



# Crowdsourcing and open innovation in drug discovery: recent contributions and future directions

David C. Thompson<sup>1</sup> and Jörg Bentzien<sup>2</sup>

<sup>1</sup>Chemical Computing Group, Montreal, QC H3A 2R7, Canada

<sup>2</sup>Alkermes, Inc. 852 Winter Street, Waltham, MA 02451-1420, USA



The past decade has seen significant growth in the use of ‘crowdsourcing’ and open innovation approaches to engage ‘citizen scientists’ to perform novel scientific research. Here, we quantify and summarize the current state of adoption of open innovation by major pharmaceutical companies. We also highlight recent crowdsourcing and open innovation research contributions to the field of drug discovery, and interesting future directions.

## Introduction

The ecosystem required to cultivate the discovery, development, distribution, and marketing of novel chemical material is a complex-adaptive system [1]. 2018 saw a record number of new drug approvals from US regulators, exceeding the previous record of 50 set in 1996 [2], while 48 new drugs were approved in 2019. It is too soon, and too complex an environment, to distinctly attribute patterns of {system, structure, and process} manifest in the ecosystem that have contributed to this significant rise in output. However, in a recent contribution looking to explore this significant increase in approvals, the role and continued promise of crowdsourcing was highlighted. Specifically: ‘In fact, it is not a stretch to imagine that, in the not-so-distant future, much of drug discovery will be crowdsourced. It is a more efficient model to explore disease space and identify opportunities than the inflexible programs of many large companies, which are driven by the needs of their marketing franchises with scant consideration to what science can deliver.’ [3].

Crowdsourcing as a common term was introduced in a 2006 issue of the popular technology magazine *Wired* [4], wherein it was described as an internet-enabled approach that harnessed the ability of agents external to an organization. However, examples of crowdsourcing pre-date the internet, with one of the best-known being the British Government’s establishment of the Longitude Act during the 18th Century. The winning solution, a device able to specify longitude at sea to an unprecedented

accuracy (the chronometer), came from an unexpected source, John Harrison, a carpenter and clockmaker by training [5]. This highlights an important feature of crowdsourcing as an organizational resource: solutions can come from unlikely quarters [6,7].

Closely related to crowdsourcing are the practices of open innovation and citizen science. A formal description of open innovation was made by Chesbrough in 2003, such that: ‘[it is] a paradigm that assumes that firms should use external ideas as well as internal ideas, and internal and external paths to market, as the firms look to advance their technology.’ [8]. This definition was subsequently refined by Chesbrough, wherein he focused on describing an intentional procedural act: ‘a distributed innovation process based on purposively managed knowledge flows across organizational boundaries, using pecuniary and non-pecuniary mechanisms in line with the organization’s business model.’ [9].

Citizen science, the involvement of the general public in research, is a specific example of crowdsourcing, focused solely on the sciences. Our prior example of the determination of longitude is an example of both the practice of citizen science and of crowdsourcing.

In January 2017, the American Innovation and Competitive-ness Act (AICA) became law, with Section 402 of the AICA, the Crowdsourcing and Citizen Science Act, providing broad authority to Federal agencies to use crowdsourcing, specifically citizen science, to advance those agency missions and to foster greater public engagement. Specifically, the Crowdsourcing and Citizen Science

Corresponding author: Thompson, D.C. (dthompson@chemcomp.com)

Act defines crowdsourcing as: ‘a method to obtain needed services, ideas, or content by soliciting voluntary contributions from a group of individuals or organizations especially from an online community.’

Furthermore, the Government Accountability Office (GAO) defines open innovation: ‘[as encompassing] activities and technologies to harness the ideas, expertise, and resources of those outside an organization to address an issue or achieve specific goals.’

That these tools have been formally codified by the US Government highlights both their utility and growing role in the practice of ‘innovation’ and broadening public participation [10]. There are numerous obvious similarities between these definitions of crowdsourcing and open innovation such that, on an abstract level, they represent structured ways for focused engagement to occur across organizational boundaries. Open innovation, as defined earlier, remains the broadest, most inclusive, approach for involvement of an ‘Other’, and the term that could be used to most readily provide a framing for the broader strategic intent of an organization. For the sake of clarity, the term ‘open innovation’ will be used throughout, and interchangeably, with the terms ‘crowdsourcing’ and ‘citizen science’, unless a specific facet of a given approach warrants explicit description. Similarly, for the remainder of this article, where definitions are important, those adopted by the US Government are used.

Here, we review a recently published contribution wherein a strategic framework was posited regarding the utilization by an organization of crowdsourcing methodologies [11]. The framework itself provides a unifying view on how organizations could meaningfully engage a group of ‘Others’, and the hope remains that the framework serves as an organizing principle for leaders to coordinate, strategize, and operationalize ongoing ‘networked activities’. We revisit this framework in the context of open innovation to validate its utility given this broader definition, making what was an implicit connection explicit in its domain of application.

We then proceed to review the use of open innovation approaches by the pharmaceutical industry, exploring the current state of adoption, highlighting recent successes and challenges appropriately. Following this, we present several recent uses of open innovation within the life sciences, with a focus on drug discovery and development. This leads into a discussion of future directions. Specifically, the following four areas are discussed: (i) the extent to which the systems, structures, and processes of open innovation are intended to enable ‘serendipity’; (ii) the increase in the number of open innovation activities where crowd-developed hypotheses are physically tested (an interesting bridging of the physical and digital); (iii) the professionalization of ‘serious games’, and the interesting approach of ‘meta gamification’: the embedding of a puzzle and/or activity, intended for engagement with the crowd into a game and/or activity that is itself a pre-existing networked platform; and (iv) the adoption of Artificial Intelligence (AI) and the possibility of ‘crowd-in-the-loop’ processes. Finally, we offer concluding remarks and discussion.

### Open innovation: a network perspective

In a recent contribution, a strategic framework was introduced that aligned crowdsourcing with more traditional elements of how

the pharmaceutical industry looks beyond its organizational boundaries [11]. What follows here is a brief recap of this framework, and some discussion on its applicability to open innovation activities. Given that open innovation is the lens through which the current work is presented (because of its more inclusive framing, as described earlier), it is appropriate to ensure that the framework, as originally suggested, is consistent with this choice, and to ensure that it is not theoretically lacking.

The grounding element of the framework is the ‘Organization’. This is defined as a collection of semi-autonomous ‘Actors’, engaged in ‘work’ that results in the delivery of a service(s) or product(s) to customers. For this work, Actors not formally part of the Organization, with no role in work directly relating to the delivery of the service(s) or product(s) consumed by the Organizations customers, are to be considered ‘Other’.

