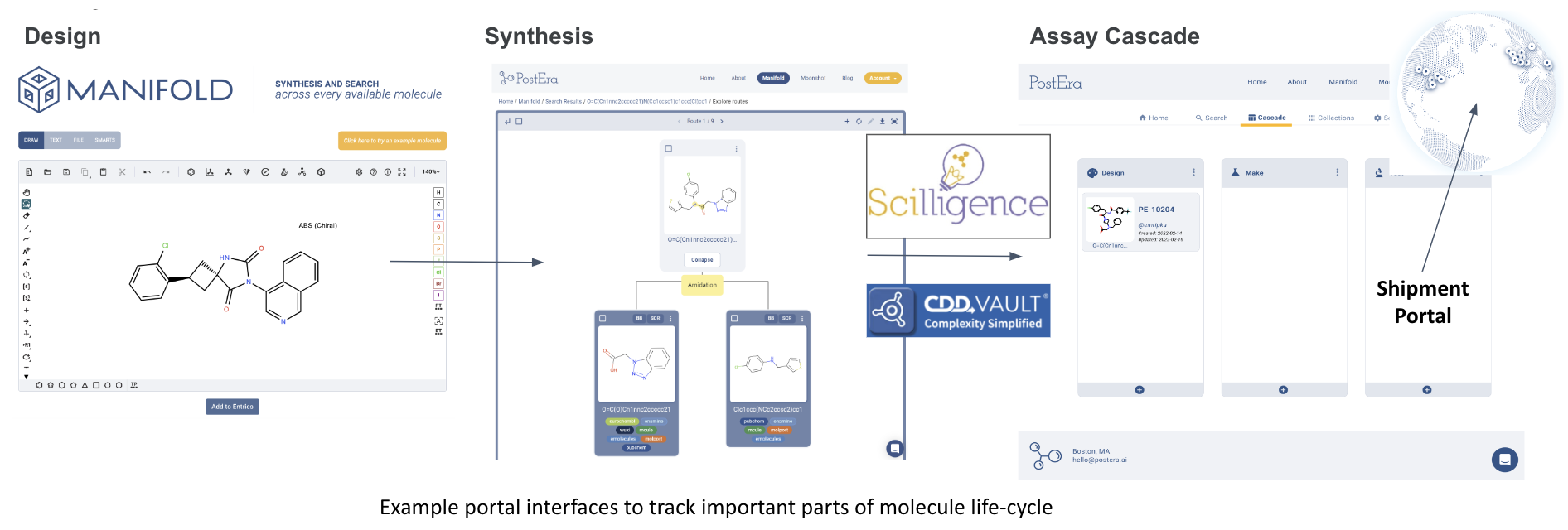
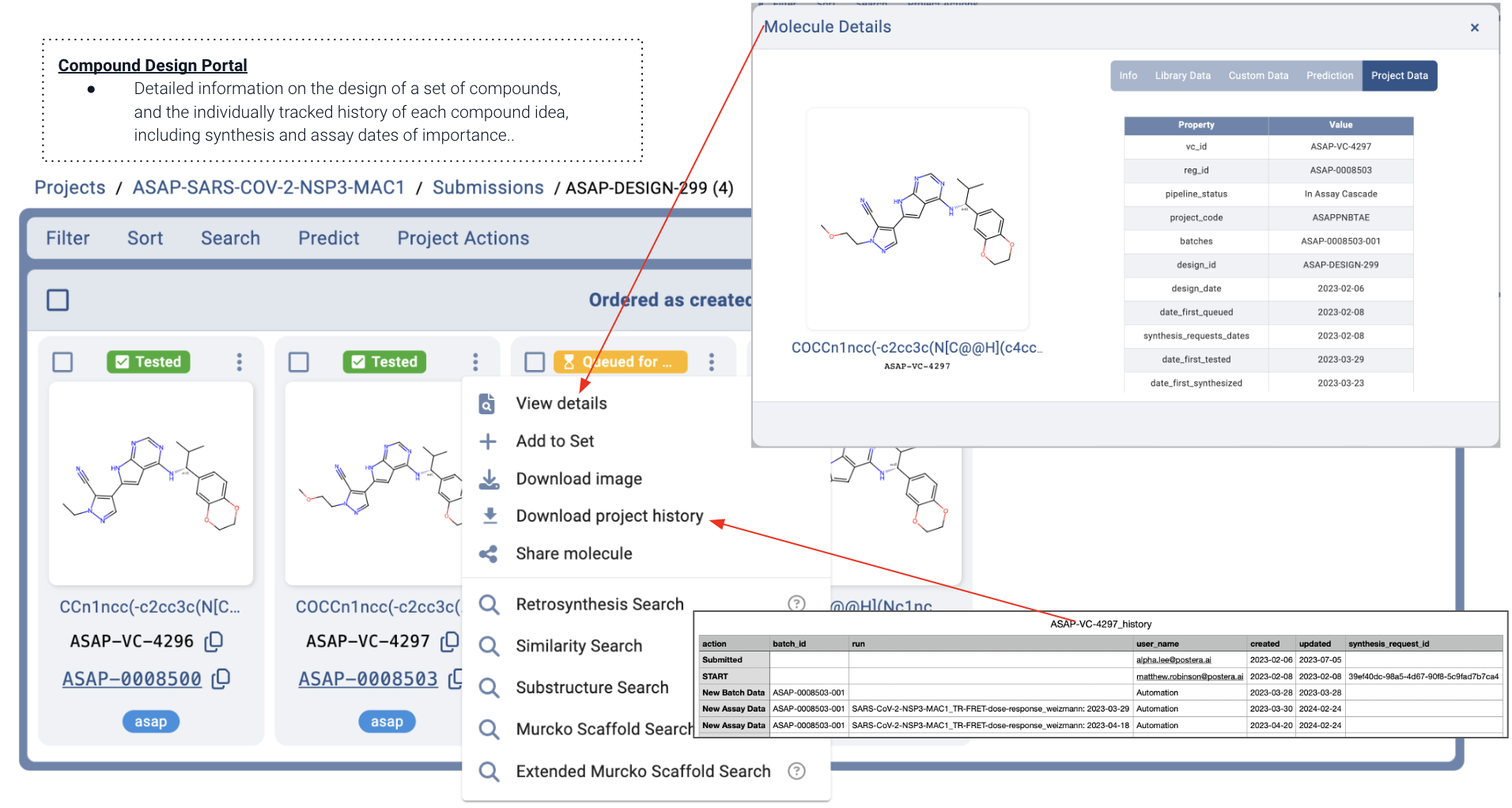
**B. Studies and Results**

