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

Trevor J. Tanner[1]

[1]Independent Researcher. Correspondence: Trevor J. Tanner (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, which integrates machine learning, public databases, and structural analysis, targeted autoimmune diseases. This method identified IL12B as a promising candidate, uncovering a potential small molecule binding site involved in dimer formation. Inhibitors targeting this site could offer benefits such as shorter half-lives and more precise tissue distribution compared to existing biologics. These findings present a new opportunity to develop small molecule therapeutics for IL12B, potentially improving the safety and efficacy of treatments for conditions like inflammatory bowel disease (IBD).

Introduction

The discovery of new medical drugs has become increasingly challenging and costly despite significant technological advancements (1). As the easier targets have been exhausted, scientific progress now requires tackling more complex challenges (2).

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" (3), 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 (4).

This study introduces a computational pipeline designed to identify protein sites lacking known small molecule binders. We incorporated additional filtering criteria to minimize the risk of failure in subsequent drug discovery campaigns. Our approach builds on the proven efficacy of the computational SiteMap tool in identifying challenging protein sites (5), further optimizing its outputs through machine learning. We also prioritized selectivity by minimizing potential off-target effects from the outset, focusing on proteins with a reduced likelihood of unintended interactions in the human body.

Our research targets autoimmune diseases for several key reasons. Autoimmune conditions continue to dominate specialty drug spending, representing the largest share of net spending in this category, which accounted for 55% of total drug expenditures in 2023 (6). Biologics are a major component within this expanding field, presenting a significant opportunity

for small molecule alternatives. Moreover, autoimmune phenotypes are well-represented in single-gene knockout mouse model databases, providing a robust foundation for our research (7). By focusing on this area, we aim to address a critical and growing medical need while offering potential new treatment options with distinct pharmacological profiles.

Methods

Homologous protein file creation

We created a custom bioinformatics pipeline using Nextflow to filter out drug targets with significant primary homology to other human proteome proteins, potentially limiting off-target drug efficacy and safety (8). We used AlphaFold structural predictions (version one) of the human proteome as input (9). We extracted FASTA protein sequences from the input PDB files and blasted them against the UniProt reference human proteome using blastp (10). We filtered hits using a custom Python script with a 1E-62 maximal e-value cutoff (11), then re-aligned the remaining hits to the query protein sequence using mafft (12). We tabulated hit counts for each query protein with a custom bash script and saved them to a reference file. Figure 1b visualizes this process.

Custom helper scripts

We used numpy, pandas, and biopython libraries for data processing and basic analysis in our internal scripts (13–15).

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. We used the Drug Gene Interaction Database (DGIdb) to identify approved drug putative targets (16). We then filtered targets to include only those with experimental structures available in the UniProt database. We excluded targets with existing small molecule probes available according to the ProbeMiner database to maximize novelty and reduce academic and commercial competitors (17). We subsequently considered targets only if large-scale human genetics studies from the gnomAD database indicated the target tolerated deleterious mutations in at least one copy of the gene with a probability of at least 50% to minimize on-target host toxicity (18). We used the calculated locus-to-gene score from the Open Targets data source to consider further targets with genes that have a greater than 50% probability of causing at least one trait from openly published genome-wide association studies in humans to increase the likelihood of a target affecting a clinically desirable phenotype (19). We then excluded targets with a known metal cofactor according to the UniProt database to reduce potential problems during computational modeling. We also excluded targets if knocking them out was known to result in embryonic lethality in mice according to the JAX database to minimize on-target host toxicity. The next filter ensured that all targets were known to result in an abnormal inflammatory response when knocked out in mice to increase the likelihood of a target modulating inflammation in mammals. 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 Figure 1c.