Accordingly, one can arrange relationship patterns between Organization and Other according to an abstract ‘coupling constant’. Such a coupling constant represents how closely connected Organization is to Other, and relates novel crowdsourcing activities of the Other to more traditional approaches of insourcing external innovation from the Other. In its original form, as presented, this contained four levels: (i) weakly coupled, (e.g., crowdsourcing activities); (ii) partially coupled (I) (e.g., consortia participation); (iii) partially coupled (II), (e.g., academic collaboration and open innovation); and (iv) strong coupled (e.g., hosted postdoctoral programs).

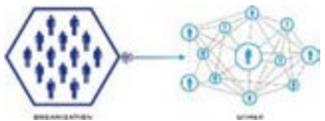
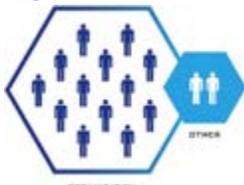
This is further summarized in Table 1, and more information can be found in [11]. In revisiting this framework, the original four elements remain consistent when viewed through the broader lens of open innovation. As such, this framework provides leadership within a life sciences environment a flexible structure for organizing, coordinating, and measuring the collective impact of open innovation activities, of which ‘crowdsourcing’ and ‘citizen science’ might be activities that occur (indeed, this framework could readily be used outside of a life sciences context with little or no modification).

Ultimately, the suggested framework allows leadership to view open innovation activities through a portfolio lens. The next question to ask might be: how is a leader best able to measure the success of both the portfolio, and the individual contributing elements? Several recent contributions provide insight into this.

The first contribution, specifically focused on understanding the impact of an open innovation platform [12], recognizes that the process of drug discovery and development contains a natural inflection point at the candidate pre- and post-selection event. Work occurring before candidate selection has a distinct ‘learning cycle’ characteristic of ‘plan–do–study–act’, and is applied to possibly thousands of distinct molecular entities (small molecule or biologic in nature). As a necessarily exploratory process at the intersection of biological–chemical space, this work is nonlinear in nature. This stands in distinct contrast to the process post-selection, wherein the focus is a well-defined single molecular entity and measurement is grounded in clinical success/attrition. Given this observation, the authors of [12] explore a general open innovation measurement structure comprising four elements: (i) investment; (ii) pipeline health; (iii) returns; and (iv) culture and capabilities. Alternate measurement schemes have previously

TABLE 1

Summary of a network perspective on open innovation<sup>a</sup>

Column heading	The four levels of networked open innovation
Weakly coupled (e.g., crowdsourcing)	
Partially coupled (I), (e.g., consortia participation)	
Partially coupled (II), (e.g., academic collaboration)	
Strongly coupled, (e.g., hosted postdoctoral program)	

<sup>a</sup> See [11] for more details.

been described [13] and could readily be layered onto the strategic framework described herein.

An important facet of the framework is the orientation of the Organization with respect to academic Other. This is especially pertinent given that it has been shown that academic institutions contribute most preapproval publications, with publication subject matter being closely aligned with the strength of the principal investigator. This broadly supports the hypothesis that enhanced investment and collaboration with the external academic community is likely to contribute to the more expedient discovery, development, and ultimate approval of innovative medicines [14].

### The current state of open innovation in practice: a strategic perspective

Although theoretical discussions are important in providing a common language to clearly transmit, with high fidelity, understandable intentions (e.g., between Organization and Other) [15], what matters most is what is actually done. The format in which companies engage 'Other' will differ between organizations and depend crucially upon their objective(s). The following analysis is intended to highlight the extent to which organizations are engaging in open innovation in a coordinated and strategic fashion. This, in turn, could be an indication of the importance that organizations are placing on these external-facing activities in the service of their ultimate objectives: the discovery, development, and delivery of novel therapeutic agents.

To that end, we surveyed the top 20 biomedical companies, those with revenues in excess of US\$10 billion. For each company, we performed a simple web-based search, using the name of the company as listed plus the phrase 'open innovation' and

the DuckDuckGo engine. For each search performed, the first page of results was explored, and the most relevant 'artefact' selected for inclusion in Table 2. Although such choices are subjective, they were guided by the following criteria: (i) the search result was the first most relevant result relating to an open innovation activity; and (ii) the hit was on the first page of the search results.

Results were explored with respect to the presence/absence of the following three related items: (i) a link relating to open innovation; (ii) within the link described in (i), a clear actionable step for a citizen scientist to engage in; and (iii) the presence of a well-articulated open innovation framework (OIF) that contains all or most of the items described in our strategic network-oriented framework (i.e. opportunities for external engagement at a variety of different levels of 'coupling').

The results of this simple investigation are detailed in Table 2. It was found that 65% (13/20) of the top 20 biomedical companies had an easily discoverable web artefact relating to open innovation and a clear actionable step for a citizen scientist to engage in [(i) and (ii) above]. We should point out that we do not differentiate 'levels of action' within this category, and note a significant amount of variance, from a simple 'Contact Us' option focused on business development opportunities (e.g., [www.amgenbd.com/](http://www.amgenbd.com/)), through to access to tool compounds (e.g., [www.openinnovation.abbvie.com/web/compound-toolbox](http://www.openinnovation.abbvie.com/web/compound-toolbox) or Boehringer Ingelheim's <https://openme.com/>). We found that 25% (5/20) of the top 20 biomedical companies by revenue have all three items [(i–iii) described above].

Table 2 was initially populated on October 22, 2019, and then re-checked on November 26, 2019. Interestingly, the original

TABLE 2

**The top 20 biomedical companies by revenue, indicating the extent to which they have clear open innovation artefacts readily found on the web, actionable next steps, and fully articulated open innovation strategies<sup>a</sup> [12,16,17,68–71].**

