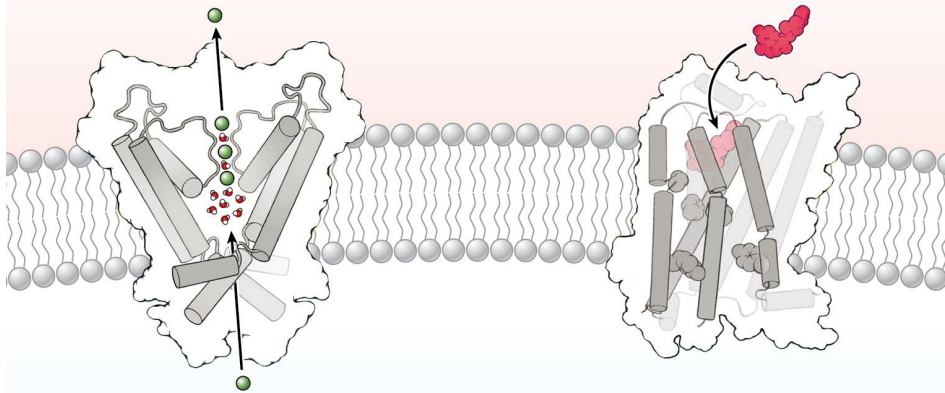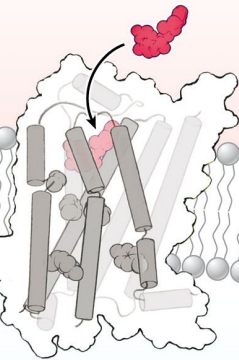# Week 12: Protein structure

- Amino-acid coevolution
  - Mutual information
  - Maximum entropy modeling
- **Molecular dynamics**

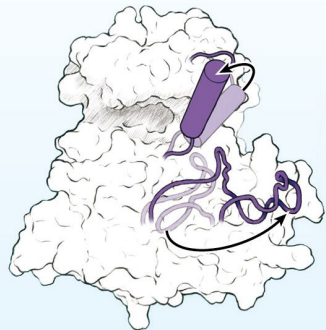# Molecular dynamics (MD) simulations = Computational microscope
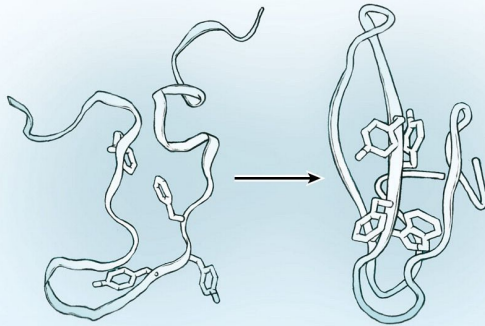

a Transport across membrane

b Ligand binding

**c** Conformational change

d Protein folding

MD simulations reveal the workings of biomolecular systems at a spatial and temporal resolution that is often difficult to access experimentally.

- Positions and velocities of atoms are computed using Newton's laws of motion.

Dror (2012) Annu. Rev. Biophys.

Molecular dynamics (MD) simulations = Computational microscope

The basic MD algorithm:
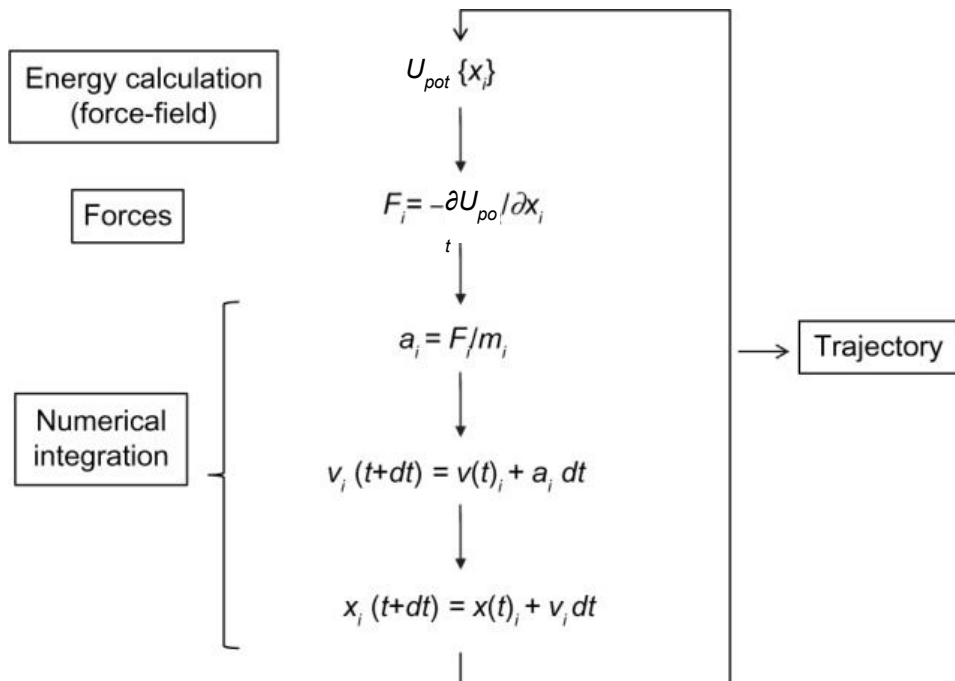
The simulation output – the trajectory – is an ordered list of 3N atom coordinates for each simulation time (or snapshot).

