In Silico Target Identification Highlights IL12B as a Candidate for Small Molecule Drug Development

Tanner Trevor J*[1]

[1]Independent Researcher – Little Rock, Arkansas, USA

*Corresponding author
Correspondence : trevor.tanner@utah.edu

ABSTRACT

Drug discovery's rising costs and complexities require innovative strategies to identify viable therapeutic targets. We developed a computational pipeline to pinpoint protein targets lacking known small molecule probes, focusing on sites traditionally considered challenging for small molecule intervention but validated by FDA-approved biologics. Our approach integrates machine learning, public databases, structural modeling, and functional annotations to prioritize novel binding pockets that overlap biologically validated interfaces. This method identified IL12B as a promising candidate, revealing a previously unexploited surface pocket that overlaps part of the briakinumab epitope. Static protein solvent mapping and dynamic fragment simulations provide convergent evidence of druggability, including fragment-binding clusters and chemically diverse hotspots. While not yet experimentally validated, this site represents a plausible target for orally available IL12B inhibitors. Such compounds could address current clinical limitations of antibody therapies - such as prolonged systemic exposure and infection risk - by enabling a shorter half-life and improved mucosal penetration in diseases like inflammatory bowel disease.

Keywords: computational drug discovery, il12b, machine learning, drug target identification, bioinformatics, small molecule inhibitors

Introduction

The discovery of new medical drugs has become increasingly challenging and costly despite significant technological advancements (Scannell et al., 2012). As the easier targets have been exhausted, scientific progress now requires tackling more complex challenges.

Our research strategy addresses these challenges by leveraging existing knowledge to uncover novel opportunities. Inspired by Sir James Black's insight that "the most fruitful basis of a new drug is to start with an old drug" (Raju, 2000), we hypothesized that protein sites traditionally deemed "difficult-to-drug" for small molecules—but already validated as therapeutic targets by FDA-approved biologics—may hold untapped potential. Small molecule drugs possess distinct pharmacokinetic and pharmacodynamic properties, such as shorter half-lives and enhanced tissue penetration, which could offer clinical advantages over biologics in certain contexts (Wan, 2016).

This study introduces a computational pipeline designed to identify protein sites lacking known small molecule binders but overlapping—at least partially—with biologic binding epitopes. Our approach builds on the proven efficacy of the SiteMap tool in identifying challenging protein sites, further refining its outputs through machine learning and filtering criteria intended to reduce the risk of failure in downstream campaigns (Liu & Altman, 2014).

To evaluate potential druggability within these sites, we combined two complementary fragment mapping tools: FTMap and SILCS-Hotspots. While FTMap has demonstrated a high true positive rate in identifying druggable regions, its rigid-body framework does not account for protein flexibility. Conversely, SILCS-Hotspots incorporates protein dynamics but tends to overpredict, leading to a higher false positive rate (Ghanakota & Carlson, 2016). By integrating both approaches, we aim to balance structural precision with conformational breadth, enabling more robust identification of functionally meaningful binding sites suitable for future small molecule targeting.

Methods

Homologous count file creation

We implemented an initial sequence-based filtering step to reduce the likelihood of off-target binding during drug-target selection. This computationally inexpensive approach prioritized candidate proteins that exhibited minimal primary sequence homology to other proteins in the human proteome. However, this method does not account for proteins that share structurally similar binding sites despite lacking primary sequence homology, a phenomenon that can still mediate significant off-target interactions (Duran-Frigola et al., 2017).

We created a custom bioinformatics pipeline using Nextflow, "bioinfo.nf", to filter out drug targets with significant primary homology to other human proteome proteins (Di Tommaso et al., 2017). We used AlphaFold structural predictions (version one) of the human proteome as input (Jumper et al., 2021). We extracted FASTA protein sequences from the input PDB files and blasted them against the UniProt reference human proteome using blastp (Altschul et al., 1990). Hits were filtered using a custom Python script, "filter_blast.py", with a 1E-62 maximal e-value cutoff (Louie et al., 2009), then re-aligned to the query protein sequence using mafft (Katoh & Standley, 2013). Hit counts for each query were tabulated using a custom bash script, "extract.sh", and saved to a reference file. Fig. 1b visualizes the process.

