Republic of the Philippines
Department of Education
REGION III
SCHOOLS DIVISION OF CITY OF SAN JOSE DEL MONTE
CITY OF SAN JOSE DEL MONTE NATIONAL SCIENCE HIGH SCHOOL
ECOPARK, MUZON, CSJDM, BULACAN

Advancing in Silico Drug Discovery Through An Enzyme Inhibition Efficiency and Entropy-based Selectivity Composite Index

Researchers:

Bartolome, Hallyneehra G.

Edic, Matthew Cedric M.

Marcaida, Dennona Ambher A.

Yola, Pauline M.

Research Adviser:

Agujetas, Bunny M.

Submitted on:

December 26, 2025

TABLE OF CONTENTS

CHAPTER I

Background of the Study

The continuous emergence of drug resistance, unwanted drug side effects, and breakthroughs across different diseases increasingly demands the development of novel drugs and biological targets. In the pre-clinical phase of drug discovery, candidate compounds require validation through *in vitro* (in a test tube), *in vivo* (in a living organism), and *in silico* (computer-based simulation) approaches—each contributing insights into drug candidacy (Singh et al., 2023).

Among these, in silico approaches frequently employ molecular docking, a computer-assisted tool for drug design that simulates the predicted mechanism of action of a molecule in relation to a biological target, such as proteins (Fan et al., 2018). Pre-clinical in silico methods have become indispensable due to their efficiency and cost-effectiveness, enabling the rapid analysis of large volumes of molecular data and thereby narrowing the drug discovery process before proceeding to experimental validation. This substantial data throughput, however, requires medicinal chemists to invest considerable time in interpreting docking outputs and disentangling potential false-positive predictions (Fan et al., 2018; Uitdehaag & Zaman, 2011).

The tedious nature of raw data analysis, coupled with the oversimplification and insufficiency of docking scores when considered in isolation, highlights the need for quantitative measures that can convert docking results across multiple targets into

more meaningful and conclusive predictions for complex biological systems. For instance, effective inhibition across multiple enzyme targets for candidate compounds or drug multitargetedness is critical for drug discovery (Macalalad & Gonzales, 2023), and such information can potentially be deduced from raw molecular docking data.

Indices are purpose-built for efficiently handling bulk data and gauging parameters holistically. In various fields of research, indices have helped condense information into a single, yet multidimensional value. The Human Development Index (HDI) is a prime example, providing an index for national development beyond the Gross Domestic Product (GDP) (Herre & Arriagada, 2023). HDI is calculated using four indicators, which are normalized and aggregated to values between 0 and 1. Water quality indices (WQIs) play a similar role, effectively unifying the multidimensional parameters of water quality. Chidiac et al. (2022) outline the methodical process involved in developing WQIs historically, which includes selecting parameters, transforming raw data, assigning weights to each parameter, and combining the parametric values into an index.

In the goal of deriving significant meaning from molecular docking scores using indices, the usage of entropy scores is a field of interest. Beyond its thermodynamic definition, entropy can be defined as a conceptual measure of uncertainty or complexity (Ribeiro et al., 2021). In bioinformatics, mathematical models involving entropy have been employed to decode genomic areas and sequences based on their level of complexity, among other applications in

experimental molecular biology (Mageed, 2024). This conceptual, statistical, and non-physical use of entropy is specifically employed in Shannon entropy, which quantifies information as the amount of unexpected or, more specifically, unintended data, as coined by Ribeiro et al. Similarly, Mageed defines Shannon entropy as a mathematical model that measures information based on its different configurations or assortments.

Uitdehaag and Zaman (2011) applied this entropy-based framework to develop a theoretical score for enzyme inhibitor selectivity using experimental inhibition data, specifically half-maximal inhibitory concentration (IC_{50}) values. In contrast, the present study seeks to quantify inhibitor selectivity using molecular docking scores. Existing studies rarely explored a viable entropy-based selectivity score, with the most significant one by Uitdehaag and Zaman dating back to 2011, which utilized experimental IC_{50} values instead of in silico molecular docking scores. Furthermore, Grossman & Adler call for the development of efficient compound screening methods to determine inhibitor selectivity.

