Victor Muñoz[1,*]
Luis Serrano
EMBL
(European Molecular Biology Laboratory), Heidelberg D-69012, Germany

Received 16 April 1996; accepted 19 August 1996

Development of the Multiple Sequence Approximation Within the AGADIR Model of α-Helix Formation: Comparison with Zimm–Bragg and Lifson–Roig Formalisms

Abstract: *In this work we present the development of the multiple sequence approximation (AGADIRms) and the standard one-sequence approximation (AGADIR1s) within the framework of AGADIR's α-helix formation model. The extensive comparison between these new formulations and the original one* [*AGADIR; V. Muñoz and L. Serrano (1994)*, Nat. Struct. Biol., *Vol. 1, pp. 399–409*] *indicates that the standard one-sequence approximation is virtually identical to the multiple sequence approximation, while the previously used residue partition function approximation* [*Muñoz and Serrano (1994); (1995)*, J. Mol. Biol., *Vol. 245, pp. 275–296*] *is less precise. The calculations of the average helical content performed with AGADIR are precise for peptides of less than 30 residues and progressively diverge from the multiple sequence formulation for longer peptides. The helicity distribution of heteropolypeptides with less than 50% average helical content is also well described, while those of quasi-homopolymers with high helical content tend to be flattened. These inaccuracies lead to an underestimation of 0.017 kcal/mol for the mean-residue enthalpic contribution in AGADIR, as compared to AGADIRms and AGADIR1s. The other energy contributions to α-helix stability are not affected by the original statistical approximation. We also discuss the particularities of the model for α-helix formation utilized in AGADIR and compare it with the classical Zimm–Bragg and Lifson–Roig theories. Moreover, we develop the mathematical relationships between the basic AGADIR energy contributions and helix nucleation and elongation, which permit the quantitative comparison between formalisms. Remarkably, the comparison between AGADIRms and the Lifson–Roig formalism shows that, despite the differences on treating helix/coil cooperativity, both theories give virtually identical results when an equivalent set of parameters is used. This indicates that the helix/coil transition is a solid theory independent of the particularities of the model for α-helix formation.* © *1997 John Wiley & Sons, Inc. Biopoly* ***41:*** 495–509, 1997.

Keywords: *Helix/coil transition; multiple sequence approximation; AGADIR; Zimm–Bragg theory; Lifson–Roig theory*

* *Present address:* Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Building 5, Room 106, Bethesda, MD-20892-0520.

Correspondence to: Luis Serrano, EMBL, Post fach 10.2209, Meyerhofstrasse 1, Heidelberg D-69012, Germany.

CCC 0006-3525/97/050495-15

INTRODUCTION

The physical principles governing α-helix formation have been a popular and important object of study, both theoretically and experimentally. As a consequence, the generic theory describing the physical properties of the transition between the random coil and the α-helix states of a polypeptide was already developed in the late 1950s and early 1960s.[1,2] This theory has proven to be robust on explaining the experimental observations of the helix/coil transition in homopolypeptides.[3] On the other hand, with the development of appropriate peptide model systems and the improvement of detection methodologies, a great amount of experimental information has been obtained from the analysis of short heteropolypeptides in aqueous solution.[4] Some of these findings have revealed that classical helix/coil theories need to be enriched to be able to describe α-helix formation in heteropolypeptides. Following these lines, several authors have developed modifications of the helix/coil transition to include effects that were neglected in the initial formulations. In particular: interactions between side chains located at positions i, $i + 3$ and i, $i + 4$,[5–9] interactions between charged groups, and the helix macrodipole[7] and capping effects.[10] The first attempt to include most of these contributions to α-helix stability in a global model for the helix/coil transition in heteropolypeptides was proposed by Finkelstein and collaborators.[11] Recently, we have presented a similar approach, in which the energy contributions are empirically obtained from previous experimental studies in designed peptides and site-directed mutagenized proteins.[12,13] Calculations carried out with this model, called AGADIR, are consistent with far-uv CD, and nmr results in designed peptides and short protein fragments.[13] The effects of temperature and pH on α-helix stability have lately been included.[14] AGADIR differs from the Lifson–Roig formalism in several aspects of the statistical approximation and of the model for α-helix formation. In this work we present the development of the multiple sequence approximation and the standard one-sequence approximation, within the framework of AGADIR's α-helix formation model. The comparison between AGADIR, its multiple sequence approximation (AGADIRms), and its one-sequence approximation (AGADIR1s) allows us to check the accuracy of the original approximation. We also discuss in depth the model of α-helix formation used in AGADIR, and develop the mathematical relationships between the basic AGADIR energy contributions and the classical helix nucleation and elongation. With these relationships we compare quantitatively AGADIR and Lifson–Roig formalisms, in order to study the putative dependence of the basic helix/coil transition parameters on the specific model of α-helix formation.

RELATIONSHIPS BETWEEN AGADIR THEORY AND ZIMM–BRAGG AND LIFSON–ROIG THEORIES

The model of α-helix formation and theory used in AGADIR depart in some aspects from the classical Zimm–Bragg (ZB) and Lifson–Roig (LR) models. In this section we will discuss the particularities of the AGADIR theory, as compared to ZB and LR, starting by comparing it with ZB, to which is more related. Knowing the relationships between these two models, the comparison of AGADIR with LR is straightforward through the use of the thorough analysis by Qian and Schellman relating ZB and LR theories.[15] AGADIR is equivalent to ZB in two basic aspects: (a) the definition of a minimal number of consecutive residues in helical dihedral angles to constitute a helical conformation; (b) the definition of the coil state of a residue as all the conformational space available, including the helical angles. In other words, the weight for the random coil (z) in ZB and AGADIR is equivalent to the sum of the weights for coil and helix alone ($z = c + v$) in Lifson–Roig.[15] The conformational space of a polypeptide chain is then simply divided into helical conformations (conformations with at least a minimal helical segment) and a single random coil conformation. This point involves an assumption, as such definition of the random coil will include the conformations with more than one residue in helical angles producing an over counting of states. This assumption, however, leads to negligible errors when the nucleation is small, as has been discussed by Qian and Schellman.[15]

On the other hand, AGADIR differs from the classical ZB in four main points: (a) In AGADIR the basic conformational unit is the residue, centered on the Cα atom, and not the peptide bond. This is also related to the two basic parameters to describe the helix/coil transition in AGADIR, which are the entropic cost of fixing the residue in ϕ, ψ helical angles ($h_n = e^{-\Delta G_{\text{intri}}/RT}$) and the mean-residue enthalpic contribution to α-helix formation ($h_e = e^{-\Delta G_{\text{HBond}}/RT}$), previously termed hydrogen bond by means of simplicity.[12,13] (b) The helix conformation is flanked by a N-cap and a C-cap, rather than

Table I

Peptide bonds	1	2	3	4	5	6	7	8	9	10
ϕ,ψ Residues	1	2	3	4	5	6	7	8	9	
ZB coding	0	0	0	0	0	1	1	1	0	0
ZB weights	1	1	1	1	1	σS	S	S	1	1
AGADIR coding	RC	N	H	H	H	H	H	C	RC	
AGADIR weights	1	n	h_n	h_n	h_n	h_n	h_n	h_e	c	1

by conformational units in random coil state. (c) The minimal helix length is four residues in helical angles (five peptide bonds) plus the two caps. (d) the helix/coil cooperativity is modeled by assuming that the first four residues in a helical conformation do not have mean-residue enthalpic contribution, compared to ZB and LR, in which the third residue already has it by virtue of the i, $i + 4$ hydrogen bond. These differences intend a description of α-helix formation more in accordance with present day's knowledge of α-helix stability. The physical reasons to include them in AGADIR will be discussed in detail in the following section. On the other hand, in order to rigorously compare AGADIR and ZB it is necessary to know the relationships between their basic helix/coil parameters. This is easily done by calculating the weight of a helical conformation embedded in a random coil with both theories and equating the results. Here we will use an example similar to the one presented by Qian and Schellman[15] to compare ZB and LR: a polypeptide chain of eleven residues with the two ends free and a helix conformation between residues 3 and 7. In this polypeptide chain there are ten peptide bonds and nine residues flanked by peptide bonds. The codings and weights for this conformation in ZB and AGADIR are shown in Table I.

