



Theoretical methods for large system

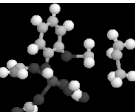
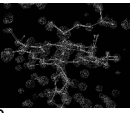




QM/MM and related methods

Winter school in theoretical chemistry
Helsinki, 2002-12-10/11



Ulf Ryde
Theoretical Chemistry
Lund University
Sweden


1. Technical implementation
2. Performance and applications
3. Quantum refinement

QM/MM and related methods
Lecture 1

Technical implementation

1. Problem
2. Solutions
3. Philosophy
4. Implementations
5. Variants


**How study
LARGE molecules
with theoretical methods**



?

- Geometry
- Energy
- Reactions
- Properties

E.g. Proteins



Related problems

Solvation	Entropy
How do we describe solvation effects?	How do we obtain free energies?
From 1 to 10^{26} molecules	Sampling of phase space
Easier: homogenous medium	Statistical mechanics
But still many atoms	Many conformations


Solutions
(= this course)

- Molecular mechanics
- Linear-scaling methods
- Supermolecular approach
 - QM/MM methods
 - Solvation methods
 - Hybrid methods

Molecular Mechanics

No electrons or Quantum Mechanics
Balls and springs
Empirical potential

+ Inexpensive and fast	- Parameterisation
+ Dynamics	- Many parameters
+ Physiological conditions	- Low accuracy



- Structures
- Energies only if heavily parameterised
 - No (electronic) properties
 - Explicit solvation
 - ΔG

Supermolecular approach

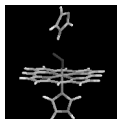
Include only **part** of the system
(active site)

Fix atoms to get a
more realistic structure

- + Accurate method
- Expensive
- One structure
- Effect of protein?
- Biased choice



- \pm Structure
- Accurate energies
- Accurate properties
- Solvation by continuum method
- ΔG from frequency calculation



QM/MM methods

Combine
Quantum Chemistry
and
Molecular Mechanics

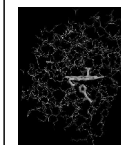
- + Accurate method
- + Objective choice
- Expensive
- One structure
- MM inaccurate

- Accurate **structure**
- Rather accurate energies
- Accurate properties
- Single-point solvation
 - No ΔG

Linear scaling methods

Usually semiempirical method
Whole system

- + Accurate inclusion of surroundings
- Expensive
- One structure
- Rather inaccurate method



- No geometry
- Enthalpies
- Accurate properties
- Explicit, fixed solvation (single point)
 - No ΔG

Solvation models

Discrete methods
Explicit solvent



- + Cheap
- + ΔG

Continuum methods

Born, Onsager, PCM, ...
Solute in continuum solvent
Dielectric constant (ϵ)



