Nucleation theory for helix unfolding in peptide chains

Ariel Fernández[1,2] and Andrés Colubri[1]
[1]*Instituto de Matemática, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional del Sur, Avenida Alem 1253, Bahía Blanca 8000, Argentina**
[2]*Department of Chemistry and The James Franck Institute, The University of Chicago, Chicago, Illinois 60637*
(Received 6 April 1999)

This work introduces a microscopic nucleation theory of helix unfolding in peptide chains aimed at obtaining a semiempirical estimation of the critical-size bubble of structural distortion which may function as the kernel for helix destruction. A dynamic nucleation model for helix-coil transition has been previously introduced as an ansatz to estimate the kinetic barrier of the helix-unfolding event [A. Fernández and A. Colubri, J. Math. Phys. **39**, 3167 (1998)]. However, the critical size of the helix-destruction bubble, empirically obtained from computer simulations of favored folding pathways, has not been hitherto justified or determined from first principles. This requires introducing a microscopic treatment of the long-time torsional dynamics to assess its bearing on the formation of structural-distortion bubbles which eventually trigger the helix-unfolding process. To reach this goal we introduce two operational tenets: (a) The torsional dynamics of the chain may be coarse grained according to a discretization of the conformational state of each unit, resolved according to its significant torsional isomers; (b) the semiempirical formulation accounts for the known dependence of the enthalpy increment due to helix unfolding on the change in the effective solvent-exposed surface area. The functional dependence on bubble size of the mean time of completion of the rate-determining step for helix unwinding is shown to be in agreement with previous *ad hoc* macroscopic models. However, in contrast with such treatments, we infer the existence of a denaturation temperature from the dynamics of critical bubble formation, rather than introducing it as an *a priori* postulate. Our determination of the critical temperature based on nucleation kinetics theory of critical bubble formation coincides with those obtained from calorimetric and spectroscopic measurements. [S1063-651X(99)11210-8]

PACS number(s): 87.15.He, 87.10.+e, 87.15.Nn

I. INTRODUCTORY REMARKS AND MOTIVATIONS

Helix-Coil transition phenomena in peptide chains are fairly well understood from a static or thermodynamic perspective, validated at a mesoscopic level of description [1–4]. However, we still lack a kinetic treatment rooted in an analysis of the long-time torsional dynamics of the chain and designed to identify the nucleation or rate-determining step for helix unwinding. A major stumbling block in such a treatment is the dearth of data on the actual potential energy surface for a folding chain, from which crucial information could be drawn on the activation energies [5–7] for the formation of the helix-destruction kernel or critical bubble. This structural distortion bubble has been inferred to occur at the ends of the helix with the highest probability [7].

In recent simulations making use of a semiempirical microscopic model of chain torsional dynamics, we have adopted the nucleation ansatz to reproduce the kinetics of helix-dismantling events [6], and ultimately generate experimentally-probed folding pathways. Such an ansatz requires an assumption on the critical size of the helix bubble warranting the cascade of destructive events that follows after nucleation and leads to a complete dismantling of the helix. Empirically, we have found the critical size $L=L^*$ to be approximately one third of N, the length of the helical segment [6].

This is not the only result that needs rigorous validation: The exponential dependence on L of the mean time of formation, $t(L)$, of an L bubble has been previously established [7] using an *ad hoc* model which does not make use of either a geometric or a topological representation of the peptide backbone to account for the ocurrence of local structural distortions. Thus, a treatment of torsional dynamics should validate this functional dependence as well.

Moreover, the previous generic model postulates the existence of a denaturation temperature $T=T^*$ as an *a priori* assumption [7], instead of deducing its existence from first principles. Thus, the model of de Gennes introduces a bubble-expansion force which sets in as $T\geqslant T^*$, and becomes a helix-restoring force when $T<T^*$. In contrast with this model which hinges upon the *ad hoc* assumption of existence of T^*, we shall infer its existence and accurately predict its value from the torsional dynamics and its bearing upon the formation of kernels triggering the helix unfolding.

Following these remarks, the need arises for theoretical underpinnings of three basic facts: (a) $L^*\approx(1/3)N$; (b) $t(L)\sim\exp(KL)$, with $K=$constant; and (c) existence of a denaturation T resulting from a sharp qualitative difference in the nucleation dynamics of both helix creation and helix destruction. This is precisely the aim of this work, where a semiempirical treatment of the long-time torsional dynamics of the chain will be adopted to serve as the theoretical framework of the nucleation model.

