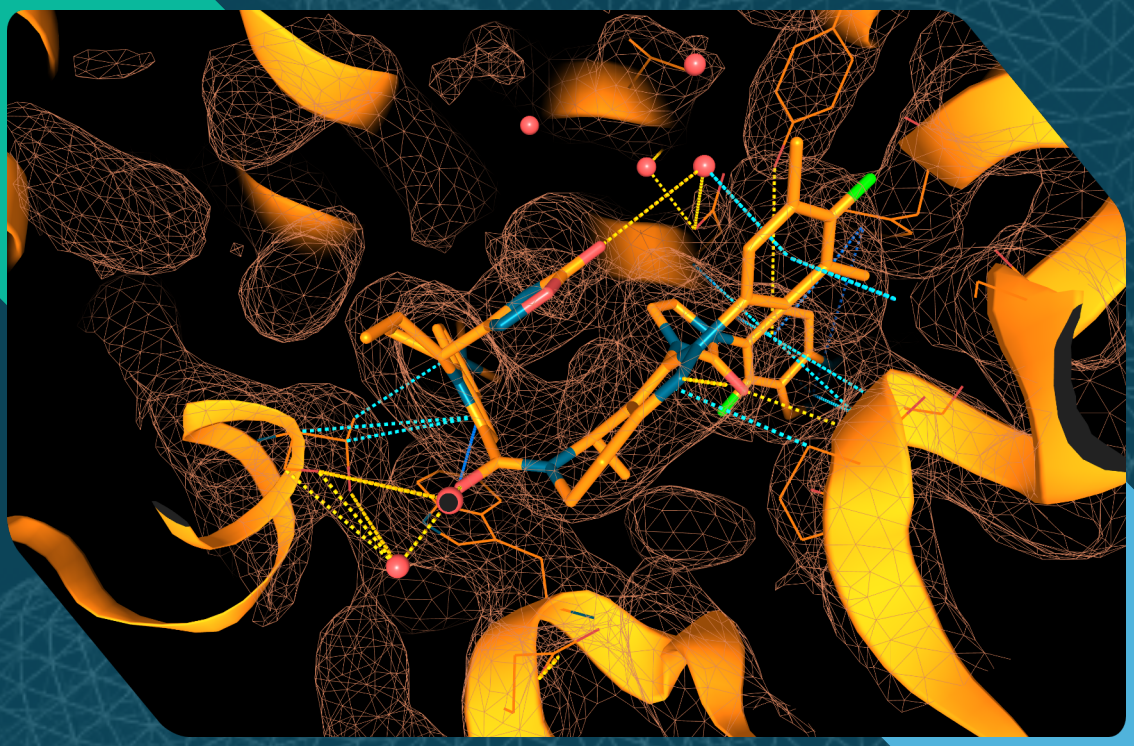
WHITEPAPER

3D PROTEIN STRUCTURES: HOW TO UNCOVER THEIR TRUE POTENTIAL TO DRIVE DRUG DISCOVERY

Elisa Martino, Marija Martini



Learn to efficiently manage big protein structure data to rapidly gather valuable insights for structure-based drug design.

3decision

discngine

OVERVIEW

Due to recent technical advances in protein structure determination techniques and the major improvements of prediction algorithms such as AlphaFold, more structural data are rapidly being produced and many more will be in the near future. In addition, in-house structures produced by pharma and biotech companies follow the same trend.

Even if this fast rise of structural data provides unprecedented opportunities for drug discovery, on the other side, it poses the challenge of managing big data.

In this whitepaper, we introduce the reasons for the growth of protein structural data and its impact on drug discovery. Then, we discuss the main difficulties of handling such a large amount of structural data. Finally, we illustrate how our data management software, 3decision®, can help address these challenges and support structure-based drug discovery projects.

Outline

Structural data avalanche in drug discovery

Challenges in managing protein structures in pharma

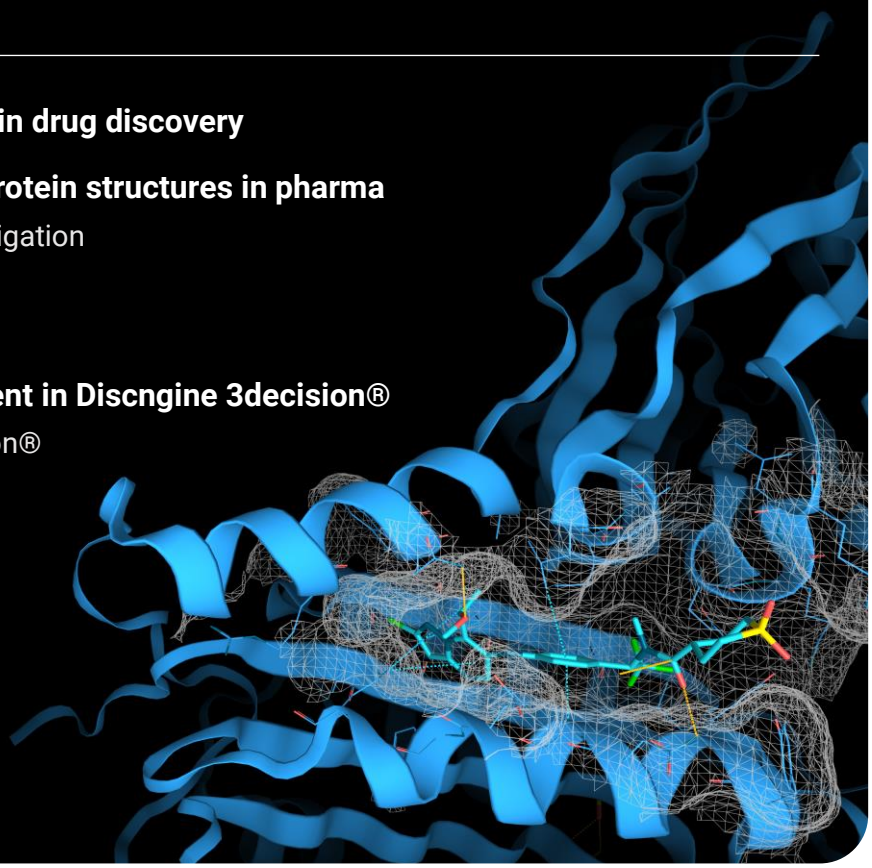
- Data storage and navigation
- Data reusability
- Data sharing
- Data analysis

