

CDD Vault - A Collaborative Data Management Platform for PROTAC Drug Discovery

Introduction

Targeted protein degraders, such as proteolysis targeting chimeras (PROTACs), leverage the cell's natural ubiquitin-proteasome system to degrade target proteins. Early work has shown that PROTACs can improve the affinity and selectivity of their parent small molecule drugs. Unlike small molecules, each PROTAC molecule can induce degradation of multiple target protein molecules. Therefore, PROTACs can be used at subnanomolar concentrations. However, for PROTACs to work *in vivo*, they must enter the target cell intact and bind to both the target protein and an E3 ligase to induce ternary complex formation.

This is no small feat as PROTACs are flexible, bifunctional molecules with a high molecular weight and high polar surface area. These properties lead to poor solubility and membrane permeability, which may negatively affect absorption. Scientists have managed to overcome these pharmacokinetic challenges and advanced two orally available PROTACs, one targeting the androgen receptor and one targeting the estrogen receptor, into Phase 2 clinical trials.

PROTACs have the potential to target the 'undruggable' proteome. Researchers are rapidly developing new PROTAC technology by solving problems associated with design, efficacy, specificity and safety.

Data Challenges for PROTAC Discovery

The data informatics requirements for PROTAC discovery researchers include:

Central Database: Creating and maintaining a reliable 'single source of truth' to store chemical and peptide molecules along with characterisation data and the ability to record and visualize connections between the different fragments, track modifications and batches.

Visualization Analysis: Visualization tools for chemical structures, amino sequences, protein crystal structures, Western Blots/ co-Immuno precipitation, amino acid sequence search and alignment, amino acid and peptide mapping to proteome, and so on.

Analysis Software: Various self-built, freeware and/or enterprise software pipelines to import, normalise, analyse and interpret raw data.

Digitalization and Secure Access: An Electronic Lab Notebook (ELN) with inter-connectivity between all data, accessed securely by selected members in real time as and when data is uploaded.

Collaboration and Connectivity: Sharing data selectively with collaborators in a secure way in real time.

Finally, connectivity among all the databases, internal and external visualisation tools and analysis software and the ELN are critical for the day-to-day functioning of any modern discovery laboratory.

CDD Vault is a comprehensive data management platform catering to the specific needs of PROTAC drug discovery researchers. This application note outlines how researchers can manage their data by implementation of CDD Vault.

Getting Started

1. Register entities and calculated properties

- CDD Vault is ideal for the registration of PROTAC molecules. This includes the registration of the tripartite PROTAC molecules as well as registration of individual sub-entities like the E3 ligands, chemical linkers and warheads which could be small molecules, antibodies, etc.
- Each type of entity can be registered in its own CDD Vault which can then be interconnected and cross-referenced with each other. For e.g. researchers can have separate CDD Vaults for E3 ligands, linkers, and warheads, and also a main PROTAC Vault containing cross-referencing links to each of the sub-entities residing within their specific Vault. See Fig.1



Fig 1: Registration of a PROTAC molecule PT-0000435, showing its chemical structure, synonyms/other nomenclature, its linker, E3 ligand, warhead, associated ELN entry and assay data.

- CDD Vault supports a range of warheads. Users can register a chemical structure of a small molecule, sequence of amino acids (both natural and unnatural) as well as potential nucleotide sequences for warheads.
- When a small molecule is registered, physicochemical properties are calculated automatically. These include molecular weight, logP, logD, Lipinski properties and so on.
- When a peptide/protein sequence is registered, CDD Vault automatically calculates properties like peptide length, molecular weight (MW), pKa, % amino acids and so on.
- Researchers can also register non-therapeutic entities with the aim of creating databases for research materials like cell lines, plasmids, commercial vectors, commercial antibodies, bacterial clones and so on (see fig 2 for an example of a cell line Vault).
- In such cases, characteristics such as base pair length (BP), %GC/AT, MW and composition are also automatically calculated where applicable.

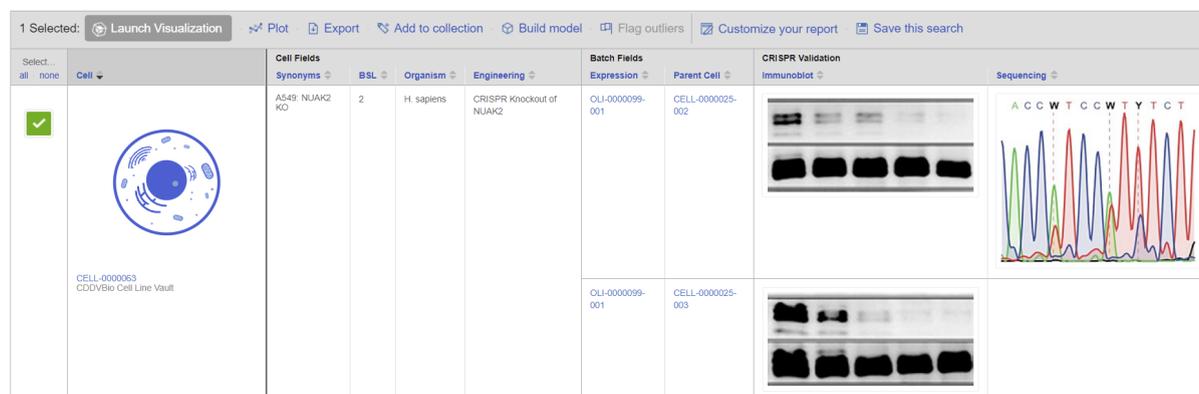


Fig 2: Screenshot of a cell line Vault customized to show an engineered cell line (the registered entity), synonym/nomenclature, biosafety level, organism, modification method such as CRISPR

Cas9, oligo expressed for cell engineering, parent cell line used as well as validation assay data showing an immunoblot and a sequencing plot.

2. Create batches and associate contextual information

- Each registered PROTAC, and each constituent entity, is assigned a unique identifier by CDD Vault which allows researchers to track batches explicitly via an entity-batch identifier.
- CDD Vault automatically searches previously registered entities and prevents duplications.
- Users can associate data in the context of specified batches so it's easy to track batch-to-batch variations.
- Ability to track batch-to-batch variability facilitates root cause analysis for any deviations.
- Researchers can also store and visualise all characterization data of a given entity in a single platform. As shown in fig 1, for a PROTAC molecule PT-0000435, chemical structure, synonyms/other nomenclature, its linker, E3 ligand, warhead, associated ELN entry and assay data are all stored and shown in a single view. Additionally, cross-references to sub-entities and ELN entries are live links to its respective record and can be clicked to discover additional details regarding each sub-entity or ELN entry.

3. Storage and visualize assays

- Each registered PROTAC can be associated with assay results originating from a wide variety of studies.
- These may include degradation assays, growth inhibition studies, ternary complex assay, Western Blots, other *in-vitro* and *in-vivo* assays (such PKPD/ADME testing etc.), FRET assays for protein-protein interactions and so on. Fig 1 shows data associated with the PROTAC molecule PT-0000435 resulting from a degradation assay, a growth inhibition study and a ternary complex assay.
- Users can track all assay data related to critical quality attributes of a PROTAC entity as a function of modifications such as altered linkers, warhead potency, antibody or peptide fragment or post-translational modifications, etc.

4. Document experimental details in the ELN

- The CDD Vaults ELN can be used by researchers to store standardized SOPs and record all experimental details across departments.
- Full integration with Microsoft Office 365 allows scientists to attach, preview and edit MS office files within the ELN.
- Image files can be displayed directly within the ELN entry in various sizes.
- Any file format, including proprietary files which require their own application, can also be stored in the ELN.
- Users can link multiple entities from the registration system within an ELN entry. An example of this is seen below (fig 3) which shows an ELN entry for the Synthesis of PROTAC PT-0000429 which contains cross-reference links (seen in blue) to PT-0000426, PT-0000428 and PT-0000427 that are the starting materials for the PROTAC synthesis.

Synthesis of PT-0000429



ID: 364450

Project
PROTAC

Study Type
Synthesis

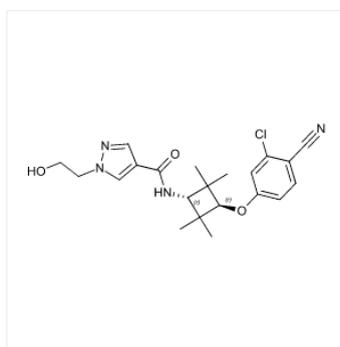
Study PI
Etta

Study Initiation Date
2021-10-11

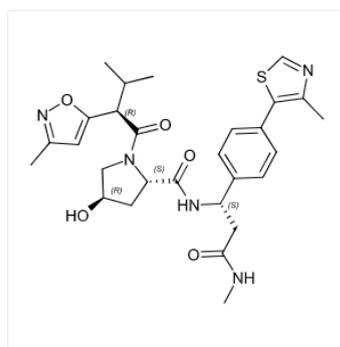
Purpose

The purpose of this entry is to detail the synthesis of PT-0000429, a potent PROTAC molecule against the Androgen Receptor (AR), which binds to the androgenic signaling molecule within the cell, using the VHL E3 Ligase for ubiquitination and degradation.

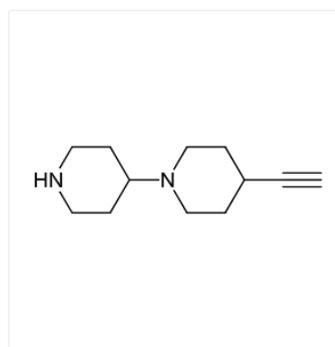
Starting Materials



PT-0000426



PT-0000428



PT-0000427

Fig 3: Partial screenshot of an ELN entry detailing an experiment for the synthesis of PROTAC PT-0000429.

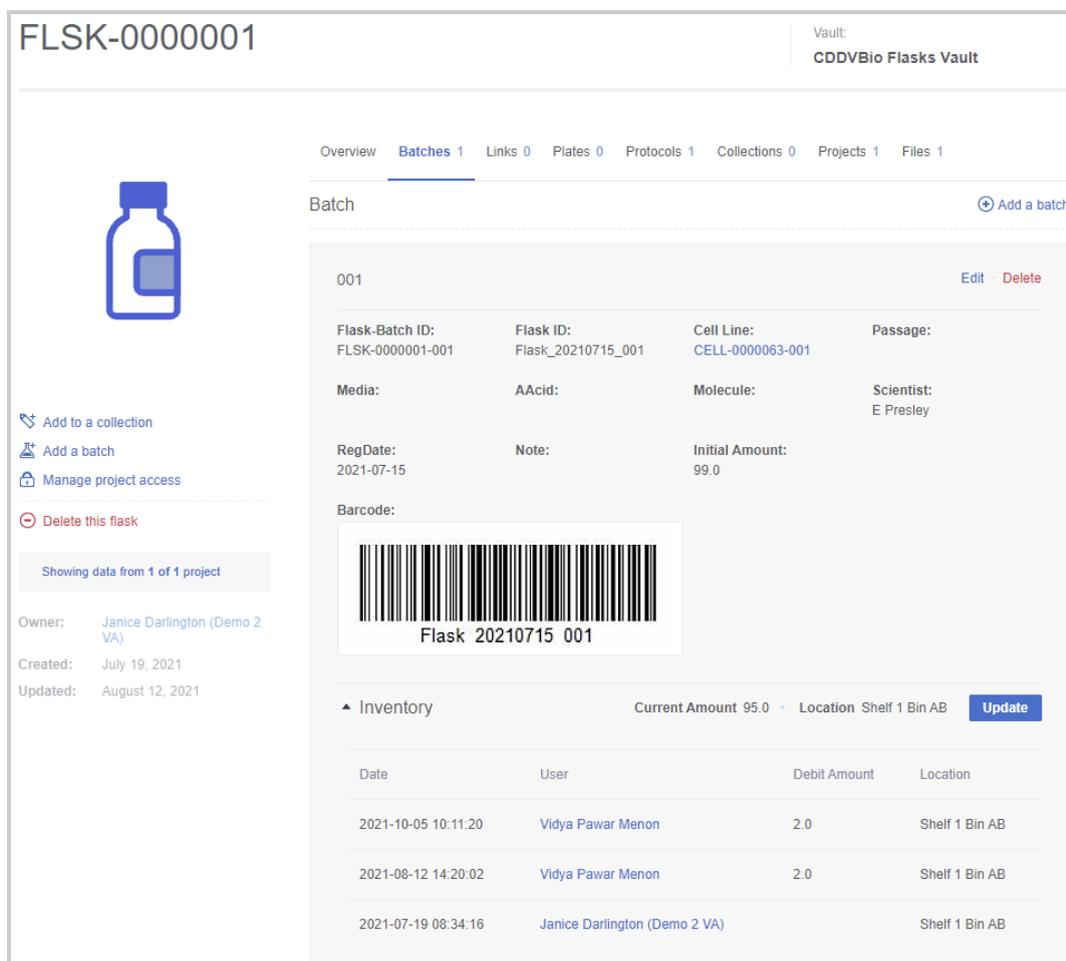
5. Mine data

- Researchers can search their data by name, chemical structure/peptide sequence and easily locate records by entities or sub-entities
- One can also search for keywords in ELN entry bodies and attachments.
- For example, researchers working with PROTACs can search for a linker and/or a warhead of choice. This will bring up all entities created with the queried linker and/or warhead along with all associated data and ELN entries.
- Users can also look up all data related to each chemical entity, for example visualize all assay results for a specific E3 ligand
- Design successful linkers by comparing candidates and searching for desirable linker characteristics.

6. Track location and quantities of entities and reagents

- Researchers can capture the location and quantities of their entities and other reagents including features such as date, vendor and batch using the inventory module (See fig 4).
- One can capture and track containers, vials, dewars, flasks, etc within the CDD Vault Inventory.
- Scientists can use barcodes to manage their inventory. One can export identifiers and key metrics to their barcode template and print labels. Users can also locate samples or pull up a list using a barcode scanner.
- A debit history provides a full record of the date, person, amount and/or location of each update.

- Users can see inventory data alongside other batch characteristics and assay data associated with the batch.



The screenshot displays the CDD Vault interface for a specific batch (FLSK-0000001). The interface is organized into several sections:

- Batch Overview:** Shows the batch ID (001) and provides options to edit or delete the batch.
- Batch Details:** A grid of key-value pairs including:
 - Flask-Batch ID: FLSK-0000001-001
 - Flask ID: Flask_20210715_001
 - Cell Line: CELL-0000063-001
 - Passage:
 - Media:
 - AAcid:
 - Molecule:
 - Scientist: E Presley
 - RegDate: 2021-07-15
 - Note:
 - Initial Amount: 99.0
- Barcode:** A barcode representing the flask, labeled "Flask 20210715 001".
- Inventory Table:** A table showing the current amount (95.0) and location (Shelf 1 Bin AB). The table includes columns for Date, User, Debit Amount, and Location.

Date	User	Debit Amount	Location
2021-10-05 10:11:20	Vidya Pawar Menon	2.0	Shelf 1 Bin AB
2021-08-12 14:20:02	Vidya Pawar Menon	2.0	Shelf 1 Bin AB
2021-07-19 08:34:16	Janice Darlington (Demo 2 VA)		Shelf 1 Bin AB

Fig 4: Snapshot of batch level inventory fields from a Vault customized to register cell line culture flasks.

7. Link to other tools

- CDD Vault's RESTful Application Programming Interface (API) can be used to integrate with external databases and analysis tools.
- Scientists can easily export data from CDD Vault in Excel, CSV or SDF file formats.
- Likewise, users can use CDD's import data wizard to load data into their Vault in bulk.

8. Collaborate and share data securely

- Data in CDD Vault can be securely shared with colleagues and collaborators.
- These can be entire peptide libraries for instance or selective assay data depending on the collaborator.
- Researchers can share data in real time and avoid sending data over unsecured email.
- Data in CDD Vault is organized by projects. Individual users are assigned to projects to easily control which data they are allowed to access.
- Each member is also allocated a user role (Read-Export versus Read-Add) which provides project permissions that control how each member interacts with the data that they can access.



CDD Vault is a modern data management platform with an integrated ELN and inventory that allows PROTAC discovery researchers to bring together data from various experiments, instruments and analysis software. Biologists, chemists and computational scientists use CDD Vault as a 'single source of truth' as well as to efficiently collaborate and share data.

To get started, please contact us at info@collaborativedrug.com to set up a free trial of CDD Vault.

REFERENCES

Arvinas Pipeline. Oncology-Immuno-oncology. (Online) Accessed July 15, 2021. Available at: <https://www.arvinas.com/pipeline-programs/pipeline>

Cantrill et al. Fundamental aspects of DMPK optimization of targeted protein degraders. *Drug Discov Today*. 2020;25(6):969-982. doi: 10.1016/j.drudis.2020.03.012.

Bondeson et al. Lessons in PROTAC Design from Selective Degradation with a Promiscuous Warhead, *Cell Chem Biol*, 2018;25(1):78-87.e5 <https://doi.org/10.1016/j.chembiol.2017.09.010>.

Ishida T et al. E3 Ligase Ligands for PROTACs: How They Were Found and How to Discover New Ones. *SLAS DISCOVERY: Advancing the Science of Drug Discovery*. 2021;26(4):484-502. doi:10.1177/2472555220965528

Li et al. An integrated bioinformatics platform for investigating the human E3 ubiquitin ligase-substrate interaction network. *Nat Commun* 2017,8: 347. <https://doi.org/10.1038/s41467-017-00299-9>

Luh et al. Prey for the Proteasome: Targeted Protein Degradation-A Medicinal Chemist's Perspective. *Angew Chem Int Ed Engl*. 2020;59(36):15448-15466. doi: 10.1002/anie.202004310.

Pettersson M, Crews CM. PROteolysis TArgeting Chimeras (PROTACs) - Past, present and future. *Drug Discov Today Technol*. 2019;31:15-27. doi: 10.1016/j.ddtec.2019.01.002.

Troup et al. Current strategies for the design of PROTAC linkers: a critical review. *Explor Target Antitumor Ther*. 2020;1:273-312. <https://doi.org/10.37349/etat.2020.00018>