THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Elucidating Physicochemical Features of Holin Proteins Responsible for Bacterial Cell Lysis

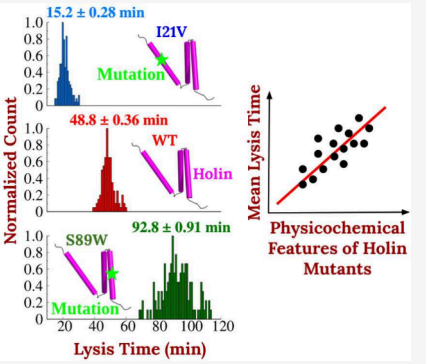
Anupam Mondal, Hamid Teimouri, and Anatoly B. Kolomeisky*

Cite This: J. Phys. Chem. B 2024, 128, 7129−7140

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Bacterial resistance to conventional antibiotics stimulated the development of so-called "phage therapies" that rely on cell lysis, which is a process of destroying bacterial cells due to their infections by bacterial viruses. For λ bacteriophages, it is known that the critical role in this process is played by holin proteins that aggregate in cellular membranes before breaking them apart. While multiple experimental studies probed various aspects of cell lysis, the underlying molecular mechanisms remain not well understood. Here we investigate what physicochemical properties of holin proteins are the most relevant for these processes by employing statistical correlation analysis of cell lysis dynamics for different experimentally observed mutant species. Our findings reveal significant correlations between various physicochemical features and cell lysis dynamics. Notably, we uncover a strong inverse correlation between local hydrophobicity and cell lysis times, underscoring the crucial role of hydrophobic interactions in membrane disruption. Stimulated by these observations, a predictive model capable of explicitly estimating cell lysis times for any holin protein mutants based on their mean hydrophobicity values is developed. Our study not only provides important microscopic insights into cell lysis phenomena but also proposes specific routes to optimize medical and biotechnological applications of bacteriophages.

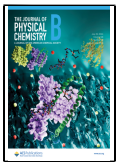


INTRODUCTION

Bacterial resistance to conventional antibiotics has already become a major global health threat.[1] These observations stimulated the search for alternative antibiotic strategies, and bacterial viruses, known as bacteriophages, have emerged as potentially powerful agents of eliminating bacterial infections.[2−4] There is a significant interest in developing so-called "phage therapies" due to the rapid rise of multidrug resistant bacteria worldwide together with the decrease in the development of new antibiotic drugs.[5] There are multiple advantages to using bacteriophages for treating bacterial infections.[4,6] However, the rational design of new phage therapies requires a comprehensive microscopic understanding of cell lysis mechanisms at the molecular level which is currently lacking.

One of the most intensely investigated cell lysis events are those that are stimulated by λ bacteriophage viruses.[7] Upon infiltrating the bacterial cell and selecting a viral replication pathway,[8] the virus triggers the host machinery to produce holin proteins, which regulate the timing of lysis and thus the duration and efficiency of the infection cycle.[9−11] These holin proteins are extremely diverse but generally possess three transmembrane domains (TMDs).[12] During the early phase of the infection cycle, holin proteins progressively accumulate within the inner membrane, originally without causing discernible effects on cell physiology or membrane integrity.[13] However, at some specific moment, when the concentration of holins surpasses a critical threshold, a sudden transition occurs, leading to the aggregation of holin proteins and the formation of irregular lesions within the membrane with an average diameter of approximately 300−350 nm.[14] These lesions serve as conduits for releasing of other type of proteins, phage endolysins, from the cytoplasm. It is essential to emphasize here that holin-induced pores are not specifically selective for endolysins but are sufficiently large to allow the passage of phage endolysins and other molecules from the cytoplasm into the periplasmic space. The primary function of these pores is to facilitate the release of endolysins, which triggers the rapid degradation of the peptidoglycan cell wall within seconds.[13] In the case of phages infecting Gram-negative bacterial hosts with inner and outer membrane walls, an additional phase in the lysis cascade unfolds, which involves the synthesis of spanin proteins. The spanin proteins facilitate the fusion of the inner and outer membranes,[11,15−17] ultimately completing the process of bacterial membranes cell fusion, thereby allowing virions to disseminate and infect other bacterial cells. The details of these events are schematically shown in Figure 1.

Received: May 8, 2024
Revised: June 20, 2024
Accepted: June 26, 2024
Published: July 10, 2024



https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129−7140

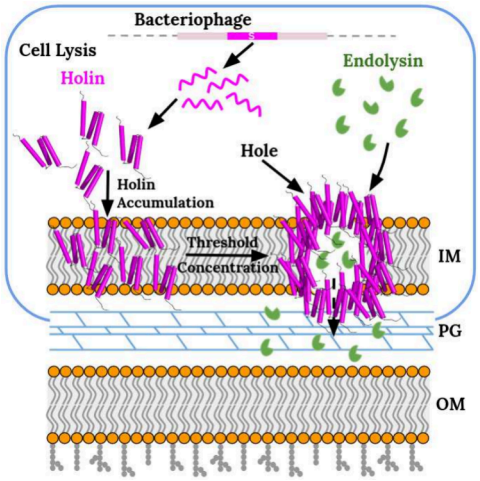


Figure 1. Schematic view of the Gram-negative bacterial cell infected by phage viruses, which undergoes a series of sequential events. Holin proteins, produced by the phage, diffuse into the inner membrane (IM), initiating the formation of holes via aggregation after reaching a threshold concentration. Subsequently, endolysin enzymes, also synthesized by the phage, traverse these holes to access the peptidoglycan (PG) cell walls, where they execute their function by breaking down the cell wall structure.

There are several experimental investigations of cell lysis processes that utilized a mutational analysis of holin proteins.[18−23] These studies found that mutations in different regions of the holin molecule might have a strong effect on the dynamics of cell lysis, most probably by impacting the oligomerization of these proteins inside the membrane. Surprisingly, the wild-type holin species did not exhibit the fastest cell lysis dynamics. Specifically, it was found that some mutations could enhance the holin activity, leading to premature lysis,[13] whereas other mutations that impaired the holin function could prolong the time required for complete cell lysis. Furthermore, it has been observed that mutations alter the structure and conformational dynamics of holin proteins, which should correlate with their functions.[24] In addition, the experimental data demonstrated that the relative hydrophobicity of TMDs might impact their externalization, meaning that these domains could be exposed outward from the membrane, affecting the activation and hole formation.[24] There are arguments that hydrophobicity might play a pivotal role in cell lysis by influencing the stability and kinetics of holin insertion into membranes.[25−29] Additionally, the mutations might also change the interactions between different TMDs, ultimately impacting the externalization of TMDs and leading to alteration of lysis triggering times.[24]

In a more quantitative experiment, a site-directed mutagenesis was employed to generate a library of *E. coli* lysogens differing from the parent allele by one or two amino acid substitutions in various regions of the three transmembrane domains of the holin protein.[30] Subsequent analysis of lysis events at the single-cell level for ∼100 cells per strain revealed a surprisingly wide spectrum of mean lysis times, spanning from 15 to 175 min across different mutant species.[30] While all these findings underscore the significant impact of holin mutations on the dynamics of cell lysis events, it remains unclear what molecular properties of holin proteins govern their abilities to break cellular membranes most efficiently.

To address this knowledge gap, we present a theoretical investigation to clarify the molecular picture of cell lysis by employing a statistical analysis. It is based on the assumption that the chemical composition and structure of holin proteins correlate with their ability to break the bacterial cellular membranes. Our results reveal strong correlations between specific physicochemical features of holin species and cell lysis dynamics, providing novel microscopic insights into the mechanisms of these complex biological processes. Notably, our analysis demonstrated an important role of local hydrophobicity around the mutation point in controlling the cell lysis dynamics. It is observed that mutations increasing the local hydrophobicity facilitate faster membrane insertion and subsequent membrane rupture by holin proteins, whereas mutations lowering the local hydrophobicity exhibit the opposite effect of slowing down the cell lysis dynamics.

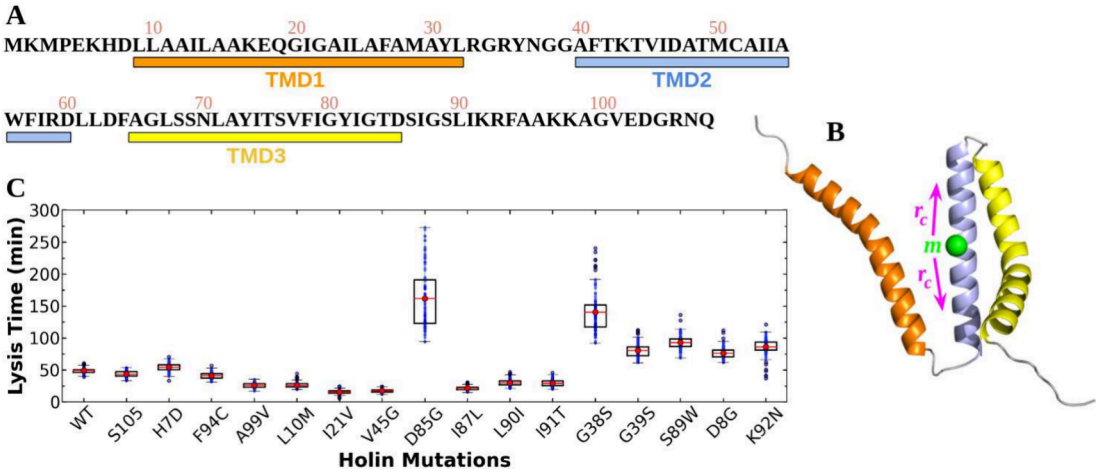


Figure 2. (A) Primary sequence of the WT holin protein. Rectangles highlight the three transmembrane domains (TMDs): TMD1 (orange), TMD2 (lightblue) and TMD3 (yellow). (B) Predicted 3D structure of the holin protein based on its primary sequence using AlphaFold.[32] A green circle denotes an arbitrary point mutation introduced at position m. A surrounding neighborhood region with radius r_c is considered to assess the impact of mutation on lysis timing. (C) Experimental lysis time data obtained from single-cell experiments[30] for various holin mutations, including its wild-type (WT) variant. Mean lysis times are indicated by red lines within box plots for each mutation. Each blue point represents the lysis time from an individual single-cell experiment.

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129−7140

These observations allow us to develop a quantitative scale based on the local hydrophobicity values to predict lysis times for newly generated mutants that have not been yet studied in experiments. These findings not only clarify the molecular details of cell lysis phenomena but also present potential avenues for developing optimized phage therapies with tailored lysis characteristics. It is also interesting to note that our approach, to a some degree, might be related to extensive studies of T4 lysozyme in which physicochemical properties of various mutants have been investigated.[31]

METHODS

Experimental Data on Cell Lysis Dynamics. The main idea of our theoretical approach is that there are correlations between the chemical composition of holin proteins and quantitative features of cell lysis dynamics. To understand such a relationship, we employed a statistical approach that connects holin sequence information with experimental data on cell lysis dynamics.

Let us first recall the experimental data on cell lysis. Figure 2A shows the primary sequence of the wild-type (WT) holin protein. While the structure of holin has not been yet determined experimentally, we utilized AlphaFold,[32] a state-of-the-art protein structure prediction tool, to generate the structural model of the holin protein based on its primary sequence, as illustrated in Figure 2B. This predicted structure shows that the holin protein has three transmembrane domains (Figure 2B), which was also confirmed in previous biochemical and genetic assays studies.[12] In a recent single-cell experiment, Kannoly et al.[30] introduced both single and double-point mutations along different positions in the holin protein and measured cell lysis dynamics. For each mutation, single-cell lysis events were recorded for approximately 90−175 cells per strain. Figure 2C displays experimental lysis time data for WT holin and different single-point mutated holin species. The data reveal that each mutation resulted in a distribution of lysis times across the mutant species, ranging from 15 to 175 min. It is important to note here that while Kannoly et al.[30] introduced both single and double-point mutations, we specifically focused on analyzing the effects of only single-point mutations on cell lysis dynamics in order to consider the systems with minimal chemical deviations from the original WT species.

Extraction of Physicochemical Features for Holin Mutants. To characterize the impact of different holin mutations, we extracted the physicochemical features from the corresponding amino acid sequences using a *propy* package.[33] This package is a comprehensive bioinformatics tool designed to extract a wide range of physicochemical features from amino acid sequences. It builds upon the foundational work of K. C. Chou, who made significant contributions to the field of protein research, particularly in the development of various sequence-based feature extraction methods.[34,35] The package incorporates 1547 physicochemical descriptors including but not limited to amino acid composition, dipeptide composition, pseudo amino acid composition, autocorrelations, physicochemical composition, transition, and distribution (see Table S1). The validity and reliability of the *propy* package have been demonstrated in multiple studies. For instance, it has been successfully employed in investigating antimicrobial peptides and their interactions with bacterial membranes.[36−38] These studies utilized the extensive set of features provided by *propy* to develop predictive models and analyze the physicochemical basis of peptide function, demonstrating the tool's utility and robustness. In addition to these applications, control studies have confirmed the accuracy of the feature extraction process by comparing the output of *propy* with other established bioinformatics tools and databases.[33] It is also important to note here that the extracted physicochemical properties generally are not independent from each other, and there is a redundancy in the reported features.

Fold-Change Analysis of Physicochemical Features. Mutation of a residue X at position n to a new residue Y is typically represented as XnY. The impact of a point mutation on protein stability and function can vary depending on the mutation's type $(X \rightarrow Y)$ and location (n).[39,40] A single mutation has only a minor effect on physicochemical properties such as amino acid and dipeptide compositions, while it can probably influence stronger other features like autocorrelations and pseudo amino acid composition, which rely on the specific positions of these amino acids in the protein sequence. Moreover, given that the holin sequence is relatively long (105 amino acids), a single mutation often does not significantly affect the features extracted using *propy* considering the entire protein sequence. To address this challenge and more accurately characterize the impact of each mutation, we analyzed only a localized region around each mutation, specifically a neighborhood defined by r_c amino acids to the left and r_c amino acids to the right within the holin sequence (see Figure 2B). This focused approach helps to isolate and assess the direct effects of mutations within a specific context.

By zooming in on a subsequence of $2r_c + 1$ amino acids instead of the whole sequence, we can discern subtle changes in physicochemical properties that may be crucial for understanding alterations in holin functions and cell lysis dynamics. To quantify the impact of mutations, we employ fold-change (FC) analysis that compares the properties of mutant holins to those of the WT holin within the neighborhood region surrounding a specific mutation. The fold-change is defined as the ratio of the property value in the mutant to that in the WT holin, given by

$$\text{Fold-Change (FC)} = \frac{P_{mutant}}{P_{WT}} \tag{1}$$

where P_{mutant} is the value of the physicochemical property being analyzed in the mutant holin and P_{WT} is the value of the same physicochemical property in the wild-type holin. It is important to note that we define fold-change for those *propy* features that are nonzero for wild-type holin, to avoid division by zero in eq 1. This expression quantifies the relative change in the physicochemical property induced by mutations compared to the WT holin. A fold-change greater than 1 indicates an increase in the property value in mutants compared to the WT, while a fold-change less than 1 indicates a decrease in the property value, thereby providing a quantitative measure of the effect of mutations on holin structure and function.

RESULTS AND DISCUSSION

Correlation Analysis of Physicochemical Features and Cell Lysis Dynamics for Holin Mutants. The experimental data in Figure 2C show some surprising observations in cell lysis dynamics for different single-point

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129−7140

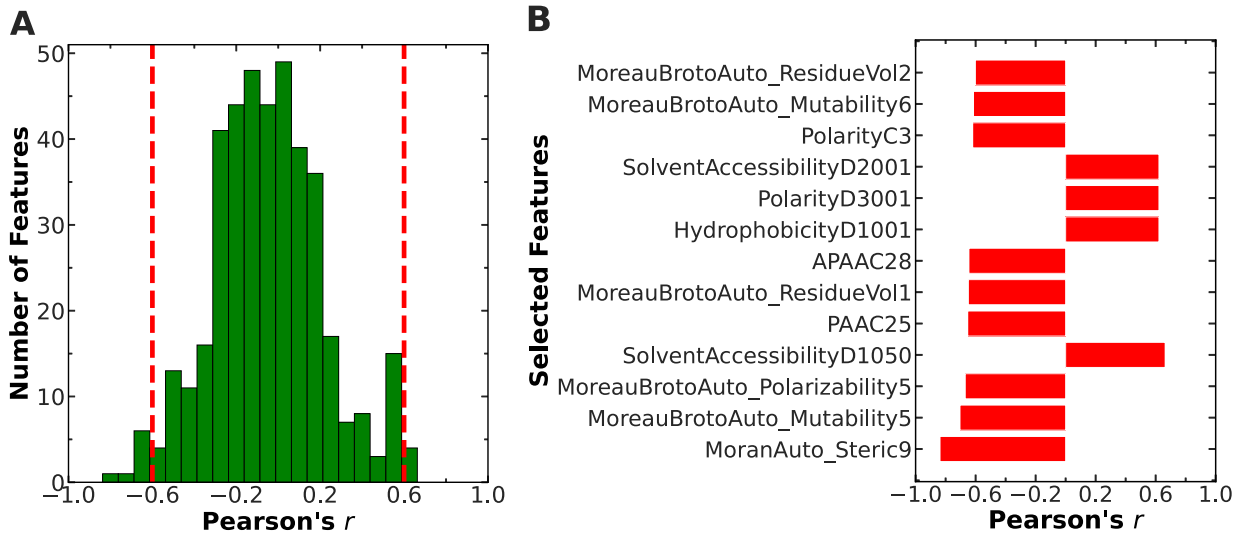


Figure 3. (A) Distribution of Pearson correlation coefficients between fold-change of physicochemical features of holin mutants and their corresponding mean lysis times. Vertical red dashed lines indicate threshold values of Pearson's correlations ($|r| = 0.6$) which are used as criteria for choosing the most relevant physicochemical properties. (B) Features selected above the threshold values are displayed along with their respective correlation coefficients. These results are based on an r_c value of 7.