The ASAP Data Core has made significant progress this past year on advancing a platform to completely capture molecular data originating from initial medicinal chemistry design to synthesis and subsequent testing in multiple downstream assays. This platform allows the ASAP discovery consortium, which involves expertise from labs around the world working in collaboration on logistically and scientifically complex medicinal chemistry projects, to efficiently progress drug discovery campaigns while also capturing data for widespread dissemination to the larger drug-discovery community. As of now, the ASAP data platform allows for detailed tracking of medicinal and computational chemistry work all the way from ideation to registration and subsequent assays, as well as the myriad shipments needed for this scientific discovery work. All of this extensive data infrastructure allows us to carefully track and organize now over 25,000 individual assay readouts from over 100 experimental assays generated by 9 different experimental labs on upwards of 10,000 physical compounds on the six different drug discovery programs that have advanced to or past hit-to-lead stages in ASAP.

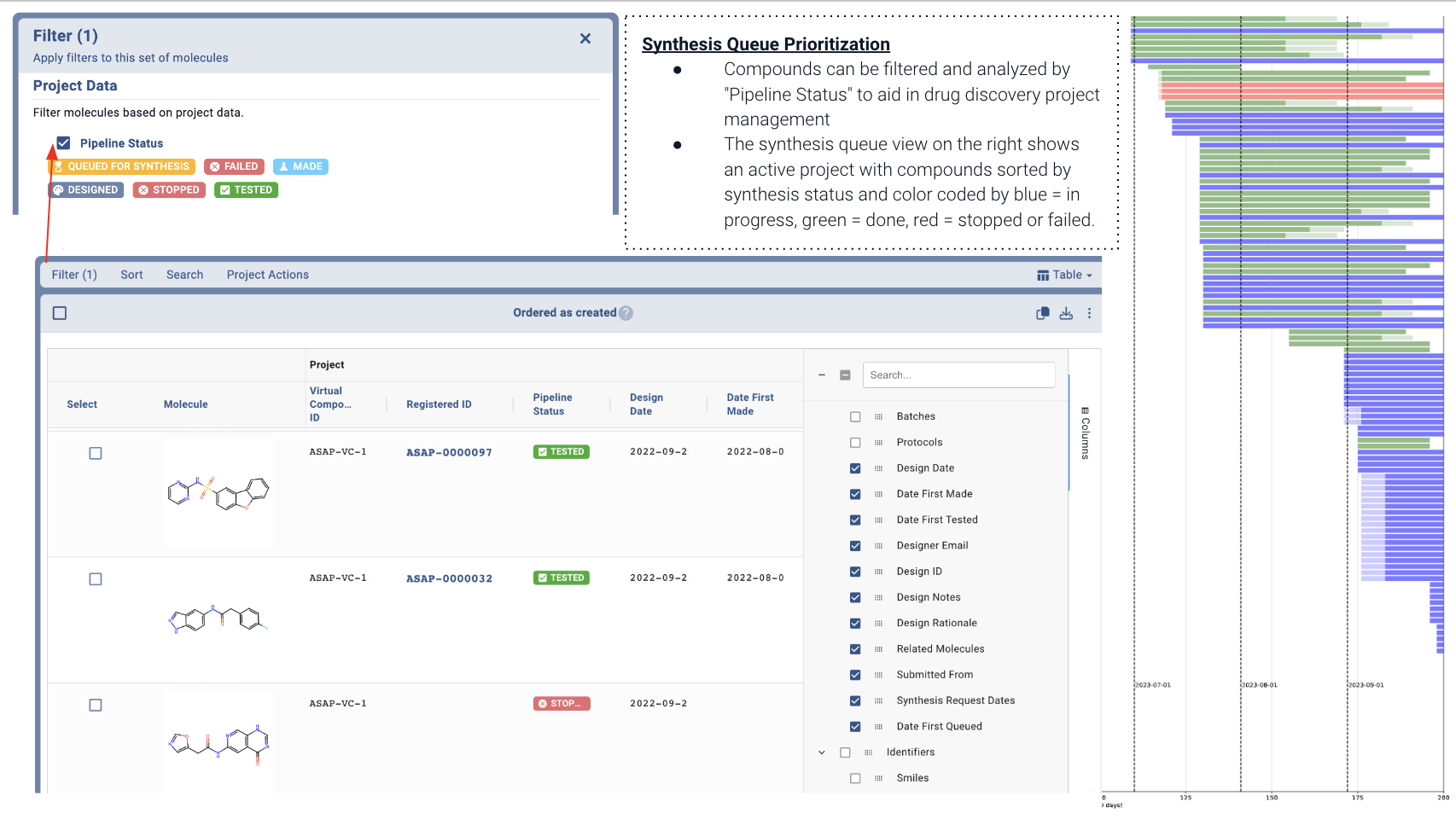


***Figure 1****: An overview of the ASAP data platform, which captures info on all facets of the design-make-test cycle. External CDD Vault software is used for compound registration and assay data uploads, while Scilligence ELN software captures synthesis data.*

Key advances to the tracking of design and synthesis priority queues have been central to the latest improvements in this data infrastructure. Medicinal chemists working on ASAP programs are using the latest techniques in machine learning driven drug discovery and parallel synthesis to drive programs in aggressively quick timelines. However, this also comes with the need for tracking potentially hundreds of compounds across parallel and singleton chemical synthesis efforts, and