



Article

# Amino-Acid Network Clique Analysis of Protein Mutation Non-Additive Effects: A Case Study of Lysozyme

Dengming Ming <sup>1,\*</sup> , Rui Chen <sup>1</sup> and He Huang <sup>1,2,\*</sup>

<sup>1</sup> College of Biotechnology and Pharmaceutical Engineering, Nanjing Tech University, Biotech Building Room B1-404, 30 South Puzhu Road, Nanjing 211816, Jiangsu, China; chenatf@gmail.com

<sup>2</sup> College of Pharmaceutical Sciences, Nanjing Tech University, 30 Puzhu South Road, Nanjing 211816, Jiangsu, China

\* Correspondence: dming@njtech.edu.cn (D.M.); biotech@njtech.edu.cn (H.H.)

Received: 14 February 2018; Accepted: 7 May 2018; Published: 10 May 2018



**Abstract:** Optimizing amino-acid mutations in enzyme design has been a very challenging task in modern bio-industrial applications. It is well known that many successful designs often hinge on extensive correlations among mutations at different sites within the enzyme, however, the underpinning mechanism for these correlations is far from clear. Here, we present a topology-based model to quantitatively characterize non-additive effects between mutations. The method is based on the molecular dynamic simulations and the amino-acid network clique analysis. It examines if the two mutation sites of a double-site mutation fall into to a 3-clique structure, and associates such topological property of mutational site spatial distribution with mutation additivity features. We analyzed 13 dual mutations of T4 phage lysozyme and found that the clique-based model successfully distinguishes highly correlated or non-additive double-site mutations from those additive ones whose component mutations have less correlation. We also applied the model to protein Eglin c whose structural topology is significantly different from that of T4 phage lysozyme, and found that the model can, to some extension, still identify non-additive mutations from additive ones. Our calculations showed that mutation non-additive effects may heavily depend on a structural topology relationship between mutation sites, which can be quantitatively determined using amino-acid network  $k$ -cliques. We also showed that double-site mutation correlations can be significantly altered by exerting a third mutation, indicating that more detailed physicochemical interactions should be considered along with the network clique-based model for better understanding of this elusive mutation-correlation principle.

**Keywords:** mutation non-additive effect; free energy change; amino-acid network;  $k$ -clique community; protein dynamics

## 1. Introduction

Successful enzyme design often hinges on a good understanding of the relationship between protein structures and their biological functions. A key step in rational design is the introduction of special amino-acid replacements at particular sites of the studied proteins, which is expected to enhance the protein thermo-stability and catalytic activity, etc. In practice, simultaneous mutations at two or more sites in the target proteins, rather than a single-site mutation, are required. Thus, one critical question concerning mutation design arises: is there any correlation between mutations at different sites in the studied proteins? If so, can we predict them? Obviously, if mutations at different sites are independent from one another, then the overall effect of the multiple-mutation can be estimated by simply summing up the effect of every single mutation and is called to be additive [1]. On the contrary,

in cases where a strong interplay between mutations at different sites exists, the overall mutation effects are unpredictable from those of single mutations and exhibit non-additive effects.

Mutation additivity effects had been studied in a variety of backgrounds in early days by many structural biologists. For example, Sandberg and Terwilliger [2] examined the additive effects of mutations in gene V proteins, and found that different types of mutations showed strong additivity. In addition, they found that mutations at sites that have intense van der Waals interactions tend to be weaker additives. Boyer and colleagues [3] suggested that the non-additivity of mutations at distant sites indicates an information communication between amino acids at these sites, and they called “thermodynamic coupling” for the enhanced thermo-stability due to this non-additive phenomenon. They used atomic resolution nuclear magnetic resonance (NMR) to examine the hydrogen exchange in the enzyme at its natural state and had attempted to determine the dynamic perturbation between the two mutation sites. They concluded that thermodynamic coupling between distal sites was caused by physical interactions between amino acids at these sites in the natural structure of the studied protein.

T4 phage lysozyme, as a model protein for studying the relationship between protein structure and their functions, was also used for the study of mutation non-additive effects in early days. Matthews and colleagues [4] observed that mutations that introduce negative charges at ends of  $\alpha$  helices in T4 phage lysozyme and produced electrostatic interactions at these sites, such as S38D and N144D, were additive. They designed a series of combinatorial mutations at distant sites that do not form direct contact. Interestingly, most of these multiple mutations are found to be very strong additives, and they can form either direct physical contacts or not. An extreme example that used the mutation additive effect is a combination of 7 mutations, S38D/A82P/N144D/I3L/V131A/A41V/N116D, which was found to have the largest melting temperature increase of 8.3 °C [5].

On the other hand, some mutations that involve direct physical interactions did exhibit strong non-additivity. For example, the double-site mutant A98V/T152S showed strong non-additivity compared with the corresponding single mutations, and melting temperature change caused by the double-site mutation was 7.6 °C less than the summation of those imposed by the two corresponding single mutations [6,7]. In the native structure, the two residues A98 and T152 orient to one another and form direct contact. Matthews and colleagues [8] suggested that the dynamic perturbation a mutation introduces will start at the mutation site and spread to its neighboring sites. According to their observation, for a given site, if its neighborhood can digest more perturbations caused by mutation at this site, then such a mutation might impose smaller changes to protein thermal stability. In the case of A98V/T152S, the two sites have strong interactions and thus their neighborhood’s responses to the mutations at these sites heavily depend on the detailed interactions: presumably mutations that enhance the two-site contacts might weaken the power of the neighboring structures to relieve the mutation perturbations, thus causing larger thermal stability changes. Interestingly, other mutations that do not involve direct interactions were also found having strong non-additivity [9].

