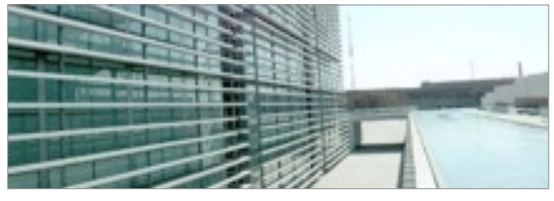
# Comparative docking for predicted molecular targets of known drugs.

## *A "kernel" for the Tropical Disease Initiative.*

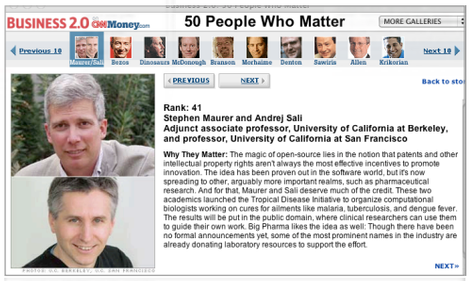

**Marc A. Marti-Renom**

http://sgu.bioinfo.cipf.es

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



# TDI *a story*



**2004**

.Steve Maurer (Berkeley) and Arti Rai (Duke)
.PLoS Medicine, Dec. 2004. Vol 1(3):e56

**2005**

.TDI web site `http://TropicalDisease.org`
.Ginger Taylor and The Synaptic Leap

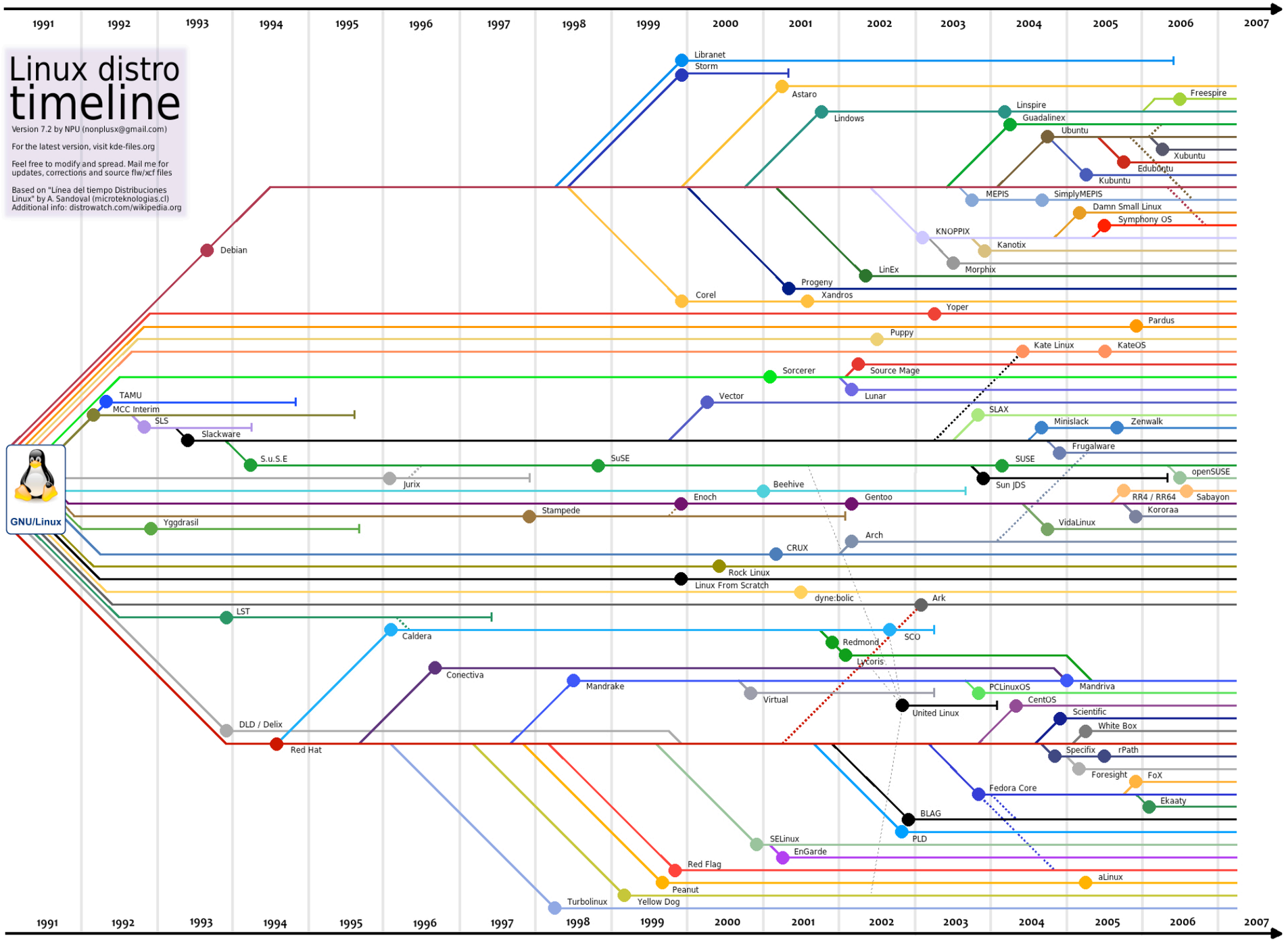
**2006**

.Maurer and Sali 41th in “50 Who Matter”
.TSL web site `http://TheSynapticLeap.org`

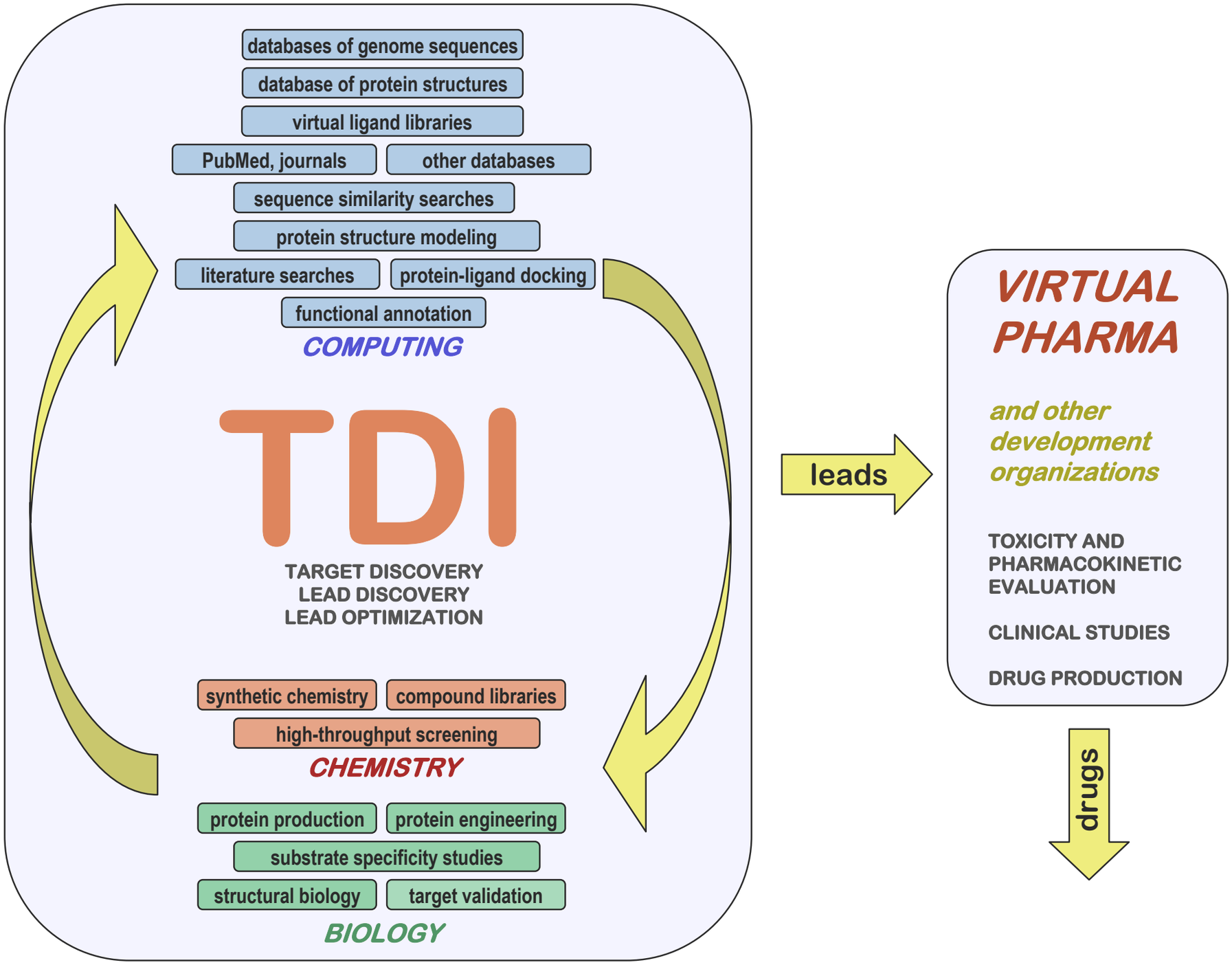
**2009**

.TDI kernel `http://TropicalDisease.org/kernel`

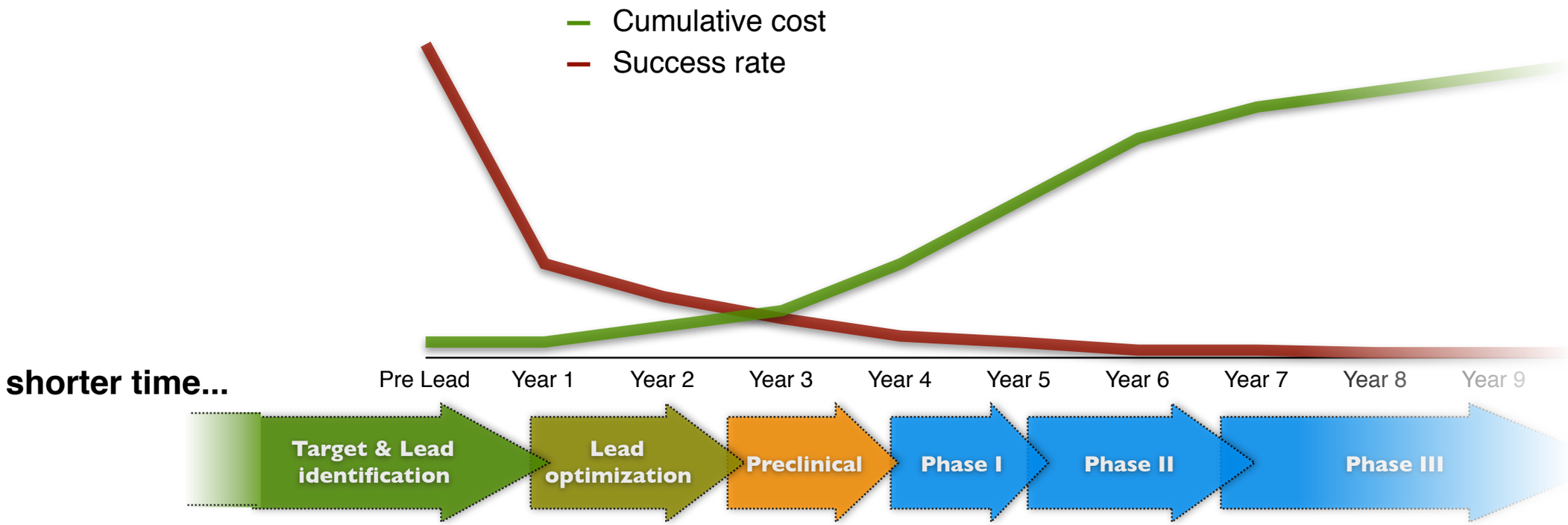
# Open Source without a Kernel?



# TDI flowchart



# Drug Discovery pipeline

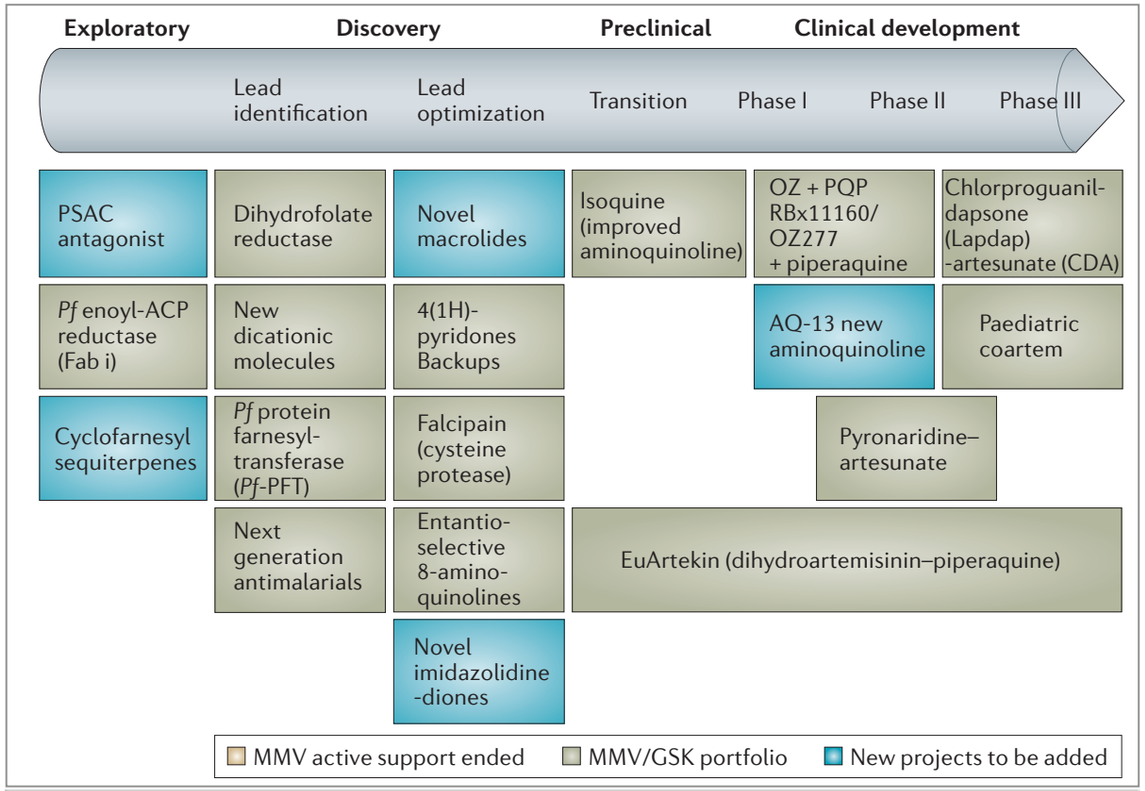


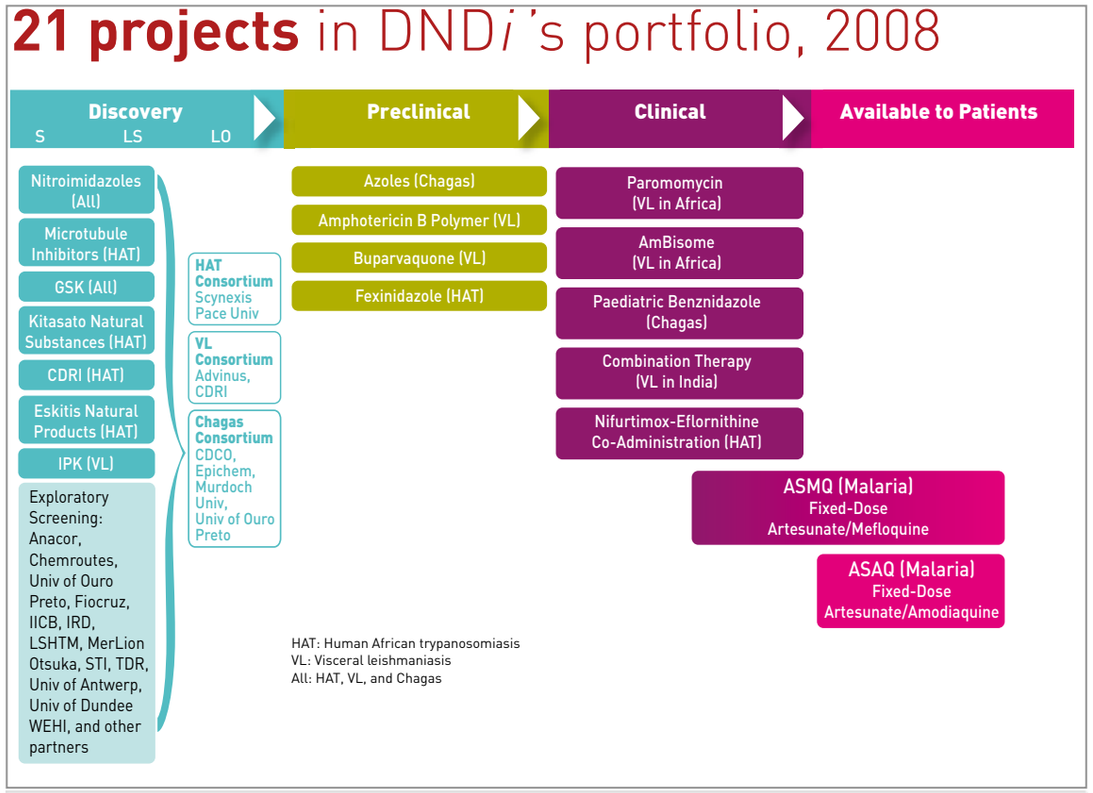


+ Completeness of genome projects (*eg*, Malaria)
+ New and more complete biological databases
+ New software and computers (cheaper and faster)
+ Internet == more people == less cost

# Non-Profit organizations

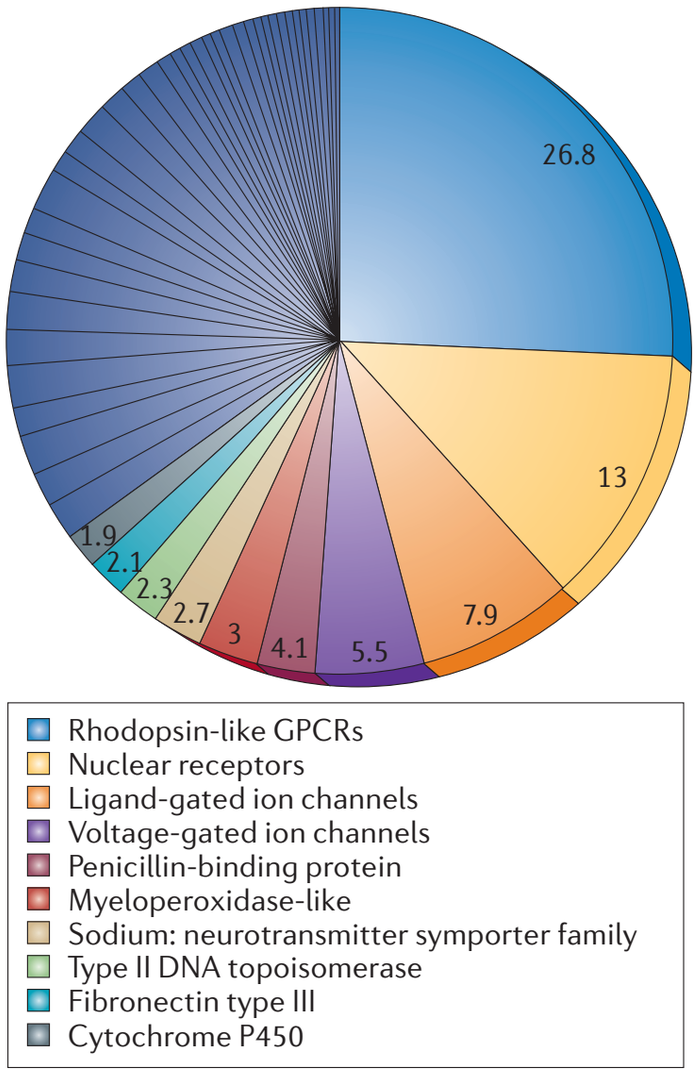
*Open-Source + Out-Source = low cost business model*




*Munos (2006) Nature Reviews. Drug Discovery.*

# Number of (new) targets?



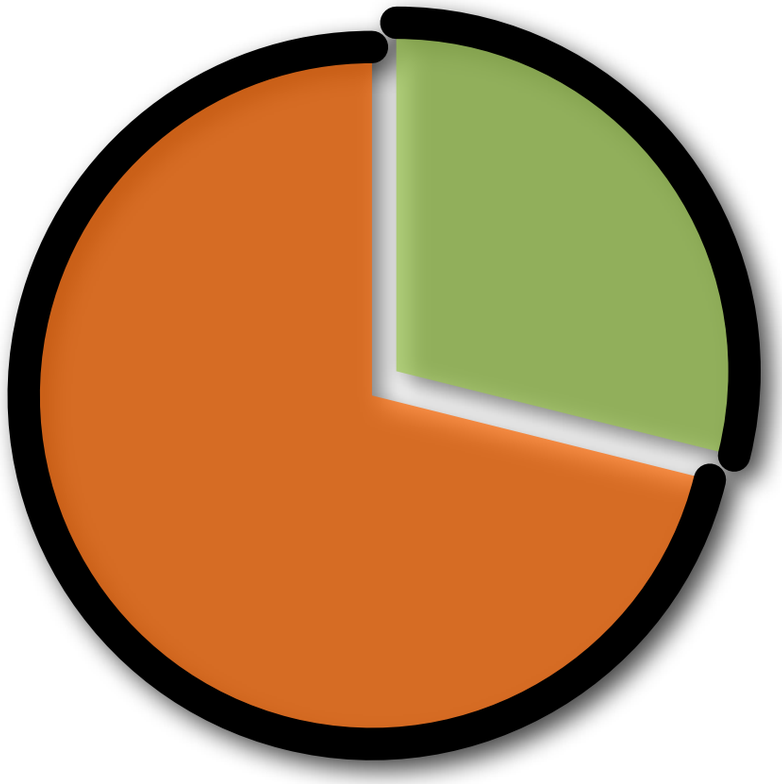

"... of 361 new molecular entities approved by the FDA between 1989 and 2000, 76% targeted a precedented drugged domain and only 6% targeted a previously undrugged domain ..."

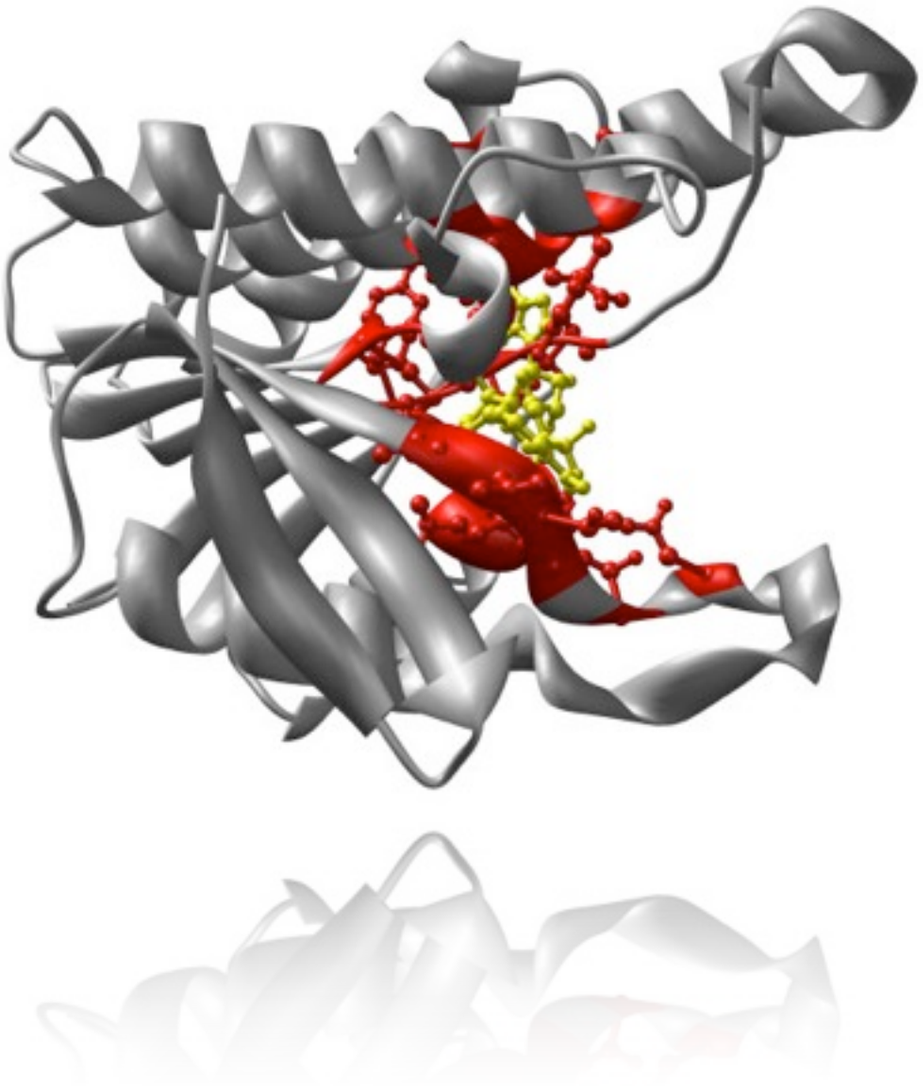
| Class of drug target | Species | Number of molecular targets |
|---|---|---|
| Targets of approved drugs | Pathogen and human | 324 |
| Human genome targets of approved drugs | Human | 266 |
| Targets of approved small-molecule drugs | Pathogen and human | 248 |
| Targets of approved small-molecule drugs | Human | 207 |
| Targets of approved oral small-molecule drugs | Pathogen and human | 227 |
| Targets of approved oral small-molecule drugs | Human | 186 |
| Targets of approved therapeutic antibodies | Human | 15 |
| Targets of approved biologicals | Pathogen and human | 76 |

**Where are new the targets for tropical diseases?**

*Overington et al. How many drug targets are there?. Nature reviews Drug discovery (2006) vol. 5 (12) pp. 993-6*

# Predicting binding sites in protein structure models of Tropical Diseases

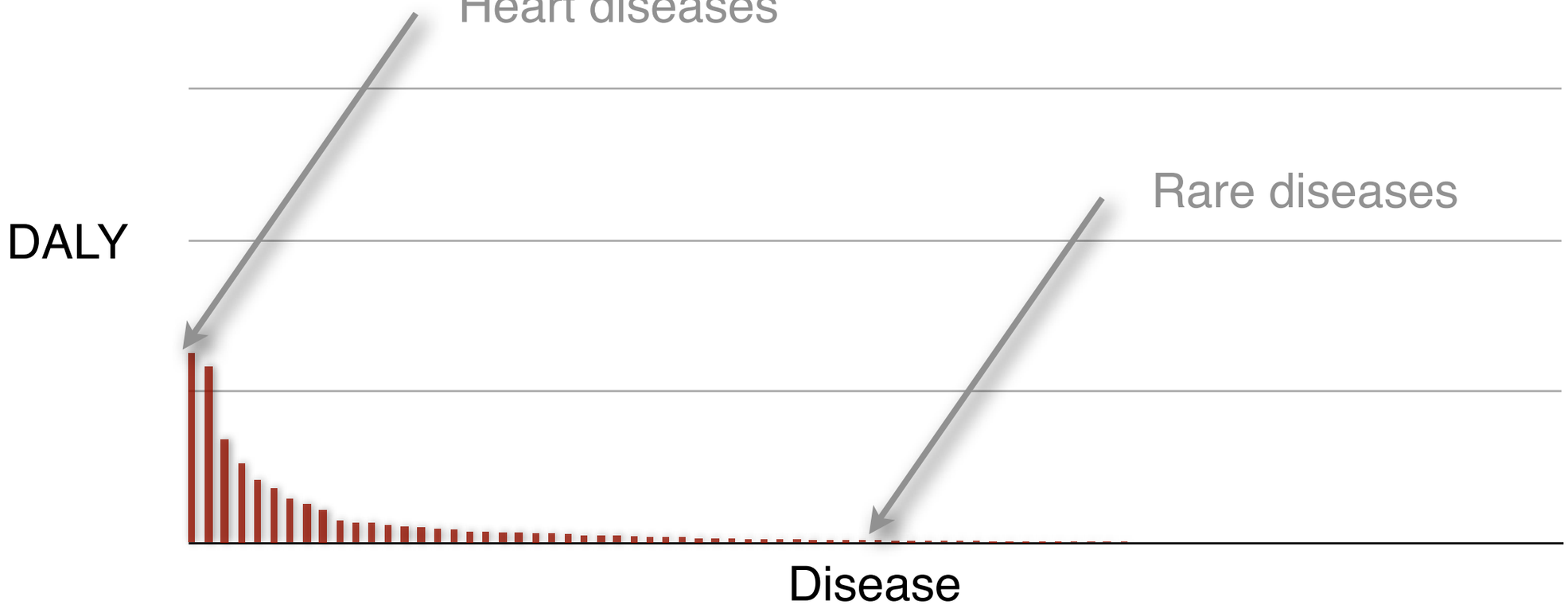




# Need is High in the Tail

DALY Burden Per Disease in Developed Countries

DALY Burden Per Disease in Developing Countries



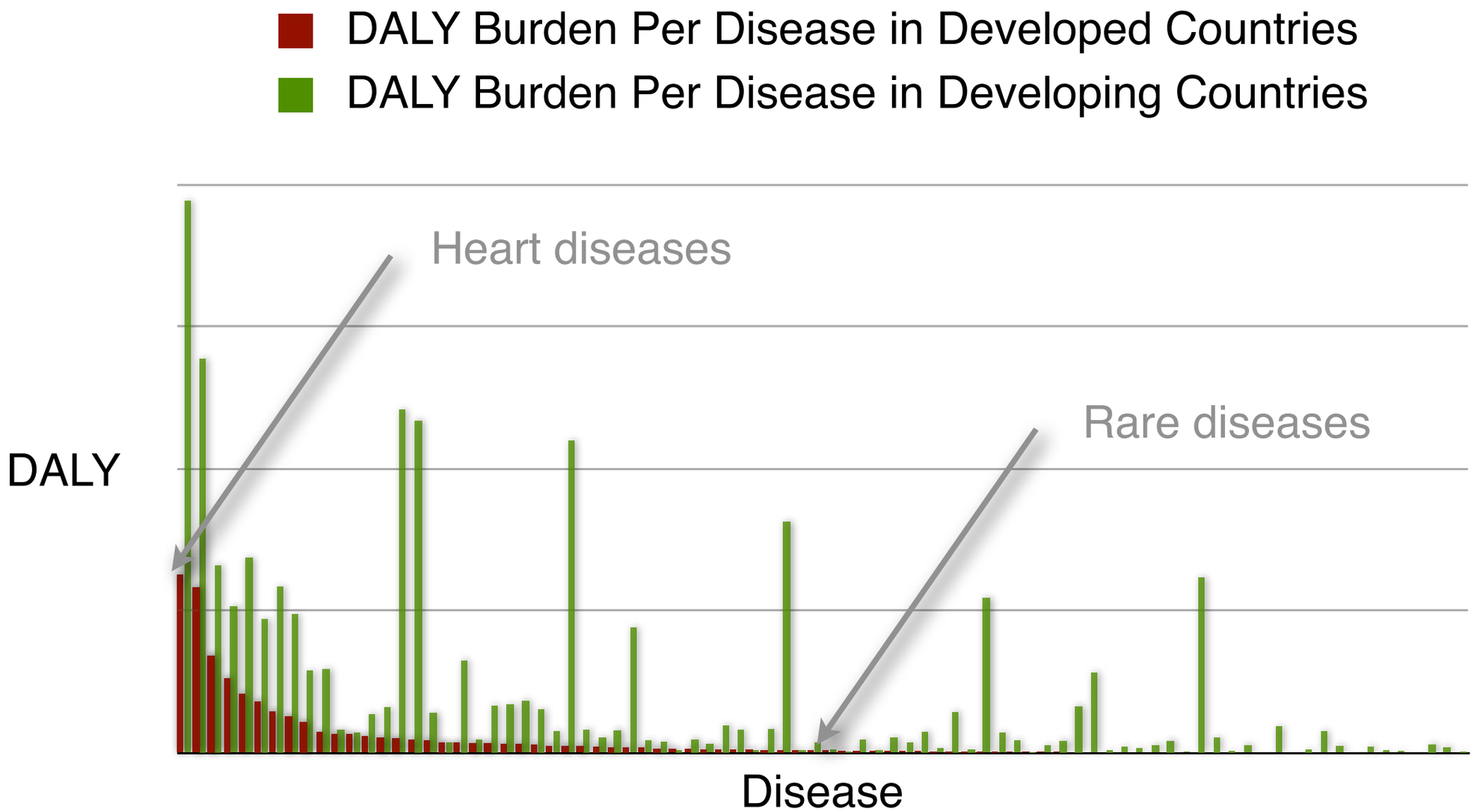

Disease data taken from WHO, *World Health Report 2004*

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




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** |

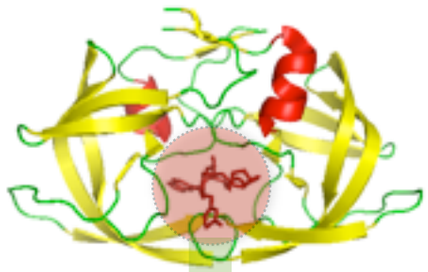
| | |
|---|---|
| Trichuriasis | 1,006 |
| Japanese encephalitis | 709 |
| **Chagas Disease*** | **667** |
| **Dengue*** | **616** |
| **Onchocerciasis*** | **484** |
| **Leprosy*** | **199** |
| Diphtheria | 185 |
| Poliomyelitise | 151 |
| Hookworm disease | 59 |

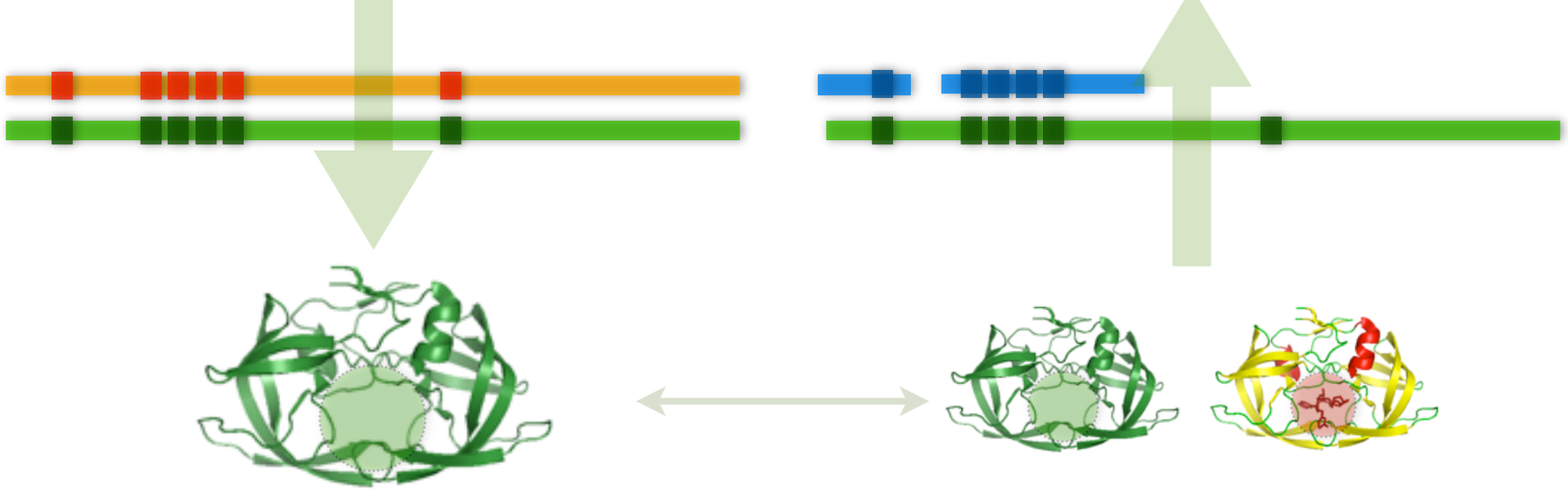
Disease data taken from WHO, *World Health Report 2004*
DALY - Disability adjusted life year in 1000’s.
* Officially listed in the WHO Tropical Disease Research disease portfolio.

# Comparative docking

**Expansion**

co-crystalized protein/ligand







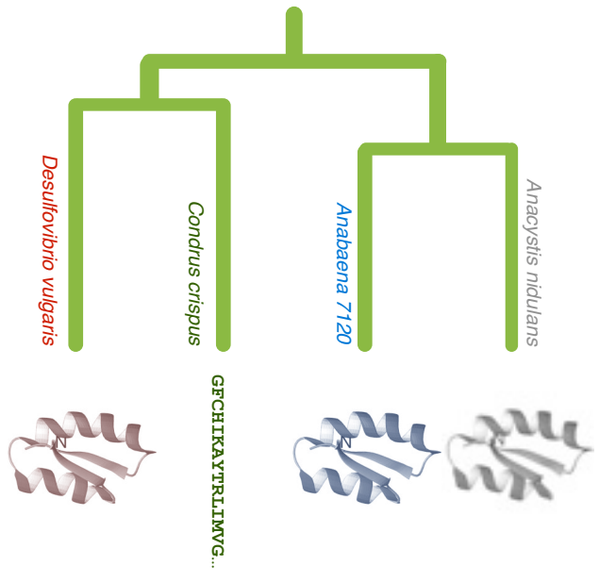
crystalized protein

**2. Inheritance**

model



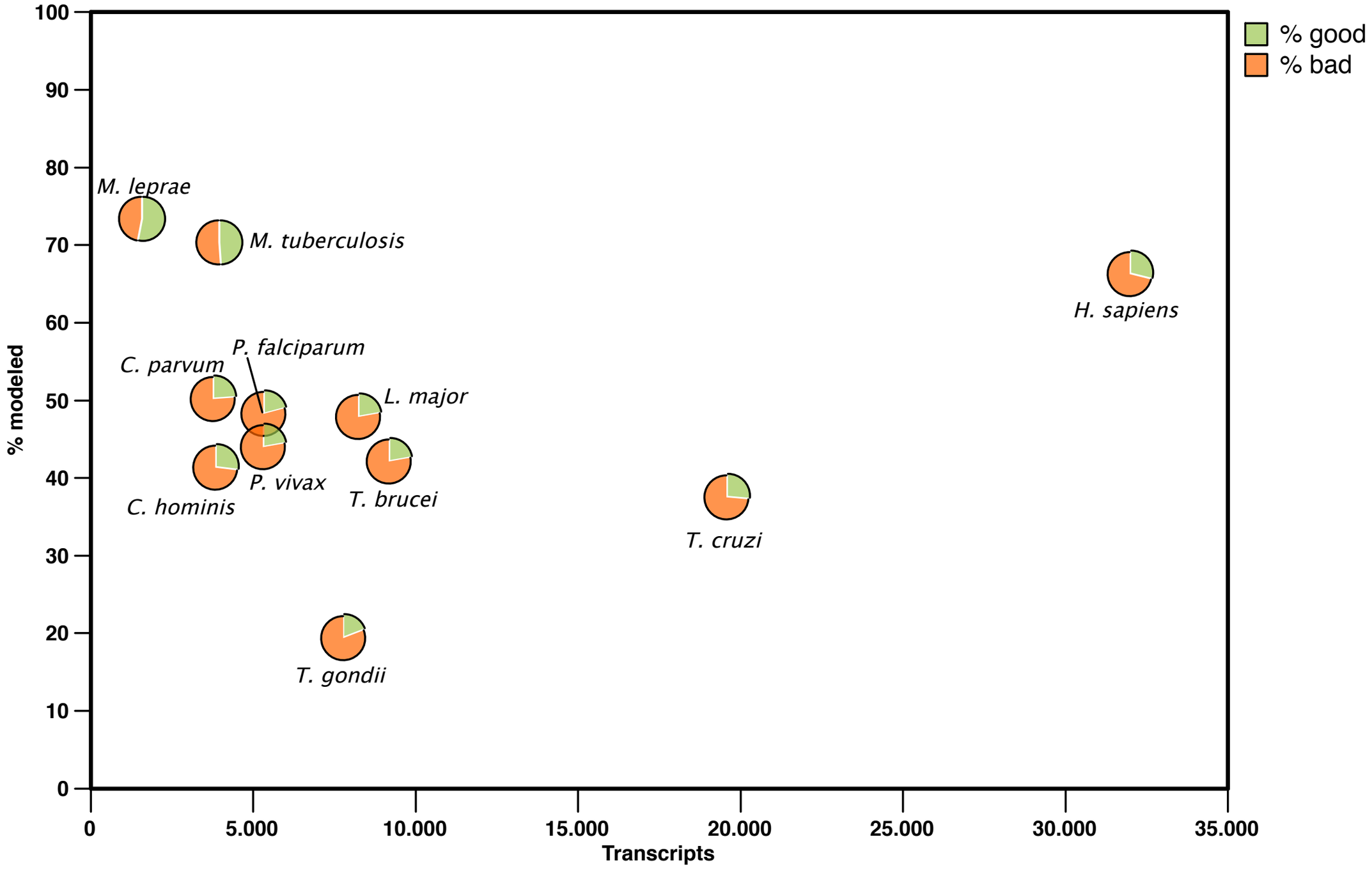
template




**1. Modeling**

# Modeling Genomes

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




*A good model has MPQS of 1.0 or higher*

# Summary table

models with inherited ligands

## 29,271 targets with good models, 297 inherited a ligand/substance similar to a known drug in DrugBank

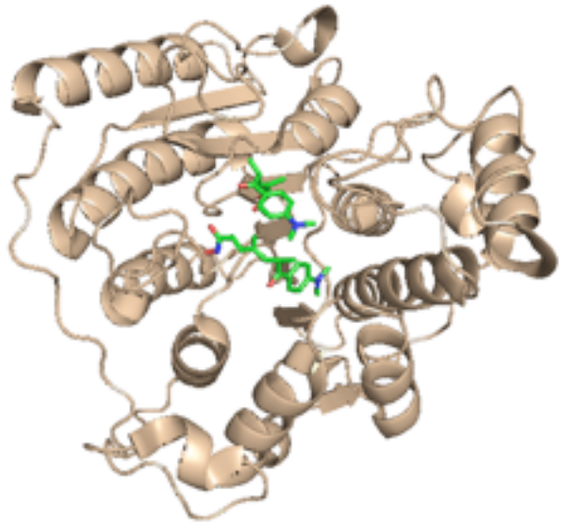
| | Transcripts | Modeled targets | Selected models | Inherited ligands | Similar to a drug | Drugs |
|---|---|---|---|---|---|---|
| ***C. hominis*** | 3,886 | 1,614 | 666 | 197 | 20 | 13 |
| ***C. parvum*** | 3,806 | 1,918 | 742 | 232 | 24 | 13 |
| ***L. major*** | 8,274 | 3,975 | 1,409 | 478 | 43 | 20 |
| ***M. leprae*** | 1,605 | 1,178 | 893 | 310 | 25 | 6 |
| ***M. tuberculosis*** | 3,991 | 2,808 | 1,608 | 365 | 30 | 10 |
| ***P. falciparum*** | 5,363 | 2,599 | 818 | 284 | 28 | 13 |
| ***P. vivax*** | 5,342 | 2,359 | 822 | 268 | 24 | 13 |
| ***T. brucei*** | 7,793 | 1,530 | 300 | 138 | 13 | 6 |
| ***T. cruzi*** | 19,607 | 7,390 | 3,070 | 769 | 51 | 28 |
| ***T. gondii*** | 9,210 | 3,900 | 1,386 | 458 | 39 | 21 |
| **TOTAL** | **68,877** | **29,271** | **11,714** | **3,499** | **297** | **143** |

http://tropicaldisease.org

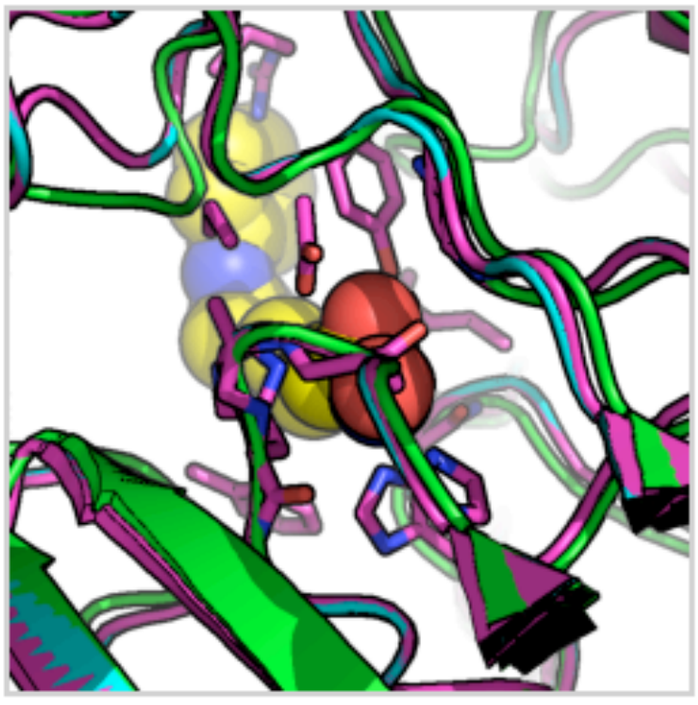
**Creative Commons (no viral!)**

# *L. major* Histone deacetylase 2 + Vorinostat

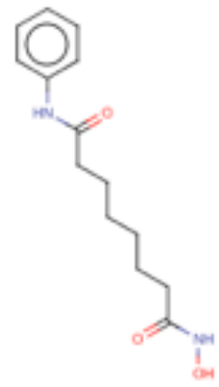
***Template 1t64A a human HDAC8 protein.***



| PDB | | Template | | Model | | Ligand | Exact | SupStr | SubStr | Similar |
|---|---|---|---|---|---|---|---|---|---|---|
| 1c3sA | 83.33/80.00 | 1t64A | 36.00/1.47 | LmjF21.0680.1.pdb | 90.91/100.00 | SHH | DB02546 | DB02546 | DB02546 | DB02546 |



**DB02546 Vorinostat**



Small Molecule; Approved; Investigational

**Drug categories:**

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

**Drug indication:**

*For the treatment of cutaneous manifestations in patients with cutaneous T-cell lymphoma who have progressive, persistent or recurrent disease on or following two systemic therapies.*

# *L. major* Histone deacetylase 2 + Vorinostat

## *Literature*

*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*

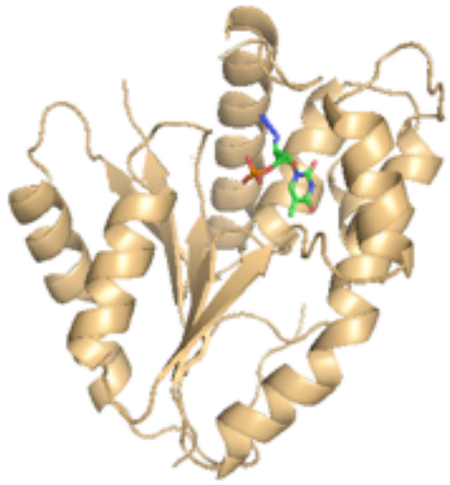
Departments of *Parasite Biochemistry and Cell Biology, ‡Medicinal Chemistry, and §Natural Products Drug Discovery, Merck Research Laboratories, P.O. Box 2000, Rahway, NJ 07065

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

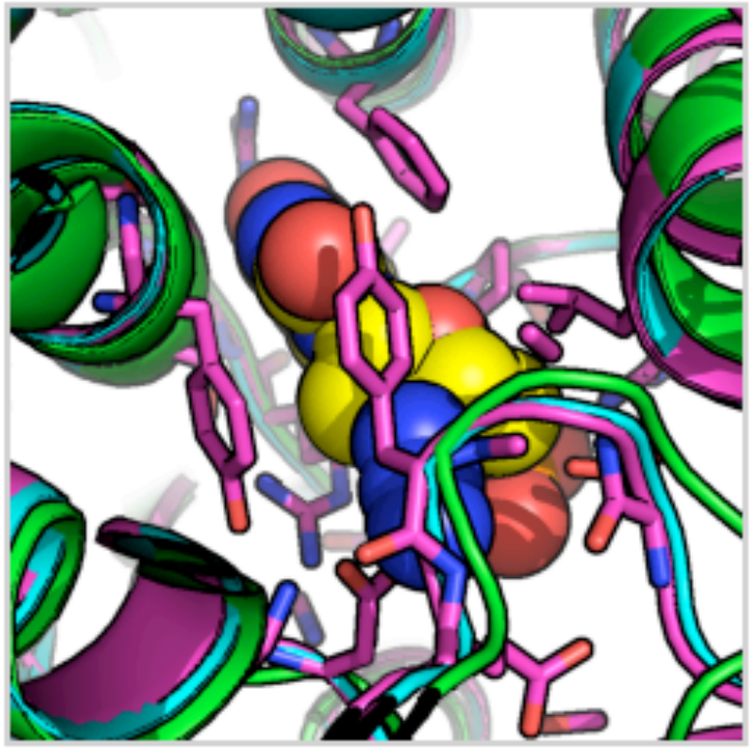

## Antimalarial and Antileishmanial Activities of Aroyl-Pyrrolyl-Hydroxyamides, a New Class of Histone Deacetylase Inhibitors

# *P. falciparum* tymidylate kinase + zidovudine

***Template 3tmkA a yeast tymidylate kinase.***



| PDB | | Template | | Model | | Ligand | Exact | SupStr | SubStr | Similar |
|---|---|---|---|---|---|---|---|---|---|---|
| 2tmkB | 100.00/100.00 | 3tmkA | 41.00/1.49 | PFL2465c.2.pdb | 82.61/100.00 | ATM | | DB00495 | | DB00495 |

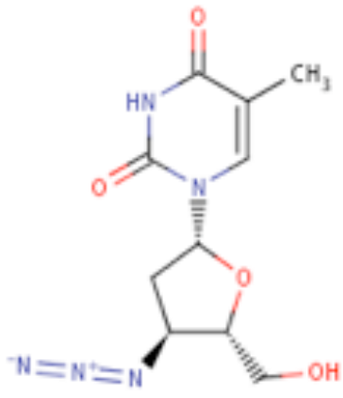


**DB00495 Zidovudine**

Small Molecule; Approved

**Drug categories:**

Anti-HIV Agents
Antimetabolites
Nucleoside and Nucleotide Reverse Transcriptase Inhibitors

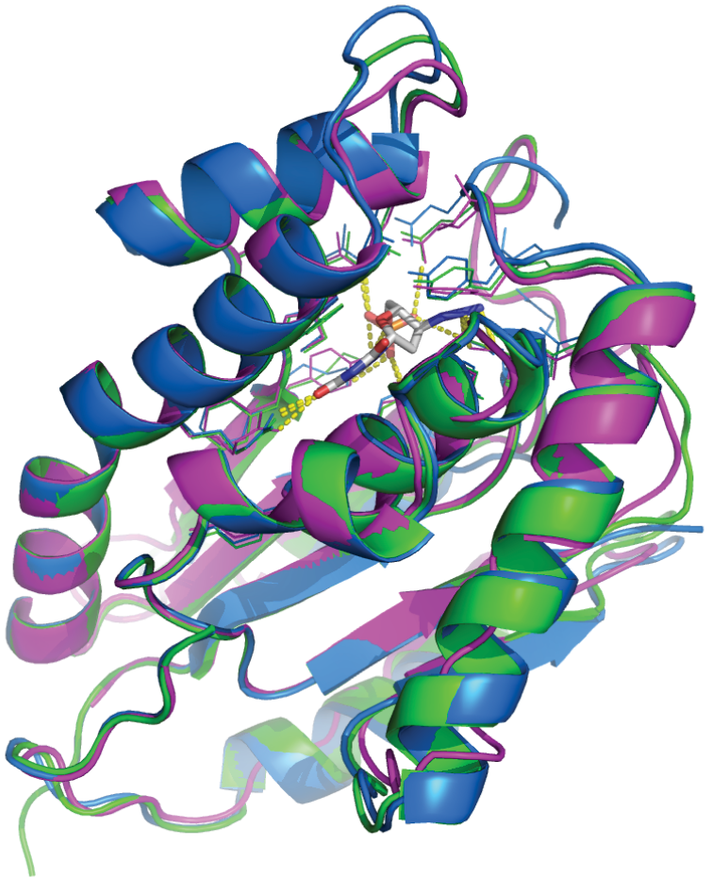


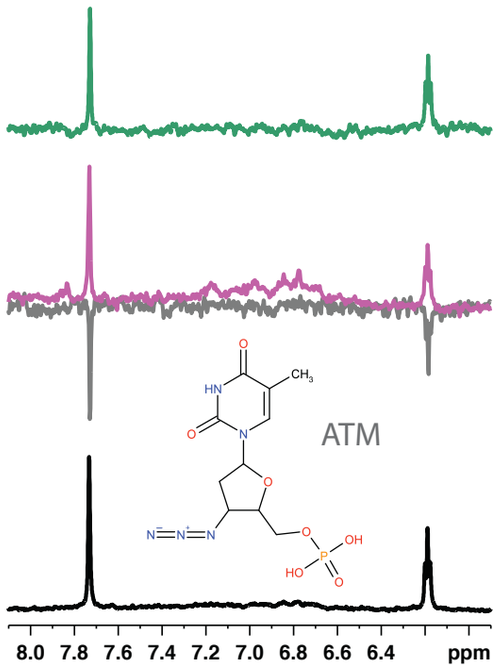

**Drug indication:**

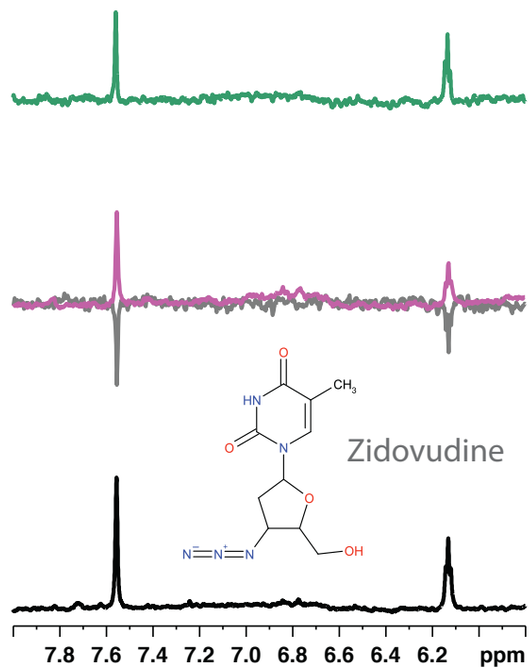
*For the treatment of human immunovirus (HIV) infections.*

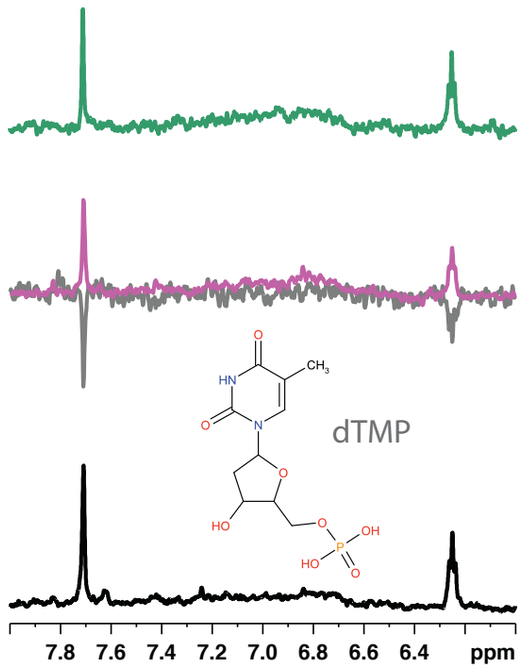
# *P. falciparum* thymidylate kinase + zidovudine

*NMR Water-LOGSY and STD experiments*





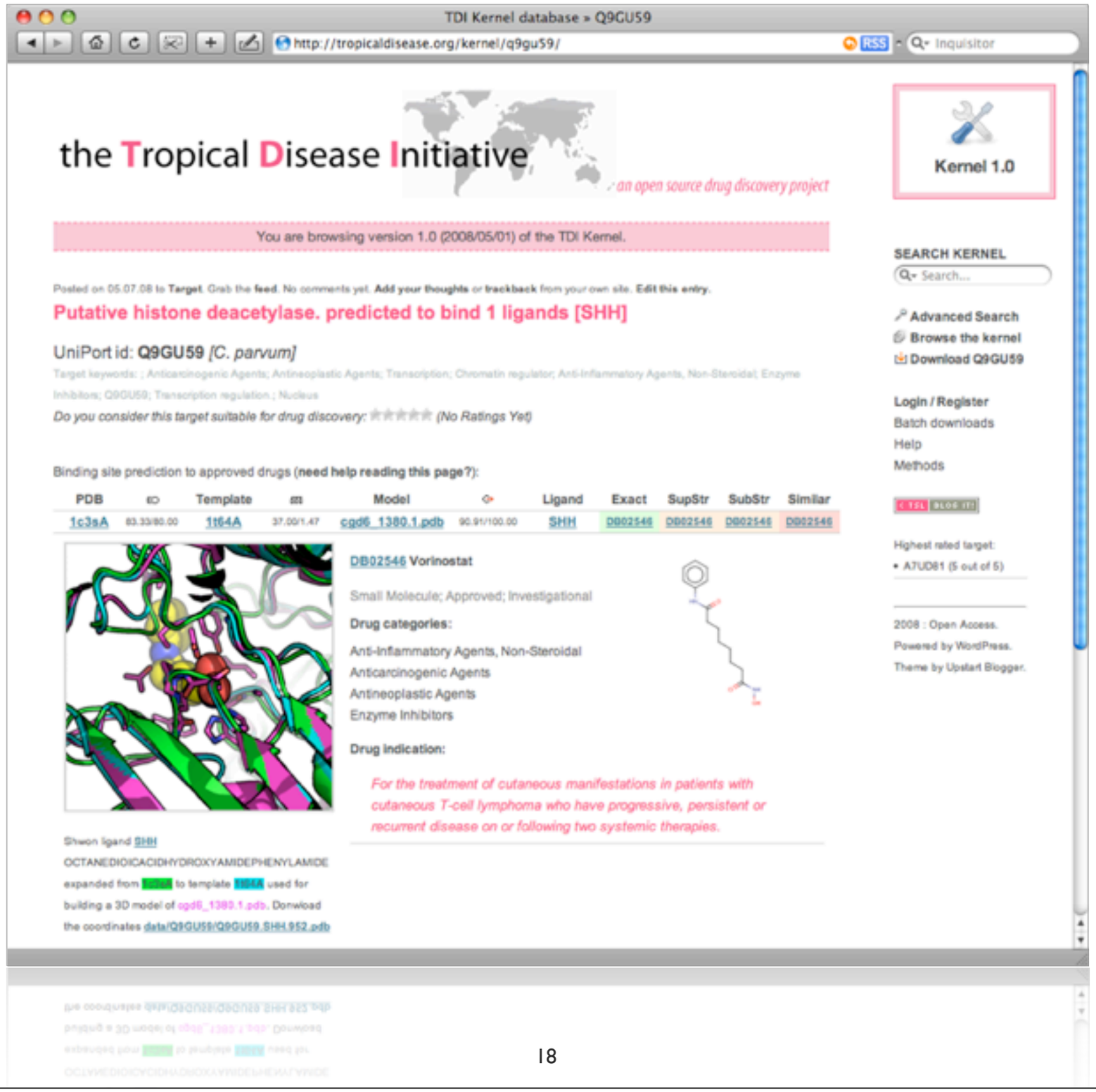




*Leticia Ortí, Rodrigo J. Carbajo, and Antonio Pineda-Lucena*

# TDI's kernel

http://tropicaldisease.org/kernel



# TDI reading...

http://tropicaldisease.org

The Economist (2004)
http://www.economist.com/node/2724420


Maurer *et al.* PLoS Medicine (2004) **1**:e56

*Open access, freely available online*

Neglected Diseases

## Finding Cures for Tropical Diseases: Is Open Source an Answer?

**Stephen M. Maurer*, Arti Rai, Andrej Sali**

Only about 1% of newly developed drugs are for tropical diseases, such as African sleeping sickness, dengue fever, and leishmaniasis [1]. While patent incentives and commercial pharmaceutical houses have made Western health care the envy of the world, the commercial model only works if companies can sell enough patented products to cover their research and development (R&D) costs. The model fails in the developing world, where few patients can afford to pay patented prices for drugs.

It is easy (and correct) to say that Western governments could solve this problem by paying existing institutions to focus on cures for tropical diseases. But sadly, there is not enough political will for this to happen. In any case, grants and patent incentives were never designed with tropical diseases in mind.

Two main kinds of proposals have been suggested for tackling the problem. The first is to ask sponsors—governments and charities—to subsidize developing-country purchases at a guaranteed price [2,3,4]. In the second approach, charities create nonprofit venture-capital firms ("Virtual Pharmas"), which look for promising drug candidates and then push drug development through contracts with corporate partners. In this article, we discuss the problems with these two approaches and suggest a third, "open source," approach to drug development, called the Tropical Diseases Initiative (TDI). We envisage TDI as a decentralized, Web-based, community-wide effort where scientists from laboratories, universities, institutes, and corporations can work together for a common cause (see www.tropicaldisease.org).

### Why Open Source?

The idea behind asking sponsors to subsidize developing country purchases at a guaranteed price is that this will prop up drug prices and restore incentives for developing new drugs [2,3,4]. In other words, it is a way of fixing the patent problem. However, subsidies have an important weakness: it is almost impossible to correctly determine how large the subsidy should be. In principle, the most cost-effective solution is to set a subsidy that just covers expected R&D costs. But how large is that? R&D costs are very poorly known, with published estimates ranging from $100 to $500 million per drug [5]. If the subsidy is set too low, companies cannot cover their R&D costs and nothing will happen. Set the subsidy too high, and the sponsor's costs skyrocket. To date, no sponsor has tried to implement these proposals.

In the "Virtual Pharma" approach, governments and philanthropies fund teams to search out and subsidize the most promising private and academic research. Examples include the Institute for One World Health (www.iowh.org), a not-for-profit pharmaceutical company funded mainly through private sources and the Gates Foundation, and the Drugs for Neglected Diseases Initiative (www.dndi.org), a public sector not-for-profit organization designed to mobilize resources for R&D of new drugs for neglected diseases.

Virtual Pharmas have clearly started to bear fruit, and are responsible for most candidate treatments for tropical diseases currently under development. For example, the Drugs for Neglected Diseases Initiative has a portfolio of nine projects spread out across the drug development pipeline for the treatment of leishmaniasis, sleeping sickness, Chagas disease, and malaria [6]. But Virtual Pharmas face three important problems. The first is similar to the problem faced by subsidy proposals: guessing private-sector R&D costs. One needs to understand what a product costs in order to negotiate the best possible price—and guessing wrong is likely to be expensive. Second, Virtual Pharma's development pipelines will run dry without more upstream

**Box 1. Possible Licenses for TDI Discoveries**

- A public-domain license that permits anyone to use the information for any purpose
- Licenses such as the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0) that permit anyone to use the information for any purpose, provided proper attribution is given
- Licenses such as the General Public License (www.opensource.org/licenses/gpl-license.php) that prohibit commercial use
- Licenses that permit commercial companies to obtain and exploit patents outside the developing world. These would allow Virtual Pharma to stretch its own R&D funds by letting corporate partners sell patented products to ecotourists, governments, and other consumers living in the industrialized world.

The Neglected Diseases section focuses attention either on a specific disease or describes a novel strategy for approaching neglected health issues in general.

**Citation:** Maurer SM, Rai A, Sali A (2004) Finding cures for tropical diseases: Is open source an answer? PLoS Med 1(3): e56.



**Abbreviations:** R&D, research and development; TDI, Tropical Diseases Initiative

Stephen M. Maurer is in the Goldman School of Public Policy, University of California, Berkeley, California, United States of America. Arti Rai is in the School of Law, Duke University, Durham, North Carolina, United States of America. Andrej Sali is in the Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry and the California Institute for Quantitative Biomedical Research, University of California, San Francisco, California, United States of America.

**Competing Interests:** The authors declare that they have no competing interests.

*To whom correspondence should be addressed. E-mail: maurer@econ.berkeley.edu

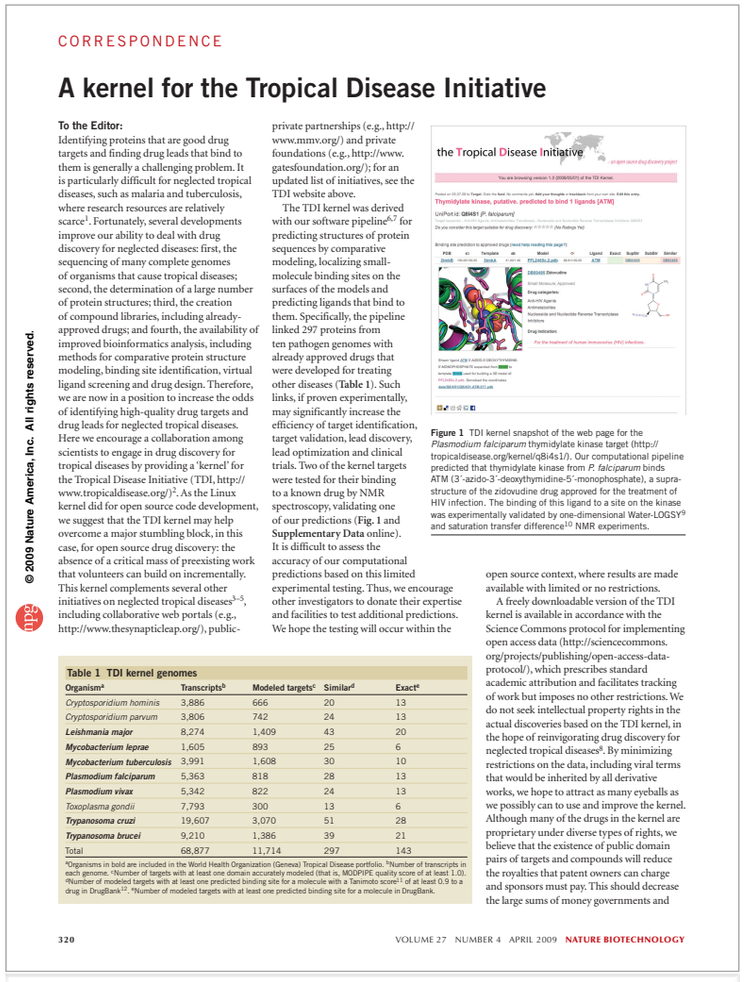
**DOI:** 10.1371/journal.pmed.0010056

# TDI's kernel

http://tropicaldisease.org/kernel

L. Orti *et al.*, *Nat Biotechnol* **27**, 320 (2009)

L. Orti *et al.*, *PLoS Negl Trop Dis* **3**, e418 (2009)



CORRESPONDENCE

## A kernel for the Tropical Disease Initiative

**To the Editor:**

Identifying proteins that are good drug targets and finding drug leads that bind to them is generally a challenging problem. It is particularly difficult for neglected tropical diseases, such as malaria and tuberculosis, where research resources are relatively scarce[1]. Fortunately, several developments improve our ability to deal with drug discovery for neglected diseases: first, the sequencing of many complete genomes of organisms that cause tropical diseases; second, the determination of a large number of protein structures; third, the creation of compound libraries, including already-approved drugs; and fourth, the availability of improved bioinformatics analysis, including methods for comparative protein structure modeling, binding site identification, virtual ligand screening and drug design. Therefore, we are now in a position to increase the odds of identifying high-quality drug targets and drug leads for neglected tropical diseases. Here we encourage a collaboration among scientists to engage in drug discovery for tropical diseases by providing a 'kernel' for the Tropical Disease Initiative (TDI, http://www.tropicaldisease.org/)[2]. As the Linux kernel did for open source code development, we suggest that the TDI kernel may help overcome a major stumbling block, in this case, for open source drug discovery: the absence of a critical mass of preexisting work that volunteers can build on incrementally. This kernel complements several other initiatives on neglected tropical diseases[3–5], including collaborative web portals (e.g., http://www.thesynapticleap.org/), public-private partnerships (e.g., http://www.mmv.org/) and private foundations (e.g., http://www.gatesfoundation.org/); for an updated list of initiatives, see the TDI website above.

The TDI kernel was derived with our software pipeline[6,7] for predicting structures of protein sequences by comparative modeling, localizing small-molecule binding sites on the surfaces of the models and predicting ligands that bind to them. Specifically, the pipeline linked 297 proteins from ten pathogen genomes with already approved drugs that were developed for treating other diseases (**Table 1**). Such links, if proven experimentally, may significantly increase the efficiency of target identification, target validation, lead discovery, lead optimization and clinical trials. Two of the kernel targets were tested for their binding to a known drug by NMR spectroscopy, validating one of our predictions (**Fig. 1** and **Supplementary Data** online). It is difficult to assess the accuracy of our computational predictions based on this limited experimental testing. Thus, we encourage other investigators to donate their expertise and facilities to test additional predictions. We hope the testing will occur within the open source context, where results are made available with limited or no restrictions.




**Figure 1** TDI kernel snapshot of the web page for the *Plasmodium falciparum* thymidylate kinase target (http://tropicaldisease.org/kernel/q8i4s1/). Our computational pipeline predicted that thymidylate kinase from *P. falciparum* binds ATM (3′-azido-3′-deoxythymidine-5′-monophosphate), a supra-structure of the zidovudine drug approved for the treatment of HIV infection. The binding of this ligand to a site on the kinase was experimentally validated by one-dimensional Water-LOGSY[9] and saturation transfer difference[10] NMR experiments.

A freely downloadable version of the TDI kernel is available in accordance with the Science Commons protocol for implementing open access data (http://sciencecommons.org/projects/publishing/open-access-data-protocol/), which prescribes standard academic attribution and facilitates tracking of work but imposes no other restrictions. We do not seek intellectual property rights in the actual discoveries based on the TDI kernel, in the hope of reinvigorating drug discovery for neglected tropical diseases[8]. By minimizing restrictions on the data, including viral terms that would be inherited by all derivative works, we hope to attract as many eyeballs as we possibly can to use and improve the kernel. Although many of the drugs in the kernel are proprietary under diverse types of rights, we believe that the existence of public domain pairs of targets and compounds will reduce the royalties that patent owners can charge and sponsors must pay. This should decrease the large sums of money governments and

**Table 1 TDI kernel genomes**

| Organism[a] | Transcripts[b] | Modeled targets[c] | Similar[d] | Exact[e] |
|---|---|---|---|---|
| *Cryptosporidium hominis* | 3,886 | 666 | 20 | 13 |
| *Cryptosporidium parvum* | 3,806 | 742 | 24 | 13 |
| ***Leishmania major*** | 8,274 | 1,409 | 43 | 20 |
| ***Mycobacterium leprae*** | 1,605 | 893 | 25 | 6 |
| ***Mycobacterium tuberculosis*** | 3,991 | 1,608 | 30 | 10 |
| ***Plasmodium falciparum*** | 5,363 | 818 | 28 | 13 |
| ***Plasmodium vivax*** | 5,342 | 822 | 24 | 13 |
| *Toxoplasma gondii* | 7,793 | 300 | 13 | 6 |
| ***Trypanosoma cruzi*** | 19,607 | 3,070 | 51 | 28 |
| ***Trypanosoma brucei*** | 9,210 | 1,386 | 39 | 21 |
| Total | 68,877 | 11,714 | 297 | 143 |

[a]Organisms in bold are included in the World Health Organization (Geneva) Tropical Disease portfolio. [b]Number of transcripts in each genome. [c]Number of targets with at least one domain accurately modeled (that is, MODPIPE quality score of at least 1.0). [d]Number of modeled targets with at least one predicted binding site for a molecule with a Tanimoto score[11] of at least 0.9 to a drug in DrugBank[12]. [e]Number of modeled targets with at least one predicted binding site for a molecule in DrugBank.





OPEN ACCESS Freely available online

PLoS NEGLECTED TROPICAL DISEASES

## A Kernel for Open Source Drug Discovery in Tropical Diseases

**Leticia Ortí[1,2], Rodrigo J. Carbajo[2], Ursula Pieper[3], Narayanan Eswar[3¤], Stephen M. Maurer[4], Arti K. Rai[5], Ginger Taylor[6], Matthew H. Todd[7], Antonio Pineda-Lucena[2], Andrej Sali[3*], Marc A. Marti-Renom[1*]**

**1** Structural Genomics Unit, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, Valencia, Spain, **2** Structural Biology Laboratory, Medicinal Chemistry Department, Centro de Investigación Príncipe Felipe, Valencia, Spain, **3** Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, California, United States of America, **4** Gould School of Law, University of Southern California, Los Angeles, California, United States of America, **5** School of Law, Duke University, Durham, North Carolina, United States of America, **6** The Synaptic Leap, San Ramon, California, United States of America, **7** School of Chemistry, University of Sydney, Sydney, New South Wales, Australia

**Abstract**

***Background:*** Conventional patent-based drug development incentives work badly for the developing world, where commercial markets are usually small to non-existent. For this reason, the past decade has seen extensive experimentation with alternative R&D institutions ranging from private–public partnerships to development prizes. Despite extensive discussion, however, one of the most promising avenues—open source drug discovery—has remained elusive. We argue that the stumbling block has been the absence of a critical mass of preexisting work that volunteers can improve through a series of granular contributions. Historically, open source software collaborations have almost never succeeded without such "kernels".

***Methodology/Principal Findings:*** Here, we use a computational pipeline for: (i) comparative structure modeling of target proteins, (ii) predicting the localization of ligand binding sites on their surfaces, and (iii) assessing the similarity of the predicted ligands to known drugs. Our kernel currently contains 143 and 297 protein targets from ten pathogen genomes that are predicted to bind a known drug or a molecule similar to a known drug, respectively. The kernel provides a source of potential drug targets and drug candidates around which an online open source community can nucleate. Using NMR spectroscopy, we have experimentally tested our predictions for two of these targets, confirming one and invalidating the other.

***Conclusions/Significance:*** The TDI kernel, which is being offered under the Creative Commons attribution share-alike license for free and unrestricted use, can be accessed on the World Wide Web at http://www.tropicaldisease.org. We hope that the kernel will facilitate collaborative efforts towards the discovery of new drugs against parasites that cause tropical diseases.

**Citation:** Ortí L, Carbajo RJ, Pieper U, Eswar N, Maurer SM, et al. (2009) A Kernel for Open Source Drug Discovery in Tropical Diseases. PLoS Negl Trop Dis 3(4): e418. doi:10.1371/journal.pntd.0000418

**Editor:** Timothy G. Geary, McGill University, Canada

**Received** December 29, 2008; **Accepted** March 23, 2009; **Published** April 21, 2009



**Funding:** MAM-R acknowledges the support from a Spanish Ministerio de Educación y Ciencia grant (BIO2007/66670). AS acknowledges the support from the Sandler Family Supporting Foundation and the National Institutes of Health (R01 GM54762, U54 GM074945, P01 AI035707, and P01 GM71790). AP-L acknowledges the support from a Spanish Ministerio de Ciencia e Innovación grant (SAF2008-01845). RJC acknowledges the support from the Ramon y Cajal Program of the Spanish Ministerio de Educación y Ciencia. We are also grateful for computer hardware gifts to AS from Ron Conway, Mike Homer, Intel, IBM, Hewlett-Packard, and NetApp. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: sali@salilab.org (AS); mmarti@cipf.es (MAM-R)

¤ Current address: DuPont Knowledge Center, Hyderabad, India

### Introduction

There is a lack of high-quality protein drug targets and drug leads for neglected diseases [1,2]. Fortunately, many genomes of organisms that cause tropical diseases have already been sequenced and published. Therefore, we are now in a position to leverage this information by identifying potential protein targets for drug discovery. Atomic-resolution structures can facilitate this task. In the absence of an experimentally determined structure, comparative modeling can provide useful models for sequences that are detectably related to known protein structures [3,4]. Approximately half of known protein sequences contain domains that can be currently predicted by comparative modeling [5,6]. This coverage will increase as the number of experimentally determined structures grows and modeling software improves. A protein model can facilitate at least four important tasks in the early stages of drug discovery [7]: prioritizing protein targets for drug discovery [8], identifying binding sites for small molecules [9,10], suggesting drug leads [11,12], and optimizing these leads [13–15].

Here, we address the first three tasks by assembling our computer programs into a software pipeline that automatically and on large-scale predicts protein structures, their ligand binding sites, and known drugs that interact with them. As a proof of principle, we applied the pipeline to the genomes of ten organisms that cause tropical diseases ("target genomes"). We also experimentally tested two predicted drug-target interactions using Nuclear Magnetic



Get our papers here: http://sgu.bioinfo.cipf.es/home/publications

# OpenPool/Lab GSK

http://ntdpool.org


Gamo *et al.* Nature (2010) vol. 465 (7296) pp. 305-10

Vol 465|20 May 2010|doi:10.1038/nature09107 nature

ARTICLES

## Thousands of chemical starting points for antimalarial lead identification

Francisco-Javier Gamo[1], Laura M. Sanz[1], Jaume Vidal[1], Cristina de Cozar[1], Emilio Alvarez[1], Jose-Luis Lavandera[1], Dana E. Vanderwall[2], Darren V. S. Green[3], Vinod Kumar[4], Samiul Hasan[4], James R. Brown[4], Catherine E. Peishoff[5], Lon R. Cardon[6] & Jose F. Garcia-Bustos[1]

**Malaria is a devastating infection caused by protozoa of the genus *Plasmodium*. Drug resistance is widespread, no new chemical class of antimalarials has been introduced into clinical practice since 1996 and there is a recent rise of parasite strains with reduced sensitivity to the newest drugs. We screened nearly 2 million compounds in GlaxoSmithKline's chemical library for inhibitors of *P. falciparum*, of which 13,533 were confirmed to inhibit parasite growth by at least 80% at 2 μM concentration. More than 8,000 also showed potent activity against the multidrug resistant strain Dd2. Most (82%) compounds originate from internal company projects and are new to the malaria community. Analyses using historic assay data suggest several novel mechanisms of antimalarial action, such as inhibition of protein kinases and host–pathogen interaction related targets. Chemical structures and associated data are hereby made public to encourage additional drug lead identification efforts and further research into this disease.**

With approximately 243 million cases and 863,000 attributed deaths reported globally in 2009 (ref. 1), malaria is one of the most severe infectious diseases, primarily affecting the world's most disadvantaged populations. Of the four typically recognized *Plasmodium* species causing disease in humans, *Plasmodium falciparum* causes most mortality, mainly in children below the age of 5, and *Plasmodium vivax* most morbidity, additionally representing a reservoir of latent infection that hampers current control and future elimination efforts[2]. No new class of antimalarials has been introduced into clinical practice since 1996 (ref. 3), owing to the intrinsic difficulties in discovering and developing new antimicrobials, as well as a relative lack of public and private resource commitment towards antimalarial research. Today, the last class of widely efficacious drugs, the artemisinins, is being compromised by the rise of *P. falciparum* strains with reduced clinical response to artemisinin-containing drug combinations[4–6]. The genomics revolution has not yet led to new antimalarial medicines and target-based lead discovery has produced disappointing results, generally for lack of whole-cell activity as documented for antibacterials[7]. To secure that property in all chemical starting points for new antimalarial leads, we have tested the approximately 2 million-compound library used for high throughput screening at GlaxoSmithKline (GSK) for inhibitors of *P. falciparum*'s intraerythrocytic cycle, the *Plasmodium* species causing the highest mortality and the parasite growth phase responsible for disease symptoms as well as being amenable to *in vitro* culture. Here we describe 13,533 compounds confirmed to inhibit parasite growth by more than 80% at 2 μM concentration. Only 15% displayed some cytotoxicity in that they inhibited proliferation of the HepG2 human hepatoma cell line by more than 50% at 10 μM. All of these proven plasmodial inhibitors, of which 82% were previously proprietary and thus unknown to the general research community, are hereby made public to accelerate the pace of drug development for malaria.

### Tres Cantos antimalarial compound set (TCAMS)

The 1,986,056 compounds present in GSK's screening collection in January 2009 were tested for inhibition of *P. falciparum* 3D7 at 2 μM under *in vitro* conditions described in Methods. 19,451 primary hits inhibiting parasite growth by more than 80% were obtained. Fresh samples of these primary hits were tested in two independent experiments and compounds displaying 80% or higher inhibition of parasite growth in at least two of the three assay runs were considered confirmed hits. 13,533 compounds were identified using this protocol (confirmation rate > 70%). We did not detect any compounds in this set as non-specific inhibitors of the biochemical readout system by testing directly for inhibition of lactate dehydrogenase (LDH) in *P. falciparum* extracts (Methods). Evidence of cytotoxicity against human hepatoma HepG2 cells (a widely used *in vitro* marker for liver toxicity[8]), or interference with the luciferase reporter system used in the cytotoxicity assay (Methods), was observed in just 1,982 of the compounds when tested at 10 μM. This relative lack of non-specific cell toxicity is probably due in part to the low (2 μM) primary screening concentration used[9]. Estimation of the concentrations producing 50% inhibition of *P. falciparum* growth ($XC_{50}$, see Methods) indicated that most compounds are sub-micromolar inhibitors. The full compound set (TCAMS) and data table (Supplementary Table 1 and available at http://www.ebi.ac.uk/chemblntd) contains 13,533 compound entries. We have detected 139 of these as variations in salt form or stereochemistry of 68 parent structures, which make good internal controls for the biological assay data. They appear as different compounds with the same structure. When the stereochemistry is resolved it shows in the SMILES structural code in Supplementary Table 1 and in the Chembl-NTD database (http://www.ebi.ac.uk/chemblntd).

Representatives from all but one class of clinically used antimalarials have been recovered in the screen, providing additional validation

[1]Tres Cantos Medicines Development Campus, GlaxoSmithKline, Severo Ochoa 2, 28760 Tres Cantos, Spain. [2]Computational and Structural Chemistry, GlaxoSmithKline, Five Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA. [3]Computational and Structural Chemistry, GlaxoSmithKline, Medicines Research Centre, Gunnels Wood Road, Hertfordshire, Stevenage SG1 2NY, UK. [4]Computational Biology, Quantitative Sciences, GlaxoSmithKline, 1250 South Collegeville Road, Collegeville, Pennsylvania 19426, USA. [5]Computational and Structural Chemistry, GlaxoSmithKline, 1250 South Collegeville Road, Collegeville, Pennsylvania 19426, USA. [6]Quantitative Sciences, GlaxoSmithKline, 709 Swedeland Road, King of Prussia, Pennsylvania 19406, USA.

# OSDD Book

***Mid 2012***

CRC Press
Taylor & Francis Group

- Dr. Lluís Ballell. GlaxoSmithKline. Madrid, Spain.
- Dr. Barry Bunin. CDD. San Francisco. CA. USA.
- Dr. Sean Ekins. CDD. San Francisco. CA. USA.
- Prof. Andrew Hopkins. Dundeee U. UK.
- Dr. Marc A. Marti-Renom. CIPF. Valencia. Spain.
- Prof. Steven Maurer. Goldman School of Public Polity. Berkeley. CA. USA.
- Bernard Munos. Corporate Strategy, Eli Lilly and Company, Indianapolis, Indiana, USA.
- Prof. John Overington. ChEMBL Group. EMBL. Germany.
- Dr. Antonio Pineda-Lucena. CIPF. Valencia. Spain.
- Prof. Arti Rai. Duke University. NC. USA.
- Prof. Andrej Sali. UCSF. San Francisco. CA. USA.
- Ginger Taylor. The Synaptic Leap. San Ramon. CA. USA.
- Dr. Zakir Thomas. Project Director. OSDD. India.
- Dr. Matthew Todd. The University of Sydney. Australia.
- Marleen Wynants. Cross Talks. Brussels. Belgium.

# GeMoA

http://gemoa-era.net

TARGET-LIGAND

**WP1**

**GSK**
High-Throughput Screening
MedChem
Synthetic chemistry

**CIPF**
Protein-ligand binding prediction

**WP4**

**EMBL**
Functional assays
Heterologous expression
Protein purification
Protein crystallization

**GSK**
MedChem
Synthetic chemistry
Enzymatic assays

GENOME-WIDE

**WP2**

**CIPF**
Comparative structure prediction
Protein-protein network analysis
Protein-ligand network analysis
Genome-wide functional annotation

**IPBS**
Transcriptomics
Genotyping

**WP3**

**IPBS**
Transcriptomics
Genotyping
Imaging

**IP**
Genomics
Mutant libraries
Genome sequencing
Imaging

# Acknowledgments


http://sgu.bioinfo.cipf.es
http://tropicaldisease.org
http://gemoa-era.net

**COMPARATIVE MODELING**
**Andrej Sali**
M. S. Madhusudhan
**Narayanan Eswar**
Min-Yi Shen
**Ursula Pieper**
Ben Webb
Maya Topf (Birbeck College)

**MODEL ASSESSMENT**
Francisco Melo (CU)
Alejandro Panjkovich (CU)

**NMR**
**Antonio Pineda-Lucena**
**Leticia Ortí**
**Rodrigo J. Carbajo**

**MAMMOTH**
Angel R. Ortiz

**FUNCTIONAL ANNOTATION**
Fatima Al-Shahrour
Joaquin Dopazo

**MODEL ASSESSMENT**
David Eramian
Min-Yi Shen
Damien Devos

**Tropical Disease Initiative**
**Marc A. Marti-Renom (CIPF)**
**Stephen Maurer (UC Berkeley)**
**Arti Rai (Duke U)**
**Andrej Sali (UCSF)**
**Ginger Taylor (TSL)**
**Matthew Todd (U Sydney)**

**GeMoA**
**Marc A. Marti-Renom (CIPF)**
**Lluís Ballell (GSK)**
**Olivier Neyrolles (IPBS)**
**Matthias Wilmanns (EMBL)**
**Brigitte Cicquel (IP)**