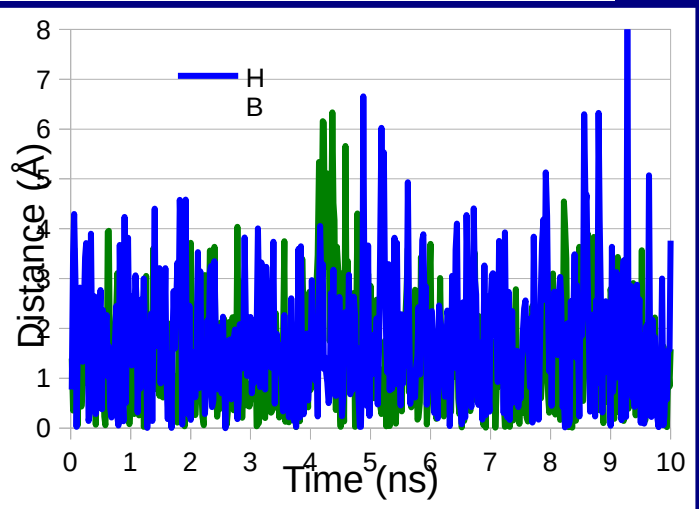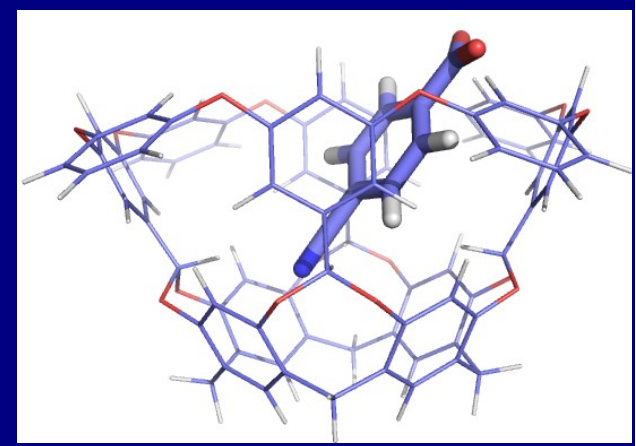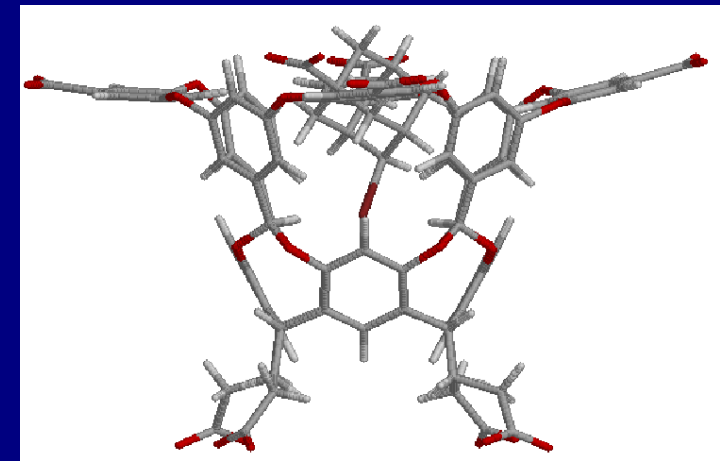# Octa-Acid Binding Affinities with QM Methods
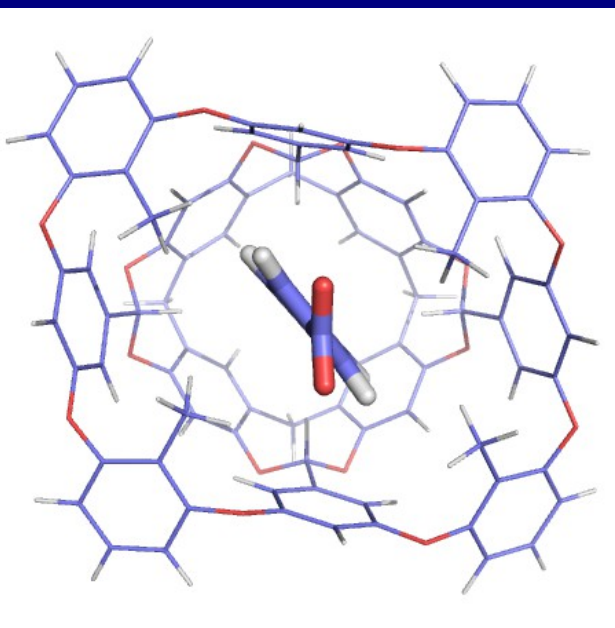
Ulf Ryde
Theoretical Chemistry
Lund University
Sweden

1. Intro
2. DFT-D3
3. CCSD(T)
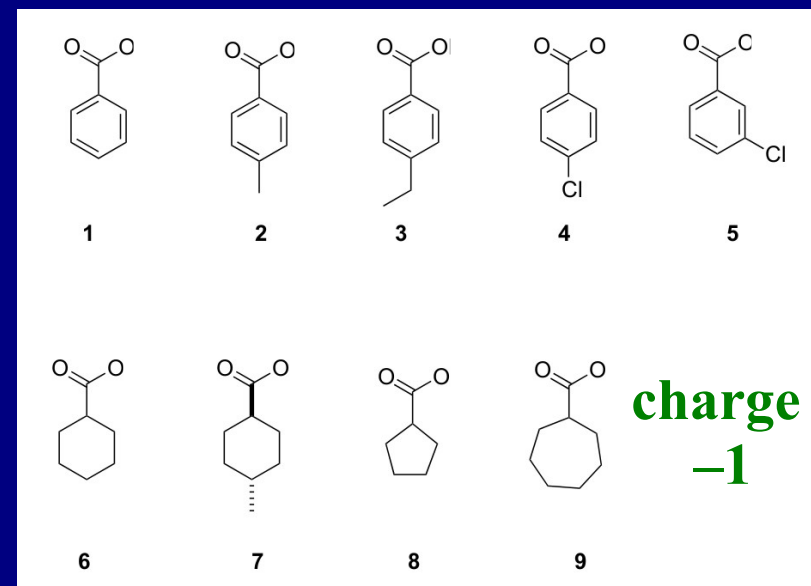4. MD snaps

1. Introduction

2. DFT Optimisation

3. DLPNO-CCSD(T)

4. MD snapshots

Results from SAMPL4

- FEP with MM
RESP or AM1-BCC charges

- QM/MM-FEP with NBB or ssEA

- DFT on optimised structures

- LCCSD(T0)-PMISP
on DFT structures

- Varying results

charge
−1

	MAD	R^2	τ_r
RESP	4	0.8	1.0
AM1-BCC	4	0.8	1.0
QM-FEP	26	0.4	-0.3
DFT opt	6	0.8	0.7
LCCSD(T0)	24	0.4	0.3

SAMPL5

- Gusts more dissimilar

- Net charge –1 or +1

$\Rightarrow$

Problems for FEP

not for DFT
continuum solvation
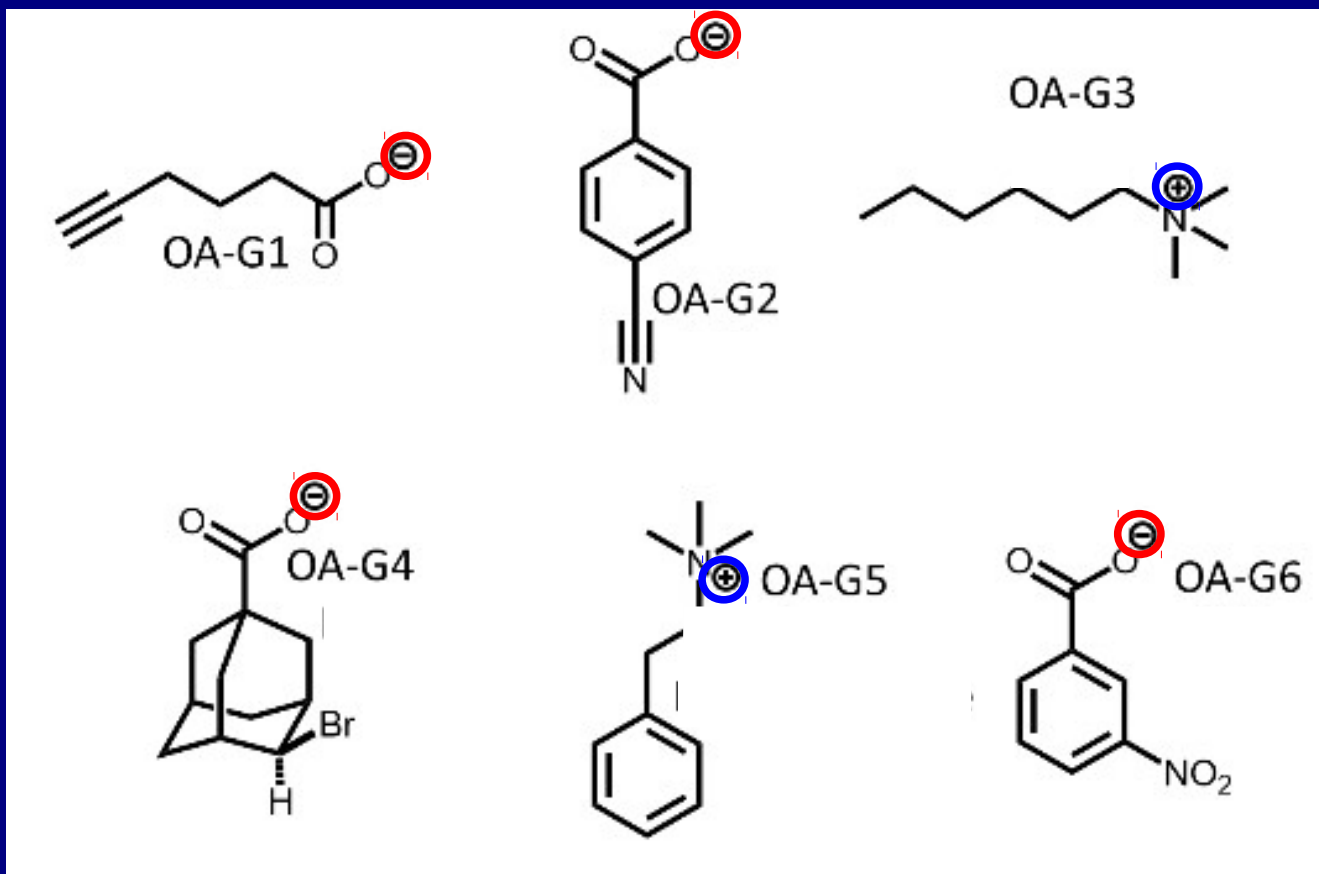
	MAD
RESP	4
AM1-BCC	4
QM-FEP	26
DFT opt	6
LCCSD(T0)	24

Philosophy

- Ligands not proper for FEP
  Dissimilar ligands
  Varying net charge

- In SAMPL4 DFT-opt gave MAD 6–9 kJ/mol

- Might be good enough
  if FEP does not work

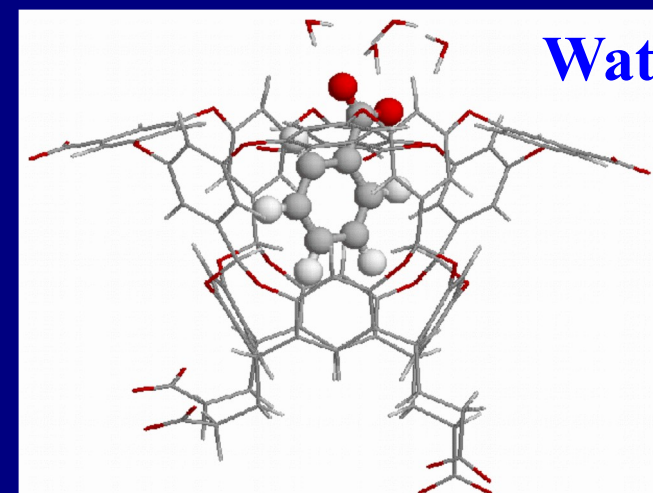- *Improve* the DFT approach
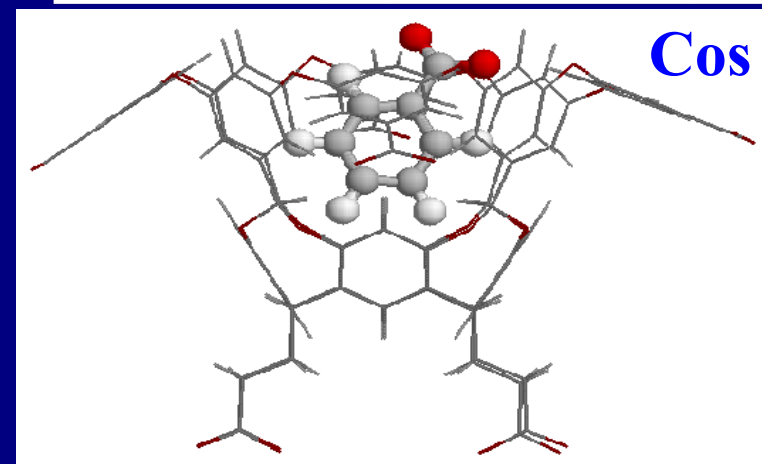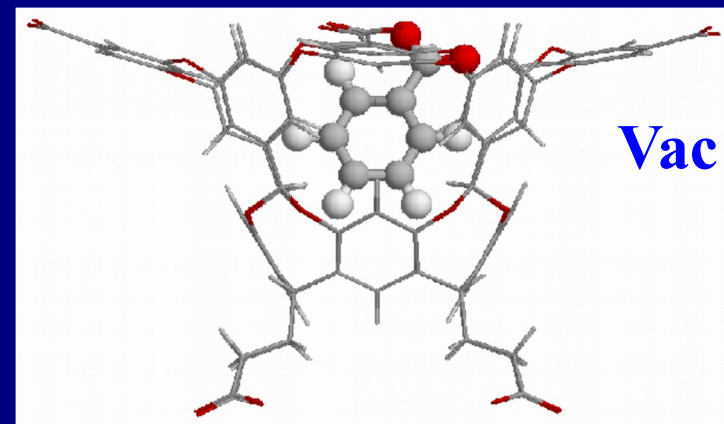 with gained *experience*

Improvements

- Minimise effect of flexibility

- Reduce effects of negative charge

- Try MD sampling

- Improve CCSD(T)
with DLPNO-CCSD(T)
(domain-based local pair natural orbital)
no fractionation needed
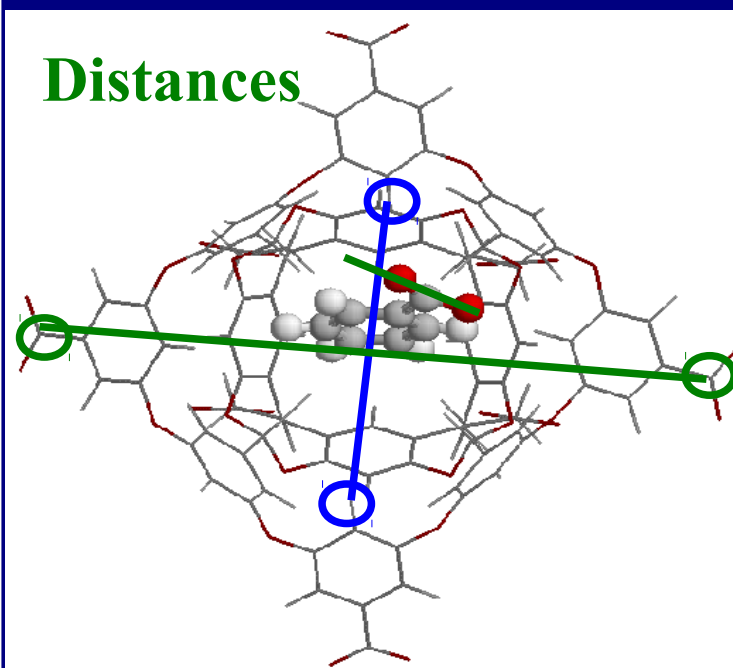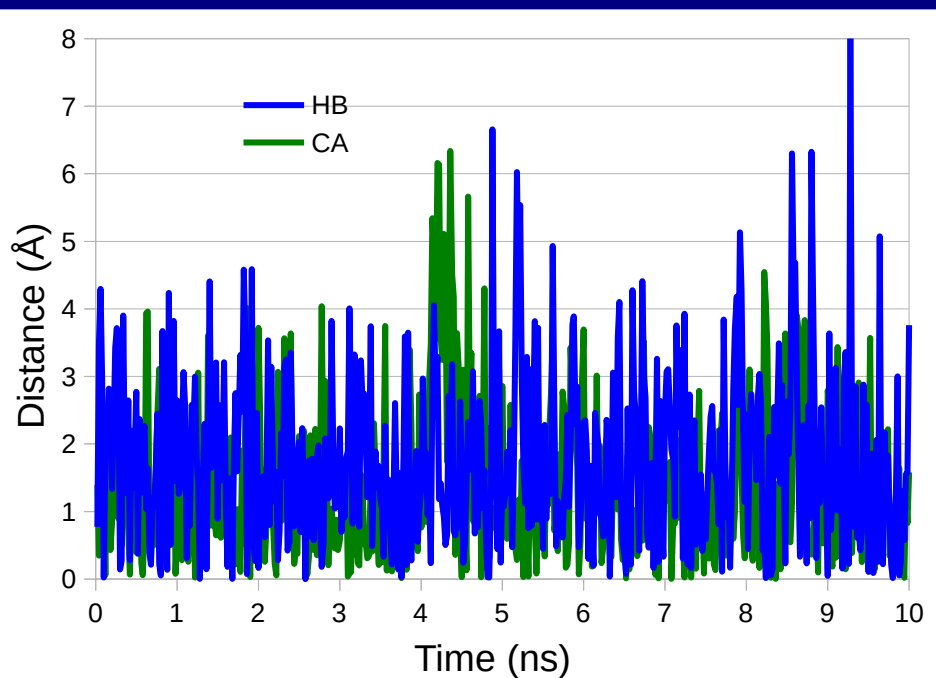
DFT-opt in SAMPL4

