Molecular Therapy
Opinion

A systematic approach to identify host targets and rapidly deliver broad-spectrum antivirals

Julien Olivet,[1,2] Sibusiso B. Maseko,[2] Alexander N. Volkov,[3,4]
Kourosh Salehi-Ashtiani,[5] Kalyan Das,[1] Michael A. Calderwood,[6,7,8]
Jean-Claude Twizere,[2,5,9] and Christoph Gorgulla[8,10,11]

https://doi.org/10.1016/j.ymthe.2022.02.015

Early vaccine development for coronavirus disease 2019 (COVID-19) was possible thanks to the prior knowledge that the main immunogenic protein of coronaviruses is the spike protein. Indeed, once the spike sequence of the initial severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variant was determined, vaccines could be developed quickly. In parallel, if virus-host protein-protein interactions (PPIs) are mapped for newly emerging or re-emerging pathogens, key interacting domains can be predicted, and validation of such interactions can help deliver small-molecule inhibitors in a relatively short period of time. Here, we propose a comprehensive pipeline to systematically identify broad-spectrum antivirals in a time-effective manner and complement traditional therapeutic strategies to fight viral infections. Following initial sequencing and virus-host interactome mapping, ultra-large virtual drug screening is conducted on the predicted interacting domains to quickly identify primary hits. The virtual hits and analogs are then experimentally verified using biophysical and biological assays, as well as structural biology techniques. This widely applicable pipeline should allow faster identification of broad-spectrum hits and leads for antiviral drug discovery.

Viruses are obligate intracellular parasites that commonly hijack cellular pathways and host protein functions to infect cells and replicate. The basic steps of the viral life cycle include attachment, entry, replication, translation, suppression of the host immune response, assembly, maturation, and particles release. Many of these steps are orchestrated via interactions with host cellular proteins. Viruses thus manipulate host cells by perturbing key macromolecules involved in the control of cell homeostasis. Despite more than 60 years of virology research, the arsenal of antiviral drugs remains alarmingly small, with only about 100 molecules currently available in clinics and most targeting enzymes from five different pathogens: HIV-1, hepatitis C virus (HCV), HBV, herpes simplex virus (HSV), and influenza.[1] Thus, efficient therapeutics are still lacking for numerous viral diseases, and the effectiveness of existing antiviral therapies is often hampered by the occurrence of resistance to treatments. However, virus-host PPIs are, in principle, valuable targets to complement viral enzymes, as their disruption should synergize the antiviral potential of enzymatic inhibitors and minimize the occurrence of drug-resistant mutants, as previously described.[2]

Viral proteins have been directly shaped through interactions with host proteomes, which has led to the development of crucial virus-host PPIs. Consequently, disruption of such PPIs by targeting the interfaces between interacting proteins provides a strong basis for the rapid development of potent antiviral drugs. Arguments to reconsider PPIs as druggable targets have been extensively reviewed elsewhere.[3,4] In summary, the ruggedness of interaction interfaces, which often display grooves and pockets; the discovery of hot spots, i.e., amino acid residues contributing most to the binding of two protein partners;[5] the development of high-performance, quantitative interaction assays;[6–8] and the design of innovative chemical libraries,[9,10] including macrocyclics[11] or peptide secondary structure mimetics,[12] have significantly boosted drug discovery efforts for this promising class of targets. One of the best-known examples of small molecules targeting virus-host PPIs is Maraviroc, which inhibits the interaction between the human CCR5 chemokine receptor and the HIV gp120 envelope.[13] This compound was approved by the US Food and Drug Administration (FDA) in 2007, but the underlying mechanisms allowing structure-based optimizations were only deciphered in 2013.[14] We argue that a systematic and unbiased pipeline, involving high-throughput virus-host protein interactome mapping, protein structure predictions, and ultra-large virtual screenings of millions of compounds (Figure 1), would allow the rapid identification of

