

PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA

Melissa F. Adasme^{1,†}, Katja L. Linnemann^{1,†}, Sarah Naomi Bolz^{1,†}, Florian Kaiser², Sebastian Salentin¹, V. Joachim Haupt² and Michael Schroeder^{1,*}

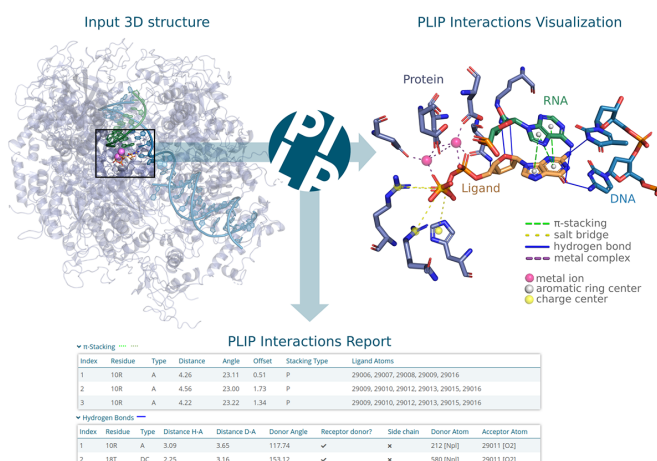
¹Biotechnology Center (BIOTEC), CMCB, Technische Universität Dresden, Tatzberg 47-49, 01307 Dresden, Germany and ²PharmAI GmbH, 01307 Dresden, Germany

Received January 29, 2021; Revised March 24, 2021; Editorial Decision April 08, 2021; Accepted April 13, 2021

ABSTRACT

With the growth of protein structure data, the analysis of molecular interactions between ligands and their target molecules is gaining importance. PLIP, the protein–ligand interaction profiler, detects and visualises these interactions and provides data in formats suitable for further processing. PLIP has proven very successful in applications ranging from the characterisation of docking experiments to the assessment of novel ligand–protein complexes. Besides ligand–protein interactions, interactions with DNA and RNA play a vital role in many applications, such as drugs targeting DNA or RNA-binding proteins. To date, over 7% of all 3D structures in the Protein Data Bank include DNA or RNA. Therefore, we extended PLIP to encompass these important molecules. We demonstrate the power of this extension with examples of a cancer drug binding to a DNA target, and an RNA–protein complex central to a neurological disease. PLIP is available online at <https://plip-tool.biotec.tu-dresden.de> and as open source code. So far, the engine has served over a million queries and the source code has been downloaded several thousand times.

GRAPHICAL ABSTRACT



MOTIVATION

The Protein Data Bank (PDB) (1) is continuously growing and recent advances in structure prediction and experimental methods such as cryoEM will further accelerate this growth. The vast majority of protein structures in PDB contain ligands and therefore it is important to understand how these ligands interact with their targets.

PLIP, the protein–ligand interaction profiler (2), addresses this need. It detects hydrogen bonds, hydrophobic contacts, π -stacking, π -cation interactions, salt bridges, water bridges, metal complexes and halogen bonds between ligands and targets. PLIP is easy to use as it requires only a PDB ID or a PDB file as input, it has been running reliably and continuously for 5 years, and it is transparent with all the source code published on GitHub. PLIP 2021 constitutes a major update with added support for nucleic acids, more flexibility by adjustable thresholds, mode and model selection and a more functional and modern design for increased usability.

The two main cases for PLIP are the analysis and visualisation of docking results and the in-depth study of ex-

*To whom correspondence should be addressed. Tel: +49 351 46340062; Fax: +49 351 463 40342; Email: michael.schroeder@tu-dresden.de