For $n = e^{-\Delta G_{\text{Ncap}}/RT}$ and $c = e^{-\Delta G_{\text{Ccap}}/RT}$.[12,13] As we indicated above, in both theories the nonhelical state is a nonrestricted random coil. This means that the weight of the random coil is the sum of the weights of the conformational unit restricted in helical angles (v in LR and h_n in AGADIR) and restricted in the remainder conformational space (c in LR). The weights shown above are normalized against the random coil weight.

As the reference state is the same for both theories it is possible to directly compare the weights for the entire polypeptide chain in this conformation. In ZB the weight is σS^3 and in AGADIR $n.h_n^5 \cdot h_e \cdot c$. Since we are only interested in comparing the basic helix/coil parameters, we can ignore the specific capping effects on AGADIR by assigning them the same weights than the random coil (the capping effects will be discussed in the following section). We obtain the equation

$$\sigma S^3 = h_n^5 h_e \tag{1}$$

Since the elongation of this helix by one conformational unit adds S to ZB weighting and $h_n h_e$ to AGADIR weighting, we can directly correlate them leading to the following relationships:

$$S\langle - \rangle h_n h_e = e^{-(\Delta G_{\text{intri}} + \Delta G_{\text{HBond}})/RT} \tag{2}$$

$$\sigma\langle - \rangle h_{n2}/h_{e2} = e^{-(2\Delta G_{\text{intri}} - 2\Delta G_{\text{HBond}})/RT} \tag{3}$$

These relationships can be used together with the following relationships between ZB and LR developed by Qian and Schellman.[15]

$$S\langle - \rangle w/(1 + v) \tag{4}$$

$$\sigma\langle - \rangle v^2/(1 + v)^4 \tag{5}$$

The relationship between AGADIR parameters and ZB elongation (S) is straightforward and physically consistent. The h_n accounts for the entropic cost of fixing the residue in helical angles, is sequence dependent, and is always smaller than one (by definition $1 = r + h_n$, r being the weight of fixing the residue in nonhelical angles, or the remainder). h_e is the weight for the mean-residue enthalpic contribution, which is sequence independent and favors helix formation (larger than one, as it comes from a negative free energy). The elongation ability of a given residue is then the product of these two numbers. If h_e predominates because h_n is close to 1, the residue has an elongating trend, but if h_n is very small (close to 0), the residue is basically a helix breaker. On the other hand, the relationship obtained for AGADIR parameters and ZB nucleation (σ) is physically inconsistent. There is no obvious physical explanation to relate σ with the quotient h_n^2/h_e^2. Moreover, the h_n^2 term results in a strong sequence dependence and the h_e^2 term adds an enthalpic component to the nucleation, in con-

Table II

Peptide bonds	1	2	3	4	5	6	7	8	9	10
ϕ,Ψ Residues	1	2	3	4	5	6	7	8	9	
ZBm coding	0	0	0	0	0	1	1	1	0	0
ZBm weights	1	1	$\sigma_n S$	S	S	S	$S\sigma_c$	1	1	

trast to the assumptions generally taken for σ in ZB and v in LR.[15] The physical inconsistency is a consequence of the combination of two differences between the two models: (a) A mismatch on the elongation starting point in both theories. In AGADIR the elongation ($h_n h_e$) starts on the fifth residue (sixth peptide bond), while in ZB and LR it starts on the third (fourth peptide bond). (b) The assignment of random coil weight to the first three helical peptide bonds in ZB (or v weight in LR to the first two residues), while they are assigned helical weight (h_n) in AGADIR. The elimination of any of the two differences vanishes the physical inconsistency. The first point is demonstrated by comparing a hypothetical AGADIR model with elongation on the third residue with ZB, on the previous example. The resulting equation is

$$\sigma S^3 = h_n^5 h_e^3 \tag{6}$$

which leads to the following relationships:

$$S\langle - \rangle h_n h_e = e^{-(\Delta G_{\text{intri}} + \Delta G_{\text{HBond}})/RT} \tag{7}$$

$$\sigma\langle - \rangle h_n^2 = e^{-(2\Delta G_{\text{intri}})/RT} \tag{8}$$

This mathematical expression has a clear physical meaning as it shows that nucleation is the entropic cost of fixing the two first residues in helical angles. Nucleation would then be entropic, but still sequence dependent. On the other hand, the fact that σ should be sequence independent is more a convenient working assumption than a requirement of ZB or LR theories. However, these relationships are important to prove the previous statement, but not to derive the nucleation from the original AGADIR theory, because in AGADIR due to its weighting procedure the elongation starts on the fifth residue.

In order to do so it is more useful to utilize the second difference between theories. Finkelstein and collaborators[11] have developed a version of ZB theory in which the main theoretical body is maintained, but the weighting is modified. This version of ZB theory (ZBm) is conceptually closer to AGADIR, uses the residue as conformational unit, and would assign the weights found in Table II to our previous example.

In this case nucleation is the product of two terms that depend on the end-helical residues (σ_n, σ_c). Equating this to AGADIR weights (above) we obtain

$$\sigma_n S^5 \sigma_c = h_n^5 h_e \tag{9}$$

which leads to the following correlations:

$$S\langle - \rangle h_n h_e = e^{-(\Delta G_{\text{intri}} + \Delta G_{\text{HBond}})/RT} \tag{10}$$

$$\sigma\langle - \rangle 1/h_e^4 = e^{-(4\Delta G_{\text{HBond}})/RT} \tag{11}$$

The physical meaning of this mathematical expression for nucleation is also clear and it is directly related to AGADIR theory. Nucleation in AGADIR comes from the fixing of the first four residues in helical angles without mean-residue enthalpic contribution. Though the nucleation process is entropic and sequence dependent, the nucleation parameter in AGADIR and ZBm is enthalpic and sequence independent, and arises from the specific weighting procedure of these models. The nucleation parameter directly accounts for the lack of four mean-residue enthalpic contributions. From these relationships it can be seen that the helix/coil cooperativity in AGADIR arises from the relative magnitudes of its two basic parameters (h_n and h_e). While there is an infinite set of h_n,h_c solutions for the elongation parameter (two numbers that compensate each other), each of these solutions produces a different nucleation, and therefore a different helix/coil cooperativity.

Summarizing, the relationships shown in Eqs. (2) and (3) are extremely useful for mathematically comparing calculations performed with AGADIR and ZB or LR [with the help of Eqs. (4) and (5)]. Equations (10) and (11), on the other hand, give the relationships between the two basic AGADIR parameters and the general processes of elongation and nucleation in the helix/coil transition. These relationships are useful for physically comparing AGADIR with theories that differ in the modeling of helix cooperativity.

AGADIR MODEL OF α-HELIX FORMATION

We have pointed out in the preceding section the differences between the model for α-helix formation presented in AGADIR and the classical ZB and LR theories. In this section we discuss the physical reasoning that in our view supports these modifications.

