STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Received 30 May 2024
Accepted 19 February 2025

Edited by J. Agirre, University of York, United Kingdom

This article is part of a collection of articles from the IUCr 2023 Congress in Melbourne, Australia, and commemorates the 75th anniversary of the IUCr.

Keywords: natural products; high-throughput crystallography; crystallographic capture; drug discovery.

PDB references: yeast 20S proteasome, soaked with MA9 crude extract, 9aw3; soaked with MA9 fraction E/F, 9aw5; soaked with MA9 fraction EF2, 9aw6; soaked with TMC-95B, 9aw7; soaked with BRA-346 fraction, 9c97; soaked with purified TMC-86A, 9c98

Supporting information: this article has supporting information at journals.iucr.org/f

High-throughput protein crystallography to empower natural product-based drug discovery

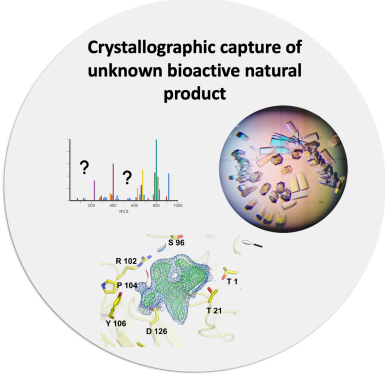
Raphael Meneghello,[a] Joane K. Rustiguel,[a] Evandro Ares de Araújo,[b] Rafael de Felício,[a] Arthur Zanetti N. Fernandes,[a] Everton L. F. Ferreira,[c] Juliana R. Gubiani,[c] Agnes A. S. Takeda,[a] Amanda Araujo,[a,d] Caio C. de Lima Silva,[a] Ariane F. Bertonha,[a] Raquel P. M. Urano,[a] Daniel M. Trindade,[a] Thiago M. Cunha,[d] Alisson C. Cardoso,[a] Roberto G. S. Berlinck,[c] Andrey F. Ziem Nascimento[b]* and Daniela B. B. Trivella[a]*

[a]Brazilian Biosciences National Laboratory (LNBio), Brazilian Centre for Research in Energy and Materials (CNPEM), Campinas, SP 13083-970, Brazil, [b]Brazilian Synchrotron Light Source National Laboratory (LNLS), Brazilian Centre for Research in Energy and Materials (CNPEM), Campinas, SP 13083-970, Brazil, [c]Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, SP 13560-970, Brazil, and [d]Centre of Research in Inflammatory Diseases, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP 14040-970, Brazil. *Correspondence e-mail: andrey.nascimento@lnls.br, daniela.trivella@lnbio.cnpem.br

Nature is a rich and largely untapped reservoir of small molecules, the latter historically being the main source of new drugs. Three-dimensional structures of proteins in complex with small-molecule ligands represent key information to progress drug-discovery projects, in particular in the hit-to-lead phase. High-throughput crystallography has been of extensive use in recent years, especially to obtain crystallographic complexes of synthetic ligands and fragments. However, the process of discovering novel bioactive natural products has experienced limitations that have long prevented large drug-discovery programs using this outstanding source of molecules. Recent technologies have contributed to the re-emergence of natural products in modern drug discovery. We present the use of high-throughput protein crystallography to directly capture bioactive natural products from unpurified biota chemical samples using protein crystals. These routines, which are currently in use at the Brazilian Centre for Research in Energy and Materials (CNPEM), are introduced with a description of crystal preparation, automated data collection and processing at the MANACÁ beamline (Sirius, LNLS, CNPEM), along with case examples of bioactive natural product capture using protein crystals. The usefulness of this pipeline, which accelerates the discovery and structural elucidation of both known and previously unknown bioactive natural products, paves the way for the development of innovative therapeutic agents, thus contributing to the new era of natural product-based drug discovery.

1. Introduction

Natural products have been used since ancient times (Leonti & Casu, 2013; Zank & Hanazaki, 2017) as primary sources of innovative molecules, inspiring the development of new drugs (Newman & Cragg, 2020; Wilson *et al.*, 2020). Historically, plants, and more recently microorganisms (van Santen *et al.*, 2022), have emerged as exceptional producers of diverse sets of small molecules. Such chemical diversity is hardly achieved by humankind due to our finite capacity to design and produce such molecules, compared with the millions of years of natural evolution. Chemical diversity is often related to biodiversity, as different species can produce different molecules. Environmental pressures also drive the evolution and production of molecules by living organisms, especially as defense and communication tools in each microenvironment. These factors contribute to the impressive chemical diversity of natural products (Atanasov *et al.*, 2021; Bruder *et al.*, 2020; Grigalunas *et al.*, 2022). Consequently, megabiodiverse countries such as



Brazil (Keil & Chase, 2019; Sabatini *et al.*, 2022) are rich sources of chemical diversity (Valli *et al.*, 2018). Brazilian biodiversity shines as a key as yet untapped source of chemical diversity and bioactive molecules with great pharmacological potential. Despite efforts to catalog and compile Brazilian natural products (Pilon *et al.*, 2017), many novel bioactive compounds remain to be discovered in Brazil and other hotspots of biodiversity around the globe.

In recent years, drug-discovery pipelines have become increasingly reliant on innovative and automated solutions, including high-throughput assays (target-based and/or cell-based screenings; Macarron *et al.*, 2011), synchrotron-based (Thomas *et al.*, 2019) and electron microscopy-based (Kühlbrandt, 2014; Robertson *et al.*, 2022) structural biology, and artificial intelligence (Wong *et al.*, 2024). Successful results have been achieved using structure-based drug design (SBDD), with examples already in the clinic (Sabe *et al.*, 2021; Talele *et al.*, 2010). More recently, artificial intelligence has entered the scene, supporting the design and selection of novel drug leads starting from novel protein–ligand complexes (Leit *et al.*, 2023; Sadybekov & Katritch, 2023). High-throughput (HT) crystallography (Blundell *et al.*, 2002) plays an important role in the new era of SBDD, as hundreds of crystals can be collected within a single shift (∼12 h), quickly delivering novel protein–ligand complexes to support innovative SBDD (Bancet *et al.*, 2020; Boby *et al.*, 2023; Fan *et al.*, 2020; Günther *et al.*, 2021; Harriman *et al.*, 2016; Mitcheltree *et al.*, 2021; Schiebel *et al.*, 2016). Fragment screening has also evolved in this sense, with six drugs already on the market (Bon *et al.*, 2022; Erlanson *et al.*, 2016).

However, the integration between these cutting-edge technologies and natural products has not yet reached its full potential. Natural products are still underutilized as samples in initial high-throughput screening (Wilson *et al.*, 2020), even though the chemical space accessed by natural product libraries is much larger compared with typical libraries of synthetic molecules (Bruder *et al.*, 2020; Lachance *et al.*, 2012; Stone *et al.*, 2022). Despite slower progress, many advancements have been made in bioactive natural product-based research, including automated liquid chromatography coupled to tandem mass spectrometry-based (LC-MS/MS) untargeted metabolomics (Wang *et al.*, 2016), (meta)genome sequencing and mining (Blin *et al.*, 2021), natural product purification, (bio)synthetic production of natural molecules (Huang *et al.*, 2023) and derivatives designed in one step and through faster routes (Teufel *et al.*, 2014). These advances have pushed the re-emergence of natural product-based drug discovery (Berlinck *et al.*, 2019). Still, the main bottleneck in natural product-based SBDD is the speed of obtaining protein–ligand complexes, which cannot rely on natural product isolation and is the longest and most resource-consuming step in bioactive natural product drug discovery (Beutler, 2009).

Aiming to quickly obtain structural information on the bioactive natural product molecule interacting with its biological target, a similar approach to that presented in fragment-based drug discovery (Huang *et al.*, 2024; Kirsch *et al.*, 2019) can be used. In this sense, unpurified natural product samples (as crude extracts and enriched fractions) can be rapidly screened almost directly from biota samples of plants and microrganisms using pre-fractionated natural product chemical libraries (de Felício *et al.*, 2021; Thornburg *et al.*, 2018; Wagenaar, 2008; Wilson *et al.*, 2020). Once screened in bioassays, the chemical samples found to be hits (the bioactive biota samples, still in complex mixtures) can be directly soaked with the target protein crystal. In this process, the crystal itself acts as an affinity phase, trapping only molecules with an affinity for a specific binding site as they pass through the solvent channels of the crystal. This approach, which was initially called 'affinity crystallography' (Aguda *et al.*, 2016) and subsequently 'crystallographic capture', has demonstrated outstanding results in our hands, being able to capture different classes of bioactive natural products in a myriad of protein types of relevant therapeutic value. This is true for both catalytic canonical sites (orthosteric sites) and novel, sometimes previously unknown, allosteric binding sites.

This work introduces the HT crystallographic capture approach for natural products, emphasizing HT crystallography experiments conducted on the MANACÁ beamline at Sirius (LNLS, CNPEM, Brazil), automated data processing at CNPEM and use of the *Pan-Dataset Density Analysis* (*PanDDA*) pipeline to highlight the presence of captured ligands even at low occupancy in the crystal structure (Pearce, Krojer, Bradley *et al.*, 2017). The *PanDDA* approach is being extended to natural product ligand capture in response to the recent availability of pre-fractionated natural product libraries (Trivella *et al.*, 2022) and the growing need for the discovery of bioactive molecules from biodiversity. The HT crystallographic capture approach can be further combined with recent advances in LC-MS/MS untargeted metabolomics dereplication and bioactive compound ranking (Baskiyar *et al.*, 2022; Bazzano *et al.*, 2024; Nothias *et al.*, 2018). With these advancements, a new era of bioactive small-molecule discovery is envisioned to push the boundaries of natural product-based drug discovery in the 21st century.

2. Materials and methods

2.1. MANACÁ beamline instrumentation

The MANACÁ (MAcromolecular micro and NAno CrystAllography) beamline at Sirius (LNLS, CNPEM, Brazil) is a fourth-generation synchrotron beamline with high photon flux, micro-size and small beam divergence (0.44 mrad). The beamline optics utilize an undulator source (Kyma) and a double-crystal monochromator (Si111 or Si311). This optical configuration incorporates both vertical and horizontal focusing mirrors, facilitating a versatile energy range from 5.6 to 20 keV. Additionally, the beam size is adjustable, ranging from 100 × 80 to 20 × 20 µm (horizontal × vertical). The standard beam size at the sample position is 20 × 20 µm, with a photon flux of 5×10^{11} photons/s/100 mA at 12.7 keV.

The experimental hutch of MANACÁ has a diffractometer base equipped with a horizontal air-bearing goniometer (in-house-developed) and an OAV microscope (BZOOM, Arinax). The beam can be attenuated in steps of at least 1%

over most of the energy range. The PILATUS 2M detector (Dectris) is mounted on a separate base and can be positioned at a distance of 85–1000 mm from the sample. These beam-condition elements allow the user to optimize the setup to collect the best data with regard to resolution and radiation damage.

The automatic sample changer was designed in-house based on a TX60 six-axis robotic arm (Stäubli) and a sample dewar with capacity of three UniPucks (48 samples). This allows data collection from more than 200 crystals per day. All of the beamline user operation, on site or remotely, is performed using the *MXCuBE* software (Gabadinho *et al.*, 2010; Oscarsson *et al.*, 2019).

