Nucleic Acids Research, 2025, **53**, D1683–D1691
https://doi.org/10.1093/nar/gkae868
Advance access publication date: 7 October 2024
Database issue

MolBiC: the cell-based landscape illustrating molecular bioactivities

Yichao Ge[1,2,3,4,†], Mengjie Yang[5,†], Xinyuan Yu[1,†], Ying Zhou[1], Yintao Zhang[1], Minjie Mou[1], Zhen Chen[1], Xiuna Sun[1], Feng Ni[6,7], Tingting Fu[1,*], Shuiping Liu[5,*], Lianyi Han[3,4,*] and Feng Zhu[1,2,*]

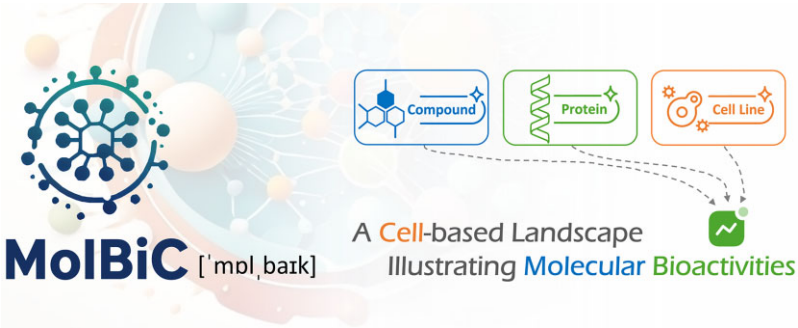
[1]College of Pharmaceutical Sciences, The Second Affiliated Hospital, Zhejiang University School of Medicine, State Key Laboratory of Advanced Drug Delivery and Release Systems, Zhejiang University, Hangzhou 310058, China
[2]Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou 330110, China
[3]Department of Dermatology, Huashan Hospital, Fudan University, Shanghai Institute of Dermatology, Shanghai 200040, China
[4]Greater Bay Area Institute of Precision Medicine, School of Life Sciences, Guangzhou, Guangzhou 511458, China
[5]Key Laboratory of Elemene Class Anti-Cancer Chinese Medicines, School of Pharmacy, Hangzhou Normal University, Hangzhou 311121, China
[6]Institute of Drug Discovery Technology, Ningbo University, Ningbo 315211, China
[7]LeadArt Biotechnologies Ltd., Ningbo 315201, China

[*]To whom correspondence should be addressed. Tel: +86 189 8946 6518; Fax: +86 571 8820 8444; Email: zhufeng@zju.edu.cn
Correspondence may also be addressed to Lianyi Han. Email: hanlianyi@ipm-gba.org.cn
Correspondence may also be addressed to Shuiping Liu. Email: lsp@hznu.edu.cn
Correspondence may also be addressed to Tingting Fu. Email: futt@zju.edu.cn
[†]The first three authors should be regarded as Joint First Authors.

Abstract

The measurement of *cell-based molecular bioactivity* (CMB) is critical for almost every step of drug development. With the booming application of AI in biomedicine, it is essential to have the CMB data to promote the learning of cell-based patterns for guiding modern drug discovery, but no database providing such information has been constructed yet. In this study, we introduce *MolBiC*, a knowledge base designed to describe valuable data on molecular bioactivity measured within a cellular context. *MolBiC* features 550 093 experimentally validated CMBs, encompassing 321 086 molecules and 2666 targets across 988 cell lines. Our *MolBiC* database is unique in describing the valuable data of CMB, which meets the critical demands for CMB-based big data promoting the learning of cell-based molecular/pharmaceutical pattern in drug discovery and development. *MolBiC* is now freely accessible without any login requirement at: https://idrblab.org/MolBiC/.

Graphical abstract



Introduction

The measurement of *cell-based molecular bioactivity* (CMB) is critical for almost every step of drug development (1–3), such as: target identification/validation, lead identification/optimization and toxicology screenings. In particular, the CMBs are more representative of human physiology than that for purified targets in linking to disease of interest during target identification/validation (4,5);

Received: August 15, 2024. **Revised:** September 13, 2024. **Editorial Decision:** September 19, 2024. **Accepted:** September 20, 2024

relate binding parameter with phenotypic outcomes in analyzing cellular structure–activity relationships during lead identification/optimization (6); facilitate toxicology screenings of drugs by identifying off-targets in native cellular contexts (7,8). All the analyses highlight the demands for the invaluable information of CMBs. With the booming applications of AI in biomedicine, it is also imperative to have a CMB-based database facilitating the learning of cell-based molecular patterns for guiding modern drug discovery (9,10).

So far, various knowledgebases have been developed to provide the molecular bioactivity related data. Some of them describe the molecular bioactivity information as parts of a broader collection of biological/pharmacological data, such as PubChem (11), BioLiP (12) and PDBbind (13); some others focus on offering the activity data of compounds on studied target, such as ChEMBL (14), DUDE (15) and BindingDB (16); the remaining ones specialize in supplying disease-specific target activities of compound, such as TTD (17), DrugBank (18) and KLIFS (19). These available databases explicitly provide the activities between compounds and their target(s), which has thus attracted great research interests and scientific discussions from research communities. However, none of those existing databases describes the valuable data of CMBs, which makes them insufficient to support the process of modern drug discovery (20–22). Therefore, it is essential to have a database that systematically describes the *cell-based molecular bioactivity* information, but no such kind of knowledge base has been constructed yet.

In this study, we developed MolBiC, a comprehensive database designed to provide invaluable data on molecular bioactivity measured within cellular contexts. First, such data were accumulated by searching such keywords as 'cellular assay', 'cellular activity' and 'intracellular target binding' in PubMed, which resulted in 550 093 experiment-validated CMBs between 321 086 compounds and 2666 targets within the context of 988 cell lines. Second, those CMBs were comprehensively collected to our database and quantitatively classified according to their activities, which resulted in 188 027, 177 755 and 73 246 data of *high* (≤100 nM), *medium* (between 100 nM and 10 μM) and *low* (>10 μM) CMBs (23,24), respectively. Finally, a total of 114 disease classes defined by the WHO ICD-11 (such as Alzheimer disease, lung cancer, diabetes mellitus, influenza, follicular lymphoma and chronic obstructive pulmonary disease), 155 biochemical classes of targets (such as cytochrome P450, neurotransmitter receptor, protein kinase, voltage-gated potassium channel and nuclear hormone receptor), and 25 species (such as *Homo sapiens*, *Mus musculus*, *Sus scrofa*, *Rattus norvegicus* and *Oryctolagus cuniculus*) were included into our database, and all data were cross-linked to various well-established molecular biology databases, including Cellosaurus (25), DrugBank (26), PubChem (27), UniProt (28), TTD (29), DrugMap (30) and so on. All in all, *MolBiC* is unique in describing the valuable data of CMB, which meets the critical demand for CMB-based big data facilitating the learning of cell-based molecular/pharmaceutical patterns when guiding modern drug discovery. Our MolBiC database is now fully accessible and free and open to all users without any login requirement at: https://idrblab.org/MolBiC/.

