

Conceitos de Modelação Molecular

O que é ?

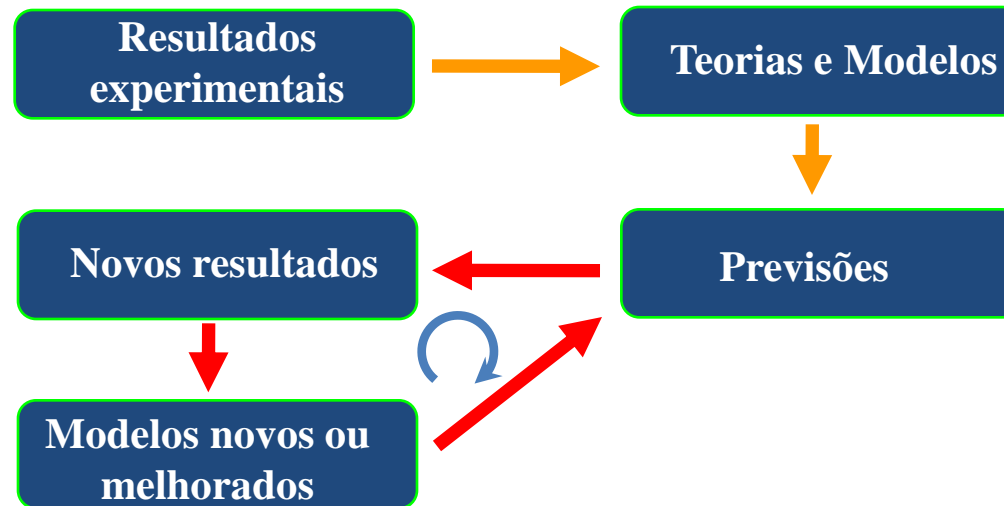
- Representação idealizada de sistemas moleculares com o fim de explicar e prever as suas propriedades físico-químicas
- Construção, análise e simulação de modelos computacionais de moléculas

Para que serve ?

- Cálculo de geometria molecular
- Visualização de propriedades moleculares
- Energética molecular
- Análise e pesquisa conformacional
- Previsão da reactividade/actividade
- Dinâmica molecular

- Estudo de propriedades inacessíveis à experimentação (exemplo: dinâmica de proteínas à escala do *picossegundo*)
- Previsão de novas propriedades moleculares (estudo do efeito de mutações)
- Previsão da interacção entre moléculas (*docking*)
- Previsão da estrutura tridimensional de proteínas (*folding*)

Como se faz ?



A modelação biomolecular baseia-se largamente em:

- Informação estrutural
 - Cristalografia de raios X
 - NMR (ressonância magnética nuclear)
 - Métodos espectroscópicos
- Modelos físicos
- Bancos de dados (estruturas, sequências)

Modelos

- Representação simplificada da realidade
- Virtualização
- Capacidade de previsão
- Analogias físicas

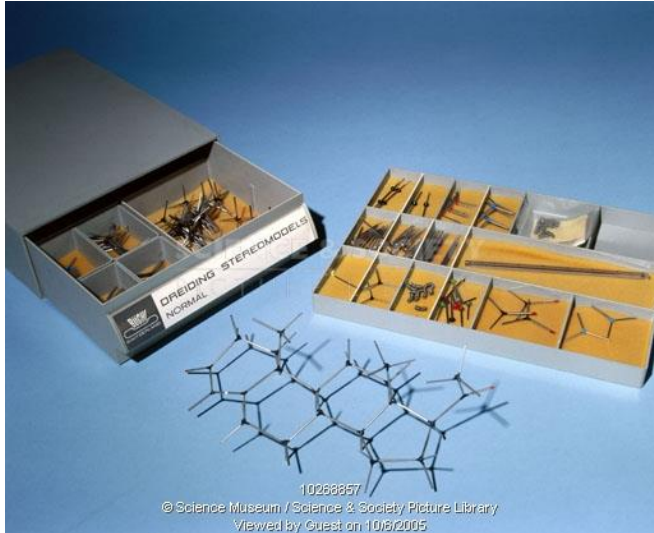


Modelo de molécula diatómica

História

- A modelação molecular tornou-se possível através do desenvolvimento das técnicas de determinação da estrutura molecular, particularmente a cristalografia de raios X, a partir do início do séc. XX
- A mecânica quântica trouxe uma descrição matemática de átomos e moléculas capaz de descrever com grande rigor as propriedades de sistemas pequenos e fornece o enquadramento teórico para a teoria da ligação química e estrutura electrónica
- Desenvolvimento dos métodos de simulação *circa* 1950 com a construção dos primeiros computadores digitais
- Até meados dos anos 60 do séc. XX a visualização era feita a usando modelos físicos de moléculas (Dreiding, CPK). A partir desta altura o desenvolvimento do grafismo computacional passa a permitir observar as moléculas directamente no écran do computador

Modelos físicos



Modelos de Dreiding

Modelos CPK (Corey & Pauling)



Molecular Models of Amino Acids, Peptides, and Proteins

ROBERT B. COREY AND LINUS PAULING

Gates and Crellin Laboratories of Chemistry,* California Institute of Technology, Pasadena, California†

(Received August 25, 1952)

A set of accurate scale models has been developed for use in studies of the structures of amino acids, peptides, and proteins. Models representing atoms or groups of atoms built from hard wood to the scale 1 in. = 1 Å are connected by a clamping device which maintains desired molecular configurations. These accurate models have been used as substitutes for calculation in investigations of the probable configuration of the polypeptide chain in proteins. Analogous models constructed of rubber-like plastic to the scale 1 in. = 2 Å and connected by snap fasteners are designed for qualitative studies of protein structure.

INTRODUCTION

FOR several years we have been working at the California Institute of Technology on the development of scale models for representing the molecules of amino acids and related compounds as an aid in our attack on the problems of the structure of proteins. The development of these models has closely paralleled our x-ray diffraction studies of amino acids, peptides, and proteins. Our first models were constructed in accordance with the dimensions experimentally obtained from complete and accurate determinations of the structures of amino

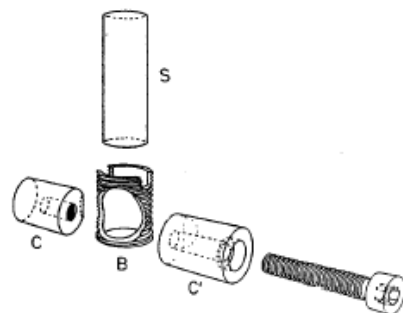


FIG. 1. An exploded view of the device used for clamping the atom models rigidly to the connecting cylindrical shafts. The threaded bushing *B* imbedded in the atom (see Fig. 2) is intersected on one side by a hole drilled in the atom and containing the two clamping (*C, C'*) jaws. The latter are cut away so as to grip firmly the $\frac{1}{8}$ -in. steel shaft *S* (see Fig. 3 and following figures).

acids. They were progressively revised and extended as more accurate data from peptides and additional amino acids became available. Recently we have been engaged in a study of the structure of proteins which has led us to propose several new configurations of the polypeptide chain. Because these proposed configurations are explicitly described in terms of atomic coordinates, their occurrence in proteins can be tested by com-

* Contribution No. 1731.

† Aided by a grant from the National Foundation for Infantile Paralysis.

parison of their theoretically calculated x-ray diffraction patterns with those derived experimentally from the proteins themselves.¹ In this study our molecular models have been of great assistance, especially in the recognition of sterically probable configurations of the polypeptide chain and the rejection of sterically improbable ones.²

With the continued refinement of x-ray diffraction measurements and the accumulation of new structural data, future revision of some dimensions will doubtless be desirable and models of additional atoms and radicals will be designed and constructed. Nevertheless the demonstrated usefulness of these models in their present form seems to justify a description of them at this time.

REQUIREMENTS AND SPECIFICATIONS

In models designed for studying probable molecular configuration and intermolecular packing the van der

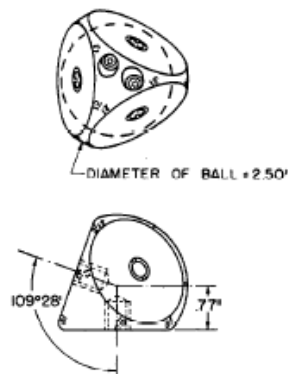


FIG. 2. A drawing of the tetrahedral carbon atom. The positions of two of the four clamps and of fiducial marks are indicated. Each face is 0.77 in. from center of ball. The ends of the steel bushings are shown in each face.

¹ Pauling, Corey, and Branson, Proc. Natl. Acad. Sci. U. S. 37, 205 (1951); L. Pauling and R. B. Corey, Proc. Natl. Acad. Sci. U. S. 37, 235, 241, 251, 256, 261, 272, 282 (1951).

² L. Pauling and R. B. Corey, Proc. Natl. Acad. Sci. U. S. 37, 729 (1951); Proc. Natl. Acad. Sci. U. S. 38, 86 (1952).

Waals radii of the atoms should conform to the intermolecular distances found in crystals and in noncrystalline solids rather than to gas collision radii or even smaller radii commonly used in models of organic molecules. Intramolecular interferences between atoms attached to the same atom or group of atoms must be avoided unless there is experimental or theoretical justification for believing that the interferences occur. If the models are intended for use as a substitute for

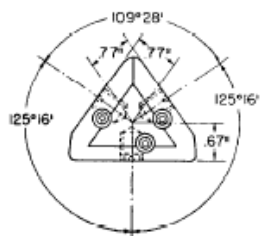


FIG. 3(a). A drawing of the double-bonded carbon atom. In the interest of clarity, only one clamp is shown in the view along the double-bond direction.

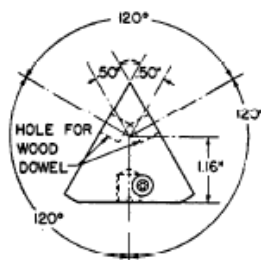
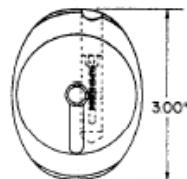
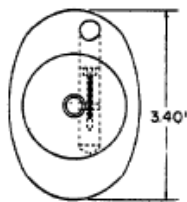


FIG. 3(b). A drawing of the aromatic carbon atom. Six atoms are fastened together permanently to form the benzene nucleus.



calculation in the examination of structures, they must be accurately built and capable of retaining their bond angles and other configurational features without deformation; the scale should be conveniently large to permit easy handling and to allow angles and distances to be measured with satisfactory precision. In order to meet even these minimum requirements, many compromises were found to be necessary.

The models are made of hard wood to the scale, 1 in. = 1 Å. Van der Waals radii, bond radii, and bond angles have the dimensions generally found in crystals of amino acids, peptides and other organic compounds. Atoms or assemblies of atoms, as the benzene nucleus and amide group, are joined together by means of short

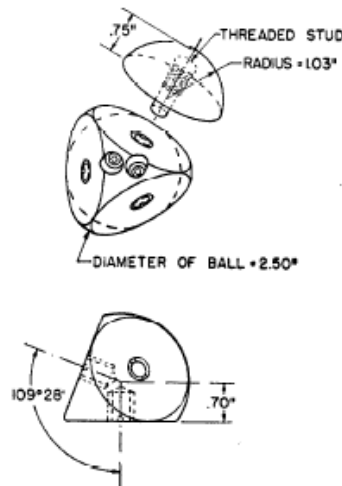


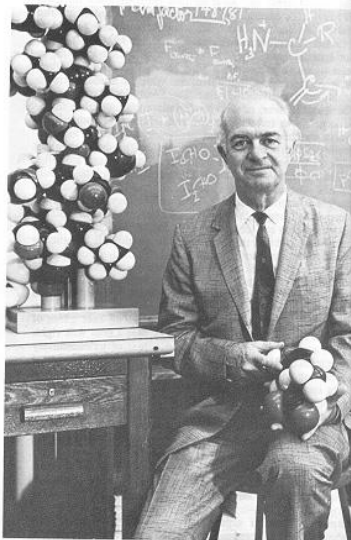
FIG. 4. A drawing of the tetrahedral positively charged nitrogen atom. For representing the neutral atom, the adapter applies an appropriate van der Waals radius to one of the bond positions.

pieces of $\frac{5}{16}$ -in. steel rod which fit into steel bushings imbedded in the atoms. For rigidly fixing the relative orientation of atoms around a bond, the bushings are locked on the rod in any desired position by means of clamping fixtures also built into the atom. An exploded assembly drawing of the steel stud, bushing, and clamping device used to connect the atoms is shown in Fig. 1. Models of carbon, nitrogen, oxygen, and hydrogen atoms, and of typical assemblies are described in the following paragraphs.

The Carbon Atom

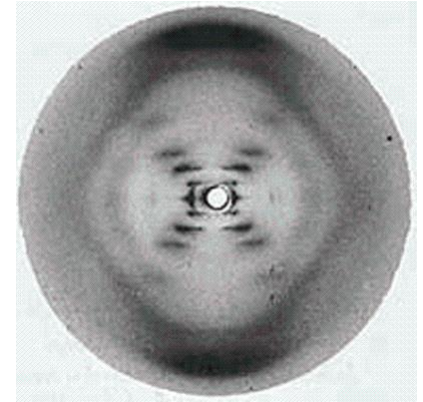
A drawing of the tetrahedral carbon atom is shown in Fig. 2. The bond radius is 0.77 in. Since the tetrahedral carbon atom is entirely surrounded by bonded atoms and so in general makes no van der Waals con-

A modelação molecular foi essencial para a descoberta da estrutura das macromoléculas biológicas

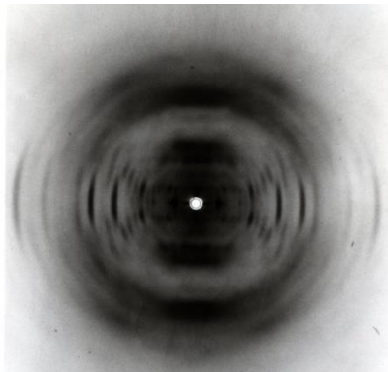


Linus Pauling with his atomic models.

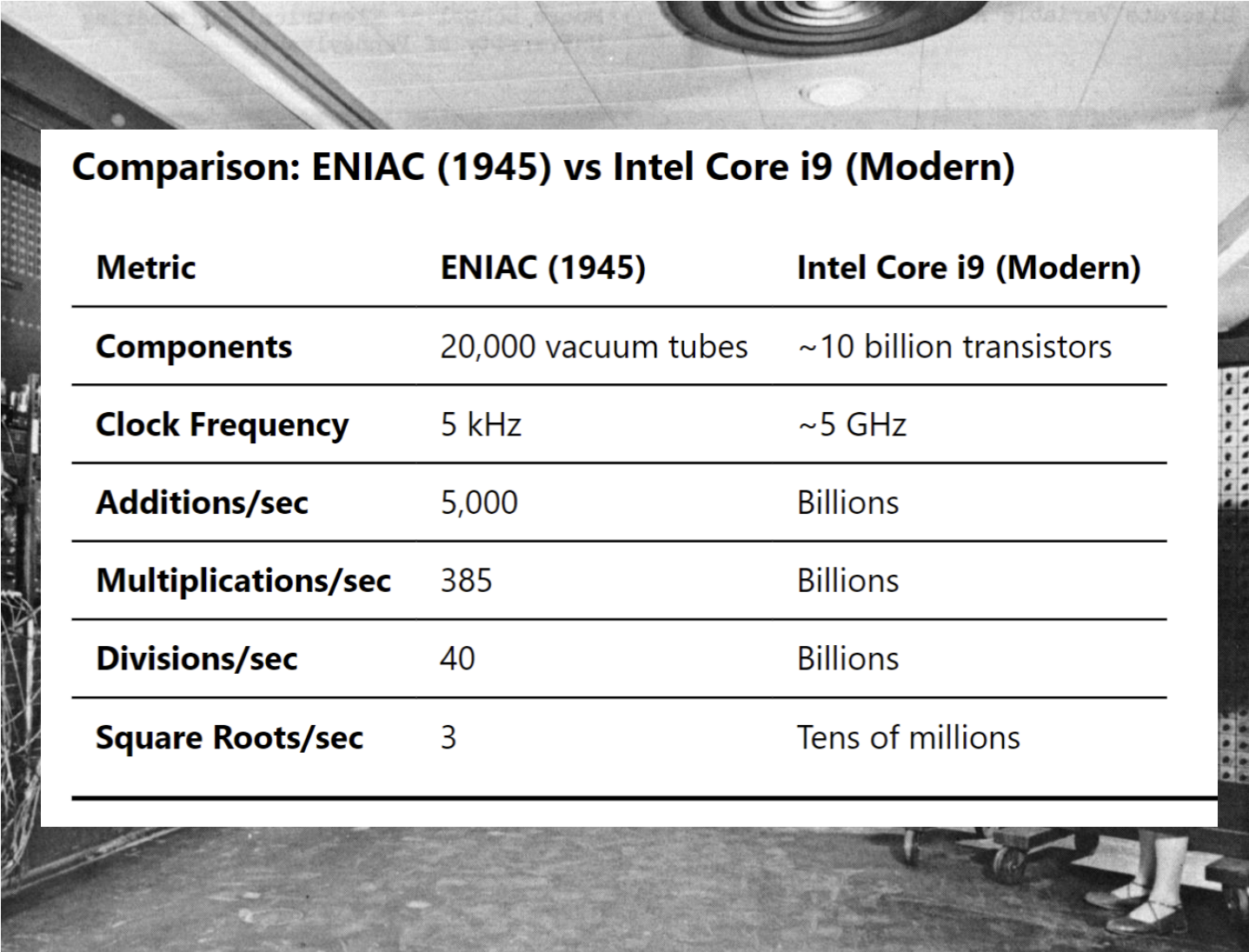
Linus Pauling descobre a estrutura da hélice alfa das proteínas em 1951. A construção de modelos foi crucial para esta descoberta



James Watson e Francis Crick usam modelos estruturais para deduzir correctamente a estrutura do DNA a partir de dados de difracção de raios X em 1953.



Primeiro computador digital: ENIAC (1945)

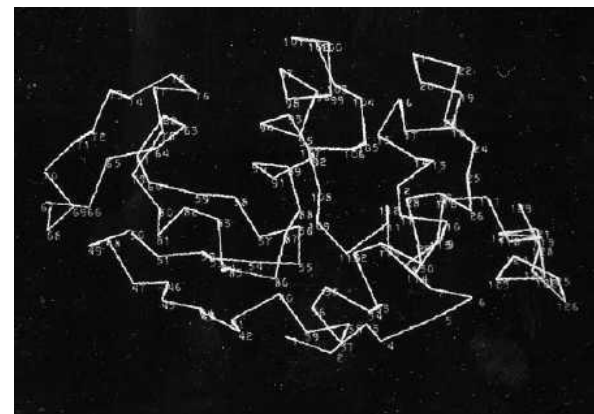


Comparison: ENIAC (1945) vs Intel Core i9 (Modern)

Metric	ENIAC (1945)	Intel Core i9 (Modern)
Components	20,000 vacuum tubes	~10 billion transistors
Clock Frequency	5 kHz	~5 GHz
Additions/sec	5,000	Billions
Multiplications/sec	385	Billions
Divisions/sec	40	Billions
Square Roots/sec	3	Tens of millions

Computação gráfica

- O primeiro sistema para a representação interactiva de gráficos moleculares foi desenvolvido no MIT na década de 60.
- Os autores descrevem o sistema como um software para “model building”.



Molecular Model-building by Computer

In which biochemists observe models of giant molecules as they are displayed on a screen by a computer and try to fold them into the shapes that they assume in nature

by Cyrus Levinthal

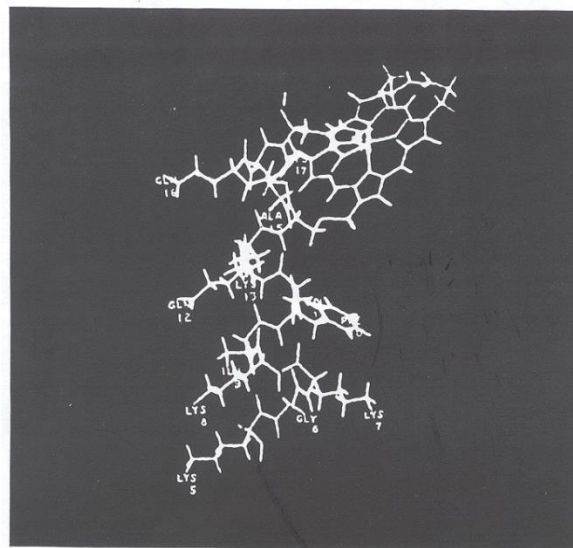
Many problems of modern biology are concerned with the detailed relation between biological function and molecular structure. Some of the questions currently being asked will be completely answered only when one has an understanding of the structure of all the molecular components of a biological system and a knowledge of how they interact. There are, of course,

a large number of problems in biology into which biologists have some insight but concerning which they cannot yet ask suitable questions in terms of molecular structure. As they see such problems more clearly, however, they invariably find an increasing need for structural information. In our laboratory at the Massachusetts Institute of Technology we have recently started using a

computer to help gain such information about the structure of large biological molecules.

For the first half of this century the metabolic and structural relations among the small molecules of the living cell were the principal concern of biochemists. The chemical reactions these molecules undergo have been studied intensively. Such reactions are specifically catalyzed by the large protein molecules called enzymes, many of which have now been purified and also studied. It is only within the past few years, however, that X-ray-diffraction techniques have made it possible to determine the molecular structure of such protein molecules. These giant molecules, which contain from a thousand to tens of thousands of atoms, constitute more than half of the dry weight of cells. Protein molecules not only act as enzymes but also provide many of the cell's structural components. Another class of giant molecules, the nucleic acids, determine what kind of protein the cell can produce, but most of the physiological behavior of a cell is determined by the properties of its proteins.

The X-ray-diffraction methods for investigating the three-dimensional structure of protein molecules are difficult and time-consuming. So far the structures of only three proteins have been worked out: myoglobin, hemoglobin and lysozyme [see "The Three-dimensional Structure of a Protein Molecule," by John C. Kendrew, *SCIENTIFIC AMERICAN*, December, 1961, and "The Hemoglobin Molecule," by M. F. Perutz, November, 1964]. In their studies of the hemoglobin molecule M. F. Perutz and his associates at the Laboratory of Molecular Biology in Cambridge, England, have observed that the structure of the molecule changes slightly when



MOLECULAR MODEL of a segment of cytochrome *c*, a protein that plays an important role in cell respiration, is shown as it is displayed on an oscilloscope screen. The protein has 104 amino acid subunits; this segment consists of units 5 through 18 (designated here by their abbreviated names). The heme group, which acts as a carrier of electrons, is known to be attached to amino acids 14 and 17. In the hypothetical structure shown here this stretch of the molecule is assumed to be in the characteristic "alpha helix" configuration.

Softwares para Modelação Molecular

- Actualmente muitos aspectos da modelação molecular podem ser realizados num computador pessoal, particularmente as tarefas de visualização
- Estão disponíveis muitas ferramentas “on-line” para modelação molecular
- Tarefas mais pesadas como o screening virtual, dinâmica molecular ou cálculos quânticos *ab initio* frequentemente necessitam de computadores mais potentes (servidores, clusters ou supercomputadores).

Softwares para Modelação Molecular

- **Visualização (instalável)**

- ✓ PyMOL
- ✓ VMD
- ✓ Chimera
- ✓ PMV
- ✓ SwissPDB Viewer

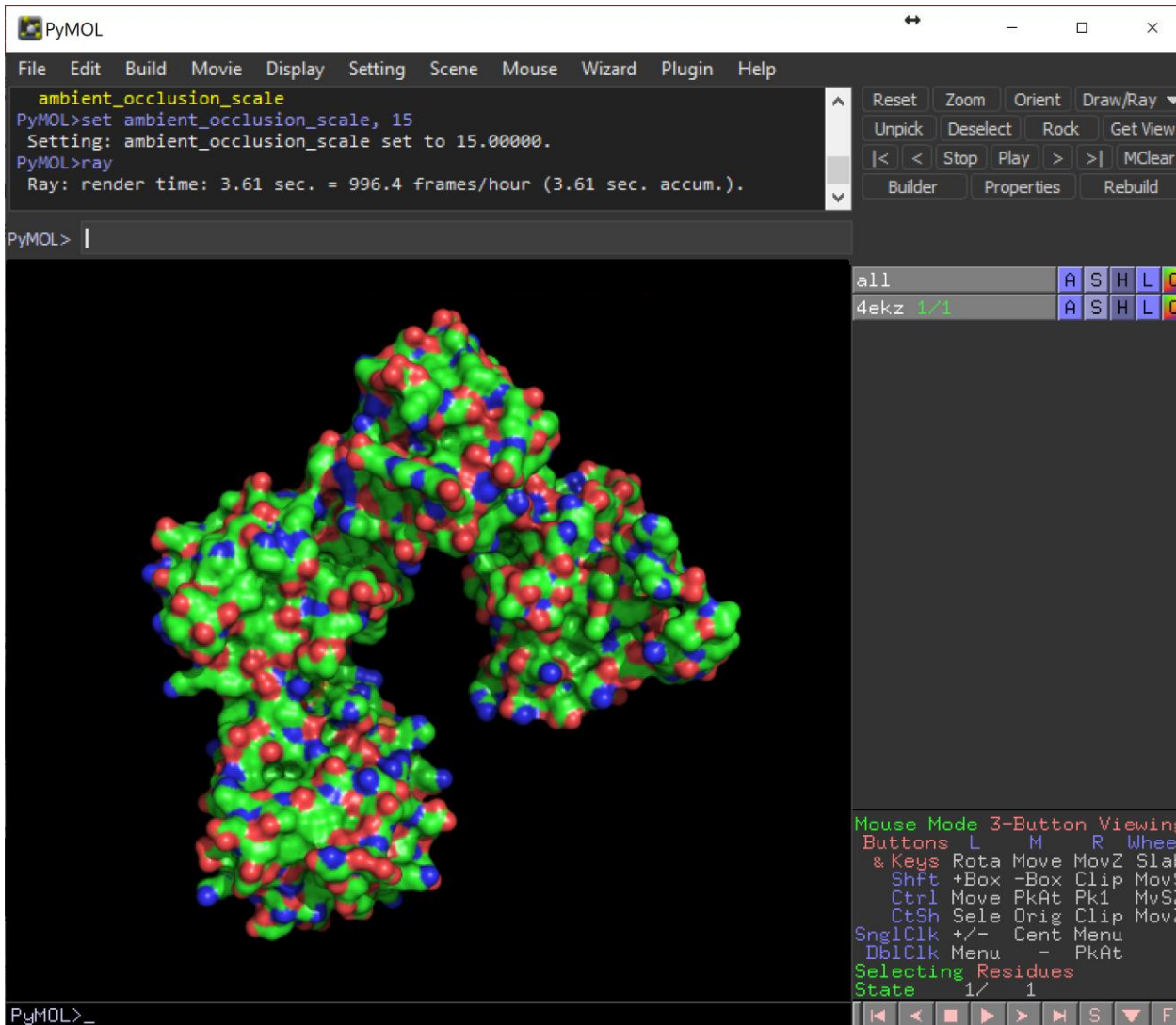
- **Visualização (on-line)**

- ✓ Jmol
- ✓ JSMOL
- ✓ NGL viewer
- ✓ Ghemical

- **Construção / Cálculos**

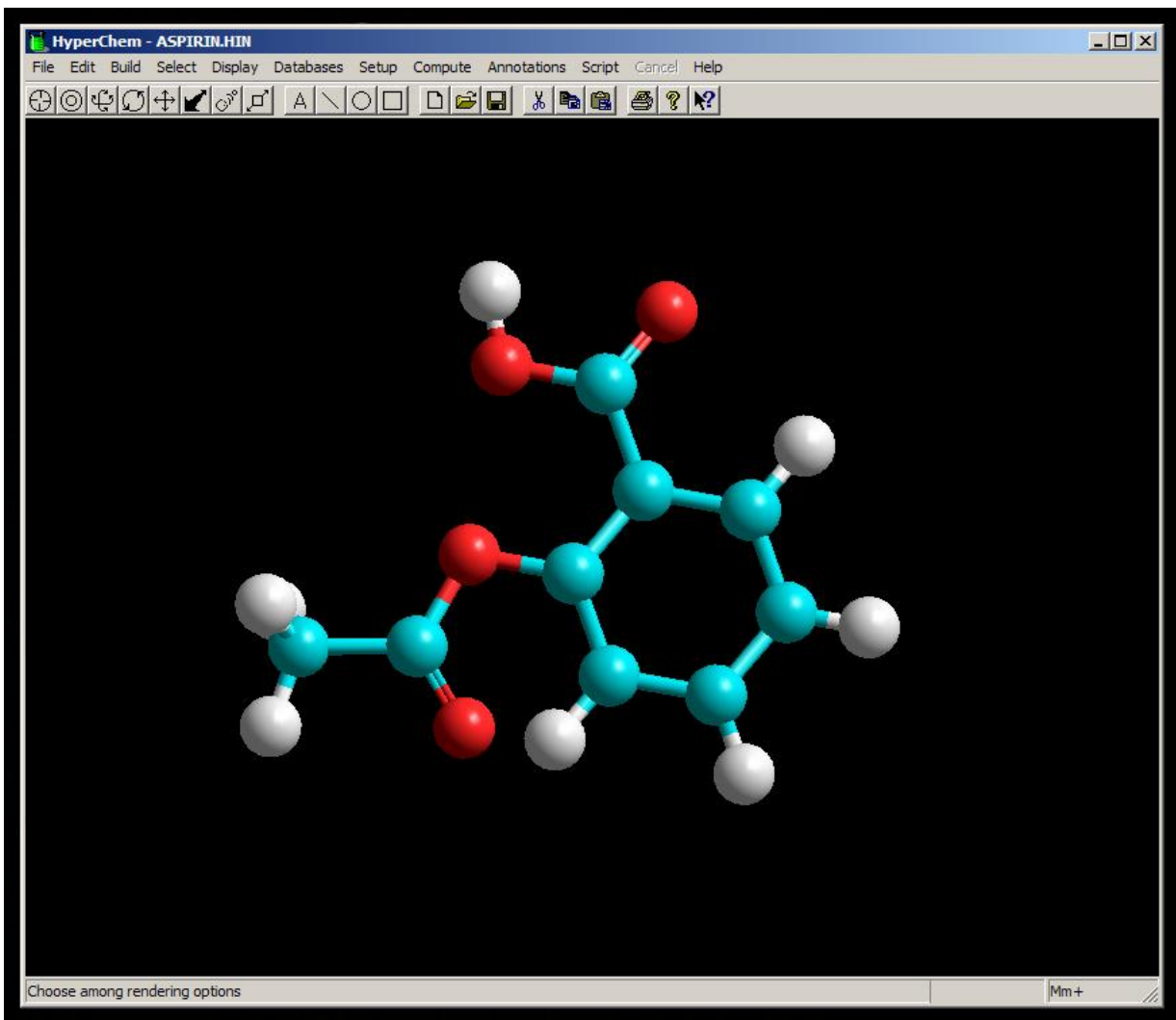
- ✓ Marvin
- ✓ Avogadro
- ✓ Hyperchem
- ✓ Schroedinger
- ✓ Insight
- ✓ Ghemical
- ✓ Gaussian/Gaussview

Visualização: PyMOL



- Visualização (macromoléculas)
- Comparações de moléculas
- Animações moleculares
- Ferramentas de análise
- Scripting
- Extensível (Plugins)
- Windows / Linux
- Open source

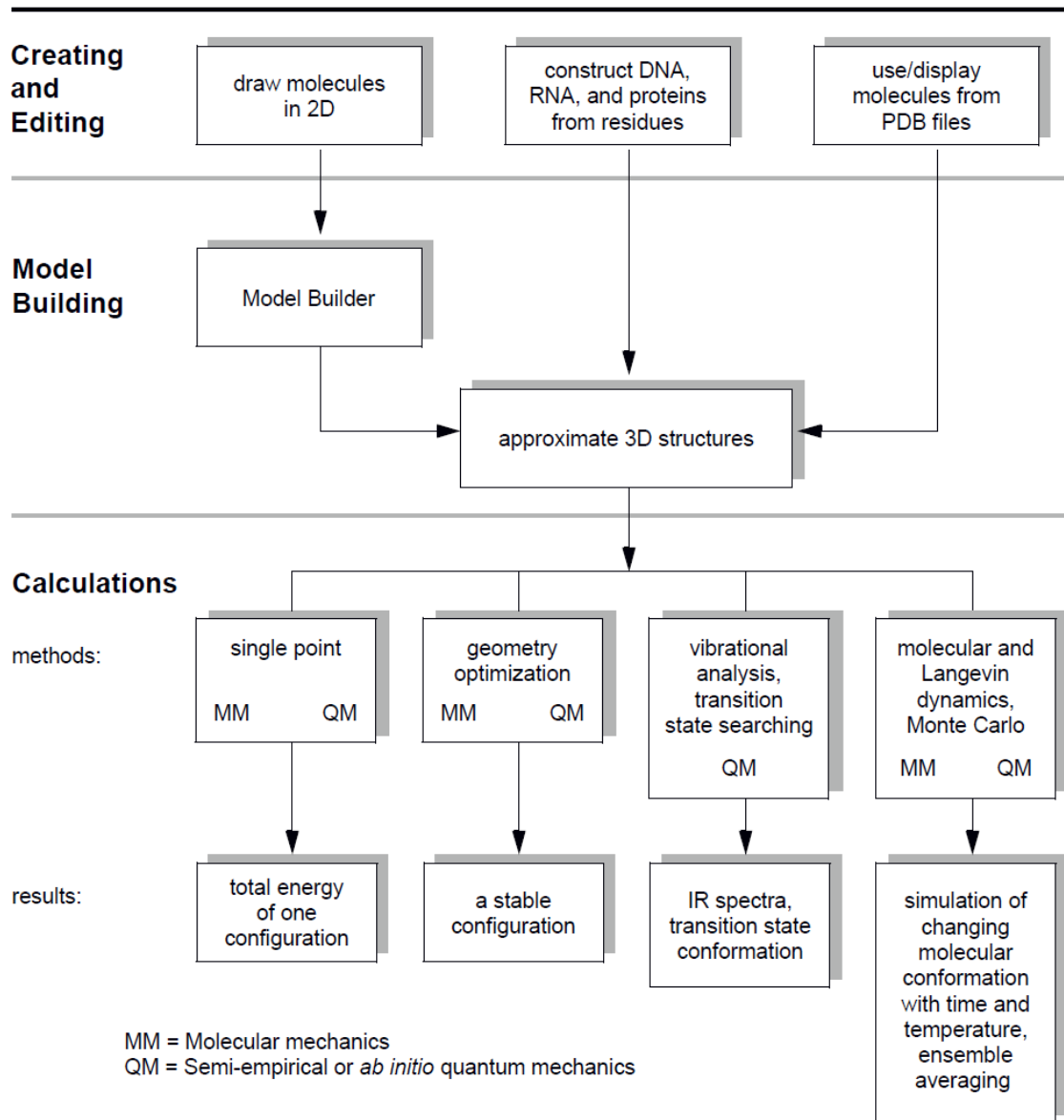
Construção: Hyperchem



- Construção de moléculas
- Optimização da geometria
- Cálculos quânticos
- Mecânica molecular
- Minimização de energia
- Pesquisa conformacional
- Dinâmica molecular
- Windows OS
- Comercial

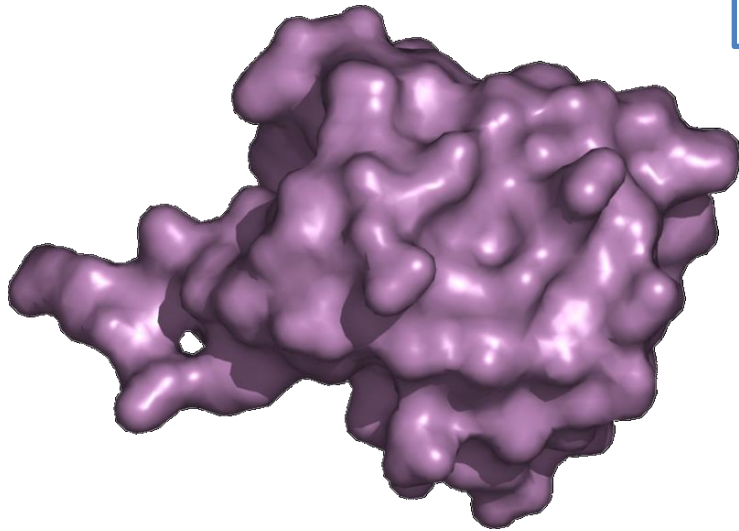
<http://www.hypercube.org>

HyperChem: Summary of Major Functions

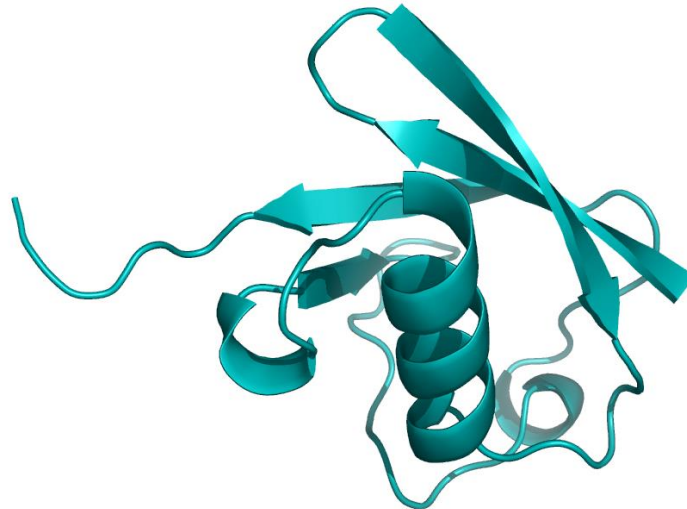


Modelos virtuais

Os modelos podem ser virtualizações que se desviam mais ou menos da “verdade” física para acentuar ou facilitar a representação de determinadas propriedades



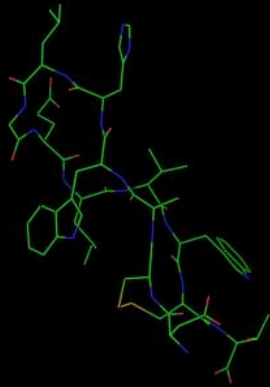
Superfície molecular de uma proteína



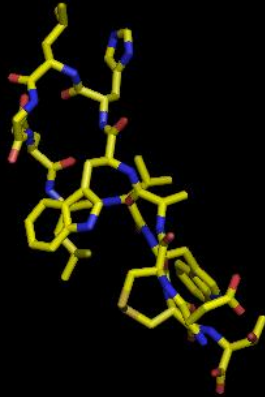
Representação esquemática da estrutura secundária

Modos de representação

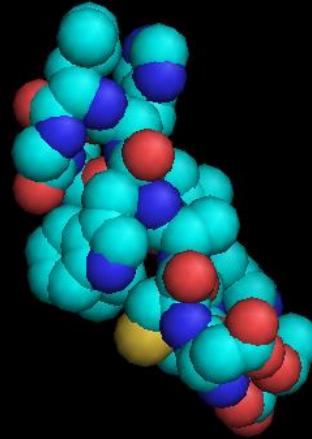
lines



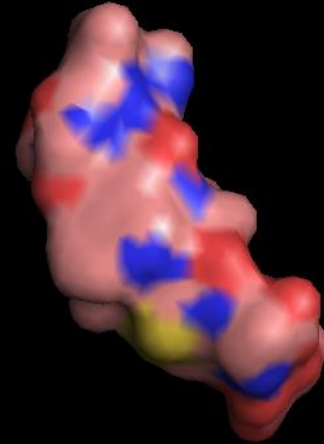
sticks



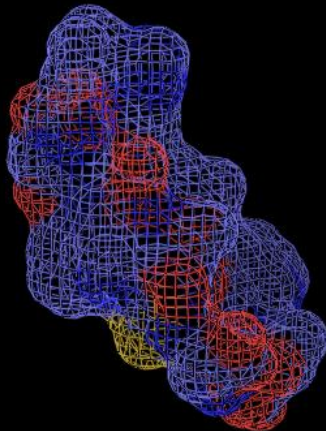
spheres



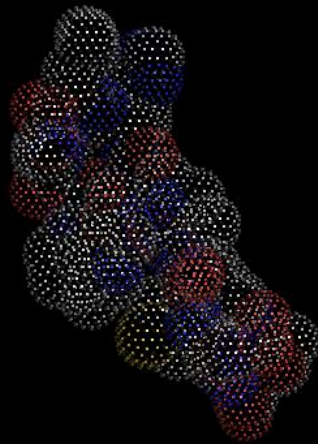
surface



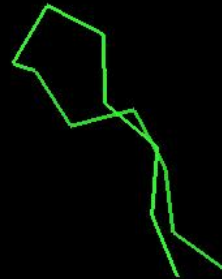
mesh



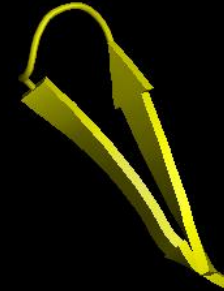
dots



ribbon



cartoon



Níveis de aproximação nos cálculos de modelação molecular

- **Ab initio:** solução da equação onda de Schrödinger para o sistema molecular, determinação da densidade electrónica e dos níveis de energia do sistema. Descrição completa e rigorosa do sistema. Permite a descrição da formação e quebra de ligações químicas.
- **Semi-empírico:** solução da equação de Schrödinger com recurso a aproximações que possibilitam o tratamento de sistemas com maior número de átomos. Descrição quase completa, mas com sacrifício do rigor em favor da simplicidade de cálculo. Permite a descrição da formação e quebra de ligações químicas.
- **Mecânica molecular:** o sistema molecular é descrito por uma *função de energia* com forma clássica, cujos parâmetros são obtidos através de uma mistura de cálculos quânticos e ajuste a resultados experimentais. Descrição incompleta e aproximada do sistemas, mas permite cálculos em sistemas com grande número de átomos tais como as *macromoléculas biológicas*.
- **QM/MM:** *combinação de métodos quânticos e mecânica molecular*

Aproximação de Born-Oppenheimer: a energia de um sistema molecular pode ser escrita como função exclusiva da posição dos centros atómicos. Por outras palavras, a energia só depende da *posição relativa e conformação* do sistema molecular.

Cálculos *ab initio*

$$\hat{H}\psi = E\psi$$

Hamiltoniano

Energia

Função de onda

Partindo de primeiros princípios

A equação de Schrodinger é uma equação **fundamental** da Física que descreve o comportamento de átomos e moléculas:

Níveis de energia - E_1, E_2, \dots, E_n

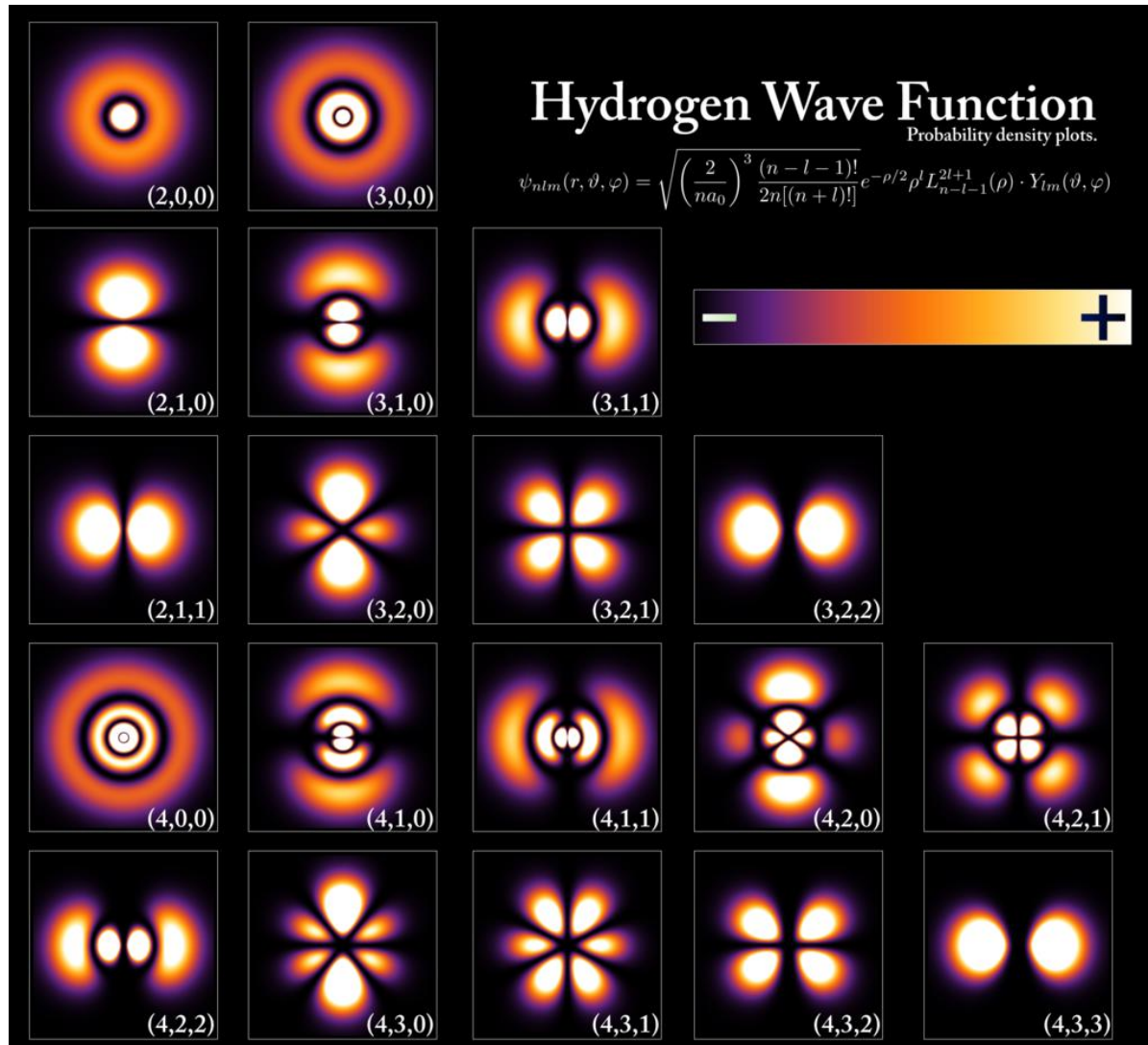
Funções de onda - $\Psi_1, \Psi_2, \dots, \Psi_n$

$|\Psi(x,y,z)|^2$ \longrightarrow Probabilidade de encontrar o electrão no ponto (x,y,z)

Equação de Schrodinger para o átomo hidrogénio

$$\frac{-\hbar^2}{8\pi^2m} \nabla^2 \Psi - \frac{e^2}{4\pi\epsilon_0 r} \Psi = E\Psi$$

Funções de onda do átomo de hidrogénio



Importância dos cálculos de química quântica no desenho de fármacos

- Permitem o cálculo preciso de propriedades moleculares tais como níveis de energia, densidades eletrônicas e comprimentos de ligação, os quais são cruciais para a compreensão do comportamento das moléculas em sistemas biológicos.
- Permitem reações químicas com quebra e formação de ligações, e fazer o cálculo da geometria e energética dos estados de transição
- Possibilitam o cálculo da afinidade de ligação de de um molécula de um fármaco ou lead ao seu alvo, essencial no desenho de fármacos
- Podem oferecer pistas para o mecanismo de ação de um fármaco, o que pode ajudar no desenvolvimento de novas moléculas com eficácia melhorada e efeitos colaterais mais reduzidos
- Podem ser usados para otimizar a estrutura de um composto lead, de forma a aumentar a sua potência e especificidade
- Podem ser usados para investigar os potenciais efeitos tóxicos de uma molécula em células humanas, o que pode ajudar a identificar potenciais riscos de segurança numa fase precoce do desenho de fármacos.

Complexidade dos cálculos *ab initio*

- A equação de Schrodinger é de difícil solução, não existindo soluções analíticas exatas para o seu caso mais geral (sistema multiatômico e polieletrónico).
- É possível resolver de forma exata a ES para o átomo de hidrogénio, e outros átomos mono-eletrónicos (He^+ , Li^{2+} , ...)
- A presença de múltiplos eletrões obriga ao uso de soluções *aproximadas*, resolvidas de forma *numérica* num computador digital.
- Num átomo ou molécula polieletrónica, os orbitais moleculares descrevendo o sistema são construído como uma *combinação linear* de orbitais atómicas mono-eletrónicas.
- A *correlação* entre os eletrões, efeitos de polarização, etc... podem ser tratadas de forma mais ou menos aproximada.
- A dificuldade do cálculo e tempo de computação envolvidos depende do nível de aproximação desejado e do tamanho do sistema molecular.

Tipos de cálculos *ab initio*

- **Hartree –Fock (HF)** método mais básico em que a correlação eletrónica é ignorada e cada eletrão sente o efeito “médio” de todos os outros
- **Moller-Plesset (MP, MP3, MP4,...)** a interação entre eletrões é aproximada por uma *perturbação* de ordem n à função de onda.
- **Coupled-Cluster (CCS, CCSD, CSDT, ...)** descreve interações (excitações) entre grupos de eletrões. Método mais rigoroso de cálculo *ab initio* mas pode ser extraordinariamente pesado.
- **Density functional theory (DFT)** Cálculo direto da densidade multi-eletrónica através da construção de um funcional de densidade, requer parametrização “empírica” mas é muito mais rápido para sistema com elevado número de eletrões. .

Funções de base

- Os orbitais atómicos de cada átomo no sistema são calculados como uma *combinação linear* de funções de base que contribuem para a descrições mais ou menos *aproximada* do comportamento de cada electrão.
- Maior complexidade das funções de base permite uma descrição mais exacta da estrutura electrónica do sistema, mas resulta em cálculos mais complexos e mais demorados.
- O rigor de um cálculo *ab initio* depende assim não apenas dos métodos anteriormente descritos, mas também da escolha do conjunto de funções de base
- A escolha do conjunto de funções de base mais adequado depende das características do sistema molecular que se pretende estudar e do tipo de propriedades que se pretende calcular

Funções de base

- “Minimal” (STO-3G) : o conjunto de funções de base mais simples, geralmente usado para cálculos preliminares. Cada orbital atômica é descrita por uma única função de base, que é uma combinação linear de 3 funções gaussianas (“3G”). “STO” significa “Slater-Type Orbital”
- Maior complexidade das funções de base permite uma descrição mais exata da estrutura eletrônica do sistema, mas resulta em cálculos mais complexos
- Tipos funções de base (basis sets):
 - Minimal (exemplo: STO-3G)
 - Polarized (exemplo: 6-31G*)
 - Diffuse (exemplo: 6-31G⁺⁺)
 - Valência dividida: double-, triple-, quadruple-zeta (exemplo: 6-31G, 6-311G)

6-31G – uma função de base de 6 gaussianas para cada orbital interna, 2 funções de base para as orbitais de valência, uma contendo 3 gaussianas, outra 1 gaussianas

Exemplos de funções de base

- **STO-3G:** é um conjunto de funções de base pequeno e minimalista que é comumente usado como ponto de partida para cálculos. Relativamente rápido de calcular, mas não muito preciso.
- **6-31G:** conjunto de funções de base de tamanho médio, o qual fornece um bom equilíbrio entre precisão e custo computacional. É comumente usado para moléculas de pequeno a médio tamanho.
- **6-311G:** conjunto de funções de base maior, o qual fornece precisão superior ao 6-31G, mas a custo de tempo computacional aumentado. É comumente usado para moléculas de médio a grande tamanho.
- **cc-pVTZ:** conjunto de funções de base muito grande que fornece alta precisão, mas é computacionalmente exigente. É comumente usado para moléculas muito grandes ou para cálculos de alta precisão.
- **LanL2DZ:** conjunto de funções de base projetado para moléculas contendo elementos Lantanídeos. Fornece precisão superior ao STO-3G para esses elementos.

Identificação de cálculo *ab initio* (nível de teoria):

HF/6-31G



Método



Conjunto de funções de base
(basis set)

Effect of Basis Set Choice on Computation Cost (cpu time)

axial-methylcyclohexane on SGI Indigo2
(*Spartan* cpu time in sec.)

<u>Method/Basis Set</u>	<u>s.p.</u>	<u>opt.</u>
AM1/STO-3G	~1	10
HF/STO-3G	72	983
HF/ 3-21G(d)	193	2214
HF/ 6-31G(d,p)	2632	34655 (9.6 h)

(approaching “HF limit”; energy [not shown] decreases w/ larger basis set)

Effect of Basis Set Choice on Computation Cost (cpu time)

Basis Set	# basis functions	Energy (au)	SCF cycles	Relative time
STO-3G	26	-189.53468869	14	0.05
3-21G	48	-190.88640754	14	0.2
6-31G	48	-191.87418982	14	0.3
6-31G*	72	-191.96061331	15	1
6-311G*	90	-192.00188312	15	3
6-311+G*	106	-192.00599408	15	6
6-311++G**	130	-192.01529556	15	25
6-311++G(2df,2pd)	226	-192.02957861	15	88
6-311++G(3df,3pd)	264	-192.03162788	31	235
cc-pCVTZ	204	-192.03289846	15	82
cc-pCVQZ	400	-192.04664288	30	3400
aug-cc-pCVQZ	712	-192.04773533	19	41000

- A precisão dos resultados depende do grau de **correlação electrónica** e do tamanho do conjunto de **funções de base**
- O peso dos cálculos aumenta rapidamente com o nível de correlação e com o tamanho do conjunto de funções de base
- Os cálculos efectuados representam sempre um compromisso entre precisão e tempo de cálculo
- O tempo de cálculo varia quadraticamente com o número de electrões do átomo ou molécula. Cálculos rigorosos para macromoléculas com milhares de átomos são impraticáveis com os computadores actuais.
- Cálculos *ab initio* puros em macromoléculas biológicas são geralmente impraticáveis, devido ao elevado número de átomos e ligações químicas

Família de métodos *ab initio*

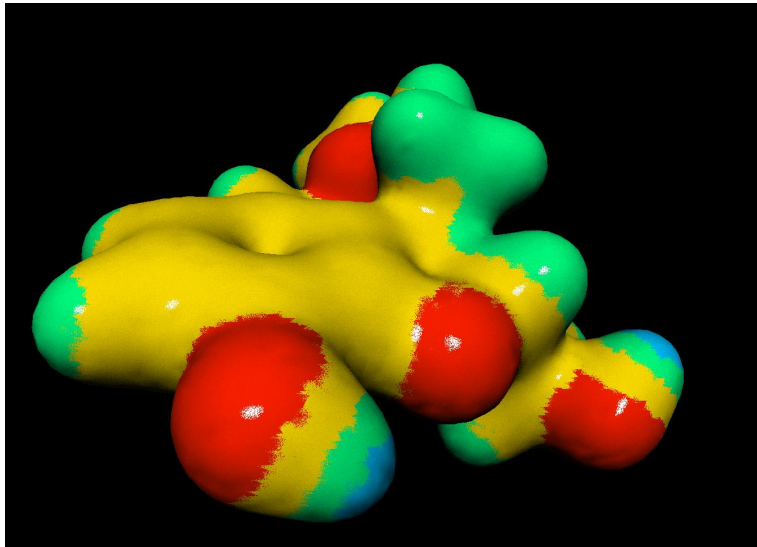
Método de cálculo

Funções de base

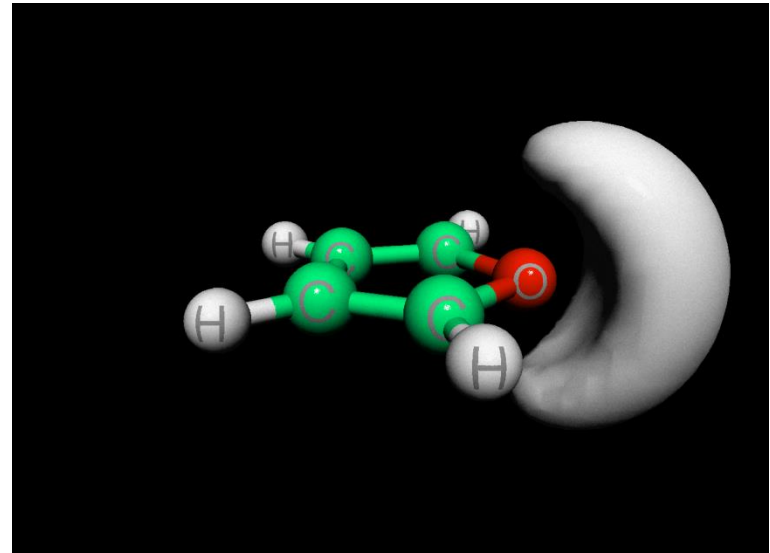
	HF	MP2	CCSD	CCSD(T)	CCSDT	...	Full CI
Minimal						...	
Split-valence						...	
Polarized						...	
Diffuse						...	
High angular momentum						...	
...
∞						...	Exact solution

Cálculos *ab initio*

Permitem o cálculo de propriedades como a *densidade electrónica total* ou o *campo electrostático molecular*, valiosas na previsão da reactividade das moléculas.

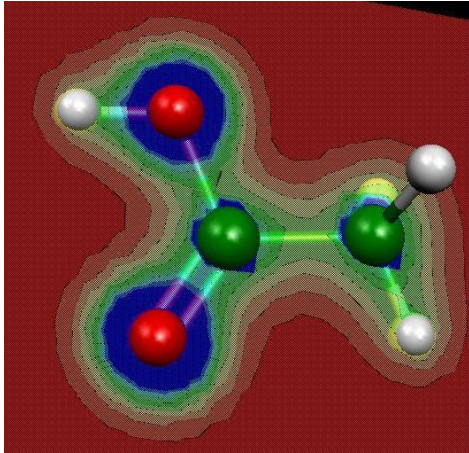


Densidade electrónica total da morfina

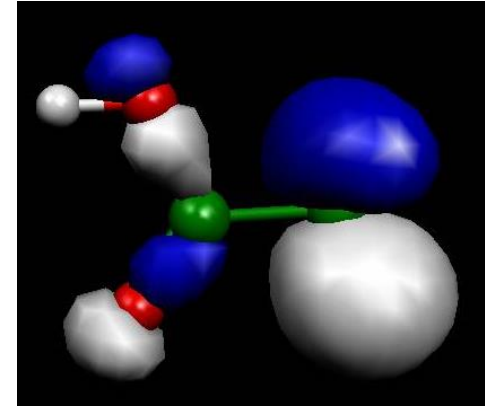


Campo electrostático do furano

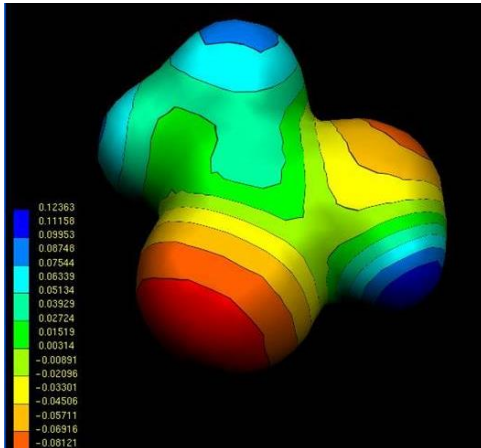
Exemplo: ácido acético



Potencial electrostático do ácido acético



Orbitais moleculares

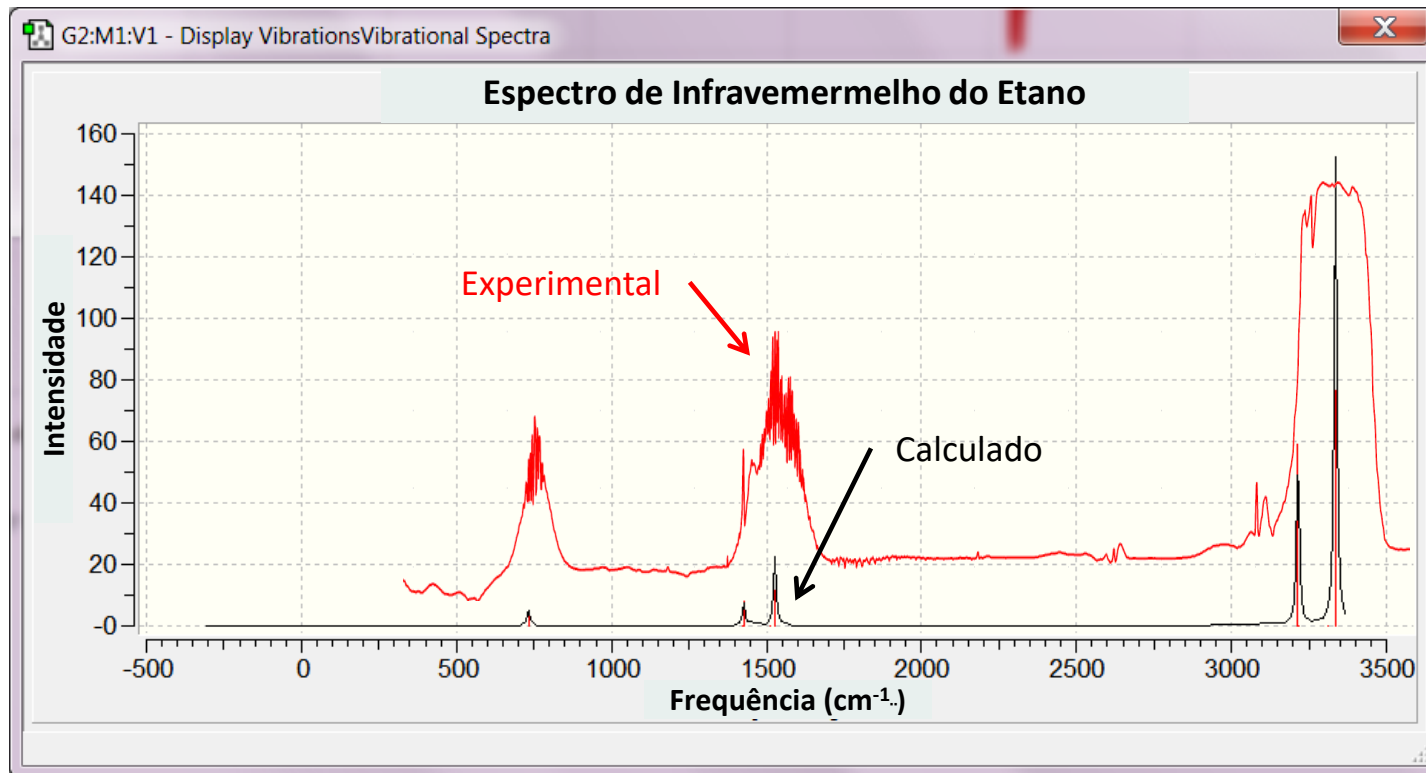
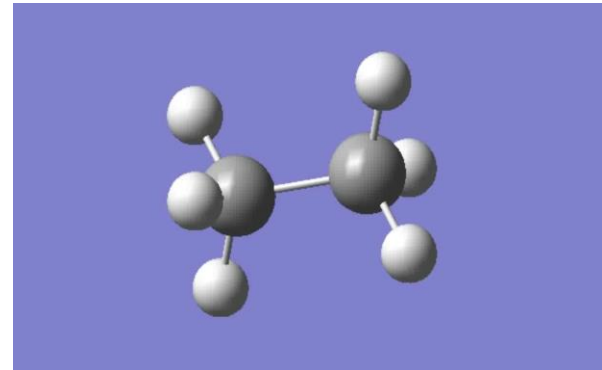


Potencial electrostático na superfície molecular

Os cálculos *ab initio* permitem facilmente o cálculo rigoroso da estrutura eletrônica de moléculas pequenas compostas de átomos leves.

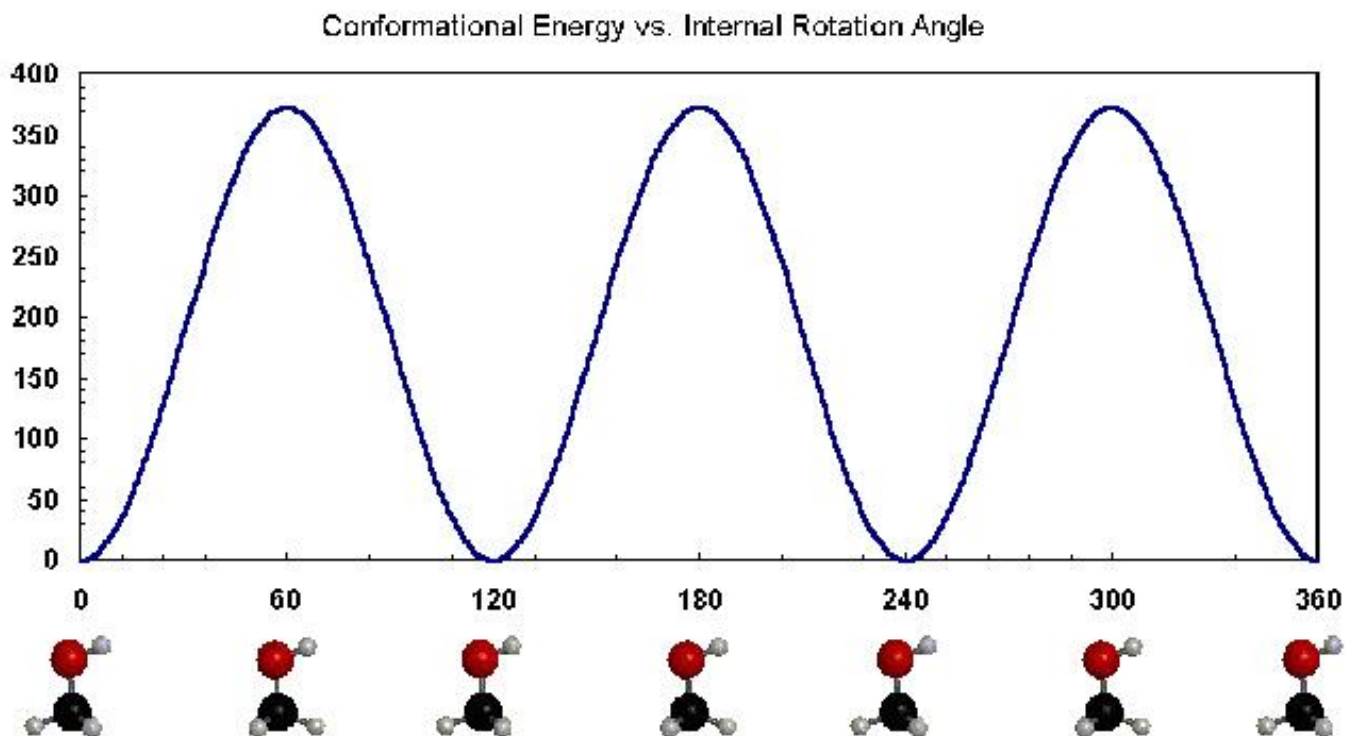
Exemplo: modos de vibração

Os cálculos quânticos podem ser usados para calcular energias vibracionais das moléculas e prever a forma dos seus espectros de IR.



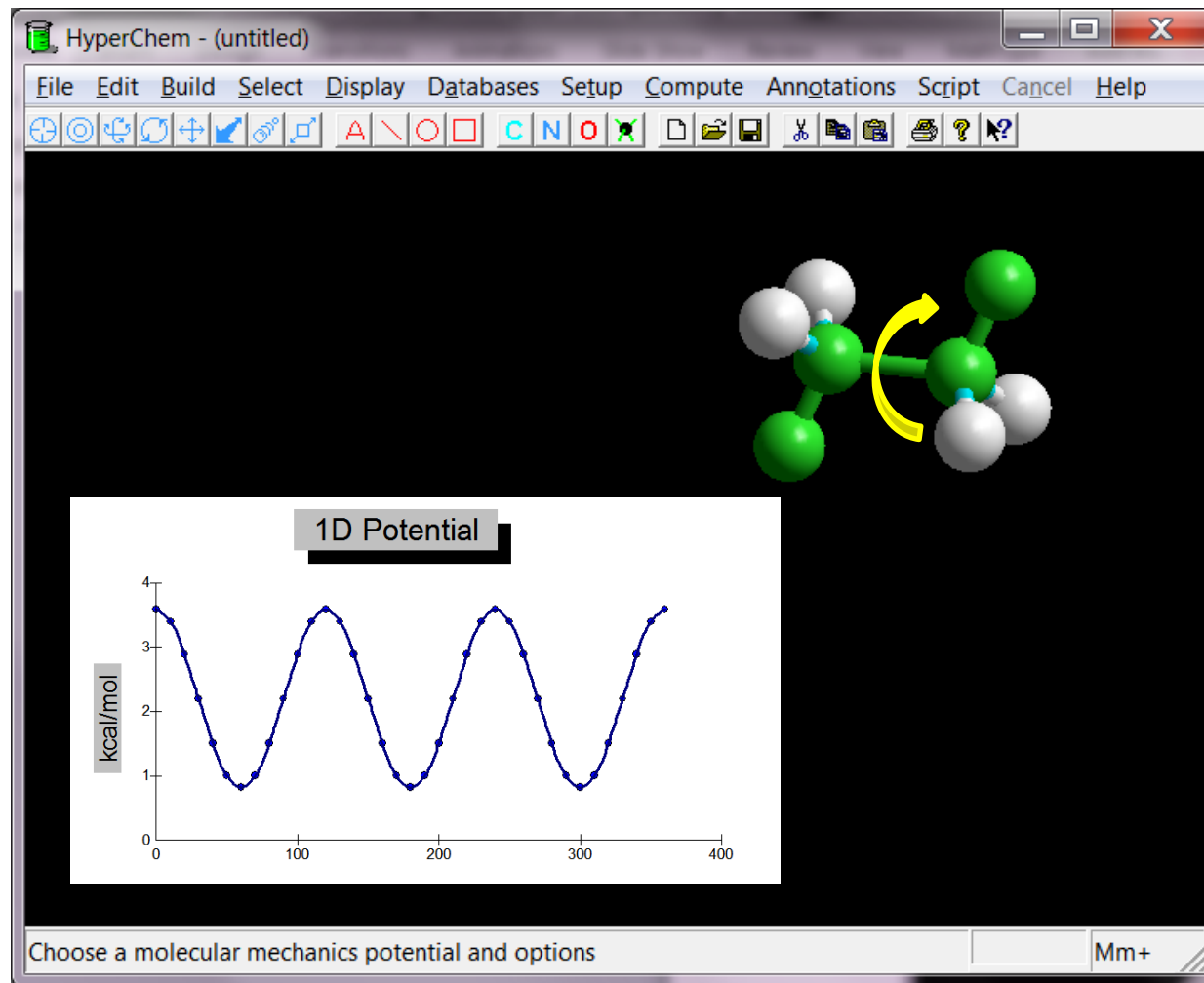
Cálculos *ab initio*

Análise conformacional: cálculo da energia de uma molécula em função da sua conformação.



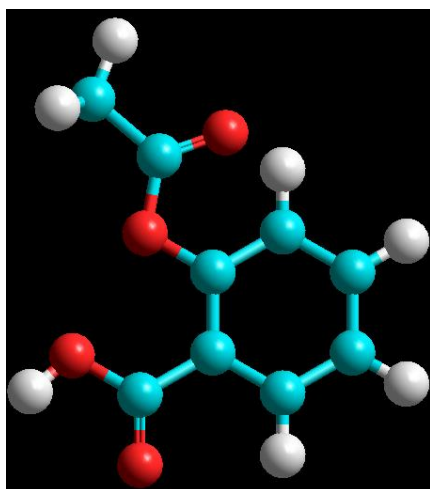
Energia da molécula de metanol em função do seu ângulo interno de rotação

Exemplo: análise conformacional

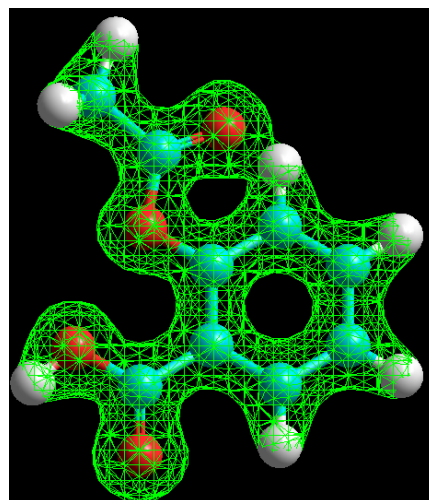
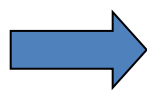


Energia da molécula de etano em função do ângulo de torção da ligação C-C

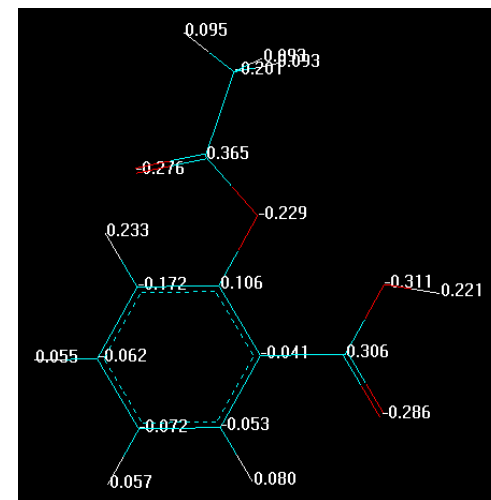
Cálculo de cargas parciais por métodos *ab initio*



Coordenadas
atómicas



Densidade
electrónica



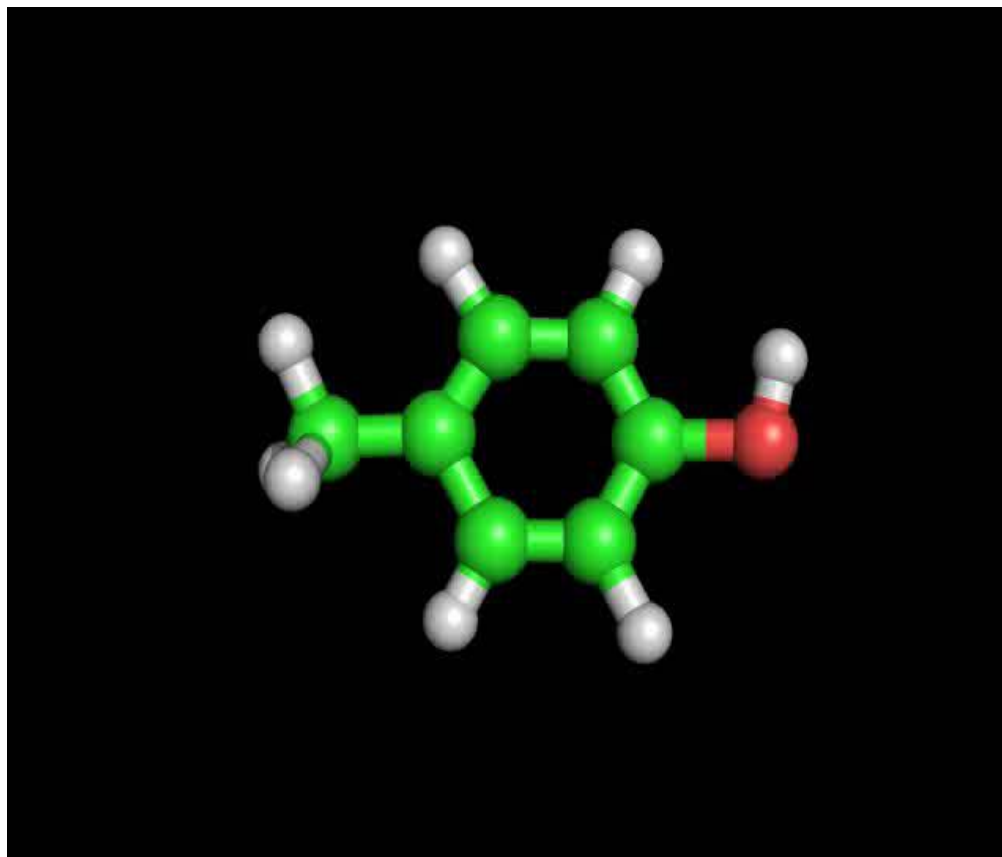
Cargas parciais
atómicas

As cargas parciais são uma aproximação que permite descrever as propriedades electrostáticas das moléculas a partir de cargas pontuais, centradas nos átomos, que reflectem as variações locais da densidade de carga electrónica.

Dinâmica molecular de Born-Oppenheimer

$$\mathbf{H}\Psi = E\Psi$$

$$\mathbf{H} = -\frac{\hbar^2}{2m}\nabla^2 + V$$



HF/3-21G

1 ps (10^{-12} s)

Nesta simulação a energia potencial do sistema é calculada usando um método quântico (Equação de Schrödinger). A partir deste potencial, são calculadas as forças actuando sobre os núcleos dos átomos, os quais são movidos de acordo com as leis da mecânica clássica ($\mathbf{F}=\mathbf{ma}$). Tratamento impraticável para sistema moleculares com um número elevado de átomos (p.ex. proteínas)

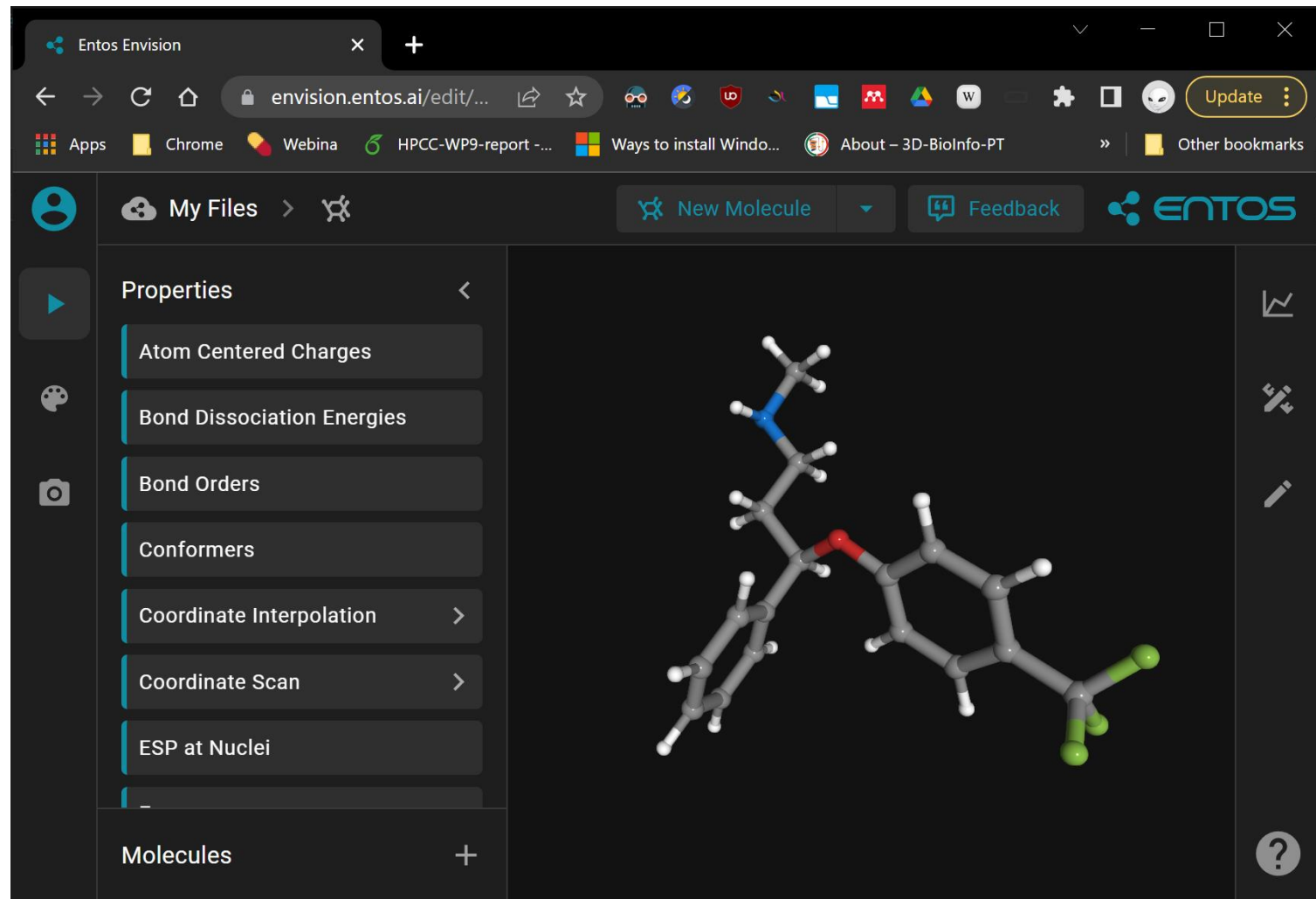
Métodos semi-empíricos

- Os métodos quânticos semi-empíricos são métodos computacionais usados em química quântica para aproximar a estrutura das electrónica das moléculas. Estes método baseiam-se numa combinação de dados empíricos e cálculos teóricos, e são tipicamente mais rápidos e menos pesados computacionalmente. Alguns pontos fundamentais:
- Baseiam-se num modelo simplificado da estrutura eletrónica das moléculas, permitindo cálculos mais eficientes
- Usam dados empíricos para ajustar o modelo, aumentando a precisão dos cálculos.
- Geralmente tratam a correlação eletrónica de forma aproximada, uma das maiores fontes de erro em cálculos quânticos
- São geralmente usados para estudar moléculas de tamanho médio, caso em que oferecem um equilíbrio entre precisão e custo computacional
- Alguns métodos semi-empíricos mais comuns: MNDO, AM1 e PM3

Métodos semi-empíricos

- **AM1:** Este método é frequentemente usado para pequenas moléculas, tais como compostos lead, sendo de computação bastante rápida. Não é muito preciso para moléculas com muitas ligações ou para estudo de alterações de conformação.
- **PM3:** Modificação de AM1 com inclusão de parâmetros adicionais que melhoram a performance para moléculas com múltiplas ligações. Geralmente usado para moléculas de tamanho médio a grande.
- **PM6:** PM3 melhorado com maior no cálculo de alterações conformacionais e propriedades relacionadas com ligações múltiplas. Usado frequentemente para moléculas de tamanho médio.
- **MNDO:** métodos comumente usado para prever as estruturas e energias de compostos organometálicos. Relativamente rápido, mas pouco preciso para moléculas com muitas ligações.
- **OM1:** MNDO melhorado, com maior precisão de cálculo em compostos organometálicos. Geralmente usado para moléculas organometálicas de tamanho médio a grande

Envision: cálculos semi-empíricos on-line



The screenshot displays the Envision web application interface. At the top, the browser address bar shows the URL envision.entos.ai/edit/.... The interface includes a navigation bar with "My Files", "New Molecule", and "Feedback" options. A sidebar on the left lists various calculation and visualization tools under the "Properties" section:

- Atom Centered Charges
- Bond Dissociation Energies
- Bond Orders
- Conformers
- Coordinate Interpolation
- Coordinate Scan
- ESP at Nuclei

The main workspace features a 3D ball-and-stick model of a complex organic molecule, rendered in a dark theme. The molecule consists of a central carbon chain with various substituents, including a blue nitrogen atom, a red oxygen atom, and a green sulfur atom. A toolbar on the right side of the workspace contains icons for zooming, rotating, and editing the model.

Método semi-empírico : **GFN1-xTB**



<https://envision.entos.ai/>

Mecânica molecular

- O sistema molecular é descrito por uma função designada *campo de forças*, que representa a energia do sistema como função da posição dos centros atômicos (aproximação de Born-Oppenheimer).
- Os átomos são descritos como pontos dotados de carga (**cargas pontuais**) e de massa (**massas pontuais**).
- As ligações covalentes são descritas por um modelo harmónico (comparável a uma mola), com uma constante de força que depende da natureza química dos átomos ligados e da ordem da ligação.
- As ligações não-covalente podem ser electrostáticas (interacções entre cargas pontuais), ou forças de van der Waals (termos de Lennard-Jones, com componentes atrativa e repulsiva).
- São usados termos adicionais para fixar os valores dos ângulos de ligação, torções (ângulos dihedro) e outros termos de energia.

Os parâmetros para os vários termos do campo de forças são obtidos através de uma mistura de cálculos quânticos e dados experimentais (principalmente medidas espectroscópicas)

Mecânica molecular

- O sistema molecular é descrito por uma função designada *campo de forças*, que representa a energia do sistema como função da posição dos centros atômicos (aproximação de Born-Oppenheimer).
- Os átomos são descritos como pontos dotados de carga (**cargas pontuais**) e de massa (**massas pontuais**).
- As ligações covalentes são descritas por um modelo harmónico (comparável a uma mola), com uma constante de força que depende da natureza química dos átomos ligados e da ordem da ligação.
- As ligações não-covalente podem ser electrostáticas (interacções entre cargas pontuais), ou forças de van der Waals (termos de Lennard-Jones, com componentes atrativa e repulsiva).
- São usados termos adicionais para fixar os valores dos ângulos de ligação, torções (ângulos dihedro) e outros termos de energia.

Os parâmetros para os vários termos do campo de forças são obtidos através de uma mistura de cálculos quânticos e dados experimentais (principalmente medidas espectroscópicas)

Mecânica molecular

- Na mecânica molecular a descrição e evolução do sistema molecular é obtida através da resolução das **equações de movimento clássicas**, equações estas muito mais fáceis de resolver que a equação de onda da mecânica quântica
- Como solução das equações de movimento da mecânica molecular, obtemos a posição dos centros atômicos do sistema em cada instante de tempo, e não uma função de onda probabilística quântica.
- O sistema é caracterizado pelo *Hamiltoniano* (H), igual à soma da energia cinética e potencial. Na mecânica molecular o Hamiltoniano de um sistema é denominado **campo de forças**.

Equações do movimento
(eqs. de Hamilton)

$$\begin{cases} \dot{p}_i = -\frac{\partial H}{\partial q_i} \\ \dot{q}_i = \frac{\partial H}{\partial p_i} \end{cases}$$

Hamiltoniano

$$H = T + V$$

Energia cinética Energia potencial

q_i - posições
 p_i - momentos

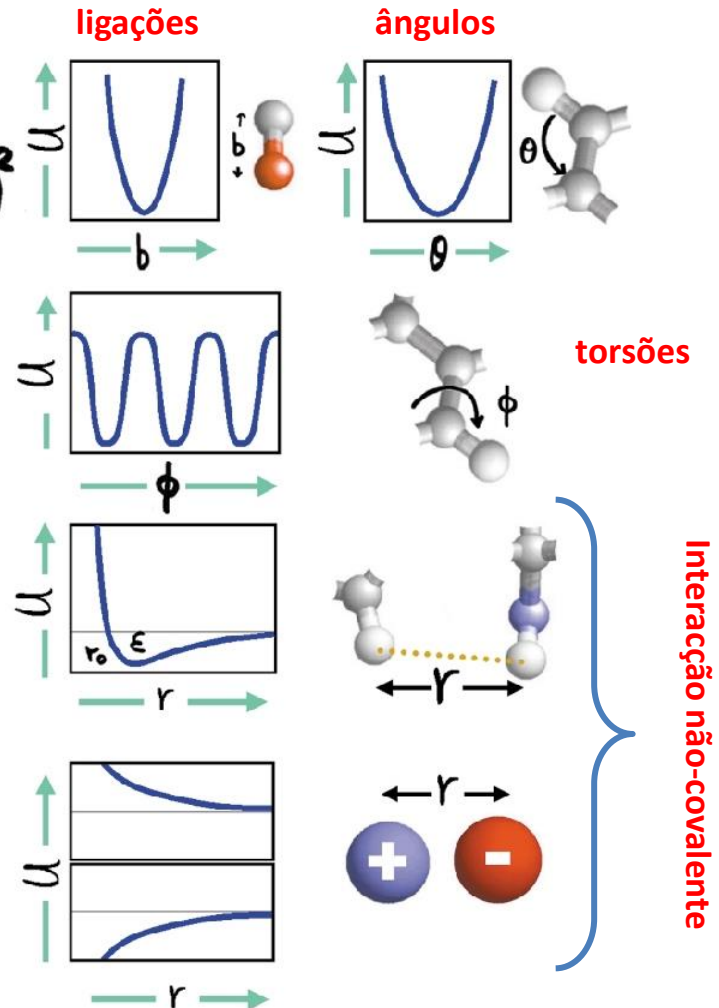
Campo de forças

$$U = \sum_{\text{All Bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{All Angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2$$

$$+ \sum_{\text{All Torsion Angles}} K_\phi [1 - \cos(n\phi + \delta)]$$

$$+ \sum_{\text{All nonbonded pairs}} \epsilon \left[\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^6 \right]$$

$$+ \sum_{\text{All partial charges}} \frac{332 q_i q_j}{r}$$



Softwares e forcefields

Softwares para mecânica molecular:

- Gromos
- Gromacs
- Amber
- Charmm
- Tinker
- Discover
- Hyperchem

Campos de forças usados em simulação biomolecular:

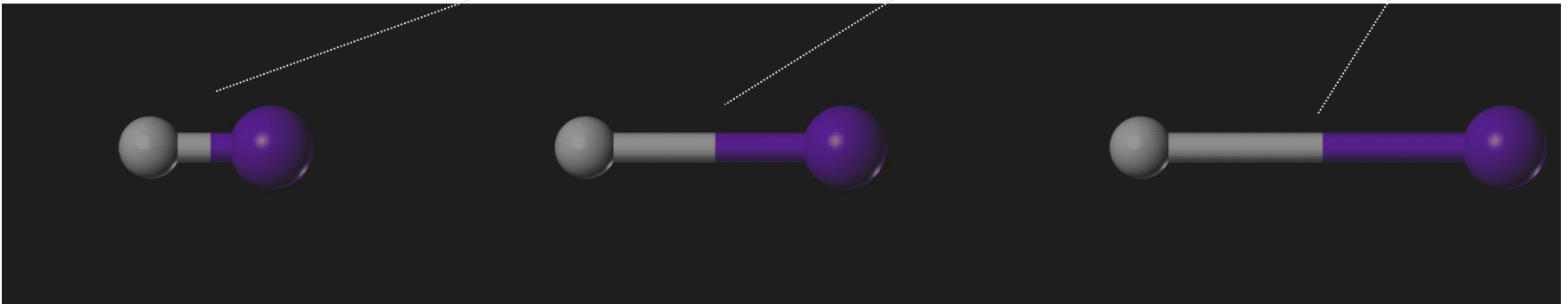
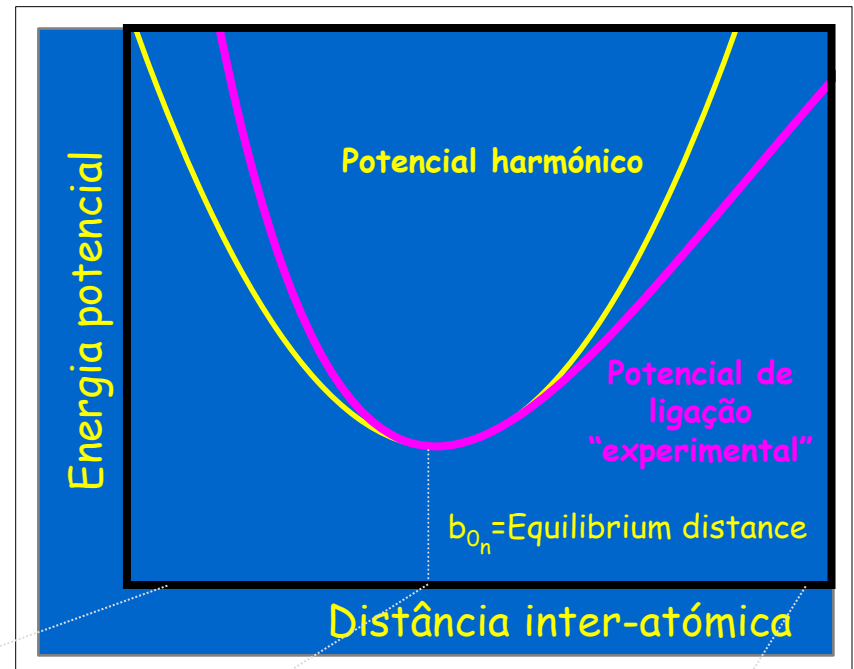
- Gromos
- CHARMM
- AMBER
- OPLS-1
- MM+
- UFS

Exemplo: campo de forças GROMOS

$$\begin{aligned} V(\mathbf{r}_1, \dots, \mathbf{r}_{N_{\text{at}}}) = & \sum_{n=1}^{N_b} \frac{1}{2} K_{b_n} (b_n - b_{0n})^2 + \\ & \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} (\theta_n - \theta_{0n})^2 + \\ & \sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} (\xi_n - \xi_{0n})^2 + \\ & \sum_{n=1}^{N_\phi} \frac{1}{2} K_{\phi_n} [1 + \cos(m_n \phi_n - \delta_n)] + \\ & \sum_{i < j}^{N_{\text{at}}} \left[\frac{C_{12}(i, j)}{r_{ij}^{12}} - \frac{C_6(i, j)}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \right] \cdot S(r_{ij}) + \\ & \text{Special terms} \end{aligned}$$

Potencial de ligação

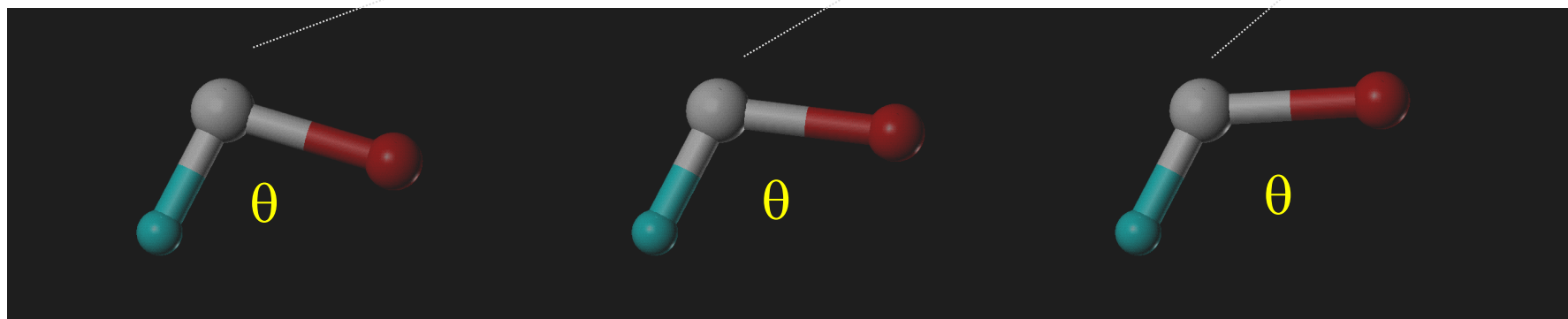
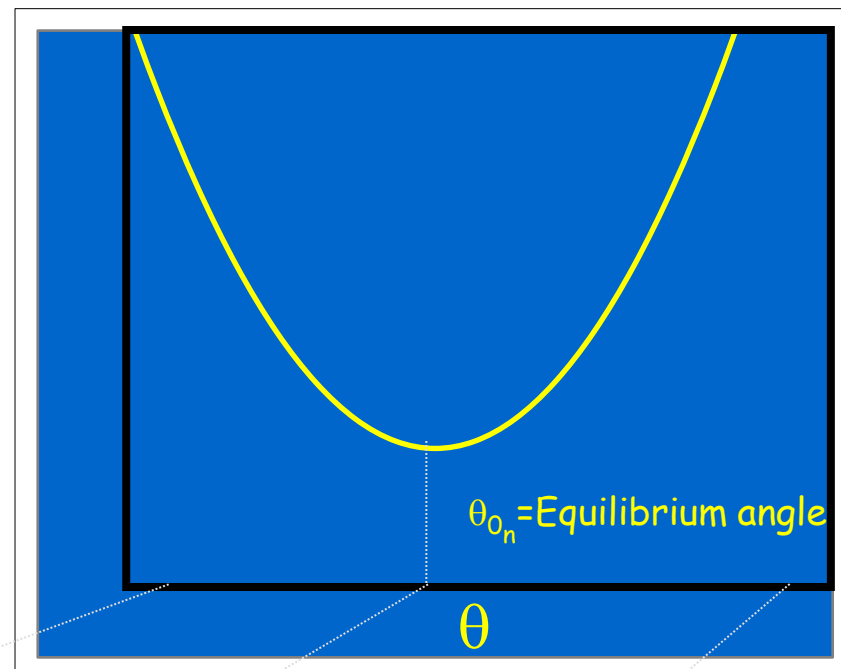
$$\sum_{n=1}^{N_b} \frac{1}{2} K_{b_n} (b_n - b_{0_n})^2$$



Não é permitida dissociação das ligações!

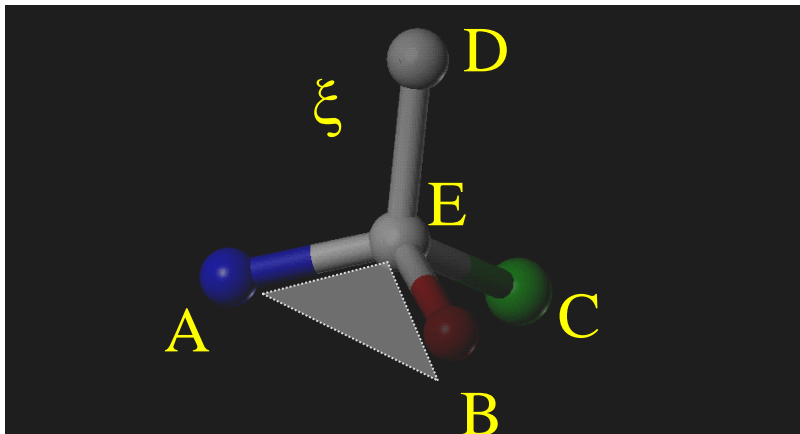
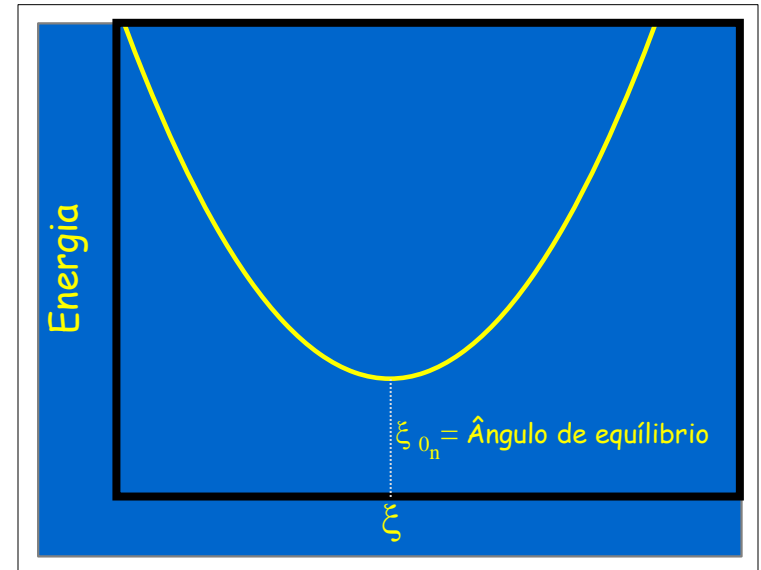
Potencial de ângulos de ligação

$$\sum_{n=1}^{N_{\theta}} \frac{1}{2} K_{\theta_n} (\theta_n - \theta_{0_n})^2$$



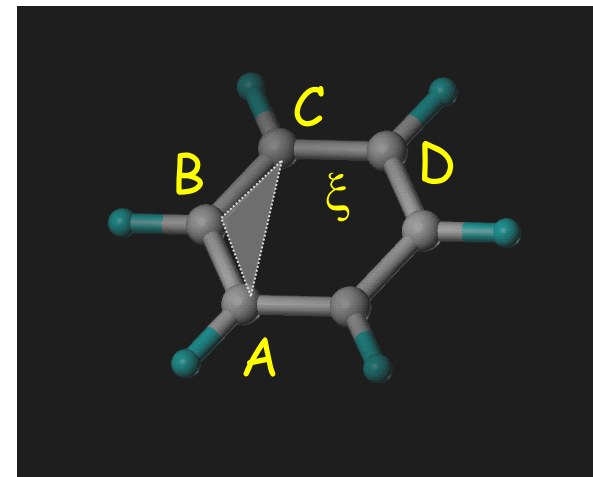
Diedros impróprios

$$\sum_{n=1}^{N_{\xi}} \frac{1}{2} K_{\xi_n} (\xi_n - \xi_{0n})^2$$



Centros quirais

Manutenção do ângulo entre E-D e o plano A-B-E



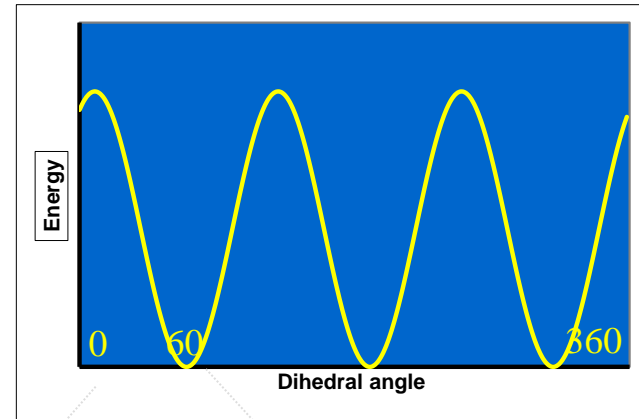
Centros planares

Manter os 4 átomos A,B,C,D no mesmo plano

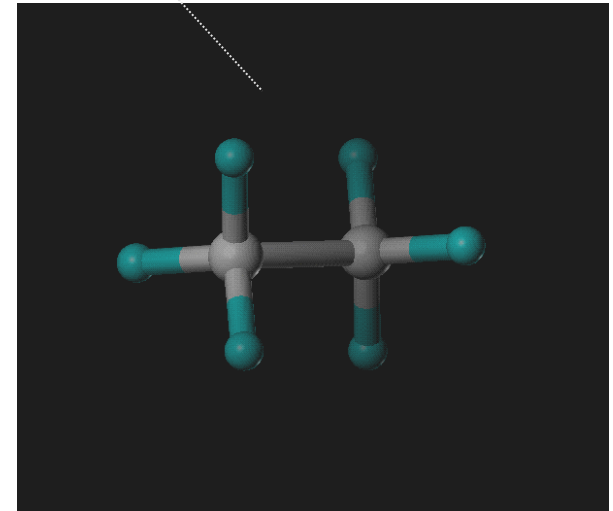
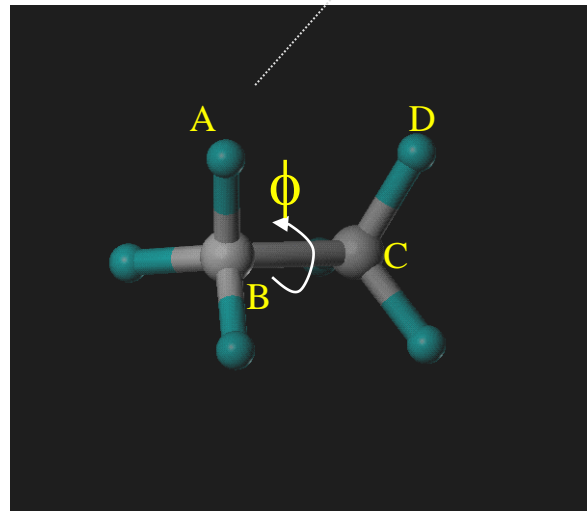
Potenciais de ângulos diedros

$$\sum_{n=1}^{N_{\phi}} \frac{1}{2} K_{\phi_n} \left[1 + \cos(m_n \phi_n - \delta_n) \right]$$

m_n - multiplicidade ; δ_n - fase



Os parâmetros destes potenciais são normalmente obtidos a partir de cálculos quânticos *ab initio*



Etano

Interacções “non-bonded”: potencial de van der Waals

Equação de Lennard-Jones:

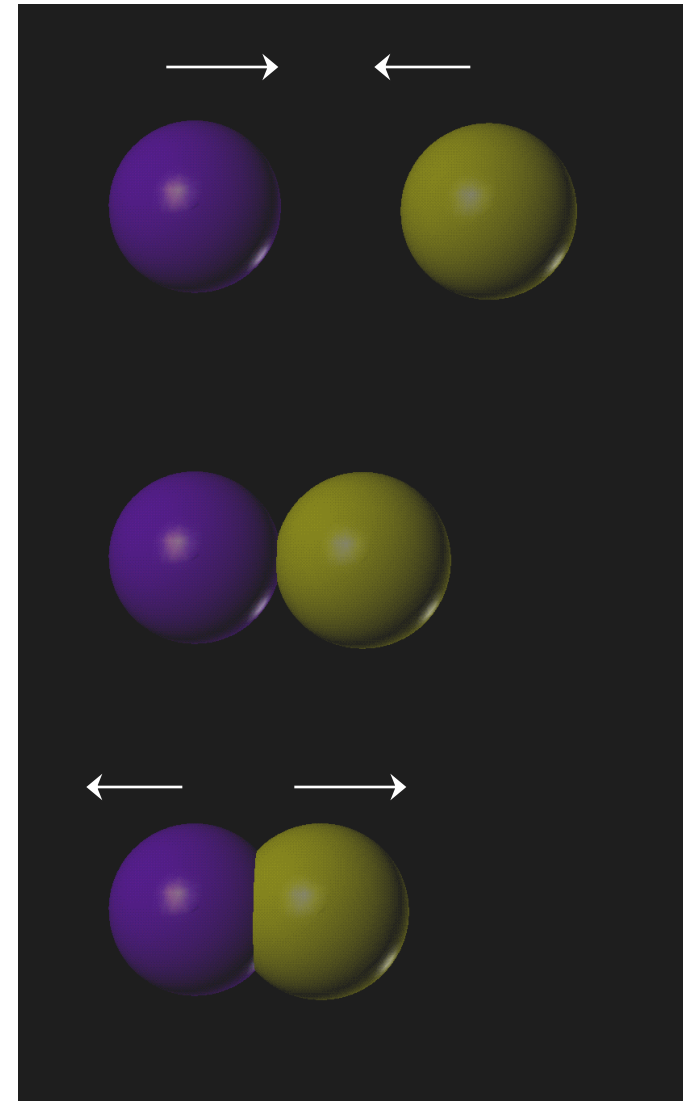
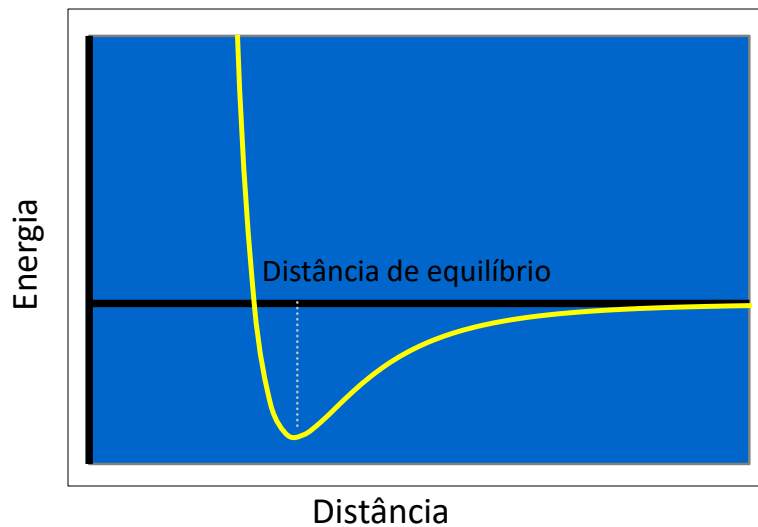
$$\sum_{i < j}^{N_{at}} \left[\frac{C_{12}(i, j)}{r_{ij}^{12}} - \frac{C_6(i, j)}{r_{ij}^6} \right]$$

C12 -Repulsivo

- Impede colisões

C6 - Atractivo

- Forças dispersivas



Interacções “non-bonded”: potencial electrostático

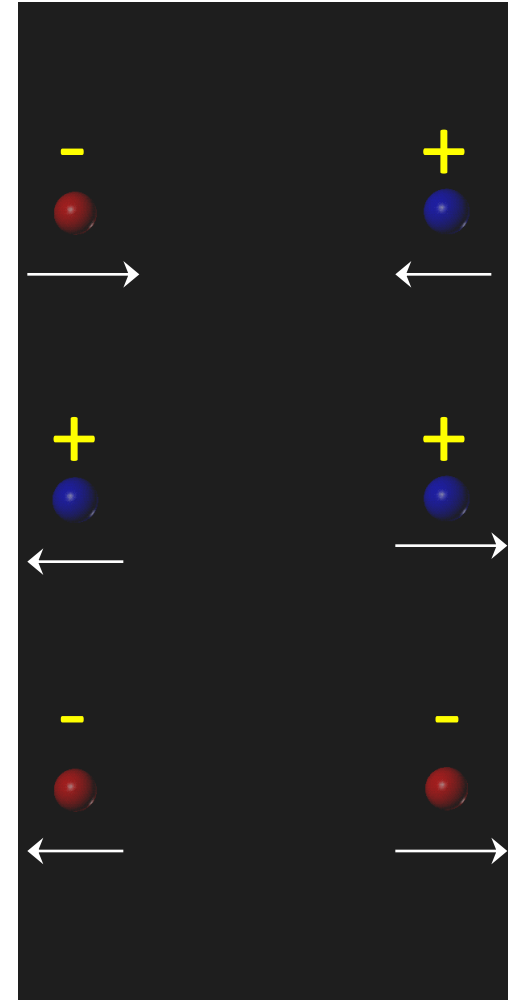
Pode ser descrito pela equação de Coulomb:

$$\sum_{i < j}^{N_{at}} \left[\frac{q_i q_j}{4 \pi \epsilon_0 \epsilon_r r_{ij}} \right]$$

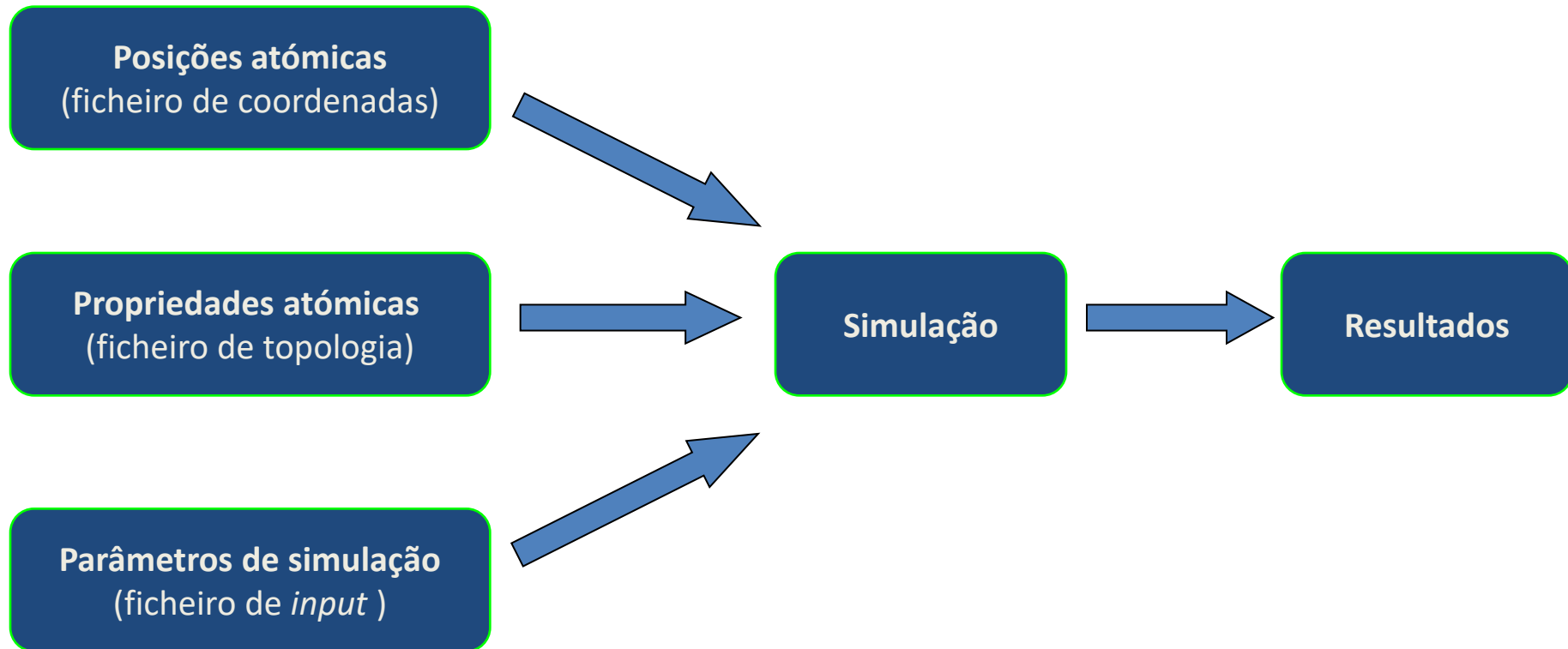
Colocam-se cargas pontuais nas posições atómicas

A interacção entre átomos ligados não é considerada (exclusão)

As cargas pontuais são geralmente obtidas por meio de cálculos quânticos



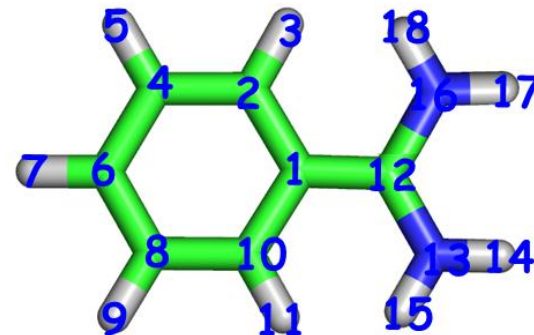
Procedimento



Topologia: depende do **campo de forças (forcefield)** usado

Simulação: pode ser feita com diferentes **aplicações de software**

Topologia da benzamida no formato Gromacs



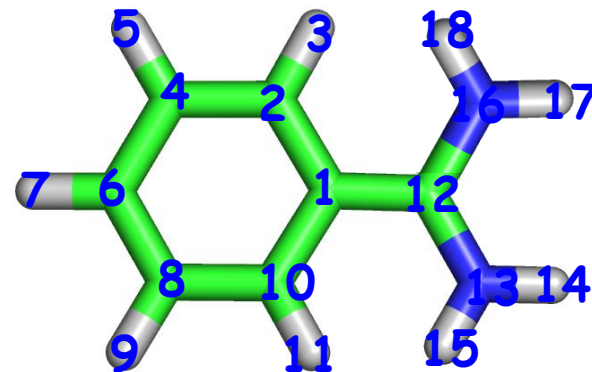
```
[ moleculetype ]  
; Name nrexcl  
Benz      3
```

```
[ atoms ]
```

; nr	type	resnr	residu	atom	cgnr	charge	mass
1	CB	1	BEN	C1	1	0	; qtot: 0
2	CR6	1	BEN	C2	2	-0.14	; qtot: -0.14
3	HCR	1	BEN	H2	2	0.14	; qtot: 0
4	CR6	1	BEN	C3	3	-0.14	; qtot: -0.14
5	HCR	1	BEN	H3	3	0.14	; qtot: 0
6	CR6	1	BEN	C4	4	-0.14	; qtot: -0.14
7	HCR	1	BEN	H4	4	0.14	; qtot: 0
8	CR6	1	BEN	C5	5	-0.14	; qtot: -0.14
9	HCR	1	BEN	H5	5	0.14	; qtot: 0
10	CR6	1	BEN	C6	6	-0.14	; qtot: -0.14
11	HCR	1	BEN	H6	6	0.14	; qtot: 0
12	C	1	BEN	C7	7	0.56	; qtot: 0.56
13	NZ	1	BEN	N1	7	-0.26	; qtot: 0.3
14	H	1	BEN	H11	7	0.24	; qtot: 0.54
15	H	1	BEN	H12	7	0.24	; qtot: 0.78
16	NZ	1	BEN	N2	7	-0.26	; qtot: 0.52
17	H	1	BEN	H21	7	0.24	; qtot: 0.76
18	H	1	BEN	H22	7	0.24	; qtot: 1

(continua)

Topologia da benzamida no formato Gromacs

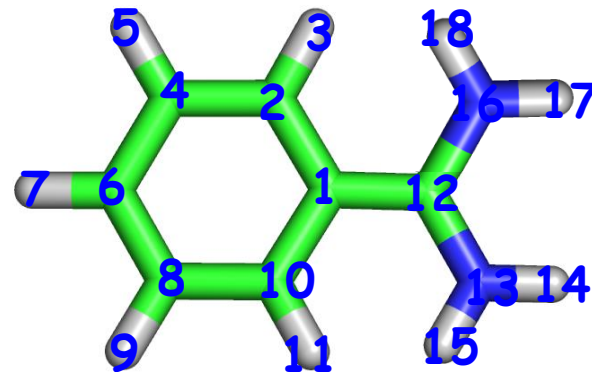


```
[ bonds ]  
; ai aj funct  
1 2 1  
1 10 1  
1 12 1  
2 3 1  
2 4 1  
4 5 1  
4 6 1  
6 7 1  
6 8 1  
8 9 1  
8 10 1  
10 11 1  
12 13 1  
12 16 1  
13 14 1  
13 15 1  
16 17 1  
16 18 1
```

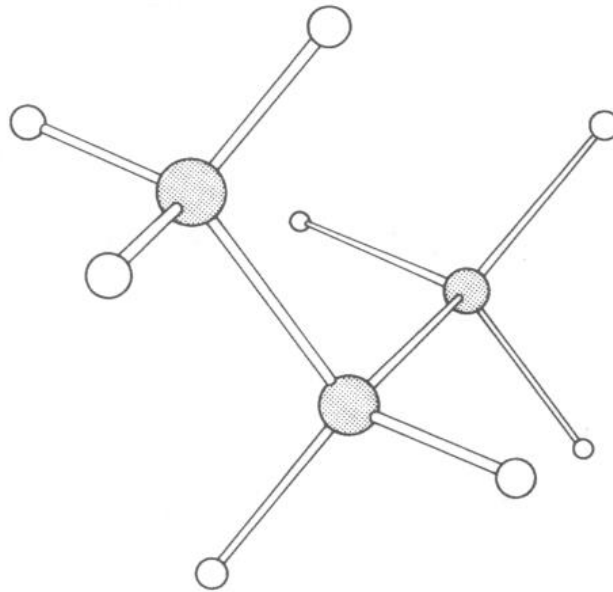
(continua)

Topologia da benzamida no formato Gromacs

```
[ angles ]
; ai   aj   ak funct
  2     1   10     1
  2     1   12     1
 10     1   12     1
  1     2    3     1
  1     2    4     1
  3     2    4     1
  2     4    5     1
  2     4    6     1
  5     4    6     1
  4     6    7     1
  4     6    8     1
  7     6    8     1
  6     8    9     1
  6     8   10     1
  9     8   10     1
  1    10    8     1
  1    10   11     1
  8    10   11     1
  1    12   13     1    120.    400.
  1    12   16     1    120.    400.
 13    12   16     1
 12    13   14     1
 12    13   15     1
 14    13   15     1
 12    16   17     1
 12    16   18     1
 17    16   18     1
```



O número de parâmetros do campo de forças cresce rapidamente com a dimensão da molécula!



10 bond terms + 18 angle terms + 18 torsion terms + 27 non-bond terms
11 atoms → 73 parameters!

Especificidade dos parâmetros do campo de forças

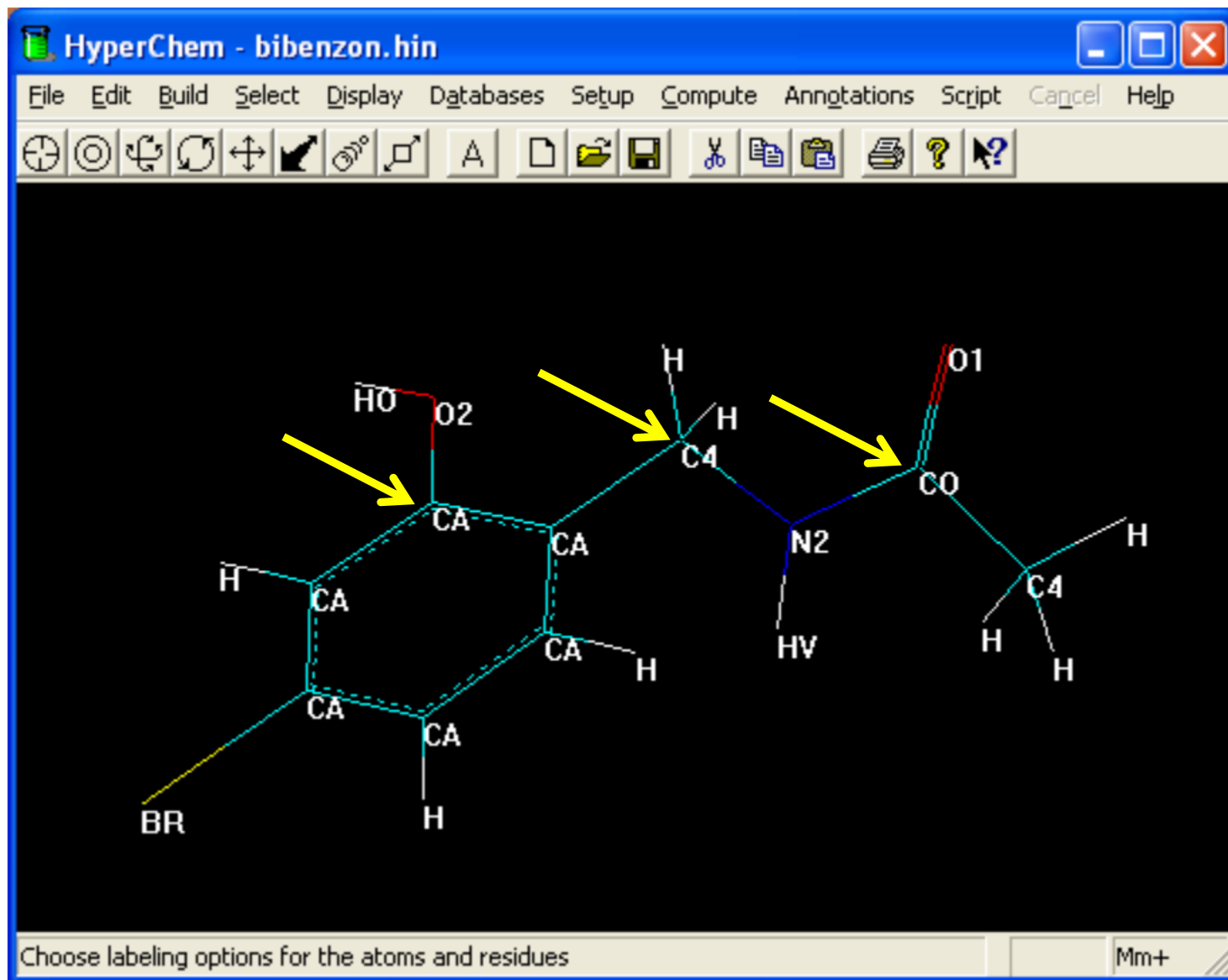
Os parâmetros do campo de forças são específicos para cada tipo de átomo (não é o mesmo que elemento) e para cada interação.

- Exemplo: Existem vários tipos de átomos de carbono, dependendo do ambiente químico circundante.

EXEMPLO: vários tipos de átomo de carbono no campo de forças GROMOS

	<i>Peptide bond carbon</i>	<i>Carbon in Carboxyl (Asp, Glu)</i>	<i>β carbon in Ser</i>	<i>β carbon in Val</i>	<i>γ carbon in His</i>
<i>Atom type</i>	C	C	CH2	CH1	C
<i>Charge</i>	0.380	0.270	0.150	0.000	-0.050

Exemplo: diferentes tipos de parâmetros para átomos de carbono no campo de forças MM+



Aplicações

A aproximação clássica do campo de forças torna muito mais simples o cálculo da energia de um sistema molecular, permitindo trabalhar com macromoléculas como proteínas e ácidos nucleicos

Análise conformacional: estudo das conformações acessíveis a uma molécula e das suas energias relativas. Determinação dos mínimos de energia em função da conformação (minimização). Cálculos de mecânica molecular.

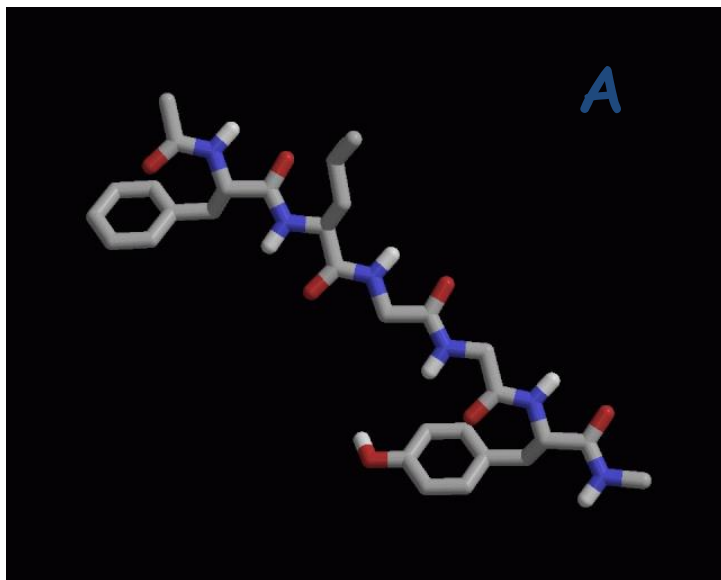
Dinâmica molecular: permite seguir a evolução temporal de um sistema molecular ou seja, as conformações que este sistema pode assumir a uma determinada pressão e temperatura, em virtude da agitação molecular. (Cálculo de frequências de vibração, transições conformacionais, etc).

Interacções proteína-ligando (docking): cálculo da interacção entre moléculas.

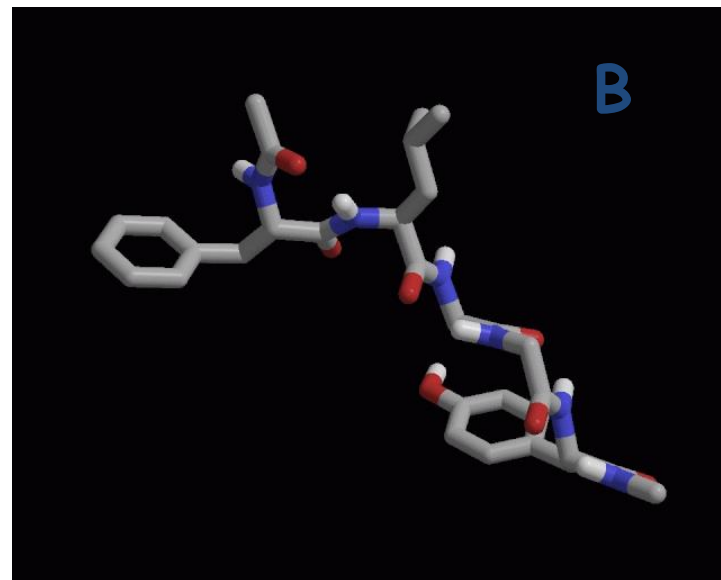
A energia é função das posições atômicas

A conformação “A” do oligopéptido tem uma energia mais alta que a “B”, logo será mais instável.

Nota: o valor de energia depende de campo de forças utilizado



V=186.2 kcal/mol

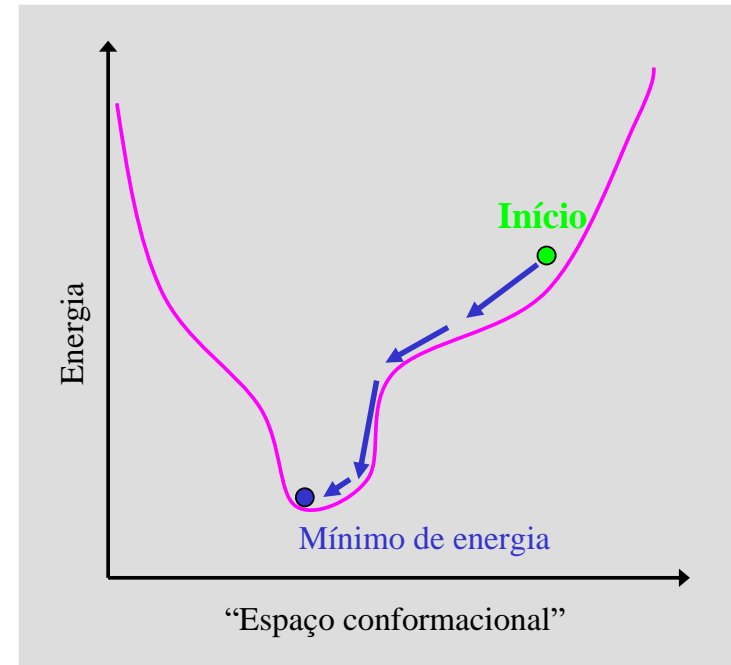


V=-51.7 kcal/mol

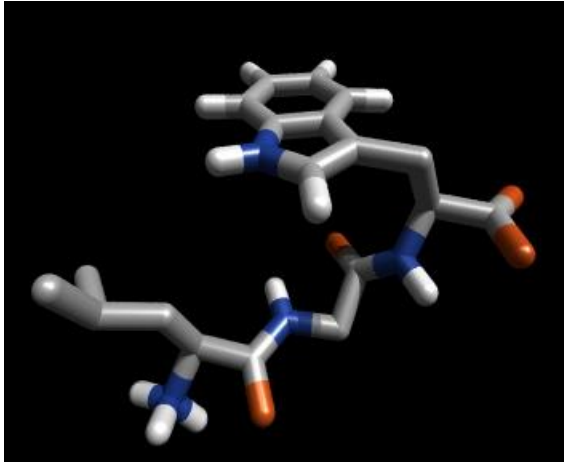
Minimização da energia

Partindo de uma conformação inicial, achar uma conformação de energia **mínima** para a molécula

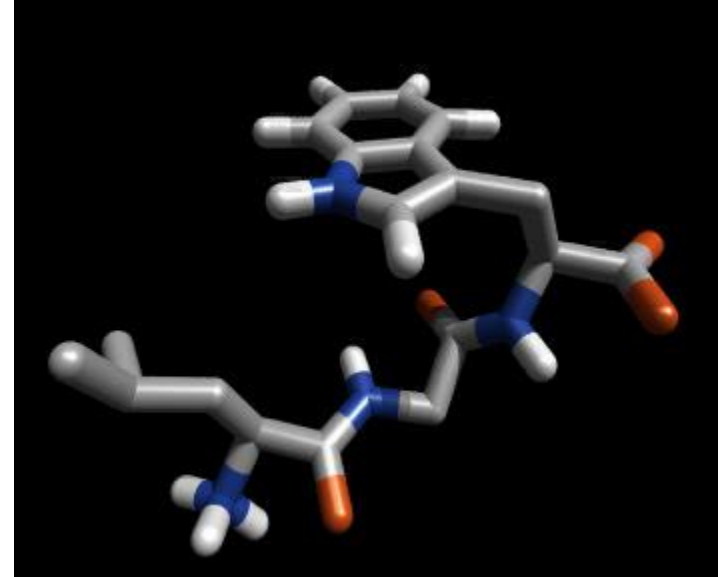
- Mover os átomos
Busca conformacional
- Calcular a energia
Campo de forças
- Encontrar conformações de energia mais baixa
Mínimos na hipersuperfície de energia potencial



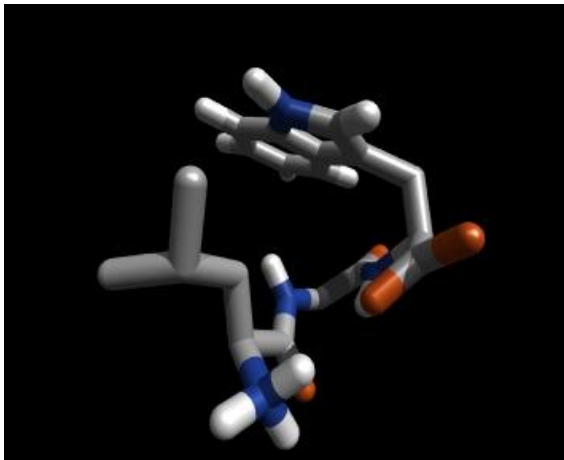
Minimização de energia



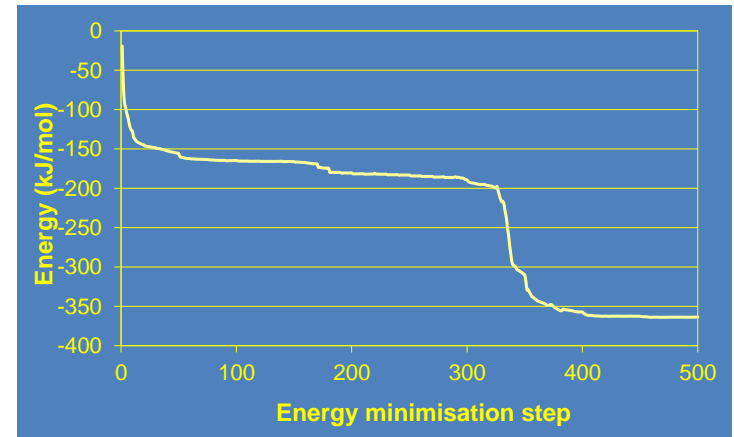
Conformação inicial



Tripéptido Leu-Gly-Trp

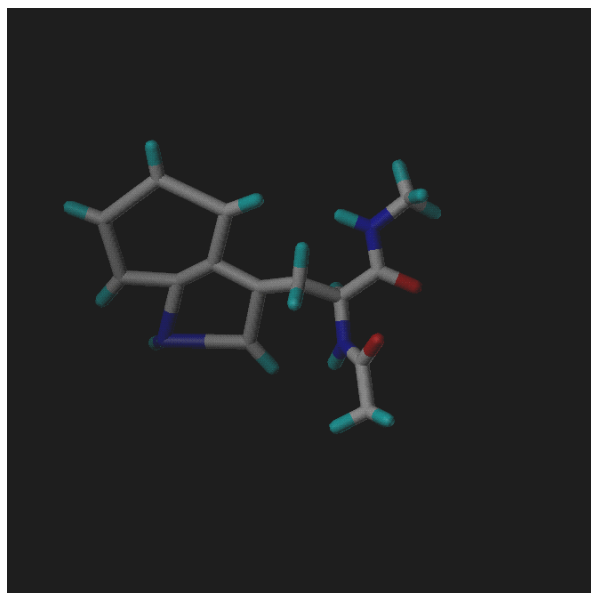


Após 500 passos de minimização

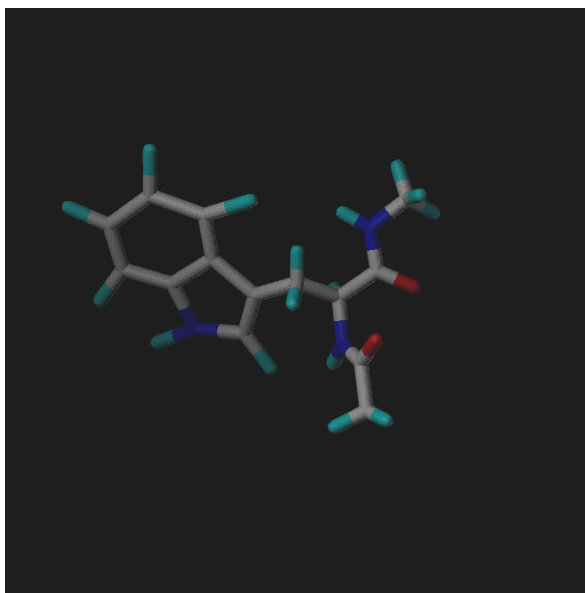


Construção de estruturas moleculares

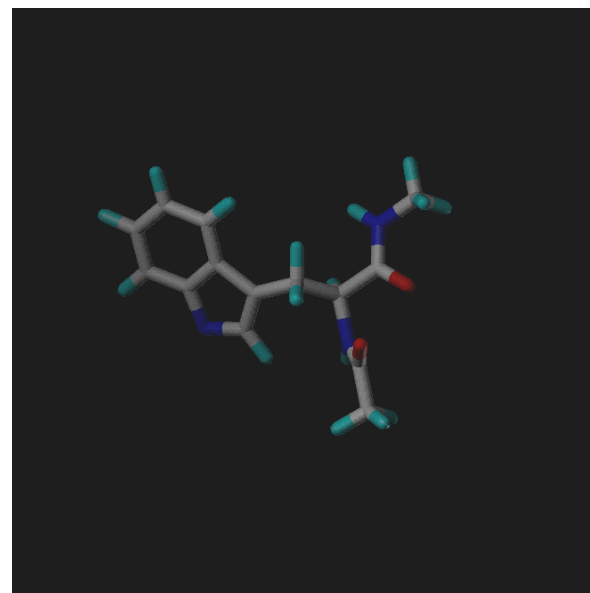
A construção interactiva de pequenas moléculas no computador é geralmente auxiliada pela minimização de energia, que conduz a uma estrutura correcta do ponto de vista da geometria molecular



Construída no écran



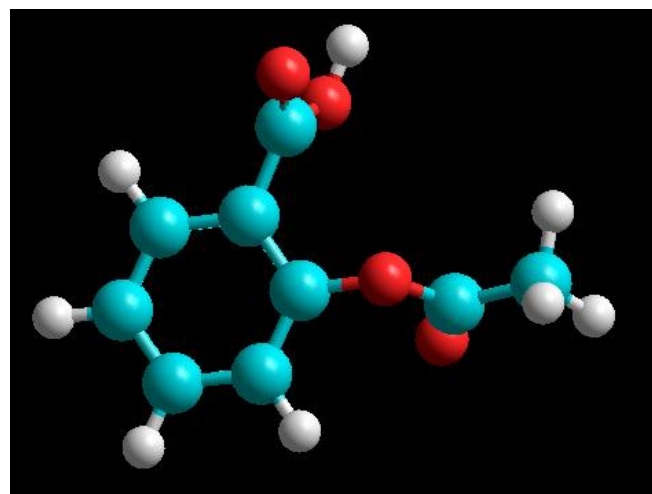
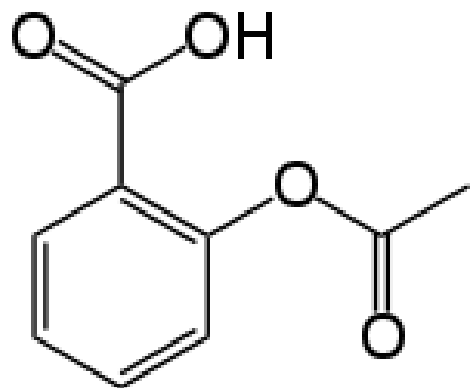
Alguns passos de
minimização



Estrutura final

As moléculas pequenas podem ser “construídas” no computador

- A partir da fórmula de um composto químico é possível construir um modelo da sua estrutura molecular no computador:



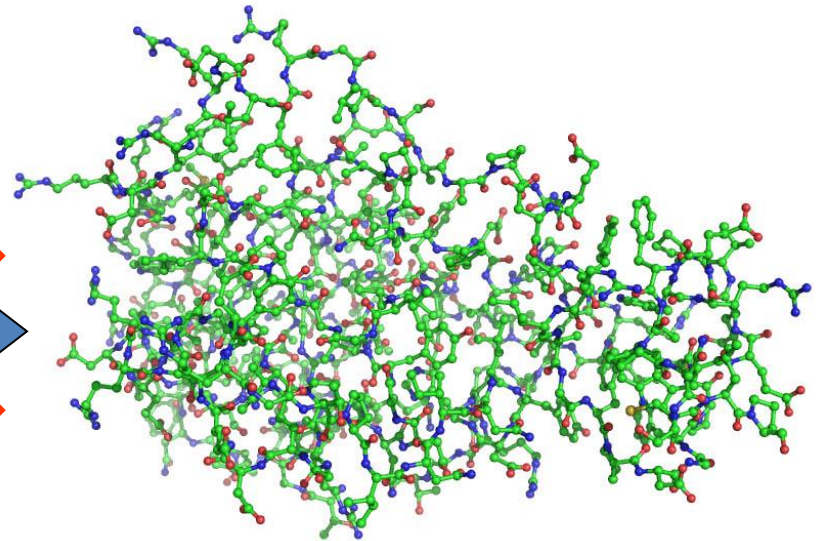
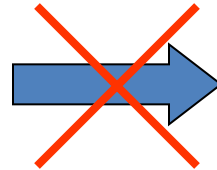
Fórmula da aspirina

Modelo molecular
da aspirina

A construção de macromoléculas biológicas é muito mais difícil!

- As macromoléculas biológicas (proteínas, ácidos nucleicos) têm um número muito elevado de átomos e de possíveis conformações, pela que sua construção directa não é geralmente possível

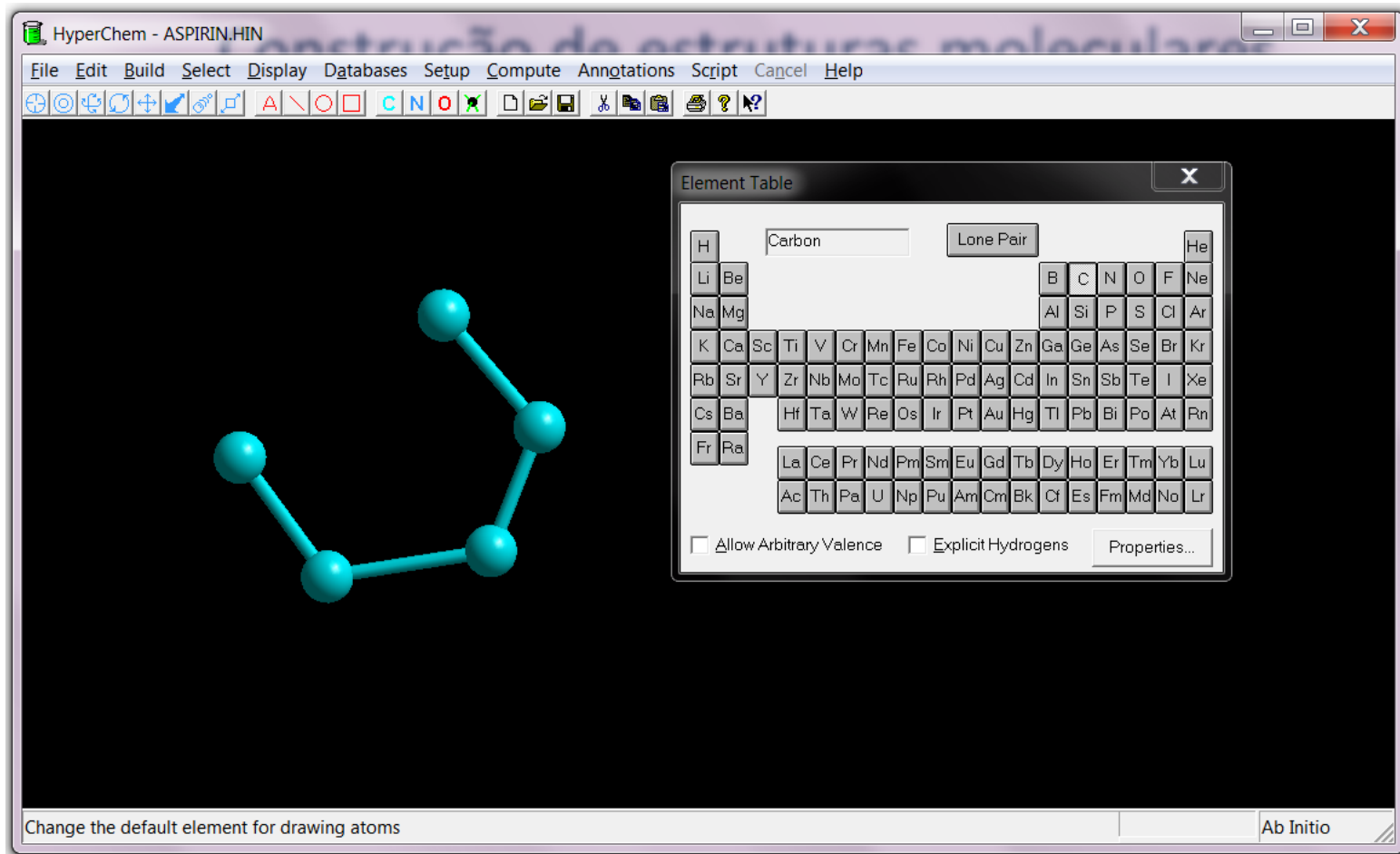
...AVAGGATILVHNQDAGEPAIVLAFG...



“Fórmula” de uma
proteína

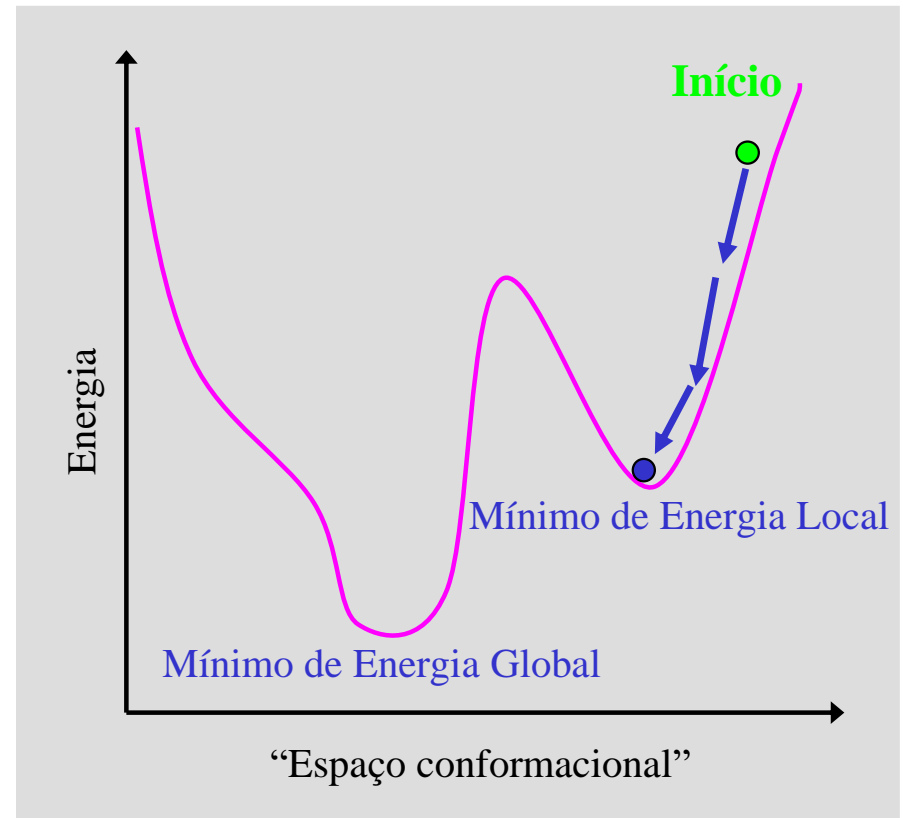
Modelo molecular
de uma proteína

Construção de estruturas moleculares usando o programa Hyperchem



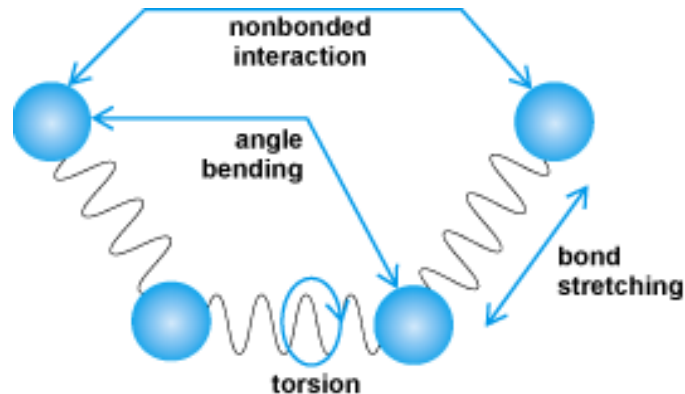
O problema da minimização

- A superfície de energia é muito complexa.
- Grande número de mínimos locais.
- É fácil ficar “encravado” num mínimo local e falhar uma estrutura mais estável para a molécula

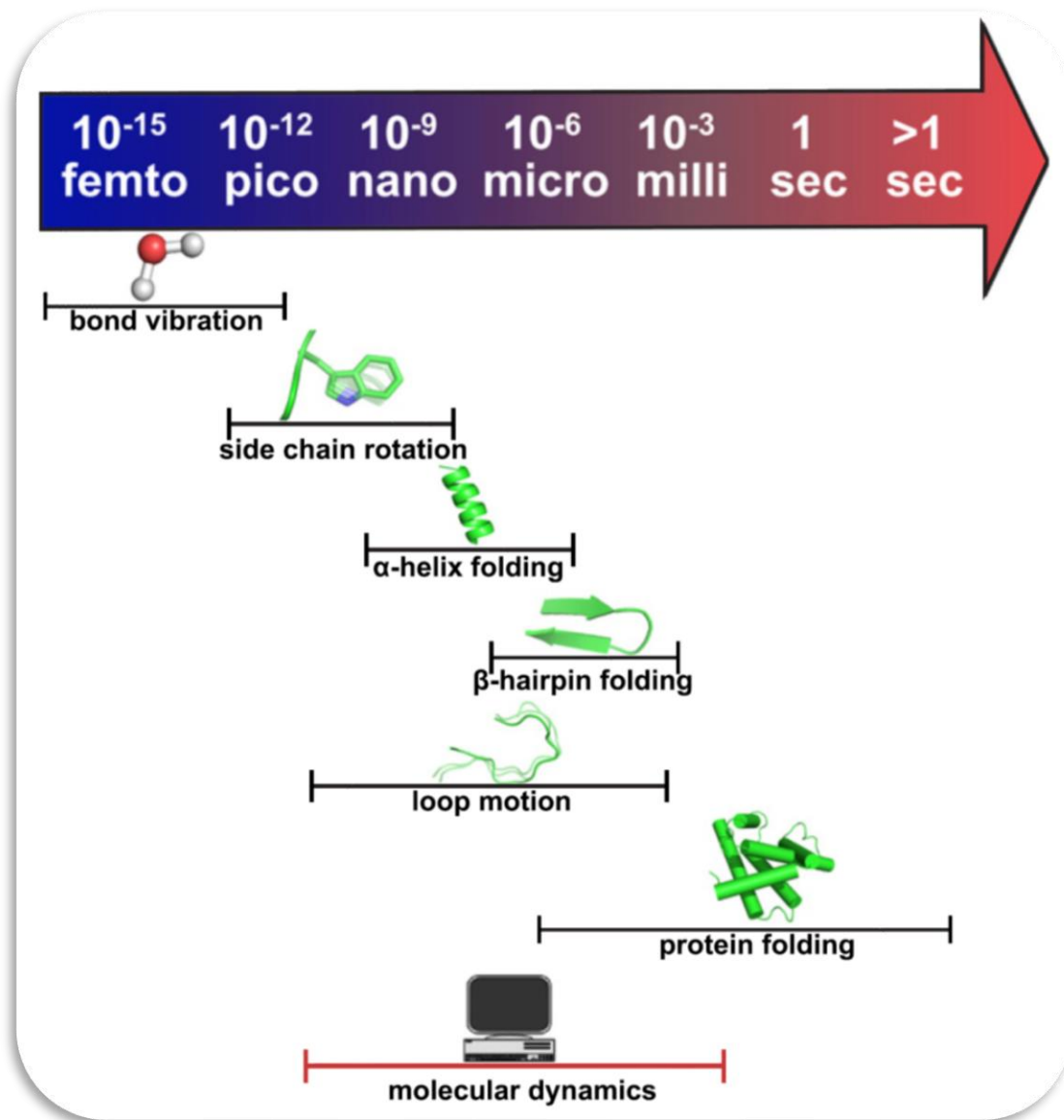


Dinâmica molecular

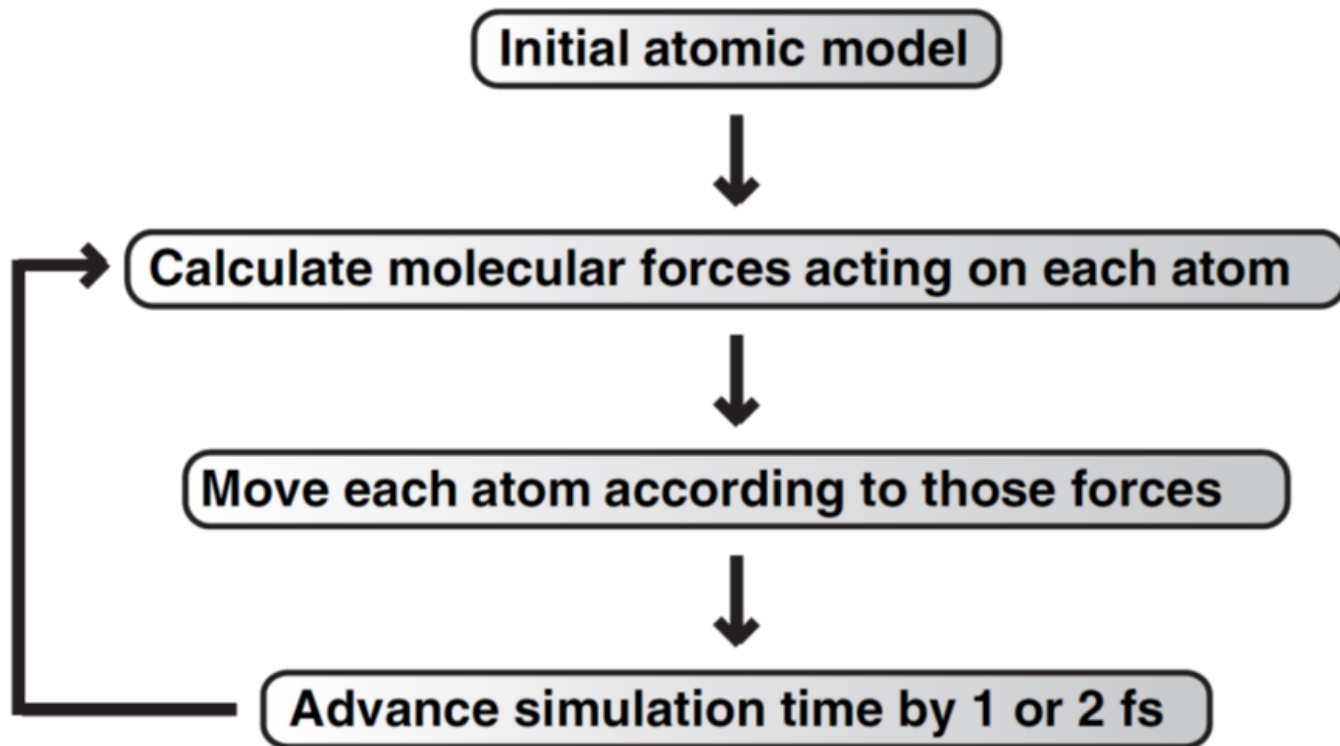
- As equações do movimento para um sistema de partículas com massas m_1, \dots, m_n sujeitas a forças F_1, \dots, F_n podem ser integradas por métodos computacionais
- A solução obtida é uma **trajectória** do sistema de partículas
- Quando as partículas são átomos e as forças são calculadas a partir de um **campo de forças**, pode-se calcular a trajectória do sistema molecular
- As simulações de MD (**M**olecular **D**ynamics) permitem explorar o espaço conformacional das moléculas, estudar a evolução temporal do sistema e calcular propriedades dependentes da entropia, como as energias livres



Escala de tempo dos movimentos moleculares



Procedimento básico da dinâmica molecular



Dinâmica molecular: teoria

Equations de Newton

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2}$$

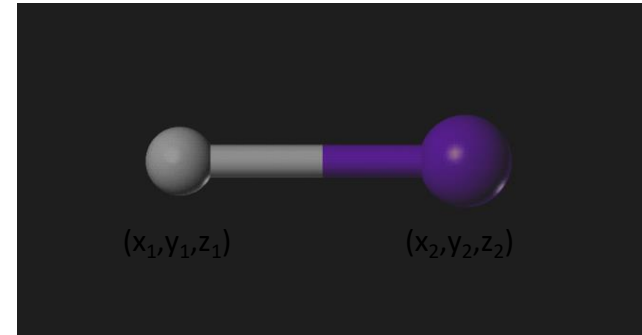
$$\frac{d^2 \mathbf{r}_i(t)}{dt^2} = \frac{\mathbf{F}_i}{m_i}$$

Derivação da força a partir do potencial:

$$\mathbf{F} = -\nabla V(\mathbf{r}_i, \dots, \mathbf{r}_n)$$

$$\mathbf{F}_i = \frac{-\partial V(\mathbf{r}_i, \dots, \mathbf{r}_n)}{\partial \mathbf{r}_i}$$

Ex: Força a partir do potencial de ligação



$$V = \frac{1}{2} K(b - b_0)^2$$

$$b = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

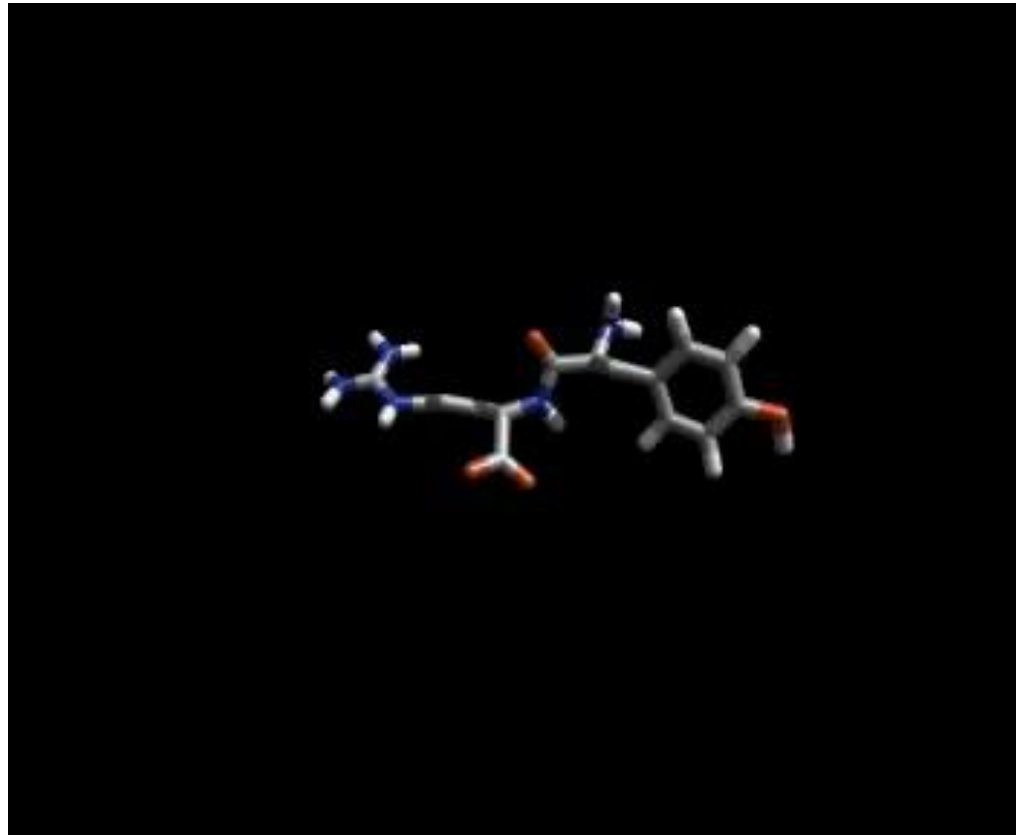
$$\frac{\partial V}{\partial x_1} = \frac{\partial V}{\partial b} \frac{\partial b}{\partial x_1}$$

$$\frac{\partial b}{\partial x_1} = \frac{(x_1 - x_2)}{b}$$

$$\frac{\partial V}{\partial b} = K(b - b_0)$$

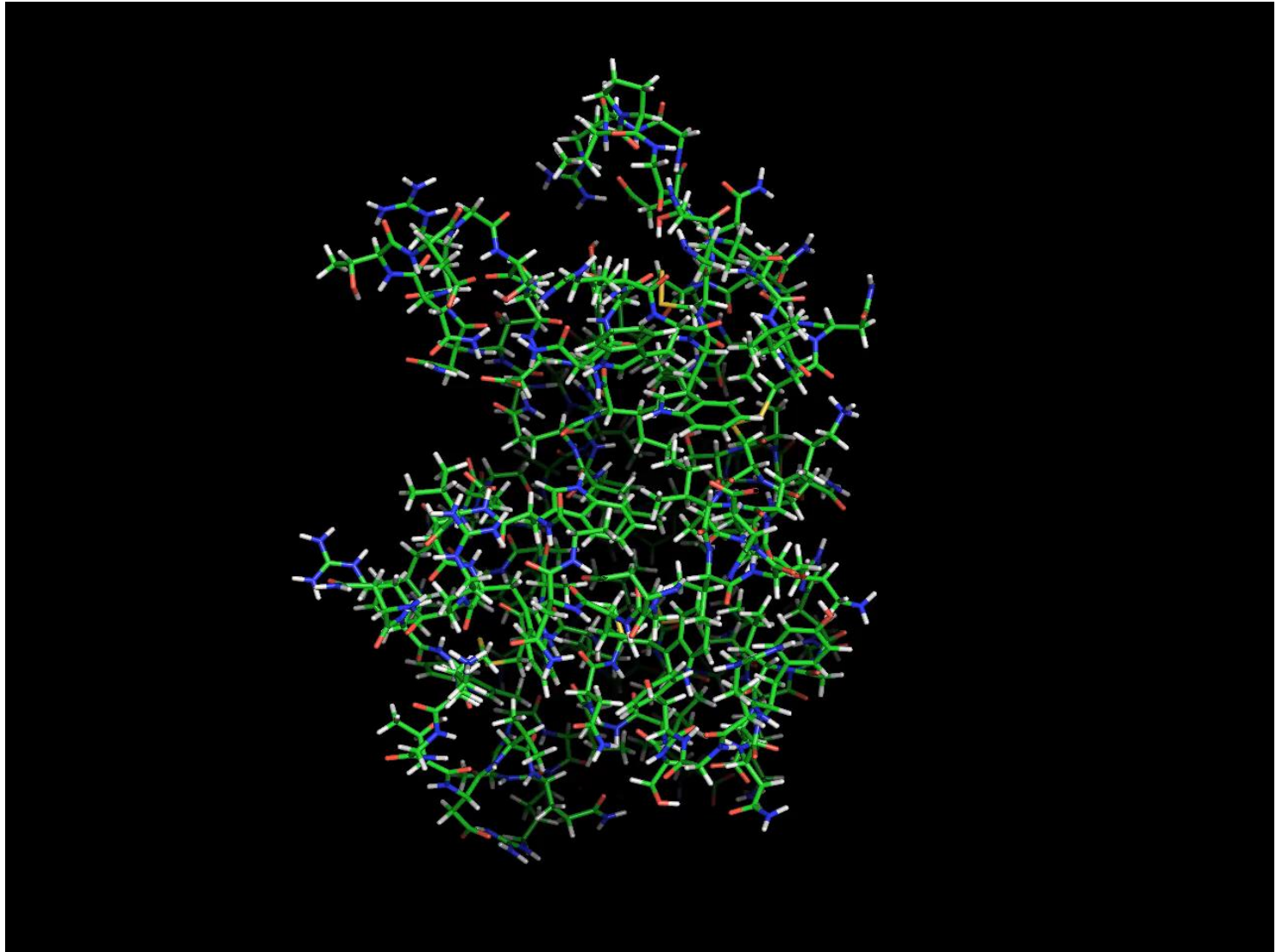
$$F_{x_1} = -\frac{\partial V}{\partial x_1} = -\frac{K(b - b_0)(x_1 - x_2)}{b}$$

Ex.:dinâmica molecular da kiotorfina (L-tyrosil-L-arginina)

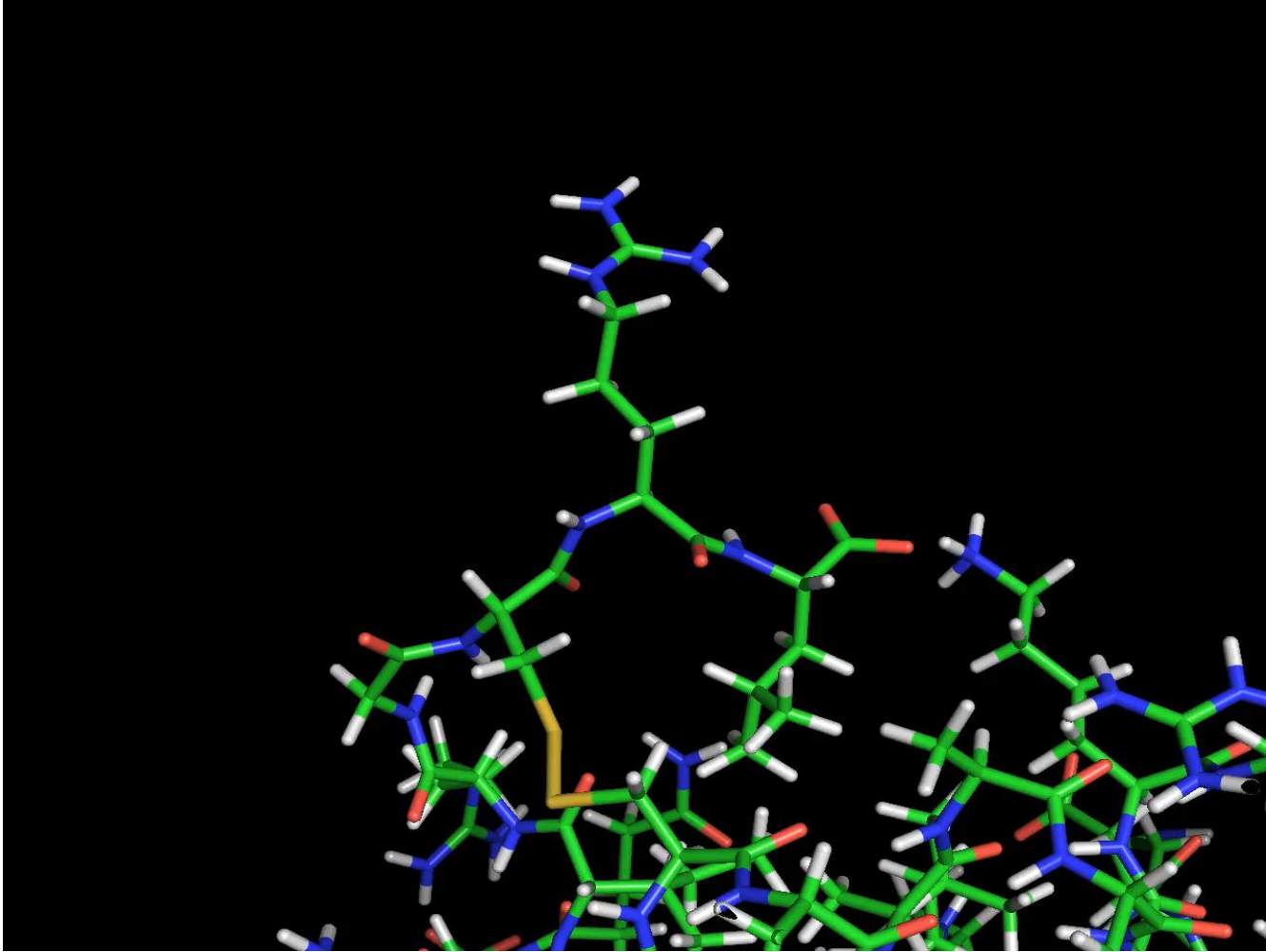


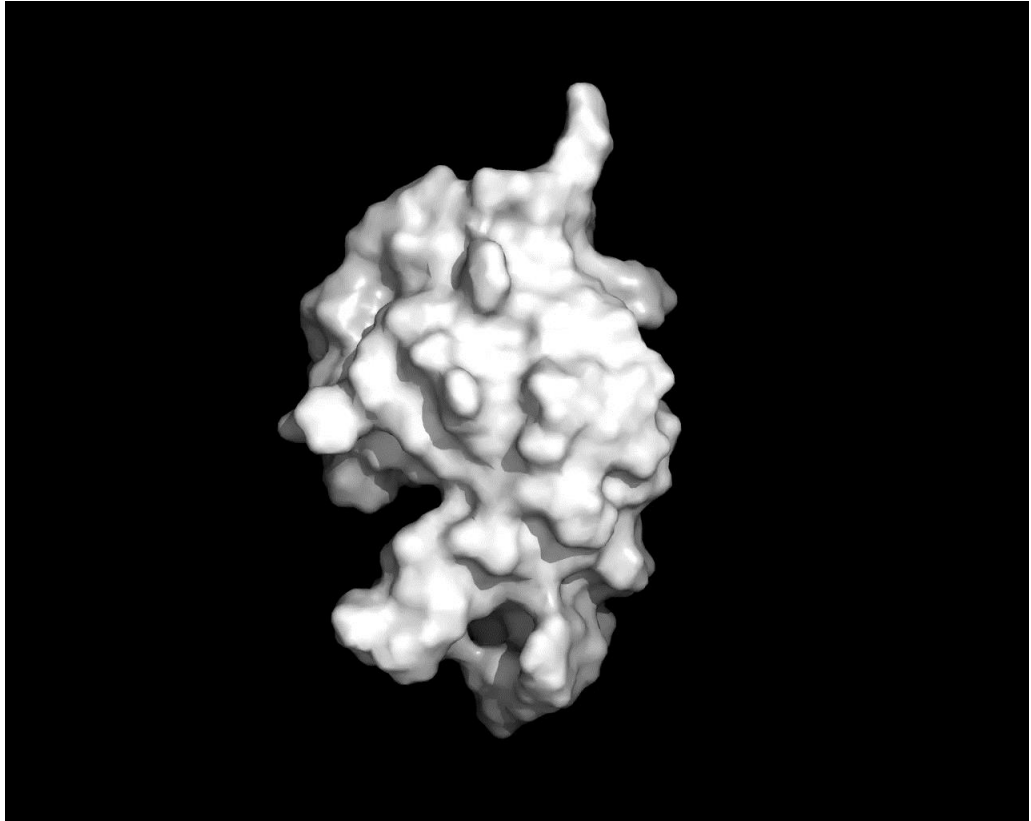
Tempo de simulação: 20ps (1ps = 10^{-12} s)

Dinâmica molecular da lisozima

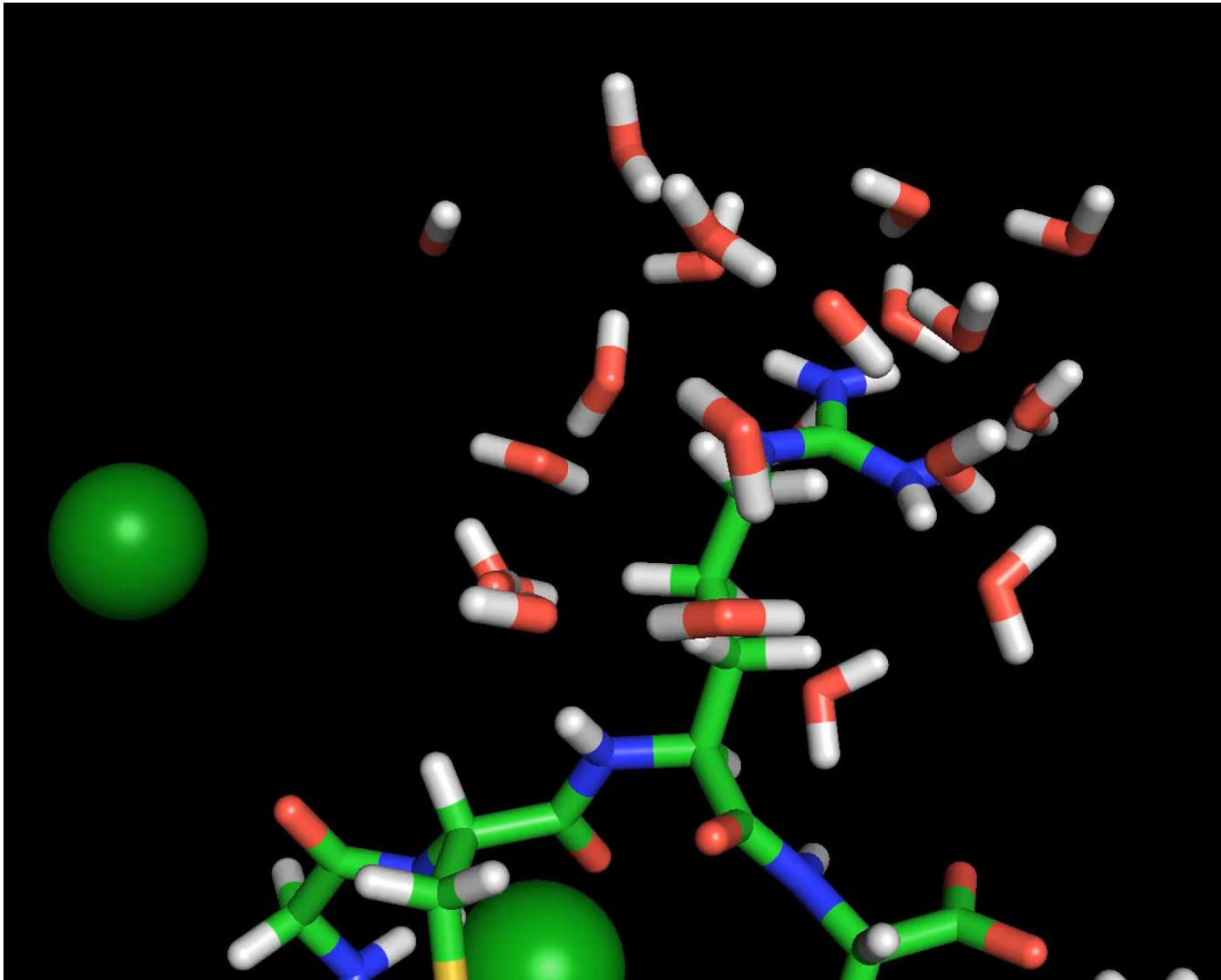


Dinâmica molecular da lisozima



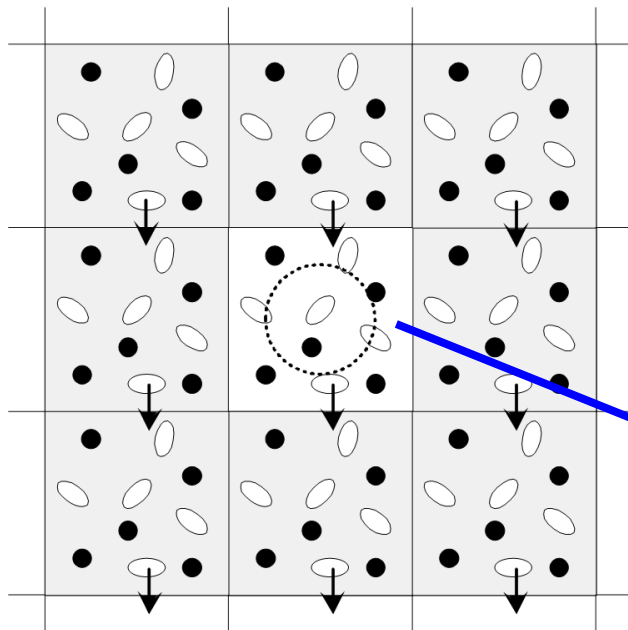


MD: Água e íons junto de um resíduo de arginina



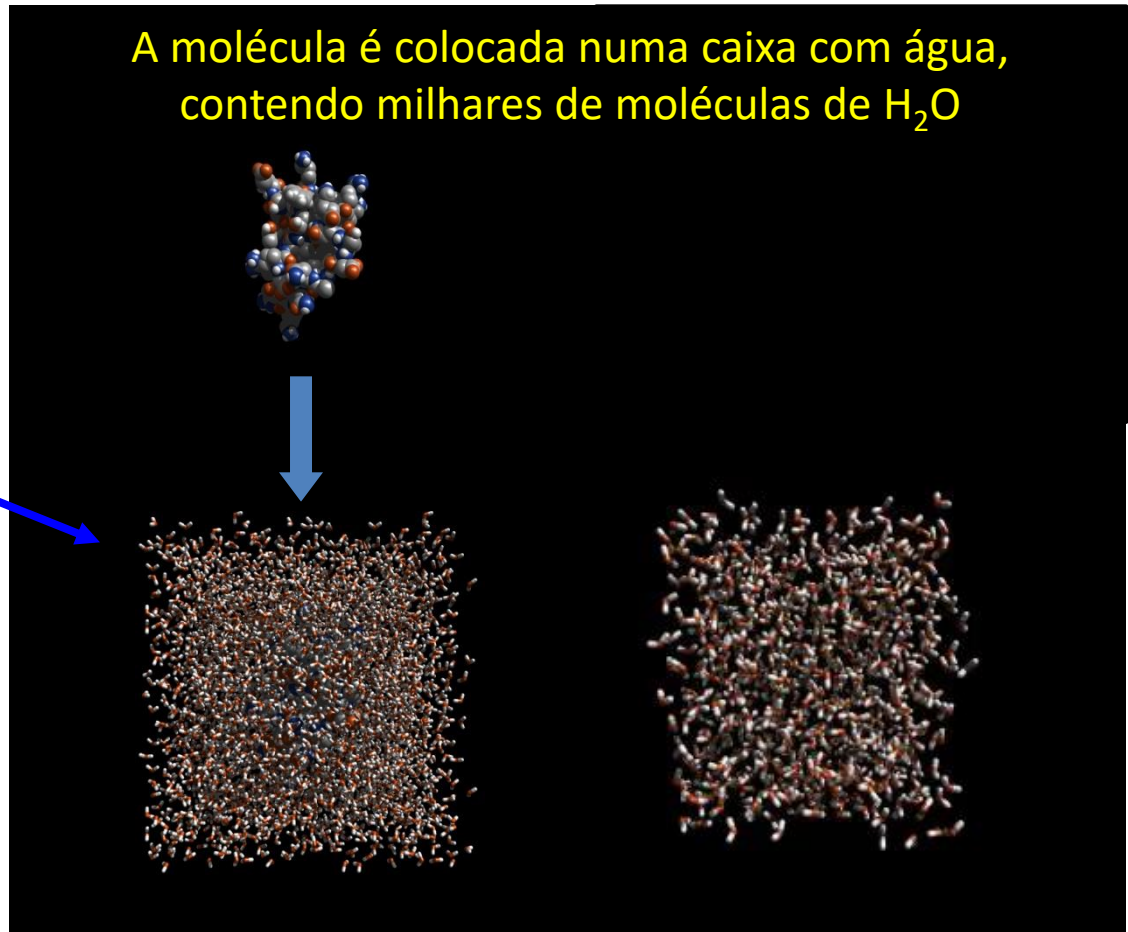
Inclusão do meio aquoso

- Uma simulação realista dos movimentos moleculares de uma molécula de proteína deve incorporar o *meio aquoso* em que a proteína se encontra imersa.



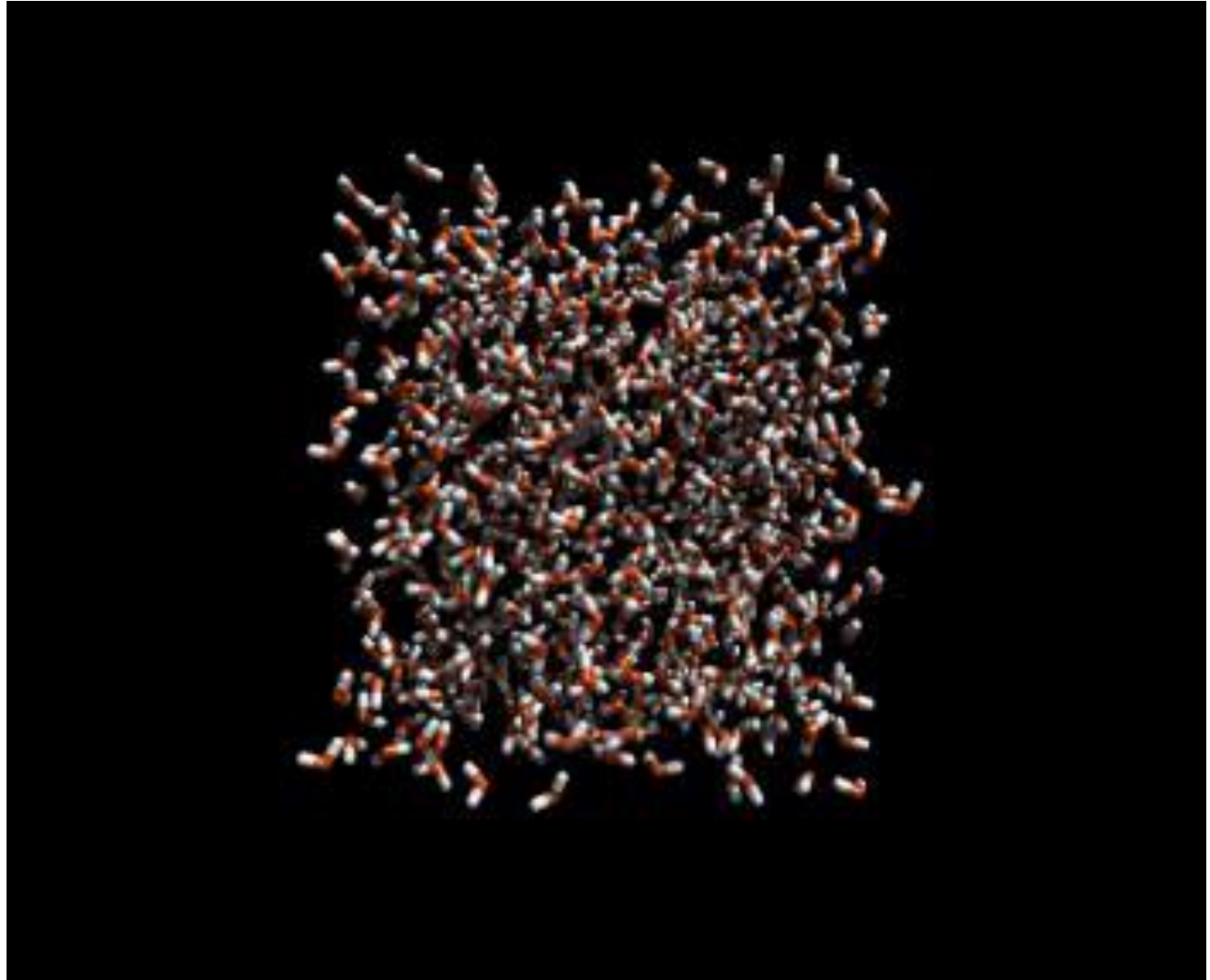
Condições fronteira **periódicas**

A molécula é colocada numa caixa com água,
contendo milhares de moléculas de H₂O

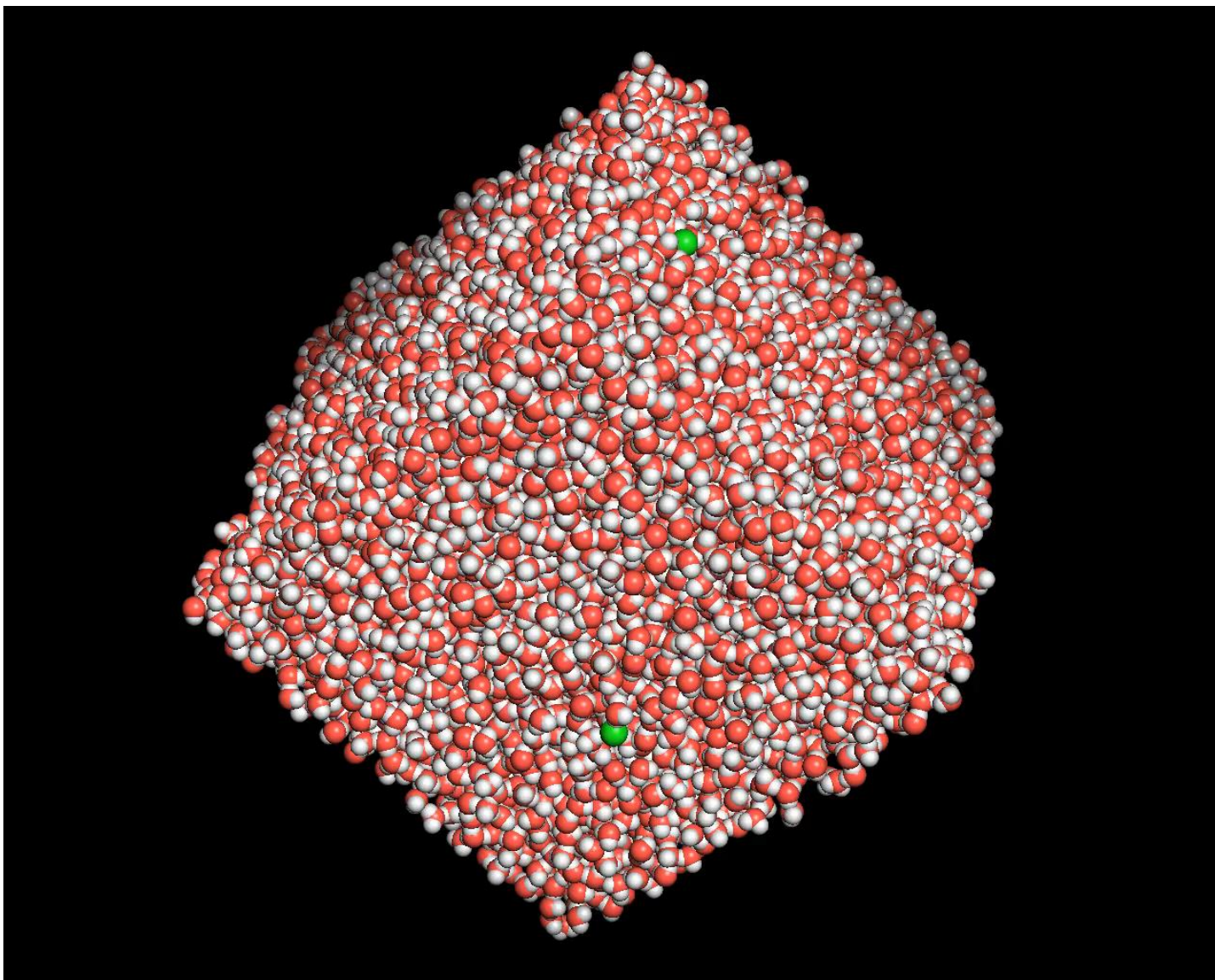


Dinâmica do meio aquoso

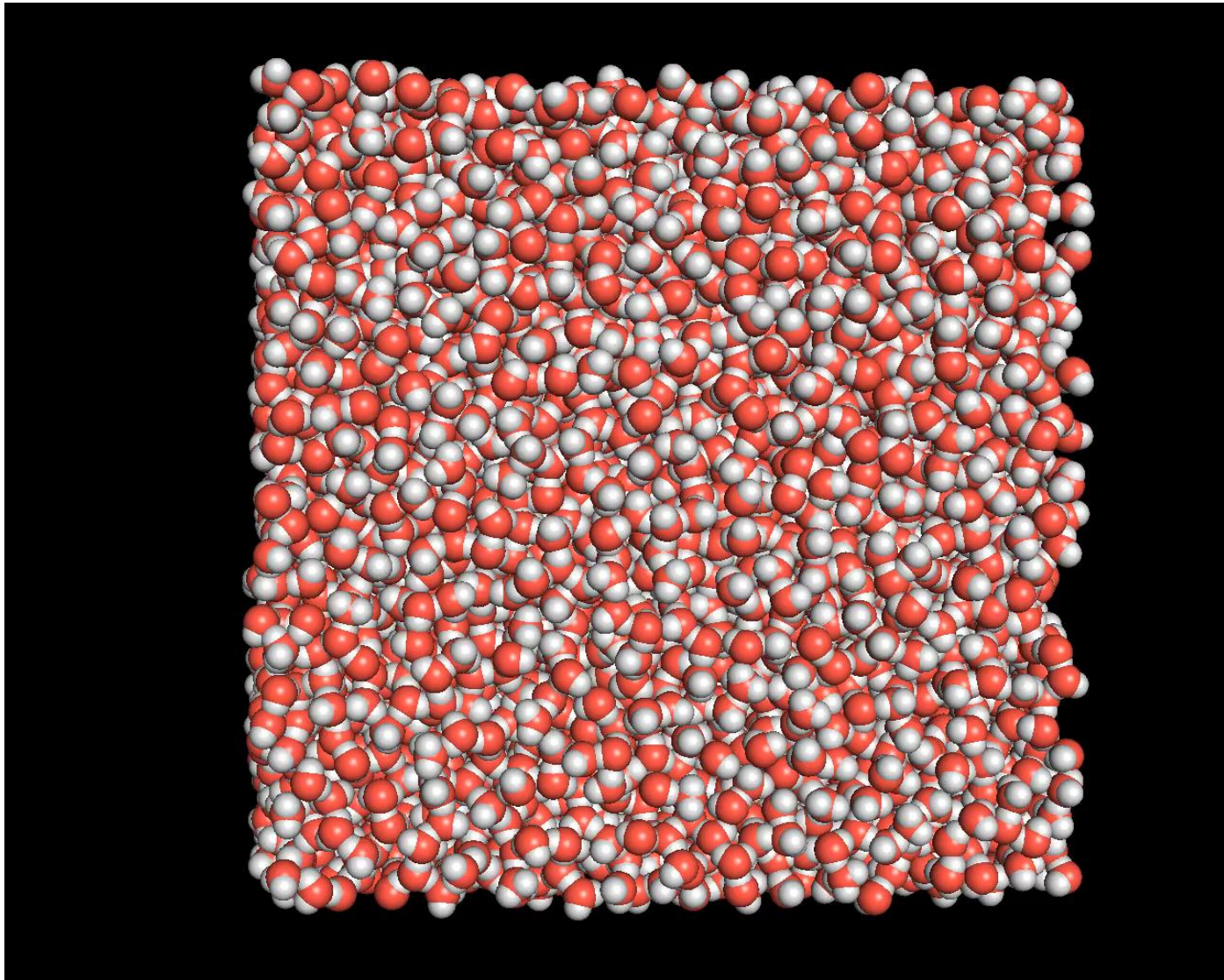
Água: 0-20ps
(1ps = 10^{-12} s)



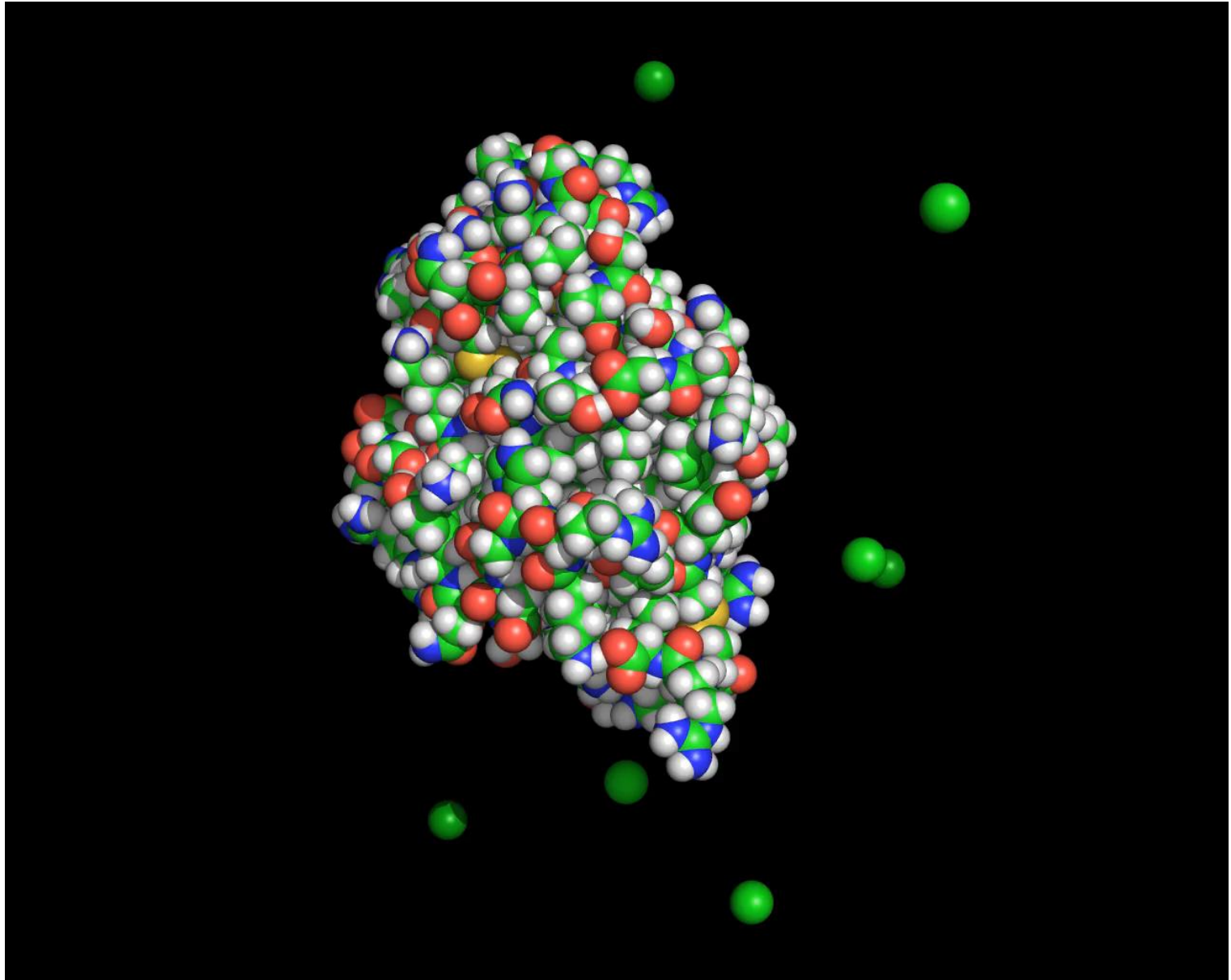
MD: caixa de solvente



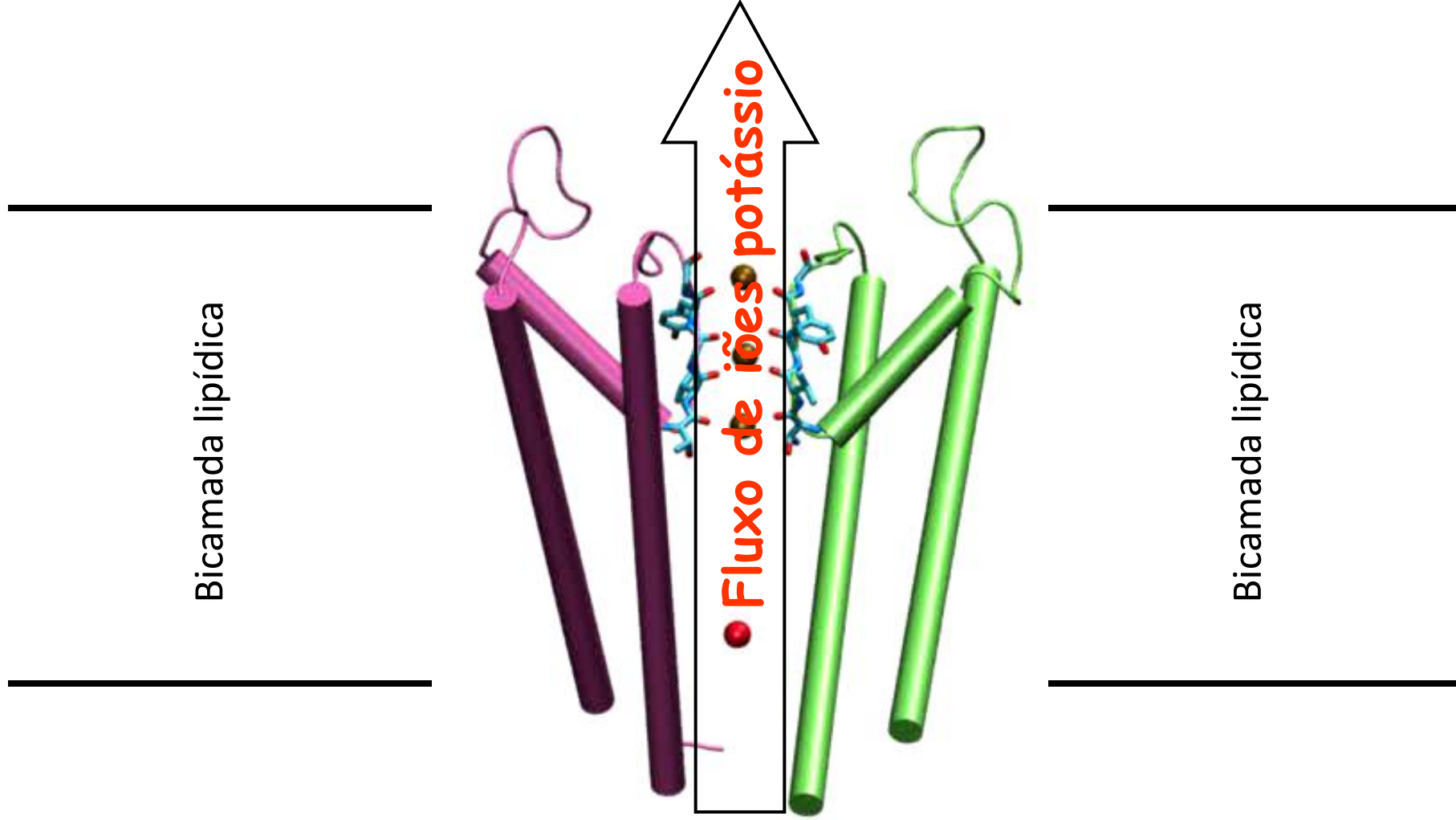
MD: caixa de solvente



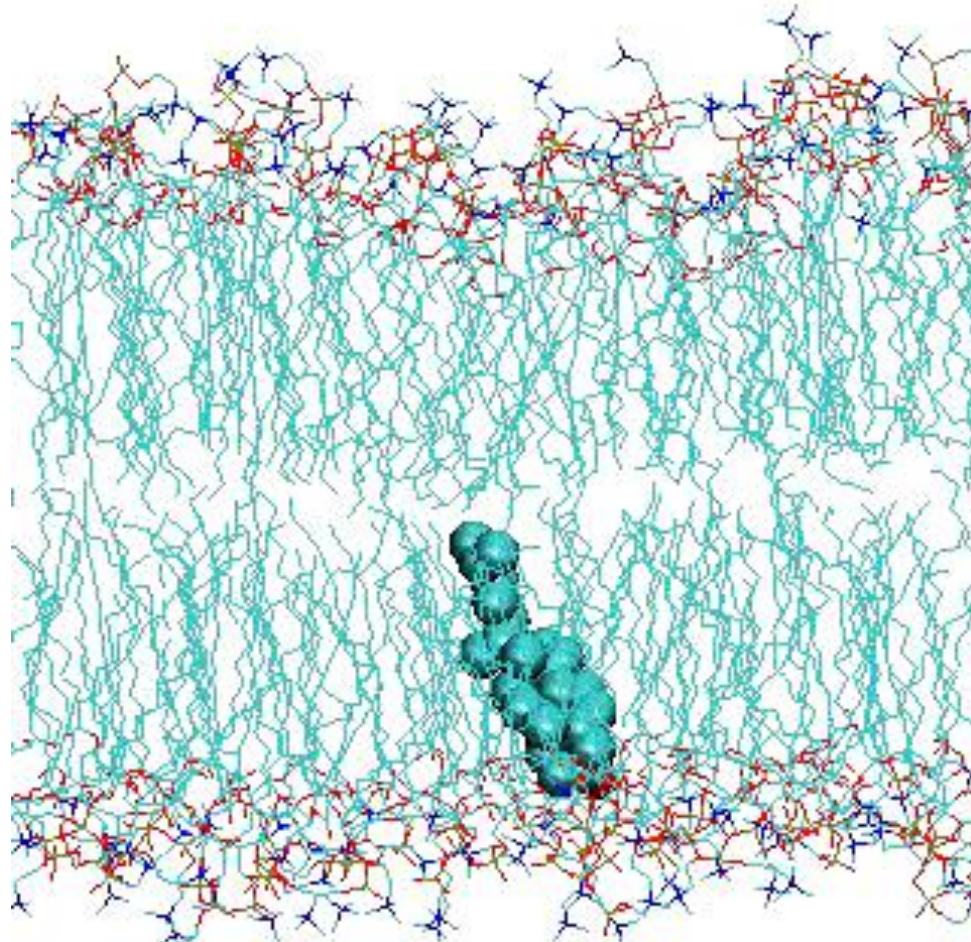
MD: Liszoima+iões cloro (10 ps)



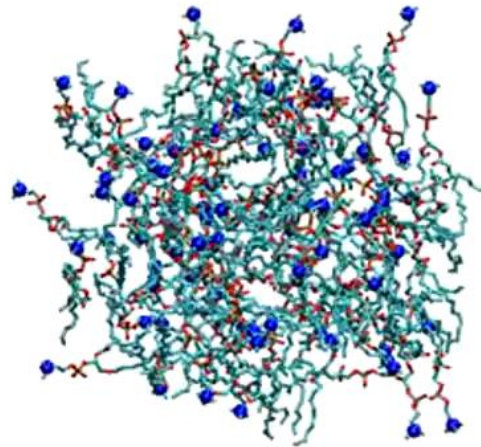
Dinâmica molecular de um canal de potássio



Ex.:dinâmica de um esteroide em membrana lipídica

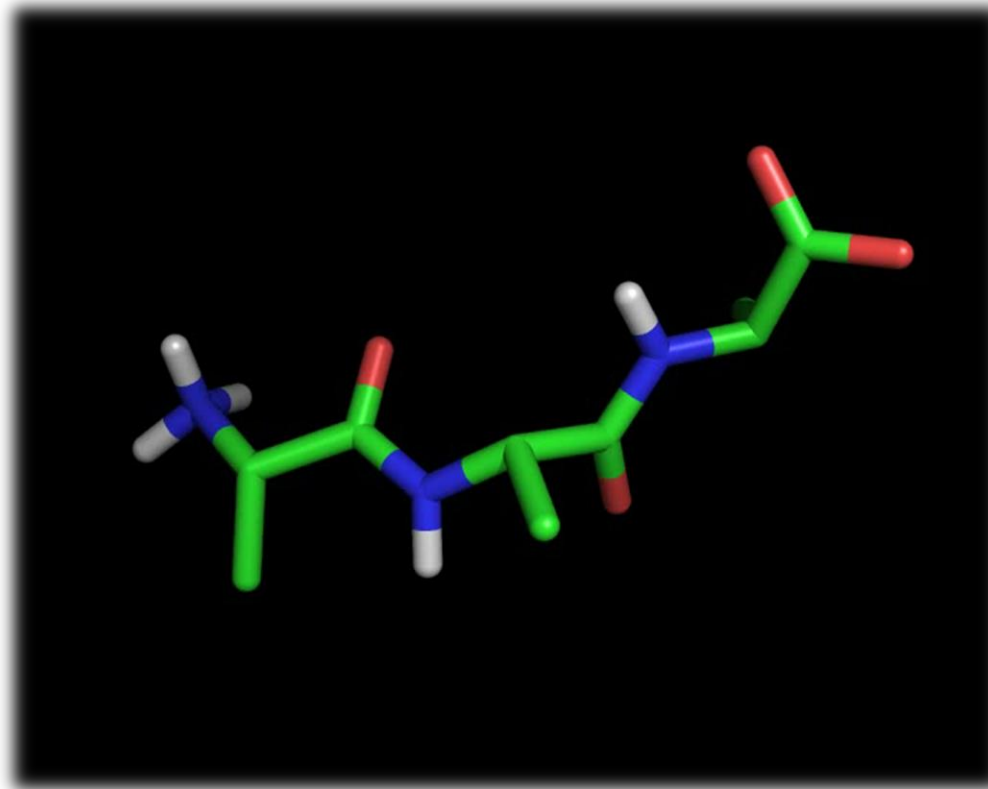


Formação espontânea de uma bicamada lipídica num simulação de MD



Importância da solvatação

Tripéptido Ala-Ala-Ala

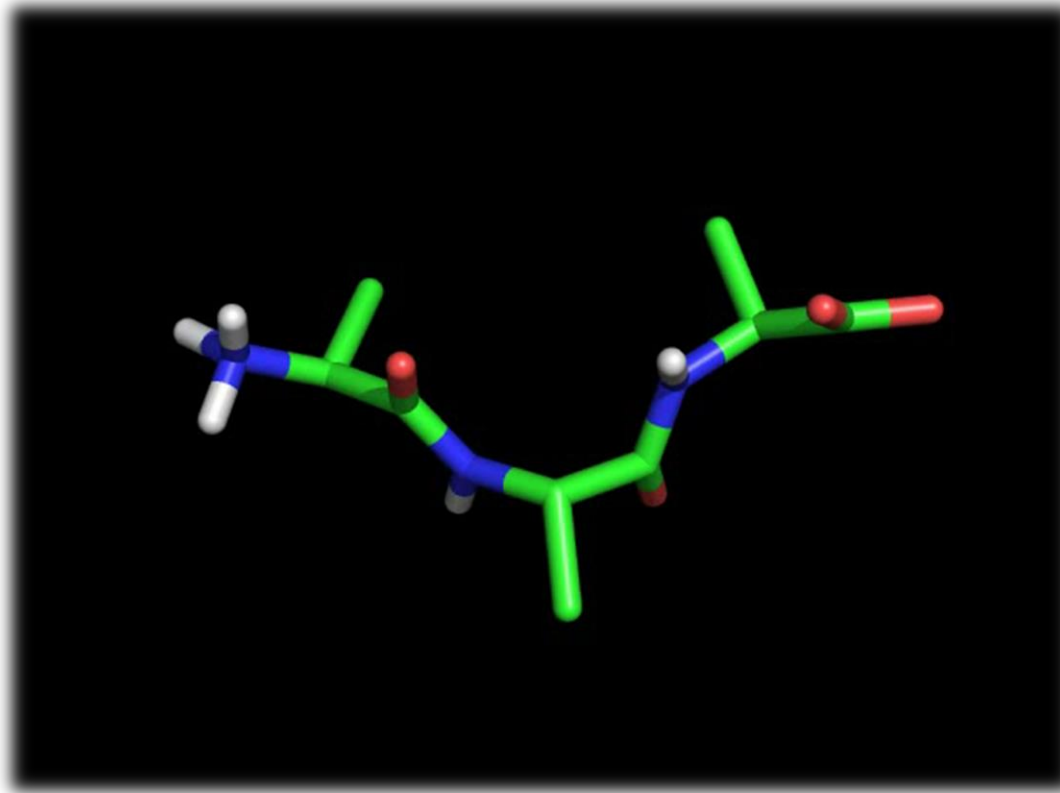


Neste caso a dinâmica molecular do tripéptido AAA é simulada em **vácuo**, e a molécula acaba por ficar “encravada” numa conformação em que as cargas terminais ($-\text{NH}_3^+$ e $-\text{COO}^-$) se encontram muito próximas

Ausência da blindagem do meio aquoso!

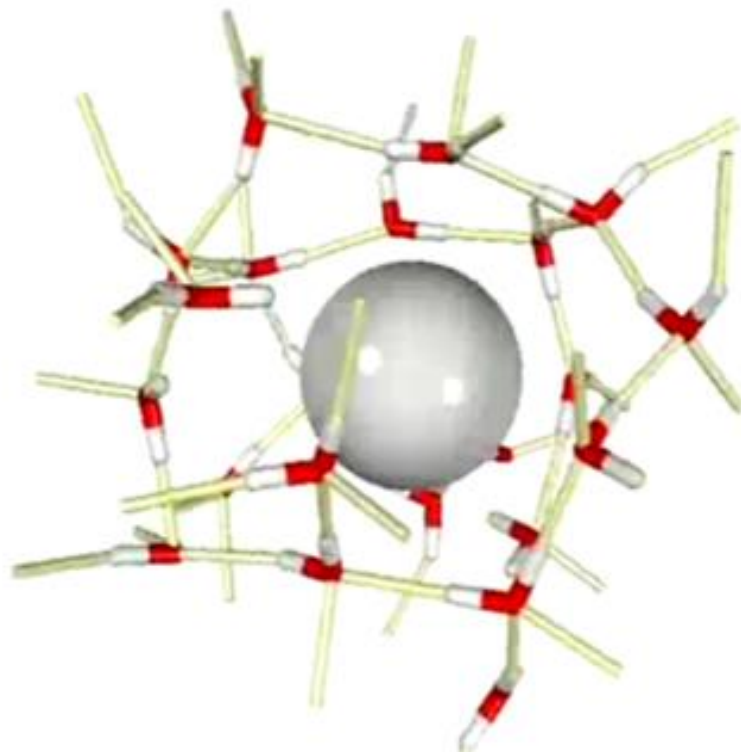
Importância da solvatação

Tripéptido Ala-Ala-Ala



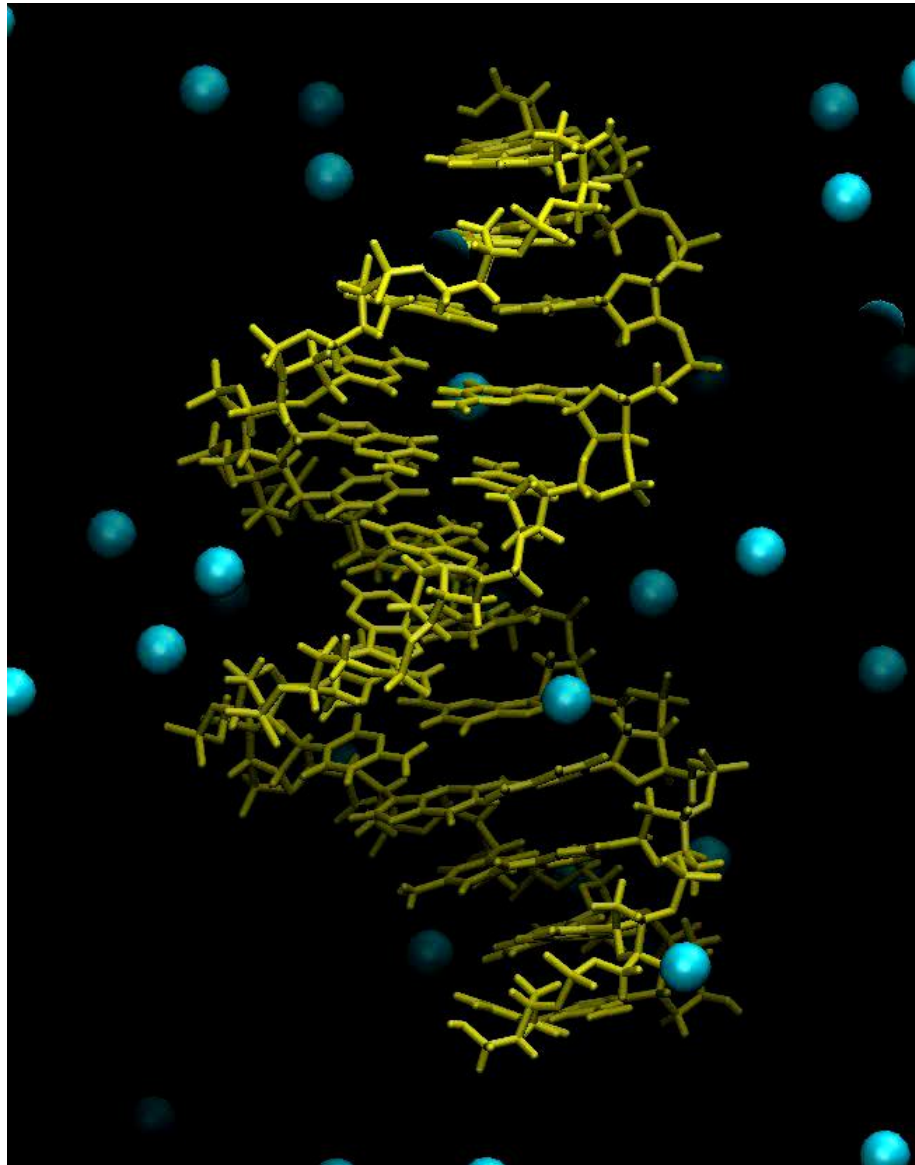
O tripéptido é simulado em solução aquosa (para facilitar a visualização as moléculas de água foram retiradas do vídeo). O efeito de blindagem do solvente atenua a atracção entre os grupos terminais

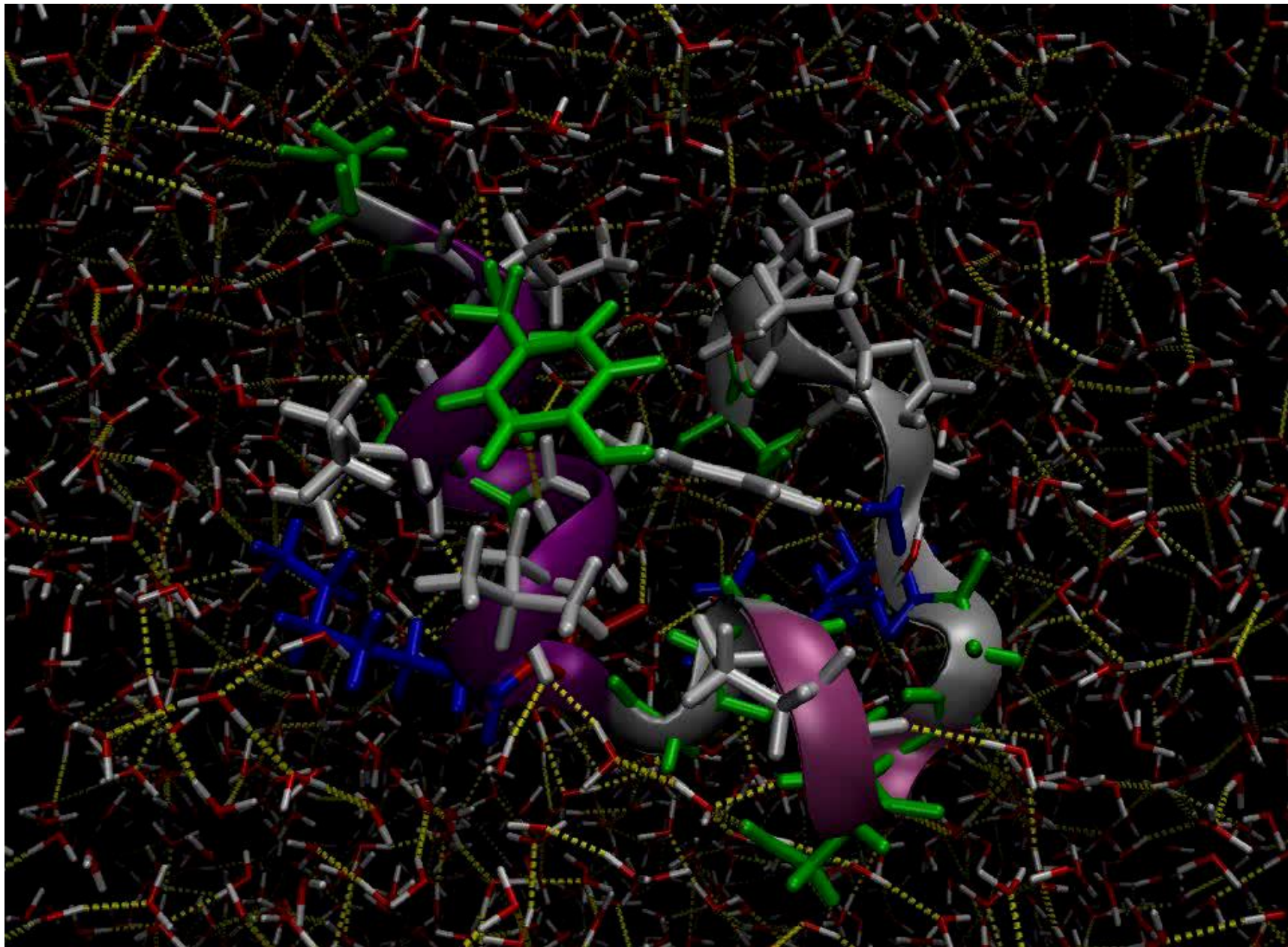
Interação das moléculas de água com um átomo de xénon



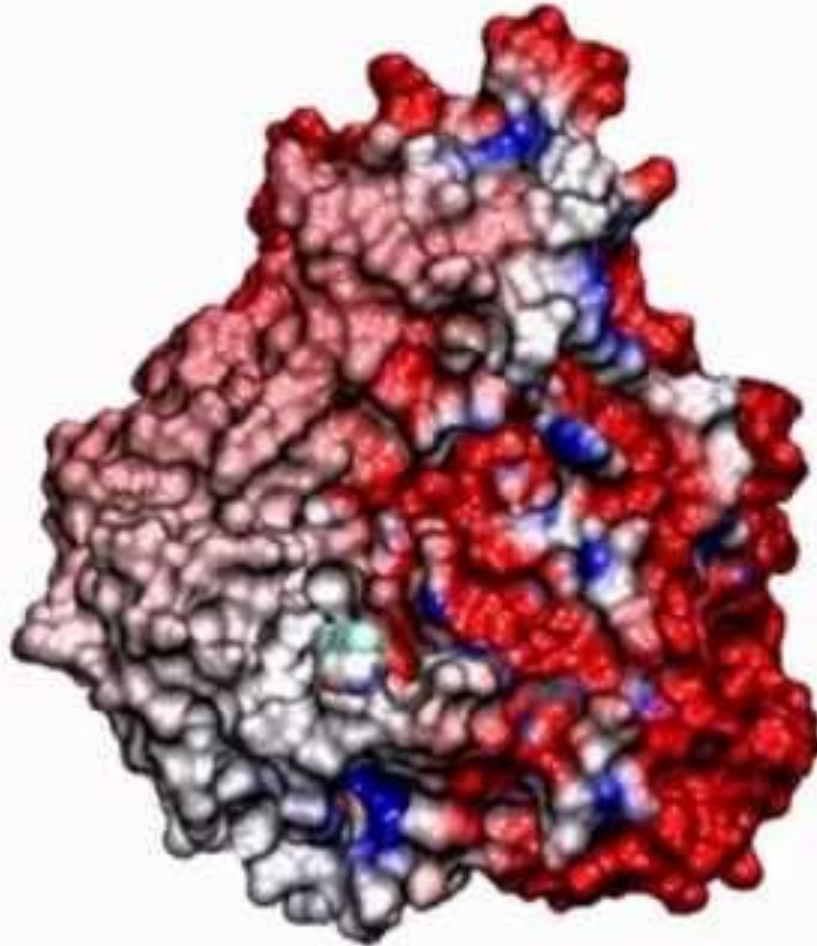
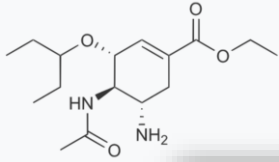
Nesta simulação de 5 ps da dinâmica da água em torno de um átomo de xénon, é possível observar o arranjo característico de ligações de hidrogénio da água sobre uma superfície hidrofóbica

Dinâmica molecular de um segmento de DNA + catiões





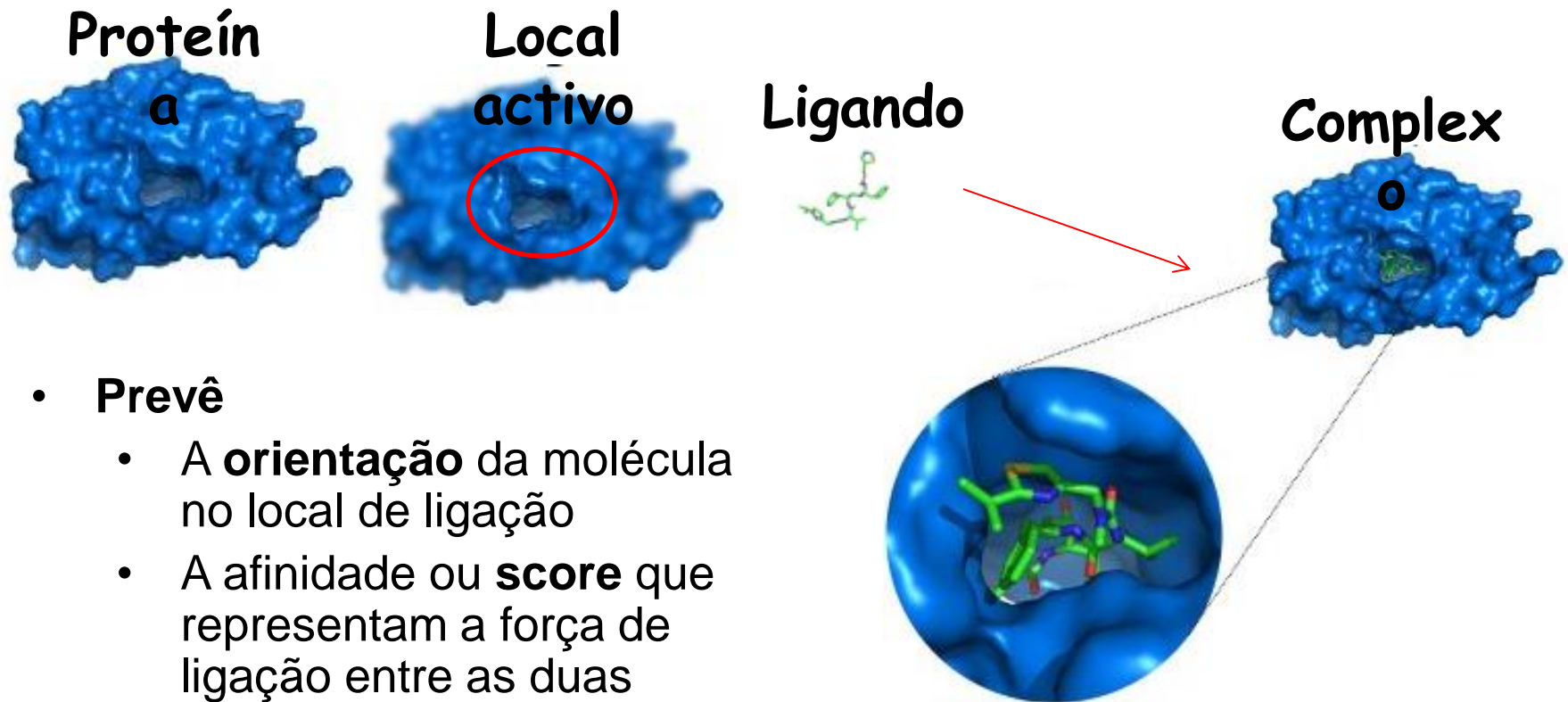
Interacção do fármaco Tamiflu com a proteínas viral Neuraminidase



Docking de ligandos a proteínas

“Docking” ligando-proteína

- Método baseado em estrutura (SBDD)
 - “estrutura” significa “usar a estrutura da proteína”
- Método computacional que mimetiza o processo físico de ligação
- **Dados os seguintes:**



- **Prevê**
 - A **orientação** da molécula no local de ligação
 - A afinidade ou **score** que representam a força de ligação entre as duas moléculas

Docking

- **Problema 1:** encontrar a conformação (ou conformações) de mais baixa energia para associação de duas moléculas
- **Problema 2:** calcular diferenças de energia livre entre ligandos de uma mesma proteína (semi-quantitativo)
- **Problema 3:** calcular energias livres *absolutas* de ligação de ligandos a proteínas (desenvolvimento de métodos)

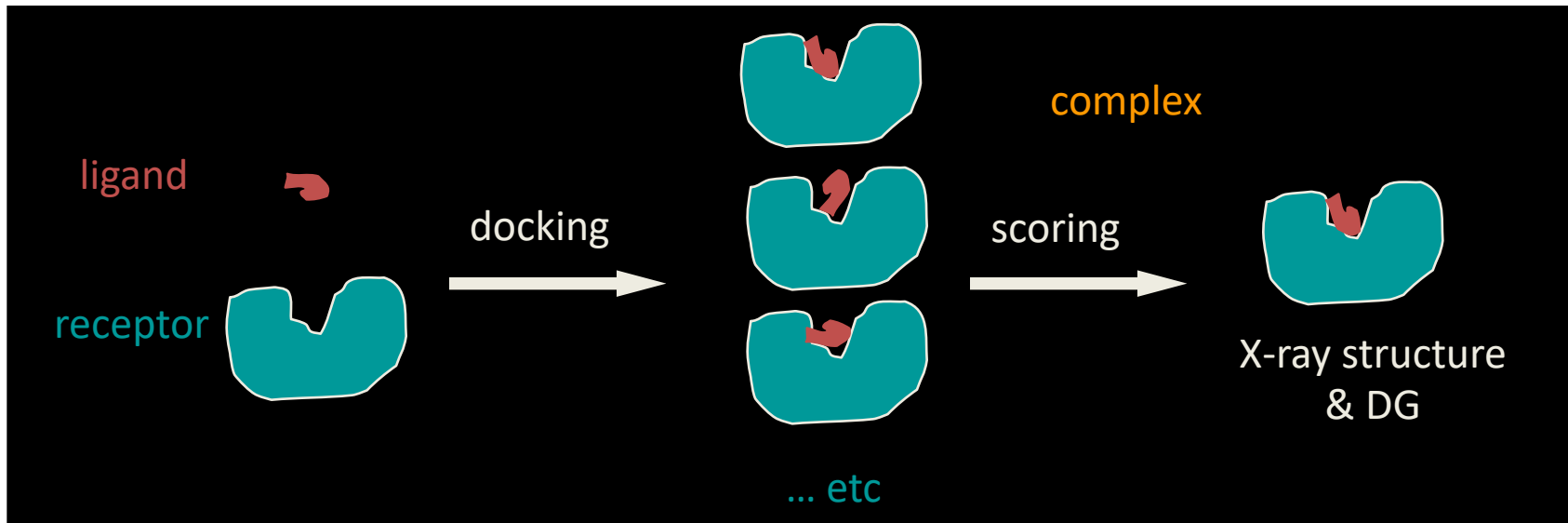
Ordem de dificuldade: **1,2,3** . Para o problema 1, o mais fácil, conseguem-se bons resultados em muitos casos. Para o problema 3 ainda estamos longe de ter uma solução rigorosa e geral.

Inhibitor/Drug	Disease/Condition	Enzyme Target
Acetazolamide	Glaucoma	Carbonic anhydrase
Acyclovir	Herpes	Viral DNA polymerase
Allopurinol	Gout	Xanthine oxidase
Argatroban	Coagulation	Thrombin
Aspirin, ibuprofen, DuP697	Inflammation, pain, fever	Prostaglandin synthase
β -Lactam antibiotics	Bacterial infections	D-Ala-D-Ala transpeptidase
Brequinar	Organ transplantation	Dihydroorotate dehydrogenase
Candoxatril	Hypertension, congestive heart failure	Atriopeptidase
Captopril	Hypertension	Angiotensin-converting enzyme
Clavulanate	Bacterial resistance	β -Lactamase
Cyclosporin	Organ transplantation	Cyclophilin/calcineurin
DuP450	AIDS	HIV protease
Enoximone	Congestive heart failure ischemia	cAMP phosphodiesterase
Finasteride	Benign prostate hyperplasia	Testosterone-5- α -reductase
FK-506	Organ transplantation, autoimmune disease	FK-506 binding protein
Fluorouracyl	Cancer	Thymidilate synthase
3-Fluorovinylglycine	Bacterial infection	Alanine racemase
(2-Furyl)-acryloyl-Gly- Phe-Phe	Lung elastin degradation in cystic fibrosis	<i>Pseudomonas</i> elastase
ICI-200,808	Emphysema	Neutrophil elastase
Lovastatin	High cholesterol	HMG CoA reductase

Docking e scoring

A descoberta das conformações ótimas receptor-ligando envolve dois passos distintos:

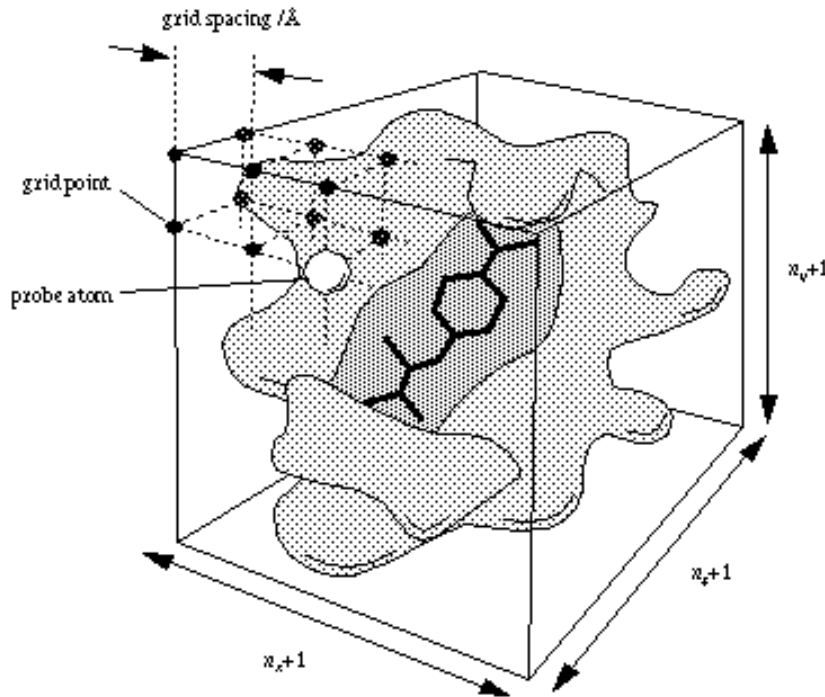
- Busca de conformações (docking): encontrar as soluções que produzem o score mais baixo. Este é um problema de minimização
- Avaliação das conformações (scoring): ordenar as conformações encontradas por ordem de energia decrescente, calculada pela função de scoring (uma espécie de campo de forças).



Importância do docking molecular

- Uma grande parte dos mecanismos bioquímicos envolve a interação específica de pequenos ligandos a moléculas de proteína (enzimas, receptores, etc.).
- A grande maioria dos fármacos existentes são inibidores enzimáticos ou interagem com receptores no organismo
- A indústria farmacêutica necessita de formas de descobrir ligandos específicos sem ter que passar pelo processo, moroso e caro, de síntese de todos os compostos em estudo
- Processos de interação proteína-proteína

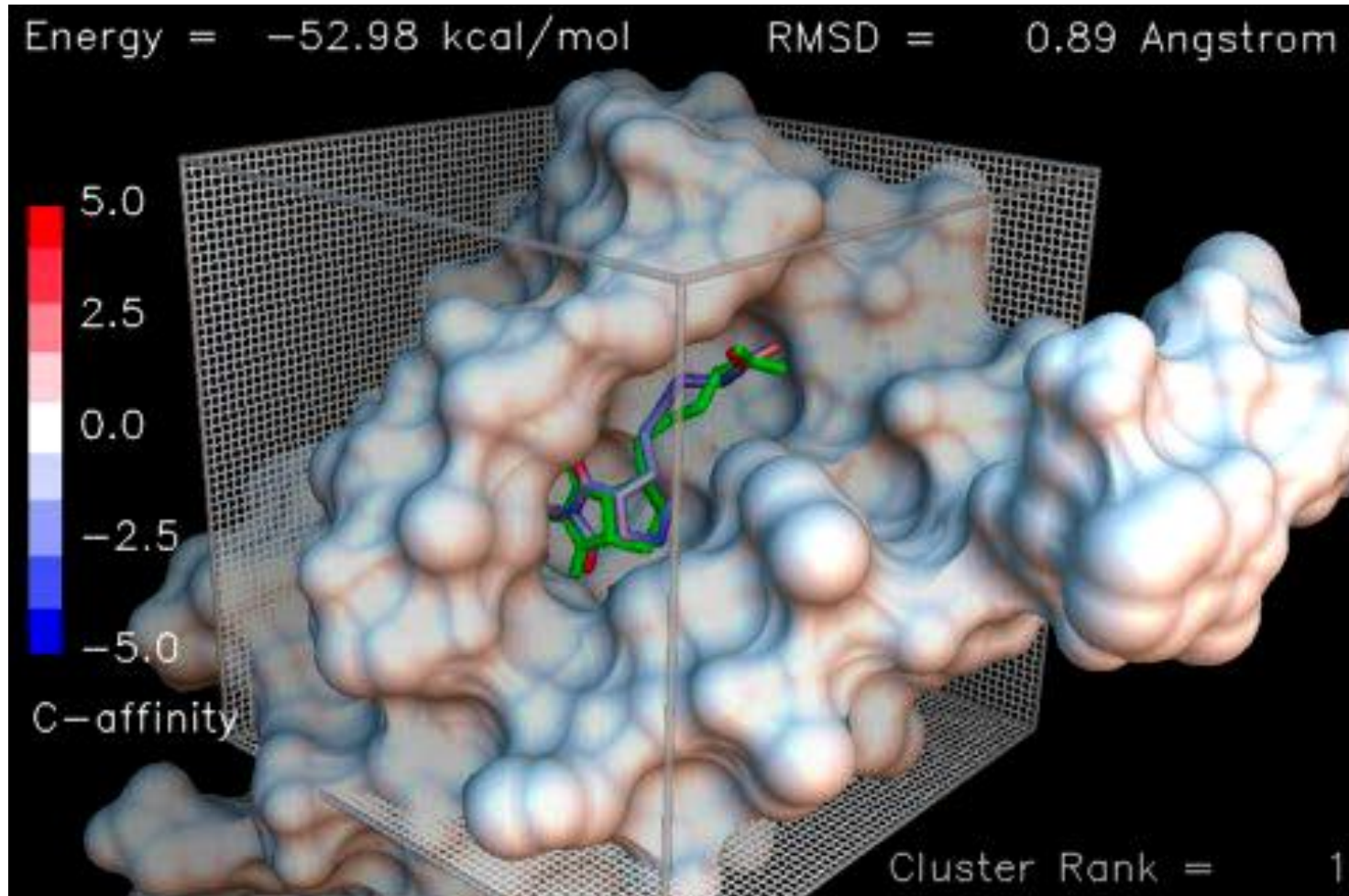
Avaliação da energia de interacção numa grelha



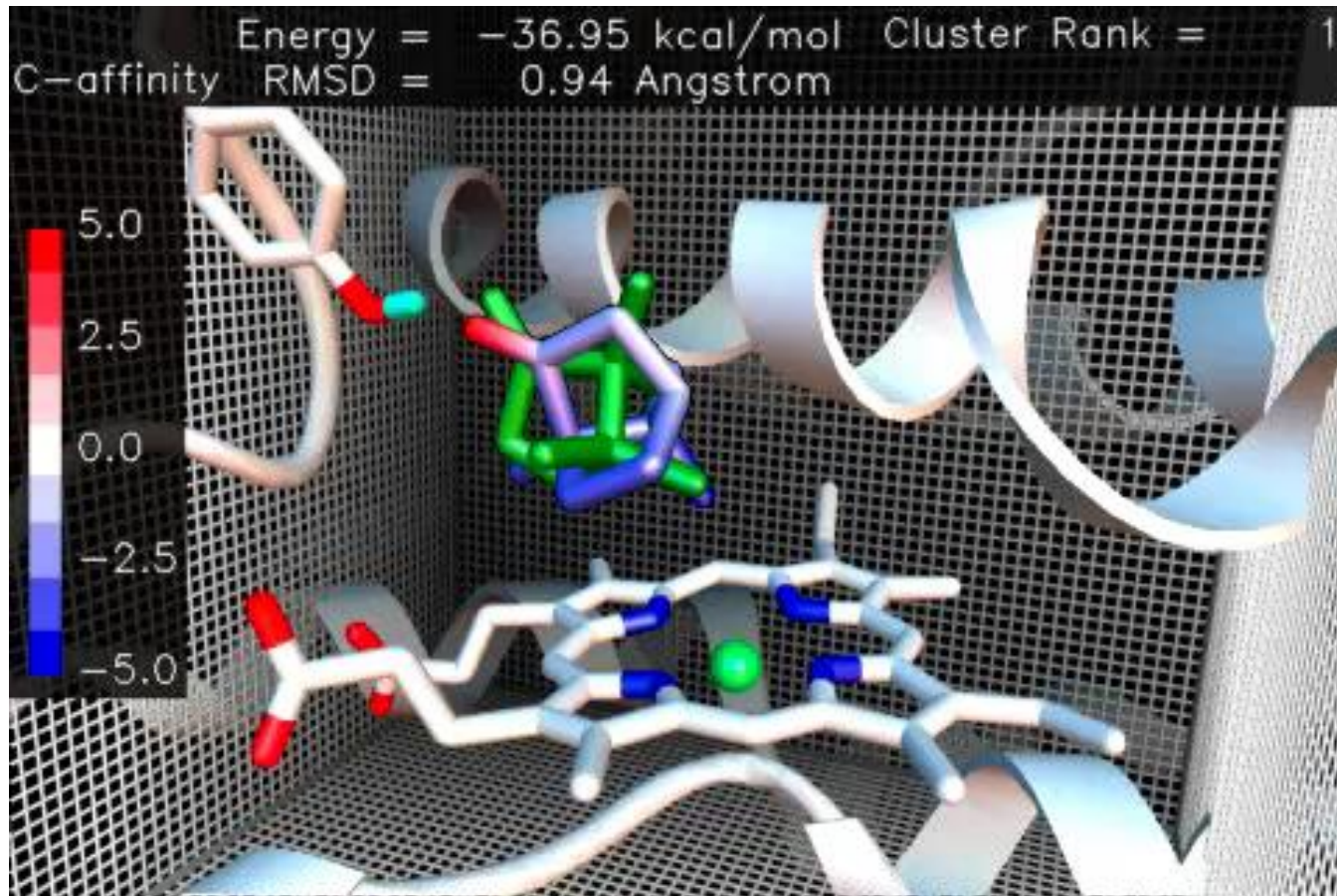
- A molécula de receptor é considerada rígida, apenas o ligando é flexível
- Quanto mais fina for a grelha, mais detalhado será o cálculo (e mais lento)
- Tem que ser calculada uma grelha para cada tipo de átomo do ligando (tipicamente C, N, O, H,...)

- Pré-cálculo energias de interacção com todos os átomos do receptor de cada um dos tipos de átomos do ligando nos nodos de uma grelha
- Para calcular a energia de uma dada conformação do ligando faz-se simplesmente interpolação dos valores calculadas na grelha. Isto é muito mais rápido do que calcular directamente a energia da cada conformação

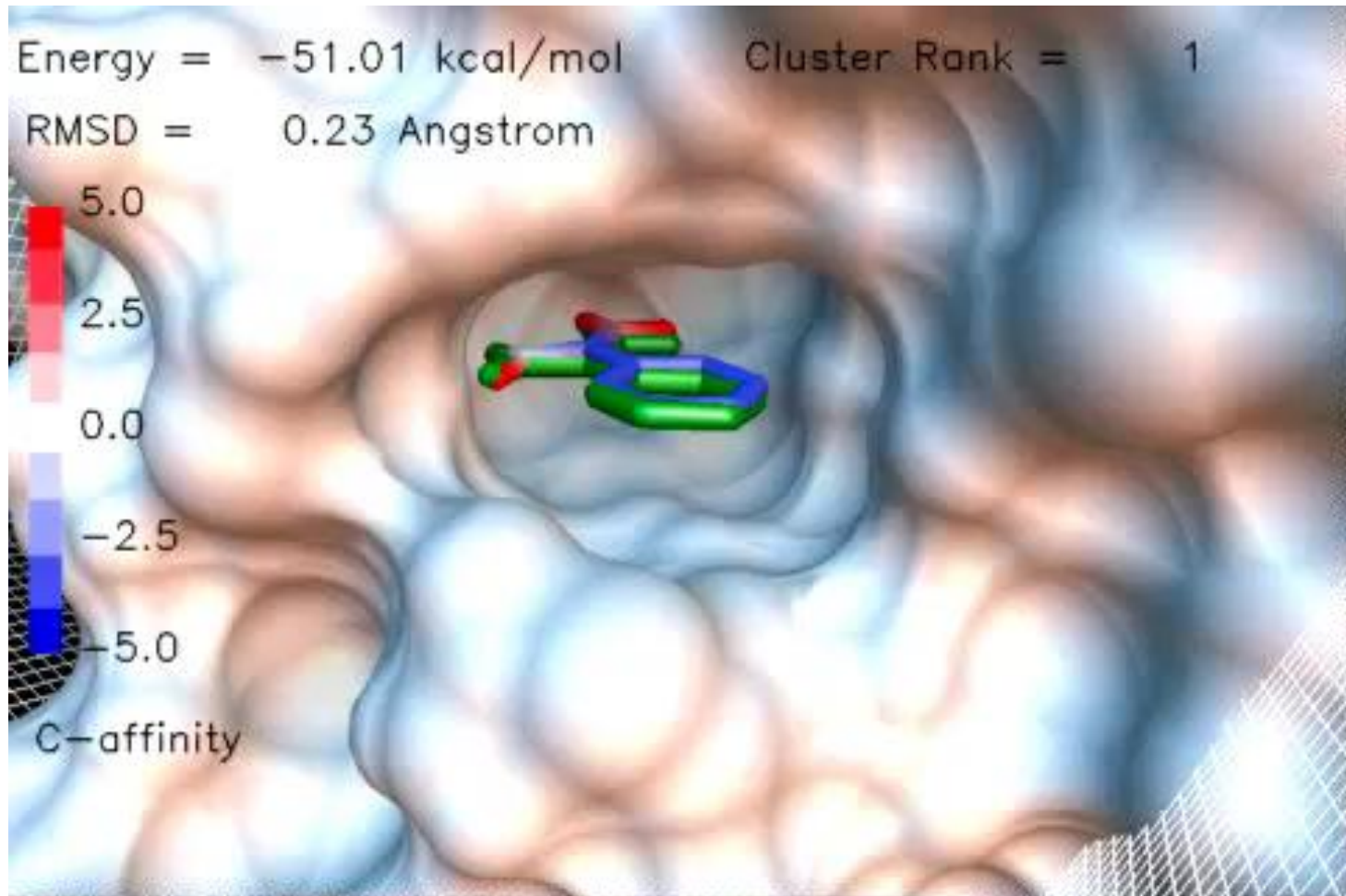
Ligação da biotina à streptavidina



Ligação da cânfora ao citocromo P450cam

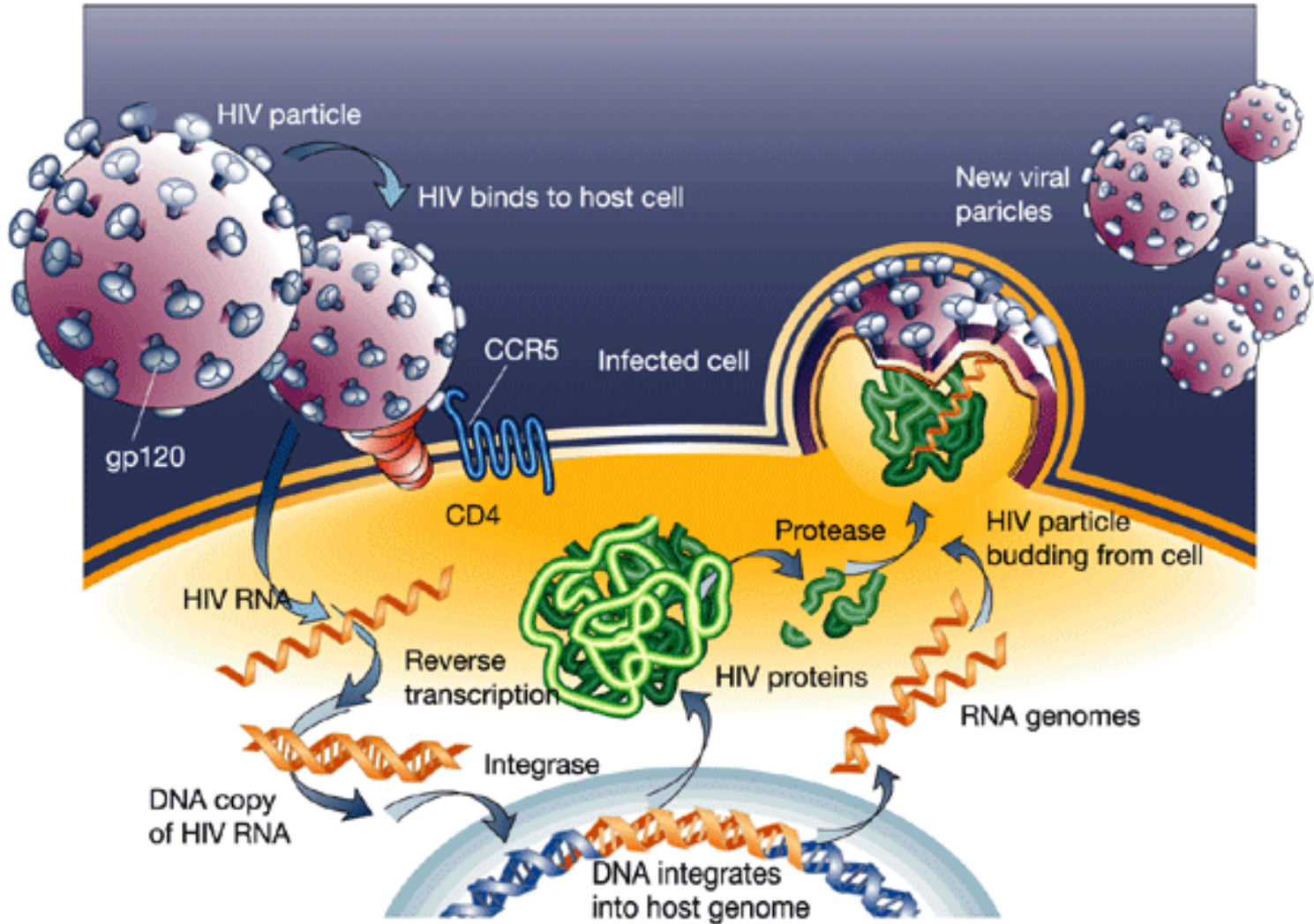


Ligação do inibidor benzamidina à tripsina

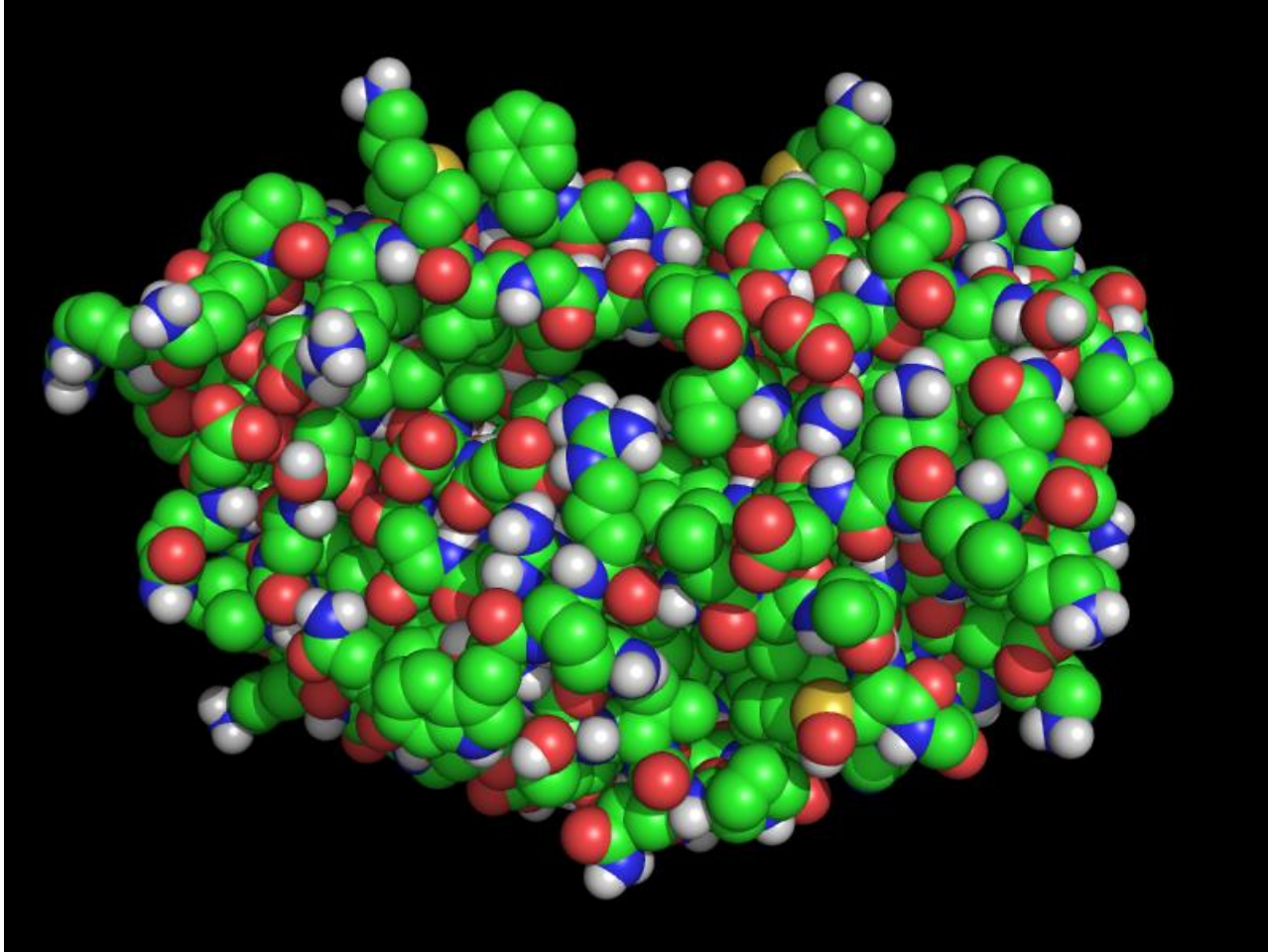


*"Docking" de um inibidor na
protease do HIV*

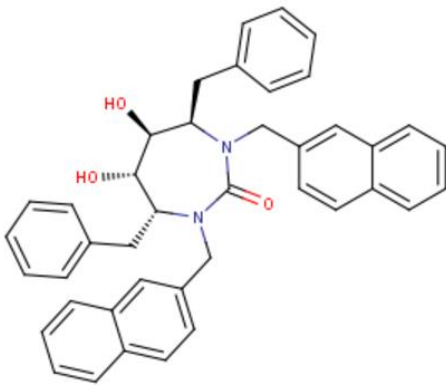
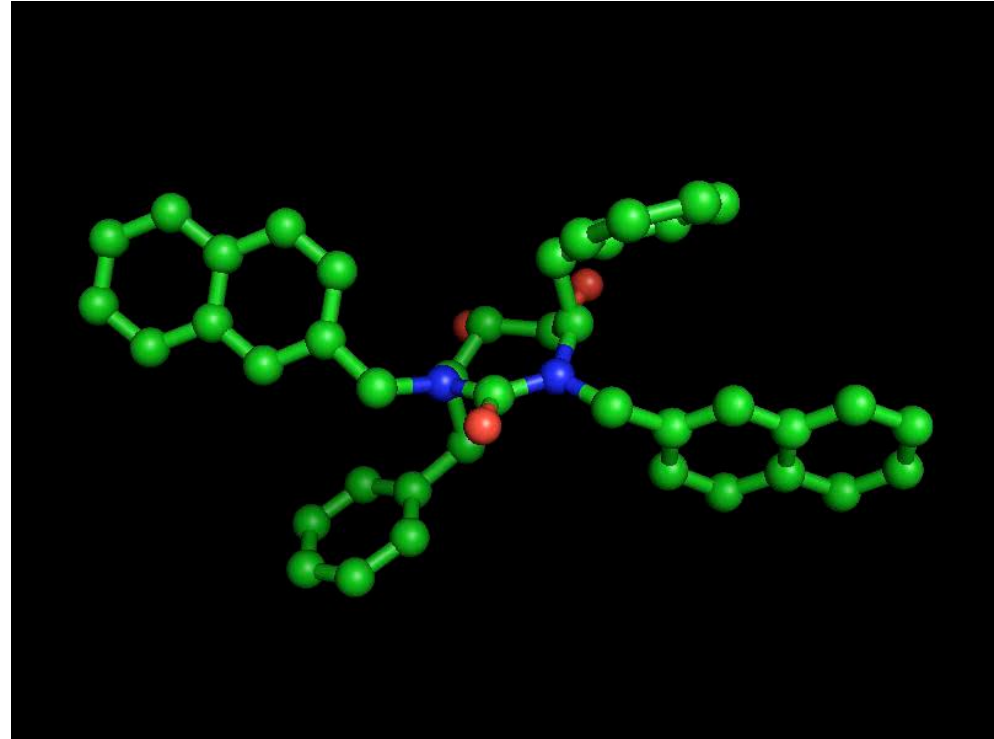
Ciclo de vida do HIV



Protease do HIV-1



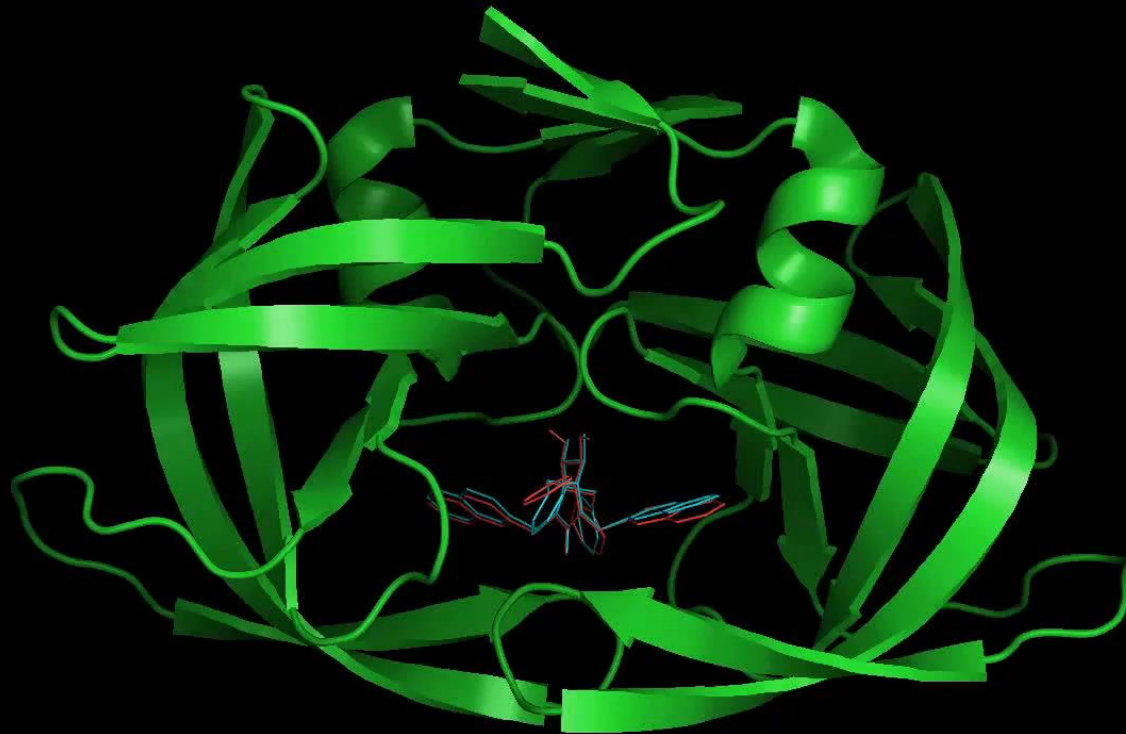
Candidato a inibidor

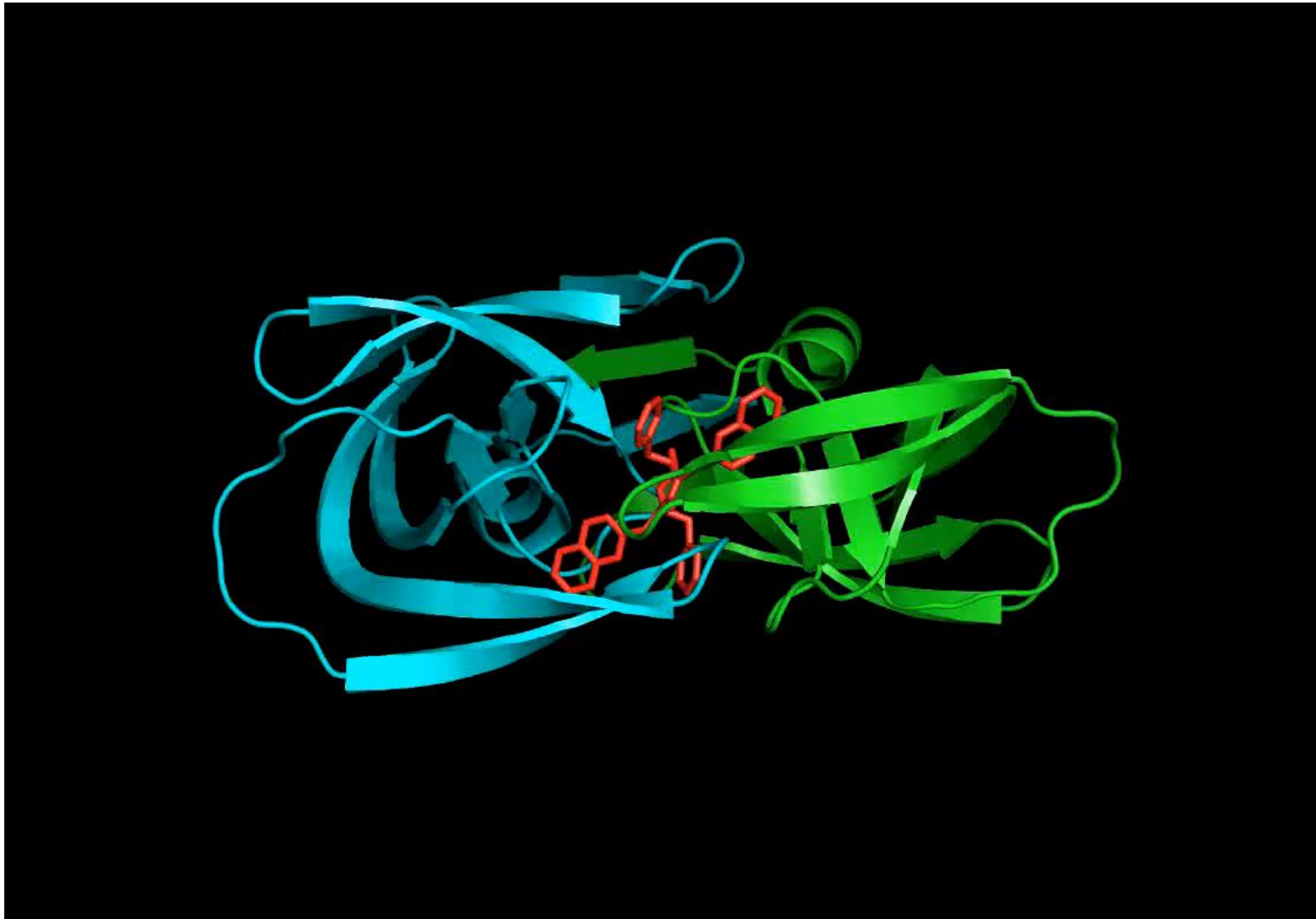


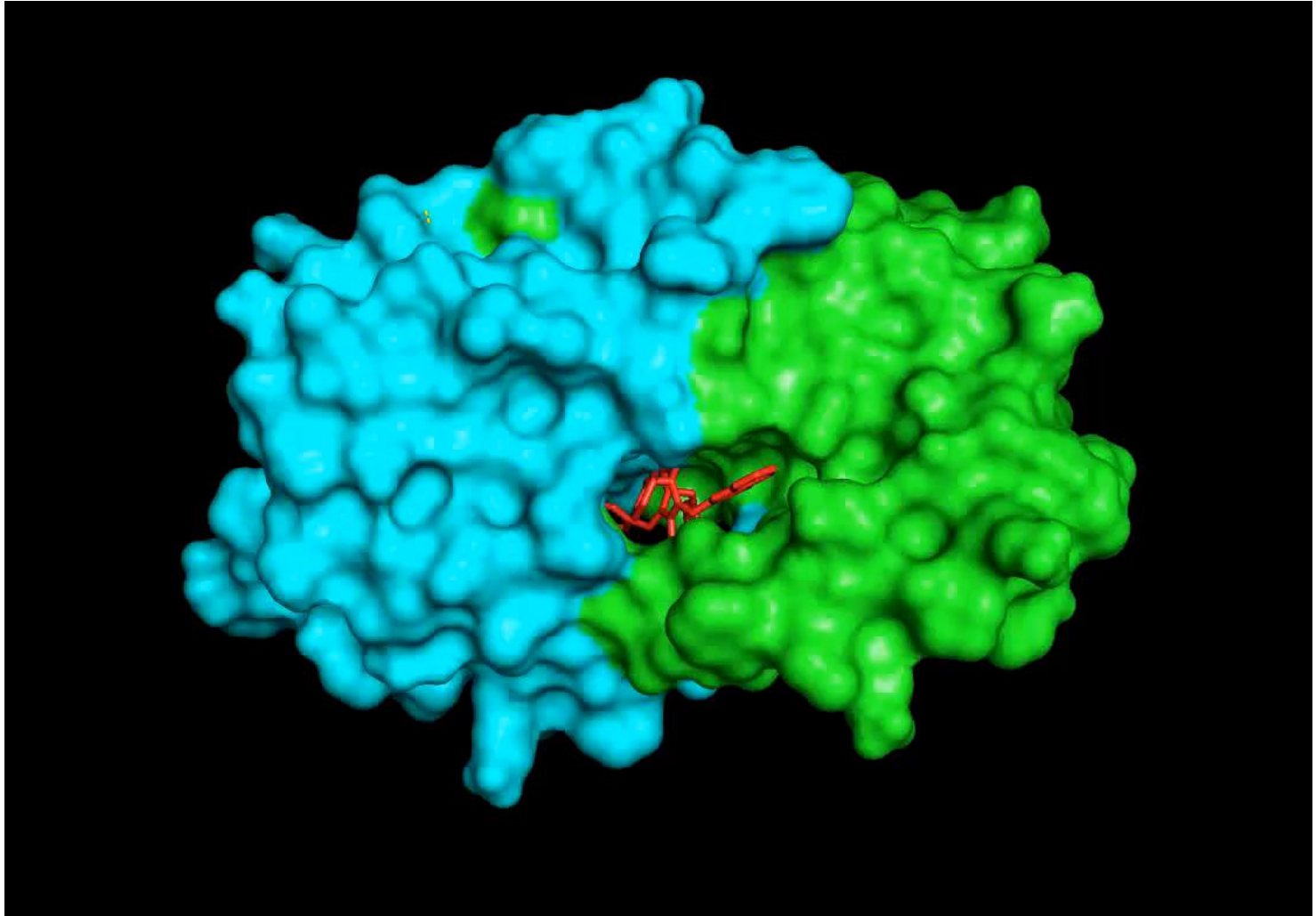
O xk263 foi um de muitos compostos modelizados, sintetizados e testados para a sua capacidade de inibir a protease do HIV.

Energias de diferentes soluções obtidas com o software Autodock Vina

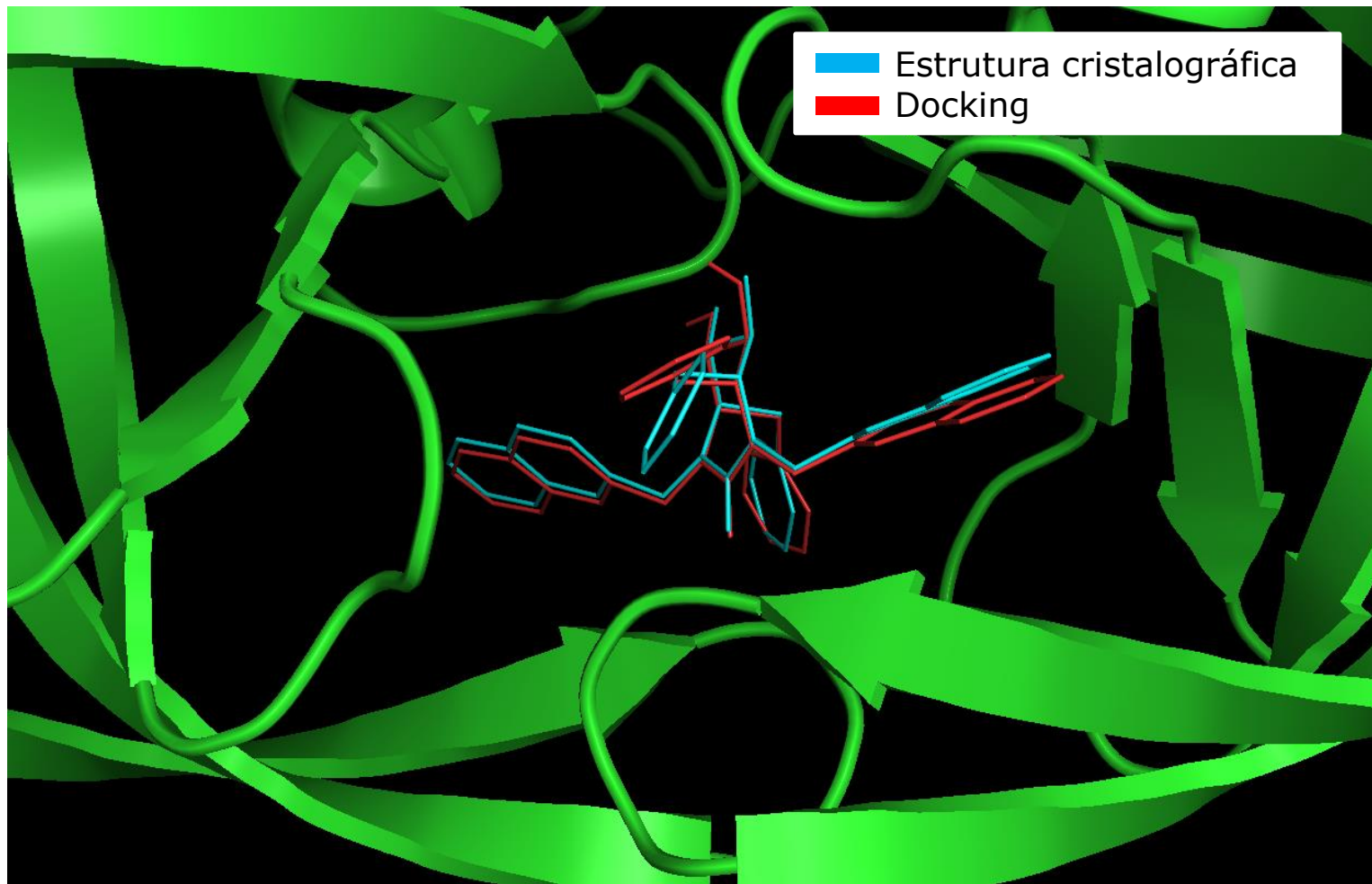
-14.6







Comparação entre a estrutura cristalográfica e docking com Autodock Vina



9. J. Cortes, S. F. Haydock, G. A. Roberts, D. J. Beville, P. F. Leadley, *Nature* **348**, 176 (1990); D. J. Beville, J. Cortes, S. F. Haydock, P. F. Leadley, *Eur. J. Biochem.* **204**, 39 (1992).
10. P. Caffrey, D. J. Beville, J. Staunton, P. F. Leadley, *FEBS Lett.* **304**, 225 (1992).
11. G. A. Roberts, J. Staunton, P. F. Leadley, *Eur. J. Biochem.* **214**, 305 (1993).
12. S. J. Wakil, *Biochemistry* **28**, 4523 (1989).
13. A. F. A. Marsden, unpublished data.
14. P. F. Leadley, *Biochem. J.* **197**, 413 (1981).
15. J. Q. Fuller and P. F. Leadley, *ibid.* **213**, 643 (1983).
16. J. F. Aparicio, P. Caffrey, A. F. A. Marsden, J. Staunton, P. F. Leadley, *J. Biol. Chem.*, in press.
17. S. Donadio, J. B. McAlpine, P. J. Sheldon, M. Jackson, L. Katz, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7119 (1993).
18. J. B. McAlpine *et al.*, *J. Antibiot.* **40**, 1115 (1987).
19. A. R. Fersht, *Enzyme Structure and Mechanism* (Freeman, New York, ed. 2, 1985), pp. 347-356.
20. D. E. Cane and C.-C. Yang, *J. Am. Chem. Soc.* **109**, 1255 (1987); S. Yue, J. C. Duncan, Y. Yamamoto, C. R. Hutchinson, *ibid.*, p. 1253.
21. B. Sadgrod *et al.*, *Eur. J. Biochem.* **75**, 481 (1977).
22. D. O'Hagan, *Nat. Prod. Rep.* **6**, 205 (1989).
23. Supported by grants from the Science and Engineering Research Council (United Kingdom) (Molecular Recognition Initiative and Biotechnology Directorate-LINK), S.F.A. was the recipient of a European Community bursary. We thank J. R. Knowles and D. A. Hopwood for their comments during preparation of the manuscript.

30 August 1993; accepted 30 November 1993

Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as HIV Protease Inhibitors

Patrick Y. S. Lam,* Prabhakar K. Jadhav, Charles J. Eyermann, C. Nicholas Hodge, Yu Ru, Lee T. Bacheler, James L. Meek, Michael J. Otto, Marlene M. Rayner, Y. Nancy Wong, Chong-Hwan Chang, Patricia C. Weber, David A. Jackson, Thomas R. Sharpe, Susan Erickson-Viitanen*

Mechanistic information and structure-based design methods have been used to design a series of nonpeptide cyclic ureas that are potent inhibitors of human immunodeficiency virus (HIV) protease and HIV replication. A fundamental feature of these inhibitors is the cyclic urea carbonyl oxygen that mimics the hydrogen-bonding features of a key structural water molecule. The success of the design in both displacing and mimicking the structural water molecule was confirmed by x-ray crystallographic studies. Highly selective, preorganized inhibitors with relatively low molecular weight and high oral bioavailability were synthesized.

Knowledge of the HIV protease (HIV PR) mechanism of action and substrate specificity has been extensively used to design a variety of transition state-based inhibitors with inhibition constants in the nanomolar or subnanomolar range (1, 2). The symmetry of the HIV PR dimer guided the design of twofold (C₂) symmetric and pseudosymmetric inhibitors (3). However, these inhibitors retain substantial peptide character, and despite many elegant structure-activity studies, it has been difficult to combine adequate potency with oral bioavailability (3, 4). The difficulty in developing such leads into useful therapeutics is challenging, for in addition to the traditional barriers encountered in the drug development process, peptide-based molecules are in general biologically unstable, poorly absorbed, and rapidly metabolized (5). This challenge is not unique to HIV PR; transition from peptide-based leads to therapeutics has proven formidable for oth-

er enzymes such as renin inhibitors (6, 7). We have previously explored a series of potent, linear C₂-symmetric inhibitors in which the transition state mimetic was a diol (8). We were unable to overcome the poor oral bioavailability of these peptide molecules and consequently sought other approaches. The technique of searching databases containing three-dimensional (3D) molecular structures has been used to identify synthetic frameworks that can serve as the starting point for the design of nonpeptide inhibitors, and this approach has been explored with HIV PR. Unfortunately, the HIV PR inhibitors designed to date on the basis of 3D database searches (9, 10) have yielded inhibitors with only micromolar potency.

Our current design of nonpeptide inhibitors (11) began with structural information available from published x-ray crystal structures of HIV-PR inhibitor complexes (12-15). A common feature observed is the presence of a tetracoordinated structural water molecule linking the bound inhibitor to the flexible glycine-rich β strands or "flaps" of the HIV PR dimer (Fig. 1). This water molecule accepts two hydrogen bonds

from backbone amide hydrogens of HIV PR residues Ile 50 and Ile 50' and donates two hydrogen bonds to carbonyl oxygens of the inhibitor, thus inducing the fit of the flaps over the inhibitor (16). Its relevance to the generation of HIV PR inhibitors has been noted (12, 13).

We hypothesized that incorporation of the binding features of this structural water molecule into an inhibitor would be beneficial because its displacement should be energetically favorable (17). In addition, conversion of a flexible, linear inhibitor into a rigid, cyclic structure with restricted conformations should provide a positive entropic effect. Finally, incorporation of a mimic for the structural water within the inhibitor should ensure specificity for the HIV PR as against other aspartic acid proteases, because this water molecule is unique to retroviral proteases. We reasoned that these effects might provide highly potent and specific binding and reduce the need for multiple interactions at the specificity pockets. This should permit design of smaller (<600 daltons) inhibitors with improved oral bioavailability.

Extensive structure-activity relations (SARs) established for C₂-symmetric diols indicated that the diol imparts significant potency as compared with corresponding mono-ol transition state analogs (3, 8). Thus, we wanted to incorporate this feature into the diol-HIV PR interaction into a 3D pharmacophore model. However, no x-ray structure of a C₂-symmetric diol-protease complex was available when this work was initiated; two independent reports of structures have since appeared (18). Therefore, computer models for C₂-symmetric diols bound to the active site of HIV-1 PR were developed from the crystal structure coordinates of a hydroxyethylene inhibitor bound to HIV PR (15) by means of distance geometry (19) (Fig. 2, A and B) and several pharmacophores were generated.

The simplest pharmacophore model (Fig. 2C) was based on two key intramolecular distances: that between symmetric hydrophobic groups, designated P1 and P1', that occupy corresponding enzyme pockets S1 and S1' and that from P1 and P1' to a hydrogen bond donor/acceptor group (or groups) that binds to catalytic aspartates. A 3D database search (20, 21) with this pharmacophore model yielded the "hit" (22) shown in Fig. 2D, which not only met the initial search criteria, but also included an oxygen that matched the structural water found in HIV PR-inhibitor complexes. This 3D search indicated that a phenyl ring could properly position groups to interact with aspartates 25 and 25' as well as to mimic the structural water (Fig. 2E). However, because a phenyl ring might not properly position all substituents in the

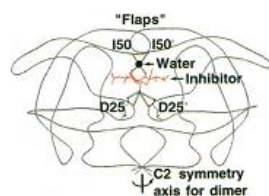


Fig. 1. Key features of the HIV PR dimer. HIV PR is a C₂-symmetric dimer, with paired aspartic acid residues located at the floor of the active site and a water molecule juxtapositioned between the inhibitor and flexible enzyme flaps. The complex shown is that of Swain *et al.* (15).

inhibitor, a cyclohexanone ring (Fig. 2F) was chosen as the initial synthetic scaffold, with the ketonic oxygen as the structural water mimic. The cyclohexanone ring was enlarged to a seven-membered ring (Fig. 2G) to incorporate the diol functionality. This synthetic target was further modified to a cyclic urea (Fig. 2H) on the basis of two considerations. First, cyclic ureas have established a precedent as excellent hydrogen-bond acceptors both in nature (for example, biotin-streptavidin interactions (23)) and in synthetic systems (24). Second, it was realized that the seven-membered cyclic urea was synthetically accessible by cyclization of the precursor (a phenylalanine-derived diaminodiol) used in the linear C₂-symmetric diol series (8).

Additional modeling studies were performed with the cyclic ureas to predict the optimal stereochemistry and conformation needed for complementary interaction with the HIV PR. The predicted optimal stereochemistry for cyclic ureas with substituents on the nitrogens is 4R, 5S, 6S, 7R (Fig. 2I), which is derived from unnatural (D) phenylalanine. This is in contrast to the linear C₂-symmetric diol inhibitors where natural (L) phenylalanine at P1/P1' provides optimal stereochemistry (25). The optimal stereoisomer prefers the conformation with pseudo diaxial benzyl groups and pseudo diequatorial hydroxyl groups (26). On the other hand, if the nitrogens of the cyclic ureas are not substituted, enantiomers derived from L- or D-phenylalanine fit equally well at the active site. These conformational preferences were subsequently confirmed by small molecule x-ray crystallography (27). Thus, the N-substituted cyclic urea ring provides a rigid framework such that with the proper stereochemistry and conformation, the P1/P1'/P2/P2' and hydroxyls should be highly complementary to the corresponding S1/S1'/S2/S2' pockets and catalytic aspartates of HIV PR.

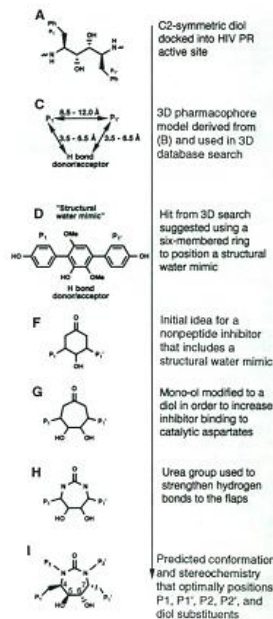


Fig. 2. Strategy and steps involved in the design of cyclic urea inhibitors of HIV PR.

REPORTS

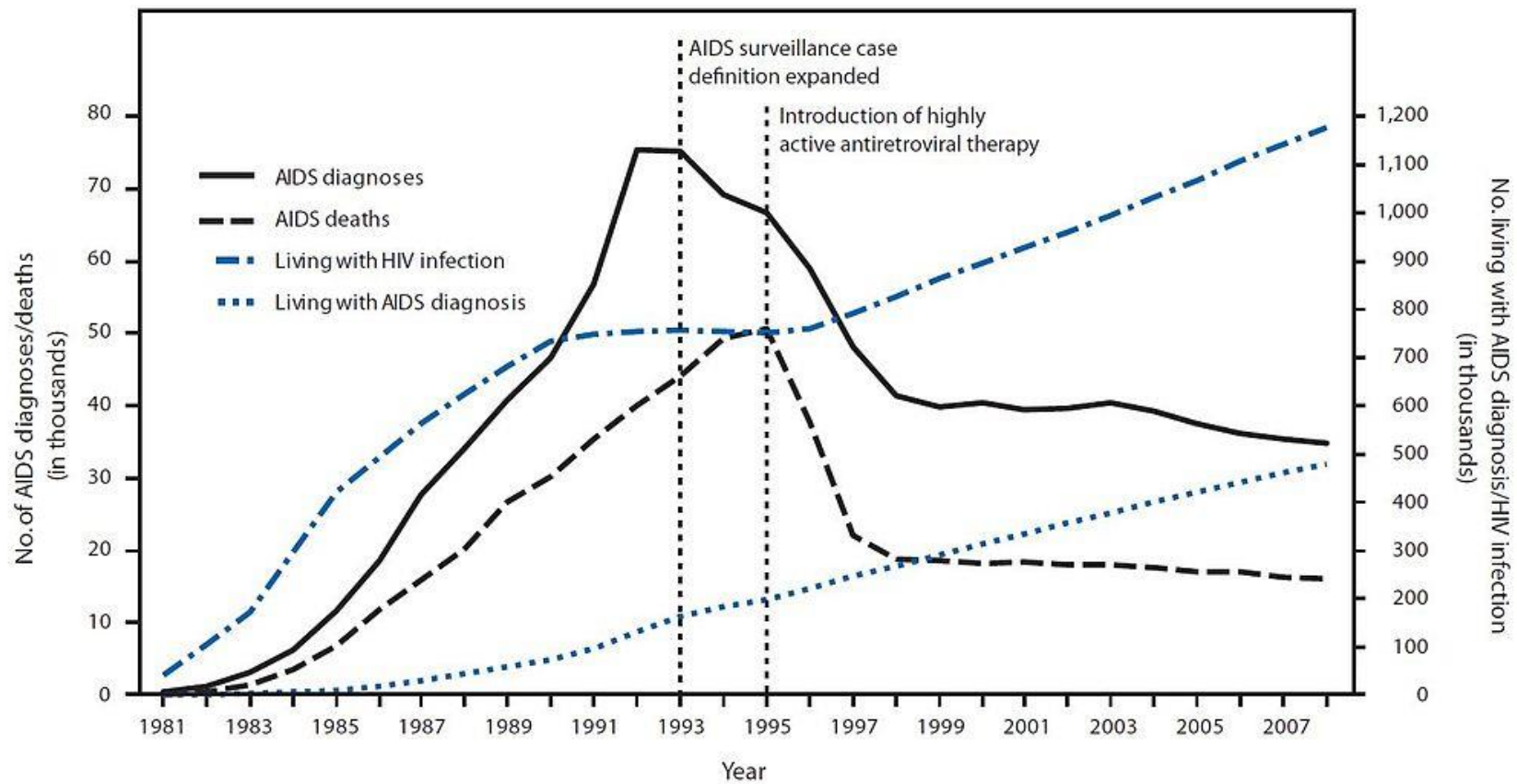
In contrast, substitution of D-phenylalanine-derived cyclic urea with alllyl groups on the nitrogens resulted in cyclic urea 1 (Table 1, K_i = 4.7 nM), which is 1000-fold more potent than the corresponding L-phenylalanine-derived cyclic urea.

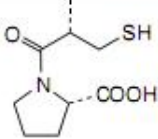
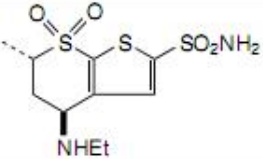
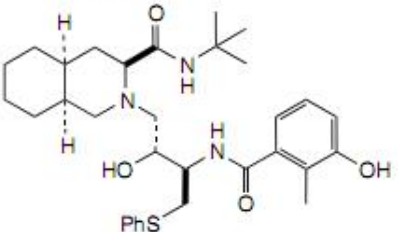
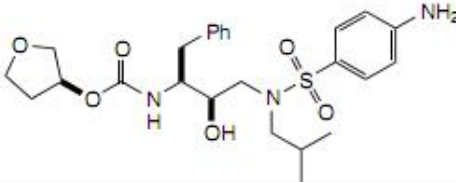
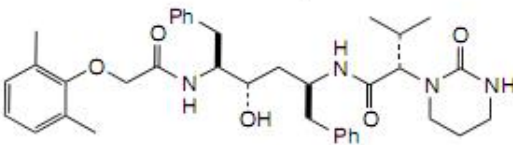
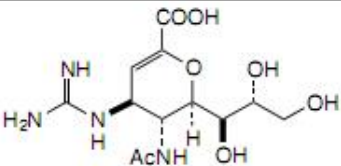
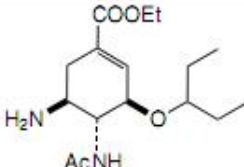
There are several factors responsible for the potency of nitrogen-substituted cyclic ureas: (i) the cyclic ureas are preorganized (28) for highly complementary binding to HIV PR, with the conformational entropic penalties typically associated with binding a linear, flexible inhibitor being "prepaid" during synthesis rather than during binding; (ii) displacement of the water molecule is probably thermodynamically favorable (17); and (iii) hydrophobic interactions between the nitrogen substituents and the S2 and S2' subsites of HIV PR are optimized with the preferred conformation and stereochemistry.

As part of our discovery program on HIV PR inhibitors, criteria for development of a compound as an oral agent for the treat-






Departments of Virology Research and Chemical and Physical Sciences, The DuPont Merck Pharmaceutical Company, Wilmington, DE 19880.

*To whom correspondence should be addressed.



Trade name ³	Generic name	Structure	Mechanism	Disease area	Date launched	Company
Capoten	Captopril		Angiotensin-converting enzyme	Hypertension	1981	Bristol-Myers Squibb
Trusopt	Dorzolamide		Carbonic anhydrase	Glaucoma	1995	Merck
Viracept	Nelfinavir		HIV protease	HIV/AIDS	1999	Agouron (Pfizer) and Lilly
Agenerase	Amprenavir		HIV protease	HIV/AIDS	1999	Vertex and GSK
Aluviran	Lopinavir		HIV protease	HIV/AIDS	2000	Abbott
Relenza	Zanamivir		Neuraminidase	Influenza	1999	Monash University and GlaxoSmithKline
Tamiflu	Oseltamivir		Neuraminidase	Influenza	1999	Gilead and Roche

Antiretrovirais inibidores de protease

Name	Trade name	Company	Patent	Notes
Saquinavir	Fortovase, Invirase	Hoffmann–La Roche	U.S. Patent 5,196,438 ↗ 	It was the first protease inhibitor approved by the FDA (December 6, 1995).
Ritonavir	Norvir	Abbott Laboratories	U.S. Patent 5,541,206 ↗ 	-
Indinavir	Crixivan	Merck & Co.	U.S. Patent 5,413,999 ↗ 	-
Nelfinavir	Viracept	Japan Tobacco	U.S. Patent 5,484,926 ↗ 	-
Amprenavir	Agenerase	GlaxoSmithKline	U.S. Patent 5,585,397 ↗ 	The FDA approved it April 15, 1999, making it the sixteenth FDA-approved antiretroviral. It was the first protease inhibitor approved for twice-a-day dosing instead of needing to be taken every eight hours. The convenient dosing came at a price, as the dose required is 1,200mg, delivered in eight very large gel capsules. Production was discontinued by the manufacturer December 31, 2004, as it has been superseded by fosamprenavir.
Lopinavir	Kaletra	Abbott	-	Is only marketed as a combination, with ritonavir.
Atazanavir	Reyataz	-	-	-
Fosamprenavir	Lexiva	GlaxoSmithKline	-	Is a pro-drug of amprenavir. The FDA approved it October 20, 2003. The human body metabolizes fosamprenavir in order to form amprenavir, which is the active ingredient. That metabolism increases the duration that amprenavir is available, making fosamprenavir a slow-release version of amprenavir and thus reduces the number of pills required versus standard amprenavir.
Tipranavir	Aptivus	Boehringer-Ingelheim	-	Also known as tipranavir disodium
Darunavir	Prezista	Tibotec	-	It was approved by the Food and Drug Administration (FDA) on June 23, 2006. Several ongoing phase III trials are showing a high efficiency for the PREZISTA/rtv combination being superior to the lopinavir/rtv combination for first-line therapy ^[1] . Darunavir is the first drug in a long time that didn't come with a price increase. It leapfrogged two other approved drugs of its type, and is matching the price of a third ^{[2][3][4]} .