

High Level Quantum Chemistry in Hungary

Péter G. Szalay
Eötvös Loránd University
Institute of Chemistry
H-1518 Budapest, P.O.Box 32, Hungary
szalay@chem.elte.hu



Importance of the theoretical information in chemistry

- **verify experiment**
 - justification for experimental findings
- **supply data not available from experiment**
 - low resolution
 - overlapping bands in spectrum
 - molecules with short lifetimes
 - complicated reaction mechanism
 - ...
- **prediction**



What are we doing in ab initio quantum chemistry?

We try to solve the electronic Schrödinger equation as accurate as possible:

$$\hat{H}\Psi = E\Psi$$

with

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_{i,A} \frac{Z_A}{r_{iA}} + \sum_{ij} \frac{1}{r_{ij}} + \sum_{AB} \frac{Z_A Z_B}{r_{AB}}$$

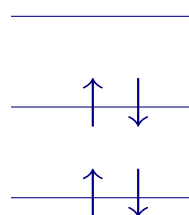
where A, B and i, j are nuclear and electronic indices, respectively. Z stands for the nuclear charge, r for the distance of the particles.

Approximations only to Ψ !!!!



THE HARTREE–FOCK METHOD

Wave function: a Slater-determinant corresponding to a configuration:



Wave function: antisymmetrized product of molecular orbitals (ϕ_i):

$$\Psi_{HF} = \hat{A}(\phi_1(1) \phi_2(2) \phi_3(3) \dots \phi_n(n))$$

Molecular orbitals (ϕ_i) obtained from Hartree-Fock equation:

$$\hat{f} \phi_i = \epsilon_i \phi_i$$



THE HARTREE-FOCK METHOD

Molecular orbitals (ϕ_i) obtained from Hartree-Fock equation:

$$\hat{f} \phi_i = \epsilon_i \phi_i$$

Using a basis for the orbitals:

$$\phi_i = \sum_a c_{ia} \chi_a \quad \chi_a : \text{basis functions}$$

we arrive at the Hartree-Fock-Roothaan equations:

$$\underline{F} \underline{c} = \epsilon \underline{c}$$



THE HARTREE–FOCK METHOD

Computational tasks:

- Calculation of two-electron integrals:

$$\langle pq|rs\rangle = \int \int \chi_p(1)\chi_q(2)\frac{1}{r_{12}}\chi_r(1)\chi_s(2)d\tau_1d\tau_2$$

with χ some kind of Gauss-functions

- Building up the Fock-matrix: matrix multiplication of two-electron integrals and density
- Diagonalization of the Fock-matrix



THE HARTREE–FOCK METHOD

Typical dimensions:

- number of basis functions: several hundred to thousands
- number of integrals: scales with the fourth power of the no. of basis functions
- matrix to diagonalize: dimension of the no. of basis functions



THE HARTREE–FOCK METHOD

Advantages:

- Independent particle approximation → concept of orbitals
- not very expensive

Problems:

- do not describe the proper interaction of electrons
→ lack of “**electron correlation**”
- accuracy is limited



INCLUSION OF THE ELECTRON CORRELATION

- **Perturbation Theory (PT)** - use HF as start
- **Configuration Interaction (CI)** - expand the wave function on several determinants
- **Coupled Cluster (CC)** - exponential expansion of the wave function
- **Density Functional Theory (DFT)** - technically similar to HF



PERTURBATION THEORY

Usual Rayleigh-Schrödinger Perturbation Theory with

$$\hat{H}_0 = \sum_i \hat{f}(i)$$

i.e. sum of the one-electron Fock-operators (Møller-Plesset partitioning)

1st order: Hartree-Fock method

2nd order: MP2 or MBPT(2) method

3rd order: MP3 or MBPT(3) method

etc.



PERTURBATION THEORY

MP2: cheap way to include electron correlation

MP3: usually not any better than MP2

MP4: often very good but expensive

Main problems:

- **series may not converge**
- **gets very expensive**



PERTURBATION THEORY

Computational tasks:

- Calculation of two-electron integrals:

$$\langle pq|rs\rangle = \int \int \chi_p(1)\chi_q(2)\frac{1}{r_{12}}\chi_r(1)\chi_s(2)d\tau_1d\tau_2 \quad (1)$$

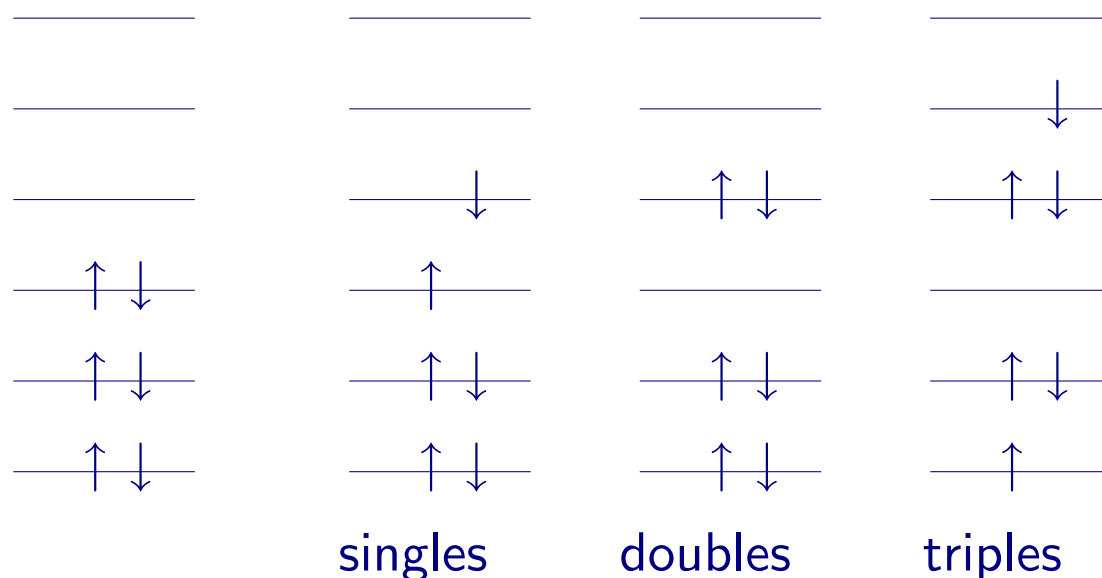
with χ some kind of Gauss-functions

- Transformation of the integrals
- Matrix multiplication of the pieces of the integral array



THE CONFIGURATION INTERACTION (CI) METHOD

Wave function: linear combination of Slater-determinants



$$\Psi_{CI} = c_0 \Psi_{HF} + \sum_{ia} c_i^a \Phi_i^a + \sum_{ijab} c_{ij}^{ab} \Phi_{ij}^{ab} + \sum_{ijkabc} c_{ijk}^{abc} \Phi_{ijk}^{abc} + \dots$$



THE CONFIGURATION INTERACTION (CI) METHOD

$$\Psi_{CI} = c_0 \Psi_{HF} + \sum_{ia} c_i^a \Phi_i^a + \sum_{ijab} c_{ij}^{ab} \Phi_{ij}^{ab} + \sum_{ijkabc} c_{ijk}^{abc} \Phi_{ijk}^{abc} + \dots$$

Solving for c 's variationally:

$$H\underline{c} = E\underline{c}$$

with

$$H_{ij} = \langle \Phi_i | \hat{H} | \Phi_j \rangle$$



THE CONFIGURATION INTERACTION (CI) METHOD

Computational tasks:

- All of the above: handling very large number of integrals
- Obtaining eigenvalues of very large matrices of dimension of hundred millions



THE CONFIGURATION INTERACTION (CI) METHOD

Approximations: Truncate after certain excitation level:

- singles and doubles: CISD
- + triples: CISDT
- + quadruples: CISTDTQ
-
- all: *Full CI*



THE CONFIGURATION INTERACTION (CI) METHOD

Approximations: Truncate after certain excitation level:

- singles and doubles: CISD
- + triples: CISDT
- + quadruples: CISTDTQ
-
- all: *Full CI*

BIGGEST PROBLEM: lack of size consistency

i.e. does not scale correctly with the size of the system



THE COUPLED-CLUSTER METHOD

Wave function:

$$\Psi_{CC} = e^T \Phi_0$$
$$T = T_1 + T_2 + \dots$$

with T_n being an excitation operator. Truncated versions:

- **CCSD** ($T = T_1 + T_2$)
- **CCSD(T)** ($T = T_1 + T_2 + \text{approximate } T_3$)
- **CCSDT** ($T = T_1 + T_2 + T_3$)
- **CCSDTQ** ($T = T_1 + T_2 + T_3 + T_4$)
- ...



