



# CACHE (Critical Assessment of Computational Hit-finding Experiments): A public–private partnership benchmarking initiative to enable the development of computational methods for hit-finding

Suzanne Ackloo<sup>1</sup>, Rima Al-awar<sup>2,3</sup>, Rommie E. Amaro<sup>4,5</sup>, Cheryl H. Arrowsmith<sup>1</sup>, Hatylas Azevedo<sup>6</sup>, Robert A. Batey<sup>7</sup>, Yoshua Bengio<sup>8</sup>, Ulrich A. K. Betz<sup>9</sup>, Cristian G. Bologa<sup>10</sup>, John D. Chodera<sup>11</sup>, Wendy D. Cornell<sup>12</sup>, Ian Dunham<sup>13,14</sup>, Gerhard F. Ecker<sup>15</sup>, Kristina Edfeldt<sup>16</sup>, Aled M. Edwards<sup>1,3</sup>, Michael K. Gilson<sup>5,17</sup>, Claudia R. Gordijo<sup>1</sup>, Gerhard Hessler<sup>18</sup>, Alexander Hillisch<sup>19</sup>, Anders Hogner<sup>20</sup>, John J. Irwin<sup>21</sup>, Johanna M. Jansen<sup>22</sup>, Daniel Kuhn<sup>23</sup>, Andrew R. Leach<sup>13,14</sup>, Alpha A. Lee<sup>24,25</sup>, Uta Lessel<sup>26</sup>, Maxwell R. Morgan<sup>1</sup>, John Moul<sup>27,28</sup>, Ingo Muegge<sup>29</sup>, Tudor I. Oprea<sup>10,30</sup>, Benjamin G. Perry<sup>31</sup>, Patrick Riley<sup>32</sup>, Sophie A. L. Rousseaux<sup>7</sup>, Kumar Singh Saikatendu<sup>33</sup>, Vijayaratnam Santhakumar<sup>1</sup>, Matthieu Schapira<sup>1,3</sup>, Cora Scholten<sup>34</sup>, Matthew H. Todd<sup>35</sup>, Masoud Vedadi<sup>1,3</sup>, Andrea Volkamer<sup>36</sup> and Timothy M. Willson<sup>37</sup>

**Abstract** | One aspirational goal of computational chemistry is to predict potent and drug-like binders for any protein, such that only those that bind are synthesized. In this Roadmap, we describe the launch of Critical Assessment of Computational Hit-finding Experiments (CACHE), a public benchmarking project to compare and improve small-molecule hit-finding algorithms through cycles of prediction and experimental testing. Participants will predict small-molecule binders for new and biologically relevant protein targets representing different prediction scenarios. Predicted compounds will be tested rigorously in an experimental hub, and all predicted binders as well as all experimental screening data, including the chemical structures of experimentally tested compounds, will be made publicly available and not subject to any intellectual property restrictions. The ability of a range of computational approaches to find novel binders will be evaluated, compared and openly published. CACHE will launch three new benchmarking exercises every year. The outcomes will be better prediction methods, new small-molecule binders for target proteins of importance for fundamental biology or drug discovery and a major technological step towards achieving the goal of Target 2035, a global initiative to identify pharmacological probes for all human proteins.

<sup>✉</sup>e-mail:  
aled.edwards@utoronto.ca;  
alexander.hillisch@bayer.com  
<https://doi.org/10.1038/s41570-022-00363-z>

Computational hit-finding is poised to make a major impact in early drug discovery<sup>1–4</sup>, enabled by leaps in computational power, increased accessibility to diverse chemical space, improved physics-based methods and the emerging potential of newer machine learning and artificial

intelligence approaches. However, despite the promise, no algorithm can currently select, design or rank potent drug-like small-molecule protein binders consistently.

Significant advances in the development of computational methods can be gained through blinded

**Hit-finding**

Identification of a small molecule that binds a target protein and that has high enough affinity and suitable physicochemical properties to qualify as a credible starting point for a drug discovery project.

**Chemical probes**

Chemical compounds used as tools to study the biological function of proteins.

benchmarking exercises, as evidenced by community progress in developing computational methods to predict protein structure from primary sequence. In 1993, when the Critical Assessment of protein Structure Prediction (CASP) exercise<sup>5</sup> was launched, humans were often better at predicting protein structures than computational methods. Now, machine learning algorithms can predict the structures of many (but not all) globular proteins as accurately as can be determined experimentally<sup>6,7</sup>, and progress is being made rapidly to predict the structures of protein complexes<sup>8,9</sup>.

In computational chemistry, organizing benchmarking exercises similar to CASP have occurred<sup>10–18</sup>, but none are currently operational. In addition, besides the TDT and DREAM benchmarking initiatives<sup>13,14,18</sup> that included a prospective arm to its prediction challenge, there has been no concerted effort to provide experimental testing of predictions, which is in large part because of the associated costs. There is no opportunity to fund the synthesis and quality control of predicted compounds and to test their binding rigorously in one laboratory under standardized conditions that facilitate head-to-head comparison of predictions. One confounding issue has been that commercial sensitivities complicate small-molecule-binding benchmarking. A large fraction of the experimental data suitable for benchmarking in silico binding predictions are generated within the pharma industry and kept confidential, rather than being released for general use. In addition, significant advances in computational chemistry technologies are taking place within companies, and massive private investment is flowing into new companies for the development of artificial intelligence methods. These companies are also likely reluctant to share their methods in any detail or see them put to the test publicly.

It is now possible to conceptualize a benchmarking exercise that can overcome some of these limitations. From a financial perspective, the creation of ultra-large libraries of chemicals that can be described in silico and procured on demand<sup>2,19</sup> significantly reduces the cost associated with accessing chemical matter to test predictions. The availability of massive amounts of computational resources facilitates data sharing and democratizes the ability to make predictions<sup>20</sup>.

From an organizational view, there is now community acceptance that public and private sectors can collaborate precompetitively in areas that were once considered commercially sensitive. The ‘open-access, open-source, open-data’ paradigm is accepted as an accelerator of biomedical science<sup>21,22</sup>. Critically, this paradigm has provided immense scientific value by normalizing the placement of chemical matter, including advanced molecules such as chemical probes, in the public domain without complex and rate-limiting intellectual property agreements<sup>21</sup>.