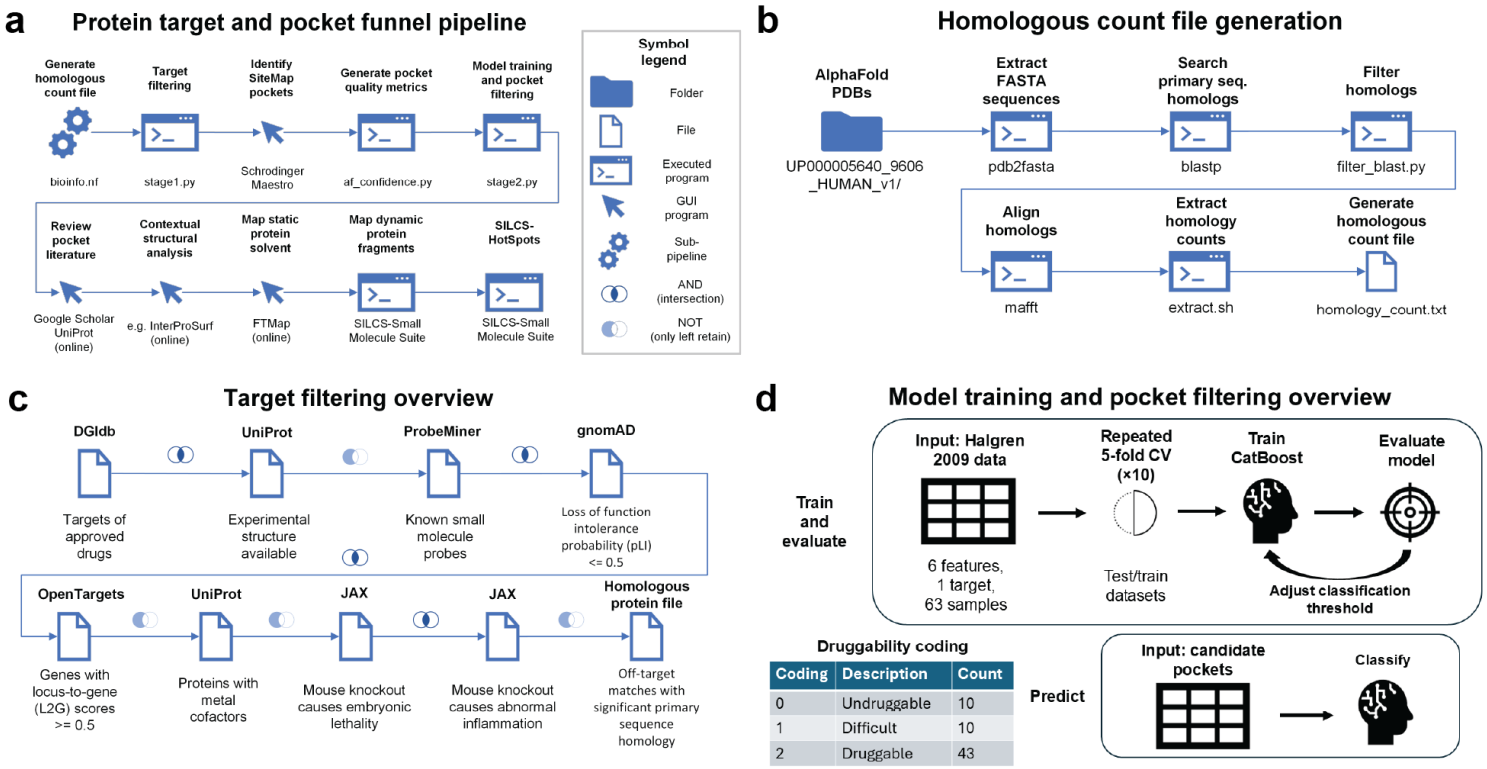


Figure 1 - Target evaluation and binding pocket prediction. (a) All intermediate steps and interactions with graphical user interface (GUI) based programs are performed manually. (b) All steps are executed using the Nextflow automation tool. Command arguments not listed. (c) All flow chart steps contained within Stage1.py. Further descriptions of databases and justifications for filtering steps found in methods. (d) All machine learning steps within Stage2.py. Target threshold refers to minimal value for differentiating between druggable and undruggable pockets. CV stands for cross validation.

Initial target filtering

We applied a series of filters using external and internal data sources to limit the number of protein candidates for structural analysis in a script titled "stage1.py". We used the Drug Gene Interaction Database (DGIdb) to identify approved drug putative targets (Cannon et al., 2024). We then filtered targets to include only those with experimental structures available in the UniProt database (The UniProt Consortium et al., 2025).

We applied an antibody structure filter using the Thera-SAbDab database to include only targets where a structure is available with a biologic binding (Raybould et al., 2020). This ensured that actual biologic binding sites were characterized and known for our analysis.

We excluded targets with existing small molecule probes available according to the ProbeMiner database to focus on unexplored targets and maximize potential for novel discoveries (Antolin et al., 2018). In the final stage of initial filtering, we considered targets only if they had no significant primary sequence homology to other proteins in the human genome to reduce the probability of off-target effects. The process of initial target filtering is visualized in Fig 1b.

SiteMap pocket identification

After initial target filtering, we manually loaded structural predictions from AlphaFold into the Schrodinger Maestro application and processed them with the Protein Preparation Wizard using default parameters and the OPLS4 force field (Madhavi Sastry et al., 2013; Lu et al., 2021). We then detected protein pockets using SiteMap with default parameters and exported them to a comma-separated file (Halgren, 2007).

Pocket quality metric generation

We processed the SiteMap pocket predictions in a custom Python script, “af_confidence.py”, that calculated the percentage of “very high” (>90 pLDDT) quality amino acid predictions in the pocket and the percentage of “very high” quality amino acids in the entire region from the first amino acid in the pocket to the last amino acid in the pocket. The calculated metrics were added as new columns to the original SiteMap dataset and saved to a new comma-separated file.

Model training

Training data points for Schrödinger SiteMap pockets and druggability labels were extracted from prior research (Halgren, 2009). The features used for model training included SiteMap’s core pocket descriptors, such as Dscore, SScore, size, enclosure, philic, and phobic values, which numerically characterize pocket geometry and physicochemical environment. Druggability labels of "undruggable," "difficult," and "druggable" were coded as 0, 1, and 2, respectively.

A CatBoost classifier was utilized as the supervised classification model, selected because its probability outputs can be directly interpreted as a measure of model confidence (Prokhorenkova et al., 2019). Model parameters included 1000 iterations, a learning rate of 0.05, a depth of 6, and a logloss function, with class weights set dynamically based on class imbalance observed during training. Random initialization state was fixed at 42.

Druggability labels were converted into binary values, marking labels as "true" if they exceeded thresholds of either 0 or 1. Model performance was evaluated using Matthews correlation coefficient (MCC. Repeated five-fold stratified cross-validation with ten repeats was used to robustly estimate model performance. Cross-validation metrics were compiled and visualized using Python libraries Pandas, Matplotlib, and Seaborn, with MCC distributions presented as boxplots with overlaid strip plots.

The final druggability threshold was selected based on achieving the highest average MCC across evaluated thresholds. The full training dataset was used to train the final CatBoost model at the

threshold that optimized MCC performance. This supervised training process is visualized in Fig 1d and implemented in the accompanying "stage2.py" file.

Pocket druggability prediction and filtering

SiteMap pockets from candidate targets were filtered based on AlphaFold structure quality metrics. Pockets were retained only if both the pocket residues and the entire region spanning from the first to last amino acid in the pocket had at least 50% of amino acids with "very high" predicted AlphaFold quality (>90 pLDDT). Druggability predictions were then made for the quality-filtered pockets using the trained CatBoost model. Pockets were classified as druggable if the model predicted a probability greater than or equal to 50%. The prediction and filtering process is implemented in the "stage2.py" file and the process is visualized in Fig 1d.

Pocket literature review

A manual literature review was conducted using Google Scholar and UniProt to identify representative natural and synthetic binders of each primary protein of interest, based on prior structural studies. To contextualize the location of these interactions, the residues involved in binding—whether natural or synthetic—were manually annotated and visualized using UCSF Chimera v1.18 (Pettersen et al., 2004) on the AlphaFold-predicted structure of the protein. Distinct sets of residues corresponding to each binder type were colored separately to enable visual differentiation of their respective binding regions. Only the residues on the parent protein structure were displayed; partner molecules were not included. These mapped residues were subsequently used to determine spatial overlap with predicted pockets during the contextual structural filtering step. Unstructured coils near the N-terminus were omitted from visualizations for clarity.