- Approximate
- ϵ often ill-defined

Lattice methods

Dipoles on a lattice
Langevin equation



- Solvation, including ΔG
- Correction to structures, energy, and properties

Poisson - Boltzmann

Areas of differing ϵ
Numerical solution

General Philosophy of QM/MM methods

A small system is studied by QM,
the rest by MM



Add QM and MM
Forces and Energies

Combines
QM accuracy
with
MM speed

Many Variants

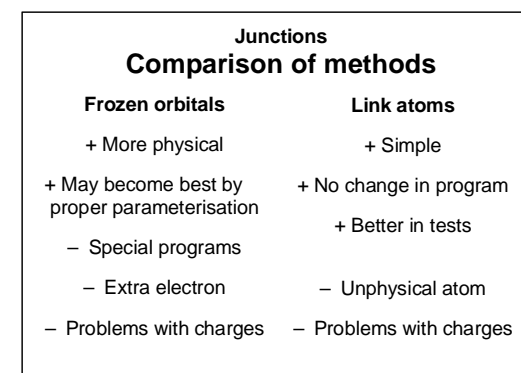
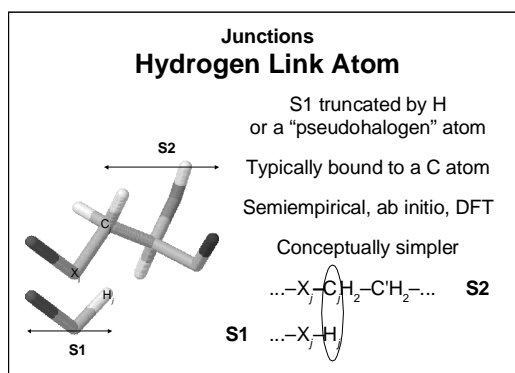
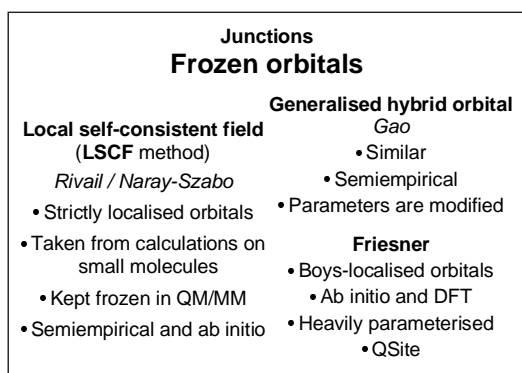
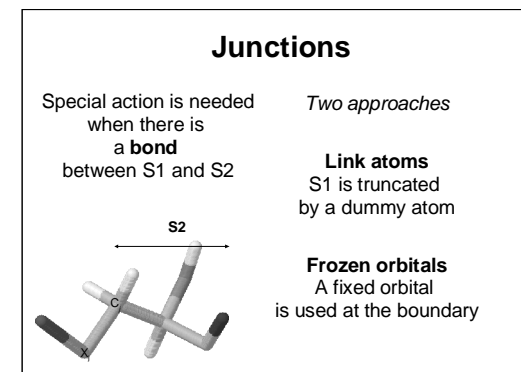
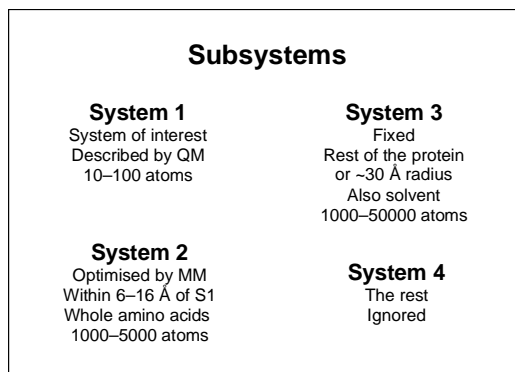
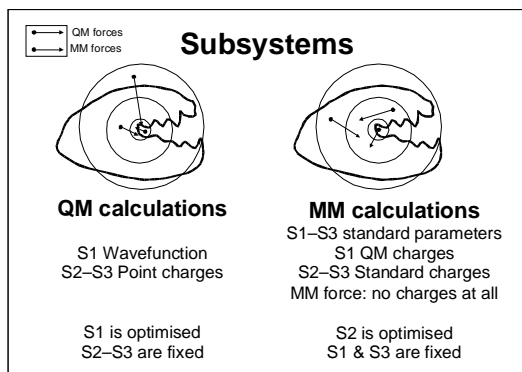
Warshel & Levitt, 1976
Semiempirical, no geometry

Singh & Kollman, 1986
Quest, ab initio, geometries

ComQum, 1992-1996

Software
Gaussian (Oniom)
CHARMM
Amber
QSite

Bash, Field & Karplus, 1987
Hillier, 1991
Vasilyev, 1991
Gao, 1992
Merz, 1993
Liu, 1994
Feller, 1994
Salahub, 1994
van Gunsteren, 1995
Morokuma, 1995
Brooks, 1996
...



