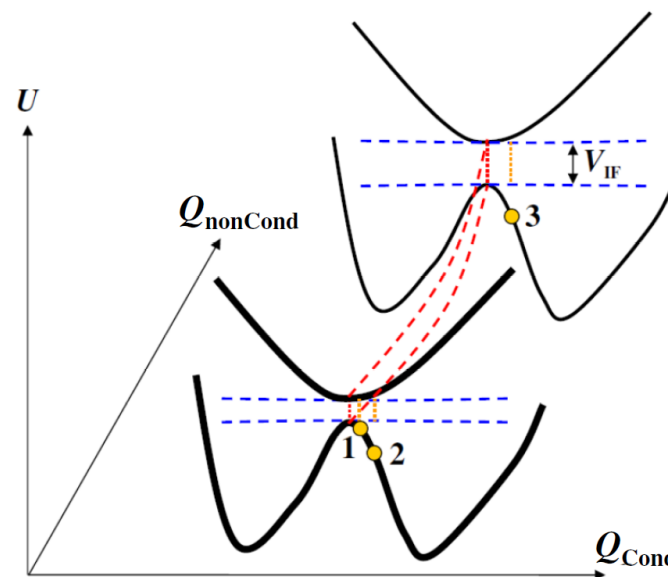
λ_{mol} = reorganization energy of the molecular bridge

System fluctuations

$$k_{\text{ET}}^{(2)} = k_{\text{ET}}^{(0)} \frac{1}{8} \left(\frac{\hbar}{\tau_c k_B T} \right)^2 \left(1 - \frac{2k_B T}{\lambda} \right) (1 - R_c)$$

t_c is the correlation time of the electronic coupling

$$1 - R_c = 1 - \frac{\langle U_{\text{IF}} \rangle^2}{\langle U_{\text{IF}}^2 \rangle} = \frac{\langle U_{\text{IF}}^2 \rangle - \langle U_{\text{IF}} \rangle^2}{\langle U_{\text{IF}}^2 \rangle} = \frac{\sigma_{U_{\text{IF}}}^2}{\langle U_{\text{IF}}^2 \rangle}$$



Troisi, A., Nitzan, A. & Ratner, M. A. (2003) *J. Chem. Phys.* **119**, 5782-5788.

Troisi, A., Ratner, M. A. & Zimmt, M. B. (2004) *JACS* **126**, 2215-2224.

Skourtis, S. S., Balabin, I. A., Kawatsu, T. & Beratan, D. N. (2005) *PNAS U.S.A.* **102**, 3552-3557.

Electronic coupling computation. Semi-empirical methods:

CNDO methods.

Tunneling pathways model:

$$|H_{\text{DA}}|^2 = A^2 \left(\prod_i \epsilon_{\text{bond}}(i) \right)^2 \left(\prod_j \epsilon_{\text{space}}(j) \right)^2 \left(\prod_k \epsilon_{\text{H-bond}}(k) \right)^2$$

$$\epsilon_{\text{bond}} = 0.6$$

$$\epsilon_{\text{H-bond}} = \epsilon_{\text{bond}}^2 \exp[-1.7 (R_{\text{bond}} - 2.8)]$$

$$\epsilon_{\text{space}} = \frac{1}{2} \epsilon_{\text{bond}} \exp[-1.7 (R_{\text{bond}} - 1.4)]$$



Beratan, Onuchic, Hopfield, *J. Chem. Phys.* **86**, 4488 (1987).

Jones, Kurnikov, Beratan, *J. Phys. Chem. A* **106**, 2002 (2002).

Balabin, Hu, Beratan, *J. Comput. Chem.* **33**, 906 (2012).

First-principles V_{IF} from non-orthogonal diabatic states

Two-state model, secular equation for the ground state ψ_-
 $\psi_{\text{I}}, \psi_{\text{F}}$: initial and final diabatic (localized) electronic states

$$\begin{vmatrix} H_{\text{II}} - \varepsilon & H_{\text{IF}} - \varepsilon S_{\text{IF}} \\ H_{\text{IF}} - \varepsilon S_{\text{IF}} & H_{\text{FF}} - \varepsilon \end{vmatrix} = 0$$

where $|\psi_-\rangle = a|\psi_{\text{I}}\rangle + b|\psi_{\text{F}}\rangle$, $E_- = \langle \psi_- | H | \psi_- \rangle$, $H_{\text{II(FF)}} = \langle \psi_{\text{I(F)}} | H | \psi_{\text{I(F)}} \rangle$, $H_{\text{IF}} = \langle \psi_{\text{I}} | H | \psi_{\text{F}} \rangle$, $S_{\text{IF}} = \langle \psi_{\text{I}} | \psi_{\text{F}} \rangle$,

