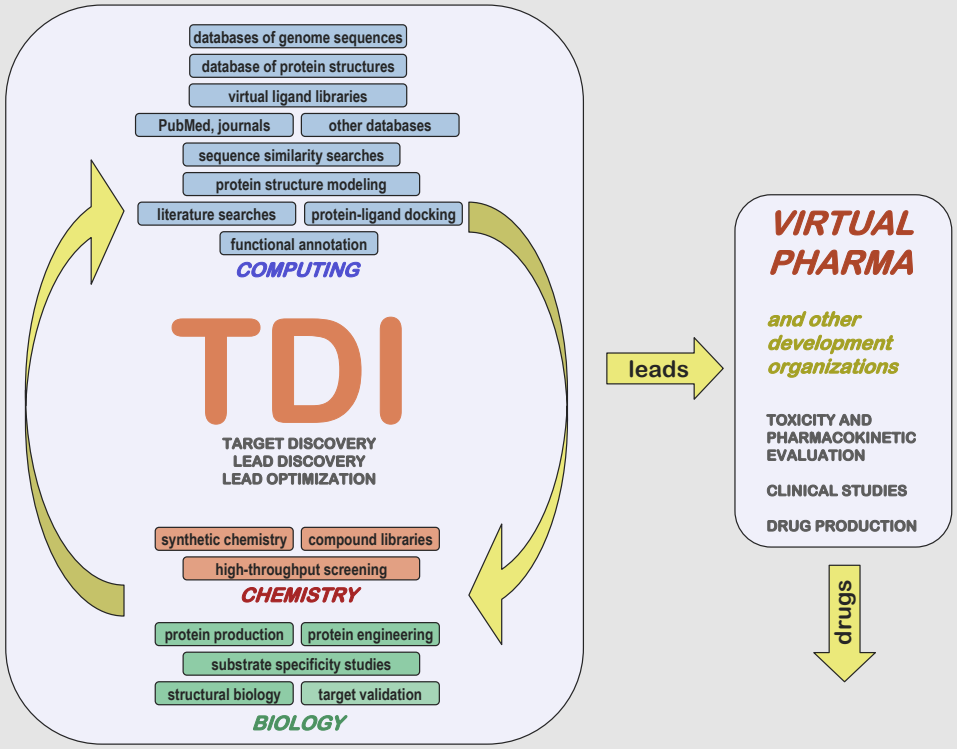
The Tropical Disease Initiative

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

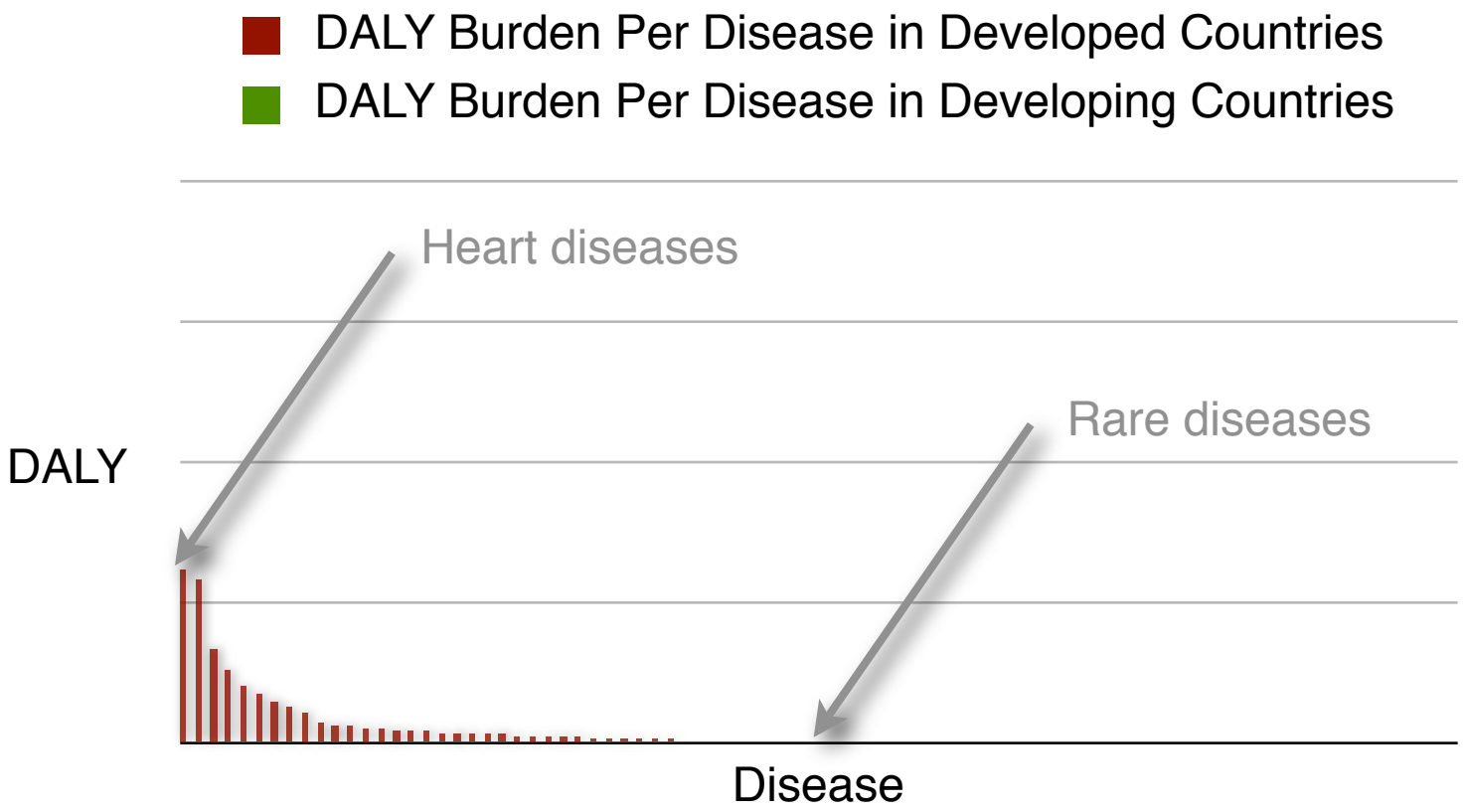


Marc A. Marti-Renom

