Task 2: Shaking the Foundations — What Did We Actually Model?

Title: "What even IS affinity and WHY does it matter?"

Seminar Report Submitted by:

Names	Matriculation Number
MAHAM AYESHA	7068890
SALEHA ATTAR	7075857

Binding affinity quantifies how strongly a ligand and its target protein form a complex at equilibrium. For a simple reversible reaction P + L ⇌ PL, this strength is described by the equilibrium dissociation constant K_D.

Mathematically, it is define as, $K_D = \frac{[P][L]}{[PL]}$

which is directly related to the standard binding free energy by $\Delta G° = RT \ln K_D$, a lower K_D therefore corresponds to a more negative $\Delta G°$ and tighter binding [1][2]. Binding affinity results from a balance of different energetic factors. Enthalpy contributions come from interactions such as hydrogen bonds, electrostatic forces, and van der Waals or hydrophobic contacts, while entropic effects arise from solvent removal and changes in the molecule's conformational, rotational, and translational freedom [3].

Modern biophysical techniques such as Isothermal Titration Calorimetry (ITC), Surface Plasmon Resonance (SPR), and fluorescence assays are widely used to measure how strongly a ligand binds to a protein and how fast these interactions occur. Since each method highlights different thermodynamic or kinetic aspects, using them together provides a more complete and trustworthy picture of the binding process [4]**.**

Isothermal titration calorimetry (ITC) measures the small heat changes that occur when a ligand binds to a protein, and surface plasmon resonance (SPR) follows binding on a sensor surface to estimate K_D. In practice, both methods are very sensitive to buffer, temperature and how the protein is immobilized, so even minor changes in the setup can give slightly different K_D value [5]. As a result, the accuracy of a reported K_D value depends heavily on the specific assay and its experimental conditions. Using an unsuitable technique or one that lacks proper controls can introduce substantial errors, ultimately distorting the true binding affinity measurement [6].

IC_{50} (half maximal inhibitory concentration) and K_i values are often measured in assays that use different conditions, so simply pooling them can make a ligand look more or less potent than it really is. Even when two studies follow similar protocols, their reported values still show noticeable variation. Mixing these activity types into a single label adds noise to the data and makes

any predictive model based on them less reliable [7]. Because IC_{50}, K_D and K_i are strongly dependent on assay specific conditions, particularly substrate concentration, and these parameters are often missing or inconsistently reported, converting between these readouts is inherently unreliable and introduces substantial label noise. Predictive models are therefore more robust and interpretable when trained on a single, well curated type of activity measurement rather than on mixed or mathematically converted values [8].

Even if two assays are run under almost the same conditions, the reported K_D and K_i values still show some experimental scatter. This built-in variation limits how precise any model based on these measurements can ever be [9]. Binding affinity models can look very accurate on standard benchmarks because hidden biases in the training data inflate their performance. When these biases are reduced by using cleaner, strictly curated datasets, the apparent accuracy often drops markedly, as seen on the challenge set [10].

Machine learning models often over fit patterns in their training data, making them fragile when the data distribution shifts. This explains why our model worked on the noisy training set but deteriorated on the cleaner, more consistent challenge dataset. Similar behavior is reported when models are tested on new protein families or altered experimental conditions, revealing their underlying limitations [11]. A recent investigation demonstrated that decreasing similarity between the training and evaluation sets leads to a pronounced decline in predictive accuracy. This observation aligns with the notion that our challenge dataset functions as a more demanding, low similarity benchmark, revealing limitations inherited from models trained on biased data distributions [12].

Treating IC_{50}, K_i, and K_D as separate but related targets help the model capture shared chemical patterns while respecting assay specific differences. A multitask framework can combine diverse bioassay data without collapsing everything into one noisy label. This reduces bias and improves robustness, especially when the model is tested on cleaner, more reliable datasets [13]. Including assay level metadata such as temperature, buffer conditions, and protein variants helps the model understand why activity values differ across experiments and adjust its predictions accordingly. This produces more stable outputs and lowers the risk of overfitting to noise in heterogeneous bioassay data [14].