THE COUPLED-CLUSTER METHOD

Computational tasks:

- All of the above: handling very large number of integrals
- Solving non-linear system of equations for the parameters
- Multiplication of very large matrices of wave function parameters and integrals



THE COUPLED-CLUSTER METHOD

Computational tasks:

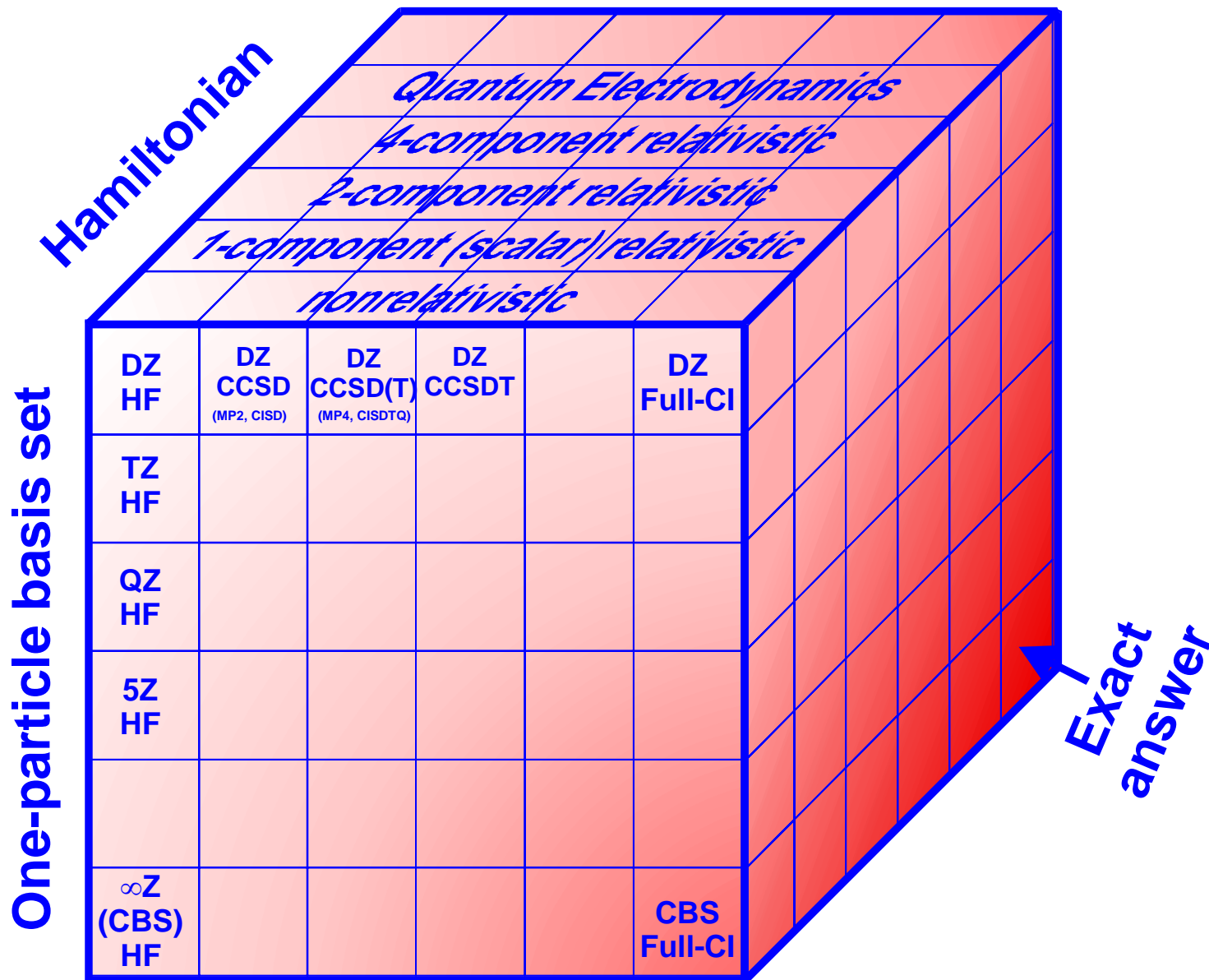
- All of the above: handling very large number of integrals
- Solving non-linear system of equations for the parameters
- Multiplication of very large matrices of wave function parameters and integrals

The big challenge:

The number of parameters grows very rapidly with the excitation level included!!



Electron correlation treatment



Computational Quantum Chemistry in Hungary

Founders (60's and 70's):

- Rezső Gáspár at Debrecen: precursor of DFT theory
- Ede Kapuy at Technical University: perturbation theory
- Péter Pulay at Eötvös University: gradient theory



Computational Quantum Chemistry in Hungary

Groups now:

- Eötvös University, Institute of Chemistry
 - method development
 - high accuracy calculations
 - structural chemistry, theoretical spectroscopy
 - structural chemical biology
- Technical University in Budapest
 - applied quantum chemistry
 - method development, high accuracy methods
- Chemical Research Center of the Hungarian Academy of Sciences
 - method development
 - applied quantum chemistry (catalysis)
 - dynamics



Computational Quantum Chemistry in Hungary

Groups now:

- University of Debrecen, Institute of Physics
 - density functional theory
- University of Szeged
 - applied quantum chemistry



Computational Quantum Chemistry in Hungary

Some computing demanding projects:

- Eötvös University, Institute of Chemistry
 - P.G. Szalay
 - G. Fogarasi
 - P.R. Surján
 - A.G. Császár
 - L. Turi
 - A. Perczel
- Chemical Research Center of the Hungarian Academy of Sciences
 - I. Pápai
- Technical University in Budapest
 - M. Kállay



Systematic determination of heat of formation

Péter G. Szalay
Eötvös University, Institute of Chemistry

Importance of knowing very accurate heat of formation for small molecules:

- combustion: modeling processes in flames and plasma
- atmospheric chemistry: modeling processes leading to ozone loss etc.
- astrophysics: molecules of interstellar medium



Systematic determination of heat of formation

High Level Extrapolated Ab initio Thermochemistry – HEAT

Geometry: all electron CCSD(T)/cc-pVQZ level

Total energy is calculated:

$$E_{HEAT} = E_{SCF} + \Delta E_{CCSD} + \Delta E_{CCSD(T)} + \Delta E_{CCSDT} + \Delta E_{CCSDTQ} \\ + E_{ZPE} + E_{DBOC} + E_{SPO} + E_{REL}$$

with:

E_{SCF} : Hartree-Fock energy, aug-cc-pCVXZ basis TQ5 extrapolated

ΔE_{CCSD} : all electron CCSD correlation energy with aug-cc-pCVXZ basis Q5 extrapolated