A. Farazdel, M. Dupuis,
 E. Clementi, A. Aviram,
 JACS 112, 4206

$$V_{\text{IF}}(Q) \equiv \frac{1}{1 - S_{\text{IF}}^2(Q)} \left| H_{\text{IF}}(Q) - S_{\text{IF}}(Q) \frac{H_{\text{II}}(Q) + H_{\text{FF}}(Q)}{2} \right|$$

A. M., J. Chem. Phys.
 131, 114113 (2009)

transfer integral or effective electronic coupling or effective ET matrix element

electronic coupling or ET matrix element

$$\begin{aligned} V_{\text{IF}} &= U_{\text{IF}}(a, b, \Delta E_{\text{IF}}) \sigma(a, b, S_{\text{IF}}) \\ &= \left| \frac{ab}{a^2 - b^2} \Delta E_{\text{IF}} \right| \left| \frac{1}{1 - S_{\text{IF}}^2} \right| \left| 1 + \frac{a^2 + b^2}{2ab} S_{\text{IF}} \right| \\ &= \left| \frac{AB}{A^2 - B^2} \Delta E_{\text{IF}} \right| \left| \frac{1}{1 - S_{\text{IF}}^2} \right| \left| 1 - \frac{A^2 + B^2}{2AB} S_{\text{IF}} \right| \end{aligned}$$

$$\Delta E_{\text{v}} \equiv E_+ - E_- = \sqrt{\frac{\Delta E_{\text{IF}}^2}{1 - S_{\text{IF}}^2} + 4V_{\text{IF}}^2}$$

$$\begin{aligned} \Delta E_{\text{IF}} &= H_{\text{II}} - H_{\text{FF}} & A &= \langle \psi_{\text{I}} | \psi_- \rangle & B &= \langle \psi_{\text{F}} | \psi_- \rangle \\ a &= \frac{A - BS_{\text{IF}}}{1 - S_{\text{IF}}^2} & b &= \frac{B - AS_{\text{IF}}}{1 - S_{\text{IF}}^2} \end{aligned}$$

➤ **Different diabatic sets can be used in AM method: two examples.**

Tensor product (TP) diabatic states, in the absence of covalent donor (\mathcal{D}) – acceptor (\mathcal{A}) bridge:

$$|\psi_I\rangle = |\mathcal{D}\rangle|\mathcal{A}\rangle \quad |\psi_F\rangle = |\mathcal{D}^+\rangle|\mathcal{A}^-\rangle$$

(using reference states for the isolated \mathcal{D} and \mathcal{A} groups in the initial and final charging states)

$$\Delta E_{\text{IF}} = (E_{\mathcal{D}} + E_{\mathcal{A}}) - (E_{\mathcal{D}^+} + E_{\mathcal{A}^-}) + W_{\mathcal{D}-\mathcal{A}} - W_{\mathcal{D}^+-\mathcal{A}^-}$$

Different levels of approximation can be used to calculate **electrostatic interactions among subsystems (and not only for computing the diabatic energy difference):**

Multipole expansion (it works for well-separated redox sites).

Complete electrostatic interaction using RESP atomic charges (e.g., it works also for DNA nucleobase stacks).

Full electrostatics using Poisson equation: especially useful in PW calculations.

$$V_{\text{IF}} = \left| \frac{AB}{A^2 - B^2} \Delta E_{\text{IF}} \left(1 - \frac{A^2 + B^2}{2AB} S_{\text{IF}} \right) \frac{1}{1 - S_{\text{IF}}^2} \right|$$

Constrained-DFT (CDFT) diabatic states can be used also in the presence of a covalent \mathcal{D} - \mathcal{A} bridge and when the resulting S_{IF} is very large

Functional under minimization: $F = \langle \psi_c | H + V_c w_c | \psi_c \rangle = E[\rho_c] + V_c \int w_c(\mathbf{r}) \rho_c(\mathbf{r}) d\mathbf{r} = E + V_c N_c$

w_c = weight function that defines the constrained property
 (e.g., it is 1 in \mathcal{D} and 0 elsewhere, thus constraining the number of electrons in \mathcal{D}).
 V_c = Lagrange multiplier implicitly determined by the specified constraint value

} $V_c w_c$ is the constraint potential

$|\psi_{\text{I}}\rangle = |\mathcal{D}, \mathcal{A}\rangle$ and $|\psi_{\text{F}}\rangle = |\mathcal{D}^+, \mathcal{A}^-\rangle$ have different $V_c \rightarrow$ they are generally not orthogonal states.

