Towards Academic Drug Development Guidelines

Version 14th July 2019, Christopher Southan

Executive summary

This report covers academic small-molecule drug development with a view to distilling guidelines. The first section covers research productivity feeding into commercial development before reviewing the literature on statistics of academic development It then considers differences between probes and drugs before discussing the role of author guidelines in medicinal chemistry and pharmacology journals. Resources for comprehensive compound and target cross-checking are then covered followed by comparisons between public and commercial databases including case studies of selected compounds. It concludes with an outline of new scientific developments that could increase the success rate of academic drug development.

Preface

As most readers of this document will know, the Karolinska Institutet (KI) is a medical university with global outreach and collaborations in education as well as research. Readers will also no doubt be aware of the International Union of Basic and Clinical Pharmacology (IUPHAR) as a voluntary, non-profit, association of pharmacology societies representing the interests of pharmacologists world-wide. Consequent to the signing of a Memorandum-of-Understanding between KI and IUPHAR in 2019 the two organisations are now jointly promoting academic drug discovery. This document has been prepared for the initial phase of collaborative outputs as specified by Per Arvidsson, Director, Drug Discovery & Development Platform at SciLifeLab and Michael Spedding IUPHAR Secretary General (see acknowledgments for links). The envisaged utilities of this initial version are;

1. Forming the basis of KI/IUPHAR recommendations
2. Instantiation as an open document to enable broad feedback (including suggestions for additional key references or illustrations)
3. Spawn one or more journal articles as decided by the stakeholders (potential publishers are welcome to make contact)
4. To seek funding for further collaborative outputs (e.g. a report specifically addressing the needs of resource-limited countries)

For ease of reading at this drafting stage (pending more detailed formatting if required later) titles are given for reference links to PubMed or other sources. A larger set of ~ 120 references have been tagged in open Zotero links. Illustrations are pasted-in with attributions.

Glossary of non-standard abbreviations

- KI = Karolinska Institutet
- IUPHAR = International Union of Basic and Clinical Pharmacology
- GtoPdb = IUPHAR/BPS Guide to Pharmacology
- ADEV = Academic Drug Development
- CDEV = Commercial Drug Development
- CI = Competitive Intelligence
- RLC = Resource-limited country
- PAINS = Pan Assay Interference Compounds
- AI/ML = Artificial Intelligence and Machine Learning
- ADDC = Academic Drug Discovery Consortium
- ELN = Electronic Lab Notebook
- FAIR = Findable, Accessible, Interoperable and Reusable
- EPMC = European PubMed Central
- RRID = Research Resource Identifier

Outline

Many of the topics addressed are big themes whose coverage depth have to be limited but key references are included. Notwithstanding the importance to academic R&D of biologicals, cell-based treatments, nucleic acid therapeutics, vaccines, gene therapy and formulation/delivery advances, this document will focus on small-molecules and endeavour to cover:

- Research productivity feeding into development
- CDEV statistics derived from Nature Reviews Drug Discovery and Citeline
- CDEV quality criteria from the literature including the Pfizer Three pillars and AstraZeneca 5R initiatives
- ADEV statistics
- ADEV quality criteria from academic organisations including NCATS
- Criteria for NIH probes and chemical biology organizations
- Journal guidelines in Medicinal Chemistry and Pharmacology

- Data resources for compound and target cross-checking
- Commercial competitive intelligence resources
- Case studies of selected ADEV published compounds
- Developments that could increase success rates
- Tabulating “compound fitness criteria” for iteration and refinement
- Appendix I is the NCATS 4D map, Appendix III is a list of ADDC operations, and Appendix II gives additional sources

****

Introduction

While much of their internal R&D data is proprietary, the efforts of the pharmaceutical industry, including the commercial drug development phase (CDEV), are extensively documented, even if not completely transparent. Evidence for this can be found in a simple PubMed query (Novartis OR AstraZeneca OR Pfizer OR Boehringer OR Glaxo, searching “affiliations” in the last 10 years) that returns 25,406 peer-reviewed journal articles (accessed July 2019). Searching WO patent assignees with the same names and time period returns 4,566 published patents. These documents disclose a significant amount of the proprietary data after a time lag of sometimes less than two years from the internal generation phase. However, terms of medicinal chemistry SAR first-disclosures this patent corpus represents a considerable (probably at least~ 5-fold) redundancy overcount due to the many ancillary applications around lead series, such as routes of synthesis, crystal structures, new indications and combinations.

In contrast to CDEV, academic drug development (ADEV) is more difficult to define. This is in part because of many different operational models that may include a commercial partner and other types of mixed funding. These cannot be easily differentiated from the general academic drug discovery-oriented activities nested within the biochemistry, pharmacology or chemistry departments of universities and institutes. Indeed, because of reputational kudos, many of these organisations declare such activity. The situation is further complicated by the practices of handing-off promising leads to another unit as well as Principle Investigators (PIs) leading the projects forming a minimal start-up (i.e. becoming a *de-facto* SME) as a route to seed funding.

Rather than trying to propose an abstract definition of ADEV the problem can be by-passed by compiling a list of operational units that would be considered as being under the broad ADEV umbrella. This has been done in Appendix III. However, output statistics, such as explicit counts of lead structures from such units, individually or collectively, are difficult to find. This is not an access problem in terms of primary disclosures (i.e. their papers, patents, press releases and clinical trials results surface alongside those of the commercial sector) but the difficulty lies on the indexing side. Much of this stems from the pragmatic aspect that commercial organisations, from the largest pharma company down to the smallest biotech, will self-assign specific names (even ‘Google clean’ where possible). Consequently, they can become explicitly identifiable as patent assignees, affiliations for journal papers, press releases, indexed in Google and, last but not least, for shareholder information. Searches will thus give useful levels of specificity.

In contrast, many ADEV operations, particularly those nested within large, multi-department academic organisations that frequently re-organise and change names, do not have ‘clean tags’

that can be used in the same way. For example, a recent medicinal chemistry patent (Phosphorodiamidates and other phosphorus derivatives of fingolimod and related S1P receptor modulators WO2019064012) is assigned to "University College Cardiff Consultants LTD". While they do have a UK Companies House entry and are assignee for 56 WO patent filings there is apparently no website, no exact match in PubMed affiliations and searching this field for 'Cardiff University' gives 16,276. This findability and searching precision problem is reflected by the fact that Citeline do not cover purely academic compounds (i.e. drugs have to be attached to a company name to be indexed, Ian Lloyd, personal communication).

Assessing research outputs feeding into development

While the focus of this document is on entry into drug development (i.e. setting the bar) consideration also needs to be given to the research output that feeds both commercial and academic pipelines. To state the obvious, in R&D the success of D is predicated on both the scale and quality of R. This applies at the global level as well as on an institutional and project team basis. As has been pointed out ('Tracking 20 years of compound-to-target output from literature and patents', 2013, PMID 24204758) quantitatively assessing R is difficult because it is not explicitly captured, even by CI sources. However, there are useful surrogates in the form of patents, papers and database entries for what are typically classified as lead compounds. Tracking of patents and papers outputs are shown in Fig.1.

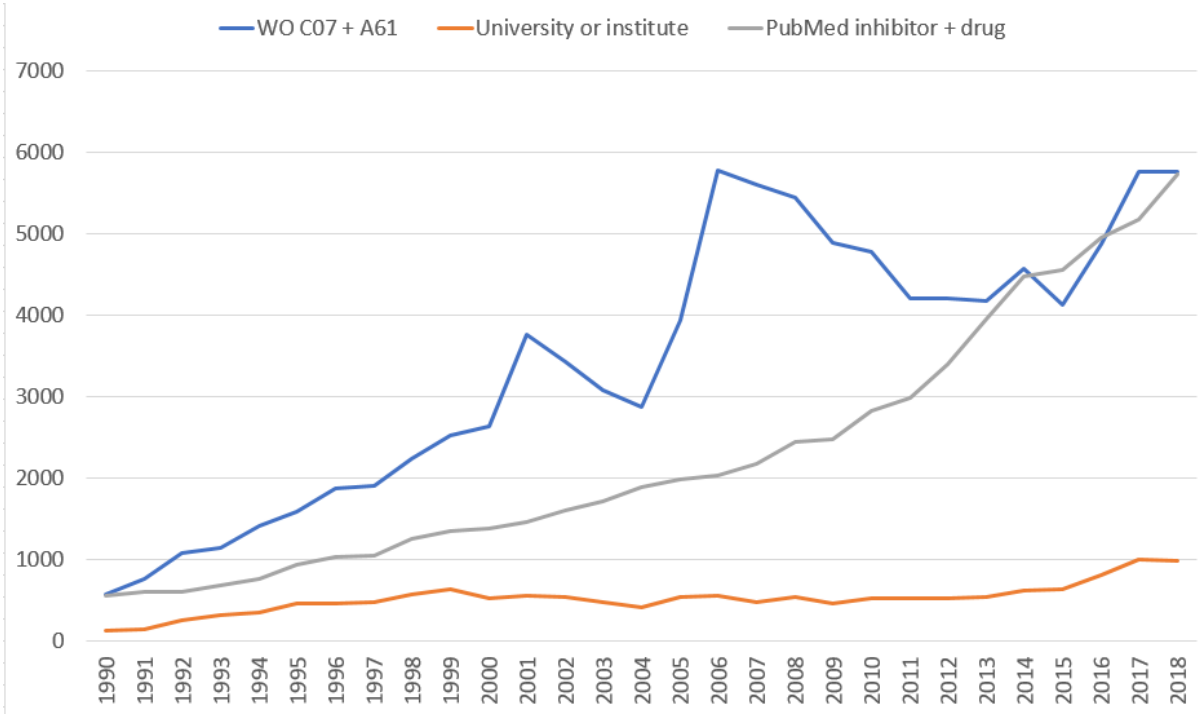


Fig.1 Open database estimates of discovery output. The uppermost blue line, generated via WIPO PatentScope, uses the two International Patent Classification (IPC) codes of C07 and A61 that are selective for medicinal chemistry patents. The lowermost orange line uses the same IPC code select as the blue line but with the additional filter of having "University or Institute" in the assignee field. The grey line is a PubMed select for "inhibitor" and "drug" in title or abstract (the slope for inhibitor alone was similar).

Possible explanations for the rise up to 2006 followed by an eight-year decline are discussed in PMID 24204758 and the accompanying 'In the Pipeline' commentary. However, while it is not clear exactly what factors might have contributed to the 2017/18 uptick, this appears to have recovered to the 2006 peak levels. It should be noted that university or institute does not provide a clean discrimination between commercial and academic drug discovery. Nonetheless, the divergence is clearly increasing because the academic (orange line) has flatlined while the commercial sector has been accelerating. There is perhaps a slight indication of an academic uptick over the last two years but note these plots have a lag time of at least 18 months extending to several years (i.e. we might speculate an increase in ADEV patenting over the last few years would not yet have come through). The breakdown of top patent applicants is shown in Fig.2.

Applicant		Pub Date	
Name	No	Date	No
F. HOFFMANN-LA ROCHE AG	1,611	2019	1,735
NOVARTIS AG	1,250	2018	5,764
BRISTOL-MYERS SQUIBB COMPANY	1,200	2017	5,657
ASTRAZENECA AB	1,121	2016	4,876
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA	1,002	2015	4,128
GENENTECH, INC.	888	2014	4,566
MERCK & CO., INC.	780	2013	4,173
BOEHRINGER INGELHEIM INTERNATIONAL GMBH	757	2012	4,213
GLAXO GROUP LIMITED	696	2011	4,216

Applicant		Pub Date	
Name	No	Date	No
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA	1,002	2019	395
BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM	332	2018	1,303
THE SCRIPPS RESEARCH INSTITUTE	292	2017	1,246
THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA	285	2016	1,025
THE JOHNS HOPKINS UNIVERSITY	238	2015	810
THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK	226	2014	847
		2013	739
THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY	217	2012	741
SHANGHAI INSTITUTE OF MATERIA MEDICA, CHINESE ACADEMY OF	209	2011	722

Fig.2. The panels are the results totals of the blue line query from Fig.1 on the left (IPC codes C70 + A61) and the orange line on the right (selecting university or institute in the assignee field).

As expected, applicants in medicinal chemistry are historically dominated by large pharma companies (n.b. the Glaxo matches are pre-GSK filings), but with the perhaps unexpected inclusion of the Regents of the University of California in 5th place. This puts them at the top of the right-hand panel that is led by US universities but note the Chinese academy in 9th position. However, inspection of the patents on the right show that many are focused more on biotech-type claims rather than medicinal chemistry composition-of-matter on the left (i.e. they tend more towards A61 than C07)

Detailed consideration of ADEV-specific patenting strategies cannot be covered here but two opposing trends play into the publication statistics above. Good quality, data-supported patents are expensive to prepare and maintain, not to mention requiring substantial experimental chemistry resources (even if contracted out) to generate sufficiently broad SAR exemplifications. While there are no detailed statistics for ADEV operations (due to lack of clean tagging) they tend to patent early towards the end of the research phase. This allows both prompt publishing after 18 months and gives ample time to approach those funders who insist on patent protection (n.b. while this includes some organisations for NTD research badged as "non-profit", in the US this is tax-based classification). By comparison, larger pharmaceutical companies usually file their novel chemistry claims later in the development cycle thereby maximising patent life and subsequent profits. They also have more resources for follow-up filings to protect back-up compounds, synthetic routes and formulations against the opportunistic filings of generic manufacturers that pop up quickly after risk-reduction is disclosed via positive efficacy data. There is an inherent tension between these two approaches. It should also be noted that ADEV operations may not have sufficient resources

for broad and deep IP protection but also if they default on maintenance charges their compounds become non-developable (n.b. the legal status and examination reports of patents are available but non-trivial to check).

Statistics of CDEV

As a prelude to assessing ADEV outputs it is important to benchmark against CDEV for two main reasons. Firstly, there is more specific and detailed information available. Secondly, most aspects of CDEV are also directly relevant to ADEV. As we know, drug development is not only difficult and expensive but also has low success rates. Compounding the challenge is the observation (reported so often in recent years that it has become a cliché) that CDEV productivity, in terms of "dollars in *vs* successful drug revenues out", has been declining over decades. Indeed, this is said to pose a threat to the long-term viability of the traditional pharmaceutical industry business model. While a landmark review pointing to the severity of the problem was published in 2009 ("Lessons from 60 years of pharmaceutical innovation PMID 19949401) a later well-cited review appeared in 2012 also in a *Nature Reviews Drug Discovery* (that included the notable "better than the Beatles" analogy) and contained the iconic chart shown below.

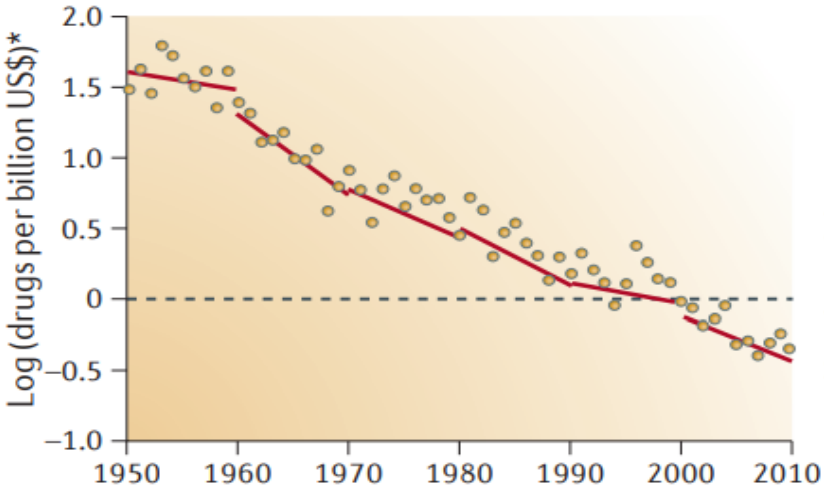


Fig.2. The rate of decline (as discontinuous red lines) in the approval of new drugs per billion US dollars over10-year periods. Adapted from "Diagnosing the decline in pharmaceutical R&D efficiency" (PMID 22378269, 2012).

Analogous statistics have been published regularly since then presenting different perspectives on the return on investment theme for large pharmaceutical companies (including additional articles by Munos). The latest of these, a report by Deloitte, declares that 2018 R&D returns have declined from 10% in 2010 to 2% in 2018, the lowest level in nine years (Fig. 2). Underlying this is the reported growing cost of bringing a drug to market which at $2.2 billion, is almost double the $1.2 billion recorded in 2010.

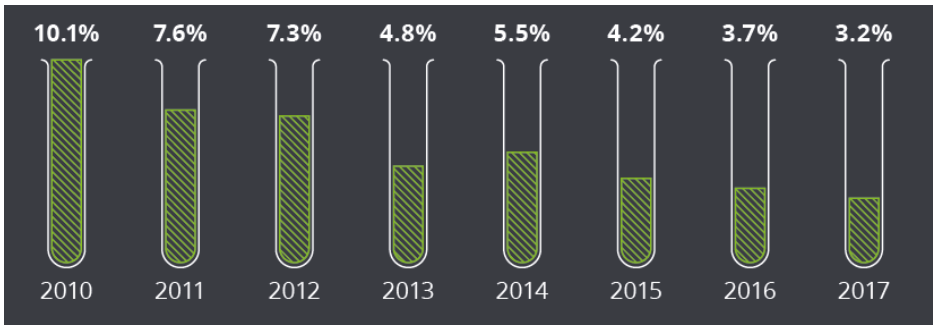


Fig. 3. Yields on investment for 12 major pharma companies. Adapted from "Measuring the return from pharmaceutical innovation" (Deloitte 2018)

However, this glass-half-empty introduction is at least partially countered by a glass-half-full valuation of the 2018 global small molecule drug discovery market as having reached ~30 billion USD with a projected 5-year growth rate of ~8%.

A useful (and free) statistical overview for CDEV is provided by the Citeline Pharma R&D Annual Review 2018. The good news is that the count of drugs in R&D from just over 4,300 companies has reached 16,181, of which 8,285 are small-molecules. Notwithstanding, circumspection is needed on interpreting this as a distinct molecular structure count since ~40% are blinded code names. Over the past three years rises of 8.4%, 2.7% and 6%, have put 2019 slightly above the three-year mean. Further grounds for optimism is the proportional increase in designations for breakthrough therapies, accelerated approvals and orphan drugs (with some of these three being for rare disease treatments).

However, the long-term less good news is encoded in the cumulative clinical raw counts for Phases I,II and III as shown in Fig.4.

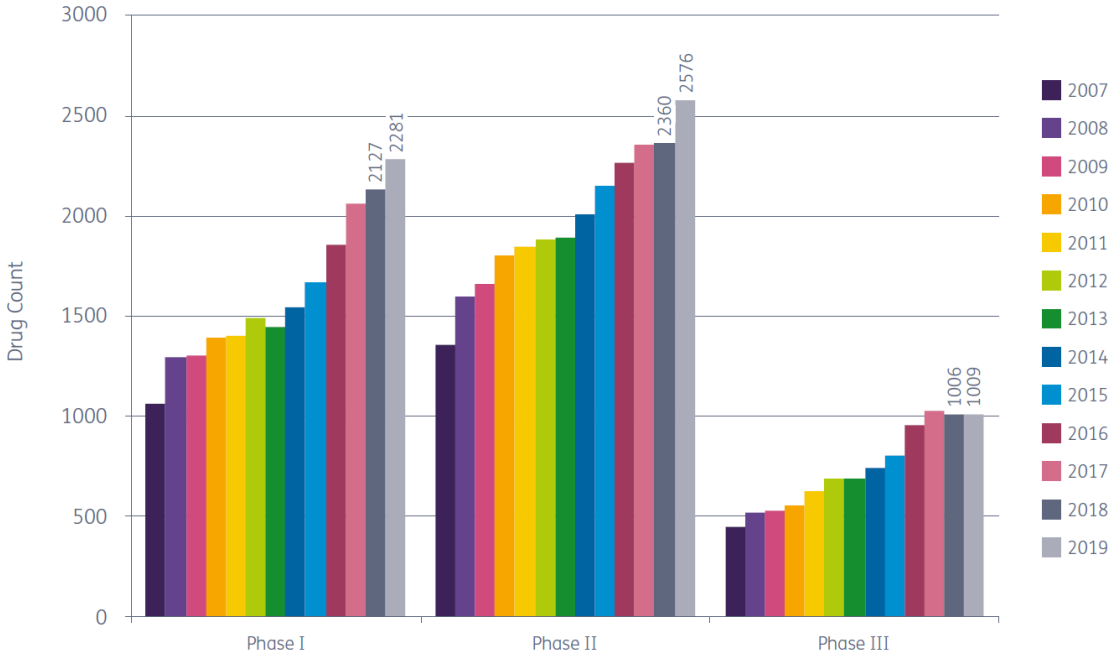


Fig.4 Numbers of drugs in each clinical stage from 2007 to 2019 (Adapted from Citeline's Pharma R&D Annual Review 2018)

This shows that Phase II-to-III transitions (colloquially known as the ‘valley of death’) has essentially not moved the dial over the last decade, even though there can be hardly any CDEV operations who have not tried their utmost to improve their own numbers. However, these are averaged figures and the top-ten pharma companies do show signs of moving above them (as expected, since this would be one of the reasons for them achieving a high ranking). This is illustrated in Fig.5.

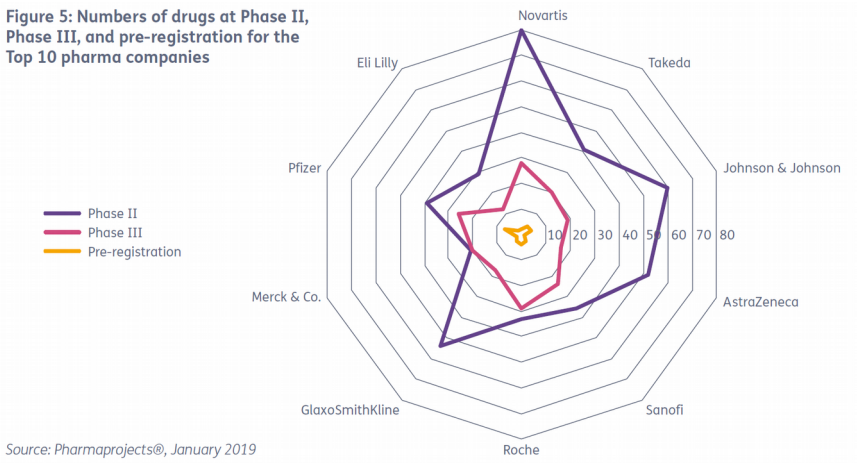


Fig.5. Numbers of drugs at Phase II, Phase III, and pre-registration for the top-10 pharma companies.

The interpretation of Fig.5 is that while Novartis currently leads in reporting more Phase II than Phase III drugs, other companies, such as Merck & Co., have similar numbers at both phases. As we zoom in towards centre with the orange data for pre-registration the numbers for all companies shrink and look more similar.

The productivity of CDEV can be viewed through different lenses. These include shareholder profits, number of quality-adjusted patient years achieved or even the continued health authority cost-saving transitions to generics and over-the-counter medicines. However, the most commonly used metric is counting the number of drug approvals. These outputs are officially collated and declared by the FDA and the EMA, with the most recent cumulative set from the former being shown below in Fig.6.

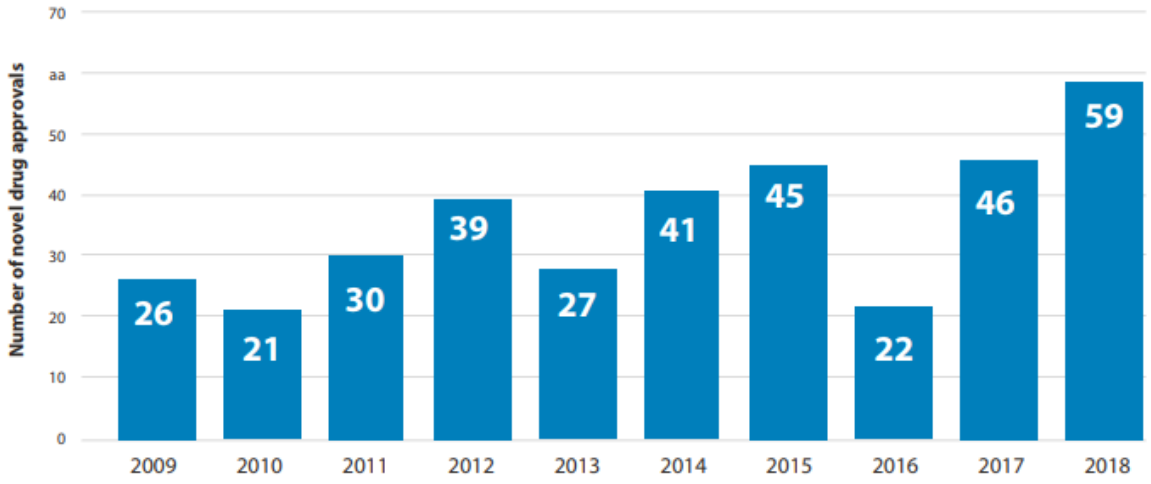


Fig.6a. The 10-year graph (adapted from the FDA Center for Drug Evaluation and Research CDER) shows approvals averaging about 33 novel drugs per year.

There is an encouraging upward trend in Fig. 6a and a detailed analysis is presented in this blogpost on the 2018 structures in PubChem. As usual, a number of pundits (referenced in the blogpost) have pronounced their judgments on this annual harvest. Those from Munos are most relevant and can be perceived as encouraging for the theme of ADEV. He interprets the increasing proportion of SMEs as signs that having to work at 'big pharma scale' is no longer so essential (although despite the desirability of new drugs for very rare diseases they have little impact on population health). He goes further to suggest the increasing commodification of big science will mean that development successes may come from scientist's ability to extract knowledge from open data (and by implication ADEV teams who can engage with this). However, it has to be noted that no academic-only drugs have appeared in the years covered by Fig.6, emphasising the necessity of commercial support to achieve FDA approval. Another recent publication from Clavariate ('Trends in clinical success rates and therapeutic focus', 2019, PMID 31267067) distils data from ~30 large, mid- sized and small companies draws similar conclusions but complements these with a useful seven-year breakdown by therapeutic area shown in Fig. 6b.

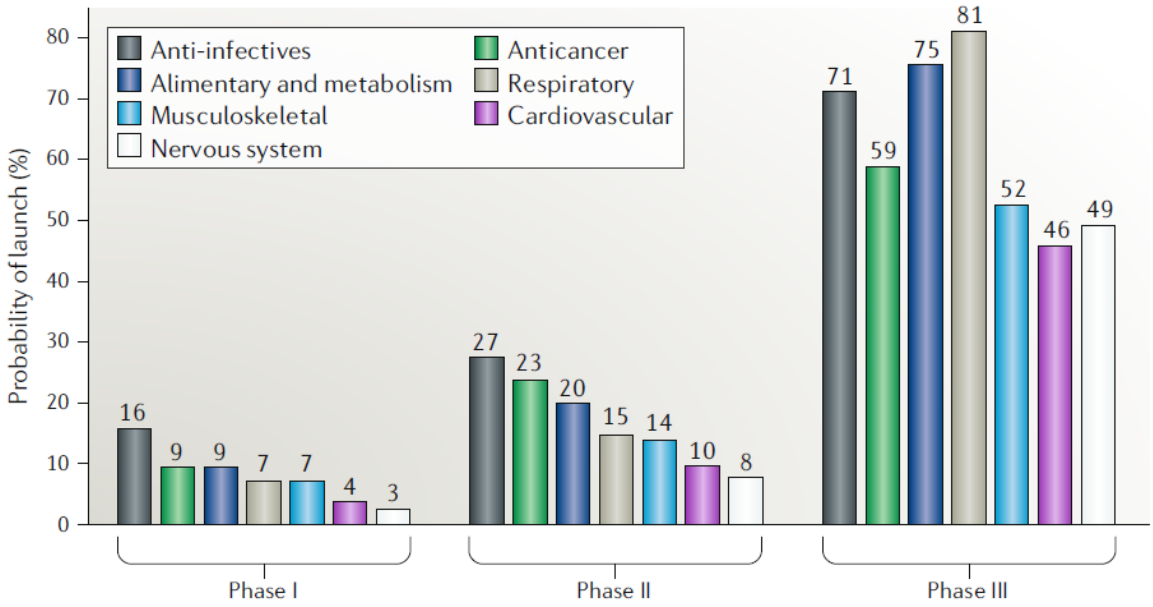


Fig. 6b. Probability of launch from Phases I, II and III by therapeutic area from 2010–2017 for new active substances (adapted from

CDEV quality criteria

So, what factors determine the successful approval of new therapeutics in the commercial sector? The literature on this is extensive, especially for the hit-to-lead optimisation phase. However, there is no magic formula since the ratios in Fig.4 have proved largely immovable. It may thus be more useful to firstly look at reasons for failure and secondly find examples of improvement strategies that have tangible evidence of a payoff.

To address the first question there is limited systematic public data on exactly why and how frequently novel agents fail in late-stage development (even in CI resources). However, historical studies and anecdotal reports (although mostly from the CDEV sector) imply that new drug applications (NDAs) most commonly fail from deficiencies in efficacy or safety. More detailed late development statistics were collated in a 2016 article 'Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results' (PMID 27723879). This concluded, that from 640 novel therapeutics, 344 (54%) failed in clinical development, 230 (36%) were approved and 66 (10%) were approved in other countries but not by the FDA. The paper went on to determine that (not particularly surprisingly) 57% of the failures were ascribed to inadequate efficacy, 17% due to safety issue and 22% for commercial reasons.