Capping Effects

The N-cap and C-cap have been introduced in AGADIR to account for the capping effects that have been experimentally reported in a variety of model systems. The inclusion of these effects in the AGADIR model of α-helix formation is rather natural, since the conformational space is defined in function of full helices. The N-cap and C-cap residues are defined as the two residues flanking the set of residues in helical dihedral angles and are required to form a helix. These residues have statistical weights different than the random coil because their ψ (for the N-cap) or ϕ (for the C-cap) dihedral angle is fixed and they contribute to the stability of the helical conformation. The maximally large helical conformation in a peptide of n residues ($n - 1$ peptide bonds and $n - 2$ residues surrounded by peptide bonds) is then constituted by $n - 2$ residues fixed in helical dihedral angles and the two caps (flanked only by one peptide bond and the other end free). This formulation is basically identical to the modification of LR theory presented by Baldwin and collaborators.[10]

Minimal Helix Length, Cooperativity, and Nucleation

AGADIR, as we have discussed above, uses a minimal helical length and a treatment of cooperativity and nucleation that depart from the classical ZB and LR theories. All these issues are closely related. In AGADIR the minimal helix length is four adjacent residues in helical angles plus the two caps (N-cap and C-cap). This is in contrast with ZB and LR formalisms, in which three adjacent residues are defined as the minimal unit to account for the fact that this nucleus is enough to make the first $i, i + 4$ main-chain–main-chain hydrogen bond.[16] Although the classical definition is very convenient for its simplicity on treating α-helix formation, there are some geometric and energetic arguments to assume a different minimal helix length. A definition of four residues is more in accordance with the natural geometry of the α-helix since a helix turn is formed by 3.6 residues. The energetic arguments are related to the physical origin of the cooperativity in the helix/coil transition. In the Lifson–Roig formalism the cooperativity is considered to arise mainly from the lack of main-chain hydrogen bonding in the first two residues of a helix.[10] However, there are other enthalpic components that should also affect the cooperativity of the helix/coil transition: interactions between the peptide dipoles and van der Waals interactions arising from the closed packing of the helix.[17] The interactions between neighboring peptide dipoles are repulsive since they are aligned parallel in the helical conformation, while those of long range are attractive, as they are aligned head to tail.[17] Equally, the van der Waals interactions between Cβ and carbonyl groups first appear in a nucleus of four and are not maximal until at least five residues are forming the helix (Figure 1). It appears that the enthalpic component of the helix/coil transition is a complex function of helix length.

The simplest model for the helix/coil cooperativity requires that all these contributions are suitably included while using the minimal set of parameters, which is two as ZB, LR, and AGADIR use. Besides, the relative contribution of these different interactions to the enthalpy of the helix/coil transition is not known. In such a simple model, the cooperative unit must be the helical segment in which the maximal change in free energy occurs, so that this free energy value is the closest to the elongation parameter. The entropic cost of fixing each residue in helical angles (h_n in AGADIR) contributes largely to the global change in free energy. However, this contribution to α-helix stability can be assumed to depend linearly on helix length (see below for a more detailed discussion on this point), as it is additive from the first residue to the last. Therefore it is less relevant for cooperativity and it is not essentially affected by the definition of the minimal helix length. The enthalpic contribution, which is the major source of the helix cooperativity and has a complex dependence on helix length, seems to be equilibrated approximately when four consecutive residues are fixed in helical angles. This can be observed schematized in Figure 1. In a nucleus of three helical residues the first hydrogen bond is made, but there is also the nearest neighbor repulsions between the three peptide dipoles. Addition of the fourth residue incorporates only one new nearest neighbor peptide dipole repulsion component, together with a new hydrogen bond, plus the van der Waals interactions between its Cβ and the carbonyl of the first residue (see Figure 1A). The incorpora-

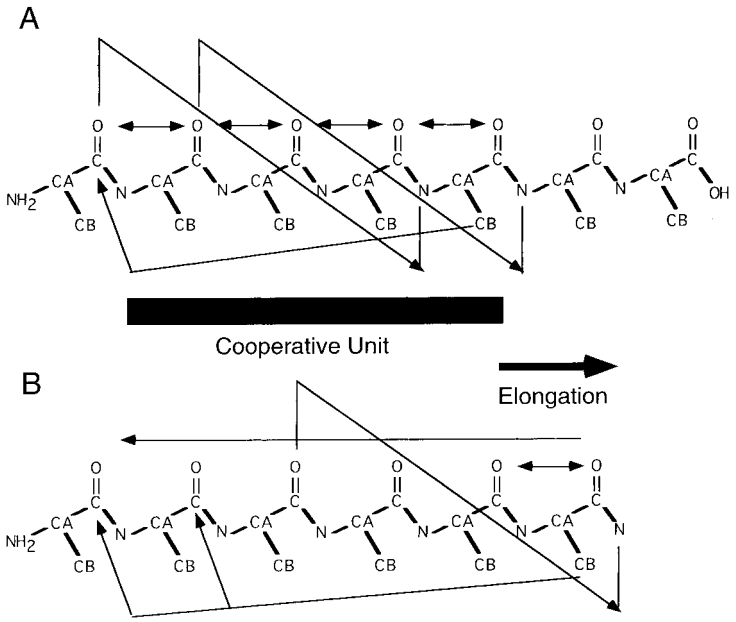

FIGURE 1 Schematic diagram showing the main contributions to helix nucleation and elongation. The atoms are indicated by name. Repulsive interactions (nearest neighbor dipole repulsions) are shown by double-headed arrows and attractive interactions (hydrogen-bond, van der Waals, and dipole attractions) by single-headed arrows. (A) Diagram showing the cooperative unit in AGADIR, with the different interactions mentioned in the text. (B) Diagram showing the new interactions added by elongating the nucleation core by one residue.

tion of the fifth residue is accompanied by a new hydrogen bond, a new net nearest neighbor peptide dipole repulsion, the first i, $i + 4$ peptide dipole attraction, and the interactions of the fifth Cβ with the first and second carbonyl groups (see Figure 1B). The pattern of the fifth residue is repeated for the sixth residue and so forth. Thus, the enthalpic contributions are equilibrated already when five consecutive residues are in helical angles and it can be considered that from then onward the α-helix is simply elongating. In AGADIR, to overcome the lack of knowledge on the relative enthalpic contributions and to maintain the number of basic parameters for the helix/coil transition on two, we have oversimplified the problem by assuming that in the cooperative unit of four residues the different conflicting contributions (repulsions and attractions) cancel out, giving a null enthalpic contribution. This means that the two hydrogen bonds and the van der Waals interactions between the fourth Cβ and the first carbonyl contribute a similar amount of enthalpy than the three repulsions between peptide dipoles. On the other hand, for the elongating residues there is a clear predominance of favorable interactions (see Figure 1B), which leads to a net favorable mean-residue enthalpy that promotes α-helix elongation. The mean residue enthalpic contribution was originally termed in AGADIR as a hydrogen-bond contribution ($\Delta G_{\mathrm{HBond}}$) by means of simplicity, and it was obtained from the values estimated experimentally by microscanning calorimetry.[18] This model is as simple as ZB and LR, but it is easily expandable when a deeper understanding of the different enthalpic contributions permit their decomposition into individual terms. Meanwhile, the error introduced by the assumption of a null enthalpic contribution for the cooperative unit is small, since it is included in the mean-residue enthalpic contribution that is averaged over all the residues forming part of the α-helix. Interestingly, despite the difference in the definition of the cooperative unit between Lifson–Roig and AGADIR, there is not a significant change in the calculation of the partition function (see the last section in this article). This is because segments shorter than four residues have an extremely small statistical weight in Lifson–Roig formalism, as they have the same cost in cooperativity than the longer segments. On the other hand, the inclusion of segments shorter than four in AGADIR would not significantly affect the partition function. The reason is that in short segments the repulsions between nearest neighbor

peptide dipoles should dominate, giving an unfavorable net enthalpic contribution that, summed up to the loss in conformational entropy from fixing the residues in α-helix, would give negligible statistical weights.

