Open data in drug discovery and development: lessons from malaria

Timothy N. C. Wells, Paul Willis, Jeremy N. Burrows and Rob Hooft van Huijsduijnen

There is a growing consensus that drug discovery thrives in an open environment. Here, we describe how the malaria community has embraced four levels of open data – open science, open innovation, open access and open source – to catalyse the development of new medicines, and consider principles that could enable open data approaches to be applied to other disease areas.

Scientific discoveries are built around cycles of thesis and antithesis; innovation is driven forwards by intensive peer review. Drug discovery should be no exception. In the early years, drug discovery was a highly collaborative effort between academia and industry, but from around the 1970s, large pharmaceutical companies became siloed and secretive organizations, and openness about early-stage drug discovery was thought to compromise the competitive edge of a chemical series or disease target. Academic research has also become more secretive, often driven by the need for high-impact, first-author publications, or unrealistic financial expectations by university technology transfer officers.

However, in recent years there has been a growing appreciation from both academic organizations and pharmaceutical companies of the value of greater openness, leading to the establishment of pre-competitive collaborations and initiatives based on various levels of openness. Overcoming the instinct to hoard requires both an understanding of the benefits of collaboration and an acceptance that intellectual property (IP) brings with it responsibilities as much as rights. Because the discovery of new antimalarials is driven primarily by humanitarian motives rather than by profit expectations, the Medicines for Malaria Venture (MMV) is in an ideal position to pioneer new models of drug development that prioritize open scientific innovation. The steps in this process may be illustrated by four levels of openness with regard to data: open science, open innovation, open access and open source, each of which reflects increasing collaboration.

Timothy N. C. Wells, Paul Willis, Jeremy N. Burrows and Rob Hooft van Huijsduijnen are at Medicines for Malaria Venture, 20 Route de Pré-Bois, Geneva 1215, Switzerland

Correspondence to T.N.C.W.
wellst@mmv.org

doi:10.1038/nrd.2016.154
Published online 12 Aug 2016;
corrected online 18 Aug 2016

Degrees of openness in malaria research

Open science. The most basic level of openness comprises the timely deposition of data in the public domain for the benefit of the community. Historically, this has covered 'hard data' such as DNA and protein sequences and crystallography data. Theoretically, it should also include all biological data reported in publications, which should also be verifiable. Without quality standards, flawed data are introduced into analyses, which has a profound cost to society[1]. MMV's strategy has been to validate primary data by testing sets of known standards, blinded to enable cross-validation of results. We have also used standardized *in vivo* pharmacology assays in our decision making, enabling clear head-to-head comparisons of molecules.

Open innovation. At this level of openness, two or more partners share all data openly within a 'walled garden', with new parties joining only by invitation. An example of this strategy is the collaboration to design improved electric car batteries within the pre-competitive space (the US Advanced Battery Consortium). Examples in drug development include the (Eli) Lilly Phenotypic Drug Discovery Initiative, the GlaxoSmithKline (GSK) clinical trial data platform, the Structural Genomics Consortium and the Single Nucleotide Polymorphism Database (dbSNP) (see Further information). Such projects need a trusted neutral broker and financial means to drive them. MMV's open innovation network of standardized assay partnerships is made available to all new project partners, thereby facilitating open and cost-effective collaboration (see Supplementary information S1 (box)). In malaria, because of the relatively large emphasis on phenotypic screening hits, several teams have identified compounds that act on the same biological pathways. Bringing groups together and sharing data is particularly beneficial in these situations and is something our partners recognize as adding value. Timing is key: sharing data too early could stifle creativity, sharing it too late risks non-productive duplication and weak differentiation. Examples of productive open innovation include project teams focused on *Plasmodium falciparum* phosphatidylinositol-4 kinase (PfPI4K) and Ca^{2+}-ATPase (PfATP4), where the availability of several different scaffolds has de-risked the portfolio by promoting clear differentiation of chemotypes.

A strict legal framework is necessary at this stage. For MMV, IP is a responsibility. MMV partners who file patents have an obligation to drive their projects forwards or pass them on to someone who can. Openness requires a realistic understanding of value; most of our projects

are pre-competitive, and many focus on new targets and assay formats. It is a mistake to keep these secret. However, patents are still key to engaging and securing investment from pharmaceutical companies. Clarity over IP ownership ensures clarity over freedom to operate, and also allows some control over the quality of generic drug copies through licences. Legal frameworks can also ensure affordable price commitments in regions with endemic diseases.

