Large-scale coupled cluster calculations on supermolecular wires

P. Jørgensen, T.Kjaergaard, K.Kristensen,, I.M.Høyvik, P. Ettenhuber,
J. J. Eriksen, J. Jakowski[a], B. Sumpter[a]

qLEAP Center for Theoretical Chemistry
Department of Chemistry
University of Aarhus

[a]Oak Ridge National Laboratory

Chemistry :

- Modeling's and simulations integrated part of experimental investigation
- Interplay between theory and experiment more important in the future

Example : *Molecular manufacturing*

- Synthesize materials of desired properties by manipulating individual atoms
- Embraces the potential of altering the future of technology

Foresight Nanotech Institute and Battelle

Report 2007 : Technology road map for productive Nano systems

p.17 “*Extending the* ***scale, scope, and accuracy*** *of atomistic modeling techniques is* ***a high priority*** *and can greatly facilitate atomically precise technologies design and implementation*”

Todays modeling on Nano systems use models of low accuracy

- Density functional theory, Force field

p.152 ‘*While quantum mechanical methods exist that approach the absolute limits of accuracy, the use of these methods is currently limited to diminutive chemical systems (< 20 atoms)”*

What have the high accuracy calculation on small systems shown ?

Small Molecules: *Accomplishments using high accuracy models*

- Calculations changed how many experimental investigation are carried out,

  A theoretical investigation often precede an experimental

- Calculations broadening the understanding of the investigated phenomena

- In some cases calculations have replaced experiment :

  Who would nowadays try to determine equilibrium
  geometries for small molecules experimentally?

Large molecules: *Challenges for high accuracy models*

- The scaling in conventional implementations prohibit applications
- Restructure calculations for lower scaling
- Even low scaling algorithms require large computational resources
- Development of low scaling methods must go hand in hand with access to increased computational resources
- Expect similar development with respect to interplay with experiment as for small molecules

Development of code for the future

Theoretical challenges

- Restructure high scaling models to become lower scaling
- The low scaling models have to be massively parallel
- Parallelism at several levels (Coarse, medium and fine grained)

Challenges in connection with implementations

- Memory available
- Data traffic
- Many many more, refer to next talk by P. Ettenhuber

Testing, testing and testing !!!!!

Granting agency :

- Which chemical problem can you solve (using available methods)
- Equally important: Development of the methods for the future

For the future :

Close contact between developers of tomorrows code and tomorrows supercomputer software and hardware

- Open new horizon we have not even dreamed about

30 years of research in electronic structure theory for small molecules have shown

a) Coupled cluster theory is the method of choice

b) Hierarchies of coupled models with improved accuracy

MP2 : Second order Møller Plesset theory , N^5 scaling

CCSD : Coupled cluster singles and doubles model , N^6 scaling

CCSD(T) : CCSD with perturbative triples corrections ,N^7 scaling

(N denote system size)

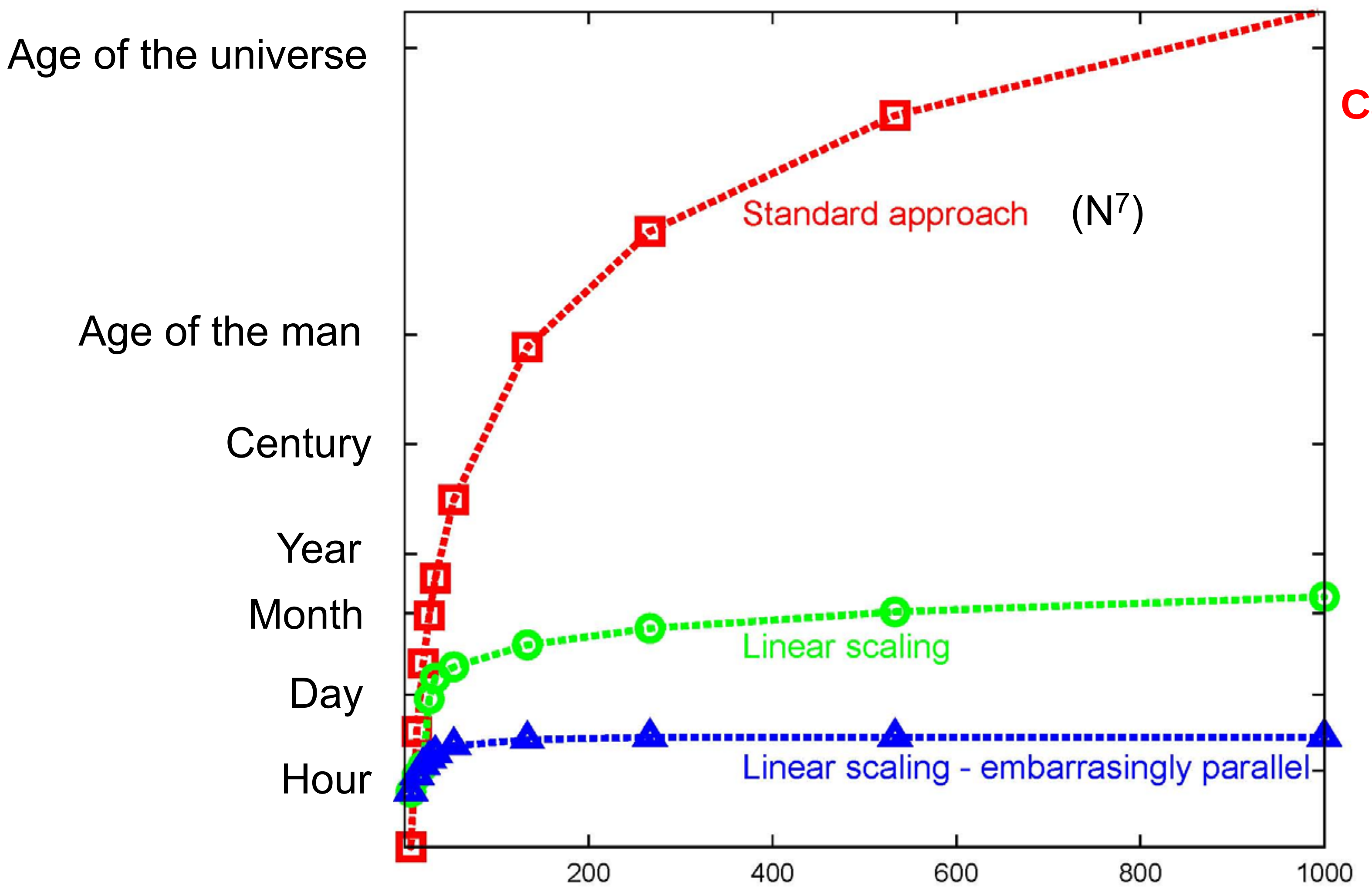
CCSD(T) is the golden standard of quantum chemistry

Many molecular properties are described to experimental accuracy or better e.g. equilibrium geometries and reaction enthalpies

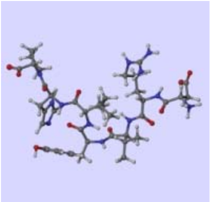
Problem: The scaling in standard implementations

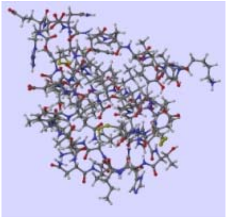
Scaling in Coupled Cluster theory CCSD(T)

Standard
Coupled Cluster
Canonical HF orbitals



DEC
Coupled Cluster
Local HF orbitals





Insuline

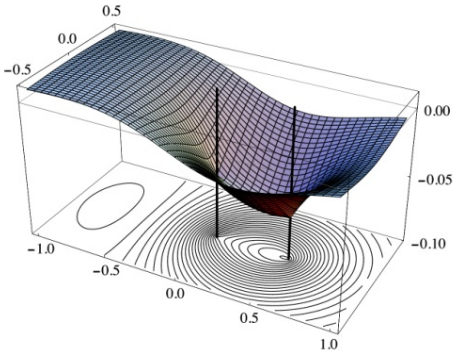
Coupled cluster wave function calculation

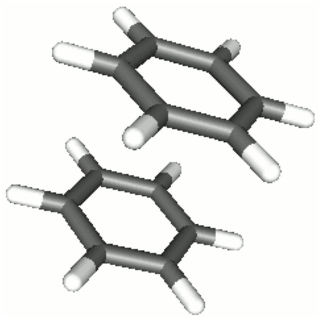
Hartree-Fock calculation determines the reference state

- Electrons move in an averaged field of the others
- Long range potential described to high accuracy
- **Give local orbitals for both occupied and virtual space**

Coupled cluster calculation describes local electron correlation effects

- Coulomb hole, Short range
- Dispersion forces, R^{-6}





Local phenomena described in **local basis** give linear scaling

The divide-expand-consolidate (DEC) coupled cluster method

The DEC coupled cluster (CC) method

Assumption

- ***Local occupied*** and ***virtual*** HF orbitals can be determined

Method

- Assign ***local occupied*** and ***virtual*** HF orbitals to atomic sites, P, Q,..