2.2. MANACÁ data collection and processing

The MANACÁ beamline currently offers conventional cryogenic data collection (Nascimento *et al.*, 2021) and a new room-temperature data-collection mode (to be published). An in-house automatic data-processing pipeline (*MNCAutoProc*) was developed in Python to process the data with minimal user input. The current version includes data-reduction and phasing options (molecular replacement, anomalous phasing and rigid-body refinement). The pipeline is based on the *XDS*, *CCP*4 and *Phenix* packages (Agirre *et al.*, 2023; Kabsch, 2010; Liebschner *et al.*, 2019). It uses *XDS* and *XSCALE* for data reduction and merging (Kabsch, 2010), *Phaser* for molecular replacement (McCoy *et al.*, 2007), *SHELXC*/*D*/*E* for anomalous phasing (Usón & Sheldrick, 2018) and *phenix.refine* for data refinement (Afonine *et al.*, 2012). Immediately after data collection, diffraction images are automatically processed without any user input. Two sets of data are generated: one without any cutoff and another cut based on $CC_{1/2}$ (Karplus & Diederichs, 2012). All of the data processing is performed using a dedicated queue with 200 threads at the high-performance computing (HPC) facility at Sirius.

All of these setups and implementations, highlighting the sample changer and data-processing pipeline, enable high-content diffraction experiments, such as crystallographic ligand screening, which were impossible on the former beamline of the Brazilian Synchrotron.

2.3. Automated data preparation for high-throughput ligand-map identification at CNPEM

Four custom scripts were developed to streamline data preparation for *PanDDA*. After the crystallographic data have been processed using *MNCAutoProc*, the first script (`01_hkl_uncut_copy.sh`) copies the uncut HKL file of each collected data set from the crystallographic data-collection folder into a new folder named `hkl_uncut`. The HKL files are then automatically renamed to a standardized format using the second script (`02_rename.sh`). The third script (`03_aimLess_parallel.sh`) generates merged MTZ files by scaling all renamed HKL files, performing a resolution cut at 1.5 $I/\sigma(I)$ and truncating them using *AIMLESS* (Evans & Murshudov, 2013). Finally, the fourth script (`04_data_preparation.sh`) uses a well refined PDB model and each processed MTZ file to run automated molecular replacement and refinement using *DIMPLE* (Wojdyr *et al.*, 2013). The refined reflection data have their missing reflections populated to achieve 100% completeness using *UNIQUEIFY*, *CAD* (Agirre *et al.*, 2023) and *phenix.maps* (Liebschner *et al.*, 2019). The last script ends by copying the refined/populated MTZ and the refined PDB files into a folder to be used by *PanDDA* (Pearce, Krojer & von Delft, 2017) as input files. The last two scripts run in a multithread manner by using the GNU parallel package (Tange, 2018). The scripts are available as supporting information.

Initial data analysis and ligand-map identification is performed by the first *PanDDA* step, *pandda.analyze*. Event-map visualization is performed using a customized *Coot* (Emsley & Cowtan, 2004) window inside the *pandda.inspect* step.

2.4. TMC case study

2.4.1. MA9 crude extract preparation, derived fractions and TMC-95B isolation

The MA9 strain was isolated from the plant *Anthurium loefgrenii* and identified as the fungus *Annulohypoxylon moriforme* (MA9) by Professor Andre Rodrigues from the Universidade Estadual Paulista (UNESP, Rio Claro, SP) based on morphological and molecular data.

The strain MA9 was grown and its compounds were extracted and partially fractionated, allowing biological screening and application of the crystallographic capture approach. The bioactive compounds were isolated to confirm their proteasome inhibition and chemical structures (Bazzano *et al.*, 2024). Briefly, the MA9 strain was inoculated in potato dextrose broth (PDB) medium (10 l) in 500 ml Schotts bottles with 250 ml medium for 14 days at room temperature (28°C) in static mode. For extraction, ethyl acetate was added to the culture medium (3×) and the organic and aqueous phases were separated. The organic phase was concentrated under low pressure at 35°C and the dried mass was dissolved in 95% methanol/H_2O and submitted to liquid–liquid partition with *n*-hexane (3×).

The methanolic final extract (882 mg, MA9 crude extract) was fractioned on a reverse-phase C18 column in a stepwise gradient of H_2O:methanol, 9:1 (A), 8:2 (B), 7:3 (C), 6:4 (D), 1:1 (E), 3:7 (F) and 0:1 (G), totalling seven fractions (A–G). Fractions MA9-11 (E) and MA9-37 (F) were grouped (fraction E/F) and subjected to a new fractionation on a C18 column using a stepwise gradient of H_2O/methanol, 7:3 (EF1), 6:4 (EF2), 1:1 (EF3), 4:6 (EF4), 2:8 (EF5) and 0:1 (EF6), totalling six subfractions (EF1–EF6).

Fractions EF2–EF4 were grouped for separation by HPLC (C18 InertSustain 5 µm column, 4.6 × 250 mm, GL Sciences) with a gradient of 79:13:08 to 20:48:32 H_2O:acetonitrile:methanol over 45 min (flow rate 2.5 ml min^{-1}), yielding a mixture of two peaks (C1 and C2). A new purification procedure (C18 X-Terra 5 µm column, 4.6 × 250 mm, isocratic 80:10:10 H_2O:acetonitrile:methanol hold for 25 min, 55:25:20 at 25–35 min, flow rate 2.0 ml min^{-1}) resulted in the purified

fraction C2 (TMC-95A/B, 1.9 mg, total yield of 0.2%). This procedure was repeated once to accumulate mass for NMR studies, starting with 18 l of fungal culture, resulting in equivalent yields of TMC-95A/B. Compounds were identified by ^{1}H and ^{13}C, HMQC and gCOSY NMR measurements (Bazzano *et al.*, 2024). All fractions were submitted to proteasome-inhibition assays and analyzed by UPLC-PDA-MS/MS to select bioactive and similar samples, respectively.

2.4.2. BRA-346 crude extract preparation, derived fractions and TMC-86A isolation

Streptomyces sp. BRA-346 was isolated from the tunicate *Euherdmania* sp. (Bazzano *et al.*, 2024; Domingues Vieira *et al.*, 2022; Furtado *et al.*, 2021). As for MA9, the BRA-346 metabolic crude extract and initial fractions were obtained and subjected to proteasome inhibition and to the crystallographic capture approach. Briefly, *Streptomyces* sp. BRA-346 was grown in A1 medium supplemented with mycelial stock for 48 h. 1 ml seed culture was then used to inoculate 100 ml A1 medium. The final culture was grown for seven days and extract by ethyl acetate liquid–liquid partition. The resulting biomass was filtered and phase-separated with ethyl acetate. The organic phase was dried and the crude extract was subjected to solid-phase separation. Crude extract and the resulting fractions were dried and resuspended in DMSO to a final concentration of 10 mg ml^{-1} for biochemical assays.

To isolate TMC-86A, heterologous *S. coelicolor* M1146 *epn*/*tmc* BRA-346 was cultivated in A1 medium followed by acidification and extraction with ethyl acetate. The organic phase was filtered, dried and subjected to reverse-phase HPLC. The bioactive component with *m*/*z* 343 was recovered, dried and purified by UPLC. The chemical structure of the bioactive molecule was determined by NMR analyses (Bazzano *et al.*, 2024; Domingues Vieira *et al.*, 2022).

2.4.3. Proteasome-inhibition assays

Proteasome-inhibition assays were conducted as described by Furtado *et al.* (2021). Each experiment was conducted in triplicate. Results are expressed as percentage of the enzyme activity obtained against the controls [positive controls for enzyme activity (100%), vehicle DMSO; negative control for enzyme activity (0%), reaction carried out in the absence of the enzyme]. For IC_{50} curves, a serial dilution of the chemical samples was used. Data were plotted and fitted with the normalized response four-parameter equation implemented in *GraphPad Prism* (GraphPad, San Diego, USA).

2.4.4. Yeast 20S purification and crystallization

Yeast proteasome purification was carried out as described previously (De Souza *et al.*, 2018). Briefly, *S. cerevisiae* lysates were purified by ion-exchange and size-exclusion chromatography, with 25 m*M* Tris pH 7.4, 150 m*M* NaCl, 10 m*M* $MgCl_2$, 1 m*M* DTT as the final buffer. The purified 20S proteasome was concentrated to 20–40 mg ml^{-1} and submitted to crystallization trials using 100 m*M* MES pH 5.5–6.9, 8–18% MPD and 20 m*M* magnesium acetate as the crystallization solution. The hanging-drop method was used with VDX48 plates (Hampton Research) filled with 200 µl crystallization buffer in the wells and drops of 1.4 µl (1:1 protein solution:crystallization buffer).

2.4.5. Soaking of proteasome crystals, data collection, processing and refinement

Crystals were soaked overnight (18°C) with BRA-346 and MA9 samples: BRA-346 enriched fraction and TMC-86A, and MA9 crude extract (fractions E/F and EF2) and TMC-95B, each at 2 mg ml^{-1}, with a final solvent (DMSO) concentration of 10%. The soaked crystals were harvested in nylon loops (Hampton Research), quickly passed through cryoprotectant solution (100 m*M* Tris pH 7.0, 20 m*M* magnesium acetate, 30% MPD) and flash-cooled in liquid nitrogen. X-ray diffraction data sets were acquired at 100 K using the rotation method on the W01B-MX2 beamline at UVX, LNLS, CNPEM and Sirius, LNLS, CNPEM. Data reduction, phase recovery, *PanDDA* protocol and refinement were applied as described in Section 2.3.

2.4.6. LC-MS/MS data acquisition

The prepared chemical samples were analyzed by liquid chromatography coupled to mass spectrometry using an Acquity UPLC H-Class system (Waters) coupled to an ESI-qQToF Impact II mass spectrometer (Bruker). For sample separation, a BEH C18 analytical column (1.7 µm, 2.1 × 100 mm, Acquity, Waters) was used with a flow rate of 0.5 ml min^{-1}, a column temperature of 40°C and a sample temperature of 20°C. All analyses were carried out in positive mode in the range 30–2000 Da with an acquisition rate of 8 Hz. Ion-source chamber parameters were 500 V end-plate offset, capillary at 4500 V, nebulizer at 4.0 bar and drying gas flow (nitrogen) at 10 l min^{-1} with a drying temperature of 200°C. For MS/MS, the collision cell was 5.0 eV, with collision energy in the range 20–70 V and an absolute fragmentation cutoff of 1000. As a fragmentation rule, ions with *m*/*z* below 200 were excluded and the 'active exclusion' function was enabled (precursor ions with more than three spectra had their fragmentation blocked for 0.3 min or until the current intensity:previous intensity ratio was greater than or equal to 1.8). In each elution, 10 m*M* sodium formate was used for internal calibration.

