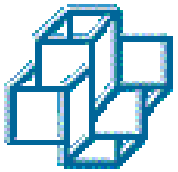
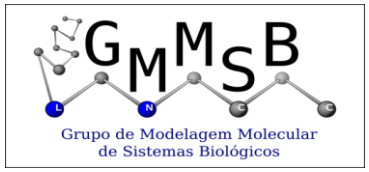
Modelagem Molecular no Planejamento de Fármacos





Laurent E. Dardenne

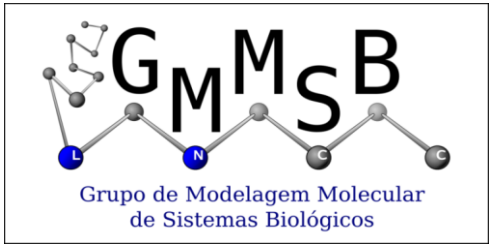
GMMSB – www.gmmsb.lncc.br
Laboratório Nacional de Computação Científica - LNCC/MCTIC

LABORATÓRIO
NACIONAL DE
COMPUTAÇÃO
CIENTÍFICA

inct
inofar

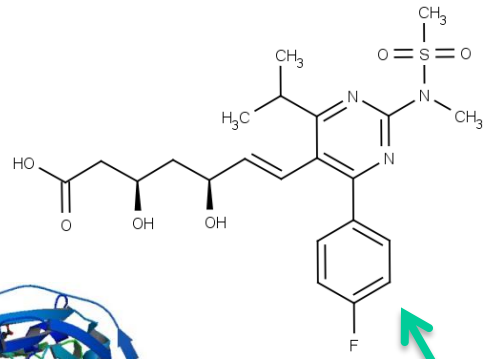
MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES

dockthor



GRUPO DE MODELAGEM MOLECULAR DE SISTEMAS BIOLÓGICOS
LNCC/MCTIC

- **Multidisciplinaridade** (física, biologia, química, computação, …);
- **Desenvolvimento** de métodos, algorítmos, programas e portais;
- **Planejamento de Fármacos** Baseado em Estrutura;
- **Predição de Estruturas e Engenharia de Proteínas**;
- **Projetos Aplicados**: Doenças negligenciadas (Doença de Chagas, Leishmaniosis), Anti-inflamatórios, Câncer, e Alzheimer.

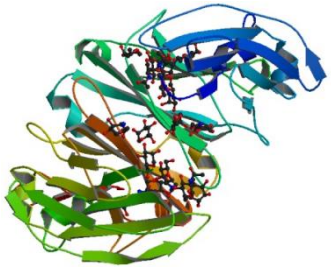


Adalimumab

fluticazona salmeterol

$8.9bi

#1



Rosuvastatina

#2

$8,5bi

Esomeprazol

#3

$8,3bi

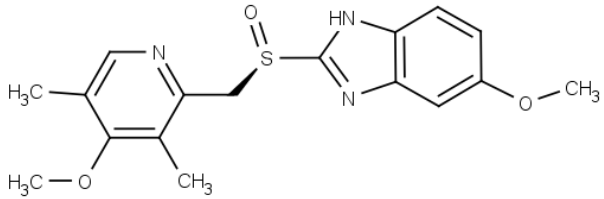
Etanercept

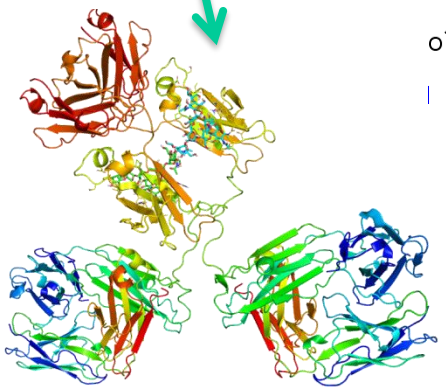
#4

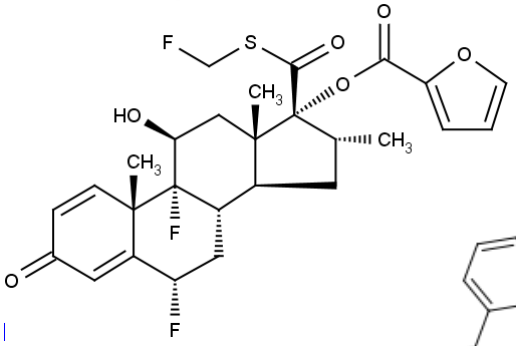
$7.5bi

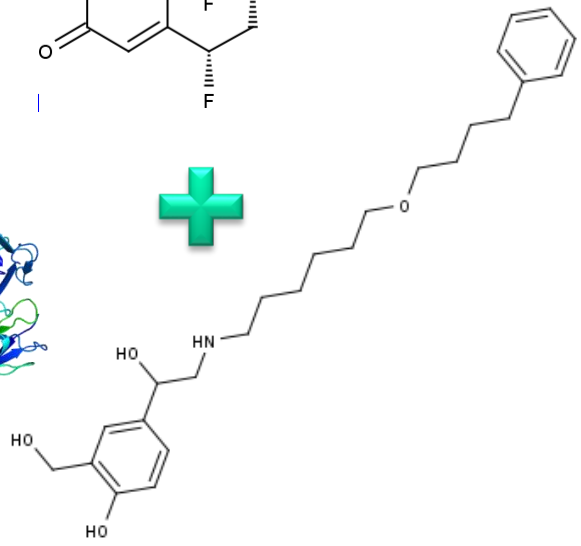
#5

$7.5bi



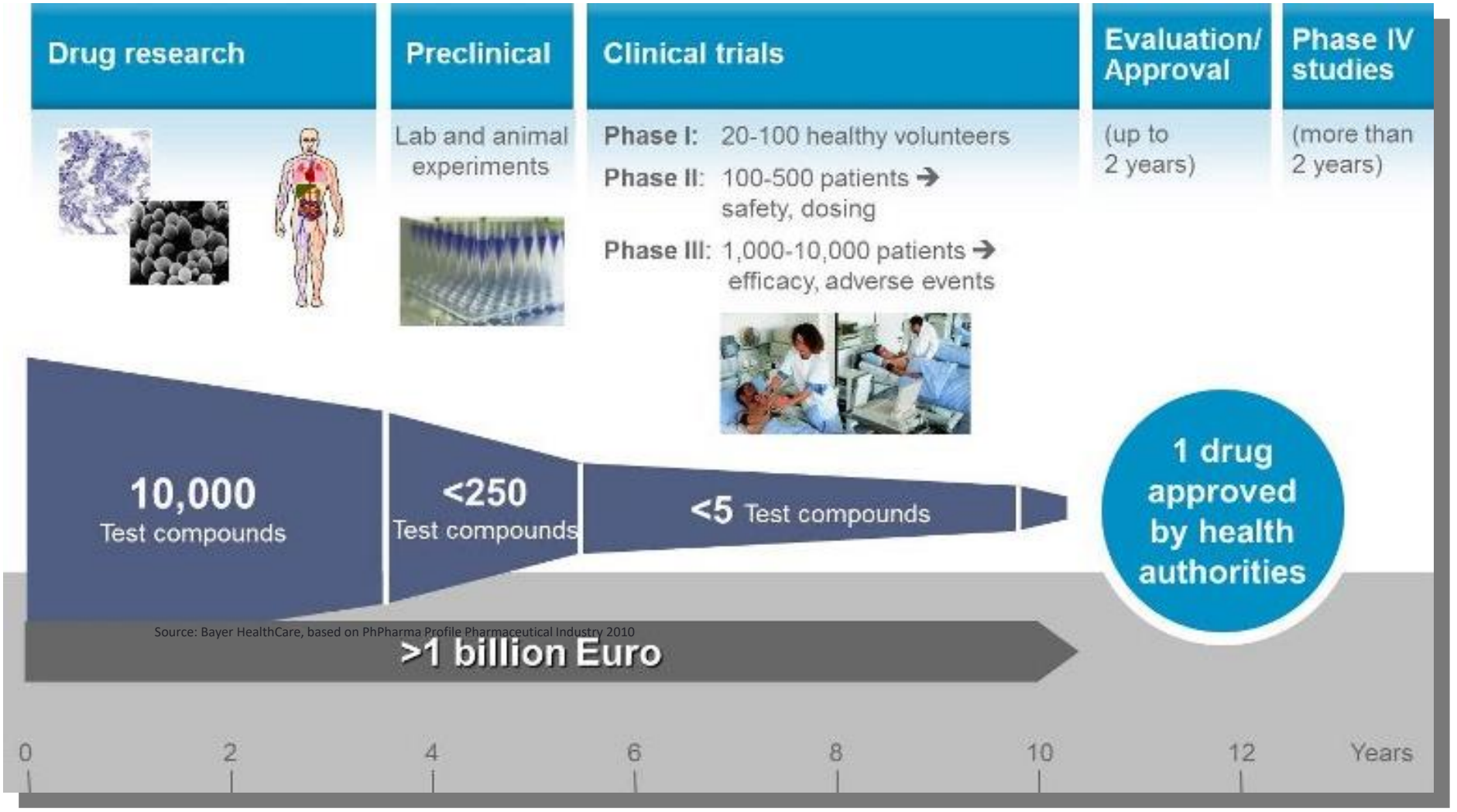






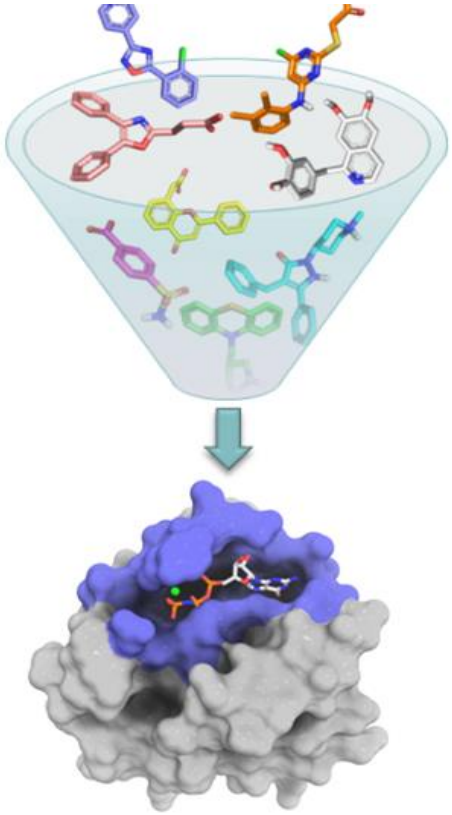
Fonte: IMS Institute for Healthcare Informatics – Dezembro de 2012

Novo fármaco: Custo x Tempo



Source: Bayer HealthCare, based on PhPharma Profile Pharmaceutical Industry 2010

Modelagem Molecular

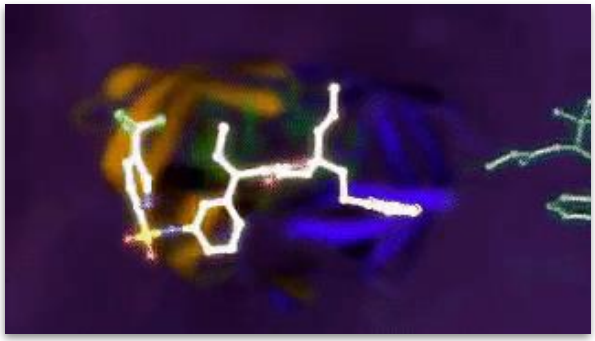


Triagem Virtual em Larga Escala

- Buca por compostos promissores
- Predizer como o alvo terapêutico e moléculas ligantes interagem
- Encontrar alvo terapêutico de compostos conhecidos

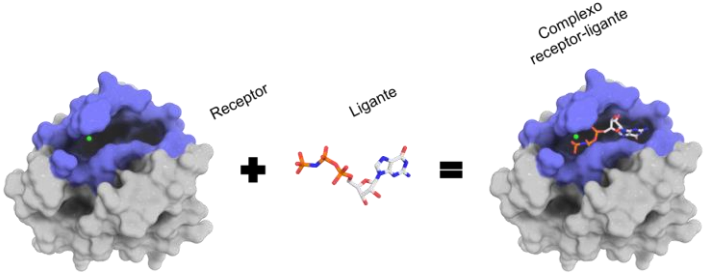
Dinâmica Molecular

- Estabilidade do complexo receptor-ligante
- Caminhos de ligação
- Impacto de mutações na estrutura



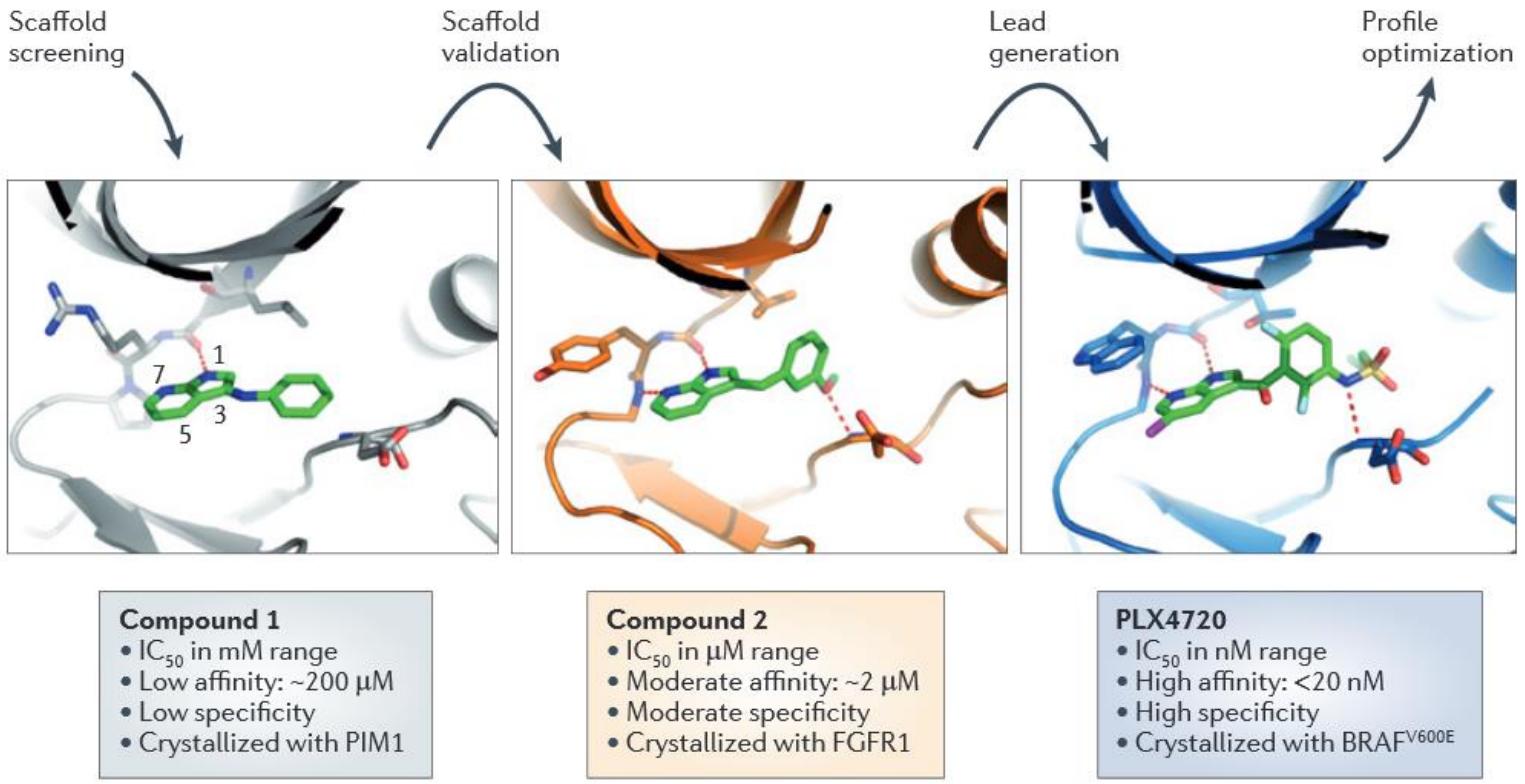
Predição da Afinidade

- Força da interação
- Potência e seletividade



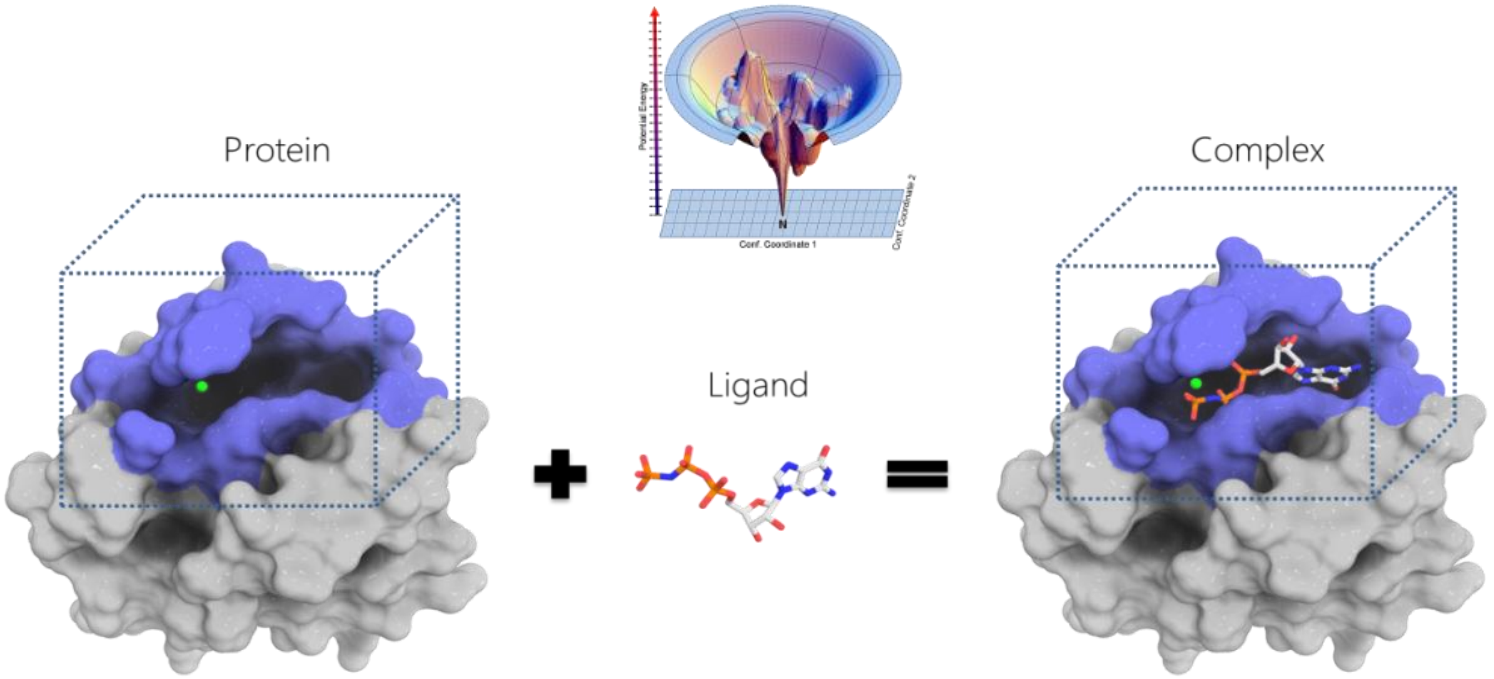
Vemurafenib (Zelboraf, Plexxikon/Roche)

- Câncer ($BRAF^{V600E}$)

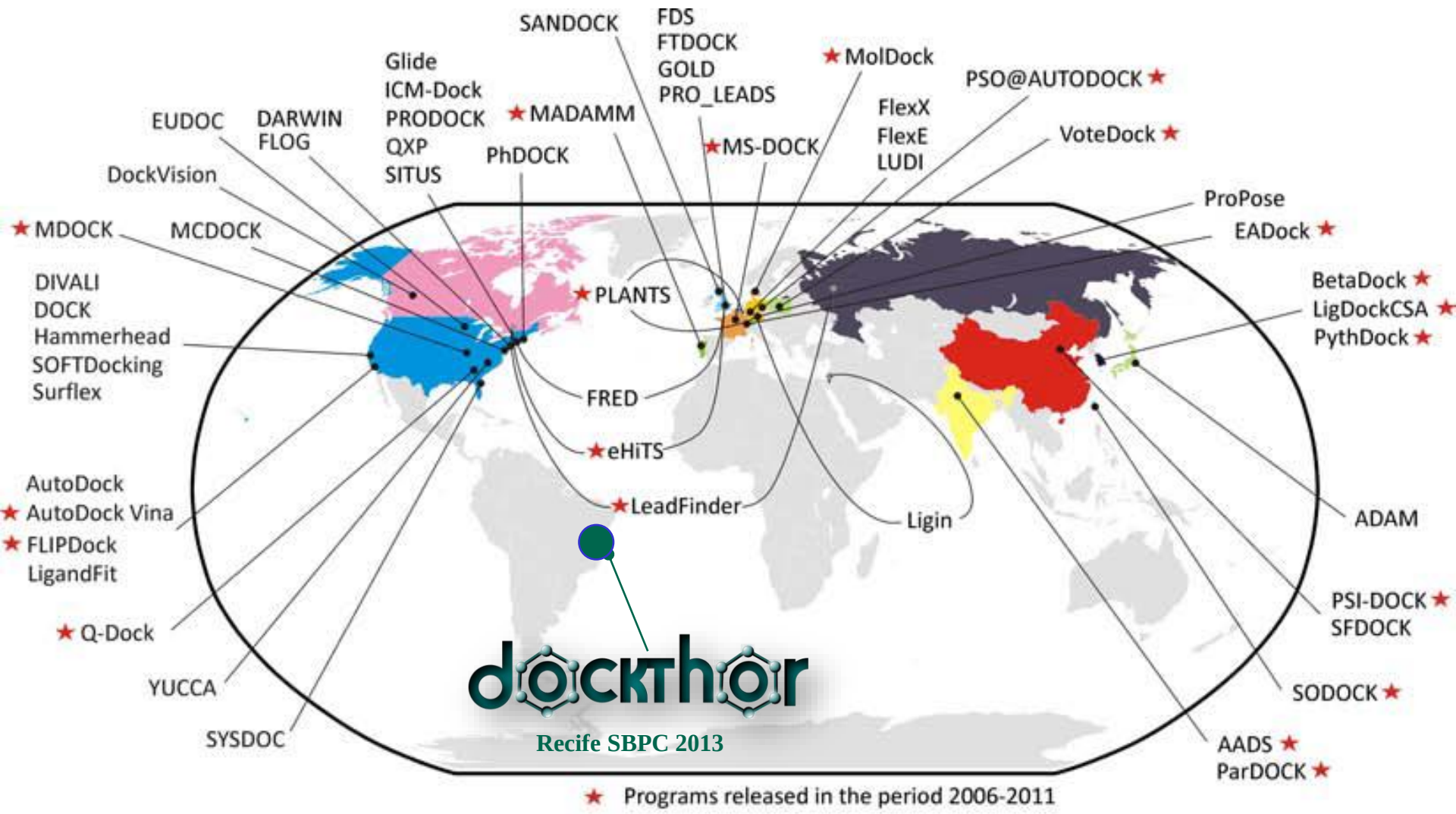


Bollag, G. et al. Vemurafenib: the first drug approved for BRAF-mutant cancer. Nat. Rev. Drug Discov. , 2012.

Docking Molecular



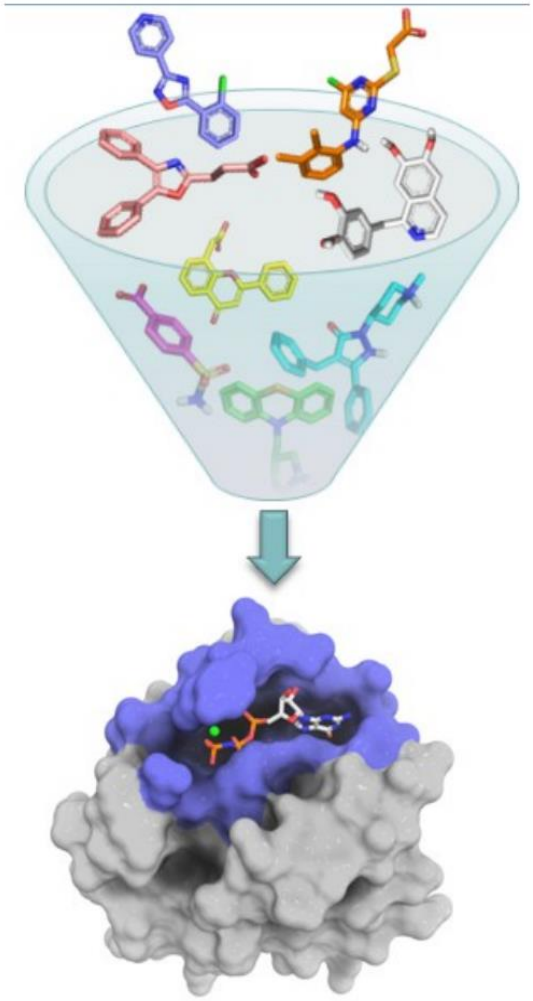
Programas de Docking no Mundo

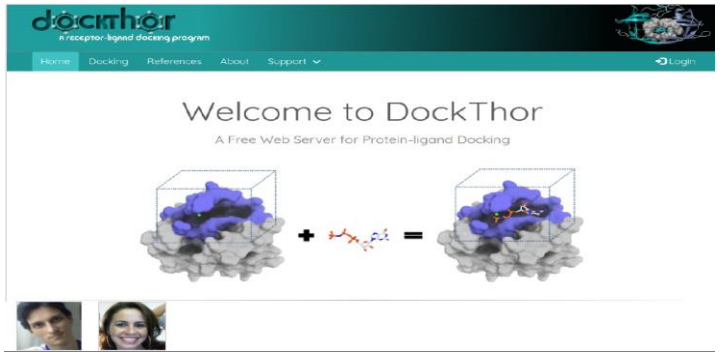


Recife SBPC 2013

Current Medicinal Chemistry. 2013. Vol. 20. No. 18

Portal Web para Triagem Virtual em Larga Escala





Supercomputador Santos Dumont

dockthor
virtual screening

- 1-100 compostos (gratuito e sem cadastro)
- 100 ↔ 1000 compostos (necessita registro/aprovação)
- >1000 compostos (projetos especiais)

Comparative Analysis - Best Pose

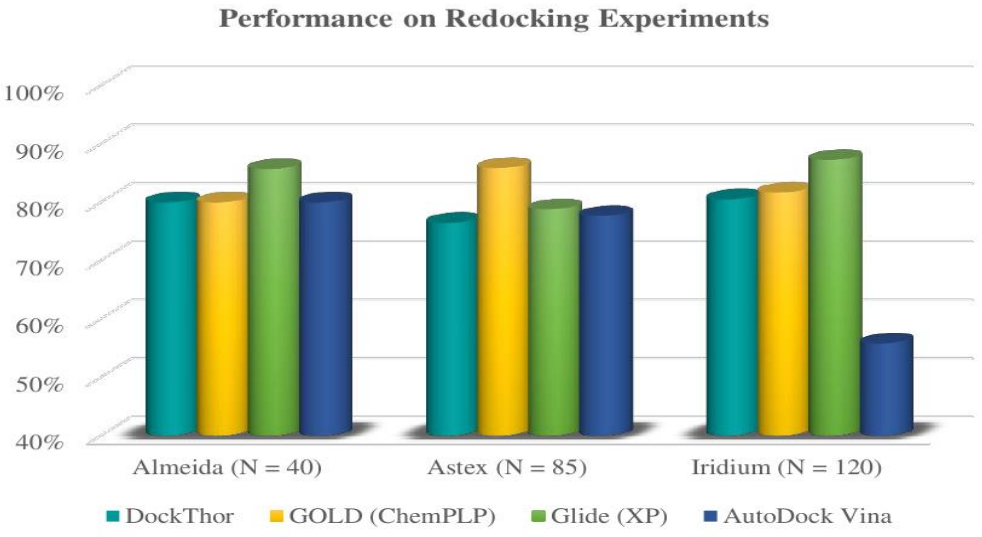
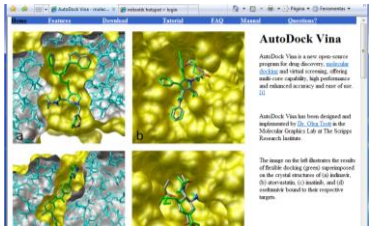
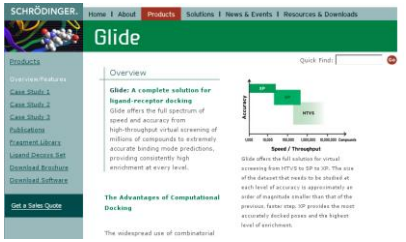
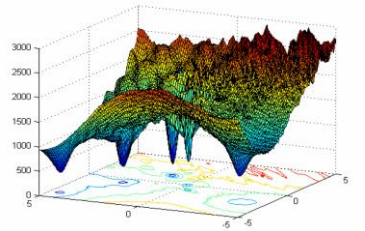
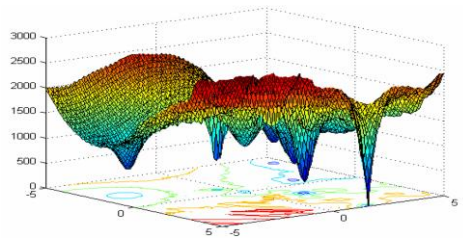


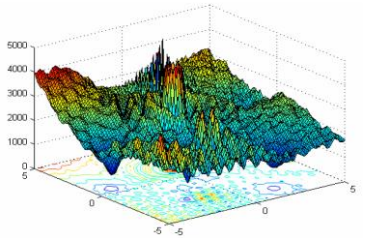
Figure 1. Comparative analyses to achieve performances of DockThor and other docking software widely used by the scientific community, Glide (XP) GOLD (ChemPLP) and AutoDock Vina, in predicting experimental binding modes through redocking experiments.

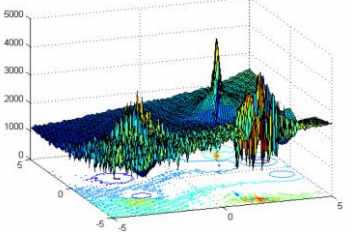


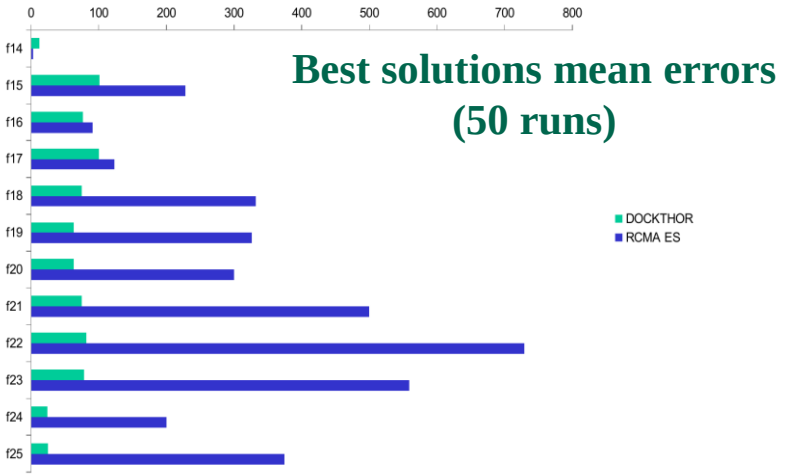








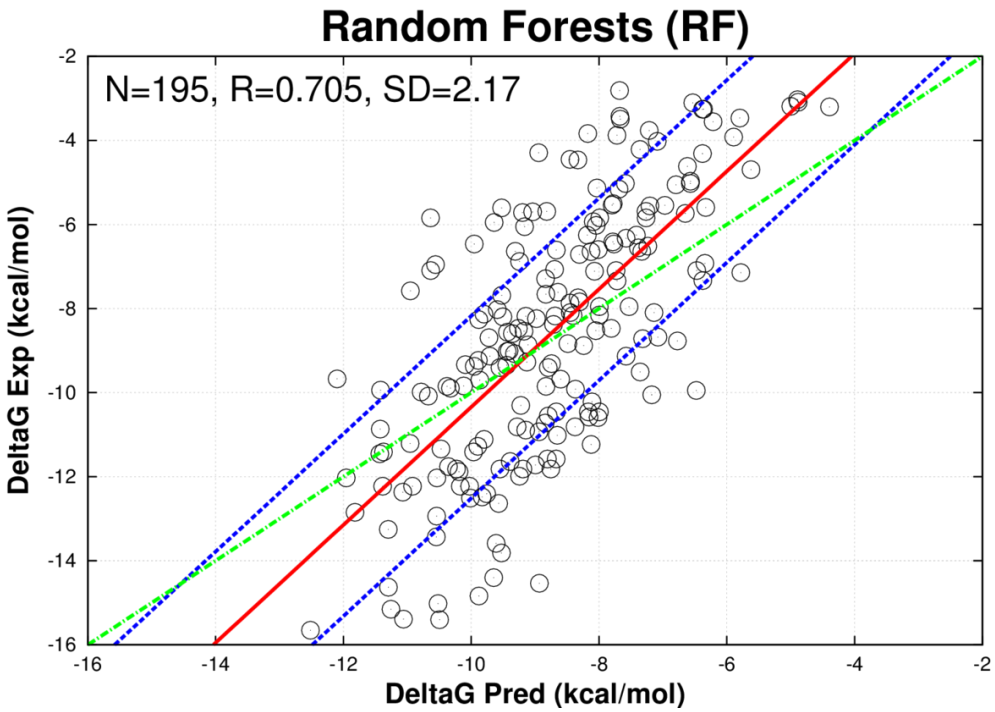




Best solutions mean errors (50 runs)

DockTScore

General Scoring Function
Nonlinear Model



[1] Li, Y. et al. J Chem Inf Model, **2014**.

PDBbind (v2013) [1]	
Scoring function	**R**
RF::VinaElem	0.752
RF::Khamis	0.704
SPA-SE	0.662
SPA	0.656
X-Score::HMScore	0.614
ΔSAS	0.606
ChemScore@SYBYL	0.592
ChemPLP@GOLD	0.579
AutoDock Vina	0.564
ChemScore@GOLD	0.536
FMW_logP	0.495
GoldScore@GOLD	0.483
FMW	0.461
GlideScore-SP	0.452
FlogP	0.321
GlideScore-XP	0.277
PMF@SYBYL	0.221
SPA-SE	0.662
SPA	0.656
X-Score::HMScore	0.614
ΔSAS	0.606

The DockThor Portal: a Free Protein-Ligand Docking Server

Guedes, I.A.[*1], **Krempser, E.**[1], **Marinho, D.**[2], **de Magalhães, C.S.**[3], **Dardenne, L.E.**[1]

[1] *Grupo de Modelagem Molecular de Sistemas Biológicos, Laboratório Nacional de Computação Científica – Petrópolis / RJ - Brasil*

[2] *Instituto de Ciências Biológicas - Universidade Federal do Pará – Pará / PA – Brasil*

[3] *Departamento de Matemática - Universidade Federal Rural do Rio de Janeiro – Rio de Janeiro / RJ - Brasil*

Patent Number **13318-3**

www.dockthor.lncc.br

14411 Jobs submetidos

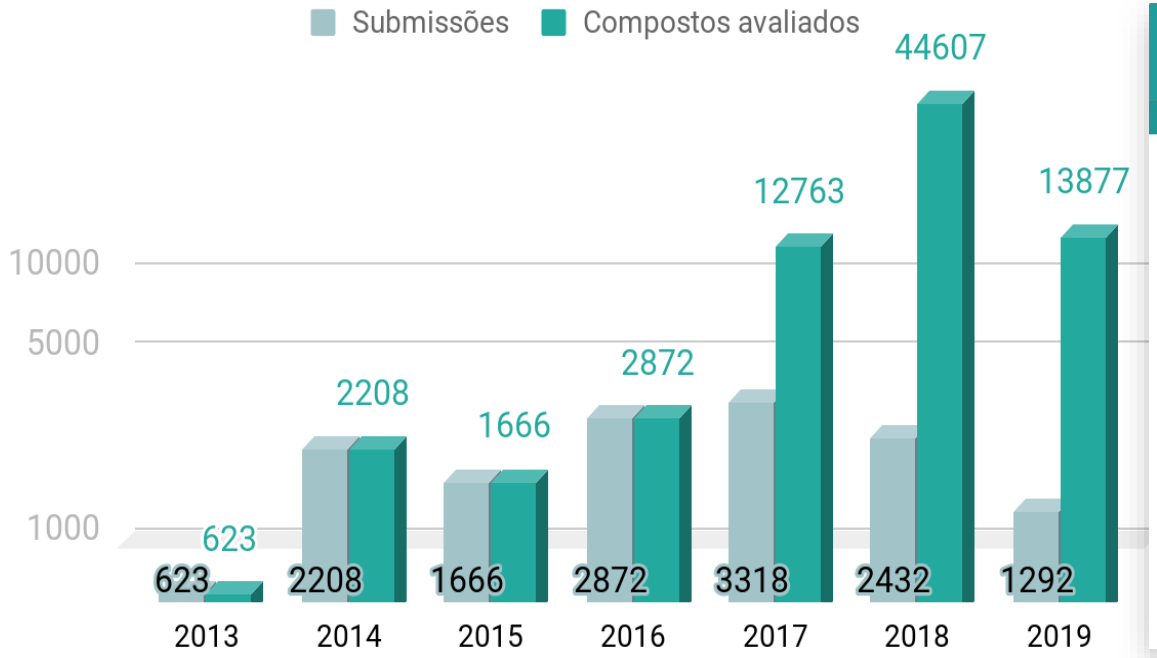
78616 Moléculas avaliadas

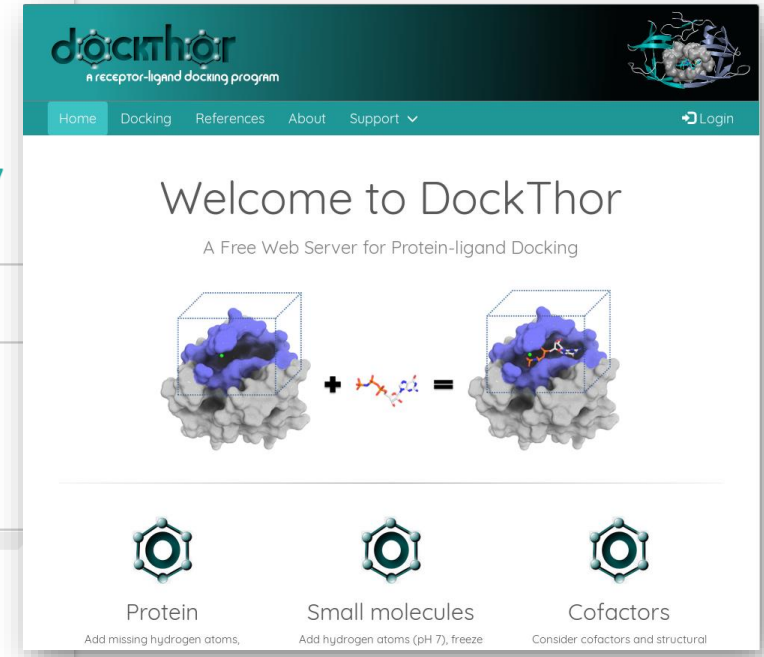
10 moléculas Média por submissão**

*Estatísticas coletadas em 25 de junho de 2019

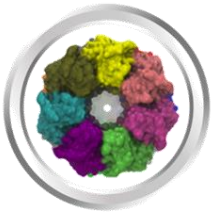
**Nova versão do portal para triagem virtual (2017)

Submissões x Compostos Avaliados





Plataforma DockThor-VS



- Terapeutic Targets
- Small molecules
- ...

- Cancer
- Neglected Diseases
- Alzheimer
- ...

Databases

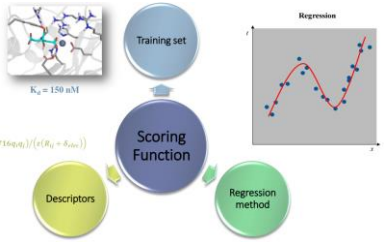
Developments

dockthor VIRTUAL SCREENING

Applicability

Collaborations

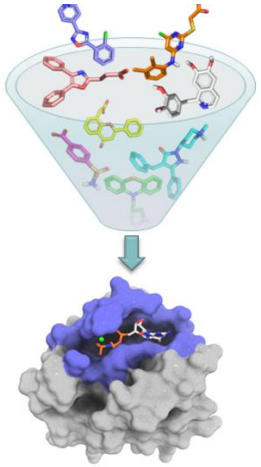
- Search algorithm
- Quantum refinement
- GPU Programming



- Protein flexibility
- Multiple binding modes
- Specific scoring functions

Machine learning:
- Affinity prediction
- *de novo* design

- INCT-Inofar
- FIOCRUZ
- Embrapa
- ...



Contents lists available at ScienceDirect

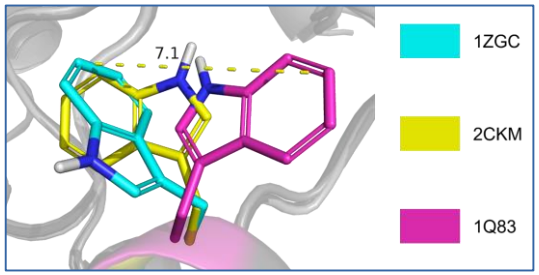
European Journal of Medicinal Chemistry

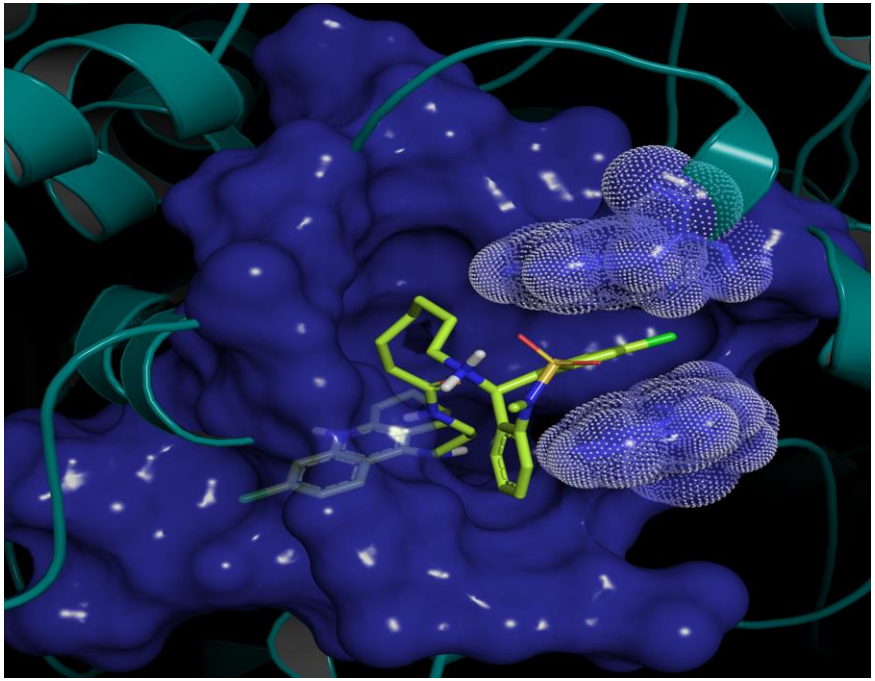
journal homepage: http://www.elsevier.com/locate/ejmech

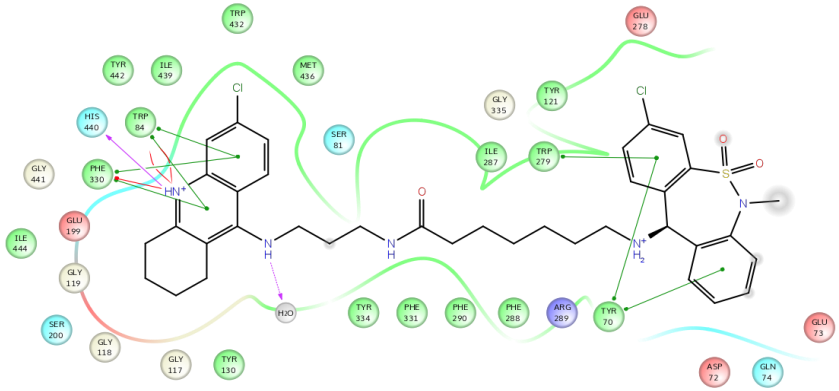
Research paper

Novel series of tacrine-tianeptine hybrids: Synthesis, cholinesterase inhibitory activity, S100B secretion and a molecular modeling approach

Marco Antonio Ceschi [a, *], Jessie Sobieski da Costa [a], João Paulo Bizarro Lopes [a], Viktor Saraiva Câmara [a], Leandra Franciscato Campo [a], Antonio César de Amorim Borges [a], Carlos Alberto Saraiva Gonçalves [b], Daniela Fraga de Souza [b], Eduardo Luis Konrath [c], Ana Luiza Martins Karl [d], Isabella Alvim Guedes [d], Laurent Emmanuel Dardenne [d, **]







INPI Patent deposi 2015
(20% LNCC and 80% UFRGS)

Compounds for Alzheimer Disease

14º Prêmio CNPq Destaque de Iniciação Científica na Área de Ciências da Vida - 2016

Chronic Inflammatory Diseases

Molecular Modeling of IKK2 Target: Structural and Ligand Binding Properties Studies

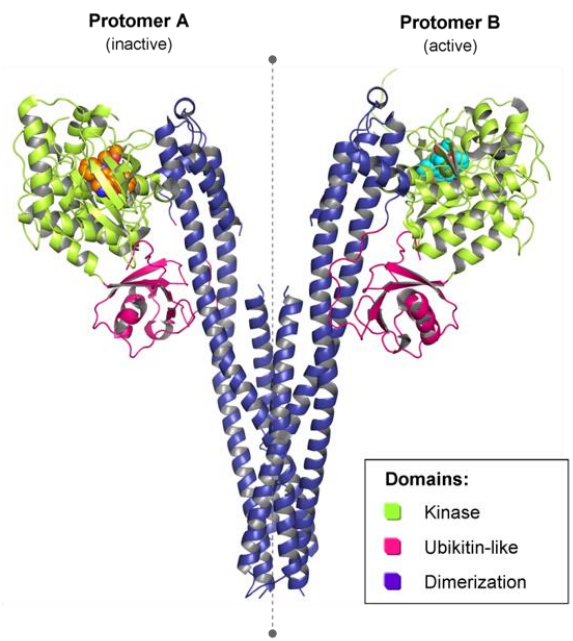
Isabella Alvim Guedes* [1], Carlos A. Manssour Fraga [2], Laurent Emmanuel Dardenne [1]

[1] Grupo de Modelagem Molecular de Sistemas Biológicos, Laboratório Nacional de Computação Científica

[2] Laboratório de Avaliação de Substâncias Bioativas, Faculdade de Farmácia - Universidade Federal do Rio de Janeiro

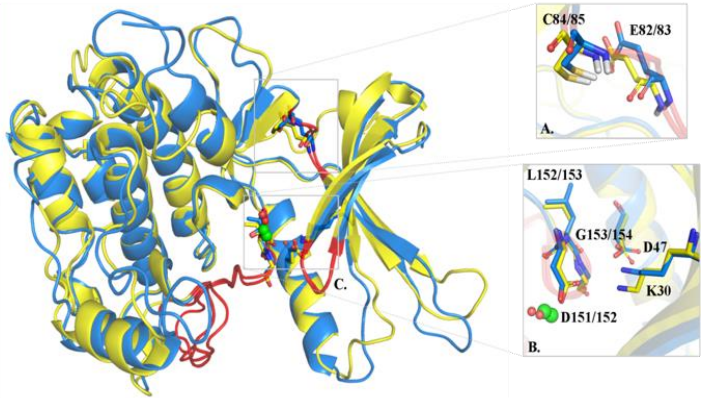
Email: *isabella@lncc.br, cmfraga@ccsdecania.ufrj.br, dardenne@lncc.br

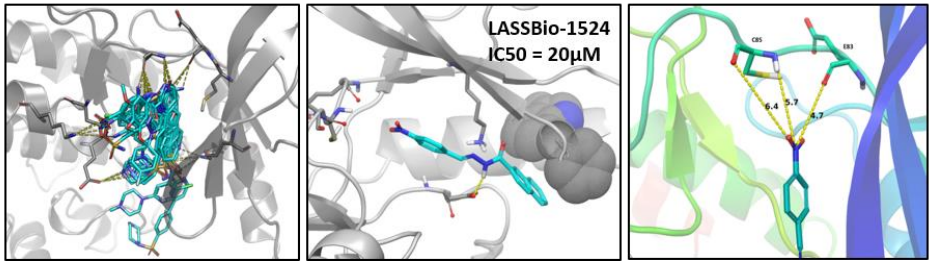
Work supported by: CNPq, FAPERJ, INCT-INOFAR.
Contract grants nº. E26/102.443/2009, E26/170.020/2008 and 573.564/2008-6.



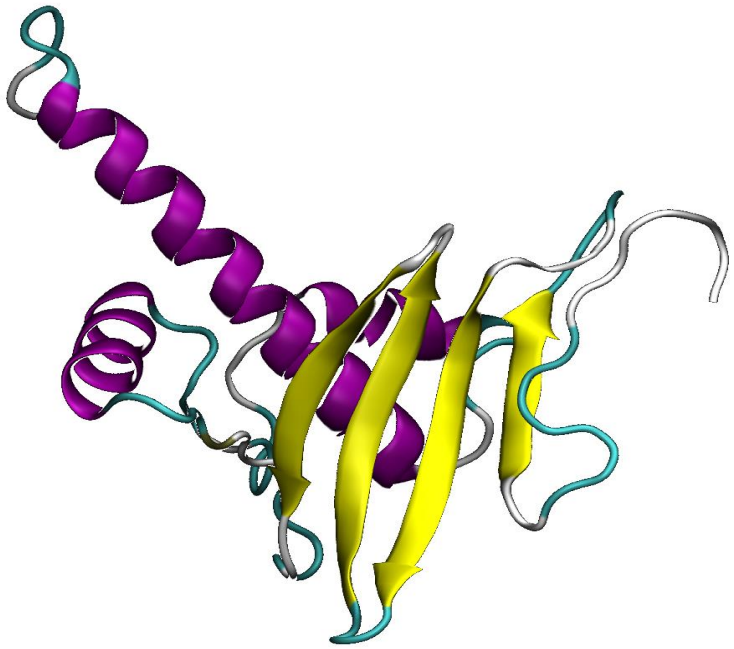
First prototype **LASSBio-1524** IC50 = 20µM

New prototype **LASSBio-1829** IC50 = 3µM





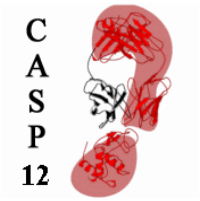
GAPF: Genetic Algorithm for Protein Folding



Predição de Estruturas de Proteínas

- **Estrutura -> Função**
- **Engenharia de Proteínas** (**Biofármacos**, processos industriais, biotecnologia)
- Importante para os estudo de doenças causadas pelovmal enovelamento de uma proteína (**doenças neurodegenerativas**: **Alzheimer's, Parkinson's, e amyotrophic lateral sclerosis**

CASP – Critical Assessment of Techniques for Protein Prediction



Biannual Community Wide Experiment

CASP12 -GAPF_LNCC_SERVER

T0942 - 487 amino acids

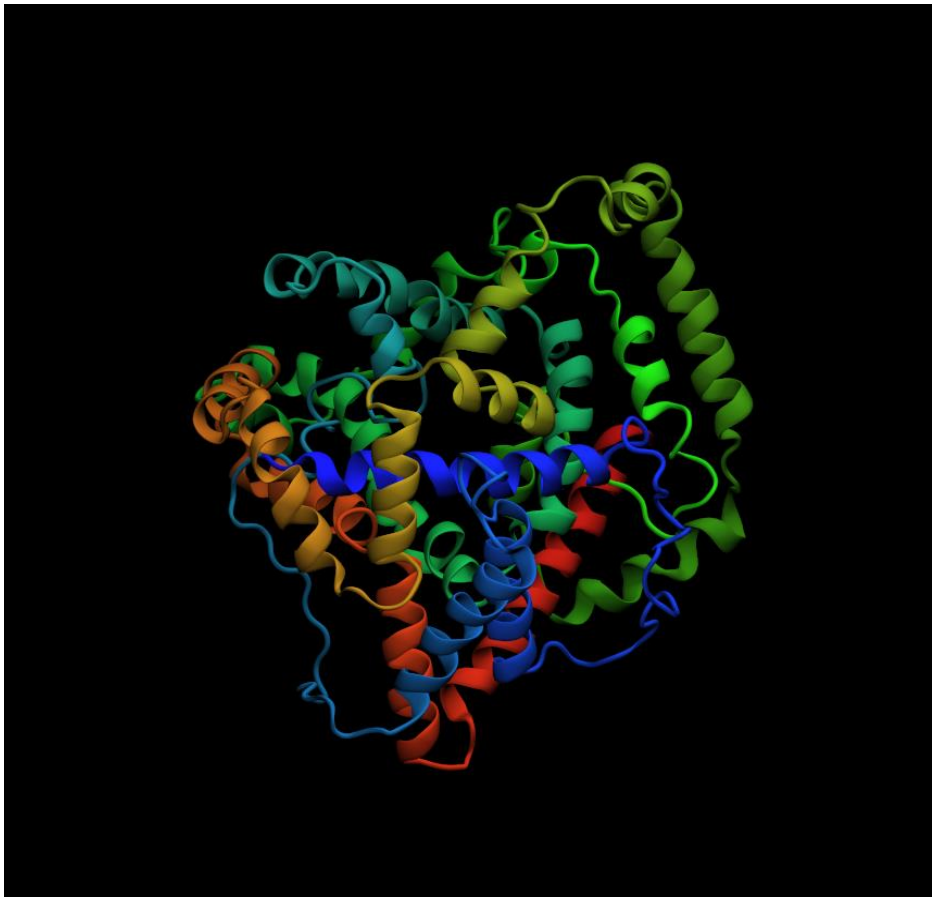
72 hours deadline

> 4.990 CPU hours…

(~7 months)

Using Santos Dumont:

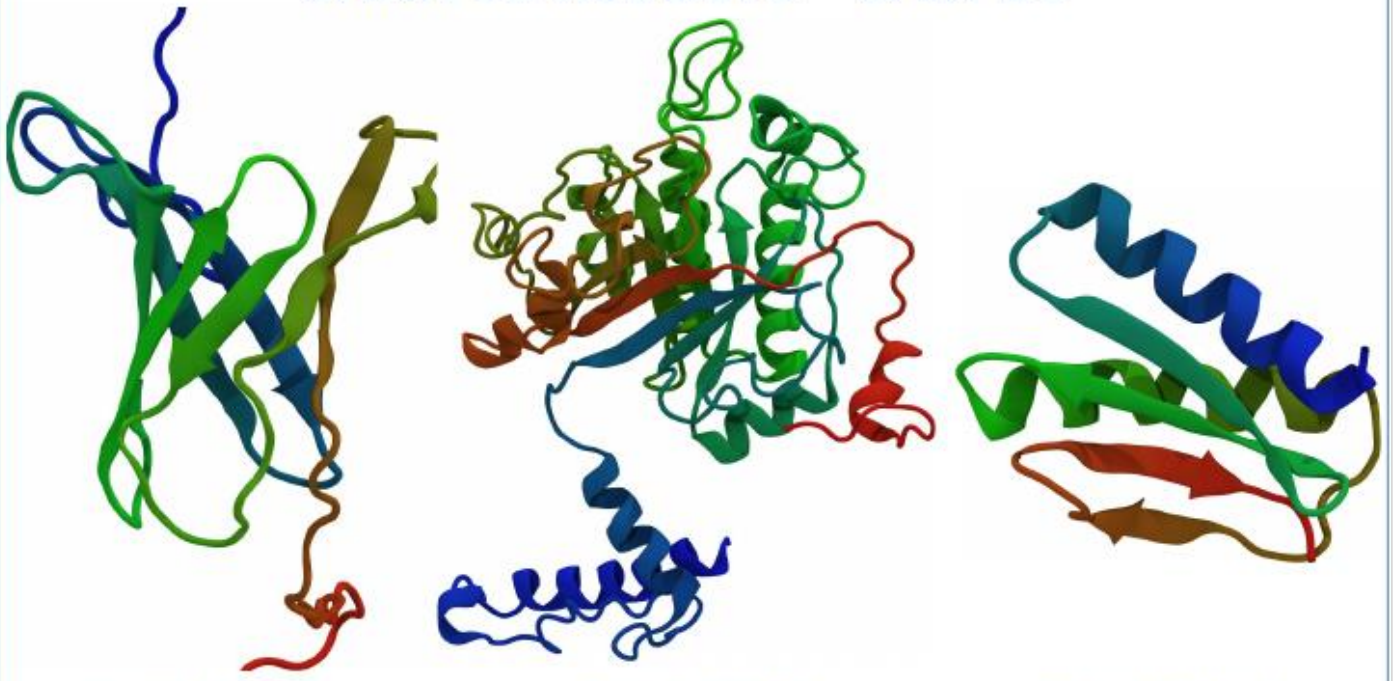
9.600 cores: ~ 8 hours



GAPF_LNCC

Foram submetidas predições de estruturas de proteínas para 56 alvos (< 200 resíduos) (Santos Dumont)

SAMPLE RESULTS - CASP13



T0992 (FM) T1005 (TBM) T1008 (TMB?)

Templates used for target T1008 with relatively high e-values (in bold).

1	3gr7_A	NADPH dehydrogenase; fl	87.4	**0.018**	3.7E-06	35.2	0.0	68	2-69	79-166 (340)
2	1z41_A	YQJM, probable NADH-dep	85.4	**0.029**	5.8E-06	34.3	0.0	68	2-69	79-166 (338)
3	3kru_A	NADH:flavin oxidoreduct	83.6	**0.04**	8.1E-06	33.7	0.0	68	3-70	78-166 (343)

GAPF_LNCC: an automated method for protein structure prediction with a multiple minima genetic algorithm

Fábio Lima Custódio[1] and Laurent E. Dardenne[2]

Laboratório Nacional de Computação Científica, Petrópolis--RJ, Brasil

[1]flc@lncc.br; [2]dardenne@lncc.br

Laboratório Nacional de Computação Científica

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES — MCTIC

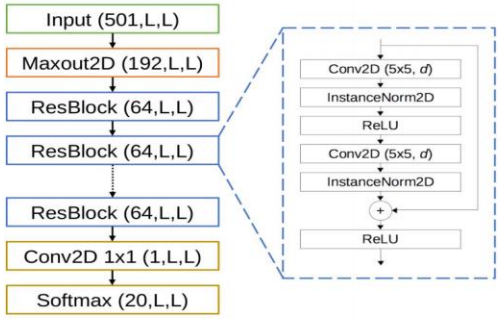


Figure 2 DMPfold model architecture. DMPfold is a deep, fully convolutional residual network. There are a total of 18 residual blocks.

Aspectos relevantes do CASP13:

- **DeepMind** venceu o CASP13!
- Outros grupos também se saíram bem utilizando Deep Learning!
- Rede Neural Profunda com cerca de **400 camadas e 21 milhões de parâmetros**!
- **É possível que avanços utilizando Deep Learning, nas mais diversas áreas, sejam mais significativos e rápidos do que se espera!**

Google's DeepMind predicts 3D shapes of proteins

The Guardian

AI program's understanding of proteins could usher in new era of medical progress

▲ Google's DeepMind artificial intelligence program, AlphaGo, plays South Korean professional Go player, Lee Sedol. Photograph: Ahn Young-joon/AP

BIOLOGICAL STRUCTURE AND FUNCTION EMERGE FROM SCALING UNSUPERVISED LEARNING TO 250 MILLION PROTEIN SEQUENCES

Alexander Rives [*†‡] **Siddharth Goyal** [*§] **Joshua Meier** [*§] **Demi Guo** [*§]
Myle Ott [§] **C. Lawrence Zitnick** [§] **Jerry Ma** [†§] **Rob Fergus** [†‡§]

ABSTRACT

In the field of artificial intelligence, a combination of scale in data and model capacity enabled by unsupervised learning has led to major advances in representation learning and statistical generation. In biology, the anticipated growth of sequencing promises unprecedented data on natural sequence diversity. Learning the natural distribution of evolutionary protein sequence variation is a logical step toward predictive and generative modeling for biology. To this end we use unsupervised learning to train a deep contextual language model on 86 billion amino acids across 250 million sequences spanning evolutionary diversity. The resulting model maps raw sequences to representations of biological properties without labels or prior domain knowledge. The learned representation space organizes sequences at multiple levels of biological granularity from the biochemical to proteomic levels. Unsupervised learning recovers information about protein structure: secondary structure and residue-residue contacts can be identified by linear projections from the learned representations. Training language models on full sequence diversity rather than individual protein families increases recoverable information about secondary structure. The unsupervised models can be adapted with supervision from quantitative mutagenesis data to predict variant activity. Predictions from sequences alone are comparable to results from a state-of-the-art model of mutational effects that uses evolutionary and structurally derived features.

[*]Equal contribution
[†]Correspondence to <arives@cs.nyu.edu>, <maj@fb.com>, and <robfergus@fb.com>
[‡]Dept. of Computer Science, New York University, USA
[§]Facebook AI Research, USA

Contribuições do GMMSB na área de Inteligência Artificial

- Métodos de Otimização para problemas complexos.

- Desenvolvimento de ferramentas baseadas em técnicas de aprendizagem de máquina.

- *Deep Learning*.

Grupo de Modelagem Molecular de Sistemas Biológicos

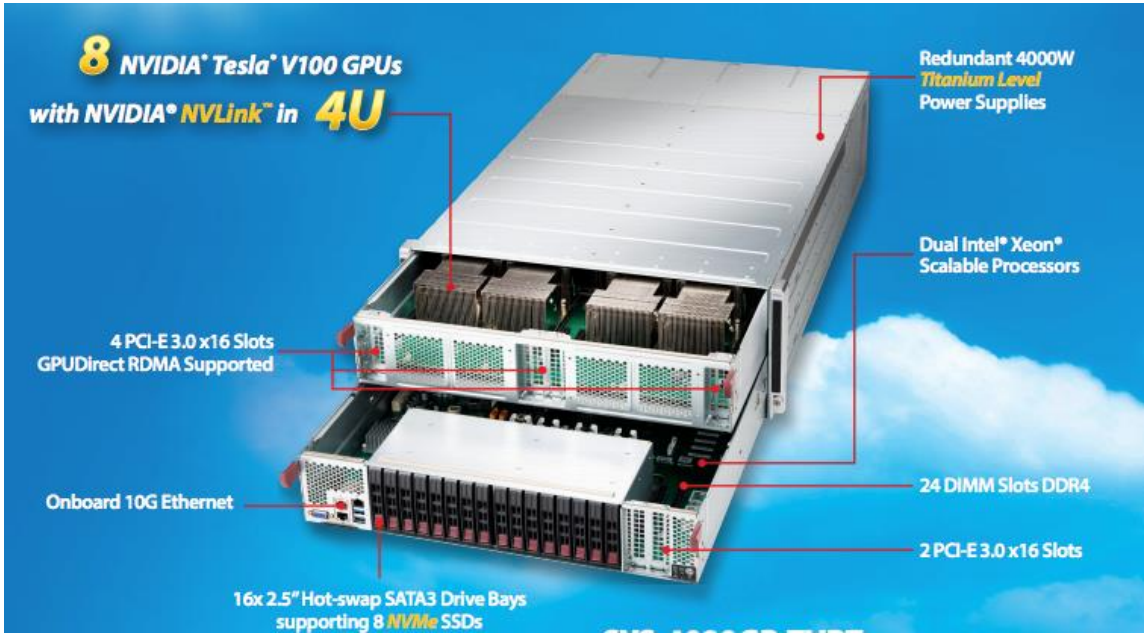
Pesquisas na área de Aprendizagem de Máquina

- Predição de estruturas de proteínas e função (Deep Learning)

- Predição de afinidade receptor-ligante visando o planejamento de fármacos (técnicas usuais e Deep Learning)

- Predição de imunogenicidade (epítopos) utilizando Deep Learning. Primeira etapa na indústria de biofármacos para o desenvolvimento de anticorpos e vacinas. Resultados preliminares mostram que a ferramenta desenvolvida é competitiva e já está sendo utilizada em estudos de casos.

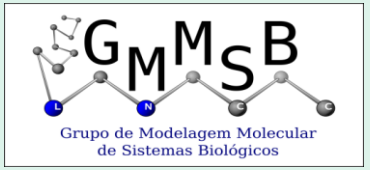
Equipamento para treinamento de redes neurais profundas (deep learning) – Acoplada ao Santos Dumont - LNCC



Características Principais:

- Artificial Intelligence
  - Big Data Analytics
  - High-performance Computing
  - Research Lab/National Lab
  - Astrophysics, Business Intelligence ...

Quais possíveis projetos em colaboração com empresas?



- **Desenvolvimento de compostos protótipos candidatos a fármacos.**
- **Desenvolvimento de biofármacos (possível colaboração com Biomanguinhos –FIOCRUZ)**
- **Desenvolvimento de produtos que possam ser importantes no desenvolvimento de fármacos e biofármacos utilizando inteligência artificial**
- **Desenvolvimento de produtos utilizando técnicas de inteligência artificial em "qualquer" área do conhecimento.**

Laboratório Nacional de Computação Científica
LNCC/MCTI

LABORATÓRIO
NACIONAL DE
COMPUTAÇÃO
CIENTÍFICA

GMMSB
Grupo de Modelagem Molecular
de Sistemas Biológicos

Pós-graduação em Modelagem Computacional - LNCC
Nível 6 – Comitê Interdisciplinar da CAPES

w w w . l n c c . br
Petrópolis - RJ

www.gmmsb.lncc.br