POLICY FORUM

DRUG DISCOVERY

Disruptive Strategies for Removing Drug Discovery Bottlenecks

Sean Ekins[1 *], Chris L. Waller[2, 3] Mary P. Bradley[4] and Antony J. Williams[5]

[1] Collaborations in Chemistry, 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526, U.S.A.

[2] Pfizer Inc., Eastern Point Road, Groton, CT 06340.

[3] School of Pharmacy, University of North Carolina, Chapel Hill, NC 27514

[4] CollaborationFinder, Harvard Square, One Mifflin Place, Suite 400, Cambridge, MA 02138

[5] Royal Society of Chemistry, 904 Tamaras Circle, Wake Forest, NC 27587.

* To whom correspondence should be addressed.
Sean Ekins, E-mail address: ekinssean@yahoo.com, Phone: +1 215 687 1320

HTS, bottlenecks and databases

Currently large pharmaceutical companies are undergoing selective disintegration, while contract research organizations (CROs) and academia are growing in influence, publicly-funded drug development programs are expanding, precompetitive efforts are increasing, along with a re-emergence of venture-backed biotechnology firms (*1*). These developments have created a dynamic ecosystem with pharma as smaller nodes in a complex network, in which collaborations have become an important business model. However, we are seeing a shift in focus away from early drug discovery, counter to what some have suggested is necessary for the industry to survive post disintegration (*2*).

This is exemplified by the shift of high throughput screening (HTS) for drug discovery from a small number of major pharmaceutical companies to a larger number of academic and institutional laboratories in the US. This seems counter intuitive as some drugs and a large percentage of leads are discovered using HTS (*3*), yet there are also examples in which HTS fails, in particular antibacterial research and other areas (*4, 5*). Learning from the pharma experience with HTS is instructive. A recent study identified 78 academic screening centers in the US focused on high risk drug targets, while there were major gaps for efficacy testing, drug metabolism, PK studies and the challenge of translation to the clinic (*6*), commonly termed the "valley of death". These gaps incidentally are all skills that pharma is removing and outsourcing. This leaves only CROs and clinically affiliated institutes able to overcome this bottleneck. Another issue identified by researchers from Bayer indicates that literature data on potential drug targets is not reproducible (*7*). Translating more compounds to the clinic from HTS screening

centers, may indicate that many would likely fail without controlling for bias in pre-clinical proof of concept studies and target-based discovery to improve clinical success (*8*). Taking HTS out of pharmaceutical companies has not achieved innovative breakthroughs. And yet, the US government through different agencies is investing heavily in large HTS initiatives such as ToxCast (*9, 10*), Tox21 (*11*), Molecular Libraries Probe Production Centers Network (MLPCN) (*12*), National Center for Translational Therapeutics (NCATS) (*13*), the LINCS project (*14*) alongside the institutional screening centers, with little apparent coordination or consideration of the outputs. We have concerns regarding simply using the HTS assays (and data) that were optimized to minimize "false negatives" for risk assessment purposes.

A crowdsourcing evaluation of MLPCN probes suggested to us that academic screening may result in a large number of dubious leads when in a drug screening mode (*12*). All of the screening efforts are generating very large quantities of data and there would be an expectation that it is freely accessible, requiring databases that can handle structures and multiple bioactivity endpoints. Recent NIH funded efforts with the NPC browser (*15*) suggest this is not straightforward (*16*) and poor data quality will severely impact the cost effective but increasingly informatics dependent tools being used for repurposing efforts (*17*). In our opinion there needs to be independent assessment and curation of the data produced across the board before embarking on more investments.

It is also unclear how such data is policed to make sure it goes out in a timely manner for maximal exposure. We are not aware of any funding agency mandating data to be published along with quality guidelines, although we have suggested granting

bodies should have minimal quality standards for published data (*16*). An extension to this would be that all data generated from publicly funded research should be openly available, within a year of generation, in high quality internet databases.

We think part of the current trend in terms of proliferation of HTS screening initiatives is due to lack of coordination of government agencies, creating duplication and overlap, as exemplified by numerous chemical databases in North America containing approved drugs (Table 1). The government agencies would argue that redundancy in funding mitigates risk, however if there is no sharing of data or experience ex post facto, then the risk of duplicative failure and unproductive expenditure increases. From what we see there is too little collaboration around databases, curation, data quality (*16*) or even openness across the board.

There has been much discussion in the context of NCATS, about the urgent need to revamp how drugs are developed, brought to market faster and what incentives can be provided to generate treatments for neglected and rare diseases (*13*). We question however whether any government or academic institute as they currently stand can adequately pursue such goals when an entire industry is struggling with the same challenges. Many of the techniques proposed (*13*), just like HTS, will not dramatically impact the process alone because this has not occurred in pharma.

Public private partnerships and translational informatics

This begs the question of how we can remove the bottlenecks impeding progress now. Academic groups could avoid the "valley of death", by working more closely with CROs and virtual pharmas to do more preclinical and development studies, who in turn

will work with pharmas to purchase the most promising compounds. To do this there needs to be an awareness of what research is going on in the screening centers, and they in turn should be aware of groups that can take their hits.

There is general agreement that the key to breakthrough success is collaboration (*18*). There is also consensus that social networking can provide an effective platform for increasing collaboration in biomedical research (*19*), yet to date this has failed to take hold. The reason is fundamental: monetization of intellectual property (IP). There is no incentive for research organizations to disclose their current research in an open social networking forum where competitors have equal access. This is even true in academic research where investigators compete for funding. The key to success of this model of collaboration is the security of IP and the ability to selectively disclose IP to a valid potential partner in a secure way that results in a mutually equitable outcome for all parties (*20*). Research collaborations are currently most advanced in the areas of neglected diseases, where funding comes primarily from public sources, data is more open, and potential profits are low or nil. The same situation is true for rare diseases (*21, 22*) and one would expect the creation of networks and ways to do more with less funding using collaborative software (*18, 23*) will be essential. In both neglected and rare diseases the partners are more likely to share IP because the monetary value of the IP ceases to be a barrier.

Given that research organizations appear to be open to embracing a new paradigm of collaboration, how is one scientist to know what other work is currently ongoing in a specific therapeutic or disease area when this is private? The key areas for success in biomedical research collaborations are for organizations to be able to "identify best-in-

class capability, evaluate opportunities presented by programs and understand the associated risks" (*24*). To date, there is a lack of support mechanisms to identify and foster collaborations, resulting in a time consuming hit-or-miss process that relies on networking, internet searching, and attending scientific conferences. New services (*25*) that provide a low cost, efficient means of finding targeted scientific connections for research and funding, while protecting intellectual property will be key to connect everyone with a role in drug discovery and development. As virtual companies will have nowhere near the resources or experience of a big pharma, much more work will need to be performed *in silico* (*17*) as well as in a collaborative manner (*18*) to ensure likely success. Another way to look at this is that a new virtual team paradigm has the potential to innovate through disruption.

There have been several collaborative public private partnerships (PPP) in Europe to share drug safety data (*26*), ontologies and models (*27*) and knowledge management of pharmacological data (*28*), all of which foster collaboration, as well as data sharing from industry and academia. In comparison the USA has nothing comparable currently ongoing in its research portfolio. Such shared knowledge could help virtual pharmas, academics and institutes alongside pre-competitive initiatives like those in informatics (*29-32*) to focus on the best ideas. The key challenge here is to ensure the delivery of tools or services to solve common problems to all parties involved and that there is coordination, progress and no overlap with the PPP initiatives described above. All of these efforts lower the cost of research and remove duplication of efforts. A direct example is the structure representation standards documented for the FDA's substance registration system (*33*) whose recommendations have largely been adopted by ChEMBL

(*34*) and will be implemented into ChemSpider (*35*) to support the OpenPHACTS project (*28*) for pharmaceutical companies that are participating in this initiative.