Company	Rank <sup>b</sup>	Open innovation artefact <sup>c</sup>	Click through <sup>d</sup>	OIF <sup>e</sup>	Relevant Refs
Johnson & Johnson	1	<a href="https://jnjinnovation.com/">https://jnjinnovation.com/</a>	Y	Y	
Roche	2	<a href="http://www.roche.com/partnering/about_partnering_at_roche.htm">www.roche.com/partnering/about_partnering_at_roche.htm</a>	Y	N	
Pfizer	3	<a href="http://www.pfizercti.com/">www.pfizercti.com/</a>	N	N	
Novartis	4	<a href="http://www.novartis.com/our-science/novartis-institutes-biomedical-research/collaborations">www.novartis.com/our-science/novartis-institutes-biomedical-research/collaborations</a>	N	N	
Bayer	5	<a href="https://innovate.bayer.com/">https://innovate.bayer.com/</a>	Y	Y	[66,67]
GlaxoSmithKline	6	<a href="https://themind.gsk.com/">https://themind.gsk.com/</a> <sup>g</sup>	Y	N	
Merck & Co.	7		N	N	
Sanofi	8	<a href="https://innovatewith.sanofi/">https://innovatewith.sanofi/</a> <sup>h</sup>	Y	N	
AbbVie	9	<a href="http://www.openinnovation.abbvie.com/">www.openinnovation.abbvie.com/</a>	Y	N	
Abbott	10	<a href="http://www.openinnovation.abbvie.com/">www.openinnovation.abbvie.com/</a>	Y	N	
Eli Lilly & Co.	11		N	N	[12,16,17]
Amgen	12	<a href="http://www.amgenbd.com/">www.amgenbd.com/</a>	Y	N	
Bristol-Myers Squibb	13	<a href="http://www.bms.com/researchers-and-partners/partnering.html">www.bms.com/researchers-and-partners/partnering.html</a>	Y	N	
Gilead	14		N	N	
AstraZeneca	15	<a href="https://openinnovation.astrazeneca.com/">https://openinnovation.astrazeneca.com/</a>	Y	Y	
Teva Pharmaceuticals <sup>i</sup>	16		N	N	
Boehringer Ingelheim	17	<a href="http://www.boehringer-ingenheim.com/innovation/research_development/open-innovation">www.boehringer-ingenheim.com/innovation/research_development/open-innovation</a>	Y	Y	[68,69]
Takeda	18		N	N	
Merck Group	19	<a href="http://www.emdgroup.com/en/research/open-innovation.html">www.emdgroup.com/en/research/open-innovation.html</a>	Y	Y	
Novo Nordisk	20	<a href="https://novonordiskinnovationchallenge.com/">https://novonordiskinnovationchallenge.com/</a>	Y	N	

<sup>a</sup>Companies shaded green are XXXXX.

<sup>b</sup>This rank is taken from a list of the largest biomedical companies by revenue, found here: [https://en.wikipedia.org/wiki/List\\_of\\_largest\\_biomedical\\_companies\\_by\\_revenue](https://en.wikipedia.org/wiki/List_of_largest_biomedical_companies_by_revenue) (Accessed October 22, 2019).

<sup>c</sup>The most relevant hit on the first page of search results returned from a DuckDuckGo search for 'company open innovation' searches performed on October 22, 2019. Validated on November 26, 2019; changes made to GSK and Novo Nordisk, as discussed in the main text.

<sup>d</sup>Is there a meaningful activity for a 'citizen scientist' to engage in, directly from the website? For example, access a challenge, submit a contact request, etc.

<sup>e</sup>Is there a readily apparent, clearly articulated OIF available from the open innovation artefact?

<sup>f</sup>Consumer healthcare focused.

<sup>g</sup>Interestingly, a search for "Teva Open Innovation" returned the following: <https://openinnovation.corteva.com/>.

Shaded entries correspond to those that have all three of the following elements present: (1) A link on the visited website relating to Open Innovation; and (2) Within the link described in (1) a clear actionable step for a citizen scientist to engage in, and (3) a well-articulated Open Innovation framework.

results from GSK and Novo Nordisk had changed, highlighting that these web artefacts are: (i) dynamically evolving, and (ii) not guaranteed to be persistent as resources, indeed the original link from the GSK search generated in October no longer resolves (i.e., <https://openinnovation.gsk.com/>).

This is particularly pointed when considering the case of Eli Lilly & Co. a pioneer of open innovation, with their early entry into, and success with, the Open Innovation and Drug Discovery (OIDD) platform, but for which seemingly no trace is now found online, despite numerous earlier publications and an open way of documenting available web resources [12,16,17] (e.g.: <https://openinnovation.lilly.com/dd/what-we-offer/>). Unfortunately, organizational learning of this type is often lost as there is no perceived benefit to it being shared externally. Indeed, there is a large literature on the concept of the 'selective reveal', specific to open innovation [18]. This is further highlighted by the fact that only 15% of the companies in Table 2 (3/20) have published on their open innovation approaches in peer-reviewed journals. Of these three, only two have an easily discoverable web artefact relating to open innovation, clear actionable steps for a citizen scientist to engage in and a well-articulated OIF. This represents 10% (2/20) of the full set, and 40% (2/5) for those companies for which all three criteria are satisfied.

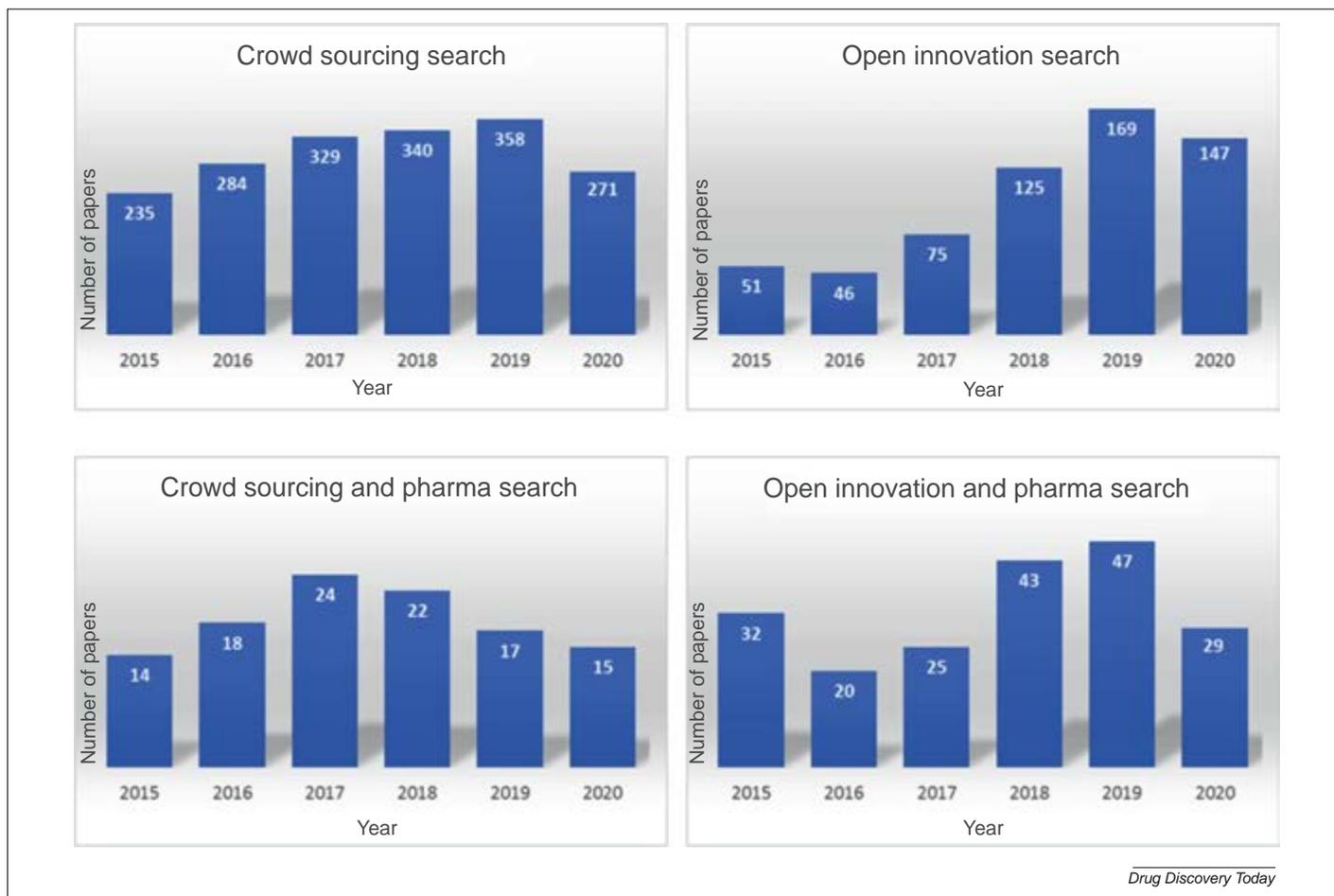
It is interesting to explore those organizations for which all three criteria are satisfied through the lens of historically benchmarked