mutations in holin proteins. For instance, some mutations may lead to faster or slower lysis times compared to the wild-type holin (see also cell lysis times histogram in Figure S1). These results suggest that specific amino acid substitutions within the holin protein can significantly impact its function and consequently the timing of cell lysis. Our main goal is to try to identify specific physicochemical features of holin proteins that influence most the cell lysis dynamics. To obtain the relationship between the physicochemical properties of different holin mutants and corresponding cell lysis times, we employ a statistical analysis based on evaluating the Pearson correlation coefficient between the mean lysis times $[T(m)]$ of a mutant m and the corresponding fold-change of a specific physicochemical feature (as a result of the mutation), $\mathrm{FC}(m)$,

$$r = \frac{\langle T(m)\mathrm{FC}(m)\rangle - \langle T(m)\rangle\langle \mathrm{FC}(m)\rangle}{\sqrt{\langle T^2(m)\rangle - \langle T(m)\rangle^2}\sqrt{\langle FC^2(m)\rangle - \langle \mathrm{FC}(m)\rangle^2}} \tag{2}$$

It is important to note that there are distributions of cell lysis times for each holin mutant (as shown in Figure S1 and Figure 2C), and we consider $T(m)$ as the mean value of the corresponding distribution for the mutation m. This analysis allows us to identify features that exhibit strong correlations with mean lysis times, providing valuable insights into the factors that lead to the variation in cell lysis dynamics due to these mutations. We analyzed the Pearson correlation in terms of the physicochemical features whose values are nonzero for the corresponding wild-type counterpart of each mutant. Here, we choose the value of r_c to be 7, thereby analyzing a total sequence length of $2r_c + 1 = 15$. It is important to note that the specific choice of the r_c value does not affect our main conclusions, as we explicitly checked in our analysis (see the Supporting Information).

The results of our correlation analysis are presented in Figure 3. More specifically, Figure 3A, exhibits the distribution of Pearson correlation coefficients (r) calculated between the FC of physicochemical features and cell lysis times. To identify features with significant correlations, we applied a cutoff of $|r| > 0.6$ (see red dotted lines in Figure 3A). Additionally, a statistical analysis with a p-value of less than 0.05 was applied to filter out irrelevant features. These combined criteria ensured that only those physicochemical properties with the strongest correlation with cell lysis dynamics are retained. Subsequently, Figure 3B illustrates the selected features plotted against their respective correlation values, exhibiting a clear visualization of the relationships observed.

Detailed Analysis of Selected Features: Autocorrelation Functions. Our statistical analysis identified several physicochemical features that exhibited strong correlations with cell lysis dynamic properties. Among those selected features, there are autocorrelation functions that characterize the spatial variation of certain physicochemical features along amino acid sequences. In *propy* package,[33] several autocorrelation functions including Moreau–Broto, Moran, and Geary have been utilized for characterizing protein sequences (see Table S1). The Moreau–Broto autocorrelation function, which measures the correlations between physicochemical properties of the residue i and the residue $i + d$ (along the contour within the neighborhood region surrounding m) is given by

$$\mathrm{MB}(d) = \frac{1}{N-d}\sum_{i=1}^{N-d}\left(\frac{P_i - \mu_P}{\sigma_P}\right)\left(\frac{P_{i+d} - \mu_P}{\sigma_P}\right) \tag{3}$$

where P_i and P_{i+d} are physicochemical properties of residue i and residue $i + d$, respectively, $d = 1, 2, ..., 30$ is the lag of the autocorrelation and $N = 2r_c + 1$. Additionally, the mean $\mu_P = \frac{1}{20}\sum_{j=1}^{20} P_j$ and the variance $\sigma_P = \left(\frac{1}{20}\sum_{j=1}^{20}(P_j - \mu_P)^2\right)^{1/2}$ of a physicochemical property P are averaged over the 20 types of amino acids. Thus, the Moreau–Broto autocorrelation function can be rewritten as,[41]

$$\mathrm{MB}(d) = \frac{1}{N-d}\sum_{i=1}^{N-d} Z_i Z_{i+d} \tag{4}$$

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129–7140

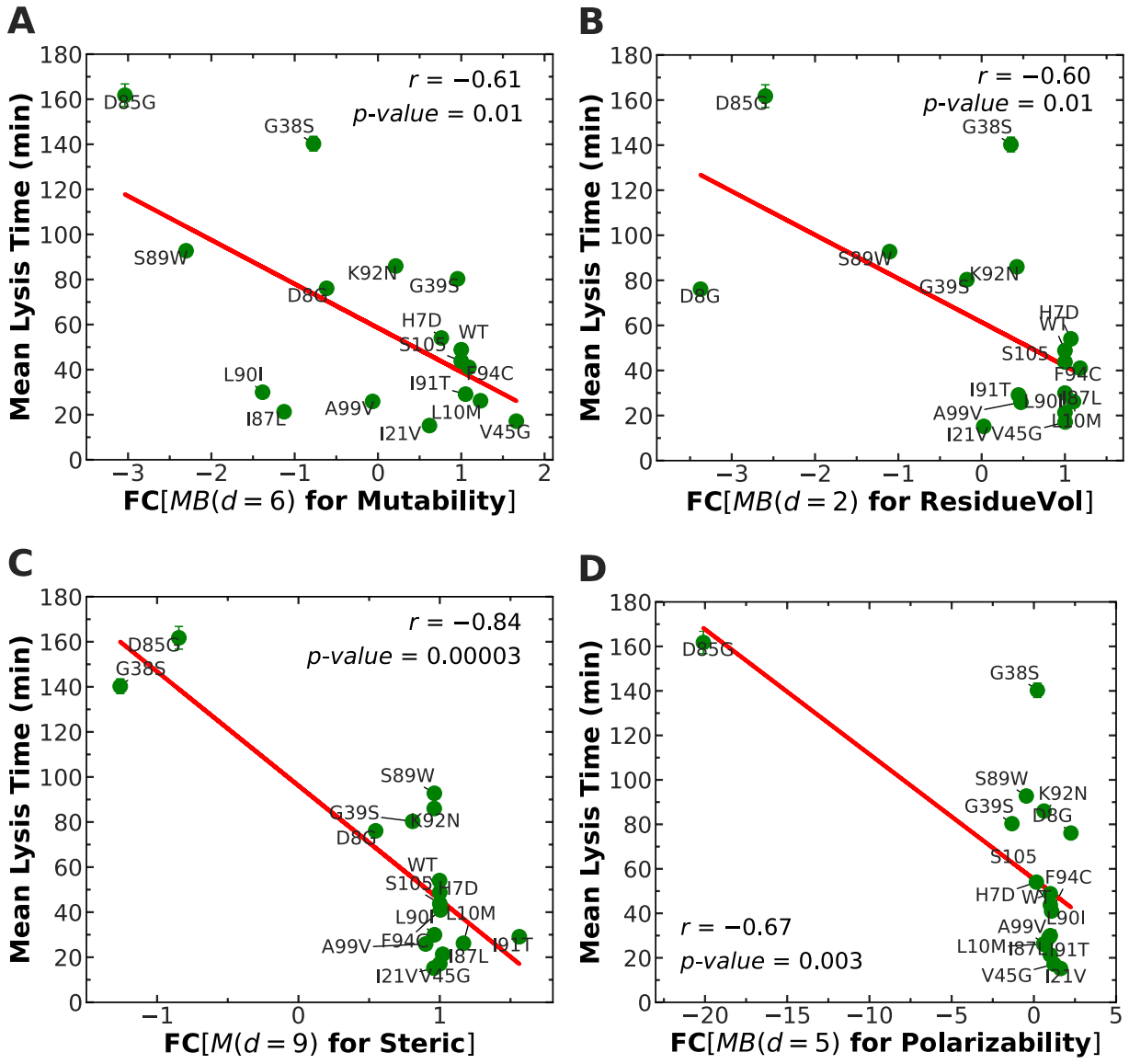


Figure 4. Mean cell lysis times for different holin mutants as a function of (A) fold-change of the Moreau−Broto autocorrelation of amino acid mutability with a lag of 6. This feature is represented as MoreauBrotoAuto_Mutability6 in *propy*[33] (see also Figure 3B). (B) Fold change of the Moreau−Broto autocorrelation of residue volume with a lag of 2, (C) fold-change of the Moran autocorrelation of steric hindrance with a lag of 9 and (D) fold-change of the Moreau−Broto autocorrelation of polarizability with a lag of 5. The linear regression fits are shown by the red lines. The parameter r represents Pearson's correlation coefficient with the associated statistical p-value.

where $Z_i = \frac{P_i - \mu_P}{\sigma_P}$ is defined as the corresponding z-score of the property P_i.

Alternatively, one can define a Moran autocorrelation function,[41] which is similar to the Pearson's correlation between the physicochemical property of a residue i and $i + d$,

$$M(d) = \frac{\frac{1}{N-d}\sum_{i=1}^{N-d}(Z_i - \overline{P})(Z_{i+d} - \overline{P})}{\frac{1}{N}\sum_{i=1}^{N}(Z_i - \overline{P})^2} \quad (5)$$

where $\overline{P} = \frac{1}{N}\sum_{i=1}^{N} Z_i$ is the average of the z-scores of the amino acid properties in the protein sequence. In the context of the Moran autocorrelation calculation, using z-scores effectively means that amino acids whose properties are close to the average property value across all amino acids (after normalization) contribute minimally to the autocorrelation function. This effectively filters out the "background noise" of average properties and highlights the impact of exceptional or distinctive properties, which can be particularly valuable in understanding protein function and structure. These two functions allow us to quantify the spatial distribution of physicochemical properties along the protein sequence, considering both local and long-range interactions. Since these methods involve centering the property P values by subtracting the mean, the resultant autocorrelation values can be both positive and negative.

Our feature selection procedure based on statistical analysis identifies autocorrelations of several physicochemical properties that are highly correlated with cell lysis times. These include the Moreau−Broto (MB) autocorrelation function for mutability, residue volume, and polarizability as well as the Moran autocorrelation function for steric hindrance (see Figure 4). Let us try to understand the physical meaning of these results. The mutability of residues in proteins

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129−7140

corresponds to the likelihood or rate at which the amino acid residues in a protein sequence change over time due to genetic mutations.[42] A negative autocorrelation mutability at the distance $d = 6$ means that the adjacent residues in the sequence (separated by five other residues) tend not to follow the same mutability patterns (Figure 4A). Likewise, a positive autocorrelation for mutability at the distance $d = 6$ could mean that residues spaced six positions apart in the protein sequence tend to exhibit similar trends in mutability (Figure 4A). This suggests that if one residue is likely to mutate (high mutability), the residue six positions away will also likely do so, and vice versa.

A fold-change for a specific holin mutant in terms of the Moreau−Broto autocorrelation function reads

$$\mathrm{FC}[\mathrm{MB}(d) \text{ for mutant } XnY] = \frac{\mathrm{MB}(d) \text{ for mutated subsequence of length } 2r_c + 1 \text{ including residue } Y}{\mathrm{MB}(d) \text{ for WT subsequence of length } 2r_c + 1 \text{ including residue } X} \quad (6)$$

This definition indicates that if a specific mutation XnY does not change the Moreau−Broto autocorrelations, then the corresponding fold-change of the holin mutant XnY is close to one. If a single mutation changes the sign of the autocorrelation function from negative to positive, this implies that the mutation has introduced a change that aligns the mutability patterns of residues that are d positions apart in the protein sequence. This alignment could mean that previously differing mutability rates at these positions have become similar, potentially stabilizing the biochemical characteristics of this segment of the protein.

The results for mean cell lysis times versus corresponding fold-change in the Moreau−Broto autocorrelation function for mutability are presented in Figure 4A. One can see that the fold-change values range from −3 to 2, indicating a significant variation in the impact of single mutations in the holin sequence. We obtained the Pearson correlation coefficient of −0.61 between mean lysis timing and fold-change for autocorrelation in mutability. The negative correlation coefficient indicates that those single mutations that make the autocorrelations function of mutability negative are more likely to make cell lysis dynamics slower. Likewise, the mutations that do not change the sign of autocorrelations function, while making correlations stronger, lead to faster cell lysis dynamics.

It is important to note that for a single mutation XnY, both the position n and the type of the new residue Y play critical roles in the cell lysis activity of the resulting holin mutant. The position of the mutated residue is particularly important for the holin protein, as it contains three transmembrane regions that contribute strongly to its interactions with the cell membrane (see Figure 2). The significance of the chemical nature of new residue Y lies in the fact that each amino acid has a unique mutability index (see Table S2); some residues are highly mutable, while others are rarely mutable.[42] In Figure S2, we present the corresponding z-scores of the mutability index for all amino acids. This analysis helps us to identify specific residues with extreme mutability values. For example, in the holin mutant $D85G$, which is associated with a mean cell lysis times of ∼162 min, the amino acid glycine (G) has a z-score of −0.75, while the aspartic acid (D) has a z-score of 1, meaning that G is rarely mutable while D is frequently mutable. We also examined the corresponding z-scores for other mutants and concluded that for all mutants XnY with cell lysis times below or above those of the wild-type (WT) holins, $Z_X Z_Y < 0$, which means that either a highly mutable residue is replaced by a rarely mutable residue, or vice versa. This is a crucial observation, especially considering the vast number of potential mutations. It might serve as a guideline for engineering mutations that optimize cell lysis dynamics.

The results for mean cell lysis times versus fold-change in the autocorrelation of two other physicochemical features, namely, residue volume and steric are given in Figures 4B and 4C. The residue volume is simply defined as the volume occupied by an amino acid residue in a protein structure,[43−45] whereas steric effects, also known as steric hindrance,[46] refers to the influence of the size, shape, and spatial arrangement of amino acid residues that modify the interactions with neighboring residues and other molecules. These effects play a significant role in shaping the overall structure, stability, and functionality of the proteins.[47] Alterations in these two features can profoundly influence the ability of holin proteins to interact with the cell membrane and govern the dynamics of bacterial cell lysis.

A positive autocorrelation in residue volume at the distance $d = 2$ means that residues separated by one residue tend to either both have large volumes or both have small volumes. On the other hand, a negative autocorrelation at this distance suggests that a residue with a small volume is typically spaced one position away from a residue with a large volume and vice versa. Consequently, introducing a single mutation XnY depending on the relative volumes of residue X and Y could change the autocorrelation sign and magnitude. As one can see in Figure 4B, for most of the single mutations FC in autocorrelation of residue volume remains close to one.

Both these quantities exhibit negative correlations between their fold-change of autocorrelation values and cell lysis times. The fold-change values range from ∼−4 to ∼2 for residue volume and range from ∼−1.5 to ∼1.5 for steric hindrance (Figures 4B and 4C). The negative fold-change values indicate a reduction in the overall volume occupied by amino acid residues and a decrease in the steric hindrance experienced by these residues within the holin protein compared to the wild-type species. These alterations arise from the mutations resulting in substitutions with smaller amino acid residues or less bulky side chains. On the contrary, the positive fold-change values suggest the increase in the residue volume or steric hindrance, potentially due to the substitutions with larger amino acid residues or bulkier side chains. The corresponding cell lysis dynamics data reveal that the holin variants exhibiting negative fold-change values generally result in prolonged lysis times compared to the wild-type holins. This suggests that mutations leading to a decrease in residue volume or in the steric hindrance (e.g., holin mutations "D85G", "G38S") impede holin's efficiency in interacting with the cell membrane, consequently delaying the process of cell lysis. It is also possible that smaller residues decrease the ability for holin proteins to aggregate inside the membrane, which is the critical step in breaking the membrane. Conversely, the holin variants with positive fold-change values (such as mutations like "L10M", "I21V", "L90I") tend to exhibit faster cell lysis times.

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129−7140

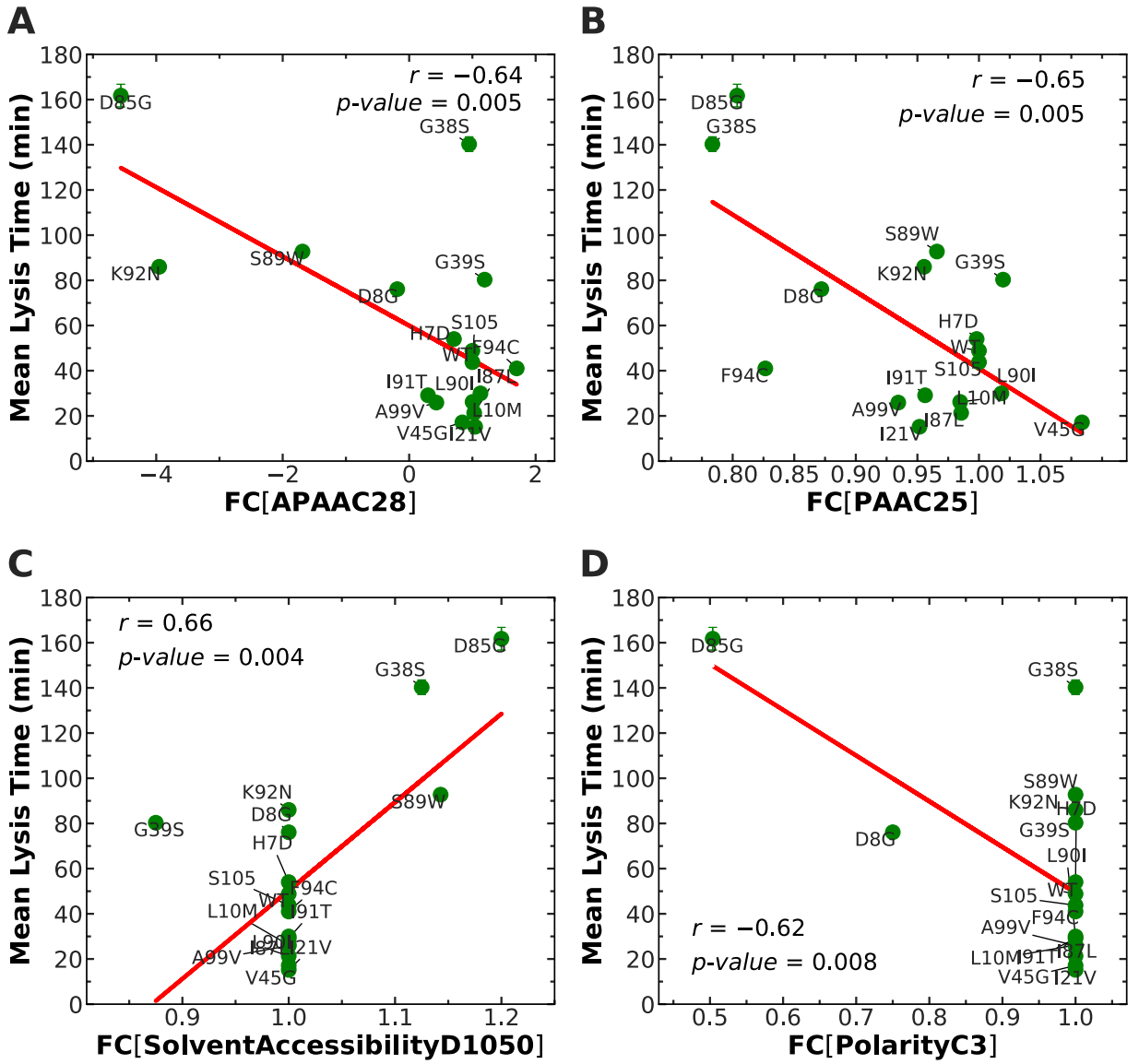


Figure 5. Mean cell lysis times for different holin mutants as a function of (A) fold-change of amphiphilic pseudoamino acid composition (APAAC), (B) fold-change of pseudoamino acid composition (PAAC), (C) fold-change of solvent accessibility and (D) fold-change of polarity. The linear regression fits are shown by the red lines. The parameter *r* represents the Pearson's correlation coefficient with the associated statistical *p*-value.

This acceleration in lysis can be attributed to enhanced membrane interactions facilitated by substitutions with larger amino acid residues or bulkier side chains.

Additionally, the analysis of correlations between the mean volume and steric hindrance of residues with the mean cell lysis times for different holin mutations, as depicted in Supporting Figures S3A and S3B, further supports this relationship. It underscores how variations in residue volume and steric hindrance directly impact the efficiency of holin proteins in inducing cell lysis, thus highlighting the intricate interplay between protein structure, physicochemical properties, and lysis dynamics. Likewise, another physicochemical feature, polarizability,[48] shows a similar trend with the mean lysis timing (see Figure 4D and Figure S3C).

Detailed Analysis of Selected Features: Pseudo-Amino Acid and Physicochemical Composition. One could also see from Figure 3B that both amphiphilic pseudoamino acid composition (APAAC)[49] and pseudoamino acid composition (PAAC)[34] correlate with the cell lysis dynamics. APAAC characterizes the amphiphilic nature of amino acids, which is crucial for interactions with membranes, while PAAC assesses the overall amino acid composition and local sequence order information. Analysis of the fold-change of these physicochemical properties reveals negative correlations with cell lysis times (see Figures 5A and 5B). The APAAC analysis indicates that negative fold-change values correspond to a decrease in the amphiphilicity compared to the WT holin (Figure 5A). This suggests alterations in the distribution of hydrophobic and hydrophilic amino acids within the holin protein, which potentially weakens the interactions with the cell membrane and delays cell lysis compared to the WT holins. Conversely, positive fold-change values indicate increased amphiphilicity compared to the WT holins, enhancing the interactions between the protein and the lipid bilayer in the cell membrane. This enhanced interaction probably facilitates more efficient membrane disruption,

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129−7140

thereby expediting the cell lysis process and resulting in accelerated cell lysis dynamics compared to the WT holins.

Similarly, PAAC analysis (Figure 5B) indicates the reduced composition of certain amino acids compared to the WT holins for fold-change values less than 1. This means that the corresponding mutations impact the structural stability, functional dynamics, and interactions of the holin proteins, resulting in delayed lysis times compared to the wild-type species. On the contrary, FC > 1 in PAAC suggests an enrichment of certain amino acids within the holin mutants. This enrichment facilitates improved interactions with cellular components, expediting the process of cell lysis compared to wild-type proteins.

Our theoretical analysis also identified other relevant physicochemical properties that are related to the distribution of particular physicochemical properties along the holin sequence (see Figures 5C and 5D and Figure S4). These features include solvent accessibility, polarity and hydrophobicity,[50] each of which is coarse grained into three categories listed in Table 2 in ref 50. The composition descriptor, *C*, measures the fraction of amino acids in the holin sequence that fall into each of the three categories. The distribution descriptor (*Di*) for each group (*i*) are defined in the following way. The fractional distance along the neighborhood sequence surrounding the mutation point that accommodates the first residue, 25%, 50%, 75% and 100% of the residues belonging to the group *i* (*i* = 1, 2, 3) are denoted as *Di*001, *Di*025, *Di*050, *Di*075, *Di*100, respectively (see Table S3 for description of the selected features).

As one can see from Figure 5, the fold-change data for solvent accessibility pattern of the holin sequence has a positive correlation with the cell lysis dynamics for various holin mutants compared to the WT (Figure 5C, Figure S4C). Solvent accessibility refers here to the extent to which amino acid residues within a protein are exposed to the surrounding solvent environment. Changes in solvent accessibility can result from alterations in the protein's conformation, surface properties, and interactions with other molecules. The wild-type holin serves as the baseline with a fold-change of 1 in the solvent accessibility. Comparing this to the holin mutants (see Figure 5C), variations in both solvent accessibility and lysis times are observed. For instance, the mutant D85G exhibits a fold-change of 1.2 in the solvent accessibility, indicating an increase in the access to solvent molecules compared to the wild-type proteins and correspondingly its lysis time is significantly larger. Conversely, mutants like G39S exhibit a fold-change of 0.88 in solvent accessibility, suggesting a decrease in solvent accessibility compared to wild-type proteins. This reduction in solvent accessibility corresponds to a relatively shorter cell lysis time. These observations suggest a correlation between the solvent accessibility and cell lysis times in the holin mutants. Alterations in solvent accessibility may influence the efficiency of holin proteins in interacting with the cellular components, consequently impacting the dynamics of cell lysis.