Three approaches

• Opt in vacuum

• Opt in COSMO continuum

• COSMO + 4 water

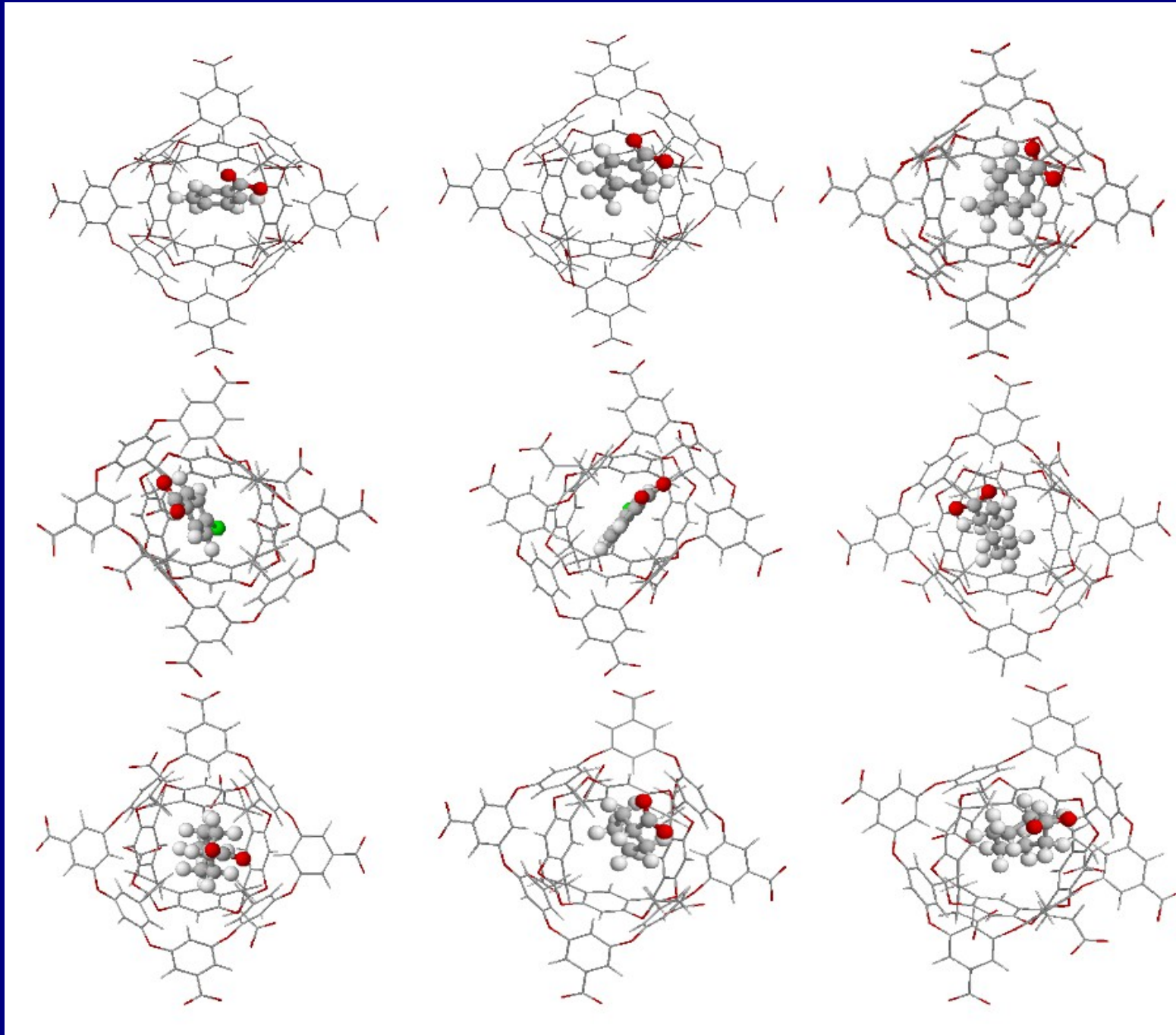|	Vac	Cos	Wat	Cons
MAD	7	9	9	6
MADtr	5	8	9	6
R^2	0.8	0.7	0.6	0.8
τ	0.7	0.7	0.7	0.7

Flexibility of host from MD

- Breathing motion

- Extensive fluctuation 8 Å

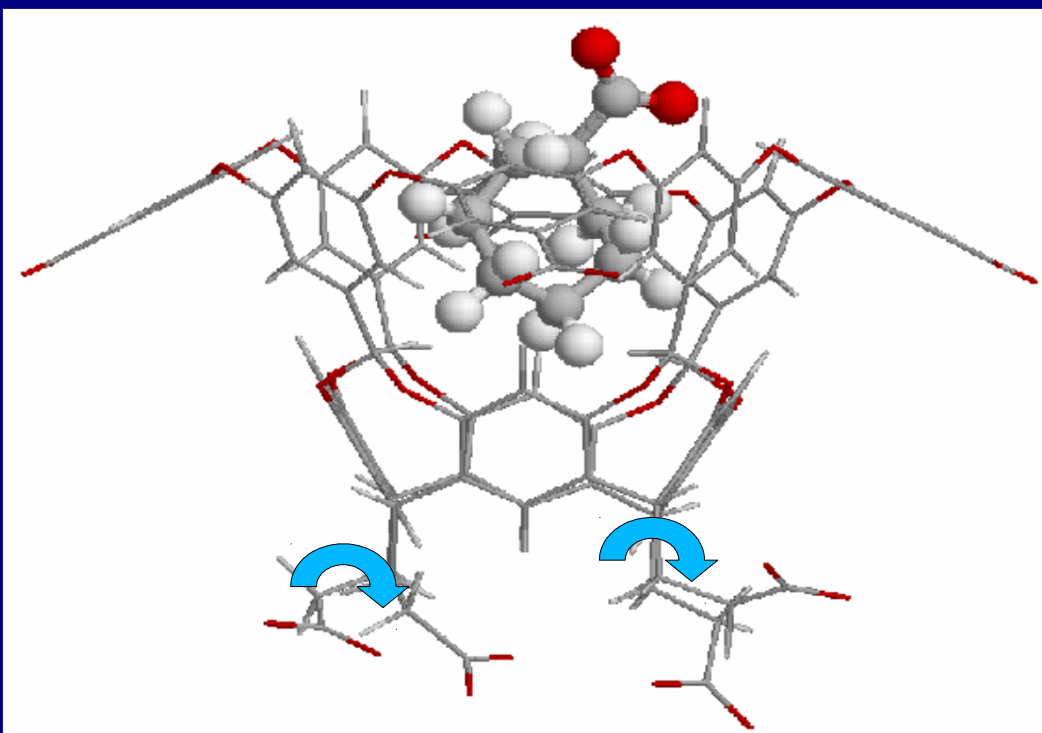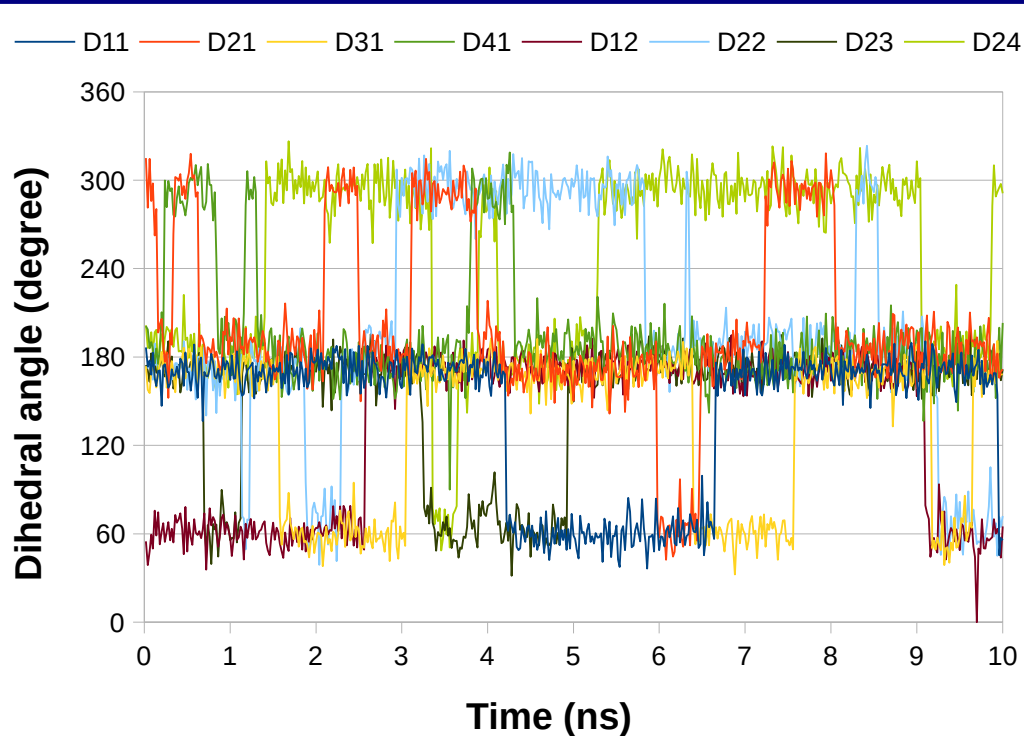- Rapid dynamics

- No difference between guests in average

	HB	CC
Bz	11.7	18.2
MeBz	11.7	18.4
EtBz	11.8	18.5
pClBz	11.6	18.4
mClBz	11.6	18.2
Hx	11.7	18.5
MeHx	11.7	18.4
Pen	11.7	18.6
Hep	11.8	18.4

