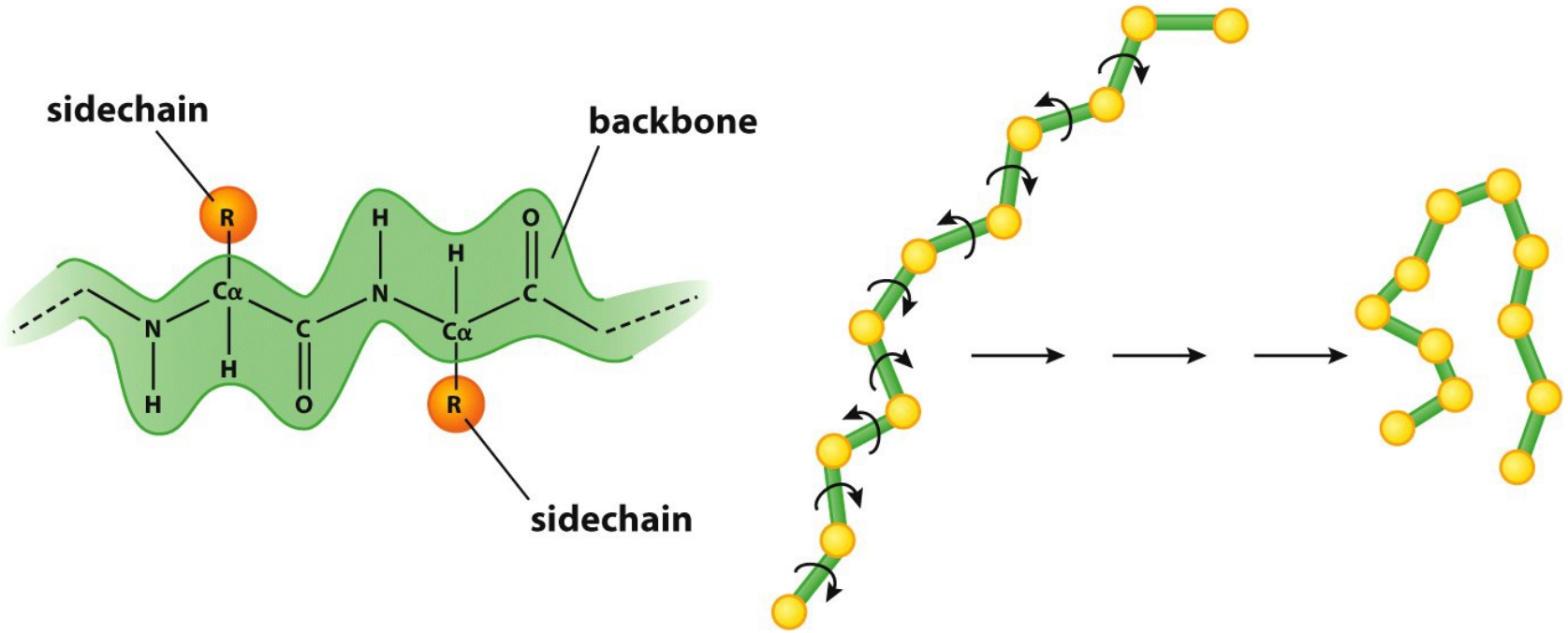
Searching

efficient sampling of protein conformations

1. conformational degrees of freedom

2. search methods

Protein folding involves rotations of the peptide backbone

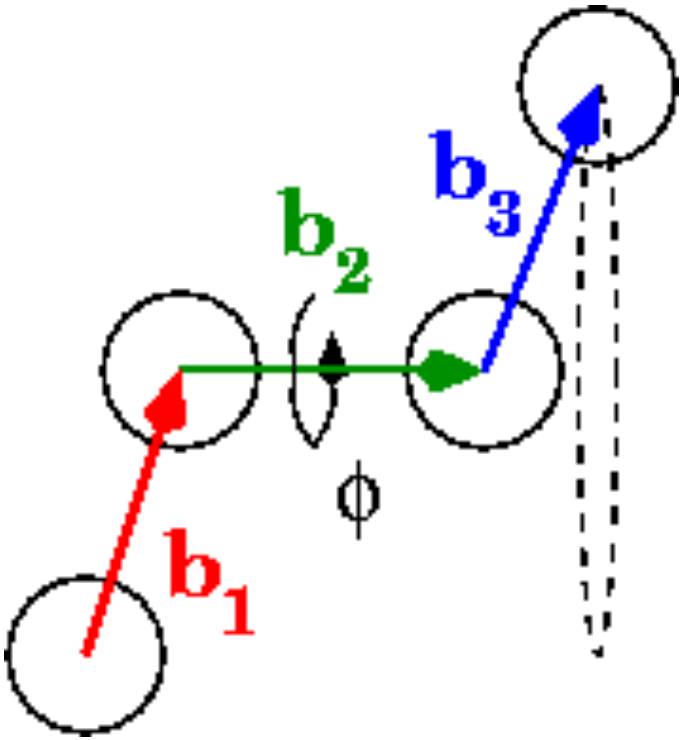


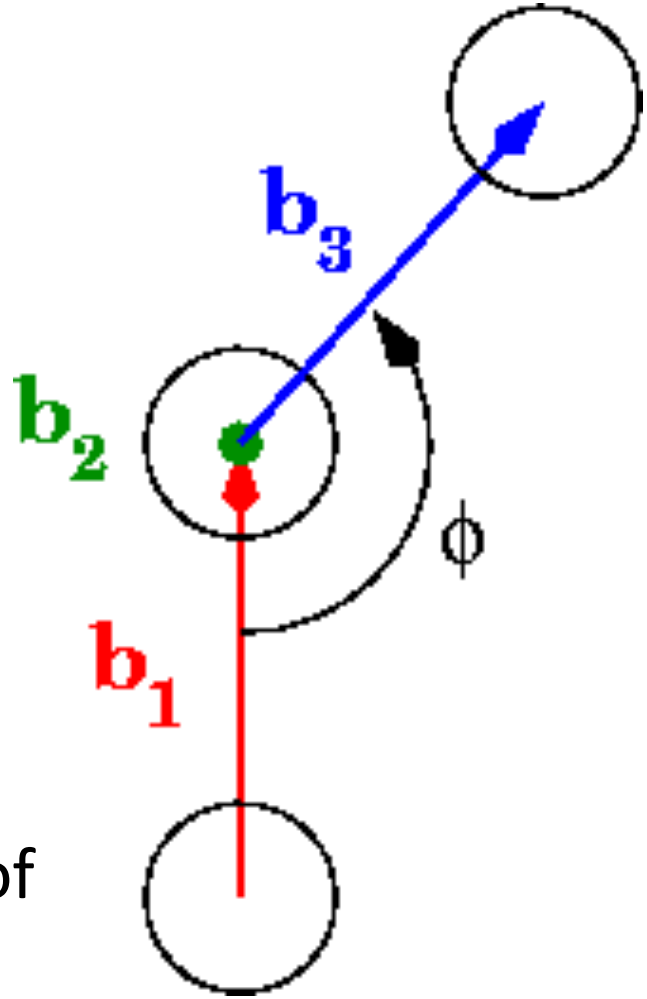
Approximation:

Peptide bond length and angles do not change

Peptide dihedral angles define the structure

Dihedral angles

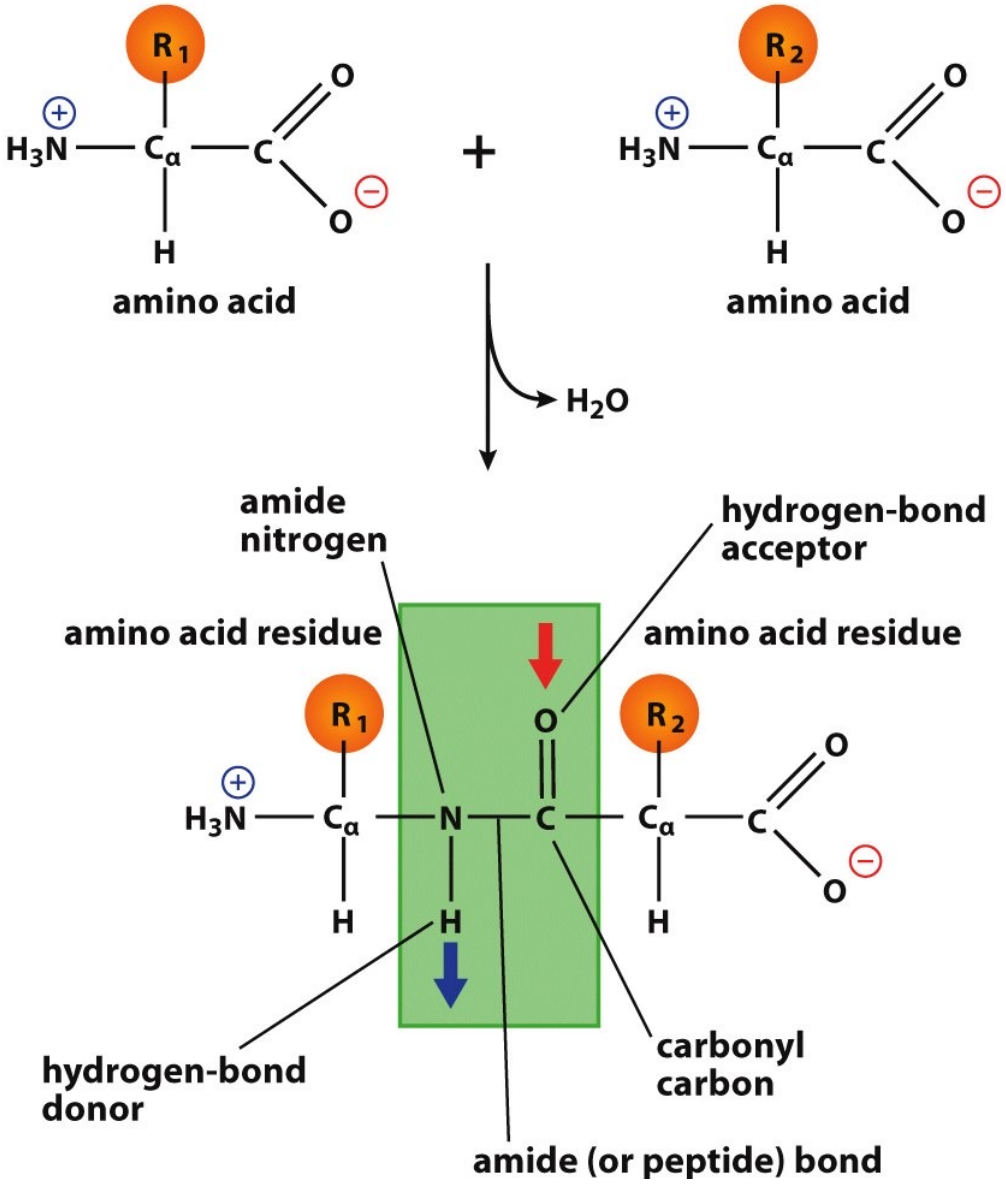




- Dihedral angle: defines geometry of 4 consecutive atoms (given bond lengths and angles)

