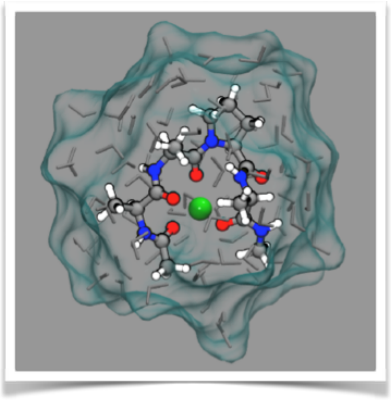
Molecular Dynamics of Peptides in Isolation and Computation of Physical Observables

Carsten Baldauf
"BioGroup", Abteilung Theorie
Fritz-Haber-Institut der MPG, Berlin



Simplified workflow to compute observables

- Rigid "simple" systems
- Structure databases e.g. PDB
- Structure search

***Generate** or find a **3D structure** or an ensemble*

- How much of the "physics" does one need to cover?

Compute Property

Molecule → *Molecular model(s)* → *Number(s)*

***How to rank** multiple models?*

- Energy-based ranking depends on the quality of the energy function.

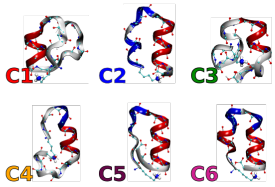
All steps might eventually involve MD simulations

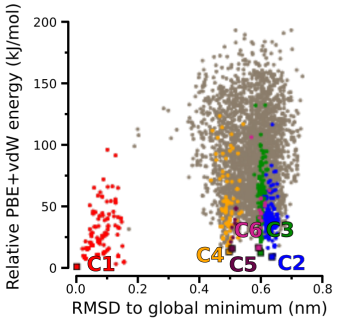
Target is the peptide Ac-LysH$^+$-Ala$_{19}$

Exploration of a large 'search space' to find minima with **force field** and ***ab initio*** replica-exchange MD

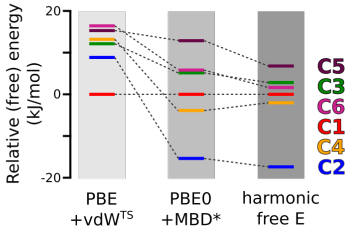
Clustering of conformers:
- Energy ordering
- RMSD clustering
- Machine learning of descriptors

Overview of the PES at a reasonable level of theory



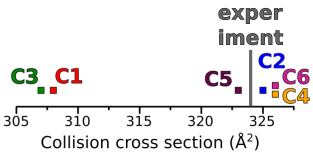


Refinement by applying *higher levels of theory* and *free energy corrections*

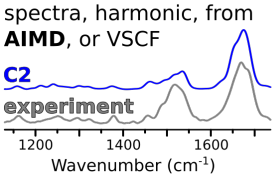


Comparison to experiment by computing physical observables

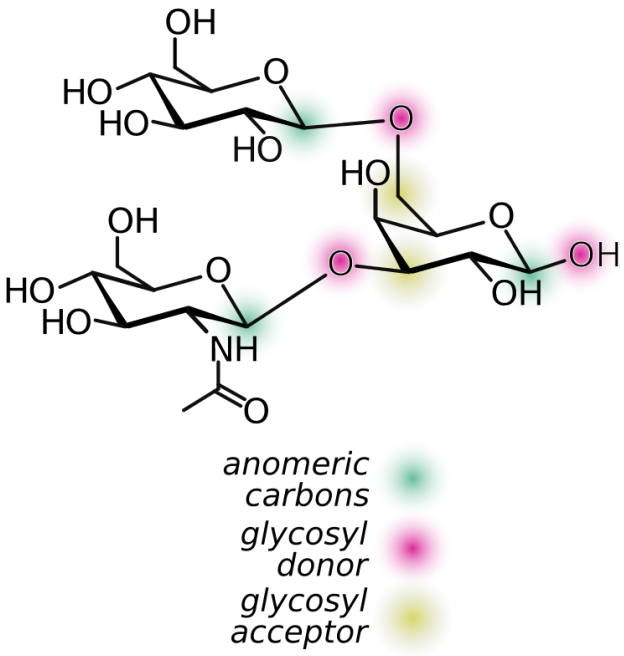
Computing CCS by means of PA, EHSS, or **TM**.

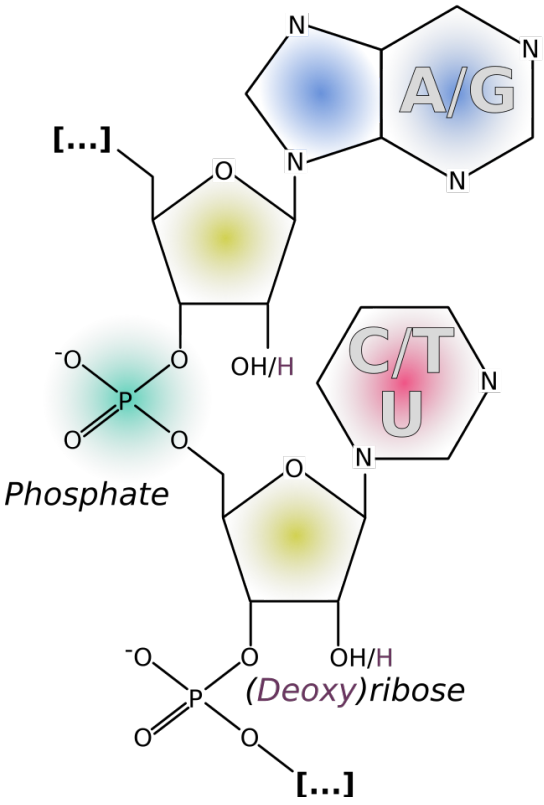


Simulating vibrational spectra, harmonic, from **AIMD**, or VSCF

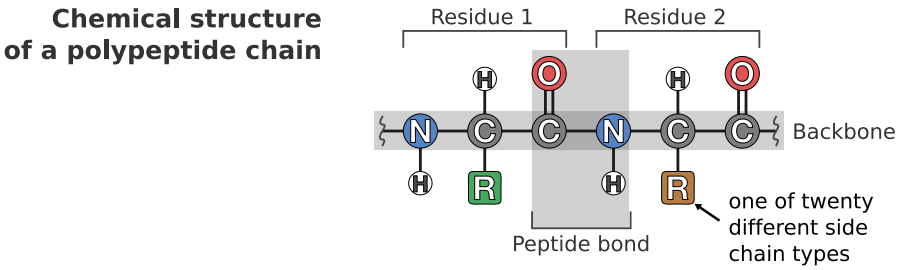


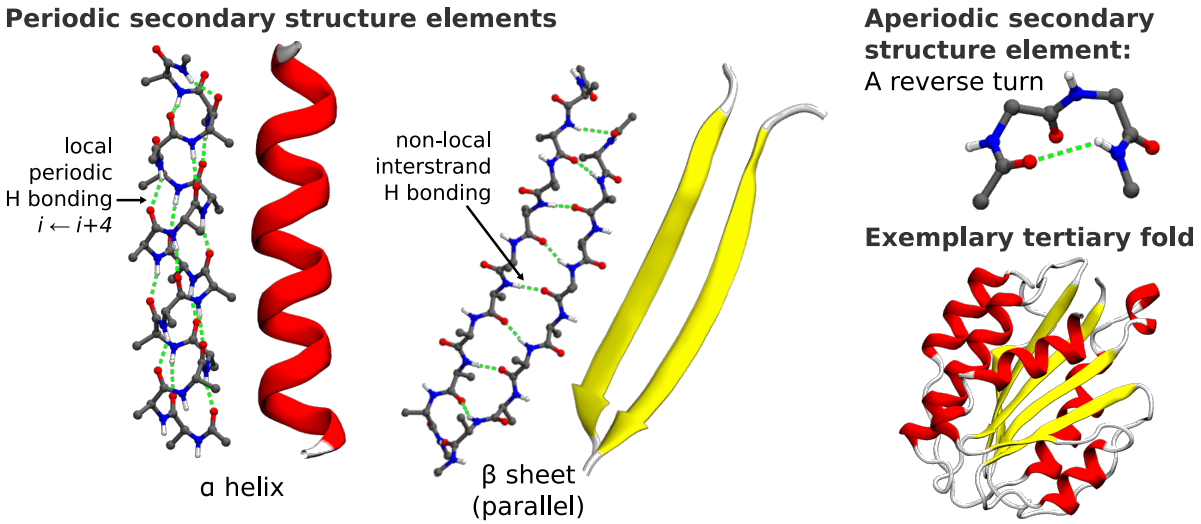
Carbohydrates and Nucleic Acids





Biomolecules: Peptides/Proteins





Why Molecular Dynamics?