A useful orthogonal study (i.e. using non-US data and not from major pharma) has looked at the classification of objections raised against human medicine marketing authorisation applications from by SMEs to the European Medicines Agency (EMA) between 2011 and 2015 and their impact on the outcome of applications (PMID 29953957). One of the informative charts this report includes is shown in Fig.7

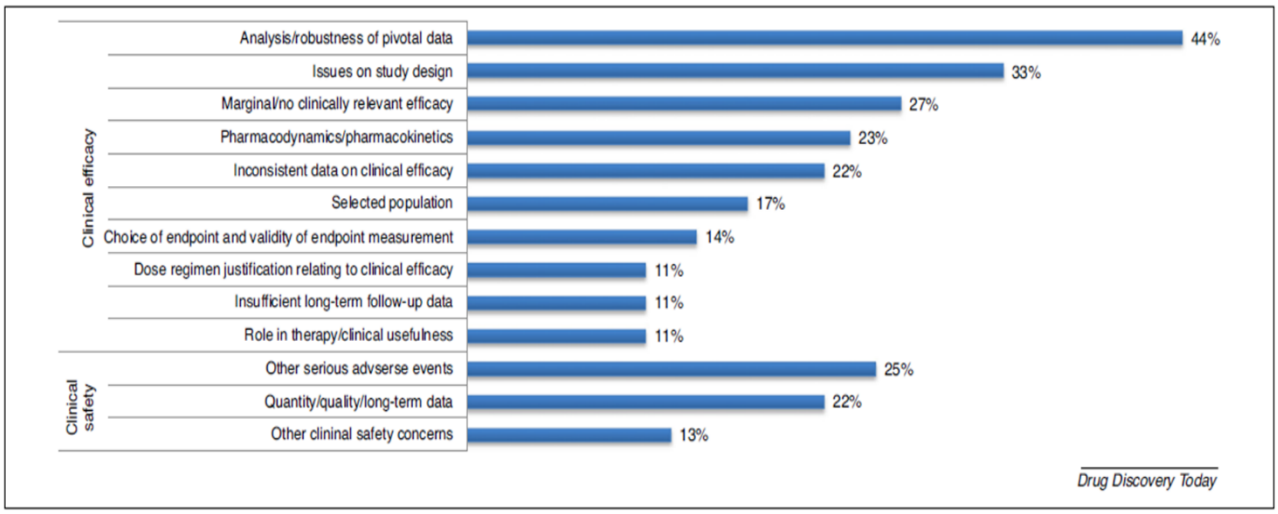


Fig.7. Percentages of quality objections (above 10%) raised in EMA marketing authorisation applications by 64 SMEs between 2011 and 2015 (adapted from PMID 29953957)

There are common themes between these two assessments but the depth of detail associated with Fig 7 is more relevant to ADEV guidelines.

Given the intense commercial imperative, at least two large pharma companied have declared improvement strategies based on a pragmatic examination of their own internal 'score-cards'. A well-cited example from Pfizer in 2012 was 'Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival' (PMID 22227532). Based on Phase II decisions for 44 programs as well as detailed case studies, the 'Three Pillars' criteria they came up with are focused on 1) understanding drug exposure, 2) target binding and 3) pharmacological activity at the target site of action through application of PD and PK modelling. A matrix-type summary is shown in Fig.7.

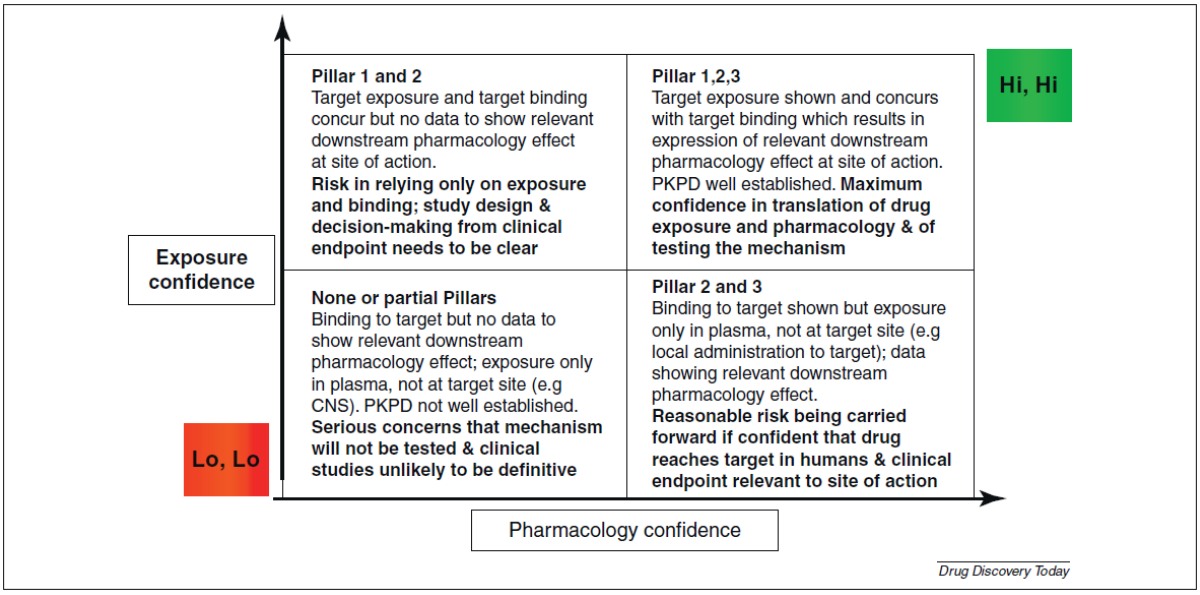


Fig.7. Clinical development risk management to assess likelihood of testing the mechanistic and quality criteria for program progression (adapted from PMID 22227532)

There are useful general points exemplified in case study details in this Pfizer exposition on their way of doing things back then. However, not only was there no disclosed quantitative data on subsequent improved outcomes but the Citeline Annual Review figures imply this actually went into reverse. In 2013 Pfizer was 5th in the global rankings with a pipeline of 202 drugs but by 2019 they had fallen to 9th with 163. It has to be said that there are many factors that affect the rankings arrived at by Citeline, including differences in therapeutic area coverage between companies. It is thus unwise to consider comparative pipeline metrics as the 'ground truth' of CDEV efficiency. Notwithstanding, there appears to be no openly declared data from Pfizer that verified improved Phase II/III transitions as a consequence of the 'Three Pillars' initiative.

Despite caveats associated with such rankings, for AstraZeneca (AZ) the Citeline metrics in this case indicate an improvement, from 8th place in 2013 with a pipeline of 157 rising to 4th in 2019 with 194. The significant corollary here is that these are declared to be (according to AZ obviously) causatively related to their 2018 publication 'Impact of a five-dimensional framework on R&D productivity at AstraZeneca' (PMID 29348681). Not unexpectedly, there is considerable overlap with 'Three Pillars' concept. However, the AZ initiative, beginning internally in 2011 and first described externally in 2014 (PMID 24833294), was not only broader in scope but the 2018 perspective does provide a detailed statistical follow-up. There is a lot of valuable detail in this publication, including case studies and descriptions of steps taken to improve the quality of leads entering development. The major distillations from the '5Rs' framework are outlined as follows:

1. *Right target* • Strong link between target and disease • Differentiated efficacy • Available and predictive biomarkers
2. *Right safety* • Differentiated and clear safety margins • Understanding of secondary pharmacology risk • Understanding of reactive metabolites, genotoxicity and drug–drug interactions • Understanding of target liability
3. *Right tissue* • Adequate bioavailability and tissue exposure • Definition of PD biomarkers • Clear understanding of preclinical and clinical PK/PD • Understanding of drug–drug interactions
4. *Right patient* • Identification of the most responsive patient population • Definition of risk–benefit for a given population
5. *Right commercial potential* • Differentiated value proposition versus future standard of care • Focus on market access, payer and provider • Personalized health-care strategy, including diagnostics and biomarkers

Internal statistics indicate the implementation has translated into success rates from candidate selection to Phase III completion, improving from 4% in 2005–2010 to 19% in 2012–2016. As would be expected, failures still persist but there is an informative breakdown as to how the reasons for these have shifted as a consequence in Fig. 8.

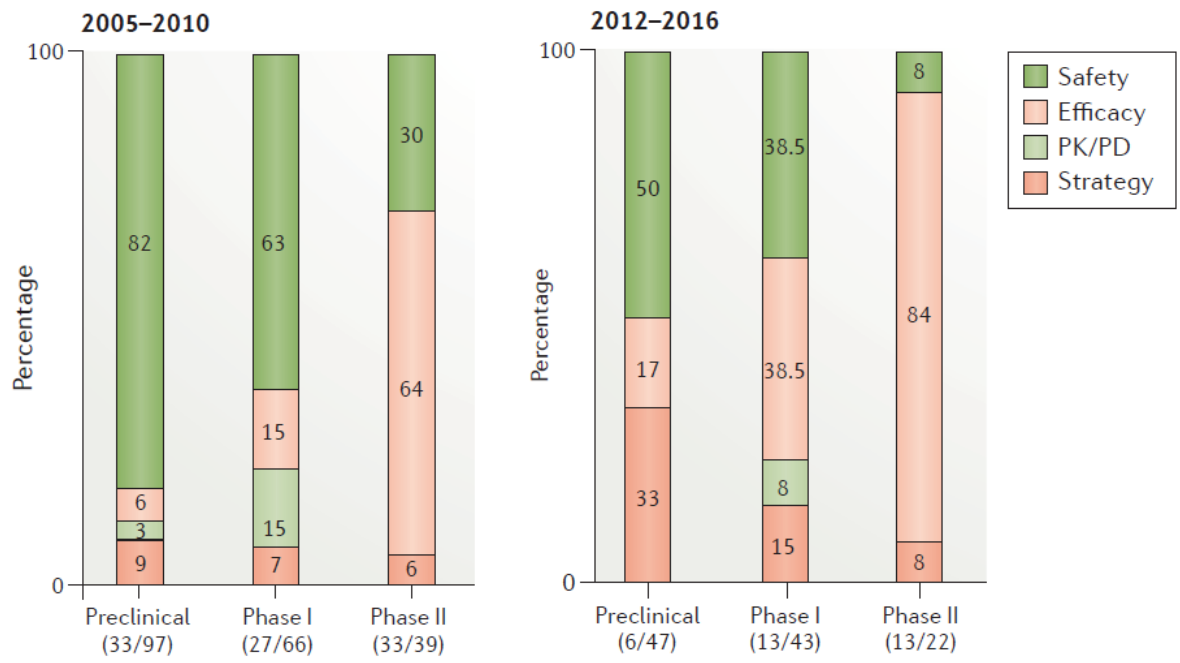


Fig.8. The lower figures are project success rates for the AZ portfolio during 2005–2010 and 2012–2016 compared with industry benchmark data. The stacked column plot compares the primary reasons for failure (adapted from PMID 29348681 where a more detailed legend can be found)

Significantly, as an independent endorsement of the real-world success of '5Rs', Dr Mene Pangolos, head of AZ R&D, was recently awarded the Prix Galien. Despite this, not all the development challenges are covered by the '5Rs'. In particular, the longitudinal statistics indicate that CDEV productivity peaked several decades ago, paradoxically when fewer molecular targets were resolved and less known about their pathways. It has thus been argued that disease-state phenotypic pharmacology was the major contribution to this historical success rate and that, against prevailing expectations, this actually declined in parallel with increasing post-genomic target-centricity. While the list of variables affecting even just drug < > receptor interaction is dauntingly large, this type of direct pathology-relevant screening can eliminate many of them (see 'Resolution of controversies in drug/receptor interactions by protein structure. Limitations and pharmacological solutions', 2011, PMID 20709088). However, as reported by one major pharma company, swinging the pendulum back towards phenotypic screening turned out not to be a panacea (see 'How Phenotypic Screening Influenced Drug Discovery: Lessons from Five Years of Practice', 2017, PMID 28800248). Notwithstanding there is a possible silver lining here for ADEV teams (including in the developing world) who, while not having access to a large portfolio of molecular assays may have developed unique types of disease-state phenotypic pharmacology, even with low-throughput capacity.

Statistics of US ADEV

As mentioned above, the caveat with analysing ADEV is that a) as a strict category it is largely 'dark' in CI databases and b) progression is usually associated with some type of commercial collaboration c) many ADEV units were only set up recently compared with the CDEV decade-plus development times. Notwithstanding, there are informative publications, the most recent of which is 'The Current Status of Drug Discovery and Development as

Originated in United States Academia: The Influence of Industrial and Academic Collaboration on Drug Discovery and Development' (PMID 29940695). The Japanese authors investigated 798 drug discovery projects between 1991 and 2015 at 36 academic institutions in the United States and recorded 75% at Phase I, 50% Phase II, 59% Phase III, and 88% NDA/biologics license application (BLA) phase. At first sight this looks encouragingly glass-half-full (e.g. compared to Fig.4) but the constitutive problem is that not only publications but also CI sources do not generate their progression statistics in exactly the same way (e.g. using different indication merges). However, the glass-half-empty message from this paper is encoded in Fig.9.

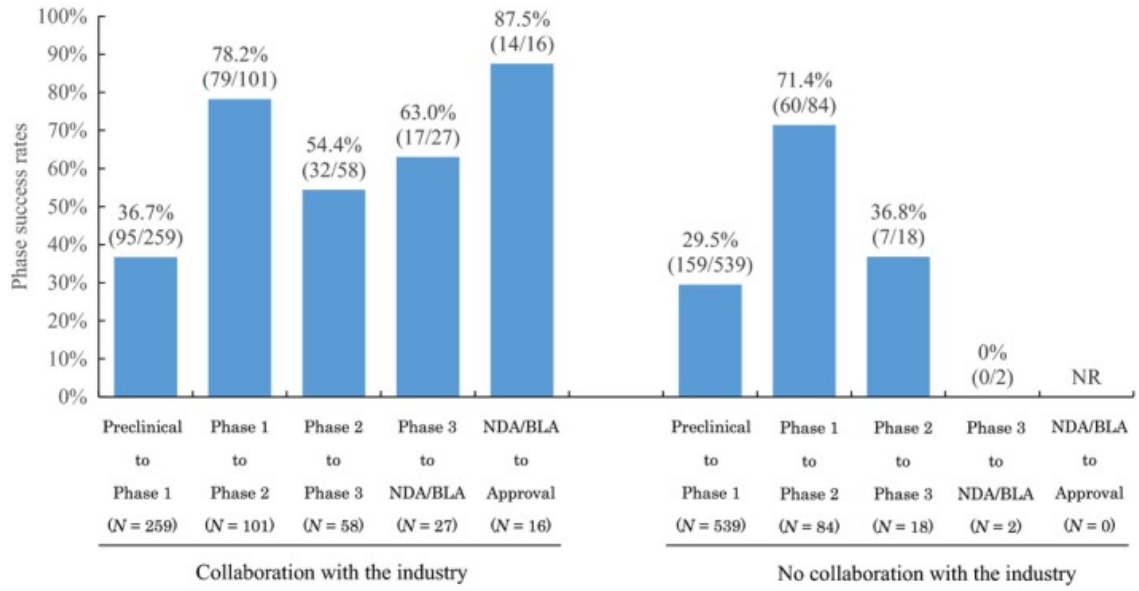


Fig 9. Comparison between with academic-industrial collaboration on the left and without on the right between 1991 and 2010 (adapted from PMID 29940695).

The academic going-it-alone success is dramatically lower (e.g. two Phase IIIs and no approvals at all) compared to industry collaboration. In the US the intersection of CDEV and ADEV, including their statistics, has become a controversial topic. The arguments pivot around the idea that public-sector scientists (with the NIH being the major funder) perform basic research on mechanisms of disease and points of intervention (but not ADEV at this early stage) whereas commercial organisations have performed the applied research resulting in the discovery of drugs for the treatment of diseases and have carried out the CDEV activities to bring them to market. The bone of contention is that too little of the profits (i.e. from ~$30 billion but not all US companies) feed back into public-funded research (that could of course boost ADEV). However, the boundaries between public and private sectors have shifted substantially since the biotech boom.

Many commentaries and papers have appeared on this theme but most take a predominantly US perspective. There is a useful summary in a Dereck Lowe 2016 blogpost 'Drugs Purely From Academia'. This includes a list of compounds distilled from PMID 21031002 where the focus was on the contribution of smaller companies as newer discovering organisations. Another publication covering some of the same ground but more specifically looking at the academic contribution was a 2011 paper 'The Role of Public-Sector Research in the Discovery of Drugs and Vaccines' (PMID 21306239), This found that over the 40 years, 93

small-molecule drugs were discovered through research carried out in the academic sector. They also state this sector has contributed to the discovery of ~10% - 20% of all NDAs from 1990 to 2007. However, these papers have not generated statistics for the direct involvement of ADEV teams.

The references above have a recent addition in the form of a *J Clin Invest* review article ('Academia and industry: allocating credit for discovery and development of new therapies', 2019, PMID 31107243, which also features as an 'In The Pipeline' commentary with additional references). While this is also US-centric, it includes a well-framed summary that is directly quoted in its entirety to round off this section (and introduces some points that will be picked up later on); 'The discovery and development of new therapies has and will likely continue to require contributions from academic institutions and the biopharmaceutical industry. Most (but not all) new insights into biology, disease, and new technologies arise in academia, funded by public grants, foundations, and institutional funds. Academic institutions identify promising discoveries and seek to initiate their development and commercialization, eventually partnering with for-profit biopharmaceutical companies, either established or venture-funded start-ups. To complete the path from molecular insight to approvable therapy, major involvement of industry is required, as it possesses the insights, culture, skills, and capital typically unavailable in the academic realm. Cooperation between academia and industry takes many forms, including transfer of information, intellectual property, and reagents; consultation with industry by academic experts; and movement of people between these domains. The two cultures are different: teamwork toward shared therapeutic goals characterizes industry, and diverse paths to individual credit characterize academia. This cultural divide sometimes impedes successful interactions. A goal of public policy should be to facilitate such interchanges to enhance the success of the combined enterprise'.

Statistics of European ADEV

While there fewer comparable descriptions of what could be classified as European ADEV one of the most significant developments has been setting up of the Innovative Medicines Initiative (IMI) as a public–private partnership between the European Union and the pharma industry. This has an overall budget of €5.3 billion covering 2008 until 2024. The objective is to boost pharmaceutical innovation in Europe and speed up the development of innovative medicines, particularly for unmet clinical needs. This has been most recently reviewed in 'The European Innovative Medicines Initiative: Progress to Date' (PMID 30174434). Interestingly (and reflecting back on the AZ PubMed count given above) they count joint project publications as output with 3800 during the period 2010–2017. More tangible achievements are declared as 35 new validated drug targets; 30 new in vitro models and tools; 300 new in silico (computer-based) models; 70 new animal models; 1500 new stem cell lines; a dozen novel imaging techniques; and 130 validated biomarkers. However, there is no listing of attributable development compounds as yet.

A specific example from this undertaking was IMI Newmeds which had the objective to explore novel methods leading to new medications in depression and schizophrenia. The constituted psychiatry consortium, comprising of multiple pharma and academic groups, was awarded 22.5M€ over five years. The project validated multiple approaches between independent partners. However, it also implicated so many mutations and copy number variants in schizophrenia that it was not practically possible to discern molecular targets for the many individual changes. This means assessing therapeutic impact at the brain circuit level (and therefore a measure of phenotypical screening) was necessary (see 'Defining the

brain circuits involved in psychiatric disorders: IMI-NEWMEDS', 2017, PMID 27811930). However, this approach has opened the way for treating depression with ketamine and related drugs.

For the UK specifically, there was a 2014 survey on drug discovery (PMID 24378794) but this was later criticised for reliance on limited questionnaire responses. This means the charts therein, while of interest, may not have been statistically representative snapshots. A re-analysis from the following year ('Academic drug discovery within the United Kingdom: a reassessment' PMID 26091271) identified 24 dedicated drug discovery facilities in the United Kingdom, some operated by universities with sources of funding including charitable organizations or connected to industrial facilities near to academic campuses. There was also a common feature between these groups of having some proportion of academic staff and being accessible to any UK academic groups. While Table 2 in this publication is an informative compilation of these facilities (with links) there is a strong focus on the those organised around the provision of discovery-phase high throughput screening (HTS) including compound collections This is not a criticism of the report but is missing clinical candidate counts as tangible ADEV outputs. The 2019 edition of the State of the Discovery Nation report by the Medicines Discovery Catapult and the UK Bioindustry Association (BIA) indicates that ~300 companies are focussed on discovering potential new medicines, of which ~ 70% cover cancer, anti-infectives and CNS. In addition, ~1,200 companies provide services and supplies. The report thus indicates strong growth since the 2014 assessment (as PMID 26091271). It also makes clear that SMEs are a critical source of innovation for new medicines but 60% have fewer than 5 staff and 80% have fewer than 20 people within the company. It also highlights cutting-edge technology to make medicines discovery more productive in the areas of Artificial Intelligence (AI) and Complex Cell Models (CCMs).

In terms of other European countries are many other drug discovery units, but, as for the UK, these tend to be focused on the provision of screening facilities. One important exception is the Science for Life Laboratory Drug Discovery and Development (SciLifeLab DDD) platform set up in Stockholm in 2014. This has three publications describing its *modus operandi*. In 2015 'Open for collaboration: an academic platform for drug discovery and development at SciLifeLab' appeared (PMID 27373760), followed by 'Institutional profile: the national Swedish academic drug discovery & development platform at SciLifeLab' in 2017 (PMID 28670468) and also in 2017 the (aptly named in the context of this document) 'On the bridge over the translational valley of death: interview with Per I Arvidsson' (PMID 28670474). As a series, these three open access papers present useful detail on the Swedish model for ADEV. The details do not need reiterating here but the state-of-the-art expert facilities that academics gain access to are listed below:

- Target product profiling and drug safety assessment
- Compound centre (screening plate distribution from < 200,000 compounds)
- Protein expression and characterization
- Biochemical and cellular screening (capacity up to 30,000 compounds)
- Human antibody therapeutics
- Biophysical screening and characterization
- Absorption, distribution, metabolism and excretion (ADME) of therapeutics
- *In vitro* and systems pharmacology for drug molecular mechanism of action (mmoa)
- Medicinal chemistry as Hit2Lead (in the Medicinal Chemistry Department at Uppsala University)

- Medicinal chemistry lead identification (also in Uppsala)
- Delivering de-risked research programs suitable for partnering according to various funding models and IP arrangements

While the issues cannot be detailed here, progressive funding is obviously critical for ADEV. Even a conservative figure for getting to early-phase human clinical trials would be ~ 1.5 Mill $. This is a major barrier for small units and underscores the need to seek commercial support.

NCATS documentation

This section crosses back over the Atlantic to consider the NIH National Center for Advancing Translational Sciences (NCATS). This has produced a range of documentation and publications that is of value not just in general terms for ADEV but also for practical guidelines. One resource that has become well appreciated by the research community is their Assay Guidance Manual (AGM). This is a best-practices online resource with the objective to robustify early-stage drug discovery results from HTS assays and focussed screens. Chapters are written by experts and edited by representatives from industry, academia and government. The eBook is thus a comprehensive reference for optimising assays with the goal of developing probes that modulate the activity of biological targets, pathways or cellular phenotypes (see later section on probes) and of course candidates for drug discovery and development. There are many sections in the manual but those on mitigation of assay artefacts and interferences, statistical validation of assay performance parameters, data standards for reporting the results of screening and SAR assays and *in vivo* assay validation provide critical guidance for ADEV. There is also a useful chapter entitled 'Early Drug Discovery and Development Guidelines: For Academic Researchers, Collaborators, and Start-up Companies' (2016, NBK92015).

NCATS have also published workflow diagrams (see 'A dynamic map for learning, communicating, navigating and improving therapeutic development' PMID 29269942). What they have termed Drug Discovery, Development and Deployment Maps (4DM) provide dynamic representations of the modern therapeutic development process (n.b. while NCATs has a mission-focus on rare diseases the 4DM concept is clearly generalisable). The maps can help those planning drug development to identify and explore solutions to their own bottlenecks as well point to where they need to establish collaborations to take a compound forward. A snapshot of the on-line interactive version is shown below Fig and a full-size diagram for small molecules as Appendix III.

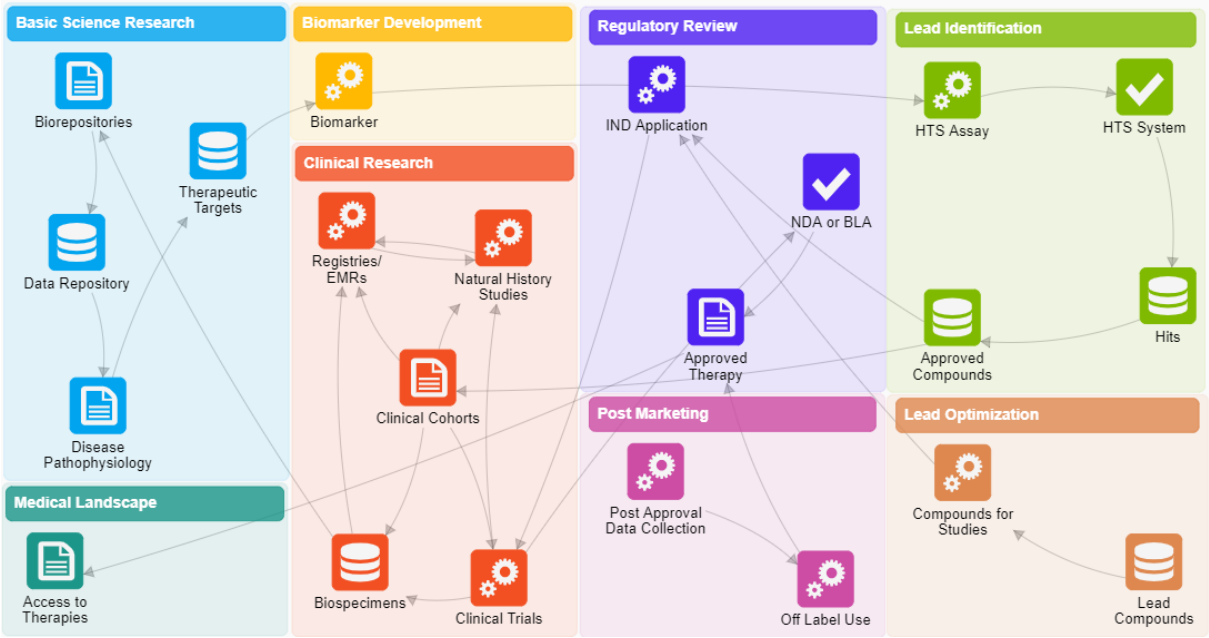


Fig. 10 4DM outline sourced from https://4dmap.ncats.nih.gov/#/. On the live version each icon expands out with a text description and links between them can be navigated.

Probes vs. drugs

Although they diverge in objectives, probes for chemical biology and leads for drug discovery converge in some of their desirable attributes. However, we can use an empirical definition that biological system insights, *in vitro* and *in vivo*, along with advantageous property profiles are the primary goals for probes. For drugs, the therapeutic outcome in human patients is obviously the overriding objective (and by definition also for ADEV). Notwithstanding, drugs need characterising by probing biological systems to determine not only the mechanistic basis of therapeutic efficacy but also that their non-efficacious effects are understood enough to be acceptable (e.g. the systems pharmacology characterisation offered by SciLifeLab DDD). Drug candidates must be mandatorily tested in rodents but relevant mechanistic insights (even unpredicted ones) may also be generated from probing experiments (a.k.a. recording functional perturbations) in yeast, *C.elegans*, Zebrafish or more recently characterised non-standard model organisms.

Probes have become popular in recent years for several reasons. Firstly, they are they are optimised for potency, target specificity, robustness and reproducibility (as inherent properties by definition) which can be lacking in published drug discovery compounds declared as leads. Secondly, both the data and the compounds themselves are typically freely shareable in the open science sense (some may have published in paywalled journals and many have IP claims but these are implicitly waived). Thirdly, they can be seen as the future saviours of the drug discovery enterprise, commercial and academic. The basis of this bold claim is their conceived potential to illuminate 100s or even pushing towards 1000 new drug targets as the concept behind the druggable genome. If realised, this would become a boom for ADEV (see 'Far away from the lamppost' PMID 30532236. However, this stands in stark contrast to targets for currently approved drugs being well below 400 proteins (i.e. those under the lamppost). Indeed the current GtoPdb statistics indicates only 328 human primary targets of approved drugs. This implied slow rate of new target validation thus seriously constrains ADEV (see 'Darkness in the Human Gene and Protein Function Space: Widely Modest or

Absent Illumination by the Life Science Literature and the Trend for Fewer Protein Function Discoveries Since 2000', 2018 PMID 30265449).

The history of chemical probes can be divided into two phases. The first NIH-centric initiative, from roughly 2004 to 2014, is described in the year-six reports of the online books series 'Probe Reports from the NIH Molecular Libraries Program' (NBK47352). It was instigated as the Molecular Libraries Probe Production Centres Network (MLPCN), a US-wide consortium of centres that were to produce probes. They solicited novel assays from the research community for high throughput assays (HTS) against what was initially a library of 350,000 compounds maintained as the Molecular Libraries Small Molecule Repository, MLSMR (this has expanded modestly over the last five years to 406,000). Most primary screens went through an orthogonal second round designated as confirmatory assays in PubChem BioAssay . Validated hits were optimised by medicinal chemistry (usually from PIs collaborating with the centres and sometimes filling patents) to produce *in vitro* chemical probes with structures and assay results eventually deposited into PubChem (indeed, this was the raison d'être for setting up PubChem in 2004).

By 2014 most of the original screening centres were no longer focussed on probe production. Thus, their generation within the historical system has ceased. The exercise become controversial in terms of the level of investment put in *vs* the yield of probes that came out. Nonetheless, a persuasive case for their impact value has been made in terms of the exploration of biological pathways and therapeutic hypotheses as a bridge between biological research and drug development (see 'Advancing Biological Understanding and Therapeutics Discovery with Small-Molecule Probes', 2015, PMID 26046436). The total number currently indexed in PubChem is shown in Fig. 11.

Fig. 11. Indexing of 243 NIH probes in PubChem. Note the right-hand facets are personal filters (e.g. matches to 26 GtoPdb entries). Controversy over this total still lingers because of the challenges of tracking down a consensus full set of that could be well above 300 (see 'Parallel worlds of public and commercial bioactive chemistry data' PMID 25415348).

The second probe phase refers to those generated outside the MLPCN network approximately post-2014 ('The promise and peril of chemical probes', 2015, PMID 26196764). Most recently a number of pharmaceutical companies have made available high-quality probes including associated data, control compounds and recommendations on use ('Donated

chemical probes for open science' PMID 29676732). This latter article includes the core probe criteria as shown below in Fig. 12.

