Exploring the trigger sequence of the GCN4 coiled-coil: Biased molecular dynamics resolves apparent inconsistencies in NMR measurements

John H. Missimer,¹ Jožica Dolenc,^{2,3} Michel O. Steinmetz,¹ and Wilfred F. van Gunsteren^{2*}

¹Laboratory of Biomolecular Research, Paul Scherrer Institute, CH-5232 Villigen, Switzerland

²Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, CH-8093, Zurich, Switzerland

³Faculty of Chemistry and Chemical Technology, University of Ljubljana, SI-1000, Ljubljana, Slovenia

Received 27 May 2010; Revised 29 September 2010; Accepted 4 October 2010

DOI: 10.1002/pro.528

Published online 15 October 2010 proteinscience.org

Abstract: Trigger sequences are indispensable elements for coiled-coil formation. The monomeric helical trigger sequence of the yeast transcriptional activator GCN4 has been investigated recently using several solution NMR observables including nuclear Overhauser enhancement (NOE) intensities and $^3J(\text{H}_\text{N}, \text{H}_{\text{C}\alpha})$ -coupling constants, and a set of 20 model structures was proposed. Constrained to satisfy the NOE-derived distance bounds, the NMR model structures do not appear to reproduce all the measured $^3J(\text{H}_\text{N}, \text{H}_{\text{C}\alpha})$ -coupling constant values, indicating that the α -helical propensity is not uniform along the GCN4 trigger sequence. A recent methodological study of unrestrained and restrained molecular dynamics (MD) simulations of the GCN4 trigger sequence in solution showed that only MD simulations incorporating time-averaged NOE distance restraints and instantaneous or local-elevation 3J -coupling restraints could satisfy the entire set of the experimental data. In this report, we assess by means of cluster analyses the model structures characteristic of the two simulations that are compatible with the measured data and compare them with the proposed 20 NMR model structures. Striking characteristics of the MD model structures are the variability of the simulated configurations and the indication of entropic stability mediated by the aromatic N-terminal residues 17Tyr and 18His, which are absent in the set of NMR model structures.

Keywords: NMR structure determination; NOE upper bounds; 3J -coupling constants; GCN4p trigger site; molecular dynamics; GROMOS; time averaging; local-elevation biasing

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Swiss National Science Foundation (SNSF); Grant number: 200020-121913; Grant sponsor: European Research Council (ERC); Grant number: 228076; Grant sponsor: Slovenian Research Agency (ARRS); Grant number: Z1-9576.

*Correspondence to: Wilfred F. van Gunsteren, Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, CH-8093 Zürich, Switzerland. E-mail: wfvgn@igc.phys.chem.ethz.ch

Introduction

The α -helical coiled coil is the most frequently encountered subunit oligomerization motif in intracellular proteins and is thus involved in many biological processes.¹ This structural motif offers the simplest system for studying both the intramolecular and intermolecular interactions that govern the folding and stability of multisubunit proteins.² It is well acknowledged that trigger sequences, autonomous helical folding units functionally conserved in a large number of native coiled-coil proteins, are indispensable for coiled-coil formation.^{2,3} Despite their

Figure 1. Deviations of model from the experimentally derived NOE upper distance bounds as a function of the NOE sequence number (left-hand panel, A) and comparison of the experimental and calculated $^3J(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constants (right-hand panel, B) for the 20 NMR model structures. The corresponding deviations (left-hand panel, C) and comparisons (right-hand panel, D) for the restrained simulations: NOE_TAR+ 3J _IR (black) and NOE_TAR+ 3J _LE (red).⁶ The corresponding deviations (left-hand panel, E) and comparisons (right-hand panel, F) for the weighted averages of the central members of the first 10 conformational clusters of the NOE_TAR+ 3J _LE simulation. The 179 NOEs and 15 3J -couplings are defined in the study of Dolenc *et al.*⁶

importance, the basis of trigger sequence structure and function is not well understood at the atomic level and remains an active field of investigation.

One of the best characterized trigger sequences is that of the parallel two-stranded leucine zipper coiled coil of the yeast transcriptional activator GCN4, denoted GCN4-p1. Recently, a multidisciplinary approach including NMR, time-resolved circular dichroism (CD) spectroscopy, and mutagenesis has proposed features of the trigger sequence that determine coiled-coil formation of GCN4-p1.⁴ The 20 model structures derived from the NMR measurements include a network of hydrogen bonds and salt bridges comprising a nonclassical cap structure at the N-terminus of the trigger sequence peptide, denoted GCN4p16-31.

In contrast to X-ray crystallography, NMR measurements do not imply atomic densities. The relatively low number of constraints provided by the measurements must be supplemented by assumptions. The solution NMR structure of the GCN4p16-31 trigger peptide presented as a set of 20 model structures obtained using a single-structure refine-

ment-simulated annealing approach with the program X-PLOR⁵ was based on 172 distance restraints derived from measured nuclear Overhauser enhancements (NOEs), and 25 assumed, standard α -helical restraints suggested by the measured $^3J(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constants, and secondary $\text{C}\alpha$ and $\text{H}\alpha$ chemical shifts.⁴

Analysis of the NOE distances and $^3J(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constants performed on the NMR model structures, displayed in Figure 1(A and B), showed that these basically satisfy the set of NOE distance bounds, but the $^3J(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constants calculated for the residues 18His, 19Leu, and 23Val deviate from the measured ones by more than 1.5 Hz.⁶