Structural data management in Discngine 3decision®

- Highlights of 3decision®

Summary

Use Cases



STRUCTURAL DATA AVALANCHE IN DRUG DISCOVERY

The discovery of novel drugs is both scientifically challenging and extremely expensive. It is estimated that the development of a new drug costs approximately 1 billion dollars,[1] with around 12-15 years for the product to enter the market, if it ever reaches this stage. The risk of failure for a drug discovery project is still around 90% during clinical development (from Phase I to approval).[2] 40-50% failure rate is due to the low therapeutic efficacy of drugs, with the highest risk associated with first-in-class drugs (with new mechanisms of action), in which the molecular basis of the disease are poorly understood.[3]

To maximize the chance of a molecule to complete all stages of drug approval, **structure-guided approaches** are now included in the pipeline of most pharmaceutical companies.

A recent example of fundamental importance was the development of the first antiviral drug against coronavirus disease (COVID-19) called Nirmatrelvir (combination of nirmatrelvir/ritonavir commercialized as PAXLOVID™) from Pfizer, which was strongly supported by structural data of the target protein.[4]

Structure-based drug design (SBDD) is an iterative process in which 3D structural knowledge of the biological target of the disease (usually a protein) drives the rational ideation and optimization of new drug candidates. The compounds are subjected to the "DMTA cycle" (**D**esign, **M**ake, **T**est, **A**nalyze) to improve their drug properties.[5]

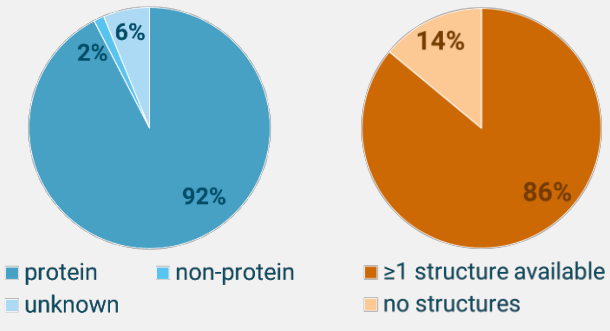
Rational design based on 3D structural insights often helps to **reduce** the number of optimization cycles needed to reach a valid candidate, making the process faster, more efficient, and, overall, cheaper. It has been estimated that SBDD projects can reduce the costs of developing a new drug by up to 50% due to the higher quality of candidates.[6]

Because of the fundamental role of biomolecule structural information for SBDD campaigns, usually, *"the first thing they [industry researchers] do when starting a new drug-discovery project is to search the PDB [protein data bank] and look hard at potential target structure(s)",* as reported by Burley."[7]

The correlation between drug approval and available target structures is evident: out of the 210 new molecular entities approved by FDA between 2010 and 2016, 92% have a known protein target, and 86% have at least one resolved 3D structure.[8] Also, in 20% of cases, more than 100 protein structures for each approved drug were available, reflecting the contribution of structure-guided approaches in the optimization of the drug (Images in the box below: Contribution of protein structures to drug discovery).

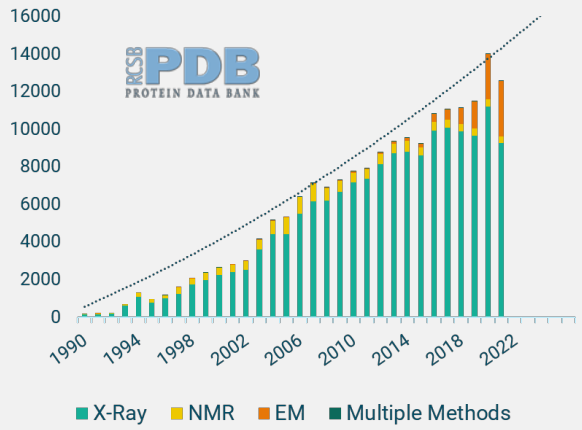
Contribution of protein structures to drug discovery

- 92% of FDA-approved new molecular entities (NME) have a known protein target
- 86% of these have at least one structure deposited in the PDB

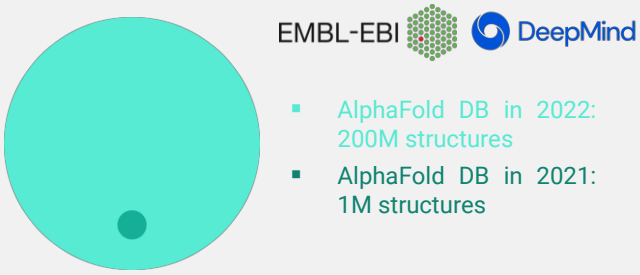


The numbers of protein structures

- In the past 10 years the number of protein structures deposited in the PDB has more than doubled, reaching almost 200,000 entries in 2022



- AlphaFold's latest release reached a total of 200 million predicted protein structures



These numbers are **very likely to increase** in the near future due to the dramatic acceleration observed in the production of structural data caused by recent technical advances in structural biology techniques.[9] Unprecedent automation and speed in protein X-ray crystallography[10] and the resolution revolution in cryo-EM,[11] impressively impacted the number of available structures in the PDB, now almost at 200,000[12]. Compared to 10 years ago, this has **more than doubled**, and the deposition rate is rapidly growing, with a historical record of 14,009 deposited structures in 2020.

