Social Choice Theory in Protein Folding

Raisa Axenie
Advisor: Sorin Istrail
Reader: Fernando Duarte

Abstract

We study the problem of estimating universal energy functions for protein folding under the thermodynamic and pairwise interaction hypotheses. We propose a novel perspective by modeling the aggregation of interaction preferences across proteins as a social choice problem. By mapping proteins to voters and interaction classes to alternatives, we show that the estimation of global energy preferences is subject to impossibility results analogous to Arrow's theorem. Our results suggest that fundamental theoretical limitations, rather than purely computational challenges, underlie the difficulty of constructing universal pairwise energy functions.

1 Introduction

Protein folding is a fundamental problem in computational biology, concerned with understanding how a protein's amino acid sequence determines its three-dimensional structure. The widely accepted thermodynamic hypothesis asserts that the native structure of a protein corresponds to a unique, stable configuration that minimizes a global energy function [1].

A central challenge in this domain is determining or estimating this energy function. A common approach assumes that protein stability can be explained through pairwise interactions between amino acids, leading to the formulation of pairwise energy functions [6]. These models aim to assign energy potentials to interaction classes such that the observed folded structures correspond to energy minima.

In this work, we investigate a fundamental limitation of this approach. Specifically, we examine whether it is possible to infer a universal pairwise energy function by aggregating interaction information across multiple proteins. We propose a novel perspective by framing this problem as one of preference aggregation, drawing a connection to Social Choice Theory [2].

By modeling proteins as voters and interaction classes as alternatives, we show that the aggregation process is subject to impossibility results analogous to Arrow's Impossibility Theorem. Our results demonstrate that any method attempting to construct a global ordering of interaction classes from individual protein data must violate at least one desirable property.

2 Motivation

The ability to accurately model protein folding has significant implications for biology, medicine, and drug discovery [4]. Central to this effort is the construction of energy functions that explain and predict protein structure.

Pairwise interaction models are particularly appealing due to their computational tractability and interpretability. By assigning energy values to interactions between amino acid pairs, these models aim to capture the essential forces governing protein folding.

However, estimating these energy values requires aggregating information across many proteins. Each protein provides evidence about the relative importance of different interaction types, typically in the form of observed contact frequencies or derived scores.

This aggregation problem is structurally similar to voting: each protein expresses a preference over interaction classes, and the goal is to derive a consensus ranking. This analogy motivates the application of Social Choice Theory [8].

3 Related Work

3.1 Protein Folding and Energy Functions

The systematic understanding of how and why proteins settle into specific three-dimensional structures is the principal concern of protein folding research. The foundation of this understanding relies heavily on the well-accepted thermodynamic hypothesis proposed by Anfinsen [1]. This hypothesis states that a protein's tertiary structure is a unique, stable, and viable state that depends solely on its amino acid sequence. The state minimizes an energy function that accounts for physiochemical utilities and is uniformly defined across all proteins based entirely on amino acid interactions.

Because obtaining all collective, grouped interactions is computationally complex, the literature frequently adopts pairwise interaction models [6]. These pairwise energy functions are defined entirely by energy potentials on pairwise interactions among amino acids. To account for physiochemical properties like hydrophobia and hydrophilia, research often relies on compiled data tables, such as the Miyazawa-Jernigan scores.

3.2 Aggregation and Statistical Methods

Many methods attempt to estimate energy functions by aggregating data from multiple proteins [4]. These approaches typically rely on statistical inference or optimization techniques.

3.3 Social Choice Theory

Social Choice Theory is primarily concerned with the systematic aggregation of individual preferences into a global social preference. The origins of this field trace back to the 13th-century philosopher Ramon Lull, who considered a voting method that counted pairwise preferences among candidates. This technique was later recovered in the 18th century by mathematicians Borda and Condorcet.

A definitive landmark in this domain is Kenneth Arrow's 1951 impossibility theorem [2]. Arrow's theorem analyzes seemingly innocuous properties that preference aggregation functions (social welfare functions) should respect, ultimately demonstrating that these properties inherently conflict. Additional theoretical extensions, such as those expanded upon by Sen [8], establish the foundations of collective decision-making. May's Theorem dictates that any positively responsive voting mechanism respecting unanimity and neutrality must rely on simple majority voting for pairwise comparisons. Furthermore, the Gibbard-Satterthwaite theorem explores manipulation in voting procedures, proving that if a non-dictatorial procedure transitively orders choices, one alternative will necessarily never be strictly preferred.

4 Model and Method

4.1 Thermodynamic Hypothesis

We adopt the thermodynamic hypothesis [1]. Under this hypothesis, the tertiary structure of a protein is unique, stable, and determined by energy minimization.

4.2 Pairwise Interaction Hypothesis

We assume that amino acid interactions can be represented as pairwise interactions [6]. This simplifies the energy function and makes estimation computationally feasible.

4.3 Interaction Classes and Preferences

An interaction class is formally defined as a tuple (X_1, X_2), where X_1 and X_2 represent names of amino acids. Because (X_1, X_2) is considered equal to (X_2, X_1), and there are 20 standard amino acids, the model accounts for 200 possible interaction classes. An interaction instance of a specific class occurs within a protein if the distance between two amino acids falls within a specified threshold τ.