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

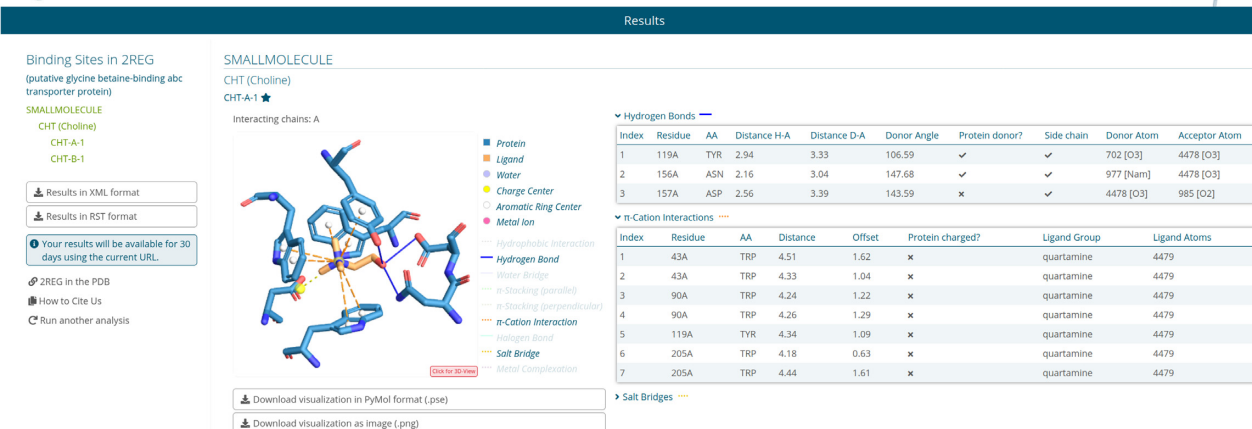
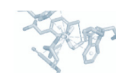


Figure 1. PLIP web tool result page. On the left is a menu for ligands and binding sites, in the middle an image of a selected binding site and on the right a table with details of the interactions.

isting structures. Elfiky uses PLIP to show docking results of a drug repurposing screen for COVID-19 (3). Furlan *et al.* run an inverse docking screen to identify targets of curcumin, a natural compound (4). They show PLIP interactions for a human folate receptor and a phosphodiesterase. PLIP is used on existing structures by Soliman *et al.* to study antibody recognition of tumour-associated glycans (5) and by Kumar *et al.* to survey specific cation– π interactions (6).

PLIP WEB SERVER AND COMMAND-LINE

The PLIP web server provides an easy-to-use interface (Figure 1), which takes as input either a PDB ID or a custom structure in PDB format (e.g. result files from docking or molecular dynamics software). Moreover, the advanced options in the input section allow the users to modify the default settings of PLIP according to their special requirements, such as adjustments to thresholds for the detection of interactions, the consideration of modified residues, the specification of which model to use in multi-model structures, the detection of intra- and inter-chain interactions and the treatment of nucleic acids.

Given an input structure, a list with all binding sites and binding ligands is provided on the left panel of the web tool (see Figure 1). On the right panel, the results for all binding sites are displayed. For each binding site, PLIP offers atom-level binding information, an image, a 3D interactive visualisation with JSmol and a PyMOL session file. In case a manual inspection or subsequent processing is necessary, a parsable XML or RST file with interaction data is available.

The web server is based on the PLIP command-line tools version 2.2.0, which is publicly available at github.com/pharmai/plip under the GNU GPLv2 licence. The command-line tool enables high-throughput computation of protein structures and can be integrated into analysis pipelines using the machine-readable result files. PLIP is easy to use for high-throughput and high-performance computing, as it is virtualised and bundled with all necessary software in Docker and Singularity containers.

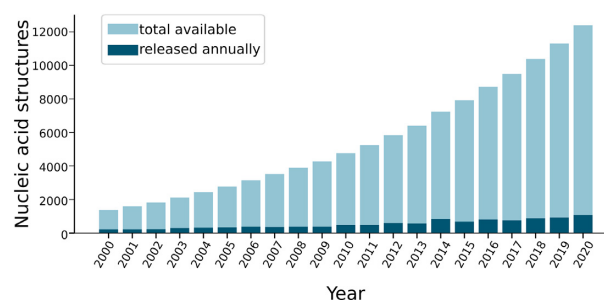


Figure 2. Nucleic acid structures growth in PDB. Number of nucleic acid structures released annually since 2000 and the total aggregated structures.

PLIP FOR DNA AND RNA

PLIP's main focus so far was the detection of interactions between small molecules and proteins. However, with the constant growth of nucleic acid structure data, a large proportion of over 7% of the PDB (12460 structures) contains DNA and RNA (Figure 2). Of these, the majority are protein–nucleic acid complexes (8837) and the rest are mostly DNA or RNA only.

Studying the interactions of nucleic acids to ligands and to proteins is important. DNA/RNA can act, for instance, as a target for novel cancer drugs such as the drug XR5944 (7) and as a ligand to proteins involved in disease such as the protein FUS, which is implicated in the neurological disorder ALS (8).

