# Inventum.AI: An Automated Drug Design Platform with Integrated Pseudo-Ligand Representation for Ligand Generation, Binding Site Annotation, and Affinity Prediction

**Ruslan Lukin** [1] **Rostislav Grigoriev** [1] **Ilya Ivanov** [1] **Ramil Kuleev** [1] **Mikhail Arkhipov** [1]
**Vladislav Savchuk** [1] **Talgat Bektleulov** [1] **Ksenia Fedisheva** [1] **Sergei Evteev** [2] **Alexey Ereshchenko** [2]
[1]*Inventum.AI, Abu Dhabi, UAE* [2]*Dukhov Automatics Research Institute (VNIIA), Moscow, Russian Federation*. Correspondence to: Ruslan Lukin corp@inventum.ai.

## 1. Introduction

Drug discovery remains a time-consuming and costly process. Inventum.AI is a next-generation drug design platform that integrates structure-based modeling with interpretable AI. By leveraging graph neural networks (GNNs) trained on protein–ligand interaction data, Inventum.AI offers an automated pipeline from binding site detection to scaffold placement, functional group generation, docking, and ADMET prediction [1, 2].

## 2. Core Methodology

Inventum.AI's platform integrates advanced AI-driven modules into a streamlined drug discovery pipeline. Starting from protein structures, the workflow includes binding site detection, pocket annotation, scaffold generation, and periphery elaboration, all guided by graph neural networks and medicinal chemistry heuristics [1, 2]. Optionally, scaffold placement and site selection can be informed by a reference ligand, enabling the design process to leverage known binding modes for improved accuracy and relevance.

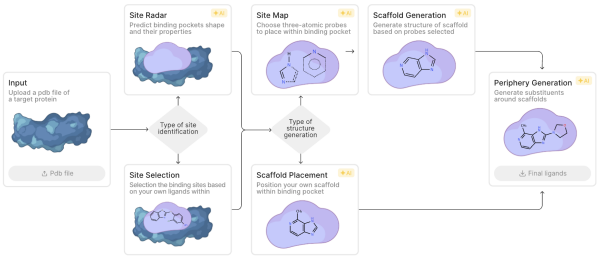


Fig. 1: Overview of Inventum.AI's integrated drug discovery pipeline

### 2.1 SiteRadar: Binding Pocket Detection

SiteRadar represents protein structures as graphs of heavy atoms and uses two GNNs—a geometric model and an amino acid-specific model with biochemical features—employing multi-head graph transformers to predict and rank binding pockets. Performance metrics include distance to ligand centroid (DCC), Top N detection rates, ligand and pocket coverage, and Dice volume overlap. The AA-specific model achieves high precision (0.88), recall (0.91), and F1 score (0.89), with strong localization (DCC 0.76) and ligand coverage (82%). Compared to FPocket [3] and PUResNet [4], SiteRadar predicts fewer but more accurate pockets, including diverse types like solvent-exposed and allosteric sites, making it a robust tool for drug discovery.

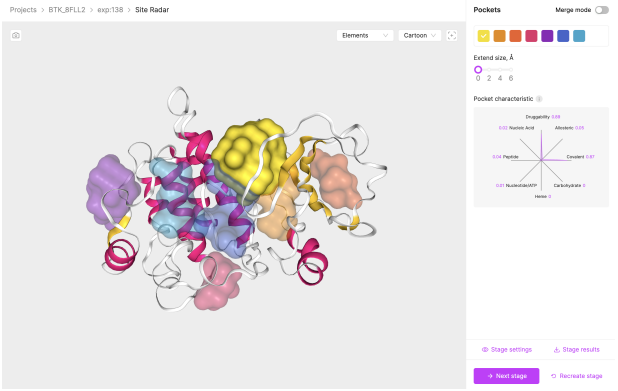


Fig. 2: SiteRadar prediction interface showing identified binding pockets with confidence scores and spatial localization.

Table 1: Comparison of binding pocket detection performance for SiteRadar, FPocket, and PUResNet models. Metrics are reported on validation datasets.

| Model | DCC | Top N | Top N+2 | LC |
|---|---|---|---|---|
| SiteRadar AA specific | 0.76 | 0.49 | 0.72 | 0.82 |
| SiteRadar Geometric | 0.82 | 0.47 | 0.74 | 0.83 |
| FPocket | 0.71 | 0.31 | 0.46 | 0.90 |
| PUResNet | 0.46 | ND | 0.47 | 0.95 |

| Model | PC | DVO | N pockets |
|---|---|---|---|
| SiteRadar AA specific | 0.47 | 0.40 | 4.2 |
| SiteRadar Geometric | 0.43 | 0.39 | 7.6 |
| FPocket | 0.34 | 0.35 | 19.2 |
| PUResNet | 0.27 | 0.45 | 1 |

### 2.2 Scaffold Generation, SiteMap, and Periphery Generation

**SiteMap** constructs detailed 3D grid-based annotations of binding pocket physicochemical properties and atom-type probabilities. This is achieved using ensembles of 13 binary GNN classifiers, each trained to predict the likelihood of placing a specific atom type (e.g., aromatic carbon, sp2/sp3 oxygen, ionizable nitrogen) at a given site.