From all of this it is also evident that the process of nucleation in AGADIR involves the fixation of the first four residues in helical angles without favorable net enthalpic contribution. As we discussed in the previous section, nucleation in AGADIR is mainly entropic and sequence dependent, but the nucleation parameter is enthalpic. AGADIR theory assumes that the entropic cost of fixing a residue in helical angles is the same for a residue nucleating a helix (forming the first nucleus of four residues plus the caps) or elongating the helix. In this way, only two parameters (h_n and h_e) describe both nucleation and elongation, and nucleation becomes a strong function of the sequence. However, this is probably an oversimplification. The geometry of the helix shows that the side chains of the residues in an α-helix are pointing to the solvent oriented toward the N-terminus. Therefore, the side chain of a residue located in the middle of a long α-helix will be leaning on the helix cylinder, which very likely is one of the reasons for the different helix propensities of the residues. On the other hand, the side chains of the residues forming the first helix nucleus do not have the support of a preformed α-helix and their microenvironment is different, being fully exposed to the solvent. This situation is not only found in the helix nucleus, but is mimicked by the N-terminal residues of all the helices of any length. Nowadays there is no experimental information indicating that this factor is non-negligible, but if it is important, it would imply that the 20 amino acids have different nucleating and elongating propensities. It is also evident from this discussion that in case this effect is found experimentally to be important, its inclusion in the body of AGADIR theory is rather natural and straightforward. Besides, the assessment of the relevance of this effect would also support a definition of the minimal helix nucleus of four residues plus the caps, because it gives more strength to the geometrical argument of a minimally symmetrical unit. The side chains of the first four residues are fully exposed to the solvent while that of the fifth is already leaning on the first helix nucleus.

THE STANDARD ONE-SEQUENCE APPROXIMATION AND THE MULTIPLE SEQUENCE APPROXIMATION

In our original work[12,13] we used a simplification of the multiple sequence partition function that differs from the standard one sequence approximation. To improve the theoretical description of the original model we have formulated a multiple sequence partition function framed in the model for α-helix formation used in AGADIR. Moreover, we have also formulated in the same context the standard one-sequence approximation. The direct comparison between the different approximations should permit to evaluate the limitations of our original approximation. A classical way to simplify the helix/coil transition is to assume that only one helical segment at a time might be present in a single molecule (one-sequence approximation). Under this approximation those conformations with more than one helical segment are neglected and the partition function is approximated as the sum of the weights for the helical segments plus the weight of the coil conformation. A standard one sequence formulation within the framework of AGADIR for a peptide with n residues is expressed in Eq. (1), where $K_{\text{hel}}{}^{ji}$ is the weight of the helical segment comprised between residues i and $i + j - 1$ inclusive, and is calculated as previously indicated.[12,13]

$$Q_{1s} = 1 + \sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} K_{\text{hel}}{}^{ji} \qquad (12)$$

Once the partition function is known, the calculation of the fraction of the time a certain residue is in helical conformation is straightforward [Eq. (13)].

$$\langle X_{\text{Hel}} \rangle = \frac{\sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} c \cdot K_{\text{hel}}{}^{ji}}{Q_{1s}}$$
$$c = 0, \text{ for } x < i \text{ or } x \geq j + i$$
$$c = 1, \text{ otherwise} \qquad (13)$$

It is at this point in which the original AGADIR approximation differs from the standard one sequence approximation. In the previous AGADIR the weights of the helical segments that do not include residue X are directly eliminated from the denominator of Eq. (13).[12,13] Thus the effective partition function used for each residue is different. In other words, the reference state for each residue depends on the position of this residue in the polypeptide chain, as it includes the helical conformations in which the residue does not participate (see below for a specific example of calculations with these two and the multiple sequence approximations).

The multiple sequence approximation, on the

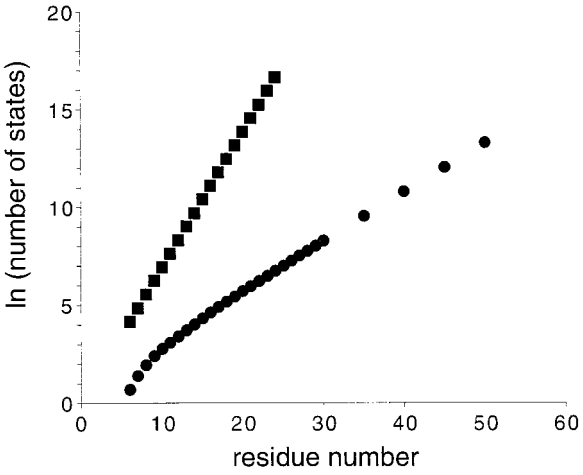

FIGURE 2 Natural logarithm of the number of states in function of peptide length. Cooperative unit of 1 residue (solid squares), AGADIRms (four helical residues plus the two caps; solid circles).

other hand, takes into account the existence of molecular conformations with more than one nonoverlapping helical segment. This approximation is more precise, as it is closer to the physical phenomena of helix formation. In the multiple sequence approximation the formation of a helical segment in one region of the polypeptide chain only interferes with another helical segment when they overlap. This considerably increases the number of calculations carried out to evaluate the partition function. With a minimal helical segment of one residue, the number of conformations that need to be accounted for grows by 2^{N-2}, N being the number of residues in the polypeptide chain. To formulate a multiple sequence approximation with the α-helix formation model of AGADIR we have considered the same minimal helix length (four helical residues plus the two caps), which largely decreases, as compared to a minimal segment of one, the number of conformations that must be calculated (see Figure 2). We have also assumed that there is no energetic coupling between nonoverlapping helical segments that are simultaneously present in the same molecule. This assumption seems rather reasonable for monomeric peptides in which there are no long- or medium-range interactions. Under these assumptions and the definition of the random coil state discussed in the previous section, the multiple sequence partition function becomes the sum of the statistical weights for all the possible combinations of helical segments (from one to the maximal number of nonoverlapping helical segments) plus the statistical weight for the random coil state. The statistical weight for each helical segment is calculated referenced to the random coil, as indicated in equation 10 of the previous work.[12] Therefore, the weight for the random coil is 1 (arises from the product of the weights of all the residues in the random coil state). As a result of the second assumption, the statistical weight of molecular conformations with more than one helical segment are simply the product of the weights of all the helical segments included on it. We have considered that, due to geometrical incompatibility, a residue cannot be the C-cap of a helical segment and simultaneously the N-cap of the following. Therefore, the maximal number of simultaneous helical segments in a polypeptide chain is the integer part of the quotient $N/6$. A formal expression of this multiple sequence approximation is given in Eq. (14).

$$Q_{\mathrm{ms}} = 1 + \sum_{j_1}\sum_{i_1} K_{j_1 i_1} \times \{1 + \sum_{j_2}\sum_{i_2} K_{j_2 i_2}[1 + \sum_{j_3}\sum_{i_3} K_{j_3 i_3} \cdots (1 + \sum_{j_{n/6}}\sum_{i_{n/6}} K_{j_{n/6} i_{n/6}})]\} \quad (14)$$

with the limits

$$j_k = 6 \text{ to } n - (i_{k-1} + j_{k-1}) + 1,\ j_0 = 0$$
$$i_k = i_{k-1} + j_{k-1} \text{ to } n - j_k + 1,\ i_0 = 1$$

where n is the number of residues in the polypeptide chain, $n/6$ stands for the integer part of the quotient, k is the number of helical segments present in a given molecular conformation, and j_k is the length of the k helical segment that has as a first residue i_k. $K_{j_k i_k}$ is the statistical weight of the helical segment that starts in i_k and ends in $i_k + j_k - 1$ and it is calculated as indicated in equation 10 of the previous work.[12] Knowing the multiple-sequence partition function, the calculation of the fraction of the time a residue in the polypeptide chain is forming part of an α-helix is straightforward, as shown in Eq. (15).