Contextual structural analysis

To further filter and prioritize candidate binding sites for small-molecule intervention, we required that predicted SiteMap pockets overlap with both (1) the binding interface of a clinically validated antibody and (2) the site of a natural binder (e.g., a physiologically interacting protein partner). Only sites meeting both criteria were retained. For the resulting case, we selected the most recent high-resolution X-ray structure of the relevant protein–binder complex and applied InterProSurf to quantify the change in solvent-accessible surface area (ΔSASA) upon complex formation (Negi et al., 2007).

Residues were labeled as high-priority contact points if they (1) fell within the region overlapping both interfaces and (2) exhibited a ΔSASA greater than 50 Å². These were interpreted as structurally validated interaction features that may inform small-molecule design aimed at mimicking therapeutically relevant contacts. ΔSASA values were visualized on the surface of the protein target of interest using ChimeraX 1.9, with a black-to-white gradient ranging from 0 Å² to the highest per-residue ΔSASA observed on the target protein (excluding binding partners).

Static protein solvent mapping

Each target that passed the pocket literature review was analyzed using FTMap in protein–protein interface mode to identify putative druggable hotspots (Kozakov et al., 2015). AlphaFold-predicted monomeric structures were used as the primary input. In cases where a known post-translational modification (PTM) was located near a predicted binding site, the most recent available monomeric X-ray structure containing a representative PTM was also analyzed using FTMap to qualitatively assess whether the modification altered the local binding environment. This step was included as a precautionary measure, given that AlphaFold predictions may implicitly incorporate structural information from PTM-containing templates. FTMap output was downloaded and analyzed locally. A pocket was considered to show evidence of small-molecule druggability if at least one hotspot

contained a minimum of 16 distinct probe types and was accompanied by one or two nearby secondary hotspots (Kozakov et al., 2011). Druggable hotspots were visualized in PyMol v2.5.4 with rainbow ribbon coloring, and unstructured N-terminal coils were hidden from all visualizations (Schrödinger, 2021).

Dynamic protein fragment mapping

We utilized a SLURM cluster on Google Cloud Platform with preemptible compute nodes in conjunction with the SilcsBio Small Molecule Suite (2022.1 release) to calculate both regular and halogen-based SILCS FragMaps of AlphaFold structural predictions for targets that passed the druggability criteria from static protein solvent mapping (Ustach et al., 2019; SchedMD, 2022). FragMaps were visualized in PyMol, rainbow ribbon coloring was used, and unstructured coils near the n-termini were hidden. FTMap results and FragMaps were aligned in PyMol and FragMap clusters were labeled if within 1 angstrom of any of the FTMap probes.

SILCS Hotspots mapping

To complement FTMap-based static solvent mapping, we used SILCS-Hotspots to evaluate potential fragment binding regions on the AlphaFold-predicted structure of each target (MacKerell et al., 2020). For targets with a known post-translational modification (PTM) near a binding site of interest, the most recent monomeric experimental structure containing a representative PTM was also analyzed to qualitatively assess whether the modification influenced predicted hotspot locations. All SILCS-Hotspots calculations were performed using the 2022.1 release of the SilcsBio Small Molecule Suite. The resulting hotspots were aligned with FTMap probe outputs in PyMol. Rainbow ribbon coloring was used and unstructured coils near n-termini were hidden. A SILCS hotspot was labeled and visualized if it fell within 5 Å of an FTMap probe molecule, based on results from either the AlphaFold or PTM-inclusive FTMap simulation. Hotspots meeting this criterion within a predicted binding site of interest were retained for further analysis and figure generation.

Custom helper scripts

We used biopython, numpy, pandas, and seaborn libraries for data processing and basic analysis in our internal scripts (Cock et al., 2009; McKinney, 2010; Harris et al., 2020; Waskom, 2021).

Results

Model performance and target prioritization

Cross-validation of the CatBoost classifier demonstrated predictive performance across both binary druggability thresholding strategies. Classification performance was higher when distinguishing druggable and difficult pockets (labels = 1 or 2) from undruggable pockets (label = 0) using a threshold of > 0, compared to distinguishing druggable pockets (label = 2) from all others using a threshold of > 1, as measured by median Matthews correlation coefficient (MCC) values (Fig. 2a). The final model used for downstream prediction was trained at the threshold > 0, which yielded the highest mean MCC across cross-validation folds.

The full target selection pipeline applied sequential filtering steps to candidate proteins: (1) identification as an approved or investigational target in DGIdb; (2) availability of at least one experimental structure in UniProt; (3) availability of at least one antibody-bound structure in the Thera-SAbDab database; (4) absence of existing small molecule ligands in the ProbeMiner database; (5) exclusion of proteins annotated with bound metal ions; (6) absence of significant primary sequence homology to other human proteins; (7) sufficient AlphaFold model quality in predicted pockets; and (8)

predicted druggability based on the trained machine learning model. IL12B was the only target that satisfied all criteria (Fig. 2b).

Site-level pocket identification in IL12B

Two predicted druggable pockets were identified in the IL12B AlphaFold structure. The CatBoost model predicted a 99.8% probability of druggability for Site A and a 98.8% probability for Site B. Site A included residues Y136, R160, Q166, E195, S197, A198, C199, P200, A201, A202, E203, E204, S205, L206, I229, R230, Y268, F269, D312, Y314, and Y315. Site B included H105, I111, W112, S113, T114, P123, K124, K126, T127, F128, L129, R130, C131, E132, W143, E221, N222, Y223, T224, S225, S226, F227, D231, I232, K234, S317, and S318 (Fig. 2c).

Literature-informed structural binder mapping

A literature-based structural review was conducted to identify known natural and synthetic binders of IL12B. Residues involved in natural binding were identified for IL12RB1, IL23R/IL12RB2, and IL12A/IL23A based on prior structural studies (Yoon et al., 2000; Bloch et al., 2018, 2024; Glassman et al., 2021; Chen et al., 2024). Residues for synthetic binders were identified for ustekinumab, briakinumab, nanobody 124C4, and peptide 23-652 (Luo et al., 2010; Desmyter et al., 2017; Bloch et al., 2018; Pandya et al., 2020). The mapped residues for both natural and synthetic binders were visualized on the AlphaFold structure of IL12B (Fig. 2d–e).

