

## RATE OF INTRA-MOLECULAR CONTACT FORMATION IN SHORT TWO-DIMENSIONAL COMPACT POLYMERS CHAINS\*

Jin Chen<sup>a</sup>, Lin-xi Zhang<sup>b\*\*</sup> and A-gen Xia<sup>a</sup>

<sup>a</sup>Department of Physics, Zhejiang University, Hangzhou 310027, China

<sup>b</sup>Department of Physics, Wenzhou Normal College, Wenzhou 325027, China

**Abstract** It is important to know the rate of intra-molecular contact formation in proteins in order to understand how proteins fold clearly. Here we investigate the rate of intra-molecular contact formation in short two-dimensional compact polymer chains by calculating the probability distribution  $p(r)$  of end-to-end distance  $r$  using the enumeration calculation method and HP model on two-dimensional square lattice. The probability distribution of end-to-end distance  $p(r)$  of short two-dimensional compact polymers chains may consist of two parts, i.e.  $p(r) = p_1(r) + p_2(r)$ , where  $p_1(r)$  and  $p_2(r)$  are different for small  $r$ . The rate of contact formation decreases monotonically with the number of bonds  $N$ , and the rate approximately conforms to the scaling relation of  $k(N) \propto N^{-\alpha}$ . Here the value of  $\alpha$  increases with the contact radius  $a$  and it also depends on the percentage of H (hydrophobic) residues in the sequences of compact chains and the energy parameters of  $\epsilon_{HH}$ ,  $\epsilon_{HP}$  and  $\epsilon_{PP}$ . Some comparisons of theoretical predictions with experimental results are also made. This investigation may help us to understand the protein folding.

**Keywords:** Two-dimensional compact polymer; HP model; Rate of contact formation.

### INTRODUCTION

Protein is composed of 20 amino acid residues and protein folding starts from an ensemble of random coil conformations to generate the native state with well-defined side-chain contacts. Protein only reaches to a certain native structure to play an important role in biologic activities. For many proteins, rapid chain collapse precedes the formation of native structure<sup>[1]</sup>. To understand how proteins fold, it is essential to know the time scales and mechanisms of the elementary processes in folding. The formation of a contact between two residues in an unfold polypeptide chain is one of the most fundamental processes and is essential for many aspects of protein folding. Many scientists have focused their studies on the diffusion-controlled formation of intra-molecular contact in proteins. Experimental<sup>[2–6]</sup> and theoretical<sup>[7–10]</sup> studies indicated that diffusion controlled search for the native contacts may be the rate limiting step in the folding of small, simple domain proteins<sup>[11–14]</sup>.

Recently the rate of formation of a contact between two ends of a protein chain as a function of the number of peptide bonds  $N$  has been measured by Lapidus<sup>[5]</sup> and Bieri *et al.*<sup>[6]</sup>. Their results show that the rate of formation decreases monotonically with  $N$  which disagrees with the theoretical results<sup>[7–9]</sup>. Moreover, Wang and Makarov<sup>[10]</sup> chose an atomistic model to display more detailed and atomistic simulations of denatured peptide chains. Their calculated rate reduces monotonically with  $N$ , surely in agreement with the experiment results<sup>[5, 6]</sup> and the rate follows the relation of  $k(N) \propto N^{-\alpha}$ <sup>[10]</sup>. However, when  $N \leq 5$ , their simulation prediction is different from the experimental data<sup>[5]</sup>. The reason may be that the experiments were limited by the rate of energy transfer

\* This research was financially supported by the National Natural Science Foundation of China (Nos. 20174036, 20274074, 20574052) and the Natural Science Foundation of Zhejiang Province (No. R404047).

\*\* Corresponding author: Lin-xi Zhang (章林溪), E-mail: lxzhang@hzcnc.com

Received December 14, 2004; Revised January 28, 2005; Accepted February 2, 2005

between the probes rather than by their diffusive approach, or their simulation models neglected the different properties of the  $N - C_\alpha$  and  $C_\alpha - C'$  bonds and the interactions between these residues. In fact, these interactions play a very important role in protein folding and intra-molecular contact formation.

In this paper, we will investigate the rate of intra-molecular contact formation in short two-dimensional compact polymers chains by calculating the probability distribution  $p(r)$  of end-to-end distance  $r$  using the enumeration calculation method and HP model on two-dimensional square lattice. In HP model<sup>[14-20]</sup>, a protein can be modeled as a specific sequence of H and P residues, here H represents the hydrophobic residue and P represents the polar or hydrophilic residue. Exhaustive exploration of full conformation space is computationally possible because molecules are modeled as short compact polymer chains on a two-dimensional square lattice. Here we investigate the rate of intra-molecular contact formation by enumeration calculation method with the chain length  $N = 2-16$  for the sequences of  $(H)_x$ ,  $(P)_x$ ,  $(HP)_x$ , and  $(HHP)_x$ , respectively. Here  $x$  is the length of repeat unit, and is related with the chain length  $N$ . In order to explore how this method may reveal the protein folding and whether there are any limitations for us to improve, we also compare our results with the experimental findings<sup>[2-6]</sup> and the previous theoretical predictions<sup>[7-10]</sup>.

## METHOD OF CALCULATION

In the two-dimensional square lattice model of compact polymer chains, chains are regarded to comprise a sequence of monomers or residues and each occupies one lattice site. Residues are numbered sequentially from 1 to  $N + 1$ . The bond lengths are normalized to unity, and bond angles are limited to  $\pm 90^\circ$  and  $180^\circ$ , respectively<sup>[14, 15]</sup>. Excluded volume is taken into account by forbidding two different residues from occupying the same lattice site. For a chain with  $N + 1$  residues, the chain length is  $N$  (the number of bonds).

For a self-avoiding chain of length  $N$  in a square lattice model, the Hamiltonian of the system is

$$H = \sum_{i < j} \varepsilon_{\mu_i \mu_j} \Delta(r_i - r_j) \quad (1)$$

Here  $\varepsilon_{\mu_i \mu_j}$  is the contact energy between residues type  $\mu_i$  and  $\mu_j$ , and  $\Delta(r_i - r_j) = 1$  if  $r_i$  and  $r_j$  are adjoining lattice sites with  $i$  and  $j$  not adjacent along the chain, and  $\Delta(r_i - r_j) = 0$  otherwise. The term  $\varepsilon_{\mu_i \mu_j}$  in Eq. (1) means that the intra-molecular contact interaction plays an important role in protein folding. A favorable contact energy,  $\varepsilon_{HH} = -1.0$  (in dimensionless  $k_B T$  units) is assigned to two nonconsecutive residues which are lattice spacing apart, while the other interactions,  $\varepsilon_{HP} = \varepsilon_{PP} = 0$ <sup>[14-20]</sup>. In addition, the other energy parameters are also assigned in our paper:  $\varepsilon_{HH} = -2.3$ ,  $\varepsilon_{HP} = -1.0$ , and  $\varepsilon_{PP} = 0.0$ <sup>[21]</sup>. Here several sequences of  $(H)_x$ ,  $(P)_x$ ,  $(HP)_x$ , and  $(HHP)_x$  are considered in order to investigate the effects of different sequences on the rate of intra-molecular contact formation.

If the contact is formed as soon as  $r$  reaches a contact radius  $a$  then the rate constant of intra-molecular contact formation can be calculated from the Szabo, Schulten and Schulten Formula<sup>[10, 22]</sup>.

$$K^{-1} = D^{-1} \int_a^\infty \frac{dx}{p(x)x^2} \left[ \int_a^\infty p(y)y^2 dy \right]^2 \bigg/ \int_a^\infty p(x)x^2 dx \quad (2)$$

Here  $p(x)$  is the probability distribution function of end-to-end distance  $x$ ,  $D$  is the diffusion coefficient and  $a$  indicates the radius of contacts. In our calculation,  $D$  is supposed to be a constant<sup>[10]</sup>.

In previous papers<sup>[10, 23-27]</sup>, scientists assume that the only degrees of freedom in a peptide are the angles  $\{\phi, \psi\}$  of rotation around the bonds  $N - C_\alpha$  and  $C_\alpha - C'$  while other geometric parameters (bond lengths and angles) remain fixed. The side groups are not allowed to overlap. In their simulations, the probability distribution  $p(r)$  of the end-to-end distance  $r$  in the absence of the force is related to the free energy.

In their model, the side groups are simply regarded as the same atoms so that they cannot discuss the diversity of different amino acid residues in protein chains. In the meantime, the interactions between the residues

also cannot be considered. Here we will consider the interactions between the P and H residues. In the two-dimensional square lattice model, every atom occupies one lattice site. So the probability distribution  $p(r)$  can be rewritten as:

$$p(r) = \frac{\sum_i e^{-E_i(r)/k_B T}}{Z} \quad (3)$$

here  $\sum_i$  is the sum of all conformations with end-to-end distance  $r$  and  $E_i(r)$  is the total energy of  $i$ -th conformation and determined by Eq. (1) with considering the contact energy of  $\varepsilon_{HH}$ ,  $\varepsilon_{HP}$  and  $\varepsilon_{PP}$ . Moreover,  $Z$  is the partition function of the system. In our calculation we change the integration to summation in Eq. (2) to calculate the rate of contact formation separately in two-dimensional square lattice model.

Here we adopt the enumeration calculation method to calculate the probability distribution function  $p(r)$  of end-to-end distance  $r$  and the rate of contact formation in two-dimensional compact polymer chains on square lattice. The results from enumerative statistics method may be more reasonable. However, the present problem is that this calculation is not yet practical for long chains because the number of conformations increases so fast that it will cost much long computer time. Therefore, we only discuss the rate of contact formation in the compact chains with the chain length  $N = 2-16$ .

## RESULTS AND DISCUSSIONS

The probability distribution of end-to-end distance  $p(r)$  is shown in Fig. 1. Because we apply the two-dimensional square lattice model to calculate the statistical properties of compact chains, the square end-to-end distance  $r^2$  is a discontinuous integer. In Fig. 1, the solid square indicates the probability distribution  $p(r)$  of the all-hydrophobic compact chain  $(H)_x$  (15 hydrophobic residues) while the hollow square shows the results of the all-hydrophilic compact chain  $(P)_x$  for chain length  $N = 14$  (15 hydrophilic residues), where the subscript  $x$  represents the number of the repeat units of residues, and there is a relationship of  $x = N + 1$  in this case. If we link these solid and hollow squares by dash, respectively, except for few separate far anomalous squares, we can approximately draw out two exquisite curves for the same compact chains. Therefore we may express  $p(r)$  as

$$p(r) = p_1(r) + p_2(r) \quad (4)$$

here  $p_1(r)$  and  $p_2(r)$  represent two curves, respectively. Therefore, the probability distribution of short two-dimensional compact chains is not Gaussian distribution. The reason may be that end-to-end distance is discontinuous for short two-dimensional lattice chains, and there may be two or three extrema for  $p(r)$ . For a  $(H)_x$  compact chain, the two curves are apart greatly at small end-to-end distance  $r$  and the separation reduces little by little with  $r$  increasing. When  $r > 8$ , two curves almost overlapped. In general, the probability distributions of

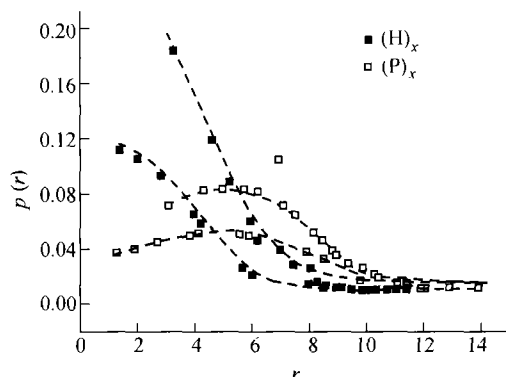


**Fig. 1** Probability distribution of end-to-end distance  $P(r)$  for  $(H)_x$  and  $(P)_x$  compact polymer chains with  $x = 15$

$p_1(r)$  and  $p_2(r)$  of the  $(H)_x$  compact chains all decreases monotonically with the end-to-end distance  $r$ . However, the  $(P)_x$  compact chain is far different from the  $(H)_x$  compact chain for the probability distribution. Although there are also two independent dashed curves, the values are not monotonic. When end-to-end distance  $r$  approaches to 5–7,  $p_1(r)$  and  $p_2(r)$  arrive at the maximum simultaneously. Similarly, two curves get together when  $r > 10$ . Generally the value of the probability distribution  $p(r)$  for the  $(H)_x$  compact chain is far larger than that for the  $(P)_x$  compact chain at small  $r$ . This indicates that the  $(H)_x$  compact chains are more compact than the  $(P)_x$  compact chains, and there are more conformations at small  $r$ . The reason may be that hydrophobic residues can form more contacts than hydrophilic residues. With the increasing of end-to-end distance  $r$ , the number of incompact structures becomes large for the  $(P)_x$  compact chains, therefore, the value of  $p(r)$  for the  $(P)_x$  compact chain is larger than that for the  $(H)_x$  compact chain. It is meaningful that the probability distribution  $p(r)$  all divides into two curves for every chain, and this property may be correlated with two-dimensional lattice model<sup>[28]</sup>.

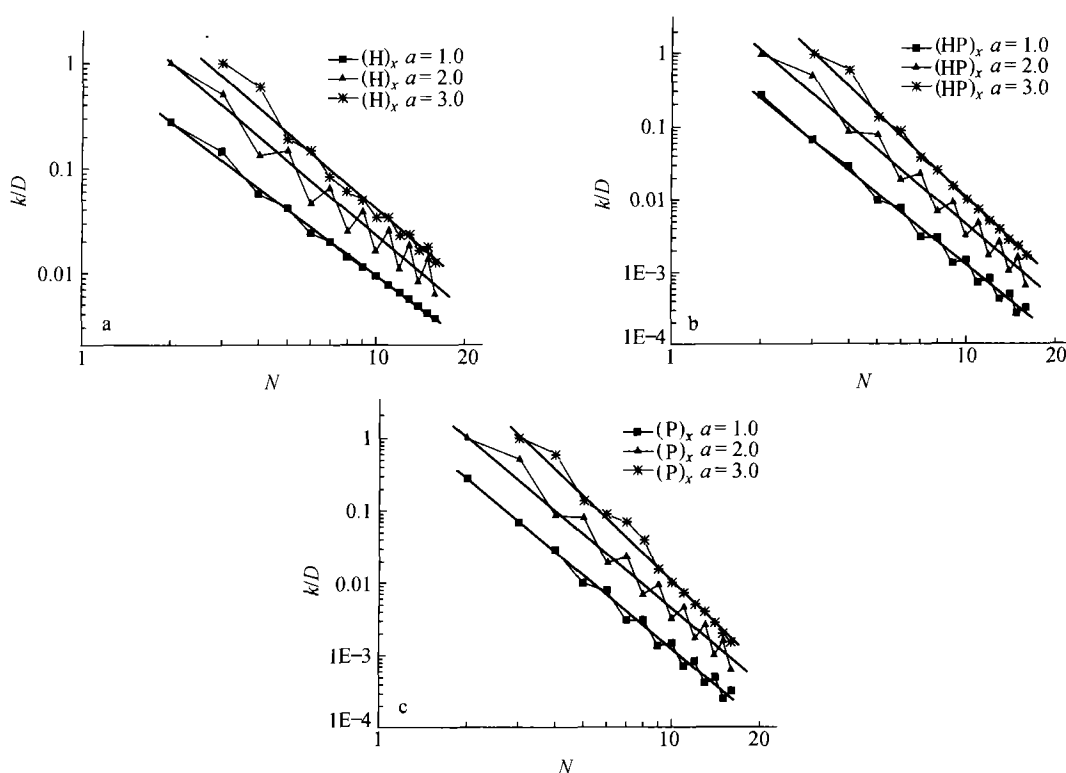


**Fig. 2** Ratio of  $k/D$  as a function of chain length  $N$  for three different values of contact radius  $a = 1.0$ ,  $2.0$ , and  $3.0$ , respectively  
Here  $k$  is the rate of contact formation,  $D$  is the diffusion coefficient and a)  $(H)_x$ ; b)  $(HP)_x$ ; c)  $(P)_x$

To investigate the effects of contact radius  $a$  and different sequence on the rate of contact formation, we first calculate the different rates of contact formation with  $a = 1.0$ ,  $2.0$ , and  $3.0$ , and three sequences of  $(H)_x$ ,  $(P)_x$ , and  $(HP)_x$ . Here the energy parameters of  $\varepsilon_{HH} = -1.0$  and  $\varepsilon_{HP} = \varepsilon_{PP} = 0$  are adopted. In Fig. 2(a), this is the case of  $(H)_x$  compact chains, and three contact radii of  $a = 1.0$ ,  $2.0$  and  $3.0$  are considered, respectively. The results show an exponential decrease of the rate of contact formation and it can be expressed as

$$k(N) \sim N^{-\alpha} \quad (4)$$

with  $\alpha_1 = 1.99$ ,  $\alpha_2 = 2.04$  and  $\alpha_3 = 2.07$  for  $a = 1.0$ ,  $2.0$  and  $3.0$ , respectively. Here the diffusion coefficient  $D$  is supposed to be a constant. In Fig. 2(a), these lines seem to be parallel and this finding is in agreement with the

characteristic property of compact chains on the two-dimensional square lattice model<sup>[28]</sup>. Similarly, we also plot straight lines with  $\alpha_1 = 3.24$ ,  $\alpha_2 = 3.45$  and  $\alpha_3 = 3.72$  for the  $(HP)_x$  compact chains in Fig. 2(b) and  $\alpha_1 = 3.34$ ,  $\alpha_2 = 3.58$  and  $\alpha_3 = 4.76$  for the  $(P)_x$  compact chains in Fig. 2(c). We find out that the value of  $\alpha$  increases with the contact radius  $a$ , especially for the  $(HP)_x$  and  $(P)_x$  compact chains. In previous papers<sup>[5, 6-10]</sup>, the measured  $\alpha$  values were about 1.36–1.50 for three-dimensional protein molecules. In the meantime, the theoretical prediction of  $\alpha$  is about 1.65 for  $N > 5$ <sup>[10]</sup>. In Fig. 2, our results are all greater than those experimental and theoretical predictions. The reason may be that the scaling value of three-dimensional compact chains is always smaller than that of two-dimensional compact chains. Of course, the value of  $\alpha$  also depends on the parameters of interaction energy  $\epsilon_{HH}$ ,  $\epsilon_{HP}$ , and  $\epsilon_{PP}$ . Otherwise, we also find that the slopes of these curves decrease with increasing the percentage of H (hydrophobic) residues in sequence. For example,  $\alpha_1 = 1.99$  for  $(H)_x$  (100%),  $\alpha_1 = 3.24$  for  $(HP)_x$  (50%) and  $\alpha_1 = 3.34$  for  $(P)_x$  (0%) with  $a = 1.0$ . The reason may be that H residues can form the valid contacts and the P residues don't contribute to the contacts in the model of  $\epsilon_{HH} = -1.0$ , and  $\epsilon_{HP} = \epsilon_{PP} = 0.0$ . This leads to a small rate of contact formation, and a slow folding rate for compact chains with a small percentage of H (hydrophobic) residue, [see Figs. 2(b) and 2(c)]. From the above discussion, we know that the rate of contact formation approximately conforms to the scaling relation of  $k(N) \propto N^{-\alpha}$  with the scaling exponent  $\alpha$  depending on the contact radius  $a$  and residue sequence, which is close to experimental results<sup>[2-6]</sup> and the theoretical findings<sup>[7-10]</sup>. Wang and Makarov's results are only fit for  $N > 5$ , however our conclusion can be applied to all  $N$  completely. Therefore, it is important to consider the interactions of residues in calculating the rate of contact formation.

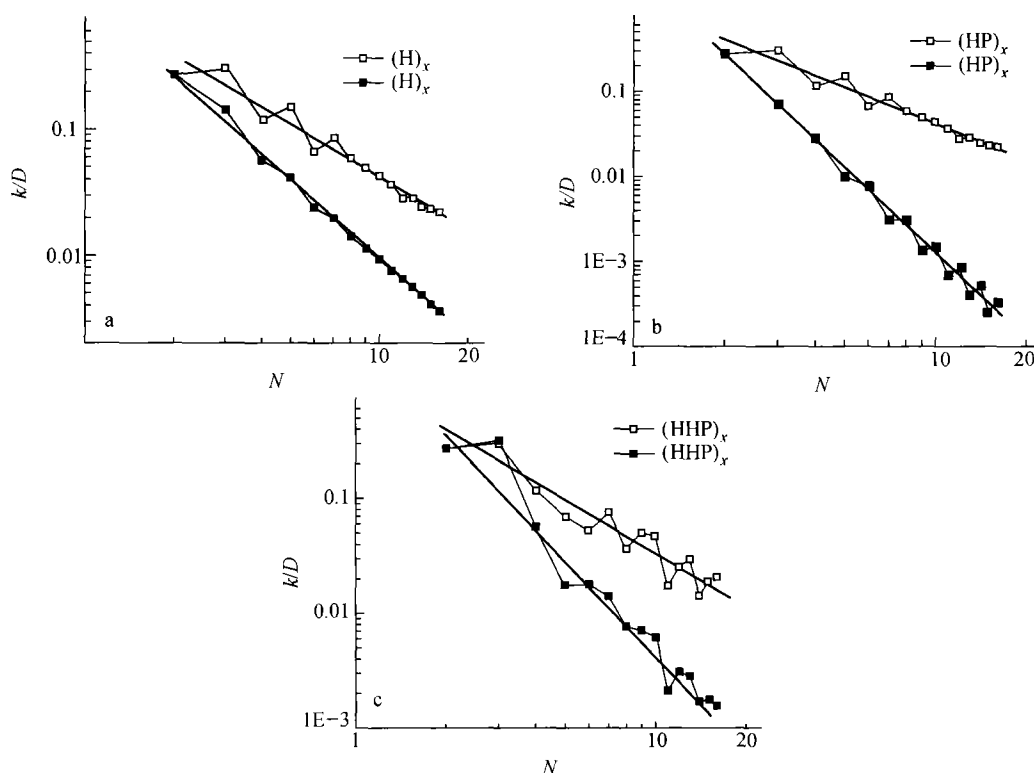


**Fig. 3** Ratio of  $k/D$  as a function of chain length  $N$  for two different values of interaction parameters with  $a = 1.0$

Here  $k$  is the rate of contact formation,  $D$  is the diffusion coefficient and a)  $(H)_x$ ; b)  $(HP)_x$ ; c)  $(HHP)_x$ .

■:  $\epsilon_{HH} = -1.0$ ,  $\epsilon_{HP} = \epsilon_{PP} = 0$

□:  $\epsilon_{HH} = -2.3$ ,  $\epsilon_{HP} = -1.0$ ,  $\epsilon_{PP} = 0$

In order to discuss the importance of interactions of residues in the rate of contact formation further, we also calculate the rate of contact formation using two different values of the energy parameters, and the results are shown in Fig. 3, here the contact radius  $a = 1.0$ . Because  $\epsilon_{pp}$  is equal to zero in two models of energy parameters, we just calculate the rate of contact formation of the  $(H)_x$ ,  $(HP)_x$  and  $(HHP)_x$  compact chains, and the  $(P)_x$  compact chain is not considered. Here we should especially point out the relationship between the number of repeat unit  $x$  and the number of bonds  $N$  for the  $(HHP)_x$  compact chains. For example, it is  $x = 2$  and  $N = 3$  for the sequence of HHPH,  $x = 2$  and  $N = 4$  for the sequence of HHPHH, and  $x = 2$  and  $N = 5$  for the sequence of HHPHHP, respectively. In Fig 3, the rate of contact formation also decreases monotonically with  $N$ , and is in agreement with the experimental results<sup>[5,6]</sup>. However, when  $\epsilon_{HH} = -2.3$ ,  $\epsilon_{HP} = -1.0$ , and  $\epsilon_{PP} = 0.0$ , the oscillation lines are more complicate than the others, and the absolute values of the slopes are  $\alpha_H = 1.37$ ,  $\alpha_{HHP} = 1.43$ , and  $\alpha_{HP} = 1.46$ , respectively, which are smaller than those of  $\alpha_H = 1.99$ ,  $\alpha_{HHP} = 2.73$ , and  $\alpha_{HP} = 3.24$  in the model of  $\epsilon_{HH} = -1.0$  and  $\epsilon_{HP} = \epsilon_{PP} = 0.0$ . The reason may be that the interactions of residues with  $\epsilon_{HH} = -2.3$  and  $\epsilon_{HP} = -1.0$  are more attractive and this leads to a faster contact formation than that with  $\epsilon_{HH} = -1.0$  and  $\epsilon_{HP} = 0.0$ . In the meantime, the value of the slope  $\alpha$  also decreases with the increase in the percentage of H (hydrophobic) residue in sequences of  $(H)_x$ ,  $(HHP)_x$ , and  $(HP)_x$ . When there is a large percentage of H residue in the sequence, there is a more attractive interaction between residues and this leads to a large value of the rate of contact formation. No matter what the different energy parameters are the rate of the contact formation in two-dimension square lattice all approximately follows the relation of  $k(N) \propto N^{-\alpha}$ , here  $\alpha$  depends on the contact radius  $a$ , the residue sequence and the energy parameter between residues simultaneously.

From the above discussion, conclusion can be drawn that the rate constant of contact formation  $k(N)$  is sensitive to the protein folding. When the interactions between residues become stronger or the percentage of H (hydrophobic) residue in the sequence increases, the values of  $\alpha$  decrease and the rate of contact formation increases, and this leads to the acceleration of protein folding greatly. Of course, many conditions may affect the protein folding, and the problem of protein folding is more complex. However, this investigation can help us to understand conformational properties of proteins and protein folding. In the future, we will study the rate of intramolecular contact formation of real proteins using Monte Carlo simulation method by considering the interactions.

## REFERENCES

- 1 Segel, D., Bachmann, A., Hofrichter, J., Hodgson, J., Doniach, S. and Kiefhaber, T., J. Mol. Biol., 1999, 288: 489
- 2 Hagen, S.J., Hofrichter, J., Szabo, A. and Eaton, W.A., Proc. Natl. Acad. Sci. U.S.A., 1996, 93: 11615
- 3 Hagen, S.J., Hofrichter, J. and Eaton, W.A., J. Phys. Chem. B, 1997, 101: 2352
- 4 Eaton, W.A., Munoz, V., Hagen, S.J., Jas, G.S., Lapidus, L.J., Henny, E.R. and Hofrichter, J., Annu. Rev. Biophys. Biomol. Struct., 2000, 29: 327
- 5 Lapidus, L.J., Eaton, W.A. and Hofrichter, J., Proc. Natl. Acad. Sci. U.S.A., 2000, 97: 7220
- 6 Bieri, O., Wirz, J., Hollrang, B., Schulkowski, M., Dnewello, M. and Keifhaber, T., Proc. Natl. Acad. Sci. U.S.A., 1999, 96: 9597
- 7 Camacho, C.J. and Thirumalai, D., Proc. Natl. Acad. Sci. U.S.A., 1995, 92: 1277
- 8 Guo, Z. and Thirumalai, D., Biopolymers, 1995, 36: 83
- 9 Thirumalai, D., J. Phys. Chem. B, 1999, 103: 608
- 10 Wang, Z.S. and Makarov, D.E., J. Phys. Chem., 2002, 117: 4591
- 11 Makarov, D.E. and Metiu, H., J. Chem. Phys., 2002, 116: 5205
- 12 Makarov, D.E., Keller, C., Plaxco, K.W. and Metiu, H., Proc. Natl. Acad. Sci. U.S.A., 2002, 99: 3535
- 13 Debe, D.A. and Goddard, W.A., J. Mol. Biol., 1999, 294: 619
- 14 Lau, K.F. and Dill, K.A., Macromolecules, 1989, 22: 3986
- 15 Zhang, L.X., Jin, J.S., Wang, X.H., Xia, A.G. and Zhao, D.L., Acta Polymerica Sinica (in Chinese), 2002, (6): 479

- 16 Chen, J., Li, J., Zhang, L.X., Jiang, Z.T., Xia, A.G. and Ye, G.X., *Acta Polymerica Sinica* (in Chinese), 2003, (2): 171
- 17 Chen, J., Dong, Z.J., Zhang, L.X., Jiang, Z.T. and Zhao, D.L., *Acta Polymerica Sinica* (in Chinese), 2003, (4): 578
- 18 Kamlekar, S., Schiffer, J.M., Xiong, H., Babik, J.M. and Hecht, M.H., *Science*, 1993, 262: 1680
- 19 Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, K.M., Thomas, P.D. and Chan, H.S., *Protein Sci.*, 1995, 4: 561
- 20 Yue, K., Fiebig, K.M., Thomas, P.D., Chan, H.S., Shacknovich, E.I. and Dill, K.A., *Proc. Natl. Acad. Sci. U.S.A.*, 1995, 92: 325
- 21 Li, H., Helling, R., Tang, H. and Wingreen, N., *Science*, 1996, 273: 666
- 22 Szabo, A., Schulten, K. and Schulten, Z., *J. Chem. Phys.*, 1980, 72: 4350
- 23 Nemethy, G. and Scheraga, H.A., *Biopolymers*, 1965, 3: 155
- 24 Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V., *J. Mol. Biol.*, 1963, 7: 95
- 25 Ramachandran, G.N., Venkatchalan, C.M. and Krimm, S., *Biophys. J.*, 1966, 6: 849
- 26 Ramachandran, G.N. and Sasisekharan, V., *Adv. Protein. Chem.*, 1968, 23: 283
- 27 Pappu, R.V., Srinivasan, R. and Rose, G.D., *Proc. Natl. Acad. Sci. U.S.A.*, 2000, 97: 12565
- 28 Chan, H. and Dill, D.A., *Macromolecules*, 1989, 22: 4559