Handling the bulk of information provided by docking-derived binding affinity estimates across multiple protein targets and addressing the acknowledged gaps, the researchers propose the Inhibition Efficiency Index (IEI), which normalizes and aggregates the docking scores into a single value. The IEI will be coupled with the Selectivity Entropy-based Index (SEI), which will be the value that measures inhibitor selectivity. Considering inhibitor selectivity is crucial in drug discovery and development, as non-selective inhibitors or compounds that bind to proteins other

than the intended target can cause undesired effects, leading to the discontinuation of a drug (Grossman & Adler, 2021). As explained by Grossman & Adler, non-selective inhibitors can still be beneficial when they can target proteins from the same pathway and cause a synergistic effect. Therefore, it is of considerable use to determine inhibitor selectivity from molecular docking scores.

Overall, molecular docking is a refined tool for the pre-clinical phase of drug discovery; however, considering the mentioned limitations and false positive-proneness of docking scores (Paggi et al., 2024), the current study aims to enhance the interpretation of raw docking scores using entropy-inspired composite indices.

Statement of the Problem

This study aims to create a computational framework using the proposed Inhibition Efficiency Index (IEI) and Selectivity Entropy Index (SEI) to derive meaningful information from raw molecular docking scores. Hence, the study seeks to address the following objectives:

1. What are the IEI values of the selected compounds across multiple protein targets?
2. What are the SEI values of the selected compounds when Shannon entropy is applied to the docking-derived binding affinity scores?
3. How different is the ranking of compounds based on the IEI values from the initial raw molecular docking scores?

4. How different is the ranking of compounds based on the SEI values from the initial raw molecular docking scores?
5. What is the existing relationship between IEI and SEI values?

Hypothesis of the Study

This study seeks to accept or reject the following hypotheses:

1. **Ho:** There is no significant difference between the ranking of compounds based on the IEI and the one based on the raw molecular docking score.

   Ha: There is a significant difference between the ranking of compounds based on the IEI and the one based on the raw molecular docking score.

2. **Ho:** There is no significant difference between the ranking of compounds based on the SEI and the one based on the raw molecular docking score.

   Ha: There is a significant difference between the ranking of compounds based on the SEI and the one based on the raw molecular docking score.

3. **Ho:** There is no significant relationship between IEI and SEI.

   Ha: There is a significant relationship between IEI and SEI.

Significance of the Study

The following stakeholders stand to benefit from this study:

Scientists and Researchers in Drug Discovery. Scientists can leverage IEI and SEI for more accurate virtual screening and lead optimization, enabling mechanistic insights into inhibitor selectivity before experimental validations and reducing the need for retrospective verification.

Pharmacologists and Medicinal Chemists. These professionals gain tools to quantify multi-target interactions, aiding in the design of selective inhibitors that mitigate side effects while preserving synergistic benefits in pathways like oncology or metabolic disorders.

Pharmacists and Healthcare Providers. Pharmacists benefit from the development of drugs with improved safety profiles, leading to fewer adverse reactions and better patient outcomes. The framework's focus on selectivity can inform prescribing practices, particularly for polypharmacology drugs, reducing discontinuation rates due to off-target effects.

Pharmaceutical Industry Professionals. Industry stakeholders, including biotech firms, can streamline R&D pipelines by integrating IEI and SEI into high-throughput screening, shortening timelines and lowering costs associated with failed candidates. AI-driven entropy methods have been shown to revolutionize drug discovery by increasing success rates (Chen et al., 2023).

Bioinformaticians and Computational Biologists. These experts can apply the entropy-inspired indices to bioinformatics applications, such as genomic sequence analysis or protein-ligand modeling, fostering interdisciplinary tools for complex data interpretation (Mageed, 2024).

Regulatory Agencies and Policymakers. Regulators gain quantitative metrics for evaluating drug selectivity during approvals, promoting evidence-based policies that prioritize patient safety and innovation in computational pharmacology.

Future Researchers. This study provides a foundational framework for subsequent investigations, such as extending IEI/SEI to machine learning models or real-time docking platforms. Future work could explore integrations with advanced simulations, building on entropy's role in thermodynamics and information theory to advance fields like personalized medicine.