Moreover, the recently developed **AI algorithm AlphaFold** has completed the prediction of 200 million structures of the human proteome and 47 other organisms. This is 200-times more than its first (already impressive) release of 1 million structures in 2021.[13] In addition to predicted single-chain proteins, AlphaFold is also being enhanced to predict multi-chain protein complexes,[14] which will expand its impact on drug discovery. For instance, it could be used for antibody development or modulators of protein-protein interactions drugs.

AI-based predictive structural biology is developing fast, especially with the latest release of the ESMFold technology (MetaAI). It predicted 600 million protein structures that haven't been categorized - the largest datasets ever seen.[15]

If the number of publicly available protein structures is already impressive (see box: The numbers of protein structures), they are just the tip of the iceberg of structural biologist contribution to drug discovery. In fact, companies produce their own proprietary structural data. From discussions with Discngine's customers and contacts in the field, we could estimate that **the production rate of protein structures in industry is at least the same as the publicly available data, if not even more**.

It is difficult to predict the exact impact of this structural data avalanche on the overall drug discovery process. However, protein structural data will most likely play a more and more central role in **accelerating the early stages of the discovery** and their number will continue to increase.

CHALLENGES IN MANAGING PROTEIN STRUCTURES IN PHARMA

> *"We were sitting on these almost 9,000 structures, and except for 2-3 people, nobody else knew where they were.*
>
> ***Rishi Gupta***
> *Associate Director,* ***Novartis***

Although a higher number of protein structures translates into tremendous opportunities for drug discovery, the structure itself does not bring much value, unless managed and used properly. Scientists rely on both internal and public sources of structural data to extract the information relevant to their research projects (e.g., ligand binding mode, binding site flexibility, protein dynamics, etc.). However, due to the growing number and complexity of structures, they are often faced with challenges associated with handling a large amount of data and their exploitation.

In the following text, we will name a few of them, reported mainly by our customers and contacts from pharma.

Data storage and navigation

One of the most critical aspects is that the storage of structural data is often **disorganized**. This is because, in pharma companies, the data are usually **stored in various locations**. The internally produced structures, for example, are kept in distinct folders, accessible only to a restricted number of people. In that case, employee turnover often results in a loss of information. Moreover, companies commonly have separate storage systems among different departments in which the data might be processed and registered using different methods and file formats.

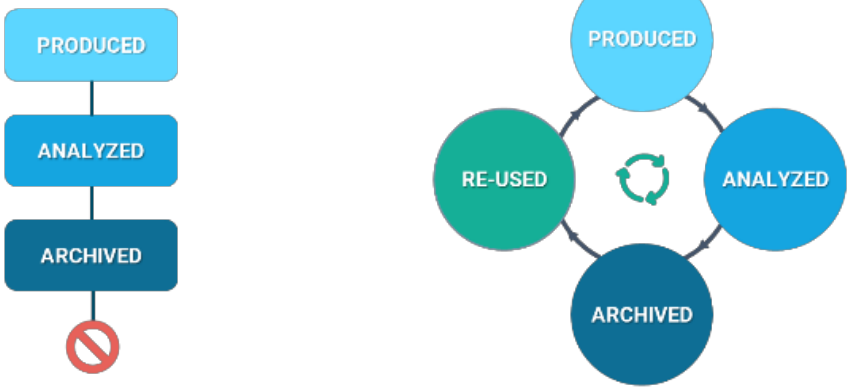
Therefore, **the data, even within the same project, might be scattered and non-uniform**. On top of this, the publicly available data are accessible from public platforms – yet another different location.

Such decentralized and unstandardized data can slow down drug discovery efforts from the very beginning. For instance, as mentioned in the section above, one of the very first steps in any SBDD campaign is the retrieval of all the structural data that might be useful for the project.[7] But because of scattered data (both public and proprietaries), the navigation is difficult, time-consuming, and sometimes even unsuccessful. This ultimately puts an entire project in danger: **if important information is left out from the beginning, there is a higher risk of failure.**

Unfortunately, the realization of such a mistake often comes too late (or never), and extra time, effort, and costs are then needed to compensate for the failure.

Data reusability

Another issue related to improper data storage is the lack of data reusability. In a research company, typically, the **life cycle of a protein structure is to be produced (with a lot of effort), then analyzed for the project purposes, and finally forgotten**. Therefore, it is very likely that a protein structure will never be re-used after its initial aim has been



Disorganized and siloed data

Centralized and standardized data

Figure 1. ***The life cycle of a protein structure:*** *disorganized and siloed storage of structural data leads to inefficient usage; centralized and standardized storage allows easy retrieval of data for future projects.*

accomplished, most of the time because historical data are easily lost in internal archives.

On the other hand, having a centralized repository in which protein structures are gathered and accessible over time, transforms structural data acquisition from a short-term into a long-term investment for drug discovery projects (Figure 1).

Data sharing

Since drug discovery requires an interdisciplinary effort and extensive collaboration among team members, it is fundamental for scientists working on the same project to easily access all the project data and at any time. However, this is not always straightforward. Data are commonly shared among different departments through email attachments that, other than being inefficient, are also subject to security issues. Email accounts are very susceptible to cyberattacks, so companies should provide their employees with **cloud-based**, software **solutions** to safely store and share their data.[16]

Moreover, scientists collaborating on the same drug discovery process produce **different types of structural metadata**: structural biologists generate electron density maps, computational chemists and modelers perform *in silico* calculations, and medicinal chemists produce data associated with the ligands.

