DORSSAA: Drug-target interactOmics Resource based on Stability/Solubility Alteration Assay

Ehsan Zangene[a], Elham Gholizadeh[a], Mohieddin Jafari[a]

a. Department of Biochemistry and Developmental Biology, University of Helsinki, Helsinki, Finland

Abstract:

The field of drug development faces significant challenges in assessing cellular-level target and off-target engagement and mechanism of action of drugs. To address these challenges, innovative approaches like Thermal Proteome Profiling and Proteome Integral Solubility Alteration have emerged. These high-throughput techniques leverage mass spectrometry to characterize drug-protein interactions on a large scale, holding the potential to revolutionize drug discovery. Here, we introduce DORSSAA (https://dorssaa.it.helsinki.fi/), an integrative resource specifically designed for the comparative analysis of stability and solubility alteration assay datasets. DORSSAA hosts an extensive database containing an impressive 258,238 records, including data from 18 distinct cell lines and organisms, 33 unique compounds, and a vast collection of 32,934 proteins. With the incorporation of multiple analysis tools, DORSSAA provides a user-friendly web application to support the comprehensive study of drug-target interactions. In a comprehensive case study, DORSSAA's utility is demonstrated by identifying and exploring protein-drug interactions, such as the prominent DHFR-Methotrexate interaction, across diverse cell lines and experimental conditions. This case study highlights the significance and cross-verification of drug targets using DORSSAA's data. This comprehensive resource accelerates drug discovery, enhances our understanding of protein target behavior, and empowers molecular biology research.

Introduction:

The process of drug discovery is a multidisciplinary endeavor that spans the fields of medicine, biotechnology, and pharmacology, aiming to identify new candidate drugs. A major challenge in drug discovery is the lack of a direct method to monitor the targets and off-targets of drugs within cells and tissues (Seashore-Ludlow et al. 2020). The FDA's approval process includes thorough evaluations such as absorption, distribution, metabolism, excretion (ADME), and animal toxicity assessments, yet a full understanding of the Mechanism of Action (MoA) for a new drug isn't always required for approval (Brown and Wobst 2020). From drug MoA point of view, identifying effective drug targets is challenging in drug development. This challenge is reflected in historically low success rates during phase II trials, where real-world drug effectiveness is assessed, presenting significant hurdles. Failing FDA approval not only wastes resources but also highlights the need for innovative strategies to improve drug efficacy (Phaneuf et al. 2023).

In the realm of drug discovery, small molecule modulators have diverse effects on target proteins, and understanding their molecular MoA is essential for improving therapeutic

interventions. While targeted techniques can assess a compound's impact on specific pathways or cellular functions, systems-level profiling is necessary for comprehensive MoA identification. RNA quantification through high-throughput sequencing technologies has been employed to characterize drug MoAs on a wide scale, as seen in the Connectivity Map (CMAP) project (Subramanian et al. 2017). However, given that most FDA-approved drugs target proteins and the weak correlation between protein and RNA levels, protein-level changes are likely the most relevant readout of compound action (Gygi et al. 1999; Liu et al. 2016; Nusinow et al. 2020).
Addressing these challenges necessitates innovative high-throughput approaches, and one such promising approach involves the characterization of drug-protein interactions using solubility/stability alteration assay combined with untargeted proteomics (Phaneuf et al. 2023). Central to this technique is a biophysical concept which is based on changes in target stability or instability upon compound binding and other disturbances such as temperature. These techniques include cellular thermal shift assay (CETSA), thermal proteome profiling (TPP), proteome integral solubility assay (PISA) , etc (Savitski et al. 2014, Gholizadeh et al. 2021). Using these methods, by subjecting samples to varying temperatures, we can earn a wealth of insights into protein behavior and interactions with drugs and compounds (Mateus et al. 2016, Perrin et al. 2020, Gaetani et al. 2019). Due to its numerous advantages, this approach has become widely used in both drug discovery and fundamental biological research.