Definition of Terms

Inhibition Efficiency Index (IEI). A composite index proposed to summarize the overall inhibitory potential of a compound across multiple protein targets. It is calculated by normalizing and combining molecular docking scores (binding free energies expressed in kcal/mol) into a single, dimensionless value ranging from 0 to 1. Higher IEI values indicate stronger predicted binding and greater overall inhibition efficiency toward the intended targets.

Selectivity Entropy Index (SEI). An entropy-based measure adapted from Shannon entropy and the selectivity entropy framework introduced by Uitdehaag and Zaman (2011). In this study, SEI is applied to docking-derived binding affinities to evaluate how selectively a compound binds across a panel of protein targets. Lower SEI values indicate high selectivity (binding concentrated on fewer targets), while higher SEI values indicate non-selective or promiscuous binding.

Molecular Docking. A computational technique used to simulate the interaction between a small molecule (ligand) and a protein target. It predicts the most favorable binding pose and estimates the binding free energy (ΔG, in kcal/mol).

Molecular Docking Score. The estimated binding free energy produced by a docking algorithm. It serves as an indirect indicator of inhibitory strength and is often used as a proxy for experimental measures such as IC_{50}, although its accuracy is limited by simplifications in scoring functions and models.

Shannon Entropy. A concept from information theory that measures the degree of uncertainty or diversity within a distribution.

Inhibitor Selectivity. The extent to which a compound preferentially binds to its intended target rather than to off-target proteins. High selectivity is generally

desirable to reduce side effects, although controlled multi-target activity can be beneficial in certain therapeutic contexts.

Composite Index. A single numerical measure created by combining multiple normalized variables to provide an overall assessment. In this study, IEI and SEI serve as composite indices for interpreting complex docking data.

Binding Affinity. The strength of interaction between a ligand and a protein. In computational studies, it is estimated as a binding free energy, while experimental equivalents include dissociation constants (Kd) or half-maximal inhibitory concentrations (IC_{50}).

In Silico. Computational methods used in drug discovery and molecular analysis, including molecular docking and virtual screening, without performing physical experiments.

Theoretical Framework

This study is based on the selectivity entropy framework developed by Uitdehaag and Zaman (2011), which provides a quantitative and theory-based method for measuring inhibitor selectivity using profiling data such as IC_{50} values. The framework treats inhibitor binding across multiple protein targets as a probabilistic distribution, where selectivity is inversely related to entropy. A low entropy value

indicates that binding is concentrated on a small number of targets, while a high entropy value reflects broad, non-selective binding.

Mathematically, Uitdehaag and Zaman derived the selectivity entropy score from experimental enzymatic association constants (K_a), which can be approximated as the inverse of Kd or IC_{50}. The fraction of inhibitor bound to each target i is defined as $\varphi_i = K_{a,i} / \Sigma K_a$, and selectivity entropy is calculated as $S = -\Sigma \varphi_i \ln \varphi_i$. Compared with alternative selectivity metrics—such as Gini scores or threshold-based selectivity measures—this entropy-based approach is non-arbitrary, less sensitive to experimental noise, and independent of reference targets, allowing consistent comparison across datasets.

Previous applications of this framework include kinase and nuclear receptor profiling, high-throughput screening data analysis, and evaluation of clinical drug candidates. These studies demonstrate that high selectivity is not always associated with greater clinical success, as several effective drugs exhibit controlled multi-target activity rather than strict single-target specificity.

By adapting the selectivity entropy concept to molecular docking data, this study extends the framework to in silico screening. The proposed IEI summarizes overall predicted inhibitory strength, while SEI quantifies selectivity based on docking-derived affinities. Together, these indices provide a clearer and more

interpretable way to evaluate both potency and selectivity in virtual screening, while acknowledging the limitations of docking scores compared with experimental data.

CHAPTER II

Review of Related Literature

Enzyme Inhibition in Biological and Pathophysiologic Pathways

mention how enzyme inhibition works and why it's important. you can use RRL from our old paper for this.

you can mention the role of enzymes in different illnesses like cancer, diabetes, etc. and how research focuses on compounds that inhibit them

Molecular Docking

everything about molecular docking and how it works (MUST BE COMPREHENSIVE), including its limitations and inaccuracy and how we're improving upon it.

https://link.springer.com/article/10.1007/s40484-019-0172-y (Fan et al., 2019)

https://www.annualreviews.org/content/journals/10.1146/annurev-biochem-030222-120000 (Paggi et al., 2024)

Macalalad, M. A. B., & Gonzales, A. A., 3rd. (2023). In silico screening and identification of antidiabetic inhibitors sourced from phytochemicals of Philippine plants against four protein targets of diabetes (PTP1B, DPP-4, SGLT-2, and FBPase). *Molecules, 28*(14), 5301. https://doi.org/10.3390/molecules28145301

Aziz, M.; Ahmad, S.; Khurshid, U.; Pervaiz, I.; Lodhi, A.H.; Jan, N.; Khurshid, S.; Arshad, M.A.; Ibrahim, M.M.; Mersal, G.A.M.; et al. Comprehensive Biological Potential, Phytochemical Profiling Using GC-MS and LC-ESI-MS, and In-Silico Assessment of Strobilanthes glutinosus Nees: An Important Medicinal Plant. Molecules 2022, 27, 6885. https:// doi.org/10.3390/molecules27206885

Composite Indices

other indices, e.g. water quality, air quality, human development index, CPI, etc.

dis one ez i think, just some examples

also, explain the process by which composite indices are created. how they begin with normalization and such

Statistical Uses of Entropy

how entropy has been used in research and why we're using it for our

EXPLAIN:

- entropy as a thermodynamic concept and the 2nd law of thermodynamics
- The Entropy Universe and how entropy is used in Information Theory
- Shannon Entropy, its development, its usage, the math involved
- How entropy-based values are used in biology, specifically bioinformatics
- entropy-based indices

can refer to these papers: (FIND MUCH MORE lol)

Ribeiro, M.; Henriques, T.; Castro, L.; Souto, A.; Antunes, L.; Costa-Santos, C.; Teixeira, A. The Entropy Universe. Entropy 2021, 23, 222. https://doi.org/10.3390/e23020222

Mageed, I. A. (2024). Entropic imprints on bioinformatics. Big data and computing visions, 4(4), 245-256.

doi:10.1186/1471-2105-12-94 Cite this article as: Uitdehaag and Zaman: A theoretical entropy score as a single value to express inhibitor selectivity. BMC Bioinformatics 2011 12:94.

CHAPTER III

Methodology

Conceptual Framework

Figure 1. *Conceptual Framework*

Materials/Tools

- RCSB Protein Data Bank
- PubChem
- PyMol
- AutoDock Vina
- AutoDockTools
- Python
- Command Prompt
- Microsoft Excel
- Personal computer

Detailed Procedure

Selection and Preparation of Protein Targets

Biologically relevant and literature-validated protein targets were selected to function as experimental representatives for the proposed IEI & SEI computational framework. Four protein targets were adapted from Macalalad & Gonzales (2023), who screened in silico potential antidiabetic inhibitors for the proteins **protein tyrosine phosphatase 1B (PTP1B)**, **dipeptidyl peptidase-4 (DPP-4)**, **sodium-glucose cotransporter 2 (SGLT-2)**, and **fructose-1,6-bisphosphatase (FBPase)**. Additional selected protein targets are **glucose transporter 2 (GLUT2)** and **α-amylase**, as validated by Medeiros et al. (2024). The effects of selectivity and nonselectivity (promiscuity) are variably sought depending on the related pathways of target enzymes (Grossman & Adler, 2021). Hence, the chosen proteins are diverse to maximize selectivity, but a multi-targeted effect—potentially quantified by IEI—remains ideal in antidiabetic drugs (Macalalad & Gonzales, 2023).

The six selected protein targets will be acquired from the Protein Data Bank (PDB) and will be prepared by adding polar hydrogens and removing superfluous prosthetic groups.