Systematic determination of heat of formation

High Level Extrapolated Ab initio Thermochemistry – HEAT

Geometry: all electron CCSD(T)/cc-pVQZ level

Total energy is calculated:

$$E_{HEAT} = E_{SCF} + \Delta E_{CCSD} + \Delta E_{CCSD(T)} + \Delta E_{CCSDT} + \Delta E_{CCSDTQ} \\ + E_{ZPE} + E_{DBOC} + E_{SPO} + E_{REL}$$

with:

$\Delta E_{CCSD(T)}$: all electron (T) energy correction with aug-cc-pCVXZ basis
Q5 extrapolated

ΔE_{CCSDT} : frozen core energy difference of CCSDT and CCSD(T) with
cc-pVXZ basis TQ extrapolated

ΔE_{CCSDTQ} : frozen core energy difference of CCSDT and CCSDTQ with
cc-pVDZ basis



Systematic determination of heat of formation

High Level Extrapolated Ab initio Thermochemistry – HEAT

Geometry: all electron CCSD(T)/cc-pVQZ level

Total energy is calculated:

$$E_{HEAT} = E_{SCF} + \Delta E_{CCSD} + \Delta E_{CCSD(T)} + \Delta E_{CCSDT} + \Delta E_{CCSDTQ} \\ + E_{ZPE} + E_{DBOC} + E_{SPO} + E_{REL}$$

with:

E_{ZPE} : zero point energy from cubic force field at CCSD(T)/cc-pVQZ level.

E_{DBOC} : diagonal Born-Oppenheimer correction at HF/cc-pVQZ level.

E_{REL} : scalar relativistic energy correction at CCSD(T)/aug-cc-pCVQZ level.

E_{SPO} : spin-orbit correction at the MR-CI/cc-pVDZ level.



Heat of formation of selected radicals (kJ mol⁻¹)

Species	HEAT	literature	Deviation
C	711.32	711.194± 0.45	-0.13
CCH	563.64	563.6±2.4	0.0
CH2	391.39	390.7 ± 1.6	-0.7
CH	592.85	592.5 ± 0.6	-0.4
CH3	150.08	150.0 ± 0.3	-0.1
F	77.22	77.28 ± 0.30	0.06
H2O	-239.11	-238.927± 0.040	0.18
H	216.23	216.035 ±0.006	-0.20
HO2	15.05	15.6 ± 1.2	0.5
N	470.54	470.82 ± 0.40	0.28
NH3	-38.30	-38.946± 0.35	-0.65
O	246.69	246.80 ± 0.10	0.11
OF	110.36	108. ± 10	-2.36
OH	36.78	37.03 ± 0.3	0.25
RMS		0.45	0.19



Systematic determination of heat of formation

High Level Extrapolated Ab initio Thermochemistry – HEAT

- Accuracy: $0.2\text{-}0.4 \text{ kJ mol}^{-1}$ ($\sim 0.1 \text{ kcal mol}^{-1}$)
- Quadruple excitations are needed to get this accuracy
- Anharmonic force field required of ZPE
- Very expensive method