- Partition correlation energy into atomic fragment E_P and pair ΔE_{PQ} energies

  Replace (i,j) or (a,b) summations in correlation energy with P and PQ summations

- Evaluate E_P and ΔE_{PQ} from small orbital fragment spaces adjusted to give the energies to a predefined fragment optimization threshold (FOT)

Features

- Adjustable orbital spaces give error control for energies and amplitudes

- Black box, linear scaling, and embarrassingly parallel

Precision of Coupled Cluster calculation

Standard Coupled Cluster

- Residual norm (***R***) of amplitude equation

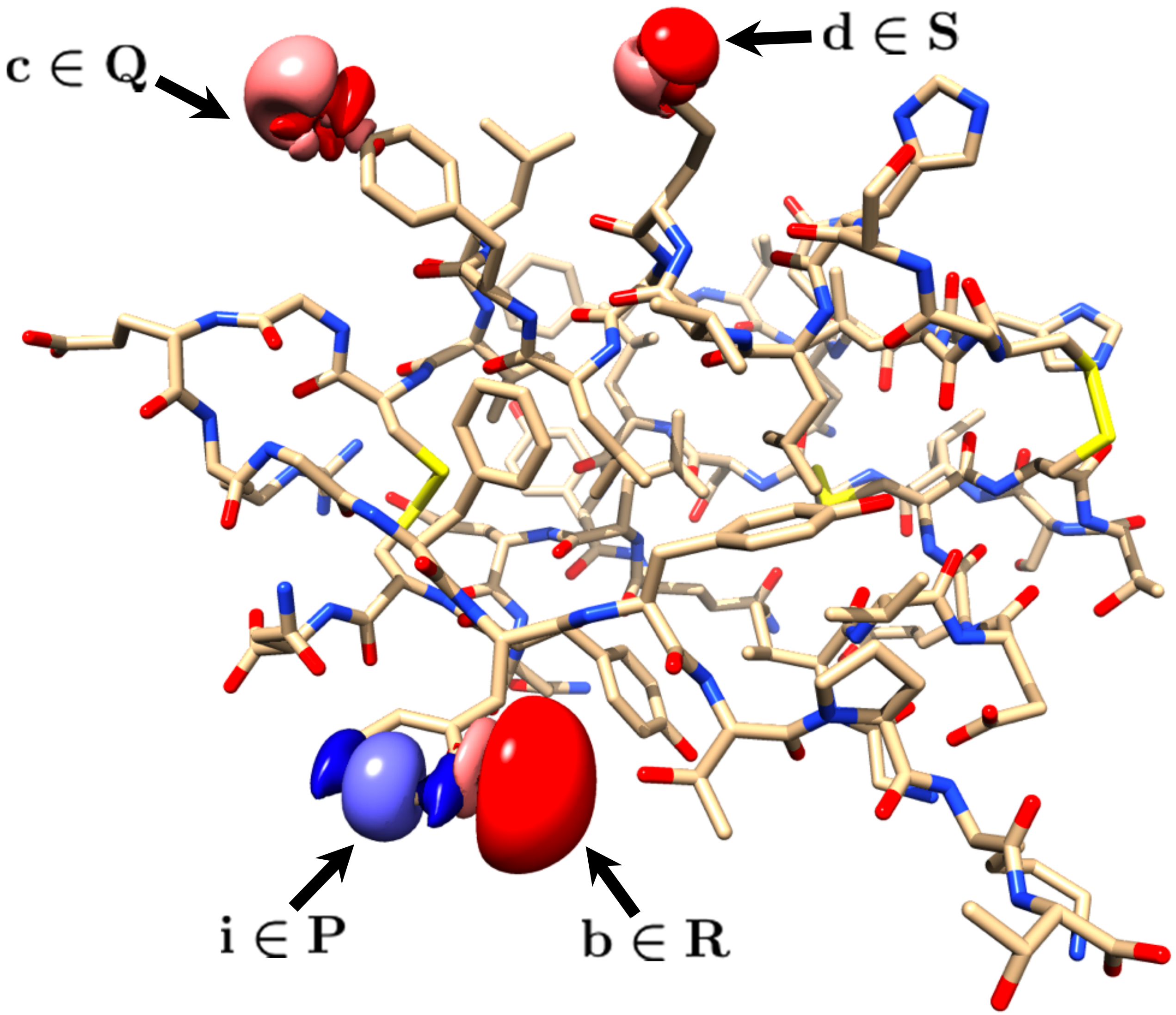
DEC Coupled Cluster (new strategy which exploit locality efficiently)

- Fragment optimization threshold (***FOT***) for atomic fragment energies

Standard and DEC Coupled Cluster on par with respect to precision

- Single parameter (***R*** or ***FOT***) define the precision of energy, density, ...

Determine local Hartree-Fock orbitals



INSULIN MOLECULE (HYDROGENS OMITTED)

Occupied orbitals:

i,j,k,l

Virtual orbitals:

a,b,c,d

Atomic sites:

P,Q,R,S

Assign orbitals to atomic sites

Correlation energy MP2 and CCSD

$$E_{corr} = \sum_{ijab} (t_{ij}^{ab} + t_i^a t_j^b)(2g_{iajb} - g_{ibja})$$

CCSD(T) energy correction

$$\Delta E^{CCSD(T)} = \sum_{AI} {}^{*}t_I^A \, {}^{*}T_I^A + \frac{1}{4} \sum_{IJ\,AB} {}^{*}t_{IJ}^{AB} \, {}^{*}T_{IJ}^{AB}$$

DEC Strategy

Assign local orbitals to atomic sites P,Q,..

Replace summation over two occupied i,j (virtual a,b) orbitals with summations over sites P and pair sites P,Q and summations over orbitals belonging to these sites

Occupied space partitioning of E_{corr}

Coulomb hole

Dispersion energy

No approximation

$$E_{corr} = \sum_{P} E_P^{\mathrm{O}} + \sum_{P>Q} \Delta E_{PQ}^{\mathrm{O}}$$

Quadratic scaling

Occupied atomic fragment energy

$$E_P^{\mathrm{O}} = \sum_{\substack{ij\in P\\ ab}} \left(t_{ij}^{ab} + t_i^a t_j^b\right)\left(2g_{iajb} - g_{ibja}\right)$$

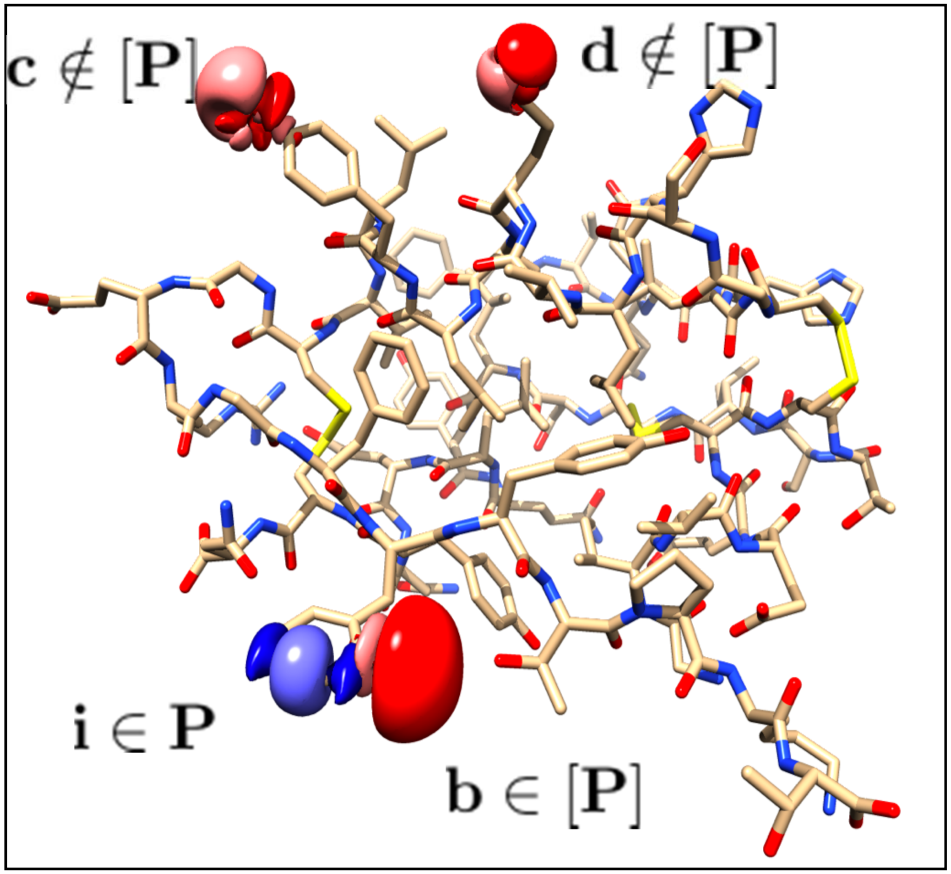
Occupied pair interaction energy

$$\Delta E_{PQ}^{\mathrm{O}} = \sum_{\substack{i\in P,\, j\in Q\\ ab}} \left(t_{ij}^{ab} + t_i^a t_j^b\right)\left(2g_{iajb} - g_{ibja}\right) + \quad P \leftrightarrow Q \ term$$

No approximations have so far been made, only a reorganization

Atomic fragment energy E_P

$$E_P^{\mathrm{O}} = \sum_{ij\in P\ ab\in[P]} \left(t_{ij}^{ab} + t_i^a t_j^b\right)\left(2g_{iajb} - g_{ibja}\right)$$



Determine [P] in a black box manner such that errors in E_P is smaller than **Fragment Optimization Threshold (*FOT*)**

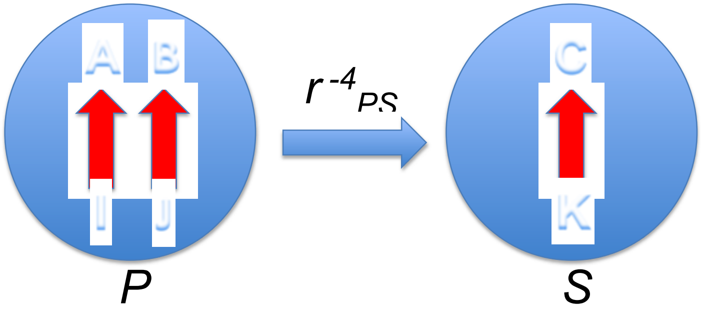
CCSD and MP2

Charge distributions in integrals determine distance decay from site P

$$g_{iajb} \quad ;\ i,j\in P \quad ,\ a,b\in[P]$$

CCSD(T)

Charge distributions together with third party excitations to site S determine distance decay from site P



Occupied space partitioning of E_{corr}

Occupied atomic fragment energy

$$E_P^{\mathrm{O}} = \sum_{\substack{ij \in P \\ ab \in [P]}} \left(t_{ij}^{ab} + t_i^a t_j^b\right)\left(2g_{iajb} - g_{ibja}\right)$$

(atomic fragment orbital space)

Occupied atomic pair interaction energy

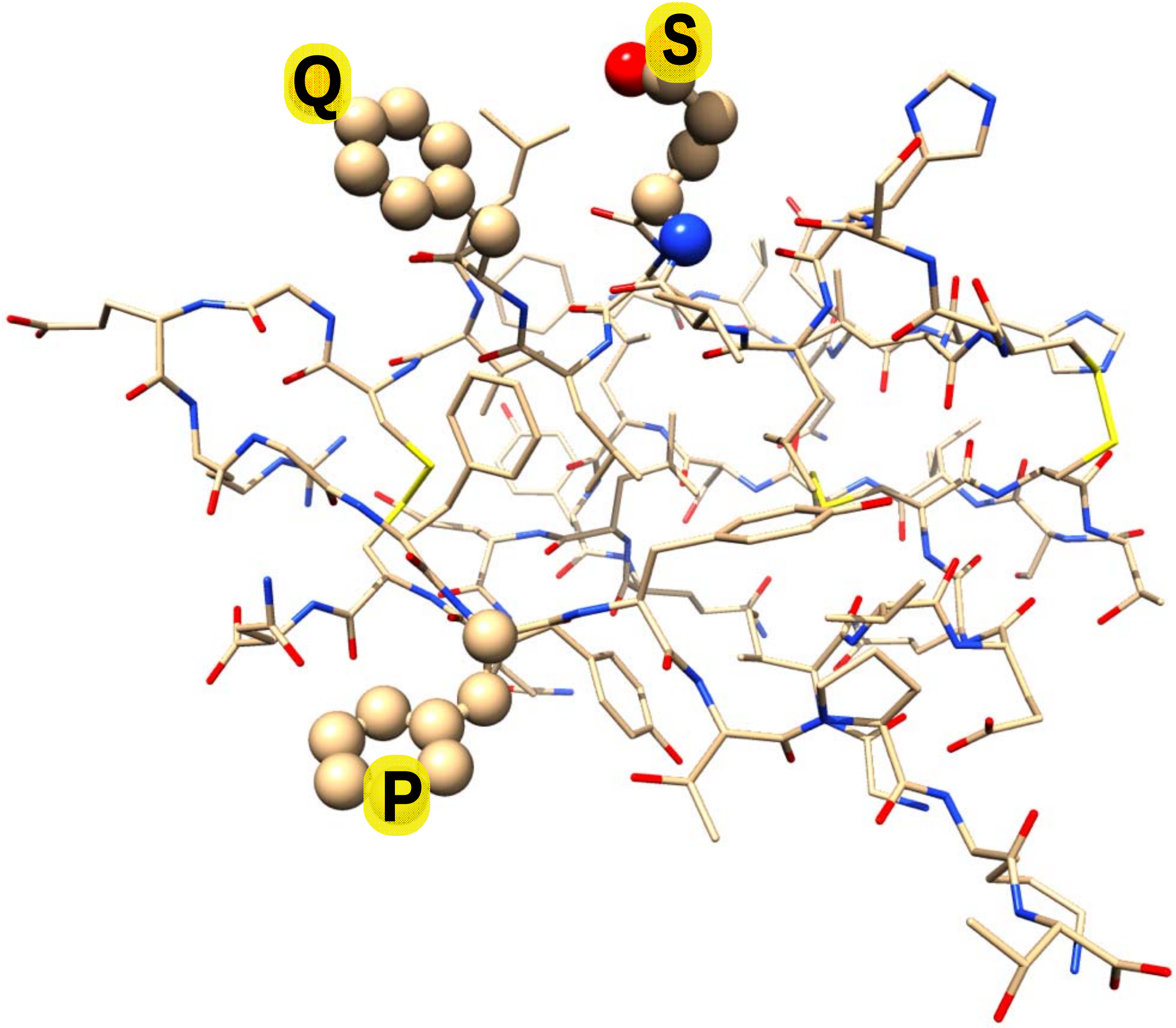
$$\Delta E_{PQ}^{\mathrm{O}} = \sum_{\substack{i \in P, j \in Q \\ ab \in [P] \cup [\Theta]}} \left(t_{ij}^{ab} + t_i^a t_j^b\right)\left(2g_{iajb} - g_{ibja}\right) + \quad P \leftrightarrow Q \ term$$

(union of atomic fragment orbital spaces)

Insulin - three examples of atomic fragments

$$E_{\mathrm{corr}} = \sum_{P} E_P + \sum_{P>Q} \Delta E_{PQ}$$

Quadratic scaling



$\mathbf{E_P}$ determined to FOT precision

↓

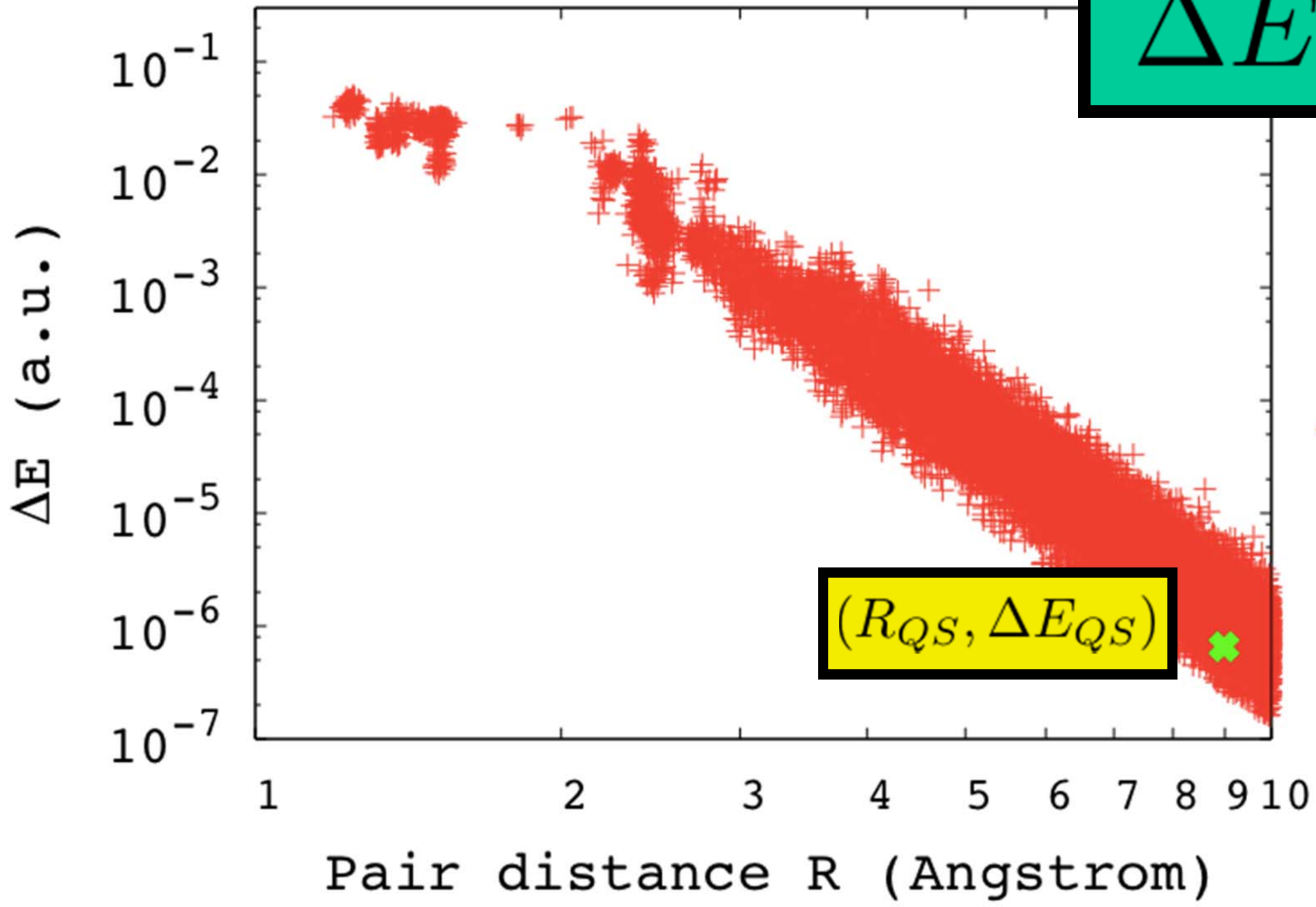
$\mathbf{\Delta E_{PQ}}$ determined to FOT precision