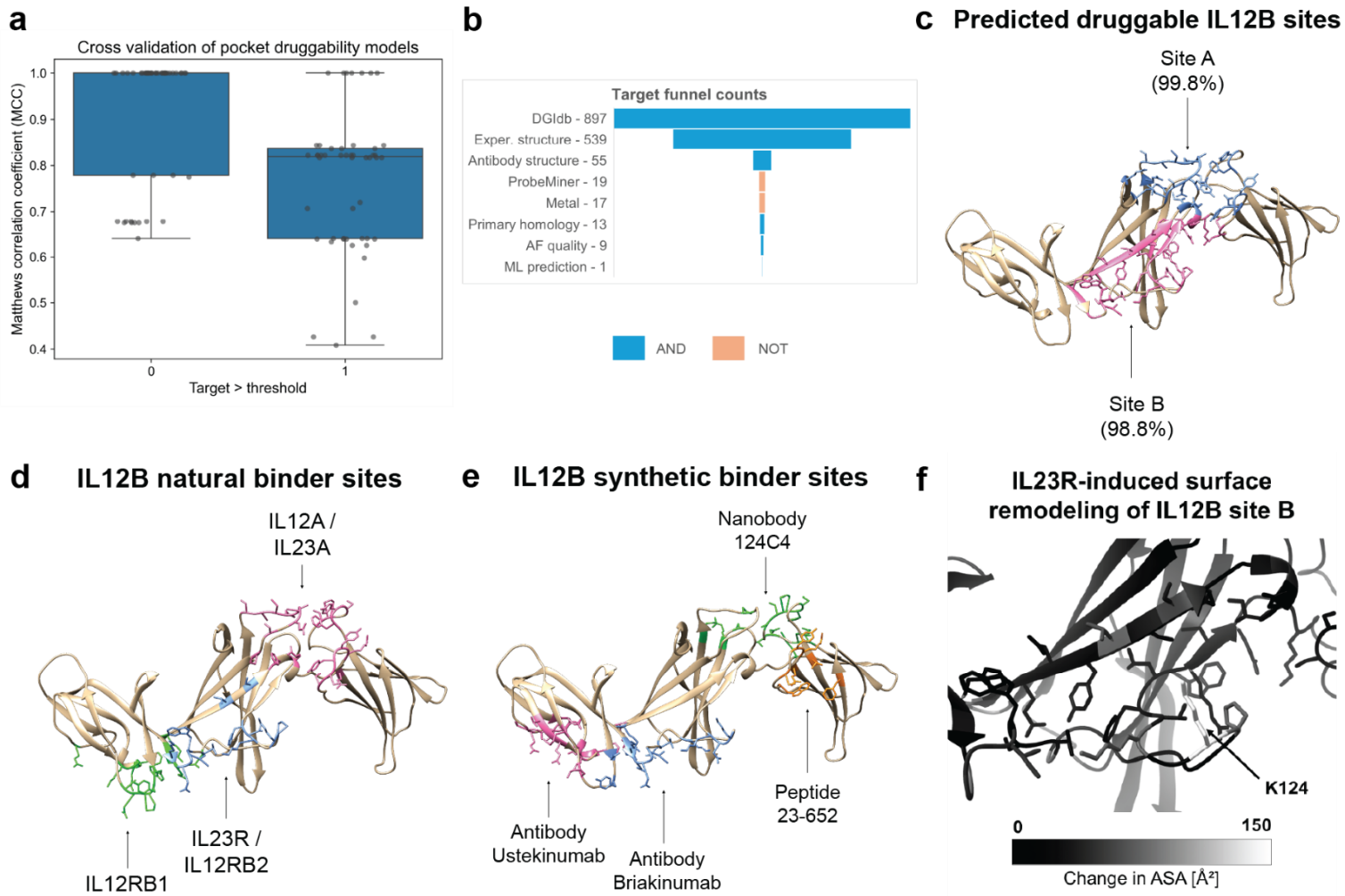
Receptor-induced remodeling of site B

The IL12B:IL23R heterodimer structure (PDB 6WDQ) was analyzed using InterProSurf to evaluate changes in solvent-accessible surface area (ASA) upon complex formation. At site B, residue K124 showed a ΔASA of 113.47 $Å^2$ (Fig. 2f).

Fragment-based solvent mapping of site B

FTMap simulations were conducted using both the AlphaFold model of IL12B (Fig. 3a) and an aligned crystal structure (PDB 1F42) that includes glycosylation near site B (Fig. 3b). In both cases, site B contained a primary cross cluster with at least 16 distinct probe types and additional nearby cross clusters.

SILCS-Hotspots analysis identified multiple predicted fragment-binding regions at site B, including hotspots 9, 24, 61, and 99 (Fig. 3c). Hotspot 99 was partially buried, as shown in the SILCS exclusion map (Fig. 3d), and located within 10 Å of other hotspots. Hotspot 99 was nearby one apolar FragMap region and a nearby hydrogen bond acceptor (hbacc) region. Additional hbacc density was observed between hotspots 24 and 61 and near hotspot 61 (Fig. 3e). SILCS-X FragMaps showed significant spatial overlap with corresponding SILCS FragMaps at site B (Fig. 3f).



(a) 1/−1 of MCC corresponds to perfect agreement/disagreement with test data. Threshold interpretation in Fig 1d. (b) AF is AlphaFold and ML is machine learning. (c) Trimmed AlphaFold model; numbers in parentheses indicate CatBoost-predicted druggability confidence scores from the model with the highest mean MCC. (d) Trimmed AlphaFold model. (e) Trimmed AlphaFold model. Overlapping residues between ustekinumab and briakinumab antibodies not colored. (f) Data from InterProSurf using PDB 6WDQ structure; labeled residue has ΔSASA > 50 and overlaps with the briakinumab binding site.

