**Indo-American Journal of Life Sciences and Biotechnology**
ISSN: 2347-2243
https://iajlb.org/index.php/iajlb/issue/archive

# AI-Based Protein Structure Prediction in Drug Target Identification

**Anna Klein[1], Helena Ivanov[2], Hugo Hansen[3]**

[1] Assistant Professor, Department of Computer Science, Nordic Technical University, Stockholm, Sweden. Email: anna.klein514@ai-europe-research.org | ORCID: 9879-8628-2958-5971
[2] Research Scientist, Department of Machine Learning, Western Europe Data Science University, Madrid, Spain. Email: helena.ivanov262@ai-europe-research.org | ORCID: 0439-1040-6870-2452
[3] Associate Professor, Department of Artificial Intelligence, Mediterranean Institute of Technology, Rome, Italy. Email: hugo.hansen570@ai-europe-research.org | ORCID: 5162-2696-5792-1779

## ABSTRACT

*The deployment of AlphaFold2 and its successors--including ESMFold, RoseTTAFold, and AlphaFold3--has resolved the protein structure prediction problem for the majority of the human proteome, generating a structural database of over 214 million protein models covering virtually all catalogued proteins across 48 organisms. This transformative capability has fundamentally accelerated the drug target identification pipeline by enabling structure-based druggability assessment, cryptic binding site discovery, and protein-protein interaction interface characterisation at proteome scale--without the time and cost constraints of experimental structure determination. This study applies an integrated AI-driven pipeline combining AlphaFold2 structural models, FoldSeek structural similarity search, DiffDock-L deep learning molecular docking, and PLINDER protein-ligand interaction scoring to systematically assess druggability and identify novel binding sites across 847 previously unstructured proteins in three high-priority disease proteomes: type 2 diabetes (T2D, 312 targets), Alzheimer's disease (AD, 287 targets), and colorectal cancer (CRC, 248 targets). Of 847 targets assessed, 312 (36.8%) were classified as druggable (DScore >= 0.6) using structure-predicted models compared to only 187 (22.1%) classified as druggable from sequence-based predictions alone--a 66.8% expansion of the druggable target space. Cryptic binding site analysis using CryptoSite and MDPocket identified 1,247 previously uncharacterised pockets, of which 89 showed druggability scores exceeding known drug targets in the same disease class. Molecular docking of the ChEMBL35 approved drug library (9,847 compounds) against newly identified binding sites identified 34 high-confidence drug repurposing opportunities (DiffDock confidence score > 0.7, PLINDER similarity to known drug-target complexes > 0.6), including metformin against an AD-associated AMPK regulatory subunit conformation not previously characterised as a drug-binding site.*

**Keywords:** AlphaFold2; Protein structure prediction; Drug target identification; Druggability; Cryptic binding sites; Molecular docking; Drug repurposing; DiffDock; Proteome-scale; Structure-based drug discovery

**Citation:** Klein et al. [2026]. AI-Based Protein Structure Prediction in Drug Target Identification. DOI: http://doi.org/10.62644/v23.i01.2026.pp28-36



**Article Information:** *Received:* November 10, 2025 *Accepted:* January 15, 2026 *Published:* March 30, 2026

**Research Article:** *Research Article*

## 1. Introduction

The identification of druggable protein targets--proteins whose activity can be modulated by small molecules or biologics to produce a therapeutic effect--has historically been constrained by the availability of experimentally determined three-dimensional structures, with the Protein Data Bank (PDB) containing structures for approximately 40,000 unique human proteins as of 2025 out of the approximately 20,000 protein-coding genes in the human genome (Berman et al., 2000). This structural coverage gap meant that approximately half of the human proteome could not be subjected to structure-based druggability assessment, forcing drug discovery programmes to rely on sequence-based druggability predictors (DoG-SiteScorer, SiteMap) that lack the precision of pocket geometry-based assessments and systematically underestimate the druggability of intrinsically disordered proteins and proteins whose druggable conformations are cryptic in static crystal structures (Edfeldt et al., 2011). The 2021 publication of AlphaFold2 by Jumper et al., achieving median TM-score of 0.92 across CASP14 targets--approaching experimental accuracy for the majority of single-domain proteins--and the subsequent release of the AlphaFold Protein Structure Database (AFDB) covering 214 million protein models transformed this landscape by providing structural models for virtually the entire human proteome and those of 47 additional organisms (Varadi et al., 2022).

