BINF 731

Protein Structure Analysis

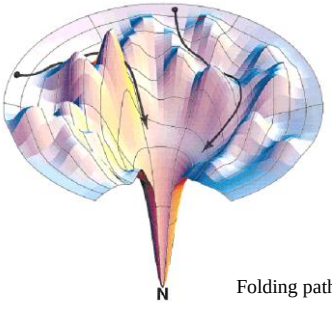
Iosif Vaisman

2023

Protein Modeling Methods

- ***Ab initio* methods:**
  solution of a protein folding problem
  search in conformational space
- **Energy-based methods:**
  energy minimization
  molecular simulation
- **Knowledge-based methods:**
  homology modeling
  fold recogniion

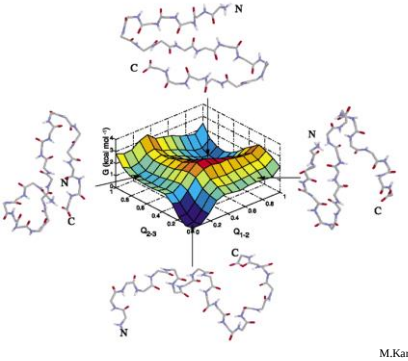
HP Lattice Models



Folding pathways

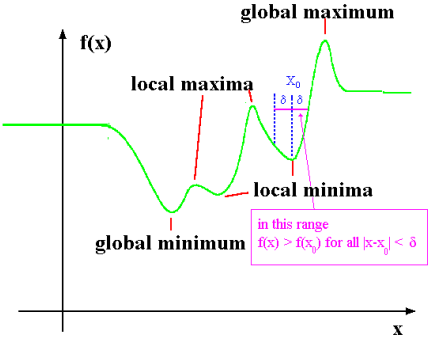
Chan & Dill, 1998

Free energy surface in protein simulation

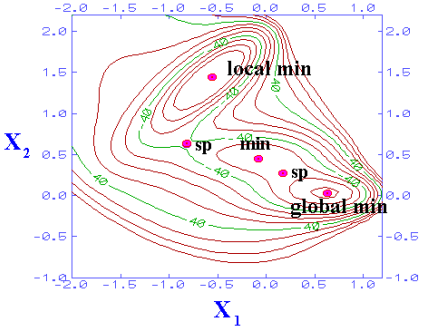


M.Karplus and J. A. McCammon, 2002

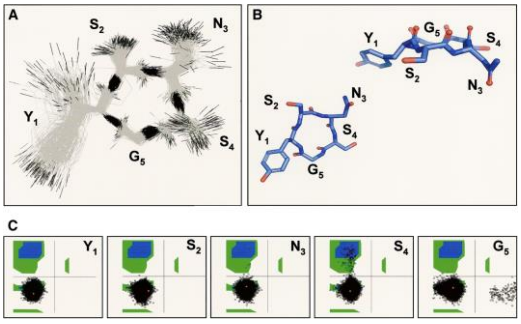
Energy Minimazation



Energy Minimazation

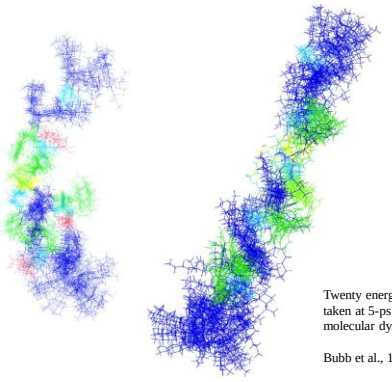


Molecular dynamics



YSNSG cyclopeptide as observed along the 20 ns molecular dynamics trajectory (Thevenard et al., 2006)

Molecular dynamics



Twenty energy minimized structures taken at 5-ps intervals along a 100-ps molecular dynamics trajectory.

Bubb et al., 1999

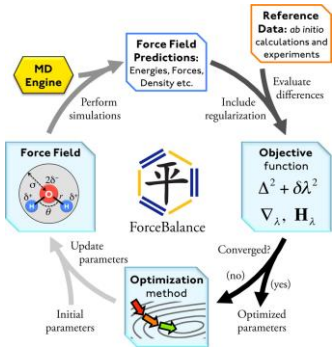
Molecular Dynamics

- Model system
- Initial conditions
- Boundary conditions
- Integration algorithm
- Constraints
- Ensemble
- Results

