

## Initial events of protein folding from an information-processing viewpoint

Henrik G. Bohr and Peter G. Wolynes

*Noyes Laboratory and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801*

(Received 1 June 1992)

The initial events of protein folding can be described by the time evolution of a neural network through the use of polymer-kinetics theory, where the fundamental variables are residue contacts. The contact energies are analogous to chemical activation potentials. The resulting recurrent neural network behaves like a partially frustrated spin system interacting with a random external field. A dynamical phase diagram is derived from numerical studies of the model, and the basins of attraction for the network are studied.

PACS number(s): 87.15.Rn, 87.15.Da

### I. INTRODUCTION

Amongst their many applications, neural networks have recently been used to predict secondary structure [1–3] and tertiary structure of proteins [4–7]. At the same time, the physical process of protein folding has been studied [8–11] using ideas from the theory of spin glasses. Many properties of neural networks can be analyzed using spin-glass theory. A family of energy functions for folding proteins called associative memory Hamiltonians has been introduced [4] with motivation from the spin-glass theory of neural nets. Most neural networks applied to protein structure, however, have not been constructed on the basis of some physical analogy.

In this paper, we will discuss a way in which neural-network equations for protein-structure prediction can be derived from the physics of protein folding while focusing on information-processing considerations. How can we imagine that the protein as a computer can perform the immense calculation leading from the initial disordered state of a chain of amino acids to an “optimized” native folded state while avoiding the trap of multiple minima? It is natural to expect the possibility that a protein carries instructions for a correct set of paths through intermediate steps, thus breaking the problem up. This will limit the number of possible minima and lead to a much simpler computational task. The question is, can most of the Levinthal problem [12] be overcome in the first stages of the folding?

We describe the early stages of protein folding using contact variables giving the proximity relations between different residues [13]. The kinetic equations for these contact variables have the same structure as feedback neural networks. The feed-forward networks that have been used represent a short-time approximation to these dynamics. The relationship between feedback neural nets and the early events in folding can be exploited in two ways. First, it can be used to design architectures for protein-structure-prediction neural networks that have some basis in physical processes. Second, one can use the relatively well-developed theory of the basins of attraction of neural networks to examine, in a crude way, some of the general features of the early stages of protein folding. This latter development complements nicely the

thermodynamic spin-glass theory, which concentrates on the long-time-scale processes.

The organization of this paper is as follows. In the next section we describe the free-energy functionals of proteins written in terms of the contact variables and show that the corresponding kinetic equations lead to a neural-network formulation. In the third section, we examine how these feed-forward neural-network equations describing the early events in folding behave in relation to the thermodynamic properties of the model. In the Conclusion, we speculate on how these ideas may be applied for real proteins.

### II. KINETIC EQUATIONS FOR CONTACT VARIABLES

There are many ways to represent the tertiary structure of a protein. The most familiar specification of this structure consists of the coordinates  $R_i$ , of each of the atoms or of the backbone atoms in a protein. Alternatively, we could give the pair distances between each of the pairs of the atoms, but this is an overdetermined set. The success of NMR structure-determination methods show that, in fact, only a small fraction of the pair distances need to be known with some degree of accuracy in order to reproduce the structure of a protein. We take in this paper, then, the basic variables to be density variables  $\sigma_{ij}^{\rho}$ , which measure the correlation between the  $i$ th and  $j$ th residue on the protein backbone. We define  $\sigma_{ij}^{\rho}$  as

$$\sigma_{ij}^{\rho} = \begin{cases} 1 & \text{if } j \text{ is within a shell of } (\rho, \rho + d\rho) \text{ from } i \\ 0 & \text{if } j \text{ is not .} \end{cases} \quad (1)$$

The description of a protein in terms of a distance matrix involves an integrated version of this variable,

$$s_{ij}^{\rho} = \int_0^{\rho} \sigma_{ij}^{\rho'} d\rho', \quad (2)$$

which becomes equal to 1 if the  $i$ th residue is close (within a distance  $\rho$ ) to the  $j$ th residue and 0 otherwise. These variables were predicted from sequence in the Copenhagen feed-forward neural networks [5]. The same variables also play a role in the approach to protein structure based on associative-memory functions [4]. The

associative-memory Hamiltonian for a protein is written in terms of the backbone coordinates as