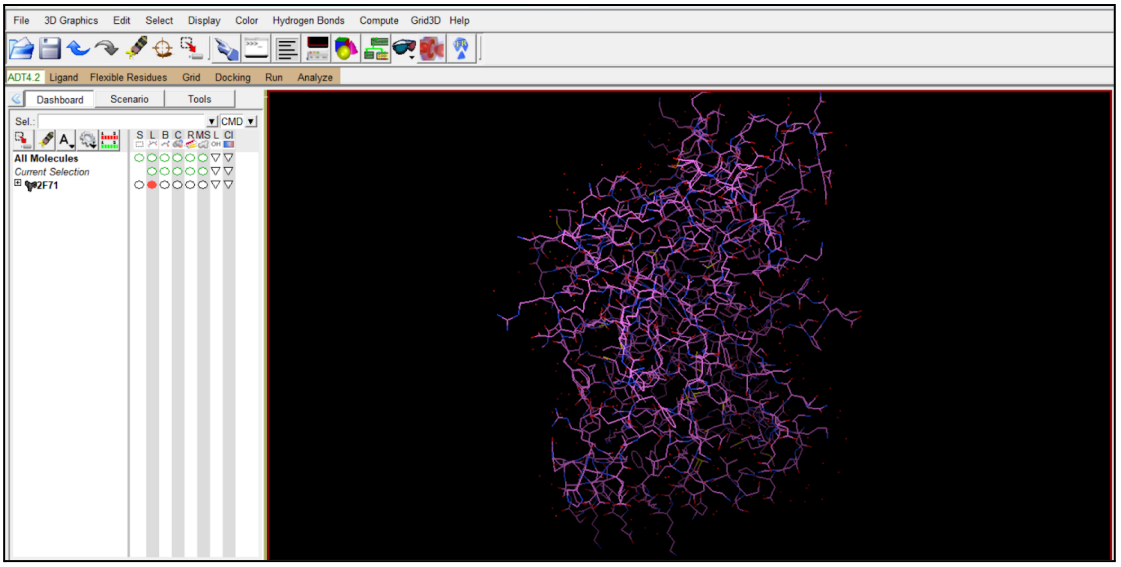

Figure 1. Optimization and preparation of protein target in .pdb file, later saved as .pdbqt for molecular docking

Selection and Preparation of Ligands

The ligands will be selected based on the final phytochemical shortlist from Macalalad & Gonzales (2023), which already have established inhibitory or antagonistic activity on four of the six protein targets. 20 ligands will be selected and subjected to necessary geometric and molecular optimizations compatible with the chosen docking software.

Molecular Docking Simulation

The optimized ligand and protein structures saved as PDBQT files will be subjected to molecular docking simulations using PyRx. The best generated docking

pose was selected based on the lowest predicted binding energy, which would serve as the raw docking results.

The processing and organization of the results were automated using Python scripts, wherein the binding affinity for each ligand-protein pair was organized into a structure matrix.

Docking Score Sign Inversion and Normalization

The raw docking scores will be subjected to the absolute value function so that higher numerical values correspond to stronger predicted binding affinity. The scores will then be normalized to enable aggregation to proportionate indices.

Computation of IEI and SEI

The IEI is purpose-built to reflect the overall magnitude of the binding affinity of a ligand across multiple protein targets. It is complementary to SEI, which will quantify the ligands' selectivity. The normalized docking scores are aggregated into a single composite efficiency index by adding their normalized docking scores on the target proteins and dividing by the number of proteins. This is the computed IEI.

The SEI will be calculated by first converting the normalized docking scores for each ligand to a probability distribution. The mathematical formula for Shannon entropy will be applied to the distribution for the final index value. Lower SEI values

correspond to a lessened assortment of information, equivalent to fewer targets, which suggests higher selectivity. Higher SEI values indicate a broader distribution of binding affinities, implying promiscuity.

Relationship and Pattern Analysis

Python was further used to generate visualizations and perform statistical analyses. The ranking of the compounds based on their raw docking scores and IEI and SEI values will be compared using statistical analyses. The existence of a relationship between IEI and SEI in its ability to distinguish selective, moderately selective, and promiscuous compounds was also evaluated. The distinguishing index value for the compound classifications will be determined using quartile-based partitioning. Furthermore, to support the interpretation of the results, scatter plots were generated to observe trends between IEI and SEI. Rank comparison plots were also created to visualize the shifting of ranks between rankings based on raw docking scores and those based on index values.

