Computational Chemistry and Materials Modeling

Lecture 8
Classical molecular dynamics

Sergei Tretiak, *Andriy Zhugayevych*

November 22, 2021

Outline

- Representing energy in classical molecular mechanics
  - Basic ideas
  - Force field (FF) elements
- Parameterization
  - Parameterization
  - Various types of empirical potentials
  - Computational efficiency
  - Typical workflow and practical guidelines
  - Illustrative examples
- Molecular dynamics
  - Molecular dynamics (MD)
  - Extending MD time scale
  - Nonadiabatic MD
- Potential energy surface (PES) exploration
  - Reminder on thermodynamics
  - Monte Carlo sampling

Reminder: Born-Oppenheimer approximation

Given separable Hamiltonian $\hat{H} = \hat{H}_1(q_1) + \hat{H}_2(q_2)$ for $\hat{H}\psi(q_1, q_2) = E\psi(q_1, q_2)$

Then $\psi(q_1, q_2) = \psi_1(q_1)\psi_2(q_2)$ (factorization) and $E = (E_1 + E_2)$ (additive)

$$\hat{H} = \hat{T}_N(\mathbf{R}) + \hat{T}_e(\mathbf{r}) + \hat{V}_{NN}(\mathbf{R}) + \hat{V}_{eN}(\mathbf{r}, \mathbf{R}) + \hat{V}_{ee}(\mathbf{r})$$

Approximately separable! $\quad \phi_T(\mathbf{r}, \mathbf{R}) = \phi_e(\mathbf{r}; \mathbf{R})\phi_N(\mathbf{R})$

$$\hat{H}\phi_e(\mathbf{r}; \mathbf{R})\phi_N(\mathbf{R}) = E_{tot}\phi_e(\mathbf{r}; \mathbf{R})\phi_N(\mathbf{R})$$

Electronic problem: $\quad \hat{H}_e = \hat{T}_e(\mathbf{r}) + \hat{V}_{eN}(\mathbf{r}; \mathbf{R}) + \hat{V}_{ee}(\mathbf{r})$

$$\hat{H}_e\phi_e(\mathbf{r}; \mathbf{R}) = \left\{ -\frac{1}{2}\sum_i \nabla_i^2 - \sum_{A,i} \frac{Z_A}{r_{Ai}} + \sum_{i>j} \frac{1}{r_{ij}} \right\} \phi_e(\mathbf{r}; \mathbf{R}) = E_e(\mathbf{R})\phi_e(\mathbf{r}; \mathbf{R})$$

Nuclei problem: $\quad \{\hat{T}_N + E_e + \hat{V}_{NN}\}\phi_N(\mathbf{R}) = E_{tot}\phi_N(\mathbf{R})$

$$\hat{H}_N\phi_N(\mathbf{R}) = \left\{ -\sum_A \frac{1}{2M_A}\nabla_A^2 + E_e(\mathbf{R}) + \sum_{A>B} \frac{Z_A Z_B}{R_{AB}} \right\} \phi_N(\mathbf{R}) = E_{tot}\phi_N(\mathbf{R})$$

i.e., the nuclei move in a potential created by the electrons.

When molecular mechanics works
("exactly" in Born-Oppenheimer approximation)

1) Not interested in electronic properties

2a) Electronic state remains unchanged (insulators)

2b) Electronic dynamics is irrelevant (metals)

Think about examples when MM doesn't work

- **Challenge** is to approximate exact PES by interatomic potentials

- **Grand challenge** is to make such potentials transferable

Think about examples when transferability is hardly possible

Molecular mechanics

Molecular Mechanics methods use classical type models (no quantum mechanics) to predict the energy of a system as a function of atomic coordinates, i.e. approximate Potential Energy Surface (PES). This approach can be used to:

- Optimize geometry of minima or transition states

- Calculate relative energies between conformers or polymorphs

but is mostly used to:

- Run molecular dynamics

- Sample PES

- As part of QM/MM
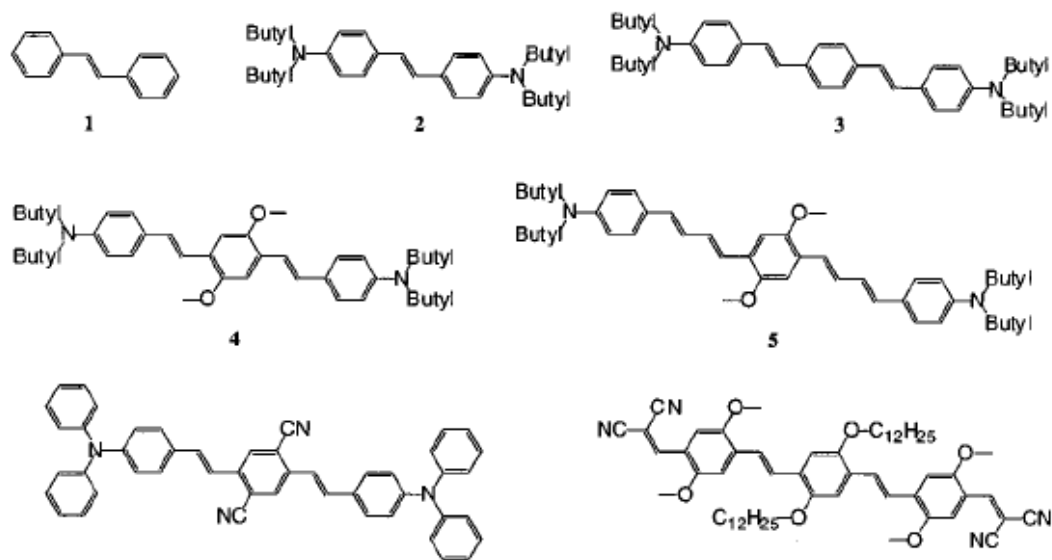
Why can we do it?

– Chemical bonding is local

– Number of local structural patterns is limited

– Nonlocal interactions are pairwise

Pair potentials, force fields, ML potentials

Pair potentials are used for gases, liquids, closely (co)packed lattices (metals, ionic solids), but not for systems with covalent bonds (molecules)

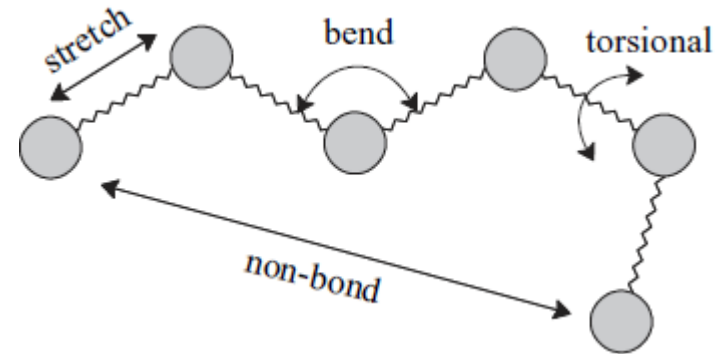- For solids embedded atom/ion model (EAM/EIM) accounts for collective (non-pairwise) interactions:
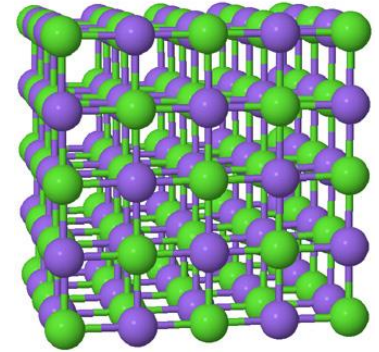
$$E_i = F_\alpha \left(\sum_{i \neq j} \rho_\beta(r_{ij}) \right) + \frac{1}{2} \sum_{i \neq j} \phi_{\alpha\beta}(r_{ij})$$

Force fields are used for molecular systems:

- Bonding interactions = stretching +
  + bending + torsion (dihedrals/impropers)

- Non-bonding interactions =
   van-der-Waals + electrostatics

- Cross-terms

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross}$$

Jensen: Illustration of the fundamental force field energy terms

ML potentials allow for systematic increase of accuracy
(will be discussed in a separate lecture)

The stretch energy

- Start with Taylor expansion of energy E(R) near the minimum:

$$E_{str}(R^{AB} - R_0^{AB}) = E(0) + \frac{dE}{dR}(R^{AB} - R_0^{AB}) + \frac{1}{2}\frac{d^2E}{dR^2}(R^{AB} - R_0^{AB})^2$$

Can be set to 0   *Vanishes at minimum*

$$E_{str}(\Delta R^{AB}) = k_2^{AB}(\Delta R^{AB})^2 + k_3^{AB}(\Delta R^{AB})^3 + k_4^{AB}(\Delta R^{AB})^4 + \cdots$$
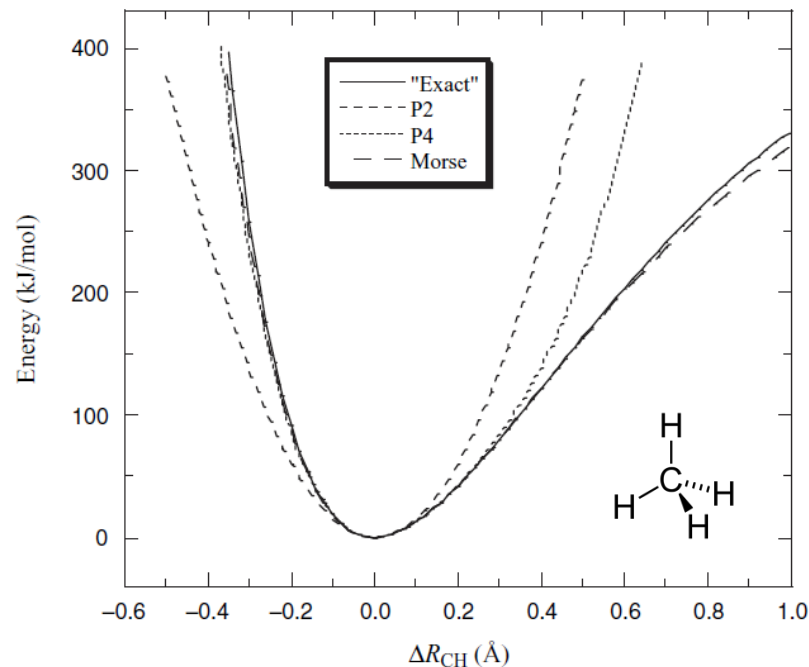
P2 term   *P4 term*

- Alternative forms include Morse potential:

$$E_{Morse}(\Delta R) = D(1 - e^{-\alpha\Delta R})^2$$

- However, numerically friendly polynomial expansion is usually used instead of Morse potential

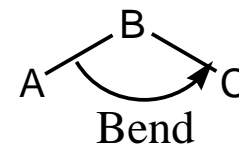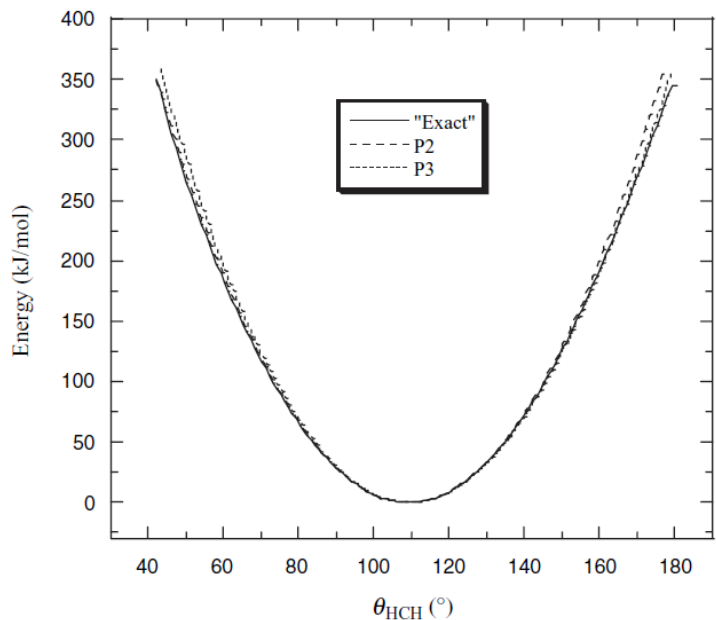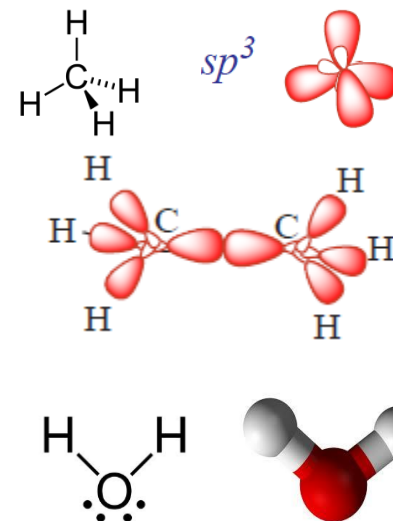- Every pair of atoms (A,B) requires at least 2 parameters: k_2^{AB} and R_0^{AB}

Jensen: The stretch energy for CH_4

7

The bending energy

$$E_{bend}(\theta^{ABC} - \theta_0^{ABC}) = k^{ABC}(\theta^{ABC} - \theta_0^{ABC})^2$$

Why? Repulsion between electrons

Jensen: The bending energy for CH_4
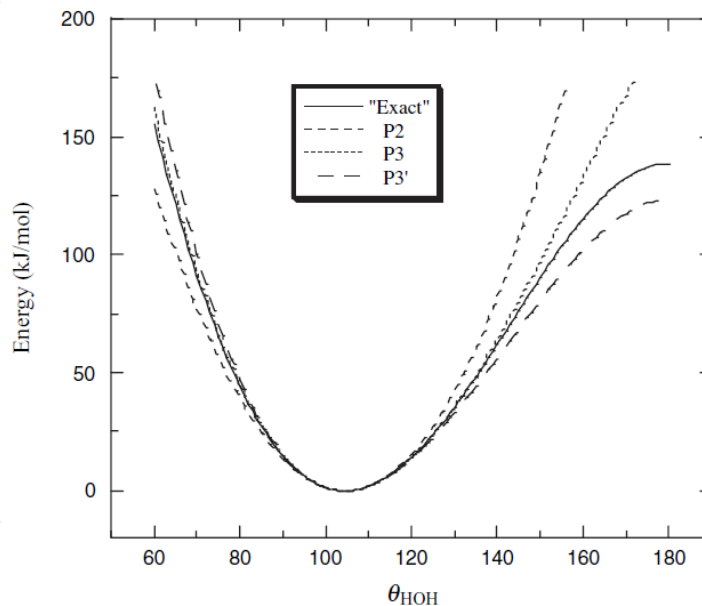
Jensen: The bending energy for H_2O

- Must be accurate in region of few kT above the minimum (accessible conformational space)

- Problems with θ=180°, multiple minima, and out-of-plane bending

- Every triple of atoms (A,B,C) requires 2 parameters: k^{ABC} and θ_0^{ABC}

The torsion energy

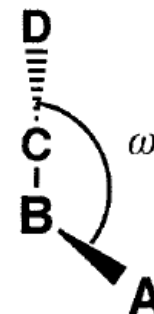$$E_{\text{tors}}(\omega^{ABCD}) = \tfrac{1}{2} V_1^{ABCD} \left[1 + \cos(\omega^{ABCD}) \right]$$
$$+ \tfrac{1}{2} V_2^{ABCD} \left[1 - \cos(2\omega^{ABCD}) \right]$$
$$+ \tfrac{1}{2} V_3^{ABCD} \left[1 + \cos(3\omega^{ABCD}) \right]$$

Dihedral angle is defined as signed angle between ABC and BCD planes

- Dihedrals are usually flexible and correspond to large-amplitude motions
- Energy scale is 1-2 orders of magnitude smaller than for stretching and bending
- Includes non-bonding interactions
- Every quartet of atoms (A,B,C,D) requires a set of parameters: V_1^{ABCD}, V_2^{ABCD}, V_3^{ABCD},...