According to the pairwise and thermodynamic hypotheses, energetic scores over these interaction instances can be estimated to derive utility values for interaction classes within each protein. Ultimately, determining the assumed pairwise universal energy function requires estimating the energy potentials for each interaction class.

Rather than estimating exact energy values directly, we consider ordinal preferences over interaction classes. Each protein induces a ranking over interaction types based on observed interactions.

4.4 Aggregation Framework

Instead of attempting to directly generate exact utility values for all interaction classes across all proteins, we address a strictly weaker problem: generating an ordinal preference among interaction classes. Let $P = \{P_1, \ldots, P_n\}$ be proteins and E interaction classes. Each protein defines a total order $I(P_i)$ over E, and the goal is to compute a global ordering O.

The estimation process can be divided into two phases:

- **Internal Aggregation:** Contact information inside proteins is examined to derive individual preferences, taking the form of either utility scores or ordinal preferences among interaction classes.
- **External Aggregation:** Individual preferences across proteins are combined to derive a global, ordinal preference over interaction classes.

Modeled after social welfare functions, the external aggregation procedure receives total orders on elements and produces a total order. In this work, we assume that internal aggregation is correctly performed and focus on the external aggregation problem.

4.5 Axiomatic Requirements

To ensure a rigorous and logical external aggregation process, the procedure must theoretically adhere to several fundamental social choice axioms [2, 8]:

- **Agreement:** The procedure must result in a single preference among interaction classes.
- **Transitivity & Totality:** Both individual and global preferences must be transitive and defined for any pairwise comparison.
- **Unanimity:** If all proteins expose a preference for a particular interaction class over another, that unanimity must be respected globally.
- **Neutrality:** All pairwise preferences over interaction classes are treated equally, meaning the procedure is agnostic to the specific classes of interactions.

- **Independence of Irrelevant Alternatives (IIA):** For any two voting scenarios where individual preferences over arbitrary elements remain identical, the algorithm output must also remain identical over those elements.
- **Anonymity:** Proteins expose preferences in an indistinguishable manner; voting information from different proteins is considered without regard to their identity.
- **Non-Dictatorship:** No single protein exclusively dictates the outcome of the voting procedure.
- **Proximity Preservation:** Small changes in the algorithm inputs must produce correspondingly small changes in the algorithm output, ensuring the external aggregation is robust against imprecise internal aggregation.
- **Monotonicity:** A positive change in the preference of one interaction class over another should induce a positive (or monotonic) change in the global order.

4.6 Impossibility Results

By applying Social Choice Theory to this protein folding framework, serious methodological difficulties are exposed. Even if the internal aggregation step is performed perfectly, the external aggregation step encounters fundamental mathematical impossibilities. We now examine whether a global aggregation procedure can satisfy the axioms defined above.

From Arrow's Impossibility Theorem [2], no aggregation function can simultaneously satisfy:

- Unanimity
- Non-dictatorship
- Independence of Irrelevant Alternatives

In addition to this classical result, further constraints arise in our setting. The application of these axioms reveals several critical conflicts:

- We cannot simultaneously respect Unanimity, Anonymity, and Proximity Preservation. If we cannot infer qualitative differences among proteins, respecting unanimity forces the procedure to fail proximity preservation, meaning it is not necessarily robust to imprecise internal aggregation methods.
- Waiving proximity preservation to maintain anonymity still frustrates the requirements, as we cannot simultaneously respect Unanimity, Anonymity, and Neutrality.

Consequently, any valid external voting procedure must waive neutrality (and the IIA), and it must also waive either anonymity or proximity preservation. Waiving neutrality presents methodological difficulties because the energetic potentials between interaction classes are our primary unknowns, prohibiting the pure reliance on precompiled physiochemical tables. Similarly, waiving anonymity is problematic because the method to correctly discriminate and favor certain "canonical" protein preferences over others is unclear a priori.

These results demonstrate that the external aggregation problem is fundamentally ill-posed under the thermodynamic and pairwise interaction assumptions. Even if individual protein preferences are correctly derived, combining them into a global ordering necessarily involves trade-offs between desirable properties. Ultimately, determining a universal energy function is inviable when relying solely on the aggregation of pairwise interactions.

5 Experiments

5.1 Implementation Details

We evaluate aggregation methods on a dataset of 15 proteins from the Protein Data Bank. For each of the proteins in the set, we perform pre-processing followed by internal aggregation of the votes based off of three methods. We perform the external aggregation step using the individual orderings of the proteins' preferences. Lastly, we apply evaluation metrics that compare the final orderings amongst each other and with each protein's individual vote.

5.1.1 Protein Data Acquisition and Pre-processing

For each of the proteins in the list, I acquired three data files from the AlphaFold Protein Structure Database(AFDB):

- .cif files: From these files, I acquired the Atomic Coordinates, in order to compute the Euclidean distance between the aminoacids, using the MMCIFParser from the Bio.PDB library.
- plDDT.json files: For each residue in the protein I acquired, I used the Predicted Local Distance Difference Test. I used this confidence measure to scale the preferences in the internal aggregation step.
- PAE.json files: For each protein, I acquired the Predicted Aligned Error, which is an NxN matrix the encaptures the predicted error in the relative position between two residues i and j.
- MJ_matrix.csv: I used the Amino Acid Index Database (AA Index) to download the Miyazawa-Jernigan (MJ) Statistical Potential matix. This matrix models the energy of interaction between the 20 residuals encapsulating Van der Waals and electrostatic forces.