U_{pot}: potential energy

t: simulation time

dt: iteration time

For each spatial coordinate of the N simulated atoms (i):

- x: atom coordinate
- F: forces component
- a: acceleration
- m: atom mass
- v: velocity.

Hospital (2016) Adv. Appl. Bioinform. Chem.

Force field and the energy function

The potential energy of N interacting atoms $U(\mathbf{r_1}, \ldots, \mathbf{r_N})$ is a function of their positions $\mathbf{r_i} = (x_i, y_i, z_i)$.

The force acting upon ith atom is determined by the gradient (vector of first derivatives) with respect to atomic displacements:

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1, \cdots, \mathbf{r}_N) = -\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i}\right)$$

Find the positions $r_i(t + \Delta t)$ at time $t + \Delta t$ in terms of the already known positions at time t.

Verlet algorithm:

$$\mathbf{r}_i(t + \Delta t) \cong 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\mathbf{F}_i(t)}{m_i}\Delta t^2$$

Force field: energy function used to compute the forces acting on atoms (due to interatomic interactions) during an MD simulation.

Force field and the energy function

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$
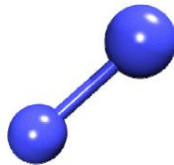
Bond

Angle

Dihedral

Improper

U_{bond}: oscillations about the equilibrium bond length

U_{angle}: oscillations of 3 atoms about an equilibrium angle
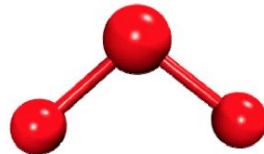
$U_{dihedral}$: torsional rotation of 4 atoms about a central bond

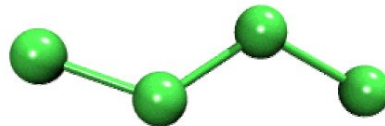$U_{nonbond}$: non-bonded energy terms (electrostatics and Lenard-Jones)

Force field: energy function used to compute the forces acting on atoms (due to interatomic interactions) during an MD simulation.

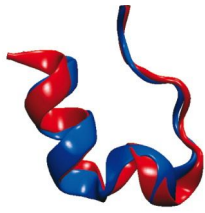# Steps in a typical MD simulation

1. **Preparing molecule**: Read in pdb and psf file

2. **Minimization**: Reconcile observed structure with force field used (T = 0)

3. **Heating**: Raise temperature of the system

4. **Equilibration**: Ensure system is stable

5. **Dynamics**: Simulate under desired conditions (NVE, NpT, etc); Collect your data

6. **Analysis**: Evaluate observables (macroscopic level properties); Or relate to single molecule experiments.
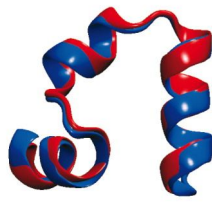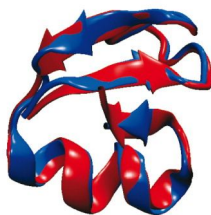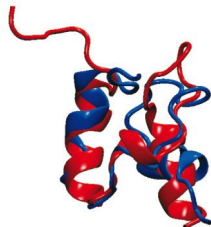
Simulations of structurally diverse proteins