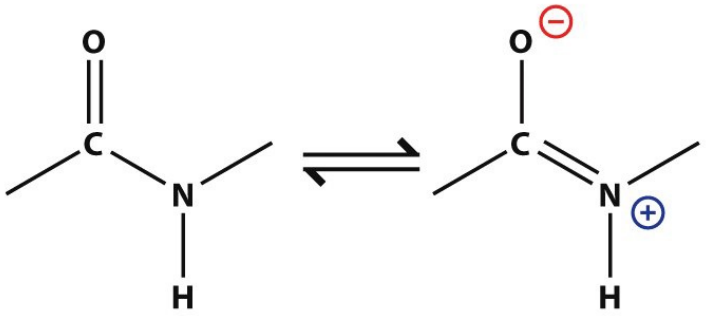
From wikipedia

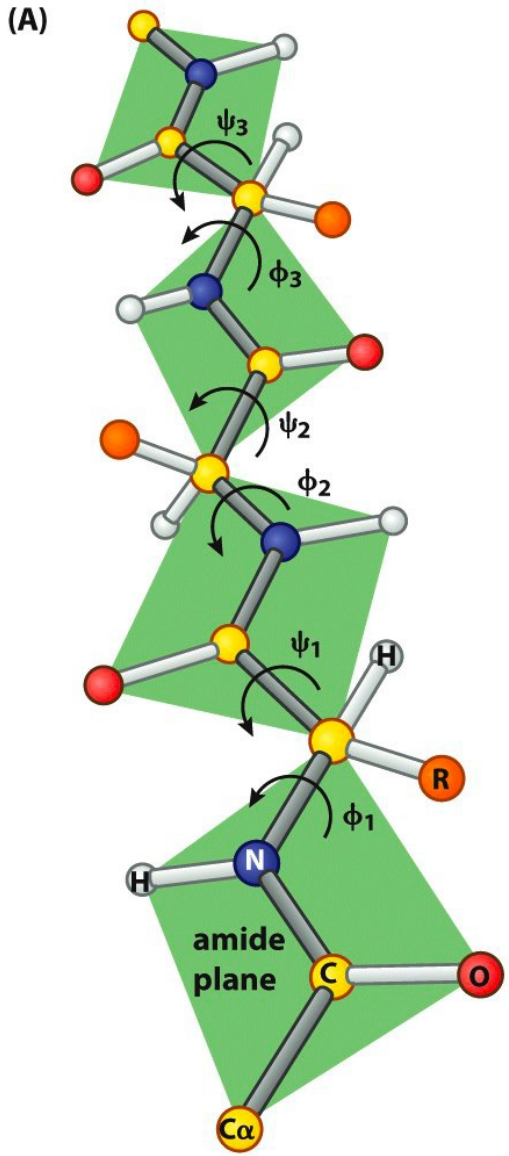
Formation of a peptide bond

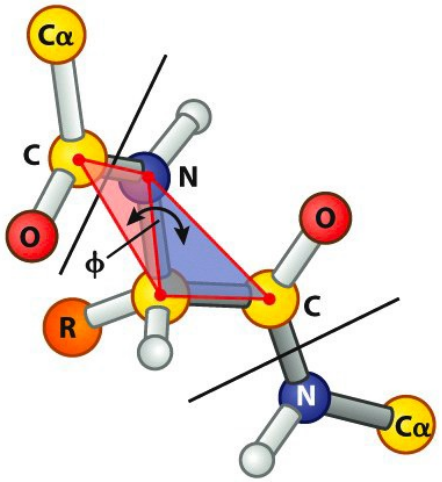


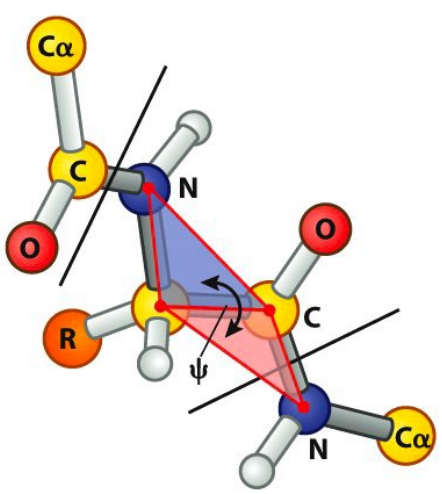
Dihedral angles Φ and Ψ define backbone conformation

The peptide bond
is **planar** and **polar**



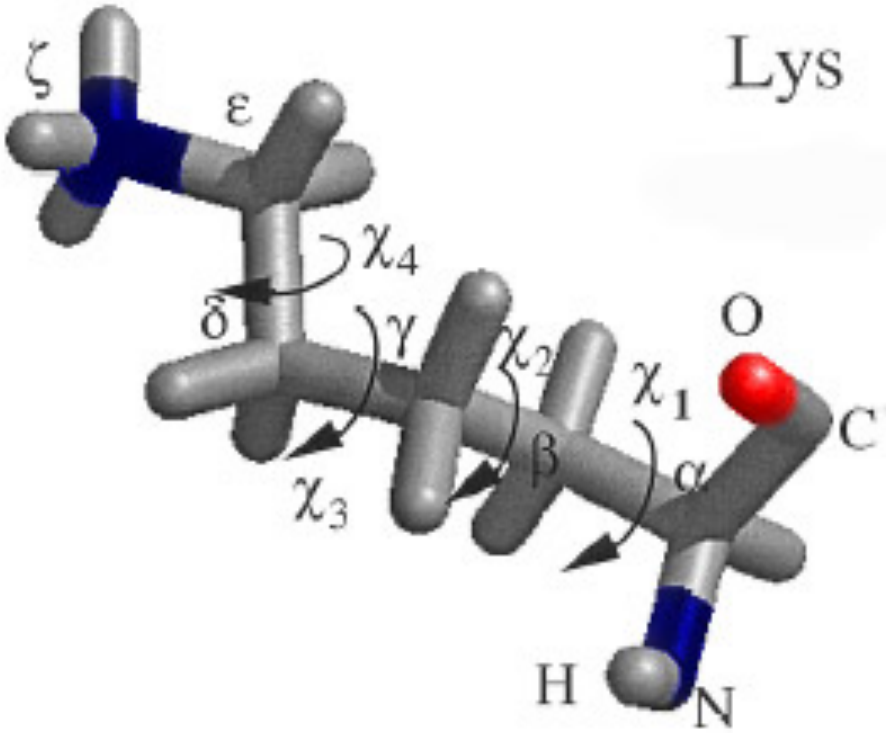


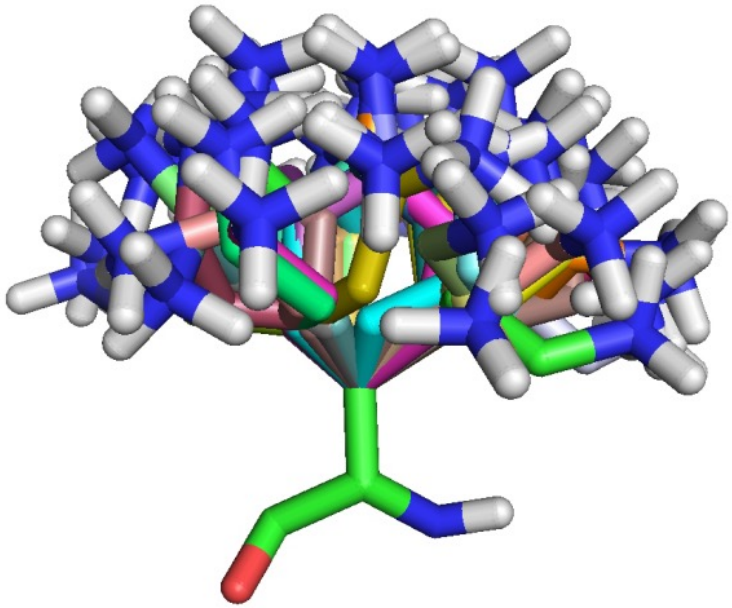




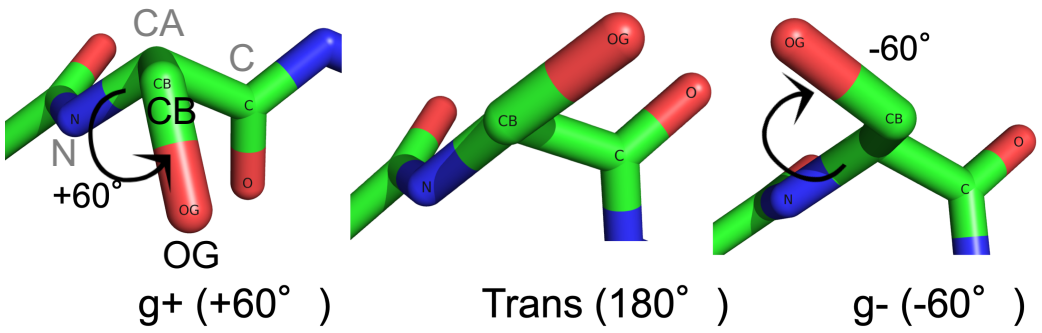
Amino acid side chains can adopt multiple conformations