$$H_{\text{am}} = \Lambda_{\text{am}} \sum_{\mu} \sum_{\substack{i,j \\ (i < j)}} \gamma_{ij}^{\mu}(q_i^{\mu}, q_i^{\mathcal{T}}, q_j^{\mu}, q_j^{\mathcal{T}}) \theta(r_{ij} - r_{ij}^{\mu}) + H_0$$

$$= \Lambda_{\text{am}} \sum_{\mu} \sum_{\substack{i,j \\ i < j}} \gamma^{\mu} \sigma_{ij}^{\mu} \sigma_{ij} + H_0, \quad (3)$$

where  $\Lambda_{\text{am}}$  is a scaling constant and  $\gamma_{ij}^{\mu}(q_i^{\mu}, q_i^{\mathcal{T}}, q_j^{\mu}, q_j^{\mathcal{T}})$  is a comparison function which was originally chosen to be related to the hydrophobic charge,  $\mu$  is an index for the memories, and  $\mathcal{T}$  stands for the target.  $\theta$  is a pairwise distance overlap function and  $H_0$  is the part of the energy that includes the constraints of backbone geometry and chain connectivity. If one notices the resemblance of the  $\sigma$  variables to spin variables, we see that this interaction energy resembles that of a Hopfield spin system [14] (but the spins already refer to pair variables and thus are fields rather than interaction terms).

At first sight it is difficult to write the free energy in terms of these contact variables because of the chain connectivity part of the Hamiltonian  $H_0$ . One can, however, introduce fields proportional to the  $\sigma_{ij}(\rho)$  and calculate the partition function of a chain with a prescribed set of fields. The usual procedures of liquid-state density-functional theory [15,16] then give us the general form of the free energy coming from the chain connectivity constraints. It has the form of

$$F^0(\sigma_{ij}) = F_{\text{PG}}(\sigma_{ij}) + \Delta F^0(\sigma_{ij}), \quad (4)$$

where  $F_{\text{PG}}(\sigma_{ij})$  is a Boltzmann "perfect-gas" entropy

$$F^0 = k_B T \left[ \sum_l \sigma_{l,l+4}^{\rho(\alpha)} \ln \sigma_{l,l+4}^{\rho(\alpha)} + \sum_{i,j} \sigma_{ij}^{\rho(\beta)} \ln \sigma_{ij}^{\rho(\beta)} + \left[ 1 - \sum_j \sigma_{ij}^{\rho(\beta)} - \sigma_{ij}^{\rho(\alpha)} \right] \ln \left[ 1 - \sum_j \sigma_{ij}^{\rho(\beta)} - \sigma_{ij}^{\rho(\alpha)} \right] \right]$$

$$\times \sum_{i,j} W_{ij}^{0\alpha} \sigma_{ij}^{\alpha} + W_{ij}^{0\beta} \sigma_{ij}^{\beta} + \sum_{i,j,k,l} W_{ijkl}^{0\alpha\beta} \sigma_{ij}^{\alpha} \sigma_{kl}^{\beta}. \quad (9)$$

While  $F^0$  is a property of polymer chains in general, the interactions must depend on the character of the residues. Thus we can write a functional Taylor-series expansion:

$$F^I = \sum_{i,j} W_{ij}^I \sigma_{ij} + \sum_{i,j} \sum_{k,l} W_{ijkl}^I \sigma_{ij} \sigma_{kl}. \quad (10)$$

The problem is to determine the expansion coefficients  $\{W_{ij}^I\}$ . The associative memory Hamiltonian ansatz gives an explicit form for  $W_{ij}^I$  and thus essentially instructs the values of the interactions. But the minimization of the total free energy gives a representation leading to a neural network and one can alternatively learn the interaction matrix by methods such as back propagation [17]. If we stop then with the linear associative memory Hamiltonian, the free-energy function reads

term:

$$F_{\text{PG}} = k_B T \sum_{i,j} \sigma_{ij} \ln \sigma_{ij} d\rho. \quad (5)$$

The higher-order terms may be expanded in orders of the contact densities

$$\Delta F^0 = \sum_{i,j} W_{ij}^0 \sigma_{ij} + \sum_{i,j} \sum_{k,l} W_{ijkl}^0 \sigma_{ij} \sigma_{kl} + \dots \quad (6)$$

The first term here measures the free-energy gain for forming the contact  $(i,j)$ , and the second term measures the free energy from forming a contact  $(i,j)$  when contact  $(l,k)$  has already been formed. One can also use the integrated contact variables  $s_{ij}$  to obtain a free-energy function

$$s_{ij}^{\rho} = \int_0^{\rho} \sigma_{ij}^{\rho} d\rho, \quad (7)$$

$$F_{\text{PG}} = k_B T \left[ \sum_{i,j} s_{ij} \ln s_{ij} + (1 - s_{ij}) \ln(1 - s_{ij}) \right],$$

$$\Delta F^0 = \sum_{i,j} W_{ij}^0 s_{ij} + \sum_{i,j} \sum_{k,l} W_{ijkl}^0 s_{ij} s_{kl}, \quad (8)$$

where, of course, the Taylor coefficients in the functional expansion will be different. These coefficients,  $W^0$ , are properties of connected polymers and do not depend explicitly on the sequence. We also point out that certain contacts are formed through hydrogen bonding, for which there are stoichiometric constraints. Thus if we choose densities corresponding to the formation of an  $\alpha$  helix,  $\sigma_{ij}^{\alpha}$  or  $\beta$  sheet,  $\sigma_{ij}^{\beta}$ , then the perfect-gas entropy should be rewritten in terms of the number of bonded contacts giving a free energy of the form

$$F = \sum_{i,j} \int d\rho k_B T [(1 - s_{ij}) \ln(1 - s_{ij}) + s_{ij} \ln s_{ij}]$$

$$- \sum_{i,j} W_{ij}^0 s_{ij} - \sum_{i,j} \sum_{k,l} W_{ijkl}^0 s_{ij} s_{kl} - \sum_{i,j,\alpha} \gamma_{ij}^{\alpha} s_{ij} s_{ij}^{\alpha}. \quad (11)$$

This equation agrees with the mean-field free energy of a magnet. The sequence information gives a field on each site and the cooperativity comes from the chain constraints in the  $W_{ij}^0$ . If contacts  $(ij)$  and  $(kl)$  are mutually favorable, this interaction is ferromagnetic. But, if, on the other hand, they are inconsistent, this would be an antiferromagnetic interaction. Thus, the equivalent magnet has a deterministic amount of frustration in it. A natural equation of motion for the contacts arises by considering gradient descent with this free-energy functional:

$$\frac{ds_{ij}}{dt} = \frac{\partial F}{\partial s_{ij}}. \quad (12)$$

We can write our contact dynamics in terms of a set of chemical kinetic equations. In that case, we consider the transition from an occupied state, with probability  $P_o$  to a nonoccupied state, with probability  $P_{no}$  and define a variable, analogous to the contact variable, as  $\sigma_{ij} = P_o - P_{no}$  and with stoichiometry  $P_o + P_{no} = 1$ .

We can write the rate of change of the probabilities in terms of the corresponding rate constants  $k_{no \rightarrow o}$  and  $k_{o \rightarrow no}$ :

$$\frac{dP_o}{dt} = k_{no \rightarrow o} P_{no} - k_{o \rightarrow no} P_o \quad (13)$$

and

$$\begin{aligned} \frac{d\sigma}{dt} &= \frac{dP_o}{dt} - \frac{dP_{no}}{dt} \\ &= (K_{no \rightarrow o} - k_{o \rightarrow no} + \sigma(-k_{no \rightarrow o} - k_{o \rightarrow no})). \end{aligned} \quad (14)$$

The energy cost to change occupation can be obtained from the free-energy function in (11). This gives the activation energy functional  $H_A^{ij}$ :

$$H_A^{ij} = W_{ij}^0 + \sum_{k,l,\rho'} W_{ijkl}^0 \sigma_{kl}' + W_{ij}^I. \quad (15)$$

Glauber kinetics then gives

$$\begin{aligned} \frac{d\sigma_{ij}}{dt} &= \text{sgn}(H_A^{ij}) (-1 + e^{-|H_A^{ij}|/k_B T}) \\ &\quad - (1 + e^{|H_A^{ij}|/k_B T}) \sigma_{ij}, \end{aligned} \quad (16)$$

or with finite time steps

$$\begin{aligned} \sigma_{ij}(t+1) - \sigma_{ij}(t) &= \text{sgn}(H_A^{ij}) (-1 + e^{-|H_A^{ij}|/k_B T}) \\ &\quad - (1 + e^{|H_A^{ij}|/k_B T}) \sigma_{ij}(t), \end{aligned} \quad (17)$$

where we have changed the differential equation into a difference equation. In the low-temperature limit we then obtain

$$\begin{aligned} \sigma_{ij}(t+1) &= \text{sgn} \left[ (k_B T)^{-1} \left[ \sum_{k,l,\rho'} W_{ijkl}^0 \sigma_{kl}' \right. \right. \\ &\quad \left. \left. + W_{ij}^0 + \dots \right] \right] f(H_A^{ij}), \end{aligned} \quad (18)$$

where

$$f(H_A) = (e^{-|H_A|/k_B T} - 1) \rightarrow -1 \quad \text{for } T \rightarrow 0.$$

The equation describes the evolution of a Boolean neural network similar to the equation of steepest descent for contact dynamics, as we shall see in the following.

We shall now proceed with the time-dependent equation for the fast downhill folding process of the protein contacts:

$$\frac{ds_{ij}}{dt} = - \frac{\partial F}{\partial s_{ij}}, \quad (19)$$

We can solve the dynamical equation by an iterative procedure where we first consider the static equation

$$\frac{\partial F}{\partial s_{ij}'} = 0. \quad (20)$$

If we insert the expression for  $F$  from before, redefine the variable  $s_{ij} \rightarrow (s_{ij} + 1)/2$ , absorb the constant term  $W_{ij}^I$  in  $W_{ij}^0$  and define  $\omega = \sum_{k,l,\alpha} q_k^\alpha q_l^\alpha s_{kl}^\alpha$  we have

$$s_{ij} = \tanh \left[ (k_B T)^{-1} \left[ W_{ij}^0 + q_i q_j \omega \frac{1}{2} \sum_k W_{ijkl}^0 s_k \right] \right], \quad (21)$$

which can be solved iteratively:

$$\begin{aligned} s_{ij,0} &= \tanh[(k_B T)^{-1} (W_{ij}^0 + q_i q_j \omega)], \\ s_{ij,1} &= \tanh \left[ (k_B T)^{-1} \left[ \bar{W}_{ij}^0 + q_i q_j \omega \right. \right. \\ &\quad \left. \left. + \frac{1}{2} \sum_{k,l,\rho'} W_{ijkl}^0 s_{kl}' \right] \right]. \end{aligned} \quad (22)$$

With the expansion of the lower-order contact variable  $s_{ij,0}$  which depends on the sidechain properties  $q$ :  $s_{ij,0} = A^0 + B^1 q_i q_j + \dots$ , it becomes an ordinary multineuron feed-forward neural network.

This network is similar to a neural network [5] constructed for distance matrices from sequence information  $q_i$ . We see that the thresholds and weights are determined partly from polymer-chain dynamics. The sequence-dependent parts of the thresholds are undetermined and thus should constitute the essential part of the training of the neural network for predicting distance matrices and hence the three-dimensional structure.

If we insert the expression for  $F$  in the time-dependent equation (19), we have

$$\frac{ds_{ij}}{dt} = -k_B T \ln \left[ \frac{s_{ij}}{1-s_{ij}} \right] + W_{ij}^0 + W_{ijkl}^0 s_{kl} + W_{ij}^I + \dots, \quad (23)$$

and if we again consider the low-temperature limit, and transform the logarithm into a sign function and consider our dynamics close to equilibrium, we obtain the equation

$$s_{ij}(t+1) = \text{sgn} \left[ (k_B T)^{-1} \left[ \sum_{k,l} W_{ijkl}^0 s_{kl}(t) + W_{ij}^0 + W_{ij}^I \right] \right], \quad (24)$$

which is a Boolean net, as in the case of the chemical activation energies.

Such a Boolean net arising from this special low-temperature limit is unlikely to be valid for real proteins, since it implies a rigid code for the protein folding. In the general case of all temperature regimes, we can instead transform the logarithms into a tanh function and obtain

$$s_{ij}(t+1) = \tanh \left[ (k_B T)^{-1} \left[ \sum_{k,l} W_{ijkl}^0 s_{kl}(t) + W_{ij}^0 + W_{ij}^I \right] \right], \quad (25)$$

representing a feed-back neural network. This neural network processes an initial structure of zero contacts into an orderly structure with patches resembling or being a combination of the patterns in the instructed set. It is well known that this equation will have fixed-point solutions if the synaptic weights are symmetric and the nonlinear function  $\tanh$  (hyperbolic tangent) is never decreasing, as in the case here.

The fixed points determine the patterns in the early structures of the folding process. What are those patterns? They are partly determined from polymer-chain connectivity through the weight factors  $W_{ij}^0$ ,  $W_{ijkl}^0$ , which have been worked out in the literature; e.g., [18]. They are basically the logarithm of the reduction factor, i.e., the ratio of the number of conformation of a chain with the given contacts over the number of all conformations. Already, here, in the case of contact interactions, we recover the known types of secondary structures of helices and antiparallel sheets as the only topological possibilities of constructing the two-loop factor  $W_{ijkl}^0$  for periodic structures as fixed-point solutions (helix if  $i < k < j < l$ , antiparallel sheets if  $i < k < l < j$ , and coil if  $i < j < k < l$ ).

To summarize, we have shown that the dynamics for residue contacts in the early events of protein folding can be formulated as evolution of neural networks. In the next section we shall consider simple numerical solutions to the evolution equation that can give us some insight about the phase diagram for formation and competition between simple secondary structures.

### III. NUMERICAL STUDIES OF COMPETING STRUCTURES

We first performed a numerical investigation of the evolution equation for protein contacts when we only consider two basic memory structures (simplified versions of an  $\alpha$  helix and an antiparallel  $\beta$  sheet) competing. We include these memories in the instructed term  $W_{ij}^I$  of the evolution equation and give them weights 1 and  $\delta$  and then study the development as time evolves with varying temperature  $T$  and weight factor  $\delta$ . The terms  $W_{ijkl}^0$  and  $W_{ij}^0$  representing the synaptic weights and neuron thresholds in neuron network terminology are stemming from polymer physics and are here approximated by the following constants:

$$\frac{1}{k_B T} W_{ij}^0 = \gamma - \ln|i-j|^{d/2}, \quad \gamma = \left[ \frac{\Delta v}{D^3} \right], \quad D = \sqrt{2} l_p, \quad (26)$$

where  $d$  is the dimension of the configuration space,  $l_p$  is the persistence length, and  $\gamma$  is a parameter that can include hydrogen bonds. The synaptic weights  $W_{ijkl}^0$  are approximated by constants close to unity.

In the actual simulation, we consider an  $\alpha$ -helical

structure as the target competing with a sheet structure that has an energetic weight  $\delta$ . In such a case we were able to “fold” a linear structure according to the memories and thus study the various phases of protein structure formation. The memories are constructed as specific patterns of contacts that are realistic. For our  $\alpha$ -helix pattern, only contacts  $s_{ij}^\alpha$ , for which  $i = j + 4$ , are different from zero. Similarly, the other type of memory is constructed as an antiparallel  $\beta$ -sheet structure with  $S_{ij}^\beta \neq 0$  for  $n - i = j - 2$ ,  $n$  being the number of residues forming contact. We chose  $n$  to be around 10, thus having 100 contact possibilities.

The actual simulation is an iterative procedure for solving the evolution equation (25) or written with the memories:

$$S_{ij}(t+1) = \tanh \left[ (k_B T)^{-1} \left[ \sum_{k,l} W_{ijkl}^0 S_{kl}(t) + W_{ij}^0 + S_{ij}^\alpha + \delta S_{ij}^\beta \right] \right],$$

where we start with a pattern  $\{S_{ij}\}$  with zero contacts, insert it as the initial condition into the evolution equation at time  $t=0$ , and insert the term  $W_{ij}^I$  as a linear combination of the two memories with weights 1 and  $\delta$ , and then obtain another pattern at  $t=1$ . The weights are chosen according to Eq. (26). In the next step, the pattern at  $t=1$  is the initial condition to be inserted in the evolution equation to find a new pattern at  $t=2$ . This procedure is repeated until the patterns have converged and a definite structure has been obtained. The folding has occurred when the correct structure ( $\alpha$ -helix pattern) has been recalled. The temperature can be varied, and then we can determine the folding temperature  $T_F$ .

In Fig. 1 we have sketched an approximate dynamical phase diagram of these numerical studies that can be compared with the usual equilibrium phase diagram for the random energy model (Refs. [19,20]). The overall features are similar, with a “folded” phase where the helix structure is completely recalled, and a distinct phase as a mixture between the helix and sheet structures at higher temperatures. These phases are divided by a transition line that determines what we would call the folding temperature. The mixture phase that is similar to the phase termed “the molten-globule phase” in Ref. [20] is again separated from a completely unfolded phase with no structure at all (less than 1% of the correct helical structure) and at very high temperatures. When increasing the disorder parameter  $\delta$  up to  $\delta_{\text{critical}} = 1.1$ , the ability to recall the correct structure breaks down, and a phase with the recall of the sheet structure starts to be present. The choice of parameters used here and given in Eq. (26) favors slightly the helix formation when the thresholds are chosen as constants in  $i$  and with  $\gamma$  dominating in  $W_{ij}^0$ . Near the critical line the helix structure can be recalled up to 40% accuracy.