- **CDFT SCF calculations, e.g., using NWChem** (Q. Wu and T. Van Voorhis, *Phys. Rev. A* **72**, 024502; *J. Chem. Phys.* **125**, 164105) $\rightarrow \{E_{\text{I}}, E_{\text{F}}\} \rightarrow \mathbf{DE}_{\text{IF}} \equiv E_{\text{I}} - E_{\text{F}}$.
- Trivial post-processing (e.g., ET module in NWChem; or **Gaussian**) $\rightarrow A, B, S_{\text{IF}}$.
- Can be used also when the resulting S_{IF} is very large.

The tunneling pathway, TP, model and the AM method [also in the prototype form from Migliore, Corni, Di Felice, Molinari, *J. Chem. Phys.* **124**, 064501] can be combined (work in progress)

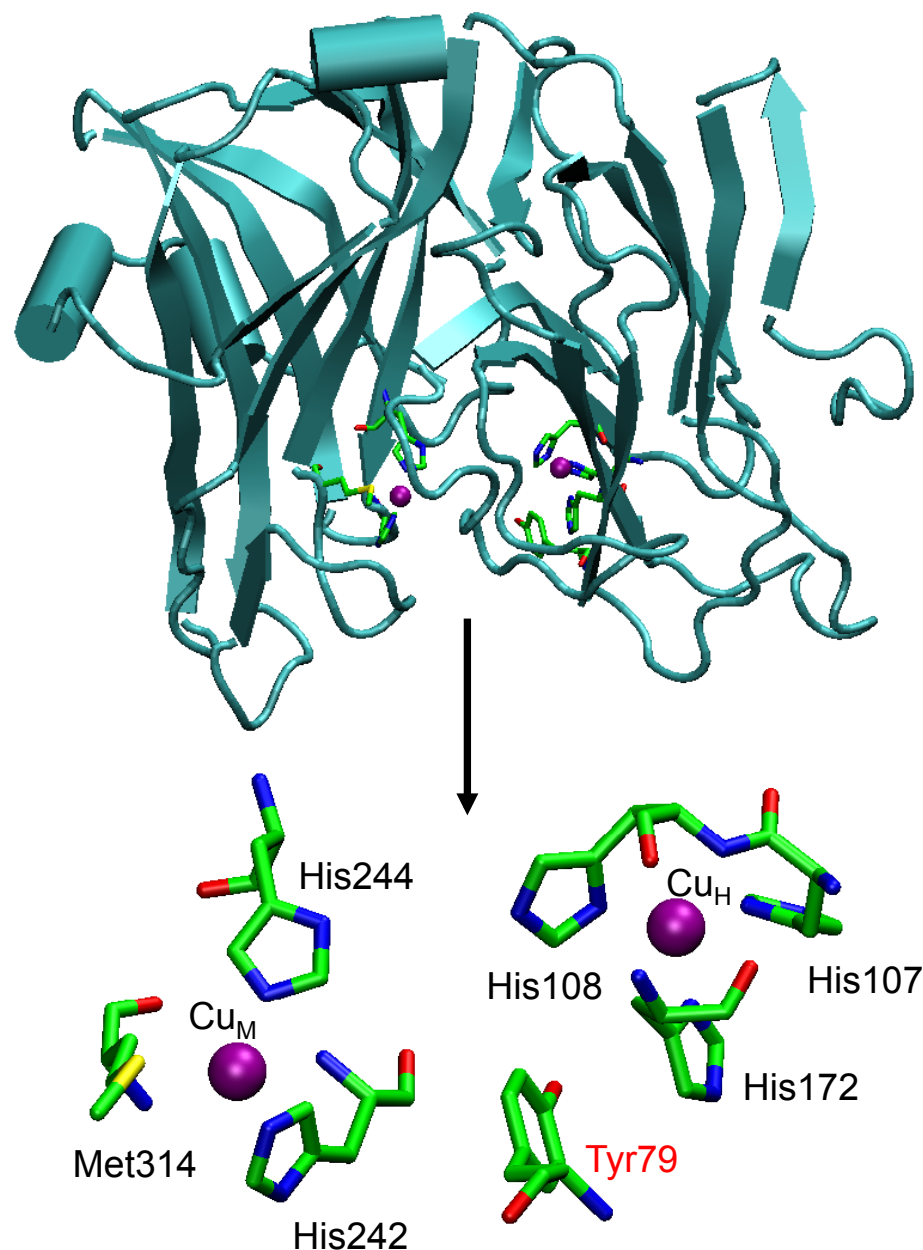
$$V_{DA}^{\text{tp-DFT}} = f_{\text{medium}}^{\text{tp}} V_{DA}^{\text{DFT}}$$