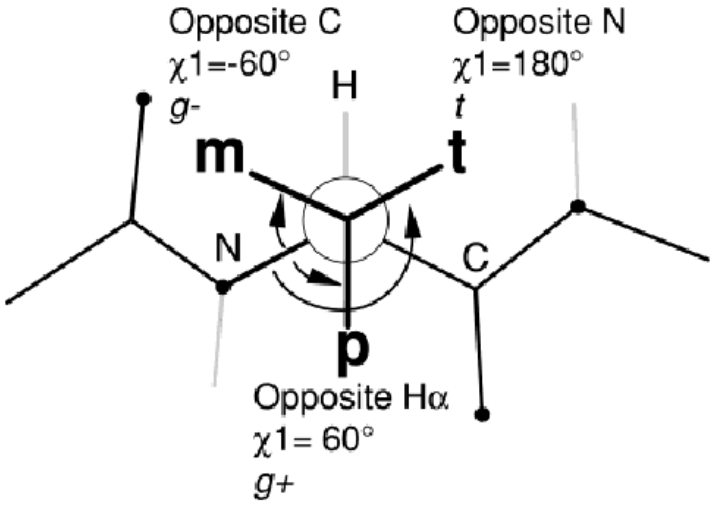
Dihedral angles χ_1–χ_4 define side chain geometry

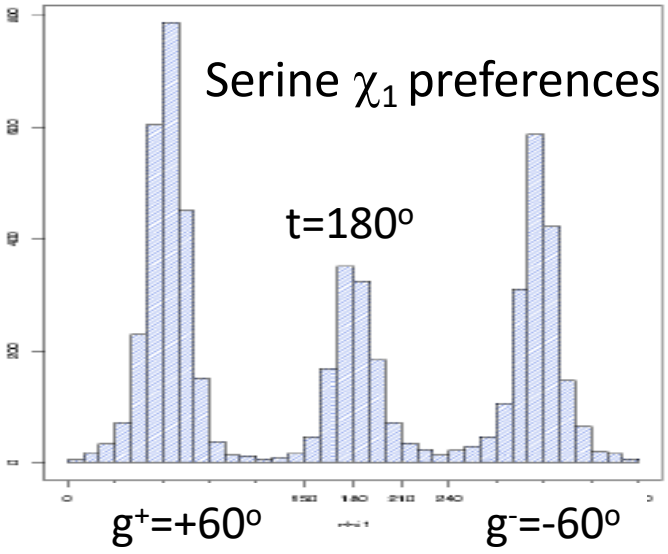




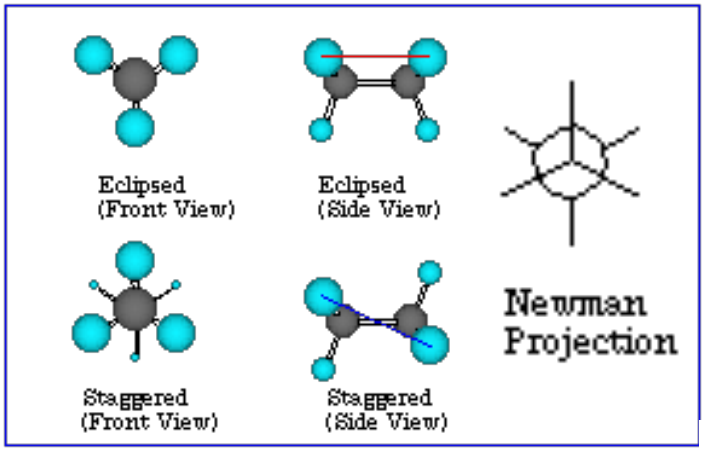
Side chains assume discrete conformations







Staggered conformations minimize collision with neighboring atoms



Lovell, 2000

Rotamer libraries contain preferred conformations

Rotamer: discrete side chain conformation defined by χ_1–χ_4

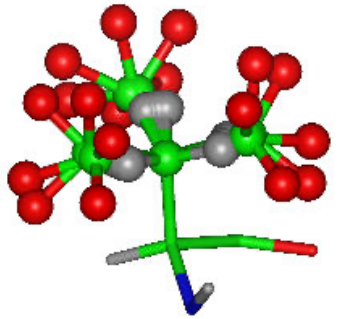


Table 1

Published rotamer libraries.

Authors	Year	Type of library	Number of proteins in library	Resolution (Å)
Chandrasekaran and Ramachandran [2]	1970	BBIND	3	NA
Janin *et al.* [4]	1978	BBIND, SSDEP	19	2.5
Bhat *et al.* [3]	1979	BBIND	23	NA
James and Sielecki [5]	1983	BBIND	5	1.8, R-factor <0.15
Benedetti *et al.* [6]	1983	BBIND	238 peptides	R-factor <0.10
Ponder and Richards [7]	1987	BBIND	19	2.0
McGregor *et al.* [8]	1987	SSDEP	61	2.0
Tuffery *et al.* [9]	1991	BBIND	53	2.0
Dunbrack and Karplus [10]	1993	BBIND, BBDEP	132	2.0
Schrauber *et al.* [11]	1993	BBIND, SSDEP	70	2.0
Kono and Doi [12]	1996	BBIND	103	NA
De Maeyer *et al.* [13]	1995	BBIND	19	2.0
Dunbrack and Cohen [14]	1997–2002	BBIND, BBDEP	850*	1.7
Lovell *et al.* [15••]	2000	BBIND, SSDEP	240	1.7
Shapovalov and Dunbrack*	2011	BBDEP	3854	1.8

** Shapovalov & Dunbrack, Structure 2011*

Dunbrack, 2002

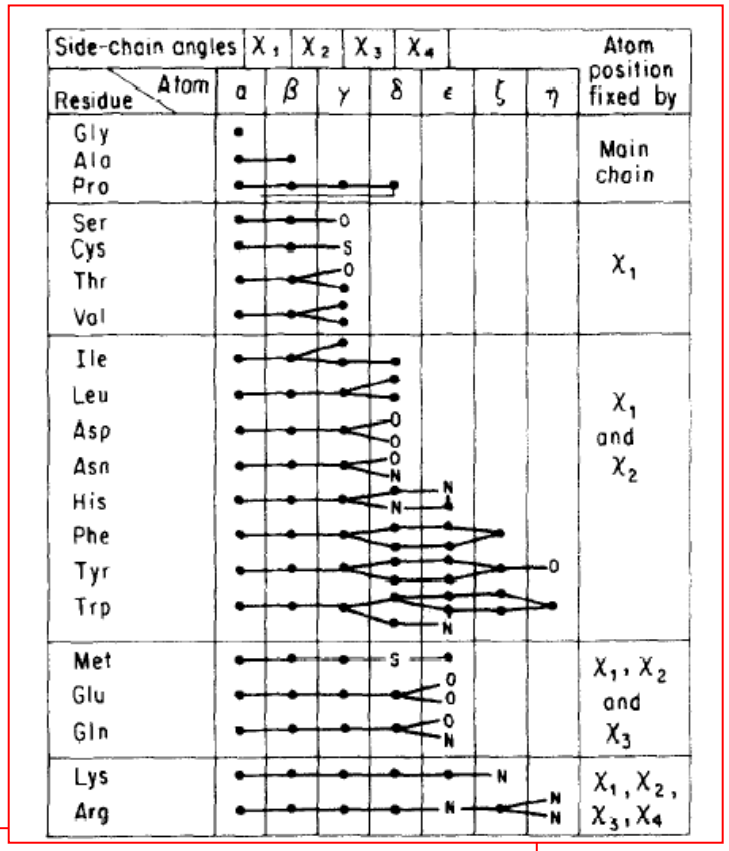
Representative rotamer libraries are surprisingly small

Ponder & Richards, 1987:

Analysis of ~20 proteins

(~2000 side chains)

67 rotamers can adequately represent side chain conformations (for 17/20aa)



Isoleucine				
− t	42	45·2	−60·9 (7·5)	168·7 (11·6)
− −	17	18·3	−59·6 (9·6)	−64·1 (14·3)
+ t	15	16·1	61·7 (5·0)	163·8 (16·4)
t t	12	12·9	−166·6 (10·1)	166·0 (8·9)
t +	3	3·2	−174·8 (24·9)	72·1 (10·5)
Other	4	4·3		