Recent advances in unbiased solubility/stability alteration assays have significantly expanded the data available for drug target interactions. However, a comparative tool for studying drug targets, deconvoluted by the latest high-throughput experimental methods, is currently lacking. Hence, we introduce DORSSAA (Drug-target interactOmics Resource based on Stability Solubility Alteration Assay) as a crucial solution to the challenge of integrating experimental differential proteome solubility data. As a central repository for drug-protein interactions, DORSSAA serves as a valuable asset for researchers in the drug discovery domain. By consolidating high-throughput data, this resource streamlines the data analysis process, enabling a more comprehensive understanding of drug-target interactions.

Materials and Methods:

Collection of Proteomics Data

In our study, we harnessed the power of publicly available research by utilizing 20 diverse datasets, encompassing a range of experiments including Thermal Proteome Profiling (TPP), Protein Interaction by Small-molecule Analysis (PISA), and Improved PISA (i-PISA), etc., involving the integration of numerous clinical samples and proteome databases from various cell lines. This extensive data compilation involved the examination of 33 different chemicals applied across 18 unique cancer cell lines or species, with comprehensive details available in Supplementary Table S1 (figure 1), allowing us to comprehensively unveil all pairwise associations between drugs and target stability/destability by reporting fold change or difference and P-values obtained from all datasets for each of the identified protein targets.

Data Normalization and Handling Missing Values

In our datasets, we encountered diverse column names stemming from different studies that employed varying characters as separators, causing inconsistency. To ensure uniformity and clarity, we uniformly replaced these separator characters with underscores. Furthermore, we aimed to establish standardized protein naming conventions, associating each protein with both a UniProt ID and a HUGO gene symbol. When UniProt IDs were unavailable, we utilized HUGO gene symbols and employed the Entrez R package to identify the corresponding UniProt IDs. To facilitate user recognition of proteins with UniProt IDs, we introduced a distinctive cloud icon (☁) alongside them. Additionally, we meticulously separated interaction records with multiple gene symbols into distinct entries within our database.

When dealing with -log10(p-value) measurements for drug-target interactions, addressing the variability arising from different studies is paramount. This variability is visually evident in the volcano plots within our web application, with each -log10(p-value) value closely tied to specific study and experimental conditions. Users can delve into these variations to assess the reliability and significance of reported interactions across diverse experimental contexts. To account for multiple replications in some datasets, we've included an additional column in the reported data table in DORSSAA, offering insight into the replication status of each interaction in the study.

Protein Query Database Interface

The Protein Query tab empowers users to search for specific compounds treated within distinct cellular contexts or organisms, allowing them to uncover either a specific drug-target interaction or explore the entirety of the database. Query results are presented in various visual formats, including bar charts, volcano plots, and comprehensive data tables containing detailed interaction information. Users can further refine their results based on the significance of reported p-values.

Moreover, users have the capability to cross-reference protein identities with UniProt records and access related databases reporting drug-target interactions such as DTC, Micha, and Drug Target Interact.

Furthermore, users can leverage the built-in gene set enrichment and GO term enrichment tools to investigate potential common features shared among proteins resulting from the analysis. The underlying hypothesis is that proteins associated with a specific chemical compound may exhibit common functional characteristics, often represented by molecular functions within gene ontology terms. To visualize these enriched terms and their relationships, a treemap and network display are available, shedding light on the interrelatedness of different GO terms with similar semantic features. In general, we anticipate that compounds targeting different protein targets may exhibit similarities in various characteristics. We also expect that these similarities may manifest in shared gene ontologies. Therefore, the ability to analyze GO terms and assess gene set enrichment similarity using semantic similarity tools would be a valuable feature for visualizing and quantifying such similarities.

Cell Lines Query Database Interface

The cell line query tab empowers users to delve into significant biological questions. It allows users to explore the extent of drug-protein interactions comprehensively. By leveraging the insights gained from the protein query tab, users can discern which cellular context, whether it be a specific cancer cell line or organism, is more conducive to fostering drug-protein interactions and which ones are not. This information enables users to identify pan-cancer interactions and cancer-specific interactions, paving the way for personalized therapeutic strategies. Similar to the protein query tab, we equip users with a range of tools and functionalities, including interaction filtering based on p-values, UniProt protein descriptions, UniProt external checks, and external drug-target interaction checks.