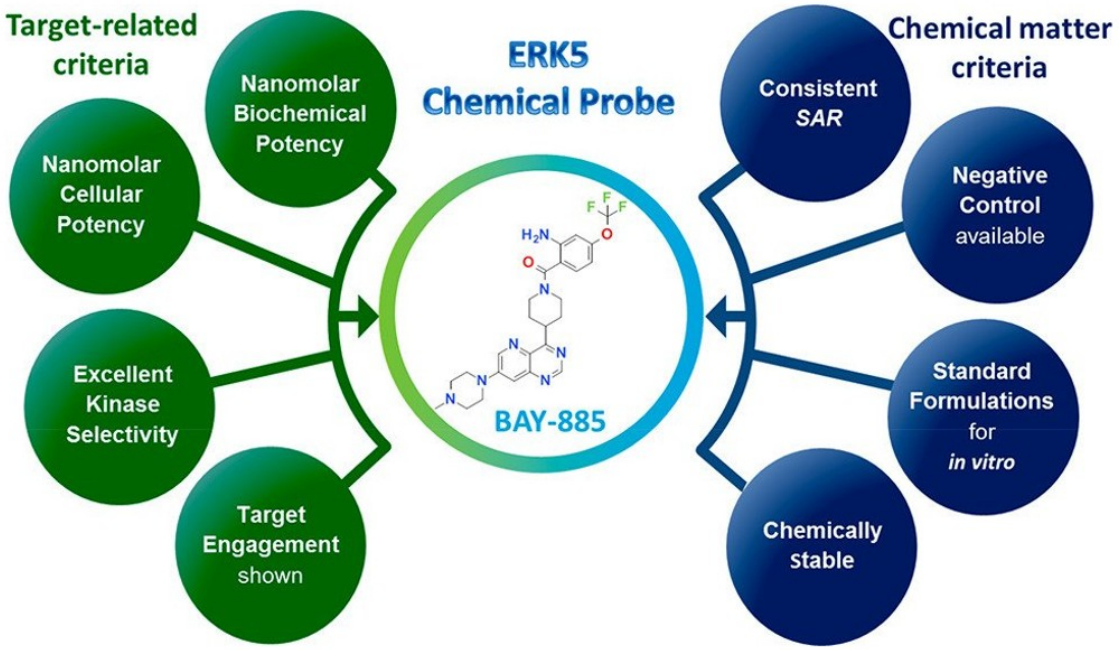


Fig 12. Outline of probe criteria as applied by the Structural Genomics Consortium (SGC) for the example of BAY-885, GtoPdb ligand ID 10183 (adapted from recent tweet).

While acknowledging the conceptual divergence between probes and drugs, ADEV has much to gain from an considering the pragmatic guideline established for the former. Two useful aspects would be firstly to adopt such guidelines during the later drug discovery stages *in vitro*. The second is that as coverage of probes expands, they provide more potential starting points for drug discovery and subsequent ADEV. Indeed, a publication on the large-scale prediction of probe-like molecules from ChEMBL indicated the chemical tools (most of which could not stringently be classified as probes) were available against 2,220 human targets ('Objective, Quantitative, Data-Driven Assessment of Chemical Probes', 2018, PMID 29249694). This publication includes a useful set of icons to illustrate the scores associated with chemical probe criteria (Fig 13).

Icon Name	Description	Image
Target Selectivity	Denoted by a histogram icon, it shows whether a compound inhibiting this protein is screened against at least one other target and has at least 10-fold selectivity against any other target	
Target Potency	Denoted by a test tube icon, it shows whether a compound inhibits this target with at least 100 nM potency	
Cell Potency	Denoted by a cell, it shows whether a compound binding to the target of interest is active in a cell line with at least 10 μM potency	
Minimum Standard	Denoted by a star, it is an aggregate of the three previous scores (which themselves are independent from each other), indicating whether there are compounds inhibiting this target with minimum standards of target potency (pActivity ≥7), selectivity (at least one tested off-target and 10-fold selectivity against off-targets) and cell potency (activity below 10 μM in at least one cell line) simultaneously. It is a key icon showing whether a compound fulfilling these minimum-quality requirements is found in publicly available databases	
SAR	Denoted by a benzene ring with an 'R' group, it indicates that there is at least one compound binding to this target that has SAR as defined by the SAR Score (see above)	
Inactive Analog	Denoted by a barred benzene ring, it indicates that there is at least one inactive analog of the compound as defined by the Inactive Analog Score (see above)	
PAINS	Denoted by a NON-PAINS icon, it shows that there is at least one compound inhibiting this target that has no PAINS alerts as defined in the PAINS Score	

Fig 13. Chemical probe icon representations of quality criteria. These may vary depending on the specific protein targeted and note that the inactive control should be of the same chemotype (adapted from PMID 29249694).

However, as for the NIH probes there are also caveats with the 'second generation' disclosures. The largest is the absence of any single, clean and independently corroborated list in PubChem. This is exemplified by the observation that the Probes & Drugs database (Probes & Drugs portal: an interactive, open data resource for chemical biology', 2017, PMID 28753599) has no less than 11 probe sets harvested and updated from various (non-PubChem) sources including supplementary data from publications (Fig 14).

Probe compound sets	ALL	UNIQUE	DUPLICATES	ALL	UNIQUE	DUPLICATES	SOURCE PAGE	DESCRIPTION	LAST UPDATE
Bromodomains chemical toolbox	25	2	0	25	2	0	A chemical toolbox for the study of bromodomains and epigenetic signaling	"Here we present a set of 25 chemical probes, selective small molecule inhibitors, covering 29 human	04/2019
Chemical Probes.org	189	25	0	189	14	0	Chemical Probes.org	"This resource is a community driven wiki-like site that recommends appropriate chemical probes for biological	03/2019
Gray Laboratory Probes	39	12	0	39	11	0	Gray Laboratory Probes	"The Gray lab believes in an open-source drug discovery model and attempts to make all of their publishea	10/2018
MLP Probes	374	128	1	374	97	1	NIH/PubChem	"The Molecular Libraries Program (MLP), a component of the NIH Common Fund, offers public sector biomedical	02/2017
Nature Chemical Biology Probes	58	16	1	58	14	1	Nature Chemical Biology	"Nature Chemical Biology provides freely available summaries of the relevant chemical, in vitro, cellular and	02/2017
Open Science Probes	47	26	0	47	21	0	Open Science Probes	"Chemical probes are validated, biologically active small molecules modulating the properties of their target	04/2019
Probe Miner (suitable probes)	2659	2389	1	2608	2189	45	Probe Miner	"Probe Miner provides computational and statistical assessment of compounds in the medicinal chemistry	10/2018
Protein methyltransferases chemical toolbox	19	0	0	19	0	0	A chemical biology toolbox to study proteinmethyltransferases and epigenetic signaling	"This chemical probe collection and associated data form a resource for the study of methylation-mediated	04/2019
SGC Probes	67	10	0	67	9	0	Structural Genomics Consortium	"SGC Chemical Probes are small, drug-like molecules which meet these criteria: in vitro IC50 or Kd < 100 nM, >	04/2019
Tool Compound Set	515	235	2	514	221	3	J. Chem. Inf. Model.	"The set of 517 best-in-class tool compounds (407 extracted from ChEMBl + 100 from ChemicalProbes.org)	11/2017
opnMe Portal	29	15	0	29	11	0	Boehringer Ingelheim Open Innovation Portal	"opnMe is an Open Innovation portal providing access to the Boehringer Ingelheim molecule library for sharing	02/2019

Fig. 14. Probe sources compiled by Probes & Drugs https://www.probes-drugs.org/compoundsets.

Combinations of these sets can be merged and mapped into PubChem but it is important to note that Probes & Drugs does not also submit to PubChem which would add confounding circularity. An extent of confusion for probes is compounded by the merging of both the Advisory Board-scored compounds and historical non-probes (i.e. previously designated as probes but subsequently considered to be inferior and potentially misleading) in the set of 466 CIDs submitted to PubChem by the Chemical Probes Portal. Other problematic aspects are highlighted in an analysis of the 189 *in situ* entries from the Chemical Probe Portal (see 'Data-Driven Exploration of Selectivity and Off-Target Activities of Designated Chemical Probes', 2018, PMID 30249057). This indicated that only ~50% should still be considered as target-selective from activity data additional to that used during the assessments for portal inclusion. However, according to this report, for the other ~ 50%, the totality of public data casts doubt on the initial claims of selectivity. The persistence of problematic probes has been most recently highlighted in an experimental study (see 'Validation and Invalidation of Chemical Probes for the Human N-myristoyltransferases', 2019, PMID 31006618 and the 'In the Pipeline' commentary).

Tool compounds, repurposing candidates and vendor reagents

In the spectrum of bioactive chemistry, probes and drugs can be regarded as broad band but others are also relevant to ADEV because their criteria overlap (even if not practically amenable to strict definitions). Of these 'Tool compound' is perhaps the loosest classification being used for any chemistry with even a low level of perturbation specificity but still deemed useful for investigating biological systems. For example, spider-venom toxin peptides can be used to probe (used in this context as verb rather than a noun) neuronal ion channels and receptors (see 'ArachnoServer 3.0: an online resource for automated discovery, analysis and annotation of spider toxins', 2018, PMID 29069336).

For repurposing candidates, the complication is that they usually refer to pharmaceutical company compounds offered for repurposing because the clinical development was halted (i.e. failed) for their intended indication but they had a safe Phase I result profile. This is distinct from reports of approved drugs being successfully repurposed because (by definition) these candidates were unsuccessful in the first place. The pilot sets appeared around 2012, predominantly from AstraZeneca (see 'Challenges and recommendations for obtaining chemical structures of industry-provided repurposing candidates', 2013, PMID 23159359). More recently analogous compounds have surfaced from other companies ('Boehringer Ingelheim experiments with open-access chemical probes', 2017, PMID 29282374 and Open Me). Having been collated by Probes & Drugs, this latter set is now indicated in Fig 13. So, are industry repurposing candidates also probes? Pragmatically, they can certainly be used as such but some may not have passed the 10-fold selectivity threshold specified in Fig 14 (and this is anyway always subject to the caveat of how extensive the selectivity screening actually was) However, they have advanced beyond the probe stage into clinical testing.

The subject of vendor compound offerings might seem peripheral but from practical point of view they are important because it is certain that some of them will be used in experiments related to ADEV investigations where they can serve as reference compounds, negative controls or for comparative system perturbation. As a quality vendor of stock compounds (as opposed to large-scale brokerage operations) Tocris provides a good example of a target-centric specialist reagent supplier with 6190 links to PubChem CIDs, of which 1876 are also

in GtoPdb with outlinks (May 2019 numbers). Not also that the former recommends the latter as part of their probe selection guide (Fig 15).

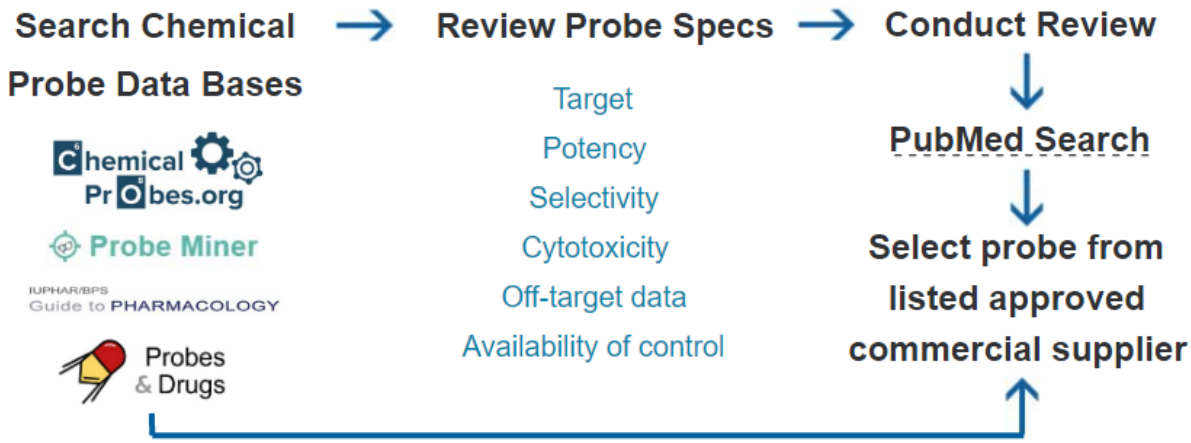


Fig.15. Guidelines for selecting a chemical probe. (adapted from the Tocris Guide to Probe Selection)

However, as already alluded to, there are caveats associated with probe classification and choice even with quality vendors. For example, from the 218 CID entries for the historical non-probes mentioned above 72 are listed by Tocris, 98 by GtoPdb and 49 by both. To be rigorous these are not explicitly endorsed as probes by either of these sources but the publications (which may have declared them as probes) have passed the threshold for curation in each case.

Journal guidelines

As the traditional repositories for data on pharmacology, medicinal chemistry, chemical biology and drug R&D, journal outputs clearly have a central role in ADEV. However, the PubChem BioAssay (AID) ratio of "Literature Extracted" to the rest (mostly screening centres) is 1043727: 23844. While this is only 2% in relative assay numbers many of these 2% are substantial screens encompassing millions of unpublished activity results linked to ~ 0.5 million CIDs in PubChem (but with the caveat that much of this is primary HTS data). It is also important to note that the erstwhile monopoly of journal papers as “primary data carriers” is further eroded by three other trends, a) the rapid proliferation of preprints (see ‘Has the Time Come for Preprints in Chemistry?’, 2017, 10.1021/acsomega.7b01190 and ‘Meta-Research: Tracking the popularity and outcomes of all bioRxiv preprints’, 2019, PMID 31017570) b) the emergence of open (web instantiated) electronic laboratory notebooks (ELNs) (see ‘Open notebook science can maximize impact for rare disease projects’, 2019, PMID 30689629 and ‘Experiences with a researcher-centric ELN’, 2015, PMID 29308130) and c) the coverage of at least the majority of exemplified structures extracted from full-text patents being openly available in PubChem that may both exceed and precede their publication in papers by several years (‘Expanding opportunities for mining bioactive chemistry from patents’, 2015, PMID 26194581).

While their status as gatekeepers of quality would be a moot point, there is nonetheless an assumption that by specifying stringent acceptance criteria for authors (i.e. journals setting

their individual bars) the higher standard ones could be considered as *de facto* gatekeepers of the essential core knowledge for CDEV and ADEV. However, an important caveat is that the terms 'lead structure', development candidate', 'clinical candidate', 'orally bioavailable', 'potent' and 'specific' are, in practice, inconsistent author-declared classifications with only implicit, rather than explicit, blessings of the journals. Leaving aside the contentious issue as to how journal quality can be defined (e.g. probably not by impact factor), we can use a useful proxy in the choice of papers selected by both the curation team and the IUPHAR target-class subcommittees for extraction into GtoPdb (see 'Challenges of Connecting Chemistry to Pharmacology: Perspectives from Curating the IUPHAR/BPS Guide to PHARMACOLOGY', 2018, PMID 30087946). The statistics are shown in Fig. 15.

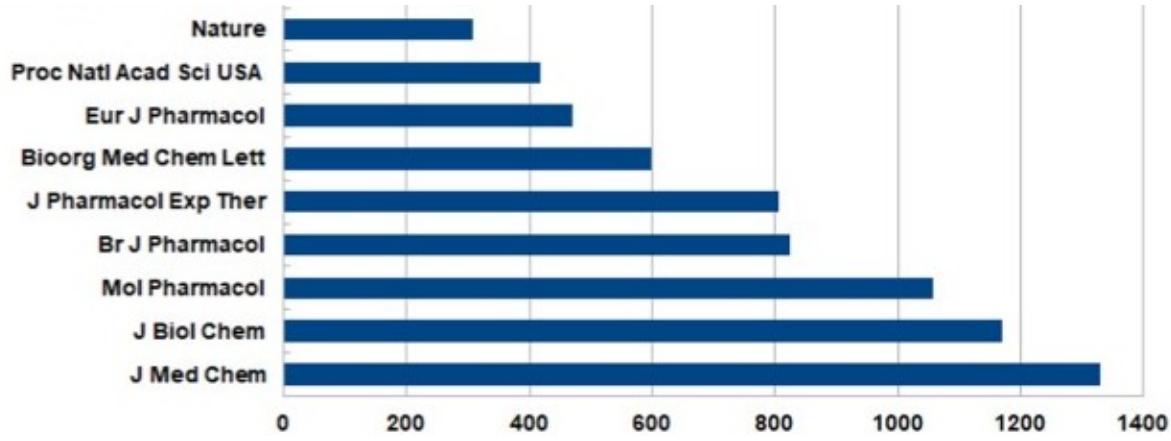


Fig. 16. Cumulative distribution of papers curated for ligand interactions in GtoPdb (release 2018.1). Numbers of papers are shown on the horizontal axis (adapted from PMID 30087946).

Of general relevance to this theme is a recent blogpost on chemistry and target specifications in pharmacology journals that collaborate with IUPHAR. What follows are submission criteria from a selection of ADEV-related from journals, including the top-four from Fig 15.

- The ACS Journal of Organic Chemistry (J Org Chem) has ~ 33,000 PubMed entries and a 2017 impact factor of 4.8. This is a gateway journal in the sense that while inherently low in bioactivity reports (i.e. having only one citation in GtoPdb as Ligand ID 9715 via PMID 23544738) many medicinal chemistry papers refer it for synthetic methods and it thus sets key criteria for authors. In particular, they are required to complete a Compound Characterization Checklist (CCC) that should be checked by Editors and referees. This includes data used to characterise the purity and identity of both new and known compounds whose preparation they describe. The criteria include the following
    - Compound, structure, or table-entry number
    - Weights and percentage yield
    - IR, UV-Vis, 1H NMR 13C NMR
    - Accurate mass (HRMS), optical rotation/ORD/CD
    - Enantiomeric/Diastereomeric ratio, X-ray (ORTEP and CIF)
    - 1H/13C NMR spectrum, chromatograms, Quant. GC, HPLC, electrophoresis
    - Cartesian coordinates or Z-matrix for crystals

- The ACS Journal of Medicinal Chemistry (J Med Chem) has 29K PubMed entries and a 2017 Impact Factor of 6.3. In terms of criteria the journal is unique in that authors are required to submit SMILES strings for molecules discussed in the manuscript along with the associated biochemical and biological data ('Digital chemistry in the Journal of Medicinal Chemistry', 2014, PMID 24521446). They recommend ChemDraw, ACD ChemSketch or Marvin Sketch to generate the SMILES. These are added to a spreadsheet template, along with basic compound information to provide a machine-readable version of the key data presented in the article's tables (n.b. while this is commendably forward thinking it is not clear if these CSV sheets submitted as Supporting Information are actually being machine-read by any operation!).

  The rest of the instructions are usefully detailed with criteria including;
  - Not unexpectedly, being both in the ACS stable, the CCC is identical to J. Org. Chem. with a recommendation of 95% purity
  - There is a section on examining active compounds for known classes of assay interference compounds with an analysis in the General Experimental section. They include (a house) reference for Pan Assay Interference Compounds (PAINS) ('The Ecstasy and Agony of Assay Interference Compounds', 2017, PMID 28244745)
  - Quantitative biological data for all tested compounds (this section goes into some detail)
  - Addressing statistical significance, variability and reliability of the results
  - Recommend IUPHAR target nomenclature

- Journal of Biological Chemistry (J Biol Chem) is published by ASBMB with just under 18K PubMed entries and a 2017 impact factor of 4.0. Surprisingly, despite its 2nd position in GtoPdb extractions, it has neither medicinal chemistry, drug discovery nor pharmacology in its topic index. Notwithstanding, the instructions include the following;
  - New compound structures and protocols for obtaining them must be provided
  - IUPAC nomenclature used for abbreviation of chemical names
  - For step-by-step protocols they encourage submissions to bio-protocol or protocols.io
  - Numerical data should be reported with the number of significant digits that corresponds to the magnitude of experimental uncertainty

- Molecular Pharmacology is an ASPET publication with ~ 11,500 PubMed entries and a 2017 impact factor of 3.9. The instructions pertaining to data are thinner than those for J Med Chem but relevant sections include the following;
  - The chemical identity or structure of novel drugs or research compounds must be provided (but there is no specification of how)
  - Purity to exceed 95% by accepted methods of mass spectrometry, chromatography, and NMR
  - Studies of extracts of natural products in unknown proportions or of unknown structure should not be submitted

- Only in rare circumstances (but not specified in these instructions) would compounds with potencies as low as 100's of micromolar EC50s be considered
- Statistical probability (p) should be expressed as *p < 0.05, **p < 0.01, and ***p < 0.001

- The Wiley/BPS British Journal of Pharmacology (Br J Pharmacol) has 24,000 PubMed entries and a 2018 impact factor of 6.8. There are several unique aspects (among this set of five) in their instructions for authors. The first of these, included in the Data and Statistical analysis section, is their own set of detailed guidelines published as an editorial ('Experimental design and analysis and their reporting II: updated and simplified guidance for authors and peer reviewers', 2018, PMID 29520785). The second is that key pharmacological ligands and targets (with IUPHAR nomenclature) are highlighted as GtoPdb links in the published article (see a GtoPdb 2017 blogpost 'GtoPdb NAR database issue 2018: Journal to database connectivity and journal to GtoPdb links'). Authors need to supply a 'Nomenclature of Targets and Ligands' statement and obtain intra-manuscript hyperlinking instructions post-acceptance. The third is the recommendation to use Research Resource Identifiers (RRID). These are ID numbers assigned key experimental resources such as antibodies, model organisms and software projects to improve transparency of research methods ('Incidences of problematic cell lines are lower in papers that use RRIDs to identify cell lines', 2019, PMID 30693867 and 'The Resource Identification Initiative: a cultural shift in publishing', 2015, PMID 27110440). The Br J Pharmacol has an extensive list of additional requirements a small selection from which is included below;

  - Chemical substances and drugs should be referred to by the generic name only. Trade names should not be used
  - Proprietary drugs referred by generic name, mentioning the proprietary name (this is unclear, the assumption could be IUPAC for the first and a code name for the second)
  - Supplier names and addresses for drugs and other chemicals
  - Synthesis and physicochemical characteristics of the new compounds summarised unless published in another journal or patent
  - Do not submit compounds of undisclosed structure or undefined mixtures such as plant extracts
  - List of Research Reporting Guidelines sources
  - Section on data storage, sharing and documentation including a reference to their publication ('Updating the guidelines for data transparency in the British Journal of Pharmacology - data sharing and the use of scatter plots instead of bar charts', 2017, PMID 28801996)

From quality to guidelines

Commendably, these four journals are clearly "on the case" in the interests of helping authors raise the quality of their submissions. However, even allowing for differences in journal scope, emphasis of affiliated organisations (e.g. ACS, ASPET, BPS, ASBMB) and the publishing house styles there are problematic aspects. The first of these is that coverage is patchy. For example, we might consider the following important enough to be adopted by all of them but they only appear explicitly in just one (e.g. a SMILES data sheet, structure sketcher recommendations, submission to protocols.io, use of RRIDs, potency limits, PAINs assessments and GtoPdb links). There is also a yawning gap in that not one of these five explicitly recommends the use of an open database identifier from PubChem or ChemSpider. Given that stable and unique compound URLs from these resources have been available since 2004 and 2008 respectively, both have been used in Wikipedia for some years and they have now reached 97 and 75 million entries respectively, this really does seems a major oversight (to be fair, BJP and BJCP do recommend GtoPdb ligand links that implicitly have stable unique database identifier URLs but the journals do not refer to explicitly to this attribute). There are also additional guidelines that would be valuable to add (e.g. 'FAIRsharing as a community approach to standards, repositories and policies' ,2019, PMID 30940948, 'STRENDA DB: enabling the validation and sharing of enzyme kinetics data', 2018, PMID 29498804 and 'Minimum information about a bioactive entity (MIABE)', 2011, PMID 21878981).

The second issue is succinctly quoted by the authors of 'Experimental design and analysis and their reporting II: updated and simplified guidance for authors and peer reviewers' 2018, PMID 29520785. This includes the quote "The main lesson learnt (from internal journal audit) that we may now share is that the guidelines that have been journal requirements since 2015 are not being routinely followed by authors and this is being missed during the peer review process. This 'non-compliance' is not unique to British Journal of Pharmacology (BJP) and is a phenomenon experienced by many other journals". This honest and important observation on the (understandable) shortcomings in author-editor-referee interactions relates back to figure 14 where even the top journals in GtoPdb represent only small proportion of the 10.7K references linked to the 9760 GtoPdb substance submissions (SIDs) in PubChem. This means there is a long tail of journals each of which have different criteria for authors and even those selected for GtoPdb curation may have compliance gaps.

The quality and reproducibility of PubMed reports in general and assessing this for bioactive compounds in particular, are major themes that cannot be addressed here in detail. However, they are crucial determinants of the data mining quality ADEV teams need to strive for. The major problem is (even if we acknowledge the quality standards of the journals referred to above) that these are by no means shared with all 29.7 million articles indexed in PubMed. Even adding the marginal stringency for those journals passing the MEDLINE acceptance criteria only reduces the total to 25.8 million (see 'MEDLINE, PubMed, and PMC (PubMed Central): How are they different?' NIH link). The problem becomes even larger because it is necessary to also be able to mine the millions of DOI-only bioscience papers that are also data-carriers.

The scale of this quality issue can be exemplified by even just a cursory examination of the 73.3K PubMed IDs retrieved with 'inhibitor' and 'drug" in the title or abstract (i.e. the same query plotted as the grey line in Fig.1). Even without detailed inspection the abstracts are very much a 'mixed bag' in terms of ADEV relevance and potential data value. While these simple queries have selectivity caveats, it was determined that only 504 of the 73.3K were included in the GtoPdb 10,798 selected citations. Another peculiarly credibility gap in the literate on lead compounds is related to the importance of PubChem BioAssay checking (as described below).

We can expand on some of the journal points above then can be distilled and harmonised into eventual IUPHAR Guidelines. With the exception of SMILES for J Med Chem there seems to be no clear mandates for the explicit specifications of chemical structures. The "holy hextet" of IUPAC names, SMILES, InChI strings, InChIKeys, SD/molfiles and rendering algorithms are show in Fig 16.

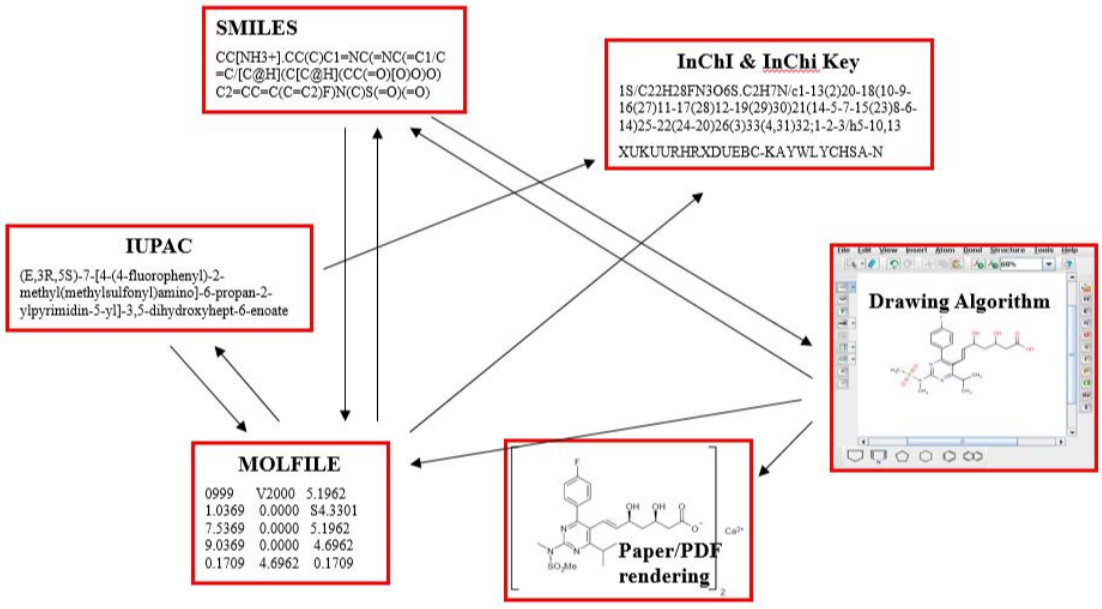


Fig 17. The six principle formal representation types for small molecules. Four of these, IUPAC names, SMILES, molefile or SD files and InChI strings are generally automatically interconvertible (but not the InChIKey). Note that a manual drawing tools (also termed sketchers) and rendering algorithms are related since they can both take as input (and usually also generate as output) either of the four interconvertible representations and subsequently convert these into screen renderings and/or PDF images of the structures (adapted from PMID: 30087946)

It is thus recommended that ADEV teams adopt the 'belt and braces' approach to specifying their structures in their internal documentation, even to the extent of storing all five representations in their local registration systems and databases.

The *In siloco* 360 (INS360) concept

INS360 is a set of guidelines (recommended here for the first time) for the selection (or equally importantly de-selection) and progression of ADEV compounds. The simple idea behind this is to generate a comprehensive informatics profile around the compound - target - disease axis that is central to the project (hence *in silico* 360). This would primarily be achieved using databases and prediction tools encompassing (but not restricted to) cheminformatics, bioinformatics, genomics, data mining and literature searching.