Problem for optimised structures

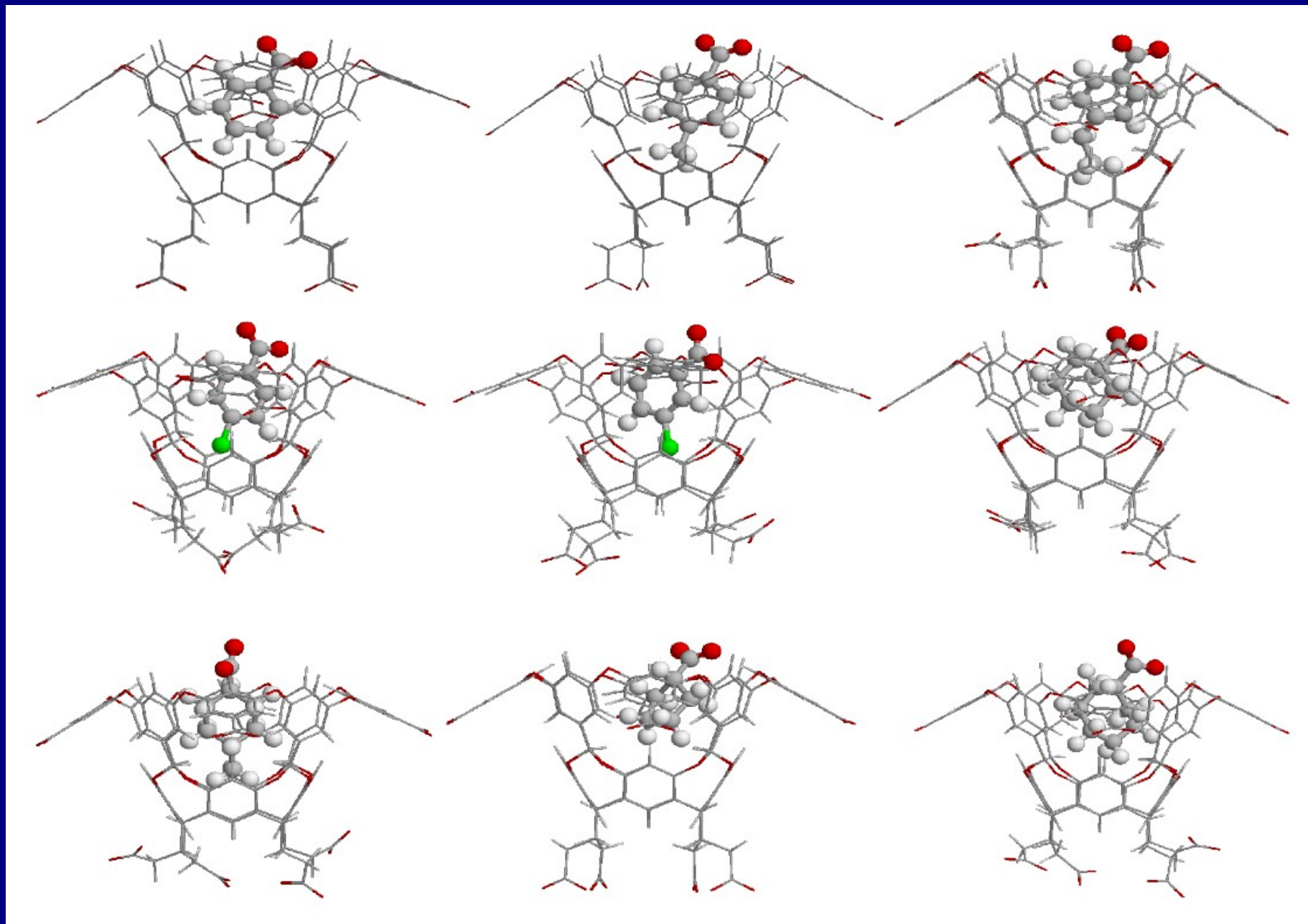# Movement of Propionate Groups

- Extensive dynamics with 3 minima

- All 8 torsions change
on a time-scale of 0.1–1.4 ns^{-1}

- All 3 minima visited, but not for all propionates
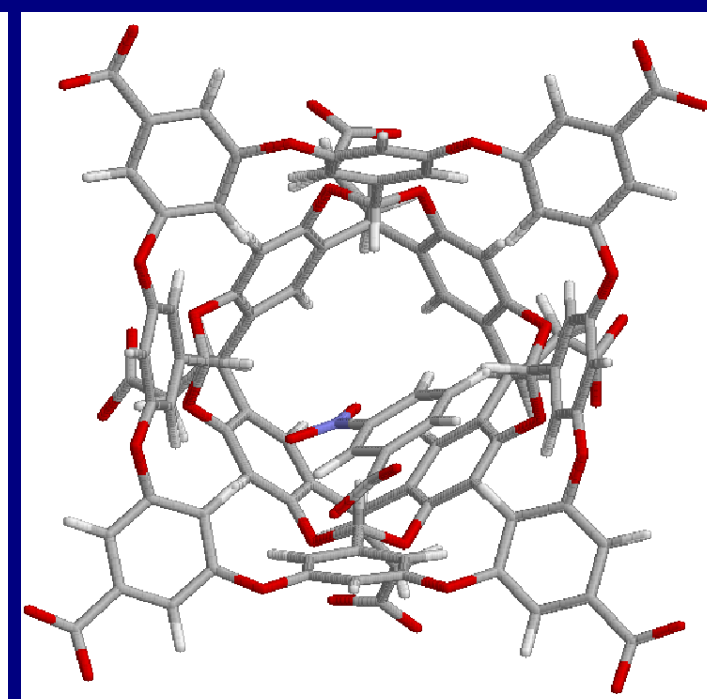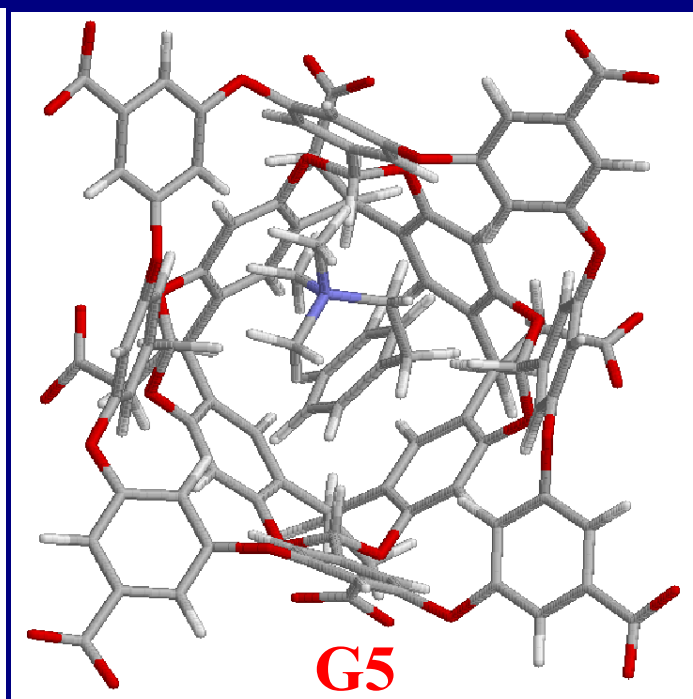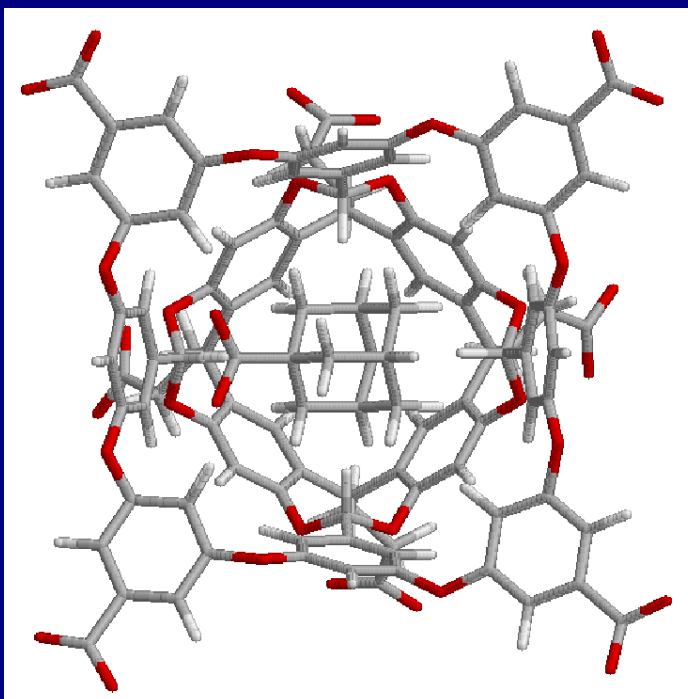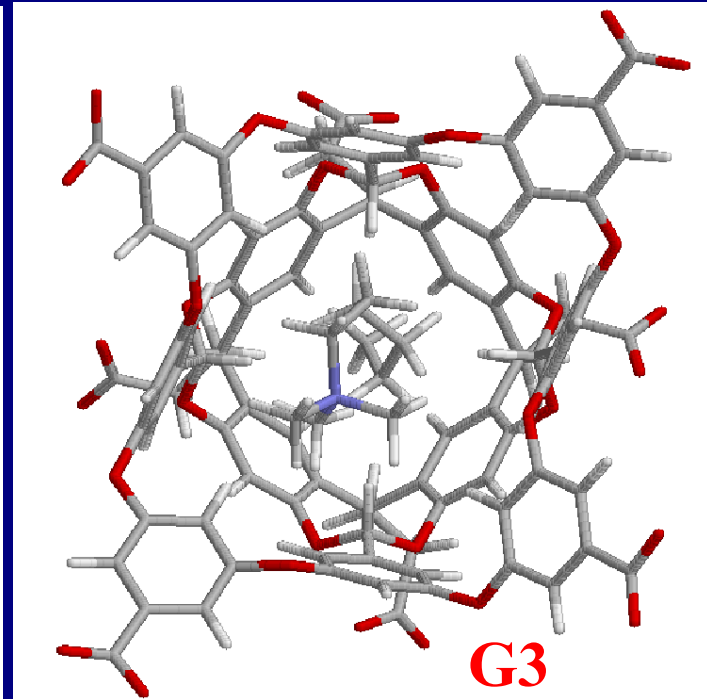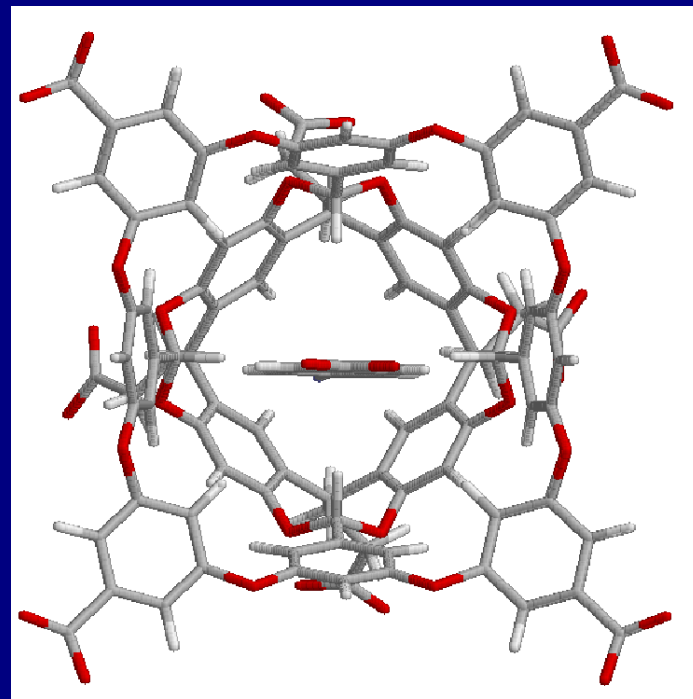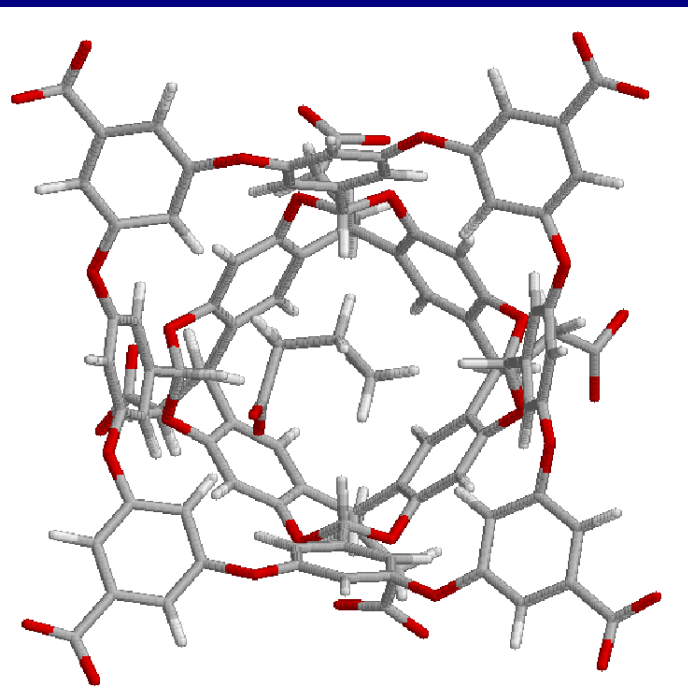
Problem also for optimised structures
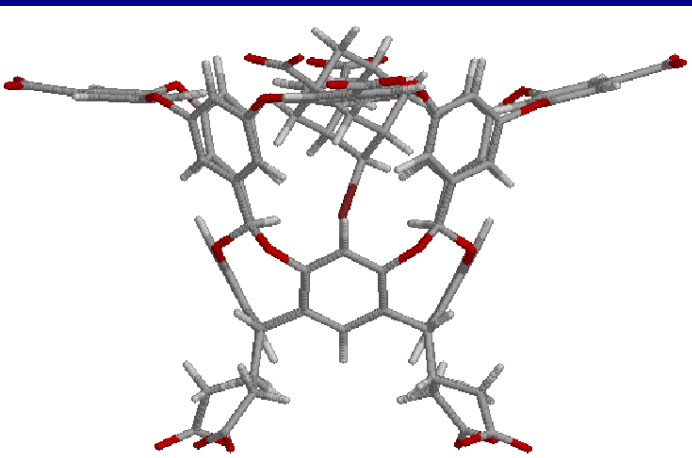

• Less for vacuum optimisations

Attempt 1
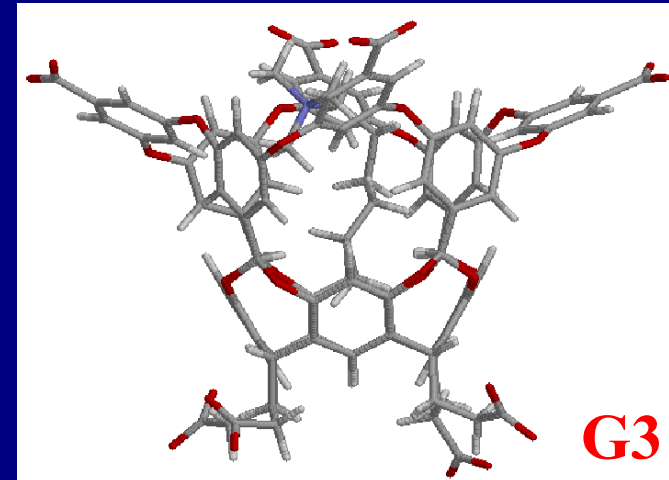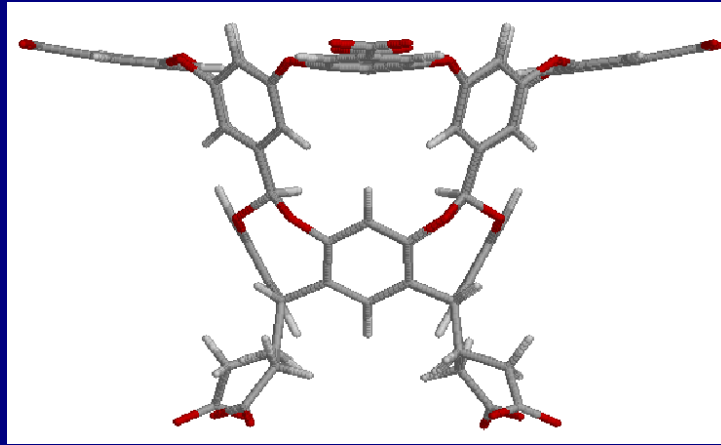
- Start from C_4 symmetric host

  - Keep all structures as symmetric and similar as possible

  - Use vacuum optimisation

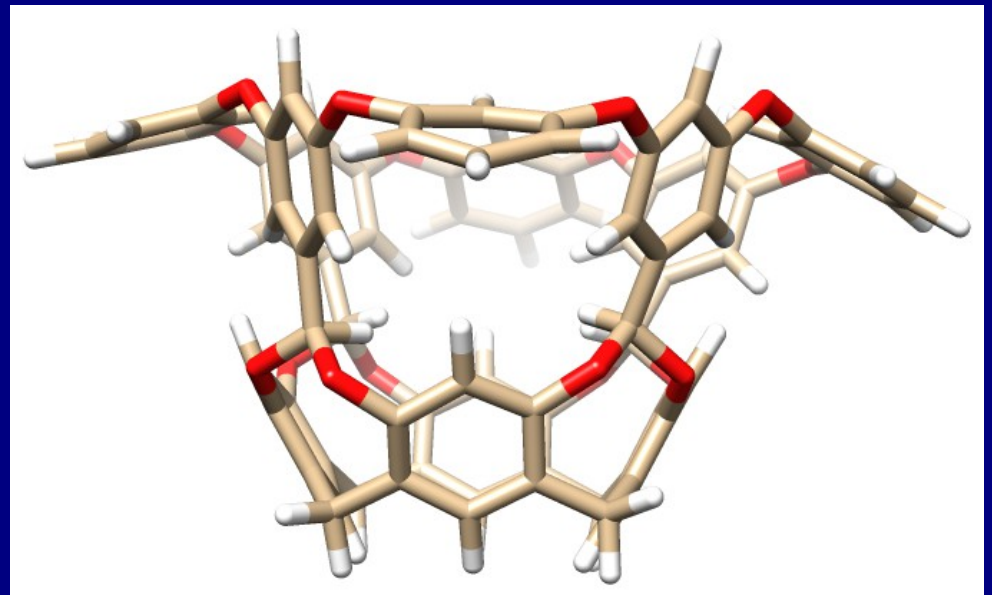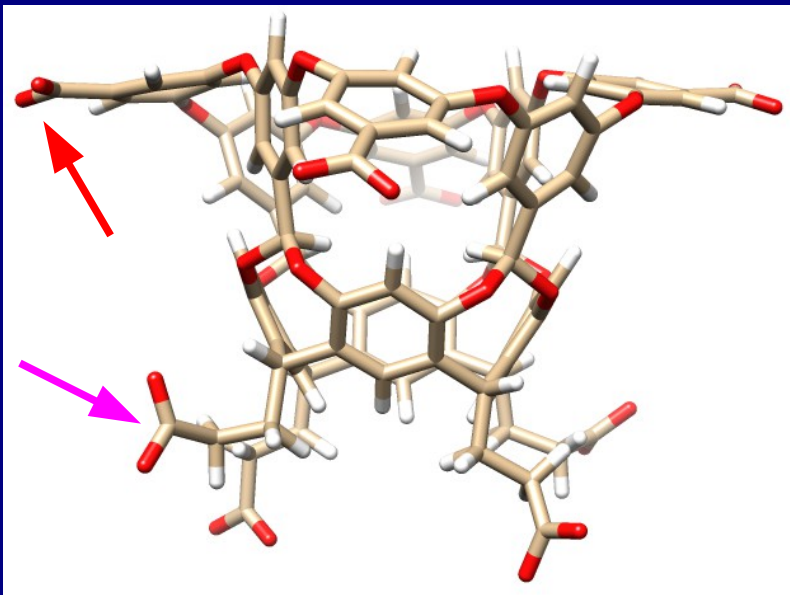|	Vac	Cos	Wat	Cons
MAD	7	9	9	6
MADtr	5	8	9	6
R^2	0.8	0.7	0.6	0.8
τ	0.7	0.7	0.7	0.7

Partly Successful

Partly Successful
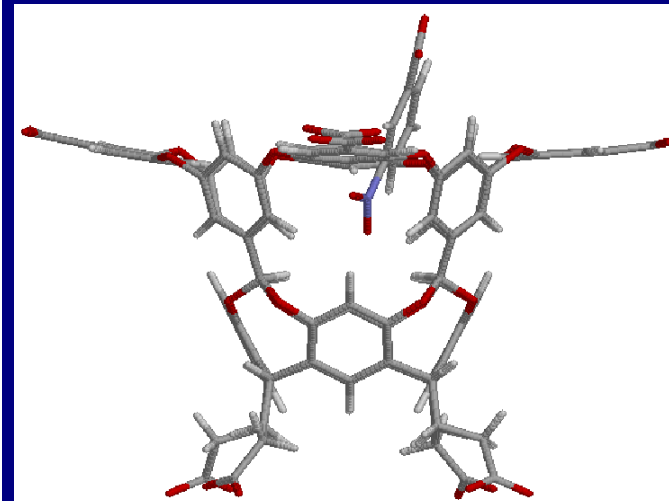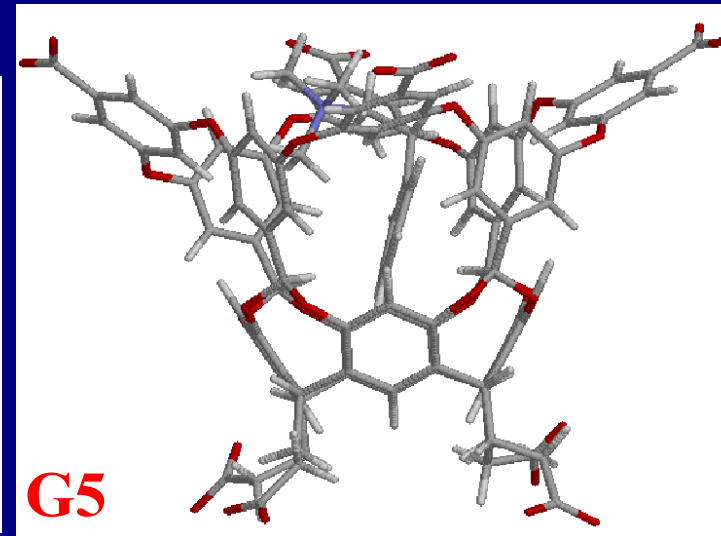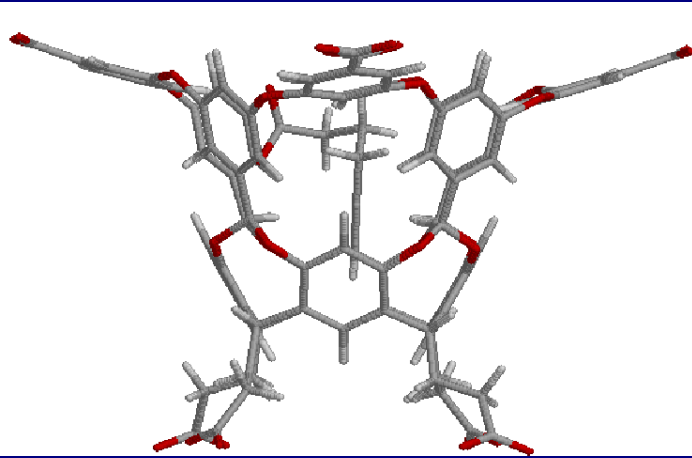
- G3 and G5 (with $-NMe_3^+$) groups still distorted

Attempt 2

• Remove all 4 propionate and 4 bezoate groups

• Reduces solvation energy
from −6600 to −300 kJ/mol

• Reduced flexibility

• Coulombic ($\varepsilon = 80$) correction for +/− series
(~23 kJ/mol)

Minimal effect at MM level

- Identical results for 5 perturbations

- 2 kJ/mol difference for 3 perturbations

- The differences come from propionate groups

- Nearly same performance compared to Exp.

	OA	NOA	HOA
MAD	3.6±0.2	4.1±0.1	3.9±0.2
R^2	0.84±0.04	0.79±0.03	0.81±0.04
τ_r	1.00±0.00	1.00±0.00	1.00±0.00
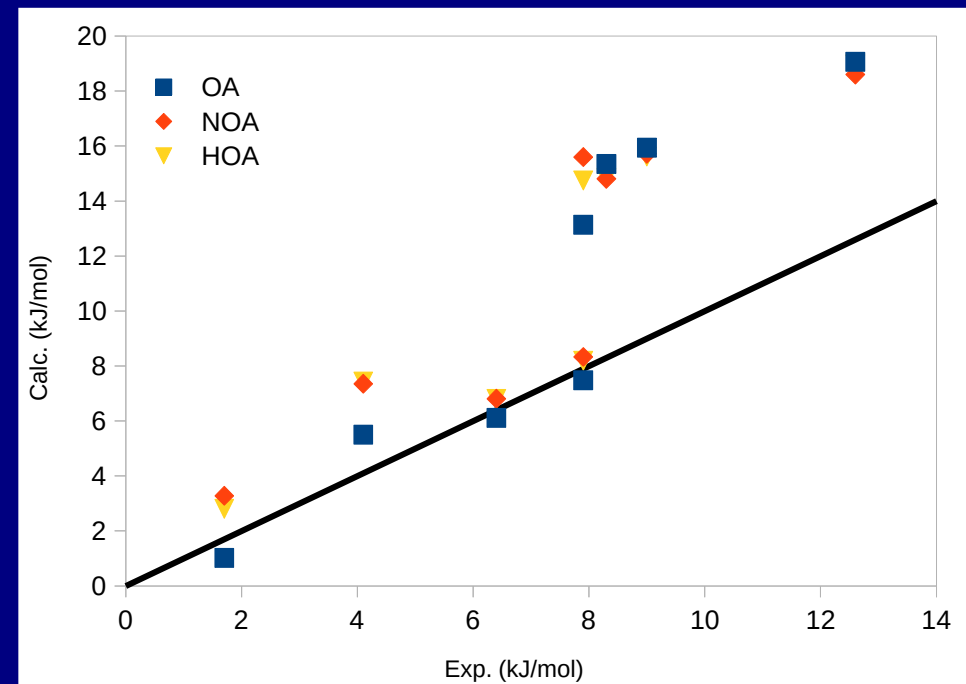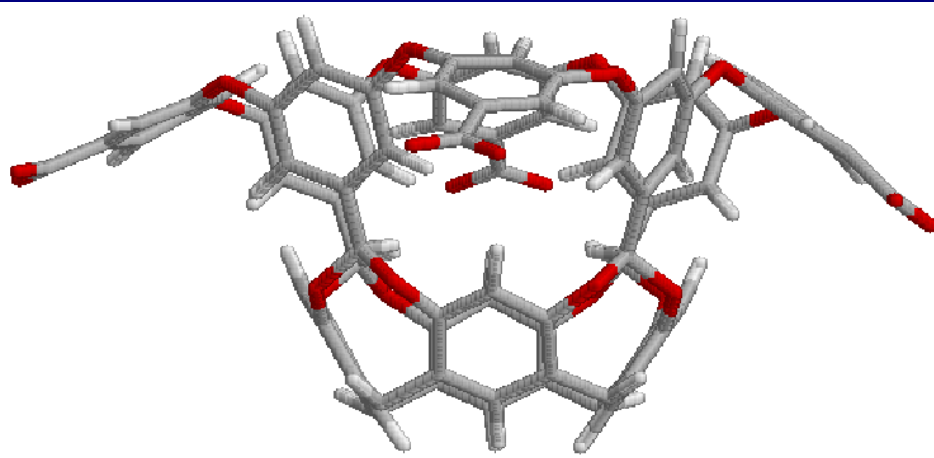
Method