Protein target and pocket funnel

After initial target filtering, we manually loaded AlphaFold structural predictions into the Schrodinger Maestro application and processed them with the Protein Preparation Wizard using default parameters and the OPLS4 force field (20,21). We then detected protein pockets using SiteMap with default parameters and exported them to a comma-separated file (22). We processed the file in a custom Python script 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. We then subjected pockets to further screening as described in the “supervised learning prediction and filtering” section. We manually submitted promising targets to the online FTMap server for solvent mapping in protein-protein interaction mode and checked if at least one hot spot had a minimum of 16 different probes and one or two additional nearby hotspots to provide evidence of druggability (23,24). We then manually checked targets with evidence of druggability for post-translational modifications at the dbPTM web service and queried for relevant structures at the RCSB protein data bank website (25,26). We used the Chimera software package to visualize the residues involved with SiteMap pockets on targets, the residues involved with targets’ known post-translational modifications, and known target multimeric quaternary states (27). We used the PyMol package to visualize relevant probe hot spots for target druggability from FTMap simulations (28). The protein target and pocket funnel are visualized in Figure 1a.

Supervised training

Training data points for Schrodinger Sitemap pockets and druggability labels were extracted from prior research (29). SiteMap pocket data was used in its original numeric form,

and druggability labels of “undruggable,” “difficult,” and “druggable” were coded as 0, 1, and 2, respectively. Outliers were removed from the training dataset using the LocalOutlierFactor function from the Scikit library (30). A balanced random forest model was utilized as a supervised classification method with a random initialization state of 42 and otherwise default model parameters from the Scikit library. Druggability labels were further coded into binary values, where a value was marked as “true” if it exceeded the threshold of 0 or 1. Model performance was evaluated using Matthew’s correlation coefficient. Five-fold stratified cross-validation was used with a random initialization state of 42, and data was shuffled for cross-validation of model performance. Cross-validation performance data was exported and imported into Microsoft Excel to create a bar plot for visualization. The full training dataset without outliers was used to train the balanced random forest using the druggability threshold that optimized model performance. The supervised training process is visualized in Figure 1d.

Supervised prediction and filtering

We checked SiteMap pockets from filtered targets of the pocket funnel pipeline to ensure they were not outliers from the main training dataset using the LocalOutlierFactor function from the Scikit library with the “novelty=True” condition. We then filtered pockets to ensure that at least 50% of the amino acids in the pocket had a predicted AlphaFold quality of “very high” (>90 pLDDT) and that at least 50% of the amino acids in the entire region from the first amino acid in the pocket to the last amino acid in the pocket were also of “very high” quality. We made druggability predictions for the remaining pockets using the balanced random forest model, and pockets arbitrarily passed filtering if the model indicated a greater than 50% probability of druggability. The supervised prediction and filtering process is visualized in Figure 1d.

Data and Code Availability

This paper's scripts, data, and results are publicly available in a GitHub repository at https://github.com/tantrev/drug-target-pocket-funnel to ensure transparency and reproducibility. The repository includes custom scripts, databases used, processed data, and specific analysis results. However, some intermediate results generated using Schrödinger's commercial software have been excluded from the public repository to comply with the terms of Schrödinger's End User License Agreement. As a result, certain scripts may not function properly without the appropriate files that contain Schrödinger-specific outputs. These scripts are provided for reference purposes. Recreating the original dataset may require a Schrödinger software license. We have endeavored to provide all other non-Schrödinger files necessary to understand and reproduce our work. The repository README file provides detailed instructions.

Results

Model performance evaluation revealed that the balanced random forest model achieved the highest accuracy when predicting both difficult and druggable pockets, as opposed to just druggable pockets (Fig. 2a).

The initial target filtering pipeline, which utilized various public databases and a custom homologous protein file, identified 10 candidate targets. Further filtering based on the structural quality of the AlphaFold predictions and the presence of druggable pockets narrowed the pool down to one target: IL12B (Fig. 2b).

IL12B is an interleukin that frequently forms dimers with other proteins and itself. The most well-studied heterodimeric forms of IL12B are IL-12, formed with IL12A, and IL-23, formed with IL23A. The monomeric form of IL12B, the homodimeric form of IL12B, and the heterodimeric forms IL-12 and IL-23 are all involved in inflammatory conditions such as autoimmune disorders and cancer (31).

SiteMap analysis of the monomeric AlphaFold structure of IL12B identified two potential binding sites. Site A is part of IL12B's canonical dimeric interface, while site B is proximal to, but not directly involved in, the dimeric interface (Fig. 2c, Fig. 2e).

Examination of known post-translational modifications (PTMs) revealed that site B has more direct PTM modifications of its pocket residues compared to site A (Fig. 2d).

FTMap simulations provided evidence for the druggability of site A, as it contains a main hot spot with 18 probe clusters and two additional nearby hotspots (Fig. 2f).

Discussion

Molecular characterization

Our computational screening identified IL12B as a promising target for small-molecule drugs. This protein, which currently lacks known small molecule probes, features a potentially druggable pocket, referred to as site A, involved in multimer formation. Both static conformational and flexible solvent-based analyses confirm the druggability of site A. Notably, site A exhibits minimal primary sequence homology with other human proteins, which could reduce the risk of off-target effects. Additionally, it has few residues involved in post-

translational modifications, suggesting that a small molecule could target multiple IL12B isoforms, irrespective of their post-translational modifications.

Developing both covalent and non-covalent probes targeting site A appears feasible. These probes could inhibit dimer formation or tag intracellular IL12B monomers for degradation using modalities such as PROTACs. The optimal probe strategy will likely depend on several factors, including the clinical indication, the specific IL12B isoform, protein location (intracellular or extracellular), and the target tissue. For instance, targeting gastrointestinal tissues could be achieved by engineering a drug with a very short half-life and utilizing an extended-release formulation to limit systemic exposure (32).

Clinical applications

The clinical utility of targeting IL12B is well established, as demonstrated by the approval of the antibody ustekinumab for treating inflammatory bowel disease (IBD), including ulcerative colitis, Crohn's disease, and psoriasis. Among these, IBD presents a stronger economic justification for small-molecule drug development, given the availability of many effective treatments for psoriasis. A small molecule inhibitor targeting IL12 B's dimeric interface could offer advantages over ustekinumab, such as a shorter half-life and more restricted tissue distribution, potentially resulting in a safer treatment option for IBD patients (33). Such a drug could be designed as a covalent probe targeting the intracellular, nascently synthesized monomeric form of IL12B.

In addition to IBD, Behcet's disease could also benefit from IL12B inhibition. Behcet's disease is a chronic inflammatory disorder characterized by recurrent oral and genital ulcers, skin lesions, and ocular inflammation (34). Mouth ulcers, which affect nearly all patients, are particularly debilitating, making eating and talking difficult. IL-12 and IL-23, which share the

IL12B subunit, play a critical role in the pathogenesis of Behcet's disease (35–37). Clinical trials have shown that the systemic inhibition of IL-12 and IL-23 reduces mouth ulcers in Behcet's patients using antibodies (38). A small molecule degrader targeting IL12B could offer a safer, more targeted approach to treating these ulcers.

Small molecule immunomodulators with limited distribution to the gastrointestinal tract, compared to systemic distribution, have been associated with fewer patient infections (39). This is particularly important because IBD medications have been linked to a 134.7-fold increased risk of infections (40). Serious infections can increase the risk of dementia; for instance, pneumonia patients have a 2.83-fold increased risk of developing dementia (41). Consequently, IBD patients have an elevated dementia risk, while those undergoing colectomy, which reduces medication use, show a decreased risk (42). Ustekinumab and IL-23 inhibitors have also been associated with major adverse cardiovascular events in patients with high cardiovascular risk (43). (39). However, a GI-limited small molecule inhibitor could be safer for elderly patients at elevated dementia risk from immunomodulators, especially if blood pressure is managed, as IL23A inhibition only causes vascular dysfunction in hypertensive mice (44).

ECCO treatment guidelines for IBD recommend discontinuing immunosuppressive therapy in cases of select infections (45). In such scenarios, a small molecule inhibitor targeting IL12B, with its shorter half-life, could offer a significant advantage over biologics by allowing for more rapid clearance from the body upon discontinuation.

Furthermore, preclinical studies suggest promising applications for an IL12B small molecule degrader that have yet to be explored with biologics. For instance, antibodies targeting the extracellular, monomeric form of IL12B have reduced tumor growth in prostate, breast, and pancreatic cancer models in mice (46–48). A PROTAC targeting the intracellular, monomeric

form of IL12B for degradation could potentially achieve more favorable tissue distribution properties than biologic alternatives.