Potential Energy Function and Force Field

$$V(\vec{R}) = \sum_{\text{bonds}} K_{\text{d}}(d-d_0)^2 + \sum_{\text{Urey-Bradley}} K_{\text{UB}}(S-S_0)^2 + \sum_{\text{angle}} K_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} K_\chi(1+\cos(n\chi-\delta)) + \sum_{\text{impropers}} K_\phi(\phi-\phi_0)^2 + \sum_{\text{nonbond}} \left\{ \in_{ij} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{12} - \left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{\in_l r_{ij}} \right\}$$

Force Field Development and Parametrization



L.-P. Wang et al., 2014

Molecular Dynamics

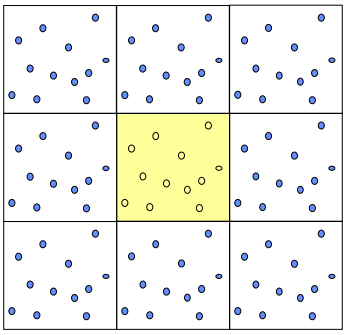
$$F_i = m_i a_i$$

$$a_i = dv_i / dt$$
$$v_i = dr_i / dt$$

$$- dE / dr_i = F_i$$

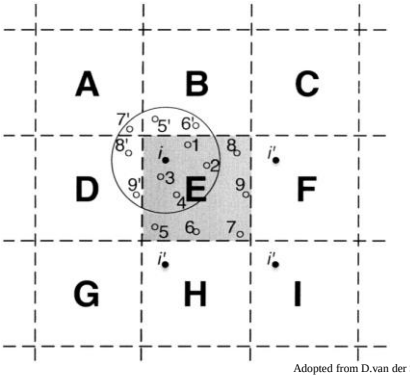
$$- dE / dr_i = m_i d^2 r_i / dt^2$$

Periodic Boundary Conditions



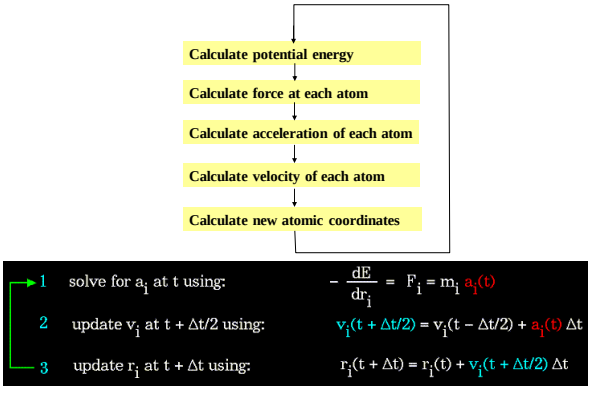
J. Jensen, animation

Periodic Boundary Conditions



Adopted from D.van der Spoel et al. (2005)

MD cycle and integration algorithm



Characteristic Time Scales for Protein Motions

event	spatial extent (nm)	amplitude (nm)	time (s)	appropriate simulations
bond-length vibration	0.2–0.5	0.001–0.01	10^{-14}–10^{-13}	QM methods
elastic vibration of globular domain	1.0–2.0	0.005–0.05	10^{-12}–10^{-11}	conventional MD
rotation of solvent-exposed side chains	0.5–1.0	0.5–1.0	10^{-11}–10^{-10}	conventional MD
torsional libration of buried groups	0.5–1.0	0.05	10^{-11}–10^{-9}	conventional MD
hinge bending (relative motion of globular domains)	1.0–2.0	0.1–0.5	10^{-11}–10^{-7}	Langevin dynamics, enhanced sampling MD methods?
rotation of buried side chains	0.5	0.5	10^{-4}–1	enhanced sampling MD methods?
allosteric transitions	0.5–4.0	0.1–0.5	10^{-5}–1	enhanced sampling MD methods?
local denaturation	0.5–1.0	0.5–1.0	10^{-5}–10^{1}	enhanced sampling MD methods?
loop motions	1.0–5.0	1.0–5.0	10^{-9}–10^{-5}	Brownian dynamics?
rigid-body (helix) motions		1.0–5.0	10^{-9}–10^{-6}	enhanced sampling MD methods?
helix–coil transitions		>5.0	10^{-7}–10^{4}	enhanced sampling MD methods?
protein association	≫1.0			Brownian dynamics

S. A. Adcock and J. A. McCammon, 2006

MD Ensemble

Microcanonical ensemble (NVE) :
The thermodynamic state characterized by a fixed number of atoms, N, a fixed volume, V, and a fixed energy, E. This corresponds to an isolated system.

Canonical Ensemble (NVT):
This is a collection of all systems whose thermodynamic state is characterized by a fixed number of atoms, N, a fixed volume, V, and a fixed temperature, T.