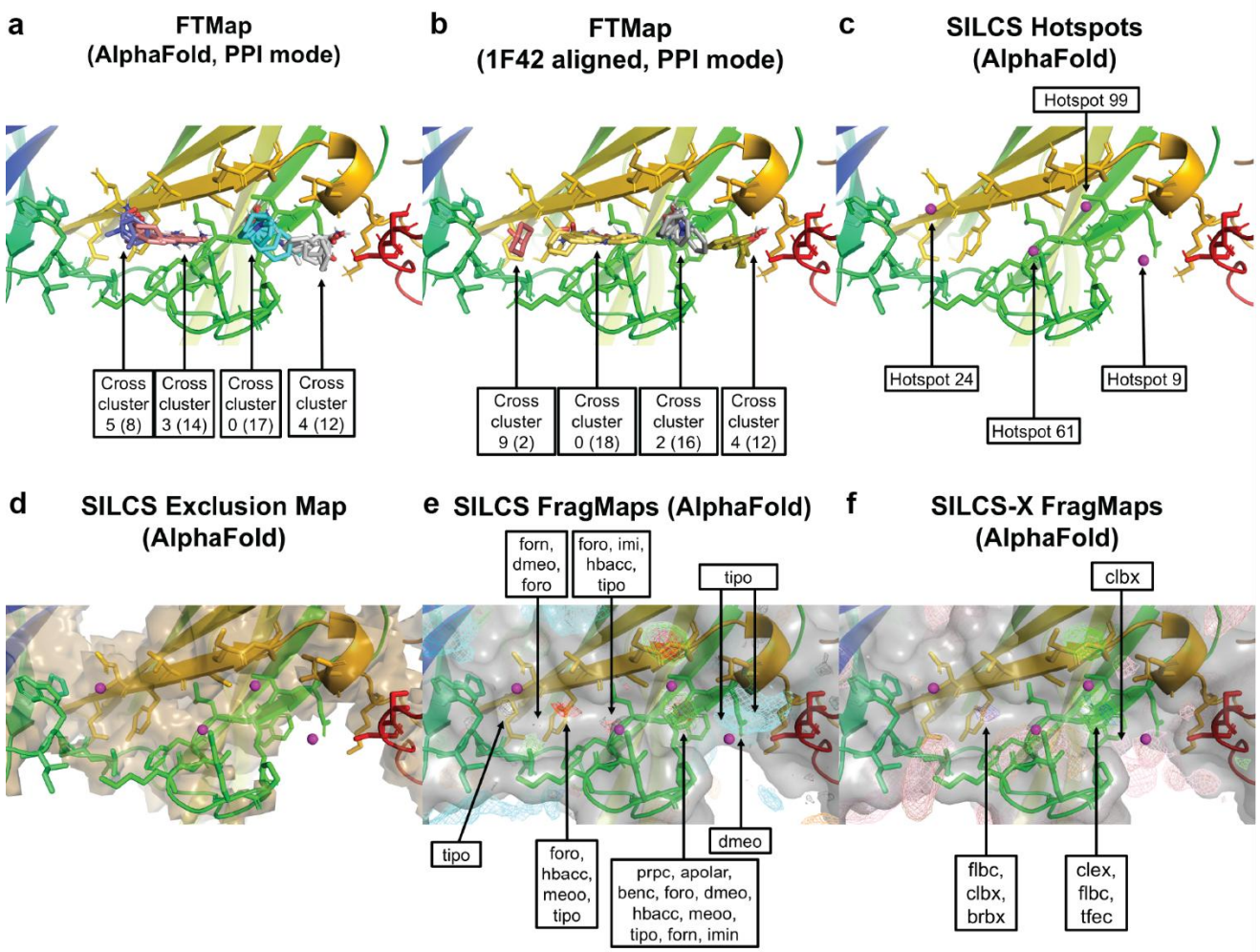


Figure 3 – Static and dynamic protein simulations of IL12B site B. Standard rainbow coloring of IL12B is used, and site B residues are shown in all subpanels. Reference labels for cross clusters (from FTMap static simulations) and hotspots (from SILCS dynamic simulations) reflect their original assignments from the respective raw simulation outputs. Numbers in parentheses for FTMap simulations are number of probes in cross cluster. AlphaFold-based models did not incorporate any post-translational modifications. (b) The 1F42-based results derive from an experimental IL12B structure that includes glycosylation near site B; cross clusters are aligned to and visualized on the AlphaFold IL12B structure for consistency. (e–f) SILCS and SILCS-X probe identities and physicochemical properties are summarized in Table 1.

FragMap Name	FragMap Color	Component Name	Associated Probe Atom(s)
apolar	Green	Generic Apolar	Generic Apolar Carbons (benc+prpc)
hbdon	Blue	Generic Donor	Generic Hydrogen Bond Donors (forn+iminh)
hbacc	Red	Generic Acceptor	Generic Hydrogen Bond Acceptors (foro+dmeo+imin)
mamn	Cyan	Positively Charged	Methylamine Nitrogen
acec	Orange	Negatively Charged	Acetate Carboxylate Carbon
meoo	Tan	Hydroxyl Oxygen	Methanol Oxygen
tipo	Black	Water Oxygen	Water Oxygen
benc	Purple	Benzene Carbon	Benzene Carbons
prpc	Lime Green	Propane Carbon	Propane Carbon
forn	Blue	Formamide Nitrogen	Formamide Nitrogen
foro	Red	Formamide Oxygen	Formamide Oxygen
dmeo	Red	Dimethylether Oxygen	Dimethylether Oxygen
iminh	Blue	Imidizole Donor Nitrogen	Imidizole Nitrogen (h-bond donor)
imin	Red	Imidizole Acceptor Nitrogen	Imidizole Nitrogen (h-bond acceptor)
flbc	Blue	Fluorobenzene	Fluorobenzene Fluorine
clbx	Pink	Chlorobenzene	Chlorobenzene Chlorine
brbx	Brown	Bromobenzene	Bromobenzene Bromine
clex	Green	Chloroethane	Chloroethane Chlorine
fetx	Green	Fluoroethane	Fluoroethane Fluorine
tfec	Green	Trifluoroethane	Trifluoroethane Carbon Bonded to Fluorines

Table 1 – SILCS and SILCS-X FragMap probe reference. Probe names, colors, component types, and key interaction atoms used in standard SILCS and halogen-augmented SILCS-X FragMap simulations. Standard probes include neutral, polar, and charged fragments; halogen probes represent specific halogenated functional groups. Colors refer to visualizations used throughout figures in this study.

Discussion

The multi-parameter prioritisation workflow reduced an initial DGIdb set of 897 drug-associated genes to a single high-confidence target, IL12B, and highlighted two predicted pockets. Site B drew particular interest because it overlaps part of the briakinumab epitope and has no reported small-molecule ligands.