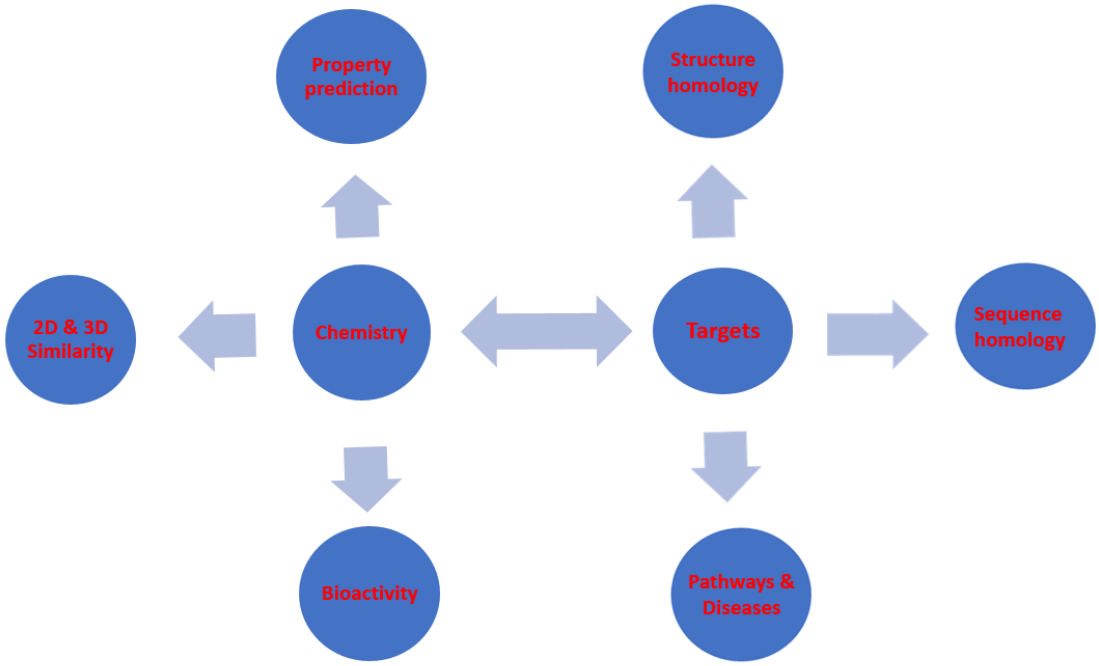


Fig.18 Diagrammatic representation of the INS360 concept. This is a simplified version in that more connectivity can be included, especially compounds of undefined molecular mechanism of action but with robust phenotypic assay signals and/or active against any type of infections disease organism.

The advantages of INS360 are

- The project team will have starting points already
- Being low-cost compared with lab work
- Faster execution than experimentation
- Presents decision-support checkpoints before further resource commitment
- It is hypothesis-neutral in the propensity to garner both good news on potential efficacy as well as bad news regarding chemistry-specific or target-specific liabilities
- Relevant data sources are expanding in content by the month
- Analyses can be pipelined and parallelised to support multiple projects
- AI/ML approaches are becoming applicable to some of the operations

- Much can be done with public databases but there are synergies with commercial sources
- The production of a “due diligence” package is essential to convince stakeholders, funders and potential collaborators
- The learning gains from this can change direction of a project and is immediately transferable to new ones

While there are no associated disadvantages, INS360 presents the challenge of effectively supporting an ADEV team (and the stakeholders thereof) with a broad range of advanced informatics that conform to at least some level of systematic best-practice. The imperative for this is generally underappreciated even when highly competent teams have impressive experimental track records by which they have achieved lead compounds with potent and selective *in vitro* target modulation. We can illustrate the challenges with some examples;

- Even experienced medicinal chemists may find the task of executing and interpreting the results of 2D, 3D and substructure searching the 98 million compounds in PubChem, the 150 million in SciFinder or the11 billion virtual enumerations in the Enamine REAL collection daunting, since the three resources are very different.
- Team members doing a good job on their in-house SAR assays may be unfamiliar with interrogating the 265 million results in PubChemBioassay as well as the 1 million results associated with the 1.9 million compounds in ChEMBL (these two sources also being very different).
- A disease biologist may be generally familiar with the UniProt entries for their protein target and pathway components. However, deep expertise is needed to interrogate resources such as Ensembl, Open Targets, and the genome Aggregation Database (gnomAD) to comprehensively explore genomic variation and possible disease association data relevant to their target including paralogues with cross-reactivity and model organism orthologues.
- Team members can doubtless execute keyword queries in PubMed and European PubMed Central (EPMC). However, they would find it difficult to fully exploit the advanced functionality of both (yet again very different portals) for navigating the deep connectivity between PubMed, PubChem, MeSH and Entrez, following out the automatic entity mark-ups and database links in EPMC or even going beyond these portals by using Natural Language Processing (NLP) text mining to extend relationship exploration (here again, despite good Help documentation on both sides of the Atlantic, there are no benchmarked best-practice guidelines).
- There are likely to have little experience either with commercial CI databases or grappling with the redundant, dense and confusing patent documents found in such sources.

What follows can only be a skim through selected sources recommended to be interrogated as part of an INS360 analysis. It must be born in mind the potential choices are large (i.e. daunting “multi-stop-shop”) some of which can, despite their value, also have non-obvious pitfalls (see ‘Caveat Usor: Assessing Differences between Major Chemistry Databases’, 2018, PMID 29451740). Since the analysis is likely to need informatics collaborators outside the immediate development team, they may of course have other sources and portal entry points they are particularly familiar with but would have similar coverage. It is convenient to divide sources into open and commercial, ordered approximately by their amount of data.

Chemistry and bioactivity

Given the context of this document it will be no great surprise that GtoPdb is a recommended 'first-stop-shop' for ADEV information gathering. As a relatively small database it nonetheless has a dense content of quantitative target-ligand molecular mechanisms of action (mmoa) via expert curation of journal papers (see 'The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY', 2018, PMID 29149325, the release notes, content statistics and Fig 18). Entry points for queries have been described in a recent how-to guide (see 'Accessing Expert-Curated Pharmacological Data in the IUPHAR/BPS Guide to PHARMACOLOGY', 2018, PMID 30040201). While name look-up of entries (including disease terms) can be used it is also possible bypass the naming equivocality problem entirely. This is done by sweeping through target sequence similarity space via the GtoPdb protein BLAST interface and ligand similarity space (for small molecules) via the Ligand search tool. With ~ 1700 human targets and ~9000 ligands (including ~7500 small-molecules) extracted from ~ 10,000 papers coverage is extensive (and includes selected probes). It should also be noted that target and ligand entries are linked to selected high-value sources. However, while human drug and clinical candidate coverage is high, not all recent lead compounds can be captured and (with the notable exception of antimalarials) coverage of antinfectives is not within the current funding mandate.

Moving from GtoPdb to ChEMBL provides a step-change up to 1.9 million compounds and 12K targets extracted from 72K papers (see 'ChEMBL: towards direct deposition of bioassay data', 2019, PMID 30398643 and Fig. 18). However, this is accompanied by a concomitant increase in navigational complexity (as can be perceived from this useful schema diagram). A key difference between GtoPdb and ChEMBL is that the former typically extracts just the lead compound from selected papers whereas the latter extracts an average of ~15 via contract curation with broad journal coverage. It should be noted that ChEMBL also imports about 0.5 million compounds from confirmatory PubChem BioAssays. While this extends the bioactivity coverage it also adds confounding circularity (see 'Redundancy in two major compound databases', 2018, PMID 29559364). Users should also note that the long release cycles result in ~2 year lag in literature extraction. Notwithstanding these minor disadvantages, ChEMBL offers a comprehensive and navigable resource for drug discovery and chemical biology. It is particularly good for discerning target "popularity". For example BACE1, as an Alzheimer's target, has no less than 8,741compounds in CHEMBL4822 with assay results against this enzyme. In contrast, GtoPdb target ID 2330 provides a selected core SAR set of 24 leads and (failed) drugs from which users can move out to the much larger SAR envelope in ChEMBL. Searching the latter thus has a high probability that ADEV compounds and targets will have at least similarity relationship matches.

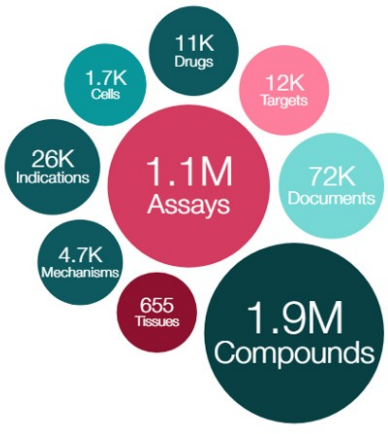
Moving up to PubChem provides not only another scale jump in compounds but also a concomitant expansion in relationship statistics (see 'PubChem 2019 update: improved access to chemical data', 2019, PMID 30371825 and Fig.19). This complexity (but also the navigational power) is magnified because of the embedding of PubChem within the NCBI Entrez network of ~ 20 interconnected databases (see 'Database resources of the National Center for Biotechnology Information', 2019, PMID 30395293). Notwithstanding the engagement challenges, there are many reasons why it is a must-stop-shop for ADEV, only a few of which can be pointed out here. Note also that PubChem and Entrez are cases where

collaboration with specialists familiar with their broad sweep of features is necessary for maximal exploitation. While UniChem (see PMID 29451740) is ~ 160 million (but cannot be similarity searched) and vendor virtual collections are now pushing towards the billions, PubChem remains the largest public source of just short of 100 million chemical structures that are (at least mostly) extant (see 'PubChem Sources by the numbers, 2016, blogpost).

There is a substantial contribution from ~65 million vendor-linked CIDs and, crucially from the prior-art aspect, ~23 million from automated patent extractions (largely from SureChEMBL and IBM) along with a smaller set from manual patent SAR extraction by BindingDB. The consequence is that, while this (free) coverage of exemplified compounds cannot be considered as complete, we could expect few chemical series from the global medicinal chemistry output to be without any similarity coverage in PubChem (see 'Managing expectations: assessment of chemistry databases generated by automated extraction of chemical structures from patents', 2015, PMID 26457120 but note coverage and indexing in PubChem has increased significantly since this paper). *Ipso facto* searches with ADEV lead series become essential and should be consolidated with substructure and superstructure searches.

Another key aspect of PubChem is the subsumation of many valuable individual collections within its 681 Data Sources. For example, GtoPdb ChEMBL, BindingDB and SureChEMBL are joined by DrugBank, DrugCentral, ChEBI and many other curated sources with ADEV utility. However, this presents a challenge in that, while complementary in aggregate (i.e. having different coverage and utilities), they each have a different look-and-feel as well as a divergent features on their stand-alone websites. Thus, a big advantage of these being merged within PubChem is that they can be searched with standardised parameters individually and/or compared (to any of the other 681 in fact). Source selection, property filtration, date cutting and Boolean intersecting can be done pre- or post-search. This means just about any desired slice-and-dice combinations can be tried out. The summary statistics of the three recommended sources are shown in Fig 19 along with a more detailed breakdown of PubChem content in Fig 20.

Category	Count in database
Number of curated binding constants	16948
Number of binding constants from large-scale screens	31207
Number of references	35878
Human targets with curated interactions	1766
Human targets with quantitative interactions	1507
Human targets with quantitative interactions with CIDs	1346
Human targets of approved drugs	631
Human primary targets of approved drugs	325
Approved drugs with quantitative interactions	907
Ligands with target interactions	8343
Ligands with quantitative interactions	7337
Ligands with CID and quantitative interactions	5795
Ligands with clincial use summary	2390
Approved drugs with clincial use summary	1423



Data Collection	Live Item Count
Compounds	98,189,235
Substances	245,449,015
BioAssays	1,340,520
Bioactivities	267,212,767
Gene Targets	58,029
Protein Targets	17,847
Taxonomy Targets	3,746
Literature	29,702,343
Patents	3,142,716

Fig. 19. Entity summary for (left to right) GtoPdb, ChEMBL and PubChem.

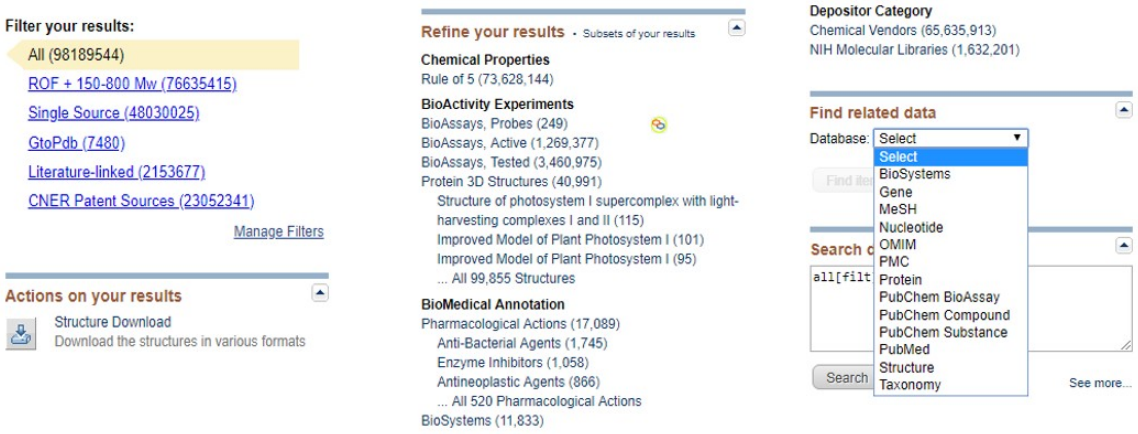


Fig. 20. Content statistics for PubChem with 98189544 CIDs. Note the example result from clicking on the 'BioAssay, Probes' link is shown in Fig.11. The left panel is a set of user-specified filters set up in MyNCBI. A category breakdown is shown on the middle panel and top-right. The lower-right shows those Entrez databases for which joins can be executed from any set of selected CIDs (but only up to 10,000).

After exploration of chemical space, the second essential cross-check is against public bioassay data. This can certainly be done in ChEMBL but in terms of efficiency PubChem can be queried directly. As can be seen in Fig 18, at 1.34 million assays this only has ~ 0.25 million that ChEMBL does not (most from MLPCN screening centres as described in the probe section above) because the latter is the by far the major submitter of activity results to the former. Notwithstanding, any matches between ADEV compound structures (including their similarity neighbours and scaffold matches) need careful inspection. This is not just to check for positive BioAssays against the indicated protein targets but also inactive test results. In addition, beyond the expected target, the data may indicate potential off-target liabilities, potential polypharmacology benefits or even repurposing indications. However, despite being an essential resource there are caveats associated with PubChem BioAssay a selection of which are outlined below:

- The count of 1.3 million "active" vs 3.4 million "tested" CIDs is very high, compared with a conventional HTS with assay hit-rates in the order of ~1%. There are several reasons for this but generous active thresholds (e.g. multi-uM) set by PubChem and their inclusion in ChEMBL extractions are the main ones
- Potencies consequently need to be checked for stringency and plausibility
- Special caution (i.e. a pinch salt) is needed for unconfirmed primary assay data
- Data is variable even extracted from papers. This means project-critical results will need at least independent support or ideally reproducible experimental verification)
- Even at over 75K papers the extraction of bioactivity by ChEMBL is only a fraction of published compound bioactivity results (the total of which remains unknown)
- As mentioned, between a curatorial backlog and long release cycles the results ChEMBL feed into BioAssay have at least a ~2yr capture lag

Three tips can be added for trawling PubChem for additional bioactivity clues that BioAssay may not contain. The first is to check if the structure may be reciprocally linked to a PubMed ID (in the right-hand facet of the abstract entry) wherein activity is reported but that paper was not captured by ChEMBL. The second is to inspect patent matches where filings may

include unpublished bioactivity data. The third is to look for similarity matches to the 28,017 PDBe CIDs derived from 87,104 protein structures. While this set has some heteroatom overcount compared to authentic binding ligands, only ~ 13K have BioAssay results. This means that for similar structures to any ligands there is a published target 3D interaction that may well track back to earlier bioactivity reports (n.b. co-crystallised metabolite similarity matches are also of high interest)

Protein targets

A 360-exploration of protein targets (i.e. extending out to literature, structure, evolution, genetic variation, alternative splicing, pathways, interaction data, disease mechanisms etc) is beyond the scope of this report. However, since ADEV teams, by definition, will be intensely focused on their portfolio over some years it is not only important to develop a deep perspective on their target biology and pathology but they also need to be actively alerted to changes triggered by new data from any angle (that could bring good or bad news). From their UniProt links, both GtoPdb and ChEMBL provide a convenient gateway into the protein informatics universe. The minor problem with PubChem is that, on their side of the Atlantic, targets, including those in BioAssay, are indexed throughout the Entrez system according to the NCBI Reference Sequence Database (RefSeq). This uses separate identifiers but the European and US systems at least cross-point. Another important aspect is that UniProtKB is composed of two sections, the expert reviewed and curated UniProtKB/Swiss-Prot section and UniProtKB/TrEMBL where the records are compiled via automatic annotation and classification. This is more significant than most ADEV teams may realise especially since there are discrepancies in the human proteins counts between sources (see 'Last rolls of the yoyo: Assessing the human canonical protein count'. 2017, PMID 28529709). As the Guide to Malaria Pharmacology team has found out, for ADEV teams working on tropical diseases, the lack of manual expert annotation on parasite protein drug targets becomes problematic (Jane Armstrong, personal communication). As an example, for the *Plasmodium falciparum* 3D7 reference proteome only 163 are Swiss-Prot reviewed while 5145 are still automated TrEMBL entries.

Literature

It is self-evidently essential for an ADEV team to explore the literature around their chemotype-target-pathway-disease-competition axis. However, there are a series of associated problems that need to be highlighted in these guidelines;

- Scientists with a project workload cannot be expected to keep up with the literature
- Even if they strive to do so, their searching methods may be a long way from best-practice (e.g. just popping keywords against PubMed has a dangerous level of false-negatives)
- Professional support for literature searching (as is generally available from Information Scientists in the larger pharma companies which extends to commercial database queries as well) is rare in an academic setting unless teams are fortunate enough to have library support
- Teams may not have awareness triggers (i.e. automated alerts) in place
- Access to large-scale text mining incorporating Natural Language Processing, NPL is rare (see 'iTextMine: integrated text-mining system for large-scale knowledge extraction from the literature' 2018, PMID 30576489)

- The three key public portals of PubMed, PMC and EPMC each offer advanced searching options, including powerful Entrez links for the first two and EBI links for EPMC. However, they all have a different-look-and-feel and the learning curves will get even longer as the NCBI roll out a new interface (see 'PubMed Labs: an experimental system for improving biomedical literature search', 2018, PMID 0239682)
- There are thousands of relevant DOI-only journal articles that are not indexed in the portals above
- As a Cinderella data source, many academics either do not search patents at all or have comparatively little experience in doing so (and may not have tried the LENS)
- As we know, the biomedical literature corpus has quality and reproducibility issues leading to search findings that can confound and mislead ADEV teams. This is clear from the example, mentioned above, of the 73.3K PubMed results with 'inhibitor' and 'drug'. Such problems have been described in recent reports (see 'Reproducible research practices, transparency, and open access data in the biomedical literature, 2015-2017', 2018, PMID 30457984)

The section above is hopefully useful for pointing out the challenges. Notwithstanding, it is not obvious which combinations can be realistically proposed as optimal solutions. At the very least, if the imperative for advanced searching message is loud and clear ADEV teams can seek training and/or institutional support. In this context it also is important to point out that an academic informatics community of practice already exists that encompasses most aspects of INS360 plus a lot more that are ADEV-relevant (e.g. omics technologies) This is ELIXIR that brings together European life science resources including databases, software tools, workflow guidance, training materials, SME and industry support, cloud storage and supercomputers with the goal of forming a single encompassing bioinformatics infrastructure. Crucially, the databases mentioned above (ChEMBL, GtoPdb, UniProt, EPMC and PDBe) are already part of this, either as Core or National Node resources (see 'Identifying ELIXIR Core Data Resources', 2016, PMID 27803796 and additional publications in F1000Research ELXIR gateway).

Commercial databases

The pharmaceutical and biotech sectors (encompassing the whole of CDEV) has spawned a substantial cottage industry producing a range of commercial databases that can be classified under the umbrella of "Competitive Intelligence" (CI). One form of this are pipeline databases offered by Citeline, Clavariate, Adis, and others. To these can be added larger-scale resources providing chemistry-to-document connectivity for patents and papers, extending from clinical trials back into the research phase, including SciFinder, Reaxys, Minesoft and Excelra (n.b. the latter has associated publications such as 'Analysis of in vitro bioactivity data extracted from drug discovery literature and patents: Ranking 1654 human protein targets by assayed compounds and molecular scaffolds', 2011, PMID 21569515). However, it is important to note that the content of such subscription products is drawn entirely from public primary sources, thus referring to these as 'proprietary, databases is thus somewhat misleading. Nonetheless, the data they contain is extracted, collated, curated and cross-indexed in various value-added ways, and typically integrated under a user-friendly front-end. They can thus be used effectively and efficiently to address the range of user questions directed towards "who's been doing what, when, why, with which structures, for what results and against which diseases".

These same general questions obviously need to be a crucial part of any ADEV INS360, While they can be certainly be partially addressed via the public chemistry and bioactivity resources described above (especially since the 'big bang' of patent extracted chemistry flowing into PubChem within the last three years) there is an element of complementarity with commercial alternatives. This was already described in 2009 (see 'Quantitative assessment of the expanding complementarity between public and commercial databases of bioactive compounds', PMID 20298516). When this was re-visited in 2015 (see 'Parallel worlds of public and commercial bioactive chemistry data' PMID 25415348) the complementarity still remained but accurate comparisons were precluded because commercial sources (understandably perhaps) ensure via licencing restrictions. However, this latter publication also made the point that the public data expansion had long passed the point where they must now be considered as essential to interrogate in parallel with commercial sources. Large pharma companies thus have the best of both worlds in being not only able to afford licences to a range of specialist CI products but are likely to have integrated these with public sources and their own internal data (see 'Making every SAR point count: the development of Chemistry Connect for the large-scale integration of structure and bioactivity data', 2011, PMID 22024215). The extensive subscription costs thus clearly put ADEV organisations (and SME's) at a considerable disadvantage for CI investigations (also termed due-diligence or freedom-to-operate)

It is not within the remit of this report to extol the features and coverage of commercial database products. This is done adequately via the company websites and publicity material but it is problematic that they rarely declare comparative entity intersect statistics with public sources (although as an unusual precedent this was permitted in PMID 22024215). However, CI products are important in the INS360 context and ADEV teams may have (or should contrive to get) access to at least some of them via their host institutions or collaborators. The licencing of SciFinder (SF) by many universities and research institutes means there is a high probability of team member access (perhaps even as their only available commercial database). The May 2019 announcement of their 150-millionth registered substance puts them ~52 million ahead of PubChem but still ~9 million behind UniChem. It is significant to note that SF increased by 50 million in four years compared to 34 million for PubChem. This suggests sustained global novel chemistry output that correlates with the sustained document growth in Fig 1 (and was confirmed via a different commercial database at least up to 2013 in PMID 24204758). One of the advantages of SF is the breadth of manually curated chemistry-to-document connectivity. This exceeds public sources by a considerable margin, even though the extent of this is unknown (n.b. arguably the scale of SF puts it beyond being classified just as a CI product since many scientists use it as their primary chemistry and literature gateway). Thus, searching a lead series in SF to provide a comprehensive interrogation of chemistry, papers and patents in parallel should be an INS360 core task.

Searchers can follow up by document-to-document 'walking' via chemistry-in-common, chemotype similarity, target-centric reports, authors/inventors and affiliations/assignees. However, like any database, SF has its shortcomings, some of which can be largely offset by public resources. The most significant of these for ADEV users are thin indexing of bioactivity and mmoa, including the absence of curated relationship between assay results and compound structures which are not numbered to allow facile cross-referencing to SAR tables in papers or patents.

Another CI product that has proven valuable in the academic as well as the commercial sectors is Clavariate Integrity. The daily-updated content stands at 457,000 compounds with documented biological activity and 297,000 patent family records. This has useful orthogonality to SF in the domains of commercial pipeline collations (but as for CiteLine also ADEV pipelines are not explicitly covered) target identification, mmoa and clinical data linked to associated diseases and therapies. The latter is particularly difficult to get cleanly from public sources (n.b. while the NLM ClinicalTrials.gov now indexes 307,479 research studies from 210 countries it does not substitute for a well-curated CI database). As an example of the importance of accessing Integrity (or any commercial offering with similar coverage) it is used to screen projects and targets being proposed for SciLifeLab DDD. It turns out that some of these proposals have been stopped because the teams were unaware of the extent of competitive impingement for their target and indication that was revealed by the Integrity queries (Per Arvidsson, personal communication).

To conclude this INS360 section the challenges presented by the informatics *de facto* multi-stop-shops need to be spelled out. As the cliché puts it "an hour of informatics can save 100 hours of experimentation". This means on the one hand ADEV teams have never had it so good for pursuing *in silico* approaches that could save a lot of lab time. On the other hand, it also means the necessary expertise needed to execute those comprehensively and efficiently continues to widen. We can take the powerful document-to-document 'walking' example mentioned above for SF. This can be done analogously across the ChEMBL, SureChEMBL, PubChem, PubMed, Entrez axis (and there is new orthogonal patent < > paper citation connectivity available at the LENS) . While public coverage is lower, this is balanced against the entire system being interlinked (i.e. we can click more or less from anywhere to anywhere). In addition, these public resources not only facilitate downloads but parts of them are also computationally accessible (e.g. GtoPdb web services, ChEMBL RDF , web services NCBI PUG and ELIXIR interoperability). Considering the essential (but expensive) complementarity of commercial products there are also integrated query options such Cortellis for Integrity and STN for SF. However, neither would supply handy tips for linking entities between the two. The problem for ADEV support is that few people have the expertise to exploit the full functionality between even just those resources mentioned above.

Competitive positioning

The execution of an INS360 in general and the inclusion of CI in particular, raises the following four issues;

- How the project intends to progress to a drug approval
- The partnership and funding model to enable this
- What type of commercialisation plan is envisaged for revenue generation
- Alternatively, they may choose the Open Source Drug Discovery route

These points cover a lot of detail but some expansion can be made. It is an inefficient use of resources with concomitantly diminished chances of commercial success if ADEV teams deliberately (as opposed to accidental CI oversight) go head-to-head against pharmaceutical companies. This certainly applies to the more crowded therapeutic areas where the available standard-of-care medicines (including generics) have achieved some level of effectiveness (i.e. the medical needs have been met to some extent). However, we can consider the following exceptions ;

- If the team can acquire compelling evidence for advances over existing drugs (see 'New drugs: where did we go wrong and what can we do better?, 2019, 366/bmj.l4340)
- Where the ADEV team can exploit some kind of unique scientific advantage for the progression of their compound. Examples might include synthesising a novel scaffold, targeting a previously unknow allosteric site, co-crystallising the target with a potent compound, pioneering a technically difficult screening assay or developing their own new disease model (e.g. a genetically engineered mouse strain or a CRISPR modified ES cell line)
- If they happen to have chosen a novel, or at least understudied, target that they are seeking to validate (e.g. in oncology)
- Stakeholders could (and arguably should) take the view that even if progression from preclinical to clinical is not successful (as is the case for the vast majority of commercial projects) the endeavour should not be seen as failure
- The output of quality FAIR publications and non-obfuscated patents would not only have an inherent payoff in an academic context by enhancing the credibility of the ADEV teams but also represent tangible scientific advances
- There will be equally important spin-off where published results and chemical structures (including flagging those that pass the probe criteria) flow into the databases mentioned including PubChem BioAssay
- There is also the benefit of novel analogues being added back into screening collections
- Publishing cases of data-supported target de-validation is a valuable community benefit

In terms of avoiding crowding by CDEV operations there are four areas to consider. The first three are Neglected Tropical Diseases (NTDs), antibiotics and orphan drugs for rare diseases. These three areas not only desperately need all the leads they can get (for obvious reasons) but also have low potential returns on investment (ROI) that make they less attractive to pharma (although to be fair they often support academic groups working in these areas in various ways). The third option is for ADEV to focus on areas that pharma has largely withdrawn from as being too challenging (i.e. little progress in biological insights, confounding genetics, poor translatability of disease models and many targets being eventually de-validated). This would include CNS research in general along with depression, schizophrenia and Alzheimer's disease (AD) in particular. Taking the examples of AD and neurodegeneration (i.e. extending to Parkinson's disease) a variety consortia have recently sprung up which either already have, or would welcome, ADEV engagements, some of which have declared a non-profit funding model (e.g. as listed in Appendix II) and include a personal initiative from Bill Gates.

Open vs closed

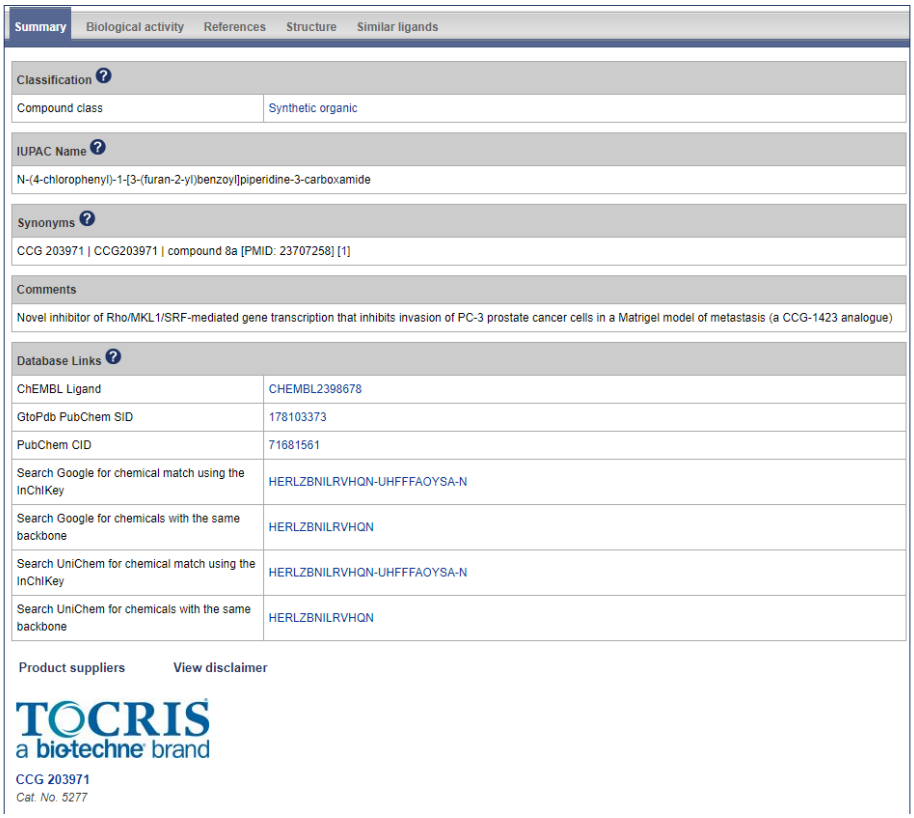
The choice is open to the team, management and stakeholders as to whether an ADEV project should work openly or closed. While the topic might have been contentious in the past, open science is now definitely established as part of the R&D landscape (see 'Open source drug discovery - a limited tutorial', 2014, PMID 23985301 and 'Is Open Science the Future of Drug Development?', 2017, PMID 28356902). Clearly this document would not recommend a position on this, one way or the other, but some aspects can be pointed out. Small-molecule antimalarial research has become the poster child for this way of working and, as an example, the Open Source Malaria initiative (OSM) has already made tangible progress (see "Open

Source Drug Discovery: Highly Potent Antimalarial Compounds Derived from the Tres Cantos Arylpyrroles", 2016, PMID 27800551, the In The Pipeline commentary and the ligand for OSM-S-38 in GtoMPdb). The key practical difference with open drug discovery is the deposition results as they are generated into an open Electronic Lab Notebook (ELN). This then gets Google-crawled and InChIKey entries can be found via a simple Google search (see InChI in the wild: an assessment of InChIKey searching in Google', 2013, PMID 23399051). Consequently, the project chooses not to file patents which leads to the following advantages;

- CDEV needs a proprietary research phase for patent generation that can extend for several years. When this is obviated by open research the development times can be considerably shortened
- The team can garner varying levels of direct support (funded or voluntary) as well as crowdsourcing feedback from experts anywhere across the globe
- They are free from the tyranny of having to pursue novel and broad composition-of-matter chemistry for IP protection
- This means they can fully exploit chemotypes with similarities in the expanding extant bioactive chemical space of at least ~ 5 million structures (although with the caveat of possible infringement issues)
- Teams, host institutions and the community can reap the benefit of rapid, full and FAIR publications
- ELNs can enable data surfacing in real time

Case studies

It is instructive to look at two ADEV examples as a practical skim through selected INS360 aspects. The hook journal article for first is 'Pharmacokinetic optimitzation of CCG-203971: Novel inhibitors of the Rho/MRTF/SRF transcriptional pathway as potential antifibrotic therapeutics for systemic scleroderma', 2017, PMID 28285914. The first reason for choosing this was that CCG-203971 was curated into GtoPdb as Ligand 6763 (show in Fig. 21) and the earlier analogue CCG-1423 that 3971was optimised from is Ligand ID 6761.