composite metrics, such as the Pharmaceutical Innovation and Inventiveness Index created and monitored by IDEA Pharma, a global strategic consulting firm. They derive measures for innovation and inventiveness (<https://www.ideapharma.com/pii>), defining innovation as the ability of a company to bring products from Phase I/II to market and successful commercialization, whereas inventiveness examines pipeline novelty. Invention defined in this way is a more forward-looking measure.

In examining the five organizations from Table 2 satisfying all three criteria described above, we find that Bayer, Boehringer Ingelheim, and the Merck Group appear to be laggards with respect to both innovation and inventiveness as defined by this index, whereas AstraZeneca leads the cohort with respect to inventiveness and is in the top 10 with Johnson & Johnson with regards to Innovation. It might be possible, in the future, to explore with respect to these indices, the performance of Bayer, Boehringer Ingelheim, and the Merck Group (clear proponents of open innovation) as a function of this inventiveness and innovation index.

### The current state of open innovation in practice: a tactical perspective

In the preceding section, we explored the wholesale adoption (or not) of the use of open innovation as a strategic practice, at the organizational level. Alternately, one could look at the more

**FIGURE 1**

Results of PubMed searches using 'crowdsourcing' and 'open innovation' search terms with and without the qualifier 'pharma'. Shown are the number of papers published between 2015 and 2020 for each of the categories indicated. (Data collected January 25, 2020.)

tactical use of open innovation and crowdsourcing, as reflected in its reported use online and in the literature. To this end, we performed PubMed searches using the key word terms: {crowdsourcing and open innovation} both with and without the additional qualifier 'pharma.' We focused our attention on 2015 through to 2020 simply because this was the last time, we, the authors, explored this topic.

Figure 1 highlights an increase in the number of papers published since 2015 containing these terms both with and without the 'pharma.' qualifier. The search term 'open innovation' showed continued and increasing interest, although it is of course realistic to expect that null results are not published, or more broadly perhaps the practice of 'selective revealing' [18] is engaged in to preserve a (perceived) competitive advantage. More sophisticated text search analysis of different areas in which crowdsourcing appears have been performed, and described the field as 'nascent' and having 'high potential' [19,20].

In addition to exploring the published material related to the use of, engagement with, and experience of crowdsourcing and open innovation by an organization, we looked at a variety of platforms that are commonly used by organizations for crowdsourcing and open innovation activities. We focus our attention

on the three largest such platforms: Kaggle, InnoCentive, and DREAM.

### *Kaggle*

Kaggle is an online community centered around participation in machine-learning competitions. The site uses many facets of gamification to enable and showcase mastery of the practice of data science. Currently, the Kaggle website ([www.kaggle.com](http://www.kaggle.com)) shows 370 competitions, 11 of which were active at the time of this publication, and none of those 11 are linked to healthcare. In Table S1 in the supplemental information online, we present a detailed view of all Kaggle competitions published online that were in some way linked to the health sciences. The year 2019 saw a large number of such competitions, primarily dealing with disease detection and image analysis. Only a few pharmaceutical companies (Boehringer Ingelheim, Merck, Pfizer, and Genentech) have openly used the Kaggle platform for crowdsourcing activities. Two of these organizations contained all three items as described in the previous section: an active weblink relating to open innovation; within that link a clear actionable step for a citizen scientist to engage in; and the presence of a well-articulated OIF. In general, a simple count from a public-facing online

resource might represent a lower bound on actual participation because it is possible to host private competitions wherein the organizational identity is blinded.

Data in Table S1 in the supplemental information online highlight a consistently high level of engagement between the Kaggle community and healthcare-related competitions, with many competitions having >1000 actively participating teams. Given the general level of enthusiasm for machine learning, data science, and artificial intelligence (see later section), and the ability to readily transform many problems in healthcare into a form tractable to analysis by data scientists [21], it is unsurprising that participation in these competitions is high.

### InnoCentive

InnoCentive is an online community, founded in 2001 as a spinout from Eli Lilly and Co.'s Internet Incubator. It is a platform wherein challenges are posted, and solvers submit solutions, and the challenges span a range from those that can be assessed against an objective assessment of 'better' or 'correct', through to a more subjective evaluation. More historic details on the platform can be found elsewhere [7].

Currently, only challenges from 2019–2020 are shown on the InnoCentive website. In Table S2 in the supplemental information online, we highlight the 49 challenges relating to healthcare. Of those, seven are open and 42 are under evaluation. A white paper published by InnoCentive during late 2019 highlighted the historic use of its platform in all phases of the drug discovery process [22]. Exploring the public-facing data shows that two companies have used this platform during the stated time period (AstraZeneca and Merck), with one of them having used it ten times (across a wide spectrum of problem classes) since 2017. Both of these organizations have exhibited all three items we use to define the presence of a fully realized open innovation strategy.