Static protein solvent mapping with FTMap provided consistent evidence that site B can sustain fragment binding. The largest cross cluster contained 17 probe types in the AlphaFold model and 18 in the experimentally derived, glycosylated structure; in both cases additional nearby cross clusters were present. This pattern exceeds the ≥16-probe threshold, with supporting clusters, that has been correlated with druggability in benchmark studies (Kozakov et al., 2011).

Dynamic fragment mapping with SILCS-Hotspots corroborated the location of the pocket by identifying at least one partially buried hotspot with neighboring hotspots within <10 Å. Although only one apolar FragMap region was detected, the co-localised hydrogen-bond acceptor density fulfils

other SILCS prioritisation criteria and indicates chemical features that could guide fragment elaboration.

Clinically, IL12B is already validated by monoclonal antibodies such as ustekinumab and briakinumab, which are effective in inflammatory bowel disease (IBD) and psoriasis (Al-Bawardy et al., 2021). Antibody therapy, however, entails parenteral dosing and prolonged systemic exposure that complicate infection management; current IBD biologics have been linked to markedly increased infection rates, leading treatment guidelines to recommend temporary withdrawal during selected infections (Zabana et al., 2019; Kucharzik et al., 2021). A small-molecule inhibitor directed to site B could, in principle, offer oral delivery, shorter half-life, and improved mucosal penetration—attributes likely to facilitate rapid dose modulation and reduce systemic immunosuppression.

Taken together, the convergent FTMap and SILCS findings nominate site B as a previously unexploited pocket for small molecule development on IL12B. Experimental characterisation of this site, including fragment screening and assessment of conformational and post-translational heterogeneity, now represents a logical next step toward broadening the therapeutic arsenal beyond existing antibody modalities.

Analytic limitations

This study is limited by its reliance on computational tools and simulations, without direct experimental validation. The machine learning model used for druggability prediction was trained on a relatively small and class-imbalanced dataset, which may affect its generalizability to uncharacterized targets.

A central limitation of this analysis involves uncertainty regarding the physiological structural state of IL12B. While IL12B can exist as a monomer, homodimer, or as part of heterodimers with IL12A or IL23A, the structural and functional relevance of each state may vary by context. The AlphaFold structural model used in this study reflects a monomeric configuration, which may not fully represent the biologically active conformations encountered in vivo.

Additionally, the post-translational modification (PTM) status of IL12B remains incompletely characterized. Glycosylation patterns, in particular, may differ across subcellular compartments (e.g., intracellular vs. extracellular) and may be influenced by tissue type, inflammatory state, or disease context. Structural models used in this study do not account for potential variety in PTMs, and they may affect both the geometry and accessibility of predicted small-molecule binding sites. The 1F42 crystal structure, one of the few experimental structures for monomeric human IL12B, lacks accompanying electron density maps and dates from before modern structure validation standards, limiting its utility for refinement. Further experimental work is needed to define the in vivo conformational landscape of IL12B in human patients, including its tissue-specific PTM states and physiologically dominant oligomeric forms.

Conclusion

Our integrated in silico workflow—combining stringent sequence, structural, and machine-learning filters with complementary solvent-mapping techniques—surfaced site B on IL12B as a tractable yet chemically unexplored pocket. FTMap clustering and SILCS-Hotspots converged on this region, together indicating a permissive blend of polar and apolar interaction motifs that should accommodate fragment growth and, ultimately, lead optimisation. Critically, site B lies outside the canonical dimer interface yet overlaps part of the briakinumab epitope, anchoring its biological relevance while remaining untouched by small-molecule medicinal chemistry.

Translating these findings could expand the therapeutic repertoire for inflammatory bowel disease and related disorders, where antibody therapies are effective but hampered by parenteral dosing and prolonged systemic exposure. A well-designed oral inhibitor of IL12B has the potential to deliver

adjustable pharmacokinetics, improved mucosal penetration, and rapid dose modulation—attributes that align with clinical needs in infection-prone patient populations.

The next milestones are clear: obtain high-resolution structures of IL12B in its predominant PTM and oligomeric states, initiate fragment screening against site B, and profile early hit series for selectivity. Success on these fronts would open a direct path toward the first small-molecule IL12B modulator, offering industry and academia a fresh entry point into a validated yet under-leveraged cytokine target.

Acknowledgements

We gratefully acknowledge the support of several organizations that made this work possible. We thank the Google for Startups Cloud Program for generously providing $100,000 of computational credits to Epilog, L.L.C. We are also appreciative of Schrödinger, L.L.C. and SilcsBio, L.L.C. for providing trials of their software packages. Chemaxon generously provided a complimentary license for their software (Chemaxon, 2024). The Chemicalize platform was usefulf for visualizing chemical structures documented in various crystallization protocols for the IL12B protein extracted from the literature. We also acknowledge Wingware for access to its Wing Pro IDE for Python development (Wingware, 2025).

Salt Lake Community College (SLCC) played a crucial role in the development of this research through two key programs. Within SLCC's STEM Learning department, Dr. Lane Law and Dr. Santosh Kiran Balijepalli offered invaluable guidance through their STEM 2010 class. Their expert feedback significantly enhanced the author's ability to articulate, refine, and communicate scientific ideas. The Everyday Entrepreneur Program, taught by Jon Beutler, provided foundational training in market research, competitive analysis, and minimum viable product development.

The author also thanks Drs. Peter Iles and Tim Beagley for sparking his interest in the life sciences through engaging introductory and cell biology courses, and Dr. Craig Thulin for strengthening his understanding of biochemical principles.

The foundational skill sets used in this work were first developed during two summers the author spent in the National Science Foundation's Research Experience for Undergraduates (REU) program (NSF award #0835713) at Harvard, working in the Aspuru-Guzik group under the mentorship of Dr. Dmitrij Rappoport. The author is especially grateful to Dr. Rappoport for his generous mentorship and for offering thoughtful feedback that helped improve this manuscript.

Finally, we acknowledge the use of advanced language models, including Claude 4 Sonnet (Anthropic) and ChatGPT (OpenAI), which assisted with editing titles, improving abstract clarity, and refining the presentation of technical content.

Funding

The author declares to have received no specific funding for this study.

