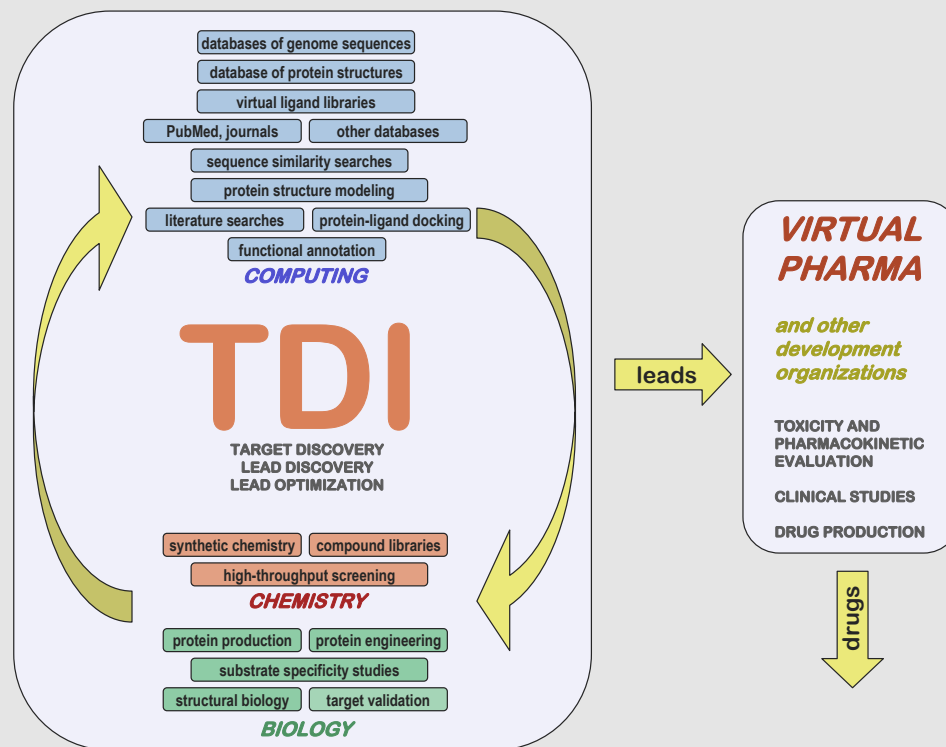


# The Tropical Disease Initiative

*Comparative docking on protein structure models from ten tropical disease genomes.*



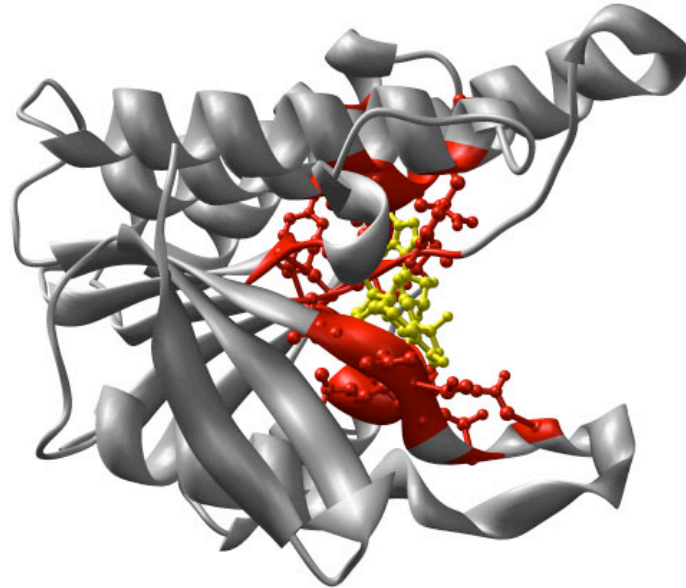
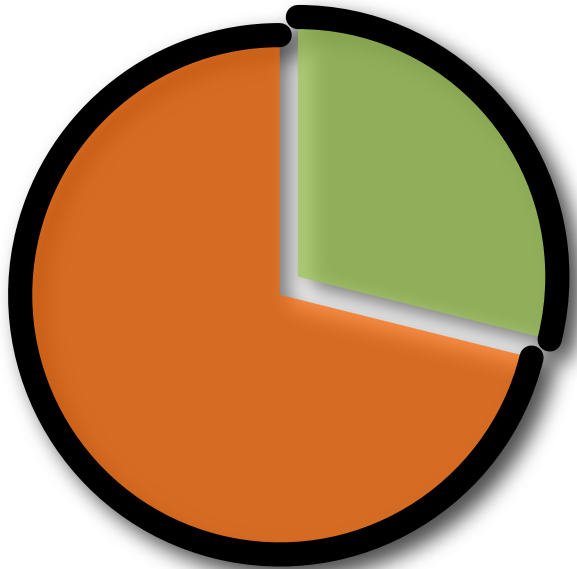
Marc A. Marti-Renom

<http://bioinfo.cipf.es/squ/>

Structural Genomics Unit  
Bioinformatics Department  
Prince Felipe Research Center (CIPF), Valencia, Spain



# Predicting binding sites in protein structure models.

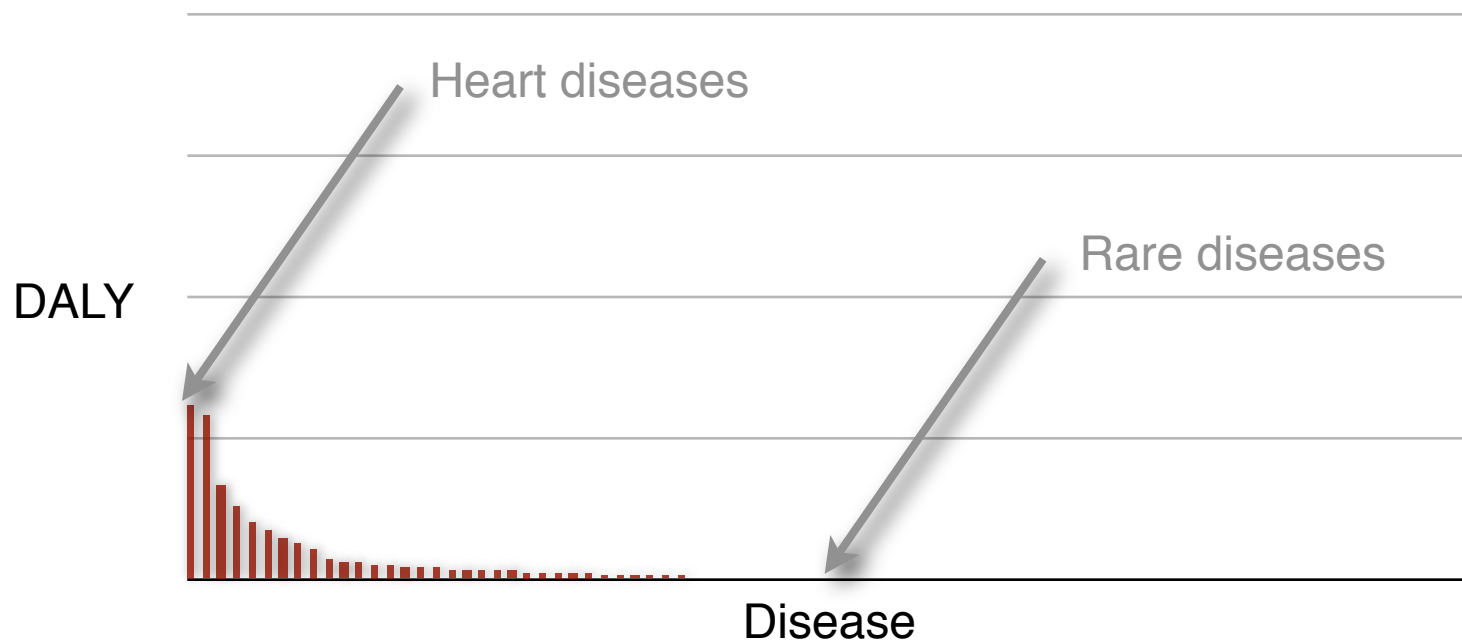


UCSF



# Need is High in the Tail

- DALY Burden Per Disease in Developed Countries
- DALY Burden Per Disease in Developing Countries



Disease data taken from WHO, *World Health Report 2004*

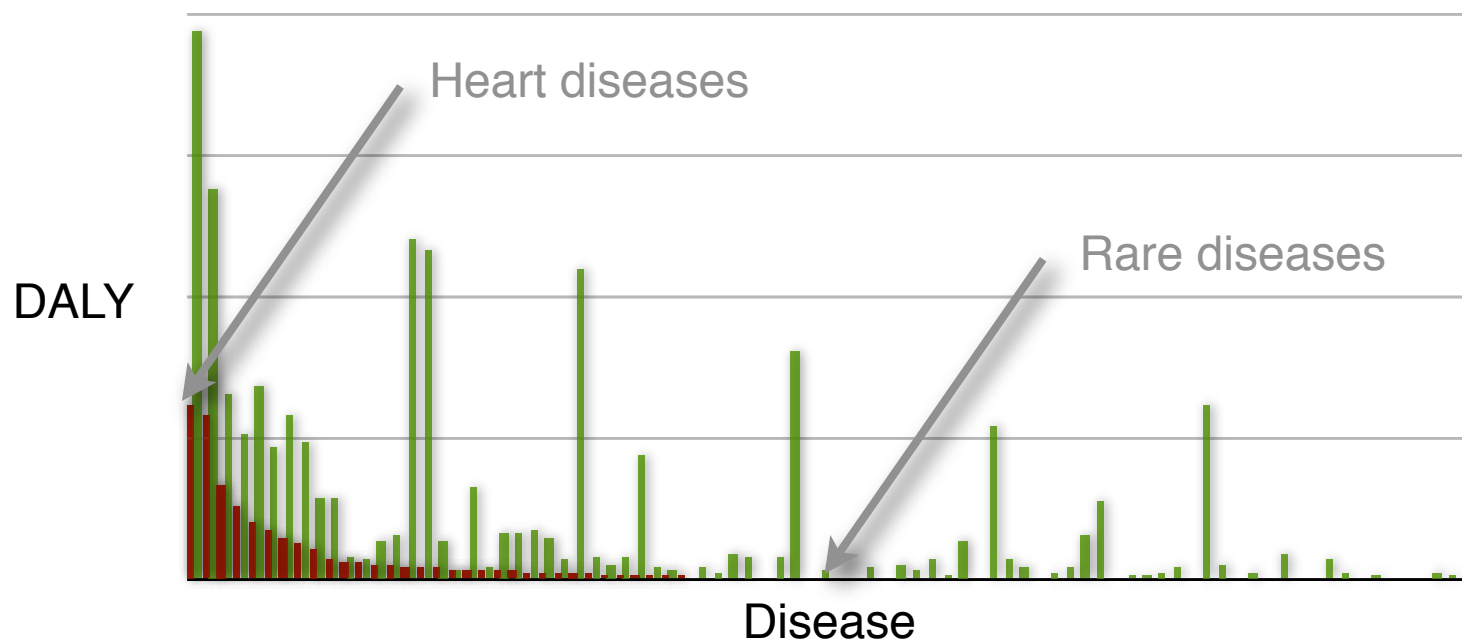
DALY - Disability adjusted life years

DALY is not a perfect measure of market size, but is certainly a good measure for importance.

*DALYs for a disease are the sum of the years of life lost due to premature mortality (YLL) in the population and the years lost due to disability (YLD) for incident cases of the health condition. The DALY is a health gap measure that extends the concept of potential years of life lost due to premature death (PYLL) to include equivalent years of 'healthy' life lost in states of less than full health, broadly termed disability. One DALY represents the loss of one year of equivalent full health.*

# Need is High in the Tail

- DALY Burden Per Disease in Developed Countries
- DALY Burden Per Disease in Developing Countries



Disease data taken from WHO, *World Health Report 2004*

DALY - Disability adjusted life years

DALY is not a perfect measure of market size, but is certainly a good measure for importance.

*DALYs for a disease are the sum of the years of life lost due to premature mortality (YLL) in the population and the years lost due to disability (YLD) for incident cases of the health condition. The DALY is a health gap measure that extends the concept of potential years of life lost due to premature death (PYLL) to include equivalent years of 'healthy' life lost in states of less than full health, broadly termed disability. One DALY represents the loss of one year of equivalent full health.*

# “Unprofitable” Diseases and Global DALY (in 1000’s)

<b>Malaria*</b>	<b>46,486</b>
Tetanus	7,074
<b>Lymphatic filariasis*</b>	<b>5,777</b>
Syphilis	4,200
Trachoma	2,329
<b>Leishmaniasis*</b>	<b>2,090</b>
Ascariasis	1,817
<b>Schistosomiasis*</b>	<b>1,702</b>
<b>Trypanosomiasis*</b>	<b>1,525</b>

Trichuriasis	1,006
Japanese encephalitis	709
<b>Chagas Disease*</b>	<b>667</b>
<b>Dengue*</b>	<b>616</b>
<b>Onchocerciasis*</b>	<b>484</b>
<b>Leprosy*</b>	<b>199</b>
Diphtheria	185
Poliomyelitis	151
Hookworm disease	59

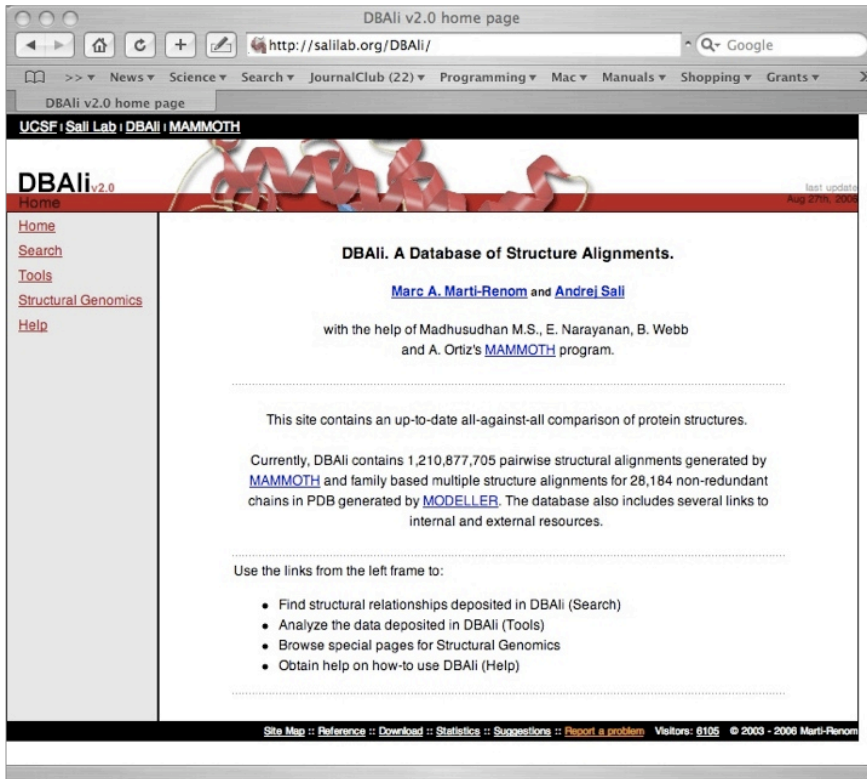
Disease data taken from WHO, *World Health Report 2004*

DALY - Disability adjusted life year in 1000’s.

\* Officially listed in the WHO Tropical Disease Research [disease portfolio](#).

# DBAli<sub>v2.0</sub> database

<http://www.dbali.org>



- ✓ Fully-automatic
- ✓ Data is kept up-to-date with PDB releases
- ✓ Tools for “on the fly” classification of families.
- ✓ Easy to navigate
- ✓ Provides tools for structure analysis

Does not provide a stable classification similar to that of CATH or SCOP

Pairwise structure alignments	
Last update:	October 6th, 2007
Number of chains:	96,804
Number of structure-structure comparisons:*	1,748,371,897
Multiple structure alignments	
Last update:	August 1st, 2007
Number of representative chains:	34,637
Number of families:	12,732

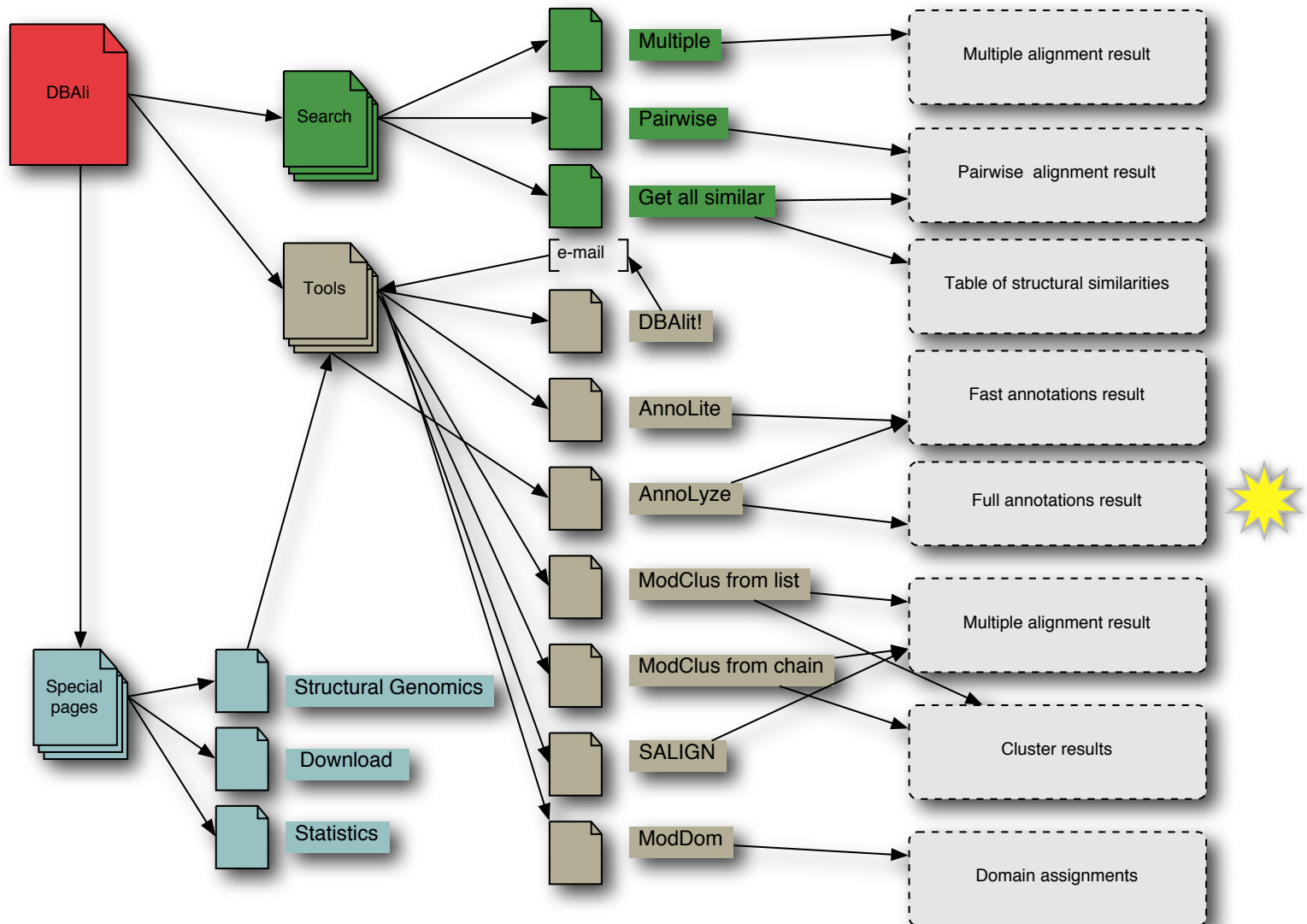
Uses MAMMOTH for similarity detection

- ✓ VERY FAST!!!
- ✓ Good scoring system with significance

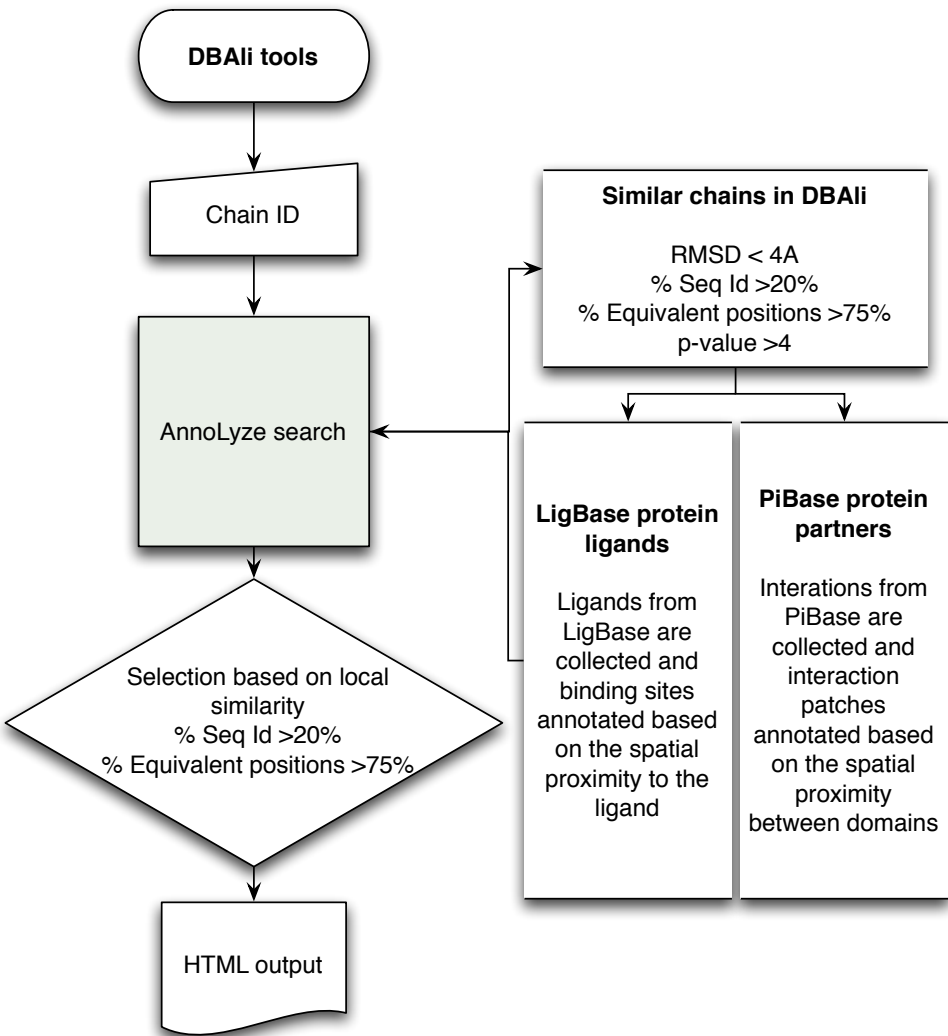
Ortiz AR, (2002) *Protein Sci.* 11 pp2606  
Marti-Renom et al. 2001. *Bioinformatics.* 17, 746

# DBAli<sub>v2.0</sub> database

<http://www.dbali.org>

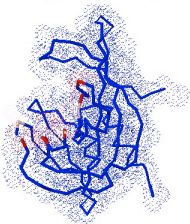


# Method



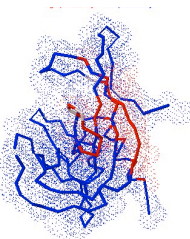
Inherited ligands: 4

Ligand	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
<a href="#">MO2</a>	59.03	<a href="#">0.185</a>	48 49 52 62 63 66 67 113 116
<a href="#">CRY</a>	20.00	<a href="#">0.111</a>	23 29 31 37 44 48 49 83 85 94 96 103 121
<a href="#">8OG</a>	20.00	<a href="#">0.111</a>	19 20 21 48 49 51 96 98 136
<a href="#">ACY</a>	15.87	<a href="#">0.163</a>	23 29 31 37 44 45 81 83 85 94 96 98 103 121 135



Inherited partners:1

Partner	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
<a href="#">d.113.1.1</a>	23.68	<a href="#">0.948</a>	19 20 50 51 52 53 54 55 56 57 58 77 78 79 80 81 82 83 84 85 93 95 97 99 134 135 138 142 145

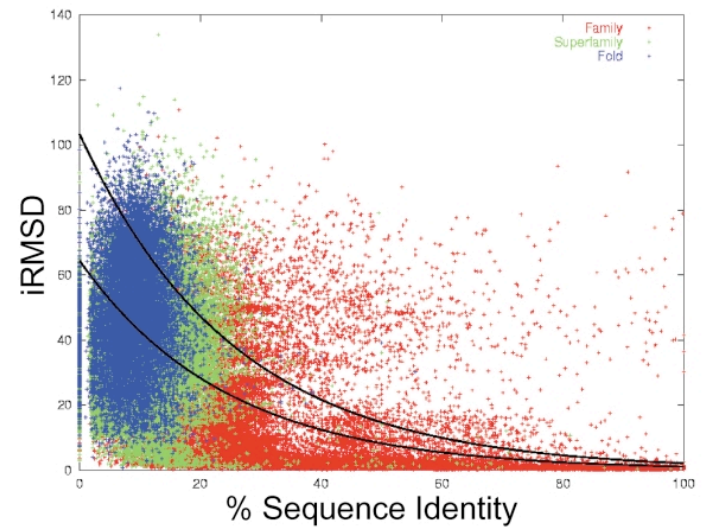
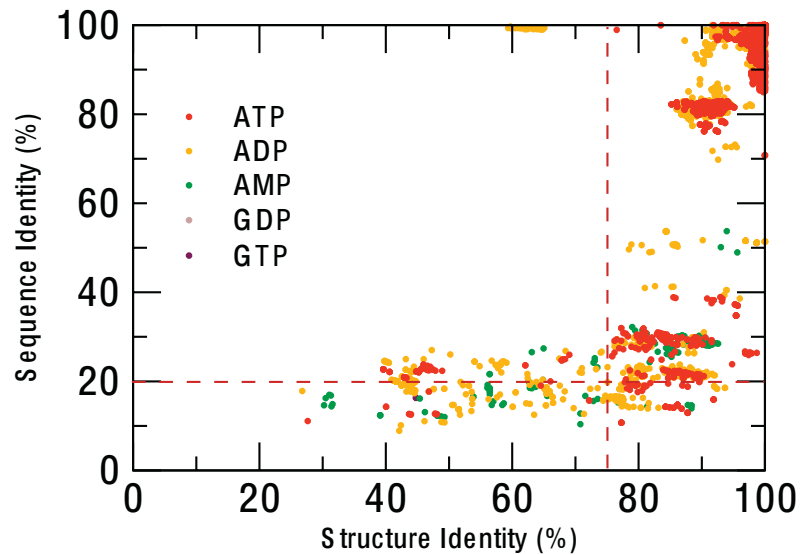




# Scoring function

Ligands

Partners



Aloy *et al.* (2003) J.Mol.Biol. 332(5):989-98.

# Benchmark

	Number of chains
<b>Initial set*</b>	78,167
<b>LigBase**</b>	30,126
<b>Non-redundant set***</b>	<b>4,948</b> (8,846 ligands)

*\*all PDB chains larger than 30 aminoacids in length (8th of August, 2006)*

*\*\*annotated with at least one ligand in the LigBase database*

*\*\*\*not two chains can be structurally aligned within 3Å, superimposing more than 75% of their Ca atoms, result in a sequence alignment with more than 30% identity, and have a length difference inferior to 50aa*

	Number of chains
<b>Initial set*</b>	78,167
<b><math>\pi</math>Base**</b>	30,425
<b>Non-redundant set***</b>	<b>4,613</b> (11,641 partnerships)

*\*all PDB chains larger than 30 aminoacids in length (8th of August, 2006)*

*\*\*annotated with at least one partner in the  $\pi$ Base database*

*\*\*\*not two chains can be structurally aligned within 3Å, superimposing more than 75% of their Ca atoms, result in a sequence alignment with more than 30% identity, and have a length difference inferior to 50aa*

# Sensitivity .vs. Precision

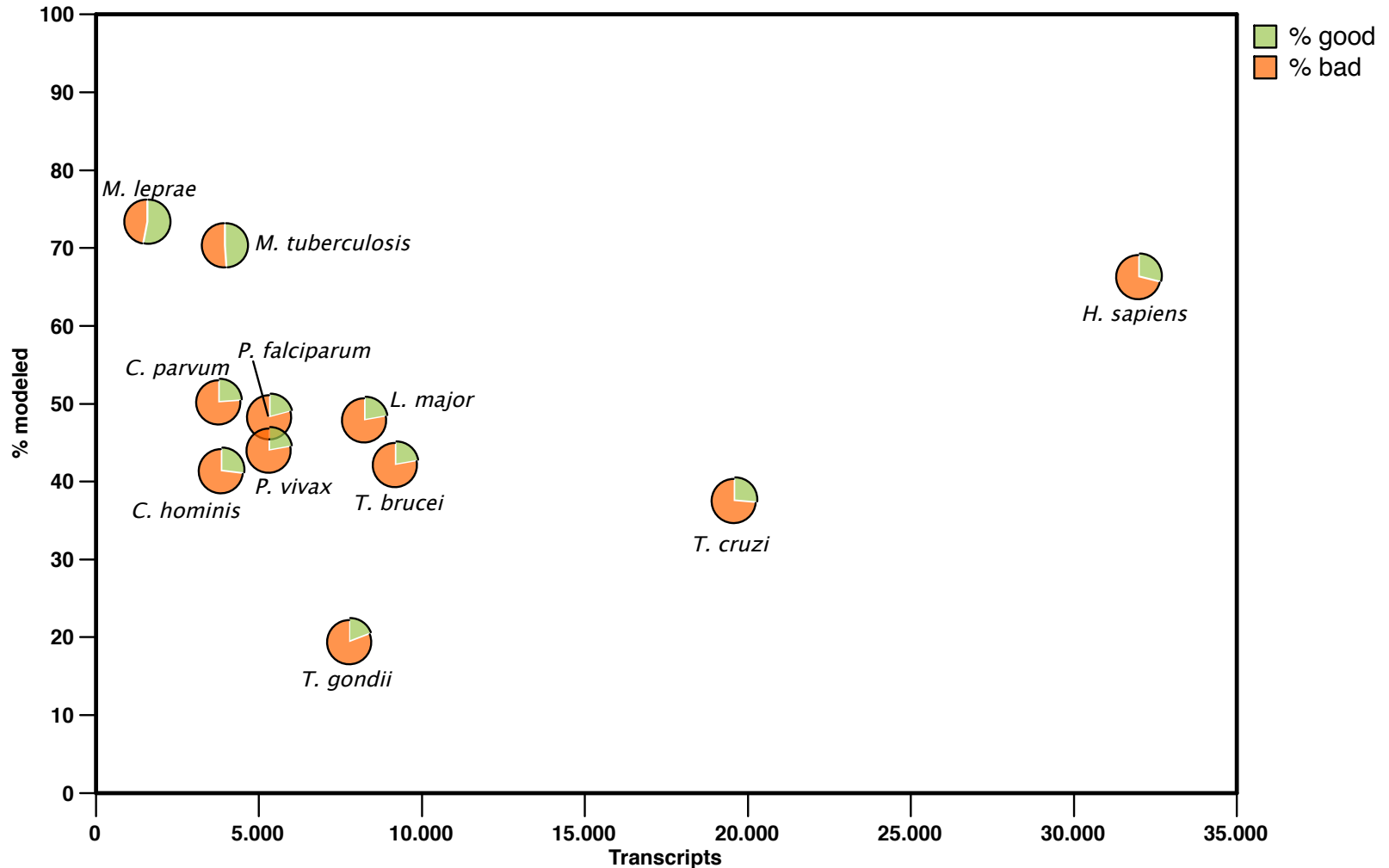
	Optimal cut-off	Sensitivity (%) Recall or TPR	Precision (%)
Ligands	30%	71.9	13.7

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{Precision} = \frac{TP}{TP + FP}$$

**~90-95% of residues correctly predicted**

# Modeling Genomes

*data from models generated by ModPipe (Eswar, Pieper & Sali)*

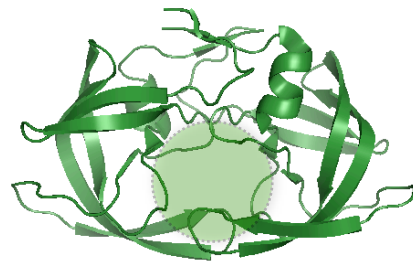
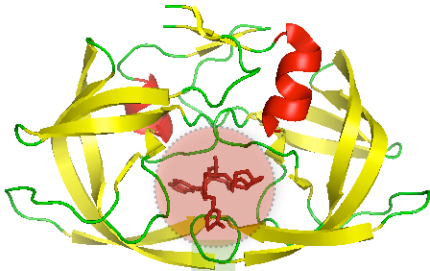


*A good model has MPQS of 1.1 or higher*

# Comparative docking

## 1. Expansion

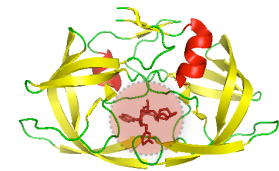
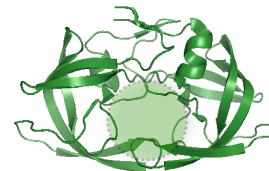
co-crystallized protein/ligand



crystalized protein

## 2. Inheritance

model



template

# Summary table

models with inherited ligands

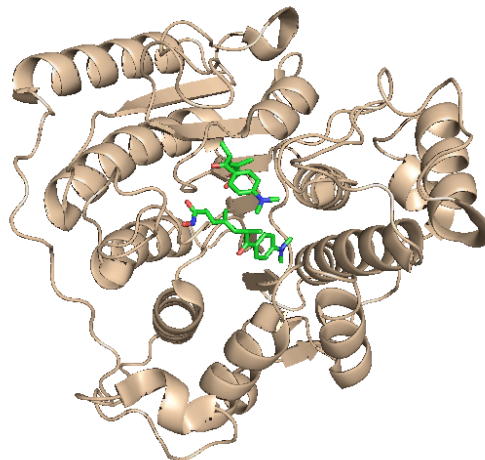
from 16,284 good models, 295 inherited a ligand/substance with at least one compound already approved by FDA and ready to be used from ZINC

	Transcripts	Good	Ligands	Lipinski	Lipinski+ZINC	FDA+ZINC
<i>C. hominis</i>	3,886	886	183	131	28	12 (10)
<i>C. parvum</i>	3,806	949	219	145	30	12 (10)
<i>L. major</i>	8,274	1,845	488	334	84	44 (34)
<i>M. leprae</i>	1,605	1,321	286	189	39	29 (25)
<i>M. tuberculosis</i>	3,991	2,887	404	285	71	44 (37)
<i>P. falciparum</i>	5,363	1,057	271	191	48	20 (16)
<i>P. vivax</i>	5,342	1,042	267	177	37	18 (15)
<i>T. brucei</i>	921	1,795	440	309	94	46 (36)
<i>T. cruzi</i>	19,607	3,915	730	493	127	62 (52)
<i>T. gondii</i>	7,793	587	174	124	28	8 (7)
<b>TOTAL</b>	<b>60,588</b>	<b>16,284</b>	<b>3,462</b>	<b>2,378</b>	<b>586</b>	<b>295 (242)</b>

# Example of inheritance (expansion)

*LmjF21.0680 from L. major “Histone deacetylase 2” (model 1)*

*Template 1t64A a human HDAC8 protein.*



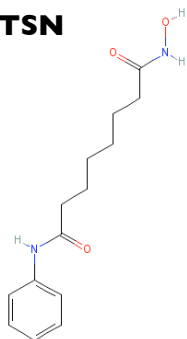
	Origen	Formula	Name	Cov.	Seq. Id. (%)
<b>ZN</b>	X-ray	Zn <sup>2+</sup>	Zinc ion	--	--
<b>NA</b>	X-ray	Na <sup>+</sup>	Sodium ion	--	--
<b>CA</b>	X-ray	Ca <sup>2+</sup>	Calcium ion	--	--
<b>TSN</b>	X-ray	C <sub>17</sub> H <sub>22</sub> N <sub>2</sub> O <sub>3</sub>	Trichostatin A	--	--
<b>SHH</b>	Expanded	C <sub>14</sub> H <sub>20</sub> N <sub>2</sub> O <sub>3</sub>	Octadenioic acid hydroxyamide phenylamide	100.00	83.8

# Example of inheritance (inheritance)

*LmjF21.0680 from L. major "Histone deacetylase 2" (model 1)*

	Formula	Name	Cov.	Seq. Id. (%)	Residues
<b>TSN</b>	C <sub>17</sub> H <sub>22</sub> N <sub>2</sub> O <sub>3</sub>	Trichostatin A	100.00	90.9	90 131 132 140 141 167 169 256 263 293 295
<b>SHH</b>	C <sub>14</sub> H <sub>20</sub> N <sub>2</sub> O <sub>3</sub>	Octadenioic acid hydroxyamide phenylamide	100.00	90.9	

**TSN**



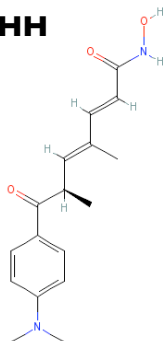
[suberoylanilide hydroxamic acid](#)

**Pharmacological Action:**

[Anti-Inflammatory Agents, Non-Steroidal](#)  
[Antineoplastic Agents](#)  
[Enzyme Inhibitors](#)  
[Anticarcinogenic Agents](#)

Inhibits histone deacetylase 1 and 3

**SHH**



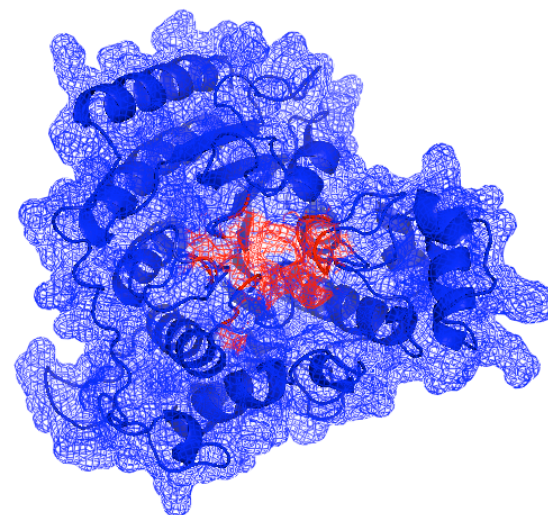
[trichostatin A](#)

**Pharmacological Action:**

[Antibiotics, Antifungal](#)  
[Enzyme Inhibitors](#)  
[Protein Synthesis Inhibitors](#)

chelates zinc ion in the active site of histone deacetylases, resulting in preventing histone unpacking so DNA is less available for transcription

	LmjF21.0680.1.pdb
<b>Template</b>	1t64A
<b>Seq. Id (%)</b>	<b>38.00</b>
<b>MPQS</b>	1.47





# Example of inheritance (CDD-Roos-literature)

*LmjF21.0680 from L. major “Histone deacetylase 2” (model 1)*

*Proc. Natl. Acad. Sci. USA*  
Vol. 93, pp. 13143–13147, November 1996  
Medical Sciences

## **Apicidin: A novel antiprotozoal agent that inhibits parasite histone deacetylase**

(cyclic tetrapeptide/Apicomplexa/antiparasitic/malaria/coccidiosis)

SANDRA J. DARKIN-RATTRAY\*<sup>†</sup>, ANNE M. GURNETT\*, ROBERT W. MYERS\*, PAULA M. DULSKI\*,  
TAMI M. CRUMLEY\*, JOHN J. ALLOCCO\*, CHRISTINE CANNOVA\*, PETER T. MEINKE<sup>‡</sup>, STEVEN L. COLLETTI<sup>‡</sup>,  
MARIA A. BEDNAREK<sup>‡</sup>, SHEO B. SINGH<sup>§</sup>, MICHAEL A. GOETZ<sup>§</sup>, ANNE W. DOMBROWSKI<sup>§</sup>,  
JON D. POLISHOOK<sup>§</sup>, AND DENNIS M. SCHMATZ\*

Departments of \*Parasite Biochemistry and Cell Biology, <sup>‡</sup>Medicinal Chemistry, and <sup>§</sup>Natural Products Drug Discovery, Merck Research Laboratories, P.O. Box 2000, Rahway, NJ 07065

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Apr. 2004, p. 1435–1436  
0066-4804/04/\$08.00+0 DOI: 10.1128/AAC.48.4.1435–1436.2004  
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

Vol. 48, No. 4

## **Antimalarial and Antileishmanial Activities of Aroyl-Pyrrolyl-Hydroxyamides, a New Class of Histone Deacetylase Inhibitors**

# TDI Models database

<http://sgu.bioinfo.cipf.es/services/TDIModels/>

The TDIModels server

<http://sgu.bioinfo.cipf.es/services/TDIModels/>

The TDIModels server

TDIModels

Results for Q7K6A1 [Q7K6A1 Histone deacetylase ]  
Number of models: 2

Model 1

This model has 2 predicted binding sites.  
Lipinski ZINC FDA Coverage Seq. Id.  
☒ SHH 100.00 90.91  
☐ TSN 100.00 90.91

SEQUENCE IDENTITY: 40.00  
MODEPIPE QUALITY SCORE: 1.43  
TEMPLATE PDB: 1t64  
TEMPLATE CHAIN: A  
TARGET LENGTH: 449  
TARGET BEGIN: 7  
TARGET END: 373  
[Download PDB file](#)

Model 2

This model has 2 predicted binding sites.  
Lipinski ZINC FDA Coverage Seq. Id.  
☐ SHH 100.00 90.91  
☐ TSN 100.00 90.91

SEQUENCE IDENTITY: 42.00  
MODEPIPE QUALITY SCORE: 1.45  
TEMPLATE PDB: 1t64  
TEMPLATE CHAIN: A  
TARGET LENGTH: 449  
TARGET BEGIN: 28  
TARGET END: 373  
[Download PDB file](#)

<- new search

HELP:

PLEASE NOTE: Our servers have been optimized for Firefox and Safari. If you are using Internet Explorer, the CSS may not be properly rendered.

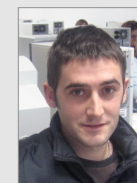
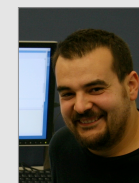
To use TDIModels you need to:

- Enter a UNIPROT id or a Keyword of interest to find models.

Applet jmolApplet1 started

17

# Acknowledgments



## COMPARATIVE MODELING

**Andrej Sali**

M. S. Madhusudhan

**Narayanan Eswar**

Min-Yi Shen

**Ursula Pieper**

Ben Webb

Maya Topf (Birbeck College)

## MODEL ASSESSMENT

David Eramian

Min-Yi Shen

Damien Devos

## FUNCTIONAL ANNOTATION

**Andrea Rossi (Rinat-Pfizer)**

**Fred Davis (Janelia Fram)**

## FUNDING

Prince Felipe Research Center

Generalitat Valenciana

STREP UE Grant

Marie Curie Reintegration Grant

## MODEL ASSESSMENT

Francisco Melo (CU)

Alejandro Panjkovich (CU)

## STRUCTURAL GENOMICS

Stephen Burley (SGX)

John Kuriyan (UCB)

NY-SGXRC

## MAMMOTH

**Angel R. Ortiz**

## FUNCTIONAL ANNOTATION

Fatima Al-Shahrour

Joaquin Dopazo

## BIOLOGY

Jeff Friedman (RU)

James Hudsped (RU)

Partho Ghosh (UCSD)

Alvaro Monteiro (Cornell U)

Stephen Krilis (St.George H)

## Tropical Disease Initiative

**Stephen Maurer (UC Berkeley)**

**Arti Rai (Duke U)**

**Andrej Sali (UCSF)**

**Ginger Taylor (TSL)**

## CCPR Functional Proteomics

Patsy Babbitt (UCSF)

Fred Cohen (UCSF)

Ken Dill (UCSF)

Tom Ferrin (UCSF)

John Irwin (UCSF)

Matt Jacobson (UCSF)

Tack Kuntz (UCSF)

Andrej Sali (UCSF)

Brian Shoichet (UCSF)

Chris Voigt (UCSF)

## EVA

Burkhard Rost (Columbia U)

Alfonso Valencia (CNB/UAM)

## CAMP

Xavier Aviles (UAB)

Hans-Peter Nester (SANOFI)

Ernst Meinjohanns (ARPIDA)

Boris Turk (IJS)

Markus Gruetter (UE)

Matthias Wilmanns (EMBL)

Wolfram Bode (MPG)