

Cite this: *RSC Med. Chem.*, 2025, 16, 457

Toward target 2035: EUBOPEN - a public-private partnership to enable & unlock biology in the open

Claudia Tredup,^{id ab} Suzanne Ackloo,^{id c} Hartmut Beck,^d Peter J. Brown,^e Alex N. Bullock,^{id f} Alessio Ciulli,^{id g} Ivan Dikic,^{id hi} Kristina Edfeldt,^{id j} Aled M. Edwards,^c Jonathan M. Elkins,^f Henner F. Farin,^{id k} Edward A. Fon,^{id l} Matthias Gstaiger,^m Judith Günther,ⁿ Anna-Lena Gustavsson,^o Sandra Häberle,^{ab} Laura Isigkeit,^{id a} Kilian V. M. Huber,^{id f} Andras Kotschy,^{id p} Oliver Krämer,^q Andrew R. Leach,^r Brian D. Marsden,^f Hisanori Matsui,^s Daniel Merk,^{id t} Florian Montel,^u Monique P. C. Mulder,^{id v} Susanne Müller,^{id ab} Dafydd R. Owen,^{id w} Ewgenij Proschak,^{az} Sandra Röhm,^{ab} Alexandra Stolz,^{id hi} Michael Sundström,^j Frank von Delft,^{id fx} Timothy M. Willson,^e Cheryl H. Arrowsmith^{*cy} and Stefan Knapp^{id *ab}

Target 2035 is a global initiative that seeks to identify a pharmacological modulator of most human proteins by the year 2035. As part of an ongoing series of annual updates of this initiative, we summarise here the efforts of the EUBOPEN project whose objectives and results are making a strong contribution to the goals of Target 2035. EUBOPEN is a public-private partnership with four pillars of activity: (1) chemogenomic library collections, (2) chemical probe discovery and technology development for hit-to-lead chemistry, (3) profiling of bioactive compounds in patient-derived disease assays, and (4) collection, storage and dissemination of project-wide data and reagents. The substantial outputs of this programme include a chemogenomic compound library covering one third of the druggable proteome, as well as 100 chemical probes, both profiled in patient derived assays, as well as hundreds of data sets deposited in existing public data repositories and a project-specific data resource for exploring EUBOPEN outputs.

Received 18th September 2024,
Accepted 5th November 2024

DOI: 10.1039/d4md00735b

rsc.li/medchem

- ^a Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany. E-mail: knapp@pharmchem.uni-frankfurt.de
- ^b Structural Genomics Consortium, BMLS, Goethe University Frankfurt, Frankfurt 60438, Germany
- ^c Structural Genomics Consortium, University of Toronto - St George Campus, 101 College Street, MaRS Center South Tower 7th Floor, Toronto, Canada
- ^d Drug Discovery Sciences, Research & Development, Pharmaceuticals, Bayer AG, Wuppertal, Nordrhein-Westfalen, Germany
- ^e Structural Genomics Consortium, University of North Carolina at Chapel Hill, Campus Box 7356, 120 Mason Farm Road, GMB 1070, Chapel Hill, North Carolina, USA
- ^f Centre for Medicines Discovery, University of Oxford, NDM Research Building, Roosevelt Drive, Oxford, Oxfordshire, UK
- ^g Centre for Targeted Protein Degradation, University of Dundee, School of Life Sciences, 1 James Lindsay Place, DD1 5JJ, Dundee, UK
- ^h Institute of Biochemistry II, Goethe University Frankfurt, Medical Faculty, Frankfurt am Main, Germany
- ⁱ Buchmann Institute for Molecular Lifesciences, Goethe University Frankfurt, Frankfurt am Main, Germany
- ^j Structural Genomics Consortium, Department of Medicine, Karolinska University Hospital and Karolinska Institutet, Stockholm, Sweden
- ^k Georg-Speyer-Haus, Institute for Tumor Biology and Experimental Therapy, Frankfurt am Main, Hessen, Germany
- ^l Department of Neurology and Neurosurgery, Montreal Neurological Institute-

- Hospital (The Neuro), McGill University, Montreal, Canada
- ^m Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, Zurich, ZH, Switzerland
- ⁿ Drug Discovery Sciences, Bayer AG, Berlin, Germany
- ^o Chemical Biology Consortium Sweden, Department of Medical Biochemistry & Biophysics, Karolinska Institute, Stockholm, Sweden
- ^p Servier Research Institute of Medicinal Chemistry, Budapest, Hungary
- ^q Discovery Research Coordination, Boehringer Ingelheim International GmbH, Binger Straße 173, 55216 Ingelheim am Rhein, Germany
- ^r European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, UK
- ^s Neuroscience Drug Discovery Unit, Takeda Pharmaceutical Company Limited, Fujisawa, Kanagawa, Japan
- ^t Ludwig-Maximilians-Universität München, München, Germany
- ^u Discovery Research Coordination, Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Straße 65, 88397 Biberach an der Riss, Germany
- ^v Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands
- ^w Pfizer Medicine Design, Cambridge, MA 02139, USA
- ^x Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK
- ^y Princess Margaret Cancer Centre, Toronto, Ontario, M5G 1L7, Canada
- ^z Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Theodor-Stern-Kai 7, 60596, Frankfurt am Main, Germany

Introduction

Current small-molecule drug development focuses on a few well-established target families that define the explored druggable proteome. Although the number of target families has increased significantly over the past few decades, many proteins within established and yet to be discovered target families remain unexplored. Sequencing effort has identified many disease-associated mutations that provide a compelling rationale for targeting these proteins for the development of new therapeutics. However, the druggability of most of these proteins has not yet been demonstrated by the development of selective and potent small molecules. Such modulators, also called chemical probes, are powerful tools for the validation of new targeting strategies.¹ Yet, the combined effort of the private sector and academic drug development community has only developed a few hundred high quality chemical probes to date. The current long development timelines preclude the development of chemical probes for every druggable protein in the near future. However, potent inhibitors or chemical activators with narrow but not exclusive target selectivity, so-called chemogenomic (CG) compounds, are a feasible and extremely useful interim solution until more selective molecules are discovered. CG tools are powerful reagents when several of these small molecules with diverse off-target activity profiles are combined into collections of compounds that allow target deconvolution based on selectivity patterns.² The combined collection of high-quality chemical probes and well annotated CG compound sets covers now a much larger target space. The selection and characterization of CG sets for the druggable protein families is one of the major goals of the EUBOPEN consortium.