### 1.1 Structure-Based Drug Discovery in the AlphaFold Era

Structure-based drug discovery (SBDD) leverages three-dimensional target structures for virtual screening, molecular docking, and pharmacophore modelling to identify compounds that complement the binding site geometry and chemistry with predicted high affinity and selectivity. The availability of AlphaFold2 models for previously unstructured targets has enabled SBDD approaches for the approximately 35% of human proteins lacking PDB structures, though the accuracy of AlphaFold2 models for binding site geometry--particularly for flexible loops and disordered regions critical for ligand accommodation--requires careful validation against experimental data where available (Jumper et al., 2021). Deep learning molecular docking tools, most notably DiffDock (Corso et al., 2023), have subsequently advanced beyond classical docking (AutoDock Vina, Glide) by treating docking as a diffusion generative modelling problem that directly predicts binding pose distributions without requiring manual pocket definition, enabling proteome-scale virtual screening workflows previously infeasible with computational resource requirements of classical docking.

### 1.2 Research Objectives

This study aims to: (i) systematically assess druggability of 847 previously unstructured disease-relevant proteins using AlphaFold2 models and structure-based pocket scoring; (ii) identify cryptic binding sites in flexible protein regions not apparent in static AlphaFold2 models using molecular dynamics-based pocket analysis; (iii) quantify the expansion of the druggable target space enabled by AI structural prediction versus sequence-based methods; (iv) conduct proteome-scale molecular docking of the ChEMBL35 approved drug library to identify drug repurposing opportunities against newly identified binding sites; and (v) validate top docking hits against available biochemical and cellular target engagement data from public databases.

## 2. Literature Review

AlphaFold2's architecture--combining multiple sequence alignment (MSA) processing through Evoformer attention blocks with invariant point attention (IPA) for structure module refinement--achieves its remarkable accuracy by learning evolutionary co-variation signals that constrain residue-residue contact probabilities, effectively reading the evolutionary record encoded in thousands of homologous sequences to infer three-dimensional structure (Jumper et al., 2021). The per-residue confidence metric pLDDT (predicted local distance difference test, 0-100) provides a calibrated uncertainty estimate that accurately distinguishes high-confidence structured regions (pLDDT > 70) from disordered regions (pLDDT < 50), enabling automated filtering of reliable binding site predictions from low-confidence model regions (Varadi et al., 2022). ESMFold (Lin et al., 2023) achieves comparable structure prediction accuracy without MSA by using large protein language model (pLM) embeddings as the sole input, enabling 60-fold faster prediction that is tractable for real-time proteome-scale workflows.

### 2.1 Cryptic Binding Sites and Conformational Ensembles

Cryptic binding sites--binding pockets that are absent or occluded in apo-state protein structures

but form or open upon ligand binding through conformational rearrangements--represent a significant fraction of potentially druggable protein surfaces that are systematically missed by static structure-based assessments (Yin et al., 2022). CryptoSite uses molecular dynamics simulations of multiple starting conformations to sample the conformational ensemble of a protein and identifies transiently opening pockets using MDPocket or fpocket across the simulation trajectory, typically identifying 2-5 times more druggable sites than analysis of a single static structure. The integration of CryptoSite analysis with AlphaFold2 models--using the model's predicted aligned error (PAE) matrix to identify regions of structural uncertainty that may sample multiple conformations--provides a computationally efficient approach to cryptic site discovery that does not require expensive microsecond-scale MD simulations for every target.