Analytic limitations

This research has limitations in its predictive power due to its computational nature, reliance on external datasets, simulations, and the absence of physical experiments. The dataset used for predicting pocket druggability is small and imbalanced across target variable classes. The AlphaFold structural predictions were also trained on potentially problematic X-ray structures. Specifically, the AlphaFold prediction for IL12B may have been derived from the 1F42 structural deposit, which dates back to 1997 and was not accompanied by electron diffraction data. The absence of this data limits the ability to assess the quality of the structure and refine it using modern software tools.

Conclusion

Our computational analysis identified IL12B as a promising target for small molecule drugs, featuring a potentially druggable site A involved in dimer formation. The unique properties of this pocket present opportunities for developing selective probes.

A crucial next step is to establish a high-quality, reproducible crystal system of monomeric IL12B. Using computational methods would enable X-ray fragment screening to identify small molecule binders and guide rational fragment library selection. Experimental validation of site A's druggability remains essential.

Both covalent and non-covalent approaches targeting site A appear viable. Small molecule inhibitors of IL12B could offer advantages over existing biologics, such as shorter half-lives and tissue-restricted distribution, potentially enhancing safety for conditions like IBD and Behcet's disease by reducing the risk of serious infections.

Although limited by its computational nature, this work lays a promising foundation for further experimental validation and drug discovery efforts targeting IL12B. Our approach integrates established clinical knowledge with innovative computational methods to identify a promising small molecule drug target.

Conflict of Interest Statement

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.

Acknowledgements

We gratefully acknowledge the support of several organizations that made this work possible. We thank 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 (49). The Chemicalize platform was invaluable for visualizing chemical structures documented in various crystallization protocols for IL12B protein that were extracted from the literature.

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. Dr. Law's and Dr. Balijepalli's guidance and expert feedback significantly enhanced the author's ability to articulate, refine, and communicate research. The Everyday Entrepreneur Program, taught by Jon Beutler, provided training in market research, competitive analysis, and minimum viable product development.

The basic foundation for the skill sets used in this work was laid during two summers Trevor J. Tanner spent in the National Science Foundation's Research Experience for Undergraduates (REU) program. This formative experience, made possible by NSF award number 0835713, took place in the Aspuru-Guzik group under the mentorship of Dr. Dmitrij Rappoport.

We gratefully acknowledge the contributions of advanced language models, specifically Claude 3.5 Sonnet (50) and GPT-4 (51), which aided in refining paper titles, enhancing

abstracts, clarifying ideas, offering valuable feedback, and assisting with various aspects of manuscript preparation throughout our research process.

References

1. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov. 2012 Mar;11(3):191–200.

2. Cowen T. The Great Stagnation: How America Ate All The Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better: A Penguin eSpecial from Dutton. Penguin; 2011. 110 p.

3. Raju TN. The Nobel Chronicles. The Lancet. 2000 Mar 18;355(9208):1022.

4. Wan H. An Overall Comparison of Small Molecules and Large Biologics in ADME Testing. ADMET DMPK. 2016 Mar 30;4(1):1–22.

5. Liu T, Altman R. Identifying Druggable Targets by Protein Microenvironments Matching: Application to Transcription Factors. CPT Pharmacomet Syst Pharmacol. 2014;3(1):93.

6. Biggs S, Long D. Insights Into the 2023 U.S. Pharmaceutical Market [Internet]. 2023 [cited 2024 Sep 2]. Available from: https://www.iqvia.com/locations/united-states/blogs/2023/07/insights-into-the-2023-us-pharmaceutical-market

7. Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ, et al. Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology. Nucleic Acids Res. 2021 Jan 8;49(D1):D981–7.

8. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017 Apr;35(4):316–9.

9. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.

10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403–10.

11. Louie B, Higdon R, Kolker E. A Statistical Model of Protein Sequence Similarity and Function Similarity Reveals Overly-Specific Function Predictions. PLoS ONE. 2009 Oct 21;4(10):e7546.

12. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr 1;30(4):772–80.

13. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009 Jun 1;25(11):1422–3.

14. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020 Sep;585(7825):357–62.

15. McKinney W. Data Structures for Statistical Computing in Python - SciPy Proceedings [Internet]. 2010 [cited 2024 Sep 2]. Available from: https://proceedings.scipy.org/articles/Majora-92bf1922-00a

16. Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, et al. DGIdb 5.0: rebuilding the drug–gene interaction database for precision medicine and drug discovery platforms. Nucleic Acids Res. 2024 Jan 5;52(D1):D1227–35.

17. Antolin AA, Tym JE, Komianou A, Collins I, Workman P, Al-Lazikani B. Objective, Quantitative, Data-Driven Assessment of Chemical Probes. Cell Chem Biol. 2018 Feb 15;25(2):194-205.e5.

18. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational constraint map using variation in 76,156 human genomes. Nature. 2024 Jan;625(7993):92–100.

19. Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res. 2021 Jan 8;49(D1):D1311–20.

20. Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, et al. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. J Chem Theory Comput. 2021 Jul 13;17(7):4291–300.

21. Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013 Mar 1;27(3):221–34.

22. Halgren T. New Method for Fast and Accurate Binding-site Identification and Analysis. Chem Biol Drug Des. 2007;69(2):146–8.

23. Kozakov D, Hall DR, Chuang GY, Cencic R, Brenke R, Grove LE, et al. Structural conservation of druggable hot spots in protein–protein interfaces. Proc Natl Acad Sci. 2011 Aug 16;108(33):13528–33.

24. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc. 2015 May;10(5):733–55.

25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235–42.

26. Li Z, Li S, Luo M, Jhong JH, Li W, Yao L, et al. dbPTM in 2022: an updated database for exploring regulatory networks and functional associations of protein post-translational modifications. Nucleic Acids Res. 2022 Jan 7;50(D1):D471–9.

27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605–12.

28. Schrödinger. The PyMOL Molecular Graphics System, Version 2.5.2. [Internet]. 2021 [cited 2024 Sep 3]. Available from: https://www.pymol.org/

29. Halgren TA. Identifying and Characterizing Binding Sites and Assessing Druggability. J Chem Inf Model. 2009 Feb 23;49(2):377–89.

30. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12(85):2825–30.

31. Yoon C, Johnston SC, Tang J, Stahl M, Tobin JF, Somers WS. Charged residues dominate a unique interlocking topography in the heterodimeric cytokine interleukin-12. EMBO J. 2000 Jul 17;19(14):3530–41.

32. Adepu S, Ramakrishna S. Controlled Drug Delivery Systems: Current Status and Future Directions. Molecules. 2021 Sep 29;26(19):5905.

33. Al-Bawardy B, Shivashankar R, Proctor DD. Novel and Emerging Therapies for Inflammatory Bowel Disease. Front Pharmacol [Internet]. 2021 Apr 14 [cited 2024 Sep 2];12. Available from: https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2021.651415/full

34. Taylor J, Glenny AM, Walsh T, Brocklehurst P, Riley P, Gorodkin R, et al. Interventions for the management of oral ulcers in Behçet's disease. Cochrane Database Syst Rev. 2014 Sep 25;2014(9):CD011018.

35. Kappen JH, Medina-Gomez C, Hagen PM van, Stolk L, Estrada K, Rivadeneira F, et al. Genome-Wide Association Study in an Admixed Case Series Reveals IL12A as a New Candidate in Behçet Disease. PLOS ONE. 2015 Mar 23;10(3):e0119085.

36. Mizuki N, Meguro A, Ota M, Ohno S, Shiota T, Kawagoe T, et al. Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behçet's disease susceptibility loci. Nat Genet. 2010 Aug;42(8):703–6.

37. Remmers EF, Cosan F, Kirino Y, Ombrello MJ, Abaci N, Satorius C, et al. Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease. Nat Genet. 2010 Aug;42(8):698–702.

