



## On the disordered SAW model for DNA denaturation

Barbara Coluzzi, Edouard Yeramian

### ► To cite this version:

Barbara Coluzzi, Edouard Yeramian. On the disordered SAW model for DNA denaturation. Philosophical Magazine, 2007, 87 (3-5), pp.517-524. <10.1080/14786430600865372>. <hal-00513733>

**HAL Id: hal-00513733**

**<https://hal.science/hal-00513733v1>**

Submitted on 1 Sep 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



## On the disordered SAW model for DNA denaturation

Journal:	<i>Philosophical Magazine &amp; Philosophical Magazine Letters</i>
Manuscript ID:	TPHM-06-May-0177
Journal Selection:	Philosophical Magazine
Date Submitted by the Author:	17-May-2006
Complete List of Authors:	Coluzzi, Barbara; Ecole Normale Supérieure, Plateforme Environnement Yeramian, Edouard; Institut Pasteur, Unité de Bio-Informatique Structurale
Keywords:	biomolecules, disordered systems, phase transitions, numerical methods
Keywords (user supplied):	DNA
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
proc_CY.tex	



On the disordered SAW model for DNA denaturation

Barbara Coluzzi\*

Plateforme Environnement (CERES), Ecole Normale Supérieure, 24 rue Lhomond, 75005 Paris (France)

Edouard Yeramian

Unité de Bio-Informatique Structurale, Institut Pasteur, 25-28 rue du Docteur Roux, 75015 Paris (France)

(Received 00 Month 200x; in final form 00 Month 200x)

We investigate numerically the transition properties for models of DNA denaturation, which can be relevant for certain classes of disordered systems. The investigation is following two, complimentary, numerical approaches: on-lattice Monte Carlo like simulations or off-lattice statistical mechanics calculations, which can extend very significantly the affordable lengths for the sequences. The on-lattice model consists of two interacting self-avoiding walks with the same origin on a three-dimensional cubic lattice. We introduce two different contact energies, for the adenine-thymine coupling and the guanine-cytosine one respectively, distributed according to a bimodal law. Whereas the transition is recognized to be of first order in the pure (homopolymer) case, the behaviour of quantities averaged over disorder suggests that the random system undergoes a second order transition.

1 Introduction

The models of DNA denaturation have been extensively studied since the discovery of the double-helix [1]. In particular, the original interest in the helix-coil model was for its biological relevance, accounting for the sequence-specific separation of the two strands. More recently, the model has become also a subject of interest to physicists, because of its relevance to problems of self-avoidance effects and disorder in “almost 1-dimensional” systems. Here we consider disordered DNA denaturation with two different (complimentary) numerical approaches, trying to bring clarifications for still debated questions concerning the transition order.

We first summarize briefly results obtained with a pure model, inspired by the classical helix-coil model (introduced by Poland and Scheraga [2]) in which self-avoidance is completely taken into account. In this pure case the transition appears to be of first order [3–5]. We consider then the effect of disorder on this transition. To do so, we compare the results of on-lattice self-avoiding walk (SAW) numerical simulations [6] with those obtained with a disordered version of the Poland-Scheraga (PS) model with loop length probability exponent  $c_p = 2.15$  (which takes into account self-avoidance).

In a nutshell, our results suggest that disorder is relevant to the transition, in agreement with recent theoretical findings [7]. It is nevertheless interesting to notice that in a study involving a slightly different disordered version of PS model with the same  $c_p$  [8], it was found evidence for a still first order transition (characterised by two different correlation length exponents, for “typical” and “averaged” quantities respectively). In this general background, we finally propose a qualitative explication (coherent with some preliminary results [9]) of the different observed behaviours.

2 The pure model

The first models for DNA denaturation were related to the 1-dimensional Ising model, with two accessible states for base pairs (open or closed) and only nearest-neighbour interactions. Long range interactions are taken into account in the PS model [2], with the introduction of length-dependent entropic weights

\*Corresponding author. Email: coluzzi@lmd.ens.fr

In three dimensions, for a random walk loop, we have the value  $c_p = 3/2$  and the transition is of second order. If we consider a self-avoiding loop (characterised by  $c_p \simeq 1.75$ ), the transition becomes sharper [11], but still of second order. Nevertheless, with on-lattice numerical studies of a model consisting of two interacting SAWs, it was found that the system undergoes a first order transition, with a linear increase of the maximum of the specific heat with the chain length [4]. It can be noticed that in this model self-avoidance is completely accounted for, since two monomers can occupy the same lattice point if and only if they occupy the same position in the two different chains.

Based on conformal theory results for polymer networks [12], it was shown theoretically [3] that self-avoidance of loops with the rest of the system can be accounted through the assignment of an appropriate value to the exponent  $c_p$ . In three dimensions the correct value for such assignment is approximately  $c_p \simeq 2.15$ , as confirmed by numerical simulations with the on-lattice model [5], which explains why the transition is of first order in the homogeneous case.

It can be stressed that the denaturation in the pure model with  $c_p > 2$  is a peculiar kind of first order transition, characterised by a diverging correlation length  $\xi \sim (T - T_c)^{-\nu_p}$  [3–5], with  $1/\nu_p = \phi_p = \min(1, c_p - 1)$ . Therefore, despite the divergence of the specific heat, it is not clear if well known criteria such as the Harris one [13] do apply in this case and if disorder is relevant. As a matter of fact, it is only recently that the effect of disorder on DNA denaturation transition with self-avoidance has been investigated [6–9].

### 3 The disordered SAW model

Base pairs heterogeneity is reasonably accounted for by considering two different contact energies, associated with adenine-thymine and guanine-cytosine pairs respectively. In [6], a disordered version of the SAW model for DNA denaturation was introduced and studied numerically. The model consists of two interacting SAWs on a three dimensional cubic lattice, with the same origin. Two monomers can occupy the same lattice point if and only if they are in the same position  $i$  in the two different chains with the corresponding (pairing) energy  $\epsilon_i$ . The  $\{\epsilon_i\}$  are quenched random variables distributed accordingly to a bimodal law  $P(\epsilon) = [\delta(\epsilon - \epsilon_{AT}) + \delta(\epsilon - \epsilon_{GC})]/2$ , with the choice of values  $\epsilon_{GC} = 2$  and  $\epsilon_{AT} = 1$ , the temperature being in  $\epsilon_{AT}$  units.

Statistics were collected over  $\mathcal{N}_s = 128$  different disorder sequences and, for each sample, 16 different temperatures around the critical one were considered. For various observables, the behaviour as function of the temperature was obtained by reweighing data at the different values of the set.

In this study, despite implementations on a parallel computer, and the usage of the Pruned-Enriched Rosenbluth Method (PERM) [14], particularly efficient for thermalizing SAWs, it was only possible to consider chain lengths up to  $N = 800$ . It appeared that such lengths were not large enough for unambiguous clarification of the situation. In particular, the maximum of the averaged specific heat, as plotted in [Fig. 1a], seems to be clearly diverging with the chain length (though less rapidly than in the pure case). A fit of all the points to the law  $\overline{c(N, \epsilon)}^{max} \simeq aN^{2\phi_r-1} + b$  gives  $\phi_r = 1/\nu_r \sim 0.85$ , with however the data still compatible with a first order transition (i.e.  $\phi_r = 1/\nu_r = 1 = \phi_p$ ) when neglecting the smallest sizes. It can be pointed out that on theoretical grounds [15] one would expect a non-diverging specific heat when disorder is relevant, that makes the observed behaviour even more intriguing.

A smoother transition is further suggested by the results concerning the averaged probability of the loop length  $\overline{P(l)}$  at  $T_c$ , with however the possibility of a still first order one not being ruled out. The data in [Fig. 1b] would suggest that  $c_r = c_p > 2$  taking into account only the smallest chain lengths but it is also

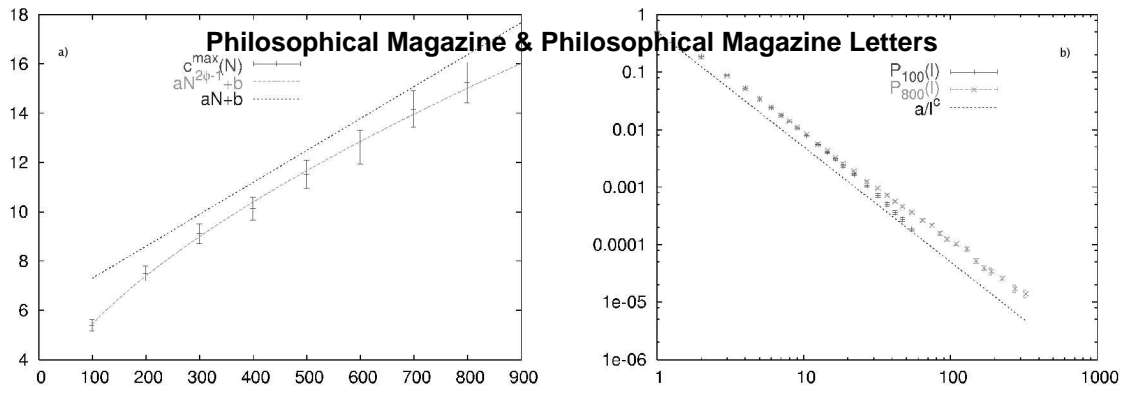


Figure 1. Disordered SAW model on lattice. a) Maximum of the averaged specific heat as function of the chain length. b) Averaged probability of the loop length at  $T_c$ .

clear upon fitting that the asymptotic value is  $c_r < 2$  when the largest sizes are considered.

As a final remark, preliminary results concerning the disordered SAW model with different values for the ratio  $\mathcal{R} = \epsilon_{GC}/\epsilon_{AT}$  suggest that the critical exponents could depend on  $\mathcal{R}$  values. In what follows we discuss in more detail such dependencies, which could be also explained by strong finite size effects, related to the details of the considered model.

#### 4 The disordered PS model

The Poland-Scheraga model for DNA denaturation has been implemented in different forms and studied in different contexts [1,2,8,16–21]. In particular, it has been successfully compared to experimental denaturation curves [18]. Recently, on biological grounds, the segmentation (helix/coil) obtained with the model was compared with the genetic segmentation (coding/non coding) demonstrating sharp correspondences for certain genomes [19,20]. Moreover, the effect of base pair mismatches on the transition was also investigated [21]. In such versions, a more detailed pairing scheme for the base pairs is adopted, taking into account the stacking energies and the cooperativity factor for loop initiations.

In the present work, we are rather interested in the thermodynamic limit behaviour of the model in the presence of disorder, and therefore we adopt as few parameters as possible. It seems indeed reasonable to suppose that the  $c_p$  exponent value should be the most relevant parameter for the determination of the universality class. It is also expected that the disordered PS model with  $c_p = 2.15$  and the disordered on-lattice SAW model, as previously considered, should belong to the same universality class. This hypothesis is nevertheless not confirmed as yet, and needs to be investigated in detail, all the more that recent results [8] suggest a different behaviour.

