BASECHEM: GUIDED LIGAND IDEATION THROUGH VISUALIZATION

Lydia Auch, Katrina Lexa, Amber Kampen; Denali Therapeutics, South San Francisco, CA

INTRODUCTION

Efficient and effective computational tools are necessary for accelerated drug discovery, particularly at the hit generation and lead optimization stages.

Relying on a modeler to provide all hypothesis-driven computational analyses on a project impedes ideation and discovery. Instead it is optimal to provide medicinal chemists with carefully developed tools that enable self-guided, immediate exploration of individual ideas.

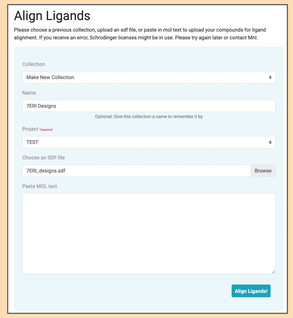
We have developed Basechem to enable this by crafting a straightforward interface centered around the idea of “Collections” of design ideas through which multiple hypotheses can be quickly assessed.

ANALYSIS DETAILS

Basechem supports multiple routes of SBDD investigation. Each analysis module is straightforward for users to operate since each form looks similar and minimal input is needed from the user.

An upload of compounds and a Project selection are minimally required to kick off an analysis. Basechem checks our internal registration system and names compounds according to existing Denali metadata.

Tasks are run in the background and users are shown a loading window while they wait, which informs them they can leave the page without delaying their results.



A submission form for creating a new Collection

PROPERTY CALCULATION

This tool is used weekly by all medicinal chemists at Denali Therapeutics to guide compound design and prioritize ideas.

Basechem calculates and displays a core set of key 2D- and 3D-properties based on user selections. At Denali, CNS-relevant properties are “on” by default for display. RDKit is used to generate most properties, however Basechem employs Schrodinger’s Ligprep and Epik to calculate the ALogD of each molecule. This decision was made to balance the open-source advantages of RDKit with the superior correlation of ALogD with LogD.

As part of property calculation, users can choose to download their results, including a file containing normalization of their stereo-enumerated compounds for registration.

LIGAND ALIGNMENT

Basechem requires a set of reference compounds (called “*Series*”) to be maintained by a subject matter expert for each project. However, this requires simply an SDF and SMILES for any new series lead(s) to be uploaded to the admin site. Basechem replaces cumbersome emailing or file sharing between project teams, acting as the single source of truth for current binding pose hypotheses.

When users upload compounds, Tanimoto similarity to all reference compounds is calculated and a *Series* is assigned. By default, a compound is aligned to the assigned *Series* but users may select another *Series* to use for exploring diverse pose hypotheses.

Basechem then performs the following steps:
(1) runs ConfGen to generate conformers for each compound uploaded
(2) fuzzily aligns a conformer to the reference using Phase Align
(3) superimposes all conformers to the Phase-Aligned structure
(4) calculates strain energy for each conformer using RDKit
(5) returns the four conformers with the lowest RMSD and the four lowest energy conformers

DOCKING

Receptor structures and binding pocket residue numbers are maintained by the computational lead for each SBDD project. This is accomplished by uploading a PDB structure and tagging its *Series*.

Denoting the binding pocket residues allows the 3D viewer to show more detail near the binding site. Basechem currently utilizes smina as the quick and open-source docking solution; however, it is easy to add or replace the docking algorithm on the backend.

Basechem uses the phase aligned structure of each compound to the default *Series* as the input pose which aids in returning top poses.

Users can interact with their 3D structures in a viewer on Basechem or download their conformers/poses as an SDF file.

RESULTS/CASE STUDY

A chemist is interested in the target Transthyretin (TTR). This target is a transport protein which was recently shown to be a causative protein in diseases which see deposition of TTR amyloid fibrils (Yokoyama, 2021). The therapeutic hypothesis suggests aggregation can be prevented by stabilization of the TTR tetramer, where two small molecule drugs are shown to bind. One drug, Bithionol, has a higher binding affinity than Triclabendazole. In this example, a chemist would like to improve the activity of the second molecule, Triclabendazole, using modeling tools.

