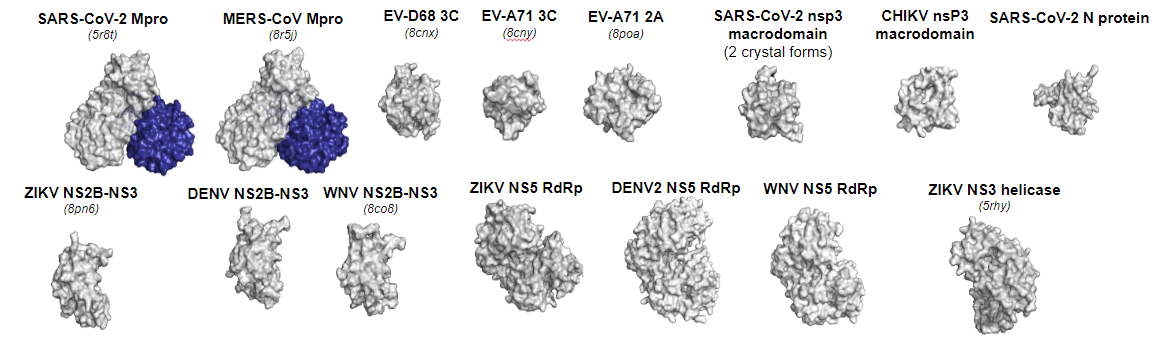
**B. Studies and Results**

The goal of Project 2 was to produce 10 Target Enablement Packages (TEPs) for viral proteins within five years, using x-ray fragment screening. Fragment-based screening is now well-established as a powerful approach to early drug ("lead") discovery. Of the many suitable biophysical techniques, X-ray crystallography was one of the first to be used, and is the most directly informative, facilitating antiviral drug discovery. A TEP provides a comprehensive data package about a target protein, including all the necessary information for structure-based drug discovery (SBDD). Finding the critical path from protein engineering to structure analysis, passing through crystal optimisation, fragment screening and fragment hit progression. To date, we have obtained 16 high-resolution crystal structures for 15 different viral target proteins, and we have completed 5 TEPs, putting us well on track to achieve the Project 2 goal.

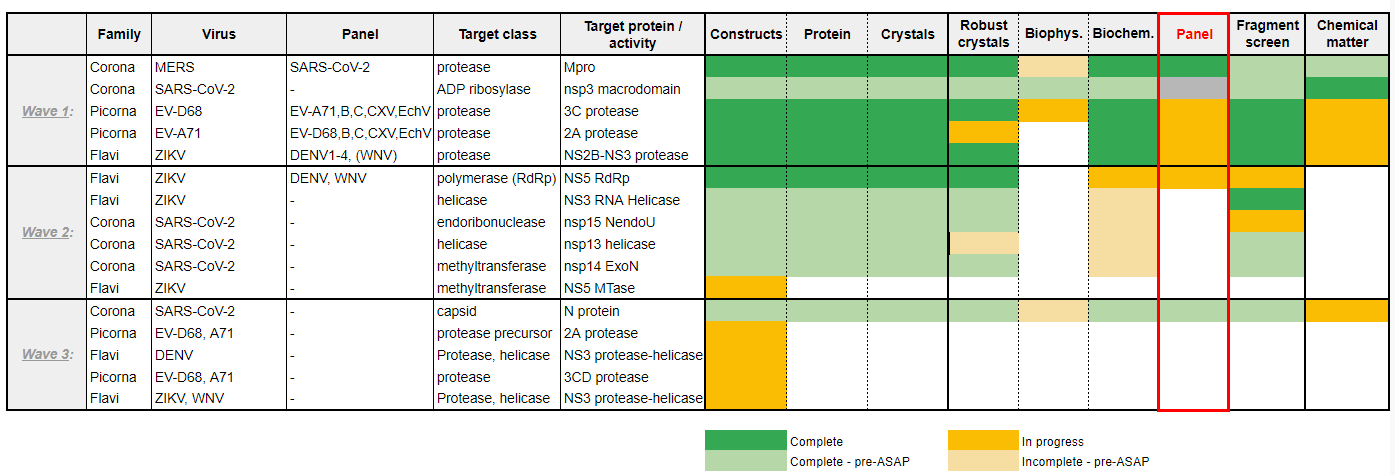


***Figure 1. Project 2 overall achievement.*** *Completed 5 Target Enablement Packages**(TEPs) and produced 16 high resolution crystal forms for 15 proteins. Resulting in 8 targeting opportunities for pandemic preparedness. All data was made public for the wider community to use and build upon the TEPs.*

This has been possible because:

* ASAP’s Project 2 has assembled a unique constellation of scientific expertise and facility infrastructure, covering virology, protein chemistry, structural biology, fragment screening, hit progression, biophysical and biochemical assay development, and computational biology.
* Project 2, in cooperation with Project 1, has developed a methodology that focuses on defining **precise targeting opportunities** (a specific molecular mechanism of action more precise than just “geneX”).
* TEPs are rapidly adaptable to new targets: panel antiviral drugs
* This effort has only modest operational costs:  $1.5m/yr (Projects 1 & 2, Structural Biology Core).
* Target enablement and immediate public data release acts as a public force multiplier
  + shrinks discovery costs for anyone (e.g., ensitrelvir discovery [1]).

To achieve broad-spectrum antiviral SBDD, we expanded the scope of our targets to include viral panels (assays and crystal structures) that allow for the development and testing of broad-spectrum antiviral drugs **(Figure 2)**. To accomplish this, we continued following a well-defined target enablement process to compile Target Enablement Packages (TEPs). Additionally, we developed a multitude of in-house technologies (explained below) to optimise TEP compilation and facilitate the development of broad-spectrum antiviral drugs targeting multiple viruses within a family.



***Figure 2. Progress of Target Enablement Packages structure-based drug discovery of broad-spectrum antivirals, targeting vital enzymatic activities.*** *The figure summarises the progress made in Project 2, which aims to produce TEPs for viral proteins from three different families: Coronaviridae, Picornaviridae, and Flaviviridae.*

**TEP is robust for structure based drug discovery (SBDD) for pandemic preparedness: Eight target opportunities were developed using TEP methodology**

1. SARS-CoV-2 Mpro

Reproducible SARS-CoV-2 main protease (Mpro) crystals suitable forfragment screening were produced.This TEP provides a comprehensive sampling of the binding properties of the active site for SARS-CoV-2 MPro by small molecule fragment molecules. Together with the established enzymatic assays, it forms the basis for SBDD. The usefulness of the data and protocols presented in this TEP have been demonstrated by the COVID Moonshot consortium successfully developing sub-micromolar binders that are now in lead optimization.

1. MERS-CoV Mpro

The highly lethal Middle East Respiratory Syndrome Coronavirus (MERS-CoV), with a mortality rate of 35%, poses a constant threat of causing another pandemic. Similar to SARS-CoV-2, MERS-CoV also possesses Mpro, which is essential for its life cycle and we investigated using the TEP methodology.

Firstly, we successfully developed methods for the expression and purification of Mpro on both small and large scales (dx.doi.org/10.17504/protocols.io.ewov194x7lr2/v1). Subsequently, we measured the specific protease activity of the enzyme in a quality control biochemical assay (dx.doi.org/10.17504/protocols.io.8epv5rzm4g1b/v1). We also obtained robust crystals suitable for fragment screening (dx.doi.org/10.17504/protocols.io.ewov194x7lr2/v1). Additionally, we studied MERS-CoV Mpro in panel assays. This TEP approach enabled the development of potential compound inhibitors against MERS-CoV Mpro.

1. SARS-CoV-2 nsp3 macrodomain

SARS-CoV-2 nsp3 macrodomain functions as a mono-ADP-ribosylhydrolase. It is a promising antiviral target as it is essential for pathogenesis and conserved throughout coronaviruses. Furthermore, recent mutagenesis studies revealed that inactivation of nsp3 macrodomain attenuates viral replication and pathogenesis. Here we have established a purification protocol for recombinant nsp3 macrodomain from SARS-CoV-2, and established a robust crystallisation protocol suitable for fragment screening. Subsequently, we have performed a substantial fragment screening campaign and identified many hits which were validated by biophysical assays performed. The developed tool compounds feature different properties, are cell-permeable and highly potent against the SARS-CoV-2 nsp3 macrodomain.

1. EV-D68 3C protease

A cause for concern is the currently emerging pathogen Enterovirus D68 (EV-D68) which primarily spreads through respiratory routes causing mostly mild to severe respiratory illness but, in severe cases, acute flaccid myelitis. The 3C protease of EV-D68 is a potential target for the development of antiviral drugs due to its essential role in the viral life cycle and high sequence conservation amongst family members. Here we have established a purification protocol [dx.doi.org/10.17504/protocols.io.n92ld8yd7v5b/v1](https://dx.doi.org/10.17504/protocols.io.n92ld8yd7v5b/v1) for recombinant 3C protease, we have tested for activity [dx.doi.org/10.17504/protocols.io.261ge54jyg47/v1](https://dx.doi.org/10.17504/protocols.io.261ge54jyg47/v1) from Enterovirus D68 (EV-D68) and established a robust crystallisation suitable for fragment screening [dx.doi.org/10.17504/protocols.io.5qpvoky29l4o/v1](https://dx.doi.org/10.17504/protocols.io.5qpvoky29l4o/v1). We have performed a substantial fragment screening campaign and identified many hits which were validated by biophysical assays established in this work.

1. ZIKV NS2B-NS3 protease

Zika virus (ZIKV) infections causes microcephaly in new-borns and Guillain-Barre syndrome in adults raising a global public health concern, yet no vaccines or antiviral drugs are available to treat or prevent ZIKV infections. The viral NS3 protease with its NS2B cofactor is essential for the cleavage of Zika polyprotein precursor into individual structural and non-structural proteins and is therefore an attractive drug target. We optimised a robust crystal system of co-expressed NS3 protease with its NS2B cofactor and used this in a crystallographic fragment screening campaign [dx.doi.org/10.17504/protocols.io.eq2lyj51mlx9/v1](https://dx.doi.org/10.17504/protocols.io.eq2lyj51mlx9/v1). Previously, crystallographic fragment screening against over 1000 fragments identified 53 binders for hit-to-lead development. Up to date, the structural Biology Core has released around 60 small molecule-bound structures, including compounds that have low micromolar IC50s against both ZIKV and DENV2 NS2B-NS3 proteases. Our structural details of protein-ligand interaction will facilitate hit-to-lead development.

1. EV-A71 2A protease

The 2A protease of Enterovirus A71 (EV-A71) is a potential target for the development of antiviral drugs due to its essential role in the viral life cycle and high sequence conservation. In our study, we have carried out several fragment screens from those provided at XChem (Diamond Light Source Ltd). The fragments screens yielded a total of 102  fragments that were shown to bind to EV-A71 2A protease, with 44 fragments directly binding to the active site of the protease. The structures of the bound fragments have been passed onto the computational chemistry Projects within ASAP for follow up designs.

1. ZIKV NS5 RdRp

The main aim of this work was to identify small molecules that bind Zika NS5 RdRp (catalytic RNA-dependent RNA polymerase domain) through X-ray fragment-based screening. The Zika NS5 RDRP domain was cloned, expressed, purified, and crystallised. Three distinct constructs were designed to address protein stability issues identified from earlier work. Suitable crystals for fragment screening were produced and optimised allowing an extensive fragment campaign to be performed. A native high-resolution structure was determined at 1.8Å and formed the basis for the fragment campaign.  Data were collected from 1026 fragment soaks and these data are currently being analysed.

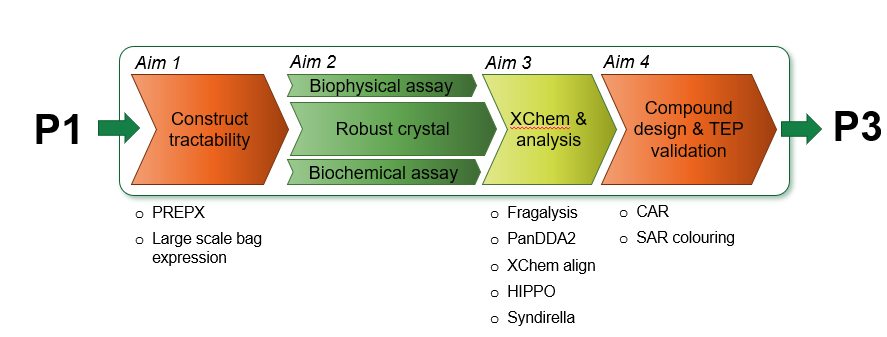
Earlier attempts to carry out a crystallographic screening of Zika catalytic RdRp domain was impeded by protein stability and degradation issues. To overcome this, three new constructs were designed. However, we focused on one of these in the first instance based on small scale purification experiments carried out on these constructs.

1. ZIKV NS3 helicase

Like other flaviviruses, Zika virus (ZIKV) presents promising drug targets within its replication machinery, notably the NS3 helicase protein (NS3hel), which plays critical roles in viral replication. Yet, a lack of structural information impedes the development of specific inhibitors targeting NS3hel. Here, by applying high-throughput crystallographic fragment screening on ZIKV NS3hel, we successfully revealed the structures of 46 fragments at multiple sites of the protein, including the identification of 8 unique fragments to the RNA-cleft site. Our findings provide crucial insights for the development of novel direct-acting antivirals against ZIKV and related flaviviruses, offering a promising avenue for combating future outbreaks.

**In-house technologies were developed during ASAP, consolidating Project 2 aims**

As summarised in **Figure 3**, the enablement of targets using the TEP methodology is only possible with the development of in-house technologies using the expertise of world leading scientists consolidated within the ASAP initiatives.

******

**Figure 3. Workflow of Target Enabling Packages (TEP) for structure based drug discovery (SBDD) and supportive in-house technologies developed during ASAP**.TEPs produced by Project 2 (P2) builds upon Project 1 (P1) viral targeting opportunity analysis and supports Project 3 (P3) hit to lead development. In-house technologies developed in the last two years consolidates the TEP methodology for antiviral SBDD.

The capabilities of Project 2 are concentrated in the UK and comprises the joint effort of Centre of Medicines Discovery (CMD) from the University of Oxford and the XChem facility, based at the Diamond Light Source. CMD was created with the ambition of developing medicines that are translated to clinical use. Within CMD, the Protein crystallography small research facility (PX-SRF) through its co-evolution with the XChem facility, contributes to this ambition by supporting structure-based drug discovery (SBDD), ranging from protein engineering, crystal optimisation, fragment screening, fragment hit progression, and structure analysis.

**Aim 1: Construct tractability**

Production of enough, consistent protein for downstream assays is crucial for the successful progression of TEPs explored by ASAP viral targets. A major blocker in SBDD is producing the right protein for the assays and crystallisation at sufficient quantities to allow complete characterization of a target. One strategy that CMD employs to address this is by exploring a diversity in constructs to probe crystallisation space (10 to >100 variants). However, this increases the experimental burden. To overcome this challenge in-house technologies were developed.

* **PREPX: Vector to simplify recombinant protein production**

PX-SRF has developed a high-throughput pipeline for efficient protein production. Synthetic DNA sequences are obtained from commercial vendors (TWIST or IDT) and cloned into a suite of expression vectors using the Golden Gate cloning method. This approach routinely achieves cloning efficiencies exceeding 90%. Recombinant proteins are subsequently expressed and purified using the PREPX method, an in-house technology developed by the CMD group [dx.doi.org/10.17504/protocols.io.yxmvm35zbl3p/v1](https://dx.doi.org/10.17504/protocols.io.yxmvm35zbl3p/v1). The highly standardised pipeline streamlines the cloning, small-scale expression, and purification of all designed constructs, enabling rapid generation of protein samples for structural and functional studies. Employing this pipeline, 92 proteins were successfully purified for further investigations in these last 12 months.

* **Large-Scale expression in bubble column bags**

To explore protein expression and crystallisation behaviour diverse constructs are designed. With this approach the experimental burden scales with the number of constructs which must be trialed. Presented is a single-use bubble column reactor array that allows for 1 L E.coli fermentation of recombinant proteins in parallel. Each reactor is constructed of a very low aspect-ratio polythene tube, sealed at the bottom, with a removable closure at the top. Mixing and aeration is supplied via a sparger, and heat by immersion in a water bath. This simple, compact system is intended to have an exceptionally low barrier to entry, whilst being highly scalable. This approach facilitates mixed operation of either single target Scale-Out, or highly parallel construct screening. [8th BioProScale Symposium 2024 Berlin (ifgb.de)](https://biotechnologie.ifgb.de/sites/biotechnologie.ifgb.de/files/2024-04/BioProScale2024_Abstracht-Book_Online_v2.pdf).

**Aim 3: XChem and analysis**

After robust crystals are obtained, the following step is to perform fragment screen using the XChem facility at Diamond Light Source. We set about addressing a major gap in the fragments field, namely the lack of a crafted, readily accessible informatics tool that allows non-specialist users to reliably progress their fragment hits to on-scale potency. This problem was not alleviated by the growing number of publications of relevant algorithms, as the barrier to usage remained high.

* **Fragalysis - fragment informatics platform for collaboration and progression**

For the rapid dissemination, curation and annotation of fragment data, we have developed Fragalysis Cloud (http://fragalysis.diamond.ac.uk), which publicly hosts all XChem’s collaborative fragment results, alongside those of a few friendly users. Its scope is deliberately fragment-specific, encompassing analysis, progression, collaboration, and dissemination.

Important features include:

* Fragment structures are pre-curated, so that fragment binding can be directly analysed and compared.
* Views of the data can be shared as URLs (snapshots) that precisely reconstitute what the user sees, and thus streamline collaboration.  This feature was hardened in the COVID [Moonshot](https://postera.ai/moonshot/), where it was heavily used to assist cycles of compound design in a globally distributed team.
* Compute-heavy algorithms can be simply executed, and results reviewed and shared, by pre-configured GUI tools. This provides a vehicle for hosting cutting-edge published algorithms, deployed and hardened in collaboration with their authors.
* Primary and secondary data is disseminated by FAIR mechanisms, in downloadable zip archives that include self-documenting PDFs and snapshot URLs to explain the context and history of the download.
* Data can be kept private, for collaboration, or made public, for dissemination.

After trialling the curation and selection, by medicinal chemists for real-world drug discovery, where algorithmic follow-up structures were rapidly determined in nine weeks, of algorithm designs uploaded to the right-hand-side of Fragalysis, further refinements were made to improve the: searchability of uploaded design sets, navigating the designed compounds for review and updating the code to meet the significant changes made to the codebase and database models for the development of the left-hand-side data views **(Figure 4).** A screenshot of a computer

Description automatically generated

***Figure 4. (A)*** *The Fragalysis Cloud platform seeks to provide comprehensive support for users to readily access leading-edge tools for fragment analysis and progression, including collaboration and FAIR dissemination tools. (B) Through the ASAP project and its medicinal chemistry expertise, we are evolving a fast hit-progression process, exploring ways to streamline it and generating experimental feedback to harden and improve the algorithms being deployed in Fragalysis.  Illustrated here is the first exemplar, performed June-August 2023.*

The code has been professionally developed as a fully open-source web-stack (https://github.com/xchem) with playbooks that make it launchable by independent laboratories with cloud expertise. Development was funded from diverse sources, including IMI/UltraDD, the Ada Lovelace Centre, the Wellcome Trust, NIH/NIAID and Diamond internal allocations. The longer-term ambition is to assemble a Consortium to provide both funding and, equally critically, industrial users and expertise.

Fragalysis provides key data infrastructure for the ASAP project, which in turn is proving a critical testing ground and thus helps rapidly harden the ecosystem. By performing all fragment-to-hit design activity within the platform, we obtain experimental data and user feedback to assess algorithmic weaknesses, prioritise features and evolve a Design-Make-Test process that can be realistically provided to XChem users as part of the user programme. This is now being trialled with friendly users, including those recruited from another large Diamond/XChem partnership, the IMI/EUbOpen consortium.

* **PanDDA2 - Fast and automated fragment finding and building**

The DPhil thesis of Conor Wild, an XChem-affiliated student at University of Oxford (2018-2023), focused on revisiting the original PanDDA formalism, also developed at XChem, firstly to understand and correct its failures and weaknesses, and secondly to implement the fully automated density modelling that had appeared feasible once its power for signal improvement had become evident.

This work led to PanDDA2 (<https://github.com/ConorFWild/pandda_2_gemmi>), a reimplementation of the original concepts, refactored to (1) bring about dramatic speed-ups; (2) implement pre-clustering of datasets according to real-space variation, as an alternative to the reciprocal space clustering method developed by Ginn; (3) introduce fast ranking of putative hits using shape recognition through machine-learning of historic XChem data **(Figure 5A)**; (4) achieve fast auto-building using a new statistically robust scoring function **(Figure 5B,C)**; and (5) robustly test the package on historic XChem data.

The new software has already helped find crucial hits for AViDD antiviral projects, and early validation work yielded crucial insights for the VP1 polyoma capsid experiment.A collage of images of a cell

Description automatically generated

***Figure 5 (A) Previously undiscovered hits revealed by evaluating multiple ground states with CNNs in PanDDA2****. (B) Autobuilding of fragments drives throughput by improving event ranking and decreasing manual building. RMSD to reference structures shows good models are produced in 2/3 cases. (C) We established that RSCC is a poor scoring function for fitting ligands into event maps because their formalism amplifies the noise in the underlying maps. Instead, we use a metric that rewards matching of the overall shape, which more closely reflects how humans do this: these shape-based metrics and signal-to-noise scores significantly improve ranking of building events****.***

* **XChem Align - Auto-analysis of sites and binding modes**

To fully understand the various binding sites revealed by a fragment screen, and to understand packing or model quality artefacts, requires the ensemble of fragments and literature structures to be placed in common reference frames, spanning diverse asymmetric units, crystal forms and binding site conformations. This includes moving electron density, protein neighbourhood and other packing artefacts to the various reference frames. This task is algorithmically complex, but now implemented in XChemAlign (XCA), which heuristically infers the various relationships and variant observations, preparing them for curation and finally upload to Fragalysis.

The way experimental structural data is displayed in Fragalysis went through a major revision with the development of XChem Align (XCA), a novel algorithm for aligning structural data by biological assembly and crystal forms, that will allow algorithm and manual compound designers to access merge and linking opportunities, not limited to crystallographic artefacts, that are based on biologically relevant observations **(Figure 6)**. This is a significant revision of Fragalysis with an entire new stack required with legacy data hosted on a separate stack.

A close-up of a diagram

Description automatically generated

***Figure 6.*** *XChem Align locally aligns binding sites to enable the analysis of merging opportunities even for fragments observed in different crystal forms.*

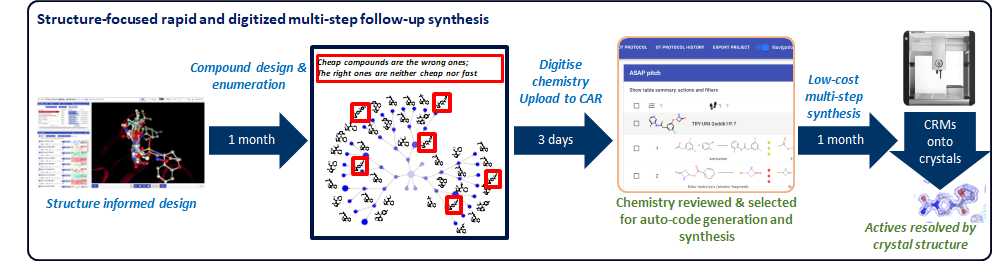
**Aim 4: Compound design and TEP validation**

Following the fragment progression, it was possible to design new compounds to be tested using the

* **Fragment progression: CAR for microscale chemistry and SAR-colouring**

To further accelerate fragment progression, we have been developing the CAR platform (“chemist-assisted robotics”), that combines AI-driven compound design derived and the testing of crude reaction mixtures by XChem. Since CAR can generate 100s-1000s of crude reaction mixtures, this therefore rapidly achieves a first iteration of the Design-Make-Test-Analyse (DMTA) cycle that can yield the volume and quality of data required to ensure the “Analyse” step can progress to high potency with very few (ideally no) additional cycles, using AI-based generative methods.

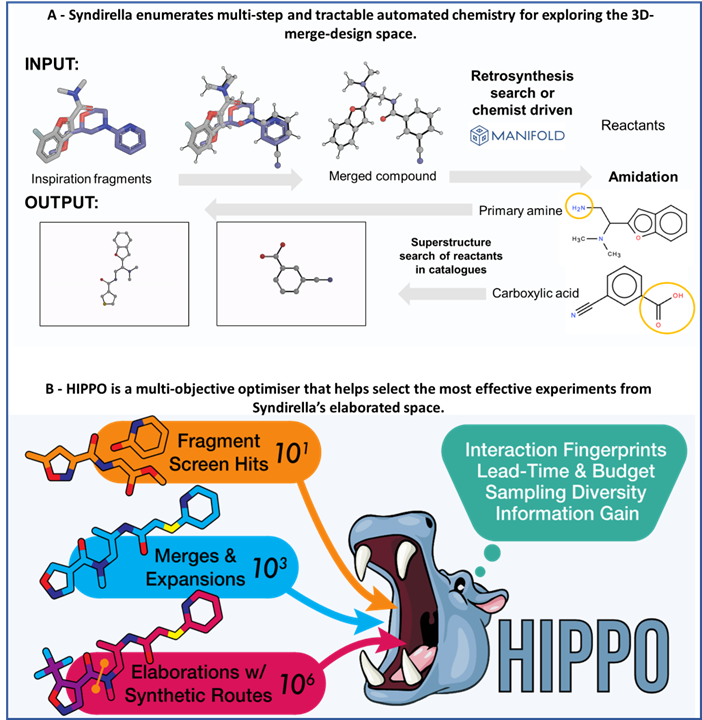
Specifically, CAR currently comprises a cloud-based web-application for curation of reactions; an interface to retrosynthesis engines; digitized chemistry (19 reactions tested and encoded); automated code generation for reviewing, selecting, and executing the automated synthesis of follow-up compounds; protocols for executing code on an OpenTrons low-cost liquid handler; and a QC tool for semi-automated analysis of MS results (**Figure 7**).

