Global optimization in chemistry: from abstract tests to practical usefulness

Bernd Hartke

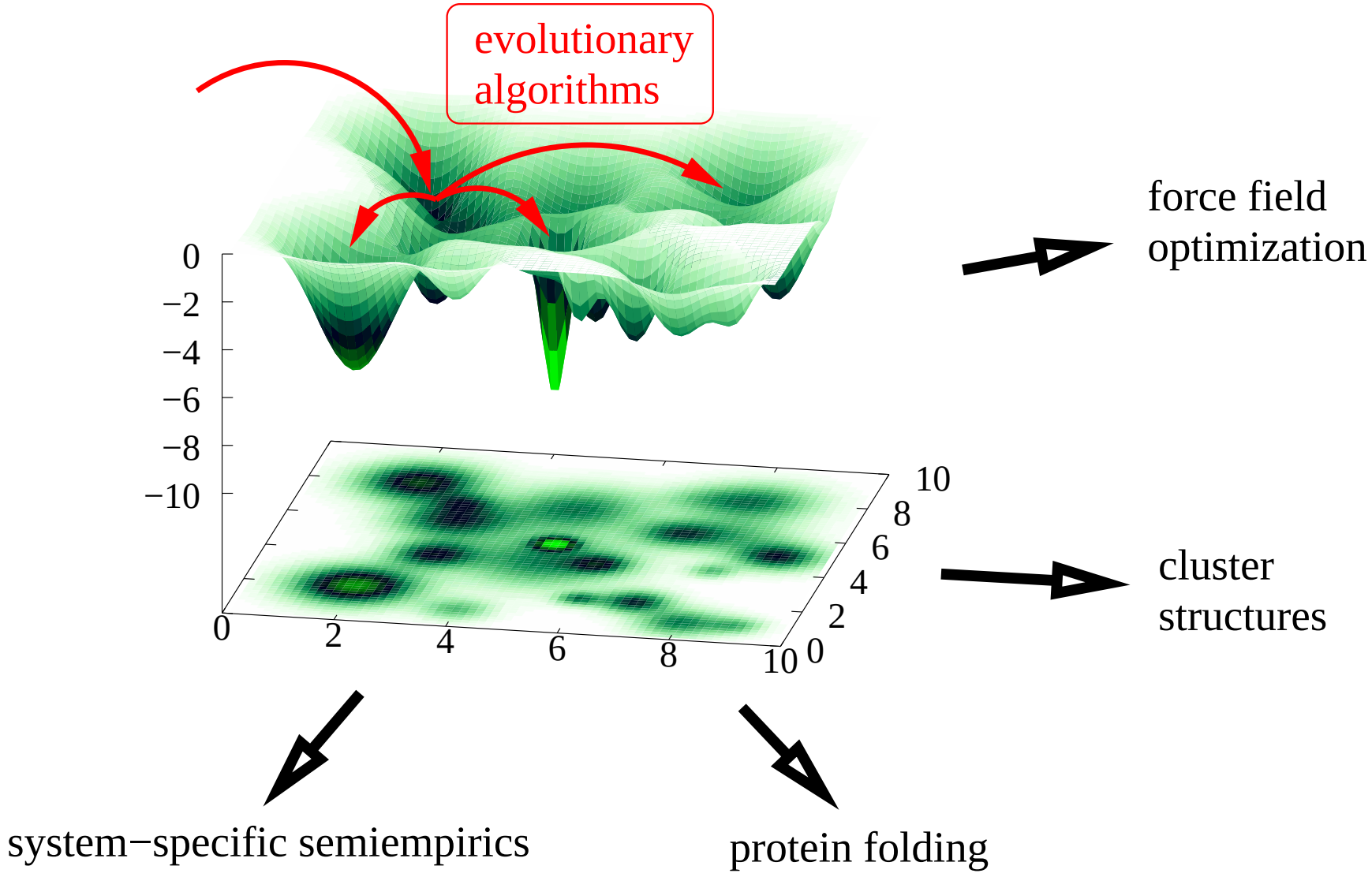
Theoretical Chemistry
Institut für Physikalische Chemie
Christian-Albrechts-Universität
Olshausenstraße 40
D-24098 Kiel, Germany

`http://ravel.phc.uni-kiel.de/`

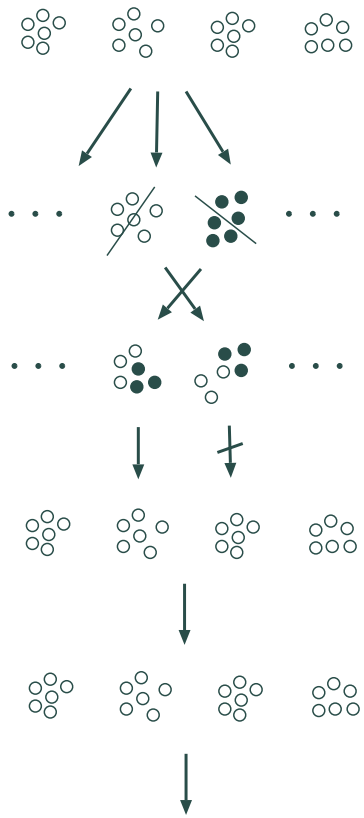
`e-mail: hartke@phc.uni-kiel.de`

Global optimization

minima grows exponentially with dimensionality



Global optimization by Evolutionary Algorithms [1] [2] [3]



generation 1 — **serial**

all possible pairs

crossover & mutation; local optimization — **parallel, 1**

fitness and geometry selection — **serial**

postprocessing — **parallel, 2**

generation 2 — **serial**

⋮

important aspects:

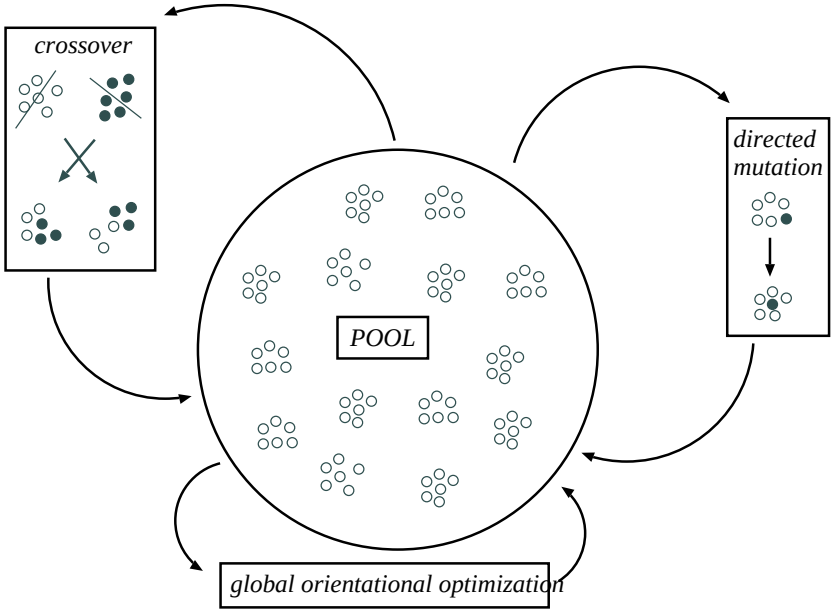
- design problem-specific crossover, exploiting near-separability
- local optimization vital but expensive; use loose thresholds initially
- "directed mutation" scans structurally similar minima
- strongly deceptive landscapes need secondary selection criteria: problem-specific niches

---

[1] B. Hartke, J. Phys. Chem. 97 (1993) 9973.
[2] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. 75 (1995) 288.
[3] B. Hartke, J. Comput. Chem. 20 (1999) 1752.

Parallel implementation[4]

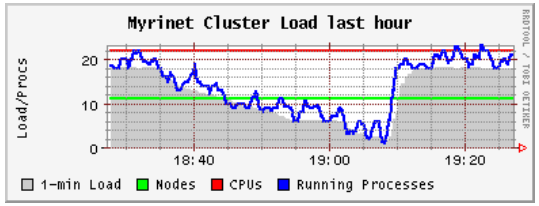
replace generational model by pool model:



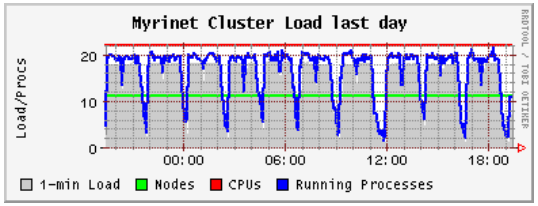
load on a 24-processor PC cluster:

old parallel generational model

load / 1 hour

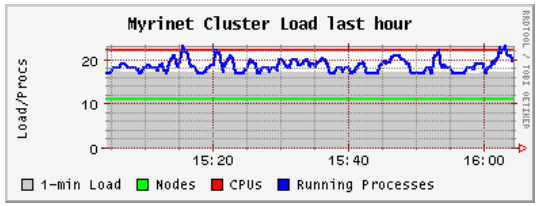


load / 1 day

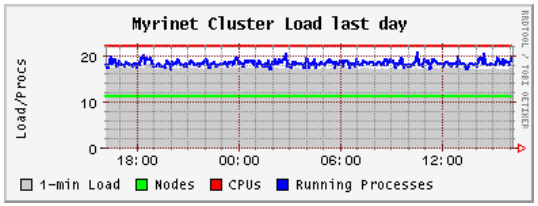


new parallel pool model

load / 1 hour



load / 1 day

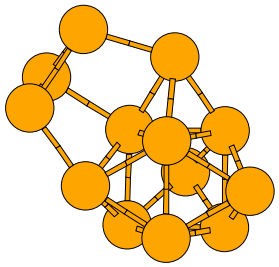


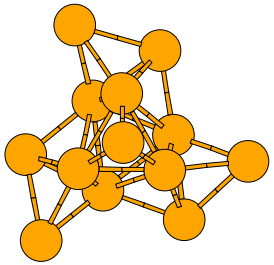
- no serial bottlenecks ⇒ perfectly even load balance during the whole run
- parallel efficiency independent of choices population size vs. # processes

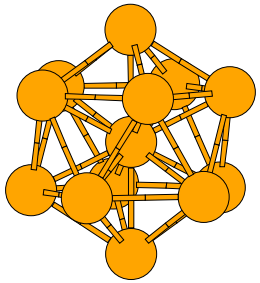
[4] B. Bandow and B. Hartke, J. Phys. Chem. A 110 (2006) 5809.

Global optimization of cluster structures: old stuff...

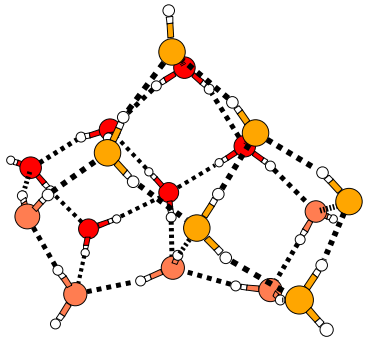
Hg_{13}:

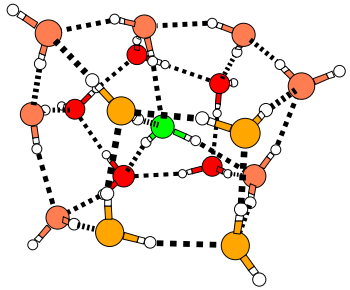


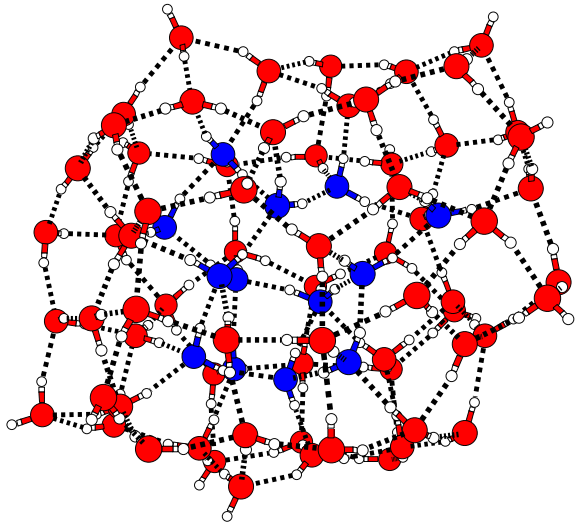




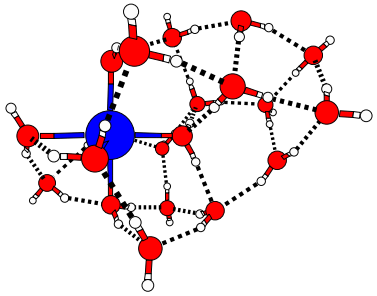
$(H_2O)_n$:

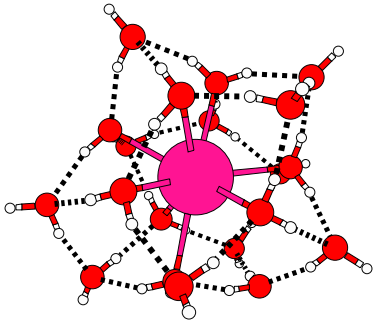


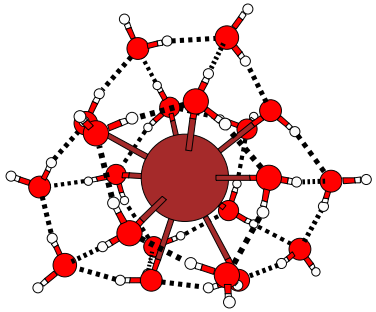




$(Na^+/K^+/Cs^+)(H_2O)_{20}$:







Challenges for global optimization in chemistry:

- larger clusters ⇐ algorithmic improvements needed
- strongly/arbitrarily mixed clusters
- of larger, flexible molecules
- easy access to force-calculation backends: from force-fields to ab-initio quantum chemistry
- prove to be useful in real-life contexts

⇒ development of the general global optimization program suite [5]

(by Johannes Dieterich)

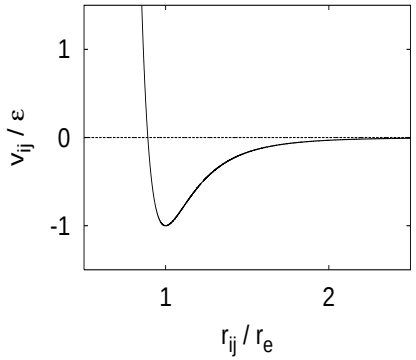
- object-oriented Java
- thread-based SMP parallelism
- MPI-based MPP parallelism
- collision detection
- dissociation detection (Warshall)
- active/passive internal/external coords
- some interal force fields
- with option for their system-specific global reparametrization
- force backends e.g. for
  - Molpro, Orca
  - MNDO, MOPAC, DFTB+
  - AMBER, NAMD, Tinker

[5] J. M. Dieterich and B. Hartke, Mol. Phys. 108 (2010) 279.

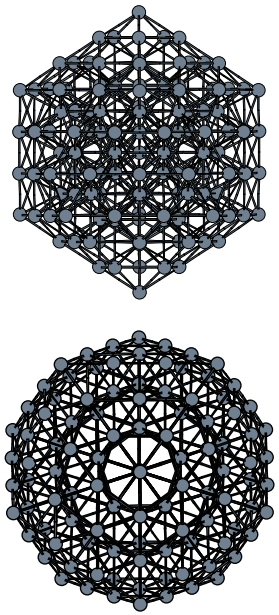
Strongly mixed clusters: from homogeneous to quinary LJ

reminder: homogeneous LJ clusters

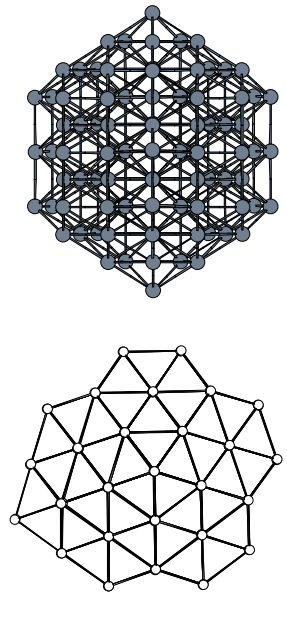
Lennard-Jones potential (imperfect model for rare gases/benchmark):
$v_{ij} = r_{ij}^{-12} - 2r_{ij}^{-6}$



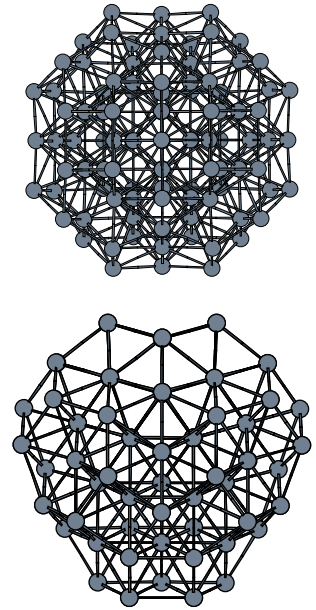
4 structural types known as global minima:



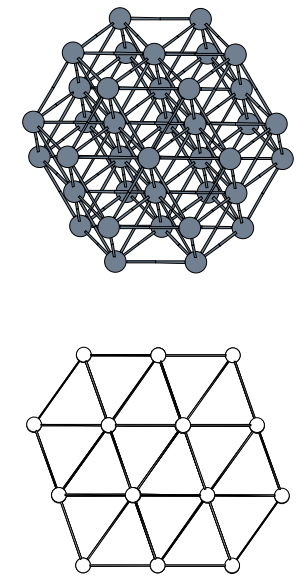
almost all $n \leq 309$
icosahedral



$n = 76 \pm 1, 103 \pm 1$
decahedral



$n = 98$
tetrahedral[6]

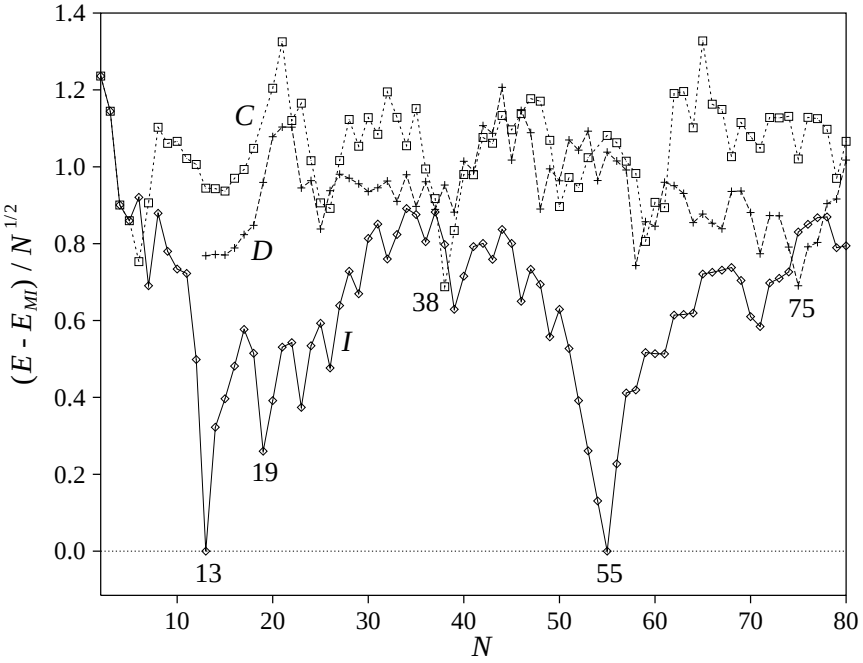


$n = 38,\ n \geq 750(?)$
fcc

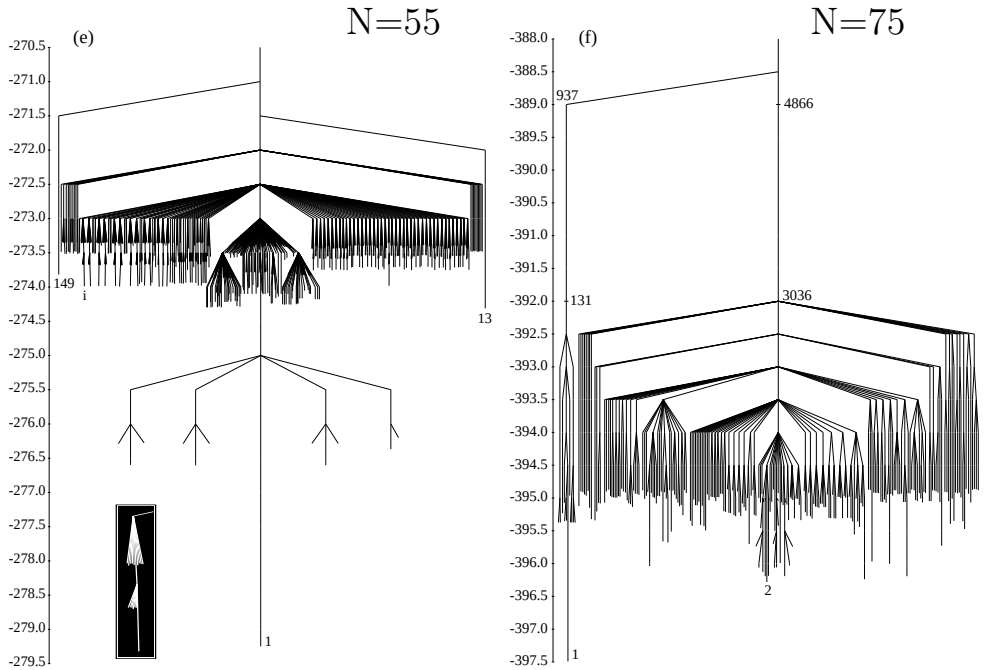
[6] R. H. Leary and J. P. K. Doye, Phys. Rev. E 60 (1999) R6320.

Competition between LJ structural types[7]

lowest-energy structure of each type,
at each cluster size N:



disconnectivity graphs of the lowest-energy minima:



(I: icosahedral, D: decahedral, C: fcc)

⇒ many LJ global minima are easy to find, but a few are very hard.

⇒ expectations for mixed LJ clusters:
sensitivity to structural transitions upon mixing may change with cluster size.

[7] J. P. K. Doye, M. A. Miller and D. J. Wales, J. Chem. Phys. 111 (1999) 8417.

Ab-initio fitted mixed LJ pair potentials [8]

generalized LJ(6,16,2) ansatz:

$$v_{ij} = 4\epsilon_{ij} \sum_{k=6,k+2}^{16} sign(\sigma_{k,ij}) \cdot \left(\frac{\sigma_{k,ij}}{r_{ij}}\right)^k \qquad (1)$$

with parameters globally fitted to CCSD(T)/aug-cc-pV5Z data.