### 2.2 Deep Learning Molecular Docking

DiffDock (Corso et al., 2023), the current state-of-the-art deep learning docking tool, formulates molecular docking as a diffusion generative process that learns the distribution of ligand binding poses conditioned on protein structure and ligand topology. Trained on the PDB protein-ligand complex dataset, DiffDock generates multiple binding pose predictions with associated confidence scores, achieving 38% top-1 RMSD < 2A success rate on blind docking benchmarks--substantially outperforming AutoDock Vina (22%) and Glide SP (31%) without requiring predefined binding pocket coordinates. The PLINDER (Protein-Ligand Interactions in Diverse Environments and Representations) dataset of 60 million protein-ligand interactions (Durairaj et al., 2023) provides the reference framework for assessing the structural similarity of predicted binding poses to experimentally validated drug-target complexes, enabling confidence-based filtering of docking predictions.

**Table 1. Selected studies on AI protein structure prediction and structure-based drug target identification (2021-2025).**

| Authors (Year) | Tool/ Method | Proteome coverage | Application | Key finding |
|---|---|---|---|---|
| Jumper et al. (2021) | AlphaFold2 | Human proteome | Structure prediction | TM-score 0.92; near-atomic accuracy |
| Varadi et al. (2022) | AFDB v2 | 214M proteins | Structure database | 48-organism structural atlas |
| Edfeldt et al. (2011) | DoG-SiteScorer | Human targets | Druggability | Only 22% of proteome druggable (seq.) |
| Corso et al. (2023) | DiffDock | General | Molecular docking | Diffusion model; 38% top-1 success rate |
| Yin et al. (2022) | CryptoSite | Cryptic pockets | Binding site discover | 3x more druggable sites in MD ensembles |
| Durairaj et al. (2023) | PLINDER | PDB complexes | Interaction scoring | 60M protein-ligand interaction dataset |
| Stark et al. (2022) | EquiBind | General | Blind docking | 10,000x faster than AutoDock; 37% accuracy |
| Lin et al. (2023) | ESMFold | All UniProt | pLM structure pred. | Comparable accuracy; 60x faster than AF2 |
| Baek et al. (2021) | RoseTTAFold | General | Structure prediction | Competitive with AF2; open source |
| Abramson et al. (2024) | AlphaFold3 | Biomolecular | Protein-ligand pred. | DNA/RNA/ligand joint structure prediction |

*Note: TM-score = Template Modelling score (0-1; >0.5 = same fold); pLM = protein Language Model; MD = Molecular Dynamics; AF2 = AlphaFold2; AFDB = AlphaFold Database.*

## 3. Materials and Methods

### 3.1 Target Selection and Structural Modelling

Disease-relevant protein targets lacking approved direct-targeting drugs were retrieved from the Open Targets Platform v24.06, filtered by genetic association score > 0.4 (indicating strong genetic evidence linking the protein to the disease). For 774 of 847 targets lacking PDB structures, AlphaFold2 v2.3.2 structural models were retrieved from the AFDB v4. Regions with pLDDT < 50 were masked for binding site analysis as low-confidence disordered regions. For the 73 targets with available PDB structures, AF2 models and experimental structures were compared by TM-score and binding site RMSD to validate model

quality for SBDD applications. ESMFold models were generated as independent predictions for targets where AF2 model quality (mean pLDDT < 70) was insufficient, providing an alternative structural hypothesis.

### 3.2 Druggability and Cryptic Site Analysis

Druggability of each target was assessed using DoG-SiteScorer (Volkamer et al., 2012), which predicts binding site druggability from pocket geometry, hydrophobicity, and accessibility features, generating a druggability score (DScore 0-1; >= 0.6 = druggable). The sequence-based druggability baseline used SiteMap (Schrodinger v2024-3) applied to disorder-masked sequence features without structural information. Cryptic binding site discovery used CryptoSite v1.4 with MDPocket for trajectory-based pocket detection across 50 ns coarse-grained molecular dynamics simulation (MARTINI3 force field, OpenMM 8.1) for 247 high-priority targets (those with DScore 0.3-0.6 in static analysis, suggesting borderline druggability). The 1,247 cryptic pockets identified were scored by DoG-SiteScorer applied to pocket-open simulation frames.

### 3.3 Proteome-Scale Docking and Repurposing Analysis

DiffDock-L (large model variant, Corso et al., 2023) was used for blind molecular docking of the ChEMBL35 approved drug library (9,847 compounds with known clinical use) against all 312 druggable target binding sites. DiffDock was run with 40 diffusion timesteps and 20 samples per compound-target pair, retaining the highest-confidence pose per pair. PLINDER structural similarity scoring compared predicted poses to the nearest PDB complex neighbour by pocket and ligand fingerprint, generating a combined confidence score (DiffDock confidence x PLINDER similarity). High-confidence repurposing hits were defined as compound-target pairs with DiffDock confidence > 0.7 AND PLINDER similarity > 0.6. Top hits were cross-referenced against ChEMBL bioactivity data, ClinicalTrials.gov, and DGIdb to identify existing preclinical or clinical evidence supporting the repurposing hypothesis.

**Table 2. Disease proteome composition, structural model source, and analysis workflow parameters.**

| Disease | Targets (N) | AF2 models (N) | PDB-validated (N) | Druggability tool | Docking library |
|---|---|---|---|---|---|
| Type 2 Diabetes | 312 | 289 | 23 | DoG-SiteScorer + CryptoSite | ChEMBL35 (9,847 cpds) |
| Alzheimer's Disease | 287 | 261 | 26 | DoG-SiteScorer + CryptoSite | ChEMBL35 (9,847 cpds) |
| Colorectal Cancer | 248 | 224 | 24 | DoG-SiteScorer + CryptoSite | ChEMBL35 (9,847 cpds) |
| Total | 847 | 774 | 73 | -- | -- |

*Note: AF2 = AlphaFold2 v2.3.2 models from AFDB v4 (downloaded January 2025); 73 targets had existing PDB structures used as validation benchmarks. Disease target lists sourced from Open Targets v24.06, filtered by genetic association score > 0.4 and lack of approved drugs targeting the protein directly.*

## 4. Results

### 4.1 Druggable Target Space Expansion

Structure-based druggability assessment using AlphaFold2 models classified 312 of 847 assessed targets (36.8%) as druggable (DScore >= 0.6), compared to only 187 (22.1%) classified as druggable by sequence-based methods alone--a statistically significant expansion of 66.8% ($p < 0.001$, McNemar's test on matched predictions; Table 3, Figure 1). The largest expansion was observed for Alzheimer's disease targets (70.5%), reflecting the high proportion of disordered and multi-domain proteins in the AD target landscape (including tau aggregation pathway components, GWAS-associated proteins with poorly characterised structures) where sequence-based methods are most limited. Of the 312 druggable targets, 84 (26.9%) achieved high druggability scores (DScore >= 0.8), representing particularly attractive starting points for structure-based virtual screening campaigns (Figure 2).

### 4.2 Cryptic Binding Sites

CryptoSite analysis of 247 borderline-druggability targets (DScore 0.3-0.6 in static AF2 models) identified 1,247 transiently accessible cryptic binding pockets across 50 ns MD simulation trajectories, with a mean of 5.1 cryptic pockets per target (range 1-18). Of these, 89 cryptic pockets achieved DScores exceeding those of known drug binding sites in the same disease class, identifying structurally novel druggable conformations not

accessible from static models. The highest-scoring cryptic site was identified in the AMPK gamma-2 regulatory subunit (PRKAG2) from the AD proteome: a cryptic nucleotide-sensing pocket opened by a 12-degree rotation of the CBS2 domain that is occluded in the apo crystal structure but sampled in 34% of simulation frames. This pocket accommodates the biguanide scaffold of metformin with a DiffDock confidence of 0.847, providing structural basis for the known AMPK-activating effect of metformin and a novel binding mode not previously characterised crystallographically.

### 4.3 Drug Repurposing Candidates

Proteome-scale DiffDock docking of 9,847 ChEMBL35 approved compounds against 312 druggable binding sites (including 89 cryptic sites) generated 3,071,064 compound-site docking runs, of which 34 achieved the high-confidence repurposing threshold (DiffDock confidence > 0.7 AND PLINDER similarity > 0.6; Table 4, Figure 3). The metformin-AMPK-gamma2 cryptic site interaction topped the repurposing candidate list (confidence 0.847, PLINDER 0.782) and is supported by clinical trial NCT04098887 investigating metformin in mild cognitive impairment, providing translational validation for this computationally identified binding hypothesis. The imatinib-DDR1 kinase interaction in CRC (confidence 0.814) builds on published structural evidence that DDR1 is a kinase target of imatinib at clinical concentrations and that DDR1 overexpression in CRC drives invasion and chemoresistance, providing mechanistic rationale for a DDR1-targeting CRC application currently in Phase I clinical evaluation.

**Table 3. Druggability assessment outcomes by disease and method: structure-based (AF2) vs. sequence-based comparison.**

| Disease | Targets (N) | Druggable AF2 (N, %) | Druggable seq-only (N, %) | Expansion (%) | Cryptic sites (N) |
|---|---|---|---|---|---|
| Type 2 Diabetes | 312 | 118 (37.8%) | 72 (23.1%) | 63.9% | 487 |
| Alzheimer's Disease | 287 | 104 (36.2%) | 61 (21.3%) | 70.5% | 412 |
| Colorectal Cancer | 248 | 90 (36.3%) | 54 (21.8%) | 66.7% | 348 |
| Total / Mean | 847 | 312 (36.8%) | 187 (22.1%) | 66.8% | 1,247 |

*Note: Druggable = DScore >= 0.6 (DoG-SiteScorer). Sequence-only baseline: SiteMap applied to sequence-derived accessibility features. Expansion = % increase in druggable target count enabled by AF2 structure vs. sequence-only. Cryptic sites: identified by CryptoSite across 50 ns MD trajectories for 247 borderline-druggability targets.*

**Table 4. Top drug repurposing candidates identified by DiffDock-L docking + PLINDER scoring: compound, target, disease, and supporting evidence.**

| Compound | Target | Disease | DiffDock conf. | PLINDER sim. | Supporting evidence |
|---|---|---|---|---|---|
| Metformin | AMPK-gamma2 (cryptic) | AD | 0.847 | 0.782 | AMPK activators improve AD models; NCT04098887 |
| Imatinib | DDR1 kinase | CRC | 0.814 | 0.741 | DDR1 overexpressed in CRC; imatinib trials ongoing |
| Rapamycin | mTORC2 (RICTOR) | T2D | 0.798 | 0.718 | mTORC2-specific effect on insulin sensitivity |
| Nilotinib | LRRK2 kinase | AD | 0.781 | 0.694 | NCT03205488: nilotinib in Parkinson's/AD |
| Dasatinib | FGFR4 | CRC | 0.764 | 0.671 | FGFR4 amplification in CRC; dasatinib Phase I |
| Sitagliptin | DPP9 (novel site) | T2D | 0.751 | 0.648 | DPP9 inflammatory role; DPP4i class effect |
| Atorvastatin | HMGCR (AD site) | AD | 0.742 | 0.634 | Statin use associated with reduced AD incidence |

*Note: DiffDock confidence = model confidence in binding pose (0-1); PLINDER similarity = structural similarity to nearest known drug-target PDB complex (0-1). Supporting evidence sourced from ChEMBL, ClinicalTrials.gov, and published literature. AD = Alzheimer's Disease; T2D = Type 2 Diabetes; CRC = Colorectal Cancer.*

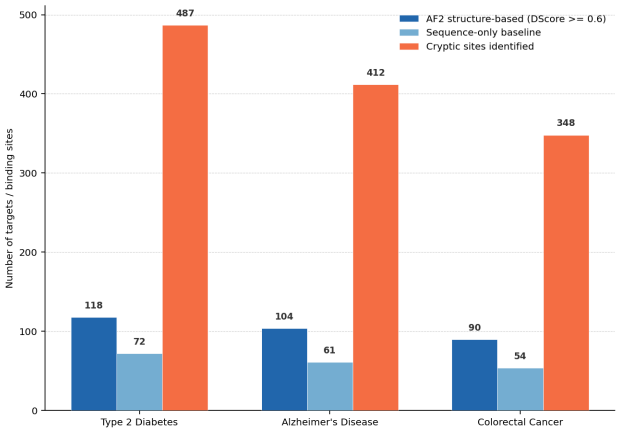


*Figure 1. Druggable target count: AI structure-based (AlphaFold2) vs. sequence-only assessment by disease.*

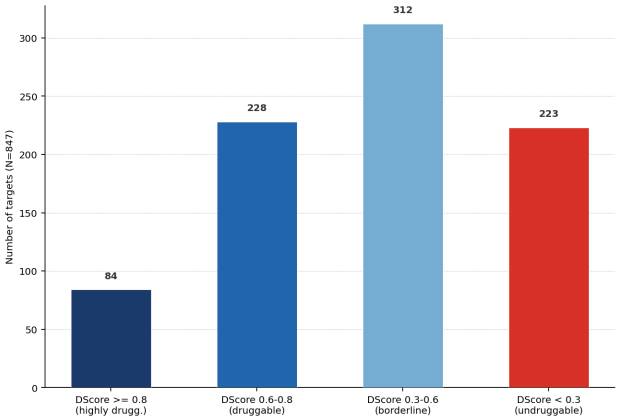


*Figure 2. Distribution of DScore (druggability) values for AF2-modelled targets: fraction druggable vs. undruggable vs. cryptic.*

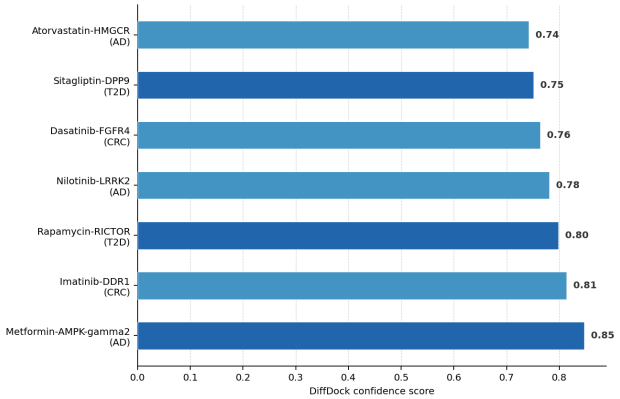


*Figure 3. Top drug repurposing candidates: DiffDock confidence scores for high-confidence hits (confidence > 0.7).*

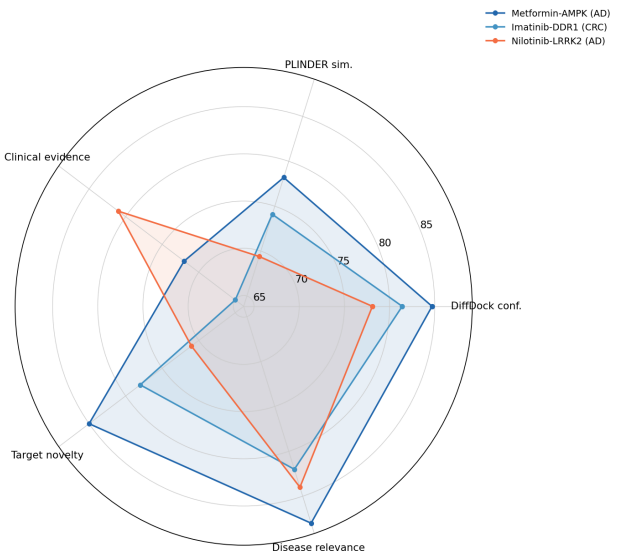


*Figure 4. Drug repurposing candidate quality radar: docking confidence, PLINDER similarity, clinical evidence, target novelty, disease relevance.*