Fig, 22 A section from the GtoPdb entry for CCG-203971. Note links to ChEMBL, PubChem and in this case the vendor Tocris. Clicking on the InChIKey entries launches searches. The entry was curated from PMID 23707258.

A second reason was that Professor Rick Neubig, as senior author on PMID 28285914, was a past Core Member of NC-IUPHAR and thus alerted GtoPdb to his earlier work that consequently picked up the entries above. Together with his colleague Scott Laursen the two of them have founded FibrosIX Inc as a start-up based on licensable assay technology assigned to The University of Michigan and Michigan State University. The company aims to develop novel anti-fibrotic agents which disrupt gene transcription (see this project plan poster). They have been awarded an SBIR grant from the National Cancer Institute of NIH to advance the development of CCG-257081 (n.b. do not overlook Google for the INS360 because this code name has a string of Google hits but no n2s in PubChem or PubMed). It should also be noted that FibrosIX Inc cannot be found as a patent assignee or PubMed (i.e. the Neubig patents are assigned to The Regents and Trustees of the University of Michigan). This is not to imply deliberate obfuscation but it emphasises the point made earlier that the academic origins, commercial affiliations and molecular outputs of ADEV teams can be difficult to track and connect up. Somewhat unusually the 2013 paper curated by GtoPdb actually precedes the 2016 WO patent (i.e. CCG-203971 was already prior art).

This type of ADEV model is not uncommon in US academic departments. The milestones declared by FibrosIX are (by implication) guidelines and commercialisation plans proposed by two leading academic pharmacologists. These include;

- Anti-fibrotic efficacy in TSK-1 mice,
- Salt selection based on physical properties
- *In vitro* ADME profiling

- Oral PK in mice
- 14-day multiple dose tolerance in mice,
- Anti-fibrotic efficacy in bleomycin-treated mice at three doses
- Oral PK in rats
- Non-GLP tox in rats
- Anti-fibrotic efficacy in TSK-1 mice.
- Application for Orphan Drug Status and Fast Track Designation
- Licencing or acquisition opportunities

Continuing the analysis of compounds implicitly associated with FibrosIX we can find a lot of information associated with 3971 in PubChem CID 71681561. For example, it indexes isomers of 3971 not mentioned in the paper. However, the R and S isomers in this case are virtual enumerations from the vendor aggregator (ZINC15) so their status as pharmacologically-relevant entities is unclear. The Nuebig team recently published a target deconvolution paper (see 'Identification of Pirin as a Molecular Target of the CCG-1423/CCG-203971 Series of Antifibrotic and Antimetastatic Compounds', 2019, Doi/10.1021/acsptsci.8b00048). The informatics gotcha here is that the paper cannot currently be found in PubMed because, as a new journal, *ACS Pharmacol. Transl. Sci.* is PubMed indexed only for two open access papers. Notwithstanding, the report declares an n2s for the FibrosIX lead compound CGC-257081 which maps to PubChem CID 121317995 but only the structure (i.e. not a n2s). Unsurprisingly, along with CCG-203971, SureChEMBL establishes this also originated from WO2016073847 as example 41. What was surprising was the Google findability of CGC-257081 n2s as a vendor listing (but, not unexpectedly, offered via custom synthesis as opposed to a stock item).

While relatively few papers have been published on Pirin function, as a core part of the INS360 assessment there are over ~50 cross-references in Swiss-Prot O00625. A skim from these is highlighted below;

- The HGNC entry for the approved human gene symbol PIR with pointers to eight orthologues supported by the Vertebrate Gene Nomenclature Committee (VGNC)
- Gene Ontology (GO) assignments for metal ion binding, quercetin 2,3-dioxygenase activity and transcription coregulator
- 11 PDB entries
- A target link to CHEMBL2010627
- The Open Targets link (a meta-portal including genetics, disease association and expression evidence)
- Evolutionary history in Ensembl GeneTree
- Human Protein Atlas (HPA) antibody and transcript tissue distribution
- The InterPro domain mark-up (shown in Fig.23)

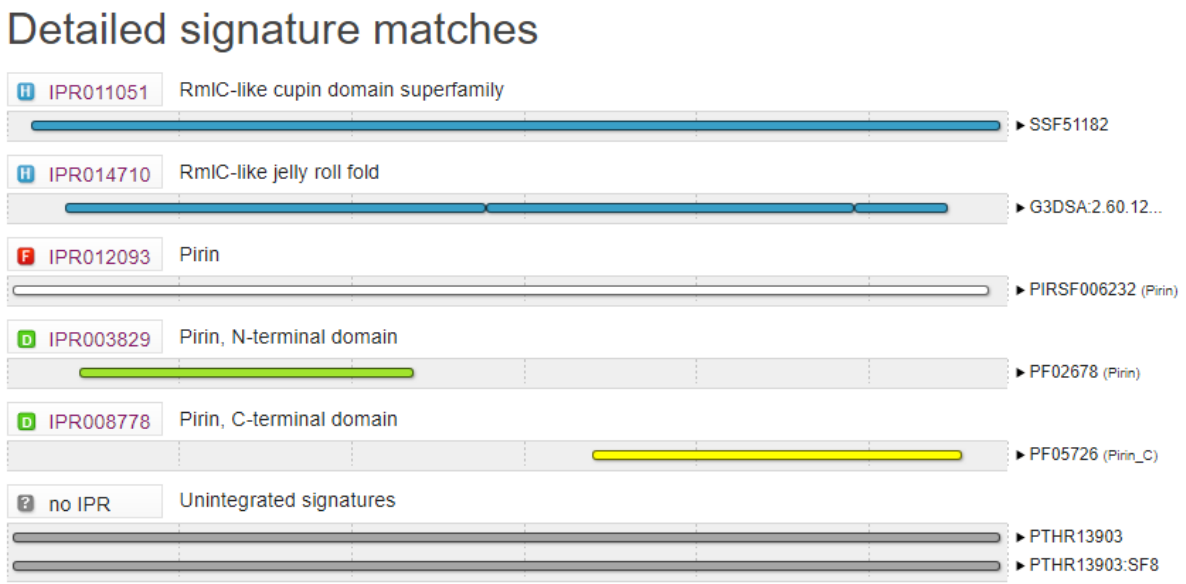


Fig. 23. InterPro matches for Pirin Swiss-Prot O00625.

The CHEMBL2010627 points to a key publication ('Discovery of a Chemical Probe Bisamide (CCT251236): An Orally Bioavailable Efficacious Pirin Ligand from a Heat Shock Transcription Factor 1 (HSF1) Phenotypic Screen', 2017, PMID 28004573). The probe structure is PubChem CID 117996749. Note that this paper was also cited as reference 28 in Doi/10.1021/acsptsci.8b00048 and the similarity between this and the Neubig series was noted. From PubChem > SureChEMBL we find example 63 from WO2015049535 ('Fused 1,4-dihydrodioxin derivatives as inhibitors of heat shock transcription factor 1' assigned to Cancer Research Technology Ltd. UK.). Notwithstanding, this chemical probe also has passed initial development criteria in being orally bioavailable and displaying efficacy in a human ovarian carcinoma xenograft model.

The CCG-1423 structure (via SureChEMBL > PubChem CID 2726015) links back to US20090264533 assigned to the Regents of the University of Michigan (not the Trustees this time). However, the first public n2s appears in a 2007 paper from the Neubig group (PMID 17699722) that apparently precedes the US patent. However, the public connectivity produces another gotcha in that SureChEMBL automated structure extractions from WO documents with poor OCR can be patchy. It was therefore necessary to cross-check the INPADOC entry to establish the patent-family connection to WO2006104751 'Methods and compositions for modulating rho-mediated gene transcription' (with yet a different assignee of just the University of Michigan) despite there being no PubChem chemistry link.

Checking these key structures in SF (via SMILES paste-across from PubChem) shows informative differences to the public domain. As expected, connectivity was more extensive with 24 documents for CCG-1423 (= RN 285986-88-1 for racemate but no documents for R and S), 4 for CCG-203971 (= RN 1443437-74-8), 2 for CCG-257081 (= RN 1922098-90-5), and 3 for the chemical probe CCT251236 (= RN 1693731-40-6). Some of this document connectivity would be difficult to track in in open sources. However, others are spurious in terms of bioactivity, for example opportunistic mentions of CCG-1423 in low-relevance patents (although the same problem occurs if automated extractions submitting to PubChem happen to have CCG-1423 in their n2s look-up dictionary). However, there are no n2s directly associated with the substance identifiers.

The second ADEV paper was chosen to exemplify different aspects to the Neubig series and is also the subject of a recent blogpost. It was published recently from a consortium of no less than 77 authors supported by six non-commercial grants centred on the University of Dundee ('Lysyl-tRNA synthetase as a drug target in malaria and cryptosporidiosis', 2019, PMID 30894487). It turns out this same malarial enzyme had already been curated by the Guide to Malaria Pharmacology (target 3059) for cladosporin (ligand ID 10247).

The lead highlighted in PMID 30894487 as compound 5 has an EC50 of 270 nM against *P. falciparum* 3D7 and is PubChem CID 134158252. Two sections from the entry are shown below.

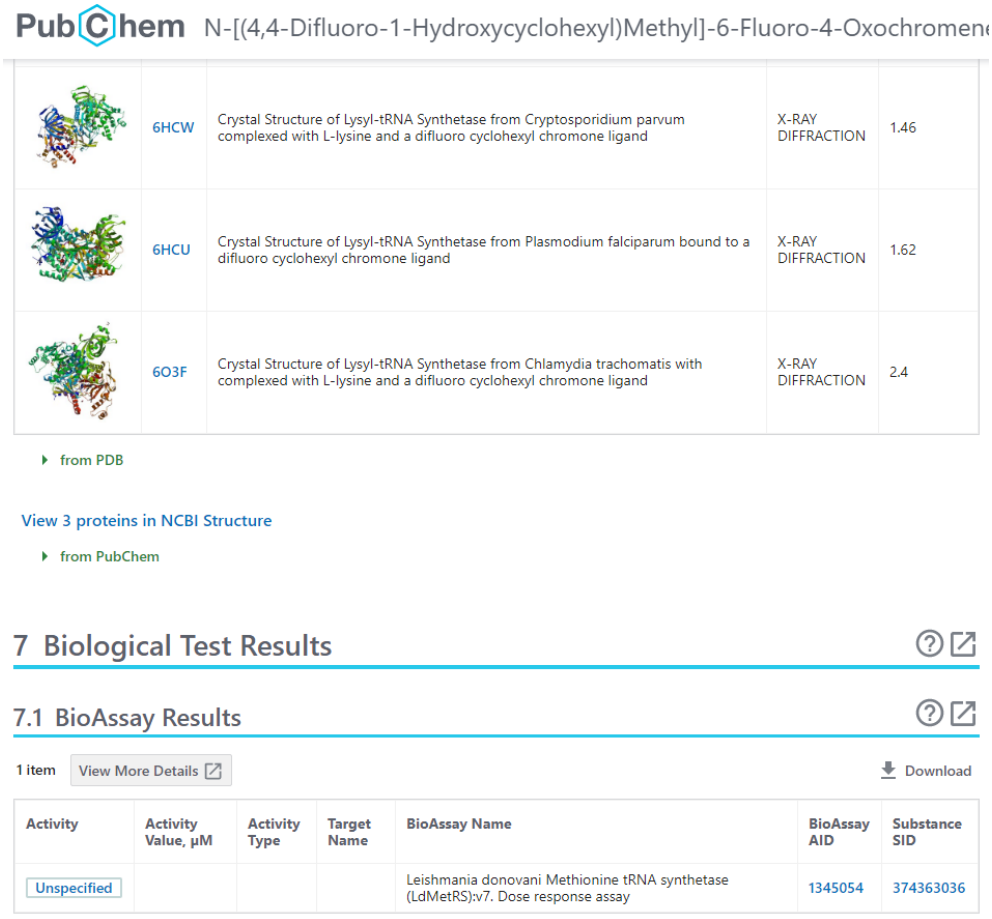


Figure 24. The three PDB entries and one BioAssay entry from CID 134158252**.**

Similar to the Nuebig CGC-257081 case from above the speed of vendor availability is surprising. In this case a lead published in April was already submitted to PubChem (SID 384293401) by the first vendor at the end of June. It should also be noted that the likelihood of these particular vendors of having licenced the sale of both these compounds from the academic patent assignee could be low.

This multi-parasite inhibitor was crystallised with the enzyme from three different species where 6HCU was the Plasmodium target. However, while the ChEMBL assay result (lower section of Fig 23) is unfortunately a dead link, a Google search finds an earlier Dundee paper ('Chemical Validation of Methionyl-tRNA Synthetase as a Druggable Target in Leishmania

donovani',2017). From this we can establish that, although it was a failed lead DDD806905 (CID 135423081) actually mapped back to WO2000071524 an old SmithKline Beacham compound directed against *S aureus* methionyl t-RNA synthetase (the minimum parsimonious assumption at this stage is that ChEMBL has redacted the erroneous assay cross reference but the deadlink remains in PubChem.

It can be established from PubChem and the patent office portals that Dundee University typically file patents on their NTD leads. It thus seemed odd that CID 134158252 had no direct patent connection in PubChem. However, it did have an indirect one. This was found via pointers to 30 CID analogues listed as 'Related Compounds with Annotation'(this is a 90% Tanimoto shell). Checking CID 132257258 gave a link to WO2017221002 'Anti-infective agents' assigned to the University of Dundee. It turns out that compound 5 is in fact example 54 as shown in Fig 25 but SureChEMBL had failed to extract this particular structure.

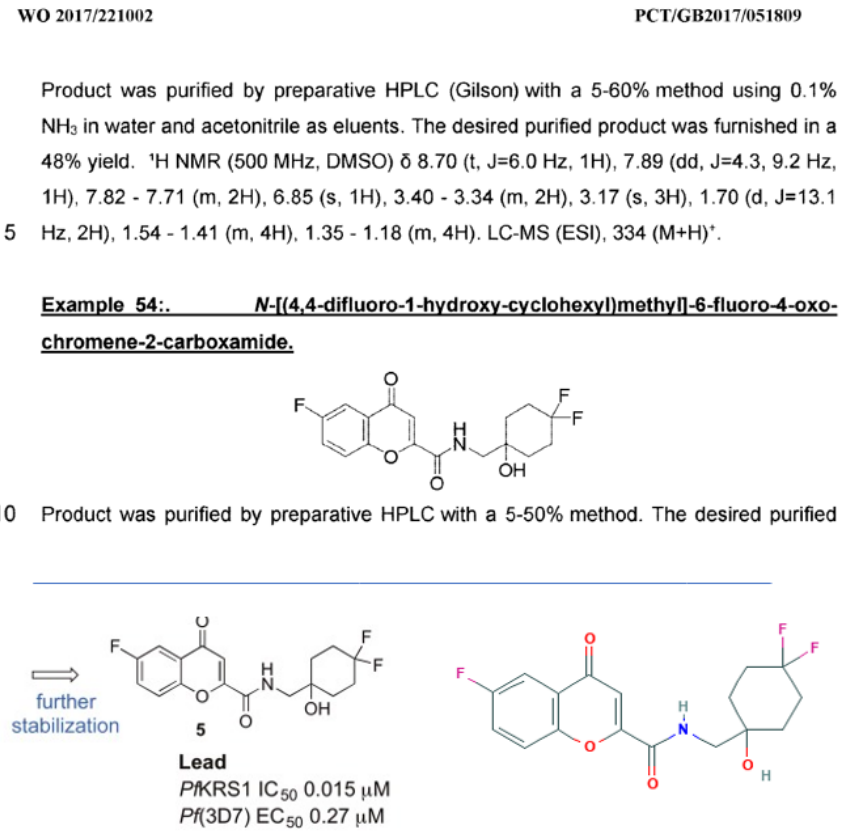


Fig 25. The identity between example 54 from WO2017221002 (upper panel) and (lower panel) compound 5 from Fig 1 in PMID 30894487 and the rendering from CID 134158252.

The team's practical choices of ADEV guidelines (in this case an antiparasitic for human use) are detailed in the Table 2 results from the paper;

- P. falciparum 3D7 EC50
- P. vivax liver schizonts/hypnozoites EC50
- P. berghei liver schizonts EC50
- P. falciparum stage V gametocytes EC50
- P. falciparum male/female gamete formation EC50 >1 μM
- Fasted state media (FaSSIF) solubility
- Microsomal stability
- Hepatocyte stability, human and mouse
- CYP inhibition
- Mouse PK intravenously
- Mouse PK orally

So far so good, but the paper contains the intriguing concluding sentence "We have identified a valuable tool compound, although further optimization is required in terms of both potency and selectivity to obtain a preclinical candidate" (i.e. not compound 5 so the exact results in Table 2 become less relevant). Since there are no published patents from Dundee claiming another series of Lysyl-tRNA synthetase inhibitors we can assume that this clinical candidate is already disclosed in WO2017221002 (but obviously not named as such). While we can only guess the likely candidate the most potent of the approximately 90 compounds with EC50 results against Pf33D7 parasite in the 10002 patent was example 93 at 8nM (compared to 320 nM for example 54). Example 93 was successfully extracted by SureChEMBL as CID 132256792. Cross checking against SF revealed a paradox. CID 134158252 (=RN 2170696-76-9) was only indexed against WO2017221002 but SF had not extracted compound 5 from PMID 30894487 from the eight substances they did curate. For the record this Dundee effort paper has been declared as the 2018 MMV Project of the year.

These two specific ADEV examples give guideline insights as well as demonstrating the utility of the INS360 approach. Importantly, they also show that both public and commercial sources are necessary to provide a complete picture.

Developments that could increase success rates

Despite the perennially gloomy pharma legacy statistics highlighted at the beginning of this report, an atmosphere of optimism can be perceived in both the academic and commercial drug R&D sectors over approximately the last 12 months. This is based on a range of scientific and technological advances that are plausibly predicted to increase the likelihood of success. Some of these (e.g. including all the "omics" options) are now offered as specialist services by companies included in Appendix II. While an evidence package based on more conventional technologies may be deemed acceptable to justify progression (e.g. for those compounds already selected and optimised) these newer orthogonal approaches may be perceived either as nice-to-have extras or even, from a more sceptical outlook, merely as passing fads in their hype-curve ascendance phases. In addition, some may be considered more relevant to early-R rather than late-D. However, the increasing importance of their utility will not only support more robust data packages for current decision-support but, collectively, they also will dictate how bright the futures of both CDEV and ADEV are going to be. There is a wide selection of such advances to choose from but only a skim introduction to some can be given below;

- Artificial Intelligence and Machine Learning (AI/ML). There is a burgeoning choice of literature but 'Can we accelerate medicinal chemistry by augmenting the chemist with Big Data and artificial intelligence? (PMID 29577971) and 'Exploiting machine learning for end-to-end drug discovery and development' (PMID 31000803) will serve as introductions.

- Functional genomics. As most will know CRISPR/Cas9 has initiated an impressive array of innovations in screening for drug target identification and validation. As a practical example this has already been introduced in the SciLifeLab High Throughput Genome Engineering (HTGE) facility. Pooled CRISPR/Cas9 screening enables parallel interrogation of up to tens of thousands of genes for involvement in biological processes. Guide libraries are used for whole genome and specifically targeted loss- and gain of function studies (CRISPR knock-out, CRISPR inhibition, CRISPR activation). They also offer generation of stable Cas9-expressing lines in users' cells of interest. CROPSeq and CRISPR-X approaches are in development.

- Target expansion. Increases in the fraction of the human proteome having tractable chemistry starting points for modulation (i.e. the data-supported druggable proteome) has been incremental. However, target classes have generally expanded, driven in part by probe expansion (see 'Target 2035: probing the human proteome', 2019, PMID 31278990) and establishing platforms that can be used to navigate understudied proteins (see 'Exploring the dark genome: implications for precision medicine', 2019, PMID 31270560). Further expansion of drugggablity is also being driven by proteolysis-targeting chimeras, PROTACs (see 'Targeted protein degradation: elements of PROTAC design', 2019, PMID 31004963). However, one overlooked aspect of small-molecule modulation is that inhibitors and antagonists remain in considerable excess over activators and agonists whereas the latter are more likely to yield a mechanistic fix for the loss-of-function that is causatively predominant for disease.

- Target engagement. As indicated in the section associated with Fig.4 , lack of efficacy as a major cause of clinical failure is often associated with insufficient target engagement. As one of the methods of measuring this the Cellular Thermal Shift Assay (CETSA) measures target engagement under physiological conditions. ed to generate modified cell lines. It can be applied against known targets of interest to confirm *in vitro* potency both in model systems and patient derived material. In addition, the approach applied proteome-wide for unknown targets (on and off targets) of a compound to be identified. Recent references on CETSA are given in this GtoPdb Hot Topics blogpost.

- Protein structure exploitation. The burgeoning field of computational chemistry, traditionally thought of as more relevant to the research phase, is increasingly important for ADEV optimisation and has already become an essential part of the INS360 process described above. The following references will give some idea of utility expansion. The EBI has launched a new resource called PDBe-KB (Knowledge Base) that enables researchers to visualise and easily access all publicly available structural data, including the 28,000 ligands and heteroatoms. The target-class with the largest recent structure expansion (boosted by cryo-EM) is clearly the GPCRs (see 'GPCRdb in 2018: adding GPCR structure models and ligands, PMID 29155946). It

should also be noted that protein modelling based on the 150,000 PDB structures is also expanding (see 'PhyreRisk: A Dynamic Web Application to Bridge Genomics, Proteomics and 3D Structural Data to Guide Interpretation of Human Genetic Variants', 2019, PMID 31075275). Last but not least, the predictive joining of the small-molecule and big molecule worlds is also moving forward (see 'Ultra-large library docking for discovering new chemotypes', 2019, PMID 30728502)

- Compound property prediction. This has a long tradition of modelling efforts in large pharma companies but, partly due to expanding open data sets, employing newer AI/ML approaches and access to more computing power, the scope of this is expanding. There is an explicit expectation that such improvements will reduce the need for ADEV teams to experimentally confirm *in silico* predictions but there are obvious caveats. A brief selection to show the directions of this are included

    - Pan Assay INterference compoundS (PAINS) and colloidal aggregators. The *ab intio* prediction of these is controversial but it is nonetheless essential to flag these up for lead compounds. As mentioned, a surfeit of positive results in PubChem BioAssay is a warning sign (see 'The Ecstasy and Agony of Assay Interference Compounds', 2017, PMID 28435522)
    - Toxicology/Safety Assessment. Whilst it may be some time before prediction displaces the experiments mandated by regulatory authorities, computational toxicology is now used in every phase of drug discovery and development (see 'Computational Toxicology and Drug Discovery', 2018, PMID 29934896)
    - ADME. Computational tools for metabolism prediction are being enhanced predict small molecules biological transformations ('BioTransformer: a comprehensive computational tool for small molecule metabolism prediction and metabolite identification', 2019, PMID 30612223)

- Disease association. Human genomics is the fastest accelerating data type of key relevance to ADEV particularly where it can support target validation and mechanistic molecular pathology for therapeutic intervention. The recent achievements of the UK Biobank provide a notable example (see 'The UK Biobank resource with deep phenotyping and genomic data', 2018, PMID 30305743). However, with ~140,000 genome wide association (GWAS) results for complex diseases from ~ 4,000 publications, the current GWAS Catalog presents an embarrassment of riches. Even when confirmed, most are outside known gene loci and have low levels of experimental orthogonal mechanistic deconvolution (see 'The support of human genetic evidence for approved drug indications', 2015, PMID 26121088, and 'Open Targets: a platform for therapeutic target identification and validation', 2017, PMID 27899665). However, consideration of just one major type 2 diabetes (T2D) study identified 33 putative functional genes, of which just three were targeted by approved drugs (see 'Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes', 2018, PMID 30054458). These explicit multi-target statistics (not unique to T2D) seriously questions the implicit ADEV assumption that single compounds can achieve sufficient efficacy. It thus presents the major challenges of deciding to push either for designed polypharmacology or drug combinations at the outset. Additional potential target opportunities are also being opened up from the ~7000 rare diseases (see 'Addressing challenges in the diagnosis and treatment of rare genetic diseases' 2018, PMID 29242613).

- Phenotypic screening in cells. This has undergone rapid advances in scale and resolution via automated image processing. It is also being extended to tissue chips, co-cultures and organoids (see ‘The Cell Painting Assay as a Screening Tool for the Discovery of Bioactivities in New Chemical Matte’, 2019, PMID 30519943 and ‘Concerns, challenges and promises of high-content analysis of 3D cellular models’, 2018, PMID 29977053). Other forms of phenotypical screening, including in vivo, may also be useful.

- Expanding the engagement to different countries (especially in the Global South). Building up research and ADEV capacity in resource-limited countries (RDLCs) will eventually generate valuable new data but needs the recruitment of more experimental and clinical pharmacologists. Discovery units such as SciLife at KI do encourage training and exchange of scientists in RDLCs. In addition, equitable deals need to be arranged for research on endogenous natural products where the Nagoya Protocol ensures redistribution of the fruits of success. Local ADEV is justly favoured under these conditions because of the consequent financial inducements for government support and training. This creates a feed-forward loop for R&D as well as healthcare in these countries. In this context, interoperability between resources for pharmacokinetics and toxicology between specialised ADEV centres (i.e. that can be shared between the Global North and South) is advocated by the Academic Drug Discovery Consortium (ADDC).

Compound and target fitness criteria

This final section presents a distillation from the coverage above (with some additions there was no room to detail) in the form of Table 1. This lists compound and target guidelines for ADEV teams and their collaborators.

Table 1. Compound and target fitness criteria (n.b. this is a paste-in from an Excel sheet that can be edited for succeeding versions)

1	Data stewardship	Ensure best practice for compound registraction system, ELNs and local SAR data storage
2	INF360	Assess outcomes and expanded knowledge
3	Disease model *in vivo*	Track record of human translatability for rodents or other
4	Disease model *in vitro*	Surrogate but relevant cell, organoid or tissue chip model
5	Efficacy	Evidence for disease amelioration, even in cell model
6	Biomarker	Biomarkers for pathway modulation and disease amelioration
7	CI assesment	Therapeutic need projection, stakeholder and funder alignment
8	IP	Does the team have composition of matter IP or are they working in the "open"
9	Popularity	Is the target competitively crowded (e.g. 1000s of cpds in ChEMBL)
10	Literature	Set up specific search profile alert for new papers and patents around target and disease
11	Chemotype	Search PubChem, BioAssay, SciFinder patents and papers
12	Validation	Evidence pakage for target validation
13	Moonlighting	Establish if the target has multiple funtional roles in different biological contexts
14	Genetics	Support for validation, efficacy, variation, nsSNP effect prediction, CNV variation, GWAS data
15	Structure	Anything in PDB at ~70% identity and/or 3D ligand similarity, structure prediction Phyre2
16	Crystalisation trials	Attempt target co-cyst with tight-binding ligand and crystalisation of lead cpd for 3D
17	Computational chemis	General support for VS consolidation, pharmacophore modeling, docking if relevant
18	Target engagemet	Evidence for thermodynamic binding and occupancy levels in lysates and whole cells
19	Potency	*In vitro* < 100nM *in cellulo* < 1 uM, < 10 mg/kg p.o. *in vivo*
20	Inactive controls	Low activity similar chemotypes (matched pairs if poss) for inclusion in most experiments
21	SAR	Could further optimisation bring major advantages
22	Mmoa	Data support for molecular mechanism of action (e.g. competive, mixed, allosteric, biased agonist, etc.
23	Pathway	Well curated and evidence underpinned normal and disease pathways identified
24	Phenotype	Pheonotypic screening in cells with image analysis
25	Omics profiling	Peuterbation patterns in cells via proteomics, transcriptomics and metabolomics
26	Paralogue specificty	Cross-screening results with extensive paralogue coverage, indications of off-target or repurposing
27	Orthologue specficity	Rodent orthologue squence identity and other models e.g. dog
28	Kinetics	Ki, Kon Koff and residence time for both human and rodent disease model proteins
29	Orthogonal assay chec	Confirm inhibtion kinetics in assay system with differnt read-outs than primary assay (e.g. SPR)
30	Chemistry robustness	Scale up, not too many steps, re-synthesis with activity re-tests, stablity in DMSO and < temperature
31	Purity	< 95% with analytical data consistant between batches and steroisomers resolved
32	Labeling	Availablöity of labled materical for metabolic studies
33	Chemical liability prec	PAINS, aggregation, reactivity, metabolism, HERG and P450 binding
34	Chemical liability test	Exeperimental exclusion or confirmation of strong predictions from above
35	Plasma stability and b	Human and rodent profiling plus *in vitro* metabolite full structure ID
36	Preclinical ADME	Bioavailability, pharmacokinetics, *in vivo* metabolite ID, activity testing of metabilites
37	Dosing studies	Dermination of dose ranges and therapeutic window
38	Predictive toxicology	Databases and prediction algorithms
39	Preclinical toxicology	Animal studies plus animal and human cells in culture
40	Pharmacogenomics	Assesing population variants (including non-European) in the target and metabolising enzymes
41	Drug combinations	Opportunities for therepeutic synergies or contra-indications

Conclusion

It is hoped this report will fulfil the utilities expected by its stakeholders but the implementation of enhancements and updates is of course envisaged. Following open publication on the ChemRxiv Preprint Server for Chemistry in July 2019 feedback from nterested parties is welcome. The three appendices listed in the outline are incorporated below.