Impropers

*Dihedral
vs
improper*
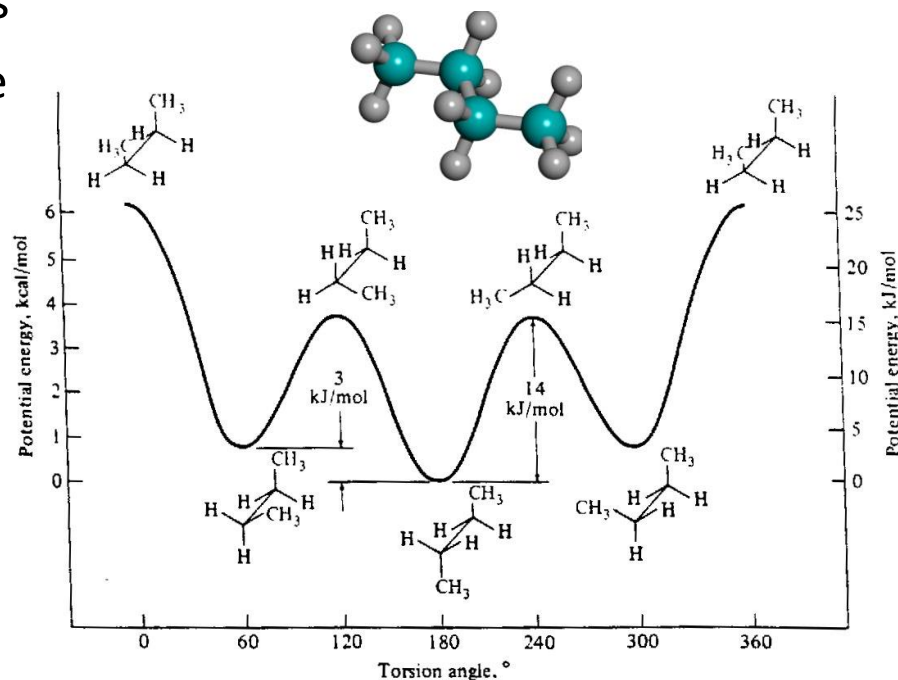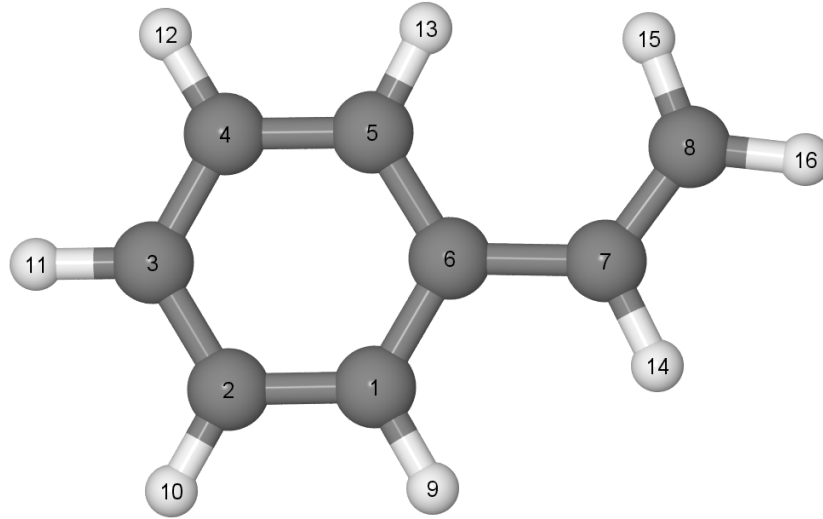
$$V_I = k_{ijkl}\left(\phi_{ijkl} - \phi_0\right)^2$$

- Correspond to small-amplitude motions, therefore harmonic potential is good enough

- Every planar quartet of atoms (A,B,C,D) requires only a single parameter: k^{ABCD}

The van der Waals (vdW) energy

- Accounts for all nonelectrostatic interaction between nonbonded atoms

- Repulsive at short distances due to steric interactions (Pauli exchange)

- Attractive at large distances due to dispersion interaction, induced dipoles

- The Lennard-Jones (LJ) potential is commonly used

- For LJ potential every pair of atoms (A,B) requires 2 parameters: ε^{AB} and R_0^{AB}

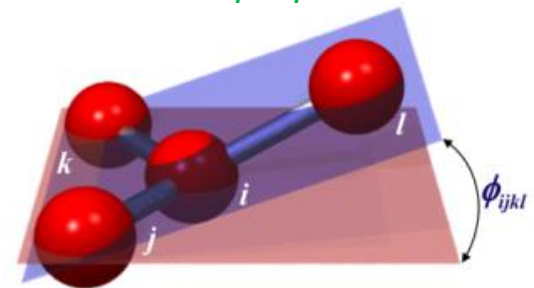$$E_{LJ}(R) = \varepsilon\left[\left(\frac{R_0}{R}\right)^{12} - 2\left(\frac{R_0}{R}\right)^6\right]$$

- To minimize number of parameters $\varepsilon^{AB} = (\varepsilon^{AA}\varepsilon^{BB})^{1/2}$

- Usually vdW interactions are excluded between bonded atoms, but problem of 'nonorthogonality' to bonded interactions exists

Some other forms:
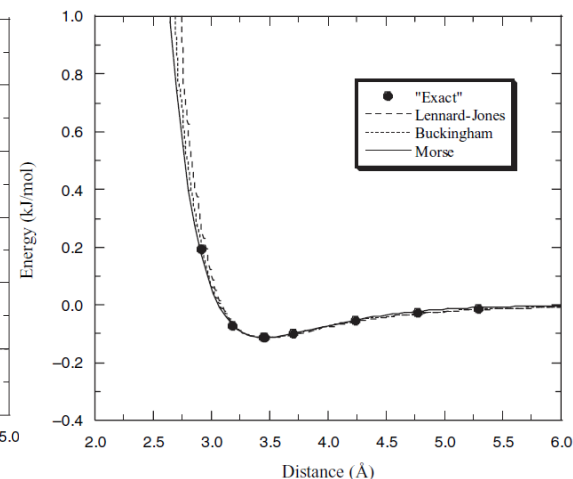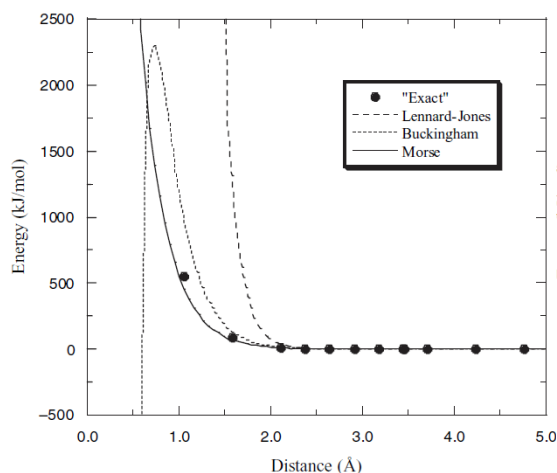- The Buckingham potential:

$$E_{Hill}(R) = \varepsilon\left[\frac{6}{\alpha-6}e^{\alpha(1-R/R_0)} - \frac{\alpha}{\alpha-6}\left(\frac{R_0}{R}\right)^6\right]$$

- Hydrogen bonding potential:

$$E_{H\text{-}bond}(R) = \varepsilon\left[5\left(\frac{R_0}{R}\right)^{12} - 6\left(\frac{R_0}{R}\right)^{10}\right]$$

Jensen: Comparison of E_{vdw} functionals for the H_2—He potential

The electrostatic energy

- Strategy 1: Assign atomic charges

Option 1. Calculate molecular electrostatic potential using electronic structure methods

$$\phi_{esp}(\mathbf{r}) = \sum_a^{N_{nuc}} \frac{Z_a}{|\mathbf{R}_a - \mathbf{r}|} - \int \frac{\Psi^2(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}'$$

then minimize error-function

$$\mathrm{ErrF}(\mathbf{Q}) = \sum_r^{N_{points}} \left(\phi_{esp}(\mathbf{r}) - \sum_a^{N_{atoms}} \frac{Q_a(\mathbf{R}_a)}{|\mathbf{R}_a - \mathbf{r}|} \right)^2$$

but the result depends on fitting domain.
Option 2. Use other quantum chemistry charges:
Hirshfeld or CM5 charges should work well.

Point-charge Coulomb interaction:

$$E_{el}(R^{AB}) = \frac{Q^A Q^B}{\varepsilon R^{AB}}$$

Dipole-dipole interaction:

$$E_{el}(R^{AB}) = \frac{\mu^A \mu^B}{\varepsilon (R^{AB})^3} (\cos\chi - 3\cos\alpha_A \cos\alpha_B)$$

- Strategy 2: Fit atomic charges (e.g. by energy)

Usually problematic because of ill-conditioned nonlinear optimization problem.

- Strategy 3: Assign bond dipoles

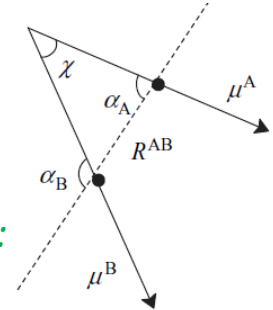How? Using experimental data or electronic structure methods.

- Strategy 4: Use multipoles and polarizabilities

See e.g. Effective Fragment Potentials (EFP) best for small molecules in a liquid:
Annu Rev Phys Chem 64, 553 (2013), github.com/ilyak/libefp

Cross-terms

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross}$$

The various terms in the total energy are not independent, and cross-terms should be considered:

$$u_i(\mathbf{x}) = u_i(0) + \sum_j x_j \left(\frac{\partial u_i}{\partial x_j} \right)_0 + \sum_{k \leqslant j} \sum_j x_j x_k \left(\frac{\partial^2 u_i}{\partial x_j \partial x_k} \right)_0$$

For example, stretching-bending:

$$E_{str/bend} = k^{ABC}(\theta^{ABC} - \theta_0^{ABC})[(R^{AB} - R_0^{AB}) - (R^{BC} - R_0^{BC})]$$

Other terms:

$$E_{str/str} = k^{ABC}(R^{AB} - R_0^{AB})(R^{BC} - R_0^{BC})$$

$$E_{bend/bend} = k^{ABCD}(\theta^{ABC} - \theta_0^{ABC})(\theta^{BCD} - \theta_0^{BCD})$$

$$E_{str/tors} = k^{ABCD}(R^{AB} - R_0^{AB})\cos(n\omega^{ABCD})$$

$$E_{bend/tors} = k^{ABCD}(\theta^{ABC} - \theta_0^{ABC})\cos(n\omega^{ABCD})$$

$$E_{bend/tors/bend} = k^{ABCD}(\theta^{ABC} - \theta_0^{ABC})(\theta^{BCD} - \theta_0^{BCD})\cos(n\omega^{ABCD})$$

However, cross-terms are rarely used because of problems with their parameterization.

Discussion

1. Slide 4: Give examples when MM doesn't work.

2. Slide 4: Give examples when transferability is hardly possible.

3. Illustrate all elements of a force field using this molecule:
   e.g. *pair of atoms (1,2) – bond stretching.*
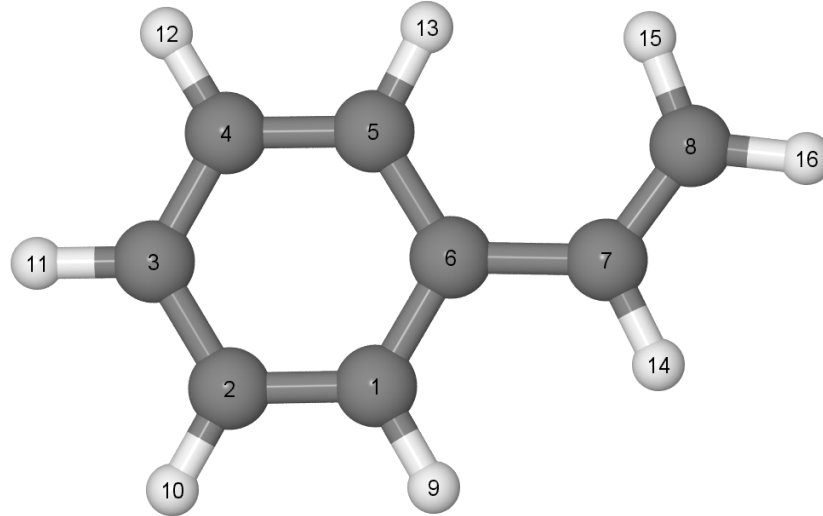
MM2 force field example: list of atom types

| Type	Symbol	Description	Type	Symbol	Description
1	C	sp^3-carbon	28	H	enol or amide
2	C	sp^2-carbon, alkene	48	H	ammonium
3	C	sp^2-carbon, carbonyl, imine	36	D	deuterium
4	C	sp-carbon	20	lp	lone pair
22	C	cyclopropane	15	S	sulfide (R_2S)
29	C·	radical	16	S^+	sulfonium (R_3S^+)
30	C^+	carbocation	17	S	sulfoxide (R_2SO)
38	C	sp^2-carbon, cyclopropene	18	S	sulfone (R_2SO_2)
50	C	sp^2-carbon, aromatic	42	S	sp^2-sulfur, thiophene
56	C	sp^3-carbon, cyclobutane	11	F	fluoride
57	C	sp^2-carbon, cyclobutene	12	Cl	chloride
58	C	carbonyl, cyclobutanone	13	Br	bromide
67	C	carbonyl, cyclopropanone	14	I	iodide
68	C	carbonyl, ketene	26	B	boron, trigonal
71	C	ketonium carbon	27	B	boron, tetrahedral
8	N	sp^3-nitrogen	19	Si	silane
9	N	sp^2-nitrogen, amide	25	P	phosphine
10	N	sp-nitrogen	60	P	phosphor, pentavalent
37	N	azo or pyridine (–N=)	51	He	helium
39	N^+	sp^3-nitrogen, ammonium	52	Ne	neon
40	N	sp^2-nitrogen, pyrrole	53	Ar	argon
43	N	azoxy (–N=N–O)	54	Kr	krypton
45	N	azide, central atom	55	Xe	xenon
46	N	nitro (–NO_2)	31	Ge	germanium
72	N	imine, oxime (=N–)	32	Sn	tin
6	O	sp^3-oxygen	33	Pb	lead
7	O	sp^2-oxygen, carbonyl	34	Se	selenium
41	O	sp^2-oxygen, furan	35	Te	tellurium
47	O^-	carboxylate	59	Mg	magnesium
49	O	epoxy	61	Fe	iron (II)
69	O	amine oxide	62	Fe	iron (III)
70	O	ketonium oxygen	63	Ni	nickel (II)
5	H	hydrogen, except on N or O	64	Ni	nickel (III)
21	H	alcohol (OH)	65	Co	cobalt (II)
23	H	amine (NH)	66	Co	cobalt (III)
24	H	carboxyl (COOH)			

MM2 example: number of parameters (Jensen)

- Each of the 71 atom types has two van der Waals parameters, R_0^A and ε^A, giving 142 parameters.
- There are approximately $\frac{1}{2} \times 30 \times 30 = 450$ possible different E_{str} terms, each requiring at least two parameters, k^{AB} and R_0^{AB}, for a total of at least 900 parameters.
- There are approximately $\frac{1}{2} \times 30 \times 30 \times 30 = 13\,500$ possible different E_{bend} terms, each requiring at least two parameters, k^{ABC} and θ_0^{ABC}, for a total of at least 27\,000 parameters.
- There are approximately $\frac{1}{2} \times 30 \times 30 \times 30 \times 30 = 405\,000$ possible different E_{tors} terms, each requiring at least three parameters, V_1^{ABCD}, V_2^{ABCD} and V_3^{ABCD}, for a total of at least 1\,215\,000 parameters.
- Cross terms may add another million possible parameters.

Table 2.3 Comparison of possible and actual number of MM2(91) parameters

Term	Estimated number of parameters	Actual number of parameters
E_{vdw}	142	142
E_{str}	900	290
E_{bend}	27\,000	824
E_{tors}	1\,215\,000	2466

Reducing number of fitting parameters

- Assign atomic charges (e.g. from DFT) instead of fitting them
- Use element-wise LJ parameterization instead of pair-wise
- Group atom types into atom classes for bonded interactions
- Omit unnecessary dihedrals and minimize number of impropers
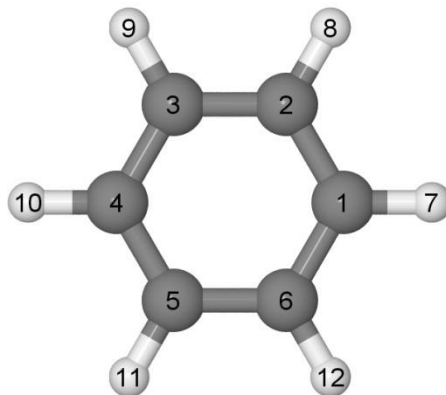
Parameterization is the bottleneck

How to do it? Define functional form, objective function, and fit parameters:

$$\text{ErrF}(\text{parameters}) = \sum_i^{\text{data}} weight_i \cdot (reference\ value - calculated\ value)_i^2$$

1) How to choose the functional form:

- The choice is not unique and depends on material, scales, and goals.
- Linear optimization form is preferable.
- Systematic improvement of accuracy is challenging for non-ML potentials.

2) What are reference values:

- Experimental data (vibrational frequencies, X-ray structures, heats of formation) were commonly used in past, but they are often incomplete and have large error bars (especially lab-to-lab and batch-to-batch variations).
- Today high-throughput DFT calculations are used, including semi-automated parameterization routines (e.g. AMBER-DFT or ML potentials). However, some observables cannot be directly calculated by DFT (e.g. melting temperature).

3) How to define and generate fitting (training) dataset:

- The dataset must be representative (cover required part of PES with proper weights) and suitable for fitting (confidence intervals for all parameters must be reasonable, parameter dependencies must be resolved).
- It can be generated at once, iteratively, in parts (e.g. separate set for dihedrals and intermolecular interactions which have complex PES and different energy scale), or "on fly" by active learning.

Overall, parameterization is a difficult process requiring good knowledge of underlying physical phenomena/interactions, chemical intuition, use of experimental data and utilization of electronic structure codes.

Various types of interatomic potentials

Different classes of materials require different approaches:

- Simple metals (Al - yes, Pu - no) – embedded-atom model (EAM)
- Ionic solids (NaCl) – embedded-ion model (EIM)
- Complex bonding – charge optimized many body potential (COMB)
- Tetrahedral semiconductors/insulators (Si, SiO_2) – Tersoff potentials
- Water – TIP3/4/5P models
- π-conjugated molecules – MM3 force field + Huckel model
- Biomolecules – well parameterized AMBER, CHARMM, GROMOS with residue-based parameterization (proteins) and united atoms
- Attempts to make universal force field for molecules – OPLS, UFF, DREIDING
- Polarizable force fields – DRUDE, AMOEBA
- Chemical reactions – reactive force field (ReaxFF)

Different states of the same molecule require different values parameters:

Anion, cation, excited singlet S_1 etc.

'Fitting-friendly' parameterization (allows for black-box fitting and systematic improvement by ML approaches): $E=\sum_i E(\text{local geometry of i-th atom})+E_{\text{long-range}}$

Computational efficiency

- Make no mistake: molecular mechanics is MUCH faster compared to any electronic structure model based on quantum mechanics!

