

生物分子计算模拟简介

要点:

- 1、描述分子体系的力场模型
- 2、分子构象分析方法
- 3、分子动力学模拟简介
- 4、分子建模与模拟的典型应用场景

稳定生物分子结构的主要作用力

1. **Hydrogen Bond**
2. **van der Waals interaction**
3. **Electrostatic interaction**
4. **Dipole-dipole interaction**
5. **π - π stacking**
6. **Hydrophobic interaction
(entropic effects)**

The Nobel Prize in Chemistry 2013



Martin Karplus



Michael Levitt



Arieh Warshel

"for the development of multiscale models for complex chemical systems"

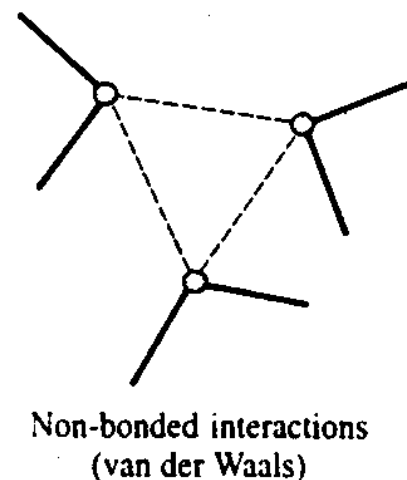
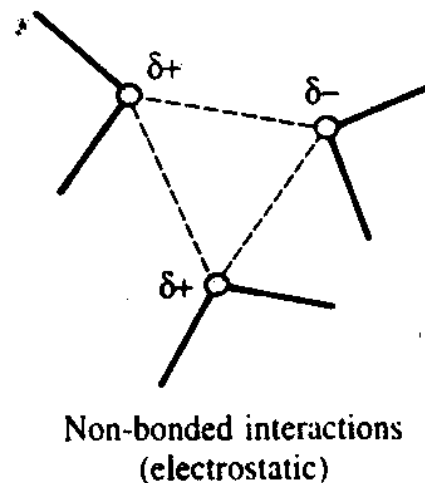
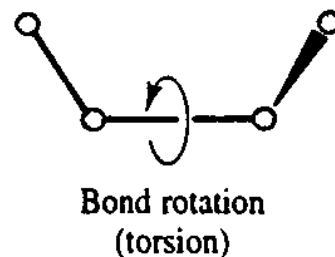
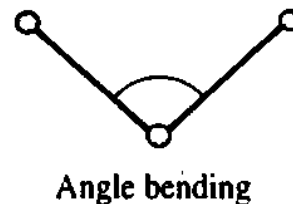
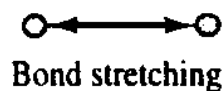
The work of Karplus, Levitt and Warshel is groundbreaking in that they managed to make **Newton's classical physics work side-by-side with the fundamentally different quantum physics.**

Previously, chemists had to choose to use either or. The strength of classical physics was that calculations were simple and could be used to model really large molecules. Its weakness, it offered no way to simulate chemical reactions. For that purpose, chemists instead had to use quantum physics. But such calculations required enormous computing power and could therefore only be carried out for small molecules.

分子力学简介

- 分子的经典力学模型：

- 分子中原子间形成的化学键具有自然的键长、键角，不成键的非键原子间存在范氏及静电作用力。分子会调整自身构象以尽量采取这些数值，给出原子核位置的最佳排布（能量最小态）。



- 分子力学的分子：

- 一套由经典力学势函数支配的原子排布

分子力学的基本假设

The Born-Oppenheimer Approximation –

原子核的运动与电子的运动可以看成是独立的

分子力学的名称

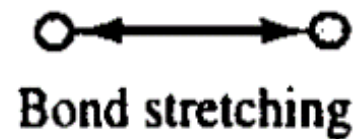
Molecular Mechanics Method

(empirical) force field calculations

Molecular simulations

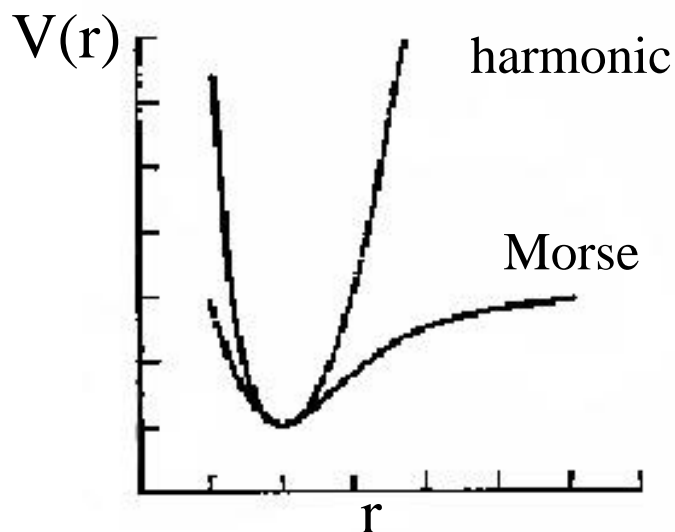
一、分子模拟的力场模型

1. 键能项 - 键的拉伸振动



谐振势函数

$$V = \frac{k_r}{2} (r - r_0)^2$$



非谐函数-加三次方项修正:

$$V = \frac{k}{2} (r - r_0)^2 [1 - k_1 (r - r_0)]$$

2. 角弯折项(Angle Bending)



Angle bending

谐振势函数

$$V = \frac{k_{\theta}}{2} (\theta - \theta_0)^2$$

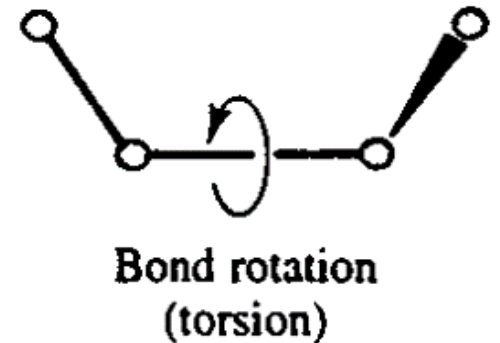
非谐函数-加三次方项修正：

$$V = \frac{k}{2} (\theta - \theta_0)^2 [1 - k_1 (\theta - \theta_0)]$$

小环体系中键长与键角相关项

3. 扭转势垒 (torsion terms)

$$V = \sum_{bonds} \frac{V_n}{2} (1 + \cos(n\varphi - \gamma))$$



V_n : barrier height, 定性表明势垒高度

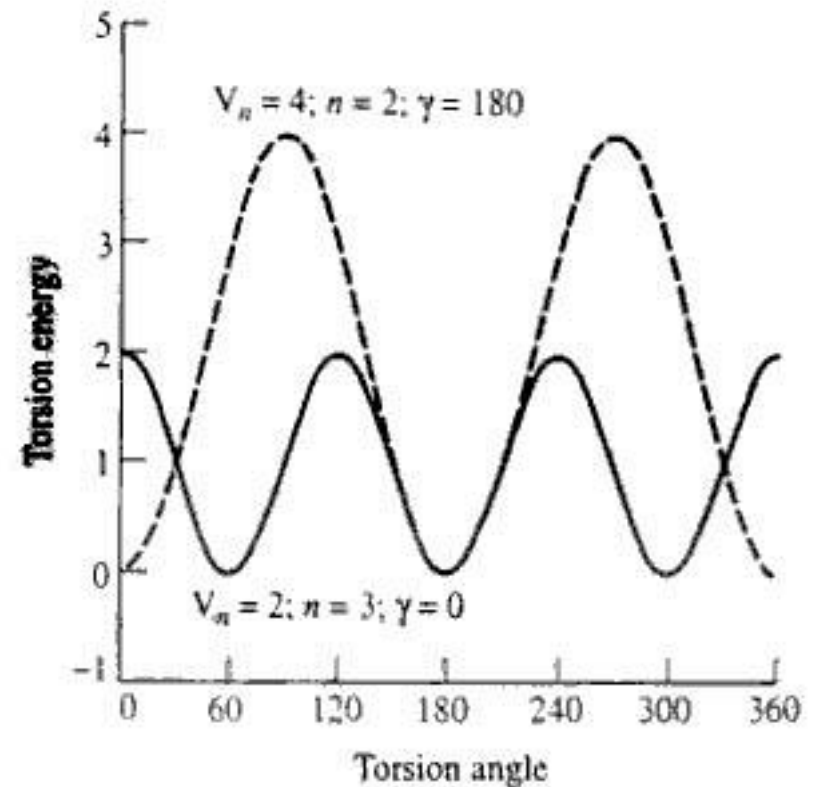
n : multiplicity

γ phase factor

e.g.: C-C single bond, $n=3, \gamma=0$,

minimum: 60, -60, 180

maximum: 0, 120, -120



4. 非键相互作用(Non-bonded Interactions)

(1) van der Waals Interaction

Lennard-Jones function

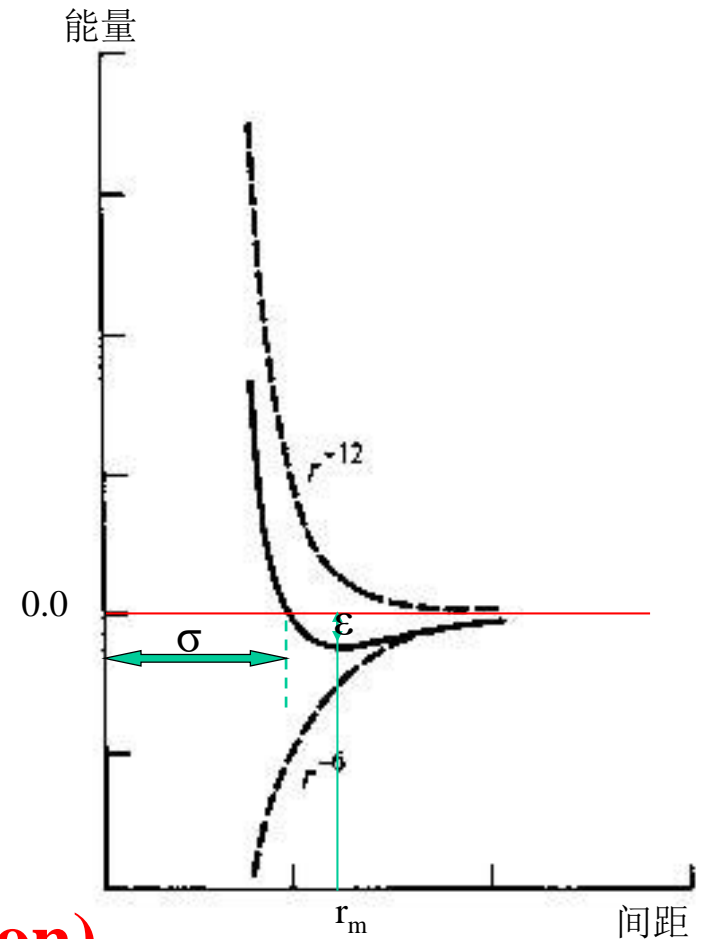
$$V = 4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^6]$$

$$V = \varepsilon[(r_m/r)^{12} - 2(r_m/r)^6]$$

$$V = A/r^{12} - C/r^6$$

近距离时的排斥力(Pauli exclusion)

远距离时的吸引力(London Dispersion force)



(2) 静电相互作用(Electrostatic Interactions)

$$V = \sum_{i < j} \left(\frac{q_i q_j}{4\pi\epsilon r_{ij}} \right) \quad \text{点电荷, 部份电荷}$$

溶剂化模型

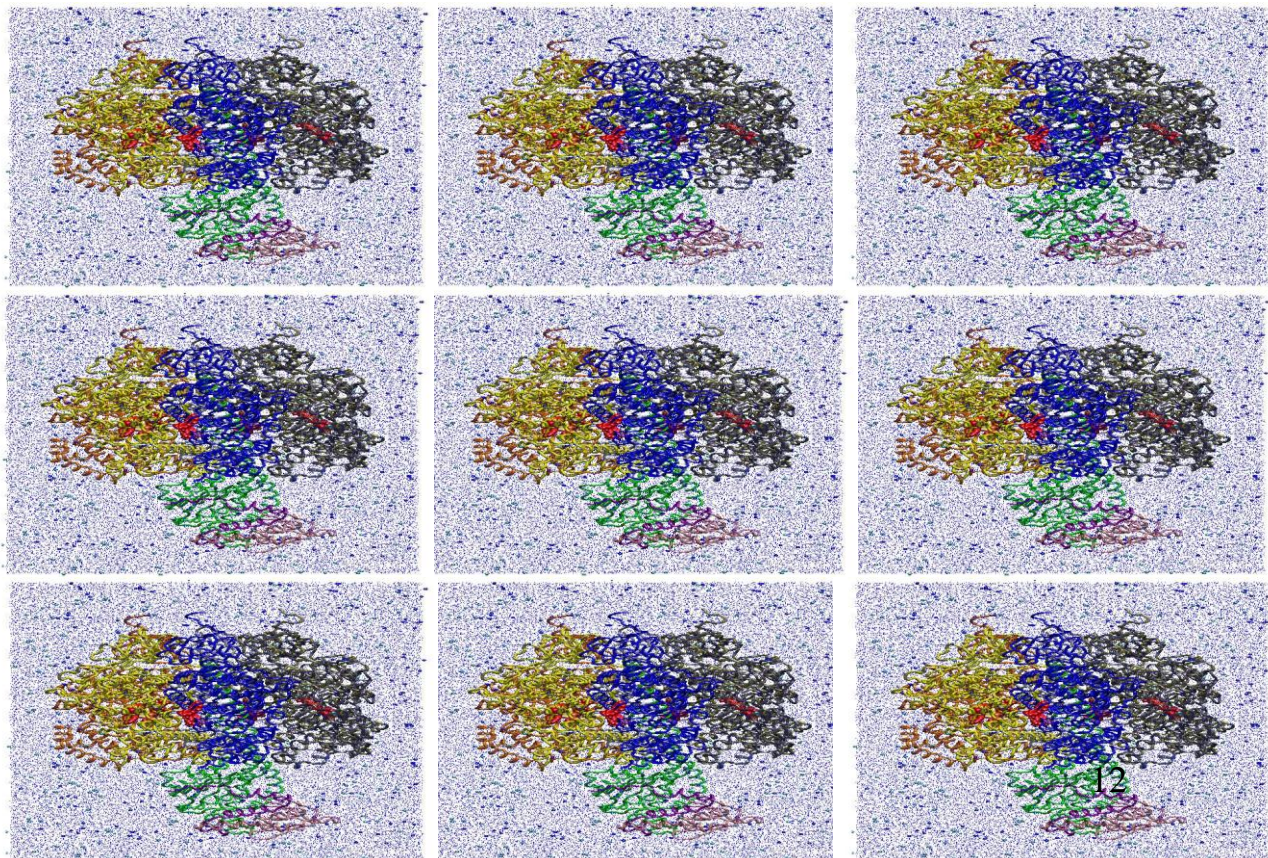
溶剂化问题

隐式水模型

Generalized Born Model, etc

显示水模型

周期边界条件

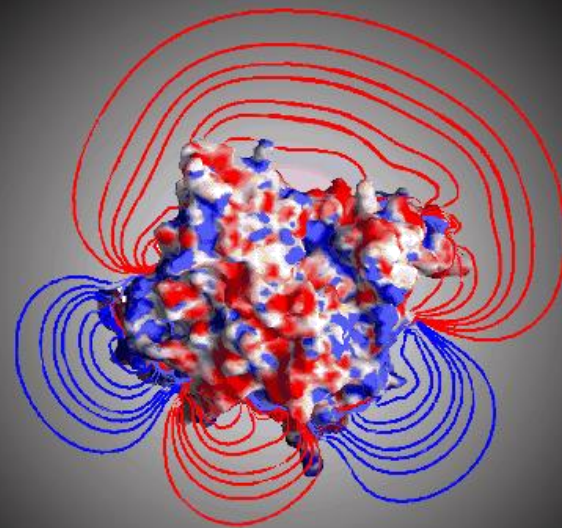


溶剂化能计算的隐式水模型 (连续介质模型)

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cav}$$

- 静电势场是长程的
- 许多生物体系带大量电荷，如DNA
- 许多结合及反应过程是静电势场加速的
- 有时需考虑氢键

Electrostatic Potential Contours of
Mouse Acetylcholinesterase



广义玻恩模型

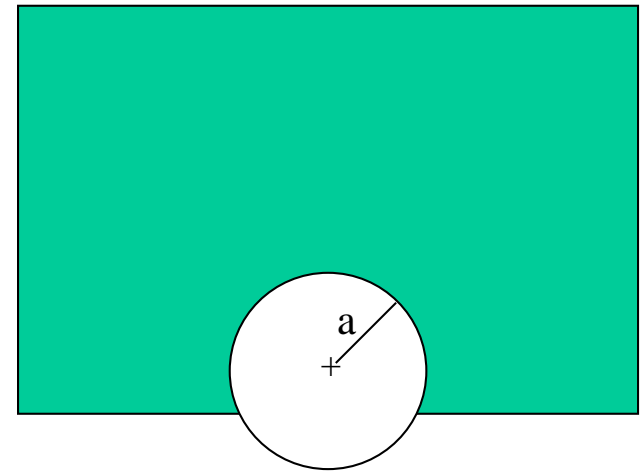
The Born Model:

$$\Delta G_{elec} = -\frac{q^2}{2a} \left(1 - \frac{1}{\epsilon}\right)$$

In reduced electrostatic units, $4\pi\epsilon_0$ was omitted.

The generalized Born equation:

$$\Delta G_{elec} = -\left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \frac{q_i^2}{a_i}$$



将电荷从真空转移到溶剂中所做的功：为分别在真空中和溶剂中给半径为 a 的质点加上电荷所做的功之差。 a 为Born半径，一般取离子半径（加上7%）。

The generalised Born/surface area model (GB/SA)

$$\Delta G_{cav} + \Delta G_{vdw} = \gamma A + b$$

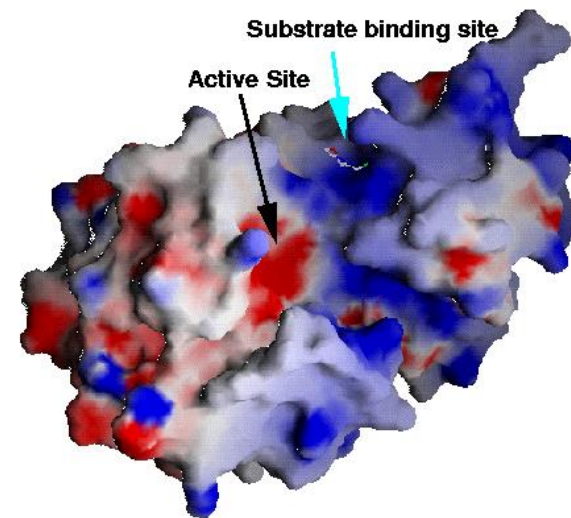
产生空穴及vdw作用均与可及表面积A成正比，设b=0

简化的溶剂化计算方法：

Atomic solvation parameters

$$\Delta G_{sol} = \sum_i a_i S_i$$

生物分子表面静电势分析



Poisson-Boltzmann方程数值解法

有限元差分方法：FDPB

(Finite Difference Poisson-Boltzmann)

代表性程序：DelPhi, GRASP

可用于计算静电势分布，静电势表面，
溶剂化能，氨基酸的表观pKa, etc

Poisson 方程 $\nabla \cdot \varepsilon(r) \nabla \phi(r) = -4\pi\rho(r)$

Debye-Hückel 理论：离子氛的分布

$$n_i = n_i^0 e^{-q_i \phi / kT}$$

$$\rho_i = q_i n_i$$

$$\nabla \cdot \varepsilon(r) \nabla \phi(r) = -4\pi[\rho_{macro}(r) + \sum q_i n_i^0 e^{-q_i \phi(r) / kT}]$$

Poisson-Boltzmann 方程

溶液中的离子:

$$\rho_{Na^+} = +1e * 0.1M * e^{-1\phi/kT}$$

$$\rho_{Cl^-} = -1e * 0.1M * e^{+1\phi/kT}$$

$$e^{+x} - e^{-x} = 2\sinh(x)$$

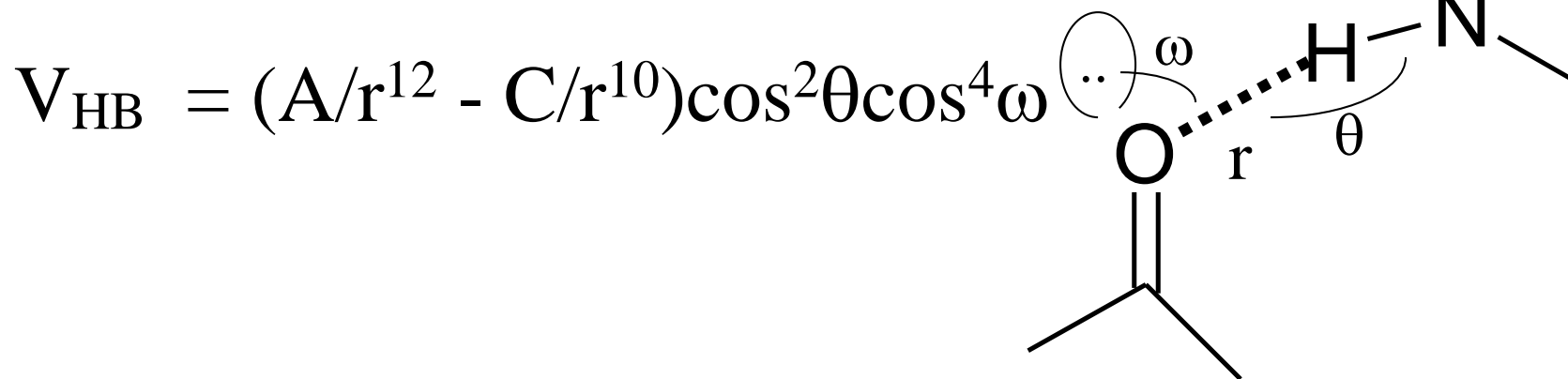
Poisson-Boltzmann 方程:

$$\nabla \cdot \epsilon(r) \nabla \phi(r) = -4\pi[\rho_{macro}(r) + q_i n_i^0 \sinh(\phi/kT)]$$

5. 氢键 (Hydrogen bonding term)

$$V(r) = A/r^{12} - C/r^{10}$$

实例：YETI 力场



目前的力场一般不单独计算氢键项，直接用静电项。

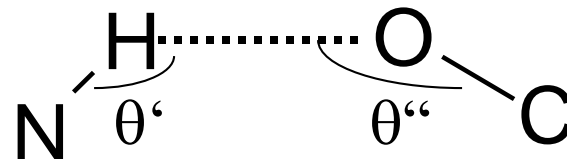
力场举例

$$V = \sum_{bonds} \frac{k_b}{2} (r - r_0)^2 + \sum_{angles} \frac{k_a}{2} (\theta - \theta_0)^2$$

$$+ \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\varphi - \gamma))^2$$

$$+ \sum_{\substack{i < j \\ r < 12A}} (4\epsilon_{ij} [(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}})$$

$$+ \sum (\frac{A'}{rm} - \frac{C'}{rp}) f(\theta', \theta'')$$



M=12 or 6, p=10 or 4

6. 力场的参数化

自洽场方法 (CFF)

尝试法(trial-and-error)

人工智能

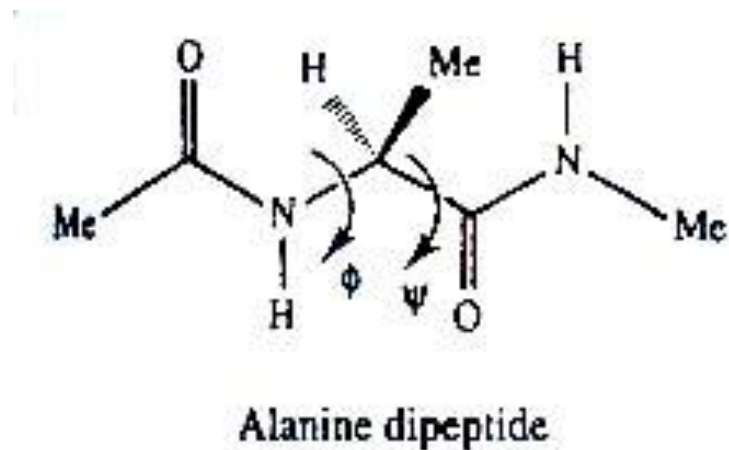
Table 4.1: Atom types.

Atom type	Description	Mass /amu
CA	Aromatic carbon	12.01
CP	Aromatic carbon (joining aromatic rings)	12.01
HA	Aromatic hydrogen	1.008
CZ	Carbon in cyano group	12.01
NZ	Nitrogen in cyano group	14.01
Cn*	Aliphatic carbon	12.01
HC	Aliphatic hydrogen	1.008
F	Fluorine	19.00

* n indicates number of attached hydrogen atoms.

http://cmt.dur.ac.uk/sjc/thesis_dlc/node91.html

分子势函数的极小化不能有效找到总体能量最小点



$$V = \sum V_b + \sum V_{\text{angle}} + \sum V_{\text{torsin}} + \sum V_{\text{vdw}} + \sum V_{\text{ele}}$$

能量函数的最小化—NP problem

能量函数的极小化：Energy minimization

一级微商算法，二级微商算法

二. 分子构象分析的常用方法

1. 系统搜索方法

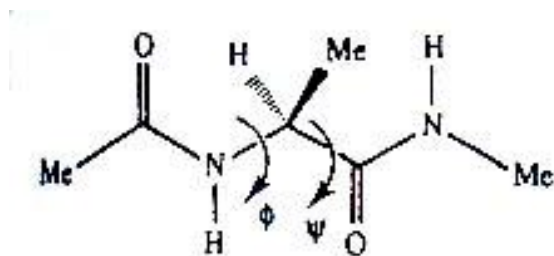
2. 随机搜索方法

- 一般随机方法
- Monte-Carlo方法
- Monte-Carlo模拟退火方法

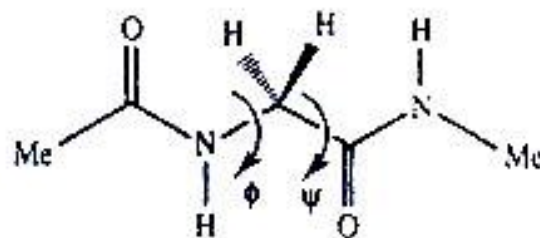
3. 距离几何算法

4. 分子动力学方法

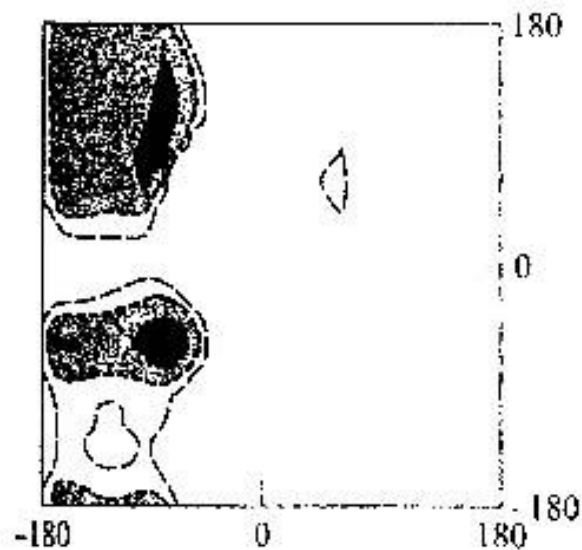
二肽的构象搜索结果- Ramachandran Plot



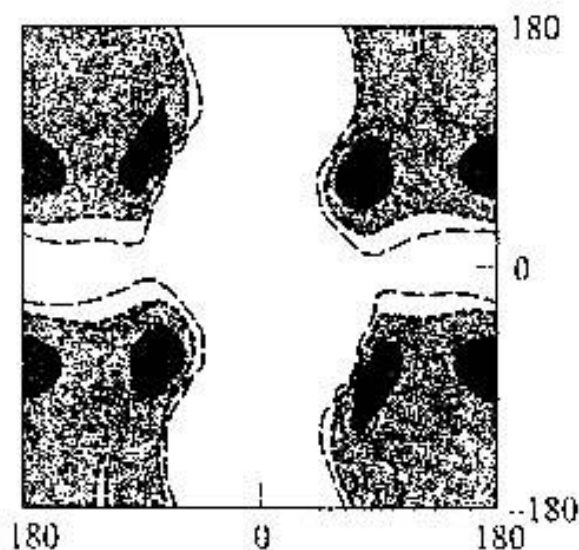
Alanine dipeptide



Glycine dipeptide



(a)



(b)

随机算法

基本算法：

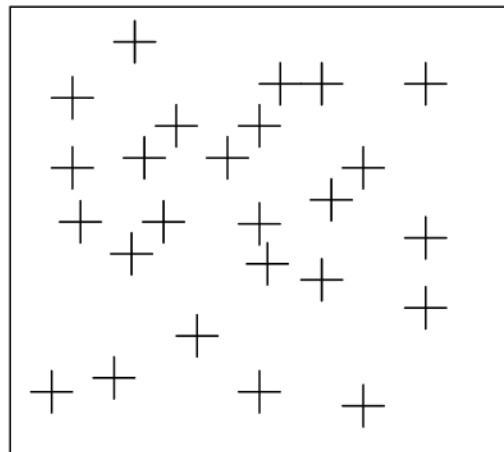
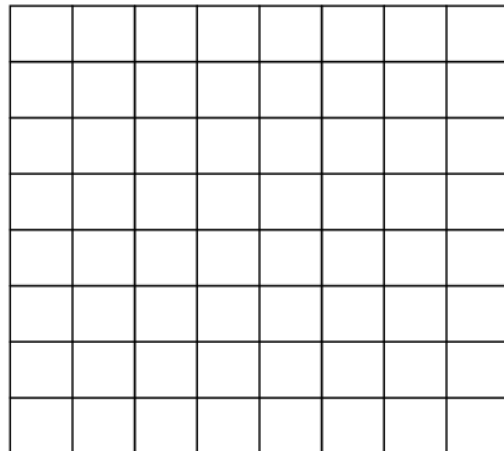
- 随机产生多个构象

(可以变化直角坐标，二面角)

- 能量优化

- 构象分组

- 选择大组中能量最低的构象



Monte-Carlo 方法

原理：Metropolis取样法，Metropolis, 1953

算法：

构象 r_i 产生微小随机变化 $\rightarrow r_{i+1}$

计算 $\Delta E = E(r_{i+1}) - E(r_i)$

if $\Delta E \leq 0$, 接受 r_{i+1}

else 在(0,1)之间产生一随机数R

if $R < \exp(-\Delta E/kT)$ 接受 r_{i+1}

不断循环，从而产生一个Boltzman集合，

此集合对低能构象的取样几率较大

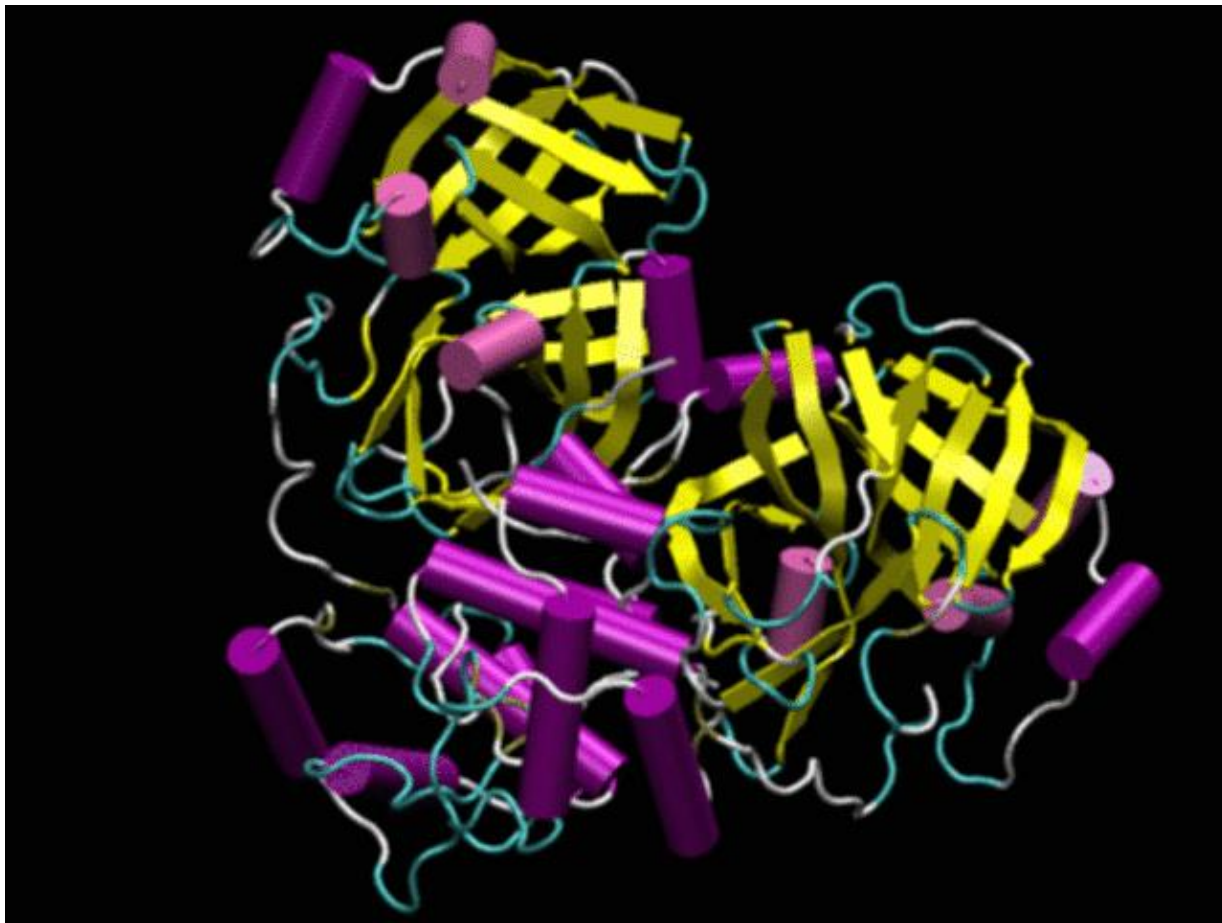
距离几何算法

使用原子间的距离矩阵来表示分子构象，运用实验中得到的原子距离信息，随机产生距离矩阵再转化为直角坐标上的分子构象。

算法：

- 1) 计算各原子间距离的上下边界值
- 2) 随机产生原子间距离矩阵
- 3) 将矩阵转化为直角坐标
- 4) 优化分子
- 5) 重复2，4至收敛

三. 分子动力学模拟



分子动力学基本原理

采用分子力学力场，假定体系中任意原子运动均符合经典牛顿力学方程。当初始坐标与速度给定后，积分后可得各原子的运动轨迹.

$$F_{x_i} = -\frac{\partial E}{\partial x_i} = m_i \frac{d^2 x_i}{dt^2}$$

$$F_{y_i} = -\frac{\partial E}{\partial y_i} = m_i \frac{d^2 y_i}{dt^2}$$

$$F_{z_i} = -\frac{\partial E}{\partial z_i} = m_i \frac{d^2 z_i}{dt^2}$$

如何开始一个分子动力学模拟

- t=0的起始结构：优化过的实验结构或模型结构
- t=0的起始速度：在给定温度下取Maxwell-Boltzmann分布

$$P(v_{ix}) = \sqrt{\frac{m_i}{2\pi k_B T}} \exp\left(-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T}\right)$$

- 加速度由势函数求出。

常见的分子运动时间尺度

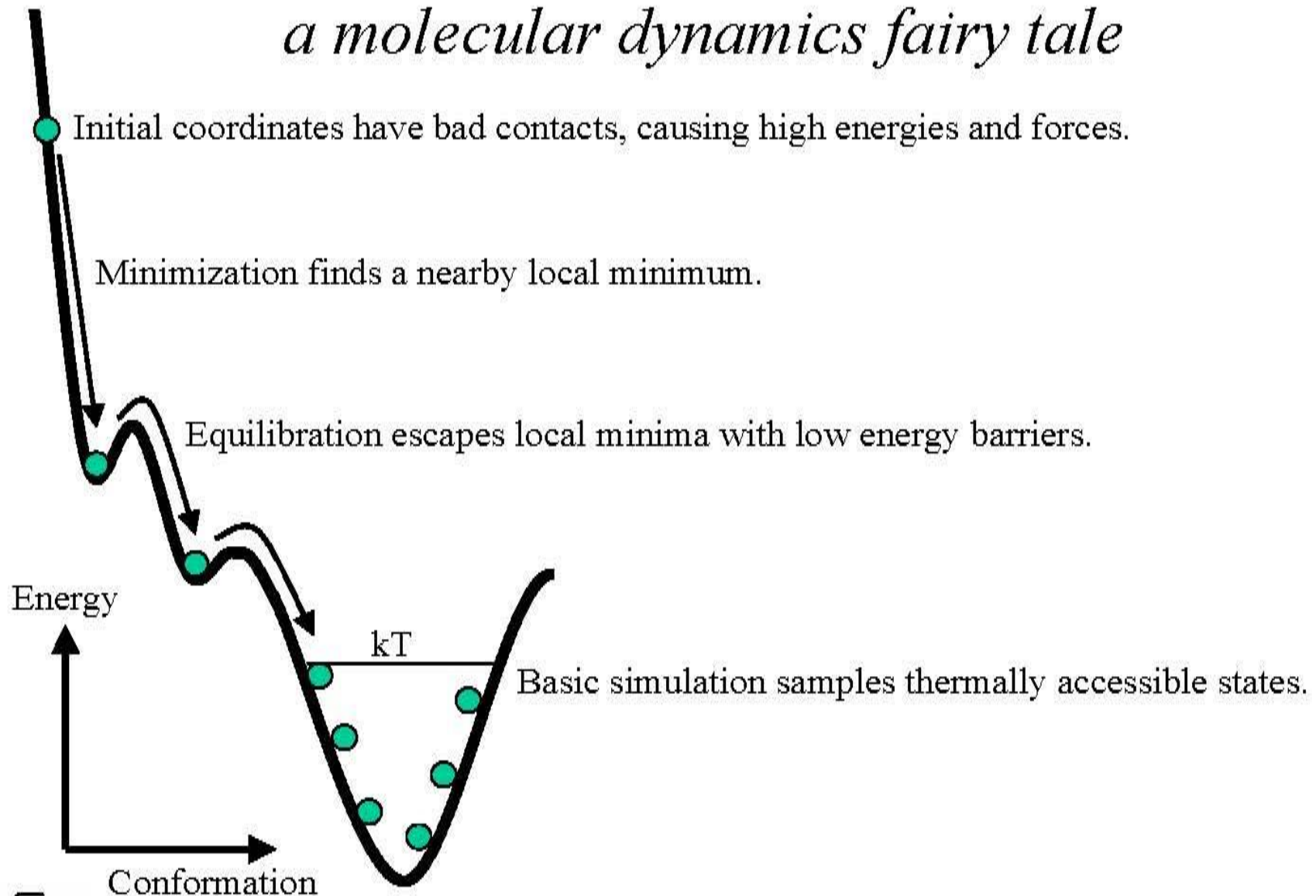
Time scale	Amplitude	Description
short femto, pico $10^{-15} - 10^{-12}\text{s}$	0.001 - 0.1 Å	- bond stretching, angle bending - constraint dihedral
medium pico, nano $10^{-12} - 10^{-9}\text{s}$	0.1 - 10 Å	- unhindered surface side chain motion - loop motion, collective motion
long nano, micro $10^{-9} - 10^{-6}\text{s}$	1 - 100 Å	- folding in small peptides - helix coil transition
very long micro, second $10^{-6} - 10^{-1}\text{s}$	10 - 100 Å	- protein folding

选择合适的步长：0.001ps–0.002ps

选择合适的模拟时间：1ns–1μs…1ms

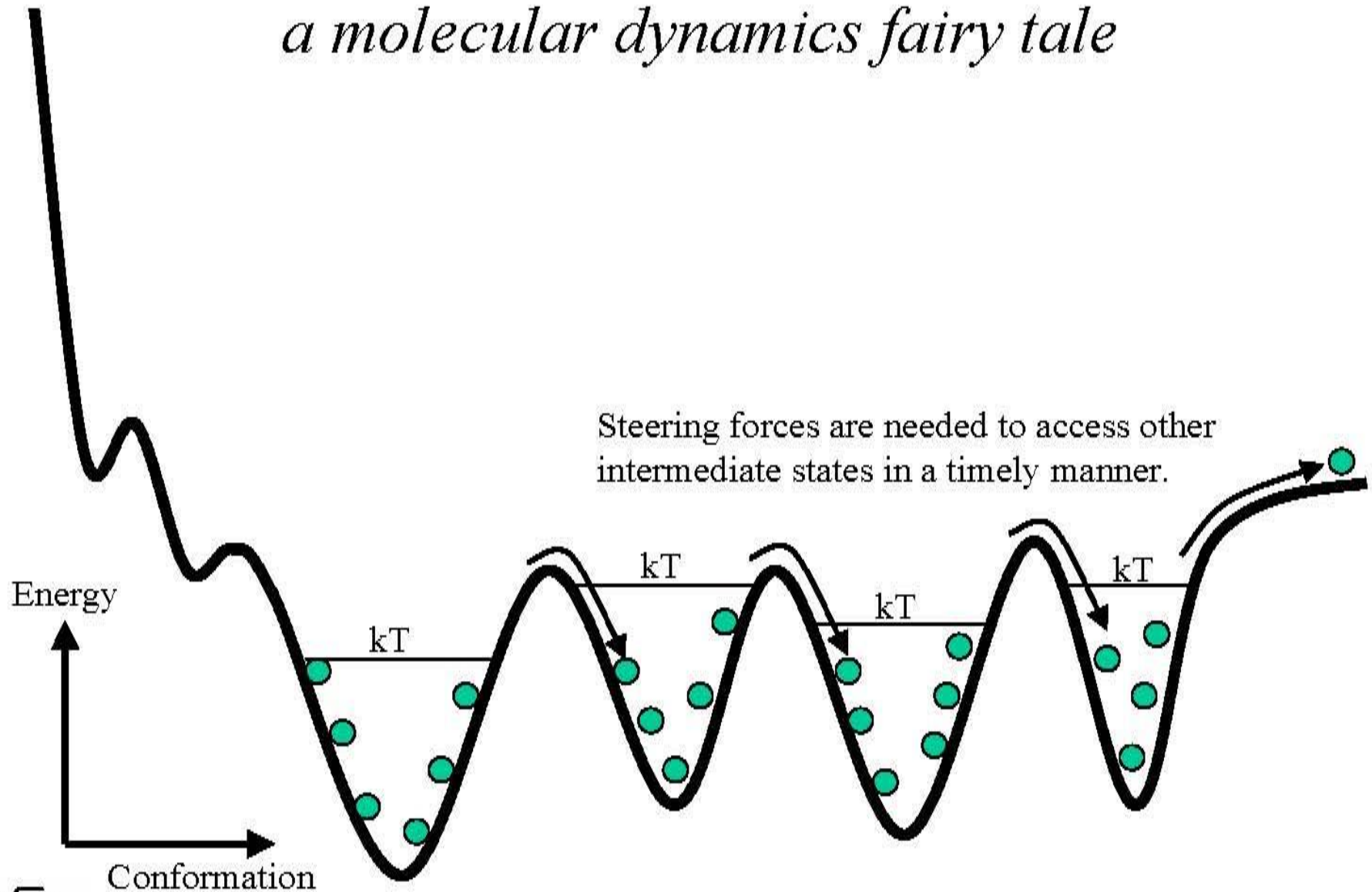
From the Mountains to the Valleys

a molecular dynamics fairy tale



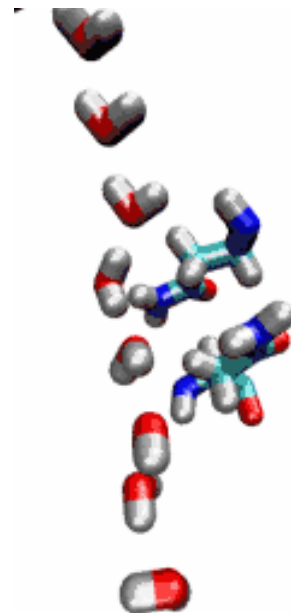
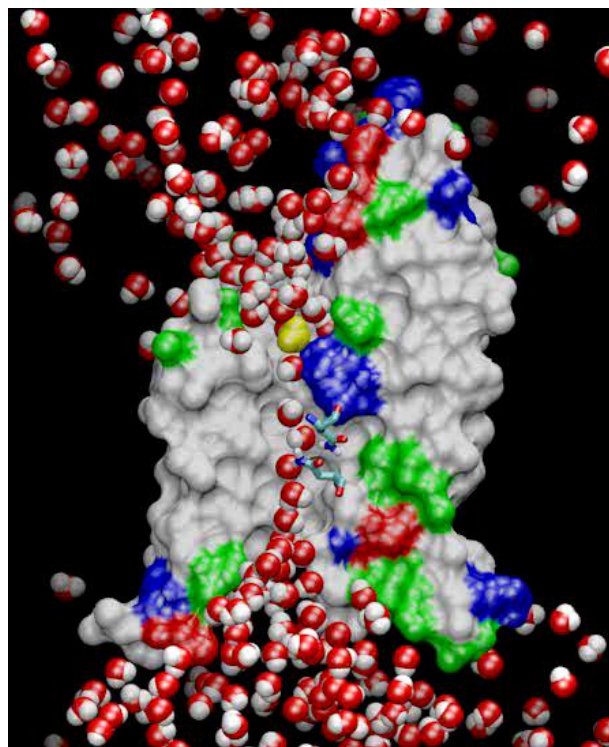
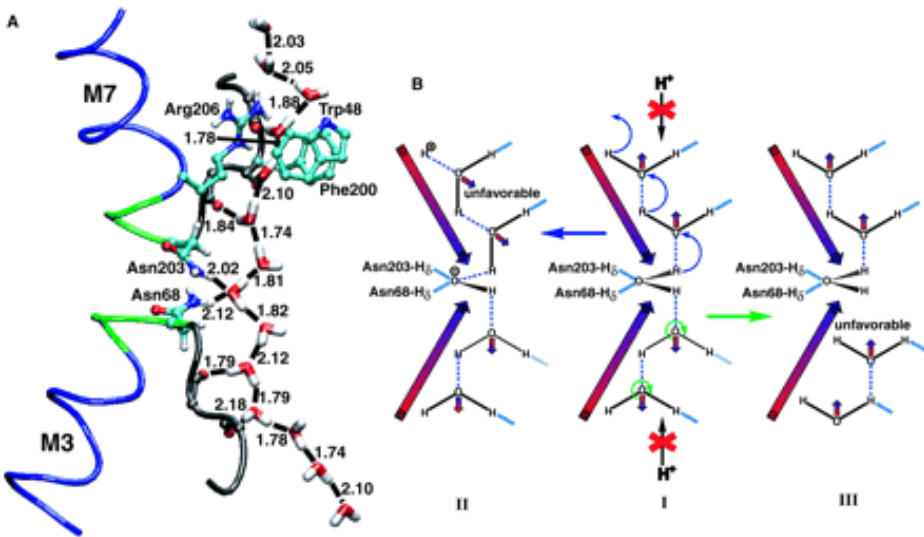
From the Mountains to the Valleys

a molecular dynamics fairy tale



复杂体系的分子动力学模拟

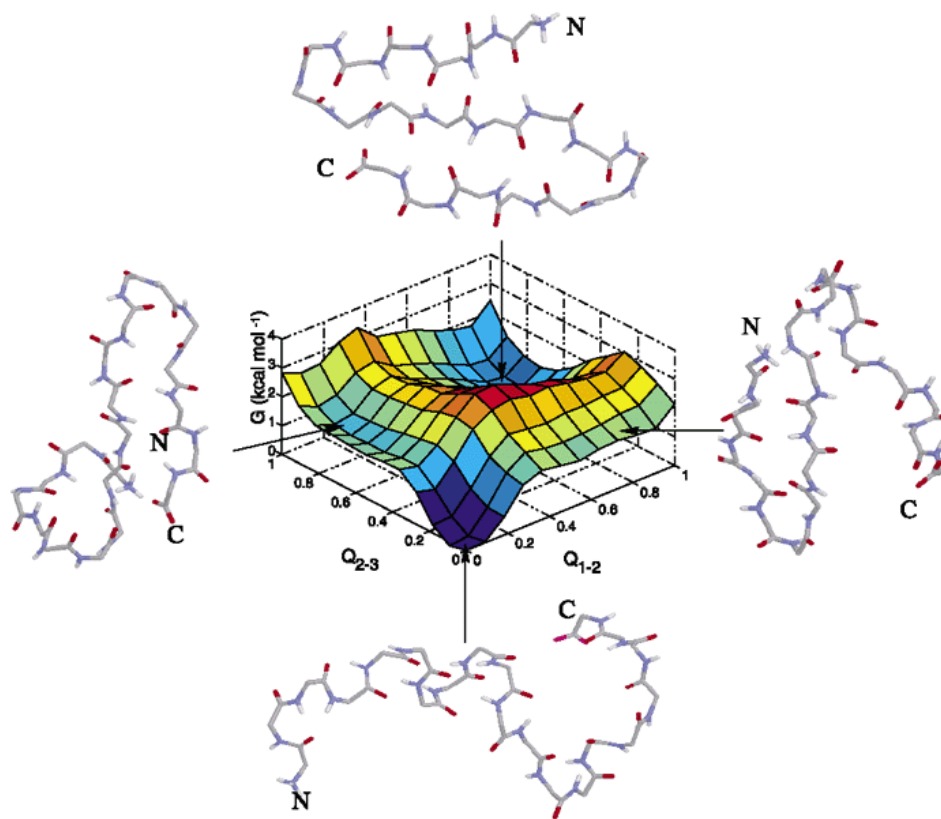
水通道的分子动力学模拟



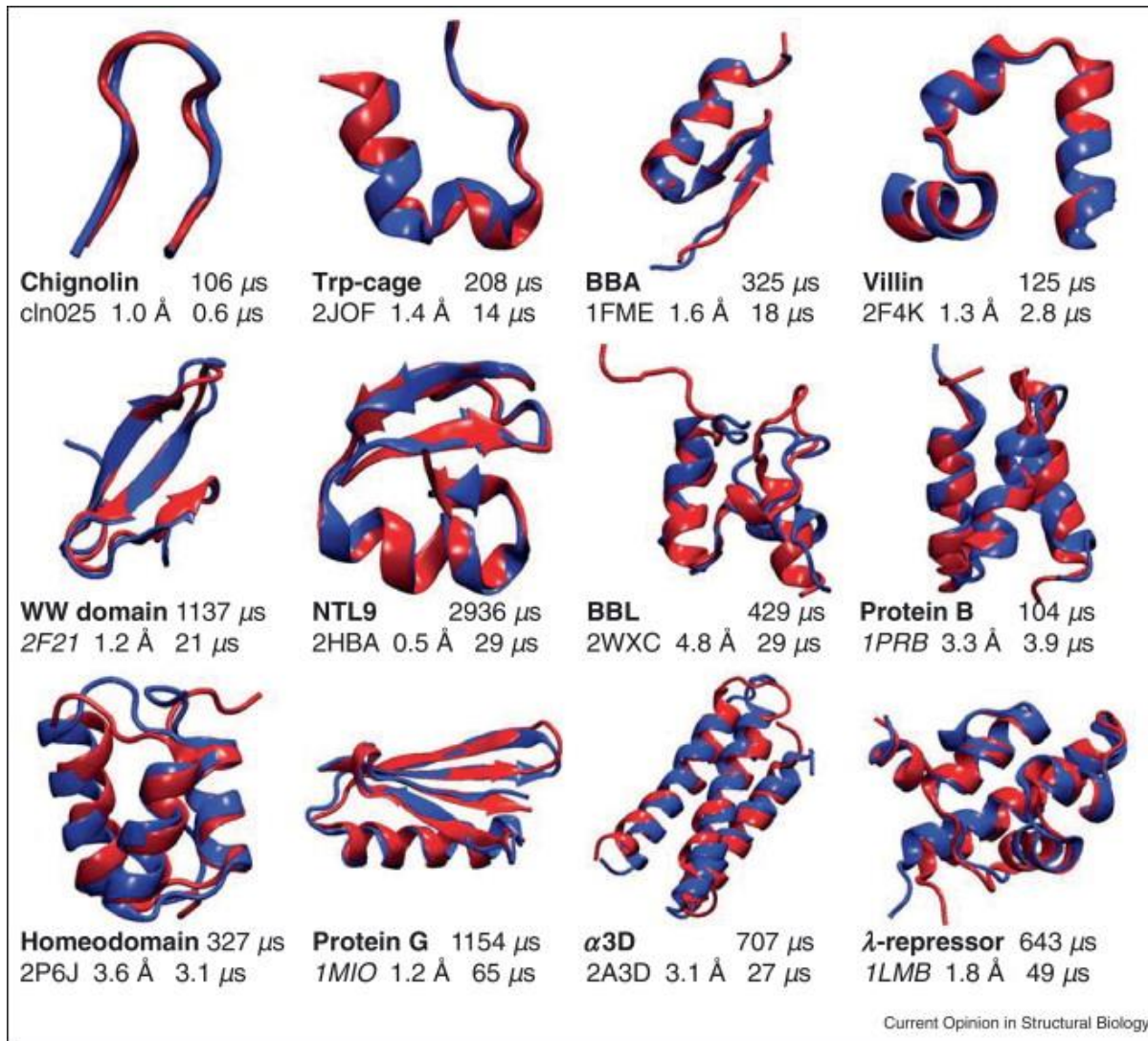
Emad Tajkhorshid,..., Klaus Schulten, *Science*, 2002, 296: 525.

长时间的分子动力学模拟进行构象采样 用于小蛋白的结构预测

Folding simulations of a three-stranded beta-sheet peptide.
330K, 2×10^5 conformations, 20 runs, $> 2 \mu s$



The State-of-Art of Protein Folding Simulations by DE Shaw Research



Anton: specially designed computer for MD simulations



David E. Shaw

K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw. How fast-folding proteins fold. Science, 334 (2011), 517
 R.B. Best. Atomistic molecular simulations of protein folding. Curr Opin Struct Biol, 22 (2012), 52

"Assessing the Accuracy of Physical Models Used in Protein-Folding Simulations: Quantitative Evidence from Long Molecular Dynamics Simulations,"

Stefano Piana, John L. Klepeis, and David E. Shaw, *Current Opinion in Structural Biology*, vol. 24, 2014, pp. 98–105.

- 1. Long simulations can accurately reproduce rates and free energies of folding.**
- 2. The structure and dynamics of folded proteins can be accurately described.**
- 3. Folding enthalpies and characteristics of the unfolded state are reproduced poorly.**

四. 分子建模与模拟的典型应用场景

1. 分子构象分析，蛋白质结构预测
2. 生物分子晶体结构、核磁共振结构等测定时的结构计算及和优化
3. 分子体系的动态变化
4. 分子间结合自由能计算，如靶标与药物
5. 蛋白质与药物的结合过程
6. 分子聚集体预测，晶体结构预测
7. etc

作业

- 1、蛋白质分子的能量分为键能、键角能、二面角扭转能、非键相互作用能等，你认为对于蛋白质结构稳定性贡献大的主要为哪些项？如果蛋白质中有二硫键，那么二硫键是否会影响蛋白质的三维结构？目前的分子动力学模拟是否可以模拟速度很慢（如秒数量级）的变化？
- 2、为什么无法利用能量极小化算法找到蛋白质的能量最低构象？分子动力学或蒙特卡洛模拟退火等方法在分子构象计算中主要解决什么问题？
- 3、蛋白质与配体结合过程的主要驱动力包括哪些？在已经形成的复合物中，两者间的主要作用力是什么？某个蛋白质P在实验条件下可以与配体L结合，但实验测出来的 $\Delta H > 0$ ，你认为驱动结合的主要因素是什么？