Undoubtedly, it is important to understand the mechanism underpinning the mutation non-additive effects, and predetermining mutation additivity at selected sites can effectively reduce experimental workload in rationale design. Recently, the quick accumulation of mutation data had stimulated the developing of methods for the prediction of mutation effects. For example, Tian et al. [10] developed a machine-learning method to predict mutation effects on protein thermo-stability based on a 3366 mutant protein database. Pires et al. [11] predicted missense mutations using some graph-based signatures. Very recently, Dehghanpoor et al. [12] compared the performance of several machine learning methods. However, up-to-now accurate prediction of mutation correlation effects, such as non-additivity, is still a very challenging task.

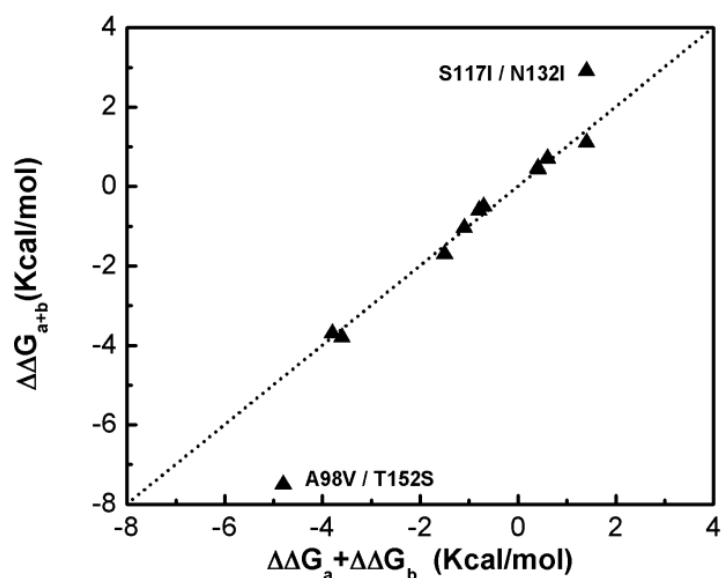
In this paper, a mathematical model based on a protein structural amino-acid network was presented that successfully isolated double-site mutations with strong non-additive effects from additive ones for T4 bacteriophage lysozyme. We had examined different factors of a protein topological network that show a strong correlation with the mutation additivity. Double-site mutations of the T4 phage lysozyme were selected if the two component single-site mutations were also measured,

and the non-additive effect of mutations at the two sites was determined based on the measured thermodynamics data [9]. The dependence of the non-additive effect on the distance between the two sites was examined. We then presented a protein topological network model based on amino acid interactions [13], and examined the network topological quantities and their relationships with the mutation non-additive effects. Finally, we presented a mathematical model based on protein network clique analysis to predict mutation additivity/non-additivity. The model was also successfully applied to a new protein of Eglin c whose structure has a different topology from that of the T4 phage lysozyme.

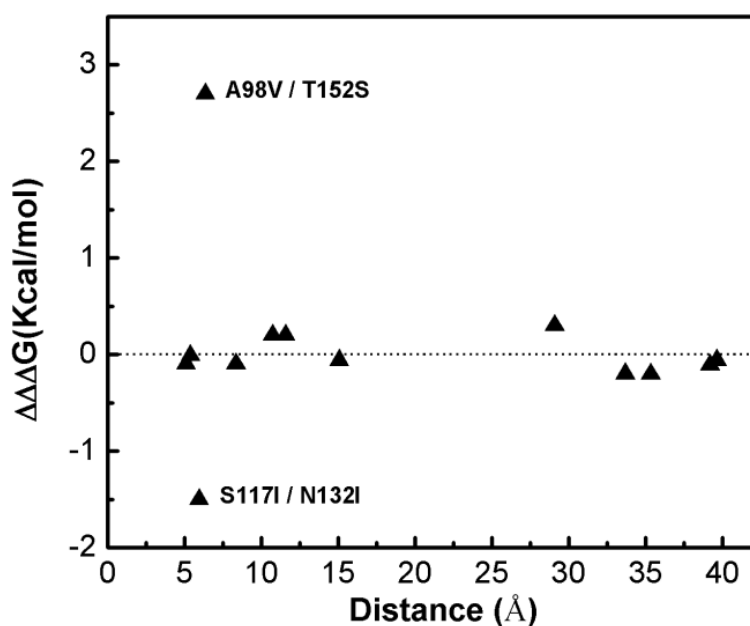
## 2. Results and Discussion

### 2.1. The Non-Additive Effects in Double-Site Mutations Are Independent of the Mutation-Site Distance

Figure 1 shows that most of the double-site mutations are strong additive, and the two double-site mutations, the S117I/N132I and A98V/T152S, show significant non-additivity. Interestingly, compared with its component single mutations the non-additive effect of the double-site mutant A98V/T152S significantly decreases the enzyme thermo-stability due to increment in free energy change, whereas the non-additive effect in S117I/N132I makes the enzyme stable due to a decrement in free energy change. Quantitatively, the non-additive effect of mutant S117I/N132I is weaker than that of A98V/T152S. There is no obvious correlation between site distance and the additivity effect of double-site mutation (Figure 2). We noticed that the distances between the sites of double-site mutations with strong non-additivity are relatively small and the reverse scenario is not necessarily true, i.e., double-site mutations exhibit very weak non-additivity whose mutation-site-distance are actually very short. No strong non-additivity effect was observed for long-distance double-site mutants. Presumably, intensive interference between the two sites of a double-site mutation might be required in order to exhibit a strong non-additivity effect, and this interferential interaction might be lacking or very weak when the two sites are well-separated. However, for a relatively short double-site mutation, it is still interesting to understand why a few of them, such as E128A/V131A, are non-additive while a majority of them are still additive.