Open access. Open access makes data, compounds and publications available to all, with the intent to maximize subsequent use. From 2010, MMV has coordinated the acquisition of data from *Plasmodium* phenotypic screening of more than 7 million compounds. Almost 20,000 hits, with their potencies and structures, were posted on the ChEMBL website[3]. Despite the previous clamour for more data to be in the public domain, this publication did not materially change the way that malaria researchers engaged in drug discovery; the main hurdle for wider development remained access to the compounds themselves. In 2011, therefore, we launched the Malaria Box. Using informatics and 'pragmatics', 400 compounds were selected, purchased, retested (to improve reproducibility) and plated for distribution and screening. These plates were freely available to all. By February 2016, more than 250 collections had been distributed, including to organizations in Africa, Asia and South America, and the set has been tested against protozoans and other pathogens, including in mechanistic screens. So far, 42 publications have ensued, and there is now a high-quality database comparing these compounds across numerous assays[4].

These projects have been particularly useful as introductions to drug discovery for small, pathogen-focused groups in disease-endemic countries. Recently, we launched the Pathogen Box: 400 compounds based on new hits from phenotypic screens against a wider variety of pathogenic organisms. We plan to build a similar compound collection against pandemic viruses and bacteria — extremely relevant in the light of the recent Ebola and Zika virus outbreaks. Using relatively modest resources, these initiatives harness the dispersed wealth of knowledge of pathogen biology and assay technologies for neglected diseases, sharing expertise and helping to build a community around high-value chemistry resources. Open access enables the movement of compounds around the world.

It is still not clear to what extent these collaborations will lead to new projects or new products. Several groups have generated new IP out of compounds from the Malaria Box (open innovation). Others in the malaria community feel an absolute imperative to ultimately put everything into the public domain (open science). Time will tell which approach will make the largest contribution to drug discovery for neglected tropical diseases.

Open source. The open source model is even more collaborative than open access: all project data and structures are laid bare, and the wider community is invited to fully engage with advice, synthesis, testing and in-kind technical support. MMV's main experiment in open source has been with the Open Source Malaria project. This collaboration, with a champion of open source chemistry, Mat Todd, has seen significant engagement and interest from scientists, although financial resources have been limited, perhaps driven by the perspective from the computing world, in which open source work can be managed for free. One example of this project's output is a collaboration with 40 undergraduate students who will prepare new analogues of a family of compounds as part of their laboratory course (see Further information). New open source projects require proper resourcing, excellent communication with the community, simple online interfaces and strong, experienced leadership and guidance.

Outlook

Open innovation has successfully brought new molecules into the clinic for malaria. During the past few years there has been a call for more open access compound collections, of which the Malaria and Pathogen Boxes are pioneers. Open access data, compounds and publications then naturally stimulate open source thinking and projects. In theory, this approach can be extended to other therapeutic areas: high-value chemical tools and compound collections can be shared, once it is accepted that discovery is a pre-competitive activity in all disease areas, and that optimization is necessary to make a drug. An example of this was the wide sharing of the bromodomain inhibitor JQ1, which spurred further drug discovery efforts for this target class[5]. Many therapeutic areas — from antibacterials to central nervous system disorders and psychiatric diseases — suffer from market failure; although they have commercial markets these are more than offset by development costs and the difficulty of linking targets to disease. Open innovation, especially in situations of market failure, is one of the keys to unlocking the potential for discovery, and may lead to a new generation of much-needed treatments.

1. Freedman, L. P., Cockburn, I. M. & Simcoe, T. S. The economics of reproducibility in preclinical research. *PLoS Biol.* **13**, e1002165 (2015).
2. Allarakhia, M. The successes and challenges of open-source biopharmaceutical innovation. *Expert Opin. Drug Discov.* **9**, 459–465 (2014).
3. Gamo, F. J. *et al.* Thousands of chemical starting points for antimalarial lead identification. *Nature* **465**, 305–310 (2010).
4. Van Voorhis, W. C. *et al.* Open source drug discovery with the Malaria Box compound collection for neglected diseases and beyond. *PLoS Pathogens* http://dx.doi.org/10.1371/journal.ppat.1005763 (2016).
5. Maxmen, A. Cancer research: open ambition. *Nature* **488**, 148–150 (2012).

Acknowledgements

The authors wish to thank R. Don (Drugs for Neglected Diseases *initiative* (DNDi)), A. Palermo (Royal Society for Chemistry (RSC)), M. Todd (University of Sydney), the MMV Science Team and the Expert Scientific Advisory Committee (ESAC) for helpful discussions.

Competing interests statement

The authors declare no competing interests.

FURTHER INFORMATION

Lilly Phenotypic Drug Discovery Initiative: https://openinnovation.lilly.com/dd/
The Structural Genomics Consortium: http://www.thesgc.org/
GlaxoSmithKline clinical trial data platform: http://www.gsk-clinicalstudyregister.com/
dbSNP: http://www.ncbi.nlm.nih.gov/projects/SNP/
Open Source Malaria project: https://opensource.com/life/14/6/international-team-open-sources-search-malaria-cure

SUPPLEMENTARY INFORMATION

See online article: S1 (box)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF