Chemical space, high throughput screening and the world of blockbuster drugs

We are constantly told that if a high throughput screen does not identify hit(s) then blame it on the compounds in your library. The community has accepted this notion and unleashed a chemical space exploration through the use of novel or pre-existing synthetic chemistries to supposedly generate better ones. Its efforts, unfortunately, did not result in a giant step for mankind to combat disease, rather this article would argue that we are still at the same start-up line as we were 30 years ago and should perhaps listen to the call of nature to revive natural product research as it has amassed the most diverse libraries over the past millions of years.

The most memorable and thought-provoking editorial was by Roger Lahana in 1999 on combinatorial chemistry (CC) and high throughput screening (HTS)[1]; and appropriately entitled 'How many leads from HTS?' It was during the phase of heavy investments by large Pharma and biotech companies alike in CC, automation, novel assay technologies and HTS as the only way forward to discover drugs and reduce overall cost and time. This revolutionary phase became obsessed with numbers and turned HTS into a numbers generator; where companies were viewed as successful if they had assembled multi-million compound libraries and generated several million data points per year. Not surprising at all that when some of the attendees of the CHI High-throughput technologies meeting, back in May of 1999 in Washington DC (USA), were asked the question of how many leads have we got from CC and HTS so far? Their overwhelming response was none![1]. A few months later, a response to Lahana's editorial was published[2]. Its author claimed that the purpose of screening is not to identify the molecule, rather to identify good leads from which clinical candidates could be synthesised using rational/classical chemistry approaches; they further added that if screening fails to identify good leads, then either the assays were no good or the quality of the library was poor[2]. Surprisingly, Fox and colleagues in 2001 reported what was thought to be the earliest signs of success of HTS as evident with up to 46 drug candidates apparently obtained from HTS over an eight-year period (1990 to 1997), as recorded through their worldwide survey

By Dr Hakim Djaballah

Drug Name	Drug Company	Drug Use	Source	FDA Approval	US Patent Expiry
Lipitor	Pfizer	Hypocholesterolemic	NP mimic	1996	2011
Abilify	Bristol-Myers Squibb, Otsuka	Neuroleptic	NP mimic	2002	2014
Plavix	Bristol-Myers Squibb/Sanofi	Anti-thrombotic	NP mimic	1997	2012
Nexium	AstraZeneca	Antiulcer	Synthetic	2000	2014
Seroquel	AstraZeneca	Neuroleptic	NP mimic	1997	2011
Singulair	Merck	Anti-asthmatic	NP mimic	1998	2012
Advair	GlaxoSmithKline	Anti-asthmatic	Synthetic	2000	2012
Crestor	AstraZeneca, Shionogi	Hypocholesterolemic	NP mimic	1996	2016
Cymbalta	Eli Lilly	Anti-depressant	NP mimic	2004	2013
Humira	Abbott Laboratories	Anti-inflammatory	Biologic	1998	2016

Table 1: The top 10 selling prescription drugs in the United States in 2011. Seven drugs have their origins from natural product sources

interviewing 52 individual HTS lab directors[3]. A year later, Fox and colleagues surprised the drug discovery world by reporting that up to 63 drug candidates, obtained from HTS efforts from the earlier 1990s to 2000, were indeed in Human clinical trials[4]. Does it mean that we can declare success of CC and HTS in providing us with the much needed drugs? Clearly, it is a sharp contrast between the Lahana editorial with no leads in sight and the Fox reports with up to 63 candidates in Human clinical trials. Almost 11 years later, one would expect to see at least a favourable outcome of some of these candidates and then we can collectively declare success. With limited information at hand, it is an impossible task to track each of the anonymous 63 clinical candidates through their trials and approval processes.

Scannell and colleagues, in a recent review, reported for the first time a thoughtful analysis of R&D efficiency over a 60-year period[5] and may provide us with a partial answer as to the success or failure of CC and HTS. They report that with an 800-fold increase in CC output during the 1980s and 1990s resulting in large chemical libraries, innovations in HTS including a 10-fold cost reduction in testing these libraries, and vast improvements to the R&D process including outsourcing, the number of new drug approvals has halved roughly every nine years since the 1950s and up to the 2010 approvals – an 80-fold fall in productivity[5]. Their analysis seems to point towards failures rather than successes of CC and HTS. However, a more optimistic view was recently presented by Berggren and colleagues at McKinsey & Company, addressing the crisis in the R&D productivity with a five-year forecast in drug innovation[6]. The probability of success (POS) is much higher in their forecast with an average annual number of approvals/launches to be above 30 during a five-year period (2011-15), and with a rapid decline to

below 10 in 2016. Their positive outlook seems to be driven by the unique number of novel molecules in development and reaching 7,709 compounds in 2011. They further estimate the total number of novel compounds between 2006 and 2011 to be more than 13,000 potential drug products, based on their pipeline analysis of the Pharmaprojects pipeline database[6]. Berggren assessment is suggestive of a success if these 13,000 compounds were indeed the fruits of CC and HTS; but their outlook of potential disaster by 2016 somehow contradicts the success and questions whether the industry can indeed generate additional 13,000 novel compounds beyond 2015?

Chemical library: good or bad?
Like many of my colleagues who are in the business of running chemical screening operations in Academia, we are often being told by our customers that the libraries we use in HTS are 'no good' since their friends were extremely lucky in finding hit(s) using someone else's library; and sometimes even criticised by colleagues on grant review panels that the libraries to be used for screening are also 'no good'. I often wondered whether this is an academic naivety problem to do with the novelty of chemical screening or there was an underlining message in there. When I joined Memorial Sloan-Kettering Cancer Center in 2003 to set up the HTS Lab, I was instructed by a particular member of faculty to purchase the ChemBridge DIVERSet E library containing 50,000 compounds and as he put it "many people have screened it and published on it" as their selection criteria for a chemical library; further adding "if no one has published on it then it will be hard to publish on it in the future". I found their instructions and insistence remarkable considering the fact that they are chemists by training. Instead, I took the simplest approach to building what I

believed would be a more comprehensive chemical library to fit all of the biologies being studied at the centre, and with four simple rules: 1) The build has to be over time; 2) The overall size should not exceed 500,000 molecules; 3) Loose exclusion criteria/rules for molecules; and 4) Enrichment in non-Lipinski compliant compounds, ie natural products, is a must. Ten years later, my chemical library is beautiful, approaching 400,000 chemicals and with a good track record in identifying actives in screening campaigns[7-18].

Surprisingly though, when Fox and colleagues sought opinions on the success of HTS from interviewing HTS lab heads back in 2001, only 12.5% felt chemical libraries were the problem and better ones were needed to be used for screening[4], hardly a majority that one would have expected. However, Scannell and colleagues would argue that the chemical library content is a huge problem considering the reality of the chemical space coverage of today's libraries being infinitesimally small[5]. They further add that Pharma and biotech merger activities in the past few years have revealed substantial overlaps between their respective chemical libraries[5]; not surprising considering that similar brains think alike, only a handful of commercial vendors to go to, and similar chemistries were applied. It also means that we are all screening similar compounds whether we are aware of it or not. If that is truly reflective of the dire state of the composition of chemical libraries and content, then how do we explain the source/origin of these 13,000 novel and unique compounds alluded to by Berggren[6]?

How do you begin to judge whether a chemical library is good or bad? In my opinion, you cannot really judge a library; you can only describe its content in terms of unique chemotypes, how big a cluster they represent, presence of nuisance chemistries or reactive groups, and if you are really desperate for a validation statement then by all means apply the Lipinski's rules of five; segregating the library into categories of 'compliance'. As an example, a recent publication by Baell addressing the coverage of lead-likes in commercial libraries using their in-house developed PAINS filters, reported that only 6,000 lead-like molecules in a vendor library of 400,000 chemicals[19]; that is a very high attrition rate of only 1 in 67 molecules is lead-like and perhaps worth screening. I would take a completely different approach and claim that each molecule in your library presents an opportunity, I do not know what it is yet and it does not really matter; the discovery is dependent on the biological question being asked, and the assay used for the screen is

highly critical. Nowadays, there are fewer and fewer reports addressing or questioning assay suitability/validity for HTS. This is unfortunate considering the overwhelming investment in screening from both the public and private sectors. We have also been led to believe that a single measure, known as the Z-factor[20], is all you need to determine whether the assay is good for HTS or not. With a Z-factor value of >0.5, your assay must be robust and ready for screening. This magic factor has industrialised the screening world and unfortunately, is the sole culprit of misguidance in assay development for HTS. The Z-factor is defined as a measure of 'statistical effect size', meaning that to be statistically relevant you need a larger data set to assess the separation of your assay signal to its noise or background. What the Z-factor does not tell us is: 1) The relevance of your assay signal to the question you are trying to study since we are fully vested in sensitivities, miniaturisations and ultra-HTS; 2) Assay dynamics and sensitivity to specific and non-specific modulators; and 3) Heterogeneity of the biology being studied. In essence, it is an abused measure and sometimes irrelevant such is the case for high content cell-based assays[18]; but we must have it to get approvals to carry out our screens, to get our manuscripts published as reviewers will more than often ask for it, and to get our grant proposals funded as it is part of the funding requirement.

Many of us do forget that screening these libraries represent the largest live casino you will ever play in, whether you are a blackjack, poker, roulette or slot-machine player, the odds are always in favour of the house; unless of course you cheat or get lucky. So, today's chemical libraries are indeed one size fits all and as far as I am concerned nothing is wrong with them; they are also the only available and accessible tools to use and should be used across a multitude of biologies and screens with the hope of getting lucky one day. Baell's comment on the sheer level of artifacts being generated by screening these libraries is duly noted[19], but biology without noise is not worth studying. I would also argue that these reported pan-active hit(s) from screening these libraries are the best metrics of suitability/validity of your assay in the first place.

Chemical space exploration

Curiosity is one of the many traits of human beings leading us to always wanting to explore and better understand our world. In very ancient times, our ancestors used to observe objects in the skies leading them to make predictions of their motions only

References

1 Lahana, R. How many leads from HTS? Drug Discov Today 4, 447-448 (1999).

2 Ramesha, CS. How many leads from HTS? – Comment. Drug Discov Today 5, 43-44 (2000).

3 Fox, S et al. High throughput screening: early successes indicate a promising future. J Biomol Screen 6, 137-140 (2001).

4 Fox, S et al. High throughput screening 2002: moving toward increased success rates. J Biomol Screen 7, 313-316 (2002).

5 Scannell, JW et al. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov 11, 191-200 (2012).

6 Berggren, R et al. Nat Rev Drug Discov 11, 435-436 (2012).

7 Antczak, C et al. High-throughput Identification of Inhibitors of the Cancer Target Human Mitochondrial Peptide Deformylase. J Biomol Screen 12, 521-535 (2007).

8 Deng, L et al. Identification of novel antipoxviral agents: Mitoxanthrone inhibits Vaccinia virus replication by blocking virion assembly. J Virol 81, 13392-402 (2007).

9 Desbordes, SC et al. High-Throughput Screening Assay for The Identification of Compounds Regulating Human Embryonic Stem Cells Self-Renewal and Differentiation. Cell Stem Cell 2, 602-12 (2008).

10 Antczak, C et al. Revisiting Old Drugs as Novel Agents for Retinoblastoma: In vitro and In vivo Antitumor Activity of Cardenolides. Invest Ophthalm Visual Science 50, 3065-3073 (2009).

11 Somwar, R et al. Identification and preliminary characterization of novel small molecules that inhibit growth of human lung adenocarcinoma cells. J Biomol Screen 14, 1176-1184 (2009).

12 Shelton, CC et al. Modulation of γ-secretase specificity using small molecule allosteric inhibitors. PNAS (USA) 106, 20228-33 (2009).

Drug Discovery

- Average cost is > $1 billion & takes up to 15 years
- Prices are under constant pressure
- Regulatory is becoming very complex
- Patent landscapes are very sensitive
- Generics have attained ~ 50% of the market
- R&D investments are becoming tough to make
  - ❖ Only 3 in 10 drugs recover R&D costs
  - ❖ Only 2 in 10 drugs are profitable to the owner
  - ❖ Only 1 in 100 drugs are blockbusters

Figure 1: Drug discovery under siege. Several constraints against investments in R&D.

if they can see them by their naked eyes. Through perseverance and technology advancements, the field has enabled the human race to go beyond the skies, the solar system and into deep space exploration using both manned and unmanned missions. The purpose of which is to unravel the mysteries of our universe, and to better understand our place in the cosmos. After several thousand years of research and a gigantic budget, have we explored or conquered space?

By comparison, have we successfully conquered the chemical space? The answer is not yet because we have only begun to explore its complexities. As disappointing as this may be, our ability to synthesize complex chemical molecules is still rather primitive as compared to what nature can and has been doing over billions of years. As an example, bacteria can make a complex molecule in a few hours versus several years of hard work of a super creative synthetic chemist with an army of students and post-doctoral researchers resulting in one high profile publication with a gigantic authorship line-up 10 to 20 years later, and with a synthetic route of more than 60 steps; hardly amenable to conventional process chemistry for scale up. So, in essence, we can replicate or mimic what nature has already made given resources and time, but is it enough for conquering space and making novel drugs? Newman and Cragg would argue that we are at least making some progress, and suggest that natural product research should be expanded significantly, as assessed through their 30 years' analysis of drug approvals: 64 New Chemical

Entities (NCEs) were natural products, 299 NCEs were derived from a natural product with a semi-synthetic modification, 268 NCEs were natural product mimics and 55 NCEs were made by total synthesis but the pharmacophore may have a natural product origin[21].

According to Scannell and colleagues, the world's chemical libraries combined represent a very tiny miniscule fraction of the vast chemical space[5], perhaps the best comparison would be the size of a rounded grain of salt versus the size of Earth; they further argue that these libraries contain redundant regions of the chemical space. Surprisingly, the CAS database counter registry contains ~68 million commercially available chemicals[22]. Based on the various suggestions as to the size of the chemical space[23-24], I would propose that empirically it may contain 10^{180} possible organic and inorganic chemical structures. I would also predict that up to 10% have evolved over billions of years to interact with living matter from bacteria, fungus, worms, scorpions, frogs, dinosaurs, plants to humans. This would leave us with up to 10^{18} possible chemical structures to explore and synthesise, clearly an unattainable task. But if I would further empirically predict that there are only 10^9 known chemicals on Earth, it would then mean that we have covered only 7%, and leaving 93% or 930 million molecules up for chemical exploration. How realistic is it to engage in making these compounds over the next 30 years? That is 31 million compound per year or 2.59 million compounds per month or 86,111 compounds per day. Even with access to the most sophisticated high output chemical synthesis strategies and robotic enabled facilities in the world, simply put this task would also be unattainable. In a recent report, Reymond and Awale predict that by using their exploration strategies including improved compound enumeration, classification and virtual screening schemes, and implementation of chemical synthesis resources, we would have an opportunity to better explore the chemical space[25]. Thirty years of CC and chemical innovations have left us stagnant at the same starting line for the race to produce better drugs, and with only 387 NCEs discovered by random screening or modifications of existing molecules[21].

Drug discovery under siege

1,073 drugs are considered unique NCEs approved over a 30-year period[21,26]. If these numbers truly hold, then the combined mighty of Pharma and biotech companies have only produced an average

of ~36 NCEs per year over a period of 30 years of R&D. At face value, this analysis does not make sense otherwise these companies would be out of business and long gone. Actually, it does make perfect sense for those companies playing the dangerous game of chasing blockbuster drugs knowing how much they will be worth in the end. A blockbuster drug is defined as the one which achieves revenues of more than $1 billion at global level for its owner per year[27]; it is estimated that 125 drugs have met the target sale. The top 10 best selling drugs in the Unites States alone generated more than $70 billion in sales in 2011[28-29]. Seven of these blockbusters (**Table 1**) have an origin linking them to products naturally produced by nature including the number one blockbuster of all time, Lipitor, generating more than $125 billion in sales for Pfizer[29].

The search for blockbuster drugs is more lucrative and important than ever; Pfizer's success with Lipitor makes the chase even more addictive and risky, though Pharma chiefs claim that recent restructuring and portfolio derisking activities are gaining traction towards rebuilding a strong and potentially profitable pharmaceutical industry[29]. I would argue that it is hardly the case, considering several indicators such as prescription costs are under pressure, patents have sensitive landscapes, generic sales has reached up to 50% of the market and only two in 10 drugs are profitable, when combined report that the drug discovery business is indeed under siege (**Figure 1**). Furthermore, the patent cliff of the best selling drugs is upon us and the recent approvals may result in unanticipated best seller replacements –

wishful thinking by the Pharma chiefs. In 2013, it is a reality that making any investments in R&D is even more challenging than gambling, leaving the dreaded question of where will the next blockbuster drug come from?

The Pharma and biotech industries have been exploring the chemical space for more than 30 years, through the use of sophisticated synthetic methodologies and CC and resulting multi-million compound libraries screened against a diverse range of targets and biologies. The outcome of these huge investments is rather disappointing with only 36% of the approved NCEs attributed to this gigantic endeavour. It is not that surprising, considering that manmade chemistries will always rely on simple coupling reactions with nitrogen enrichment in the final molecules. **Figure 2** summarises the unreasonableness of chemical exploration by synthetic chemistries leading to success. The numbers clearly bias those molecules with natural source origins and contributing to 64% of the approved NCEs. Newman and Cragg alluded to various academic chemistry groups modifying active natural product skeletons as leads for novel drugs[21], but failed to caution against the continuous use of combinatorial chemistry approaches with a new coat of diversity paints. If these approaches were to be successful, then we would have observed their impact by now. This leaves us with only one choice and that is to return to Mother Nature as a golden source of novel agents and drugs. Billions of years of evolution can only make better chemical molecules than mankind will ever come close to making synthetically.

Continued from page 75

13 Shum, D et al. High-content assay to identify inhibitors of dengue virus infection. Assay Drug Dev Technol 8, 553-570 (2010).
14 Antczak, C et al. Identification of benzofuran-4,5-diones as novel and selective non-hydroxamic acid, non-peptidomimetic based inhibitors of human peptide deformylase. Bioorg Med Chem Lett 21, 4528-4532 (2011).
15 Somwar, R et al. Superoxide dismutase 1 is a probable target for a small molecule identified in a screen for inhibitors of the growth of lung adenocarcinoma cell lines. PNAS (USA) 108, 16375-16380 (2011).
16 Feldman, T et al. Class of Allosteric Caspase Inhibitors Identified by High-Throughput Screening. Mol Cell 47, 585-95 (2012).
17 Lee, G et al. IKBKAP Expression Rescue in Neural Crest of Familial Dysautonomia-iPSC Cells by Novel RT-PCR Based High Throughput Screening. Nature Biotech 30, 1244-1248 (2012).
18 Shum, D et al. An Image-Based Biosensor Assay Strategy to Screen for Modulators of the microRNA 21 Biogenesis Pathway. Comb Chem High Through Screen 15, 529-541 (2012).
19 Baell, JB. Broad Coverage of Commercially Available Lead-like Screening Space with Fewer than 350,000 Compounds. J Chem Inf Model 53, 39-55 (2013).
20 Zhang, JH et al. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4, 67-73 (1999).
21 Newman, DJ, Cragg, GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 75, 311-35 (2012).
22 CAS Database Counter accessed on 14 February 2013; http://www.cas.org/content/counter.

- Chemical Space Estimate 10^{180}
- Likely Candidates 10^{18}
- Known Chemicals 1,000,000,000
- Commercial Chemicals 68,000,000
- Filtered for Lead-Like* 1,014,926
- Approved Drugs (NCEs) 1,073
  - HTS or modification of an existing molecule 387
  - Synthetic with Natural Product pharmacophore 55
  - Natural Product with semi-synthetic modification 299
  - Natural Product mimic 268
  - Natural Product 64
- Blockbuster Drugs to Date 124

Figure 2: Chemical space exploration an unattainable task. * *PAINS filters as described by Baell[19]*

Drug Discovery

23 Brown, D. Future pathways for combinatorial chemistry. Mol Divers 2, 217-222 (1996).
24 Drew, KLM et al. Size estimation of chemical space: how big is it? J Pharm Pharm 64, 490-495 (2012).
25 Reymond, JL, Awale, M. Exploring chemical space for drug discovery using the chemical universe database. ACS Chem Neurosci 3, 649-57 (2012).
26 Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations. Accessed on 14 February 2013; http://www.accessdata.fda.gov/scripts/cder/ob/default.cfm.
27 In European Commission, Pharmaceutical Sector Inquiry, Preliminary Report (DG Competition Staff Working Paper), November 28, 2008.
28 Nisen, M. The 10 Best Selling Prescription Drugs in the United States. In the Business Insider, June 28, 2012; http://www.businessinsider.com/10-best-selling-blockbuster-drugs-2012-6?op=1#ixzz2KzPtaOWz.
29 Mullin, R. Before the Storm. Chem Eng News 89, 12-18 (2011).
30 Cragg, GM et al. The impact of the United Nations Convention on Biological Diversity on natural products research. Nat Prod Rep 29, 1407-1423 (2012).

A call to revive natural product research

Will the Pharma industry answer the call? Most likely not as reflected in their recent activities of consolidations, layoffs and site closures. I would also speculate that they would eventually eliminate the 'R' from their R&D efforts and shift it towards the public sector; in turn, government agencies, academics and non-for-profit research organisations will answer the call and hopefully revive natural product research through a global initiative to benefit mankind. Combined, they would provide Pharma with unprecedented access to the largest research enterprise in the world at little or no cost to them. Academic and non-for-profit screening operations throughout the world are better positioned than ever to screen natural products as extracts or purified molecules. The global nature of this initiative would require special collaborative agreements protecting the rights of all parties involved, especially those biodiversity rich source countries providing access to their resources and in return, be rewarded in a fair and equitable manner. Such efforts have been ongoing for many years now with the hope that we are indeed close to the ratification of the Nagoya Protocol, which requires signatures from 50 countries[30].

Acknowledgements

Dr Hakim Djaballah, molecular pharmacologist and technologist, has been the Director of the HTS Core Facility at Memorial Sloan-Kettering Cancer Center since its establishment in 2003. In 1992, he received his PhD in biochemistry from the University of Leicester, England. He was the recipient of the 2007 Robots and Vision User Recognition Award.