Conflict of interest disclosure

Trevor J. Tanner, the founder of Epilog, L.L.C., received computational resources through the Google Cloud for Startups program. Some of these resources, along with academic resources and affiliations, contributed to the research process used in the identification of the IL12B target described in this work. Epilog, L.L.C. asserts no rights associated with any of the findings presented herein. The author declares no financial conflicts of interest related to this work.

Data, scripts, and code availability

Data, scripts, and code are available online: https://doi.org/10.5281/zenodo.15765054. Some intermediate results generated using Schrödinger's commercial software have been excluded from the public repository to comply with Schrödinger's End User License Agreement. Certain scripts may not function without Schrödinger-specific outputs and are provided for reference. Recreating the original dataset may require a Schrödinger software license. All other non-Schrödinger files necessary for understanding and reproducing our work are included. The repository README provides detailed instructions.

References

Al-Bawardy B, Shivashankar R, Proctor DD (2021) Novel and Emerging Therapies for Inflammatory Bowel Disease. Frontiers in Pharmacology, 12. https://doi.org/10.3389/fphar.2021.651415

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology, 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Antolin AA, Tym JE, Komianou A, Collins I, Workman P, Al-Lazikani B (2018) Objective, Quantitative, Data-Driven Assessment of Chemical Probes. Cell Chemical Biology, 25, 194-205.e5. https://doi.org/10.1016/j.chembiol.2017.11.004

Bloch Y, Bouchareychas L, Merceron R, Składanowska K, Bossche LV den, Detry S, Govindarajan S, Elewaut D, Haerynck F, Dullaers M, Adamopoulos IE, Savvides SN (2018) Structural Activation of Pro-inflammatory Human Cytokine IL-23 by Cognate IL-23 Receptor Enables Recruitment of the Shared Receptor IL-12Rβ1. Immunity, 48, 45-58.e6. https://doi.org/10.1016/j.immuni.2017.12.008

Bloch Y, Felix J, Merceron R, Provost M, Symakani RA, De Backer R, Lambert E, Mehdipour AR, Savvides SN (2024) Structures of complete extracellular receptor assemblies mediated by IL-12 and IL-23. Nature Structural & Molecular Biology, 31, 591–597. https://doi.org/10.1038/s41594-023-01190-6

Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, McMichael JF, Kuzma K, Morrissey D, Cotto K, Mardis ER, Griffith OL, Griffith M, Wagner AH (2024) DGIdb 5.0: rebuilding the drug–gene interaction database for precision medicine and drug discovery platforms. Nucleic Acids Research, 52, D1227–D1235. https://doi.org/10.1093/nar/gkad1040

Chemaxon (2024) Chemicalize. Chemicalize. URL: https://chemicalize.com Accessed 3 September 2024.

Chen H, Ge X, Li C, Zeng J, Wang X (2024) Structure and assembly of the human IL-12 signaling complex. Structure (London, England: 1993), 32, 1640-1651.e5. https://doi.org/10.1016/j.str.2024.07.010

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25, 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Desmyter A, Spinelli S, Boutton C, Saunders M, Blachetot C, Haard H de, Denecker G, Roy MV, Cambillau C, Rommelaere H (2017) Neutralization of human interleukin 23 by multivalent nanobodies explained by the structure of cytokine–nanobody complex. Frontiers in Immunology, 8, 1–10. https://doi.org/10.3389/fimmu.2017.00884

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C (2017) Nextflow enables reproducible computational workflows. Nature Biotechnology, 35, 316–319. https://doi.org/10.1038/nbt.3820

Duran-Frigola M, Siragusa L, Ruppin E, Barril X, Cruciani G, Aloy P (2017) Detecting similar binding pockets to enable systems polypharmacology. PLoS computational biology, 13, e1005522. https://doi.org/10.1371/journal.pcbi.1005522

Ghanakota P, Carlson HA (2016) Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics. Journal of Medicinal Chemistry, 59, 10383–10399. https://doi.org/10.1021/acs.jmedchem.6b00399

Glassman CR, Mathiharan YK, Jude KM, Su L, Panova O, Lupardus PJ, Spangler JB, Ely LK, Thomas C, Skiniotis G, Garcia KC (2021) Structural basis for IL-12 and IL-23 receptor sharing reveals a gateway for shaping actions on T versus NK cells. Cell, 184, 983-999.e24. https://doi.org/10.1016/j.cell.2021.01.018

Halgren T (2007) New Method for Fast and Accurate Binding-site Identification and Analysis. Chemical Biology & Drug Design, 69, 146–148. https://doi.org/10.1111/j.1747-0285.2007.00483.x

Halgren TA (2009) Identifying and Characterizing Binding Sites and Assessing Druggability. Journal of Chemical Information and Modeling, 49, 377–389. https://doi.org/10.1021/ci800324m

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE (2020) Array programming with NumPy. Nature, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution, 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D, Vajda S (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nature Protocols, 10, 733–755. https://doi.org/10.1038/nprot.2015.043

Kozakov D, Hall DR, Chuang G-Y, Cencic R, Brenke R, Grove LE, Beglov D, Pelletier J, Whitty A, Vajda S (2011) Structural conservation of druggable hot spots in protein–protein interfaces. Proceedings of the National Academy of Sciences, 108, 13528–13533. https://doi.org/10.1073/pnas.1101835108

Kucharzik T, Ellul P, Greuter T, Rahier JF, Verstockt B, Abreu C, Albuquerque A, Allocca M, Esteve M, Farraye FA, Gordon H, Karmiris K, Kopylov U, Kirchgesner J, MacMahon E, Magro F, Maaser C, de Ridder L, Taxonera C, Toruner M, Tremblay L, Scharl M, Viget N, Zabana Y, Vavricka S, on behalf of the European Crohn's and Colitis Organisation [ECCO] (2021) ECCO Guidelines on the Prevention, Diagnosis, and Management of Infections in Inflammatory Bowel Disease. Journal of Crohn's and Colitis, 15, 879–913. https://doi.org/10.1093/ecco-jcc/jjab052

Liu T, Altman R (2014) Identifying Druggable Targets by Protein Microenvironments Matching: Application to Transcription Factors. CPT: Pharmacometrics & Systems Pharmacology, 3, 93. https://doi.org/10.1038/psp.2013.66