3. Results and discussion

3.1. Pipeline for the crystallographic capture of unknown bioactive natural products

The CNPEM pipeline for capturing natural products using HT crystallography is summarized in Fig. 1. Using pre-fractionated natural product libraries (de Felício *et al.*, 2021; Trivella *et al.*, 2022), high-throughput screening methods (typically enzyme-inhibition assays) are employed to identify natural samples with bioactivity in proteins (activators or inhibitors). Hit samples are confirmed using the primary assay in concentration–response curves. LC-MS/MS analyses are

performed for all samples, aiming to characterize their chemical contents (Bazzano *et al.*, 2024). The confirmed hits (crude extracts or enriched fractions of natural products, which are still complex mixtures of unknown compounds) are directly soaked with the target protein crystals, aiming to capture the bioactive compound from the natural sample and directly characterize its binding to the targeted protein.

Typical stock concentrations used for soaking natural product bioactive samples with their target protein crystals start at 100 mg ml^{-1} and can be adjusted based on crystal stability. The final concentration in the crystal drops ranges from 10 to 0.1 mg ml^{-1}. Recent upgrades to automated soaking with an Echo 650 liquid handler (Beckman) at the Compound Management Laboratory (LGC, LNBio, CNPEM, Campinas, SP, Brazil) allow the soaking of a several hundred crystals within minutes. Crystals are incubated for 4–48 h at controlled temperature (4 or 18°C). Three or four crystals soaked with each bioactive natural product sample are then selected, harvested and cryocooled for data collection.

Conventional cryogenic data-collection experiments were conducted on the MANACÁ beamline. Typically, each data-collection run involved mounting 48 cryocooled crystal samples on the UniPuck devices, which were subsequently stored under liquid nitrogen within the beamline sample dewar. The samples were mounted onto the goniometer head using the six-axis robotic arm installed at the experimental hutch. Crystal harvesting, alignment and data collection were controlled through the *MXCuBE*3 web interface.

Data acquisition occurs at a cryogenic temperature of 100 K, with a monochromatic X-ray beam typically set to an energy of 12.688 keV. Full data sets generally cover a 360° oscillation range to guarantee data completeness and redundancy, with a minimum angular step of 0.1° to provide fine-slicing (Dauter, 2017; Mueller *et al.*, 2012). A complete data-set collection usually takes up to 6 min using 0.07 s of X-ray exposure per image.

Automated HT crystallographic data processing was designed to include three modules as described in Sections 2.2 and 2.3: the automated data-processing module *MNCAutoProc* (on the fly), a data-preparation module comprising four scripts and the automated event map obtained using *PanDDA* (Pearce, Krojer & von Delft, 2017). After data collection at the MANACÁ beamline, the complete CNPEM pipeline for HT crystallographic data processing can be run within a day for 100 natural product-soaked crystallographic data sets. The data-preparation module takes about a couple of hours to complete using a personal computer (eight-core processor and 16 GB RAM). The use of high-performance computing (HPC) is recommended to run the *PanDDA* module. At CNPEM, the *PanDDA* module takes 12–18 h to complete map calculation for 100 data sets using a cluster with a dedicated queue with 150 threads and 4 TB RAM available at the Throughput Enhanced Processing Unit (TEPUI), Computing Platforms (COMP), Sirius, LNLS, CNPEM.

A set of crystallographic ground-state data sets must be provided by the user to allow the calculation of *PanDDA* maps. Conventionally, to attain a robust average ground-state map and facilitate the extraction of *PanDDA* maps, the collection of 40 apo data sets is recommended (Pearce, Krojer, Bradley *et al.*, 2017). These data sets are acquired under comparable conditions and in the same space group as the soaked data sets. This extensive set is deemed to be necessary to ensure a representative sampling of the ground state, enhancing the statistical robustness and reliability of the resultant maps (Pearce, Krojer & von Delft, 2017).

Upon completion of the entire data processing, the user can inspect the resulting event maps using *Coot* and an internal *PanDDA* navigation tool to directly focus the structure on the event blob.

3.2. Case examples: capture of TMC-86A (*Streptomyces* sp. BRA-346) and TMC-95A/B (*A. moriforme* MA9) into yeast 20S proteasome crystals

We selected MA9 and BRA-346 as representative case studies to demonstrate the application of our natural product crystallographic capture methodology.

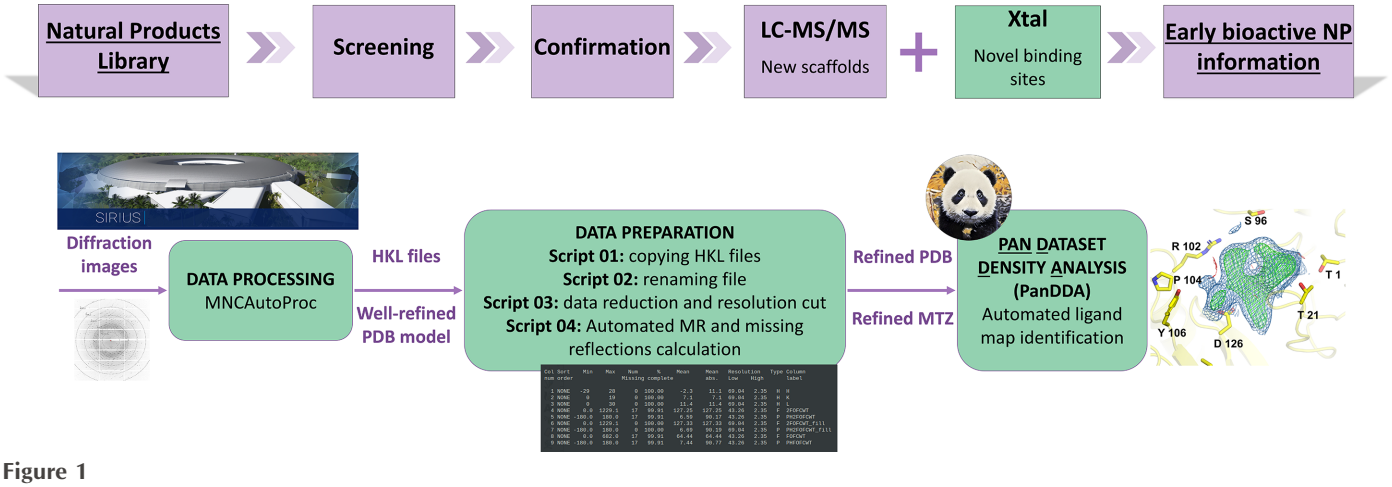


Figure 1
Pipeline for crystallographic capture of unknown bioactive natural products. The CNPEM pipeline for HT crystallographic data processing is shown in the bottom panel. It comprises three modules (green boxes), which include *MNCAutoProc* data processing, data preparation and automated ligand-map identification with *PanDDA*. Input files for each module are described (purple).

Table 1
Data collection and processing.

Values in parentheses are for the outer shell.

	BRA-346 enriched fraction F48	TMC-86A	MA9 extract	MA9 fraction E/F	MA9 fraction EF2	TMC-95B
Diffraction source	Sirius, LNLS	Sirius, LNLS	UVX, LNLS	UVX, LNLS	UVX, LNLS	UVX, LNLS
Wavelength (Å)	0.9772	0.9772	1.458	1.458	1.458	1.458
Temperature (K)	100	100	100	100	100	100
Detector	PILATUS 2M	PILATUS 2M	PILATUS 2M	PILATUS 2M	PILATUS 2M	PILATUS 2M
Crystal-to-detector distance (mm)	250	250	180	217	272	216
Oscillation range per image (°)	0.1	0.1	0.2	0.5	0.2	0.1
Total rotation range (°)	360	360	158	180	349	360
Exposure time per image (s)	0.1	0.1	10	5	10	1.5
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a, b, c (Å)	117.51, 299.92, 143.91	117.31, 298.28, 144.57	131.96, 301.46, 142.65	134.89, 299.29, 144.21	135.37, 299.67, 144.66	134.62, 299.91, 144.66
α, β, γ (°)	90, 108.43, 90	90, 108.79, 90	90, 111.98, 90	90, 112.78, 90	90, 112.89, 90	90, 112.6, 90
Mosaicity (°)	0.13	0.10	0.12	0.57	0.29	0.14
Resolution range (Å)	49.11–3.33 (3.39–3.33)	49.19–3.04 (3.09–3.04)	29.84–3.42 (3.54–3.42)	48.80–3.44 (3.50–3.44)	49.56–3.43 (3.49–3.43)	30.04–2.91 (3.01–2.91)
Total No. of reflections	394217 (19403)	1215367 (61188)	418357 (18869)	410932 (15789)	930824 (46154)	1591363 (79070)
No. of unique reflections	132592 (6605)	179850 (8943)	136674 (6705)	135780 (6315)	140402 (6840)	228968 (11282)
Completeness (%)	96.5 (97.4)	99.9 (100)	98.4 (97.3)	97.6 (92.0)	99.3 (98.6)	99.2 (98.5)
Multiplicity	3.0 (2.9)	6.8 (6.8)	3.1 (2.8)	3.0 (2.5)	6.6 (6.7)	7.0 (7.0)
$\langle I/\sigma(I)\rangle$	3.0 (1.1)	6.6 (1.5)	7.2 (1.6)	6.6 (1.6)	6.8 (1.6)	7.4 (1.5)
R_{meas}	0.483 (0.973)	0.283 (1.509)	0.190 (0.988)	0.215 (0.669)	0.321 (1.285)	0.289 (1.647)
Overall B factor from Wilson plot (Å^2)	38.34	51.73	73.21	63.54	68.00	47.38

The MA9 samples were obtained from *A. moriforme* (MA9), an endophytic fungus isolated from the endemic plant *A. loefgrenii* found on the south-eastern coast of Brazil (Bazzano *et al.*, 2024). Initially, the MA9 crude extract (MA9 extract) and a series of C18 chromatographic fractions (A–G) were assayed against the proteasome (Bazzano *et al.*, 2024). The crude extract and fractions E and F demonstrated potent inhibition of the yeast 20S proteasome (IC_{50} values of 3971 ng ml^{-1} for the MA9 extract, 156 ng ml^{-1} for fraction E and 98 ng ml^{-1} for fraction F). Fractions E and F were combined (fraction E/F) and repurified, resulting in bioactive fractions EF2 and EF3 that exhibited enhanced inhibition (IC_{50} values of 2 ng ml^{-1} for EF2 and 49 ng ml^{-1} for EF3; Bazzano *et al.*, 2024). The MA9 bioactive natural product samples were then soaked against 20S proteasome core-particle crystals. For confirmation, the isolated TMC-95B was further soaked with proteasome crystals.

The BRA-346 samples were obtained from the actinobacterium *Streptomyces* sp. isolated from the Brazilian endemic tunicate *Euherdmania* sp.. Inhibition assays revealed that both the crude extract and the enriched fraction F48 of BRA-346 exhibited potent proteasome inhibition (Furtado *et al.*, 2021). The crude extract and the bioactive fraction F48 were individually soaked with proteasome crystals, aiming to capture the proteasome inhibitor that was present in these bioactive samples. In addition, the purified compound TMC-86A obtained after several steps of natural product purification from BRA346 cultures (Bazzano *et al.*, 2024) was also soaked with proteasome crystals for confirmation.

The MA9 X-ray diffraction data were collected at UVX, LNLS and the BRA-346 data were collected at Sirius, LNLS. Independently of the beamline used, the CNPEM automated processing pipeline was applied to both data sets. Data-collection and processing statistics are shown in Tables 1 and 2, respectively.

For *PanDDA* analysis, we used 26 refined structures for ground-state map characterization. Specifically for the proteasome, the *Z*-map characterization was restricted to the active site, using a mask size of 15 Å. This was necessary due to the large protein size of the 20S proteasome core particle.

3.2.1. *A. moriforme* (MA9) case study

The MA9 case study was systematically followed up as a proof of concept of the unpurified natural product capture approach, aiming at a comprehensive structural exploration of the natural product capture process at different stages. The MA9 data set (Fig. 2) revealed a natural ligand interacting in the proteasome catalytic site. The ligand is positioned in close proximity to the catalytic Thr1 residue, with no evidence of covalent interaction. The electron density relative to the ligand captured from the crude extract was partially defined, but became clearly delineated, probably due to higher occupancy, using the first fractionation step (fraction E/F). Each enriched fraction significantly contributed to the refinement of a more detailed difference map (Fig. 2*a*). Application of the *PanDDA* protocol corroborated these observations, clearly highlighting the presence of a natural ligand binding to the proteasome chymotrypsin-like catalytic site (Fig. 2*b*), even for fraction E/F, which was the first fractionation step.

Despite the partially defined electron density observed from the MA9 crude extract, important information was still obtained about the presence of a genuine protein ligand in the natural sample, including the binding site and binding mode of the bioactive natural product. On the other hand, using the enriched fraction E/F, much clearer electron density for the

Table 2
Structure solution and refinement statistics.

Values in parentheses are for the outer shell.

	BRA-346 enriched fraction F48	TMC-86A	MA9 extract	MA9 fraction E/F	MA9 fraction EF2	TMC-95B
PDB code	9c97	9c98	9aw3	9aw5	9aw6	9aw7
Resolution range (Å)	49.11–3.33 (3.39–3.33)	49.19–3.04 (3.09–3.04)	29.84–3.42 (3.54–3.42)	48.80–3.44 (3.50–3.44)	49.56–3.43 (3.49–3.43)	30.04–2.91 (3.01–2.91)
Completeness (%)	96.5 (97.4)	99.9 (100)	98.4 (97.3)	97.6 (92.0)	99.3 (98.6)	99.2 (98.5)
No. of reflections, working set	132030	179706	136333	135580	138909	228749
No. of reflections, test set	6630	8947	6917	6587	6749	11262
Final R_{cryst}	0.2003	0.1746	0.2148	0.2006	0.1817	0.1945
Final R_{free}	0.2520	0.2201	0.2594	0.2518	0.2357	0.2320
No. of non-H atoms						
Protein	49321	49387	49061	50274	49539	49373
Ligand	229	501	22	411	589	758
Water	316	337	177	332	319	252
Total	49866	50225	49260	50274	50447	50383
R.m.s. deviations						
Bond lengths (Å)	0.002	0.003	0.002	0.002	0.002	0.002
Angles (°)	0.48	0.59	0.48	0.49	0.54	0.54
Average B factors (Å^2)						
Overall	38.34	60.44	100.7	75.72	85.16	51.82
Protein	38.41	60.35	100.86	75.89	85.35	51.81
Ligand	57.9	77.95	100.20	83.59	89.86	57.19
Water	12.94	46.44	55.49	39.84	47.79	37.83
Ramachandran plot						
Most favored (%)	96.48	97.06	96.91	96.68	96.08	97.27
Allowed (%)	3.49	2.93	3.09	3.29	3.92	2.73
Outliers (%)	0.03	0.02	0	0.03	0	0

ligand was observed, allowing the interpretation of its chemical structure based on the shape of the electron density. The crystal structure obtained for the proteasome soaked with fraction E/F clearly shows the macrocycle ligand binding noncovalently to the proteasome active site (Fig. 2*b*).

Both the *m*/*z* of the bioactive compound (*m*/*z* 679.27; Fig. 2*c*) and a clear electron density were only obtained using the enriched fraction E/F. This highlights the value of using pre-fractionated natural product libraries in natural product-based drug-discovery pipelines. Pre-fractionation improves the detection of bioactivity, enhances LC-MS/MS sensitivity and increases captured ligand occupancies in crystal structures. Therefore, pre-fractionated natural product libraries are being used more frequently in natural product-based drug-discovery programs (de Felício *et al.*, 2021; Thornburg *et al.*, 2018; Trivella *et al.*, 2022; Wagenaar, 2008).

LC-MS/MS is a key technique that is widely used in natural product-based drug-discovery pipelines for detecting and dereplicating natural product samples (Wang *et al.*, 2016; Kurita *et al.*, 2015; Hight *et al.*, 2022). However, it is noteworthy that the bioactive compound was not detected in the crude extract by LC-MS/MS, being progressively detected in the purified samples (Fig. 5*c*). This indicates that the bioactive compound (*m*/*z* 679.27) is present in the crude extract in trace amounts, making it undetectable by LC-MS/MS. Nonetheless, our crystallographic capture technique was able to detect the MA9 crude extract ligand at the active binding site of the proteasome, which was further highlighted by *PanDDA* analysis (Fig. 2*b*). Despite the electron density for the bioactive compound still not being clear enough to define its chemical structure in the crude extract, it was sufficient to show the presence of a genuine ligand binding to the catalytic site of the proteasome. This suggests that the demonstrated crystallographic capture approach can complement LC-MS/MS in the early detection of bioactive compounds in natural product-based drug-discovery pipelines, even when the bioactive molecule is present in trace amounts. Moving from the MA9 crude extract to its E/F enriched fraction, in which the bioactive compound starts to be detected by LC-MS/MS (*m*/*z* 679.27; Fig. 2*c*), it is possible to clearly observe the presence of the ligand in the $F_o - F_c$ map (Fig. 2*a*). The ligand electron density is further defined in the *PanDDA* analysis (Fig. 2*b*), allowing interpretation of its chemical structure. It is also important to comment that additional bioactive compound purification did not enhance the definition of the ligand in the *PanDDA* maps. Therefore, the chemical structure interpretation of the MA9 captured ligand could already be performed from the initial enriched fraction E/F, which was achieved much more quickly and using significantly smaller amounts (<10 µg) of the natural product sample than the isolated compound.

For validation purposes, the bioactive compounds isolated from MA9 were identified by analysis of spectroscopic data as TMC-95A or TMC-95B (Fig. 3). TMC-95A/B could be perfectly fitted into the electron densities observed in the proteasome catalytic site. These results unequivocally confirm the observations made for the ligand captured from the MA9 E/F fraction crystal structure.

TMC-95A and TMC-95B are known peptide macrocycles that display low-nanomolar noncovalent proteasome inhibition (Groll *et al.*, 2001). TMC-95A/B are interconvertible diastereomers (Fig. 3*c*) that show equivalent proteasome

inhibition (Fig. 3*d*). They differ in the position of a hydroxyl group at chiral center #2 (Fig. 3*a*). The other two diastereomers, TMC-95C and TMC-95D, differ from TMC-95A/B in the position of the hydroxyl group at chiral center #1 (Fig. 3*b*). TMC-95C and TMC-95D were also found in the purified MA9 samples, but at lower concentrations (i1 and i2 in Fig. 3*c*).

Interestingly, the crystallographic capture approach further assisted in isomer identification and provided insights into the significance of chiral centers for target enzyme inhibition. Observing the crystal structure (Fig. 3*a*), it is noted that chiral center #2 does not affect ligand binding, as it is positioned on the left side of the binding site, opposed to the catalytic center. This region of the binding site consists of a large chamber with no significant contacts between the ligand and the protein. In contrast, chiral center #1 is positioned deep within the catalytic center, close to the catalytic Thr1 residue. The main protein–ligand contacts are found in this region. Changing chiral center #1 would disrupt contacts with the Gly47 main chain, generate steric clashes with the protein, affect the ligand position and compromise protein–ligand contacts. Indeed, TMC-95C and TMC-95D are tenfold less active against the proteasome than TMC-95A or TMC-95B (Koguchi *et al.*, 2000). Therefore, knowing the binding site and 3D shape of the captured ligand provides valuable insights into protein–ligand interactions, including the assignment of isomers and the significance of chiral centers for protein binding.

3.2.2. *Streptomyces* sp. (BRA346) case study

The second case study presented here, *Streptomyces* sp. BRA346, shows a covalent ligand captured in the proteasome catalytic site. The BRA346 bioactive fraction, which is equivalent to the samples present in modern pre-fractionated natural product libraries, was soaked into yeast proteasome crystals. Electron density ($2mF_{obs} - DF_{model}$ and $mF_{obs} - DF_{model}$ *Phenix* maps) and the *PanDDA* event map for BRA-346 are presented in Fig. 4, along with the corresponding analytical LC-MS/MS chromatogram of the respective natural product samples.