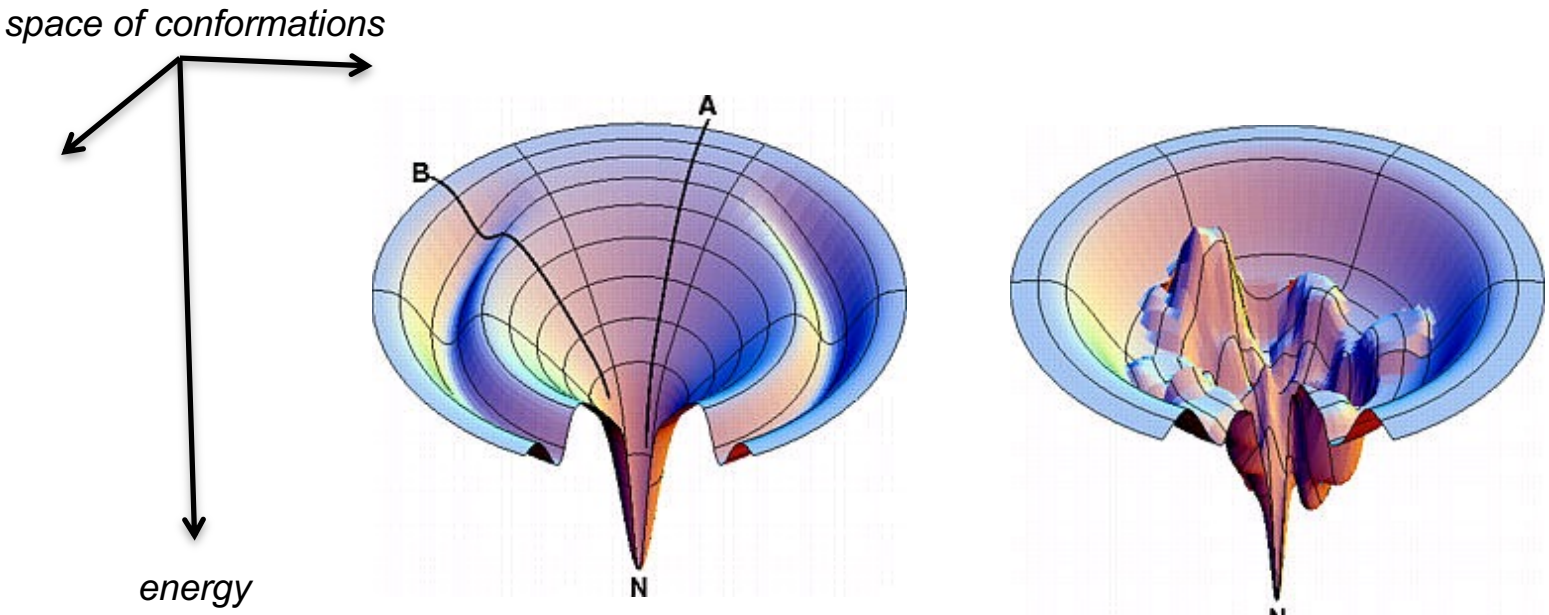
Searching

efficient sampling of protein conformations

1. conformational degrees of freedom

2. search methods

Characteristics of the Protein Conformational Energy Landscape

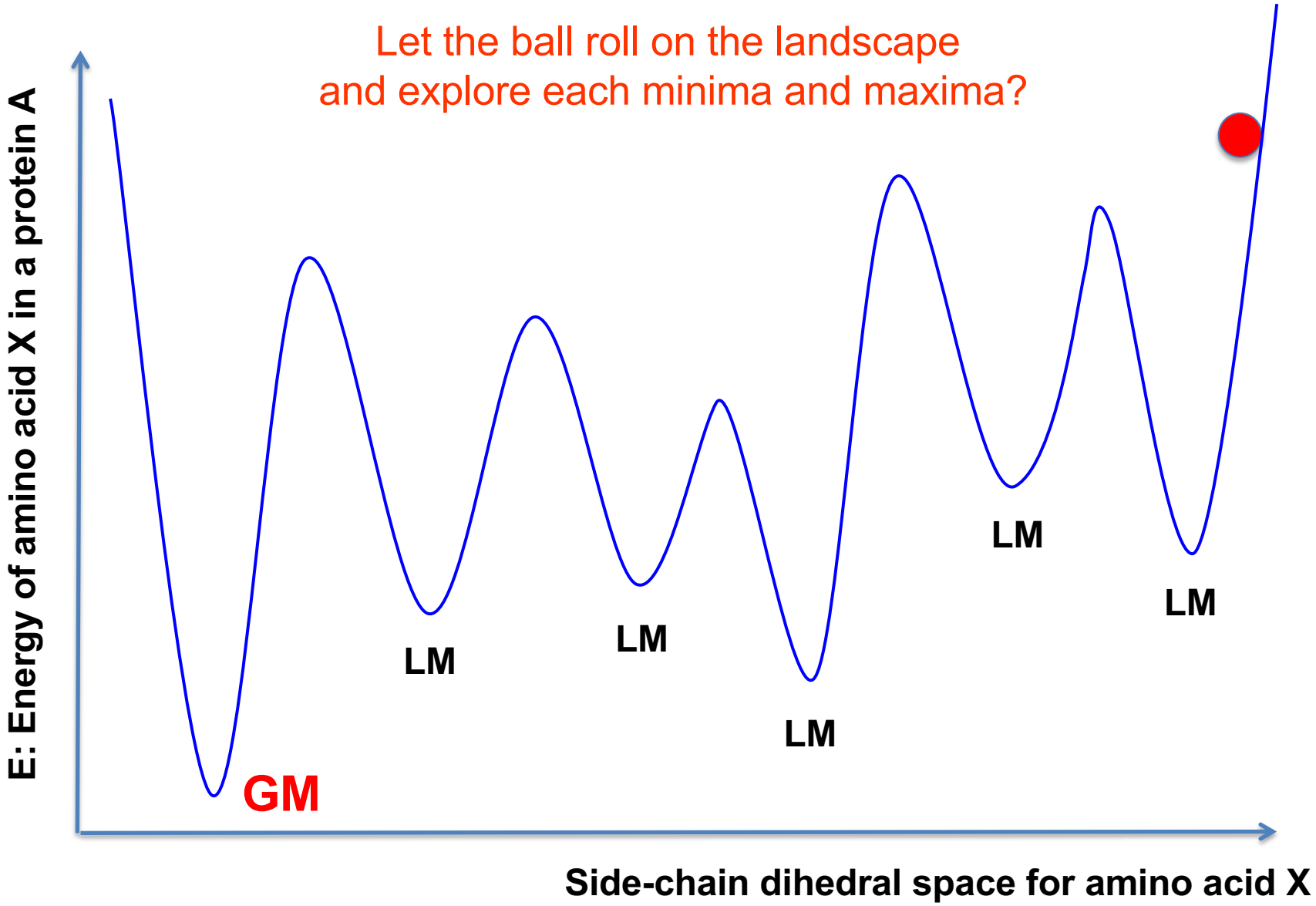


smooth?

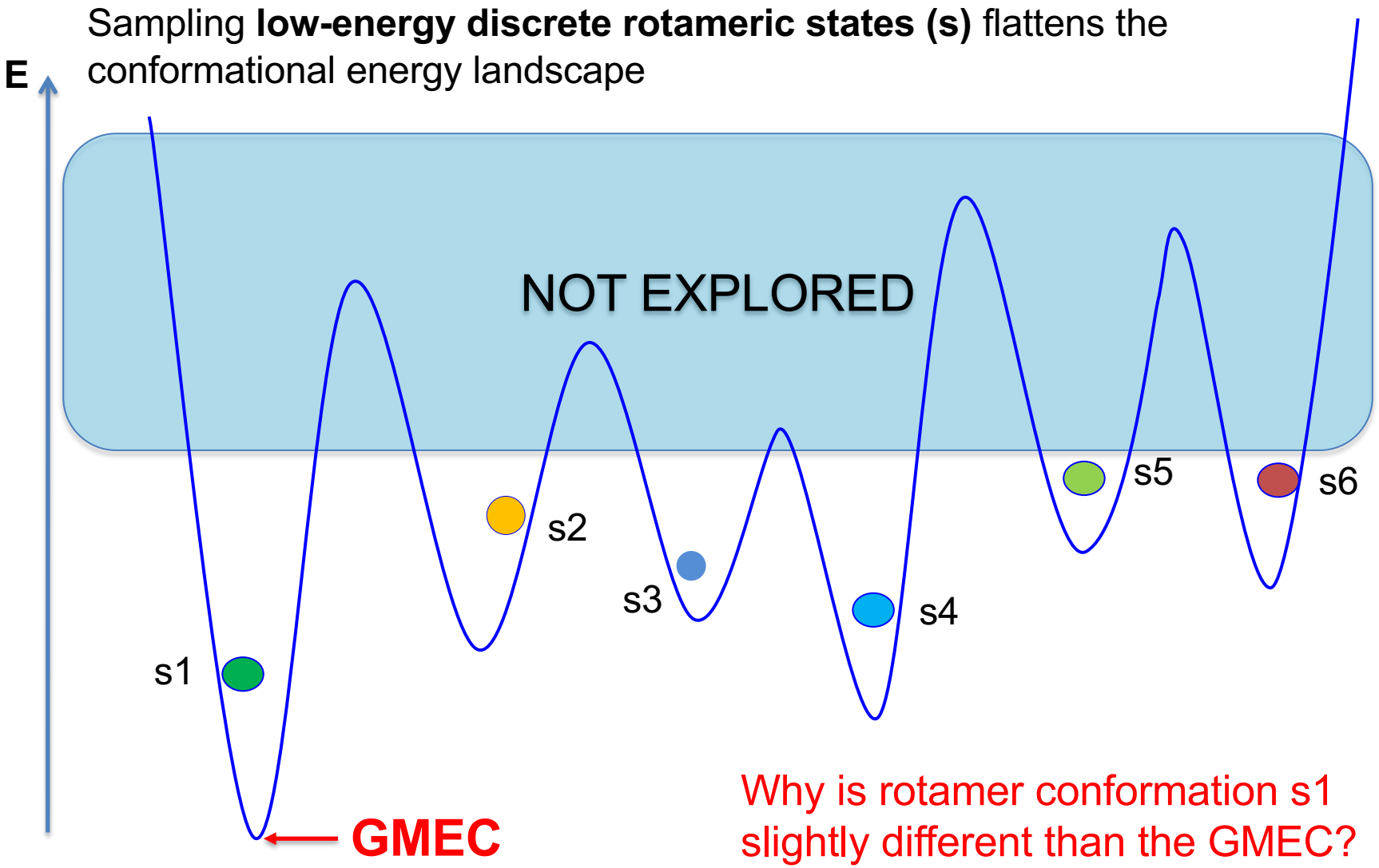
rugged?