Comparison: standard LJ(6,12,6) potentials

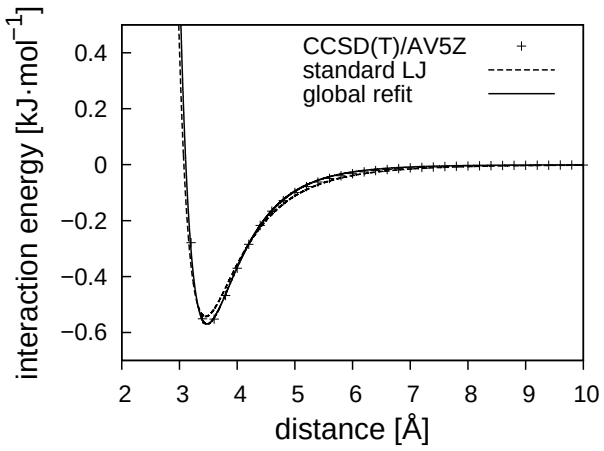
$$v_{ij} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad (2)$$

with Lorentz-Berthelot mixing rules:

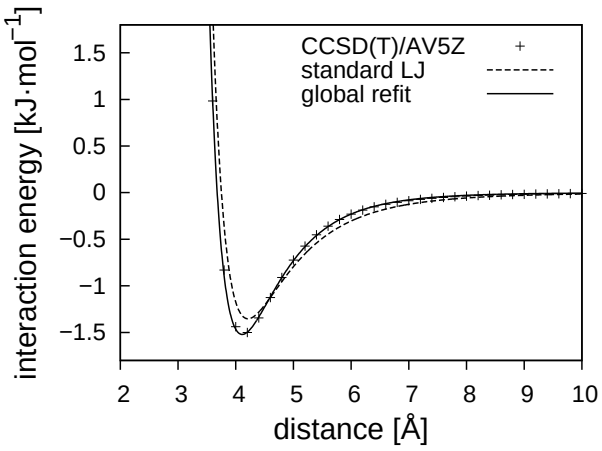
$$\epsilon_{ij} = \sqrt{\epsilon_i \cdot \epsilon_j} \qquad (3)$$

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \qquad (4)$$

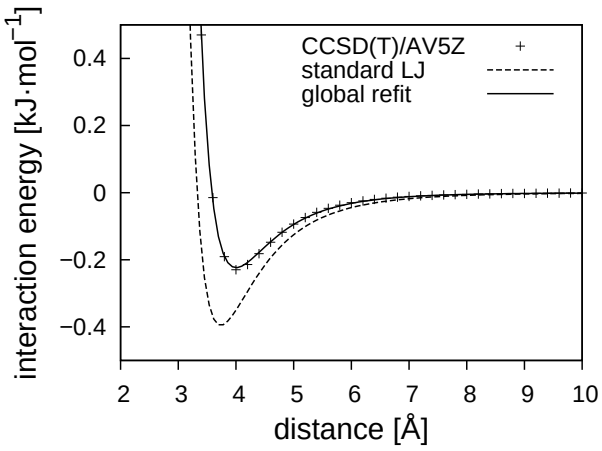
exemplary comparisons for some pair potentials:



Ne–Ar



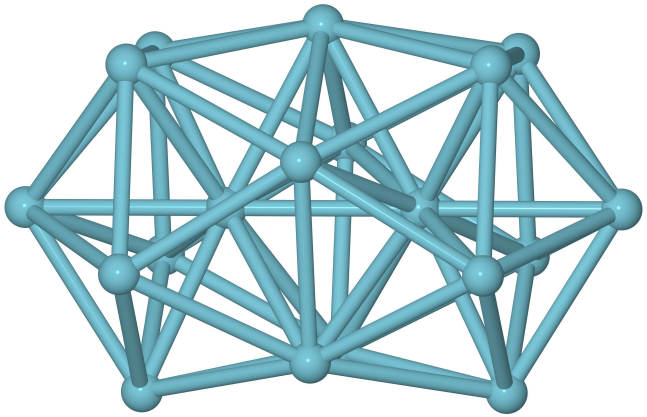
Ar–Xe



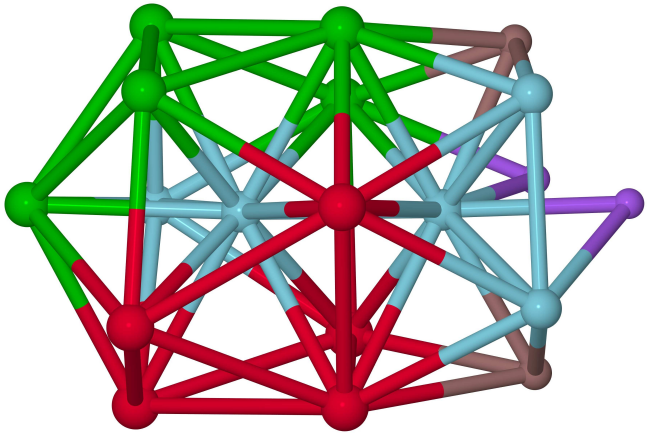
He–Xe

[8] J. M. Dieterich and B. Hartke, manuscript in preparation.

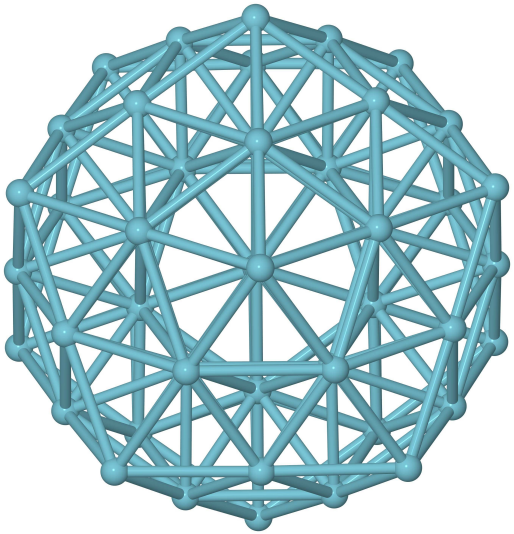
Structural stability under strong mixing [9]



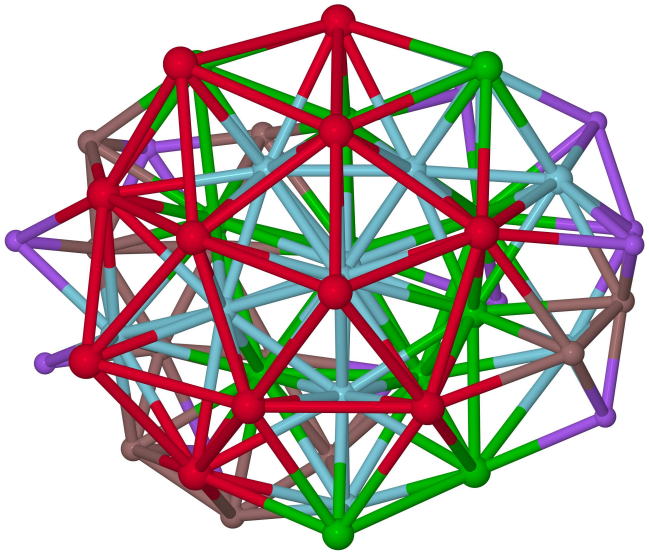
Ar_{19}



$Ar_5Kr_5Xe_5He_2Ne_2$



Ar_{55}

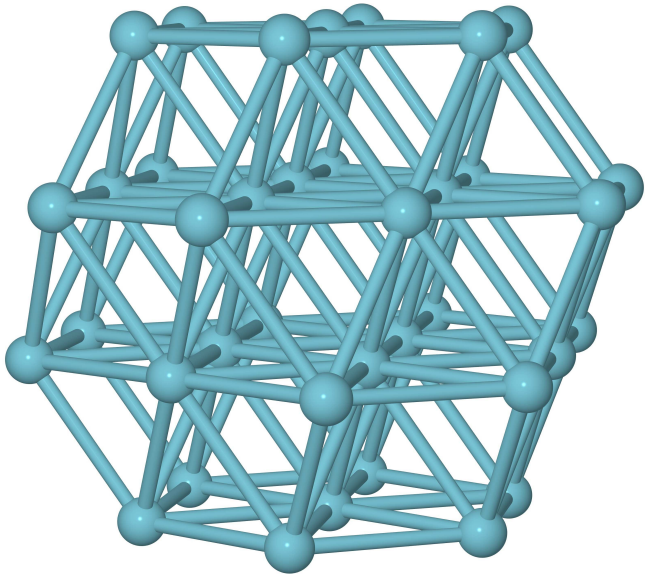


$Ar_{11}Kr_{11}Xe_{11}He_{11}Ne_{11}$

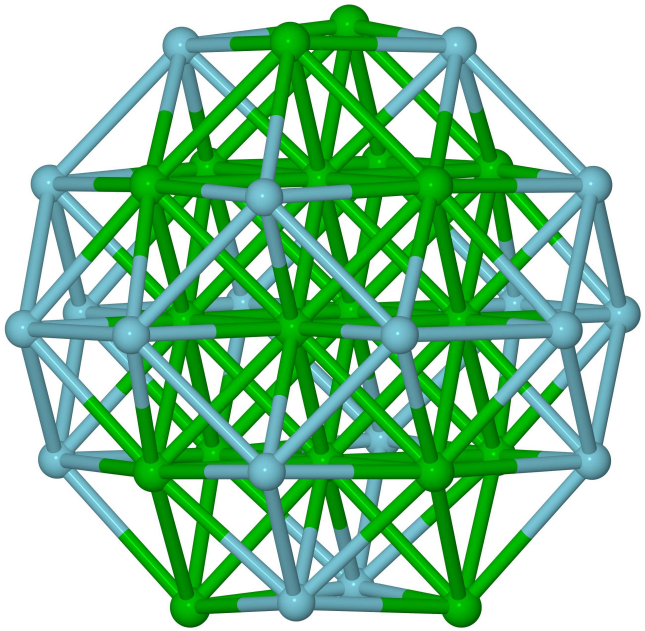
(global minima)

[9] J. M. Dieterich and B. Hartke, manuscript in preparation.

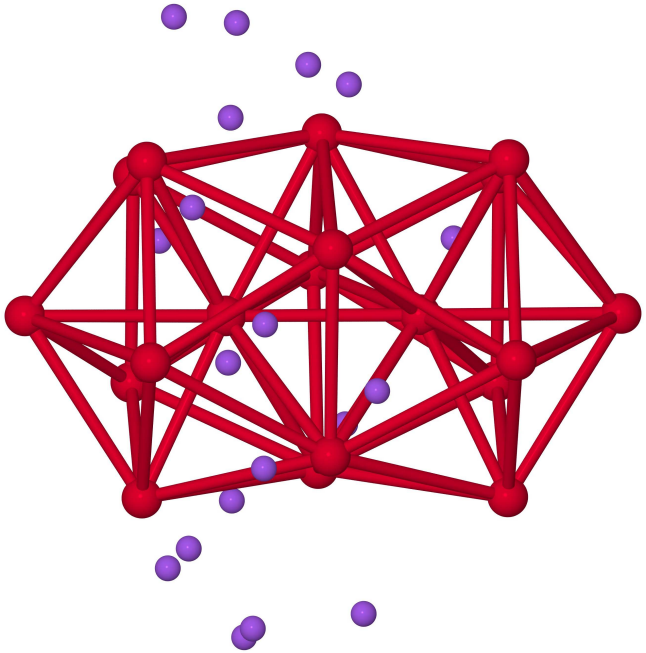
Mixing-induced structural transitions in LJ_{38} [10]



Ar_{38}
fcc



$Ar_{19}Kr_{19}$
still fcc

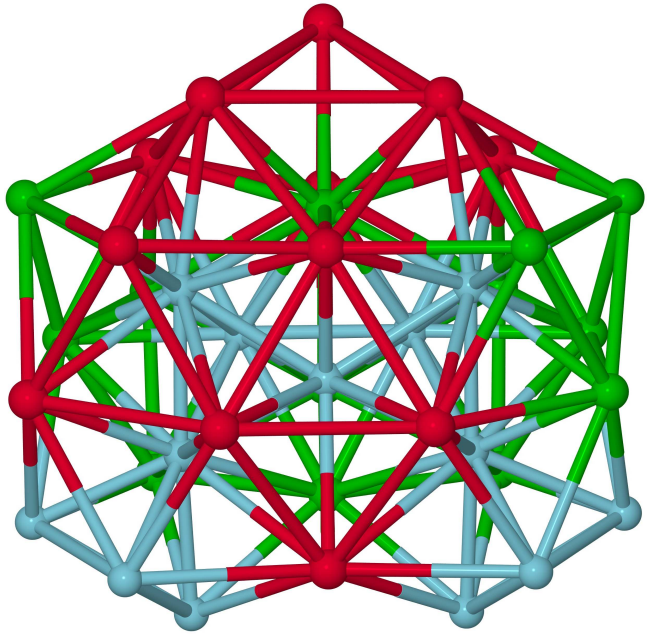


$He_{19}Xe_{19}$
core-shell icosahedral

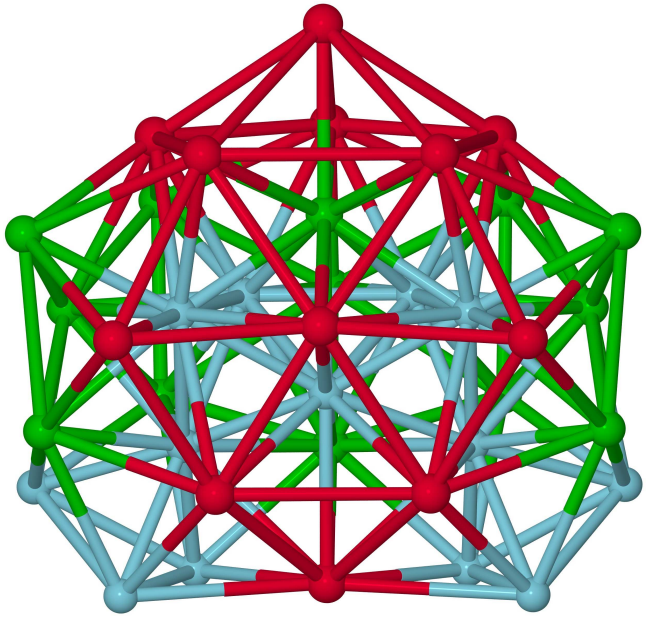
[10] J. M. Dieterich and B. Hartke, manuscript in preparation.

Mixing-induced structural transitions in LJ_{38} [11]

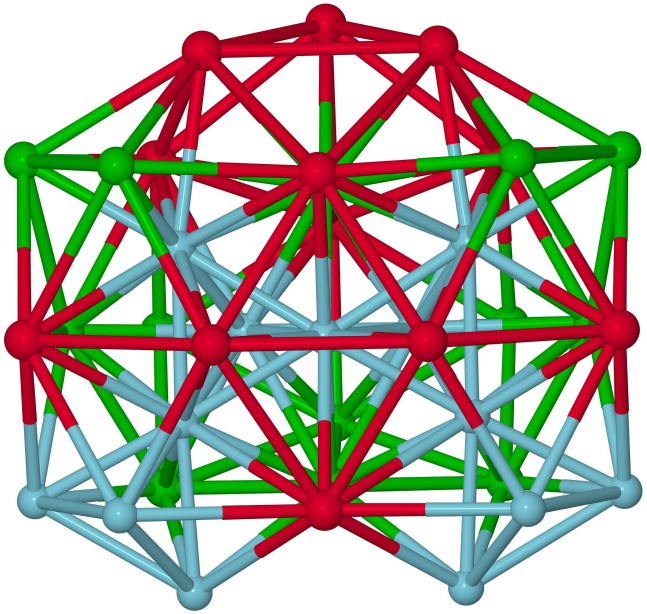
$Ar_{13}Kr_{13}Xe_{12}$: strongly icosahedrally dominated



-280.4455 kJ/mol
(global minimum, rank 0)



-280.2024 kJ/mol
(rank 1)

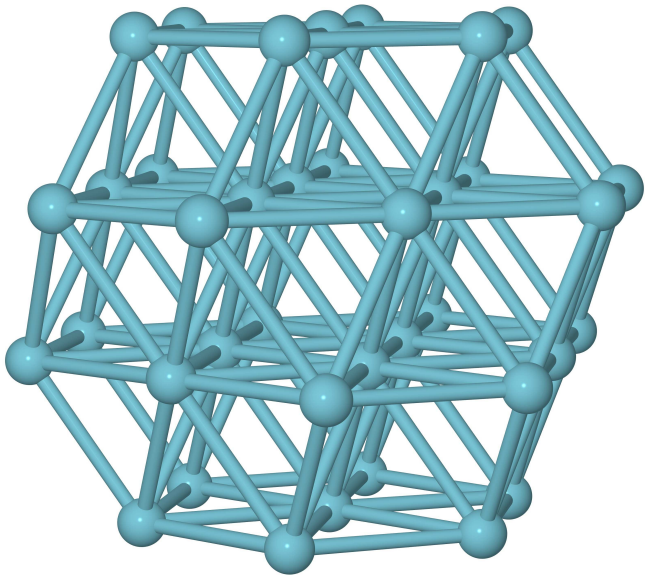


-280.1620 kJ/mol
(rank 2)

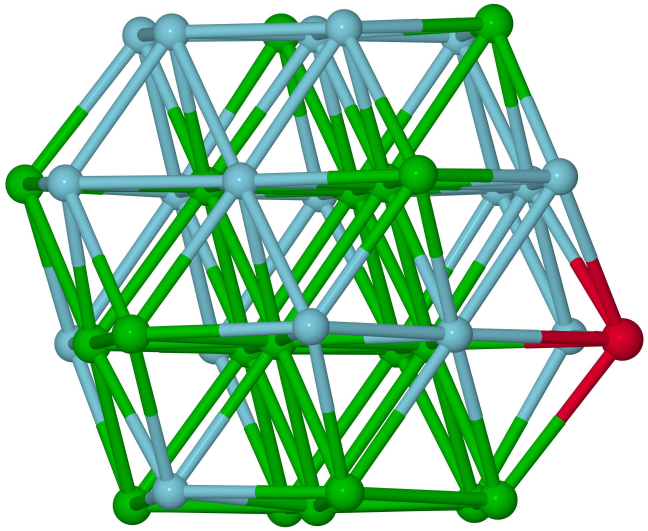
[11] J. M. Dieterich and B. Hartke, manuscript in preparation.

Mixing-induced structural transitions in LJ_{38} [12]

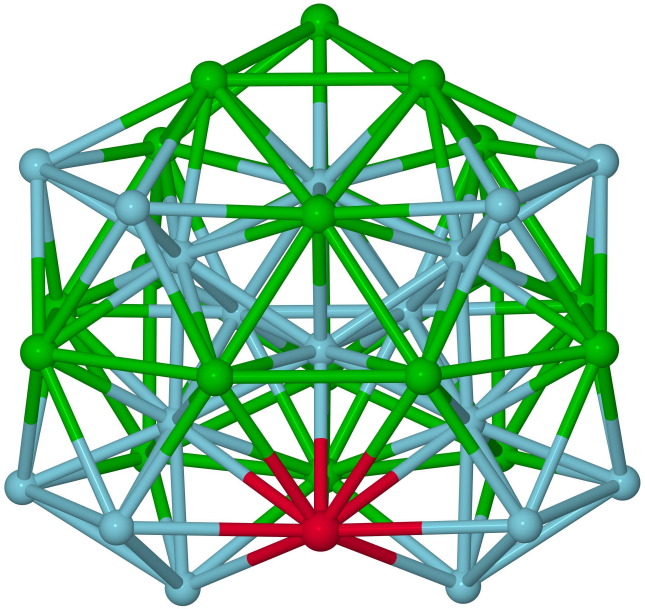
$Ar_{18}Kr_{19}Xe$: global minimum icosahedral, but low-energy fcc present



Ar_{38}



-242.5240 kJ/mol
(best fcc)

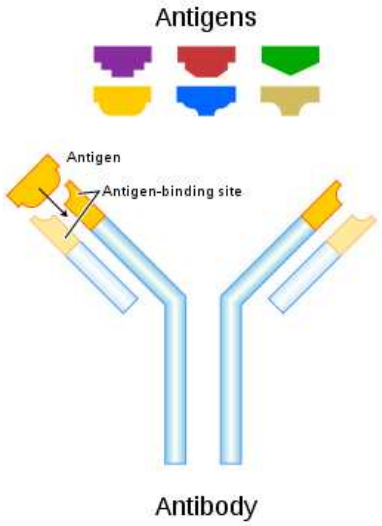


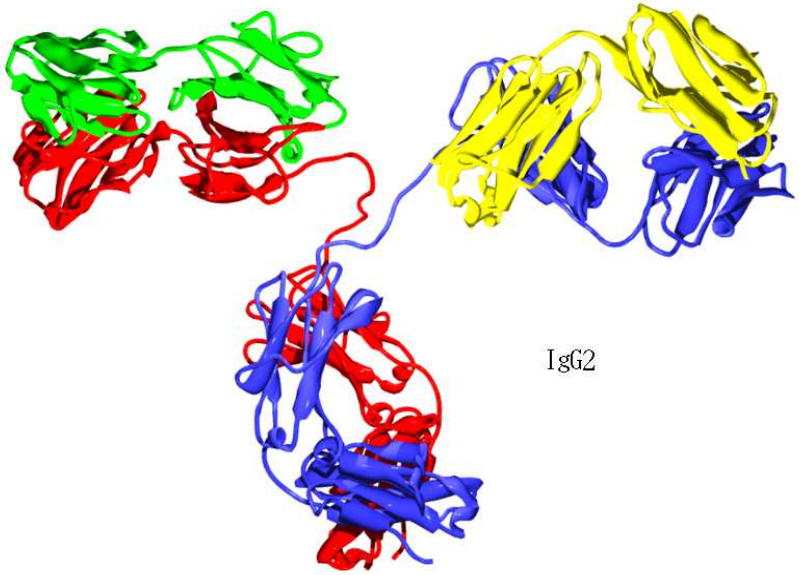
-248.8864 kJ/mol
(global minimum)

[12] J. M. Dieterich and B. Hartke, manuscript in preparation.