The deficiencies of this refinement protocol prompted us to perform a series of unrestrained and restrained molecular dynamics (MD) simulations of GCN4p16-31 in solution using the thermodynamically calibrated GROMOS force fields at room temperature and explicit solvation. Comparison of the NOE distances and $^3J(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constants calculated from the simulated MD trajectories with the primary, measured NMR data provided the

Figure 2. Time evolution of secondary structure²¹ and occupation frequencies for each residue in restrained simulations: NOE_TAR+³J_IR (top panel) and NOE_TAR+³J_LE (lower panel). The color code is displayed on the upper right of the top panel.

quality criteria. We concluded that only simulations incorporating both NOE distance restraints effected via time-averaging and instantaneous or local-elevation ³J-value restraints could satisfy the entire set of the experimental data, Figure 1(C and D).⁶ This report presents the structural implications of these simulations, especially concerning the network of hydrogen bonds, salt bridges, and polar contacts at the N-terminus of GCN4p16-31. We advocate the results of the simulations incorporating local-elevation biasing because they enhance sampling by enabling barrier transitions and compensate force-field deficiencies.

Results

NMR model structures

The solution NMR structure of the C-terminal peptide of GCN4p16-31⁴ is represented as a set of 20 model structures computed using a simulated annealing approach with the program X-PLOR.⁵ Analysis of the NOE distances and ³J(H_N-H_{C α})-coupling constants performed on the NMR model structures shows that these satisfy the set of NOE distance bounds with minor violations not exceeding 0.1 nm associated with the following proton pairs:

H_N-H _{α} of the residues 21Asn and 17Tyr, H_N-H _{γ} of the residues 21Asn and 25Arg, H _{ϵ} -H _{δ} and H _{δ} -H _{δ} of the residues 18His and 19Leu, H _{δ} -H _{β} and H _{δ} -H _{γ} of the residues 18His and 20Glu, and H _{ϵ} -H _{γ} of the residues 17Tyr and 20Glu [Fig. 1(A)]. However, a comparison of the ³J(H_N-H_{C α})-coupling constants calculated from the set of 20 NMR model structures with the corresponding experimental values shows that the calculated ³J(H_N-H_{C α})-coupling constants for the residues 18His, 19Leu, and 23Val deviate from the measured ones by more than 1.5 Hz, that is, by 4.0, 3.4, and 1.8 Hz, respectively [Fig. 1(B)].

NOE_TAR+³J_IR and NOE_TAR+³J_LE simulations

As is evident from the two panels in Figure 1C and D, both refinement protocols, NOE_TAR+³J_IR and NOE_TAR+³J_LE, successfully reproduce NOE distance bounds and ³J-coupling constants.⁶ We note that the deviation of 1 Hz for the NOE_TAR+³J_LE ensemble in Figure 1D (red dots) reflects the ± 1 Hz flat-bottom function used in ³J-LE biasing. The differences in the conformational space sampled by NOE_TAR+³J_IR and NOE_TAR+³J_LE simulations are reflected in the secondary structure analysis

Figure 3. Time evolution of hydrogen bonds and of major clusters observed in restrained simulations: NOE_TAR+ $^3\text{J}_{\text{IR}}$ (top panel) and NOE_TAR+ $^3\text{J}_{\text{LE}}$ (lower panel). For the hydrogen bonds, green denotes $i, i-4$ main-chain bonds; dark green, $i, i-3$ main-chain bonds; cyan, $i, i-2$ main-chain bonds; blue, $i, i-5$ main-chain bonds; and indigo, bonds involving side chains.

presented in Figure 2. When ^3J -coupling constants are enforced as instantaneous restraints, GCN4p16-31 exhibits α -helical propensities exceeding 80% in six mid peptide residues and 60% in 10 of the 16 residues. Only 18His evidences a dominant nonhelical secondary structure. On the other hand, applying ^3J -coupling constant restraints via local elevation yields a decrease of the total helical propensity to less than 60% in all residues, and the nonhelical secondary

structure-denoted bend becomes dominant in four residues near the terminal ends. Moreover, helical secondary structures associated with the 3_{10} -helix occur with increased frequency near the N-terminal end.

The prominence of α -helical conformations in the NOE_TAR+ $^3\text{J}_{\text{IR}}$ simulation is also evident in the analysis of hydrogen bonds (Fig. 3 and Table I) which reveals dominant backbone hydrogen bonds, $i, i-4$ characteristic of α -helices, between most mid