Diversity based or scaffold aware splits ensure that test compounds differ structurally from the training set. This gives a more realistic measure of how well a model can generalize to new chemical space, unlike overly easy random splits. Such challenging partitions help reveal whether the model learns genuine structure activity relationships instead of memorizing training patterns [15]. Model performance drops when bioactivity values from different assays are mixed without explaining how each measurement was obtained. Adding assay aware descriptors such as assay type, detection method, or protein construct stabilizes predictions and improves generalization from noisy training data to cleaner external benchmarks [16].

REFERENCES:

[1] I. Jarmoskaite, I. Alsadhan, P. P. Vaidyanathan, S. Chem-h, and U. States, "How to measure and evaluate binding affinities," 2020.

[2] M. Thafar, A. Bin Raies, S. Albaradei, and M. Essack, "Comparison Study of Computational Prediction Tools for Drug-Target Binding Affinities," vol. 7, no. November, pp. 1–19, 2019, doi: 10.3389/fchem.2019.00782.

[3] H. Gohlke and G. Klebe, "Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular ReceptorsNo Title," *Angew. Chemie Int. Ed.*, vol. 41, no. 15, pp. 2644–2676, 2002, doi: 10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O.

[4] X. Du *et al.*, "Insights into Protein – Ligand Interactions : Mechanisms , Models , and Methods," no. i, pp. 1–34, 2016, doi: 10.3390/ijms17020144.

[5] D. A. Mlsna and A. L. Patrick, "Biophysical characterization of traditional and nontraditional equilibria," vol. 1, 2020.

[6] M. Jing and M. T. Bowser, "Analytica Chimica Acta Methods for measuring aptamer-protein equilibria : A review," vol. 686, p. 2011, 2011, doi: 10.1016/j.aca.2010.10.032.

[7] G. A. Landrum and S. Riniker, "Combining IC 50 or," 2024, doi: 10.1021/acs.jcim.4c00049.

[8] N. Schapin, M. Majewski, A. Varela-rial, C. Arroniz, and G. De Fabritiis, "Machine learning small molecule properties in drug discovery," *Artif. Intell. Chem.*, vol. 1, no. 2, p. 100020, 2023, doi: 10.1016/j.aichem.2023.100020.

[9] C. A. Hern, "Artificial Intelligence in the Life Sciences Experimental Uncertainty in Training Data for Protein-Ligand Binding Affinity Prediction Models," vol. 4, no. September, 2023, doi: 10.1016/j.ailsci.2023.100087.

[10] D. Graber, P. Stockinger, F. Meyer, S. Mishra, C. Horn, and R. Buller, "Resolving data bias improves generalization in binding affinity prediction," *Nat. Mach. Intell.*, vol. 7, no. October, 2025, doi: 10.1038/s42256-025-01124-5.

[11] A. Chatterjee *et al.*, "Improving the generalizability of protein- ligand binding predictions with AI-Bind," no. May 2022, 2023, doi: 10.1038/s41467-023-37572-z.

[12] J. Buhmann, A. Bilsland, and H. Triendl, "Improving generalisability of 3D binding affinity models in low data regimes," no. Ml, 2024.

[13] J. Yan *et al.*, "Multi-task bioassay pre-training for protein-ligand," vol. 25, no. 1, pp. 1–17, 2024.

[14] M. Rayka, M. Mirzaei, and A. M. Latifi, "An ensemble-based approach to estimate confidence of predicted protein – ligand binding affinity values," no. February, pp. 1–16, 2024, doi: 10.1002/minf.202300292.

[15] X. Ouyang, Y. Feng, C. Cui, Y. Li, L. Zhang, and H. Wang, "Data and text mining Improving generalizability of drug – target binding prediction by pre-trained multi-view molecular," vol. 41, no. June 2024, pp. 1–9, 2025.

[16] L. Schoenmaker, E. G. Sastrokarijo, L. H. Heitman, J. B. Beltman, W. Jespers, and G. J. P. Van Westen, "Toward Assay-Aware Bioactivity Model(er)s: Getting a Grip on Biological Context," 2025, doi: 10.1021/acs.jcim.5c00603.