5.1.2 Internal Aggregation

To rank each protein's ordinal preference on the amino acid interaction, I proposed three distinct methods of aggregating the preferences:

1. **Simple Proximity(Count**): For this method, we simply count the number of times each interaction happens, meaning that the distance between two residuals. Specifically, we rank the interaction classes by the number of residual pairs of the two amino acids that are under a threshold τ
2. **Miyazawa-Jernigan Potential**: For this method, for each pair of residuals where their distance is under a threshold τ, we add the value $(1 - \frac{dist}{\tau})(-MJ(i,j))$. This way, we account both for physio-chimical properties of the interaction classes and scaled by a decay of the distance. The reason for this is that a distance of 2Å represents a stronger interaction than a distance of 7Å
3. **Hybrid**: For this method, we adapt the Miyazawa-Jernigan Potential method and scale by the confidence matrices of plDTT and PAE. We calulate the utility of an interaction as $U(i,j) = Confidence(i,j) \times Phyics(i,j) \times Distance(i,j)$, where:
   (a) $Confidence(i,j) = \frac{plDDT_i + plDDT_j}{200} \times e^{-\frac{-pae_{ij}}{tau}}$ the confidence function is increasing in plDDT, as they are confidence measures, we also scale by the relative positional certainty of the interaction

(b) $Physics(i,j) = (-M_{ij}) + 1$, with this, we regularize the energy function to prevent neutral interactions

(c) $Distance(i,j) = 1 - \frac{dist}{\tau}$, we also scale by the linear decay as we did in the Miyazawa-Jernigan Potential method.

With this type of aggregation, the aim is to encapsulate into the preference both physico-chemical properties, distance measures, and confidence metrics so that the internal votes represent more informed preference lists

5.1.3 External aggregation

In the external aggregation phase we aim to obtain a global social choice from the individual rankings of the proteins. To do so, I implemented two voting counts:

- **Borda Count**: The Borda Count is a simple voting method that associates points to the different interaction classes given their position in each voter's rankings. For each protein input I(Pk),$1 \leq i \leq N$, the top-ranked interaction class would receive 210 points, the second-ranked element would receive 209 points, and so on. The least-ranked interaction class would receive 1 point. The points are aggregated among all voters. For an interaction class I, we denote the number of points attributed by protein Pk as $Points_k(I)$, and the number of points aggregated in total as points(I). Hence, for an interaction class ranked j-th at protein Pk, $Points_k(I) = m - j + 1$. Note that this voting method requires that ties are broken regarding internal scoring. If two interaction classes have the same internal scoring, we prefer the one with smaller Miyazawa Jernighan score. In the unlikely scenario that ties are still present, we then break them arbitrarily. In regard to Arrow's restrictions, this voting method satisfies Transitivity, Respect to Unanimity and Non-Dictatorship. It does not satisfy, however, the Independence of Alternatives.

- **Schulze method**: The Schulze method consists on associating points to different pairwise comparisons among interaction classes, across all protein inputs, resolving the final ordering such that it avoids cyclic dependencies that break transitivity.

  More specifically, for each protein input $I(P_k)$, if $I_x >_{I(P_k)} I_y$, we add one point to I_x in the $I_x \leftrightarrow I_y$ subcount. Formally, call $N(I_x, I_x \leftrightarrow I_y)$ the number of points assigned to I_x in the $I_x \leftrightarrow I_y$ count, in total [7]. If $N(I_x, I_x \leftrightarrow I_y) > N(I_y, I_x \leftrightarrow I_y)$, we say that I_x is the winner of $I_x \leftrightarrow I_y$. The concept of loser is defined analogously [3]. The difference between I_x and I_y is denoted $D(I_x \leftrightarrow I_y)$, defined as $|N(I_x, I_x \leftrightarrow I_y) - N(I_y, I_x \leftrightarrow I_y)|$.

  Then, we create a graph where the vertexes are the interaction classes, and if I_x is the winner of $I_x \leftrightarrow I_y$, we add an edge from I_x to I_y with weight $N(I_x, I_x \leftrightarrow I_y)$. After all edges were added, we say the relation $I_x \leftrightarrow I_y$ has weight w_{xy} if w_{xy} is the minimum weight of an edge in a particular path between I_x and I_y, where this particular path is the path that maximizes w_{xy} in the graph. The problem of finding such weight is known as the *widest path problem*, and it can be solved in polynomial time using a modified version of the Floyd-Warshall dynamic programming algorithm [5, 9].

6 Metrics

We ran the simulation for different τ ($\tau \in \{5\mathring{A}, 7\mathring{A}, 10\mathring{A}\}$). For each simulation, I have 3 different aggregation methods and 2 different external aggregation, and I used metrics to compare the methods both with each other, and within the methods to assess the similarity between the final orderings and the individual votes of the proteins. The metrics used assess the similarity in ordering, the directional consistency and the overlap in top ranking pairs.

6.0.1 Spearman Rank-Order Correlation Coefficient (ρ)

For each protein, the Spearman score is calculated by comparing the ranks of the interaction classes in the **Protein Ranking** (R_i) against the **Global Ranking** (S_i):

$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$

Where:

- $d_i = R_i - S_i$ (The difference between the rank the protein gave an interaction and its "true" rank in the global order).
- $n = 210$ (The number of possible amino acid pair combinations).

6.0.2 Kendall's Rank Correlation Coefficient (τ)

For each protein, the Kendall coefficient assess the directional consistency of our model. Kendall's τ looks at every possible pair of interaction classes (e.g., comparing CYS-CYS vs ALA-ALA) and checks if the protein rank and the consensus agree on which one is "better".

- Concordant Pair: Both the model and the consensus agree that Interaction A > Interaction B.
- Discordant Pair: the model says A > B, but the consensus says B > A.

The formula used to calculate the coefficient for each protein is:

$$\tau = \frac{\text{Number of Concordant Pairs-Number of Discordant Pairs}}{\frac{1}{2}n(n-1)}$$

6.0.3 nDCG (Normalized Discounted Cumulative Gain)

For each protein, this metric analyzes the overlap between the top 20 interaction classes in the individual ranking and the total orders.

Relevance (Gain): We assign a "relevance score" to each interaction. High-utility interactions, the ones listed at the top in the global ranking have high relevance.

Logarithmic Discounting: The "gain" you get for finding a relevant interaction is divided by the logarithm of its position.

- If you find the most important interaction at Rank 1, you get 100% of the points.
- If you find it at Rank 10, the score is "discounted" by $\log_2(10+1)$, making it worth much less.

To normalize the score (on a 0-to-1 scale), we calculate the **Ideal DCG (IDCG)**: the score you would get if the ranking was a perfect 1-to-1 match with the ground truth. We then divide the protein's score by this ideal:

$$nDCG = \frac{DCG}{IDCG}$$

7 Results

The primary objective of this experiment is to see to what extent a global consensus list can be found, such that it respects all axioms above and consistently aligns with each of the 15 proteins' individual preferences of the proteins. By aggregating the final lists for each combination of

internal and external aggregation methods, the following lists were extracted for the top 20 interaction classes.

COUNT_Schulze	**COUNT_BORDA**	**MJ_Schulze**	**MJ_BORDA**	**Hybrid_Schulze**	**Hybrid_BORDA**
(ALA, LEU)	(ALA, LEU)	(ALA, LEU)	(ALA, LEU)	(GLU, LYS)	(GLU, LYS)
(LEU, VAL)	(LEU, VAL)	(LEU, VAL)	(ILE, LEU)	(ASP, LYS)	(ASP, LYS)
(GLY, LEU)	(GLY, LEU)	(ILE, LEU)	(LEU, VAL)	(ARG, LYS)	(LYS, LYS)
(ILE, LEU)	(ILE, LEU)	(LEU, LEU)	(ALA, VAL)	(LYS, LYS)	(ARG, LYS)
(LEU, SER)	(LEU, PHE)	(GLY, LEU)	(GLY, LEU)	(ASN, LYS)	(ASN, LYS)
(LEU, LEU)	(LEU, SER)	(LEU, SER)	(ILE, VAL)	(ARG, GLU)	(ARG, GLU)
(ILE, VAL)	(ALA, VAL)	(ALA, VAL)	(LEU, PHE)	(ASP, GLU)	(ASP, GLU)
(ALA, GLY)	(ILE, VAL)	(ILE, VAL)	(LEU, SER)	(GLY, LYS)	(LYS, PRO)
(ALA, VAL)	(LEU, THR)	(LEU, PHE)	(LEU, LEU)	(LYS, PRO)	(GLY, LYS)
(LEU, PHE)	(GLY, SER)	(ALA, GLY)	(ALA, SER)	(GLU, GLU)	(GLU, GLU)
(GLY, VAL)	(LEU, LEU)	(ALA, ILE)	(LEU, THR)	(GLN, LYS)	(ASN, GLU)
(LEU, THR)	(ALA, SER)	(GLY, ILE)	(ALA, ILE)	(ASN, GLU)	(GLN, LYS)
(ALA, SER)	(GLY, VAL)	(GLY, PHE)	(GLY, SER)	(GLU, PRO)	(GLU, PRO)
(GLY, SER)	(GLU, LEU)	(GLY, SER)	(ALA, ALA)	(ARG, ASP)	(ARG, ASP)
(GLU, LEU)	(ALA, GLY)	(GLY, VAL)	(ALA, GLY)	(ASP, ASP)	(ASP, ASP)
(ASP, SER)	(LEU, LYS)	(LEU, THR)	(GLY, VAL)	(ARG, ARG)	(ARG, ARG)
(LEU, LYS)	(ALA, ILE)	(ALA, SER)	(SER, VAL)	(ARG, PRO)	(ARG, PRO)
(ALA, GLU)	(ASP, SER)	(ALA, ALA)	(GLY, PHE)	(ARG, ASN)	(ARG, ASN)
(GLU, GLY)	(LEU, PRO)	(ILE, SER)	(THR, VAL)	(GLN, GLU)	(GLN, GLU)
(GLU, LYS)	(GLU, VAL)	(SER, VAL)	(ILE, SER)	(ASP, PRO)	(ASP, PRO)

Table 1: Top 20 interaction classes for each method ($\tau = 7$Å)

Looking at the lists, we can see that there is high overlap for the lists that came from the sae internal aggregation method but that different in the external one. We can also see that there is high similarity in the simple count and the Miyazawa-Jernigan Potential method, which are both very different than the Hybrid version that includes the confidence scores. One possible explanation for this is that in the simple count and MJ Potential method, we can see that pairs containing Leucine (LEU) tend to appear in higher position due to the fact that this aminoacid is one of the most prevalent in protein structures. So its high frequency makes it statistically more likely to be closer to other amino acids. On the other hand, for the Hybrid method, we see pairs like (GLU, LYS) or (ASP, LYS) which are pairs of amino acids of opposite charge that form Salt Bridges. From a physico-chemical perspective, they form strong bonds that make the final protein structure more stable.

7.1 Analysis of Spearman Scores

The Spearman Alignment Scores reveal that across all distance thresholds, the hybrid method achieves a higher median score $\approx 0.75-0.80$ compared with $\approx 0.6-0.7$ for the other two methods. Although this method shows a stronger peak and median, it also has a higher variance and the lowest minimum. The increased spread can be attributed to the weighting of the confidence scores: for proteins in which the structures have high confidence scores, the model performs well, while for the proteins that have large areas with low confidence scores, the model mutes the interaction, effectively lowering their influence in the aggregation, whereas for the classical models, all interactions are treated the same.

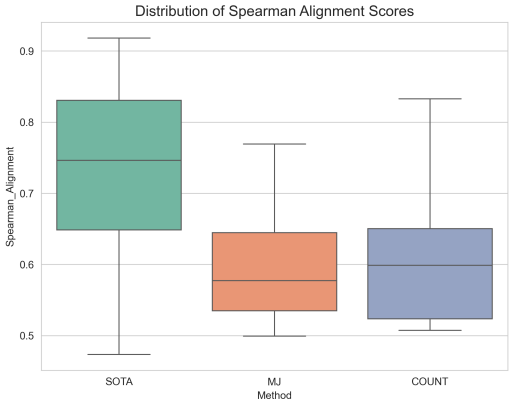

Figure 1: Distribution of Spearman alignment scores for t=5A

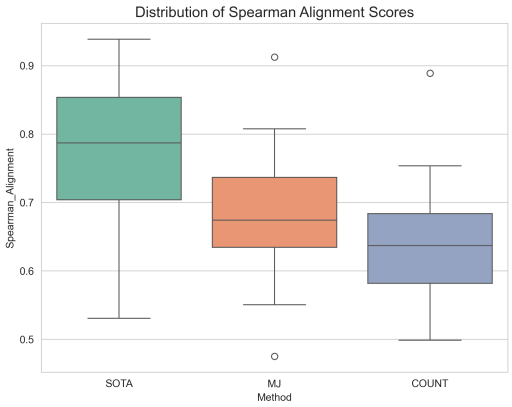

Figure 2: Distribution of Spearman alignment scores for t=7A

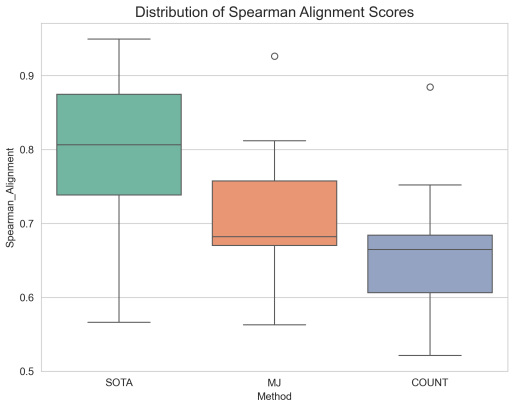

Figure 3: Distribution of Spearman alignment scores for t=10A

In the plots below, we can see the distribution of scores across protein sizes (as expressed in the number of amino acids in the protein chain). We can see that for smaller proteins, the models hae similar scores, while for larger proteins, the Hybrid model significantly outperform the other models. With larger proteins, the structure becomes noisier, with coordinates being

less accurate, so adding the gating terms ensure that only the most reliable interactions are counted, making the overall ranking more aligned.