Table I. Analysis of Hydrogen Bonds

Donor	Acceptor	NMR model	IR	IR 1	IR 2	LE	LE 1	LE 2
		20	10,000	5792	2353	10,000	3190	1379
16Asn NH	17Tyr O					16	38	
16Asn N H	18His ND1		9	12		19		55
16Asn ND2 HD22	20Glu OE1	5						
17Tyr N H	18His ND1		12	17				9
18His CD2 HD2	17Tyr O					22	58	
18His NE2 HE2	22Glu OE1		17	21	10	5		9
18His NE2 HE2	22Glu OE2		13	19	6	8		27
19Leu NH	16Asn O		6		22			
19Leu NH	17Tyr O		8		24			
19Leu NH	His18 ND1		11	15		11		40
19Leu NH	22Glu OE1					6		
19Leu NH	22Glu OE2					5		
20Glu NH	16Asn O	55	5		18			
20Glu NH	17Tyr O	100			9	9	17	
20Glu NH	18His O		22	25	17	28	19	51
21Asn NH	18His O		57	75	27	31	34	31
21Asn NH	20Glu OE1					7	16	8
21Asn NH	20Glu OE2					9	13	6
21Asn ND2 HD22	20Glu OE1				6	16	40	
21Asn ND2 HD22	20Glu OE2					17	38	
22Glu NH	18His O	100	30	12	73	16	28	11
22Glu NH	19Leu O		24	32	6	12	11	29
22Glu NH	20Glu O					33	26	24
23Val NH	19Leu O	100	48	42	64	43	55	40
23Val NH	20Glu O		10	11	7	7	7	7
24Ala NH	20Glu O	10	28	27	26	33	24	35
24Ala NH	21Asn O		6	6	8	10	10	12
24Ala NH	22Glu O					5	7	
25Arg NH	21Asn O	100	95	95	97	33	38	53
25Arg NE HE	21Asn OD1	10	9	7	13	6	6	8
25Arg NH2 HH22	21Asn O					6	5	7
25Arg NH2 HH22	21Asn OD1		6	5		7		10
25Arg NH	22Glu O					9		
25Arg NH2 HH22	22Glu OE1		11	12	10	5	6	9
25Arg NH2 HH22	22Glu OE2		8	9	7		5	10
25Arg NH	23Val O					8		
26Leu NH	22Glu O	100	93	93	94	40	31	62
26Leu NH	23Val O					17	17	11
26Leu NH	24Ala O					6	9	
27Lys NH	23Val O	90	71	72	76	38	38	39
27Lys NH	24Ala O		6	9		10	6	23
27Lys NH	25Arg O					9	9	8
28Lys NH	23Val O					6		
28Lys NH	24Ala O	10	30	35	23	15	15	18
28Lys NZ HZ2	24Ala O	5						
28Lys NH	25Arg O	20	21	26	13	18	20	20
28Lys NH	26Leu O		19	16	26	14	19	15
29Leu NH	24Ala O		7		17		5	
29Leu NH	25Arg O	80	39	42	39	12	12	19
29Leu NH	26Leu O		11	12	9	9	13	7
29Leu NH	27Lys O					7	9	
30Val NH	25Arg O		5					
30Val NH	26Leu O	30	64	73	50	40	48	39
31Gly NH	26Leu O	5	19	13	37	14	28	
31Gly NH	27Lys O	10	11	16	9		5	
31Gly NH (C-ter)	30Val O	15						

In the header, the specification of the bond, the simulation, and in the case of the restrained simulations, IR and LE, the two dominant clusters, denoted 1 and 2, and the respective number of configurations analyzed are listed. The hydrogen bonds were defined by the condition that the donor-H-acceptor angle exceed 135° and the H-acceptor distance be less than 0.25 nm. Only those bonds are listed for which the frequency of occurrence in the complete simulations exceeded 5%; the percentages quoted for the clusters are relative to the respective number of configurations. Bonds characteristic of the clusters are shaded in gray.

Table II. *Analysis of Salt Bridges*

Positive charge	Negative charge	NMR model	IR	IR 1	IR 2	LE	LE 1	LE 2
		20	10,000	5792	2353	10,000	3190	1379
25Arg HE	22Glu OE1	70	53	56	53	39	47	50
25Arg HE	22Glu OE2	50	49	53	43	36	46	59
25Arg HH11	20Glu OE1	5	0	0	0	0	0	0
25Arg HH11	22Glu OE1	45	27	29	26	14	17	24
25Arg HH11	22Glu OE2	30	22	26	17	14	16	29
25Arg HH12	22Glu OE1	70	20	20	23	12	15	16
25Arg HH12	22Glu OE2	50	17	18	16	12	12	22
25Arg HH21	22Glu OE1	40	55	59	53	36	42	50
25Arg HH21	22Glu OE2	25	51	57	44	36	43	59
25Arg HH22	22Glu OE1	40	77	76	70	55	64	70
25Arg HH22	22Glu OE2	40	74	74	63	56	67	78

In the header, the specification of the contact, the simulation, and in the case of the restrained simulations, IR and LE, the two dominant clusters, 1 and 2, with the respective number of configurations analyzed are listed. The salt bridges were defined by the condition that the distance between positive and negative charges be less than 0.60 nm. Only those contacts are listed for which the frequency of occurrence in the complete simulations exceeded 5%; the percentages quoted for the clusters are relative to the respective number of configurations.

peptide residues. These hydrogen bonds are even more prominent in the NMR model structures. However, *i*, *i* – 3 hydrogen bonds characteristic of 3_{10} -helices and *i*, *i* – 5 hydrogen bonds characteristic of π -helices, which are absent with three exceptions in the NMR model structures, also occur with notable frequencies of occurrence. These latter hydrogen bonds appear with increased frequency and number of residues involved in the NOE_TAR+³J_LE simulation, confirming increased flexibility of the backbone. Comparison of the trajectories of the two simulations in Figure 3 further confirms this assessment, suggesting a continuous breaking and forming of hydrogen bonds in the latter simulation. The analysis of hydrogen bonds also evidences a proliferation of interactions between backbone and side chains and between side chains in the two restrained simulations not found in the NMR model structures. Two of three such hydrogen bonds found in the NMR model, between 16Asn and 20Glu and between 28Lys and 24Ala, are not among the 17 hydrogen bonds observed in the simulations. Among the 17 hydrogen bonds are three involving the aromatic ring of 18His, which mediate hydrogen bonds with 16Asn, 17Tyr, and 19Leu, four involving the side-chain oxygens of 20Glu with main- and side-chain hydrogens of 21Asn, and six involving the side-chain oxygens of 22Glu with side chains of 18His and 25Arg and with the main chain of 19Leu.

