

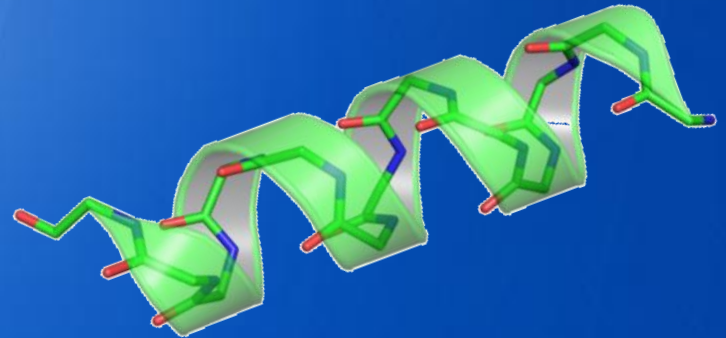
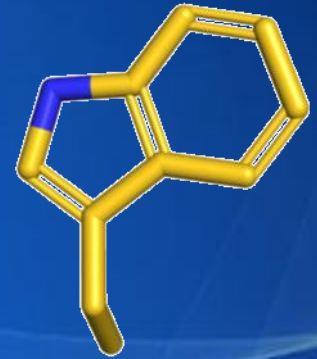
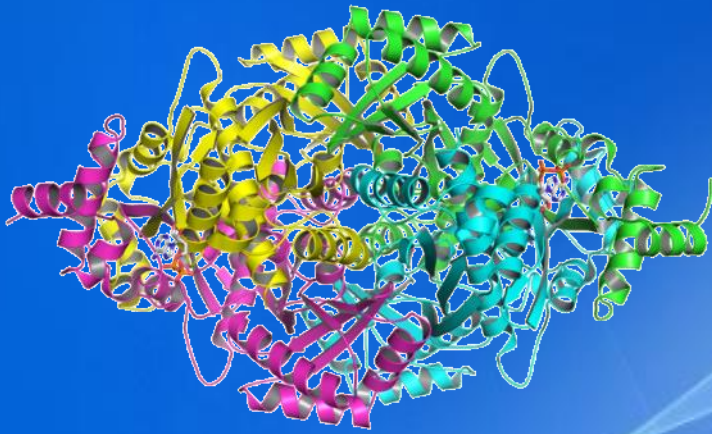
FFI0750 – Biologia Molecular Estrutural

Prof. Rafael V. C. Guido
rvcguido@ifsc.usp.br

Aula 02

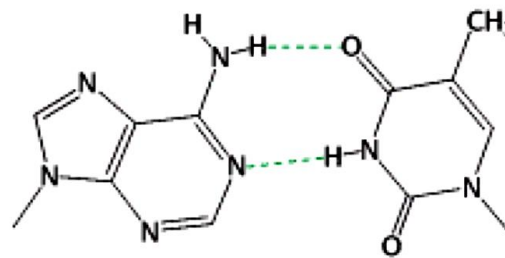
Bacharelado em Ciências Físicas e Biomoleculares
Instituto de Física de São Carlos - USP

Por que estudamos proteínas?



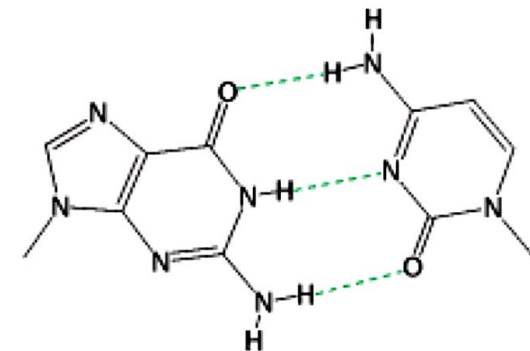


GACTTCACTTCTAATGATGATTATGGGAGAACTGGAGCCTTCAGAGGGT
AAAAATTAAGCACAGTGGAAGAATTCATTCTGTTCTCAGTTTTCTGGA
TTATGCCTGGCACCATTAAAGAAAATATCTTTGGTGTTCCTATGATGAAT
ATAGATACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAG...



Adenine (A)

Thymine (T)



Guanine (G)

Cytosine (C)

- Mais de 3 bilhões de nucleotídeos
- O genoma humano se fosse escrito em páginas de papel A4 necessitaria de mais de 500.000 páginas



GenBank



National Library of Medicine
National Center for Biotechnology Information

<https://www.ncbi.nlm.nih.gov/genbank/>

GenBank

Nucleotide

Search

GenBank

Submit

Genomes

WGS

Metagenomes

TPA

TSA

INSDC

Other

GenBank Overview

What is GenBank?

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan;41(D1):D36-42). GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the [ftp site](#). The [release notes](#) for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for [previous GenBank releases](#) are also available. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An [annotated sample GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

GenBank Resources

[GenBank Home](#)

[Submission Types](#)

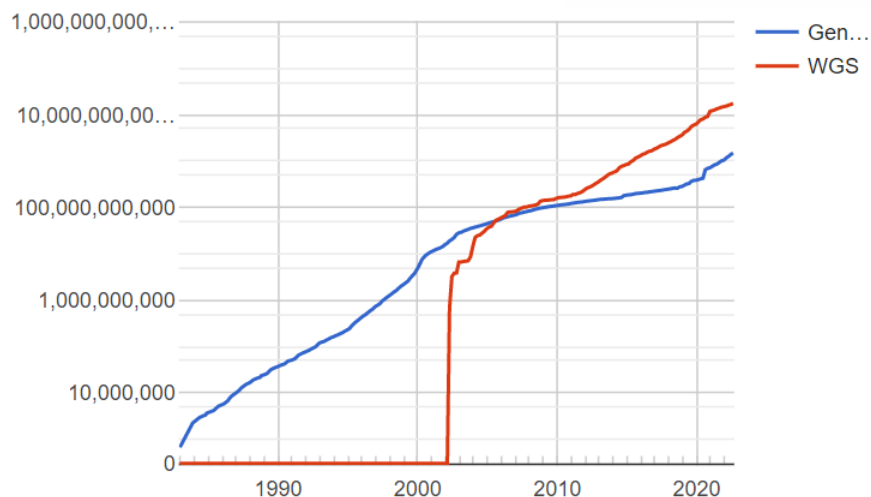
[Submission Tools](#)

[Search GenBank](#)

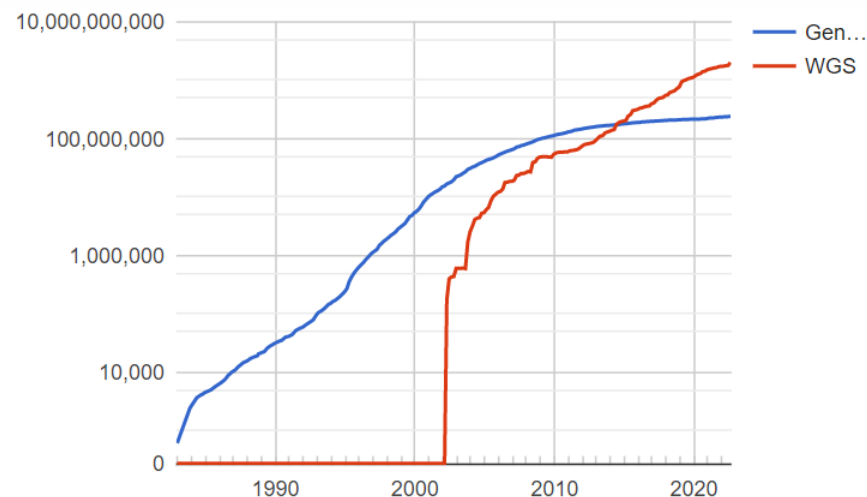
[Update GenBank Records](#)

GenBank and WGS Statistics

Bases

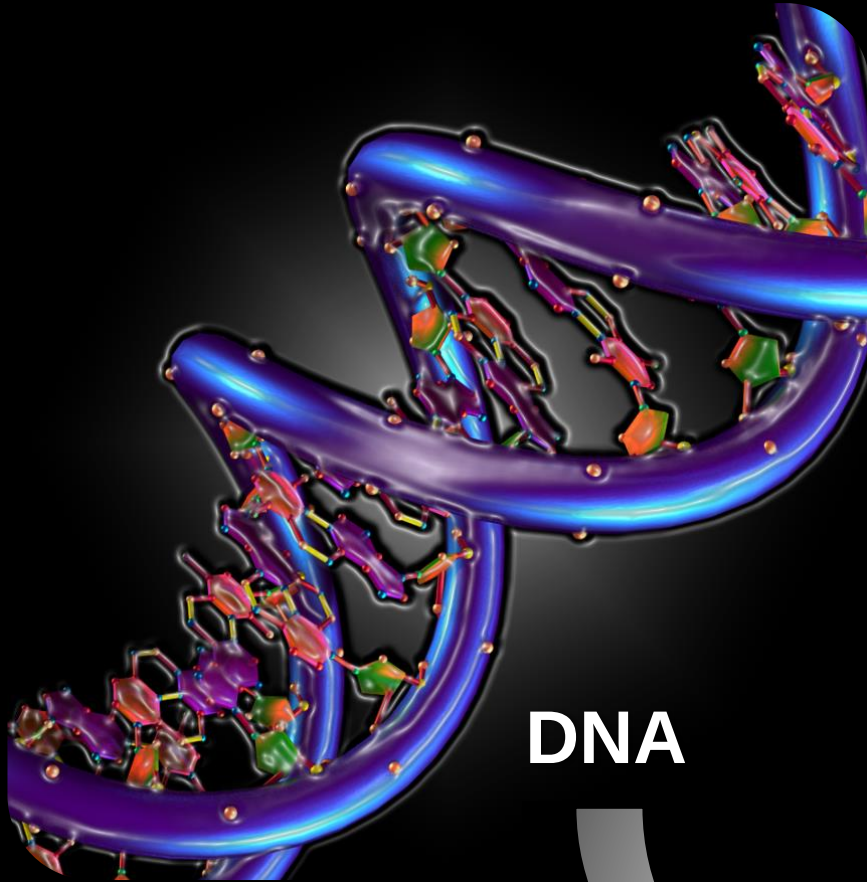


Sequences

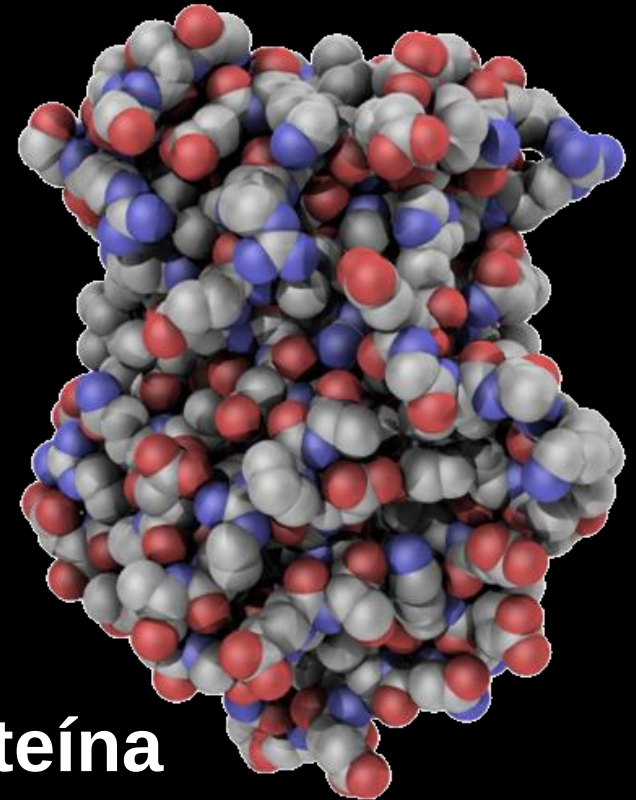


<https://www.ncbi.nlm.nih.gov/genbank/statistics/>

DNA → RNA → Proteínas



DNA



Proteína

RNA

Protein data bank (PDB)

<https://www.rcsb.org/>

RCSB PDB

Deposit ▾

Search ▾

Visualize ▾

Analyze ▾

Download ▾

Learn ▾

More ▾

Documentation ▾

Careers

MyPDB ▾

RCSB PDB

PROTEIN DATA BANK

194259 Biological
Macromolecular Structures
Enabling Breakthroughs in
Research and Education

▼ PDB Archive ?

Advanced Search | Browse Annotations

Help

PDB-101

WORLDWIDE PDB
PROTEIN DATA BANK

EMDataResource
Unified Data Resource for 3DEM

NUCLEIC ACID
DATABASE

Worldwide
Protein Data Bank
Foundation

f t y d

Developers: Join the RCSB PDB Team [Explore Open Positions](#)

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.

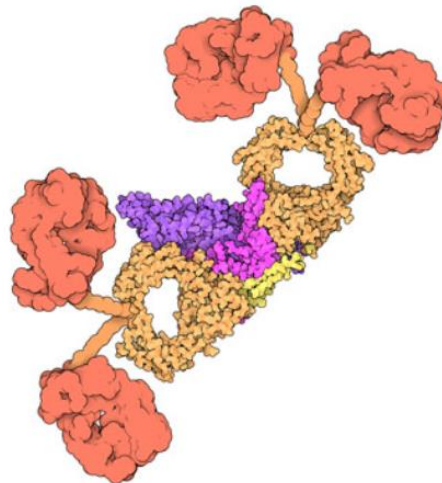
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

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.

COVID-19
CORONAVIRUS
Resources

Join the
RCSB PDB
Team

August Molecule of the Month



Secretory Antibodies

Protein data bank (PDB)

<https://www.rcsb.org/>

RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

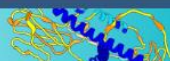
RCSB PDB PROTEIN DATA BANK

208,702 Structures from the PDB
1,068,577 Computed Structure Models (CSM)

3D Structures ? Enter search term(s), Entry ID(s), or sequence Include CSM ?

Advanced Search | Browse Annotations Help

PDB-101 PDB EMDatResource NAKB wwPDB Foundation PDB-Dev



New: More Computed Structure Models (CSM) available

[Learn more](#)

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

Experimentally-determined 3D structures from the **Protein Data Bank (PDB)** archive

Computed Structure Models (CSM) from AlphaFold DB and ModelArchive

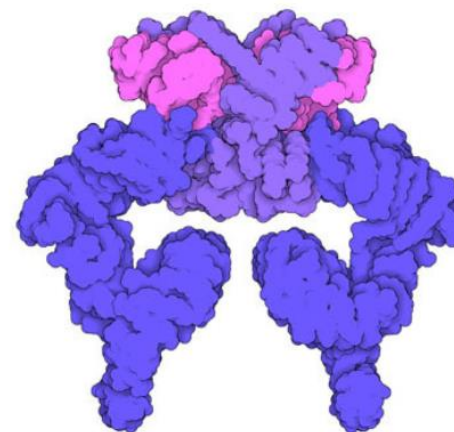
These data can be explored in context of external annotations providing a structural view of biology.

Explore
NEW
Features



PDB-101
Training
Resources

August Molecule of the Month



ATM and ATR Kinases

Latest Entries

As of Tue Aug 15 2023

Features & Highlights

News

Publications+



STRUCTURAL
BIOLOGY

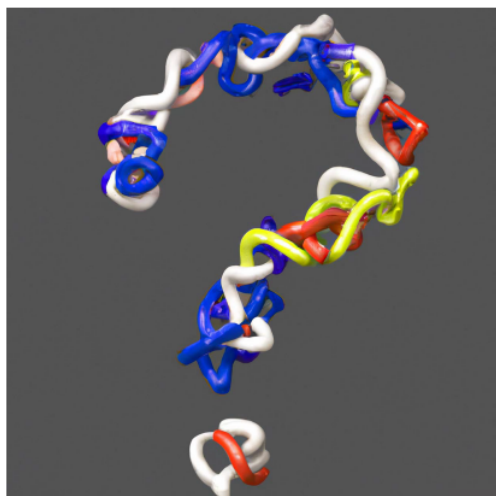
ISSN 2059-7983

AlphaFold and the future of structural biology¹

Randy J. Read,^{a*} Edward N. Baker,^{b*} Charles S. Bond,^{c*} Elspeth F. Garman^{d*} and Mark J. van Raaij^{e*}

Structural biologists quickly embraced the advantages that these models bring (Perrakis & Sixma, 2021), including insights into improved construct design, the generation of compelling hypotheses about interacting proteins or domains, and the acceleration of structure solution by molecular replacement (Millán *et al.*, 2021) for crystallography or docking for electron cryo-microscopy (cryo-EM). A particularly notable example was integrative modelling of the nuclear pore complex by combining *AlphaFold* (Jumper *et al.*, 2021) and *RoseTTAFold* (Baek *et al.*, 2021) models with cryo-electron tomography and complementary data (Mosalaganti *et al.*, 2022).

Most structural biologists are aware that the models are not actually as accurate as experimental structures. The backbone accuracy measured in CASP does not ensure the accuracy of all coordinates including side chains. An objective evaluation of how well different models explain experimental diffraction data showed that experimental structures from alternative crystal forms (in spite of different crystal-packing interactions) are generally better than *AlphaFold* models (Terwilliger *et al.*, 2023). It has also been reported that *AlphaFold* models perform less well than experimental structures as targets for the computational docking algorithms used in drug design (Karelina *et al.*, 2023).

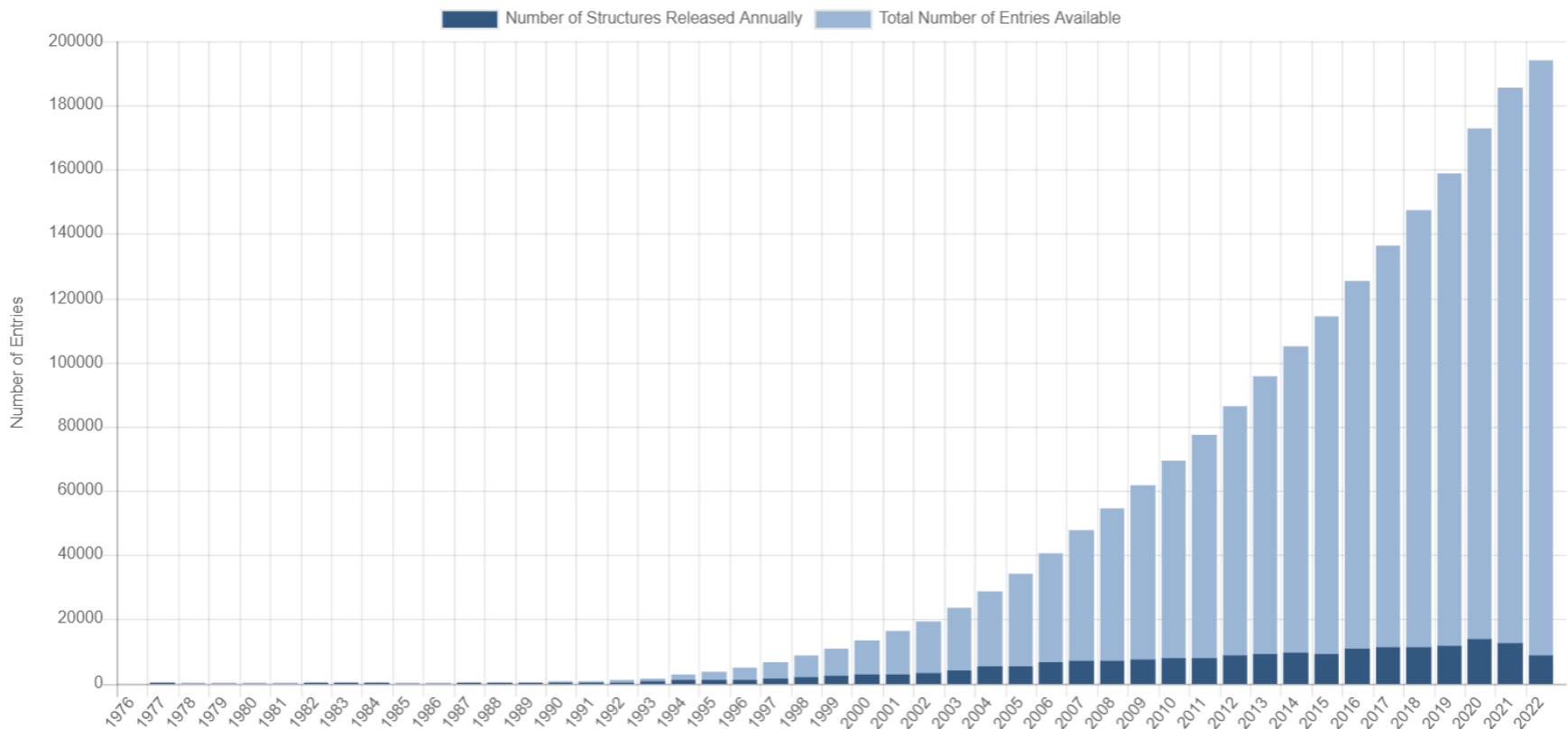


Protein data bank (PDB)

Proteínas

PDB Statistics: Overall Growth of Released Structures Per Year

All Statistics



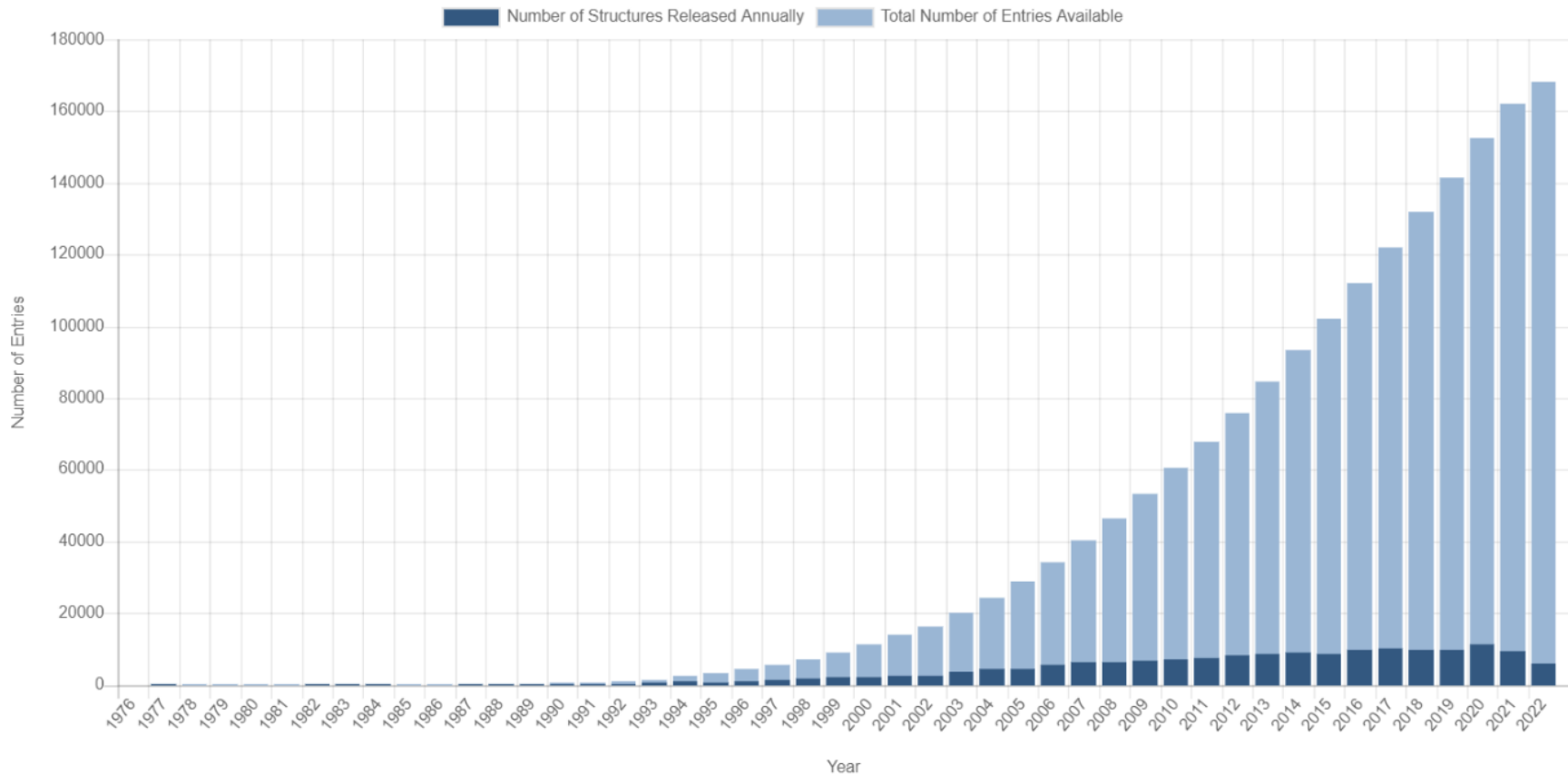
<https://www.rcsb.org/stats/growth/growth-released-structures>

Proteínas

Protein data bank (PDB)

PDB Statistics: Growth of Structures from X-ray Crystallography Experiments Released per Year

Experimental methods such as [X-ray crystallography](#), [NMR spectroscopy](#), and [3D electron microscopy](#) are used to determine the location of each atom relative to each other in the molecule.



<https://www.rcsb.org/stats/growth/growth-released-structures>

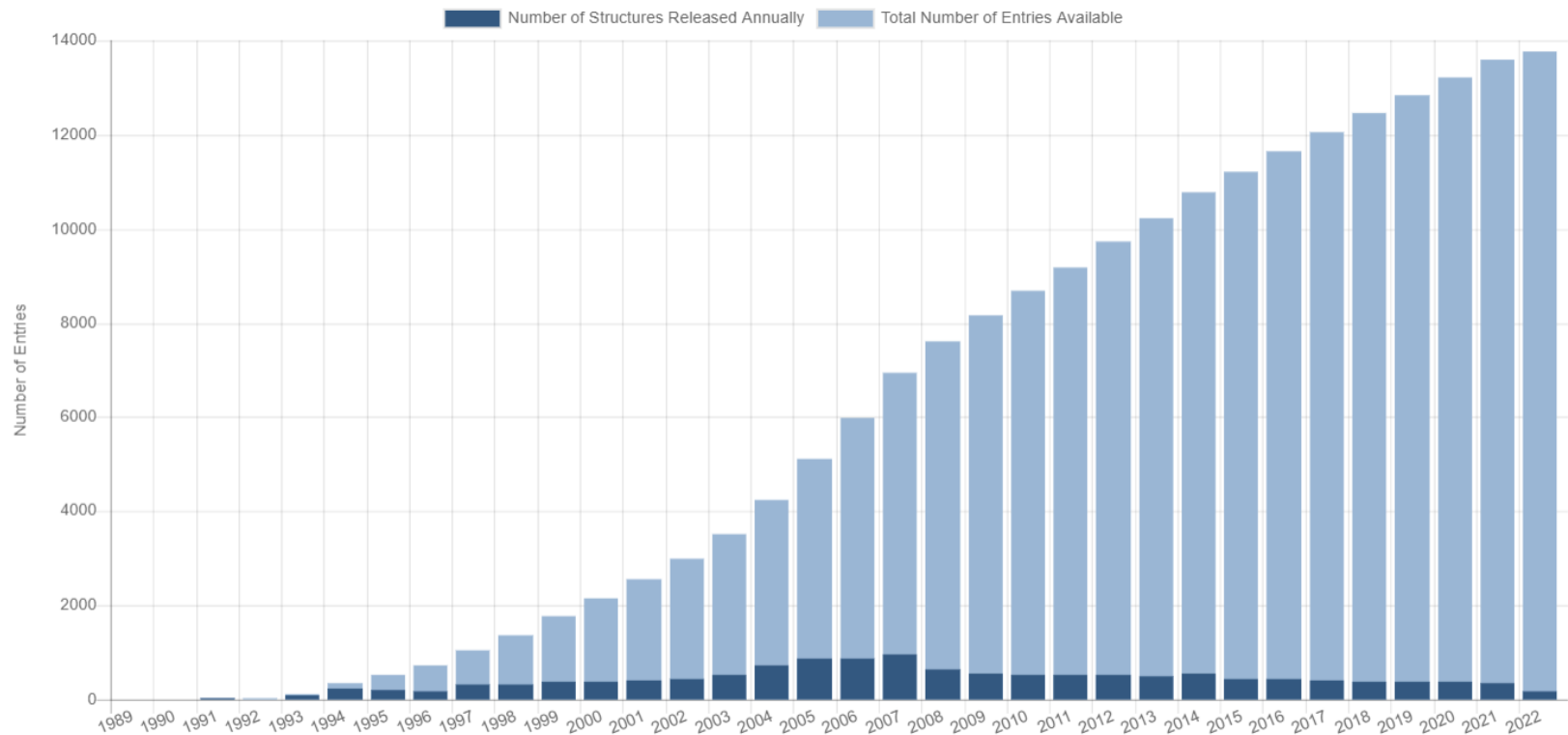
Protein data bank (PDB)

Proteínas

All Statistics

PDB Statistics: Growth of Structures from NMR Experiments Released per Year

Experimental methods such as X-ray crystallography, NMR spectroscopy, and 3D electron microscopy are used to determine the location of each atom relative to each other in the molecule.



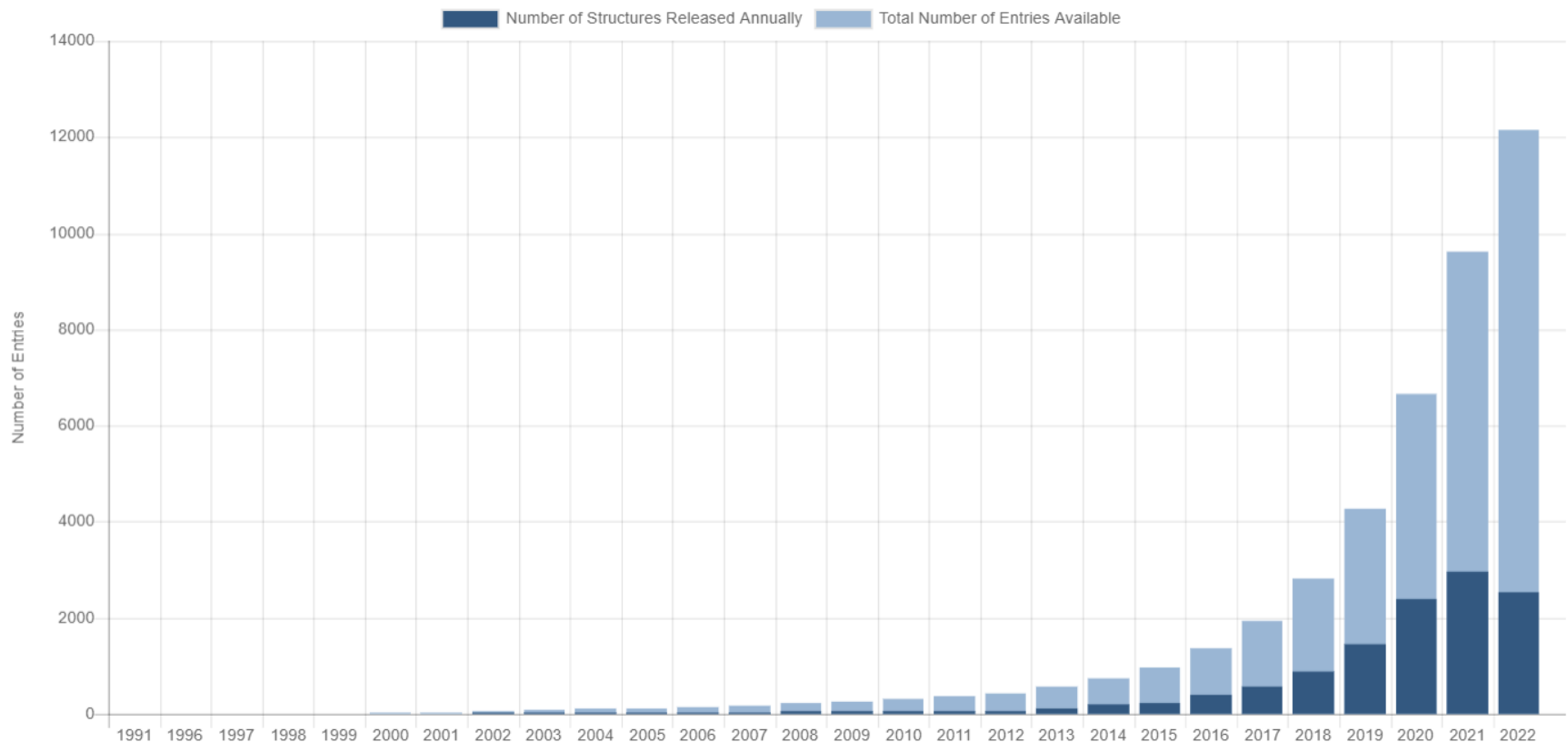
<https://www.rcsb.org/stats/growth/growth-released-structures>

Protein data bank (PDB)

[All Statistics](#)

PDB Statistics: Growth of Structures from 3DEM Experiments Released per Year

Experimental methods such as [X-ray crystallography](#), [NMR spectroscopy](#), and [3D electron microscopy](#) are used to determine the location of each atom relative to each other in the molecule.



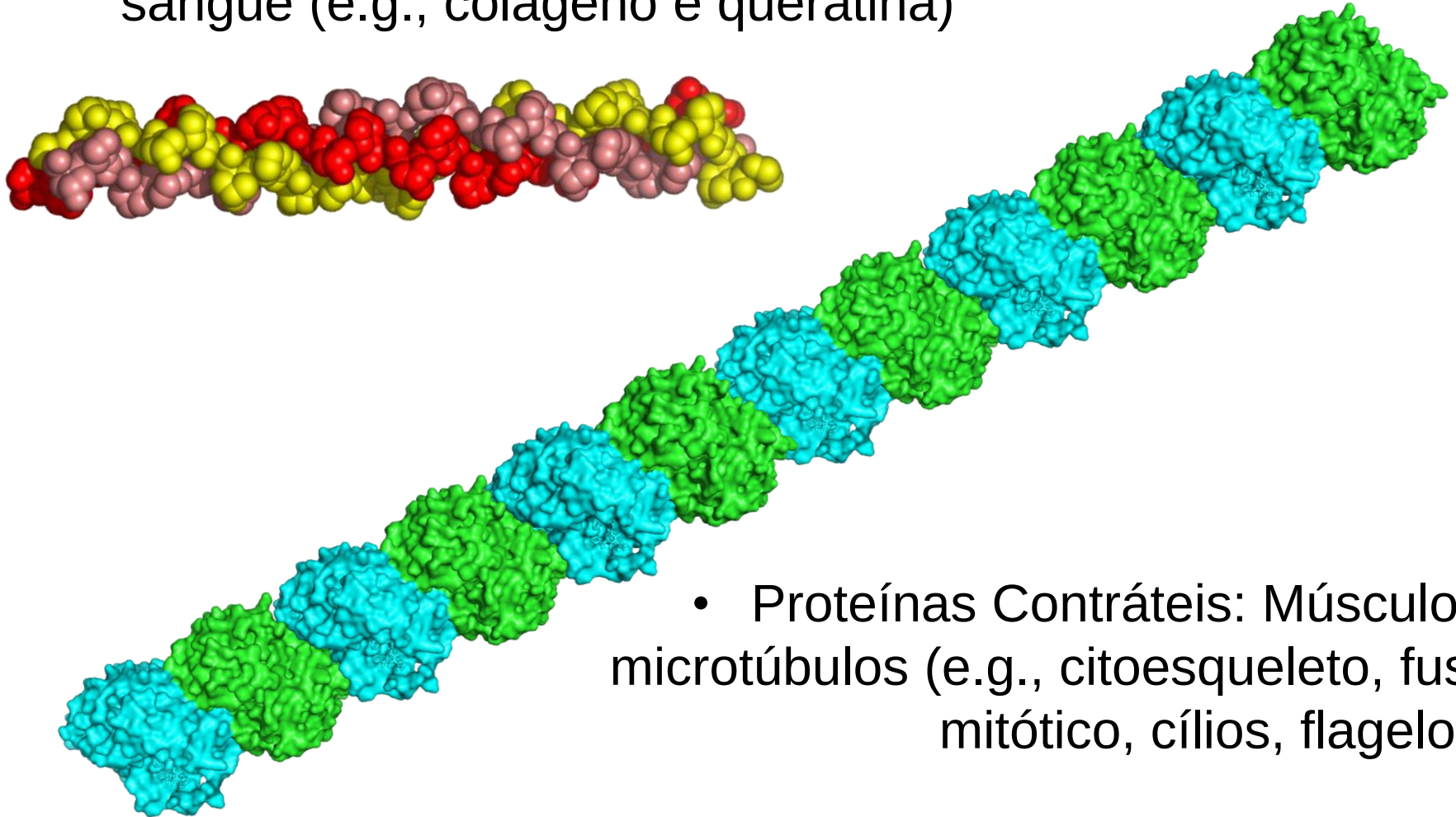
<https://www.rcsb.org/stats/growth/growth-released-structures>

Classificação de Proteínas por Função

- **CATALÍTICA:** enzimas
- **ARMAZENAMENTO:** ovalbumina (ovos), caseína (leite), zeína (milho)
- **TRANSPORTE:** hemoglobina
- **COMUNICAÇÃO:** hormônios (insulina) e neurotransmissores
- **CONTRÁTEIS:** actina, miosina
- **PROTEÇÃO:** imunoglobulina, fibrinogênio, fatores de coagulação do sangue
- **TOXINAS:** veneno de serpente
- **ESTRUTURAL:** queratina (cabelo), colágeno

Proteínas Fibrosas

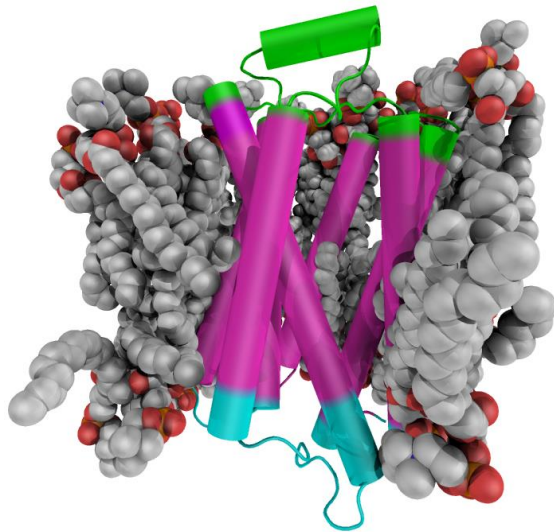
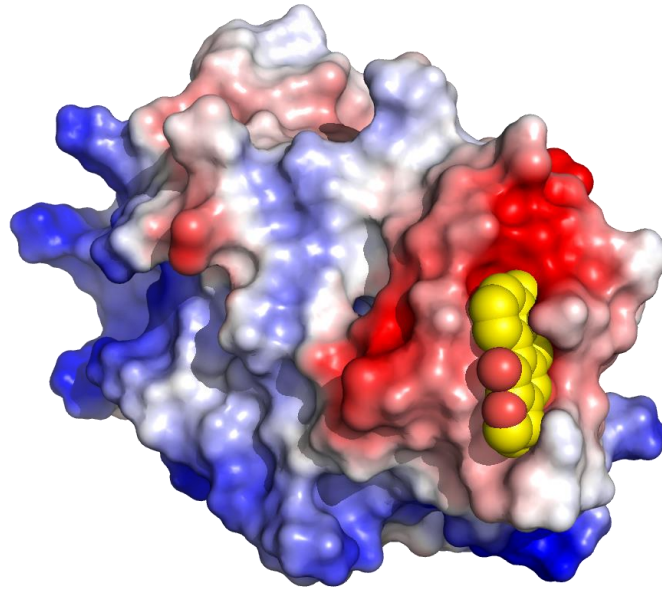
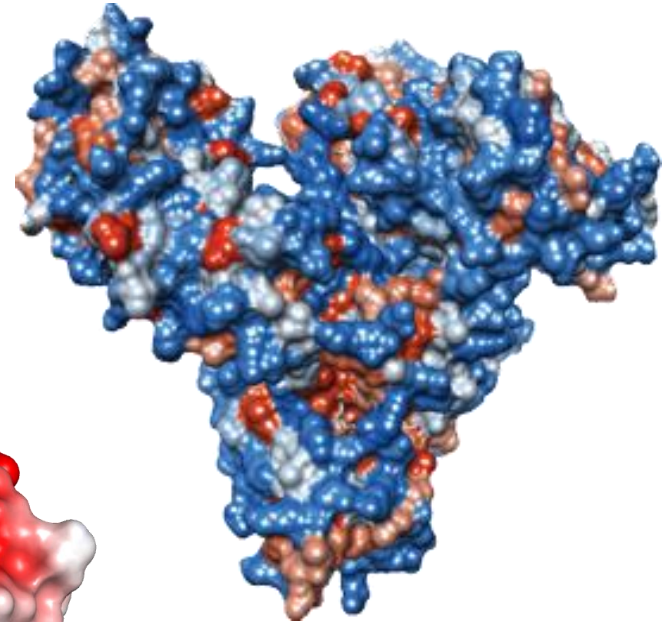
- Proteínas Estruturais: Tendões, ligamentos, coágulos de sangue (e.g., colágeno e queratina)



- Proteínas Contráteis: Músculos, microtúbulos (e.g., citoesqueleto, fuso mitótico, cílios, flagelos)

Proteínas Globulares

- Proteínas de armazenamento (e.g., albumina)



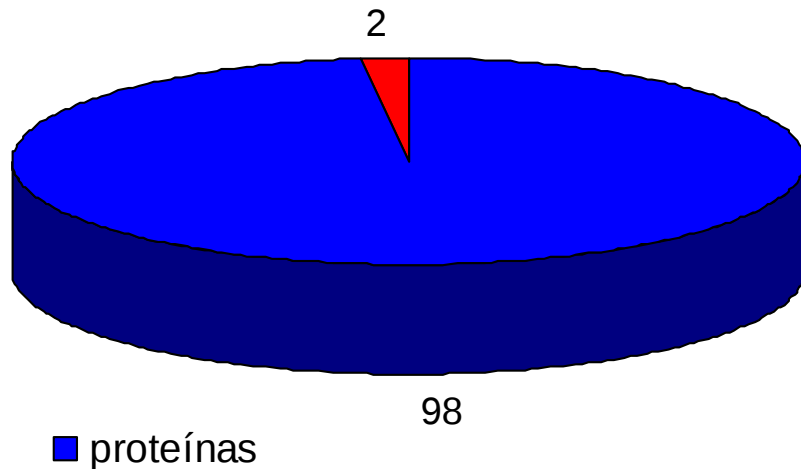
- Proteínas catalíticas, transporte, proteção: enzimas, hemoglobina, imunoglobulinas, receptores de membrana

Alvos Moleculares para Fármacos



Principais Alvos Biológicos

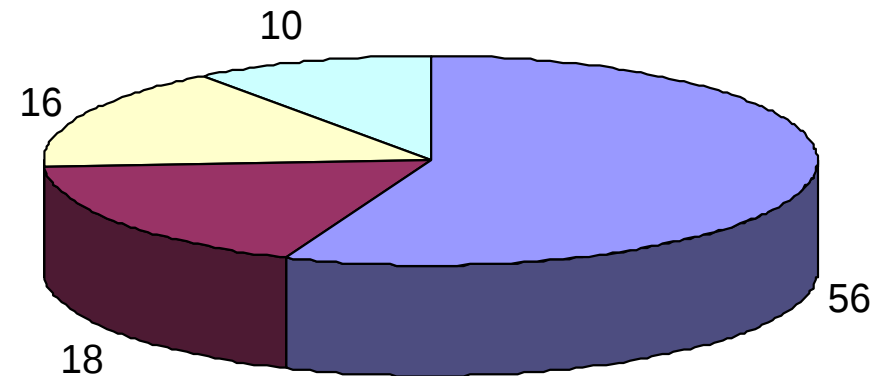
1.357 fármacos em uso
terapêutico



■ proteínas

■ ácidos nucleicos + lipídeos +
polissacarídeos

Considerando os 100 fármacos de
maior sucesso no mundo tem-se:



■ enzimas

■ receptores transmembrânicos

■ canais iônicos

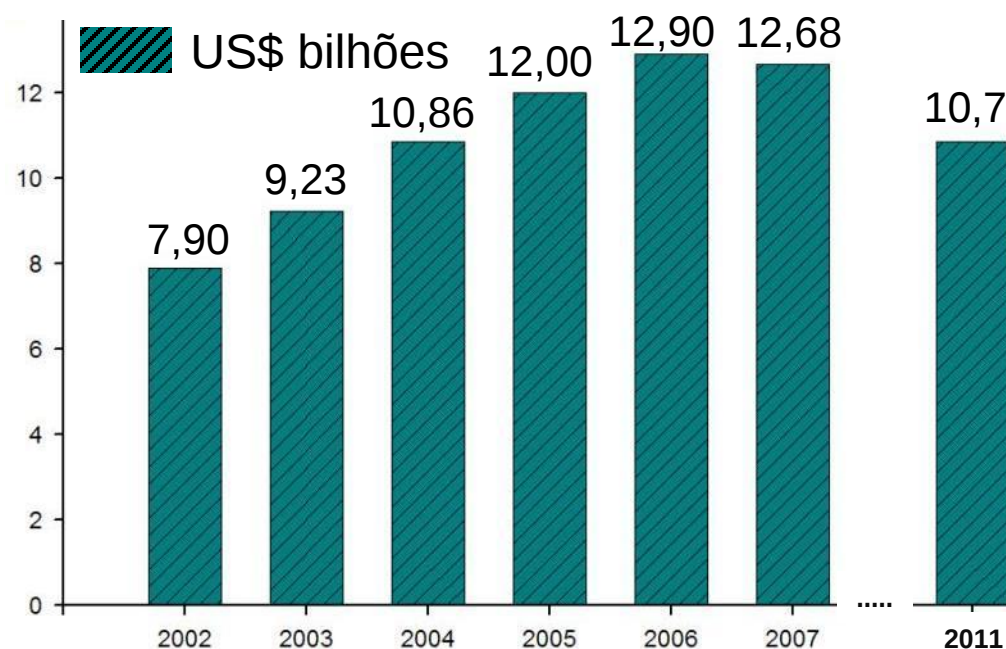
■ receptores nucleares

O fármaco mais bem sucedido da história

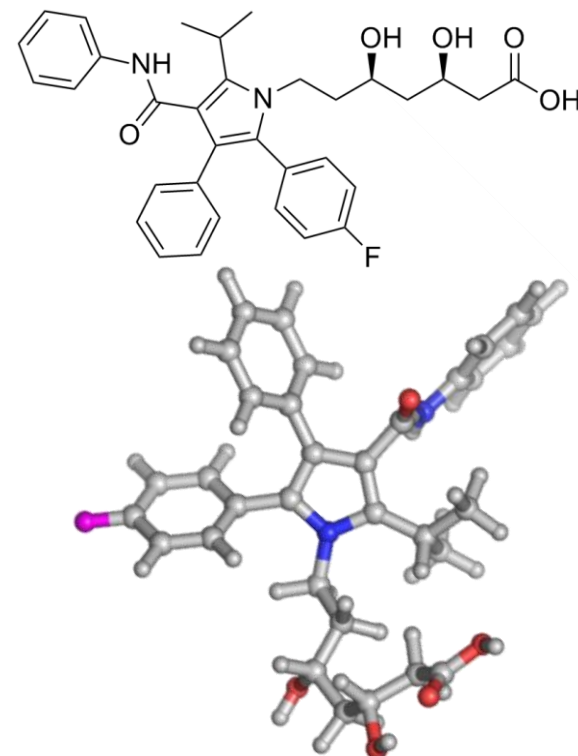


Lípor® atorvastatina cálcica

- Redutor de colesterol;
- O maior sucesso de vendas da história



$C_{33}H_{35}FN_2O_5 = 76$ átomos
US\$ 140,79 milhões/átomo!!!



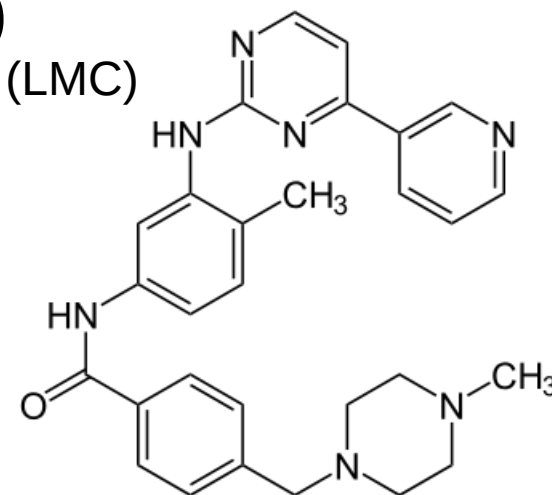
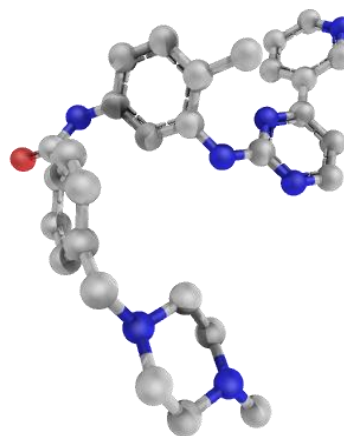
Glivec (Imatinib)



Gleevec/Glivec®
mesilato de imatinib

Anticâncer (Via Oral)

- leucemia mieloide crônica (LMC)
- tumores gastrointestinais



ANO	VENDAS (US\$ BILHÕES)
2007	3,05
2008	3,67
2009	3,94
2010	4,30
2012	1,70

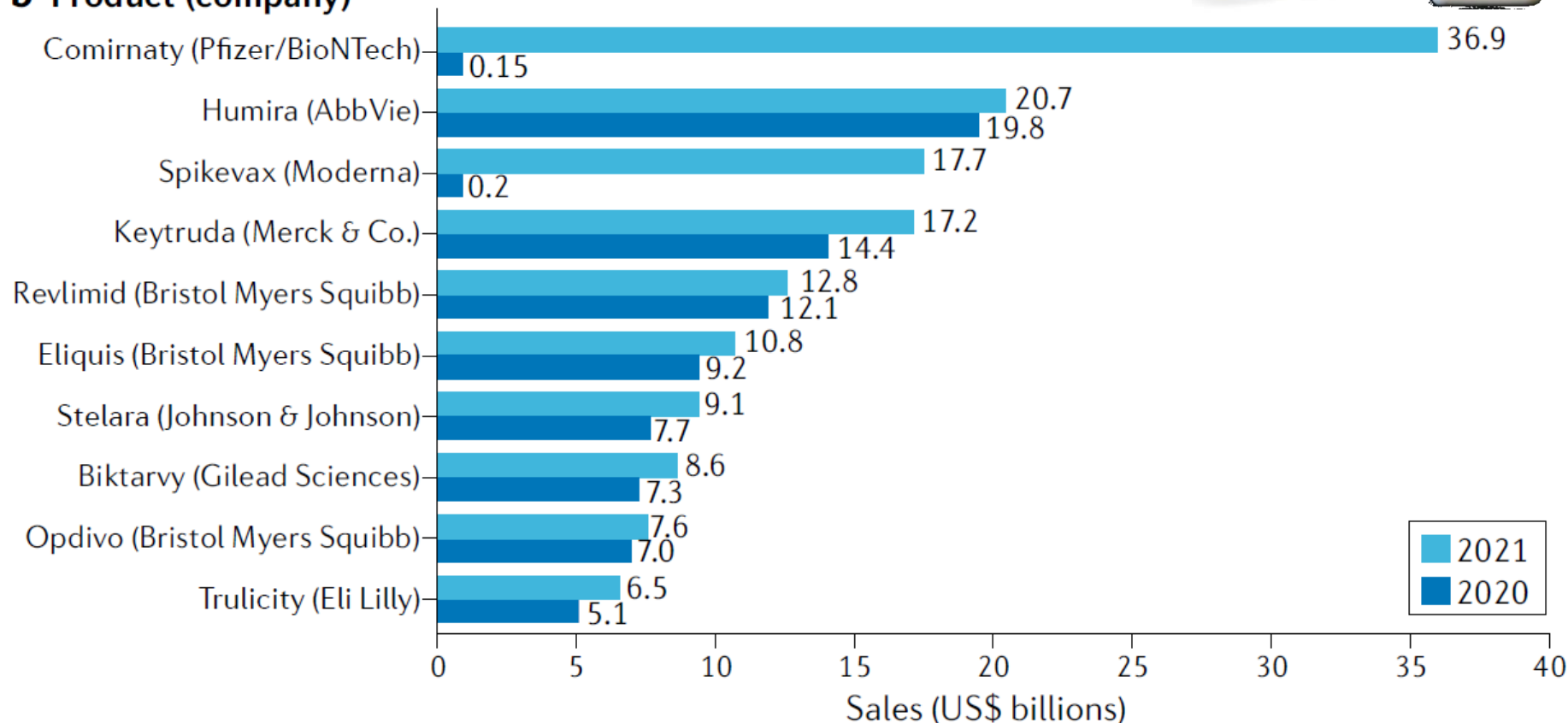


R\$ 15.835,52 / 30 cprs.

Top 10 fármacos/vacinas em 2021



b Product (company)



Por que estudamos proteínas?

Talvez a melhor resposta que possa ser dada seja porque as proteínas sustentam todos os aspectos da atividade biológica.

- As proteínas são essenciais para os organismos
- As proteínas constituem, em média, 15% da massa corporal
 - ✓ Grande parte do tecido do nosso corpo é constituído de proteína.
 - ✓ Proteínas desempenham funções vitais que mantêm o organismo funcionando adequadamente
- Algumas proteínas são economicamente atrativas

