CDD Vault - A Collaborative Data Management Platform for Protein & Peptide Drug Discovery

Introduction

Protein and peptide therapeutics commonly include engineered native proteins, enzymes or hormones; synthetic peptide analogs; polyclonal (PAb) and monoclonal antibodies (MAb), nanobodies or other antibody fragments; antibody drug conjugates (ADC); recombinant fusion peptides; and so on. The global market for protein therapeutics is growing, both in terms of newly approved drugs and market size. Protein based biologics accounted for almost 25% of the new drugs approved by the FDA's Center for Drug Evaluation and Research in 2020. These new drugs contributed to the worldwide protein therapeutics market, worth an estimated $160 billion in 2020, and predicted to grow 6.5% a year.

Data Challenges for Protein Therapeutics Discovery

Given the diversity of peptide molecular structures, ranging from classical molecules (like native proteins or enzymes, PAbs, MAbs) or novel molecules (such as nanobodies) to multiple entity molecules (such as ADCs, alphamers, aptamers, etc.), the data informatics requirements for their discovery present unique challenges including:

Central Database: Creating and maintaining a reliable 'single source of truth' to store 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 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 modern informatics platform that enables scientists working with protein and peptide biologics to store, organize and manage their drug discovery data in a central location and share their data in a secure manner from the comfort of their workspace.

Getting Started

1. **Register protein/peptide entities**

   - CDD Vault supports the registration of a wide variety of protein entities which usually include mABs, pABs, ADCs, peptide fragments, fusion peptides, antibody fragments, nanobodies,

peptide aptamers, alphamers, and so on (See fig 1 for an example of registration of an ADC molecule).

- Users can register a sequence of amino acids, both natural and unnatural.
- When a peptide/protein sequence is registered, CDD Vault automatically calculates properties like peptide length, amino acid composition percentages, Molecular Weight (MW), and hydrophobicity.
- 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, etc. (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.

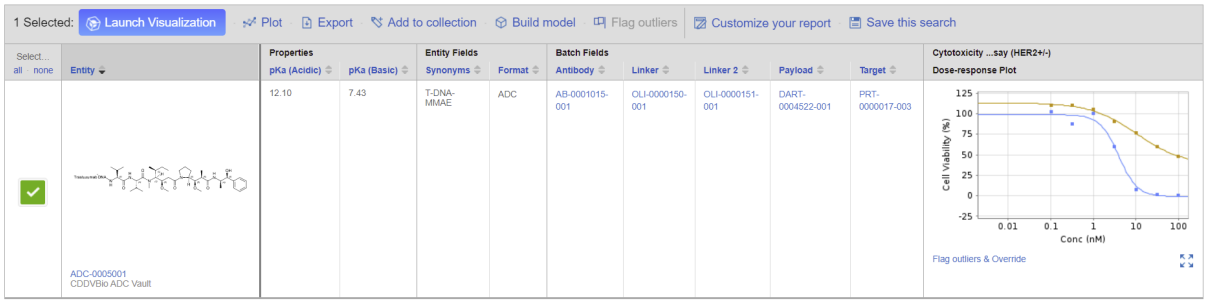

Fig 1: Registration of an example ADC molecule showing calculated properties such as pKa (acidic and basic), entity level fields such as synonyms and format, batch-level linking to constituent moieties - antibody, linkers and payload/small molecule.