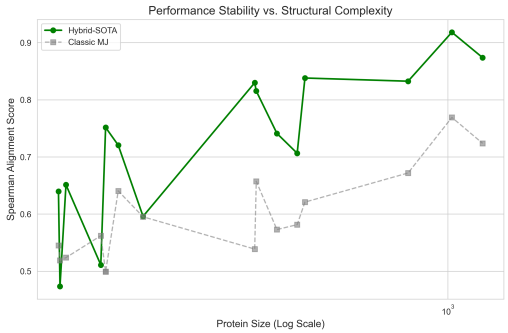

Figure 4: Performance relative to protein size, for t=5A

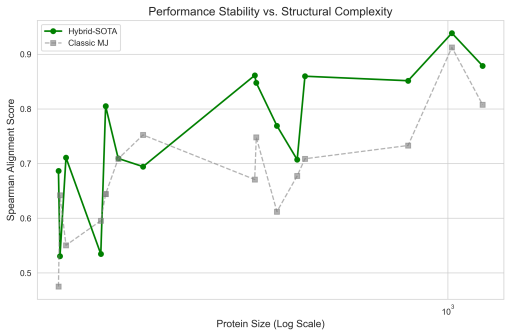

Figure 5: Performance relative to protein size, for t=7A

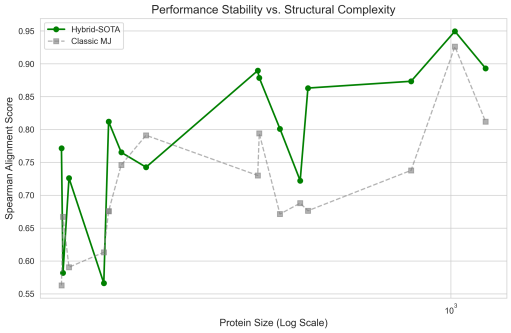

Figure 6: Performance relative to protein size, for t=10A

7.2 Analysis of Kendall Coefficients

While the Spearman scores provides a global overview of the alignment of two lists, the Kendall's τ coefficient assesses the model's pairwise consistency. The metric evaluates the probability that the model keeps the same order between two interaction classes as in the individual protein list. Again, for the Hybrid method we have higher scores (median, highest and lowest score), meaning that we can potentially extract a sublist of interaction classes that appear in the same order across all proteins and in the global ordering (especially in the Schulze Method as it satisfies Independence of Irrelevant Alternatives).

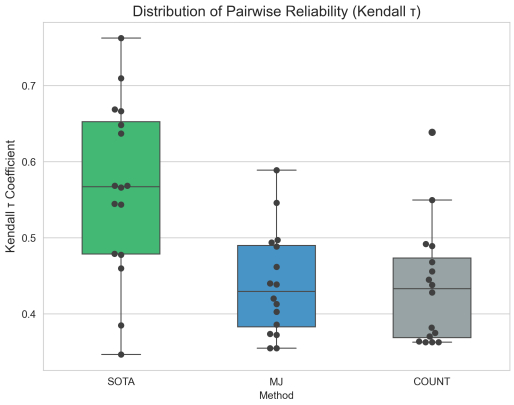

Figure 7: Distribution of Kendall Coefficient for t=5A

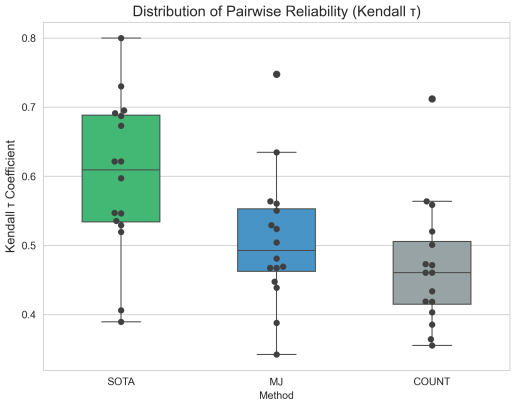

Figure 8: Distribution of Kendall Coefficient, for t=7A

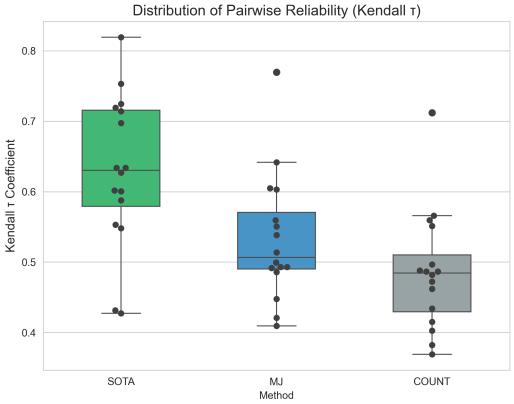

Figure 9: Distribution of Kendall Coefficient, for t=10A

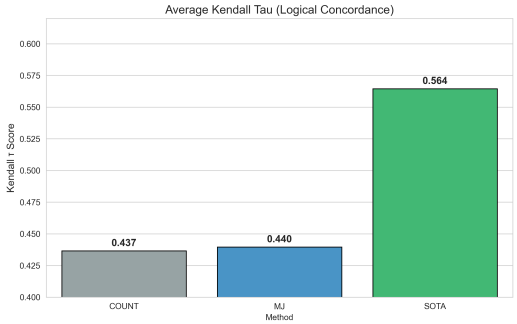

Figure 10: Average Kendall Coefficients for t=5A

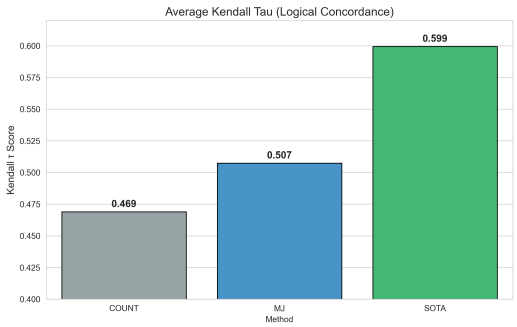

Figure 11: Average Kendall Coefficients, for t=7A

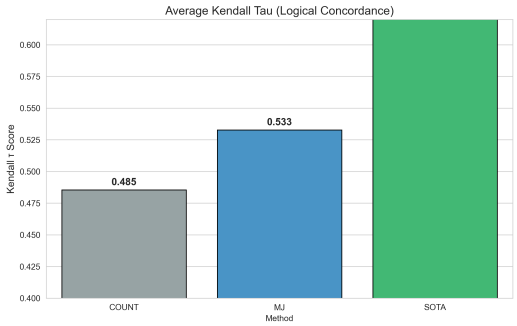

Figure 12: Average Kendall Coefficients, for t=10A

To determine the pairwise stability of the orderings, we also compared the Kendall Coefficients across the different protein sizes, and observed a similar effect with the one for the Spearman Scores, namely that the scores seem to be the same across methods for smaller proteins, with the Hybrid method scoring significantly higher for larger proteins, suggesting that the logical orderings remains valid for larger proteins as well.