Chignolin

Trp-cage

BBA

Villin

WW domain

NTL9

BBL
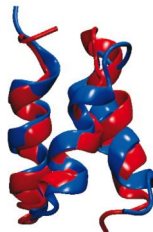
Protein B
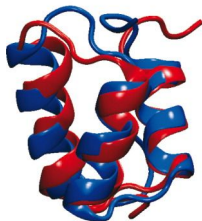
Homeodomain

Protein G
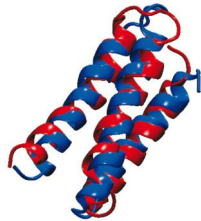
α3D
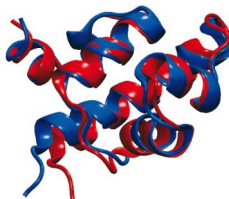
λ repressor

Simulations with a single force field.

- 12 structurally diverse proteins fold spontaneously to a structure (blue) closely resembling that determined experimentally (red).

- Simulation snapshots chosen automatically based on a **clustering analysis** that did not exploit knowledge of the experimental structure.

- Total simulation time per protein: 104 – 2,936 µs – allowing observation of at least 10 folding & 10 unfolding events for each protein.

Dror (2012) Annu. Rev. Biophys.

Beta-blockers binding spontaneously to the β2-adrenergic receptor

Beta blockers – aka beta-adrenergic blocking agents – reduce blood pressure by blocking the effects of epinephrine (adrenaline).

Root mean square deviation (rmsd) of the ligand in simulation from its position in the crystal structure.

Dror (2012) Annu. Rev. Biophys.
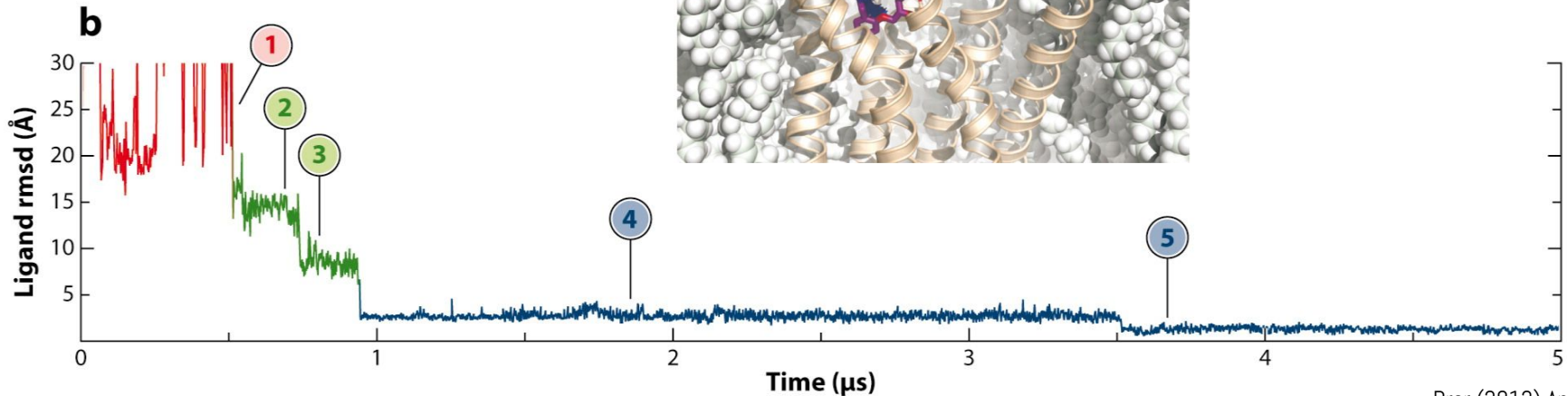
Beta-blockers binding spontaneously to the β2-adrenergic receptor
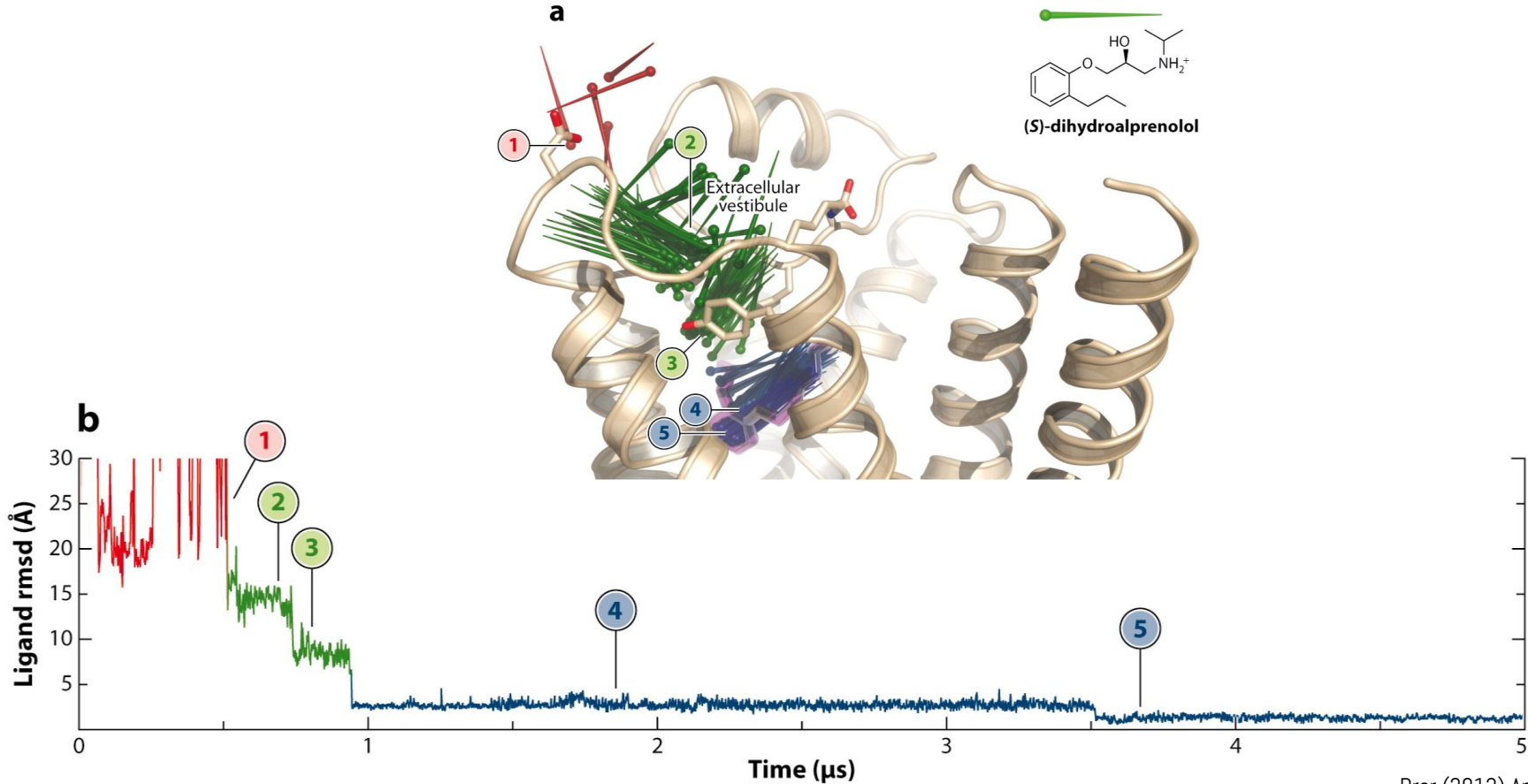
Beta-blockers binding spontaneously to the β2-adrenergic receptor

Metastable Intermediate stages of beta blocker binding.

1: Ligand moves from bulk solvent…

2, 3: … into the extracellular vestibule, and finally…

4, 5: … into the binding pocket.

Dror (2012) Annu. Rev. Biophys.

A. Modulating Protein Stability

I. Perform simulations in native and nonnative environments

Native MD

Nonnative MD
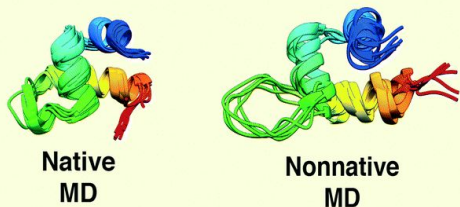
II. Inform designs by analyzing the dynamics of unstable residues

Unstable Residues
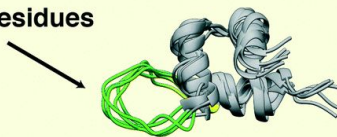
III. Perform simulations of designs to determine the impact of mutations on protein stability