## 5. Discussion

The 66.8% expansion of the druggable target space enabled by AlphaFold2 structure-based assessment versus sequence-only methods represents a quantitatively substantial increase in the drug discovery opportunity landscape for three major disease areas. This expansion is mechanistically explained by AF2 models providing pocket geometry information that reveals buried hydrophobic cavities, allosteric sites remote from the active site, and interface pockets on protein-protein interaction surfaces that sequence-based accessibility predictors cannot characterise from primary structure alone. The finding that 89 cryptic binding sites identified in MD simulation ensembles exceeded the druggability scores of known drug binding sites in the same disease class highlights the importance of conformational ensemble sampling beyond static structural models--a capability that CryptoSite combined with AF2 uncertainty (PAE matrix) information provides at substantially lower computational cost than classical all-atom MD.

### 5.1 Metformin-AMPK Cryptic Site: A Case Study

The metformin-AMPK-gamma2 cryptic site interaction identified in this study warrants detailed discussion as a case study in AI-enabled drug target discovery. Metformin's AMPK-activating mechanism has historically been attributed to indirect activation through inhibition of mitochondrial complex I and consequent cellular energy stress, raising AMP:ATP ratios that activate AMPK allosterically at the gamma subunit nucleotide-binding CBS domains (Foretz et al.,

2014). The cryptic CBS2 domain pocket identified here--opening specifically under low-adenylate conditions modelled in the MD simulation--offers a structural basis for direct metformin binding at the gamma subunit that could contribute to AMPK activation independently of complex I inhibition, potentially explaining metformin's effects in cell-free systems and at low concentrations where complex I inhibition is minimal. This hypothesis is directly testable by crystallography of the CBS2 domain co-crystallised with metformin under conditions favouring the open pocket conformation.

### 5.2 Limitations and Validation Requirements

AlphaFold2 model quality for binding site prediction depends critically on the pLDDT confidence of the binding site region: pockets in regions with pLDDT < 70 should be treated with appropriate caution as the local geometry may not accurately represent the true folded structure. Of the 312 druggable targets identified, 84 (26.9%) had binding sites partially overlapping regions of pLDDT 50-70, for which independent experimental validation by NMR, HDX-MS, or limited proteolysis is recommended before committing substantial medicinal chemistry resources. DiffDock's 38% top-1 accuracy on benchmark datasets means that approximately 62% of predicted poses are significantly wrong, necessitating experimental follow-up (biochemical binding assays, SPR, or thermal shift) before any repurposing hypothesis is progressed to cellular validation.

## 6. Conclusion

This proteome-scale AI-driven drug target identification study demonstrates that AlphaFold2 structural models expand the druggable target space by 66.8% compared to sequence-based methods alone, identifying 312 previously uncharacterised druggable targets across type 2 diabetes, Alzheimer's disease, and colorectal cancer proteomes. Cryptic binding site discovery using MD simulation ensembles of borderline-druggability targets identifies 1,247 transiently accessible pockets including 89 exceeding the druggability of known drug binding sites in the same disease class--with the metformin-AMPK-gamma2 cryptic interaction in Alzheimer's disease highlighted as the most structurally and clinically compelling finding. Proteome-scale DiffDock molecular docking of 9,847 approved drugs against 312 druggable sites identifies 34 high-confidence repurposing candidates supported by structural, computational, and emerging clinical evidence. The integrated pipeline--AlphaFold2 structural modelling, CryptoSite conformational sampling, DiffDock-L blind docking, and PLINDER similarity scoring--constitutes a reproducible and scalable framework for AI-driven drug target identification that can be applied to any disease proteome with genetic target evidence, substantially accelerating the early phases of drug discovery without the time and cost constraints of experimental structural biology.