Our goal will be achieved by introducing two basic operational tenets: (a) The torsional dynamics are coarse grained according to a ''cis-trans'' (or compact-extended) rotameric description of the conformation of each individual

*Present and permanent address.

residue [8]; (b) the semiempirical formulation must account for the dependence of the enthalpy increment due to helix unfolding on the change in the effective solvent-exposed surface area [3,9]. Finally, the local torsional constraints obtained from tenet (a) are imposed onto the mesoscopic level of description of the dynamics governed by tenet (b) in order to obtain an operational framework to derive the critical extent of structural distortion responsible for the destruction of the helix and the T-dependent time required to reach such a distortion.

The outline of the work is as follows: Sec. II is devoted to the semiempirical construction of a coarse conformation manifold upon which we may draw a coarse nucleation rate theory for the folding/unfolding of an oligopeptide, later specialized to the helix-coil transition context. Section III deals with the estimation of the critical bubble size L^* using a coarse nucleation rate theory. Section IV is concerned with the nucleation dynamics, establishing the existence of a denaturation temperature at which the formation of the critical bubble becomes feasible within physically-relevant timescales.

II. COARSE TORSIONAL CONFORMATION MANIFOLD: AN OPERATIONAL REPRESENTATION

As stated above, our primary goal is to develop a semiempirical dynamic nucleation theory of helix unfolding by taking into account the long-time limit of torsional dynamics. The vast timescale gap between free torsions ($\sim$10 ps) and bubble formation within a helix ($\sim 1\ \mu$s to seconds) requires a suitable simplification of conformation space [6,8,10] which nevertheless captures the main features of the dominant torsional fluctuation ultimately responsible for the mesoscopic transition. Precisely at the mesoscopic level, the concept of intrachain contact must be operationally introduced *vis-a-vis* the maximum distance ($\sim$7 Å) for a meaningful long-range interaction with associated decrease in heat content larger or equal to $1/2k_BT$ (k_B=Boltzmann constant). Each contact pattern (CP) of the chain represents a set of constraints applied on the residues engaged in the contacts (those in proximity as well as those forming the concurrent loops and turns), while free residues may adopt a discrete number of rotameric forms. Thus, the relative entropic content of each CP may be easily determined by coarse-graining the possible local torsional states.

To specify the fine structure of a CP-basin, we shall start by discretely codifying the torsional states of a peptide chain according to the local constraints to be fulfilled in order to form each possible CP. Thus, we shall discretize the soft-mode torsional dynamics of the chain and, following standard procedures, integrate out as conformational entropy [6] the hard-mode dynamics–vibrational and angular stretching–which operates on incommensurably faster (10^{-15}–10^{-11} s) timescales. The next step is to resolve each CP-basin in the potential energy surface (PES) as an ensemble of substates [10–12]. Each substate represents one–possibly many–of the discretized torsional states of the chain realizing the CP. At this point, we may introduce an ''adiabatic ansatz'' valid in the long-time limit of torsional dynamics if and only if thermal equilibration within CP-basins is incommensurably faster than transitions between CP's. This approximation, whose domain of validity will be confirmed in this work, represents an entrainment of the dynamics to the folding pathway coarsely resolved at the CP level.

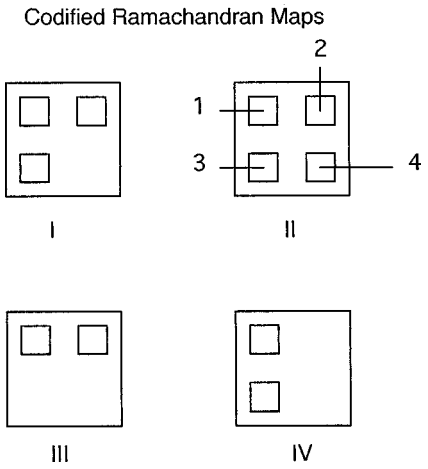

FIG. 1. Discrete codification of local torsional states of aminoacids (residues) by indicating the basin (1, 2, 3, or 4) in the Ramachandran map where the torsional coordinates Φ, Ψ lie. There are four types of maps I–IV, depending on whether the residue is L-alanyl-like (I), glycine (II), precedes a proline (III), or proline (IV). Thus, a Ramachandran discrete variable $R(\mathbf{y},n)=1,2,3,4$, indicates the basin for the nth residue in the conformation roughly defined by the contact pattern $\mathbf{y}$.