- Approach suggested by Grimme
  Chem Eur J 18(12)9955

ΔG_{bind}

$=$

E_{DFT}

$+$

E_{disp}

$+$

ΔG_{solv}

$+$

ΔG_{freq}

- Optimise structures with TPSS-D3/def2-SV(P)

  - Single-point TPSS/def2-QZVP'
    No counter-poise correction

  - DFT-D3
    Becke–Johnson damping and 3rd-order terms

  - COSMO-RS solvation energy
    from BP/TZVP in vacuum and in COSMO($\varepsilon = \infty$)

  - ZPE, entropy & thermal corrections
    from HF-3c optimisation & frequency calculation
    Rigid-rotor harmonic-oscillator ideal-gas approximation
    low-lying modes with free-rotor approximation

Rigid Interaction Energies

- Results strongly stabilised when using ***rigid interaction energies*** i.e. with the geometry of the complex
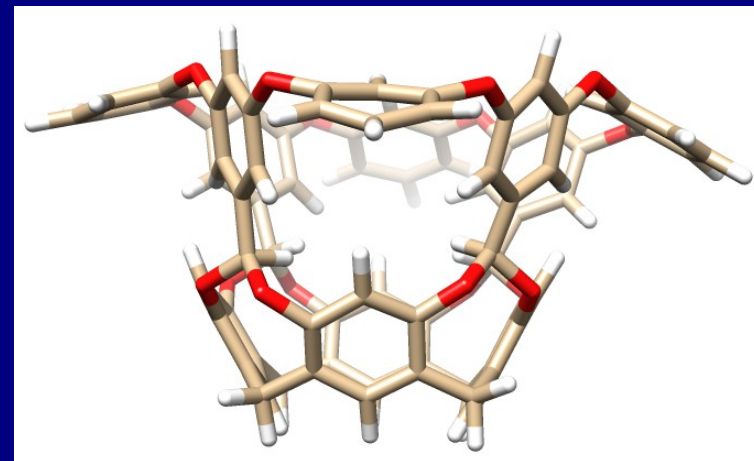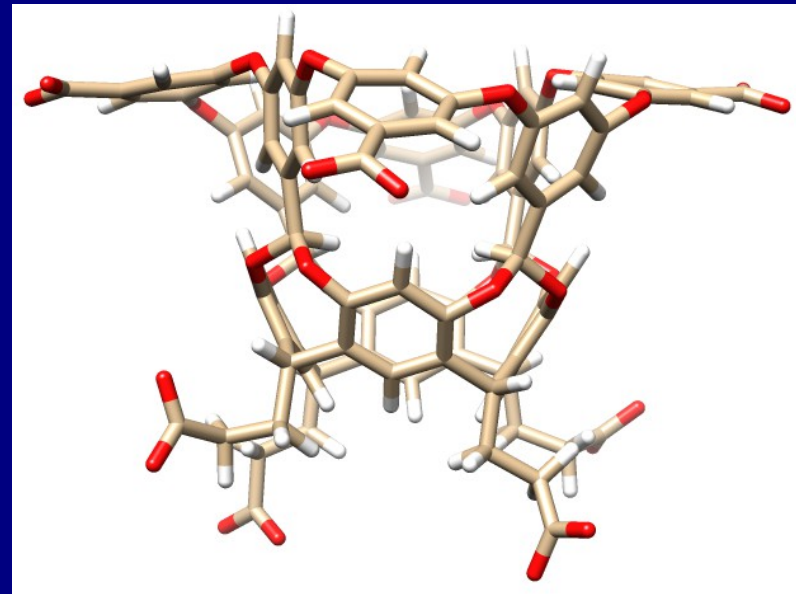
- $E_{bind} = E\text{(HG@HG)} - E\text{(H@HG)} - E\text{(G@HG)}$

- Relaxation energy of guests
0–8 kJ/mol for G2, G4–G6
G1 & G3 13–37 kJ/mol
Deteriorates the results

- Reduces the effect of host flexibility

Four sets

- Fully charged hosts (199–228 atoms)

- Neutralised hosts

- DLPNO-CCSD(T) calculations on DFT structures neutralised hosts

- MD snapshots for neutralised host

DFT Results

ΔG_{bind}

=

E_{DFT}

+

E_{disp}

+

ΔG_{solv}

+

ΔG_{freq}

- Charged host
Large variation in DFT binding energy
−1062 to +973 kJ/mol

Compensated by solvation energy
−906 to +1156

- Neutral host
Less variation in DFT energy
−72 to +30 kJ/mol

Also solvation energy
62–197 kJ/mol

- Sum shows small correlation with Exp. (0–0.4)

DFT Results

ΔG_{bind}

$=$

E_{DFT}

$+$

E_{disp}

$+$

ΔG_{solv}

$+$

ΔG_{freq}

• Dispersion energy
−165 to −94 kJ/mol
Somewhat more variation with charged host

Some correlation with Exp.
$R = 0.3–0.4$

• Entropy and thermal corrections
68–89 kJ/mol

No correlation with Exp.

Total Energies

- Poor results

- All binding energies 41–88 kJ/mol too positive

- MADtr = 18–27 kJ/mol

$R^2 = 0.1$–0.6

$$\Delta G_{bind}$$
$$=$$
$$E_{DFT}$$
$$+$$
$$E_{disp}$$
$$+$$
$$\Delta G_{solv}$$
$$+$$
$$\Delta G_{freq}$$
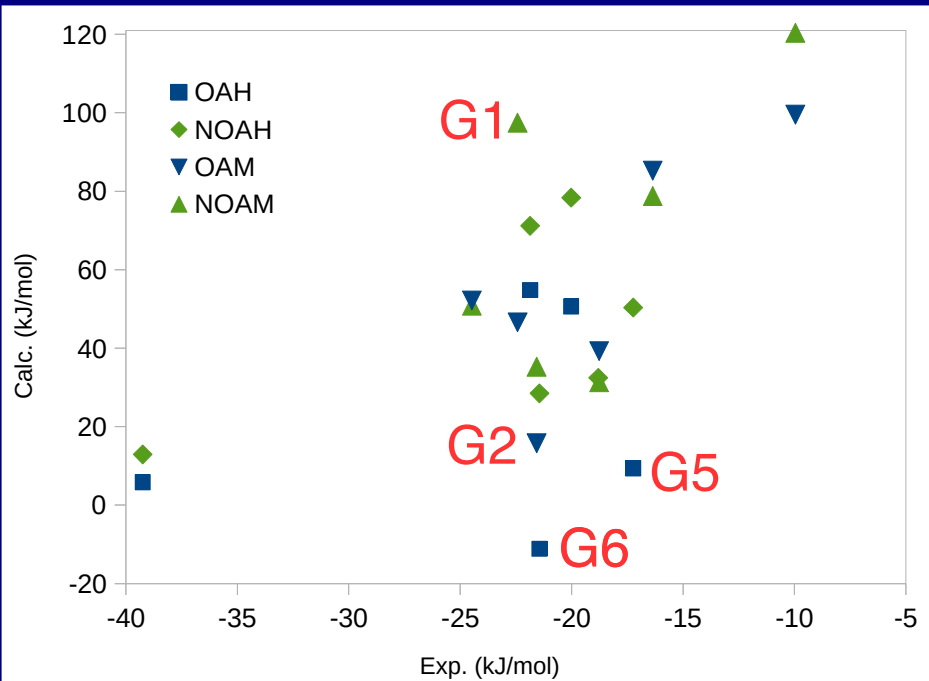
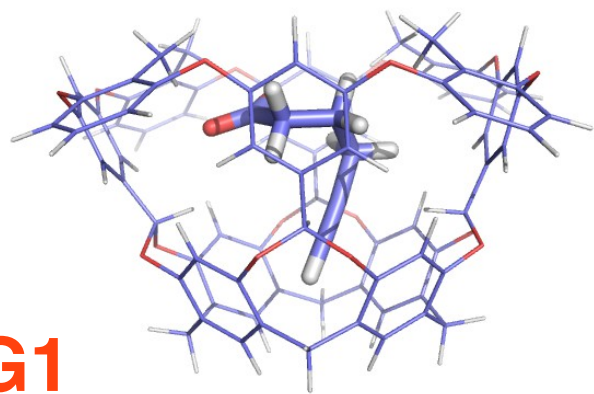|	OAH	OAMe	NOAH	NOAMe
MADtr	23	21	18	27
R^2	0.07	0.60	0.34	0.37

Sometimes Strange Structures
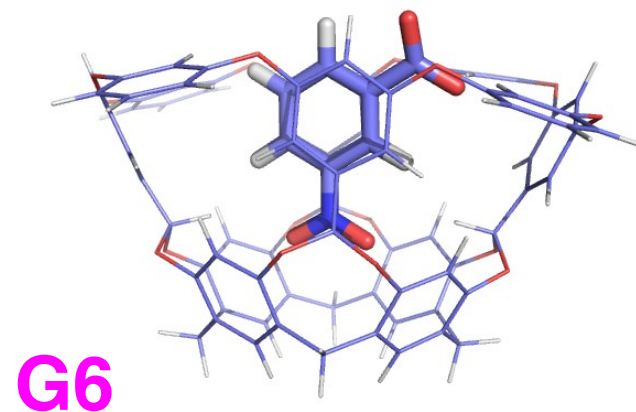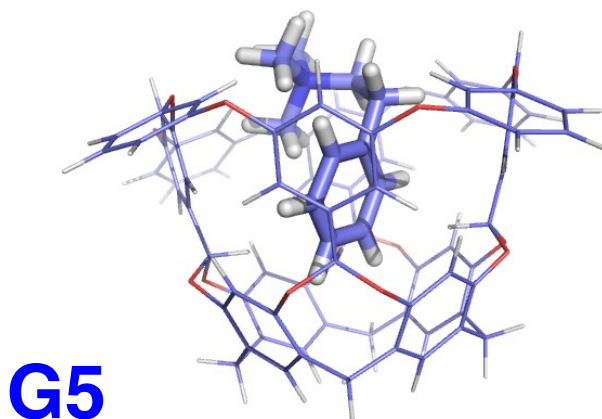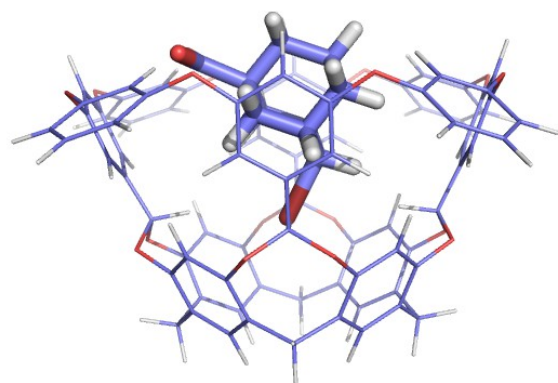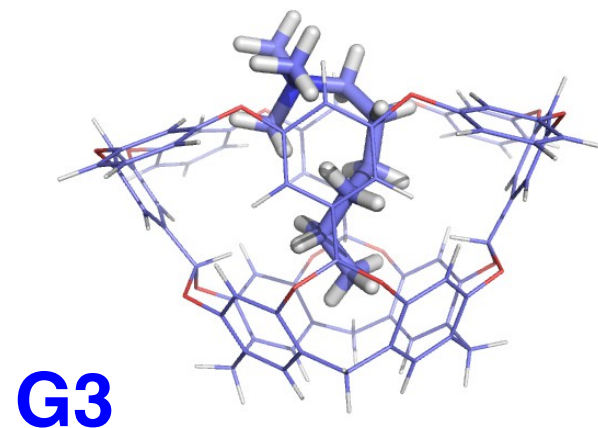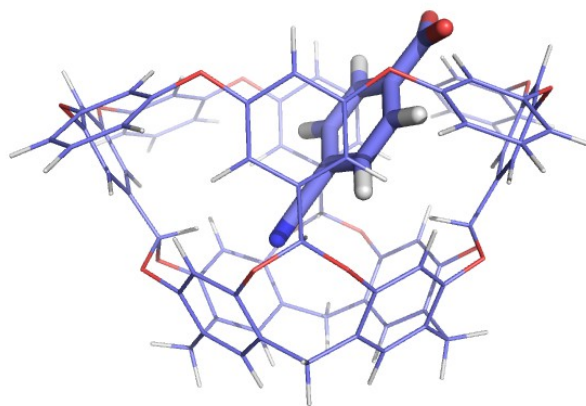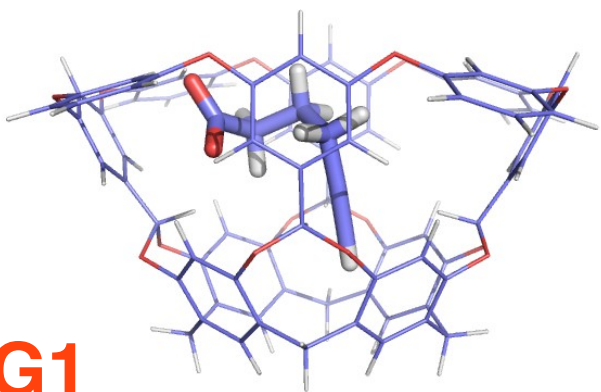
- G1 carboxylate inside host
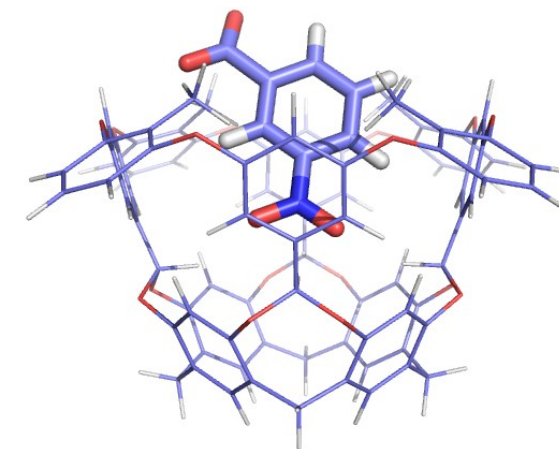- G3 & G5 –NMe$_3$ partly inside
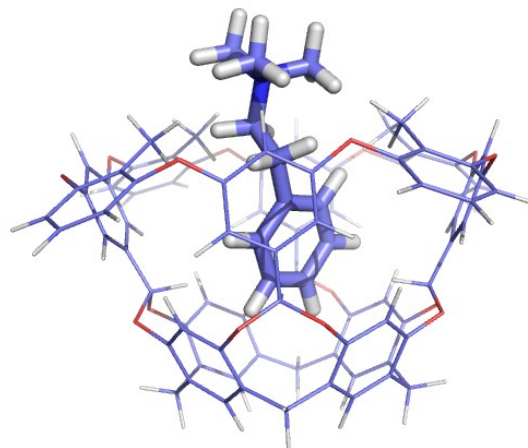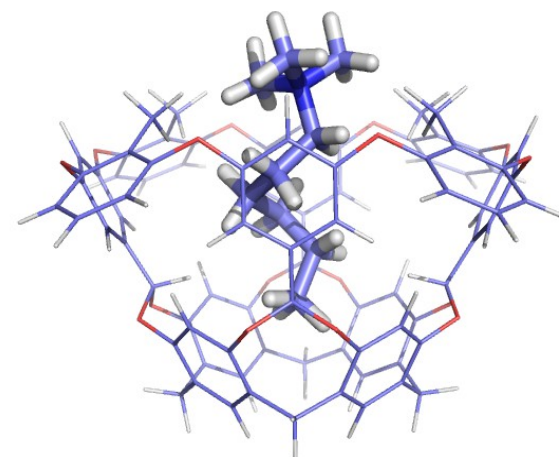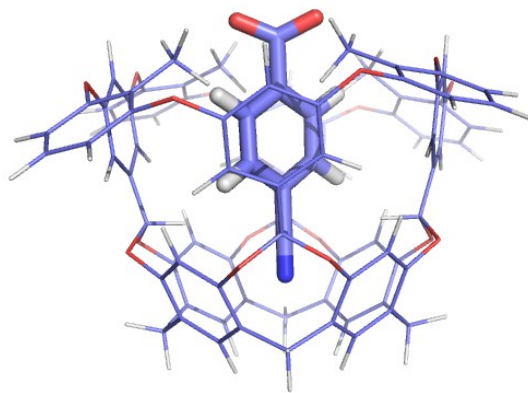- G6 also too deep