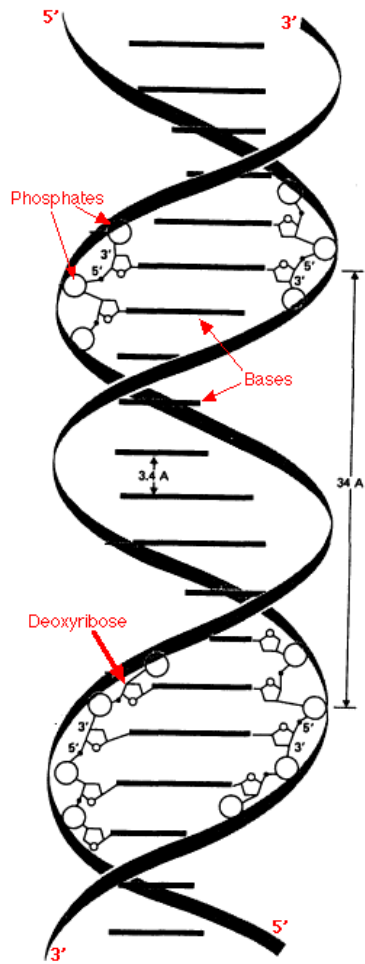


Water-mediated tautomerization of
cytosine
observed *in silicio*

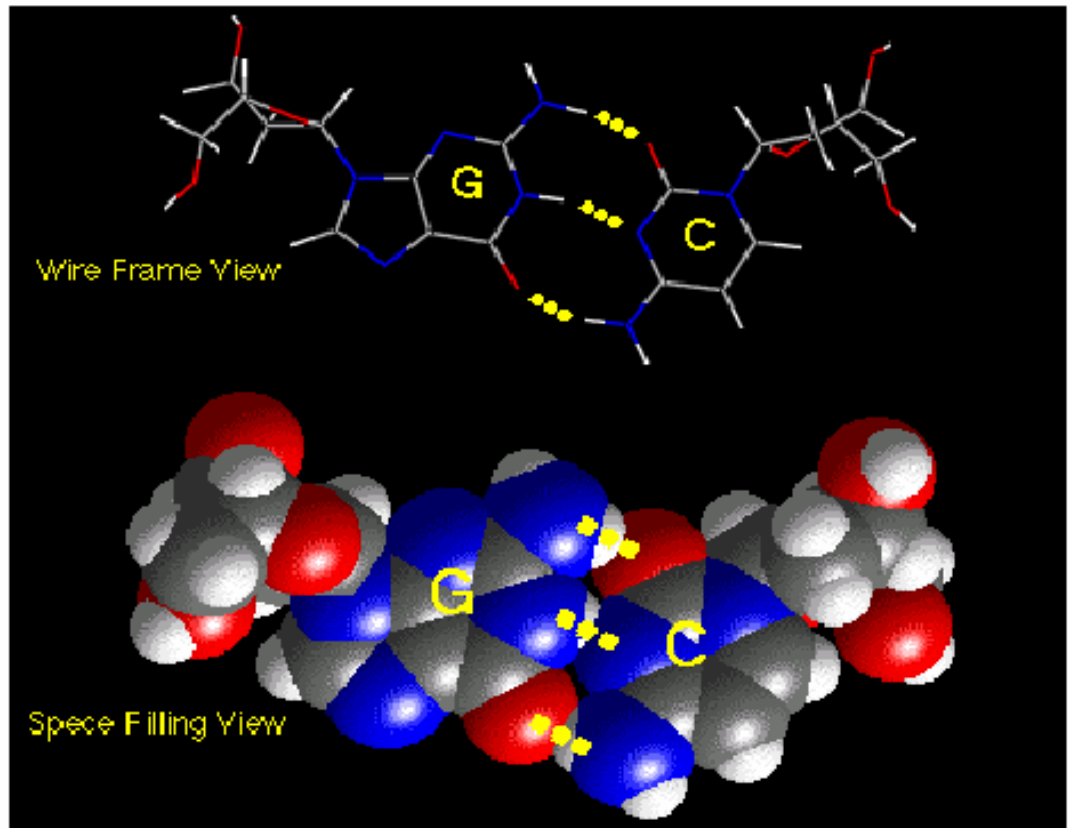
Geza Fogarasi

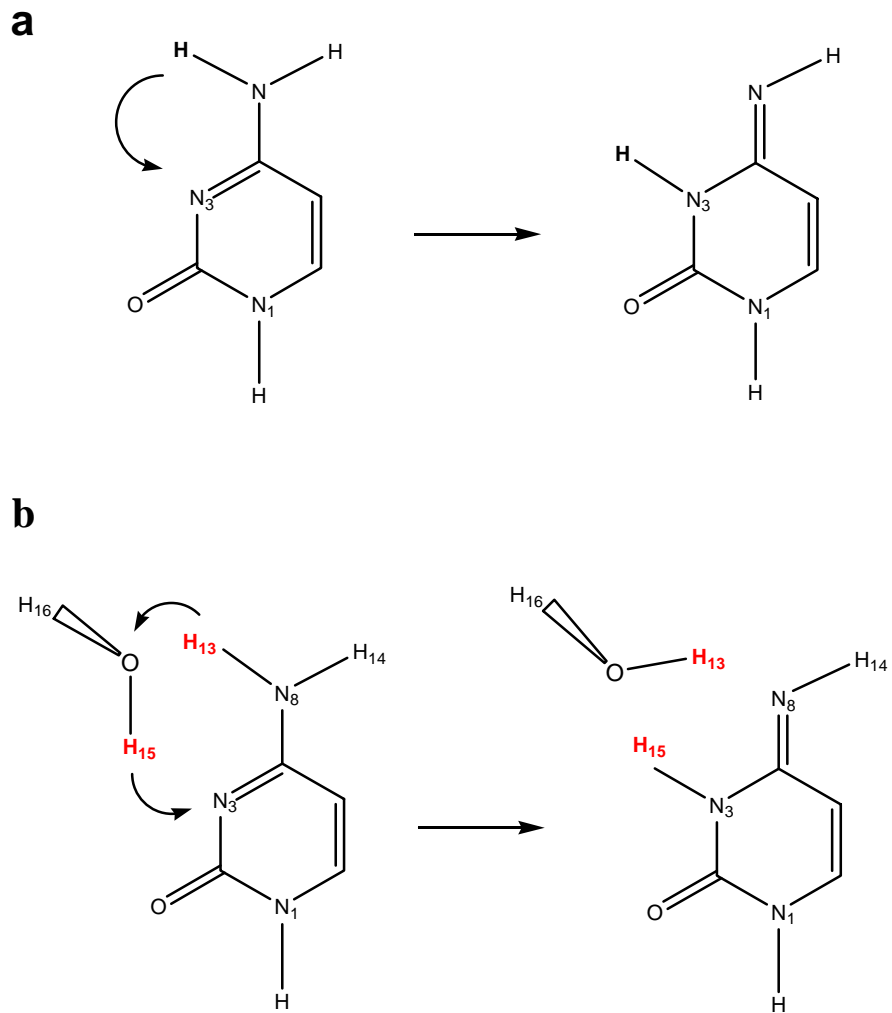
Department of Chemistry, Eötvös University, H-1518 Pf. 32, Budapest, Hungary

Significance of the project: cytosine is one of the four bases in **DNS**;
tautomers would form different H-bonds \Rightarrow **MUTATION**



Example of dG-dC base pair as found within DNA double helix





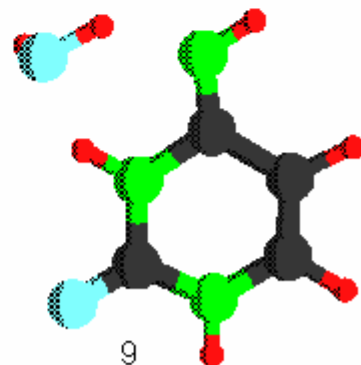
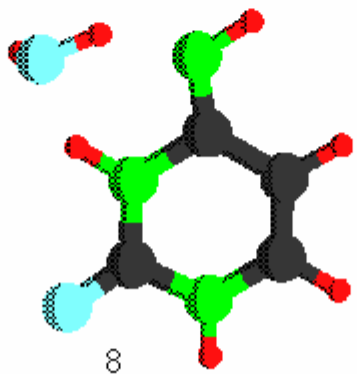
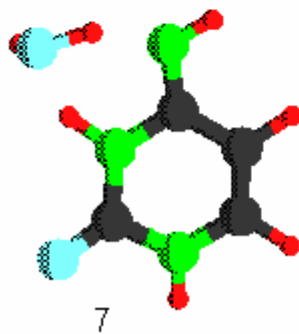
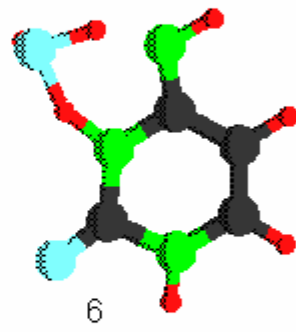
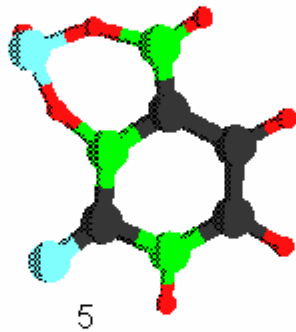
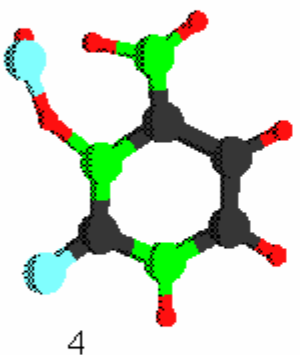
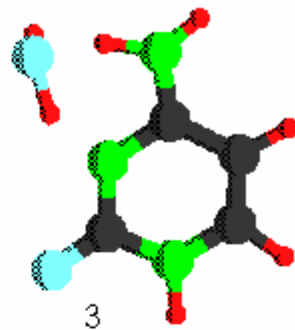
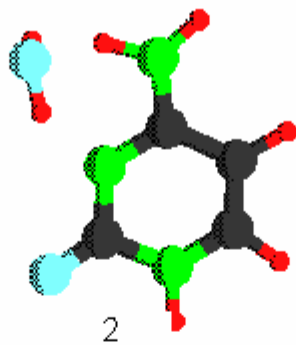
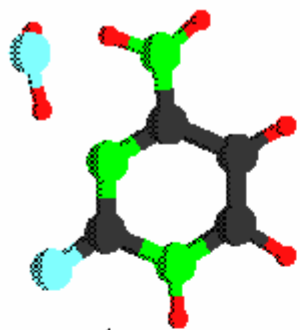
Scheme 1 Tautomerization without water and with water mediation

Computations: ab initio dynamics at B3LYP/3-21G ; ~ 400 trajectories, 3000-5000 steps each, 1 fs resolution.

About **1.5 million** complete wave-function calculations, including derivatives.

(1 calcn ~ 1' on one dualproc. PC)

Resource at present: 10 dualprocessor PCs;
several months computer time



Computation-demanding projects in the group of P.R. Surján

Eötvös University, Budapest

- Linear scaling ab initio calculations

Phys. Rev. Lett. 95 13002 (2005)

- Solution of tight-binding models for nano-structures

Phys. Rev. A. 68 062503 (2003)

(Laplace-transformed PT denominators)

- Multiconfiguration perturbation theory

J.Chem.Phys. 119 1922-1928 (2003)

J.Chem.Phys. 122 114104 (2005)