The electronic coupling from *ab initio* computation can be suitably “dressed” with a factor determined by means of the TP method.

Important for quantitative V_{IF} analysis.

Expected to be fruitful in studying ET pathways and their fine tuning by ET-relevant features in protein-DNA.

Quantum portions can be pruned as in corresponding QM/MM investigation.



Molecular dynamics

MD techniques efficiently employed to study the thermodynamics of electron flow in proteins. An emblematic one is reported below, about Deca-heme Cytochrome MtrF.

The **thermodynamic integration** protocol is used to obtain the reaction (Gibbs) free energy difference ΔG between two equilibrium states.

MD vertical excitation energies can be used to compute ΔG^0 , and also λ .

$$\Delta G_{\text{oxid.}} = \int_0^1 \left\langle \frac{\partial E_\eta}{\partial \eta} \right\rangle_\eta d\eta \quad E_\eta = \eta E_{\text{ox}} + (1-\eta) E_{\text{red}} : \text{not allowed by NAMD; instead, for the}$$

changing charges,

$$q_{j\eta} := \eta q_j^{\text{ox}} + (1-\eta) q_j^{\text{red}}$$

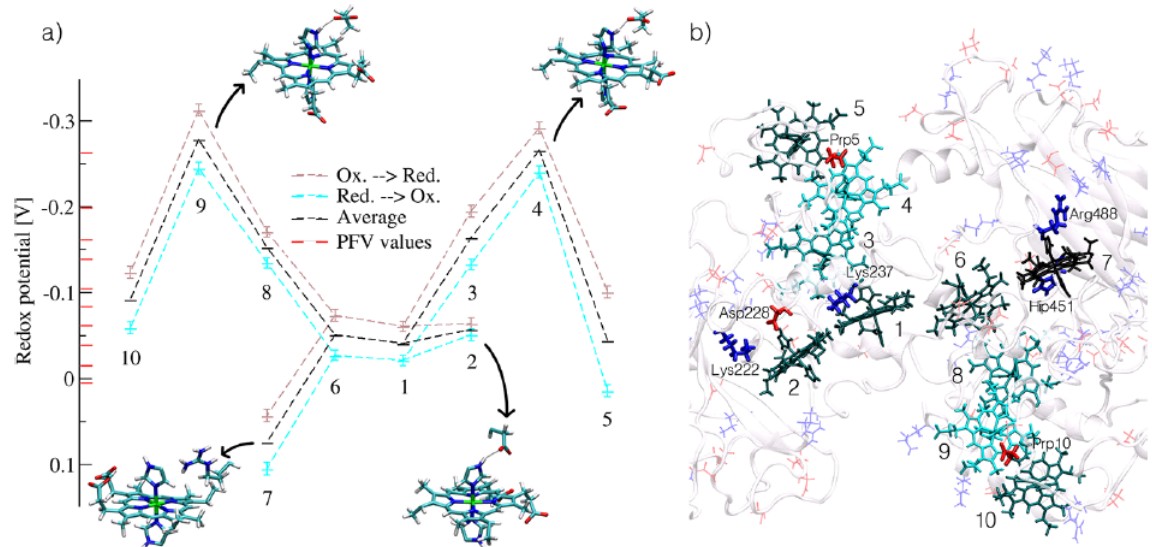
and thus

$$\frac{\partial E_\eta}{\partial \eta} = E_{\text{ox}} - E_{\text{red}} + (2\eta - 1) E_\Delta$$