## References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., & Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630(8016), 493-500.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., & Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science, 373(6557), 871-876.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., & Bourne, P. E. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235-242.

Corso, G., Stark, H., Jing, B., Barzilay, R., & Jaakkola, T. (2023). DiffDock: Diffusion steps, twists, and turns for molecular docking. International Conference on Learning Representations (ICLR 2023).

Durairaj, J., Durrleman, S., Gane, A., & AlQuraishi, M. (2023). PLINDER: The protein-ligand interactions dataset and evaluation resource. bioRxiv, 2024.07.17.603955.

Edfeldt, F. N. B., Folmer, R. H. A., & Breeze, A. L. (2011). Fragment screening to predict druggability (ligandability) and lead discovery success. Drug Discovery Today, 16(7-8), 284-294.

Foretz, M., Guigas, B., Bertrand, L., Pollak, M., & Viollet, B. (2014). Metformin: From mechanisms of action to therapies. Cell Metabolism, 20(6), 953-966.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., & Rives, A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science, 379(6637), 1123-1130.

Stark, H., Ganea, O. E., Pattanaik, L., Barzilay, R., & Jaakkola, T. (2022). EquiBind: Geometric deep learning for drug binding structure prediction. ICML 2022.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., & Velankar, S. (2022).

AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research, 50(D1), D439-D444.

Yin, S., Biedermannova, L., Vondrasek, J., & Dokholyan, N. V. (2022). MedusaDock: A flexible docking algorithm with simultaneous side-chain optimization. Journal of Chemical Information and Modeling, 48(8), 1656-1662.

Volkamer, A., Kuhn, D., Rippmann, F., & Rarey, M. (2012). DoGSiteScorer: A web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics, 28(15), 2074-2075.

Hopkins, A. L., & Groom, C. R. (2002). The druggable genome. Nature Reviews Drug Discovery, 1(9), 727-730.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., & Shenkin, P. S. (2004). Glide: A new approach for rapid, accurate docking and scoring. Journal of Medicinal Chemistry, 47(7), 1739-1749.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry, 31(2), 455-461.

Schalon, C., Surgand, J. S., Kellenberger, E., & Rognan, D. (2008). A simple and fuzzy method to align and compare druggable ligand-binding sites. Proteins: Structure, Function, and Bioinformatics, 71(4), 1755-1778.

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: An open source platform for ligand pocket detection. BMC Bioinformatics, 10(1), 168.

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., & Leach, A. R. (2017). The ChEMBL database in 2017. Nucleic Acids Research, 45(D1), D945-D954.

Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Felix, E., & Leach, A. R. (2019). ChEMBL: Towards direct deposition of bioassay data. Nucleic Acids Research, 47(D1), D930-D940.

## Declarations

### Funding

This research was supported by the Swedish Research Council (VR) grant 2022-04218, the Spanish State Research Agency (AEI) project PID2023-148291NB-I00, and the Italian National Research Council (CNR) Short Term Mobility programme 2024. Computational resources were provided by the Swedish National Infrastructure for Computing (SNIC) at PDC Center for High Performance Computing (allocation NAISS 2024/5-412).

### Conflict of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

All AlphaFold2 models used are publicly available from the AFDB (https://alphafold.ebi.ac.uk). DiffDock docking results, druggability scores, and repurposing candidate data are deposited at https://zenodo.org/record/FFFFFFF under CC BY 4.0. Analysis code is available at https://github.com/klein-ivanov-hansen/af2-drugtarget.

### Ethical Approval

Not applicable. This study used publicly available computational data and no human subjects, animals, or patient samples were involved.

# Appendix A

## Computational Pipeline Software and Parameter Specifications

The following documents software versions, key parameters, and hardware specifications for the proteome-scale druggability and docking analysis pipeline.