Computer simulation of the folding of coiled coils

Antonio Rey[a)]
Departamento de Química Física I, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain

Jeffrey Skolnick
Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037

(Received 2 August 1993; accepted 13 October 1993)

A simple model capable of providing possible folding pathways of two stranded, coiled coil peptides is described and simulated using an off-lattice dynamic Monte Carlo algorithm. Short sequences of very regular repetitive blocks of amino acids are studied. The regularity of the sequence is enhanced by a simplified interaction scale between pairs of residues. Following the transition from two isolated chains in a random conformation to the folded dimeric structure, the main features capable of obtaining a parallel, in-register, unique conformation, are examined. These include the geometrical representation of the model, the cooperative development of secondary and tertiary structures, and the role of tertiary interactions stabilizing the coiled coil geometry. The influence of introducing disulfide bridges in certain locations of the sequence is also discussed.

I. INTRODUCTION

The theoretical modeling of the protein folding process constitutes one of the most active research fields of contemporary theoretical biophysics. From this point of view, computer simulation techniques nowadays constitute a powerful tool, that can complement experiment for the detailed consideration of the dynamic processes involved in the folding and unfolding pathways of a polypeptide chain.[1] Most simulations considering the protein folding problem have mainly focused on globular proteins, and considered the transition of a single chain of amino acids from a broad set of extended random coil conformations to a unique compact native shape. Thus, the folding process involves the simultaneous development of secondary and tertiary structure, during the cooperative transition that yields the native conformation. This cooperativity is one of the most fundamental points to be considered, and represents one of the controversial topics as well.[2]

Different global mechanisms proposed for the folding pathways have postulated several alternatives, including the collapse of the chain to a dense structure, the diffusion of preformed elements of secondary structure, and others. At the bottom of all these hypotheses lies the intrinsic complexity of the folding process, and the difficult interpretation of an increasing, but still reduced, set of experimental facts.

In this paper, we use a simplified model to investigate the folding pathways of simple proteins. Our aim is to keep the topology as simple as possible, which will permit the detailed exploration of the main features of the dynamic pathway. In particular, we shall examine the importance of the geometrical details of the model, especially the side chain description, the interplay between local secondary structure tendencies and tertiary interactions to achieve a unique folded structure, and the role of disulfide bridges in both folding pathways and stability of the resulting structures.

In order to achieve these objectives, we have focused on protein structures having completely regular secondary structure, and a very uniform three-dimensional fold. Specifically, we have chosen the structure of the coiled coil fibrous proteins, which are involved in biologically important functions such as muscle regulation, or DNA binding properties.[3,4] These are dimeric structures composed of two right-handed helical polypeptide chains, which are parallel, in-register, and coil about one another. The number of residues per turn is equal to 3.5, instead of the 3.6 residues per turn one finds in α-helices in globular proteins. This difference is created by the slight left-handed super twist, and gives a highly defined hydrophobic face in every helix. This way, the full structure is stabilized by both intrachain and interchain interactions. Intrachain interactions are, in this case, short ranged, and mainly induce (though are not the only responsible for) the helical secondary structure. Interchain interactions, on the other hand, play the role of tertiary interactions in globular proteins. In coiled coils, however, they are responsible for the stabilization of the quaternary structure of the molecule (strictly speaking, there is no tertiary structure in a coiled coil).[5] Therefore, model coiled coils can be used to investigate the noncovalent interactions involved in the stabilization of the three-dimensional structure of a protein, a feature that is common to both globular and fibrous proteins.[6]

Both synthetic and real coiled coils have been experimentally and theoretically used as test structures of the protein folding process for a number of years.[5–18] Recently, the atomic coordinates of one of these structures were even theoretically predicted, being very close to the crystallographic structure determined afterwards.[19] Nevertheless, the starting point for that treatment was composed of two

a)To whom correspondence should be addressed.

 0021-9606/94/100(3)/2267/10/$6.00

perfectly regular, parallel helices. Here, we present a model which precisely explores the way in which such a conformation can be reached, starting from two separated chains, in random conformations. Thus, the model cannot only provide useful information related to the main features of protein folding, but also constitutes one of the simplest cases of multimeric assembly.

Our simple model allows as well for the introduction of cysteine sulfur bridges that covalently link the two chains. As expected, the stability of the resulting conformations is enhanced when they are present. However, there are occasions in which these sulfur bridges can create a configurational stress into the system, producing rather tortuous initial stages in some folding pathways. We shall describe these pathways, and their differences with the folding process of non-cross-linked chains.

It is important to remark that we do not try here to present an algorithm capable of providing a low resolution structure similar to the real folded protein, starting from the amino acid sequence alone. The model we employ is far too crude for that, and as it will be described below, it contains certain contributions biased towards the desired final conformation. The final structure we get cannot be considered, from this point of view, as an absolute prediction. On the other hand, we shall present features of the folding pathways obtained with our model that could well represent a plausible physical pathway for the assembly of two chain coiled coils.

In the next sections, we describe the characteristics of the model and the simulation algorithm. Then, we present some results corresponding to specifically designed sequences which mimic real coiled coils. In the last section of the article, we summarize the main conclusions and possibilities of the model.

II. DESCRIPTION OF THE MODEL

A full description of the theoretical model developed includes the consideration of its geometrical, dynamic, and interaction features.

The geometry of the model is kept very simple. We are not interested in getting a full atom folded structure, and therefore all the atomic details characteristic of the chemical structure are suppressed. Instead, a series of spheres representing sets of atoms is employed. Specifically, every amino acid is represented by two spheres. One corresponds to the backbone atoms and is centered at the α-carbon position of every residue. The distance between two neighbor backbone spheres along the sequence equals 3.8 Å, the distance between two contiguous α-carbons linked through a *trans* peptide bond. These spheres are identical for every possible residue in the sequence. The second sphere, which appears in all the residues with the exception of glycines, constitutes a crude representation of the side chain. It is positioned at the center of mass of the real atoms comprising the side chain. Its size and orientation depend on the chemical nature of the residue under consideration, i.e., on the primary sequence of the peptide chain. Statistical analyses of crystallographic structures have shown that the possible coordinates of the center of mass of the side chain with respect to the backbone, for a given local conformation of the latter, are not randomly distributed. Instead, they are discretely located.[20] Therefore, an amino acid-dependent rotamer library has been constructed from the aforementioned statistics, and has been used to locate the sidechain sphere of every amino acid in the model once the backbone conformation (the first sphere position for three contiguous residues) is known.[21] This model is very simple, but allows us to reproduce the gross amino acid-dependent geometrical features of a polypeptide chain in a compact and computationally tractable way. In some initial tests of the model, a single rotamer was used to represent the side chain of every residue, as would correspond to a poly-alanine geometry. In the following sections, we will briefly mention the effect of such a further simplification of the geometry of the model in the observed folding characteristics.