---

[1]Structural Biology Unit, Laboratory of Virology and Chemotherapy, Rega Institute for Medical Research and Department of Microbiology, Immunology and Transplantation, Catholic University of Leuven (KU Leuven), Leuven, Belgium; [2]Laboratory of Viral Interactomes Networks, Unit of Molecular Biology of Diseases, Interdisciplinary Cluster for Applied Genoproteomics (GIGA Institute), University of Liège, Liège, Belgium; [3]VIB-VUB Center for Structural Biology, Flemish Institute of Biotechnology (VIB), Brussels, Belgium; [4]Jean Jeener NMR Centre, Free University of Brussels (VUB), Brussels, Belgium; [5]Division of Science and Math, New York University Abu Dhabi, Abu Dhabi, UAE; [6]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute (DFCI), Boston, MA, USA; [7]Department of Genetics, Blavatnik Institute, Harvard Medical School (HMS), Boston, MA, USA; [8]Department of Cancer Biology, Dana-Farber Cancer Institute (DFCI), Boston, MA, USA; [9]TERRA Research and Teaching Centre, Microbial Processes and Interactions (MiPI), Gembloux Agro Bio-tech, University of Liège, Gembloux, Belgium; [10]Department of Biological Chemistry and Molecular Pharmacology, Blavatnik Institute, Harvard Medical School (HMS), Boston, MA, USA; [11]Department of Physics, Faculty of Arts and Sciences, Harvard University, Cambridge, MA, USA

Correspondence: Jean-Claude Twizere, PhD, Laboratory of Viral Interactomes Networks, Unit of Molecular Biology of Diseases, Interdisciplinary Cluster for Applied Genoproteomics (GIGA Institute), University of Liège, Liège, Belgium.
E-mail: jean-claude.twizere@uliege.be
Correspondence: Christoph Gorgulla, PhD, Department of Cancer Biology, Dana-Farber Cancer Institute (DFCI), Boston, MA, USA.
E-mail: cgorgulla@g.harvard.edu

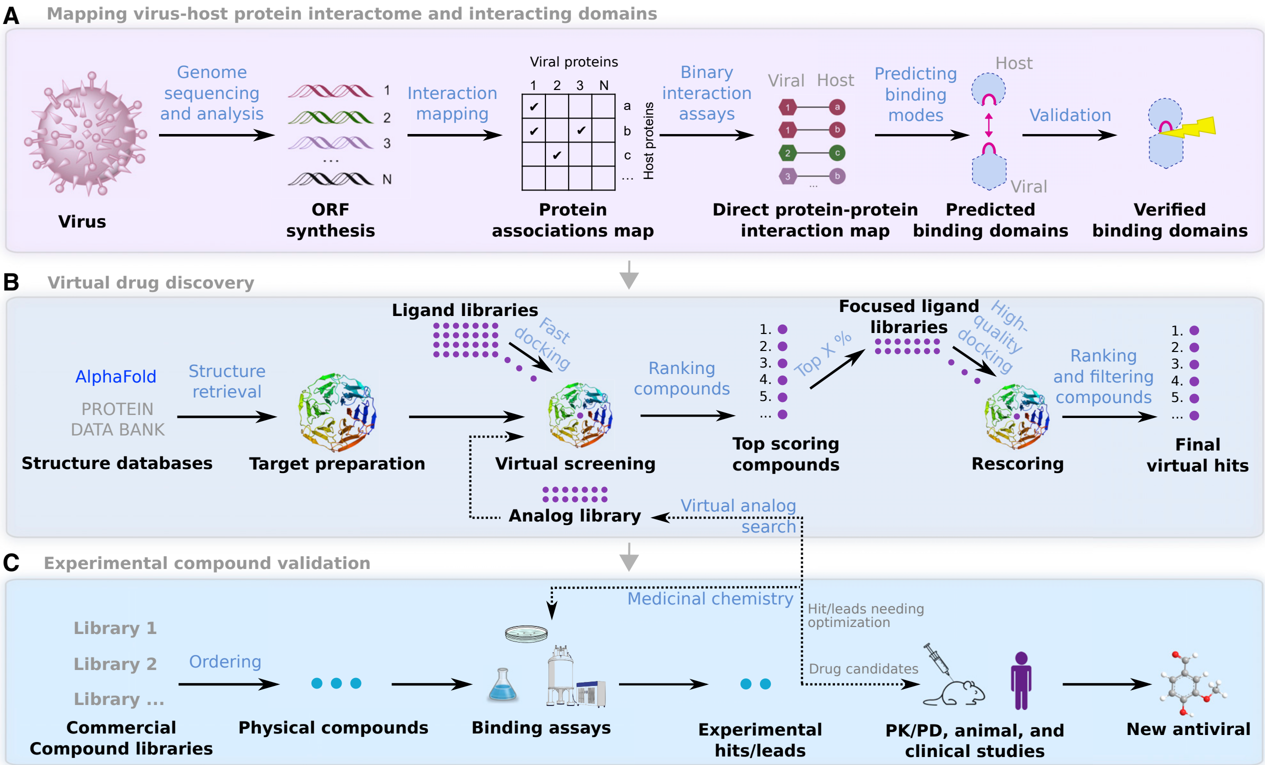
A Mapping virus-host protein interactome and interacting domains
Genome sequencing and analysis
Interaction mapping
Viral proteins
1 2 3 N
a b c ...
Host proteins
Binary interaction assays
Viral Host
Predicting binding modes
Host
Viral
Validation
Virus
ORF synthesis
Protein associations map
Direct protein-protein interaction map
Predicted binding domains
Verified binding domains
B Virtual drug discovery
AlphaFold
PROTEIN DATA BANK
Structure retrieval
Ligand libraries
Fast docking
Ranking compounds
Top X %
Focused ligand libraries
High-quality docking
Ranking and filtering compounds
Structure databases
Target preparation
Virtual screening
Top scoring compounds
Rescoring
Final virtual hits
Analog library
Virtual analog search
C Experimental compound validation
Library 1
Library 2
Library ...
Ordering
Medicinal chemistry
Hit/leads needing optimization
Drug candidates
Commercial Compound libraries
Physical compounds
Binding assays
Experimental hits/leads
PK/PD, animal, and clinical studies
New antiviral