http://bioinfo.cipf.es/sgu/

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

Need is High in the Tail



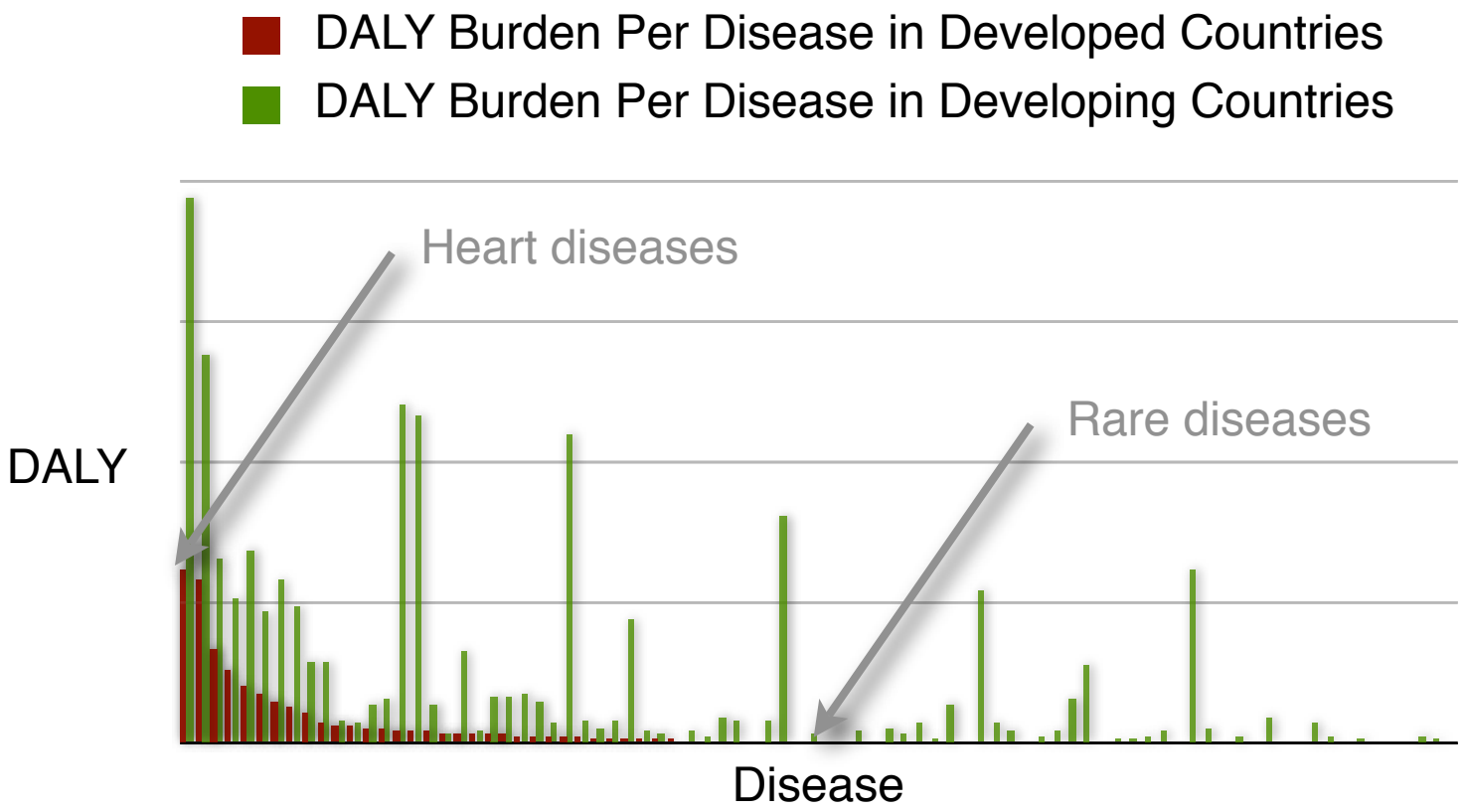
Disease data taken from WHO, *World Health Report 2004*

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



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)

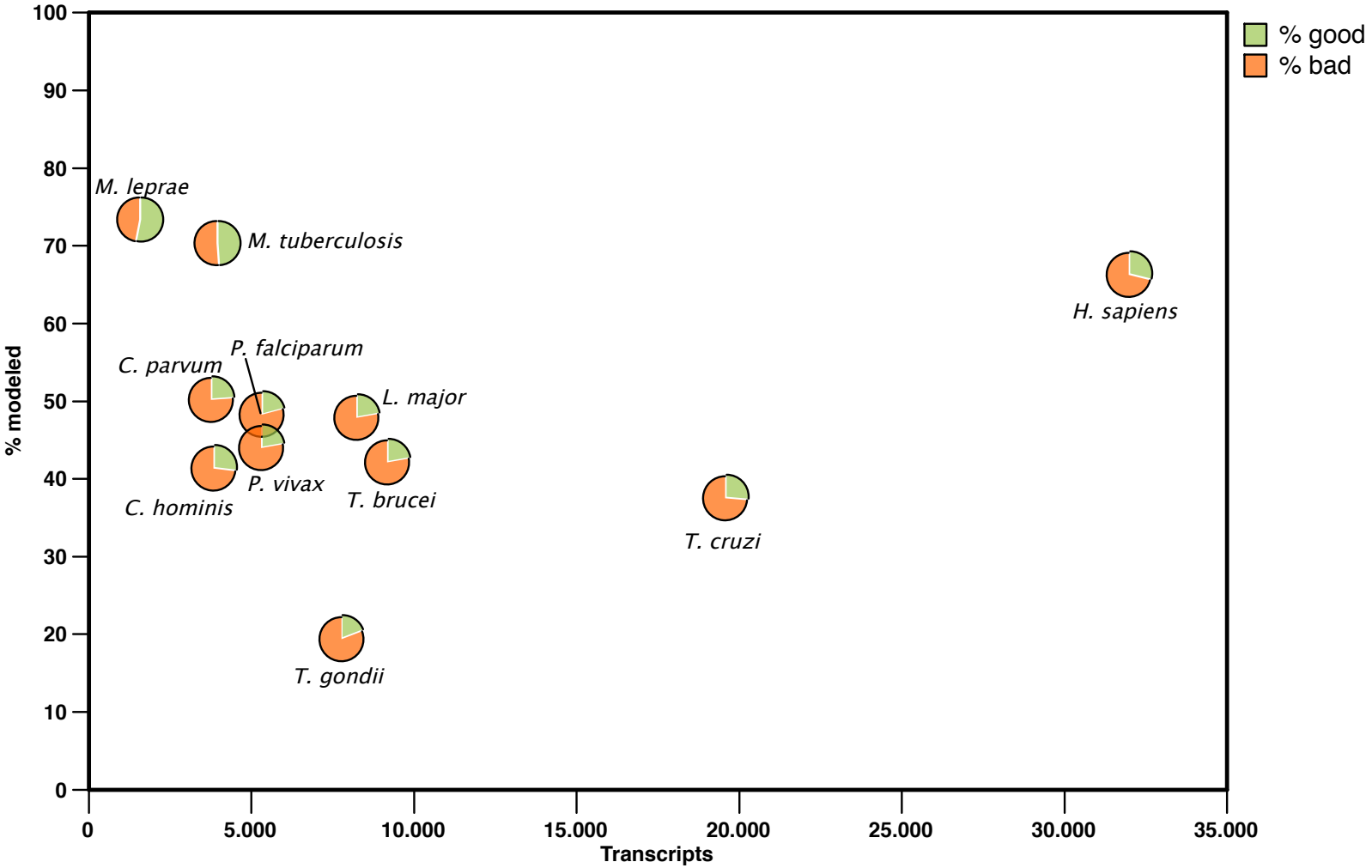
Malaria*	**46,486**
Tetanus	7,074
Lymphatic filariasis*	**5,777**
Syphilis	4,200
Trachoma	2,329
Leishmaniasis*	**2,090**
Ascariasis	1,817
Schistosomiasis*	**1,702**
Trypanosomiasis*	**1,525**

Trichuriasis	1,006
Japanese encephalitis	709
Chagas Disease*	**667**
Dengue*	**616**
Onchocerciasis*	**484**
Leprosy*	**199**
Diphtheria	185
Poliomyelitise	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.

Modeling Genomes

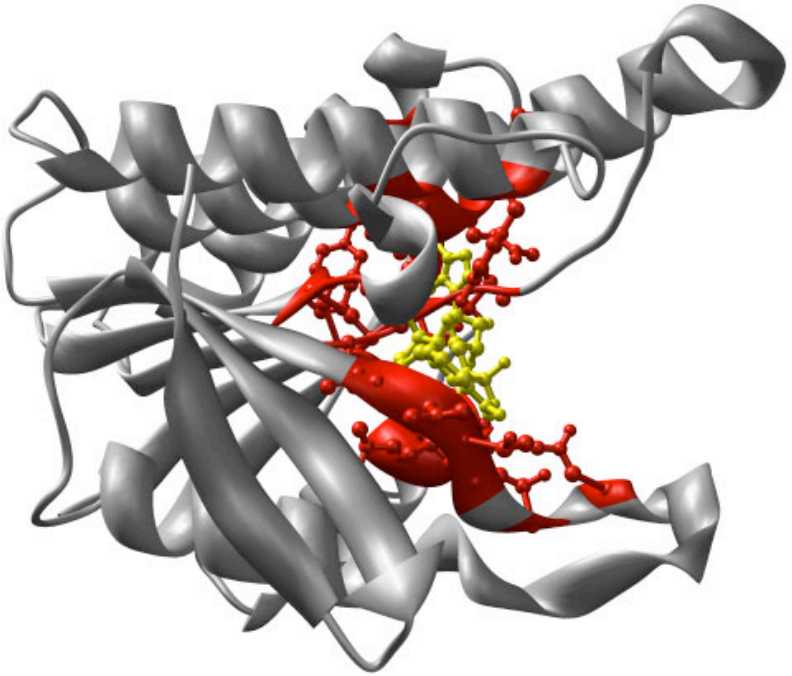
data from models generated by ModPipe (Eswar, Pieper & Sali)



A good model has MPQS of 1.1 or higher

AnnoLyze

Predicting binding sites in protein structure models.

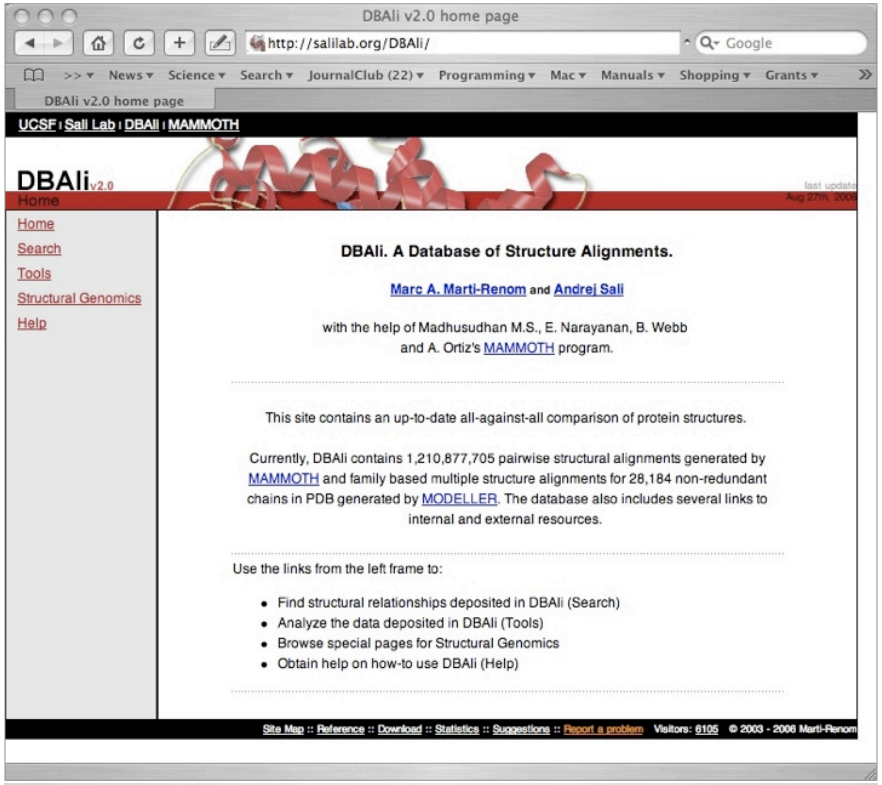


http://bioinfo.cipf.es/sgu/services/DBAli/

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

DBAli$_{v2.0}$ database

http://bioinfo.cipf.es/sgu/services/DBAli/
http://www.salilab.org/DBAli/



- ✓ 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:	July 5th, 2007
Number of chains:	93,307
Number of structure-structure comparisons:*	1,617,719,157
Multiple structure alignments	
Last update:	March 22nd, 2007
Number of representative chains:	31,848
Number of families:	11,900

Uses MAMMOTH for similarity detection

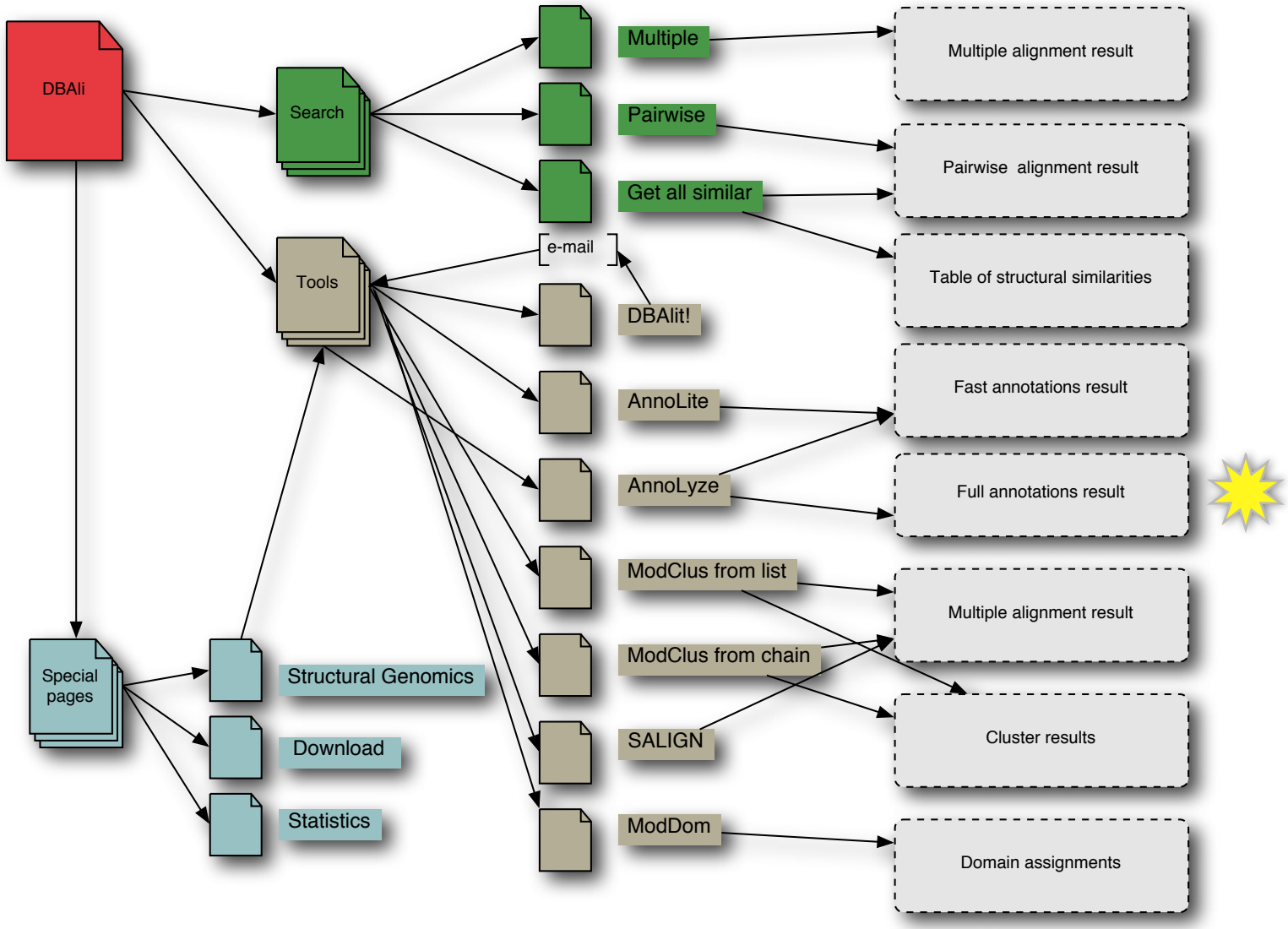
- ✓ VERY FAST!!!
- ✓ Good scoring system with significance

Ortiz AR, (2002) Protein Sci. 11 pp2606

Marti-Renom et al. 2001. Bioinformatics. 17, 746

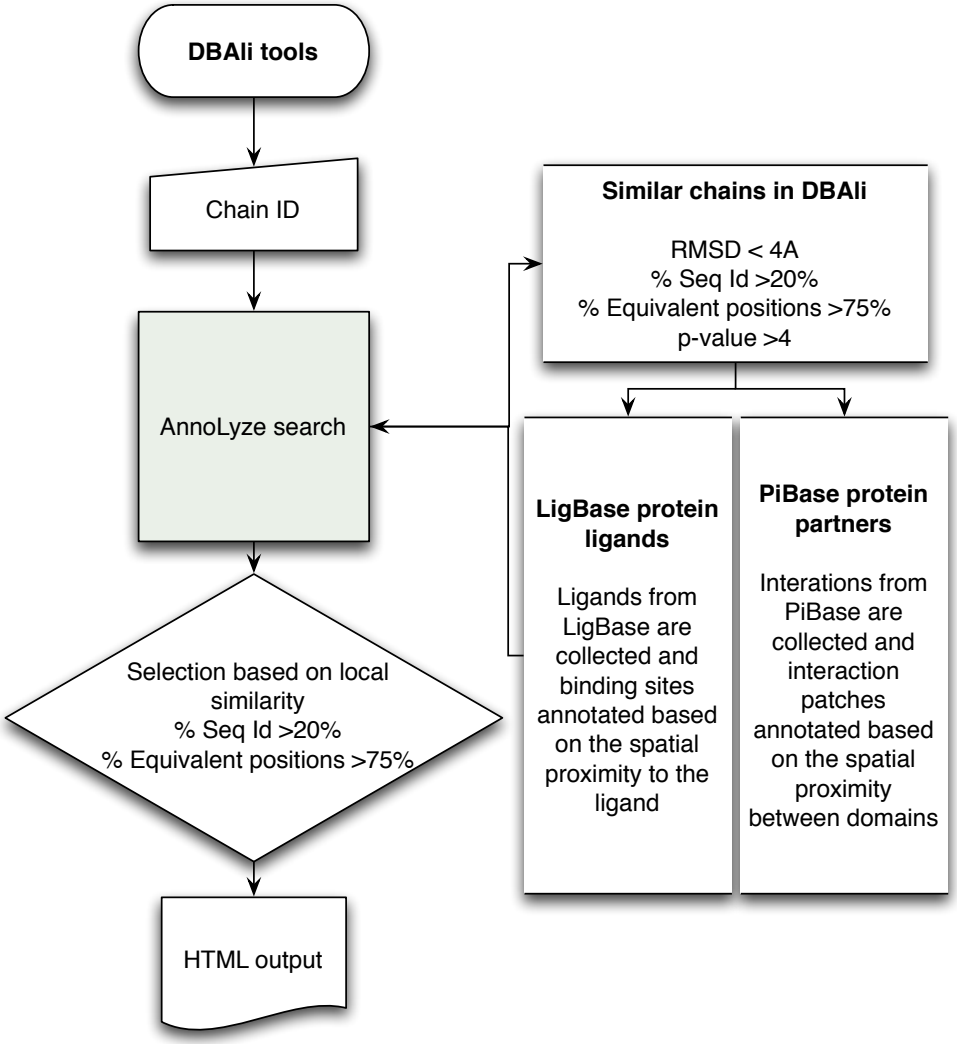
DBAli$_{v2.0}$ database

http://bioinfo.cipf.es/sgu/services/DBAli/
http://www.salilab.org/DBAli/



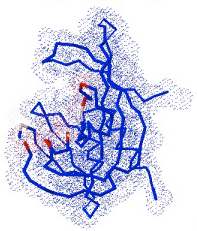
Marti-Renom et al. BMC Bioinformatics (2007) Volume 8. Suppl S4

Method



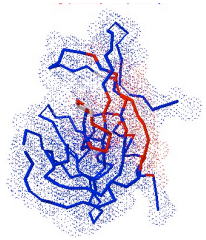
Inherited ligands: 4

Ligand	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
MO2	59.03	0.185	48 49 52 62 63 66 67 113 116
CRY	20.00	0.111	23 29 31 37 44 48 49 83 85 94 96 103 121
8OG	20.00	0.111	19 20 21 48 49 51 96 98 136
ACY	15.87	0.163	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)
d.113.1.1	23.68	0.948	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



Sensitivity .vs. Precision

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

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

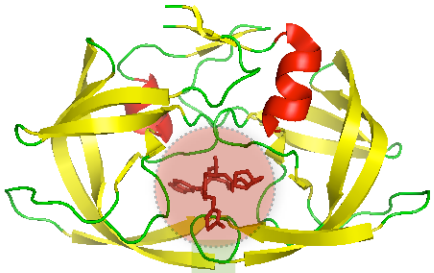
~90-95% of residues correctly predicted

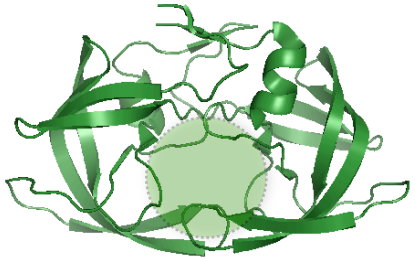
Marti-Renom et al. BMC Bioinformatics (2007) Volume 8. Suppl S4

Comparative docking

1. Expansion

co-crystalized protein/ligand

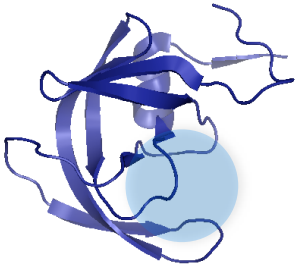


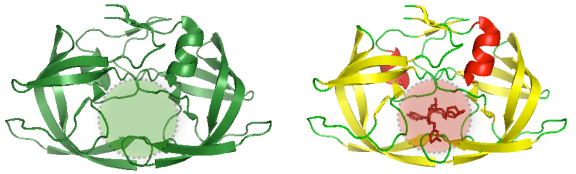


crystalized protein

2. Inheritance

model





template

Summary table

models with inherited ligands

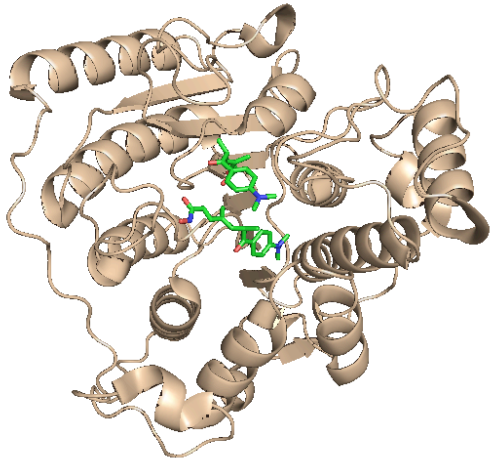
from 16,284 good models, 295 inherited a ligand/substance with at least one compound already approved by FDA and ready to be used from ZINC

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

Example of inheritance (expansion)

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

Template 1t64A a human HDAC8 protein.



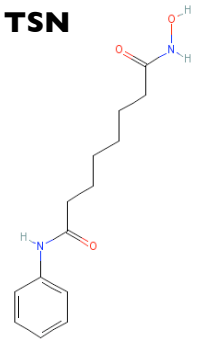
	Origen	Formula	Name	Cov.	Seq, Id. (%)
ZN	X-ray	Zn^{2+}	Zinc ion	--	--
NA	X-ray	Na^{+}	Sodium ion	--	--
CA	X-ray	Ca^{2+}	Calcium ion	--	--
TSN	X-ray	$C_{17}H_{22}N_2O_3$	Trichostatin A	--	--
SHH	Expanded	$C_{14}H_{20}N_2O_3$	Octadenioic acid hudroxyamide phenylamide	100.00	83.8

Example of inheritance (inheritance)

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

	Formula	Name	Cov.	Seq, Id. (%)	Residues
TSN	$C_{17}H_{22}N_2O_3$	Trichostatin A	100.00	90.9	90 131 132 140 141 167 169 256 263 293 295
SHH	$C_{14}H_{20}N_2O_3$	Octadenioic acid hudroxyamide phenylamide	100.00	90.9	

TSN



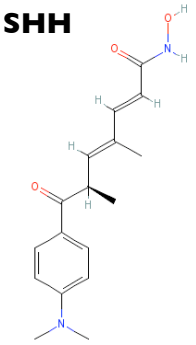
suberoylanilide hydroxamic acid

Pharmacological Action:

Anti-Inflammatory Agents, Non-Steroidal
Antineoplastic Agents
Enzyme Inhibitors
Anticarcinogenic Agents

Inhibits histone deacetylase I and 3

SHH



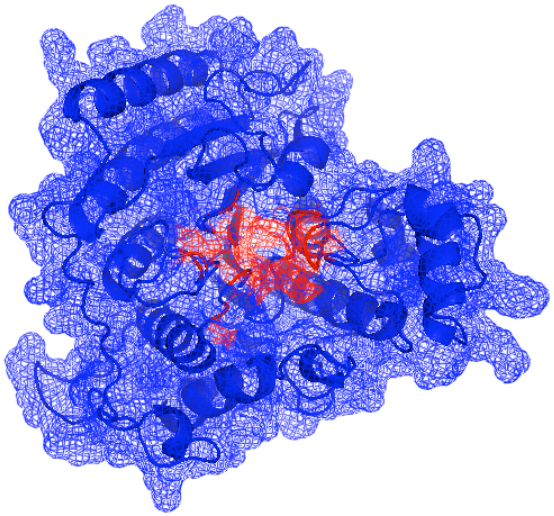
trichostatin A

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
Template	1t64A
Seq. Id (%)	**38.00**
MPQS	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*†, ANNE M. GURNETT*, ROBERT W. MYERS*, PAULA M. DULSKI*, TAMI M. CRUMLEY*, JOHN J. ALLOCCO*, CHRISTINE CANNOVA*, PETER T. MEINKE‡, STEVEN L. COLLETTI‡, MARIA A. BEDNAREK‡, SHEO B. SINGH§, MICHAEL A. GOETZ§, ANNE W. DOMBROWSKI§, JON D. POLISHOOK§, AND DENNIS M. SCHMATZ*

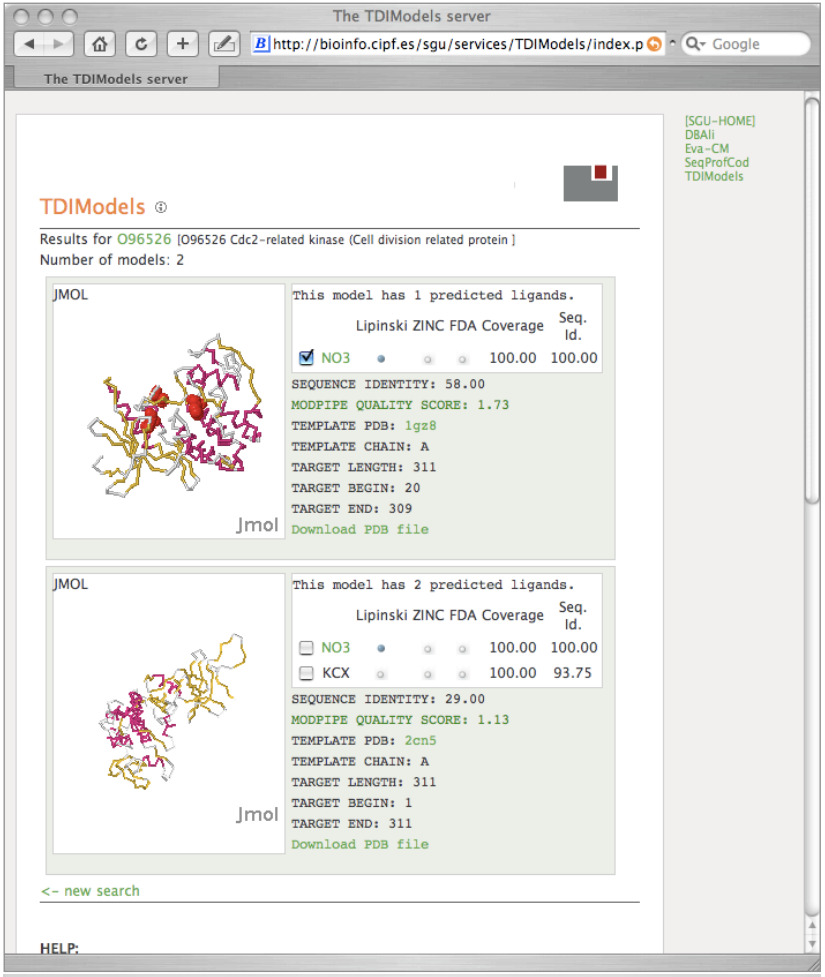
Departments of *Parasite Biochemistry and Cell Biology, ‡Medicinal Chemistry, and §Natural Products Drug Discovery, Merck Research Laboratories, P.O. Box 2000, Rahway, NJ 07065

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Apr. 2004, p. 1435–1436 Vol. 48, No. 4
0066-4804/04/$08.00+0 DOI: 10.1128/AAC.48.4.1435–1436.2004

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

Models database

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



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)

http://bioinfo.cipf.es/sgu/