Mixed clusters of flexible molecules: Theoretical chemistry in medicine

adaptive immune system:

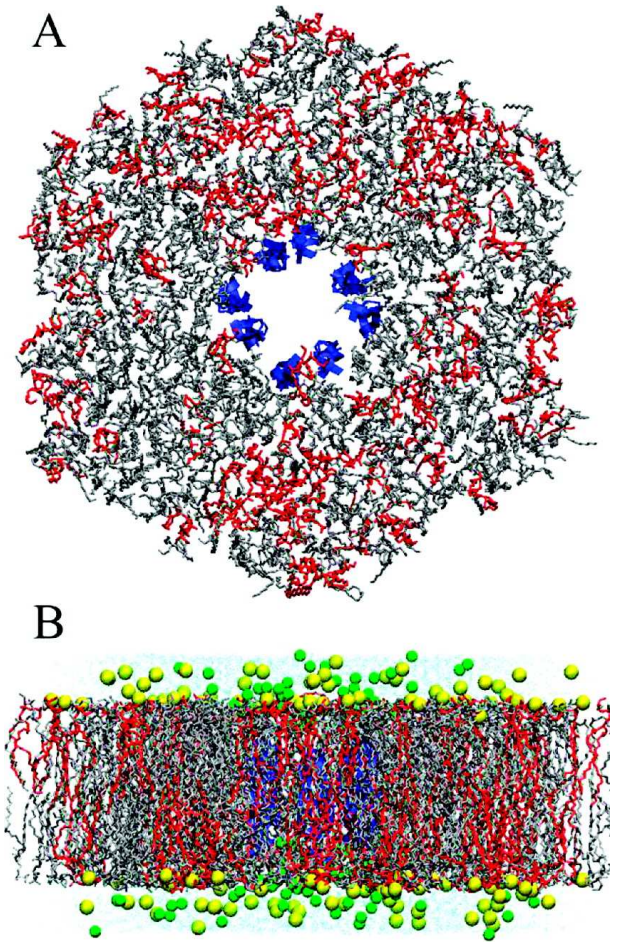




IgG2

innate immune system:

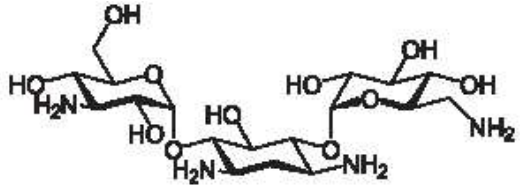
pathogen-associated molecules (PAMs) $\rightarrow$
antimicrobial peptides (AMPs) $\rightarrow$ cell wall damage:



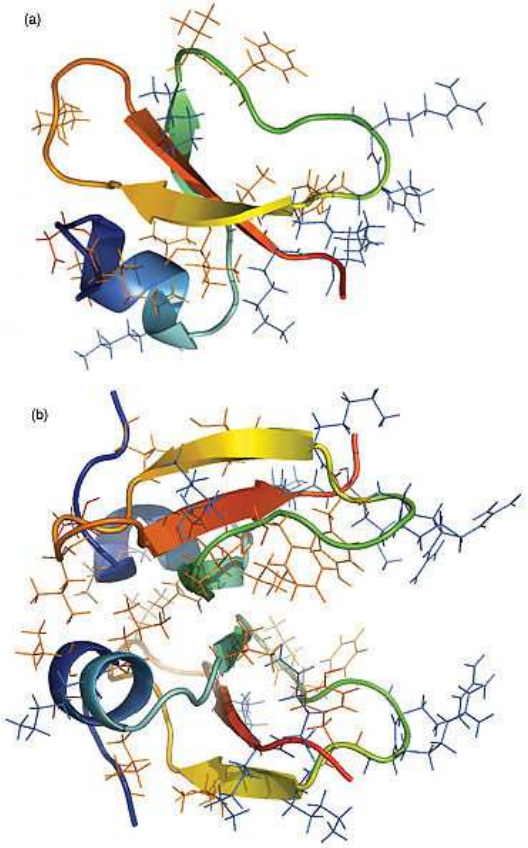
protegrin-1 octamer pore in lipid bilayer

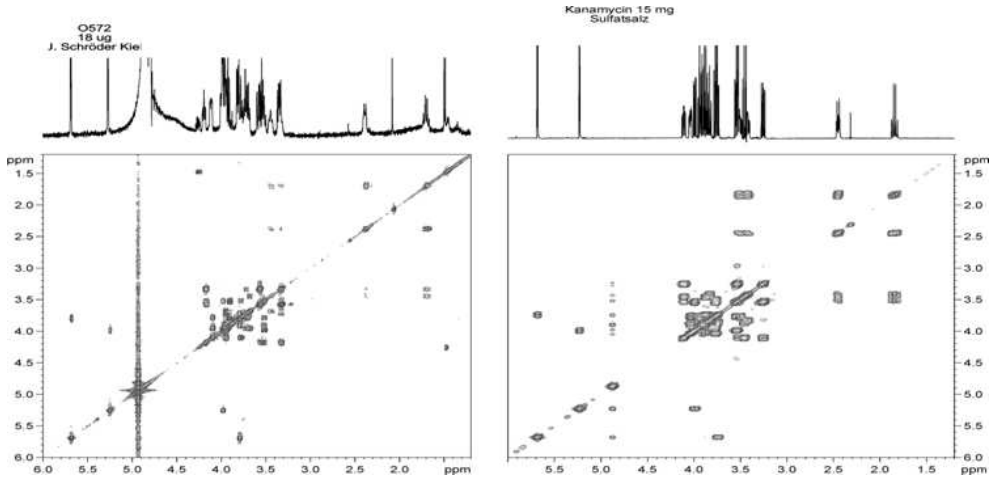
Pseudomonas aerigunosa $\Rightarrow$ unknown PAM

J.-M. Schröder (dermatology, Univ Kiel):
clusters of Kanamycin A (with Na^+, K^+, Cu^{2+})

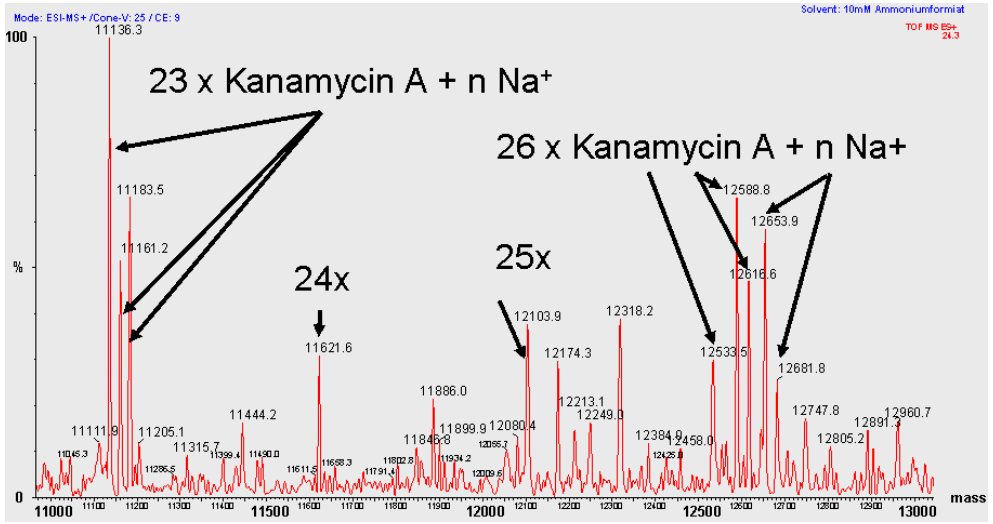


induce human β-defensin in epithelial cells



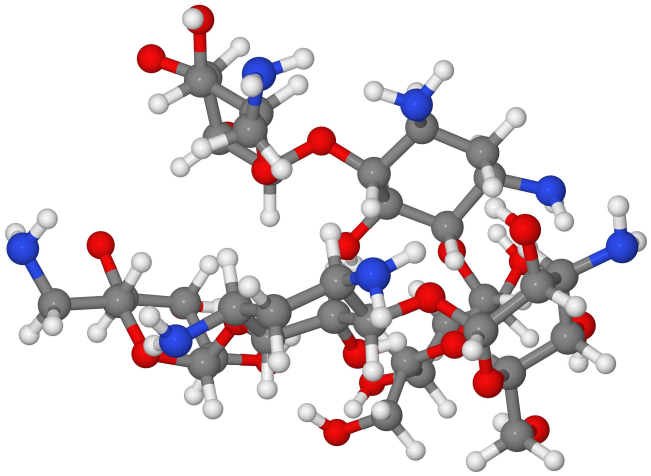


COSY-^{1}H-NMR of active PAM (left), Kanamycin A (right)

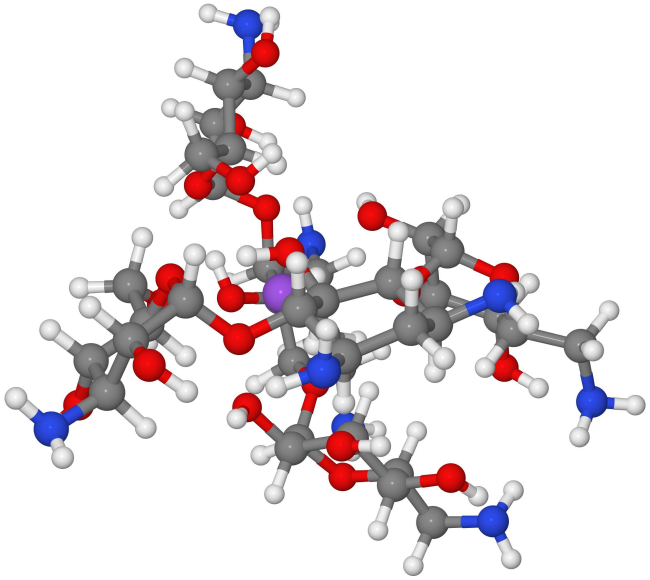


ESI-MS analysis of active PAM (with some assignments)

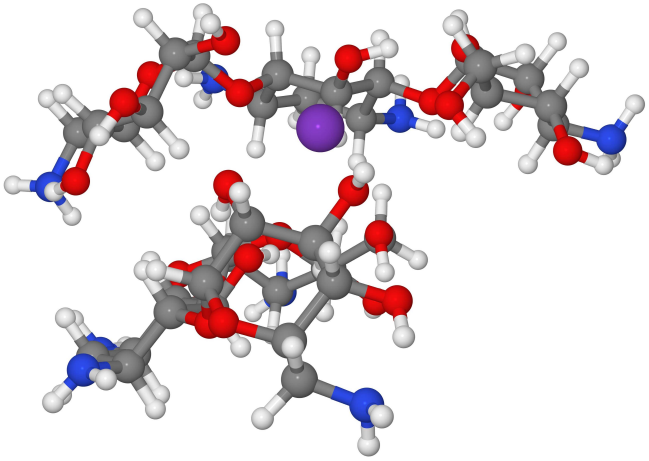
global structure optimization of Kanamycin-A dimers [13]



$(\text{Kanamycin A})_2$



$(\text{Kanamycin A})_2\text{Na}^+$

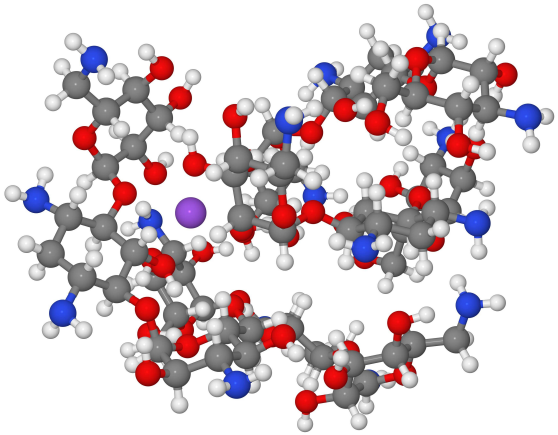


$(\text{Kanamycin A})_2\text{K}^+$

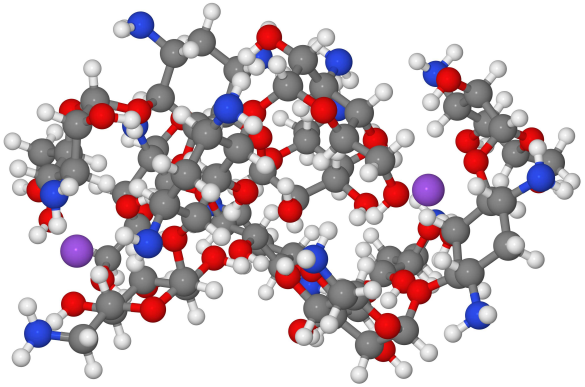
- exp. and theory agree: preference of Na^+ over K^+
- predicted IR and NMR spectra show experimentally accessible signatures of aggregation

[13] J. M. Dieterich, U. Gerstel, J.-M. Schröder and B. Hartke, manuscript in preparation.

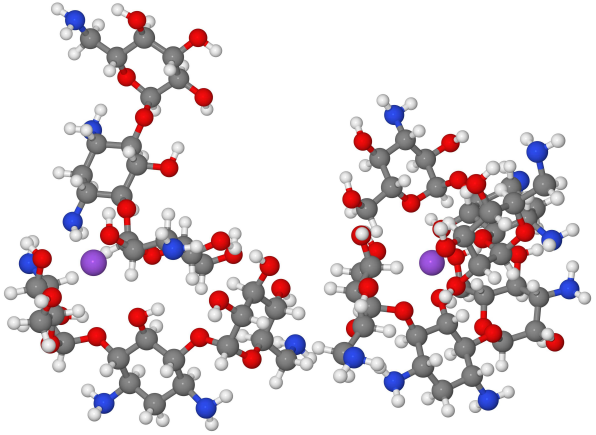
larger Kanamycin-A clusters [14]



$(\text{Kanamycin A})_4\text{Na}^+$
very stable



$(\text{Kanamycin A})_2\text{Na}_2^+$
barely stable



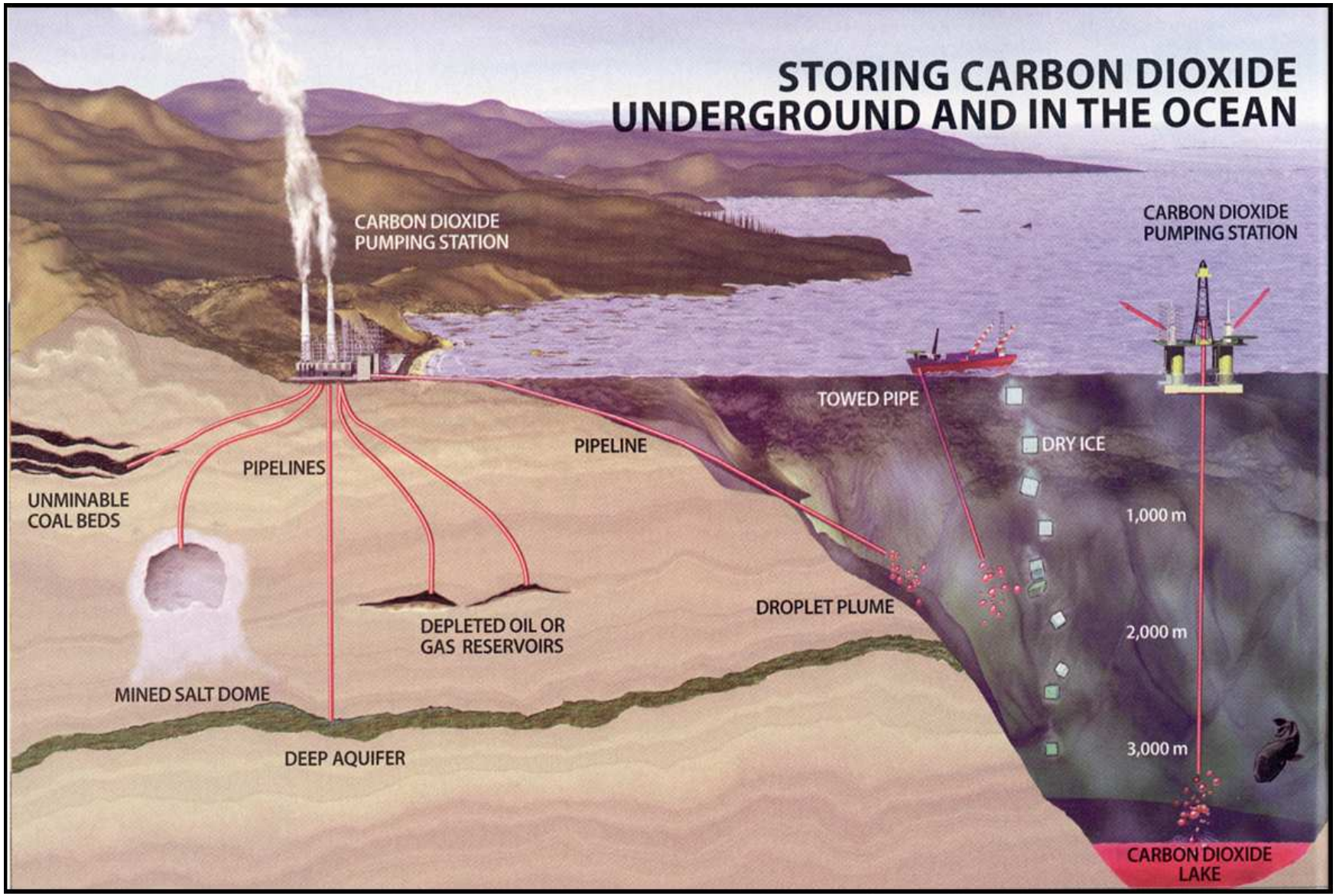
$(\text{Kanamycin A})_2\text{Na}_2^+$
dissociated

future work:

- larger clusters, with more/different ions
- prediction of which minor constitutional isomer is likely to be present experimentally
- design of low-molecular weight AMP-inducing agents

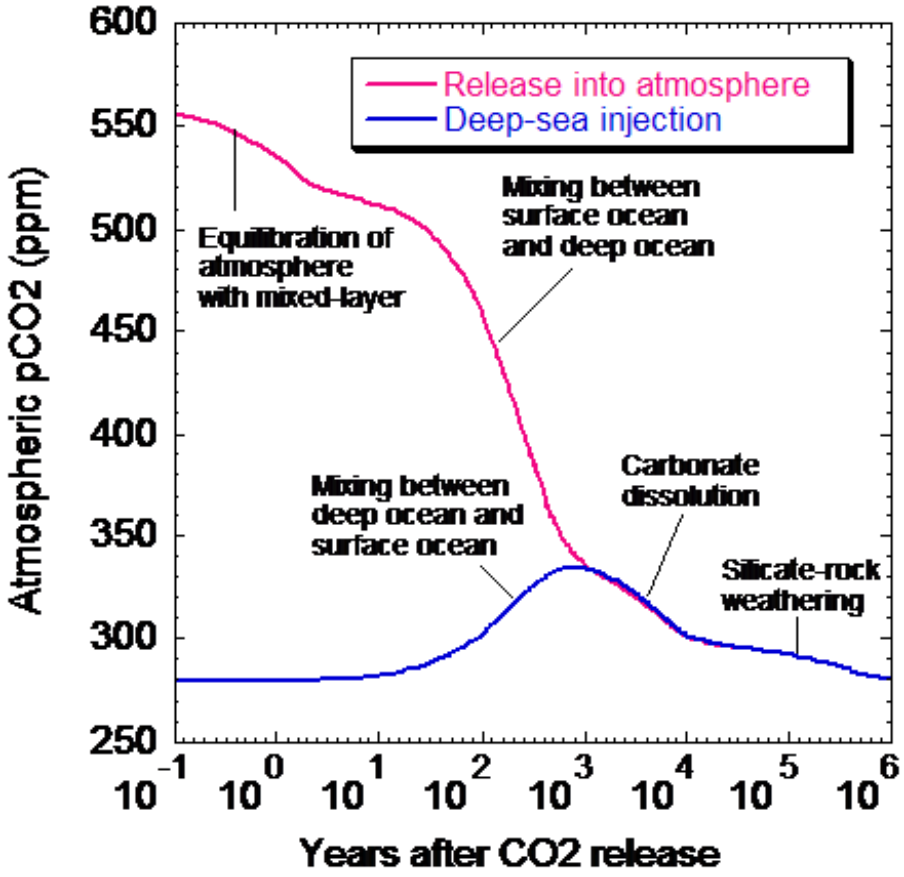
[14] J. M. Dieterich and B. Hartke, Mol. Phys. 108 (2010) 279.

Cluster of Excellence: Future Ocean, project A5: CO_2 sequestration



Future Ocean A5: CO_2 sequestration

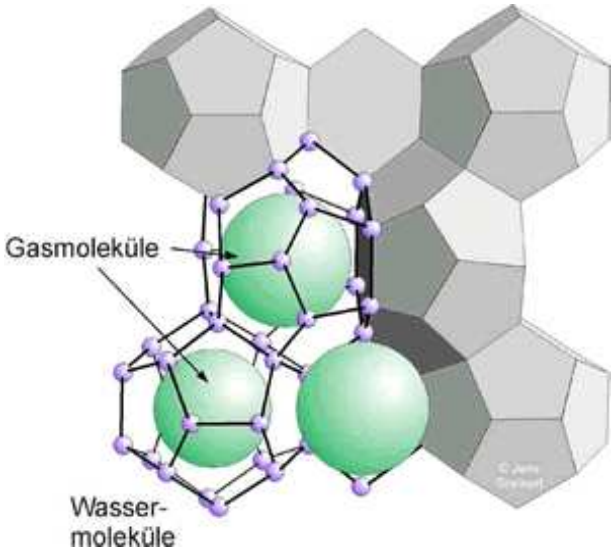
Global simulation of CO_2 distribution
(Ken Caldeira, Lawrence Livermore National Lab):
instantaneous addition of a mass of CO_2 equal to the pre-industrial amount, introduced either to the atmosphere or the deep ocean.

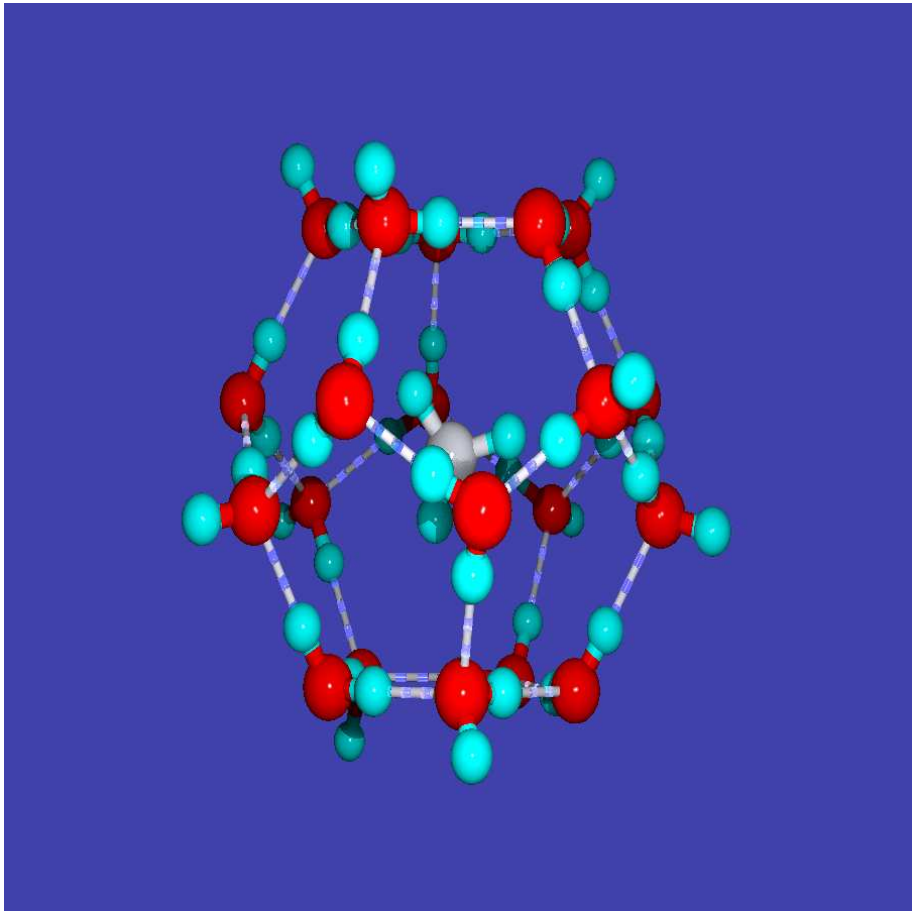


- equilibrium CO_2 uptake by the ocean 90% anyway: same state after relaxation of the system (note time scale: 10^6 years!)
- *but "benefit":*
  ocean sequestration avoids atmospheric "detour" of CO_2 and hence (hopefully) most of its adverse climatic effects
  (on a time scale of 100 years)

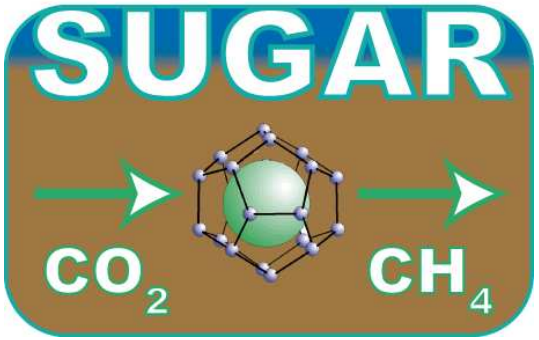
Global optimization in marine/climate research: clathrate hydrates

- water cages around CH_4, CO_2,...