Based on this new landscape, we are creating a public-private partnership called Critical Assessment of Computational Hit-finding Experiments (CACHE) to benchmark computational approaches for the identification of a small molecule that binds a targeted protein with high enough affinity and suitable physicochemical properties to qualify as a credible starting point for a drug discovery project. Modelled after CASP, CACHE will organize hit-finding challenges against selected biologically relevant targets and participants will use various computational methods to predict hits. However, unlike CASP, which was able to piggyback experiments being done in the structural biology community, CACHE will have an experimental arm testing predictions prospectively. Each challenge will typically include two testing iterations to enable refinement and forward application of successful predictive models. Upon completion of a hit-finding challenge, all data generated by CACHE,

**Author addresses**

<sup>1</sup>Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada.

<sup>2</sup>Ontario Institute for Cancer Research, Toronto, Ontario, Canada.

<sup>3</sup>Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada.

<sup>4</sup>Department of Chemistry and Biochemistry, UC San Diego, La Jolla, CA, USA.

<sup>5</sup>Drug Design Data Resource, University of California, San Diego, La Jolla, CA, USA.

<sup>6</sup>Aché Laboratórios Farmacêuticos, Guarulhos, São Paulo, Brazil.

<sup>7</sup>Department of Chemistry, University of Toronto, Toronto, Ontario, Canada.

<sup>8</sup>Mila, University of Montreal, Montreal, Québec, Canada.

<sup>9</sup>Merck Healthcare KGaA, Darmstadt, Germany.

<sup>10</sup>Department of Internal Medicine, University of New Mexico School of Medicine, University of New Mexico Albuquerque, Albuquerque, NM, USA.

<sup>11</sup>Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

<sup>12</sup>Healthcare & Life Sciences Research, IBM TJ Watson Research Center, Yorktown Heights, NY, USA.

<sup>13</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK.

<sup>14</sup>Open Targets, Wellcome Genome Campus, Hinxton, UK.

<sup>15</sup>Department of Pharmaceutical Sciences, University of Vienna, Vienna, Austria.

<sup>16</sup>Structural Genomics Consortium, Department of Medicine, Karolinska University Hospital and Karolinska Institutet, Stockholm, Sweden.

<sup>17</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, UC San Diego, La Jolla, CA, USA.

<sup>18</sup>Sanofi-Aventis Deutschland GmbH, R&D, Integrated Drug Discovery, Frankfurt am Main, Germany.

<sup>19</sup>Bayer AG, Pharmaceuticals, Research and Development, Wuppertal, Germany.

<sup>20</sup>Medicinal Chemistry, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

<sup>21</sup>Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, USA.

<sup>22</sup>Novartis Institutes for BioMedical Research, Emeryville, CA, USA.

<sup>23</sup>Computational Chemistry & Biologics, Merck Healthcare KGaA, Darmstadt, Germany.

<sup>24</sup>PostEra Inc., San Francisco, CA, USA.

<sup>25</sup>Department of Physics, University of Cambridge, Cambridge, UK.

<sup>26</sup>Medicinal Chemistry, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany.

<sup>27</sup>Institute for Bioscience and Biotechnology Research, Rockville, MD, USA.

<sup>28</sup>Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD, USA.

<sup>29</sup>Alkermes, Inc., Waltham, MA, USA.

<sup>30</sup>University of New Mexico Comprehensive Cancer Center, Albuquerque, NM, USA.

<sup>31</sup>Drugs for Neglected Diseases initiative (DNDi), Geneva, Switzerland.

<sup>32</sup>Relay Therapeutics, Boston, MA, USA.

<sup>33</sup>Global Research Externalization, Takeda California, Inc., San Diego, CA, USA.

<sup>34</sup>Bayer AG, Pharmaceuticals, Open Innovation — Public Private Partnerships, Berlin, Germany.

<sup>35</sup>School of Pharmacy, University College London, London, UK.

<sup>36</sup>In Silico Toxicology and Structural Bioinformatics, Institute of Physiology, Charité Universitätsmedizin Berlin, Berlin, Germany.

<sup>37</sup>Structural Genomics Consortium, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

including all screening data and chemical structures, will be publicly available without intellectual property restrictions.

### The genesis of the CACHE concept

Prompted by recent developments and interest in computational methods, including deep learning, as well as the challenges in identifying the best performing methods, ~80 scientists from industry, academia and funding agencies met virtually in November 2020 to consider potential areas of drug discovery that might benefit from coordinated benchmarking. Of the many areas that were identified, the group prioritized hit-finding as particularly suitable and practical, and an excellent area to begin. To advance the idea, a set of ~30 representatives developed a draft concept for CACHE in four working groups, which focused on: target selection and prioritization; virtual library construction; measuring outcomes; and governance. These groups' ideas for the CACHE project are presented in this Roadmap.

### The CACHE concept

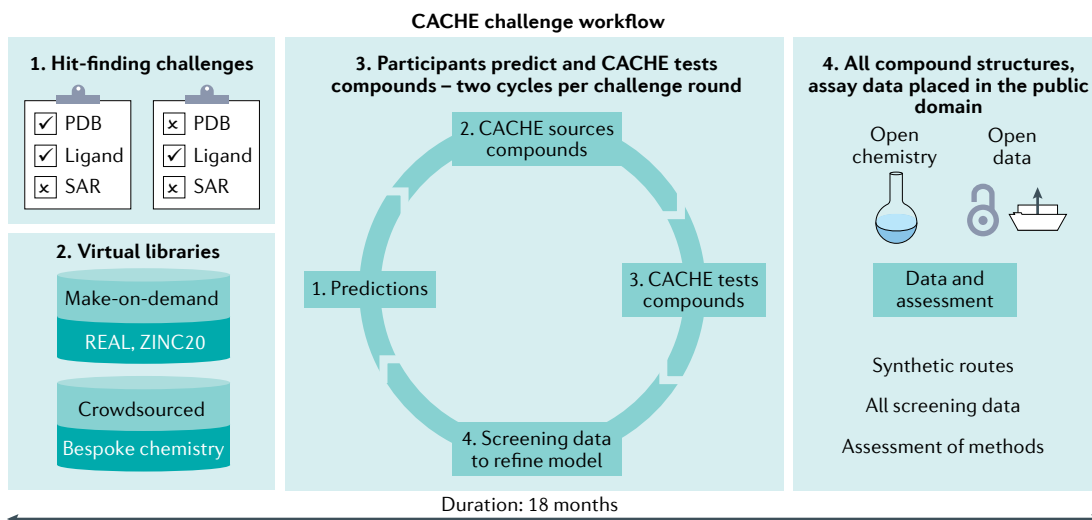
CACHE will present and organize a variety of hit-finding challenges to the community. As a part of this, and as described in detail below, CACHE will identify suitable protein targets, curate the virtual chemical libraries, define success parameters for generated predictions and solicit predictions for hit compounds. For evaluation, CACHE will purchase or otherwise procure the compounds that are predicted to bind, experimentally measure their binding to their intended target, calculate other key properties of the active compounds and share

the outcomes openly with the scientific community (FIG. 1). We envision that CACHE, like CASP, will organize multiple rounds of challenges, providing ongoing opportunities for computational scientists, molecular modellers, algorithm developers etc. to improve and test their methods.

### CACHE challenges and target selection

CACHE will organize hit-finding challenges that represent the common scenarios encountered in hit-finding (FIG. 2b). The CACHE target selection committee will select targets appropriate for each of these five scenarios. They will define the acceptance criteria for targets in each scenario and use bioinformatics tools to compile a longlist of targets that meet these criteria. Subsequently, they will create a mechanism or mechanisms for the community, including the funders of CACHE, to prioritize from this list of potential targets those that will be included in the benchmarking challenges.

Only targets having two orthogonal, cost-effective direct binding assays that can provide rapid, validated, high-quality experimental feedback will be considered. From this list, CACHE and its funders will use a prioritization scheme that maximizes both the structural diversity of the target proteins and the opportunity to discover new biological insights. The aim is for CACHE to benefit both the computational as well as the pharmaceutical communities. We anticipate that a funder (such as a disease-focused charity) might consider CACHE as an attractive funding opportunity through the mobilization of a wide global network of computational chemists to focus on their priority target(s) (FIG. 2a). We also



**Fig. 1 | CACHE challenge workflow.** 1. Hit-finding challenges: Critical Assessment of Computational Hit-finding Experiments (CACHE) presents a variety of hit-finding challenges to the community, including assessment criteria. 2. Virtual libraries: CACHE will establish and host two virtual libraries: a make-on-demand library (REAL, ZINC20) and a library comprising compounds synthetically accessible by chemists in academia or industry (bespoke chemistry). 3. Participants predict chemical matter and CACHE experimentally tests compounds: each participant will have the opportunity to make two cycles of predictions per round. CACHE will procure and assay the predicted compounds. At this stage, structures of compounds will be made available to all participants, but screening data will be provided only to the specific participant and competition management, in order to serve as a starting point for an additional cycle of predictions. 4. Compounds and data placed in the public domain: once the second cycle is complete, the data package, including all structures and screening data, as well as an assessment of each compound, will be made available to all, without restriction. PDB, Protein Data Bank; SAR, structure–activity relationship.

## a Target selection, prioritization

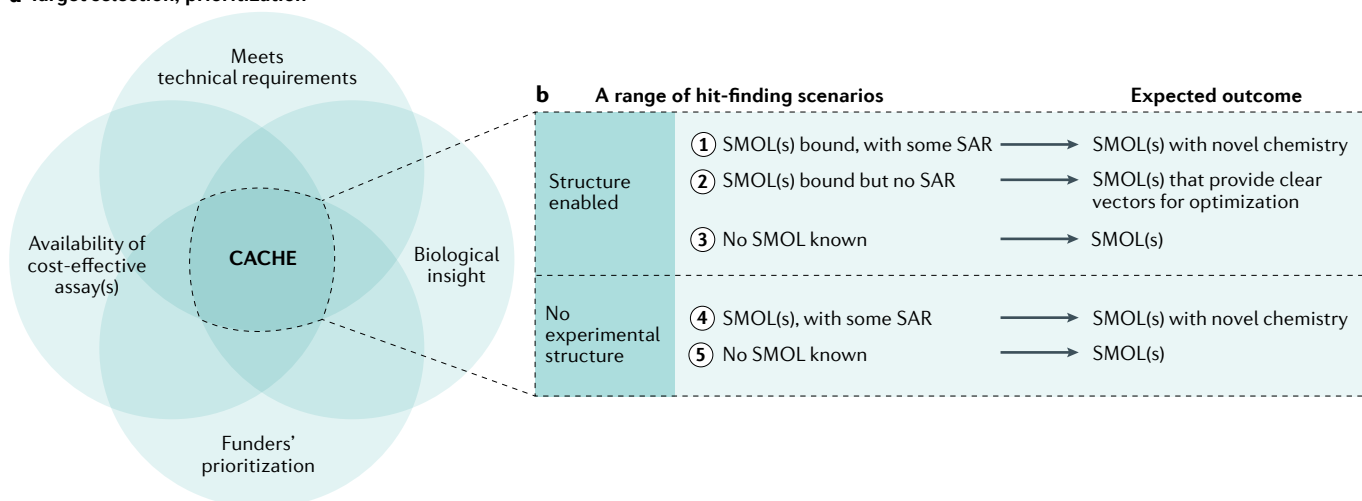


Fig. 2 | **Target selection consideration and classes of CACHE challenges.** **a** | Targets will be selected from a longlist of proteins that represent a range of scenarios of varying technical difficulty, are experimentally enabled (for example, there must be a robust binding assay) and, where possible, represent opportunities to make new biological or medical discoveries. Funders can prioritize targets within each challenge. **b** | The five potential hit-finding scenarios that address key technical questions in computational chemistry. CACHE, Critical Assessment of Computational Hit-finding Experiments; SAR, structure–activity relationship; SMOL, small molecule.

imagine that, in lieu of providing direct financial support, funders, foundations or companies might also offer in-kind support for CACHE, for example, by offering to evaluate all predictions for a given target or provide access to computational resources, assay reagents and/or laboratory equipment. Over a 5-year period, we aspire to provide CACHE with the resources to pursue 15 targets, representing each of the five hit-finding scenarios to enable it to fulfil its goals.