Figure 1. General pipeline for the rapid and systematic identification of potent antiviral small molecules

selective virus-host PPI inhibitors to complement vaccines and other therapeutic tools.

Viral genome sequencing and virus-host protein interactome mapping

With recent advances in DNA sequencing, it has become possible to map and analyze viral genomes within hours or days. New or re-emerging viruses can thus be classified quickly and assigned to a family following phylogenetic studies. Using genomic information, open reading frames (ORFs) can also be rapidly identified, synthesized, and cloned into many expression systems to test viral proteins in the desired experimental conditions (Figure 1A). In particular, ORFs can be cloned into entry vectors to be rapidly shuttled into various destination plasmids using the high-throughput Gateway cloning technology.[15,16] This strategy has allowed systematic expression of proteins and mapping of PPIs in different assays and multiple environments.[7,8]

There are two main complementary approaches to identify protein interactions at the proteome scale: affinity purification followed by mass spectrometry (AP-MS) and binary interaction assays, such as yeast two-hybrid (Y2H).[15,16] AP-MS has been widely used to confidently identify the composition of distinct cellular complexes[17] and understand how viral proteins can disrupt them.[18] However, as recently described in Trepte et al.,[8] it remains challenging to differentiate direct interactions (i.e., proteins sharing a physical interface) from indirect associations (i.e., proteins that do not directly interact) within the different complexes identified by AP-MS. Structural biology techniques can provide such information and generate high-resolution 3D models of directly interacting proteins, but their modest throughput currently limits their applications for proteome-wide efforts. In this context, several complementary, quantitative binary interaction assays, such as Y2H, NanoLuc two-hybrid (N2H), or luminescence-based two-hybrid (LuTHy), can be applied to rapidly identify direct PPI targets.[6–8]

To systematically map direct virus-host PPIs with binary interaction assays, two strategies can be applied in parallel. First, all possible pairwise combinations between the viral proteins and the ∼20,000 human reference protein-coding genes can be mapped and validated, similar to what was done for SARS-CoV-2.[16,19] Second, AP-MS from different cell lines expressing tagged viral proteins can be performed and binary interaction assays applied between the relevant viral proteins and the members of the identified disrupted complexes. This will output a list of high-quality, direct virus-host PPI targets, from which the interacting domains can be predicted *in silico* using motif-domain and/or domain-domain complementarity approaches, as described in Maseko et al.[20] Indeed, many viral proteins contain short linear motifs (SLiM), which are predicted to bind known complementary host protein domains. Alternatively, structure-based protein-protein docking can be carried out using experimental structures available in the Protein Data Bank or predicted structures in the AlphaFold Protein Structure Database.[21] The best candidates for potentially interacting domains can then be systematically tested and the corresponding interactions confirmed experimentally using several *in vitro* (e.g., fluorescence polarization, isothermal titration calorimetry [ITC], etc.) and/or *in cellulo* binary interaction assays (e.g., N2H, LuTHy, Y2H, etc.).[7] Finally, the validated viral and host protein domains can be used to conduct ultra-large virtual drug screenings and potentially identify small-molecule inhibitors of key virus-host PPIs.