- Evaluation of the non-bonded energy is the most time-consuming step growing as ~N^2. Can be reduced to ~N using cut-off distances or advanced summation techniques (e.g. Fast Multipole method).

Table 2.6 Number of terms for each energy contribution in $CH_3(CH_2)_{n-2}CH_3$ *(from Jensen)*

n	N_{atoms}	E_{str}	E_{bend}	E_{tors}	E_{vdw}
10	32	31 (5%)	30 (10%)	81 (14%)	405 (70%)
20	62	61 (3%)	60 (6%)	171 (8%)	1710 (83%)
50	152	151 (1%)	300 (3%)	441 (4%)	11025 (93%)
100	302	301 (1%)	600 (1%)	891 (2%)	44550 (96%)
	N	$(N-1)$	$2(N-2)$	$3(N-5)$	$\frac{1}{2}N(N-1)-3N+5$

Benchmarks by LAMMPS

Potential	System	# Atoms	Timestep	Neighs/atom	Memory	CPU	LJ Ratio
Granular	chute flow	32000	0.0001 tau	7.2	33 Mb	2.08e-7	0.26x
FENE bead/spring	polymer melt	32000	0.012 tau	9.7	8.4 Mb	2.86e-7	0.36x
Lennard-Jones	LJ liquid	32000	0.005 tau	76.9	12 Mb	8.01e-7	1.0x
DPD	pure solvent	32000	0.04 tau	41.3	9.4 Mb	1.22e-6	1.53x
EAM	bulk Cu	32000	5 fmsec	75.5	13 Mb	1.87e-6	2.34x
REBO	polyethylene	32640	0.5 fmsec	149	33 Mb	3.18e-6	3.97x
Stillinger-Weber	bulk Si	32000	1 fmsec	30.0	11 Mb	3.28e-6	4.10x
Tersoff	bulk Si	32000	1 fmsec	16.6	9.2 Mb	3.74e-6	4.67x
ADP	bulk Ni	32000	5 fmsec	83.6	25 Mb	5.58e-6	6.97x
EIM	crystalline NaCl	32000	0.5 fmsec	98.9	14 Mb	5.60e-6	6.99x
Peridynamics	glass fracture	32000	22.2 nsec	422	144 Mb	7.46e-6	9.31x
SPC/E	liquid water	36000	2 fmsec	700	86 Mb	8.77e-6	11.0x
CHARMM + PPPM	solvated protein	32000	2 fmsec	376	124 Mb	1.13e-5	14.1x
MEAM	bulk Ni	32000	5 fmsec	48.8	54 Mb	1.32e-5	16.5x
Gay-Berne	ellipsoid mixture	32768	0.002 tau	140	21 Mb	2.20e-5	27.5x
BOP	bulk CdTe	32000	1 fmsec	4.4	74 Mb	2.51e-5	31.3x
AIREBO	polyethylene	32640	0.5 fmsec	681	101 Mb	3.25e-5	40.6x
ReaxFF/C	PETN crystal	32480	0.1 fmsec	667	976 Mb	1.09e-4	136x
COMB	crystalline SiO2	32400	0.2 fmsec	572	85 Mb	2.00e-4	250x
eFF	H plasma	32000	0.001 fmsec	5066	365 Mb	2.16e-4	270x
ReaxFF	PETN crystal	16240	0.1 fmsec	667	425 Mb	2.84e-4	354x
VASP/small	water	192/512	0.3 fmsec	N/A	320 procs	26.2	17.7e6
VASP/medium	CO2	192/1024	0.8 fmsec	N/A	384 procs	252	170e6
VASP/large	Xe	432/3456	2.0 fmsec	N/A	384 procs	1344	908e6

Other thoughts and facts

- Blind use of empirical potentials is very dangerous and can be in error.

- Well-parameterized potentials can give more accurate geometries and relative energies than low-level QM methods.

- Often FF are bound to specific MD code (AMBER, CHARMM etc) so that parameters are not easily transferable between codes.

Typical workflow

1) Choose the functional form of empirical potentials:

- define atom types;

- decide on type of potentials and number of parameters
  (e.g. not all dihedrals are usually needed, parameterization of LJ
  interactions can be chosen element-wise or pair-wise);

- which parameters are fitted and which are assigned
  (e.g. often it is bad idea to fit charges in force fields, for intramolecular
  dynamics LJ parameters can be assigned from generic tables).

2) Force fields require topology – define it.

3) Provide initial parameterization

(from generic force field like OPLS or from DFT calculations).

4) Parameterize.

Molecular topology

Force fields require topology: atomic types and connectivity

- Normally you fix topology for a given system

- There is no commonly accepted format for storing topology

- Protein Databank (PDB) format store topology information

- XYZ format can be used for storing topology
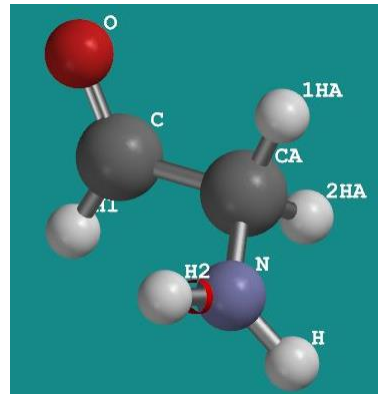
The Protein Data Bank (PDB) format

Provides a standard representation for macromolecular structure data derived from X-ray diffraction and NMR studies. Broadly used in computational and visualization software (e.g. NAMD and VMD: http://www.ks.uiuc.edu/research/namd)

Example from Wikipedia

```
HEADER    EXTRACELLULAR MATRIX                    22-JAN-98   1A3I
TITLE     X-RAY CRYSTALLOGRAPHIC DETERMINATION OF A COLLAGEN-LIKE
TITLE    2 PEPTIDE WITH THE REPEATING SEQUENCE (PRO-PRO-GLY)
...
EXPDTA    X-RAY DIFFRACTION
AUTHOR    R.Z.KRAMER,L.VITAGLIANO,J.BELLA,R.BERISIO,L.MAZZARELLA,
AUTHOR   2 B.BRODSKY,A.ZAGARI,H.M.BERMAN
...
REMARK 350 BIOMOLECULE: 1
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 350   BIOMT1   1  1.000000  0.000000  0.000000        0.00000
REMARK 350   BIOMT2   1  0.000000  1.000000  0.000000        0.00000
...
SEQRES   1 A    9  PRO PRO GLY PRO PRO GLY PRO PRO GLY
SEQRES   1 B    6  PRO PRO GLY PRO PRO GLY
SEQRES   1 C    6  PRO PRO GLY PRO PRO GLY
...
ATOM      1  N   PRO A   1       8.316  21.206  21.530  1.00 17.44           N
ATOM      2  CA  PRO A   1       7.608  20.729  20.336  1.00 17.44           C
ATOM      3  C   PRO A   1       8.487  20.707  19.092  1.00 17.44           C
ATOM      4  O   PRO A   1       9.466  21.457  19.005  1.00 17.44           O
ATOM      5  CB  PRO A   1       6.460  21.723  20.211  1.00 22.26           C
...
HETATM  130  C   ACY   401       3.682  22.541  11.236  1.00 21.19           C
HETATM  131  O   ACY   401       2.807  23.097  10.553  1.00 21.19           O
HETATM  132  OXT ACY   401       4.306  23.101  12.291  1.00 21.19           O
...
```

TXYZ-format: MM3 force field in Tinker

(http://zhugayevych.me/soft/tinker/test1.inp)

```
16   Styrene -------- coordinates ----------    type     --- bonds ---
  1  C       0.000000     0.000000     0.000000    2     2     6     9
  2  C       0.000000     0.000000     1.500000    2     1     3    10
  3  C       1.299038     0.000000     2.250000    2     2     4    11
  4  C       2.598076     0.000000     1.500000    2     3     5    12
  5  C       2.598076     0.000000     0.000000    2     4     6    13
  6  C       1.299038     0.000000    -0.750000    2     1     5     7
  7  C       1.299038     0.000000    -2.250000    2     6     8    14
  8  C       2.598076     0.000000    -3.000000    2     7    15    16
  9  H      -0.952628     0.000000    -0.550000    5     1
 10  H      -0.952628     0.000000     2.050000    5     2
 11  H       1.299038     0.000000     3.350000    5     3
 12  H       3.550704     0.000000     2.050000    5     4
 13  H       3.550704     0.000000    -0.550000    5     5
 14  H       0.346410     0.000000    -2.800000    5     7
 15  H       3.550704     0.000000    -2.450000    5     8
 16  H       2.598076     0.000000    -4.100000    5     8
```
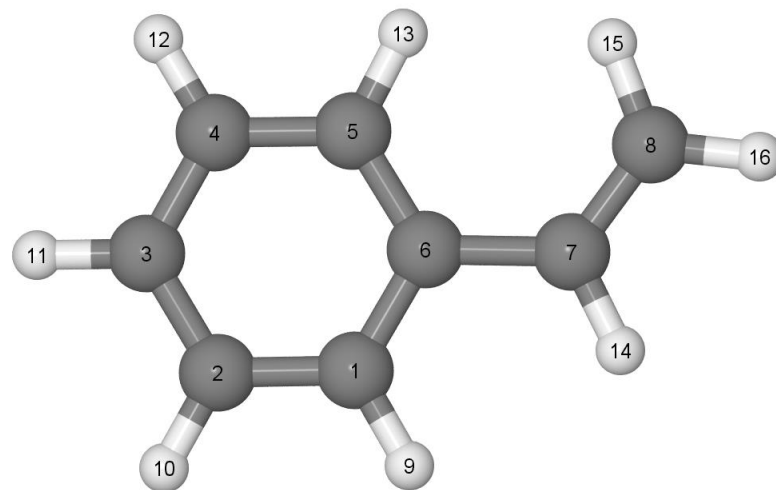
Type assignment is the most nontrivial
part of preparing input files
   → use special software

XYZ-format for storing topology

(http://zhugayevych.me/soft/template_topo.xyz)

```
51 -------- coordinates ------------------------------- type --- bonds ---
P3HT polymer optimized by CAM-B3LYPp2p + atom types and connectivity
   S         1.89846705      -1.00721949       0.00000000     `7      [2,3]
   C         0.69873877       0.25898281       0.00000000     `3      [1,4,28]
   C         3.21473987       0.13086628       0.00000000     `3      [1,5,27]
   C         1.29764141       1.50338361       0.00000000     `9      [2,5,7]
   C         2.71418575       1.40870908       0.00000000     `5      [3,4,6]
   H         3.35659890       2.27994179       0.00000000     `6      [5]
   C         0.55527024       2.81771042       0.00000000     `11     [4,8,9,10]
   C         1.43502545       4.06804021       0.00000000     `14     [7,11,12,13]
.......................................................................
   H         6.07735555      -9.19821743      -0.88348350     `15     [45]
   H         4.81961625     -10.07258964       0.00000000     `15     [45]
   H         6.07735555      -9.19821743       0.88348350     `15     [45]
   Tv        7.82695786       0.00000000       0.00000000
```

# From topology to force field
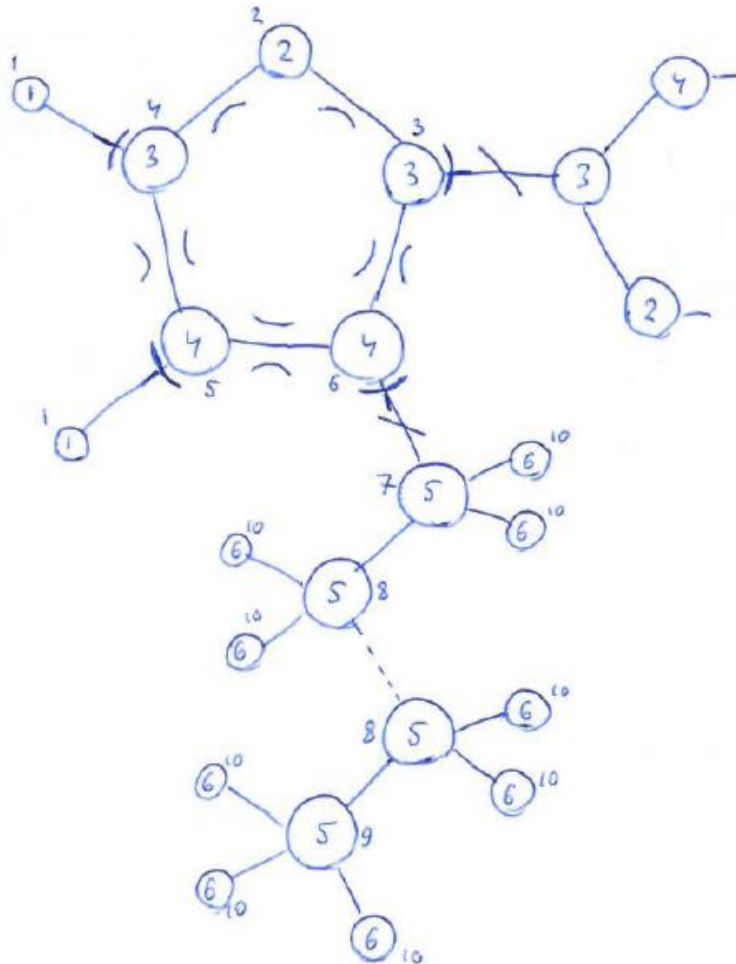
Using generic force fields

- Most likely you will need to add some parameters

See e.g. modified MM3 at http://zhugayevych.me/soft/tinker/_prm/mm3.prm


```
# added by Andriy Zhugayevych

atom         165    I-      "IODIDE (ANION)"              53    126.900     0

vdw          165                2.2800      0.4950    # as for Xe

charge       165                -1.0000

angle         37    2    11     0.810     115.50      119.00        0.00    # as for 2-2-11
angle          2    2    39     0.600     122.00      125.00        0.00    # as for 2-2-8
angle          1   39     2     0.719     107.114     110.00      112.00    # as for 1-39-1
angle          2   39    48     0.521     105.949     107.425     108.524   # as for 1-39-48
angle5         2   19     2     0.600     104.50                            # =angle

torsion        2    3     6    24     0.000 0.0 1    5.390 180.0 2    1.230 0.0 3    # as for 1-3-6-24
torsion        2   37     2    11     0.000 0.0 1   14.500 180.0 2    0.000 0.0 3    # as for 2-2-2-11
torsion        1    1     2    42    -0.700 0.0 1   -0.200 180.0 2   -0.550 0.0 3    # as for 1-1-2-2
torsion        2    2    39     1     0.000 0.0 1   10.000 180.0 2    0.000 0.0 3    # as for 5-2-39-1
torsion        2    2    39    48     0.000 0.0 1   10.000 180.0 2    0.000 0.0 3    # as for 5-2-39-1
torsion       37    2     2    42     1.250 0.0 1    8.500 180.0 2    2.250 0.0 3    # average of 37-2-2-37 & 42-2-2-42
torsion        2    2    37    42     0.000 0.0 1   10.000 180.0 2    0.000 0.0 3    # constrained to 0 or 180 deg
torsion       19    2     2    42     0.000 0.0 1   10.000 180.0 2    0.000 0.0 3    # constrained to 0 or 180 deg
torsion5       2    2     2    19     0.000 0.0 1   10.000 180.0 2    0.000 0.0 3    # constrained to 0 or 180 deg
torsion5       2    2    19     2     0.000 0.0 1   10.000 180.0 2    0.000 0.0 3    # constrained to 0 or 180 deg

opbend        39    2     0     0                0.100    # as for 8-2-0-0
```
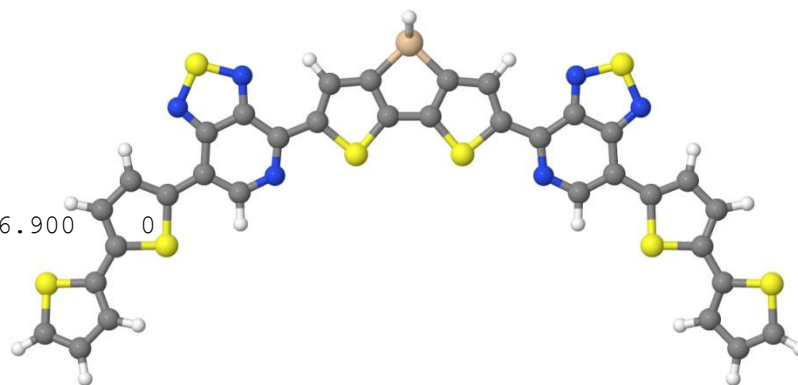
LAMMPS-friendly OPLS force field file format
(http://zhugayevych.me/soft/prm.htm)

```
# comment line

any LAMMPS input

special_bonds lj 0 0 1 coul 0 0 1 (1 1 for Dreiding, 1/2 5/6 for AMBER, 0 0 for CHARMM)

      type  mass     class symbol coordination   charge     description
mass  1   1.008    #  1     H          1          0.1475    HC-alpha/beta THIOPHENE
mass  2   31.972   #  2     S          2          0.0090    S polyTHIO centr-HR
...

pair_style lj/cut/coul/cut 10 10  (refine parameters and consider other coul styles)

            i ≤ j  epsilon   sigma (here ij are types)
pair_coeff  1   1   0.0300   2.4200
pair_coeff  2   2   0.2500   3.5500
...

bond_style harmonic

            id      K      bondlength   i ≤ j (from now on ijkl are classes)
bond_coeff   1    367.0    1.082    #  1    3
bond_coeff   2    367.0    1.082    #  1    4
...

angle_style harmonic
(it is angle between ji and jk, so that j is central atom)
            id    K     angle       i    j    k (i≤k)
angle_coeff   1   60.0   91.00    #  3    2    3
angle_coeff   2   60.0  119.00    #  1    3    2
...

dihedral_style opls
(dihedral is angle between ijk and jkl planes, so that jk is central bond)
              id    K1      K2       K3       K4      i   j ≤ k   l
dihedral_coeff   1   0.000   7.250   0.000   0.000   #  3   2    3    1
dihedral_coeff   2   0.000   7.250   0.000   0.000   #  3   2    3    3
...

improper_style harmonic
(improper is angle between ijk and jkl planes, i is central atom bonded to j,k,l)
              id     K  angle    i   j ≤ k   l
improper_coeff   1    2.2   0   #  3   2    4    1
improper_coeff   2    2.2   0   #  3   2    4    3
...
```
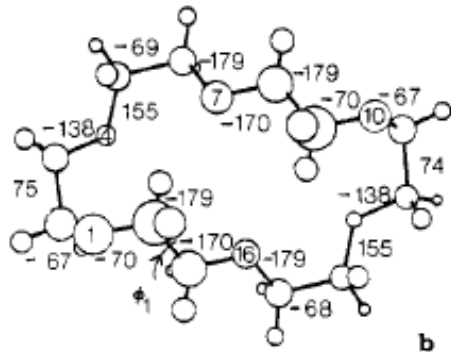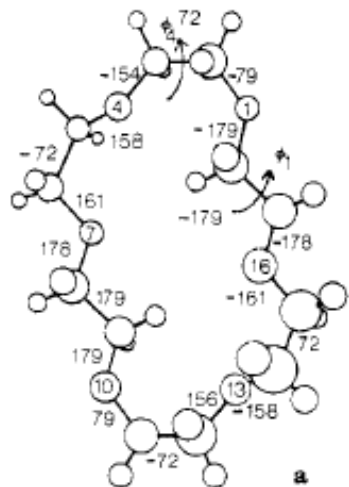
Example 1: Historical