The **complete set of data** necessary to support a structure-based drug discovery project should be ideally located in the same place. That way, all scientists involved in a project, even if with different roles, can easily have a complete overview of the project state and maximize the efficiency of the drug discovery campaign.

> *"What we miss is a repository for datasets; only a few people have access to all data. I want to be able to share to modelers and chemists maps and other data rather than just PDBs so if they don't believe me they can have a look themselves.*
>
> **Chiara Rapisarda**
> Group leader in Cryo-EM, **Sanofi**

Data analysis

Having protein structures well-organized inside a single database, readily accessible to all scientists collaborating on the project, is already a good starting point for any structure-based drug discovery project.

Nevertheless, simply tidily storing structural data does not fully exploit their full value. Protein structures alone are not sufficient to support drug discovery: they need to be **associated** with structural metadata (such as sequence information, structural annotations, mutation data, or chemical data of the ligands) and **analyzed** to answer the specific questions of the project and support decision-making (Figure 2).

However, this process does not come without its hurdles either. Usually, protein structures are not well integrated with other data sources and the analyses of structural data are performed with different tools - often individually by scientists in different departments. This results in low speed and reduced efficiency of the data analysis process and, therefore of the drug discovery campaign.

In conclusion, failing to efficiently store, share and analyze your structural data leads to a huge waste of resources and overall, to a low return on investment.

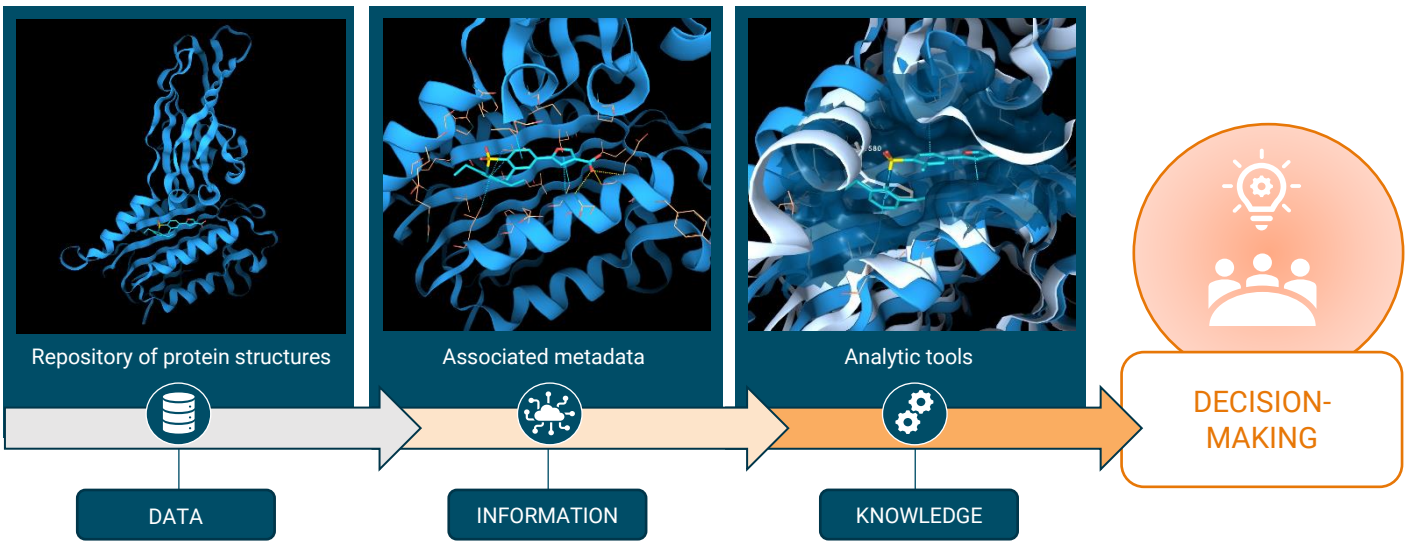


*Figure 2. **From data to knowledge:** structural data must be associated with structural metadata and analyzed to be converted into new knowledge for the drug discovery project and to fully support scientists in decision-making.*

STRUCTURAL DATA MANAGEMENT IN DISCNGINE 3DECISION®

All these issues associated with the handling of large structural datasets, highlighted, and discussed in the previous section, have motivated scientists at Discngine to develop a **comprehensive software solution called 3decision®** that ultimately supports SBDD research.

Here we will present the technology behind 3decision®, as well as describe two use-cases for the application of the software to drug discovery projects.

3decision® is a web-based **protein structure repository that allows the complete management of structural data – from small to large datasets** (Figure 3). It overcomes the challenges related to data storage, navigation, analysis, and collaboration by allowing scientists to **use a single platform to:**

- Access all the available structural data
- Rapidly and efficiently search through the massive amount of protein structures
- Perform on the fly analyses
- Exchange data with the whole drug discovery team in real-time

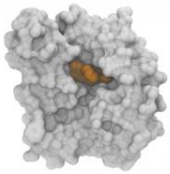
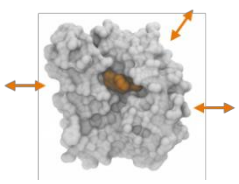
3decision® *knowledge database*		
Protein structures 	**Public**	Protein Data Bank Public *in silico* models, for example: • AlphaFold; • ESMFold
	Proprietary	in-house produced experimental and *in silico*
Sequence and annotations K D L S C K T S G L S T S D A A L A T G A T I G A G Y V	UniProt, Pfam, KLIFS, PDEStrIAn	
Associated structural metadata 	fpocket	Pockets on the molecular surface of the protein
	3decision	Ligand-protein interactions
	PDB	Electron density maps
	ChEMBL	Ligands information
		Additional relevant project data

Figure 3. Discngine solution for structural data management: 3decision®

Highlights of 3decision®

The core value of 3decision® is the **centralization and organization of all the structural data and metadata** in one single place. When a protein structure is uploaded to the 3decision® database (details in the box below: Processing of structural data in 3decision®) it is associated with its sequence and annotations, and with all the structural metadata available. In this way, what we provide is not only a simple repository for 3D protein structures, but a **complete structural knowledge database**.