Factual content and data retrieval

Systematic accumulation of cell-based molecular bioactivities into MolBiC

The CMB information were collected into MolBiC using the following procedure. First, cell-based assays were defined as ligand-binding assays conducted within living cells (31). The CMBs were identified based on a comprehensive research literature review. Relevant publications were discovered via PubMed search using such keyword combinations as 'intracellular binding assay', 'cellular activity', 'cellular functional assays', 'intracellular target binding' and 'cell-based drug discovery'. Second, the newly identified literatures underwent systematic validation to ensure their credibility by following a strict data collection protocol proposed in previous studies (32,33). Particularly, the protocol included two important criteria: (a) compound-protein interaction must be observable in living cells, rather than *in vitro* purified proteins/cell lysates, (b) experimentally determined bioactivity (e.g. IC_{50}, EC_{50}) must be available, which should be derived from diverse cell-based assays, such as *cellular thermal shift assay* (CETSA), *fluorescence resonance energy transfer* (FRET) and *activity-based protein profiling* (ABPP). **Third**, all collected CMB data were categorized according to their role in the modern drug development pipeline: target identification (TI), lead identification (LI), lead optimization (LO), toxicology screening (TS) and so on.

As a result, the curation process resulted in a total of 550 093 CMBs. These CMBs were stratified by their activity level: 188 027 *high* (≤100 nM), 177 755 *medium* (between 100 nM and 10 μM) and 73 246 *low* (>10 μM) activities. The database included 18 139 CMB data detected using cell lysates. MolBiC encompassed 321 086 unique molecules, 2666 protein targets and 988 cell lines. As shown in Figure 1, there were a total of 550 093 CMBs of 321 086 molecules against 2666 protein targets based on 988 cell lines. Among those molecules, there were 1111 approved, 1289 clinical trial and 463 preclinical drugs together with 319 223 experimental agents. For those cell lines, they originated from 23 organisms, covered 137 diseases and were from 51 tissues/organs. For those protein targets, they belonged to 155 GO functional families, and 2206 of them were with reported clinical importance (34). Besides, based on the UniProt subcellular location classification system, a total of 2666 proteins in MolBiC were classified into four groups: 1298 membrane-based, 1193 cytoplasmic-based, 612 nuclear-based and 210 extracellular-based. Among these data, a total of 851 proteins were found to be associated with multiple subcellular locations. Moreover, the activities of the studied molecules against purified protein targets were also included in MolBiC, allowing for a comparison between their activities against purified proteins and cell-based assays. Our database comprised a total of 61 345 molecular activity records against purified protein targets.

Substantial variation of CMB data induced by different cellular contexts

The CMB information has shown significant variability due to the influence of the cellular environment, which has therefore attracted great interest from research community (35,36). There are two key variations in the CMB data. First, the variations induced by cellular context. Cellular environment could substantially alter the protein conformation and in turn

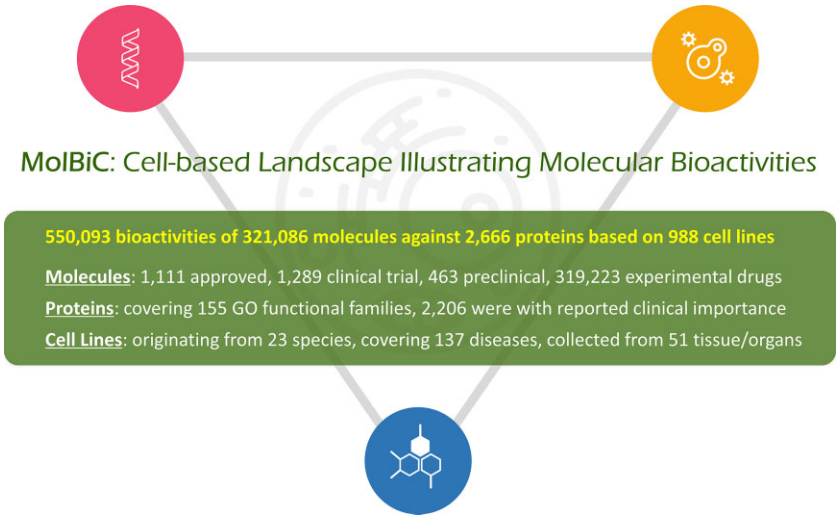


Figure 1. The statistics of all data collected into MolBiC. In total, there were 550 093 *cell-based molecular bioactivities* (CMBs) of 321 086 molecules against 2666 protein targets based on 988 cell lines. For those molecules, there were 1111 approved, 1289 clinical trial and 463 preclinical drugs together with 319 223 experimental agents. For those protein targets, they belonged to 155 GO functional families, and 2206 of them were with reported clinical importance. For those cell lines, they originated from 23 species, covered 137 diseases, and were from 51 tissues/organs.

the resulting molecular interactions compared to those observed with purified proteins (37–39). Taking the approved drug *vorinostat* as an example, the drug exhibited an IC_{50} of 21.6 nM when targeting purified protein HDAC7 (40). However, in a cell-based assay using Sf9 cells, the IC_{50} against the same target increased dramatically to 35 160 nM (41). This significant discrepancy underscored the profound impact of cellular context on drug efficacy. Second, the variations caused by differential expression. The CMB values could vary extensively due to the differential expression of molecule's target protein in distinct cell lines (42–44). Taking the molecule CHEMBL190274 as an example, its binding to protein CASP3 varied substantially among colon cell *DLD-1* (440 nM), pleural effusion cell *T-47D* (1210 nM), ascites cell *ZR-75–1* (3340 nM) and lymph node cell *LNCaP* (10 000 nM). These differences have been attributed to the variable expression levels of CASP3 in these distinct cell lines (45,46). All in all, it is essential to collect the valuable CMB data together with its variations among different cell lines.

To explicitly describe such variations, MolBiC was therefore developed to provide the CMB data in different cellular contexts. Particularly, a total of 550 093 CMBs were systematically provided, which covered 988 cell lines for 137 disease classes. Moreover, the activity information between molecules and purified protein targets was collected, which resulted in a total of 60 483 activities covering 29 003 molecules and 1143 protein targets. In essence, MolBiC data is critical for understanding the occurrence/avoidance of the influence induced by cellular variation.

Diverse CMBs classified based on the different stages of drug development

CMB quantification plays a key role in almost every phase of the drug development, such as target identification, lead identification/optimization, and toxicology screening (47,48). The MolBiC serves as an invaluable repository giving the comprehensive information from multiple perspectives. First, target identification (TI) is considered as the initial stage of drug discovery, which focuses on identifying the function of a target and its role in disease (49–51). CMB data can provide crucial cellular evidence and establish links between prospective target and disease mechanism (52–54). An example is the identification of new multi-kinase targets for indole-based compound via target fishing assays in HeLa cells (55). Second, lead identification (LI) involves the identifications of the chemicals that are synthetically feasible, stable and drug-like (56,57). CMB evaluations in lead identifications can greatly enhance the correlation between binding responses and disease phenotypes (58–60). An illustrative example is the discovery of potent LIMK1/2 inhibitors using the cellular NanoBRET assays of HEK293 (61). Third, lead optimization (LO) focuses on refining the lead by enhancing potency, selectivity and drug-like properties, ultimately producing drug candidates for clinical discovery (62,63). CMB assays facilitate meaningful interpretations of structure–activity relationship in cellular contexts, enabling researchers to fine-tune molecules that facilitate the optimization of efficacies and maximization of successful clinical translations (64,65). A good example is the optimization of an intestine-selective FXR partial agonist derived from *fexaramine* scaffold via NHR assay of CHO-K1 (66). Fourth, toxicology screening (TS) involves the assessment of molecular safety and adverse effects (67–69). CMB evaluations in toxicology screening give insights into the compound behavior across different cell types and physiological contexts, offering a comprehensive safety profile (70–72). An exemplar case includes the assessments of A1R agonists' off-target activities in HEK293 (73). All in all, the CMB assays allowed the unbiased discovery of molecule-target interaction to confirm cellular mode-of-action, serving as an indispensable tool throughout the drug discovery and development process.