The salt bridges between 25Arg and 22Glu exhibited by the NMR model structures also emerge in the restrained simulations (Table II). The NMR model structures yield bridges involving protons of the NH1 atom with higher occurrence frequencies than the restrained simulations; the latter yield bridges involving protons of the NH2 atom with higher frequencies than the NMR model structures. The NOE_TAR+³J_IR simulation yields frequencies

consistently larger than the NOE_TAR+³J_LE simulation. Salt bridges between 25Arg and 20Glu occur in the restrained simulations with frequencies of occurrence not exceeding 1.6%. The distance distributions of the dominant salt bridges in the restrained simulations (not shown) are broad, ranging between 0.2 and 1 nm with distinct peaks at 0.5 nm.

The polar contacts, particularly those involving the tyrosine residue, discriminate very clearly between the NMR model structures and the restrained simulations (Table III and Fig. 4). In the NMR model structures, side-chain hydrogens of the tyrosine residue interact frequently with nitrogens of the histidine residue, the main-chain tyrosine hydrogen with the main-chain nitrogen of 16Asn, and the main-chain tyrosine oxygen with the main-chain hydrogens of 19Leu and 21Asn. These interactions are absent or much less frequent in the restrained simulations in which the main-chain atoms of tyrosine interact rather with 16Asn and 18His. The main- and side-chain atoms of 18His evidence numerous interactions in the restrained simulations not found in the NMR model structures. As the distance distributions of Figure 4 show for the NOE_TAR+³J_LE simulation, the nonhydrogen bonded polar contacts, with one exception, exhibit little deviation from 0.3 nm, suggesting rigid contacts despite the flexibility of the backbone. In contrast, the contacts designated hydrogen bonds appear to be quite broad with secondary peaks in the distributions at about 0.6 nm and extending to 1 nm. These observations are also valid for the NMR model structures and NOE_TAR+³J_IR simulation.

To illustrate the conformational differences among the set of NMR model structures and the ensembles of the NOE_TAR+³J_IR and NOE_TAR+³J_LE simulations, a joint conformational cluster analysis of heavy side-chain atoms is summarized in Figure 5.

Table III. Analysis of Electrostatic Contacts Between N-terminal residues (16–22 and 25) not Included in the List of Hydrogen Bonds or Salt Bridges

Positive partial charge	Negative partial charge	NMR model	IR	IR 1	IR 2	LE	LE 1	LE 2
		20	10,000	5792	2353	10,000	3190	1379
16Asn HD22	17Tyr N	10	1	0	2	0	0	0
16Asn H	18His ND1		14	20	0	23	3	66
17Tyr H	16Asn N	65	15	13	22	2	0	9
17Tyr H	16Asn O		33	32	21	80	89	44
17Tyr HD1	16Asn OD1		24	39	1	2	3	4
17Tyr HD1	18His N	45	1	0	3	1	0	0
17Tyr HD2	18His N	55	0	0	0	2	3	0
17Tyr HH	18His ND1	55	3	0	2	2	4	0
18His H	16Asn O		10	0	33	6	9	0
18His H	17Tyr O		73	100	15	34	12	77
18His HD2	19Leu N		42	40	53	32	35	39
19Leu H	17Tyr N		18	28	4	2	0	2
19Leu H	17Tyr O	100	13	0	40	5	4	4
19Leu H	18His N	100	36	18	72	15	6	28
19Leu H	18His NE2		10	13	2	12	3	44
19Leu H	18His O		16	11	22	54	65	20
19Leu H	20Glu N		13	8	21	9	3	11
20Glu H	19Leu N	100	94	97	91	87	96	90
20Glu H	19Leu O		8	7	6	20	28	10
20Glu H	21Asn N		22	22	20	36	32	24
21Asn H	17Tyr O	100	5	0	15	4	4	8
21Asn H	18His O	100	85	89	86	61	87	47
21Asn H	19Leu O		10	12	4	11	4	13
21Asn H	20Glu N	100	92	97	85	85	77	95
21Asn HD22	20Glu OE1	5	4	1	7	23	54	1
21Asn HD22	20Glu OE2		3	2	6	24	54	2
21Asn H	22Glu N		27	25	32	21	22	22
21Asn HD21	25Arg NE	25	0	0	0	0	0	0
22Glu H	19Leu O	100	67	74	56	40	45	63
22Glu H	20Glu O		10	12	4	41	35	32
22Glu H	21Asn N	100	99	100	100	83	86	95
22Glu H	21Asn O		8	8	5	18	18	10
25Arg H	22Glu O		47	51	39	48	40	56

In the header, the specification of the contact, the simulation, and in the case of the restrained simulations, IR and LE, the two dominant clusters, 1 and 2, with the respective number of configurations analyzed are listed. The electrostatic contacts were defined by the condition that the distance between positive and negative partial charges be less than 0.30 nm. Only those contacts are listed for which the frequency of occurrence in the complete simulations exceeded 10%; the percentages quoted for the clusters are relative to the respective number of configurations. Contacts characteristic of the clusters are shaded in gray.

The analysis shows that the first conformational cluster is dominated by NMR model configurations, the second by configurations of the NOE_TAR+³J_IR simulation, and the third by configurations of the NOE_TAR+³J_LE simulation. These three sets of structures show similar features in the C-terminal part of the peptide, while differing in the N-terminal part. Reflecting the variability of the NMR model structures and NOE_TAR+³J_IR ensemble, the first cluster includes 95% of the NMR configurations and the second cluster 57% of the NOE_TAR_3J_IR configurations; the partitioning into two clusters indicates the distinction between the two sets of configurations. In contrast, the first three clusters include less than 50% of the NOE_TAR_3J_LE ensemble, confirming its conformational variability.

Performed to deduce a few model structures characteristic of the restrained simulations, separate conformational cluster analyses of the side-

chain atoms of the restrained simulations produced two clusters representing 81% of the NOE_TAR+³J_IR trajectories and 10 clusters representing 75% of the NOE_TAR+³J_LE trajectories, of which the first two describe 46% and the succeeding three 6% each. In the Tables and Figures we denote as IR1 and IR2 and as LE1 and LE2, the two most populated clusters of the trajectories NOE_TAR+³J_IR and NOE_TAR+³J_LE, respectively. The occurrence of trajectory structures included in these clusters is displayed in panels below the hydrogen bonds in Figure 3. Representations of the central members of the two IR and 10 LE most-populated clusters (coordinates deposited in the Protein Data Bank, PDB ID: 2L56) with specification of prominent hydrogen bonds are shown in Figure 6. The individual NOE distance bound violations and ³J-coupling constants for the 10 central members are shown in Supporting

Figure 4. Distance distributions of selected hydrogen bonds (Table I) and polar contacts (Table III) observed in restrained simulation NOE_TAR+³J_LE.

Information. The sizeable NOE distance bound violations and ³J-coupling constants in the individual model structures differ among each other and so manifest the extent of conformational space explored by the local-elevation search. The average NOE distance bound violations and ³J-coupling constants for the 10 central members are shown in Figure 1(E and F); the averages are weighted by the occurrence frequencies of the corresponding clusters. As expected, the agreement with the measured data improves but does not approach the level of the complete ensemble [Fig. 1(C and D)].

For NOE_TAR+³J_IR, the first cluster appears associated with the occurrence of hydrogen bonds between the aromatic hydrogens of 18His with the side-chain oxygens of 22Glu and the aromatic nitrogens of 18His with 16Asn, 17Tyr, and 19Leu. The i, i – 3 hydrogen bond between the main-chain atoms of 21Asn and 18His is also evident. Polar contacts between the side chains of tyrosine and the adjacent asparagine and between the main chains of tyrosine and histidine are prominent. The network of contacts suggested by Tables I and III and Figure 3 is reflected in Figure 7. The second cluster describing

Figure 5. Populations of the first 15 conformational clusters obtained from joint cluster analysis of the 10⁴ NOE_TAR+³J_IR trajectory structures (black), the 10⁴ NOE_TAR+³J_LE trajectory structures (gray), and 500 copies of the 20 NMR model structures (hatched) using a side-chain heavy atom-positional root mean square deviation (RMSD) of 0.25 nm as similarity criterion.

Figure 6. Views of the central members of the two major conformational clusters of the restrained simulation NOE_TAR+³J_IR and of the first 10 conformational clusters of the restrained simulation NOE_TAR+³J_LE. The frequency of occurrence of cluster structures and their most populated hydrogen bonds are specified.

NOE_TAR+³J_IR appears associated with the occurrence of *i, i - 3* and *i, i - 4* main-chain hydrogen bonds characteristic of 3_{10} - and α -helices: 19Leu-16Asn, 20Glu-16Asn, 22Glu-18His, and 23Val-19Leu. However, the *i, i - 2* hydrogen bond between 19Leu-17Tyr and *i, i - 5* bonds between 29Leu-24Ala and 31Gly-26Leu also occur. Prominent polar contacts include interactions between 16Asn and 18His and between 17Tyr or 18His and 19Leu. Transitions between the two clusters appear associated in Figure 3 with dissolution of hydrogen bonds between the main-chain atoms of 16Asn and 17Tyr and the aromatic nitrogen ND1 of 18His and formation of hydrogen bonds between the main-chain atoms of the former two residues with those of 19Leu and 20Glu. The network of contacts is shown in Figure 7.

The occupancy patterns of NOE_TAR+³J_LE are less distinct than those of NOE_TAR+³J_IR, but the first cluster appears associated with the prominence of hydrogen bonds between the main-chain hydrogens of 16Asn and 20Glu or side-chain nitrogen of 18His and the main-chain oxygen of 17Tyr, as well as between the side-chain atoms of 21Asn and 20Glu. The fifth cluster also appears related to this set of bonds. Polar interactions between 16Asn and

17Tyr as well as between 18His and 19Leu are also notable. In contrast to the tyrosine-centred network of the first and fifth clusters, the second cluster appears centred on the histidine residue with prominent hydrogen bonds between the main-chain hydrogens of 16Asn and 19Leu and the aromatic nitrogens of 18His, between the side-chain atoms of 18His and 22Glu and between main-chain atoms of 20Glu and 18His. The third and fourth clusters also appear related to this set of hydrogen bonds. Polar contacts between the main chains of 18His and 17Tyr and between a side-chain nitrogen of 18His and main chain of 19Leu are also notable. Transitions between clusters appear associated in Figure 3 with dissolution of a pair of hydrogen bonds between the main chain of 16Asn and side chain of 18His and the main chain of 17Tyr with formation of hydrogen bonds between the main chain of 16Asn and 19Leu with the side chain of 18His. The networks of contacts characterizing the two clusters suggested by the Tables and Figure 3 are shown in Figure 7.

Discussion

Motivated by a discrepancy between the measured ³J-coupling constants and those yielded by the NMR model structures based on NOE and other distance

Figure 7. Views of the networks of hydrogen bonds and polar contacts for the restrained simulations NOE_TAR+³J_IR and NOE_TAR+³J_LE. The panel on the upper right shows in the same orientation the central members of the two major conformational clusters of NOE_TAR+³J_IR; the panel on the upper left shows the same two structures rotated by 180°. The two lower panels show the identical views of the two dominant conformational clusters of NOE_TAR+³J_LE.

and torsional angle restraints, we performed simulations of the GCN4p16-31 coiled coil trigger peptide incorporating two methods of ³J-coupling constant restraining to reproduce the observables: instantaneous restraining and local-elevation search. The NOE distance restraints were imposed by time averaging in both simulations, because instantaneous NOE distance restraining failed to reproduce the measured data.⁶ We have shown previously that neither time-averaged NOE restraints nor ³J-coupling constant restraints are alone sufficient in MD structure refinement protocols to reproduce the entire set of 194 experimental NMR data.⁶ These failures are due to force-field deficiencies or to energetic barriers that inhibit the proper sampling of conformational space. Therefore, the application of local-elevation biasing serves a dual function in a simulation: (1) enhancing sampling by enabling the transition of barriers much larger than $k_B T$; (2) compensating

force-field deficiencies by building up local Gaussian potential energy surfaces.

The differences between the NMR solution structures and the configurations produced by the NOE_TAR+³J_IR simulation, shown in the Tables and in Figure 5, are most likely due to the assumption of standard α -helical hydrogen bonds and ϕ -angle restraints in the single-structure refinement that produced the NMR model structures. These assumptions bias the sampling of the GCN4p16-31 conformational space towards α -helical configurations, producing a set of closely related structures, which produce weak discrepancies with the measured data in the NOE bounds and marked discrepancies in the ³J-coupling constants for just those N-terminal residues purported to be essential for understanding the cap structure of the trigger. In fact, the NOE_TAR+³J_IR simulation also favors mid peptide α -helical configurations but is consistent

with the measured data and implies a cap structure quite different from the one seen in the NMR model structures.

As the Tables and Figure 5 also show, the conformational space of the backbone of GCN4p16-31 sampled by the NOE_TAR+³J_LE refinement is notably larger than those of the NOE_TAR+³J_IR or conventional single-structure NMR refinements. Thus, the NOE_TAR+³J_LE simulation demonstrates that the experimentally derived NOE distance bounds and ³J-coupling constants do not restrict GCN4p16-31 to a rigid α -helical conformation and, in fact, permit substantial flexibility in the backbone.

Implications concerning the structure of GCN4p16-31 are found in the conformational cluster analyses. Salt bridges between the side chains of 22Glu and 25Arg are prominent features in the NMR model structures and both simulations, although markedly less prominent in NOE_TAR+³J_LE. The hydrogen bond between the side chain of 25Arg and 21Asn regarded as essential to establish the network in the NMR model structures occurs rarely in the simulations, whereas the *i, i* – 4 hydrogen bond between the main chains of the two residues occurs frequently in the NMR model structures and the simulations, as do those between 24Ala and 20Glu and between 23Val and 19Leu. Cluster analyses of the two simulations evidence N-terminal networks that deviate notably from those in the NMR model structures and from each other. A remarkable feature of both simulations is the transition between networks, characteristic of a conformational cluster of structures. The transition in the NOE_TAR+³J_IR appears associated with the dissolution of hydrogen bonds between the main-chain atoms of 16Asn and 17Tyr and the aromatic nitrogen ND1 of 18His and formation of hydrogen bonds between the main-chain atoms of 19Leu and 20Glu with the backbone of the former two residues. The transition in the NOE_TAR+³J_LE appears associated with the dissolution of a pair of hydrogen bonds between the main chain of 16Asn and the side chain of 18His with the main chain of 17Tyr and creation of hydrogen bonds between the main chain of 16Asn and 19Leu with the side chain of 18His. The networks corresponding to the two conformational clusters are centred on the tyrosine and histidine, respectively. The transitions occurring in both simulations suggest an entropic stability reminiscent of the salt bridge networks observed in a study of thermal stability in a trimeric coiled-coil protein,⁷ which could not be captured by the NMR model structures. This result illustrates that single-structure refinement artificially restricts molecular motion. It also illustrates that single-structure refinement based on measured NMR data, in combination with atom–atom distance and torsional angle

restraints that do not have a strong basis in measured data, may lead to a set of NMR model structures that do not capture the Boltzmann ensemble of conformations.

Our results contribute to a better understanding of the role of electrostatic interactions and hydrogen bonds in the stability of the monomeric, helical GCN4-p1 trigger sequence. The analyses indicate that the essential role of the salt bridges between 25Arg and 22Glu is complemented by the networks of hydrogen bonds and polar contacts involving 18His and 17Tyr. The networks involve not only interactions between side chains discussed by Steinmetz *et al.*⁴; hydrogen bonds and additional polar interactions, some quite inflexible, between main chain atoms and between main and side chains are equally prominent. The networks formed around these crucial residues are complex, and the fluctuations between networks in both simulations suggest an entropic stability that exploits this complexity. Moreover, the simulation using local-elevation search evidences stability despite notable deviations from helical structure which may facilitate coiled-coil formation.

