

Mutation Patterns in Lysostaphin

Chunhong Zhang¹, Weisheng Ye², Shaomin Yan³ , Guang Wu³ 

¹Department of Spinal Surgery IV, Tianjin Hospital, Tianjin, China; ²Department of Orthopedic Trauma, Tianjin Hospital, Tianjin, China; ³State Key Laboratory of Non-Food Biomass Enzyme Technology, Guangxi Key Laboratory of Biorefinery, National Engineering Research Center for Non-Food Biorefinery, Guangxi Academy of Sciences, Nanning, China

Correspondence to: Guang Wu, hongguanglishibahao@gxas.cn

Keywords: Amino-Acid Pair, Distribution Probability, Lysostaphin, Mutation Pattern, Predictability

Received: May 14, 2019

Accepted: June 17, 2019

Published: June 20, 2019

Copyright © 2019 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

ABSTRACT

Lysostaphin is widely used in clinical settings against *Staphylococcus aureus*, but its mutants can abolish its killing activity. The difficulty in studies of mutations in lysostaphin is the shortage of data, which may need many decades to collect, although lysostaphin is so important for clinical therapeutics and drug development. In order not to passively wait for the accumulation of new data, in this study 1) the 23,442 mutations in 1408 proteins from databank were used to determine whether the mutations in lysostaphin follow the general mutation trend obtained from the databank, 2) the amino-acid pair predictability was used to explore the underlined mechanism for lysostaphin mutations, and 3) the amino-acid distribution probability was used to associate the mutation with dysfunction of lysostaphin. The results show that the mutations in lysostaphin follow the general trend of mutations in proteins; the underlined mechanism for mutations in lysostaphin is explainable from a viewpoint of randomness, and a mutation with increased distribution probability would have a larger chance to dysfunction lysostaphin. This study provides useful information for future design of anti-*S. aureus* drug and enzyme engineering.

1. INTRODUCTION

Lysostaphin is widely used in clinical settings because it was found to have a specific lytic action against *Staphylococcus aureus* half a century ago [1]. Structurally and functionally, lysostaphin is a zinc metalloenzyme [2], which includes three enzymes, glycyglycine endopeptidase, endo-beta-N-acetylglucosaminidase and N-acteylmuramyl-L-alanine amidase. Glycyglycine endopeptidase specifically cleaves the glycine-glycine bonds, which is unique to the interpeptide cross-bridge of *S. aureus* cell wall [3].

With time going on, the interaction between lysostaphin and *S. aureus* leads to two outcomes, *i.e.*

mutations in *S. aureus* lead to drug-resistant [4] whereas mutations in lysostaphin lead to the dysfunction of lysostaphin. Indeed, previous studies have suggested that the site-directed mutants in lysostaphin can abolish its killing activity, *i.e.* the killing activity of glycyglycine endopeptidase against *S. aureus* [5].

These imply that the therapeutic effects of lysostaphin face serious challenges, because lysostaphin may someday lose its efficacy due to unpredictable mutations either in *S. aureus* or in itself. Nevertheless, these still suggest that efforts should be directed to study the mutations in both *S. aureus* and lysostaphin. Technically, studies on mutations either in *S. aureus* or in lysostaphin are not easy, because the structure and mechanism in *S. aureus* are very complicated, which make it difficult to figure out the real force and reason behind the mutations, whereas lysostaphin has few mutations although its biological structure and function are simple. However, it would be preferred to investigate the mutations in lysostaphin because its mutation pattern should have less unexplainable than that in *S. aureus*.

A question is whether the mutations in lysostaphin have some patterns, which could reveal the explainable mechanism for the general trend in lysostaphin dysfunctional mutations. Doubtless, this question is meaningful not only for the clinical therapeutics but also for drug developments. Unfortunately, not many mutations have been documented in literature for lysostaphin, especially, glycyglycine endopeptidase, over decades. Therefore, the researchers are facing to the problem regarding the mutations in lysostaphin: either we are passively waiting for hundreds of years to collect sufficient mutations, or we could conduct a study using these limited data, which is not fashionable nowadays because currently researchers are accustomed to using tens of thousands of data to draw a conclusion.

However, the lack of data should be weighed by the facts that not only such studies at worst can throw a clue to clinical therapeutics and drug development, but also other mutation data can be added. Therefore, we compare the general mutation patterns obtained from almost all the mutations in proteins documented in databank to lysostaphin mutations. If the handful mutations in lysostaphin could follow the general mutation patterns, then we would be in the position to address the question, what is the probability that a future mutation in lysostaphin would lead to its dysfunction?

On the other hand, the shortage of data suggests that we might not be able to use the standard techniques to investigate mutations because they generally require a large amount of data. In this study, we use the computational mutation approach [6-10] because it is more suitable for our purpose. Thus the aim of this study is designed to use the computational mutation approach to investigate the possible mutation pattern in lysostaphin.

2. MATERIALS AND METHODS

2.1. Data

Forty-one sequences of lysostaphin (EC 3.4.24.75) were documented in the Comprehensive Enzyme Information System BRENDA [11], of which 10 are glycyglycine endopeptidases. However, only two glycyglycine endopeptidases documented their mutations: O33599 with 5 mutations [12] and P10547 with 7 mutations [5] (the first and second columns in Table 1). Their sequences were obtained from UniProt [13].

Still, there are another 1408 proteins with 23,442 missense point mutations in UniProt [13], and these 1408 proteins are composed of 52 to 34,350 amino acids. Hopefully these 1408 proteins with 23,442 mutations would provide a general pattern to compare with mutations in lysostaphin.

2.2. Mutations in Terms of Predictable Proportion of Amino-Acid Pairs

Because there are 400 types of amino-acid pairs from combinations of 20 kinds of amino acids, we use the amino-acid pair predictability to classify the amino-acid pairs in a lysostaphin as predictable and unpredictable [6-10]. For example, lysostaphin P10547 has 51 glutamic acids (E), 38 glycines (G), and 48 valines (V). According to the permutation, the amino-acid pair GV should appear 4 times ($38/493 \times 51/492 \times 492 = 3.70$); actually there are four GVs in lysostaphin P10547, so the appearance of GV is predictable. Yet, the amino-acid pair VE should appear 5 times ($48/493 \times 51/492 \times 492 = 4.97$), however, it appears 30

times in lysostaphin P10547, so the appearance of VE is unpredictable. In this way, all amino-acid pairs in lysostaphin P10547 are classified, and its predictable proportion is 44.50%.

In this way, we can easily determine whether it occurs in predictable or unpredictable amino-acid pairs when a mutation occurs in a lysostaphin. Consequently we can estimate the change in predictable proportion of amino-acid pairs. All of the data can be calculated in the website [14].

In general, a point missense mutation results in two amino-acid pairs to mutate to another two amino-acid pairs. For instance, a mutation at position 210 in lysostaphin O33599 substitutes histidine (H) with alanine (A) and thus inactivates its own activity [12]. This mutation makes two amino-acid pairs AH and HY change to the pairs AA and AY, because the amino acid is alanine (A) at position 209 and tyrosine (Y) at position 211. Taking two original amino-acid pairs as an example, the actual and predicted frequencies were 3 and 2 for AH and 2 and 1 for HY before mutation (Row 4, Table 2), and were 2 and 2 for AH and 1 and 1 for HY after mutation (Row 17, Table 2). Taking two mutated amino-acid pairs into account, the actual and predicted frequencies were 3 and 3 for AA and 2 and 3 for AY before mutation (Row 30, Table 2), and were 4 and 4 for AA and 3 and 3 for AY after mutation (Row 43, Table 2). Thus, we can estimate their change in frequency to find the mutation effects on these amino-acid pairs.

Table 1. Increase (↑) and decrease (↓) in predictable proportion of amino-acid pairs before and after mutation.

Lysostaphin	Mutation	Before mutation	After mutation	Difference	Change
O33599	N117A	60.25	59.50	-0.75	↓
	H210A	60.25	61.50	1.25	↑
	D214A	60.25	60.00	-0.25	↓
	H291A	60.25	60.50	0.25	↑
	H293A	60.25	60.75	0.50	↑
	H279A	44.50	44.75	0.25	↑
	D283A	44.50	44.75	0.25	↑
P10547	Y327A	44.50	45.75	1.25	↑
	H329A	44.50	44.75	0.25	↑
	T353A	44.50	44.75	0.25	↑
	H360A	44.50	44.75	0.25	↑
	H362A	44.50	45.00	0.50	↑

Table 2. Actual frequency (AF), predicted frequency (PF) and their difference $\Sigma(PF-AF)$ in original and mutated pairs before and after mutation (the abbreviations of amino acid are given in Pair I and Pair II columns).

Mutation	Pair I	AF	PF	Pair II	AF	PF	$\Sigma(PF-AF)$
1) Original pairs before mutation							
O33599 N117A	DN	2	2	NN	5	3	-2
O33599 H210A	AH	3	2	HY	2	1	-2
O33599 D214A	VD	3	1	DY	3	2	-3
O33599 H291A	PH	1	0	HV	1	1	-1
O33599 H293A	VH	1	1	HF	1	0	-1
P10547 H279A	MH	2	0	HY	2	0	-4

Continued

P10547 D283A	VD	1	1	DF	1	0	-1
P10547 Y327A	WY	1	0	YM	1	0	-2
P10547 H329A	MH	2	0	HL	2	0	-4
P10547 T353A	ST	5	4	TG	4	4	-1
P10547 H360A	PH	1	1	HL	2	0	-2
P10547 H362A	LH	1	0	HF	1	0	-2
2) Original pairs after mutation							
O33599 N117A	DN	1	2	NN	4	3	0
O33599 H210A	AH	2	2	HY	1	1	0
O33599 D214A	VD	2	1	DY	2	1	-2
O33599 H291A	PH	0	0	HV	0	1	1
O33599 H293A	VH	0	1	HF	0	0	1
P10547 H279A	MH	1	0	HY	1	0	-2
P10547 D283A	VD	0	1	DF	0	0	1
P10547 Y327A	WY	0	0	YM	0	0	0
P10547 H329A	MH	1	0	HL	1	0	-2
P10547 T353A	ST	4	4	TG	3	4	1
P10547 H360A	PH	0	0	HL	1	0	-1
P10547 H362A	LH	0	0	HF	0	0	0
3) Mutated pairs before mutation							
O33599 N117A	DA	4	2	AN	3	3	-2
O33599 H210A	AA	3	3	AY	2	3	1
O33599 D214A	VA	0	1	AY	2	3	2
O33599 H291A	PA	1	1	AV	1	1	0
O33599 H293A	VA	0	1	AF	0	1	2
P10547 H279A	MA	0	1	AY	0	2	3
P10547 D283A	VA	1	5	AF	0	1	5
P10547 Y327A	WA	0	1	AM	0	1	2
P10547 H329A	MA	0	1	AL	3	2	0
P10547 T353A	SA	3	4	AG	5	4	0
P10547 H360A	PA	0	3	AL	3	2	2
P10547 H362A	LA	2	2	AF	0	1	1
4) Mutated pairs after mutation							
O33599 N117A	DA	5	2	AN	4	3	-4
O33599 H210A	AA	4	4	AY	3	3	0
O33599 D214A	VA	1	2	AY	3	3	1
O33599 H291A	PA	2	1	AV	2	2	-1
O33599 H293A	VA	1	2	AF	1	1	1
P10547 H279A	MA	1	1	AY	1	2	1

Continued

P10547 D283A	VA	2	5	AF	1	1	3
P10547 Y327A	WA	1	1	AM	1	1	0
P10547 H329A	MA	1	1	AL	4	2	-2
P10547 T353A	SA	4	4	AG	6	4	-2
P10547 H360A	PA	1	3	AL	4	2	0
P10547 H362A	LA	3	2	AF	1	1	-1

2.3. Mutation in Terms of Amino-Acid Distribution Probability

To search for the probability that a mutation in lysostaphin in future will lead to its dysfunction, we use the amino-acid distribution probability [9, 10], where the position of amino acid in a protein is viewed as the problem of the occupancy of subpopulations and partitions [15] with the following equation, $r!/(q_0 \times q_1 \times \dots \times q_n!) \times r!/(r_1 \times r_2 \times \dots \times r_n!) \times n^{-r}$. For our propose, in this equation! is the factorial function, r is the number of a type of amino acid, q is the number of partitions with the same number of amino acids and n is the number of partitions in the protein for a type of amino acid. For example, there are 8 aspartic acids (D) in glycyglycine endopeptidase of lysostaphin P10547, and their positions are 37, 74, 93, 131, 177, 204 and 210. Their amino-acid distribution probability is calculated according to the above equation [15]:

$$\begin{aligned} & \frac{r!}{(q_0 \times q_1 \times \dots \times q_n!)} \times \frac{r!}{(r_1 \times r_2 \times \dots \times r_n!)} \times n^{-r} \\ &= \frac{7!}{2 \times 3 \times 2 \times 0 \times 0 \times 0 \times 0 \times 0!} \times \frac{7!}{0 \times 1 \times 2 \times 1 \times 1 \times 2 \times 0!} \times 7^{-7} \\ &= \frac{5040}{24} \times \frac{5040}{4} \times \frac{1}{823543} = 0.3213 \end{aligned}$$

A mutation at position 37 leads D to mutate to A, which is clinically dysfunctional, so we have

$$\begin{aligned} & \frac{6!}{2 \times 3 \times 0 \times 1 \times 0 \times 0 \times 0!} \times \frac{6!}{0 \times 1 \times 1 \times 1 \times 3 \times 0!} \times 6^{-6} \\ &= \frac{720}{12} \times \frac{720}{6} \times \frac{1}{46656} = 0.1543 \end{aligned}$$

These two probabilities concern the original amino acid, similarly we can also compute two probabilities for the mutated amino acid, which are 0.0621 and 0.1544 before and after mutation, and the effect of mutation is $(0.1543 - 0.3213) + (0.1544 - 0.0621) = 0.0609$, suggesting that this mutation increased the amino-acid distribution probability. In this way, we quantify the mutation numerically with respect to original and mutated amino acids [16].

2.4. Statistics

The *Chi*-square test was used to compare the number of amino-acid pairs, and the Mann-Whitney *U*-test was used to compare the difference between predicted and actual frequencies.

3. RESULTS

As there are not many mutations in lysostaphin, we necessarily compare its mutations against the general mutation patterns of all proteins with mutations in databank. If the mutation pattern in lysostaphin follows the general pattern of mutations in all proteins with mutations in databank, then we could

suggest that the mutations in lysostaphin would have a similar mechanism as other proteins and the studies on mutations in other proteins would shed light on mutations in lysostaphin. If this is not the case, we should consider that the mutations in lysostaphin have different mechanisms from other proteins, and then the research would follow different directions.

Mutations in different proteins are so different that we use the ratio of mutation position versus protein length as an indicator to compare mutations in lysostaphin. **Figure 1** and **Figure 2** show such comparison. As shown by arrows, there is some similarity between general mutation patterns and mutations in lysostaphin. This means that a likely mutation would be positioned around six tenths of lysostaphin lengths. At least, these positions could be defined as so-called “hotspots” for mutations [17]. Other hotspots could be positions around ratio of 0.56 to 0.57, which also have relatively more mutations in **Figure 1**.

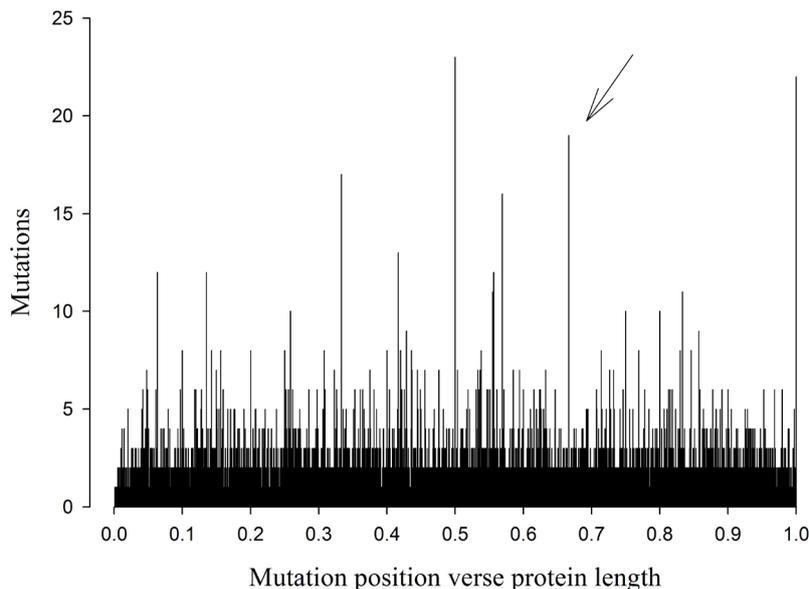


Figure 1. Ratio of 23,442 mutations versus 1408 protein lengths.

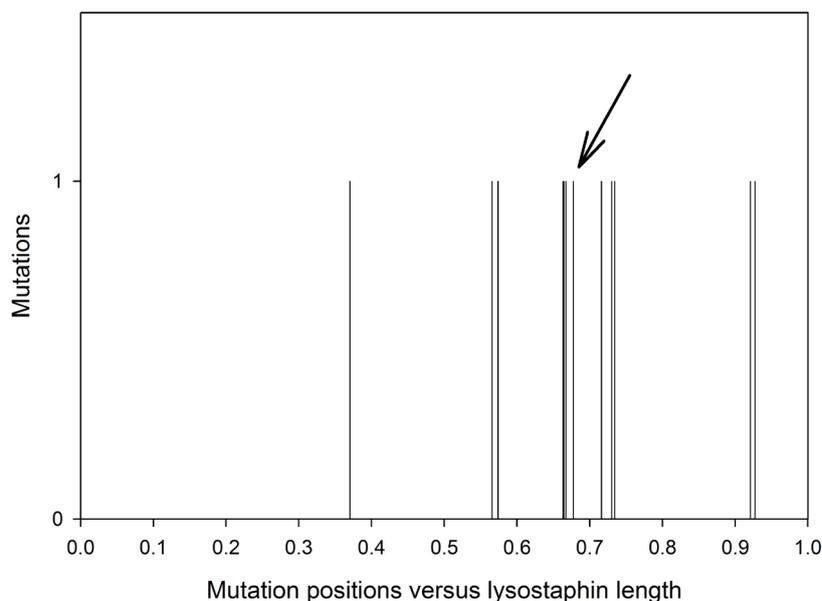


Figure 2. Ratio of 12 mutations versus 2 lysostaphin lengths.

However, the comparison between **Figure 1** and **Figure 2** could not explain the documented mutation positions beyond the ratio of 0.7 (arrows in **Figure 1** and **Figure 2**) although the general mutation patterns also have less intensive mutations around this region. This difference is very suggestive because lysostaphin might have mutated so many times in this region, so we would expect the mutations in future in other regions. Alternatively, the mutations in other regions of lysostaphin might have led to the complete dysfunction of lysostaphin at very early stage, and then lysostaphin has been cleaved by proteases so those mutations were not documented.

Following these general trends, we can look at what the computational mutation approach tells us. **Table 1** shows the changes in predictable proportions of amino-acid pairs before and after mutation. As can be seen, 10 out of 12 mutations increase the predictable proportion in mutant lysostaphin, and 2 mutations decrease the predictable proportion, indicating that most mutant lysostaphin has a larger predictable proportion of amino-acid pairs (*Chi-square test*, $P < 0.05$).

Hereafter, we can analyze the changes of amino-acid pairs to find what effects the mutations have. **Table 2** details the actual frequency, predicted frequency and their difference in original and mutated amino-acid pairs before and after mutation. The unpredictable pair characterizes one of targeted amino-acid pairs with the actual frequency larger than its predicted one so that the frequency differences are negative in the original pairs before mutation (Rows 3-14, **Table 2**). After mutation, the original pairs have smaller frequency difference and become more predictable (Rows 16-27, **Table 2**).

For the mutated amino-acid pairs, 10 of 12 mutations were involved at least one unpredictable pair whose actual frequency is smaller than its predicted one so that the frequency differences are more likely to be positive before mutation (Rows 29 - 40, **Table 2**). After mutation in general, the mutated pairs have smaller frequency difference and become more predictable (Rows 42 - 53, **Table 2**).

Figure 3 illustrates the mutation effects on the targeted amino-acid pairs. As can be seen, the median of frequency difference is -2 for original pairs and 1.5 for mutated ones before mutation. After mutation, the median of frequency difference is zero for both original and mutated pairs (the Mann-Whitney *U*-test, $P < 0.01$).

In order to find the probability that a mutation leads to dysfunction of lysostaphin, we use the amino-acid distribution probability to quantify the documented mutations in lysostaphin, of which lysostaphin P10547 documented 7 mutations and lysostaphin O33599 documented 5 mutations. **Table 3** and **Table 4** show that most mutations lead the increase in difference before and after mutation in terms of amino-acid distribution probability. This observation is similar to the findings using amino-acid distribution probability in other proteins [9, 10].

Accordingly, we can conduct the cross-impact analysis to get more insight into mutations in lysostaphin [18]. **Figure 4** shows the scheme based on cross-impact analysis for events with respect to mutations and dysfunction. At the level of amino-acid distribution probability, $P(2)$ and $P(\bar{2})$ are the decreased and increased probabilities induced by mutations, and 2 and 10 mutations result in the distribution probability decreased and increased, respectively. At the level of classification: 1) $P(1|\bar{2})$ is the impact probability (conditional probability) that the mutation activates the glycyglycine endopeptidase of lysostaphin occurs under the condition of increased distribution probability, and one mutation has such an effect. 2) $P(\bar{1}|\bar{2})$ is the impact probability that the glycyglycine endopeptidase of lysostaphin dysfunctions under the condition of increased distribution probability, and 9 mutations work in such a manner. 3) $P(1/2)$ is the impact probability that the mutation activates the glycyglycine endopeptidase of lysostaphin under the condition of decreased distribution probability, and no mutation plays such a role. 4) $P(\bar{1}|2)$ is the impact probability that the glycyglycine endopeptidase of lysostaphin dysfunctions under the condition of decreased distribution probability, and 2 mutations fall into this category. At the level of combined events, we see the combined results with their frequency.

Table 5 lists the computed probabilities in scheme in **Figure 4**, from which several interesting points can be found. 1) As $P(\bar{2})$ is larger than $P(2)$, a mutation has 10/12 chance of increasing the distribution probability in mutant lysostaphin. 2) As $P(1|\bar{2})$ is far much smaller than $P(\bar{1}|\bar{2})$, a mutation that increases the distribution probability has little chance to activate lysostaphin. 3) As $P(1/2)$ is remarkably

smaller than $P(\bar{1}|2)$, a mutation that decreases the distribution probability has no chance to activate lysostaphin.

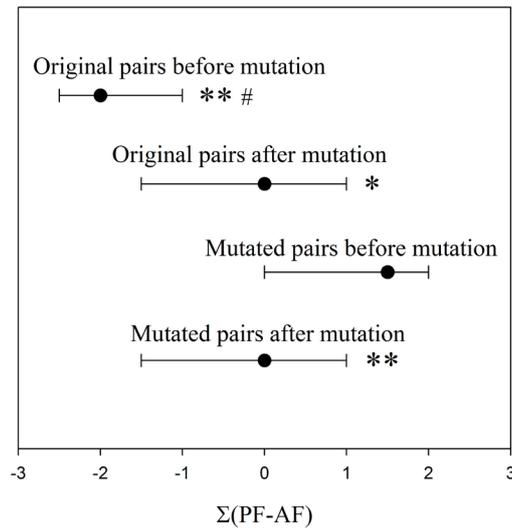


Figure 3. Sum of differences between predicted frequency and actual frequency $\Sigma(\text{PF-AF})$ in original and mutated pairs before and after mutation. The data are presented as median with interquartile. * and ** indicate the statistical difference at $P < 0.05$ and $P < 0.01$ level compared with the mutated pairs before mutation. # indicates the statistical difference at $P < 0.01$ level compared with the original pairs after mutation.

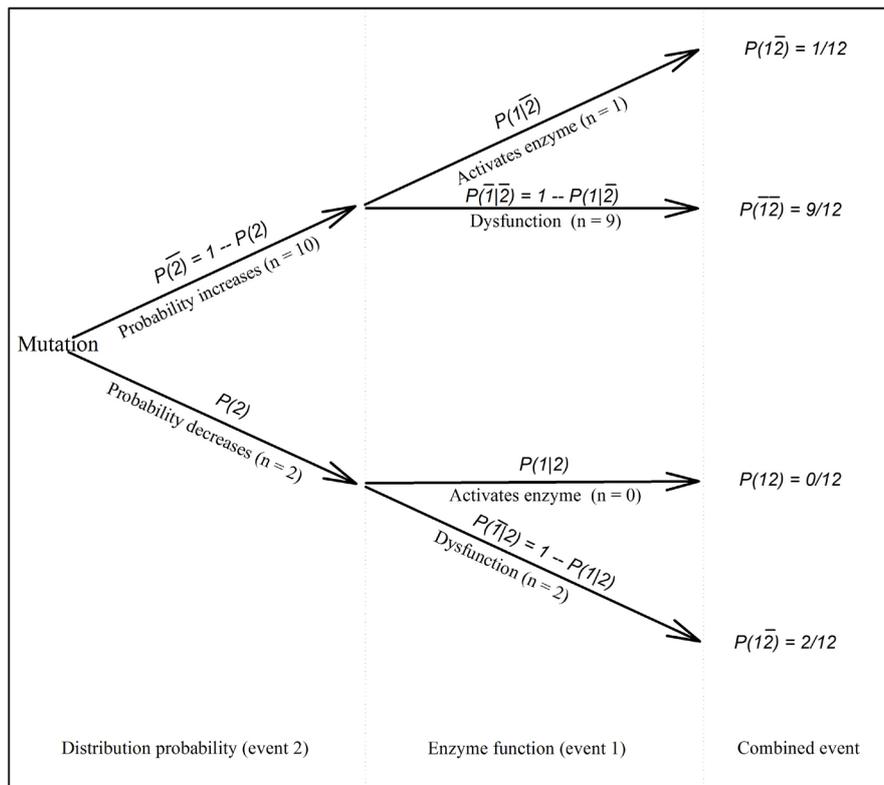


Figure 4. Cross-impact relationship among mutation, function and combined events in lysostaphin.

Table 3. Increase (↑) and decrease (↓) in amino-acid distribution probability of seven mutations in P10547 lysostaphin before and after mutation.

Mutation	Distribution Probability for Original Amino Acid		Distribution Probability for Mutated Amino Acid		Difference	Change
	Before mutation	After mutation	Before mutation	After mutation		
D → A	0.3213	0.1543	0.0621	0.1544	-0.0747	↓
H → A	0.0369	0.1682	0.0621	0.0515	0.1207	↑
H → A	0.0369	0.1682	0.0621	0.0386	0.1078	↑
H → A	0.0369	0.0105	0.0621	0.1544	0.0659	↑
H → A	0.0369	0.1682	0.0621	0.1544	0.2236	↑
T → A	0.0183	0.0177	0.0621	0.0515	-0.0112	↓
Y → A	0.0341	0.0336	0.0621	0.1544	0.0918	↑

Table 4. Increase (↑) and decrease (↓) in Amino-acid distribution probability of five mutations in O33599 lysostaphin before and after mutation.

Mutation	Distribution Probability for Original Amino Acid		Distribution Probability for Mutated Amino Acid		Difference	Change
	Before mutation	After mutation	Before mutation	After mutation		
N → A	0.0309	0.0415	0.0169	0.0079	0.0016	↑
H → A	0.0374	0.0618	0.0169	0.0268	0.0343	↑
D → A	0.1280	0.1192	0.0169	0.0268	0.0011	↑
H → A	0.0374	0.1031	0.0169	0.0134	0.0622	↑
H → A	0.0374	0.1031	0.0169	0.0134	0.0622	↑

Table 5. Probability in cross-impact analysis in Figure 4.

$P(2) = 2/12 = 0.1667$
$P(\bar{2}) = 1 - P(2) = 1 - 0.1667 = 0.8333 = 10/12$
$P(1 \bar{2}) = 1/10 = 0.1$
$P(\bar{1} \bar{2}) = 1 - P(1 \bar{2}) = 1 - 0.1 = 0.9 = 9/10$
$P(1/2) = 0/2 = 0$
$P(\bar{1} 2) = 1 - P(1/2) = 1 - 0 = 1 = 2/2$
$P(1\bar{2}) = P(1 \bar{2}) \times P(\bar{2}) = 1/10 \times 10/12 = 0.0833 = 1/12$
$P(\bar{1}\bar{2}) = P(\bar{1} \bar{2}) \times P(\bar{2}) = 9/10 \times 10/12 = 0.9167 = 9/12$
$P(12) = P(1/2) \times P(2) = 0/2 \times 2/12 = 0 = 0/12$
$P(\bar{1}2) = P(\bar{1} 2) \times P(2) = 2/2 \times 2/12 = 0.1667 = 2/12$

4. DISCUSSION

In this study, we attempted to gain the insight of mutations in lysostaphin in three ways. First, we compare the mutations in lysostaphin against almost all the available mutations in proteins from databank to determine whether the mutations in lysostaphin follow the general trend in mutations. Second, we use the amino-acid pair predictability to explore the possible underlined mechanism for the mutations in lysostaphin. Third, we use the amino-acid distribution probability to determine the probability of dysfunction of lysostaphin when a mutation occurs.

Here, we need more elaborations on the underlined mechanism for mutations in lysostaphin. The results reveal the mutation pattern in lysostaphin based on the current data: the mutations are more likely to target the unpredictable amino-acid pairs, and reduce the difference between actual and predicted frequencies, leading to the increase in the predictable proportion of amino-acid pairs in mutant lysostaphin. This is very meaningful because it is consistent with the mutation patterns obtained from other proteins by means of computational mutation approaches [6-10].

The hotspot sites [17, 19-21], and function disruption [5, 22, 23] can explain why some amino acids are mutated more frequently than others. From the random viewpoint, the predictable amino-acid pairs suggest that their construction needs the least time and energy with the maximal probability of occurrence. Natural parsimony demands enzyme construction with least energy and time to adapt for the evolution, which would lead to mutations [6-10]. The lysostaphin mutations do diminish the difference between predicted and actual frequencies, which generally lead to the dysfunction of lysostaphin.

On the other hand, the unpredictable pairs suggest that nature spends more time and energy to deliberately construct them so the functional amino-acid pairs should be deliberately evolved, and maintained the difference between actual frequency and predicted one. Interestingly there is a mutation from lysostaphin O33599, which changes the asparagine at position 117 to alanine, and is the only one activating the enzyme [12]. Our study shows that this mutation enlarges the frequency difference of mutated pairs to -4 after mutation (Row 42, Table 2) from -2 before mutation (Row 29, Table 2), and decreases the predictable proportion of amino-acid pairs in the mutant (Row 2, Table 1).

In conclusion, the significance of this study is that we get three pieces of information regarding the mutations in lysostaphin: 1) the mutations in lysostaphin follow the general trend of mutations in proteins, 2) the underlined mechanism for mutations in lysostaphin can be explainable from the viewpoint of randomness, and 3) a mutation with increased distribution probability would have a larger chance to dysfunction lysostaphin. These results provide useful information for future design of anti-*S. aureus* drug and enzyme engineering.

FUND

This study was supported by National Natural Science Foundation of China (31460296 and 31560315), and Key Project of Guangxi Scientific Research and Technology Development Plan (AB17190534).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this paper.

REFERENCES

1. Browder, H.P., Zygmunt, W.A., Young, J.R. and Travormina, P.A. (1965) Lysostaphin: Enzymatic Mode of Action. *Biochemical and Biophysical Research Communications*, **19**, 383-389. [https://doi.org/10.1016/0006-291X\(65\)90473-0](https://doi.org/10.1016/0006-291X(65)90473-0)
2. Cohen, D.N., Sham, Y.Y., Haugstad, G.D., Xiang, Y., Rossmann, M.G., Anderson, D.L. and Popham, D.L. (2009) Shared Catalysis in Virus Entry and Bacterial Cell Wall Depolymerization. *Journal of Molecular Biology*, **387**, 607-618. <https://doi.org/10.1016/j.jmb.2009.02.001>

3. Bastos, M.D., Coutinho, B.G. and Coelho, M.L. (2010) Lysostaphin: A Staphylococcal Bacteriolysin with Potential Clinical Applications. *Pharmaceuticals*, **3**, 1139-1161. <https://doi.org/10.3390/ph3041139>
4. Vimberg, V., Cavanagh, J.P., Benada, O., Kofroňová, O., Hjerde, E., Zieglerová, L. and Balíková Novotná, G. (2018) Teicoplanin Resistance in *Staphylococcus haemolyticus* Is Associated with Mutations in Histidine Kinases VraS and WalK. *Diagnostic Microbiology and Infectious Disease*, **90**, 233-240. <https://doi.org/10.1016/j.diagmicrobio.2017.11.007>
5. Bardelang, P., Vankemmelbeke, M., Zhang, Y., Jarvis, H., Antoniadou, E., Rochette, S., Thomas, N.R., Penfold, C.N. and James, R. (2009) Design of a Polypeptide FRET Substrate That Facilitates Study of the Antimicrobial Protease Lysostaphin. *Biochemical Journal*, **418**, 615-424. <https://doi.org/10.1042/BJ20081765>
6. Wu, G. and Yan, S. (2002) Randomness in the Primary Structure of Protein: Methods and Implications. *Molecular Biology Today*, **3**, 55-69. <https://doi.org/10.1063/1.2408446>
7. Wu, G. and Yan, S. (2006) Fate of Influenza A Virus Proteins. *Protein & Peptide Letters*, **13**, 377-384. <https://doi.org/10.2174/092986606775974474>
8. Wu, G. and Yan, S. (2006) Mutation Trend of Hemagglutinin of Influenza A Virus: A rEview from Computational Mutation Viewpoint. *Acta Pharmacologia Sinica*, **27**, 513-526. <https://doi.org/10.1111/j.1745-7254.2006.00329.x>
9. Wu, G. and Yan, S. (2008) Lecture Notes on Computational Mutation. Nova Science Publishers, New York.
10. Yan, S.M. and Wu, G. (2010) Creation and Application of Computational Mutation. *Journal of Guangxi Academy of Sciences*, **26**, 130-139.
11. http://www.brenda-enzymes.info/php/result_flat.php4?ecno=3.4.24.75
12. Odintsov, S.G., Sabala, I., Marcyjaniak, M. and Bochtler, M. (2004) Latent LytM at 1.3Å Resolution. *Journal of Molecular Biology*, **335**, 775-785. <https://doi.org/10.1016/j.jmb.2003.11.009>
13. UniProt Consortium (2019) UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Research*, **47**, D506-D515. <https://doi.org/10.1093/nar/gky1049>
14. <http://www.gxas.cn/pp.htm>
15. Feller, W. (1968) An Introduction to Probability Theory and Its Applications. 3rd Edition, Vol. 1, Wiley, New York.
16. <http://www.gxas.cn/dp.htm>
17. Kim, S. and Jeong, S. (2019) Mutation Hotspots in the β -Catenin Gene: Lessons from the Human Cancer Genome Databases. *Molecules and Cells*, **42**, 8-16.
18. Sage, A.P. (1977) Methodology for Large-Scale Systems. McGraw-Hill, New York, 165-203.
19. Stern, D.L. and Orgogozo, V. (2009) Is Genetic Evolution Predictable? *Science*, **323**, 746-751. <https://doi.org/10.1126/science.1158997>
20. Parsons, B.L., Myers, M.B., Meng, F., Wang, Y. and McKinzie, P.B. (2010) Oncomutations as Biomarkers of Cancer Risk. *Environmental and Molecular Mutagenesis*, **51**, 836-850. <https://doi.org/10.1002/em.20600>
21. Tock, A.J. and Henderson, I.R. (2018) Hotspots for Initiation of Meiotic Recombination. *Frontiers in Genetics*, **9**, 521. <https://doi.org/10.3389/fgene.2018.00521>
22. Tiemann-Boege, I., Schwarz, T., Striedner, Y. and Heissl, A. (2017) The Consequences of Sequence Erosion in the Evolution of Recombination Hotspots. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**, 1736. <https://doi.org/10.1098/rstb.2016.0462>
23. Dornan, G.L. and Burke, J.E. (2018) Molecular Mechanisms of Human Disease Mediated by Oncogenic and Primary Immunodeficiency Mutations in Class IA Phosphoinositide 3-Kinases. *Frontiers in Immunology*, **9**, 575. <https://doi.org/10.3389/fimmu.2018.00575>