3decision® database can store public and proprietary protein structures, such as:

- PDB releases and public *in silico* models (*AlphaFold, EMSFold, SwissModel, etc)*
- in-house produced experimental and *in silico* structures

Having these data centralized in one single place makes **navigation through structures easy and rapid**. 3decision® contains several search options (see Usecase 1) and various post-search filters, to quickly retrieve datasets for SBDD campaigns. This also makes **historical data reusable** over time with bigger overall impact of a single structure on multiple projects.

Moreover, structures can be organized in **projects** and easily shared with the entire team, to **promote collaboration** among scientists with different backgrounds working on the same project. 3decision®'s user-friendly interface has been built to maximize efficiency in ideation and decision-making.

Even if collaboration is a key aspect to any drug discovery effort, companies often need to regulate the access to confidential structures and projects. This can be handled in 3decision® by the user **privileges**, which allow the creation of user groups with different management capabilities inside the application.

Protein structures can also be **visualized and analyzed** directly in the same platform, without

Processing of structural data in 3decision®

What really makes 3decision® unique is the way that structures are registered by the software when they are uploaded to the database:

- ✓ the structure coordinates are mapped back to the **UniProt sequence**, and all the **annotations** correlated to the sequence are associated with the structure
- ✓ all putative and known **binding sites** on the molecular surface of the protein are identified and characterized, using the **f-pocket** algorithm[14]
- ✓ ligand-protein **molecular interactions** are computed (hydrophobic, aromatic, electrostatic, polar, hydrogen bond, halogen bond, etc.)
- ✓ **pharmacophoric features** of the binding site are indexed for binding site **comparisons**

Figure 4. Visualization options in 3decision®. A: ligand-protein interactions can quickly be toggled in the 3Dviewer; B: electron density maps fetching; C: annotation browser where the protein structure is colored corresponding to the selection.

the need to download the selected files from the search and open them in different programs. The user interface contains several layouts that are specifically adapted to perform various actions. Some of them are (Figure 4):

A. 3D visualization and selection of the **protein-ligand interactions** (automatically calculated inside the application)
B. retrieval of **electron density maps** (from RCSB database)
C. visualization of the **annotations** of a protein in the 3D structure (e.g. domains, numbering scheme, etc.)
D. **on-the-fly superposition** of multiple protein structures (see Usecase 2)

Finally, 3decision® is a **cloud-based application** which does not require a physical installation on the user's computer. Therefore, it facilitates the access by people working from different sites. The cloud-based deployment[17] of the solution guarantees quick installation and easy delivery of software updates, in addition to the highest standards in terms of **security** (ISO 27001 certification coming soon).

SUMMARY

To keep up with the fast growth and increased complexity of protein structures, pharma and biotech companies need novel technologies to support big structural data management and decision making.

If overlooked, the massive amount of protein structures quickly becomes disorganized and decentralized. This unfavorably impacts the outcome of SBDD campaigns, leading to missed opportunities and time-waste. Moreover, it creates collaboration bottlenecks, since efficient systems for sharing all the structure-related data among scientists from different departments, are often missing.

Finally, a protein structure alone does not bring a real contribution to the drug discovery effort: it is its integration with the complete set of structural metadata that really release its full potential for drug discovery.

The introduction of a centralized structural data management system early in the drug discovery pipeline can significantly impact on the success of the SBDD projects. In Discngine, we developed one such technology (3decision® software) with the aim of quickly transforming massive structural data into valuable insights for drug discovery.

In the following table, we collected some of the most common challenges in structural data management addressed by the Discngine's data management system called 3decision®.

Challenge	3decision® solution
Scattered structural data	Centralized platform for all the structural data (integration of both public and proprietary databases)
Loss of historical data	Registration of in-house produced data in the same centralized platform
Navigation through big data	Many search options and query combinations to maximize the value of the search results
Retrieval of structural metadata	Structural metadata (sequence info, structural analysis details, electron density maps, ligand information, annotations, and more) associated with every structure to be easily fetched
Use of several software	Visualization and analytic tools (e.g., on-the-fly superposition) integrated into the same platform; ligand-protein interactions calculated upon structure registration and automatically displayed
Sharing data among the team	Structures and associated data are stored in the same project folder inside the application with the possibility to manage access privileges
Security issues	On-cloud deployment with the highest standards in terms of security

USE CASES

1. Retrieve structural information to generate new compound design ideas

The design of drugs able to penetrate the blood-brain barrier (BBB) is still one of the major challenges in medicinal chemistry. A recent study showed that the acrylamide moiety can improve the overall physicochemical properties of drugs and increase BBB penetration.[18]

If you want to develop a new Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase Inhibitor (TKI) with an acrylamide group using a structure-based approach, you need to gather more knowledge on the following:

- which acrylamide-containing ligands have already been observed in a complex with a protein
- how does this functional group typically interact with the protein

*In 3decision®, you can answer these questions quickly with the **advanced search feature**.*

Combining the two queries of interest: gene name “EGFR” and chemical substructure search of the moiety acrylamide, 3decision® will quickly retrieve all the structures responding to both queries. In less than a minute, you get the list of the 45 EGFR structures (out of 287 available in the public domain) in complex with ligands containing an acrylamide group. By opening them in the 3decision® 3D Viewer, you can compare the binding mode of the compounds in the different complexes and focus on the type of interactions this moiety can have with the binding site residues. With this, you are quickly gathering instant knowledge to help you validate or further refine your design ideas.

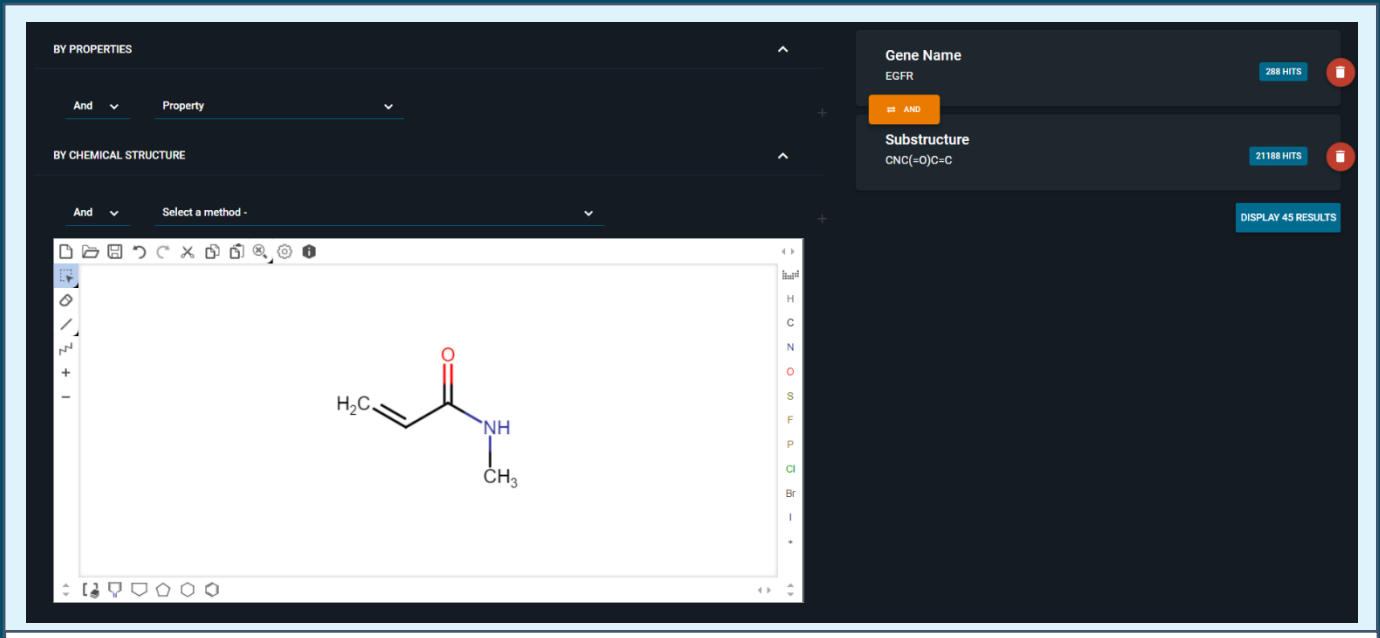
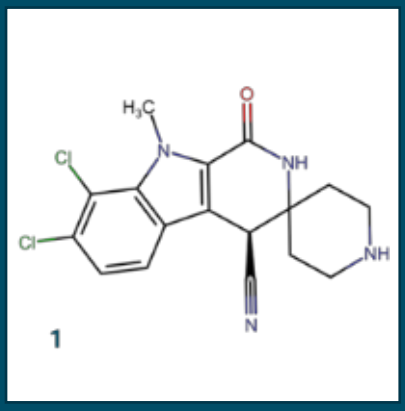


Figure A: Advanced search in 3decision®. You can combine multiple queries by properties (gene name, protein Uniprot, experimental method, etc.) with 2D chemical structure search (by exact, similar or sub-structure).

2. Compare ligand binding sites in CKL family proteins to rationalize selectivity



ATP-mimic kinase inhibitors are often associated with selectivity issues due to high sequence homology in the ATP binding site. However, the major structural difference between the CLK family (CLK 1-4) members sits at the DFG-1 position. In CLK3, the Alanine residue is found as opposed to Valine in other CLKs. This difference could be exploited to develop specific inhibitors for a single subtype of CLK. Inhibitor 1, for example, is selective towards the CLK family. Therefore, comparing the binding mode of this inhibitor in CLK1 and CLK3 could help to elucidate the role of the residues at the DFG-1 position and guide the design of new, selective drugs.[19]

In 3decision®, you can quickly assess differences in the binding mode by ***on-the-fly pocket-based superposition****.*

The comparison of the complexes of inhibitor 1 with CLK1 and CLK3 (Figure A) clearly shows that the bulkier valine residue at the DFG-1 position induces a different binding mode in CLK1. While in CLK3, the inhibitor shows a canonical binding mode, in CLK1, the inhibitor is flipped and forms uncommon halogen bonds with the hinge region (interactions automatically calculated in 3decision®, shown in purple in Figure A-i), and extensive interactions with the back pocket (Figure A-ii).