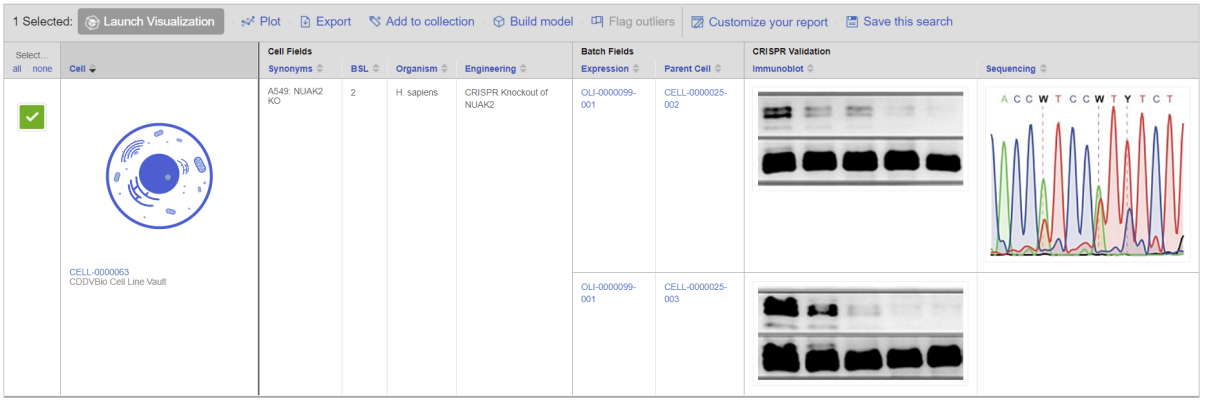

Fig 2: Screenshot of a cell line Vault customized to show an engineered cell line (the registered entity), synonym/nomenclature, biosafety level, organism, modification method such as CRISPR Cas9, oligo expressed for cell engineering, parent cell line used as well as validation assay data showing an immunoblot and a sequencing chromatogram.

2. Create batches and associate contextual information

- Each registered protein/peptide entity is assigned a unique identifier by CDD Vault which allows researchers to track batches explicitly via an entity-batch identifier.
- CDD Vault automatically checks previously registered entities to prevent 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 entity characterization within a single platform including protein structures, antibody heavy and light chains, B/T cell epitopes, MHC binding

predictions, proteasomal cleavage site predictions, immunogenicity predictions, surface hydrophobicity/hydrophilicity, isoelectric point, etc.

- Likewise, constituent moieties of fusion or conjugated molecules (such as ADC's) can be registered as individual entities which can then be batch-linked to each other. Each of the constituent moieties (i.e. antibody, linker and payload) can be registered as individual entities in separate Vaults and then linked together within an ADC Vault (See fig 1).

3. Store and visualize assays

- Each registered peptide/protein entity can be associated with assay results originating from a wide variety of studies.
- These may include *in vitro* cytotoxicity assays, *in vivo* studies, PKPD/ADME testing, Western Blots for confirmation of protein size and components, FRET assays for protein-protein interactions and so on. Fig 1 above also shows the plot from a cell viability assay.
- Users can track all assay data related to critical quality attributes of a therapeutic entity as a function of modifications, such as altered linkers, payload, antibody fragment, peptide fragment or post-translational modification, etc.

Fig 3: Partial screenshot of an ELN entry detailing an experiment for the design and validation of linkers used for ADC generation.

4. Document experimental details in the ELN

- CDD Vault's ELN is 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, edit, and share 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 be stored in the ELN.
- Users can link multiple entities in the registration system within an ELN entry. An example of this is seen in fig 3, where cell lines registered in the cell line vault and an ADC molecule registered in the ADC vault have been referred to, cross-referenced and linked in the ELN's Materials section.

5. Mine data

- Search data by sequence or name and easily locate records by sub-sequences or fragments of multi-partite entities.
- Search for keywords in ELN entry bodies and file attachments.
- For example, researchers working with ADC's 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 unnatural amino acids, which can be useful in studying and comparing the predicted effects of amino acid substitutions on protein properties

6. Track location and quantities of entities and reagents

- Capture the location and quantities of entities and other reagents including features such as date, vendor and batch using the inventory module (See fig 4).
- Store and track containers/vials/dewars/flasks.
- Users 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.