Protein–RNA interaction: FUS binding UGGUG

FUS is a nuclear RNA-binding protein involved in RNA metabolism. It can undergo liquid–liquid phase separation, forming membraneless organelles (9–11). Aberrant phase transition of FUS into pathological solid aggregates has been linked to neurodegenerative diseases (12,13). Recent studies have shown that high RNA concentrations prevent phase separation of FUS (8) and mutations in FUS can cause aberrant phase separation (14). Loughlin *et al.* have

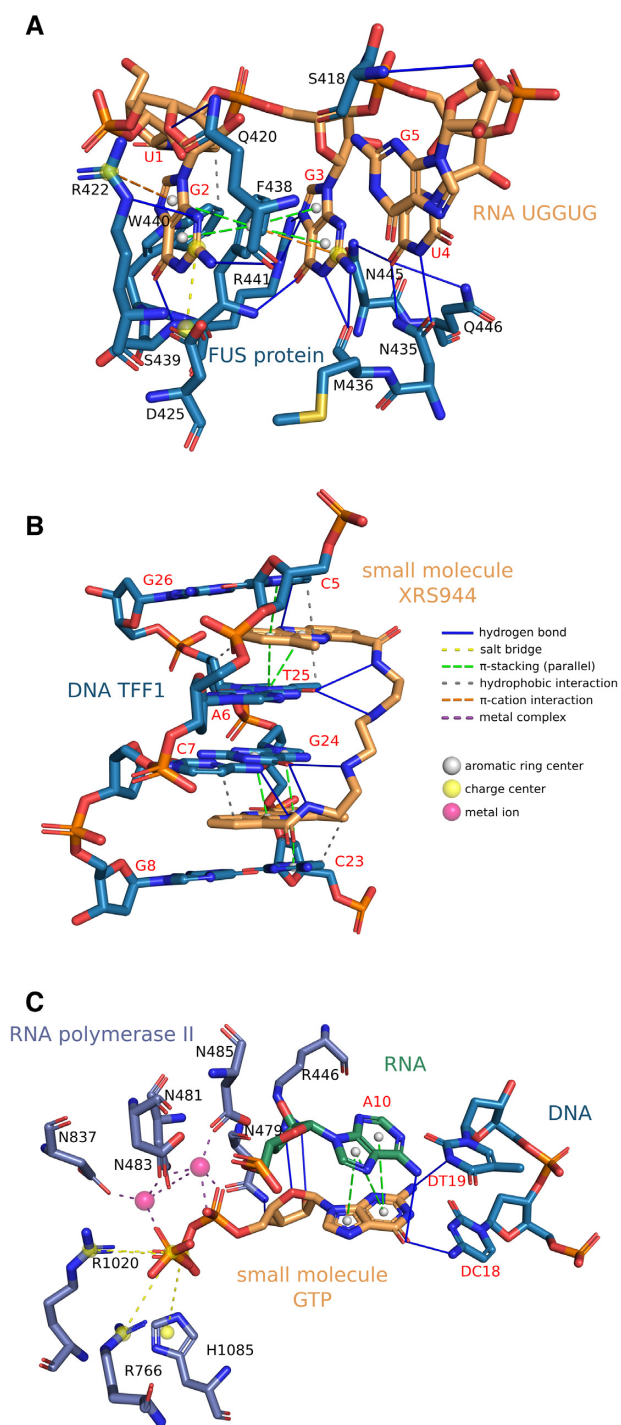


Figure 3. Examples. (A) FUS binding the UGGUG (PDB ID: 6G99), (B) XR5944 binding the TFF1 estrogen response element (PDB ID: 2MG8) and (C) GTP binds RNA polymerase II and DNA/RNA of the elongation complex (PDB ID: 2E2H). Ligands are shown in orange and receptors in blue, green or purple gray. Protein residues are labeled in black and DNA/RNA bases in red.

solved the structure of FUS bound to RNA (15). We applied the new version of PLIP to identify the interactions between the zinc finger of FUS as receptor and the RNA sequence UGGUG as ligand (Figure 3A). PLIP detects π -

stacking interactions of the central Phe438 with the flanking G2 and G3 of the RNA. G2 is additionally anchored by a π -cation contact with Arg422, a salt bridge with Asp425, and several hydrogen bonds with Arg422, Trp440, Ser439 and Gln420. Multiple sequence-specific hydrogen bonds are also formed between G3 and Phe438, Arg441 and Met436. U4 binds the FUS zinc finger via hydrogen bonds with Asn435, Gln446 and Asn445. The zinc finger recognises U1 by a hydrophobic interaction. Illuminating the RNA-binding mode of FUS paves the way for a deeper understanding of the mechanisms that lead to the formation of solid aggregates in neurodegenerative diseases.