38. Mirouse A, Barete S, Desbois AC, Comarmond C, Sène D, Domont F, et al. Long-Term Outcome of Ustekinumab Therapy for Behçet's Disease. Arthritis Rheumatol. 2019;71(10):1727–32.

39. Pallmer K, Timeus S, Laoun R. P500 The use of budesonide results in lower risk for infections compared to systemic glucocorticoids. J Crohns Colitis. 2022 Jan 1;16(Supplement_1):i462–3.

40. Zabana Y, Rodríguez L, Lobatón T, Gordillo J, Montserrat A, Mena R, et al. Relevant Infections in Inflammatory Bowel Disease, and Their Relationship With Immunosuppressive Therapy and Their Effects on Disease Mortality. J Crohns Colitis. 2019 Jul 25;13(7):828–37.

41. Chu CS, Liang CS, Tsai SJ, Bai YM, Su TP, Chen TJ, et al. Bacterial pneumonia and subsequent dementia risk: A nationwide cohort study. Brain Behav Immun. 2022 Jul 1;103:12–8.

42. Garcia-Argibay M, Hiyoshi A, Montgomery S. Association between dementia risk and ulcerative colitis, with and without colectomy: a Swedish population-based register study. BMJ Open. 2023 Dec 1;13(12):e074110.

43. Tzellos T, Kyrgidis A, Trigoni A, Zouboulis CC. Association of ustekinumab and briakinumab with major adverse cardiovascular events: An appraisal of meta-analyses and industry sponsored pooled analyses to date. Dermatoendocrinol. 2012 Jul 1;4(3):320–3.

44. Balasubbramanian D, Goodlett BL, Mitchell BM. Is IL-12 pro-inflammatory or anti-inflammatory? Depends on the blood pressure. Cardiovasc Res. 2019 May 1;115(6):998–9.

45. Kucharzik T, Ellul P, Greuter T, Rahier JF, Verstockt B, Abreu C, et al. ECCO Guidelines on the Prevention, Diagnosis, and Management of Infections in Inflammatory Bowel Disease. J Crohns Colitis. 2021 Jun 1;15(6):879–913.

46. Kundu M, Roy A, Pahan K. Selective neutralization of IL-12 p40 monomer induces death in prostate cancer cells via IL-12–IFN-γ. Proc Natl Acad Sci. 2017 Oct 24;114(43):11482–7.

47. Kundu M, Raha S, Roy A, Pahan K. Regression of Triple-Negative Breast Cancer in a Patient-Derived Xenograft Mouse Model by Monoclonal Antibodies against IL-12 p40 Monomer. Cells. 2022 Jan;11(2):259.

48. Sheinin M, Mondal S, Pahan K. Neutralization of p40 Homodimer and p40 Monomer Leads to Tumor Regression in Patient-Derived Xenograft Mice with Pancreatic Cancer. Cancers. 2023 Jan;15(24):5796.