Virtual drug screenings to identify small-molecule binders

Ultra-large virtual screening using structure-based molecular dockings has several key advantages over experimental high-throughput approaches, as highlighted by recent successes for protein structures with resolutions above 3Å.[22–25] First, it can predict potent compounds with dissociation constants (K_D) often lying in the lower nanomolar range. This is a direct consequence of sampling a vast chemical space, which allows the identification of exceptionally well-fitting molecules. This aspect leads to the second advantage: the ability to find sufficiently strong binders to targets traditionally considered challenging, such as PPI interfaces and allosteric sites. The third major advantage is that a large-scale screen improves the true hit rate, i.e., the number of active compounds divided by the number of experimentally tested compounds, since the error robustness increases when scaling up.[26] Furthermore, ultra-large virtual screenings can be carried out in days and thus save a substantial amount of time and costs compared with traditional high-throughput screens.

Published in 2020, VirtualFlow is the first freely accessible, ultra-large virtual screening platform able to routinely sample billions of commercially available compounds.[26] VirtualFlow also provides ultra-large ligand libraries in a ready-to-dock format. These include 1.48 billion compounds from the ZINC library and ~1.4 billion molecules from the REAL library of Enamine.[26] Platforms such as VirtualFlow can thus be used to rapidly find small-molecule binders to the sites identified in systematic protein interactome mapping studies (Figure 1B). Furthermore, VirtualFlow can perform multi-stage screenings, beneficial for the proteins exhibiting high flexibility at the target site. In such cases, a conventional docking to a rigid protein is carried out first, followed by rescoring of the top hits with a flexible protein model. The starting protein structures can be obtained from the Protein Data Bank (PDB) or by using structure prediction methods, such as AlphaFold.[21] When needed, the molecular dynamic simulations can be performed during the protein structure preparation step prior to docking. Once the virtual screening step is completed, the highest ranked hits can be tested in the lab. If an experimentally confirmed compound is promising but requires further optimization, a virtual analog library can be created and screened and the best virtual analogs then re-tested in the same setting. Alternatively, classical medicinal chemistry and structure-activity relationship (SAR) studies can be implemented to further optimize the compounds, which is particularly suitable at the later stages of the optimization procedure.

Structural studies and biological validations of interacting molecules

As soon as interacting virus-host protein domains are experimentally validated (Figure 1A), they can be expressed and purified to conduct follow-up binding assays and structural biology studies. For example, ITC, surface plasmon resonance (SPR), and/or nuclear magnetic resonance (NMR) spectroscopy can be employed to validate the virtual screening results and confirm binding of the small molecules to their targets.[26] Since viruses often acquire resistance by developing mutations in response to inhibitors, *in vitro* resistance selection antiviral assays can also be conducted to validate the binding sites of the compounds. In parallel, inhibition of the corresponding virus-host PPIs by the candidate small molecules can be evaluated using the aforementioned experimental techniques (Figure 1A). Analogs of a hit compound can be screened to identify more potent molecules, i.e., those with higher binding affinities for their targets (Figure 1B). In addition to NMR and X-ray crystallography, single-particle cryoelectron microscopy (cryo-EM) can also be applied to solve the 3D structures of the desired PPI complexes.

In parallel to those biophysical assays, primary cell-based experiments can be conducted to evaluate whether the candidate small molecules can cross cellular membranes while avoiding cytotoxicity. Comparative proteomic, transcriptomic, and metabolomic analyses can then be implemented to model potential perturbed cellular functions and anticipate off-target and side effects *in vivo*. For example, genome-scale metabolic network models could be used to identify targets that alter cellular metabolism in a disease state.[27,28] Currently, *Recon3D*,[29] a reconstruction of the human metabolic network with over 3,200 ORFs, 13,500 metabolic reactions, and 12,800 protein structures, serves as a comprehensive computational resource to predict proteins that should contract metabolic perturbations following viral infections or disruptions by other agents, including small molecules. These *in silico* evaluations can further reduce the time and costs of wet-lab experimentations. Follow-up studies can then be performed to confirm that the small molecules indeed impact the viral life cycle and/or enhance the resistance of the target cells to infection at low micromolar to low nanomolar concentrations.

Based on these results, the best PPI inhibitors can then be used to conduct pharmacokinetic (PK) and pharmacodynamic (PD) assays and initiate pre-clinical and clinical studies to rapidly deliver novel antivirals (Figure 1C). This pipeline presents two obvious challenges: (1) complicated structural optimization of the PPI inhibitors at early discovery stages and (2) potential side effects resulting from blocking other functions of the targeted host proteins. These might potentially be overcome by early determination of 3D structures of the PPI targets and by testing molecules in relevant animal models, respectively. Importantly, our methodology should allow the identification of broad-spectrum, small-molecule PPI inhibitors, since the same host protein domains are often targeted by viruses from distant families.