Link-atom junctions Positions

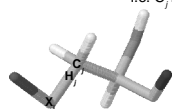
Junction carbon C_j

$$r_{Cj} = r_{Xj} + (r_{Hj} - r_{Xj}) r_{CX} / r_{HX}$$

r_{CX} is the C-X bond length in the MM library

r_{HX} is the optimal H-X bond length with the same QM method

i.e. C_j moves as a scaled image of H_j optimised by QM



Hydrogens bound to C_j
Determined in the MM step

Energy Function

In general

$$E_{tot} = E_{QM} + E_{MM} + E_{QM/MM}$$

E_{QM} = QM energy

E_{MM} = MM energy

$E_{QM/MM}$ = Energy of interface between QM and MM

COMQUM Energy Function

$$E_{tot} = E_{QM} + E_{MM123} - E_{MM1}$$

E_{QM} Standard QM energy of S1 with electrostatics of S2-S3

E_{MM1} energy of S1 with H junctions

E_{MM123} energy of S1-S3 with C junctions (no electrostatics)

Same philosophy as Oniom

↓
MM energy of S1 cancels
 E_{MM} corresponds to C junction

Forces

$$F_{tot} = F_{QM} + F_{MM123} - F_{MM1}$$

in principle, but

$$\nabla E_{tot}(r_1, r_{Hj}, r_{2-3}) = \nabla E_{QM}(r_1, r_{Hj}) + \nabla E_{MM123}(r_1, r_{Cj}, r_{2-3}) - \nabla E_{MM1}(r_1, r_{Hj})$$

so the formulae must be *corrected* using the *chain rule*

$$r_{Cj}(r_{Hj}) = r_{Xj} + k(r_{Hj} - r_{Xj})$$

Forces on the S2-S3 atoms are discarded

Charges

QM calculations

Point charges from MM library

Junction neighbours $q = 0$

Other charges in the amino acid are uniformly scaled (change < 0.01 e / atom)



MM calculations

S2-S3: standard MM charges

S1 ESP charges or Mulliken charges scaled to ESP

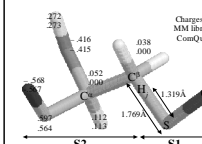
C_j charge adapted

⇒

- Charge transfer allowed
- q_{Cj} typical C-charge (not H)

Example Cys-C β junction

	QM	MM ₁	MM ₁₂₃	
Cu-S	x	(-x)	x	cancel exactly
Cu-S-H ₁	(x)	(-x)	x	cancel approximately
Cu-S-C β			x	
Cu-S-C β -H β			x	
Cu-S-C β -C γ			x	
S-H ₁		(x)	(-x)	cancel approximately
S-C β			x	
S-C β -H β			x	
S-C β -C γ			x	
C β -C γ			x	



$$r_{Cj} = r_{Xj} + k(r_{Hj} - r_{Xj})$$

$$dr_{Cj} = k dr_{Hj}$$

Forces

$F(\text{Cu}) = F_{QM}(\text{Cu}) + F_{MM1}(\text{Cu}) - F_{MM123}(\text{Cu})$
 $F(\text{S}) = F_{QM}(\text{S}) + F_{MM1}(\text{S}) - F_{MM123}(\text{S}) + (1-k)F_{MM123}(\text{C}\beta)$
 $F(\text{H}) = F_{QM}(\text{H}) + kF_{MM1}(\text{C}\beta) - F_{MM1}(\text{H})$
 $F(\text{C}\beta) = F_{MM1}(\text{C}\beta)$

Program flow

```

Evaluate QM wave function
Repeat
  Evaluate the QM forces (from S1-S3 onto S1)
  Evaluate the MM forces (from S1-S3 onto S1)
  Add the forces
  Relax the geometry of S1 using these forces
  Change the coordinates of S1 in MM representation
  If S2 is to be relaxed
    Calculate the QM charges of S1
    Insert them into the MM representation
    Relax S2 by MM minimisation with S1,S3 fixed
    Change the coordinates of S2 in QM representation
  Evaluate QM wave function (and energy) of S1
  Evaluate MM potential energy (S2-S1)
  Add the QM and MM energy
until convergence
    
```

QM program MM program Interface

Requirements

QM calculations

Many point charges
2 000 – 20 000

Forces available in a file

MM calculations

Forces available in a file

Charges can be changed

Possible to fix atoms

You do not have to change the QM and MM programs

Present ComQum Implementation

QM calculations

Turbomole

HF, DFT, MP2, ...