A Molecular Mechanics Study of 18-Crown-6 and Its
Alkali Complexes: An Analysis of Structural Flexibility,
Ligand Specificity, and the Macrocyclic Effect
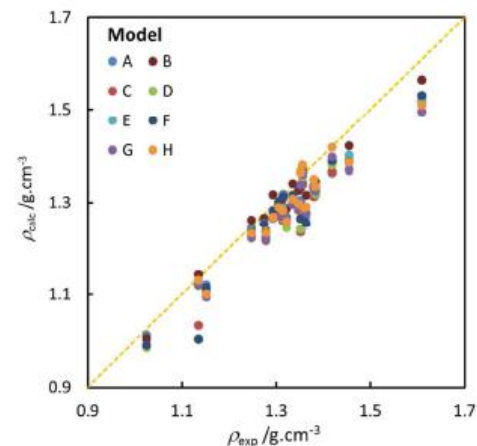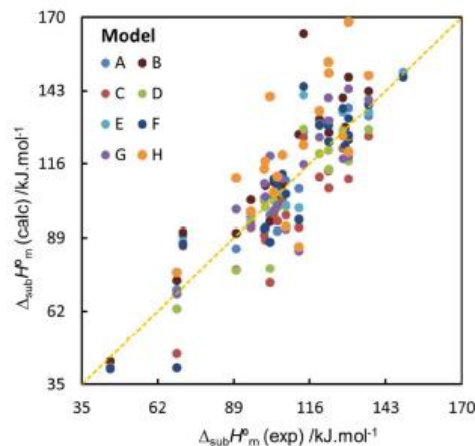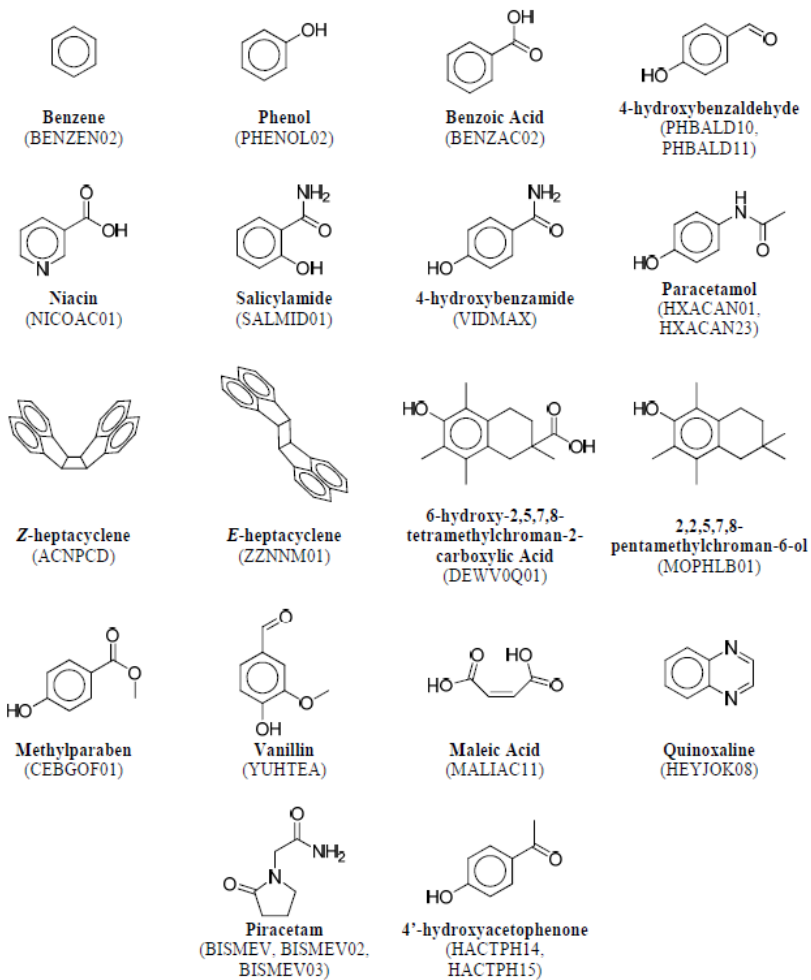
Georges Wipff, Paul Weiner, and Peter Kollman*
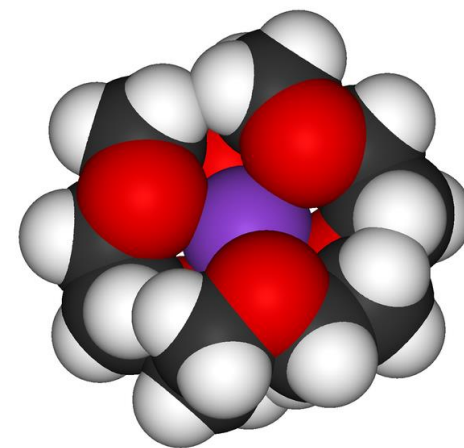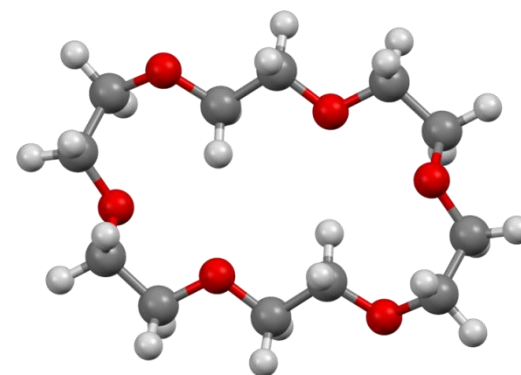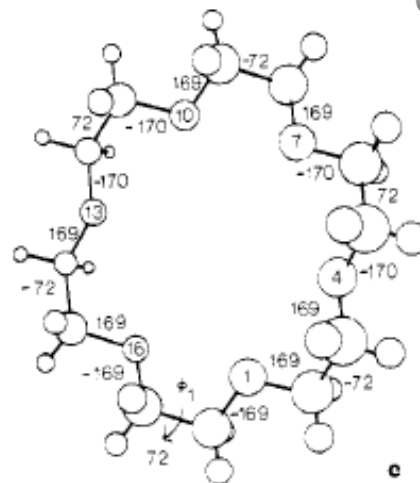
Synthesis crown ethers (cyclic polyethers) lead to the Nobel Prize in Chemistry in 1987 (Charles Pedersen, Donald Cram and Jean-Marie Lehn). The donut-shaped molecules were the first in a series of extraordinary compounds that form stable structures with alkali metal ions (used as phase transfer catalysts).

31

Example 2: Accuracy

Benzene
(BENZEN02)

Phenol
(PHENOL02)

Benzoic Acid
(BENZAC02)

4-hydroxybenzaldehyde
(PHBALD10,
PHBALD11)

Niacin
(NICOAC01)

Salicylamide
(SALMID01)

4-hydroxybenzamide
(VIDMAX)

Paracetamol
(HXACAN01,
HXACAN23)

Z-heptacyclene
(ACNPCD)

E-heptacyclene
(ZZNNM01)

6-hydroxy-2,5,7,8-
tetramethylchroman-2-
carboxylic Acid
(DEWV0Q01)

2,2,5,7,8-
pentamethylchroman-6-ol
(MOPHLB01)

Methylparaben
(CEBGOF01)

Vanillin
(YUHTEA)

Maleic Acid
(MALIAC11)

Quinoxaline
(HEYJOK08)

Piracetam
(BISMEV, BISMEV02,
BISMEV03)

4'-hydroxyacetophenone
(HACTPH14,
HACTPH15)

Conclusions:

- OPLS-AA force field without any modification, can be a good starting point for the study of the main features of crystal structures, but an assessment is always recommended.

- Use charges from quantum chemistry.

- Polarizable modifications of OPLS-AA are more accurate.

32

Example 3: Polymorphism

Explain bulk structure of

	X	Y
1	N	CH
2	CH	N
3	CF	CH
4	CH	CF
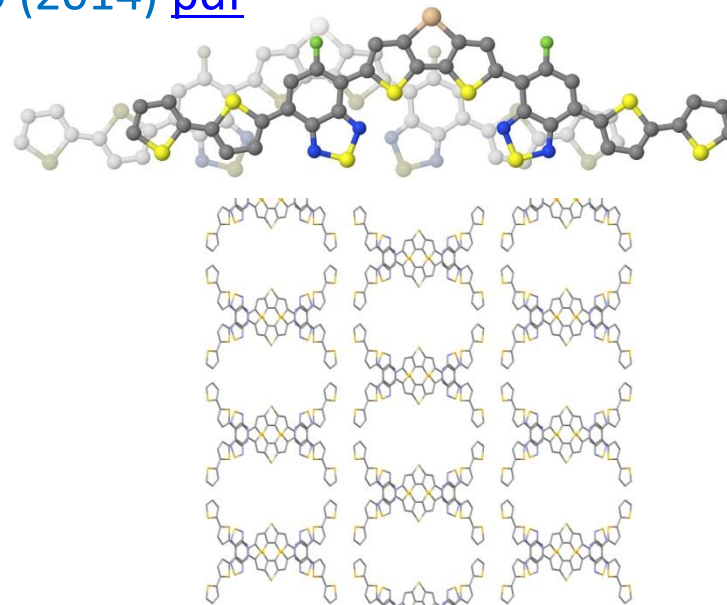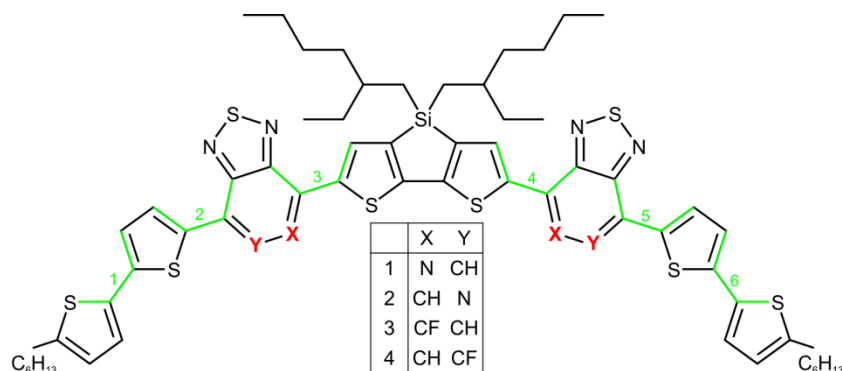
Table 2. Summary of Computed Properties of Crystals Optimized by MM3 (Values Based on Observed Structures in Parentheses)

entry	conformation energy (eV)	binding energy (eV)	intrastack binding energy (eV)	interstack binding energy (eV)	exciton coupling (meV)	hole coupling (meV)
1a	0	3.39	2.04	0.45	65 (63)	53
1b	0.16	3.10	1.88	0.41	69	121
2a	0	3.28	2.06	0.41	45	40
2b	0.04	3.23	1.86	0.46	51	126
3a	0	2.99	2.08	0.30	57	65
3b	0.06	3.34	2.15	0.40	43 (30)	127
4a	0	3.14	2.06	0.36	65	44
4a'[a]	0	3.16	2.07	0.36	102 (113)	100
4a"[a]	0.02	3.18	2.02	0.39	80 (88)	75
4b	0.03	3.31	2.03	0.43	56 (46)	105

[a]4a' and 4a" represent experimentally determined crystal structures for **4** that adopt type *a* geometry, and prime and double prime accommodate two observed disorder contributions. Binding energies are reported per-molecule.
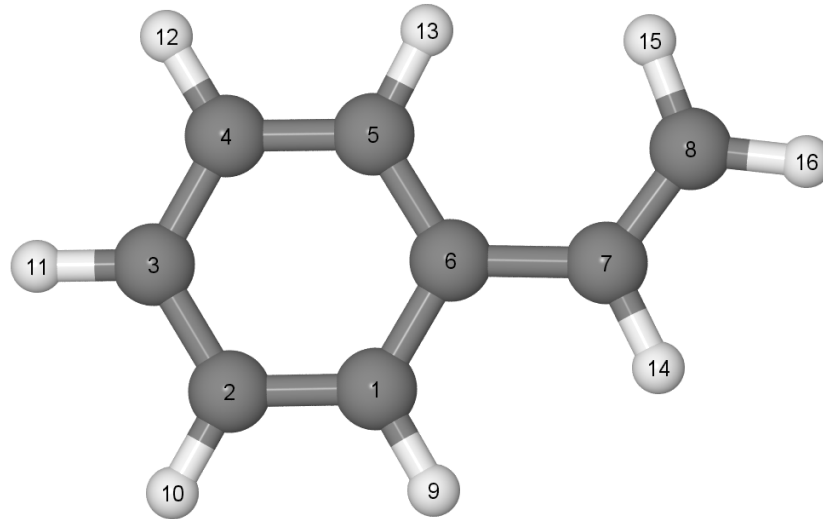
Example 4: Complex polymorphism

J Phys Chem C 122, 9141 (2018) pdf

Use simple force field to prescreen possible low-energy polymorphs
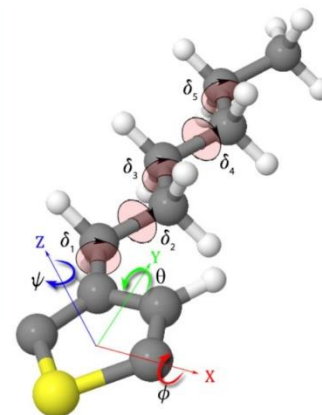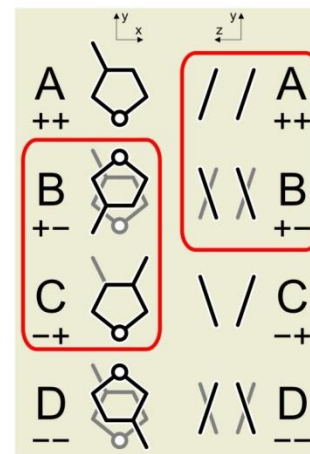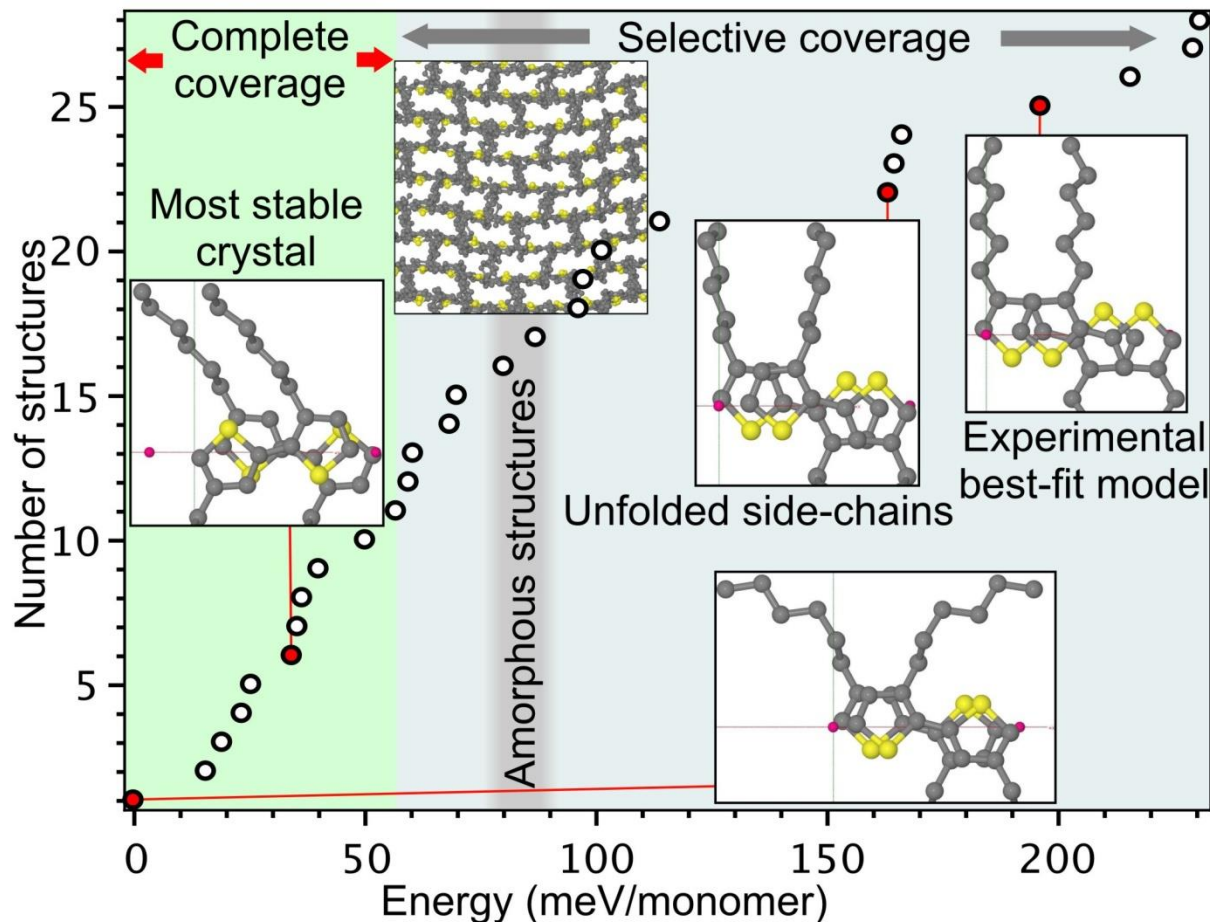
Discussion

1. List all elements of a force field for this molecule:
   e.g. *??? atom types; ??? bond types: 6 aromatic CC, …*

2. What problems do we expect for parameterization of angle bendings and impropers at atom #1, and dihedrals at bond 1-2?