**

***Figure 7.*** *CAR is a digitised chemistry platform for the rapid selection of follow-ups for synthesis, via automated code generation, on a low-cost liquid handler.*

The compound-design, we are additionally developing, seeks to augment the algorithm for pure fragment merges. These use and retain only the original fragment information, and therefore they cannot on their own sufficiently explore the merge-design space, since adding or removing fragment bits may have a significant impact on the structure-activity relationships (SAR), that ideally needs to be understood early in the early discovery.

Instead, we can vastly expand the merge-design space, to “colour in” the SAR (hence “SAR-colouring”). We combine merge-enumeration with retrosynthesis prediction from PostEra’s Manifold tool: the merge algorithm provides the iteration of designs, whereafter each design’s synthetic pathway(s) are enumerated by Manifold, and finally the reactions are elaborated through searching for purchasable analogues of the building blocks and substructure searches (**Figure 8A**). Finally, the most effective experiments are selected (**Figure 8B**), and fed to synthesis by CAR.



***Figure 8.*** *(A). Syndirella elaborates multi-step and tractable chemistry for CAR though an exhaustive substructure search, accessed through Postera’s Manifold system, for massively expanding the chemical space for merge-based-designs. (B) HIPPO effectively samples, using a multi-objective optimization of weighted properties such as building block cost, lead time and fragment interactions, for execution on CAR*.

**C. Significance** (1 page max) Write only about the completed TEPs in Y2

Crystallographic fragment screening is one of many techniques used for fragment hit identification, accounting for 8% of examples of hit-to-lead projects published in J. Med. Chem. in 2022 (<https://doi.org/10.1021/acs.jmedchem.3c02070>). A recent analysis of 18 successful fragment-to-lead case studies in 2022 reported that 83% (15/18) generated a structure for the fragment, 78% (14/18) generated a structure for the lead and 100% (18/18) of entries used structural information during lead optimization.A collage of a graph

Description automatically generated

Furthermore, in 156 hit-to-clinical campaigns published in J. Med. Chem. between 2018 and 2021 65% of campaigns reported used Structure-Based Drug Design as a key enabling technology. This increased to 100% when starting with fragment hits (<https://doi.org/10.1021/acs.jmedchem.3c00521>).

This illustrates that structural enablement is critical for efficient hit-to-lead development and lead optimization.

ASAP utilizes structure-enabled drug discovery processes and the structural biology core is essential in providing the capability of using high-throughput X-ray crystallography to structurally enable antiviral targets pursued by its discovery programs, providing X-ray structures of apo and liganded targets to support Projects 2, 3, 4, and 5.

*A screenshot of a project

Description automatically generated*

Additionally, rapid public structural data release acts as a public force multiplier to reduce costs to discovery as was demonstrated by the discovery of the SARS-CoV-2 Main protease inhibitor Ensitrelvir (<https://doi.org/10.1021/acs.jmedchem.2c00117>)).

**D. Plans**

Plans for the next project period include:

* Finalize and fully document the full list of Targeting Opportunities, and initiate all corresponding TEPs.
* Fully complete 8 advanced TEPs (goal is 10 total) with complete sign-off by SAB and handover to Biochemistry and Structural Biology Cores
* Validate a robust biophysical experimental assay using Creoptix platform for testing candidate compounds with antiviral activity
* Finalize deployment of Electronic Lab Notebook with streamlined TEP publication
* Develop and demonstrate (or disprove) new methodology for generating and exploiting diverse crystal polymorphs for each target protein.
* Progress fragment hits from three (3) TEPs using the novel SAR colouring approach with CAR microscale chemistry and hit enumeration by Fragment Knitting and Syndirella..