ACKNOWLEDGMENTS

J.O. thanks Wallonie-Bruxelles International (WBI-World Excellence Fellowship), the Belgian F.R.S.-FNRS-Télévie (FC27371, Crédit no. 7454518F), and the Rega Institute for Medical Research, KU Leuven for the financial support. This work was in part supported by NIAID R21 AI153738 awarded to M.A.C. and by the NYU Abu Dhabi Faculty Research Fund AD060 awarded to K.S.-A. J.-C.T. is a Maître de Recherche of the Belgian F.R.S.-FNRS. C.G. thanks ChemAxon for a free academic license. This work was in part supported by NIH grants CA200913, AI037581, and GM129026 awarded to Gerhard Wagner, as well as by the ARO grant W911NF1910302 to Arthur Jaffe.

AUTHOR CONTRIBUTIONS

J.O., J.-C.T., and C.G. wrote the manuscript with critical insight from S.B.M., A.N.V., K.S.-A., K.D., and M.A.C. All authors discussed the proposed methodology and experimental approaches.

DECLARATION OF INTERESTS

C.G. is the cofounder of two companies, Virtual Discovery, Inc. and Quantum Therapeutics, Inc., which are engaged in drug discovery activities, in part using VirtualFlow.

REFERENCES

1. Flint, S.J., Racaniello, V.R., Rall, G.F., Hatziioannou, T., and Skalka, A.M. (2020). Principles of virology. In Pathogenesis and Control, *Vol. 2* (John Wiley & Sons).
2. Christ, F., Shaw, S., Demeulemeester, J., Desimmie, B.A., Marchand, A., Butler, S., Smets, W., Chaltin, P., Westby, M., Debyser, Z., et al. (2012). Small-molecule inhibitors of the LEDGF/p75 binding site of integrase block HIV replication and modulate integrase multimerization. Antimicrob. Agents Chemother. *56*, 4365–4374.
3. Milroy, L.G., Grossmann, T.N., Hennig, S., Brunsveld, L., and Ottmann, C. (2014). Modulators of protein-protein interactions. Chem. Rev. *114*, 4695–4748.
4. Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., and Shi, J. (2020). Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. Signal Transduct Target Ther *5*, 213.
5. Clackson, T., and Wells, J.A. (1995). A hot spot of binding energy in a hormone-receptor interface. Science *267*, 383–386.
6. Trepte, P., Kruse, S., Kostova, S., Hoffmann, S., Buntru, A., Tempelmeier, A., Secker, C., Diez, L., Schulz, A., Klockmeier, K., et al. (2018). LuTHy: a double-readout bioluminescence-based two-hybrid technology for quantitative mapping of protein–protein interactions in mammalian cells. Mol. Syst. Biol. *14*, e8071.
7. Choi, S.G., Olivet, J., Cassonnet, P., Vidalain, P.O., Luck, K., Lambourne, L., Spirohn, K., Lemmens, I., Dos Santos, M., Demeret, C., et al. (2019). Maximizing binary interactome mapping with a minimal number of assays. Nat. Commun. *10*, 3907.
8. Trepte, P., Secker, C., Choi, S.G., Olivet, J., Ramos, E.S., Cassonnet, P., Golusik, S., Zenkner, M., Beetz, S., Sperling, M., et al. (2021). A quantitative mapping approach to identify direct interactions within complexomes. Preprint at bioRxiv. https://doi.org/10.1101/2021.08.25.457734.
9. Milhas, S., Raux, B., Betzi, S., Derviaux, C., Roche, P., Restouin, A., Basse, M.J., Rebuffet, E., Lugari, A., Badol, M., et al. (2016). Protein-protein interaction inhibition (2P2I)-oriented chemical library accelerates hit discovery. ACS Chem. Biol. *11*, 2140–2148.
10. Torchet, R., Druart, K., Ruano, L.C., Moine-Franel, A., Borges, H., Doppelt-Azeroual, O., Brancotte, B., Mareuil, F., Nilges, M., Ménager, H., et al. (2021). The iPPI-DB initiative: a community-centered database of protein-protein interaction modulators. Bioinformatics *37*, 89–96.
11. Dougherty, P.G., Qian, Z., and Pei, D. (2017). Macrocycles as protein-protein interaction inhibitors. Biochem. J. *474*, 1109–1125.