- Time-averaged properties of a system, e.g. distribution and population of different states

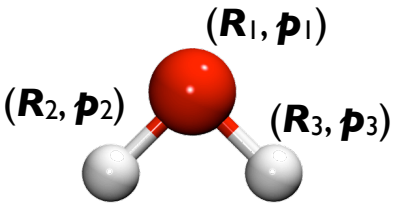
  —> Statistical mechanics, free energies

  —> Structure prediction, sampling

- Time dependence and correlation of properties

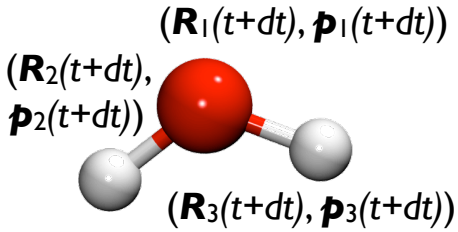
  —> e.g. vibrational spectra

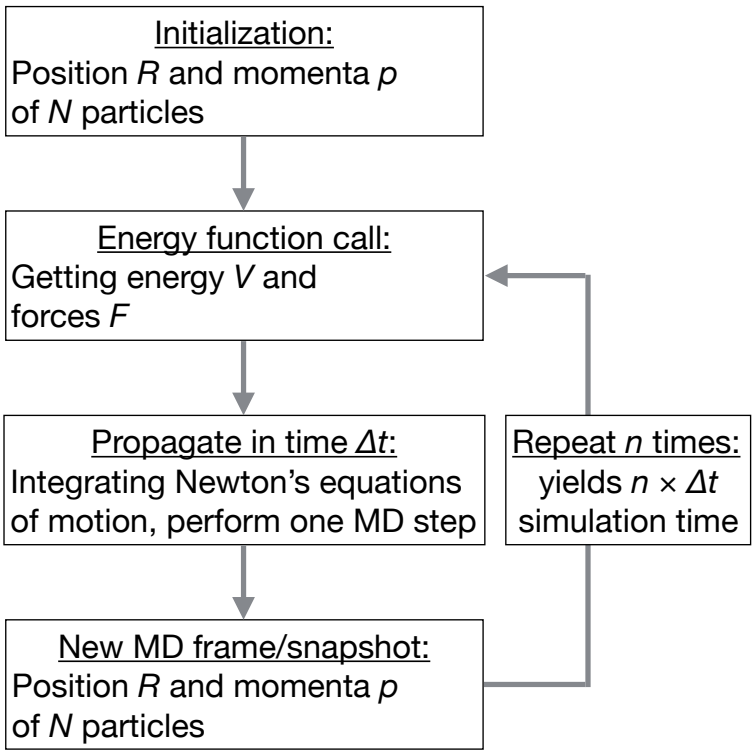
How do we perform MD?



Like an experiment:
1. Sample preparation
2. Equilibration
3. Measurement

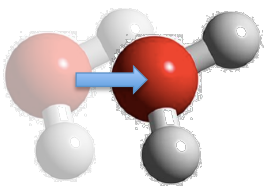
[Smit and Frenkel]





The time step

Translation



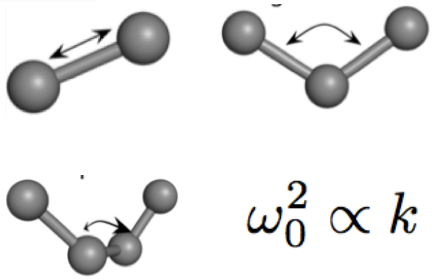
$$\frac{1}{2}mv^2 = \frac{3}{2}k_\mathrm{B}T$$

$$v \approx 600\mathrm{m/s}$$

time to travel its own size:

$$\tau \approx \frac{0.2\,\mathrm{nm}}{v} \approx 0.3\,\mathrm{ps}$$

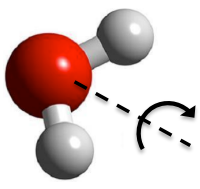
Vibrations



$$\omega_0^2 \propto k$$

high frequency for stiff potentials

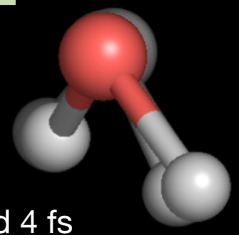
Rotation



$$\frac{1}{2}I\omega^2 = \frac{1}{2}k_\mathrm{B}T$$

$$\tau_\mathrm{rot} \approx 0.4\,\mathrm{ps}$$

Choice of time step



between 0.5 and 4 fs

Integrating the Equations of Motion

- Taylor expansion, 2 points in time

$$R(t+\Delta t) = R(t) + \frac{p(t)}{m}\Delta t + \frac{\dot{p}(t)}{2m}\Delta t^2 + \dddot{R}(t)\frac{\Delta t^3}{3!} + O(\Delta t^4)$$

Error

- Verlet algorithm, 3 points in time

Force

$$R(t+\Delta t) = R(t) + \frac{p(t)}{m}\Delta t + \frac{\dot{p}(t)}{2m}\Delta t^2 + \dddot{R}(t)\frac{\Delta t^3}{3!} + O(\Delta t^4)$$

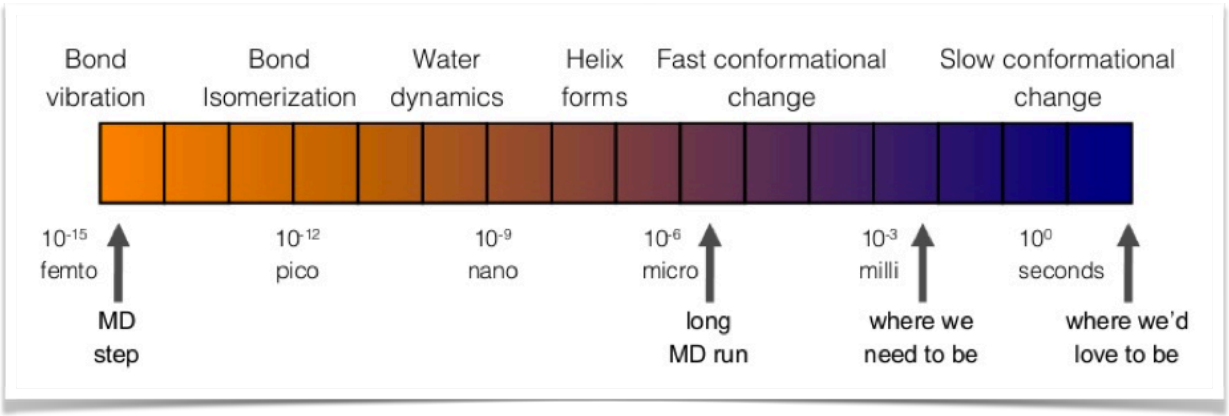
$$R(t-\Delta t) = R(t) - \frac{p(t)}{m}\Delta t + \frac{\dot{p}(t)}{2m}\Delta t^2 - \dddot{R}(t)\frac{\Delta t^3}{3!} + O(\Delta t^4) \quad +$$

$$R(t+\Delta t) + R(t-\Delta t) = 2R(t) + \frac{\dot{p}(t)}{m}\Delta t^2 + O(\Delta t^4)$$

Error

$$R(t+\Delta t) \approx 2R(t) - R(t-\Delta t) + \frac{\dot{p}(t)}{m}\Delta t^2$$

Simulated time is a multiple of time steps Δt



Ensembles, how realistic?

- Micro-canonical (*NVE*) ensemble

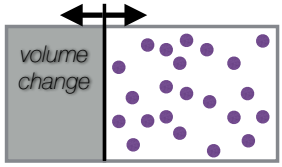
  Number of particle *N*, volume *V*, and energy *E* kept constant, rather 'local'

- Canonical (*NVT*) ensemble

  Coupled to a heat bath to ensure constant *T*

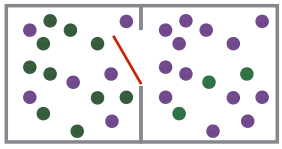
- Isobaric, isothermic (*NPT*) ensemble

  Coupled to a heat and pressure bath (thermostat and barostat)

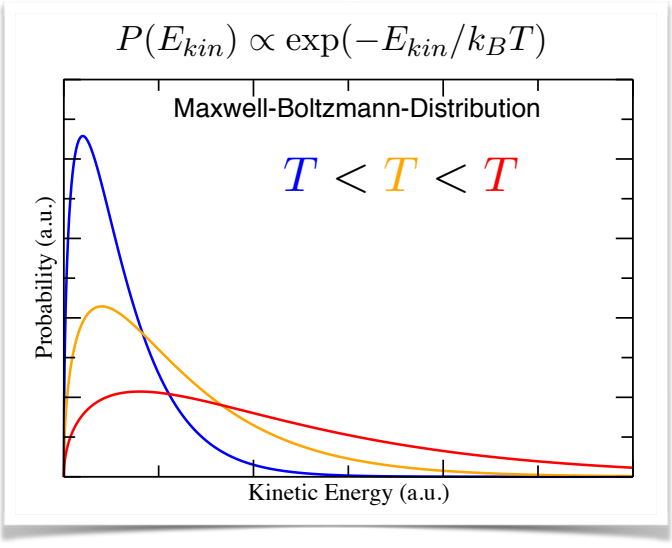


- Grand-canonical ensemble

  Energy and particle exchange



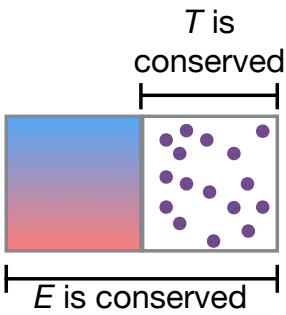
What is temperature?



$$\langle T \rangle = \frac{2\langle E_{kin} \rangle}{3Nk_B}$$

instantaneous temperature

How and why to keep Temperature constant?