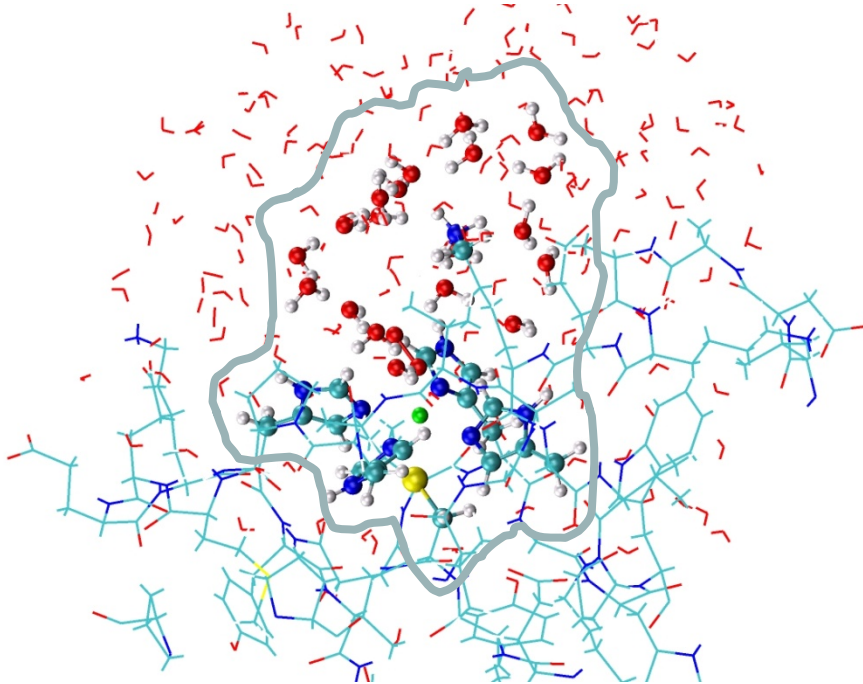
small \uparrow

redox potential $\varepsilon = \Delta G/F + C$

cf. experiment \rightarrow offset C



QM/MM method



$$\hat{H}_{\text{QM/MM}}^{\text{el}} = - \sum_{s=1}^S \sum_{\substack{I=1 \\ I \neq s}}^N \frac{eq_s}{r_{Is}} + \sum_{s=1}^S \sum_{m=1}^M \frac{q_s Z_m}{R_{ms}}$$

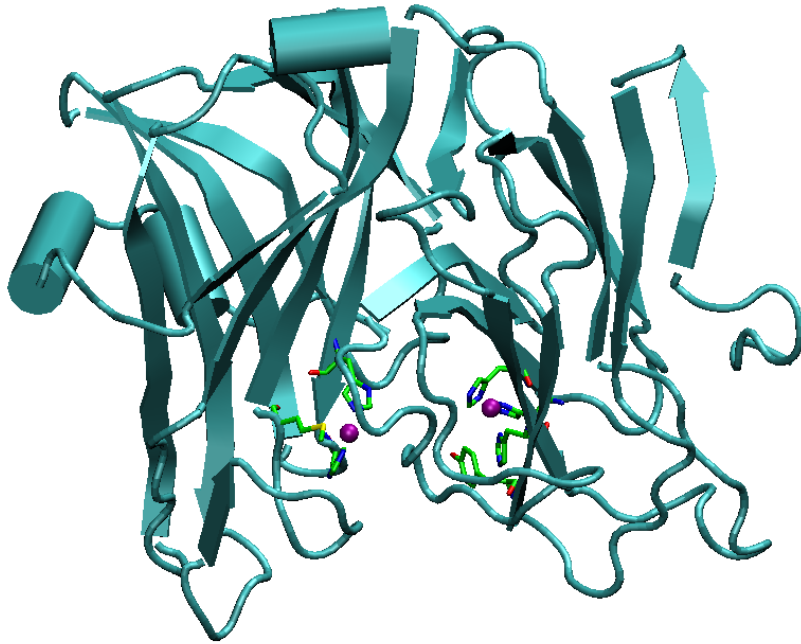
$$\hat{H}_{\text{QM/MM}}^{\text{vdW}} = \sum_{s=1}^S \sum_{m=1}^M 4\epsilon_{ms} \left[\left(\frac{\sigma_{ms}}{R_{ms}} \right)^{12} - \left(\frac{\sigma_{ms}}{R_{ms}} \right)^6 \right]$$

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}}^{\text{p}} + \hat{H}_{\text{QM/MM}}^{\text{el}} + \hat{H}_{\text{QM/MM}}^{\text{vdW}} + \hat{H}_{\text{MM}}$$

Warshel and Levitt, *J. Mol. Biol.* **103**, 227 (1976).