12. Wang, X., Ni, D., Liu, Y., and Lu, S. (2021). Rational design of peptide-based inhibitors disrupting protein-protein interactions. Front. Chem. *9*, 682675.
13. Dorr, P., Westby, M., Dobbs, S., Griffin, P., Irvine, B., Macartney, M., Mori, J., Rickett, G., Smith-Burchnell, C., Napier, C., et al. (2005). Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. Antimicrob. Agents Chemother. *49*, 4721–4732.
14. Tan, Q., Zhu, Y., Li, J., Chen, Z., Han, G.W., Kufareva, I., Li, T., Ma, L., Fenalti, G., Li, J., et al. (2013). Structure of the CCR5 chemokine receptor–HIV entry inhibitor maraviroc complex. Science *341*, 1387–1390.
15. Luck, K., Kim, D.K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B., et al. (2020). A reference map of the human binary protein interactome. Nature *580*, 402–408.
16. The ORFeome Collaboration (2016). The ORFeome collaboration: a genome-scale human ORF-clone resource. Nat. Methods *13*, 191–192.
17. Huttlin, E.L., Bruckner, R.J., Navarrete-Perea, J., Cannon, J.R., Baltier, K., Gebreab, F., Gygi, M.P., Thornock, A., Zarraga, G., Tam, S., et al. (2021). Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell *184*, 3022–e28.
18. Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature *583*, 459–468.
19. Kim, D.-K., Weller, B., Lin, C.-W., Sheykhkarimli, D., Knapp, J.J., Kishore, N., Sauer, M., Rayhan, A., Young, V., Marin-de la Rosa, N., et al. (2021). A map of binary SARS-CoV-2 protein interactions implicates host immune regulation and ubiquitination. Preprint at bioRxiv. https://doi.org/10.1101/2021.03.15.433877.
20. Maseko, S., Van Molle, I., Blibek, K., Gorgulla, C., Olivet, J., Blavier, J., Vandermeulen, C., Skupiewski, S., Saha, D., Ntombela, T., et al. (2021). Interactome and structural basis for targeting the human T-cell leukemia virus Tax oncoprotein. Preprint at bioRxiv. https://doi.org/10.1101/2021.08.25.457680.
21. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature *596*, 590–596.
22. Kim, H.H., Hyun, J.S., Choi, J., Choi, K.E., Jee, J.G., and Park, S.J. (2018). Structural ensemble-based docking simulation and biophysical studies discovered new inhibitors of Hsp90 N-terminal domain. Sci. Rep. *8*, 368.
23. Hori, K., Ajioka, K., Goda, N., Shindo, A., Takagishi, M., Tenno, T., and Hiroaki, H. (2018). Discovery of potent disheveled/dvl inhibitors using virtual screening optimized with NMR-based docking performance index. Front. Pharmacol. *9*, 983.
24. Hughes, T.E.T., Del Rosario, J.S., Kapoor, A., Yazici, A.T., Yudin, Y., Fluck, E.C., III, Filizola, M., Rohacs, T., and Moiseenkova-Bell, V.Y. (2019). Structure-based characterization of novel TRPV5 inhibitors. eLife *8*, e49572.
25. Cole, B.A., Johnson, R.M., Dejakaisaya, H., Pilati, N., Fishwick, C.W.G., Muench, S.P., and Lippiat, J.D. (2020). Structure-based identification and characterization of inhibitors of the epilepsy-associated $K_{Na}1.1$ (KCNT1) potassium channel. iScience *23*, 101100.
26. Gorgulla, C., Boeszoermenyi, A., Wang, Z.-F., Fischer, P.D., Coote, P.W., Das, K.M.P., Malets, Y.S., Radchenko, D.S., Moroz, Y.S., Scott, D.A., et al. (2020). An open-source drug discovery platform enables ultra-large virtual screens. Nature *580*, 663–668.
27. Yizhak, K., Gabay, O., Cohen, H., and Ruppin, E. (2013). Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. Nat. Commun. *4*, 2632.
28. Venayak, N., von Kamp, A., Klamt, S., and Mahadevan, R. (2018). MoVE identifies metabolic valves to switch between phenotypic states. Nat. Commun. *9*, 5332.
29. Brunk, E., Sahoo, S., Zielinski, D.C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G.A., Aurich, M.K., et al. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat. Biotechnol. *36*, 272–281.