### DREAM

Unlike the Kaggle and InnoCentive Crowdsourcing platforms that host challenges across a wide spectrum of industries and sciences, the DREAM challenges are focused on biology and medicine. These challenges started in 2006 and are supported by Sage Bionetworks. In a recent correspondence introducing the DREAM Malaria challenge, the authors not only highlight the potential of crowdsourcing, but also pointed to some biases with respect to disease areas [23]. As such, they reported that, among the open Kaggle challenges at that time, only one was focusing on a disease, cancer, and of the DREAM challenges, none focused on infectious diseases but 15 on cancer. Some of this might have to do with the ease of availability of large data but is most likely a reflection of where scientific focus is in general, and not a reflection of the applicability of crowdsourcing *per se* in certain areas. The DREAM challenges are mostly organized by academia and not-for-profit organizations and the incentives are mostly non-monetary in the form of co-authorship of highly ranked literature publications and conference attendance. AstraZeneca appears to be the only for-profit pharmaceutical organization to have utilized a DREAM challenge, further highlighting the commitment of this company for engagement with external Other(s).

Table S3 in the supplemental information online highlights challenge title, launch and close dates, challenge poster, incentives

(where applicable), and participant numbers where stated. This information was taken directly from the DREAM website (<http://dreamchallenges.org/>). The DREAM challenges and the concept of crowdsourcing have been reviewed more thoroughly elsewhere [24]. The concept of community and its role in advancing science is a key theme running through the presentation of the DREAM challenges, and this is further realized through the use of the Synapse platform ([www.synapse.org](http://www.synapse.org)) to support transparent communication regarding the challenges, their progress, and community goals against the stated aim.

BioMed X, an innovative incubator company, is another approach used by pharmaceutical companies to find innovative solutions. The company currently lists Abbvie, Boehringer Ingelheim, Janssen, Johnson & Johnson, Merck, and Roche as sponsors on its website (<https://bio.mx/>). BioMed X is active in a diverse set of therapeutic areas, including oncology, immunology, neuroscience, and respiratory. Several projects have been completed. This approach appears to be successful and is active, having announced a new research program with Boehringer Ingelheim for schizophrenia and with Merck for autoimmune diseases.

In the context of the network-oriented strategic framework, the above represent specific examples of 'weakly correlated' engagement: the interests of the participants engaging in open innovation might not be wholly aligned with that of the broader organization, and the motivation might be driven by financial, positional, or reputational considerations.

### Precompetitive agreements or consortia

Another tactical avenue frequently explored by organizations is that of precompetitive agreements or consortia. In these examples, the interests of the parties are clearly more aligned, and we described this earlier as an example of a 'partially coupled' networked interaction (Table 1). Here, we summarize several recently published examples of such partially coupled Open innovation activities.

The Investigative Toxicology Leaders Forum (ITLF) is working to develop investigative toxicology concepts and practices for decision-making related to early safety [25]. Precompetitive agreements appear to be often used in cases of general scientific interest where intellectual property is less of a concern. The Open PHACTS Discovery Platform, which looks at the analysis of biological pathways, is one example in this area [26]. A very successful consortium is the Structural Genomics Consortium (SGC), which enabled a large number of new targets by solving and providing X-ray structures to the community [27]. There are several cases where companies share scientific probes and data with the community [28]. The traditional drug discovery model is such that data sharing between competitive organizations is not typically incentivized. To allow data sharing even in these situations, a third party can be set up as a trusted broker, who communicates between the parties using a cheminformatic approach, for matched molecular pair analysis [29].

Sometimes, these precompetitive collaborations are set up under the umbrella of a public-private partnership (PPP). One large PPP is the Innovative Medicines Initiative (IMI) by the European Union [30]. Several projects are part of the IMI, such as the U-BIOPRED program, which addresses fundamental issues around asthma [31] or eTOX, which generated ~200 *in silico* models for

diverse end-points of toxicological interest [32], and the European Lead Factory (ELF), which has developed a large compound library available for screening [33]. WIPO Re:Search is a PPP co-led by BIO Ventures for Global Health (BVGH) and the World Intellectual Property Organization (WIPO). WIPO Re:Search supports research for neglected tropical diseases, an area of limited commercial incentives but high unmet need [34].

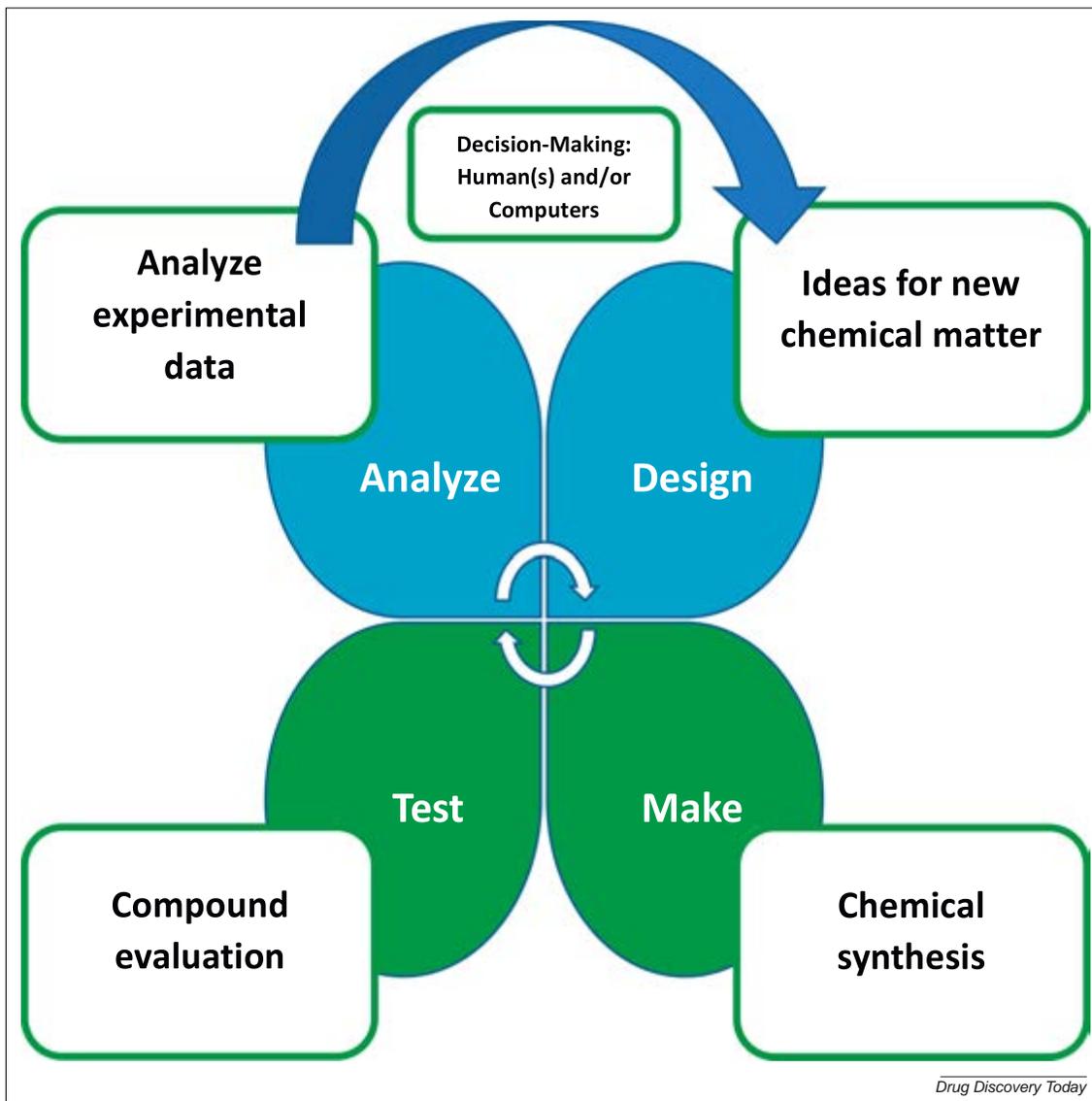
### Future trends: opportunities and challenges