Gathering of Data

The following are the computational processes the current study aims to employ:

Let:

$A_{ij} = \ predicted\ binding\ affinity\ (kcal/mol)\ of\ compound\ i\ against\ protein\ target\ j$

$i\ =\ 1,\ 2,\ 3,\ \ldots,\ M\ compounds$

$j\ =\ 1,\ 2,\ 3,\ \ldots,\ N\ protein\ targets$

Docking Score Sign Inversion

$$S_{ij} = -A_{ij}$$

Where:

$S_{ij} = sign - inverted\ docking\ score$

This is conducted to reflect the intended magnitude of higher binding affinity corresponding to a higher numerical value.

Normalization

Min-max normalization was applied to the inverted docking scores to produce proportionality in the distribution of the predicted binding affinity of a compound on varying proteins.

$$N_{ij} = \frac{S_{ij} - min_i(S_j)}{max_i(S_j) - min_i(S_j)}$$

Where:

$N_{ij} \in [0, 1]$

$max_i(S_j) = highest\ docking\ score\ of\ a\ compound\ i\ for\ protein\ j$

$min_i(S_j) = lowest\ docking\ score\ for\ a\ compound\ i\ protein\ j$

Inhibition Efficiency Index (IEI)

The study proposes the following weighted average formula for the aggregation of multidimensional binding affinity scores across different protein targets:

$$IEI_i = \frac{1}{N} \sum_{j=1}^{N} N_{ij}$$

Where:
$N = no.\ of\ protein\ targets$

Higher IEIs correspond to stronger predicted binding affinity.

Selectivity Entropy Index (SEI)

To compute selectivity, the normalized scores are first converted to a probability distribution.

$$p_{ij} = \frac{N_{ij}}{\sum_{k=1}^{M} N_{ik}}$$

Where:

$p_{ij} = probability\ inhibitor\ i\ is\ bound\ to\ a\ specific\ protein\ target\ j$

$k = all\ proteins$

$N_{ik} = normalized\ docking\ score\ for\ a\ compound\ i\ across\ all\ proteins$

Shannon entropy is then employed, which will be the SEI.

$$SEI_i = -\sum_{j=1}^{M} p_{ij} ln(p_{ij})$$

In the case of a probability of 0, the convention $0\ ln\ 0 = 0$, which would otherwise be undefined, was followed.

Methods and Techniques

Research Instrument

Data Analysis

BIBLIOGRAPHY

Herre, B., & Arriagada, P. (2023). The Human Development Index and related indices: What they are and what we can learn from them. Our World in Data. https://archive.ourworldindata.org/20251220-152415/human-development-index.html

Uitdehaag, J. C., & Zaman, G. J. (2011). A theoretical entropy score as a single value to express inhibitor selectivity. BMC Bioinformatics, 12, Article 94. https://doi.org/10.1186/1471-2105-12-94

Chidiac, S., El Najjar, P., Ouaini, N., El Rayess, Y., & El Azzi, D. (2023). A comprehensive review of water quality indices (WQIs): history, models, attempts and perspectives. Re/views in environmental science and bio/technology, 22(2), 349–395. https://doi.org/10.1007/s11157-023-09650-7

Fan, J., Fu, A., & Zhang, L. (2019). Progress in molecular docking. Quantitative Biology, 7, 83–89. https://doi.org/10.1007/s40484-019-0172-y

Paggi, J. M., Pandit, A., & Dror, R. O. (2024). The art and science of molecular docking. Annual Review of Biochemistry, 93, 389–410. https://doi.org/10.1146/annurev-biochem-030222-120000

Singh N., Vayer, P., Tanwar, S., Poyet, J.-L., Tsaioun, K., & Villoutreix, B. O. (2023). Drug discovery and development: Introduction to the general public and patient groups. Frontiers in Drug Discovery, 3, Article 1201419. https://doi.org/10.3389/fddsv.2023.1201419

Grossman, M., & Adler, E. (2021). Protein Kinase Inhibitors - Selectivity or Toxicity? IntechOpen. doi: 10.5772/intechopen.98640

Mageed, I. A. (2024). Entropic imprints on bioinformatics. Big data and computing visions, 4(4), 245-256.

Ribeiro, M.; Henriques, T.; Castro, L.; Souto, A.; Antunes, L.; Costa-Santos, C.; Teixeira, A. The Entropy Universe. Entropy 2021, 23, 222. https://doi.org/10.3390/e23020222