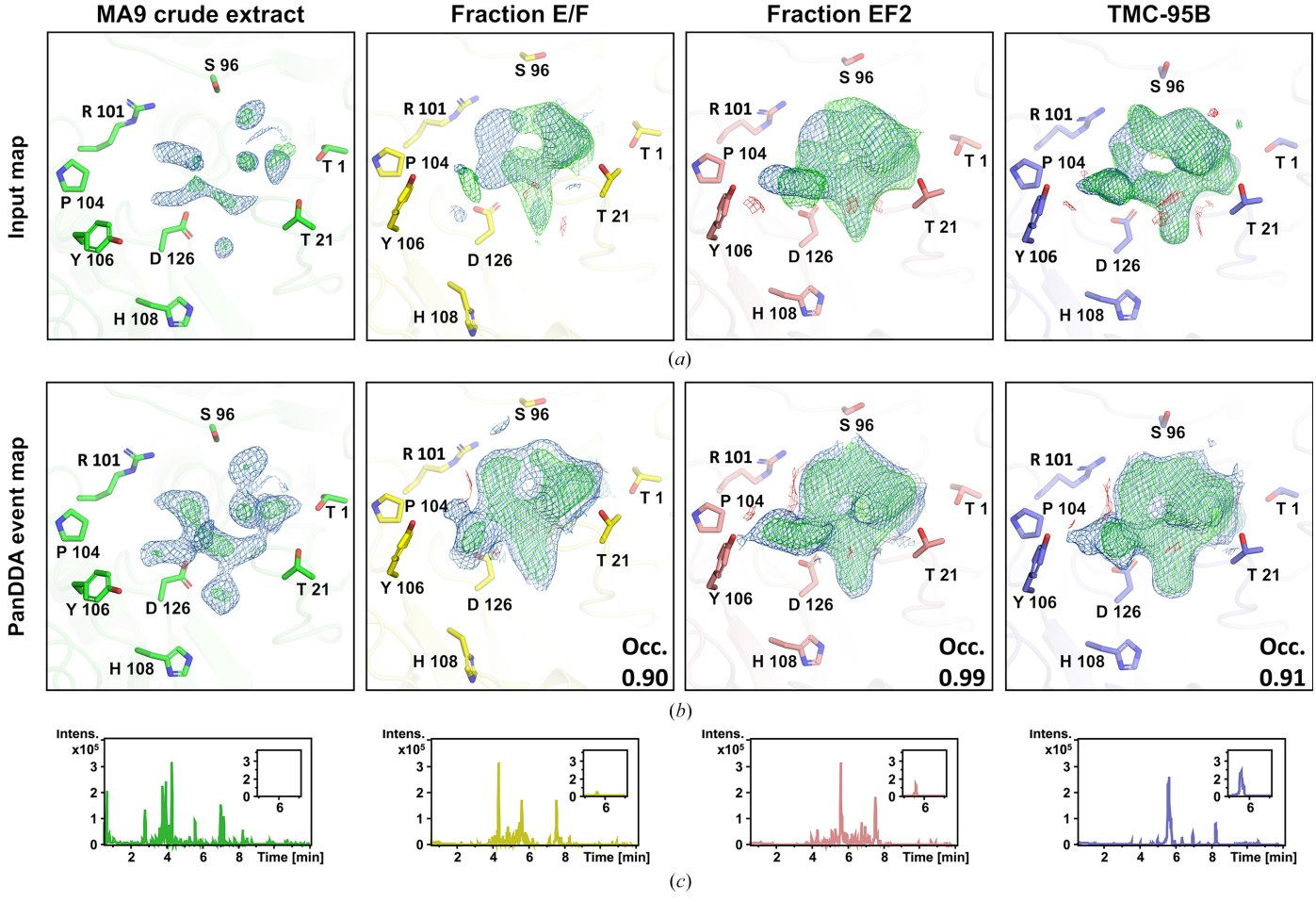


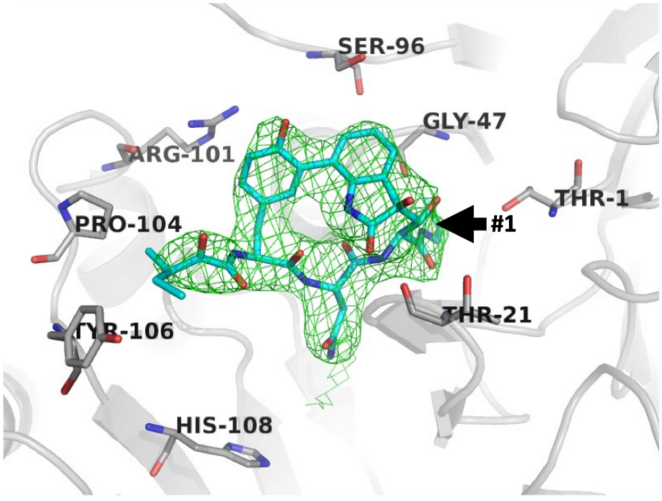
Figure 2
A. moriforme (MA9) crystallographic series. The panels show the yeast 20S proteasome active site (β5 subunit) binding a captured natural product from MA9 crude extract (left column) and fractions E/F and EF2, along with the isolated bioactive compound TMC-95B (right column). (*a*) Input maps are $2F_o - F_c$ (blue) and $F_o - F_c$ (green/red) contoured at 1σ and $\pm 3\sigma$, respectively. (*b*) *PanDDA* event (blue) and *Z*-maps (green/red) are contoured at 1.5σ and $\pm 3\sigma$, respectively. Ligand refined occupancies at subunit β5 are 0.90, 0.99 and 0.91 for fractions E/F, EF2 and isolated TMC-95B, respectively. (*c*) LC-MS/MS base-peak chromatograms (BPCs) are further provided for each MA9 natural product sample (MA9 crude extract, fraction E/F, fraction EF2 and TMC-95B). The insets are the extracted ion chromatograms (EICs) for *m*/*z* 679.27 (TMC-95).

The resultant structures from *DIMPLE* (input maps) were inspected using *Coot*, which revealed a discernible electron density in the difference maps within the chymotrypsin-like active site of the proteasome. For BRA-346 samples, the electron density clearly showed a covalently bound ligand to the catalytic Thr1, even in the enriched fraction, which still contained a significant number of molecules. All generated samples underwent LC-MS/MS analysis, bioactivity correlation and chemical structure dereplication, indicating the component with *m*/*z* 343.19 (TMC-86A) to be the bioactive compound (Bazzano *et al.*, 2024).

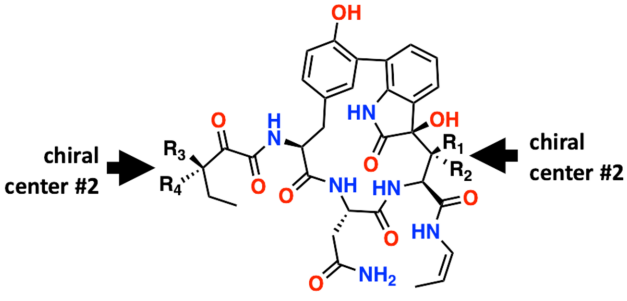
TMC-86 is a linear peptide that covalently binds to the active site of the proteasome and inspired the development of carfilzomib and oprozomib, two drugs used in cancer treatment (Huang *et al.*, 2023; Johnson *et al.*, 2018; Kim & Crews, 2013).

The bioactive molecule TMC-86A was then isolated for proof-of-concept purposes and soaked into new yeast proteasome crystals. The resulting ligand electron density (Fig. 5) confirmed that the crystallographic capture approach can recover the bioactive ligand, even in the early stages of natural product drug discovery.

Taken together, crystallography and LC-MS/MS metabolomics are orthogonal techniques that provide complementary information on unpurified bioactive natural products and can be used in a high-throughput mode. These techniques should be combined in modern natural product-based drug-discovery pipelines. By using LC-MS/MS dereplication and bioactivity-ranking pipelines (Baskiyar *et al.*, 2022; Bazzano *et al.*, 2024; Nothias *et al.*, 2018), the project benefits from mass detection, fragmentation pattern-based chemical structure annotation and bioactive compound ranking directly from unpurified

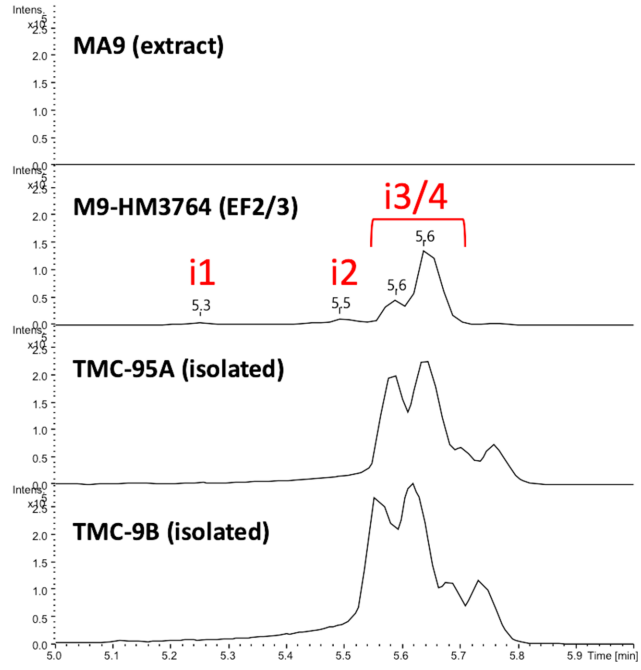


(*a*)

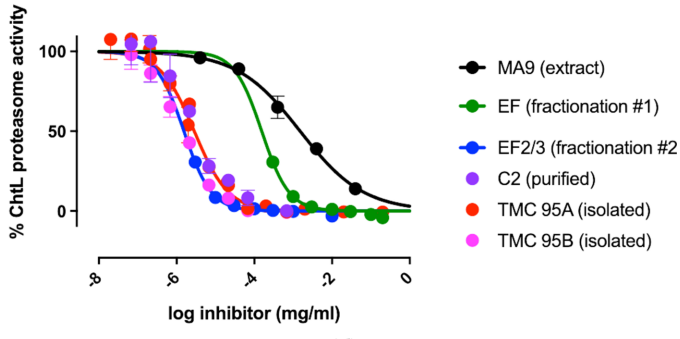


	R_1	R_2	R_3	R_4
TMC-95A	H	OH	CH_3	H
TMC-95B	H	OH	H	CH_3
TMC-95C	OH	H	CH_3	H
TMC-95D	OH	H	H	CH_3

(*b*)



(*c*)



(*d*)

Figure 3
Crystallographic data can differentiate isomers of the captured natural product, further supporting SBDD and (bio)synthetic route design. (*a*) Refined TMC-95B structure using the TMC-95B data set. Ligand occupancy was refined to 0.91 at subunit β5. (*b*) Structure of TMC-95 diastereomers. (*c*) EIC for *m*/*z* 679.27 in MA9 samples. (*d*) Chymotrypsin-like (subunit β5) yeast 20S proteasome-inhibition curves. Experimental data are shown as a normalized percentage of proteasome enzymatic activity. Values shown are the average ± SD of three experimental replicates.

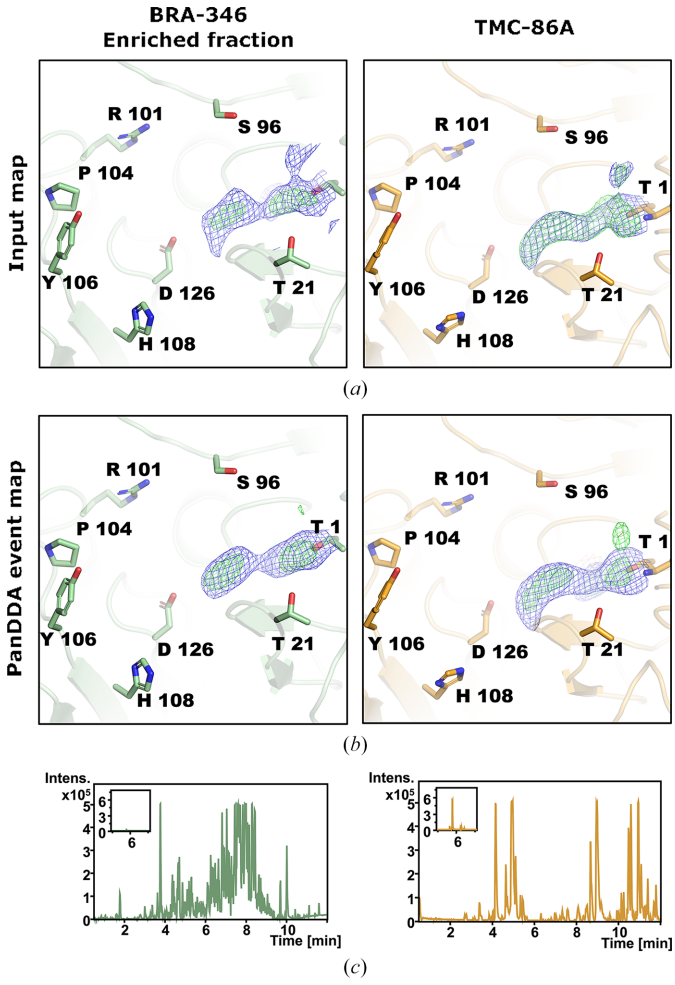


Figure 4
Crystallographic capture is overly sensitive, providing ligand structural information and the protein binding sites in the very early stages of natural product-based drug discovery. (*a*) Input maps are $2F_o - F_c$ (blue) and $F_o - F_c$ (green/red) contoured at 1σ and $\pm 3\sigma$, respectively. (*b*) *PanDDA* event (blue) and *Z*-maps (green/red) are contoured at 1.5σ and $\pm 3\sigma$, respectively. (*c*) Base-peak chromatograms (BPCs) are provided for positive LC-MS/MS detection of the BRA-346 enriched fraction and purified TMC-86A). The insets are the extracted EICs for *m*/*z* 343.19.

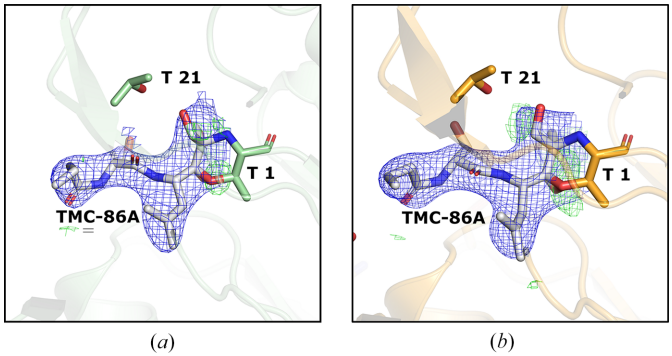


Figure 5
(*a*) Proposed bioactive molecule (*m*/*z* 343.19) present in the enriched fraction and (*b*) the purified bioactive molecule (*m*/*z* 343.19) covalently bound to the catalytic Thr1 of the chemotrypsin-like active site. Refined maps are $2F_o - F_c$ (blue) and $F_o - F_c$ (green/red) contoured at 1σ and $\pm 3\sigma$, respectively.

mixtures of natural products. The crystallographic capture approach further enhances the comprehension of the structural features of the bioactive natural product by revealing the 3D image of the ligand at near-atomic resolution, its binding site in the target protein and key protein–ligand interactions, all of which are essential for structure-based drug discovery (SBDD).

3.3. Success rates for automated data processing and bioactive natural product capture from complex mixtures

We have successfully captured several natural products that bind to different classes of proteins using the 'crystallographic capture of unknown bioactive natural products' pipeline. Four of these projects are highlighted in Fig. 6.

Protein identities and ligands are not disclosed at this stage due to active development of these projects. However, some statistics regarding the success rates of the pipeline are presented as follows. For automated data processing we achieved a success rate of 33–98% (78 ± 13) for data sets suitable for *PanDDA*. These data sets passed the automated molecular-replacement step and their refined PDB and MTZ files have R_{free} and R_{work} below 0.4 (Supplementary Table S1). The average success rate for capturing bioactive natural products was 48%. In these four projects, we started with 65 validated HTS hits from pre-fractionated libraries of natural products. These samples, consisting of biota with unknown mixtures of natural products, conferred target enzyme inhibition with IC_{50} values better than 10 µg ml^{-1}. We successfully captured 31 natural product ligands in the crystal structures, with success rates ranging from 42 to 100%, depending on the target protein that was used (Fig. 6, Supplementary Table S2).

Interestingly, we have captured ligands in both active and allosteric sites, with some ligands probing previously unknown allosteric binding sites (highlighted in Fig. 6). Notably, the captured natural products exhibited a wide range of polarities and bound to neutral, positively charged and negatively charged binding sites. This revealed a diverse array of natural chemicals that can be used as structural probes for drug development.

The occupancies of the captured ligands ranged from 0.2 to 1.0. It is important to emphasize that the samples used were pre-fractionated natural product libraries, with a preliminary reverse-phase chromatography step performed directly from the crude extracts (de Felício *et al.*, 2021; Trivella *et al.*, 2022). Despite often being minor compounds in the fractions, the captured ligands were successfully visualized within the protein crystal structures.

4. Conclusions

High-throughput protein crystallography is a key tool in natural product-based drug-discovery pipelines. Crystals act as 'filters' that directly capture bioactive natural products from intricate mixtures of unpurified and unidentified compounds. Determining the crystal structure of complexes soaked with enriched natural product fractions facilitates the early iden-

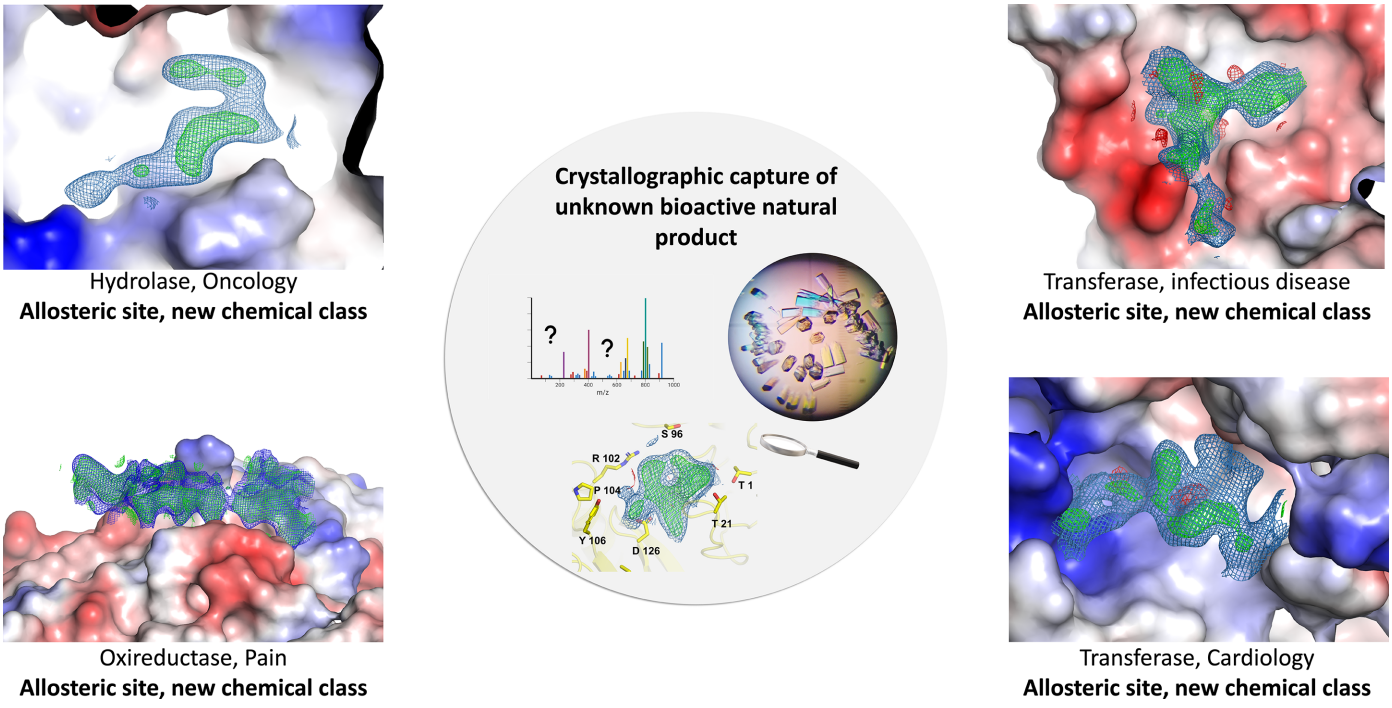


Figure 6
Case examples of captured natural products in allosteric sites of different protein targets, using the crystallographic capture approach directly from enriched fractions of natural products used in high-throughput screening with pre-fractionated natural product chemical libraries. *PanDDA* event maps (blue) and *Z*-maps (green/red) are contoured at 1σ and $\pm$3σ, respectively. The protein surface is colored according to its electrostatic potential in vacuum (gradient red–white–blue from $-100\ kT$/e to $+100\ kT$/e, negative potential in red and positive potential in blue). Binding-site figures and electrostatic potential calculations were generated with *PyMOL* (Schrödinger).

tification of novel binding sites. This approach provides valuable insights into the structural characteristics of the captured bioactive natural products, including their shapes and interactions with target proteins. In addition, direct *in crystallo* capture of natural products can quickly reveal novel allosteric binding sites and new chemical scaffolds, thereby feeding SBDD pipelines with innovative information.

Moreover, this approach can be combined with LC-MS/MS dereplication, benefiting from the latest advances in high-throughput data-generation and data-mining approaches applicable to natural product-based drug discovery. New medicines are needed in many therapeutic areas, and natural products are excellent sources of novel chemistry and probes for novel binding sites and mechanisms of action. The as yet untapped chemical diversity of natural products can be readily investigated using the pipeline presented here, especially in the case of Brazil, *in loco*.

Acknowledgements

The Brazilian Centre for Research in Energy and Materials (CNPEM) is acknowledged for the use of its facilities, especially the Chemistry and Natural Product Laboratory (LQPN, LNBio, CNPEM), the Bioassay Laboratory (LBE, LNBio, CNPEM), the UVX-MX2 (UVX, LNLS, CNPEM) and Sirius-MANACÁ (Sirius, LNLS, CNPEM) beamlines, and the HPC-TEPUI (Sirius, LNLS, CNPEM). The authors also thank Professor Dr Leticia Veras Costa-Lotufo and Dr Luciana C. Furtado from the University of São Paulo for providing *Streptomyces* sp. BRA-346 samples.

Funding information

This research was funded by Serrapilheira Institute grant No. Serra-1709-19681 (to DBBT), São Paulo Research Foundation (FAPESP) grant Nos. 2014/10753-9 (to DBBT), 2015/01017-0 and 2019/17721-9 (to RGSB), and bursary Nos. 2023/01133-6, 2021/01309-1 (to AFB) and 2019/13497-7 (to CCLS), and the Ministry of Science, Technology and Innovation – MCTI (CNPEM). AZNF received a fellowship from the Coordination for the Improvement of Higher Education Personnel (CAPES).

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C. B., Baslé, A., Beilsten-Edmands, J., Borges, R. J., Brown, D. G., Burgos-Mármol, J. J., Berrisford, J. M., Bond, P. S., Caballero, I., Catapano, L., Chojnowski, G., Cook, A. G., Cowtan, K. D., Croll, T. I., Debreczeni, J. É., Devenish, N. E., Dodson, E. J., Drevon, T. R., Emsley, P., Evans, G., Evans, P. R., Fando, M., Foadi, J., Fuentes-Montero, L., Garman, E. F., Gerstel, M., Gildea, R. J., Hatti, K., Hekkelman, M. L., Heuser, P., Hoh, S. W., Hough, M. A., Jenkins, H. T., Jiménez, E., Joosten, R. P., Keegan, R. M., Keep, N., Krissinel, E. B., Kolenko, P., Kovalevskiy, O., Lamzin, V. S., Lawson, D. M., Lebedev, A. A., Leslie, A. G. W., Lohkamp, B., Long, F., Malý, M., McCoy, A. J., McNicholas, S. J., Medina, A., Millán, C., Murray,

J. W., Murshudov, G. N., Nicholls, R. A., Noble, M. E. M., Oeffner, R., Pannu, N. S., Parkhurst, J. M., Pearce, N., Pereira, J., Perrakis, A., Powell, H. R., Read, R. J., Rigden, D. J., Rochira, W., Sammito, M., Sánchez Rodríguez, F., Sheldrick, G. M., Shelley, K. L., Simkovic, F., Simpkin, A. J., Skubak, P., Sobolev, E., Steiner, R. A., Stevenson, K., Tews, I., Thomas, J. M. H., Thorn, A., Valls, J. T., Uski, V., Usón, I., Vagin, A., Velankar, S., Vollmar, M., Walden, H., Waterman, D., Wilson, K. S., Winn, M. D., Winter, G., Wojdyr, M. & Yamashita, K. (2023). *Acta Cryst.* D**79**, 449–461.

Aguda, A. H., Lavallee, V., Cheng, P., Bott, T. M., Meimetis, L. G., Law, S., Nguyen, N. T., Williams, D. E., Kaleta, J., Villanueva, I., Davies, J., Andersen, R. J., Brayer, G. D. & Brömme, D. (2016). *J. Nat. Prod.* **79**, 1962–1970.

Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., Orhan, I. E., Banach, M., Rollinger, J. M., Barreca, D., Weckwerth, W., Bauer, R., Bayer, E. A., Majeed, M., Bishayee, A., Bochkov, V., Bonn, G. K., Braidy, N., Bucar, F., Cifuentes, A., D'Onofrio, G., Bodkin, M., Diederich, M., Dinkova-Kostova, A. T., Efferth, T., El Bairi, K., Arkells, N., Fan, T., Fiebich, B. L., Freissmuth, M., Georgiev, M. I., Gibbons, S., Godfrey, K. M., Gruber, C. W., Heer, J., Huber, L. A., Ibanez, E., Kijjoa, A., Kiss, A. K., Lu, A., Macias, F. A., Miller, M. J. S., Mocan, A., Müller, R., Nicoletti, F., Perry, G., Pittalà, V., Rastrelli, L., Ristow, M., Russo, G. L., Silva, A. S., Schuster, D., Sheridan, H., Skalicka-Woźniak, K., Skaltsounis, L., Sobarzo-Sánchez, E., Bredt, D. S., Stuppner, H., Sureda, A., Tzvetkov, N. T., Vacca, R. A., Aggarwal, B. B., Battino, M., Giampieri, F., Wink, M., Wolfender, J., Xiao, J., Yeung, A. W. K., Lizard, G., Popp, M. A., Heinrich, M., Berindan-Neagoe, I., Stadler, M., Daglia, M., Verpoorte, R. & Supuran, C. T. (2021). *Nat. Rev. Drug Discov.* **20**, 200–216.

Bancet, A., Raingeval, C., Lomberget, T., Le Borgne, M., Guichou, J.-F. & Krimm, I. (2020). *J. Med. Chem.* **63**, 11420–11435.

Baskiyar, S., Ren, C., Heck, K. L., Hall, A. M., Gulfam, M., Packer, S., Seals, C. D. & Calderón, A. I. (2022). *J. Chem. Inf. Model.* **62**, 6378–6385.

Bazzano, C. F., de Felicio, R., Alves, L. F. G., Costa, J. H., Ortega, R., Vieira, B. D., Morais-Urano, R. P., Furtado, L. C., Ferreira, E. L. F., Gubiani, J. R., Berlinck, R. G. S., Costa-Lotufo, L. V., Telles, G. P. & Trivella, D. B. B. (2024). *Anal. Chem.* **96**, 7460–7469.

Berlinck, R. G. S., Monteiro, A. F., Bertonha, A. F., Bernardi, D. I., Gubiani, J. R., Slivinski, J., Michaliski, L. F., Tonon, L. A. C., Venancio, V. A. & Freire, V. F. (2019). *Nat. Prod. Rep.* **36**, 981–1004.

Beutler, J. A. (2009). *Curr. Protoc. Pharmacol.* **46**, 9.

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H. & Weber, T. (2021). *Nucleic Acids Res.* **49**, W29–W35.

Blundell, T. L., Jhoti, H. & Abell, C. (2002). *Nat. Rev. Drug Discov.* **1**, 45–54.

Boby, M. L., Fearon, D., Ferla, M., Filep, M., Koekemoer, L., Robinson, M. C., Chodera, J. D., Lee, A. A., London, N., von Delft, A., von Delft, F., Achdout, H., Aimon, A., Alonzi, D. S., Arbon, R., Aschenbrenner, J. C., Balcomb, B. H., Bar-David, E., Barr, H. & Zvornicanin, S. N. (2023). *Science*, **382**, eabo7201.

Bon, M., Bilsland, A., Bower, J. & McAulay, K. (2022). *Mol. Oncol.* **16**, 3761–3777.

Bruder, M., Polo, G. & Trivella, D. B. B. (2020). *Nat. Prod. Rep.* **37**, 488–514.

Dauter, Z. (2017). *Methods Mol. Biol.* **1607**, 165–184.

Domingues Vieira, B., Niero, H., de Felício, R., Giolo Alves, L. F., Freitas Bazzano, C., Sigrist, R., Costa Furtado, L., Felix Persinoti, G., Veras Costa-Lotufo, L. & Barretto Barbosa Trivella, D. (2022). *Front. Microbiol.* **13**, 786008.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. (2016). *Nat. Rev. Drug Discov.* **15**, 605–619.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.

Fan, L., Tan, L., Chen, Z., Qi, J., Nie, F., Luo, Z., Cheng, J. & Wang, S. (2020). *Nat. Commun.* **11**, 1074.

Felício, R. de, Ballone, P., Bazzano, C. F., Alves, L. F. G., Sigrist, R., Infante, G. P., Niero, H., Rodrigues-Costa, F., Fernandes, A. Z. N., Tonon, L. A. C., Paradela, L. S., Costa, R. K. E., Dias, S. M. G., Dessen, A., Telles, G. P., da Silva, M. A. C., de Souza Lima, A. O. & Trivella, D. B. B. (2021). *Metabolites*, **11**, 107.

Furtado, L. C., Bauermeister, A., de Felicio, R., Ortega, R., Pinto, F., Machado-Neto, J. A., Trivella, D. B. B., Pessoa, O. D. L., Wilke, D. V., Lopes, N. P., Jimenez, P. C. & Costa-Lotufo, L. V. (2021). *Front. Mar. Sci.* **8**, 644730.

Gabadinho, J., Beteva, A., Guijarro, M., Rey-Bakaikoa, V., Spruce, D., Bowler, M. W., Brockhauser, S., Flot, D., Gordon, E. J., Hall, D. R., Lavault, B., McCarthy, A. A., McCarthy, J., Mitchell, E., Monaco, S., Mueller-Dieckmann, C., Nurizzo, D., Ravelli, R. B. G., Thibault, X., Walsh, M. A., Leonard, G. A. & McSweeney, S. M. (2010). *J. Synchrotron Rad.* **17**, 700–707.

Grigalunas, M., Brakmann, S. & Waldmann, H. (2022). *J. Am. Chem. Soc.* **144**, 3314–3329.

Groll, M., Koguchi, Y., Huber, R. & Kohno, J. (2001). *J. Mol. Biol.* **311**, 543–548.

Günther, S., Reinke, P. Y. A., Fernández-García, Y., Lieske, J., Lane, T. J., Ginn, H. M., Koua, F. H. M., Ehrt, C., Ewert, W., Oberthuer, D., Yefanov, O., Meier, S., Lorenzen, K., Krichel, B., Kopicki, J.-D., Gelisio, L., Brehm, W., Dunkel, I., Seychell, B., Gieseler, H., Norton-Baker, B., Escudero-Pérez, B., Domaracky, M., Saouane, S., Tolstikova, A., White, T. A., Hänle, A., Groessler, M., Fleckenstein, H., Trost, F., Galchenkova, M., Gevorkov, Y., Li, C., Awel, S., Peck, A., Barthelmess, M., Schlünzen, F., Lourdu Xavier, P., Werner, N., Andaleeb, H., Ullah, N., Falke, S., Srinivasan, V., França, B. A., Schwinzer, M., Brognaro, H., Rogers, C., Melo, D., Zaitseva-Kinneberg, J. I., Knoska, J., Peña-Murillo, G. E., Mashhour, A. R., Hennicke, V., Fischer, P., Hakanpää, J., Meyer, J., Gribbon, P., Ellinger, B., Kuzikov, M., Wolf, M., Beccari, A. R., Bourenkov, G., von Stetten, D., Pompidor, G., Bento, I., Panneerselvam, S., Karpics, I., Schneider, T. R., Garcia-Alai, M. M., Niebling, S., Günther, C., Schmidt, C., Schubert, R., Han, H., Boger, J., Monteiro, D. C. F., Zhang, L., Sun, X., Pletzer-Zelgert, J., Wollenhaupt, J., Feiler, C. G., Weiss, M. S., Schulz, E. C., Mehrabi, P., Karničar, K., Usenik, A., Loboda, J., Tidow, H., Chari, A., Hilgenfeld, R., Uetrecht, C., Cox, R., Zaliani, A., Beck, T., Rarey, M., Günther, S., Turk, D., Hinrichs, W., Chapman, H. N., Pearson, A. R., Betzel, C. & Meents, A. (2021). *Science*, **372**, 642–646.

Harriman, G., Greenwood, J., Bhat, S., Huang, X., Wang, R., Paul, D., Tong, L., Saha, A. K., Westlin, W. F., Kapeller, R. & Harwood, H. J. (2016). *Proc. Natl Acad. Sci. USA*, **113**, E1796–E1805.

Hight, S. K., Clark, T. N., Kurita, K. L., McMillan, E. A., Bray, W., Shaikh, A. F., Khadilkar, A., Haeckl, F. P. J., Carnevale-Neto, F., La, S., Lohith, A., Vaden, R. M., Lee, J., Wei, S., Lokey, R. S., White, M. A., Linington, R. G. & MacMillan, J. B. (2022). *Proc. Natl Acad. Sci. USA*, **119**, e2208458119.

Huang, C., Zabala, D., de Los Santos, E. L. C., Song, L., Corre, C., Alkhalaf, L. M. & Challis, G. L. (2023). *Nucleic Acids Res.* **51**, 1488–1499.

Huang, C.-Y., Metz, A., Lange, R., Artico, N., Potot, C., Hazemann, J., Müller, M., Dos Santos, M., Chambovey, A., Ritz, D., Eris, D., Meyer, S., Bourquin, G., Sharpe, M. & Mac Sweeney, A. (2024). *Acta Cryst.* D**80**, 123–136.

Johnson, H. W. B., Lowe, E., Anderl, J. L., Fan, A., Muchamuel, T., Bowers, S., Moebius, D. C., Kirk, C. & McMinn, D. L. (2018). *J. Med. Chem.* **61**, 11127–11143.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 133–144.

Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.

Keil, P. & Chase, J. M. (2019). *Nat. Ecol. Evol.* **3**, 390–399.

Kim, K. B. & Crews, C. M. (2013). *Nat. Prod. Rep.* **30**, 600.