Here, we synthesize some interesting future trends, highlighting opportunities and possible challenges with respect to the use of open innovation within a pharmaceutical context, specifically drug discovery and development. The presentation of these trends is subjective, and necessarily reflect our own interests and exposure. They are not intended to be comprehensive, but instead suggestive of possible areas for accelerating the adoption, application, and hoped for success of open innovation within drug discovery and development.

### Serendipity in science

The dichotomy of exploitation versus exploration remains a useful construct from the management literature to view most research activities [35,36]. Novel findings, almost tautologically, exist at the intersection of what we currently know (and have incorporated into a world view, or 'global understanding'), and what we are beginning to know. When the novel finding and/or understanding is of a surprising nature, or is somewhat unexpected, it is described as serendipitous in its nature, and, if popularized, a lore around said discovery becomes as important as the discovery itself (consider, for instance, Alexander Fleming's discovery of penicillin [37] or Barry Marshall's identification of gut bacteria as a principal source of stomach ulceration [38]).

Recently, several academics have begun to explore serendipity rigorously, providing frameworks and taxonomies to categorize and ground the concept epistemologically [39–41]. Such foundational scholarship is crucial as we look to understand the conditions required to best enhance for the likelihood of serendipitous



**FIGURE 2** Schematic of a 'Human-in-the-loop' Design–Make–Test–Analyze' (DMTA) cycle in drug discovery.

discovery outcomes. This was recognized during the recent allocation of a sizeable grant from the European Research Council (1.4 million; \$1.6 million) to gather evidence on the role of serendipity in science [42]. The recognition of these developments is important for practitioners of open innovation, because, as observed: 'A serendipitous discovery process may involve several unexpected observations and events, and may entail the formation of a network of interactions between individuals from various communities, backgrounds, and even times.' [39].

This is precisely the environment constructed, the space held, when an organization enters into any kind of mediated interaction with 'Other', as described by our network-oriented OIF. This remains conjecture at present (that such structures are sufficient to enrich for the likelihood of serendipitous discovery), but organizations would perhaps benefit from ensuring that networked interactions are viewed in as broad a perspective as possible, to ensure that such possibilities did not go unexplored (a recent example, with serendipitous discovery as a goal can be found in [43]).

In general, constructing open innovation activities ensuring a perspective of 'generative doubt' [44] might be of utility; this would represent a purposeful search for understanding, while recognizing the limitation of what is currently known. Again, such thinking makes what was historically difficult to define or 'handle', somewhat more tangible and amenable to intentional design.

#### *'Turtles all the way down' [45]*

In a previous contribution, open innovation tactics were described that borrow elements of games (gamification), to marshal, contain, and facilitate the production of novel insight to scientific problems [46]. The interested reader is referred to [47,48] for a more recent summary of the use of 'Scientific Discovery Games' (SDG) in the context of biomedical research.

Of particular note in recent developments regarding gamification is the use of gaming contexts to host meta-gaming challenges, that themselves encode actual science. This is spearheaded by the MMOS program ([www.mmos.ch](http://www.mmos.ch); a project of the GAPARS Horizon 2020 Grant), which looks to connect scientific research and video games as a seamless gaming experience. The creation of an image classification mini-game, and its embedding into the popular EVE Online platform saw participation of >300 000 gamers providing >30 million classifications of fluorescence microscopy images. In addition, the developers of the mini-game were able to combine annotations of the participants with deep learning methodologies to create readily improvable image classification models [49].

This represents an alternate open innovation tactic, that of packaging the act of science, and embedding it in a place of high foot traffic (taking advantage of the 'platform' presented) [50]. To our knowledge, at this time, no major pharmaceutical company is participating in such activities.

#### *Bridging the digital divide*

How readily can one translate novel crowd-sourced/open innovation insights into actual testable physically manifest science? The EteRNA platform allows participants to remotely perform experiments to verify computational predictions of how RNA molecules fold. Not only has the problem of RNA folding been abstracted to

an engaging game open to non-subject matter experts, but the translation of the creation of digital insight has also been made manifest in the real world [51,52]. Recently, this approach was brought to scale, leveraging crowdsourced RNA design to enable the discovery of reversible, efficient, and diverse self-contained molecular sensors [53].

#### *Artificial intelligence and 'crowd-in-the-loop'*

Recent advances in computer science and the dramatic increase in the amount of, often publicly, available data resulted in the increased use of machine-learning techniques and the introduction of what is summarized by the buzz-word of artificial intelligence (AI). As such, as in many high-technology fields, the pharmaceutical and biotechnology industry is exploring the potential of AI with respect to drug discovery. There are examples of AI techniques being used in all stages of drug discovery, from target identification [54] to lead identification and lead optimization [55,56], compound synthesis planning [57,58], and chemical synthesis [59,60] as well as appearing in differing aspects of clinical trials [61–63]. The use of AI could drastically revolutionize the drug discovery process and the sheer number of newly founded companies leveraging such technology is a barometer for the excitement and opportunity surrounding AI [64].

The recent and balanced perspective highlighting five 'grand challenges' facing the use of AI in drug design is a good start for a comprehensive overview [55]. What is highlighted is that, unlike in the fields of image and natural language processing, wherein AI does a good job in an unsupervised fashion, the complexity of drug design and paucity of relevant and appropriately annotated data suggest that there will be optimal use of AI approaches in direct 'collaboration' with human actors. As such, AI-enabled workflows with 'human-in-the-loop' elements could yield benefits, reducing cycle times (see, for example, the recent prospective application of deep learning and human selection/prioritization approaches to the development of DDR1 inhibitors [65]). One could imagine, at least conceptually, a significantly automated design-make-test-analyze cycle with a varying degree of human intervention (Fig. 2).