Images by Ken Dill

The Problem: Find Local (LM) and Global Minima (GM) on a Rugged **One** Dimensional Surface

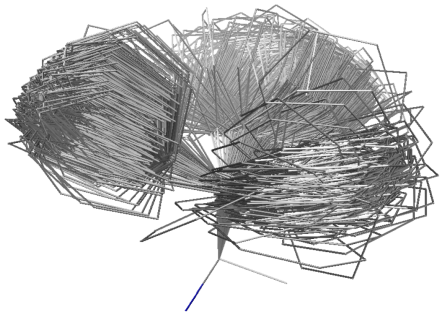


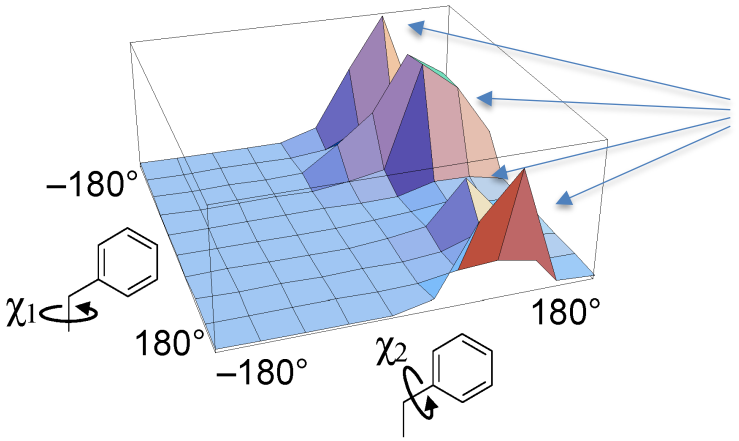
Sampling low-energy discrete rotameric states simplifies the search in conformational space and allows rapid exploration of local minima



Rotamer conformation accuracy

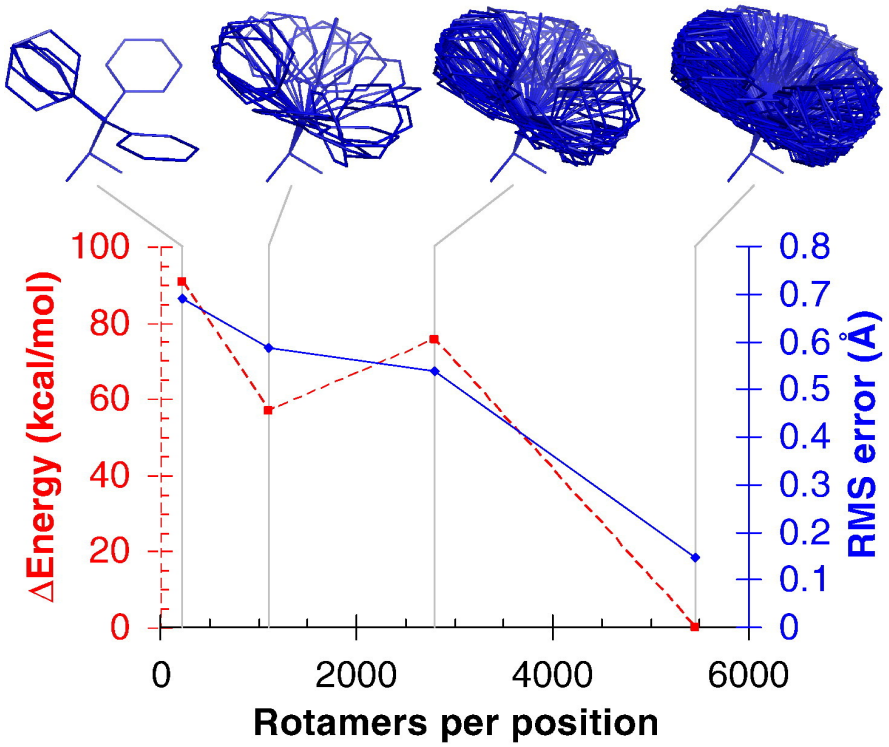
PDB





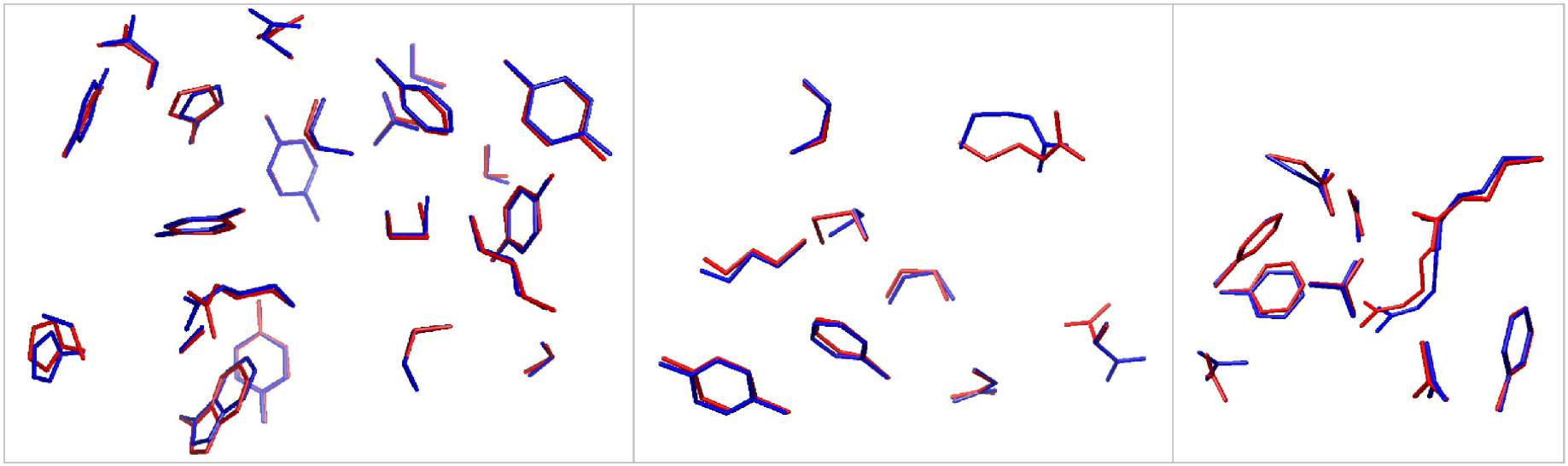
4 major rotameric states

Effect of the resolution of the rotamer library on the conformation prediction accuracy



Boas & Harbury , 2007

Rotamers are an approximation to the true side-chain conformation



bevacizumab-VEGF
RMS error: 0.621 Å

VEGF (unbound)
RMS error: 1.11 Å

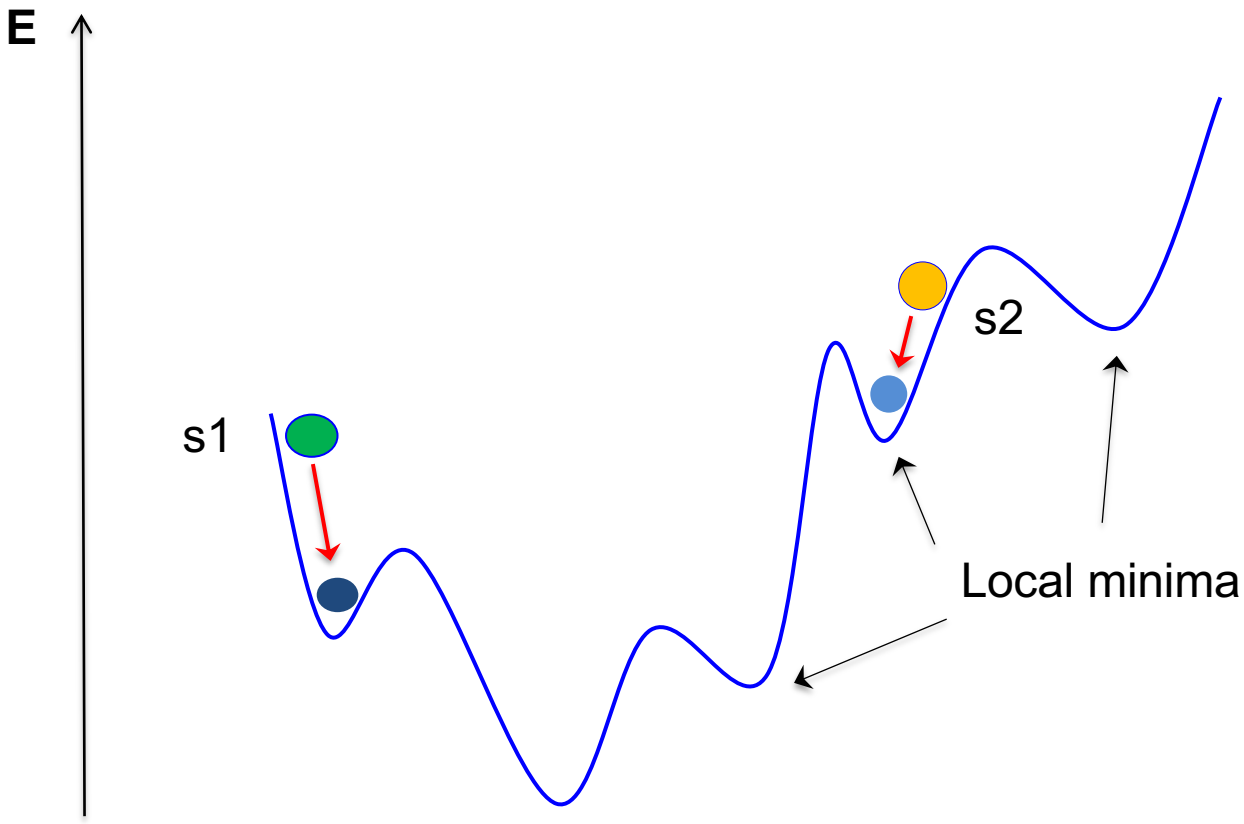
RBP (unbound)
RMS error: 0.483 Å

Crystal structure / **Predicted structure**

Boas & Harbury , 2007

Sampling method: Finding near-by local minima

- Derivative-based methods (Gradient Descent, Newton's method, DFP) are excellent at ***finding*** near-by local minima

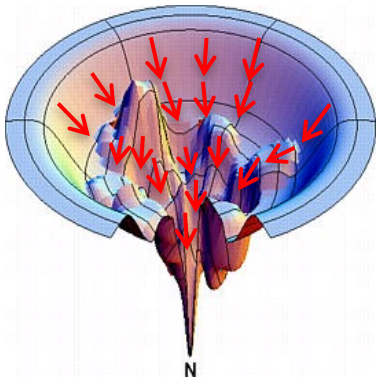


Gradients and Hessians

Gradients and Hessians generalize the first and second derivatives (respectively) of multi-variate scalar functions (= functions from vectors to scalars)

Energy = f(x_1, y_1, z_1, … , x_n, y_n, z_n)

$$\vec{\nabla}_i = \frac{\partial E}{\partial \vec{r}_i} = \begin{pmatrix} \frac{\partial E}{\partial x_i} \\ \frac{\partial E}{\partial y_i} \\ \frac{\partial E}{\partial z_i} \end{pmatrix}$$



$$\mathbf{h}ij = \frac{\partial^2 E}{\partial \vec{r}_i \partial \vec{r}_j} = \begin{pmatrix} \frac{\partial^2 E}{\partial x_i \partial x_j} & \frac{\partial^2 E}{\partial x_i \partial y_j} & \frac{\partial^2 E}{\partial x_i \partial z_j} \\ \frac{\partial^2 E}{\partial y_i \partial x_j} & \frac{\partial^2 E}{\partial y_i \partial y_j} & \frac{\partial^2 E}{\partial y_i \partial z_j} \\ \frac{\partial^2 E}{\partial z_i \partial x_j} & \frac{\partial^2 E}{\partial z_i \partial y_j} & \frac{\partial^2 E}{\partial z_i \partial z_j} \end{pmatrix}$$