Compound query database interface

The compound query tab addresses another pivotal biological question, specifically focusing on drug repurposing. Within this tab, we elucidate how one can target a protein using various drugs, shedding light on which drugs hold the potential for use against a particular protein within a specific context.

In parallel with the cell lines query tab, we provide users with a suite of essential checks and tools, encompassing interaction filtering based on p-values, UniProt protein descriptions, UniProt external checks, and external drug-target interaction assessments. This feature, central to our web application, facilitates the discovery of alternative drug therapies and even allows for the design of drug combination therapies tailored to specific protein targets within distinct contexts.

Database Implementation:

The DORSSAA backend server was meticulously constructed and accessed utilizing the Rstudio shiny package, leveraging the capabilities of the R programming language (version 4.3.2). SQL database functionalities were seamlessly integrated into the system through the DBI R package, facilitating database implementation and querying. Additional functionalities were enhanced through the incorporation of JavaScript code on both the UI and server sides. All data analyses were executed using R. DORSSAA underwent rigorous testing across multiple well-known web browsers, including Google Chrome (recommended), Firefox, and Apple Safari, to ensure optimal performance and compatibility (figure 1).

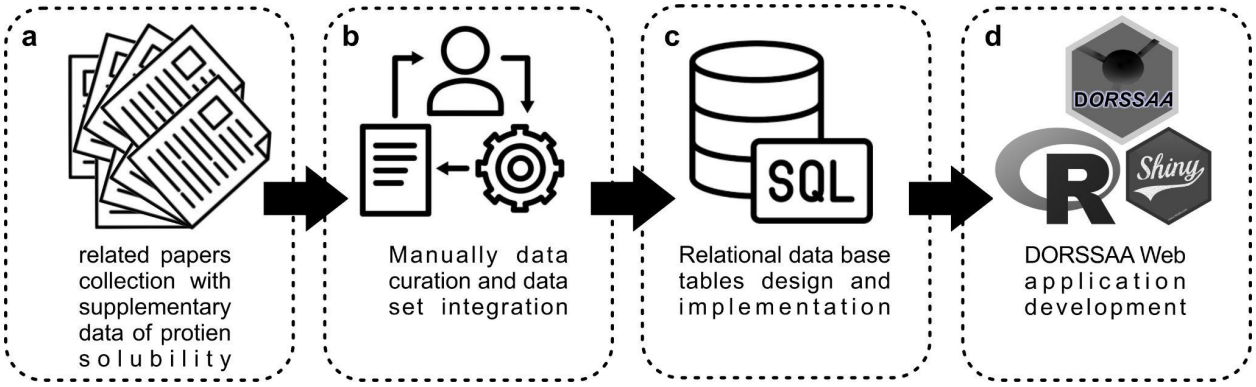

Figure 1: The comprehensive workflow for DORSSAA database development, a) encompassing the collection of related papers, b) manual data curation, dataset integration, c) relational database tables design, and d) implementation, and DORSSAA web application development.

Database Content:
DORSSAA houses a diverse array of data components that significantly augment its value for researchers:

A. Number of Records: The primary "Drug-Protein Interaction" table within the database boasts an impressive 258,238 records. These entries encompass data related to 18 distinct cell lines and organisms, 33 unique compounds, and an extensive collection of 32,934 proteins.
B. Data Accessibility: Accessing the wealth of information in DORSSAA is user-friendly and effortless. Researchers can utilize the DORSSAA web application, an interactive and visually intuitive web-based interface, to easily retrieve results for any conceivable query related to drug-target interactions based on proteome solubility assay. This accessible interface is readily available via web browsers, ensuring widespread and open access to valuable data at https://dorssaa.it.helsinki.fi/.
C. Ongoing Updates: DORSSAA maintains its currency and relevance through regular updates. As new publications with publicly available TPP datasets emerge, the database is promptly refreshed to incorporate the latest research findings, ensuring that users have access to the most current information.
D. Long-term Availability: The graphical user interface (GUI) for database querying not only provides free access but will remain available indefinitely. This commitment ensures that researchers can rely on DORSSAA as a steadfast and enduring resource for their molecular biology investigations.

