



ACCURATE PROTEIN–LIGAND INTERACTION ENERGIES BY COMBINING CLASSICAL AND FRAGMENT MODELS

Pär Söderhjelm, Francesco Aquilante, and Ulf Ryde

par.soderhjelm@teokem.lu.se; http://www.teokem.lu.se/~per

Department of Theoretical Chemistry, Lund University, Sweden

LUND UNIVERSITY



1. Objective

We propose a new method, *polarizable multipole interaction with supermolecular pairs* (PMISP), for approximating the quantum-mechanical interaction energy of a protein–ligand system. The method is designed for second-order perturbation theory (MP2) or coupled cluster (CC) theory with large basis sets. The method will be used in conjunction with an implicit solvent model and molecular dynamics sampling (the MM-PBSA approach) to improve the accuracy of binding affinity predictions.

2. Interaction energies for large systems: Previous approaches

I. Classical modeling

Estimate the electrostatic, induction, dispersion, and repulsion contributions separately.

Examples: NEMO (Karlström et al.), EFP (Gordon et al.), PFF (Friesner et al.), AMOEBA (Ponder et al.), SIBFA (Gresh et al.)

Problem: Transferability of the dispersion and repulsion terms.

II. Fragmentation

Divide the protein into smaller fragments whose interaction energy with the ligand can be accurately computed by quantum chemistry.

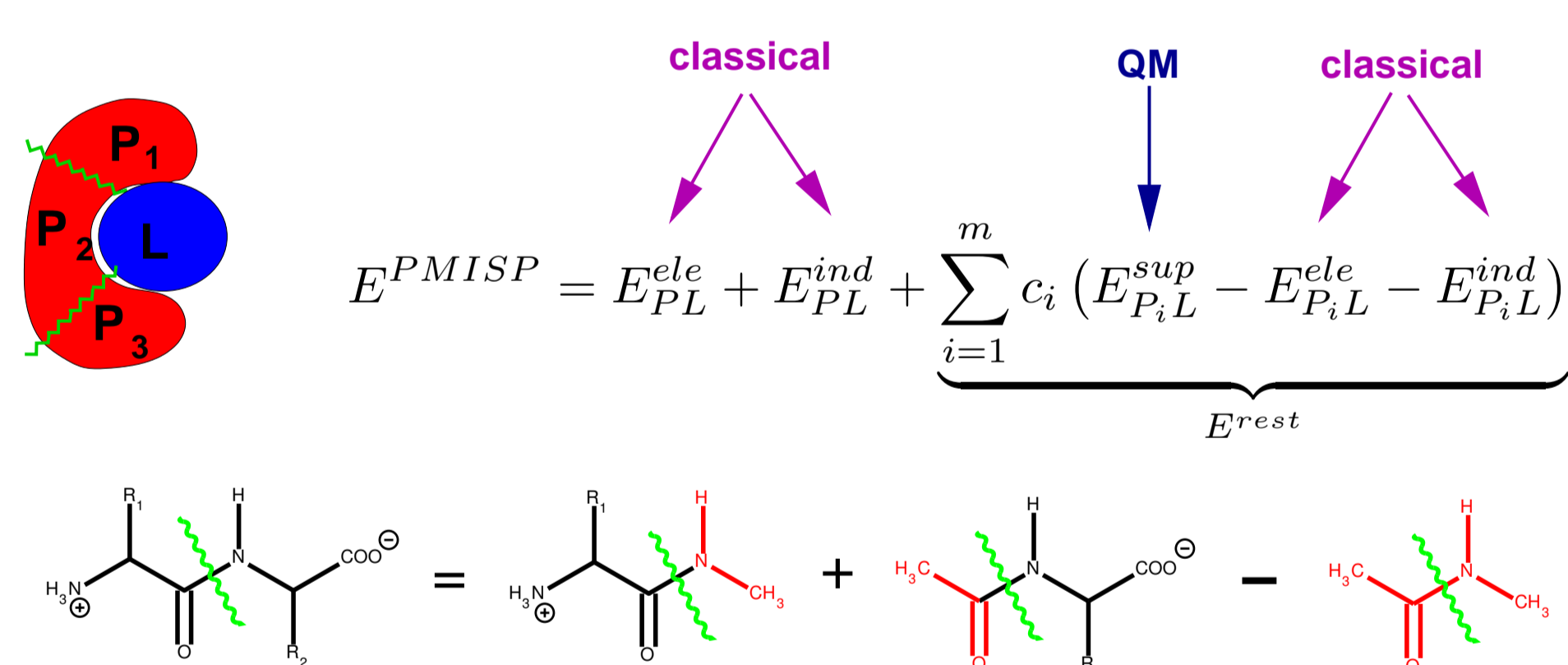
Examples: FMO (Kitaura et al.), MFCC (Zhang et al.), EE-PA (Truhlar et al.)

