

Polymers with Random Self-Interactions.

Y. KANTOR(*) and M. KARDAR(**)

(*) *School of Physics and Astronomy, Tel Aviv University - 69978 Tel Aviv, Israel*

(**) *Physics Department, Massachusetts Institute of Technology
Cambridge, MA 02139, USA*

(received 18 September 1990; accepted in final form 17 December 1990)

PACS. 36.20 – Macromolecules and polymer molecules.

PACS. 64.60 – General studies of phase transitions.

PACS. 75.10N – Spin-glass models.

PACS. 87.10 – General, theoretical, and mathematical biophysics (inc. logic of biosystems, quantum biology, and relevant aspects of thermodynamics, information theory, cybernetics, and bionics).

Abstract. – We examine macromolecules with quenched random (on the average vanishing) charges interacting either via long- or short-range interactions. If charges of the same sign repel each other, at low space dimensions such interactions produce swollen states. For short-range one-dimensional interactions, we find the lowest-energy configuration by a complete enumeration of all states for chains up to length $L = 22$. Averaging over the quenched charges results in a swelling exponent $\nu \approx 0.57$. For power law interactions an exact value of the exponent ν is found, which in the particular case of Coulomb interactions in $d = 3$ gives $\nu = 1$.

Complex structural conformations of biological macromolecules [1] have inspired a number of attempts to use concepts from the statistical mechanics of random systems, such as spin glasses [2], or neural networks, to understand some generic aspects of their behavior. The models usually schematically represent the macromolecules as a specified linear sequence of L elements (amino acids, bases, etc.) subject to a variety of complicated interactions, which eventually determine its three-dimensional shape. The interactions, in increasing order of sophistication, fall into the following groups: *a*) The strengths and sign of the $L(L-1)/2$ interactions between pairs of monomers are chosen at random [3, 4]. *b*) A particular interaction is selected for a specified pair of elements, and the randomness resides in the sequence of elements [5]. *c*) In neural-network-inspired models [6, 7] the interactions are «learned» by using data from known structures. While in *c*) the main concern is to reproduce the structure of specific molecules, approaches *a*) and *b*) attempt to describe the generic nature of freezing into such folded structures. In this paper we investigate a model described by interactions in approach *b*), and point out how the nature of these interactions (range, sign, symmetry) controls the possible shapes of resulting macromolecules. In the majority of cases, the molecules «fold» into compact structures [5]. However, we find that if like elements repel and unlike elements attract, the resulting configurations tend to be swollen. The swelling is described by the exponent ν , which relates

the radius of gyration (r.m.s. size) R_g of the chain to its linear size (number of monomers) L : $R_g \sim L^\nu$. For Coulomb interactions in three dimensions the exponent ν obtains its maximum value of unity. For short-range interactions we used a numerical approach to establish a value of $\nu \approx 0.57$ in one dimension.

Let us examine the simplest construction of such a random chain. The elements, henceforth referred to as charges, will belong to two species. The primary sequence of the chain is specified by the set $\{q_i = \pm 1\}$. The actual configuration of the chain in an external d -dimensional space is specified by positions r_i of the monomers. These configurations occur with probabilities proportional to the Boltzmann weights of a Hamiltonian

$$\mathcal{H}(\{q_i\}; \{r_i\}) = \sum_{0 \leq i < j \leq L} U_{q_i q_j}(r_i - r_j). \quad (1)$$

The two-body interaction can be short-ranged (SR), or long-ranged (LR), and we have neglected more than two-body interactions. As pointed out by Obukhov [8], the most general form of the two-body interaction is

$$U_{q_i q_j} = b + u(q_i + q_j) + V q_i q_j. \quad (2)$$

The constant b is the effective two-body interaction, and randomness is relevant only for $b \approx 0$. The importance of the second term, proportional to u , has been discussed by Obukhov: it leads to compact shapes, since irrespective of the sign of u charges of one sign are attracted to each other, while charges of the opposite sign repel. One can then construct a state in which all attracting charges compactify to a point, while the repelling charges form a cloud around them. The energy of such a configuration is proportional to uL^2 , and, thus, the collapse of a long chain will occur at any finite temperature (since the entropy can only grow as L). If the interaction is symmetric under the operation $\{q_i\} \rightarrow \{-q_i\}$, as in the case of Coulomb interactions, u has to be zero. Then $U_{q_i q_j}(r_i - r_j) = q_i q_j V(r_i - r_j)$. Even in this simple case, there are two distinct possibilities.

a) Like charges attract ($V < 0$). Such a situation occurs for similar hydrophobic monomers which effectively attract each other in an attempt to minimize their interaction with a solvent. For SR forces with $V = -v_0 \delta_{r_i r_j}$ ($v_0 > 0$), it is then easy to see that the system minimizes its energy by simply collapsing into a configuration consisting of two spatial groups (one containing all the positive charges, and the other all the negative ones) separated by the interaction range. Again this type of state has energy proportional to $v_0 L^2$, and its collapse is inevitable at any finite temperature. To mitigate such a behavior, Garel and Orland [5] assume that v_0 scales as $1/L$, thus making the ground-state energy simply proportional to the system size L , and permitting a finite-temperature phase transition from a collapsed to an extended (random-walk-like) state.

b) Like charges repel ($V > 0$). This is the case when the monomers interact via a (LR or screened) Coulomb potential; it also describes the effective repulsion between identical bases of RNA. The macromolecule now behaves quite differently: it prefers to be extended and its free energy is extensive in L . The difference in behavior is easily understood by examining a single chain (or a portion thereof). While the average charge is statistically zero, individual chains will have net positive or negative charge. In either case the self-repulsion of charges leads to an expansion. In more realistic models of protein folding [5, 6] it is recognized that a single «charge» is not sufficient to characterize a monomer, and its interactions with other monomers. The model is then extended by characterizing each

element by a set of attributes $\{q_i^\alpha = \pm 1\}$, where $\alpha = 1, 2, \dots, m$ enumerates the attributes, and using a generalized interaction $q_i^\alpha q_j^\alpha V_\alpha(r_i - r_j)$. As the number of such attributes m increases, the model approaches the random pair potential limit of approach a) above. There can now be a competition between the above two types of interactions (and also uniform two-body interactions). However, as described above, the relative importance of the attractive interactions grows with the chain length L , and such interactions dominate the resulting behavior.

We thus see that practically all choices of random interactions ($u \neq 0$, or $V < 0$) lead to compact folded state. Since such situations have been the focus of most previous work, here we examine the swollen polymers obtained for the repulsive case ($V > 0$). The repulsive interactions also have the virtue of having a proper thermodynamic behavior (*i.e.* extensive free energies), without the need to scale interaction parameters with the chain length L . We considered both LR (power law) and SR repulsive interactions. The continuum version of these interactions is given by a generalized Edwards Hamiltonian

$$\frac{\mathcal{H}}{k_B T} = \frac{K}{2} \int_0^L dx \left(\frac{d\mathbf{r}}{dx} \right)^2 + \frac{1}{2} \int_0^L \int_0^L dx dx' q(x) q(x') V(\mathbf{r}(x) - \mathbf{r}(x')), \quad (3)$$

where K is the (entropic) elasticity constant, while the discrete summation over the monomer indices i in (1) has been replaced by an integration over the internal coordinate x , and \mathbf{r} is the position vector in a d -dimensional embedding space. One may use $V(r) \sim 1/r^{d-\alpha}$, with $\alpha = 2$ for Coulomb interactions, and $V(r) \sim \delta^d(\mathbf{r})$ for SR interactions. The quenched random sequence is assumed to be uncorrelated, *i.e.* $\langle q(x)q(x') \rangle = \delta(x - x')$, where $\langle \rangle$ represents the averaging over quenched charge configurations. Let us now estimate the swelling exponent ν , by a dimensional analysis of eq. (3). Under a rescaling $x \rightarrow \lambda x$ and $\mathbf{r} \rightarrow \lambda^\nu \mathbf{r}$, the entropic force constant K scales by a factor λ^{y_K} , with $y_K = 2\nu - 1$, while the repulsive LR interaction strength scales as $\lambda^{y_v^{(lr)}}$, with $y_v^{(lr)} = 1 - (d - \alpha)\nu$. For SR interactions the latter exponent is $y_v^{(sr)} = 1 - d\nu$, *i.e.* it corresponds to the limit $\alpha = 0$. In the absence of interactions, the usual random walk behavior is obtained with $\nu = 1/2$ which leaves K scale invariant. At this fixed point, the LR interactions become relevant below the upper critical dimension $d_u^{(lr)} = 2 + \alpha$. (For SR interactions $d_u^{(sr)} = 2$.)

Below the upper critical dimension, an *approximate Flory estimate* [9] of the exponent ν is obtained by equating y_K and y_v leading to $\nu_F^{(lr)} = 2/(d + 2 - \alpha)$ for LR interactions, and $\nu_F^{(sr)} = 2/(d + 2)$ for SR interactions. However, in performing a renormalization group (RG) treatment of power law interactions for homogeneously charged polymers Pfeuty *et al.* [10] observed that there is no perturbative correction to the LR term, thus permitting an *exact* calculation of ν . Similar arguments have been used for generalized LR interactions for D -dimensional manifolds [11]. This observation is simply illustrated by considering two polymer segments, a distance r apart. The net interaction between the two segments falls off asymptotically as $Q_1 Q_2 / r^{d-\alpha}$, where Q_1 and Q_2 are the net charges of the two segments. Various rearrangements of charges within each segment do not modify its total charge. However, under rescaling the segment is replaced by an elementary renormalized monomer which has to be described by its charge, as well as by higher «multipole moments». But since the interactions between higher multipoles decay faster than the interaction between net charges, they are irrelevant in the RG sense. This argument, generalized from the case of uniform charges, then implies the nonrenormalizability of the leading interaction, *i.e.* $y_v^{(lr)} = 0$, which leads to

$$\nu_{RG}^{(lr)} = 1/(d - \alpha). \quad (4)$$

According to eq. (4), $\nu = 1$ for $\alpha = d - 1$. This is in fact the case for Coulomb interactions in $d = 3$. The random chain is thus stretched enough to maintain a specific orientation on average. Of course the extent of transverse fluctuations can still be quite large [11]. For $\alpha > d - 1$, ν remains equal to unity, which is its maximum allowed value. Note that in this range, the Flory estimate still predicts a crumpled chain ($\nu < 1$). We indirectly verified the predictions of RG by numerical simulations of the one-dimensional chains with α varying between 0 and 1. In this range we expect $\nu = 1$, while $\nu_{\text{F}}^{(\text{sr})}$ varies between $2/3$ and 1. The numerical results were consistent with $\nu = 1$ as predicted by RG. As α is reduced from $d - 1$, ν decreases according to eq. (4). However, it cannot go below the value of $\nu = \nu^{(\text{sr})}$. This is because at this point the basic assumption that the asymptotic interaction at long distances is unrenormalized is no longer valid; corrections generated by short-range effects will now dominate the potential, and $\nu = \nu^{(\text{sr})}$ for $\alpha > d - 1/\nu^{(\text{sr})}$. Since negative values of α are meaningless in the presence of a lattice cut-off, the above restrictions on the validity of eq. (4) are summarized by the inequalities $d - 1 > \alpha > \max(d - 1/\nu^{(\text{sr})}, 0)$. Thus for $d = 3$, $\nu = 1$ for $\alpha \geq 2$, $\nu = 1/(3 - \alpha)$ for $1 < \alpha < 2$, and $\nu = 1/2$ for $\alpha \leq 1$ (assuming $\nu^{(\text{sr})} = 1/2$), while for $d = 1$, $\nu = 1$ for all the allowed values of α .

To obtain the correct behavior of polymers with SR interactions (corresponding to $V(r_i - r_j) = \delta_{rr_j}$ in eq. (1)) in $d \leq 2$ requires a more complicated RG treatment which is not attempted here. Due to the «repulsive nature» of the three-monomer interactions, we expect an exponent $\nu > 1/2$. As is characteristic of many other systems where randomness is relevant, the energy landscape in configuration space is expected to have deep minima. Hence the conformation, and R_g for a given selection of quenched charges is determined by the lowest-energy configuration, and the thermodynamics of the chain is reduced to the investigation of the behavior of the lowest-energy state. The mean energy of the ground state is expected to increase linearly with its size L , *i.e.* there is a self-averaging for the energy (and also the free energy). Sample-to-sample fluctuations in the minimal energy $\sigma(E)$, however, can increase as L^θ . Indeed, in the space dimension $d = 0$, all charges are located at the same point, and $H = \sum_{i < j} q_i q_j$ depends only on the charge distribution $\{q_i\}$.

Obviously the mean energy vanishes, while its standard deviation increases linearly with L , *i.e.* $\theta = 1$. We expect θ to decrease with increasing d .

We numerically investigated the behavior of θ and ν in $d = 1$. Our model is a discrete one-dimensional L -step [$(L + 1)$ -site] random walk, with $L = 2, 3, 4, \dots, 22$. The quenched values of q_i are selected from a standard Gaussian distribution. For each quench $\{q_i\}$ all possible 2^{L-1} configurations are examined. Note that the charge neutrality of the chain is enforced only in the average over the quenches. (The L -step random walk is characterized by specifying the directions of $L - 1$ steps; the first step is taken in a fixed direction.) From all these configurations we find the one with the lowest energy $E(\{q_i\})$ and also record its squared end-to-end distance $R^2(\{q_i\})$. We then calculate the quenched averages $E \equiv \langle E(\{q_i\}) \rangle$ and $R^2 \equiv \langle R^2(\{q_i\}) \rangle$ by averaging the results over 8192 independent quenches $\{q_i\}$. The computational effort in such an exhaustive search for the absolute minimum grows exponentially: for $L = 22$ we examined more than 10^{10} different configurations, requiring several days of CPU time on a Sun 4/110 minicomputer. Therefore, our results are limited to L 's of moderate size.

The minimum-energy configurations exhibit surprisingly little departure from a random walk: fig. 1 depicts the dependence of R^2/L on L in a logarithmic scale. This representation exposes the departure from the random-walk behavior, since the slope of the curve is $2\nu - 1$. The scatter of the data points is due to the finite number of quenches. From this data we find that $\nu = 0.574 \pm 0.006$, where the error bars represent the statistical uncertainty due to the scatter of the data points. This result clearly excludes $\nu_{\text{F}}^{(\text{sr})} = 2/3$, predicted by the Flory-type theory. Note that we have assumed that the end-to-end distance scales the same as the

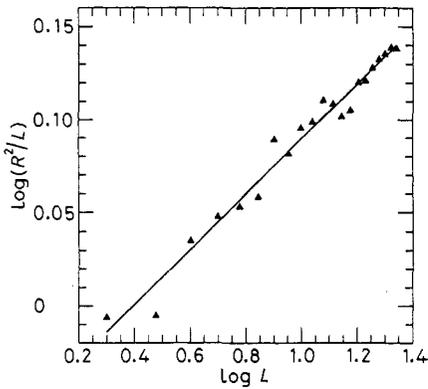


Fig. 1.

Fig. 1. – Logarithmic plot of the R^2/L vs. L , where R^2 is the mean squared end-to-end distance of the minimum-energy configuration for a chain of length L . The slope $2\nu - 1$ of least-square-fitted line corresponds to $\nu = 0.574$.

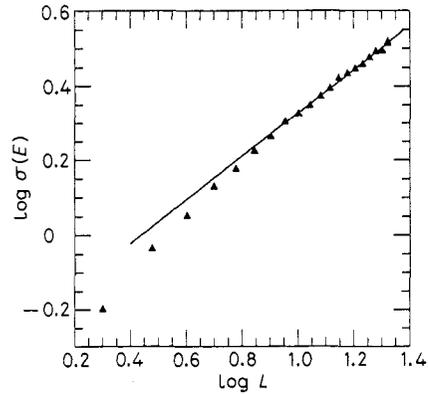


Fig. 2.

Fig. 2. – Logarithmic plot of the standard deviation $\sigma(E)$ in the energy of minimum-energy configurations vs. the chain length L . The straight line is a least-squares fit through the last 10 data points, with slope $\theta = 0.58$.

radius of gyration. One can imagine a multiscaling distribution in which this is not the case. However, we saw no anomalous behavior for the standard deviation of R^2 , lending support to the assumption of a single scaling exponent ν . As expected, the mean minimal energy E increases linearly with L , but fluctuates very strongly from one quench to another. The statistical distribution of the minimal energies for the values of L under consideration ranges from 0 to some negative value, and is peaked close to its lower end. The standard deviation of the minimal energies grows with L , as can be seen from fig. 2. The data points are quite scattered and display a rather strong curvature. For these two reasons it is difficult to reliably extrapolate θ to large values of L . A reasonable estimate is obtained by drawing a straight line representing the least-square fit to the last ten data points. Its slope is $\theta = 0.58$.

We finally note some similarities, and differences between this problem, and that of the fluctuations of a directed polymer in a random medium [12]. Both problems exhibit superdiffusive behavior ($\nu > 1/2$), and large fluctuations in sample-to-sample free energy ($\theta > 0$) due to impurities. For the pinning of the directed polymer, it is believed that the elastic energy of stretching scales the same way as the typical energy fluctuations. This leads to the important exponent identity [13] $2\nu - 1 = \theta$ which is verified by numerical simulations. The values of θ and ν quoted above from the simulations in $d = 1$ clearly do not satisfy this identity, and it would be interesting to understand the origin of this difference. Both problems have the same critical dimension of two for the relevance of small randomness. However, for directed polymers the randomness is marginally relevant in $d = 2$, and consequently a phase transition between weak and strong pinning behavior is predicted [12] in $d > 2$. In the absence of a similar RG treatment of the randomly self-interacting polymer, we cannot rule out such a scenario, *i.e.* a transition from a random-walk behavior to a swollen frozen state as randomness increases in $d > 2$. It is plausible to assume that the random chain cannot be more swollen than the regular self-avoiding walk. Thus, for $d > 4$ the value $\nu = 1/2$ seems to be the only possibility. Also, in a mean-field approach [5], we did not observe a phase transition. Despite these indications, the possibility of a $\nu > 1/2$ in

$2 < d < 4$ is not excluded, and due to the relevance of this question to polymers in three dimensions, it may be worthwhile to undertake further numerical and RG studies of randomly self-interacting polymers. Another interesting question is whether our results are to be modified if for each polymer we add the global constraint of strict charge neutrality.

* * *

This research was supported by the U.S.-Israel Binational Science Foundation under Grant No. 87-00010 (YK) and by the NSF Grant No. DMR-90-01519, and the PYI program (MK).

REFERENCES

- [1] See, e.g., CREIGHTON T. E., *Proteins: Their Structure and Molecular Properties* (Freeman, San Francisco, CA) 1984.
- [2] STEIN D. L., *Proc. Natl. Acad. Sci. U.S.A.*, **82** (1985) 3670; BRYNGELSON J. D. and WOLYNES P. G., *Proc. Natl. Acad. Sci. U.S.A.*, **84** (1987) 7524.
- [3] GAREL T. and ORLAND H., *Europhys. Lett.*, **6** (1988) 307.
- [4] SHAKHNOVICH E. I. and GUTIN A. M., *Europhys. Lett.*, **8** (1989) 327.
- [5] GAREL T. and ORLAND H., *Europhys. Lett.*, **6** (1988) 597; GAREL J.-R., GAREL T. and ORLAND H., *J. Phys. (Paris) Lett.*, **50** (1989) 3067.
- [6] FRIEDRICHS M. S. and WOLYNES P. G., *Science*, **246** (1989) 371.
- [7] BOHR H. J., BURNAK S., COTTERILL R. M. J., FREDHOLM H., LAUTRUP B. and PETERSON S. B., *FEBS Lett.* (1990), in press.
- [8] OBUKHOV S. P., *J. Phys. A*, **19** (1986) 3655.
- [9] FLORY P., *Statistics of Chain Molecules* (Interscience Publishers, New York, N.Y.) 1969.
- [10] PFEUTY P., VELASCO R. M. and DE GENNES P. G., *J. Phys. (Paris) Lett.*, **38** (1977) L-5.
- [11] KANTOR Y. and KARDAR M., *Europhys. Lett.*, **9** (1989) 53.
- [12] KARDAR M. and ZHANG Y.-C., *Phys. Rev. Lett.*, **58** (1987) 2087.
- [13] HUSE D. A. and HENLEY C. L., *Phys. Rev. Lett.*, **54** (1985) 2708.