- Experiments happen at finite T
- Conformational transition become possible since *E* in the sample can rise

- Instantaneous *T(t)* depends on E_{kin} depends on *v(R)*

Idea 1: Rescale all *v* in order to match *T* to desired value
… Berendsen thermostat
… no canonical ensemble, "*v* not Boltzmann distributed"

Idea 2: at every n-th step, adjust a particle's *v* from a Boltzmann distributions at desired *T* (stochastic collisions with heat bath)
… Andersen thermostat
… canonical, but takes long to equilibrate

Nosé-Hoover Thermostat

Extend the **Original System** by an **Oscillator**

$$\mathcal{H}_{NH} = \sum_I \frac{\mathbf{p}_I^2}{2M_I} + V(\mathbf{R}) + \frac{p_\eta^2}{2Q} + 3Nk_BT\eta$$

- *Q* is the mass of the fictitious oscillator
  - too low, to high frequency of harm. motion
  - too high, slow equilibration
- Canonical ensemble sampled, augmented total energy is conserved
- But: trajectories still feel the harmonic oscillation
- Solution: Chain of oscillators

S.Nosé, *J.Chem.Phys.* **81**,511(1984) & W.G.Hoover, *Phys.Rev.A* **31**,1695(1985).

Flavors of MD w.r.t. the energy function

- First-principles MD
  Solving the Schrödinger eq., Born-Oppenheimer or Car-Parrinello MD

electrons
1000 atoms
100 ps … 1ns

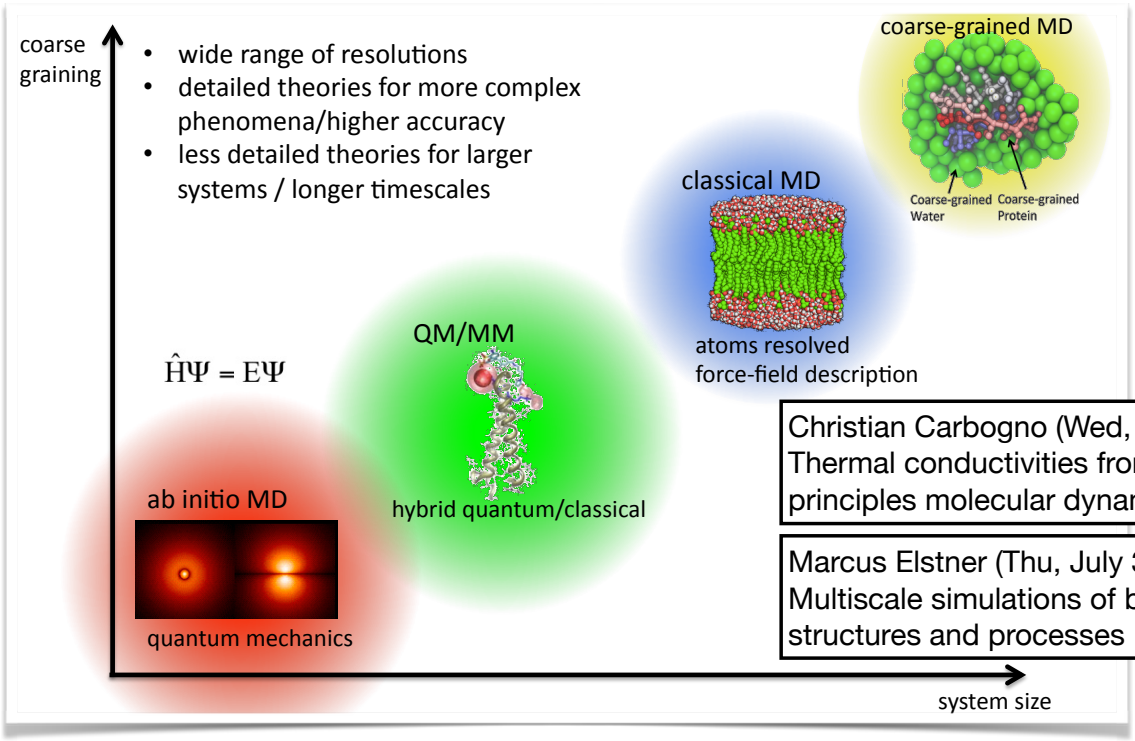
- Empirical potentials (a.k.a. force fields)
  Classical mechanics, fitted to experimental data or higher level theory
  How valid outside of their parametrization?

atoms
100,000s of atoms
µs … ms

- Coarse-grained models
  From bottom: Combine multiple atoms into one entity, parametrize from small simulations
  From top: Break a large system into a granular structure that one can parametrize to reach experimental accuracy

blobs
[… depends …]
long

Flavors of MD w.r.t. the energy function



Christian Carbogno (Wed, July 29)
Thermal conductivities from first-principles molecular dynamics

Marcus Elstner (Thu, July 30)
Multiscale simulations of biological structures and processes

Fazit

Good starting point?

Energy function valid?

Like an experiment:
1. Sample preparation
2. Equilibration
3. Measurement

[Smit and Frenkel]

Right conditions?

Long enough?

Numerical settings OK?

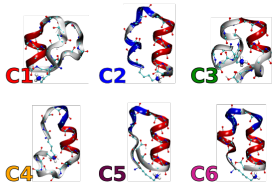
Single run enough?

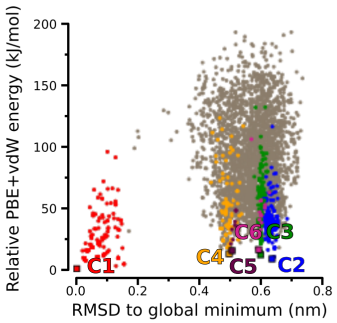
Target is the peptide Ac-LysH$^+$-Ala$_{19}$

Exploration of a large 'search space' to find minima with **force field** and ***ab initio*** replica-exchange MD

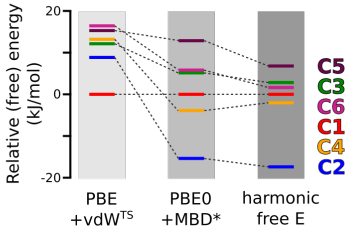
Clustering of conformers:
- Energy ordering
- RMSD clustering
- Machine learning of descriptors

Overview of the PES at a reasonable level of theory



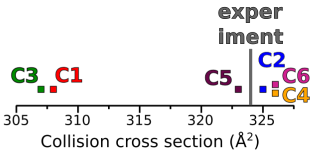


Refinement by applying *higher levels of theory* and *free energy corrections*

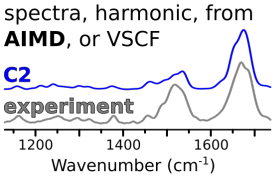


Comparison to experiment by computing physical observables

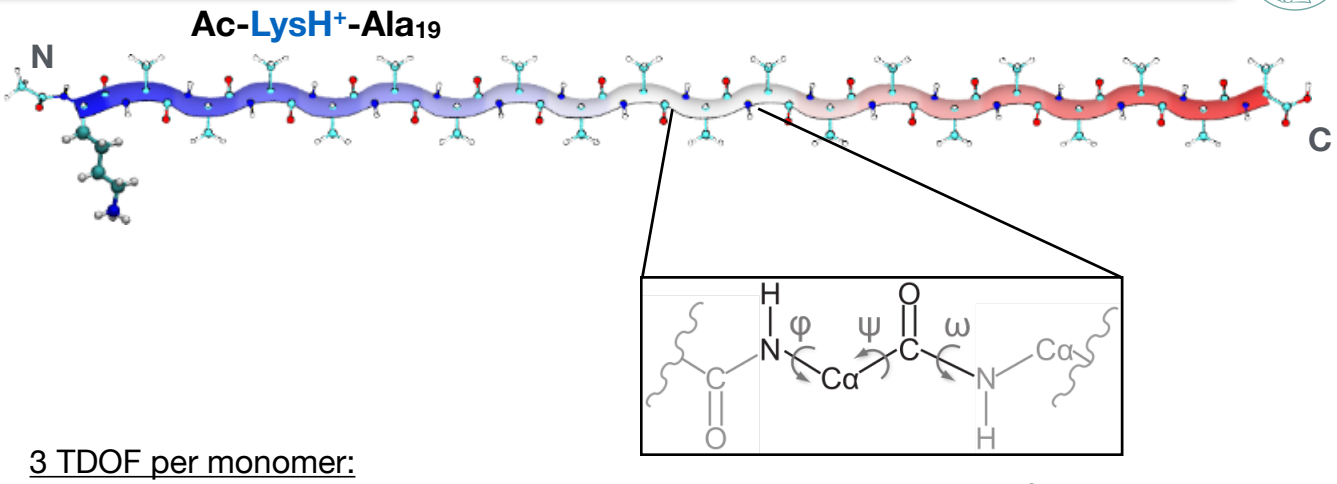
Computing CCS by means of PA, EHSS, or **TM**.



Simulating vibrational spectra, harmonic, from **AIMD**, or VSCF



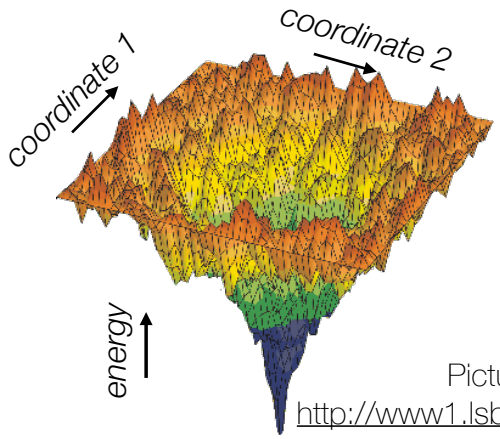
First of all, it's a search problem



3 TDOF per monomer:

- 60 degrees grid for single bonds
- *cis* or *trans* for peptide bond
- 2 × 6 × 6 = 72

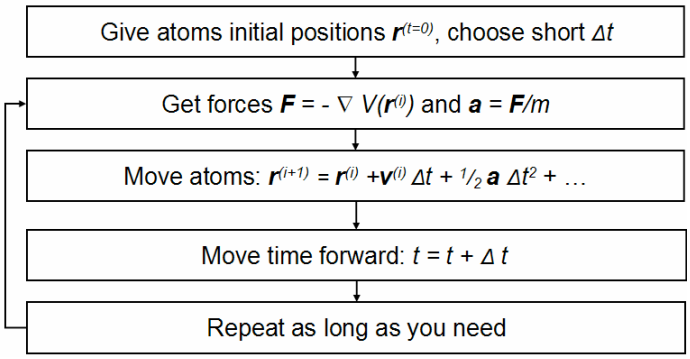
Results in 72^{20} trial structures in a 20mer



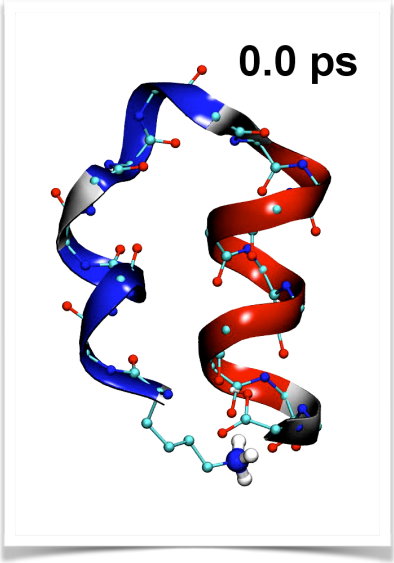
Picture from http://www1.lsbu.ac.uk

Molecular dynamics (MD) sampling

- Numerically solving Newton's equations of motion



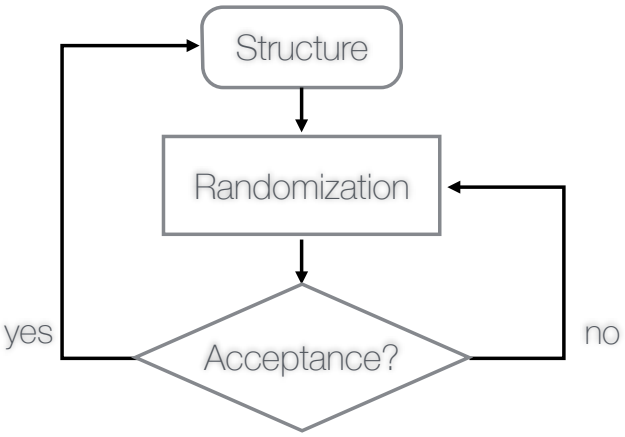
Scheme from wikipedia

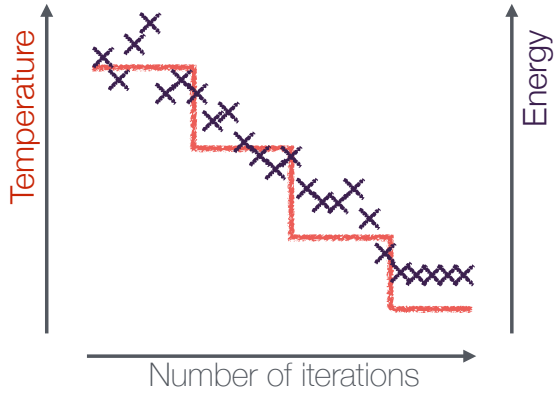


Video by F. Schubert

- **Pro:** generates a meaningful ensemble of conformers
- **Contra:** rather limited sampling of conformational space; for short trajectories, mainly around starting point

Simulated annealing





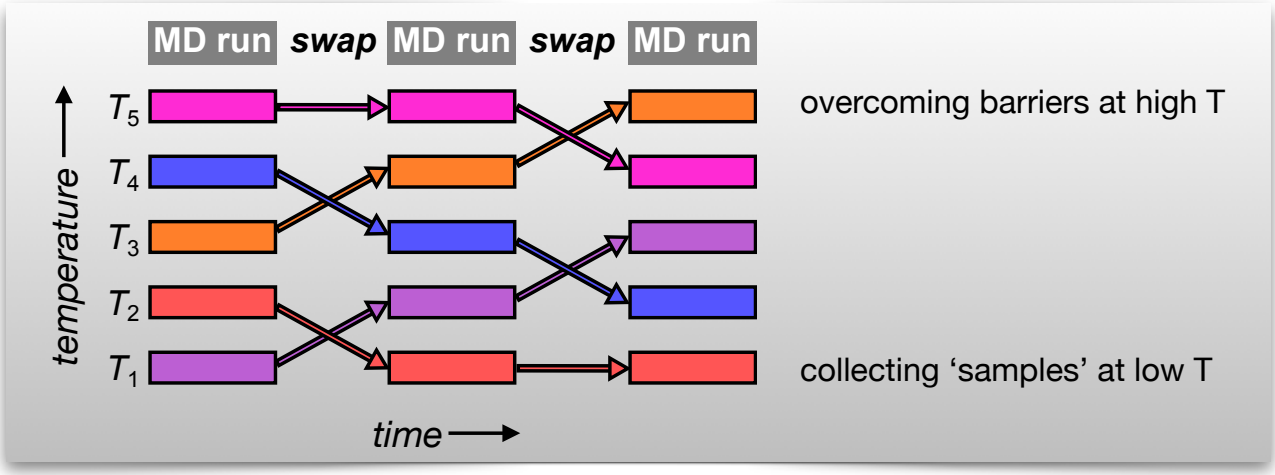
Metropolis criterion

$$P(\Delta E) = \exp(-\Delta E / k_B T)$$

Stepwise reduction of T

- overcoming barriers at high T
- converging with lowering T

Replica-exchange MD



- **Pro:** (i) improved sampling, (ii) generates a meaningful ensemble of conformers, (iii) opens access to free energy properties, (iv) assumption free
- **Contra:** still we are mainly sampling known territory