Shell-script driver
Modulated
Easy to add keywords

Interface: 6 small programs that transfer information

MM calculations

Amber 7.0

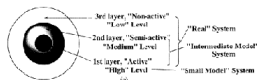
MM, MD, FEP, ...

Standard potential:
Harmonic bonds and angles
Trigonometric dihedrals
Point charges
Lennard-Jones

ONIOM

"Our own N-layered Integrated MO and MM approach"
Svensson, Morokuma, et al., J Phys Chem 100 (1996) 19357

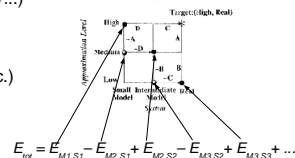
- Generalised method
- Any number of layers (CCSD(T)/B3LYP/MM/...)



- Any methods (not only QM/MM, but also QM/QM, etc.)

- In Gaussian-98

- No point charges at present



QM/MM literature

- Warshel A, Levitt M: J Mol Biol 1976, 103: 227
- Singh UC, Kollman PA: J Comput Chem 1986, 7: 718
- Monard G, Merz KM: Acc Chem Res 1999, 32: 904
- Orozco M, Luque FJ: Chem Rev 2000, 100:4187
- Wang W, Donini O, Reyes CM, Kollman PA: Annu Rev Biophys Biomol Struct 2001, 30: 211
- Gogonea V, Suárez D, van der Vaart A, Merz KM: Curr Opin Struct Biol 2001, 11:217
- Mulholland AJ: In Theoretical biochemistry – processes and properties of biological systems. Eriksson LA, ed. Elsevier; 2001: 597
- Gao J, Truhlar DG: Annu Rev Phys Chem 2002, 53:467
- Ryde U: Curr Opin Chem Biol 2003, in press.
- Ryde U: J Comput-Aided Mol Design, 1996, 10:153
- Field MJ: J Comput Chem 2002, 23:48
- Maseras F: Chem Commun 2000, 1821
- Åqvist J, Warshel A: Chem Rev 1993, 93:2523
- Villa J, Warshel A: J Phys Chem B 2001, 105:7887

QM/MM and related methods

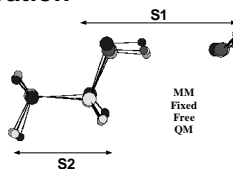
Lecture 2

Performance and Applications

1. Calibration
2. Is QM/MM needed?
3. Free Energies
4. Free or fixed surroundings
5. Electrostatics
6. Choice of QM system
7. Applications

Calibration

system	QC	ComQum		MM	
	(pm)	fixed	free	bond	rms
MeOH-OH ₂	205.2	2.3	-3.8	2.1	-0.2
HOH-OHMe	201.3	3.9	0.4	2.7	3.3
MeOH-OHMe	202.6	2.6	-2.9	3.1	0
EtOH-OH ₂	205.8	2.6	-1.1	7.1	-1.7
HOH-OHEt	200.9	2.4	2.7	13.7	-0.3
MeOH-NH ₂	211.1	11.5	-1.7	4.6	-0.7
H ₂ NH-OHMe	234.4	2.6	-0.2	3.3	-0.6
HSH-OHMe	227.5	1.9	1.2	1.9	1.6
HOH-SMe	245.8	1.5	2	2.1	2.4
Average with sign		3.5	1.8	4.5	1.2
			-0.4	+0.4	39.4



- No systematic error
- Much better than MM
- Free and Fixed Similar (depends on starting structure)

Calibration

Philipp & Friesner, J Comput Chem 20 (1999) 1468
Murphy, Philipp & Friesner Chem Phys Lett 321 (2000) 113

Ala ₄ conformational energies				QM/MM error (kJ/mole) in protonation energy of C ^δ in Leu Capitalised residue is the junction	
Energy (kJ/mole)	MP2	HF	HF/MM	MM	
	0.00	4.60	5.44	6.86	
Rotamer RMS deviations (kJ/mole) compared to MP2				Peptide Error	
B3LYP/MM	MM				
Phe	0.58	0.75			2 ^o Ala-LEU-Nma 2.43
His	4.60	4.52			2 ^o Ala-ALA-Leu-Nma 1.42
Asn	5.69	9.58			2 ^o Ala-ALA-Ala-Leu-Nma 0.79
Val	1.72	2.59			6 ^o Ala-LEU-Nma 4.14

□. Junction should be far from interesting system

The effect of surroundings

Is QM/MM really necessary?