prioritizing thousands of designs that have been scored by machine learning models. Thus, we have had to build a design tracking portal that allows for efficient handling of these data management tasks. Figure 2 below shows an example of an individual "design set" and the detailed compound tracking contained therein. Figure 3 additionally shows some of the functionality built for prioritizing queues for compound synthesis and individual assays.



***Figure 2****: An overview of a design submission within the ASAP design portal data infrastructure. Within the design are 4 compounds suggested by the medicinal chemist to be considered for testing of the design hypothesis. Each compound has easily accessible details regarding its submission (top-right) as well as an exportable table that will give a complete history of every design, synthesis, and testing action performed on the individual compound.*



***Figure 3****: An overview of the platform infrastructure for prioritizing compounds to advance in through the design-make-test pipeline and larger assay cascade. The Gantt-like chart on the right provides a live look of synthesis load on each chemistry team and the observed cycle time for progressing from design to registration.*

Beyond building infrastructure to extensively track design and synthesis efforts, the Data Core team has also put significant effort into onboarding partner organizations to its data platform, such that all essential compound data enters a robust, organized database rather than living in scattered spreadsheets. These organizational initiatives allow for data obtained from biochemical, antiviral, and ADMET testing to quickly go from experiment to inputs into machine learning and computational chemistry algorithms as well as the eyes of medicinal chemists for further design of compounds. Overall, this results in data from over 100 assays on over 10,000 compounds being easily retrievable at any given time by all ASAP chemists.