Speculatively, one wanders to what extent a 'crowd-in-the-loop' enabled process would even further boost performance and outcomes, wherein 'human' intervention is generalized to input from a crowd. It has been suggested that such 'human computation' strategies can be deployed on problems of arbitrary complexity, and might even be considered the natural endpoint of the innovation in crowdsourcing and open innovation we continue to witness [66].

#### **Concluding remarks**

Here, we have revisited a previously posited strategic framework through the lens of open innovation. This framework provides leadership within drug discovery and development an ability to orient and manage ongoing research activities along a networked continuum, leveraging a portfolio approach. Using this framework, along with an objective checklist, it is observed that a quarter of the top 20 biomedical companies (by revenue) have public-facing comprehensive open innovation strategies consistent with this framework, and the use and participation in open innovation practices continues to grow.

Indeed, since the original submission of this work, the use of open innovation has been applied in a variety of ways to contribute to the effort of understanding the ongoing Coronavirus 2019 (COVID-19) pandemic. Specifically, scientists have rallied to crowdsourcing platforms, such as folding@home (<https://foldingathome.org/>), to donate computer cycle time to the problem of simulating biomolecular interactions implicit in the virus pathology, and in the design of novel inhibitors (e.g., <https://covid.postera.ai/covid>) [67]. In addition, several COVID-19-related challenges have been added to the three platforms, Kaggle, InnoCentive, and DREAM, discussed in this work. In general, these challenges have focused on forecasting, gamification efforts to attempt to prevent or diminish the spread of the virus, and strategies to find the most useful solutions or services for people impacted by COVID-19. SageBionetworks is also hosting a COVID-19 electronic health records challenge, to help identify risk factors leading to positive test results. In general, these

challenges represent the platforms and their owners themselves wanting to contribute their participants time and effort, and do not reflect the use of drug discovery companies themselves using these platforms.

We conclude with a variety of adjacent future trends being highlighted. These are speculative and inherently subjective, reflecting our interests and exposure, but all of which might contribute to an increased adoption and application of open innovation in drug discovery and development.

The use of open innovation approaches continues to grow, and it is our expectation that the comprehensive practice of organizing activities to engage the 'Other' will remain a robust component of the strategy of any high-performing organization.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at [doi:https://doi.org/10.1016/j.drudis.2020.09.020](https://doi.org/10.1016/j.drudis.2020.09.020).

## References

- Dooley, K. (1996) Complex adaptive systems: a nominal definition. *Chaos Netw.* 8, 2–3
- FDA (2018) *FDA Novel Drug Approvals for 2018*. FDA
- Munos, B. (2019) 2018 new drugs approvals: an all-time record, and a watershed. *Forbes* January 14
- Howe, J. (2006) The rise of crowdsourcing. *Wired* June 14
- Sobel, D. (2007) *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. Bloomsbury
- Boudreau, K.J. and Lakhani, K.R. (2013) Using the crowd as an innovation partner. *Harv. Bus. Rev.* 91, 60–69
- Jeppesen, L.B. and Lakhani, K.R. (2010) Marginality and problem-solving effectiveness in broadcast search. *Organ. Sci.* 21, 1016–1033
- Chesbrough, H. (2003) *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business School Press
- Chesbrough, H. and Bogers, M. (2014) Explicating open innovation: clarifying an emerging paradigm for understanding innovation. In *New Frontiers in Open Innovation* (Chesbrough, Henry, Vanhaverbeke, Wim, West, Joel, eds), pp. 3–28, Oxford University Press
- Anon (2019) *Implementation of Federal Prize and Citizen Science Authority: Fiscal Years 2017–18*. US Government Office of Science and Technology Policy
- Bentzien, J. *et al.* (2015) Crowdsourcing in pharma: a strategic framework. *Drug Discov. Today* 20, 874–883
- Carroll, G.P. *et al.* (2017) Measuring the effectiveness and impact of an open innovation platform. *Drug Discov. Today* 22, 776–785
- Erkens, M. *et al.* (2014) Measuring open innovation: a toolkit for successful innovation teams. *Performance* 6, 12–23
- Liu, X. *et al.* (2019) The impact of external innovation on new drug approvals: a retrospective analysis. *Int. J. Pharm.* 563, 273–281
- Nilsson, N. and Minssen, T. (2018) Unlocking the full potential of open innovation in the life sciences through a classification system. *Drug Discov. Today* 23, 771–775
- Alvim-Gaston, M. *et al.* (2014) Open innovation drug discovery (OIDD): a potential path to novel therapeutic chemical space. *Curr. Top. Med. Chem.* 14, 294–303
- Martinez-Grau, M.A. and Alvim-Gaston, M. (2019) Powered by open innovation: opportunities and challenges in the pharma sector. *Pharm. Med.* 33, 193–198
- Henkel, J. (2006) Selective revealing in open innovation processes: the case of embedded Linux. *Res. Policy* 35, 953–969
- Wazny, K. (2017) Crowdsourcing's ten years in: a review. *J. Glob. Health* 7, 10502
- Wazny, K. (2018) Applications of crowdsourcing in health: an overview. *J. Glob. Health* 8 (June), . <http://dx.doi.org/10.7189/jogh.08.010502>
- Wagstaff, K. (2012) Machine learning that matters. *ArXiv* 2012 ArXiv12064656
- Bingham, R. (2019) *Crowdsourcing Drug Development*. Innocentive
- Davis, S. *et al.* (2019) Leveraging crowdsourcing to accelerate global health solutions. *Nat. Biotechnol.* 37, 848–850
- Saez-Rodriguez, J. *et al.* (2016) Crowdsourcing biomedical research: leveraging communities as innovation engines. *Nat. Rev. Genet.* 17, 470–486
- Beilmann, M. (2018) Optimizing drug discovery by Investigative Toxicology: Current and future trends. *ALTEX* 36, 289–313
- Miller, R.A. *et al.* (2018) Explicit interaction information from WikiPathways in RDF facilitates drug discovery in the Open PHACTS Discovery Platform. *F1000Research* 7, 75
- Ng, J.T. *et al.* (2016) Lessons from ten years of crystallization experiments at the SGC. *Acta Crystallogr. Sect. Struct. Biol.* 72, 224–235
- Müller, S. *et al.* (2018) Donated chemical probes for open science. *eLife* 7, e34311
- Kramer, C. *et al.* (2018) Learning medicinal chemistry absorption, distribution, metabolism, excretion, and toxicity (ADMET) rules from cross-company matched molecular pairs analysis (MMPA): miniperspective. *J. Med. Chem.* 61, 3277–3292
- Laverty, H. and Meulien, P. (2019) The Innovative Medicines Initiative – 10 years of public-private collaboration. *Front. Med.* 6, 275
- Riley, J.H. *et al.* (2018) U-BIOPRED: evaluation of the value of a public-private partnership to industry. *Drug Discov. Today* 23, 1622–1634
- Pastor, M. *et al.* (2018) Development of an infrastructure for the prediction of biological endpoints in industrial environments. Lessons learned at the eTOX project. *Front. Pharmacol.* 9, 1147
- Karawajczyk, A. *et al.* (2017) The European lead factory: a blueprint for public-private partnerships in early drug discovery. *Front. Med.* 3, 75
- Manner, C. *et al.* (2019) WIPO Re:Search: catalysing public-private partnerships to accelerate tropical disease drug discovery and development. *Trop. Med. Infect. Dis.* 4, 53
- March, J.G. (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2, 71–87
- Gupta, A.K. *et al.* (2006) The interplay between exploration and exploitation. *Acad. Manage. J.* 49, 693–706
- ACS (1999) *American Chemical Society International Historic Chemical Landmarks. Discovery and Development of Penicillin*. ACS
- Marshall, B. (2006) *Helicobacter* connections. *ChemMedChem* 1, 783–802
- Copeland, S. (2019) On serendipity in science: discovery at the intersection of chance and wisdom. *Synthese* 196, 2385–2406
- Yaqub, O. (2018) Serendipity: towards a taxonomy and a theory. *Res. Policy* 47, 169–179
- Cunha, M.P.E. (2005) *Serendipity: Why Some Organizations Are Luckier than Others*. SSRN
- Anon (2019) The serendipity test. *Nature* 554, 5
- Trouille, L. *et al.* (2019) Citizen science frontiers: efficiency, engagement, and serendipitous discovery with human-machine systems. *Proc. Natl. Acad. Sci. U. S. A.* 116, 1902–1909
- Pina e Cunha, M. *et al.* (2015) The dialectics of serendipity. *Eur. Manag. J.* 33, 9–18 [https://en.m.wikipedia.org/wiki/Turtles\\_all\\_the\\_way\\_down](https://en.m.wikipedia.org/wiki/Turtles_all_the_way_down).
- Bentzien, J. *et al.* (2013) Crowd computing: using competitive dynamics to develop and refine highly predictive models. *Drug Discov. Today* 18, 472–478
- Das, R. *et al.* (2019) Scientific discovery games for biomedical research. *Annu. Rev. Biomed. Data Sci.* 2, 253–279
- Graafland, M. and Schijven, M. (2018) How serious games will improve healthcare. In *Digital Health* (Rivas, H. and Wac, K., eds), pp. 139–157, Springer International Publishing