B. Engineering Functional Regions

I. Perform simulations that capture functional dynamics

Native MD

II. Inform designs by analyzing the dynamics of functional residues

Functional Residues
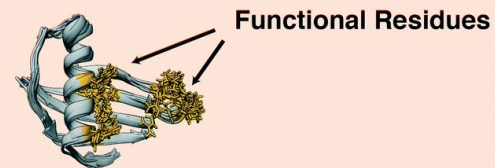
III. Perform simulations of designs to assess the impact of mutations on function regions

Childers & Daggett (2017) Mol. Syst. Des. Eng.,

C. Insights From Folding Pathways

I. Simulate the unfolding/folding pathway and partition the trajectory into conformational states
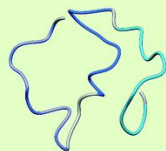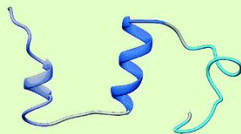
Denatured State ⟷ Folding Intermediates ⟷ Transition State ⟷ Native State
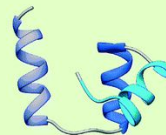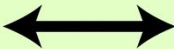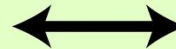
II. Inform designs with insights from specfic conformations or transitions along the folding pathway

Destabilize the Denatured State | Stabilize Folding Intermediates | Design Fast Folding Variants | Probe / Alter the Folding Pathway

III. Perform simulations of designs to assess the impact of mutations on the folding landscape

Childers & Daggett (2017) Mol. Syst. Des. Eng.,

MD simulations for protein design

Direct

1. Simulate the design target over a range of temperatures; then analyze the simulations to resolve atomistic details of flexible sites.

MD, 300 K   MD, 330 K   MD, 360 K

Native

Non-Native

Increasing Temperature, Increasing Flexibility
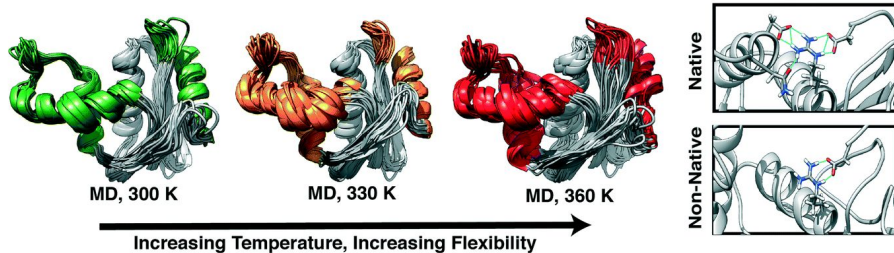
2. Select mutations by combining insights from MD simulations with available data.

Flexible Residues
N-   -C

Potential Disulfide Bonds
N-   -C

Conserved Residues
N-   -C

Candidate Mutations
N-   -C
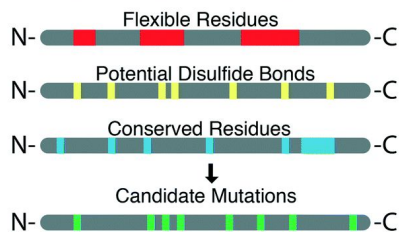
3. Assess design(s) computationally,

iteratively design, simulate, and assess

MD, 360 K

4. Evaluate the design experimentally

Indirect

1. Construct multiple designs using available data

Design 1   Design 2   Design 3   . . .   Design N
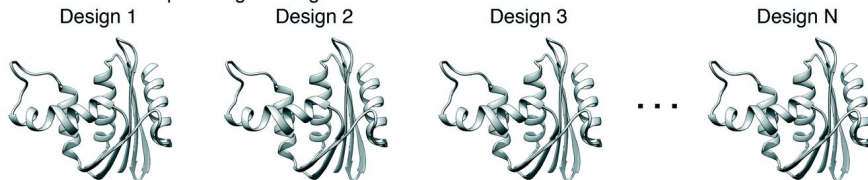
2. Perform simulations of designs at high temperature and rank designs according to their flexibility and other structural or dynamic criteria

MD, 360 K   MD, 360 K   MD, 360 K   MD, 360 K
❌ Flexible   ✅ Rigid   ❌ Flexible   ✅ Rigid
✅ Native-like Interactions   ✅ Native-like Interactions   ❌ Non-Native-like Interactions   ❌ Non-Native-like Interactions

Best Design
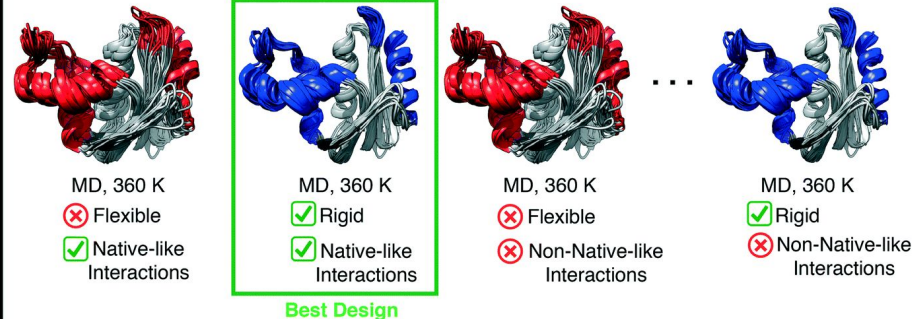
3. Evaluate the top scoring design(s) experimentally

Childers & Daggett (2017) Mol. Syst. Des. Eng.,

Virtual screening: docking & MD simulations

A. Use MD trajectory to explore the receptor conformational space.

B. Extract several snapshots from the trajectory; clustering to eliminate redundancy.

C. From each cluster, select a representative structure (e.g., medoid).

D. Carry out virtual ligand screening independently at each representative conformation.

E. Return activity predictions by independent runs and combine together in a global ranking.

A) Molecular Dynamics — Structural Features vs Time

B) Cluster Analysis

C) Representatives Selection — Structural Features vs Time

D) Virtual Screening

E) Combined Ranking

✓ Improved Hit Retrieval

✓ Increased Diversity among Hits

Rank	Molecule	Score
1	CMP_390	-43.2
2	CMP_185	-40.7
3	CMP_71	-39.6
4	CMP_13	-39.3
5	CMP_26	-38.1
6	CMP_45	-36.9

Results from Individual Runs

Rank	Molecule	Score
	71	-39.6
	13	-38.2
	26	-35.8

Rank	Molecule	Score
	390	-43.2
	185	-40.7
	13	-39.3

Rank	Molecule	Score
1	CMP_26	-38.1
2	CMP_45	-36.9
3	CMP_390	-35.4

De Vivo (2016) J. Med. Chem.
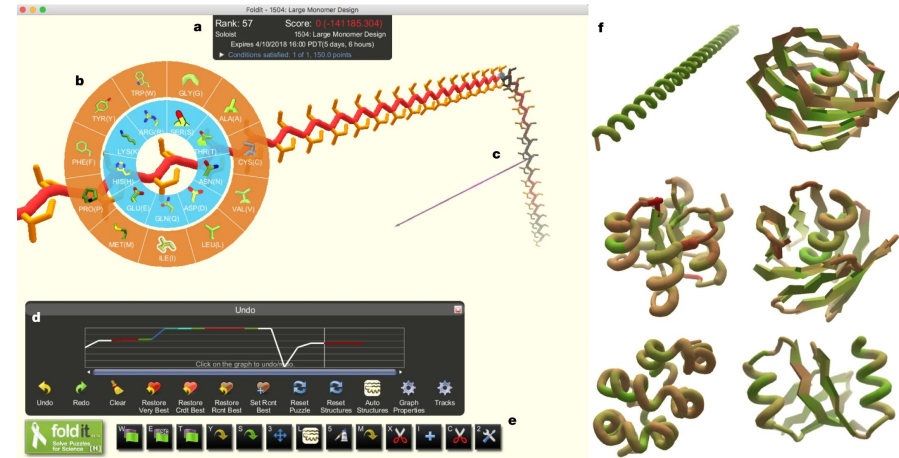
Distributed computing & Crowdsourcing

Folding@home: foldingathome.org

- Distributed computing project for MD simulations (e.g., protein folding, computational drug design).

- Uses the idle resources of personal computers owned by volunteers from all over the world.

- https://foldingathome.org/2020/02/27/foldinghome-takes-up-the-fight-against-covid-19-2019-ncov/

Foldit: fold.it

- An online game that poses complex puzzles about how proteins fold.

- Helped solve the structure of a protein-sniping enzyme critical for reproduction of the AIDS virus within 3 weeks; Identified targets for drugs to neutralize it.