The explicit description of the molecules and their affiliated information

It is important to assess the discrepancies among compound–protein interaction profiles observed between *in vitro* purified proteins and intracellular environments (74–76). A critical consideration in this context was the potential changes in protein conformation, post-translational modification and molec-

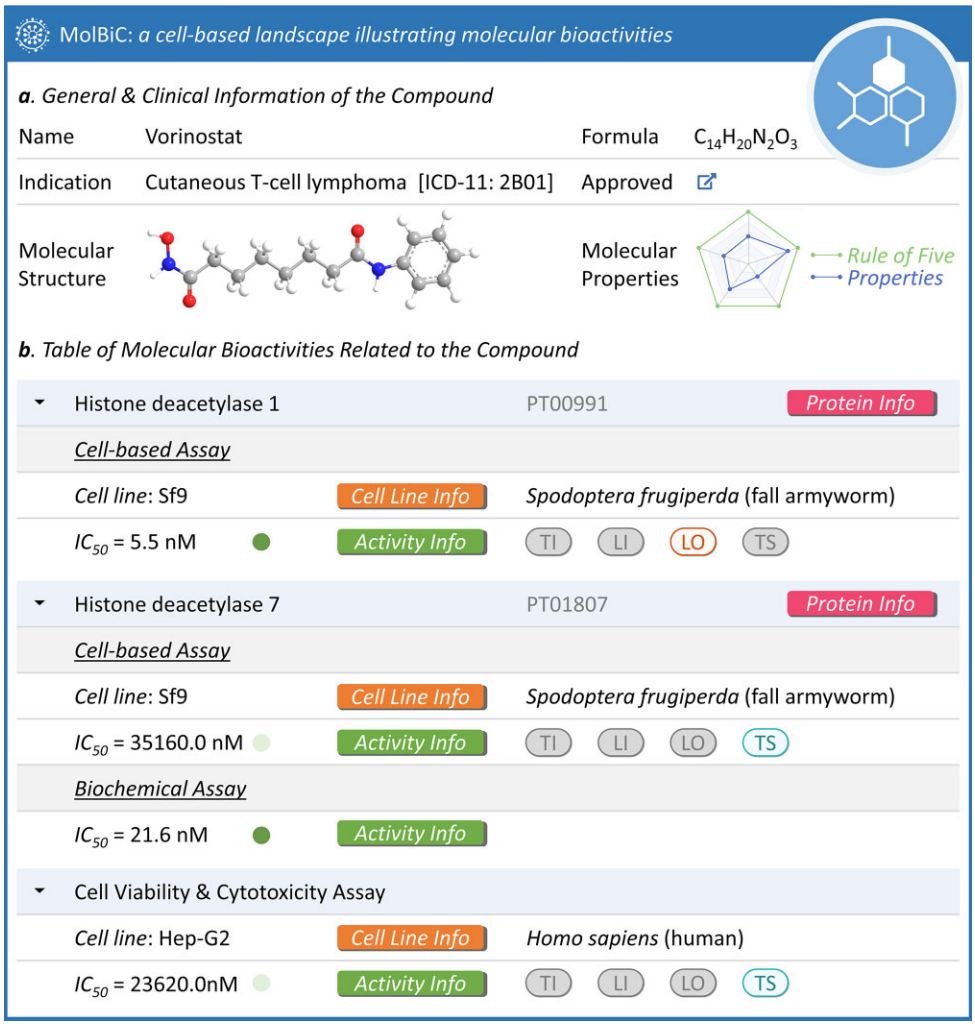


Figure 2. The introduction and visualization of the analyzed molecules collected to MolBiC. (***a***) the general and clinical information of the molecules (such as molecule name, formula, structure, *lipinski's rule of five* characteristics, disease and clinical developmental status). (***b***) the molecular bioactivities of this molecule, including the binding affinity against purified/cellular protein and the pharmacological/toxicological responses in a variety of studied cell lines.

ular interaction (77–79). Moreover, evaluating drug toxicity arising from multi-target effects remained a critical component of the drug development pipeline (80,81). To facilitate the essential investigations, the compound page in MolBiC was meticulously designed to describe the comprehensive data. As illustrated in Figure 2, the drug *vorinostat* was approved for treating cutaneous T-cell lymphoma, which demonstrated potent inhibitory activities against its primary therapeutic target HDAC1 (IC_{50} = 5.5 nM) when expressed in cell (82). To elucidate the potential multi-target binding effects, the CMBs of *vorinostat* against HDAC7 were carefully investigated, which displayed a marked difference in inhibitory potencies between cellular and *in vitro* purified conditions with different IC_{50} values of 35 160 and 21.6 nM (40,41), respectively.

To enhance users' accessibility and utilization of MolBiC data, compound page was meticulously designed to describe the comprehensive array of molecular, clinical and CMB data for a spectrum of molecules. The web interface was constructed to present general and clinical information. The *general information* included molecule names, synonyms, molecular structure, chemical formula, chemical identifiers (including SMILES and InChIKey), analysis of RO5 violations, CAS number and external database cross-references. Those *clinical information* contained drug names, clinical trial statues, disease indications, molecular targets and mechanism of action. CMB data were visualized based on *Sankey diagram* and *tabular format*. The *Sankey diagram* illustrated a four-tiered hierarchical structure. In the *tabular format*, CMB data were mainly stratified by protein name, with subsequent differentiation between cell-based and biochemical assay. Each CMB entry included activity value and classifications. Additionally, the table was supplemented with data from cell viability or cytotoxicity assays for any molecule of interest. To improve the user experience and facilitate comprehensive understanding, the page incorporated the hyperlinked buttons for cell line, protein and bioactivity. These features enabled the seamless navigation of the detailed information for each entity, which allowed reader to conduct in-depth exploration of any specific component within our MolBiC.

The In-depth description of protein targets and their affiliated information

The protein page in MolBiC was clearly described in Figure 3, which provided the reliable data. Taking the HDAC1 as an example. Figure 3a depicted it as a pivotal epigenetic regu-

