Proceedings of ICBioMed 2026 Symposium: Biomedical Engineering and Smart Medicine
DOI: 10.54254/2753-8818/2026.HZ38370

Impact of Hydrophobic Residue Ratio and Hydrophobic Distribution Pattern on Protein Folding in a Two-Dimensional Hydrophobic-Polar Lattice Model: A Simulated Annealing Simulation

Ningyu Qiu

Academy of Life and Natural Science, Xi'an Jiaotong-Liverpool University, Suzhou, China
Ningyu.Qiu24@student.xjtlu.edu.cn

Abstract. The process of protein folding depends on the interaction between sequence composition and conformational search. This study, in detail, the effect of the hydrophobic residue ratio and the spatial distribution pattern on thermodynamic stability and search efficiency by applying minimalistic two-dimensional hydrophobic-polar (HP) lattice model. Four different hydrophobic ratios (30 percent, 40 percent, 50 percent, and 60 percent) and three distribution modes (aggregated, alternating, and random) were used to generate 300 independent sequences of fixed length 30. The folding simulations were done with simulated annealing protocol using corner-flip and end-rotation moves on a square lattice. As results indicate, increasing the hydrophobic content reduces the conformational energy and radius of gyration monotonically, hence establishing the hydrophobic effect as the main driving force of chain compaction. Kinetic accessibility, however, has a different trend: aggregated distributions have the lowest absolute energies but take long convergence times, low success rates and high conformational degeneracies, particularly, at high hydrophobic ratios. On the other hand, alternating distributions have small thermodynamic stability but fast convergence and high success rates. These results show that sequence architecture has a role in controlling the ruggedness of folding energy landscape and underscore the disconnect between thermodynamic optimality and kinetic reliability in simple protein models.

Keywords: hydrophobic-polar model, protein folding, simulated annealing, sequence pattern, energy landscape

1. Introduction

The protein folding phenomenon is the mechanism through which a polypeptide chain attains its functional three-dimensional shape and it has remained one of the core issues in molecular biology [1]. Since all-atom simulations have very large computational requirements, simpler lattice models are used most commonly to represent the critical physics of the folding process [2, 3]. The two-dimensional hydrophobic-polar (HP) lattice model developed by Dill and collaborators is one of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ38370

these models that give the minimum but still rich framework of the study of how sequence composition determines the conformational results [4-6]. In this model residues are divided into two categories hydrophobic (H), polar (P); the chain is constrained to a square lattice with the exclusion of self-contacts and energy is given by the negative sum of non-consecutive H-H interactions [4]. Maximization of buried hydrophobic contact is thus minimized in energy, an effect similar to hydrophobic effect in natural proteins [5].

Earlier computational experiments based on HP models have also frequently analyzed the impact of the hydrophobic residue ratio or the spatial distribution in isolation [6, 7]. Nonetheless, with a fixed chain length, the combined effect of H-ratio and H-distribution pattern on thermodynamic stability and search efficiency has not been compared experimentally. This work will fill this gap in this paper by creating 300 independent sequences and folding them using simulated annealing folding simulations. This work aims at explaining the role of sequence architecture in determining both the final structure and the intermediate steps involved in the folding process.

2. Methodology

2.1. Sequence generation

The entire set of test sequences had a uniform chain length of 30 residues. This chain length was selected to trade-off between computational accessibility and non-trivial folding behavior, including the development of a compact core and the existence of more than two local energy minima.

The sequences were built based on four hydrophobic residue ratios per experimental condition: 30 percent (9 H and 21 P), 40 percent (12 H and 18 P), 50 percent (15 H and 15 P), and 60 percent (18 H and 12 P). Three different distributions were established. The combined (clustered) regime is where hydrophobic residues are organized into either one or two consecutive chunks placed at certain parts of the chain resembling a continuous hydrophobic core. The alternating distribution is where hydrophobic and polar residues are distributed in an alternating fashion with the greatest regularity possible by the given ratio, creating a kind of periodic-copolymer structure. In the random regime, hydrophobic residues are allocated randomly and strictly keeping the precise amount that was calculated by the ratio, and every random permutation forms a separate sequence implementation.

To each of these combinations of hydrophobic ratio and distribution pattern there corresponded 25 independent sequences resulting in a total of 300 sequences (4 ratios x 3 patterns x 25 sequences). The repetition helps to ascertain that the observed effects are the result of the sequence characteristics, not the accidental features of the particular sequences.

2.2. Energy function and structural metrics

The simulations were performed on a two-dimensional square lattice. Every residue is assigned an individual lattice site and neighboring residues of the backbone are joined by unit-length bonds in the axial directions. The self-avoidance restriction is strictly applied to the system: two different residues cannot be located at the same lattice site at the same time.The conformational energy adopts the standard HP model energy function, as follows:

$$E = -N_{HH}, \tag{1}$$

DOI: 10.54254/2753-8818/2026.HZ38370

With the symbol of N_{HH} indicating the count of non-consecutive H-H contacts. A contact is defined as two H residues at adjacent lattice positions (one unit apart horizontally or vertically), which are also not adjacent in the primary sequence. All such contacts make -1 contribution to the total energy; thus, the global energy minimum is achieved with the conformation that has the maximum number of buried hydrophobic contacts. Radius of gyration (R_g) is determined as the root-mean-square distance between all residues to the center of geometry and it is used as a measurement of structural compactness.

2.3. Simulated annealing algorithm

The simulation of folding uses simulated annealing algorithm which is a stochastic optimization method based on the physical annealing process in metallurgy [8-10]. This approach is especially useful in dealing with the complex energy landscapes that have several local minima, which makes it ideal to solve the HP lattice folding problem.

The procedure of simulation is described below. To start with, an initial conformation is created by positioning the sequence in the form of self-avoiding random-walk on the lattice. The initial temperature T_{max} is chosen such that it is high enough so that about 80 percent of suggested movements are accepted to explore the conformational space extensively in the initial stages.

At every step, the algorithm makes a local conformational step. The types of moves available are corner flip (a single residue rotates about an axis determined by its neighbors), and end rotation (one of the terminal residues is relocated to a neighboring unoccupied lattice site). When the suggested move leads to the self-intersection, the process is stopped right away; if not, the energy difference ΔE is determined. The probability of acceptance p is calculated using the Metropolis criterion:

$$p = min(1, exp(-\frac{\Delta E}{(k_B T)}). \tag{2}$$

Here $k_B T$ is the Boltzmann constant (scaled to 1 in reduced units), and T is the current temperature. The system will always accept those changes that decrease the energy; with those that increase the energy, the probability of acceptance goes down exponentially as the energy increases and it is inversely related to the temperature.

The temperature drops down to T_{min} starting with T_{max} after N_{total} iterations according to a linear annealing schedule. The particular parameter values used are T_{max} = 5.0, T_{min} = 0.01 and the total number of iterations N_{total} is 500,000. These parameter values were selected by preliminary tests to even out exploration at high temperatures and exploitation at low temperatures.

3. Results

3.1. Effect of hydrophobic residue ratio

With a fixed distribution of amino acids and increasing proportions of hydrophobic residues (30-60 percent), the least energy achieved is lower and the radius of gyration is decreased. It means that a higher hydrophobic concentration causes the conformation to be more compact, and raises the number of buried H-H contact points. Nevertheless, the number of iterations needed to achieve the minimum energy is not shown in a monotonic decreasing fashion; the precise correlation relies on the interaction between the density of the hydrophobic residues and the induced ruggedness of the

DOI: 10.54254/2753-8818/2026.HZ38370

energy-landscape. This implies that even too much hydrophobic content could complicate the search space.

3.2. Effect of distribution pattern

With a constant amount of hydrophobic residues, the different distribution patterns result in considerable variability of energy and search behavior. In each hydrophobic ratio, the clustered distribution always attains the lowest mean energy. E.g. at hydrophobic ratio of 50% the average minimum energy of the clustered pattern is -7.8 compared to -6.9 of the random patterns and -6.5 of the alternating patterns. The explanation behind this benefit is that uninterrupted hydrophobic segments are able to fold into a dense core without the need to rearrange the chains over long distances. Furthermore, the clustered pattern is also associated with the lowest average radius of gyration which suggests that it has a very compact globular structure.Nevertheless, the kinetics of the aggregated mode are more disadvantageous. On average, it requires more convergence steps than the alternating mode does, and has a lower success rate after 40 independent runs in particular when the hydrophobic ratio is 60 percent. With certain aggregation-distributed sequences with significant hydrophobicity, the simulation is prone to become stuck in a compact but suboptimal conformation, which cannot be easily escaped at low temperature.The alternating distribution has the opposite tendency. Although its average energy is the largest of the three modes at every hydrophobic ratio, it shows the quickest convergence rate and the highest success rate. It seems that a periodic pattern of hydrophobic and polar residues could be used to stop the premature creation of local structures that are too stable, facilitating the exploration of larger conformational space by the chain during the high-temperature phase of simulated annealing. Concerning random distribution, its results in all metrics are in between those of the other two, which acts as a reference point indicating the presence of systematic differences due to the ordered arrangements and aggregations patterns.

3.3. Interaction effects

Heatmaps and two-factor analysis were used to study the interaction of hydrophobic ratios and distribution patterns. It was found that there are significant interactions between energy and convergence steps. The changes in the distribution patterns were insignificant at small hydrophobic ratios (30-40 percent) since the number of hydrophobic residues is limited, thus, limiting the range of possible contact formations regardless of the sequence order. With the hydrophobic ratio increased to 50, the clustering and alternating distribution patterns became noticeably different. The interaction reached its peak at 60% of the hydrophobic ratio: at the combination of a high level of hydrophobicity and a cluster configuration, the lowest absolute energy was obtained but at the highest computational cost, which included the maximum number of convergence steps and minimum success rates. However, at 60 percent, the alternating sequences remained fairly efficient in their searches although energy use increased slightly. Such results imply that the penalty of excessive hydrophobic packing is regulated by sequence architecture. Precisely, the clustered pattern increases the thermodynamic benefits and the kinetic costs caused by the large hydrophobic content, whereas the alternating pattern successfully separates the two factors, trading away some thermodynamic optimality in favor of better search reliability.

DOI: 10.54254/2753-8818/2026.HZ38370

4. Discussion

4.1. Hydrophobic effect being the main driving force and how it differs in complete folding

It has been found that there exists a very good empirical monotonic relationship between a hydrophobic residue and thermodynamic stability, which is highly consistent with the preliminary theoretical model of Dill et al. [4]. It is commonly known that already in his early research on HP lattice model, Dill demonstrated that hydrophobic effect, i.e., the preference of nonpolar residues to be accessible to aqueous solvent to the smallest possible level, is the main thermodynamic driving force behind protein folding. The present observations are consistent with this idea: in each type of distribution, a larger proportion of hydrophobic residues produces more favorable H-H interactions reducing the size of conformational energy and enhancing the structural packing which is quantified through the radius of gyration. These findings have crucial implications to the knowledge of coarse-grained models. Even a minimalist model that does not take into account any chemical information beyond the hydrophobic-polar distinction reveals the tendency of hydrophobic residues to form clusters in the protein core. The fact that this trend is found in 300 independent sequences suggests that it is inherent and largely independent of sequence-specific peculiarities. The present findings therefore validate the existing point of view that hydrophobic effect cannot be viewed as a simple additive but as an organizing entity behind the low-resolution behavior of collapsing polypeptide

However, despite the high thermodynamic advantages of large hydrophobic content, the kinetic data present a far subtler image. The non-monotonicity of the hydrophobic ratio/search efficiency correlation and the particular behavior of clustered and alternating distributions justify the theoretical dichotomy between collapse and full folding of hydrophobic residues postulated by Haran [8]. Haran suggested that despite a rapid and early collapse of a polypeptide chain into a compact and inaccessible structure to the solution, this does not necessarily imply reaching a native topology. In recent simulations, sequences with large hydrophobic ratios and clustered architectures rapidly collapse into compact structures of small radius of gyration. The collapsed states do not however always correspond to global energy minima, but to local minima whereby hydrophobic blocks make favorable, yet not optimal, internal interactions. Such configurations need much more iterations to be escaped by the simulated annealing algorithm. Sometimes, the system fails to overcome the energy barrier entirely until the temperature reduces below the energy required to overcome it. This is particularly obvious at 60 percent clustered, where the success rates are extremely low despite the apparent thermodynamic preference of the target state.

The alternate distribution scheme provides an important comparative paradigm. The use of hydrophobic residues and polar spacers prevents the chain from forming large, stable hydrophobic complexes prematurely. As a result, the polypeptide has more conformational flexibility in its early phases of the simulation, thus exploring a larger set of compact conformations before it commits to a particular collapsed state. This behavior is typical of a folding funnel with decreased local ruggedness, where the path to the native state is kinetically accessible even when the chain is compacted. Taken together, these results suggest that although the hydrophobic effect is a potent initiator of chain collapse, the particular arrangement of hydrophobic residues controls whether this collapse can be used to guide the search towards the global energy-free minimum, or leads to a kinetic stall in a local trap. The thermodynamic drive and the kinetic availability of the native state become uncoupled features of the folding landscape, a conclusion that is consistent with the theoretical expectations of both Dill and Haran on the subject of hydrophobic stabilization and protein folding as a multi-step process, respectively.

DOI: 10.54254/2753-8818/2026.HZ38370

4.2. The cost of excessive hydrophobic packing

A fascinating fact is that the thermodynamically optimal conditions, namely, a significant proportion of hydrophobic residues and their spatial distribution in clusters, do not always give the strongest results of folding. The outcome is consistent with the experimental evidence of the natural proteins reported by the authors; including Ventura et al. who showed that overly large or too close-packed hydrophobic cores caused the introduction of conformational strain and destabilization [9]. There have been instances where these configurations may cause proteins to aggregate rather than assisting them to fold efficiently. Despite the fact that the HP model is unable to capture all of the chemical complexities involved in these phenomena, it can also produce similar effects on the energy-landscape level: the high density of hydrophobic patches causes energetically deep but kinetically trapped local minima that prevent the system from reaching the global free-energy minimum.

The lowest-energy conformation degeneracy provided in this paper provides a further reinforcement of such an interpretation. Groups of sequences with 60 percent hydrophobicity and clustering distribution have a tendency to show multiple conformations with the same or nearly equal energy levels. Such conformational degeneracy means that there are several competing energy wells of comparable depth in the energy landscape and the energy barrier between them cannot be easily crossed during the simulated annealing process. Conversely, sequences having an alternating distribution or moderate hydrophobic fractions generally converge to a single or almost singular lowest-energy state indicating that their energy landscape is less rough and more funnel-shaped. The implication of these findings is that the natural proteins could be exposed to a trade-off whereby the maximal burial of hydrophobic residues is sought but a smooth enough energy landscape is maintained to allow predictable folding on biologically relevant times. In conclusion, evolutionary selection can select sequence patterns that have sufficient thermodynamic stability without making the energy landscape too rugged, even when these patterns do not provide maximum absolute counts of hydrophobic contacts.

4.3. Limitations

A number of drawbacks in the current research should be mentioned. Firstly, although the simulated annealing iteration number is used as an indicator of search efficiency, it is not equivalent to experimental folding kinetics or rate. Computational optimization algorithm timescales are by their nature incommensurate with microsecond-to-second regimes of biological protein folding. Secondly, the topological limitations imposed by two-dimensional lattice geometry are of a completely different character than those of three-dimensional space. It is important to note that, in 2D, the excluded volume effects and allowed contact geometries are more restricted, and this may cause exaggerated kinetic penalties due to clustered hydrophobic residues. Lastly, the binary H/P alphabet does not consider the chemical variety of natural amino acids, electrostatic interactions, hydrogen bonding, backbone conformational preferences, all of which play a critical role in determining the folding energy landscape.