3. Why EIM potential a priori should be slower than EAM?

Molecular dynamics

- Solve Newtonian equation of motion for N classical particles (3N coupled equations)
- For now, let limit ourselves by natural NVE ensemble
- The force depends on positions only (not velocities)
- Each particle is allowed to interact simultaneously with every other particle and can experience an additional external potential
- A single point in a 6N-dimensional phase space (*p*,*r*) represents our dynamical system

$$\mathbf{F} = m\mathbf{a}$$

$$E_{\text{tot}} = \sum_{i=1}^{N} \tfrac{1}{2} m_i \mathbf{v}_i^2 + V(\mathbf{r})$$

$$-\frac{dV}{d\mathbf{r}} = m\frac{d^2\mathbf{r}}{dt^2}$$

$$\langle A \rangle = \lim_{t \to \infty} \frac{1}{t} \int_{t_0}^{t_0+t} A(\tau)d\tau$$

$$\langle A \rangle = \frac{1}{M} \sum_{i}^{M} A(t_i)$$

Our dimensionality: N particles, 3N-dim vectors

- Coordinates $r = q = (x_1, y_1, z_1, x_2, y_2, z_2, ...)$
- Velocities $v = dr/dt$
- Momenta $p = mv$
- Accelerations $a = dv/dt$
- Potential energy $V(r)$
- Kinetic energy $K(\mathbf{p}) = \sum m_i \mathbf{v}_i^2 / 2 = \sum \mathbf{p}_i^2 / 2m_i$
- Forces $\mathbf{F}(\mathbf{r}) = dV / d\mathbf{r} = \nabla V = (dE_p / dx_1, ...)$
- Particle masses m_i
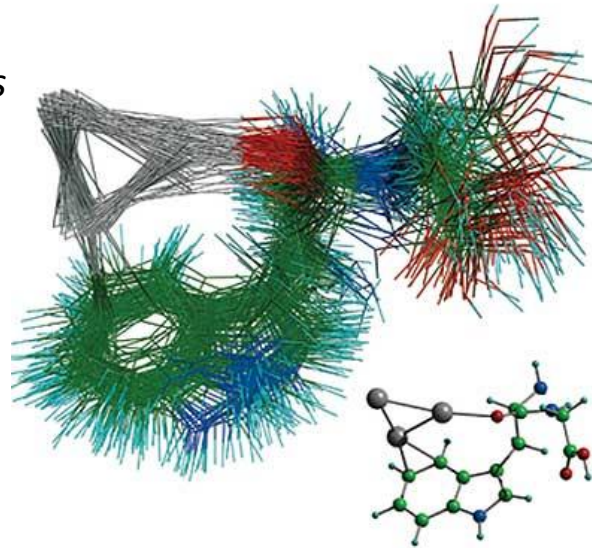
Ensemble of possible structures of a tryptophan-glycine–silver cluster cation complex, obtained by superimposing the configurations sampled from MD trajectory simulation.
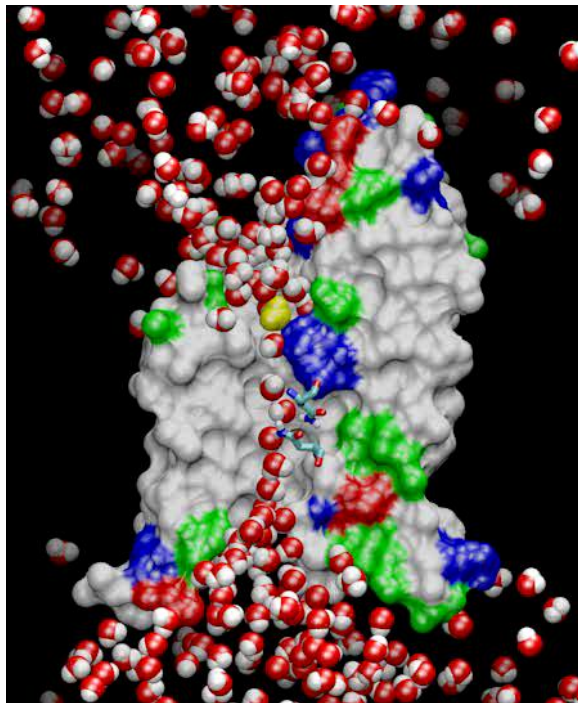Phys Rev Lett 101, 213001 (2008)

Molecular dynamics: a typical algorithm

- *Initialize:* select positions and velocities

- *Propagate:* compute all forces, and then determine new positions

- *Equilibrate:* let the system reach equilibrium for a given thermodynamic ensemble (e.g. *NVE*) and 'forget' about initial conditions

- *Sample (average):* accumulate long enough trajectory and calculate quantities of interest

Every second about a billion water molecules (red and white spheres) pass through a channel formed in the middle of an aquaporin protein as shown by white (nonpolar) and green (polar) areas, and charged areas in blue (positive) and red (negative). The yellow sphere highlights the path of a single water molecule.
Courtesy of Beckman Institute for Advanced Science and Technology
http://www.ks.uiuc.edu/

Algorithm:

1. Give particles initial positions $r_0=r(t=0)$, velocities $v_0=v(t=0)$. Calculate and store energy $E_0=E(t=0)$ and other quantities at $t=0$. Choose short time-step Δt (typical ~0.1-1fs)

2. Get forces $F(t)$ and accelerations $a(t)$ (see *)

3. Move particles, i.e. compute $r(t+\Delta t)$ and $v(t+\Delta t)$

4. Move time forward $t=t+\Delta t$

5. Calculate and store energy $E(t)$ and other quantities at t.

6. Repeat as long as you need

*Note: Propagator (or integrator) steps 2-4 frequently require some stored vectors r, v, a, F at $t-\Delta t$ and t to calculate the next set of variables at $t+\Delta t$

Molecular dynamics: initialization

Selecting initial positions:

- Interconnection topology should agree with chemical structure
- Avoid short distances – huge energy penalty ($\sim 1/r^{12}$ for Lenard-Jones)
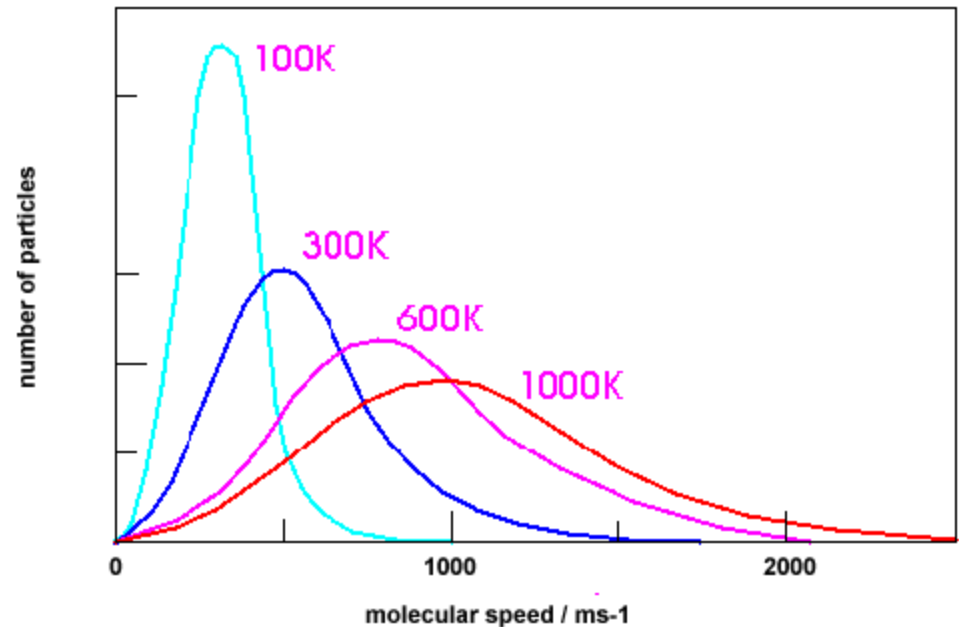- Avoid highly non-equilibrium conditions

Selecting initial velocities:

- Start with v=0, then allow to equilibrate/thermalize with an increase of temperature OR
- Start with some distribution (e.g. Maxwell-Boltzmann distribution)

$$f(\mathrm{v}) = \sqrt{\left(\frac{m}{2\pi kT}\right)^3 4\pi \mathrm{v}^2 \exp\left(-\frac{m\mathrm{v}^2}{2kT}\right)}$$

$$\frac{dN}{N} = f(\mathrm{v})dv \qquad \mathrm{v}_p = \sqrt{\frac{2kT}{m}}$$

$$\langle \mathrm{v} \rangle = \sqrt{\frac{8kT}{\pi m}} \qquad \mathrm{v}_{rms} = \sqrt{\frac{3kT}{m}}$$

Question: Is it "legal" to use Maxwell-Boltzmann distribution for interacting particles?

Molecular dynamics: propagation

Verlet type integrators (typical accuracy up to $O(t^4)$): Most common in the MD land!
Simple to calculate, well preserves the energy along the trajectory (i.e. time-reversible)

As usual, start with Taylor expansion $\mathbf{r}_{i+1} = \mathbf{r}_i + \frac{\partial \mathbf{r}}{\partial t}(\Delta t) + \frac{1}{2}\frac{\partial^2 \mathbf{r}}{\partial t^2}(\Delta t)^2 + \frac{1}{6}\frac{\partial^3 \mathbf{r}}{\partial t^3}(\Delta t)^3 + \ldots$

Add $\mathbf{r}_{i+1} = \mathbf{r}_i + \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 + \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \ldots$ *and* $\mathbf{r}_{i-1} = \mathbf{r}_i - \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 - \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \ldots$

The original Verlet

$$\mathbf{r}_{i+1} = (2\mathbf{r}_i - \mathbf{r}_{i-1}) + \mathbf{a}_i(\Delta t)^2 + \ldots$$

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i} = -\frac{1}{m_i}\frac{dV}{d\mathbf{r}_i}$$

To initiate $\mathbf{r}_{-1} = \mathbf{r}_0 - \mathbf{v}_0\Delta t$

Note velocities ($\mathbf{v}(t)=[\mathbf{r}(t+\Delta t)-\mathbf{r}(t-\Delta t)]/2\Delta t$) are not necessary but useful. Also ($2\mathbf{r}_i-\mathbf{r}_{i-1}$) is a large difference! Require 9N variables for storage, for i=1,…,N (compact!)

The leap-frog Verlet

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \mathbf{v}_{i+\frac{1}{2}}\Delta t$$

$$\mathbf{v}_{i+\frac{1}{2}} = \mathbf{v}_{i-\frac{1}{2}} + \mathbf{a}_i\Delta t$$

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i} = -\frac{1}{m_i}\frac{dV}{d\mathbf{r}_i}$$

To initiate $\mathbf{r}_{-1} = \mathbf{r}_0 - \mathbf{v}_0\Delta t$

Note velocities ($\mathbf{v}_{i+1/2}$)'leap' over coordinates half-step. Explicit velocities is a plus. But $\mathbf{r}$ and $\mathbf{v}$ are out of phase.

The Velocity Verlet

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \mathbf{v}_i\Delta t + \frac{1}{2}\mathbf{a}_i\Delta t^2$$

$$\mathbf{v}_{i+1} = \mathbf{v}_i + \frac{1}{2}\{\mathbf{a}_i + \mathbf{a}_{i+1}\}\Delta t$$

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i} = -\frac{1}{m_i}\frac{dV}{d\mathbf{r}_i}$$

To initiate $\mathbf{r}_{-1} = \mathbf{r}_0 - \mathbf{v}_0\Delta t$

Best numerical performance and compact storage make it method of choice for MD codes!

Runge–Kutta or Gear predictor-corrector propagators: rarely used in the MD simulations – numerically expensive, usually non time-reversible, and advantages of extrapolation are frequently lost

Molecular dynamics: thermostat

NVE (microcanonical) ensemble is natural: energy is conserved being redistributed along potential and kinetic parts along the trajectory

What about NVT (canonical) ensemble? Need thermostat (e.g. Anderson, Nosé–Hoover, Berendsen, Langevin etc.) allowing an exchange of the energy with a bath (e.g. solvent)

Langevin equation of motion
$$m\frac{d^2\mathbf{r}}{dt^2} = -\zeta\frac{d\mathbf{r}}{dt} + \mathbf{F}_{intra} + \mathbf{F}_{random}$$

Compared to the standard Newtonian equation, we have new terms:
1) $\zeta\mathbf{v}$ – friction term with a friction coefficient ζ (~$2ps^{-1}$), which removes energy from the system
2) $\mathbf{F}_{random}$ – random force associated with the temperature, which balances energy of the system

T=10K

T=300K
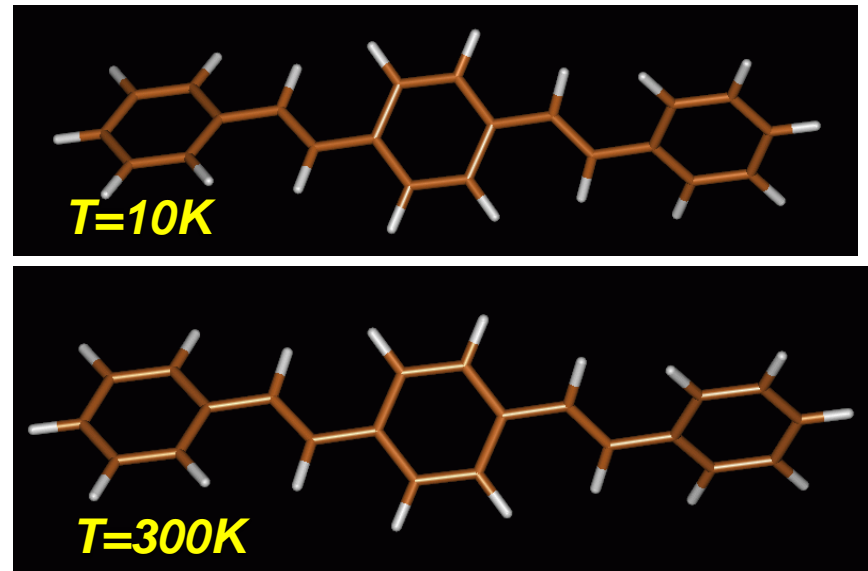
Implementation

Berendsen thermostat and barostat

J Chem Phys 81, 3684 (1984)

$$\sum_i \frac{|\boldsymbol{p}_i|^2}{m_i} = 3NkT, \qquad \sum_i \left(\frac{|\boldsymbol{p}_i|^2}{m_i} + \sum_{j>i} \boldsymbol{r}_{ij} \boldsymbol{F}_{ij} \right) = 3pV$$

Rescale velocities by λ and coordinates (or unit cell) by μ, where

$$\lambda^2 = 1 + \gamma \Delta t \left(\frac{T_0}{T} - 1 \right), \quad \mu^3 = 1 - \beta \Delta t \left(p_0 - p \right)$$

Advanced thermostat, barostat, and integrator see in program manuals

LAMMPS users see `fix nvt/npt/nph` command

Final notes on MD

Important scales: 1000 atoms - 2 nm, 10^6 - 20 nm, 10^9 - 200 nm

Classical MD (here 'classical' has double meaning):

- 10^9 particles at ps-scale OR   Laboratory times for smaller systems
- Metadynamics methods increase time-scale
- Performance of "Anton" supercomputer in 2010:

  20 mks/day for 24000 atoms (protein+water)
- Si by Tersoff potentials: 1 ns for 10^9 atoms [10.1002/jcc.26113]
- Limited by accuracy of empirical potentials and classical mechanics

Ab initio MD (adiabatic MD with explicit electrons):

- e**X**tended **L**agrangian approach: CPMD (**Car-Parrinello**) and then XLBOMD (**Born-Oppenheimer**)
- DFT level: 10^4 particles at ps-scale, use DFTB for larger scales
- Limited by accuracy of e-structure method and classical mechanics