Problem: How to treat many-body effects without calculating trimers?

3. PMISP: Combine the two approaches

Our idea: Use the classical approach for the electrostatic and induction terms but the MFCC fragmentation approach for the rest term (mainly dispersion and exchange–repulsion).

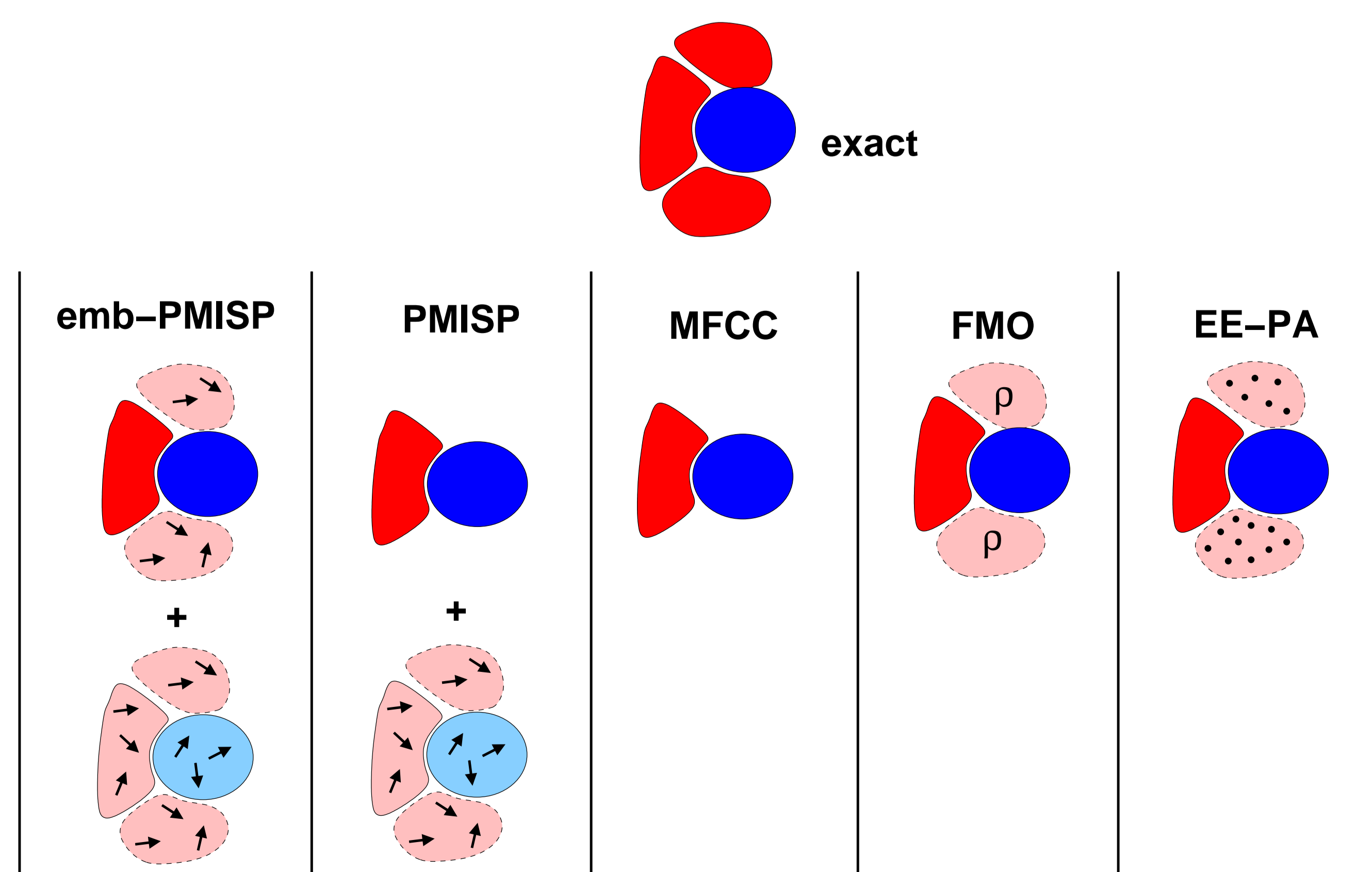
Assumption: The rest term is pairwise additive, i.e. all many-body effects reside in the induction term.



Two types of QM calculations needed:

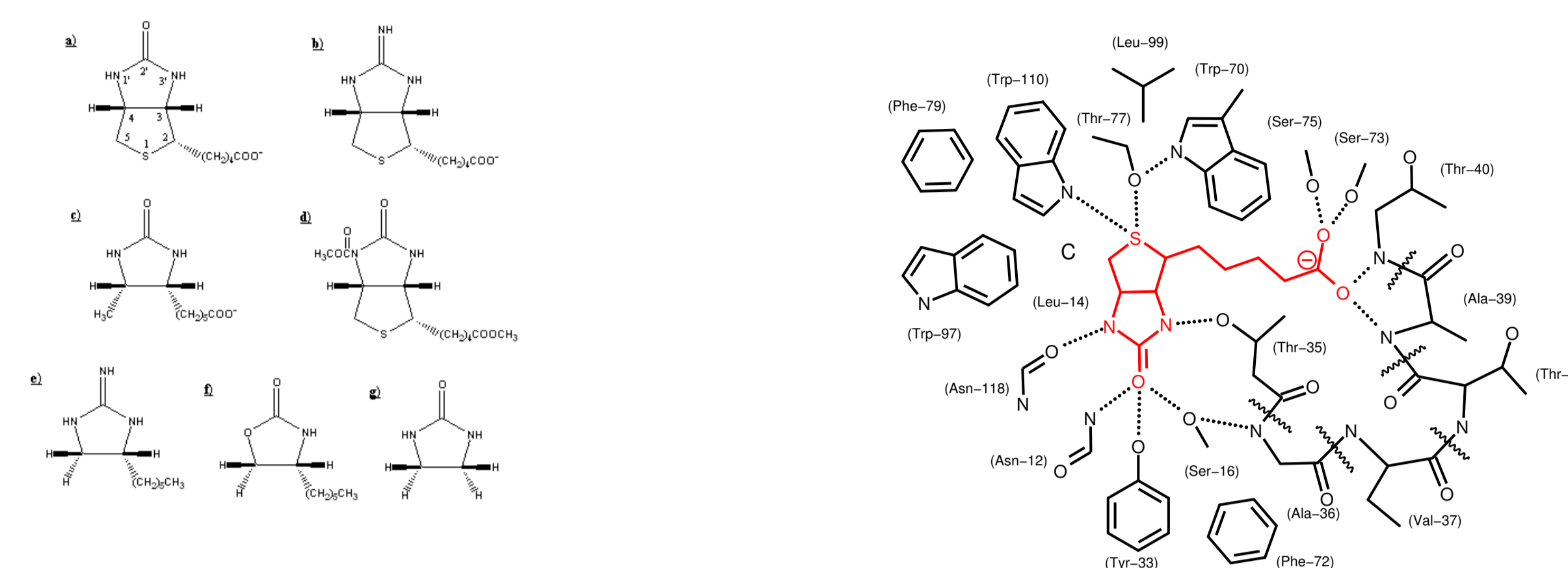
- Properties (multipoles and polarizabilities) for L and each P_i fragment
- Counterpoise-corrected supermolecular interaction energy for each P_i – L dimer

4. PMISP versus other methods



5. Results: Test case

Seven ligands (a = biotin) interacting with a model of avidin



Error (\pm geometrical variation) in kJ/mol relative to the QM reference.

Method	Ligand						
	a ⁻	b ⁻	c ⁻	d	e	f	g
QM reference (HF/6-31G*)	-266 \pm 22	-269	-227	-32	-3	-47	-63
QM reference (MP2/6-31G*)	-422 \pm 21	-433	-372	-184	-125	-148	-119
PMISP (HF)	-10.0 \pm 2.5	-11.3	-10.2	+1.3	+4.9	-1.0	-1.6
PMISP (MP2)	-9.0 \pm 3.3	-9.0	-12.6	+7.2	+9.1	+2.4	+1.5
MFCC (HF)	-28.3 \pm 7.0	-27.2	-42.0	+9.9	+18.2	+8.6	+7.1
FMO (HF)	+14.4 \pm 2.4	+10.5	-0.6	+7.5	+4.8	+5.4	+6.8
EE-PA (HF)	+16.4 \pm 2.2	+13.0	+2.4	+8.7	+5.0	+6.9	+7.2
emb-PMISP (HF)	+2.0 \pm 2.6	+5.9	+6.5	+2.9	+9.0	+5.0	+5.3

- Important to include many-body effects (MFCC significantly worse).
- Exact electrostatic embedding (FMO) *not* better than classical modelling of many-body effects (PMISP). Possible reason: neglect of Pauli effects.
- PMISP error \sim 10 kJ/mol for the charged ligands (a–c), smaller for the neutral ones (d–g). The error is due to neglect of coupling between induction and repulsion and can be reduced by embedding (emb-PMISP).
- MP2 dispersion nearly pairwise additive (PMISP error not higher for MP2).
- Error from the fragmentation procedure negligible (1.1 kJ/mol for electrostatic energy of the 25-fragment model)
- High multipole level (charges, dipoles, and quadrupoles) and anisotropic polarizabilities essential for good results.

6. Results: Full avidin–biotin system

System size: 7 832 atoms. E^{rest} contributions from distant fragments (>4 Å from L) are estimated by the Van der Waals term of a standard molecular mechanics force-field.

	PMISP/MM (MP2)		AMBER	
	6-31G*	aug-cc-pVTZ	ff94	ff02
E^{ele}	-1120	-1126	-1300	-1096
E^{ind}	-190	-285		-113
E^{rest}	49	-9	-143	-143
E_{tot}	-1261	-1419	-1443	-1353

- Protein–ligand interaction energies with 6-31G* are useless
- Error from long-range MM estimate (\sim 10 kJ/mol) similar to the inherent PMISP error.
- MP2 properties can be replaced by B3LYP properties without loss of accuracy.
- Properties computed with small basis set (6-31G*) introduce an error of \sim 10 kJ/mol
- The two AMBER force fields differ significantly; both give large error.
- Computationally, PMISP is intermediate between QM and MM. Using the Cholesky decomposition technique for the MP2 calculations, the total CPU time is \sim 10 CPU-days (the corresponding QM calculation would have 270 974 basis functions).

7. Conclusion

Our method (PMISP) uses quantum chemistry for short-range interactions and therefore corresponds to a polarizable force-field with perfect parameters. It estimates the interaction energy of a protein–ligand system with an accuracy of \sim 15 kJ/mol or better. Further improvement requires inclusion of coupling between induction and exchange–repulsion.