Recently, new modalities such as molecular glues, PROTACs (PROteolysis TArgeting Chimeras) and other proximity inducing small molecules such as molecular glues have also emerged as new chemical modulators with unique properties.^{3–5} This has led to a large expansion of the druggable proteome and new opportunities for drug development. However, only a few of the large family of E3 ligases have been successfully exploited so far, limiting our ability to develop the next generation of chemical degraders. The development of new E3 ligase ligands and the identification of linker attachment points, known as E3 handles, for degrader design has also been a major goal of EUBOPEN, and the first new E3 ligands have now been published by the consortium.^{6–8}

Target 2035 is an international open science initiative that aims to generate chemical or biological modulators for nearly all human proteins by 2035.^{1,9–12} Initially defined by scientists from academia and the pharmaceutical industry and driven by the Structural Genomics Consortium (SGC), this initiative has grown into a global effort with the goal of making chemical and biological tools and data freely available to the research community. Importantly, the developed tools are peer reviewed, to promote confidence that they are fit for purpose in studying the target of interest.^{1,13,14} A major contribu-

tor to these efforts is the European Union Innovative Medicines Initiative project “Enabling and Unlocking Biology in the OPEN”, or ‘EUBOPEN’ for short. Here, we provide an overview and summary of recent achievements of EUBOPEN, and comment on how they enhance Target 2035 activities.

EUBOPEN as a major contributor to target 2035

EUBOPEN is a public–private partnership with ambitious goals to create, distribute and annotate the largest openly available set of high-quality chemical modulators for human proteins <https://www.eubopen.org/>. EUBOPEN consists of 22 partners from academia and the pharmaceutical industry, working together on projects that include novel chemical probes for emerging target areas such as solute carriers (SLCs), E3-ubiquitin ligases (E3s) and other understudied target families, as well as a chemogenomic (CG) library covering one third of the druggable genome. All compounds in the collections are comprehensively characterised for their potency, selectivity, and cellular activity. The compounds are annotated with a suite of biochemical and cell-based assays, including those derived from primary patient cells. Diseases of particular focus for the development of primary cell assays include inflammatory bowel disease, cancer and neurodegeneration. Finally, the development of new technologies to significantly shorten the hit identification and the hit-to-lead optimisation processes, aim to provide the foundation for future efforts toward Target 2035.

Developing, validating and distributing peer reviewed chemical probes

Chemical probes represent the gold standard amongst chemical tools. They are highly characterised, potent and selective, cell-active small molecules that modulate the function of a protein.^{1,15,16} Commonly, these molecules are inhibitors or degraders, but for some target families, such as G-protein coupled receptors (GPCRs) other forms of modulations, such as agonist or antagonist activities, exist. Strict criteria for chemical probes have been agreed upon by the EUBOPEN consortium members. These criteria include potency measured in *in vitro* assays of less than 100 nM, a selectivity of at least 30 fold over related proteins, evidence of target engagement in cells at less than 1 μ M or 10 μ M for shallow protein–protein interaction targets, and a reasonable cellular toxicity window (unless cell death is target mediated). EUBOPEN aims to deliver 50 new, collaboratively developed chemical probes within the 5-year programme with a specific focus on E3 ligases and SLCs. EUBOPEN scientists have ongoing collaborations with many academic researchers outside of the consortium as well as five pharmaceutical companies that are members of the European Federation of Pharmaceutical Industries and Associations (EFPIA). These collaborations are



conducted in a pre-competitive manner with EUBOPEN scientists from multiple academic institutions. All chemical probes developed by the EUBOPEN consortium are peer reviewed by an external committee and released with a structurally similar inactive negative control compound.

The focus of EUBOPEN on novel challenging target classes, in particular ubiquitin E3 ligases, given their roles as attractive targets in their own right, and as the enzymes hijacked/co-opted by degrader molecules such as molecular glues and PROTACs, has pushed the field to revisit criteria and to include new modalities *e.g.* covalent binders and PROTACs, and E3 ligase handles.¹⁶ A recent output from EUBOPEN illustrates a fitting case study and a compelling example and template for developing chemical probes that meet these evolving qualifying utility criteria. Small-molecule covalent inhibitors of the Cul5-RING ubiquitin E3 ligase substrate receptor subunit SOCS2 were designed by Ramachandran *et al.*¹⁷ to target the hard-to-drug SH2 domain and shown to block substrate recruitment *in vitro* and within cells. Starting from phosho-tyrosine as an anchor bound fragment, crystallographic structure-based design guided the optimisation of compounds with high ligand efficiency. The qualified E3 ligase handle/probe molecule featured effective covalent modification of a specific cysteine residue on the SOCS2 SH2 domain binding site, and a pro-drug strategy to mask the phosphate group and aid cell permeability.¹⁷

EUBOPEN aims to collate and make openly available additional 50 high-quality chemical probes from the community. In this unique project, the Donated Chemical Probes (DCP) project, chemical probes developed by academics and/or industry are peer-reviewed by two independent committees on the chemical probe criteria and made available to researcher worldwide without any restrictions on use.^{15,18} So far, EUBOPEN is on track to generate or collect in total 100 high quality chemical probes from the community by May 2025. All compounds will be freely and openly available and can be requested *via* <https://www.eubopen.org/chemical-probes>. An information sheet with key data and recommendations for their use in cellular assays is provided to avoid or minimise off-target effects when using these compounds.¹⁹ To date, EUBOPEN distributed more than 6000 samples of chemical probes and controls to researchers around the world without restrictions. This initiative aims to accelerate target validation and serve as a foundation for drug discovery. By concentrating on targets where the link to diseases is poorly understood, new applications for small molecule ligands could be uncovered.