Louie B, Higdon R, Kolker E (2009) A Statistical Model of Protein Sequence Similarity and Function Similarity Reveals Overly-Specific Function Predictions. PLoS ONE, 4, e7546. https://doi.org/10.1371/journal.pone.0007546

Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, Ross GA, Dahlgren MK, Russell E, Von Bargen CD, Abel R, Friesner RA, Harder ED (2021) OPLS4: Improving Force Field Accuracy on

Challenging Regimes of Chemical Space. Journal of Chemical Theory and Computation, 17, 4291–4300. https://doi.org/10.1021/acs.jctc.1c00302

Luo J, Wu SJ, Lacy ER, Orlovsky Y, Baker A, Teplyakov A, Obmolova G, Heavner GA, Richter HT, Benson J (2010) Structural Basis for the Dual Recognition of IL-12 and IL-23 by Ustekinumab. Journal of Molecular Biology, 402, 797–812. https://doi.org/10.1016/j.jmb.2010.07.046

MacKerell AD, Jo S, Lakkaraju SK, Lind C, Yu W (2020) Identification and characterization of fragment binding sites for allosteric ligand design using the site identification by ligand competitive saturation hotspots approach (SILCS-Hotspots). Biochimica et Biophysica Acta (BBA) - General Subjects, 1864, 129519. https://doi.org/10.1016/j.bbagen.2020.129519

Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. Journal of Computer-Aided Molecular Design, 27, 221–234. https://doi.org/10.1007/s10822-013-9644-8

McKinney W (2010) Data Structures for Statistical Computing in Python - SciPy Proceedings. URL: https://proceedings.scipy.org/articles/Majora-92bf1922-00a Accessed 2 September 2024.

Pandya P, Sayers RO, Ting JP, Morshedian S, Torres C, Cudal JS, Zhang K, Fitchett JR, Zhang Q, Zhang FF, Wang J, Durbin JD, Carrillo JJ, Espada A, Broughton H, Qian Y, Afshar S (2020) Integration of phage and yeast display platforms: A reliable and cost effective approach for binning of peptides as displayed on-phage. PLoS ONE, 15, 1–19. https://doi.org/10.1371/journal.pone.0233961

Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2019) CatBoost: unbiased boosting with categorical features. https://doi.org/10.48550/arXiv.1706.09516

Raju TN (2000) The Nobel Chronicles. The Lancet, 355, 1022. https://doi.org/10.1016/S0140-6736(05)74775-9

Raybould MIJ, Marks C, Lewis AP, Shi J, Bujotzek A, Taddese B, Deane CM (2020) Thera-SAbDab: the Therapeutic Structural Antibody Database. Nucleic Acids Research, 48, D383–D388. https://doi.org/10.1093/nar/gkz827

Scannell JW, Blanckley A, Boldon H, Warrington B (2012) Diagnosing the decline in pharmaceutical R&D efficiency. Nature Reviews Drug Discovery, 11, 191–200. https://doi.org/10.1038/nrd3681

SchedMD (2022) Slurm on Google Cloud Platform.

Schrödinger (2021) The PyMOL Molecular Graphics System, Version 2.5.2. (version 2.5.2). Programmers: _:n796 URL: https://www.pymol.org/

The UniProt Consortium, Bateman A, Martin M-J, Orchard S, Magrane M, Adesina A, Ahmad S, Bowler-Barnett EH, Bye-A-Jee H, Carpentier D, Denny P, Fan J, Garmiri P, Gonzales LJDC, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Joshi V, Jyothi D, Kandasaamy S, Lock A, Luciani A, Luo J, Lussi Y, Marin JSM, Raposo P, Rice DL, Santos R, Speretta E, Stephenson J, Totoo P, Tyagi N, Urakova N, Vasudev P, Warner K, Wijerathne S, Yu CW-H, Zaru R, Bridge AJ, Aimo L, Argoud-Puy G, Auchincloss AH, Axelsen KB, Bansal P, Baratin D, Batista Neto TM, Blatter M-C, Bolleman JT, Boutet E, Breuza L, Gil BC, Casals-Casas C, Echioukh KC, Coudert E, Cuche B, De Castro E, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gaudet P, Gehant S, Gerritsen V, Gos A, Gruaz N, Hulo C, Hyka-Nouspikel N, Jungo F, Kerhornou A, Mercier PL, Lieberherr D, Masson P, Morgat A, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Poux S, Pozzato M, Pruess M, Redaschi N, Rivoire C, Sigrist CJA, Sonesson K, Sundaram S, Sveshnikova A, Wu CH, Arighi CN, Chen C, Chen Y, Huang H, Laiho K, Lehvaslaiho M, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Y, Zhang J (2025) UniProt: the Universal Protein Knowledgebase in 2025. Nucleic Acids Research, 53, D609–D617. https://doi.org/10.1093/nar/gkae1010

Ustach VD, Lakkaraju SK, Jo S, Yu W, Jiang W, Mackerell AD (2019) Optimization and Evaluation of Site-Identification by Ligand Competitive Saturation (SILCS) as a Tool for Target-Based Ligand Optimization. Journal of Chemical Information and Modeling, 59, 3018–3035. https://doi.org/10.1021/acs.jcim.9b00210

Wan H (2016) An Overall Comparison of Small Molecules and Large Biologics in ADME Testing. ADMET and DMPK, 4, 1–22. https://doi.org/10.5599/admet.4.1.276

Waskom M (2021) seaborn: statistical data visualization. Journal of Open Source Software, 6, 3021. https://doi.org/10.21105/joss.03021

Wingware (2025) Wing Pro.

Yoon C, Johnston SC, Tang J, Stahl M, Tobin JF, Somers WS (2000) Charged residues dominate a unique interlocking topography in the heterodimeric cytokine interleukin-12. The EMBO Journal, 19, 3530–3541. https://doi.org/10.1093/emboj/19.14.3530

Zabana Y, Rodríguez L, Lobatón T, Gordillo J, Montserrat A, Mena R, Beltrán B, Dotti M, Benitez O, Guardiola J, Domènech E, Garcia-Planella E, Calvet X, Piqueras M, Aceituno M, Fernández-Bañares F, Esteve M (2019) Relevant Infections in Inflammatory Bowel Disease, and Their Relationship With Immunosuppressive Therapy and Their Effects on Disease Mortality. Journal of Crohn's and Colitis, 13, 828–837. https://doi.org/10.1093/ecco-jcc/jjz013