### Participation guidance and support

**Virtual compound libraries availability.** To enable rapid and cost-effective testing of predictions, CACHE will establish a well-defined and robust core make-on-demand virtual library comprising compounds that are readily accessible from commercial vendors, at reasonable cost. A combination of *Enamine REAL* (now providing 21 billion make-on-demand compounds) and *ZINC20* (REF.<sup>19</sup>) (containing over 750 million purchasable compounds) might comprise the core of this library.

CACHE will annotate compounds in the library with predicted physical properties, such as *cLogP*, polar surface area and the fraction of  $sp^3$  carbon atoms ( $F_{sp^3}$ ), among others, which will be assessed in the challenge's success criteria. Ideally, these annotated properties should enable participants to select individual subsets and/or apply relevant filtering as they see best fit for their challenge, while ensuring any such pre-filtering or subset restrictions can be accounted for in any subsequent evaluation and comparison of approaches. CACHE will also create subsets within the initial library, as this classification may be required to account for the needs of specific CACHE participants. For example, a 1% diversity set or a 10% diversity set might be preferred when examining computationally intensive approaches, and so on. The libraries will evolve, such that more

compounds will be added as they become commercially available or accessible, and additional library subsets will be created as a function of their performance.

To accommodate *de novo* design methods, which are not selecting compounds from commercial vendors but designing new molecules, CACHE will test custom-synthesized compounds if the compounds can be procured by participants within 3 months of the completion of the *in silico* selection step. In later challenges, CACHE may also incrementally explore mechanisms to provide participants access to a virtual library containing new chemistry, where synthetic chemists within academia or industry would be offered the opportunity to contribute to a virtual library that covers new chemical space. In this initiative, chemists would add compounds that they would be willing to synthesize on demand in a timely manner, using emerging synthetic chemistry protocols and their own resources.

At regular and defined intervals over the course of the CACHE benchmarking exercises, the CACHE virtual libraries committee will evaluate the impact of library choice, composition and nature (diversity, size) on both virtual screening capabilities and on general screening success, and recommend changes accordingly.

**Evaluating predictions experimentally.** At the core of the CACHE initiative will be an experimental hub that will provide rapid, high-quality testing of the predicted hits. Predicted compounds will be submitted to the experimental hub, which will procure the compounds and evaluate them using a binding assay selected to be most appropriate for the protein target. Each compound will be assayed at a single concentration in duplicate, and each positive will be retested in dose–response mode, as well as in an orthogonal biophysical assay, which is critical for the robustness of the experimental results. Feedback will

#### *cLogP*

Calculated partition coefficient of a chemical compound between water and 1-octanol.

#### Polar surface area

Surface sum over all polar atoms (namely, oxygen, nitrogen, phosphor and polar hydrogen) in a chemical compound.

#### Chemical space

Ensemble of all possible chemical compounds adhering to a given set of principles and boundary conditions, for drug-like small molecules estimated to be  $10^{60}$  compounds.

#### Experimental hub

Platform where predicted compounds are tested experimentally.

**Surface plasmon resonance**  
Label-free method that can be used to measure the binding of a small molecule to a protein immobilized on a chip.

**Dynamic light scattering**  
Method that can be used to measure the solubility or aggregation of molecules in solution.

**Pan-assay interference (PAINS) compounds**  
Chemical compounds often giving false positive results in high-throughput screens as they interact nonspecifically with numerous biological molecules.

**Differential scanning fluorimetry**  
Experimental method to measure protein unfolding by monitoring changes in fluorescence as a function of temperature.

**oralPhysChemScore (oPCS)**  
Combined score based on certain molecular properties, roughly estimating the suitability of a compound as the lead structure for an orally administered drug.

**Corrected molecular weight**  
Surrogate parameter for molecular volume, correcting the molecular weight of molecules containing halogen atoms.

be given first to the participant(s), and participants who made successful predictions will have the opportunity to improve on them by submitting a new set of predictions.

Each CACHE challenge round will take ~18 months, with two cycles of predictions per round in order to give participants the opportunity to incorporate learnings from the first round into their next designs. The timing and sequence of the proposed challenge round is shown in FIG. 3. Challenges will be staggered in order to avoid overwhelming the experimental hub. As part of each challenge, participants will be asked to make predictions from a small library constituting the combined list of predicted compounds contributed to the first cycle by all participants. Experimental testing of these compounds and then comparing with predictions will facilitate inter-algorithm benchmarking.

### CACHE benchmarking

Benchmarking computational hit-finding methods poses a challenge, because no single measure, or even combination of measures, can be used to unambiguously quantify the success of virtual screens, let alone determine which binder among many is the best. The affinity of compounds that are active in a primary screen, typically in a surface plasmon resonance assay, will be evaluated with an orthogonal biophysical method. Although binding affinity to the desired protein will be the main benchmarking criterion, selectivity against specific off-targets will be tested if called for in the challenge. The solubility and colloidal aggregation<sup>23</sup> of hit molecules will be determined experimentally by dynamic light scattering. Insoluble and aggregating compounds will be flagged because precipitation and aggregation are confounders in nearly all binding assays. Common pan-assay interference (PAINS) compounds<sup>24</sup>, predicted, for instance, by a strong indication of promiscuity with Badapple<sup>25</sup>, will also be flagged. Method-specific patterns of binding or inhibition that could be associated

with nonspecific interaction or aggregation will also be monitored. These include high Hill slopes of IC<sub>50</sub> determination plots, linear fitting of surface plasmon resonance data and unreasonable stabilization of proteins measured by differential scanning fluorimetry. Experimental hits will also be subjected to rigorous analytical quality control to confirm the purity of the samples. CACHE will seek to solve the crystal structure of validated hits in complex with their target when robust crystallization protocols are available.