To ensure the robustness of our analysis, we performed additional analysis with a larger subsequence window $r_c = 8$, and the corresponding results are presented in the Supporting Information (see Figures S5 and S6). We found that the conclusions on correlations are very similar to those that we already discussed above.

Local Hydrophobicity Plays a Critical Role in Determining Cell Lysis Times. While our theoretical analysis identified several physicochemical properties that most strongly correlate with cell lysis dynamics, it is important to analyze separately the effect of hydrophobicity since it provides a clearer physical picture of the holin proteins' ability to break the cellular membranes.

Hydrophobic interactions are fundamental to protein−membrane interactions and are known to play an important role in membrane disruption processes.[53−55] Recent experiments suggest that mutations alter the hydrophobicity of transmembrane domains of the holin proteins, impacting the externalization of transmembrane domains.[24] This means that the degree of hydrophobicity can markedly affect the efficiency and dynamics of the cell lysis process. To explore this, we analyzed the local mean hydrophobicity of the neighborhood sequence of length $2r_c + 1$ surrounding the mutations, as defined in eq 7. Our goal is to elucidate how local hydrophobicity influences the efficiency and speed of cell lysis.

$$\langle H \rangle = \frac{1}{2r_c + 1} \sum_{i=1}^{2r_c+1} H_i \tag{7}$$

For precise quantification of hydrophobicity, we employed the Wimley−White hydrophobicity scale[51] (see also Table S4), focusing on the amino acid residues within the holin protein sequence. The results, as depicted in Figure 6A, reveal diverse

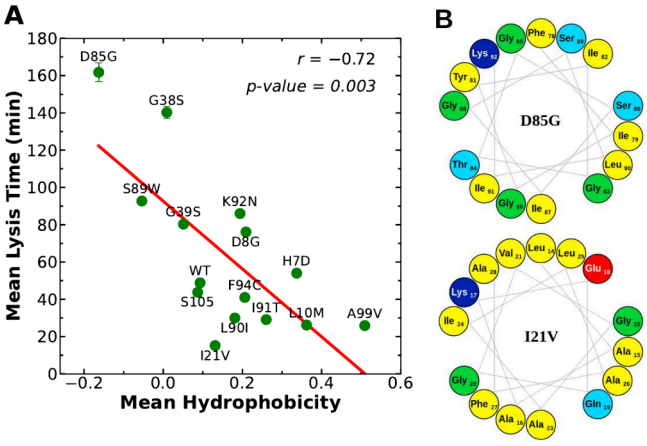


Figure 6. (A) Mean cell lysis times as a function of local mean hydrophobicity (using eq 7) for various single-point mutants.[30] The mean hydrophobicity is calculated according to the Wimley−White hydrophobicity scale[51] for the neighborhood region of radius r_c around the mutation point. The calculations assume $r_c = 7$. The linear regression fit is shown by the red line. The parameter *r* represents Pearson's correlation coefficient with the associated statistical *p*-value. (B) The helical wheel diagram[52] for amino acid sequences around the mutation point for mutations "D85G" and "I21V". Residues are color-coded: yellow - hydrophobic; light blue - polar; dark blue - cationic; red - anionic; green - Gly/Pro.

changes in the mean hydrophobicity values (ranging from −0.2 to 0.6) across different experimentally studied mutations.[30] Notably, we observe a strong inverse correlation between the mean hydrophobicity of holin mutants and their mean cell lysis times (Figure 6A). This emphasizes the important role of hydrophobicity in cell lysis. Holin mutants with higher hydrophobicity tend to exhibit shorter lysis times, while mutants with lower hydrophobicity exhibit longer lysis times. For instance, the mutations I21V, A99V, and L10M, characterized by relatively high hydrophobicity values,

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129−7140

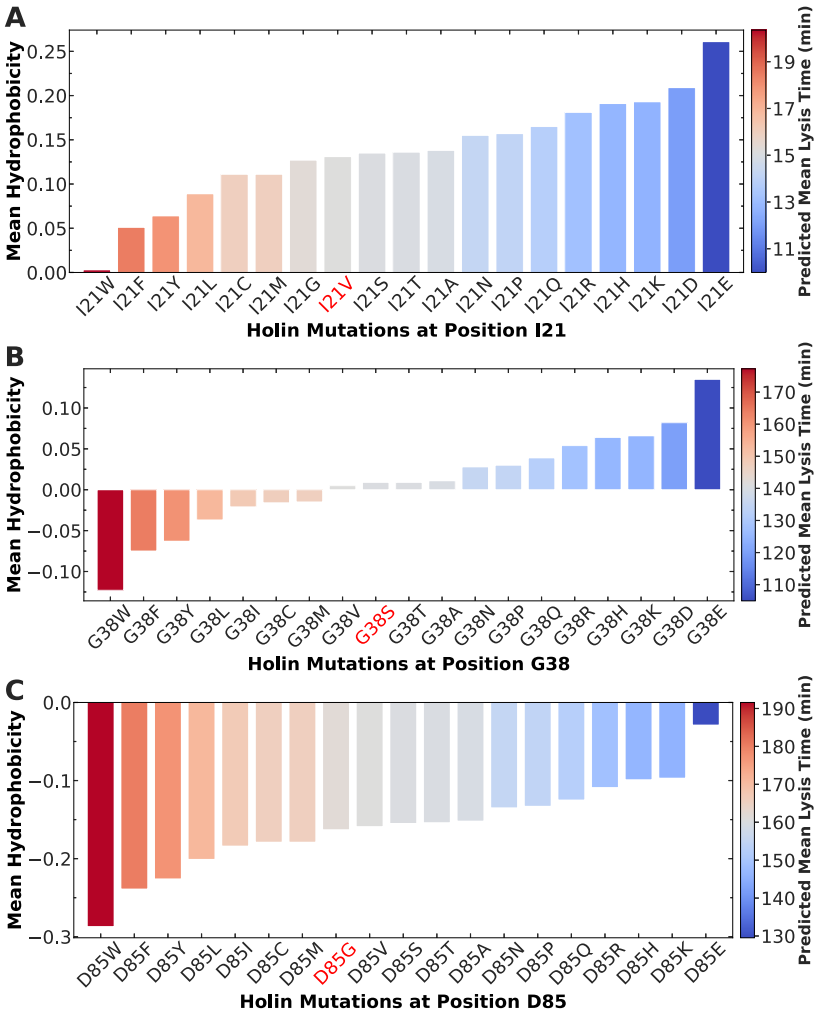


Figure 7. 2D contour plot of predicted mean cell lysis times based on the mean hydrophobicity values for all possible amino acid substitutions at the position (A) I21, (B) G38, and (C) D85. The mean lysis times for the red colored mutations I21V, G38S and D85G were originally determined by Kannoly et al.[30] The mean hydrophobicity is calculated according to the Wimley–White hydrophobicity scale[51] for the neighborhood region of radius r_c = 7 around the mutation point.

demonstrated significantly shorter mean lysis times compared to the wild-type holins. Conversely, the mutations with lower hydrophobicity, such as D85G and S89W, exhibited longer lysis times than the WT variant.

These observations can be explained with the following arguments. Mutations with negative hydrophobicity values typically include more hydrophilic or polar residues (to better visualize this see the helical wheel diagram in Figure 6B for the mutation "D85G"), indicating a stronger affinity for the aqueous cytoplasmic environment rather than the hydrophobic lipid bilayer of the cell membrane. Consequently, such mutations exhibit reduced membrane affinity that eventually leads to slower or less efficient membrane disruption and cell lysis. On the other hand, mutations with positive hydrophobicity values tend to be more hydrophobic (see Figure 6B for the mutation "I21V"), showing a preference for interaction with hydrophobic regions of the cell membrane. In addition, it can be argued that proteins with larger hydrophobicity might interact strongly with each other, stimulating the protein aggregation that should help to break the cellular membranes more easily. Thus, mutations with higher hydrophobicity values exhibit enhanced membrane affinity, facilitating faster membrane disruption and more efficient cell lysis.

In addition to analyzing the observed relationship between mean hydrophobicity and lysis times based on the holin mutations introduced by Kannoly et al.,[30] we extended our analysis to include a completely separate set of experimental cell lysis times data for various other holin mutations performed by To et al.[21] It is worth noting that the experimental conditions in the study by To et al.[21] were entirely different from those in the study by Kannoly et al.,[30] indicating variability across the data sets. Despite this variation, our analysis again revealed a strong correlation between mean hydrophobicity and cell lysis times for the mutations introduced by To et al.[21] (see Figure S7). This result further supports the notion that local hydrophobicity plays a crucial role in determining the efficiency and dynamics of cell lysis, regardless of experimental conditions. The correlation coefficient obtained for this additional data set is similar to that obtained from the Kannoly et al.[30] data set, indicating the robustness of our findings across different experimental setups. Furthermore, to additionally assess the robustness of our analysis, we also investigated the relationship between the mean hydrophobicity and the cell lysis times using a larger neighborhood sequence radius (r_c = 14) on both data sets (see Figure S8). Remarkably, even with this increased radius, the

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129–7140

correlation between mean hydrophobicity and lysis times remained strong, albeit with a slightly reduced correlation coefficient. This observation underscores the consistency of our theoretical analysis and the universality of correlations between the hydrophobicity and cell lysis dynamics.

Prediction of Cell Lysis Times for Arbitrary Holin Mutants Based on Local Hydrophobicity. The advantage of our theoretical method is that not only it can quantitatively analyze the correlations between the physicochemical properties of holin proteins and their ability to break cellular membranes, but it can also make explicit predictions about cell lysis dynamics for any holin mutants, even for those that have not been studied yet. This clearly provides an important tool to test the validity of our theoretical approach.

To illustrate this, we proceed to predict the mean cell lysis times for newly generated holin mutants based on their mean hydrophobicity ($\langle H \rangle$) values. From the linear regression fitting between mean hydrophobicity and mean lysis times depicted in Figure 6A, one can evaluate the slope ($a = -181.8$) and intercept ($b = 92.6$) parameters. Then, assuming that correlations between the hydrophobicity and cell lysis times are universal, the predicted times can be estimated from

$$\langle T_{predicted} \rangle = a\langle H \rangle + b \tag{8}$$

More specifically, we selected three different positions along the protein sequence, namely, I21, G38, and D85, each corresponding to one of the three transmembrane domains (see the protein structure in Figure 2). Notably, for the mutations I21V, G38S, and D85G, Kannoly et al.[30] had originally determined the cell lysis times experimentally. However, our predictive model allows us to estimate cell lysis times for all possible amino acid substitutions at these positions, and the corresponding results are presented in Figure 7. It quantifies how different mutations at these locations are affecting the cell lysis dynamics. These mutations yield varying effects on the mean local hydrophobicity, even within the same mutation type but at different positions. For example, the mutation I21G results in a positive mean hydrophobicity value (Figure 7A), whereas for the mutation D85G, it is negative (Figure 7C). This positional specificity suggests that certain regions of the holin sequence may be more sensitive to mutations affecting hydrophobicity and thus depending on the nature and location of the mutation, the dynamics of cell lysis can be either accelerated or delayed. This phenomenon likely stems from the mutation's proximity to critical functional domains involved in the membrane interactions and rupture. While our analysis demonstrates the importance of hydrophobicity in determining the kinetics of cell lysis in holin mutants, our model also allows us to design new holin mutants with tailored lysis kinetics that could accelerate the development of novel antimicrobial strategies targeting bacterial lysis.

CONCLUSIONS

In conclusion, our theoretical study determines the correlations between certain physicochemical properties of holin proteins and the corresponding cell lysis dynamics, clarifying some aspects of the underlying mechanisms governing bacterial membrane disruption. Through correlation statistical analyses, we identified specific physicochemical features, such as mutability, residue volume, steric hindrance, amphiphilicity, solvent accessibility, polarity, polarizability, and hydrophobicity, that significantly impact the efficiency and kinetics of cell lysis. Based on these results, it is argued that the critical role of these properties is in modulating the interactions between holin proteins and the cell membrane, ultimately influencing the dynamics of bacterial cell lysis.

Our analysis emphasizes the important role of hydrophobicity in governing the efficiency and speed of cell lysis. Strong inverse correlation between the local hydrophobicity and cell lysis times is observed, emphasizing the crucial role of hydrophobic interactions in membrane disruption processes. By leveraging this relationship, we developed a quantitative predictive model capable of explicitly estimating cell lysis times for holin mutants that have not been yet experimentally studied. This predictive model not only deepens our understanding of the underlying principles governing bacterial lysis but also offers practical applications in the design of novel antimicrobial strategies targeting bacterial membranes.

It is also important to discuss the limitations of the current study. First, our analysis was performed on very limited sets of data where fluctuations and systematic errors might significantly affect the conclusions. However, given the fact that our analysis produces the same results for experimental observations done under very different conditions, it suggests that our approach is quite robust. Due to the enormous number of possible mutations for holin, it will be critical to test our predictions with more cell lysis data for other holin mutants. Second, the Pearson correlation method can be heavily influenced by outliers. A single outlier can significantly affect the value of the Pearson coefficient, misleading the interpretation of the linear relationship. There are other methods to quantify the correlations, and we plan to explore them in the future. Furthermore, it will be important to investigate this system using computer simulations with atomic resolution. In addition, there are large error bars in the experimental measurements of cell lysis dynamics that might also complicate the analysis. However, despite these issues, our theoretical method provides a simple quantitative method of identifying the physicochemical properties that are the most relevant for understanding these complex biological processes. Importantly, our study not only clarifies many aspects of the molecular mechanisms of holin-mediated cell lysis, but also provides possible directions for engineering advanced medical therapies with tailored cell lysis kinetics.

ASSOCIATED CONTENT

Data Availability Statement

The data obtained in this work and the in-house scripts are available on GitHub at the following URL: https://github.com/anupam-rice/JPCB_data_cell_lysis_dynamics.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.4c03040.

Description of feature extraction by *propy* package; values of mutability, residue volume, steric hindrance and polarizability for the 20 amino acid residue; description of selected features from the correlation analysis; whole residue hydrophobicity scale for the 20 natural amino acids; distribution of experimental lysis time data for WT holin and two other holin mutations; *z*-score of amino acid residues for mutability, residue volume, steric hindrance and polarizability; mean lysis time for different holin mutations as a function of the

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129−7140

residue volume, steric hindrance and polarizability of the residue *Y* in the mutation *XnY*; mean lysis times for different holin mutations as a function of the fold change of different physicochemical features for r_c = 7; distribution of Pearson correlation coefficients between fold change of physicochemical features of holin mutants and their corresponding lysis timing and also the selected features for r_c = 8; mean lysis times for different holin mutations as a function of the fold change of different physicochemical features for r_c = 8; mean lysis time as a function of mean local hydrophobicity for various single-point mutations introduced by To et al. for r_c = 10; mean lysis time as a function of mean local hydrophobicity for various single-point mutations introduced by Kannoly et al. and To et al. for r_c = 14 (PDF)

AUTHOR INFORMATION

Corresponding Author

Anatoly B. Kolomeisky − *Center for Theoretical Biological Physics, Department of Chemistry, and Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-5677-6690; Email: tolya@rice.edu

Authors

Anupam Mondal − *Center for Theoretical Biological Physics and Department of Chemistry, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-8436-5618

Hamid Teimouri − *Center for Theoretical Biological Physics and Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.4c03040

Author Contributions

A.M., H.T. and A.B.K. designed the research. A.M. and H.T. performed the research. A.M., H.T. and A.B.K. wrote the article. All authors reviewed the article.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (C-1559), by the NIH (R01GM148537 and R01HL157714-04), by the NSF (CHE-2246878), and by the Center for Theoretical Biological Physics sponsored by the NSF (PHY-2019745).

REFERENCES

(1) O'Neill, J. Tackling drug-resistant infections globally: final report and recommendations. 2016.

(2) Ye, M.; Sun, M.; Huang, D.; Zhang, Z.; Zhang, H.; Zhang, S.; Hu, F.; Jiang, X.; Jiao, W. A review of bacteriophage therapy for pathogenic bacteria inactivation in the soil environment. *Environ. Int.* **2019**, *129*, 488−496.

(3) Lin, D. M.; Koskella, B.; Lin, H. C. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World J. Gastrointest Pharmacol Ther* **2017**, *8*, 162−173.

(4) Principi, N.; Silvestri, E.; Esposito, S. Advantages and limitations of bacteriophages for the treatment of bacterial infections. *Front Pharmacol* **2019**, *10*, 513.

(5) Perros, M. Infectious disease. A sustainable model for antibiotics. *Science* **2015**, *347*, 1062−1064.

(6) Kakasis, A.; Panitsa, G. Bacteriophage therapy as an alternative treatment for human infections. A comprehensive review. *Int. J. Antimicrob. Agents* **2019**, *53*, 16−21.

(7) Young, R. Bacteriophage lysis: mechanism and regulation. *Microbiol Rev.* **1992**, *56*, 430−481.

(8) Golding, I. Infection by bacteriophage lambda: an evolving paradigm for cellular individuality. *Curr. Opin Microbiol* **2018**, *43*, 9−13.

(9) Wang, I. N.; Smith, D. L.; Young, R. Holins: the protein clocks of bacteriophage infections. *Annu. Rev. Microbiol.* **2000**, *54*, 799−825.

(10) Young, R. Bacteriophage holins: deadly diversity. *J. Mol. Microbiol Biotechnol* **2002**, *4*, 21−36.

(11) Young, R. Phage lysis: do we have the hole story yet? *Curr. Opin Microbiol* **2013**, *16*, 790−797.

(12) Grundling, A.; Blasi, U.; Young, R. Biochemical and genetic evidence for three transmembrane domains in the class I holin, λ S. *J. Biol. Chem.* **2000**, *275*, 769−776.

(13) Gründling, A.; Manson, M. D.; Young, R. Holins kill without warning. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 9348−9352.

(14) Dewey, J. S.; Savva, C. G.; White, R. L.; Vitha, S.; Holzenburg, A.; Young, R. Micron-scale holes terminate the phage infection cycle. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 2219−2223.

(15) Summer, E. J.; Berry, J.; Tran, T. A.; Niu, L.; Struck, D. K.; Young, R. Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. *J. Mol. Biol.* **2007**, *373*, 1098−1112.

(16) Berry, J.; Summer, E. J.; Struck, D. K.; Young, R. The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes. *Mol. Microbiol.* **2008**, *70*, 341−351.

(17) Young, R. Phage lysis: three steps, three choices, one outcome. *J. Microbiol* **2014**, *52*, 243−258.

(18) Cahill, J.; Holt, A.; Theodore, M.; Moreland, R.; O'Leary, C.; Martin, C.; Bettridge, K.; Xiao, J.; Young, R. Spatial and temporal control of lysis by the lambda holin. *mBio* **2024**, *15*, e0129023.

(19) Pang, T.; Park, T.; Young, R. Mutational analysis of the S21 pinholin. *Mol. Microbiol.* **2010**, *76*, 68−77.

(20) Dennehy, J. J.; Wang, I. N. Factors influencing lysis time stochasticity in bacteriophage λ. *BMC Microbiol* **2011**, *11*, 174.

(21) To, K. H.; Young, R. Probing the structure of the S105 hole. *J. Bacteriol.* **2014**, *196*, 3683−3689.

(22) Kannoly, S.; Singh, A.; Dennehy, J. J. An optimal lysis time maximizes bacteriophage fitness in quasi-continuous culture. *mBio* **2022**, *13*, e0359321.

(23) Raab, R.; Neal, G.; Garrett, J.; Grimaila, R.; Fusselman, R.; Young, R. Mutational analysis of bacteriophage lambda lysis gene S. *J. Bacteriol.* **1986**, *167*, 1035−1042.

(24) Ahammad, T.; Khan, R. H.; Sahu, I. D.; Drew, D. L.; Faul, E.; Li, T.; McCarrick, R. M.; Lorigan, G. A.; Pinholin, S. Pinholin S^{21} mutations induce structural topology and conformational changes. *Biochim Biophys Acta Biomembr* **2021**, *1863*, 183771.

(25) Marx, D. C.; Fleming, K. G. Local bilayer hydrophobicity modulates membrane protein stability. *J. Am. Chem. Soc.* **2021**, *143*, 764−772.

(26) Elazar, A.; Weinstein, J. J.; Prilusky, J.; Fleishman, S. J. Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 10340−10345.

(27) Pajtinka, P.; Vácha, R. Amphipathic helices can sense both positive and negative curvatures of lipid membranes. *J. Phys. Chem. Lett.* **2024**, *15*, 175−179.

(28) Shai, Y. Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by α-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochim. Biophys. Acta* **1999**, *1462*, 55−70.

(29) Ladokhin, A. S.; White, S. H. Folding of amphipathic α-helices on membranes: energetics of helix formation by melittin. *J. Mol. Biol.* **1999**, *285*, 1363−1369.

(30) Kannoly, S.; Gao, T.; Dey, S.; Wang, I. N.; Singh, A.; Dennehy, J. J. Optimum threshold minimizes noise in timing of intracellular events. *iScience* **2020**, *23*, 101186.
(31) Baase, W. A.; Liu, L.; Tronrud, D. E.; Matthews, B. W. Lessons from the lysozyme of phage T4. *Protein Sci.* **2010**, *19*, 631−641.
(32) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; dek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.
(33) Cao, D. S.; Xu, Q. S.; Liang, Y. Z. propy: a tool to generate various modes of Chou's PseAAC. *Bioinformatics* **2013**, *29*, 960−962.
(34) Chou, K. C. Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins* **2001**, *43*, 246−255.
(35) Chou, K. C. Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology. *Curr. Proteomics* **2009**, *6*, 262−274.
(36) Medvedeva, A.; Teimouri, H.; Kolomeisky, A. B. Predicting antimicrobial activity for untested peptide-based drugs using collaborative filtering and link prediction. *J. Chem. Inf Model* **2023**, *63*, 3697−3704.
(37) Teimouri, H.; Medvedeva, A.; Kolomeisky, A. B. Bacteria-specific feature selection for enhanced antimicrobial peptide activity predictions using machine-learning methods. *J. Chem. Inf Model* **2023**, *63*, 1723−1733.
(38) Medvedeva, A.; Teimouri, H.; Kolomeisky, A. B. Differences in relevant physicochemical properties correlate with synergistic activity of antimicrobial peptides. *J. Phys. Chem. B* **2024**, *128*, 1407−1417.
(39) Leninger, M.; Sae Her, A.; Traaseth, N. J. Inducing conformational preference of the membrane protein transporter EmrE through conservative mutations. *Elife* **2019**, *8*, e48909.
(40) Frenz, B.; Lewis, S. M.; King, I.; DiMaio, F.; Park, H.; Song, Y. Prediction of protein mutational free energy: benchmark and sampling improvements increase classification accuracy. *Front Bioeng Biotechnol* **2020**, *8*, 558247.
(41) Ong, S. A.; Lin, H. H.; Chen, Y. Z.; Li, Z. R.; Cao, Z. Efficacy of different protein descriptors in predicting protein functional families. *BMC Bioinformatics* **2007**, *8*, 300.
(42) Jones, D. T.; Taylor, W. R.; Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci* **1992**, *8*, 275−282.
(43) Bigelow, C. C. On the average hydrophobicity of proteins and the relation between it and protein structure. *J. Theor. Biol.* **1967**, *16*, 187−211.
(44) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* **1999**, *290*, 253−266.
(45) Goldsack, D. E.; Chalifoux, R. C. Contribution of the free energy of mixing of hydrophobic side chains to the stability of the tertiary structure of proteins. *J. Theor. Biol.* **1973**, *39*, 645−651.
(46) Charton, M. Protein folding and the genetic code: an alternative quantitative model. *J. Theor. Biol.* **1981**, *91*, 115−123.
(47) Richards, F. M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.
(48) Charton, M.; Charton, B. I. The structural dependence of amino acid hydrophobicity parameters. *J. Theor. Biol.* **1982**, *99*, 629−644.
(49) Chou, K. C. Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. *Bioinformatics* **2005**, *21*, 10−19.
(50) Li, Z. R.; Lin, H. H.; Han, L. Y.; Jiang, L.; Chen, X.; Chen, Y. Z. PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. *Nucleic Acids Res.* **2006**, *34*, W32−37.
(51) Wimley, W. C.; White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **1996**, *3*, 842−848.
(52) Reißer, S.; Prock, S.; Heinzmann, H.; Ulrich, A. S. Protein ORIGAMI: A program for the creation of 3D paper models of folded peptides. *Biochem Mol. Biol. Educ* **2018**, *46*, 403−409.
(53) White, S. H.; Wimley, W. C. Hydrophobic interactions of peptides with membrane interfaces. *Biochim. Biophys. Acta* **1998**, *1376*, 339−352.
(54) White, S. H.; Wimley, W. C. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319−365.
(55) Tanford, C. The hydrophobic effect and the organization of living matter. *Science* **1978**, *200*, 1012−1018.

https://doi.org/10.1021/acs.jpcb.4c03040
J. Phys. Chem. B 2024, 128, 7129−7140