Acknowledgments

Anders Gustafsson Academic Vice President for Research, and the Committee for Research at KI for a start-up grant awarded in the context of the KI-IUPHAR Memorandum of Understanding signed in February 2019 that enabled this work to be carried out by TW2Informatics.

Per Arvidsson, Science for Life Laboratory Drug Discovery and Development Platform and Division of Translational Medicine and Chemical Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet.

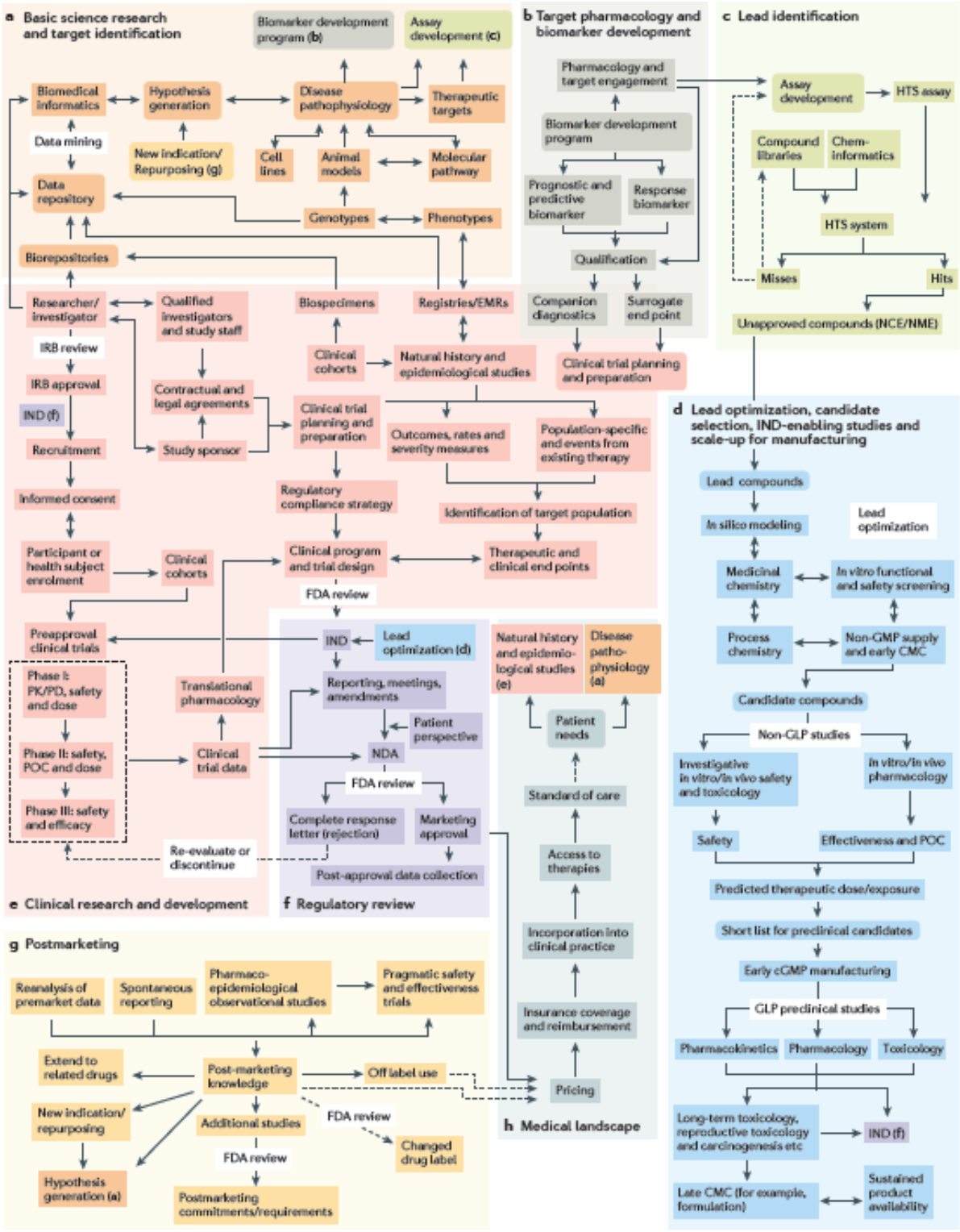
Michael Spedding IUPHAR Secretary General and Director of Spedding Research Solutions.

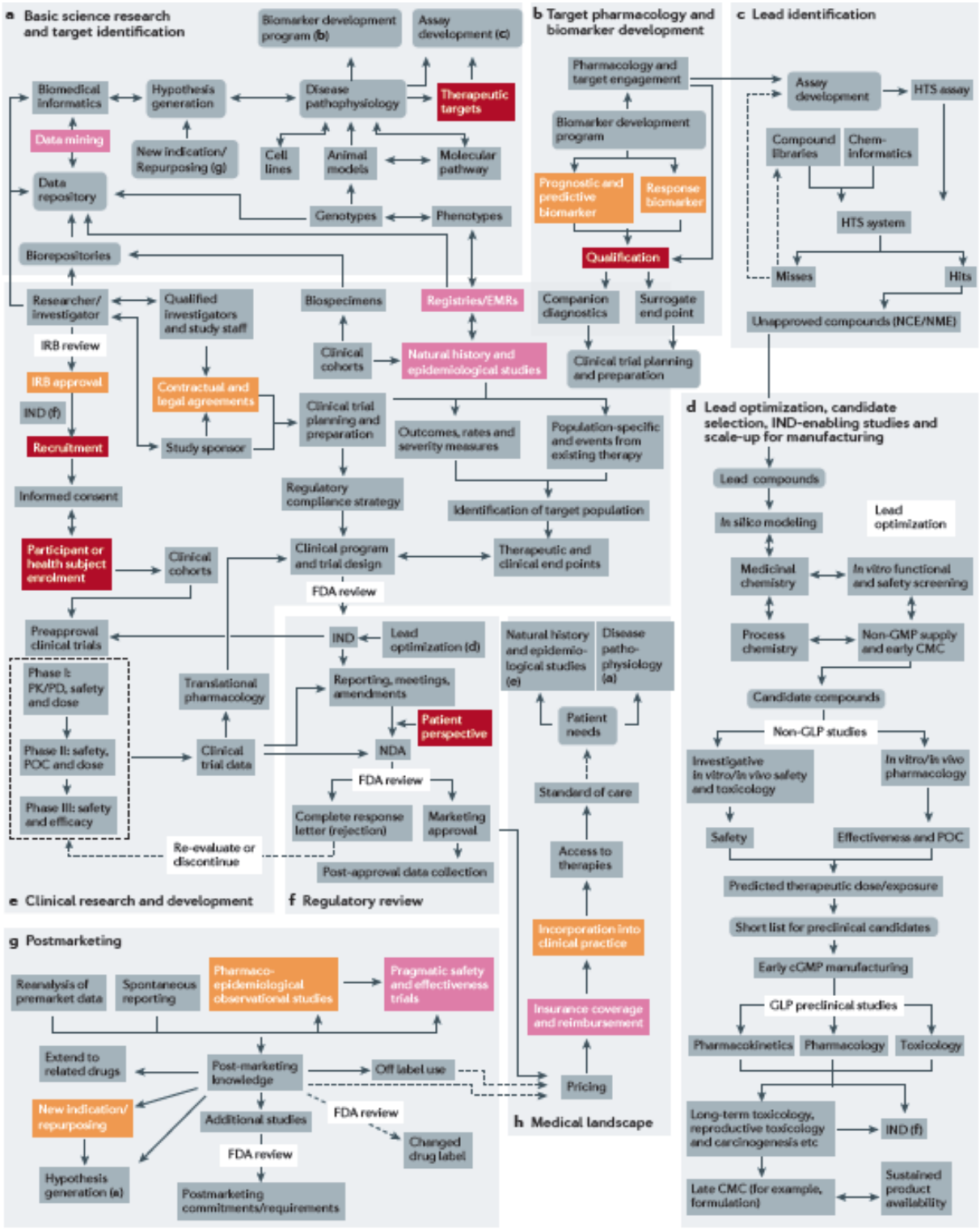
Gaelle Woestelandt of Spedding Solutions for preparing Appendix III.

Members of the Edinburgh IUPHAR/BPS Guide to Pharmacology (GtoPdb) team.

****

Appendix I. 4DM maps for small-molecule drug development (adapted from PMID 29269942). The upper figure shows colour-coded neighbourhoods. The lower figure shows s 'traffic map' version. Steps associated with the greatest cost, time or likelihood of failure are highlighted in dark red. Less problematic roadblocks are identified in pink and orange. All other steps are indicated in grey.




a Basic science research and target identification
Biomarker development program (b)
Assay development (c)
Biomedical informatics
Hypothesis generation
Disease pathophysiology
Therapeutic targets
Data mining
New indication/ Repurposing (g)
Cell lines
Animal models
Molecular pathway
Data repository
Genotypes
Phenotypes
Biorepositories
b Target pharmacology and biomarker development
Pharmacology and target engagement
Biomarker development program
Prognostic and predictive biomarker
Response biomarker
Qualification
Companion diagnostics
Surrogate end point
Clinical trial planning and preparation
c Lead identification
Assay development
HTS assay
Compound libraries
Chem-informatics
HTS system
Misses
Hits
Unapproved compounds (NCE/NME)
Researcher/ investigator
Qualified investigators and study staff
Biospecimens
Registries/EMRs
IRB review
Clinical cohorts
Natural history and epidemiological studies
IRB approval
Contractual and legal agreements
Clinical trial planning and preparation
IND (f)
Study sponsor
Outcomes, rates and severity measures
Population-specific and events from existing therapy
Recruitment
Regulatory compliance strategy
Informed consent
Identification of target population
Participant or health subject enrolment
Clinical cohorts
Clinical program and trial design
Therapeutic and clinical end points
FDA review
d Lead optimization, candidate selection, IND-enabling studies and scale-up for manufacturing
Lead compounds
In silico modeling
Lead optimization
Medicinal chemistry
In vitro functional and safety screening
Process chemistry
Non-GMP supply and early CMC
Candidate compounds
Non-GLP studies
Investigative in vitro/in vivo safety and toxicology
In vitro/in vivo pharmacology
Safety
Effectiveness and POC
Predicted therapeutic dose/exposure
Short list for preclinical candidates
Early cGMP manufacturing
GLP preclinical studies
Pharmacokinetics
Pharmacology
Toxicology
Long-term toxicology, reproductive toxicology and carcinogenesis etc
IND (f)
Late CMC (for example, formulation)
Sustained product availability
Preapproval clinical trials
IND
Lead optimization (d)
Natural history and epidemiological studies (e)
Disease pathophysiology (a)
Phase I: PK/PD, safety and dose
Translational pharmacology
Reporting, meetings, amendments
Patient perspective
Patient needs
Phase II: safety, POC and dose
Clinical trial data
NDA
Standard of care
Phase III: safety and efficacy
FDA review
Access to therapies
Re-evaluate or discontinue
Complete response letter (rejection)
Marketing approval
Incorporation into clinical practice
Post-approval data collection
Insurance coverage and reimbursement
e Clinical research and development
f Regulatory review
Pricing
h Medical landscape
g Postmarketing
Reanalysis of premarket data
Spontaneous reporting
Pharmaco-epidemiological observational studies
Pragmatic safety and effectiveness trials
Extend to related drugs
Post-marketing knowledge
Off label use
New indication/ repurposing
Additional studies
FDA review
Changed drug label
FDA review
Hypothesis generation (a)
Postmarketing commitments/requirements

Appendix II. References and other links related to KI/IUPHAR drug development report

Chris Southan, 12th July.

References https://www.zotero.org/cdsouthan/items/tag/drug_development This is an open reference collection tagged with "drug_development". The tag has not been applied particularly stringently in the bibliometric sense but has simply been used to conveniently index papers and other source material specifically for the main report. The web listing below is by posting date and suggestions for additions are welcome.

Title	Creator	Year	Date Added
Applications of machine learning in drug discovery and devel...	Vamathevan et al.	2019	7/10/2019 11:36 AM
Transforming cancer drug discovery with Big Data and AI	Workman et al.	2019	7/9/2019 3:38 PM
Resolution of controversies in drug/receptor interactions by...	Spedding	2011	7/7/2019 11:28 AM
Ligand- and Structure-Based Drug Design and Optimization usi...	Mazanetz et al.	2019	6/29/2019 3:12 AM
Targeted protein degradation: elements of PROTAC design	Paiva and Crews	2019	6/7/2019 10:40 PM
Studying how genetic variants affect mechanism in biological...	Raimondi and Russell	2018	6/2/2019 12:28 PM
Advances and Challenges in Computational Target Prediction	Sydow et al.	2019	6/2/2019 10:09 AM
Toxicology testing steps towards computers	Kling	2019	6/2/2019 10:07 AM
MolTarPred: A web tool for comprehensive target prediction w...	Peón et al.	2019	6/2/2019 9:39 AM
Preclinical candidate for the treatment of visceral leishman...	Wyllie et al.	2019	6/1/2019 8:40 PM

****

Source links: below is an *ad hoc* collection (but may get some structure later) that were collated as groundwork for the review document. It includes mainly resources and announcements relevant to ADEV, as well as some key publications (n.b. the more hyperbolic or verbose wording from websites and press releases has been edited out. There is some overlap with Appendix I because the same organisations release announcements via different channels).

****

http://www.cd3.eu/ Centre for Drug Design and Discovery (CD3) CD3 is an investment fund and technology transfer platform aimed at promoting the discovery and development of innovative medicines for all kinds of diseases. It builds on a pool of innovative research and technologies of universities and spin-off companies. By providing the necessary expertise and financial resources, CD3 ensures that fundamental biomedical research carried out by universities and small biotech companies is translated into more usable results that can then be further developed by pharmaceutical companies or form the basis for new spin-off companies. CD3 was set up at the end of 2006 by KU Leuven Research & Development (LRD) and the European Investment Fund (EIF) with a starting capital of 8 million euro.

https://www.mcgill.ca/neuro/article/open-science-research/4m-open-science-drug-screening $4M for open science drug screening. The Neuro (Montreal Neurological Institute and Hospital) will lead an open science partnership to develop precision drugs for amyotrophic lateral sclerosis (ALS) and Parkinson's disease (PD). A grant from Quebec's Ministry of Economy and Innovation, under its Fonds d'accélération des collaborations en santé (FACS), worth up to $2M, will fund this unique public-private partnership that includes The Neuro, Takeda Pharmaceuticals and the Structural Genomics Consortium (SGC). Takeda, SGC, and

philanthropist J. Sebastian van Berkom will match the public funds providing up to a further $2M. Called NeuroOme, the partnership will use a multi-omics approach that incorporates several genetic datasets for more precise screening of promising drug targets.

https://drive.google.com/file/d/1WQK67Cxib3AtOr4FPZM1eBUIl2ABowx0/view
June 3, 2019, A new consortium of pharmaceutical, technology and academic partners has announced the launch of the "MELLODDY" (Machine Learning Ledger Orchestration for Drug Discovery) project which aims, for the first time, to use machine learning methods on the chemical libraries of 10 pharma companies and to develop a platform creating more accurate models to predict which compounds could be promising in the later stages of drug discovery and development. The project, which demonstrates a new model of collaboration between traditional competitors in drug discovery, involves an unprecedented volume of competitive data.

http://www.mauricewilkinscentre.org/about-us/what-we-do.aspx The Maurice Wilkins Centre produces world class research and makes exciting scientific discoveries. We harness and link New Zealand's outstanding expertise in biomedical research to develop cutting-edge drugs and vaccines, tools for early diagnosis and prevention, and new models of disease.

https://ncats.nih.gov/tissuechip/projects NCATS, in collaboration with other NIH Institutes and Centers and the Food and Drug Administration, is leading the Tissue Chip for Drug Screening program to develop human tissue chips that accurately model the structure and function of human organs such as the lungs, liver and heart, to help predict drug safety in humans more rapidly and effectively. The program has focused on developing physiologically relevant models for toxicity testing, disease modelling and efficacy testing.

https://www.simulations-plus.com/software/admetpredictor/ ADMET Predictor is a software tool that quickly and accurately predicts over 140 properties including solubility, logP, pKa, sites of CYP metabolism, and Ames mutagenicity. The ADMET Modeler™ module in ADMET Predictor allows one to rapidly and easily create high quality QSAR/QSPR models based on your own data. The newest module offers advanced data mining, clustering, and matched molecular pair analysis. The program has an intuitive user interface that allows one to easily manipulate and visualize data.

https://www.scibite.com/news/keeping-up-with-the-life-sciences-literature-how-semantic-enrichment-is-changing-the-way-we-search/ For keeping up with the Life Sciences literature Semantic Enrichment is changing the way we search. This blog discusses the challenges life sciences companies, like LifeArc, face in keeping up-to-date with scientific literature, and how semantic enrichment technology can automate this process to reduce the time spent mining data by up to 80%.

https://endpts.com/1-3b-keytruda-windfall-will-be-used-to-accelerate-translational-drug-rd-in-the-uk/ The nonprofit LifeArc has struck a deal to sell a large block of its Keytruda royalties to a pension firm for $1.3 billion it will now use to support scientists at work developing new drugs and devices. The rich UK R&D scene has been a fertile ground for early stage work. This windfall of blockbuster cash will now bolster LifeArc's work in the field.

https://www.bioinfogate.com/bioinfogate/fda-and-bioinfogate-agree-to-extend-the-material-transfer-agreement-mta-to-access-the-off-x-translational-safety-intelligence-portal/ FDA and Bioinfogate agree to extend the material transfer agreement (MTA) to access the OFF-x

translational safety intelligence portal. The agreement aims to evaluate new approaches to enhance safety assessments of human pharmaceuticals and their associated molecular targets, supporting the FDA's mission of protecting public health. Under this MTA, the FDA has agency-wide access to Bioinfogate's translational safety intelligence portal, OFF-X. The primary objective of the agreement is to evaluate the potential use of Bioinfogate OFF-X™ as a research tool to anticipate adverse events associated with molecular targets and evaluate its utility in the regulatory review process. Under the recently extended agreement, the FDA will continue to provide feedback on the utility of the OFF-X database and/or any aspects related to its functioning and use. Unexpected safety issues constitute one of the most disruptive events in clinical research. A translational approach to drug safety that integrates preclinical and clinical data can significantly reduce patient burden and avoid costly failures. In this context, it is essential to detect as early as possible in the drug R&D process the potential safety liabilities associated to both new targets and drugs under development. May 2019 numbers; targets, 7,707, other biomolecules of interest: 6,000, drugs & biologics, 9,594 and safety alerts, 322,540 (n.b. Prous was the historical developer of Claviariate Integrity but not yet clear how much overlap there may be with OFF-X for targets and compounds. Also see below)

https://clarivate.com/blog/news/clarivate-analytics-signs-agreement-with-melbourne-biomedical-precinct-to-provide-comprehensive-drug-discovery-intelligence/ Clarivate Analytics providing insights to accelerate the pace of innovation, has signed an agreement to provide drug discovery intelligence across the Melbourne Biomedical Precinct (MBP). Under the agreement, The University of Melbourne, Monash University, BioCurate and The Walter Eliza Hall Institute of Medical Research will have enterprise-wide access to Cortellis drug discovery intelligence to support their strategic research practices from early science through drug discovery. Cortellis is a suite of life science solutions from Clarivate Analytics that inform decision-making across the drug/device development and commercialization spectrum

https://www.broadinstitute.org/therapeutics/areas-focus Broad Institute areas of focus: chemical biology and therapeutics science has evolved rapidly over the past decade. CBTS is leveraging breakthroughs in the use of macromolecules to study biology and modify the genome and proteome in therapeutically relevant ways; development of technologies to explore the small-molecule and macromolecular structure space and rapidly identify molecules with remarkable biological properties; innovative means of using chemistry to provide an unprecedented view of previously unseen processes in living systems. These efforts combine with disease-focused research conducted by Broad scientists to find therapeutic solutions to the biological challenges posed by human disease.

https://www.gvkbio.com/media/press_releases/lifearc-gvk-bio-celebrate-decade-partnership-transforming-research-life-saving-medicines/ GVK BIO, a leading global Contract Research and Development Organisation (CRDO), and LifeArc, one of UK's top medical research charities, are pleased to announce the 10th anniversary of a successful partnership. The collaboration between LifeArc and GVK BIO brings together two organizations with a singular vision: advancing novel research ideas into potential medicines that can make a significant difference to patients. LifeArc's novel approach involves levering partnerships with academia and leading service organizations to advance discoveries into promising medicines. Over the past decade, LifeArc scientific teams have focused their efforts on modulating novel targets with therapeutic potential in diseases where there is patient need,

with the GVK BIO team assisting in synthesizing complex molecules rapidly and cost effectively to drive iterative improvement in efficacy and safety. Promising novel chemical entities discovered from this partnership have been either advanced by LifeArc into clinical development, or have been licensed out to leading global pharmaceutical companies, including a recently announced deal in neuroscience. As the partnership evolved from a tactical relationship into a preferred partnership, the GVK BIO team has augmented LifeArc's internal biology efforts, by supporting biochemical assays, carrying out drug metabolism and pharmacokinetics profiling studies, and soon, will be developing cell-based assays and running protein expression and purification studies.

https://s3-eu-west-1.amazonaws.com/media.newmd.catapult/wp-content/uploads/2019/05/09112422/sodn19.pdf This second State of the Discovery Nation report provides key insights from Medicines Discovery Catapult's (MDC) market research, surveys, interviews, and experience from the drug discovery community. MDC has two main purposes: 1) Identifying, industrialising and driving the adoption of technologies and methods that will improve productivity and predictability of medicines discovery.Two areas of Research and Development (R&D) selected to industrialise new tools are Humanised preclinical models with combinations of advanced bioanalytical tools and Artificial Intelligence (AI)/Machine Learning systems. 2) Small or medium-sized enterprises (SMEs) and translational academics require efficient access to UK infrastructure, both public and private, to support their R&D. MDC has therefore developed 3 platforms; Virtual R&D: Brokered access to outsourced expertise and experimental support, Samples & Data: Brokered access to consented patient samples and clinical data sets, Syndicates: Managed consortia of medical research charities and other providers to affect portfolios of patient-driven discovery projects. UK drug discovery is a large, diverse, vibrant and growing sector. Turnover increased by £3.3bn and created 47 new businesses between 2016 and 2017, 300 companies are focussed on discovering potential new medicines; 70 percent are working in the areas of cancer, anti-infectives and CNS, 1,200 companies provide vital services and supplies, SMEs are a critical source of innovation for new medicines with 60% fewer than five staff and 80% fewer than 20.

https://md.catapult.org.uk/news/medicines-discovery-catapult-and-lifearc-launch-strategic-rd-partnerships-in-biomarker-discovery-and-proteomics-technology/ Today (9th May 2019), Medicines Discovery Catapult and LifeArc, leading forces driving UK innovation in drug discovery, announce their strategic partnership. The identification and development of novel biomarkers that detect efficacy and safety of novel drug molecules, which can be used in human trials, will be the focus of the partnership. LifeArc has a growing portfolio of innovative therapeutic candidates across priority disease areas where there is high patient need. Medicines Discovery Catapult has established platforms and expertise in the discovery and validation of new biomarker approaches. The partnership will use the expertise of both organisations to progress candidate molecules towards clinical evaluation with an optimised biomarker strategy. LifeArc will fund a select number of full-time employees, who will be embedded within the Catapult to provide strategic biomarker discovery and development.

https://www.genomicsplc.com/ Four leading geneticists formed Genomics plc in 2014 to combine cutting-edge science with commercial delivery. We have built the world's largest genomic database, which continues to expand in size and diversity every week. Our algorithms overcome the limitations of other methods, and are capable of working at an unprecedented scale of data to tell us how any one of millions of genetic variants affects

thousands of molecular, cellular, physiological and disease traits. We use this powerful engine to find new drug targets for serious diseases, and to deliver precision health.

https://www.scilifelab.se/news/researching-cells-through-colorful-imaging/ The Cell Profiling facility at SciLifeLab has a unique library of more than 25 000 validated antibodies, which are used to create images describing the inner workings of cells. The antibodies come from the Human Protein Atlas project, a project that started in 2003 with the aim of mapping all human proteins in cells and tissues. Having access to the library means that researchers can use the facility to visualize many different proteins.

http://www.e-tox.net/ The eTOX project as IMI consortium has been completed with the accomplishment of an effective synergic sharing of historical toxicological data within the pharmaceutical industry. It created a series of models to support toxicity prediction. Both data and models are integrated in the platform developed in the project, the eTOXsys.

https://www.lbic.com/about The creation of the London BioScience Innovation Centre in 2000 was a direct response to the evident shortage of accommodation for life sciences activity in central London with the objective of developing a commercial life sciences cluster around the capital's world-class knowledge base. It provides a focus for life sciences activity in the capital, offering laboratory and office facilities. Owned by the Royal Veterinary College, LBIC is home to more than 50 biotechnology and life science companies, including small start-ups and established companies. (see WO2013186546, Polytherics Ltd London Bioscience Innovation C, Drugs for Neglected Diseases Initiative 'Complexes of amphotericin B with poly glutamic acid')

https://www.img.cas.cz/2018/04/cz-openscreen-is-a-founding-member-of-eu-openscreen/ The Czech Republic CZ-OPENSCREEN became the member of the European research Infrastructure Consortium EU-OPENSCREEN ERIC. EU-OPENSCREEN ERIC, the infrastructure for chemical biology established European Commission on March 20, 2018. Its host country is Germany (with the headquarters based in Berlin), and the Czech Republic together with Finland, Latvia, Norway, Poland and Spain belongs to the founder states.

https://www.ft.com/content/4da1c6e4-603d-11e9-9300-0becfc937c37 Achaogen is not a company most people have heard of even if its successful antibiotic product (plazomycin PubChem CID 42613186) may have saved their life. Its recently announced bankruptcy is one of the most significant and worrying corporate failures of this decade.

https://www.abdn.ac.uk/kosterlitz/about/centre-strategy/ The Kosterlitz (center, Aberdeen, UK) oversees projects which focus on target validation and drug discovery. The team works alongside Research & Innovation and the lead biologist for each project. These have generally started with biological innovation, e.g. relating to the discovery of a new receptor or pathway and involvement in a disease.

https://www.ncbi.nlm.nih.gov/pubmed/29564125 The 'Therapeutics discovery: From bench to first in-human trials' conference, held at the King Abdullah International Medical Research Center (KAIMRC), Ministry of National Guard Health Affairs (MNGHA), Kingdom of Saudi Arabia (KSA) from October 10–12, 2017, provided a unique opportunity for experts worldwide to discuss advances in drug discovery and development, focusing on phase I clinical trials.

https://www.eurekalert.org/pub_releases/2019-04/jh-daa041519.php DNDi and Atomwise collaborate to advance drug development using AI for neglected diseases. Efforts to discover novel compounds against Chagas disease see early successes -- part of Atomwise's Artificial Intelligence Molecular Screen (AIMS) Awards program to fast track drug development.

http://fibrosix.com/ FibrosIX Inc. is a start-up based on technology from The University of Michigan and Michigan State University that aims to develop novel anti-fibrotic agents which disrupt gene transcription. They have been awarded an SBIR grant from the National Cancer Institute of NIH to advance the development of our compounds in fibrotic diseases (r43 ca235823-01 demonstrating the feasibility of CGC-257081 as a preventive therapy to inhibit bleomycin-induced lung fibrosis in mice (see PMID 28285914 and other Neubig RR papers such as PMID 30951312)

http://www.edctp.org/ The European & Developing Countries Clinical Trials Partnership (EDCTP) funds clinical research to accelerate the development of new or improved drugs, vaccines, microbicides and diagnostics against HIV/AIDS, tuberculosis and malaria as well as other poverty-related infectious diseases in sub-Saharan Africa, with a focus on phase II and III clinical trials. http://www.edctp.org/news/wanecam-consortium-kicks-off-clinical-study-for-next-generation-antimalarial-drug/ West African Network for Clinical Trials of Antimalarial Drugs (WANECAM), a consortium of ten academic organisations in Africa and Europe, starts off the 'WANECAM 2' study in Bamako, Mali with a meeting on 15 April 2019. EDCTP invests EUR 10 million in the project. WANECAM will conduct clinical trials of a novel antimalarial combination comprising KAF156 (ganaplacide) http://www.guidetopharmacology.org/GRAC/LigandDisplayForward?ligandId=9946 and lumefantrine in a new once-daily formulation. The project is led by Professor Abdoulaye Djimdé of the Université des Sciences, Techniques et Technologies de Bamako. The consortium will collaborate with the not-for-profit product development partnership Medicines for Malaria Venture (MMV) and the pharmaceutical company Novartis to develop its compound KAF156 in combination with lumefantrine. The aim is to advance the development of a much-needed new antimalarial therapy while strengthening clinical trial development capabilities in Africa.

https://www.monash.edu/pharm/research/pharmaceutical-sciences Monash Institute of Pharmaceutical Sciences (MIPS) Excellence in translational drug discovery and development. Drug Discovery Biology, Deep understanding of drug targets to enable the design and development of better medicines, Medicinal Chemistry The design and synthesis of bioactive molecules for chemical-biology and drug discovery, Drug Delivery Disposition and Dynamics, Next-generation drug-delivery systems to enhance drug exposure and action, Centre for Drug Candidate Optimisation, Translating ligands into drugs by optimising ADME and pharmacokinetic properties, Centre for Medicine Use and Safety, Conducting research into preventive, acute and chronic care, optimising medication management and patient safety, Research areas explained.