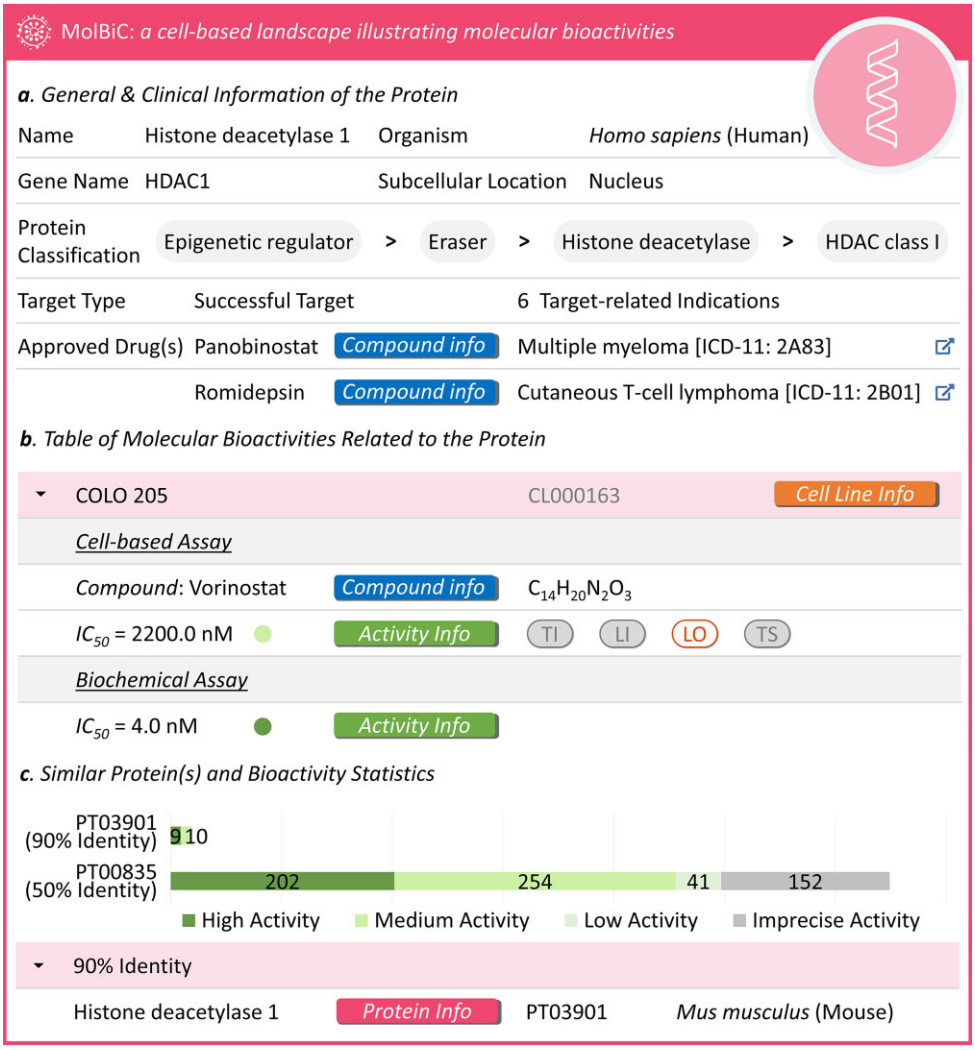


Figure 3. Integrated profile of protein targets in MolBiC. (***a***) The general and clinical information of the target (such as the target name, organism, gene name, subcellular location, function family, clinical status, disease indication and the associated drugs of the studied target). (***b***) The molecular bioactivities related to the studied target (such as cell-based assay and biochemical assay results). (***c***) The activity profiles of the proteins similar to the studied target. Similarity level between target and protein was measured using the *Identity* value generated using BLAST (90), and the activity level were grouped into *high, medium, low* and *imprecise* based on the previous reports (91).

lator treating in multiple diseases (such as multiple myeloma and cutaneous T-cell lymphoma). Figure 3b described the approved drug *vorinostat* that exhibited greatly differential inhibitory potencies against HDAC1 when expressed in distinct cell lines. Particularly, this drug demonstrated an IC_{50} of 13.0 nM in human embryonic kidney (HEK293) cells, whereas in colorectal adenocarcinoma (COLO 205) cells, the IC_{50} raised dramatically to 2200 nM (83,84). Interestingly, when assessed against the purified protein HDAC1, the *vorinostat* exhibited a IC_{50} of 4.0 nM (85). Furthermore, as described in Figure 3c, comparative sequence analysis revealed that the HDAC1 derived from *M. musculus* and the HDAC2 from *H. sapiens* shared 90% and 50% sequence identity with target protein, respectively. From the perspective of CMB data frequency, the latter demonstrated higher prevalence in database. Such information was key for understanding xenobiotic cross-species reactivity and paralog-mediated pharmacological effects in the current drug development.

Protein page was comprehensively curated to encompass extensive molecular, clinical and CMB data for a broad spectrum of protein targets. Each protein page was constructed to present general and clinical information. The *general information* comprised the protein names, synonyms, gene names, protein classification, functional annotations, subcellular localizations, primary sequence data and external database cross-references. The *clinical information* offered protein's target type, related diseases, clinical status, and associated drugs. CMB data were visualized using *Sankey diagrams* and *tabular formats*, facilitating intuitive data interpretation. Additionally, the interface provided statistical analyses of CMB data for homologous proteins, with the identity threshold established to 100%, 90% and 50% (86). This feature enabled researchers to explore potential homolog-mediated effects, thereby enhancing the detailed understanding of drug–protein interaction.

The thorough description of the cell lines and their affiliated information

The cell line page in MolBiC was thoroughly described in Figure 4, which meticulously showed the comprehensive data of each cell line. Taking HeLa as an exemplar, its general infor-

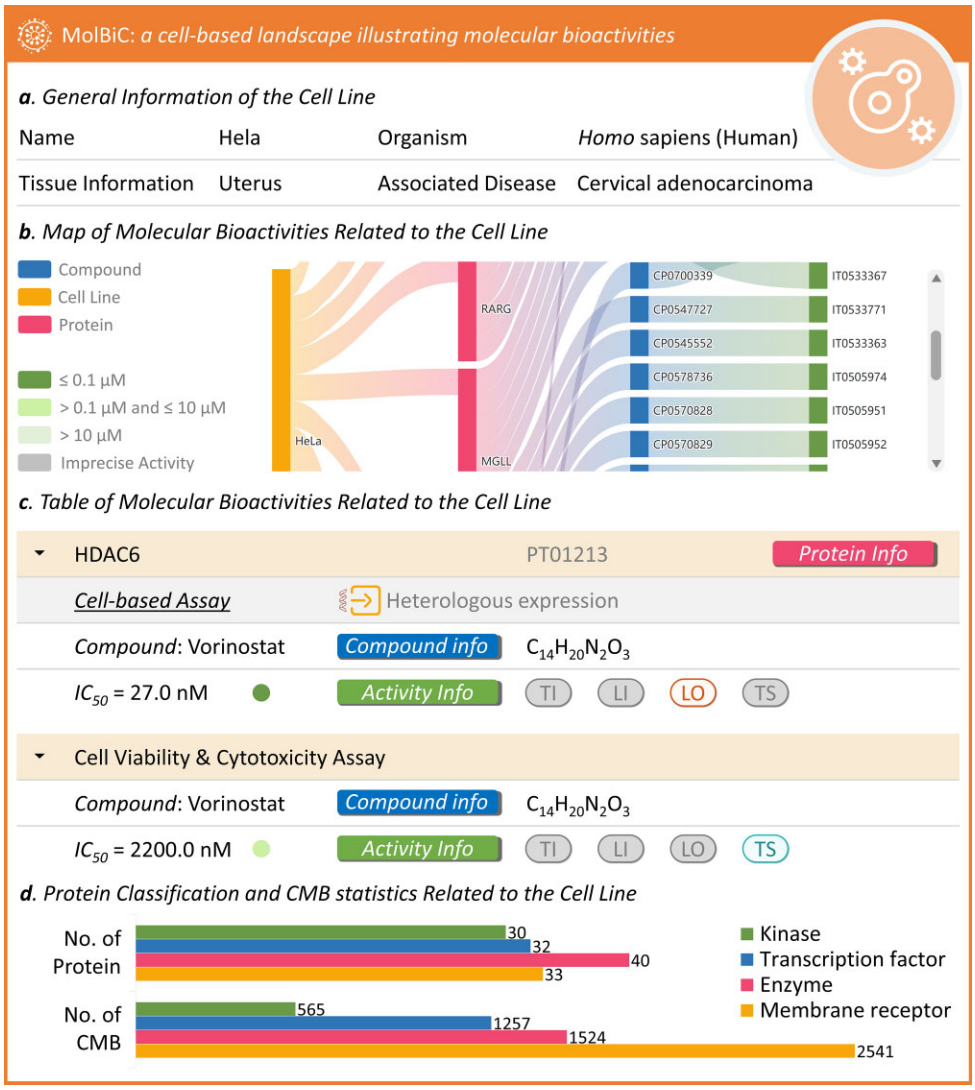


Figure 4. Descriptions and visualization of the cell lines in MolBiC. (***a***) The general information of the cell line (such as cell line name, species origin, tissue and disease). (***b***) The molecular bioactivities of the cell line including the compound-target interactions within the studied cellular context. (***c***) Detailed molecular bioactivity (such as cellular compound-protein binding assay and cytotoxicity together with cell viability information). (***d***) The target classification and activity statistics related to the studied cell line (such as target functional family and the number of bioactivities).