Gradient

Hessian

Analytical Energy Gradient

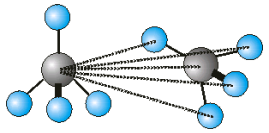
(i) Cartesian Coordinates

$$E = f(x_1, y_1, z_1, \ldots, x_n, y_n, z_n)$$

$$\vec{\nabla} = \left(\frac{\partial E}{\partial x_1} \frac{\partial E}{\partial y_1} \frac{\partial E}{\partial z_1} \cdots \frac{\partial E}{\partial x_n} \frac{\partial E}{\partial y_n} \frac{\partial E}{\partial z_n} \right)$$

Example:

Van der-Waals energy between pairs of atoms – $O(n^2)$ pairs:

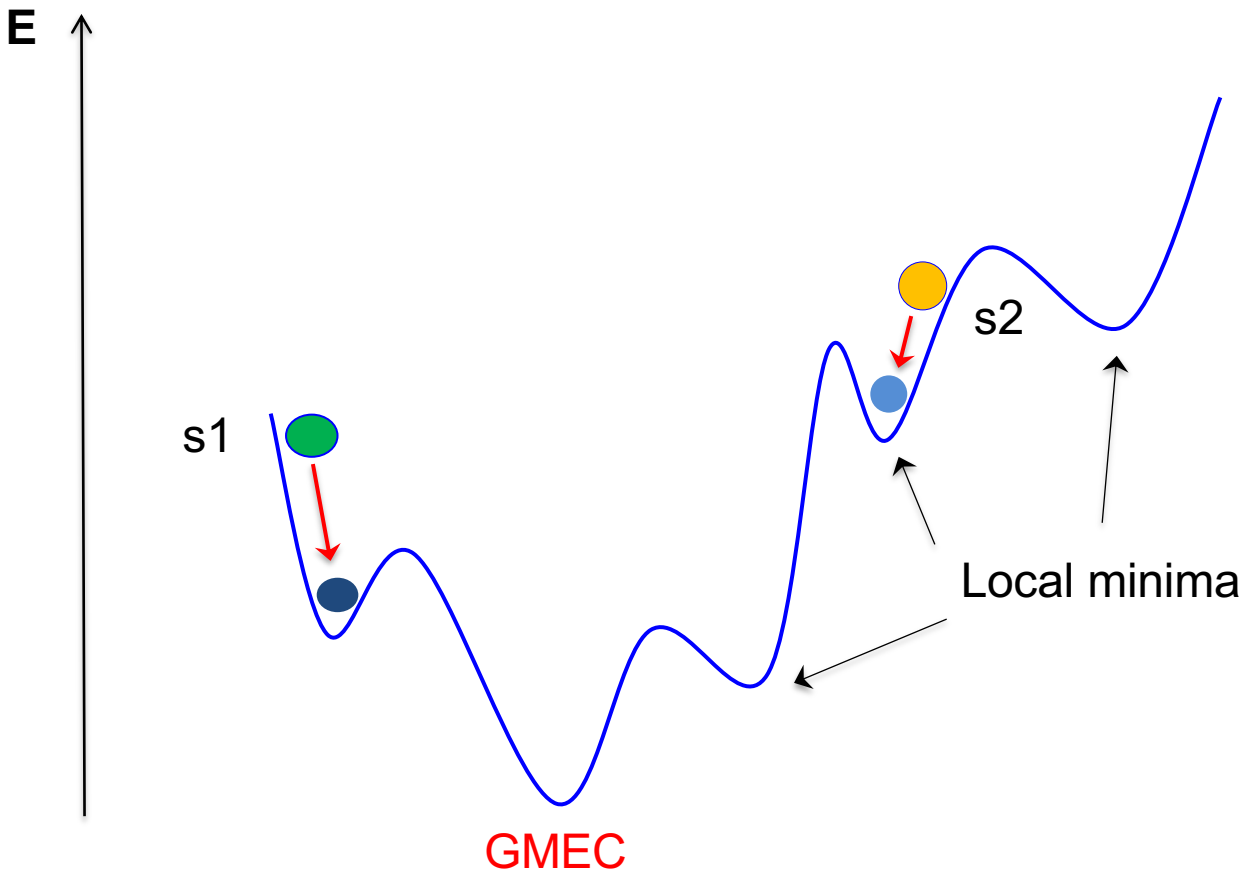


$$E_{VdW} = \sum_{i,j} \frac{A}{\vec{R}_{ij}^{12}} - \frac{B}{\vec{R}_{ij}^{6}} \Rightarrow \frac{\partial E_{VdW}}{\partial R_{ij}} = \frac{-12A}{R_{ij}^{13}} + \frac{6B}{R_{ij}^{7}}$$

$$R_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

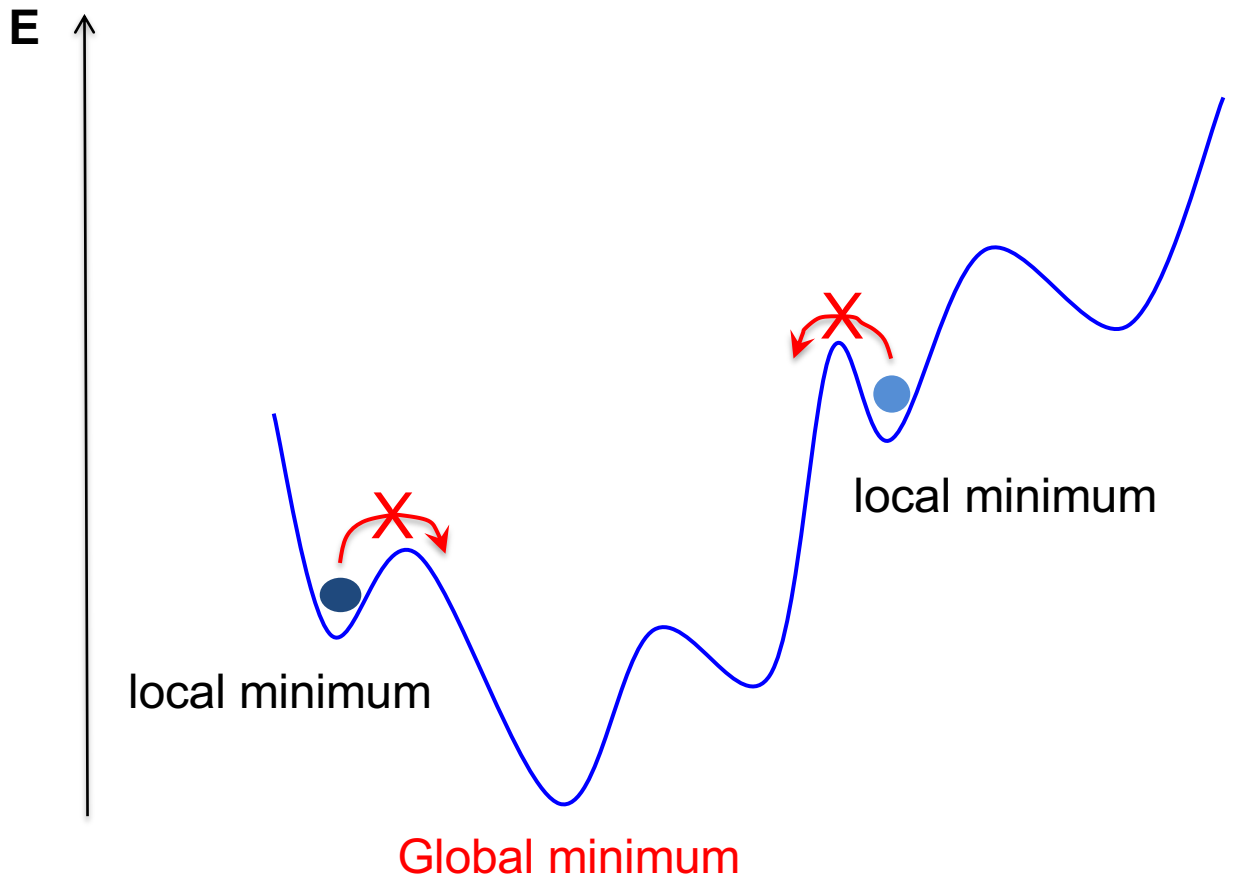
Sampling method: Finding near-by local minima

How can we now find the GMEC?