Web interface:

The DORSSAA web application provides a user-friendly interface, facilitating data retrieval, visualization, and download (Figure 2). DORSSAA is organized into three primary tabs: Protein Query, Cell Lines Query, and Compound Query. For each of these tabs, users initiate their query by selecting options from three dropdown lists, including Uniprot_ID selection, cell line selection, and protein selection dropdown lists. Upon making these selections, a bar chart and a volcano

plot are generated in the main panel of DORSSAA, providing visual representations of the queried data. Additionally, a comprehensive data table is presented below the visualizations, offering detailed information on all possible interactions (Figure 2). These interactive tools empower users to explore drug-target interactions, quantify protein changes, and gain valuable insights into the molecular mechanisms at play. The DORSSAA web application thus equips researchers with the means to efficiently analyze and visualize data for various research inquiries, advancing the study of drug-target interactions and related phenomena.

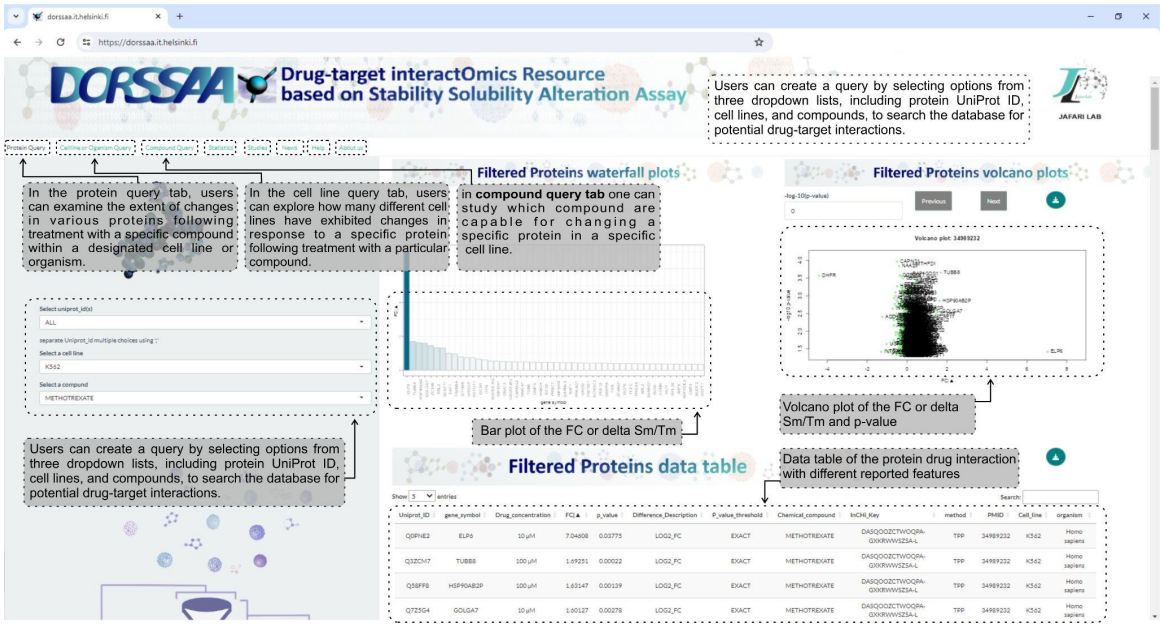

Figure 2: The DORSSAA web application provides a user-friendly interface with three primary tabs: Protein Query, Cell Lines Query, and Compound Query. Users can select Uniprot_IDs, cell lines, and proteins from dropdown lists to initiate queries. Visualizations, including bar charts and volcano plots, are generated in the main panel, offering insights into drug-target interactions and protein changes. Below the visualizations, a comprehensive data table presents detailed information on all possible interactions, enabling efficient data exploration and analysis.

