Low-temperature properties of directed walks with random self interactions

B Derrida and P G Higgs†

CEA, Service de Physique Théorique‡, CE-Saclay, F-91191 Gif-sur-Yvette Cedex, France

Abstract. We consider a simple model for a chain with random self-interactions. Each monomer on the chain carries a charge, which is a random variable, and the configurations of the chain are directed walks. We show that at zero temperature the end-to-end distance R and the energy E of chains with total charge $Q \sim N^x$ scale as $R \sim N^{\nu(x)}$ and $E \sim N^{\theta(x)}$. The exponents thus depend on the total charge through the parameter x. These exponents are measured numerically and compared to the predictions of some simple arguments. At a finite temperature, R is extensive and appears to be independent of x.

1. Introduction

Many models have been proposed to understand the problem of folding of heteropolymers like proteins or DNA molecules [1, 5]. For many of these models, ideas coming from the mean-field theory of spin glasses have proved to be very useful. Recently, it was shown that even very simplified one-dimensional models of heteropolymers [6, 7] could reveal interesting properties. Kantor and Kardar [6] considered a simple model where the heteropolymer consists of a sequence of N random charges $q_i = \pm 1$, the allowed configurations of which are the 1D random walks $\{r_i\}$ with $r_{i+1} - r_i = +1$ or -1. The energy $\tilde{E}(\{r_i\})$ of a configuration $\{r_i\}$ of the polymer is defined by

$$\tilde{E}(\{r_i\}) = \sum_{1 \leqslant i < j \leqslant N} q_i q_j \delta_{r_i, r_j}. \tag{1}$$

Despite its simplicity, this model is not easy to attack either by analytical or by numerical approaches. At zero temperature, using an exact enumeration procedure to find the ground-state configurations, Kantor and Kardar predicted power laws for the end-to-end distance $R \sim N^\nu$ and the fluctuation of the ground-state energy $\Delta\tilde{E} \sim N^\theta$ (with $\nu \simeq 0.574$ and $\theta \simeq 0.58$). Because there is so far no alternative to the exact enumeration procedure, it is difficult to check the validity of these results for sizes $N \geqslant 20$.

More recently, an even simpler model was considered in a joint work with Griffiths [7]. The model is very similar to the one studied by Kantor and Kardar, the only difference being that the configurations $\{r_i\}$ are directed walks ($r_{i+1} - r_i = 0$ or 1). It is possible to obtain much more information when the walk is directed because transfer-matrix techniques can be used, allowing for the numerical study of much longer chains (10^4–10^5).

---

† Present address: Department of Physics, University of Sheffield, Sheffield S3 7RH, UK.
‡ Laboratoire de la Direction des Sciences de la Matière du Commissariat à l'Energie Atomique.

When the charges q_i are $+1$ or -1, the results [7] revealed the existence of a weak freezing transition temperature below which some degrees of freedom are frozen out. At zero temperature, however, the problem looks simple since the ground-state energy of a given chain is always related to the absolute value of the total charge of the chain. This can be seen by rewriting the energy $\tilde{E}(\{r_i\})$ defined by (1) as

$$2\tilde{E}(\{r_i\}) + \sum_i q_i^2 = E(\{r_i\}) = \sum_r [Q(r)]^2 \tag{2}$$

where

$$Q(r) = \sum_i q_i \delta_{r_i, r} \tag{3}$$

is the total charge on site r. For the sake of convenience, we will hereafeter refer to $E(\{r_i\})$, defined by (2), as the *energy of the system*. Obviously, when $q_i = \pm 1$, the ground-state energy is always equal to the absolute value of the total charge $|\sum^i q_i|$, since one can easily build ground-state configurations where each occupied site r has a total charge $Q(r)$ which is 0 or ± 1, so that $\sum_r Q^2(r) = |\sum_i q_i|$. However, the ground state being usually very degenerate, it is not easy to obtain the end-to-end distance R at zero temperature since one would need to average over all the ground-state configurations. (Only when the total charge $Q = 0$, one can show that $R \sim N^{1/2}$ because the calculation of R can be formulated as the number of returns to the origin of a random walker [7].)

In this paper, we discuss the low-temperature properties of this one-dimensional directed model [7] when the charges q_i have a continuous distribution that we will take to be a Gaussian. For continuous charges q_i, even the N dependence of the ground-state energy of directed walks is a non-trivial problem. We will see that an important parameter is the way the total charge Q of the chain scales with N. So in what follows, we study chains of N random Gaussian charges q_i with the constraint that the sum Q is given by

$$Q = \sum_{i=1}^N q_i = N^x \tag{4}$$

where x is a fixed parameter when we vary N. It is easy to generate numerically N Gaussian numbers q_i satisfying the constraint (4). This can be done by choosing N independent Gaussian numbers z_i of variance 1 and by generating the N charges q_i using the formula

$$q_i = z_i + \frac{Q}{N} - \frac{1}{N}\sum_{j=1}^N z_j \qquad \text{where} \qquad \rho(z_j) = \frac{1}{\sqrt{2\pi}} \exp\left(-z_j^2/2\right). \tag{5}$$

Once the $\{q_i\}$ of a given chain are chosen, one can use transfer matrices [7] to calculate the partition function Z_N. This partition function Z_N (which is, as usual, the sum of the Boltzmann weights of all the directed walks) of a chain of N charges satisfies the recursion

$$Z_N = \sum_{i=0}^{N-1} Z_i \exp\left[-\left(\sum_{j=i+1}^N q_j\right)^2 \Big/ T\right] \tag{6}$$

where T is temperature and $Z_0 = 1$. This recursion expresses the fact that any configuration $\{r_i\}$ of a chain of N charges can be thought of as a site occupied by the the last $N - i$

charges of the chain and a chain consisting of the first i charges of the chain. Clearly, the calculation (6) of the partition function Z_N of a chain of N charges only requires the storage of the $N-1$ numbers Z_i for $1 \leqslant i \leqslant N-1$, thus rather long chains can be studied. The calculation of the average size R_N (more precisely R_N is the *thermal* average number of sites occupied by the chain) can also be done recursively

$$R_N = 1 + \frac{1}{Z_N} \sum_{i=0}^{N-1} R_i \ Z_i \ \exp\left[-\left(\sum_{j=i+1}^{N} q_j\right)^2 / T\right] \tag{7}$$

with $R_0 = 0$.

At zero temperature, (6) induces a very simple recursion for the ground-state energy E_N:

$$E_N = \min_{0 \leqslant i \leqslant N-1}\left[E_i + \left(\sum_{j=i+1}^{N} q_j\right)^2\right] \tag{8}$$

with $E_0 = 0$. Note that in contrast to the case of $q_i = \pm 1$, the ground state is non-degenerate for continuous variables q_i.

It is clear from recursions (6)–(8) that the time required to calculate the properties of a chain of length N increases at most as N^2.

Figure 1 shows the structure of the ground state for a random chain of $N = 2000$ monomers with various choices of x (see (4)). For each x, the plot shows the total charge Q_i up to the ith charge

$$Q_i = \sum_{j=1}^{i} q_j \tag{9}$$

versus i and the vertical bars are the break points in the ground-state configuration (i.e. all the monomers between two successive vertical bars are on the same site). Clearly the number of vertical bars is R.

We see in figure 1 that as x decreases, R decreases (since the number of bars decreases) and the fluctuations of the distances ℓ between successive bars become more and more important.

2. The scaling of the ground-state energy E and of the number of occupied sites R

When one averages the ground-state energy E or the number of occupied sites R (at $T = 0$) over many samples (here, 5000 samples), one finds evidence that both E and R show power-law behaviour as N with exponents $\theta(x)$ and $\nu(x)$, which depend on x (see figure 2).

Because one can write the energy as

$$E = \sum_{r=1}^{R}[Q(r)]^2 \qquad \text{and} \qquad Q = N^x = \sum_{r=1}^{R} Q(r) \tag{10}$$
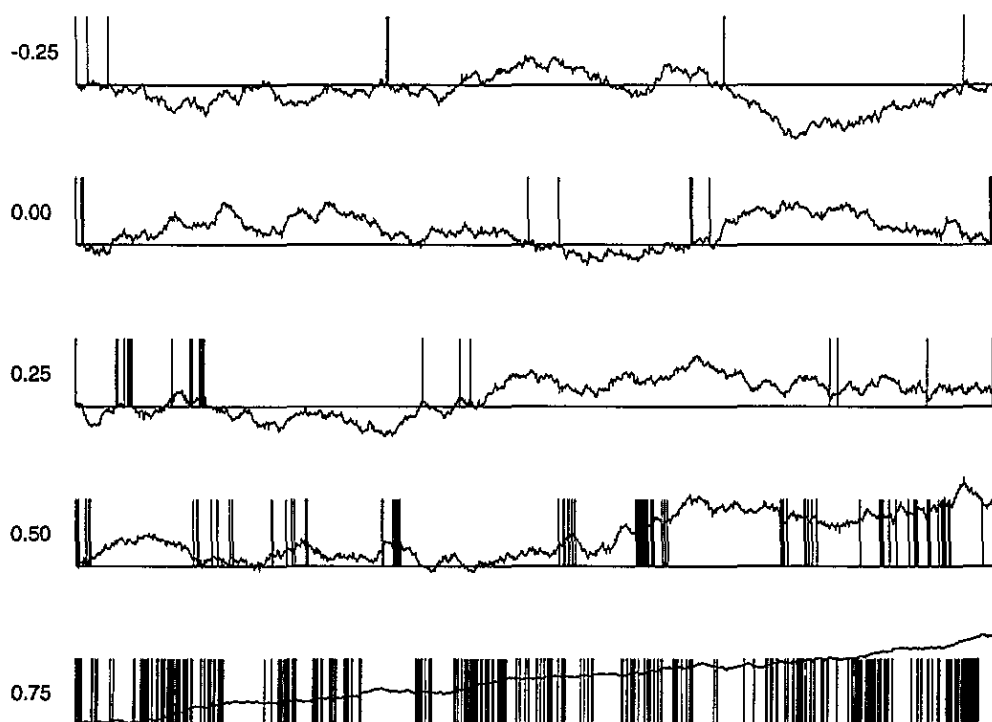
one always has

$$E \geqslant Q^2 / R \tag{11}$$

Figure 1. The cumulated charge Q_i defined by (9) versus i for chains of length $N = 1000$ and total charge $Q = N^x$ for various choices of x. The vertical bars mark the break points in the ground-state configuration.

which implies that

$$\theta(x) \geqslant 2x - \nu(x). \tag{12}$$

In figure 2(c), the ratios ER/Q^2 versus N is shown for various values of x. In all cases, the ratio seems to reach a limit as N increases, indicating that (12) is an equality rather than an inequality. The equality would mean that the total charge Q is evenly distributed among the R sites since a charge Q/R on each site would give an energy $E = R(Q/R)^2$.

As always, it is difficult to extract very reliable exponents from log–log plots. To estimate $\theta(x)$ and $\nu(x)$, we generated 2000 chains of sizes 50, 500 and 5000 and estimated these exponents by comparing $(\theta = \log(E_{500}/E_{50})/\log 10)$ the sizes 50 and 500 (white circles) and the sizes 500 and 5000 (black circles). Both for $\theta(x)$ and $\nu(x)$, the results (figure 3) show little difference between the two sets of data, giving confidence that $N \geqslant 50$ is close to the asymptotic regime.

The problem, of course, is to understand the x dependence of these exponents. We have not been able to find an exact expression for $\theta(x)$ and $\nu(x)$ for all x. Instead, we can give an argument which leads to
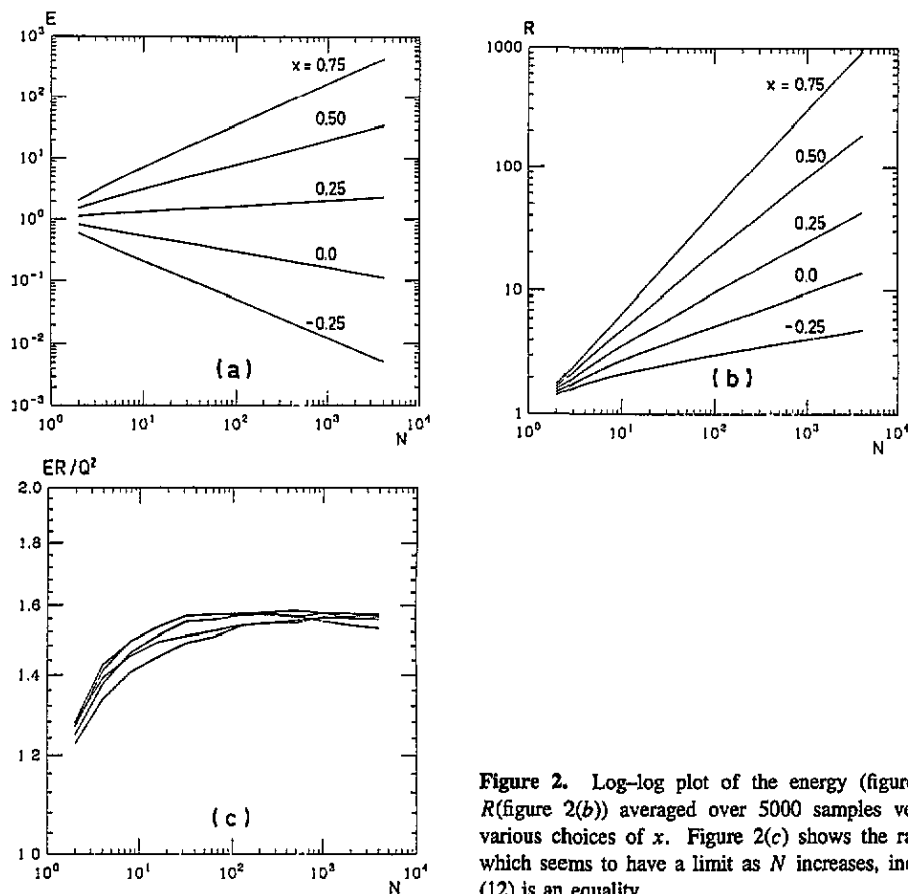
Figure 2. Log–log plot of the energy (figure 2(a)) and R(figure 2(b)) averaged over 5000 samples versus N for various choices of x. Figure 2(c) shows the ratio ER/Q^2 which seems to have a limit as N increases, indicating that (12) is an equality.
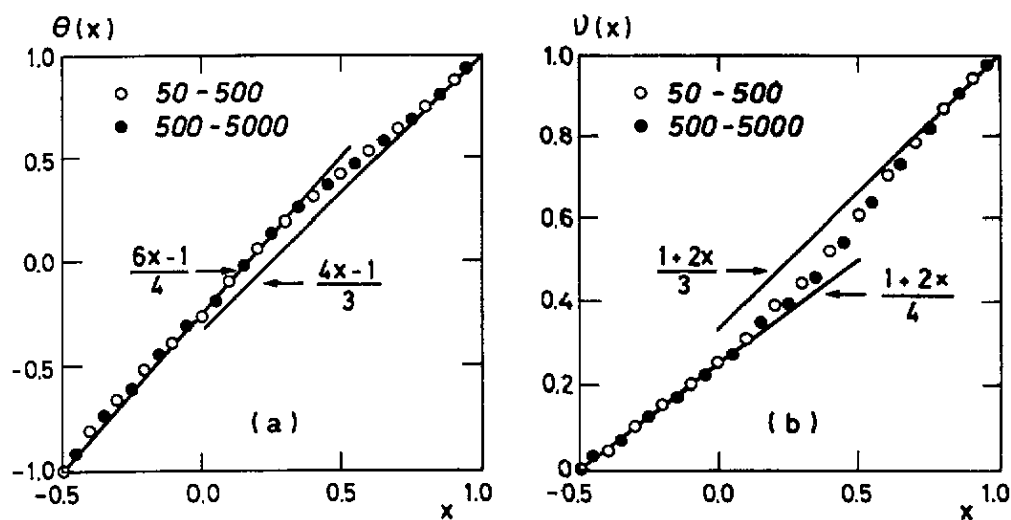

Figure 3. Estimates of the exponents $\theta(x)$ (figure 3(a)) and $\nu(x)$ (figure 3(b)) obtained by comparing the results for $N = 50$ and 500 (white circles) and $N = 500$ and 5000 (black circles). The statistics were on 2000 chains.

$$v(x) = \frac{1+2x}{3} \qquad \text{for } x \text{ close to } 1$$

$$v(x) = \frac{1+2x}{4} \qquad \text{for } x \text{ close to } -\tfrac{1}{2}. \tag{13}$$

These predictions together with the corresponding predictions for $\theta(x)$ using the equality in (12) are the straight lines shown in figure 3. They seem to be in good agreement with all data except for the neighborhood of $x = 1/2$ where there seems to be a crossover from one regime to the other.

Let us first present the argument which gives $v(x) = (1+2x)/3$. To obtain the ground-state energy, the best is to divide the total charge $Q = N^x$ as much as possible. If the chain occupies R sites, the best would be to have on each site a charge Q/R and the resulting energy would be

$$E = Q^2/R. \tag{14}$$

This expression tells us that to decrease E, R has to be as large as possible (and one would conclude that $R = N$ in the best choice). However, this is not possible because if $R \sim N$, there would be of the order of one monomer per site and hence a charge of the order of one per site, which is not consistent with $Q/R \sim N^{x-v(x)}$ as required to satisfy (14). It is therefore necessary to find the maximal value of R consistent with a charge of the order of Q/R on each site.

Here, it is useful to mention a simple fact on the sum S_ℓ of ℓ continuous random variables:

$$S_\ell = \sum_{i=1}^{\ell} q_i. \tag{15}$$

If one asks what the minimum value $S_{\min}$ of $|S_k|$ is for $1 \leqslant k \leqslant \ell$, one finds that

$$S_{\min} = \min_{1 \leqslant k \leqslant \ell} |S_k| \sim \ell^{-1/2}. \tag{16}$$

This is because the average number of points S_k in an interval of size ΔS around the origin varies like $\ell \, \Delta S/\sqrt{2\pi \ell}$. When $\Delta S \simeq S_{\min}$, this number is of order 1.

So if the chain is divided into R occupied sites and *if one assumes that the number ℓ of monomers does not fluctuate too much from site to site*, then ℓ is given by $\ell \sim N/R$, and the charge on each occupied site is of the order of $\ell^{-1/2} \sim (N/R)^{-1/2}$. Then the energy E can be obtained using (14) and the fact that

$$Q/R \sim (N/R)^{-1/2} \tag{17}$$

i.e.

$$R \sim N^{(1+2x)/3} \qquad \text{and} \qquad E \sim N^{(4x-1)/3}. \tag{18}$$

These predictions shown on figures 3 are in good agreement with the numerical simulations for x close to 1, but below $x = 1/2$ the agreement is clearly not satisfactory.

An important assumption used to arrive at (18) was that the number ℓ of monomers on each occupied site does not fluctuate too much. This seems not to be true according to figure 1 (at least for the smaller values of x). Therefore, one needs to know if large

fluctuations of ℓ are sufficient to invalidate the relation $\ell = N/R$. If one considers the total charge Q_i up to site i (9), Q_i performs a random walk and when $|Q_{i+\ell} - Q_i|$ is small enough, a break point (a vertical bar in figure 1) becomes possible. So the problem of the distance ℓ between successive vertical bars is similar to that of the first return close to the origin of a random walk. One knows from the theory of random walks [8, 9] that the probability of returning close to the origin for the first time after ℓ steps behaves like $\ell^{-3/2}$ for large ℓ. Therefore, if there are R occupied sites and if a typical number of monomers is ℓ_0 on each site, one expects the site with the largest number of monomers to have the number of monomers ℓ_{max} given by

$$R \int_{(\ell_{max}/\ell_0)}^{\infty} \frac{d(\ell/\ell_0)}{(\ell/\ell_0)^{3/2}} \sim 1 \qquad \text{i.e.} \qquad \ell_{max} \sim R^2 \ell_0. \tag{19}$$

This implies that most monomers are located on a very small number of sites, since $\ell_{max} \gg \ell_0 R$. Therefore, one arrives at

$$N \sim \ell_{max} \sim R^2 \ell_0 \tag{20}$$

which together with the relation we had before that $Q/R \sim \ell_0^{-1/2}$ and (14) leads to

$$R \sim N^{(1+2x)/4} \qquad \text{and} \qquad E \sim N^{(6x-1)/4}. \tag{21}$$

These predictions are in good agreement with the simulations at least for $x \leqslant 0$ (see figure 3). It is of course clear that $\nu(-1/2) = 0$ since if $Q = N^{-1/2}$, there is at most of order one other point with Q_i as small as $N^{-1/2}$.

Of course, one can wonder why (18) is more likely to be valid for x close to 1 and (21) for x close to $-1/2$. There is no very strong argument for that except for one more comparison with the statistics of the returns of a random walker close to the origin.

The main difference between the discussion which leads to (21) and the case $x > 1/2$ is that for a random walker, a strong enough bias (when $x > 1/2$) makes the probability of first return close to the origin after ℓ steps decay much faster than $\ell^{-3/2}$ for large ℓ. So the fluctuations of the number ℓ of monomers from site to site are expected to be much less important for $x > 1/2$, and, considering that the ℓ are all roughly the same, it is not too bad an approximation.

Even if the numerical results of figures 3 are well fitted for $x > 1/2$ and $x < 0$ by expressions (18) and (21), the intermediate range does not seem to agree with either of these formulae and it is not clear to us whether another expression for the exponents should hold in this intermediate regime or whether (18) and (21) could be valid for $\frac{1}{2} < x < 1$ and $-\frac{1}{2} < x < \frac{1}{2}$, respectively, the discrepancy in figure 3 being due to a slow convergence problem.

3. Conclusion

We saw that for the problem of directed random heteropolymers with a continuous distribution of charges, the energy E and the end-to-end distance R at $T = 0$ seem to scale like power laws $E \sim N^{\theta(x)}$ and $R \sim N^{\nu(x)}$ with exponents $\theta(x)$ and $\nu(x)$ which vary with the total charge $Q = N^x$ of the chain. We have obtained approximate expressions of these exponents (18) and (21) which agree rather well with the results of simulations at least in some ranges of value of x. According to figure 3, the exponent ν equals $\nu(1/2)$
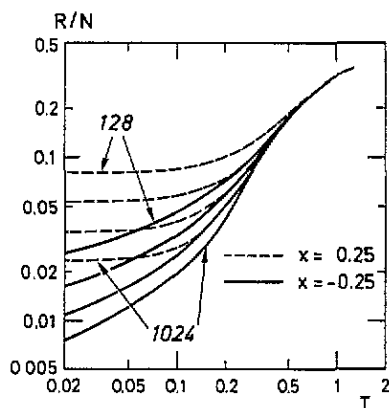
Figure 4. R averaged over 1000 samples for sizes $N = 128, 256, 512$ and 1024 and with $x = 0.25$ and $x = -0.25$. As N increases, the range of temperatures where finite-size effects are visible decreases.

$\simeq 0.6$ for unconstrained chains (i.e. when Q is not fixed). It is worth noting that this is rather close to the prediction of Kantor and Kardar for the undirected case.

One can wonder how the x dependences of the zero temperature properties can manifest themselves at non-zero temperature. Figure 4 shows the temperature dependence of R for various sizes N and various x. The results indicate that the finite-size effects and the x dependence seem to occur in a range of temperature which shrinks as $N \to \infty$.

We thus think that a picture consistent with the results shown in figure 4 is that when $N \to \infty$ at a fixed temperature, the number of occupied sites R is extensive,

$$\frac{R}{N} \to r(T) \tag{22}$$

and $r(T)$ is independent of x. The temperature range in which the x dependence is noticeable varies probably as a power law $N^{-\lambda(x)}$ with an exponent $\lambda(x)$, which should vary with x; hence one would expect for $T \sim N^{-\lambda(x)}$ the dependence

$$R \sim N^{\nu(x)} h\left(T N^{\lambda(x)}, x\right). \tag{23}$$

We did not find a good way of extracting the exponent $\lambda(x)$ from our data of figure 4 and so we do not present here any prediction for $\lambda(x)$. However, if the picture of the two regimes (22) and (23) is right, then for R/N to be independent of x for large $T N^{\lambda(x)}$, one would need that $(1 - \nu(x))/\lambda(x)$ be independent of x.

A similar low-temperature range where the exponents are x-dependent and which shrinks as N increases seems to be present in several other models; we hope to discuss this in more detail in the future.

If this picture is right, it would be the existence of these low-temperature regimes which are responsible for the non-trivial exponents $\theta(x)$ and $\nu(x)$. In these low-temperature regimes, the sample-to-sample fluctuations would become important. It is worth noting that in these low-temperature regimes, the exponents $\nu(x)$ are due to non-extensive changes of energy. This is rather reminiscent of the fact that for homopolymers with excluded volume, the Flory approximation predicts an exponent ν by calculating contributions to the free energy which are non-extensive.

Acknowledgments

References

[1] Obukhov S P 1986 *J. Phys. A: Math. Gen.* **19** 3655
[2] Bryngelson J D and Wolynes P G 1987 *Proc. Natl. Acad. Sci.* **84**; 1990 *Biopolymers* **30** 177
[3] Garel T and Orland H 1988 *Europhys. Lett.* **6** 307, 597
[4] Shakhnovich E I and Gutin A M 1989 *Europhys. Lett.* **8** 327; 1989 *J. Phys. A: Math. Gen.* **22** 1647
[5] Higgs P G and Joanny J F 1991 *J. Chem. Phys.* **94** 1543
[6] Kantor Y and Kardar M 1991 *Europhys. Lett.* **14** 421
[7] Derrida B, Griffiths R B and Higgs P G 1992 *Europhys. Lett.* **18** 361
[8] Montroll E W and West B J 1979 *Fluctuations Phenomena* ed Montroll-Lebowitz (Amsterdam: North Holland)
[9] Haus J W and Kehr K W 1987 *Phys. Rep.* **150** 265
[10] Foster D P, C. Van der Zande and Yeomans J 1992 *J. Stat. Phys.* **69** 857
[11] Victor J M and Imbert J B 1993 *Europhys. Lett.* **24** 189
[12] Wittmer J, Johner A and Joanny J F 1993 *Europhys. Lett.* **24** 263
[13] Gutin A M and Shakhnovich E I 1993 *J. Chem. Phys.* **98** 8174
[14] Kantor Y, Kardar M and Li H 1994 *Phys. Rev. E* **49** 1383
[15] Joanny J F 1994 *Europhys. Lett.* (to be published)