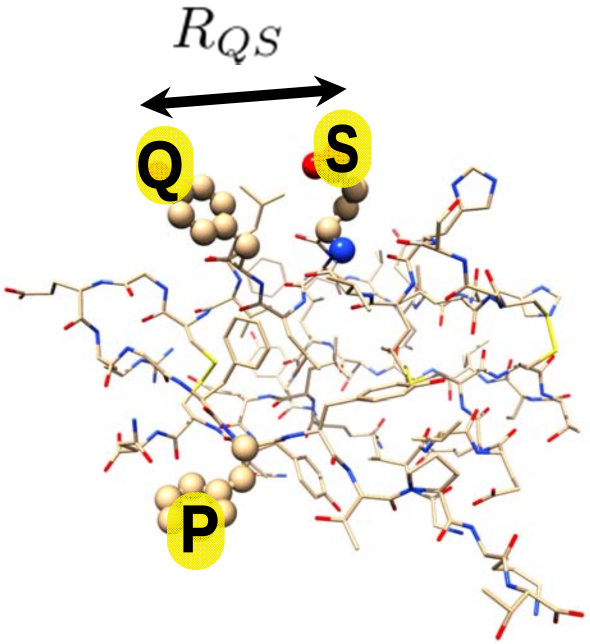
↓

$\mathbf{E_{corr}}$ determined to FOT precision

Insulin - pair interaction energies ΔE_{PQ}

$$\Delta E \propto R^{-6}$$



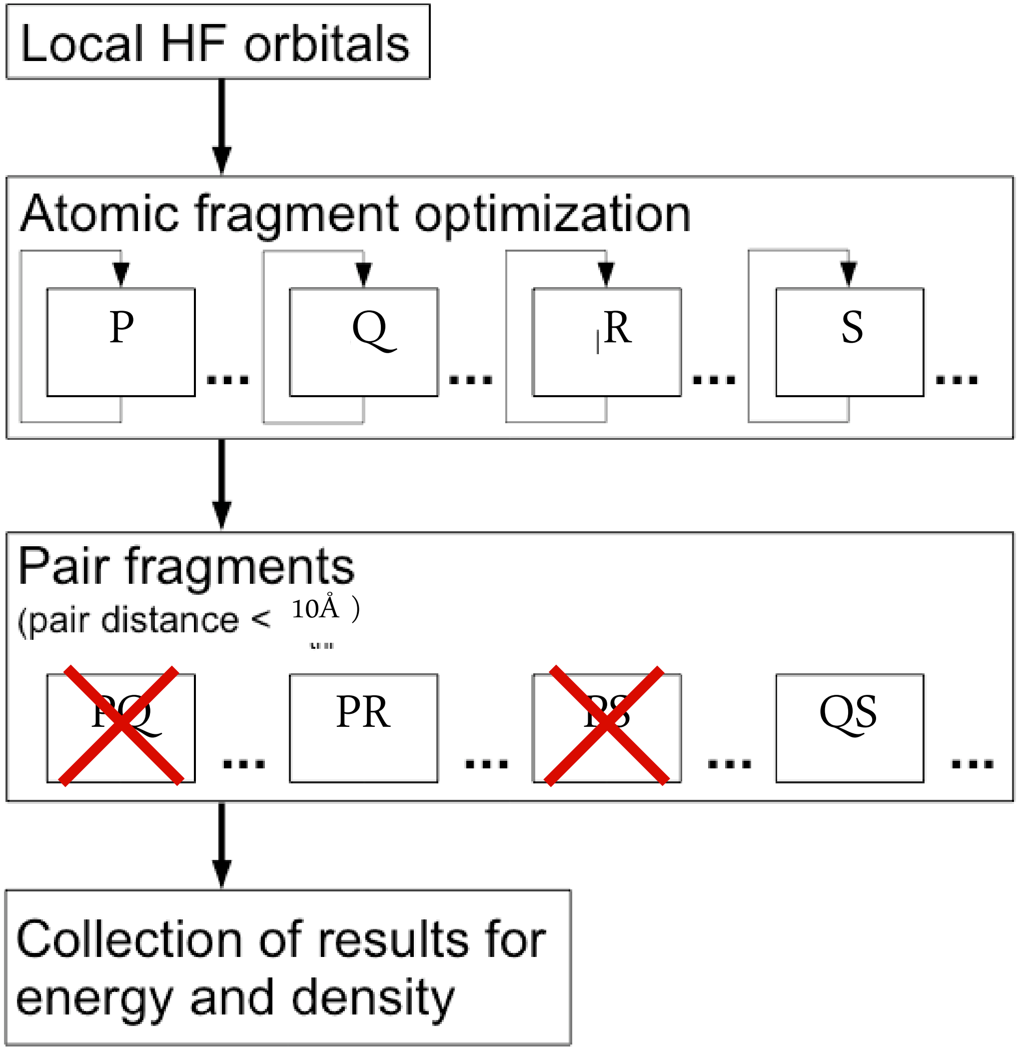


Pairs separated by more than ~10 Å can be neglected without affecting the precision of the calculation

Linear scaling algorithm

19L

Summary of the DEC scheme



Local HF orbitals

Atomic fragment optimization

P … Q … R … S …

Pair fragments

(pair distance < 10Å)

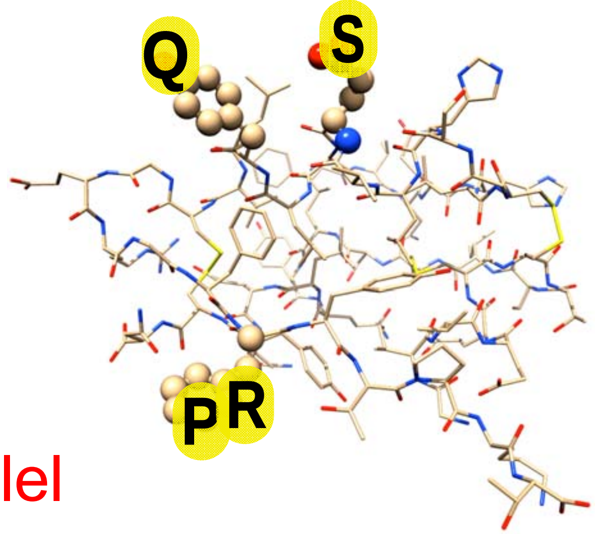
~~PQ~~ … PR … ~~PS~~ … QS …

Collection of results for energy and density

N atomic fragments

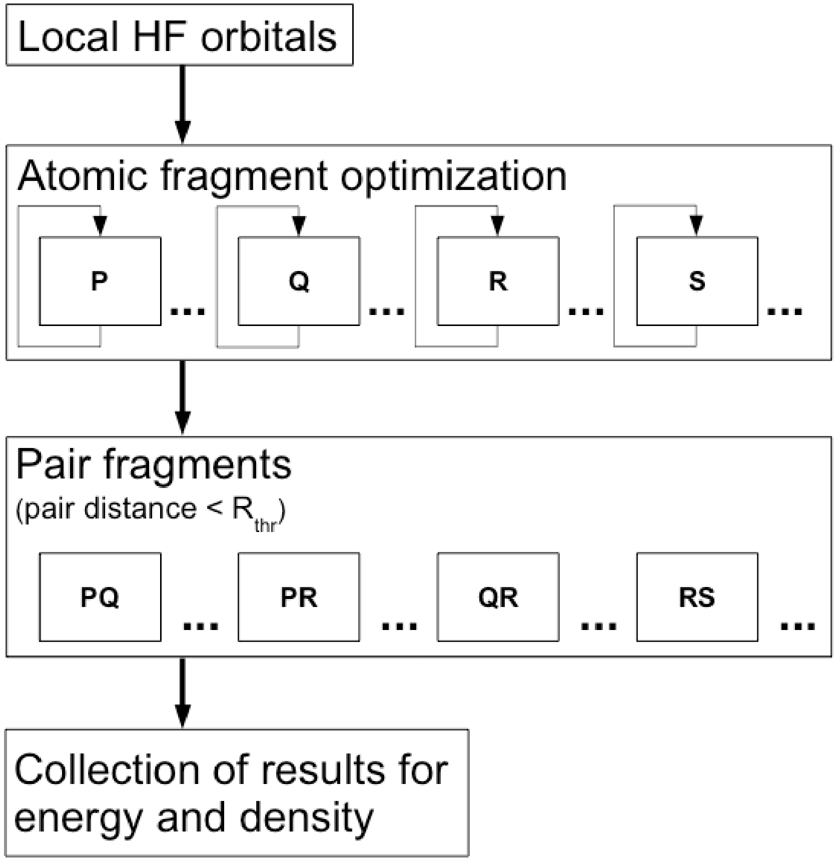
Formally: N*(N-1)/2 pair fragments

Use cut-off: const*N pair fragments



DEC is linear scaling and embarrassingly parallel

Parallelism in DEC calculation


Local HF orbitals

Atomic fragment optimization

P … Q … R … S …

Pair fragments
(pair distance < R_{thr})