**Figure 1.** Compare the free energy changes of double-site mutations with that derived by the summation of the two corresponding single-site mutations of the T4 phage lysozyme.



**Figure 2.** The dependence of the double-site mutation effects on the distances between the two corresponding single sites of the T4 phage lysozyme.

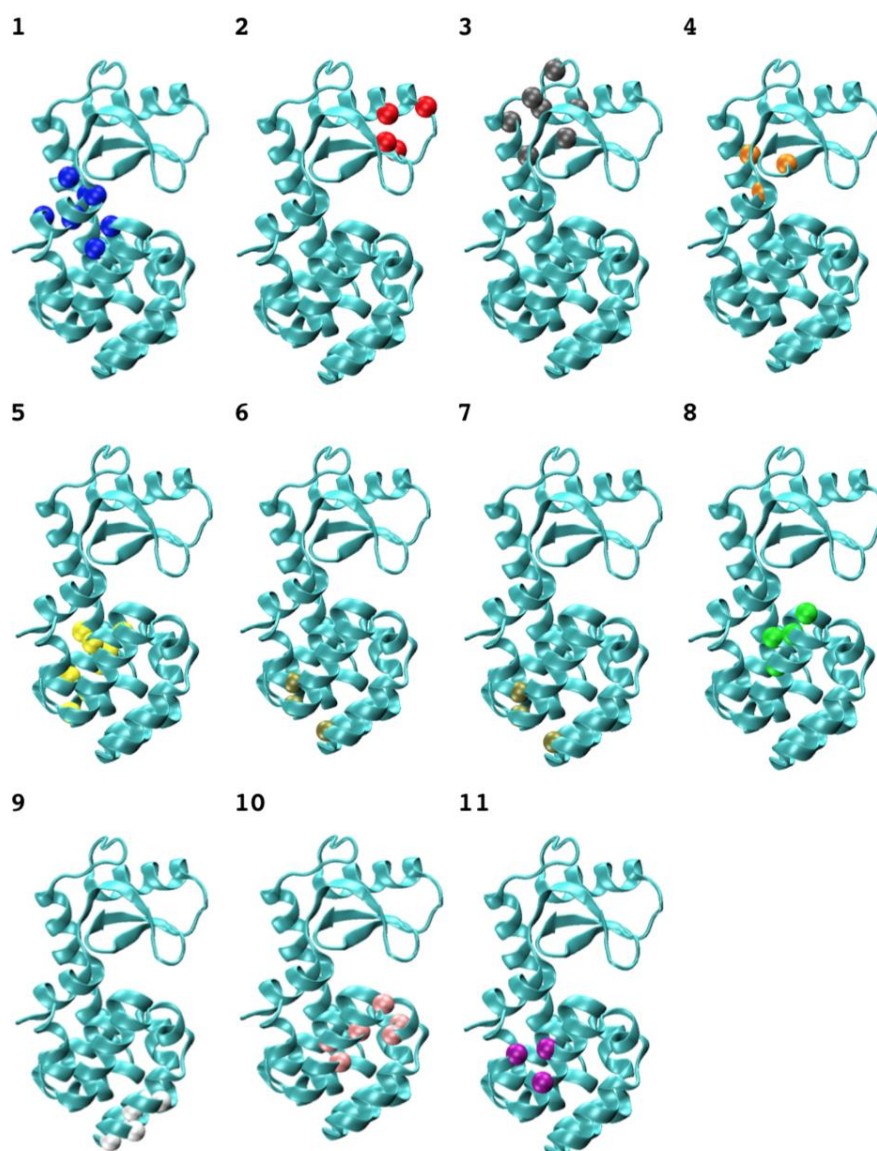
### 2.2. Double-Site Mutations That Have Strong Non-Additive Effects Tend to Have Their Sites Located within A 3-Clique and That of Weak Correlation Have Their Sites Located in Different 3-Cliques

The calculated 3-cliques of T4 bacteriophage lysozyme vary in size and locations (Table 1). It is interesting to examine the mutation effects at every site in each clique and the additivity properties between these sites within each clique. For this sake, we first considered the additive double-site mutants, and found each such mutant whose mutated two sites did not belong to any 3-clique. We then checked the site distribution for all the non-additive double-site mutants and found that the two mutation sites of each such mutant could be identified in some 3-clique community (Cliques 7 and 9). Spatial arrangements of clique members in the lysozyme network (Figure 3) indicate that 3-cliques have a relative uniform distribution, while larger cliques tend to form at the linker area that joins the two lobes of the enzyme.

**Table 1.** Typical 3-cliques found in a T4 phage lysozyme amino-acid network\*.

Clique	Nodes (Residue Index)
1	67, 4, 101, 70, 7, 104, 71, 11
2	34, 38, 42, 25
3	66, 46, 50, 54, 58, 27, 62
4	66, 70, 31
5	96, 99, 100, 103, 75, 78, 88, 91
6	91, 126, 95
7	<b>98, 152</b> , 156, 94
8	101, 105, 145, 149
9	128, 129, <b>132, 117</b> , 120, 125
10	138, 139, 142, 146, 149, 150
11	160, 148, 151

\* The bold numbers indicate non-additive double-site mutations found in reference [9].



**Figure 3.** The typical 3-cliques of an amino-acid network of the T4 phage lysozyme. The clique nodes are represented by  $\alpha$ -carbon as shown in a ball, and the labels correspond to the factions listed in Table 2.

$P_{ab}$  of the additive double-site mutations were determined to be 0 (see Table 2), indicating that the probability of these mutations in the wild-type T4 bacteriophage lysozyme was 0 in the presence of 3-cliques. Therefore, the results of the model calculations suggest that there is no non-additive effect in the combination of these mutations. This is consistent with the experimental observation [1]. Although  $P_{85,96}$  is not zero it is so small that this double-site mutation can hardly be assumed to be non-additive.  $P_{ab}$  of non-additive double-site mutations are remarkably different from those of additive double-site mutations, having a value varied mostly from 0.4 to 0.5 depending on the structural models of the enzyme. Essentially, the  $P_{ab}$  values calculated based on the wild-type structure are representative to measure the non-additive effects of the double-site mutations in the lysozyme. These results suggest that the additive effects of the double-site mutation can be closely related to the topological feature of the protein amino-acid network and are less dependent on the detailed physico-chemical interactions involved inside the protein.

**Table 2.**  $P_{ab}$  for double-site mutations derived from the different T4 phage lysozyme models. WT stands for wild type lysozyme structure (PDB code 2LZM [14]), K16E for the mutant structure (PDB code 1L42 [15]), R154E for the mutant structure (PDB code 1L47 [15]), the four structures of K16E/R154E, S117I, N132I, S117I/N132I are homology models derived by MODELLER [16] based on the structure of wild type lysozyme with the corresponding mutations of K16E/R154E, S117I, N132I, S117I/N132I.

Double Mutations	Models						
	WT	K16E	R154E	K16E/R154E	S117I	N132I	S117I/N132I
(116, 119)	0	0	0	0	0	0	0
(117, 132)	0.48	0.45	0.28	0.56	0.60	0.39	0.52
(119, 135)	0	0	0	0	0	0	0
(128, 131)	0	0	0.01	0	0	0	0
(135, 147)	0	0	0	0	0	0	0
(16, 119)	0	0	0	0	0	0	0
(16, 135)	0	0	0	0	0	0	0
(16, 154)	0	0	0	0	0	0	0
(38, 144)	0	0	0	0	0	0	0
(41, 131)	0	0	0	0	0	0	0
(85, 96)	0.07	0.05	0.05	0.05	0.12	0.11	0
(89, 96)	0	0	0	0	0	0	0
(98, 152)	0.43	0.40	0.24	0.03	0.02	0.56	0.12

It is interesting to notice that  $P_{ab}$  values do vary when measured with different mutant structures. For example, while  $P_{85,96}$  for the non-additive dual-mutation at sites 85 and 96 was determined close to 0 in three mutants, K16E, R154E, and K16E/R154E, it reaches 0.1 in both mutants S117I and N132I, and, at the same time, this value decays to zero again in dual-mutant S117I/N132I. Another case happens in A98V/T152S, while  $P_{98,152}$  has significant value for most of the examined structures, it almost drops to zero in mutants K16E/R154E and S117I. These results suggest that the non-additive effects of a double-site mutation might be affected by a third (or a fourth) mutation.

### 2.3. Eglin C

We examined the 3-cliques in the amino-acid network of a new Eglin c protein whose topology is distinct from that of lysozyme (Table 3). It is interesting to notice that compared with the two zero probability double-site mutations V18I/L27I and V34L/P58Y, the non-additive mutation V18A/V54A does show non-zero probability, indicating that the 3-clique relationship between the two mutation sites may still play a role in distinguishing non-additive mutations from additive one. However, the relatively small value of  $P_{18,54}$  suggests that some information is still missing to fully isolate this non-additive dual mutation from the other two, which is necessary for explaining the non-additive effects of the examined mutation sites.

**Table 3.**  $P_{ab}$  for double-site mutations for Eglin c protein. The simulation is based on wild type Eglin c (PDB code 1EGL [17]), and the distance is measured between the two mutation sites.

Double Mutation	Distance (Å)	$\Delta\Delta\Delta G$ (kJ/mol)	$P_{ab}$
(18, 27)	14.2	$-0.12 \pm 0.20$	0
(18, 54)	8.2	$-0.95 \pm 0.19$	0.11
(34, 58)	17.2	$-0.38 \pm 0.13$	0

### 2.4. Compared with the Maestro Calculations

Although our model is based on the 3-clique occurrence values  $P_{ab}$ , we only make an overall qualitative prediction by stating whether a given double-mutation is additive or non-additive. Recently, freely available webservers became available for modeling free energy changes due to single or multiple mutations. Tables 4 and 5 list free energy changes of dual-mutations in lysozyme and Eglin c predicted by MAESTROweb [18,19]. In both cases, *maestro* correctly found the non-additive

dual mutations as reported by the experiment measurements. However, the overall correlations between *maestro* values and the observed were small (<0.35). Particularly, *maestro* tends to assign bigger free-energy-change values for almost all dual-mutations whose two mutation sites are relatively close to each other. An extreme case is the lysozyme E128A/V131A mutation, for which *maestro* reports a relatively large free-energy change but only a negligible value is observed. Our  $P_{ab}$  calculations shows that these two residues do not form some 3-clique structure with a third one even though they are very close, thus they cannot form a stable interaction and have a very weak non-additive effect.

**Table 4.** The free energy changes of 13 double mutations in T4 phage lysozyme by MAESTROweb.

Mutations	$\Delta\Delta\Delta G$ (Kcal/Mol)	$\Delta\Delta\Delta G$ (Kcal/Mol)	$C_{\alpha}$ Atom Distance (Å)
S117I/N132I	−1.5	1.12	5.95
K16E/K135E	−0.2	0.40	33.69
K16E/R154E	−0.2	0.69	35.33
A41V/V131A	−0.11	−0.21	39.20
N116D/R119M	−0.1	−0.60	5.09
D89A/R96H	−0.1	0.87	8.35
R119E/K135E	−0.06	0.07	15.06
K16E/R119E	−0.06	0.32	39.63
E128A/V131A	−0.01	−1.11	5.38
K135E/K147E	0.2	−0.04	11.58
K85A/R96H	0.2	−0.40	10.73
S38D/N144D	0.3	0.14	29.07
A98V/T152S	2.7	1.84	6.35

**Table 5.** The free energy changes of 13 double mutations in T4 phage lysozyme by MAESTROweb.

Double Mutation	$\Delta\Delta\Delta G$ (Kj/Mol)	Distance (Å)
(18, 27)	−0.42	14.2
(18, 54)	0.11	8.2
(34, 58)	1.66	17.2

### 3. Materials and Methods

#### 3.1. Preprocessing the Experimental Mutation Data and Selection

The T4 bacteriophage lysozyme mutation data are taken from reference [9]. We first ignored those data lacking thermodynamic measurement or melting temperature changes. Then, the experimental data of double-site mutants were examined; they were kept for further analysis if two corresponding single-site mutations that make up the double-site mutation were present. The additive effect of a double mutation was measured based on the thermodynamic quantities as follows:

$$\Delta\Delta G_{sum} = \Delta\Delta G_i + \Delta\Delta G_j \quad (1)$$

$$\Delta\Delta\Delta G_{ij} = \Delta\Delta G_{ij} - \Delta\Delta G_{sum} \quad (2)$$

where  $\Delta\Delta G_i$ ,  $\Delta\Delta G_j$  are the Gibbs free energy changes due to the single point mutation at site  $i$  and  $j$ , and  $\Delta\Delta G_{ij}$  is that due to the double-site mutations at both sites  $i$  and  $j$ .  $\Delta\Delta G_{sum}$  evaluates the total effect due to the two single-site mutations in an ideal case when the two mutations are completely independent.  $\Delta\Delta\Delta G_{ij}$  measured the difference between the observed double-site mutation effect and the ideal effect when the corresponding two single-site mutations are additive. In other words,  $\Delta\Delta\Delta G_{ij}$  reveals how far a double-site mutation deviates from a perfect additive one. In this sense, the larger the absolute value of  $\Delta\Delta\Delta G_{ij}$ , the less likely the studied double-site mutation is additive and the greater its chance to be non-additive. To examine the possible dependence of the double-site mutation non-additive effect on the distance between the two involved sites, we defined the distance

as the length of a virtual edge linking the two C $\alpha$  atoms in the wild-type lysozyme structure (PDB code 2LZM [14]).

### 3.2. The Equilibrium Dynamics Conformation Ensemble

The 3D structure deposited in PDB usually captures the frozen snapshot of a protein in a typical crystal-packing state, which might be significantly different from its functioning conformations. Here, we derived a series of conformation ensembles using conventional molecular dynamics simulation techniques. All simulations were performed using the simulation package of GROMACS (version 4.5.4) [20] and the Oplsaa all atom field [21], with the lysozyme placed in a cubic water box. The starting lysozyme conformations with different mutations were taken from X-ray structures whose PDB entry codes are listed in reference [9]. For mutant structures whose structures were not solved and not available in PDB, we built their structural models based on that of wild-type protein (PDB code 2LZM [14]) using the homology modeling program of MODELLER version 9.4 [16]. All simulations were carried out in a temperature of 320 K, a pressure of 1 atm, a time step of 2 fs and a non-bond cutoff of 12 Å. For each simulation system, a certain number of Na<sup>+</sup> ions were added to neutralize the system and a layer of water molecules with thickness of about 1 nm was added to solvate the proteins. The PME (Particle, Mesh, Ewald, PME) algorithm was applied in calculating the long-range electrostatic interactions. After an initial minimization and a 1 ns steric relief equilibrium simulation, each system then performed a total 100 ns productive simulation. We collected the snapshots of each system every 10 ps, and recorded a total of 10 thousand conformations for each mutant, which were used for further analyses.

### 3.3. The Amino-Acid Interaction Network

To understand the mutation non-additive effect, we examined the topologies of the studied protein structures and focused on the amino acid networks. The amino-acid networks had been used in studying different biophysical problems, such as the protein folding, catalysis, as well as the mutation perturbations [22–25]. We built the network based on the amino acid interactions, which were determined using the program RING-2.0 (Residue Interaction Network Generator [13]). The program determined the most common types of physicochemical interactions that are indispensable in maintaining the protein 3D structure, including hydrogen bonds, disulfide bonds, Van der Waals interaction, electrostatic interaction,  $\pi$ - $\pi$  stacking interactions, and  $\pi$ -cation interactions. Table 6 listed typical parameters in determining the interactions and energetics. The network was created by using  $\alpha$  carbon atoms as nodes and the edges were generated between two neighboring nodes whose amino acids were found to form direct interactions. Thus, in such a network most amino acids are connecting to one another, and two amino acids are found either directly linked through an edge or indirectly connected via some intermediate linkers. In some cases, there are also scattered a few isolated nodes where the amino acids have no connection with any surrounding residues. The distance between two given amino acids was counted as the number of edges in the shortest path linking the two nodes within the network.

### 3.4. The $k$ -Clique Community in the Amino-Acid Network

One interesting topic concerning amino-acid networks is to examine the geometric pattern that emerges from protein structural topology and to analyze their meaning in the sense of biological function. We analyzed the network pattern using the Networkx package version 1.11 [26] developed for programming language Python version 3.6. A network can be divided into a few domains, and nodes inside a domain tend to form dense connections among one another while those belonging to different domains show very weak connections. The nodes in a domain build up the so-called community, which is further divided into a series of connected sub-graphs, called the  $k$ -cliques, using a clique percolation method [27,28].



Specifically, a  $k$ -clique is a complete sub-graph of  $k$  nodes in which each pair of nodes is connected by an edge, indicating a strong and intense mutual interaction among amino-acids on these nodes. Two cliques are regarded as adjacent if there are  $k-1$  edges linking the two cliques. Further, two-cliques are regarded as interconnected if one can find a way to connect one  $k$ -clique to the second one through several intermediate adjacent  $k$ -cliques. A collection of all interconnected  $k$ -cliques in a given network defines a  $k$ -clique community. In this sense, a network can be simplified by dividing it into a few  $k$ -clique communities. There may be nodes that belong to different clique communities that are not connecting with each other. Considering the dynamic feature of the protein structures, it is likely that the  $k$ -clique community distribution of the amino-acid network may be perturbed to some extent due to thermodynamic fluctuations. Thus, we calculated the ensemble of  $k$ -clique communities for each mutant structure derived from the molecular dynamic simulations, and to examine the mutation's non-additive effects we compared  $k$ -clique community distributions and their changes upon different mutations. For simplicity, in the rest of the paper a  $k$ -clique is simply referred to as a  $k$ -clique community.

**Table 6.** The type of interaction bonds and their energetic parameters used in defining protein amino-acid interaction network.

Bonds	Cutoff (Å)	Energy (KJ/Mol)
Hydrogen bond	3.5	17.0/40.0/115.0 (on distance)
Van der Waals	0.5	6.0
Disulfide bond	2.5	167.0
electrostatics	4.0	20.0
$\pi$ - $\pi$ stacking	6.5	9.4
$\pi$ -cation	5.0	9.6

Note: the distance in hydrogen bonds refers to that between the hydrogen donor and acceptor atoms. The distance of van der Waals interaction is that between the surfaces of two atoms. The distance in sulfur bonds refers to that between the two sulfur atoms. The distance used in electrostatic interaction calculations are measured between mass centers of the two oppositely charged groups. The distance in  $\pi$ - $\pi$  stacking interaction refers to those of the geometric centers of benzene rings of the aromatic residues. The distance in a  $\pi$ -cation interaction is measured from the mass center of the positively charged group in a residue to that of the benzene ring in another residue. The energy of the action is averaged over the various cases of the same type of interaction, which is a rough approximation of the corresponding real interaction.

### 3.5. The 3-Clique and Non-Additive Effects in Double-Site Mutations

Considering that an amino-acid within the same clique tends to have a tighter connection than does an amino-acid belonging to different cliques, it is interesting to examine whether or not mutations with higher non-additive effects tend to be locatable in the same clique. Specifically, for a given double-site mutation we calculated the probability by which the two mutation sites belong to the same 3-clique and the non-additive effect involving the double-site mutations as follows:

1. Generating an ensemble of protein conformation from a 100 ns equilibrium dynamics simulation of the studied enzyme. The combination of ten thousand snapshots in aqueous solution should be a better representation of the interactions within the protein in functioning conditions.
2. Determining 3-cliques for a network of each snapshot using the Networkx package [26].
3. Calculating the proportion  $P_{ab}$  of snapshots in the ensemble in which the two sites ( $a$ ,  $b$ ) of the studied double-site mutation belong to a 3-clique:

$$P_{ab} = \frac{\sum_{i=1}^N C_{ab}(i)}{N} \quad (3)$$

where  $N$  represents the total number of equilibrium snapshots, here it equals to  $10^4$ , and

$$C_{ab}(i) = \begin{cases} 1 & a \text{ and } b \text{ in a 3-clique} \\ 0 & \text{else} \end{cases} \quad (4)$$

$P_{ab}$  measures the probability that two sites ( $a, b$ ) are kept in some 3-clique due to either direct or indirect interactions among amino-acid interactions. The closer  $P_{ab}$  is to 1, the more likely  $a$  and  $b$  tends to have a tight connection from the topological perspective. In this work, the condition of  $P_{ab} \geq 0.1$  is used to evaluate if site  $a$  and  $b$  form some stable 3-clique.

### 3.6. The Validation Models

To evaluate the relationship between clique-probability  $P_{ab}$  and the non-additive effects of a double-site ( $a, b$ ) mutation, which is quantified by the value of additivity, we first investigated the additivities of the 13 double-site mutations of the T4 phage lysozyme, and then determined  $P_{ab}$  for each of them from the simulations and network modeling. We also examined the Eglin c protein [3] which has a distinct structural topology with that of the T4 phage lysozyme. Finally, we also studied double-site mutations in which the two involved sites are far from each other but have a high  $P_{ab}$  value, which, according to our prediction, might have a high probability to have non-additivity effect.

## 4. Conclusions

Protein mutation effects have become a popular topic in cell biology due to recently developed deep scanning techniques, which creates large-scale mutagenesis data that associates intrinsic protein structures and functions with the consequences of relevant genetic variation [29]. A critical question that arises from this scenario is how natural selection works with the innumerable yet almost random mutations in the so-called evolution process? In this paper, we examined possible intrinsic correlations between random protein mutations based on protein structural network calculations. We analyzed the additivity effects of 13 double-site mutations of the T4 bacteriophage lysozyme, and found that mutations at distal sites are usually strongly additive while those occurring at neighboring sites can be either additive or non-additive. To systematically estimate the non-additive effects of double-site mutations, we investigated the amino-acid network structures for each mutant and determined the topological quantities of these networks. We generated equilibrium configuration ensembles of the studied proteins using conventional simulations and built the amino-acid network for each structure. We then analyzed the topological characteristics of the protein networks, such as the distribution of  $k$ -cliques, and found significant correlation between 3-faction associations and the double-site mutant additivity: non-additive mutations tended to happen between sites belonging to the same 3-clique structure. It was found that the clique model could significantly separate non-additive double-site mutations from those additive ones for the examined proteins. Our calculations also suggested that such correlation probabilities can be changed to some extent by applying a third mutation.

Although the faction group model used here is very simple it does work very well for lysozyme structures. However, we also noticed that the model cannot explain mutation non-additive effects for some different proteins, such as myoglobin [30]. Another weak point with the model is that it tends to create very few 3-cliques for many proteins, especially for those protein whose network topology are relatively sparse, which usually resulted in false negative predictions. It becomes even more complicated when considering the perturbation due to a third mutation. Thus, we expect to refine the models in the near future by combining the simple network analysis as shown in this work with a detailed physico-chemistry characterization and provide a fruitful understanding of protein mutation effects. Considering that counter-examples always exist in biological phenomena, we regard our model as a simple yet rudimentary picture to understand the mutation-correlation puzzle, on which many more details may be added for a deeper understanding.

**Author Contributions:** Conceived and designed the experiments: D.M. Performed the experiments: R.C. Analyzed the data R.C., D.M. Contributed reagents/materials/analysis tools: R.C., D.M. wrote the paper: D.M. Organized and supervised this work: H.H.

**Acknowledgments:** One of the author (D.M.) thanks Hong Qian in University of Washington, Seattle for his helpful discussion on the lysozyme mutation data. This work was supported, in part, by the National Key Research and Development Program of China (Grant No. 2017YFC1600900) and by the Key University Science Research Project of Jiangsu Province (Grant No. 17KJA180005).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Wells, J.A. Additivity of mutational effects in proteins. *Biochemistry* **1990**, *29*, 8509–8517. [[CrossRef](#)] [[PubMed](#)]
2. Sandberg, W.S.; Terwilliger, T.C. Engineering multiple properties of a protein by combinatorial mutagenesis. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8367–8371. [[CrossRef](#)] [[PubMed](#)]
3. Boyer, J.A.; Clay, C.J.; Luce, K.S.; Edgell, M.H.; Lee, A.L. Detection of native-state nonadditivity in double mutant cycles via hydrogen exchange. *J. Am. Chem. Soc.* **2010**, *132*, 8010–8019. [[CrossRef](#)] [[PubMed](#)]
4. Nicholson, H.; Becktel, W.; Matthews, B. Enhanced protein thermostability from designed mutations that interact with  $\alpha$ -helix dipoles. *Nature* **1988**, *336*, 651–656. [[CrossRef](#)] [[PubMed](#)]
5. Zhang, X.-J.; Baase, W.A.; Shoichet, B.K.; Wilson, K.P.; Matthews, B.W. Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. *Protein Eng.* **1995**, *8*, 1017–1022. [[CrossRef](#)] [[PubMed](#)]
6. Daopin, S.; Alber, T.; Baase, W.; Wozniak, J.; Matthews, B. Structural and thermodynamic analysis of the packing of two  $\alpha$ -helices in bacteriophage T4 lysozyme. *J. Mol. Biol.* **1991**, *221*, 647–667. [[CrossRef](#)]
7. Pucci, F.; Bourgeas, R.; Rooman, M. High-quality Thermodynamic Data on the Stability Changes of Proteins Upon Single-site Mutations. *J. Phys. Chem. Ref. Data* **2016**, *45*. [[CrossRef](#)]
8. Lipscomb, L.A.; Gassner, N.C.; Snow, S.D.; Eldridge, A.M.; Baase, W.A.; Drew, D.L.; Matthews, B.W. Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme. *Protein Sci.* **1998**, *7*, 765–773. [[CrossRef](#)] [[PubMed](#)]
9. Baase, W.A.; Liu, L.; Tronrud, D.E.; Matthews, B.W. Lessons from the lysozyme of phage T4. *Protein Sci.* **2010**, *19*, 631–641. [[CrossRef](#)] [[PubMed](#)]
10. Tian, J.; Wu, N.; Chu, X.; Fan, Y. Predicting changes in protein thermostability brought about by single- or multi-site mutations. *BMC Bioinform.* **2010**, *11*, 370. [[CrossRef](#)] [[PubMed](#)]
11. Pires, D.E.V.; Ascher, D.B.; Blundell, T.L. mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **2014**, *30*, 335–342. [[CrossRef](#)] [[PubMed](#)]
12. Dehghanpoor, R.; Ricks, E.; Hursh, K.; Gunderson, S.; Farhoodi, R.; Haspel, N.; Hutchinson, B.; Jagodzinski, F. Predicting the Effect of Single and Multiple Mutations on Protein Structural Stability. *Molecules* **2018**, *23*. [[CrossRef](#)] [[PubMed](#)]
13. Piovesan, D.; Minervini, G.; Tosatto, S.C. The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* **2016**, *44*, W367–W374. [[CrossRef](#)] [[PubMed](#)]
14. Weaver, L.H.; Matthews, B.W. Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* **1987**, *193*, 189–199. [[CrossRef](#)]
15. Sun, D.; Soderlind, E.; Baase, W.A.; Wozniak, J.A.; Sauer, U.; Matthews, B.W. Cumulative site-directed charge-change replacements in bacteriophage T4 lysozyme suggest that long-range electrostatic interactions contribute little to protein stability. *J. Mol. Biol.* **1991**, *221*, 873–887. [[CrossRef](#)]
16. Marti-Renom, M.A.; Stuart, A.C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325. [[CrossRef](#)] [[PubMed](#)]
17. Hyberts, S.G.; Goldberg, M.S.; Havel, T.F.; Wagner, G. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci. Publ. Protein Soc.* **1992**, *1*, 736–751. [[CrossRef](#)] [[PubMed](#)]
18. Laimer, J.; Hofer, H.; Fritz, M.; Wegenkittl, S.; Lackner, P. MAESTRO—Multi Agent Stability Prediction upon Point Mutations. *BMC Bioinform.* **2015**, *16*, 116. [[CrossRef](#)] [[PubMed](#)]
19. Laimer, J.; Hiebl-Flach, J.; Lengauer, D.; Lackner, P. MAESTROweb: A web server for structure-based protein stability prediction. *Bioinformatics* **2016**, *32*, 1414–1416. [[CrossRef](#)] [[PubMed](#)]

20. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M.R.; Smith, J.C.; Kasson, P.M.; van der Spoel, D.; et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854. [[CrossRef](#)] [[PubMed](#)]
21. Jorgensen, W.L.; Maxwell, D.S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236. [[CrossRef](#)]
22. Ghosh, A.; Brinda, K.V.; Vishveshwara, S. Dynamics of lysozyme structure network: Probing the process of unfolding. *Biophys. J.* **2007**, *92*, 2523–2535. [[CrossRef](#)] [[PubMed](#)]
23. Dhulesia, A.; Gsponer, J.; Vendruscolo, M. Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. *J. Am. Chem. Soc.* **2008**, *130*, 8931–8939. [[CrossRef](#)] [[PubMed](#)]
24. Axe, J.M.; Yezdimer, E.M.; O'Rourke, K.F.; Kerstetter, N.E.; You, W.; Chang, C.E.; Boehr, D.D. Amino acid networks in a ( $\beta/\alpha$ )<sub>8</sub> barrel enzyme change during catalytic turnover. *J. Am. Chem. Soc.* **2014**, *136*, 6818–6821. [[CrossRef](#)] [[PubMed](#)]
25. Boehr, D.D.; Schnell, J.R.; McElheny, D.; Bae, S.H.; Duggan, B.M.; Benkovic, S.J.; Dyson, H.J.; Wright, P.E. A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. *Biochemistry* **2013**, *52*, 4605–4619. [[CrossRef](#)] [[PubMed](#)]
26. Schult, D.A.; Swart, P. Exploring network structure, dynamics, and function using NetworkX. In Proceedings of the 7th Python in Science Conferences (SciPy 2008), Pasadena, CA, USA, 19–24 August 2008; pp. 11–16.
27. Derenyi, I.; Palla, G.; Vicsek, T. Clique percolation in random networks. *Phys. Rev. Lett.* **2005**, *94*, 160202. [[CrossRef](#)] [[PubMed](#)]
28. Palla, G.; Derenyi, I.; Farkas, I.; Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **2005**, *435*, 814–818. [[CrossRef](#)] [[PubMed](#)]
29. Fowler, D.M.; Fields, S. Deep mutational scanning: A new style of protein science. *Nat. Methods* **2014**, *11*, 801–807. [[CrossRef](#)] [[PubMed](#)]
30. Lin, L.; Pinker, R.J.; Phillips, G.N.; Kallenbach, N.R. Stabilization of myoglobin by multiple alanine substitutions in helical positions. *Protein Sci. Publ. Protein Soc.* **1994**, *3*, 1430–1435. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).