Small molecule–DNA interaction: XR5944 binding estrogen response element

The cancer drug candidate XR5944 is a DNA bis-intercalator that strongly binds DNA in a sequence-specific manner (7,16–18). Because of its structural similarity to other topoisomerase inhibitors, XR5944 was originally thought to inhibit topoisomerase I and II (19), but was later shown to be a DNA transcription inhibitor instead (20,21). Interestingly, XR5944 specifically recognises the estrogen response element (ERE), the target DNA sequence of estrogen receptor- α (ER α), making it a promising drug for the treatment of ER α -positive breast cancer (18,22,23). ER α are present in the regulatory region of several genes, including Trefoil factor 1 (TFF1). Lin *et al.* published the structure of two XR5944 molecules bound to the naturally occurring DNA TFF1-ERE (18). Thus, TFF1 acts as receptor and XR5944 as ligand, which PLIP treats accordingly (Figure 3B). PLIP identifies parallel π -stacking interactions between the phenazine rings of the XR5944 molecules and the flanking base pairs. In addition, it detects several hydrophobic interactions at the intercalation sites. Furthermore, the carboxamide aminoalkyl linkers of the XR5944 molecules form site-specific hydrogen bonds with the base pairs in the major groove. The non-covalent interactions formed by XR5944 with the TFF1-ERE determine its binding mode and sequence specificity. Uncovering these interactions provides a structural basis for the rational design of drugs that target ER α .

Small molecule–protein/DNA/RNA interaction: GTP binding RNA polymerase II and the DNA/RNA strands

Transcription, the synthesis of RNA from a DNA template, is one of the most important steps in the control of cell growth and differentiation (24,25). During transcription, the information in a strand of DNA is copied into a new molecule of mRNA, mainly carried out by the enzyme RNA polymerase II and a number of accessory proteins called transcription factors (26). The RNA polymerase begins mRNA synthesis by unwinding the DNA helix and adding complementary bases to the RNA strand (elongation). The substrates for RNA synthesis are the four nucleoside triphosphates ATP, GTP, CTP and UTP. Cleavage of the high-energy phosphate bond between the phosphate groups in the nucleoside triphosphate structure provides the necessary energy for the addition of nucleotides into the growing RNA chain (27). Wang *et al.* obtained the X-ray

structure revealing the transcribing complex in the ‘post-translocation’ state with the nucleotide added to the RNA transcript (28). Figure 3C shows how PLIP analyses such a complex and characterises the binding of GTP (orange) to the RNA polymerase (purple grey), RNA strand (green) and DNA strand (blue), all at the same time. The guanine group in GTP binds to the DNA strand with three hydrogen bonds and to the RNA strand via parallel π -stacking and one hydrogen bond. Moreover, the phosphate groups in GTP bind to the RNA polymerase residues via a salt bridge and a metal complex. The characterisation of the GTP binding mode in the elongation complex provides a structural understanding of the transcription mechanism.

PLIP algorithm for DNA/RNA detection

PLIP uses four steps to detect and report the relevant interactions of a complex: structure preparation, functional characterisation and rule-based matching and filtering of interactions. In the preparation step, the input structure is hydrogenated and the ligands are extracted along with their binding sites. If the advanced option for analysing DNA or RNA as receptor is selected, PLIP will automatically detect the nucleic acid residues as part of the receptor and exclude them from the ligand molecules. The RNA and DNA residues to be considered are detected by their name U, A, C, G and DT, DA, DC, DG, respectively. After the correct detection of receptor and ligands, PLIP continues with the subsequent interaction detection and visualisation as previously described in (2).

PLIP AND OTHER TOOLS

There are a number of tools (29–37) analysing ligand–target interactions, but apart from one, they have a different focus and therefore suffer from some limitations: (29–33) detect only a limited set of interaction types and do not offer 3D visualisation, (34) requires extensive preparation of input files, (38) is only commercially available and (35,36) are limited to binary interaction fingerprints and do not cover DNA and RNA. Arpeggio (37) is most closely related to PLIP offering comparable functionality except for flexible adjustment of thresholds.