As big pharma relies more on the CROs and academics, they will focus on translational informatics (integrated software solutions to manage the logistics, data integration and collaboration) and other efforts such as Pfizer's ePlacebo. This uses placebo dosing data from previously executed clinical trials to augment or potentially supplant the need for placebo control groups in clinical trials. A cross-pharma data sharing consortium would dramatically impact the cost associated with clinical trial recruitment and execution of placebo dosing. In an effort to stimulate data sharing of this type the FDA has announced an overhaul of its IT infrastructure (*36*). A first step is the effort to make the historical clinical data in the FDA's vaults public to be followed by a vast amount of de-identified post market surveillance data. By doing this, the FDA hopes that the open access movement will stimulate the creation of public private partnerships aimed at sharing data relevant to other drug development stages. Could they go further and mandate all de-identified clinical data be made public as part of the cost of doing business? Although some groups are pro (*37*) and others con (*38*) this approach could be universally useful for health research. We should be aware of potential barriers to data sharing and collaboration. Data and information silos exist at all levels of organizations. Allowing for data/information integration across silos is not a technological problem, regardless of issues of taxonomies and ontologies, but those will be much easier to surmount than the cultural, societal, and behavioral barriers to effective collaboration (*18*). Such non-technical issues generally inhibit translational data analysis on a broad

scale. With all the distributed research efforts we do not want to see creation of new data silos.

Mining by swarm and finding the best collaborators

While the FDA and the NHS (*39*) have discussed the 'big data' or 'analytics' future involving analysis of patient data. We are also moving into the era of drug safety analysis, drug repurposing and marketing by sentiment analysis using social media stream mining tools (*40-42*). Swarm intelligence is a new subfield of bio-inspired artificial intelligence offering solutions to complex problems like pooled health-related data from different organizations as well as real time data from social networks (*43*). Emerging and likely disruptive technologies that listen to the crowd passively do not appear to be on the agenda (*36*).

In summary, if we are to remove bottlenecks we need to provide more confidence that lead compounds will have efficacy *in vivo* and be safe. Some of these aspects could be considered using predictive models already assembled and exclusive to the pharmaceutical companies. Sharing precompetitive data and models (*44*, *45*), whether through a PPP or collaborations, could provide more confidence in the quality of the leads produced such that they will attract investment. At the same time the fringes of industry and academia may harbor the real innovators that should be funded to transform R&D. Both governments and pharmas could use software like Collaboration Finder (*25*) to find the best researchers to fund and collaborators to work with on strategic priorities. This would enable NIH to fund continuous innovation, rather than rebuilding academia in

the shape of big pharma. Disruption of the pharmaceutical industry may begin by a fundamental rethink of how to reward collaborative researchers in any organization.

References

1. G. M. Robertson, L. M. Mayr, *Future Med Chem* **3**, 1995 (Dec, 2011).
2. S. M. Paul *et al.*, *Nat Rev Drug Discov* **9**, 203 (Mar, 2010).
3. R. Macarron *et al.*, *Nat Rev Drug Discov* **10**, 188 (Mar, 2011).
4. D. A. Payne, M. N. Gwynn, D. J. Holmes, D. L. Pompliano, *Nat Rev Drug Disc* **6**, 29 (2007).
5. G. M. Rishton, *Med Chem* **1**, 519 (Sep, 2005).
6. S. Frye, M. Crosby, T. Edwards, R. Juliano, *Nat Rev Drug Discov* **10**, 409 (Jun, 2011).
7. F. Prinz, T. Schlange, K. Asadullah, *Nat Rev Drug Discov* **10**, 712 (Sep, 2011).
8. M. D. Lindner, *Pharmacol Ther* **115**, 148 (Jul, 2007).
9. D. J. Dix *et al.*, *Toxicol Sci* **95**, 5 (Jan, 2007).
10. R. Judson *et al.*, *Environ Health Perspect* **117**, 685 (May, 2009).
11. Tox21. http://www.rsc.org/chemistryworld/News/2011/December/chemical-toxicity-human-health-testing.asp.
12. T. I. Oprea *et al.*, *Nat Chem Biol* **5**, 441 (Jul, 2009).
13. F. S. Collins, *Sci Transl Med* **3**, 90cm17 (Jul 6, 2011).
14. Lincsproject. www.lincsproject.org.
15. R. Huang *et al.*, *Sci Transl Med* **3**, 80ps16 (Apr 27, 2011).
16. A. J. Williams, S. Ekins, *Drug Disc Today* **16**, 747 (2011).
17. S. Ekins, A. J. Williams, M. D. Krasowski, J. S. Freundlich, *Drug Disc Today* **16**, 298 (2011).
18. S. Ekins, M. A. Z. Hupcey, A. J. Williams, *Collaborative computational technologies for biomedical research* (Wiley, Hoboken, NJ, 2011), pp.
19. V. Gewin, *Nature* **468**, 993 (2010).
20. B. A. Bunin, S. Ekins, *Drug Disc Today* **16**, 643 (2011).
21. R. C. Griggs *et al.*, *Mol Genet Metab* **96**, 20 (Jan, 2009).
22. G. J. Brewer, *Transl Res* **154**, 314 (Dec, 2009).
23. F. Bost, R. T. Jacobs, P. Kowalczyk, *Curr Opin Drug Discov Devel* **13**, 286 (May, 2010).
24. A. Jones, L. Clifford, *Nat Rev Drug Discov* **4**, 807 (Oct, 2005).
25. CollaborationFinder. www.collaborationfinder.com.
26. eTox. http://www.e-tox.net/index.html.
27. B. Hardy *et al.*, *J Cheminform* **2**, 7 (2010).
28. OpenPHACTS. http://www.openphacts.org/.
29. Pistoia Alliance. http://pistoiaalliance.org/.
30. BioIT Alliance. http://bioitalliance.org/.
31. PRISME. http://prismeforum.org/.