49. Chemaxon. Chemicalize. 2024 [cited 2024 Sep 3]. Chemicalize. Available from: https://chemicalize.com

50. Anthropic. Claude 3 Opus [Internet]. 2024 [cited 2024 Sep 3]. Available from: https://claude.ai/

51. OpenAI. ChatGPT-4 [Internet]. 2024 [cited 2024 Sep 3]. Available from: https://chatgpt.com/

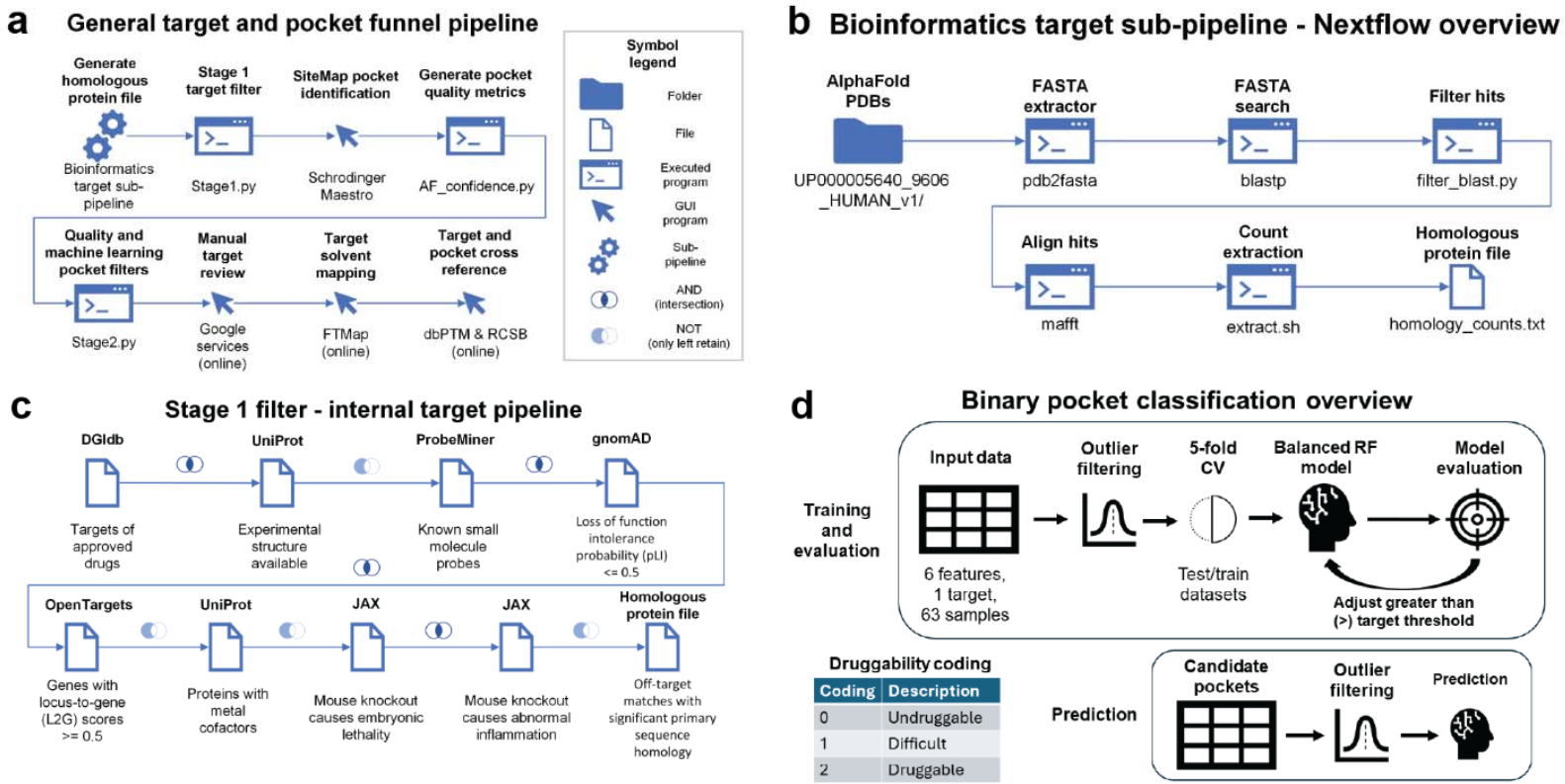

Fig 1. | Target and pocket identification and prediction. a, All intermediate steps and interactions with graphical user interface (GUI) based programs are performed manually. Command arguments not listed. **b**, All steps are executed using the Nextflow automation tool. Command arguments not listed. **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 contained within Stage2.py. Target threshold refers to minimal value for differentiating between druggable and undruggable pockets. CV stands for cross validation and RF stands for random forest.