$$\langle X_{\mathrm{Hel}} \rangle = \frac{\sum_{j_1}\sum_{i_1} cK_{j_1 i_1}\{1 + \sum_{j_2}\sum_{i_2} cK_{j_2 i_2} \times [1 + \sum_{j_3}\sum_{i_3} cK_{j_3 i_3} \cdots \times (1 + \sum_{j_{n/6}}\sum_{i_{n/6}} cK_{j_{n/6} i_{n/6}})]\}}{Q_{\mathrm{ms}}} \quad (15)$$

with the limits:

$$j_k = 6 \text{ to } n - (i_{k-1} + j_{k-1}) + 1,\ j_0 = 0$$
$$i_k = i_{k-1} + j_{k-1} \text{ to } n - j_k + 1,\ i_0 = 1$$
$$c = 0 \text{ for } x < i_k \text{ or } x \geq j_k + i_k$$
$$c = 1 \quad \text{otherwise}$$

The differences on the calculation of the average helical content of a residue with these three approximations are better appreciated by comparing them on a specific example. Suppose there is a short polypeptide chain of 14 residues with the two ends nonprotected (13 peptide bonds and 12 residues bounded by peptide bonds). Let us also assume that this polypeptide chain is a polyalanine to simplify the mathematical expression of the partition function, since this does not alter the differences between calculations. In this example there might be a maximum of two helical segments simultaneously present in a molecule, because the minimal helix length is six residues (four in helical angles plus the two caps). The statistical weight of a helical segment of j residues is simply the product $nh_n^j h_e^{j-4} c$. Therefore, the average helical content of the residue 5 of this polypeptide chain is calculated with the multiple sequence approximation as

$$\begin{aligned}\langle 5_{\text{Hel}} \rangle = \{[&4nh_n^4c + 5nh_n^5h_e^1c + 6nh_n^6h_e^2c \\ &+ 6nh_n^7h_e^3c + 5nh_n^8h_e^4c + 4nh_n^9h_e^5c \\ &+ 3nh_n^{10}h_e^6c + 2nh_n^{11}h_e^7c + 1nh_n^{12}h_e^8c \\ &+ 6(nh_n^4cnh_n^4c) + 5(nh_n^4cnh_n^5h_e^1c) \\ &+ 2(nh_n^4cnh_n^6h_e^2c) + 1(nh_n^5h_e^1cnh_n^5h_e^1c)]/ \\ [1 &+ 9nh_n^4c + 8nh_n^5h_e^1c + 7nh_n^6h_e^2c \\ &+ 6nh_n^7h_e^3c + 5nh_n^8h_e^4c + 4nh_n^9h_e^5c \\ &+ 3nh_n^{10}h_e^6c + 2nh_n^{11}h_e^7c + 1nh_n^{12}h_e^8c \\ &+ 6(nh_n^4cnh_n^4c) + 5(nh_n^4cnh_n^5h_e^1c) \\ &+ 2(nh_n^4cnh_n^6h_e^2c) + 1(nh_n^5h_e^1cnh_n^5h_e^1c)]\}\end{aligned}$$

with the standard one-sequence approximation:

$$\begin{aligned}\langle 5_{\text{Hel}} \rangle = \{[&4nh_n^4c + 5nh_n^5h_e^1c + 6nh_n^6h_e^2c \\ &+ 6nh_n^7h_e^3c + 5nh_n^8h_e^4c + 4nh_n^9h_e^5c \\ &+ 3nh_n^{10}h_e^6c + 2nh_n^{11}h_e^7c + 1nh_n^{12}h_e^8c]/ \\ [1 &+ 9nh_n^4c + 8nh_n^5h_e^1c + 7nh_n^6h_e^2c \\ &+ 6nh_n^7h_e^3c + 5nh_n^8h_e^4c + 4nh_n^9h_e^5c \\ &+ 3nh_n^{10}h_e^6c + 2nh_n^{11}h_e^7c + 1nh_n^{12}h_e^8c]\}\end{aligned}$$

with the residue partition function approximation:

$$\begin{aligned}\langle 5_{\text{Hel}} \rangle = \{[&4nh_n^4c + 5nh_n^5h_e^1c + 6nh_n^6h_e^2c \\ &+ 6nh_n^7h_e^3c + 5nh_n^8h_e^4c + 4nh_n^9h_e^5c \\ &+ 3nh_n^{10}h_e^6c + 2nh_n^{11}h_e^7c + 1nh_n^{12}h_e^8c]/ \\ [1 &+ 4nh_n^4c + 5nh_n^5h_e^1c + 6nh_n^6h_e^2c \\ &+ 6nh_n^7h_e^3c + 5nh_n^8h_e^4c + 4nh_n^9h_e^5c \\ &+ 3nh_n^{10}h_e^6c + 2nh_n^{11}h_e^7c + 1nh_n^{12}h_e^8c]\}\end{aligned}$$

COMPARISON BETWEEN STATISTICAL APPROXIMATIONS

The development of the multiple-sequence (AGADIRms) and standard one-sequence (AGADIR1s) approximations allows us to compare the three theories and to check the accuracy of our original approximation and the dependence of AGADIR energy contributions on it. We will use the multiple-sequence formalism as the reference to compare the other two approximations, since it is the most precise. Besides, we will put the emphasis on the comparison between AGADIR and AGADIRms because our major interest is to determine the limitations of the residue partition function approximation and its effect on the determination of the energy contributions.

The standard one-sequence approximation is a simplification of the partition function in which the weights of helical conformations with more than one helical segment are eliminated (see previous example). The effects expected from this simplification are an underestimation of the helical content (many helical conformations are disregarded) that grows with increasing polypeptide length, and more pronounced helix fraying at the ends. However, the error should be small if the cooperativity of the helix/coil transition is strong and the calculations are performed for not very long polypeptide chains. The main reason is that those conformations have much lower weights than conformations with equivalent number of helical residues distributed in only one helical segment, because they have the cost of several nucleations (see previous example). Therefore, the helical term of the partition function is dominated by conformations with a single helical segment. Moreover, the shorter the polypeptide chain the smaller the number of conformations that are neglected by this simplification. Baldwin and collaborators have already studied the effect of this approximation on the calculation of the partition function of quasi-homopolymers of different lengths.[7] They found that the effect is negligible

for sequences of 50 residues or less. We have also extensively compared the standard one-sequence and the multiple-sequence versions of AGADIR. The results of our comparisons between AGADIR1s and AGADIRms, both in quasi-homopolymers and heteropolymers, coincide with those reported by Baldwin and collaborators. The calculations of AGADIR1s coincide within 0.3% error with those of AGADIRms in all the cases we have tested (polypeptide chains shorter than 56 residues). The value for the mean-enthalpic contribution (reflects the cooperativity) in AGADIR1s is the same than in AGADIRms, and the same happens for the rest of the energy contributions. Moreover, the results of the tests for the effect of the dependence on peptide length, helical content, and the distribution of the helical population at a residue level were almost identical for AGADIR1s and AGADIRms (data not shown). These results indicate that the standard one-sequence approximation is very precise in lengths shorter than 56 residues, even when applied to complex heteropolypeptides (protein fragments).

The approximation of the original AGADIR includes in the random coil the helical conformations in which a given residue does not participate (see previous example), and this is done in the same way for each residue. This is equivalent to assuming that different helical segments can be simultaneously present in the same molecule, even when they are overlapping in sequence. In contrast, the multiple sequence approximation only takes into account those multiple helical segments conformations that are physically feasible (nonoverlapping). This approximation is expected to have opposite effects to those described for the standard one-sequence approximation. The helical content is overestimated, because the effective partition function (residue partition function) includes less helical conformations than the real partition function. This effect will also grow with polypeptide length and will affect the helix/coil cooperativity. Besides, this simplification will underestimate the helix fraying of the ends. The reason is that the random coil of the residues at the ends of the polypeptide chain include more helical conformations than the random coil of residues centered in the chain, since the end residues participate in less helical conformations than the central residues. The direct comparison between AGADIR and AGADIRms reveals some differences, in contrast to the comparison between AGADIR1s and AGADIRms. Thus indicating that the residue partition function approximation is less precise than the standard one-sequence approximation. AGADIR produces an overestimation of the helical content, which is reasonably small in short peptides (less than 2%), but becomes more important in peptides longer than 25 residues (data not shown). The fitting of AGADIRms to the same set of experimental data used to fit AGADIR[13] results in a change in the mean-residue enthalpic contribution ($\Delta G_{\mathrm{HBond}}$) from -0.775 kcal/mol ($h_e = 4.17$ at 273K) to -0.792 kcal/mol ($h_e = 4.3$ at 273K). On the other hand, the other energy contributions to α-helix stability used in AGADIR[13] are not affected by the use of the residue partition function approximation. Therefore, the errors in the determination of the energy contributions due to the residue partition approximation only affect the mean-residue enthalpic contribution and are very small (less than 0.02 kcal/mol). The calculations of the average helical content on the whole peptide database (more than 500 peptides[13]) performed with AGADIRms (with an enthalpic contribution of -0.792 kcal/mol) reproduce the values of AGADIR. The overall linear correlation coefficient between experimental and calculated values is slightly better than the original ($r = 0.97$ with a slope of 1.04) and the error margins for the calculation are identical (data not shown). Similar results are found when the dependence of specific energy contributions, rather than the general agreement, is investigated. The intrinsic propensities of the amino acids are not affected by the statistical approximation and the same can be said for more complex contributions like: side-chain–side-chain interactions, interactions with the helix macrodipole and capping effects. As an example, we show in

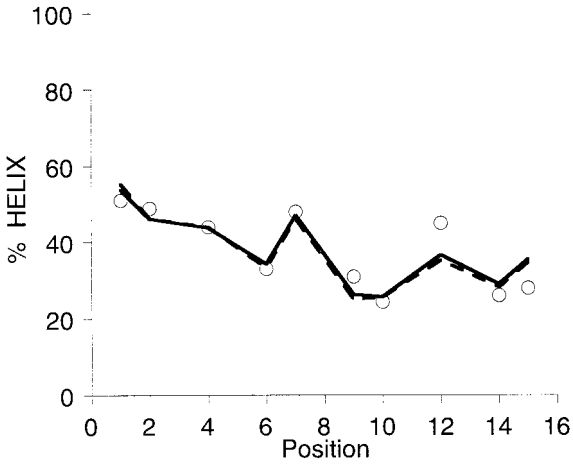

FIGURE 3 Changes in the helical content of polyalanine-based peptides upon placing Asp residues at different positions.[19] Experimental data (open circles), calculation with AGADIR (dashed line), calculation with by AGADIRms (continuous line).

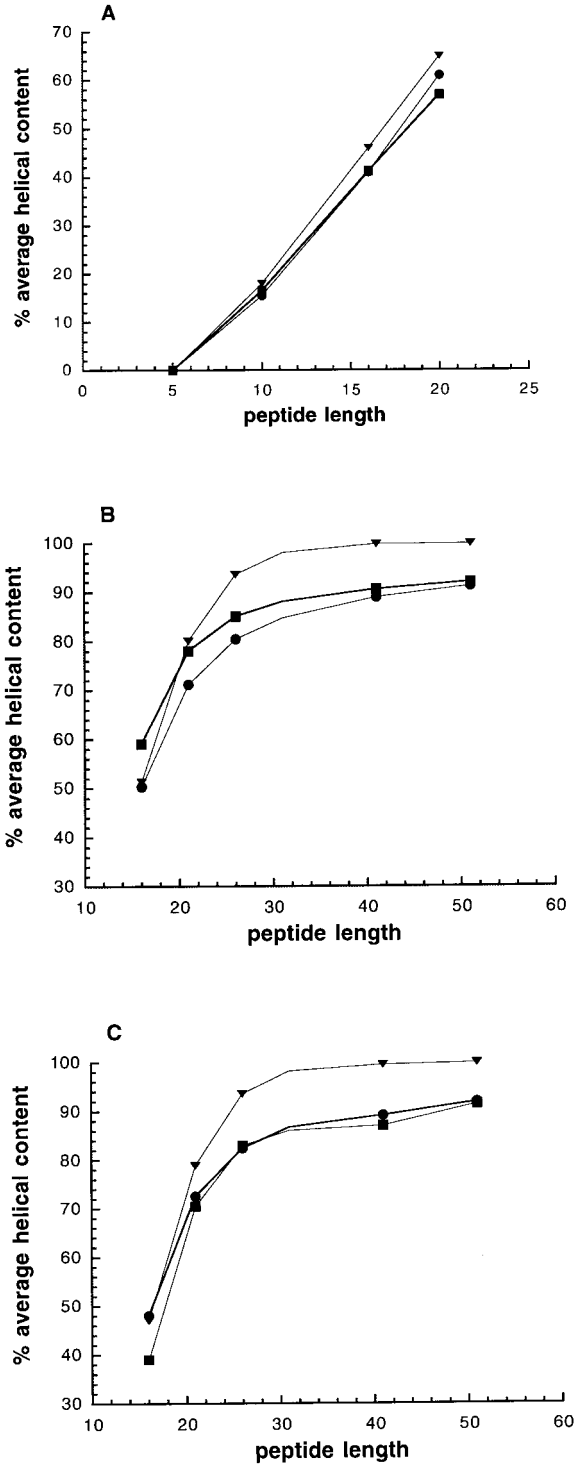

FIGURE 4 Dependence of the helical content of polyalanine-based peptides on the length of the peptide. Experimental data (squares), AGADIR (triangles), and AGADIRms (circles). (A) $(AAQAA_n)$[12,30]; (B) $(AAKAA)_n$.[21] The experimental average helical contents of this set of peptides have been derived from the fractional occupancies of the amide protons measured by Rohl et al.[21] using their equation 5. The numbers shown here correspond to the average of the values at four different times and referenced to the six residue peptide.[21] (C) $(AEAAKA)_n$.[20]

Figure 3 the comparison between the experiments of Huyghues-Despointes et al.[19] to study the interaction of Asp with the helix macrodipole and the calculations performed with AGADIR and AGADIRms. From these results we can conclude that the unique important bias produced by the original approximation on the energy contributions for α-helix formation is in the $\Delta G_{\mathrm{HBond}}$ value. This bias is very weak, influences the cooperativity of the helix/coil transition, especially in sequences longer than 30 residues, and is sequence independent. From now on, the comparisons between AGADIR and AGADIRms will be between versions using -0.775 and -0.792 kcal/mol, respectively, for the mean-residue enthalpic contribution ($\Delta G_{\mathrm{HBond}}$). The comparisons with AGADIR1s are not shown because AGADIR1s gives results nearly identical to AGADIRms in all the tests performed, as we discuss above.

