

Structural Genomics in Drug Discovery: An Overview

Karri Sowjanya, Chandrashekar Girish

Department of Pharmacology, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

Abstract

Structural genomics, technology and methodology driven field has emerged in the past 20–25 years and mainly aims at coming up with as many three-dimensional structures of proteins as possible by combining experimental (X-ray crystallography and nuclear magnetic resonance) and comparative modeling methods at a minimal cost in comparison to the traditional approaches. Structure aided drug discovery which is used since long mainly focuses on a single target protein whereas structural genomics targets and investigates multiple proteins simultaneously, thus cutting down time and cost. These activities are done at specialized structural genomic centers available worldwide that help to interpret the biology of humans and pathogenic organisms in a better way leading to an increased productivity and saving time. Virtual screening methods like docking have also brought a great impact to quicken the drug discovery process. In diseases of public health concern (like tuberculosis), structural genomics helped greatly by giving important insights toward a better comprehension of the pathways associated in its pathogenesis and in guiding drug discovery. Several newer targets such as Scaff10-8, MS2734 (6), and GSK-J4 were developed showcasing their beneficial effects in a wide variety of disorders including diabetes insipidus, Parkinson's disease, and autoimmune diseases, respectively. Furthermore, these programs guide to understand the root cause and the mechanism associated with various human diseases and their pathogens. This review tries to look into the high throughput methods of structure determination, the newer targets developed in the past few years and the benefits offered that can pave a way for faster and efficient drug discovery.

Keywords: Docking, modeling, virtual screening, X-ray crystallography

INTRODUCTION

Structural genomics, as the term denotes is the field of genomics which is mainly determined to speculate the three-dimensional (3D) structure of the genes or proteins in a high-throughput way by the usage of either experimental or computational techniques. Most of the times, it is associated with 3D-protein structure determination and is hence referred to as “Structural Proteomics.”^[1] These were introduced around 1990s, and their major intention was to devise high throughput (HTP) methods, enhance the structure quality and to bring down the cost per structure of macromolecules.^[2] Initially, the determination of protein structure is done either experimentally by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy; or with the help of computational modeling techniques that involve the demonstration of homology to proteins of known structure.^[1]

Now, the prediction of protein function or its molecular mechanism from its structure gives an idea regarding the binding of drugs at a specific location (to some extent) in the protein structure, their rearrangements and also the relationship

with other ligands. Later, the structural genomics has taken a step forward in drug discovery by revealing certain novel targets with the aid of certain virtual screening methods such as ligand-based and structure-based (docking) screening methods for which an insight about the fixed and flexible parts on the protein surface is considered imperative.^[3] The macromolecular docking may be dealt in different ways that include – by its atoms (denoting its types and positions); and by the conversion of information into either grids or maps three-dimensionally.^[3] Structural genomics mainly focuses on the challenging proteins, and it aims to address the major confrontations in the field of structural molecular biology. This further helps in the refinement of biomedical research by the usage of advanced techniques and data processing methods for the societal benefit.^[2]

Address for correspondence: Dr. Chandrashekar Girish,
Department of Pharmacology, Jawaharlal Institute of Postgraduate Medical
Education and Research, Puducherry, India.
E-mail: gcnx@rediffmail.com

Access this article online

Quick Response Code:



Website:
www.jpharmacol.com

DOI:
[10.4103/jpp.JPP_117_18](https://doi.org/10.4103/jpp.JPP_117_18)

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Sowjanya K, Girish C. Structural genomics in drug discovery: An overview. *J Pharmacol Pharmacother* 2019;10:1-6.

Received: 05-10-2018 **Revised:** 05-12-2018 **Accepted:** 05-01-2019

Structural genomics versus functional genomics versus comparative genomics

Structural genomics mainly looks into the static aspects such as genome sequences and helps in the structure determination of as many proteins as possible that further helps in determining the structure of other macromolecules based on homology modeling. A homology model with >30% sequence identity is contemplated to be significant. By combining the structural information and sequence identity, it also helps to establish novel relationships between proteins that are related distantly.^[4] Functional genomics is concerned with the dynamic aspects by describing their functions through the application of genome-wide association studies. It gathers data from DNA and RNA sequences and employs this to study transcription, translation, and other interactions (such as DNA-protein, RNA-protein, and protein-protein interactions) to understand regarding the expression of genes, cell differentiation, and cell cycle progression. In short, it covers transcriptomics, proteomics, metabolomics, and interactomics.^[5] Comparative genomics, however, relates to the sequence identity which remains conserved across multiple or distantly related species that implies sharing of a common function but not vice versa. It also helps in understanding the evolutionary mechanisms, the biological and pathological processes that might be responsible for a specific biological function in those with identical sequence.^[6]

Industrial versus academic structural genomics versus structural genomics consortium

The aims governing to conduct structural genomics projects differ to some extent between academics and industries. A collaboration between them often plays a crucial role in drug discovery. However, there are some conflicts which have to be looked through for it to be successful. First, the major conflict is regarding the target selection. The academic projects are mainly focused to work on those protein structures which are easily expressed as their major concern is successful production of increased number of structures. Hence, the drug targets that are actually produced are comparatively less. Second, in academic projects, once a structure of a protein in a particular family is solved, they move on to another family. This is particularly difficult to find out the new antibiotics in some instances of bacterial resistance as there are multiple structures associated within the same family with different pathogens. Third, integral membrane proteins which are the major drug targets and play a prime role in health care are not focussed on.^[7]

Structural Genomics Consortium (SGC) is an open access public-private collaboration that mainly works with the 3D-structure determination of the less focused and studied areas in a human genome by keeping a special attention on antibodies, chemical probes, and epigenetics. It is a large scale organization with expertise in various fields and a huge amount of resources that can help maintain the quality and reproducibility thereby increasing its efficiency. It receives its

funding from the public, private, and charitable associations and repays them by providing a membership in the SGC board. The funding by the public sector helps to maintain an open-access whereas that by the private sector gains them by preventing duplication between the pharmaceutical companies thereby reducing the effort and the cost in producing new molecules. In the recent years, the research in SGC has progressed very rapidly and efficiently in comparison to either industries or academics by exploring new areas.^[8]

Structural genomic centers

There are numerous structural genomic centers worldwide to determine the protein structure among which Protein Structure Initiative (PSI) was one of the towering programs that made an extensive contribution by laying the preparatory measures necessary for drug discovery. Later, many programs were initiated as a part of PSI like the Tuberculosis SGC (TBSGC). All these SG groups deposit structures in the Protein Data Bank (PDB) for easy accessibility to conduct research and pursue to establish the structure of as many target proteins as feasible even before their identification in the disease pathogenesis. Despite the tedious efforts by these centers, there are hardly any drugs developed which are solely attributed to structural genomics as the field of structural genomics started evolving recently from a decade compared to the drug development process which is thriving since long. Some of the important SG centers are mentioned in Table 1.^[9,10]

Protein structure determination

Initially, protein structure is determined that helps in predicting the protein function guiding in drug discovery by molecular docking techniques. First, the protein structure determination is done either by *de novo* (experimental) methods or by modeling-based methods (by comparing with a protein of known structure).

The *de novo* method is a traditional method of protein structure determination either by X-ray crystallography, NMR spectroscopy, or electron microscopy.^[9,11] Among these, X-ray crystallography is more accurate and is considered to be more precise for structure determination. This can either be an HTP protein crystallography or a low throughput protein crystallography. The physical principles and methodological basis remain the same between the two types; however, they only differ in terms of automation, cost, and the integration process.^[12] NMR spectroscopy is a substitute to X-ray crystallography for proteins of small-to-medium size. HSQC (heteronuclear single quantum coherence) spectra are the main deciding factor to use NMR for structure determination (NMR is used for proteins with good HSQC spectra). To fasten up the process, newer techniques were developed such as ultra-high field magnet, chilled probe technology, transverse relaxation optimization spectroscopy, and isotope labeling techniques.^[13] Electron microscopy and atomic force microscopy can also be done to anticipate the structure at a very low resolution which can be later confirmed by X-ray crystallography.^[10]

The modeling-based methods are done by comparing with the proteins of the PDB through profile-profile matching, model building, or threading. In profile-profile matching, a PSI-Basic Local Alignment and Search Tool search is done to identify the closely related sequences of the query compound in the database. In modeling-based approach, a comparative approach is made, and the targets are selected so that the remaining sequences can be changed to attain sufficient accuracy. Threading is another method of matching based on the fold prediction method (fold similarities between the molecules).^[11,14]

Protein function determination

The next step followed by structure determination is to predict its function. The function of a gene and its encoded protein can often be predicted by several ways that include sequence similarity, fold similarity, structural motifs, attachment to binding sites, 3D-template methods, and protein-protein interaction sites. Sequence similarity is done by the comparison of the gene sequences with those in the database. The amino acid sequence helps to determine the protein structure, and from the structure, its function is determined; proteins sharing a similar sequence generally exhibit similar biochemical functions, although found in distantly related organisms. At present, the determination of what a newly discovered protein does begins with a search of the proteins identified previously that are similar in their amino acid sequences.^[15,16] Another way is by fold similarity in which proteins having the same folds might elicit a similar function due to a common ancestry. Structural motifs are those that exhibit a particular function like the DNA binding motifs such as zinc fingers and leucine zippers. Hence, based on their similarity, its function is

determined. In 3D-template methods, a thorough search of the protein structure is made to find out the clusters of residues that perform a specific function. The best example of a 3D-template method to predict function is the Ser-His-Asp catalytic triad of serine proteases.^[15]

A number of databases are also available that can help provide information to do the functional analysis. These include the motif-based databases (PRINTS, PROSITE), domain-based databases (Pfam, Gene3D, SUPERFAMILY, TIGRFAMs, SMART), gene-based databases (SYSTEMS, SWISS-PROT), and other miscellaneous ones (GO, WIT, KEGG). As a thorough search of each of these databases separately takes a lot of time, “meta-servers” such as Motifscan, Interpro, and CDsearch (by NCBI) were developed to merge all the available data from the diverse sources concomitantly.^[17]

Virtual screening methods

With the rapid advancement in structural genomics generating protein structures substantially at a higher rate, it is of utmost importance to screen potential targets and anti-targets that can guide in drug discovery by the usage of virtual screening methods. These methods employ computers to coalesce and analyze the data from different sources and databases to identify the potential hits and to complement HTP screening. The various types of virtual screening (or *in silico* screening) methods as listed in Figure 1.^[18,19]

The central idea of virtual screening is to limit the virtual space of the molecules and to screen against a particular target to help synthesize a lead molecule that can guide in drug discovery. It mainly aims in evaluating the drug-likeness of the molecules as a majority of the lead molecules fail during clinical trials due to problems associated with its pharmacokinetic parameters and other toxicity related issues. In brief, drug-likeness can be evaluated by the method of simple counting based on their molecular properties such as molecular weight, charge, lipophilicity; or by comparing the hits with molecules in the database to identify specific functional groups that can result in toxicity; or by similarity testing with their structure by the identification of characteristic motifs; or by means of pharmacophore in which the functional groups, atoms or the geometry of the query compound are matched with the compounds in database that can facilitate in drug discovery.^[19]

Table 1: Overview of the common structural genomic centers and associated target proteins

SG centers	Target proteins
PSI and ID centers	<i>Shigella flexneri</i> , <i>Vibrio cholerae</i> , <i>Vibrio parahaemolyticus</i> , <i>Bacillus anthracis</i> , <i>Salmonella typhimurim</i> , <i>Listeria monocytogenes</i>
JCSG	<i>Thermotoga maritima</i> , baculovirus, adenovirus, human GPCRs expression in <i>Escherichia coli</i>
MCSG	Bacterial and eukaryotic proteins
NESG	<i>Saccharomyces cerevisiae</i> , <i>Drosophila melanogaster</i> , <i>Caenorhabditis elegans</i>
NYSGXRC	Yeast and bacterial proteins and their expression
SGC	Eukaryotic proteins of public health importance (cancers, malaria, diabetes)
TBSGC	Mycobacterium tuberculosis proteins and expression in <i>Escherichia coli</i>
RSGI	Proteins of the mouse, <i>Thermus thermophilus</i> and <i>Arabidopsis thaliana</i>
MePNet	GPCRs; expression in <i>Escherichia coli</i> and mammalian cells

PSI=Protein structure initiative, ID=Infectious disease, SG=Structural genomic, JCSG=Joint Center for SG, MCSG=Midwest Center for SG, NESG=Northeast SG Consortium, NYSGXRC=New York SG Research Consortium, SGC=SG Consortium, TBSGC=Tuberculosis SG Consortium, RSGI=RIKEN SG Initiative, MePNet=Membrane protein network, GPCRs=G protein coupled receptors

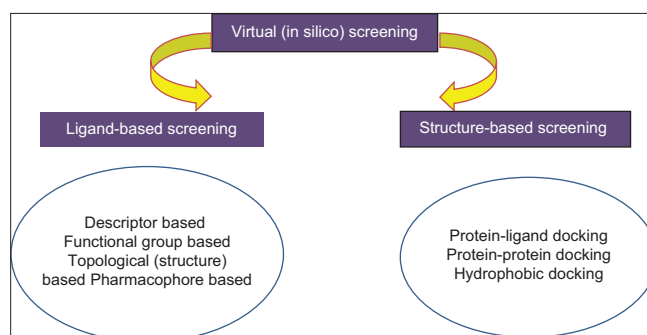


Figure 1: A brief enumeration of the types of virtual screening

Structure-based drug design is by means of docking in which the complementarity of the steric and electrostatic configurations is made use of to understand the relationships between the receptor and ligand. The ligand often is a small molecule or a protein, and the receptor is a protein of interest or a specific DNA or RNA sequence. The foremost step before the initiation of docking is to identify the specific binding site on the receptor so as to limit the search thereby limiting the degrees of freedom (both translational and rotational) on the receptor surface to be searched. This is followed by a scoring function considering the ability of the ligand to bind with its receptor.^[20]

The various search algorithms include-grid search, descriptor matching, and by energy-based methods. First, the grid search as the name implies, makes a thorough search with the ligand in all the available degrees of freedom so as to identify the binding site on the receptor. The Fast Fourier transform methods are used to accelerate the searching process. However, this method is not successful to dock small molecules as the search area is too large due to the flexibility of most of the ligands. The second method is by descriptor matching in which descriptors with some specific properties are placed in the receptor site so that they can be matched with the ligand for efficient interaction and binding which is later scored. Third, energy-based methods make use of molecular dynamics by doing a local search to find out the energy minima for the ligand-receptor interaction. Scoring is done after the calculation of the binding free energy. Previously, it is either done by thermodynamic integration or by free energy perturbation methods. Due to its time constraints, other approaches such as molecular mechanics force field, empirical and semi-empirical methods, knowledge-based are being followed which mainly focus on the interaction ability and bonding.^[20]

Drug discovery

Structural biology has paved a way in the process of drug discovery for quite a long time. Now, the focus is shifted toward structure-based drug discovery. The knowledge about its structure helps us identify its potency as well as its selectivity thereby reducing the incidence of serious adverse effects. The initiatives made in structural genomics helps in the discovery of newer drug targets to be used as a tool in drug designing. The major attention is toward the membrane proteins as they constitute the majority of the drug targets (about 70%) and markedly affects the drug discovery.^[10] A brief review on the drug discovery by structural genomics is depicted in Figure 2.

Newer developments

With the advances in structural genomics, numerous drug targets have been developed with diverse functions. Some of the recent developments are mentioned in Table 2.

Benefits and future direction

Structural genomics programs helped in upgrading the understanding of molecular foundation of human diseases. They have also contributed significantly to the structural coverage of genomes of pathogens associated with some

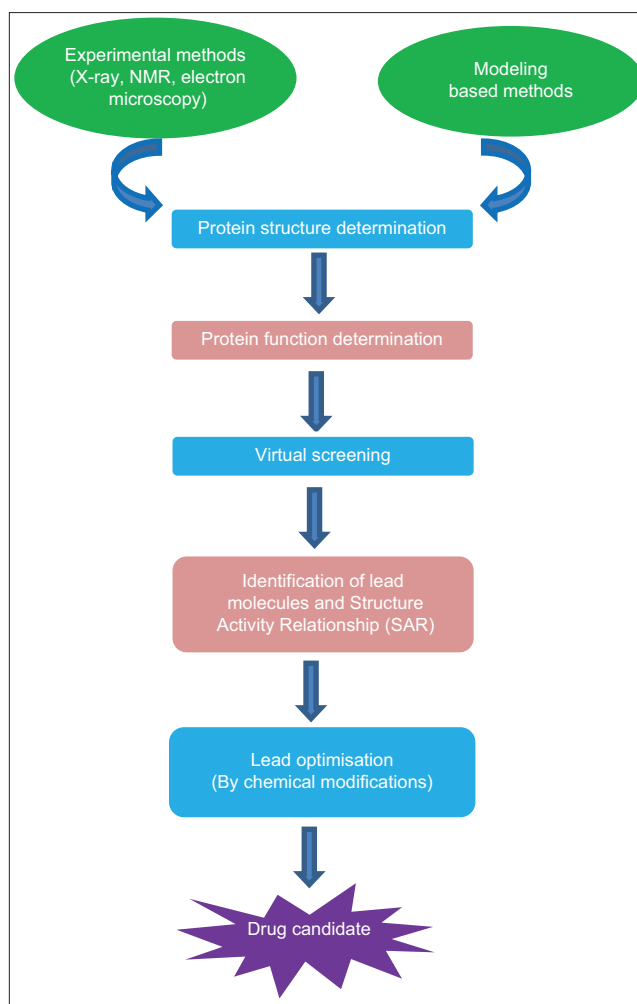


Figure 2: Structural genomic-based drug discovery process

infectious diseases as well as structurally uncharacterized biological processes in general. Most importantly, these programs aimed to develop new methodologies at almost all the steps throughout the process of structure determination; not alone for determining the novel structures highly and efficiently but also to help in screening the protein/ligand interactions.^[9] The protein structures of hundreds and thousands of medically relevant targets from infectious disease organisms are found and are available freely on the websites. This information about the newer structures provides both academic and industrial scientists with an opportunity to hasten the development of chemotherapeutic agents against these pathogens.^[26]

However, the existing databases have to be modified regularly to allow the easy dissemination of results from these fragment screens and a serious effort is needed to encourage the small and big pharmaceutical industries to release the coordinates of drug targets of globally important infectious disease organisms. It is also mandatory for the structural biologists to collaborate with the medicinal chemists and other molecular biologists to help convert these fragments from promising leads to effective drugs. Together, all these steps should be integrated in an effective manner to release numerous structures that can help

Table 2: Newer targets with their function and utility in disease conditions

Target/drug	Functions	Disease conditions
Scaff10-8	AKAP-Lbc-RhoA interaction inhibitor Promotes translocation of aquaporin-2 to the plasma membrane of renal collecting duct principal cells	Water balance disorders like diabetes insipidus ^[21]
MS2734 (6)	Bisubstrate NNMT inhibitor NNMT and its methylated products are involved in pathogenesis of human diseases	Parkinson's disease; tool to study the enzymatic mechanisms ^[22]
GSK-J4	Histone H3K27 demethylase inhibitor Modulates the inflammatory response by decreasing IFN- γ , TNF α , GM-CSF, IL-10 levels ^[23]	Autoimmune diseases like rheumatoid arthritis
4-Anilinoquinolines KDM2/7 inhibitors	Inhibitors of cyclin GAK that regulate the entry of bacteria and into the host cells Targets cancer cells at the transcription level with selectivity and good cell permeability	In infectious diseases as a preventive therapy ^[24] Gastric carcinoma and small cell carcinoma of the lung ^[25]

AKAP-Lbc=A-kinase anchoring protein, NNMT=Nicotinamide N-Methyltransferase, IFN- γ =Interferon gamma; TNF α =Tumor necrosis factor alpha, GM-CSF=Granulocyte monocyte colony stimulating factor, IL=Interleukin; GAK=G associated kinase, KDM=Histone lysine demethylases

provide a tremendous resource to improve health in both the developed and developing countries.^[26]