User enters Basechem to Align design ideas to a reference

↓

User Uploads *Compounds* in the form of Mol text or .sdf file

↓

Basechem cleans structures and **assigns** *Series*

↓

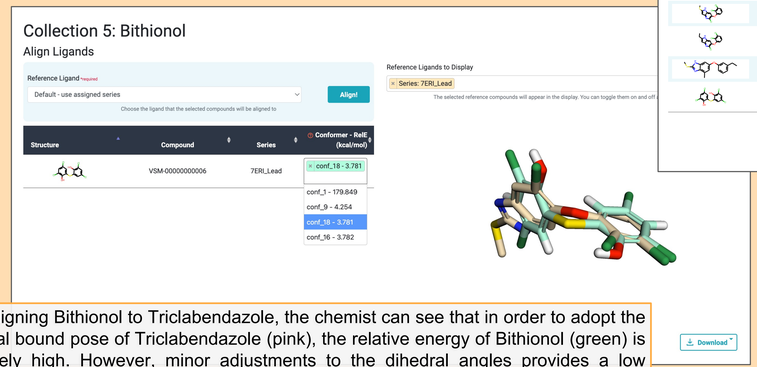
Basechem creates *CompoundOccurrences* to track owner and *Compounds* if they don’t exist

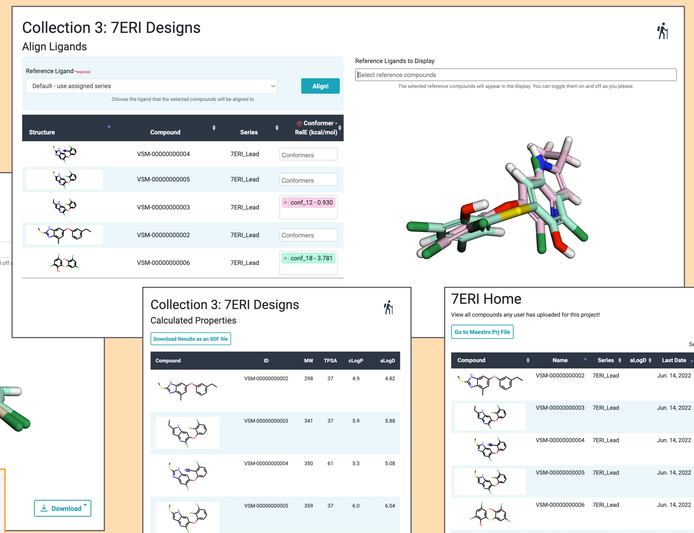
↓

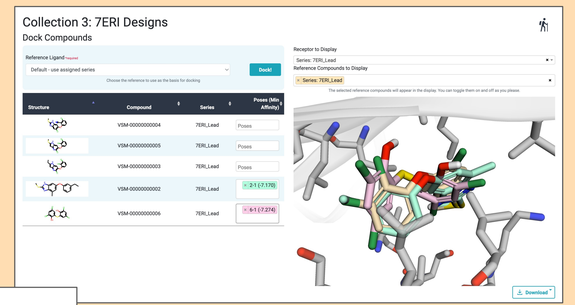
Basechem searches internal database to determine if *Compounds* are registered and attaches relevant identifiers

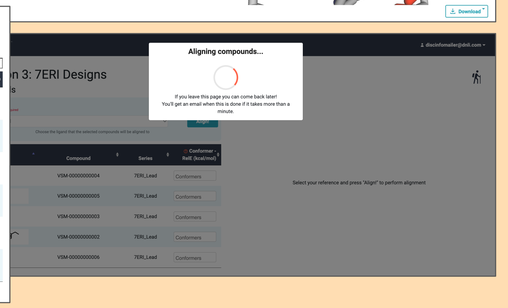
↓

Basechem creates a *Collection* and directs to the Align analysis page









After aligning Bithionol to Triclabendazole, the chemist can see that in order to adopt the identical bound pose of Triclabendazole (pink), the relative energy of Bithionol (green) is extremely high. However, minor adjustments to the dihedral angles provides a low energy conformation that is highly similar to the bound state. Based on this assessment of allowed states, further compounds can be designed to favor the bound pose.