https://www.monash.edu/pharm/research/areas/optimisation The Centre for Drug Candidate Optimisation (CDCO) is a collaborative research centre based within the Monash Institute of Pharmaceutical Sciences. We provide expertise and infrastructure in physicochemical property evaluation, drug metabolism and pharmacokinetics to multidisciplinary drug discovery teams for improved compound design, selection and progression. Established in 2003, we have collaborated with numerous drug discovery groups that have progressed 28

novel drug candidates into clinical development across disease indications including cancer, CNS disorders, cardiovascular disease and infectious diseases.

http://biocurate.com/about-biocurate BioCurate Pty Ltd was formed jointly in June 2016 by The University of Melbourne and Monash University, with support from the State of Victoria. As an independent 'venture catalyst' and technology accelerator, BioCurate operates in the under-resourced, yet critical early phases of new drug development (a phase often referred to as the 'valley of death' due to the high rate of attrition). BioCurate will provide the commercial focus, expertise and funding necessary to address the barriers to success, and drive early decisions, with the intent of boosting the successful development of new medicines. A portfolio of quality candidates has commenced development and spans a range of therapeutic areas and modalities. Strong and collaborative linkages are well progressed with the Universities and other key partners in health, research and industry in Australia and internationally.

https://www.sib.swiss/about-sib/news/10318-one-million-jobs-on-sib-s-computer-aided-drug-design-software-platform?utm_source=newsletter&utm_medium=email&utm_campaign=swiss-bioinformatics-09_04_2019&utm_content=SwissDrugDesign On 16 December 2018, the count of total jobs run on the SIB Resource SwissDrugDesign passed the symbolic one-million mark, highlighting a remarkable acceleration of usage over the past years.

https://www.openforrare.com/ The 'Open for Rare' program uses a collaborative, open science approach to improve the intelligibility of the molecular defects that cause rare diseases in order to accelerate the development of medicines made available at affordable prices and the democratization of science.

https://www.warf.org/for-uw-inventors/warf-therapeutics/warf-therapeutics.cmsx WARF Therapeutics mission to partner with UW-Madison and Morgridge Institute for Research Principal Investigators (PIs) interested in translational research, invest to discover & develop novel drug like molecules that modulate validated targets and improve human disease, and partner with Bio-centric companies to develop and commercialize.

https://www.nature.com/articles/d41586-019-01054-4?utm_source=twt_nnc&utm_medium=social&utm_campaign=naturenews&sf210373931=1 India's clinical-trial rules to speed up drug approvals. Drugs that have been approved in some countries will no longer require trials in India.

https://www.theguardian.com/commentisfree/2017/apr/19/why-open-source-pharma-is-the-path-to-both-cheaper-and-new-medicines Why open source pharma is the path to both new and cheaper medicines, by Matthew Todd

https://www.sygnaturediscovery.com/ Sygnature Discovery is a leading 'centre of excellence' for integrated drug discovery, offering significant know-how and expertise across a broad range of therapeutic areas and biological target classes. As we approach candidate selection our experienced drug metabolism and pharmacokinetics (DMPK) scientists will help to establish in-vitro to in-vivo and PK/PD correlations and predictions of human dose for promising candidates.

https://www.wipo.int/research/en/about/ WIPO Re:Search is a public-private partnership administered by the World Intellectual Property Organization (WIPO) in collaboration with BIO Ventures for Global Health (BVGH). Members include some of the world's largest pharmaceutical companies, prestigious academic institutions and product development partnerships. Mission is improved global health through innovation that mobilizes intellectual property and the power of private and public sector collaborations (see 2019, PMID 30917506).

https://www.medchemica.com/ MedChemica Founded in 2012 by experienced large Pharma medicinal and computational chemists to accelerate drug hunting by exploiting data driven knowledge, domain leaders in SAR knowledge extraction, experienced in secure large scale cloud computing, creators of the largest ever documented database of medicinal chemistry ADMET knowledge.

https://invivoai.com/en/about-us#about InVivo AI interdisciplinary team of scientist-entrepreneurs with extensive experience in machine learning, drug discovery, and building and scaling impactful technology businesses.

http://insitro.com/ Insitro will integrate cutting-edge machine learning techniques with the ground-breaking innovations that have occurred in life sciences that enable the creation of the large, high-quality data sets, collect and use a range of very large data sets to train ML models to help address key problems in the drug discovery and development process.

http://www.euretos.com/ Euretos is an AI platform used by pre-clinical researchers for in-silico discovery & validation of targets and biomarkers. World leading pharma, biotech and academic institutions use it to accelerate their research. Available as subscription, consulting service or direct access.

https://benevolent.ai/about-us BenevolentAI has capability from early discovery to late stage clinical development. Our platform of computational and experimental technologies and processes, drawing on vast quantities of mined and inferred biomedical data, can advance the entire drug development process. The platform is built and used by our world-class scientists, researchers, and technologists, working side-by-side. Our strength comes from this integrated, end-to-end approach, combined with a relentless pursuit of scientific and technological excellence.

http://www.atomwise.com/aims-awards/ Artificial Intelligence Molecular Screen (AIMS) Awards program by Atomwise for academic researchers seeking novel compounds to treat disease.

https://www.exscientia.co.uk/#home-section At the forefront of small molecule drug discovery we are the first company to automate drug design, surpassing conventional human endeavour. Our AI driven systems actively learn best practice from vast repositories of discovery data and are further enhanced with knowledge acquired from seasoned drug hunters. Exscientia has already delivered exceptional productivity, generating candidates in roughly one-quarter of the time of traditional approaches.

https://www.dndi.org/2019/media-centre/press-releases/gardp-set-up-as-independent-legal-entity/ The Global Antibiotic Research and Development Partnership (GARDP) https://www.gardp.org/about/vision-mission/ is now an independent legal entity following a

successful three-year incubation, hosted by the Drugs for Neglected Diseases initiative (DNDi). During this time, GARDP has already begun working with partners (inc. WHO) to develop antibiotics to tackle drug-resistant infections which pose a threat to global health and development, including the achievement of the Sustainable Development Goals.

https://pharmaintelligence.informa.com/products-and-services/data-and-analysis/citeline Citeline is the industry's most comprehensive, reliable and current global R&D intelligence suite of solutions. 265,000+ clinical trials, 400,000 investigators, 68,000+ drugs, 117,000+ sites, 235+ diseases in 8 major therapeutic areas across 166 countries from 250+ therapeutic area experts and analysts.

https://www.excelra.com/discovery/#gostar GOSTAR is the world's largest manually curated knowledgebase of biologically active small molecules, with annotated pharmacological profiles. It singularly provides comprehensive, actionable intelligence into SAR, ADME, toxicity, and pharmacokinetic profiles of compounds in discovery, development, and market (see poster)

https://clarivate.com/products/integrity/ Integrity and their Biomarkers module provide the key scientific information and insight needed to prioritize drug development candidates, benchmark competitors/market leaders and assess the competitive landscape earlier than any other drug R&D database directly supporting scientific discovery. ~ 500,000 Drugs & biologics > 2,100,000 Experimental pharmacology records > 74,000
Experimental model records

http://prosilico.com/ PROSILICO is a Swedish company focusing on the research and development of innovative technologies to provide high quality estimates of human ADME/PK directly from chemical structure.

https://pharmaintelligence.informa.com/resources/product-content/pharma-rd-annual-review-2018 Citeline's Pharmaprojects' Pharma R&D Annual Review 2018 expert insights and illuminating trends that affected pharmaceutical R&D throughout 2017

https://www.bio.org/press-release/bio-releases-largest-study-ever-clinical-development-success-rates Biotechnology Innovation Organization (BIO) released a study of clinical development success rates 2006-2015. This analyzed 9,985 clinical and regulatory phase transitions, across 1,103 companies.

https://www.drugdiscoverychemistry.com/lead-optimization/ Lead optimization for drug metabolism & safety tools and strategies for predicting, evaluating and building safety into drug design (conference abstracts)

http://addconsortium.org/index.php Academic Drug Discovery Consortium (ADDC) collaborative network among the growing number of university-led drug discovery centers and programs.

http://nd4bb-enable.eu/ European Gram Negative AntiBacterial Engine (ENABLE). The ENABLE consortium is working to advance the development of promising new antibiotic compounds targeting Gram-negative bacteria such as Escherichia coli. The goal of the project is to develop attractive antimicrobial candidates for extensive preclinical studies and First-in-Human clinical trials.

https://www.ft.com/content/5541440c-38ff-11e9-b856-5404d3811663?segmentid=acee4131-99c2-09d3-a635-873e61754ec6 Kelly Chibale: Leading the way in African drug discovery

https://www.cell.com/cell/pdf/S0092-8674(19)30231-4.pdf The Missing Diversity in Human Genetic Studies

https://knect365.com/next-generation-therapeutics/article/106ee157-8da3-499c-9dd2-7effe3f4e93f/boston-non-profit-seeding-labs-connect-scientists-worldwide?utm_source=twitter&utm_medium=social&utm_campaign=nextgencontent&utm_content=bio Boston non-profit connecting scientists worldwide

https://info.africarxiv.org/ AfricArxiv is a free, open source and community-led digital archive for African research. We provide a non-profit platform for African scientists to upload their working papers, pre-prints, accepted manuscripts (post-prints), and published papers. We also provide options to link data and code, and for article versioning.

https://www.rx-network.com/channels/1923-spotlight-on-preclinical-testing/posts/44891-infographic-preclinical-testing-survey Infographic: preclinical testing survey

https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm633500.htm
https://www.biopharmadive.com/news/fdas-gottlieb-calls-for-greater-adoption-of-new-clinical-trial-tools/550598/ Statement by FDA Commissioner Scott Gottlieb, M.D., on new strategies to modernize clinical trials to advance precision medicine, patient protections and more efficient product development

https://md.catapult.org.uk/opportunities/secure-a-share-of-up-to-20-million-for-your-medicines-rd-innovation/ Medicines Discovery Catapult to help push projects towards commercial viability, with access to our teams connections – across industry, academia, finance, government, and research networks

https://www.broadinstitute.org/center-development-therapeutics-cdot The goal of the Center for the Development of Therapeutics (CDoT) is to carve a path to novel medicines for the treatment of important human diseases. CDoT's role is to demonstrate how biological insights gained at the Broad (Institute) can be leveraged to create small molecules or biologics that can eventually become medicines.

https://clinicaltrials.gov/ct2/results/map?cntry=US&map= World map of clinical trials

https://calgaryherald.com/health/diet-fitness/keenan-making-drugs-to-save-lives-not-to-make-money Aled Edwards, professor of medical genetics at the University of Toronto, certainly thinks so. With business partner Owen Roberts, he's created two not-for-profit companies targeting different diseases. M4K 2009
https://pubchem.ncbi.nlm.nih.gov/substance/375272374

https://www.researchresearch.com/news/article/?articleId=1379716
How African scientists can boost the visibility of their research

https://www.lifesci.dundee.ac.uk/research/wcair

The vision for the Wellcome Centre for Anti-Infectives Research (WCAIR) is to help tackle the urgent unmet medical need and lack of drug discovery research for Neglected Tropical Diseases (NTDs).

https://blogs.lse.ac.uk/impactofsocialsciences/2019/03/06/plan-s-and-the-global-south-what-do-countries-in-the-global-south-stand-to-gain-from-signing-up-to-europes-open-access-strategy/ Plan S and the Global South – What do countries in the Global South stand to gain from signing up to Europe's open access strategy?

https://saferworldbydesign.com/saferskin/app/
Integrated testing strategies (ITS) for regulatory safety assessments serve as a replacement for stand-alone tests and typically combine information from different sources (in vitro, in silico) to arrive at a prediction for a specified endpoint. Skin sensitisation is one such endpoint that has been prioritized by the OECD and other public agencies

https://www.research4life.org/wp-content/uploads/2016/07/R4L_strategy_report_web.pdf
Research4Life is a collaborative public-private partnership. The initiative, designed to support research conducted in lower income countries, provides free or low-cost online access to an extensive range of online scientific publications in key areas related to improving quality of life. The publications include subscription-based and open access resources.

https://www.thesgc.org/chemical-probes Potent, selective and cell-permeable inhibitors of protein function ("chemical probes") are valued reagents in both fundamental and applied biological research, and they are essential for the early stages of drug discovery by allowing preclinical target validation in both academic and industrial laboratories. 22,186 samples of SGC's open chemical probes have now been distributed/bought, and 13,159 papers mention an SGC probe by name.

https://www.dndi.org/2019/media-centre/news-views-stories/news/hcv_rnd_status_2019/
In 2016, DNDi launched a Phase II/III study in Malaysia and Thailand to assess the efficacy, safety, tolerability, pharmacokinetics, and acceptability of 12- and 24-week regimens containing the drug candidate ravidasvir combination with sofosbuvir in patients with HCV. A total of 301 patients were included: 220 in Malaysia and 81 in Thailand.

https://www.bindingdb.org/jsp/dbsearch/PrimarySearch_patent.jsp?patent=US9518026&patent_submit=Search New NDT patent example US9518026206, 5-aminopyrazole-4-carboxamide inhibitors of CDPK1 from T. gondii and C. parvum
University of Washington Through its Center For Commercialization

https://github.com/OpenSourceMycetoma Open Source Mycetoma, New Medicines for Mycetoma without Secrecy

http://www.emro.who.int/media/news/global-conference-on-mycetoma.html
Delegates attending the Sixth International Conference on Mycetoma in Khartoum, Sudan, have endorsed a "call for action" urging the global community to work together with multilateral agencies, partners, research institutions and pharmaceutical companies to address the devastating consequences of this disease.

https://github.com/OpenSourceMalaria http://opensourcemalaria.org/
The Open Source Malaria project is trying a different approach to curing malaria. Guided by open source principles, everything is open and anyone can contribute.

https://www.myeloma.org.uk/news/collaborative-research-supporter-funding-key-finding-cure-myeloma/ Collaborative research and supporter funding key to finding a cure for myeloma

https://www.afd.fr/en/dndi-distinctive-illustration-commons-area-public-health
A new research paper recently published by the AFD (Agence Française de Développement) examines how DNDi, a distinctive Product Development Partnership (PDP), can illustrate what can be presented as a "commons" in the area of public health.

https://www.surechembl.org/document/WO-2018229683-A1/
5,6-fused-bicyclic compounds and compositions for the treatment of parasitic diseases. Novartis patent for proteasome inhibitors against leishmaniasis and Chagas disease.

https://www.bioarctic.se/en/about-bioarctic-1881/ BioArctic is a Swedish research intensive biopharma company aimed at developing new treatments that address the causes of diseases that affect the Central Nervous System

http://alzheimereurope.newsweaver.com/ConferenceAnnouncement/8w4ryiy1ssc
The Innovative Medicines Initiative launches a public-private coordination and support action (NEURONET) to develop an operational platform for its neurodegeneration projects

https://www.alzheimer-europe.org/Alzheimer-Europe Alzheimer Europe is a non-governmental organisation (NGO) aiming to provide a voice to people with dementia and their carers, make dementia a European priority, promote a rights-based approach to dementia, support dementia research and strengthen the European dementia movement

https://www.imprind.org/outcomes/ Blocking aggregate propagation in neurodegenerative diseases IMPRiND – Inhibiting Misfolded protein Propagation In Neurodegenerative Diseases – is an international consortium that aims to map and target critical steps in the propagation of misfolded tau and α-synuclein, considered the main culprits of neurodegeneration in Alzheimer's and Parkinson's disease respectively.

https://www.dementiasplatform.uk/about/our-organisation Dementias Platform UK (DPUK) was established to enhance research into dementia, so we find effective treatment sooner.

https://go.chanzuckerberg.com/NCN The CZI Neurodegeneration Challenge Network brings together experimental scientists from diverse research fields, along with computational biologists and physicians, to understand the fundamental biology of neurodegenerative disorders. Their shared aim is to develop new strategies for the treatment and prevention of neurodegenerative diseases.

http://www.pharmtech.com/fda-approves-novartis-drug-tropical-disease The approval expands use of the drug, Egaten (triclabendazole), which has been by the WHO since 2005 for treating liver fluke infestation.

https://atlasantibodies.com/ Validated antibodies

http://agencia.fapesp.br/sonda-quimica-permite-regular-via-de-sinalizacao-essencial-nas-celulas-/29699/ A group of researchers at the Center for Medicinal Chemistry (CQMED) discovered a way to regulate the WINT pathway. Created with support from FAPESP through the Program for Support of Research in Partnership for Technological Innovation (PITE), the CQMED is a unit of the Brazilian Industrial Research and Innovation Company (Embrapii) specialized in the field of biopharmaceuticals and pharmaceuticals. T

https://www.dndi.org/wp-content/uploads/2019/02/DNDi_Mycetoma_2019.pdf
DNDI mycetoma

https://www.cigionline.org/articles/innovation-open-science-means-open-business
How Canadians manage and mismanage intellectual property (IP) lies at the heart of Canada's innovation woes. As federal and provincial governments unroll their innovation strategies – after decades of neglect – they need to build on Canadian successes while putting aside tired and failed strategies. One groundbreaking deal announced last week points to a way forward

https://www.uq.edu.au/news/node/122254 A team of Queensland researchers developing new antibiotics to treat tuberculosis has been awarded $1.45 million by the US Department of Defense.

https://www.collaborativedrug.com/industrial-and-academic-neglected-disease-drug-discovery-collaborations-gsk-and-northeastern-university-cdd-webinar-video/
CDD webinar recording from Julio Martin and his colleagues at the Tres Cantos GSK Open Laboratory are working collaboratively on diseases like African Sleeping Sickness, Chagas Disease, Leishmaniasis, Tuberculosis, and Malaria in an industrial environment. Michael Pollastri brings big pharma experience to multi-center academic collaborations. Their experiences are relevant for both commercial and neglected disease drug discovery. They are both experimenting while pushing boundaries in new modes combining sensitive and open data.

https://www.utoronto.ca/news/u-t-researcher-s-open-science-drug-discovery-model-expands-neurodegenerative-diseases With the help of a $1.5-million grant, U of T's Aled Edwards (right) and business partner Owen Roberts (left) will forgo patents and profits to tackle diseases like Parkinson's and ALS

https://ncats.nih.gov/pubs/features/4dm-2019?utm_source=Twitter&utm_medium=Social&utm_campaign=CenterPromotions_20190215
NCATS Releases Updated Translational Road Maps

http://cibfar.ifsc.usp.br/english/ Center for Innovation in Biodiversity and Drug Discovery (CIBFar)

https://www.ncbi.nlm.nih.gov/books/NBK401307/ NCATS Assay manual

https://www.collaborativedrug.com/therapeutically-relevant-bioassays-cdd-webinar/
Slides and transcript from CDD "Therapeutically Relevant Bioassays" Webinar

https://www.scitecheuropa.eu/microfluidics/92579/ Microfluidics

https://www.telegraph.co.uk/news/0/revolutionary-sleeping-sickness-tablet-approved-use-drc/

New drug which promises to revolutionise the treatment of sleeping sickness is set to be introduced in the Democratic Republic of Congo (DRC).

https://blogs.sciencemag.org/pipeline/archives/2016/02/02/drugs-purely-from-academia Academia (vs. Industry) Drugs purely from academia, By Derek Lowe Feb 2016

https://stories.dndi.org/sleepingsickness-doctors-dream/#group-intro-752cq70HVt Fexinidazole, https://pubchem.ncbi.nlm.nih.gov/compound/68792 the first all-oral drug for both stages of sleeping sickness was approved in November 2018 Here are the stories of the doctors, patients, and scientists who contributed to this story

https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0007005 Ghana: Accelerating neglected tropical disease control in a setting of economic development

https://carb-x.org/ CARB-X is accelerating global antibacterial innovation by investing in the development of new antibiotics and other life-saving products to combat the most dangerous drug-resistant bacteria. Our projects represent the world's largest pre-clinical and early development pipeline of antibiotics and other therapeutics, diagnostics, microbiome, vaccines, and devices.

https://www.dndi.org/2019/media-centre/press-releases/wellcome-commits-10million-develop-new-generation-oral-drugs-leishmaniasis/ Wellcome commits £10 million to DNDi to develop new generation of oral drugs to treat leishmaniasis

https://www.collaborativedrug.com/antibiotic-discovery-getting-through-the-discovery-wall-cdd-webinar-video/ CDD webinar discussion with front-lines of antibiotic drug discovery: Dr. Lynn Silver (ex-Merck) and Dr. Johannes Zuegg (CO-ADD in Australia). Co-moderated by Joe Thomas (The Pew Charitable Trusts).

https://m4kpharma.com/about/#mission At M4K Pharma, we are using open science to revolutionize how affordable new treatments are discovered and developed. Through our Medicines4Kids program, we will aggregate and align the work of global academics, foundations and industry researchers to advance new cures for childhood diseases not well-served by current business models.

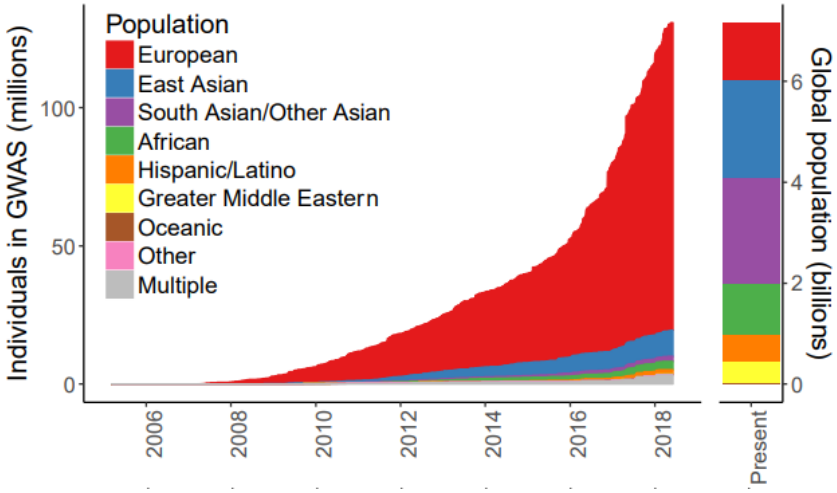
https://www.nature.com/articles/s41573-018-0009-9 The virtuous cycle of human genetics and mouse models in drug discovery

https://www.ncbi.nlm.nih.gov/pubmed/29877628 Assay Guidance Manual: Quantitative Biology and Pharmacology in Preclinical Drug Discovery.

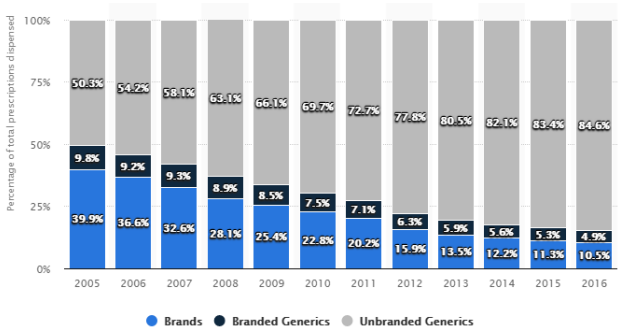
https://www.europeanleadfactory.eu/ The European Lead Factory has successfully accelerated the process of drug discovery. A recently published public summary of results so far clearly shows the benefits of collaborative drug discovery in the European arena.

https://www.fraxa.org/category/clinical-trials+current-research-grants/
Finding effective treatments and ultimately a cure for Fragile x syndrome

https://www.biorxiv.org/content/10.1101/441261v1 Ethical and scientific challenge surrounding clinical implementation of polygenic risk scores is the observation that they are currently of far greater predictive value in individuals of recent European ancestry than others.



https://www.zotero.org/cdsouthan/items/itemKey/52JY66LN Statistics and facts on the U.S. pharmaceutical industry



Appendix III: Academic Drug Discovery Centres (non US)

This list was prepared in Jan 2019 by Michael Spedding and Gaelle Woestelandt of Spedding Solutions

MS has established links to the following (e.g. is on their SAB or has good contacts and could thus be involved at an early stage).

- H3D South Africa http://www.h3d.uct.ac.za/
- Drug Discovery Unit (DDU) Dundee http://www.drugdiscovery.dundee.ac.uk/home
- Eurofins https://www.discoverx.com/home
- National phenotypical screening centre, Dundee, http://npsc.ac.uk/
- National Facility For Drug Discovery Complex, Rajkot, India (Anamik Shah)
- Monash Institute of Pharmaceutical Science
- Melbourne https://www.monash.edu/pharm/research/pharmaceutical-sciences
- Medicines for Malaria Venture MMV

Note that the large number of centres based in the USA are indexed by the Academic Drug Discovery Consortium (ADDC).

****

The non-US Centres below are listed in ADDC. The grey highlights links from the ADDC website.

Sweden : Chemical Biology Consortium Sweden

UK : MRC Protein Phosphorylation & Ubiquitylation Unit
Centre for Therapeutics Discovery
UCL TRO Drug Discovery Group
The Institute of Cancer Research London
University of Sussex Translational Drug Discovery Group

Russia : MIPT Life Science Center
Germany: European ScreeningPort GmbH
PharmaCenter Bonn
TuCADD Tuebingen Center for Academic Drug Discovery
Czechoslovakia : CZ-OPENSCREEN: National Infrastructure for Chemical Biology

France : PCBIS (Platform of Chemical Biology & ADME of Strasbourg)
Switzerland : IOR Institute of Oncology Research
Italy : P4T (Probes for Targets)

Spain: Centre of Applied Medical Research

Health Research Institute of Santiago de Compostela
Spanish National Research Council/Medicinal Chemistry Institute (IQM-CSIC)

Spanish National Cancer Research Centre (CNIO) / Experimental Therapeutics
Portugal: Research Institute for Medicines (iMed.ULisboa)
Cyprus: Experimental Pharmacology Laboratory
Israel BLAVATNIK CENTER for Drug Discovery

The Nancy and Stephen Grand Israel National Center for Personalized Medicine

Saudi Arabia: Genetic engineering and biotechnology
Japan: Drug Discovery Initiative
Australia: CO-ADD (Community for Open Antimicrobial Drug Discovery)
ACRF Drug Discovery Centre
ACRF Rational Drug Discovery Centre
Canada: Institut de Pharmacologie de Sherbrooke (IPS)

****

The main section of this document includes summary descriptions from the ADDC

Chemical Biology Consortium Sweden

http://www.cbcs.se/
Therapeutic Approaches: Small Molecules
Therapeutic Areas of Expertise: Immunology, Infectious Diseases, Metabolic Disorders, Oncology, Pain
Center/Program Highlights: CBCS is established as a non-profit national infrastructure dedicated to the advancement of chemical biology as a critical component of biomedical research in Sweden. The consortium aims to assist and collaborate with cutting-edge academic research for the discovery and development of chemical probes (small organic molecules) with which to unravel new knowledge related to protein and cellular function and signaling. CBCS will supply expertise within assay development, high-throughput and high-content screening, development of molecular libraries, computational chemistry, medicinal chemistry optimisation, targe identification/validation, cheminformatics and in vitro pharmaceutical profiling to academic researchers at Swedish universities. By helping to bring together people, ideas and technology from all aspects of biomedical research in Swedish academia, CBCSâ€™s envisions to become an internationally recognized contributor to the advancement of medical science.
Sources of Funding: Federal/State, University

CBCS CHEMICAL BIOLOGY CONSORTIUM SWEDEN

Institutional Affiliation(s): Karolinska Institute
Address:
LCBKI
Solnavägen 1
Stockholm, 17177
Sweden
Center/Program Start Date: 2010
Number of Employees: 15

MRC Protein Phosphorylation & Ubiquitylation Unit

https://www.ppu.mrc.ac.uk/
Therapeutic Approaches: Small Molecules, Biomarkers, Antibodies, Stem Cells
Therapeutic Areas of Expertise: Cardiology, Immunology, Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology, Regenerative Medicine
Center/Program Highlights: The MRC Protein Phosphorylation and Ubiquitylation Unit (PPU) is a major research centre that focuses on the understanding of the biological roles of phosphorylation and ubiquitylation and how disruption of these processes cause human diseases such as neurodegeneration, cancer, hypertension and immune disorders. The ultimate goal of the Units research programmes is to help develop new improved strategies to treat disease.
A vital asset of the MRC-PPU is the Division of Signal Transduction Therapy (DSTT) which was established in 1998. This division operates as a unique collaboration between scientists in the MRC-PPU, signalling researchers at the University of Dundees College of Life Sciences and six of the worlds leading pharmaceutical companies -- AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Janssen Pharmaceutica, Merck Serono and Pfizer.