Laio, VandeVondele and Rothlisberger, *J. Chem. Phys.* **116**, 6941 (2002).

- Study of the reaction dynamics and thermodynamics.
- Refined computation of free energy parameters and possibly electronic couplings.
- Effects of specific quantum residues and water on the reaction rate.



After MD: hybrid-DFT computation of electronic coupling on snapshots

$$|\psi_I\rangle = |D, A\rangle \quad |\psi_F\rangle = |D^+, A^-\rangle$$

$$|\psi_{GS}\rangle = a|\psi_I\rangle + b|\psi_F\rangle,$$

with and without:

- bridge,
- relevant moieties

Number of required computations:

Direct bare coupling: 3×4

Bridge-mediated coupling: 3×2

Stepwise:

First step: 3×2

Second step: 3×2

Total: **30**

Two XC density functionals: **60**

Two basis sets: **120**

Two structures: **240**

Electronic couplings
Excitation energies
Inner-sphere /

active sites
pruning

charge donor $|D\rangle, |D^+\rangle$

stepwise

direct tunneling

charge transfer bridge

other relevant groups

charge acceptor $|A\rangle, |A^-\rangle$

60 – 110 atoms treated quantum mechanically

depending on the model and presence of bridge, strategical residues, mutation simulations.

e.g., **2.5 production MD run**

Ideal, statistically meaningful **quantum computation on 100-250 snapshots**

Minimum number of quantum chemistry computations per snapshot:

3 (mediated tunneling from donor to acceptor)

6 (stepwise charge transfer; two steps)

Total ideal number of quantum coupling computations: 900 – 2250.

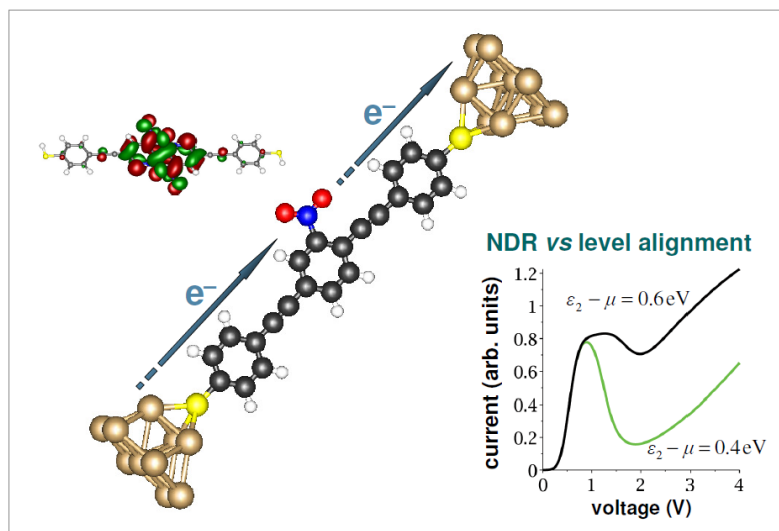
Computations using **semi-empirical methods** on higher number of snapshots.

Other computations (using MD or QM/MM):

Free energy of reaction (MD or QM/MM)

Reorganization energy: outer-sphere (MD), inner-sphere (QM).

ET in electrochemical setup



Analytical expressions of the ET rates at metal-molecule interfaces, recently obtained to study charge transport properties of redox junctions, can be used and have been applied to study the voltammetric setup.

Accessible computational quantities involved.

$$R_{\text{metal} \rightarrow \text{molecule}} = \frac{\gamma}{4} S(\lambda, T, \alpha) \exp\left[-\frac{(\alpha - \lambda)^2}{4\lambda k_B T}\right]$$

$$R_{\text{molecule} \rightarrow \text{metal}} = \frac{\gamma}{4} S(\lambda, T, \alpha) \exp\left[-\frac{(\alpha + \lambda)^2}{4\lambda k_B T}\right]$$

g = coupling strength, depending on the electronic coupling; V = overpotential; $\alpha = \mu - E_{AB} + eV$

$$S(\lambda, T, \alpha) = \sum_{n=0}^{\infty} \frac{1}{2^n} \sum_{j=0}^n (-1)^j \binom{n}{j} [\chi_j(\lambda, T, \alpha) + \chi_j(\lambda, T, -\alpha)]$$

$$\chi_j(\lambda, T, \alpha) = \exp\left\{\frac{[(2j+1)\lambda + \alpha]^2}{4\lambda k_B T}\right\} \operatorname{erfc}\left[\frac{(2j+1)\lambda + \alpha}{2\sqrt{\lambda k_B T}}\right]$$

A Migliore, A Nitzan,
ACS Nano **5**, 6669 (2011);
J. Electroanal. Chem 671, 99 (2012)

Computation of outer-sphere /for molecule-metal interfaces.

