



Article

Effects of Residual Composition and Distribution on the Structural Characteristics of the Protein

Qiaoling Song, Zhenan Wu, Chenghao Jin, Zhichao Yu, Peng Xu  and Zhouting Jiang *

Department of Applied Physics, China Jiliang University, Hangzhou 310018, China

* Correspondence: z.jiang@cjlu.edu.cn; Tel.: +86-571-8683-5749

Abstract: The effect of ratio and consecutive number of hydrophobic residues in the repeating unit of protein chains was investigated by MD simulation. The modified off-lattice HNP model was applied in this study. The protein chains constituted by different HNP ratios or different numbers of consecutively hydrophobic residues with the same chain length were simulated under a broad temperature range. We concluded that the proteins with higher ratio or larger number of sequentially hydrophobic residues present more orientated and compact structure under a certain low temperature. It is attributed to the lower non-bonded potential energy between H-H residual pairs, especially more hydrophobic residues in a procession among the protein chain. Considering the microscopic structure of the protein, more residue contacts are achieved with the proteins with higher ratios and sequential H residues under the low temperature. Meanwhile, with the ratio and consecutive number of H residues increasing, the distribution of stem length showed a transition from exponential decline to unimodal and even multiple peaks, indicating the specific ordered structure formed. These results provide an insight into 3D structural properties of proteins from their residue sequences, which has a primary structure at molecular level and, ultimately, a practical possibility of applying in biotechnological applications.



Citation: Song, Q.; Wu, Z.; Jin, C.; Yu, Z.; Xu, P.; Jiang, Z. Effects of Residual Composition and Distribution on the Structural Characteristics of the Protein. *Int. J. Mol. Sci.* **2022**, *23*, 14263. <https://doi.org/10.3390/ijms232214263>

Academic Editors: Istvan Simon and Csaba Magyar

Received: 13 October 2022

Accepted: 14 November 2022

Published: 17 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: HNP model; molecular dynamics simulation; hydrophobicity

1. Introduction

Proteins, as compact polymers, show various structural diversity since they have numerous internal degrees. The biological function of a protein strongly depends on its unique structure determined by complicated interactions among residues. The protein chain conformation, which leads to multiple mechanical and thermodynamical properties, even to biological functions, is extensively interesting for industrialists and scientists. The structure formation of polymer chains, such as polymer crystallization, protein folding, and self-assembly of block copolymers, has recently become the focus of attention in chemistry, physics, biology, and material science [1–3]. From the viewpoint of polymer statistics, every polymer chain may occasionally return to itself. That is, some points of the free chain trajectory may come within a short reach of one and another, forming a closed loop, also named contacts. The folding of protein chains into compact, unique three-dimensional structures is guided and stabilized by intramolecular interactions between the constituent amino acid residues along the chain [4]. According to Anfinsen's denaturation and renaturation experiments, it is known that the three-dimensional conformation of a protein is directly determined by its primary structure, the amino acid sequence, according to the global minimum on free energy landscape of a protein [5]. So, in the case of proteins, the residue–residue contacts, which are mainly induced by van der Waals interactions, play an important role in the protein folding process. The residue contacts are defined as two heavy atoms spatially close but far away in amino acid sequences. Residue contact has been proposed to describe the effect of long-range interactions, such as hydrogen bond, hydrophobic–hydrophobic, aromatic–aromatic, or aromatic–polar interactions on

the structural level of the protein [6,7]. However, the determination of the three-dimensional structure of a protein molecule from its amino acid sequence remains difficult because the energy that stabilizes the folded conformation is poorly understood and the number of possible structures is extensive [8]. As each protein has a unique and exact amino acid sequence, it is particularly important to study the effect of amino acids sequence on protein structure.

Computational simulation is a powerful method to discover the detailed mechanism of structural information at an atomistic level. Currently, fruitful research has been carried out in the field of structural formation of the macromolecular system by computer simulation to overcome experimental difficulties [9–11]. Despite the rapid progress of computer power and the efforts made by many researchers, numerous basic questions, such as the nature of the interactions involved in protein folding, the thermodynamic states involved as a protein folds, and the structural properties of each thermodynamic state, remain unanswered [12]. Molecular dynamics (MD) simulations at atomic resolution could, in principle, monitor molecular interactions with a high accuracy. However, the simulation on realistic all-atom representations of a protein with the numerous freedom degrees limits the running time because of high computational costs. To solve this problem, the coarse-grained model is applied to represent proteins, which preserves the biological information of the protein and saves computational resources. In the coarse-grained model, each protein is viewed as a chain of beads, each bead representing an amino acid residue and each amino acid residue is replaced by a corresponding C^α atom [13]. One of the widely used simplified lattice models is the HP model. Two kinds of beads named H (hydrophobic residues) and P (polar residues) construct a protein chain [14]. The hydrophobic energies are attractive in a coarse-grained model, which means H-H contacts play a major role in the protein-structured process. Later, protein engineering experiments suggest that instead of two, three kinds of residues can be effectively substituted for the twenty amino acids [15]. Miyazawa–Jernigan (MJ) modified the interaction matrix based on this model [16]. According to the MJ matrix, the standard HP lattice model was improved by the off-lattice model with three kinds of residues, namely hydrophobic residues (H), neutral residues (N), and hydrophilic residues (P) [17]. The nonlocal interaction term depends on the type of residues. The interaction between H-H residues has a deep primary minimum that gives rise to defining the ordered structure of a protein. The other residue pairs interact repulsively with different strengths. In our model, the backbone of the protein was constituted by simplified residues in a continuous space instead of a cubic lattice. While lacking the chemical detail of all-atom representations, the simplified models are able to characterize both kinetics and thermodynamics during the protein folding process [18–20].

In this work, we apply MD simulations on protein chains with different chain lengths, residue composition, and repeat units. The aim of this study is to investigate the effect of hydrophobic properties of residues on the structural features and conformational transition of proteins. The rest of this paper is organized as follows. The simulation results regarding the effects of the chain length, ratio of hydrophobic residues, and repeat units are discussed in Section 2. The methodology, including the construction of the protein chain model and molecular dynamics simulation details, are presented in Section 3, and the paper is concluded in Section 4.

2. Results and Discussion

2.1. Effect of the Chain Length

2.1.1. Energy

The chain length dependence of potential energy of each bond in a protein chain with specific H3N1P1 as repeating segment is investigated. According to the simulation model, the total bonded energies of each protein are summed of all the bonds. Meanwhile, the non-bonded energy of the protein sample is summed of all residue pairs. To compare the change of potential energy with the temperature rather than the chain length, the four potential energies of the whole protein chain were divided by the number of bonds, the number of bond angles, or the number of dihedral angles of each protein chain with different chain lengths.

Figure 1a,b plot the simulation results of bonded potential energy of each bond versus the reduced temperature T^* with chain length $N = 50, 120, 240,$ and 300 residues, respectively. As shown in Figure 1a, the change of bond-stretching/bending energy per bond as the function of temperature is independent of chain length. Both energies are almost linearly increasing with the temperature increases, which is approved by the theorem of energy equipartition, i.e., the average energy associated with each individual degree of freedom (bond length/angle) is linearly proportional to the system temperature. The bond-torsional energy per bond of proteins with different chain lengths as the function of temperature is shown in Figure 1b. As the similar results show in Figure 1a, the change of bond-torsional energy per bond as the function of temperature is independent of chain length. Meanwhile, the increasing tendency of the bond-torsional energy is more obvious when the protein at the low-temperature region as the deviation from a straight line. The increased non-bonded energy of each bond with the increasing temperature is shown in Figure 1c, except for the protein at the low temperature. Additionally, the increasing tendency is more obvious when the protein with longer chain as the number of residue–residue pairs is increased with the chain length. The simulation results of the total potential energy including bonded and non-bonded interactions with different chain lengths as the function of temperature is plotted in Figure 1d. The same tendency that the total potential energy is increasing with temperature increases is also presented in Figure 1d. So, we can conclude that the potential energy of the whole chain is only dependent on the chain length when the protein at the certain temperature. In addition, the non-bonded interactions are related to the hydrophobic property of primary sequence of the protein. As the equilibrium state of each protein is mainly due to non-bonded interactions, it will be discussed in detail in the following context.

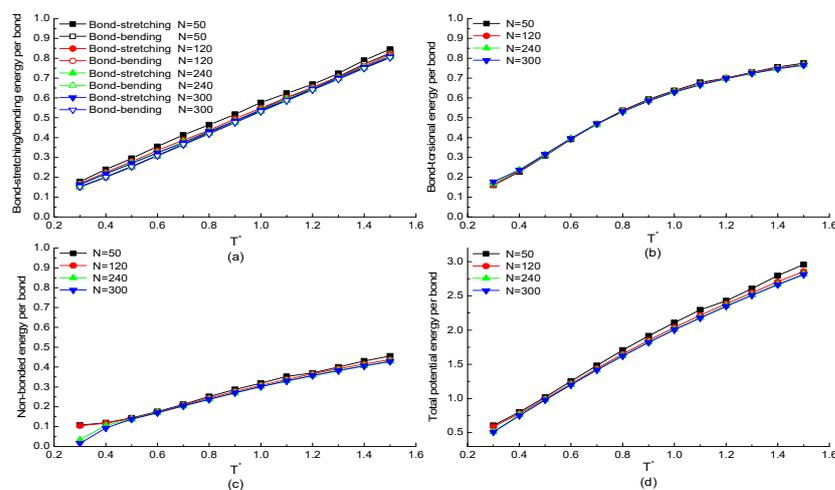


Figure 1. The temperature dependencies of the potential energies (a) bond-stretching/bending energy, (b) bond-torsional energy, (c) non-bonded energy, (d) total potential energy of each bond in H3N1P1 proteins with different chain lengths.

2.1.2. Radius of Gyration

The radius of gyration, which shows the structural information of a single protein during the simulation process, is mainly discussed. The temperature dependence of the total and components of radius of gyration of a protein chain with the same repeating segment H3N1P1 but different chain lengths are plotted in Figure 2. The curves in Figure 2a show that the z-component of radius of gyration R_{gz} of H3N1P1 protein with chain length $N = 50$ residues is increased rapidly with the temperature decreasing when $T^* < 0.8$ after the gradual increase in the high temperature region. Meanwhile, R_{gx} (or R_{gy}) keeps constant value or increases slightly with the temperature decreasing, and then has a decreasing trend especially $T^* < 0.6$. At low temperatures $T^* < 0.6$, the difference between R_{gx} (or R_{gy}) and R_{gz} becomes more and more obvious as the temperature reduces, which indicates that the protein chain presents anisotropic configuration. Figure 2b–d show the simulation

results on the H3N1P1 proteins with chain length of 120, 240, and 300 beads. The curves of total and components of radius of gyration show the similar trends shown in Figure 2a. Comparing these four figures, for H3N1P1 proteins, the longer the chain length is, the larger the radius of gyration presents. These results indicate that *trans* state is dominated at low temperatures, which makes the ordered structure with obvious anisotropy. Instead, *gauche* states are located everywhere, and a protein takes a random coil configuration when the protein is under the high temperature condition.

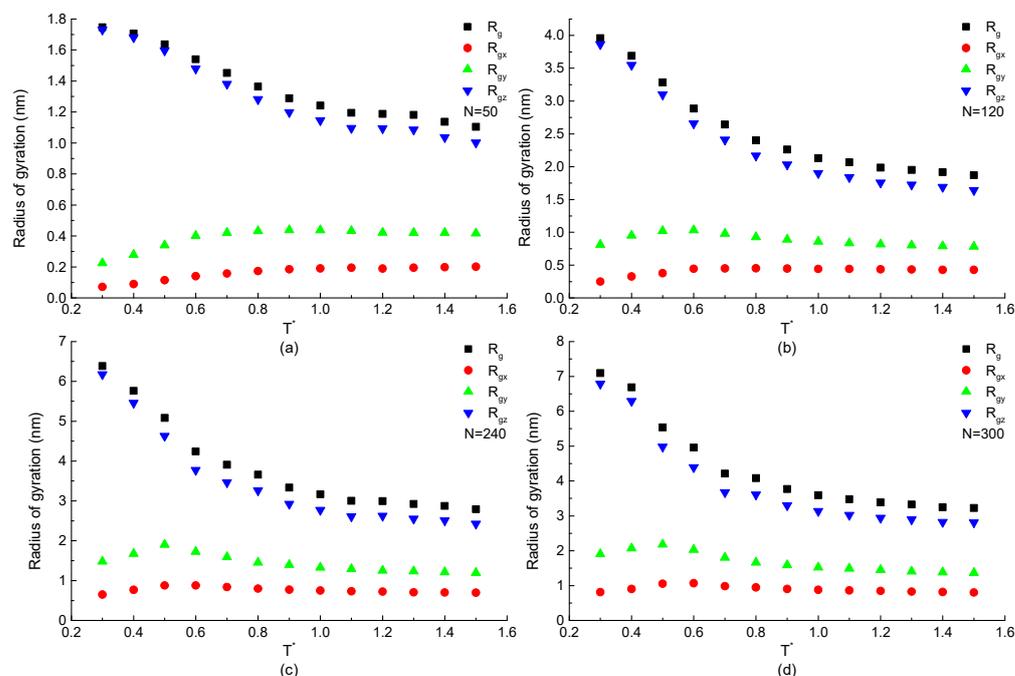


Figure 2. The total and components of radius of gyration versus reduced temperature T^* of H3N1P1 proteins with chain length (a) $N = 50$, (b) $N = 120$, (c) $N = 240$, and (d) $N = 300$ residues.

2.1.3. Global Orientational Order Parameter

Figure 3 shows the global orientational order parameter P of H3N1P1 proteins with the chain length of 50, 120, 240, and 300 residues. It takes a value near 0 under high temperatures, which indicates the protein takes a random coil configuration. After the value of P gradually increases with temperature decrease in the range of $T^* > 0.8$, it increases sharply with decreasing temperature leading to a large growth rate of the orientational order. The increasing order parameter P with decreasing temperature indicates the transition of the protein chain from a random coil to an ordered structure. The simulation results on H3N1P1 proteins with different chain lengths show the same tendency. Meanwhile, the shorter chain takes the larger value of P , which indicates the bonds are almost paralleled, but the longer protein is partially paralleled under a certain temperature as more continuous bonds influence each other to hardly form a good orientation structure.

We can summarize briefly that the large size of a protein constructed by more residues has the higher non-bonded interactions between residue–residue pairs and low global orientational order indicating the partially paralleled bonds among the protein. The conformational transition from random coil to ordered structure occurs when the temperature drops low enough, and the transition temperature is independent of chain length.

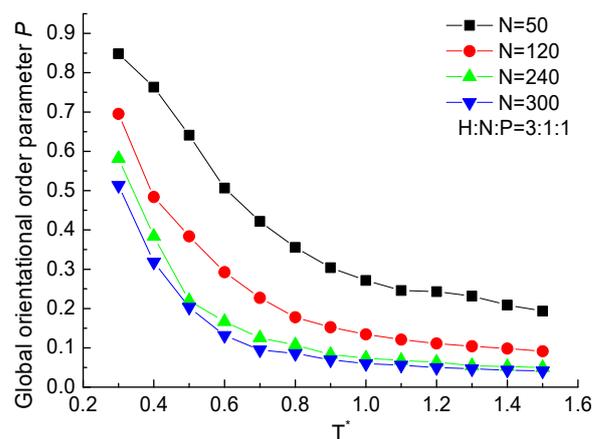


Figure 3. The global orientational order parameter P versus reduced temperature T^* of H3N1P1 proteins with different chain lengths.

2.2. Effect of the HNP Ratios

2.2.1. Energy

In this section, the MD simulation results on the proteins with the same chain length but different residue ratios are mainly discussed. As mentioned above, the non-bonded energy plays a main role to present a specific structure. Then, the temperature dependence of the non-bonded energy of proteins with the different residue ratios are plotted in Figure 4. As the same tendency shown in Figure 1c, the value of the non-bonded energy decreases with the decreasing temperature in all three types of proteins with H4N1P1, H3N1P1, and H4N2P2 residue components. As shown in Figure 4a,b, the higher ratio of hydrophobic residues (H) among the protein chain, the lower non-bonded energy is obtained when the protein is under a certain temperature. The straightforward reason is that only the interaction between H-H residue pairs could be a minus value according to the Equation (4) and parameter values in table in Section 3. When the higher proportion of hydrophobic residues in the repeating unit leads to the larger number of H-H pairs with the attractive interaction between hydrophobic residues, and the non-bonded energy of the whole protein chain is correspondingly lower. It means that the interaction between hydrophobic residues plays an important role in stabilizing the protein structure. Comparing the simulation results in Figure 4a,b, the non-bonded energy decreases more significantly in the case of longer protein $N = 240$ and even presents the minus value under the low temperature when the ratio of hydrophobic residues is high enough.

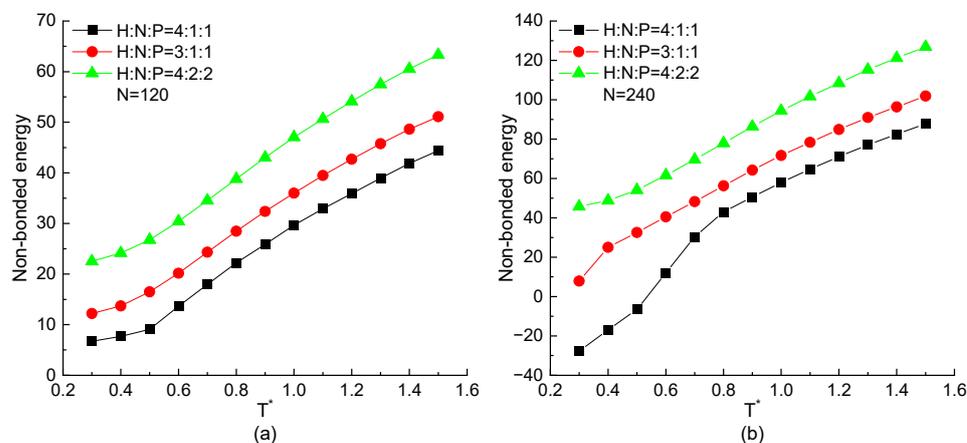


Figure 4. The non-bonded energy versus reduced temperature T^* of H4N1P1, H3N1P1, and H4N2P2 proteins with chain length of (a) $N = 120$, (b) $N = 240$.

2.2.2. Radius of Gyration

The curves of the radius of gyration as the function of temperature of H4N1P1, H3N1P1, and H4N2P2 proteins with the chain length of 120 and 240 beads are plotted in Figure 5. All curves show that the radius of gyration of the protein chain increases with the decrease of temperature. This tendency is independent of chain length and repeating unit. But comparing the proteins constructed by different repeating units, for the chain with higher proportion of H residues in a repeating unit may lead to the lower value of the radius of gyration presenting the compact structure comparatively. On the other hand, the structural size of the protein constructed by the low percentage of hydrophobic residues changes more obviously than the one formed by high percentage of H beads. It means the configuration of the protein with low ratio of hydrophobic residues is flexible and easily influenced by temperature. The protein constituted by more hydrophobic residues could form a relatively stable structure that is hardly influenced by temperature. It also demonstrates that more H-H pairs interactions could maintain the stability of the protein structure.

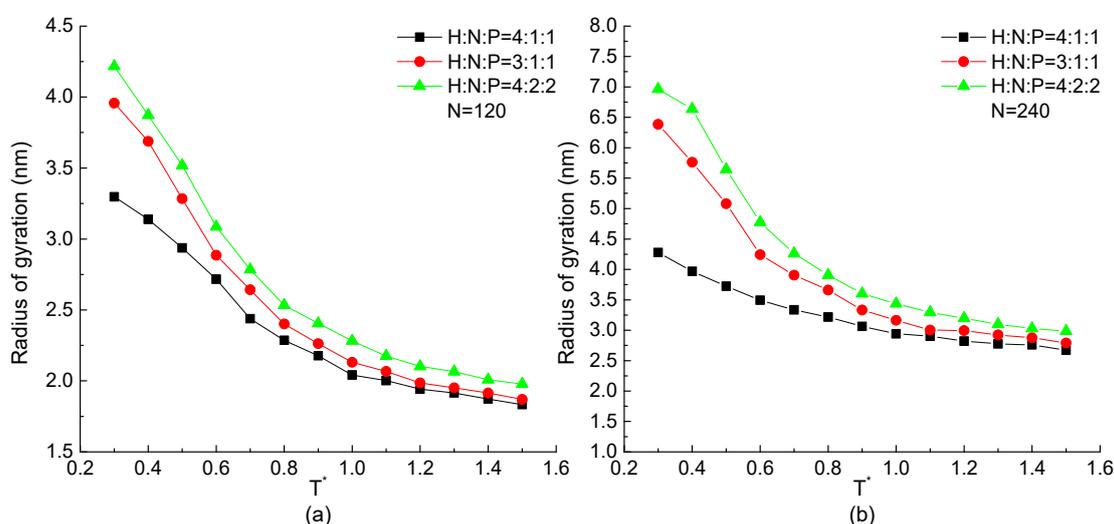


Figure 5. The radius of gyration versus reduced temperature T^* of H4N1P1, H3N1P1, and H4N2P2 with chain length of (a) $N = 120$, (b) $N = 240$.

2.2.3. Global Orientational Order Parameter

In order to reveal the effect of HNP proportion on the orientation of the protein chain, the temperature dependence of global orientational order parameter P with chain length $N = 120$ and 240 beads but different HNP ratios are plotted in Figure 6. When the protein is in the high temperatures ($T^* > 0.8$) condition, it takes a value near 0, indicating the chain takes a random coil configuration and the numerical difference of P among three proportional ratios of HNP residues is slight, especially in the case of longer chain. When $T^* < 0.8$, P increases rapidly with the decreasing temperature, and the higher the H residue proportion in a repeating unit, the higher value of P , which indicates a more ordered structure formed. For the longer protein chain, such tendency is more obvious and the numerical difference between protein chains with different HNP proportions is larger than the ones in short protein chains at the low temperature region.

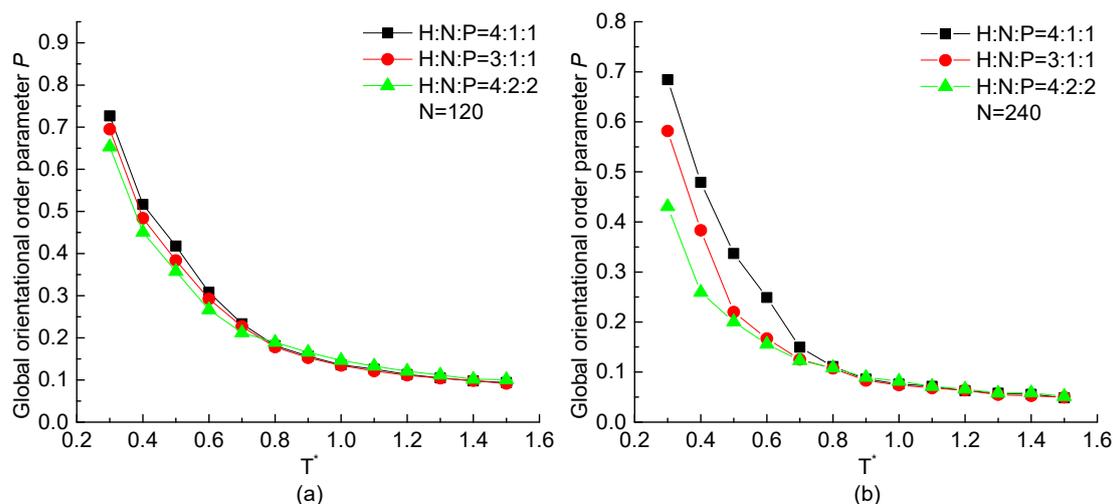


Figure 6. The global orientational order parameter P versus reduced temperature T^* of H4N1P1, H3N1P1, and H4N2P2 with chain length of (a) $N = 120$, (b) $N = 240$.

By analyzing the non-bonded energy, the radius of gyration and the global orientational order parameter of the protein, results show that the proteins with the same chain length but different HNP segment proportions in a repeating unit present the different size and orientational characteristics. The one with higher proportion of H residues receives the lower non-bonded energy, the smaller value of radius of gyration, but the larger value of global orientational order, which indicates that the protein chains with a higher proportion of H have more compact and parallel conformation than others.

2.3. Effect of the Serial Number of Three Types of Residues with Same Ratio

2.3.1. Energy

The protein chain with the same HNP distribution ratio but different serial residue numbers in the repeating unit are mainly discussed in this part. The non-bonded energies of the protein with different chain lengths as the function of temperature are shown in Figure 7a,b. In these figures, the tendency that the non-bonded energy keeps decreasing with the decreasing temperature is the same in all kinds of proteins regardless of the chain length, residue distribution ratio, and serial number of the same type of residues. For the certain chain length, the protein chain with the higher proportion of H residues in a repeating unit has the lower non-bonded potential energy, which is the same result as mentioned above. On the other hand, the curves with a hollow symbol are always lower than the ones with a solid symbol in any simulation cases. It means that for the protein chains with the same HNP proportion but different serial numbers of hydrophobic, neutral, or hydrophilic residues in the repeating unit, the more H residues in succession along the protein chain can obtain the lower non-bonded energy. It is due to the series of hydrophobic residues forming consecutive contacts easily. Additionally, the negative interaction between hydrophobic residues leads to the lower value of non-bonded energy. Comparing the results presented in Figure 7a,b, although the chain length is doubled, the same tendency that the chain connected by more consecutive H residues has the lower non-bonded energy than the ones with the same ratio but smaller number of consecutive H residues is achieved, especially when the is protein under the low temperature region. We also noticed that in the high temperature region ($T^* > 0.7$), the non-bonded energy of the H8N4P4 protein is little lower than the ones of the H6N2P2 protein. Here, although the ratio of H residues is lower, the serial number of H residues is larger. It is the synergistic effect of the residue distribution factor and the consecutive number of hydrophobic residues.

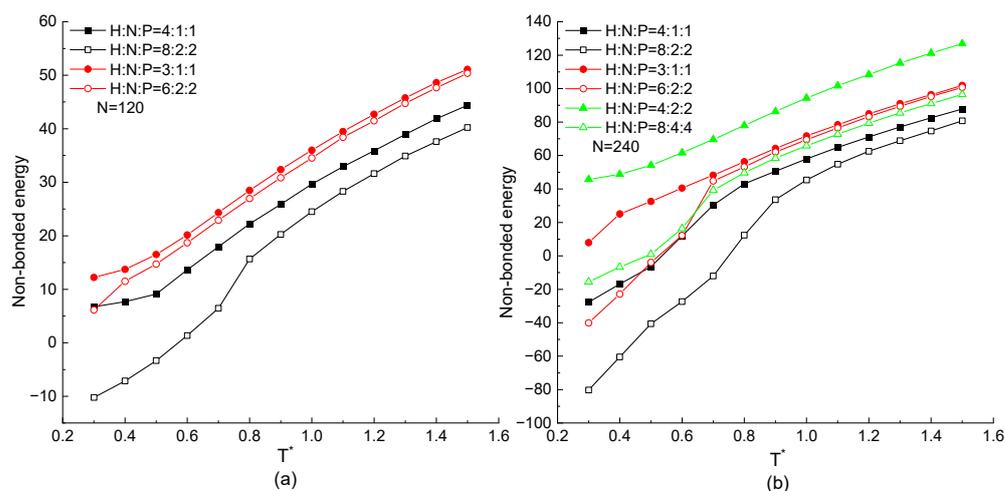


Figure 7. The non-bonded energy versus reduced temperature T^* of different ratio and serial number of H, N, and P residues with protein chain length (a) $N = 120$, (b) $N = 240$.

2.3.2. Radius of Gyration

The total size of the protein chain constituted by $N = 120$ or 240 residues with the same H, N, and P distribution proportion but different serial residue numbers in the repeating unit is analyzed in this section. The radius of gyrations of each protein chain as the function of temperature are shown in Figure 8. In these two figures, the radius of gyration increases with the decreasing temperature in any simulation cases, which is the same result as in Figures 2 and 5. The protein chain with higher H residue distribution takes the lower value of radius of gyration, which is also obtained in Figure 8. Considering on the effect of serial number of three types of residues with same ratio in a repeating unit, we found that the hollow symbol curves are always lower than the same shape of solid curves. In both Figure 8a,b, the minimum radius of gyration occurs on the protein chain with the repeating unit of H8N2P2, which has a higher H ratio and more sequential hydrophobic residues. It means that for a protein with the same HNP proportion but different repeating units, the one with higher numbers of consecutive H in a repeating unit takes the lower value of radius of gyration because of the compact conformation formed by the mutual attraction between consecutive hydrophobic residues. Briefly, it concluded that the protein chain with higher proportion of H and greater number of consecutive H residues has a tighter configuration.

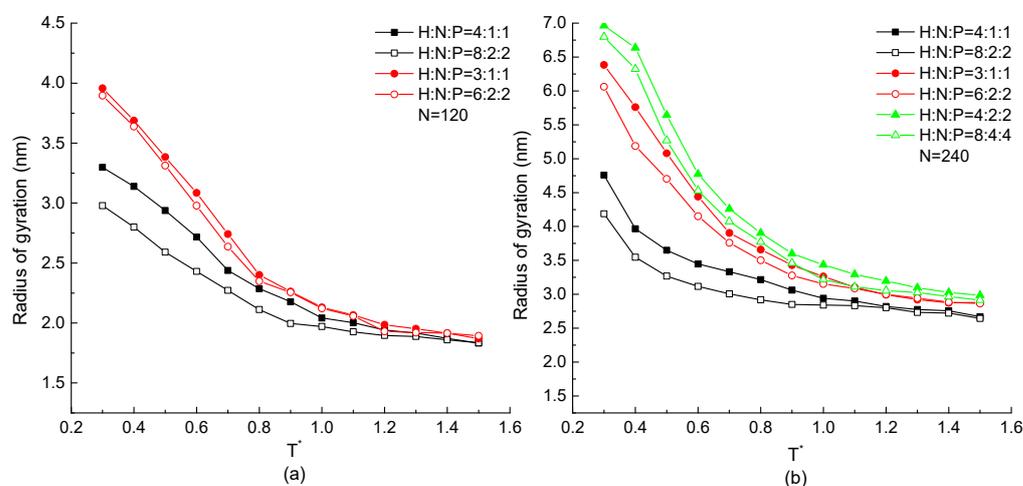


Figure 8. The radius of gyration versus reduced temperature T^* of different ratio and serial number of H, N, and P residues with protein chain length (a) $N = 120$, (b) $N = 240$.

2.3.3. Global Orientational Order Parameter

The global orientational order parameter P of the protein with the same HNP proportion but different serial residue numbers in the repeating unit as a function of temperature is plotted in Figure 9. As the same tendency shown in Figures 3 and 6, P increases with the decreasing temperature, which indicates the structural transition from a random coil to an ordered state. For the same chain length, the protein with a larger number of consecutive H residues or a higher proportion of H residues in the repeating unit has the higher value of P , indicating the orientationally ordered structure is formed. The longer the chain length is, the more pronounced difference between proteins with different HNP proportions or serial number of same type of residues occurs, especially in the condition of the low temperature region.

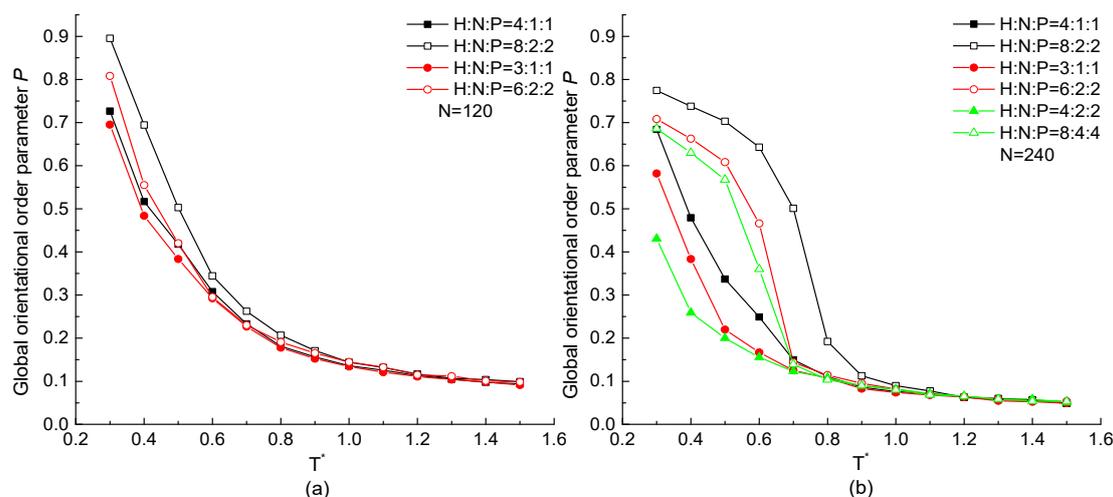


Figure 9. The global orientational order parameter P versus reduced temperature T^* of different ratio and serial number of H, N, and P residues with protein chain length (a) $N = 120$, (b) $N = 240$.

Concluding from the above calculation results, for the protein with the higher proportion of H residues and the larger number of serial H residues in a repeating unit, the interaction of hydrophobic residues is more obvious and continuous leading to the lower non-bonded energy, the smaller radius of gyration, and the larger value of P . The formation of ordered and compact configuration could be easily obtained in such sequential condition.

2.3.4. Stem Length

In this section, we discuss the conformational change of a protein chain by microscopic analysis. The bond of a protein chain takes the *trans* or *gauche* state, which is defined by the dihedral angle $|\varphi| \leq \frac{\pi}{3}$ or $|\varphi| > \frac{\pi}{3}$, respectively. We introduce the concept of a stem defined by more than three consecutive all-*trans* bonds and investigate the stem length distribution at different temperatures. A protein chain consists of several stems with different stem length N_{tr} (i.e., the number of consecutive *trans* bonds) and connected by loops. The distribution of stem length N_{tr} was analyzed as the function of temperature.

The distribution of the stem size $P(N_{tr})$ is normalized and shown in Figure 10. An exponential decrease with the increase of stem size is observed at high temperature $T^* \geq 1.0$ in all protein samples. The reason for this exponential decrease is that each bond takes the *trans* or *gauche* state randomly with no spatial correlation under the high temperature. It also presents that the rate of exponential decline increases with the increasing temperature especially in the case of H3N1P1 protein in Figure 10a. Besides the exponentially decreased tendency observed in Figure 10b–d, highly non-exponential behavior is also shown when the temperature is low enough. The linear-lg plot of stem length distributions $P(N_{tr})$ of H4N1P1, H6N2P2, and H8N2P2 proteins are deviated from exponential decrease at low temperature $T^* \leq 0.5$ and peaks are visible. The unimodal distribution of H4N1P1 protein

samples is shown in Figure 10b. Additionally, the peak located at the stem length about 60 bonds becomes more obvious with the temperature decreasing. It indicates that the protein chain is ordered with a specific length at low temperature. Meanwhile, multimodal distributions are shown in the case of H6N2P2 and H8N2P2 protein samples at the low-temperature region in Figure 10c,d. The difference between them is the successive bond number corresponding to the peaks. It means that the protein is ordered with different stem lengths when the proportion of H residues is higher and the number of serial H residues is larger in a repeating unit. Then, both the proportion and the consecutive number of H residues in a repeating unit would affect the stem length of the ordered protein. According to the simulation results, we may infer that not only the H-occupancy ratio but also the number of consecutive hydrophobic residues affect the stem length and its distribution. With the number of consecutive H increases, there will be a transition from no peak to single peak and even multiple peaks. Most of the stems are fixed to one or several specific lengths, indicating the ordered protein chain with specific lengths under the low temperature.

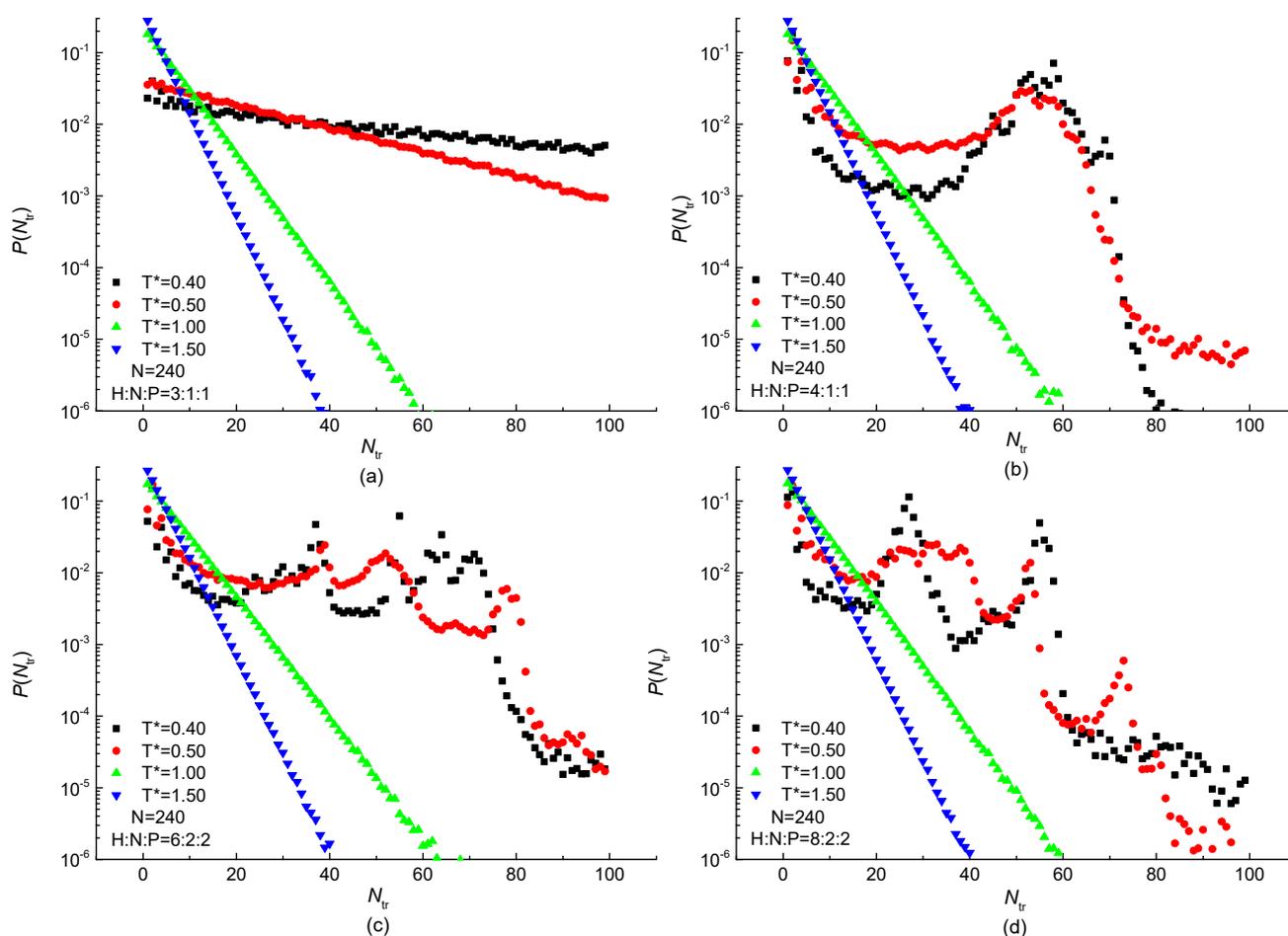


Figure 10. Distribution of stem length under different temperatures in linear-lg plot of (a) H3N1P1, (b) H4N1P1; (c) H6N2P2 and (d) H8N2P2 proteins with chain length of $N = 240$ residues.

2.3.5. Contacts Number

Two residues that are close in space and far in the sequence are often defined as contact to describe the effect of long-range interactions on the structural characteristics of the protein. The number of contacts on the structural level of protein samples with different ratios and sequential hydrophobic residues were statistically analyzed. Nonlocal residue–residue contacts are defined as two C^α atoms within a cutoff distance $R_{cut} = 0.65$ nm and separated by at least a residue separation cutoff value $l_{cut} = 3$. The statistical results are plotted in Figure 11. It presents that the contacts number of the protein H4N2P2 and H3N1P1, which

is constituted by the low ratio and small sequential number of hydrophobic residues, is independent of the temperature. For the other protein chains in Figure 11, contacts number is independent of the temperature when the protein in the high temperature range $T^* \geq 1.0$. With the temperature decreasing, more contacts formed according to the lower potential energy between C^α atoms. It is the similar tendency with the global orientational order parameter as the function of temperature in Figure 9b. More contacts between residues could make the chain orientated to the ordered configuration. In Figure 11, it also shows clearly that proteins made up with a higher constituent ratio of hydrophobic residues are easy to form contacts. Meanwhile, comparing the protein chains with the same ratio but different repeat units, the protein with large sequential number of hydrophobic residues are easy to form contacts. Here, the discussion of contact number could also evaluate the long-range interactions among the protein and present the connection of its structure.

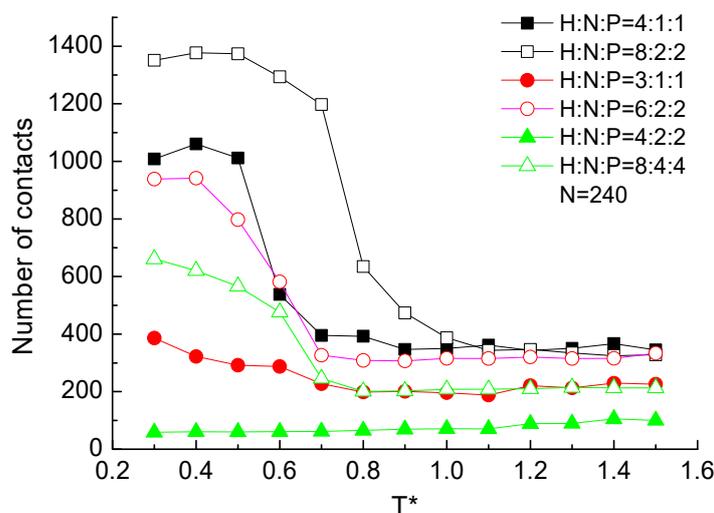


Figure 11. The number of contacts versus reduced temperature T^* of different ratio and serial number of H, N, and P residues with protein chain length $N = 240$.

3. Materials and Methods

To minimize the running time during simulation processes and generalize the general properties of proteins instead of a specific one, the modified off-lattice HNP model is applied in this study [21]. Twenty types of residues are classified into three groups, namely hydrophobic residues (H) consisting of Phe, Met, Ile, Leu, Trp, Val, Cys, and Tyr, neutral residues (N) consisting of His, Ala, Gly, and Thr, and hydrophilic residues (P) consisting of Lys, Asp, Asn, Glu, Gln, Ser, Pro, and Arg [22]. Then, a single protein chain is described by a sequence of three types of beads, H, N, and P. Each bead in a protein molecule is represented by the position of a C^α atom. The beads interact via the non-bonded potential and bonded potentials, which including bond-stretching, bond-bending, and bond-torsional potentials. The potential energy expressions and the parameter values of the coarse-grained model are adopted as in the literature [21,23,24]. It can be thought of a simple version used in the study of the structure and dynamics of real proteins. The behavior of a single protein chain with different distributions of H, N, and P beads has been studied by means of MD simulations here. The four potential functions of a protein are as follows:

- (I) The bond-stretching potential energies for every couple of adjacent beads connected by covalent bonds,

$$E_{stretch} = \sum_{i=2}^N \frac{1}{2} k_d (l_i - l_0)^2 \quad (1)$$

where the equilibrium bond length $l_0 = 0.153$ nm and l_i is the bond length between beads $i - 1$ and i . The bond-stretching potential energy constant $k_d = 7 \times 10^4$ kcal/nm²mol.

(II) The bond-bending potential energies are defined for every triplet of adjacent beads,

$$E_{bend} = \sum_{i=3}^N \frac{1}{2} k_{\theta} (\theta_i - \theta_0)^2 \quad (2)$$

where the equilibrium bond angle $\theta_0 = 1.231$ rad and θ_i is the bond angle between three beads $i - 2$, $i - 1$, and i . The bond-bending potential energy constant $k_{\theta} = 100$ kcal/rad²mol.

(III) The bond-torsional potential energies are defined for every quadruplet of adjacent beads,

$$E_{torsion} = \sum_{i=4}^N \frac{1}{2} k_{\varphi} (1 + \cos(3\varphi_i)) \quad (3)$$

where φ_i is the dihedral angle of two planes consisting of four consecutive beads $i - 3$, $i - 2$, $i - 1$, and i along the backbone. The constant of torsional energy $k_{\varphi} = 2.0$ kcal/mol. The threefold torsional energy of a bond in a polymer generally exhibits several stable states: *trans* (t), *gauche plus* (g⁺), and *gauche minus* (g⁻) separated by barriers. The continuous *trans* states in the polymer chain show the rigidity of chain molecules with parallel ordering configuration.

(IV) The non-bonded interaction involves beads that are at least four residues apart in the amino acid sequence. For two beads, the non-bonded interaction is given by the truncated 12-6 Lennard-Jones potential,

$$E_{LJ} = \sum_{i=1}^N \sum_{j=1}^N 4\epsilon \left[\left(\frac{\sigma}{R_{ij}} \right)^{12} - \Lambda \left(\frac{\sigma}{R_{ij}} \right)^6 \right] (j - i \geq 4) \quad (4)$$

where r_{ij} is the distance between atoms i and j . ϵ is the parameter that defines the energy strength and Λ sets the nonlocal interaction that is attractive or repulsive. Both of these two parameters depend on the nature of the two residues involved. The values of the van der Waals interaction parameters are shown in Table 1. Here, the values of potential parameters are $\epsilon_h = 0.1984$ kcal/mol and $\sigma = 0.3624$ nm.

Table 1. Parameters ϵ and Λ in the non-bonded interaction term of the potential function depending on the nature of the two residues involved.

	H	N	P
H	$\epsilon = \epsilon_h; \Lambda = 1$	$\epsilon = \frac{2}{3}\epsilon_h; \Lambda = 0$	$\epsilon = \frac{7}{12}\epsilon_h; \Lambda = 0$
N	$\epsilon = \frac{2}{3}\epsilon_h; \Lambda = 0$	$\epsilon = \frac{1}{3}\epsilon_h; \Lambda = 0$	$\epsilon = \frac{1}{4}\epsilon_h; \Lambda = 0$
P	$\epsilon = \frac{7}{12}\epsilon_h; \Lambda = 0$	$\epsilon = \frac{1}{4}\epsilon_h; \Lambda = 0$	$\epsilon = \frac{1}{6}\epsilon_h; \Lambda = -1$

The mass of protein chain is fixed to cancel overall chain translation. The coordinate system with the three principal axes of inertia is introduced to analyze the anisotropic structure. The radius of gyration R_g is a basic measurement of the overall size of a protein chain. The change in the structure of a protein during MD simulations can be quantified by the radius of gyration R_g , which is defined as:

$$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^N |r(i) - r_{center}|^2} \quad (5)$$

where N is the number of protein atoms. $r(i)$ and r_{center} are the coordinates of a bead i and the center of mass, respectively. For in-depth understanding of the average shape of a protein chain, a three inertial principal axis coordinate system is introduced, and

the relation $R_g^2 = R_{gx}^2 + R_{gy}^2 + R_{gz}^2$ holds. We also calculate the bond-orientational order parameter P defined by

$$P = \frac{1}{N-2} \sum_{i=3}^N \left(\frac{3 \cos^2 \varphi_i - 1}{2} \right) \quad (6)$$

where φ_i is the angle between the bond vector and the z-axis of the protein chain coordinate system and N is the number of beads. The parameter P would take a value of 1 for a protein whose bonds are perfectly parallel, and that of 0 for randomly oriented. We have monitored the value of P for the whole chain.

The same as the previous work, all simulation results were presented in terms of reduced units [25]. It follows that the reduced temperature T^* is equal to $k_B T / \epsilon$, the reduced energy is E / ϵ , and the reduced time is $t \sqrt{\epsilon / m \sigma^2}$. A single protein chain consisting of three types of beads with different chain lengths is exposed to a vacuum. The MD simulations are carried out from random configuration of the protein at high temperature. Then, the protein is quenched to a certain low temperature, and 500 ns (1×10^7 time steps) simulations are carried out under several temperature conditions, which takes T^* from 0.3 to 1.5. The equations of motion for all atoms are solved numerically using the velocity version of the Velocity–Verlet algorithm [26]. The Nosé–Hoover method is applied to keep the system temperature constant [27,28]. The integration time step is 2.5 fs and a relaxation constant is 25 fs, respectively. The cutoff distance is 0.95 nm in our simulation. The total linear and angular momentums are taken to be zero in order to cancel overall translation and rotation of the chain.

Based on the idea of the HNP model, proteins are described as a linear polymer consisting of three beads with different properties as basic units. For example, H3N1P1 can be described as a protein chain with three hydrophobic residues, one neutral residue, and one hydrophilic residue connected as a repeat unit. According to this model, we investigate the effects of the composition and distribution of amino acid in protein primary sequence on the 3D structure of the protein through calculating the energies and structural parameters. The potential and structural properties of the protein chain with different lengths and HNP components in terms of both energy and structural parameters have been discussed in this paper.

4. Conclusions

In this article, we performed MD simulations for the protein chain with different HNP segment proportions and the same HNP proportion but different repeating units over a wide range of temperatures. We investigated the potential energy and structural parameters such as radius of gyration, global orientational order parameter, stem length distribution, and contact number of the protein. The current results indicate:

- (1) The four potential energy of each bond in the protein chain is independent of chain length. The linear relationship between the bond-stretching/bending energy per bond and the temperature is obtained. It is consistent as the equipartition theorem of energy that the average value of bond-stretching/bending energy associated with each independence degree of freedom bond length l or bond angle θ . The increasing tendency of the bond-torsional or non-bonded energy has a little deviation from the straight line. Overall, the total potential energy of the whole chain is only dependent on the chain length when the protein is at a certain temperature.
1. The total and components of the radius of gyration with different protein chain lengths as the function of temperature show that R_g increases with decreased temperature. R_{gz} shows a similar tendency with R_g , while R_{gx} (or R_{gy}) presents the opposite tendency with R_g , especially when $T^* < 0.6$. The differences between R_g / R_{gz} and R_{gx} / R_{gy} indicate that the protein takes anisotropic configuration at the low temperature. The increasing global orientational order parameter P with the decreasing temperature indicates the paralleled bonds among the protein. The shorter chain takes the larger value of P than the one of longer protein, as the longer protein is partially paralleled

under a certain temperature with more continuous bonds influencing each other to hardly form a good orientated structure.

- (3) The effect of the hydrophobic, neutral, and polar residue proportions under certain chain length was also investigated. The simulation results show that the protein with the higher proportion of hydrophobic residue in a repeating unit takes the lower value of non-bonded energy, the lower value of radius of gyration and the higher value of global orientational order when protein chains with the same length under the certain temperature. It indicates that the protein with higher ratio of hydrophobic residues can easily transform from a random coil to an oriented and compact structure with the decreasing temperature.
- (4) From the perspective of the successive number of hydrophobic residues in a repeating unit, we concluded that the protein with a higher number of consecutive H residue receives the lower value of the non-bonded energy, the lower value of the radius of gyration and the larger value of global orientational order as the sequential hydrophobic residues make more consecutive H-H interaction pairs to form tight and ordered configuration. The analysis of contact number also reflects the long-range interaction among the residues.
- (5) Exponential decline of stem length distribution is shown in any proteins under the high temperature situation. However, with the temperature decreasing, different proteins present a diverse tendency of stem length distribution. For example, the stem length distribution of H3N1P1 protein chains can keep exponentially decreased in the low temperature, while H4N1P1, H6N2P2, and H8N2P2 present peaks located at different stem lengths from 30 to 60 bonds. It indicates the specific ordered lengths of the ordered protein chain. The protein with higher proportion or larger number of consecutive hydrophobic residue H in a repeating unit present the transition of the stem length distribution from exponential decline to unimodal peak and even multiple peaks. Thus, the proportion and the number of consecutive hydrophobic residue influence the stem length distribution simultaneously.

Author Contributions: Methodology, P.X.; Software, Q.S.; Formal analysis, Z.W.; Investigation, Z.W. and Z.Y.; Resources, C.J.; Data curation, Q.S., Z.W., C.J. and Z.Y.; Writing—original draft, Q.S.; Supervision, Z.J.; Project administration, Z.J.; Funding acquisition, P.X. and Z.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 21873087 and 51876196).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We also appreciate the computing resource provided by the cluster at Key Laboratory of Intelligent Manufacturing Quality Big Data Tracing and Analysis of Zhejiang Province, China Jiliang University.

Conflicts of Interest: The authors declared no conflict of interest.

References

1. Hagita, K.; Fujiwara, S. Single-chain folding of a quenched isotactic polypropylene chain through united atom molecular dynamics simulations. *Polymer* **2019**, *183*, 121861. [[CrossRef](#)]
2. Farris, A.C.K.; Seaton, D.T.; Landau, D.P. Effects of lattice constraints in coarse-grained protein models. *J. Chem. Phys.* **2021**, *154*, 084903. [[CrossRef](#)] [[PubMed](#)]
3. Jafari, M.; Doustdar, F.; Mehrnejad, F. Molecular self-assembly strategy for encapsulation of an amphipathic α -helical antimicrobial peptide into the different polymeric and copolymeric nanoparticles. *J. Chem. Inf. Model.* **2019**, *59*, 550–563. [[CrossRef](#)] [[PubMed](#)]
4. Jiang, Z.; Zhang, L.; Chen, J.; Xia, A.; Zhao, D. Effect of amino acid on forming residue–residue contacts in proteins. *Polymer* **2002**, *43*, 6037–6047. [[CrossRef](#)]
5. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230. [[CrossRef](#)]

6. Plaxco, K.W.; Simons, K.T.; Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **1998**, *277*, 985–994. [[CrossRef](#)]
7. Ivankov, D.N.; Garbuzynskiy, S.O.; Alm, E.; Plaxco, K.W.; Baker, D.; Finkelstein, A.V. Contact order revisited: Influence of protein size on the folding rate. *Protein Sci.* **2003**, *12*, 2057–2062. [[CrossRef](#)]
8. Zhong-Can, O.-Y.; Su, Z.-B.; Wang, C.-L. Coil Formation in Multishell Carbon Nanotubes: Competition between Curvature Elasticity and Interlayer Adhesion. *Phys. Rev. Lett.* **1997**, *78*, 4055–4058. [[CrossRef](#)]
9. Gumerov, R.A.; Potemkin, I.I. Computer simulations of comb-like macromolecules with responsive diblock copolymer side chains. *Colloid Polym. Sci.* **2020**, *299*, 407–418. [[CrossRef](#)]
10. Nikam, R.; Xu, X.; Kanduc, M.; Dzubiella, J. Competitive sorption of monovalent and divalent ions by highly charged globular macromolecules. *J. Chem. Phys.* **2020**, *153*, 044904. [[CrossRef](#)] [[PubMed](#)]
11. Xu, B.; Jin, H.; Li, H.; Guo, Y.; Fan, J. Investigation on the Evolution of the Coal Macromolecule in the Process of Combustion With the Molecular Dynamics Method. *J. Energy Resour. Technol.* **2020**, *142*, 052301. [[CrossRef](#)]
12. Ferreon, A.C.; Deniz, A.A. Protein folding at single-molecule resolution. *Biochim. Biophys. Acta* **2011**, *1814*, 1021–1029. [[CrossRef](#)] [[PubMed](#)]
13. Hattori, L.T.; Pinheiro, B.A.; Frigori, R.B.; Benitez, C.M.V.; Lopes, H.S. PathMolD-AB: Spatiotemporal pathways of protein folding using parallel molecular dynamics with a coarse-grained model. *Comput. Biol. Chem.* **2020**, *87*, 107301. [[CrossRef](#)]
14. Dill, K.A. Theory for the folding and stability of globular proteins. *Biochemistry* **1985**, *24*, 1501–1509. [[CrossRef](#)]
15. Regan, L.; DeGrado, W.F. Characterization of a helical protein designed from first principles. *Science* **1988**, *241*, 976–978. [[CrossRef](#)]
16. Miyazawa, S.; Jernigan, R.L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **1996**, *256*, 623–644. [[CrossRef](#)]
17. Sangha, A.K.; Keyes, T. Protein folding and confinement: Inherent structure analysis of chaperonin action. *J. Phys. Chem. B* **2010**, *114*, 16908–16917. [[CrossRef](#)]
18. Li, B.; Lin, M.; Liu, Q.; Li, Y.; Zhou, C. Protein folding optimization based on 3D off-lattice model via an improved artificial bee colony algorithm. *J. Mol. Model.* **2015**, *21*, 261. [[CrossRef](#)]
19. Bošković, B.; Brest, J. Two-phase protein folding optimization on a three-dimensional AB off-lattice model. *Swarm Evol. Comput.* **2020**, *57*, 100708. [[CrossRef](#)]
20. Bardis, N.; Todorin, I.; Yanev, N.; Traykov, M.; Yurukov, B. A new off-lattice HP model with side-chains for protein folding problem. *ITM Web Conf.* **2018**, *16*, 02007. [[CrossRef](#)]
21. Jiang, Z.-T.; Dou, W.-H.; Shen, Y.; Sun, T.-T.; Xu, P. Residual occurrence and energy property of proteins in HNP model. *Chin. Phys. B* **2015**, *24*, 116802. [[CrossRef](#)]
22. Miyazawa, S.; Jernigan, R.L. Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules* **2002**, *18*, 534–552. [[CrossRef](#)]
23. Mayo, S.L.; Olafson, B.D.; Goddard, W.A. DREIDING: A generic force field for molecular simulations. *J. Phys. Chem.* **2002**, *94*, 8897–8909. [[CrossRef](#)]
24. Li, H.; Fujiwara, S.; Nakamura, H.; Mizuguchi, T.; Yasunaga, T.; Otsuka, T.; Kenmotsu, T.; Hatano, Y.; Saito, S. Structural Changes in Tritium-Substituted Polymeric Materials by Beta Decays: A Molecular Dynamics Study. *Plasma Fusion Res.* **2019**, *14*, 3401106. [[CrossRef](#)]
25. Jiang, Z.-T.; Xu, P.; Sun, T.-T. A study on the structural transition of a single polymer chain by parallel tempering molecular dynamics simulation. *Chin. J. Polym. Sci.* **2011**, *30*, 45–55. [[CrossRef](#)]
26. Swope, W.C.; Andersen, H.C.; Berens, P.H.; Wilson, K.R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982**, *76*, 637–649. [[CrossRef](#)]
27. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519. [[CrossRef](#)]
28. Hoover, W.G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A Gen. Phys.* **1985**, *31*, 1695–1697. [[CrossRef](#)]