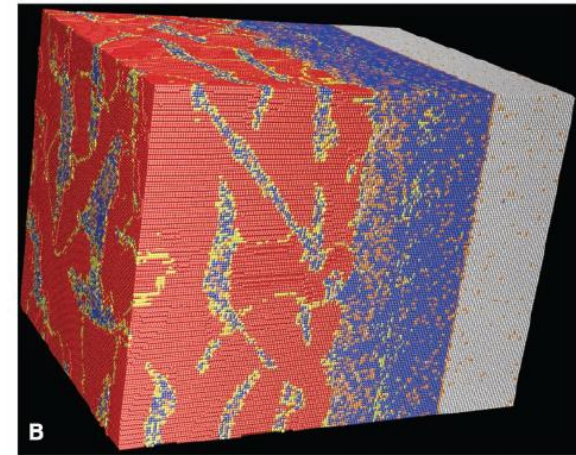**Next level is Non-Adiabatic MD (NAMD) and then Quantum MD**
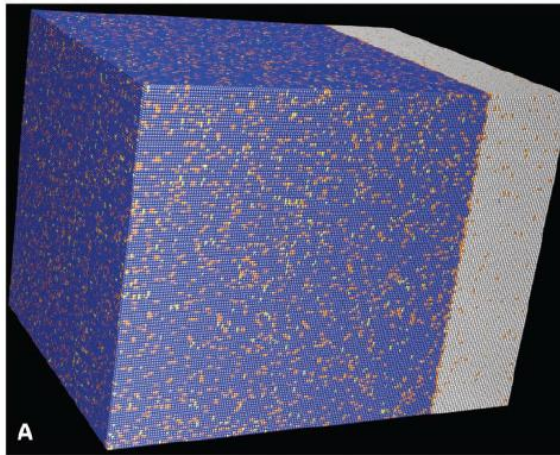
Example 5: Shocks in metals

Microscopic View of Structural Phase Transitions Induced by Shock Waves

EAM potentials, 10^7 atoms

Kai Kadau,[1,2]* Timothy C. Germann,[3] Peter S. Lomdahl,[1] Brad Lee Holian[1]

Fig. 1. Shocked samples (shock fronts propagate from left to right) after 8.76 ps for four different shock strengths in the bcc [001] direction as follows: piston velocities u_p are (**A**) 362 m/s (**B**) 471 m/s (**C**) 689 m/s. Atoms are color-coded by the number of neighbors n within 2.75 A. Gray, unshocked bcc (n=8); blue, uniaxially compressed bcc (n=10); and red, the transformed close-packed grains (n=12) separated by yellow (n=11) grain boundaries.
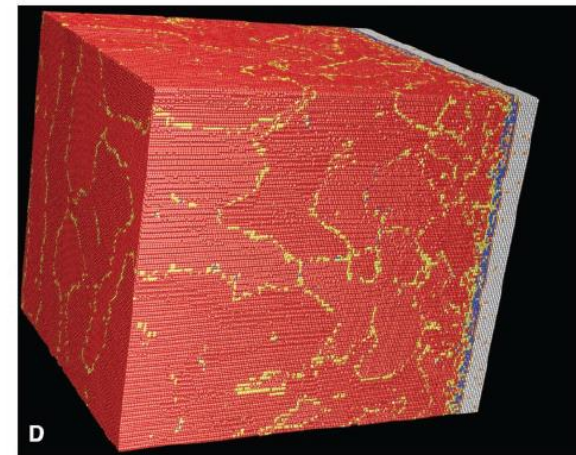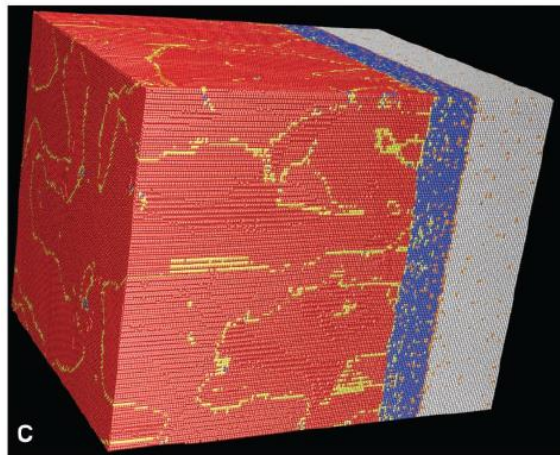
Example 5: Shocks in metals

Microscopic View of Structural Phase Transitions Induced by Shock Waves

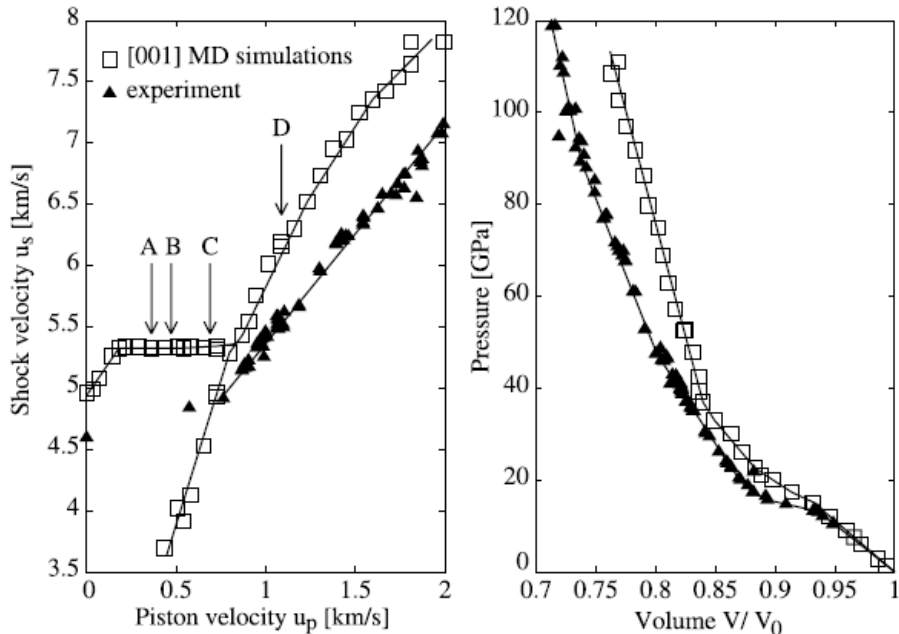Kai Kadau,[1,2]* Timothy C. Germann,[3] Peter S. Lomdahl,[1] Brad Lee Holian[1]

Fig. 3. Measured shock velocities u_s as a function of piston velocity u_p demonstrating the existence of split two wave-shock structure, due to a structural transition. Triangles, experimental polycrystal data; squares, perfect single crystal MD simulations in the bcc [001] direction.
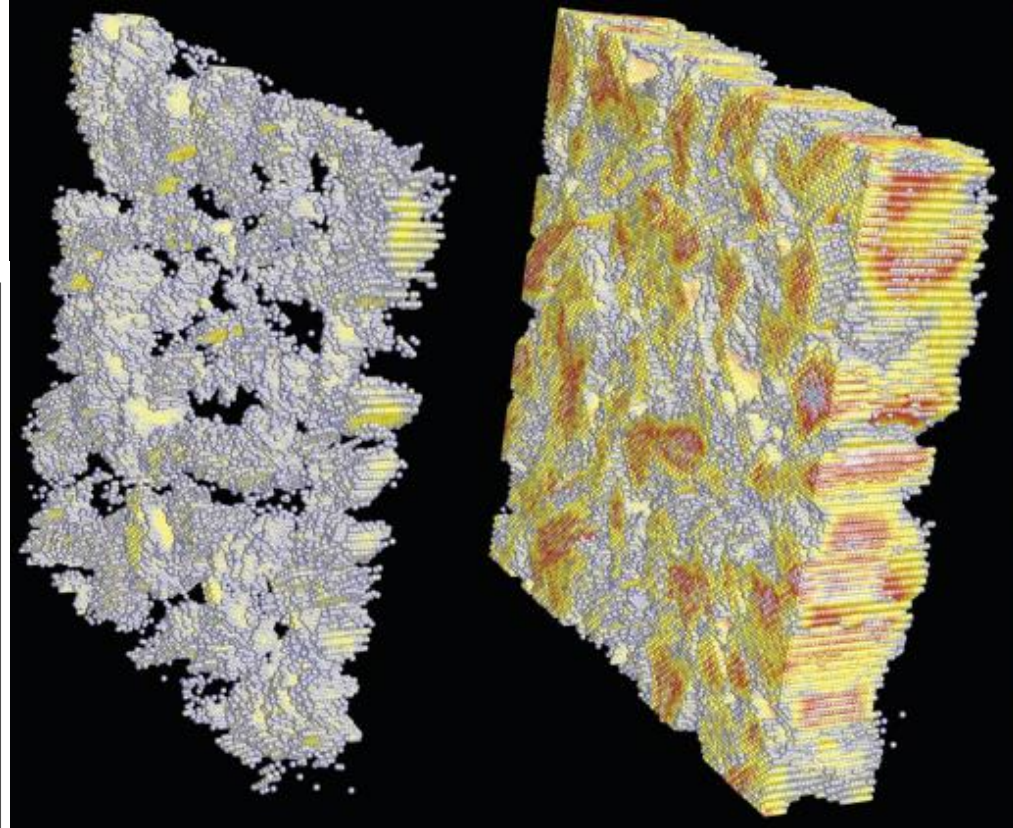

Fig. 2. Nucleation of close-packed material in the bcc matrix for a shock strength above the transformation threshold [u_p=471m/s, (Fig. 1B)]. Colored atoms show nucleation centers induced by statistical thermal fluctuations. After 1.095 ps (left), small nucleation centers build the transformation front (right, after 2.19 ps).

44

Extend MD time scale: coarse graining

Idea: simulate coarse grained model, then get back to all-atom description to refine local geometry

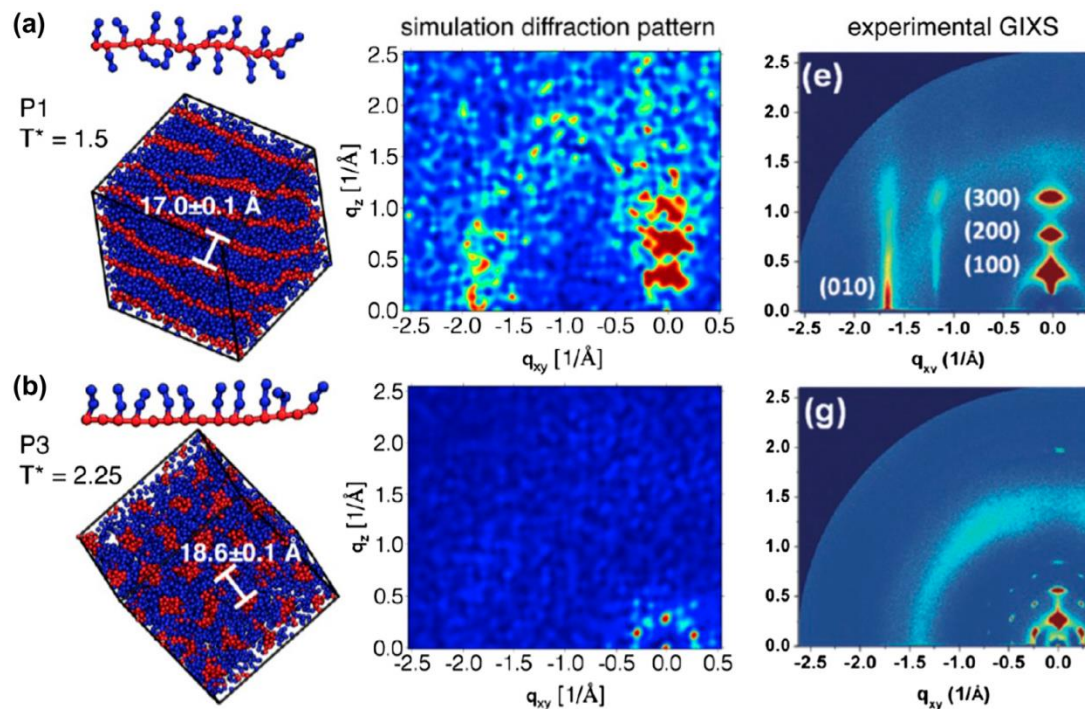**Example**: π-conjugated polymers

Figure: P3HT polymer from M L Jones, E Jankowski, Computationally connecting organic photovoltaic performance to atomistic arrangements and bulk morphology, Molec Simul 43, 756 (2017)
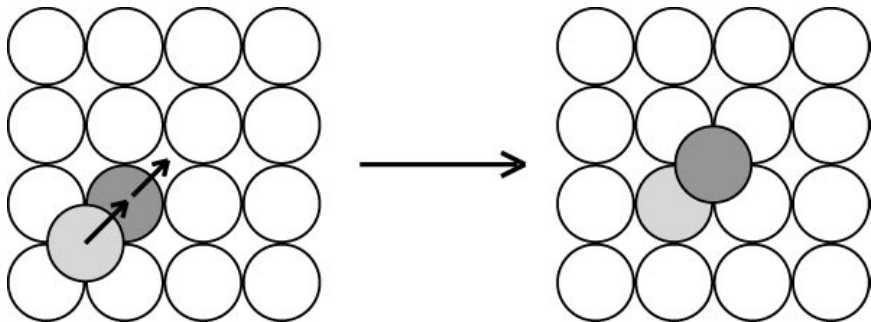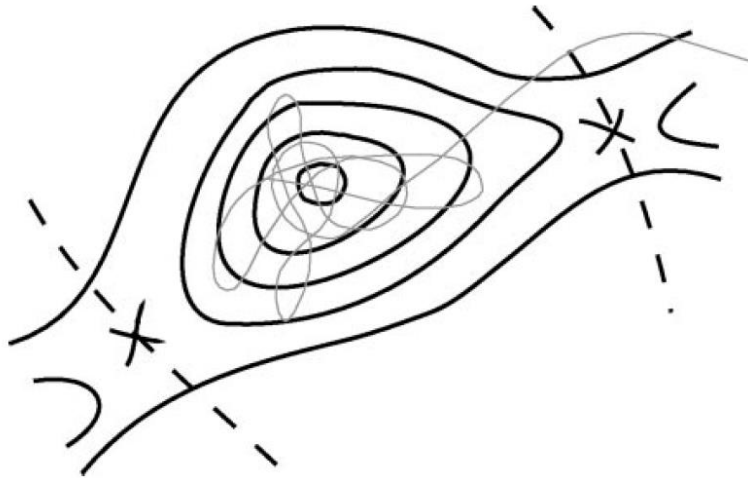
Extend MD time scale: accelerated dynamics

Idea: accelerate trivial dynamics and focus on nontrivial events

Infrequent Event Systems:
The correlation time (τ_{corr}) is the duration of the system memory.
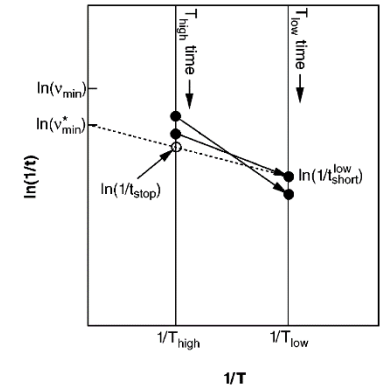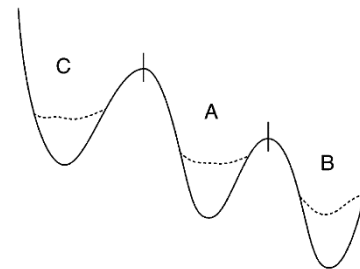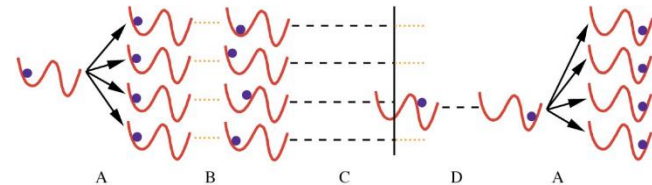**Infrequent event**: escape time >> τ_{corr}

EXTENDING THE TIME SCALE IN ATOMISTIC SIMULATION OF MATERIALS

Arthur F. Voter, Francesco Montalenti, and Timothy C. Germann

Annu. Rev. Mater. Res. 2002. 32:321–46

- Parallel replica MD
- Hyperdynamics
- Temperature-accelerated MD

Extend MD time scale: KMC

Idea: do not simulate transition events as MD – calculate rates and simulate the coarse grained model as Markov chain, i.e. do **Kinetic Monte Carlo** (KMC) simulations [Not atomistic modeling → out of scope of this course. Important: in many cases the KMC model can be solved without a simulation]

Example: single-atom diffusion of add-ions
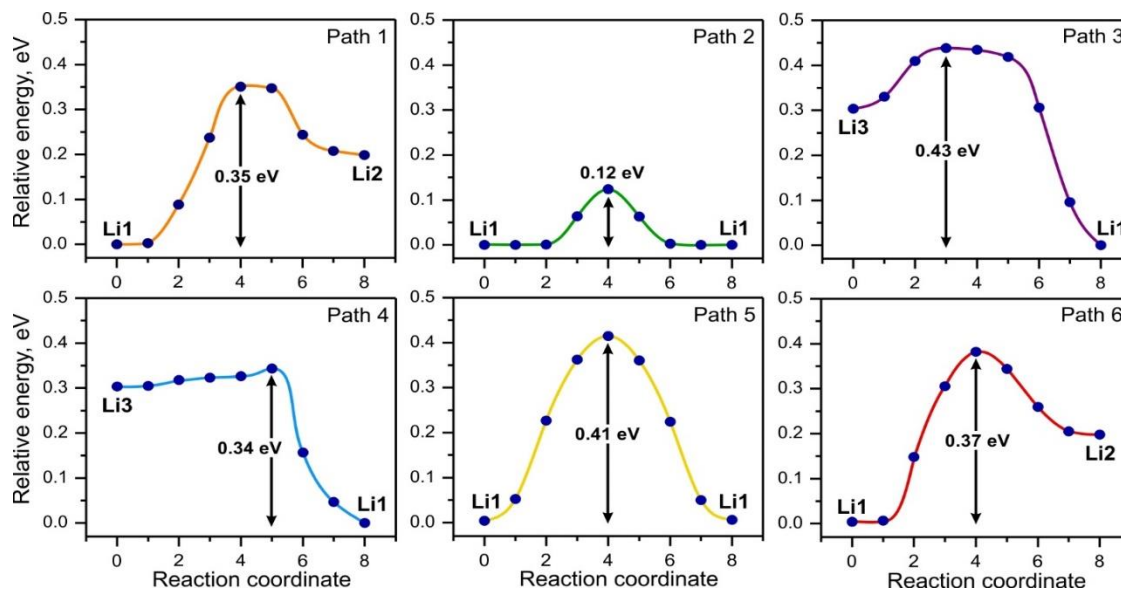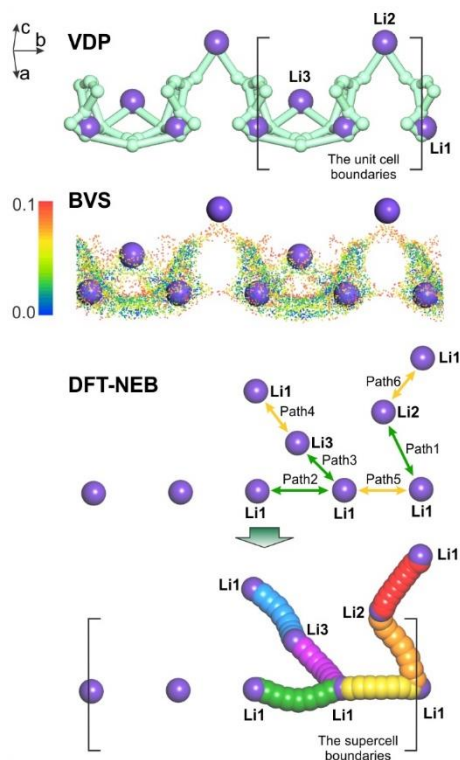
Figures: Li-ion transport in Li_2CoPO_4F high-voltage cathode material for Li-ion batteries, J Phys Chem C 121, 3194 (2017)

Extend MD time scale: effective Hamiltonian

Idea: isolate important degrees of freedom and parameterize appropriate effective Hamiltonian – solve that Hamiltonian by other methods
[Not atomistic modeling → out of scope of this course]

Example: small molecules intercalated in cage crystals

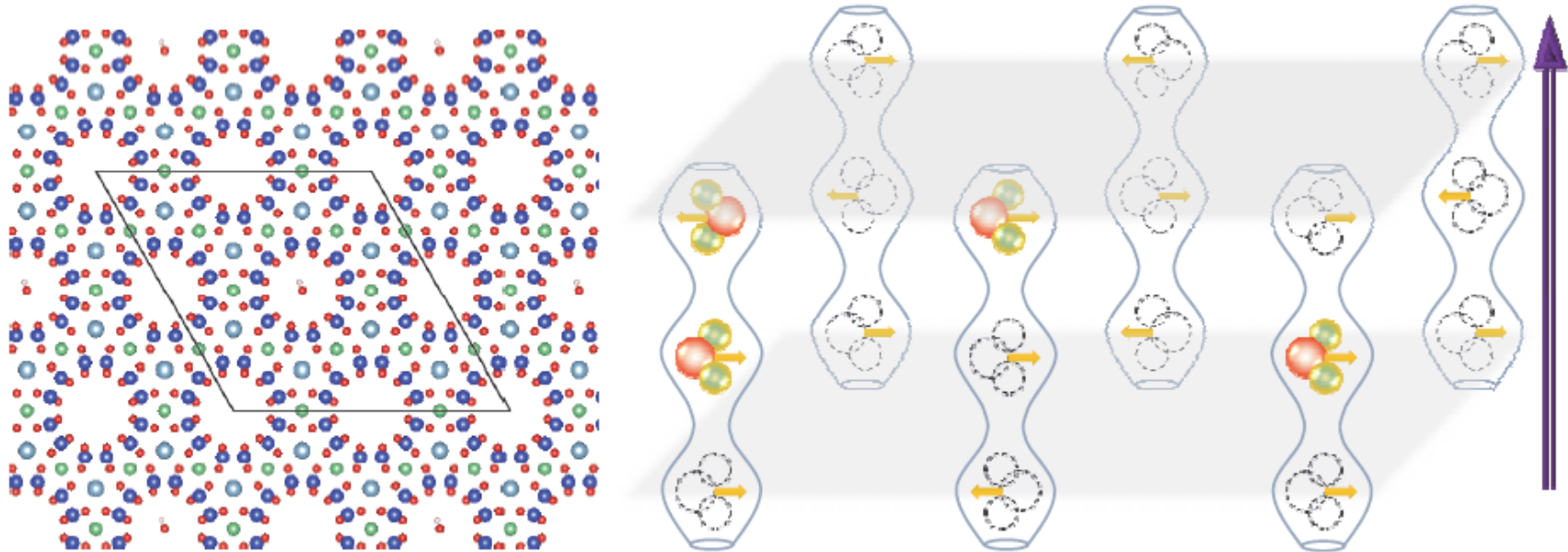**Figure**: water molecule in beryl from M A Belyanchikov (course alumnus) et al, Vibrational states of nano-confined water molecules in beryl investigated by first-principles calculations and optical experiments, Phys Chem Chem Phys 19, 30740 (2017)

Example 6: Protein folding – challenge for MD

Navigating the Folding Routes

Peter G. Wolynes, Jose N. Onuchic, D. Thirumalai
SCIENCE • VOL. 267 • 17 MARCH 1995

The extent of a protein energy landscape is huge. Before folding, each residue can take on about 10 different conformations; thus, a 60-residue protein can be in any of 10^{60} states. An unguided search, like a drunk playing golf, would take practically forever.

Concept of "funneled energy landscapes"
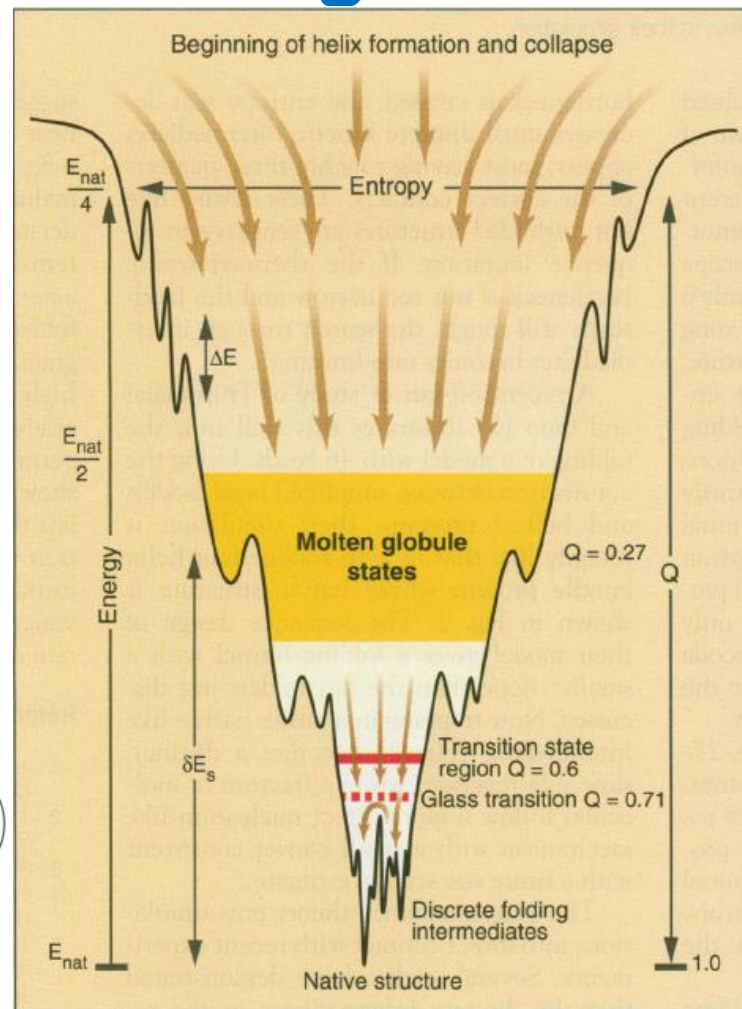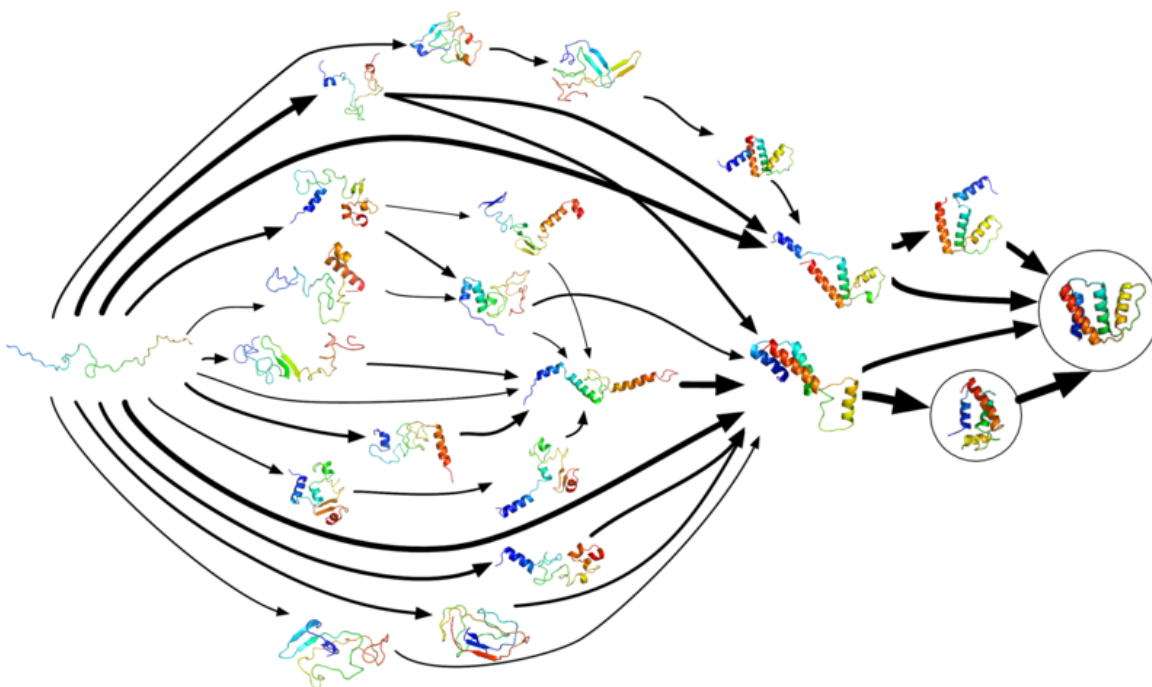
Fig. 1. Schematic of the folding funnel for a fast-folding 60-residue helical protein according to Onuchic *et al.* (*2*). The width of the funnel represents entropy, and depth, the energy. The flow of the molecule through the molten globule, folding bottleneck, or transition state ensemble and a glass transition region where discrete pathways emerge are indicated. The fraction of native contacts correctly made, *Q*, is indicated for each collection of states.

Example 6: Protein folding – challenge for MD

Folding a protein in a computer: An atomic description of the folding/unfolding of protein A

Angel E. García[†‡] and José N. Onuchic[§]

- 10–55 helical fragment B of protein A from *Staphylococcus aureus*
- Replica Exchange Molecular Dynamics (REMD), 82 replicas
- A cubic box containing 5,107 water molecules and 16,055 atoms
- T= 277-548 K, AMBER code.

Fig. 2. Contour maps of the free energy in the folded state ΔG(folded T =387 K). The ''folded'' state basin has two minima separated by a small barrier. These two minima correspond to the native state ($Q >$ 0.8, rmsd < 2 Å) and a nearly folded state with a hydrated core (0.30<Q < 0.8). The population in these two folded basins is equal at 387 K.
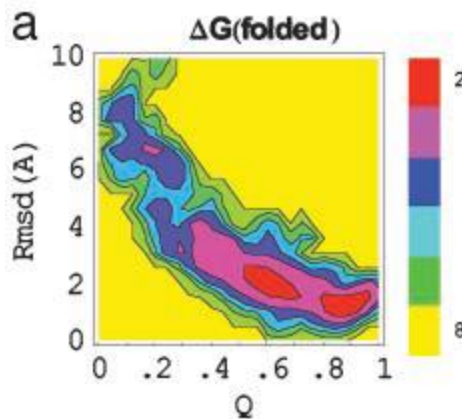
Folding a protein in a computer is now possible without the help of any structural constraints, other than those imposed by the limitation of the simulation box size, which may increase stability of the folded state but does not affect the folding mechanism.
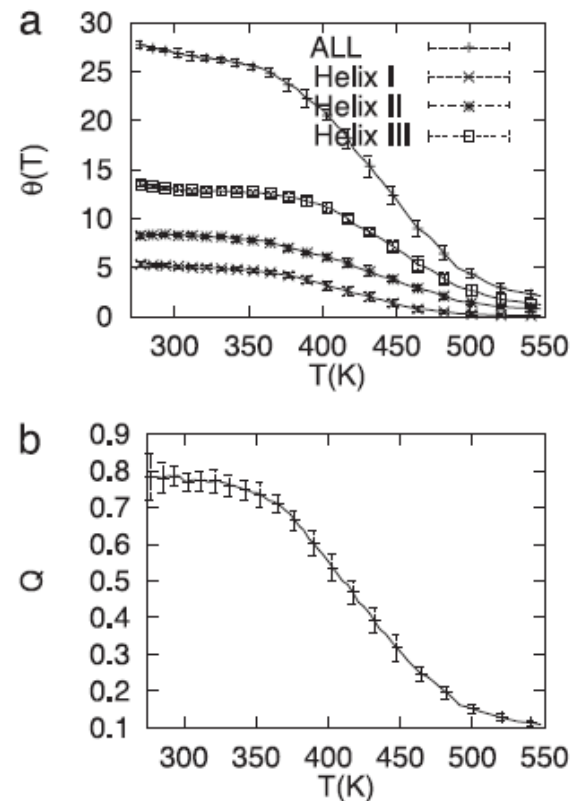
Fig. 1. (*a*) Average number of amino acids in the a-helices as a function of temperature for all amino acids (ALL) and for amino acids in a-helices I, II, and III. The temperature stability of the helices are helix III ~ helix II > helix I. (*b*) Average fraction of native contacts as a function of temperature.

Nonadiabatic MD

T Nelson, S Fernandez-Alberti, A E Roitberg, S Tretiak, Nonadiabatic Excited-State Molecular Dynamics: Modeling Photophysics in Organic Conjugated Materials, Acc Chem Res 47, 1155 (2014)
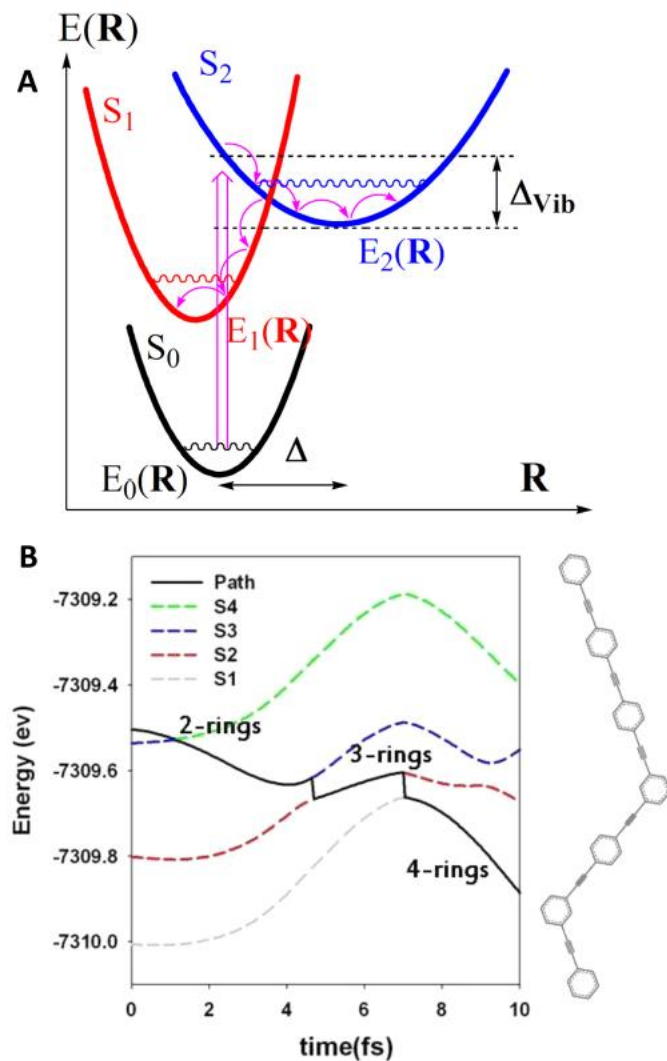
The total time-dependent electronic wave function is a mixed state, expanded in terms of adiabatic basis functions[18]

$$\Psi(\mathbf{r}, \mathbf{R}, t) = \sum_{\alpha} c_{\alpha}(t)\phi_{\alpha}(\mathbf{r}; \mathbf{R}(t)) \tag{1}$$

where $\mathbf{r}$ and $\mathbf{R}$ are electronic and nuclear coordinates, respectively, and $c_{\alpha}(t)$ are time-dependent expansion coefficients. The equation of motion for $c_{\alpha}(t)$ simplifies in the adiabatic Hamiltonian eigenstates, ϕ_{α}, as[18]

$$i\hbar\dot{c}_{\alpha}(t) = c_{\alpha}(t)E_{\alpha}(\mathbf{R}) - i\hbar\sum_{\beta}c_{\beta}(t)\dot{\mathbf{R}}\cdot\mathbf{d}_{\alpha\beta} \tag{2}$$

where $\mathbf{d}_{\alpha\beta} = \langle\phi_{\alpha}(\mathbf{r}; R)|\nabla_{R}\phi_{\beta}(\mathbf{r}; R)\rangle$ is the NA coupling vector (NACR) and the scalar NA coupling term (NACT) is $\dot{\mathbf{R}}\cdot\mathbf{d}_{\alpha\beta} = \langle\phi_{\alpha}(\mathbf{r}; \mathbf{R})|(\partial\phi_{\beta}(\mathbf{r}; \mathbf{R}))/(\partial t)\rangle$.[18,40] The time-dependent elements of the density matrix are $a_{\alpha\beta}(t) = c_{\alpha}^{*}(t)c_{\beta}(t)$, where diagonal terms provide the occupation probabilities of adiabatic states.