PQ … PR … QR … RS …

Collection of results for energy and density

Parallelity at three levels

Coarse grained parallelism

All fragment calculations E_P and ΔE_{PQ} carried out independently

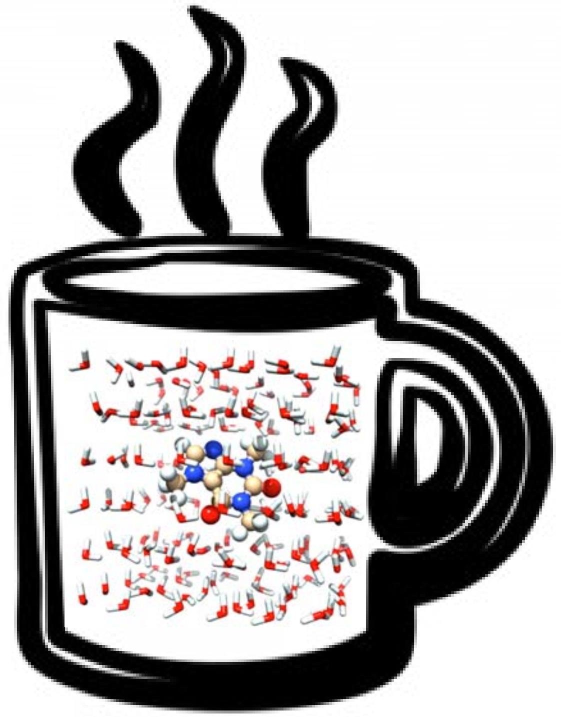
Medium and fine grained parallelism

Individual fragment calculations parallelized at two levels (MPI and OpenMP)

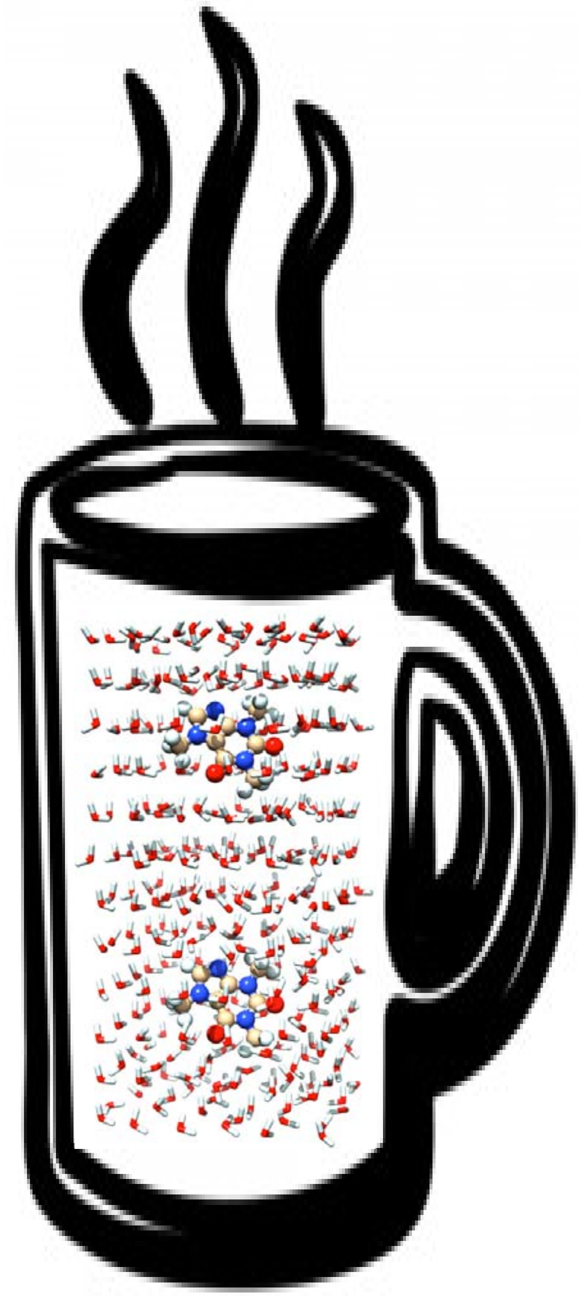
(as in conventional implementations)

“A nano-cup of coffee...”

- Calculate MP2 correlation energy and density using cc-pVDZ basis.
- Nanospresso (system **1**): 528 atoms (4278 BF.)
- Nanospresso Doppio (system **2**): 1056 atoms (8556 BF.)



Nanospresso



Nanospresso doppio

Time to solution (TTS)

System	#fragments	#nodes	TTS(hours)
1	7136	5890	1.66
1	7136	11780	0.93
2	16151	5890	4.49
2	16151	11780	2.37

- Calculations on Titan, Oak Ridge National Laboratory* using 8 OpenMP threads per MPI process. Nodes refer to NUMA nodes

- **Scaling with system size?**

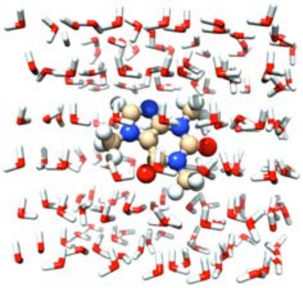
  #frags(2) / #frags(1) = 2.26 (ideal)
  TTS(2) / TTS(1) = 2.70 (5890 nodes)
  TTS(2) / TTS(1) = 2.55 (11780 nodes)

- **Parallel (strong) scaling** (ideal: 0.5)

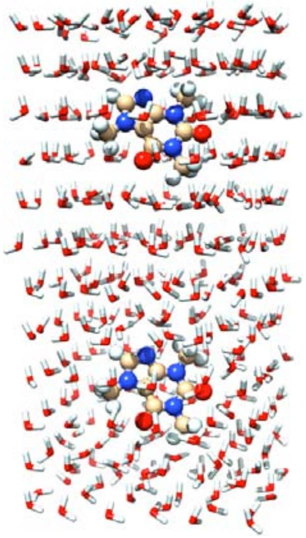
  TTS(11780) / TTS(5890) = 0.56 (system 1)

  TTS(11780) / TTS(5890) = 0.53 (system 2)

1



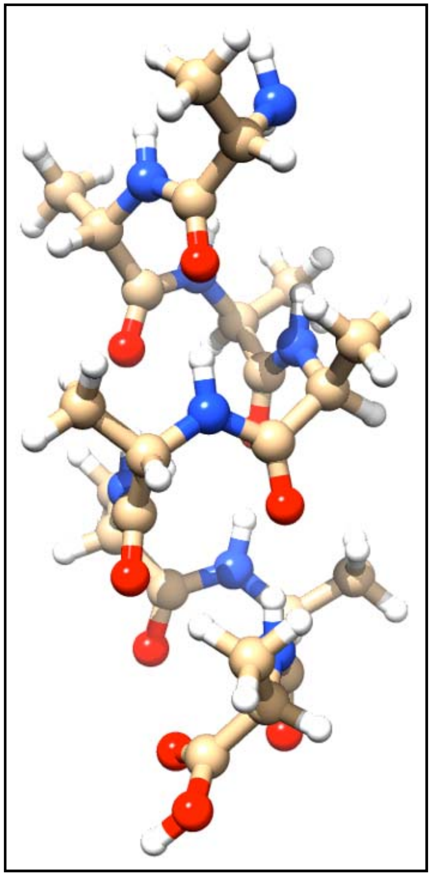
2



*https://www.olcf.ornl.gov/

Test molecule: Energy errors vs. FOT

MP2/cc-pVDZ calculation: Errors in DEC correlation energies (a.u.)