mation and associated CMB visualization were depicted in Figure 4a and b, respectively. Notably, as highlighted in Figure 4c, *vorinostat* demonstrated differential potencies, showing potent inhibitory activity against HeLa cells expressing HDAC6 (IC_{50} = 27.0 nM), while maintaining an acceptable cytotoxicity profile towards the HeLa cell line itself (IC_{50} = 2200.0 nM) (87–89). Moreover, as shown in Figure 4d, protein classification of HeLa expression system revealed a preponderance of membrane receptor, which exhibited the diversity of CMBs.

Cell line page was comprehensively constructed to offer extensive cellular and CMB data among a diverse array of cell lines. Each cell line was developed to depict general information (such as cell line's name, synonyms, categorization, species origin, related diseases and external database cross-references. CMB visualization was implemented using *Sankey diagram* and *tabular formats*, facilitating intuitive data interpretation. The page incorporated statistical analyses of protein classes and their associated CMBs, enabling users to elucidate the distribution of protein targets within cellular context. To facilitate tissue-specific study, the platform provided a curated compendium of cell lines from similar tissues, thereby enhancing the understanding of drug-protein interaction in biological system. All in all, the online page was capable of elucidating tissue-specific profiles and identifying the optimal cellular systems that facilitated the target-based drug discovery.

Conclusion and perspectives

Distinct from the traditional ligand-protein interaction profiles observed between *in vitro* purified proteins/cell lysates, CMB has emerged to be critical in drug discovery and development pipeline. In this study, we introduce a novel database entitled 'MolBiC', designed to provide detailed biological activity data within a cellular context. Given the significance of these newly collected data, our MolBiC is expected to attract broad interests from the relevant research communities and boost advances for diverse cutting-edge directions in modern drug discovery.

Data availability

All CMBs along with associated compound, protein and cell line could be viewed, accessed and downloaded from MolBiC, which could be freely accessible at: https://idrblab.org/MolBiC/.

Funding

National Natural Science Foundations of China [82373790, 22220102001, 81872798, U1909208]; Natural Science Foundation of Zhejiang Province [LR21H300001]; Fundamental Research Funds for Central Universities [2018QNA7023]; National Key R&D Program of China [2022YFC3400501]; Double Top-Class University [181201*194232101]; Key R&D Program of Zhejiang [2020C03010]; Westlake Laboratory (Westlake Lab of Life Sciences and Biomedicine); Leading Talent of the 'Ten Thousand Plan' National High-Level Talents Special Support Plan of China; Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare; Alibaba Cloud; The Information Technology Center of Zhejiang University. Funding for open access charge: Natural Science Foundation of Zhejiang Province [LR21H300001].

Conflict of interest statement

None declared.

References

1. Cornwell,J.A., Crncec,A., Afifi,M.M., Tang,K., Amin,R. and Cappell,S.D. (2023) Loss of CDK4/6 activity in S/G2 phase leads to cell cycle reversal. *Nature*, **619**, 363–370.
2. Baker,Y.R., Thorpe,C., Chen,J., Poller,L.M., Cox,L., Kumar,P., Lim,W.F., Lie,L., McClorey,G., Epple,S., *et al.* (2022) An LNA-amide modification that enhances the cell uptake and activity of phosphorothioate exon-skipping oligonucleotides. *Nat. Commun.*, **13**, 4036.
3. Barkley,D., Moncada,R., Pour,M., Liberman,D.A., Dryg,I., Werba,G., Wang,W., Baron,M., Rao,A., Xia,B., *et al.* (2022) Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat. Genet.*, **54**, 1192–1201.
4. Siriwongsup,S., Schmoker,A.M., Ficarro,S.B., Marto,J.A. and Kim,J. (2024) Bioorthogonally activated reactive species for target identification. *Chem.*, **10**, 1306–1315.
5. Balikci,E., Marques,A.M.C., Hansen,J.S. and Huber,K.V.M. (2023) Open resources for chemical probes and their implications for future drug discovery. *Expert Opin. Drug Discov.*, **18**, 505–513.
6. Rufer,A.C. (2021) Drug discovery for enzymes. *Drug Discov. Today*, **26**, 875–886.
7. Emmerich,C.H., Gamboa,L.M., Hofmann,M.C.J., Bonin-Andresen,M., Arbach,O., Schendel,P., Gerlach,B., Hempel,K., Bespalov,A., Dirnagl,U., *et al.* (2021) Improving target assessment in biomedical research: the GOT-IT recommendations. *Nat. Rev. Drug Discov.*, **20**, 64–81.
8. Balikci,E., Marques,A.M.C., Bauer,L.G., Seupel,R., Bennett,J., Raux,B., Buchan,K., Simelis,K., Singh,U., Rogers,C., *et al.* (2024) Unexpected noncovalent off-target activity of clinical BTK inhibitors leads to discovery of a dual NUDT5/14 antagonist. *J. Med. Chem.*, **67**, 7245–7259.
9. Payne,N.C., Kalyakina,A.S., Singh,K., Tye,M.A. and Mazitschek,R. (2021) Bright and stable luminescent probes for target engagement profiling in live cells. *Nat. Chem. Biol.*, **17**, 1168–1177.
10. Lu,M.K., Yin,J.Y., Zhu,Q., Lin,G.L., Mou,M.J., Liu,F.Y., Pan,Z.Q., You,N.X., Lian,X.C., Li,F.C., *et al.* (2023) Artificial intelligence in pharmaceutical sciences. *Engineering*, **27**, 37–69.
11. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B., *et al.* (2023) PubChem 2023 update. *Nucleic Acids Res.*, **51**, D1373–D1380.
12. Zhang,C., Zhang,X., Freddolino,P.L. and Zhang,Y. (2024) BioLiP2: an updated structure database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.*, **52**, D404–D412.
13. Su,M., Yang,Q., Du,Y., Feng,G., Liu,Z., Li,Y. and Wang,R. (2019) Comparative assessment of scoring functions: the CASF-2016 update. *J. Chem. Inf. Model.*, **59**, 895–913.
14. Zdrazil,B., Felix,E., Hunter,F., Manners,E.J., Blackshaw,J., Corbett,S., de Veij,M., Ioannidis,H., Lopez,D.M., Mosquera,J.F., *et al.* (2024) The ChEMBL database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.*, **52**, D1180–D1192.
15. Stein,R.M., Yang,Y., Balius,T.E., O'Meara,M.J., Lyu,J., Young,J., Tang,K., Shoichet,B.K. and Irwin,J.J. (2021) Property-unmatched decoys in docking benchmarks. *J. Chem. Inf. Model.*, **61**, 699–714.
16. Gilson,M.K., Liu,T., Baitaluk,M., Nicola,G., Hwang,L. and Chong,J. (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.*, **44**, D1045–D1053.
17. Zhou,Y., Zhang,Y., Zhao,D., Yu,X., Shen,X., Zhou,Y., Wang,S., Qiu,Y., Chen,Y. and Zhu,F. (2024) TTD: therapeutic Target Database describing target druggability information. *Nucleic Acids Res.*, **52**, D1465–D1477.
18. Knox,C., Wilson,M., Klinger,C.M., Franklin,M., Oler,E., Wilson,A., Pon,A., Cox,J., Chin,N.E.L., Strawbridge,S.A., *et al.* (2024) DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Res.*, **52**, D1265–D1275.
19. Kanev,G.K., de Graaf,C., Westerman,B.A., de Esch,I.J.P. and Kooistra,A.J. (2021) KLIFS: an overhaul after the first 5 years of supporting kinase research. *Nucleic Acids Res.*, **49**, D562–D569.
20. Xie,X., Yu,T., Li,X., Zhang,N., Foster,L.J., Peng,C., Huang,W. and He,G. (2023) Recent advances in targeting the "undruggable" proteins: from drug discovery to clinical trials. *Signal Transduct. Target. Ther.*, **8**, 335.
21. Li,Y.H., Li,X.X., Hong,J.J., Wang,Y.X., Fu,J.B., Yang,H., Yu,C.Y., Li,F.C., Hu,J., Xue,W.W., *et al.* (2020) Clinical trials, progression speed differentiating features and swiftness rule of the innovative targets of first in class drugs. *Brief. Bioinf.*, **21**, 649–662.
22. Xue,W., Yang,F., Wang,P., Zheng,G., Chen,Y., Yao,X. and Zhu,F. (2018) What contributes to serotonin-norepinephrine reuptake inhibitors' dual-targeting mechanism? The key role of transmembrane domain 6 in human serotonin and norepinephrine transporters revealed by molecular dynamics simulation. *ACS Chem. Neurosci.*, **9**, 1128–1140.
23. Bui-Thi,D., Riviere,E., Meysman,P. and Laukens,K. (2022) Predicting compound-protein interaction using hierarchical graph convolutional networks. *PLoS One*, **17**, e0258628.
24. Chen,L.F., Fan,Z.S., Chang,J., Yang,R.R., Hou,H., Guo,H., Zhang,Y.H., Yang,T.B., Zhou,C.M., Sui,Q.B., *et al.* (2023) Sequence-based drug design as a concept in computational drug design. *Nat. Commun.*, **14**, 4217.
25. Bairoch,A. (2018) Cellosaurus, a cell line knowledge resource. *J. Biomol. Tech.*, **29**, 25–38.
26. Wishart,D.S., Feunang,Y.D., Guo,A.C., Lo,E.J., Marcu,A., Grant,J.R., Sajed,T., Johnson,D., Li,C., Sayeeda,Z., *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
27. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B., *et al.* (2021) PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.*, **49**, D1388–D1395.
28. UniProt, C. (2023) UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.
29. Wang,Y., Zhang,S., Li,F., Zhou,Y., Zhang,Y., Wang,Z., Zhang,R., Zhu,J., Ren,Y., Tan,Y., *et al.* (2020) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.*, **48**, D1031–D1041.

30. Li,F., Yin,J., Lu,M., Mou,M., Li,Z., Zeng,Z., Tan,Y., Wang,S., Chu,X., Dai,H., *et al.* (2023) DrugMAP: molecular atlas and pharma-information of all drugs. *Nucleic Acids Res.*, **51**, D1288–D1299.
31. Stefaniak,J. and Huber,K.V.M. (2020) Importance of quantifying drug-target engagement in cells. *ACS Med. Chem. Lett.*, **11**, 403–406.
32. Zhu,F., Shi,Z., Qin,C., Tao,L., Liu,X., Xu,F., Zhang,L., Song,Y., Liu,X.H., Zhang,J.X., *et al.* (2012) Therapeutic target database update 2012: a resource for facilitating target oriented drug discovery. *Nucleic Acids Res.*, **40**, D1128–D1136.
33. Robers,M.B., Friedman-Ohana,R., Huber,K.V.M., Kilpatrick,L., Vasta,J.D., Berger,B.T., Chaudhry,C., Hill,S., Muller,S., Knapp,S., *et al.* (2020) Quantifying target occupancy of small molecules within living cells. *Annu. Rev. Biochem.*, **89**, 557–581.
34. Gene Ontology, C., Aleksander,S.A., Balhoff,J., Carbon,S., Cherry,J.M., Drabkin,H.J., Ebert,D., Feuermann,M., Gaudet,P., Harris,N.L., *et al.* (2023) The gene ontology knowledgebase in 2023. *Genetics*, **224**, iyad031.
35. Yin,J., Li,F., Zhou,Y., Mou,M., Lu,Y., Chen,K., Xue,J., Luo,Y., Fu,J., He,X., *et al.* (2021) INTEDE: interactome of drug-metabolizing enzymes. *Nucleic Acids Res.*, **49**, D1233–D1243.
36. Tang,J., Fu,J., Wang,Y., Li,B., Li,Y., Yang,Q., Cui,X., Hong,J., Li,X., Chen,Y., *et al.* (2020) ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Briefings Bioinf.*, **21**, 621–636.
37. Blay,V., Tolani,B., Ho,S.P. and Arkin,M.R. (2020) High throughput screening: today's biochemical and cell-based approaches. *Drug Discovery Today*, **25**, 1807–1821.
38. Yang,Q., Wang,Y., Zhang,Y., Li,F., Xia,W., Zhou,Y., Qiu,Y., Li,H. and Zhu,F. (2020) NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data. *Nucleic Acids Res.*, **48**, W436–W448.
39. Yang,Q., Li,B., Chen,S., Tang,J., Li,Y., Li,Y., Zhang,S., Shi,C., Zhang,Y., Mou,M., *et al.* (2021) MMEASE: online meta-analysis of metabolomic data by enhanced metabolite annotation, marker selection and enrichment analysis. *J. Proteomics*, **232**, 104023.
40. Luan,Y., Li,J., Bernatchez,J.A. and Li,R. (2019) Kinase and histone deacetylase hybrid inhibitors for cancer therapy. *J. Med. Chem.*, **62**, 3171–3183.
41. Hsu,K.C., Chu,J.C., Tseng,H.J., Liu,C.I., Wang,H.C., Lin,T.E., Lee,H.S., Hsin,L.W., Wang,A.H., Lin,C.H., *et al.* (2021) Synthesis and biological evaluation of phenothiazine derivative-containing hydroxamic acids as potent class II histone deacetylase inhibitors. *Eur. J. Med. Chem.*, **219**, 113419.
42. Jin,M.Z. and Jin,W.L. (2020) The updated landscape of tumor microenvironment and drug repurposing. *Signal Transduct. Target. Ther.*, **5**, 1478–1493.
43. Kao,K.C., Vilbois,S., Tsai,C.H. and Ho,P.C. (2022) Metabolic communication in the tumour-immune microenvironment. *Nat. Cell Biol.*, **24**, 1574–1583.
44. Yin,J., Sun,W., Li,F., Hong,J., Li,X., Zhou,Y., Lu,Y., Liu,M., Zhang,X., Chen,N., *et al.* (2020) VARIDT 1.0: variability of drug transporter database. *Nucleic Acids Res.*, **48**, D1042–D1050.
45. Yuan,J., Najafov,A. and Py,B.F. (2016) Roles of caspases in necrotic cell death. *Cell*, **167**, 1693–1704.
46. Zhang,H.Z., Kasibhatla,S., Kuemmerle,J., Kemnitzer,W., Ollis-Mason,K., Qiu,L., Crogan-Grundy,C., Tseng,B., Drewe,J. and Cai,S.X. (2005) Discovery and structure activity relationship of 3-aryl-5-aryl-1,2,4-oxadiazoles as a new series of apoptosis inducers and potential anticancer agents. *J. Med. Chem.*, **48**, 5215–5223.
47. Zhang,Y., Zhou,Y., Zhou,Y., Yu,X., Shen,X., Hong,Y., Zhang,Y., Wang,S., Mou,M., Zhang,J., *et al.* (2024) TheMarker: a comprehensive database of therapeutic biomarkers. *Nucleic Acids Res.*, **52**, D1450–D1464.
48. Yin,J., Chen,Z., You,N., Li,F., Zhang,H., Xue,J., Ma,H., Zhao,Q., Yu,L., Zeng,S., *et al.* (2024) VARIDT 3.0: the phenotypic and regulatory variability of drug transporter. *Nucleic Acids Res.*, **52**, D1490–D1502.