Linear scaling calculations

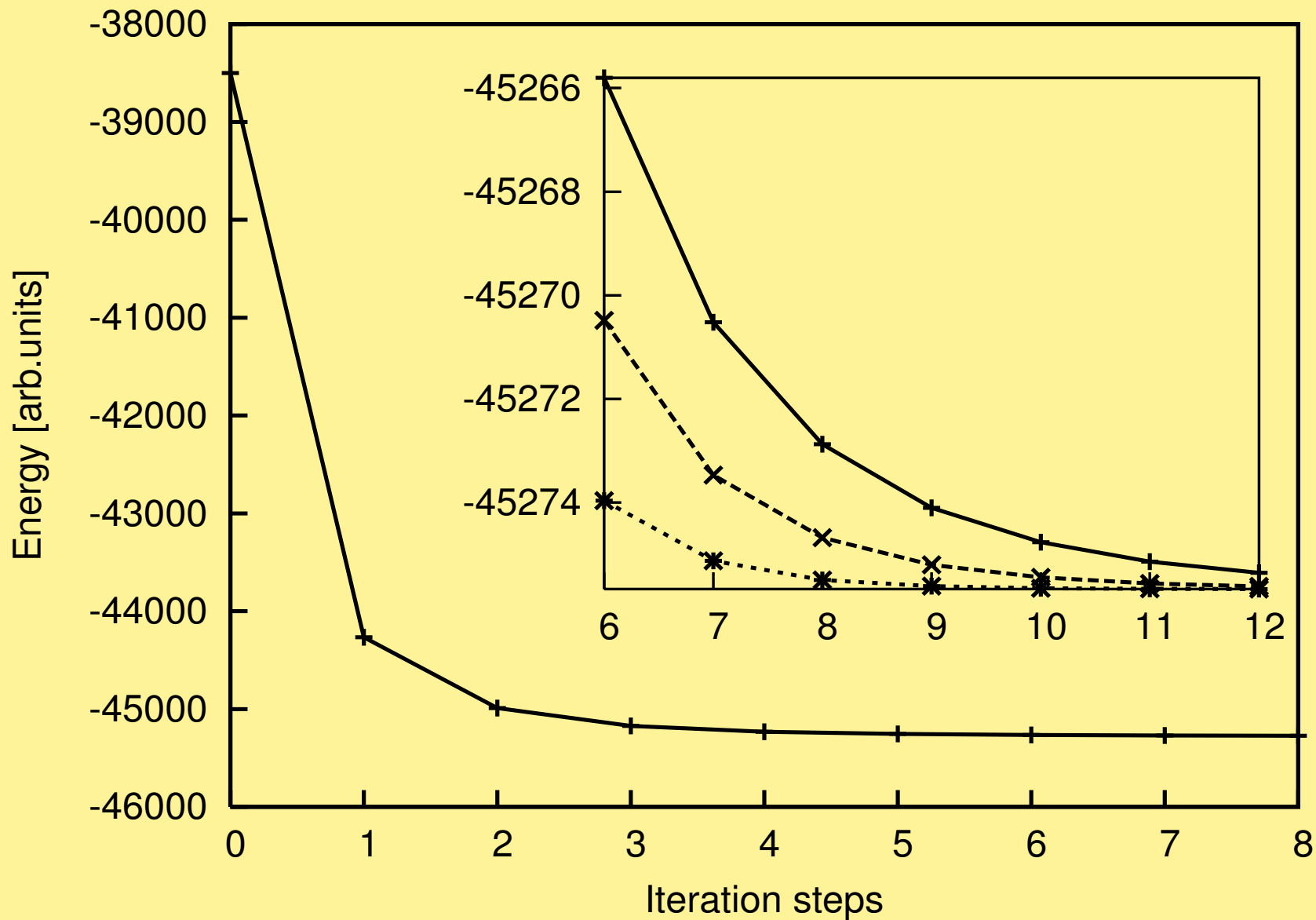
An idempotency-conserving iteration scheme for the density matrix:

$$P' = P + \eta (FP - PFP)$$

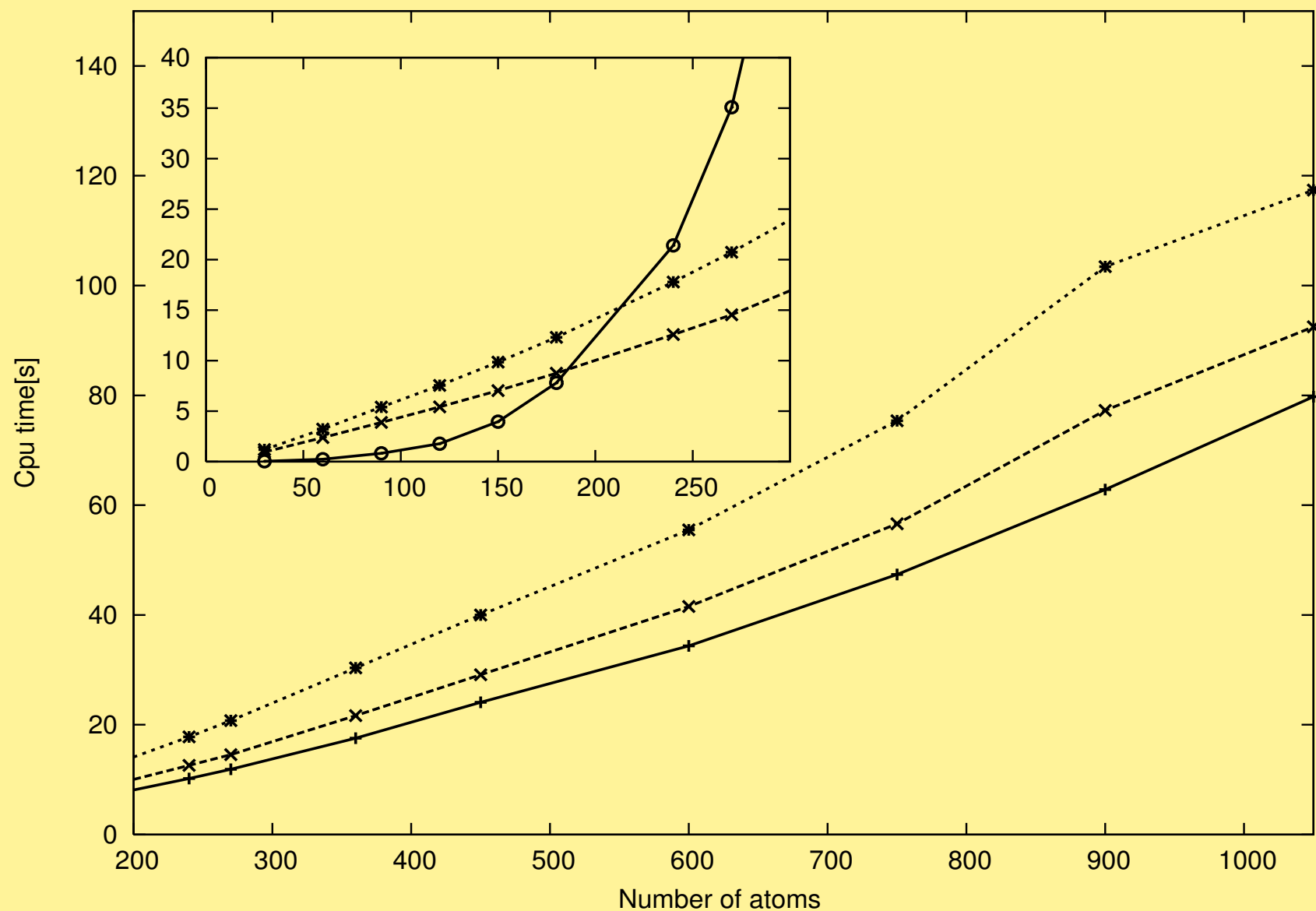
P : Density matrix

F : Fockian

Polyacetylene: tight-binding chain of 35 000 atoms - energy convergence



Chain of water molecules: CPU measure



(ab initio Hartree Fock)

Toward Complete Spectroscopy of Small Molecules

Attila Császár

Eötvös University, Institute of Chemistry

Electronic structure calculations



Potential energy (hyper)surface (**PES**)
[energies (and derivatives) over a grid]



Property (*e.g.*, dipole) surface
(**DMS**) over a grid

Nuclear motion calculations



All energy levels, wave functions,
and dipole transition moments



COMPLETE SPECTRA

Incremental buildup of the potential energy surface of water

Valence-only problem

nearly complete basis set (extrapolated) ICMRCI

Core-correlation correction

Relativistic corrections

one-electron mass-velocity and Darwin (MVD1)

two-electron Darwin (D2)

Gaunt and Breit corrections

quantum electrodynamics (QED)

Adiabatic and nonadiabatic Born-Oppenheimer corrections

O. L. Polyansky, A. G. Császár, J. Tennyson, P. Barletta, S. V. Shirin, N. F. Zobov, D. W. Schwenke, and P. J. Knowles, *Science* **299**, 539 (2003).