A possible approach consists in the atomistic description, at a QM level, of the protein active site, coupled to the other components of the system described as continuous media according to the polarizable continuum model (PCM): see S. Corni, The reorganization energy of Azurin in bulk solution and in the electrochemical scanning tunneling microscopy setup. *J. Phys. Chem. B* **109**, 3423 (2005) (implemented in GAMESS).

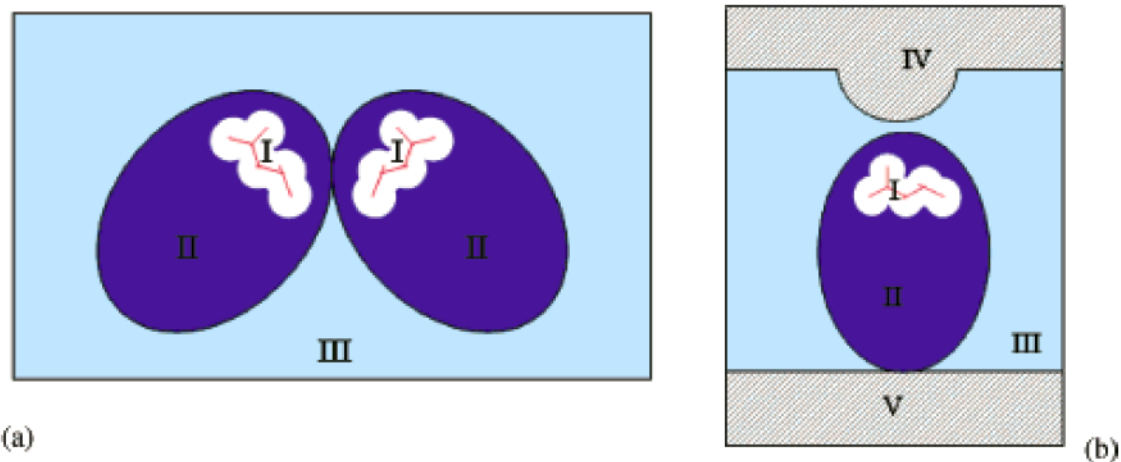


Figure 1. Schematic representations of the models used in the present work: (a) ESE reaction of Az; (b) Az in the ECSTM geometry. The numbering refers to regions with different dielectric properties: I, active site cavity (the atomistic active site, treated quantum-mechanically, is sketched in red); II, protein scaffold; III, solvent; IV, STM tip; V, STM substrate. Drawings are out of scale.

Computational tools

- **NAMD (1)**
MD: snapshots for coupling computation, reorganization energy, reaction free energy
- **NWChem (2), Gaussian (3), GAMESS (4) and Quantum Espresso (5)**
Electronic couplings, excitation energies, other electronic properties, inner-sphere reorganization free energy
- **VMD program, and Pathways plugin (6)**
Visualization, and exploring biological electron transfer pathway dynamics
- **CP2K (7), QM4D (8)**
Average packing density between redox sites
- **HARLEM program (9)**
Other semi-empirical : average packing density between redox sites

(1) <http://www.ks.uiuc.edu/Research/namd/>

(2) http://www.nwchem-sw.org/index.php/Main_Page; (3) <http://www.gaussian.com/>;

(4) <http://www.msg.ameslab.gov/gamess/>; (5) <http://www.quantum-espresso.org/>

(6) <http://www.ks.uiuc.edu/Research/vmd/>

and I. A. Balabin, X. Hu, D. Beratan, *J. Comput. Chem.* **33**, 906 (2012).

(7) <http://www.cp2k.org/>; (8) <http://www.qm4d.info/trac/QM4D/login>

(9) I. V. Kurnikov, <http://www.kurnikov.org/>

Conclusion

Qualitative and quantitative progress requires large-scale quantum chemistry computation, and not just its marginal use for some refinement and secondary electronic structure computation. Nowadays, such a large-scale computation is feasible, but suitable computational resources are needed.

Acknowledgements

Thank you for your attention