Isobaric-Isothermal Ensemble (NPT):
This ensemble is characterized by a fixed number of atoms, N, a fixed pressure, P, and a fixed temperature, T.

Grand canonical Ensemble (μVT):
The thermodynamic state for this ensemble is characterized by a fixed chemical potential, m, a fixed volume, V, and a fixed temperature, T.

Temperature in molecular dynamics

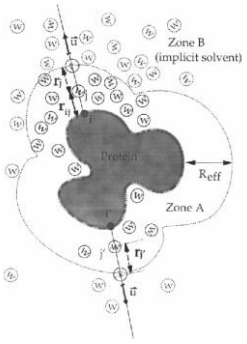
$$U_{kin} = \sum \frac{1}{2} m_i v_i^2 = \frac{3}{2} NkT$$

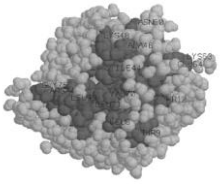
N – *number of atoms*

k – *Boltzmann constant*

T – *absolute temperature*

MD of proteins: Solvent model

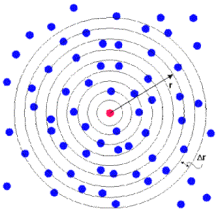


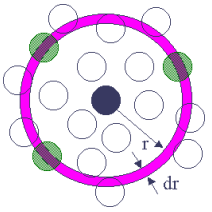


MD simulation of ubiquitin
(400 ps)

Adopted from V.Daggett (1999)

MD of proteins: radial distribution functions

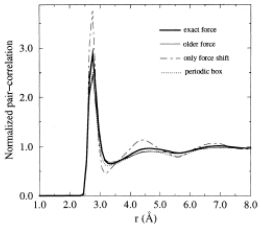




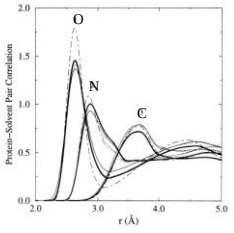
$$N_{ideal}(r) = \frac{N}{V} \times V_{shell}(r) = \rho \times 4\pi r^2 dr$$

$$\mathbf{g(r)} = \frac{\mathbf{N(r)}}{\mathbf{N_i(r)}} = \frac{\mathbf{N(r)}}{(\boldsymbol{\rho} \times \mathbf{4\pi\, r^2 dr})}$$

MD of proteins: radial distribution functions



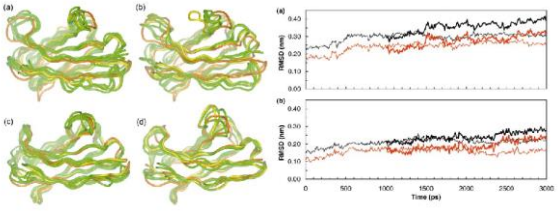
Water O-O rdf



Protein-water rdf

Adopted from V.Daggett (1999)

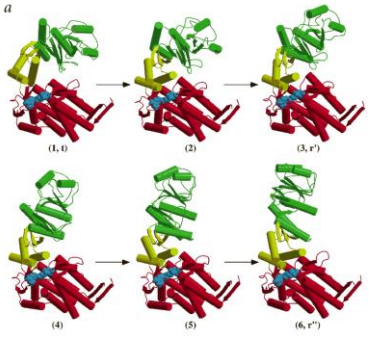
MD of proteins: mobile regions



Snapshots of V_H domain simulation
at 300 and 340 K

Adopted from W.F.VanGunsteren (2001)

MD of proteins: Conformational change

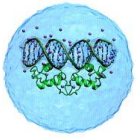


A set of structures on the reaction path showing the behavior of a single subunit (apical domain, green; intermediate domain, yellow; equatorial domain, red; and ATP, blue). Structures 1−3 correspond to the first stages associated with ATP binding; 4−6 correspond to the second stage involving GroES binding. The early downward motion of the intermediate domain (compare [1,t] with [2] and [6,r"]) is the trigger for the overall transition.