A realistic MD based search scheme

16 × 500 ns
FF REMD

Step 1: REMD search
OPLS-AA force field

Step 2: DFT relaxation
PBE+vdW, light

Global search

16 × 20 ps
ab initio REMD

Step 3: i) *ab initio* REMD
PBE+vdW, light, 200-800K

Clustering

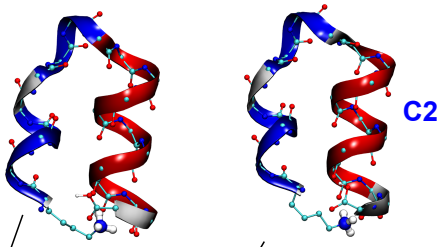
ii) DFT relaxation
PBE+vdW, light

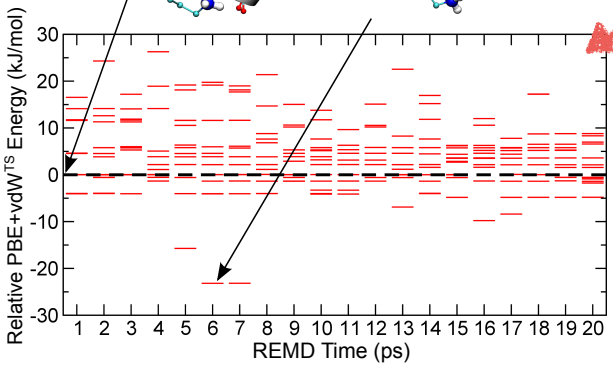
Clustering

iii) DFT relaxation
PBE+vdW, tight

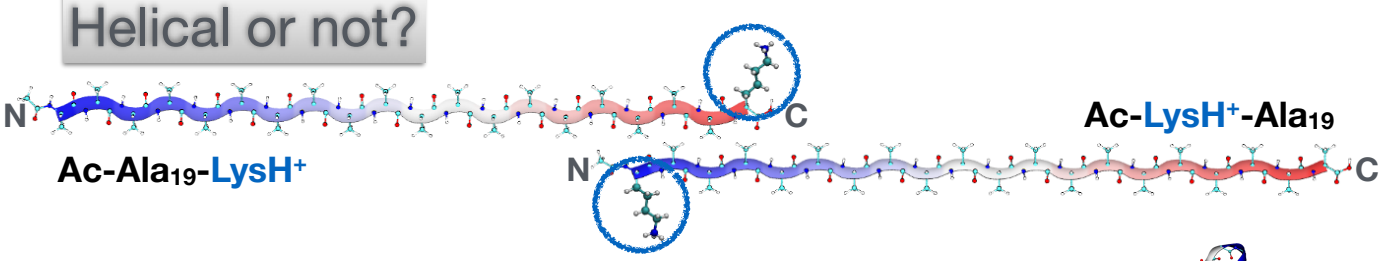
Refinement

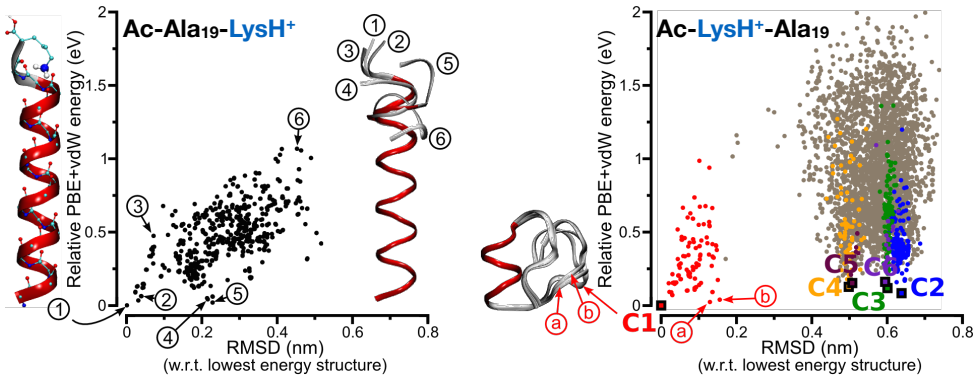
initial geometry

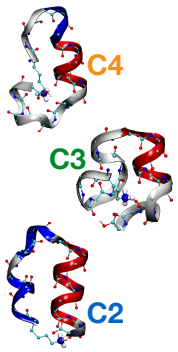




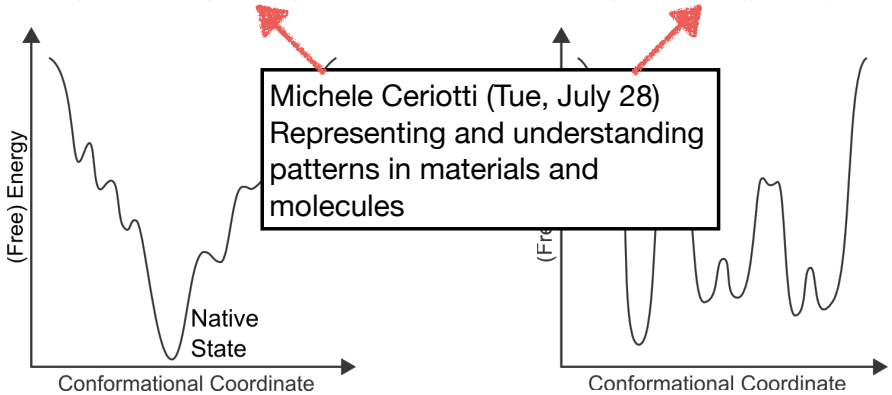
Helical or not?



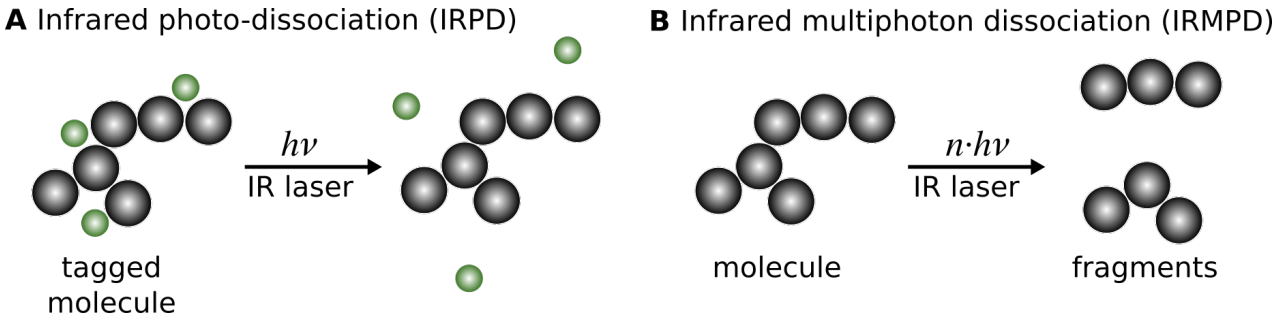


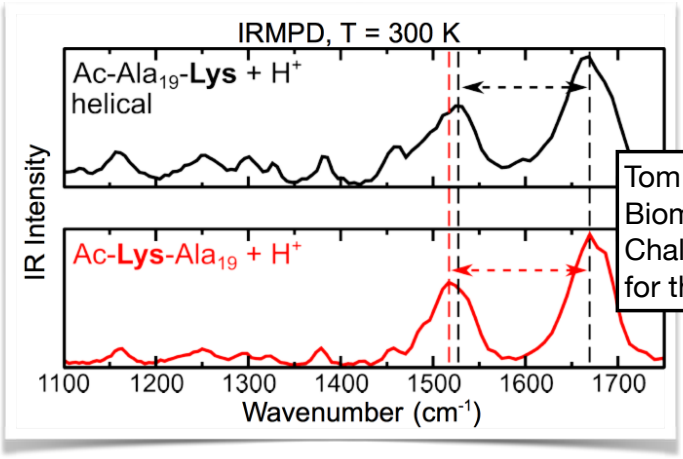


Michele Ceriotti (Tue, July 28)
Representing and understanding patterns in materials and molecules



Gas-phase IR, action spectroscopy

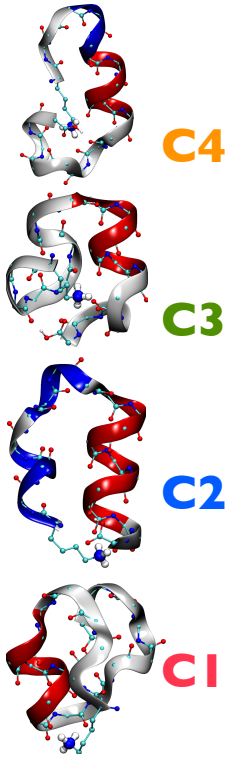




Tom Rizzo (Tue, July 28)
Biomolecules in isolation — Challenges and benchmarks for theory

Experiment vs. theory

Ac-LysH⁺-Ala19

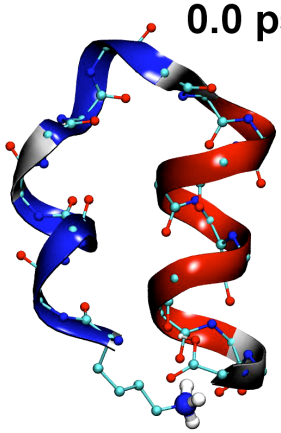


Infrared spectra

$$I(\omega) \propto \omega^2 \int_{-\infty}^{\infty} \underbrace{\langle \vec{\mu}(0) \cdot \vec{\mu}(t) \rangle}_{\text{Dipole-dipole time correlation function}} \mathrm{e}^{-i\omega t} \mathrm{d}t$$