On algorithmic grounds, for the implementation of the PS model, we write a “forward” recursive equation in terms of canonical partition functions (following [16], instead of recursions concerning specific conditional probabilities as in [17]):  $Z_f(i+1, \epsilon) = \exp(\beta\epsilon_{i+1} - \log \mu)[Z_f(i, \epsilon) + \sum_l Z_f(i-l, \epsilon)/l^{c_p}]$ , with  $Z_f(i, \epsilon)$  the partial partition function associated with configurations up to position  $i$  and  $i$  in the closed state. An analogous recursion can be written for the “backward” partition function  $Z_b$ . With appropriate boundary conditions, and for strict correspondence with the on-lattice conditions, we restrict the calculations to configurations with base pair at position 1 in the closed state (with the allowance for non-penalized free-ends at the other extremity). The total canonical partition function is simply  $Z(N, \epsilon) = Z_b(1, \epsilon)$  and the probability  $P_N(i, \epsilon)$  for a base pair in position  $i$  in a chain of length  $N$  with disorder sequence  $\{\epsilon_i\}$  to be closed is given by:  $P_N(i, \epsilon) = Z_f(i, \epsilon)Z_b(i, \epsilon)/[Z(N, \epsilon) \exp(\beta\epsilon_i - \log \mu)]$ . For a given disordered sequence we can evaluate from the set of  $P_N(i, \epsilon)$  values at different temperatures, the various quantities of interest, such as the energy density, the specific heat, the fraction of closed base pairs  $\theta(T, N, \epsilon) = \langle n \rangle / N$  (which is the order parameter) and the susceptibility  $\chi(T, N, \epsilon)$ .

In the forward-backward recursive equations which define the model, the only thermodynamic parameters are the possible values of the coupling energies  $\{\epsilon_i\}$  and  $\log \mu$ . In order to ensure as much as possible strict

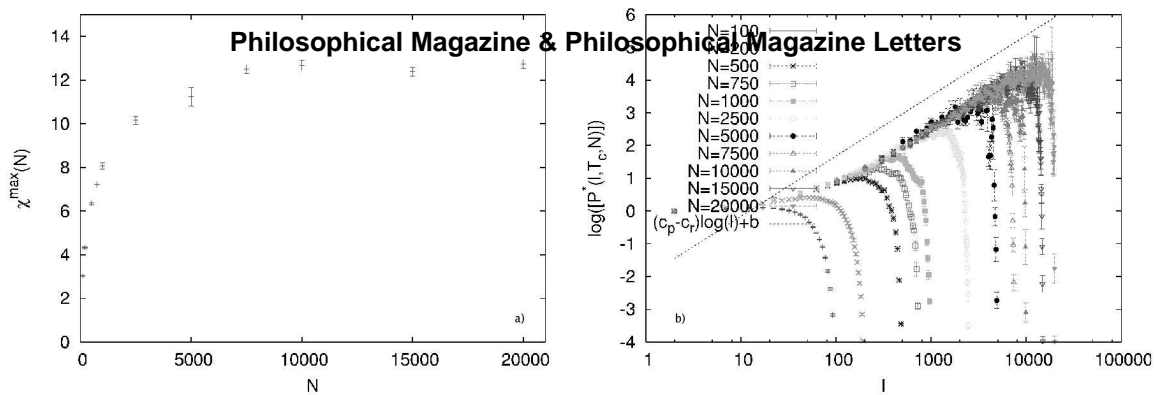


Figure 2. Disordered PS model with  $c_p = 2.15$ . a) Maximum of the averaged susceptibility as function of the chain length. b) Logarithm of the averaged  $P^*(l) = l^{c_p} P(l)$  at  $T_c$ .

correspondence with the disordered on-lattice SAW model, we adopted also in the PS implementation energies distributed according to the bimodal law with  $\epsilon_{GC} = 2$  and  $\epsilon_{AT} = 1$  (the temperature in  $\epsilon_{AT}$  units) and choose the value  $\log \mu \simeq 1.55$ , from the connectivity constant of SAWs on a three dimensional cubic lattice. As discussed in more detail in the following section, the finite size behaviour of the model could be strongly influenced by these particular values.

For the efficient calculation of partition functions following the recursions above, with algorithmic complexity reduced by one order of magnitude, we adopted a SIMEX scheme [16,17] (SIMulation with EXponentials, after an idea first expressed in [17] for conditional probabilities), which has been recently applied to the study of base pair mismatches [21] and to a slightly different version of disordered PS model [8]. The method relies on the numerical representation of the long-range effect as  $1/l^{c_p} \simeq \sum_{k=1}^{k_{max}} a_k \exp(-l b_k)$  and such approximations (for example up to  $k_{max} = 14$  in [16,19], for  $c_p = 1.95$ ) can be obtained through the Padé-Laplace method [22]. Here, for  $c_p = 2.15$ , and in order to make results comparable, we used the values of the parameters in [8] (up to  $k_{max} = 15$ ). For sufficient accuracy in the obtainment of the different observables, for each chain length at least  $N_s = 500$  disordered sequences were generated, and each sample was studied at 250 temperature values.

The data for the maximum of the averaged susceptibility as a function of  $N$  are shown in [Fig. 2a]. A striking similarity in the behaviour of this quantity with the one concerning the maximum of the specific heat is observed, suggesting that the two observables may be described by the same critical exponents and corrections to scaling. This figure also clearly demonstrates the existence of strong finite size corrections to the asymptotic behaviour, which would suggest divergence of  $\chi^{max}(N)$  for chain lengths  $N \leq 2500$ , as previously observed on the maximum of the specific heat for the disordered on-lattice SAW model (with  $N \leq 800$ ). Instead, the quantity is saturating, in agreement with the theoretically expected result  $\nu_r \geq 2/d$  [15].

The analysis of  $\overline{P(l)}$  data at  $T_c$  (plotted in [Fig. 2b]), leads to the same qualitative results. Also in this case, for chain lengths smaller than  $N \sim 2500$ , the asymptotic behaviour is not reached and in particular the smallest sizes would be in agreement with a still first order transition as in the pure case. On the contrary, when looking at the largest sizes it is clear that for the exponent of the random system  $c_r \ll c_p$ . In detail, the asymptotic behaviour suggests  $c_r \sim 1.35$ , which should correspond to a correlation length exponent  $\nu_r = 1/(c_r - 1) \sim 3 > 2/d = 2$ , again in agreement with a non-diverging specific heat (and susceptibility).

## 5 Discussion and perspectives

The results obtained here for the disordered PS model considered, with  $c_p = 2.15$ , appear to confirm that disorder is relevant to this first order phase transition, as suggested from recent theoretical findings [7]. Moreover, this model and the disordered SAW model on-lattice should belong to the same universality class and they could also be characterised by the same kind of strong finite size corrections to the asymptotic



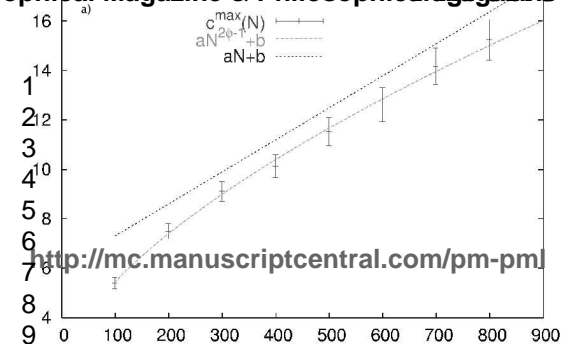
of the disordered PS model with the same  $c_p$  value, suggested a transition which is still of first order as in the pure case but characterised by two different correlation length exponents. These studies performed an analysis in terms of the pseudo-critical temperature  $T_c(N, \epsilon)$  [23], appropriately defined for a given disorder sequence and chain length. It was found that the averaged value and the fluctuations of this quantity appear to behave as  $\overline{T_c(N, \epsilon)} \simeq T_c^\infty + \text{const}/N^{\nu_{r,1}}$  and  $\delta T_c(N, \epsilon) \simeq \text{const}/N^{\nu_{r,2}}$  respectively, with the different exponents  $\nu_{r,1} = \nu_p = 1$  and  $\nu_{r,2} = 2/d = 2$ , as one expects in the case of irrelevant disorder. Correspondingly, the behaviour of the typical sample should be governed by the correlation length  $\xi_1 \sim (T - T_c)^{-\nu_{r,1}}$ , whereas when considering averaged quantities the behaviour is dominated by the fluctuations and one should find  $\xi_2 \sim (T - T_c)^{-\nu_{r,2}}$ .

At present, we are performing the same kind of analysis in terms of  $T_c(N, \epsilon)$  for the disordered PS model discussed here and preliminary results [9] show that in our case the same exponent  $\nu_{r,1} = \nu_{r,2} \sim 3$  could apply for  $\overline{T_c(N, \epsilon)}$  and  $\delta T_c(N, \epsilon)$  data, also compatible with the value obtained independently from the behaviour of  $\overline{P(I)}$  at  $T_c$ , therefore confirming that disorder is relevant. Interestingly, a similar value was found in [8] for the slightly different disordered PS model in the study, but for the case  $c_p = 1.75$ .

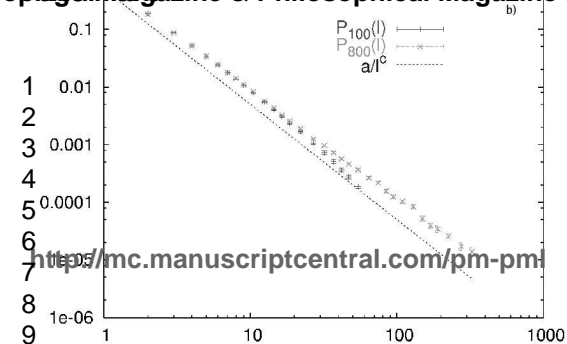
A possible explanation for the different observations is that the various disordered PS models are in the same universality class, which could not depend on the  $c_p$  value as soon as  $c_p > 3/2$  (i.e., the specific heat in the pure case is diverging and disorder should be relevant following the Harris criterion [13]) but there are strong finite size effects depending on the particular model. More in detail, the effect of disorder could be related to the possible presence of “rare regions” in the typical sample, and correspondingly it could become evident only at chain lengths larger than some value  $N^*$  that one could expect [9] to increase exponentially with a parameter  $x$  depending on model details. If this is the case, despite the very large sizes studied in [8], it is possible that the asymptotic regime was still not reached because of finite size effects amplified in a very significant way. On the contrary, in our case the asymptotic regime appears to be reached even with sequence lengths in the range  $N \sim 5000$ . In any event, more careful studies are necessary in order to confirm our findings and hypothesis. Moreover, it will be important and interesting to investigate in detail the differences in the models for an overall clarification of the situation.

## References

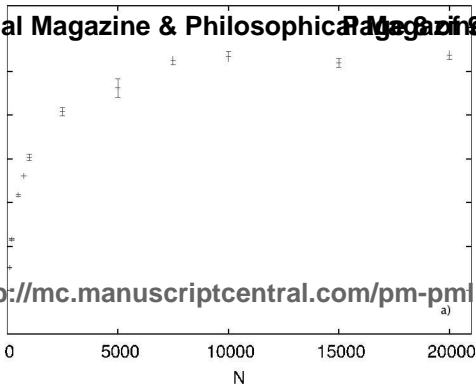
- [1] D. Poland and H.A. Scheraga (eds.), *Theory of Helix-Coil Transitions in Biopolymers*, (Academic, New York, 1970); R.M. Wartell and A.S. Benight, *Phys. Rep.* **126**, 67 (1985); for a recent review in which the effects of self-avoidance in Poland-Scheraga models are critically discussed see C. Richard and A.J. Guttman, *J. Stat. Phys.* **115**, 943 (2004).
- [2] D. Poland and H.A. Scheraga, *J. Chem. Phys.* **45**, 1456, 1464 (1966).
- [3] Y. Kafri, D. Mukamel and L. Peliti, *Phys. Rev. Lett.* **85**, 4988 (2000); *Eur. Phys. J. B* **27**, 135 (2002).
- [4] M.S. Causo, B. Coluzzi and P. Grassberger, *Phys. Rev. E* **62**, 3958 (2000).
- [5] E. Carlon, E. Orlandini and A.L. Stella, *Phys. Rev. Lett.* **88**, 198101 (2002); M. Baiesi *et al*, *Phys. Rev. E*, **67**, 021911 (2003).
- [6] B. Coluzzi, *Phys. Rev. E*, **73**, 011911 (2006).
- [7] G. Giacomin and F.L. Toninelli, *Smoothing effect of quenched disorder on polymer depinning transition*, math.PR/0506431, to appear on *Commun. Math. Phys*; *Phys. Rev. Lett.* **96**, 060702 (2006); F.L. Toninelli, *Critical properties and finite-size estimates for the depinning transition of directed random polymers* cond-mat/0604453.
- [8] T. Garel and C. Monthus, *J. Stat. Mech.*, P06004 (2005); C. Monthus and T. Garel, *Eur. Phys. J. B* **48**, 393 (2005); C. Monthus, *Random walks and polymers in the presence of quenched disorder*, cond-mat/0601332.
- [9] B. Coluzzi and E. Yeramian, *On the relevance of disorder to the Poland-Scheraga model for DNA denaturation transition*, work in progress.
- [10] M.E. Fisher, *J. Stat. Phys.* **34**, 667 (1984).
- [11] M.E. Fisher, *J. Chem Phys.* **45**, 1469 (1966).
- [12] B. Duplantier, *Phys. Rev. Lett.* **57**, 941 (1986); *J. Stat. Phys.* **54**, 581 (1989).
- [13] A.B. Harris, *J. Phys. C* **7**, 1671 (1974).
- [14] P. Grassberger, *Phys. Rev. E* **56**, 3682 (1997).
- [15] J.T. Chayes, L. Chayes, D.S. Fisher and T. Spencer, *Phys. Rev. Lett.* **57**, 2999 (1986).
- [16] E. Yeramian, F. Schaeffer, B. Caudron, P. Claverie and H. Buc, *Biopolymers* **30**, 481 (1990); E. Yeramian, *Europhys. Lett.* **25**, 49 (1994).
- [17] M. Fixman and J.J. Freire, *Biopolymers* **16**, 2693 (1977).
- [18] R.D. Blake *et al*, *Bioinformatics*, **15**, 370 (1999).
- [19] E. Yeramian and L. Jones, *Nucl. Acid Res.* **31**, 3843 (2003); E. Yeramian, *Gene* **255**, 139 (2000); *Gene* **255**, 151 (2000).
- [20] E. Yeramian, S. Bonnefoy and G. Langsley, *Bioinformatics* **18**, 190 (2002).
- [21] T. Garel and H. Orland, *Biopolymers* **75**, 453 (2004).
- [22] E. Yeramian and P. Claverie, *Nature* **326**, 169 (1987).
- [23] A. Aharony, A.B. Harris and S. Wiseman, *Phys. Rev. Lett* **81**, 252 (1998).







1  
2  
3  
4  
5  
6  
7  
8  
9



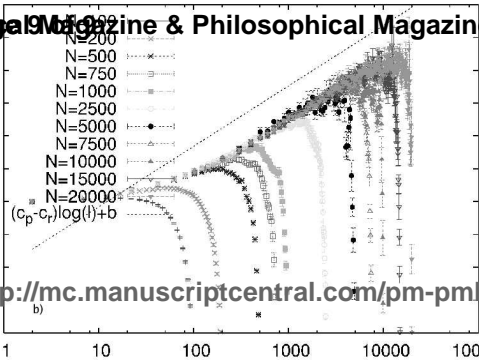
<http://mc.manuscriptcentral.com/pm-pml>

$\log(IP^*(l, T_0 N))$

- N=200
- N=500
- N=750
- N=1000
- N=2500
- N=5000
- N=7500
- N=10000
- N=15000
- N=20000

$$(c_p - c_r) \log(l) + b$$

b)



1 10 100 1000 10000 100000

l