Stationary points on PESs

Equilibrium structures of water

	$r_e/\text{\AA}$	$\Theta_e/\text{degrees}$
Born-Oppenheimer	0.957 82	104.48 ₅
Adiabatic (H ₂ ¹⁶ O)	0.957 85	104.50 ₀
(D ₂ ¹⁶ O)	0.957 83	104.49 ₀
Spectroscopic	0.957 77	104.48

A. G. Császár, G. Czakó, T. Furtenbacher, J. Tennyson, V. Szalay, S. V. Shirin, N. F. Zobov, O. L. Polyansky, *J. Chem. Phys.* **122**, 214305 (2005).

$J = 1$ rotational term values for the ground vibrational state of water from the *ab initio* PESs

	H_2^{16}O		H_2^{18}O	
	Theory	Experiment	Theory	Experiment
1_{01}	23.795	23.7944	23.756	23.7549
1_{11}	37.138	37.1371	36.749	36.7486
1_{10}	42.372	42.3717	42.024	42.0234
4_{22}	315.799	315.779		

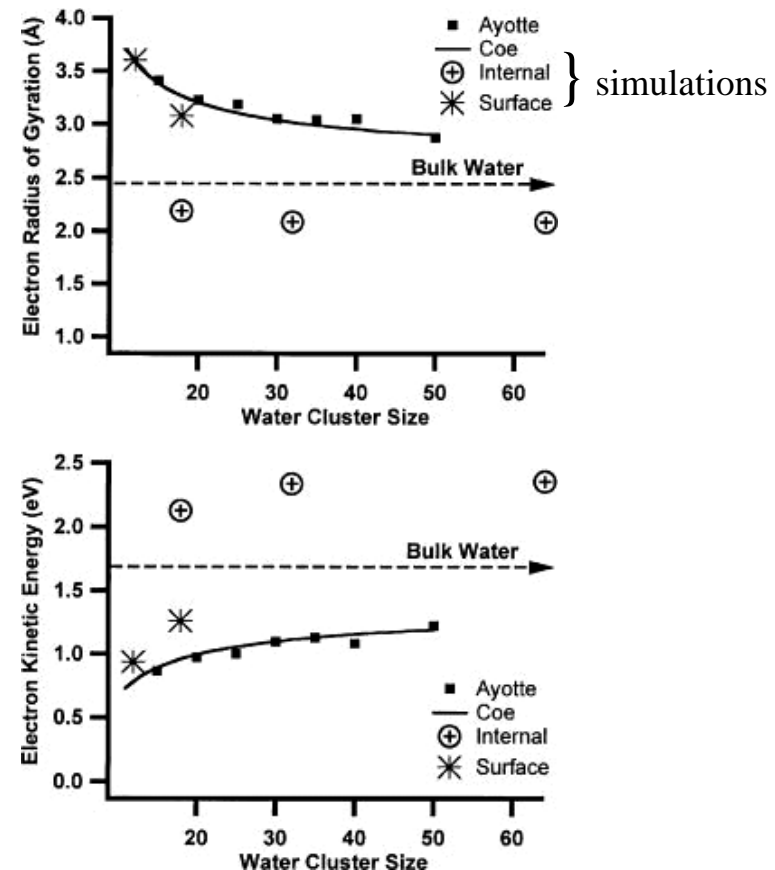
Quantum Molecular Dynamics Simulations of $(\text{H}_2\text{O})_n^-$ Water Cluster Anions

Dr. László Turi
Eötvös University, Institute of Chemistry

Question:

Where is the solvated electron?

- *on the surface*
- *inside the cluster*
(*internal states*)



Method of the calculation:

- Mixed quantum-classical simulation: **quantum e- + classical nuclei**
- New e- - water pseudopotential: Turi, Borgis *J. Chem. Phys.* 117, 6186 (2002)
 - electrostatics (**SPC water**)
 - local repulsion
 - local exchange
 - induced polarization of solvent

Computational details:

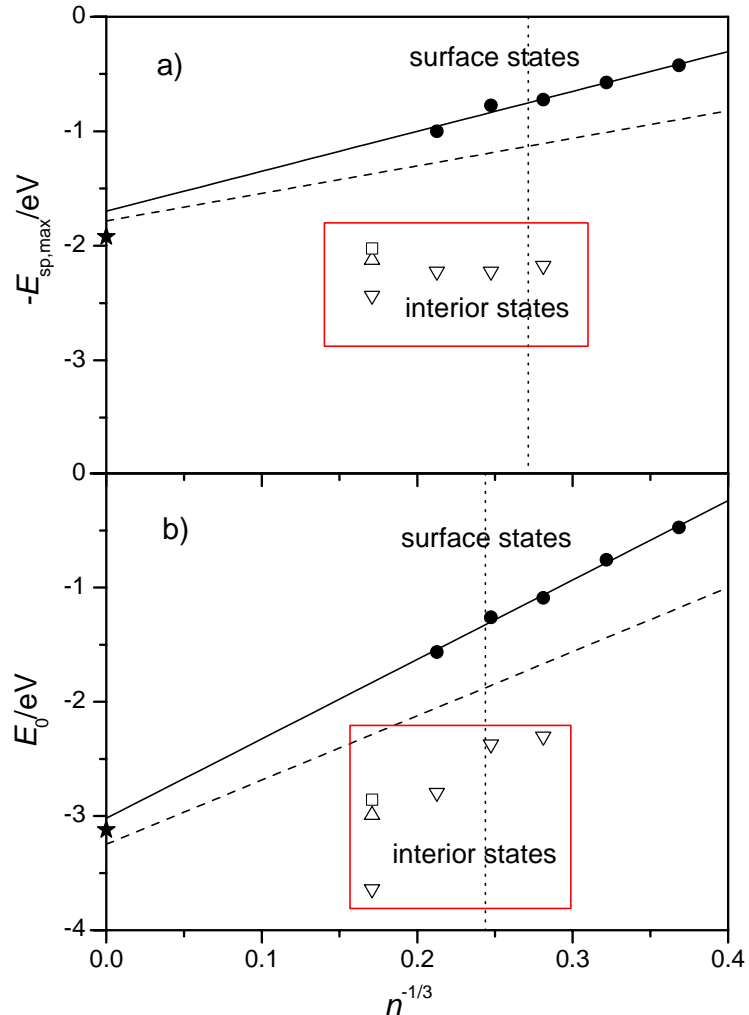
- Water cluster anions with n=20,30,45, 66, 104, 200
- Temperature: T=100K, 200K, 300K

Hardware:

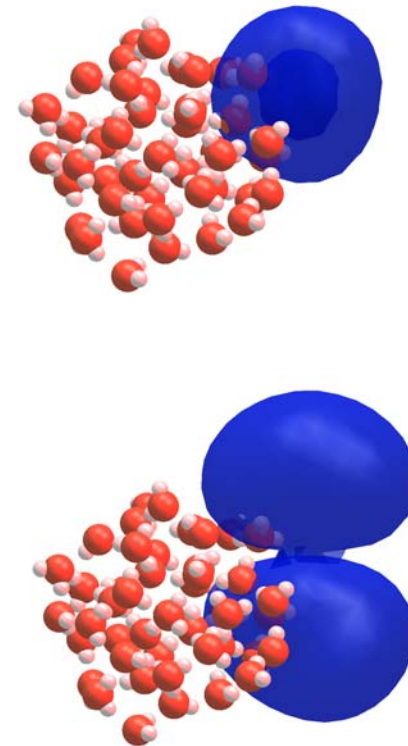
- INTEL Xeon 2.2 GHz Cluster of 24 processors
- CPU: approximately 1 month

Simulation results:

surface state extrapolation to the bulk reproduces experiment (VDE, spectrum)



Surface states of $(H_2O)_{45}^-$



Structure and stability of secondary structure elements of proteins



András Perczel

*Department of Chemistry
Eötvös Loránd University*

Budapest

2005 June 17

Why studying secondary structural properties?

To know more about

- the theoretical properties of structural building units,
- the structurally disordered (unstructured) regions
- the folding path and mechanism
of peptides and proteins

-To speed up

- genomics and proteomics,
- structure assignment of secondary structures,
- structure determination
based on CD-, IR-, VCD- and NMR-spectroscopy

Strategy

Study the relative stability (relative energy, ΔE , ΔG) of conformers as function of the length of the polypeptide and compare with experimental data

Methods

- computation (*ab initio*)

levels of theory used: RHF, DFT (B3LYP), MP2, CCSD(T)

basis sets applied: dzp, tzp, tz2p

e.g.: RHF/6-311++G(d,p)

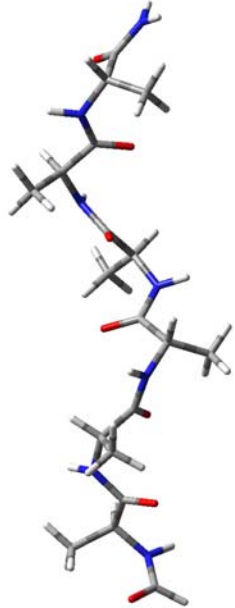
- data base analysis

X-ray structures of **1211** non-homologous proteins

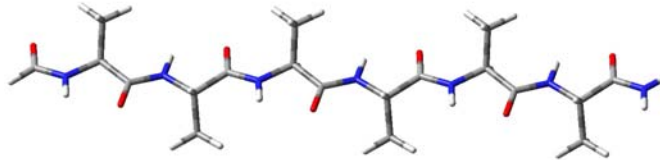
Perczel et al . J.Comp.Chem. 2003

Build up of *homo*-conformers from their conformational units

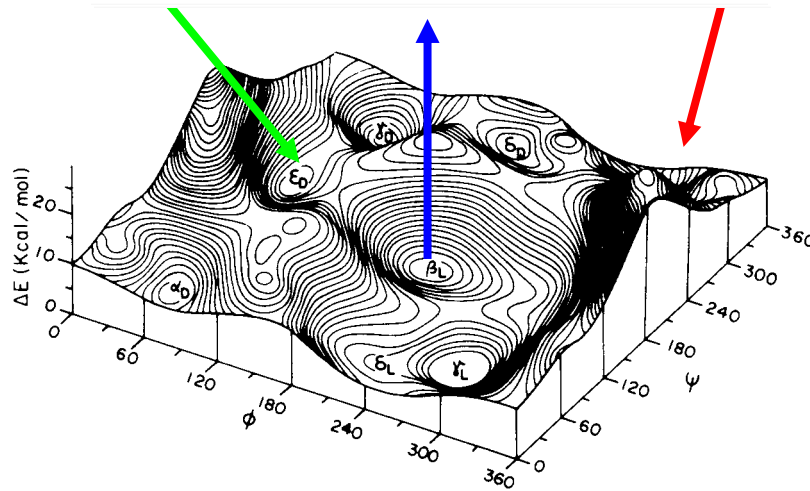
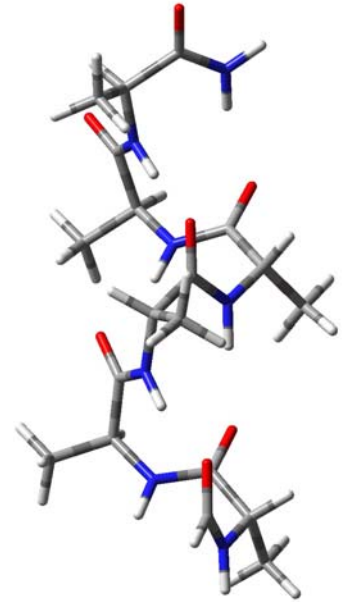
$(\epsilon_D)_n$



β -sheet



α -helix



Mechanistic studies on organocatalytic reactions

Imre Pápai, Chemical Research Center of the Hung. Acad. Sciences

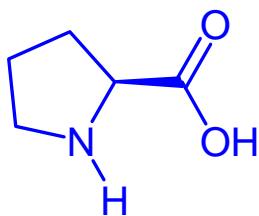
Organocatalysis: Activation without metals (with organic compounds)

Advantages (as compared to metal-based catalysis):

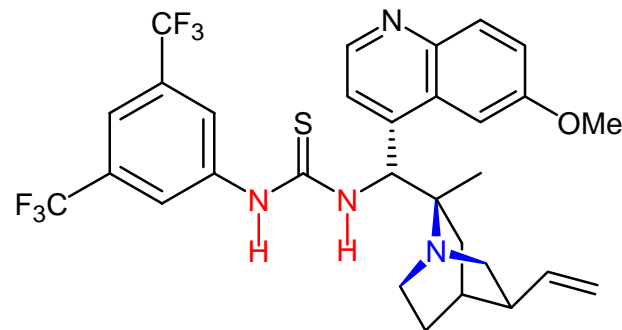
preparative (cheap, stable, easy-to-construct)
environmental (non-toxic, no product contamination)

Main goal: Design and prepare organocatalysts that promote enantioselective C-C bond formation

Examples:



L-proline

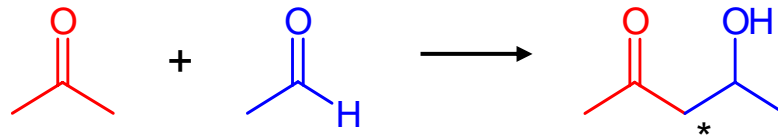


thiourea-based catalyst

Unknown reaction mechanism → need for theoretical studies

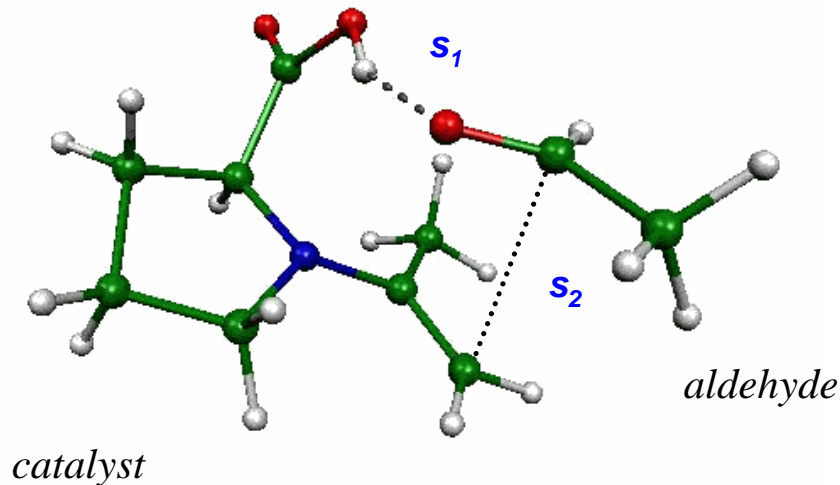
Asymmetric aldol reaction

Investigated reaction:



(asymmetric aldol reaction)

Applied method: Car-Parrinello MD simulation to describe the mechanism



Metadynamics with collective variables:

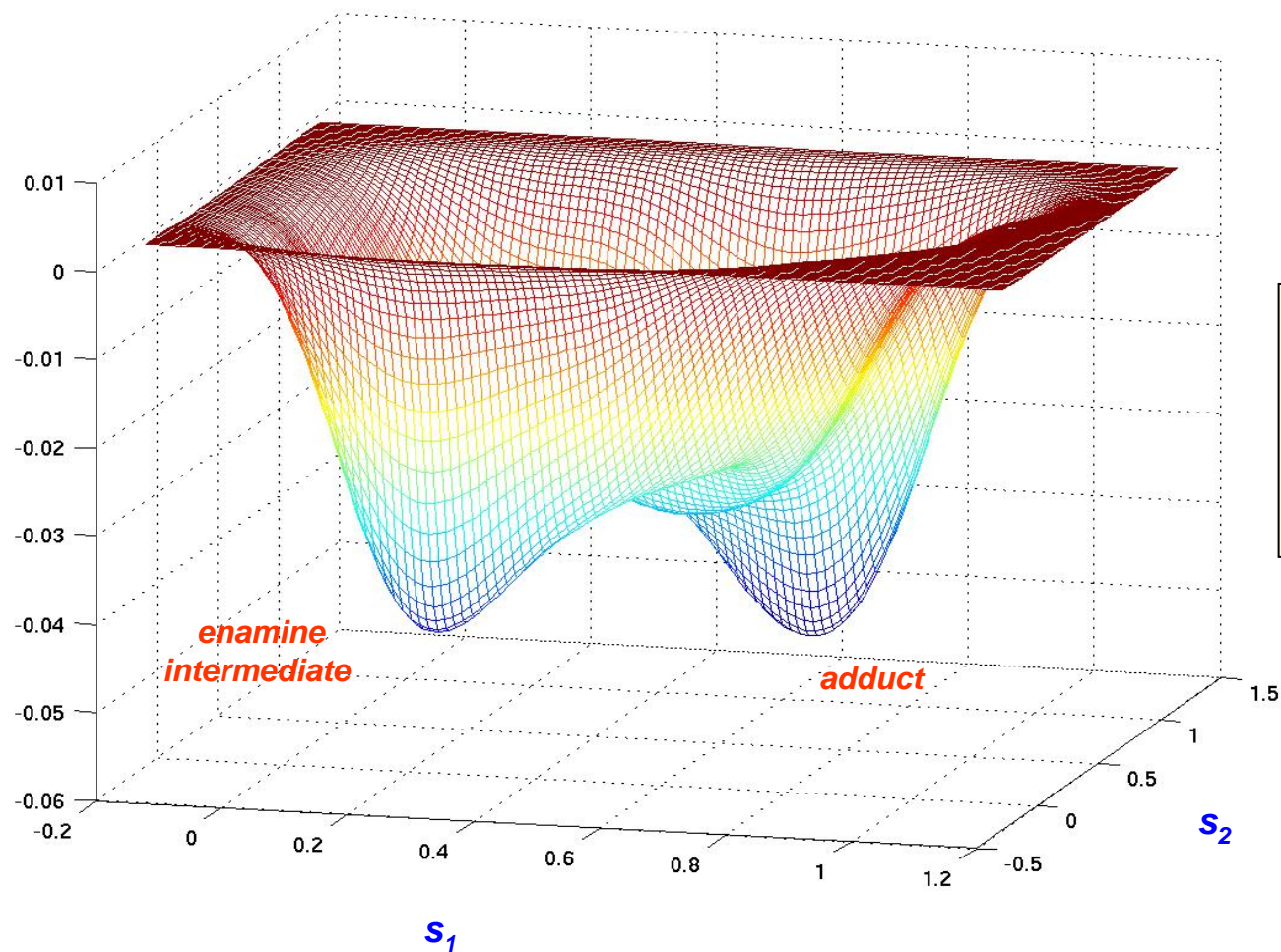
s_1 : H–O bond formation

s_2 : C–C bond formation



$F(s_1, s_2)$: free energy surface
is derived

Free energy surface derived from metadynamics

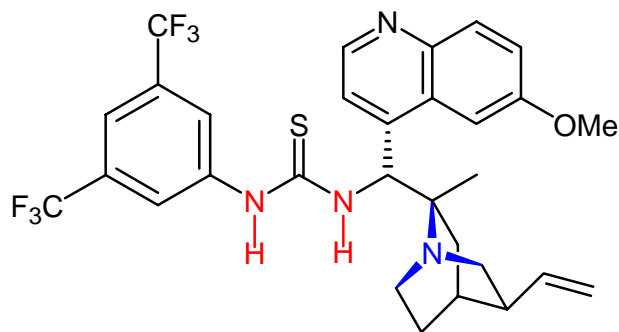


Computations:

- CPMD 3.9 (parallel)
- SGI Origin 2000
- 1 week job on 8 CPU

Need for HPC

Newly developed catalysts – mechanism unknown



Explicit inclusion of solvent molecules – large simulation cell, longer simulations

String based general Coupled-Cluster method and code

Mihály Kállay

Technical University, Budapest

Department of Physical Chemistry

String-based algorithms

A string is an ordered set of (spin-)orbital indices (p):

$$\mathcal{P} = p_1 p_2 p_3 \dots \quad (p_1 < p_2 < p_3 < \dots).$$

General implementation of quantum chemical methods: arbitrary number of indices.

Solution: string-based formalism. Quantities in terms of strings, loops over strings.

The major advantage of using strings is that it enables one to implement quantum chemical methods in a general way without explicit reference to the number of indices of the wave function parameters and other quantities.

Attainable properties for general (SR and MR) CC and CI methods implemented by string-based techniques

- **First-order properties (analytic gradients): geometries, dipole and higher moments**
- **Second-order properties (analytic second derivatives): harmonic vibrational frequencies, NMR chemical shifts, static and frequency-dependent polarizabilities, g-tensors**
- **Higher-order properties (numerical differentiation): anharmonic force fields, higher-order polarizabilities**
- **Excited-state properties (linear-response functions): excitation energies, transition moments, first-order excited-state properties**

Parallelization of approximate CC methods

Approximate CC methods can simply be parallelized using replicated memory and disk strategies: the outmost loop is split up and the workload is distributed among the nodes.

Non-iterative methods: only one number is communicated
Iterative $CC(n)$ approximations: a reduction of an array of T_{n-1} size is required.

Parallelization employing MPI and Open MP technologies

Example: N_2 , CCSDT(Q)/cc-pVTZ, Opteron cluster, MPI

Number of nodes	1	2	4	8	12
CPU (min)	76.73	38.36	19.20	10.04	6.62
Parallel. efficiency (%)	-	100	100	97	97

Benchmark calculations on a 12-node Opteron cluster

- **Butadiene: CCSDT(Q), cc-pVDZ basis, 10 atoms, 82 MOs, 22 electrons, 9 billion quadruple excitations, 3.1 hours**
- **Benzene: CCSDT(Q), cc-pVDZ basis, 12 atoms, 108 MOs, 30 electrons, 47 billion quadruples, 32 hours**

Calculated contributions to the atomization energy of the
trans-butadiene molecule (kJ/mol)

	Contrib.	Source
HF	3191.57	aug'-cc-pV(T,Q,5)Z extr.
CCSD(T)-HF	1027.37	aug'-cc-pV(Q,5)Z extr.
CCSDT-CCSD(T)	-3.30	cc-pVTZ
CCSDT(Q)-CCSDT	3.84	cc-pVDZ
ZPE	-223.99	CCSD(T)/cc-pVTZ
Anharmonicity	3.48	CCSD(T)/cc-pVDZ
Core-correlation	19.75	CCSD(T)/cc-pCVQZ
Scalar relativistic	-2.72	CCSD(T)/cc-pCVTZ
Spin-orbit splitting	-1.42	Experiment
DBOC	0.83	HF/cc-pVQZ
Total	4015.41	
Experiment	4014.09	± 0.96
	4017.18	± 0.79