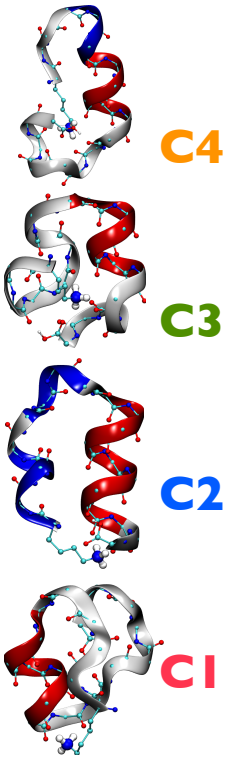
25 ps Born-Oppenheimer MD, PBE+vdW

0.0 ps



Experiment vs. theory

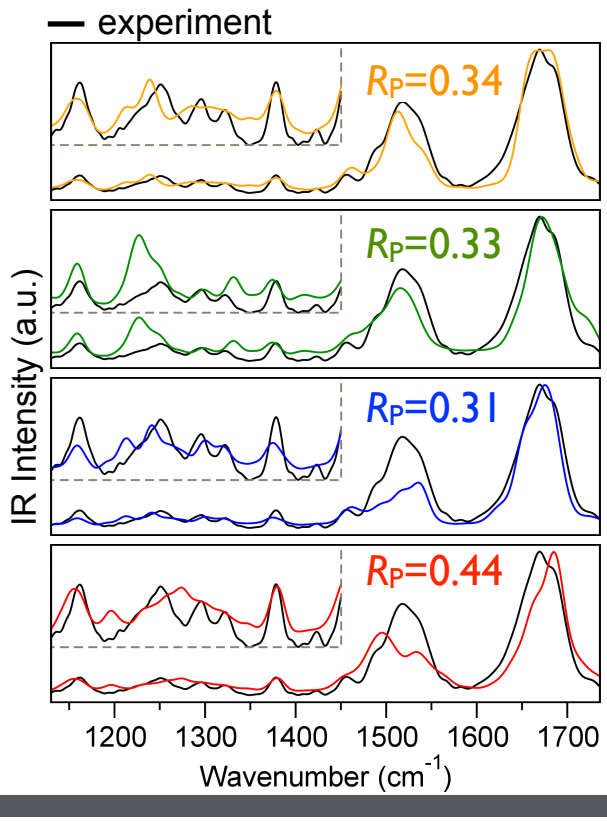
Ac-LysH$^+$-Ala$_{19}$



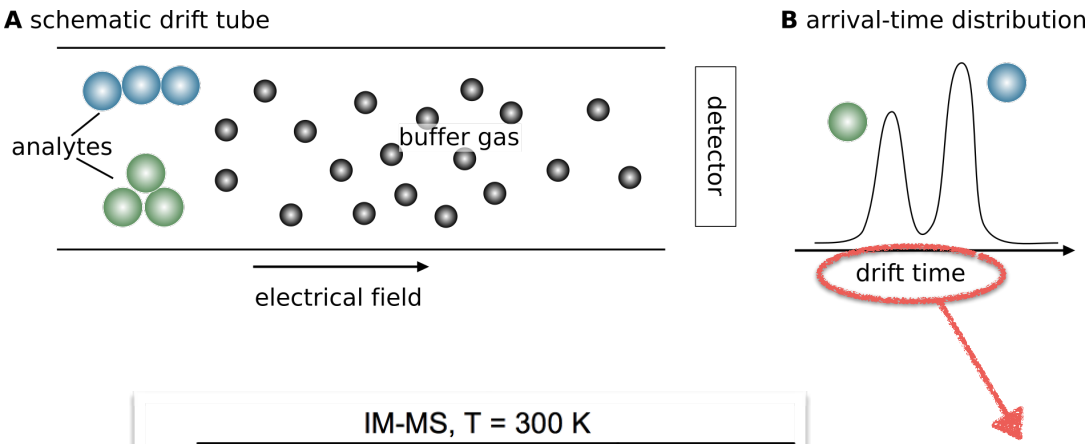
Infrared spectra

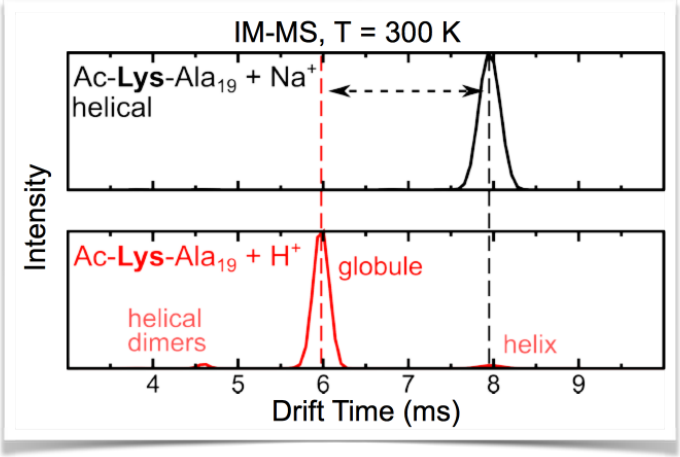
✓ ✓ ✓ ✗

R_P=0: perfect agreement
R_P=1: uncorrelated



Ion mobility-mass spectrometry

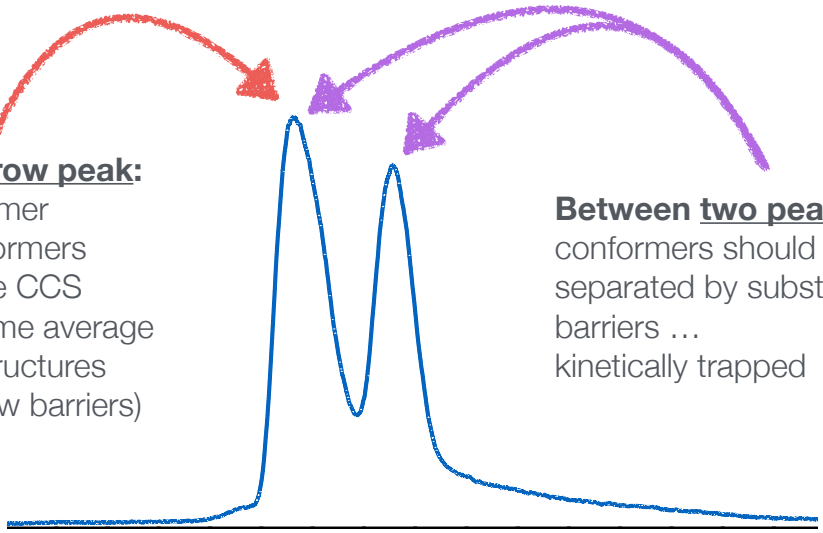




Can be converted to a **collision cross section** (CCS):

- transferable property of an ionic species
- depends on buffer gas
- independent of the machine

Peak considerations



Within <u>one narrow peak</u>:

1. Single conformer
2. Multiple conformers with the same CCS
3. Converged time average of different structures (sufficiently low barriers)

Between <u>two peaks</u>, conformers should be separated by substantial barriers … kinetically trapped

Three basic approaches

PA

$$\Omega_{\text{avg}}^{(1,1)} \approx \frac{1}{4\pi^2}\int_0^{2\pi} d\theta \int_0^{\pi} d\varphi \sin\varphi \int_0^{2\pi} d\gamma\, \pi b_{\min}^2$$

Shape of the ion

EHSS

$$\Omega_{\text{avg}}^{(1,1)} = \frac{1}{4\pi^2}\int_0^{2\pi} d\theta \int_0^{\pi} d\varphi \sin\varphi \int_0^{2\pi} d\gamma \times \int_0^{\infty} db\, 2b(1-\cos\chi(\theta,\varphi,\gamma,b))$$

Scattering

TM

$$\Omega_{\text{avg}}^{(1,1)} = \frac{1}{4\pi^2}\int_0^{2\pi} d\theta \int_0^{\pi} d\varphi \sin\varphi \int_0^{2\pi} d\gamma \frac{\pi}{8} \times \left(\frac{\mu}{k_B T}\right)^3 \int_0^{\infty} dg\, e^{-\mu g^2/2k_B T} g^5 \times \int_0^{\infty} db\, 2b(1-\cos\chi(\theta,\varphi,\gamma,g,b)).$$

Interaction between ion and buffer gas

Formulas from: Shvartsburg & Jarrold *Chem. Phys. Lett.* **1996**, 261, 86.

Projection approximation (PA)

- Pairwise potential including
  - Lennard-Jones potential
  - Charge-dipole interaction
- Tabulated pairwise collision parameters (radii of spheres for atom types)

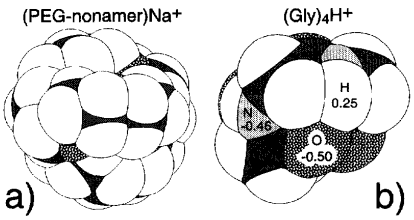
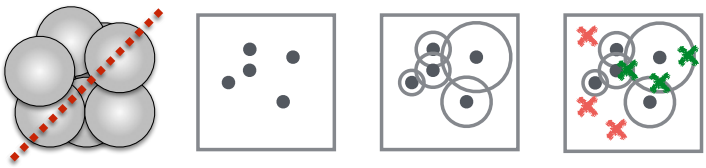


Figure 2. Space-filling models of (a) sodiated PEG nonamer, (b) protonated tetraglycine (labeled are one amid-N, -O, and -H with their MM point charges). (The nuclear positions of the atoms are fixed at the locations determined from molecular mechanics calculations. The size of each atomic radius, however, is set as the appropriate collision radius determined by the fits described in this work. For example, the space-filling radius for H is really the H–He collision radius determined from fitting the temperature dependence of the PEG systems. These space-filling models then directly reflect the collisional properties of the molecule for He as a collision partner.) C, white; C, dark gray; N, light gray; O, dotted.

- Random selection of a plane
- Projection of nuclear positions
- Drawing of the collision radii
- Random selection of points in area A enclosing the projection



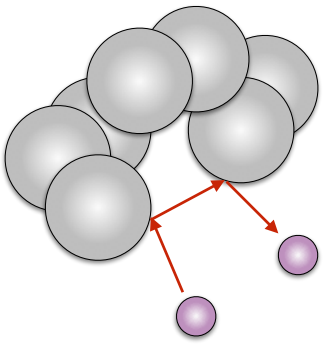
CCS_{Proj1} = (hits/tries) x Area

Wyttenbach, von Helden et al. *J. Am. Soc. Mass Spectrom.* **1997**, 8, 275.