49. Vu,V., Szewczyk,M.M., Nie,D.Y., Arrowsmith,C.H. and Barsyte-Lovejoy,D. (2022) Validating small molecule chemical probes for biological discovery. *Annu. Rev. Biochem.*, **91**, 61–87.
50. Mateus,A., Kurzawa,N., Perrin,J., Bergamini,G. and Savitski,M.M. (2022) Drug target identification in tissues by thermal proteome profiling. *Annu. Rev. Pharmacol. Toxicol.*, **62**, 465–482.
51. Cai,H., Liao,Y., Zhu,L., Wang,Z. and Song,J. (2024) Improving cancer survival prediction via graph convolutional neural network learning on protein-protein interaction networks. *IEEE J. Biomed. Health Inform.*, **28**, 1134–1143.
52. Yang,Q., Li,B., Tang,J., Cui,X., Wang,Y., Li,X., Hu,J., Chen,Y., Xue,W., Lou,Y., *et al.* (2020) Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data. *Brief. Bioinf.*, **21**, 1058–1068.
53. Zhang,S., Amahong,K., Sun,X., Lian,X., Liu,J., Sun,H., Lou,Y., Zhu,F. and Qiu,Y. (2021) The miRNA: a small but powerful RNA for COVID-19. *Brief. Bioinf.*, **22**, 1137–1149.
54. Zhang,S., Amahong,K., Zhang,C., Li,F., Gao,J., Qiu,Y. and Zhu,F. (2021) RNA-RNA interactions between SARS-CoV-2 and host benefit viral development and evolution during COVID-19 infection. *Brief. Bioinf.*, **23**, bbab397.
55. Ostacolo,C., Di Sarno,V., Lauro,G., Pepe,G., Musella,S., Ciaglia,T., Vestuto,V., Autore,G., Bifulco,G., Marzocco,S., *et al.* (2019) Identification of an indol-based multi-target kinase inhibitor through phenotype screening and target fishing using inverse virtual screening approach. *Eur. J. Med. Chem.*, **167**, 61–75.
56. McGibbon,M., Shave,S., Dong,J., Gao,Y., Houston,D.R., Xie,J., Yang,Y., Schwaller,P. and Blay,V. (2023) From intuition to AI: evolution of small molecule representations in drug discovery. *Brief. Bioinf.*, **25**, bbad422.
57. McNair,D. (2023) Artificial intelligence and machine learning for lead to candidate decision-making and beyond. *Annu. Rev. Pharmacol. Toxicol.*, **63**, 77–97.
58. Xue,W., Fu,T., Deng,S., Yang,F., Yang,J. and Zhu,F. (2022) Molecular mechanism for the allosteric inhibition of the human serotonin transporter by antidepressant escitalopram. *ACS Chem. Neurosci.*, **13**, 340–351.
59. Lian,X., Zhang,Y., Zhou,Y., Sun,X., Huang,S., Dai,H., Han,L. and Zhu,F. (2024) SingPro: a knowledge base providing single-cell proteomic data. *Nucleic Acids Res.*, **52**, D552–D561.
60. Zhang,Y., Sun,H., Lian,X., Tang,J. and Zhu,F. (2023) ANPELA: significantly enhanced quantification tool for cytometry-based single-cell proteomics. *Adv. Sci.*, **10**, e2207061.
61. Collins,R., Lee,H., Jones,D.H., Elkins,J.M., Gillespie,J.A., Thomas,C., Baldwin,A.G., Jones,K., Waters,L., Paine,M., *et al.* (2022) Comparative analysis of small molecule LIMK1/2 inhibitors: chemical synthesis, biochemistry, and cellular activity. *J. Med. Chem.*, **65**, 13705–13713.
62. Kenakin,T. (2024) Know your molecule: pharmacological characterization of drug candidates to enhance efficacy and reduce late-stage attrition. *Nat. Rev. Drug Discov.*, **23**, 626–644.
63. Subbaiah,M.A.M., Rautio,J. and Meanwell,N.A. (2024) Prodrugs as empowering tools in drug discovery and development: recent strategic applications of drug delivery solutions to mitigate challenges associated with lead compounds and drug candidates. *Chem. Soc. Rev.*, **53**, 2099–2210.
64. Stefaniak,J., Lewis,A.M., Conole,D., Galan,S.R.G., Bataille,C.J.R., Wynne,G.M., Castaldi,M.P., Lundback,T., Russell,A.J. and Huber,K.V.M. (2018) Chemical instability and promiscuity of arylmethylidenepyrazolinone based MDMX inhibitors. *ACS Chem. Biol.*, **13**, 2849–2854.
65. Musella,S., D'Avino,D., Peltner,L.K., Di Sarno,V., Cerqua,I., Merciai,F., Vestuto,V., Ciaglia,T., Smaldone,G., Di Matteo,F., *et al.* (2023) Design, synthesis, and pharmacological characterization of a potent soluble epoxide hydrolase inhibitor for the treatment of acute pancreatitis. *J. Med. Chem.*, **66**, 9201–9222.