Structure Problems
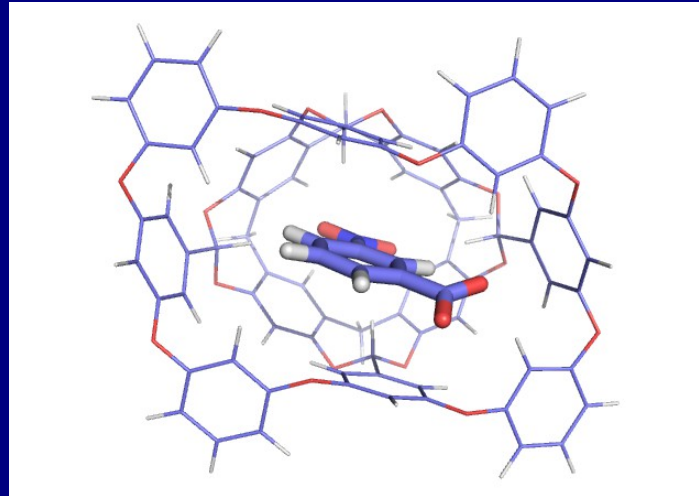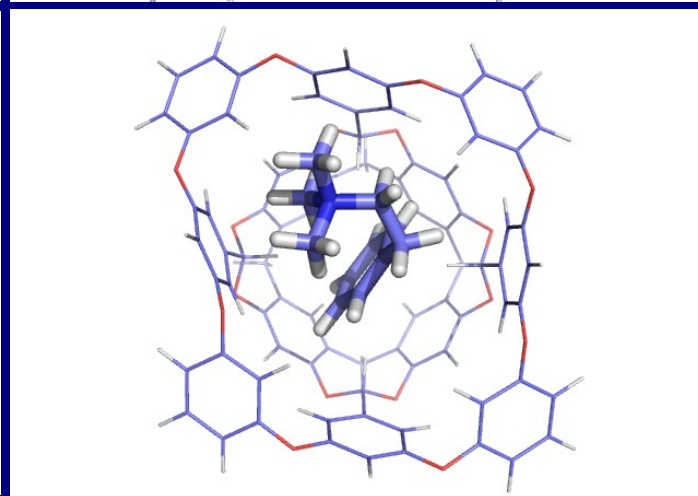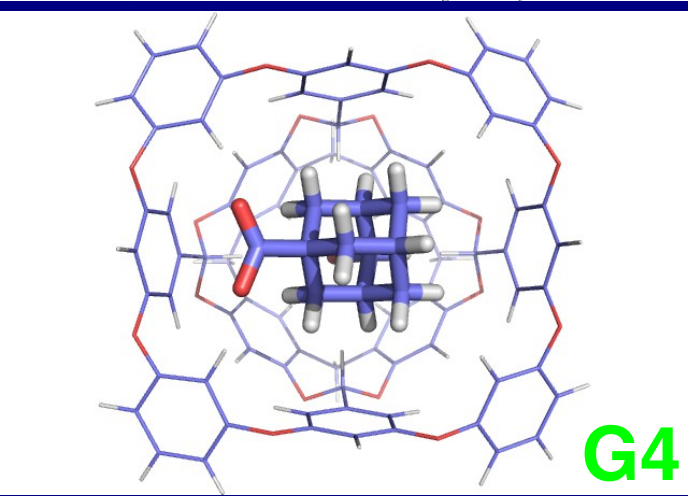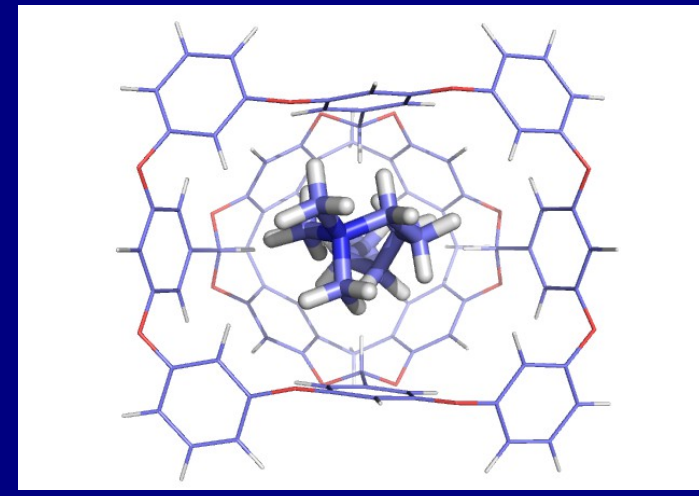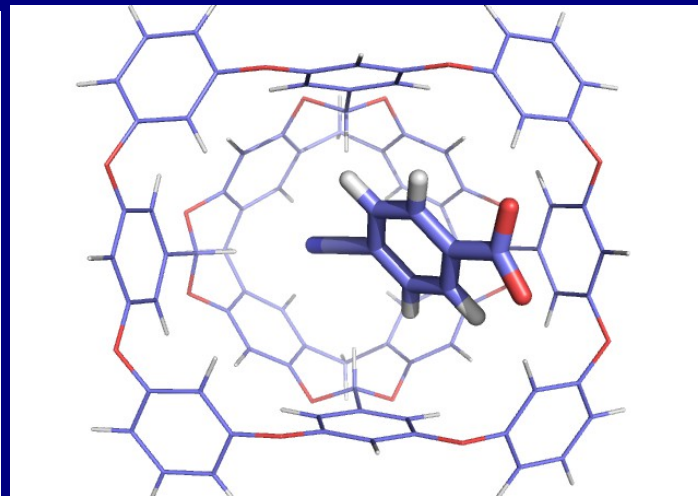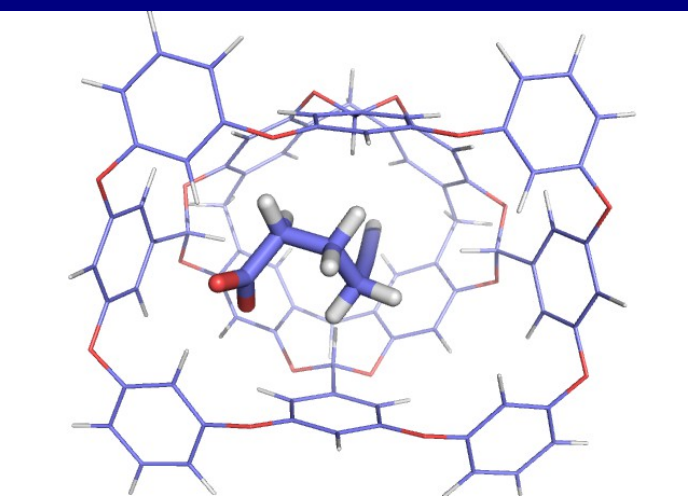
- More distorted than charged hosts
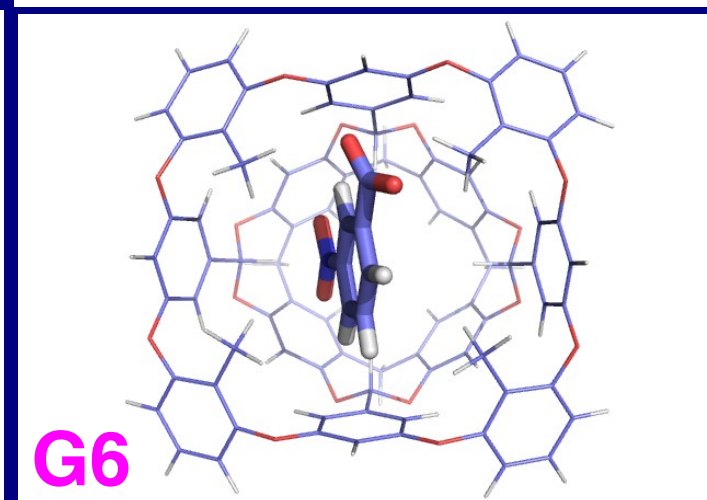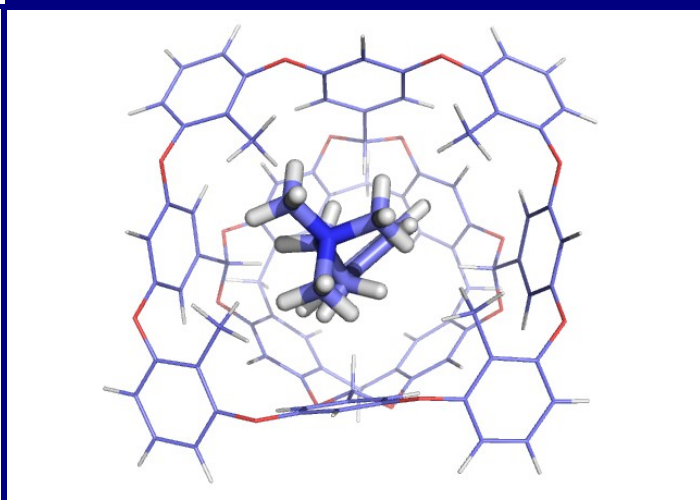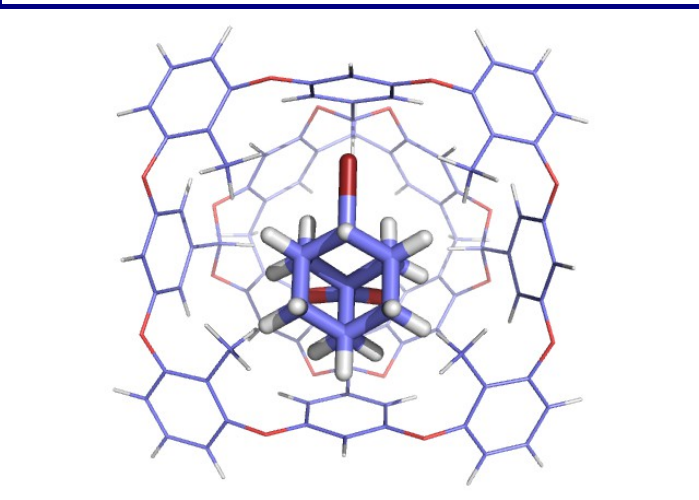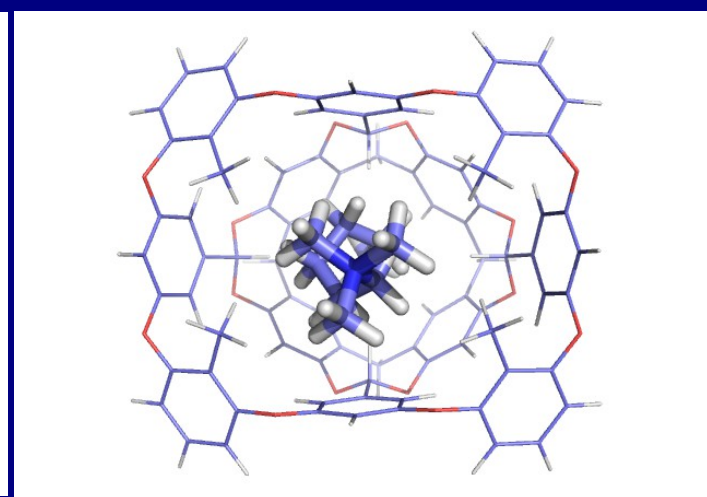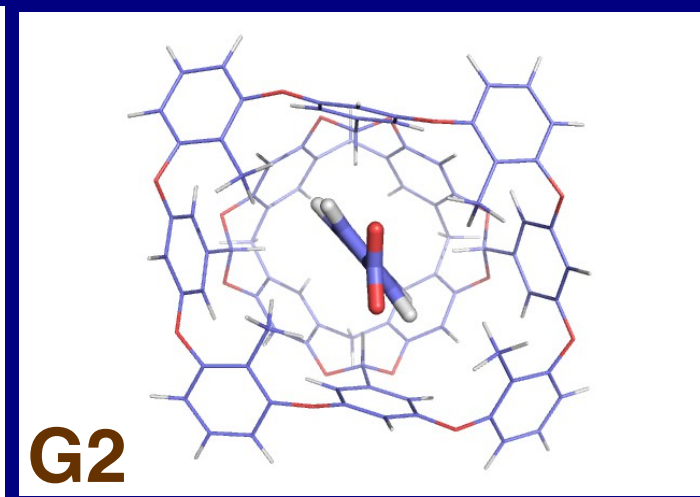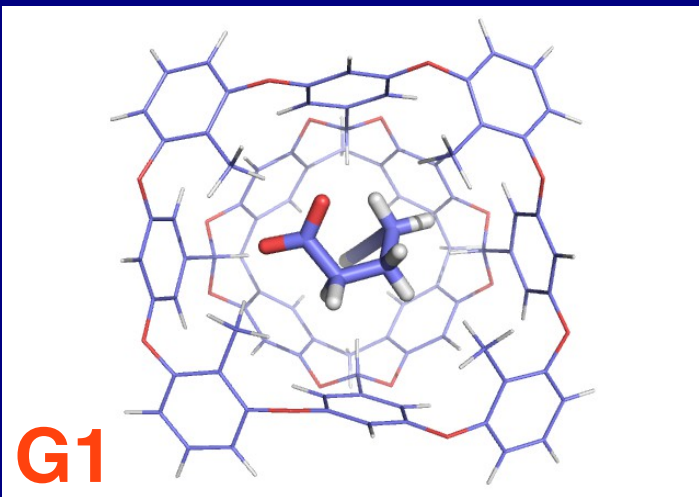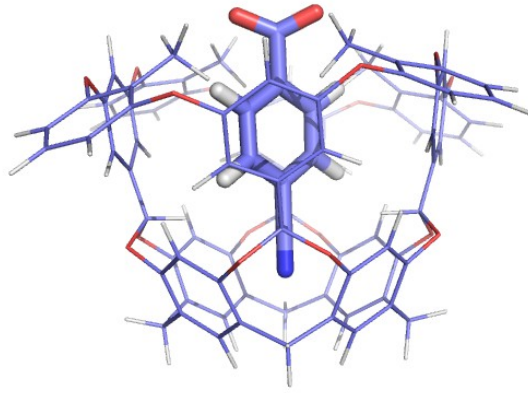
- All distorted, except G4

OAMe structures

G1

- G4 does not bind

- G1 carboxylate still inside host

OAMe structures

G1

G2

G6

- G1 & G6 less distorted

- G2 more distorted

OAH vs. OAMe Differences
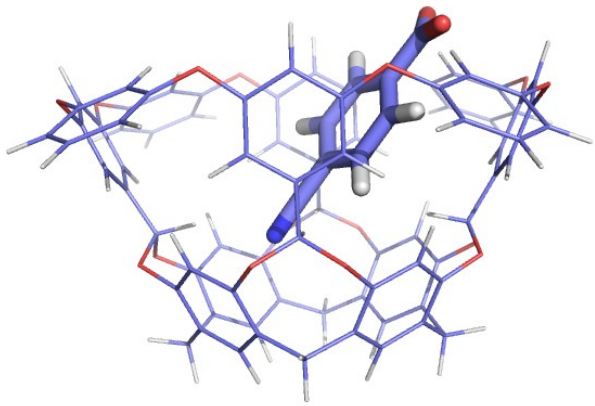
- G2 upright in OAMe
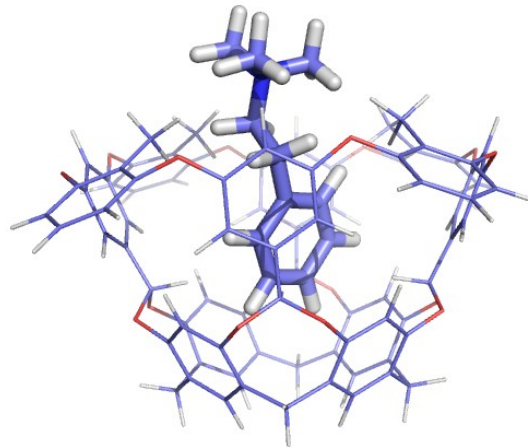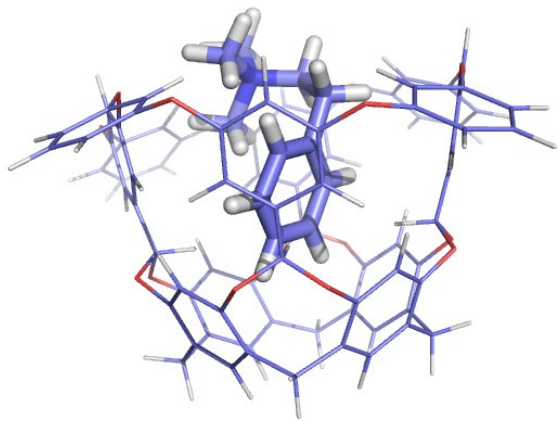
- G3 & G5 –NMe$_3$ more inside in OAH

- G6 more inside in OAH

DLPNO-CCSD(T) Calculations

ΔG_{bind}

$=$

E_{DFT}

$+$

E_{disp}

$+$

ΔG_{solv}

$+$

ΔG_{freq}

- Tried to improve the DFT-D3 results with DLPNO-CCSD(T) calculations
  Neese et al. J Chem Theory Comput 2011, 7 (2011) 33
  Tight PNO thresholds

- Full complex (159–188 atoms)

- Counter-poise corrections

- CBS extrapolation with def2-SVP and def2-TZVP TZ→QZ on way

- Br in G4 treated by ZORA

- CCSD(T) replaced the TPSS/def2-QZVP'+DFT-D3 energies

ΔG_{bind}

$=$

$E_{CCSD(T)}$

$+$

ΔG_{solv}

$+$

ΔG_{freq}

DLPNO-CCSD(T) Results

- Quite similar to DFT results
  −7 to +11 kJ/mol difference
  MAD = 5 kJ/mol

- Much less than in SAMPL4
  LCCSD(T)+PMISP
  24–60 kJ/mol
  Charged host?
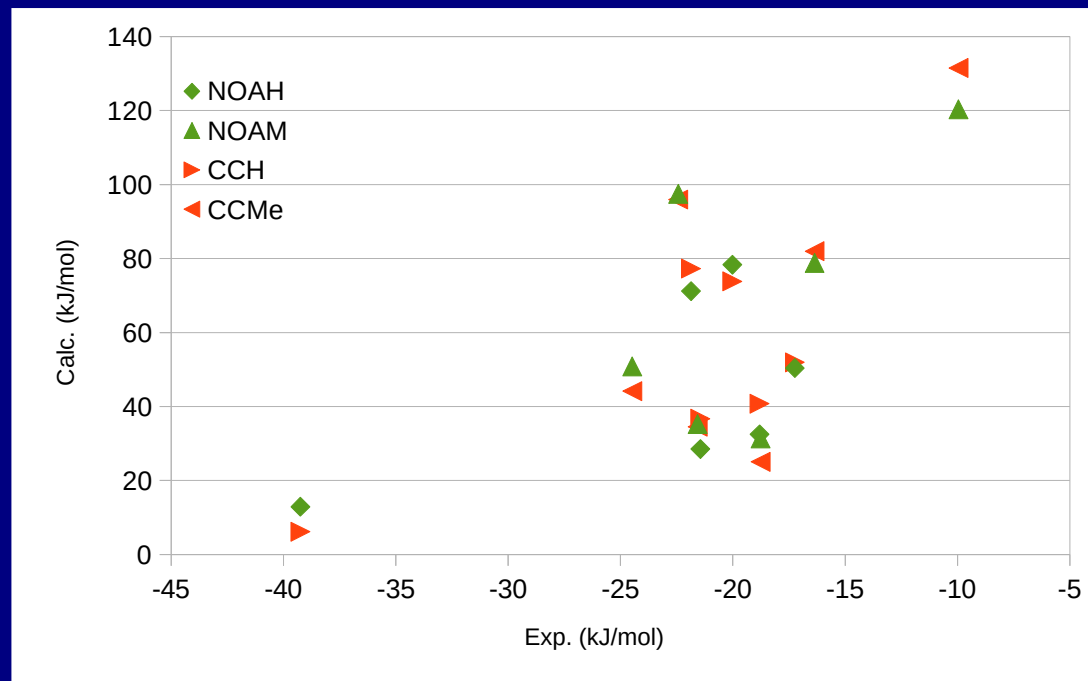
DLPNO-CCSD(T) Results