In general yes

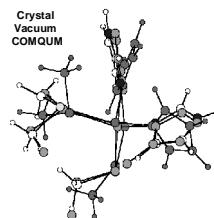
Enhances the H-abstraction step in cytochrome P450 by **10** orders of magnitude
H-bonds to the haem propionate side chains

Stabilises the binding of O₂ to haemerythrin by **57** kJ/mole

Pre-formed cavity and enhanced H-bonding

Reduces activation energy of phospholipase A₂ by **20** kJ/mole

Effect of surroundings Geometry

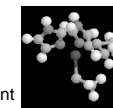


Reduced plastocyanin

- Dihedrals improved
- Orientations improved

Low-energy modes

Effect of surroundings Blue copper proteins



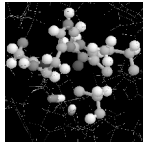
Improved bond lengths Cu(II)-S_{Met} excellent

Improved angles No change without bonds

Cu(I)-S_{Met} too long No change with crystals

System	Cu Protein	Fix Con	Distance to Cu (pm)				Angle around Cu (°)				φ (°)
			S _{Cu}	N	S _{aa}	N-N	S _{Cu} -N	S-S	S _{aa} -N		
I Vacuum			232	214-215	237	109	105-108	115	107-113	89	
	Pe red	No Yes	221	203-208	339	103	120-134	104	78-101	76	
II Vacuum			218	204	267	103	120-122	116	94-95	90	
	Crystal		211-217	203-239	287-291	91-118	110-141	99-114	85-110	74-80	
Pe ox	Yes	Yes	213	197-198	286	101	124	105	88-109	78	
	No	Yes	214	197-198	290	103	123-125	105	86-106	78	
Crystal			207-221	189-222	278-291	96-104	112-144	102-110	85-108	77-89	

Effect of surroundings Reorganisation Energy



Protein	Fix	Con	λ_{vac}	λ	
Plastocyanin red	Yes	Yes	23	4	27
	No	Yes	26	0	26
	No	No	-	-	-
Plastocyanin ox	Yes	Yes	19	19	38
	No	Yes	26	5	31
	Vacuum		33	29	62
Nitrite reductase	Yes	Yes	22	9	31
	No	Yes	29	5	34
	No	No	23		
Vacuum			38	40	78
Cucumber basic protein	Yes	Yes	18	8	27
	No	Yes	15	42	57
	No	No	3		
Azurin	Yes	Yes	15	13	28
	No	Yes	23	91	114
Vacuum			41	43	84

Unit: kJ/mole

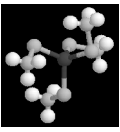
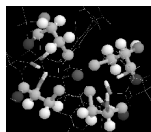
Changes by 24–56 kJ/mole

Half as much as in vacuum

Because of H-bonds to S_{Cys}

Covalent strain *increase* λ !