In order to define a coarse version of the PES of real operational value, we start by defining the coarse conformation manifold for a discretized version of the soft-mode dynamics. To reach this goal, we codify the local torsional state of each residue (chain unit) according to the basin of attraction where its two torsional dihedral coordinates Φ, Ψ lie within the local potential energy surface. This local surface, known as the Ramachandran plot [2], consists of a map of the 2-torus, or local (Φ,Ψ)—coordinate space, onto the energy real line. This local surface has a finite and small number of basins of attraction of local minima, and its topography depends on the kind of residue at the particular contour position along the chain. As indicated in Figs. 1 and 2, we may regard the torsional conformation of each reside modulo Ramachandran basins, and this simplification begets a dis-

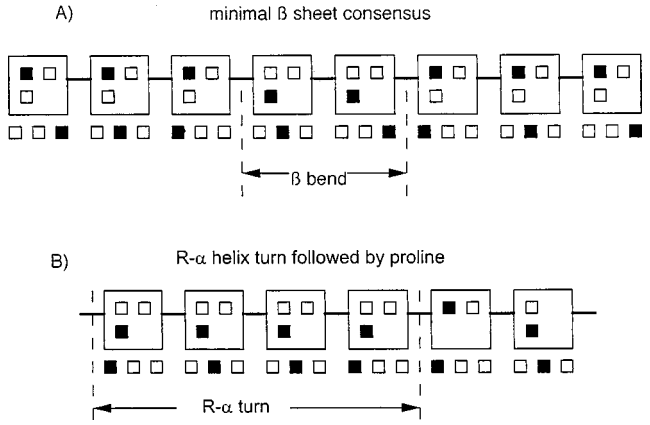

FIG. 2. (a) Consensus window for a minimal β-sheet structural motif. (b) Consensus window for a right-handed α-helix turn interrupted by a proline (and a residue preceding proline).

cretization of the torsional dynamics of the chain whose torsional state may be now given in a binary matrix form. In simple terms, we are replacing detailed geometric information by a coarse topological representation.

In turn, this discretization is possible because equilibration within Ramachandran basins is incommensurably faster than inter-basin transitions [6,8]. In essence, the discretized Ramachandran map coarsely governs the local torsional dynamics of a single aminoacid residue of the peptide chain. Such dynamics are not correlated with those of nearest neighbor residues due to the torsional rigidity of the backbone bond (CO—NH) linking adjacent residues [2]. Furthermore, the general topological features of each Ramachandran map are preserved during the folding or unfolding process [1,2,6], that is they are not essentially distorted (no basins are fused or created) by long-range interactions developed. These considerations lead us naturally to define a ''Ramachandran variable,'' denoted $R(\mathbf{y},n)$, which indicates the basin of attraction of the two dihedral torsional coordinates for residue n in substate $\mathbf{y}$.

In our codification of the local torsional dynamics we classify residues or aminoacids as follows: *L*-alanyl-like, glycine, proline, and any residue preceding proline (Fig. 1). Thus, each residue is classified essentially according to the topology of its Ramachandran map, and there are four such topologies. Thus, since an *L*-alanyl-like residue (the most common kind of aminoacid in the protein sequence) with contour position n has three basins of attraction [2], we would get three possible values depending on $\mathbf{y}$:$R(\mathbf{y},n)=1,2,3$, while if glycine is at the nth position, we would get $R(\mathbf{y},n)=1,2,3,4$, again depending on $\mathbf{y}$. On the other hand, if proline happens to be the nth residue, we would get $R(\mathbf{y},n)=1,3$, while if the nth residue precedes proline, we would obtain $R(\mathbf{y},n)=1,2$. This codification is consistent with the existence of local torsional isomers coarsely represented as basins of attraction in the Ramachandran plots. Generically speaking, we provide a consistent dynamical picture in which long-range intra-chain interactions are determined by a set of local torsional constraints whose fulfillment is demanded by the need to avoid frustrated conformations containing torsional incongruities (for instance, an extended local conformation within a loop, a β-bend or an α-helix turn; or a compact local conformation within one strand of a β-sheet).

Following these tenets, the discretized torsional dynamics within a CP-basin may be computed according to the following formal scheme:

(a) We introduce a ternary variable $G(n)=1,2,3$ indicating respectively whether the nth residue along the chain is hydrophobic, neutral or hydrophilic (polar).

(b) We determine the type of Ramachandran plot (I–IV), as indicated in Fig. 1, for each residue $n=1,...,N$.

(c) We define substate $\mathbf{y}$ by two rows $\{R(\mathbf{y},n),G(n)\}_{n=1,...,N}$, as illustrated in Fig. 2. Thus, $R(\mathbf{y},n)=1$ indicates that the nth residue has adopted the extended conformation compatible with a β-sheet; $R(\mathbf{y},n)=2$ indicates that either the nth residue has adopted a locally compact conformation compatible with a β-bend (zero pitch), or with a left-handed α helix; finally, $R(\mathbf{y},n)=3$ indicates that the conformation of the nth residue is compatible with the formation of a β-bend or with a right-handed α helix [2].