SOFTWARE DESIGN

Basechem is a Dockerized Django application that leverages a mix of open-source libraries and licensed integrations. The code is largely organized into two sub applications, one to store models and views and one to handle analysis utilities. This modular code structure offers scientists who are not software developers, such as cheminformaticians or modelers, a straightforward path to contribute to the code base. Models are adaptable and Django handles database join operations painlessly. The model structure is described here:

Collection
- A group of *CompoundOccurrences*
- Associated with a *Project* and *User*
- Handles cleaning and creation of new *Compounds* and *CompoundOccurrences*
- Dispatches analyses

Compound
Describes a unique molecule and keeps track of high-cost files associated with the molecule

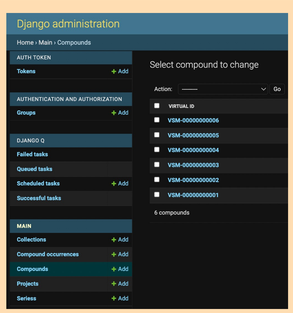
CompoundOccurrence
Pointer to a *Compound* that is associated with a specific *User*

Series
Admin uploaded molecule that is a lead compound for a *Project*

Project
A label associated with a target

User
An individual account for a Denali scientist

New analyses are easy to integrate. For example, all upload forms for each analysis share a basic layout, which parses multiple compounds from a block of MOL text or an SDF file. Basic cleaning of the structures is performed and only simple metadata fields need to be added to any additional new forms. Once uploaded structures are processed, a user-chosen analysis task (e.g., align) is asynchronously initiated. Establishing a new analysis is also modular, since analysis views share the same code for retrieving results and managing user experience. While async tasks run, users are directed to an analysis page with a window that informs them they can safely leave the page if desired and will be notified by email when their job is complete.



The Django Admin panel showing Compounds

We heavily utilize Django-Q as a lightweight task processing queue. It is seamlessly integrated into the Docker container using Supervisor to manage the Django server and Django-Q processes at the same time. Task information is stored in the database for up to a default 2.5K tasks, providing immediate results for recent analyses. This storage can be configured based on database resources and how often users return to tasks. Django-Q is visible in the Django Admin panel which, combined with our human readable naming scheme, makes it very easy to see task failures, error messages, and admin submissions of new tasks.

The Denali implementation of Basechem includes tight coupling with Dotmatics API. This enables pushing properties automatically through overnight asynchronous tasks and pulling of registered identifiers upon structure upload.

By creating a Dockerized application, we equipped Basechem with flexibility in when and where to start a server. The base Docker container is built on Centos7 with Conda for scientific package management. Additional file systems such as AWS EFS or large installations like Schrodinger can be mounted into the container. In this way, no package versions can conflict and the server can be locally run on any Mac, Linux or Windows machine that has Docker supported. Basechem uses the same Dockerfile for building development servers and cloud-based deployment. The package also comes with docker-compose files to supply “single click” start-up of a development environment.

CONCLUSIONS & COMING SOON

Thanks to the Django framework and our application architecture, customization of Basechem is straightforward for both software developers and less experienced programmers. Further, we have found that increasing user familiarity before implementing intensive analyses by building an intuitive user experience is helpful for establishing a returning user base.

Some immediate areas for customization include email templates that provide extra links, further connections between an internal registration database and Basechem, or customizable *Project* pages to view additional project details.

The Denali team has plans to continue expanding Basechem analyses. This year we will expand the property calculation tool to include machine learning predictions and launch a quantum mechanics module, starting with torsion analysis (including atom selection and dihedral plots) and geometry optimization. This module will enable immediate action on align/dock results to explore ground state/bound state energetics.

Longer-term expansion ideas include matched molecular pair analyses, CSD information, and electrostatic surface potential (ESP) maps for both ligands and receptors. We also plan to open source the Basechem codebase on GitHub.

Acknowledgements: We would like to thank the Denali Small Molecule Discovery team for their insight on feature design and testing.

Schrödinger Release 2022-2: LigPrep, Schrödinger, LLC, New York, NY, 2021.
RDKit: Open-source cheminformatics; http://www.rdkit.org
Repositioning of the Anthelmintic Drugs Bithionol and Triclabendazole as Transthyretin Amyloidogenesis Inhibitors
Takeshi Yokoyama, Mirai Kashihara, and Mineyuki Mizuguchi *Journal of Medicinal Chemistry* **2021** *64* (19), 14344-14357
DOI: 10.1021/acs.jmedchem.1c00823