Chemogenomic compound collections covering one third of the druggable proteome

The generation of chemical probes can be both costly and challenging, particularly when it comes to optimising a compound for high selectivity within the target family. Achieving

this level of selectivity is not always feasible, which limits the potential to create chemical probes for a wide range of targets. To overcome these limitations and address a significant portion of the druggable genome, a chemogenomics strategy has been applied. Chemogenomic compounds, in contrast to chemical probes, may bind to multiple targets but are still valuable due to their well-characterized target profiles. By leveraging chemogenomic compound sets, researchers can systematically explore interactions between small molecules and a broad spectrum of biological targets, providing insights into druggable pathways and enhancing the efficiency of drug discovery. In addition to data from the literature, EUBOPEN has set up several selectivity panels for different target families to further annotate these compounds. By using a set of these well-characterised compounds with overlapping target profiles, the target responsible for a specific phenotype can be identified. Family-specific criteria have been defined with the help of an external expert committee from different target areas, taking into account the availability of well-characterised compounds and screening possibilities, ligandability of different targets and the possibility to collate more than one chemotype per target. The criteria can be viewed at <https://www.eubopen.org/chemogenomics/chemogenomics-criteria>.

Hundreds of thousands of bioactive compounds generated by medicinal chemistry efforts in the industrial and academic sectors, together with the availability of public compound/bioactivity databases, facilitated the assembly of highly annotated CG libraries for large target families. When EUBOPEN was launched in 2020, prominent public repositories contained a total of 566 735 compounds with a target-associated bioactivity $\leq 10 \mu\text{M}$ covering 2899 human target proteins as CG compound candidates.²⁰ Kinase inhibitors and G-protein coupled receptor (GPCR) ligands dominate the annotated compounds, as these have long been the focus of medicinal chemistry, but several other target families are sufficiently represented for the development of high-quality CG sets. Moreover, at the time of writing, 7818 candidate compounds are commercially available, enabling the assembly of sustainable libraries with broad accessibility for the scientific community. Benefitting from these valuable resources, we set out to assemble CG sets for several large target families. Based upon data from leading public repositories and an in-house curated bioactivity database,²⁰ we identified CG compound candidates meeting our stringent family-specific quality criteria. The candidates underwent rigorous chemical and biological quality control, were validated for their intended activities and profiled for selectivity before potential inclusion in a CG set. Despite considerable attrition due to non-reproducibility of published bioactivity values, we were able to identify up to five chemically diverse CG compounds per target for orthogonality.²¹

Efforts to develop a dedicated kinase CG set have faced a high attrition rate of 30%, mainly due to insufficient potency and selectivity of kinase inhibitors, despite promising published profiles. All compounds are additionally characterised



in biochemical and/or cellular assays and we have developed a gateway <https://gateway.eubopen.org> to make the compound characterisation data available to the public.^{21–25} This comprehensive effort ensures that the data is accessible for further research and development, supporting the scientific community in their endeavors to understand and target various biological pathways.

Patient-derived disease tissue assays for target discovery and validation

Translational medical studies rely on data from disease relevant models. In EUBOPEN, we use our high-quality tool compounds to discover novel target-disease associations. Our specific aims are to develop models and assays based on samples from patients and control subjects (*e.g.* blood, tissue) and test the compounds in unbiased studies. We focus on inflammatory bowel diseases (IBD), colorectal cancer (CRC), liver fibrosis (LF) and multiple sclerosis (MS). In these areas, we developed and validated several patient-derived cell assays (PCAs) <https://www.eubopen.org/tissue-assays>, cooperating closely with hospitals, clinicians, and patients.

We have established patient-derived organoids (PDOs) from healthy and CRC tissues to study tumour-specific responses. Screening of chemical probes revealed patient-specific tumour vulnerabilities.¹⁸ Viability data from other cell systems allows for the identification of compounds with a favourable toxicity profile. PDOs were further co-cultured with matched fibroblasts to mimic native growth conditions. Subsequent co-cultures from 29 patients allowed us to study therapy resistance in the tumour microenvironment and the identification of compound combinations to overcome resistance.²⁶ Furthermore, our focus is on the identification of chemical probes that can restore compromised intestinal epithelial barrier, promoting mucosal healing. Thus, we have established an automated screening platform using gastrointestinal PDOs co-cultured with matched fibroblasts and are testing the effects of chemical probes. Other developments include gut biomimetic scaffolds for controlled organoid growth²⁷ and identification of pathways that uncouple intestinal regeneration from tumorigenesis. In addition, spheroid models for metabolic dysfunction-associated steatohepatitis (MASH) from patient-derived cells have been established and characterised, which showed that they recapitulate interactions between hepatocytes and immune cells and closely resembles the disease tissue using multi-dimensional omics.²⁸ The system has been optimised for compound screening with phenotypic and functional readouts, *e.g.* inflammation and fibrogenesis. We have identified novel potential intervention points for therapy, including a novel GPCR-ion channel axis and other signalling networks controlling human hepatocyte regeneration.²⁹

The important role of B cells and epigenetics in multiple sclerosis³⁰ prompted us to establish a B cell assay using primary patient cells to test and measure effects of chemical probes and chemogenomic compounds on *e.g.* pro-inflammatory cytokines.

Preliminary results have revealed interesting target-disease associations, which are now undergoing validation studies. In summary, we have established validated PCAs in four main disease areas and demonstrated the value of EUBOPEN chemogenomic libraries and chemical probes for the identification of new potential disease targets and biomarkers for further studies.