Before each challenge, CACHE will publish the corresponding success criteria (activity, selectivity, aqueous solubility, lipophilicity, novelty etc.) and how these will be combined into an overall multi-objective score<sup>26,27</sup>, similar to the oralPhysChemScore (oPCS)<sup>28</sup>. Binding affinity, aqueous solubility and logD will be measured. Calculated properties include: corrected molecular weight; polar surface area<sup>29</sup>; number of rotatable bonds; Fsp<sup>3</sup> (REF.<sup>30</sup>); and novelty. This novelty parameter will be defined as the Tanimoto distance relative to most similar structures binding that target, as calculated from RDKit <http://www.rdkit.org>. These novelty thresholds were chosen based on previous work with circular fingerprints<sup>18,31</sup>. CACHE will provide the workflows and scripts that were used to calculate the different descriptors. In one possible scheme (TABLE 1), active compounds will not be ranked per se but, rather, will be classified into three buckets (green, yellow and red) by summing up the traffic light values for each property. The scoring scheme used to assess a compound's physical and molecular properties will be similar across the challenges, but the values for potency and selectivity may change, depending on the challenge. For example, compounds with weaker affinity might be acceptable for targets that are more difficult to identify hits against and have no reported precedent, but higher affinities might be the aim if the challenge is to identify novel chemotypes for precedented targets. As stated above, to facilitate comparison among the methods, all predictions from all participants for a given target will be combined into a single small virtual library, and all participants will also be asked to rank these compounds.

Top-scoring molecules (TABLE 1) will be further analysed by a panel of experienced medicinal chemists in order to provide additional annotation to the molecules, including opinion on the suitability of the hits to serve as a starting point for potential drug discovery programmes. This includes human experience on reactivity, synthesizability, chemical stability, potential toxicity, off-target activity etc. Their reflections will not influence the score but, rather, will help contextualize the output and provide insight for refinement of the scoring process for future challenge iterations.

### CACHE output sharing

CACHE will generate three main outputs for the community: screening data, chemical structures and algorithm performance (BOX 1). CACHE's mandate is to ensure that the screening data and the chemical structures are available to the community without intellectual property or other restrictions on use, and in a digitally readable format according to FAIR principles<sup>32</sup>.

#### Box 1 | CACHE output

##### List of methods and strategies

- Anonymized list of participants, along with a description of their approach.

##### Predicted structures from each participant

- Experimentally determined and calculated properties for all predicted compounds (TABLE 1), for each of the two cycles.

##### Performance of algorithms on a common set of compounds

- Create a virtual library that comprises predictions made by all participants in cycle 1, and each participant will rank the compounds in that library using their algorithm.

##### Set of top structures

- Top-ranked structures, including structure–activity relationship if available.

##### Crystal structures

- Coordinates of all complexes of targets and predicted binders.

##### Synthetic routes for top-ranked set of structures

- Aim to provide synthetic routes with a summary of experimental methods and primary data (yields, purities etc.).

##### Assay data (screening)

- Primary screening data for all predictions and orthogonal confirmation data for active molecules.

##### Quality control data for compounds NMR, high-performance

- Liquid chromatography, mass spectrometry, solubility.

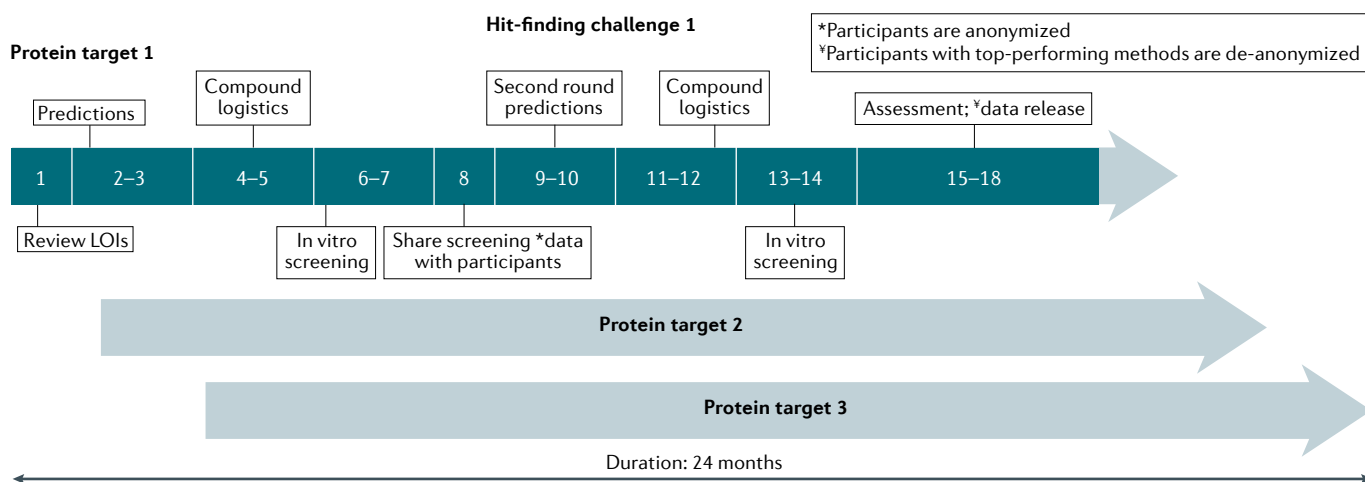


Fig. 3 | **The timelines of challenge activities.** After reviewing the letters of intent (LOIs), each complete challenge round will take ~18 months, with the various stages outlined.

These data will also include the composition of the virtual libraries screened, all predicted small molecules (including negative data), all experimental screening results and all screening methods.

CACHE will mandate that participants disclose their computational approaches in sufficient detail to enable an expert in the area to understand the methodology and algorithms. These methodology descriptions will be double-blind peer reviewed by other participants to ensure they contain sufficient information according to the standards of the field. In the interest of encouraging participation from all sectors, participants will not be required to provide access to their code and can remain anonymous. However, CACHE will encourage participants to share their software code and, as stated below, intends to provide a range of financial incentives for those participants who release their code, algorithms and workflows under permissive open-source license terms and, ideally, who also submit their fully automated workflows. In addition, participants must agree that the identity of those who submit top-performing methods (as determined by prespecified criteria agreed to by CACHE and the participants) will automatically be de-anonymized when the screening data and compound structures are publicly released. Participants who agree to share workflows, code and methodology must do so in a FAIR manner<sup>32</sup>.