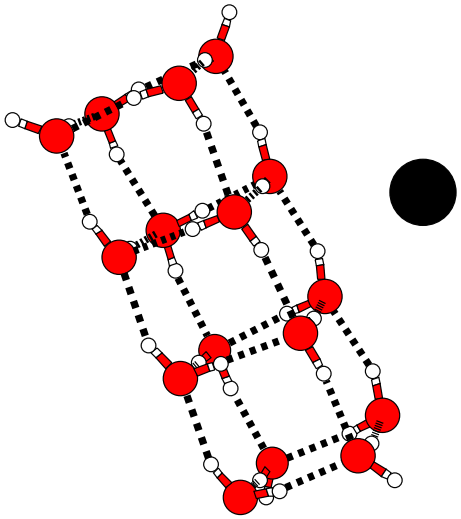


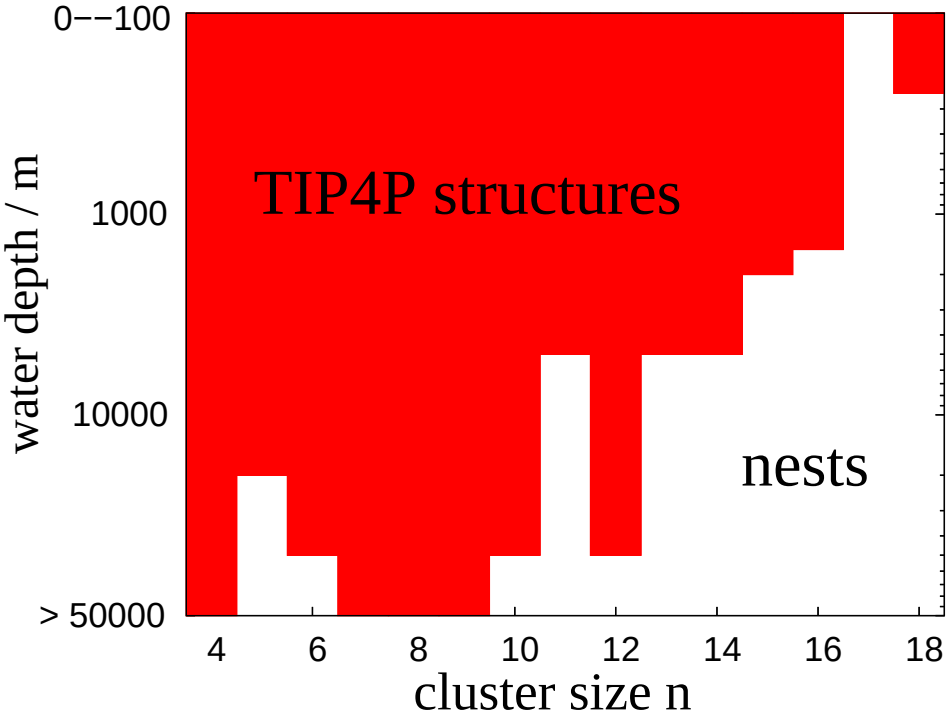


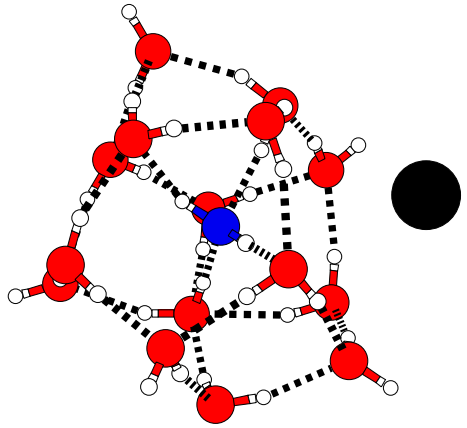
- stable at low T & high p $\Rightarrow$ large deposits of CH_4 hydrate on the seafloor (continental shelves)
- carbon content $\approx$ fossil fuels
- possibility to get rid of CO_2 (sequestration, CCS)
- $\Rightarrow$ current commercial dream:



Methane hydrate clusters: global optimization under pressure [15]

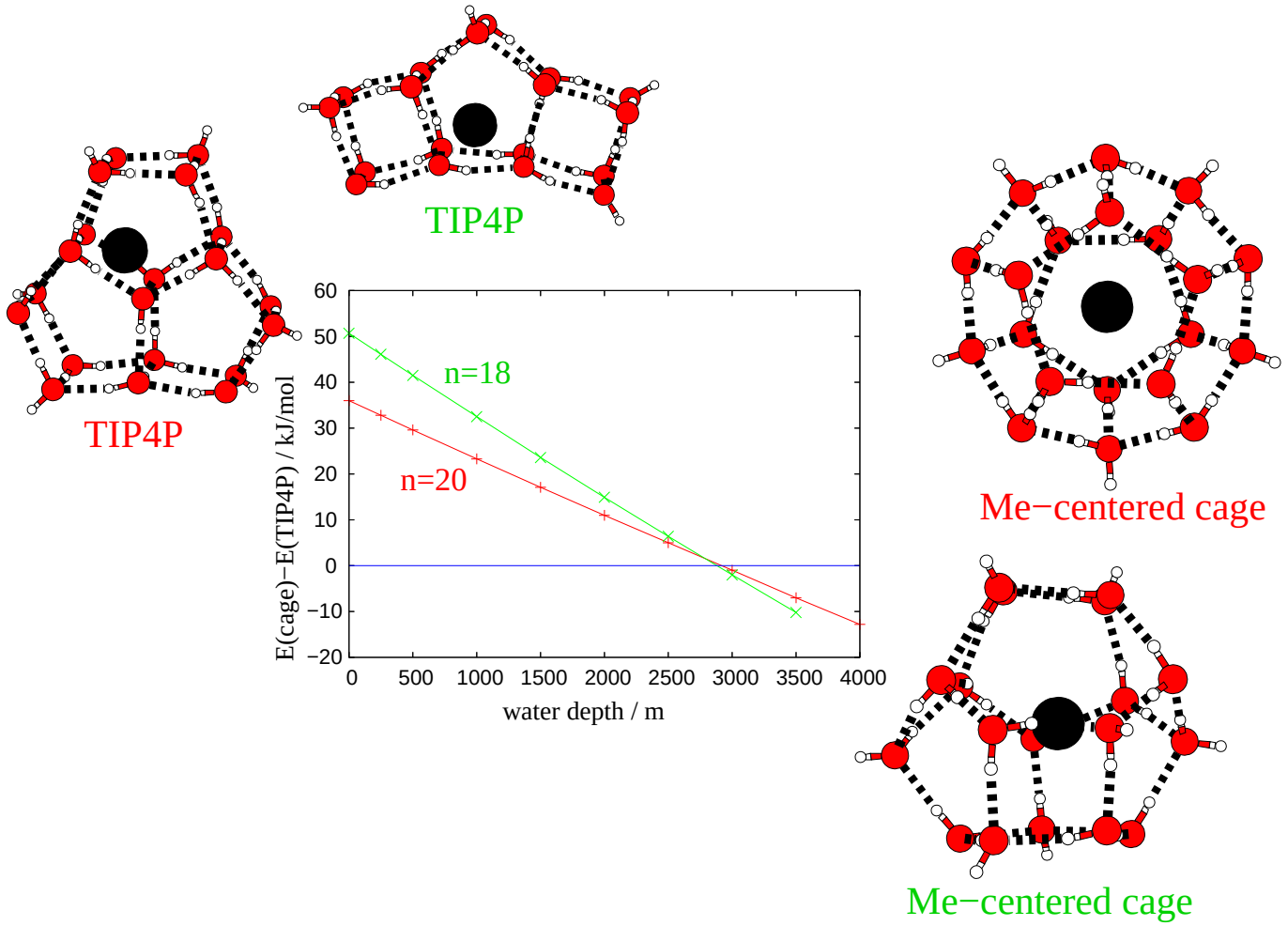






[15] B. Hartke, J. Chem. Phys. 130 (2009) 024905

outlook to larger clusters:

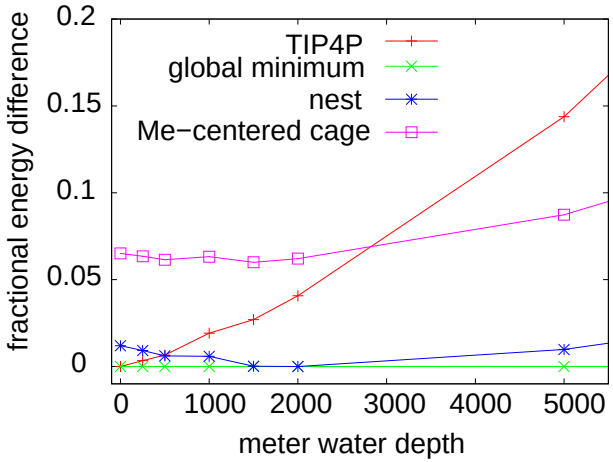


transition to Me-centered cages expected for n≈20 and 3000m

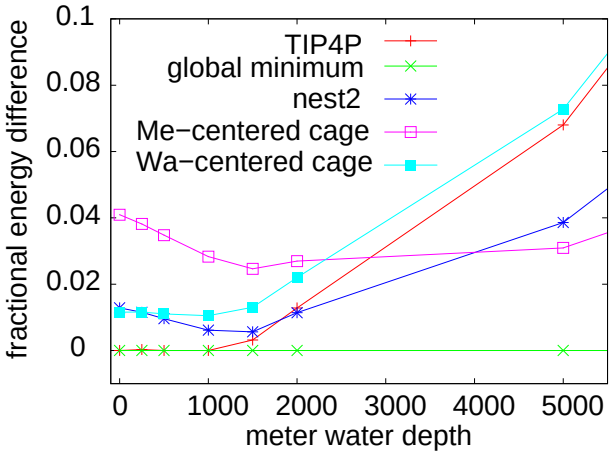
but this is wrong! (local optimization of given structures, no global optimization)

what really happens at n=18,20 after global optimization:

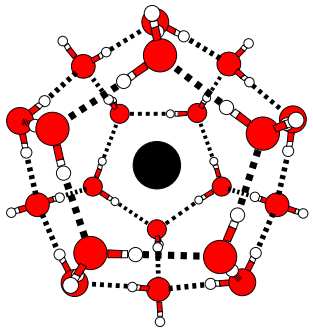
n=18



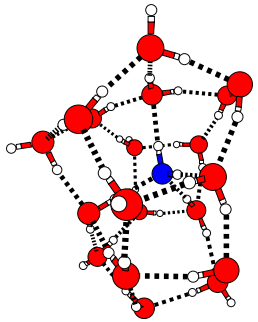
n=20



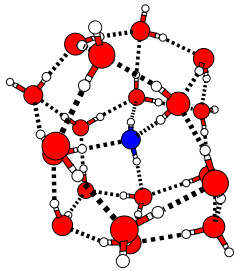
n=20:



5^{12} cage



nest

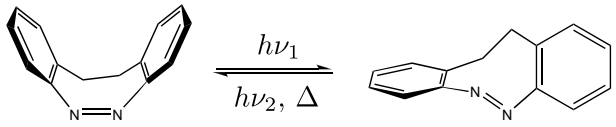


water-centered cage

There *is* a crossover between the TIP4P structures and clathrate-like Me-centered cages at 3000m — but global optimization finds many better nests.

Global optimization for the design of photochemically switchable molecules

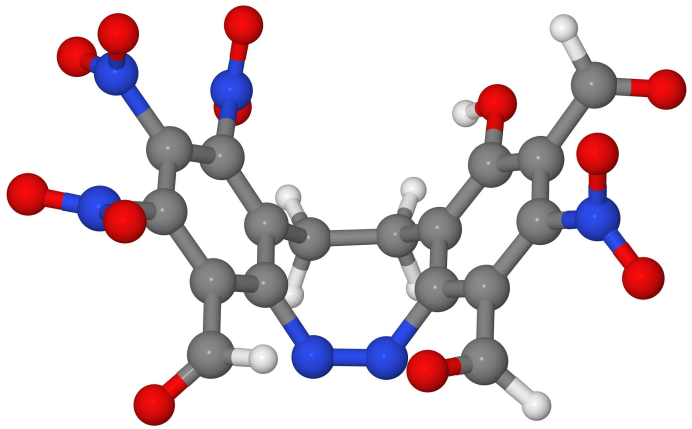
In recent work [16], a bridged azobenzene derivative was found to have superior photophysical properties compared to the parent compound:



1Z **1E**

Application of OGOLEM to globally optimize substituent patterns (8 allowed types);
aim: move absorption maxima from 425nm (cis) / 579nm (trans)
to the easily accessible 405nm / 532nm (commercial laserpointers).

Optimized compound [17]: multiple nitro and carbonyl groups, absorption maxima: 405nm / 532nm

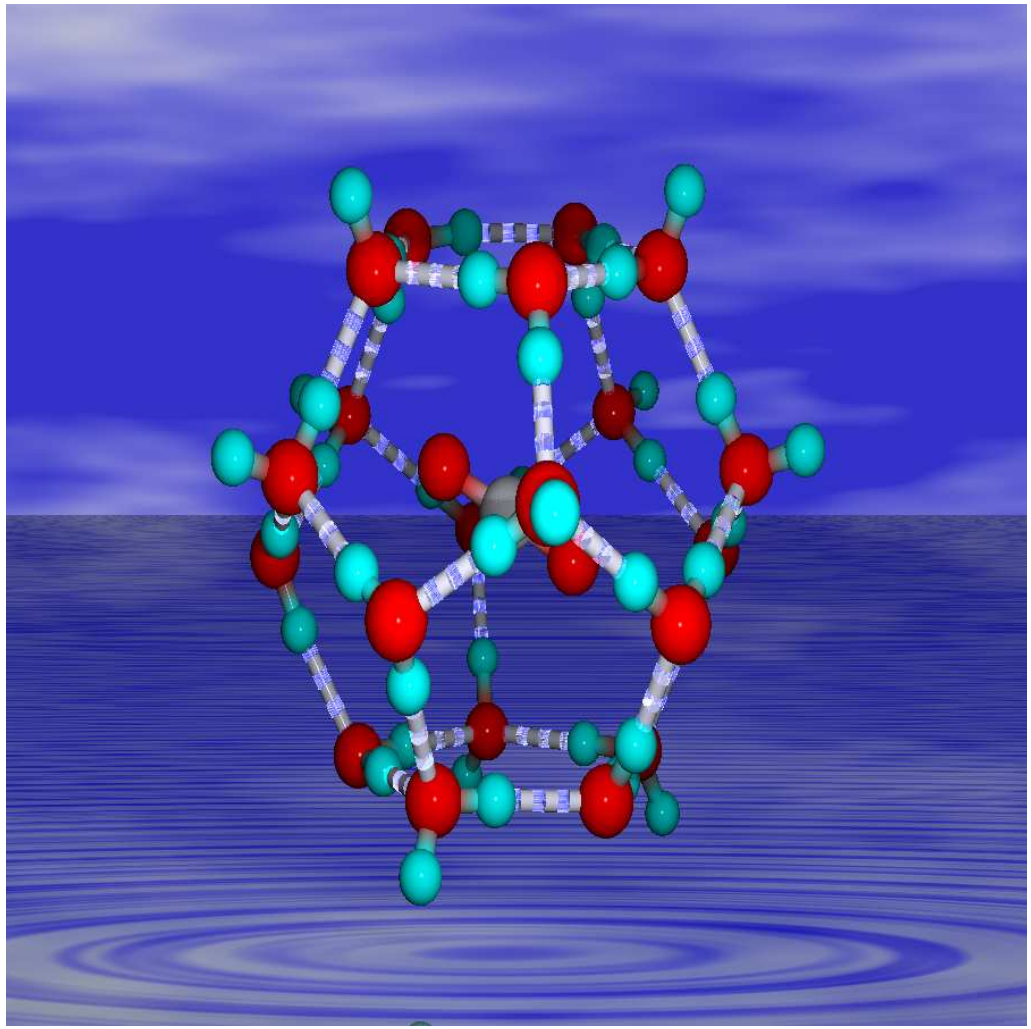


[16] R. Siewertsen, H. Neumann, B. Buchheim-Stehn, R. Herges, C. Näther, F. Renth, and F. Temps, J. Am. Chem. Soc 131 (2009) 15594.

[17] N. O. Carstensen, J. M. Dieterich and B. Hartke, manuscript in preparation.

acknowledgements:

- J. Dieterich, Ole Carstensen (AK Hartke)
- J.-M. Schröder (dermatology, Univ Kiel)
- A. Neumaier (Mathematics, Univ Wien)
- S. T. Bromley (Univ Barcelona)
- B. Abel (MPI and Univ Göttingen)
- A. Lüchow (RWTH Aachen)
- V. Buch † (Jerusalem)
- J. W. Ponder (St.Louis/Missouri)
- J. M. Lisy (Urbana-Champaign/Illinois)
- U. H. E. Hansmann (Houghton/Michigan)



cpu time: computing center Univ Kiel, North-German supercomputing alliance (HLRN Berlin/Hannover)
€€€: DFG (Normalverfahren, SFB, Exzellenzcluster);
Innovationsfonds des Landes Schleswig-Holstein (Numerik-Zentrum Uni Kiel)

Basic problem of global cluster structure optimization: computational complexity

atomic clusters:
possibly exponential increase in number of local minima with cluster size:

- empirical scaling formula for number of minima in Lennard-Jones clusters (Hoare/McInnes):

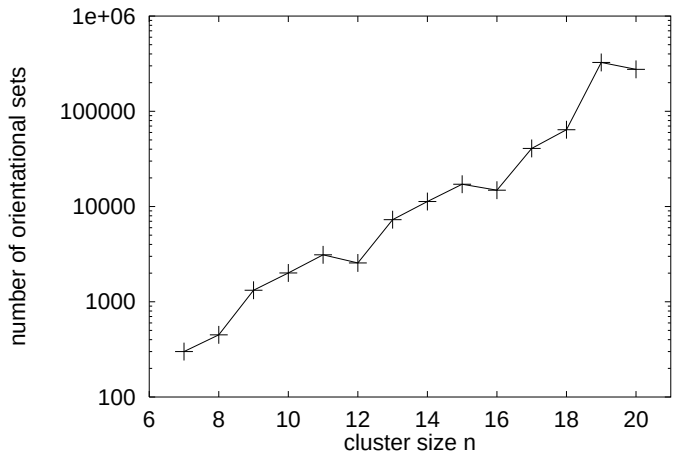
$$\exp\left(0.028\,n^2 + 0.3572\,n - 2.5176\right) \qquad (5)$$

  $\Rightarrow$ 98 LJ atoms have 10^{130} local minima

- other estimates vary wildly, e.g.:
  98 LJ atoms have 10^{40} local minima

- abstract proof: NP-hard problem[18],
  but proof not quite correct[19]

molecular clusters:
3 additional orientational degrees of freedom per particle.
example: purely orientational subspace of water clusters[20]



$\Rightarrow$ combination of two interdependent exponentially scaling problems

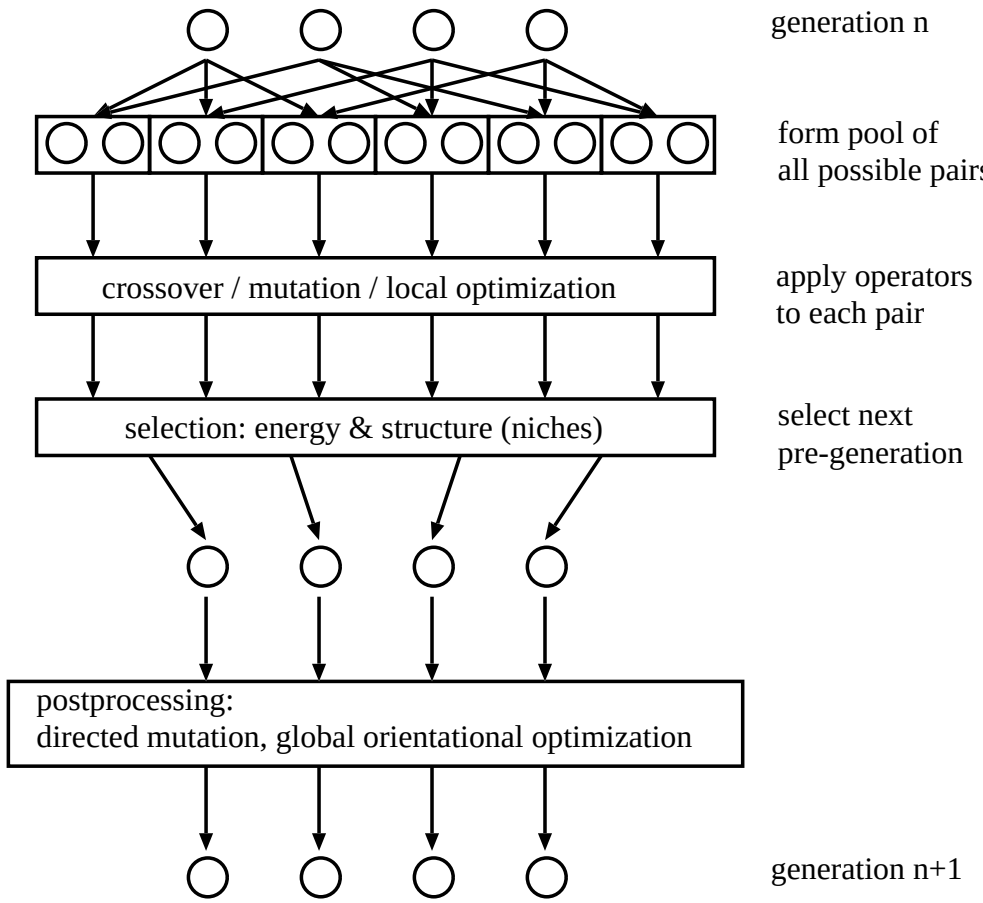
- deterministic global optimization is catching up, but still not applicable to clusters of nontrivial size
  $\Rightarrow$ alternative: stochastic-heuristic algorithms

- $\Rightarrow$ clusters of interesting sizes can be studied,
  but guarantee for finding global minima is lost.

---

[18] L. T. Wille and J. Vennik, J. Phys. A 18 (1985) L419.
[19] A. Srivastav, Univ. of Kiel, personal communication.
[20] B. Hartke, Z. Phys. Chem. 214 (2000) 1251.

Global optimization by Evolutionary Algorithms [21] [22] [23]



important aspects:

- design problem-specific crossover, exploiting near-separability
- local optimization vital but expensive; use loose thresholds initially
- "directed mutation" scans structurally similar minima
- strongly deceptive landscapes need secondary selection criteria: problem-specific niches

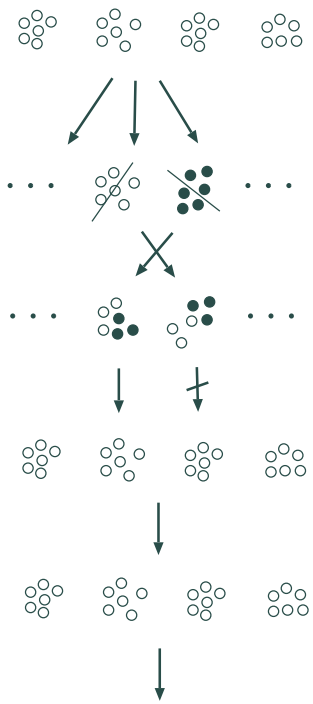
[21] B. Hartke, J. Phys. Chem. 97 (1993) 9973.
[22] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. 75 (1995) 288.
[23] B. Hartke, J. Comput. Chem. 20 (1999) 1752.

Parallel implementation[24]

properties of the algorithm

- local optimizations need >90% of total CPU time;
- but they are independent of each other, and
- data communication demands are very low.

⇒ "embarrassingly parallel":



realization:

- Fortran90 with simple, explicit standard-MPI calls;
- "master-slave" model
- time for local optimizations varies
  → implicit load-balancing by # locOpt ≫ # processes

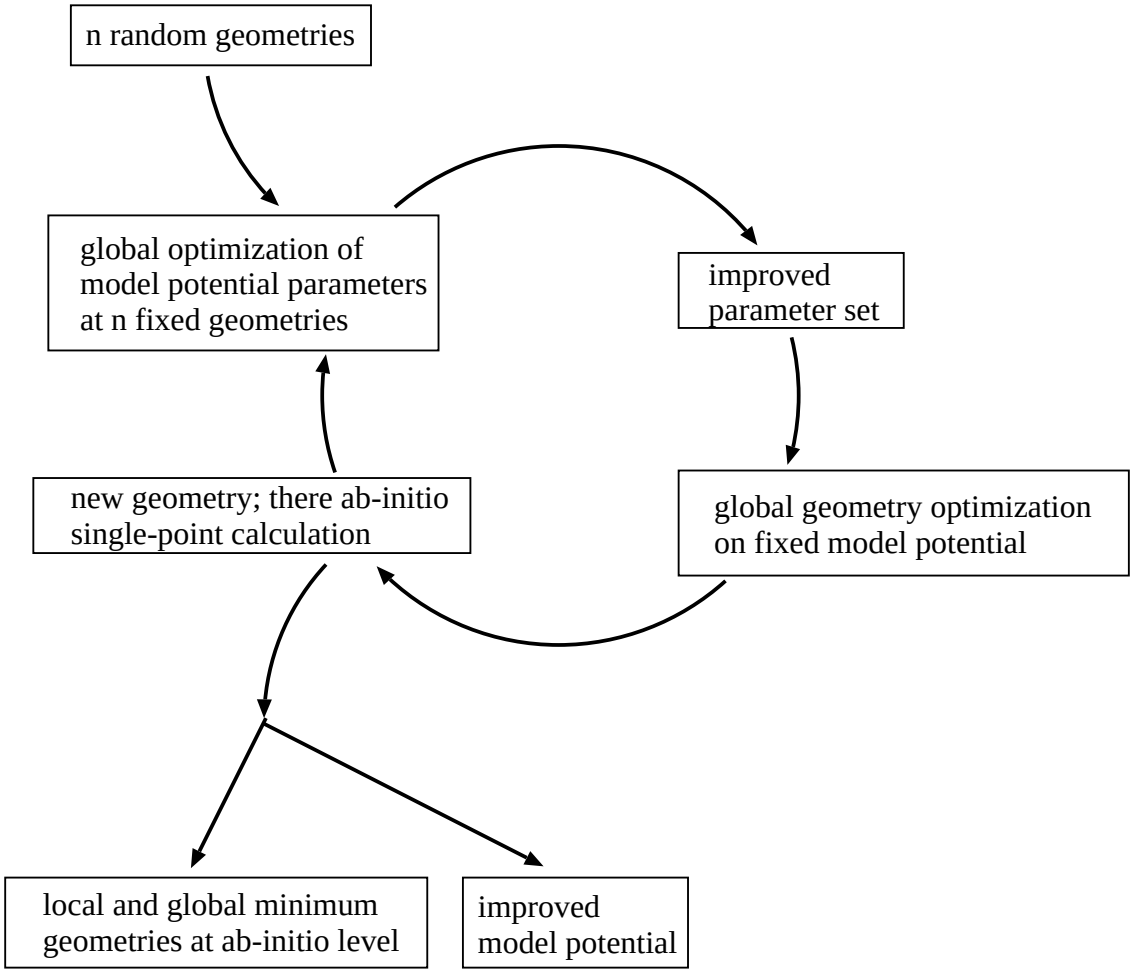
Problem: lengths of local optimization tasks vary, but serial bottlenecks in generational scheme need synchronization of all processes ⇒ significant idle times in most processes

[24] project at the national computer center HLRN (Berlin/Hannover)

Global geometry optimization of clusters on ab-initio potentials, guided by empirical potentials

Problem:
Every known heuristic is too expensive on ab-initio potentials.

Solution:[25]
Globally optimize cluster geometries on a model potential that is simultaneously fit to ab-initio single-point calculations (*surrogate function optimization*):



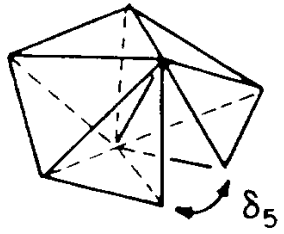
work in progress:

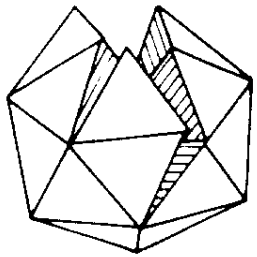
- inclusion of gradient information;
- surrogate function with smooth interpolation between ab-initio data and model data;
- universal surrogate functional forms.

[25] B. Hartke, Chem. Phys. Lett. 258 (1996) 144.

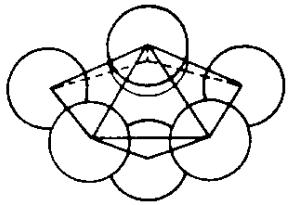
Hard sphere packing

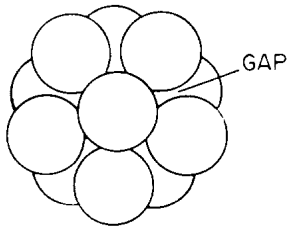
packing of *a few* hard spheres
→ tetrahedron at $n = 4$.





Packing of regular tetrahedrons
almost leads to larger regular forms:





$n = 7$ $n = 13$

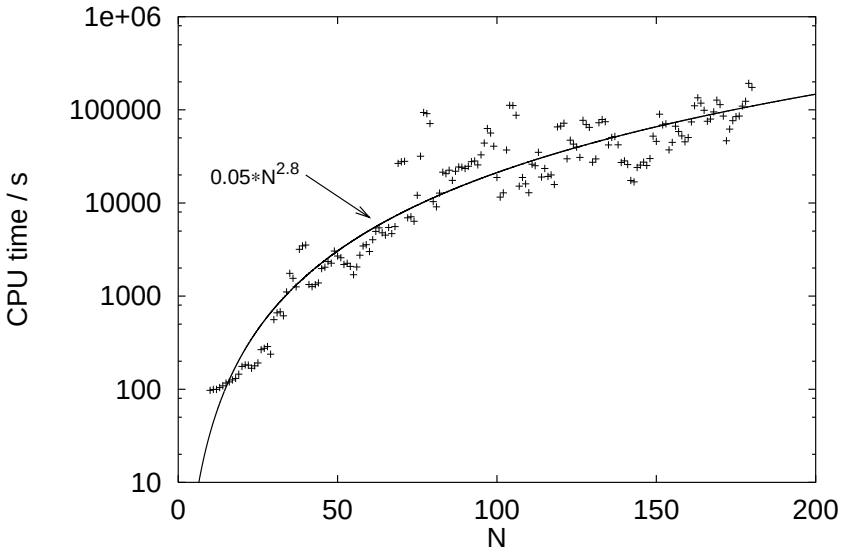
pentagonal bipyramid icosahedron

- "frustration" → polytetrahedral[26] quasi-crystals, glasses, ...
- five-fold symmetry axes; icosahedral growth
- impossible in periodic crystals ⇒ n-dependent structural transitions

[26] F. C. Frank, Proc. Royal Soc. London 215A (1952) 43; H. S. M. Coxeter, Illinois J. Math. 2 (1958) 746; J. D. Bernal, Proc. Royal Soc. London Ser. A280 (1964) 299.

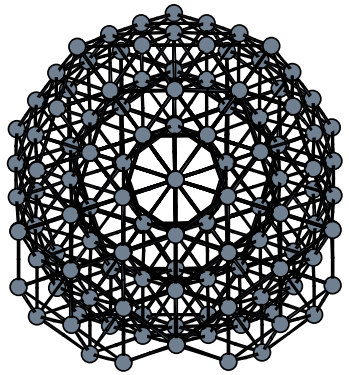
Performance in the LJ cluster benchmark:[27]

- all 4 structural types found without prior information
- *cubic* scaling of CPU time for reaching global minima:

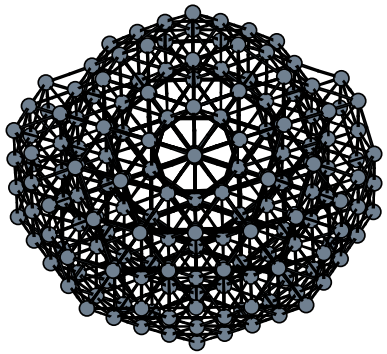


$\Rightarrow$ *large* systems can be treated:

- improved global minima found for $n = 185, 186, 187$;
- published global minima confirmed up to $n = 250$:



$(\mathrm{LJ})_{186}$, E = -1132.669966

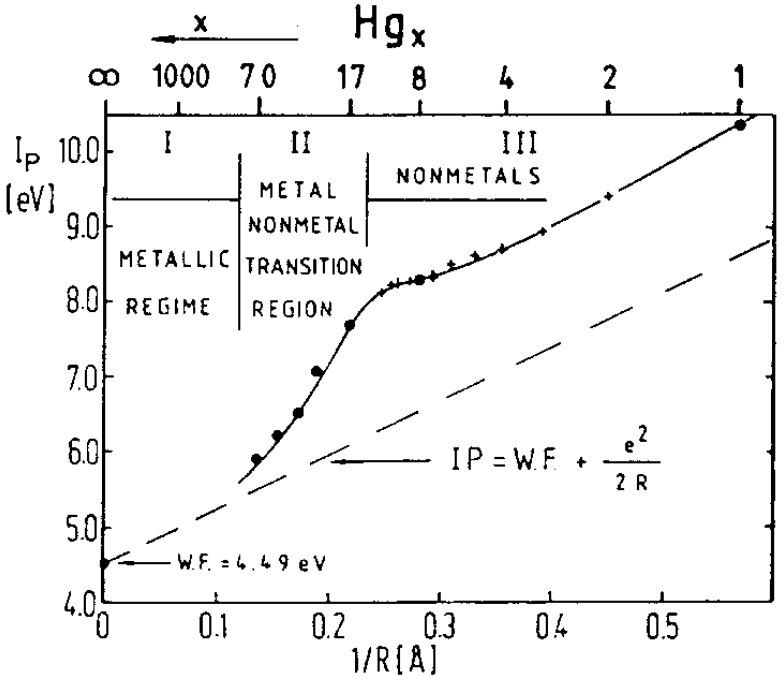


$(\mathrm{LJ})_{250}$, E = -1579.794975

open problem: structural transition from dominance of icosahedral structures to bulk-like fcc; suspected at approx. $n = 750 - 1500$

[27] B. Hartke, J. Comput. Chem. 20 (1999) 1752.

Application example: Mercury clusters



hypothesis:
two size-dependent transitions in bond type

hybrid model (Dolg & Flad, Phys.Rev.B 61 (2000) 2362.):

$$E(\mathrm{Hg}_n) = E_{HF,ECP,CPP}(\mathrm{Hg}_n) - \sum_{i<i}^{n} \frac{C_6 f_6(R_{ij})}{R_{ij}^6}$$

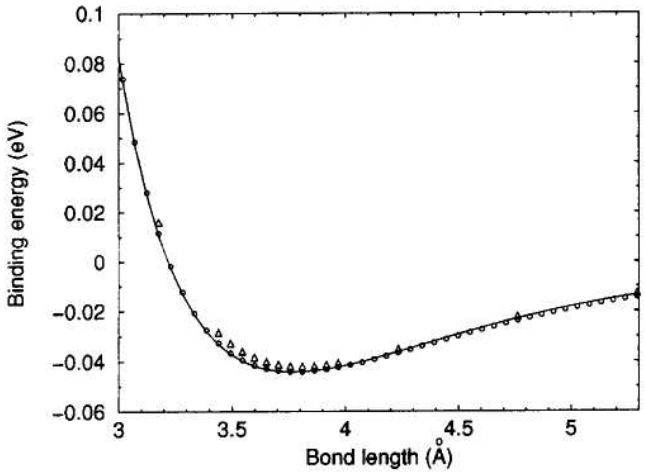
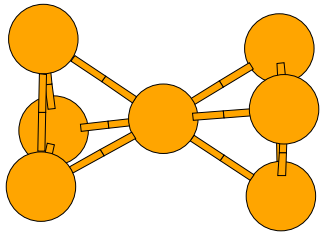


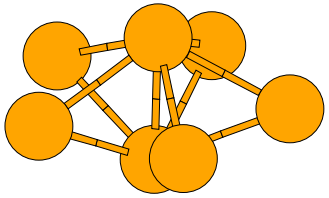
FIG. 1. Potential energy curve of Hg_2 obtained from CCSD(T) calculations using a 2-valence-electron pseudopotential plus core-polarization potential ○ and a 20-valence-electron pseudopotential △. The solid line represents the hybrid model adjusted to the aforementioned type of calculation.

Mercury cluster structures in the hybrid model: [28]

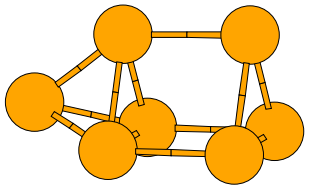
Hg_7



$\frac{E_{coh}}{atom} =$ 12.3 kJ/mol



10.7 kJ/mol

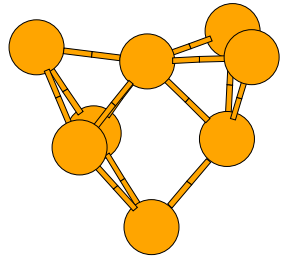


9.0 kJ/mol

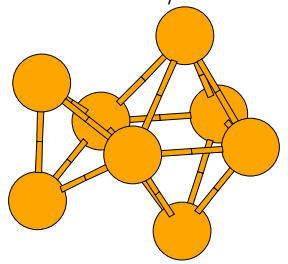
3-body forces compress the pentagonal bipyramid:

	dihedral angle/ degrees
Lennard-Jones	73.1382
Morse	72.8817
this potential	67.1189
pot. w/o 3-body	85.6078

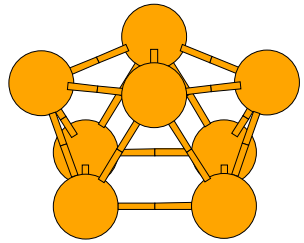
Hg_8



$\frac{E_{coh}}{atom} =$ 12.8 kJ/mol

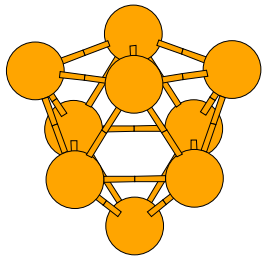
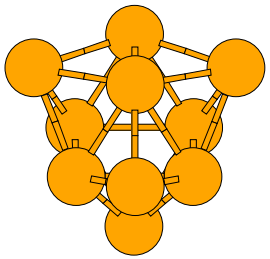
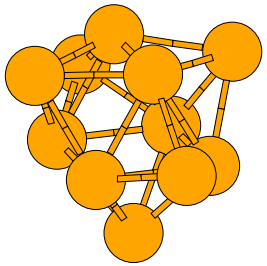
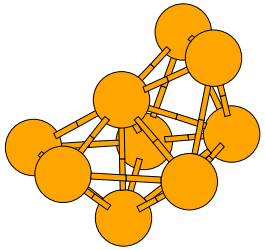
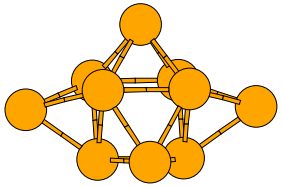
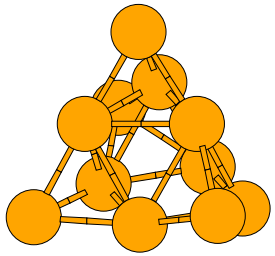
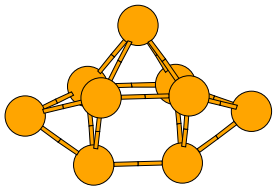
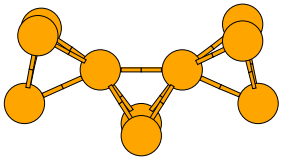
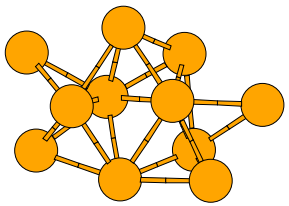


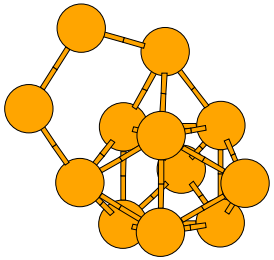
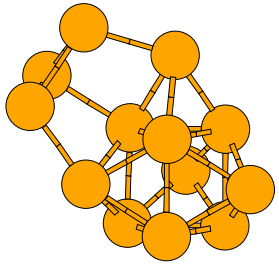
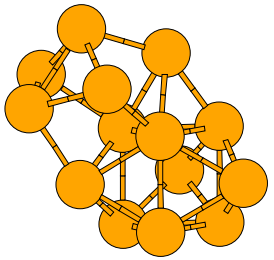
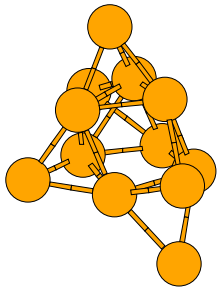
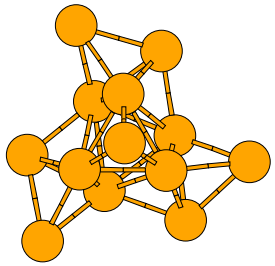
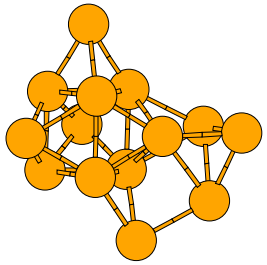
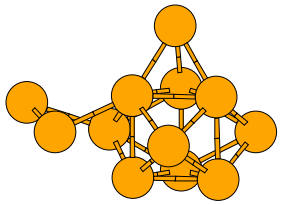
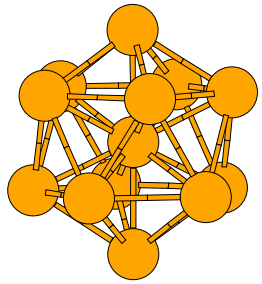
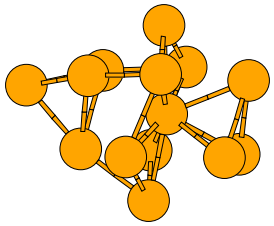
10.6 kJ/mol



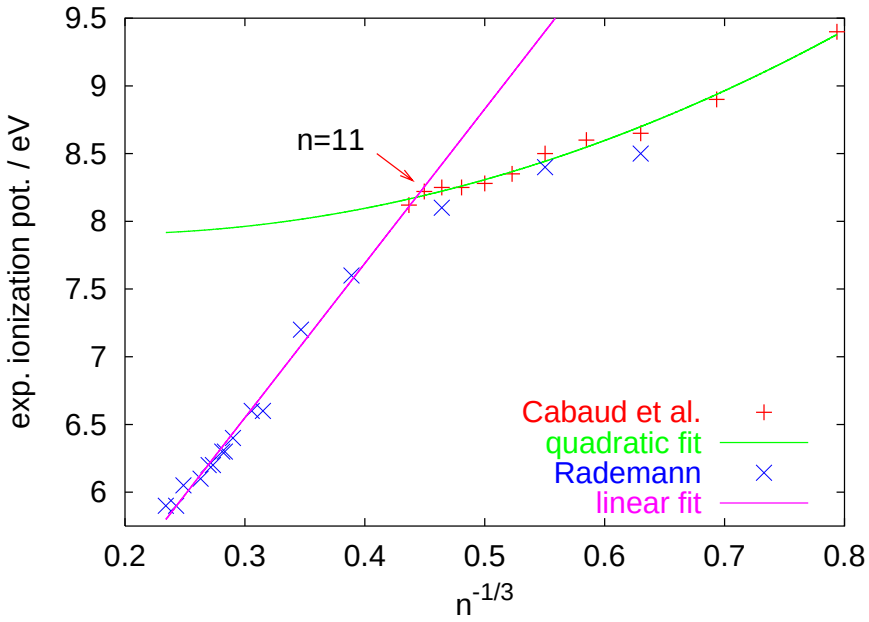
10.5 kJ/mol

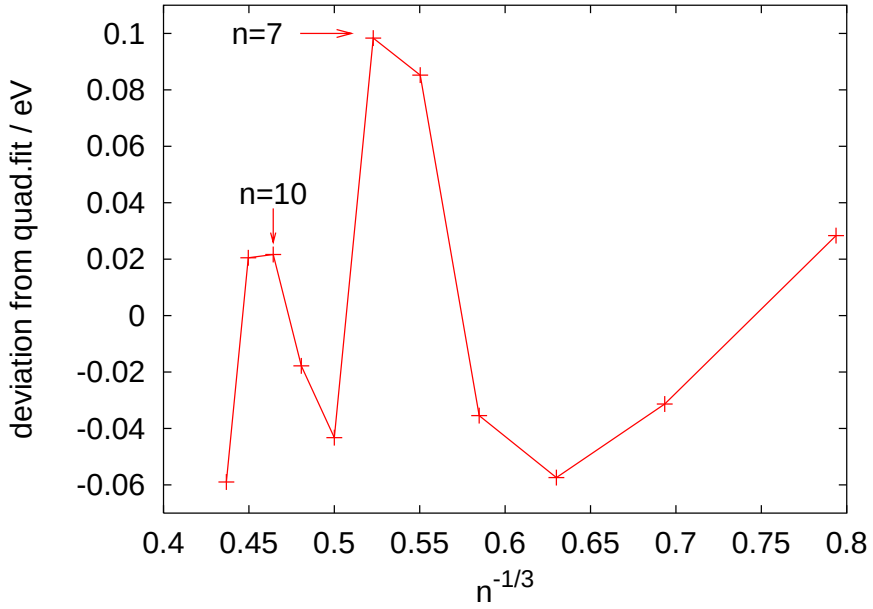
[28] B. Hartke, H.-J. Flad and M. Dolg, Phys. Chem. Chem. Phys. 3 (2001) 5121.