CONCLUSION

Interactions between nucleic acids and small molecules play a crucial role in various biological processes and have a strong impact on drug discovery. In that regard, the PLIP web tool has been adapted to recognise nucleic acid molecules as receptors and provide a comprehensive analysis and visualisation of non-covalent interactions with a one-click loading of structures. Moreover, its novel design displays a user-friendly platform, which facilitates the intuitive use of the tool’s features and exploration of the results for a better understanding of the biological processes behind the binding mechanism. Furthermore, the availability of PLIP source code enables local batch processing, customisation of the algorithm for special applications as well as active development of the tool in the community.

ACKNOWLEDGEMENTS

The authors would like to thank Sven Schreiber, Alexandre Mestiashvili and Christoph Leberecht for their technical support and valuable feedback.

FUNDING

Funding for open access charge: Technische Universität Dresden.

Conflict of interest statement. None declared.

REFERENCES

- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Salentin, S., Schreiber, S., Haupt, V.J., Adasme, M.F. and Schroeder, M. (2015) PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Res.*, **43**, W443–W447.
- Elfiky, A.A. (2020) Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. *Life Sci.*, **248**, 117477.
- Furlan, V., Konc, J. and Bren, U. (2018) Inverse molecular docking as a novel approach to study anticarcinogenic and anti-neuroinflammatory effects of curcumin. *Molecules*, **23**, 3351.
- Soliman, C., Yuriev, E. and Ramsland, P.A. (2017) Antibody recognition of aberrant glycosylation on the surface of cancer cells. *Curr. Opin. Struct. Biol.*, **44**, 1–8.
- Kumar, K., Woo, S.M., Siu, T., Cortopassi, W.A., Duarte, F. and Paton, R.S. (2018) Cation– π interactions in protein–ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.*, **9**, 2655–2665.
- Dai, J., PUNCHIHEWA, C., Mistry, P., Ooi, A.T. and Yang, D. (2004) Novel DNA bis-intercalation by MLN944, a potent clinical bisphenazine anticancer drug. *J. Biol. Chem.*, **279**, 46096–46103.
- Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovskiy, A., Poser, I., Bickle, M., Rizk, S., Guill’en-Boixet, J., Franzmann, T.M. *et al.* (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science*, **360**, 918–921.
- Bergeron-Sandoval, L.-P., Safaee, N. and Michnick, S.W. (2016) Mechanisms and consequences of macromolecular phase separation. *Cell*, **165**, 1067–1079.
- Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovskiy, A., Drechsel, D. *et al.* (2018) A molecular grammar governing the driving forces for phase separation of Prion-like RNA binding proteins. *Cell*, **174**, 688–699.
- Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L. *et al.* (2018) Protein phase separation: a new phase in cell biology. *Trends Cell Biol.*, **28**, 420–435.
- Ramaswami, M., Taylor, J.P. and Parker, R. (2013) Altered ribostasis: RNA–protein granules in degenerative disorders. *Cell*, **154**, 727–736.
- Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M. *et al.* (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell*, **162**, 1066–1077.
- Niaki, A.G., Sarkar, J., Cai, X., Rhine, K., Vidaurre, V., Guy, B., Hurst, M., Lee, J.C., Koh, H.R., Guo, L. *et al.* (2020) Loss of dynamic RNA interaction and aberrant phase separation induced by two distinct types of ALS/FTD-linked FUS mutations. *Mol. Cell*, **77**, 82–94.
- Loughlin, F.E., Lukavsky, P.J., Kazeeva, T., Reber, S., Hock, E.-M., Colombo, M., Von Schroetter, C., Pauli, P., Cléry, A., Méhlemann, O. *et al.* (2019) The solution structure of FUS bound to RNA reveals a bipartite mode of RNA recognition with both sequence and shape specificity. *Mol. Cell*, **73**, 490–504.
- Harris, S., Mistry, P., Freathy, C., Brown, J. and Charlton, P. (2005) Antitumour activity of XR5944 in vitro and in vivo in combination with 5-fluorouracil and irinotecan in colon cancer cell lines. *Brit. J. Cancer*, **92**, 722–728.