Implementing FAIR principles in the EUBOPEN data management strategy

Publication of high quality, open datasets is also a core deliverable of EUBOPEN. Primarily this consists of data and metadata describing small molecules as part of chemogenomic or chemical tool collections and contains a broad range of data types derived from biophysical, biochemical and cellular assays. Additionally, data from structural biology, proteomics and biological probe generation are also made available.

Our experience has been that it is critical to capture data and metadata from day one in a well-defined and accessible manner to ensure close adherence to FAIR (findable, accessible, interoperable, and re-usable) principles. Within EUBOPEN, we have leveraged our data management approaches from the SGC and Oxford's Centre for Medicines Discovery, employing a laboratory information management system (LIMS), Scarab, from MolSoft LLC. This platform, which we call EUBScarab, is able to capture unstructured data, such as lab notebooks, as well as structured and workflow data. Importantly, Scarab can be adapted rapidly to accommodate new data types without needing any changes in its codebase. The use of EUBScarab has meant that the deposition path of chemical biology data into both ChEMBL and BioImage Archive has been relatively frictionless, as we have been able to adapt our data capture approaches to map tightly with the existing ChEMBL database schema, for example. Nevertheless, close collaboration with colleagues at the EMBL's European Bioinformatics Institute (EBI) has resulted in fruitful solutions to provide new ways of capturing and presenting EUBOPEN data to the community, ensuring long-term sustainability and adherence to FAIR principles.

EUBOPEN datasets are stored in the public resource most relevant to the data type so, for example, pharmacology data from bioassays is stored in ChEMBL and protein structures are stored in the wwPDB. The EUBOPEN Gateway is a bespoke web interface and supporting infrastructure that provides the global user community with facile access to EUBOPEN data. The gateway is designed to enable users who are not informatics experts to easily answer common questions and use-cases. These range from the relatively straightforward, such as “what chemical probes are available for my target?” and “what pharmacology data is available for this chemogenomic set compound?” to more complex queries such as “I have a number of active compounds from screening the chemogenomic set in my phenotypic assay: what target(s) are likely to be responsible?”

Various screenshots from the currently released version of the gateway can be seen in Fig. 1. Functions are provided to



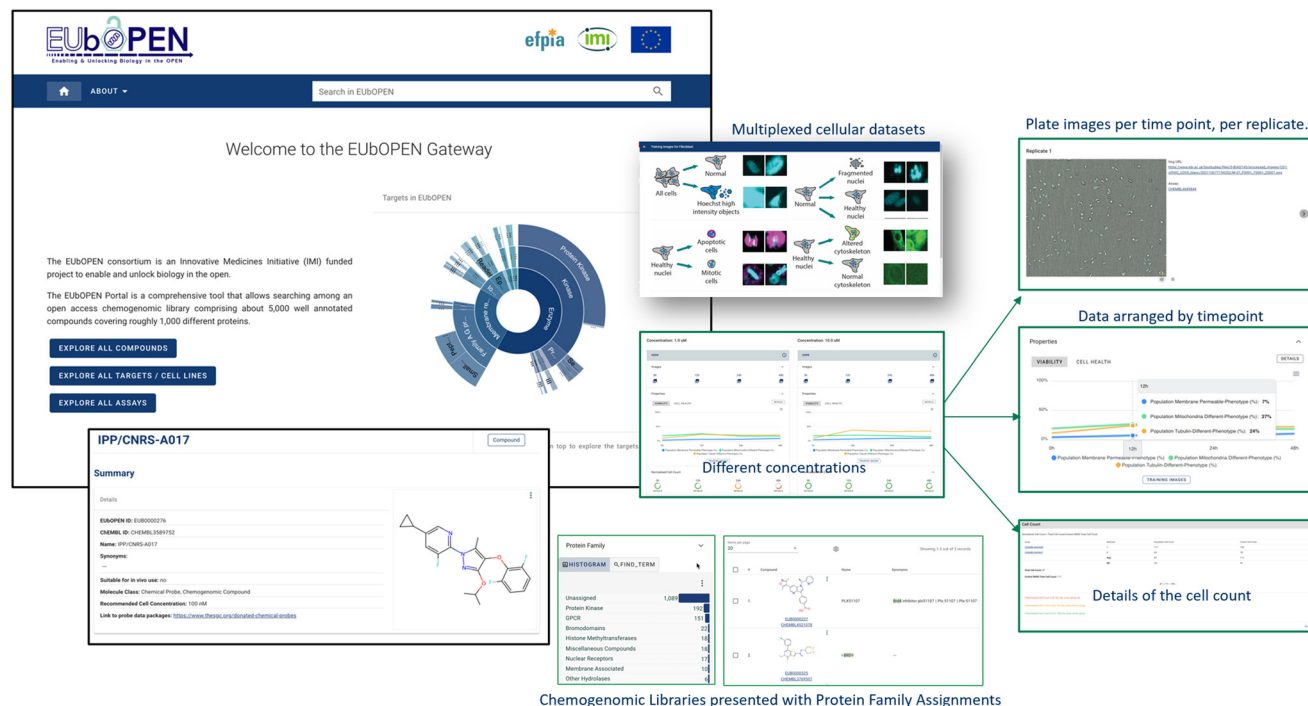


Fig. 1 Screenshots from the EUBOPEN Gateway.

allow users to search, filter and visualise results across molecules, targets and assays. Of particular note is the development of capabilities to display complex multiplexed data, for example from cell-health assays. This example also illustrates how different data types are stored in the most relevant resource (here, pharmacology data in ChEMBL and imaging data in the BioImage Archive). The gateway moreover provides links to the underlying core databases to enable more detailed access, interrogation and to download the data for *in silico* applications.

To date, >600 000 activity data points from >14 000 assays covering >1500 targets have been deposited into ChEMBL. The EUBOPEN datasets constitute a highly valuable, high-quality, resource for chemical biology and drug discovery that will continue to evolve and grow to the end of the project and beyond. We anticipate that it will prove to be invaluable for advanced machine learning methods in the chemical biology space.

