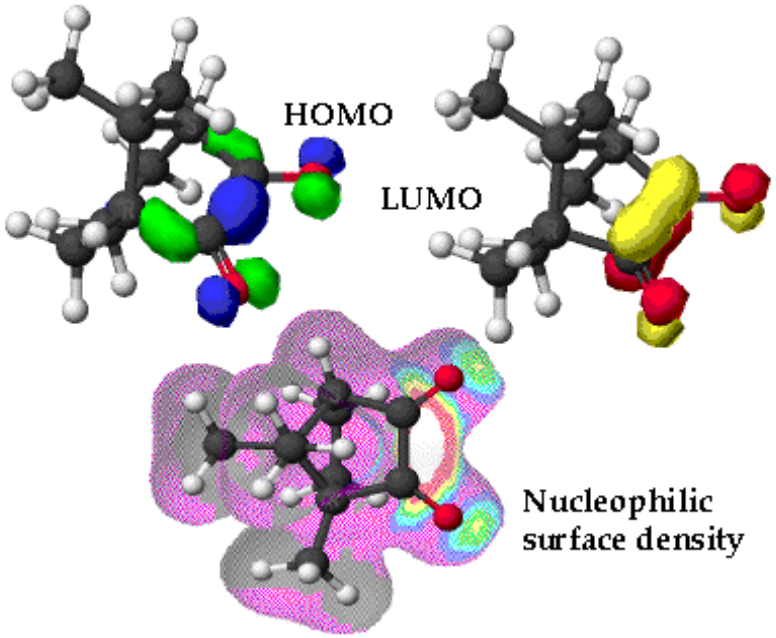
第七讲
分子模拟方法的应用

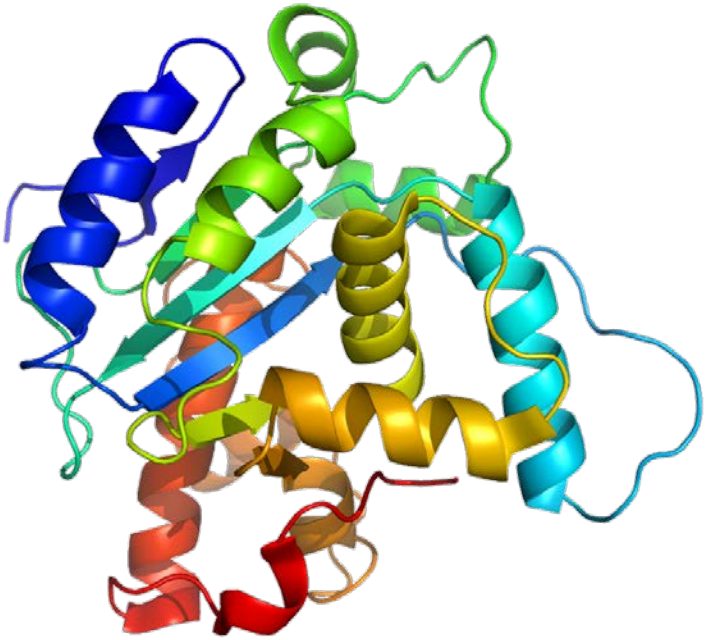
授课教师：王任小 研究员

中国科学院上海有机化学研究所
Shanghai Institute of Organic Chemistry

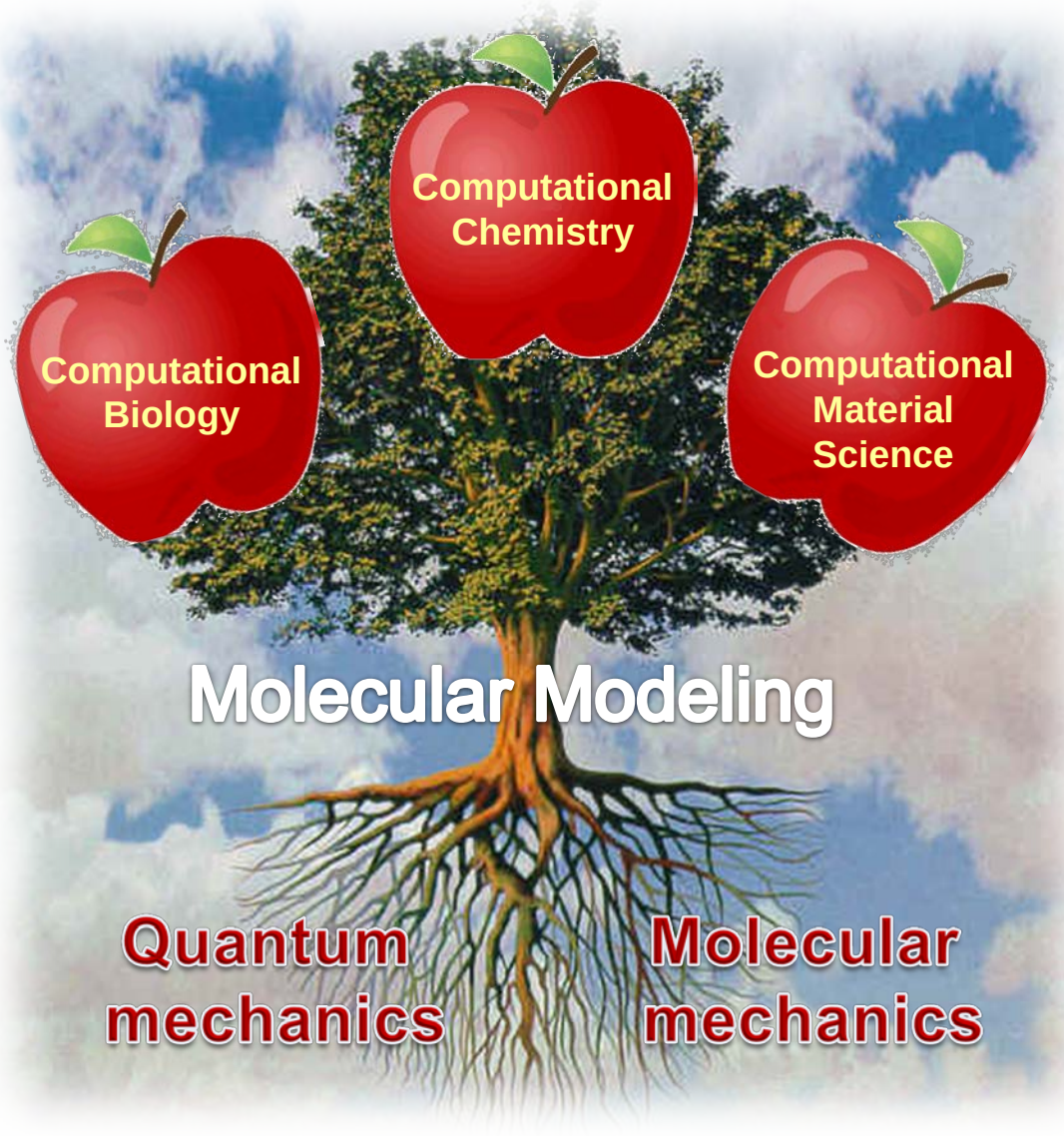
引言：分子模拟的基本概念

分子模拟：将分子体系描述为可操作的数学模型，借助计算机技术来研究分子体系的结构、能量和其他性质。

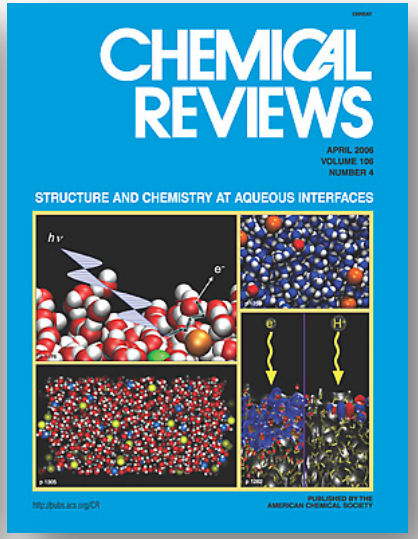


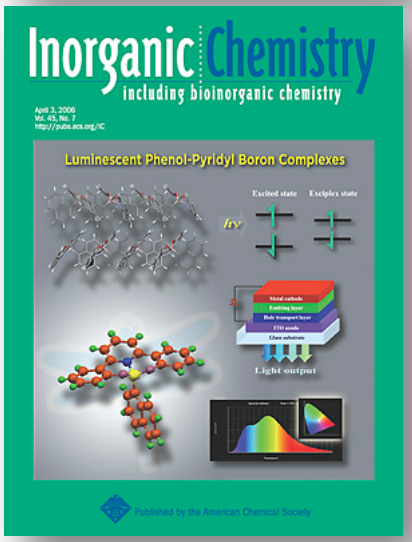


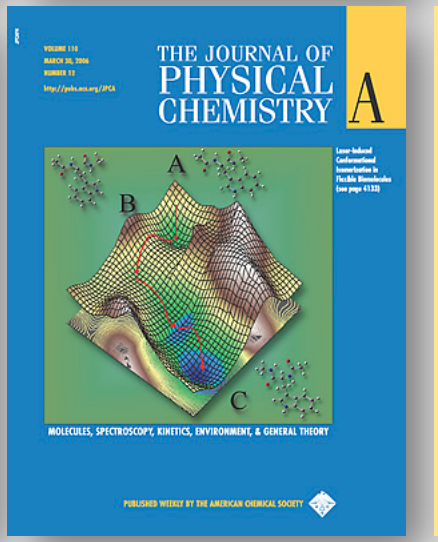
Molecular Modeling As a Central Technique

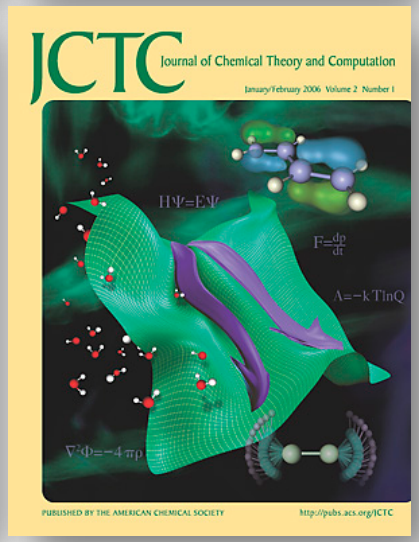


Molecular Modeling As a Central Technique







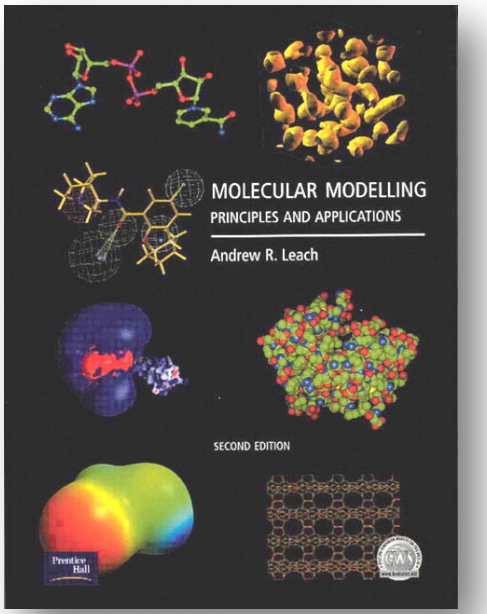


授课提纲

一、计算化学方法用于研究有机反应机理

二、分子动力学模拟用于研究生物大分子体系

三、基于靶标结构的药物分子设计



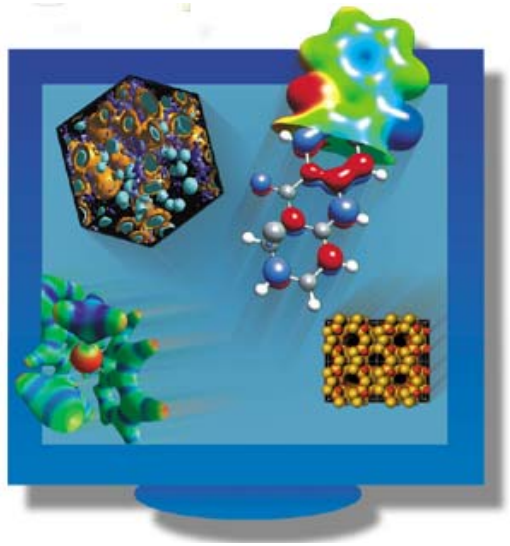
补充参考教材：

Andrew R. Leach, *Molecular Modeling: Principles and Applications*, 2nd Ed., Prentice Hall, 2001.

I. Computational Organic Chemistry

理论计算以及分子模拟在有机化学研究中可以用于解决以下问题：

- 对有机分子进行构象分析，确定其低能构象以及构象变化的规律。
- 研究有机分子的各种物理性质：标准生成焓、偶极矩、简正振动模式等。
- 研究有机反应机理，确定反应的过渡态，表征有机反应的热力学以及动力学性质。



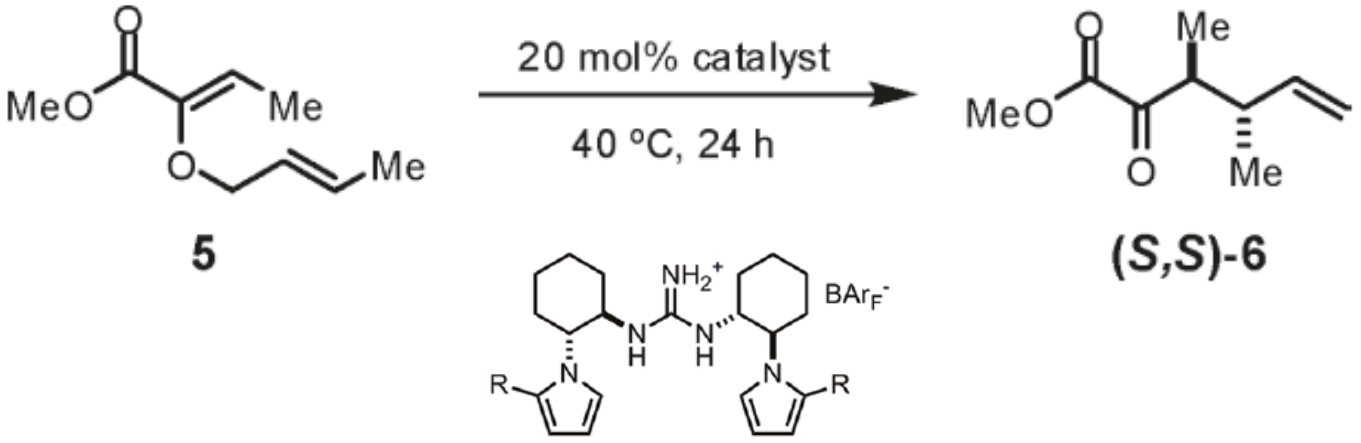
An Example: Claisen Rearrangement

J. Am. Chem. Soc. **2011**, *133*, 5062–5075.

Transition-State Charge Stabilization through Multiple Non-covalent Interactions in the Guanidinium-Catalyzed Enantioselective Claisen Rearrangement

Christopher Uyeda and Eric N. Jacobsen*

Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States

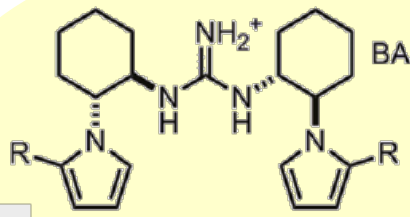


5

(*S*,*S*)-6

Guanidinium as Catalyst

Simple guanidinium ion derivatives were identified as effective catalysts for O-allyl α-ketoesters rearrangement in non-polar organic solvents.

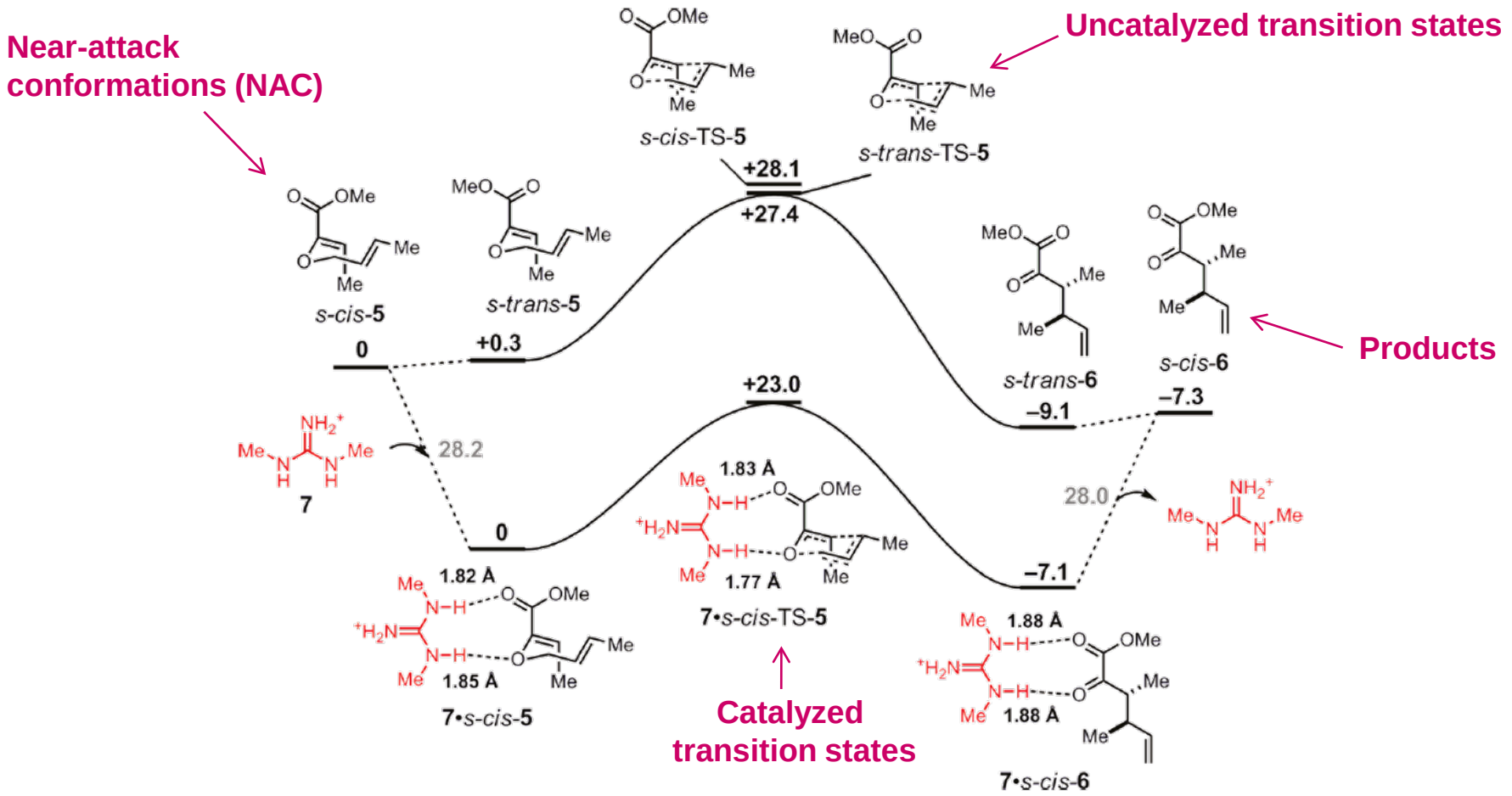


(*R*,*R*)-2: R = Ph
(*R*,*R*)-3: R = Me

entry	catalyst	solvent	conversion (%)[a]	ee (%)[b]
1	(*R*,*R*)-**2**	hexanes	85	73
2	(*R*,*R*)-**3**	hexanes	59	41
3	(*R*,*R*)-**2**	toluene	82	72
4	(*R*,*R*)-**2**	CH_2Cl_2	83	65
5	(*R*,*R*)-**2**	$CDCl_3$	79	66
6	(*R*,*R*)-**2**	TBME	16	19

[a] Conversions were determined from crude reaction mixtures by ^1H NMR signal integration. All rearrangements afforded product **6** with a >20:1 dr. [b] Enantiomeric excesses of purified products were determined by GC analysis using commercial chiral columns.

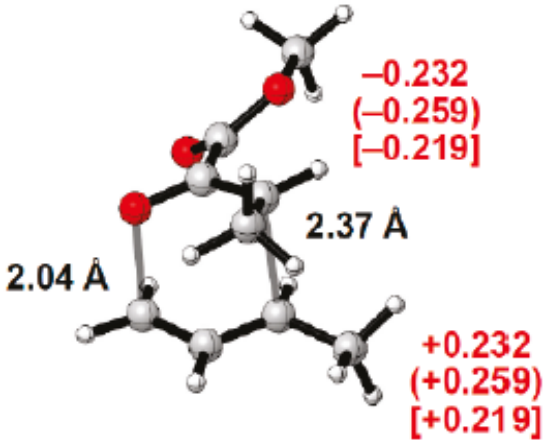
Reaction Diagram



Energy diagram for the uncatalyzed and N, N' -dimethylguanidinium (**7**)-catalyzed rearrangements of ***5*** to **6**. All stationary points are fully optimized at the B3LYP/6-31G(d) level of theory and verified by frequency analysis. Uncorrected electronic energies in kcal/mol are relative to the lowest-energy structure of the substrate or catalyst-substrate complex.

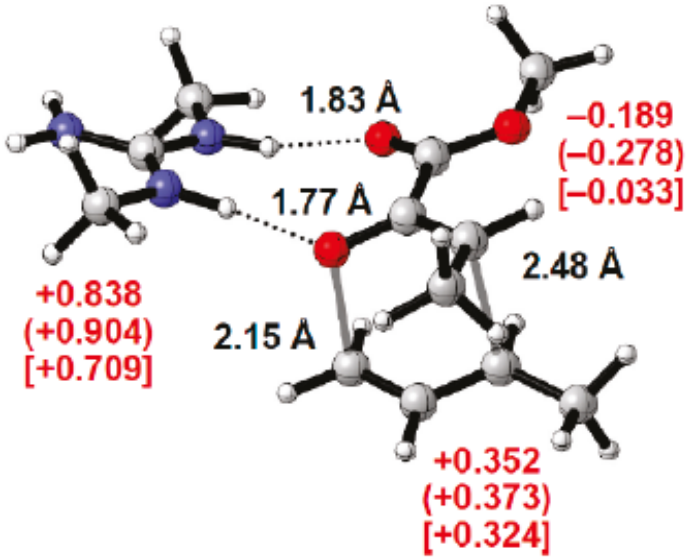
Stabilization of the Transition State

TS in the uncatalyzed arrangement



s-cis-TS-**5**
$\Delta E^{\ddagger}$ = 27.4 kcal/mol

TS in the catalyzed arrangement

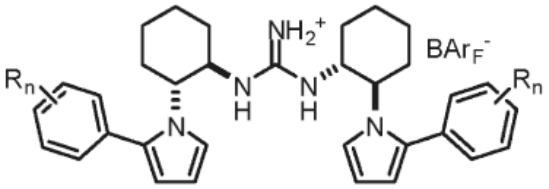


7•*s-cis*-TS-**5**
$\Delta E^{\ddagger}$ = 23.0 kcal/mol

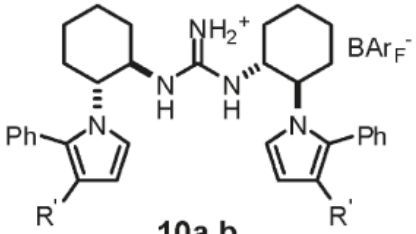
Calculated transition structures at the B3LYP/6-31G(d) level of theory for the rearrangement of 5. Distances for the breaking C-O and forming C-C bonds as well as hydrogen bonds are in angstroms. Mulliken charges, NBO charges in parentheses, and CHelpG charges in square brackets for the oxallyl and allyl fragments as well as the guanidinium ion are shown in red.

A Secondary Factor in TS Stabilization

For rearrangements catalyzed by **2**, different substitutions on the arene moiety exhibited noticeable difference in enantiomeric ratios.



9a–d

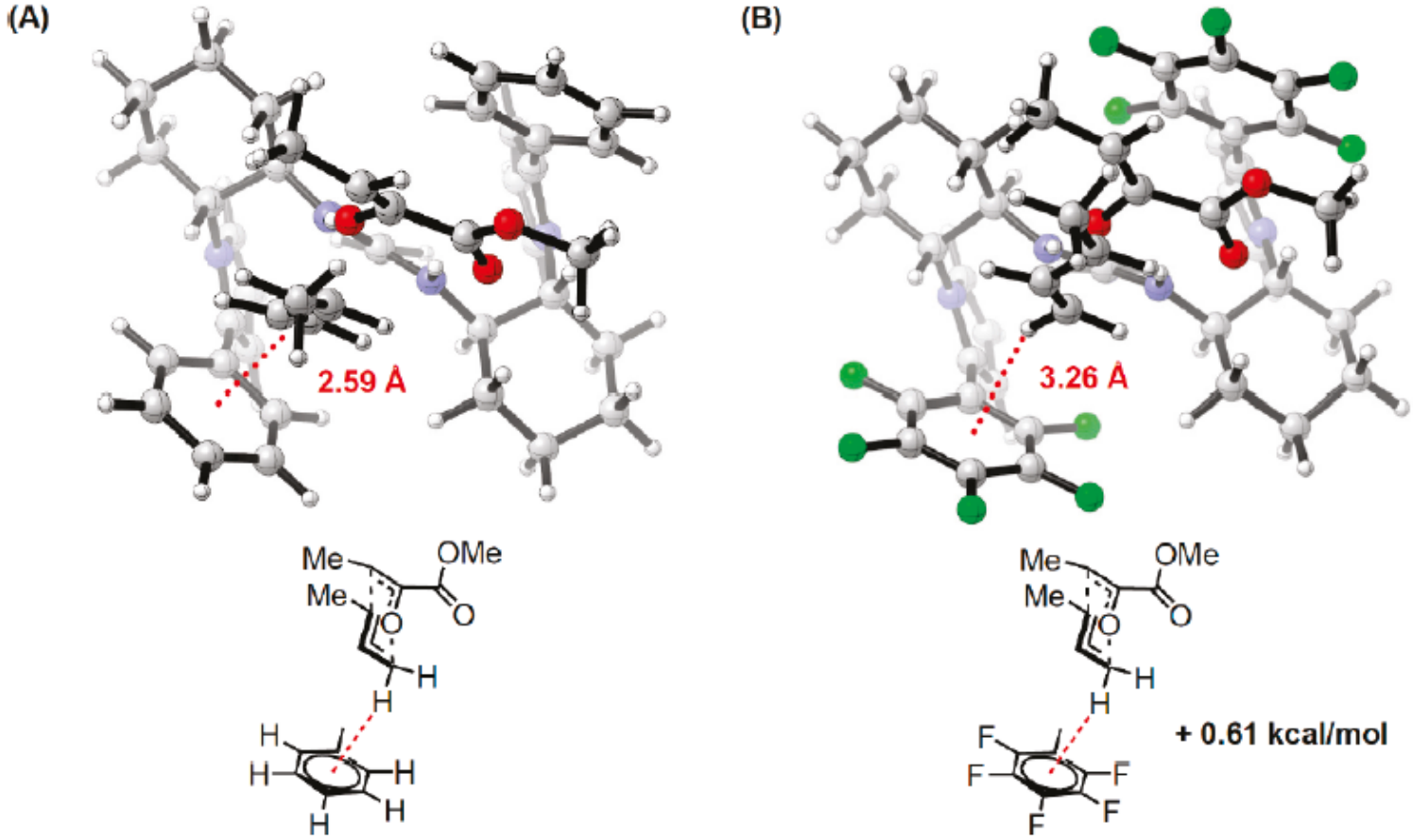


10a,b

catalyst	catalyst substituent	exptl er[a]	exptl $\Delta\Delta G^{\ddagger}$ (kcal/mol)[b]
2	—	6.33 ± 0.05	1.15 ± 0.01
9a	R = 4-fluoro	5.05 ± 0.12	1.01 ± 0.01
9b	R = 4-dimethylamino	8.01 ± 0.20	1.29 ± 0.02
9c	R = 3,4,5-trifluoro	2.61 ± 0.01	0.597 ± 0.002
9d	R = 2,3,4,5,6-pentafluoro	3.80 ± 0.01	0.830 ± 0.001
10a	R′ = methyl	7.03 ± 0.03	1.21 ± 0.01
10b	R′ = trifluoromethyl	4.40 ± 0.05	0.92 ± 0.01

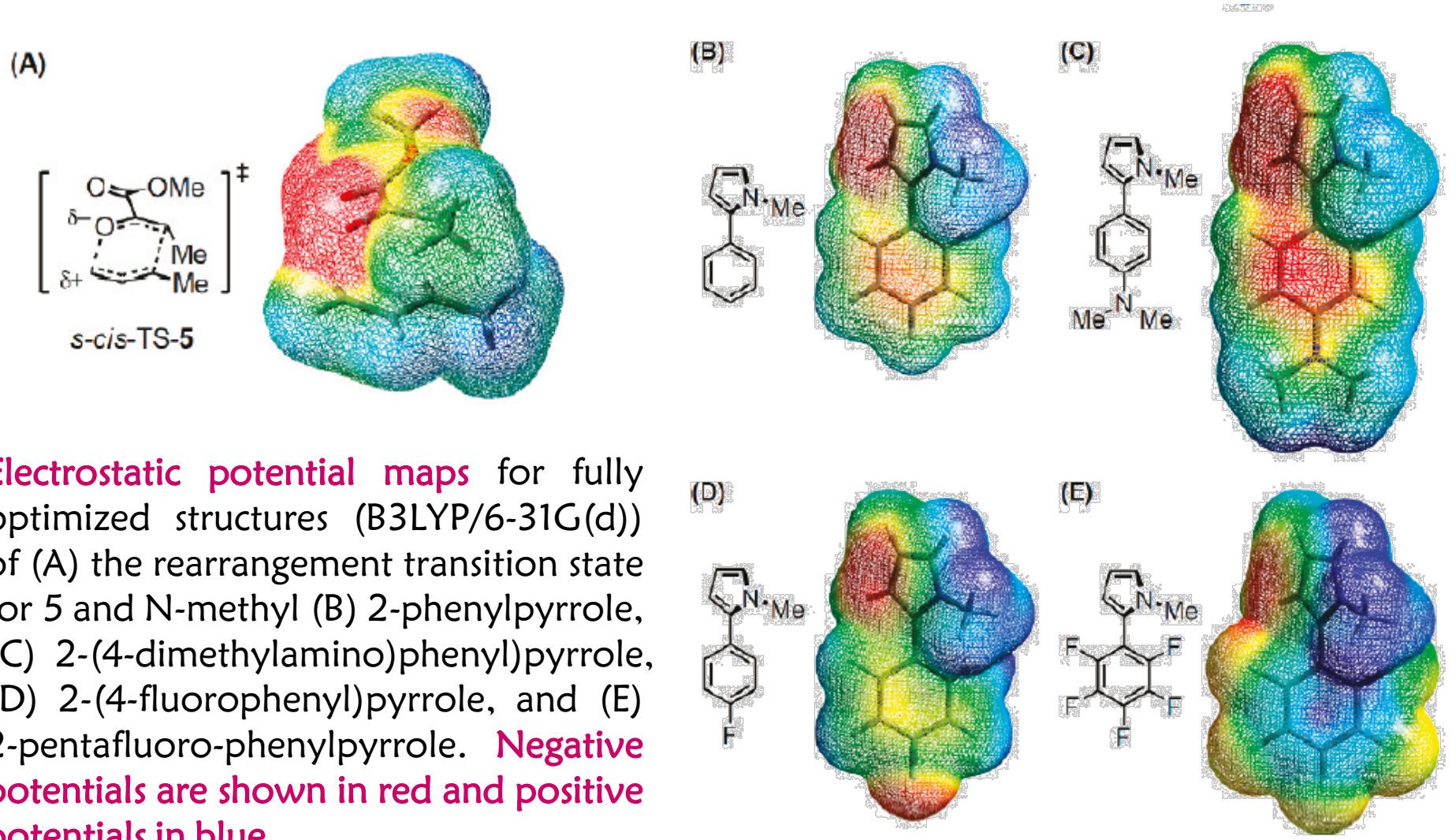
[a] Enantiomeric ratios are averages of two experiments, with the error bars representing the range of results. [b] Relative activation free energies were estimated according to classical transition-state theory ($\Delta\Delta G^{\ddagger} = -RT \ln([(S,S)\text{-}\mathbf{6}]/[(R,R)\text{-}\mathbf{6}])$, T = 313.15 K).

A Secondary Factor in TS Stabilization



pro-(*S*,*S*) Transition structures for (A) catalyst **2** and for the pentafluoro-substituted catalyst **9d** highlighting interactions of the cationic allyl fragment with the π-face of the aromatic moiety.

A Secondary Factor in TS Stabilization



Electrostatic potential maps for fully optimized structures (B3LYP/6-31G(d)) of (A) the rearrangement transition state for 5 and N-methyl (B) 2-phenylpyrrole, (C) 2-(4-dimethylamino)phenyl)pyrrole, (D) 2-(4-fluorophenyl)pyrrole, and (E) 2-pentafluoro-phenylpyrrole. Negative potentials are shown in red and positive potentials in blue.

Claisen Rearrangement: Conclusions

"The phenylpyrrole-substituted guanidinium **catalyst 2 induces a 3.6 kcal/mol lowering of the activation free energy** for the rearrangement of 5, as compared to the thermal rearrangement in hexanes, **corresponding to a rate acceleration of approximately 250-fold.** "

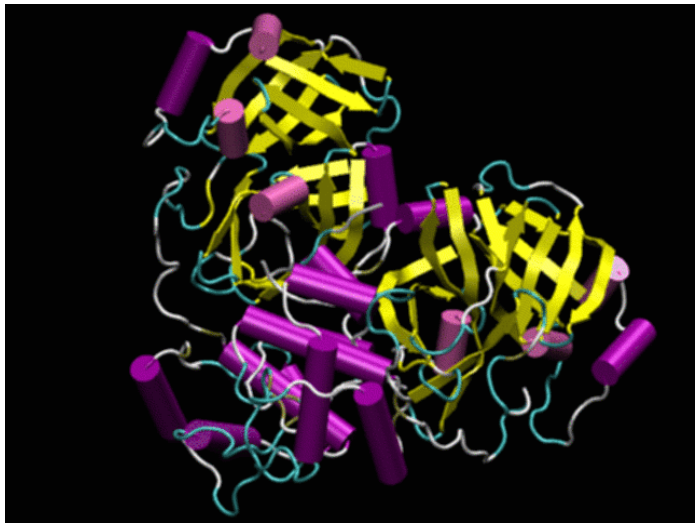
"In computational models, **guanidinium catalysts are seen to interact with the allyl vinyl ether substrate through hydrogen bonds** with both the ether oxygen atom and the pendant ester group. This interaction allows **stabilization of the developing negative charge in the transition state**. "

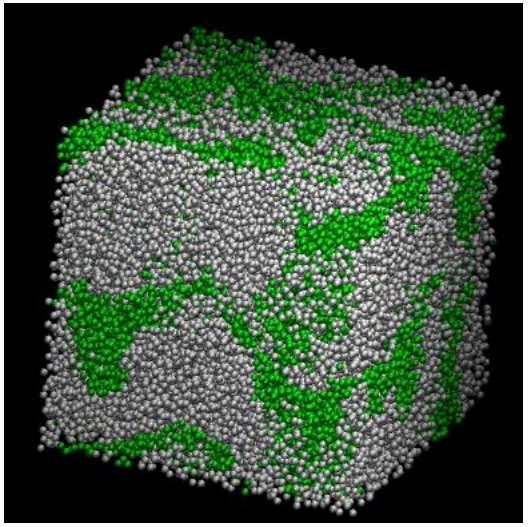
"For rearrangements catalyzed by 2, **a secondary interaction is evident in the major diastereomeric transition state** between the π-system of the catalyst phenyl substituent and the cationic allyl fragment of the substrate. Furthermore, the strength of this interaction is rationally tunable through substitution of the arene."

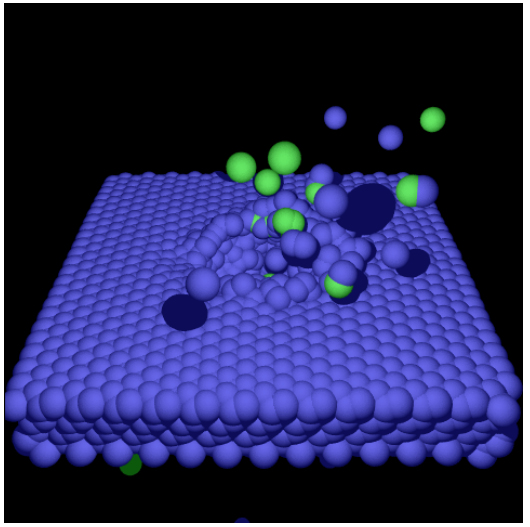
J. Am. Chem. Soc. **2011**, *133*, 5062–5075.

II. Molecular Dynamics

分子动力学(molecular dynamics, MD) 使用牛顿经典力学来模拟分子体系的运动，在由分子体系不同状态构成的系综中抽取样本，从而计算体系的构型积分，并以此为基础计算体系的热力学量和其他宏观性质，并研究体系的性质随时间的变化规律。







结构比较复杂的分子体系处于系列构象的互变平衡之中。分子发挥其功能的同时往往伴随着构象的运动。

A Brief History of Molecular Dynamics

- 1957年，Alder和Wainwright首次进行了基于刚球势的分子动力学模拟。1964年，Rahman采用Lennard-Jones势函数模拟了液态氩的结构与性质。1971年Rahman 和Stillinger报导了液态水的分子动力学模拟结果。这些是首批具有实际意义的分子动力学模拟工作。
- 1977年，Karplus等发表了第一个蛋白质分子(458 atoms)的分子动力学模拟结果。尽管所采用的方法很粗糙，模拟的时间尺度也很短(～10ps)，但是这一工作却极大地改变了人们对蛋白质结构的认识，即蛋白质并不是完全刚性的分子，其构象变化在实现蛋白质的功能中可以起到关键作用。
- 目前分子动力学可以模拟的体系超过100万个原子，可以模拟的时间尺度最长可至到1 ms。

The First MD Simulation of Protein

articles

Dynamics of folded proteins

J. Andrew McCammon, Bruce R. Gelin & Martin Karplus

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

The dynamics of a folded globular protein (bovine pancreatic trypsin inhibitor) have been studied by solving the equations of motion for the atoms with an empirical potential energy function. The results provide the magnitude, correlations and decay of fluctuations about the average structure. These suggest that the protein interior is fluid-like in that the local atom motions have a diffusional character.

RESULTS of X-ray crystallography provide a picture of a globular protein in its native conformation as a well defined, densely-packed structure. Other experimental data[1–10] and theoretical considerations[11–13] indicate that there is considerable local motion inside a protein at ordinary temperatures. Moreover, the structural data themselves show that significant residue or subunit displacements have an important role in the activity of proteins (for example, enzyme catalysis[14], haemoglobin cooperativity[15], immunoglobulin action[16]). To obtain a more complete understanding of proteins, it is essential to have a detailed knowledge of their dynamics. In spite of the considerable effort directed toward protein folding[17], very little has been done to investigate the motions of a protein in the neighbourhood of its equilibrium configuration. For certain cases in which the displacements along a suitably chosen coordinate can be isolated (for example, aromatic side chain rotations in the pancreatic trypsin inhibitor[18], the opening and closing of the active site cleft in lysozyme[12]), it has been demonstrated that empirical energy functions can be used to analyse the motion involved. Here we undertake a more general examination of the internal dynamics of a folded globular protein. The approach used is of the molecular dynamics type[19]. In such a study the classical equations of motion for all the atoms of an assembly are solved simultaneously for a suitable time period and detailed information is extracted by analysing the resulting atomic trajectories. The full interatomic potential can be used to obtain the forces on the atoms, so that the method is applicable even when the system is highly anharmonic. The molecular dynamics approach has been very successful in revealing structural and dynamical characteristics of fluids[19]. As we demonstrate in what follows, molecular dynamics can have a corresponding role for the internal motions of proteins.

Bovine pancreatic trypsin inhibitor (PTI) (see Fig. 1*a*) was selected for study because of its small size (58 amino acid residues), high stability and accurately determined X-ray structure[20,21]. Four water molecules, which are strongly bound in the interior, were included in the dynamical simulation. The potential energy was represented by an empirical function[18] composed of a sum of terms associated with bond lengths, bond angles, dihedral angles, hydrogen bonds, and non-bonded (van der Waals and electrostatic) interactions. Hydrogen atoms are not explicitly considered, but are combined with the heavy atoms to which they are bonded by a suitable adjustment of atomic parameters. This use of 'extended atoms' reduces the number of interactions which must be calculated and also permits larger steps in the trajectory calculation since the high frequency hydrogen vibrations have been eliminated. Integration of the equations of motion was performed by means of the Gear algorithm[22] with time steps of 9.78×10^{-16} s. X-ray coordinates[23] were used for the initial positions and the initial velocities were set equal to zero. After 100 equilibration steps, the stresses in the initial structure had partly relaxed and the system had an internal kinetic energy corresponding to a temperature of 140 K. At this point, all velocities were multiplied by a factor of 1.5 and 250 more equilibration steps were taken. The added kinetic energy (250.6 kcal mol^{-1}) partitioned itself between kinetic and potential terms during this interval, and an average kinetic temperature of 285 K was reached. The actual simulation consisted of 9,000 additional steps, corresponding to 8.8 ps. Some equilibration of the bond lengths, bond angles and electrostatic interactions continues during the first 3 ps of the simulation, producing a small rise in temperature. Over the whole simulation, the average temperature is 295 K and the total energy is well conserved, changing by only 0.7 kcal mol^{-1}.

In what follows we first describe results concerned with the relation between the time-averaged and X-ray structures and with the magnitudes of structural fluctuations. We then examine the details of the dynamics, the correlation and damping of fluctuations, and some examples of the larger changes that occur. Of primary importance are the conclusions reached concerning the fluid-like nature of the internal motions, which will clearly have to be considered in developing dynamical models for biological processes.

Time-averaged structure and fluctuations

The time-averaged structure obtained in the dynamics run is near the X-ray structure but not identical with it; the root mean square (r.m.s.) deviation of the α carbons is 1.2 Å and that for all the atoms is 1.7 Å. The largest deviations come from the two ends of the molecule, the external loop (residues 25–28) and exposed sidechains. The two parts of the β sheet (residues 18–24, 29–35) and the α helix (residues 47–56) have significantly smaller deviations.

Use of the X-ray structure as the starting configuration corresponds to 0.4 kcal mol^{-1} of kinetic energy per atom after equilibration (140 K); for the main dynamics run (295 K), the mean kinetic energy per atom is about 0.9 kcal mol^{-1}. Although there is no tendency to unfold, the dynamic development with this kinetic energy permits the molecule to make small rearrangements throughout; a picture of the structure after 3.2 ps is shown in Fig. 1*b*. Apparently the PTI molecule samples a series of neighbouring conformations, the average of which is not identical with the X-ray structure. It is not clear how much of the difference is a real effect (in the sense that the structure in the crystal is constrained) and how much is due to inaccuracies in the potential function and the absence of solvent. In lysozyme, comparison of the X-ray structures for the triclinic and tetragonal crystal forms[23,24] shows an r.m.s.

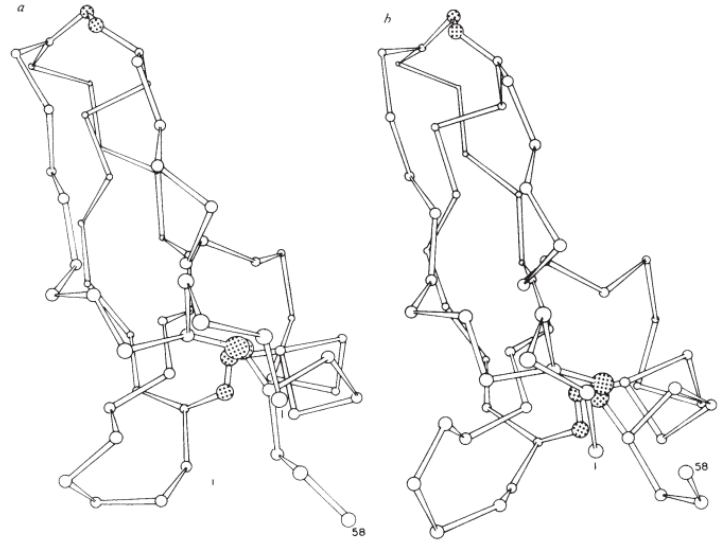


Fig. 1 The peptide backbone (α carbons) and disulphide bonds of PTI. *a*, X-ray structure[21]. *b*, Time evolved structure after 3.2 ps of dynamical simulation.

deviation of backbone atoms equal to 0.5 Å, with the largest changes at the surface of the protein. Also, a variety of other studies[25–27] support the picture of a protein accommodating its structure to different environments. In any case, the X-ray and dynamic average structures are sufficiently similar that the present results should give a meaningful approximation to the short-time dynamics of PTI.

Figure 2 shows the r.m.s. displacements of the α carbons, relative to the dynamical average structure. The three residues at the carboxyl end of the chain have particularly large fluctuations; these are due to the presence of excess kinetic energy resulting from local strain in the initial structure. Some of this kinetic energy was transferred to the nearby amino end of the chain, which also exhibits relatively large r.m.s. displacements. The important fluctuations in residues 25–28 of the loop connecting the two strands of β sheet seem to reflect an intrinsic softness in this part of the molecule. The smaller fluctuations of the β sheet and α helical regions are also evident. The sidechain atom displacements are somewhat larger than those of the α carbons, but generally show a similar pattern of variation. The average r.m.s. fluctuation of all atoms is 0.9 Å. Particularly large r.m.s. displacements (1.8–3.1 Å) occur in the sidechains of Lys 26, Arg 39, Arg 42, Lys 46 and Met 52; all of these are at the surface of the protein. Overall, the large fluctuation regions correspond to those deviating most from the X-ray structure. In principle, the temperature factors from the X-ray analysis can be compared with the calculated fluctuations. The r.m.s. atomic fluctuation for the α carbons obtained from the experimental temperature factor is 0.4 Å. Also, some of the more mobile regions correspond to groups of atoms which were not located or had unusually high temperature factors (J. Deisenhofer, personal communication).

The range of variation of the internal coordinates in the static structure can be compared with the dynamical averages and fluctuations. Since the X-ray results for bond lengths and angles correspond to standard values for the different amino acids[21], we use for comparison the energy-refined geometry (ERG)[18], which reflects the effect of the environment on the coordinates. Selected examples for backbone coordinates are listed in Table 1. It can be seen that the ERG and dynamical averages and r.m.s. variations are in close correspondence. It should also be noted that the average r.m.s. dynamical fluctuation of any given coordinate is two to five times larger than the variation in the coordinate throughout the molecule. The results for the side chains are similar. For the dihedral angle ω, neither the ERG refinement nor the dynamic results show any significant distortions or unusual fluctuations for residues 14–17; this differs from the X-ray results[21].

The dynamic averages for individual φ and ψ dihedral angles differ from the corresponding ERG results by less than 15° for 70 of the 115 backbone angles and by somewhat larger values for the rest (up to 30°). The behaviour of the sidechain dihedral angles is similar to φ and ψ.

The r.m.s. dynamic fluctuations of the bond lengths, the bond angles, and the ω angles are essentially constant along the backbone except for slightly (~ 20%) larger fluctuations in

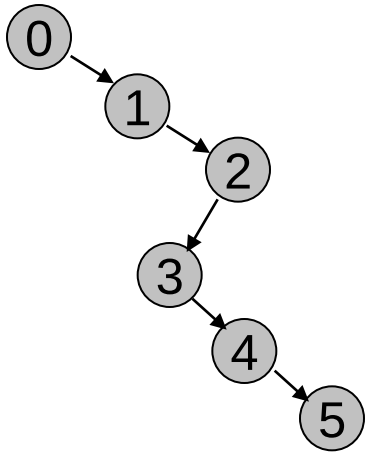
Basic Principles of Molecular Dynamics

分子动力学模拟通常采用分子力场计算体系的能量，假定体系中每个原子的运动均符合经典牛顿力学方程。当体系的初始坐标与速度给定后，进行积分后即可得到各原子的运动轨迹，进而得出体系的各种性质。

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

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

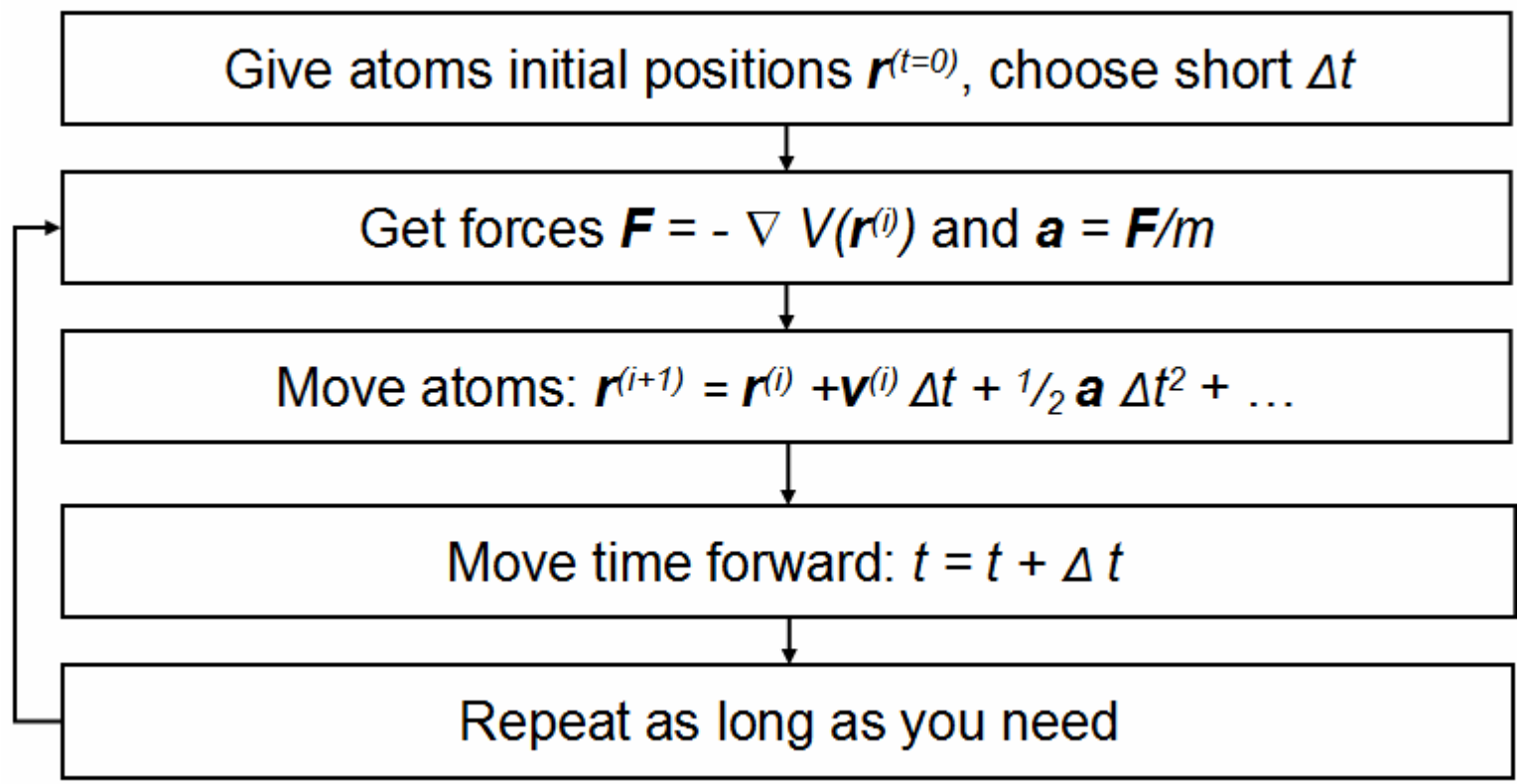
$$Fz_i = m_i \frac{d^2 z_i}{dt^2} = -\frac{\partial E}{\partial z_i}$$



Deterministic Markov chain

微分动力学系统：研究体系的性质随时间演变的特性。

MD Simulation: The Overall Process



MD Simulation: The Finite Difference

有限差分方法(finite difference techniques): 体系的位置及动力学性质可以用Taylor级数展开。

Brook Taylor
(1685-1731)

The Taylor expansion of $f(x)$ when $x \to x_0$

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}\delta x + \frac{f''(x_0)}{2!}\delta x^2 + \frac{f^3(x_0)}{3!}\delta x^3 + \cdots$$

位置： $$r(t+\delta t) = r(t) + \delta t \cdot v(t) + \frac{1}{2}\delta t^2 \cdot a(t) + \frac{1}{6}\delta t^3 \cdot b(t) + \ldots$$

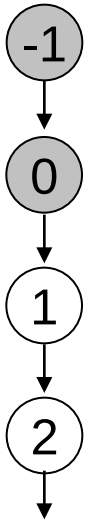
速率： $$v(t+\delta t) = v(t) + \delta t \cdot a(t) + \frac{1}{2}\delta t^2 \cdot b(t) + \frac{1}{6}\delta t^3 \cdot c(t) + \ldots$$

MD Simulation: The Finite Difference

Verlet 积分方法：下一时刻的位置可以从当前时刻及前一时刻的位置求出。

$$r(t+\delta t) = r(t) + \delta t \cdot v(t) + \frac{1}{2}\delta t^2 \cdot a(t) + \ldots$$

$$r(t-\delta t) = r(t) - \delta t \cdot v(t) + \frac{1}{2}\delta t^2 \cdot a(t) - \ldots$$



以上两公式相加：

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 \cdot a(t)$$

当前时刻的速率可以从坐标位置求出：

$$v(t) = [r(t+\delta t) - r(t-\delta t)] / 2\delta t$$

MD Simulation: The Initial State

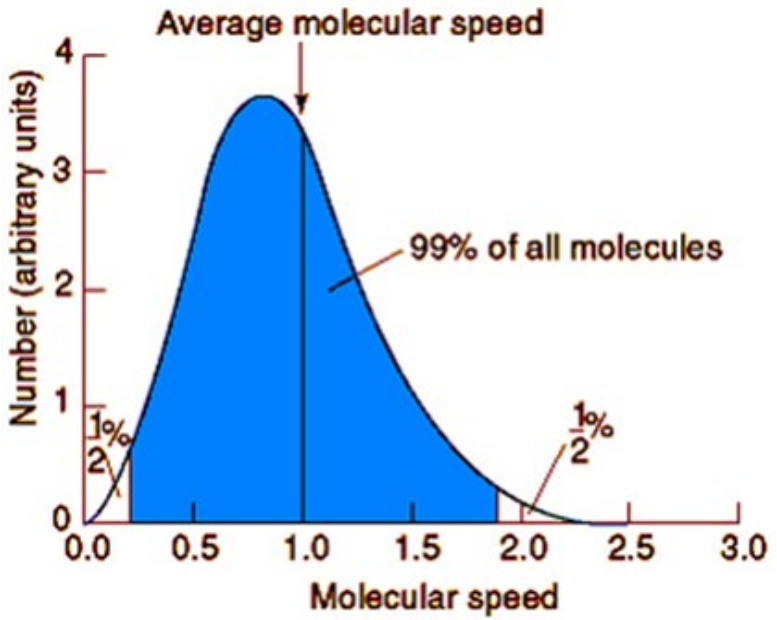
- t=0的起始结构：由实验方法测得的结构或由理论方法导出的模型结构。

- t=0的起始速度：在给定温度下取Maxwell-Boltzmann分布：

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

- t=0的加速度可由体系的势能函数求出：

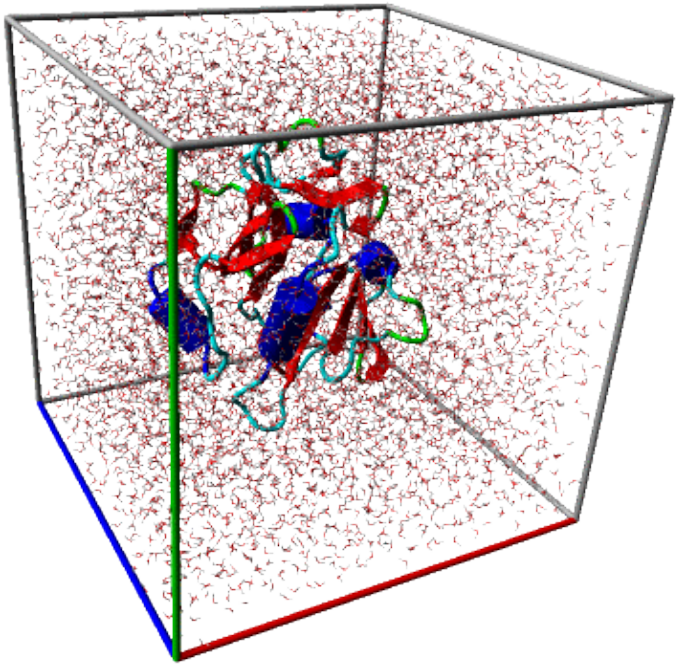
$$F_i = -\frac{\partial E}{\partial \mathbf{r_i}} = m_i \frac{d^2 \mathbf{r}_i}{dt^2}$$



MD Simulation: The Parameters

进行分子动力学模拟需要设置的基本参数：

- 模拟时间尺度(total length): 100 ps – 1000 ns
- 合适的步长(step size)：1-2 fs
- 体系的平衡温度(T)和压力(P)
- 升温过程的控制: from 0 *K* to 300K
- 溶剂模型的选择



MD Simulation: The Length

分子体系各种运动所对应的时间尺度：

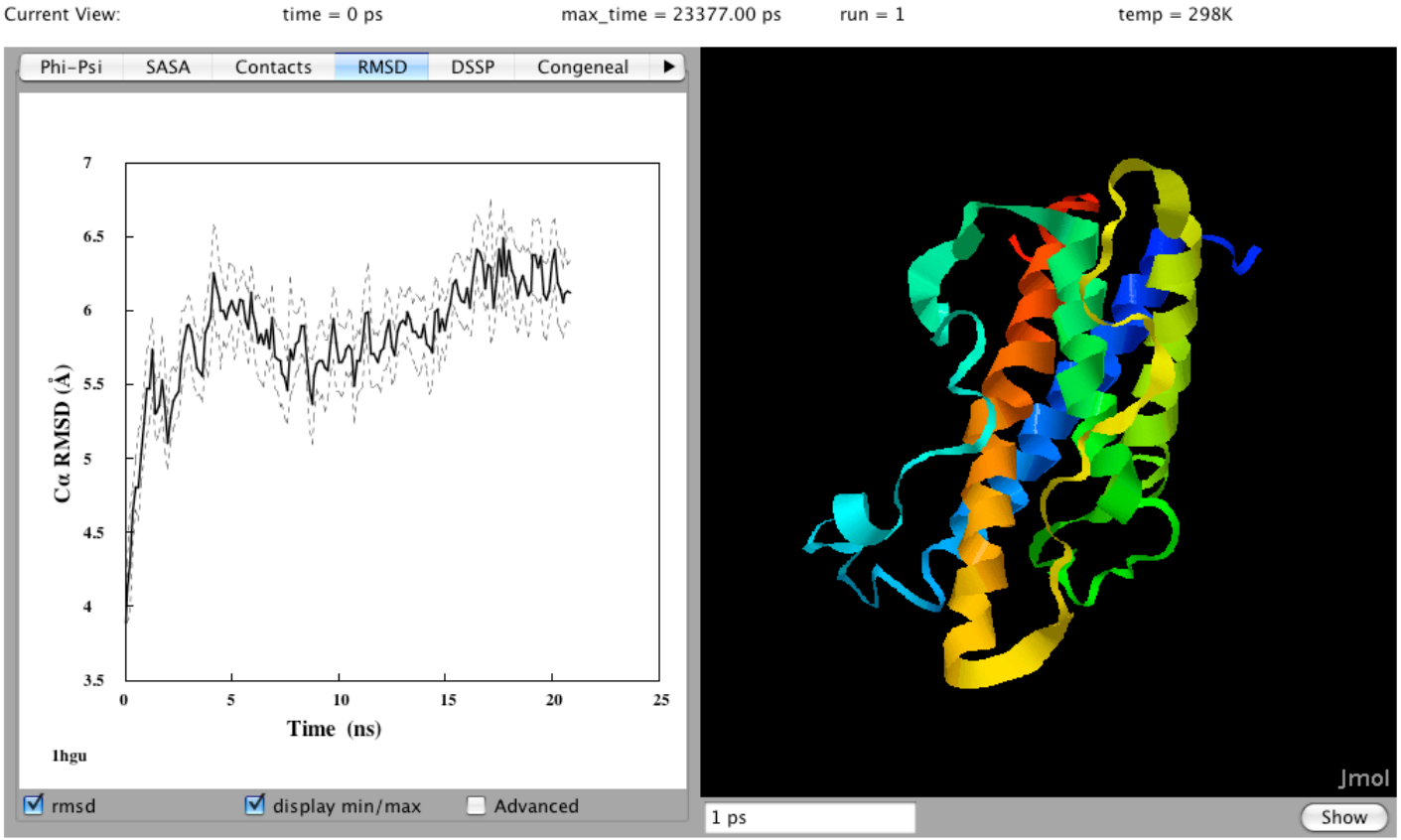
Time scale	Amplitude	Description
short: femto~pico-second		
10^{-15} - 10^{-12} s	0.001 - 0.1 Å	- bond stretching, angle bending - constraint dihedral
medium: pico~nano second		
10^{-12} - 10^{-9} s	0.1 - 10 Å	- unhindered surface side chain motion - loop region motion, collective motion
long: nano~micro-second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

step size for MD = 1 *fs*

step size for MD = 2 *fs*

MD Simulation: The Outputs

分子动力学模拟给出体系的动力学轨迹(trajectory)，即整个体系的结构随着模拟时间的变化结果，从中可以提取体系的各种性质随模拟时间的变化结果。

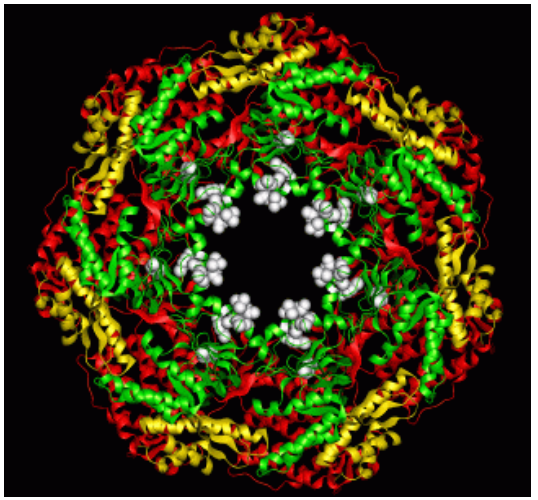


Applications of MD Simulation

- 对分子体系进行构象空间的采样,经常用于对核磁共振、X射线衍射和电镜方法测定的生物大分子结构进行优化和模拟。

- 描述分子体系在平衡状态的性质,例如:衡量蛋白质结构的稳定性、计算体系的自由能差值等。

- 提供分子体系随时间演化的动态过程,如蛋白质构象的变化、蛋白质折叠和去折叠过程、受体－配体之间的分子识别机制和生物体系中的离子输运过程等。

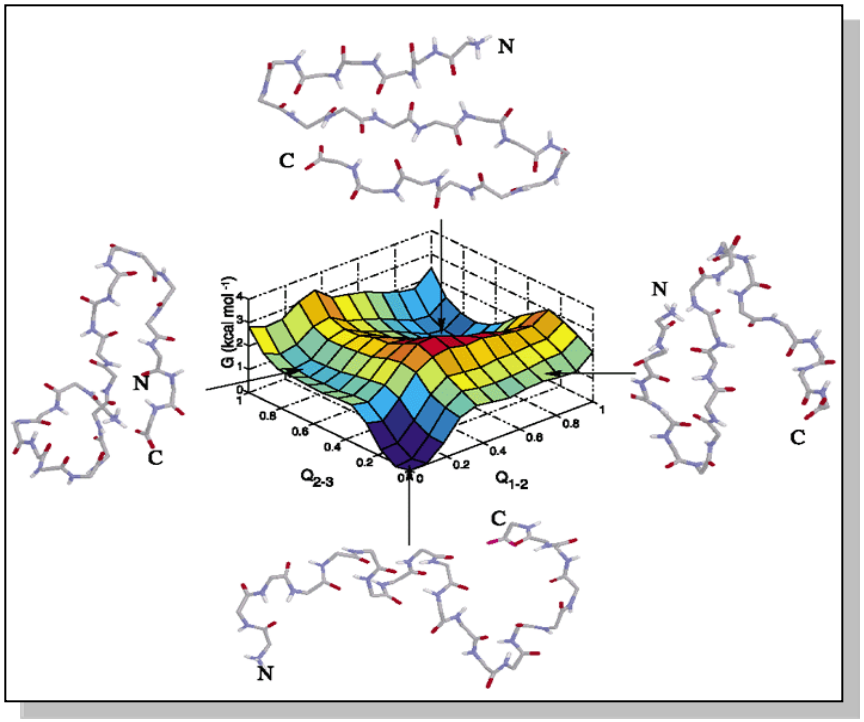


MD Simulation for Conformational Analysis

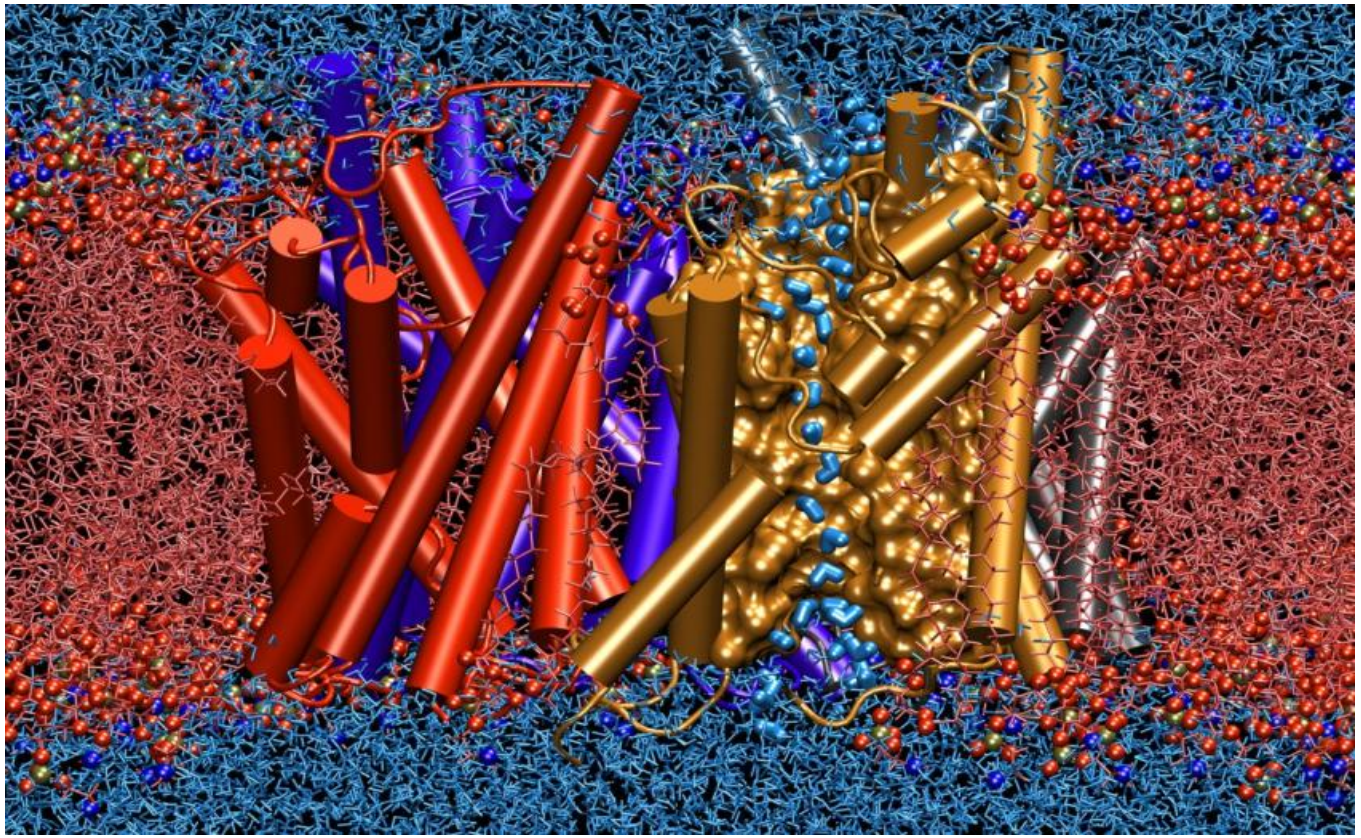
基本做法：对所研究的体系进行分子动力学模拟，在获得的轨迹上每隔一定时间间隔截取分子构象，得到构象集合并进行聚类分组。为了对构象空间进行充分采样，通常需要进行若干组平行的分子动力学模拟。

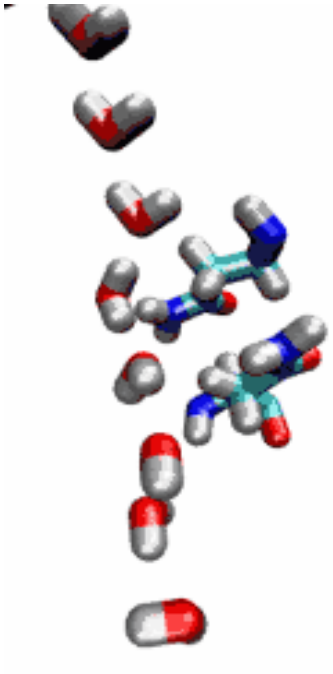
An Example: Folding simulations of a three-stranded beta-sheet peptide (20 runs, each run produced 2 x 10^5 conformations, > 2 micro-second)

M Karplus, JA McCammon, "Molecular dynamics simulations of biomolecules", *Nature Structural Biology*, **2002**, *9*, 646-652.



MD Simulation for Mechanism Study





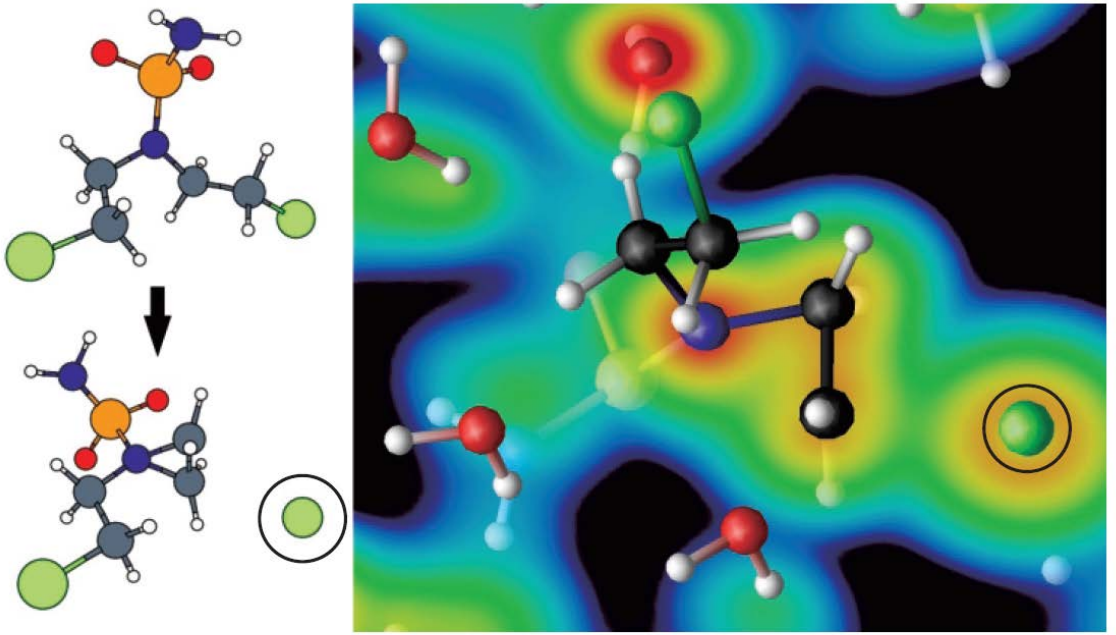
Emad Tajkhorshid et al, "Control of the selectivity of the aquaporin water channel family by global orientational tuning", *Science*, **2002**, *296*, 525-530.

Quantum Molecular Dynamics!

分子动力学模拟同样可以在量子力学的基础上进行(QM-MD)，目前在模拟简单的化学反应、液相以及固相材料的性质等方面得到应用。

Computed with *ab initio* or DFT methods

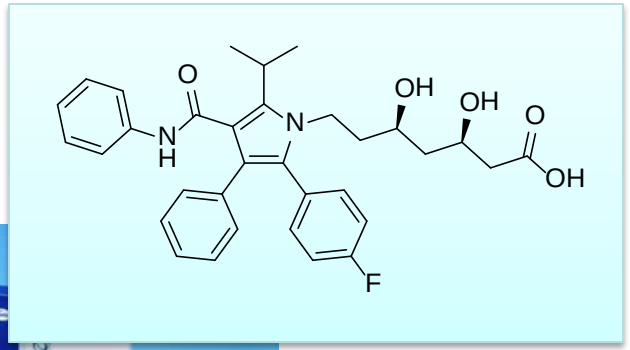
$$Fx_i = -\frac{\partial E}{\partial x_i} = m_i \frac{d^2 x_i}{dt^2}$$

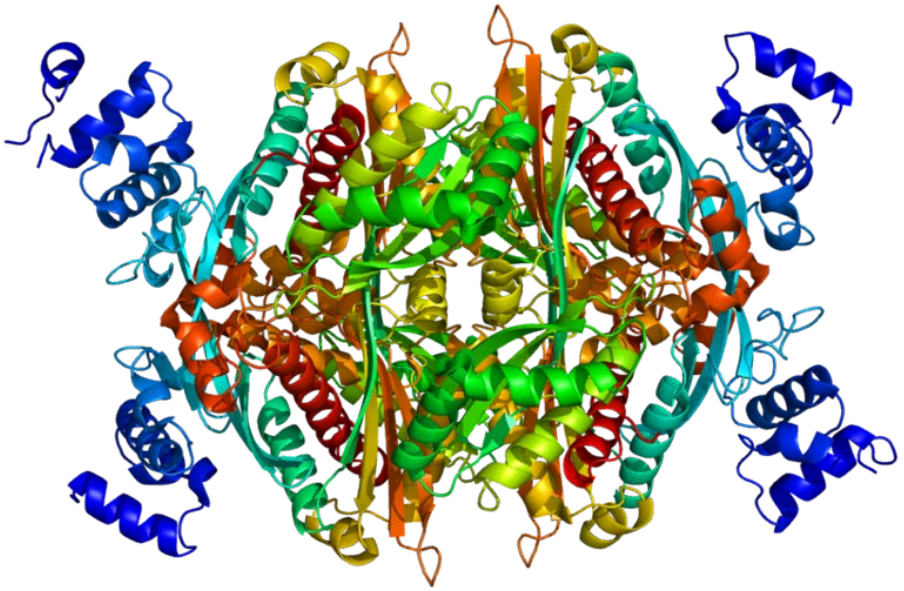


The cyclization of phosphoramide mustard in solution. (left) As the new carbon–nitrogen bond is formed, a chloride ion (circled) leaves the mustard and (right) is solvated by the surrounding water molecules.

III. Structure-Based Drug Design

药物产生药效的分子基础：药物分子与生物靶标分子立体形状互补、化学性质相匹配，二者通过各种非共价或共价作用相结合而形成稳定的复合物，从而影响生物靶标分子的功能和性质，导致机体微环境产生一系列相应的生物反应，从而产生需要的生理效应。

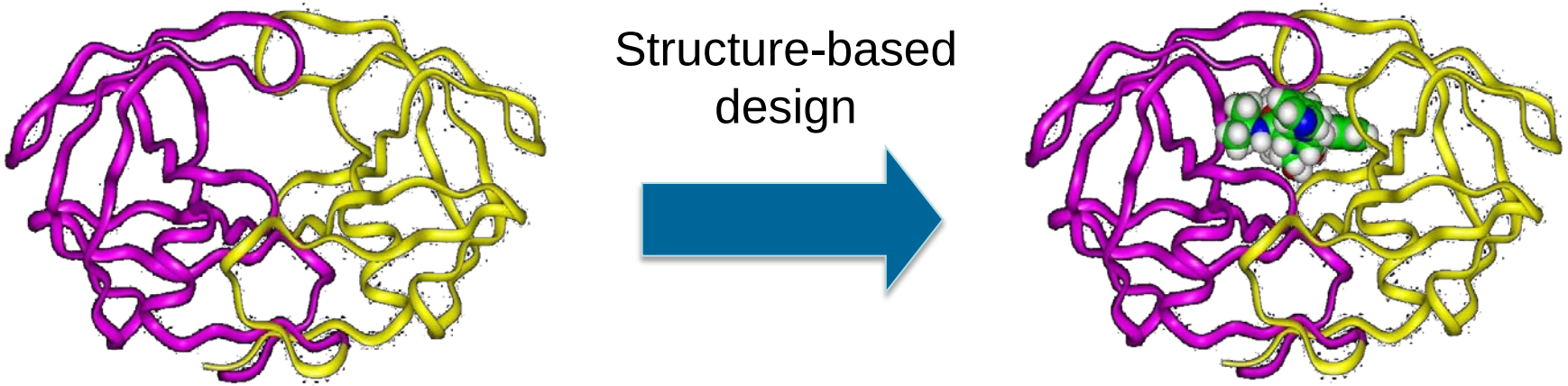




Structure of 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMG-CoA reductase)

Structure-Based Drug Design

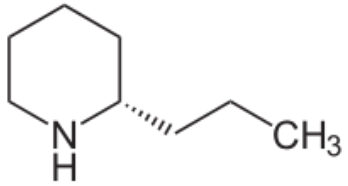
基于靶标结构的药物分子设计：如果生物靶标分子的三维结构已知，则可应用基于结构的设计方法，寻找或设计与靶标分子的结合部位立体形状互补以及化学性质匹配的配体分子。



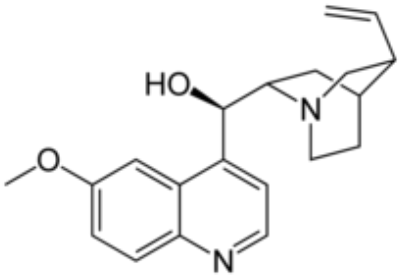
基于靶标结构的药物分子设计是分子识别的原理和方法在药物研发领域中的具体运用。

药物研发的历史：药物化学的诞生

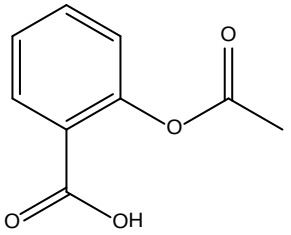
19世纪上半叶诞生了近代有机化学，提供了必要的理论基础以及技术方法，使得药物研究从直接利用天然产物阶段过渡到从天然产物中提取活性有效成分的阶段。



Coniine



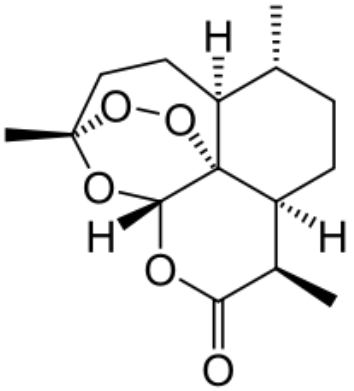
奎宁



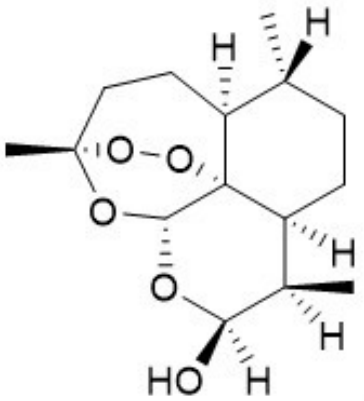
Aspirin

毒芹碱：第一个化学结构被确定的天然产物。1826年被分离出来；1870年Schiff确定其结构；1889年Ladenburg合成了该化合物。

2015 Nobel Medicine Prize!



青蒿素



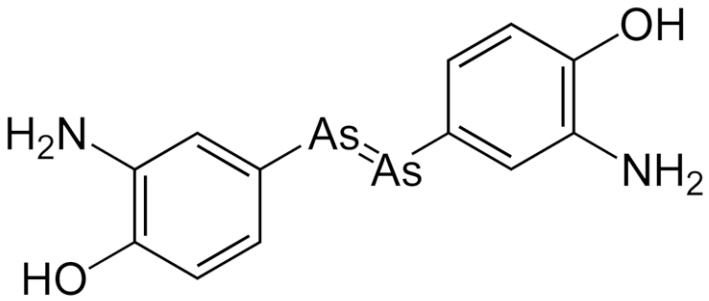
双氢青蒿素

屠呦呦 (1930 --)

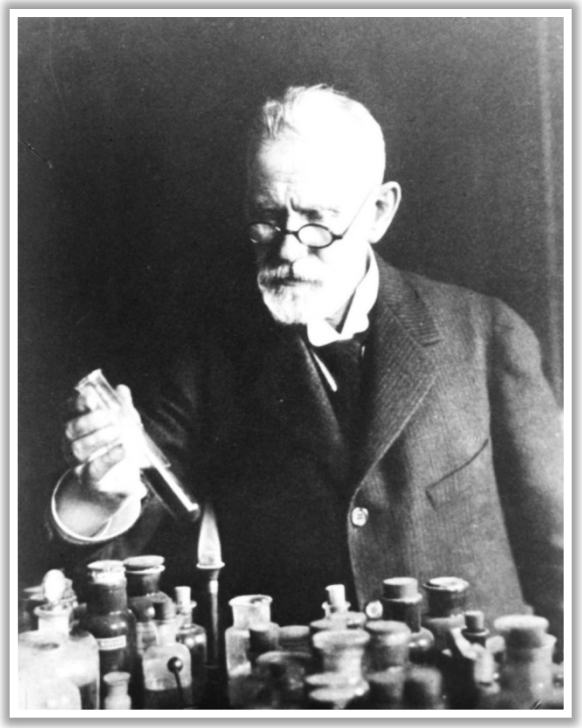
" … for her discoveries concerning a novel therapy against Malaria"

药物研发的历史：药物化学的诞生

19世纪下半叶人们逐渐发现人工合成的化学物质可以用作治疗疾病的药物(chemotherapy)，并相继提出了药物作用于特定的“受体”等概念，药物化学从此得到蓬勃的发展。

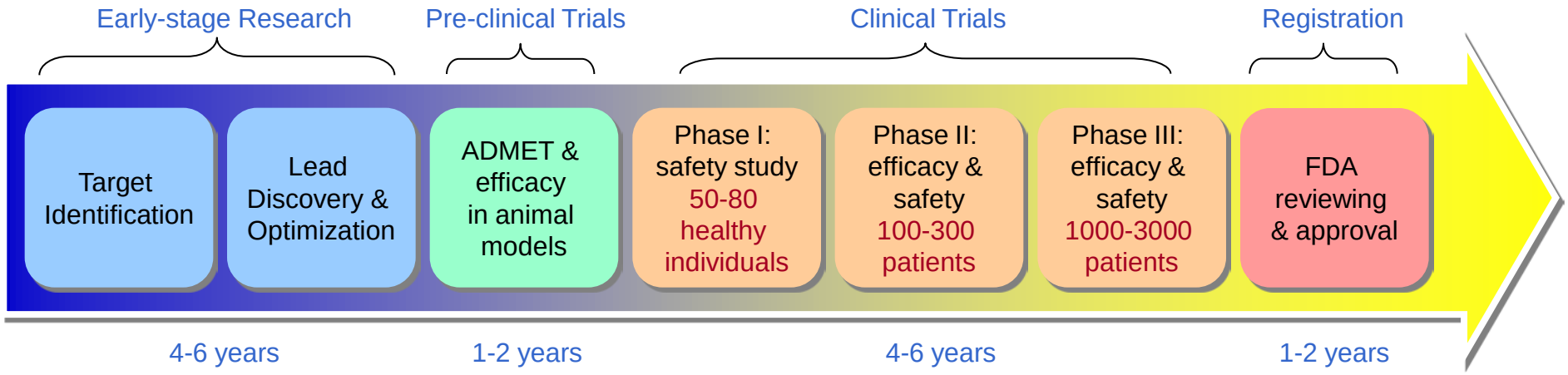


Arsphenamine (Salvarsan, also known as “606”)
The first “magic bullet” for treating diseases



Paul Ehrlich (1854-1915)
Winner of the Nobel Prize in Physiology/ Medicine in 1908

当代药物研发的典型流程



- 经过实验室研究和临床前研究的每10000个化合物中，平均有5个化合物可以进入后续的临床试验，最终可能有1个化合物成功上市。
- 新药研发的整个流程平均需要10年以上，花费约8-10亿美元，而且该数字还在逐步增长中。

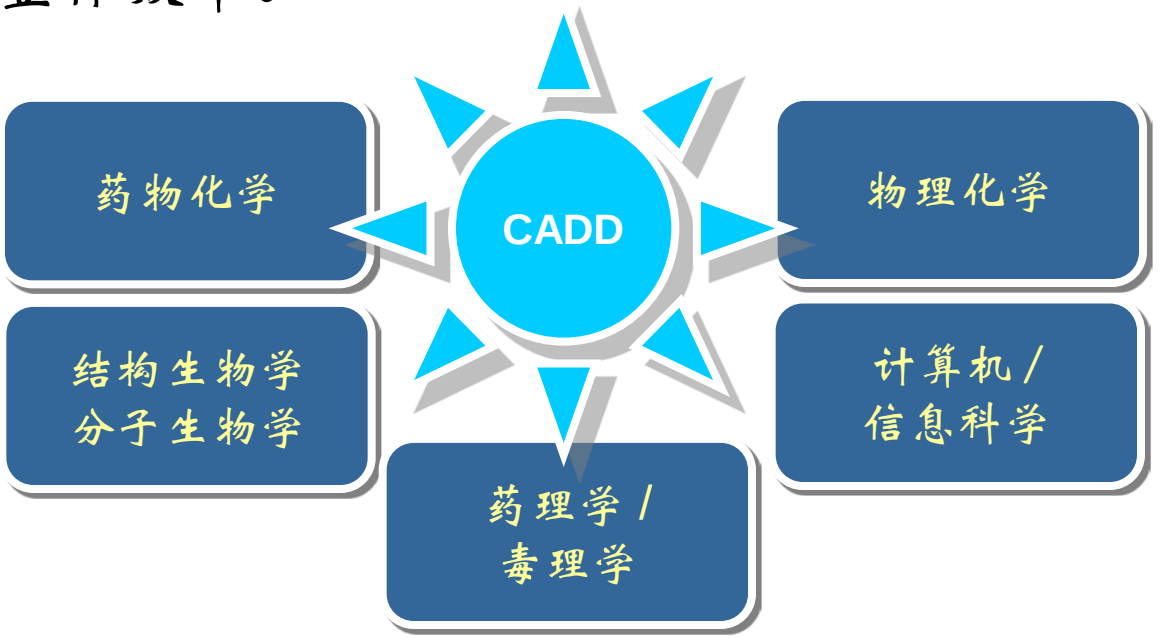
- Zambrowicz, B. P. & Sands, A. T., *Nature Rev. Drug Discov.* **2003**, *2*, 38-51.
- Lipinski, C.; Hopkins, A. *Nature* **2004**, *432*, 855-861.
- Ashburn, T. T. et al. *Nat. Rev. Drug Discov.* **2004**, *3*, 673-683.

传统的新药研发方法中存在的弊病

- 先导化合物的发现主要依靠偶然发现，对先导化合物的改造和优化主要依靠trial-and-error。但是这种凭运气的发现充满了未知性，造成新药发现中存在效率低下，成本高昂的普遍缺点。
  - 先导化合物(Lead Compound)是指具有一定药理活性的、可通过结构改造来优化其药理特性而可能导致药物发现的特殊化合物。
  - 先导物的来源主要有从天然产物中提取、偶然发现、随机筛选、老药新用等。
- 上世纪九十年代初出现的组合化学和高通量筛选由于盲目性较大，并没有达到人们预期的效果，促使人们重新回归理性。
- 现代药物发现倾向于采用理性方法，在掌握与疾病相关的靶标或生理过程以及药物作用机理的基础上，进行有目的的药物设计。

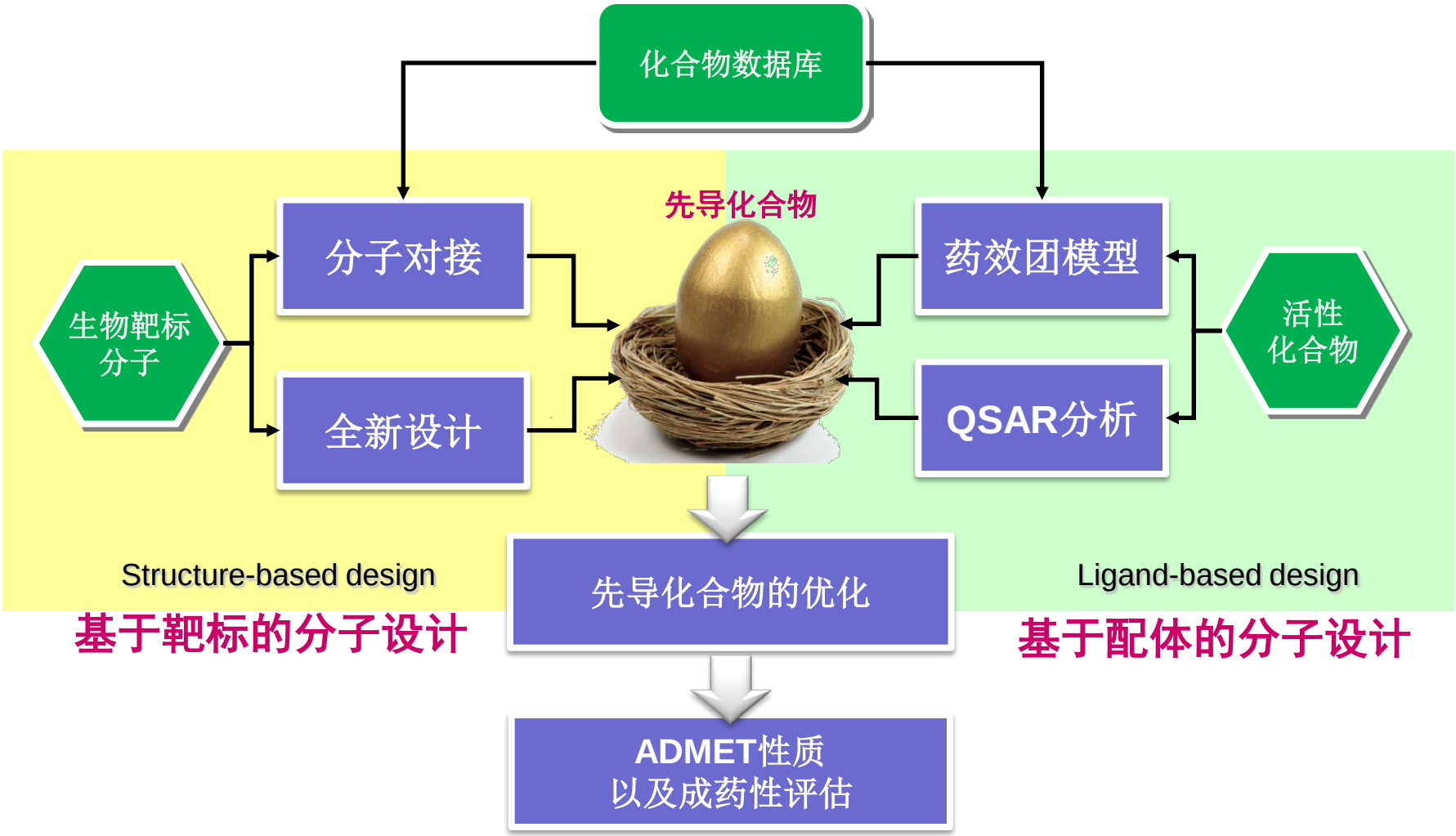
什么是计算机辅助药物设计？

计算机辅助药物设计(Computer-Aided Drug Design, CADD)以计算机为工具，依靠分子模拟、化学信息学和生物信息学等技术手段，对药物分子的结构-活性关系、疾病的生物靶标分子的性质等进行计算、分析和预测，从而指导新型药物的发现，减少盲目性，提高新药研发的整体效率。



计算机辅助药物分子设计是一门高度交叉的学科

计算机辅助药物设计中的常用方法



基于靶标结构的药物分子设计方法

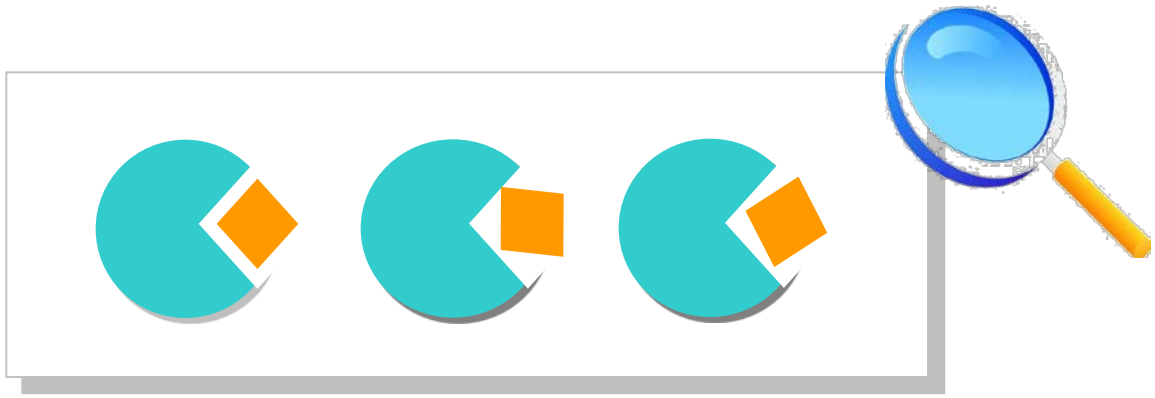
基于靶标结构的药物分子设计：如果生物靶标分子的三维结构已知，则可应用基于结构的设计方法，寻找或设计与靶标分子的结合部位立体形状互补以及化学性质匹配的配体分子。

基于结构的药物设计中的基本方法：

- 分子对接(molecular docking)：Find the proper ones among the known molecules!
- 从头设计(*de novo* design)：Build up whatever is proper from the scratch!

Molecular Docking

分子对接(molecular docking)这种理论计算方法用来研究一个配体分子与另一个受体分子之间的最佳结合模式。这样的结合模式可以用来评价两个分子之间的相互作用以及预测亲合性的强弱等性质。



分子对接是基于结构的药物设计中应用最普遍的方法

Docking: The Lock-and-Key Model

The lock-and-key model: rigid docking

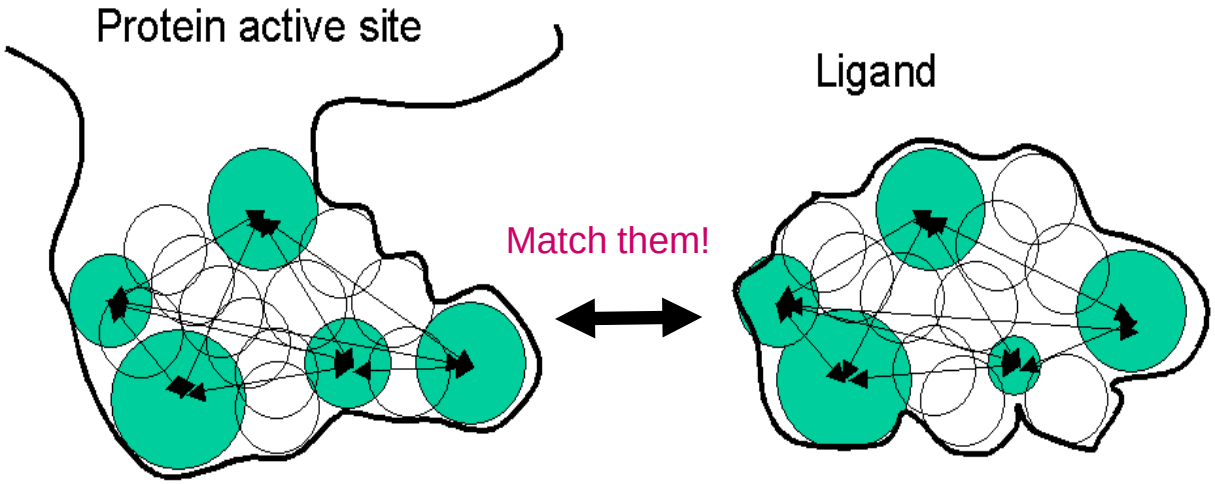
"The specificity of an enzyme (the lock) for its substrate (the key) arises from their geometrically complementary shapes"

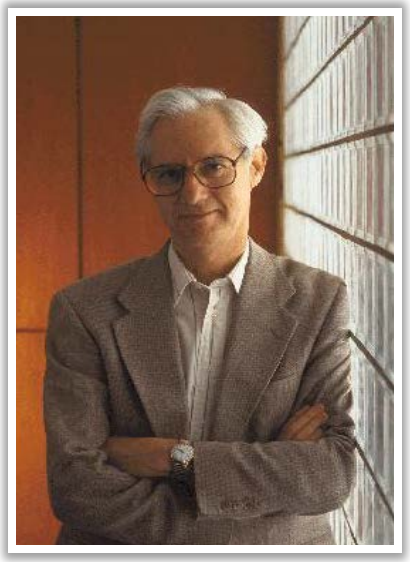
(E. Fischer, *Ber Dtsch. Chem. Ges.*, **1894**, *27*, 2985)

Hermann E. Fischer (1852-1919)
Winner of the Nobel Chemistry Prize in 1902

分子对接程序DOCK

世界上首例实现自动分子对接的计算机程序是由UCSF大学Kuntz教授小组发展的DOCK程序(http://dock.compbio.ucsf.edu/)。





Prof. Irwin Kuntz

Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. "A geometric approach to macromolecule-ligand interactions", *J. Mol. Biol.* **1982**, *161*(2), 269-288.

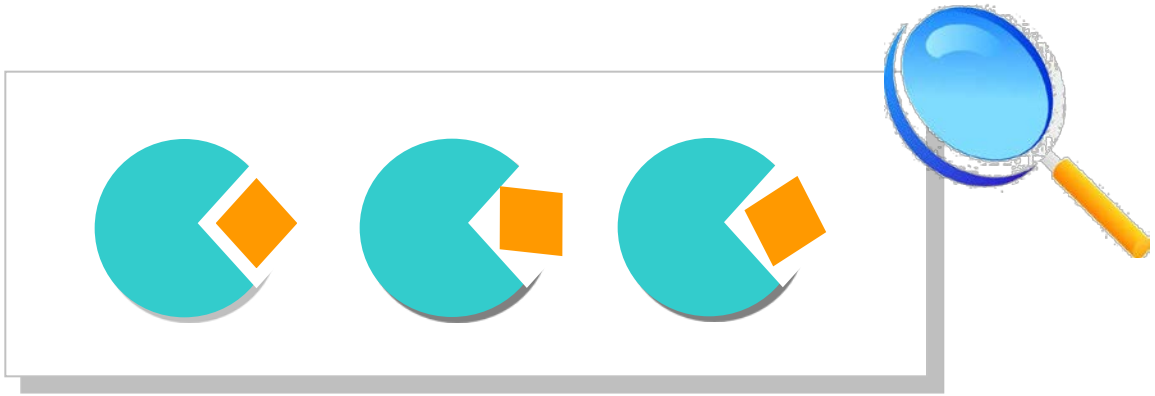
Docking: The Induced-Fit Model

- Induced Fit: flexible docking
  - Flexible ligand + rigid protein
  - Flexible ligand + protein with flexible side chains
  - Flexible ligand + fully flexible protein
  - Flexible ligand docked into several representative conformations of the target protein

- Popular molecular docking programs today
  - **DOCK** (University of California San Francisco)
  - **AutoDock** (Scripps Institute)
  - **GOLD** (Cambridge Crystallographic Data Center)
  - **GLIDE** (Schrodinger Inc.)

Essential Problems in Molecular Docking

采样问题(sampling problem):在考虑受体分子结合部位空间限制的情况下，对配体分子可能采取的构象进行采样。生成配体分子构象的方法可以采用蒙特卡洛方法、基因算法以及分子动力学等方法。

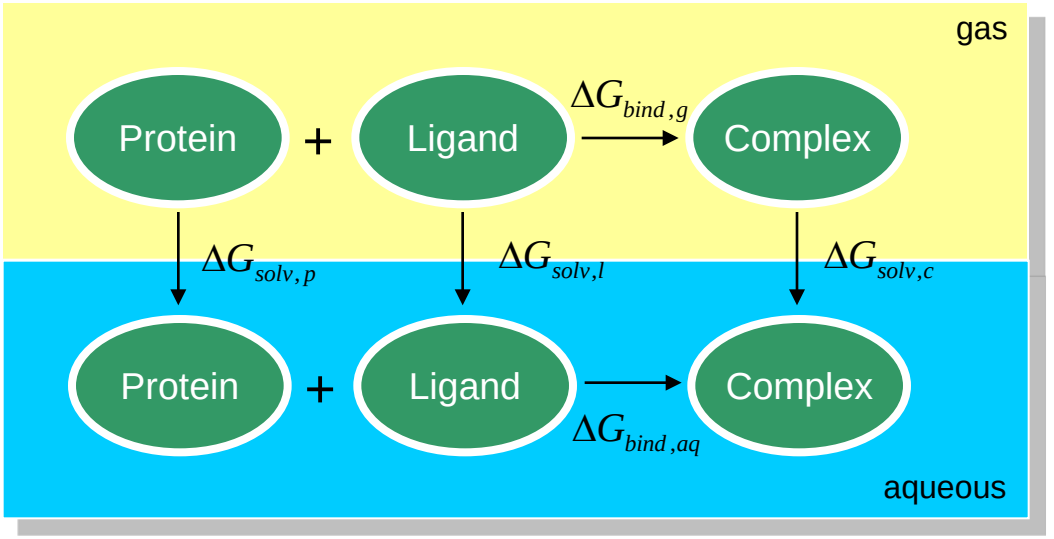


Essential Problems in Molecular Docking

打分问题(scoring problem): 计算配体分子与受体分子的结合能，可以采用分子力场、量子力学计算或经验方法，需要考虑分子间相互作用(离子对、氢键、偶极-偶极、范德华相互作用等)、溶剂化效应等因素。

$$K_a = \frac{[P \cdot L]}{[P][L]}$$

$$\Delta G^0_{binding} = -RT \ln K_a$$

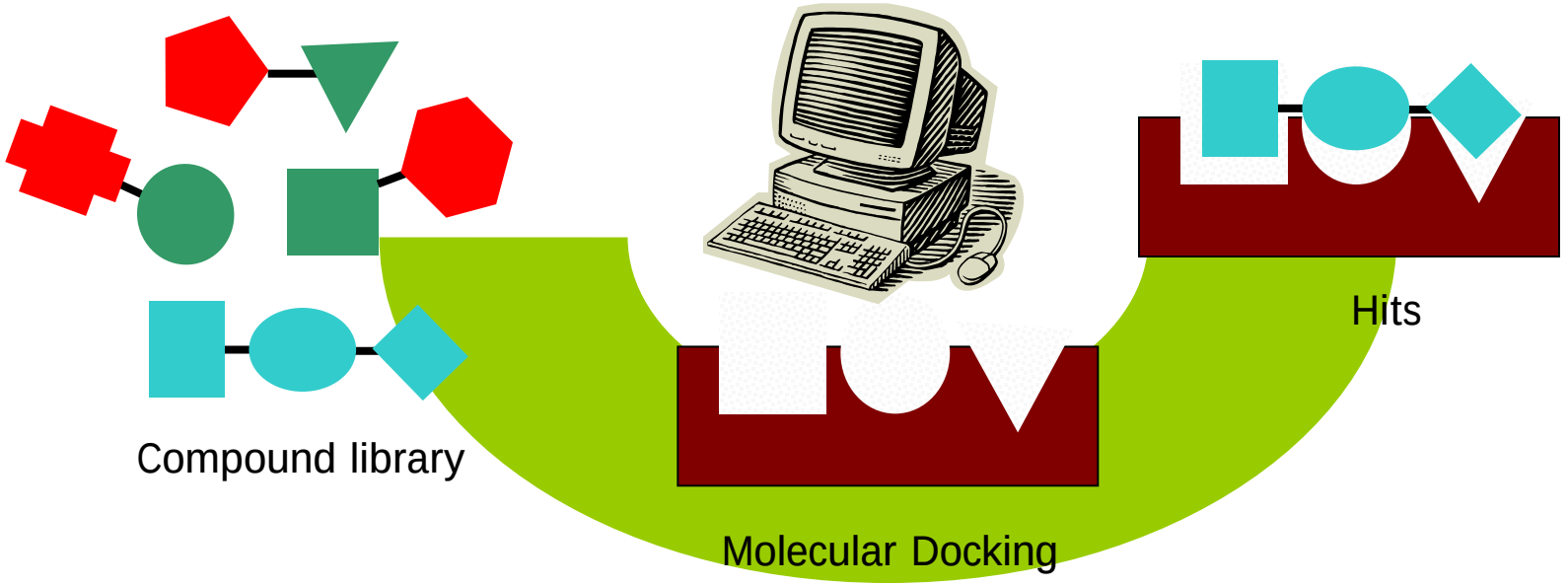


蛋白-配体结合过程的热力学循环

$$\Delta G_{bind,aq} = \Delta G_{bind,gas} + (\Delta G_{solv,c} - \Delta G_{solv,p} - \Delta G_{solv,l})$$

基于分子对接的虚拟筛选

如果药物作用靶标分子的三维结构已知，虚拟筛选(virtual screening)通过分子对接技术考察一批化合物与靶标分子结合的能力，预测候选化合物的生理活性。

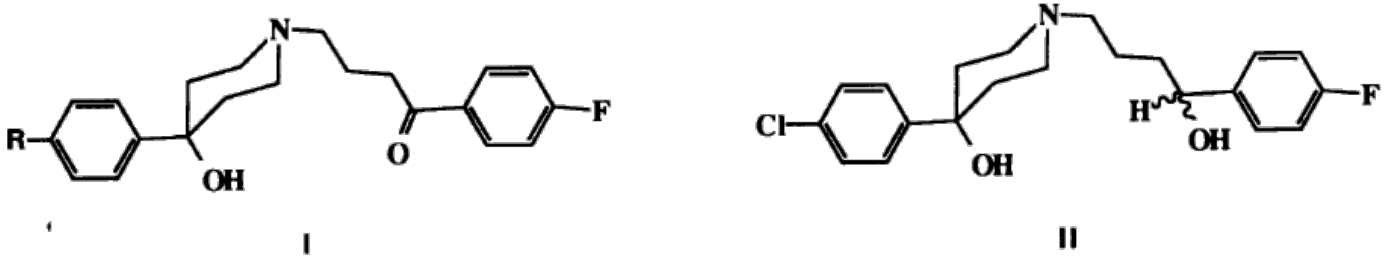


虚拟筛选已成为在药物设计中发现新型先导化合物广泛使用的有效方法，与传统的盲目筛选相比可以大大提高成功率。

A Successful Example

HIV-1蛋白水解酶新型抑制剂的虚拟筛选：

利用DOCK3.0程序，从剑桥晶体结构数据库(Cambridge Structural Database)中约10,000个分子中选出一个感兴趣的化合物bromperidol。其类似物haloperidol是已知的抗精神类药物。

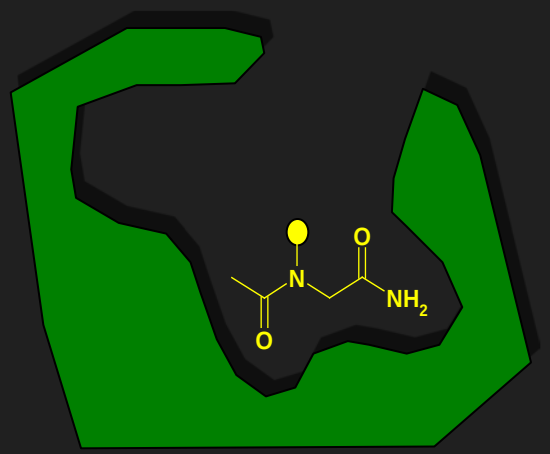


(1) Bromperidol (R=Br); Haloperidol (R=Cl). (2) A derivative of haloperidol

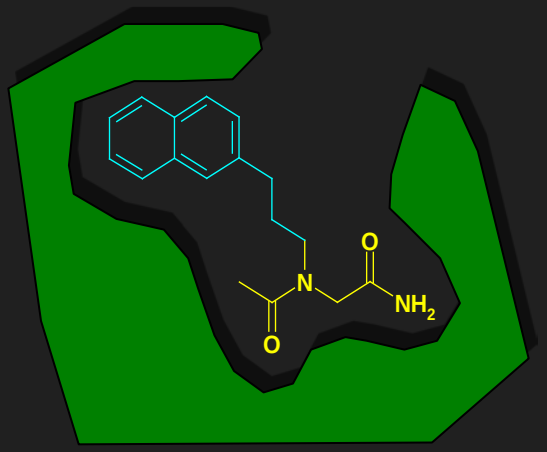
随后对该化合物进行了化学合成、体外酶抑制剂活性和细胞活性试验。证明该化合物的抑制常数(K_i)约为100 μM。虽然这些化合物的活性并不高，但这是DOCK程序在虚拟筛选中的首例成功应用。

R L Desjarlais, et al. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease, *Proc. Natl. Acad. Sci.* **1990**, *87*, 6644- 6648.

药物分子的从头设计 (*de novo* design)

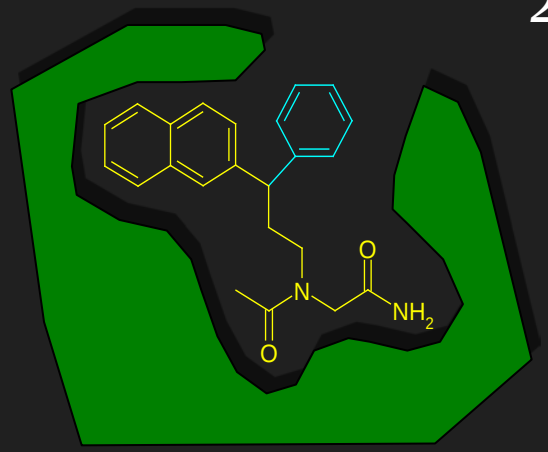


1. 放入“种子”片段



2. “生长”出另一个片段

3. 继续生长直至填满结合部位



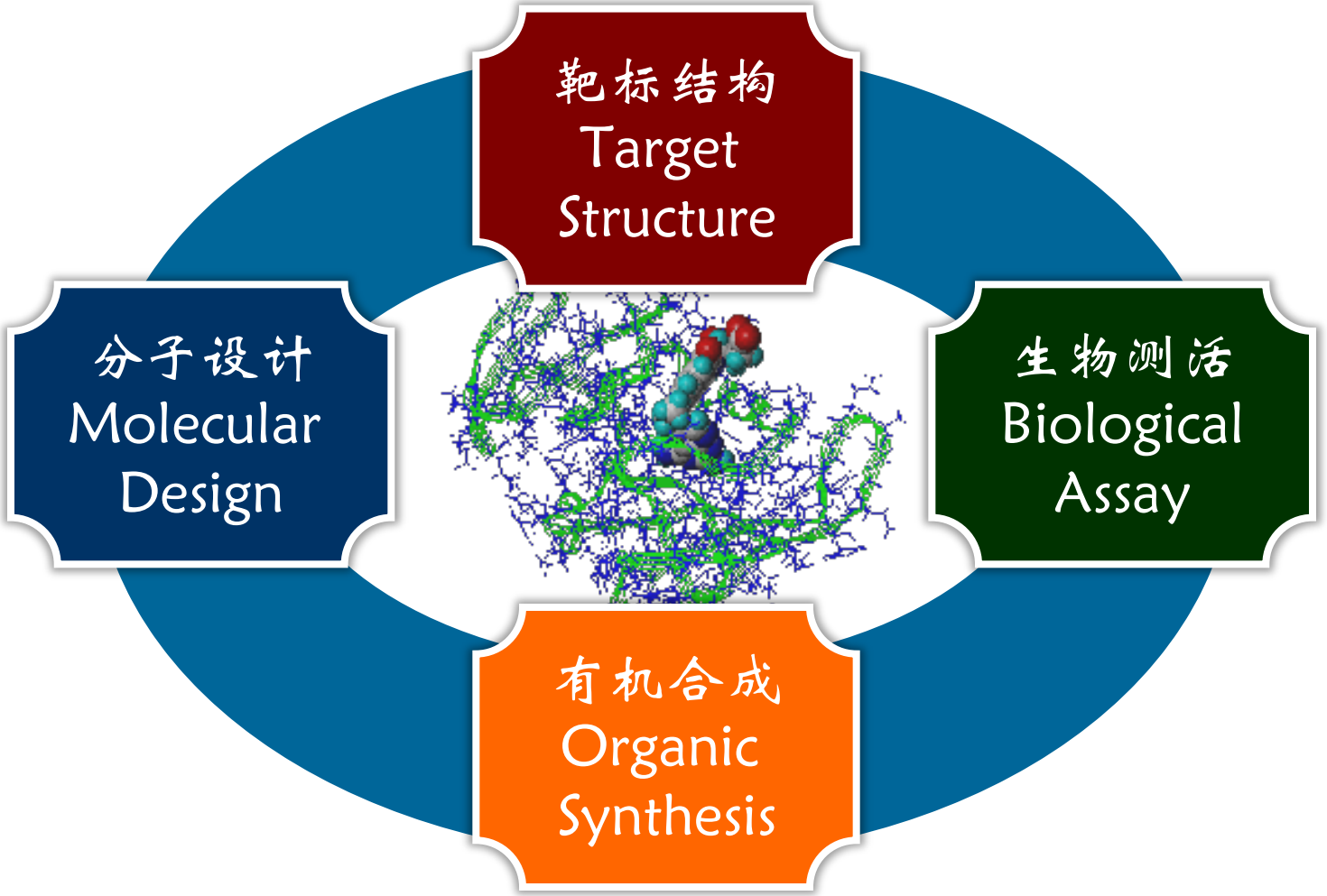
4. 对结果进行聚类分析，选择出最好的配体结构。

比较两类先导化合物的发现方法

- 虚拟筛选：从已知有机化合物中寻找
  - 优点：有可能迅速获得化合物的样品进行生物活性测定。
  - 缺点：受限于已有的知识；所筛选的化合物库中未必包含有适合给定靶标的配体分子；即使包含有这样的分子，分子对接程序也未必一定能够把它们识别出来。

- 从头设计：创造全新的分子
  - 优点：不受现有知识的约束；该方法既可用于发现先导化合物，也可用于对先导化合物的结构进行优化。
  - 缺点：所设计出来的分子可能不易进行有机合成。

Flowchart of Structure-Based Drug Design

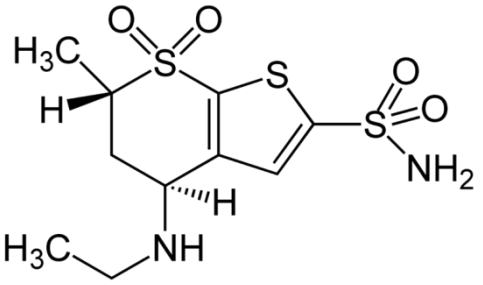


Drugs by Rational Design

药物分子设计技术已经应用到了药物研究的各种环节，提高了药物研发效率。据统计由于药物分子设计技术的广泛应用，新药研发的周期缩短了0.9年，直接研发费用降低了1.3亿美元。应用理论方法设计并最终进入临床研究的化合物已有40余个，并有若干成功上市的案例。

计算机辅助药物设计成功例子：

药物	靶标	公司
Dorzolamide	碳酸酐酶	Merck
Saquinavir	HIV蛋白水解酶	Roche
Relenza	神经氨酸苷酶	GlaxoSmithKline
AG85/337/331	胸腺核酸合成酶	Agouron
Ro466240	凝血酶	Roche
Gleevec	Abl-酪氨酸激酶	Novartis



Dorzolamide: The first marketed drug by structure-based design

Acknowledgments

- 感谢研究生部和分院教育基地的大力协助！
- 特别致谢中科院上海有机化学研究所陈敏伯研究员、黎占亭研究员、北京大学来鲁华教授以及华东理工大学唐赟教授提供本教程的部分素材！
- 感谢各位同学的积极参与和支持！