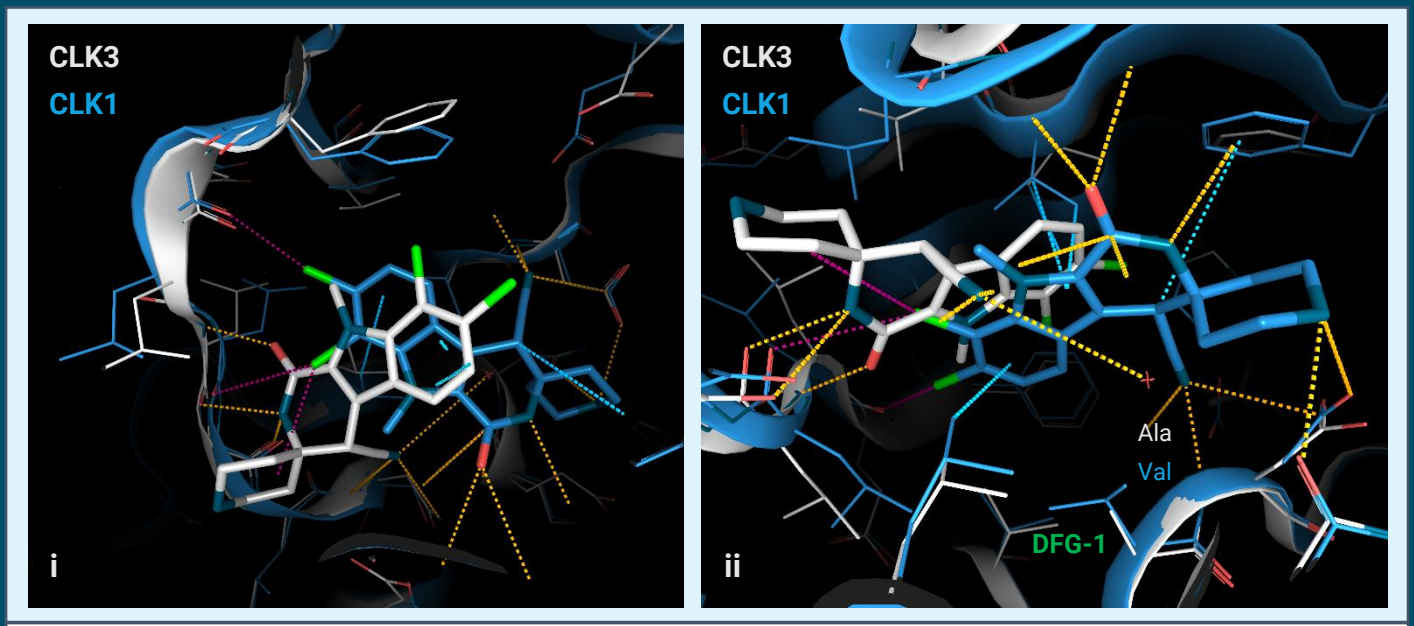


Figure B: inhibitor 1 bound to CLK1 (blue, PDB: 6ytg) and CLK3 (white, PDB: 6yu1). In complex with CLK3, shows canonical binding mode, with hydrogen bonds with the hinge region; in complex with CLK1, it is flipped and shows non-canonical interactions: (i) halogen bond (pink) with the hinge region and (ii) back pocket interactions.

To furtherly prove that the difference in the binding is due to the bigger steric hindrance of the Valine residue, you can compare the binding mode of an A319V CLK3 mutant in complex with 1. To help you assess the differences among the three structures, you can use the 3decision® highlight mode, which will clean up the view showing only the differences (RMSD can be adjusted to be more or less restrictive) among the superimposed structures. With the A319V CLK3 mutant, the ligand shows the same binding mode observed in CLK1 (Figure B), which is stabilized by an additional interaction with the valine.

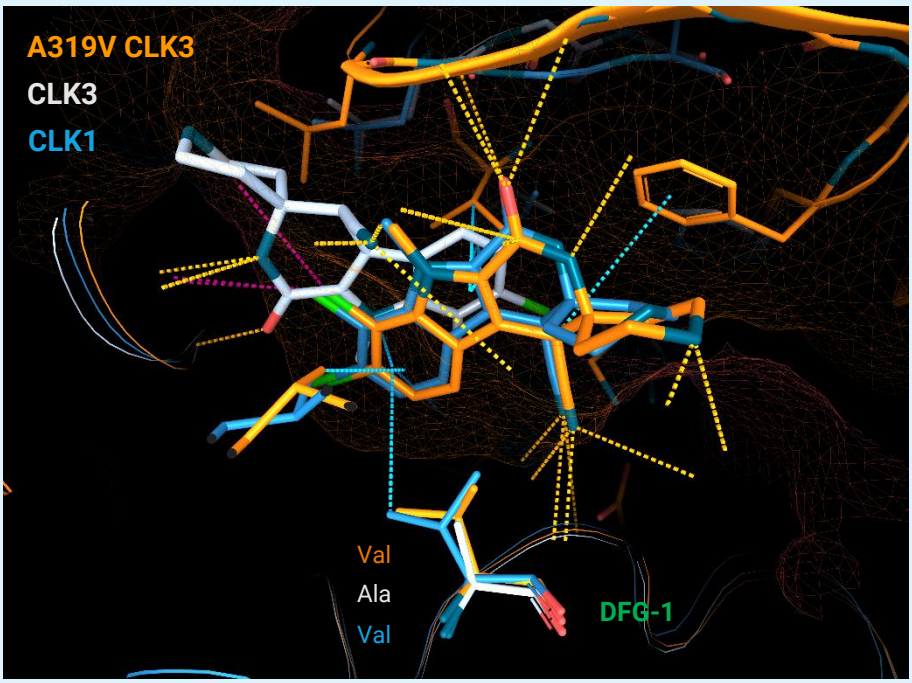


Figure C: inhibitor 1 bound to CLK1 (blue, PDB: 6ytg), CLK3 (white, PDB: 6yu1), A319V CLK3 (orange, PDB: 6z2v). The mutation to A319V at the DFG-1 position induces the same non-canonical binding mode observed in CLK1.