(d) We determine the CP-basin to which a particular torsional substate belongs by identifying consensus windows in the matrix representation of the chain conformation. By consensus we simply mean a region of the chain where the local topological constraints associated to the formation of a particular folding or structural element or motif are fulfilled. In other words, all residues belonging to the consensus window are in the ''correct'' torsional state (or Ramachandran basin) to form the particular pattern. In this way, a consensus window emerges as a pattern of local structural signals encoded in some portion of the aminoacid sequence. The broad latitude (from 30° up to 60°, [2]) in the possible values of the local torsional coordinates within the Ramachandran basin, and the vast structural distorsion it leads to, implies that the discrete codification cannot be implemented at the geometric level. Rather, the inter-basin transitions are meant to mimic changes in the local topological constraints to which the flexible chain is subject in order to reach specific structural patterns.

(e) The dominant secondary structure motifs can be identified as recognizable patterns in the substate $\mathbf{y}$. Thus, the right-handed α-helix requires a window of residues with $R(\mathbf{y},n)=3$. Without loss of generality and for the sake of notation, we shall identify this motif by a window in $\mathbf{y}$ with $R(\mathbf{y},n)=3$ and a periodic $G(n)=1=G(n+3)$ or $G(n)=1=G(n+4)$ hydrophobicity (Fig. 2). Because of the highly helix-disruptive tendencies of glycine [2], if its local diagram (of type II) appears within a consensus window, the entire helix turn containing glycine is obliterated from the CP. The disrupting tendencies of the proline, on the other hand, do not require special instructions as the $R(\mathbf{y},n)=3$ value cannot occur for a residue preceding proline, as shown in Figs. 1 and 2. A consensus window translating into the (R) α-helix is indicated in Fig. 2. Similarly, for a left-handed α-helix, we must demand permanence of $R(\mathbf{y},n)=2$ value, while retaining all other conditions regarding hydrophobic periodicity along the chain.

Likewise, being pleated structures, β-sheets are characterized by the persistence of the extended local conformation basin marked by $R(\mathbf{y},n)=1$. In order to fulfill hydrophobic/polar compatibilities, the G-values must be preserved in a parallel or antiparallel fashion, depending on the relative orientation of the strands in the β-sheet (Fig. 2).

Turns and bends may be a determinant of the β-sheet or simply required to form hydrophobic contacts, thus they will be treated generically, regardless of whether or not they realize β-sheet topologies. Should such motifs require closure of chain loops, they would require a $R(\mathbf{y},n)=2$ or $R(\mathbf{y},n)=3$ consensus window.

A description of the fine structure of CP-basins *vis-a-vis* our discretized codification of torsional states requires that we compute the substate multiplicity $\Omega(i)$ of each CP(i), that is, the number of possible substates translatable into CP(i). Thus, we obtain:

$$\Omega(i)=\prod_{n=1,...N} q(i,n), \tag{1}$$

where $q(i,n)$ indicates the number of possible values of the

nth Ramachandran variable $R(\mathbf{y},n)$ for all substrates $\mathbf{y}$'s translatable into CP(i). In addition, we have: $q(i,n)=1$, if the nth residue is engaged in a structural element recorded in CP(i); or $q(i,n)=2$, 3, or 4, if the nth residue is *not* engaged in any structural element of CP(i), *and* the nth residue is proline or a residue preceding proline $[q(i,n)=2]$, the nth residue is

$$\text{alanyl-like } [q(i,n)=3],$$

$$\text{or the } n\text{th residue is glycine } [q(i,n)=4]. \quad (2)$$

III. COARSE RATE THEORY OF NUCLEATION FOR HELIX UNFOLDING

This section is devoted to the development of a nucleation rate theory rooted in the coarse description of the torsional conformation manifold given in Sec. II. Due to the essentially enthalpic nature of the helix-dismantling kinetic barriers [6], two paramount thermodynamic parameters of the helix unfolding process must be estimated: The enthalpy change, denoted $-\Delta H(L^*)$, associated to critical bubble formation within an N-helix, and the enthalpy change, denoted $-\Delta H(N-L^*)$, involved in the full unfolding of the remaining helix after the critical intermediate has formed. It should be emphasized that our system consists in an ensemble of protein molecules not encompassing the solvent, although our treatment thoroughly incorporates its effect upon the system.