Technology development for accelerating hit-to-lead chemistry

While chemical probes have proven to be powerful reagents,¹ their discovery and development remains an ongoing challenge in medicinal chemistry and chemical biology. As in traditional drug discovery programmes, the process starts with the identification of bioactive small molecules, followed by iterative cycles of medicinal chemistry for lead generation. Additionally, characterising their selectivity profile is crucial to ensuring high-quality chemical probes.³¹ This laborious, time, and cost-intensive workflow emphasises the signifi-

cance of automation and new technologies in expediting and enhancing the generation and characterisation of chemical probes, a focal point of the EUBOPEN project.

Computer aided drug design (CADD), fragment-based drug discovery (FBDD)³² and DNA-encoded library (DEL) technology³³ have emerged as powerful hit-finding approaches, enabling the exploration of challenging targets and broadening the accessible boundaries of chemical space. Despite these advancements, challenges persist. While automation is widely used across various stages of drug discovery workflows,³⁴ integration across adjacent stages remains uncommon. Moreover, the reliance on a limited reaction toolkit dominates molecular discovery, resulting in an uneven exploration of chemical space within drug discovery programmes.³⁵

Initiatives within EUBOPEN focus on enhancing hit-to-lead timelines by expanding reaction toolkits and fragment sets for plate-based chemistry. Employing a structure-blind, function-driven approach, arrays of reactions are conducted *via* plate-based chemistry, with crude products screened for biological function, facilitating activity-directed discovery.³⁶ Here, only bioactive products are purified for structure elucidation. Unlike conventional workflows favouring specific designed molecules, this approach allows exploration of chemistry with multiple regiochemical outcomes and functionalisation of different heteroatoms (*e.g.* metal-carbenoid chemistry).^{37–39}

Another successful development involves methods for progressing hits from XChem crystallographic fragment screens. In an early proof-of-concept experiment based on fragment hits against the second bromodomain of PHIP (PHIP(2)),⁴⁰ a diverse 2000-compound follow-up library was synthesised by low-cost robotics, and all products were tested



as crude reaction mixtures by crystallography, as a second round XChem experiment. This effort yielded micromolar hits in both biochemical and biophysical assays,⁴¹ providing actionable structure–activity relationship (SAR) data from the crystal structures.⁴² This experiment worked because the crystallised protein can extract (purify) the active species from solution, and crystallography is sensitive enough to observe it. The result forms the basic premise underlying a new platform, Fast-forward fragments (FFF), which combines an experimental workflow with a new algorithmic approach for computing fragment merges, the fragment knitwork.⁴³

Finally, as covalent libraries are highly valuable in biology and drug discovery, a platform for acoustic high-throughput automated synthesis⁴⁴ facilitates rapid generation of diverse covalent screening libraries and parallel screening without purification.⁴⁵ This method streamlines the synthesis of focused screening libraries, reducing the need for extensive medicinal chemistry efforts typically associated with fragment based drug discovery programmes.

Integrating automation with algorithm-assisted design enhances efficiency in drug discovery by optimizing reaction conditions, predicting molecular properties, and prioritizing compounds for further investigation.^{46,47} This integration also holds promise in addressing probe selectivity by enabling high-throughput screening against diverse targets and facilitating the prediction of selectivity profiles based on molecular properties and structural features. Combining these approaches with other EUBOPEN initiatives towards detailed probe selectivity profiles ensures the reliability and specificity of these reagents. Family-wide cellular screening platforms,^{48,49} as well as new mass spectrometry based technologies⁵⁰ help streamline the process of probe development.⁵¹ Overall, this integrated approach accelerates the discovery of reliable and selective chemical tools for biological system interrogation.

Outlook

By the end of the EUBOPEN project in May 2025, this initiative will have contributed more than 100 novel chemical probes and collected chemogenomic modulators covering 1000 targets. For each chemical probe, characterisation data will be available on the EUBOPEN Gateway as well as in the ChEMBL database enabling both human readability and machine learning. The current progress of this ambitious project indicates that these milestones are within reach, providing the chemical biology community with powerful tools for the exploitation of novel disease mechanisms.

Despite this major contribution of EUBOPEN to the exploration of the druggable proteome, much remains to be done. Many targets in established or previously unknown target families remain unexplored. In the future, more efficient and cost effective hit-finding strategies and advances in machine learning and AI technologies are expected to play pivotal roles in this area in the future. As these technologies rely on robust data for algorithm refinement and optimisation, the

EUBOPEN Gateway will provide a crucial resource for machine learning technologies. In summary, while EUBOPEN will significantly enhance the chemical biology toolkit by 2025, continued innovation and investment will be critical for advancing drug discovery and understanding complex biological systems in the years to come.

Disclaimer

This communication reflects the views of the authors and neither IMI nor the European Union, EFPIA or any Associated Partners are liable for any use that may be made of the information contained herein.

Data availability

No primary research results, software or code have been included and no new data were generated or analysed as part of this review.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This project has received funding from 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 Högskolan, Diamond Light Source Limited.

References

- 1 C. H. Arrowsmith, J. E. Audia, C. Austin, J. Baell, J. Bennett, J. Blagg, C. Bountra, P. E. Brennan, P. J. Brown, M. E. Bunnage, C. Buser-Doepner, R. M. Campbell, A. J. Carter, P. Cohen, R. A. Copeland, B. Cravatt, J. L. Dahlin, D. Dhanak, A. M. Edwards, M. Frederiksen, S. V. Frye, N. Gray, C. E. Grimshaw, D. Hepworth, T. Howe, K. V. Huber, J. Jin, S. Knapp, J. D. Kotz, R. G. Kruger, D. Lowe, M. M. Mader, B. Marsden, A. Mueller-Fahrnow, S. Muller, R. C. O'Hagan, J. P. Overington, D. R. Owen, S. H. Rosenberg, B. Roth, R. Ross, M. Schapira, S. L. Schreiber, B. Shoichet, M. Sundstrom, G. Superti-Furga, J. Taunton, L. Toledo-Sherman, C. Walpole, M. A. Walters, T. M. Willson, P. Workman, R. N. Young and W. J. Zuercher, *Nat. Chem. Biol.*, 2015, **11**, 536–541.
- 2 C. I. Wells, H. Al-Ali, D. M. Andrews, C. R. M. Asquith, A. D. Axtman, I. Dikic, D. Ebner, P. Ettmayer, C. Fischer, M. Frederiksen, R. E. Futrell, N. S. Gray, S. B. Hatch, S. Knapp, U. Lucking, M. Michaelides, C. E. Mills, S. Muller, D. Owen, A. Picado, K. S. Saikatendu, M. Schroder, A. Stolz, M. Tellechea, B. J. Turunen, S. Vilar, J. Wang, W. J. Zuercher,



