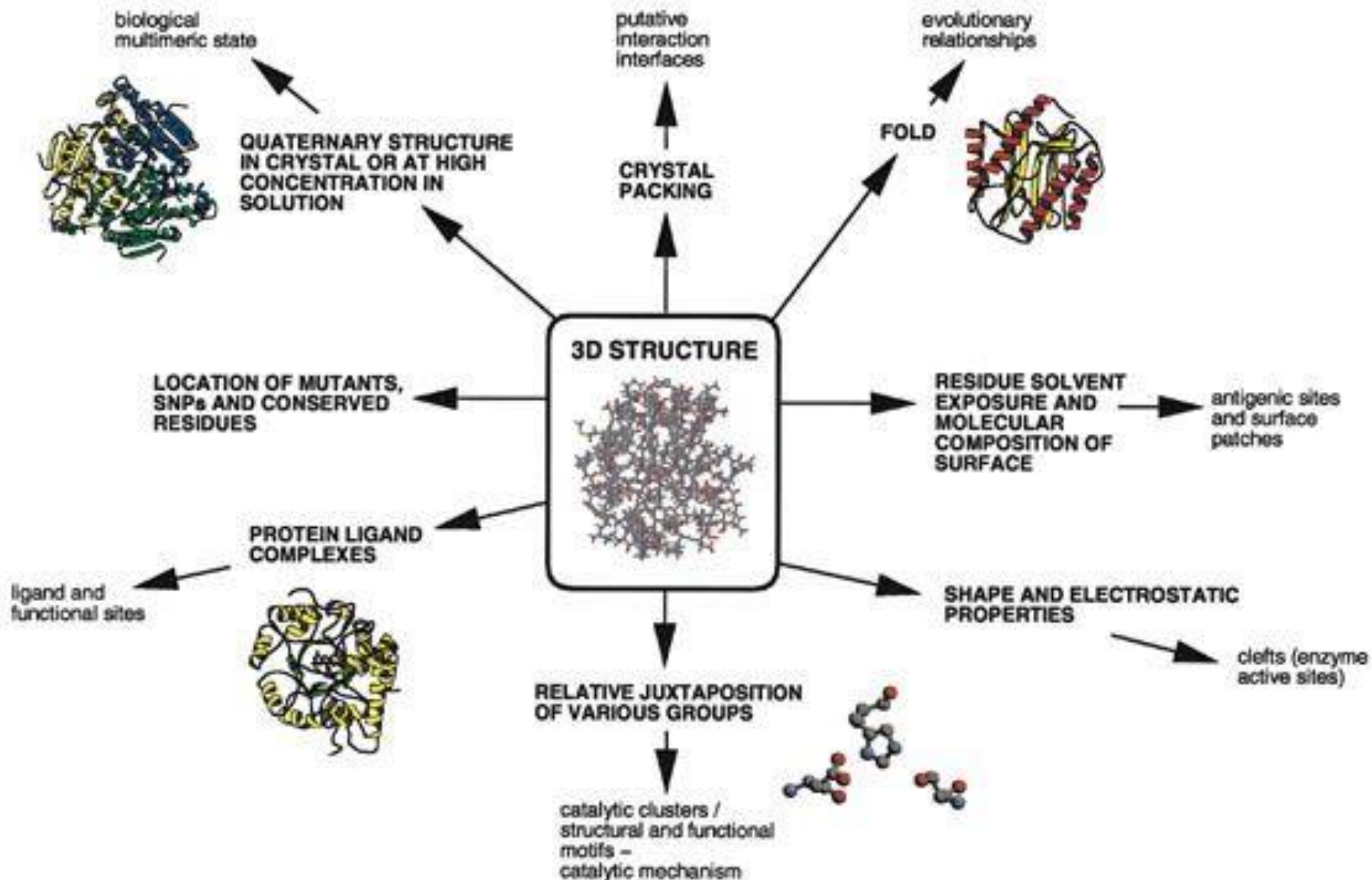


Bioinformática estrutural

Lucianna Santos
luciannahss@gmail.com

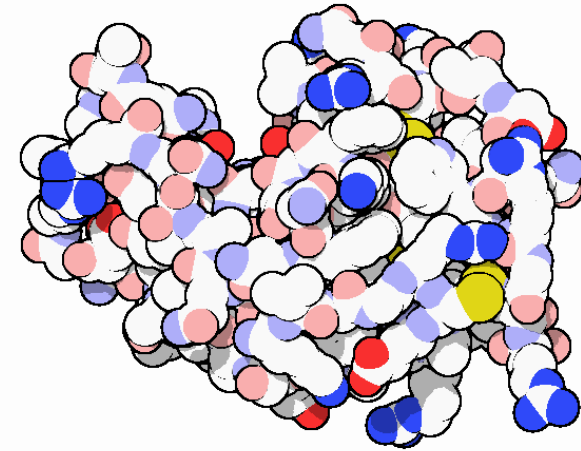
Quais informações retiramos de estruturas



Vamos focar em ...

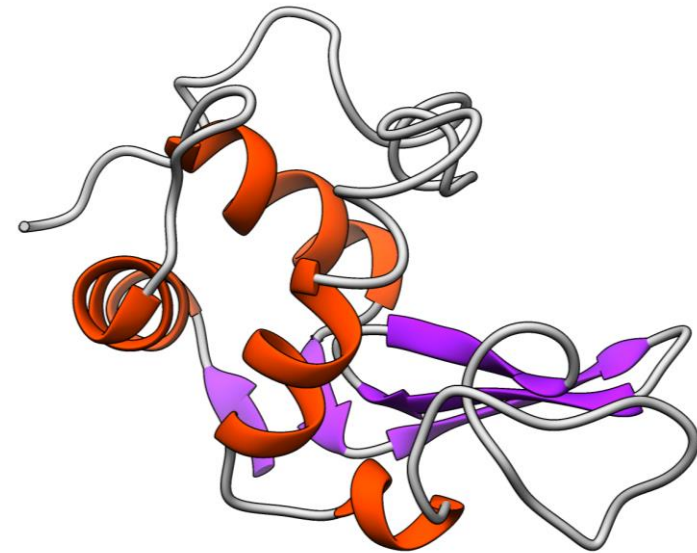
Proteínas

- Proteínas são as **macromoléculas biológicas** mais abundantes e versáteis dos seres vivos, sendo essenciais em inúmeros processos biológicos.



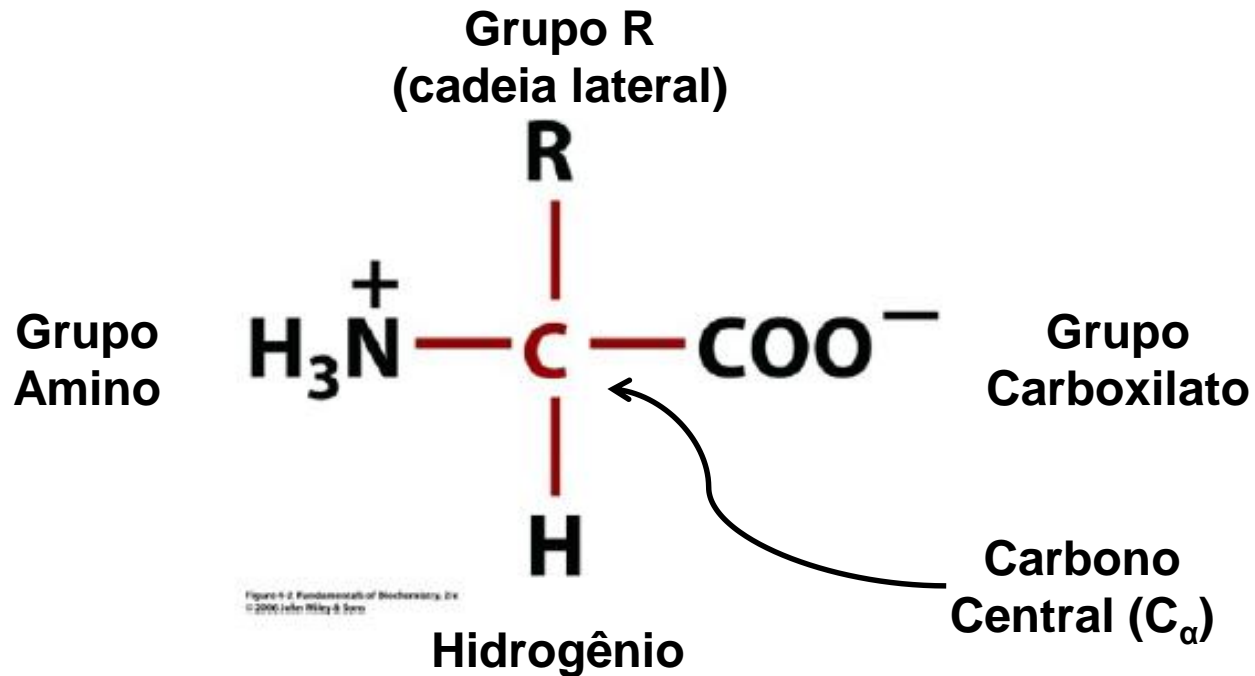
Possuem ampla diversidade de funções:

- Estrutural ou mecânica;
- Catalisador de reações;
- Regulação do metabolismo;
- Proteção imunológica;
- Transportador;
- Sinalização celular e comunicação.



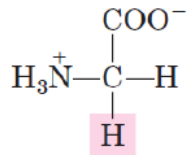
Componentes primários: Aminoácidos

- As proteínas são polímeros lineares formados por subunidades monoméricas, os **aminoácidos**, ligadas covalentemente entre si.
- São **20 aminoácidos** podendo ser combinados de diversas formas dando **origem a sequências** e, portanto, **proteínas diferentes**.

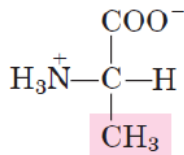


Aminoácidos

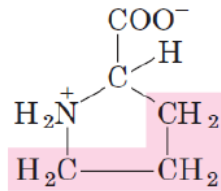
Nonpolar, aliphatic R groups



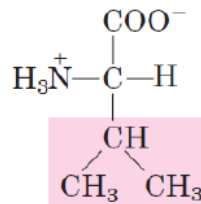
Glycine



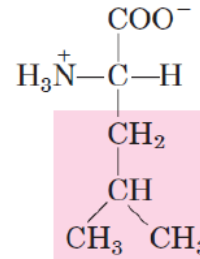
Alanine



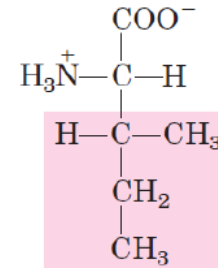
Proline



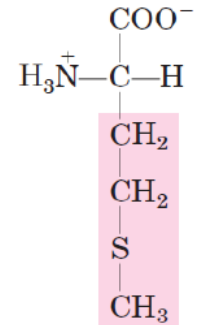
Valine



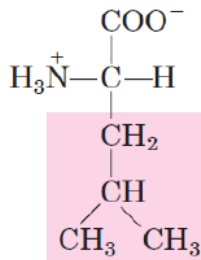
Leucine



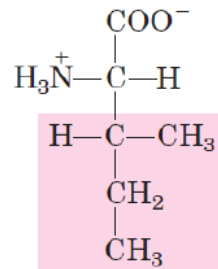
Isoleucine



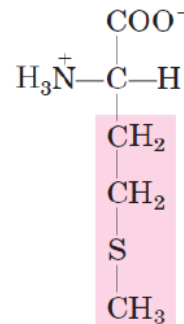
Methionine



Leucine

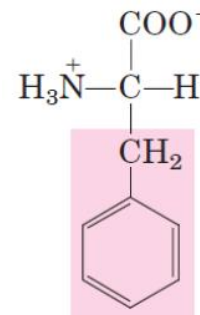


Isoleucine

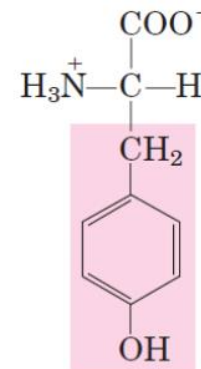


Methionine

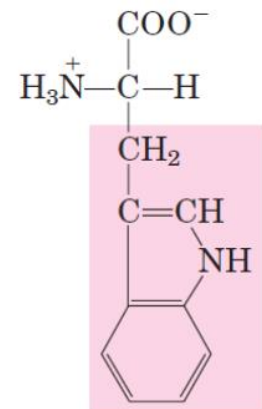
Aromatic R groups



Phenylalanine



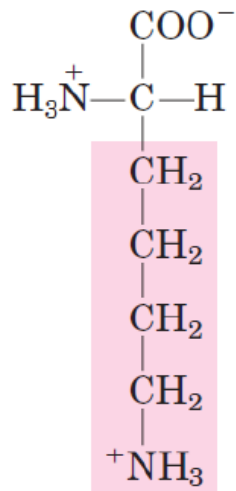
Tyrosine



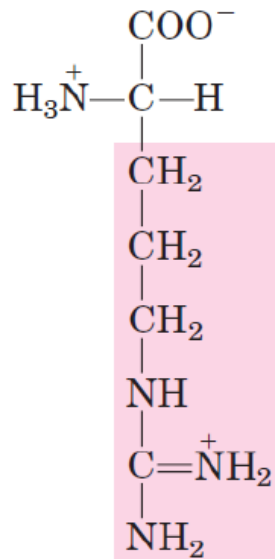
Tryptophan

Aminoácidos

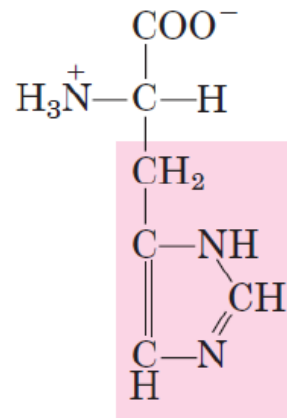
Carga positiva



Lysine

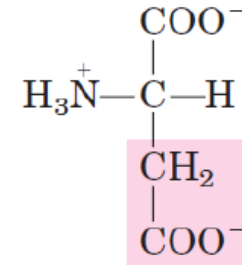


Arginine

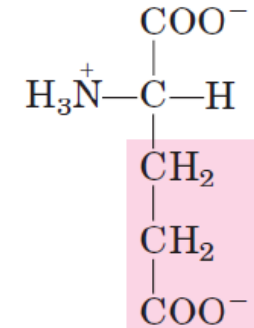


Histidine

Carga negativa

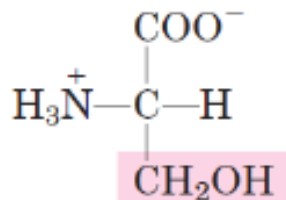


Aspartate

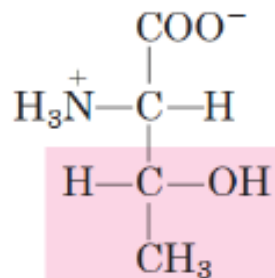


Glutamate

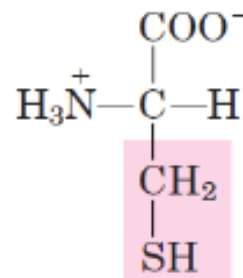
Polar, uncharged R groups



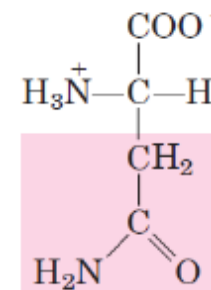
Serine



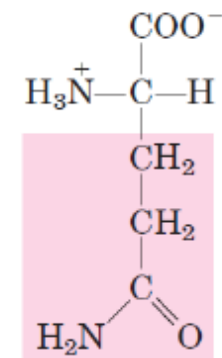
Threonine



Cysteine

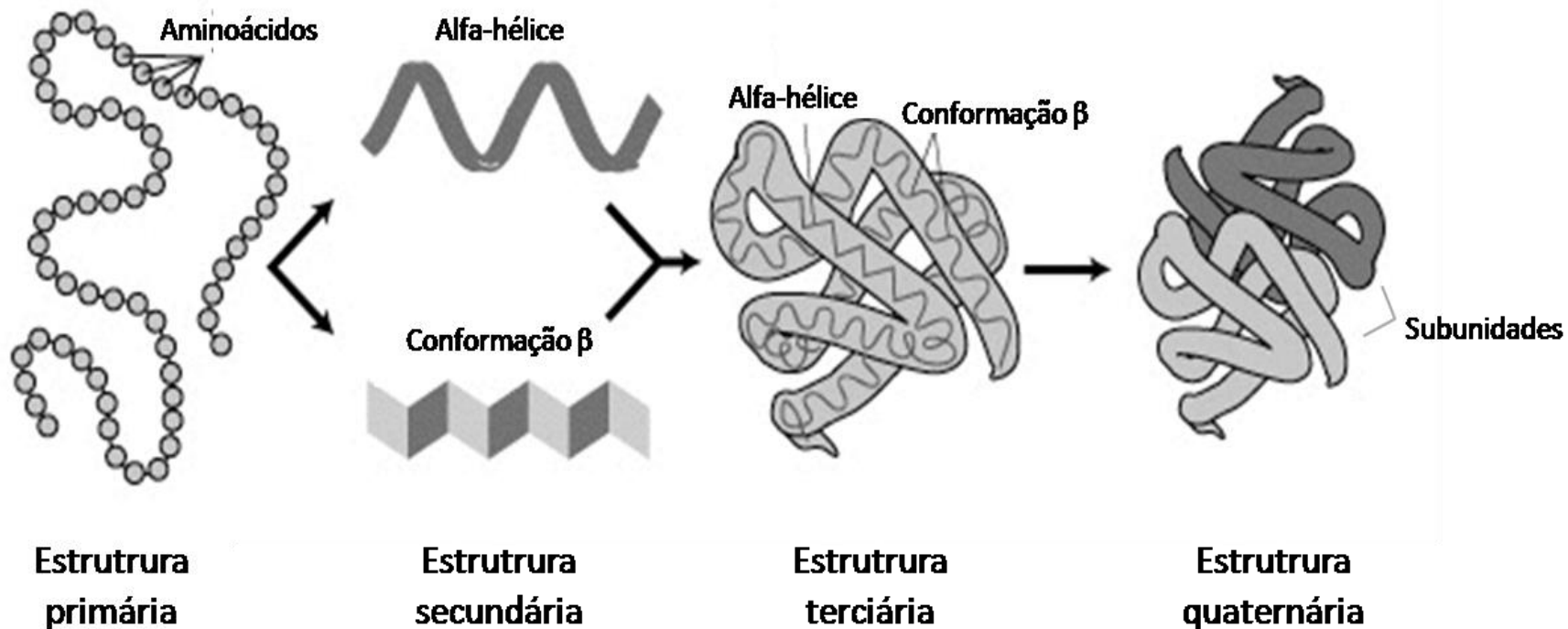


Asparagine

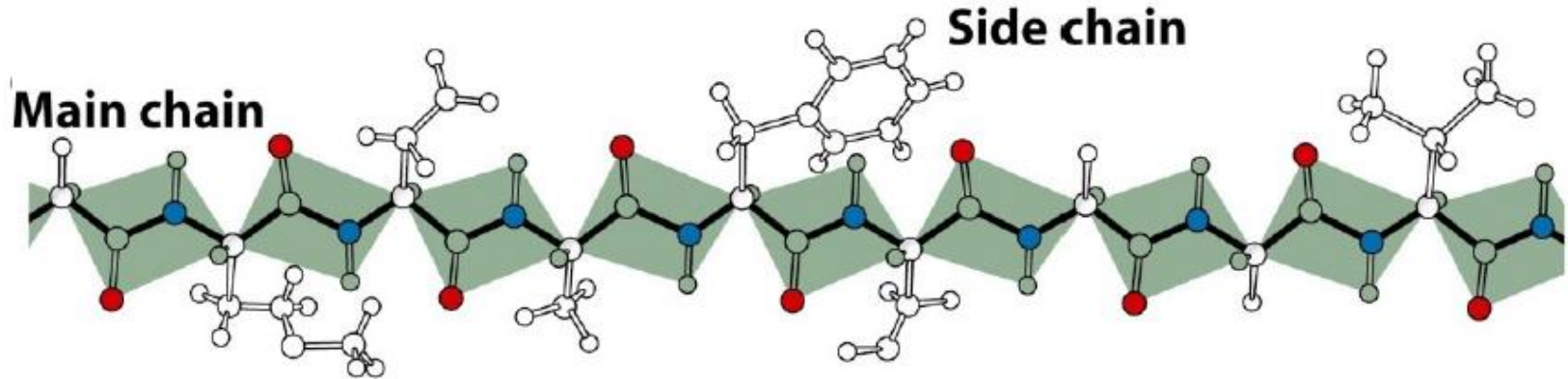


Glutamine

Níveis de organização estrutural de proteínas



Cadeia peptídica



- **Cadeia principal** que se repete regularmente (*main chain*), atuando como o “**esqueleto da proteína**” (*backbone*).
- **Cadeia lateral** (*side chain*) é **variável**.
- A conformação da ligação peptídica é definida por **ângulos de torção** ou **ângulos diedros**, que são os ângulos de interseção entre dois planos, denominados ϕ (phi), ψ (psi).

Ângulos Φ e Ψ determinam a conformação da cadeia polipeptídica

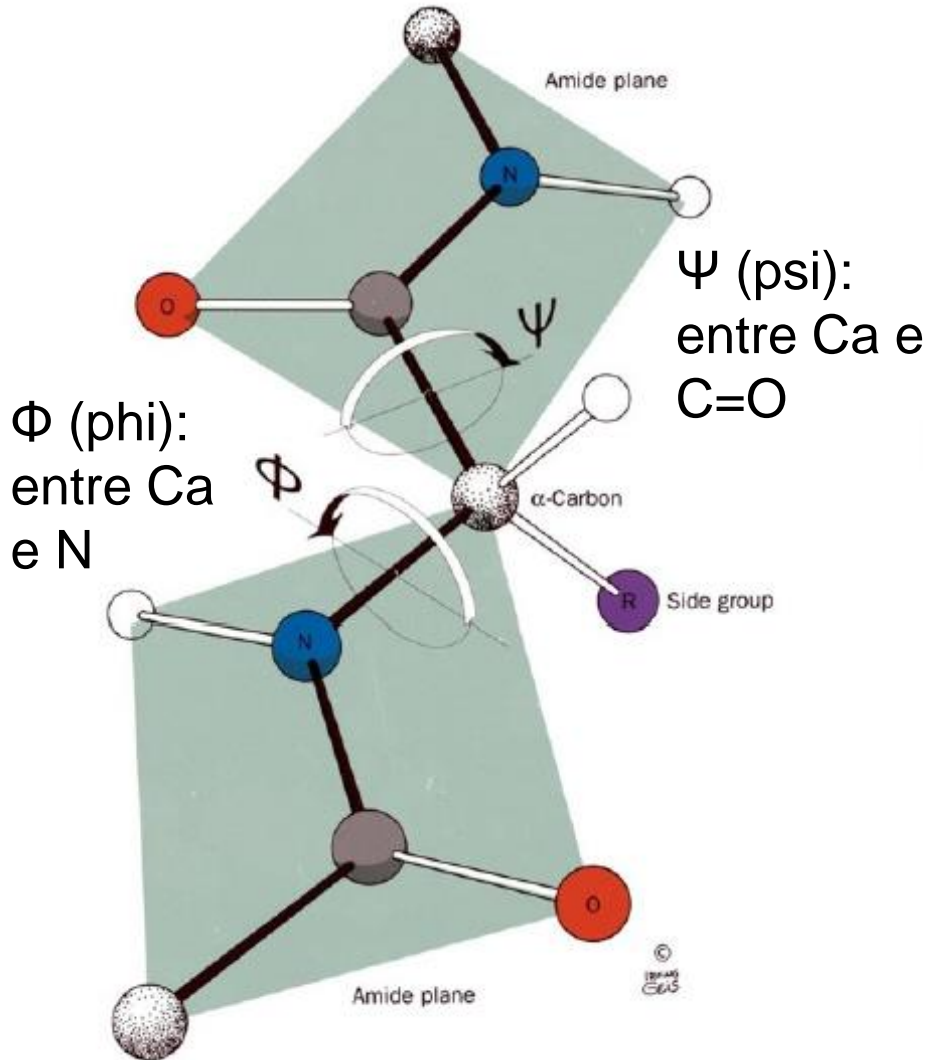


Figure 6-4 Fundamentals of Biochemistry, 2/e

Os ângulos ϕ e ψ podem assumir valores entre -180° e $+180^\circ$

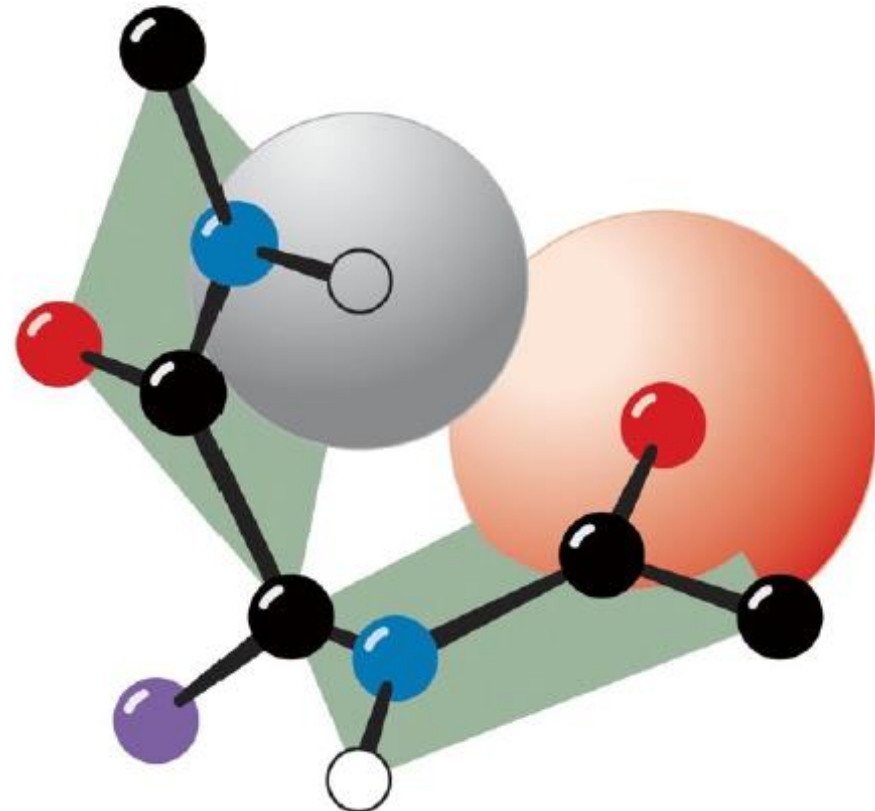


Figure 6-5 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Impedimentos estéricos

Valores permitidos de Φ e Ψ podem ser visualizados no gráfico de Ramachandran

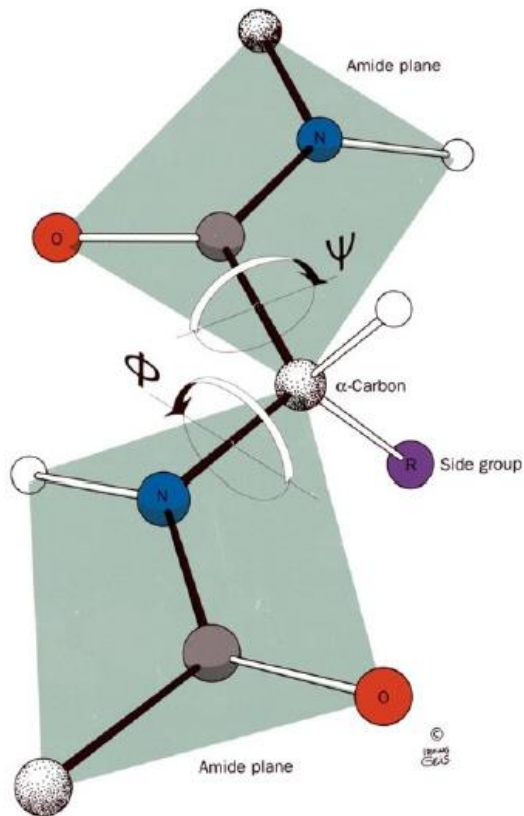
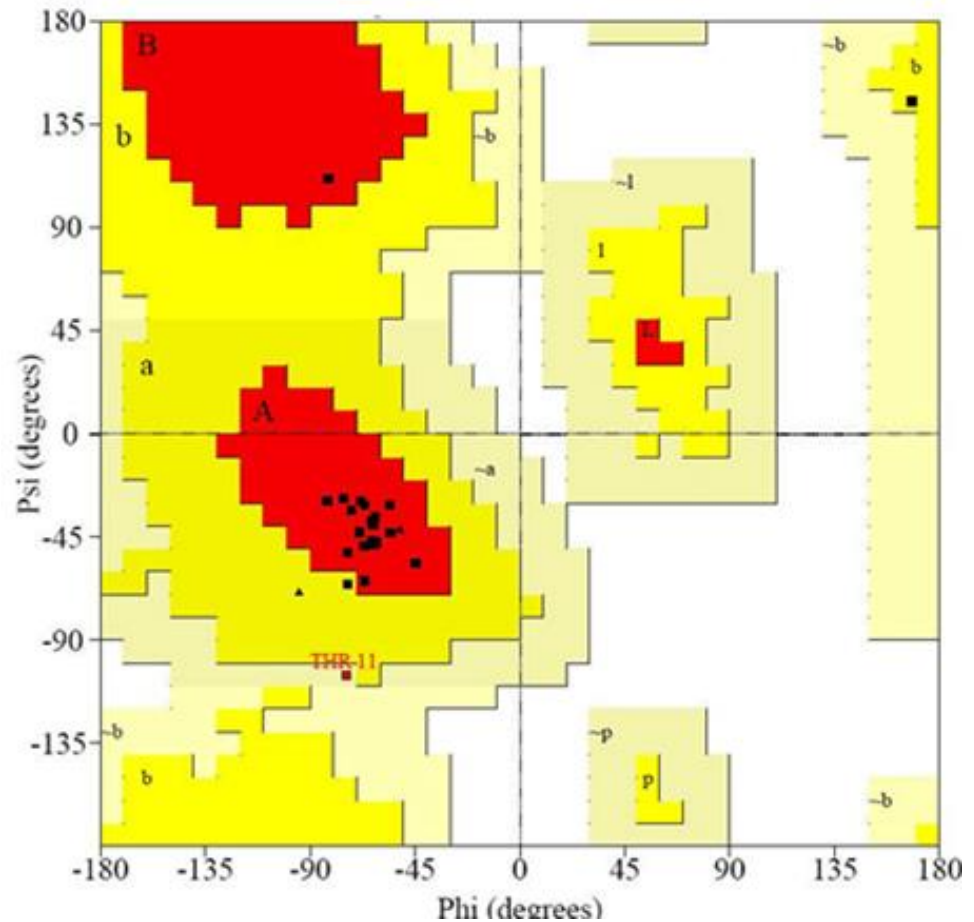


Figure 6-4 Fundamentals of Biochemistry, 2/e

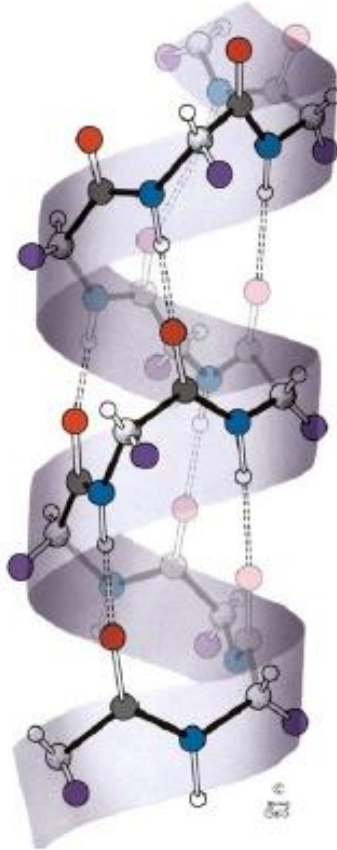


As regiões representadas em **vermelho**, são as **mais favoráveis**, as representadas em **amarelo**, são as **favoráveis**, as representadas em **bege**, são as **menos favoráveis** e as regiões **brancas**, são as **desfavoráveis**. 11

Estruturas secundárias regulares de proteínas

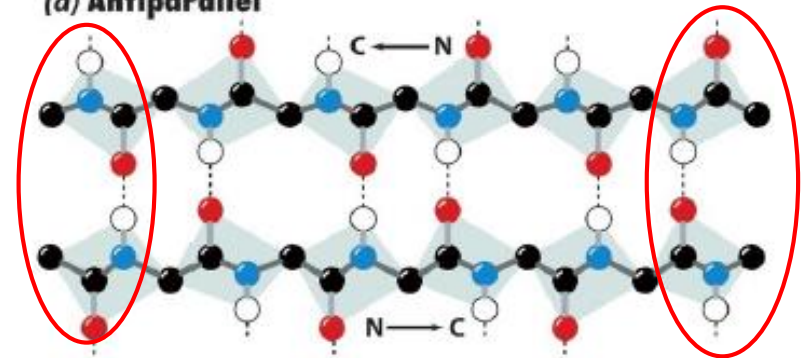
hélices alfa

- Ângulos de torção com $\varphi = -57^\circ$ e $\psi = -47^\circ$
- Volta composta por 3,6 resíduos
- Lig. de H entre os grupos C=O de cada aminoácido com o N-H



folhas beta

(a) Antiparallel



(b) Parallel

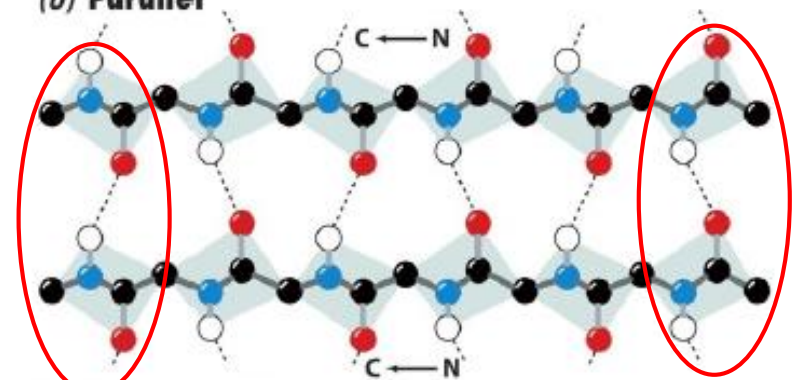
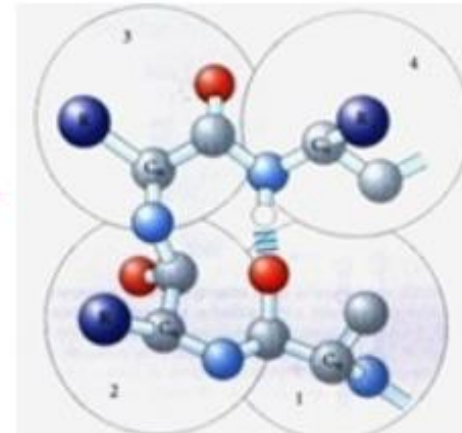


Figure 6-9 Fundamentals of Biochemistry, 2/e

- Conformação mais estendida e na forma de zigue-zague
- repetições nos ângulos de torções e nas ligações de hidrogênio.

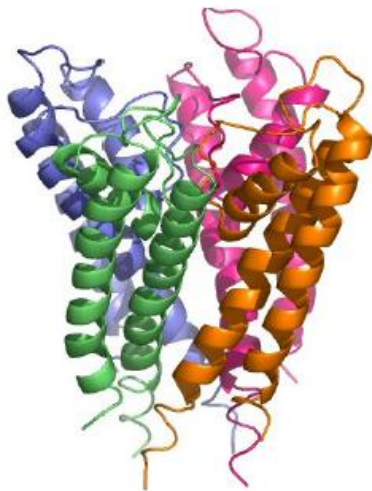
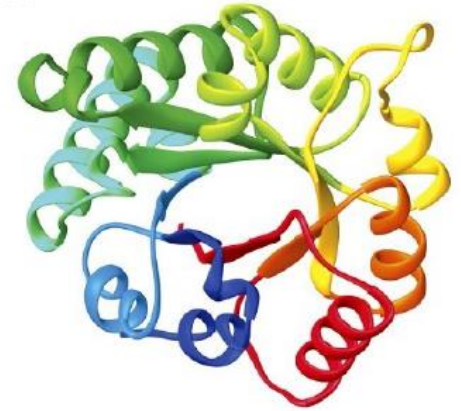
Loops / Dobras / Alças

- Conexão entre folhas β e α -hélices
- Mais **flexíveis** e **expostas** ao solvente



Estruturas terciária e quaternária

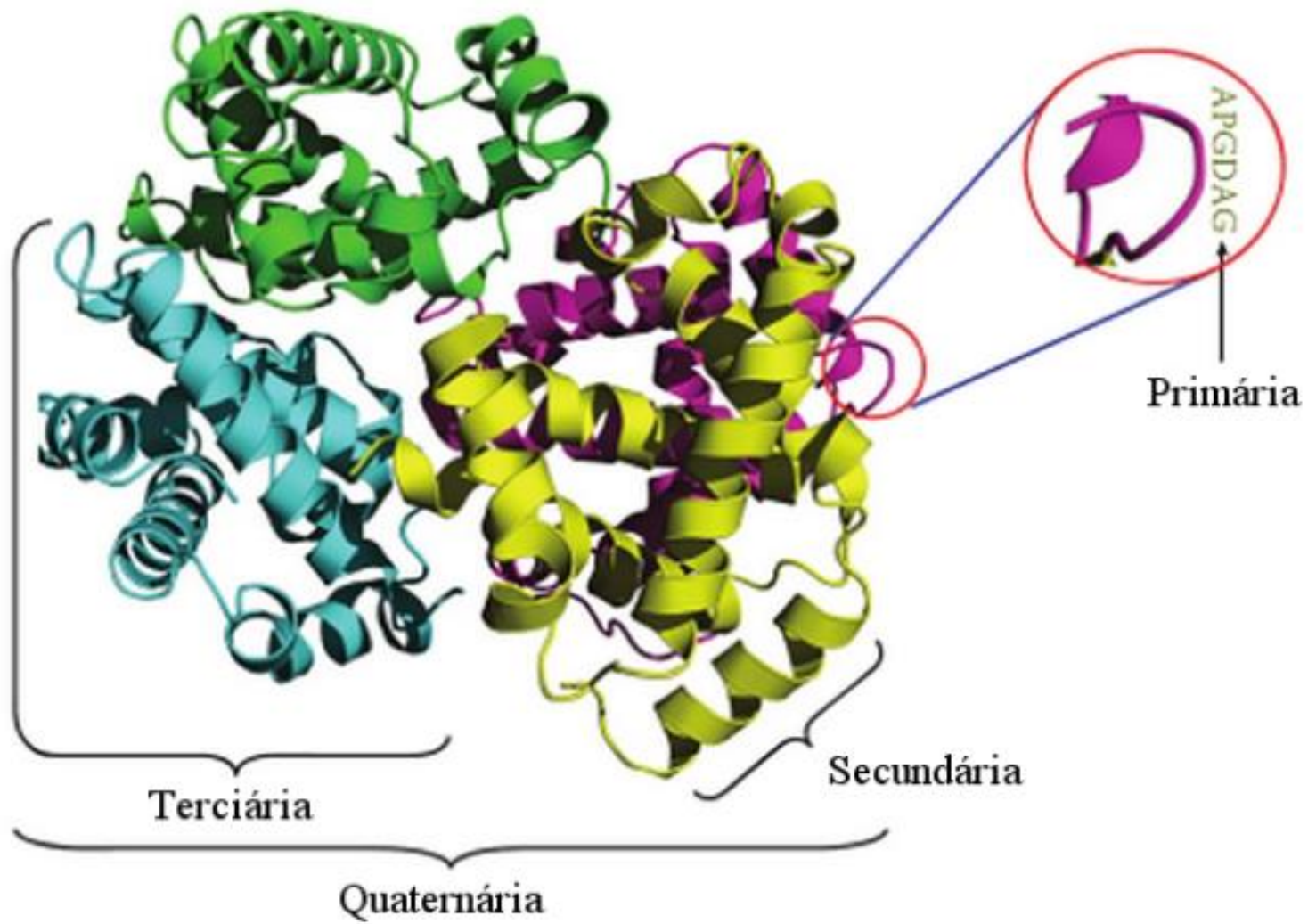
- **Estruturas terciárias** constituem a **conformação global** de uma cadeia.
- Elementos de estrutura secundária se **dobram** e se **organizam** até que a proteína atinja a sua **conformação final**, no qual a proteína assumirá a sua função biológica.
- **Cada proteína** apresenta uma **estrutura terciária** característica.



Canal de potássio

- **Estrutura quaternária** constitui uma proteína **com duas ou mais cadeias polipeptídicas** ou **subunidades iguais**.
- **Homodímero**, no caso de **duas subunidades iguais** ou **Heterodímero**, no caso de **duas subunidades distintas**.

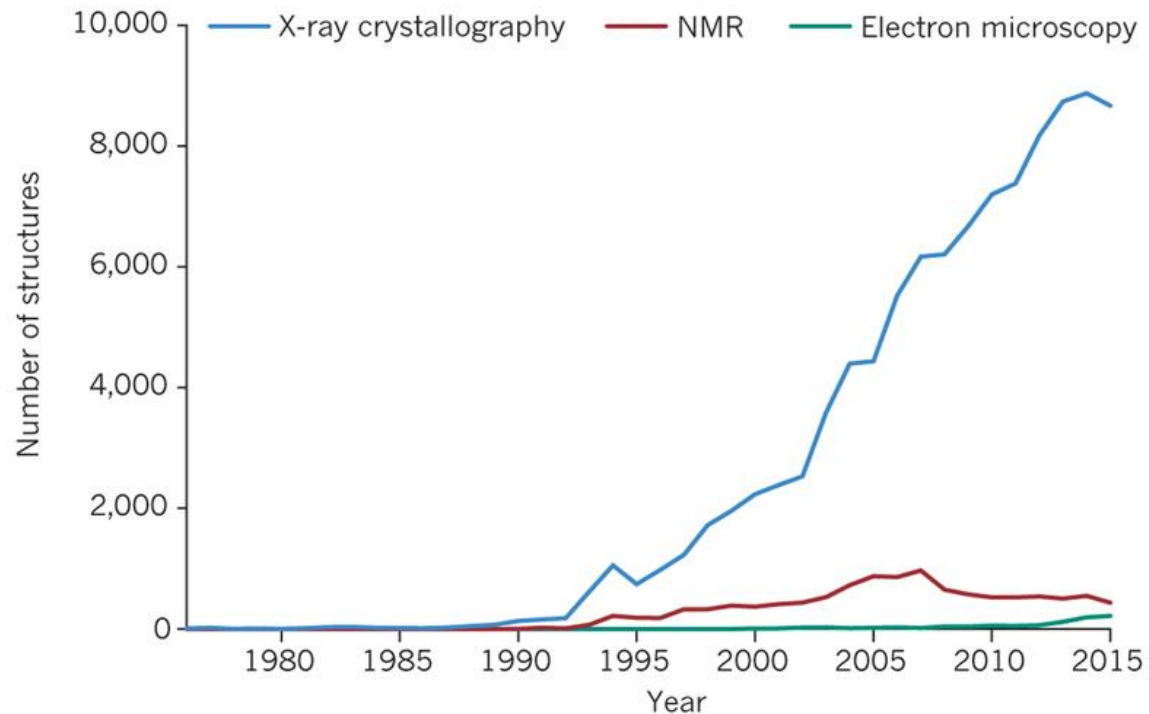
No final ...



Como obter estruturas?

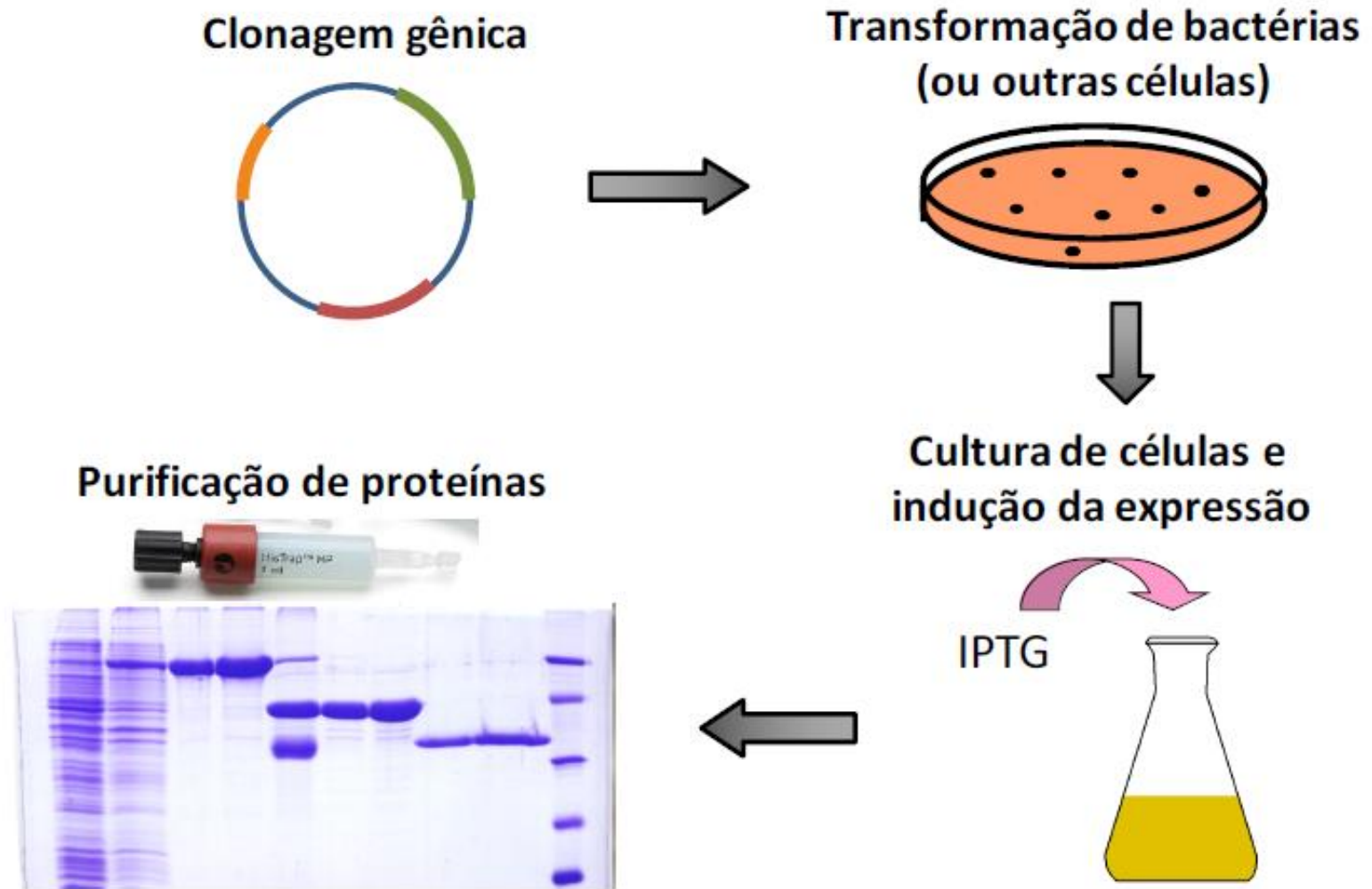
- **Determinação experimental**

- Cristalografia
- RMN
- Criomicroscopia eletrônica (Cryo-EM)
- entre outras



- **Modelagem**

Como obter estruturas?



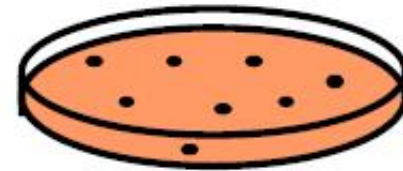
Processo de produção e purificação da proteína alvo

Como obter estruturas?

Clonagem gênica

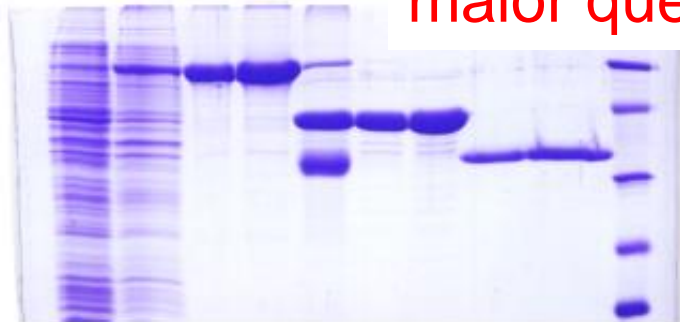


Transformação de bactérias
(ou outras células)



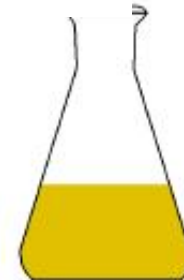
Quantidade e pureza suficiente para os estudos cristalográficos (em torno de miligramas de proteína com teor de pureza maior que 95%).

Purificação de



as e
essão

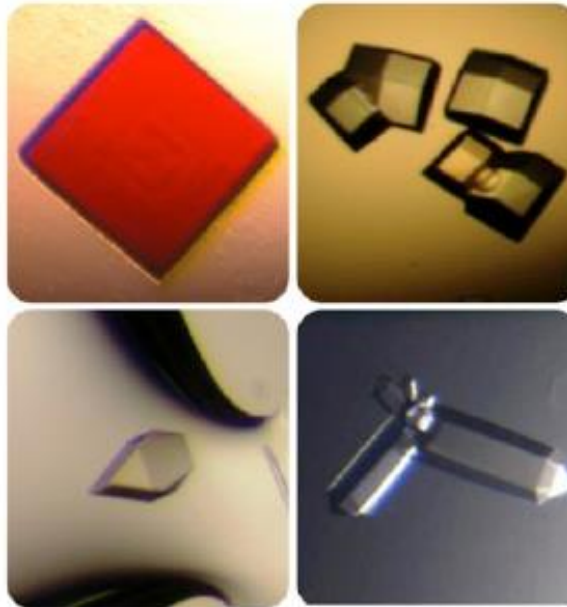
IPTG



Processo de produção e purificação da proteína alvo

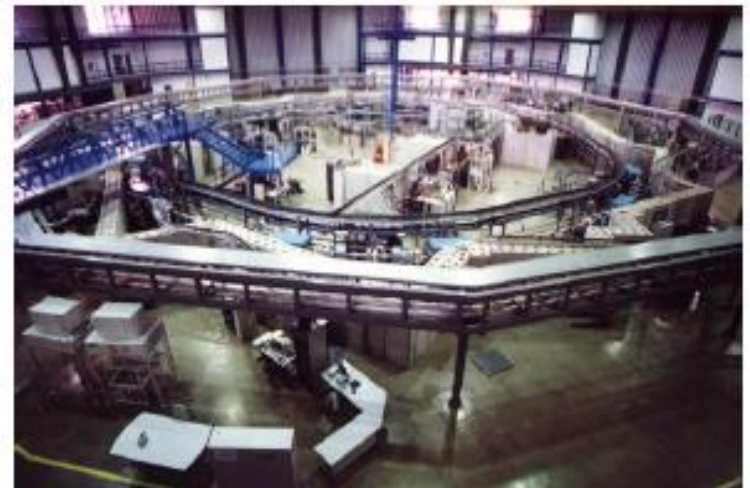
Cristalografia

- A cristalografia é uma técnica física capaz de dar informações estruturais sobre um material, desde que este esteja **disponível** na forma de um **crystal**.
- Os métodos cristalográficos dependem da análise dos padrões de **difração** de uma amostra direcionada por um **feixe de algum tipo**. **Raios-X** são mais comumente usados; outros tipos usados incluem **elétrons** ou **nêutrons**.



Coleta de dados

- Cada tipo de feixe usado possui um método associado: **cristalografia de raios-X, difração de nêutrons e difração de elétrons**. Estes três tipos de radiação interagem de maneiras diferentes.
- Tais dados podem ser coletados em **difratômetros de laboratório** ou em **linhas de luz sincrotron**.
- A informação obtida corresponde a um **mapa de densidade eletrônica**, que pode ser interpretado de forma a produzir um modelo da estrutura da molécula que compõe o cristal.



Coleta de dados

Sirius: As mentes por trás do maior acelerador de partículas do Brasil

Felipe Souza
Da BBC News Brasil em São Paulo

13 novembro 2018

f     Compartilhar

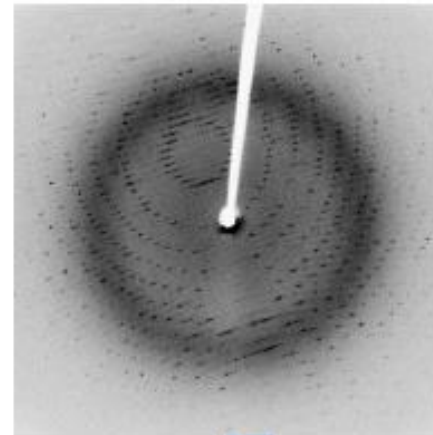


- Orçado em R\$ 1,8 bilhão, é a maior construção científica já feita no Brasil.
- A fonte de luz síncrotron é um tipo de radiação eletromagnética de alto fluxo e alto brilho produzida quando partículas carregadas, aceleradas a velocidades próximas à velocidade da luz, têm sua trajetória desviada por campos magnéticos.

A conclusão da montagem dos aceleradores do Sirius está prevista para o final de 2018 e o início da operação, para 2019. Já a conclusão do projeto, incluindo 13 estações de pesquisa, é previsto para 2020.

Cristalografia

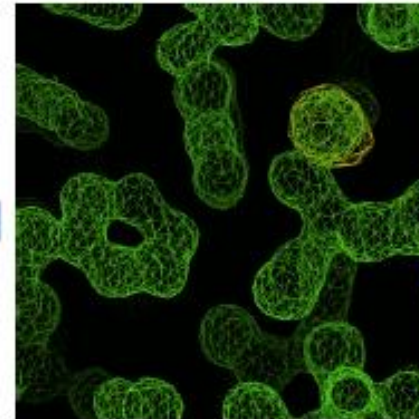
Obter cristais



Coletar dados de difração

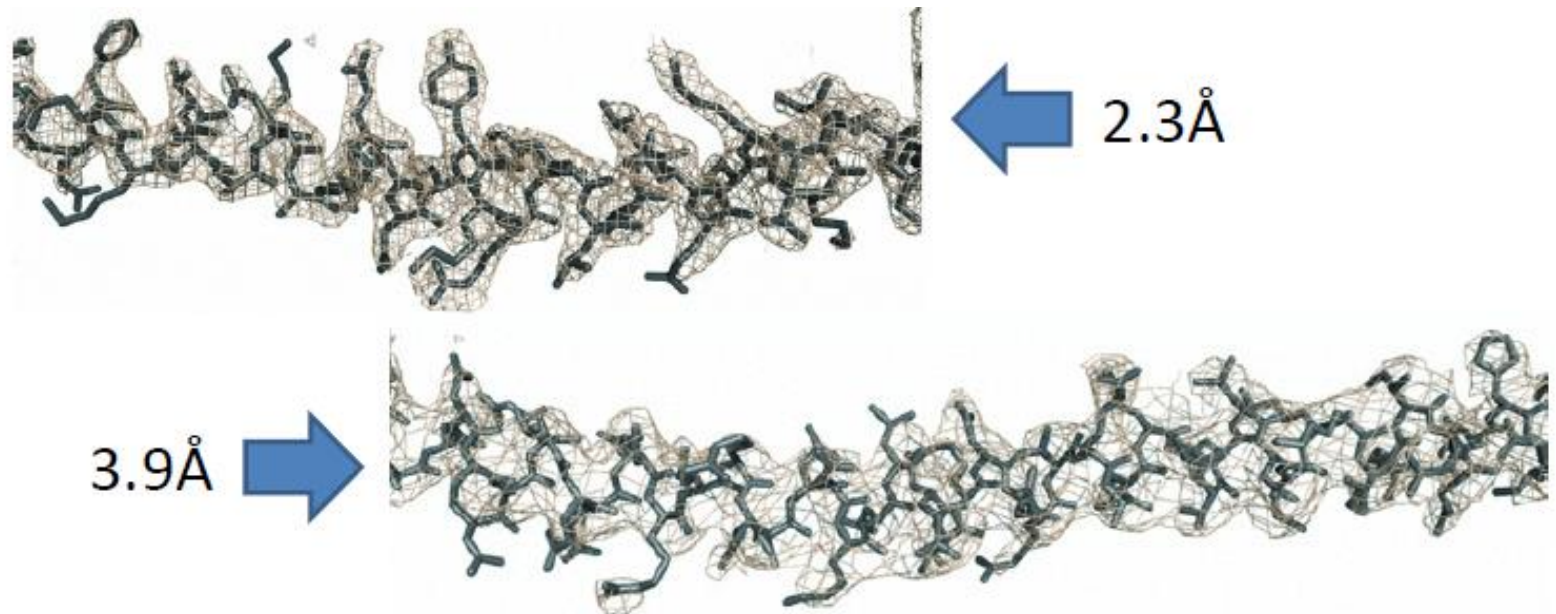


Interpretar dados



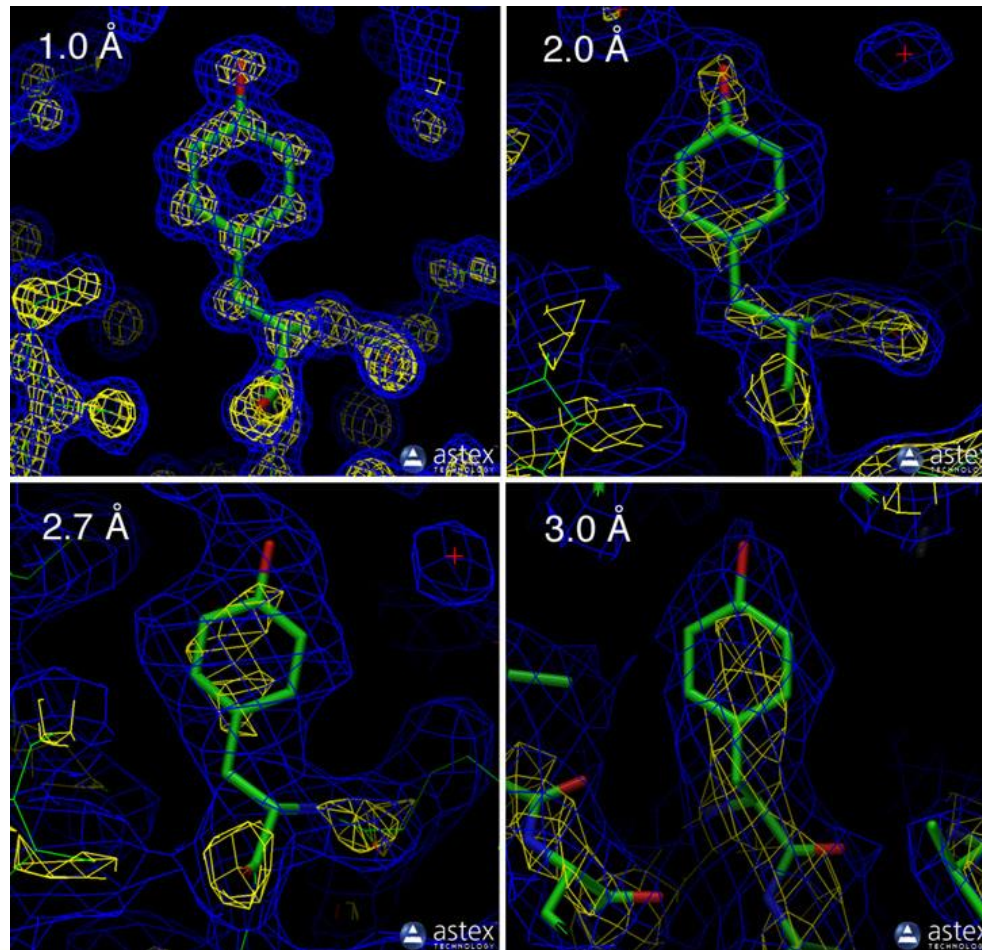
Qualidade do modelo

- Se todas as proteínas no cristal estiverem alinhadas de maneira idêntica, formando um **cristal muito perfeito**, todas as proteínas espalharão os raios X da mesma maneira, e o padrão de difração mostrará **os detalhes finos do cristal**.
- Por outro lado, se as proteínas no cristal **forem ligeiramente diferentes**, devido à **flexibilidade local ou ao movimento**, o padrão de difração não conterá tanta informação detalhada.
- Assim, a **resolução** é uma medida do nível de **detalhe** presente no **padrão de difração** e que será visto quando o mapa de densidade eletrônica é calculado.



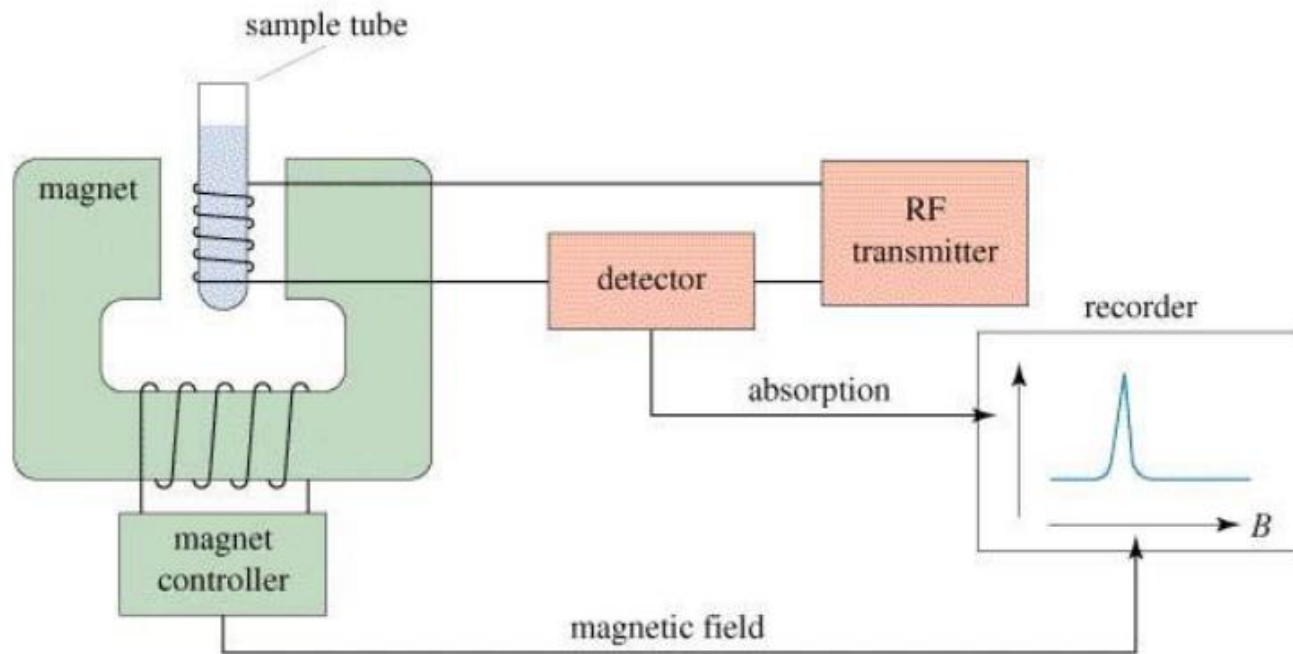
Qualidade do modelo

- O mapa de densidade eletrônica tem qualidade compatível com a **resolução**, parâmetro relacionado ao poder de difração



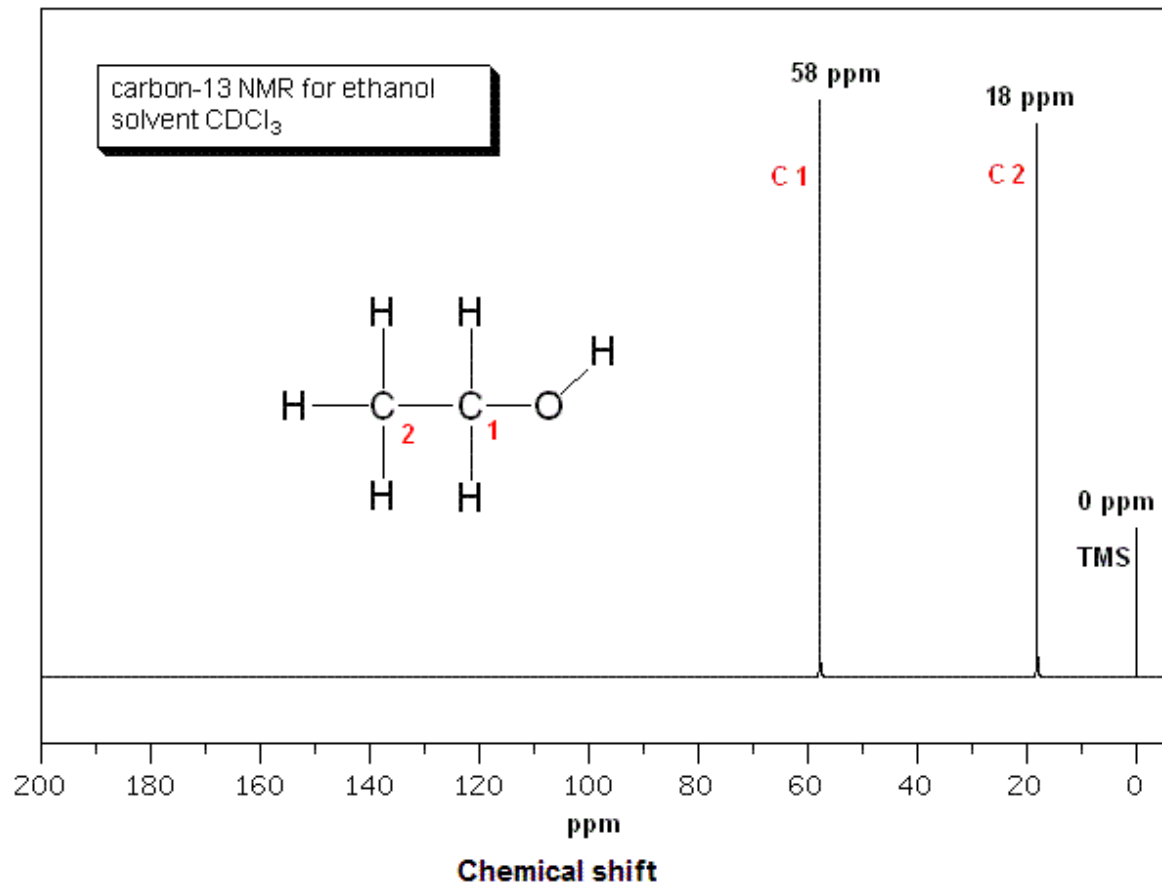
Ressonância magnética nuclear

- A **espectroscopia por RMN**, é uma técnica de pesquisa que explora as propriedades magnéticas de certos núcleos atômicos para determinar propriedades físicas ou químicas de átomos ou moléculas nos quais eles estão contidos.
- Baseia-se no fenômeno da ressonância magnética nuclear.
- A espectroscopia por RMN usa **proteína em solução**.



Ressonância magnética nuclear

- Ao invés de gerar um mapa de densidade eletrônica, os dados de RMN geram um conjunto de **restrições espaciais**, isto é, **distâncias interatômicas específicas**, com um dado grau de incerteza do cristal.



Ressonância magnética nuclear

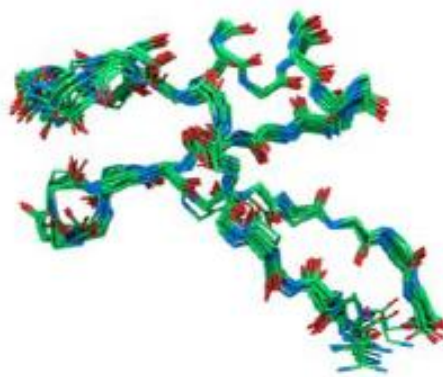
- Muitas vezes ha **vários modelos** capazes de descrever um conjunto de restrições espaciais. Por isso, os espectroscopistas não depositam uma, mas **várias** (entre dez e vinte) estruturas.

A



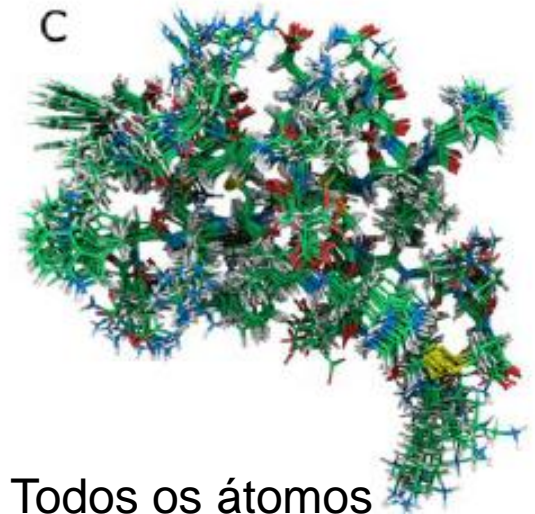
Estrutura Secundária

B



Cadeia Principal

C



Todos os átomos

Cristalografia X RMN

- **Prós:**

- produz modelos de alta resolução (0.5-10Å, tipicamente ~2.0Å) e confiabilidade
- é possível avaliar quais regiões são mais ou menos confiáveis.
- processamento de dados mais rápido

- **Contras:**

- necessita da obtenção de cristais, o que nem sempre é possível

- **Prós:**

- usa proteína em solução, pode dar alguns parâmetros dinâmicos.

- **Contras:**

- resolução mais baixa,
- limite prático de tamanho (~300 aa)
- validação menos estabelecida que na cristalografia
- Processamento de dados demorado

Criomicroscopia eletrônica (Cryo-EM)


NATURE | NEWS



Cryo-electron microscopy wins chemistry Nobel

Jacques Dubochet, Joachim Frank and Richard Henderson share the prize for developing a technique to image biomolecules.

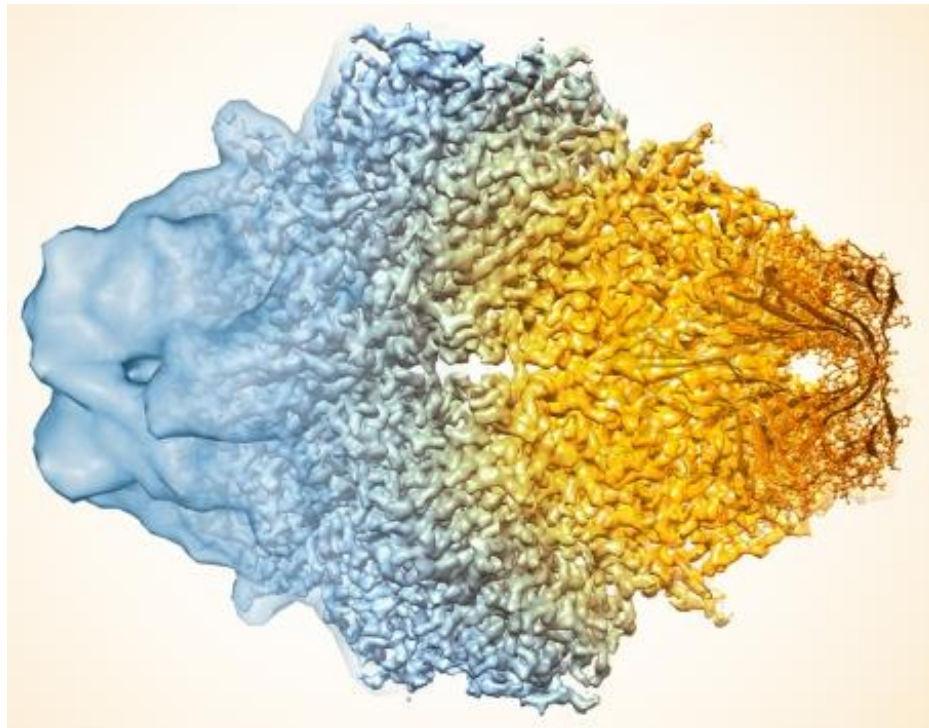
The Royal Swedish Academy of Sciences has decided to award the
2017 NOBEL PRIZE IN CHEMISTRY



Jacques Dubochet
Joachim Frank
Richard Henderson

"for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution"

Illustrations: Niklas Elmehed, Nobel Prize Media; © 2017 Nobel Foundation. Photos: Lennax Engblom.



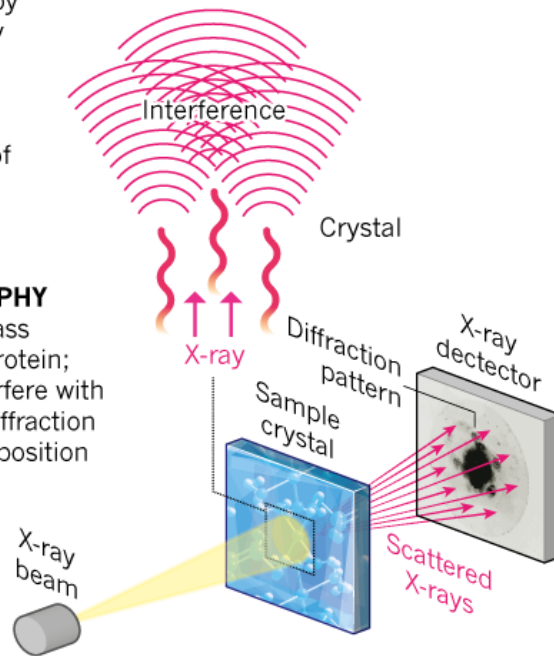
Cristalografia X Cryo-EM

THE RISE OF CRYO-ELECTRON MICROSCOPY

Cryo-electron microscopy is taking over from X-ray crystallography as a method to deduce high-resolution protein structures, particularly of large molecules.

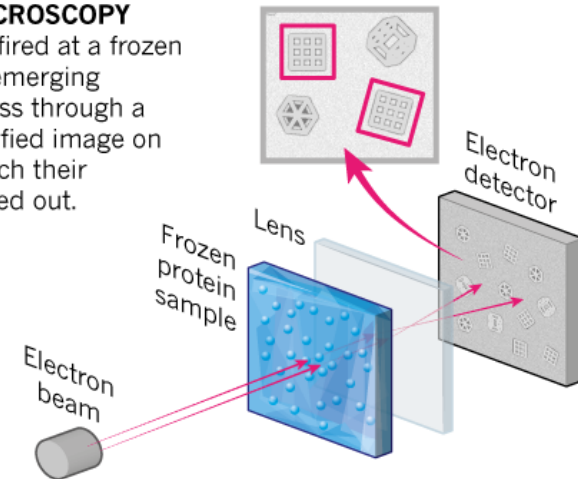
X-RAY CRYSTALLOGRAPHY

X-rays scatter as they pass through a crystallized protein; the resulting waves interfere with each other, creating a diffraction pattern from which the position of atoms is deduced.



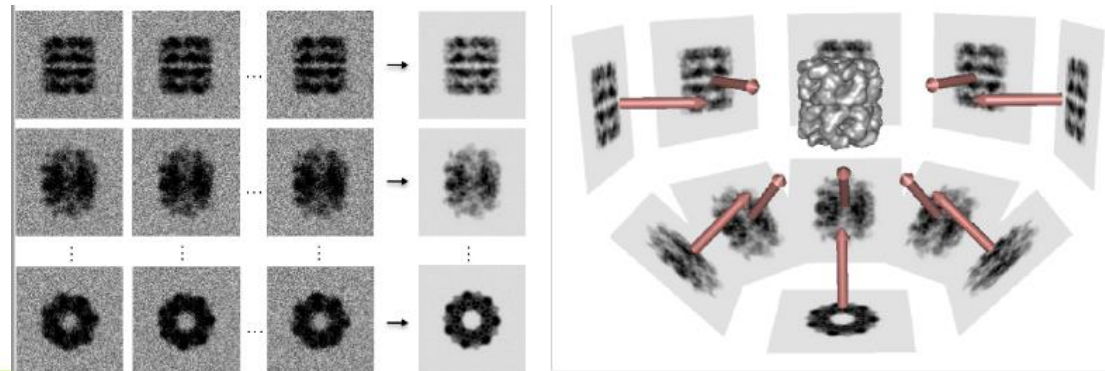
CRYO-ELECTRON MICROSCOPY

A beam of electron is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, from which their structure can be worked out.



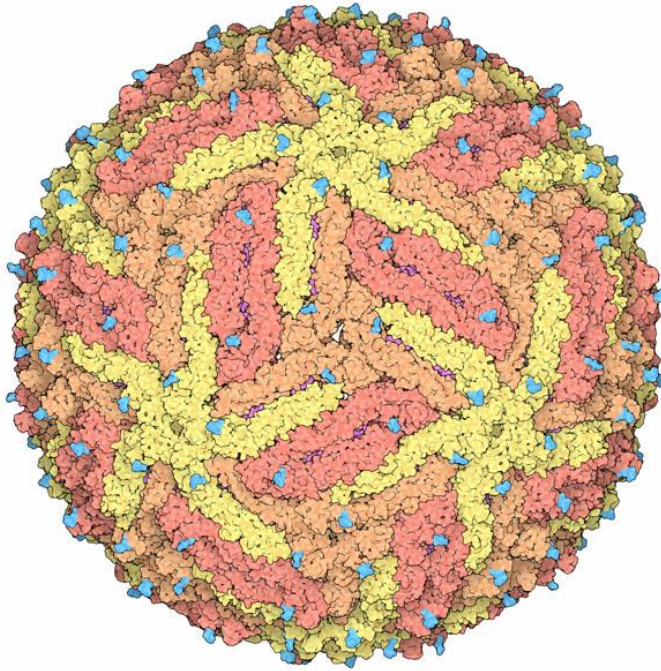
©nature

Reconstrução da imagem

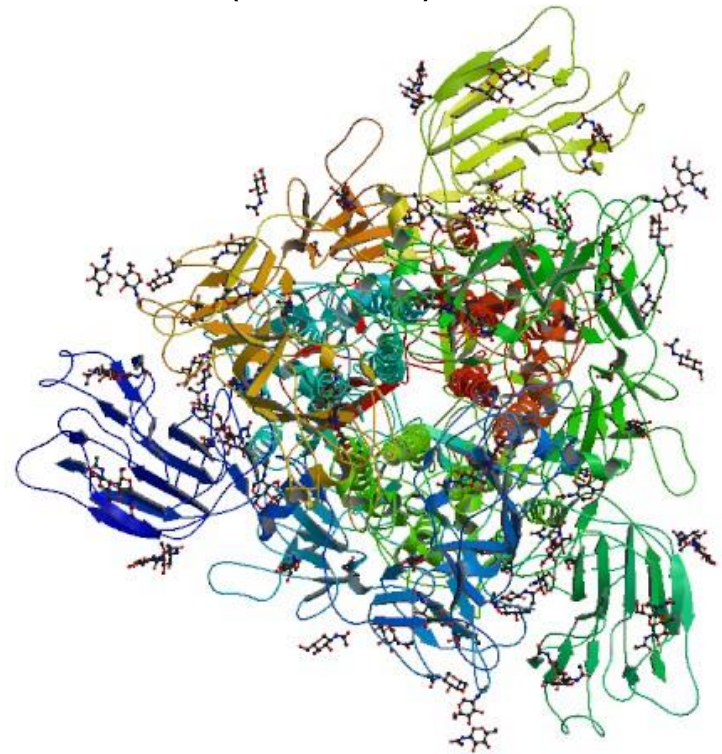


Desafios Cryo-EM

- Moléculas do ar também desviam elétrons.
- Preparação de amostra em vácuo.
- Fonte de elétrons tende a causar danos por radiação.
- Amostra preparada a temperaturas baixas (-200 °C).



Zika Virus



Structure of infectious bronchitis
coronavirus spike protein 31

Onde estão as estruturas?

O PDB (Protein Data Bank) **cataloga, trata e disponibiliza** estruturas de macromoléculas biológicas para as quais é possível, ao menos, traçar a cadeia polipeptídica.

www.rcsb.org

The screenshot shows the RCSB PDB website homepage. At the top, there is a navigation bar with links for Deposit, Search, Visualize, Analyze, Download, Learn, and More. A search bar is prominently displayed with the text "Search by PDB ID, author, macromolecule, sequence, or ligands" and a "Go" button. Below the search bar, there are logos for PDB-101, PDB, EMDatabank, Nucleic Acid Databases, and Worldwide Protein Data Bank Foundation. The main content area is divided into three sections: a left sidebar with navigation options (Welcome, Deposit, Search, Visualize, Analyze, Download, Learn), a central section titled "A Structural View of Biology" which describes the resource's purpose and provides links for "Advanced Search" and "Browse by Annotations", and a right section titled "June Molecule of the Month" featuring a 3D molecular model of a protein and nanoparticles.

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

RCSB PDB PROTEIN DATA BANK 141010 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

PDB-101 PDB EMDatabank NUCLEIC ACID DATABASES Worldwide Protein Data Bank Foundation

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

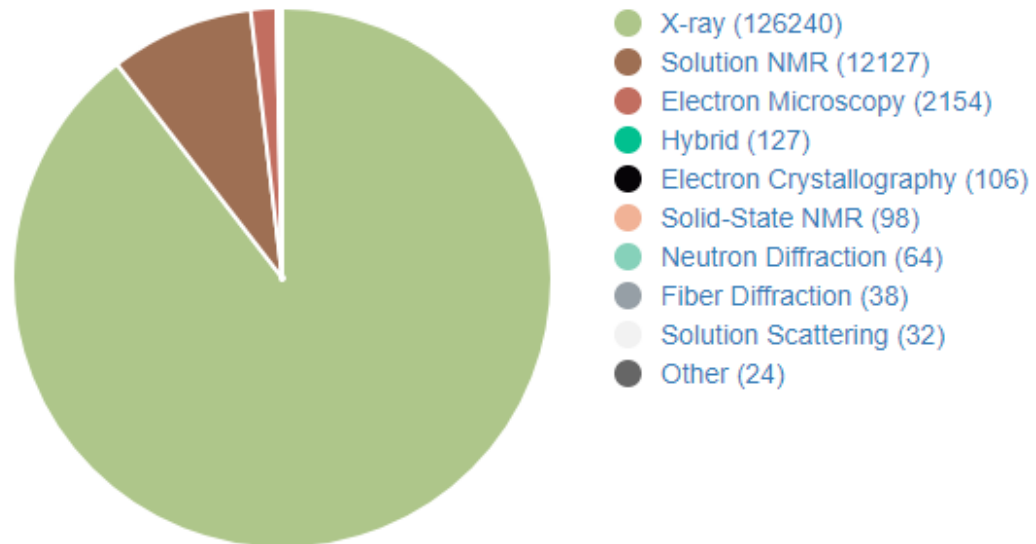
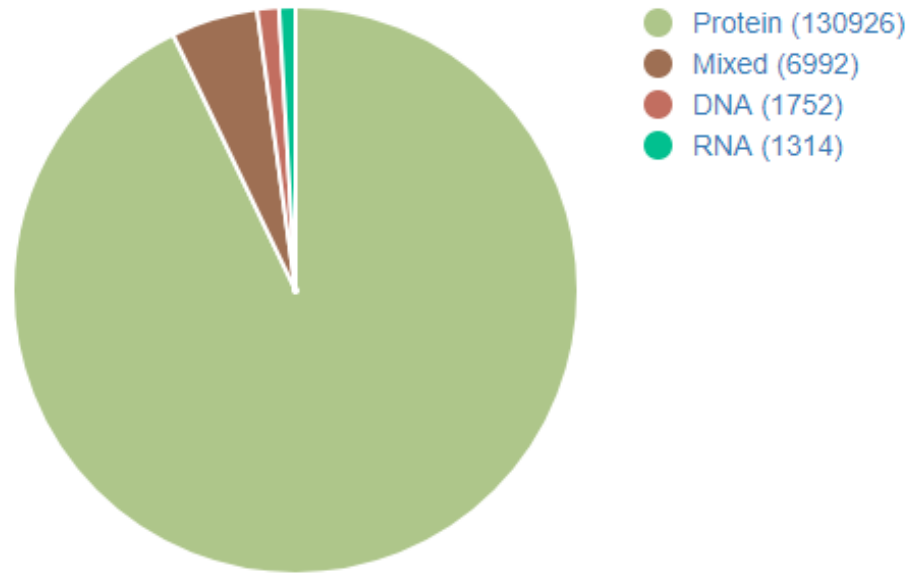
The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

RCSB PDB Services and Impact

June Molecule of the Month

Proteins and Nanoparticles

Onde estão as estruturas?



Como é organizado?

Structure Summary

3D View

Annotations

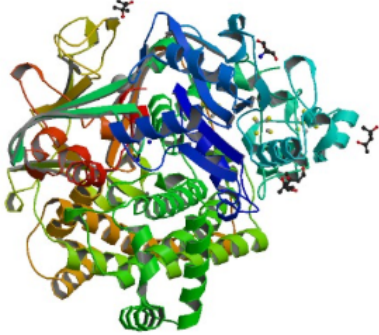
Sequence

Sequence Similarity

Structure Similarity

Experiment

Biological Assembly 1 ?



3D View: [Structure](#) | [Electron Density](#) | [Ligand Interaction](#)

Standalone Viewers

5XLE

Crystal structure of anaerobically purified and anaerobically crystallized *D. vulgaris* Miyazaki F [NiFe]-hydrogenase

DOI: 10.2210/pdb5XLE/pdb

Classification: [OXIDOREDUCTASE](#)

Organism(s): [Desulfovibrio vulgaris \(strain Miyazaki F / DSM 19637\)](#)

Deposited: 2017-05-10 Released: 2018-06-06

Deposition Author(s): [Nishikawa, K.](#), [Mochida, S.](#), [Hiromoto, T.](#), [Shibata, N.](#), [Higuchi, Y.](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.69 Å

R-Value Free: 0.176

R-Value Work: 0.141

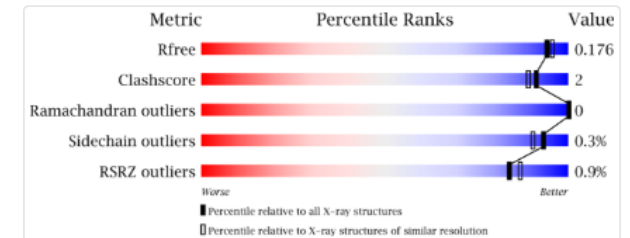
Display Files

Download Files

wwPDB Validation

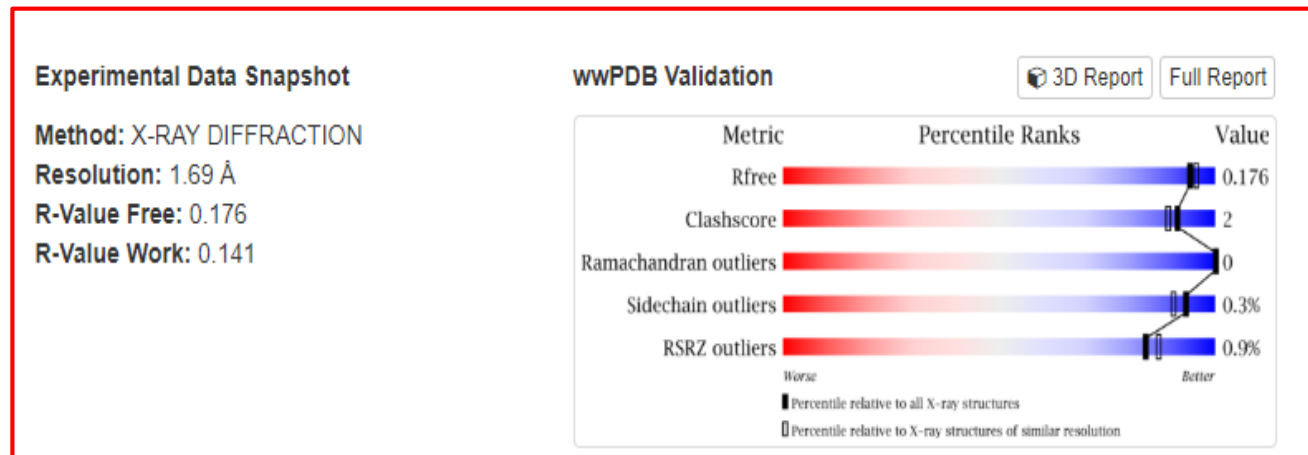
3D Report

Full Report



Como é organizado?

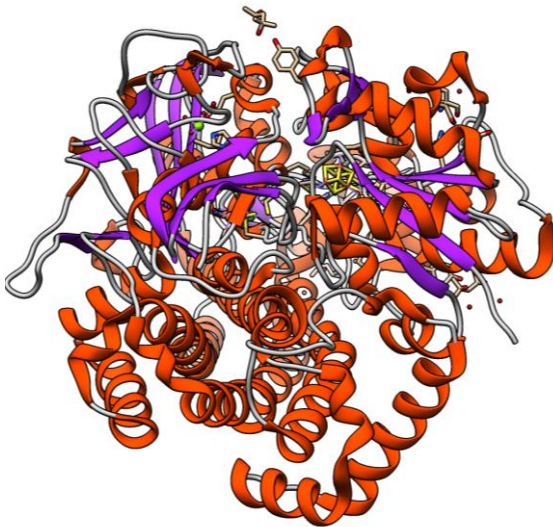
- **Resolução**: quanto **menor** for o valor numérico da resolução (diz-se quanto *maior* a resolução), mais bem definido estará o mapa de densidade eletrônica, e mais confiável a estrutura.
- R_{factor} : mede o erro do modelo proposto em relação aos dados experimentais.
- R_{free} : similar ao R_{free} , mas utiliza um pequeno conjunto de dados experimentais **não** usado no refinamento, para evitar ajustes “forçados”.



O que é depositado?

O formato **.pdb** é um **arquivo texto de estrutura fixa** que representa um modelo dos dados coletados experimentalmente. Contém:

- coordenadas atômicas
- características químicas e bioquímicas
- detalhes experimentais da determinação estrutural
- características estruturais
 - estrutura secundária, ligações de hidrogênio e sítio de ligação



SCALE3		0.000000	0.000000	0.007838	0.000000					
ATOM	1	N	PRO S	4	-33.864	93.461	47.905	1.00	35.98	N
ATOM	2	CA	PRO S	4	-33.125	94.691	47.584	1.00	39.57	C
ATOM	3	C	PRO S	4	-34.014	95.682	46.830	1.00	30.96	C
ATOM	4	O	PRO S	4	-34.600	95.304	45.811	1.00	35.78	O
ATOM	5	CB	PRO S	4	-31.998	94.189	46.679	1.00	38.71	C
ATOM	6	CG	PRO S	4	-32.569	92.990	46.002	1.00	37.32	C
ATOM	7	CD	PRO S	4	-33.545	92.362	46.975	1.00	40.42	C
ATOM	8	N	ARG S	5	-34.129	96.914	47.322	1.00	26.94	N
ATOM	9	CA	ARG S	5	-35.042	97.883	46.717	1.00	25.69	C
ATOM	10	C	ARG S	5	-34.554	98.316	45.333	1.00	27.01	C
ATOM	11	O	ARG S	5	-33.379	98.636	45.156	1.00	25.07	O
ATOM	12	CB	ARG S	5	-35.227	99.103	47.617	1.00	26.69	C
ATOM	13	CG	ARG S	5	-36.274	100.061	47.101	1.00	31.38	C
ATOM	14	CD	ARG S	5	-36.565	101.188	48.073	1.00	30.50	C
ATOM	15	NE	ARG S	5	-37.801	101.866	47.692	1.00	33.57	N
ATOM	16	CZ	ARG S	5	-39.011	101.474	48.084	1.00	35.07	C
ATOM	17	NH1	ARG S	5	-39.145	100.416	48.880	1.00	34.62	N
ATOM	18	NH2	ARG S	5	-40.086	102.144	47.687	1.00	37.64	N
ATOM	19	N	ARG S	6	-35.456	98.307	44.353	1.00	17.98	N

O arquivo .pdb (cabeçalho)

```
HEADER      OXIDOREDUCTASE                               10-MAY-17   5XLE
TITLE      CRYSTAL STRUCTURE OF ANAEROBICALLY PURIFIED AND ANAEROBICALLY
TITLE      2 CRYSTALLIZED D. VULGARIS MIYAZAKI F [NIFE]-HYDROGENASE
CAVEAT     5XLE      MPD L 603 HAS WRONG CHIRALITY AT ATOM C4
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: PERIPLASMIC [NIFE] HYDROGENASE SMALL SUBUNIT;
COMPND     3 CHAIN: S;
COMPND     4 FRAGMENT: UNP RESIDUES 51-317;
COMPND     5 SYNONYM: NIFE HYDROGENLYASE SMALL CHAIN;
COMPND     6 EC: 1.12.2.1;
COMPND     7 MOL_ID: 2;
COMPND     8 MOLECULE: PERIPLASMIC [NIFE] HYDROGENASE LARGE SUBUNIT;
COMPND     9 CHAIN: L;
COMPND    10 FRAGMENT: UNP RESIDUES 1-552;
COMPND    11 SYNONYM: NIFE HYDROGENLYASE LARGE CHAIN;
COMPND    12 EC: 1.12.2.1
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM_SCIENTIFIC: DESULFOVIBRIO VULGARIS (STRAIN MIYAZAKI F / DSM
SOURCE     3 19637);
SOURCE     4 ORGANISM_TAXID: 883;
SOURCE     5 STRAIN: MIYAZAKI F / DSM 19637;
SOURCE     6 MOL_ID: 2;
SOURCE     7 ORGANISM_SCIENTIFIC: DESULFOVIBRIO VULGARIS (STRAIN MIYAZAKI F / DSM
SOURCE     8 19637);
SOURCE     9 ORGANISM_TAXID: 883;
SOURCE    10 STRAIN: MIYAZAKI F / DSM 19637
KEYWDS     [NIFE]-HYDROGENASE D. VULGARIS, OXIDOREDUCTASE
EXPDTA     X-RAY DIFFRACTION
AUTHOR     K.NISHIKAWA,S.MOCHIDA,T.HIROMOTO,N.SHIBATA,Y.HIGUCHI
REVDAT    1   06-JUN-18 5XLE      0
JRNL       AUTH   K.NISHIKAWA,S.MOCHIDA,T.HIROMOTO,N.SHIBATA,Y.HIGUCHI
JRNL       TITL   NI-ELIMINATION FROM THE ACTIVE SITE OF THE STANDARD
JRNL       TITL 2 [NIFE]&#8208;HYDROGENASE UPON OXIDATION BY O2.
JRNL       REF   J INORG BIOCHEM                V 177    435 2017
```

O arquivo .pdb (Observações)

```
REMARK 2
REMARK 2 RESOLUTION.      1.69 ANGSTROMS.
REMARK 3
REMARK 3 REFINEMENT.
REMARK 3   PROGRAM      : PHENIX 1.8.2_1309
REMARK 3   AUTHORS     : PAUL ADAMS, PAVEL AFONINE, VINCENT CHEN, IAN
REMARK 3                 : DAVIS, KRESHNA GOPAL, RALF GROSSE-KUNSTLEVE,
REMARK 3                 : LI-WEI HUNG, ROBERT IMMORMINO, TOM IOERGER,
REMARK 3                 : AIRLIE MCCOY, ERIK MCKEE, NIGEL MORIARTY,
REMARK 3                 : REETAL PAI, RANDY READ, JANE RICHARDSON,
REMARK 3                 : DAVID RICHARDSON, TOD ROMO, JIM SACCHETTINI,
REMARK 3                 : NICHOLAS SAUTER, JACOB SMITH, LAURENT
REMARK 3                 : STORONI, TOM TERWILLIGER, PETER ZWART
REMARK 3
REMARK 3   REFINEMENT TARGET : ML
REMARK 3
REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.69
REMARK 3 RESOLUTION RANGE LOW  (ANGSTROMS) : 39.06
REMARK 3 MIN(FOBS/SIGMA_FOBS)      : 1.360
REMARK 3 COMPLETENESS FOR RANGE    (%) : 99.9
REMARK 3 NUMBER OF REFLECTIONS     : 94656
REMARK 3
REMARK 3 FIT TO DATA USED IN REFINEMENT.
REMARK 3 R VALUE      (WORKING + TEST SET) : 0.143
REMARK 3 R VALUE      (WORKING SET)       : 0.141
REMARK 3 FREE R VALUE                               : 0.176
REMARK 3 FREE R VALUE TEST SET SIZE (%)    : 5.000
REMARK 3 FREE R VALUE TEST SET COUNT      : 4732
REMARK 3
REMARK 3 FIT TO DATA USED IN REFINEMENT (IN BINS).
REMARK 3   BIN  RESOLUTION RANGE  COMPL.    NWORK  NFREE    RWORK  RFREE
REMARK 3     1  39.0655 - 5.2471    0.99    3219   169  0.1748 0.1893
REMARK 3     2   5.2471 - 4.1663    1.00    3088   162  0.1223 0.1526
```

O arquivo .pdb (Resíduos faltantes)

REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)

REMARK 465

REMARK 465 M RES C SSSEQI

REMARK 465 LEU S 1

REMARK 465 MET S 2

REMARK 465 GLY S 3

REMARK 465 MET L 1

REMARK 465 SER L 2

REMARK 465 GLY L 3

REMARK 465 CYS L 4

REMARK 465 ARG L 5

REMARK 465 ALA L 6

REMARK 465 GLN L 7

REMARK 465 ASN L 8

REMARK 465 ALA L 9

REMARK 465 PRO L 10

REMARK 465 GLY L 11

REMARK 465 GLY L 12

REMARK 465 ILE L 13

REMARK 465 PRO L 14

REMARK 465 VAL L 15

REMARK 465 THR L 16

REMARK 465 PRO L 17

REMARK 465 LYS L 18

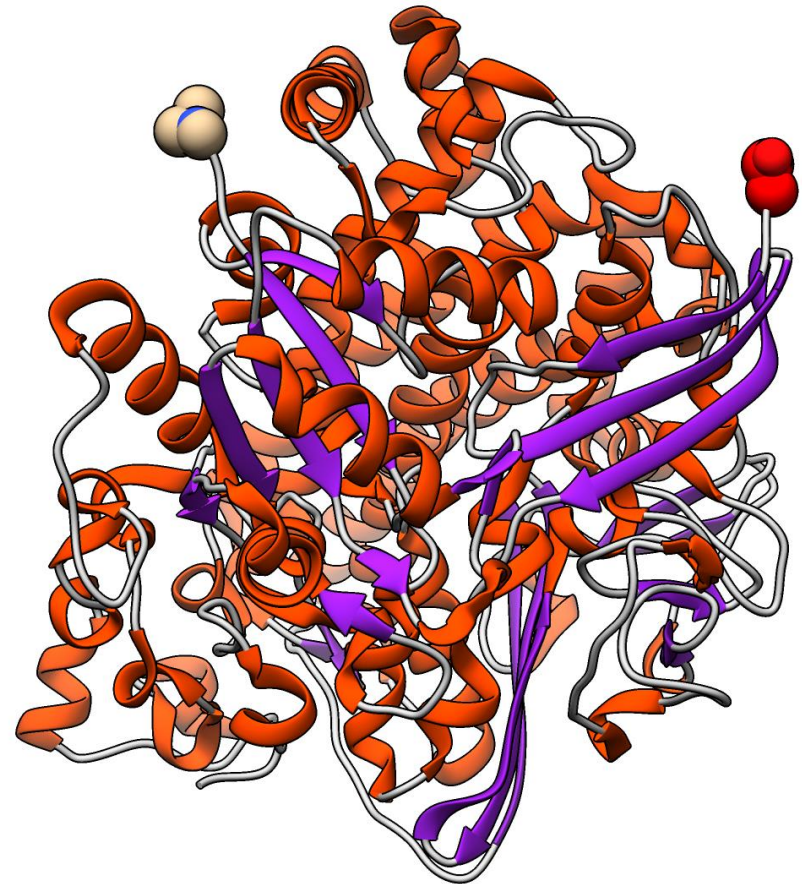
REMARK 465 SER L 19

REMARK 500

REMARK 500 GEOMETRY AND STEREOCHEMISTRY

REMARK 500 SUBTOPIC: TORSION ANGLES

REMARK 500



O arquivo .pdb (Ocupância)

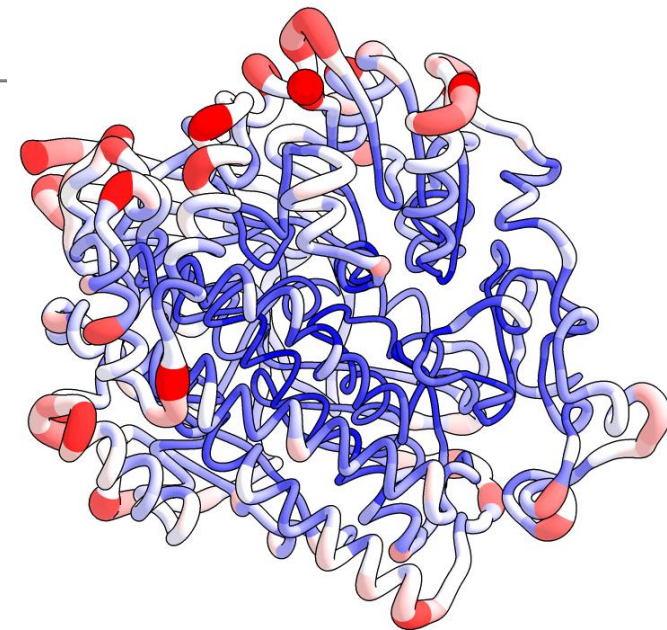
- Cristais são compostos de **muitas moléculas individuais** embaladas em um arranjo simétrico. Em alguns cristais, existem **pequenas diferenças** entre cada uma dessas moléculas.
- Quando os pesquisadores constroem o modelo atômico, eles podem usar a **ocupância** para estimar a **quantidade de cada conformação** que é observada no cristal.
- Para a maioria dos átomos, a **ocupação recebe um valor de 1**, indicando que o átomo é encontrado em todas as moléculas no mesmo lugar no cristal.

ATOM	6225	CB	VAL	L	551	-55.237	115.626	18.270	1.00	7.98	C
ATOM	6226	CG1	VAL	L	551	-55.144	117.164	18.223	1.00	11.07	C
ATOM	6227	CG2	VAL	L	551	-56.325	115.195	19.279	1.00	8.60	C
ATOM	6228	N	HIS	L	552	-53.296	114.936	15.939	1.00	8.23	N
ATOM	6229	CA	HIS	L	552	-52.126	115.438	15.184	1.00	9.08	C
ATOM	6230	C	HIS	L	552	-51.912	116.952	15.368	1.00	11.69	C
ATOM	6231	O	HIS	L	552	-52.016	117.403	16.504	1.00	9.01	O
ATOM	6232	CB	HIS	L	552	-52.135	115.015	13.689	1.00	8.21	C
ATOM	6233	CG	HIS	L	552	-53.489	115.043	13.042	1.00	10.81	C
ATOM	6234	ND1	HIS	L	552	-54.435	114.057	13.241	1.00	9.20	N
ATOM	6235	CD2	HIS	L	552	-54.055	115.942	12.203	1.00	9.80	C
ATOM	6236	CE1	HIS	L	552	-55.535	114.362	12.572	1.00	10.05	C
ATOM	6237	NE2	HIS	L	552	-55.333	115.506	11.940	1.00	8.57	N
ATOM	6238	OXT	HIS	L	552	-51.658	117.739	14.438	1.00	10.44	O
TER	6239		HIS	L	552						

O arquivo .pdb (B-factor)

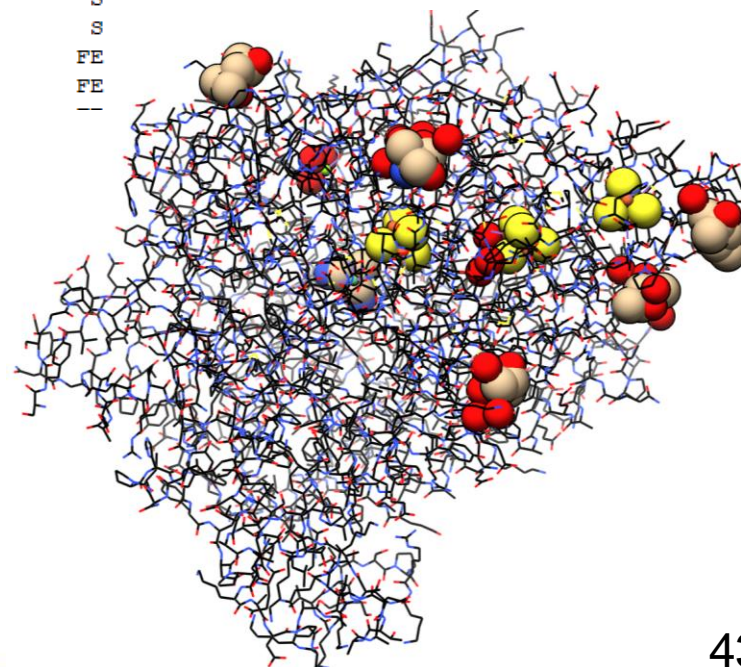
- **Movimentos** são incorporados ao modelo atômico por valores de **B-factor** ou **fator de temperatura**.
- Valores **abaixo de 10** indicam que o átomo **não se move** muito e está na **mesma posição** em todas as moléculas do cristal.
- Valores **acima de 50** indicam que o átomo **se move tanto** que mal pode ser visto. Este é frequentemente o caso de átomos na **superfície das proteínas**, onde longas cadeias laterais são livres para se movimentar na água.

ATOM	6225	CB	VAL	L	551	-55.237	115.626	18.270	1.00	7.98	C
ATOM	6226	CG1	VAL	L	551	-55.144	117.164	18.223	1.00	11.07	C
ATOM	6227	CG2	VAL	L	551	-56.325	115.195	19.279	1.00	8.60	C
ATOM	6228	N	HIS	L	552	-53.296	114.936	15.939	1.00	8.23	N
ATOM	6229	CA	HIS	L	552	-52.126	115.438	15.184	1.00	9.08	C
ATOM	6230	C	HIS	L	552	-51.912	116.952	15.368	1.00	11.69	C
ATOM	6231	O	HIS	L	552	-52.016	117.403	16.504	1.00	9.01	O
ATOM	6232	CB	HIS	L	552	-52.135	115.015	13.689	1.00	8.21	C
ATOM	6233	CG	HIS	L	552	-53.489	115.043	13.042	1.00	10.81	C
ATOM	6234	ND1	HIS	L	552	-54.435	114.057	13.241	1.00	9.20	N
ATOM	6235	CD2	HIS	L	552	-54.055	115.942	12.203	1.00	9.80	C
ATOM	6236	CE1	HIS	L	552	-55.535	114.362	12.572	1.00	10.05	C
ATOM	6237	NE2	HIS	L	552	-55.333	115.506	11.940	1.00	8.57	N
ATOM	6238	OXT	HIS	L	552	-51.658	117.739	14.438	1.00	10.44	O
TER	6239		HIS	L	552						



O arquivo .pdb (HETATM)

```
ATOM 6238 OXT HIS L 552 -51.658 117.739 14.438 1.00 10.44 O
TER 6239 HIS L 552
HETATM 6240 FE1 SF4 S1001 -42.133 110.858 18.656 1.00 10.41 FE
HETATM 6241 FE2 SF4 S1001 -40.926 113.030 19.576 1.00 10.27 FE
HETATM 6242 FE3 SF4 S1001 -39.781 111.618 17.669 1.00 10.57 FE
HETATM 6243 FE4 SF4 S1001 -39.961 110.567 20.207 1.00 10.28 FE
HETATM 6244 S1 SF4 S1001 -38.784 112.330 19.552 1.00 8.09 S
HETATM 6245 S2 SF4 S1001 -40.350 109.575 18.240 1.00 9.70 S
HETATM 6246 S3 SF4 S1001 -41.907 111.493 20.786 1.00 8.80 S
HETATM 6247 S4 SF4 S1001 -41.738 112.711 17.547 1.00 8.96 S
HETATM 6248 FE1 SF4 S1002 -22.033 110.222 4.742 1.00 12.01 FE
HETATM 6249 FE2 SF4 S1002 -24.624 109.844 4.196 1.00 12.35 FE
HETATM 6250 FE3 SF4 S1002 -22.972 107.754 4.427 1.00 12.14 FE
HETATM 6251 FE4 SF4 S1002 -22.735 109.394 2.245 1.00 13.27 FE
HETATM 6252 S1 SF4 S1002 -24.442 108.101 2.836 1.00 11.05 S
HETATM 6253 S2 SF4 S1002 -21.113 108.603 3.532 1.00 12.47 S
HETATM 6254 S3 SF4 S1002 -23.259 111.331 3.268 1.00 11.37 S
HETATM 6255 S4 SF4 S1002 -23.548 109.211 6.002 1.00 11.60 S
HETATM 6256 FE1 F3S S1003 -30.051 109.735 10.927 1.00 10.90 FE
HETATM 6257 FE3 F3S S1003 -29.852 107.856 12.825 1.00 10.83 FE
-----
```

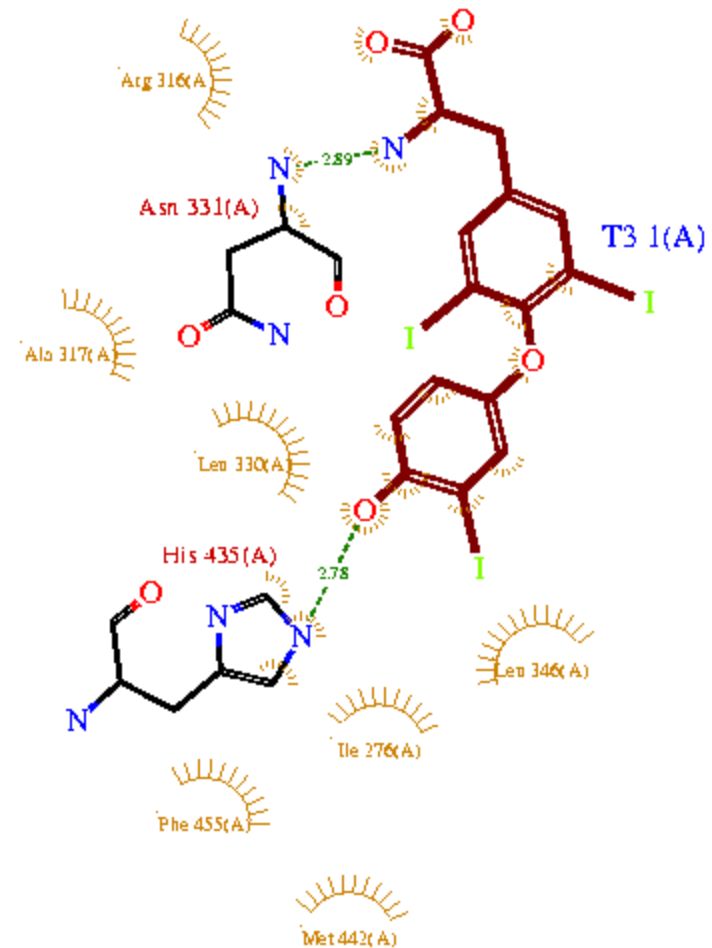


Outras bases de dados de proteínas

3D structure databases	DisProt	Database of Protein Disorder	http://www.disprot.org/
	MobiDB	Database of intrinsically disordered and mobile proteins	http://mobidb.bio.unipd.it/
	ModBase	Database of Comparative Protein Structure Models	http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi
	PDBe ⁻	Protein Data Bank at Europe	http://www.ebi.ac.uk/pdbe/
	PDBj ⁻	Protein Data Bank at Japan	http://pdbj.org/
	PDBsum	Pictorial database of 3D structures in the Protein Data Bank	http://www.ebi.ac.uk/pdbsum/
	ProteinModelPortal	Protein Model Portal of the PSI-Nature Structural Biology Knowledgebase	http://www.proteinmodelportal.org/
	SMR	Database of annotated 3D protein structure models	http://swissmodel.expasy.org/repository/

Exemplo: PDBsum

- Fornece uma série de informações úteis a partir de uma entrada no PDB (ou enviada pelo usuário).
- Na opção **ligands** é possível obter o gráfico de interações para heteroátomos como produzido pelo programa **Ligplot**.



Como obter estruturas?

- **Determinação experimental**

- Cristalografia

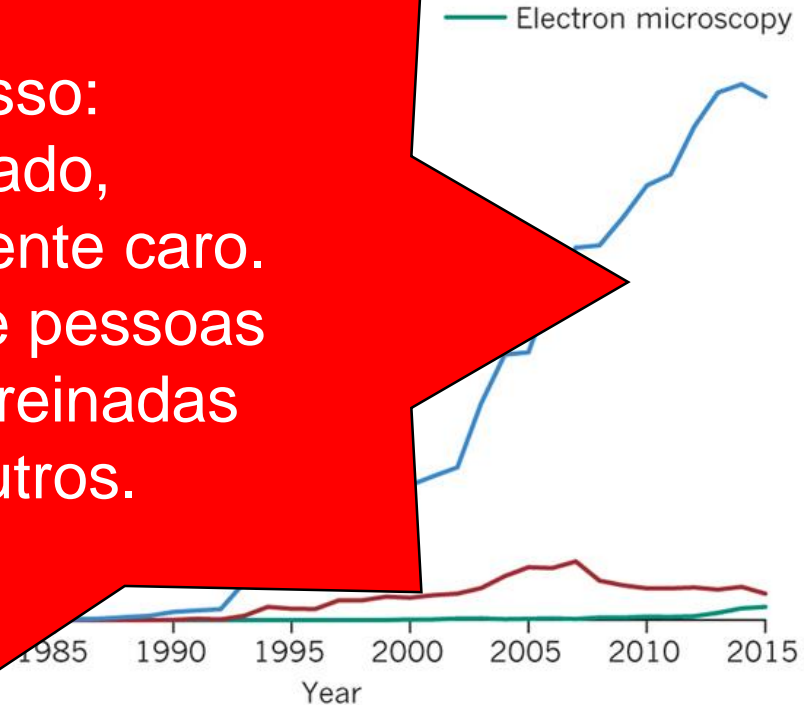
- RMN

- Cri

- eletrônica (C

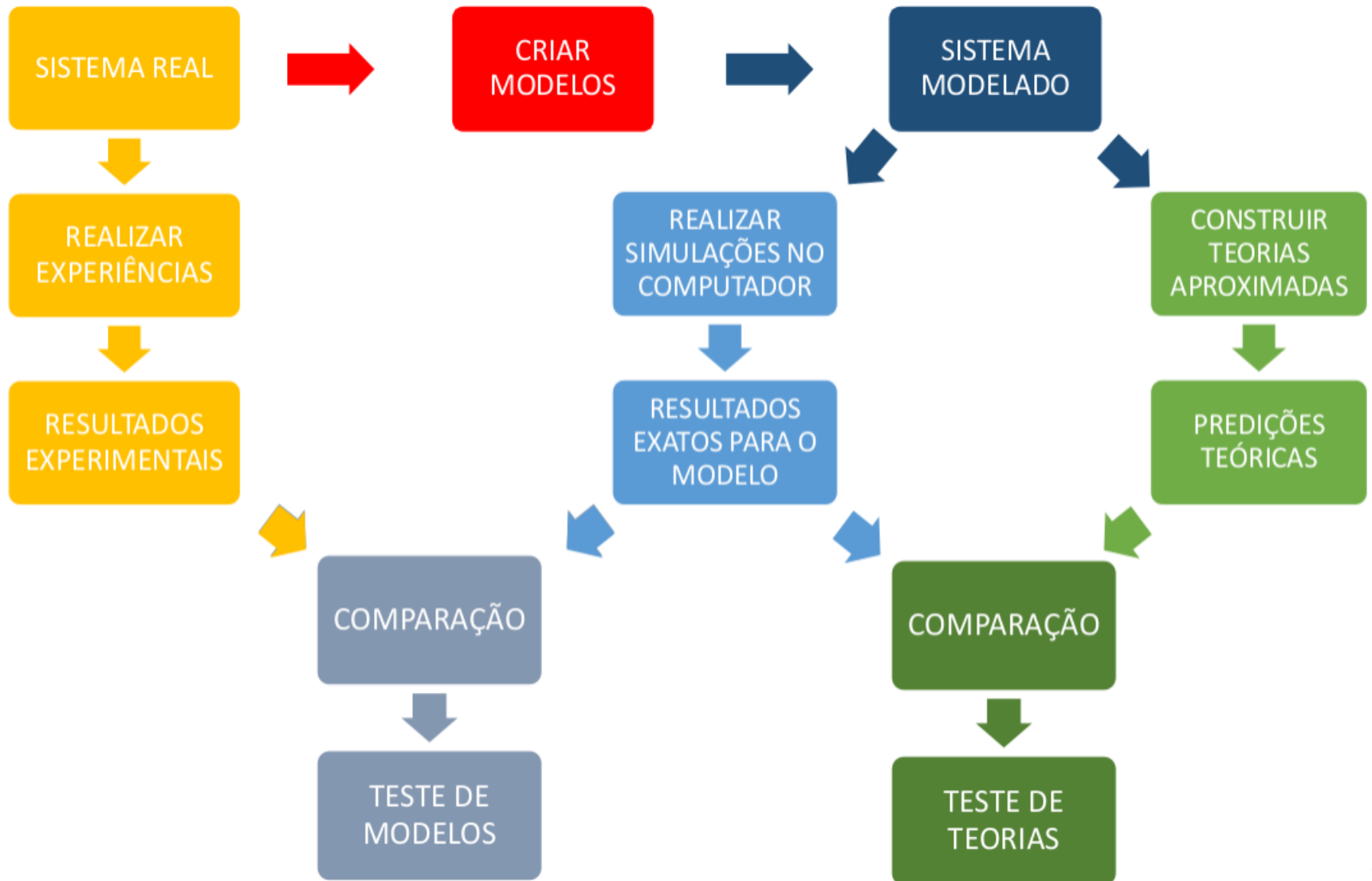
- entre outras

- Processo:
- demorado,
- financeiramente caro.
- necessita de pessoas altamente treinadas
- entre outros.



- **Modelagem**

De onde parte a modelagem?



Que problemas podemos resolver a partir da modelagem?

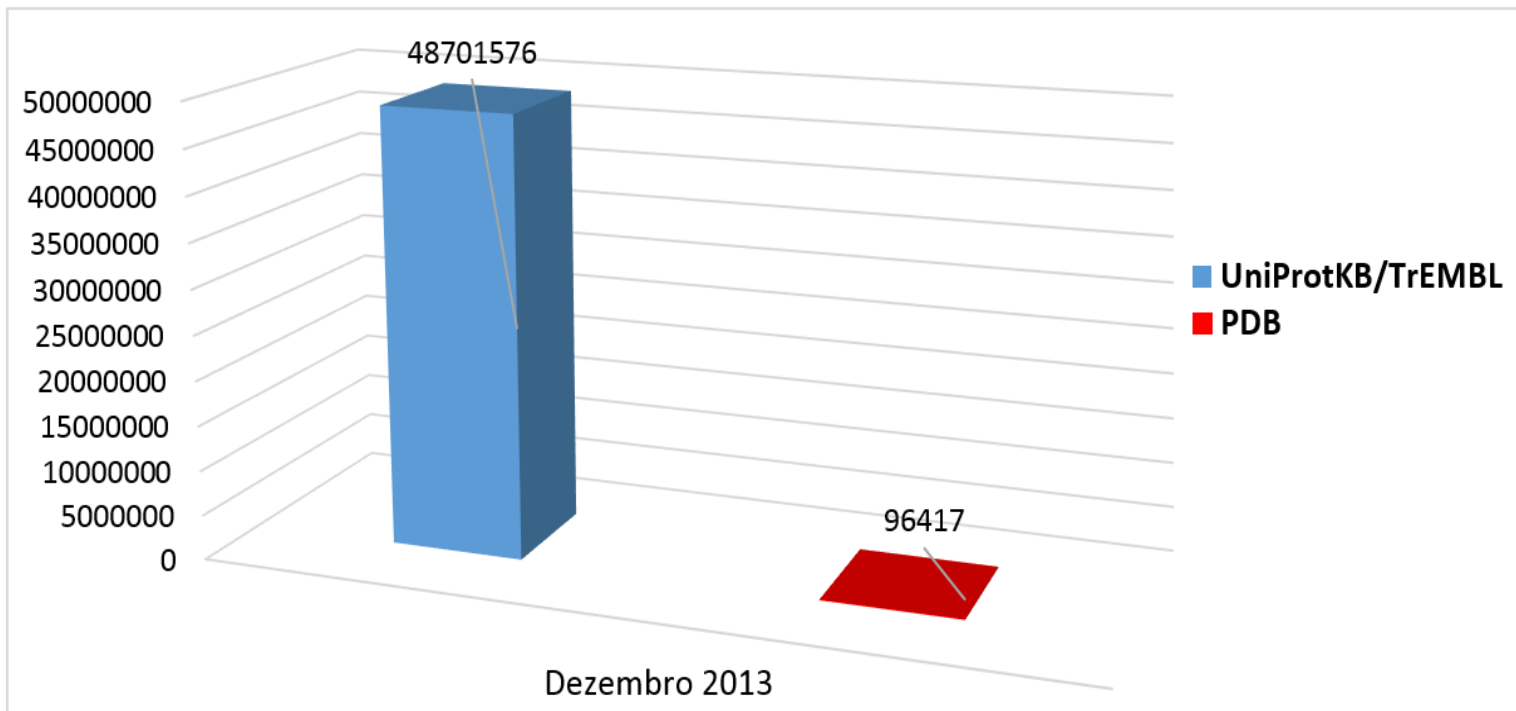
- Predição de estruturas de proteínas.
- Predição de complexos proteína-ligante.
- Descoberta de novos ligantes proteicos.
- Análise da evolução temporal e das conformações assumidas por um sistema.

Que problemas podemos resolver a partir da modelagem?

- **Predição de estruturas de proteínas.**
- Predição de complexos proteína-ligante.
- Descoberta de novos ligantes proteicos.
- Análise da evolução temporal e das conformações assumidas por um sistema.

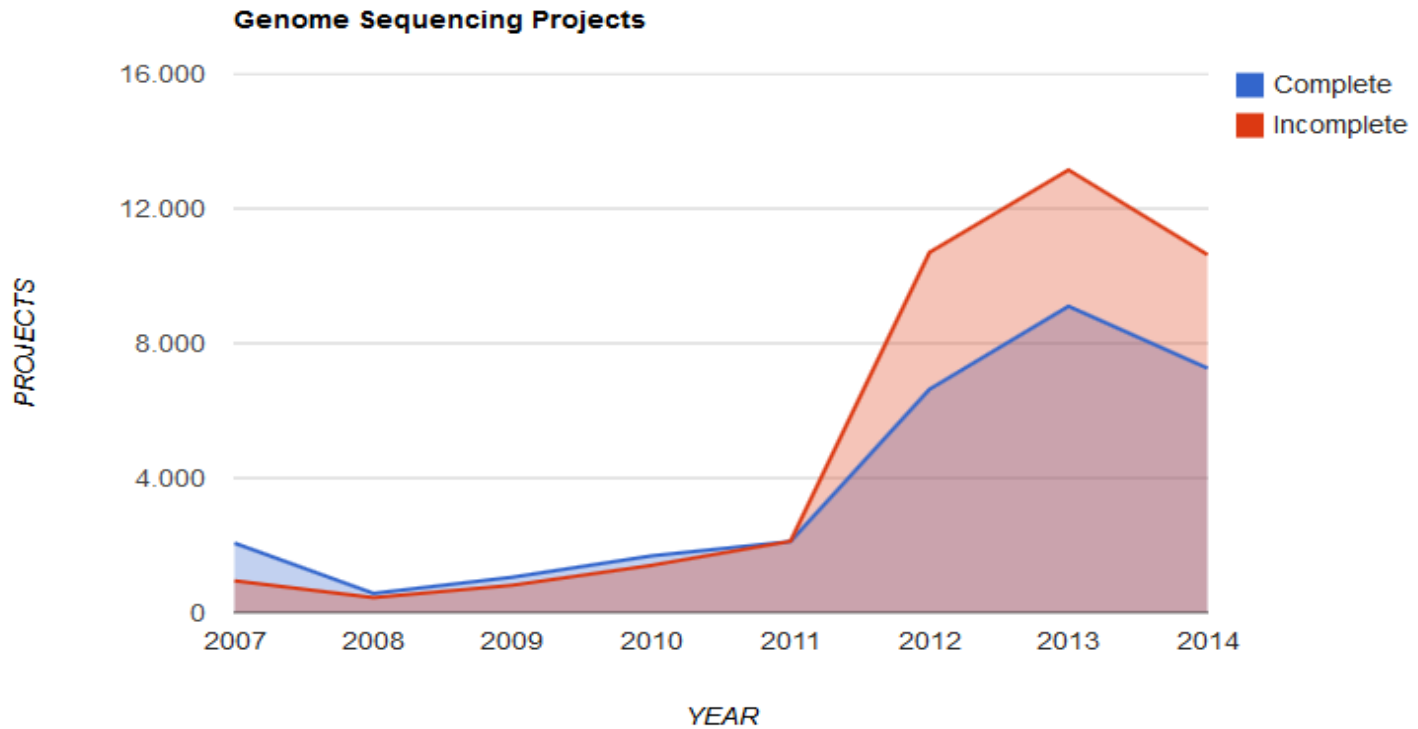
Por que prever estruturas?

Não é possível acompanhar a taxa de crescimento das bases de dados de sequência:



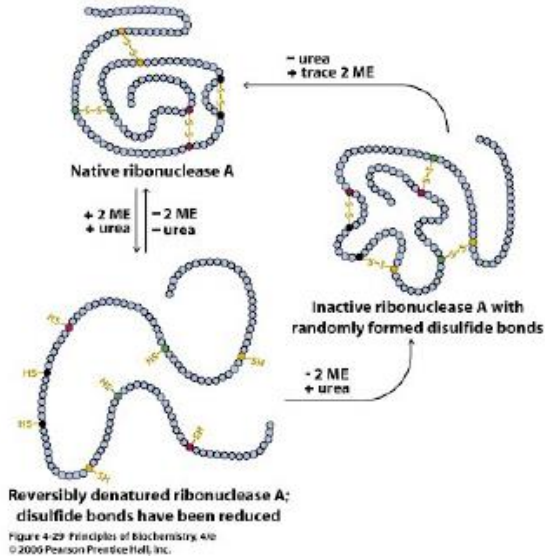
O que nos leva a prever estruturas?

Aumento do número de projetos de sequenciamento de genomas
+
Limitações da predição experimental de estruturas proteicas
=
Aumento das técnicas de predição teórica

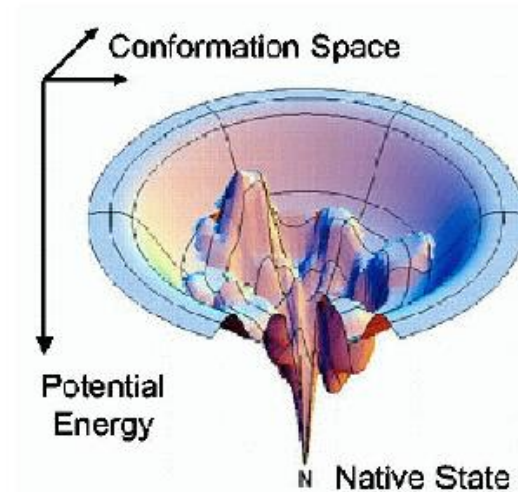


O que possibilita a predição?

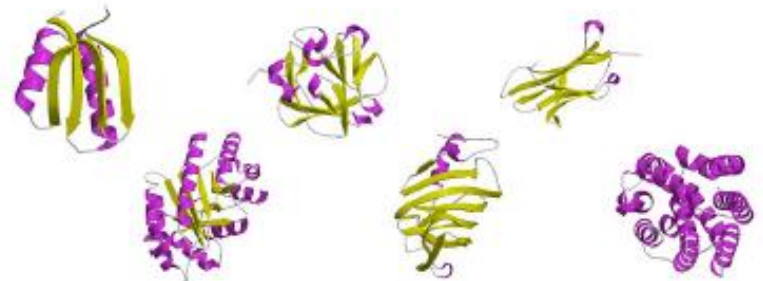
A informação para o enovelamento está contida na sequência da proteína



A estrutura enovelada pode ser distinguida por sua energia potencial



Centenas de enovelamentos conhecidos



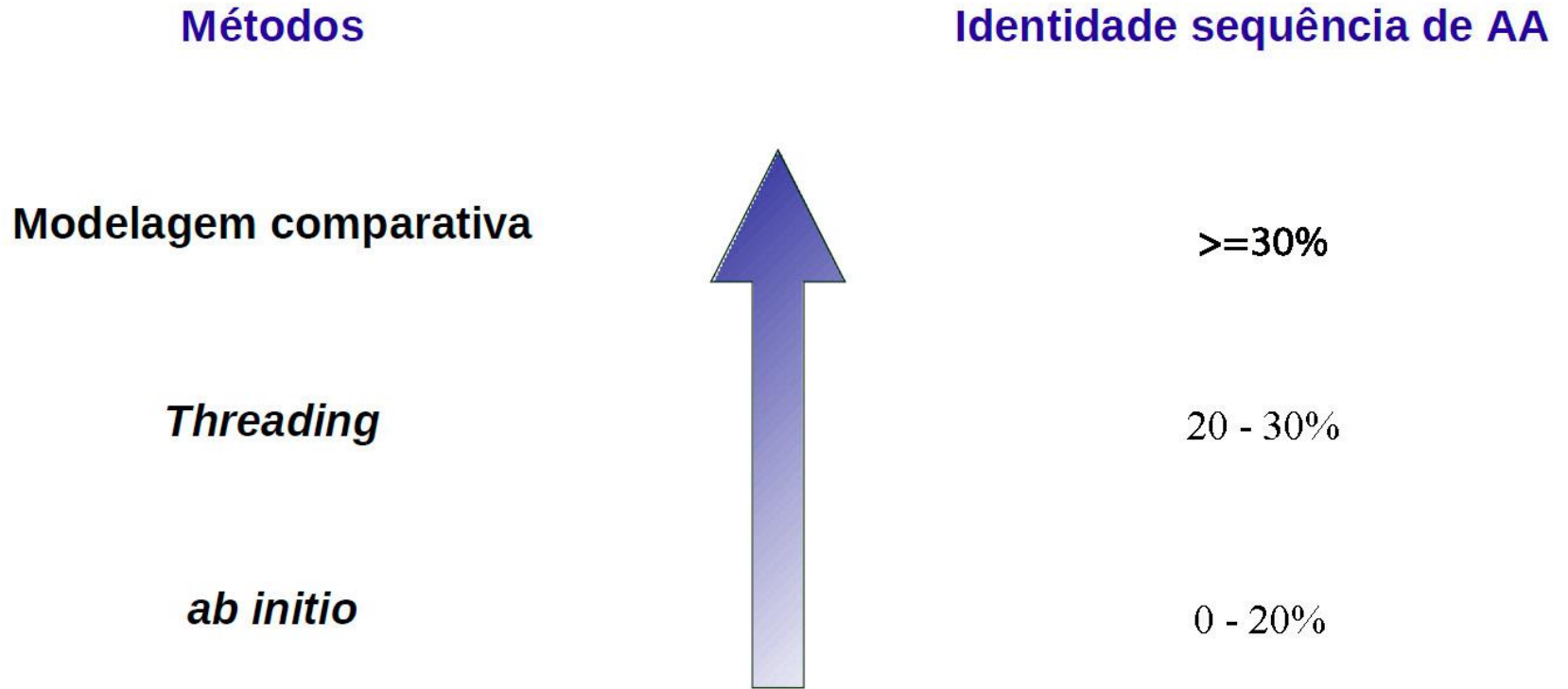
Milhares de estruturas disponíveis no PDB

RCSB **PDB**
PROTEIN DATA BANK

Estratégias de predição de estrutura

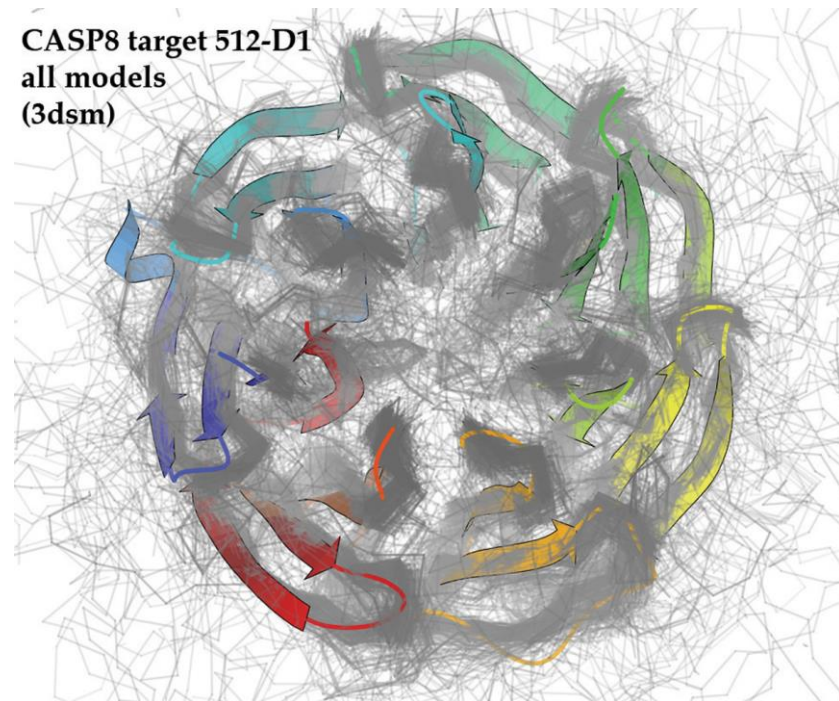
Métodos <i>dependentes</i> de molde	Modelagem comparativa (ou por homologia)
	<i>Threading</i>
Métodos <i>independentes</i> de molde	<i>Predição de novo</i>
	<i>Ab initio</i>

Limites dos Métodos de Predição de Estrutura 3D



Competição CASP

- **Critical Assessment of protein Structure Prediction**
- Ajudar a avançar os métodos de identificação de estrutura tridimensional de proteína a partir de sua sequência de aminoácidos.
- “Copa do Mundo” da predição de estrutura.



“Como”?

- Modeller
- SWISS-MODEL
- I-tasser
- RaptorX
- MOE
- Prime
- entre outros

Modeller

Program for Comparative Protein
Structure Modelling by Satisfaction
of Spatial Restraints



```
A I L V G S M P R R D G M E R K D L L K A N V K I F K C Q G A  
V E V C P V D C F Y E G P N F L V I H P D E C I D C A L C E P  
L A C K P E C P V N I Q G S - - Y A I D A D S C I D C G S  
G - - L A C G A C K P E C P V N I Q G S - - Y A I D A D S
```

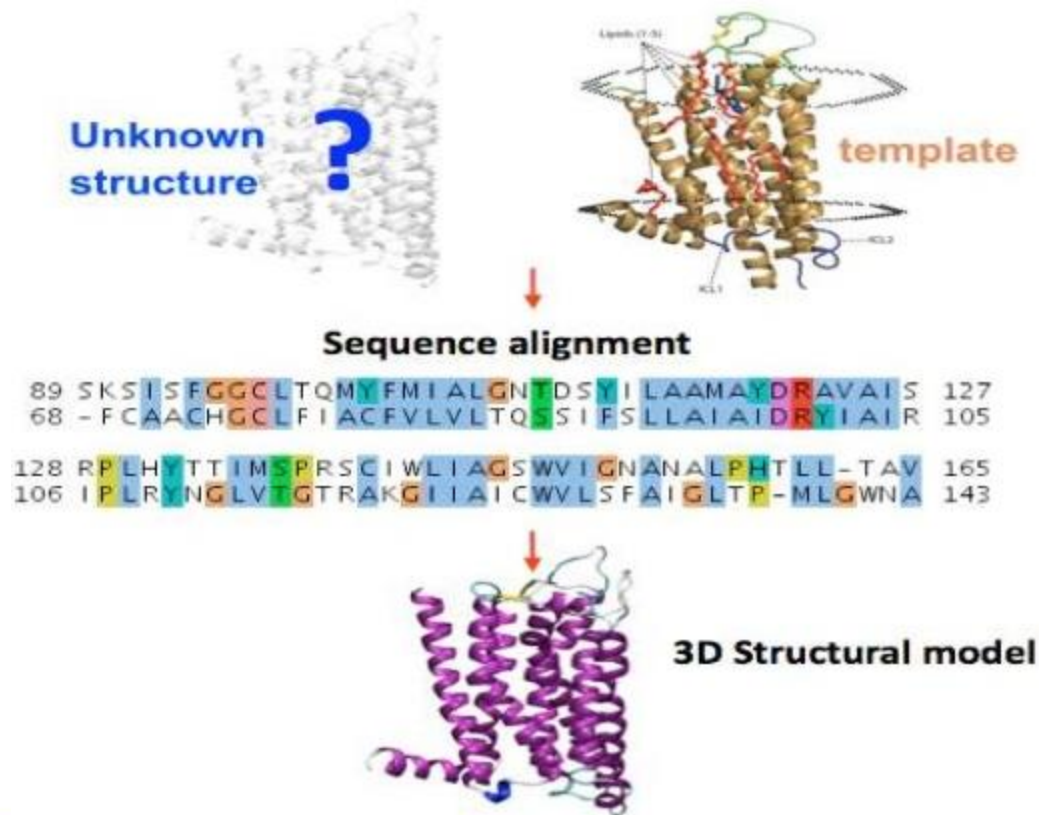


Estratégias de predição de estrutura

<i>Métodos dependentes de molde</i>	Modelagem comparativa (ou por homologia)
	<i>Threading</i>
<i>Métodos independentes de molde</i>	<i>Predição de novo</i>
	<i>Ab initio</i>

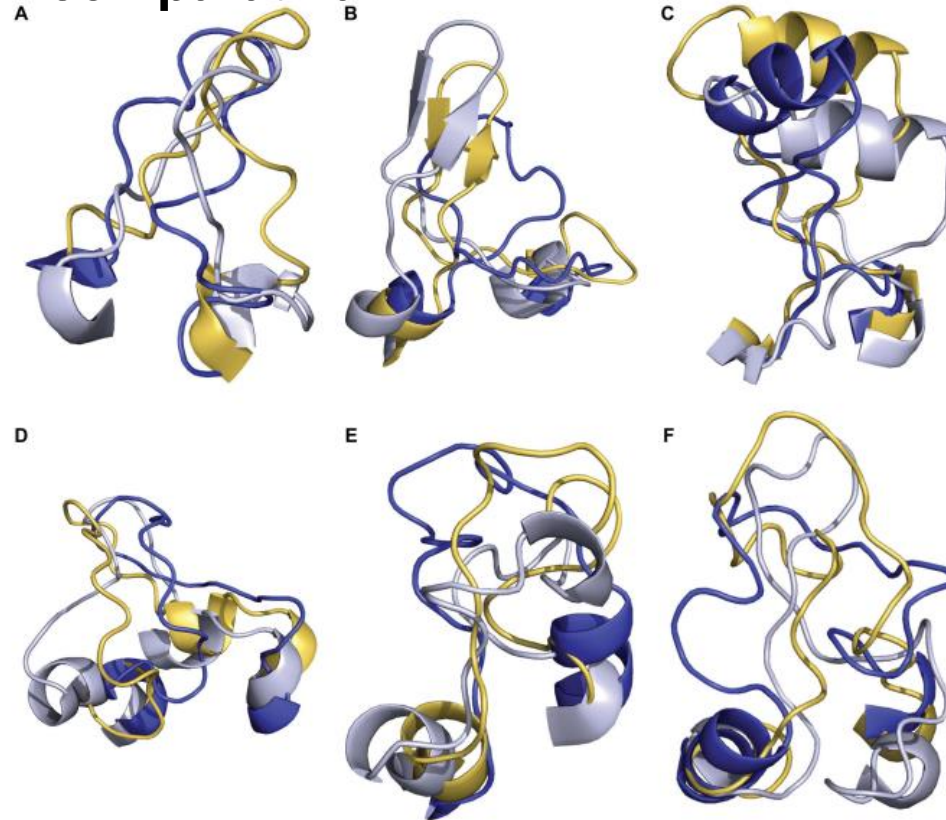
Modelagem Comparativa

- A estratégia baseia-se no conhecimento de que a **conformação estrutural** de uma proteína é **mais conservada** que sua **sequência de aminoácidos** durante o **processo evolutivo**, e que pequenas **mudanças na sequência**, em geral, resultam em, apenas, **sutis modificações na estrutura tridimensional**.



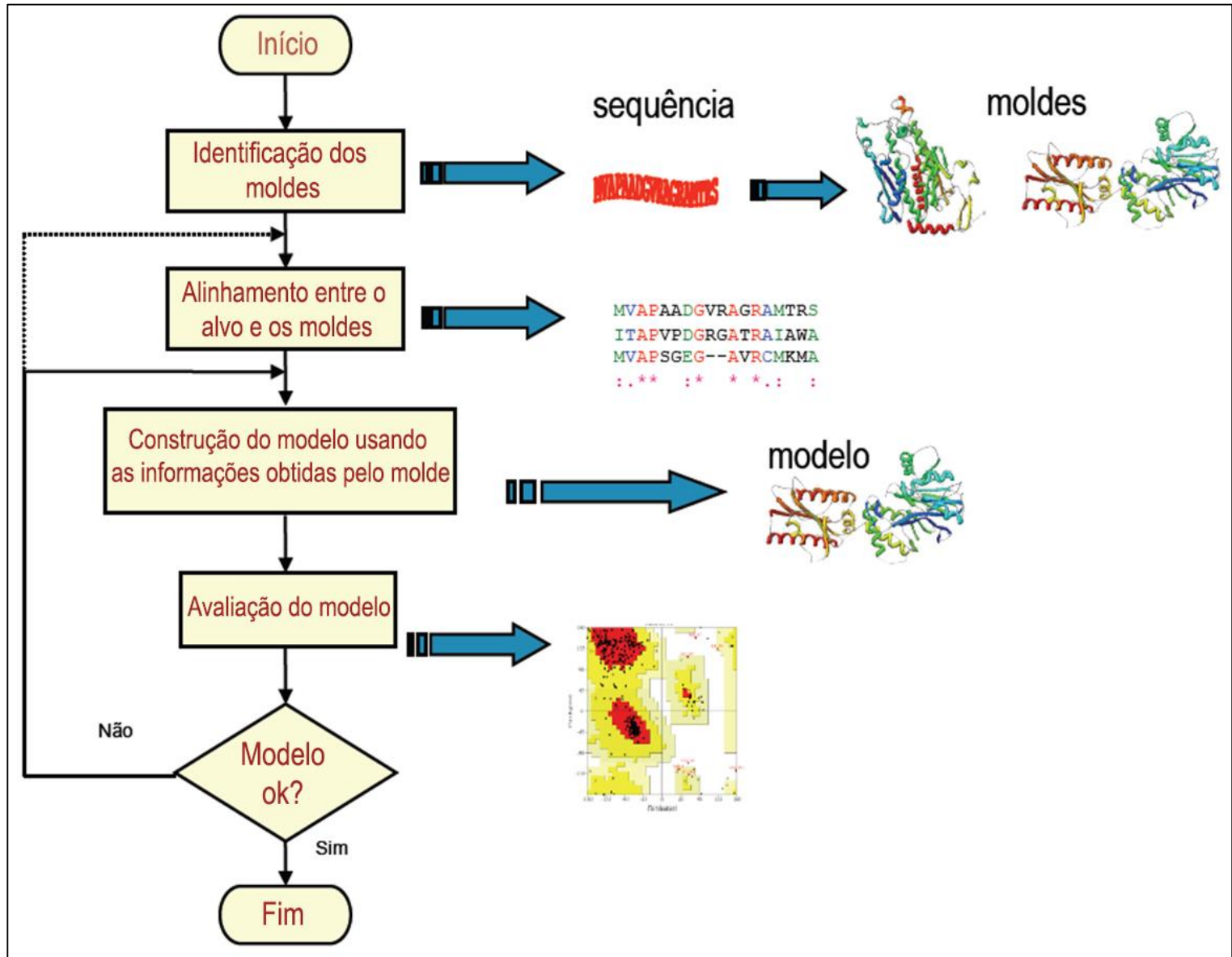
Modelagem Comparativa

- Se pelo menos uma **sequência homóloga** para qual a estrutura tridimensional resolvida esteja **disponível** é encontrada, o método de escolha para predição da estrutura tridimensional de uma proteína-alvo é a **modelagem comparativa**.



Representações estruturais da alça extracelular a partir de modelos comparativos comparados a estruturas experimentais. A estrutura experimental está em cinza, o modelo mais preciso amostrado está em amarelo e o modelo melhor ranqueado está em azul. 59

Modelagem Comparativa



Que problemas podemos resolver a partir da modelagem?

- Predição de estruturas de proteínas.
- **Predição de complexos proteína-ligante.**
- Descoberta de novos ligantes proteicos.
- Análise da evolução temporal e das conformações assumidas por um sistema.

Complexo - Proteína-ligante?

- Proteínas são os principais alvos de pesquisas farmacêuticas.
- Há grande interesse no desenvolvimento de **moléculas moduladoras da atividade** de uma proteína alvo.



- Essa “**ligação**” ou “**encaixe**” entre duas moléculas pode ser simulada através de técnicas computacionais como o atracamento molecular.

Como simular essa “ligação”?

Atracamento molecular (*molecular docking*)

- Predição dos modos de ligação de um ligante ativo conhecido;
- Identificação de novos ligantes usando *virtual screening*;
- Predição das afinidades de ligação de compostos relacionados de uma série ativa conhecida.

Contribuições:

- Detalhes moleculares da interação receptor-ligante;
- Determinar as poses de ligação do ligante;
- Importante na descoberta de novos fármacos

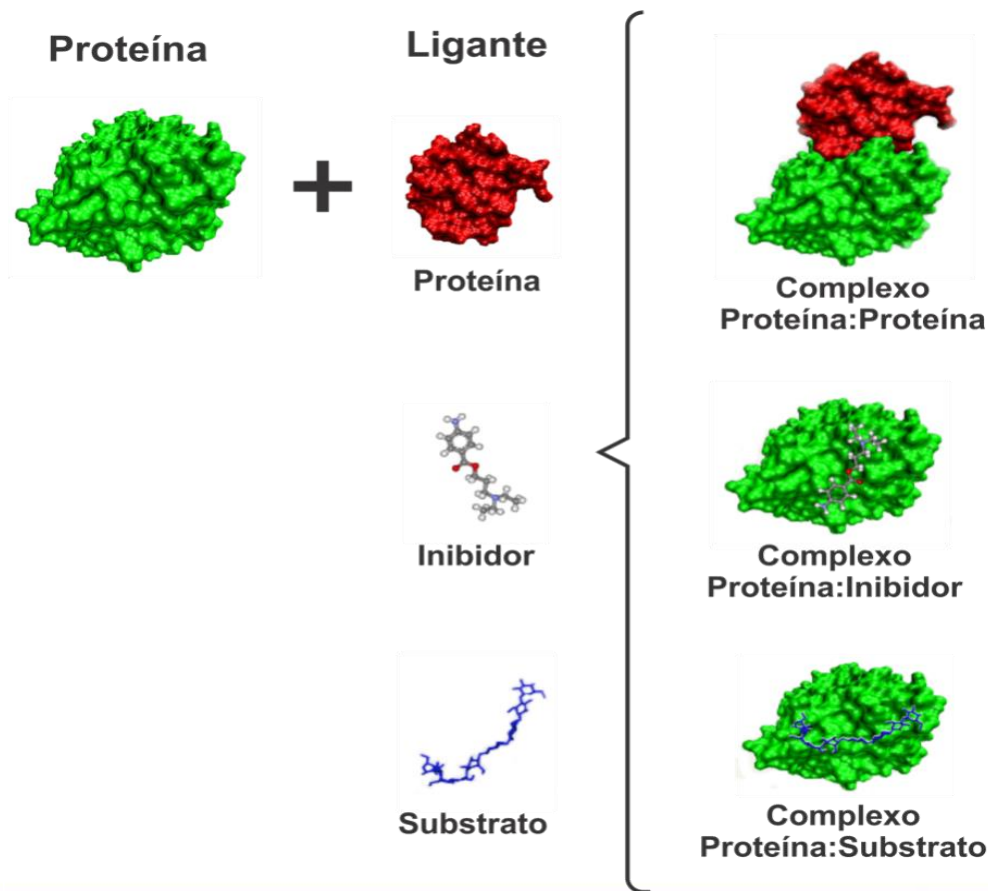
O que é atracamento molecular?

Atracamento (*Docking*) molecular é um **método físico-químico computacional** utilizado para **predizer a orientação preferencial** de uma **molécula** em relação a uma segunda a qual se **pretende ligar**, formando um possível **complexo químico estável**.

Receptor?

...

Ligante?



O que é o atracamento?



Meu pai



Foto ilustrativa



O que é o atracamento molecular?

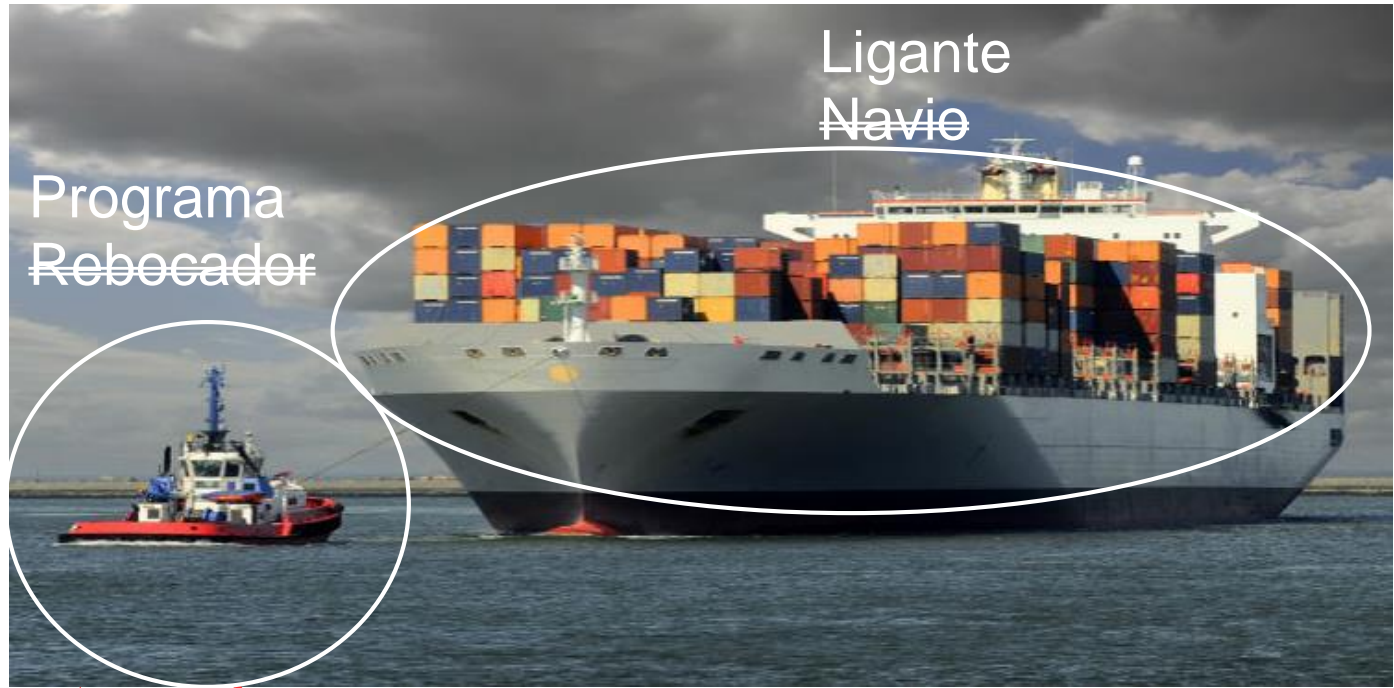
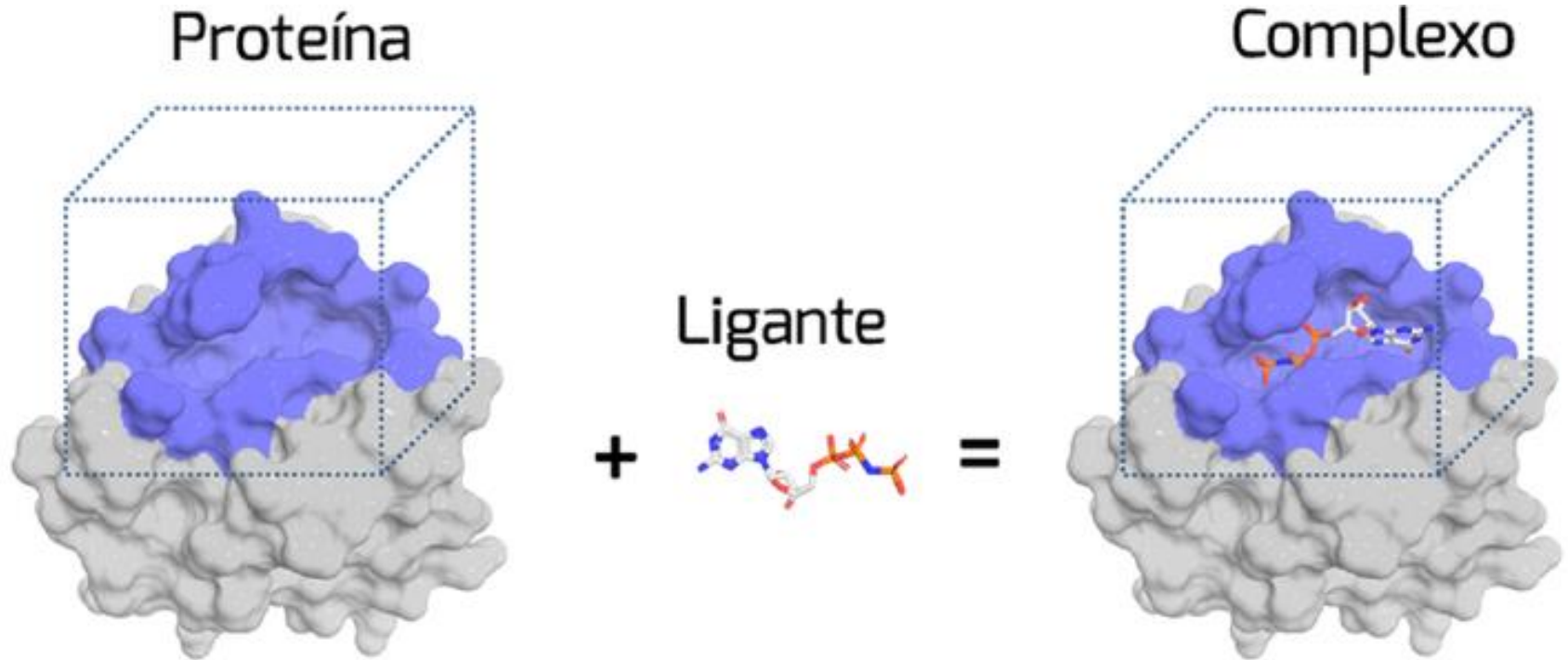
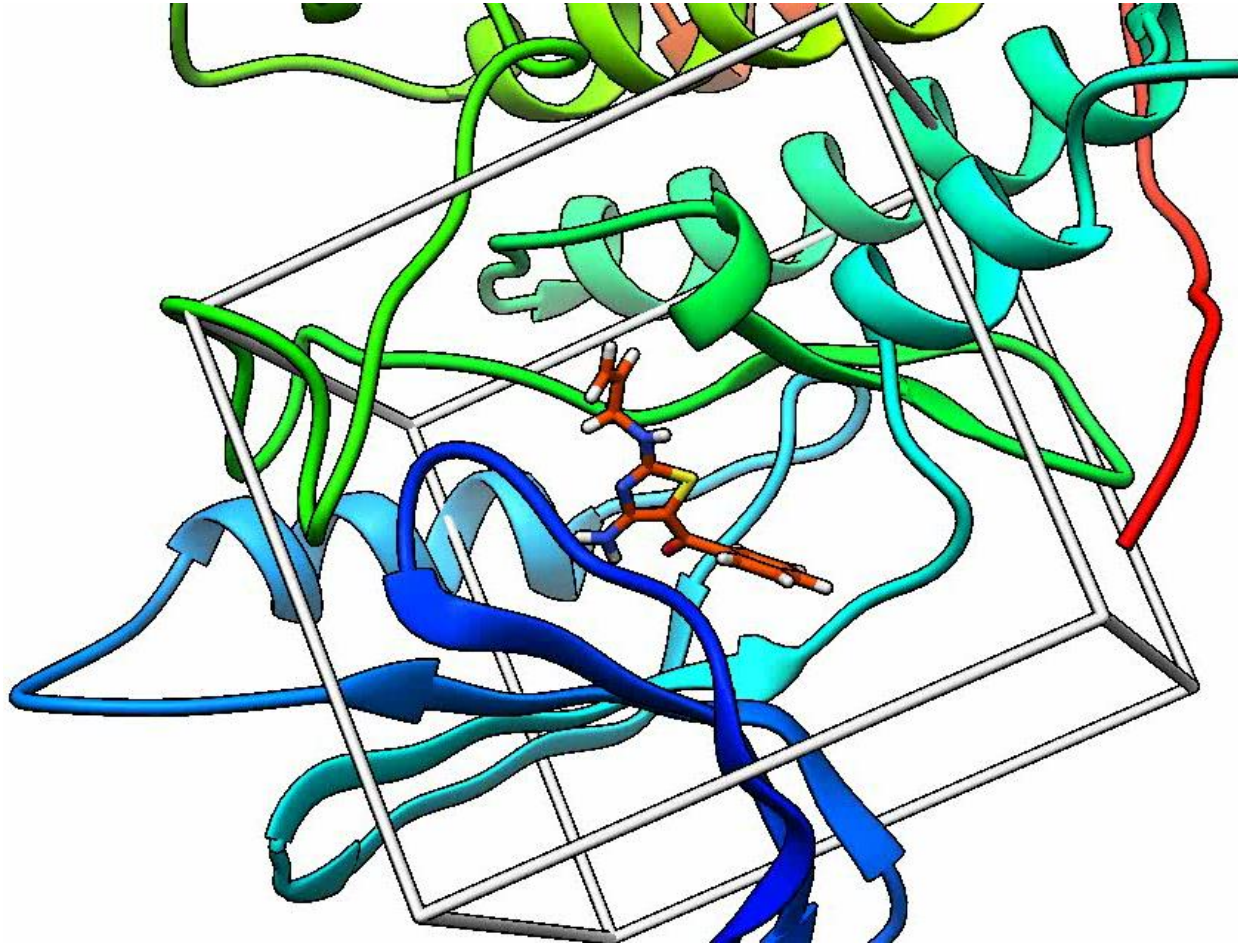


Foto ilustrativa

No computador ...

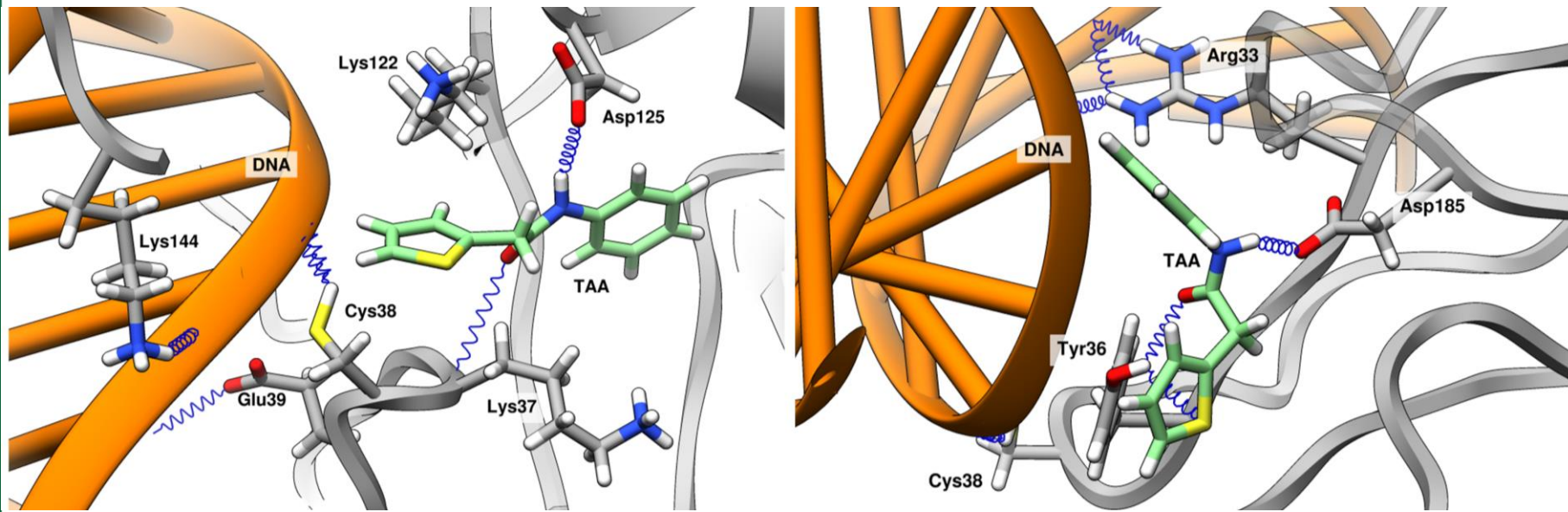


Depois de escolhida particularidades...



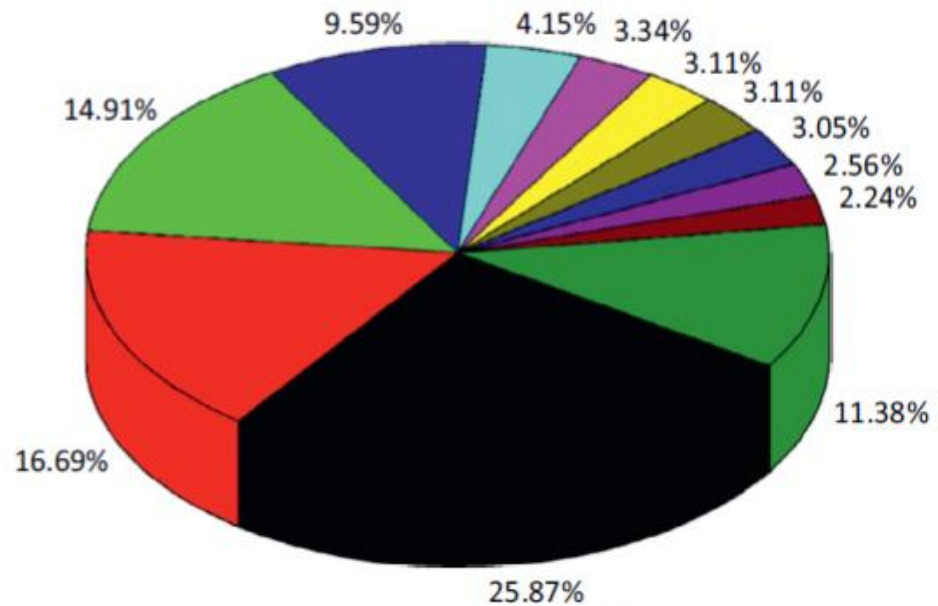
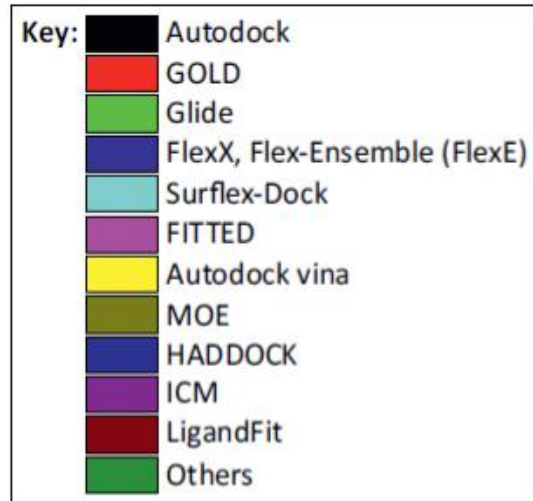
Mas no final escolhemos ...

Uma ou mais do que chamamos de poses. As poses correspondem ao **possível modo de ligação** entre o receptor e o ligante.



“Como?”

- **Programas** de computadores **específicos** permitem prever ou simular uma **possível interação** entre duas moléculas baseado nas **estruturas tridimensionais** das mesmas.



TRENDS in Pharmacological Sciences

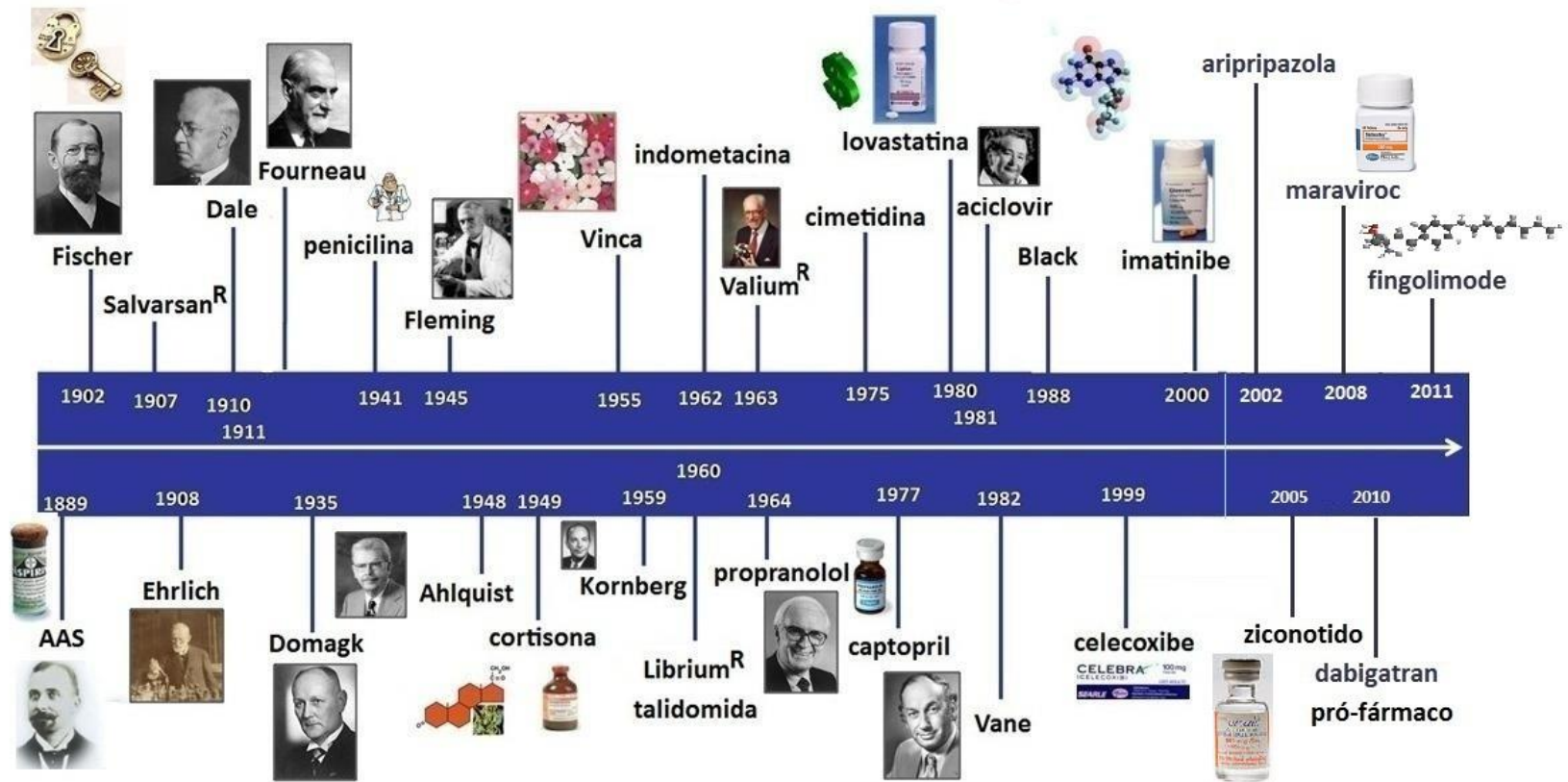
Programas mais usados em publicações de 1990 a 2013.

Que problemas podemos resolver a partir da modelagem?

- Predição de estruturas de proteínas.
- Predição de complexos proteína-ligante.
- **Descoberta de novos ligantes proteicos.**
- Análise da evolução temporal e das conformações assumidas por um sistema.

Descoberta de fármacos no passado ...

- A descoberta de novos fármacos envolvia uma abordagem de **tentativa e erro** no teste de materiais e substâncias de origem natural.
- Ou, fármacos eram descobertos **ao acaso**.



Descoberta de fármacos no passado ...

- Com a descoberta ao acaso da **penicilina** veio a **triagem** de microorganismos, resultando em um grande número de antibióticos de fontes bacterianas e fúngicas.
- Este processo de **triagem aleatório**, embora ineficiente, levou à identificação de novos compostos chaves (*lead*).
- Com o desenvolvimento exponencial da **biologia molecular** por um lado e da **informática** por outro, tornou-se possível colocar a descoberta de drogas numa base racional.

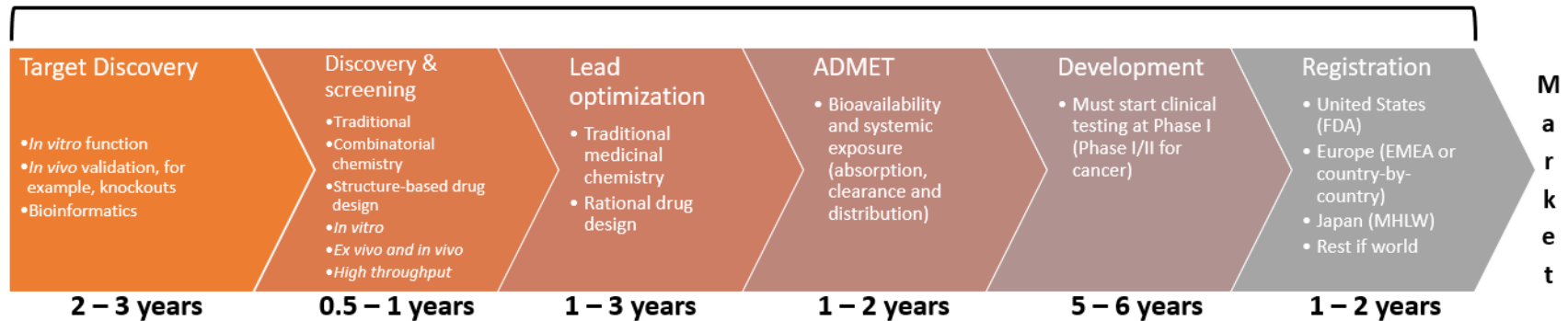


Alexander Fleming, o descobridor da penicilina.

Processo usual de descoberta

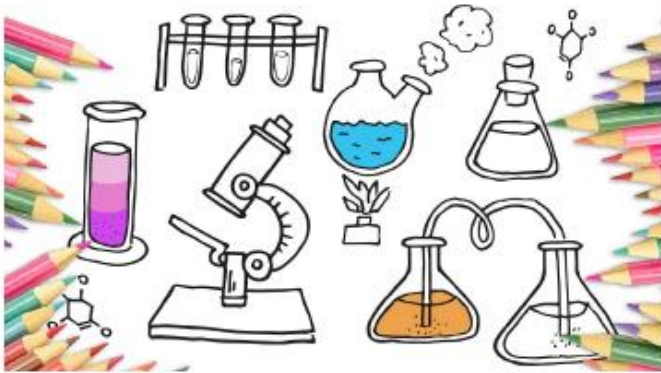
De novo drug discovery and development

- 10 – 17 years process
- < 10% overall probability of success



Processo que leva de 10-17 anos da ideia inicial ao lançamento no mercado. A probabilidade de sucesso é menor que 10%.

Qual a ideia?



- As moléculas disponíveis são submetidas a triagem utilizando ***high throughput screening (HTS)*** para decifrar a atividade biológica dos compostos.
- Embora **a triagem é rápida (100.000 compostos por dia)**, a técnica e material são caros e não costuma produzir compostos com condição de fármacos e sim identifica '**hits**' para futura otimização.

Qual a ideia?



Triagem Virtual

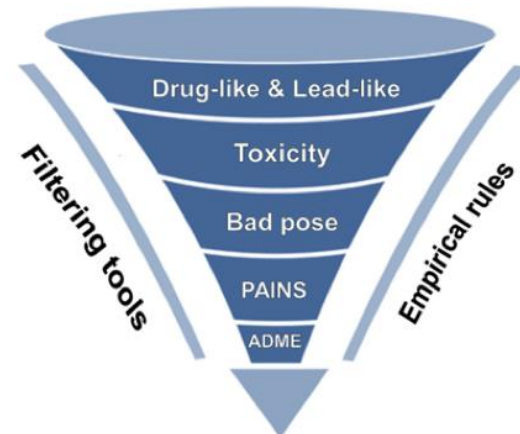
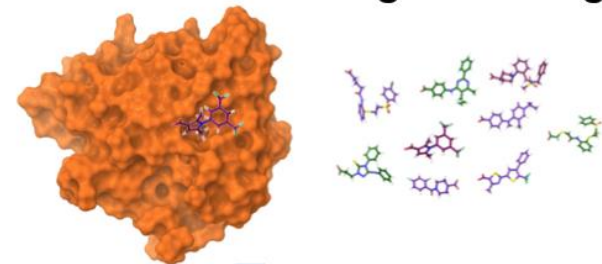
- Muitas vezes, as técnicas de triagem virtual são empregadas **em paralelo com ou em substituição** de métodos tradicionais de HTS, particularmente dentro de laboratórios acadêmicos.



Mas o que é triagem virtual?

- A triagem virtual é uma **técnica computacional** utilizada na descoberta de fármacos para **filtrar bibliotecas de moléculas pequenas** de modo a identificar os compostos que são mais susceptíveis de se **ligar a um alvo de fármaco**, tipicamente uma proteína.
- Pode ser **baseada no ligante** ou baseado na **estrutura do alvo**.

Virtual Screening & Scoring

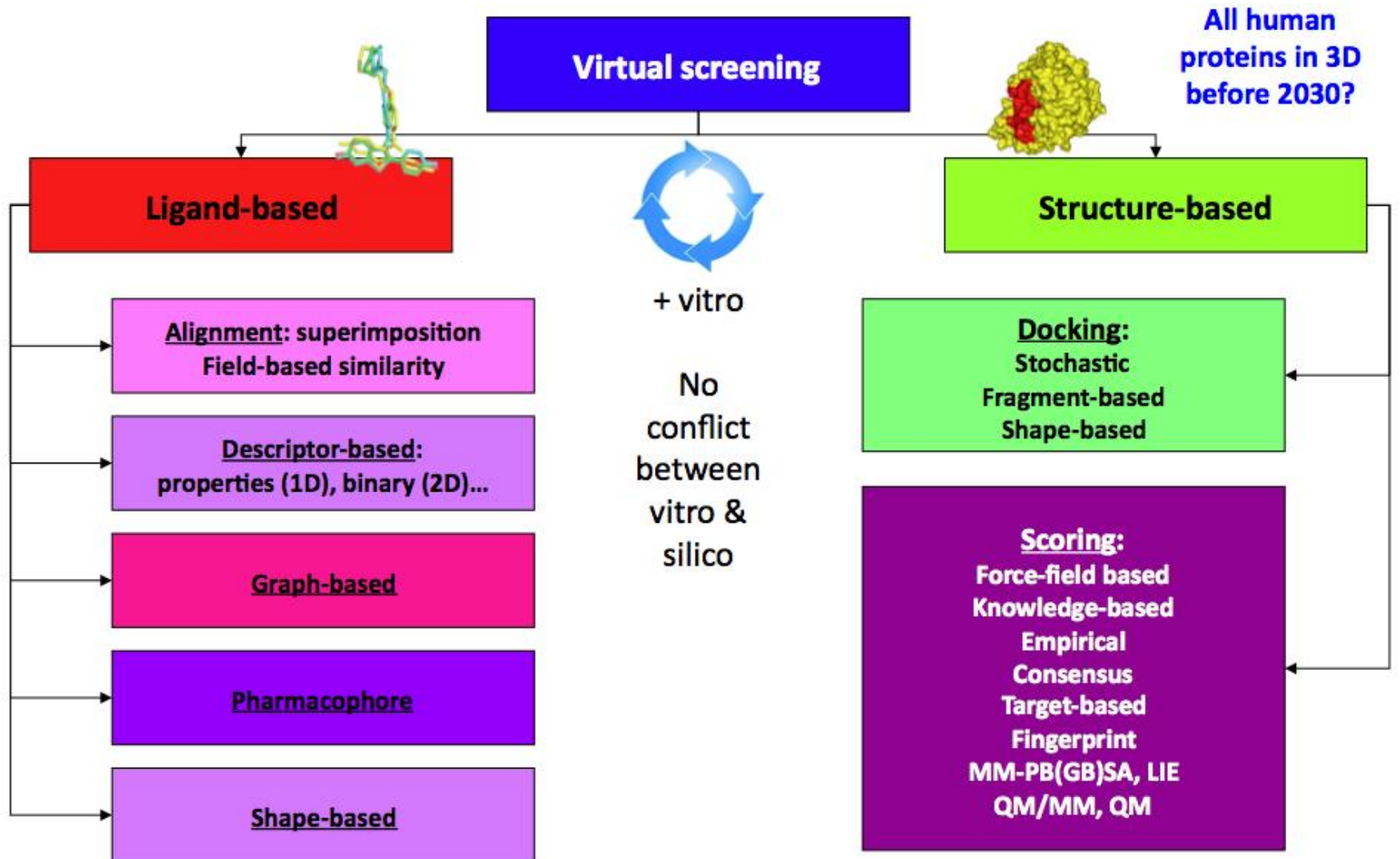


Re-ranking Virtual Screening results

Compound selection & *in vitro* assays



Mas o que é triagem virtual?

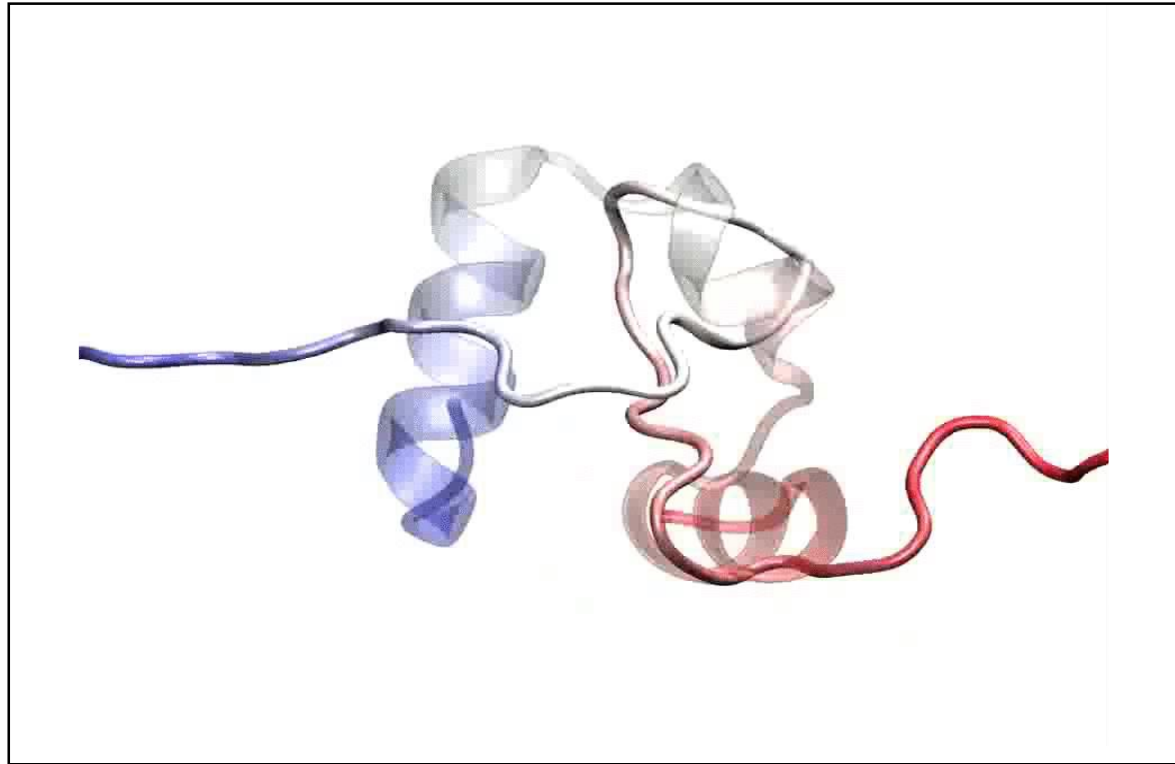


Que problemas podemos resolver a partir da modelagem?

- Predição de estruturas de proteínas.
- Predição de complexos proteína-ligante.
- Descoberta de novos ligantes proteicos.
- **Análise da evolução temporal e das conformações assumidas por um sistema.**

Chamado de dinâmica molecular

A dinâmica molecular é um **procedimento de simulação** que consiste na **computação do movimento dos átomos** em uma molécula ou de átomos individuais ou moléculas em sólidos, líquidos e gases, de acordo com as leis de movimento de Newton.

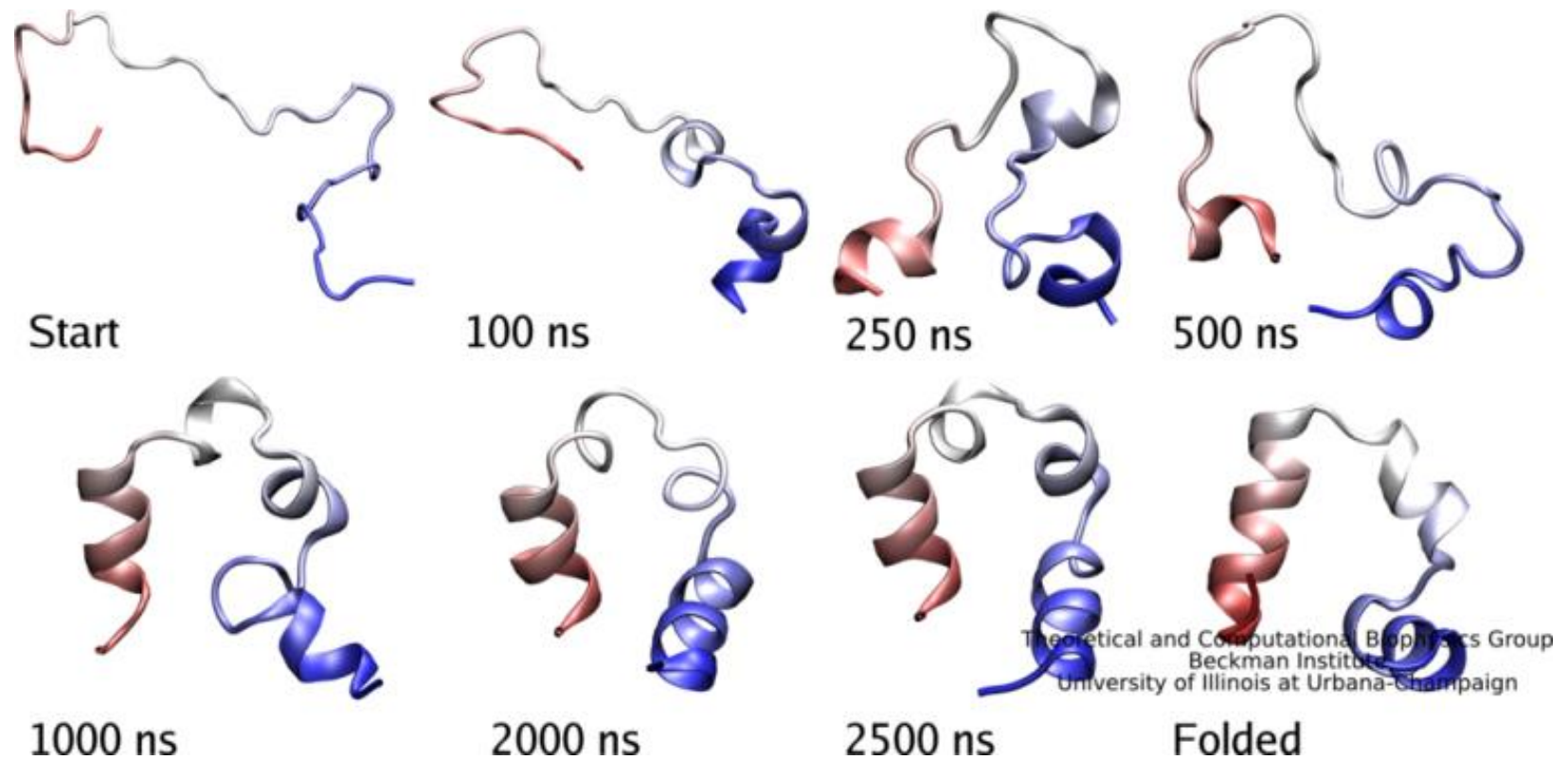


Seis microssegundos de simulação de enovelamento.

A descrição conformacional oferecida pela DM, para uma determinada molécula ou conjunto de moléculas, baseia-se na solução da 2ª Lei de Newton

$$F_i(t) = m_i a_i$$

onde F_i é a força que atua sobre cada partícula do sistema em um instante de tempo t , e a_i é a aceleração do átomo i de massa m_i .



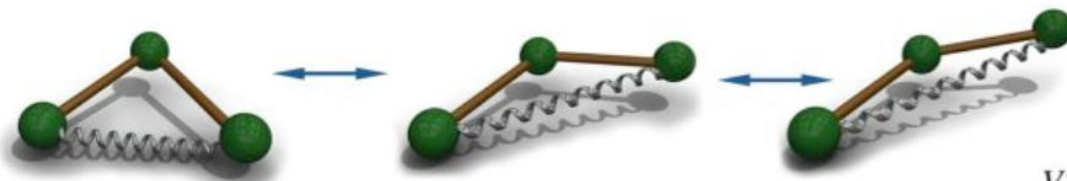
Campos de Força

Um conjunto de funções e parametrização usadas em cálculos de mecânica molecular



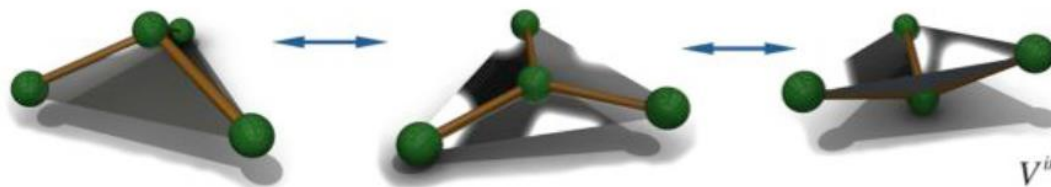
estiramento de ligações químicas

$$V^{\text{ligação}}(\vec{r}(t)) = \frac{1}{4} \sum_{n=1}^{N_b} K_{b_n} [b_n^2(t) - b_{0_n}^2]^2$$



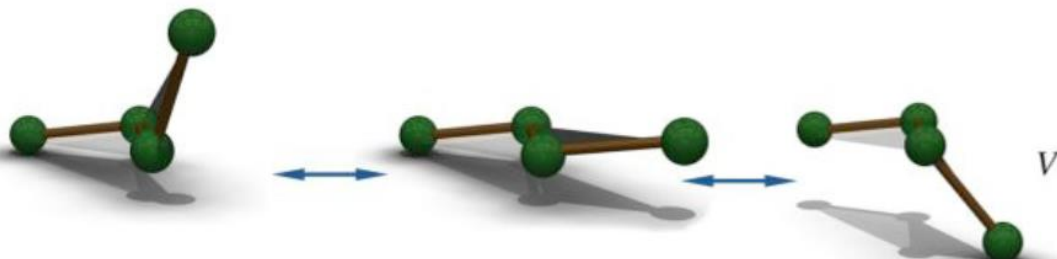
deformação de um ângulo

$$V^{\text{ângulo}}(\vec{r}(t)) = \frac{1}{2} \sum_{n=1}^{N_\theta} K_{\theta_n} [\cos(\theta_n(t)) - \cos(\theta_{0_n})]^2$$



torção de um diedro impróprio

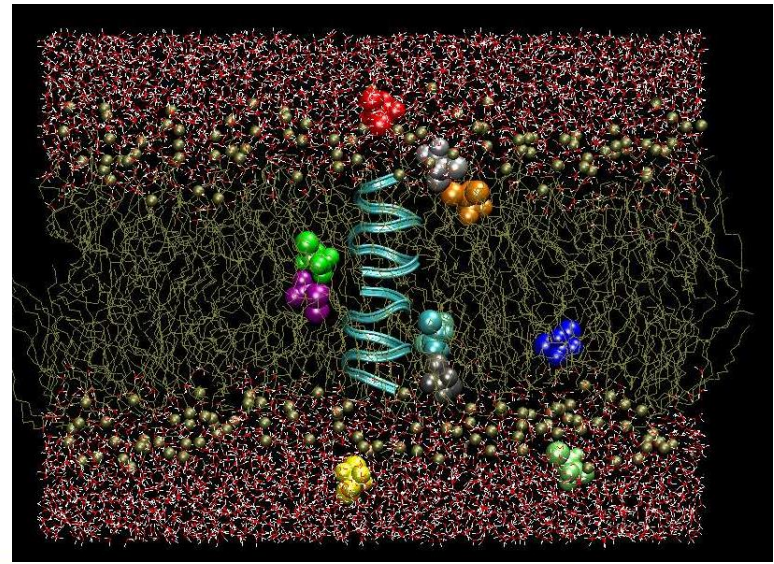
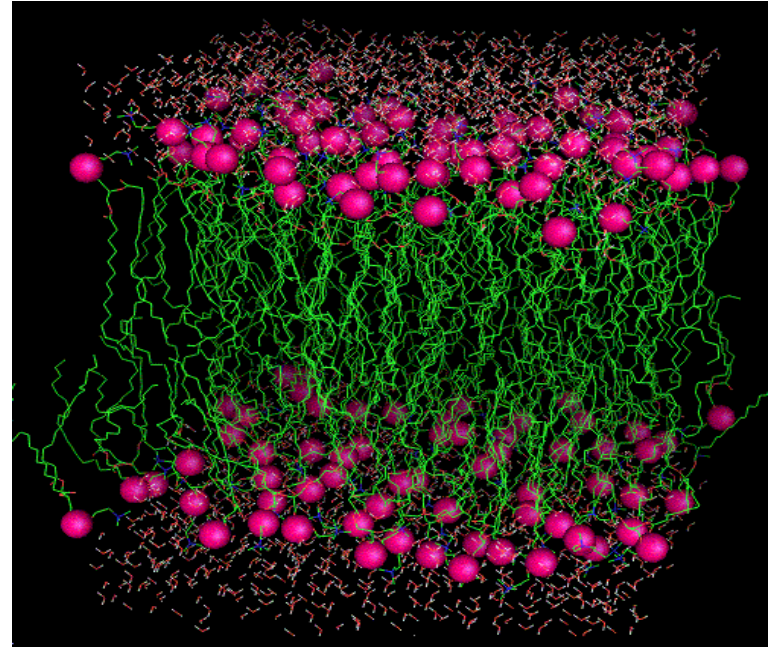
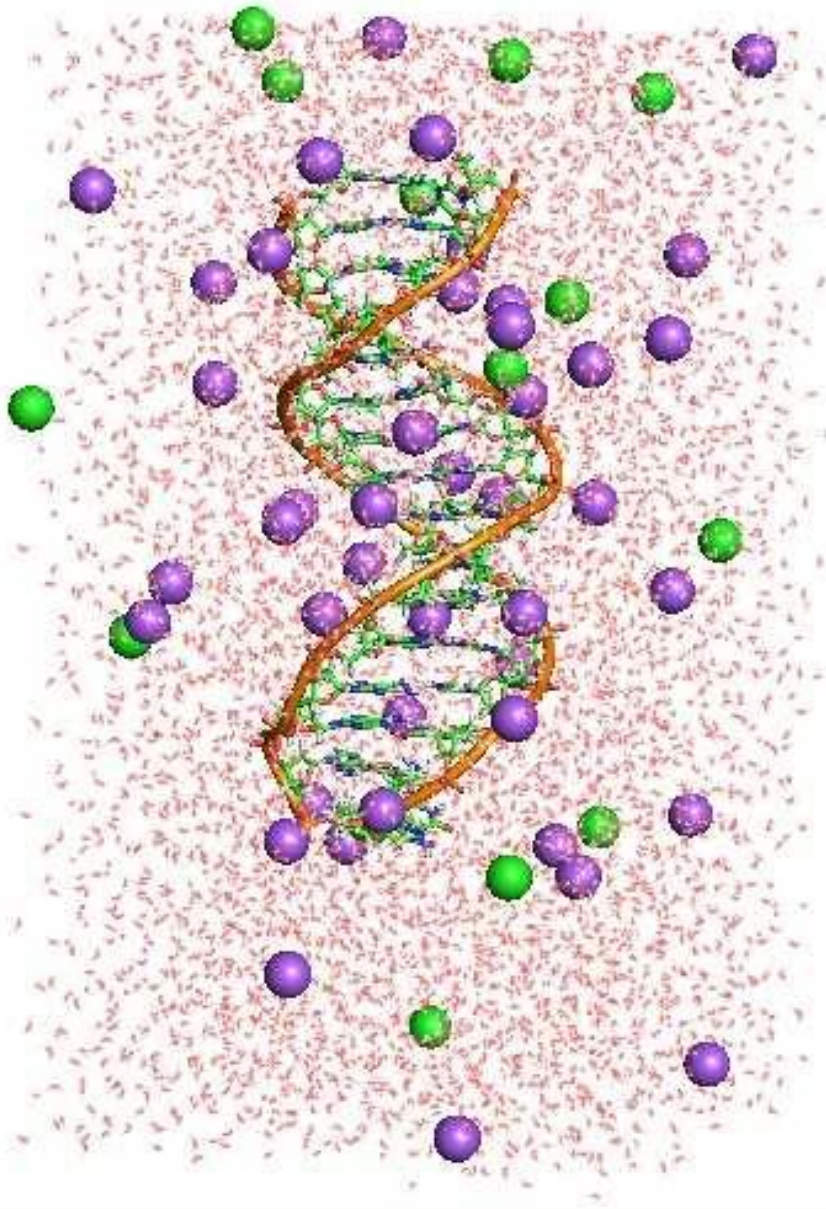
$$V^{\text{impróprio}}(\vec{r}(t)) = \frac{1}{2} \sum_{n=1}^{N_\xi} K_{\xi_n} [\xi_n(t) - \xi_{0_n}]^2$$



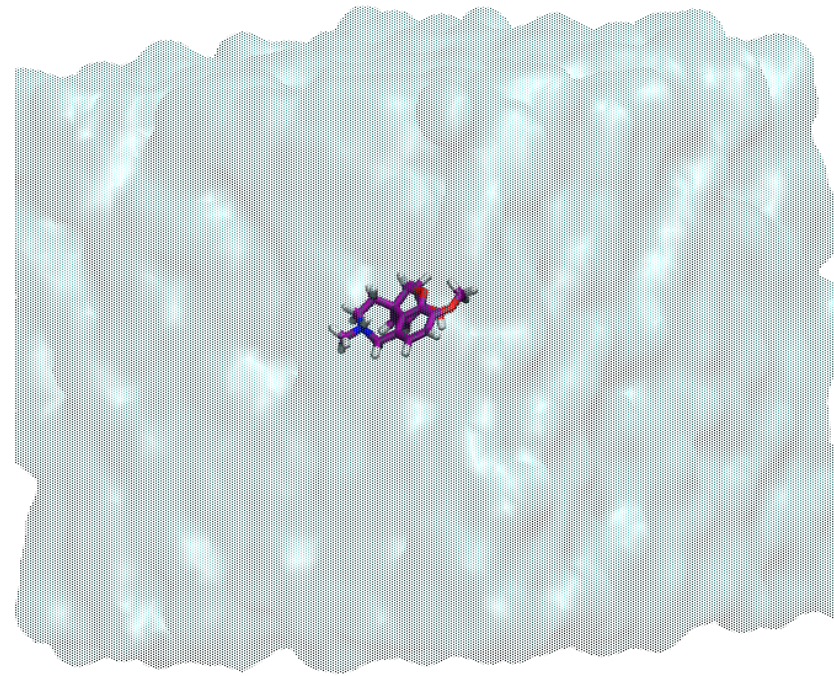
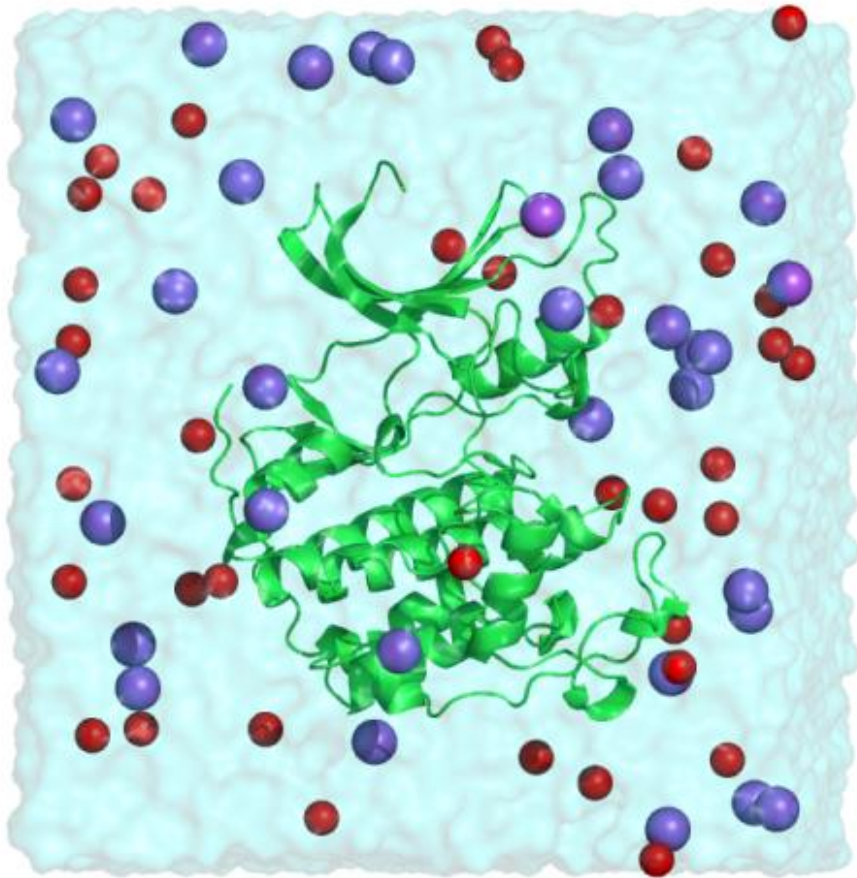
torção de um diedro próprio

$$V^{\text{próprio}}(\vec{r}(t)) = \sum_{n=1}^{N_\phi} K_{\phi_n} [1 + \cos(\delta_n) \cos(m_n \phi_n(t))]$$

Sistemas possíveis de ser simulados...

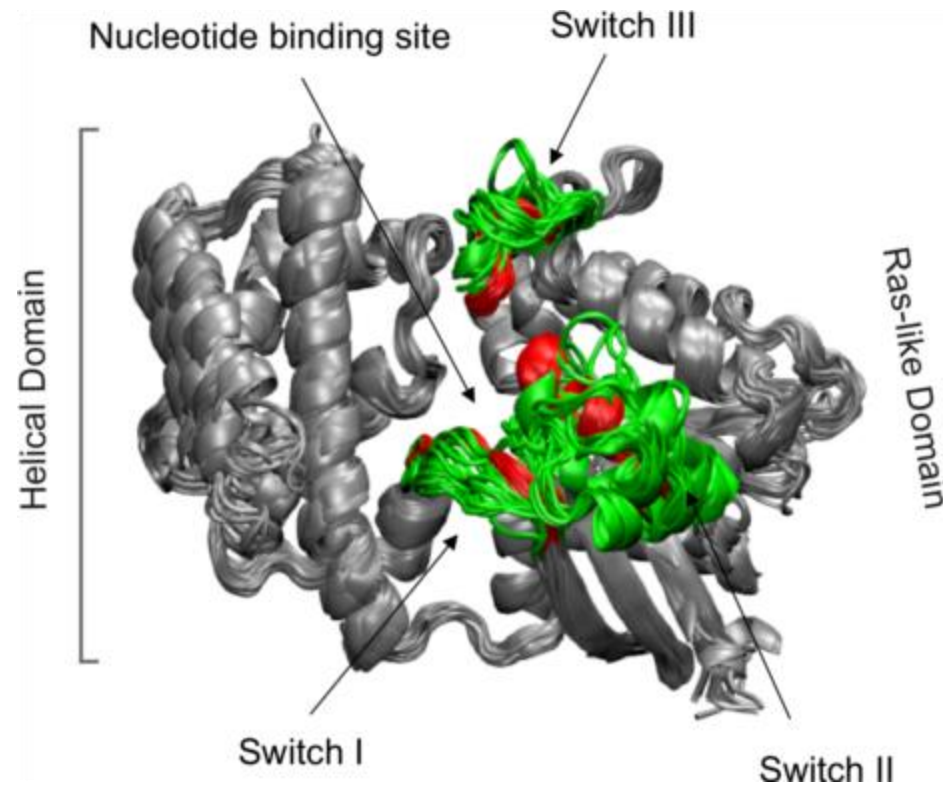


Sistemas possíveis de ser simulados...



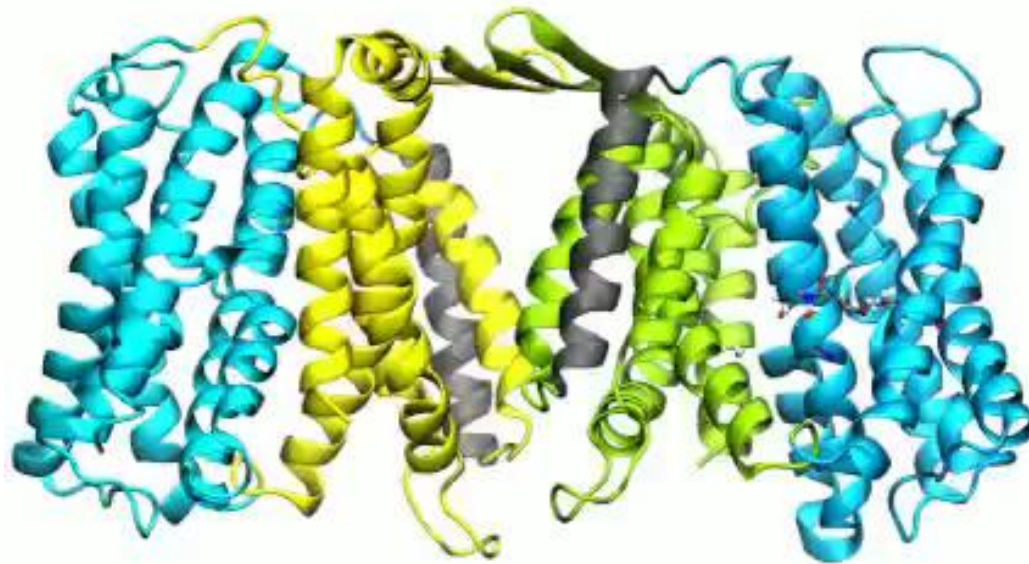
Por que fazer simulações?

A DM possibilita obter modelos de moléculas muito mais próximos da realidade biológica, pois inclui diretamente características como a **flexibilidade molecular** (através da variação temporal de propriedades) e a **temperatura** (através da aceleração dos átomos).



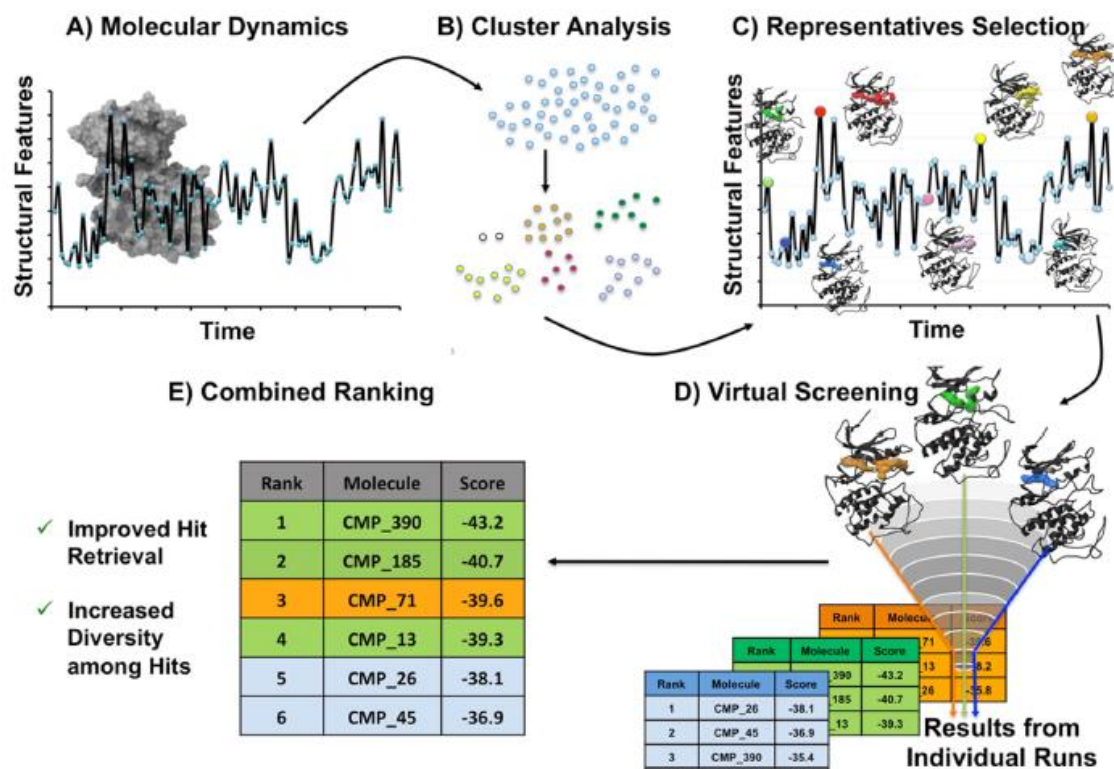
Por que fazer simulações?

A DM possibilita obter modelos de moléculas muito mais próximos da realidade biológica, pois inclui diretamente características como a **flexibilidade molecular** (através da variação temporal de propriedades) e a **temperatura** (através da aceleração dos átomos).



Por que fazer simulações?

- Determinação dos estados acessíveis a um sistema.
- Observação de progressão temporal de processos.
- Compreensão de alosteria.
- Avaliação da estabilidade de resultados de atracamento.
- Cálculos de ΔG .



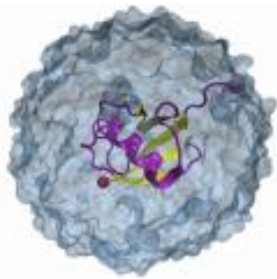
“Como”?

- NAMD; AMBER; GROMACS; entre outros

University of Illinois at Urbana-Champaign
NIH Center for Macromolecular Modeling and Bioinformatics
Beckman Institute
Computational Biophysics Workshop

NAMD TUTORIAL

Windows Version



NAMD Developers: James Phillips, David Hardy

NAMD Tutorial Contributors: Tim Legro, James Phillips, Marcos Sotomayor, Elizabeth Villa, Hang Yu, David Tanner, Yanxin Liu, Zhe Wu, David Hardy

April 2017

A current version of this tutorial is available at
<http://www.ku.xinc.edu/Training/Tutorials/>
Join the tutorial-18ka.uiuc.edu mailing list for additional help.

Amber 2016 Reference Manual

(Covers Amber16 and AmberTools16)



GROMACS

Groningen Machine for Chemical Simulations



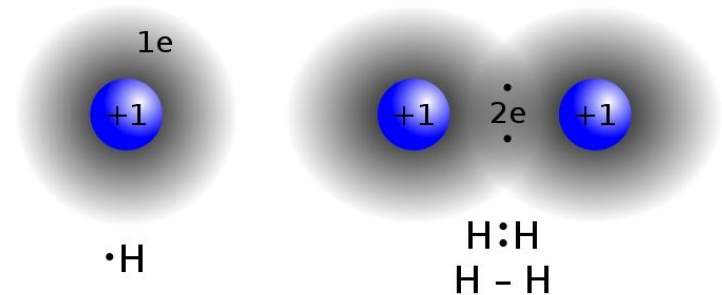
Reference Manual

Version 2016

Limitações atuais da DM

Ausência de elétrons: cálculos baseados na mecânica molecular (campos de força)

- **Não são capazes de descrever reações químicas.**
- Alternativa: métodos híbridos entre a mecânica molecular e a mecânica quântica.

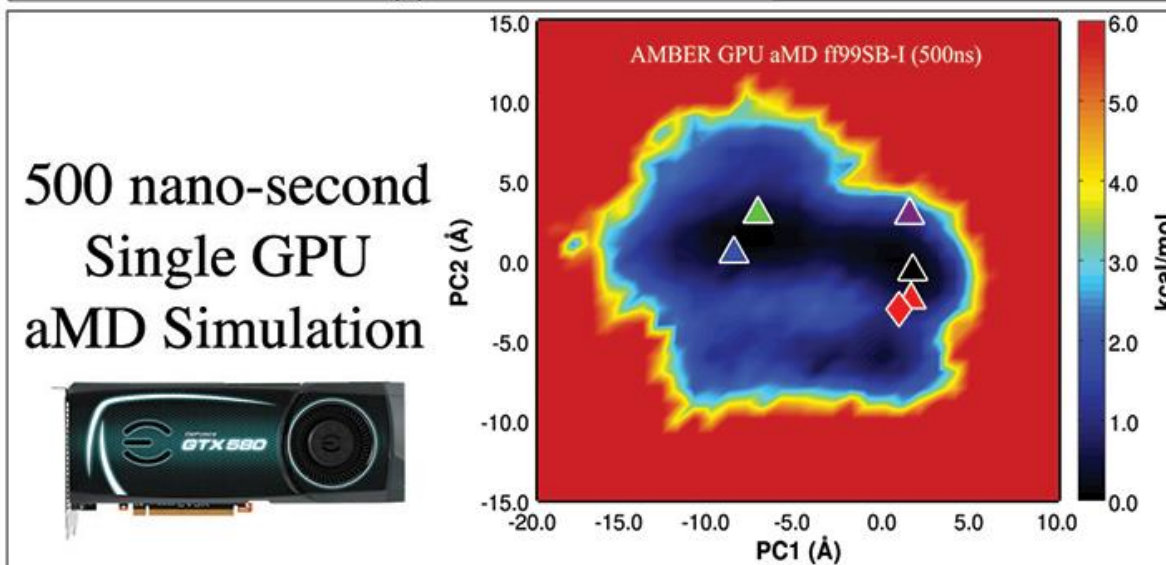
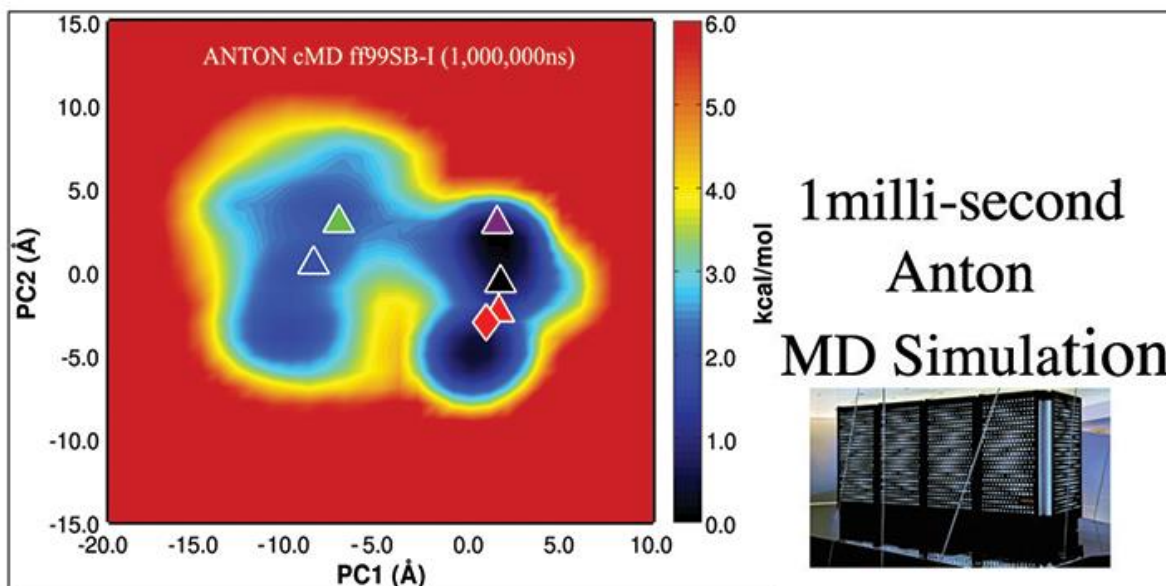


Tempo computacional: obter amostragens compatíveis com fenômenos observáveis em experimentos ou fisiologicamente.

- Ainda não é possível chegar em escalas de tempo compatíveis com o comportamento de proteínas em soluções biológicas.

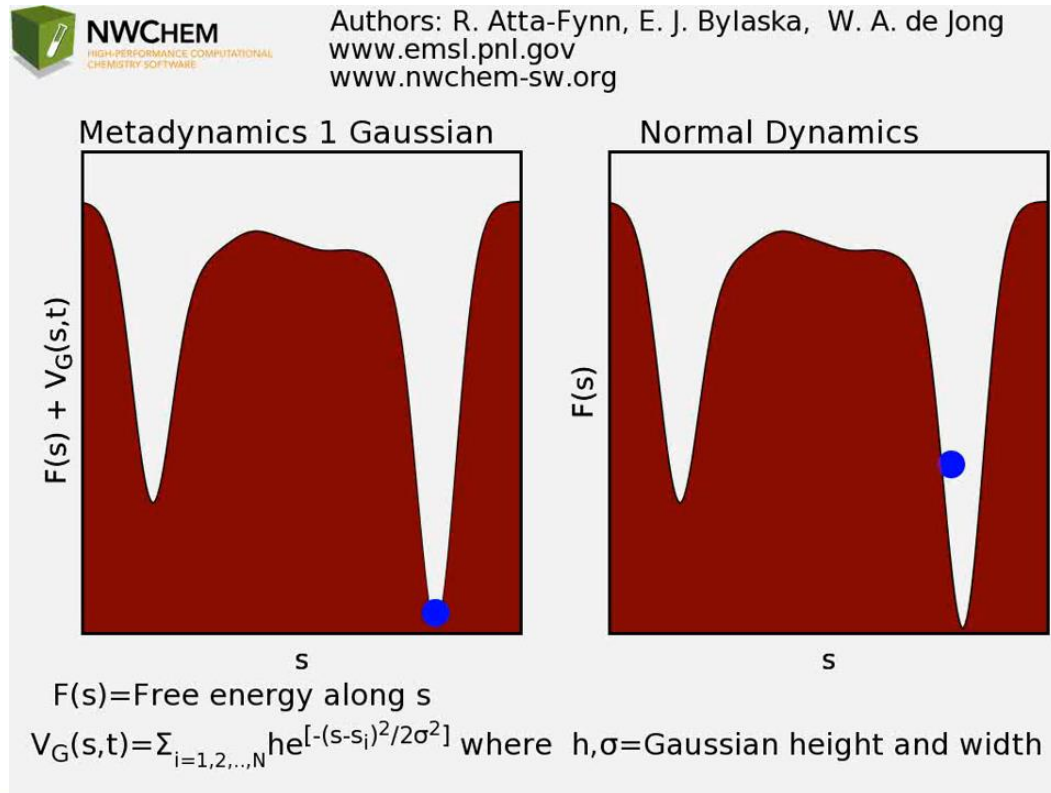


Métodos que “aceleram” a escala de tempo



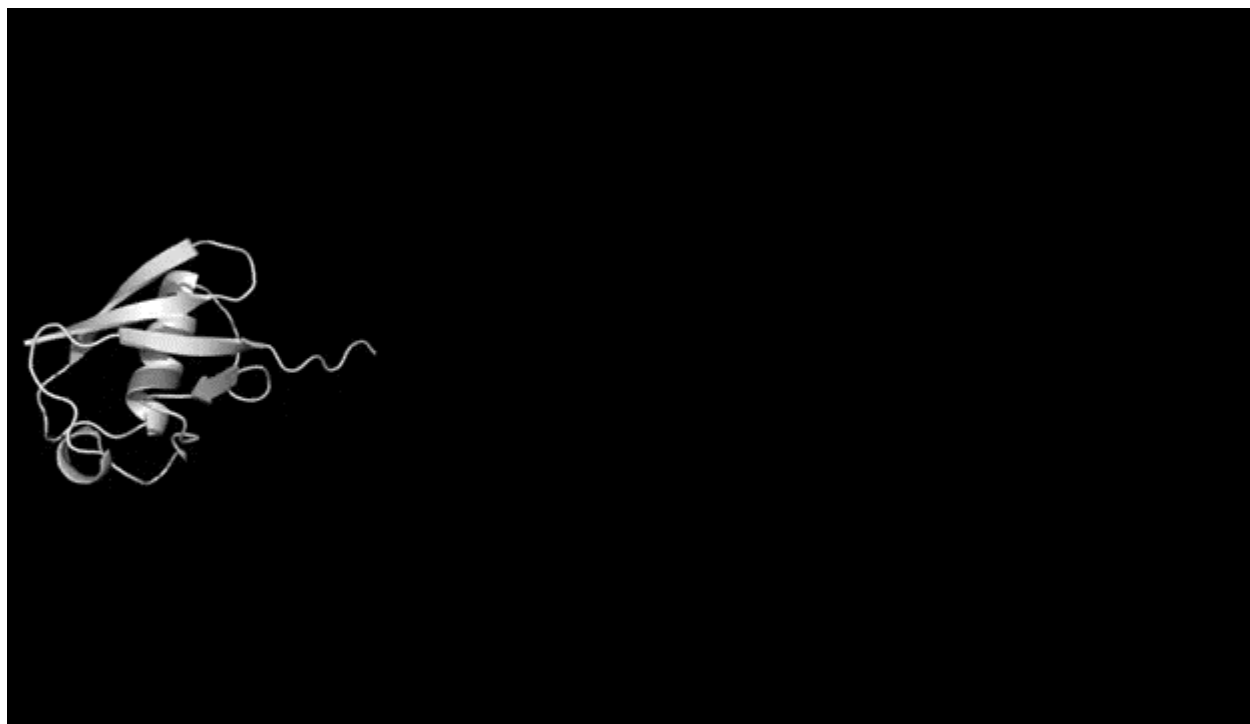
Chamados *enhanced sampling methods*

- **Metadinâmica**
- Dinâmica acelerada
- Dinâmica direcionada
- entre outros.



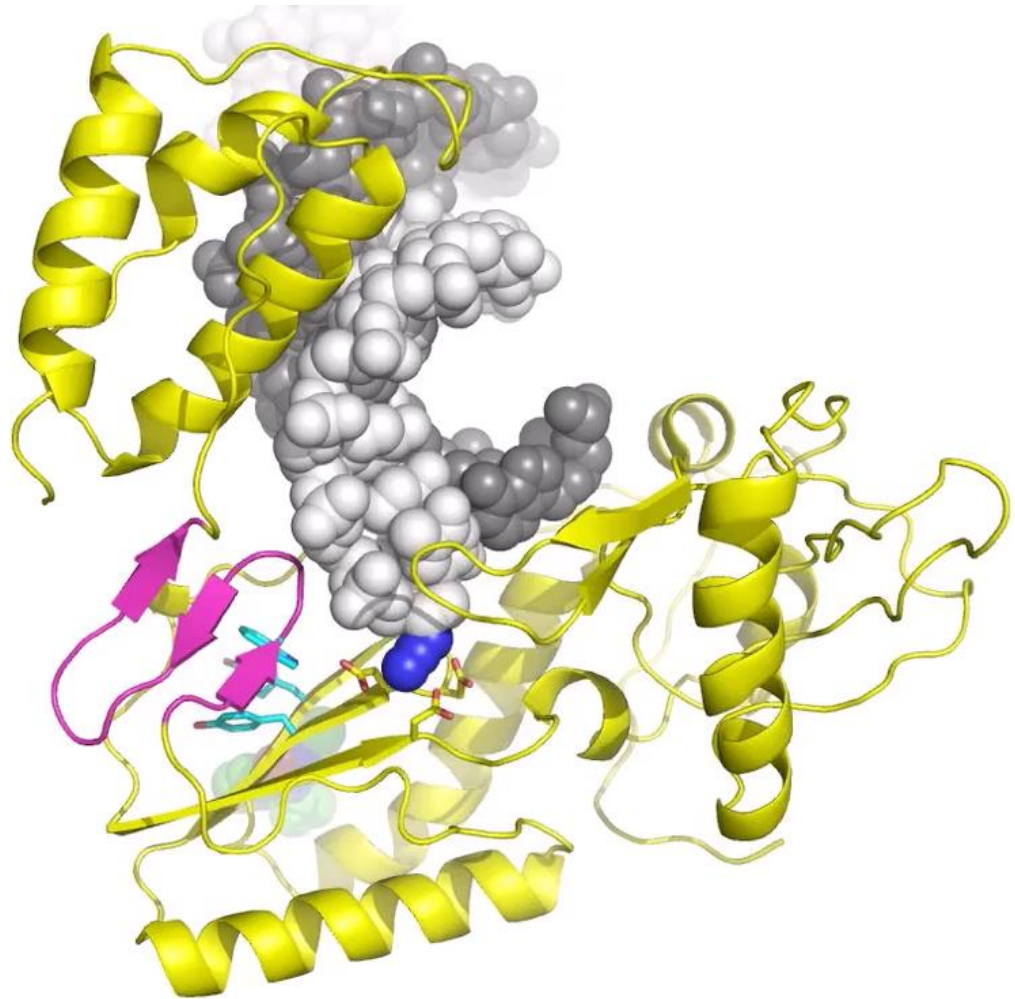
Chamados *enhanced sampling methods*

- Metadinâmica
- Dinâmica acelerada
- **Dinâmica direcionada**
- entre outros.



Ou métodos que aproximam

- **Modos normais**
- MMPB/GB-SA
- entre outros



Cronograma

Aulas práticas de:

- Visualização de estruturas com o Pymol
- Atracamento molecular
- Modelagem comparativa

Obrigada.

Dúvidas?

Contato: luciannahss@gmail.com