	$\mathbf{Hg}_9$	$\mathbf{Hg}_{10}$	$\mathbf{Hg}_{11}$
$\frac{E_{coh}}{atom} =$			
	16.1 kJ/mol	19.1 kJ/mol	19.7 kJ/mol
			
	13.7 kJ/mol	15.4 kJ/mol	18.9 kJ/mol
			
	12.6 kJ/mol	13.6 kJ/mol	14.5 kJ/mol

	$\mathbf{Hg}_{12}$	$\mathbf{Hg}_{13}$	$\mathbf{Hg}_{14}$
			
$\frac{E_{coh}}{atom} =$	18.9 kJ/mol	20.2 kJ/mol	21.6 kJ/mol
			
	18.5 kJ/mol	18.2 kJ/mol	20.1 kJ/mol
			
	18.3 kJ/mol	15.1 kJ/mol	16.7 kJ/mol

tentative connection to experiment:





B. Cabaud, A. Horeau and P. Melinon, J. Phys. D: Appl. Phys. 13 (1980) 1831.

K. Rademann, Ber. Bunsenges. Phys. Chem. 93 (1989) 653.

Application example: pure neutral water clusters

literature results for small clusters $(H_2O)_n$:

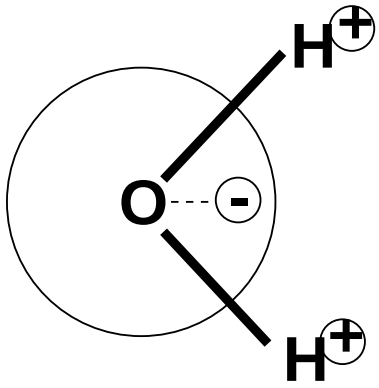
- experimental VRT spectra of size-selected clusters n=3–6 simulated theoretically[a] with excellent agreement;
- IR spectra in the OH-stretch region of size-selected clusters n=7–10 measured and simulated[b], with good agreement.

larger clusters n >10:

- only a few ab-initio/DFT calculations for selected geometries
- global optimization[29]only for TIP4P, up to n=22.

water in the TIP4P model:

- rigid monomers
- point charges
- not polarizable
- no many-body terms



[a] E. M. Maas, R. Bukowski, K. Szalewicz, G. C. Groenenboom, P. E. S. Wormer and A. van der Avoird, J. Chem. Phys. 113 (2000) 6678, 6702.

[b] J. Sadlej, V. Buch, J. K. Kazimirski and U. Buck, J. Phys. Chem. A 103 (1999) 4933.

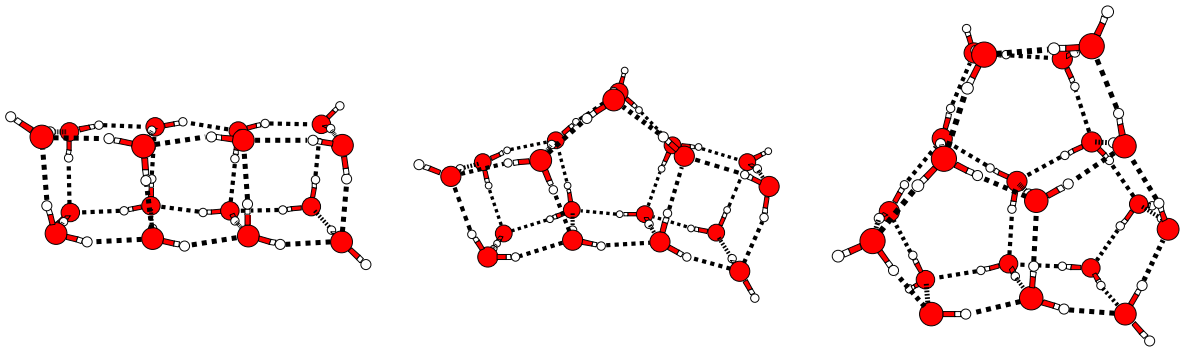
[29] D. J. Wales and M. P. Hodges, Chem. Phys. Lett. 286 (1998) 286; B. Hartke, Z. Phys. Chem. 214 (2000) 1251.

Results of previous studies with TIP4P:

- structures of small clusters $n \leq 10$ qualitatively correct.

But for larger clusters:

- strange tendency to fused cubes and pentagonal prisms
- no systematic structural trends
- all molecules at the cluster surface, up to at least $n = 22$.



n=16 n=18 n=20

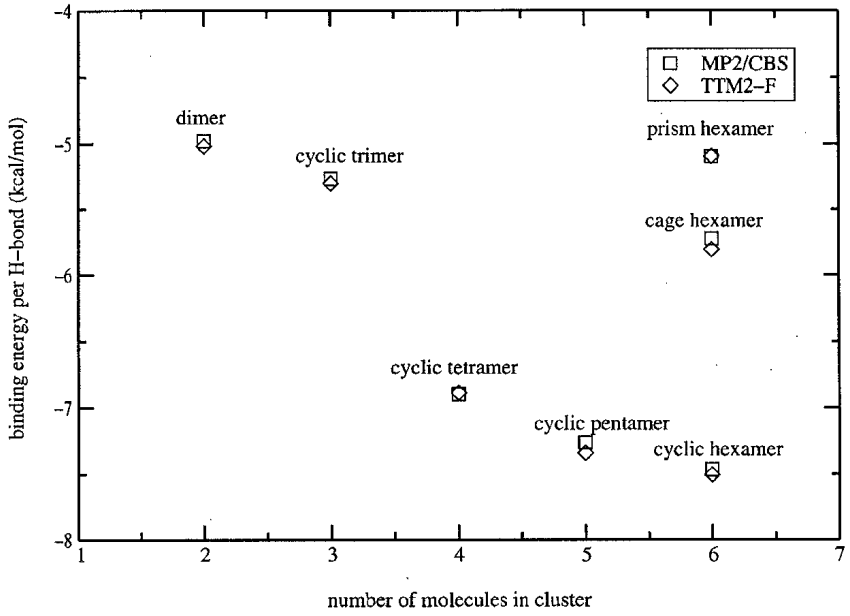
$\Rightarrow$ artifacts of the TIP4P model?

The TTM2-F model: [30]

important properties:

- flexible monomers, with a
- highly exact, internal ab-initio potential;
- smeared partial charges;
- smeared polarizabilities;
- explicit many-body interactions via iterative induction.

quantitative agreement with MP2-CBS for small clusters:



. . . but computationally more expensive than TIP4P by a factor of 20 !

[30] C. J. Burnham and S. S. Xantheas, J. Chem. Phys. 116 (2002) 5115.

Global minimum structures with TTM2-F: [31]

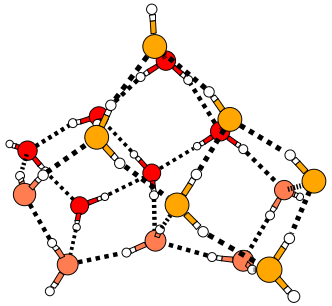
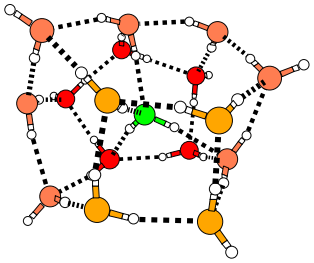
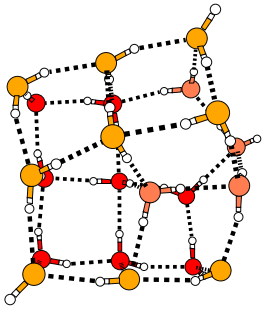
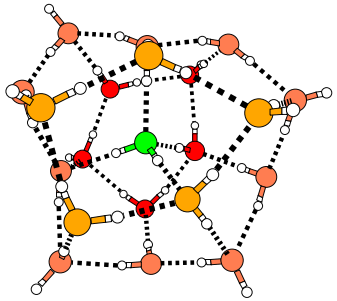
some qualitative agreements between TIP4P and TTM2-F:

- same global and local minima up to n=11;
- some global minima identical up to n=20;
- tendency to cubes and pentagonal prisms.

but also qualitative differences:

- different orientational isomers already at n=12,13;
- big differences at n=17,21,22,...:

Our TTM2-F results were later supported by MP2/aug-cc-pVTZ results [32]

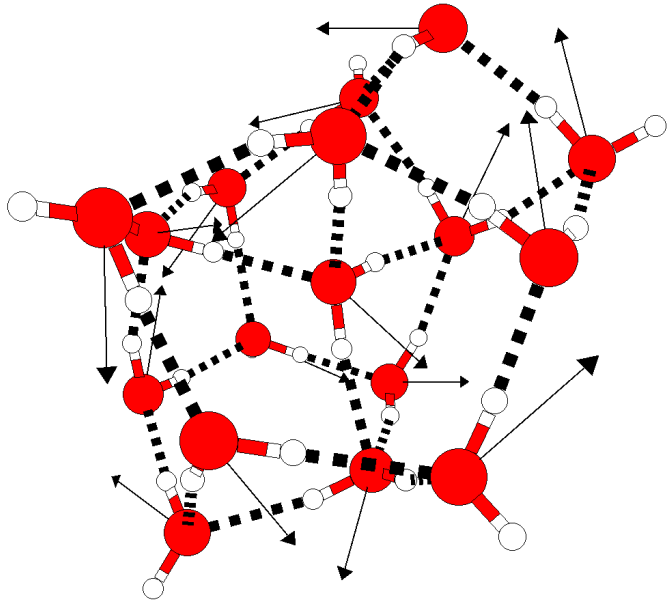
	TIP4P global minimum	TTM2-F global minimum
n=**17**	 all-surface	 centered cage
n=**21**	 all-surface	 centered cage

[31] B. Hartke, Phys. Chem. Chem. Phys. 5 (2003) 275.

[32] A. Lagutschenkov, G. S. Fanourgakis, G. Niedner-Schatteburg, S. S. Xantheas, J. Chem. Phys. 122 (2005) 194310.

Reasons for structural differences

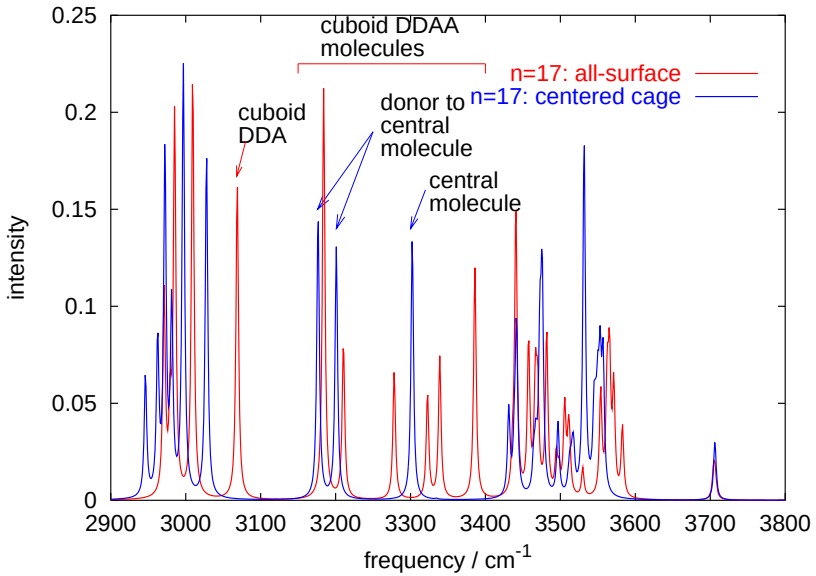
- monomer flexibility is not important (comparison to TTM2-R): at most 16% of the effect, and often even contraproductive.
- adaptive size and orientation of monomer multipoles in TTM2-F leads to
  - antiparallel homodromy in ring stacks is less bad (TIP4P favors parallel homodromy, n=12,15,16)
  - cages less bad, by deviation of dipoles from HOH-plane and from HOH angle bisector $\Rightarrow$ better relative orientations of the dipoles

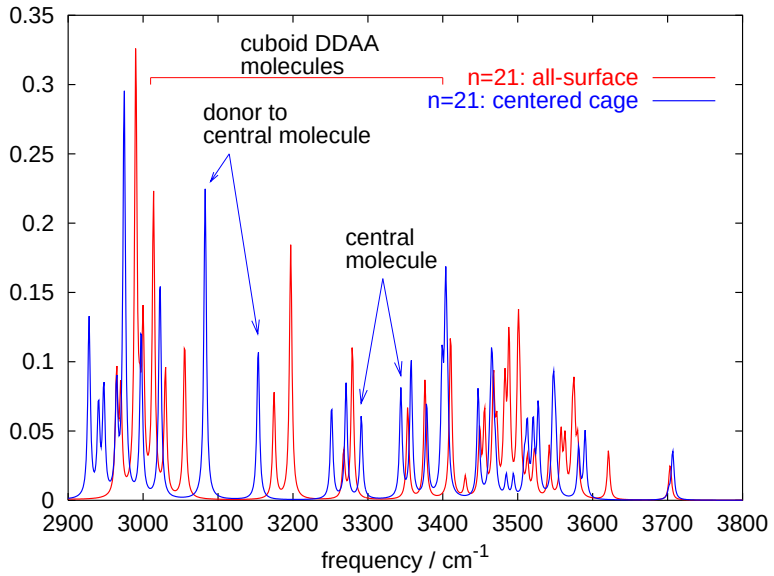


Connection to experiment

Calculation of IR frequencies and intensities of OH-strech vibrations, using the anharmonic empirical model of Buch et al. (good agreement with experiment for small clusters[33])

Characteristic DDAA region allows for differentiation between all-surface structures and centered cage structures:



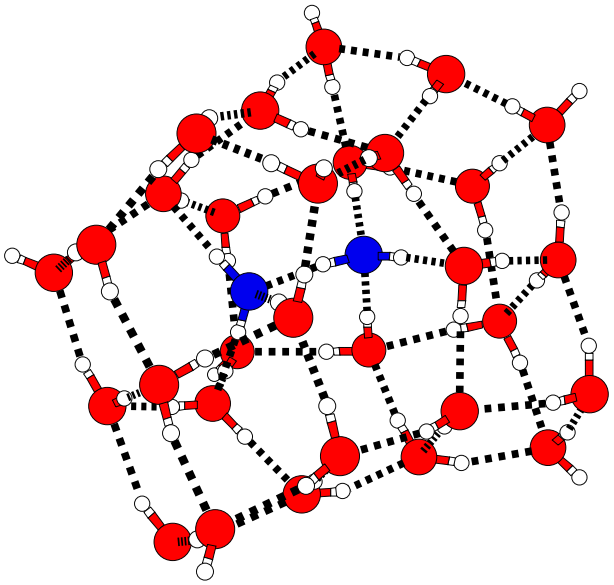


[33] J. Sadlej, V. Buch, J. K. Kazimirski and U. Buck, J. Phys. Chem. A 103 (1999) 4933.

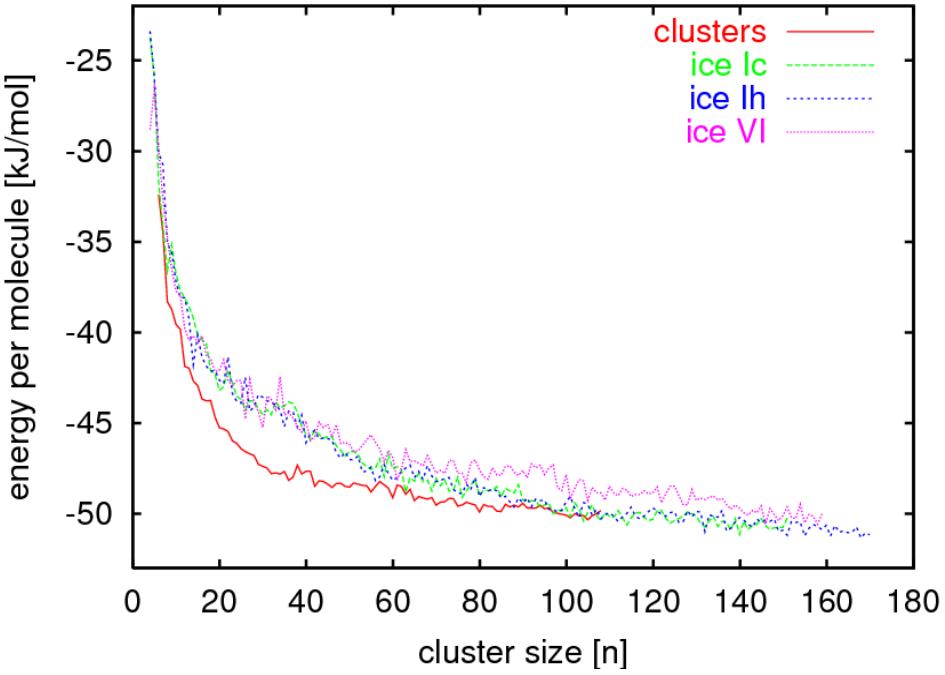
Larger water clusters on their way to ice[34]

- TTM2-F water potential [35] (20 times more expensive than TIP4P); quantitative agreement with MP2/CBS
- parallel pool model on 8–64 processors (local PC clusters, national high-performance computing center HLRN Berlin/Hannover)
- up to 110 molecules; seeding with spherical ice cut-outs

n=32, cage with 2 internal mol.

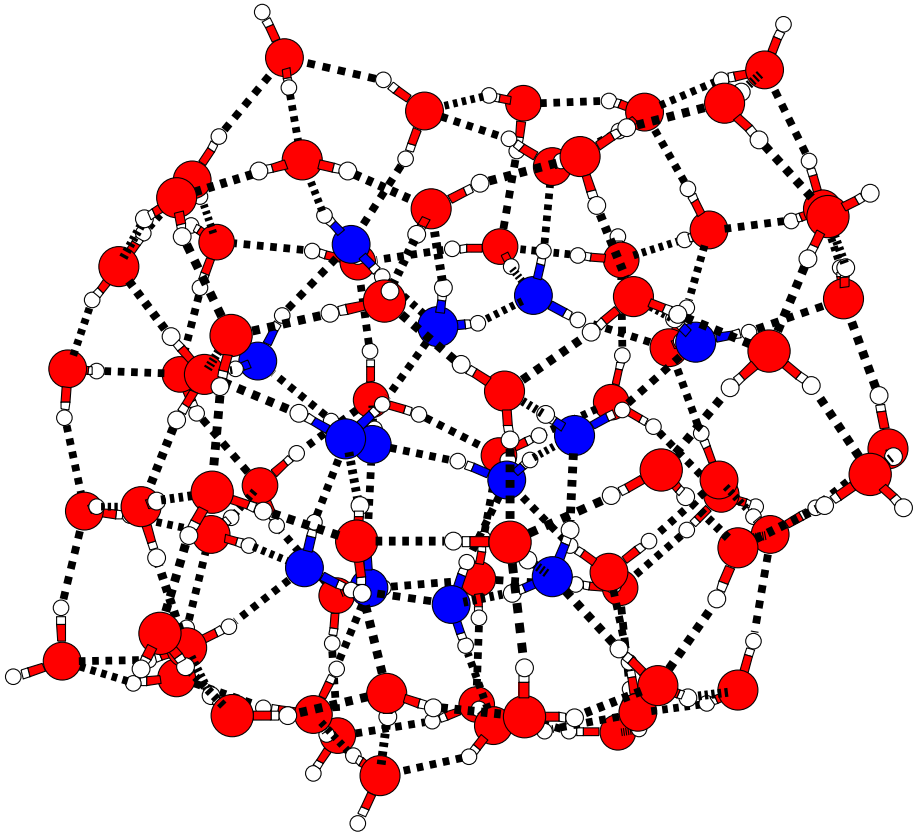


clusters ↔ locally relaxed ice cutouts;
⇒ isoenergetic at n=90 (??)

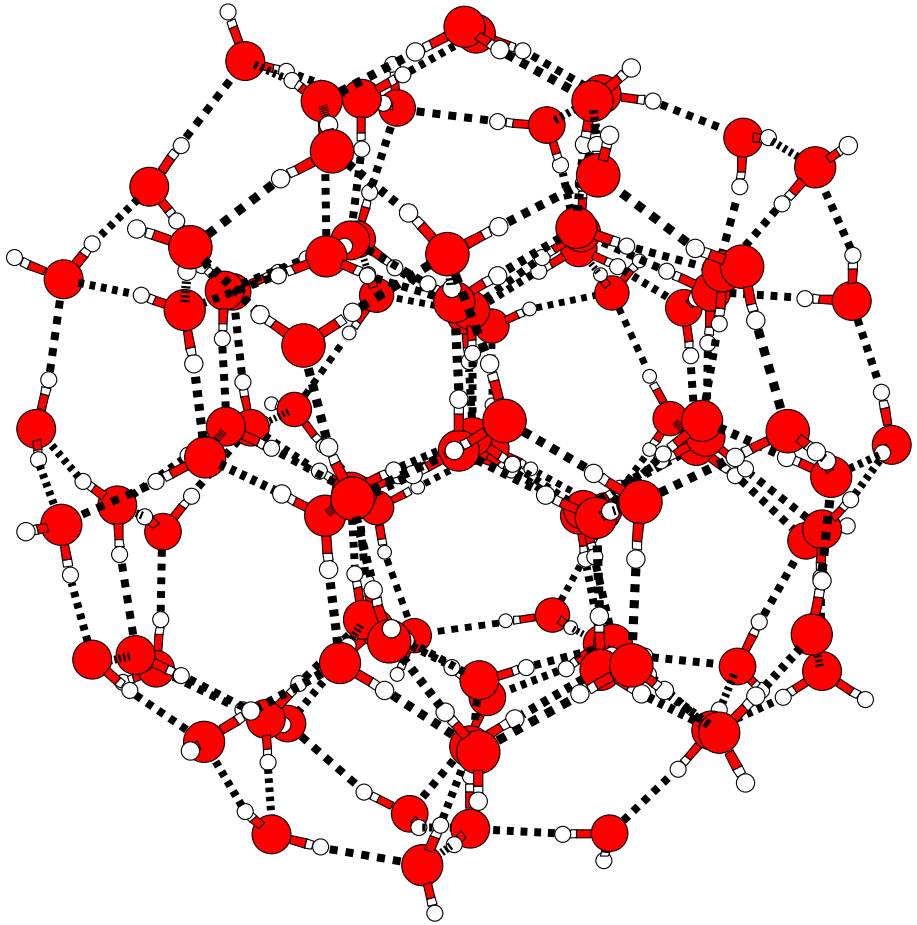


[34] B. Bandow and B. Hartke, J. Phys. Chem. A 110 (2006) 5809.
[35] C. J. Burnham and S. S. Xantheas, J. Chem. Phys. 116 (2002) 5115.

n=80, with 13 internal molecules,
after 20 generations unbiased globopt



n=100, ice Ih seeded; very likely
only a locally relaxed seed

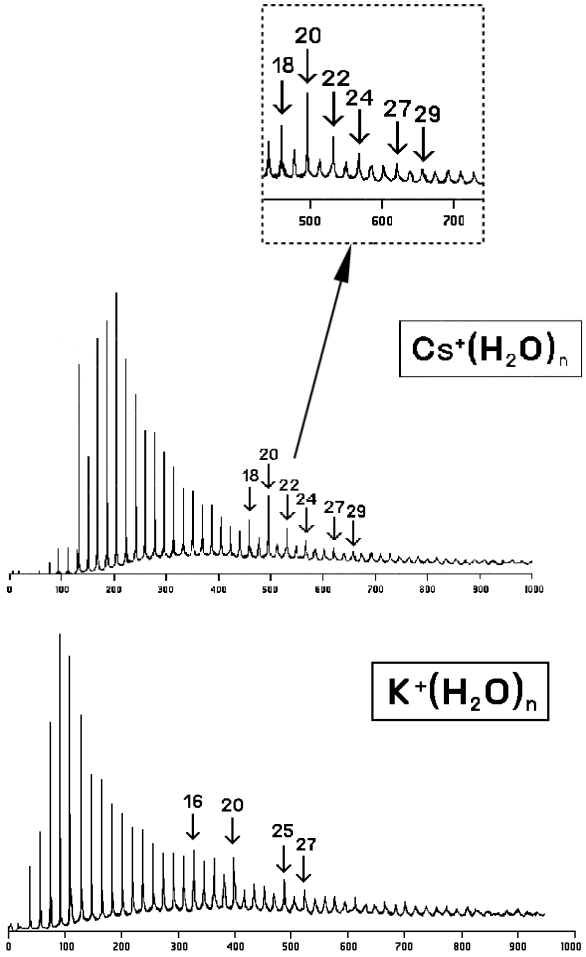


difficulties:

- global optimization at these cluster sizes extremely expensive, even in massively parallel pool mode;
- recognition of a minimal-size ice core non-trivial.

Application example: microhydration clusters of alkali cations

robust experimental result[36]:
(weak) magic numbers



standard explanation[37]:
"clathrate" formation
(n=20: dodecahedron)

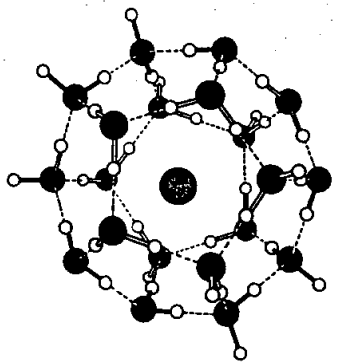
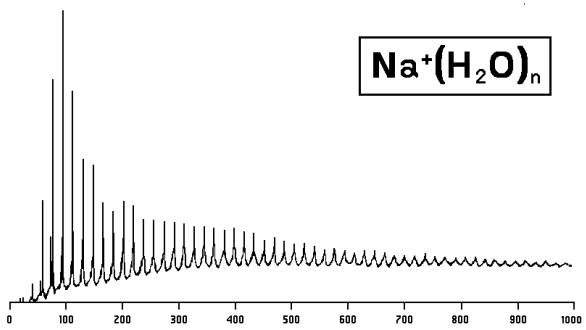


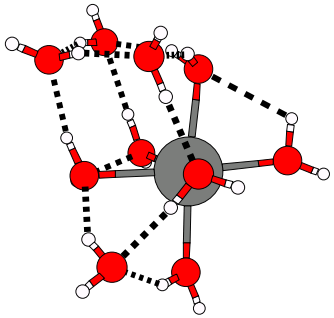
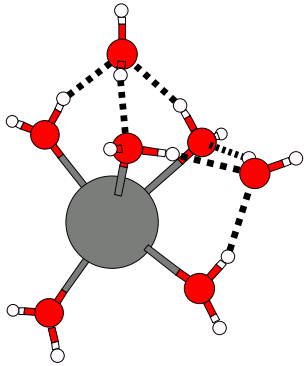
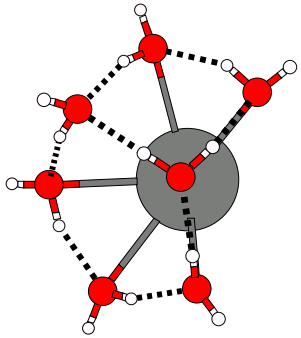
Figure 2. Proposed structure for $Cs^+(H_2O)_{20}$. This structure comprises 12 pentagons. White spheres, hydrogen; black spheres, oxygen; gray central sphere, Cs^+; dashed lines, hydrogen bonds.

...but not for Na^+:



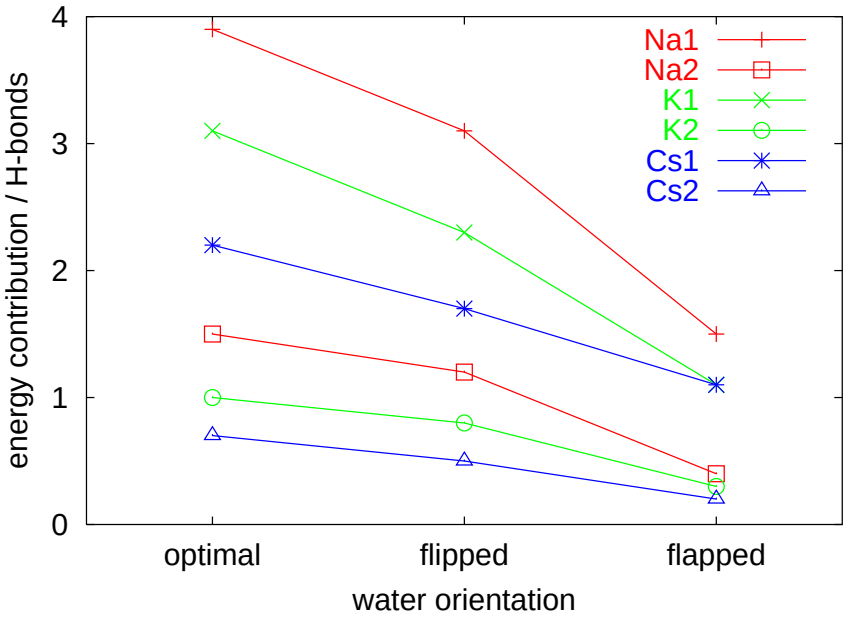
[36] B. Brutschy et al., Int. J. Mass Spectrom. 185/186/187 (1999) 271.
[37] A. Selinger and A. W. Castleman, Jr., J. Phys. Chem. 95 (1991) 8442

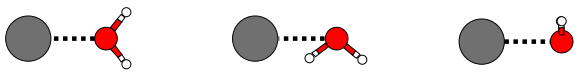
Unbiased global optimization: [38] structural trends:

$Na^+(H_2O)_n$	$K^+(H_2O)_n$	$Cs^+(H_2O)_n$
		
• coordination polyhedra around the ion • competition between – ion centered – ion off-center • no cages	• for $n \leq 9$ as $Na^+(H_2O)_n$ • for $n \geq 10$ as $Cs^+(H_2O)_n$	• no coordination polyhedra • instead: water rings • closing to cages at $n = 18$

[38] F. Schulz and B. Hartke, Chem. Phys. Chem. 3 (2002) 98.

Rationalization of different structural preferences:





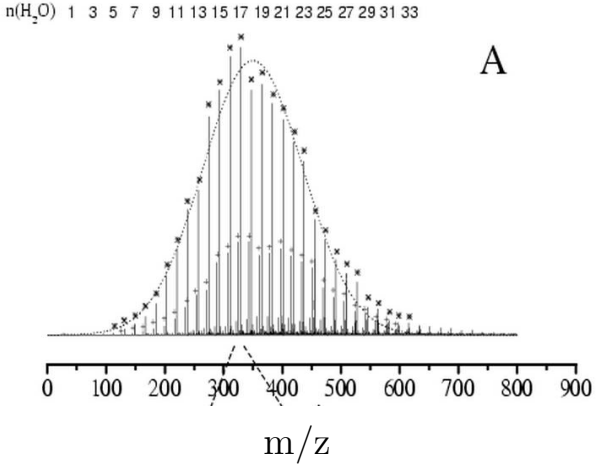
optimal "flip" "flap"

TIP4P/OPLS energy contributions in kJ/mol and in comparison to an optimal water-dimer H-bond (column "H"), for different ion-water distances

ion-water distance [Å]	optimal orientation	H	flipped	H	flapped	H
Na						
2.32	102.64	3.9	80	3.1	40	1.5
4.0	40	1.5	30	1.2	10	0.4
K						
2.68	79.59	3.1	60	2.3	30	1.1
5.0	25	1.0	20	0.8	8	0.3
Cs						
3.20	58.08	2.2	45	1.7	28	1.1
6.0	17	0.7	14	0.5	5	0.2

Na^+ is strongly and far-reachingly orienting, hard to beat by a good water network.
For Cs^+, the opposite applies.

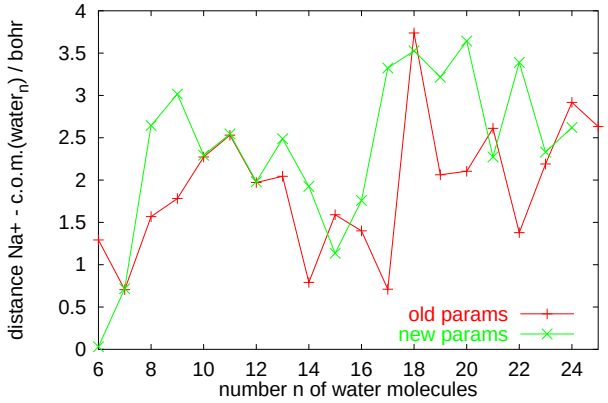
Can this be checked experimentally?



$\Longleftarrow$

"jump" in the mass spectrum[41] of $Na^+(H_2O)_n$ at $n = 17$ agrees with our structural observations.

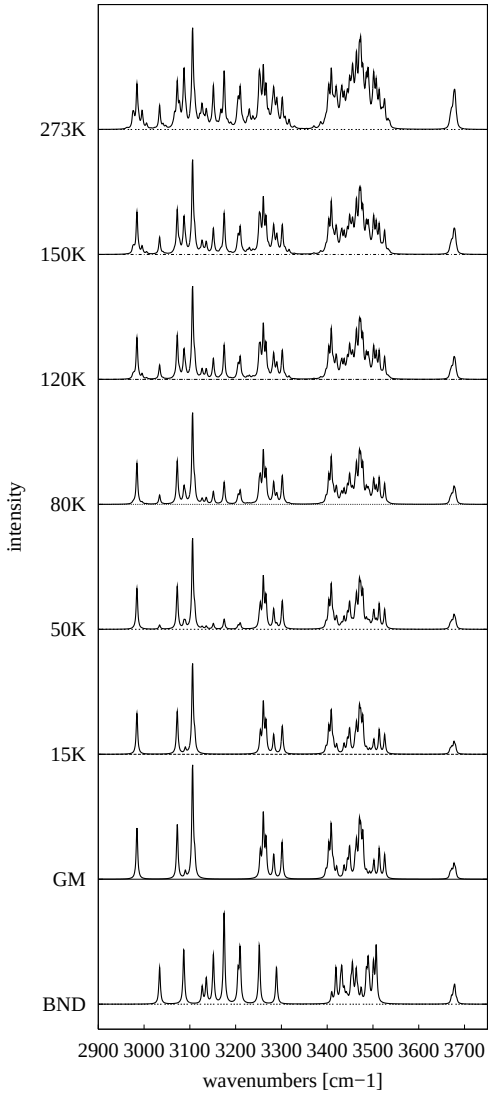
$\Longrightarrow$



T-dependence[42] of IR spectra in OH-stretch region (anharmonic empirical model of Victoria Buch):

$Cs^+(H_2O)_{20}$
GM: global minimum, T=0K
BND: best non-dodecahedron, T=0K

structural differences detectable

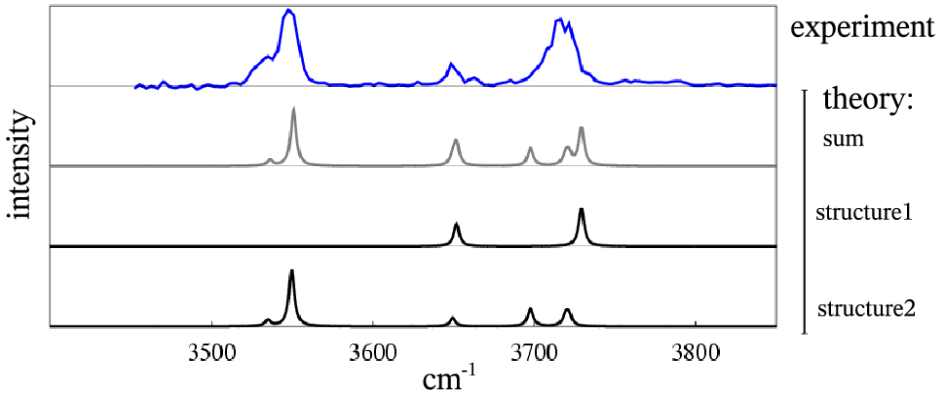


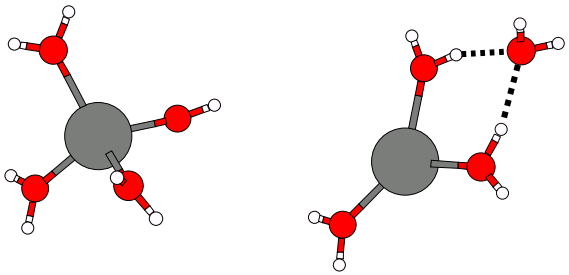
[41] B. Hartke, A. Charvat, M. Reich and B. Abel, J. Chem. Phys. 116 (2002) 3588.
[42] F. Schulz and B. Hartke, Phys. Chem. Chem. Phys. 5 (2003) 5021.

Direct structural comparison with experiment: IR spectra

Experiment: James M. Lisy, University of Illinois at Urbana–Champaign
Theory: anharmonic empirical model by Victoria Buch[43], for our global and best local minima structures

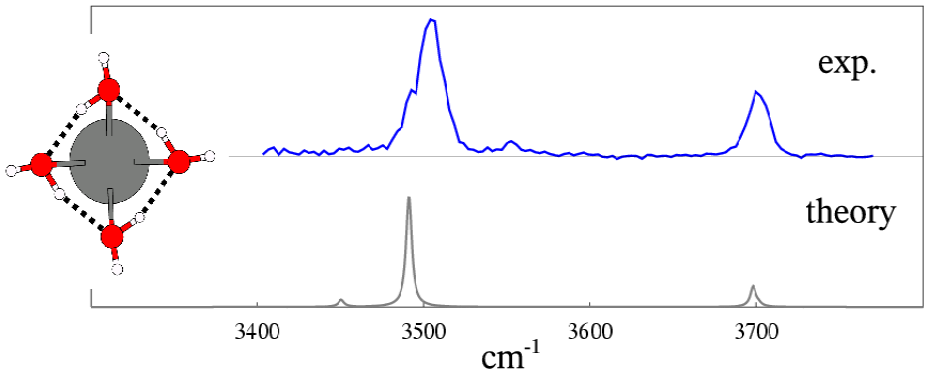
$Na^+(H_2O)_4$:

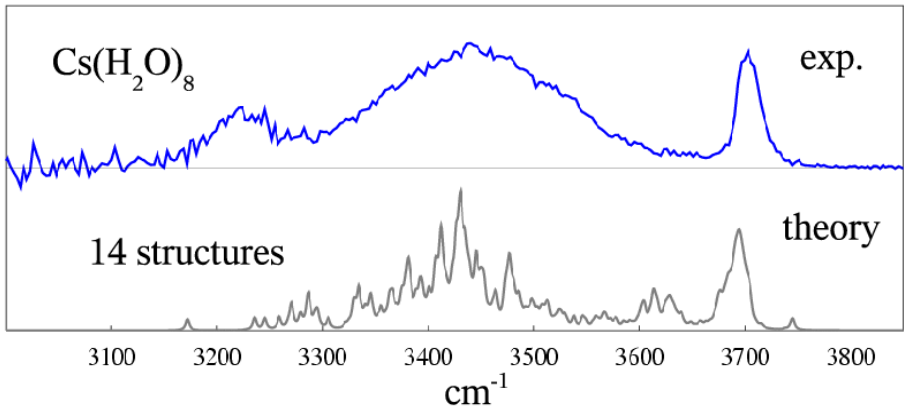




structure 1 structure 2

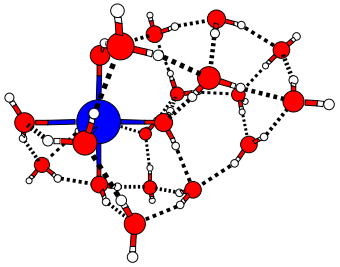
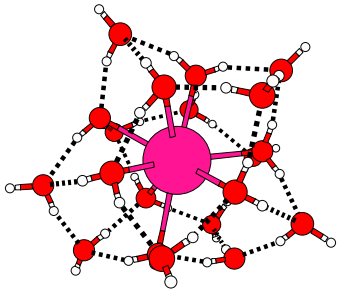
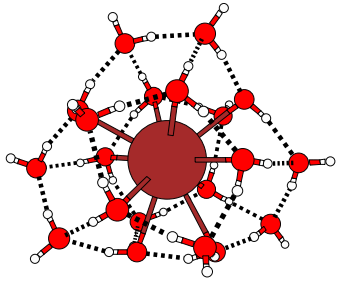
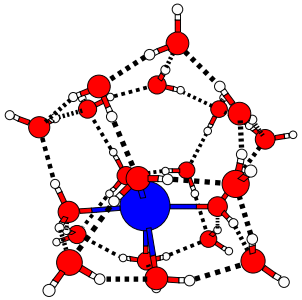
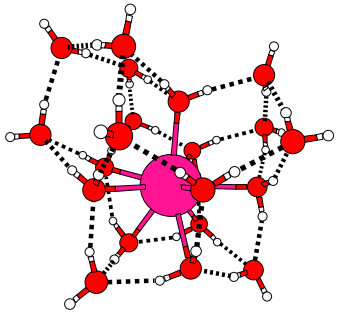
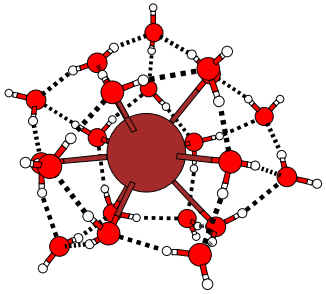
$Cs^+(H_2O)_4$:





[43] J. Sadlej, V. Buch, J. K. Kazimirski and U. Buck, J. Phys. Chem. A, 1999, **103**, 4933.

Are dodecahedra important? [44]

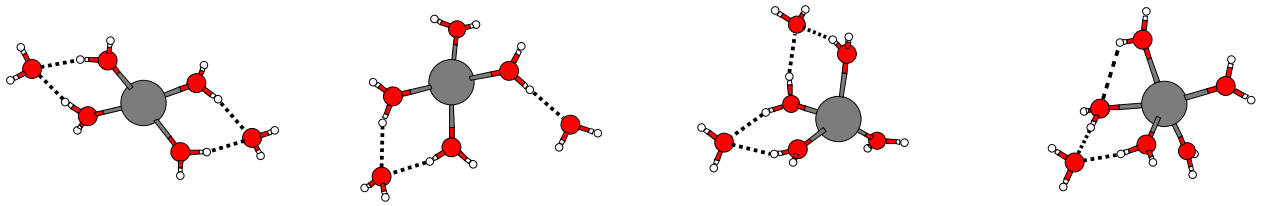
	$Na^+(H_2O)_{20}$	$K^+(H_2O)_{20}$	$Cs^+(H_2O)_{20}$
global minimum			
	irregular, off-center	distorted dodecahedron	distorted dodecahedron
local minimum			
	dodecahedron: +13 kJ/mol	no cage: +3.3 kJ/mol	4-/5-/6-rings: +0.88 kJ/mol

[44] F. Schulz and B. Hartke, Chem. Phys. Chem. 3 (2002) 98.

Stability and dynamics: free dynamcis

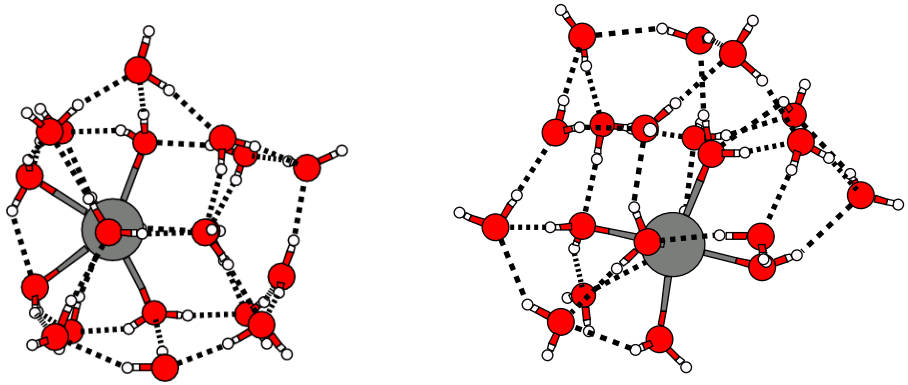
molecular dynamics (canonical ensemble, 40 K $\leq$ T $\leq$ 150 K); input: global & low-energy local minima

- prejudice: very rapid fluctuations, "structure" fluxonial: true only for $Na(H_2O)_n$, $n \leq 10$, T = 120–150K: visits various coordination polyhedra:



globmin, t=0 | t = 0.3 ps | t = 0.4 ps | t = 5.6 ps

- for larger cases of all ions, *no qualitative structural changes* even at 150 K
- big exception: $Na(H_2O)_{20}$ at T $\geq$ 120 K:

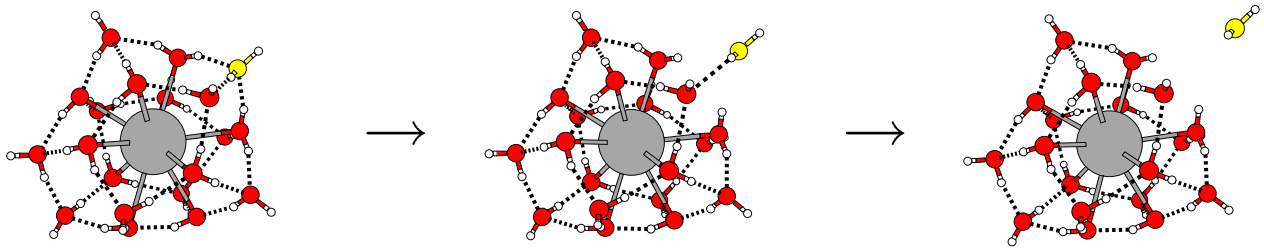


t = 0
dodecahedral local min.

t = 3.3 ps
close to non-dod. global min.