- T. M. Willson and D. H. Drewry, *Int. J. Mol. Sci.*, 2021, **22**, 566.
- 3 X. Liu and A. Ciulli, *ACS Cent. Sci.*, 2023, **9**, 1269–1284.
 - 4 D. A. Nalawansha, K. Mangano, W. den Besten and P. R. Potts, *ChemBioChem*, 2024, **25**, e202300712.
 - 5 S. Singh, W. Tian, Z. C. Severance, S. K. Chaudhary, V. Anokhina, B. Mondal, R. Pergu, P. Singh, U. Dhawa, S. Singha and A. Choudhary, *Chem. Soc. Rev.*, 2023, **52**, 5485–5515.
 - 6 D. D. G. Owens, M. E. R. Maitland, A. Khalili Yazdi, X. Song, V. Reber, M. P. Schwalm, R. A. C. Machado, N. Bauer, X. Wang, M. M. Szewczyk, C. Dong, A. Dong, P. Loppnau, M. F. Calabrese, M. S. Dowling, J. Lee, J. I. Montgomery, T. N. O'Connell, C. Subramanyam, F. Wang, E. C. Adamson, M. Schapira, M. Gstaiger, S. Knapp, M. Vedadi, J. Min, G. A. Lajoie, D. Barsyte-Lovejoy, D. R. Owen, C. Schild-Poulter and C. H. Arrowsmith, *Nat. Chem. Biol.*, 2024, **20**, 1164–1175.
 - 7 A. S. M. Li, S. Kimani, B. Wilson, M. Nouredin, H. Gonzalez-Alvarez, A. Mamai, L. Hoffer, J. P. Guilingier, Y. Zhang, M. von Rechenberg, J. S. Disch, C. J. Mulhern, B. L. Slakman, J. W. Cuzzo, A. Dong, G. Poda, M. Mohammed, P. Saraon, M. Mittal, P. Modh, V. Rathod, B. Patel, S. Ackloo, V. Santhakumar, M. M. Szewczyk, D. Barsyte-Lovejoy, C. H. Arrowsmith, R. Marcellus, M. A. Guie, A. D. Keefe, P. J. Brown, L. Halabelian, R. Al-Awar and M. Vedadi, *J. Med. Chem.*, 2023, **66**, 5041–5060.
 - 8 C. J. Munoz Sosa, C. Lenz, A. Hamann, F. Farges, J. Dopfer, A. Kramer, V. Cherkashyna, A. Tarnovskiy, Y. S. Moroz, E. Proschak, V. Nemec, S. Muller, K. Saxena and S. Knapp, *ACS Chem. Biol.*, 2024, **19**, 1638–1647.
 - 9 A. J. Carter, O. Kraemer, M. Zwick, A. Mueller-Fahrnow, C. H. Arrowsmith and A. M. Edwards, *Drug Discovery Today*, 2019, **24**, 2111–2115.
 - 10 M. Zwick, O. Kraemer and A. J. Carter, *Data Brief*, 2019, **25**, 104284.
 - 11 S. Muller, S. Ackloo, A. Al Chawaf, B. Al-Lazikani, A. Antolin, J. B. Baell, H. Beck, S. Beedie, U. A. K. Betz, G. A. Bezerra, P. E. Brennan, D. Brown, P. J. Brown, A. N. Bullock, A. J. Carter, A. Chaikuad, M. Chaineau, A. Ciulli, I. Collins, J. Dreher, D. Drewry, K. Edfeldt, A. M. Edwards, U. Egner, S. V. Frye, S. M. Fuchs, M. D. Hall, I. V. Hartung, A. Hillisch, S. H. Hitchcock, E. Homan, N. Kannan, J. R. Kiefer, S. Knapp, M. Kostic, S. Kubicek, A. R. Leach, S. Lindemann, B. D. Marsden, H. Matsui, J. L. Meier, D. Merk, M. Michel, M. R. Morgan, A. Mueller-Fahrnow, D. R. Owen, B. G. Perry, S. H. Rosenberg, K. S. Saikatendu, M. Schapira, C. Scholten, S. Sharma, A. Simeonov, M. Sundstrom, G. Superti-Furga, M. H. Todd, C. Tredup, M. Vedadi, F. von Delft, T. M. Willson, G. E. Winter, P. Workman and C. H. Arrowsmith, *RSC Med. Chem.*, 2022, **13**, 13–21.
 - 12 S. Ackloo, A. A. Antolin, J. M. Bartolome, H. Beck, A. Bullock, U. A. K. Betz, J. Bottcher, P. J. Brown, M. Chaturvedi, A. Crisp, D. Daniels, J. Dreher, K. Edfeldt, A. M. Edwards, U. Egner, J. Elkins, C. Fischer, T. Glendorf, S. Goldberg, I. V. Hartung, A. Hillisch, E. Homan, S. Knapp, M. Koster, O. Kramer, J. Llaveria, U. Lessel, S. Lindemann, L. Linderorth, H. Matsui, M. Michel, F. Montel, A. Mueller-Fahrnow, S. Muller, D. R. Owen, K. S. Saikatendu, V. Santhakumar, W. Sanderson, C. Scholten, M. Schapira, S. Sharma, B. Shireman, M. Sundstrom, M. H. Todd, C. Tredup, J. Venable, T. M. Willson and C. H. Arrowsmith, *RSC Med. Chem.*, 2023, **14**, 1002–1011.