The probability of hopping from the current state α to another state during the time interval Δt is related to the probability flux $\dot{a}_{\alpha\alpha}(t) = \sum_{\beta\neq\alpha}b_{\alpha\beta}$ (see eq 5 in section 2.3) where $b_{\beta\alpha}(t) = -2\mathrm{Re}(a_{\alpha\beta}^{*}\dot{\mathbf{R}}\cdot\mathbf{d}_{\alpha\beta})$[18] and hops are accepted or rejected stochastically.[26] Following a hop, nuclei evolve on the PES of the new state, and energy is conserved by rescaling nuclear velocities along the direction of NACR.[41] If the nuclear kinetic energy is insufficient to allow a hop to higher energy, then the hop is *classically forbidden* and is rejected.

NAMD: practical considerations

Current challenge: Make it computationally tractable and practically relevant (1000 atoms for 1 ps)

A. Simplify e-structure method to semiempirical or DFTB

See **T Nelson**, S Fernandez-Alberti, A E Roitberg, **S Tretiak**, Nonadiabatic Excited-State Molecular Dynamics: Modeling Photophysics in Organic Conjugated Materials, Acc Chem Res 47, 1155 (2014)
ES-NAMD code

B. Simplify MD-propagation method to 'state-average' forces

See L Wang, **A Akimov**, **O V Prezhdo**, Recent Progress in Surface Hopping: 2011-2015, J Phys Chem Lett 7, 2100 (2016)
Pyxaid code

Next level, Quantum MD, is of limited use (10 atoms) → out of scope

NAMD: example

Idea: conduct non-adiabatic molecular dynamics simulations to explore energy transfer processes from dendrimer's periphery to the center following photoexcitation

Chemical structure of 2G1m-Eper dendrimer and the 1000 different structures obtained from QM-MM dynamics surface exploration (AM1+explicit solvent)
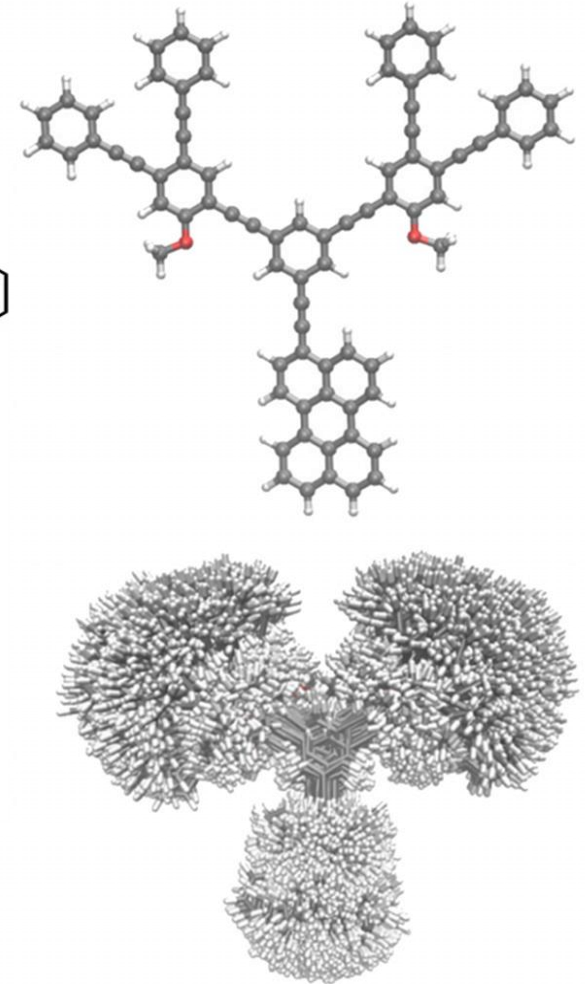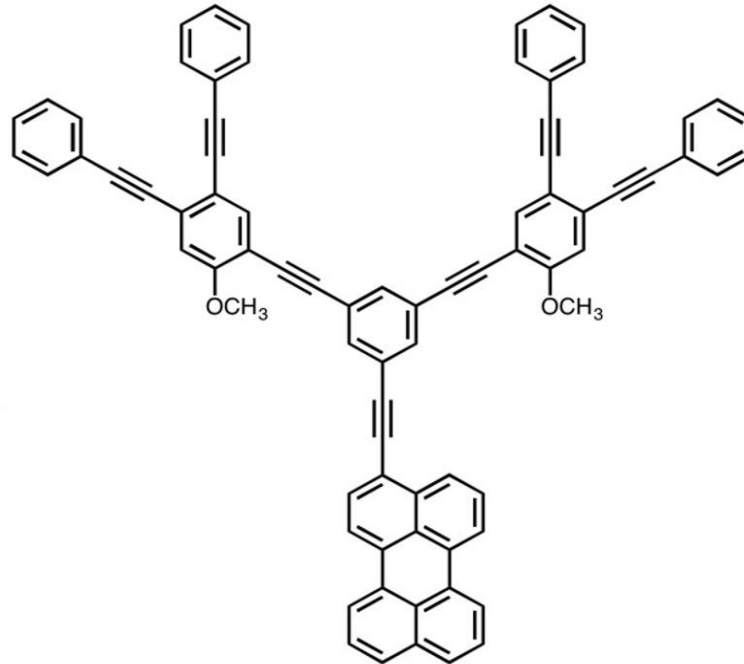
J F Galindo, E Atas, A Altan, D G Kuroda, S Fernandez-Alberti, S Tretiak, A E Roitberg, V Kleiman, Dynamics of energy transfer in a conjugated dendrimer driven by ultrafast localization of excitations, JACS 137, 11637 (2015)

Discussion

1. What would be a reasonable scale for time step: as, fs, ps, ns, μs?

2. What would be a reasonable size of the supercell?

3. How long should you run single MD trajectory?

4. Is there a reason to run multiple trajectories?

5. You started MD but the output looks unrealistic from the very beginning of your trajectory. What would you do?

Potential Energy Surface (PES) exploration

1) Geometry relaxation and derivatives – point-wise exploration of PES
(local minima, saddle points, conformers, vibrations)

2) Real-time dynamics – local exploration of PES
(chemical reactions, conformational dynamics, crack propagation)

3) Thermodynamics – global exploration of PES
(ensemble averages, phase diagram)

Reminder on statistical mechanics

Thermodynamics variables: volume (V), pressure (P), temperature (T), number of particles (N), energy (E), chemical potential (μ)

Statistical mechanics: connection between properties of a microscopic system and a macroscopic sample

Ensemble: collection of microscopic states consistent with thermodynamic boundary conditions; defined by 3 variables (NVT) or (NVE) or (NPT) or (μVT)

Boltzmann probability: relative probability to be in a state with an energy E at a temperature T and is proportional to exp($-E/kT$)

Partition function is a key quantity in statistical mechanics, a normalization factor for the Boltzmann probability distribution:

$$Q = \sum_{i}^{\infty} e^{-E_i/kT} = \int e^{-E(\mathbf{r},\mathbf{p})/kT} \, d\mathbf{r} d\mathbf{p}$$

Everything can be expressed through Q, e.g., for a canonical ensemble (NVT)

- Internal energy $\quad U = kT^2 \left(\dfrac{\partial \ln Q}{\partial T} \right)_V$

- Enthalpy $\quad H = U + PV = kT^2 \left(\dfrac{\partial \ln Q}{\partial T} \right)_V + kTV \left(\dfrac{\partial \ln Q}{\partial V} \right)_T$

- Entropy $\quad S = \dfrac{U - A}{T} = kT \left(\dfrac{\partial \ln Q}{\partial T} \right)_V + k \ln Q$

- Gibbs free energy $\quad G = H - TS = kTV \left(\dfrac{\partial \ln Q}{\partial V} \right)_T - kT \ln Q$

- Helmholtz free energy $\quad A = U - TS = -kT \ln Q$

Thermodynamic ensembles

Microcanonical (NVE): *Newtonian system (N=const) in box (V=const) with elastic walls (or periodic boundary conditions)*

Canonical (NVT): *Newtonian system (N=const) in box (V=const) with non-elastic walls (walls are equilibrated with T=const - thermostat)*

Isothermal-isobaric (NPT): *Newtonian system (N=const) in box with varying volume (keeping P=const - barostat) and non-elastic walls (keeping T=const - thermostat)*

Grand-canonical (μVT): *Open system (number of particles is not conserved but their energy in the reservoir is fixed at μ)*
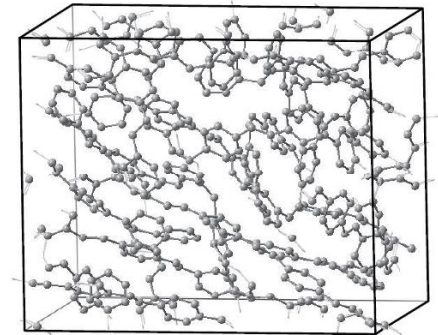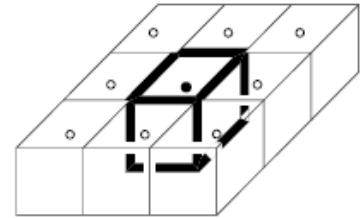
Table 14.1 Constants in different ensembles, and corresponding equilibrium states (From Jensen)

N	P	V	T	E	μ	Acronym	Equilibrium	Name
×		×	×			NVT	A has minimum	Canonical
×		×		×		NVE	S has maximum	Micro-canonical
×	×		×			NPT	G has minimum	Isothermal-isobaric
		×	×		×	$VE\mu$	(PV) has maximum	Grand canonical

Question: At what number of particles NVE≈NVT?

Statistical sampling: MD vs MC

Ergodicity: Time-average of a property can be replaced by a suitable average over collection of possible microscopic states (***ensemble average***):

$$\langle A \rangle = \int\int A(\mathbf{q}, \mathbf{p}) P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} \qquad P(\mathbf{q}, \mathbf{p}) = Q^{-1} e^{-E(\mathbf{q}, \mathbf{p})/k_B T}$$

$$Q = \int\int e^{-E(\mathbf{q}, \mathbf{p})/k_B T} d\mathbf{q} d\mathbf{p}$$

Given a potential energy surface $E_p(r)$, what are the possible modeling avenues for generating ensembles?

Molecular Dynamics (MD): propagate Newtonian equation of motion, analyze trajectories – generally gives rates and time constants.

Monte-Carlo (MC): 'flip a coin' statistical approach (*T* is an essential component) – gives information on the thermodynamics, no rates and time scales.

Table 14.2 Differences between Monte Carlo and molecular dynamics methods   (From Jensen)

Property	MC	MD
Basic information needed	Energy	Gradient
Particles moved in each step	One	All
Coordinates	Any	Cartesian
Constraints	Easy	Difficult
Atomic velocities	No	Yes
Time dimension	No	Yes
Deterministic	No	(Yes)
Sampling	Non-physical	Physical
Natural ensemble	*NVT*	*NVE*

MC idea: importance sampling

Replace strategy
"First pick, then weight" with
"First weight, then pick"

by penalizing states with high energy by a Boltzmann factor **exp (-ΔE/kT)**

How?

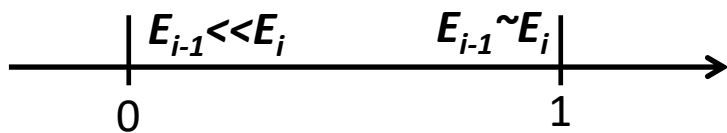Say on (**i-1**)-th step, the energy of accepted configuration is E_{i-1}
on **i**-th step, the energy of new configuration is E_i

Case 1: $E_i < E_{i-1}$: Good! We are going lower, accept **i**

Case 2: $E_i > E_{i-1}$: We are going higher, calculate **exp (-ΔE/kT)=exp (-(E_i-E_{i-1})/kT)**, pick a random number ζ from [0,1]
If $\zeta > $ **exp (-ΔE/kT)** reject it
If $\zeta < $ **exp (-ΔE/kT)** accept it

$$\underset{\substack{| \\ 0}}{\Big|} \overset{E_{i-1} \ll E_i}{} \overset{E_{i-1} \sim E_i}{\underset{\substack{| \\ 1}}{\Big|}} \longrightarrow$$

Metropolis (+ Fermi, Ulam, von Neumann) Monte-Carlo (1952): "walks" through phase space (Markov chain of states) visiting each state with proper probability (in the infinite time limit)

Algorithm:

1. *Generate trial configuration q_1, its energy E_1=E(q_1) and store the desired property A_1=A(q_1)*
2. *For **i=2,…,M** do*
3. *Perturb the system, get new configuration q_i*
4. *Calculate energy of q_i : E_i=E(q_i)*
5. *Test E_i vs E_{i-1} for acceptance (see left)*
6. *If rejected, discard q_i*
7. *If accepted, calculate and store A_i=A(q_i)*
8. *If **i<M**, go to Step 2*
9. *Calculate* $\left\langle \mathrm{A(M)} \right\rangle = \dfrac{1}{\mathrm{M}} \sum_{i=1}^{\mathrm{M}} \mathrm{A}_i$

MC: some practical considerations

1) The natural Monte-Carlo ensemble is (NVT)

2) The "Devil is in details": The art of running an MC calculation lies in defining the perturbation step(s). If the steps are very small, then the volume of phase space sampled will increase only slowly over time, and the cost will be high in terms of computational resources. If the steps are too large, then the rejection rate will grow so high that again computational resources will be wasted by an inefficient sampling of phase space.

2) Monitoring convergence of <A(M)> and possibly <E(M)> with M is necessary. This will tell you a lot about convergence rate and possible transition to a different PES valley.

3) Multiple MC variations were developed, including different ensembles, annealing (varying T) algorithms, simulating timescales (e.g. kinetic MC), etc.

For example, (NPT) isothermal-isobaric ensemble, G (Gibbs free energy) at min (frequent in experimental conditions)

$$\Delta E \rightarrow \Delta E + P\Delta V - NkT\ln(1 + \Delta V/V)$$
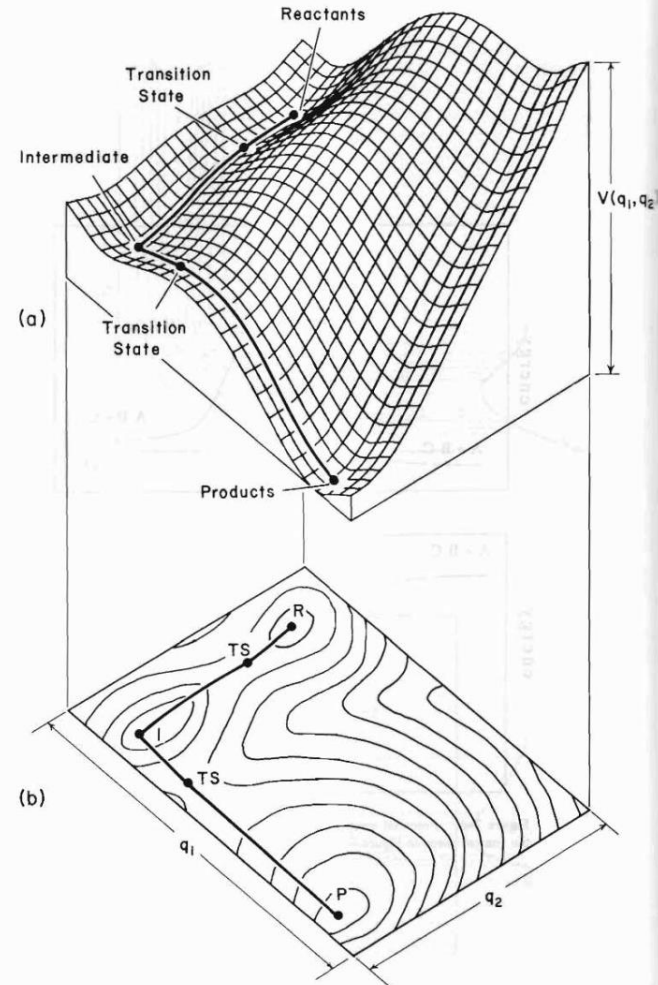
LAMMPS users see [fix gcmc](#) command

Figure 7–7. A three-dimensional perspective (a) and contour map (b) for a model chemical reaction. The solid line is the reaction path. [Adapted from G. M. Maggiora and R. E. Christofferson, in *Transition States of Biochemical Processes*, ed. R. D. Gandour and R. L. Schowen (New York: Plenum, 1978).]

PES example from Steinfeld, Francisco, and Hase