These data infrastructure efforts have involved extensive software engineering and chemoinformatics development efforts; however, this detailed, robust data organization is central to the ASAP Consortium's mission of providing useful data to the larger antiviral discovery community. ASAP has been able to release a subset of key data on key targets including the MERS-CoV Mpro, SARS-CoV-2 Mpro, and SARS-CoV-2 nsp3 macrodomain programs. These data releases, for example the [MERS-CoV/SARS-CoV-2 Main Protease (Mpro) program: March 2024](https://asapdiscovery.notion.site/MERS-CoV-SARS-CoV-2-Main-Protease-Mpro-program-March-2024-35dfdcf42ffb4d1ebc8a164d681b68eb) release, provide extensively annotated data to the community and also link out to [well-defined experimental protocol information](https://www.protocols.io/view/mers-cov-main-protease-mpro-fluorescence-dose-resp-eq2ly7r1rlx9/v3) that may be of further interest to the community. Further efforts are on going to directly deposit the [SARS-CoV-2 nsp3 macrodomain data](https://asapdiscovery.notion.site/SARS-COV-2-Macrodomain-Mac1-Inhibitor-Data-Release-9e454e794ee44df693251477ff8686e9) into industry leading data repositories such as ChEMBL. All of these efforts hope to help the larger antiviral drug discovery community benefit from the extensive ASAP data core efforts to build out a next-generation data platform.

**C. Significance**

Modern small-molecule drug discovery has progressed greatly due to the advent of improved machine-learning and computational chemistry design tools, as well as a vast, decentralized network of research organizations offering synthesis and experimental testing capabilities. With these advances comes increased logistical complexity and the need for tooling to support this rapid data generation across the globe. Chemists are burdened with prioritizing medicinal chemistry designs, tracking ongoing synthesis efforts, and arranging the further characterization of promising compounds through downstream assays. With these tasks comes the additional need for organizing robust data cleaning and upload procedures, while also providing details such as shipment tracking information. The ASAP Data Core has built a next-generation drug discovery platform that enables our chemists to perform these tasks using well-defined, carefully-tracked methodology that delivers key decision-making data when and where they need it.

Overall, the Data Core provides this carefully annotated, well-organized data to primarily support the following ASAP projects:

* **PROJECT 2 (P2)** : TARGET ENABLEMENT
* **PROJECT 3 (P3)** : FRAGMENT-TO-LEAD AND TARGET VALIDATION
* **PROJECT 4 (P4)** : COVALENT TARGETING STRATEGIES
* **PROJECT 5 (P5)** : LEAD OPTIMIZATION
* **PROJECT 6 (P6)** : PRECLINICAL DEVELOPMENT AND TRANSLATION

showing that the Data Core is involved in many facets of molecular data storage and annotation all the way from early fragment screen follow-up to the advanced characterization efforts of advanced preclinical compounds. Furthermore, our aforementioned efforts in public releases of data on ongoing programs shows our commitment to providing valuable data to the broader antiviral drug discovery community.

**D. Plans**

In pursuit of the Data Core's overall aims to provide state-of-the-art data infrastructure to ASAP consortium drug-discovery programs, the following goals are planned for the upcoming year:

Improving the "Assay Requests Framework" for chemists

* Currently, requests for Assays and Synthesis are made via a combination of several different forms. Our goal is to unify this into a centralized requests framework.
* This goal will be we started by completing the details around moving Synthesis requests fully into the main platform workspace, followed by assay specific requests, which will then be tracked more easily than by current methodology.

Improved validation of protocol definitions and uploaded data

* Currently verification of uploaded assay data remains somewhat manual. We have built prototypes of computational tools that can automatically verify the validity of experimental data uploads from third parties.

Depositing data directly into industry leading data repositories for use by the drug discovery community

* Data from the SARS-CoV-2 nsp3 macrodomain program will be deposited directly into the ChEMBL data depository. This will involve data from over 1000 compounds on over 10 experimental assays including biochemical, antiviral, and ADMET data.

More detailed inventory tracking

* As more advanced enter into advanced preclinical characterization, this often involves measurements from upwards of 30 distinct assays on a compound from multiple experimental sites. Thus, there is a need for improved tracking of which specific vials of which specific batch of compound went to each experimental testing site. Future efforts will aim to greatly improve this sample tracking.