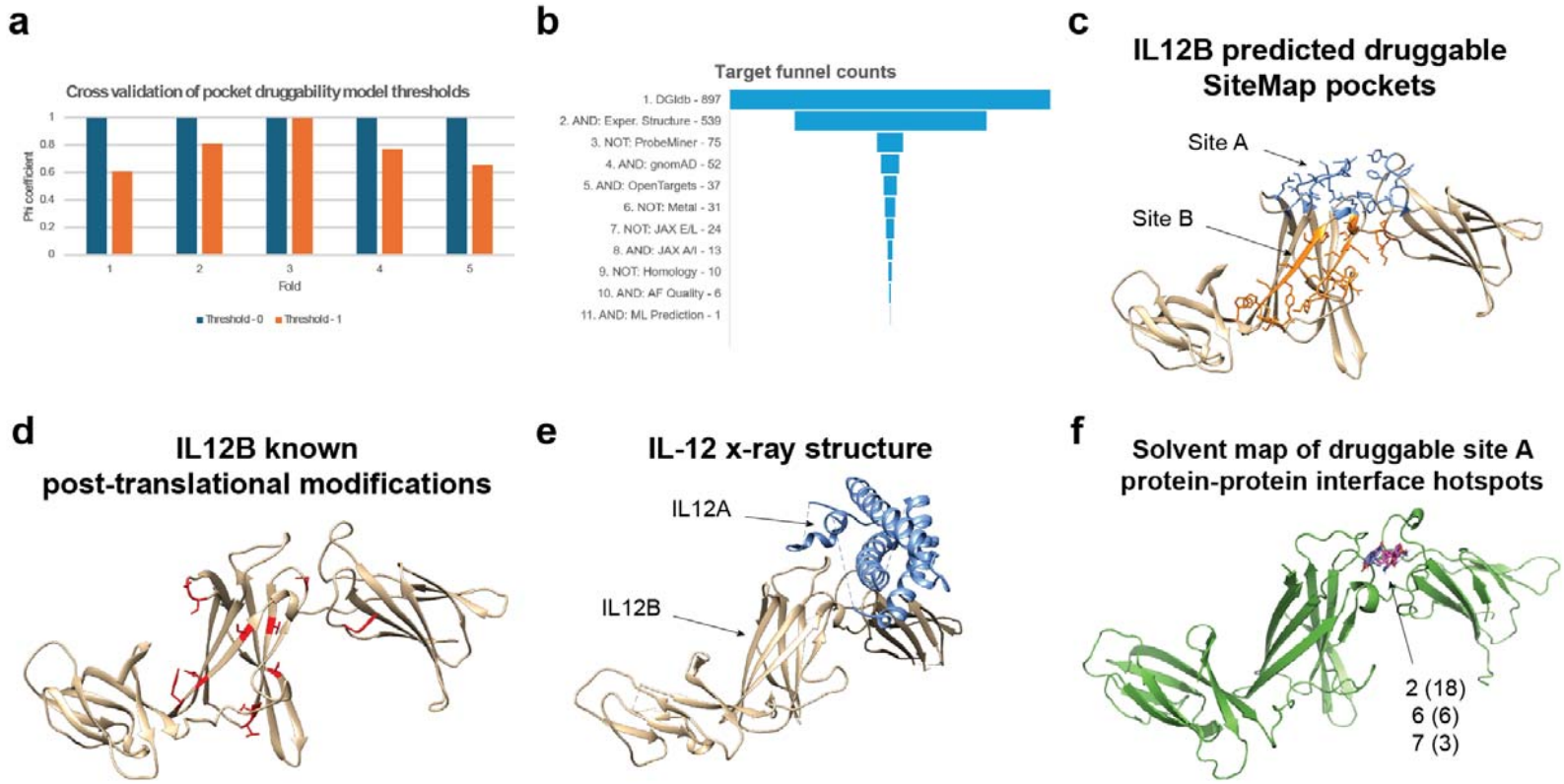

Fig. 2 | Model, target, and pocket results. a) Phi coefficient is Matthew's correlation coefficient (1/-1 is perfect agreement/disagreement). Threshold of zero yields perfect predictions. **b)** E/L stands for embryonic lethal and A/I for abnormal inflammation. **c)** AlphaFold model. Shared AA color indicates pocket membership. **d)** AlphaFold model. Red AAs are known PTMs. Shows site A has less nearby PTMs than site B. **e)** Crystal structure 1F45 shows AAs involved in IL12B heterodimer formation. **f)** Crystal structure 1F42 used in FTMap. Only hotspots near site A shown. Each row of numbers indicates first the hotspot rank and the number inside parentheses indicates how many probe clusters are in the hotspot. Details in discussion explain evidence of druggability.