17. Harris,S., Scott,J., Brown,J., Charlton,P. and Mistry,P. (2005) Preclinical anti-tumor activity of XR5944 in combination with carboplatin or doxorubicin in non-small-cell lung carcinoma. *Anti-cancer Drug*, **16**, 945–951.
18. Lin,C., Mathad,R.I., Zhang,Z., Sidell,N. and Yang,D. (2014) Solution structure of a 2:1 complex of anticancer drug XR5944 with TFF1 estrogen response element: insights into DNA recognition by a bis-intercalator. *Nucleic Acids Res.*, **42**, 6012–6024.
19. Stewart,A., Mistry,P., Dangerfield,W., Bootle,D., Baker,M., Kofler,B., Okiji,S., Baguley,B., Denny,W. and Charlton,P. (2001) Antitumor activity of XR5944, a novel and potent topoisomerase poison. *Anti-cancer Drug*, **12**, 359–367.
20. Sappal,D.S., McClendon,A.K., Fleming,J.A., Thoroddsen,V., Connolly,K., Reimer,C., Blackman,R.K., Bulawa,C.E., Osheroff,N., Charlton,P. *et al.* (2004) Biological characterization of MLN944: a potent DNA binding agent. *Mol. Cancer Ther.*, **3**, 47–58.
21. Byers,S.A., Schafer,B., Sappal,D.S., Brown,J. and Price,D.H. (2005) The antiproliferative agent MLN944 preferentially inhibits transcription. *Mol. Cancer Ther.*, **4**, 1260–1267.
22. Punchihewa,C., De Alba,A., Sidell,N. and Yang,D. (2007) XR5944: a potent inhibitor of estrogen receptors. *Mol. Cancer Ther.*, **6**, 213–219.
23. Sidell,N., Mathad,R.I., Jue Shu,F., Zhang,Z., Kallen,C.B. and Yang,D. (2011) Intercalation of XR5944 with the estrogen response element is modulated by the tri-nucleotide spacer sequence between half-sites. *J. Steroid. Biochem.*, **124**, 121–127.
24. Young,R.A. (2011) Control of the embryonic stem cell state. *Cell*, **144**, 940–954.
25. Lee,T.I. and Young,R.A. (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.
26. Hahn,S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.*, **11**, 394–403.
27. Alberts,B., Johnson,A., Lewis,J., Raff,M., Roberts,K. and Walter,P. (2002) In: *Molecular Biology of the Cell*. 4th edn, Garland Science, NY.
28. Wang,D., Bushnell,D.A., Westover,K.D., Kaplan,C.D. and Kornberg,R.D. (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*, **127**, 941–954.
29. Laskowski,R.A. and Swindells,M.B. (2011) LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.*, **51**, 2778–2786.
30. Clark,A.M. and Labute,P. (2007) 2D depiction of protein–ligand complexes. *J. Chem. Inf. Model.*, **47**, 1933–1944.
31. Stierand,K. and Rarey,M. (2010) Drawing the PDB: protein–ligand complexes in two dimensions. *ACS Med. Chem. Lett.*, **1**, 540–545.
32. de Beer,T. A.P., Berka,K., Thornton,J.M. and Laskowski,R.A. (2013) PDBsum additions. *Nucleic Acids Res.*, **42**, D292–D296.
33. Gallina,A.M., Bisignano,P., Bergamino,M. and Bordo,D. (2013) PLI: a web-based tool for the comparison of protein–ligand interactions observed on PDB structures. *Bioinformatics*, **29**, 395–397.
34. Durrant,J.D. and McCammon,J.A. (2011) BINANA: a novel algorithm for ligand-binding characterization. *J. Mol. Graph. Model.*, **29**, 888–893.
35. Marcou,G. and Rognan,D. (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.*, **47**, 195–207.
36. Radifar,M., Yuniarti,N. and Istyastono,E.P. (2013) PyPLIF: Python-based protein–ligand interaction fingerprinting. *Bioinformatics*, **9**, 325–328.
37. Jubb,H.C., Higuero,A.P., Ochoa-Montano,B., Pitt,W.R., Ascher,D.B. and Blundell,T.L. (2017) Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.*, **429**, 365–371.
38. Weisel,M., Bitter,H.-M., Diederich,F., So,W.V. and Kondru,R. (2012) Prolix: rapid mining of protein–ligand interactions in large crystal structure databases. *J. Chem. Inf. Model.*, **52**, 1450–1461.