Published in final edited form as:

Nat Biotechnol. 2009 April ; 27(4): 320–321. doi:10.1038/nbt0409-320.

A kernel for the Tropical Disease Initiative

Leticia Ortí^{1,2}, Rodrigo J. Carbajo², Ursula Pieper³, Narayanan Eswar^{3,†}, Stephen M. Maurer⁴, Arti K. Rai⁵, Ginger Taylor⁶, Matthew H. Todd⁷, Antonio Pineda-Lucena², Andrej Sali^{3,*}, and Marc A. Marti-Renom^{1,*}

¹ Structural Genomics Unit, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, 46012 Valencia, Spain

² Structural Biology Laboratory, Medicinal Chemistry Department, Centro de Investigación Príncipe Felipe, 46012 Valencia, Spain

³ Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California at San Francisco, San Francisco, CA 94158-2330, USA

⁴ University of Southern California, Gould School of Law, Los Angeles, CA 90089-0071, USA

⁵ School of Law, Duke University, Durham, NC 27708-0360, USA

⁶ The Synaptic Leap, <http://www.thesynapticleap.org>

⁷ School of Chemistry, University of Sydney, Sydney NSW2006, Australia

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 scarce¹. Fortunately, several developments improve our ability to deal with drug discovery for neglected diseases: (i) the sequencing of many complete genomes of organisms that cause tropical diseases; (ii) the determination of a large number of protein structures; (iii) the creation of compound libraries including already approved drugs; and (iv) 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. Based on the impact of the Linux kernel on the open source code development, we suggest that the TDI kernel may help overcome a major stumbling block for open source drug discovery: the absence of a critical mass of preexisting work that volunteers can build on incrementally. This kernel complements a number of other initiatives on neglected tropical diseases^{3–5} including collaborative Web portals (e.g., The Synaptic Leap; <http://www.thesynapticleap.org>), Public-Private-Partnerships (e.g., Medicines for Malaria Venture; <http://www.mmv.org>), and private foundations (e.g., Bill and Melinda Gates Foundation;

*Corresponding authors: Marc A. Marti-Renom, Structural Genomics Unit, Bioinformatics and Genomics Department., Centro de Investigación Príncipe Felipe., Av. Autopista del Saler, 16, 46012 Valencia, Spain, Tel: +34 96 3289680, Fax: +34 96 3289701, mmarti@cipf.es, Andrej Sali, Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California, Institute for Quantitative Biosciences, Byers Hall 503B, 1700 4th Street, University of California at San Francisco, San Francisco, CA 94158-2330 US, Tel +1 (415) 514-4228; Fax +1 (415) 514-4231 sali@salilab.org.

[†]Current address: DuPont Knowledge Center, Hyderabad - 500 078, India.

<http://www.gatesfoundation.org>); for an updated list of initiatives, see <http://www.tropicaldisease.org>.

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 (Figure 1). 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.

The TDI kernel is freely downloadable in accordance with the Science Commons protocol for implementing open access data (<http://sciencecommons.org/projects/publishing/open-access-data-protocol/>) that 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 to reinvigorate drug discovery for neglected tropical diseases⁸. 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. While 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 foundations need to invest to turn validated targets and candidate drugs into actual treatments.

Our list of likely drug leads and their targets must be validated and extended using additional lines of evidence by computation and, most importantly, wet lab experiments. We are committed to helping other researchers add their protocols and analyses to the current kernel. For example, computational docking, biophysical analysis, activity assays, site-directed mutagenesis, and synthetic chemistry could be performed for all predicted targets. Unfortunately, such techniques are usually very expensive and thus not feasible on a genomic scale by a single research group. The main goal of our exercise was to narrow down the number of targets and identify their putative ligands for experimental follow-up, so that the overall process is faster, more thorough, and less expensive. The TDI kernel's list of “hits” does not exhaust the ten target genomes. Researchers who want TDI to investigate additional candidates should contact us or engage in online discussions at our collaborative portal (The Synaptic Leap, <http://www.thesynapticleap.org>).

Acknowledgments

We acknowledge the support from the Spanish Ministerio de Educación y Ciencia (BIO2007/66670 and SAF2008-01845), the NIH (R01 GM54762, U54 GM074945, P01 AI035707, and P01 GM71790), and the Sandler Family Supporting Foundation.

References

1. Nwaka S, Ridley RG. Virtual drug discovery and development for neglected diseases through public-private partnerships. *Nat Rev Drug Discov*. 2003; 2:919–928. [PubMed: 14668812]

2. Maurer SM, Rai A, Sali A. Finding cures for tropical diseases: is open source an answer? *PLoS Med.* 2004; 1:e56. [PubMed: 15630466]
3. Kepler T, et al. Open Source Research - The power of Us. *Aust J Chem.* 2006; 59:291–294.
4. Singh S. India takes an open source approach to drug discovery. *Cell.* 2008; 133:201–203. [PubMed: 18423188]
5. Agüero F, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov.* 2008; 7:900–907. [PubMed: 18927591]
6. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993; 234:779–815. [PubMed: 8254673]
7. Marti-Renom MA, et al. The AnnoLite and AnnoLyze programs for comparative annotation of protein structures. *BMC Bioinformatics.* 2007; 8 (Suppl 4):S4. [PubMed: 17570147]
8. Munos B. Can open-source R&D reinvigorate drug research? *Nat Rev Drug Discov.* 2006; 5:723–729. [PubMed: 16915233]
9. Gower JC. A general coefficient of similarity and some of its properties. *Biometrics.* 1971; 27:857–871.
10. Wishart DS, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008; 36:D901–906. [PubMed: 18048412]
11. Dalvit C, et al. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J Biomol NMR.* 2000; 18:65–68. [PubMed: 11061229]
12. Meyer B, Peters T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl.* 2003; 42:864–890. [PubMed: 12596167]

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 1D Water-LOGSY¹¹ and Saturation Transfer Difference¹² NMR experiments.

Table 1

TDI kernel genomes.

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

^aOrganisms in bold are included in the WHO Tropical Disease portfolio.^bNumber of transcripts in each genome.^cNumber of targets with at least one domain accurately modeled (*i.e.*, MODPIPE quality score of at least 1.0).^dNumber of modeled targets with at least one predicted binding site for a molecule with a Tanimoto score⁹ of at least 0.9 to a drug in DrugBank¹⁰.^eNumber of modeled targets with at least one predicted binding site for a molecule in DrugBank.