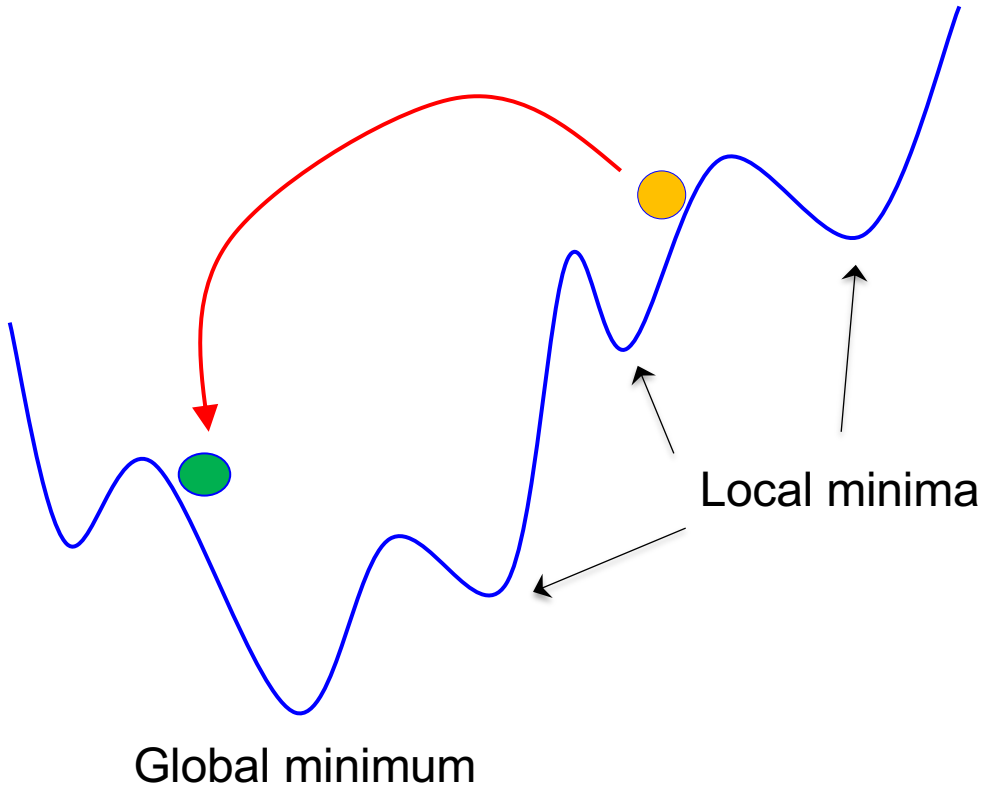
Sampling method: Finding Global minima

- Derivative-based methods ***cannot escape*** near-by local minima



Sampling method: jumping between discrete minima

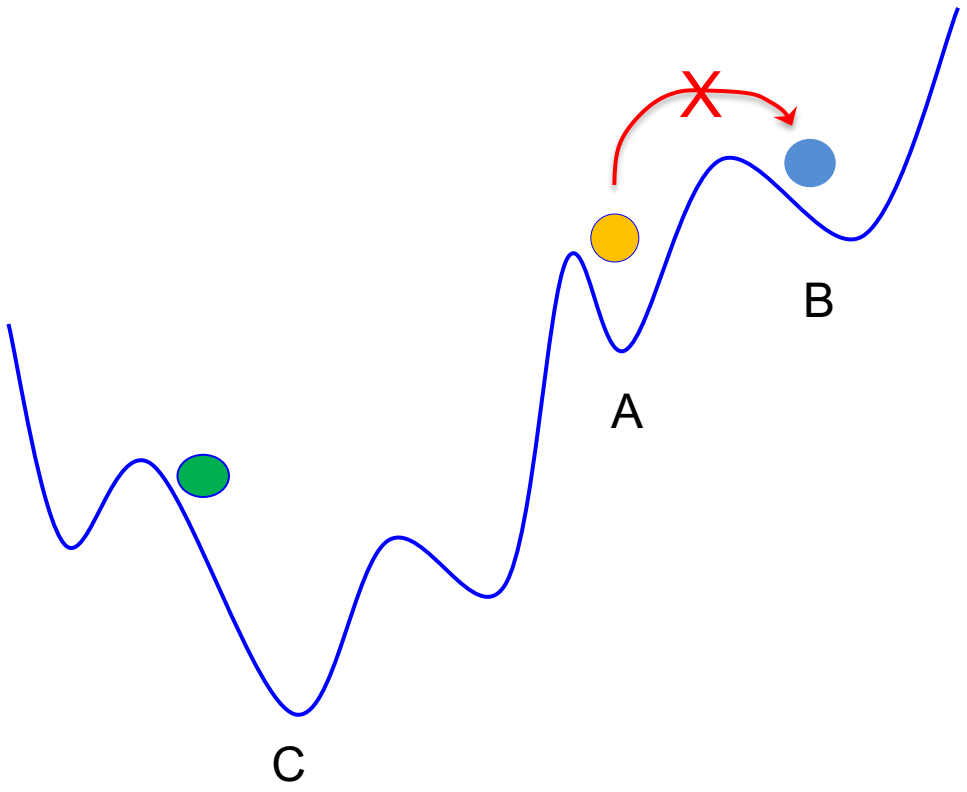
- Several algorithms can be used to ***jump*** between discrete ***near-by local minima*** (e.g. Monte Carlo, Genetic algorithm, Self consistent Mean Field, Dead End Elimination)



Monte Carlo: stochastic sampling of discrete minima

Basic steps:

Random move: A->B
$E_B < E_A$?
No: move rejected

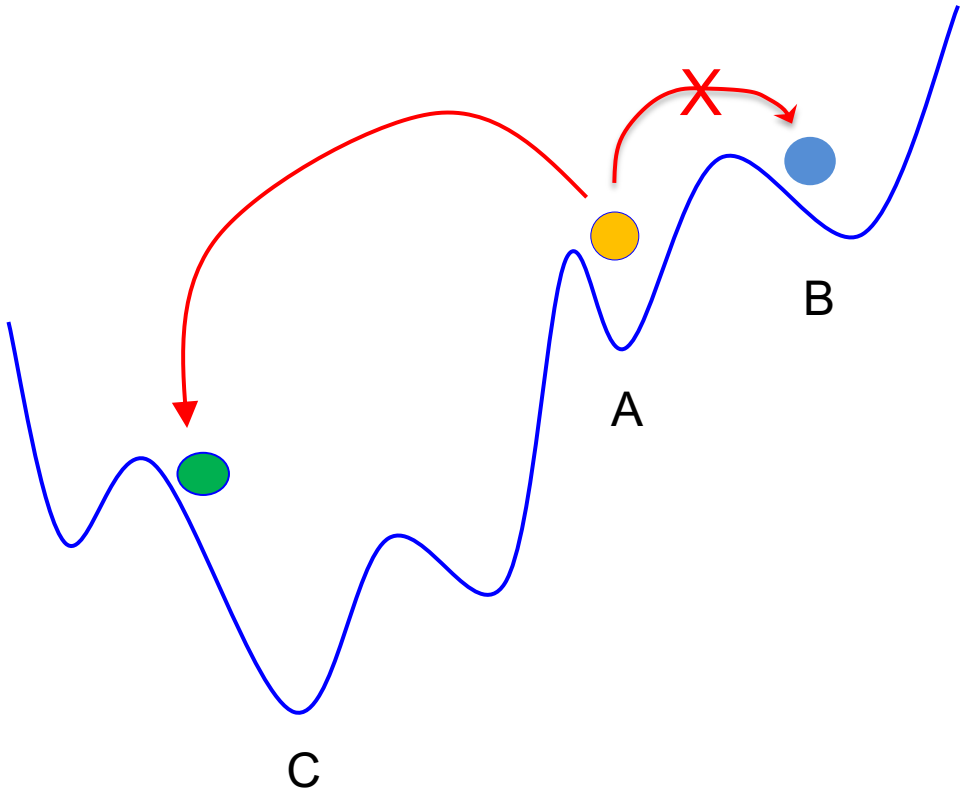


Monte Carlo: stochastic sampling of discrete minima

Basic steps:

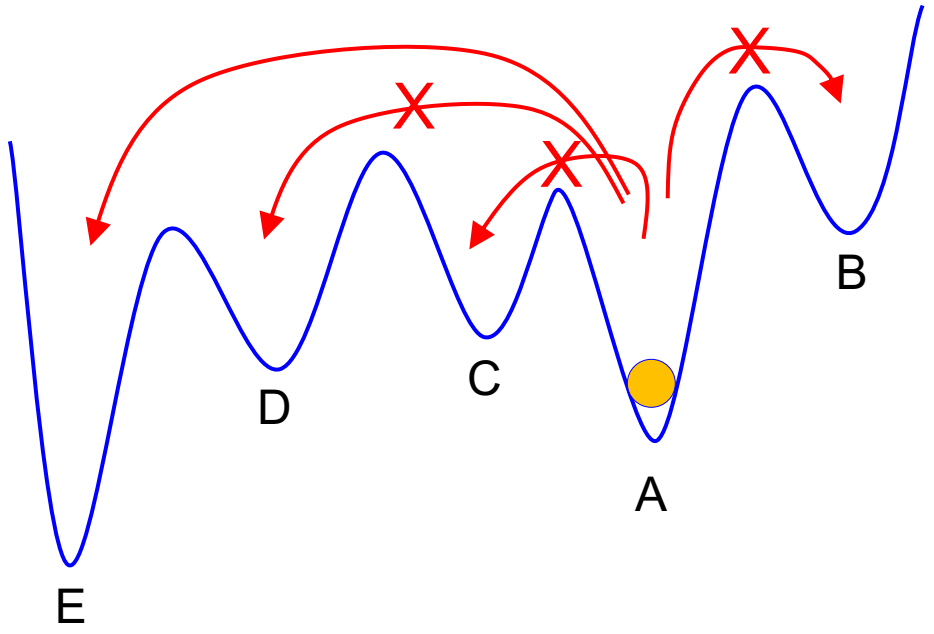
Random move: A->B
$E_B < E_A$?
No: move rejected

Random move: A->C
$E_C < E_A$?
Yes: move accepted



Metropolis Monte Carlo

Very low probability
to move out of A
and reach the
Global Minimum E



Acceptance with probability $P = e^{-\Delta E / k_B T}$ (Boltzmann weight), $\Delta E = E_B - E_A$