Due to the simplifications in the geometrical representation, the dynamics of the model has to be formulated in equivalent terms. We have chosen a dynamic Monte Carlo algorithm to mimic the conformational transitions taking place during the folding process. This algorithm is defined through a series of arbitrary moves that affect a variable number of model units (both for the backbone and the side chain spheres). These motions can move the model units to any position in the space, always maintaining chain connectivity and rigid virtual bond lengths. Therefore, no underlying lattice is used in these Monte Carlo simulations.

The internal moves (i.e., those affecting a single chain) include spike and end moves, which change the position of a single residue, and shifting moves, which move a large portion of the chain to a close parallel position.[22] All these motions also affect the corresponding side chain positions. The side chains themselves can be modified as well with motions which keep the α-carbon trace untouched.

In addition to these internal moves, identical to those used in single chain simulations,[22] global motions affecting a whole chain are included in the present algorithm. These occur by rigid body translations or rotations of a single chain with respect to the other. The amplitude of these moves is small (less than 3 Å for the displacement of the atoms during translations, and less than 4 degrees for the Euler angles controlling the rotation, centered at a randomly chosen backbone unit of the model). The full set of Monte Carlo moves creates a very physical picture for the dynamics of the system. Different motions affect different portions and different lengths of each chain. Even then, the individual moves retain a local character, in the sense that the conformational changes induced by them, on an individual basis, are rather small. None of these individual moves represents physical conformational transitions occurring in real chains, though spike moves can probably represent localized torsional transitions as those occurring in hydrocarbon chains. This is not important, since we do not try to study the very fast local dynamics of the system. We just try to define a set of moves that can span all the conformational space accessible to the protein in this simplified representation. In addition, through the combination of a large sequence of such motions, real dynamic

behavior emerges, which is similar to Brownian dynamics simulations performed for simple systems.[23]

Probably, the definition of the energetic interactions that try to mimic the free energy landscape of the conformational space is the most important feature of the model. In this initial study of simple coiled coils, since the secondary structure is perfectly regular, we are mainly interested in the interactions responsible for the quaternary structure. Therefore, the main contribution to the potential defined for our model is that corresponding to the nonlocal (or tertiary) interactions, i.e., interactions between pairs of amino acids not directly connected along the sequence.

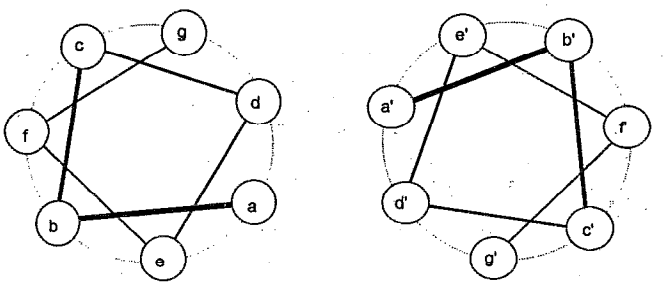
These are not, however, the only components of the free energy function. Local interactions defined at the level of virtual bond angles and virtual torsion angles (defined, respectively, by three and four consecutive α-carbons) are also included.[22] These are minor contributions that permit control of the rigidity of the chain. Through them, we have introduced a local bias towards helical states, by constraining the virtual bond angle to a value of about 90°, and the virtual torsional angle close to 60°. Soft harmonic potentials are used for these contributions. This could be considered as a flaw in the model, since it might appear as an arbitrary contribution without any physical background. However, there are some considerations to be taken into account: First, these terms alone do not induce stable helices in isolated chains. In the simulated trajectories we have obtained, the helical population of isolated chains never amounts to more than about 30 percent. Only when the dimer forms and the coiled coil structure appears, are the helical conformations perfectly stable. Second, for the amino acid sequences of the peptides we have considered, any secondary structure prediction algorithm[24,25] provides an overwhelming preference towards helical states. Therefore, our biased contributions in the potential could be partially contemplated as a mean field representation of intrinsic preferences already incorporated in the primary sequence of the peptides under consideration.

Local interactions also include angular correlations between the orientation of the side chains (in the second, third, and fourth closest neighbors) with respect to the backbone. These interactions only depend on the geometry of the residues, and therefore on their chemical nature. It has been previously demonstrated[20] that this contribution to the potential is able to induce the formation of regular elements of secondary structure.

As stated above, the most important contribution to the free energy, and of course the most interesting, is the one corresponding to nonlocal interactions between pairs of residues, when they are separated by at least three residues along the chain backbone. There are several pairwise potential scales between amino acids that try to mimic real interactions between residues in a protein.[20,26,27] Most are based on different hydrophobicity scales, or on the statistics of contacts found in crystallographic structures of globular proteins. In principle, any of them could be a candidate to be used in our model for coiled coils (maybe with some caution, since the ratio of buried to surface residues is rather different in a coiled coil and a globular protein). However, at the moment we have chosen a simpler approach, reducing the interactions among 20 different amino acids to only four groups of them. This is inspired by the coiled coil designed sequences of Hodges *et al.*,[6,9,10] which are themselves based on sequence analysis of real coiled coil fibrous proteins, mainly tropomyosin.[28] It has long been known that hydrophobic residues in the core of the protein are mainly responsible for the stabilization of coiled coils. In addition, charged residues with alternating positive and negative charges stabilize the parallel conformation of the coiled coil against the antiparallel situation. The residues facing the exterior of the protein can have a different nature. Here, nevertheless, it is enough to consider them as hydrophilic residues, which interact favorably with the solvent, and avoid the possible aggregation of several fibers into more complex multimeric structures. Thus, it would in principle be possible to consider, from the point of view of nonlocal interactions, only four types of residues, namely hydrophobic, hydrophilic, positively charged, and negatively charged. Hydrophobic residues attract other hydrophobic residues, and repel any other residue type. Hydrophilic residues repel hydrophobic ones, but are neutral to other members of the same type and to charged residues (though a strong short distance repulsion due to excluded volume effects is effective between any pair of residues, independent of their nature). Finally, charged residues attract charged residues with opposite charge, repel charged residues with the same charge, as well as hydrophobic residues, and are neutral towards hydrophilic amino acids. Table I shows the mathematical definitions we have employed for these attractive and repulsive pair potentials, together with the potential parameters used in our trajectories. We do not claim that the interactions between real amino acids can be correctly reproduced by Lennard-Jones type potentials, in which neither the spherical symmetry, nor the distance dependency, correspond to real interactions. We only try to define a

TABLE I. Contributions to the potential for nonlocal interactions defined in the model. They depend on the nature of the pair of residues. A pair of residues is considered to be interacting when the distances between their side chains centers of mass, r, is less than r_{cut}. To avoid very large energy contributions from small overlaps, for distances $r<r_{min}$, the potential takes the value defined at r_{min}.

Interaction pair	Expression	r_{cut}/Å	σ/Å	r_{min}/Å	ϵ/k_BT
Pho-Pho	$4\epsilon\left[\left(\frac{\sigma}{r}\right)^8-\left(\frac{\sigma}{r}\right)^6\right]$	5.8	2.7	2.5	8.0–10.0
Pho-Phi *Pho-Ch*$^{\pm}$	$4\epsilon\left(\frac{\sigma}{r}\right)^8$	5.8	2.7	2.5	4.0—5.0
Phi-Phi *Phi-Ch*$^{\pm}$	ϵ	2.9	···	···	4.0–4.5
Ch$^+$-*Ch*$^-$	$4\epsilon\left[\left(\frac{\sigma}{r}\right)^4-\left(\frac{\sigma}{r}\right)^2\right]$	8.6	5.2	5.0	2.2–2.5
Ch$^+$-*Ch*$^+$ *Ch*$^-$-*Ch*$^-$	$4\epsilon\left(\frac{\sigma}{r}\right)^4$	8.6	5.2	5.0	2.4–2.7



Position	Geometry residue	Interactive residue
a	Leu	*Pho*
b	Glu	*Phi*
c	Ala	*Phi*
d	Leu	*Pho*
e	Glu	Ch^-
f	Gly	*Phi*
g	Lys	Ch^+

FIG. 1. Projection view of the two helices in a coiled coil, from the *N*-terminus. The heptapeptide sequence corresponding to a double helical turn is shown. The double assignment of the residues, according to their geometry (their real chemical nature) and their interaction properties, is indicated at the bottom of the diagram.

model in which the number of parameters is kept as low as possible, yet it can reproduce some fundamental features in the physical folding pathways.

To define which type of interaction is associated with each residue in the chain, we have used a criterion based on the putative position of the amino acid in the final coiled coil structure, and not on the chemical nature of the residues (which, nevertheless, is used to define the size and geometry of the side chain). Thus, a peptide whose sequence were based on the repetition of the heptad Lys-Leu-Glu-Ala-Leu-Glu-Gly, as appears in some of the Hodges' peptides,[9] would correspond in our model to an interacting sequence of the type Ch^+-*Pho-Phi-Phi-Pho-*Ch^--*Phi* (where *Pho* represents a hydrophobic residue, *Phi* a hydrophilic residue, and $Ch^{\pm}$ a charged residue of positive or negative charge). If we use the standard notation for the seven positions of a double helical turn in a coiled coil (whose cross section is represented in Fig. 1), positions **a** and **d** will be occupied by hydrophobic residues, positions **e** and **g** are occupied by charged residues of opposite signs, and positions **b**, **c**, and **f** are occupied by hydrophilic residues. It might be thought that the consideration of Glu in position **b** and Ala in position **c** as hydrophilic residues is too crude even for a simplified model. Actually, it is possible to consider that the role of these two residues in the designed sequence is to favor helical states, since both are well-known helix inducing amino acids. Since our model explicitly accounts for the formation of helical states independently of the amino acid nature, we believe the aforementioned approximation does not introduce unphysical simplifications into the model.

Nonlocal interactions defined this way do not distinguish whether the pair of interacting amino acids are included in the same chain or they belong to different chains. This way, the formation of the dimer against the collapse of individual chains can be explored.

When one defines a molecular system in which more than a single molecule (polypeptide chain, in our case) is considered, there must be a way to avoid the situation when chains diffuse away from one another; i.e., the chain concentration must be controlled. This is usually achieved through the introduction of periodic boundary conditions.[29] In our model, having only two chains, we have opted for a simpler (and computationally more efficient) way to treat this problem, and consists of the introduction of an additional term in the potential energy function, dependent on the separation between the two chains. When their centers of mass are separated by a distance larger than a certain threshold, a strong harmonic constraint appears. This constraint completely vanishes if the distance between centers of mass is below the threshold. This way, a spherical space is defined in which the chains move freely, without being forced to be joined together. The size of this spherical cavity is always large enough (depending on the chain dimensions, and ultimately on the length of the amino acid sequences involved) to allow for a considerable freedom of motion of the chains in the unfolded, i.e., not helical and not associated, form. The radius of the cavity has a value of about 40 Å for the longer chains used in our simulations. This translates in a concentration of the order of 5 mM, larger than the values used in experimental studies. Still, a cavity of this size represents a huge space for two single chains to move. This makes the simulation somewhat slower, since an important percentage of the computer time is wasted exploring the (now) uninteresting dynamics of individual random coils. We believe this is a better method to study a model of a real system than to force the chains to be close in space from the very beginning, with the possible distortions of the folding pathways that could create.

III. FOLDING PATHWAYS

Following the experiments of Hodges *et al.*,[9] we have initially considered regular sequences of the type Gly-(Lys-Leu-Glu-Ala-Leu-Glu-Gly)$_n$, with $n=2$, 3, 4, and 5. Therefore, the total length of the peptides studied is 15, 22, 29, and 36 amino acids, for each of the two chains considered. The Gly residue added at the *N*-terminus of the original Hodges' peptide allows the model to define the side chain position of the second residue (since, as mentioned previously, three α-carbons are necessary to define the backbone conformation on which the side chains are built).

The starting conformation of the simulation corresponds to two random coils, usually in rather extended conformations, with their centers of mass considerably separated. From that initial situation, the system evolves according to an asymmetric Metropolis scheme. The sequential set of conformations appearing along the calculation constitute the dynamic trajectory of the system. Some an-

nealing trajectories were computed in order to properly determine the temperature of the unfolding-native conformation transition. All the results presented here correspond to isothermal calculations performed at this transition temperature.

The computed trajectories show a fundamental difference between the folding of short peptides (with 15 or 22 amino acids) and the longer ones (29 and 36 amino acids). The former, due to the reduced length of the chains, show very few intramolecular contacts (i.e., contacts within the same peptide chain). The intrinsic rigidity of the chain backbone, in both the random and helical conformations, precludes the possibility of these contacts to frequently occur. On the other hand, when the length of the chain grows, intramolecular contacts become more frequent. Under these conditions, a competition appears between intramolecular and intermolecular (native or not) contacts. In these cases, the same conditions that favor greater coiled coil stability (namely, a strong set of hydrophobic interactions) make it more difficult to reach the same structure from the unfolded state. Often, the individual chains separately collapse on themselves, forming dense globules without regular secondary structure. This contrasts with the results of other simplified models for the folding of globular proteins, whose conclusion has been that the driving forces for the collapse of a peptide chain contribute as well to the development of the correct secondary structure.[30] Our results do not support this conclusion.

In addition, since the collapsed globules present an outer face mainly composed of hydrophilic residues, it is very difficult for the two globules (or a globule and an extended chain) to interact with each other by unwrapping the collapsed compact conformations and growing the coiled coil structure. We do not really know how real chains solve this competition. Probably, the hydrophobic interactions are not so large in these sequences to stabilize single chain globular conformations, and only by forming dimeric long fibers in-register is the molecule stable. We have been unable to find any experimental data describing the dimensions or conformation of individual chains in the unfolded state of a coiled coil protein. Nevertheless, it is rather plausible that the lack of specificity of our potential energy functions is mainly responsible for the observed collapse.

To reduce this effect, we have introduced a soft coupling between nonlocal interactions and secondary structure. That is, the strength of the nonlocal interactions between any pair of residues is scaled according to the local conformation they present. Only when both amino acids present a helical conformation do the tertiary interactions display their whole value. When none is in a helical conformation, only 20% of the interaction is retained. Between these two limits, a smooth linear scale is defined according to the "helicity" of the two residues involved (defined through the virtual torsion angles of the backbone around them[22]). It is important to mention that this scale affects any possible nonlocal interaction in the model, independently of the nature and position of the pair of amino

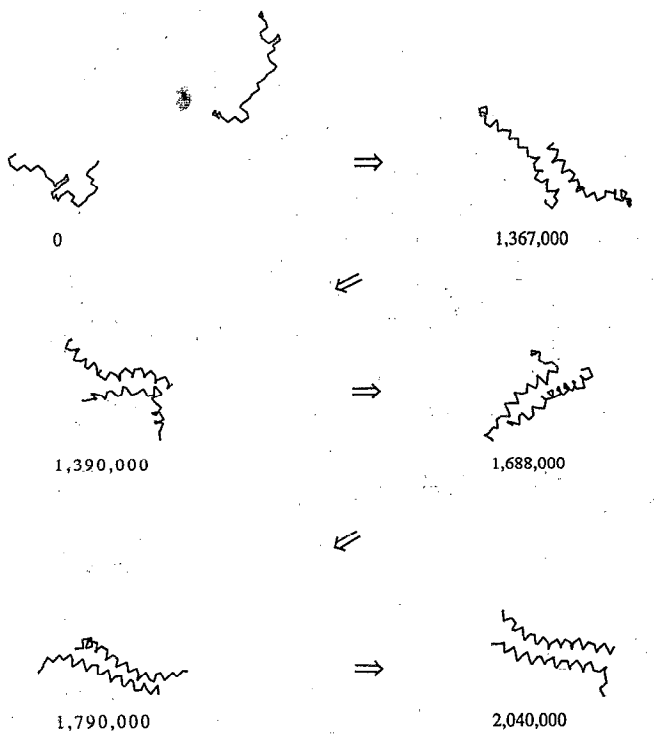


FIG. 2. Characteristic snapshots along a Monte Carlo dynamic trajectory for two independent chains of 36 residues each. The number of Monte Carlo steps from the beginning of the trajectory is indicated. Notice that almost one half of the trajectory involves conformational transitions prior to the beginning of a successful folding event.

acids, including whether they are in the same chain or in different chains.

We recognize this approach introduces a second biased term into the potential function towards the desired final structure. Again, it just tries to cover the defects imposed by the simplifications of the model. It does not avoid the appearance of non-native contacts, nor does it preclude intrachain long lived contacts in the longer peptides. On the other hand, it is very similar in spirit to the use of the ω° parameter introduced in some analytical theories developed for the formation of two-stranded, coiled coils.[15]

Using this new feature for the longer chains (it has no effect in the short chains), the folding pathways follow a series of general steps, independent on the size of the peptide chains. These can be monitored by obtaining snapshots of instantaneous conformations along the trajectories, and by observing the evolution of some characteristic properties for the system. Among them, we have chosen as most representative the helical content of the chains, the global energy of the system, and the distance between the centers of mass of the two chains involved in a given trajectory. For an isothermal trajectory of two chains comprising 36 units each, Fig. 2 shows some representative snapshots with important conformations along the folding pathway. The evolution of the aforementioned properties is shown in Fig. 3.

At the beginning of the trajectory, the two chains are separated (with a considerable distance between their centers of mass), and therefore they do not interact. We

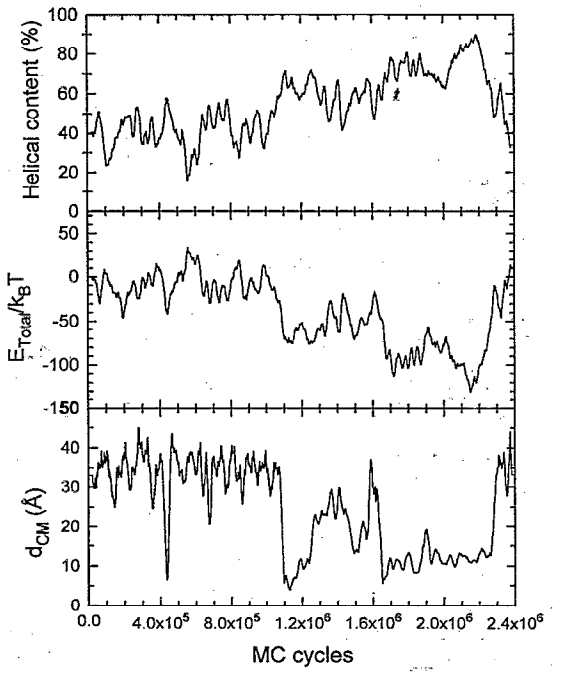


FIG. 3. Evolution of the helical content, the reduced total energy, and the distance between centers of mass along a trajectory for two independent chains of 36 residues each.

should also mention that the initial conformations were rather extended, giving a very small number of intrachain contacts. Consequently, the energy of the system is essentially zero. The two individual chains, when separated, show a certain preference toward helical states, that never amounts to more than about 30% of the torsional states. The configuration essentially corresponds to a random coil. Promptly, as the system evolves, nonlocal contacts begin to appear among hydrophobic residues. When the contacts are intramolecular, they usually form and dissolve quickly, though some of them survive during short periods of the simulation trajectory. These contacts create the fluctuations appearing in the global energy and the helical content. The same happens with intermolecular contacts. Many form and dissolve during the first stages of the trajectory. These involve all kinds of combinations of the two chains: any unit in one of the chains can approach every unit in the other, creating the fluctuations in the interchain distance. Obviously, the actual folding process commences at one of these intermolecular contacts. Most of the time, as one would expect just by probability considerations, this is a non-native contact, i.e., a contact between two hydrophobic residues (two Leu amino acids) which are not in-register in the coiled coil structure. In addition, this contact frequently corresponds to a crossed configuration of the two peptide chains, without any parallel or anti-parallel arrangement. As a consequence, the system is not stable enough to retain the interacting configuration, and the two chains run apart. If, on the other hand, the parallel arrangement grows, then the contact may represent the beginning of a successful folding pathway. This is what happens, in the particular simulation shown in Figs. 2 and 3, after almost one half of the trajectory has elapsed. What one observes, in addition to a sudden drop in the interchain distance, is an increase in the number of hydrophobic contacts, which is due to a double effect, clearly reflected in Fig. 3: the growth of the helical secondary structure in the individual chains, up to a total close to 90%, and the nestling of their Leu side chains when the two chains adopt a parallel arrangement, thereby reducing the energy of the system.

In a model with such a simplified geometry as ours, this nestling could in principle be obtained as well with an anti-parallel arrangement of both chains. In order to check the reason for the parallel arrangement, we have computed some trajectories with a simpler model, in which only hydrophobic and hydrophilic residues are considered, and charge interactions are ignored. In these cases, as expected, the coiled coil structure is also obtained in our simulations, but both parallel and anti-parallel arrangements appear with equal probability. In real chains, the interaction between the helix dipoles, not present in our model, might tend to favor anti-parallel configurations. Nevertheless, as has been supposed from experimental results,[9,31] our model confirms that it is the presence of charged residues of opposite sign that produces the parallel disposition of coiled coil structures.

The snapshots in Fig. 2 and the evolution of properties in Fig. 3 clearly indicate that the transition from the unfolded to the folded state is not a two-step process. On the contrary, the correct fold propagates in a continuous way by increasing the number of hydrophobic contacts and the helical content of the peptide chains simultaneously, until the maximum number of contacts is reached. This is indicated by a maximum in the helical content plot, and a minimum in the total energy. The full process can be a long one, with a large number of intermediate states (as some theoretical studies had previously postulated[18]). This result is not contradictory with the all-or-none folding transitions proposed for globular proteins, in which the transition from the unfolded to the folded state is much more abrupt. However, our results support the idea that the use of this two-state model for the interpretation of folding pathways in coiled coils does not seem to be the best option.

It is interesting to consider in more detail how the development of the correct folded structure takes place. Since the first contact at which the process is initiated does not usually correspond to a native one, the chain at this stage is out of register (see the snapshots in Fig. 2). The free ends of the chains not involved in nonlocal contacts show random conformations, without any special tendency toward helical states. This confirms our belief that the nonlocal interactions are mainly responsible for the final stabilization of the observed secondary structure. The intrinsic tendency toward helical states of our model (and of real amino acid sequences appearing in coiled coils) is not able on its own to create stable helices. Experiments in aqueous solution have been repeatedly interpreted using this assumption, whose validity is suggested by the simulation (see, for example, Ref. 5, and references therein), though

the experimental verification of very low helical content in individual chains has been never established.

In order to achieve the registration of the two chains, a shift of the interchain contacts has to occur. We have to remember that the hydrophobic contacts result from a certain interdigitation of the side chains, as happens in globular proteins as well. Therefore, this shift in registration is not straightforward. We observe that it is accomplished by a "scissorslike" motion. That is, the angle formed by the helical portions in contact, initially close to zero (as would correspond to a perfect parallel disposition), slightly increases while the shift in registration takes place. The procedure continues step after step, opening and closing the angle formed by the two chains until the perfect registration, and the maximum number of favorable tertiary contacts, is achieved. In many occasions, especially for longer peptides, in which the registration shifts can be rather large, if the initial contact is very far from native, the growing angle between the two chains significantly reduces the number of contacts, and then the two chains fall apart, with immediate loss of most of their helical character.

Since no experimental information is available to verify this propagation mechanism, we slightly modified the model in order to check whether this observed pathway could be considered as a physical one. In order to do so, we have run some trajectories keeping the interaction definition of the residues as hydrophobic, hydrophilic, and charged, as in the original sequence, but simplifying the geometry. What we did is to reduce all the amino acid geometries to that of Ala, i.e., a small side chain sphere rigidly attached to the backbone. This suppresses the internal degrees of freedom represented by the set of side chain rotamers. For this model, the shifts in registration are achieved in a much simpler way. Actually, a very small modification of the interaction angle among the two peptides is observed, but the shifts mostly happen through a continuous sliding of the hydrophobic surfaces one over the other. From our results, it seems that the hydrophobic face created by side chains with Ala geometry is much smoother than that created by side chains with Leu geometry. We must remember that, in the framework of our model, both geometries are considered as single spheres. The main difference between them is their size, and the ability of Leu side chains to acquire different conformations (or rotamers) with respect to the backbone, a possibility which does not exist for Ala side chains. Real side chains have a much more complex geometry, and the interdigitation between interacting pairs of hydrophobic residues will be far more complicated than that appearing between our spherical units. Therefore, we believe that the "scissors" movement shown by our model during shifts in registration is much closer to reality than any smooth sliding of hydrophobic faces. In addition, this reinforces our belief that a certain level of accuracy in the representation of the side chains is fundamental to any model which aims at extracting useful conclusions about folding pathways.[22] (Incidentally, a model in which the sphere representing the side chains is suppressed, resulting in a poly-glycine geometry, in which the tertiary interactions are centered at the α-carbon positions, was not adequate for this study. In order to get the correct separation between the two chains, the interactions, both attractive and repulsive, have to be very long ranged, and compact structures of globular single chains were the outcome of the simulations in most of the cases.)

All together, the observed mechanism of assembly in our trajectories is fully consistent with the unimolecular rate-determining mechanism previously proposed from stopped flow circular dicroism (SFCD) experiments on tropomyosin.[13,14] This mechanism involves two steps: In the first step, very fast, unfolded chains rapidly form a dimeric, partially folded coiled coil; the second step involves the slow conversion to the native coiled coil. A possible alternative mechanism in which the fast step would involve the formation of stable helices, that subsequently associate slowly in a second step to form the native structure, was discarded on the basis of the observed SFCD experimental results. We have never observed in our trajectories this kind of assembly either. The global time scale of the simulations presented in this article mainly corresponds to the early stages of folding, and is therefore probably included within the dead time of the experimental SFCD instrument. However, the overall conclusions we get from our trajectories are almost identical to the interpretation of the experimental results.

There is a second result in the simulated trajectories which indicates that the side chain representation included in the model captures the main geometric features of real side chains. It is well known[3,4] that actual coiled coils show a left-handed super twisting with a very long pitch (in comparison to that of an α-helix). The structural reason for this super twist is not clear, but it could well be related to the geometry of the side chains involved in the stripe of hydrophobic residues that forms the core of the fibrous protein.

For the chains having the Leu geometry for the **a** and **d** residues, most of the folded states we obtain are very mobile, with the tails of the helices continuously unfolding and refolding, and the angle formed by the two helical chains changes without showing any clear trend. Nevertheless, if the temperature of the system is slightly reduced once the coiled coil structure has appeared, the mobility drastically drops. And, in this case, a clear super twisting immediately appears. What is more, when the side chain geometry of the real sequence is used, the super twisting always shows the expected left-handed orientation. An example of this situation is shown in Fig. 4, where the folded conformation of one of the trajectories is shown in a top view. The angle of super coiling is a bit too large in comparison with real peptides. We have not been able to determine the exact value of the pitch, since the twisted structure is still mobile enough inside a broad basin of left-handed coiled states to preclude an accurate determination. This is what one should expect of a simple model as ours, both at the level of secondary and quaternary structures. However, the role of side chain geometry in the super twisting of coiled coils seems to be captured even in this model. In contrast, when the Ala geometry is used for

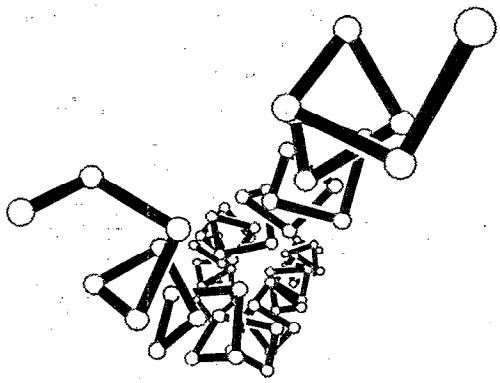


FIG. 4. Top view of the folded, coiled coil conformation resulting from a dynamic Monte Carlo trajectory. The left handed super twist is clearly appreciated.

the side chains, both left-handed and right-handed super twisting is observed.

The length of the peptide has also an influence on the peptide stability. The longer the peptide, the larger the stability. This can be measured through the mean life of the coiled coil structure once it forms in a simulation at the folding-unfolding transition temperature, as the one considered in Figs. 2 and 3. The last steps of this trajectory correspond to the beginning of the unfolding process, in which the chains separate, the helical content abruptly drops, and the energy of the system grows. This process is much faster than the folding mechanism, but it proves the reversibility of the transition, and confirms once more that isolated helices are not stable in the absence of long range (quaternary) interactions. As one would expect, when the length of the peptide chain grows, the number of possible hydrophobic contact increases, and the same happens with the stability of helical states.[9] On the other hand, the formation of the folded structure usually takes a larger time for the longer peptides. This is not a surprise either, since the folding pathway of coiled coils follows a continuous transition through many intermediate states. This stands in contrast with the all-or-none folding transition one observes in globular proteins.

IV. CROSS-LINKED CHAINS

Finally, we wanted to check the effect of cross-links on the behavior of our model. Experimental and theoretical studies on peptides in which a Leu residue is substituted by a Cys in every chain have shown that the stability of the resulting coiled coils can be enhanced: both the registration and the parallel disposition of the two chains are automatically preserved.[6,17] In addition, from a practical point of view, the "effective concentration" of peptides is increased. Since the two chains are chemically bonded, all regions of configurational space in which they are separated are not accessible any more. On the other hand, by slightly distorting the coiled coil structure, cross-links can also induce a local stress, which can somehow distort the folding pathway with respect to the situation in which these links are not present.

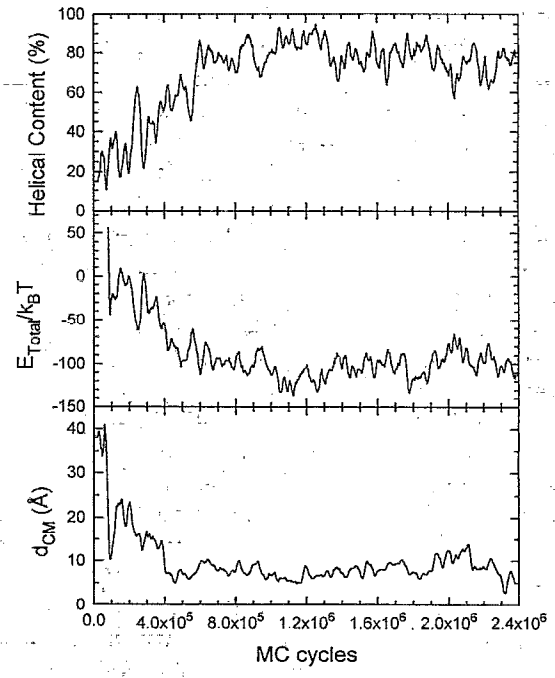


FIG. 5. As in Fig. 3, but for a sequence including a CYS residue at the middle of the chain.

We have defined a Cys amino acid by changing the interaction definition of the corresponding unit of the model. For us, a Cys residue can only interact with another (only one) Cys residue, when their side chains are at a distance less than 10 Å. This interaction is defined through a strong harmonic potential, with a minimum at a distance of about 4 Å (which roughly corresponds to the average distance between side chains centers of mass of two covalently bonded Cys, as found in crystallographic structures). With this definition, we have run simulations under identical conditions to those previously defined. We have considered three cases: a single Cys substitutes one Leu at one end of the peptide chain; the substitution takes place in the middle of the peptide chain; and two Cys substitute two Leu residues at both ends of the chain. Obviously, the primary sequence after the substitution is identical for both chains in the simulated system.

In Fig. 5 we show the evolution along a Monte Carlo trajectory of the helical content, the total energy and the distance between centers of mass of the two chains, for peptides of 36 units each, with one cysteine at the middle of every chain. The simulation begins again with two separated chains in random conformations. With the present sequences, however, due to the special potential defined between Cys residues, once the two chains become close in space the formation of the disulfide bridge is almost immediate, as reflected by the sudden drop in the distance between centers of mass. The total energy drops as well, due to the great energy reduction defined in the model for the formation of the disulfide cross link. If the chains are small (15 or 22 residues), the development of the coiled coil structure is very fast from that moment on. By growing the secondary and quaternary structure in two directions, if

the Cys residues are in the middle of the chain, or only in one direction, when Cys residues occupy one end of the sequence, very stable folded structures appear in much shorter simulation periods than those required by the peptides without Cys. Such growth was conjectured previously based on loop entropy considerations.[17]

In longer chains (29 and 36 residues), some problems appear in certain trajectories once the disulfide bridge is formed. These are caused by a configurational stress created in the chains by the constraint imposed by the disulfide bridge. What we observe is a certain collapse of one of the chains in the vicinity of the disulfide bridge. This collapse usually appears when, at the formation of the disulfide cross link, the two chains have an antiparallel or crossed configuration. Instead of rotating one of the chains around the common bond to create the parallel arrangement, a fragment in one chain may crumple over the linking position, allowing its tail or tails to grow the coiled coil structure in a parallel disposition to the other chain. Of course, since several residues form part of this wrongly folded portion of the chain, a correct registration cannot be achieved in this situation. In addition, the presence of the disulfide bridge precludes any possibility of correcting the chains registration by means of the "scissors" motions we observed in non-cross-linked chains. Those facts explain the fluctuations in the different properties which appear during the first part of the trajectory after the formation of the crosslink, as shown in Fig. 5. Nevertheless, all these difficulties are finally overridden, and the folding pathway continues as in the shorter chains we mentioned above. The stability of the cross-linked coiled coil structure is far larger than that of the independent chains, as shown by the small fluctuations present during most of the trajectory, mainly due to tail unfoldings and refoldings, and sometimes to small modifications in the left-handed super twist of the folded structure.

Another different problem related with the folding of covalently linked chains appears when we have disulfide bridges at both ends of the chains (we have studied this case only for the peptides of 29 and 36 residues per chain). The problem is almost equivalent to getting a coiled coil structure from a ring molecule. The algorithm is able to find this structure, which appears to be completely stable once it is formed. However, its formation is difficult as well. It always begins at one of the sulfur bridges. From that point, the propagation is not an easy task. Since the ends of the unfolded portions of the chains are joined together by a second disulfide bridge, it is relatively difficult for them to develop the helical secondary structure that creates the most favorable quaternary interactions, or vice versa. This, of course, could be related to an imperfect definition of the Monte Carlo motions included in the model. Anyway, the algorithm is able to find the correct structure. Whether the problems observed with one and two disulfide bridges indicate deficiencies in the algorithm, or are a physical effect that just mimics the negative effects induced by the configurational restrictions created by the disulfide bridges, is at the moment difficult to definitively address. Experiments can be very helpful to clarify this point, although the situation of the mentioned problems among the initial events of the folding process could make this investigation terribly complicated.

V. SUMMARY AND CONCLUSIONS

In this work, we have explored the possible folding pathways of coiled coil peptides by means of a simple model and an off-lattice dynamic Monte Carlo algorithm. The model includes a geometrical representation of the peptide backbone and the side chains based on real amino acids. The potential energy scale defining the nonlocal interactions between pairs of residues, on the other hand, has been completely simplified. It has been reduced to the consideration of only hydrophobic, hydrophilic, and charged residues, according to their positions in the sequence.

Beginning from random conformations of separated chains, the algorithm is able to provide coiled coil structures, following rather plausible folding pathways. These pathways show a continuous spectrum of intermediate states, in contrast to the all-or-none transition one usually gets in globular proteins, at least of small to medium size. The final structures reproduce, under appropriate conditions, the left-handed super twist shown by real coiled coils. This seems to be related to the geometry of the hydrophobic side chains that stabilize the dimer structure. The stability of the formed structures is related to the chain length. The longer the sequence, the larger is the stability of the coiled coil.

The introduction of cysteine residues in certain positions of the sequence greatly enhances the stability of the resulting folded conformations. The disulfide bridges avoid the possibility of nonparallel or out of register conformations. In some occasions, though, initial stages of the folding pathways suffer from certain complications, probably related to the configurational stress imposed by the presence of the covalent bond between both chains.

This model represents a first attempt to investigate the folding pathways of a peptide chain dimer with a very simple structure. The results are rather encouraging, and open up the possibility of studying more complicated multimeric assemblies. Of course, many further refinements of the model are necessary. Most importantly, we plan to use a scale for the nonlocal interactions based only on the amino acid nature, and not in its position along the sequence. We have tested the possibility of changing the interactive definition of Glu in position **b** (Fig. 1) as a charged, instead of hydrophilic, residue (the consideration of Ala in position **c** as a hydrophilic residue is probably less severe). The coiled coil conformation is still the outcome in most of the trajectories, though there are serious problems to obtain it in a reasonable amount of computer time for 36 amino acid chains. It seems that the repulsion between charged groups with the same charge, close in space, produces a bending of the chain. The pitch of the left-handed super twisting is considerably reduced, and would probably make it very difficult for longer chains to be folded. Therefore, more specific and refined scales for the nonlocal interactions, dependent on the exact nature of the residues and not only on their "class," are required. Work

is in progress in our group to explore this possibility and its application to more realistic sequences appearing in fibrous proteins, or in coiled coil fragments of globular ones.

ACKNOWLEDGMENTS

Most of this work was performed while A. Rey was at The Scripps Research Institute. The support of a Postdoctoral MEC/Fulbright Scholarship is gratefully acknowledged. We thank the referee for his very helpful suggestions, and his great insight on the experimental and theoretical aspects of coiled coils research. This work was supported in part by Grant No. GM-38794 from the Division of General Medical Sciences, National Institutes of Health (USA), and by Grant No. PB92/0227, from DGICYT (Spain).

[1] C. L. Brooks, M. Karplus, and B. M. Pettitt, Adv. Chem. Phys. **71**, 1 (1988).
[2] C. R. Matthews, Curr. Opin. Struct. Biol. **1**, 28 (1991).
[3] C. Cohen and D. A. D. Parry, Trends Biochem. Sci. **11**, 245 (1986).
[4] C. Cohen and D. A. D. Parry, Proteins: Struct. Funct. Genet. **7**, 1 (1990).
[5] J. A. Talbot and R. S. Hodges, Acc. Chem. Res. **15**, 224 (1982).
[6] N. E. Zhou, B.-Y. Zhu, C. M. Kay, and R. S. Hodges, Biopolymers **32**, 419 (1992).
[7] A. Holtzer, R. Clark, and S. Lowey, Biochemistry **4**, 2401 (1965).
[8] E. Woods, Biochemistry **8**, 4336 (1969).
[9] R. S. Hodges, A. K. Saund, P. C. S. Chong, S. A. St.-Pierre, and R. E. Reid, J. Biol. Chem. **256**, 1214 (1981).
[10] S. Y. M. Lau, A. K. Taneja, and R. S. Hodges, J. Biol. Chem. **259**, 13253 (1984).
[11] M. E. Holtzer, A. Holtzer, and J. Skolnick, Macromolecules **16**, 173 (1983).
[12] A. Holtzer and M. E. Holtzer, Biopolymers **30**, 1231 (1990).
[13] J. Mo, M. E. Holtzer, and A. Holtzer, Proc. Natl. Acad. Sci. USA **88**, 916 (1991).
[14] J. Mo, M. E. Holtzer, and A. Holtzer, Biopolymers **31**, 1417 (1991).
[15] J. Skolnick and A. Holtzer, Macromolecules **15**, 303 (1982).
[16] J. Skolnick, Macromolecules **18**, 232 (1984).
[17] J. Skolnick, Macromolecules **19**, 1153 (1986).
[18] A. Holtzer, M. E. Holtzer, and J. Skolnick, in *Protein Folding*, edited by L. M. Gierasch and J. King (American Association for the Advancement of Science, Washington, DC, 1990).
[19] M. Nilges and A. T. Brünger, Proteins: Struct. Funct. Genet. **15**, 133 (1993).
[20] A. Kolinski and J. Skolnick, J. Chem. Phys., **97**, 9412 (1992).
[21] A. Kolinski, A. Godzik, and J. Skolnick, J. Chem. Phys. **98**, 7420 (1993)
[22] A. Rey and J. Skolnick, Proteins: Struct. Funct. Genet. **16**, 8 (1993).
[23] A. Rey and J. Skolnick, Chem. Phys. **158**, 199 (1991).
[24] P. Prevelige and G. D. Fasman, in *Prediction of Protein Structure and the Principles of Protein Conformation*, edited by G. D. Fasman (Plenum, New York, 1989).
[25] J. Garnier and B. Robson, in *Prediction of Protein Structure and the Principles of Protein Conformation*, edited by G. D. Fasman (Plenum, New York, 1989).
[26] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).
[27] M. Sippl, J. Mol. Biol. **213**, 859 (1990).
[28] A. D. McLachlan, M. Stewart, and L. B. Smillie, J. Mol. Biol. **98**, 281 (1975).
[29] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1989).
[30] H. S. Chan and K. A. Dill, Annu. Rev. Biophys. Chem. **20**, 447 (1991).
[31] M. E. Holtzer, A. Holtzer, and J. Skolnick, Macromolecules **16**, 462 (1983).