PROTEINS: Structure, Function, and Genetics 52:292–302 (2003)

Possible Locally Driven Folding Pathways of TC5b, a 20-Residue Protein

Gregory V. Nikiforovich,[1*] Niels H. Andersen,[2] R. Matthew Fesinmeyer,[2] and Carl Frieden[1]
[1]*Department of Biochemistry and Molecular Biophysics, Washington University, St. Louis, Missouri*
[2]*Department of Chemistry, University of Washington, Seattle, Washington*

ABSTRACT **A novel computational procedure for modeling possible locally driven folding pathways by stepwise elongations of the peptide chain was successfully applied to TC5b, a 20-residue miniprotein. Systematic exploration of the possible locally driven pathways showed that the Trp-cage structure of TC5b could be obtained by stepwise elongation starting from the noncentral local nucleation centers preexisting in the unfolded state of TC5b. The probable locally driven folding pathway starts with folding of α-helical fragment 4-9, followed by formation of the proper three-dimensional structure of fragment 4-12, and then 4-18. Accordingly, the Trp-cage-forming interactions emerge successively, first Trp^6–Pro^{12}, then Trp^6–Pro^{18}, and then Trp^6–Tyr^3. The Trp-cage-like structures of TC5b found in this study by independent energy calculations are in excellent agreement with the NMR experimental data. The same procedure rationalizes the incomplete Trp-cage formation observed for two analogs of TC5b. Generally, the success of this novel approach is encouraging and provides some justification for the use of computational simulations of locally driven protein folding. Proteins 2003;52:292–302.** © 2003 Wiley-Liss, Inc.

Key words: miniprotein; local interactions; peptides; stepwise elongation

INTRODUCTION

Two problems still dominate the structural biology of proteins, namely "Prediction of (3D) Structure From Sequence" and "Determination of How a Polypeptide Folds to the Native State."[1] Recently, there have been significant advances toward solving the first problem both from the basic principles of physics and by successful application of the knowledge-based approaches such as homology modeling. This study considers only the second problem, which makes the assumption that the unique native 3D structure of a protein-polypeptide is already known.

Many factors determine the complexity of this folding problem; perhaps the single most significant factor is the size of proteins. Obviously, the smaller a protein that folds into a stable native 3D structure, the easier it is to follow its folding pathways both by experimental and theoretical approaches. Typical protein sequences that fold into a uniquely stable 3D structure, however, are often comprised of hundreds of amino acid residues, whereas peptide sequences up to 20–25 residues generally exist in solution as an ensemble of multiple conformers with comparable stability. Therefore, design and discovery of relatively short sequences folding into a single predominant structure opens new prospects for studies of the molecular mechanisms of protein folding. Examples of those identified recently are a 28-residue zinc finger mimic,[2] about 40-residue β-sheet-like WW domains,[3] and a 20-residue β-sheet (Betanova).[4]

A 20-residue "miniprotein" Asn^1-Leu^2-Tyr^3-Ile^4-Gln^5-Trp^6-Leu^7-Lys^8-Asp^9-Gly^{10}-Gly^{11}-Pro^{12}-Ser^{13}-Ser^{14}-Gly^{15}-Arg^{16}-Pro^{17}-Pro^{18}-Pro^{19}-Ser^{20} (TC5b) was designed recently[5] by optimizing the sequence of the 39-residue peptide extendin-4 that has a partially populated fold designated as the "Trp-cage."[6] In TC5b, the Trp-cage motif consists of the indole ring of Trp^6 shielded from solvent by the aliphatic side-chains of Pro^{18} and Pro^{12}, as well as by the aromatic ring of Tyr^3 (Fig. 1). This structure was derived by using the NOE connectivities obtained in NMR studies, which afforded a well-converged 38-structure ensemble for TC5b (see the PDB entry 1L2Y). The Trp-cage structure was shown to be highly populated in TFE as well as in water, and TC5b was shown to fold cooperatively.[5] All these features make TC5b an excellent model protein for computational studies of possible folding pathways.

Computational studies of folding pathways for relatively short but stable protein fragments are most often performed by using molecular dynamics (MD). In typical MD simulations, one starts from an experimental structure and unfolds the structure under various conditions. The process of folding is simply considered as the reverse of unfolding. This technique has been used, for example, for Betanova[7] and for many medium-size proteins (see the extensive discussion on advantages and disadvantages of this approach in an excellent recent review[8]). MD simulations of folding from arbitrary starting structures have rarely been performed [e.g., for the model peptide $(Val)_4$-D-Pro-Gly-$(Val)_4$[9]]. However, recently such simulations were successfully performed for TC5b[10] and, through distrib-

Grant sponsor: National Institutes of Health; Grant numbers: GM 53630, GM 59658, DK 13332.

*Correspondence to: Gregory V. Nikiforovich, Department of Biochemistry and Molecular Biophysics, Campus Box 8036, Washington University, St. Louis, MO 63110. E-mail: gregory@ccb.wustl.edu

Received 3 October 2002; Accepted 12 December 2002

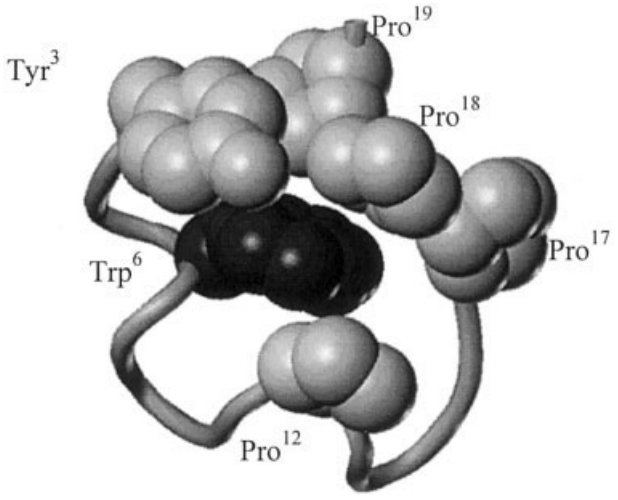

Fig. 1. Sketch of the NMR structure of TC5b. Only side-chains of the Trp-cage-forming residues (Trp in black) are shown.

uted computing projects, for at least three other polypeptides ranging in size from 12 to 36 amino acid residues.[11,12]

Despite the apparent successes of MD simulations, direct study of protein folding by this approach still has some inherent shortcomings. For instance, the path of a typical MD trajectory directly depends on energy values calculated for the saddle points that divide local energetic minima. The problem is that the force-field parameters used for energy calculations are usually calibrated by experimental data measured for the thermodynamically stable states, not for the saddle points that are, by definition, the thermodynamically unstable states. The recent MD simulations of TC5b converge to the Trp-cage structure found by NMR, but they stopped short of suggesting the actual folding pathways of TC5b.[10] Even more recently, MD simulations of BBA5, a 23-residue ββα zinc finger, suggested a two-state folding pathway that starts with collapse to native-like structure within 20 ns.[11] The BBA5 folding time was estimated as about 6 μs, in good agreement with the experimental value of 7.5 ± 3.5 μs.[13] It is not yet completely clear whether the MD trajectories that led to the folding event(s) provide an accurate description of the folding pathway characteristic of this particular miniprotein, because other systems studied by this method (including a 36-residue three-helix bundle villin headpiece protein and a 12-residue tryptophan zipper β-hairpin) also went through an initial collapse in the first 20 ns of simulation.[11,12]

This article presents a different approach based on the general suggestion that elements of the native structure may already be present in the unfolded states of proteins. There is much experimental data (mostly from NMR spectroscopy) supporting this assertion for proteins of various sizes and types,[14–25] from the 56-residue protein G[20] to the 128-residue CheY,[22] and from the β-sheet protein IFABP[15] to the α-helical apomyoglobin.[25] Recently, we developed a computational algorithm for identification of these elements of native structure in the unfolded states of proteins.[26] We also showed that these elements in some cases may be regarded as local nucleation centers that may be elongated in the process of further folding. Therefore, the basic assumptions used in the present approach are as follows:

1. There may be native-like locally stable fragments in the TC5b structure that maintain at least some of their low-energy conformations during folding.
2. Some of these TC5b fragments possess higher than average relative propensities to adopt the native-like structure and may be regarded as local nucleation centers.
3. The various locally driven folding pathways of TC5b can be modeled by alternative stepwise elongations of the peptide chain starting from different local nucleation sites (i.e., we assume that computational protocols using stepwise elongations may model possible locally driven pathways if, of course, the protocols converge to the native-like structure).

The main limitation here is that the above assumptions do not consider the possible existence of nonlocal interactions at the early stages of folding (e.g., "hydrophobic collapse"[27]). Therefore, only the folding pathways of TC5b that arise from local interactions can be reproduced by this approach. On the other hand, this approach clearly ensures fairly wide sampling of conformational space corresponding to a diversity of potential folding pathways. In other words, this study addresses only two very specific questions: (i) whether it is possible to fold the TC5b sequence to obtain the Trp-cage 3D structure consistent with the NMR data by successive stepwise elongations of the peptide chain (i.e., only through locally driven folding pathways) and, if yes, (ii) which locally driven folding pathway of TC5b appears more probable than others.

MATERIALS AND METHODS

Conformational Energy Calculations

Energy calculations for all peptide fragments were performed by using the ECEPP/2 potential field[28,29] assuming rigid valence geometry with planar *trans*-peptide bonds for all residues, including proline. In all cases, the fragments were considered acetylated at the N-terminal and N-methylamidated at the C-terminal. The conventional N- and C-terminal groups (NH_2 and COOH) were introduced only when considering the entire molecule of TC5b. Aliphatic and aromatic hydrogens were generally included in the unified atomic centers of CH_n type; H^{α}-atoms and amide hydrogens, as well as H^{δ}-atoms in prolines were described explicitly. All calculations were performed with the value of the dielectric constant $\epsilon = 80$ (the macroscopic ϵ value for water) to mimic, to some extent, the effect of water.

The starting points for energy calculations for penta-, hexa-, hepta-, and octapeptides (regarded as the initial fragments for possible further elongations) were all possible combinations of selected local energetic minima in the Ramachandran map.[30] The following local minima were included: φ, ψ values of −140°, 140°; −75°, 140°; −75°, 80°; −60°, −60°; and 60°, 60° for all non-glycine and

non-proline residues, with the addition of the φ, ψ values of 140°, −140°; 140°, −80°; −140°, 80°; 75°, −140°; and 75°, −80° for glycines (in most cases); and for prolines, the φ, ψ values of −75°, 140°; −75°, 80°; and −75°, −60° were included. This selection covers all significant peptide backbone conformers. The side-chain dihedral angle values were optimized before energy minimization to achieve their most favorable spatial arrangements by using an algorithm described previously.[31] The exception was the side-chain of Trp^6 for which each χ_1 rotamer in some cases was considered separately (see below).

Subsequent Elongation of Peptide Chain

First, energy minimization was performed for each starting conformation of every initial fragment. In total, the numbers of different starting conformations of the peptide backbone for different hexapeptides were between 15,000 and 70,000; for octapeptides, the numbers were up to 650,000. As a result, all low-energy conformers of the peptide backbone were selected (those with relative conformational energies $\Delta E = E - E_{min} \leq 1$ kcal/mol per residue; for justifying this criterion see Ref. 30). Then, several different computational protocols were used, which included either subsequent elongation of the peptide chain by one or two amino acid residues (all local minima for these additional amino acid residues were considered) or splicing two peptide fragments together (see below) starting from the selected conformations of the initial fragments of TC5b. The actual computational protocols are described in detail in Results.

Local Fragments: Relative Inherent Propensity to Adopt The Native Structure

According to the computational procedure developed previously,[26] the inherent propensity for a given fragment in the protein sequence to adopt the native-like structure is defined as a ratio between the number of low-energy conformers geometrically similar to the native structure (n_i) and the total number of low-energy conformers (N_i). Low-energy conformers are calculated for the isolated fragment and are selected by a chosen threshold ΔE of relative conformational energy. Low-energy conformers similar to the native structure are those satisfying a criterion of the chosen root-mean-square deviation (RMSD) between a selected conformer and the known native structure (the NMR structure in TC5b). Generally, such definition of the inherent propensity tacitly assumes that the computational procedure always finds the correct sets of low-energy conformers for isolated fragments. In reality, this is not the case, mainly because of systematic errors in modeling of the diverse interatomic interactions existing in peptides by the uniformly parameterized force fields. To alleviate this problem, the ratios were normalized by the average ratio for all fragments of the same size overlapping within the sequence in question to obtain measures of relative propensities to adopt the native structure for each given fragment. The relative propensities are calculated as $(n_i/N_i)/(\Sigma(n_i/N_i))/M)$, where index i relates to the number of a given fragment in the protein sequence, and M is the total number of fragments of the same size in the sequence. Obviously, any fragment that has a relative propensity value >1 has an above average propensity to adopt the native structure compared with other fragments of the same size in the same protein sequence.

RESULTS

Local Nucleation Centers in TC5b

In our previous study,[26] all overlapping hexapeptide segments in four proteins of about 130 residues were searched for possible local nucleation centers. Hexapeptide fragments were selected in that early study as the basic search units for two reasons. First, the length of six residues is sufficient to distinguish between 3D structures for most of the expected nucleation centers (β-strand, α-helix, or β-turn) at the chosen criteria of geometrical similarity with RMSDs up to < 2 Å (Cα atoms only). Second, complete conformational sampling of hexapeptides in about 130-residue sequences was readily attainable in terms of available computational resources.

On the other hand, larger initial fragments for stepwise elongation increase the likelihood to identify a local nucleation site, because a larger number of potentially important interactions will be considered at the very first stage. In fact, if it were possible to make the entire TC5b the "initial fragment" (i.e., to perform energy calculations for all combinations of local energetic minima of the peptide backbone for all 20 residues of TC5b (ca. 10^{13} combinations), all of the conformational possibilities for TC5b could be observed. In the absence of this, it is important to estimate the minimal initial fragment size that is suitable for stepwise elongation studies.

Figure 2 compares the relative propensities for the penta-, hexa-, hepta-, and octapeptides of TC5b to adopt NMR-like structures. Fragments containing residues 1 and 20 are excluded, because their terminal groups were not the same in the calculations as in the experimental system (see Materials and Methods). Energy calculations were performed for all fragments and those with RMSDs (Cα for residues 3–18 vs structure 1 of PDB entry 1L2Y) <2.0 Å were considered NMR-like. The penta-, hexa-, and heptapeptides fragments display three areas of increased relative propensity. By identifying the fragments by their starting residue, these areas correspond to fragments 4–6, 9–10, and 13–14 for the hexapeptides and 4–5, 9, and 12–13 for the heptapeptides. For pentapeptides, the peaks are slightly shifted (6–7, 10–11, and 13–15), and for octapeptides there are only two peaks close to the termini corresponding to fragments starting with residues 4–5 and 11–12. On this basis and in light of the available computational resources, it seemed reasonable to retain hexa- or heptapeptides as the initial fragments for further stepwise elongation procedures.

Systematic Preliminary Exploration of Possible Locally Driven Folding Pathways in TC5b

This section describes the estimates of relative propensities to adopt the native structure for all hepta-, nona-, undeca-, and tridecapeptides of TC5b. All combinations of local minima in the Ramachandran maps were used as the starting points for energy minimization for heptapeptides.

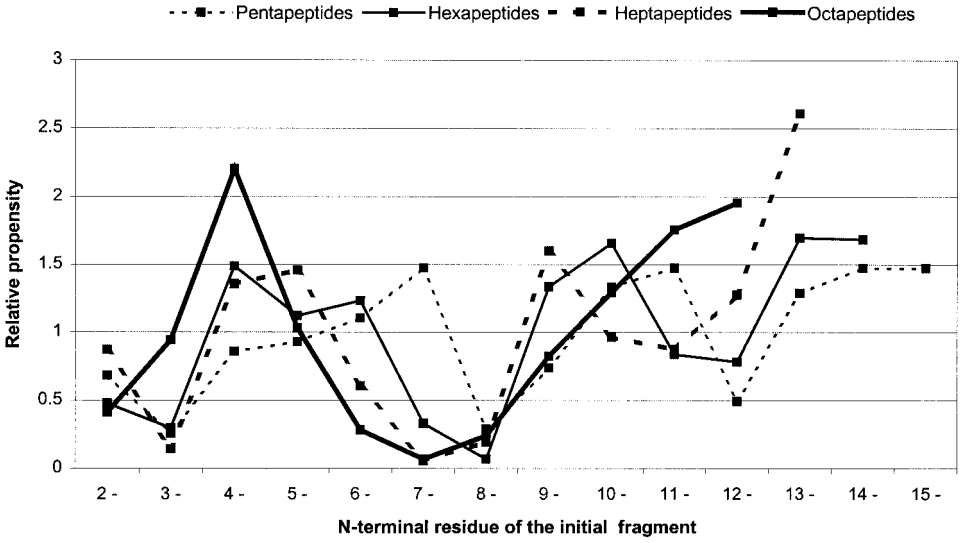

Fig. 2. Relative propensities of penta-, hexa-, hepta-, and octapeptides of TC5b to adopt the NMR structure.

TABLE I. Results of Systematic Stepwise Elongation of the TC5b Peptide Chain Starting From Heptapeptides

Heptapeptides				Nonapeptides				Undecapeptides				Tridecapeptides			
Type	a)	b)	c)	Type	a)	b)	c)	Type	a)	b)	c)	Type	a)	b)	c)
2–8	100	28	0.87												
3–9	197	9	0.14	2–10	94	1	0.07								
4–10	195	85	**1.35**	3–11	329	147	**3.19**	2–12	1281	795	**1.65**				
5–11	350	164	**1.45**	4–12	643	195	**2.17**	3–13	890	356	**1.06**	2–14	2262	505	**1.54**
6–12	385	75	0.60	5–13	686	30	0.31	4–14	945	165	0.46	3–15	3877	344	0.61
7–13	1381	25	0.05	6–14	479	22	0.33	5–15	1950	418	0.57	4–16	2404	245	0.70
8–14	434	27	0.19	7–15	2470	9	0.03	6–16	1874	436	0.61	5–17	2388	237	0.68
9–15	408	210	**1.60**	8–16	1302	44	0.24	7–17	2198	174	0.21	6–18	755	68	0.62
10–16	271	84	0.96	9–17	429	62	**1.03**	8–18	196	96	**1.30**	7–19	823	220	**1.84**
11–17	395	111	0.87	10–18	899	130	**1.03**	9–19	556	440	**2.10**				
12–18	212	87	**1.27**	11–19	584	133	**1.62**								
13–19	75	63	**2.61**												

[a]Number of low-energy conformers.
[b]Number of conformers geometrically similar to the NMR structure; [c]Relative propensity to adopt the NMR structure.

Then, three elongation steps, each involving simultaneous addition of two residues (the N- and C-terminal ones), from heptapeptides to nonapeptides to undecapeptides to tridecapeptides were applied. Only low-energy conformations possessing relative conformational energies of <1 kcal/mol per residue were selected for the next step. One can suggest then that the resulting low-energy 3D structures of each tridecapeptide emerge from locally driven folding pathways that started from the corresponding heptapeptides. For each fragment, the number of low-energy backbone conformations geometrically similar to the 3D structure of the corresponding hexapeptide fragment in the NMR structure of TC5b (structure 1 from the PDB entry 1L2Y) has been determined by the geometrical criteria of RMSD $<$ 2.0 Å (3–18 Cα atoms only) for hepta- and nonapeptides, and RMSD $<$ 3.0 Å for undeca- and tridecapeptides.

Table I summarizes the calculation data along with the relative propensities for various fragments of TC5b to adopt the NMR structure. The N- and C- terminal fragments featuring residues 1 and 20 are not included in Table I for reasons explained in the previous section. The results in Table I may be regarded as a fairly systematic exploration of all possible locally driven folding pathways in the TC5b molecule up to tridecapeptide fragments.

The relative propensity values >1 are shown in bold in Table I. Although heptapeptides fragments with higher than average propensity to adopt the NMR structure are located close to the termini (4–10, 5–11, 12–18, and 13–19) as well as in the central part of TC5b (9–15), the elongated fragments, starting with nonapeptides, retain higher than average propensity to adopt the NMR structure only when they are close to the termini. At the level of tridecapeptides, only two terminal fragments retain this feature, namely 2–14 and 7–19.

These observations also agree well with the results of energy calculations of octapeptides as the initial fragments (Fig. 2), because both suggest that locally driven pathways should not start with a nucleation site located in the center of the sequence, (e.g., the hexapeptide frag-

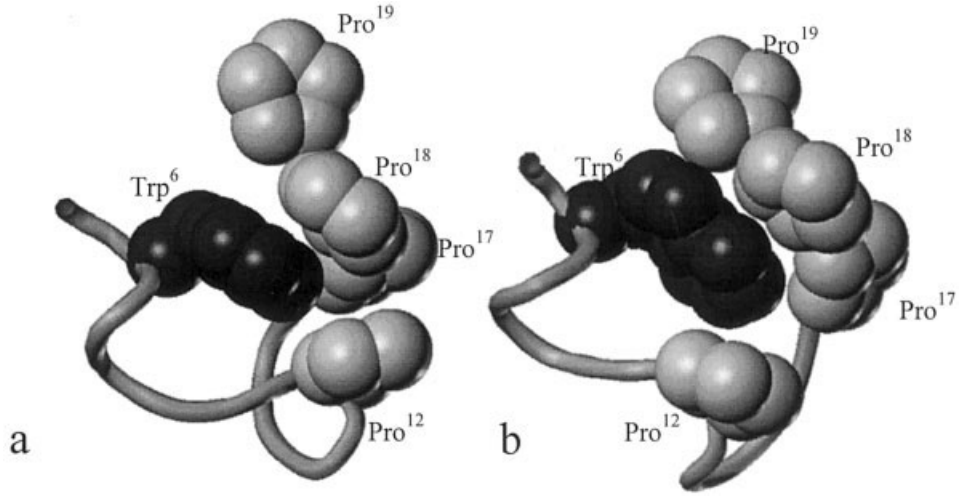

Fig. 3. Sketches of the Trp-cage-like structures of fragment 4-19 obtained by stepwise elongation starting either from the N-terminal local nucleus (**a**) or from the C-terminal local nucleus (**b**).

ments 9–14 or 10–15), but with terminal fragments (e.g., with hexapeptides 4–9, 5–10, and 6–11 that correspond to the α-helical fragment in the NMR structure (see Fig. 1) or with the C-terminal fragments 14–19). These possibilities are explored below in an attempt to build the Trp-cage structure beginning with terminal nucleation center candidates.

Elongation Pathways Starting From The N-terminal Helical Fragments: 4-11 to 4-15 to 4-19

As the first step, low-energy conformations for fragment 4-11 were determined by standard energy calculations. Starting points for energy minimization were obtained by splicing together low-energy conformers with differences of up to 10° in the overlapping dihedral rotation angles ϕ and ψ found in three hexapeptide fragments possessing above average relative propensities to adopt the native-like structure, namely 4–9, 5–10, and 6–11. Of 439 low-energy conformations for fragment 4-11 ($\Delta E < 8$ kcal/mol), 139 were geometrically similar to the corresponding fragment in the NMR structure (the RMSD < 1.5 Å). As the next step, low-energy conformers of fragment 4-11 were spliced with low-energy conformers of fragment 10-15 (the conformers with differences of up to 10° in the overlapping dihedral rotation angles ψ_{10} and ϕ_{11} were spliced). Of the resulting 327 low-energy conformers ($\Delta E <$ 11 kcal/mol) for fragment 4-15, 10 conformers were geometrically similar to the corresponding fragment in the NMR structure up to the RMSD < 1.5 Å (47 conformers at RMSD < 2.0 Å).

Finally, low-energy conformers of fragment 4-15 were spliced with low-energy conformers of fragment 14-19, which is the C-terminal fragment possessing the above average relative propensity to adopt the native-like structure (the conformers with differences of up to 10° in the overlapping dihedral rotation angles ψ_{14} and ϕ_{15} were spliced). At this stage, all three possible χ_1 rotamers of the side-chain of Trp^6 were considered separately (the dihedral angles of the other side-chains were optimized as previously described[31]). Of the resulting 268 low-energy conformers ($\Delta E < 12$ kcal/mol) for fragment 4-19, two conformers were geometrically similar to the Trp-cage structure with the RMSDs < 1.5 Å. One of them is shown in Figure 3(a).

Elongation Pathways Starting From the C-terminus: 14-19 to 10-19 to 4-19

Here the first step consisted of splicing together low-energy conformers of two hexapeptide fragments with the above average propensities to adopt the native structure, 14-19 and 10-15 (the conformers with differences of up to 10° in the overlapping dihedral rotation angles ψ_{14} and ϕ_{15} were spliced). Of the resulting 1157 low-energy conformers ($\Delta E < 10$ kcal/mol) for fragment 10-19, 102 conformers were geometrically similar to the corresponding fragment in the NMR structure with the RMSDs < 1.5 Å. Then, low-energy conformers of fragment 10-19 were spliced with low-energy conformers of fragment 4-11 (see above); the conformers with differences of up to 10° in the overlapping dihedral rotation angles ψ_{10} and ϕ_{11} were spliced. Again, all three possible χ_1 rotamers of the side-chain of Trp^6 were considered separately at this stage. Of the resulting 620 low-energy conformers ($\Delta E < 12$ kcal/mol) for fragment 4-19, only one conformer was geometrically similar to the Trp-cage structure with the RMSD < 1.5 Å. It is shown in Figure 3(b).

The above calculations strongly suggest that it is possible to build the Trp-cage structure out of the preselected sets of low-energy conformations of the noncentral local nucleation centers of the TC5b molecule starting from either of them. Two types of described stepwise elongation of the peptide chain correspond roughly to two different families of locally driven folding pathways toward the Trp-cage structure. More detailed comparison of calculated structures for fragment 4-19 in Figures 3(a) and (b) with the NMR structure in Figure 1 shows that the structure obtained by stepwise elongation from the N-terminal fragment (from 4-11 to 4-15 to 4-19) is closer to

the NMR structure in the important conformation of fragment 8-12, which contains two highly flexible Gly^{10} and Gly^{11} residues [Fig. 3(a)]. Finally, we selected the locally driven folding pathway that starts from the N-terminal local nucleation center 4-9 for more detailed studies.

Locally Driven Folding Pathway of TC5b Starting From Fragment 4-9 to 4-12 to 4-16 to 4-18 to 1-20

We started this computational procedure of stepwise elongation of the peptide chain by combining the low-energy conformations of the local nucleation center 4-9 with low-energy conformations of fragment 6-12, which, in turn, were obtained by one-residue elongation of fragment 6-11. As expected, the Trp-cage-forming interactions between the side-chains of Trp^6 and Pro^{12} were present in some low-energy conformations of fragment 6-12. In total, 57 of 1519 low-energy conformations ($\Delta E < 7$ kcal/mol) were geometrically similar to the NMR structure (the RMSD < 1.5 Å). Combination of low-energy conformations of fragments 4-9 and 6-12 was performed by splicing conformers with differences in the overlapping dihedral rotation angles ϕ and ψ up to 10°. The resulting fragment 4-12 yielded 1106 low-energy conformers ($\Delta E < 10$ kcal/mol); 151 of them were geometrically similar to the NMR structure (RMSD < 1.5 Å). Thus, the ratio of the NMR-like structures has increased in fragment 4-12 compared to fragment 6-12. This seems quite reasonable, because the Ile^4 residue, which additionally limits possible rotations of the Trp^6 side-chain in the Trp-cage, is present in fragment 4-12, but not in 6-12.

The next step consisted of one-residue elongation of fragment 4-12 to 4-13, which resulted in 1558 low-energy structures ($\Delta E < 10$ kcal/mol), 64 of them being geometrically similar to the NMR structure (RMSD < 1.5 Å). Separately, fragment 12-16 was calculated, which yielded 92 NMR-like structures out of 387 low-energy conformers ($\Delta E < 10$ kcal/mol). Then the low-energy conformations of fragment 4-13 were combined with the low-energy conformations of fragment 12-16 by splicing conformers with differences of up to 5° in the overlapping dihedral rotation angles ϕ and ψ. The resulting fragment 4-16 possesses 845 low-energy conformers ($\Delta E < 10$ kcal/mol); 13 (139) of them were NMR-like at the level of RMSD < 1.5 (2.0) Å. The Trp-cage-like structures of fragment 4-16 are stabilized by interacting side-chains of Asp^9 and Arg^{16} that form an additional obstacle for the Trp side-chain to escape the cage.

Then low-energy conformers of fragment 4-16 were combined with low-energy conformations of fragment 15-18 (deduced from low-energy conformers of the local nucleation center, 14-19) by splicing conformers with differences of up to 10° in the overlapping dihedral rotation angles ϕ and ψ. Fragment 4-18 features almost all the Trp-cage-forming interactions suggested by experimental studies, namely Trp^6-Pro^{12} and Trp^6-Pro^{18}. It possesses 314 low-energy structures ($\Delta E < 12$ kcal/mol), 15 of them being geometrically similar to the NMR structure (RMSD $<$ 2.0 Å).

Finally, stepwise elongations were performed for fragment 4-18 to fragment 3-19 to 2-20 and then to 1-20, which is the entire molecule of TC5b. As the last step, low-energy conformations of TC5b (994 conformers with $\Delta E < 12$ kcal/mol) were reminimized starting from the g^- rotamer ($\chi_1 \sim -60°$) of the Leu^7 side-chain to further restrict the Trp^6 side-chain within the Trp-cage. These calculations resulted in 409 low-energy conformers of TC5b ($\Delta E < 10$ kcal/mol), which represent the independently calculated ensemble of possible conformations of TC5b.

The above computational procedure describes the probable locally driven folding pathway of TC5b leading to the Trp-cage structure as consisting of the following steps: (i) at least one turn of α-helical structure from Ile^4 to Leu^7 is formed; (ii) the Trp^6 side-chain forms a complex with Pro^{12}; (iii) side-chains of Asp^9 and Arg^{16} interact with each other stabilizing the characteristic structure of the Gly-rich fragment 10-15; (iv) Pro^{18} docks to the still available surface of the Trp^6 side-chain; (v) the side-chain of Tyr^3 interacts with Pro^{19} covering the Trp-cage. The steps are depicted in Figure 4.

DISCUSSION

The ensemble of 409 low-energy conformations obtained by energy calculations does not consist only of structures geometrically similar to the NMR structure. It can be divided into six conformational clusters. Each of the clusters contains structures that are similar to each other at the level of RMSD < 1.5 Å (C^α atoms of the Trp-cage forming fragment 3-18). The most populated predominant cluster I consists of 278 structures, cluster II of 68 structures, and cluster III of 34 structures. The last three clusters together consist only of 27 structures.

Figure 5 depicts structures within the first three clusters that most closely resemble the NMR structure, along with the NMR structure itself. It is obvious that selected structures from the most populated cluster (cluster I) bear the greatest similarity to the NMR structure. Indeed, there are 17 structures belonging to cluster I that are geometrically similar to the NMR structure (the NMR-like structures) with an RMSD < 1.5 Å (C^α atoms of fragment 3-18). This is a key finding of this study, because it demonstrates that it is possible to reach the experimentally observed Trp-cage by modeling folding pathways driven by local interactions.

More quantitative comparison of calculation results to the NMR data are given in Table II, where average interproton distances were calculated by assuming that all structures possess equal statistical weights in solution. The distances are compared with the NMR restraints from NOE measurements[5] that are summarized in the PDB entry 1L2Y. Because the computational protocol of energy calculations involved determination of only one of the plausible rotamers of the side-chains (except the Trp^6 side-chain), only those NMR restraints that could be unambiguously compared with the calculation results were selected. They include 47 restraints on interproton distances within the peptide backbone (two of them, αH_3–$\beta 2H_{19}$ and αH_3–$\delta 2H_{19}$, technically belong to the backbone–side-chain distances) and 27 restraints defining the spatial position of the Trp^6 side-chain.

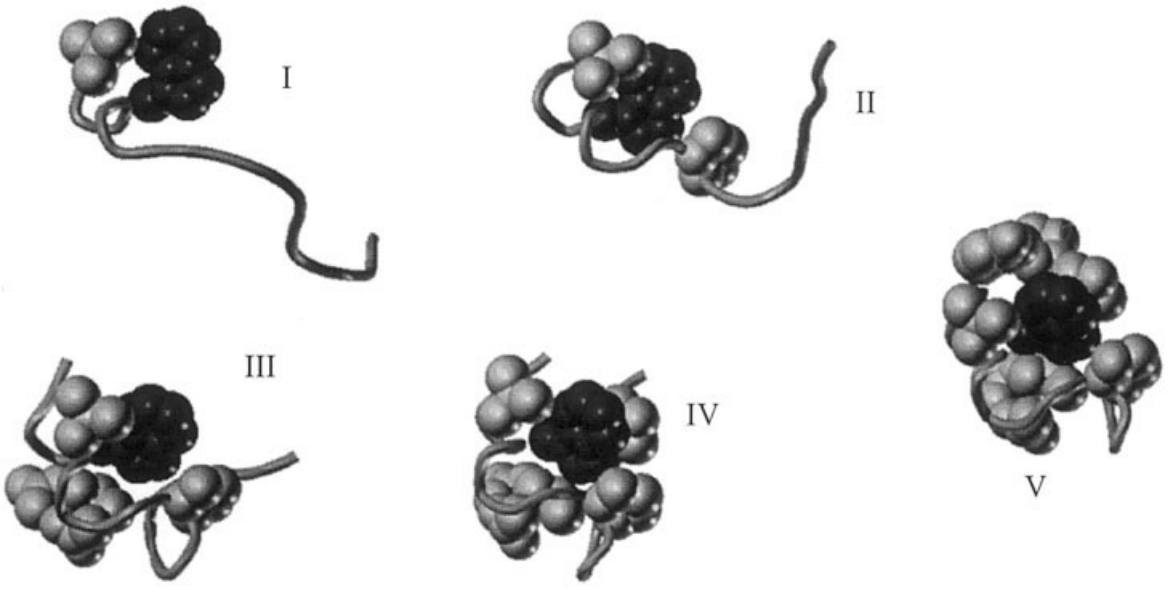

Fig. 4. Suggested steps of a locally driven folding pathway leading to the Trp-cage structure. Only the side-chains of the "cage-forming" residues mentioned in text are shown (Trp in black).

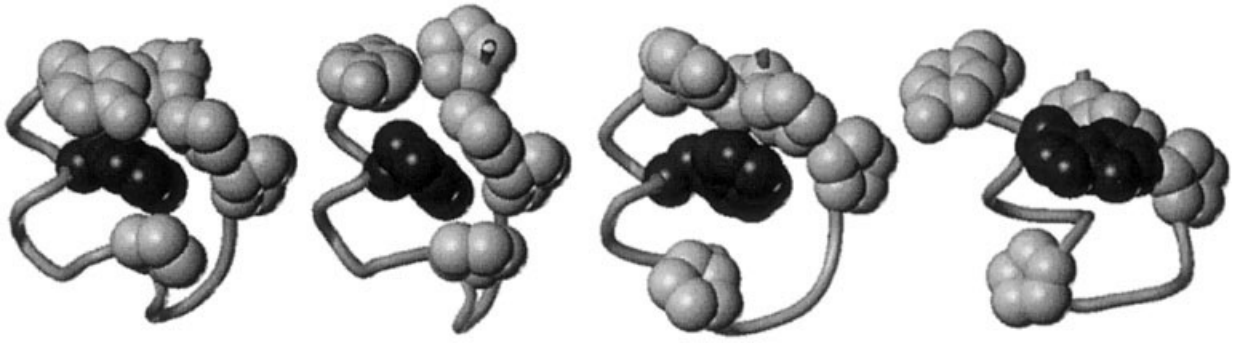

Fig. 5. The NMR structure of TC5b (**left**) compared to low-energy conformers from cluster I, II, and III most close to the NMR structure (the RMSDs are 1.09 Å, 2.15 Å, and 3.17 Å, respectively). Only side-chains for residues Tyr^3, Trp^6 (in black), Pro^{12}, Pro^{17}, Pro^{18}, and Pro^{19} are shown.

In Table II, average distances that differ from the NMR restraints by more than 0.5 Å are shown in bold. One can see that those discrepancies are rare among distances within the peptide backbone. The most populated cluster I features 10 such distances, and the set of NMR-like structures features only three, namely, distances αH_7–$\alpha 3H_{11}$, αH_{13}–NH_{15}, and αH_3–$\beta 2H_{19}$. The first of these distances satisfies the NMR restraint in structures of clusters II and III. At first sight, this could suggest that the calculated NMR-like structures may not be totally predominant in water solution. The αH_{13}–NH_{15} and αH_3–$\beta 2H_{19}$ distance are out of limits set by the NMR restraint in all clusters. However, the difference of 0.7 Å for the αH_{13}–NH_{15} distance (as that for the NMR-like structures) hardly can be regarded as significant, because this particular distance involves a presumably flexible glycine residue (Gly^{15}). The most sequentially remote contact αH_3–$\beta 2H_{19}$ is also problematic by the same reasons (involves Gly^{10}, Gly^{11}, and Gly^{15}).

For the distances defining the spatial position of the Trp^6 side-chain within the Trp-cage, they are in somewhat lesser agreement with the NMR restraints. However, in this case, too, only 4 of 27 distances calculated for the NMR-like structures display significant difference (>1 Å) with the NMR restraints, namely, ζ_2H_6–αH_{12}, η_2H_6–αH_{12}, η_2H_6–δH_{12}, and $\epsilon 1H_6$–NH_{16}. All of them are long-range interproton contacts, which can change dramatically with small dihedral angle variations (as αH_3–$\beta 2H_{19}$ and αH_3–$\delta 2H_{19}$). Overall, the agreement of the independently calculated data (the NMR-like structures) with those deduced from the NOE measurements is excellent: of 74 examined distances, only 7 show some discrepancy with the experimental data.

The experimental NMR studies revealed some conformational features of TC5b complementary to the measured NOEs, namely, the abnormal ring current shifts of $\beta 3H_{18}$ and $\alpha 2H_{11}$, as well as almost complete protection from the

TABLE II. Average Interproton Distances (Å) Calculated for Various Clusters of Low-Energy Structures of TC5b Compared to the NMR Restraints (the NMR-Like Are 17 Structures From Cluster I)

Distance	Total	Cluster I	Cluster II	Cluster III	NMR-like	NMR restraints
Distances within peptide backbone						
αH_1–HN_2	2.8	2.8	2.9	2.8	3.2	2.5–3.5
HN_2–NH_3	3.9	3.8	4.0	3.9	2.6	2.5–3.5
αH_2–HN_3	2.8	2.7	2.7	2.9	3.5	2.9–4.0
NH_2–NH_4	5.4	**5.6**	5.1	4.5	4.0	3.3–5.0
αH_2–NH_5	5.0	5.2	4.2	4.1	3.4	3.3–5.0
NH_3–NH_4	3.2	3.4	2.7	2.7	2.6	2.0–3.0
αH_3–NH_4	3.5	3.5	3.5	3.6	3.6	2.9–4.0
αH_3–NH_6	4.2	4.5	3.4	3.2	3.3	2.9–4.0
NH_4–NH_5	2.8	2.9	2.5	2.5	2.5	2.5–3.5
αH_4–NH_5	3.8	3.9	3.6	3.6	3.6	2.9–4.0
NH_4–NH_6	4.3	4.6	3.8	3.8	3.8	3.3–5.0
αH_4–NH_7	**4.5**	**4.9**	3.3	3.3	3.4	2.5–3.5
αH_5–NH_6	3.8	4.0	3.5	3.5	3.5	2.9–4.0
NH_5–NH_7	4.6	4.8	4.1	4.0	4.0	3.3–5.0
αH_5–NH_8	**4.4**	**5.0**	3.2	3.1	2.9	2.5–3.5
NH_6–NH_7	3.0	3.0	2.5	2.5	2.5	2.5–3.5
αH_6–NH_7	4.0	4.1	3.6	3.6	3.6	2.9–4.0
αH_6–NH_9	4.9	5.4	3.4	3.4	3.4	3.3–5.0
NH_7–NH_8	2.9	3.1	2.6	2.5	2.4	2.5–3.5
αH_7–NH_8	3.9	4.2	3.6	3.6	3.6	2.9–4.0
NH_7–NH_9	4.9	5.3	4.1	4.1	3.9	3.3–5.0
αH_7–NH_9	5.4	**5.8**	4.5	4.4	4.2	3.3–5.0
αH_7–NH_{10}	5.4	5.5	5.5	5.5	3.2	3.3–5.0
αH_7–NH_{11}	**5.8**	**6.7**	3.5	**3.6**	3.0	2.0–3.0
αH_7–$\alpha 3H_{11}$	**7.6**	**9.1**	3.6	3.8	**5.3**	2.9–4.0
αH_8–NH_9	4.2	4.5	3.6	3.6	3.6	2.9–4.0
NH_8–NH_9	3.2	3.5	2.7	2.6	2.5	2.0–3.0
NH_8–NH_{10}	5.4	5.2	**6.1**	**6.1**	3.8	3.3–5.0
αH_9–NH_{10}	3.5	4.1	2.2	2.2	3.5	2.9–4.0
NH_9–NH_{10}	3.2	3.1	3.9	3.9	2.5	2.0–3.0
NH_9–NH_{11}	5.2	**5.7**	4.2	4.4	4.1	3.3–5.0
NH_{10}–NH_{11}	3.1	3.2	2.8	3.0	2.4	2.0–3.0
$\alpha 2H_{10}$–NH_{11}	2.9	2.7	3.6	3.6	2.8	2.9–4.0
$\alpha 3H_{10}$–NH_{11}	3.4	3.8	2.8	2.6	3.6	2.9–4.0
$\alpha 2H_{11}$–NH_{11}	2.4	2.3	2.8	2.8	2.3	2.5–3.5
$\alpha 3H_{11}$–NH_{11}	2.7	3.0	2.2	2.2	2.9	2.0–3.0
αH_{12}–NH_{13}	3.6	4.0	3.4	3.5	3.4	2.9–4.0
αH_{12}–NH_{14}	4.7	4.8	4.1	4.5	3.7	3.3–5.0
αH_{12}–NH_{15}	**5.6**	4.5	**6.2**	**7.6**	2.8	3.3–5.0
NH_{13}–NH_{14}	3.4	3.4	2.3	2.4	2.8	2.0–3.0
αH_{13}–NH_{14}	4.3	4.1	3.5	3.6	3.6	2.9–4.0
αH_{13}–NH_{15}	**6.8**	**6.1**	**6.4**	**5.2**	**4.7**	2.9–4.0
NH_{14}–NH_{15}	**3.7**	2.9	**4.2**	3.1	2.5	2.0–3.0
αH_{14}–NH_{15}	4.1	4	2.2	2.6	3.6	2.9–4.0
NH_{15}–NH_{16}	3.3	2.9	**4.2**	2.3	2.3	2.0–3.0
αH_3–$\beta 2H_{19}$	**9.5**	**9.2**	**8.4**	**11.9**	**5.6**	3.3–5.0
αH_3–$\delta 2H_{19}$	**10.0**	**10.6**	**7.5**	**9.8**	2.5	2.9–4.0
Distances defining spatial position of the Trp^6 side-chain						
αH_3–$\beta 3H_6$	**4.8**	**5.7**	2.7	2.5	2.4	2.9–4.0
αH_3–$\varepsilon 3H_6$	**4.8**	**5.0**	3.7	4.8	3.3	2.9–4.0
αH_6–$\beta 2H_6$	2.8	3.0	2.4	2.4	2.4	2.5–3.5
αH_6–$\beta 3H_6$	2.8	2.7	3.0	3.0	3.0	2.0–3.0
$\beta 2H_6$–NH_6	2.5	2.5	2.7	2.6	2.6	2.0–3.0
$\beta 3H_6$–NH_6	3.1	3.5	2.3	2.3	2.3	2.0–3.0
$\delta 1H_6$–αH_6	2.9	2.6	3.0	3.7	2.8	2.5–3.5
$\delta 1H_6$–$\beta 3H_6$	3.7	3.6	3.9	3.2	3.9	2.5–3.5
$\varepsilon 3H_6$–NH_6	4.7	4.9	4.3	4.5	4.4	3.3–5.0
$\beta 2H_6$–NH_7	3.8	3.5	**4.1**	**4.1**	**4.1**	2.5–3.5

TABLE II. (Continued)

Distance	Total	Cluster I	Cluster II	Cluster III	NMR-like	NMR restraints
$\varepsilon 3H_6$–αH_7	**5.8**	**6.6**	3.6	**5.3**	4.3	2.9–4.0
$\varepsilon 3H_6$–NH_7	**5.0**	**5.6**	3.3	**4.7**	3.9	2.5–3.5
$\zeta 2H_6$–αH_{12}	**10.3**	**11.4**	**6.9**	**7.3**	**5.5**	2.0–3.0
$\eta 2H_6$–αH_{12}	**11.0**	**12.3**	**7.6**	**8.3**	**7.9**	3.3–5.0
$\eta 2H_6$–$\delta 2H_{12}$	**10.2**	**12.0**	**6.0**	**6.8**	**4.5**	2.5–3.5
$\varepsilon 1H_6$–NH_{16}	**9.3**	**9.3**	3.9	**9.5**	**6.1**	3.3–5.0
$\varepsilon 1H_6$–αH_{17}	**9.6**	**9.5**	**7.2**	**8.2**	3.5	2.9–4.0
$\zeta 2H_6$–αH_{17}	**9.8**	**10.0**	**6.2**	**8.8**	**5.6**	3.3–5.0
$\zeta 2H_6$–$\delta 2H_{18}$	**10.4**	**10.9**	**8.2**	**7.7**	3.8	2.9–4.0
$\zeta 2H_6$–$\delta 3H_{18}$	**10.1**	**10.1**	**9.2**	**9.3**	5.3	3.3–5.0
$\zeta 2H_6$–$\beta 2H_{18}$	**10.9**	**11.7**	**8.8**	**8.5**	4.4	3.3–5.0
$\zeta 2H_6$–$\gamma 2H_{18}$	**11.4**	**12.1**	**10.1**	**8.0**	3.2	3.3–5.0
$\eta 2H_6$–$\beta 2H_{18}$	**10.9**	**11.7**	**8.9**	**8.6**	4.4	3.3–5.0
$\varepsilon 1H_6$–αH_{18}	**9.4**	**9.6**	**7.3**	**8.9**	5.0	2.9–4.0
$\eta 2H_6$–$\gamma 2H_{18}$	**11.2**	**12.0**	**9.7**	**8.1**	4.0	3.3–5.0
$\delta 1H_6$–αH_{18}	**8.9**	**9.2**	**6.7**	**7.4**	4.1	3.3–5.0
$\delta 1H_6$–$\delta 3H_{19}$	**9.4**	**9.6**	**7.7**	**9.5**	**5.8**	3.3–5.0

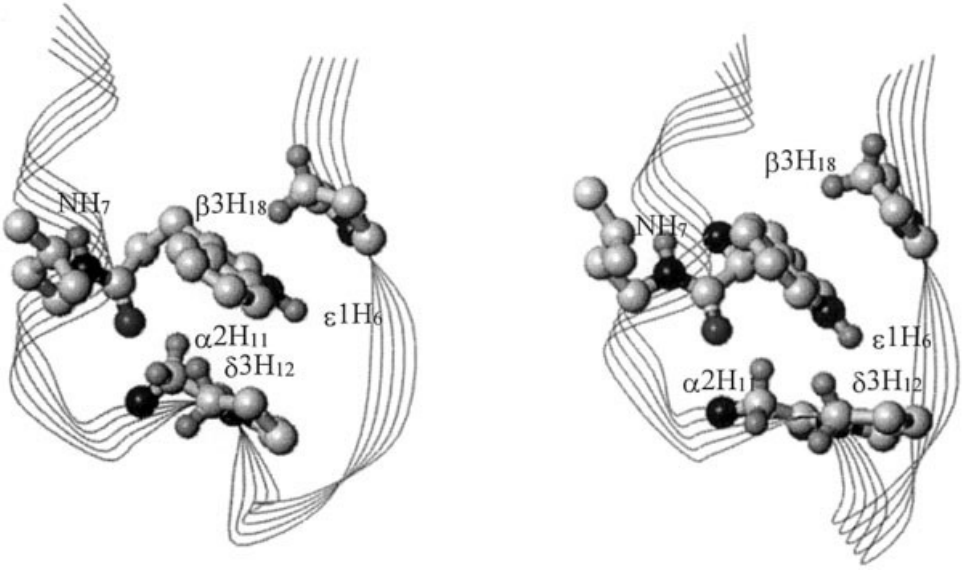

Fig. 6. Detailed comparison of the experimental structure (**left**) and a representative calculated structure. Only residues mentioned in the text are shown. Protons mentioned in the text are labeled.

solvent of $\epsilon 1H_6$ and NH_7.[5] Figure 6 shows a more detailed comparison of the NMR structure to the NMR-like structure closest to the NMR data (RMSD = 1.09 Å). The very subtle difference between them is in the tighter packing of the Trp-cage-forming residues in the NMR structure. One can see that spatial positions of all mentioned protons in the calculated structure completely agree with these experimental observations.

NMR experiments also revealed the dramatic stereospecific chemical shifts for α-protons of Gly_{11} ($\alpha 2H_{11}$ was shifted much more significantly than $\alpha 3H_{11}$), as well as moderate shifting for $\delta 3H_{12}$ during the thermal loss of tertiary structure.[5] Both may be interpreted as an indication of changes in Trp^6–Pro^{12} interactions during TC5b unfolding. These observations also are in good agreement with the predicted folding pathway of TC5b (step II in Fig. 4).

A recent experimental study of folding kinetics of TC5b found that it folds in 4 μs, which is much faster than what has been observed for other small proteins.[32] It was specifically noted[32] that such fast folding cannot be explained in terms of the "contact order" hypothesis, which assumes the early formation of long-range tertiary contacts.[33] Locally driven folding of TC5b could rationalize this experimental observation.

Summarizing, our computational study points out that among the low-energy structures of TC5b found by the independent energy calculations following probable locally driven folding pathways for TC5b, there are structures that agree very well with the experimental data regarding

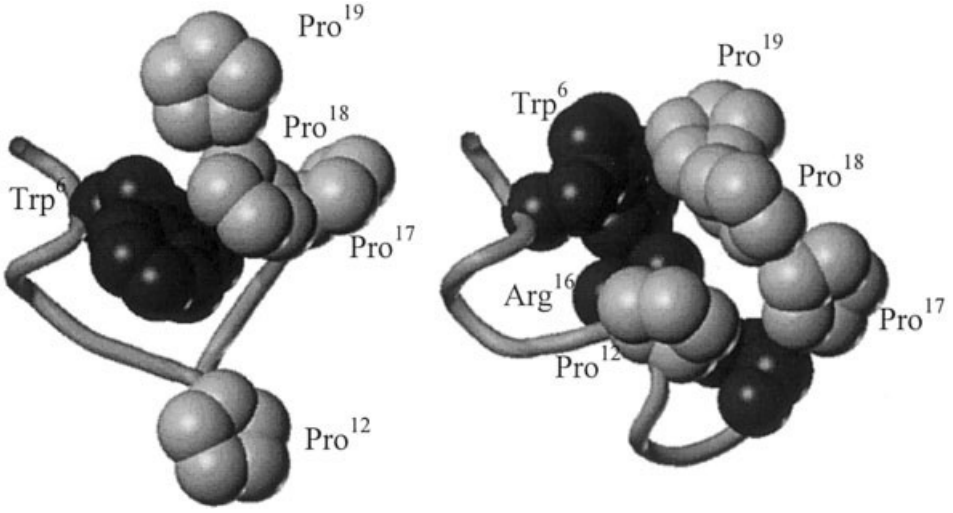

Fig. 7. Low-energy conformers of TC3b (**left**) and TC4c (**right**) most similar to the NMR structure of TC5b. Only the side-chains of the "cage-forming" residues mentioned in text are shown (Trp and Arg in black).

the fold of TC5b in water. However, although structures that are most similar to the NMR-ensemble belong to the predominant cluster I, the energy calculations did not exclude the existence of other structures such as those from clusters II or III. On the other hand, the force field that was used in our calculations did not include the effect of solvent; the only parameter modified to approximate the water environment was the macroscopic dielectric constant (see Materials and Methods). Roughly, solvation effects can be estimated simply by calculation of solvent-accessible surface areas (SASA) for different low-energy conformations of TC5b. Such estimations were performed for the 10 lowest-energy NMR-like structures and for the 10 lowest-energy structures from each clusters II and III. The readily available and commonly used GETAREA computational procedure (the URL address is http://www.scsb.utmb.edu/sgi-bin/get_a_form.tcl) was used. The average SASA values for a water environment were 1620 $\pm$ 20 Å^2, 1742 $\pm$ 14 Å^2 and 1741 $\pm$ 13 Å^2, respectively [i.e., structures in clusters II and III possessed slightly higher (by ca. 7%) SASA values than the NMR-like structures]. At the same time, their "apolar" SASA values (as defined by GETAREA) were by 11% and 16% higher than the same value for the NMR-like structures, mainly due to higher exposure to solvent of Leu^2, Ile^4, and Pro^{17} in cluster II, and Leu^2, Tyr^3, Trp^6, and Pro^{17} in cluster III (see also Fig. 5). Although these are a rough estimates of solvation effects, they do suggest that only the NMR-like structures of cluster I should be favored in water as was observed experimentally.[5]

Another corroboration of our method comes from energy calculations performed for two analogs of TC5b, which show substantially less Trp-cage formation by NMR, namely, Asn^1-Leu^2-Tyr^3-Ile^4-**Glu^5**-Trp^6-Leu^7-Lys^8-**Asn^9**-Gly^{10}-Gly^{11}-Pro^{12}-Ser^{13}-Ser^{14}-Gly^{15}-**Ala^{16}**-Pro^{17}-Pro^{18}-Pro^{19}-Ser^{20} (TC3b) and **Lys^0-Gly^1**-Leu^2-Tyr^3-Ile^4-**Glu^5**-Trp^6-Leu^7-Lys^8-**Asn^9**-Gly^{10}-Gly^{11}-Pro^{12}-Ser^{13}-Ser^{14}-Gly^{15}-Arg^{16}-Pro^{17}-Pro^{18}-Pro^{19}-Ser^{20} (TC4c)[5]; the amino acid replacements in TC3b and TC4c compared to TC5b are shown in bold. The plots of relative propensities for hexapeptides of TC3b and TC4c to adopt the NMR structure characteristic for TC5b are similar to that in Figure 2 (data not shown); the only significant difference is that the peak at fragment 14-19 is less distinct for TC3b. For fragments 4-19 of both TC3b and TC4c, we performed the same procedures of stepwise elongation of the peptide chains toward either the N- or C-termini that have been performed for the TC5b fragment 4-19 (i.e., either from 4-12 to 4-15 to 4-19, or from 14-19 to 10-19 to 4-19).

For TC3b, both types of stepwise elongation found some low-energy conformers geometrically similar (the RMSDs $<$ 1.5 Å) to the NMR structures of TC5b, but, remarkably, these conformers are not of the Trp-cage type, because they lack a crucial Trp-cage interaction, that of Trp^6–Pro^{12} (see Fig. 7, left). And for TC4b, stepwise elongation toward the N-terminus found one low-energy conformer similar to the NMR structures of TC5b (the RMSD $<$ 1.5 Å), where the cage-like "sandwich" formed by Pro^{12} and Pro^{18} is the same as in TC5b, but spatial position of the Arg^{16} side-chain is different. Instead of interaction with the side-chain of Asp^9, which is replaced by Asn^9 in TC4b, the amino groups of the Arg^{16} side-chain form the hydrogen bonds with carbonyls of Trp^6 and Pro^{18}. As a result, the Arg^{16} side-chain prevents the Trp^6 side-chain from entering the cage between Pro^{12} and Pro^{18} (see Fig. 7, right).

CONCLUSIONS

This study presents a computational protocol of stepwise elongation of the peptide chain of the TC5b sequence as a tool to locate and evaluate possible locally driven folding pathways of the TC5b miniprotein. It does not discuss any possibility of folding pathways of TC5b involving early forming of nonlocal long-range contacts. Systematic exploration of the possible pathways showed that the Trp-cage structure of TC5b could be obtained by stepwise elongation starting from the noncentral local nucleation centers pre-existing in the unfolded state of TC5b. Of the locally driven folding pathways that lead to the Trp-cage structure, the

formation of an α-helical fragment 4-9, followed by the formation of the proper 3D structure of fragment 4-12 and then 4-18 appears the most probable. According to this folding pathway, key Trp-cage-stabilizing interactions emerge successively, first Trp^6–Pro^{12}, then Trp^6–Pro^{18}, and then Trp^6–Tyr^3. Early formation of the Trp^6–Pro^{12} interaction is also suggested by NMR data for the partially melted state.

Stepwise elongation of the peptide chain corresponding to the pathway in question resulted in the ensemble of low-energy conformations of TC5b, which contains structures possessing the characteristic fold of the Trp-cage observed for TC5b by NMR spectroscopy. The Trp-cage-like structures found by calculations are in excellent agreement with the NMR experimental data. They satisfy 67 of 74 interproton NOE distance restraints. They are also in good agreement with the observed ring current shifts for selected protons in TC5b.

This novel approach was successfully applied to TC5b. It also explains why two analogs of TC5b, namely TC3b and TC4c, do not form the same Trp-cage structures as exclusively as TC5b. Applications of this approach to other miniproteins should provide more information on its potential and limitations.

ACKNOWLEDGMENTS

The authors thank Dr. Garland R. Marshall for reading the manuscript and for extremely useful discussion and Dr. Alan Grossfield for his help in efforts to estimate solvation energies for TC5b conformers. Grant sponsor: National Institutes of Health; Grant numbers: GM 53630 (G.V.N.), GM 59658 (N.H.A. and R.M.F.), DK 13332 (C.F.)

REFERENCES

1. Dinner AR, Karplus M. Comment on the communication "The key to solving the protein-folding problem lies in an accurate description of the denatured state" by van Gunsteren et al. Angew Chem Int Ed 2001;40:4615–4646.
2. Dahiyatt BI, Mayo SL. De novo protein design: fully automated sequence selection. Science 1997;278:82–87.
3. Jager M, Nguen H, Crane JC, Kelly JW, Gruebele M. The folding mechanism of a β-sheet: the WW domain. J Mol Biol 2001;311:373–393.
4. Kortemme T, Ramirez-Alvarado M, Serrano L. Design of a 20-amino acid, three-stranded β-sheet protein. Science 1998;281:253–256.
5. Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue protein. Nat Struct Biol 2002;9:425–430.
6. Neidigh JW, Fesinmeyer RM, Prickett KS, Andersen NH. Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. Biochemistry 2001;40:13188–13200.
7. Colombo G, Roccatano D, Mark AE. Folding and stability of the three-stranded β-sheet peptide betanova: insights from molecular dynamics simulations. Proteins 2002;46:380–392.
8. Daggett V. Molecular dynamics simulations of the protein unfolding/folding reactions. Acc Chem Res 2002;35:422–429.
9. Wang H, Varady J, Ng L, Sung S-S. Molecular dynamics simulations of β-hairpin folding. Proteins 1999;37:325–333.
10. Simmerling C, Stockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. J Am Chem Soc 2002;124:11258–11259.
11. Zagrovic B, Snow CD, Khaliq S, Shirts MR, Pande VS. Native-like mean structure in the unfolded ensemble of small proteins. J Mol Biol 2002;323:153–164.
12. Zagrovic B, Snow CD, Shirts MR, Pande VS. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. J Mol Biol 2002;323:927–937.
13. Snow CD, Nguen H, Pande VS, Gruebele M. Absolute comparison of simulated and experimental protein-folding dynamics. Nature 2002;420:102–106.
14. Kortemme T, Kelly MJS, Kay LE, Forman-Kay J, Serrano L. Similarities between the spectrin SH3 domain denatured state and its folding transition state. J Mol Biol 2000;297:1217–1229.
15. Hodsdon ME, Frieden C. Intestinal fatty acid binding protein: the folding mechanism as determined by NMR studies. Biochemistry 2001;40:732–742.
16. Kazmirski SL, Wong K-B, Freund SMV, Tan Y-J, Fersht AR, Daggert V. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. Proc Natl Acad Sci USA 2001;98:4349–4354.
17. Wong K-B, Clarke J, Bond CJ, Neira JL, Freund SMV, Fersht AR, Daggert V. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. J Mol Biol 2000;296:1257–1282.
18. Neri D, Billeter M, Wider G, Wuthrich K. NMR determination of residual structure in a urea-denatured protein, the 434-repressor. Science 1992;257:1559–1563.
19. Logan TM, Theriault Y, Fesik SW. Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. J Mol Biol 1994;236:637–648.
20. Sari N, Alexander P, Bryan PN, Orban J. Structure and dynamics of an acid-denatured protein G mutant. Biochemistry 2000;39:965–977.
21. Bai Y, Chung J, Dyson HJ, Wright PE. Structural and dynamic characterization of an unfolded state of poplar apo-plastocyanin formed under nondenaturing conditions. Protein Sci 2001;10:1056–1066.
22. Garcia P, Serrano L, Durand D, Rico M, Bruix M. NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. Protein Sci 2001;10:1100–1112.
23. Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8M urea. Biochemistry 1997;36:8977–8991.
24. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H. Long-range interactions within a nonnative protein. Science 2002;295:1719–1722.
25. Yao J, Chung J, Eliezer D, Wright PE, Dyson HJ. NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. Biochemistry 2001;40:3561–3571.
26. Nikiforovich GV, Frieden C. The search for local native-like nucleation centers in the unfolded state of β-sheet proteins. Proc Natl Acad Sci USA 2002;99:10388–10393.
27. Fersht AR. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. Proc Natl Acad Sci USA 1995;92:10869–10873.
28. Nemethy G, Pottle MS, Scheraga HA. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. J Phys Chem 1983;87:1883–1887.
29. Dunfield LG, Burgess AW, Scheraga HA. Energy parameters in polypeptides. 8. Empirical potential energy algorithm for the conformational analysis of large molecules. J Phys Chem 1978;82:2609–2616.
30. Nikiforovich GV. Computational molecular modeling in peptide design. Int J Pept Protein Res 1994;44:513–531.
31. Nikiforovich GV, Hruby VJ, Prakash O, Gehrig CA. Topographical requirements for delta-selective opioid peptides. Biopolymers 1991;31:941–955.
32. Qiu L, Pabit SA, Roitberg AE, Hagen SJ. Smaller and faster: the 20-residue Trp-cage protein folds in 4 μs. J Am Chem Soc 2002;124:12952–12953.
33. Plaxco KW, Simons KT, Baker D. Contact order transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985–994.