The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2145-19

Spring College on Computational Nanoscience

17 - 28 May 2010

QM and Hybrid QM-MM Simulation of Biomolecules
Part I

Dario A. ESTRIN
Univ. of Buenos Aires
Argentina

QM simulation of biomolecules

Darío A. Estrin

University of Buenos Aires, Argentina

Biomolecules: Wikipedia

Any organic molecule that is produced by a living organism, including large polymeric molecules such as proteins, polysacharides, and nucleic acids, as well as small molecules such as primary metabolites, secondary metabolites, and natural products.

Biomolecules contain: C, N, O, H, and to smaller extent P, S, and metals, such as Fe, Cu, Mo, etc.

QM based simulations:

Essential in cases in which there is a process which needs electronic level description.

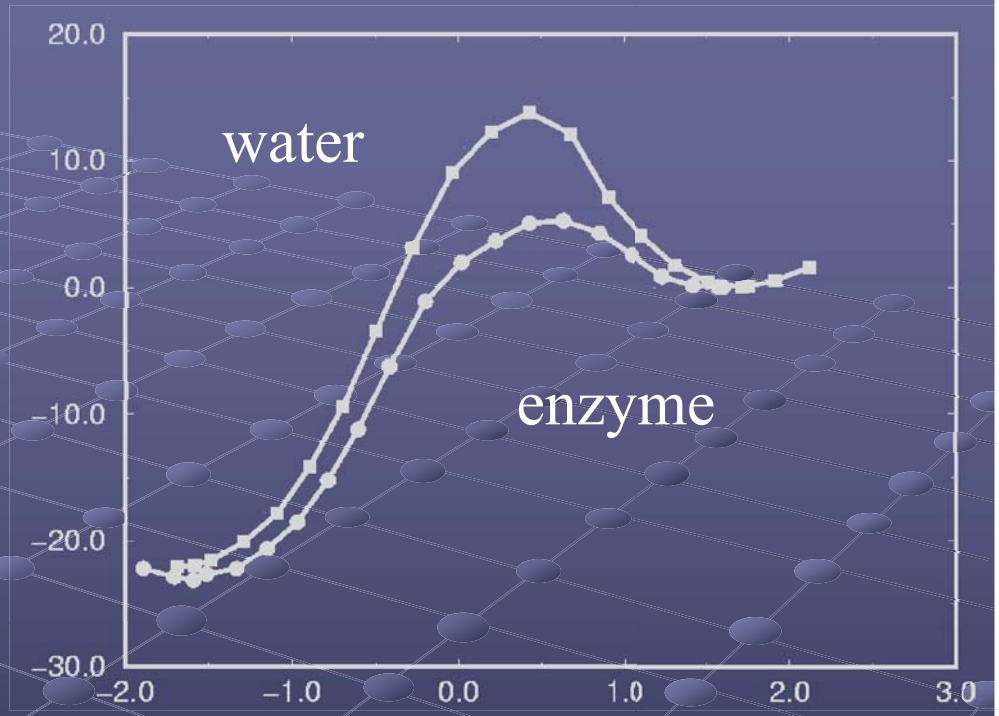
In this context: the protagonist biomolecules are proteins which typically act catalyzing chemical reactions

Final goal: to answer:

Structural properties: geometry optimizations, molecular dynamics (fluctuations!).

Thermodynamical properties: energetic changes associated with reactions, energy profiles, free energy profiles.

Example: chorismate mutase



Catalytic effects!!!!

- We can understand effects: environment, role of specific aminoacids, build in silico mutations, design systems with a given property!

Key question:

To investigate small model systems at high levels of theory or to describe more complex systems at lower levels.

Illustrative cartoon taken from A. Ghosh

in Current Opinion in Chem. Biology, 2003

Guest section: Computational bioinorganic chemistry. Part III. The tools of the trade: from high-level ab initio calculations to structural bioinformatics

Everyone's working on MMO these days. So what have you got here?

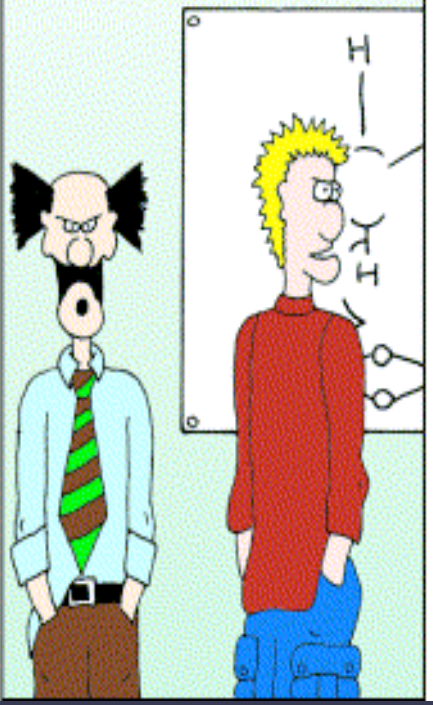
Well, I'm a quantum CHEMIST, just helping out my Biochemistry colleagues with some DFT calculations on the mechanism...

I'm a QUANTUM chemist too... and I have also had my eyes on the MMO problem but I don't want to do DFT. Y'know that's really a semiempirical method and I want to get the right answer for the right reason.

(How many times have I heard that?)

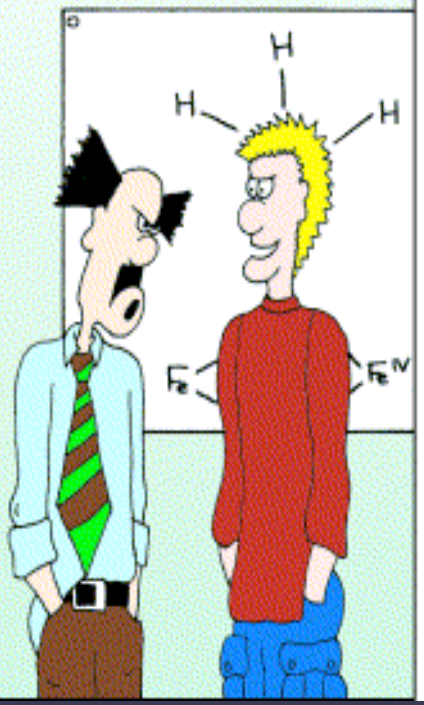
Well, the results from my calculations seem all right; they agree with all the experimental constraints...

I personally find it more interesting when a calculated result doesn't agree with the experiment...



(Do they always spout these cliches?)

Well, my adviser says quantum chemistry methods development has gone as far as it will and now's the time to take DFT and do all the exciting applications in biology.

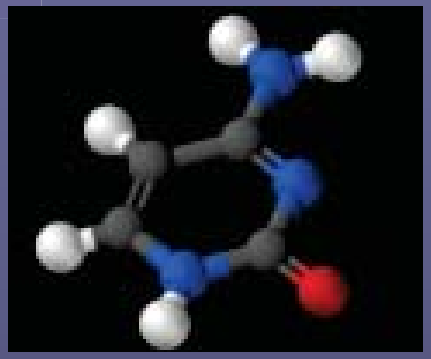


Your adviser? Oh, he! He used to be one us, you know, did careful work on small molecules. Now he's completely gone over to the bio camp, just wallowing in bio-goo.

Bio-goo?!?

Y'know, proteins and DNA and whatnot. Talking about bio-goo, here come the bioinformaticists, I'd better clear out!

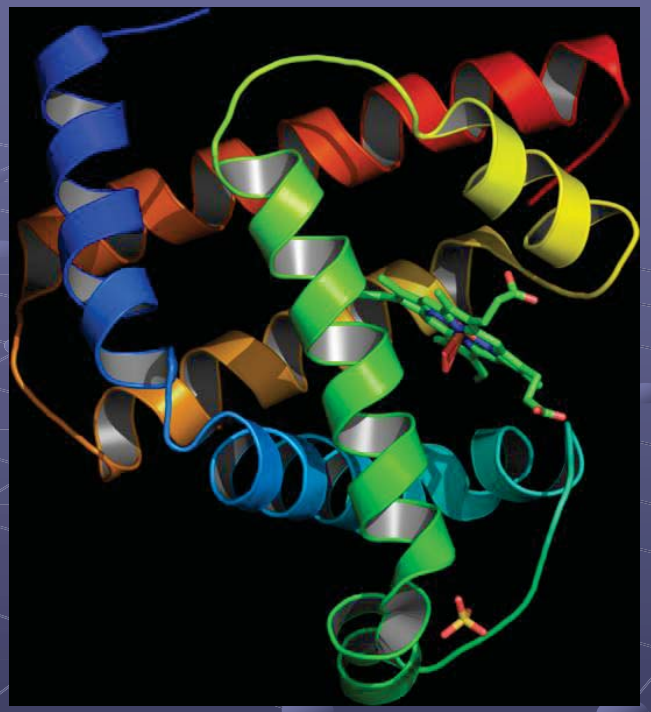
Small model systems: QM enough



One can resort to any quantum method, and employ standard optimizations, saddle point searches, etc.
(convenional approaches used in Quantum Chemistry)

More complex systems:

Standard approaches are not enough.
Adequate consideration of sampling and environment effects are crucial!



Why? because potential energy surfaces become more complex

Validating QM method in small model systems is always good as a first step before including environment and thermal fluctuations (next talks!).....

So, performing pure QM calculations is necessary

Quantum Models

- Models based on Quantum Mechanics:
  Solve (or tray!) Schrödinger´s equation.