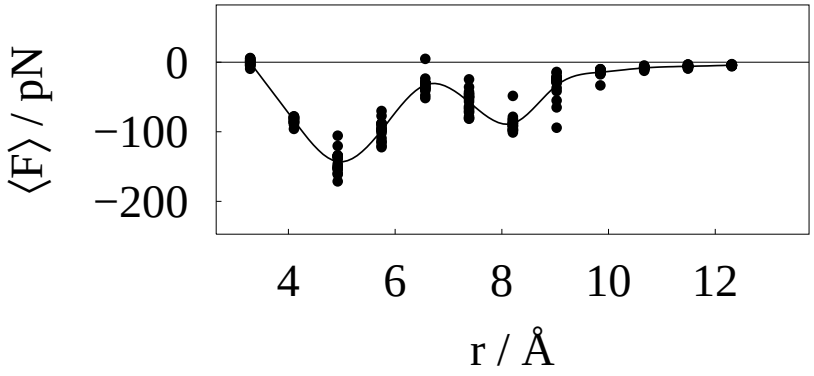
Stability and dynamics: constrained dynamics

- dissociation of single water molecule enforced by fixing its distance to the central ion, for increasing values:

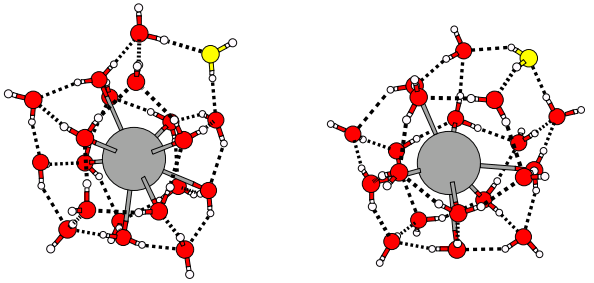


- calculation of the average force $\langle F \rangle$ at each distance
- integration of $\langle F \rangle \rightarrow$ free energy ΔG of dissociation

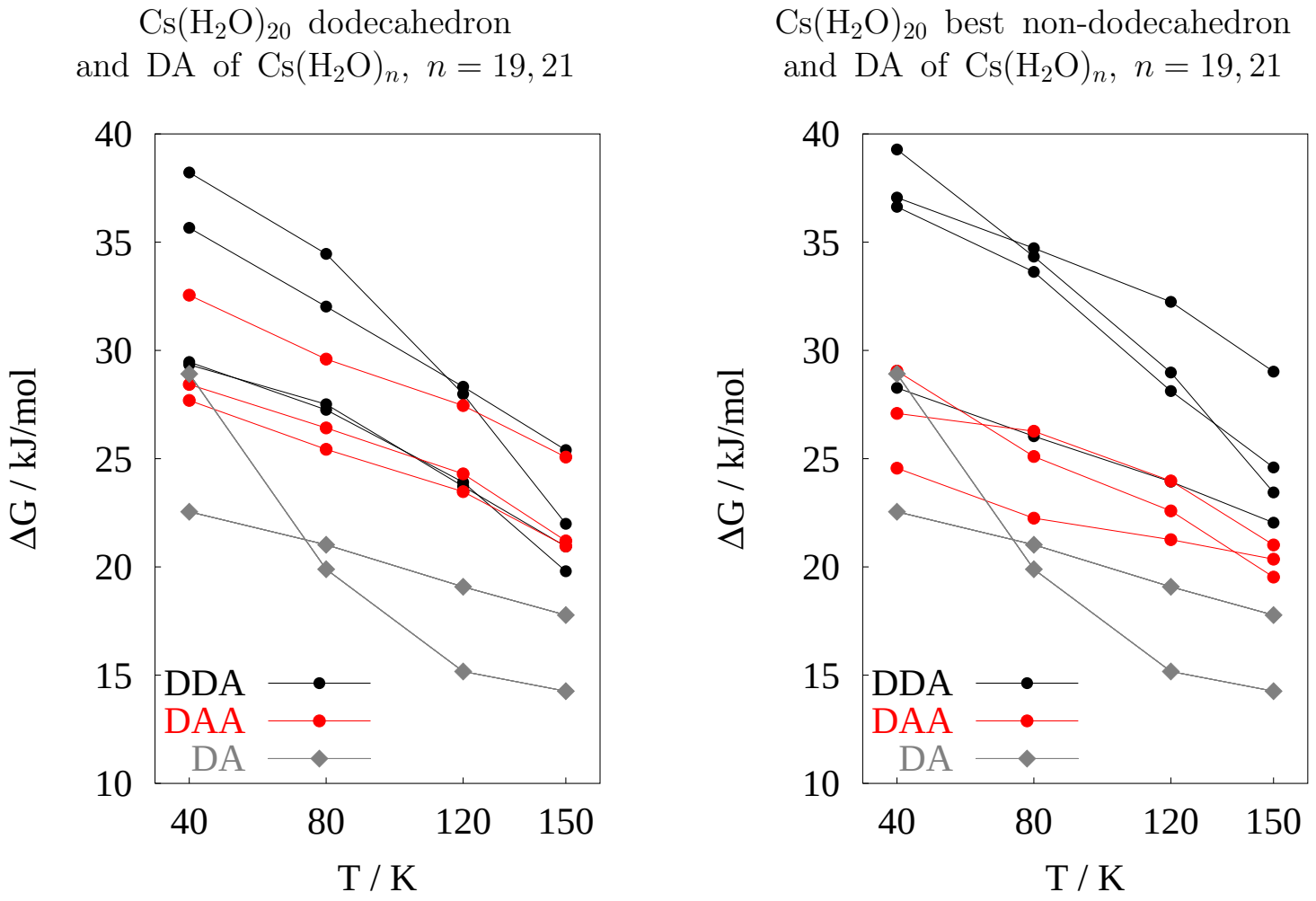
Example: $Cs(H_2O)_{20}$, T = 120 K
ΔG = 28.4 kJ/mol



scatter at each distance due to different reaction pathways:



representative free energy values, as a function of temperature, H-bonding environment, and cluster structure:



- no significant differences between dodecahedron and non-dodecahedron
- resistance to dissociation: DA $<$ DAA $\leq$ DDA

⇒ new hypothesis:[45] magic-number status depends on presence/absence of DA molecules

(note: this is a combination of structural and dynamical propensities)

	single	A	D	DA	AA	DD	DDA	DAA	DDAA
$Na^+(H_2O)_4$	4	–	–	–	–	–	–	–	–
$Na^+(H_2O)_5$	5	–	–	–	–	–	–	–	–
$Na^+(H_2O)_6$	–	–	4	–	2	–	–	–	–
$Na^+(H_2O)_7$	1	–	4	–	1	–	–	–	–
$Na^+(H_2O)_8$	1	–	3	1	–	–	–	3	–
$Na^+(H_2O)_9$	1	–	2	1	–	–	1	4	–
$Na^+(H_2O)_{10}$	–	–	2	2	–	–	2	4	–
$Na^+(H_2O)_{11}$	1	–	2	1	–	1	1	5	–
$Na^+(H_2O)_{12}$	–	–	2	1	–	1	2	6	–
$Na^+(H_2O)_{13}$	–	–	–	1	–	2	3	7	–
$Na^+(H_2O)_{14}$	–	–	–	–	–	–	7	7	–
$Na^+(H_2O)_{15}$	–	–	–	1	–	2	4	8	–
$Na^+(H_2O)_{16}$	–	–	–	1	–	1	6	8	–
$Na^+(H_2O)_{17}$	–	–	–	1	–	2	5	9	–
$Na^+(H_2O)_{18}$	–	–	–	1	–	2	4	8	3
$Na^+(H_2O)_{19}$	–	–	–	2	–	2	4	8	3
$Na^+(H_2O)_{20}$	–	–	–	1	–	2	5	9	3
$Na^+(H_2O)_{21}$	–	–	2	–	–	1	4	9	4

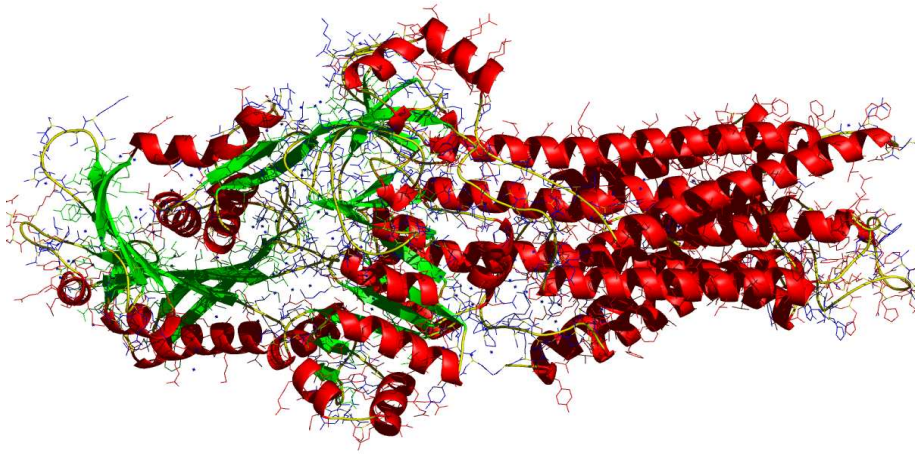
	single	A	D	DA	AA	DD	DDA	DAA	DDAA
$Cs^+(H_2O)_4$	–	–	–	4	–	–	–	–	–
$Cs^+(H_2O)_5$	–	1	1	3	–	–	–	–	–
$Cs^+(H_2O)_6$	2	–	–	4	–	–	–	–	–
$Cs^+(H_2O)_7$	–	–	–	5	–	–	1	1	–
$Cs^+(H_2O)_8$	–	–	1	5	–	–	1	1	–
$Cs^+(H_2O)_9$	–	–	–	5	–	–	2	2	–
$Cs^+(H_2O)_{10}$	–	–	–	6	–	–	2	2	–
$Cs^+(H_2O)_{11}$	–	–	–	5	–	–	3	3	–
$Cs^+(H_2O)_{12}$	–	–	–	4	–	–	4	4	–
$Cs^+(H_2O)_{13}$	–	–	–	5	–	–	4	4	–
$Cs^+(H_2O)_{14}$	–	–	–	4	–	–	5	5	–
$Cs^+(H_2O)_{15}$	–	–	–	3	–	–	6	6	–
$Cs^+(H_2O)_{16}$	–	–	–	2	–	–	7	7	–
$Cs^+(H_2O)_{17}$	–	–	–	3	–	–	7	7	–
$Cs^+(H_2O)_{18}$	–	–	–	–	–	–	9	9	–
$Cs^+(H_2O)_{19}$	–	–	–	1	–	–	9	9	–
$Cs^+(H_2O)_{20}$	–	–	–	–	–	–	10	10	–
$Cs^+(H_2O)_{21}$	–	–	–	1	–	–	10	10	–
$Cs^+(H_2O)_{22}$	–	–	–	–	–	–	11	11	–
$Cs^+(H_2O)_{23}$	–	–	–	1	–	–	11	11	–
$Cs^+(H_2O)_{24}$	–	–	–	–	–	–	11	11	2

[45] F. Schulz and B. Hartke, Theor. Chem. Acc. 114 (2005) 357.

Protein folding

The next grand challenge after the human genome project. NP-hard. Currently favored strategies:

- **databank-based:** homology modeling, threading, ...
  - *advantage:* sometimes very successful
  - *disadvantages:*
    - premise "similar primary sequences fold similarly" may be wrong;
    - we know 3D structures for some but not all primary sequences.
- **ab-initio:** (without using databank information)
  - *advantage:* should work for unknown sequences
  - *disadvantages:*
    - so far limited to short oligopeptides;
    - performance hard to judge independent of force field;
    - accuracy much worse than for good cases of threading.

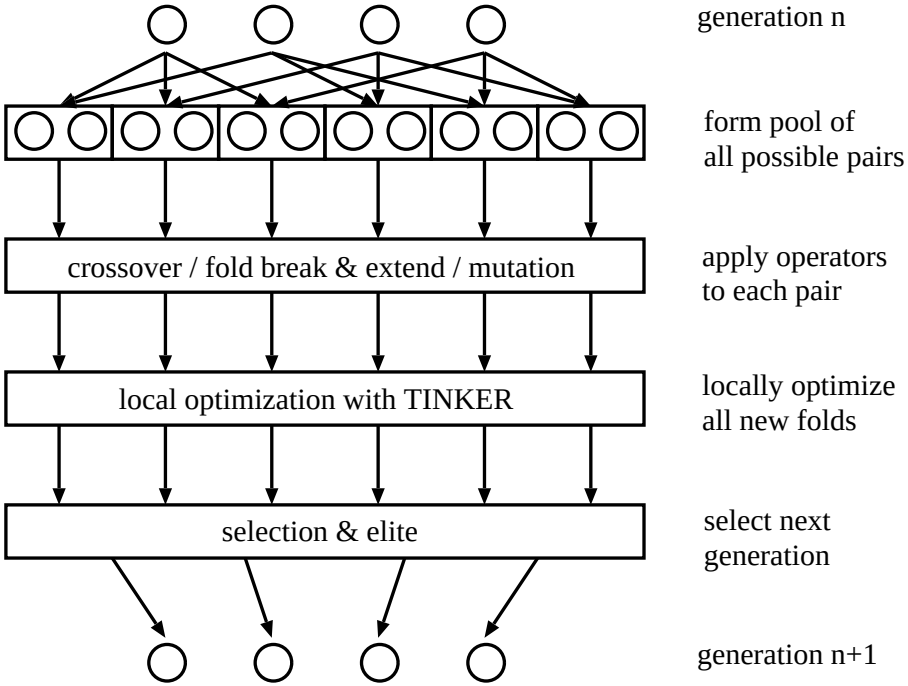


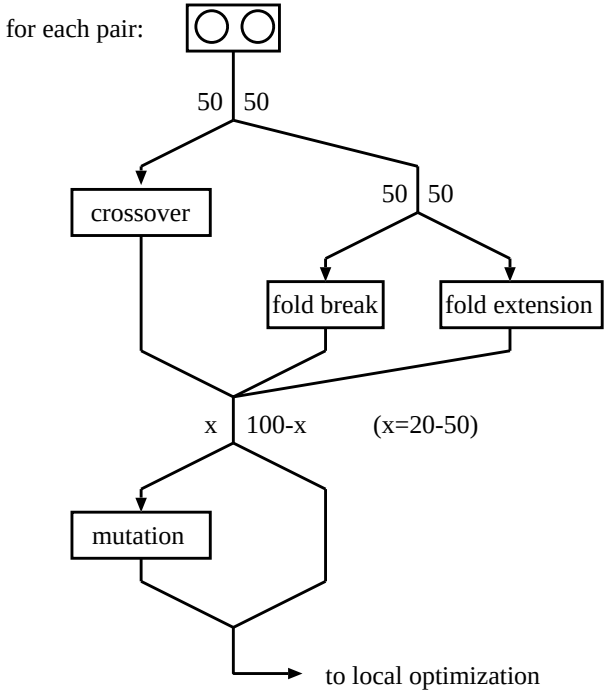
New algorithm PROFET[46], based on our cluster structure evolutionary algorithm:

- It does not operate on backbone dihedral angles exclusively
- instead, secondary structure elements are defined as correlated dihedral angle sets (ϕ, ψ, ω)
- operations of the algorithm introduce, extend, split, and destroy these secondary structure elements
- in all operations, clashes are avoided by construction.

[45] F. Koskowski and B. Hartke, J. Comput. Chem. 26 (2005) 1169.

PROtein Folding with Evolutionary Techniques: PROFET[47]



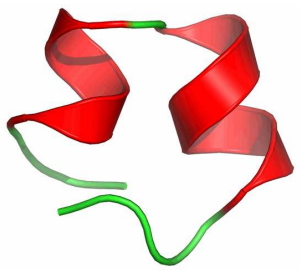


- v1.0: good secondary structure elements (but not their location) as a-priori input information
  v2.0: during the folding, algorithm learns which dihedral angle pairs are good
- side chains explicitly present only during local optimization
- clash avoidance: (1) finite thickness of backbone in global moves,
  (2) no dihedral angle value is enforced, but only optimally approached
- default folding moves not random, but towards maximum compactness

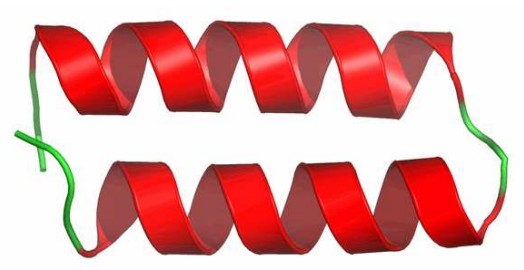
[47] F. Koskowski and B. Hartke, J. Comput. Chem. 26 (2005) 1169.

Test application: Polyalanines

- OPLSAA and AMBER94/96 force fields, as implemented in TINKER
- in vacuo, and with implicit solvent model (distance dependent dielectric)
- ala_n, $n = 20 - 60$



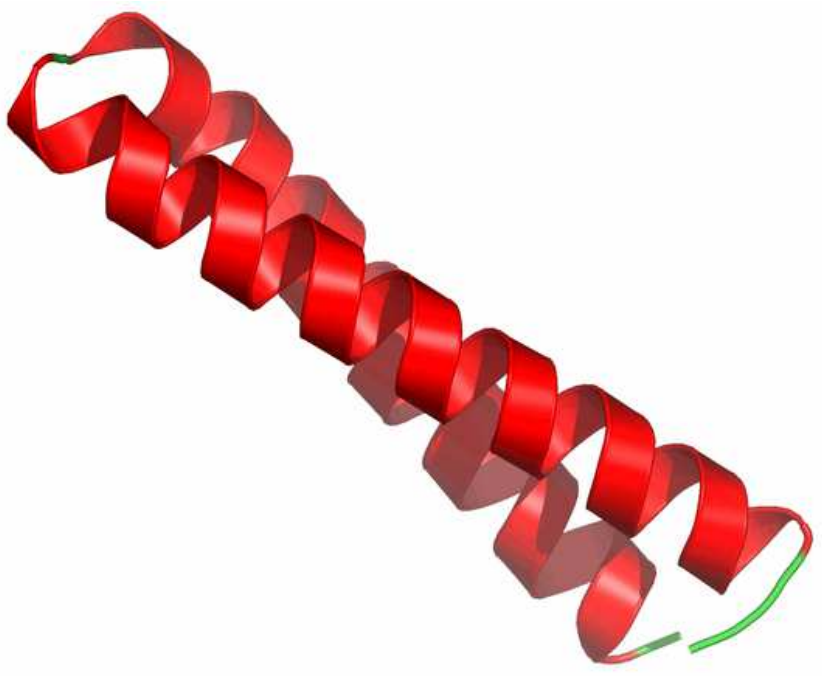
ala_{20}



ala_{40}

Result:
overwhelming preference for bundles of 2 alpha-helices.

This agrees with theoretical expectations and with experimental observations (mobility experiments[a])

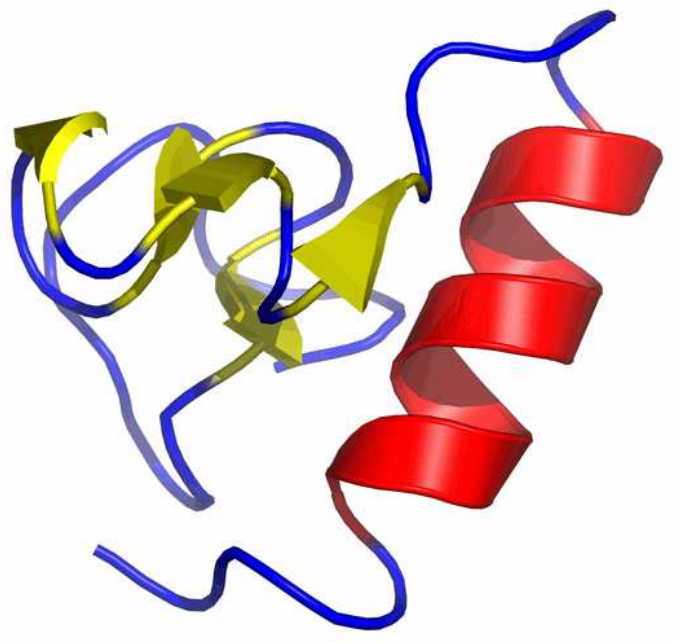


ala_{60}

[a] R. R. Hudgins and M. F. J. Jarrold, J. Am. Chem. Soc. 121 (1999) 3494; A. E. Counterman and D. E. Clemmer, J. Phys. Chem. B 107 (2003) 2111.

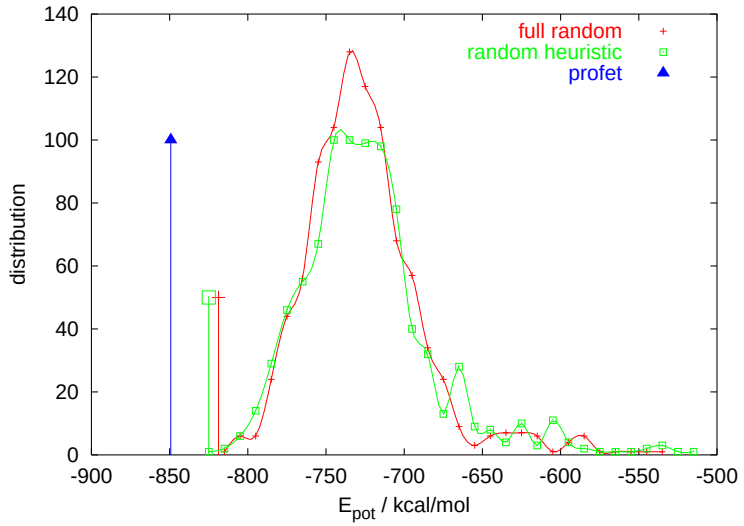
Test application: random sequences

totally different results: globular random coils, sometimes with short secondary structure elements:



⇒ PROFET does *not* enforce the given secondary structure elements!

PROFET clearly beats multiple local optimizations, both without and with random secondary structure elements initialization:



⇒ not the given secondary structure elements are important, but the global optimization algorithm.

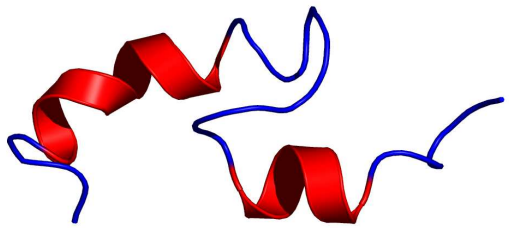
Real life tests: PDB sequences

Connection of PROFET with the SMMP suite[48] as force-field engine for ECEPP and FLEX.

benchmark case 1:

biologically active PTH fragment (residues 1–34) of the human parathyroid hormone (84 residues)
X-ray structure (PDB code: 1ET1):

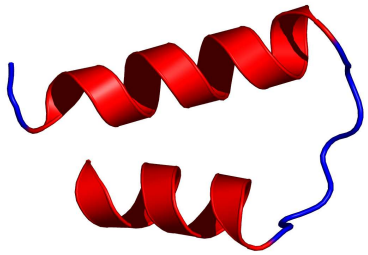
NMR solution structure (PDB code: 1ZWA):



large-scale simulated-annealing/multicanonical sampling study by Hansmann[49] results in RMS=0.8Å from the crystal structure:

We find RMS=1.2Å with FLEX and $\epsilon(r)$:

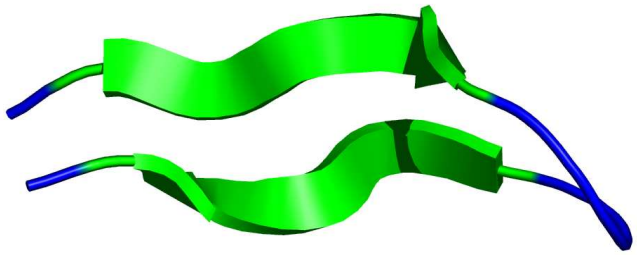
and differing structures with ECEPP:



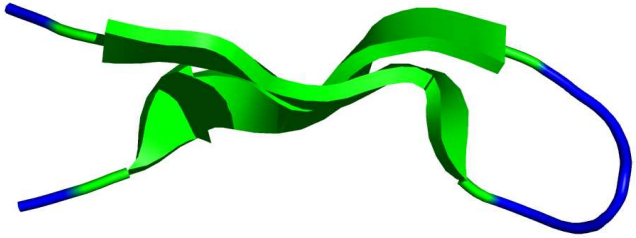
[48] F. Eisenmenger, U. H. E. Hansmann, Sh. Hayryan and C.-K. Hu, Comput. Phys. Commun. 138 (2001) 192.
[49] U. H. E. Hansmann, J. Chem. Phys. 120 (2004) 417.

benchmark case 2: cis proline turn linking two beta-hairpin strands in HIV-1IIIB V3, 18 residues

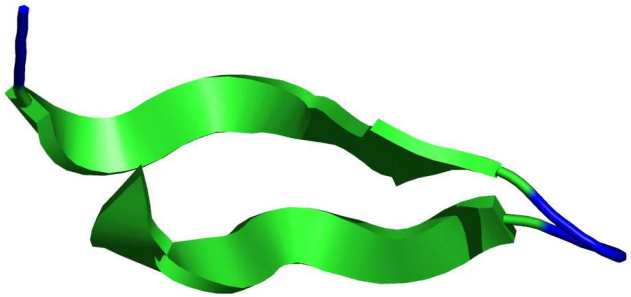
NMR solution structure (PDB entry 1B03)
top view:



side view:



PROFET result with FLEX/$\epsilon(r)$: RMS=4.4Å
top view:



side view: