Energy level structure of two-dimensional self-interacting random chains

1992 J. Phys. A: Math. Gen. 25 L121

(http://iopscience.iop.org/0305-4470/25/3/007)

J. Phys. A: Math. Gen. 25 (1992) L121–L126. Printed in the UK

LETTER TO THE EDITOR

Energy level structure of two-dimensional self-interacting random chains

Masataka Fukugita†, David Lancaster‡ and Mark G Mitchard‡

† Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606, Japan
‡ Department of Physics, Kyoto University, Kyoto 606, Japan

Received 18 November 1991

Abstract. We study the energy level structure of the lowest-lying states for a two-dimensional version of the random chain model proposed by Iori, Marinari and Parisi. We find that the multifold degeneracy of the ground state is lifted as the random noise interaction is switched on and a solitary global minimum emerges in a glassy phase as the strength of the noise term is increased. This indicates that the model mimics the characteristics expected of a native protein. We also show that the transition from the globule state (with no noise term) to the glassy phase takes place smoothly.

Protein folding provides us with an interesting problem, not only in view of its practical importance in biochemistry, but also in the statistical mechanics sense that it exhibits a curious frustration similar to spin glasses. It is tempting to study the universal physical properties that determine the behaviour of this system. Recently, Iori, Marinari and Parisi (IMP) [1] proposed a self-interacting random chain model, which seems to have some of the characteristics expected of real proteins. For a chain of N sites their Hamiltonian is

$$H = \sum_{1 \leq i < j \leq N} \left(\delta_{i,j+1} r_{i,j}^2 - \frac{A}{r_{ij}^6} + \frac{R}{r_{ij}^{12}} + \frac{\varepsilon \eta_{ij}}{r_{ij}^6} \right).$$

The first term is a harmonic force that represents the bond holding adjacent sites together. The second and third terms are the conventional Lennard-Jones potential, consisting of a van der Waals attraction and a repulsive hard core. The last term can be regarded as a noise applied to the system, which models the varying potentials between pairs of atoms in a protein. Specifically, η_{ij} is an uncorrelated set of values taken from a uniform distribution of mean zero and variance 1, with ε determining the strength of the interaction.

Using standard Monte Carlo techniques, IMP have shown that systems in which ε is sufficiently large display a so-called glassy phase with several local minima. In long Monte Carlo simulations the system samples a small number of metastable states which can persist for extended periods, and the same state is visited after an interval of many tens of millions of updates.

In this communication we attempt to elucidate the low lying energy level structure of this system, and to clarify the nature of the states. We work in two dimensions, which we argue provides an even simpler laboratory for studying the problem [2]. In two dimensions the phase space is greatly restricted, so the chain folding is more easily

0305-4470/92/030121+06$04.50

frustrated than in the three-dimensional case. We therefore hope that simulations with rather short chains will exhibit some of the behaviour of much longer chains in three dimensions. The two-dimensional model also has the great advantage that the folded configurations are easy to visualize. We have used both simulated annealing [3] and Monte Carlo techniques, choosing whichever is more appropriate to the problem in hand. We also employ a deterministic quenching procedure to bring configurations to the nearest local energy minimum, and this allows us to accurately classify states. We fix our chain to have $N=10$ points, take $A=2$, $R=2$, and work mainly with one set of η_{ij}'s. The strength of the random part of the potential, ε, is varied between 0 and 5.

Let us first justify that the two-dimensional version of the model retains the generic characteristics found by IMP. In figure 1(a) we show a Monte Carlo time history of the mean square deviation δ^2 of the chain from a reference configuration, where δ^2 between configurations α and β is defined by

$$\delta^2_{\alpha,\beta}=\frac{1}{N}\sum_{i=1,N}(\boldsymbol{r}_i^{(\alpha)}-\boldsymbol{r}_i^{(\beta)})^2$$

minimized over translations, rotations and reflections. The simulation is made at $\beta=2$ and $\varepsilon=\sqrt{10}$. We clearly see the same features as IMP, signalling a glassy phase. Figure 1(b) shows the same data after quenching to $\beta=\infty$. This removes irrelevant thermal noise from the system, and allows us to see the jumping behaviour directly from a graph of the energy, which is otherwise impossible: figure 1(c). In order to explore the ground state of the system we employed simulated annealing as the best method [4]. Typically we cooled from $\beta=0.25$ to $\beta=4.0$ and 50 temperature steps and 20 000 Monte Carlo sweeps at each temperature step. Studies with a variety of different cooling schedules, where we count the frequency with which states appear, lead us to be confident that we have obtained the ground state for each of our chosen values of ε. In addition, we have found the ten or so lowest-lying states in each case, at least insofar as we are not missing states which resemble, and are only separated by low potential barriers from, other states in our list.

By comparing our list of the low-lying states with those in the Monte Carlo runs above, we immediately notice that the latter does not contain the lowest states. This discrepancy suggests that IMP are quite likely to have missed the ground state of their model. If the simulation is carried out at substantially higher temperature, $\beta=1$ say, the time history is very noisy and does not resemble the plot obtained by IMP. However, using the quenching technique described above, we are able to identify the metastable states which are accessed. In this case the set of states does contain the ground state, and in fact the few lowest-lying states agree with the list obtained by simulated annealing. This further corroborates our claim to have discovered the lowest-lying states of the system.

Having established the model and techniques, we proceed to explore the energy level structure of the lowest-lying states. For $\varepsilon=0$ the system displays a multifold degeneracy in energy, both at the ground state and at the excited states, which are separated from the former by substantial gaps. The origin of this is clear when the configurations are examined: the sites of the chain lie on the vertices of a triangular lattice, a typical configuration being shown in figure 2(a). To the extent that the R/r^{12} term in the potential can be regarded as providing a hard sphere repulsion, the sites of the folded chain will pack as hard spheres in a plane under the influence of the attractive terms of the potential. The minimum energy configurations then live on the lattice in which the number of adjacent sites is maximized: for $N=10$ this is a unique

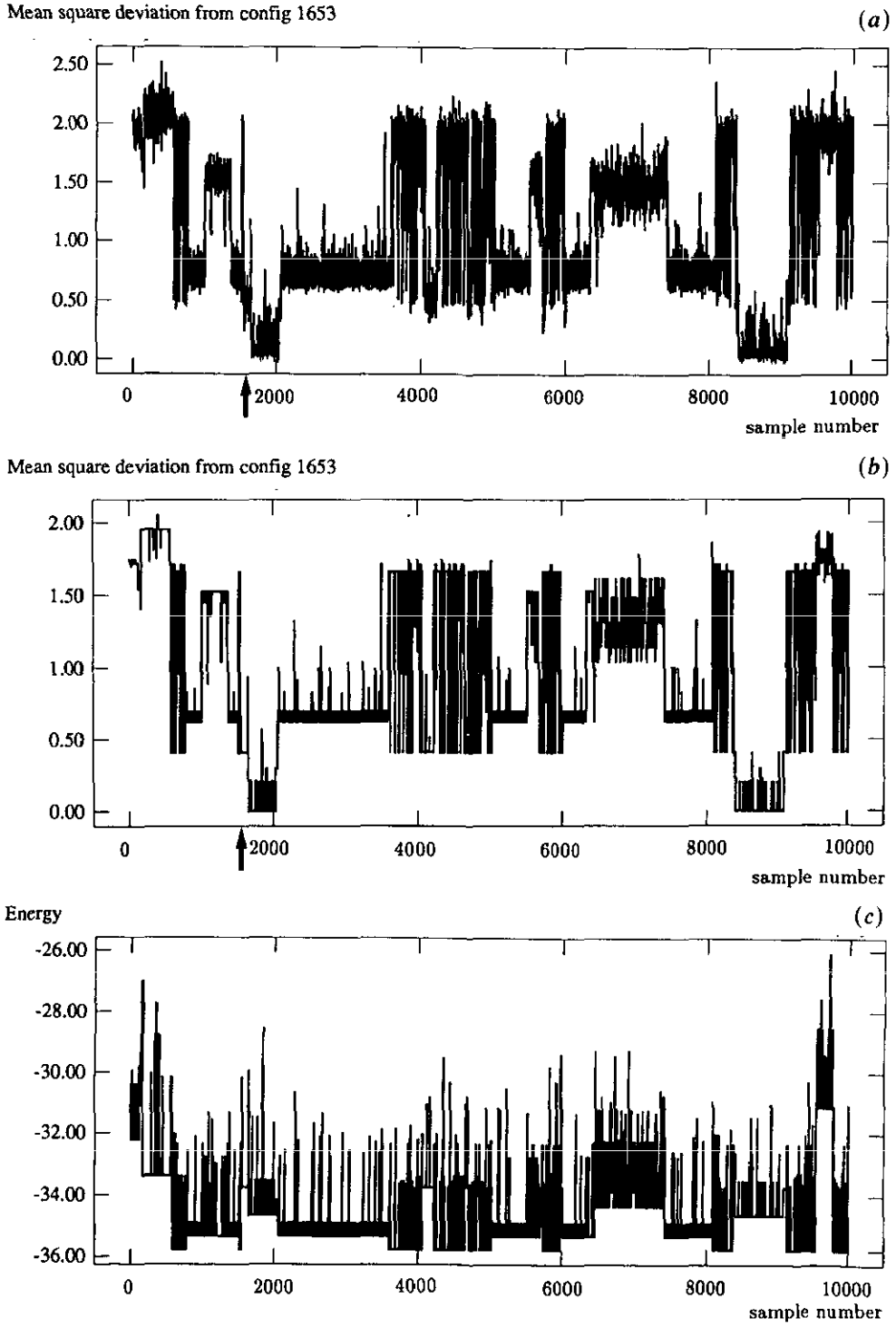

Figure 1. Monte Carlo evolution of the chain at $\varepsilon = \sqrt{10}$. Samples are taken every 10 000 complete chain updates, up to a total of 10^8 updates. (*a*) Mean squared deviation from an arbitrary reference configuration (in this case sample number 1653, marked with an arrow). (*b*) Same quantity as in (*a*), calculated after quenching. (*c*) Energy of quenched data. Note that the ground state energy at this value of ε is -37.7, and this state is not accessed in this run.

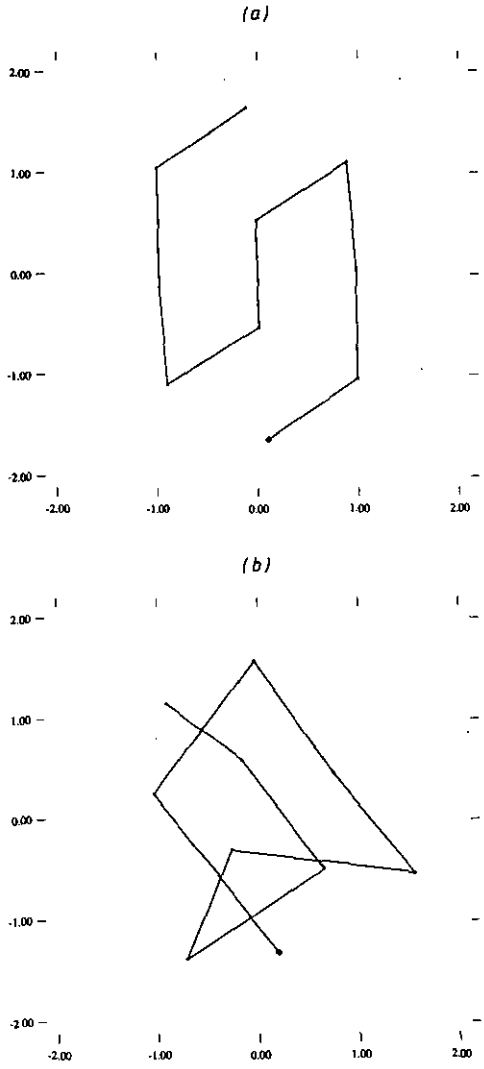

Figure 2. (*a*) Typical ground state configuration at $\varepsilon = 0$. (*b*) Ground state configuration at $\varepsilon = \sqrt{10}$.

lattice with 19 nearest neighbour pairs. We have counted some 100 possible ways of putting the chain on this lattice. The next set of degenerate states correspond to lattices in which the total number of adjacent sites (18) is one fewer. We also see configurations with 17 nearest neighbour pairs, but none with fewer than 17. We have found that configurations on lattices with 16 and fewer pairs of adjacent sites are generally unstable.

When the noise is applied to the system, this degeneracy is lifted. In figure 3 we present the energy level structure as explored by a series simulated annealing runs for each of $\varepsilon = 0$, $\sqrt{0.4}$, $\sqrt{2}$ and $\sqrt{10}$. For the last three values we also show a plot of the energy levels obtained from a high temperature Monte Carlo run followed by quenching. The Monte Carlo method is more efficient at generating states, but the simulated annealing runs are necessary to ensure that the Monte Carlo is finding the ground state. Simulated annealing is specifically designed to find low lying states, so it is

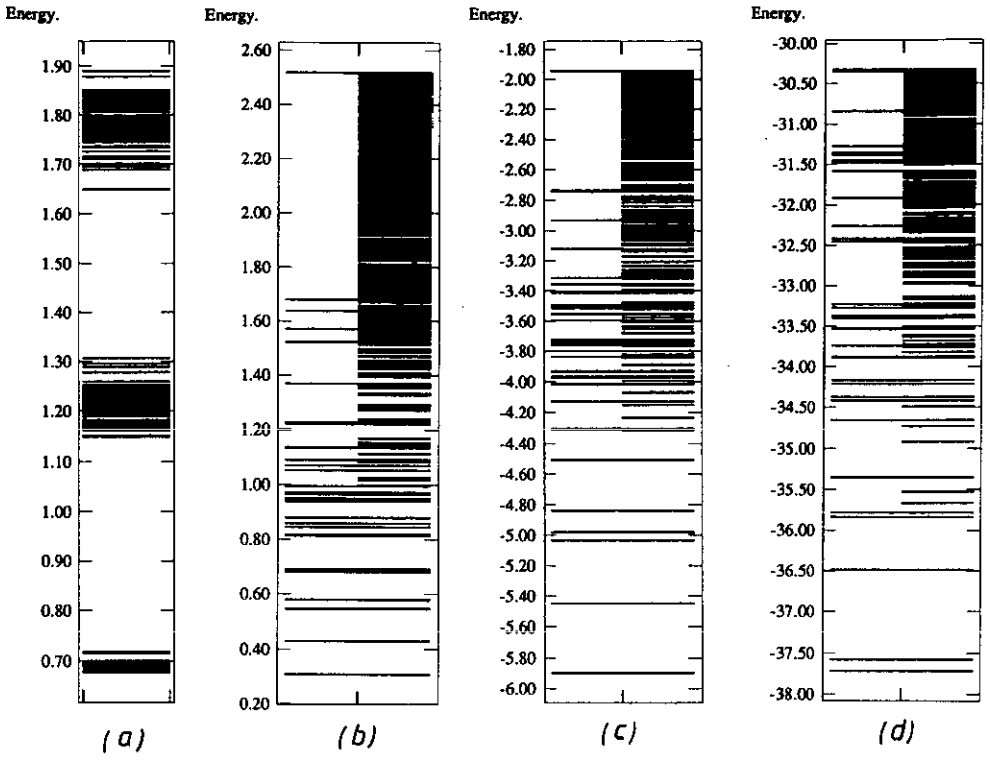

Figure 3. Energy spectrum at various values of ε, derived from simulated annealing and quenched high temperature Monte Carlo data. The simulated annealing data in each case represent all the configurations obtained in 250 trials. (*a*) $\varepsilon = 0$ spectrum from simulated annealing (232 states shown). (*b*) $\varepsilon = \sqrt{0.4}$ spectrum from simulated annealing (left, 32 states) and quenched Monte Carlo (right, lowest 584 states shown). (*c*) $\varepsilon = \sqrt{2}$ spectrum from simulated annealing (left, 32 states) and quenched Monte Carlo (right, lowest 233 states shown). (*d*) $\varepsilon = \sqrt{10}$ spectrum from simulated annealing (left, 36 states) and quenched Monte Carlo (right, lowest 260 states shown).

natural that the Monte Carlo finds additional states at high energy. At lower energies it is also reasonable since we expect the Monte Carlo to find states separated by only shallow barriers from lower energy states, whereas simulated annealing will avoid such states.

The most conspicuous feature of figure 3 is that an increase in ε leads to the appearance of a single deep global minimum which is separated from the other local minima by a large energy gap, whereas at higher energies there are a large number of local minima which characterize the glassy phase. This feature is what Anfinsen's experiment [5] leads us to expect for a native protein with a few hundred bases: the native state minimizes the free energy of the system, and is therefore determined by the totality of the interatomic interactions. Our simple model seems to realize the condition which the experiment requires.

To examine the lowest energy state we display in figure 2(*b*) this state for $\varepsilon = \sqrt{10}$. It is interesting to recognize that the lattice structure which we saw so clearly at $\varepsilon = 0$ persists, although now the configuration of the chain on the lattice is self-crossing. The lattice structure is also apparent in the low lying excited states. We expect that a similar structure plays a part in the three-dimensional simulations of IMP, although

defects are possibly more likely and it is rather hard to recognize in the two-dimensional projections they display.

There is a qualitative difference between the low energy states at small and large ε. The former consist exclusively of configurations with bonds of almost equal length connecting adjacent lattice sites, whereas for large ε typical states have some bonds between non-adjacent sites as well as bonds which cross each other. Nonetheless the transition from the $\varepsilon = 0$ system (known as the 'globule' state since the chain compactifies under the influence of the A/r^6 interaction) to that with a large ε where 'glassy' phases are detected takes place smoothly. There is no abrupt transition which might separate two distinct phases. This conclusion is reached by tracking the behaviour of individual states as ε is varied. Starting with a state at a given value of ε, we change ε by a small amount and then quench to return the state to the minimum of the updated potential. Iterating this procedure, generally we see a smooth change in the energy, the gyration radius and other such parameters. However, discontinuities do occur when a change in ε removes a potential barrier between two states allowing a state to fall into a lower well. These discontinuities are specific to individual states and there is no value of ε where such discontinuities accumulate.

The authors would like to thank N Go for discussions. DL and MGM acknowledge the Japan Society for the Promotion of Science for postdoctoral fellowships, and MGM thanks the Ministry of Education, Science and Culture for the Grant-in-Aid (#01795056).

References

[1] Iori G, Marinari E and Parisi G 1991 Random self-interacting chains: a mechanism for protein folding *Preprint ROM2F-91-4*

[2] Go N and Taketomi H 1979 *Proc. Natl. Acad. Sci. USA* **75** 559

[3] Kirkpatrick S, Gelatt C D Jr and Vecchi M P 1983 *Science* **220** 671

[4] Kawai H, Okamoto Y, Fukugita M, Nakazawa T and Kikuchi T 1991 *Chem. Lett.* **1991** 213

[5] Anfinsen C B 1973 *Science* **181** 223