- Basic approximation: Born-Oppenheimer

Different levels of approximation:
  semiempirical, Hartree-**Fock, post Hartree Fock, DFT….**

- Indispensable for: chemical reactivity, excited states

- In principle: valid for any system

- In practice: size limitations due to computational expense

First approximation: Hartree Fock

- Wave function: Slater determinant
  Assignment of a molecular orbital for each electron

- Mathematically: Hartree-Fock equations: not a good choice if transition metals are present.
- Nowadays practically not used

Post Hartree-Fock methods

- Correct HF assumptions
- Much more expensive computationally

- MP2 (Perturbation Theory)
- CI (variational). The best choice for excited states. Variations: CASSCF, etc
- The best choice for transition metals

Example: CASPT2 multiconfiguration + perturbation:

difficult to use....

Most employed method: DFT

- Density Functional Theory: ground state theory
- Different flavors: LDA, GGA (i.e. PBE), hybrid functionals (i.e. B3LYP)
- Efficient computationally. Reasonable results. OK with metals….
- Excited states extension: TDDFT

QM: *DFT*

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{ne}[\rho]$$

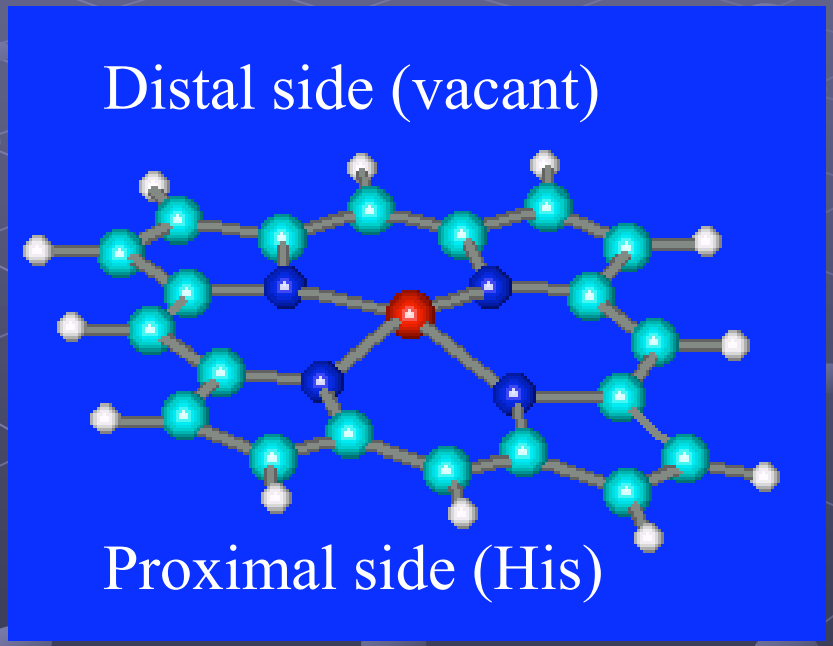
$$E[\rho] = -\frac{1}{2}\langle\chi_i^{KS}|\sum_i \nabla_i^2|\chi_i^{KS}\rangle + \int v(\mathbf{r})\,\rho(\mathbf{r})\,d\mathbf{r}$$

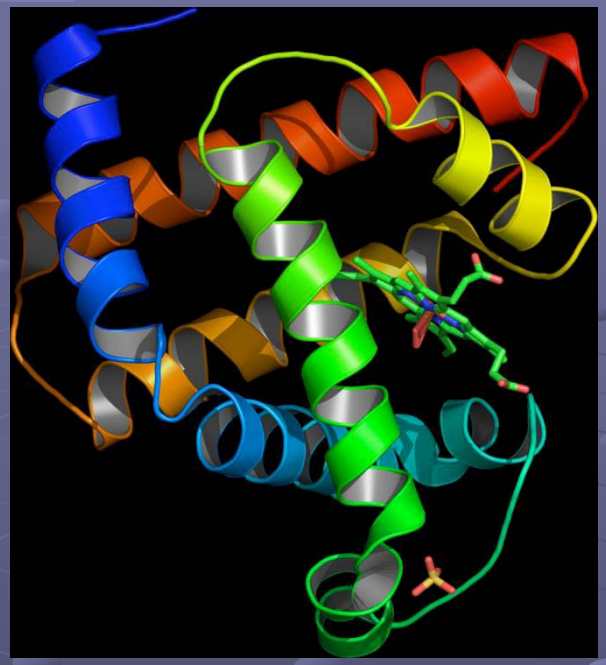
$$+\frac{1}{2}\int\int \frac{\rho(\mathbf{r_1})\rho(\mathbf{r_2})}{\mathbf{r_{12}}} d\mathbf{r_1} d\mathbf{r_2} + E_{xc}[\rho]$$

$$\rho = \sum_{i=1}^{N} |\chi_i^{KS}|^2$$

Nice example: Hemeproteins:

Active site: heme
Fe porphyrin





Very different roles:

Transport of O_2

Hormone Biosynthesis

Detoxification

Electron Transfer

Example: Heme-O_2 Binding

Old classical problem: is it $Fe(II)O_2$ or $Fe(III)O_2^-$?

Pauling, Nature, 203, 182, 1964.

Still controversial:

CASPT2, JIB, 2005, 99, 45-54

Says it is multiconfigurational...

DFT: predicts $Fe(III)O_2^-$

Open problem:
Cases in which different spin states are accesible
Spin gaps. Crucial issue in determining accurate interaction energies

Fe(II)P (quintuplet)+O_2(triplet) yields Fe(II)PO_2 which is a singlet!!!!

7384 *J. Phys. Chem. B* **2007**, *111*, 7384−7391

Simulation of Heme Using DFT + U: A Step toward Accurate Spin-State Energetics

Damián A. Scherlis,*,† Matteo Cococcioni,‡ Patrick Sit,§ and Nicola Marzari§

Correlated methods are fine, but you need to validate a lower cost method if you want to include environment and thermal effects DFT+U may be a good choice!

TABLE 1: Experimental and Calculated Electronic Ground States of Five and Six-Coordinated Iron Porphines (FeP), with the Following Axial Ligands: O_2, CO, Imidazole (Im), and Chloride

	six-coordinated		five-coordinated		
	FeP(Im)(O_2)	FeP(Im)(CO)	FeP(CO)	FeP(Im)	FeP(Cl)
experimental	singlet	singlet	singlet	quintuplet	sextet
Hartree−Fock	quintuplet	quintuplet	quintuplet	quintuplet	sextet
DFT-GGA	singlet	singlet	singlet	triplet	quartet
B3LYP	singlet	singlet	singlet	triplet	quartet/sextet

TABLE 2: Spin-Transition Energies (kcal/mol) for the Low-Lying Spin States of $Fe^{II}P(Im)$ Calculated with Several Density Functionals and with DFT + U, Using $U_{sc} = 3.9$ eV

	singlet	triplet	quintuplet
	DFT		
PBE[a]	7.8	0.0	7.9
BP86[b]	8.3	0.0	6.5
B3LYP[c]	5.8	0.0	1.9
	DFT + U		
	20.9	4.9	0.0

Results very sensitive to level of theory!

TABLE 4: Spin-Transition Energies (kcal/mol) for the Low-Lying Spin States of $Fe^{II}P(Cl)$ Calculated with Highly Correlated Methods and Density-Functional Theory, Including DFT + U (U_{xc} = 4.0 eV)

	quartet	sextet
CASPT2[a]	19.6	0.0
RCCSD(T)[b]	16.1	0.0
DFT-PBE	0.0	5.6
DFT-PW91[a]	0.0	8.1
DFT + U	9.2	0.0

Note that results depend a lot on the level of theory used: Though problem!

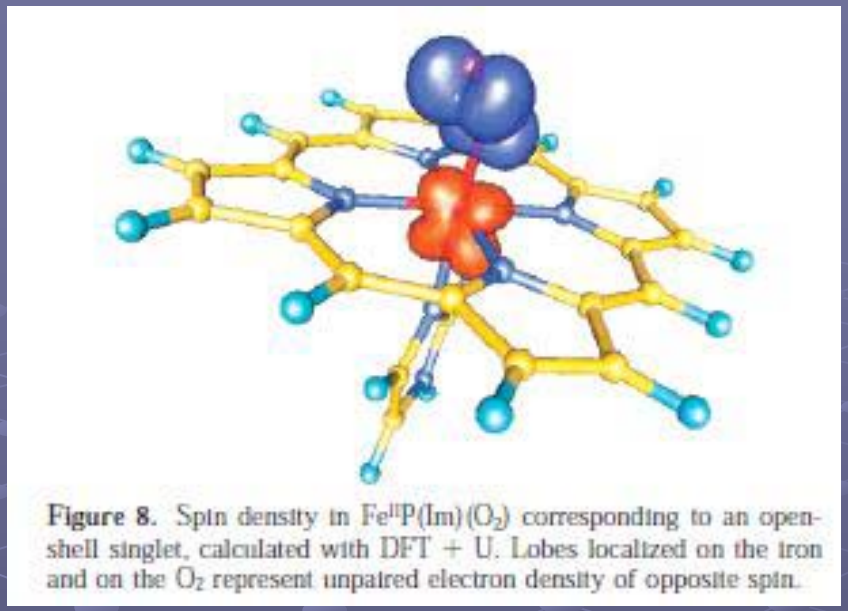
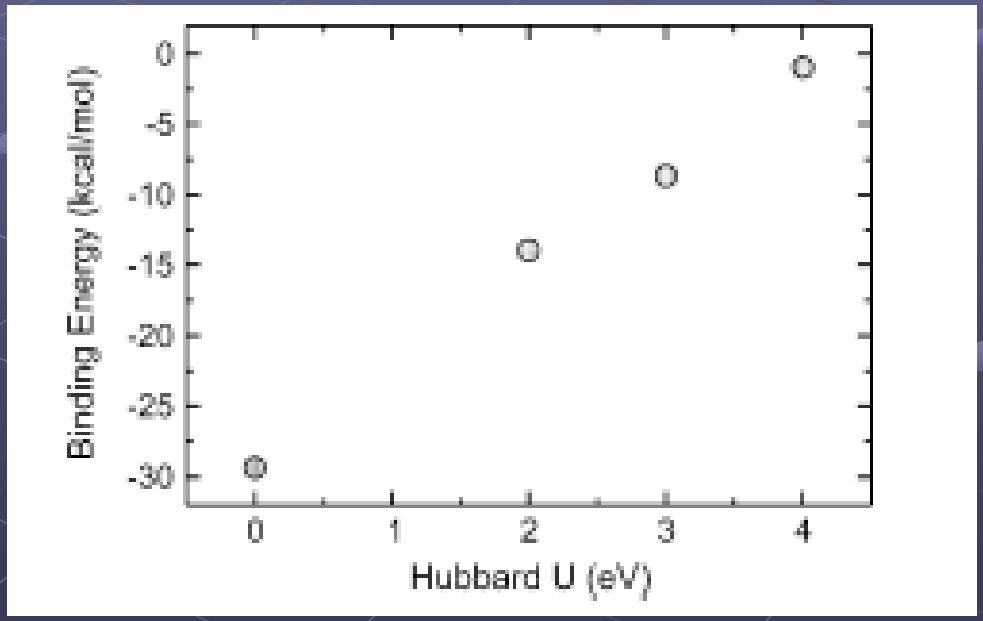