This section is divided into two subsections, the first dealing with the estimation of relevant thermodynamic parameters, while the second deals with the coarse rate theory.

A. Exposed surface area and bubble formation

Considerable efforts to estimate thermodynamic folding parameters have revealed that the enthalpy change in a folding event should be related to changes in the solvent accessible surface area. This line of thought was pioneered by Sinanoglu and Fernández [9], reviewed in [2], and has been recently resurfaced [3], thus it will only be sketched here. Let us first compute the quantity $\Delta H(N)$, the enthalpy change of formation of an N-helix. For simplicity, we shall assume that all residues are of the most common L-alanyl-like type. The length N may be written as $N=r\cdot w$, where r is the number of units per turn and w is the number of turns (r is assumed fixed for a given helix, while $r=3.6$ when averaged over a statistical ensemble of helices [2]). In order to incorporate the packing effect at a mesoscopic level, we may adopt a geometric representation of stacking by regarding residues as cylinders linked as indicated in Fig. 3. The actual geometry of residues is immaterial, as only the effective changes in solvent-accessible areas become relevant. In agreement with [9], the variation of cavity free energy associated to the overall folding event leading to N-helix formation is

$$\Delta G(N)\approx\gamma\Delta s(N). \quad (3)$$

Here γ denotes the microthermodynamic surface tension operating down to molecular dimensions, as determined by Sinanoglu [13] and measured chromatographically by Horvath and coworkers [14]; and $\Delta s(N)$ denotes the effective change

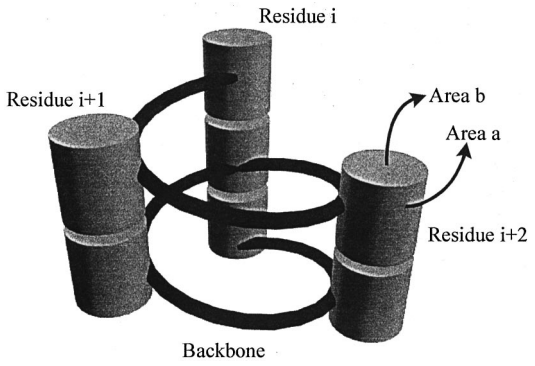

FIG. 3. Schematic representation of the packing of an α-helix in order to minimize the effective solvent-exposed surface area. The actual geometry of each residue is immaterial within the context of the theory, only the parameter b becomes relevant.

in exposed surface area. For fixed w, the change Δs may be evaluated as $\Delta s=w\Delta s'$, where $\Delta s'$ is the surface area change for a single pile in the helix, as depicted in Fig. 4. Since $\Delta s'=-2b(w-1)$, where $-2b$ is the loss in exposed area per residue, we get

$$\Delta G(N)=-2\gamma bN+2\gamma br. \quad (4)$$

On the other hand, using Eq. (1), the entropic change associated to constraining N residues to their structurally-relevant Ramachandran basin is

$$\Delta S(N)=R\ln[\Omega(\text{random coil})/\Omega(N\text{-helix})]=R\ln 3^{-N}$$
$$=-NR\ln 3. \quad (5)$$

where Ω represents the number of microscopic realizations of a given structural pattern. From Eqs. (4) and (5) we get

$$\Delta H(N)=-(2\gamma b+RT\ln 3)N+2\gamma br. \quad (6)$$

To estimate the parameter $2\gamma b$ we may take into account the nucleation event in the zipper model of helix formation [2]. Since the helix formation kernel consists in a single piling of two residues obtained after a single turn, we get

$$\Delta G(\text{nuc.})=-2\gamma b=h+(r+1)RT\ln 3, \quad (7)$$

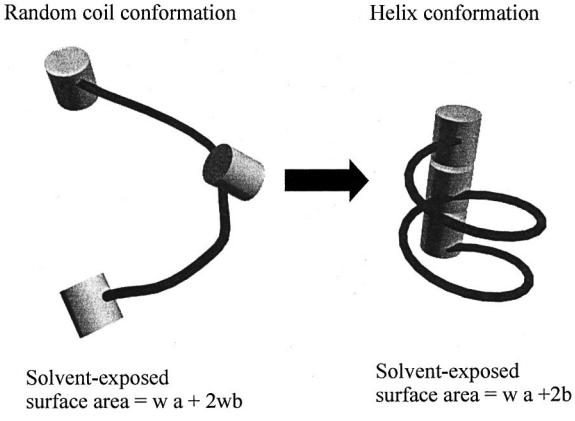

FIG. 4. Change in solvent-exposed surface area associated to the formation of a single pile within the helix.