Alanine(8)

FOT	Occupied energy error	Virtual energy error	Lagrangian energy error	Average relative error*	Average percentage of E_{corr}
10^{-2}	$1.38*10^{-1}$	$2.51*10^{-1}$	$2.00*10^{-1}$	3.1*FOT	96.9%
10^{-3}	$2.83*10^{-2}$	$2.76*10^{-2}$	$1.87*10^{-2}$	3.9*FOT	99.6%
10^{-4}	$3.20*10^{-3}$	$1.87*10^{-3}$	$1.95*10^{-3}$	3.6*FOT	99.96%
10^{-5}	$2.69*10^{-4}$	$2.56*10^{-4}$	$1.78*10^{-4}$	3.7*FOT	99.996%
10^{-6}	$3.25*10^{-5}$	$1.09*10^{-5}$	$2.01*10^{-5}$	3.3*FOT	99.9997%

DEC MP2 insulin calculation: Energy

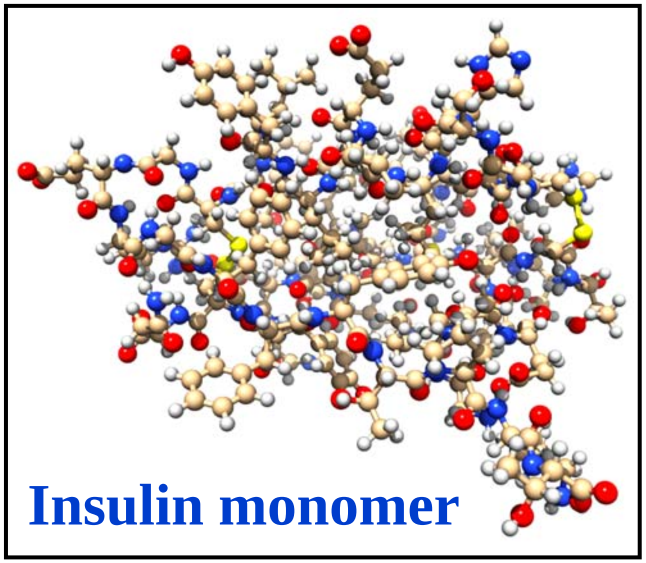
cc-pVDZ: 7604 basis functions

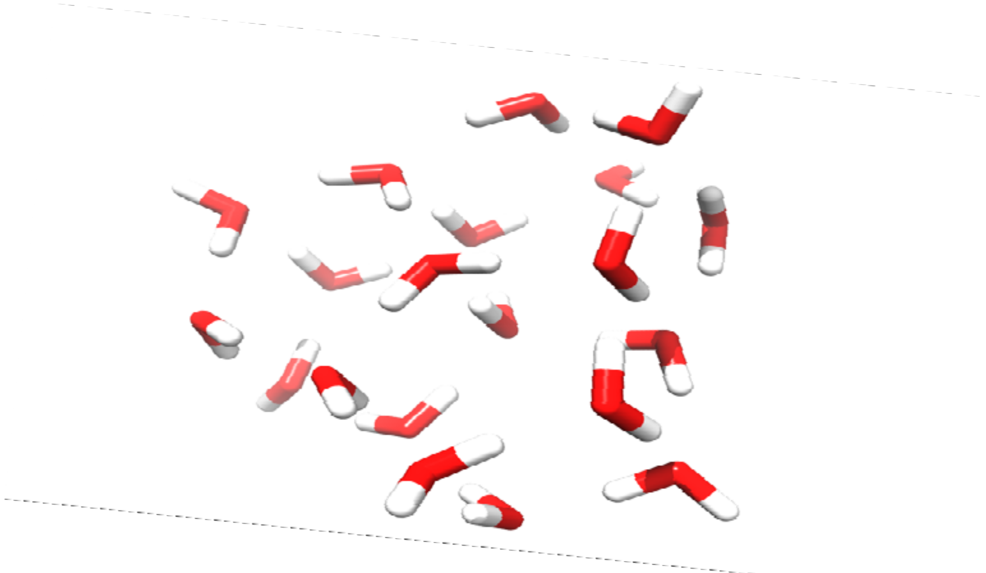
FOT	Occupied correlatione nergy	Virtual correlatione nergy	Lagrangian correlation energy	Estimated absolute error*	Estimated relative error**	Estimated percentage of E_{corr}
10^{-4}	-61.755	-61.783	-61.763	0,028	4.5*FOT	99.95%

- Consistent with errors for smaller test molecules
- E.g. alanine(8) for FOT=10^{-4}: Relative error = 3.6*FOT (99.96% of E_{corr})

DEC is a black box method:

The relative errors are system-independent, regardless of system size


Insulin monomer



20 water cluster

6-31G*

FOT	MP2	CCSD	(T)	CCSD(T)
10^{-3}	$2.1*10^{-2}$	$1.4*10^{-2}$	$7.0*10^{-3}$	$2.2*10^{-2}$
10^{-4}	$2.3*10^{-3}$	$4.4*10^{-4}$	$2.4*10^{-3}$	$2.8*10^{-3}$
10^{-5}	$2.3*10^{-4}$	$4.7*10^{-5}$	$3.1*10^{-4}$	$3.6*10^{-4}$
10^{-6}	$2.3*10^{-5}$	$-3.5*10^{-5}$	$9.0*10^{-5}$	$5.5*10^{-5}$

20 water cluster

cc-pVTZ

FOT	CCSD	(T)	CCSD(T)
10^{-3}	$2.8*10^{-2}$	$1.6*10^{-2}$	$4.4*10^{-2}$
10^{-4}	--	--	--

Largest conventional CCSD(T) calculation $(H_2O)_{20}$

E. Apra *et al.,* **SC**09, submission for Gordon Bell prize (2009)

Also:

CCSD(T) calculation carried out on one "Nanospresso"

with FOT=10^{-3}.

4278 basis functions on 3201 nodes in 4 hours and 20 minutes

Conclusion for DEC

- New strategy for CC energy and properties which explore locality efficiently
- Full system in terms of CC calculations on small fragments of total orbital space
- Linear scaling and massive parallel algorithm
- Full control of errors in energy and cluster amplitudes
- Black box method

- Performance (speed) depends on locality of HF orbitals
- DEC on par with standard CC method,
  FOT defines precision as residual norm in standard CC

Acknowledgements

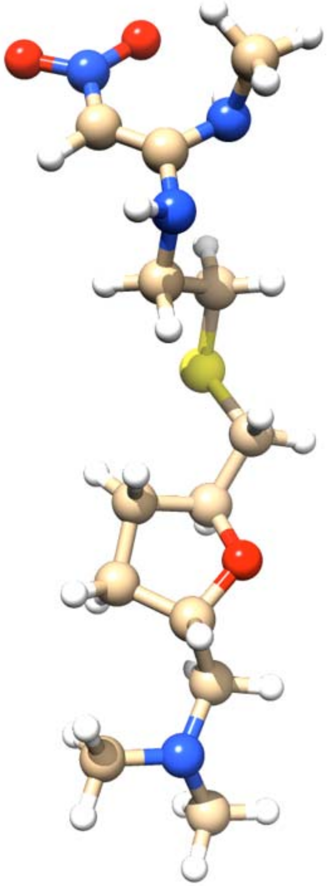
Work supported by:

- US Department of Energy
- Oak Ridge Leadership Computing Facility
- European Research Council
- Danish Council for Independent Research – Natural Sciences
- Aarhus University

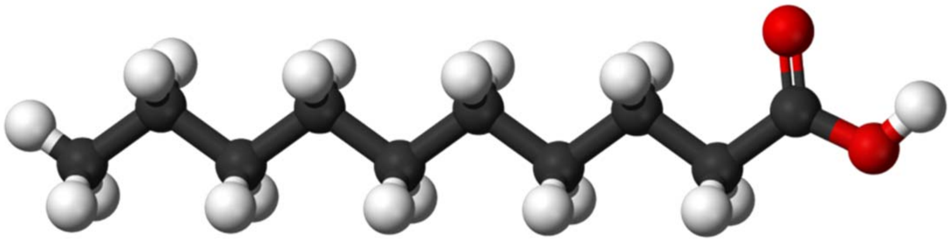
DEC MP2 molecular gradient errors

cc-pVTZ basis, all errors are given in a.u.

FOT	Energy error	RMS error for gradient
10^{-3}	$9,8*10^{-3}$	$1,5*10^{-3}$
10^{-4}	$9.1*10^{-4}$	$4,5*10^{-4}$
10^{-5}	$6.9*10^{-5}$	$2,7*10^{-5}$
10^{-6}	$7.5*10^{-6}$	$8,3*10^{-6}$