 - 13 M. M. Mader, J. Rudolph, I. V. Hartung, D. Uehling, P. Workman and W. Zuercher, *Cancer Discovery*, 2023, **13**, 2150–2165.
 - 14 O. Hsia, M. Hinterndorfer, A. D. Cowan, K. Iso, T. Ishida, R. Sundaramoorthy, M. A. Nakasone, H. Imrichova, C. Schatz, A. Rukavina, K. Husnjak, M. Wegner, A. Correa-Saez, C. Craigon, R. Casement, C. Maniaci, A. Testa, M. Kaulich, I. Dikic, G. E. Winter and A. Ciulli, *Nature*, 2024, **627**, 204–211.
 - 15 S. Muller, S. Ackloo, C. H. Arrowsmith, M. Bauser, J. L. Baryza, J. Blagg, J. Bottcher, C. Bountra, P. J. Brown, M. E. Bunnage, A. J. Carter, D. Damerell, V. Dotsch, D. H. Drewry, A. M. Edwards, J. Edwards, J. M. Elkins, C. Fischer, S. V. Frye, A. Gollner, C. E. Grimshaw, A. Ijzerman, T. Hanke, I. V. Hartung, S. Hitchcock, T. Howe, T. V. Hughes, S. Laufer, V. M. Li, S. Liras, B. D. Marsden, H. Matsui, J. Mathias, R. C. O'Hagan, D. R. Owen, V. Pande, D. Rauh, S. H. Rosenberg, B. L. Roth, N. S. Schneider, C. Scholten, K. Singh Saikatendu, A. Simeonov, M. Takizawa, C. Tse, P. R. Thompson, D. K. Treiber, A. Y. Viana, C. I. Wells, T. M. Willson, W. J. Zuercher, S. Knapp and A. Mueller-Fahrnow, *eLife*, 2018, **7**, e34311.
 - 16 I. V. Hartung, J. Rudolph, M. M. Mader, M. P. C. Mulder and P. Workman, *J. Med. Chem.*, 2023, **66**, 9297–9312.
 - 17 S. Ramachandran, N. Makukhin, K. Haubrich, M. Nagala, B. Forrester, D. M. Lynch, R. Casement, A. Testa, E. Bruno, R. Gitto and A. Ciulli, *Nat. Commun.*, 2023, **14**, 6345.
 - 18 C. Tredup, B. Ndreshkjana, N. S. Schneider, A. Tjaden, A. M. Kemas, S. Youhanna, V. M. Lauschke, B. T. Berger, A. Kramer, L. M. Berger, S. Rohm, S. Knapp, H. F. Farin and S. Muller, *ACS Chem. Biol.*, 2023, **18**, 822–836.
 - 19 J. Sterling, J. R. Baker, A. McCluskey and L. Munoz, *Nat. Commun.*, 2023, **14**, 3228.
 - 20 L. Isigkeit, A. Chaikuad and D. Merk, *Molecules*, 2022, **27**, 2513.
 - 21 L. Isigkeit, E. Schallmayer, R. Busch, L. Brunello, A. Menge, L. Elson, S. Muller, S. Knapp, A. Stolz, J. A. Marschner and D. Merk, *Nat. Commun.*, 2024, **15**, 5201.
 - 22 L. Elson, A. Tjaden, S. Knapp and S. Muller, *Methods Mol. Biol.*, 2023, **2706**, 75–88.
 - 23 A. Tjaden, S. Knapp and S. Muller, *Methods Mol. Biol.*, 2023, **2706**, 59–73.
 - 24 V. Nemec and S. Knapp, *Methods Mol. Biol.*, 2023, **2706**, 51–58.
 - 25 A. Chaikuad and D. Merk, *Methods Mol. Biol.*, 2023, **2706**, 1–10.
 - 26 H. F. Farin, M. H. Mosa, B. Ndreshkjana, B. M. Grebbin, B. Ritter, C. Menche, K. B. Kennel, P. K. Ziegler, L. Szabo, J. Bollrath, D. Rieder, B. E. Michels, A. Kress, M. Bozlar, T. Darvishi, S. Stier, I. M. Kur, K. Bankov, R. Kesselring, S. Fichtner-Feigl, B. Brune, T. O. Goetze, S. E. Al-Batran, C. H. Brandts, W. O. Bechstein, P. J. Wild, A. Weigert, S. Muller, S.