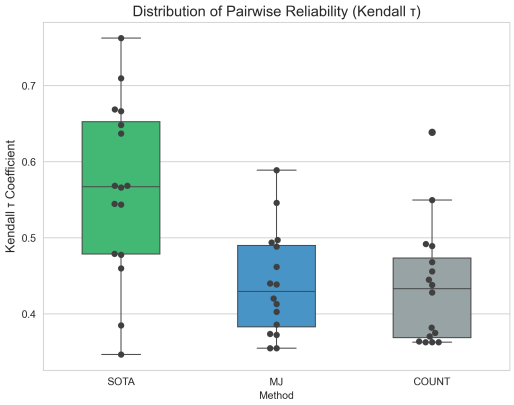

Figure 13: Distribution of Kendall Coefficients across Protein Sizes for t=5A

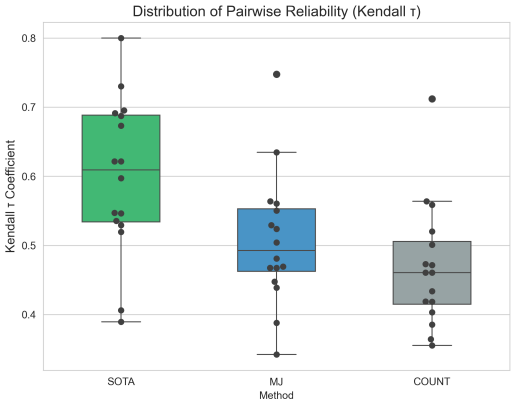

Figure 14: Distribution of Kendall Coefficients across Protein Sizes for t=7A

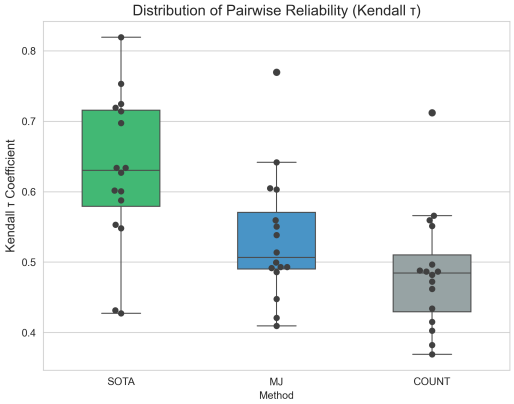

Figure 15: Distribution of Kendall Coefficients across Protein Sizes for t=10A

7.3 Analysis of the nDCG Score

We employed the Normalized Discounted Cumulative Gain with a focus on the top 20 interaction classes in the individual orderings and the global consensus ranking. In the context of protein

folding, the top 20 interaction classes should represent the stabilizing interactions of the fold, such as the hydrophobic core and the salt bridges. While all methods achieved relatively high scores (above 0.8), we can see that the hybrid method achieved clustered scores around 0.99. This suggests that the interaction classes that were prioritized by the Schulze method match the rank and relevance of the interaction classes in the individual rankings.

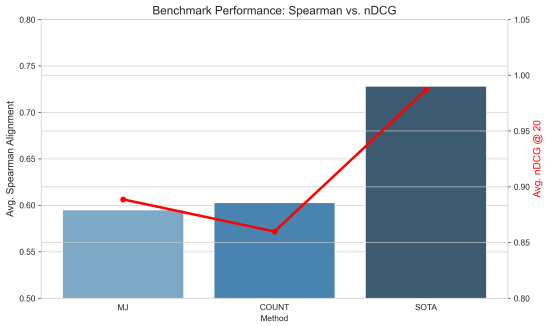

Figure 16: Average nCDG scores for t=5A

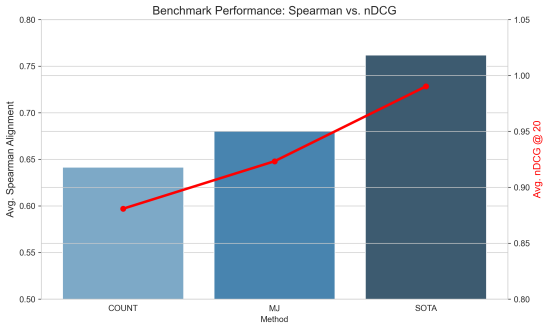

Figure 17: Average nCDG scores for t=7A

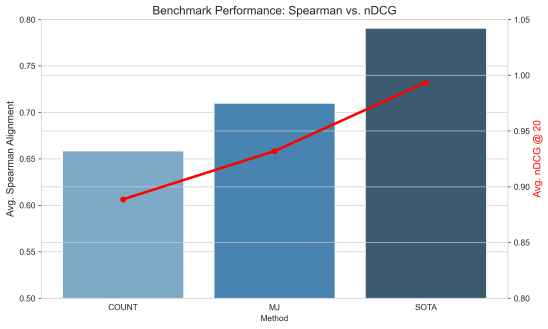

Figure 18: Average nCDG scores for t=10A

There is a significant performance gap across methods for this coefficient, as the variance is almost zero for the Hybrid method. This almost zero variance suggests that we can infer that the top 20 interaction classes remain invariant across proteins. On the other hand, for the other methods, although they show relatively high medians, there is greater variance across proteins, suggesting that a list of top 20 interaction classes cannot be extracted from those internal aggregation methods.