Methods

MD simulations

The MD simulations reported in this article were carried out using the GROMOS biomolecular simulation package^{8–10} and the 53A6^{11,12} GROMOS force-field parameter set. The GCN4p16-31 peptide comprises the sequence: Ac-16Asn-17Tyr-18His-19Leu-20Glu-21Asn-22Glu-23Val-24Ala-25Arg-26Leu-27Lys-28Lys-29Leu-30Val-31Gly-NH₂. The His residue is protonated at NE2, the Arg and Lys side chains are protonated with charge +e. Coordinates of the first model structure of the NMR set of structures (PDB entry 2ovn) were taken as the starting coordinates for MD simulations. The last residue (32Glu) of the model structure was removed because it was not present in the NMR experiment.⁴ The simulations were performed using periodic boundary conditions. After steepest descent energy minimization, the structure was solvated in a rectangular box of approximately 3000 pre-equilibrated simple point charge water molecules¹³ with a minimal solute-to-wall distance of 1.0 nm. The system was relaxed by performing a steepest-descent energy minimization with harmonic positional restraints on all solute atoms (force constant 2.5×10^4 kJ mol^{–1} nm^{–2}) followed by a 100-ps long equilibration, in which the positional restraints were gradually released reducing the force constant to 0.0 kJ mol^{–1} nm^{–2} and the temperature was raised from 60 to 278 K. The initial atomic velocities were taken from a Maxwell distribution at 60 K. The equations of motion were

integrated using the leap-frog algorithm with a time step of 2 fs. Centre of mass motion was stopped every 2 ps. Bond lengths of the peptide and the geometry of the water molecules were constrained by applying the SHAKE algorithm¹⁴ with a relative geometric tolerance of 10^{-4} . The temperature and pressure were maintained at 278 K and 1 atm using the Berendsen thermostat with a coupling time $\tau_T = 0.1$ ps and barostat with a coupling time $\tau_P = 0.5$ ps and an isothermal compressibility of 4.575×10^{-4} ($\text{kJ mol}^{-1} \text{nm}^{-3}$)⁻¹.¹⁵ A reaction-field approach was used to treat the electrostatics using a triple-range cutoff scheme, with cutoffs of 0.8 and 1.4 nm and a dielectric permittivity of 66.6.¹⁶ The pairlist was updated every five steps. The 179 NOE distance restraints in which GROMOS pseudoatom corrections⁸ were included and 15 $^3\text{J}(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constant restraints are presented and discussed in detail elsewhere.⁶ Time-averaged NOE distance restraints were imposed with a force constant of $6000 \text{ kJ mol}^{-1} \text{nm}^{-2}$ and a memory relaxation time τ_{qr} of 20 ps.^{17,18} The $^3\text{J}(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constant instantaneous restraints were imposed with a force constant of $10 \text{ kJ mol}^{-1} \text{Hz}^{-2}$; the LE ^3J -coupling biasing used a memory relaxation time τ_{qr} of 5 ps.¹⁹ Additionally, due to the uncertainty in the $^3\text{J}(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constants calculated from the corresponding ϕ angles via the Karplus relation, a flat-bottom restraining energy term with a 2 Hz wide well ($\Delta J = 1 \text{ Hz}$) was used in the LE ^3J -coupling biasing simulations.¹⁹ The number of LE Gaussian functions per dihedral angle was set to $N_{\text{le}} = 36$, and the restraints were imposed with a force constant $K^{\text{Jres}} = 0.005 \text{ kJ mol}^{-1} \text{Hz}^{-4}$. The parameters of the Karplus relation are specified below.

From the equilibrated structure, a MD simulation of 1 ns duration was started in which the distance restraints were imposed as instantaneous restraints. From the final structure of that simulation, two 10-ns long MD simulations were started using the 53A6 force field in which NOE distance restraints were imposed as time-averaged restraints, and $^3\text{J}(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constant restraints were imposed either as instantaneous restraints, NOE_TAR+ $^3\text{J}_{\text{IR}}$, or restraints using time averaging together with LE biasing of the conformational search, NOE_TAR+ $^3\text{J}_{\text{LE}}$.

Analysis

The trajectory configurations were saved every 0.5 ps. Interproton distances derived from the NOE cross-peak intensities were compared with the average interproton distances calculated from the simulated and NMR model structures using $\langle r^{-6} \rangle^{-1/6}$ averaging. The results are presented as distance bound violations, that is, as a difference between the distances averaged over the simulation and the corresponding NMR-derived upper distance bounds.

Because the GROMOS force fields make use of united atoms, positions of aliphatic hydrogen atoms of interest were constructed based on standard geometries.⁸ If a NOE upper bound involved nonstereospecifically assigned protons, a pseudoatom was constructed.⁸ The pseudoatom bound corrections used in the original NOE list⁴ were subtracted from the upper bounds and the GROMOS pseudoatom bound corrections were applied. The list of NOE hydrogen-atom pairs, the corresponding NOE upper bounds and violations calculated for the 20 NMR model structures and the NOE_TAR+ $^3\text{J}_{\text{IR}}$ and NOE_TAR+ $^3\text{J}_{\text{LE}}$ trajectories and the corresponding list of the $^3\text{J}(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constants and violations are provided in the Online Resource of Dolenc *et al.*⁶ The $^3\text{J}(\text{H}_\text{N}-\text{H}_{\text{C}\alpha})$ -coupling constants were calculated for the simulated and NMR model structures using the Karplus relation with the parameters $a = 6.4 \text{ Hz}$, $b = -1.4 \text{ Hz}$, and $c = 1.9 \text{ Hz}$.²⁰ The secondary structure assignment was done with the program DSSP, based on the Kabsch-Sander rules.²¹

Occurrence frequencies and distance distributions for hydrogen bonds, polar contacts, and salt bridges were calculated from the 20,000 saved trajectory configurations. Hydrogen bonds were defined by a minimum donor-H-acceptor angle of 135° and a maximum H-acceptor distance of 0.25 nm. Similarly, polar contacts were defined by a maximum distance of 0.3 nm between the protons of hydrogen and nitrogen or oxygen, and salt bridges by a maximum distance of 0.6 nm between the partially charged HE or HH protons and OE oxygens of 25Arg and 20Glu or 22Glu, respectively.

A cluster analysis of peptide conformations was performed on every second frame of the simulated trajectories, that is, at 1-ps intervals. The clustering algorithm, which uses the atom-positional root mean square deviation (RMSD) as similarity criterion, has been described in previous studies of peptide dynamics.^{22,23} For this analysis, we performed a translational superposition of centres of mass and a rotational least-squares fit for every pair of configurations using all side-chain atoms of the peptides and calculated the corresponding atom-positional RMSD for the same set of atoms. The cluster criterion was a side-chain atom-positional RMSD of less than 0.25 nm. The dominant clusters, that is, those describing the largest fractions of the trajectories, were analyzed with respect to temporal evolution, hydrogen bonds, salt bridges, and polar contacts. The joint conformational clustering of the two MD trajectories, NOE_TAR+ $^3\text{J}_{\text{IR}}$ and NOE_TAR+ $^3\text{J}_{\text{LE}}$ and the set of 20 NMR model structures was performed using 10^4 structures of each MD trajectory and 500 copies of each of the 20 NMR model structures, which lead to equal weighting of the three sets of structures in the clustering. For the visual analysis, the visual molecular dynamics (VMD)²⁴ and PyMOL programs were used.

Acknowledgments

Financial support by the National Centre of Competence in Research (NCCR) in Structural Biology is gratefully acknowledged. We would like to thank Jane R. Allison for help with the local-elevation biased ^3J -coupling restraining, and Andrei Alexandrescu and Wolfgang Jahnke for their constructive comments concerning the measurements.

References

1. Burkhard P, Stetefeld J, Strelkov SV (2001) Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol* 11:82–88.
2. Lupas AN, Gruber M (2005) The structure of alpha-helical coiled coils. *Adv Protein Chem* 70:37–78.
3. Parry DAD (2005) Microdissection of the sequence and structure of intermediate filament chains. *Adv Protein Chem* 70:113–142.
4. Steinmetz MO, Jelesarov I, Matousek WM, Honnappa S, Jahnke W, Missimer JH, Frank S, Alexandrescu AT, Kammerer RA (2007) Molecular basis of coiled-coil formation. *Proc Natl Acad Sci USA* 104:7062–7067.
5. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160:65–73.
6. Dolenc J, Missimer JH, Steinmetz MO, van Gunsteren WF (2010) Methods of NMR structure refinement: molecular dynamics simulations improve the agreement with measured NMR data of a C-terminal peptide of GCN4-p1. *J Biomol NMR* 47:221–235.
7. Missimer JH, Steinmetz MO, Baron R, Winkler FK, Kammerer RA, Daura X, van Gunsteren WF (2007) Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Sci* 16:1349–1359.
8. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Zurich: Vdf Hochschulverlag AG an der ETH Zurich.
9. Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennel J, Torda AE, Huber T, Krüger P, van Gunsteren WF (1999) The GROMOS biomolecular simulation program package. *J Phys Chem A* 103:3596–3607.
10. Christen M, Hünenberger PH, Bakowies D, Baron R, Bürgi R, Geerke DP, Heinz TN, Kastenholz MA, Kräutler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF (2005) The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* 26:1719–1751.
11. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 25:1656–1676.
12. Oostenbrink C, Soares TA, van der Vegt NFA, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. *Eur Biophys J Biophys Lett* 34:273–284.
13. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J, Interaction models for water in relation to protein hydration. In: Pullman B, Ed. (1981) *Intermolecular forces*. Dordrecht, The Netherlands: Reidel, pp 331–342.
14. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23:327–341.
15. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690.
16. Glättli A, Daura X, van Gunsteren WF (2002) Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L. *J Chem Phys* 116:9811–9828.
17. Torda AE, Brunne RM, Huber T, Kessler H, Vangunsteren WF (1993) Structure refinement using time-averaged J-coupling constant restraints. *J Biomol NMR* 3:55–66.
18. Nanzer AP, Torda AE, Bisang C, Weber C, Robinson JA, van Gunsteren WF (1997) Dynamical studies of peptide motifs in the *Plasmodium falciparum* circumsporozoite surface protein by restrained and unrestrained MD simulations. *J Mol Biol* 267:1012–1025.
19. Allison JR, van Gunsteren WF (2009) A method to explore protein side chain conformational variability using experimental data. *Chemphyschem* 10:3213–3228.
20. Pardi A, Billeter M, Wüthrich K (1984) Calibration of the angular-dependence of the amide proton- C^α proton coupling constants, $^3J_{\text{HN}\alpha}$, in a globular protein. Use of $^3J_{\text{HN}\alpha}$ for identification of helical secondary structure. *J Mol Biol* 180:741–751.
21. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
22. Daura X, van Gunsteren WF, Mark AE (1999) Folding–unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations. *Proteins: Struct Funct Genet* 34:269–280.
23. Keller B, Daura X, van Gunsteren WF (2010) Comparing geometric and kinetic cluster algorithms for molecular simulation data. *J Chem Phys* 132:074110.
24. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38.