- Reproduce Exp equally poorly
MADtr = 17–32 kJ/mol
R^2 = 0.4–0.5

- Same DFT structures
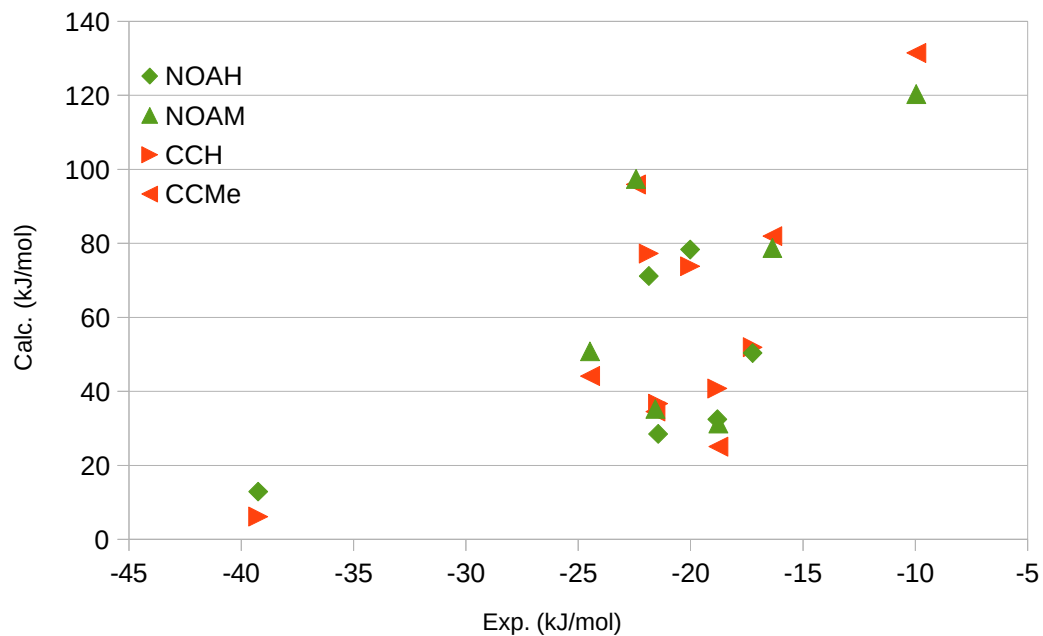
	NOAH	NOAMe	CCH	CCMe
MADtr	18	27	17	32
R^2	0.34	0.37	0.51	0.44

1. Introduction

2. DFT structures

3. DLPNO-CCSD(T)

4. MD snapshots

Structures from MD

$$\Delta G_{bind}$$

$$=$$

$$E_{DFT}$$

$$+$$

$$E_{disp}$$

$$+$$

$$\Delta G_{solv}$$

$$+$$

$$\Delta G_{freq}$$

- Run 10 ns MD simulation of each complex with MM

- Took 10 snapshots

- Optimised with HF-3c

- TPSS/def2-QZVP'
DFT-D3 BJ 3rd-order dispersion
COSMO-RS solvation energy (BP/TZVP)
HF-3c ZPE, entropy & thermal corrections
(*same* as DFT-opt)

Variation of binding free energies

- Restricted variation for most guests; 3–15 kJ/mol
  $\Rightarrow$ Standard Error = 0.3–1.4 kJ/mol

- Sometimes a one outlier

- G6 large variation (28–52 kJ/mol; SE 3–5 kJ/mol)

- G4 bimodal in OAH
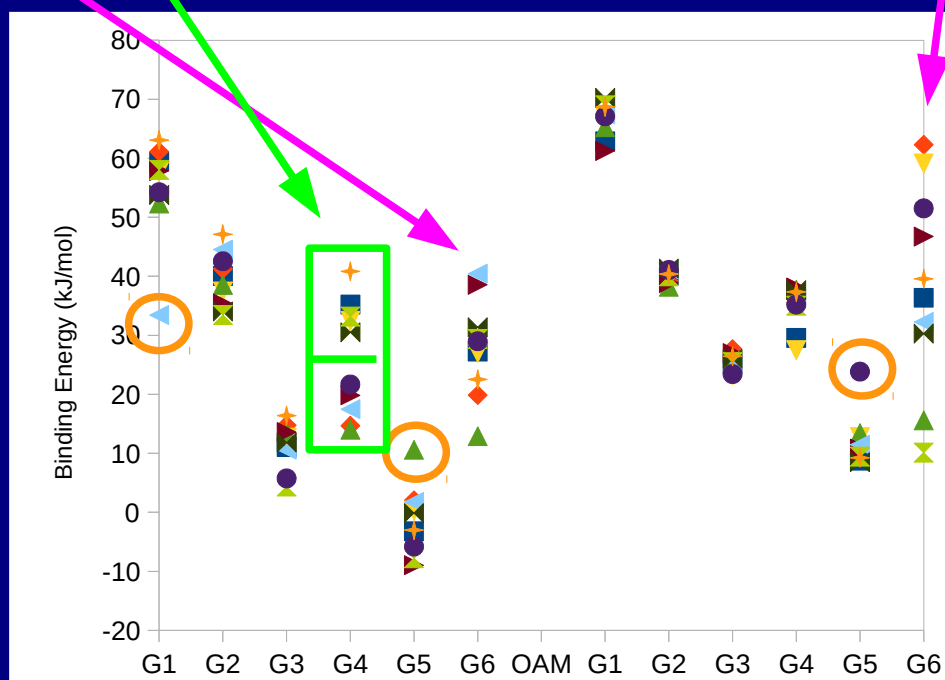
Structures different (NOAH)
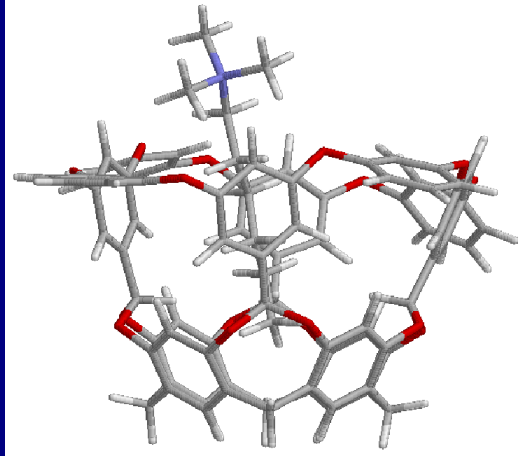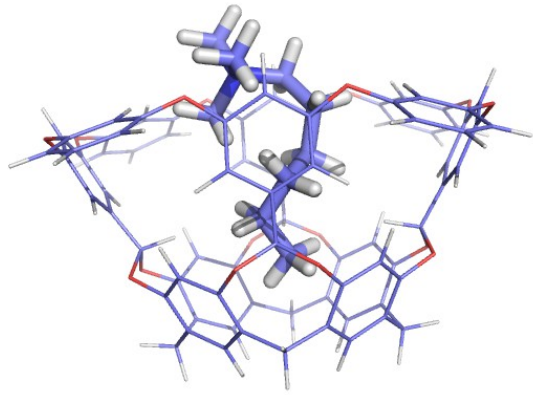
- G3 much more outside
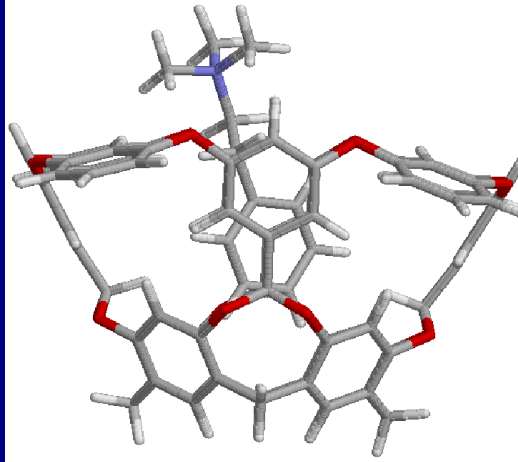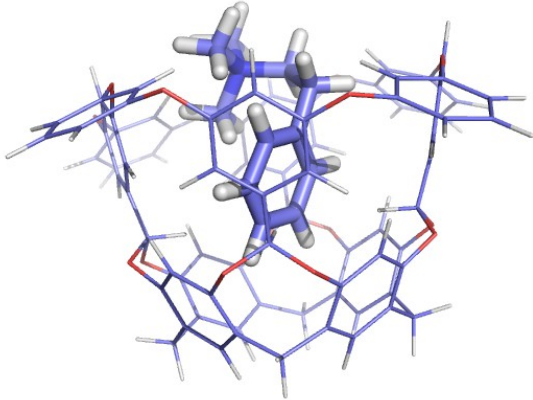
- G5 much more outside
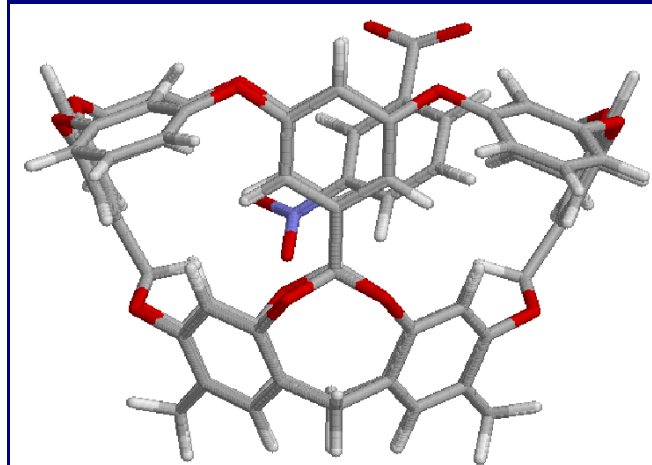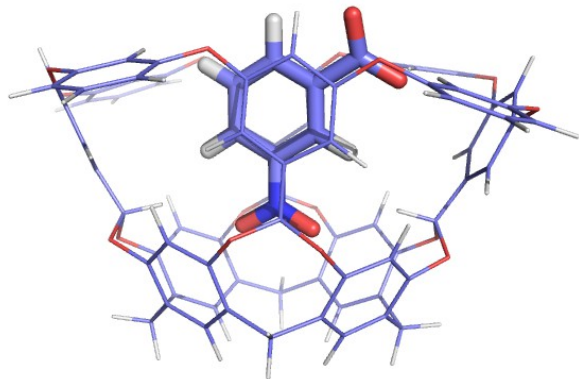
- G6 sometimes other orientation

Structures different (NOAMe)
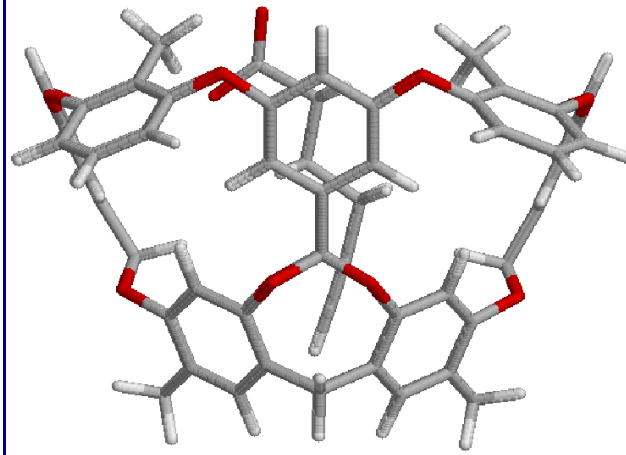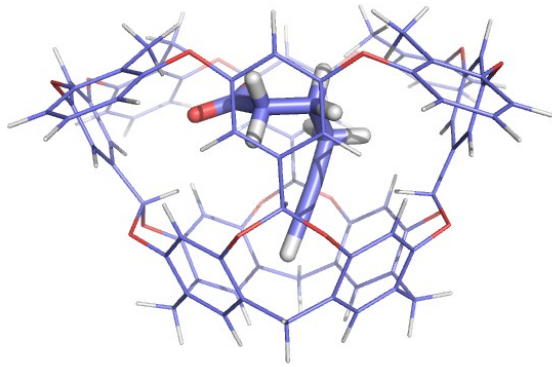
- G1
less inside
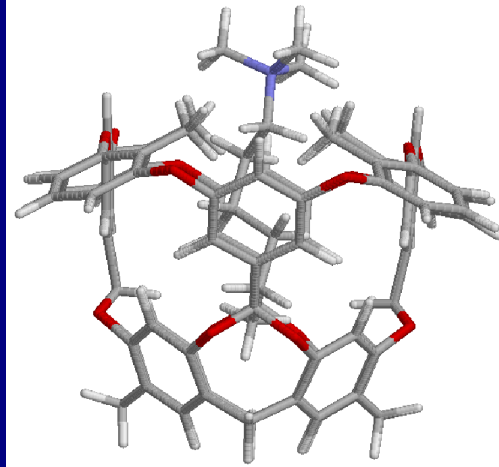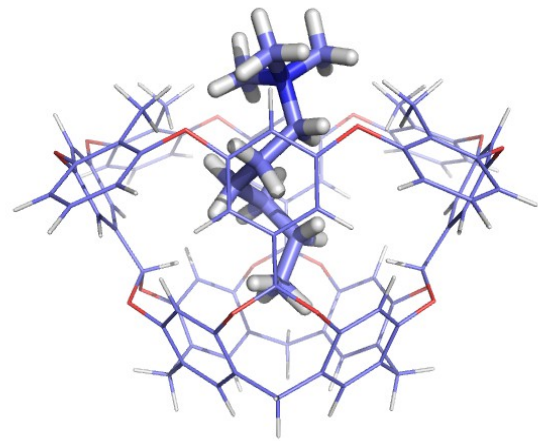
- G3 & G5
rather similar
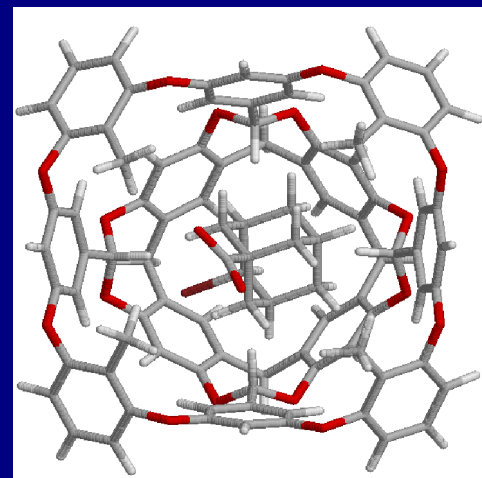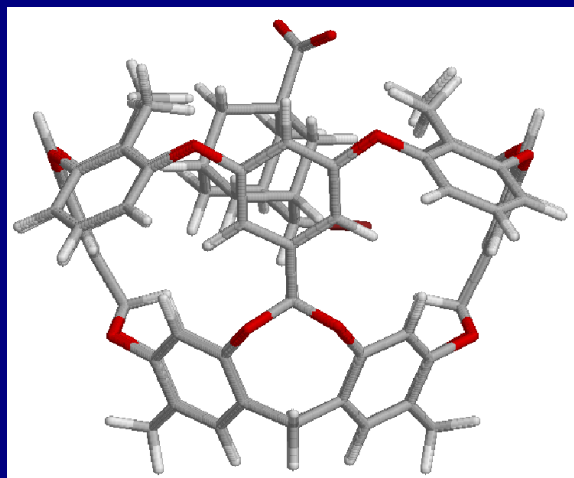
- G4
forced inside in MD
and stayed there
Distorted host

Results

- R^2 much worse
  0.0 to –0.1

- partly owing to
  forced binding of G4
  (host distortion E missing)

- MAD better
  14–17 kJ/mol

- G3 & G5 too strong

- G1 too weak
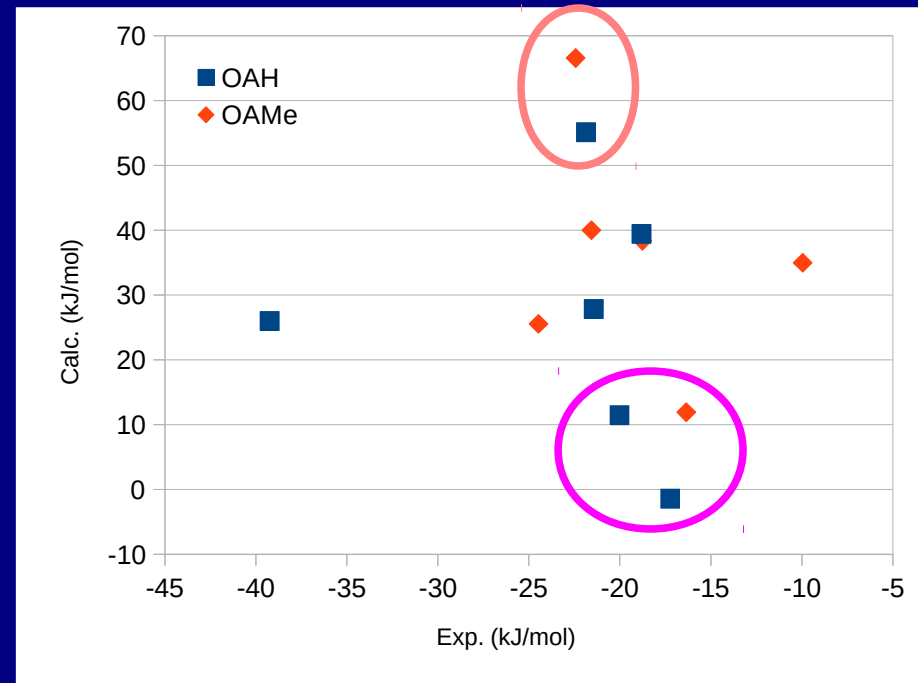
	NOAH	NOAMe	MDH	MDMe
MADtr	18	27	17	14
R^2	0.34	0.37	-0.02	-0.08

Conclusions MD

- Promising approach

- More black-box

- Gives a precision of 1 (–5) kJ/mol

- Trying better methods
  for optimisation

Improved Optimisation Methods
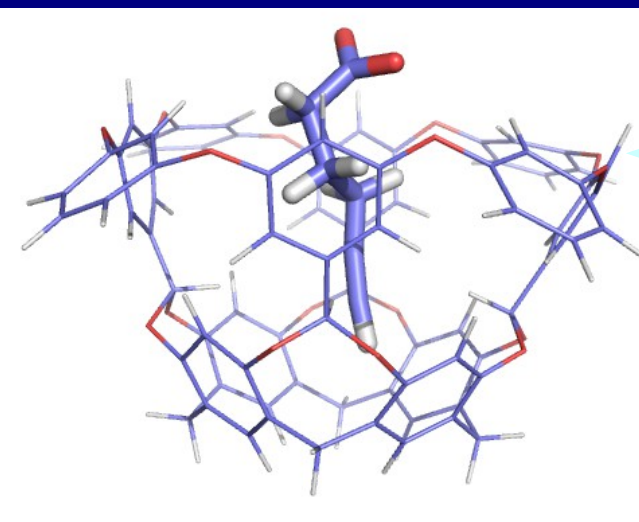
- MD snapshot

- Vacuum opt.
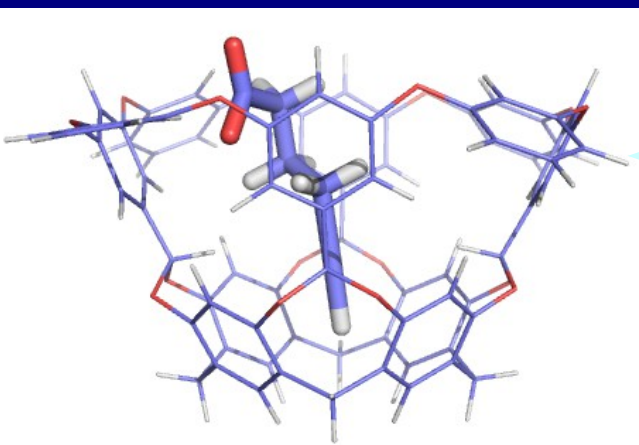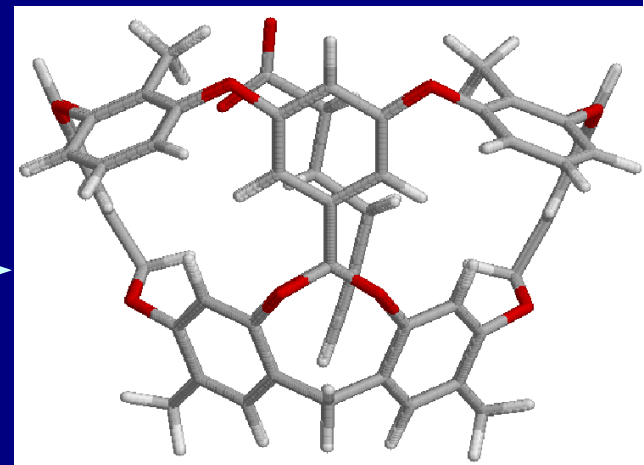
- COSMO opt.

- COSMO + 4 Water
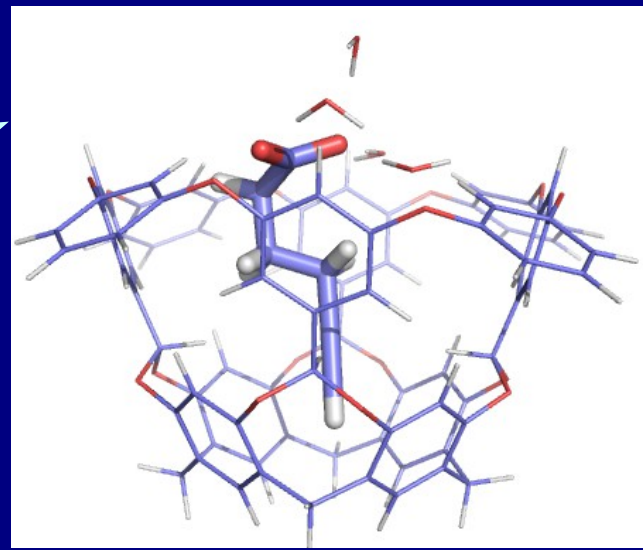
Conclusions

- Poor results

- Partly owing to the use of vacuum structures

- Possibly additional un-recovered problems (unexperienced PhD)

- Flexibility a major problem

- Optimisation time-consuming (~1 month)

- Need much nursing

- MD approach may solve some problems

Worth Investigating

- Alternative QM/MM FEP
with MM reference potential

- Latest results
with SAMPL4 data

- QM systems of 158–224 atoms

- NOA and PM6-DH2X

- 720 000 QM calculations are needed
for convergence to 1 kJ/mol

- DFT optimisation
~200 QM calculations
1/4000 !

$$RL_1^{QM/MM} \xrightarrow{\Delta G_{L_1 \to L_2}^{QM/MM}} RL_2^{QM/MM}$$

$$\Delta G_{L_1}^{MM \to QM} \uparrow \qquad \uparrow \Delta G_{L_2}^{MM \to QM}$$

$$RL_1^{MM} \xrightarrow{\Delta G_{L_1 \to L_2}^{MM}} RL_2^{MM}$$
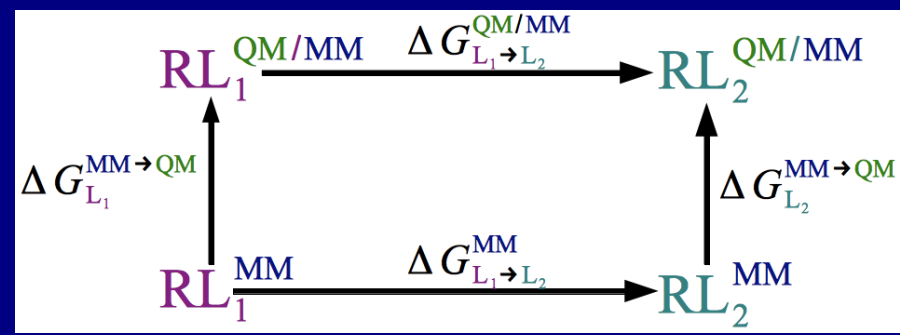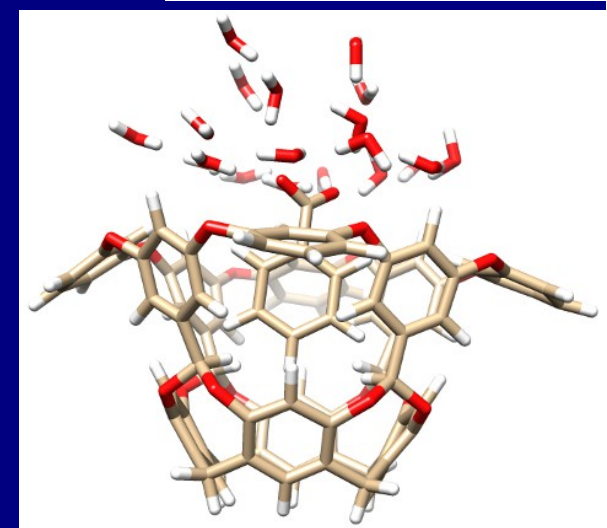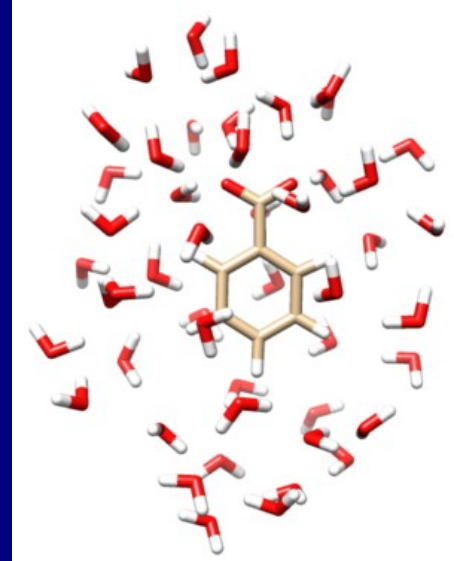
Acknowledgements

Octav Caldararu
Martin Olsson

Dr. Christoph Riplinger
Prof. Frank Neese

Swedish Research Council

Lunarc
HPC2N

Submitted Data

We submitted three sets of data

1. Charged host with DFT

2. Neutral host with DFT

3. Neutral host with DLPNO-CCSD(T)//DFT

- OAH, not rigid energies
(missing at submission time)

- With guest relaxation energies
(suboptimal)

- Error in isolated G2 energy

- MD not finished in time

- No Coulombic correction

MM-MD structures
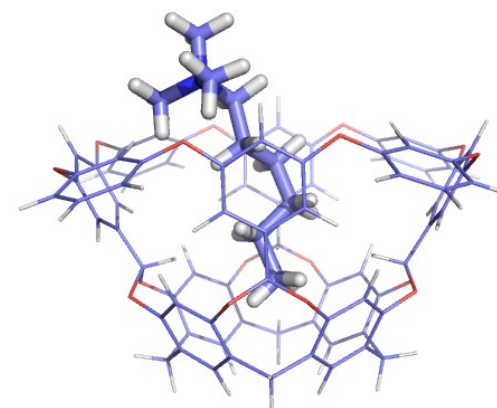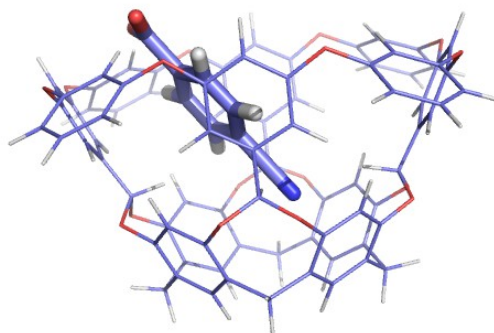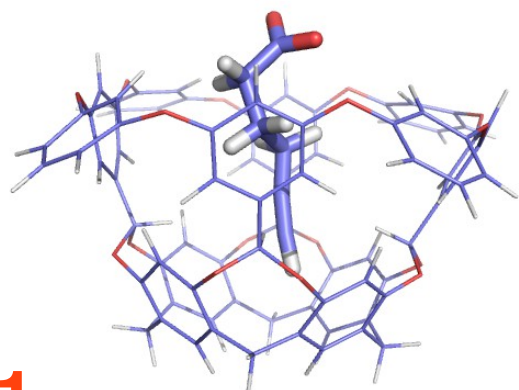
G1
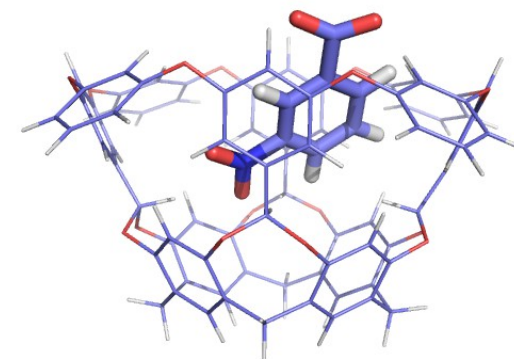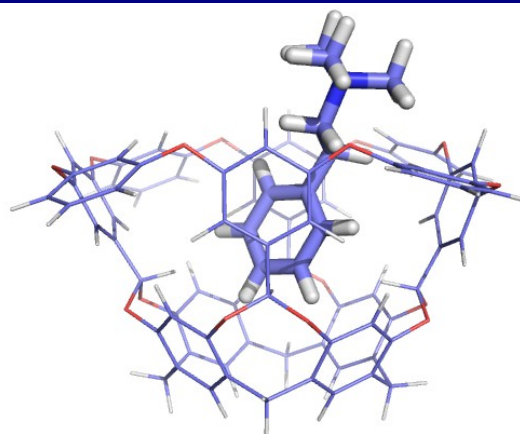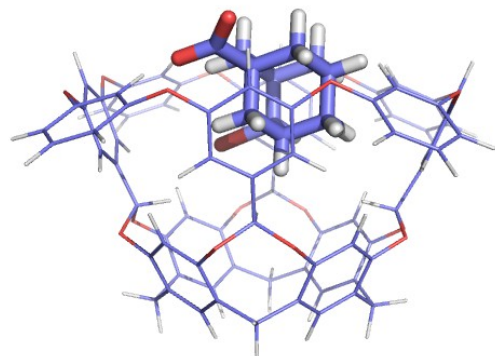
- **G1** carboxylate actually inside host

MM-MD structures OAMe
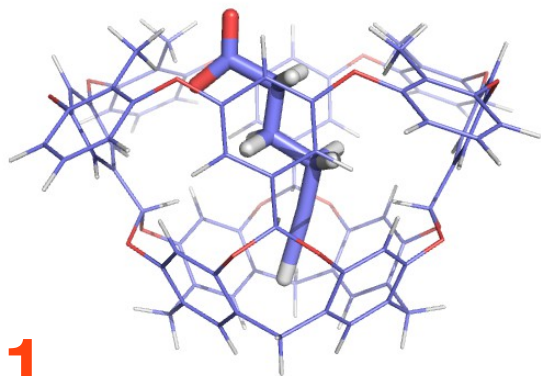
G1
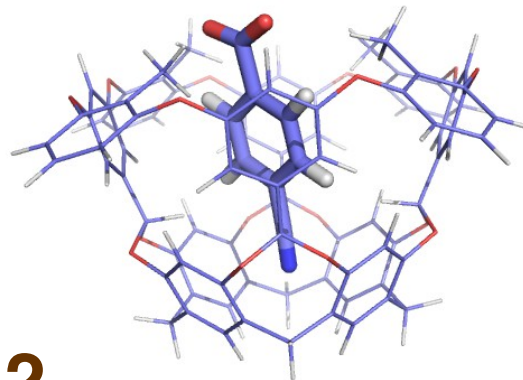
G2
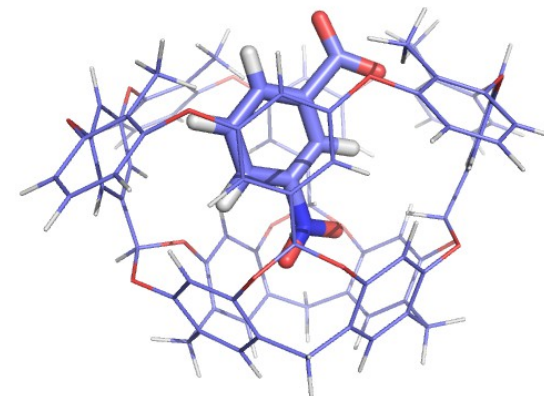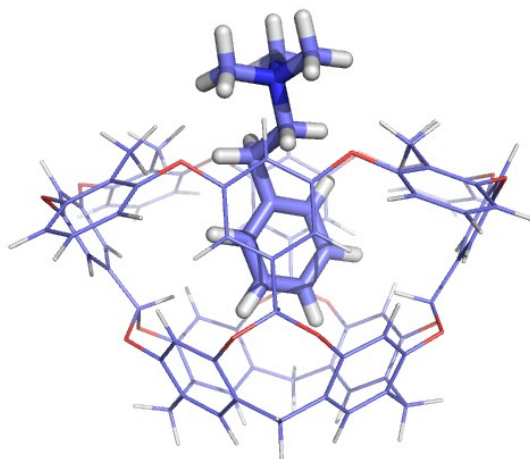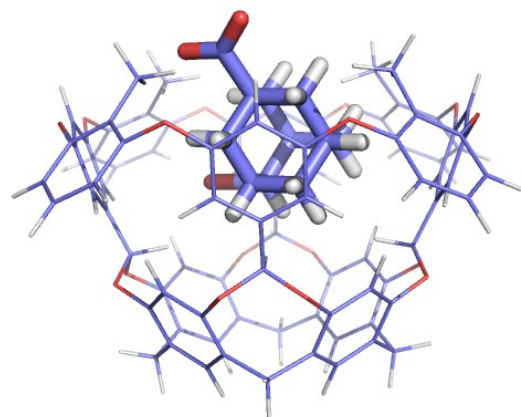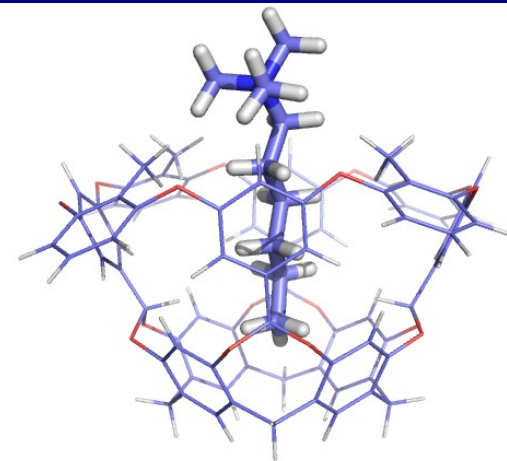
- G1 carboxylate more exposed

- G2 more up-right

LCCSD(T)–PMISP

- Tried to improve the DFT-D3 results with
Local CCSD(T) calculations
Hampel & Werner, J Chem Phys 104 (1996) 6286

- Employing the PMISP approach
Polarised Multipolar Interactions with Supermolecular Pairs
Söderhjelm & Ryde, J Phys Chem A 113 (2009) 617

- Based on the Cos structures
solvation energies, and thermal corrections

- Replaced the
TPSS/def2-QZVP'/TZVP+DFT-D3 energies
with
LCCSD(T) energies

LCCSD(T) calculations

- Together with Prof. Ricardo Mata, Göttingen

- cc-pVTZ basis set

- Extrapolated to basis-set limit
at conventional MP2 level with
aug-cc-pVTZ and aug-cc-pVQZ basis sets
and n^{-3} scheme