Although such simplifications are made, the value of the HP lattice model lies exactly in its ability to isolate the effects of hydrophobic patterning in isolation of other complicated confounding factors. The similar nature of the trends found in hundreds of independent sequences also indicates that the association between the hydrophobic ratio and the distribution is a natural physical phenomenon and not an artifact of the simulation model.

DOI: 10.54254/2753-8818/2026.HZ38370

5. Conclusion

Both percentage as well as spatial distribution were found to influence the way proteins fold in 2D HP lattice model. These findings mean that the thermal stability increases linearly as the hydrophobic residue amount is increased with the lower conformational energy and reduced radius of gyration. How it contributes to the present perspective is that it helps sustain the concept that the hydrophobic effect is the main cause of polypeptide chain collapse and compaction.

Nevertheless, the research suggests that thermodynamic optimality doesn't imply kinetic accessibility. The arrangement of hydrophobic residues in space has much effect on search efficiency. To be more precise, although a distribution of hydrophobic residues which are clustered may lead to minimum energy requirement because they will have to centralize the hydrophobic core, these distributions are usually associated with slower convergence time, lower success rates and greater conformational degeneracy (especially at high number of hydrophobic residues). On the other hand, an alternating distribution would sacrifice some aspects of thermodynamic stability to achieve faster and more stable convergence. This means that sequence patterns are important predictors of ruggedness of energy landscape.

The interaction between hydrophobicity and their spatial distribution also reflects on the complexity of the sequence-structure relationship. At low hydrophobicity levels, the selection of the distribution mode plays a very minor role. But as more hydrophobicity occurs, high hydrophobic content and clustered structure will be combined to form an energy landscape with extremely large local minima and poor global search properties. The findings are discussed within three theoretical models: the paramount significance of hydrophobic interactions that lead to chain compaction, the mechanistic distinction between hydrophobic collapse and cooperative folding, and the hypothetical instability that could be caused by excessive hydrophobic packing.

When these results are considered as a whole, it becomes evident that the protein folding in simple models is not only determined by an amount of hydrophobic residues, but also by the spatial localization of those residues in the primary sequence. This model can be applied to three-dimensional lattice tasks in the future, additional interaction terms can be added to better approximate the native amino acid chemistry, or evolutionary algorithms can be used to determine sequence motifs which are the most thermodynamically stable and the most resistant to folding. Such extensions will expand information on how sequence decides the end point as well as the path of folding.

References

[1] Levinthal, C. (1969). How to fold graciously. In: *Mössbauer spectroscopy in biological systems*, University of Illinois Press, 22–24.
[2] Bryngelson, J. D., Onuchic, J. N., Socci, N. D., & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics, 21*(3), 167–195.
[3] Dill, K. A., & MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *Science, 338*(6110), 1042–1046.
[4] Dill, K. A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry, 24*(6), 1501–1509.
[5] Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Advances in Protein Chemistry, 14*, 1–63.
[6] Lau, K. F., & Dill, K. A. (1989). A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules, 22*(10), 3986–3997.
[7] Chan, H. S., & Dill, K. A. (1991). Polymer principles in protein structure and stability. *Annual Review of Biophysics and Biophysical Chemistry, 20*, 447–490.
[8] Haran, G. (2012). How, when and why proteins collapse: The relation to folding. *Current Opinion in Structural Biology, 22*(1), 14–20.

DOI: 10.54254/2753-8818/2026.HZ38370

[9] Ventura, S., Vega, M. C., Lacroix, E., Angrand, I., Spagnolo, L., & Serrano, L. (2002). Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nature Structural Biology, 9*(6), 485–493.
[10] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics, 21*(6), 1087–1092.