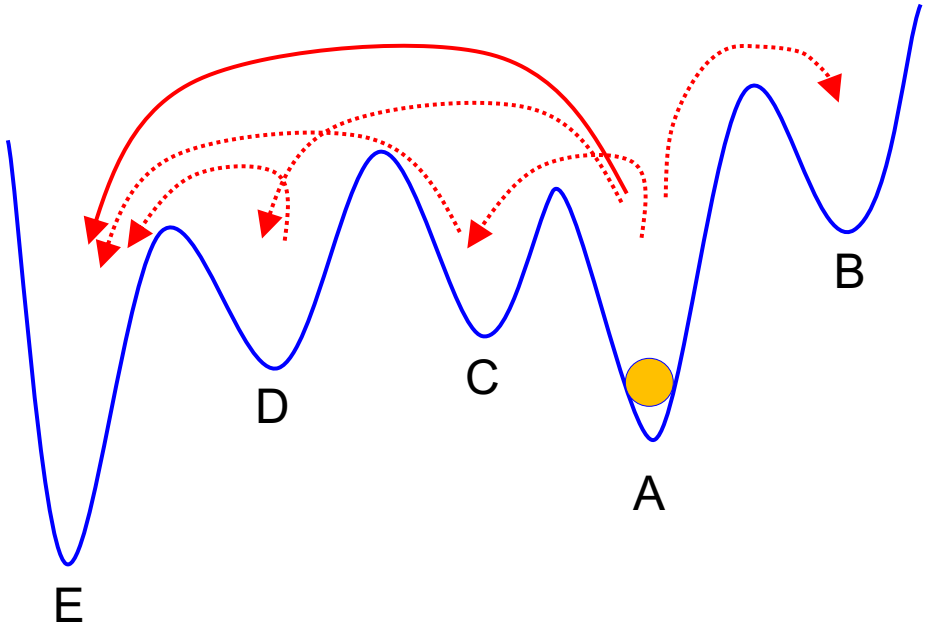
P = 1 if $E_B < E_A$

P = 1 if $E_B > E_A$ && $e^{-\Delta E / k_B T} \geq rand(0,1)$

P = 0 if $E_B > E_A$ && $e^{-\Delta E / k_B T} \leq rand(0,1)$

Metropolis Monte Carlo

Probability P
to move out of A
depends on the
temperature T



$$P = e^{-\Delta E / k_B T}$$

P = 1 if $E_B < E_A$

P = 1 if $E_B > E_A$ && $e^{-\Delta E / k_B T} \geq rand(0,1)$

P = 0 if $E_B > E_A$ && $e^{-\Delta E / k_B T} \leq rand(0,1)$

Metropolis Monte Carlo with Simulated Annealing

The Boltzman distribution depends on the *in-silico* temperature **T**:

- At low temperatures, we will get stuck in local minima (we will always get zero acceptance if the energy rises even slightly)
- At high temperatures, acceptance will always be 1 (always jump between conformations regardless of their energies).

P = 1 if Ej > Ei && $e^{-\Delta E / k_B T} \geq rand(0,1)$

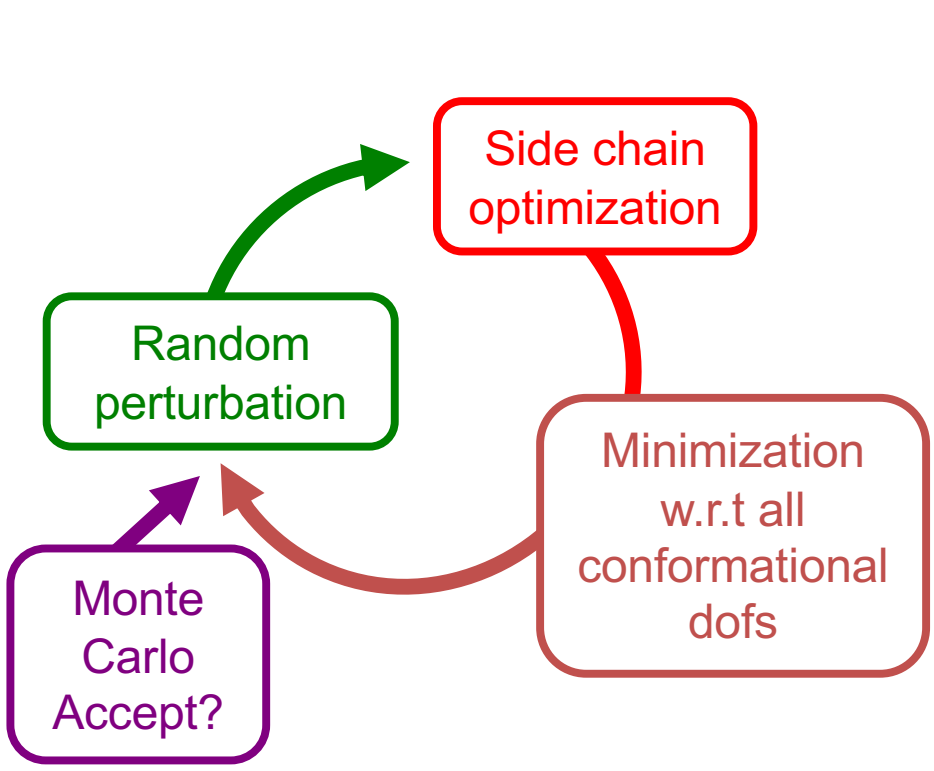
P = 0 if Ej > Ei && $e^{-\Delta E / k_B T} \leq rand(0,1)$

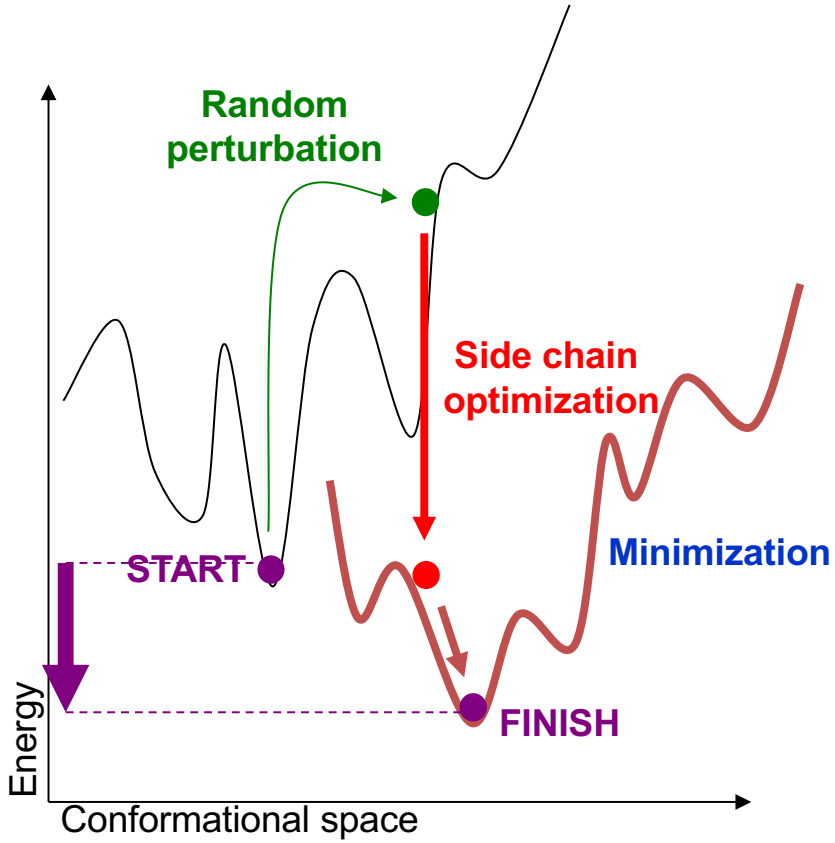
In simulated annealing, we gradually decrease ("cool down") the virtual temperature factor, until we converge to a minimum point

Basic design cycle step in Rosetta

Start: target structure

Random perturbation: amino acid substitution



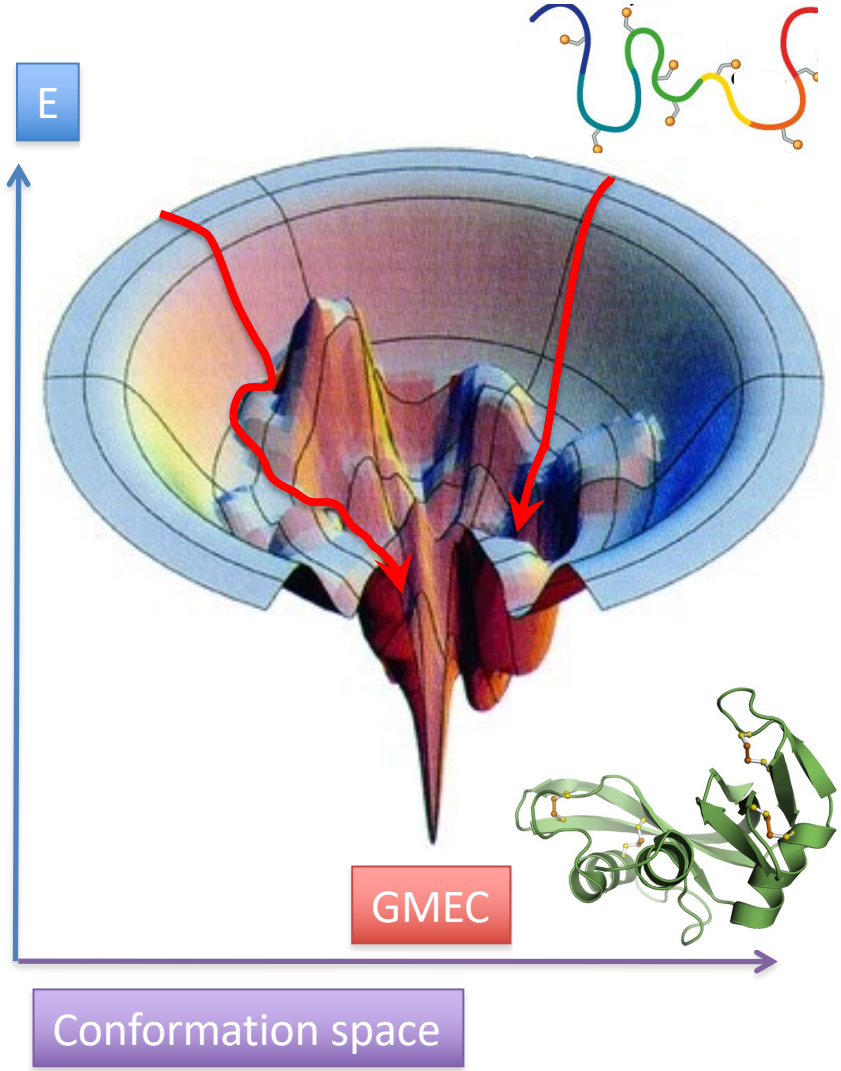


Take home message

Computational protein design is challenging but made possible thanks to:

1. An energy function to rapidly rank sequences

2. An efficient search technique to find the GMEC



Main points to remember

1. Protein design challenges: explore astronomically large space of possibilities

Approximations to accelerate physics-based protein design calculations based on:

2. Simple description physical interactions

3. Discretization of conformational space

4. Efficient continuous and stochastic search technique to find the GMEC