Case Study: In this study, we observed concordance in protein targets found in DORSSAA output and reviewed proteins in uniprot and some other databases. To ensure our retrieved data is accurate and meaningful, we applied strict filters to select only the most relevant and statistically significant interactions. One particular drug, methotrexate, has been extensively investigated within the K562 cell line. Intrigued by this filtration, we used DORSSAA's Protein query tool to find a detailed list of proteins linked to methotrexate in K562 cells. Then by searching on uniprot as an external database, we observed a high alignment between uniprot and DORSSAA representing the accuracy of drug targets. We also checked the alignment of these proteins in other databases which the results were acceptable. As the next step we investigated the similarity of gene ontology among these significant proteins targeted with

methotrexate in the K562 cell line. Numerous proteins were categorized as having comparable molecular function based on gene set enrichment analysis. These proteins were shown to be engaged in 26 distinct molecular functions. As an example 23 proteins were significantly enriched in pyrophosphatase activity and 19 proteins were significantly enriched in ribonucleoside triphosphate phosphatase function.

In the multitude of outcomes, 661 protein targets, one protein, Dihydrofolate Reductase (DHFR), adorned with a UniProt ID of P00374, emerged as a standout candidate. Notably, DHFR exhibited the highest log2-FC, accompanied by a highly significant log10(p-value). These findings compellingly underscored the strong association between DHFR and methotrexate within the domain of K562 cells. DHFR significantly contributes to the biosynthesis of amino acids and folic acid. Its involvement in reducing dihydrofolate to tetrahydrofolate, facilitated by the nicotinamide dinucleotide phosphate cofactor, has been substantiated through diverse clinical investigations (Sehrawat et al. 2023). DHFR has been identified as a therapeutic target in the treatment of cancer and various bacterial infections and is well-known as a target of methotrexate in many studies (Sehrawat et al. 2023; Heppler et al. 2022; Maksimovic et al. 2020). Interestingly, in DORSSAA database this protein has been targeted with 5 different drugs in two studies with different methods, can be seen in compound query tab. These drugs include staurosporine, sns032, raltitrexed, panobinostat. In exploring the MoA of methotrexate, an analysis of biological process enrichment revealed a noteworthy alignment of enriched terms, providing insights into methotrexate's MoA, particularly its role in inhibiting transmethylation reactions.

Keen to ascertain the universality of the DHFR-Methotrexate interaction across diverse cell lines and experimental contexts, we conducted a series of subsequent queries within DORSSAA. This crucial interaction found confirmation across a spectrum of cell lines and experimental studies, further solidifying its significance. The collective validation was facilitated through an array of experimental methodologies, including TPP, and STPP, in distinct projects involving cell lines such as K562, and 293T. Surprisingly these two cells, K562 and 293T, originated from different sources, embryonic kidney cells and chronic myelogenous leukemia, respectively. In this DHFR case, DORSSA offers the benefit of identifying additional drugs that target this protein in various cell lines, potentially facilitating drug repurposing. Furthermore, the database reveals that this protein is subjected to different stability solubility alteration assays at varying drug concentrations.

This comprehensive case study underscores DORSSAA's prowess in facilitating the identification and exploration of protein-drug interactions across diverse cell lines and experimental conditions. It not only sheds light on potential drug targets but also showcases their significance through cross-verification across various studies and methodologies.

Conclusion:

DORSSAA represents a groundbreaking and vital resource in the realm of drug discovery and protein research. In a landscape where the challenges of assessing cellular-level target engagement and off-target binding are ever-present, DORSSAA emerges as a beacon of

innovation. It addresses these challenges by harnessing the power of thermal proteome profiling (TPP) and Proteome Integral Solubility Alteration (PISA) techniques, which leverage mass spectrometry to characterize drug-protein interactions.

The main objective of DORSSAA is to accelerate drug discovery and enhance our understanding of protein behavior under different conditions. Its comprehensive database compiles differential proteome solubility data sets from various sources, offering a rich and diverse proteomic landscape for researchers to explore. The user-friendly web application of DORSSAA provides an intuitive interface for studying drug-target interactions, enabling researchers to visualize, analyze, and interpret the proteins involved in different experimental contexts.

DORSSAA excels in its ability to answer critical biological questions. Researchers can explore how various proteins change in response to specific compound treatments within designated cell lines or organisms. They can also investigate the extent of changes in different cell lines in response to specific proteins following treatment with particular compounds. Additionally, the compound query tab provides insights into which drugs have the potential to affect specific proteins in specific cell lines, opening doors to drug repurposing and combination therapy strategies.