Kirsch, P., Hartman, A. M., Hirsch, A. K. H. & Empting, M. (2019). *Molecules*, **24**, 4309.

Koguchi, Y., Kohno, J., Nishio, M., Takahashi, K., Okuda, T., Ohnuki, T. & Komatsubara, S. (2000). *J. Antibiot.* **53**, 105–109.

Kühlbrandt, W. (2014). *Science*, **343**, 1443–1444.

Kurita, K. L., Glassey, E. & Linington, R. G. (2015). *Proc. Natl Acad. Sci. USA*, **112**, 11999–12004.
Lachance, H., Wetzel, S., Kumar, K. & Waldmann, H. (2012). *J. Med. Chem.* **55**, 5989–6001.
Leit, S., Greenwood, J., Carriero, S., Mondal, S., Abel, R., Ashwell, M., Blanchette, H., Boyles, N. A., Cartwright, M., Collis, A., Feng, S., Ghanakota, P., Harriman, G. C., Hosagrahara, V., Kaila, N., Kapeller, R., Rafi, S. B., Romero, D. L., Tarantino, P. M., Timaniya, J., Toms, A. V., Wester, R. T., Westlin, W., Srivastava, B., Miao, W., Tummino, P., McElwee, J. J., Edmondson, S. D. & Masse, C. E. (2023). *J. Med. Chem.* **66**, 10473–10496.
Leonti, M. & Casu, L. (2013). *Front. Pharmacol.* **4**, 92.
Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.
Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., Green, D. V. S., Hertzberg, R. P., Janzen, W. P., Paslay, J. W., Schopfer, U. & Sittampalam, G. S. (2011). *Nat. Rev. Drug Discov.* **10**, 188–195.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Mitcheltree, M. J., Pisipati, A., Syroegin, E. A., Silvestre, K. J., Klepacki, D., Mason, J. D., Terwilliger, D. W., Testolin, G., Pote, A. R., Wu, K. J. Y., Ladley, R. P., Chatman, K., Mankin, A. S., Polikanov, Y. S. & Myers, A. G. (2021). *Nature*, **599**, 507–512.
Mueller, M., Wang, M. & Schulze-Briese, C. (2012). *Acta Cryst.* D**68**, 42–56.
Nascimento, A., Araujo, E., Hagio, C., Almeida, S., Rodrigues, A. C., Barretto Barbosa Trivella, D., Bruder, M., Rustiguel, J. K., Dislich Ropke, C., Ravanelli Pessa, L., Azevedo, H., Cardoso, M. & Zeri, A. C. (2021). *Synchrotron Radiat. News*, **34**, 3–10.
Newman, D. J. & Cragg, G. M. (2020). *J. Nat. Prod.* **83**, 770–803.
Nothias, L. F., Nothias-Esposito, M., da Silva, R., Wang, M., Protsyuk, I., Zhang, Z., Sarvepalli, A., Leyssen, P., Touboul, D., Costa, J., Paolini, J., Alexandrov, T., Litaudon, M. & Dorrestein, P. C. (2018). *J. Nat. Prod.* **81**, 758–767.
Oscarsson, M., Beteva, A., Flot, D., Gordon, E., Guijarro, M., Leonard, G., McSweeney, S., Monaco, S., Mueller-Dieckmann, C., Nanao, M., Nurizzo, D., Popov, A., von Stetten, D., Svensson, O., Rey-Bakaikoa, V., Chado, I., Chavas, L., Gadea, L., Gourhant, P., Isabet, T., Legrand, P., Savko, M., Sirigu, S., Shepard, W., Thompson, A., Mueller, U., Nan, J., Eguiraun, M., Bolmsten, F., Nardella, A., Milàn-Otero, A., Thunnissen, M., Hellmig, M., Kastner, A., Schmuckermaier, L., Gerlach, M., Feiler, C., Weiss, M. S., Bowler, M. W., Gobbo, A., Papp, G., Sinoir, J., McCarthy, A., Karpics, I., Nikolova, M., Bourenkov, G., Schneider, T., Andreu, J., Cuní, G., Juanhuix, J., Boer, R., Fogh, R., Keller, P., Flensburg, C., Paciorek, W., Vonrhein, C., Bricogne, G. & de Sanctis, D. (2019). *J. Synchrotron Rad.* **26**, 393–405.
Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., Marsden, B. D., Kelm, S., Shi, J., Deane, C. M. & von Delft, F. (2017). *Nat. Commun.* **8**, 15123.
Pearce, N. M., Krojer, T. & von Delft, F. (2017). *Acta Cryst.* D**73**, 256–266.
Pilon, A. C., Valli, M., Dametto, A. C., Pinto, M. E. F., Freire, R. T., Castro-Gamboa, I., Andricopulo, A. D. & Bolzani, V. S. (2017). *Sci. Rep.* **7**, 7215.
Robertson, M. J., Meyerowitz, J. G. & Skiniotis, G. (2022). *Trends Biochem. Sci.* **47**, 124–135.
Sabatini, F. M., Jiménez-Alfaro, B., Jandt, U., Chytrý, M., Field, R., Kessler, M., Lenoir, J., Schrodt, F., Wiser, S. K., Arfin Khan, M. A. S., Attorre, F., Cayuela, L., De Sanctis, M., Dengler, J., Haider, S., Hatim, M. Z., Indreica, A., Jansen, F., Pauchard, A., Peet, R. K., Petřík, P., Pillar, V. D., Sandel, B., Schmidt, M., Tang, Z., van Bodegom, P., Vassilev, K., Violle, C., Alvarez-Davila, E., Davidar, P., Dolezal, J., Hérault, B., Galán-de-Mera, A., Jiménez, J., Kambach, S., Kepfer-Rojas, S., Kreft, H., Lezama, F., Linares-Palomino, R., Monteagudo Mendoza, A., N'Dja, J. K., Phillips, O. L., Rivas-Torres, G., Sklenář, P., Speziale, K., Strohbach, B. J., Vásquez Martínez, R., Wang, H., Wesche, K. & Bruelheide, H. (2022). *Nat. Commun.* **13**, 4683.
Sabe, V. T., Ntombela, T., Jhamba, L. A., Maguire, G. E. M., Govender, T., Naicker, T. & Kruger, H. G. (2021). *Eur. J. Med. Chem.* **224**, 113705.
Sadybekov, A. V. & Katritch, V. (2023). *Nature*, **616**, 673–685.
Santen, J. A. van, Poynton, E. F., Iskakova, D., McMann, E., Alsup, T. A., Clark, T. N., Fergusson, C. H., Fewer, D. P., Hughes, A. H., McCadden, C. A., Parra, J., Soldatou, S., Rudolf, J. D., Janssen, E. M.-L., Duncan, K. R. & Linington, R. G. (2022). *Nucleic Acids Res.* **50**, D1317–D1323.
Schiebel, J., Krimmer, S. G., Röwer, K., Knörlein, A., Wang, X., Park, A. Y., Stieler, M., Ehrmann, F. R., Fu, K., Radeva, N., Krug, M., Huschmann, F. U., Glöckner, S., Weiss, M. S., Mueller, U., Klebe, G. & Heine, A. (2016). *Structure*, **24**, 1398–1409.
Souza, R. T. M. P. de, Freire, V. F., Gubiani, J. R., Ferreira, R. O., Trivella, D. B. B., Moraes, F. C., Paradas, W. C., Salgado, L. T., Pereira, R. C., Amado Filho, G. M., Ferreira, A. G., Williams, D. E., Andersen, R. J., Molinski, T. F. & Berlinck, R. G. S. (2018). *J. Nat. Prod.* **81**, 2296–2300.
Stone, S., Newman, D. J., Colletti, S. L. & Tan, D. S. (2022). *Nat. Prod. Rep.* **39**, 20–32.
Talele, T., Khedkar, S. & Rigby, A. (2010). *Curr. Top. Med. Chem.* **10**, 127–141.
Tange, O. (2018). *GNU Parallel 2018*. Morrisville: Lulu.
Teufel, R., Kaysser, L., Villaume, M. T., Diethelm, S., Carbullido, M. K., Baran, P. S. & Moore, B. S. (2014). *Angew. Chem. Int. Ed.* **53**, 11019–11022.
Thomas, S. E., Collins, P., James, R. H., Mendes, V., Charoensutthivarakul, S., Radoux, C., Abell, C., Coyne, A. G., Floto, R. A., von Delft, F. & Blundell, T. L. (2019). *Phil. Trans. R. Soc. A.* **377**, 20180422.
Thornburg, C. C., Britt, J. R., Evans, J. R., Akee, R. K., Whitt, J. A., Trinh, S. K., Harris, M. J., Thompson, J. R., Ewing, T. L., Shipley, S. M., Grothaus, P. G., Newman, D. J., Schneider, J. P., Grkovic, T. & O'Keefe, B. R. (2018). *ACS Chem. Biol.* **13**, 2484–2497.
Trivella, D. B. B., Bruder, M. C. P., Oliveira, F. C. B., Porcaro, R., Rustiguel, J. K., Ribeiro, L. B., Felicio, R., Cunha, M. G., Nascimento, A. F. Z., Zeri, A. C. M., Pessa, L. R., Mascarello, A., Guimarães, C. R. W., Azevedo, H., Perfeito, M. L. G., Pagani, E. & Ropke, C. D. (2022). *Rev. Fitos*, **16**, 176–192.
Usón, I. & Sheldrick, G. M. (2018). *Acta Cryst.* D**74**, 106–116.
Valli, M., Russo, H. M. & Bolzani, V. S. (2018). *An. Acad. Bras. Cienc.* **90**, 763–778.
Wagenaar, M. (2008). *Molecules*, **13**, 1406–1426.
Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W.-T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C. C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C. C., Yang, Y. L., Humpf, H. U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C. B. P. C. A. B., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S.,

Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P. M., Phapale, P., Nothias, L. F., Alexandrov, T., Litaudon, M., Wolfender, J. L., Kyle, J. E., Metz, T. O., Peryea, T., Nguyen, D. T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K. M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P. R., Palsson, B. O., Pogliano, K., Linington, R. G., Gutiérrez, M., Lopes, N. P., Gerwick, W. H., Moore, B. S., Dorrestein, P. C. & Bandeira, N. (2016). *Nat. Biotechnol.* **34**, 828–837.

Wilson, B. A. P., Thornburg, C. C., Henrich, C. J., Grkovic, T. & O'Keefe, B. R. (2020). *Nat. Prod. Rep.* **37**, 893–918.

Wojdyr, M., Keegan, R., Winter, G. & Ashton, A. (2013). *Acta Cryst.* A**69**, s299.

Wong, F., Zheng, E. J., Valeri, J. A., Donghia, N. M., Anahtar, M. N., Omori, S., Li, A., Cubillos-Ruiz, A., Krishnan, A., Jin, W., Manson, A. L., Friedrichs, J., Helbig, R., Hajian, B., Fiejtek, D. K., Wagner, F. F., Soutter, H. H., Earl, A. M., Stokes, J. M., Renner, L. D. & Collins, J. J. (2024). *Nature*, **626**, 177–185.

Zank, S. & Hanazaki, N. (2017). *PLoS One*, **12**, e0174731.