Participants will be encouraged to seek peer-reviewed open-access publication of the results of their submissions and detailed analyses of their performance, and to work together to share learnings and identify differentiators of performance. CACHE will organize a workshop following each challenge and coordinate the open-access publication of overview papers for each challenge, perhaps with dedicated special issues of relevant journals to provide a wider forum for participants.

#### CACHE organization and management

CACHE will be structured as an independent, not-for-profit entity or fiscally governed by a not-for-profit organization with aligned goals, such as the [Structural Genomics Consortium \(SGC\)](#) or the [Open Group](#).

CACHE or its parent organization will receive funding as described below and subcontract other organizations (academic, government or industry) to carry out CACHE activities, all under terms that mandate open data sharing. CACHE will create a secretariat to handle administration, fundraising, project management and logistics.

CACHE will be funded in part by members, who will have the opportunity to influence the strategic directions of CACHE through appointments to a governing board (FIG. 4). The governing board will be responsible for making operational decisions, including target selection, participation rules and use of funds. An external scientific advisory board will be appointed by the governing board to provide outside advice on scientific questions such as the strategy for target selection and the metrics for success.

CACHE plans to launch challenges for each of the five hit-finding scenarios shown in FIG. 2, each challenge occurring at least once over 2 years (FIG. 3). There will be periodic public open calls for participation. For the first rounds, letters of intent will be solicited to better understand the needs and goals of potential participants. All potential participants would be asked to submit brief applications detailing their qualifications to participate and general intended approach. For inclusivity, the initiative should strive to accept every reasonable application, paying attention to use resources efficiently.

For each challenge, CACHE will contribute a challenge lead, who will be responsible for the coordination of experiments and logistics. The challenge lead will ensure that best practices are used in challenge design, execution and assessment, and codified in iteratively revised documents. For instance, these documents could be similar to the living reviews found in the [Living Journal of Computational Molecular Science](#) or made as contributions to the [NCATS Assay Guidance Manual](#). Challenge leads, in consultation with the governing board, will determine the details of specific challenges and what compound properties — experimental or computed — beyond affinity for the




#### Tanimoto distance

Statistic used for gauging the similarity and diversity of sample compound sets.

#### Circular fingerprints

Fingerprints representing molecular structures by means of circular atom neighbourhoods.

Table 1 | Example Critical Assessment of Computational Hit-finding Experiments (CACHE) traffic light scoring scheme for one arbitrary target protein

Traffic light score	Score	Binding affinity <sup>a</sup> (μM)	Solubility in water <sup>a</sup> (mg l <sup>-1</sup> )	logD (pH 7.5) <sup>a</sup>	MWcorr	PSA (Å <sup>2</sup> )	Number of rotatable bonds	Fsp <sup>3</sup>	Novelty <sup>b</sup>
	2	>10	<10	>4	>500	>140	≥11	<0.2	>0.6
	1	1–10	10–50	3–4	400–500	120–140	8–10	0.2–0.3	0.4–0.6
	0	<1	≥50	<3	≤400	≤120	≤7	>0.3	<0.4

Fsp<sup>3</sup>, fraction of sp<sup>3</sup> hybridized carbon atoms, calculated based on Murcko scaffolds. <sup>a</sup>Measured experimentally. <sup>b</sup>Tanimoto distance relative to most similar structures binding that target, as calculated from RDKit. PSA, polar surface area.

target will be incorporated into the overall performance scores.

Challenge leads will also be responsible for determining and executing or delegating the execution of appropriate baseline methods to be run centrally to avoid duplication for participants running many similar baselines. These methods would likely include random local search, simple similarity matching or vanilla docking methods, where applicable. Challenge leads will have the support of the scientific advisory board in making all of these decisions.

### CACHE funding strategy

CACHE intends that its activities, including governance, management, logistics and data sharing, will be supported by a pool of government, industry and charitable funders. Ideally, CACHE funding would also be used to provide subsidies for participants from resource-poor

environments, providing an overall more inclusive approach.

The funding of the challenges themselves will be shared among interested funders and participants. Funders, such as a disease foundation, could support challenges of particular interest to them. As CACHE matures, participants will be expected to pay a participation fee reflective of a portion of per-compound costs (including synthesis/procurement and assays). To facilitate this, CACHE will develop a transparent cost structure for each challenge. In the interest of encouraging transparency, CACHE aspires to be able to subsidize the cost of participation for participants who agree to share their methods, code or methodologies.

By centralizing the experimentation, CACHE will not only provide standardized data but will also provide logistical and cost savings over carrying out the activities in individual labs. Within CACHE, we estimate that the costs of rigorous experimental testing for 100 compounds is approximately US \$25,000; this includes purchasing of the compounds, quality control, protein purification, equipment time, primary biophysical assays and hit confirmation using orthogonal assays. CACHE will procure the compounds on behalf of all participants to facilitate logistics as well as to provide the opportunity to negotiate bulk pricing.

In the first two competitions, CACHE aims to secure sufficient seed funding to purchase and evaluate ~100 compounds for every qualified participant, but, in subsequent rounds, these costs will be transferred to participants. If participants wish to test more than 100 compounds, or if the number of participants exceeds the initial available funding, participants may also be required to fund some portion of per-compound costs.

CACHE will also be well positioned to collaborate with other successful community initiatives in order to increase the impact of CACHE. For example, if CACHE includes a viral target among the challenges, then the CACHE predictions might input into community antiviral development initiatives, such as the COVID Moonshot initiative<sup>20</sup>. Predicted compounds that pose synthetic challenges can be turned into additional community challenges, such as [Merck's Compound Synthesis Challenge](#), to design and predict the most efficient synthetic pathway for a given small molecule. Confirmed hits could also be used as starting points to develop new [chemical probes](#).