Figure 8. Spin density in $Fe^{II}P(Im)(O_2)$ corresponding to an open-shell singlet, calculated with DFT + U. Lobes localized on the iron and on the O_2 represent unpaired electron density of opposite spin.

GGA+U : predicts open shell Singlet for the $Fe(II)O_2$ species



Binding energy depends on the Hubbard parameter….. It can be calibrated

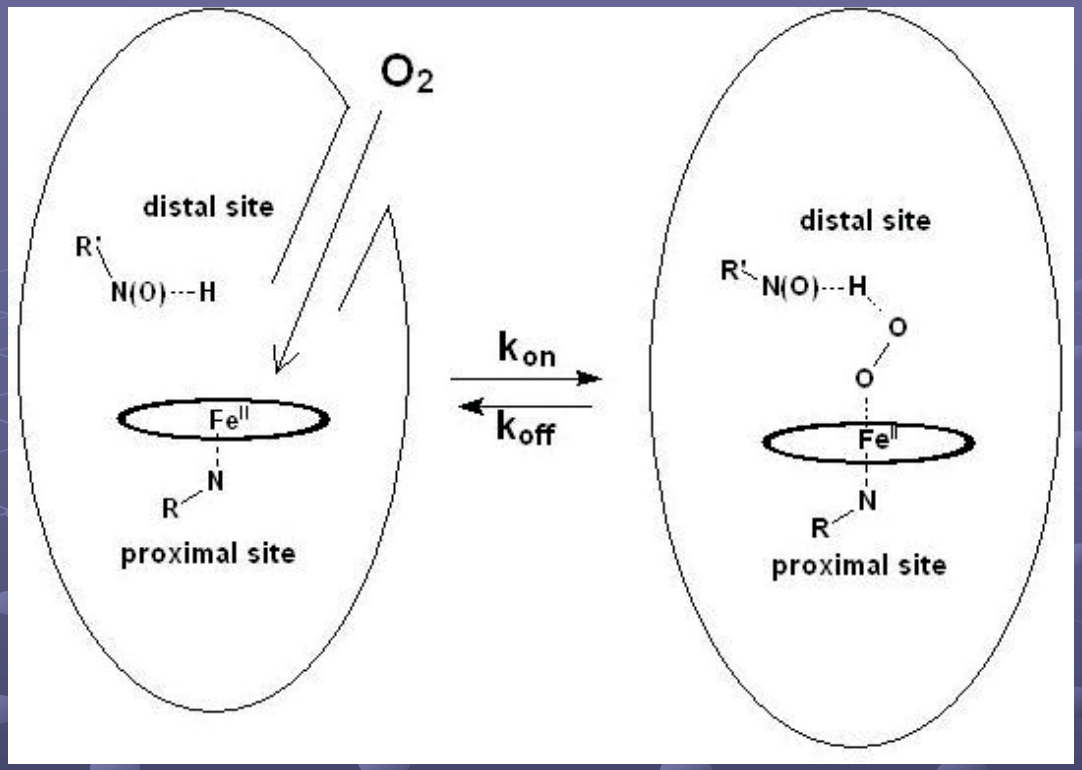
Experimental estimation: about -5 kcal/mol

However, if the interest is on trends, not so good methods (such as standard DFT at PBE level) may be useful

Nice example: see how the protein environment affects a given property. For example, oxygen affinity.

This can be tackled by QM-MM calculations (next talk), but in a first approach it can be studied by considering small QM models.

Dioxygen binding in Globins



Association process

$O_2(sn) + Prot \rightarrow O_2\ Prot$

$O_2\ Prot \rightarrow O_2\text{-}Prot$

k_{on}: mainly related with ligand migration

Classical MD simulations

k_{off}: mainly related with bond breaking:

Distal (direct) effects

Proximal (indirect) effects

Dissociation process

$O_2\text{-}Prot \rightarrow O_2\ Prot$

$O_2 + Prot \rightarrow O_2(sn) + Prot$

QM or QM-MM calculations

Proximal effects

Histidine rotational position

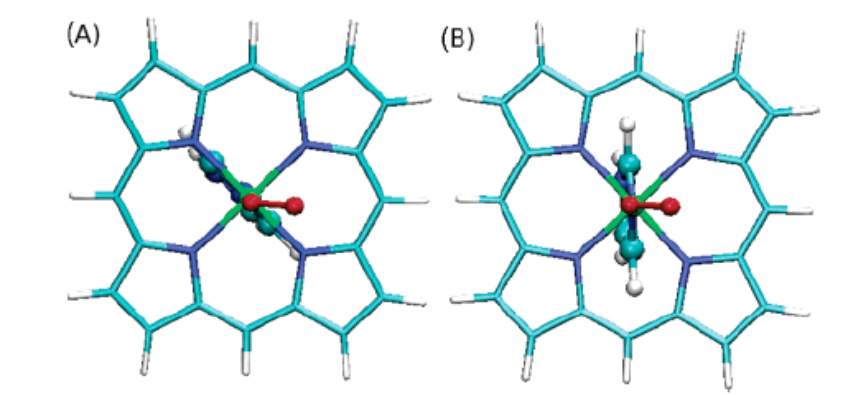
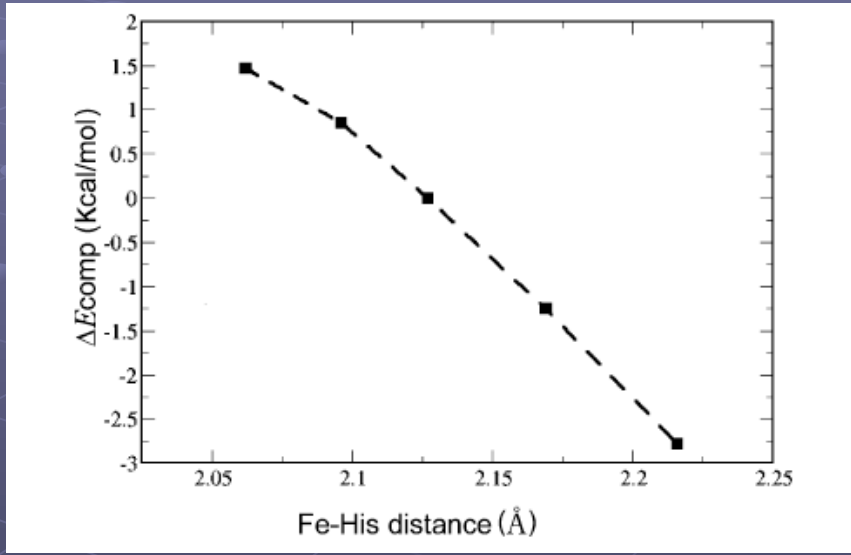


Figure 2. Histidine rotational position: (A) eclipsed and (B) staggered.

Staggered: oxygen affinities from 1 to 3 kcal/mol larger than eclipsed

Capece et al, J. Am. Chem. Soc. 128, 12455, (2006).

Fe-His distance

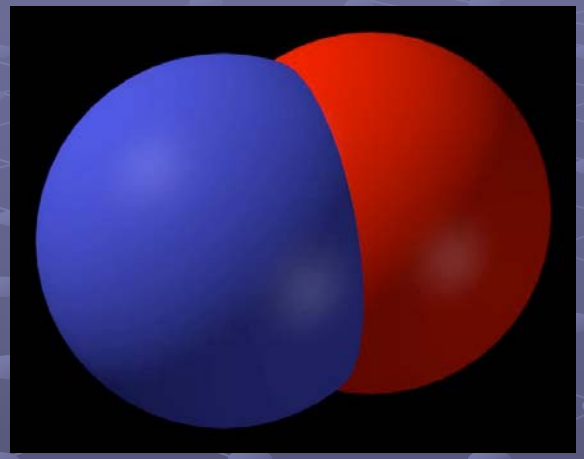


Binding energy increases when distance is constrained at smaller values

Message: even if absolute values of binding energy are not very good, the method is good enough to represent correctly the environment effects.

In some cases, actual values may be wrong (specially binding energies), but the overall trends are correct

Another physiologically relevant diatomic molecule: nitric oxide (NO)



Very reactive, free radical

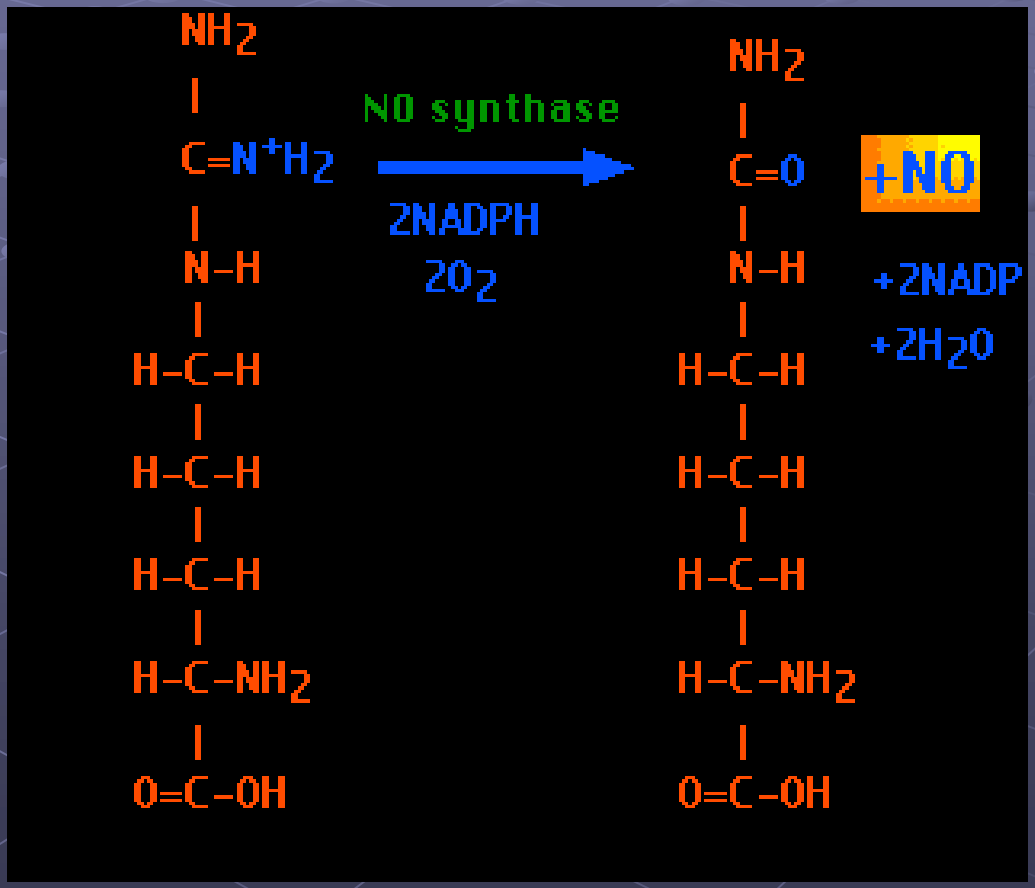
Inmense importance in physiology, pathology, and inmunology

Molecule of the year 1992

Physiological targets of NO: heme groups, thiol groups (SH)

NO biosynthesis:

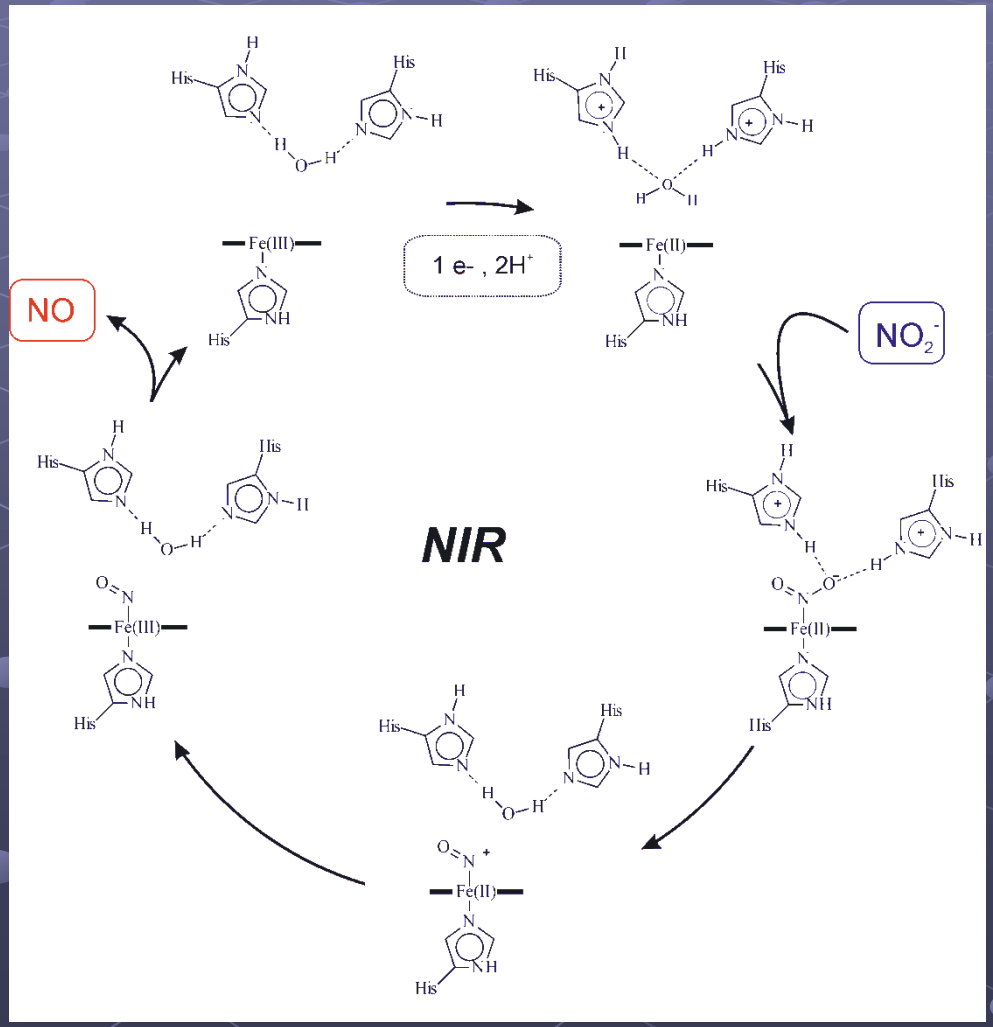
Most accepted pathway involves heme protein called NOS(nitric oxide synthase)



arginine cytrulline

An alternative NO generation mechanism: Operative in anoxia (absence of oxygen)

NO_2^- yields NO



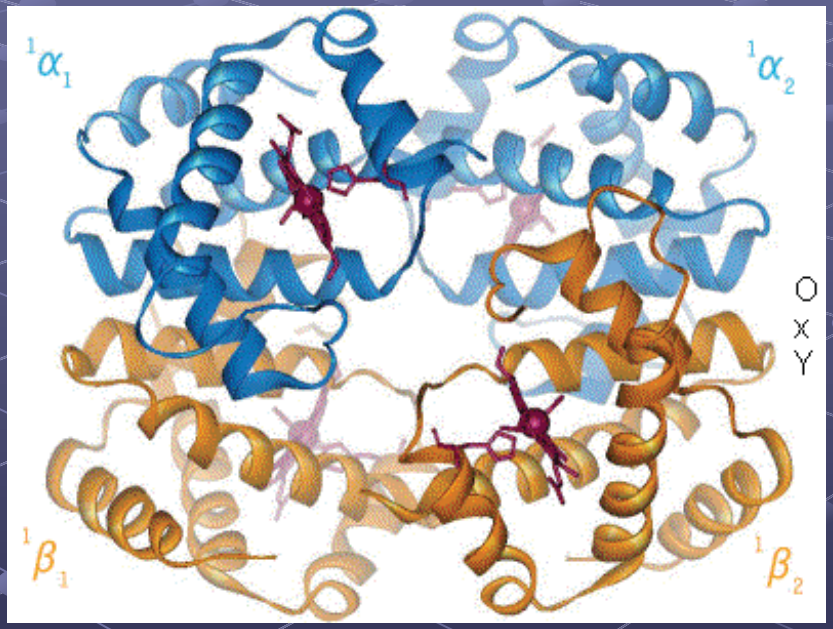
Known mechanism for Bacteria

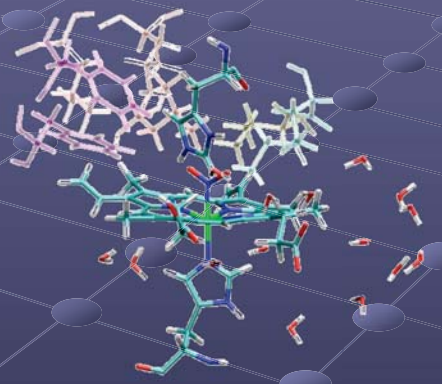
Enzime: called NIR

It has been shown recently that a similar reaction may be catalyzed by human hemoglobin

$NO_2^- + HbFe^{2+}(deoxy\text{-}Hb) + 2H^+ \rightarrow HbFe^{3+}(metHb) + NO + H_2O$