Effect of surroundings Rubredoxin

Fe	Protein	Protein	Fe-S _n	Reorg.	Strain
		Fixed?	pm	Energy	energy
II	Red	Yes	236.2	15.9	20.6
		No	236.7	6.7	18.1
	Ox	Yes	237.7	14.4	27.4
		No	238.5	19	37.2
	Vac.		242.3	21.4	
	Exp.			232	
III	Ox	Yes	231.0	8.2	25
		No	230.9	4	36.6
	Red	Yes	230.1	3.2	15.9
		No	230.2	16.1	14.1
	Vac.		232.4	18.3	
	Exp.			226	

Units: kJ/mole

Decreases Fe-S by 2 (Fe^{III}) to 5 (Fe^I) pm

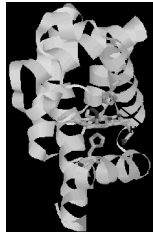
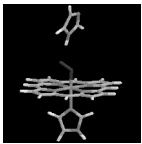
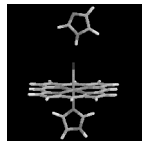
Reduce reorganisation energy by ~20 kJ/mole

More similar to experiments

- NH-S hydrogen bonds
- Negative charge of Fe(SCH₃)₄^{-1/2}

Not always large effect of surroundings O₂ and CO binding to Myoglobin

Difference in the H-bond strength between His64 and O₂ or CO

21 kJ/mole in vacuum

24 kJ/mole in vacuum with **constraints**

21–22 kJ/mole with QM/MM

The effect of surroundings Is QM/MM really necessary?

Depends on the protein and the property of interest

Sometimes accurate results can be obtained *once some crucial residues are included*

But which?

Can only be decided in an *unbiased* way by QM/MM

Strain Energy

Can be calculated as

$$\Delta E_{S1} = E_{QM;S1}(QM/MM) - E_{QM;S1}(vacuum)$$

Experimental geometry better, but normally not accurate enough

Estimates how much the protein changes the site
Stabilises alternative configuration

Problematic

Strain Energy

$\Delta E_{S1} = E_{QM;S1}(QM/MM) - E_{QM;S1}(vacuum)$

Ryde, *Rec Res Devel Prot Engin*, 2 (2002) 65

Varies a lot (kJ/mole)

35–172 Alcohol dehydrogenase

33–66 Blue copper proteins

14–110 Fe-S cluster proteins

10–21 Ferrochelatase

31–83 Myoglobin

48–197 MMO and RNR

Covalent strain

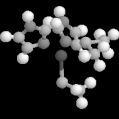
33 Alcohol dehydrogenase

21–29 Blue copper proteins

Depends on

- Size and
- Charge of QM system
- Polar groups
- Reference state vacuum or solvent
- Which interactions bonds, angles, dihedrals, and repulsive v d Waals

Strain Energies Blue copper proteins



System	Fix		Con		ΔE_{str} (kJ/mole)
	Protein	Pe red	Pe ox	Nir	
I	Yes	Yes	Yes	Yes	30.6
	No	Yes	Yes	Yes	44.7
	No	No	No	No	21.3
II	Pe ox	No	Yes	Yes	46.8
	Pe red	Yes	Yes	Yes	36.1
	Pe red	No	Yes	Yes	34.7
III	Pe ox	Yes	Yes	Yes	36.0
	Pe ox	No	Yes	Yes	32.6
	Nir	Yes	Yes	Yes	40.7
II	Nir	No	Yes	Yes	40.6
	CBP	No	No	Yes	19.7
	CBP	No	Yes	No	15.5
I	Azarin	Yes	Yes	Yes	60.6
	Azarin	No	Yes	Yes	47.2
	Azarin	Yes	Yes	Yes	39.2
II	Azarin	No	Yes	Yes	33.8
	Azarin	Yes	Yes	Yes	33.8

33–61 kJ/mole

Lower with relaxed protein
(as expected)

16–21 kJ/mole without bonds
↓
Strain \approx 20 kJ/mole

Also this includes non-strain terms



Problems

Energies unstable

Local-minima problem

Many local minima

A H-bond far away may change the energy by 20 kJ/mole

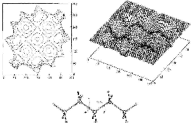




Solutions

Run forth and back until convergence

Small radius for S2

Include dynamics entropy ΔG

How to obtain Free Energies

Free Energy Perturbation

$$\Delta G = -k_B T \ln \exp \frac{-H_1 - H_0}{k_B T}$$

Can be based on QM/MM structures (kept fixed)