Exact hard-sphere scattering (EHSS)

- PA neglects scattering and multiple collisions of a gas atom with the ion
- Still PA is correct for fully convex ions

- EHSS explicitly considers scattering



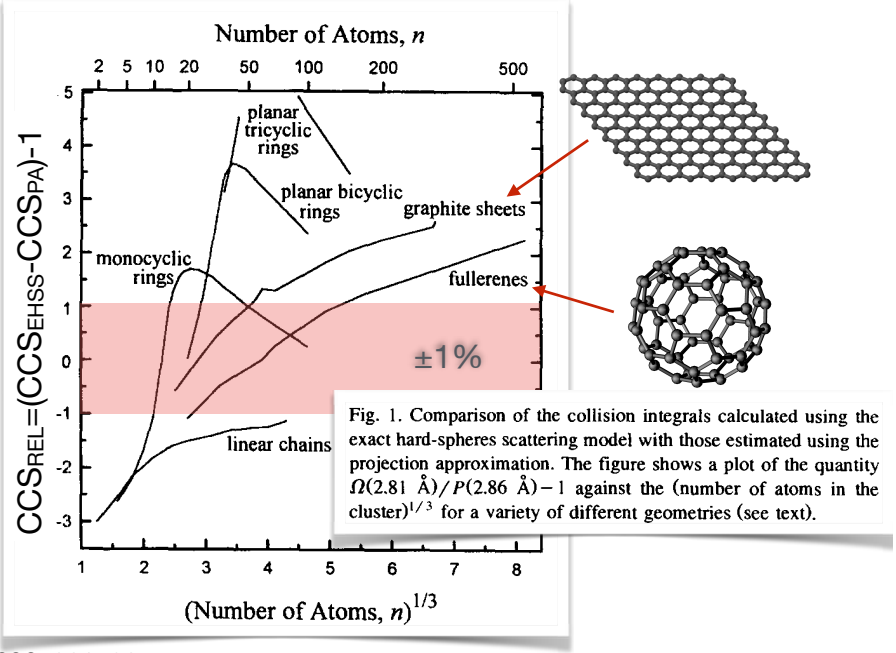
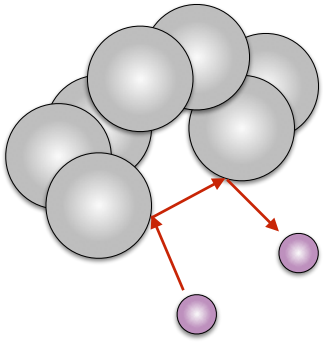


Fig. 1. Comparison of the collision integrals calculated using the exact hard-spheres scattering model with those estimated using the projection approximation. The figure shows a plot of the quantity $\Omega(2.81\ \text{Å})/P(2.86\ \text{Å})-1$ against the (number of atoms in the cluster)$^{1/3}$ for a variety of different geometries (see text).

Shvartsburg & Jarrold *Chem. Phys. Lett.* **1996**, 261, 86.

Trajectory method (TM)

- EHSS plus long-range interactions



$$\Phi(\theta,\phi,\gamma,b,r) = 4\epsilon\sum_i^n\left[\left(\frac{\sigma}{r_i}\right)^{12} - \left(\frac{\sigma}{r_i}\right)^{6}\right] - \frac{\alpha}{2}\left(\frac{ze}{n}\right)^2\left[\left(\sum_i^n \frac{x_i}{r_i^3}\right)^2 + \left(\sum_i^n \frac{y_i}{r_i^3}\right)^2 + \left(\sum_i^n \frac{z_i}{r_i^3}\right)^2\right]$$

Lennard-Jones Potential

Ion-induced Dipole Interaction

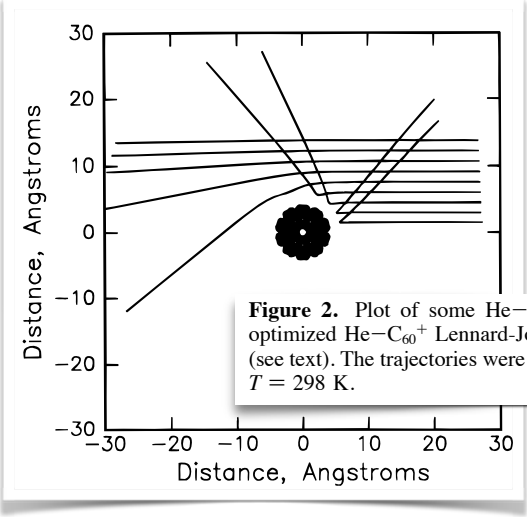
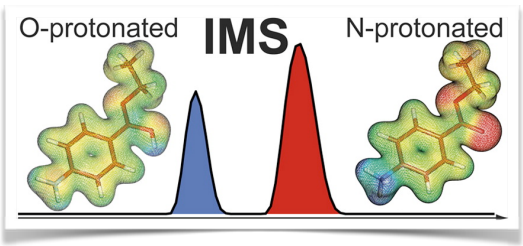


Figure 2. Plot of some $He-C_{60}^+$ trajectories calculated using the optimized $He-C_{60}^+$ Lennard-Jones plus ion-induced dipole potential (see text). The trajectories were run with a collision energy of k_BT with $T = 298$ K.

Mesleh et al. *J. Phys. Chem.* **1996**, 100, 16082.

CCS calculations compared to experiment

Structure	$CCS_{PA/PSA}$ in Å^2	CCS_{TM} in Å^2	CCS_{Exp} in Å^2
Ac-Ala_6-Lys(H^+) from ref. [99]			
α helix	180	181	180
compact	172	171	
Ac-β^2hAla_6-Lys(H^+) from ref. [99]			
H12	203	204	190
H16	191	193	
H20	182	182	
compact	183	182	
Benzocaine from ref. [47]			
O-prot./trans	131.1	132.7	135
O-prot./gauche	131.6	132.5	
N-prot./trans	133.3	144.0	155
N-prot./gauche	129.5	144.0	



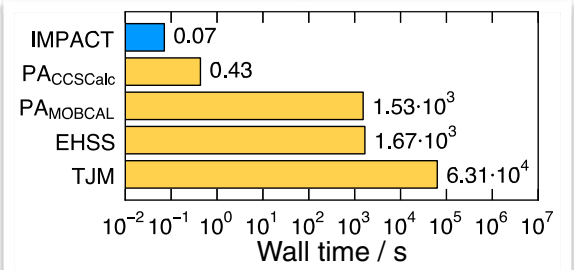
[47] Warnke S, Seo J, Boschmans J, Sobott F, Scrivens J H, Bleiholder C, Bowers M T, Gewinner S, Schöllkopf W, Pagel K and von Helden G 2015 *Journal of the American Chemical Society* **137** 4236–4242

[99] Schubert F, Pagel K, Rossi M, Warnke S, Salwiczek M, Koksch B, Helden G v, Blum V, Baldauf C and Scheffler M 2015 *Phys. Chem. Chem. Phys.* **17** 5376–5385

Summary and performance

	PA		EHSS		TM
Shape	✓		✓		✓
Scattering	✗		✓		✓
Multiple collisions	✗		✓		✓
Long-range interactions	✗		✗		✓
Computational cost	PA	≈	EHSS	≪	TM

Performance benchmarks:
asymmetric unit of the Norwalk virus capsid (PDB code 1IHM) calculated to 1% precision.
Timings range from below **1s to ~18h**



Marklund et al. *Structure* **2015**, 23, 1.