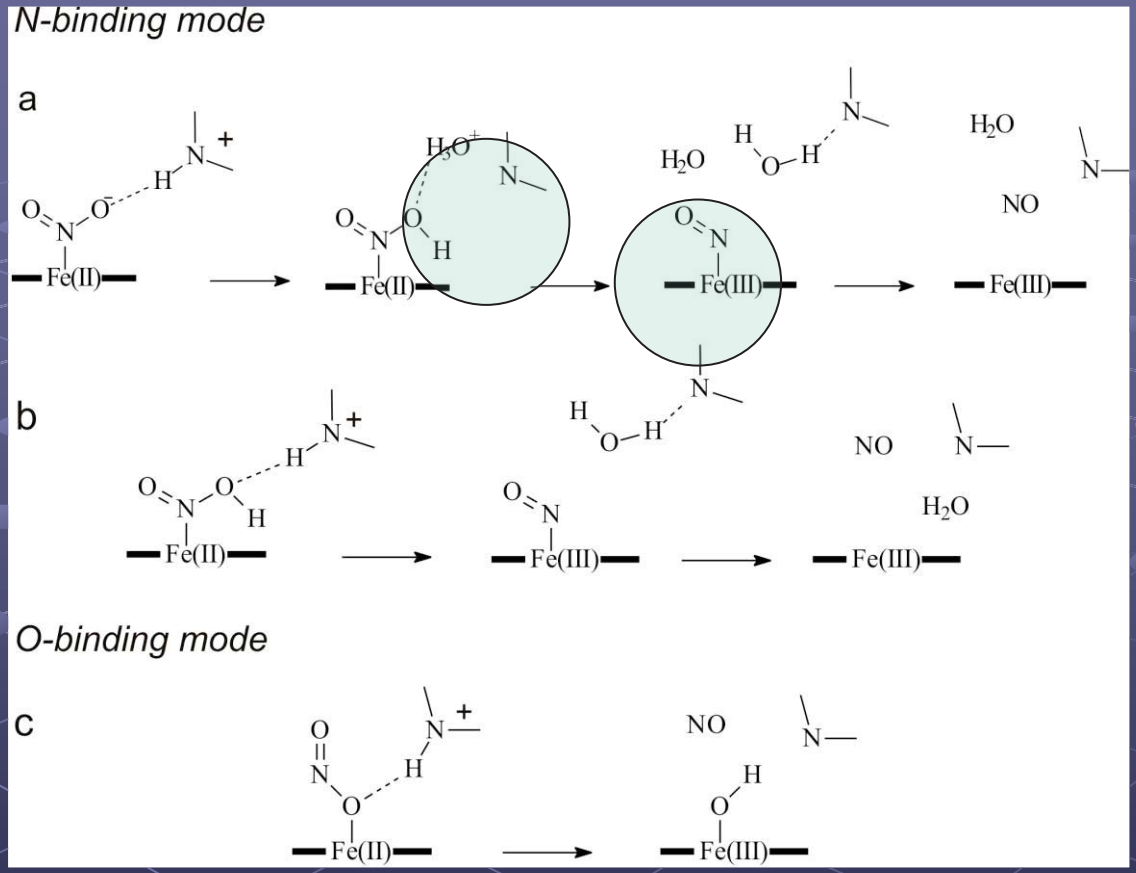
Mechanism: involves proton transfer steps and a final reduction of the active site





However, Hb has only one histidine residue, so mechanism should be different

NO generation



Key questions:

Active species
NO_2^- or HONO ?

pKa=3.14, 25 °C

$[NO_2^-]/[HONO]=9500$

Gets coordinated through N or O atoms?

What about proton Transfer?

Rate limiting step?

Silaghi-Dumitrescu, R., *Inorg. Chem.* **2004,** 43, (12), 3715-3718.

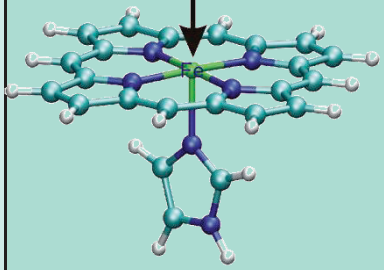
Copeland, D. M.; Soares, A. S.; West, A. H.; Richter-Addo, G. B., *J. Inorg. Biochem.* **2006,** 100, (8), 1413-1425.

N or O coordination?

-QM:

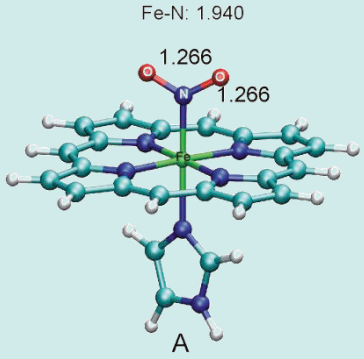
Model systems

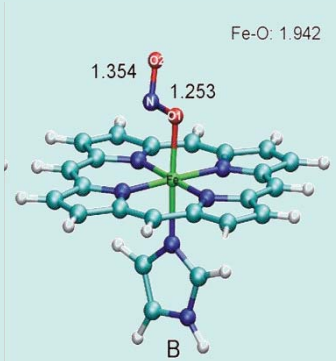
Site of binding

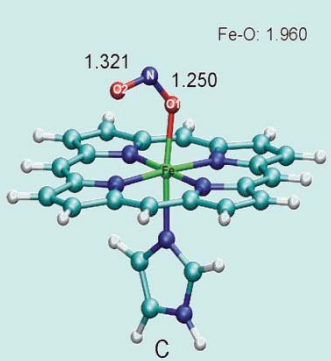


DFT: pbe
SIESTA

Energía (kcal/mol)





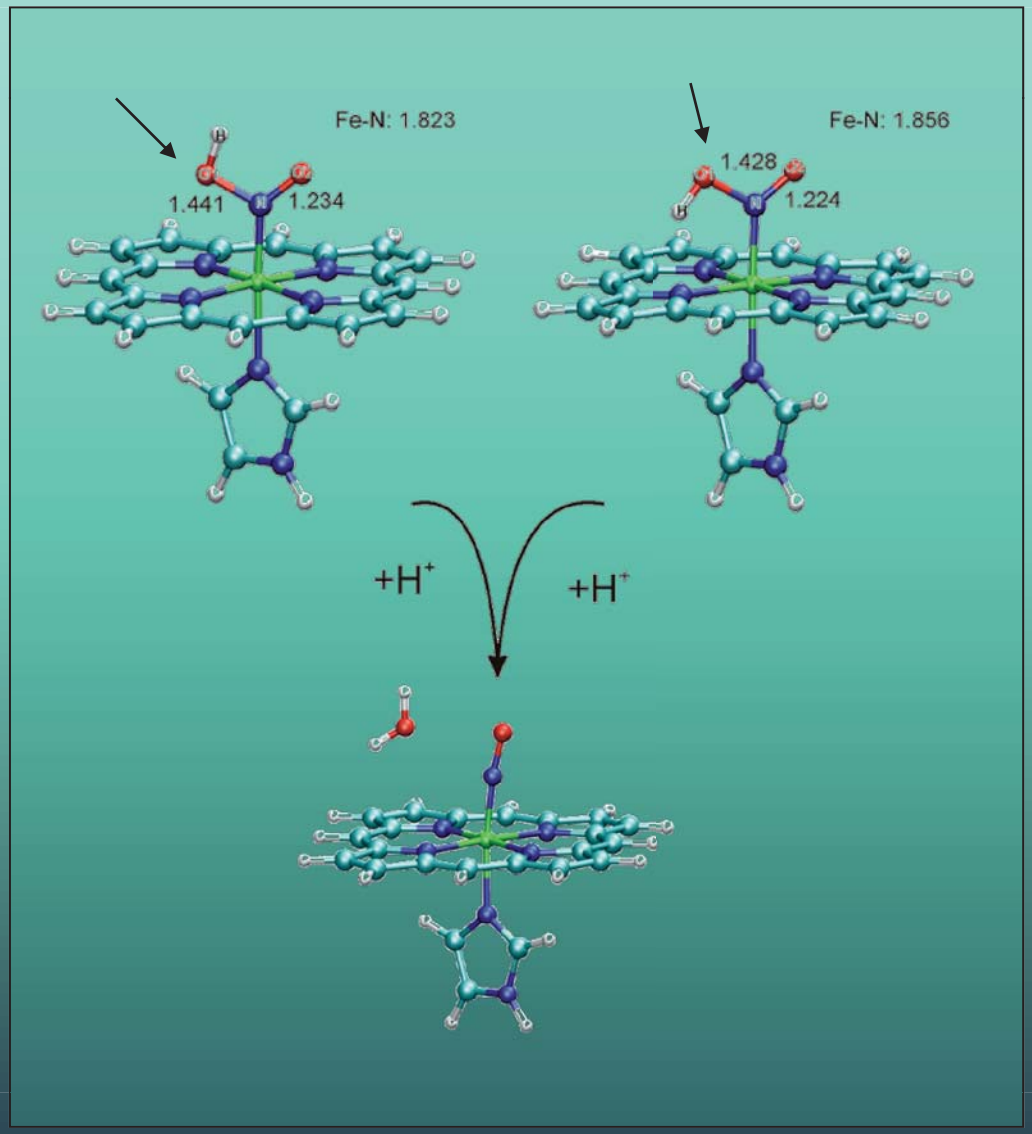


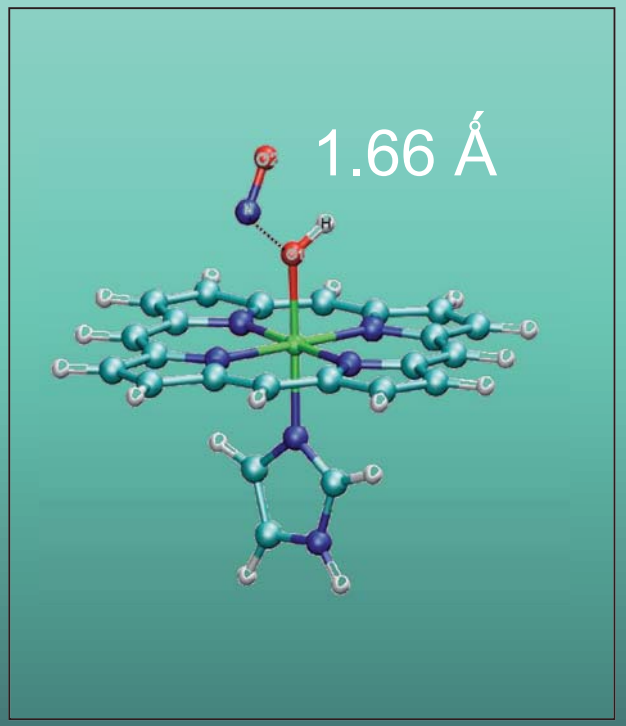
~1 kcal/mol

~10 kcal/mol

What happens inside the protein? Requires QM-MM!