An important point to test is the effect of the polypeptide chain length on the helical content of a polyalanine-based peptide, which is a measure of the cooperativity of the helix/coil transition. This has been experimentally investigated in different polyalanine peptide series.[12,20,21] In Figure 4 is given the comparison between the experimental data and AGADIR and AGADIRms calculations. In Figure 4 it can be observed that AGADIR is rather similar to AGADIRms for peptide lengths shorter than 30 amino acids, but AGADIR overpredicts the helical content of longer peptides. However, due to the experimental uncertainty produced by intrinsic errors in the measurements, different peptide series, and detection methodologies, the original calculations were in reasonably good agreement with the experimental data set. Finally, to investigate in more detail the accuracy of the original approximation, we can use a rather restrictive test, which is to compare calculations at a residue level performed with the different approximations. An incorrectly calculated distribution of helical content along the polypeptide chain modifies the shape of the distribution function. However, this kind of error is difficult to detect by simply comparing peptide helical contents because it is compensated by the averaging among all the residues of the molecule. We have calculated the helical content at a residue level with the different AGADIR approximations for all the protein fragments available in the

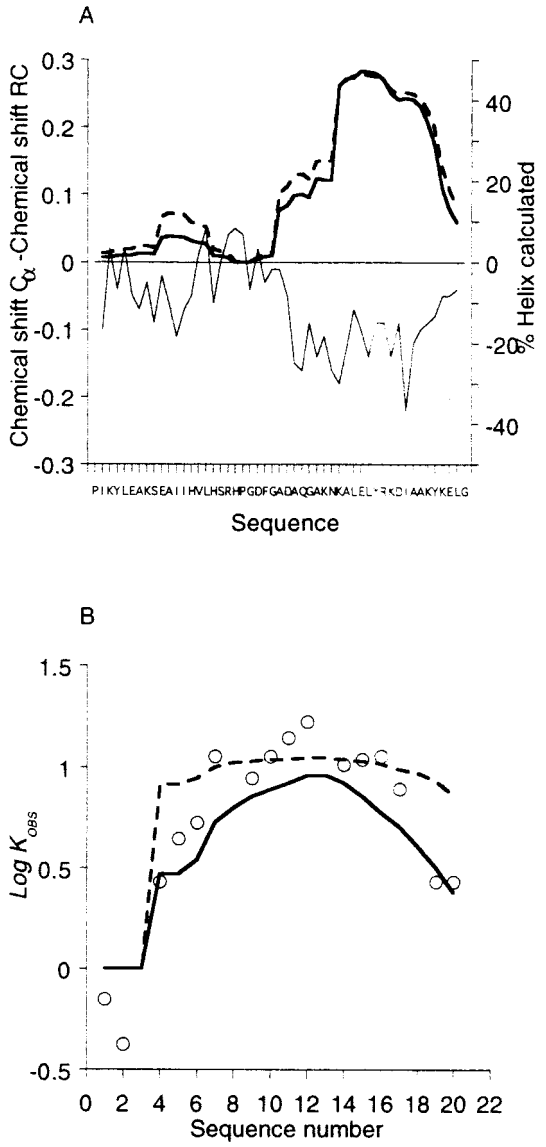

FIGURE 5 (A) Comparison of the individual-residue helical contents calculated with AGADIR (dashed line) and AGADIRms (continuous line) and the conformational shifts of the Cα protons (thin line).[22] (B) Comparison of the theoretical and experimental amide protection rates from the base-catalyzed exchange in a polyalanine based peptides.[23] Experimental data (open circles), AGADIR (dashed line), and AGADIRms (continuous line).

literature with significant helical content and nmr data. We have found that in complex heteropolypeptides with not very high helical contents (less than 50%, as it is normally found in protein fragments[14]), the calculated distribution of helicity along the polypeptide chain is very similar for all the approximations. The small differences found are much less than the experimental error on the determination of the helicity distribution by nmr. In Figure 5A we show, as an example of such comparisons, a model case protein fragment that has a significant helical population concentrated in one region of the peptide.[22] Nevertheless, in these peptides the helicity distribution is more determined by factors like capping residues and bad intrinsic propensities of the amino acids than by the helix fraying of the helix/coil transition. It is important to carry out the same kind of comparisons for polyalanine-based peptides, in which the helical content is very high and the helicity distribution is spread along the whole polypeptide chain. When our original work was published, there was no such an experimental information available. Recently, Rohl et al.[23] measured the base- and acid-catalyzed rates of amide proton exchange by one-dimensional nmr in a selectively ^{15}N-labeled polyalanine-based peptide and compared them with calculations performed with the Lifson–Roig formalism. This peptide is reported to have 80% average helical content as measured by far-uv CD. The protection rates can also be calculated with AGADIR, AGADIR1s, and AGADIRms with the equations that relate individual-residue helical contents with the rates of exchange of amide protons, which depend on the formation of hydrogen bonds[23] (Figure 5B). It can be observed that the profile calculated with AGADIR is flatter, predicting less fraying at the ends than AGADIRms (AGADIR1s is almost identical to AGADIRms) and likely than that experimentally observed. The last is hard to say, though, because the noise in this experiment is large. The profile for amide protection calculated with AGADIRms is almost identical to the one calculated by the authors with the Lifson–Roig formalism. Therefore, while in peptides with less than 50% helical content the original approximation renders helical distributions very close to those calculated with the multiple-sequence approximation, it diverges when the helical content is higher.

COMPARISON BETWEEN ELONGATION AND NUCLEATION IN AGADIR AND LIFSON–ROIG

Since the model of α-helix formation is quite different in AGADIR and ZB and LR formalisms, it is important to compare them in order to determine how dependent are the basic helix/coil parameters on the specific model used. In the first section of this article we have developed the mathematical relationships between AGADIR theory and ZB and LR. These relationships

Table III Elongation Parameters (*S*) at 0°C for Ala and Gly Intrinsic Propensities and AGADIR Free Energy Contributions Obtained from the Fitting of Different Helix/Coil Transition Approximations to The Polyalanine-Based Peptides of Baldwin and Co-workers[26 a]

	L-R[b]	AGADIR[b]			AGADIRms[b]		
	Value	ΔG	Value[d]	Values[e]	ΔG	Value[d]	Value[e]
S(Ala)	1.54	0.56	1.534	1.534	0.56	1.534	1.534
S(Gly)	0.05	2.34	0.0575	0.0575	2.41	0.05	0.05
σ	0.0029	−0.792	0.0069	0.0029	−0.792	0.0069	0.0029

	AGADIR[c]			AGADIRms[c]		
	ΔG	Value[d]	Value[e]	ΔG	Value[d]	Value[e]
S(Ala)	0.61	1.355	1.355	0.60	1.43	1.43
S(Gly)	1.94	0.117	0.117	1.94	0.12	0.12
σ	−0.775	0.0061	0.0033	−0.792	0.0059	0.00285

[a] The experimental results for Ala and Gly elongation parameters and the nucleation parameter are obtained from the fitting of several polyalanine-based peptides studied at 0°C, pH 7.0, and 1 *M* NaCl.

[b] The fitting has been carried out assuming the same elongation parameter for Ala and Lys, without taking into consideration capping and blocking effects.

[c] The fitting was done using and elongation of 1.06 for Lys and the capping and blocking effects described in AGADIR.[12]

[d] The elongation and nucleation parameters were obtained from AGADIR energy contributions using the mathematical correlations 2 and 3.

[e] The elongation and nucleation parameters were obtained from AGADIR energy contributions using the mathematical correlations (10) and (11).

can then be used to compare the theories on the same set of experimental data. In this regard, Baldwin's group has published nucleation and elongation parameters for most of the amino acids using a Lifson–Roig model for fitting the data.[24–26] Particularly, we have chosen a set of peptides studied by Chakrabartty et al.,[25,26] designed to obtain the nucleation parameter and elongation of Ala and Gly. The fitting of Lifson–Roig theory to the set of data was made by these authors neglecting N- and C-capping contributions and assuming the same elongation for Ala and Lys.[26] We have carried out fittings mimicking those conditions with the three different approximations of AGADIR theory and maintaining constant the h_e parameter (h_e = 4.3 at 273 K > $e^{0.792/RT}$). The basic AGADIR helix/coil parameters obtained from these fittings are shown in Table III. This table also shows the σ and *S* parameters obtained with Eqs. (4) and (5)[15] from the *v* and *w* values reported by Chakrabartty et al.[26] The calculation with equation 2 (or 10 since they are equivalent) of the *S* parameter upon the fitted AGADIR parameters shows that LR and AGADIR theories are identical on their description of the elongation process. The AGADIRms and AGADIR1s derived *S* parameter are the same than that obtained by Chakrabartty et al. with a LR algorithm. The original AGADIR also renders a very similar value, the small discrepancy being caused by the residue partition function approximation. Interestingly, if the fitting with AGADIR approximations is carried out including the original capping effects,[13] the difference in elongation of Gly and Ala becomes less drastic (see Table III). On the other hand, the calculation of σ with Eq. (3) ($\sigma = h_n^2/h_e^2$) renders a value that approximately doubles the one reported by Chakrabartty et al., regardless of the AGADIR approximation used (see Table III). The reason for this is a different minimal helix caught in AGADIR than in Z-B and L-R. Thus in order to achieve a mathematically correct expression, the relationship of Eq. (3) loses its physical meaning and Eq. (3) does not account for the process of α-helix nucleation in AGADIR theory. The relationship obtained with Eq. (3) really corresponds in ZB and LR theory to the nucleation of the helix plus the elongation by two residues ($h_n^2/h_e^2 = \sigma S^2$). In

fact, dividing the quantity h_n^2/h_e^2 (obtained from the fitting of AGADIR in any of its approximations) by the square of S [calculated from h_n and h_e with Eq. (2)] renders a value of nucleation of 0.00291 at 273 K, which is identical to the one previously reported using LR theory. Similarly, Eq. (11), which relates σ with only h_e and is the formal expression of dividing h_n^2/h_e^2 by S^2, produces directly the same value of nucleation than LR theory. We conclude from this that the process of nucleation is also virtually identical for AGADIR and ZB and LR, despite the different definition of minimal helix length and different treatment of cooperativity. This conclusion is of great importance because it means that the specific model for α-helix formation does not substantially affect the basic parameters of the helix/coil transition. The helix/coil transition is then a very robust theory that has physical relevance beyond the particularities of the model used to describe it. However, the differences between the models are most important when the helix/coil parameters are decomposed into the individual free energy contributions. The reason is the decomposition on individual free energy contributions depends on the specific model of the helix/coil cooperativity. A different cooperativity lead to discrepancies in the mean-residue enthalpic contribution. These discrepancies are transferred to the cost in conformational entropy when it is derived with analogs of eq. (2).[27] It is not straightforward to say a priori which model is closer to the physical reality, since very short helices of 3 or 4 residues in length are not experimentally observable. Lifson–Roig's has the clear advantages of its simplicity and wide acceptance, while AGADIR is more coincident with recent theoretical studies of α-helix formation.[17] Besides, AGADIR is easily expandable, when a better understanding of the different enthalpic contributions and of the process of nucleation is achieved. Finally, it is important to point out that the calculation of the energy contributions to α-helix formation not involved in α-helix cooperativity, like side-chain–side-chain interactions, do not show this dependence on the model for α-helix formation. The free energies of these factors calculated with the two models should be fully consistent and, indeed, some experimental examples show that this is the case.[28,29]

Remarks. The algorithm implementing both the multiple sequence and the standard one sequence approximations are available upon request. The algorithm implementing the standard one sequence approximation can also be freely run through the Internet at the web site: http://www.embl-heidelberg.de/ExternalInfo/Serrano/.

We are indebted to Prof. R. L. Baldwin, Dr. A. Doig, and specially C. Rohl for their critical reading and important remarks on the manuscript.

REFERENCES

1. Zimm, B. H. & Bragg, J. K. (1959) *J. Chem. Phys.* **31,** 526.
2. Lifson, R. & Roig, A. (1961) *J. Chem. Phys.* **34,** 1963–1974.
3. Scholtz, M., Qian, H., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991) *Biopolymers* **31,** 1463.
4. Chakrabartty, A. & Baldwin, R. L. (1995) *Adv. Protein Chem.* **46,** 141–176.
5. Vasquez, M. & Scheraga, H. A. (1988) *Biopolymers* **27,** 41–58.
6. Gans, P. J., Lyu, P. C., Manning, M. C., Woody, R. W. & Kallenbach, N. R. (1991) *Biopolymers* **31,** 1605–1614.
7. Scholtz, J. M., Qian, H., Robbins, V. H. & Baldwin, R. L. (1993) *Biochemistry* **32,** 9668–9676.
8. Stapley, B. J., Rohl, C. A. & Doig, A. J. (1995) *Protein Sci.* **4,** 2383–2391.
9. Shalongo, W. & Stellwagen, E. (1995) *Protein Sci.* **4,** 1161–1166.
10. Doig, A. J., Chakrabartty, A., Klinger, T. M. & Baldwin, R. L. (1994) *Biochemistry* **33,** 3396–3403.
11. Finkelstein, A. V., Badretinov, A. Y. & Ptitsyn, O. B. (1991) *Proteins Struct. Funct. Genet.* **10,** 287–290.
12. Muñoz, V. & Serrano, L. (1994) *Nat. Struct. Biol.* **1,** 399–409.
13. Muñoz, V. & Serrano, L. (1995) *J. Mol. Biol.* **245,** 275–296.
14. Muñoz, V. & Serrano, L. (1995) *J. Mol. Biol.* **245,** 297–308.
15. Qian, H. & Schellman, J. A. (1992) *J. Phys. Chem.* **96,** 3987–3994.
16. Cantor, C. R. & Schimmel, P. R. (1971) *The Behaviour of Biological Macromolecules,* 3rd ed., W. H. Freeman. New York, NY.
17. Yang, A. & Honig, B. (1995) *J. Mol. Biol.* **252,** 351–365.
18. Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M. & Bolen, D. W. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2854–2858.
19. Huyghues-Despointes, B. M. P., Scholtz, J. M. & Baldwin, R. L. (1993) *Protein Sci.* **2,** 1604–1611.
20. Scholtz, M., Qian, H., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991) *Biopolymers* **31,** 1463–1470.
21. Rohl, C. A., Scholtz, J. M., York, E. J., Stewart,

J. M. & Baldwin, R. L. (1992) *Biochemistry* **31,** 1263–1269.
22. Munier, H., Blanco, F. J., Precheur, B., Dieisis, E., Nieto, J. L., Craescu, C. T. & Barzu, O. (1993) *J. Biol. Chem.* **268,** 1695–1701.
23. Rohl, C. A. & Baldwin, R. L. (1994) *Biochemistry* **33,** 7760–7767.
24. Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994) *Protein Sci.* **3,** 843–852.
25. Chakrabartty, A., Schellman, J. A. & Baldwin, R. L. (1991) *Nature* **351,** 586–588.
26. Chakrabartty, A., Kortemme, T., Padmanabhan, S. & Baldwin, R. L. (1993) *Biochemistry* **32,** 5560–5565.
27. Scholtz, J. M. & Baldwin, R. L. (1992) *Ann. Rev. Biophys. Biomol. Struct.* **21,** 95–118.
28. Doig, A. J. & Baldwin, R. L. (1994) *Protein Sci.* **4,** 1325–1336.
29. Viguera, A. & Serrano, L. (1995) *Biochemistry* **34,** 8771–8779.