- Full counter-poise corrections

- Density fitting

- Pipek–Mezey localisation

PMISP

$$E_{tot} = E_{ele} + E_{ind} + E_{other}$$

$$E_{other} =$$

$$\Sigma \, c_j(E_{QM}(BA_i) - E_{ele}(BA_i) - E_{ind}(BA_i)$$
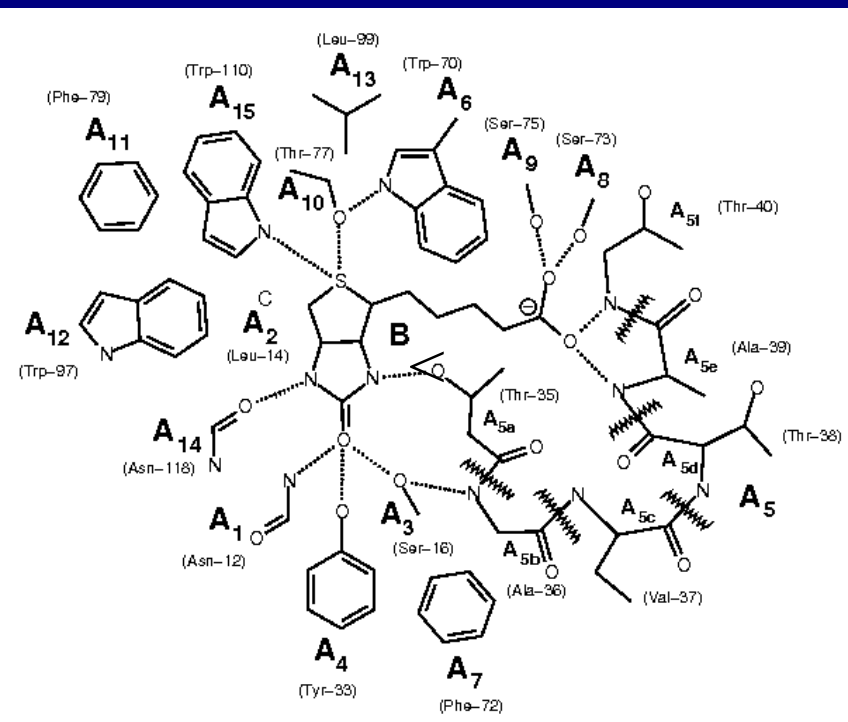
MM (NEMO) electrostatics and induction

QM calculations on
Guest (B) and Host fragments (A$_j$)
(to get multipole expansion and polarisabilities)

and all BA$_i$ pairs
(to get dispersion, repulsion, penetration, CT, ... = "other")

PMISP details

- Multipoles up to octupoles
and
anisotropic polarisabilities
in atomic centres and all bond midpoints

- Calculated at the MP2/cc-pVTZ level

- Interaction energies between guest and host
calculated for guest+fragment
at the LCCSD(T)/CBS level

- Many-body effects treated at the MM (NEMO) level
with the multipoles and polarisabilities

Fractionation

- Host fractionated according to the MFCC approach
Molecular Fractionation with Conjugate Caps

- Adapted to the complicated host

- 24 fragments
32 conjugate fragments
4 doubly-conjugate fragments
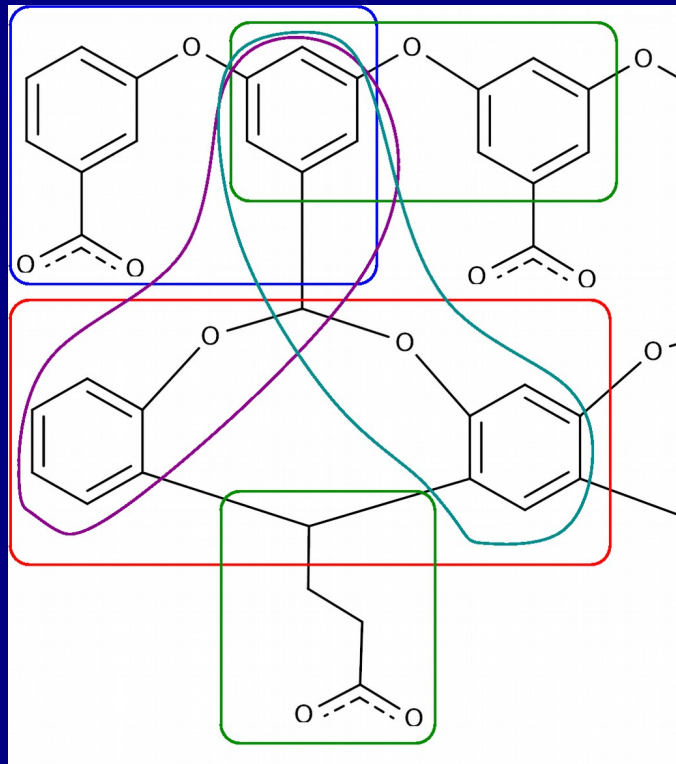
Results

- CCSD(T) correction rather large and negative –27 (Hx) to –54 (mClBz) kJ/mol

- All quality measures significantly worse

- Constant shift + exaggerate differences

- Submitted absolute affinities with one outlier

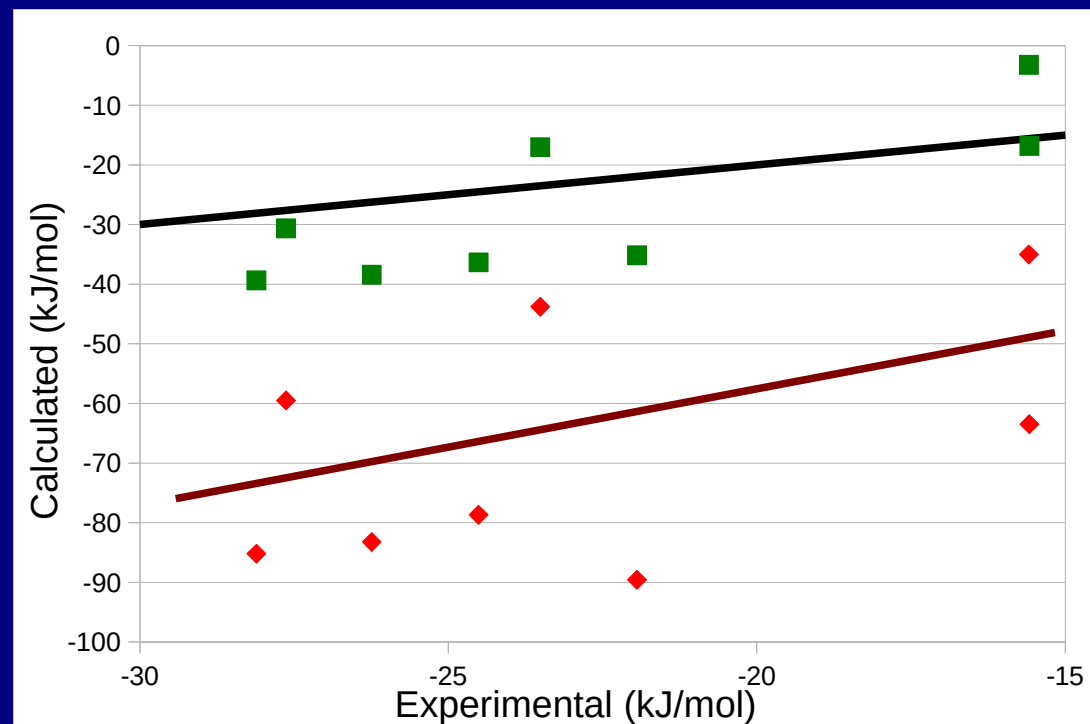|	Cos	CC	corr
Bz	-17	-64	-47
MeBz	-36	-79	-42
EtBz	-38	-83	-45
pClBz	-39	-85	-46
mClBz	-35	-90	-54
Hx	-17	-44	-27
MeHx	-44	-79	-35
Pen	-3	-35	-32
Hep	-31	-60	-29

	Cos	CC
MAD	9	48
MADtr	8	24
R^2	0.7	0.4
τ	0.7	0.3
τ_{90}	0.8	0.3
PI	0.9	0.5
RMSD	10	50
MSD	-5	-37
slope	2	4
intercep	21	41

Conclusions

- MM–FEP acceptable results   *Best*
    - Not much better than for proteins
  MAD = 6.0 kJ/mol for 91 transformations in 10 proteins
- No difference between RESP & BCC charges

- DFT-FEP has severe convergence problems   *Poor*

- Optimised DFT intermediate performance
    - Severe problems with flexibility   *Intermediate*
      - Hard to improve

- LCCSD(T) possible for ligand binding with PMISP
    - But poor results – why?   *Worst*

Structures different (NOAH)
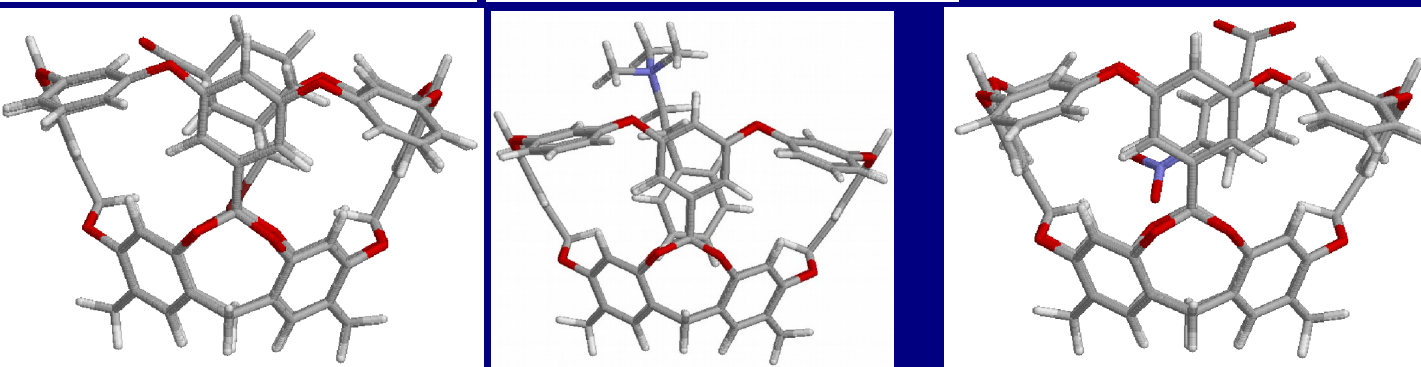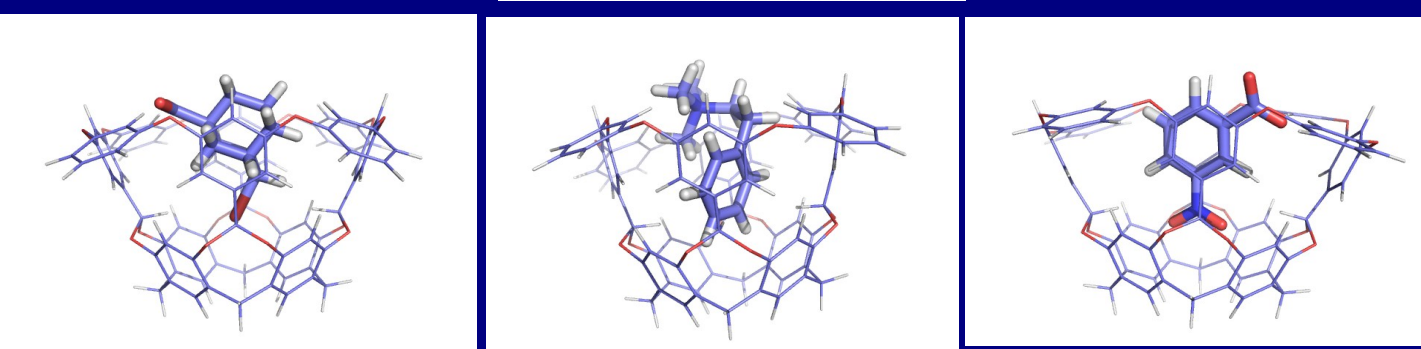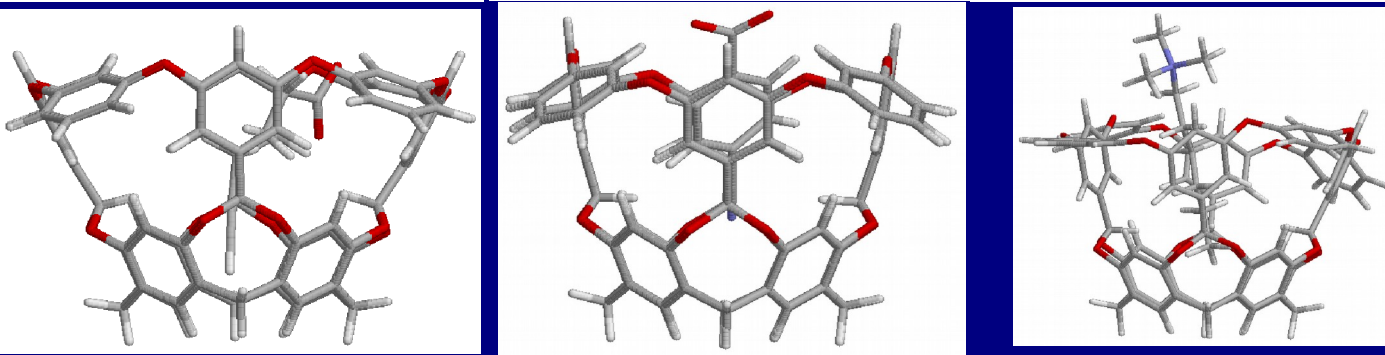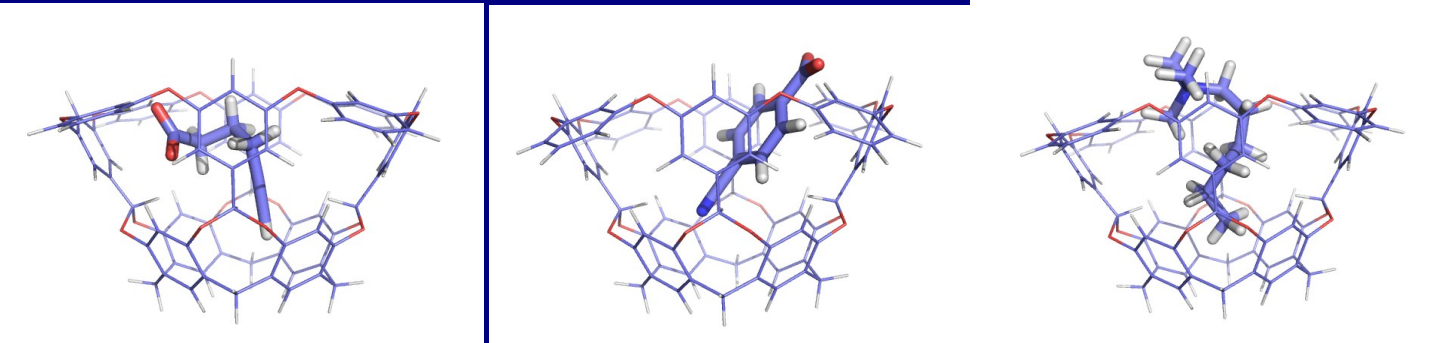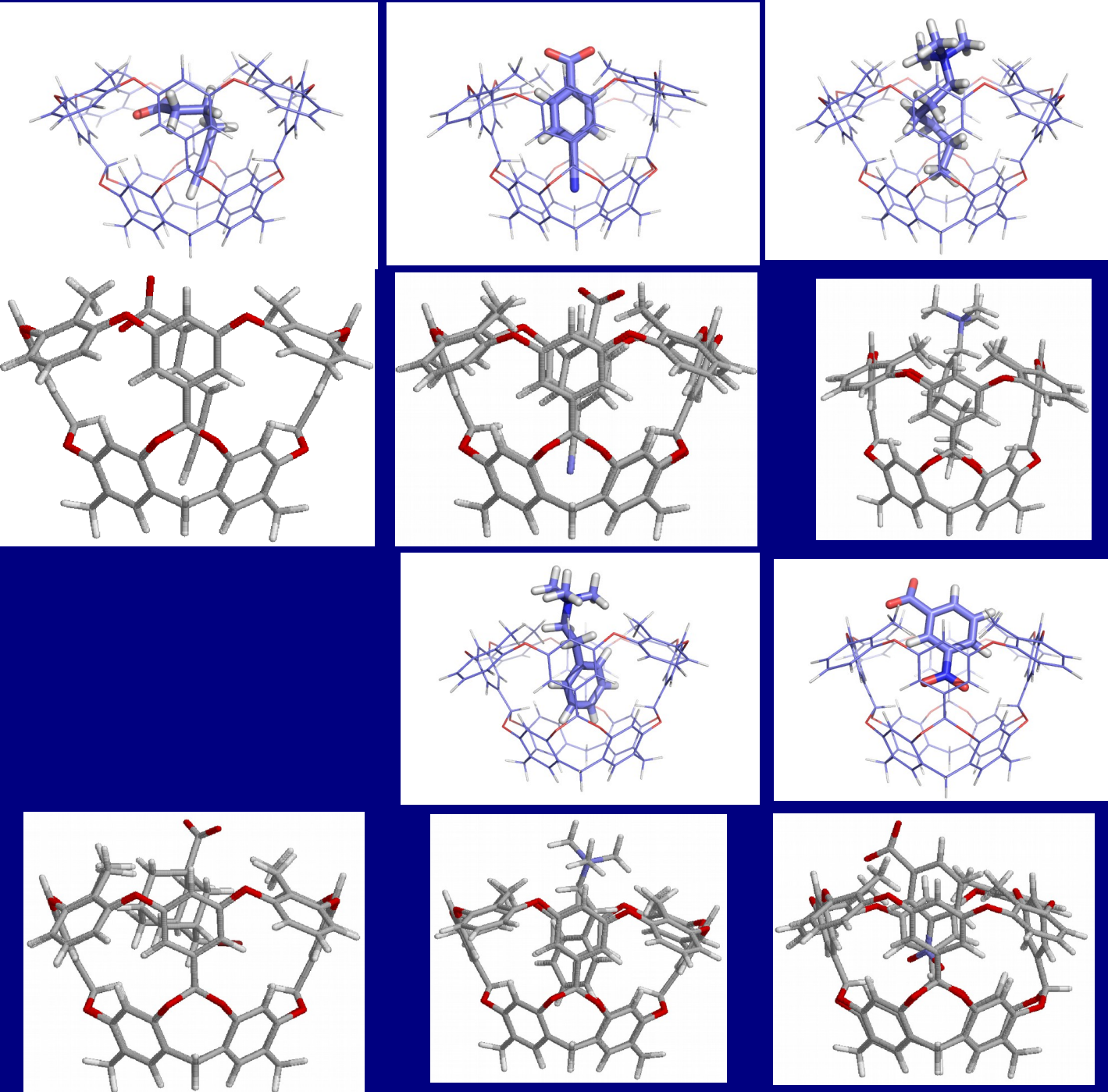

- G1 deeper
- G2 similar
- G3 much more outside
- G4 similar
- G5 much more outside
- G6 other orientation

Structures different (NOAMe)

- G1 less inside
- G2 similar
- G3 rather similar
- G4 forced inside
- G5 similar
- G6 similar