66. Shim,S., Krishnaiah,M., Sankham,M.R., Kim,I., Lee,Y., Shin,I., Oh,A.R., Lee,H.J., Vu,T.N.L., Park,J., *et al.* (2022) Discovery of (E)-3-(3-((2-Cyano-4'-dimethylaminobiphenyl-4-ylmethyl)cyclohexanecarbonylamino)-5-fluorophenyl)acrylic Acid Methyl Ester, an intestine specific, FXR partial agonist for the treatment of nonalcoholic steatohepatitis. *J. Med. Chem.*, **65**, 9974–10000.
67. Zhang,Y., Liu,X., Li,F., Yin,J., Yang,H., Li,X., Liu,X., Chai,X., Niu,T., Zeng,S., *et al.* (2024) INTEDE 2.0: the metabolic roadmap of drugs. *Nucleic Acids Res.*, **52**, D1355–D1364.
68. Sun,X., Zhang,Y., Li,H., Zhou,Y., Shi,S., Chen,Z., He,X., Zhang,H., Li,F., Yin,J., *et al.* (2023) DRESIS: the first comprehensive landscape of drug resistance information. *Nucleic Acids Res.*, **51**, D1263–D1275.
69. Ge,Y.C., Zhou,C.Z., Ma,Y.H., Wang,Z.H., Wang,S.F., Wang,W. and Wu,B. (2024) Advancing natural product discovery: a structure-oriented fractions screening platform for compound annotation and isolation. *Anal. Chem.*, **96**, 5399–5406.
70. Zheng,L., Shi,S., Lu,M., Fang,P., Pan,Z., Zhang,H., Zhou,Z., Zhang,H., Mou,M., Huang,S., *et al.* (2024) AnnoPRO: a strategy for protein function annotation based on multi-scale protein representation and a hybrid deep learning of dual-path encoding. *Genome Biol.*, **25**, 41.
71. Mou,M., Pan,Z., Zhou,Z., Zheng,L., Zhang,H., Shi,S., Li,F., Sun,X. and Zhu,F. (2023) A transformer based ensemble framework for the prediction of protein-protein interaction sites. *Research*, **6**, 240.
72. Li,F., Yin,J., Lu,M., Yang,Q., Zeng,Z., Zhang,B., Li,Z., Qiu,Y., Dai,H., Chen,Y., *et al.* (2022) ConSIG: consistent discovery of molecular signature from OMIC data. *Brief. Bioinf.*, **23**, bbac253.
73. Preti,B., Suchankova,A., Deganutti,G., Leuenberger,M., Barkan,K., Manulak,I., Huang,X., Carvalho,S., Ladds,G. and Lochner,M. (2022) Discovery and structure–activity relationship studies of novel adenosine A(1) receptor-selective agonists. *J. Med. Chem.*, **65**, 14864–14890.
74. Buyanova,M. and Pei,D. (2022) Targeting intracellular protein-protein interactions with macrocyclic peptides. *Trends Pharmacol. Sci.*, **43**, 234–248.
75. Li,F., Zhou,Y., Zhang,Y., Yin,J., Qiu,Y., Gao,J. and Zhu,F. (2022) POSREG: proteomic signature discovered by simultaneously optimizing its reproducibility and generalizability. *Brief. Bioinf.*, **23**, bbac040.
76. Fu,J., Zhang,Y., Wang,Y., Zhang,H., Liu,J., Tang,J., Yang,Q., Sun,H., Qiu,W., Ma,Y., *et al.* (2022) Optimization of metabolomic data processing using NOREVA. *Nat. Protoc.*, **17**, 129–151.
77. Comez,D., Glenn,J., Anbuhl,S.M., Heukers,R., Smit,M.J., Hill,S.J. and Kilpatrick,L.E. (2022) Fluorescently tagged nanobodies and NanoBRET to study ligand binding and agonist-induced conformational changes of full length EGFR expressed in living cells. *Front. Immunol.*, **13**, 1006718.
78. Kilpatrick,L.E., Alcobia,D.C., White,C.W., Peach,C.J., Glenn,J.R., Zimmerman,K., Kondrashov,A., Pfleger,K.D.G., Ohana,R.F., Robers,M.B., *et al.* (2019) Complex formation between VEGFR2 and the beta(2)-adrenoceptor. *Cell Chem Biol*, **26**, 830–841.
79. Li,B., Tang,J., Yang,Q., Li,S., Cui,X., Li,Y., Chen,Y., Xue,W., Li,X. and Zhu,F. (2017) NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.*, **45**, W162–W170.
80. Chowdhury,S., Zielinski,D.C., Dalldorf,C., Rodrigues,J.V., Palsson,B.O. and Shakhnovich,E.I. (2023) Empowering drug off target discovery with metabolic and structural analysis. *Nat. Commun.*, **14**, 3390.
81. Sadri,A. (2023) Is target based drug discovery efficient? Discovery and "off target" mechanisms of all drugs. *J. Med. Chem.*, **66**, 12651–12677.
82. Rabal,O., Sanchez-Arias,J.A., Cuadrado-Tejedor,M., de Miguel,I., Perez-Gonzalez,M., Garcia-Barroso,C., Ugarte,A., Estella-Hermoso de Mendoza,A., Saez,E., Espelosin,M., *et al.* (2018) Design, synthesis, biological evaluation and in vivo testing of dual phosphodiesterase 5 (PDE5) and histone deacetylase 6 (HDAC6) selective inhibitors for the treatment of Alzheimer's disease. *Eur. J. Med. Chem.*, **150**, 506–524.
83. Beckers,T., Mahboobi,S., Sellmer,A., Winkler,M., Eichhorn,E., Pongratz,H., Maier,T., Ciossek,T., Baer,T., Kelter,G., *et al.* (2012) Chimerically designed HDAC and tyrosine kinase inhibitors. A series of erlotinib hybrids as dual selective inhibitors of EGFR, HER2 and histone deacetylases. *MedChemComm*, **3**, 829–835.
84. Wang,H., Lim,Z.Y., Zhou,Y., Ng,M., Lu,T., Lee,K., Sangthongpitag,K., Goh,K.C., Wang,X., Wu,X., *et al.* (2010) Acylurea connected straight chain hydroxamates as novel histone deacetylase inhibitors: synthesis, SAR, and in vivo antitumor activity. *Bioorg. Med. Chem. Lett.*, **20**, 3314–3321.
85. Chen,X., Zhao,S., Li,H., Wang,X., Geng,A., Cui,H., Lu,T., Chen,Y. and Zhu,Y. (2019) Design, synthesis and biological evaluation of novel isoindolinone derivatives as potent histone deacetylase inhibitors. *Eur. J. Med. Chem.*, **168**, 110–122.
86. Suzek,B.E., Wang,Y.Q., Huang,H.Z., McGarvey,P.B., Wu,C.H. and UniProt Consortium (2015) UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, **31**, 926–932.
87. Mahboobi,S., Dove,S., Sellmer,A., Winkler,M., Eichhorn,E., Pongratz,H., Ciossek,T., Baer,T., Maier,T. and Beckers,T. (2009) Design of chimeric histone deacetylase and tyrosine kinase inhibitors: a series of imatinib hybrides as potent inhibitors of wild type and mutant BCR-ABL, PDGF-rbeta, and histone deacetylases. *J. Med. Chem.*, **52**, 2265–2279.
88. Guerrant,W., Patil,V., Canzoneri,J.C., Yao,L.P., Hood,R. and Oyelere,A.K. (2013) Dual acting histone deacetylase topoisomerase I inhibitors. *Bioorg. Med. Chem. Lett.*, **23**, 3283–3287.
89. Li,X., Wang,C., Chai,X., Liu,X., Qiao,K., Fu,Y., Jin,Y., Jia,Q., Zhu,F. and Zhang,Y. (2024) Discovery of potent selective HDAC6 inhibitors with 5-phenyl-1H-indole fragment: virtual screening, rational design, and biological evaluation. *J. Chem. Inf. Model.*, **64**, 6147–6161.
90. Camacho,C., Boratyn,G.M., Joukov,V., Vera Alvarez,R. and Madden,T.L. (2023) ElasticBLAST: accelerating sequence search via cloud computing. *BMC Bioinf.*, **24**, 117.
91. Oprea,T.I., Bologa,C.G., Brunak,S., Campbell,A., Gan,G.N., Gaulton,A., Gomez,S.M., Guha,R., Hersey,A., Holmes,J., *et al.* (2018) Unexplored therapeutic opportunities in the human genome. *Nat. Rev. Drug Discov.*, **17**, 317–332.

Received: August 15, 2024. **Revised:** September 13, 2024. **Editorial Decision:** September 19, 2024. **Accepted:** September 20, 2024