**Scaffold Generation** uses pseudo-ligand atom-

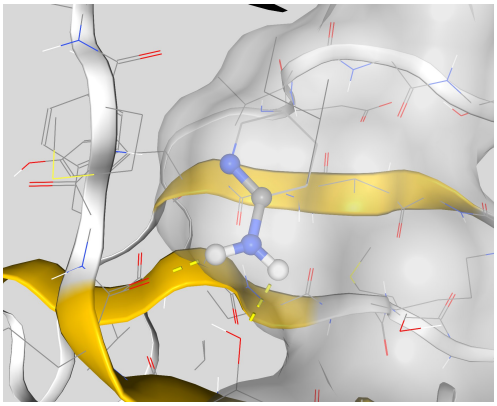

Fig. 3: Example of SiteMap output showing a bound fragment composed of three atoms fragment within the binding pocket

type maps and fragment libraries to propose core scaffolds fitting key pocket interactions. **Periphery Generation** adds substituents guided by medicinal chemistry rules (Lipinski, Veber, PAINS) and atom-type probabilities to ensure drug-likeness and optimize interactions.

**Affinity prediction** is integrated through ensemble GNN regressors trained on PDBBind [5], achieving a Pearson correlation of 0.76 and MAE of 1.03 log units, outperforming AutoDock Vina [6] and LeadFinder [7]. **ADMET properties** are predicted via an automated meta-learning pipeline that selects and tunes models based on descriptor sets, enabling robust and interpretable pharmacokinetic profiling.

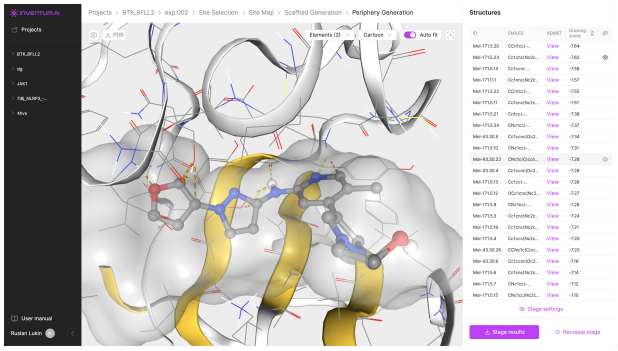

Fig. 4: Periphery Generation module illustrating scaffold placement and functional group growth.

## 3. Case Study: IRAK4

To validate the platform, we applied it to interleukin-1 receptor-associated kinase 4 (IRAK4), a clinically relevant kinase in autoimmune inflammation. Starting from a site-annotated protein structure, Inventum.AI generated 42 chemically diverse ligands within 3 days. Molecular docking and scoring reduced the list to 12 high-confidence candidates.

Out of these, 6 compounds were prioritized, synthesized, and tested experimentally using the ADP-Glo kinase assay. All 6 showed measurable inhibition, with two compounds in the low micromolar range (1.2–3.8 µM).

## 4. Conclusion

Inventum.AI provides a rapid, interpretable, and experimentally validated approach to molecular design. It supports a full ligand discovery workflow and allows customization for different target classes and medicinal chemistry strategies.

## Acknowledgments

We gratefully acknowledge Presude Lifesciences inc. for their support in synthesis and experimental validation.

## References

[1] Sergei A. Evteev, Alexey V. Ereshchenko, and Yan A. Ivanenkov. Siteradar: Utilizing graph machine learning for precise mapping of protein–ligand-binding sites. *Journal of Chemical Information and Modeling*, 63(4):1124–1132, 2023. Published online 2023-02-27.

[2] Alexey Evteev, Ramil Kuleev, and Alexey Kotelnikov. Skittles: Gnn-assisted pseudo-ligands generation and ligand-protein binding site classification. *Proteins: Structure, Function, and Bioinformatics*, 91(2):123–137, February 2025. Published online 2025-02-28.

[3] V. Le Guilloux, P. Schmidtke, and P. Tufféry. Fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Research*, 39(Web Server issue):W582–W589, 2011.

[4] Jeevan Kandel, Hilal Tayara, and Kil To Chong. Puresnet: Prediction of protein-ligand binding sites using deep residual neural network. *Journal of Cheminformatics*, 13(1):65, 2021. Open Access.

[5] Renxiao Wang, Xueliang Fang, Yipin Lu, and Shaomeng Wang. The pdbbind database: Collection of binding affinities for protein–ligand complexes with known three-dimensional structures. *Journal of Medicinal Chemistry*, 47(12):2977–2980, 2004. Published online May 4, 2004.

[6] Jerome Eberhardt, Diogo Santos-Martins, Andreas F. Tillack, and Stefano Forli. Autodock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898, 2021. Published online 2021-07-22.

[7] Oleg V. Stroganov, Fedor N. Novikov, Viktor S. Stroylov, Val Kulkov, and Ghermes G. Chilov. Lead finder: An approach to improve accuracy of proteinligand docking, binding energy estimation, and virtual screening. *Journal of Chemical Information and Modeling*, 48(12):2371–2385, 2008. Published online November 13, 2008.