The DSTT is widely regarded as a model for how academia should interact with industry. This collaboration agreement has been renewed for the period 2012-2016 and currently focuses on mammalian protein kinases and the ubiquitylation system. Its main aim is to help accelerate the development of future drugs that target specific phosphorylation and ubiquitylation systems for the improved treatment of disease. A key remit of our research is to help define and validate new drug targets with the aim of obtaining sufficiently convincing results to persuade pharmaceutical companies to develop drugs against these targets. Our Unit also helps generate reagents that will help the DSTT companies set up screens to identify and characterise inhibitors. Participating companies have access to our unpublished results, technology, expertise and reagents, as well as first rights to license our intellectual property.
The MRC Protein Phosphorylation Unit at Dundee pioneered analysis of the selectivity of protein kinase inhibitors by setting up the first service to tackle this problem in 1998. This procedure, termed "kinase profiling" proved to be of great help to the pharmaceutical industry, speeding up the development of specific protein kinase inhibitors with therapeutic potential. Working with leading experts from Pharma, we have implemented the industry level of standard operating procedures that guide all our services. Since 1998, we have collaborated with 8 different large pharmaceutical companies including Pfizer, GlaxoSmithKline, Boehringer-Ingelheim, Merck-Serono, Astrazeneca, Janssen Pharmaceutica NV, Novo-Nordisk and Merck & Co. We offer this service world-wide to scientists in both academia and industry: http://www.kinase-screen.mrc.ac.uk/. Our aim for the MRC-PPU is to operate as one of the worlds leading centers for the investigation of signal transduction pathways involving protein phosphorylation and ubiquitylation networks. We are also keen to make our range of essential resources and services available to aide with worldwide research in this area. To this end, we have decided to set up a simple user-friendly, web-based system whereby not-for-profit research groups can request clones, antibodies and proteins our unit possesses. These can be accessed via: https://mrcppureagents.dundee.ac.uk/
Sources of Funding: Federal/State, Philanthropy, Industry Partnerships, Disease Foundations, University, Revenue

Institutional Affiliation(s): The University of Dundee
Address:
The University of Dundee
Sir James Black Centre, Dow Street
Dundee
Scotland, DD1 5EH
United Kingdom
Center/Program Start Date: 1990
Number of Employees: 270

Centre for Therapeutics Discovery

https://www.lifearc.org/working-with-us/drug-discovery/
Therapeutic Approaches: Small Molecules, Biomarkers, Antibodies
Therapeutic Areas of Expertise: Cardiology, Immunology, Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology, Ophthalmology, Psychiatric Disorders, Pain
Center/Program Highlights: LifeArc is a charitable foundation that specialises in life science translational research and licensing. We offer early stage drug discovery, antibody engineering, diagnostics development and intellectual property management and commercialisation. In our laboratories we have more than 80 scientists (including an ex-pharma team of medicinal chemists and a biologics group with nearly a quarter century antibody engineering experience) working to create drug-like molecules to innovative new drug targets. We have a proven track record of success having produced four humanized antibodies with regulatory approval: Elan/Biogen Idecs Tysabri and Chugai/Roches Actemra, Millenium's Entyvio and Merck & Cos Keytruda.
Sources of Funding: Revenue

Institutional Affiliation(s): LifeArc
Address:
SBC Open Innovation Campus
Stevenage
Hertfordshire, SG1 2FX
United Kingdom
Center/Program Start Date: 2006
Number of Employees: 80

UCL TRO Drug Discovery Group

https://www.ucl.ac.uk/translational-research/drug-discovery-group
Therapeutic Approaches: Small Molecules
Therapeutic Areas of Expertise: Cardiology, Immunology, Infectious Diseases, Neurological Disorders, Oncology, Ophthalmology, Psychiatric Disorders, Pain, Regenerative Medicine
Center/Program Highlights:

UCL Translational Research Office

(TRO) builds on an increasingly vibrant translational culture across the School of Life and Medical Sciences (SLMS) and the wider university community by providing integrated support for translational research, industrial partnerships and drug discovery.

Through working closely with a broad spectrum of investigators, industry partners and external funding bodies, the TRO facilitates the translation of emerging research into therapies, techniques and medical products with therapeutic value. The TRO aims to provide practical help and expertise to investigators wishing to explore the translational pathway for their idea / project and to negotiate the hurdles and barriers that are inevitably encountered during the progression from idea to health benefit.

The TRO also develops strategies for engaging effectively with global and national companies and institutions and responds to corporate and academic needs through networking opportunities, tailored partnerships, projects, and brand-building initiatives.

The Drug Discovery Group (DDG) provides theoretical and practical expertise to support the progression of selected UCL projects. Activities include the assessment of target tractability; development of project, medicinal chemistry, and screening strategies; compound design and synthesis; hit discovery and validation; and hit to lead optimisation. The DDG is based in established labs in the UCL School of Pharmacy and is part of the wider UCL Drug Discovery Cluster. The Drug Discovery group also hosts the Innovative Therapeutics Postdoctoral Training Programme (ITPTP).

Sources of Funding: Industry Partnerships, Disease Foundations, University

Institutional Affiliation(s): University College London
Address:
UCL School of Pharmacy,
29-39 Brunswick Square
London
London, WC1N 1AX
United Kingdom
Center/Program Start Date: 2012
Number of Employees: 11

The Institute of Cancer Research London

https://www.icr.ac.uk/

Therapeutic Approaches: Small Molecules

Therapeutic Areas of Expertise: Oncology

Center/Program Highlights: The Cancer Research UK Cancer Therapeutics Unit is a comprehensive small molecule drug discovery center specializing in oncology. Since 2005 we have discovered 16 drug candidates, 6 of these have entered the clinic and our drug abiraterone was approved for late stage prostate cancer. Other drug candidate in the clinic include inhibitors of HSP90, Pi3 kinase and AKT/PKB.

Sources of Funding: Federal/State, Philanthropy, Industry Partnerships, Disease Foundations, University, Revenue

Institutional Affiliation(s): Cancer Research UK
Address:
The Institute of Cancer Research
15 Cotswold Road
London
Surrey, SM2 5NG
United Kingdom
Center/Program Start Date: 1997
Number of Employees: 163

University of Sussex Translational Drug Discovery Group

http://www.sussex.ac.uk/sddc/

Therapeutic Approaches: Small Molecules

Therapeutic Areas of Expertise: Infectious Diseases, Neurological Disorders, Oncology

Center/Program Highlights: The Translational Drug Discovery Group was established in 2011 as part of a major strategic investment within the School of Life Sciences. The group comprises medicinal chemistry and molecular pharmacology laboratories and plays a pivotal role in the Schools translational research strategy which aims to translate a fundamental understanding of disease processes and molecular targets, whether that

knowledge resides within the University of Sussex or in UK-based and international academic and industrial institutions, into novel drugs.
Sources of Funding: Federal/State, Industry Partnerships, Disease Foundations, University, Wellcome Trust, EU

University of Sussex
Translational Drug Discovery Group

Institutional Affiliation(s):
Address:
University of Sussex
Falmer
Brighton
East Sussex, BN1 9QJ
United Kingdom
Center/Program Start Date: 2011
Number of Employees: 20

MIPT Life Science Center

http://www.bio.mipt.ru/
Therapeutic Approaches: Biologics, Small Molecules, Vaccines, Biomarkers, Antibodies, Stem Cells, Gene Therapy, Drug Delivery
Therapeutic Areas of Expertise: Gene-Therapy, Immunology, Infectious Diseases, Neurological Disorders, Oncology, Regenerative Medicine
Center/Program Highlights: MIPT Life Sciences Center (LSC) is the MIPT department specializing in life sciences. The LSC carries out fundamental and applied research in the fields of pharmaceutics, development of new medical technologies, innovative diagnostic and treatment technologies, agro-biotechnologies. The LSC infrastructure includes centers of excellence (analytical laboratories and MIPT Genome center) .
Sources of Funding: Federal/State, Industry Partnerships, University
Institutional Affiliation(s): Moscow Institute of Physics and Technology(MIPT)
Address:
Institutskiy per., 9
Dolgoprudny
Moscow Region , 141700
Russian Federation
Center/Program Start Date: 2014
Number of Employees: > 200

European ScreeningPort GmbH

https://www.ime.fraunhofer.de/
Therapeutic Approaches: Small Molecules, Biomarkers, Stem Cells, Drug Delivery
Therapeutic Areas of Expertise: Cardiology, Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology, Regenerative Medicine
Sources of Funding: Industry Partnerships, Disease Foundations, University, Revenue

European ScreeningPort

Institutional Affiliation(s):
Address:
Schnackenburgallee 114
Hamburg, D-22525
Germany
Center/Program Start Date: 2007
Number of Employees: 22

PharmaCenter Bonn

https://www.pharmazentrum.uni-bonn.de/
Therapeutic Approaches: Biologics, Small Molecules, Biomarkers, Stem Cells, Gene Therapy, Drug Delivery, Drug Devices
Therapeutic Areas of Expertise: Cardiology, Gene-Therapy, Immunology, Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology, Psychiatric Disorders, Pain, Regenerative Medicine, CNS Disorders
Center/Program Highlights: About us
The major aim of the PharmaCenter Bonn (PZB) is to **investigate pharmaceutical principles** and to **develop novel therapeutic approaches**. The latter encompasses the development of new drugs, the study of mechanism

of action and research on and development of experimental therapies, for example, based on stem cells and RNA interference.

As an **interface of basic research to application** (in the form of new drugs and therapies), the PZB plays an important role in the **integration** of all major participants in this process. In particular, the PZB provides a link between basic biomedical research at the University of Bonn and other research institutions (ceasar, Max-Planck-Gesellschaft, Helmholtz research centers) to pharmaceutical and biotechnology companies (especially Life & Brain GmbH) and to the Federal Institute for Drugs and Medical Devices (BfArM).
Finally, it is an important goal to initiate **cooperation** between the PharmaCenter and the neighbouring universities (ABCD-Universities: Aachen-Bonn-Cologne-Duesseldorf), the Max-Planck Institute for the Biology of Ageing in Cologne and the Fraunhofer-Institute in St. Augustin.
Sources of Funding: Federal/State, Industry Partnerships, University

Institutional Affiliation(s): Institutes from the Faculty of Mathematics and Natural Sciences, Institutes from the Medical Faculty
Address:
Bruehler Strasse 7
Room 266
Bonn
North-Rhine Westfalia, 53119
Germany
Center/Program Start Date: 2007
Number of Employees: 250

TuCADD Tuebingen Center for Academic Drug Discovery

https://www.medchem.uni-tuebingen.de/
Therapeutic Approaches: Small Molecules
Therapeutic Areas of Expertise: Immunology, Oncology, Regenerative Medicine
Center/Program Highlights: Our centre combines strong expertises in RNAi based target discovery, modelling/virtual screening and medicinal chemistry. New therapeutic targets (main programs: cancer and regenerative medicine) are functionally identified via in vitro or in vivo shRNA screening. Prioritized targets are subjected to thorough functional validation in predictive preclinical mouse models. Genetically validated targets are then subjected to an established virtual screening plattform. Hits can be validated by re-synthesis and wet lab assays.
The centre has access to fully established medicinal chemistry units with a track record of 3 first into man projects.Further information can be found at http://www.medchem.uni-tuebingen.de/
Sources of Funding: Federal/State, University

Institutional Affiliation(s): Eberhard Karls University and University Hospital Tuebingen
Address:
Auf der Morgenstelle 8
Tuebingen
Baden-Wuerttemberg, 72076
Germany
Center/Program Start Date: 2012
Number of Employees: 20

CZ-OPENSCREEN: National Infrastructure for Chemical Biology

https://openscreen.cz/
Therapeutic Approaches: Small Molecules, Stem Cells
Therapeutic Areas of Expertise: Immunology, Oncology, Regenerative Medicine
Center/Program Highlights:
CZ-OPENSCREEN: National Infrastructure for chemical biology
Sources of Funding: Federal/State, Industry Partnerships

Institutional Affiliation(s): Institute of Molecular Genetics AS CR v.v.i.

Address:
Videnska 1083
Prague 4
Prague, 14220
Czech Republic
Center/Program Start Date: 2011
Number of Employees: 12

PCBIS (Platform of Chemical Biology & ADME of Strasbourg)

http://www.pcbis.fr/
Therapeutic Approaches: Biologics, Small Molecules, Stem Cells
Therapeutic Areas of Expertise: Cardiology, Immunology, Infectious Diseases, Metabolic Disorders, Neurological Disorders, Psychiatric Disorders, Pain, Rare diseases
Center/Program Highlights: The Platform of Integrative Chemical Biology of Strasbourg (PCBIS) gives access to High Throughput Screening (HTS), chemical libraries, target libraries and early ADME-Tox to academic and private laboratories
Starting from biological targets (cells, proteins, nucleic acids, ...), PCBIS is able to set up specific high throughput screening in accordance with the associated composed problem and then to making out at high throughput its chemical and/or natural libraries or the molecules furnished in respect to ISO 9001 Quality Management System. Moreover PCBIS performs early ADME-Tox study which is necessary to determine or to predict properties as early as possible to avoid failures and to limit costs in chemical synthesis, in vivo studies, formulation.

Sources of Funding: Federal/State, Disease Foundations, University, project grants

Institutional Affiliation(s): CNRS & University of Strasbourg
Address:
UMS 3286 CNRS - Université de Strasbourg
Pôle API - Boulevard Sébastien Brant
ILLKIRCH
, 67412
France
Center/Program Start Date: 1997
Number of Employees: 13

IOR Institute of Oncology Research

http://ior.iosi.ch/site/
Therapeutic Approaches: Biologics, Small Molecules, Biomarkers, Stem Cells, Drug Delivery
Therapeutic Areas of Expertise: Oncology
Center/Program Highlights: **The IOR Institute of Oncology Research is part of the Research Division of the Oncology Institute of Southern Switzerland (IOSI), the comprehensive cancer center of the Ente Ospedaliero Cantonale (EOC), and is the main facility for basic and translational research at the IOSI. The Institute was established in 2003 formerly as the IOSI Laboratory of Experimental Oncology.**
The Institute comprises different groups working in the areas of tumor biology, genomics, molecular oncology and experimental therapeutics. The expertise available in the Institute spans from cell and molecular biology to molecular pharmacology, genomics and bioinformatics.
IOR serves also as a training facility for Swiss and international undergraduate and graduate students, post-doctoral and clinical fellows in collaboration with Swiss and Italian Universities.
The Institute is an integral component of the clinical research activities of the IOSI participating in translational studies and multi-center clinical trials and providing expertise, facilities and technical support for pharmacodynamics, pharmacogenomics and molecular follow-up studies.
Sources of Funding: Federal/State, Philanthropy, Industry Partnerships

Institutional Affiliation(s):

Address:
via Vela 6
Bellinzona
TI, 6500
Switzerland
Center/Program Start Date: 2003
Number of Employees: 50

P4T (Probes for Targets)
http://p4t.farmacia.unipr.it/
Therapeutic Approaches: Small Molecules
Therapeutic Areas of Expertise: Immunology, Infectious Diseases, Neurological Disorders, Oncology
Center/Program Highlights:
P4T (Probes for Targets) is an Academic Medicinal Chemistry Team specialized in the development of novel small-molecule of biological interest by integrating molecular modeling and combinatorial chemistry approaches.
Our mission is to provide new insights in the orphan and neglected diseases fields, with the goal of fulfilling unmet medical needs and bringing affordable new medicines to the developing world.
Our expertise lies in the following fields: anti-infectives, anticancer, neurological diseases, kinases targeting agents, drug-design and combinatorial chemistry.
Sources of Funding: Philanthropy, Industry Partnerships, University

Institutional Affiliation(s): University of Parma (IT)
Address:
University of Parma, Dipartimento di Farmacia
Viale delle Scienze, 27/A, 43124 Parma (Italy
Parma
Italy, 43124
Italy
Center/Program Start Date: 2011
Number of Employees: na

Centre of Applied Medical Research
https://cima.unav.edu/
Therapeutic Approaches: Biologics, Small Molecules, Vaccines, Biomarkers, Gene Therapy
Therapeutic Areas of Expertise: Cardiology, Gene-Therapy, Immunology, Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology, Hepatology;
Center/Program Highlights: The Center for Applied Medical Research (CIMA) is a biomedical research institution of the University of Navarra, based in Pamplona, Spain.
CIMAs research on cellular and molecular biology proudly aims to a deep translational patient orientation to solve unmet medical needs.
CIMA synergizes with the contiguous University Hospital and benefits from the experience of the University Schools of Medicine, Sciences and Pharmacy. This organization of academic, research and clinical knowledge represents an ideal environment for the translation of biological science into its clinical application for the benefit of the patient.
The Small Molecule Discovery platform: de-risking the drug-discovery process: Chemical Biology (identification of chemical probes hits- as pharmacological tools for target validation); Medicinal Chemistry - multifactorial optimization of proprietary hit compounds to achieve lead molecules and perform in-vivo PoC, in terms of preliminary efficacy and safety.
Sources of Funding: Federal/State, Philanthropy, Industry Partnerships, Disease Foundations
Institutional Affiliation(s): Foundation for Applied Medical Research, University of Navarra., University Hospital of the University of Navarra
Address:
Av Pio XII - 55
CIMA Building
Pamplona
Navarra, 31008
Spain
Center/Program Start Date: 2002
Number of Employees: 400

Health Research Institute of Santiago de Compostela

http://www.idisantiago.es/

Therapeutic Approaches: Small Molecules, Vaccines, Biomarkers, Stem Cells, Gene Therapy, Drug Delivery, eHealth/mHealth

Therapeutic Areas of Expertise: Cardiology, Gene-Therapy, Immunology, Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology, Ophthalmology, Psychiatric Disorders, eHealth/mHealth

Center/Program Highlights: The Health Research Institute of Santiago de Compostela (IDIS) is a traslational research, innovation and knowledge transfer centre located in Galcia, Northwest Spain. It was born as a joint venture between the University of Santiago de Compostela and the University Hospital complex of the same city. Created in January the 31st, 2008, it was awarded research centre of the Spanish National Health System in March 2010 by the Instituto de Salud Carlos III.

[701] persons integrate **[63]** research groups and **[1]** support group, organized as **[6]** research areas (Endocrinology, Epidemiology, Genetics and Systems Biology, Inflammation, Neurosciences and Oncology).

Our main aim: to make health research closer to your life! Our keys for success: to work close to the patient, within a hospital, secondary and primary health research environment and the efficient management of the (Fundaci□am□om□uez), body in charge of research management in the rea de Gesti□ntegrada de Santiago de Compostela (Servizo Galego de Sade) and the IDIS. Our upper lever of integration is CAMPUS VIDA, awarded International Excellence Campus in 2009 by the Spanish Ministerio de Educaci□Cultura y Deporte.

Sources of Funding: Federal/State, Philanthropy, Industry Partnerships, Disease Foundations, University

Institutional Affiliation(s): University of Santiago de Compostela, SERGAS (Galician Health Services Public Provider)

Address:
Clinical University Hospital of Santiago de C
Spain
Santiago de Compostela
A Coruña, 15706
Spain

Center/Program Start Date: 2008

Number of Employees: 706

Spanish National Research Council/Medicinal Chemistry Institute (IQM-CSIC)

http://www.iqm.csic.es/

Therapeutic Approaches: Small Molecules, Drug Delivery

Therapeutic Areas of Expertise: Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology, Pain

Center/Program Highlights: The Institute of Medical Chemistry (IQM), created in 1973, is an Institute of the Higher Council for Scientific Research (CSIC), belonging to the Department of Science and Chemical Technologies, which is located in the Center of Organic Chemistry "Manuel Lora-Tamayo ". The research activity of IQM, framed in the field of R & D Pharmaceutical aims the design and synthesis of new drugs. The training of research staff and collaborating with pharmaceutical companies are also activities in the late IQM. The result of their work is reflected in publications, patents, doctoral theses and doctoral courses and specialization.

Sources of Funding: Federal/State, Industry Partnerships, Disease Foundations

Institutional Affiliation(s): Department of Pharmacology, University of Alcalá (Spain) , Department of Pharmacology, University Rey Juan Carlos(Spain)

Address:
Juan de la Cierva 3
Madrid
Madrid, E-28006
Spain

Center/Program Start Date: 1974

Number of Employees: 70

Spanish National Cancer Research Centre (CNIO) / Experimental Therapeutics

https://www.cnio.es/en/

Therapeutic Approaches: Small Molecules

Therapeutic Areas of Expertise: Oncology

Center/Program Highlights: The Experimental Therapeutics Programme (ETP) within the CNIO (Spanish National Cancer Centre) was established in 2006 to foster translation of scientific breakthroughs into novel and more effective ways of prevent, diagnose and treat cancer. Its main role is to translate fundamental understanding of cancer into novel drugs.

For the past ten years, the ETP was completely focused on delivering balanced and optimized lead compounds (activity, off-target selectivity and ADME-Tox properties) with demonstrated in vivo proof-of-concept (PoC) after oral administration (mechanism of action and efficacy in tumor models).

Our multidisciplinary team of medicinal chemists, biologists, pharmacologists, crystallographers, etc. has proven skills in early drug discovery ranging from High-Throughput Screening (HTS) assay development (library of 50K chemically diverse compounds), medicinal chemistry for hit generation, hit-to-lead and lead optimization, to in vivo pharmacokinetics and PoC studies in animal models of human cancer.

By working closely with CNIO and other external basic research scientists, several lead compounds for PIM, PI3K and ATR kinases were progressed into preclinical phases, which have triggered licensing agreements with companies such as Inflection Biosciences and Merck-Serono.

The current ETP-CNIO pipeline encompasses targeted projects directed towards the discovery of novel inhibitors for CDK8 (Lead Optimization), HASPIN and MASTL kinases (Hit Generation phase) and projects based on phenotypic screenings such as the identification of TRF1 (shelterin) modulators and novel targets and inhibitors to erradicate Cancer Stem Cells (CSCs). In addition to classical drug discovery approaches for targeted projects, our team apply target deconvolution techniques for phenotypic projects, including the design of affinity and imaging chemical probes to help in the identification of molecular targets responsible for a given phenotype.
Sources of Funding: Federal/State, Industry Partnerships, Revenue

cnio Centro Nacional de Investigaciones Oncológicas

Institutional Affiliation(s):
Address:
Melchor Fernandez Almagro, 3
Madrid
Madrid, 28029
Spain
Center/Program Start Date: 2006
Number of Employees: 400/23

Research Institute for Medicines (iMed.ULisboa)

http://imed.ulisboa.pt/
Therapeutic Approaches: Biologics, Small Molecules, Vaccines, Biomarkers, Antibodies, Stem Cells, Gene Therapy, Drug Delivery
Therapeutic Areas of Expertise: Gene-Therapy, Immunology, Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology
Center/Program Highlights: The Research Institute for Medicines Instituto de Investiga磯 do Medicamento (iMed.ULisboa) is a multidisciplinary R&D Unit in Life and Health Sciences, hosted at Faculdade de Farm殀a, Universidade de Lisboa (FF/ULisboa).
Our mission is to develop innovative medicines and benefit human health through top-class multidisciplinary research.
Objectives
- Encourage and support high-quality multidisciplinary research to improve human health and benefit quality of life;
- Promote commercial exploitation of our discoveries to improve economic competitiveness;
- Attract and train highly skilled researchers;
- Outreach the general public and schools;
- Promote good practices, striving for continuous improvements, including evaluation of all aspects of performance.
Sources of Funding: Federal/State, Industry Partnerships, University

iMed. ULisboa Research Institute for Medicines

Institutional Affiliation(s): Faculty of Pharmacy, University of Lisbon
Address:
Av. Prof Gama Pinto
Lisbon

Lisbon, 1649-003
Portugal
Center/Program Start Date: 0000
Number of Employees: 000

Experimental Pharmacology Laboratory

https://www.dietislab.org/
Therapeutic Approaches: Small Molecules, Drug Repurposing
Therapeutic Areas of Expertise: Neurological Disorders, Oncology, Pain, Antimicrobial resistance
Center/Program Highlights: The University of Cyprus *Experimental Pharmacology Laboratory* is a molecular and behavioural pharmacology laboratory based in the Medical School of the University of Cyprus.
Our research revolves around novel drug discovery and drug repurposing screenings, both in vitro and in vivo. We focus on the exploration of the various effects of drugs on neural pathology (such as pain, cancer and neurodegeneration) and on the associated organism behaviours. We apply this knowledge to discover novel drugs or find new therapeutic applications for existing clinical drugs against chronic diseases. Our lab is housed in a newly-built laboratory facility
at the "Siakolas Educational Centre for Clinical Medicine" building in Nicosia, Cyprus.
Sources of Funding: University

Institutional Affiliation(s): University of Cyprus Medical School
Address:
1 University Avenue
Aglantzia, Nicosia
Nicosia
Nicosia, 2109
Cyprus
Center/Program Start Date: 2016
Number of Employees: 8

BLAVATNIK CENTER for Drug Discovery

https://bcdd.tau.ac.il/
Therapeutic Approaches: Small Molecules, Peptides, Image-based HCS analysis, HT qRT-PCR and miRNA microarray assays, Target-structure modeling, Molecular modeling, Structure based drug design
Therapeutic Areas of Expertise: Neurological Disorders, Oncology, Rare Diseases
Center/Program Highlights: The BLAVATNIK CENTER for Drug Discovery (BCDD) is a fully integrated center for translational research, based at Tel Aviv University, that comprises medicinal chemistry, biological target identification and computer-assisted drug design units under one roof. The BCDDs commitment is to provide support for early-stage drug discovery and translating academic innovations into new drug candidates.
The BCDD provides state-of-the-art research environment and academic platform, expertise and and infrastructure to help researchers develop disease models/biological assays for drug screening and to advance promising early-stage drug candidates through the optimization/target validation pipeline.
Sources of Funding: Federal/State, Philanthropy, University

Institutional Affiliation(s): Tel Aviv University
Address:
BLAVATNIK CENTER for Drug Discovery
Tel Aviv University
Tel Aviv
Israel, 6997801
Israel
Center/Program Start Date: 2016
Number of Employees: 14

The Nancy and Stephen Grand Israel National Center for Personalized Medicine

http://g-incpm.weizmann.ac.il/drug-discovery/maurice-and-vivienne-wohl-institute-drug-discovery
Therapeutic Approaches: Small Molecules

Therapeutic Areas of Expertise: Cardiology, Immunology, Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology, Psychiatric Disorders, Regenerative Medicine
Center/Program Highlights: The facility is organized to meet the needs expressed by the Weizmann Institute scientific community, researchers from other academic organizations, as well as biotech and pharmaceutical companies. These needs range from high-throughput screening of chemical compounds on complex cellular readouts, to functional testing of a limited subset of molecules in doses or combinations in a defined model, to supporting medicinal chemistry on an existing compound associated with the need to develop an appropriate bioassay
Sources of Funding: Federal/State, Philanthropy, University, Revenue

המרכז הישראלי הלאומי לרפואה מותאמת אישית ע"ש ננסי וסטיבן גרנד
The Nancy & Stephen Grand Israel National Center for Personalized Medicine

Institutional Affiliation(s): The Weizmann Institute of Science
Address:
234 Herzl Street
Rehovot
Israel, 7610001
Israel
Center/Program Start Date: 2013
Number of Employees: 6

Genetic engineering and biotechnology

Therapeutic Approaches: Biologics, Small Molecules, Biomarkers, Gene Therapy
Therapeutic Areas of Expertise: Gene-Therapy, Metabolic Disorders, Oncology, Regenerative Medicine
Center/Program Highlights: Personalize medicine. For alzahmar disease in Saudi
Sources of Funding: University
Institutional Affiliation(s): Shaqra university
Address:
Shaqra university / ksa
Riyadh
Riyadh, 11353/49459
Saudi Arabia
Center/Program Start Date: 2012
Number of Employees: 12

Drug Discovery Initiative

https://www.ddi.u-tokyo.ac.jp/en
Therapeutic Approaches: Small Molecules
Therapeutic Areas of Expertise: Infectious Diseases, Metabolic Disorders, Neurological Disorders, Oncology
Center/Program Highlights: Drug Discovery Initiative (DDI) has collected 220K chemical compounds so far. These are available for screening and shipped to researchers around Japan. DDI also provides advice and the use of its facilities.
Sources of Funding: Federal/State

Drug Discovery Initiative DDI

Institutional Affiliation(s): The University of Tokyo
Address:
7-3-1 Hongo, Bunkyo-ku
Tokyo
Tokyo, 113-0033
Japan
Center/Program Start Date: 2007
Number of Employees: 19

CO-ADD (Community for Open Antimicrobial Drug Discovery)

http://www.co-add.org/
Therapeutic Approaches: Small Molecules
Therapeutic Areas of Expertise: Infectious Diseases
Center/Program Highlights: CO-ADD is a not-for-profit initiative led by academics.

Our goal is to screen compounds for antimicrobial activity for academic research groups from anywhere in the world for free. We make no claim to IP or results.
We aim to help researchers around the world to find new, diverse compounds to combat the superbug crisis.
CO-ADD is a global initiative of The University of Queenslands Institute for Molecular Bioscience and is funded by the Wellcome Trust and The University of Queensland. We are reaching out to chemists in academia and research organisations who have compounds sitting on shelves that were not designed as antibiotics and would not otherwise be screened for antimicrobial activity.from academic research groups
Sources of Funding: Philanthropy, University

Institutional Affiliation(s): The University of Queensland, Wellcome Trust
Address:
Institute for Molecular Bioscience
The University of Queensland Bld 80
Brisbane
QLD, 4072
Australia
Center/Program Start Date: 2015
Number of Employees: 15

ACRF Drug Discovery Centre

https://ccia.org.au/
Therapeutic Approaches: Small Molecules
Therapeutic Areas of Expertise: Immunology, Metabolic Disorders, Oncology
Sources of Funding: Federal/State, Philanthropy, Industry Partnerships, Disease Foundations, University, Revenue
Institutional Affiliation(s): Children's Cancer Institute Australia for Medical Research
Address:
PO Box 81
Randwick
Sydney
NSW, 2031
Australia
Center/Program Start Date: 2010
Number of Employees: 3

ACRF Rational Drug Discovery Centre

https://www.svi.edu.au/research_themes/structural_biology
Therapeutic Approaches: Biologics, Small Molecules
Therapeutic Areas of Expertise: Infectious Diseases, Neurological Disorders, Oncology
Center/Program Highlights: The ACRF Rational Drug Discovery Centre at St Vincent's Institute focuses on early stage drug discovery for cancer therapeutics with other programs in neuroscience and infectious disease. The Centre has expertise in structural biology, biophysical analysis (SPR, MST, ITC) and computational drug design.
Sources of Funding: Federal/State, Philanthropy, Industry Partnerships

Institutional Affiliation(s): St Vincent's Institute for Medical Research
Address:
9 Princes Street
Melbourne
Victoria, 3065
Australia
Center/Program Start Date: 2013
Number of Employees: 15

Institut de Pharmacologie de Sherbrooke (IPS)

https://www.usherbrooke.ca/ips/fr/

Therapeutic Approaches: Biologics, Small Molecules, peptides, macrocycles
Therapeutic Areas of Expertise: Cardiology, Infectious Diseases, Oncology, Pain

Center/Program Highlights: Institut de Pharmacologie de Sherbrooke focuses on the discovery of new drugs and new diagnostic agents , as well as the validation of emerging drug targets and new technologies to support drug discovery. Major areas of focus include cardiovascular diseases, pain, brain and prostate cancer, and infectious diseases. A particular area of expertise is GPCRs, proteases, transcription factors, peptidic and macrocyclic chemistry as well as various imaging modalities.
Sources of Funding: Federal/State, Industry Partnerships, University

Institutional Affiliation(s): Université de Sherbrooke
Address:
Institut de pharmacologie de Sherbrooke
3001, 12e av nord
Sherbrooke
Québec, J1H 5N4
Canada
Center/Program Start Date: 1998
Number of Employees: 250