- Knapp, Z. Trajanoski and F. R. Greten, *Cancer Discovery*, 2023, **13**, 2192–2211.
- 27 N. Pentinmikko, R. Lozano, S. Scharaw, S. Andersson, J. I. Englund, D. Castillo-Azofeifa, A. Gallagher, M. Broberg, K. Y. Song, A. Sola Carvajal, A. T. Speidel, M. Sundstrom, N. Allbritton, M. M. Stevens, O. D. Klein, A. Teixeira and P. Katajisto, *Sci. Adv.*, 2022, **8**, eabm1847.
- 28 E. Barreby, B. Strunz, S. Nock, L. Naudet, J. X. Shen, H. Johansson, I. Sonnerborg, J. Ma, E. Urgard, L. J. Pallett, Y. Hu, A. Fardellas, V. Azzimato, A. Vankova, L. Levi, C. Morgantini, M. K. Maini, P. Stal, S. P. Rosshart, J. M. Coquet, G. Nowak, E. Naslund, V. M. Lauschke, E. Ellis, N. K. Bjorkstrom, P. Chen and M. Aouadi, *Nat. Metab.*, 2023, **5**, 1188–1203.
- 29 N. Oliva-Vilarnau, C. M. Beusch, P. Sabatier, E. Sakaraki, A. Tjaden, L. Graetz, F. A. Buttner, D. Dorotea, M. Nguyen, F. Bergqvist, Y. Sundstrom, S. Muller, R. A. Zubarev, G. Schulte, C. Tredup, R. Gramignoli, U. J. F. Tietge and V. M. Lauschke, *Hepatology*, 2024, **79**, 1337–1351.
- 30 E. Ewing, L. Kular, S. J. Fernandes, N. Karathanasis, V. Lagani, S. Ruhrmann, I. Tsamardinos, J. Tegner, F. Piehl, D. Gomez-Cabrero and M. Jagodic, *EBioMedicine*, 2019, **43**, 411–423.
- 31 M. E. Bunnage, E. L. Chekler and L. H. Jones, *Nat. Chem. Biol.*, 2013, **9**, 195–199.
- 32 D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discovery*, 2016, **15**, 605–619.
- 33 C. A. Reiher, D. P. Schuman, N. Simmons and S. E. Wolkenberg, *ACS Med. Chem. Lett.*, 2021, **12**, 343–350.
- 34 G. Schneider, *Nat. Rev. Drug Discovery*, 2018, **17**, 97–113.
- 35 J. Bostrom, D. G. Brown, R. J. Young and G. M. Keseru, *Nat. Rev. Drug Discovery*, 2018, **17**, 922.
- 36 G. Karageorgis, S. Warriner and A. Nelson, *Nat. Chem.*, 2014, **6**, 872–876.
- 37 G. Karageorgis, M. Dow, A. Aimon, S. Warriner and A. Nelson, *Angew. Chem., Int. Ed.*, 2015, **54**, 13538–13544.
- 38 A. Leggott, J. E. Clarke, S. Chow, S. L. Warriner, A. J. O'Neill and A. Nelson, *Chem. Commun.*, 2020, **56**, 8047–8050.
- 39 A. I. Green, F. Hobor, C. P. Tinworth, S. Warriner, A. J. Wilson and A. Nelson, *Chemistry*, 2020, **26**, 10682–10689.
- 40 H. Grosjean, A. Aimon, S. Hassell-Hart, W. Thompson, L. Koekemoer, J. Bennett, C. Anderson, C. Wild, W. Bradshaw, E. A. FitzGerald, T. Krojer, A. Bradley, O. Fedorov, P. C. Biggin, J. Spencer and F. von Delf, *ChemRxiv*, 2024, preprint, DOI: [10.26434/chemrxiv-2023-6m2s0-v2](https://doi.org/10.26434/chemrxiv-2023-6m2s0-v2).
- 41 O. B. Cox, T. Krojer, P. Collins, O. Monteiro, R. Talon, A. Bradley, O. Fedorov, J. Amin, B. D. Marsden, J. Spencer, F. von Delft and P. E. Brennan, *Chem. Sci.*, 2016, **7**, 2322–2330.
- 42 S. Wills, R. Sanchez-Garcia, T. Dudgeon, S. D. Roughley, A. Merritt, R. E. Hubbard, J. Davidson, F. von Delft and C. M. Deane, *J. Chem. Inf. Model.*, 2023, **63**, 3423–3437.
- 43 S. Wills, R. Sanchez-Garcia, S. D. Roughley, A. Merritt, R. E. Hubbard, F. von Delft and C. M. Deane, *bioRxiv*, 2024, preprint, DOI: [10.1101/2024.08.02.606367](https://doi.org/10.1101/2024.08.02.606367).
- 44 F. Sutanto, S. Shaabani, C. G. Neochoritis, T. Zarganes-Tzitzikas, P. Patil, E. Ghonchepour and A. Domling, *Sci. Adv.*, 2021, **7**, eabd9307.
- 45 Y. Jia, M. Oyken, R. Q. Kim, R. T. N. Tjokrodrijo, A. H. de Ru, A. P. A. Janssen, S. M. Hacker, P. A. van Veelen, P. P. Geurink and A. Sapmaz, *J. Med. Chem.*, 2024, **67**, 7935–7953.
- 46 D. Francis, S. Farooque, A. Meager, D. Derks, A. Leggott, S. Warriner, A. J. O'Neill and A. Nelson, *Org. Biomol. Chem.*, 2022, **20**, 9672–9678.
- 47 K. L. Saar, W. McCorkindale, D. Fearon, M. Boby, H. Barr, A. Ben-Shmuel, C. M. Consortium, N. London, F. von Delft, J. D. Chodera and A. A. Lee, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**, e2214168120.
- 48 M. P. Schwalm, L. M. Berger, M. N. Meuter, J. D. Vasta, C. R. Corona, S. Rohm, B. T. Berger, F. Farges, S. M. Beinert, F. Preuss, V. Morasch, V. V. Rogov, S. Mathea, K. Saxena, M. B. Robers, S. Muller and S. Knapp, *Front. Cell Dev. Biol.*, 2022, **10**, 886537.
- 49 M. B. Robers, J. M. Wilkinson, J. D. Vasta, L. M. Berger, B. T. Berger and S. Knapp, *STAR Protoc.*, 2021, **2**, 100822.
- 50 V. Reber and M. Gstaiger, *Methods Mol. Biol.*, 2023, **2706**, 177–190.
- 51 D. D. G. Owens, M. E. R. Maitland, A. Khalili Yazdi, X. Song, V. Reber, M. P. Schwalm, R. A. C. Machado, N. Bauer, X. Wang, M. M. Szweczyk, C. Dong, A. Dong, P. Loppnau, M. F. Calabrese, M. S. Dowling, J. Lee, J. I. Montgomery, T. N. O'Connell, C. Subramanyam, F. Wang, E. C. Adamson, M. Schapira, M. Gstaiger, S. Knapp, M. Vedadi, J. Min, G. A. Lajoie, D. Barsyte-Lovejoy, D. R. Owen, C. Schild-Poulter and C. H. Arrowsmith, *Nat. Chem. Biol.*, 2024, **20**, 1164–1175.