These structural insights support the idea that bulk amino acids at the DFG-1 position can stabilize a non-canonical binding mode of the inhibitors. Since inhibitor 1 shows higher affinity towards the CLK1, this provides a rationale to design new and more selective kinase inhibitors that should be developed to favor the non-canonical binding pose. In just a few clicks, you can retrieve valuable information to drive drug design.

For more information on 3decision® contact us:
3decision@discngine.com
https://3decision.discngine.com/

3decision

REFERENCES

(1) Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **2020**, *323* (9), 844–853. https://doi.org/10.1001/JAMA.2020.1166.

(2) Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of Clinical Drug Development Fails and How to Improve It? *Acta Pharm. Sin. B* **2022**, *12* (7), 3049–3062. https://doi.org/10.1016/J.APSB.2022.02.002.

(3) Yamaguchi, S.; Kaneko, M.; Narukawa, M. Approval Success Rates of Drug Candidates Based on Target, Action, Modality, Application, and Their Combinations. *Clin. Transl. Sci.* **2021**, *14* (3), 1113–1122. https://doi.org/10.1111/CTS.12980.

(4) Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J.; Dantonio, A.; Di, L.; Eng, H.; Ferre, R. A.; Gajiwala, K. S.; Gibson, S. A.; Greasley, S. E.; Hurst, B. L.; Kadar, E. P.; Kalgutkar, A. S.; Lee, J. C.; Lee, J.; Liu, W.; Mason, S. W.; Noell, S.; Novak, J. J.; Obach, R. S.; Ogilvie, K.; Patel, N. C.; Pettersson, M.; Rai, D. K.; Reese, M. R.; Sammons, M. F.; Sathish, J. G.; Singh, R. S. P.; Steppan, C. M.; Stewart, A. E.; Tuttle, J. B.; Updyke, L.; Verhoest, P. R.; Wei, L.; Yang, Q.; Zhu, Y. An Oral SARS-CoV-2 Mpro Inhibitor Clinical Candidate for the Treatment of COVID-19. *Science* **2021**, *374* (6575), 1586–1593. https://www.science.org/doi/10.1126/science.abl4784.

(5) Plowright, A. T.; Johnstone, C.; Kihlberg, J.; Pettersson, J.; Robb, G.; Thompson, R. A. Hypothesis Driven Drug Design: Improving Quality and Effectiveness of the Design-Make-Test-Analyse Cycle. *Drug Discov. Today* **2012**, *17* (1–2), 56–62. https://doi.org/10.1016/J.DRUDIS.2011.09.012.

(6) Stevens, R. *The cost and value of three-dimensional protein structure*. https://www.ddw-online.com/the-cost-and-value-of-three-dimensional-protein-structure-1114-200308/.

(7) Burley, S. K. Impact of Structural Biologists and the Protein Data Bank on Small-Molecule Drug Discovery and Development. *J. Biol. Chem.* **2021**, *296*, 100559–100560. https://doi.org/10.1016/J.JBC.2021.100559.

(8) Westbrook, J. D.; Burley, S. K. How Structural Biologists and the Protein Data Bank Contributed to Recent FDA New Drug Approvals. *Structure* **2019**, *27* (2), 211–217. https://doi.org/10.1016/j.str.2018.11.007.

(9) Martini, M. *5 fast-growing techniques in structural biology and their impact on drug discovery*. https://3decision.discngine.com/blog/2021/11/30/5-fast-growing-techniques-in-structural-biology-and-their-impact-on-drug-discovery.

(10) Maveyraud, L.; Mourey, L. Protein X-Ray Crystallography and Drug Discovery. *Mol. 2020, Vol. 25, Page 1030* **2020**, *25* (5), 1030. https://doi.org/10.3390/MOLECULES25051030.

(11) Scapin, G.; Potter, C. S.; Carragher, B. Cryo-EM for Small Molecules Discovery, Design, Understanding, and Application. *Cell Chem. Biol.* **2018**, *25* (11), 1318–1325. https://doi.org/10.1016/j.chembiol.2018.07.006.

(12) https://www.rcsb.org/stats/growth/growth-released-structures

(13) https://alphafold.ebi.ac.uk/

(14) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer. *bioRxiv* **2022**, 2021.10.04.463034. https://doi.org/10.1101/2021.10.04.463034.

(15) Callaway E. AlphaFold's new rival? Meta AI predicts shape of 600 million proteins. *Nature* **2022**, *611*, 211-212. https://doi.org/10.1038/d41586-022-03539-1.

(16) Lacombe, J. *Improving data management in the life sciences industry - European Pharmaceutical Manufacturer*. https://pharmaceuticalmanufacturer.media/pharmaceutical-industry-insights/improving-data-management-in-the-life-science-industry/.

(17) Gillet Markowska, A. *Five advantages of 3decision AWS Quick Start for Pharma and Biotech*. https://3decision.discngine.com/blog/2022/4/11/eumr0r1n6zizdpj64ikjclyj28my9o.

(18) Wu, K. Da; Chen, G. S.; Liu, J. R.; Hsieh, C. E.; Chern, J. W. Acrylamide Functional Group Incorporation Improves Drug-like Properties: An Example with EGFR Inhibitors. *ACS Med. Chem. Lett.* **2019**, *10* (1), 22–26. https://pubs.acs.org/doi/abs/10.1021/acsmedchemlett.8b00270

(19) Schröder, M.; Bullock, A. N.; Fedorov, O.; Bracher, F.; Chaikuad, A.; Knapp, S. DFG-1 Residue Controls Inhibitor Binding Mode and Affinity, Providing a Basis for Rational Design of Kinase Inhibitor Selectivity. J. Med. Chem. 2020, 63 (18), 10224–10234. https://pubs.acs.org/doi/full/10.1021/acs.jmedchem.0c00898.