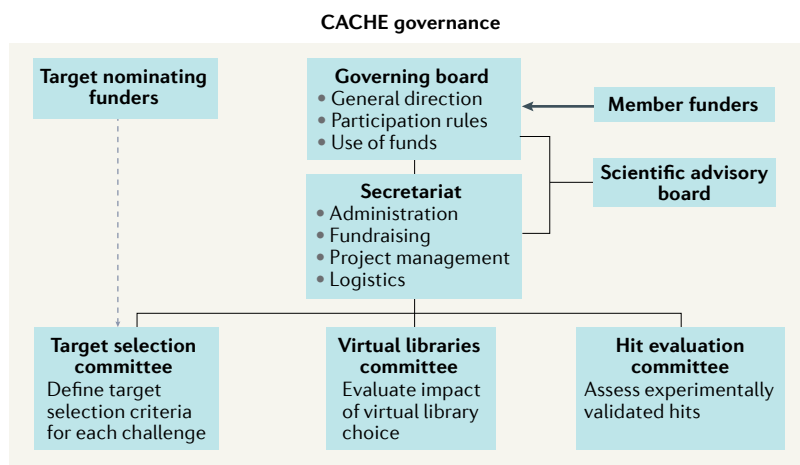


Fig. 4 | **CACHE governance.** Critical Assessment of Computational Hit-finding Experiments (CACHE) will be structured as an independent, not-for-profit entity. The CACHE governance will include: a governing board constituted by funders (members) and two independent members selected with input from the scientific community: an external scientific advisory board and a secretariat who will oversee day-to-day operations. The governing board will create three scientific committees: the target selection committee will select protein targets (with the final decision impacted by the governing board); the virtual libraries committee will define the virtual chemistry libraries to be screened; and the hit evaluation committee will create the metrics of success and assess performance against the metrics. Funders who do not wish to play an active role in governance can nominate targets for consideration by the target selection committee.

**CACHE success criteria**

CACHE will be a long-term project that will be assessed against success metrics of organizational capabilities and community engagement in the short term (1–3 years) and scientific accomplishments in the longer term (year 3 and beyond). Organizational success will be achieved by running the entire workflow of target selection for several rounds. For example, we expect six rounds to run over ~2 years, where a round includes hit prediction, chemical synthesis, biochemical/biophysical testing of the compounds and analysis/dissemination of the results (FIG. 3). Community engagement success will be defined as generating a constant flow of targets, hit proposals and experimental results from an increasing number of community members over time. Scientific success can likely be analysed only after 12 rounds (year 4), after which all five types of challenges are performed at least two to three times with different targets. Scientific success metrics will include providing unbiased comparisons of which computational methods deliver suitable hits (chemotypes) as starting points for drug discovery and the number and quality of novel chemical matter for biologically interesting new targets.

With respect to quantitative metrics, we aspire for CACHE to have deposited experimental screening data for 12 proteins and 30,000 drug-like molecules selected by over 100 participants in the public domain after

4 years. Over this period, we also expect that computational methods will predict unprecedented hits for 25% of the nominated novel targets. We also expect CACHE to provide clearer guidance as to which computational approaches are most promising for identifying novel small molecules active substances and, thus, significantly influence computational hit-finding-method development on a global scale.

**Summary and next steps**

A group of ~50 scientists from the public and private sectors intend to launch a benchmarking initiative to accelerate the development of computational methods to predict small molecules that bind to proteins. The initiative will comprise experimental and data hub(s), which will support a community of participants in their predictions. All data, including chemical structures, will be made available without restriction on use. The initiative intends to attract funding from industry, governments and foundations to support the infrastructure and challenge-specific funding in order to give disease-focused funders the opportunity to enable a community-wide effort to target proteins of interest to them. The intention is to launch the first **CACHE challenge** in early 2022.

Published online 15 February 2022

- Gorgulla, C. et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663–668 (2020).
- Lyu, J. et al. Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).
- Walters, W. P. & Wang, R. New trends in virtual screening. *J. Chem. Inf. Model.* **60**, 4109–4111 (2020).
- Grebner, C. et al. Virtual screening in the cloud: how big is big enough? *J. Chem. Inf. Model.* **60**, 4274–4282 (2020).
- Moult, J., Pedersen, J. T., Judson, R. & Fidelis, K. A large-scale experiment to assess protein structure prediction methods. *Proteins* **25**, ii–iv (1995).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
- Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at *BioRxiv* <https://doi.org/10.1101/2021.10.04.463034> (2021).
- Humphreys, I. R. et al. Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805 (2021).
- Gaieb, Z. et al. D3R Grand Challenge 3: blind prediction of protein–ligand poses and affinity rankings. *J. Comput. Aided Mol. Des.* **33**, 1–18 (2019).
- Parks, C. D. et al. D3R grand challenge 4: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* **34**, 99–119 (2020).
- Gaieb, Z. et al. D3R Grand Challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* **32**, 1–20 (2018).
- Jansen, J. M., Cornell, W., Tseng, Y. J. & Amaro, R. E. Teach–Discover–Treat (TDT): collaborative computational drug discovery for neglected diseases. *J. Mol. Graph. Model.* **38**, 360–362 (2021).
- Jansen, J. M., Amaro, R. E., Cornell, W., Tseng, Y. J. & Walters, W. P. Computational chemistry and drug discovery: a call to action. *Future Med. Chem.* **4**, 1893–1896 (2012).
- Gathiaka, S. et al. D3R grand challenge 2015: evaluation of protein–ligand pose and affinity predictions. *J. Comput. Aided Mol. Des.* **30**, 651–668 (2016).
- Yin, J. et al. Overview of the SAMPL5 host–guest challenge: Are we doing better? *J. Comput. Aided Mol. Des.* **31**, 1–19 (2017).
- Bannan, C. C. et al. Blind prediction of cyclohexane–water distribution coefficients from the SAMPL5 challenge. *J. Comput. Aided Mol. Des.* **30**, 927–944 (2016).
- Xiong, Z. et al. Crowdsourced identification of multi-target kinase inhibitors for RET- and TAU-based disease: the Multi-Targeting Drug DREAM Challenge. *PLoS Comput. Biol.* **17**, e1009302 (2021).
- Irwin, J. J. et al. ZINC20 — a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **60**, 6065–6073 (2020).
- von Delft, F. et al. A white-knuckle ride of open COVID drug discovery. *Nature* **594**, 330–332 (2021).
- Edwards, A. M., Bountra, C., Kerr, D. J. & Willson, T. M. Open access chemical and clinical probes to support drug discovery. *Nat. Chem. Biol.* **5**, 436–440 (2009).
- Müller, S. et al. Target 2035—update on the quest for a probe for every protein. *RSC Med. Chem.* **13**, 13–21 (2022).
- McGovern, S. L., Helfand, B. T., Feng, B. & Shoichet, B. K. A specific mechanism of nonspecific inhibition. *J. Med. Chem.* **46**, 4265–4272 (2003).
- Baell, J. B. & Nissink, J. W. M. Seven year itch: pan-assay interference compounds (PAINS) in 2017 — utility and limitations. *ACS Chem. Biol.* **13**, 36–44 (2018).
- Yang, J. J. et al. Badapple: promiscuity patterns from noisy evidence. *J. Cheminformatics* **8**, 29 (2016).
- Wager, T. T., Hou, X., Verhoest, P. R. & Villalobos, A. Central nervous system multiparameter optimization desirability: application in drug discovery. *ACS Chem. Neurosci.* **7**, 767–775 (2016).
- Cummins, D. J. & Bell, M. A. Integrating everything: the molecule selection toolkit, a system for compound prioritization in drug discovery. *J. Med. Chem.* **59**, 6999–7010 (2016).
- Lobell, M. et al. In silico ADMET traffic lights as a tool for the prioritization of HTS hits. *ChemMedChem* **1**, 1229–1236 (2006).
- Ertl, P., Rohde, B. & Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **43**, 3714–3717 (2000).
- Lovering, F., Bikker, J. & Humblet, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52**, 6752–6756 (2009).
- Muchmore, S. W. et al. Application of belief theory to similarity data fusion for use in analog searching and lead hopping. *J. Chem. Inf. Model.* **48**, 941–948 (2008).
- Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).