What about nitrous acid coordination?





Up to now:

It seems that both coordination forms are feasible in principle.
It may also coordinate as nitrous acid.

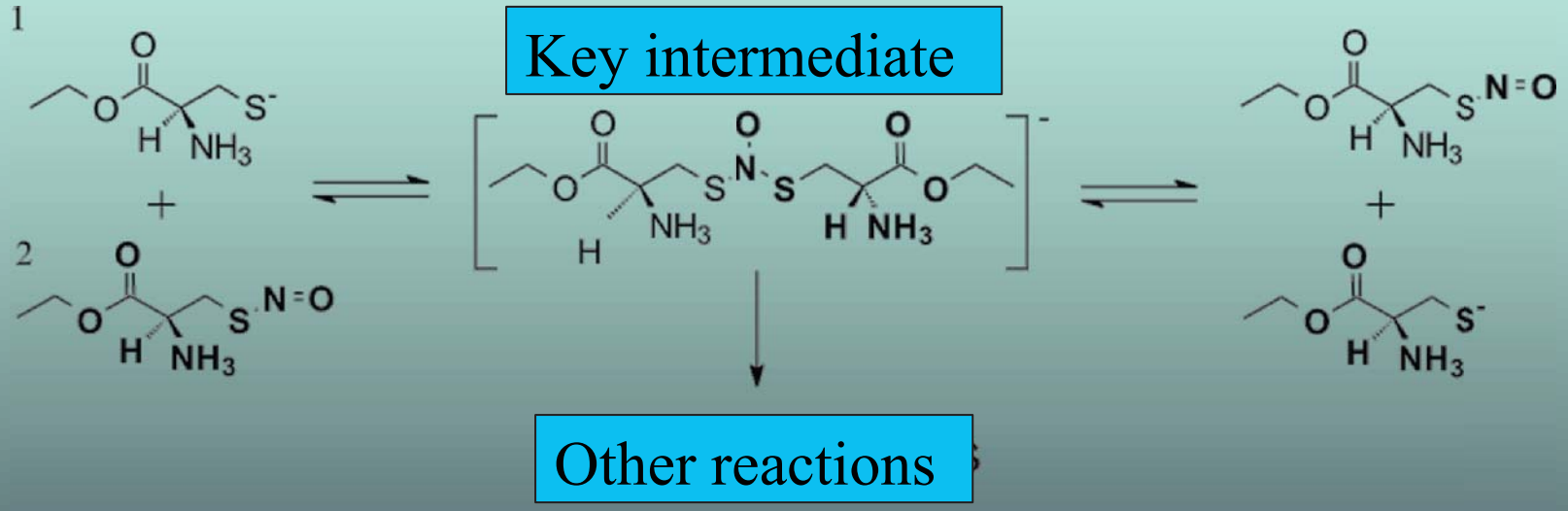
Final answers require consideration of protein effects….

Classical MD followed by QM-MM simulations!

Perissinotti et al, Biochemistry, 47, 9793, 2008.

Other problem related to NO:

NO transport, it is assumed that NO travels transported by nitrosothiols (mainly cysteine, aminoacid containing SH group)



Transnitrosation reaction

Limiting step

$$\mathbf{RS}^- + \mathbf{R}'\mathrm{SNO} \rightleftharpoons [\mathbf{RS}(\mathrm{NO})\mathbf{SR}']^-$$

$$[\mathbf{RS}(\mathrm{NO})\mathbf{SR}']^- \rightleftharpoons \mathbf{R}'\mathrm{S}^- + \mathbf{RSNO}$$

Reactive species

$$\mathrm{RSH} \overset{K_a}{\rightleftharpoons} \mathrm{RS}^- + \mathrm{H}^+$$

- **Active species : RS$^-$**

Second order law kinetics

Rate law

$$v = k[R'SNO][RS^-]$$

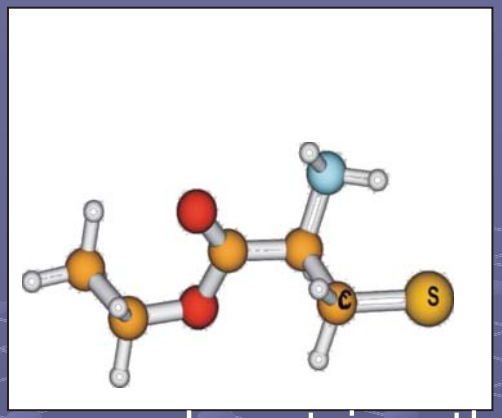
$$v = \frac{kK_a[RSH][R'SNO]}{K_a + [H^+]}$$

pH=cte k'

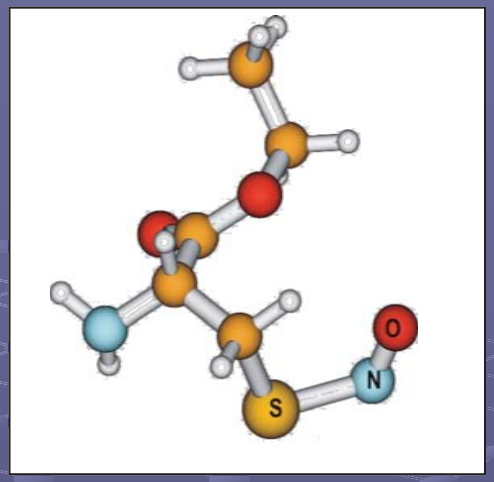
K. Wang et al. / Bioorg. Med. Chem. Lett. 11 (2001) 433±436

Singh, P. S. ; Wishnok, J. S. ; Keshive, M.; Deen, W. M.; Tannebaum, S. R. *Proc. Natl. Acad. Sci. U.S.A.*, **1996**, 93, 14428.

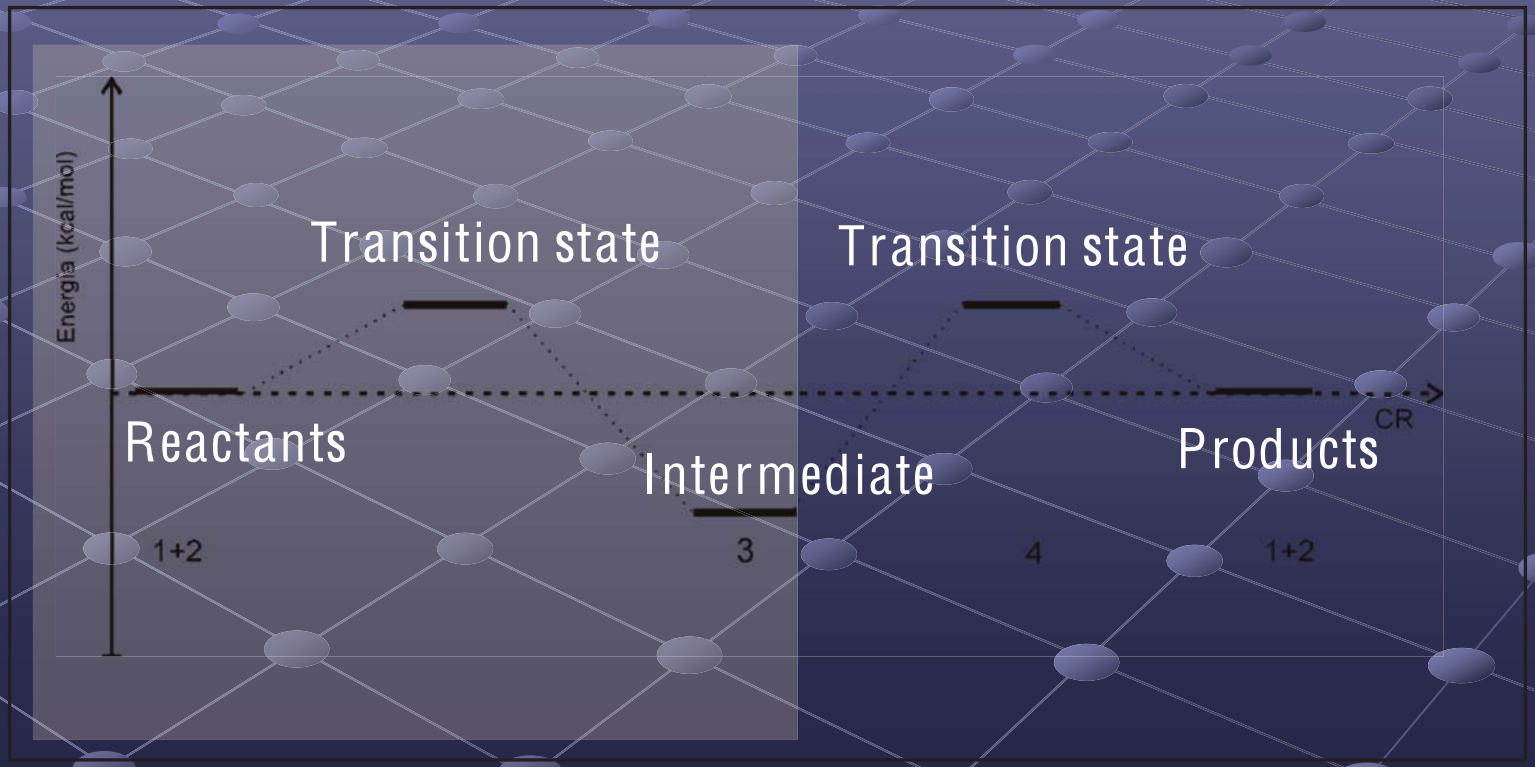
CHOICE OF THE MODEL



L-cysteine ethyl ester

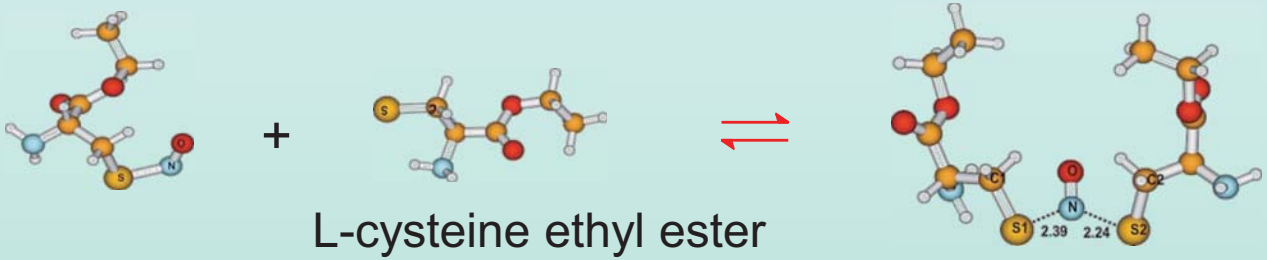


CR: S-N



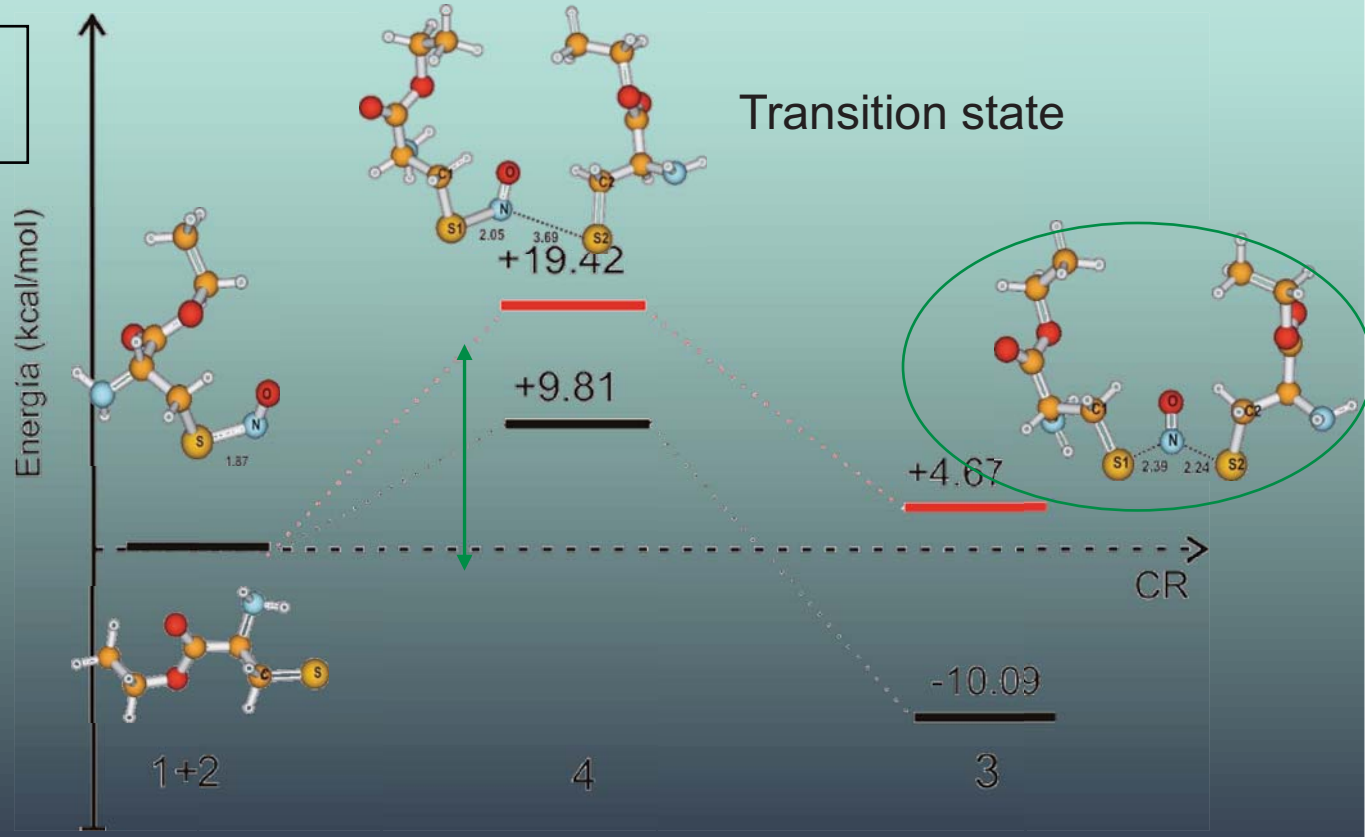
Clancy, R.; Cederbaum, A.; Stoyanovsky, D. *J. Med. Chem.*, **2001**, 44, 2035-2038.

S-nitroso L- cysteine ethyl ester



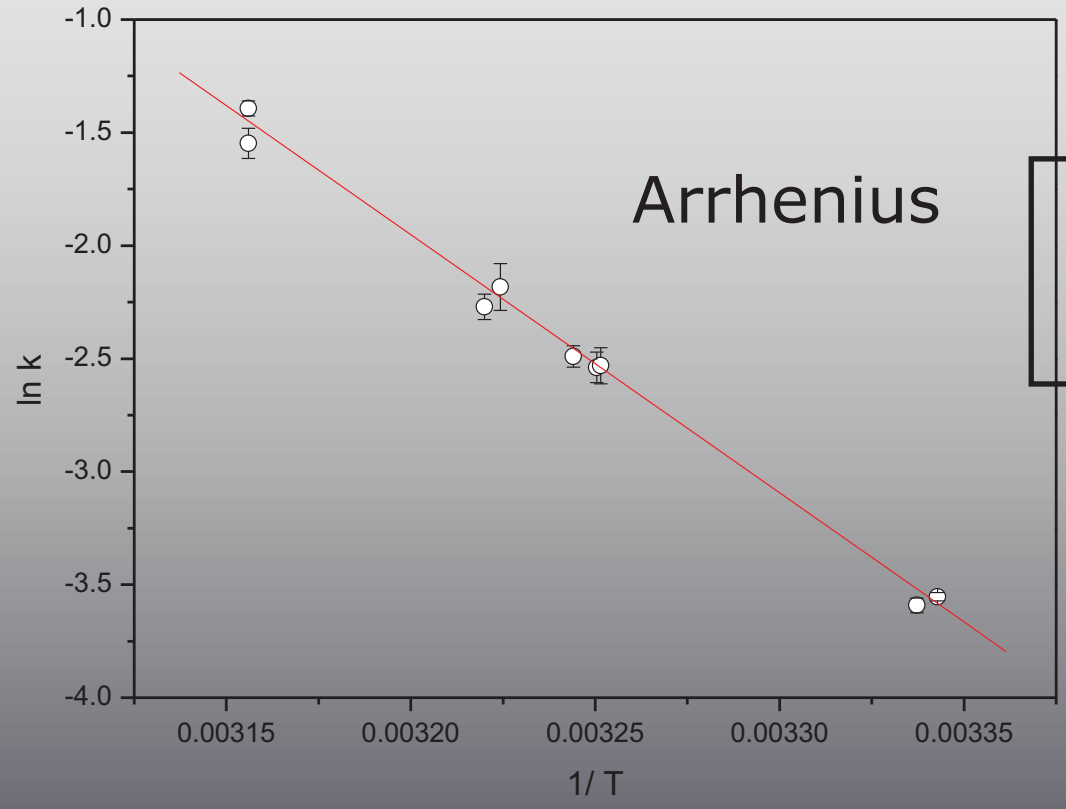
B3LYP/6-311+G*

Intermediate



Perissinotti, L. L.; Turjanski, A. G.; Estrin, D. A.; Doctorovich, F. *J. Am. Chem. Soc.;* **2005**; *127*(2); 486-487

Experimental corroboration: reaction performed in methanol (in water there were lots of side products)



E_a= 22.7 ± 0.2 kcal/mol
R=0.99822

19.42 kcal/mol

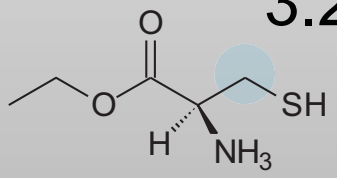
Reaction rate constants measured at different temperatures: From Arrhenius law $k=Aexp(-E_a/RT)$ E_a can be extracted

Perissinotti, L. L.; Turjanski, A. G.; Estrin, D. A.; Doctorovich, F. *J. Am. Chem. Soc.;* **2005**; *127*(2); 486-487

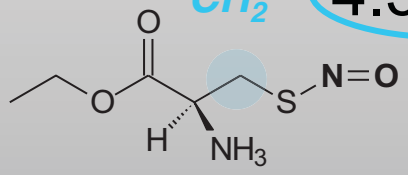
In the search of the intermediate…….

RS$^-$ + **R**´SNO $\rightleftharpoons$ [**R**S(NO)S**R**´]$^-$

3.2 ppm

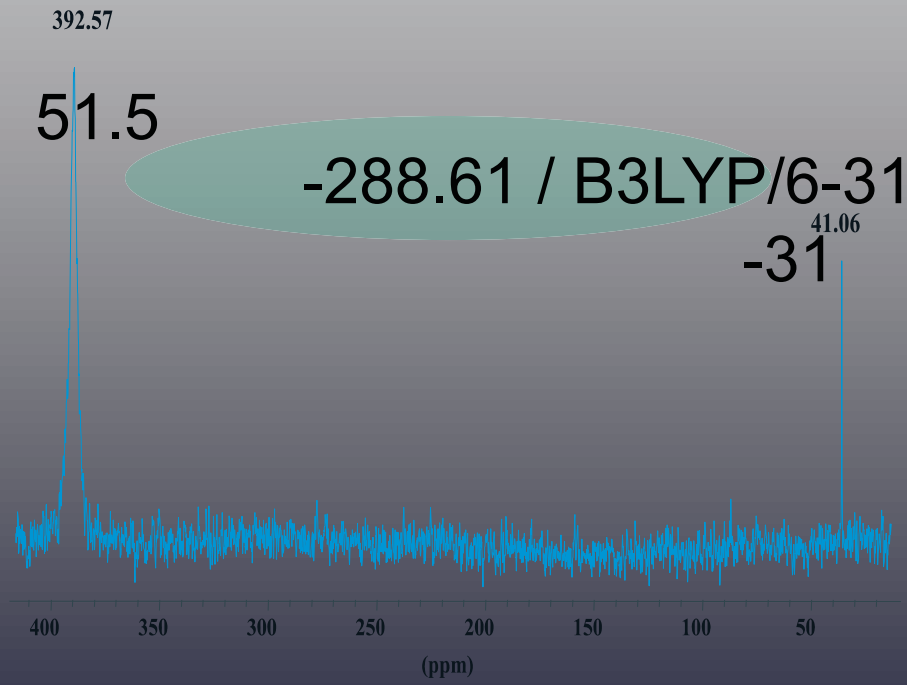


CH_2 4.3 ppm



^{1}H, ^{13}C RMN

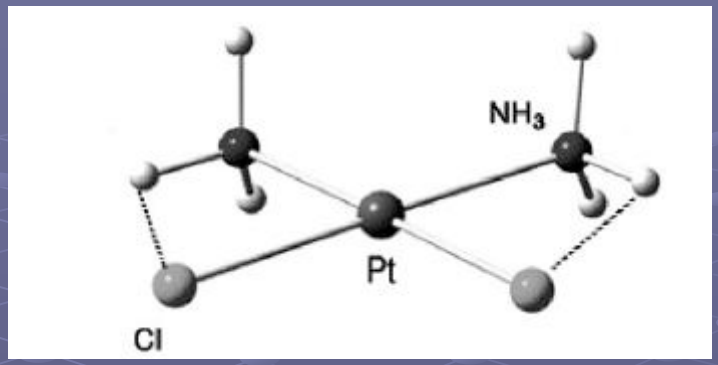
^{15}N RMN



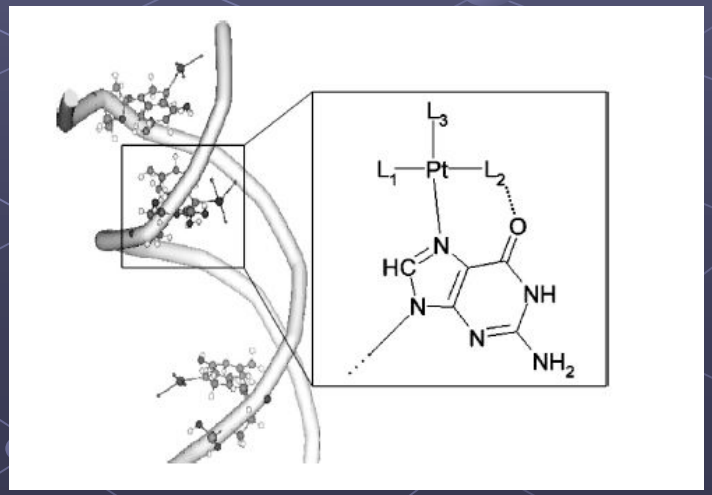
-288.61 / B3LYP/6-311+G*

Perissinotti, L. L.; Turjanski, A. G.; Estrin, D. A.; Doctorovich, F. *J. Am. Chem. Soc.;* **2005**; *127*(2); 486-487

Cisplatine: anticancer drug



It undergoes aquation, and then interacts with nucleic acids



Cisplatine: anticancer drug

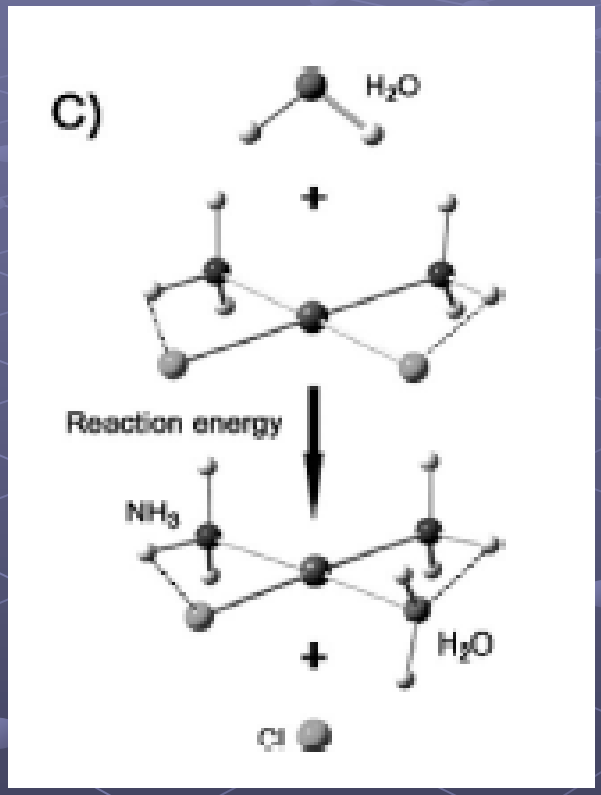


Table 7. Calculated Reaction Energy (kcal/mol) for the First Aquation Process of Cisplatin at the Different Levels of Theory over Optimized Isolated Species in the Gas Phase

	ECP/basis set employed		
level of theory	LANL2/ LANL2DZ	SBK/ CEP-31G	Troullier–Martins/Dζ
PBE1PBE/ PBE_{SIESTA}	115 (−5)	116 (−4)	142 (22)[a] 127 (7)[b]
mPW1PW91	115 (−5)	116 (−4)	
B3LYP	114 (−6)	115 (−5)	
B3PW91	115 (−5)	116 (−4)	
B3P86	119[c] (−1)	116 (−4)	
MP2(FC)	116[c] (−4)	116 (−4)	
mp2(full)	113 (−7)	115 (−5)	
HF	108 (−12)	109 (−11)	
G3-type strategy[d]	120		

Cisplatine: simpler case, results are much more robust

Same predictions for different methods

Dans et al, J. Chem. Theory and Computation, 2008, 4, 740

Model calculation: Cisplatin 9 methyl guanine adduct

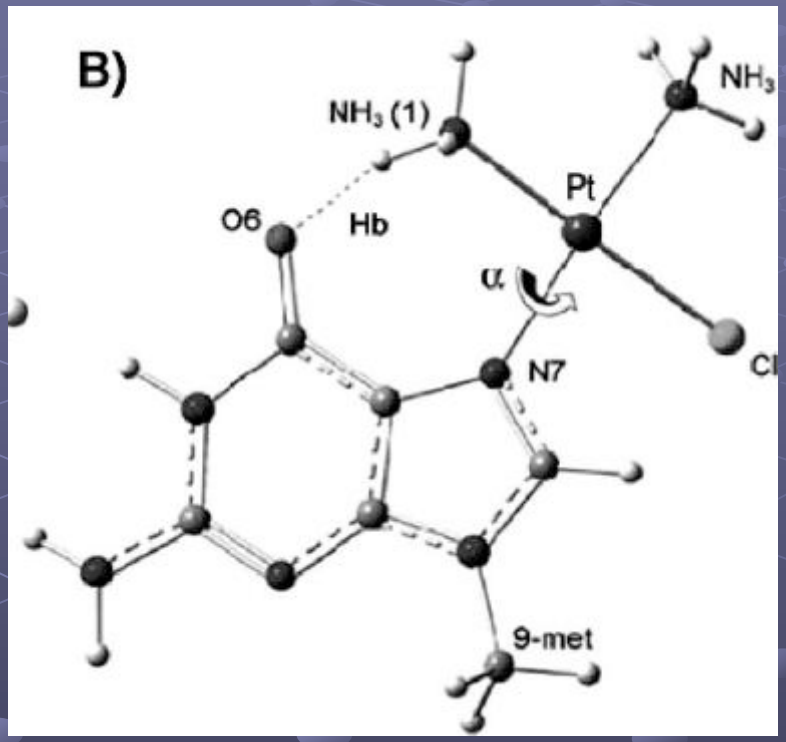


Table 6. *cis*-$[Pt(NH_3)_2Cl(9\text{-met-guanine})]^+$ Structure—Distances in angstroms; Angles in deg—Calculated with SIESTA at the PBE_{SIESTA} Level of Theory with Troullier–Martins Pseudopotentials, Deviations, OMPBDs and OMPGDs, and Reference Experimental Data

	Troullier-Martins/Dζ		
	PBE_{SIESTA}	exp (mean)	exp (range)
Pt–X[a]	2.329		
Pt–N(2)	2.066(−0.055)[b]	2.121	1.999–2.230
Pt–N(1)	2.078(0.023)	2.055	1.814–2.247
Pt–N7	2.031(−0.206)	2.237	2.164–2.315
N(1)···O(Hb)	1.745		
X[a]–Pt–N(2)	85.2(−14.4)	99.6	95.5–104.7
X[a]–Pt–N7	90.2(−5.9)	96.1	79.0–105.6
N(2)–Pt–N(1)	93.6(11.9)	81.7	78.7–85.1
N(1)–Pt–N7	91.0(10.4)	80.6	69.7–102.0
N(1)···O(Hb)	163		
α	37 (11)	26	13–44

Structural predictions at the DFT level are usually good

Can we believe to the results?

Hard to obtain reliable and robust predictions, specially regarding energetics

It is crucial to analyze critically the obtained results. Typically, the results do not allow for quantitative answers

A good practice is always compare similar systems and trying to understand trends instead of doing absolute predictions