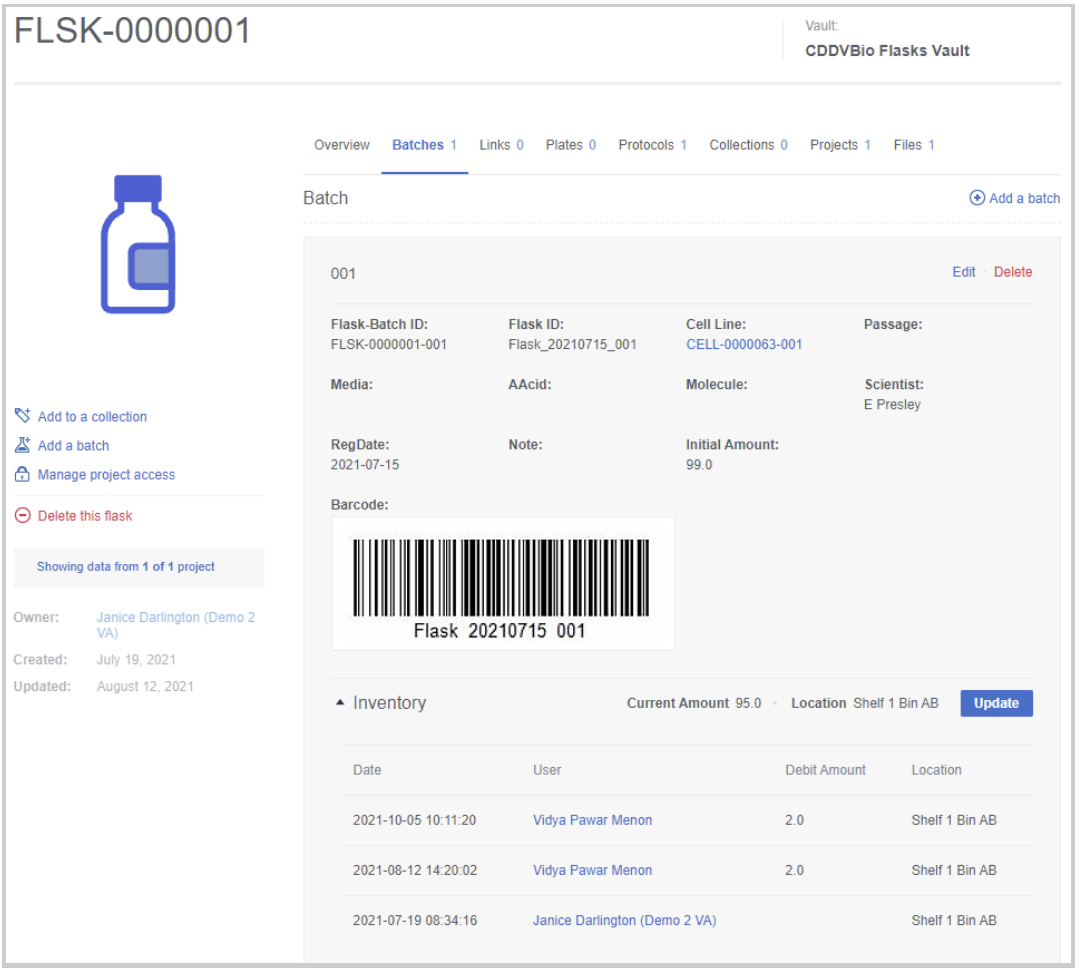

Fig 4: Snapshot of batch level inventory fields from a Vault customized to register cell line culture flasks.

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 via unsecure email.
- Data in CDD Vault is organized by projects. Members are assigned to projects to easily control data access and visibility.
- Each member can also be allocated further roles where roles act as 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 protein and peptide researchers to bring together data from various experiments, instruments and analysis software. Biologists, Chemists and computational scientists can use CDD Vault not only as a 'single source of truth' but also for efficient collaboration and data sharing.

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

REFERENCES

de la Torre and Albericio. The Pharmaceutical Industry in 2020. An Analysis of FDA Drug Approvals from the Perspective of Molecules. *Molecules* 2021, 26, 627. https://doi.org/10.3390/molecules 26030627

Fu et al. Immunogenicity of Protein Therapeutics: A Lymph Node Perspective. *Front. Immunol.* 14 May 2020 https://doi.org/10.3389/fimmu.2020.00791

Heaton. Challenges of Developing Novel Vaccines With Particular Global Health Importance. *Front. Immunol*. 14 October 2020 https://doi.org/10.3389/fimmu.2020.517290

Jawa et al. T-Cell Dependent Immunogenicity of Protein Therapeutics Pre-clinical Assessment and Mitigation—Updated Consensus and Review 2020 *Front. Immunol*. 30 June 2020 https://doi.org/10.3389/fimmu.2020.01301

Lau and Dunn. Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg. Med. Chem.* 26 (10): 2700-2707 https://doi.org/10.1016/j.bmc.2017.06.052

Xu X, Huang Y, Pan H, et al. Quantitation and modeling of post-translational modifications in a therapeutic monoclonal antibody from single- and multiple-dose monkey pharmacokinetic studies using mass spectrometry. *PLoS One*. 2019;14(10):e0223899. doi:10.1371/journal.pone.0223899

Zinsli et al. Deimmunization of protein therapeutics - Recent advances in experimental and computational epitope prediction and deletion. *Comput Struct Biotechnol J*. 2020;19:315-329. doi:10.1016/j.csbj.2020.12.024