- 49 Sullivan, D.P. *et al.* (2018) Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nat. Biotechnol.* 36, 820–828
- [50] An early example of this, albeit one with more of a public relations orientation to it can be found in Boehringer Ingelheim's proposed use of Facebook to host a 'farmville' like game to introduce people to the complex context of pharmaceutical research. Despite a significant amount of press, the game remained unreleased (<https://web.archive.org/web/20120128125623/https://www.mmm-online.com/boehringers-pharma-ville-could-salve-industrys-bad-reputation/printarticle/215350/>) For those still wanting to experience Big Pharma., they may wish to turn to the recently released game of the same name on the PlayStation 4 console: (<https://www.playstation.com/en-us/games/big-pharma-ps4/>).
- 51 Bohannon, J. (2014) Online video game plugs players into remote-controlled biochemistry lab. *Science* 343 475–475
- 52 Lee, J. *et al.* (2014) RNA design rules from a massive open laboratory. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2122–2127
- 53 Andreasson, J.O.L. *et al.* (2019) Crowdsourced RNA design discovers diverse, reversible, efficient, self-contained molecular sensors. *bioRxiv*. <http://dx.doi.org/10.1101/2019.12.16.877183> Published online December 16, 2019
- 54 Wang, Q. *et al.* (2017) A novel framework for the identification of drug target proteins: combining stacked auto-encoders with a biased support vector machine. *PLoS One* 12, e0176486
- 55 Schneider, P. *et al.* (2020) Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* 19, 353–364. <http://dx.doi.org/10.1038/s41573-019-0050-3> Epub 2019 Dec 4
- 56 Yang, X. *et al.* (2019) Concepts of artificial intelligence for computer-assisted drug discovery. *Chem. Rev.* 119, 10520–10594
- 57 de Almeida, A.F. *et al.* (2019) Synthetic organic chemistry driven by artificial intelligence. *Nat. Rev. Chem.* 3, 589–604
- 58 Segler, M.H.S. *et al.* (2018) Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* 555, 604–610
- 59 Coley, C.W. *et al.* (2019) A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* 365, eaax1566
- 60 Steiner, S. *et al.* (2019) Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* 363, eaav2211
- 61 Gligorijevic, J. *et al.* (2019) Optimizing clinical trials recruitment via deep learning. *J. Am. Med. Inform. Assoc.* 26, 1195–1202
- 62 Haenssle, H.A. *et al.* (2018) Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* 29, 1836–1842
- 63 McKinney, S.M. *et al.* (2020) International evaluation of an AI system for breast cancer screening. *Nature* 577, 89–94
- 64 Boerner, Leigh Krietsch, ed. (2019) *Artificial Intelligence in Drug Discovery*, ACS In: <https://cen.acs.org/sections/discovery-reports/AI-drug-discovery.html>
- 65 Zhavoronkov, A. *et al.* (2019) Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 37, 1038–1040
- 66 Michelucci, P. and Dickinson, J.L. (2016) The power of crowds. *Science* 351, 32–33
- 67 Chodera, J. *et al.* (2020) Crowdsourcing drug discovery for pandemics. *Nat. Chem.* 12 (July), . <http://dx.doi.org/10.1038/s41557-020-0496-2> 581–581
- 68 Dorsch, H. *et al.* (2015) Grants4Targets: an open innovation initiative to foster drug discovery collaborations. *Nat. Rev. Drug Discov.* 14 (January), 74–76. <http://dx.doi.org/10.1038/nrd3078-c2>
- 69 Lessl, M. *et al.* (2011) Grants4Targets – an innovative approach to translate ideas from basic research into novel drugs. *Drug Discov. Today* 16, 288–292
- 70 Carter, A.J. *et al.* (2017) Establishing a reliable framework for harnessing the creative power of the scientific crowd. *PLoS Biol.* 15, e2001387
- 71 Carter, A.J. and Wood, C.R. (2016) Opening the innovation process in research at Boehringer Ingelheim to discover new medicines. *Nature* 533, 5