TABLE I. Temperature dependence of the stability of the helix-formation kernel, as measured by $\Delta G(\text{nuc.})$, the change in free energy associated to forming a single helix turn.

T [°K]	ΔG_{nuc} [Kcal/mol]
298	−0.12
303	−0.07
308	0.02
313	0.03
318	0.08
323	0.13
328	0.18
333	0.23
338	0.28
343	0.33
348	0.38

where $\Delta G(\text{nuc.})$ represents the free energy change associated to the nucleation event, which is typically very small, of the order of 1/10 kcal/mol [15]. On the other hand $\Delta H(\text{nuc.}) = h \approx -3.1$ kcal/mol. [15]. Equation (7) allows us to infer the existence of a denaturation temperature, taken to be the value at which $\Delta G(\text{nuc.}) \approx 0$, and beyond which no nucleation kernel for helix formation becomes stable. The values given in Table I reveal that $T^* \approx 313\,°\text{K}$. A consistent theory should yield and identical value for the denaturation temperature regarded as the temperature at which the critical size bubble forms expeditiously within physically-relevant timescales. This is indeed the case, as shown in the next section. Furthermore, the value given is in perfect agreement with that obtained from calorimetric measurements of denaturation criticality in proteins of comparable size [16].

B. Coarse nucleation rate theory

In order to determine the rate-limiting step in the helix unfolding transition, we assume that a bubble of length L in a helix of contour length N forms at one extremity [7] associated to the contour region $[N-L+1,N]$. As we know, such bubbles may easily reverse back to the helical conformation [2] unless they reach a critical size $L=L^*$, which has been empirically estimated to be about 30% N. In the latter case, the bubble becomes equally prone to expansion along the helix, triggering the helix coil transition. In order to compute L^*, we must determine the Arrhenius transition rates $k(-,L)$ and $k(+,L)$, where $k(-,L)$ denotes the unimolecular rate of dismantling the remaining helix with contour region $[1,N-L]$, and $k(+,L)$ denotes the unimolecular rate of full helix recovery, with both rates determined relative to the L bubble as the starting point. Thus, L^* must satisfy the equation:

$$k(-,L^*) = k(+,L^*). \tag{8}$$

The kinetic barrier $B(+,L)$ associated to helix recovery (or L-bubble destruction) is of entropic origin [6,10], since L units must be frozen in a single Ramachandran basin to restore the $[N-L+1,N]$-helix. Thus, we obtain:

$$B(+,L) = -T\Delta S(+,L) = LRT \ln 3. \tag{9}$$

TABLE II. The critical bubble size L^* as a function of the contour length N of the helix.

N	L^*
10	5.47
20	8.58
30	11.83
40	15.13
50	18.45
60	21.79
70	25.13
80	28.47
90	31.84
100	35.2
1000	339.86

On the other hand, the pre-exponential factor $f(+,L)$ to $k(+,L)$ is

$$f(+,L) = 3^{-L} \times 10^{11}\,\text{s}^{-1} \tag{10}$$

where $f = 10^{11}\,\text{s}^{-1}$ is the mean free torsional frequency of a (Ψ,Φ) degree of freedom [17], and 3^{-L} is the probability that L discretized torsions adopt simultaneously the helical Ramachandran basin from within three possible choices. Thus, we get

$$k(+,L) = f(+,L)\exp[-B(+,L)/RT] = 3^{-2L} \times 10^{11}\,\text{s}^{-1}. \tag{11}$$

Now we must compute $k(-,L)$. As indicated in Sec. III A, we get $B(-,L) = -\Delta H(N-L)$. On the other hand, the pre-exponential factor $f(-,L)$ becomes now proportional to $2(N\text{-}L)$, the sum of all torsional possibilities which are successful in shifting any of the $(N\text{-}L)$ local conformations to one of the two Ramachandran basins different from the one compatible with the helix. Thus, we get

$$f(-,L) = 2(N\text{-}L) \times 10^{7}\,\text{s}^{-1} \tag{12}$$

where $10^7\,\text{s}^{-1}$=torsional frequency for a residue engaged in an α-helix [6].

Using Eqs. (6), (9)–(12) we can translate Eq. (8) into a working equation to determine the critical size L^* of the helix-destruction kernel at temperature T as:

$$\begin{aligned} &3^{-2L^*} \times 10^{11}\,\text{s}^{-1} \\ &\quad = 2(N-L^*) \times 10^{7}\,\text{s}^{-1} \\ &\qquad \times \exp\{[-(2\gamma b + RT\ln 3)(N\text{-}L^*) + 2\gamma br]/RT\}. \end{aligned} \tag{13}$$

Equation (13) is transcendental and thus it cannot be effectively manipulated algebraically: it will be solved numerically, as indicated in Sec. IV.

IV. RESULTS

The numerical solution of Eq. (13) obtained at a typical renaturation temperature $T = 308\,°\text{K}$ is displayed in Table II. Direct inspection of the results reveal the rapidly converging

TABLE III. Temperature dependence of the critical bubble size and of the time $t(L^*)$ of formation of the critical bubble. The results were obtained for $N=60$.

T [°K]	L^*	$t(L^*)$ [s]
298	23.49	18 484.5
303	22.65	1229.24
308	21.79	98.16
313	20.92	9.36
318	20.03	1.06
323	19.12	1.4×10^{-1}
328	18.2	2×10^{-2}
333	17.26	4×10^{-3}
338	16.31	8.7×10^{-4}
343	15.34	2.2×10^{-4}
348	14.35	6.3×10^{-5}

behavior $L^*\approx(1/3)N$, in virtually perfect agreement with the previous empirical dependence determined from computer simulations of favored folding pathways that yielded a predictive folding algorithm [6].

The actual T-dependence of the critical bubble size is displayed in Table III for $N=60$. This length, in turn, has been adopted since it corresponds to the shortest helix for which the empirically found asymptotic dependence is reproduced to within 2% accuracy (cf. Table II). As expected, the critical size of the bubble gets smaller as the temperature increases, implying that thermal fluctuations make a smaller bubble prone to expansion within the helix, while the same bubble would not trigger the helix unfolding at lower temperatures.

The actual timespan of the helix unfolding process is kinetically determined by the rate-determining step, that is, the formation of the critical kernel or bubble. According to the coarse nucleation theory developed in Sec. III, this timespan is $t(L^*)$, the time it takes to form the critical bubble, which is given by:

$$t(L^*)=(2L^*)^{-1}\times10^{-7}\ \mathrm{s}\times\exp\{[(2\gamma b+RT\ln 3)L^*-2\gamma br]/RT\}. \tag{14}$$

The dominant exponential dependence, $t(L^*)\sim\exp(KL^*)$, with $K=(2\gamma b+RT\ln 3)$ is functionally in agreement with previous generic estimations rooted in ad-hoc mesoscopic models [cf. Ref. [7], Eq. (57)]. However, in contrast with such kinetic models that introduce *a priori* the existence of a denaturation temperature, our nucleation theory does not necessitate of this assumption: *The existence of the denaturation temperature is grounded in the kinetics of formation of the helix-destruction kernel itself*. Inspection of Table III reveals a sharp drop of over five orders of magnitude in $t(L^*)$ as the temperature is raised from 298 °K to 323 °K, while the actual formation of the critical bubble becomes accessible within meaningful over-all folding times (~10 s for $N=60$ [6]) only at $T=T^*=313$ °K.

This high sensitivity of $t(L^*)$ on T is responsible for the qualitatively dramatic effect, characterized dynamically by the feasibility of helix unfolding within relevant timescales. This fact explains the existence of a denaturation temperature from the perspective of nucleation kinetics. Furthermore, by contrasting Tables I and III, we can establish the consistency of our theory, since *precisely the same physically realistic temperature that makes it possible to form a helix-destruction bubble* ($T=T^*=313$ °K) *is the temperature that makes the kernel for helix formation (a single turn) unstable*. Thus the critical temperature is unambiguously identified and the coincidence of both inferences (cf. Tables I and III) determines the validity of our approach.

Furthermore, our theoretical estimate is in solid agreement with experimental determinations of the denaturation temperature of proteins of comparable size, and is identical to the critical value obtained from calorimetric measurements by Privalov and coworkers [16].

V. DISCUSSION

In contrast with previous thermodynamic approaches or *ad hoc* kinetic models, this work introduces a kinetic theory of nucleation for helix-coil transition phenomena rooted in a coarse microscopic description of torsional dynamics in the long-time limit. The theory makes use of a discrete codification of the local torsional states of the chain, while casting the entropically-driven solvophobic packing of the helix in terms of the microthermodynamic surface tension. In this way, the microscopic level of description, which allows us to determine the local torsional constraints, is combined with a mesoscopic level required to account for nonlocal interactions. This is effectively carried out by transferring the local torsional constraints to the coarser level upon which surface tension effects become meaningful.

The theory does not *assume* the existence of a denaturation temperature: *It deduces its existence from first principles and yields accurate critical values*. The critical temperature is obtained from dynamic considerations: It is defined as the lowest temperature at which the helix-destruction kernel or bubble is formed within physically-relevant timescales (less than the overall folding timescales), while the nuclear helix turn required to initiate the cascade of helix-growth events becomes unstable.

Besides accurately predicting the existence and concrete value of the denaturation temperature, the theory validates the empirical computational tenet that a bubble with size of about one third of the total contour length of the helix becomes a critical helix-destruction kernel [6]. Finally, our theory is shown to be consistent with previous *ad hoc* mesoscopic models in which the existence of a denaturation temperature is postulated, as opposed to being inferred from the basic tenets themselves [7].

While the stability of the helix-creation kernel (a single turn) is obviously independent of the length of the helix, the feasibility of formation of the critical bubble is strongly N-dependent. This observation leads us to generically define two [$T^*(c.)$ and $T^*(d.)$] critical values instead of one (T^*), a distinction which becomes more and more apparent in the limit of long chains $N>100$. The critical temperature $T^*(c.)$, beyond which the creation kernel becomes unstable is in general lower than the critical temperature for helix destruction $T^*(d.)$. At $55\leqslant N\leqslant 60$, $T^*(c.)$ and $T^*(d.)$ are within half a degree difference and therefore a single value $T^*=T^*(c.)=T^*(d.)$ is identifiable given the experimental

uncertainty of ~ 1 degree [16–18]. However, for $N>60$, a metastable folding phase of ''slow denaturation'' emerges for T within the region $T^*(c.)<T<T^*(d.)$. This ''slow denaturation'' regime has been identified in spectroscopic–and obviously not in thermodynamic–measurements [18]. According to our theory, it corresponds to the slow formation of the critical bubble which, as demonstrated in this work, takes times incommensurably longer than overall folding times for $T<T^*(d.)$ (cf. Table III). Furthermore, however long it may take to form the destruction kernel, there is no possibility of restoring the helix because the formation kernel is unstable for $T>T^*(c.)$: *The helix cannot be recovered once dismantled, however slow the latter process might be*. In forthcoming work we shall attempt to construct a phase diagram to characterize the kinetic metastability regime $T^*(c.)<T<T^*(d.)$ as a function of chain length.

ACKNOWLEDGMENTS

A.F. is principal investigator of the National Research Council of Argentina (CONICET). Financial support from the J. S. Guggenheim Foundation and the Alexander-von-Humboldt Stiftung is gratefully acknowledged. The kind hospitality of Professor R. Stephen Berry during A.F.'s stay in Chicago is gratefully acknowledged.

[1] Y. M. Bai and S. W. Englander, Proteins: Struct., Funct., Genet. **24**, 145 (1996).
[2] C. Cantor and P. Schimmel, *Biophysical Chemistry* (Freeman, New York, 1980), Vols. I–III.
[3] V. J. Hilser, J. Gomez, and E. Freire, Proteins: Struct., Funct., Genet. **26**, 123 (1996).
[4] J. A. Daquino, J. Gomez, V. J. Hilser, K. H. Lee, L. M. Amzel, and E. Freire, Proteins: Struct., Funct., Genet. **25**, 143 (1996).
[5] G. Schwartz, J. Mol. Biol. **11**, 64 (1965).
[6] A. Fernández and A. Colubri, J. Math. Phys. **39**, 3167 (1998).
[7] P. G. de Gennes, J. Stat. Phys. **12**, 463 (1975).
[8] J. G. Saven and P. G. Wolynes, J. Mol. Biol. **257**, 199 (1996).
[9] O. Sinanoglu and A. Fernández, Biophys. Chem. **21**, 167 (1985).
[10] A. Fernández and A. Colubri, Physica A **248**, 336 (1998).
[11] K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[12] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[13] O. Sinanoglu, *Molecular Interactions*, edited by H. Ratajczak and W. J. Orville-Thomas (Wiley, New York, 1982), Vol. 3, p. 284.
[14] C. Horvath, W. Melander, and I. Molnar, J. Chromatogr. **125**, 123 (1976).
[15] D. A. Brant, Macromolecules **1**, 291 (1968).
[16] P. L. Privalov, *Protein Structure and Protein Engineering*, edited by E. L. Winnacker and R. Huber (Springer, Berlin, 1988), pp. 6–15.
[17] C. Brooks, III, M. Karplus, and B. Montgomery Pettitt, *Proteins: A Theoretical Perspective on Dynamics, Structure and Thermodynamics*, Advances in Chemical Physics, Vol. LXXI (Wiley, New York, 1988).
[18] A. Ginsburg and W. R. Carroll, Biochemistry **4**, 2159 (1965).