M.Karplus and J. A. McCammon, 2002

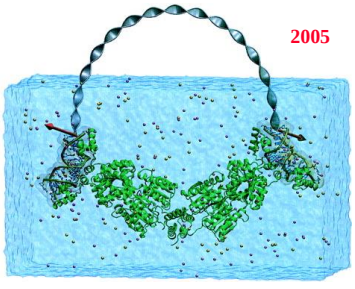
MD of proteins: scale of simulation

1997



estrogen receptor
DNA-binding domain
36,000 atoms
100 ps

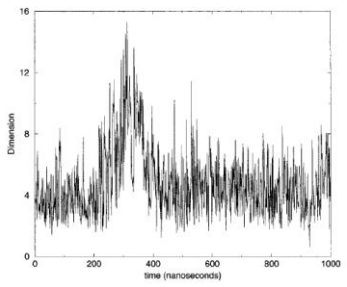
2005



LacI-DNA complex
314,000 atoms
10 ns

Adopted from J.C.Phillips et al. (2005)

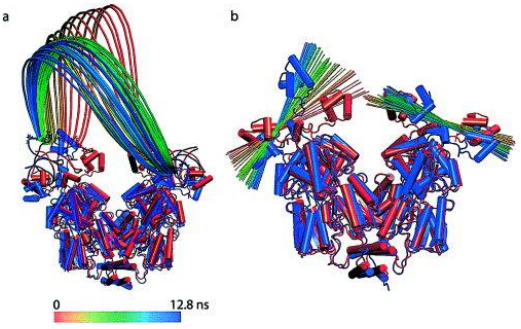
MD of proteins: long runs



1 microsecond simulation of villin

Adopted from I.D.Kuntz and P.Kollman (2001)

MD of proteins: long runs



Adopted from J.C.Phillips et al. (2005)

MD of proteins: performance

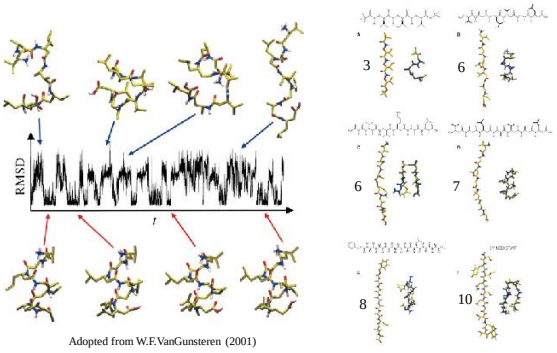
1999

Simulation setup						Performance, ps/day		
System	FF	virtH	Water	Coulomb	LJ	ia32	x86-64	ppc
Vil	G	no	TIP3P	cutoff 0.8	cutoff 0.8	9744	9574	14,385
Vil	G	yes	TIP3P	cutoff 0.8	cutoff 0.8	16,900	16,895	23,681
Vil	G	yes	TIP3P	RF 1.0	cutoff 1.0	10,308	9719	12,934

2014

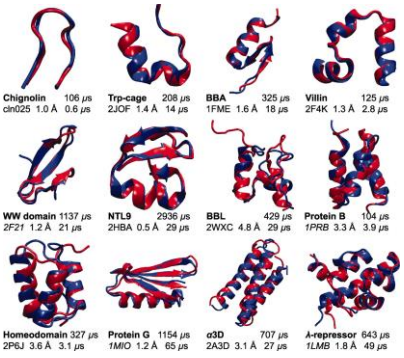
System (PDB ID)	Number of atoms	Approximate performance (μs/machine-day)
*DHFR (5DFR)	23,558	17.4
aSFP (1SFP)	48,423	11.7
FtsZ (1FSZ)	98,236	5.7
T7Lig (1A01)	116,650	5.5
bILAP (1BPM)	132,362	4.8

MD: Reversible folding of peptides



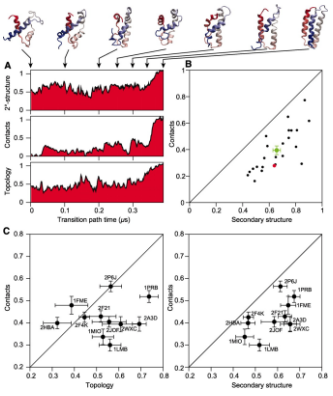
Adopted from W.F.VanGunsteren (2001)

MD: Reversible folding of small proteins



K. Lindorff-Larsen et al., 2011

MD: Reversible folding of small proteins



Formation of topology, native contacts, and secondary structure during protein folding.

(A) The three panels show the accumulation of native secondary structure, nonlocal native contacts, and native topology during a single folding event for α3D

(B) The 24 transitions of α3D in a scatter plot are represented, with each of the black points corresponding to the time series integral for a single folding event

(C) Each point shows the average value over all folding and unfolding events observed for one protein. Each point is labeled with the PDB code of the relevant protein. Most proteins fall below the diagonal in these plots, showing that topology and secondary structure develop earlier than the full set of native contacts.

K. Lindorff-Larsen et al., 2011