ZANTAC



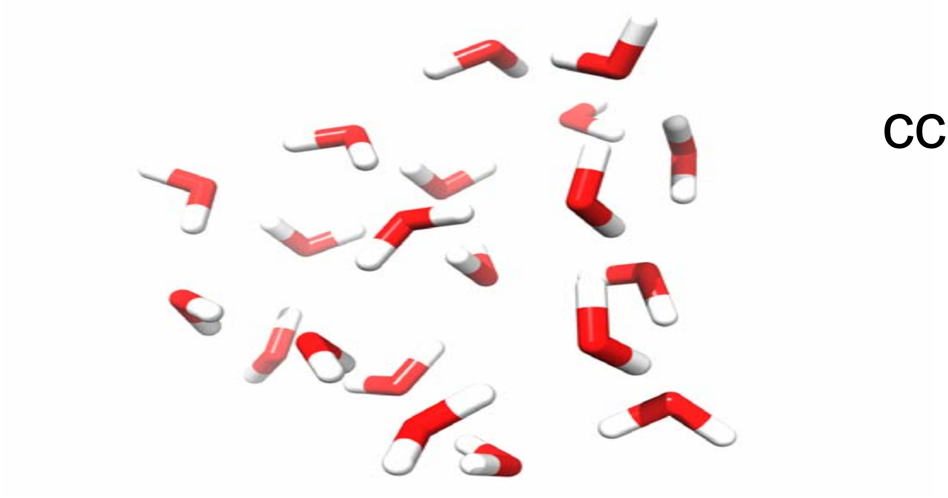
Decanoic acid

6-31G*

FOT	MP2 occ	MP2 vir	CCSD occ	CCSD vir	CCSD(T) occ	CCSD(T) vir
10^{-3}	$1.9*10^{-2}$	$1.5*10^{-2}$	$-2.0*10^{-3}$	$-1.1*10^{-2}$	$5.1*10^{-3}$	-4.410^{-3}
10^{-4}	$1.3*10^{-3}$	$1.5*10^{-3}$	$-3.8*10^{-3}$	$-3.5*10^{-3}$	$-1.6*10^{-3}$	$-2.0*10^{-3}$
10^{-5}	$1.2*10^{-4}$	$1.4*10^{-4}$	$-1.2*10^{-3}$	$-1.3*10^{-3}$	$-3.6*10^{-4}$	$-9.0*10^{-4}$
10^{-6}	$1.2*10^{-5}$	$1.3*10^{-5}$	$-3.3*10^{-4}$	$-3.6*10^{-4}$	$-1.1*10^{-4}$	$-2.3*10^{-4}$

Largest conventional CCSD(T) calculation $(H_2O)_{20}$

E. Apra *et al.*, **SC**09, submission for Gordon Bell prize (2009)



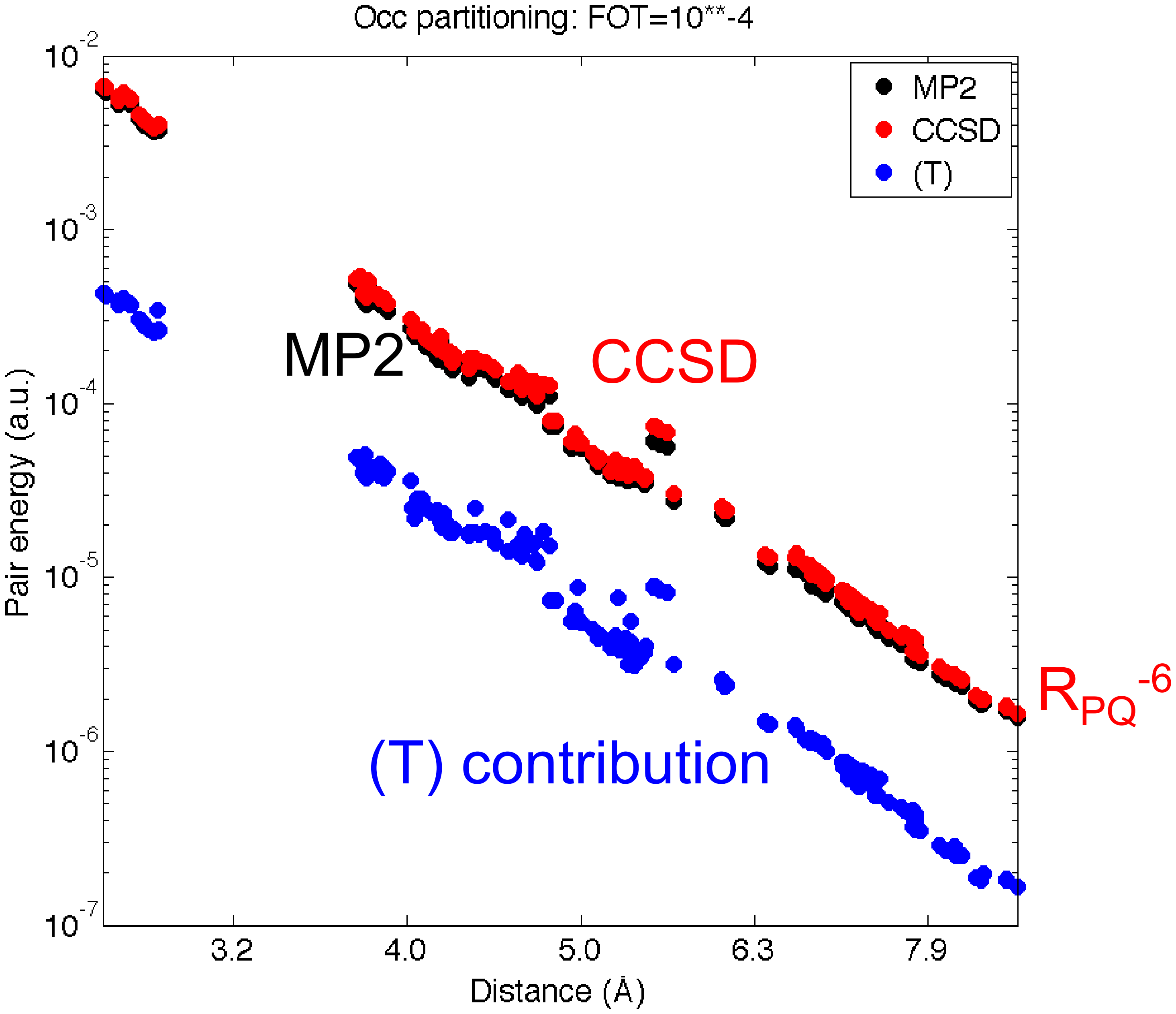
cc-pVTZ

Model \Error	MP2	CCSD(T)
FOT=10-3	1.5e-2	1.5e-2
FOT=10-4	1.6e-3	4.6e-4

MP2, CCSD, and CCSD(T) pair interaction energies

$(H_2O)_{20}$

cc-pVTZ



DEC perspective

MP2

- Energy, density, molecular gradient and geometry optimizer ***Done***
- F12 energies, Nuclear shifts, ... ***In progress***

CCSD

- Energy implemented ***Ready to be tested for massive parallelism***
- Future work, density, molecular gradient, F12 energy ...

CCSD(T)

- Energy implemented ***Ready to be tested for massive parallelism***
- Future work, density, molecular gradient, ...

DEC CCSD(T) energy

$$\Delta E^{CCSD(T)} = \sum_{AI} {}^{*}t_I^A \, {}^{*}T_I^A + \frac{1}{4} \sum_{IJ\,AB} {}^{*}t_{IJ}^{AB} \, {}^{*}T_{IJ}^{AB}$$

Two occupied and two virtual indices

$$\Delta E^{CCSD(T)} = \sum_{P} {}^{(T)}E_{\underline{P}} + \sum_{P>Q} {}^{(T)}E_{\underline{PQ}}$$

Triples fragment and pair fragment energies

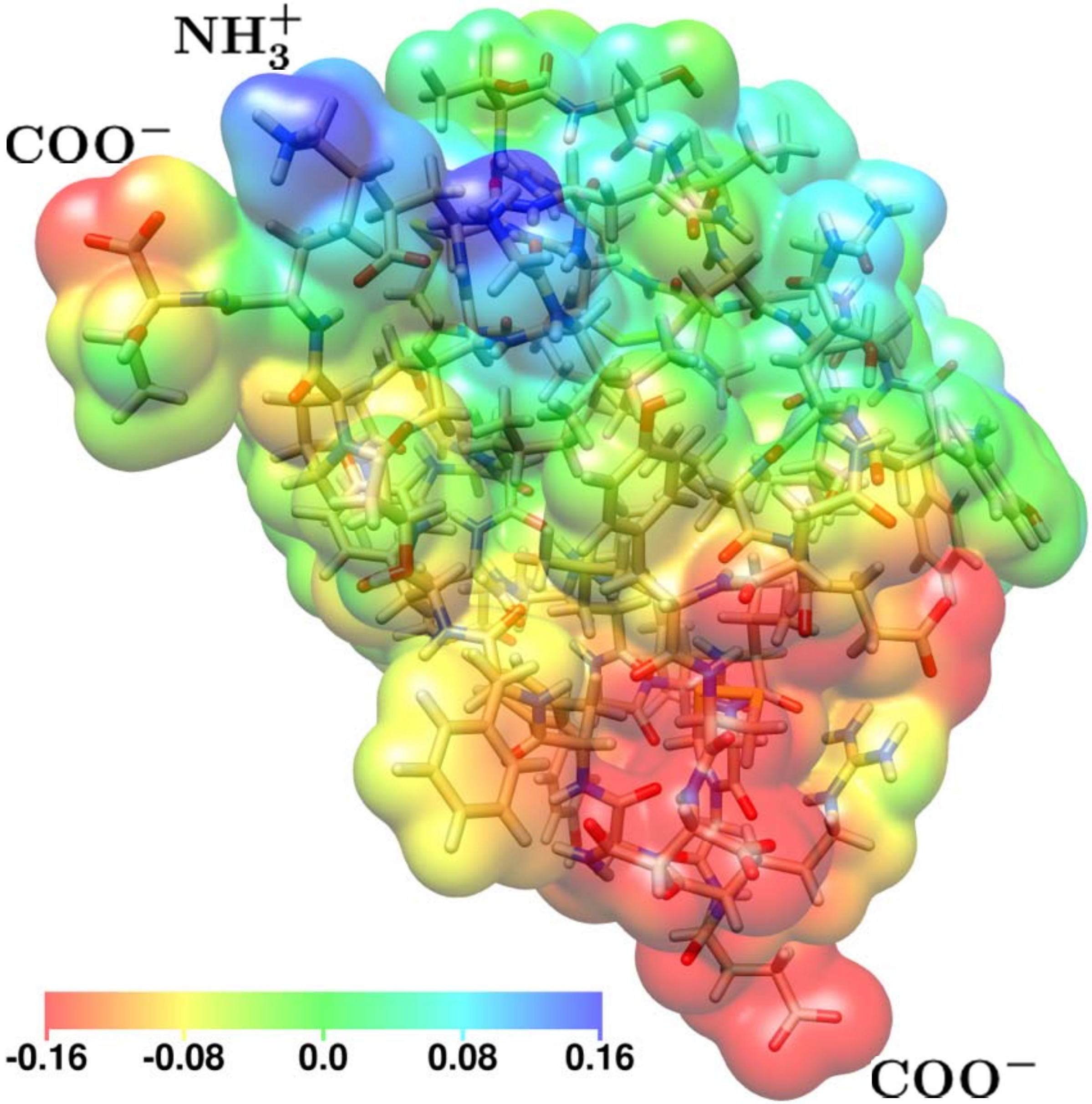
$${}^{*}T_I^A = \left\langle AI \left[\Phi, {}^{*}T_3 \right] \middle\| HF \right\rangle = \frac{1}{12} \sum_{\substack{CD \\ KL}} ({}^{*}t_{IKL}^{ACD} - 2\,{}^{*}t_{LKI}^{ACD}) g_{KCLD}^{a}$$

$${}^{*}T_{IJ}^{AB} = \left\langle ABIJ \left[\Phi, {}^{*}T_3 \right] \middle\| HF \right\rangle = \frac{2}{3} [\sum_{CDK} [g_{DKBC}^{a} \, {}^{*}t_{IJK}^{ACD} - g_{DKBC} \, {}^{*}t_{KJI}^{ACD}] - \sum_{CKL} [g_{LCKJ}^{a} \, {}^{*}t_{IKL}^{ABC} - g_{LCKJ} \, {}^{*}t_{LKI}^{ABC}]]$$

$$\varepsilon_{IJK}^{ABC} \, {}^{*}t_{IJK}^{ABC} = -P(IJK)P(ABC)[\sum_{D} g_{CKBD}^{a} \, {}^{*}t_{IJ}^{AD} - \sum_{L} g_{CKLJ}^{a} \, {}^{*}t_{IL}^{AB}]$$

$$P(QRS)f(Q,R,S) = f(Q,R,S) + f(S,Q,R) + f(R,S,Q)$$

DEC MP2 electrostatic potential for insulin



Insulin monomer:

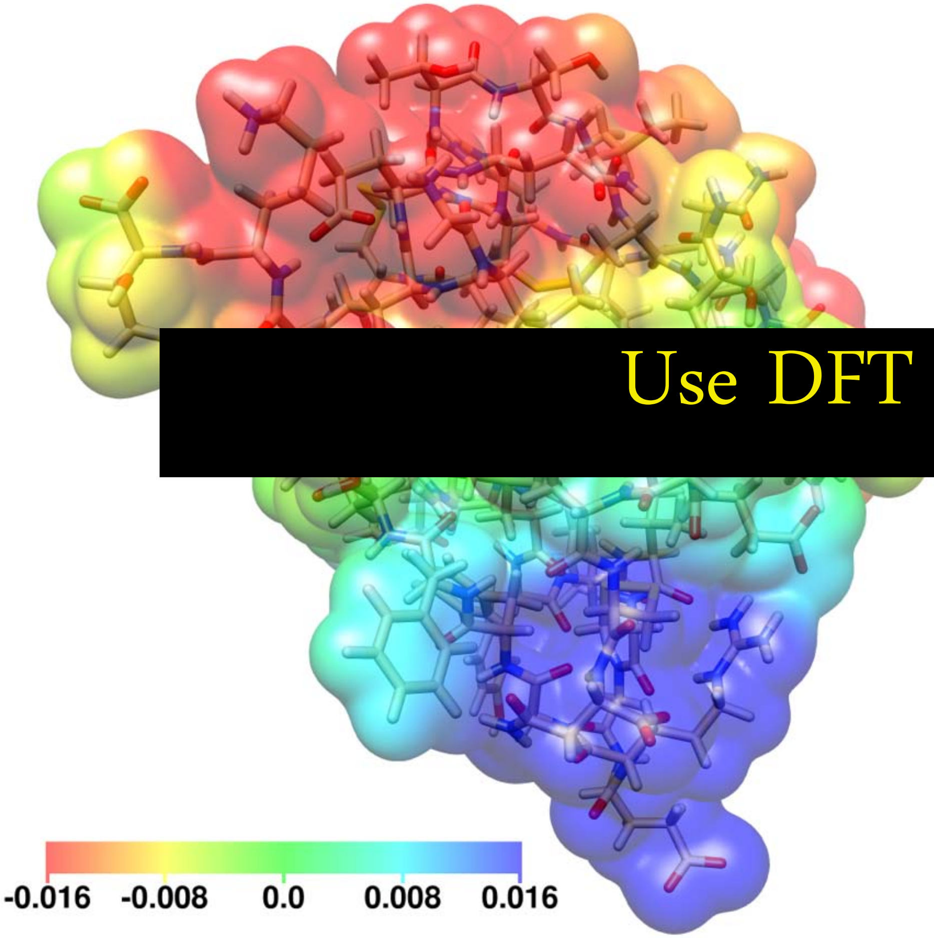
$C_{257}N_{65}O_{77}S_6H_{382}^-$

Red/blue regions indicate high/low potential energy for a positive point charge

Benchmarking other models:
DFT - MP2 difference in electrostatic potential

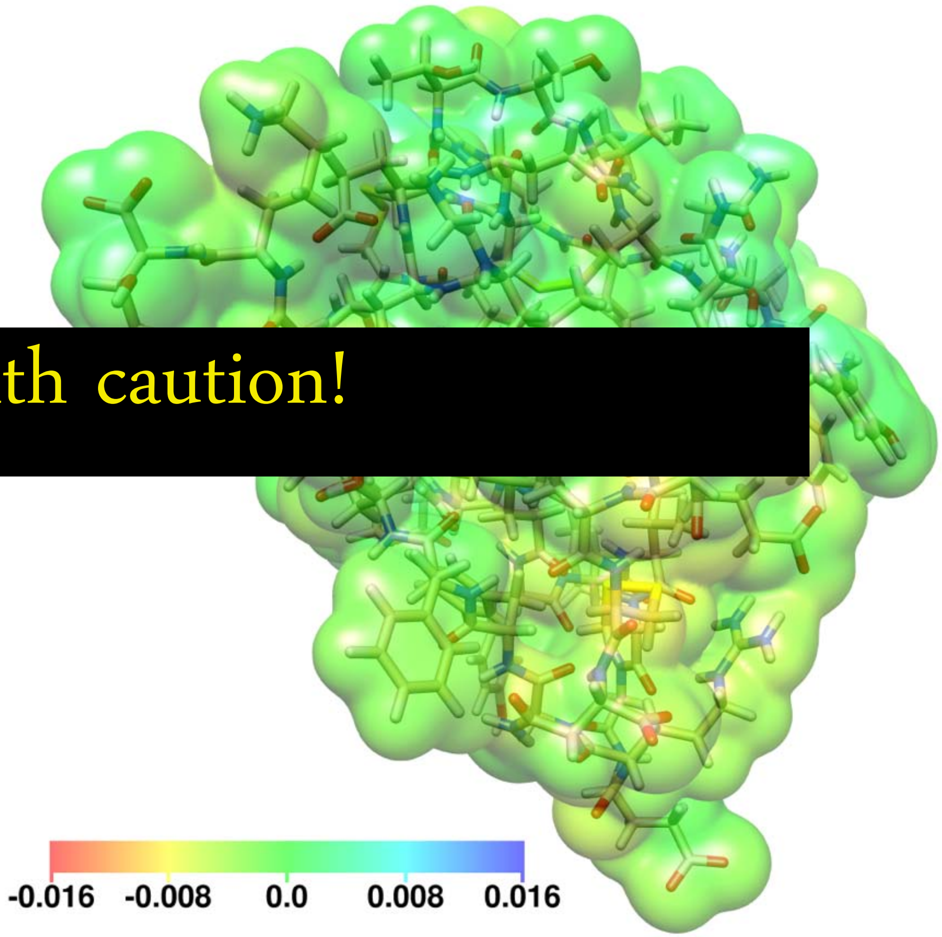
B3LYP - MP2

(no long-range correction)



CAMB3LYP - MP2

(long-range correction)



Use DFT with caution!

Red/blue regions correspond to increased/decreased electrostatic potential for DFT compared to MP2