Thermodynamic integration

$$\Delta G = G_1 - G_0 = \int_0^1 \delta \frac{H_\lambda}{\delta \lambda} d\lambda$$

or Integrated by QM/MM (Quantum dynamics)

A system is changed slowly to another via unphysical mixed system

Normally by MD or MC

- Expensive
- Tedious
- Hard to converge

Jorgensen, J Am Chem Soc 107 (1985) 154
Kollman, Acc Chem Res 34 (2001) 72
Yang, J Am Chem Soc 122 (2000) 6560
Flekt, J Am Chem Soc 122 (2000) 7688

How to obtain Free Energies Simplified methods

Linear free energy relationships

Supermolecular approach
Entropy from frequencies
Solvation from continuum model

Poisson – Boltzmann method

Langevin dipoles on a grid

Empirical Valence Bond method

Free Energy example MM/PBSA

$$G = E_{MM} + PBSA / GBSA - TS_{MM}$$

- Averages over short MD simulation
- MM strain energy of protein and ligand
- Entropy from MM frequency calculation
 - Solvation effects by Poisson–Boltzmann or Generalised Born model
- Hydrophobic effects from surface area

Free vs fixed protein

Effect

Geometry: small
 E_{QM} moderate (<20 kJ/mole)
 E_{MM} large (>500 kJ/mole)

Cost

Flexible S2
about 3 times more expensive than fixed S2

Perhaps better to relax S2 only a few times
?

Free vs fixed protein

Recent results indicate that MM is quite inaccurate
Ryde, et al. J Comput Chem, 23 (2002) 1058
 Titmus, et al. Chem Pys Lett, 320 (2000) 169

S2 diverges from crystal structure

⇒

Relax as little as possible

Electrostatic interactions

<p>Type A No electrostatics between QM and MM Works for organic models (strain) ONIOM, IMMOM</p>	<p>Type C Also polarisation in MM Polarisabilities in MM First force fields available Amber, Tinker, ...</p>
<p>Type B QM is polarised by MM Point-charge model in QM Most common Most important term included</p>	<p>Type D QM and MM polarises each other self-consistently Dalton New QM code</p>

Choice of QM system

Rules of thumb

- As large as possible
- Do not break double bonds or conjugation
- Include at least a CH₃ group around polar atoms
- Break only C–C bonds (no charge transfer)
 - Include all metal ligands
- Include H-bond partners, if possible

Offset forces

Fogarasi & Pulay J. Am. Chem. Soc. 114 (1992) 8191

Simple way to correct systematic errors

1. Calculated geometry with accurate method or use experimental data
2. Calculate forces at this geometry with standard method
3. Subtract these forces from all later forces

Applications

Protein reaction

Catechol O-methyl transferase
 Substrate – Product
 Three frozen states
 230 – 213 – 195 ps
 1. s. later than in vacuum

Preorientation

Vacuum ↓

ComQum ↓

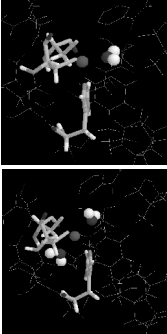
Applications

Coordination number of Zn in Alcohol Dehydrogenase

Ryde, J. Comput.-Aided Mol. Design, 10 (1996) 153

Ligand	Rel. Energy (kJ/mole)	Strain Energy
H ₂ O	76	42 144
(H ₂ O) ₂	128	63 131
OH	84	61 172

Five-coordination possible
Disfavoured by protein
Confirmed later



PAC on Cd-alcohol dehydrogenase

Perturbed angular correlation of γ rays
Ryde & Hemmingsen, J. Inorg. Biol Chem. 2 (1997) 567

Field gradient measured, but what is the structure?

	Exp V_{zz}	Error	Ligand
1	1.56	0.35	H ₂ O
2	0.74	0.01	OH ⁻
3	1.03	0.15	Glu-68
4	0.7	0.03	OH ⁻
5	1.88	0.07	Glu-68
6	1.92	0.17	H ₂ O
7	1.99	0.1	H ₂ O

Glu-68 coordinates to Zn