Finally, we shall discuss the basin of attraction for the fast downhill processes where there are many competing structures. Consider the case where the linear associative memory rule is used to develop the interaction Hamil-

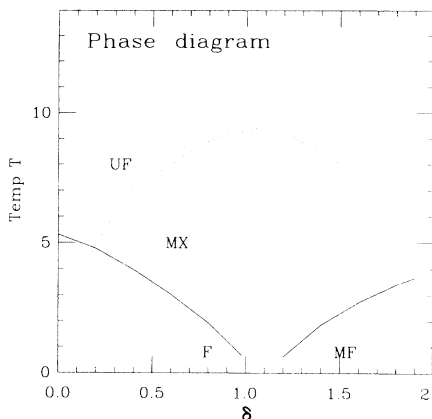


FIG. 1. A phase diagram in temperature and  $\delta$  ( $\delta$  is the disorder parameter, i.e., the weight on the undesired sheet structure) arising from a simulation of the network performance defined by the evolution equation (25) with two types of structures, a desirable helix structure and an undesirable sheet structure. The various phases found are the following: UF, unfolded phase; MX, mixed phase; F, folded phase; MF, misfolded phase. These phases are separated by lines and defined as follows: The folded state appears when the correct helix structure is recalled up to 1% accurately and an order of magnitude better than the wrong one, the sheet structure. The misfolded state is defined similarly but with the sheet structure instead of the helix structure. The mixed state is a distinct phase and defined when the good structure is recalled up to 75% correct and simultaneously the bad structure up to 25%. The unfolded phase covers the rest of the cases. The correct memory consisted of five  $\alpha$ -helical patterns, the incorrect memory of five antiparallel  $\beta$ -sheet patterns. The number of residues for each unit of pattern was 6 and the synaptic weights were fixed to a constant. Units for "Temp T" are in  $k_B$ .

tonian. In this case we take the influence of the incorrect memories to be equivalent to a Gaussian field. In the contact representation, we have seen that the Hamiltonian is the same as a regular partially frustrated spin glass with an external field that is basically ferromagnetic, i.e., favoring a single configuration and a random part coming from other memories:

$$H = - \sum_{i,j,k,l} W_{ijkl}^0 S_{ij} S_{kl} + \sum_{i,j} W_{ij}^0 S_{ij} + \sum_{\alpha,i,j} \gamma_{ij}^\alpha S_{ij}^\alpha S_{ij} , \quad (27)$$

where the second term  $\sum_{i,j} W_{ij}^0 S_{ij}$  represents an external field, and the third term  $\sum_{\alpha,k,l} \gamma_{ijkl}^\alpha S_{kl}^\alpha S_{ij}$  is the random part where there are many conflicting memories. The corresponding evolution equation for the spins in the fast downhill motion is

$$s_{ij}(t+1) = \text{sgn} \left[ (k_B T)^{-1} \left( \sum_{k,l} W_{ijkl}^0 S_{kl}(t) + W_{ij}^0 + \sum_{\alpha,k,l} \gamma_{ijkl}^\alpha S_{kl}^\alpha \right) \right] = \text{sgn}(h_{ij} + h_{ij}^0 + h_{ij}) . \quad (28)$$

This equation corresponds to a single spin reorienting the field, partly coming from the correct instructions, partly from the neighboring interacting spins, and a random part from the incorrect instructions. At this level, the evolution equation, therefore, looks like the evolution equation in a conventional feedback neural net that carries out association, and we can follow precisely the analysis given by Amit, Gutfreund, and Sompolinsky [21], Krauth, Nadel, and Mezard [22], and Gardner and Derrida [23] to write down an equation for the overlap between a contact and the correct pattern (in the approximation of dilute nets):

$$m_{ij}(t+1) = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp(z^2/2) \times \tanh \left[ \beta \left( \sum_{k,l} W_{ijkl}^0 m_{kl}(t) + h_{ij}^0 + \sqrt{\alpha} z \right) \right] , \quad (29)$$

or in the more familiar notation, in terms of the error function  $\Phi$ ,

$$m_{ij}(t+1) = \Phi \left( \frac{\sum_{k,l} W_{ijkl}^0 m_{kl} + h_{ij}^0}{\sqrt{\alpha r}} \right) , \quad (30)$$

similar to the spin overlap equation of conventional neural networks, where  $r$  is the parameter of overlap with noncondensed memories.

The size of the basins of attraction is determined from the slope of the error function near the fixed points  $m_f$ . The size of the basins of attraction is thus proportional to  $(1 - m_f)$ , which means the fraction of nonhelical contacts that can be recalled to a helix.

Numerical methods can again be applied to the solution of this equation. As before, we shall use an iterative simulation technique, where we first solve the equation for  $m(t=1)$  with  $m(t=0)$  inserted as zero. In the next step,  $m(t=1)$  is inserted, and the equation is solved for  $m(t=2)$ , and so on, until a stationary value is reached. We have assumed a uniform pattern of overlaps as in the case of one long  $\alpha$  helix and dropped the indices in the overlap, so  $m_{ij} = m$ . Therefore, the first equation for  $m(t=1)$  is

$$m(1) = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh[\beta(h^0 + \sqrt{\alpha}z)] \quad (31)$$

which is inserted in the equation at the next time  $t=2$ :

$$m(2) = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \times \tanh\{\beta[nW^0 m(1) + h^0 + \sqrt{\alpha}z]\} , \quad (32)$$

where  $n$  is the number of contacts in the overlap. The integration is done by a numerical forward Euler algorithm. Again, the synaptic weights are chosen to be constants and the threshold  $h_0$  is chosen to have realistic values between 0.1 and 0.5. The temperatures are varied over a wide range from 0 to 20 in  $k_B$  units. To find the fixed points, we calculate  $m(t+1)$  as a function of  $m(t)$

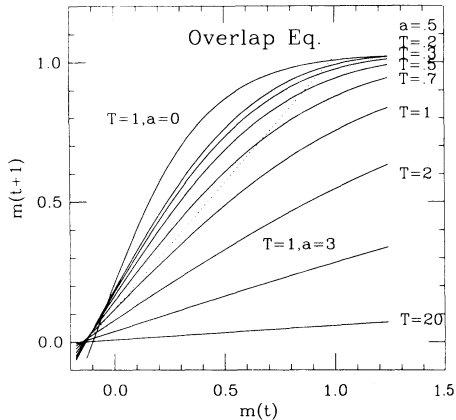


FIG. 2. Graphs of the numerical solutions to the overlap Equation (32), where the overlap  $m(t+1)$  is shown as a function of  $m(t)$ . (The overlap  $m$  has been made homogeneous in the indices  $i, j$  according to an  $\alpha$ -helix structure.) The various curves represent evaluations for different values of the temperature  $T$  and the randomness parameter  $\alpha$ . A straight dashed line describing  $m(t+1)=m(t)$  has been added in order to determine the fixed points that are given by the crossing of the straight line and the curves. For high temperatures and for large randomness, no nontrivial fixed points can occur, and the corresponding net cannot recall correct patterns.

and see where that curve crosses the straight line  $m(t+1)=m(t)$ . For temperatures below 2 and for the random factor  $\sqrt{\alpha}$  less than a critical value  $\sqrt{\alpha} \sim 1.5$ , there exists a unique fixed point for which the “correct” helical structure can be recalled.

In Fig. 2 we have drawn the curves for  $m(t+1)$  as a function of  $m(t)$  at different temperature  $T$  and random parameter  $\sqrt{\alpha}=a$ . As we expected, the position of the nontrivial fixed point  $m_f (m_f \neq 0)$  moves up towards  $m=1$  as temperature and random parameter  $a$  decrease, having helix patterns containing less noise. Furthermore, if we increase the thresholds, which here means increasing the  $\gamma$  factor and thus increasing the hydrogen-bond content, we enhance the helix-forming tendencies (the size of basins of attraction is increased) and the neural net can tolerate much more noise in making a perfect recall.

#### IV. CONCLUSIONS

This paper describes a way in which the initial events of protein folding can be thought about using the concepts of neural-network statistical mechanics. We have shown that a natural expansion of the free-energy function for a biomolecule in terms of contacts leads to a kinetic equation for the formation of contacts that resembles a recurrent neural network. The dynamics of the network imitate the initial events of folding in which no collective barrier-crossing processes are involved. Thus the theory complements the approaches to protein folding based on spin glasses that discuss long-time events involving the surmounting of barriers. Just as in the theory of neural nets, we have shown how to analyze the case of a dynamical competition between a finite number of structures. The case in which one structure dominates and many other structures can be thought of as random noise was also analyzed using methodology analogous to that of [22,23] for neural nets. This gives an estimate of the size of the basin of attractions as a function of the noise. The analysis was simplified by assuming a spatially uniform degree of overlap of the initial structure with the final structure. Some of the directions for future exploitation of this viewpoint include the use of recurrent network learning algorithms in order to develop practical structure prediction schemes. In addition, considerable work needs to be done on the problem of the size of basins of attraction. The analysis here for  $\alpha$  helices needs to be generalized to include the non-mean-field nature and the spatial inhomogeneity of the degree of overlap. In addition, the analysis of  $\beta$ -sheet formation using these methods would seem to be possible. Finally, it should be possible in this neural-network formulation to use some of the work of Gardner [24] on the phase space of interactions and get information-theoretical constraints on the possibility of a protein-folding code that operates in the extremely rapid regime of initial folding events.

#### ACKNOWLEDGMENTS

We wish to thank Dr. J. Hopfield, Dr. R. A. Goldstein, Dr. A. Caliri, Dr. Z. Schulten, Dr. J. Irwin, Dr. D. Evenisky, Dr. J. Clark, Dr. J. Bryngelson, Dr. K. E. Kürten, Dr. M. Sasai, Dr. R. Wade, and Dr. A. Herz for their helpful remarks and inspiring discussions. This work was supported by a National Institutes of Health Grant No. PHS-1-R01,GM44557.

- [1] N. Qian and T. J. Sejnowski, *J. Mol. Biol.* **202**, 265 (1988).
- [2] H. Bohr, J. Bohr, S. Brunak, R. M. J. Cotterill, B. Lautrup, L. Norskov, O. Olsen, and S. B. Petersen, *FEBS Lett.* **241**, 223 (1988).
- [3] L. H. Holley and M. Karplus, *Proc. Nat. Acad. Sci. USA* **86**, 152 (1989).
- [4] M. S. Friedrichs and P. G. Wolynes, *Science* **246**, 371 (1989).
- [5] F. Bohr, J. Bohr, S. Brunak, R. M. J. Cotterill, B. Fredholm, B. Lautrup, and S. B. Petersen, *FEBS Lett.* **261**, 43 (1990).
- [6] J. D. Bryngelson, J. Hopfield, and S. N. Southard, Jr., *Tetrahedron Comput. Methodol.* **3**, 129 (1990).
- [7] G. L. Wilcox, M. Poliac, and M. N. Liebman, *Tetrahedron Comput. Methodol.* **3**, 191, (1990).
- [8] P. G. Wolynes, in *Biologically Inspired Physics*, edited by L. Peliti (Plenum, New York, 1991), p. 15.
- [9] E. I. Shakhnovich and A. V. Finkelstein, *Biopolymers* **28**, 1667 (1989).
- [10] D. Stein, *Proc. Nat. Acad. Sci. USA* **82**, 3670 (1985).
- [11] A. Ansari *et al.*, *Proc. Nat. Acad. Sci. USA* **82**, 5000 (1985).
- [12] C. J. Levinthal, *Chem. Phys.* **65**, 99 (1968).
- [13] H. G. Bohr and P. G. Wolynes, in *Proceedings of the Workshop “Neural Networks: From Biology to High Energy Physics,” Elba, Italy, 1991*, edited by O. Behan, C. Bosio,

- P. del Giudice, and E. Tabet (ETS Editrice, Pisa, 1991).
- [14] J. J. Hopfield, Proc. Nat. Acad. Sci. USA **79**, 2554 (1982).
- [15] K. F. Freed, *Renormalization Group Theory of Macromolecules* (Wiley, New York, 1987).
- [16] Y. Oono and K. F. Freed, J. Phys. A **15**, 1931 (1982).
- [17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Nature **323**, 533 (1986).
- [18] H. S. Chan and K. A. Dill, J. Chem. Phys. **92**, 3118 (1990).
- [19] J. D. Bryngelson and P. G. Wolynes, Proc. Nat. Acad. Sci. USA **84**, 7524 (1987).
- [20] M. Sasai and P. G. Wolynes, Phys. Rev. Lett. **65**, 2740 (1990).
- [21] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. **173**, 30 (1987).
- [22] W. Krauth, J. P. Nadel, and M. J. Mezard, J. Phys. A **21**, 2995 (1988).
- [23] E. Gardner and B. Derrida, J. Phys. A **21**, 271 (1988).
- [24] E. Gardner, J. Phys. A **21**, 257 (1988).