32. W3C. http://www.w3.org/.
33. SRS. http://www.fda.gov/ForIndustry/DataStandards/SubstanceRegistrationSystem-UniqueIngredientIdentifierUNII/default.htm.
34. ChEMBL. http://www.ebi.ac.uk/chembldb/index.php.
35. A. J. Williams, *Drug Discov Today* **13**, 502 (Jun, 2008).
36. Anon, "Advancing regulatory science at FDA" (U.S. Food and Drug Administration, 2011).
37. Monash research guidelines. http://www.researchdata.monash.edu/guidelines/deposit.html.
38. A. Silversides, *BMJ* **342**, d2570 (2011).
39. NHS. http://www.rsc.org/chemistryworld/News/2011/December/08121101.asp.
40. Teranode. http://www.teranode.com.
41. Ceiba. http://www.ceibasolutions.com/solutions/#social.
42. Swarmology. http://www.swarmology.com/), .
43. D. Martens, B. Baesens, T. Fawcett, *Mach Learn* **82**, 1 (2011).
44. R. R. Gupta *et al.*, *Drug Metab Dispos* **38**, 2083 (2010).
45. O. Spjuth, E. L. Willighagen, R. Guha, M. Eklund, J. E. Wikberg, *J Cheminform* **2**, 5 (2010).

46. Conflicts of Interest**:** SE Consults for Collaborative Drug Discovery and is on the board of directors for the Pistoia Alliance, AJW is employed by the Royal Society of Chemistry which owns ChemSpider and associated technologies and is involved with the OpenPHACTS project, CW is an employee of Pfizer, MB is an employee of CollaborationFinder.

Table 1. North American small molecule databases containing FDA approved drugs

Database name	Funding	Content and details	URL
PubChem	NIH	>30M molecules includes FDA approved drugs	http://pubchem.ncbi.nlm.nih.gov/
NPC Browser	NIH	~10,000 compounds includes FDA approved drugs	http://tripod.nih.gov/npc/
ToxCast	EPA	>1000 compounds includes some drugs and drug like molecules	http://epa.gov/ncct/toxcast/
DailyMed	FDA	>31,942 labels – many labels for the same drug	http://dailymed.nlm.nih.gov/dailymed/about.cfm
ChemIDplus	NIH	> 295,000 structures including many FDA small molecule approved drugs	http://chem.sis.nlm.nih.gov/chemidplus/
DrugBank	Canadian	6707 drug entries including 1436 FDA-approved small molecule drugs (this may be underestimated).	http://www.drugbank.ca/