**Acknowledgements**

We are grateful to Ali Hazrat (Karolinska Institutet, Stockholm, Sweden) for developing the CACHE website: [www.cache-challenge.org](http://www.cache-challenge.org). Advice on chemoinformatic aspects from Hans Briem (Bayer AG, Berlin, Germany) is gratefully acknowledged. The Structural Genomics Consortium is a registered charity (no. 1097737) that receives funds from Bayer AG, Boehringer Ingelheim, Bristol Myers Squibb, Genentech, Genome Canada through Ontario Genomics Institute (OGI-196), Janssen, Merck KGaA (aka EMD in Canada and USA), Pfizer, Takeda and the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement no. 875510. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and Ontario Institute for Cancer Research, Royal Institution for the Advancement of Learning McGill University, Kungliga Tekniska Hogskolan and Diamond Light Source Limited. This communication reflects the views of the authors and the JU is not liable for any use that may be made of the information contained herein. M.K.G. acknowledges funding from the National Institute of General Medical Sciences (GM061300). J.J.I. acknowledges funding from the National Institute of General Medical Sciences (GM133836). J.D.C. acknowledges funding from the National Institute of General Medical Sciences (R01GM124270) and the National Cancer Institute (P30CA008748). These findings are solely of the authors and do not necessarily represent the views of the NIH. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence



Berkeley National Laboratory, operated under contract no. DE-AC02-05CH11231. The Lee laboratory at the University of Cambridge receives funding from multiple sources, including Pfizer, AstraZeneca, the Engineering and Physical Sciences Research Council and the Winton Programme for the Physics of Sustainability. T.I.O. and C.G.B. acknowledge funding from the National Institutes of Health Common Fund programme, Illuminating the Druggable Genome (CA224370 and TR002278). R.A.B. acknowledges funding from the Natural Science and Engineering Research Council (NSERC) of Canada. B.G.P. acknowledges the following donors for contributing to DNDi's overall mission: UK Aid, UK; Médecins Sans Frontières, International; and the Swiss Agency for Development and Cooperation (SDC), Switzerland.

#### Competing interests

M.K.G. has an equity interest in and is a co-founder and scientific adviser of VeraChem LLC. J.J.I. is a co-founder of Blue Dolphin LLC, which undertakes fee-for-service ligand discovery. A. Hillisch is on the board of directors of the Structural Genomics Consortium (SGC) and the scientific advisory board of Cresset. A.A.L. is the chief scientific officer and a

shareholder of PostEra Inc. T.I.O. has received honoraria from or consulted for Abbott, AstraZeneca, Chiron, Genentech, Infinity Pharmaceuticals, Merz Pharmaceuticals, Merck Darmstadt, Mitsubishi Tanabe, Novartis, Ono Pharmaceuticals, Pfizer, Roche, Sanofi and Wyeth, and is on the scientific advisory board of ChemDiv and InSilico Medicine. J.D.C. is a current member of the scientific advisory boards for OpenEye Scientific Software, Redesign Science, Interline Therapeutics and Ventus Therapeutics, and holds equity interests in Redesign Science and Interline Therapeutics. B.G.P. is on the board of directors of Evolia Therapeutics S.A. and the scientific advisory board of Spirochem A.G. All remaining authors declare no competing interests.

#### Peer review information

*Nature Reviews Chemistry* thanks M. Kostic, C. W. Murray, B. Shoichet and the other, anonymous, reviewer for their contribution to the peer review of this work.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### RELATED LINKS

CACHE challenge: <https://cache-challenge.org/>

Critical Assessment of protein Structure Prediction (CASP): <https://predictioncenter.org/>

Develop new chemical probes: <https://www.chemicalprobes.org/>

Enamine REAL: <https://enamine.net/compound-collections/real-compounds/real-space-navigator>

Living Journal of Computational Molecular Science: <https://livecomsjournal.org/index.php/livecoms/index>

Merck's Compound Synthesis Challenge: <http://compoundchallenge.merckgroup.com/>

NCATS Assay Guidance Manual: <https://ncats.nih.gov/expertise/preclinical/agm>

Open Group: <https://www.opengroup.org/>

Protein Data Bank (PDB): <https://www.rcsb.org/>

RDKit: <https://rdkit.org/>

Structural Genomics Consortium (SGC): <http://thesgc.org/>

TDT (Drug Design Data Resource, D3R): <https://drugdesigndata.org/>

© Springer Nature Limited 2022