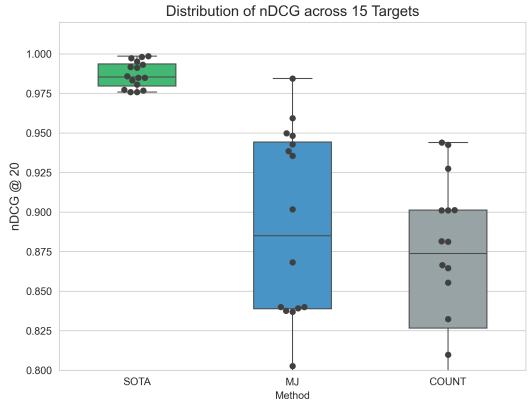

Figure 19: Distribution of nCDG scores for t=5A

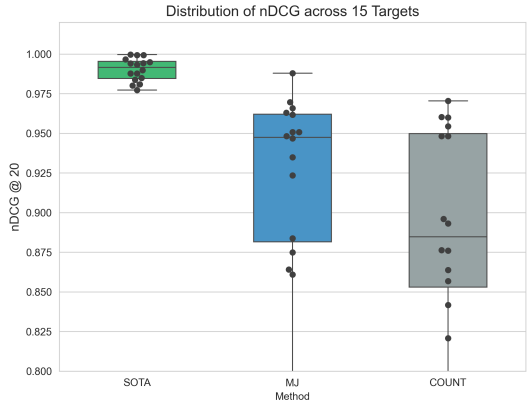

Figure 20: Distribution of nCDG scores for t=7A

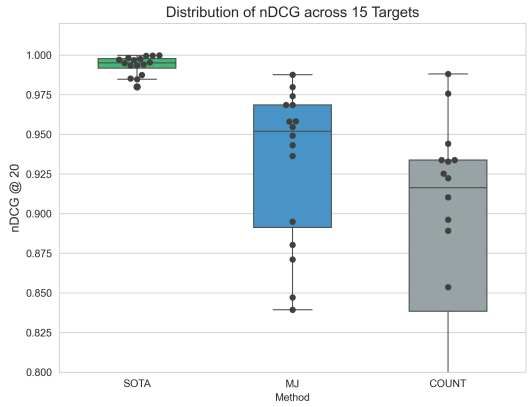

Figure 21: Distribution of nCDG scores for t=10A

8 Conclusion and future work

We presented a social choice formulation of protein folding energy estimation. Our results show that fundamental impossibility theorems apply, limiting the construction of universal pairwise

energy functions. This highlights that the challenge is not only computational but also theoretical. However, our results show that even though a universal pairwise energy function is impossible to derive across all interaction classes, because the Hybrid method shows invariance across the top 20 interaction classes, we can use that ordering as a baseline for the most important pairwise amino acid interaction across proteins.

Future work includes exploring richer interaction models, for deriving better internal aggregation lists and relaxing aggregation assumptions. Another direction for future work is studying misfolded proteins and seeing how their votes would compare to the global orderings, to assess from where the misfolding came. As we have shown invariance in the Hybrid method for the top 20 interaction classes, having a misfolded protein rank different interaction classes very high would suggest that the protein folded in such a way that it favored those pairwise interaction instead of the globally favored ones.

References

[1] Christian B. Anfinsen. "Principles that govern the folding of protein chains". In: *Science* 181.4096 (1973), pp. 223–230.

[2] Kenneth J. Arrow. *Social Choice and Individual Values*. Wiley, 1951.

[3] Jean-Charles de Borda. "Mémoire sur les élections au scrutin". In: *Histoire de l'Académie Royale des Sciences* (1781).

[4] Ken A. Dill and Justin L. MacCallum. "Protein folding problem". In: *Science* 338.6110 (2012), pp. 1042–1046.

[5] Robert W Floyd. "Algorithm 97: Shortest path". In: *Communications of the ACM* 5.6 (1962), p. 345.

[6] Sanzo Miyazawa and Robert L. Jernigan. "Estimation of effective interresidue contact energies from protein crystal structures". In: *Macromolecules* 18.3 (1985), pp. 534–552.

[7] Markus Schulze. "A new monotonic, clone-independent, reversal symmetric, and Condorcet-consistent single-winner election method". In: *Social Choice and Welfare* 36.2 (2011), pp. 267–303.

[8] Amartya Sen. *Collective Choice and Social Welfare*. Holden-Day, 1970.

[9] Stephen Warshall. "A theorem on Boolean matrices". In: *Journal of the ACM* 9.1 (1962), pp. 11–12.