CCS packages

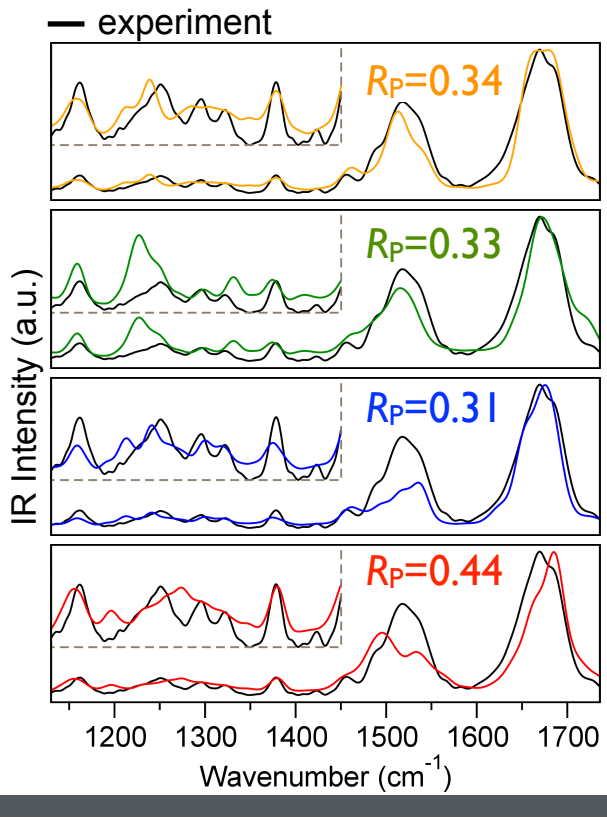
- MOBCAL, Jarrold group (the only one that you can simply download)
  http://www.indiana.edu/~nano/software.html
  (PA, EHSS, TM)
- Sigma, Bowers group (available upon request)
  http://bowers.chem.ucsb.edu/theory_analysis/cross-sections/sigma.shtml
  (PA, EHSS)
- IMPACT, Benesch lab (to be available soon)
  http://benesch.chem.ox.ac.uk/resources.html
  (PA, promises to be very fast)
- CCSCalc from Waters (never tried, not found)
- There are several advanced methods published but with limited availability of code

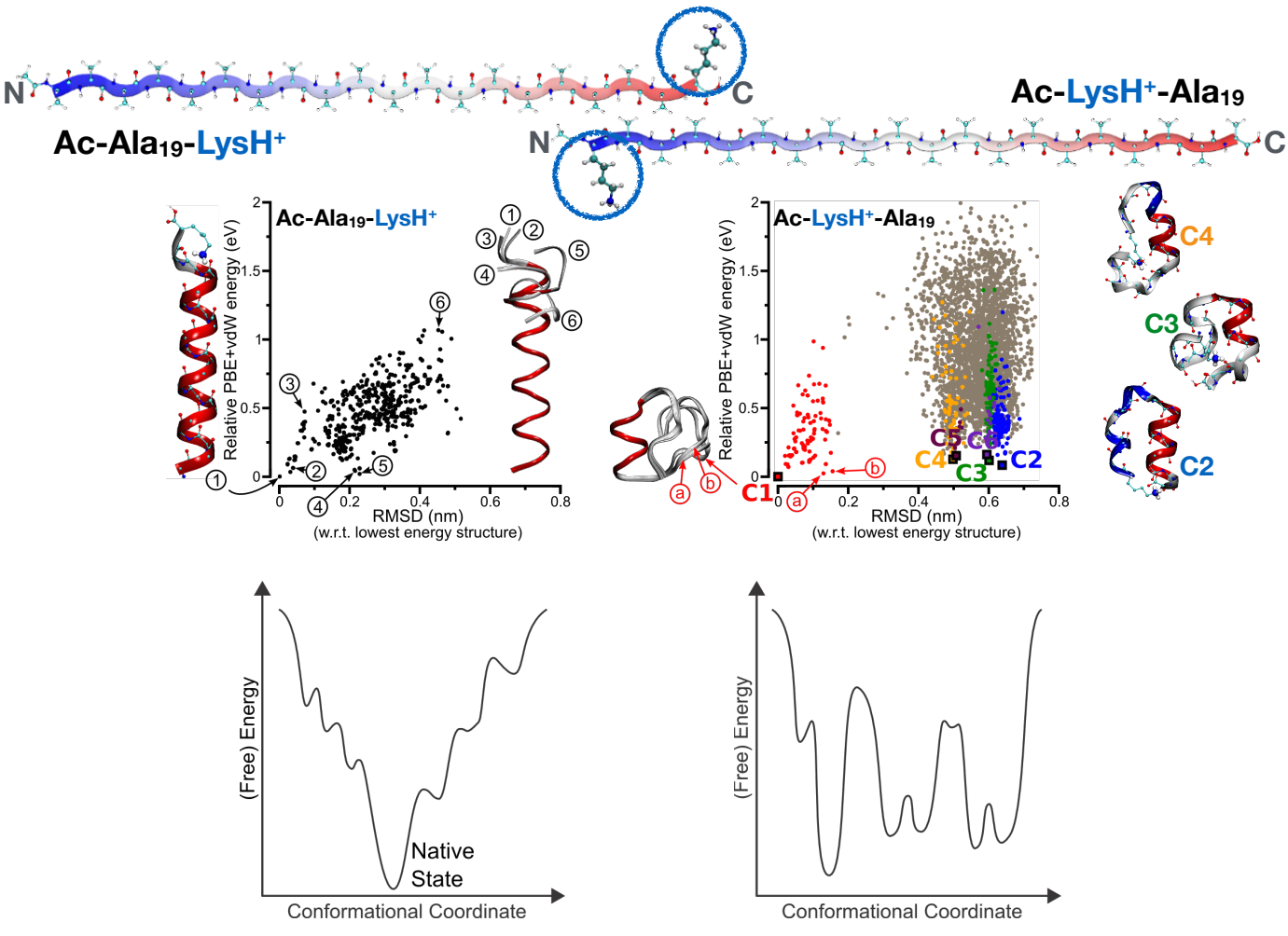
Experiment vs. theory

Ac-LysH⁺-Ala₁₉

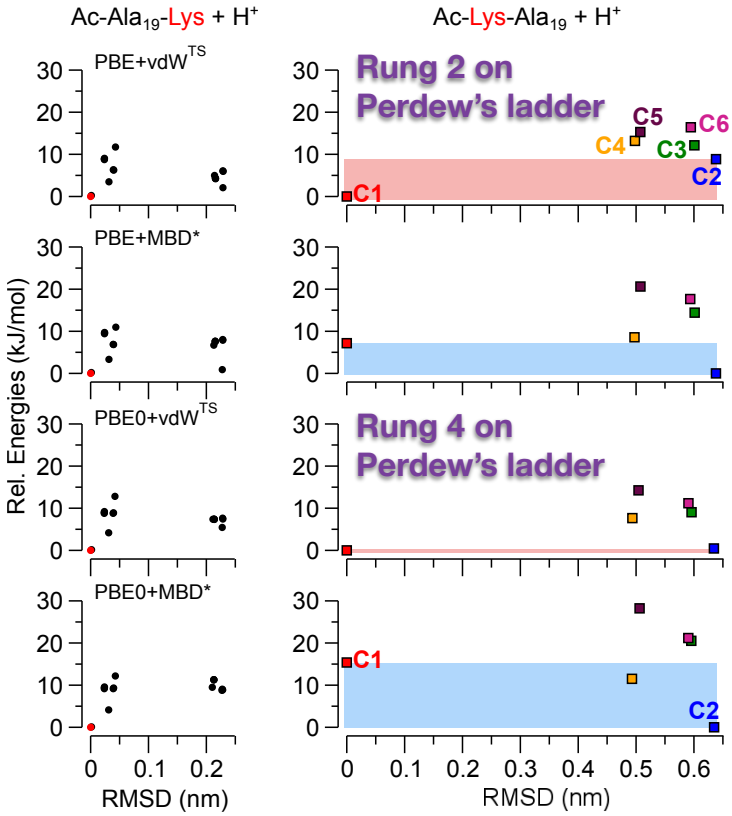
	exp. **CCS** from IM-MS (324 Å²)	Infrared spectra
C4	✓ (326 Å²)	✓
C3	✗ (307 Å²)	✓
C2	✓ (325 Å²)	✓
C1	✗ (308 Å²)	✗

R_P=0: perfect agreement
R_P=1: uncorrelated



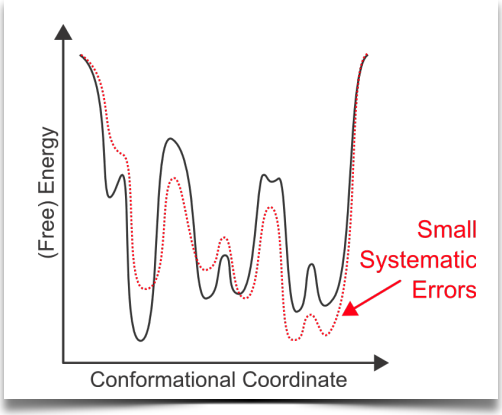


The correct level of theory



Almost no dispersion (van der Waals) below 5th rung of Perdew's ladder!
Solution:
Augment DFA by a correction:

- pairwise TS scheme
- many-body scheme (MBD*)



References

Talk by C. Carbogno (July 2011):
http://th.fhi-berlin.mpg.de/th/Meetings/DFT-workshop-Berlin2011/presentations/2011-07-19_Carbogno_Christian.pdf

Talk by M. Rossi (August 2013):
http://th.fhi-berlin.mpg.de/sitesub/meetings/DFT-workshop-2013/uploads/Meeting/Lecture_12_Rossi_HandsOnDFT2013.pdf
http://th.fhi-berlin.mpg.de/sitesub/meetings/DFT-workshop-2013/uploads/Meeting/2013-11-19_Videos_ICTP_Trieste_640x360/2013_08_09_C.mp4

Talk by R. Netz (July 2015):
http://th.fhi-berlin.mpg.de/sitesub/meetings/dft-workshop-2015/uploads/Meeting/July-17_T11_Netz.pdf

Understanding Molecular Simulation
From Algorithms to Applications
by Daan Frenkel and Berend Smit
ISBN: 978-0-12-267351-1