CONCLUSION

Although the structure-based drug design has a long history to start with, to date only a few drugs have come out into the market through this approach. This is attributed to a delay in the identification of lead compounds from the available structural knowledge and in bringing them into the market successfully. However now, the field of structural genomics has progressed rapidly. The investment by the pharmaceutical companies in structural genomics has helped to explore the structural biology, and its assistance with bioinformatics further helped in guiding drug discovery. The structural genomics programs established worldwide are diligent in providing the structures for many infectious pathogens. These structures will further guide and accelerate drug discovery very soon in the future. Although the structure of each and every human protein cannot be determined, some of the structures in each family can be obtained that can help to predict its function to an extent, finally developing a drug candidate.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- Brenner SE. A tour of structural genomics. *Nat Rev Genet* 2001;2:801-9.
- Joachimik A. High-throughput crystallography for structural genomics. *Curr Opin Struct Biol* 2009;19:573-84.
- Abagyan R. Problems in computational structural genomics. In: Sundstrom M, Norin M, Edwards A, editors. *Structural Genomics and High Throughput Structural Biology*. New York: Taylor and Francis Group; 2006. p. 223-50.
- Goldsmith-Fischman S, Honig B. Structural genomics: Computational methods for structure analysis. *Protein Sci* 2003;12:1813-21.
- Bunnik EM, Le Roch KG. An introduction to functional genomics and systems biology. *Adv Wound Care (New Rochelle)* 2013;2:490-8.
- Alföldi J, Lindblad-Toh K. Comparative genomics as a tool to understand evolution and disease. *Genome Res* 2013;23:1063-8.
- Russell RB, Eggleston DS. New roles for structure in biology and drug discovery. *Nat Struct Biol* 2000;7 Suppl: 928-30.
- Jones MM, Castle-Clarke S, Brooker D, Nason E, Huzair F, Chataway J. The structural genomics consortium: A knowledge platform for drug discovery: A summary. *Rand Health Q* 2014;4:19.
- Grabowski M, Chruszcz M, Zimmerman MD, Kirillova O, Minor W. Benefits of structural genomics for drug discovery research. *Infect Disord Drug Targets* 2009;9:459-74.
- Lundstrom K. Structural genomics and drug discovery. *J Cell Mol Med* 2007;11:224-38.
- Baker D, Sali A. Protein structure prediction and structural genomics. *Science* 2001;294:93-6.
- Rupp B. High throughput protein crystallography. In: Sundstrom M, Norin M, Edwards A, editors. *Structural Genomics and High Throughput Structural Biology*. New York: Taylor and Francis Group; 2006. p. 61-104.
- Lee W, Yee A, Arrowsmith CH. NMR spectroscopy in structural genomics. In: Sundstrom M, Norin M, Edwards A, editors. *Structural Genomics and High Throughput Structural Biology*. New York: Taylor and Francis Group; 2006. p. 49-60.
- Teichmann SA, Chothia C, Gerstein M. Advances in structural genomics. *Curr Opin Struct Biol* 1999;9:390-9.
- Laskowski RA. Determining function from structure. In: Sundstrom M, Norin M, Edwards A, editors. *Structural Genomics and High Throughput Structural Biology*. New York: Taylor and Francis Group; 2006. p. 163-84.
- Vitkup D, Melamud E, Moul J, Sander C. Completeness in structural genomics. *Nat Struct Biol* 2001;8:559-66.
- Watson JD, Todd AE, Bray J, Laskowski RA, Edwards A, Joachimik A, *et al*. Target selection and determination of function in structural genomics. *IUBMB Life* 2003;55:249-55.
- Taboureau O, Baell JB, Fernández-Reco J, Villoutreix BO. Established and emerging trends in computational drug discovery in the structural genomics era. *Chem Biol* 2012;19:29-41.
- Vyas V, Jain A, Jain A, Gupta A. Virtual screening: A fast tool for drug design. *Sci Pharm* 2008;76:333-60.
- Brooijmans N. Docking methods, ligand design, and validating data sets in the structural genomics era. In: Jenny G, Philip E, editors. *Structural Bioinformatics*. New York: John Wiley and Sons; 2009. p. 635-63.
- Schrade K, Tröger J, Eldahshan A, Zühlke K, Abdul Azeez KR, Elkins JM, *et al*. An AKAP-Lbc-RhoA interaction inhibitor promotes the translocation of aquaporin-2 to the plasma membrane of renal collecting duct principal cells. *PLoS One* 2018;13:e0191423.
- Babault N, Allali-Hassani A, Li F, Fan J, Yue A, Ju K, *et al*. Discovery of bisubstrate inhibitors of nicotinamide N-methyltransferase (NNMT). *J Med Chem* 2018;61:1541-51.
- Cribbs A, Hookway ES, Wells G, Lindow M, Obad S, Oerum H, *et al*. Inhibition of histone H3K27 demethylases selectively modulates

- inflammatory phenotypes of natural killer cells. *J Biol Chem* 2018;293:2422-37.
24. Asquith CR, Laitinen T, Bennett JM, Godoi PH, East MP, Tizzard GJ, *et al.* Identification and optimization of 4-anilinoquinolines as inhibitors of cyclin G associated kinase. *ChemMedChem* 2018;13:48-66.
25. Gerken PA, Wolstenhulme JR, Tumber A, Hatch SB, Zhang Y, Müller S, *et al.* Discovery of a highly selective cell-active inhibitor of the histone lysine demethylases KDM2/7. *Angew Chem Int Ed Engl* 2017;56:15555-9.
26. Van Voorhis WC, Hol WG, Myler PJ, Stewart LJ. The role of medical structural genomics in discovering new drugs for infectious diseases. *PLoS Comput Biol* 2009;5:e1000530.