The database's robust implementation is built upon R programming, SQL database functions, and JavaScript, ensuring a seamless and efficient user experience. Rigorous testing across popular web browsers guarantees optimal performance and compatibility.

DORSSAA's rich content encompasses an extensive number of records, including data on 18 distinct cell lines and organisms, 33 unique compounds, and a vast collection of 32,934 proteins. Access to this wealth of information is user-friendly through the DORSSAA web application, ensuring open and widespread access to valuable data.

The commitment to ongoing updates ensures that DORSSAA remains current and relevant in the fast-evolving landscape of proteomics and drug discovery. As new publications with publicly available differentially proteome solubility datasets emerge, the database promptly incorporates the latest research findings.

In conclusion, DORSSAA emerges as a primary indispensable tool for researchers seeking to unravel the complexities of drug-protein interactions and understand protein behavior in different biological contexts. Its user-friendly interface, extensive data coverage, and commitment to updates make it a valuable asset in the pursuit of innovative drug discovery and protein research. DORSSAA empowers researchers to accelerate their investigations and make data-driven discoveries, contributing to advancements in molecular biology and therapeutic development.

Brown, Dean G., and Heike J. Wobst. 2020. “Opportunities and Challenges in Phenotypic Screening for Neurodegenerative Disease Research.” *Journal of Medicinal Chemistry* 63 (5): 1823–40.

Dwyer, Donard S., Eric Aamodt, Bruce Cohen, and Edgar A. Buttner. 2014. “Drug Elucidation: Invertebrate Genetics Sheds New Light on the Molecular Targets of CNS Drugs.” *Frontiers in Pharmacology* 5 (July): 177.

Gaetani, Massimiliano, Pierre Sabatier, Amir A. Saei, Christian M. Beusch, Zhe Yang, Susanna L. Lundström, and Roman A. Zubarev. 2019. “Proteome Integral Solubility Alteration: A High-Throughput Proteomics Assay for Target Deconvolution.” *Journal of Proteome Research* 18 (11): 4027–37.

Gholizadeh, Elham, Reza Karbalaei, Ali Khaleghian, Mona Salimi, Kambiz Gilany, Rabah Soliymani, Ziaurrehman Tanoli, et al. 2021. “Identification of Celecoxib-Targeted Proteins Using Label-Free Thermal Proteome Profiling on Rat Hippocampus.” *Molecular Pharmacology* 99 (5): 308–18.

Heppler, Lisa N., Sanaz Attarha, Rosanne Persaud, Jennifer I. Brown, Peng Wang, Boryana Petrova, Isidora Tošić, et al. 2022. “The Antimicrobial Drug Pyrimethamine Inhibits STAT3 Transcriptional Activity by Targeting the Enzyme Dihydrofolate Reductase.” *The Journal of Biological Chemistry* 298 (2): 101531.

Maksimovic, V., Z. Pavlovic-Popovic, S. Vukmirovic, J. Cvejic, A. Mooranian, H. Al-Salami, M. Mikov, and S. Golocorbin-Kon. 2020. “Molecular Mechanism of Action and Pharmacokinetic Properties of Methotrexate.” *Molecular Biology Reports* 47 (6): 4699–4708.

Perrin, Jessica, Thilo Werner, Nils Kurzawa, Anna Rutkowska, Dorothee D. Childs, Mathias Kalxdorf, Daniel Poeckel, et al. 2020. “Identifying Drug Targets in Tissues and Whole Blood with Thermal-Shift Profiling.” *Nature Biotechnology* 38 (3): 303–8.

Sehrawat, Renu, Priyanka Rathee, Sarita Khatkar, Esraküpeli Akkol, Maryam Khayatkashani, Seyed Mohammad Nabavi, and Anurag Khatkar. 2023. “DihydrofolateReductase (DHFR) Inhibitors: A Comprehensive Review.” *Current Medicinal Chemistry*, March. https://doi.org/10.2174/0929867330666230310091510.