

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 scarce¹. 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.mmwv.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

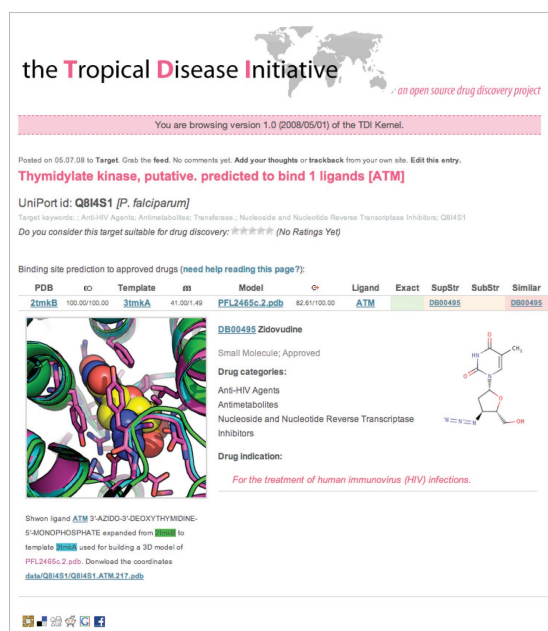


Figure 1 TDI kernel snapshot of the web page for the *Plasmodium falciparum* thymidylate kinase target (<http://tropicaldisease.org/kernel/q8i4s1/>). Our computational pipeline predicted that thymidylate kinase from *P. falciparum* binds ATM (3'-azido-3'-deoxythymidine-5'-monophosphate), a suprastructure 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-LOGSY⁹ and saturation transfer difference¹⁰ NMR experiments.

open source context, where results are made available with limited or no restrictions.

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 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. 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
<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 World Health Organization (Geneva) Tropical Disease portfolio. ^bNumber of transcripts in each genome. ^cNumber of targets with at least one domain accurately modeled (that is, 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 score¹¹ of at least 0.9 to a drug in DrugBank¹². ^eNumber of modeled targets with at least one predicted binding site for a molecule in DrugBank.

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 (<http://www.thesynapticleap.org/>).

Note: Supplementary information is available on the *Nature Biotechnology* website.

ACKNOWLEDGMENTS

We acknowledge the support from the Spanish Ministerio de Ciencia e Innovación (BIO2007/66670 and SAF2008-01845), the National Institutes of Health (R01 GM54762, U54 GM074945, P01 AI035707, and P01 GM71790), and the Sandler Family Supporting Foundation.

Leticia Orti^{1,2}, Rodrigo J Carbajo², Ursula Pieper³, Narayanan Eswar^{3,8}, Stephen M Maurer⁴, Arti K Rai⁵, Ginger Taylor⁶, Matthew H Todd⁷, Antonio Pineda-Lucena², Andrej Sali³ & Marc A Marti-Renom¹

¹Structural Genomics Unit, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, Valencia, Spain. ²Structural Biology Laboratory, Medicinal Chemistry Department, Centro de Investigación Príncipe Felipe, Valencia, Spain. ³Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California at San Francisco, San Francisco, California, USA. ⁴University of Southern California, Gould School of Law, Los Angeles, California, USA. ⁵School of Law, Duke University, Durham, North Carolina, USA. ⁶The Synaptic Leap, San Ramon, California, USA. ⁷School of Chemistry, University of Sydney, Sydney, Australia. ⁸Present address: DuPont Knowledge Center, Hyderabad, India. e-mail: mmarti@cipfes.org or sali@salilab.org

- Nwaka, S. & Ridley, R.G. *Nat. Rev. Drug Discov.* **2**, 919–928 (2003).
- Maurer, S.M., Rai, A. & Sali, A. *PLoS Med.* **1**, e56 (2004).
- Kepler, T. *et al. Aust. J. Chem.* **59**, 291–294 (2006).
- Singh, S. *Cell* **133**, 201–203 (2008).
- Aguero, F. *et al. Nat. Rev. Drug Discov.* **7**, 900–907 (2008).
- Sali, A. & Blundell, T.L. *J. Mol. Biol.* **234**, 779–815 (1993).
- Marti-Renom, M.A. *et al. BMC Bioinformatics* **8 Suppl.** **4**, S4 (2007).
- Munos, B. *Nat. Rev. Drug Discov.* **5**, 723–729 (2006).
- Dalvit, C. *et al. J. Biomol. NMR* **18**, 65–68 (2000).
- Meyer, B. & Peters, T. *Angew. Chem. Int. Ed. Engl.* **42**, 864–890 (2003).
- Gower, J.C. *Biometrics* **27**, 857–871 (1971).
- Wishart, D.S. *et al. Nucleic Acids Res.* **36**, D901–D906 (2008).

Of Newtons and heretics

To the Editor:

In a commentary for *Nature Biotechnology* last year¹, Joachim Boldt and Oliver Müller argue that synthetic biology poses ethical issues beyond those of traditional genetic engineering. This is because synthetic biology aims, among other things, to create biological systems with features that might never have been part of living organisms before. Moreover, the authors point out that synthetic biology represents a radical shift from manipulation to creation—a shift to organisms, “significant portions” of which are “designed by humans,” conferring new responsibilities on human beings as creators and questioning our understanding of life and living organisms. According to data we describe here, however, this interpretation contrasts with the views of the scientific community itself. As part of the European SYNBIOSAFE project on the ethics and safety of synthetic biology², we have interviewed 20 European researchers working in synthetic biology to explore their ideas and attitudes regarding the ethical implications of their work, a discussion we hoped would contribute to the ethical, safety and regulatory discourse in Europe. To complement the discussion begun by Boldt and Müller, we present the results of this survey below.

The survey was carried out between June 2007 and January 2008 (see **Supplementary Methods** online), during which we interviewed leading scientists from the European synthetic biology community, which we defined as those persons and institutions that coordinate (or participate in) one of the EC-FP6-NEST³-funded synthetic biology projects. As the synthetic biology community in Europe was still, at

the time, defining its identity, we wanted to survey the opinions and expectations of researchers regarding their field. Moreover, given the past debates about biotechnologies in Europe, scientists here might have had, we believed, a significant contribution to make regarding the ethics, regulation and

perception of a new biotech that may or may not match the prevailing views in the US community. The survey took the form of an interview, in which we asked whether there were ethical issues associated with synthetic biology, whether the creation of artificial organisms posed ethical problems and whether synthetic biology ethical debates resembled those associated with earlier

biotechnologies (interested readers should contact us for a copy of the interview guide). Answers are listed in full in **Supplementary Methods** online.

According to most of the interviewees, and quite to the contrary of what Boldt and Müller suggest, the prevailing view was that synthetic biology raises no particular ethical issues in itself. Escaping the natural is part of what man does and many felt that where there are ethical questions, they are almost entirely restricted to unintended negative effects and intended misuse. Where respondents did recognize other moral implications, these were almost exclusively related to practical applications of synthetic biology. Some respondents, for example, felt that the only important issue was the application of synthetic biology to higher organisms or that the manipulation of the human genome should be prohibited. Others had more difficulty in pinpointing specific ethical issues but showed discomfort at bringing wholesale changes to the genome, worries perhaps related to the ‘instrumentalization’ of living organisms.

