Structure, Vol. 9, 503–511, June, 2001, ©2001 Elsevier Science Ltd. All rights reserved. PII S0969-2126(01)00611-6

Structural Characterization of the LEM Motif Common to Three Human Inner Nuclear Membrane Proteins

Cédric Laguri,[1] Bernard Gilquin,[1] Nicolas Wolff,[1] Régine Romi-Lebrun,[1] Karine Courchay,[1] Isabelle Callebaut,[2] Howard J. Worman,[3] and Sophie Zinn-Justin[1,4]

[1]Département d'Ingénierie et d'Etudes des Protéines
CEA Saclay
91191 Gif-sur-Yvette
France

[2]Laboratoire de Minéralogie-Cristallographie Paris
CNRS UMR 7590
Universités Paris 6/Paris 7
Case 115
4 place Jussieu
75252 Paris Cedex 05
France

[3]Departments of Medicine and Anatomy and Cell Biology
College of Physicians and Surgeons
Columbia University
New York, New York 10032
USA

Summary

Background: Integral membrane proteins of the inner nuclear membrane are involved in chromatin organization and postmitotic reassembly of the nucleus. The discovery that mutations in the gene encoding emerin causes X-linked Emery-Dreifuss muscular dystrophy has enhanced interest in such proteins. A common structural domain of 50 residues, called the LEM domain, has been identified in emerin MAN1, and lamina-associated polypeptide (LAP) 2. In particular, all LAP2 isoforms share an N-terminal segment composed of such a LEM domain that is connected to a highly divergent LEM-like domain by a linker that is probably unstructured.

Results: We have determined the three-dimensional structures of the LEM and LEM-like domains of LAP2 using nuclear magnetic resonance and molecular modeling. Both domains adopt the same fold, mainly composed of two large parallel α helices.

Conclusions: The structural LEM motif is found in human inner nuclear membrane proteins and in protein-protein interaction domains from bacterial multienzyme complexes. This suggests that LEM and LEM-like domains are protein-protein interaction domains. A region conserved in all LEM domains, at the surface of helix 2, could mediate interaction between LEM domains and a common protein partner.

[4]Correspondence: szinn@cea.fr

Introduction

In eukaryotic cells, the nuclear envelope consists of three major components, the nuclear lamina, the inner and outer nuclear membranes, and the nuclear pore complexes. The outer nuclear membrane faces the cytoplasm and is continuous with the peripheral rough and smooth endoplasmic reticulum, whereas the inner nuclear membrane is tightly associated with lamina. Several integral inner nuclear membrane proteins that bind to the lamina have been characterized. In particular, lamina-associated polypeptide (LAP) 1, isoforms of LAP2, lamin B receptor (LBR), and emerin attach the lamina to the inner nuclear membrane [1–6]. These proteins are part of the nuclear envelope architecture and play important roles in postmitotic nuclear reassembly [7–9]. LAP2 isoforms and LBR also bind to chromatin, either to DNA [5–6, 10–11] or other chromatin-associated proteins [3, 12–14]. The chromatin-inner nuclear membrane protein interactions may localize heterochromatin to the nuclear periphery in interphase and may underlie the reassembly of nuclear envelopes around decondensing chromosomes at the end of mitosis. In particular, LAPs are phosphorylated during mitosis, and phosphorylation of LAP2 by mytotic cytosol inhibits its binding to both lamin B1 and chromosomes [4]. The dynamics of LAP2 during mitosis are probably closely linked to the processes of nuclear envelope disassembly and reformation. Finally, mutations in emerin have been described in X-linked Emery-Dreifuss muscular dystrophy, providing the first connection between the inner nuclear membrane and a human disease [15]. Despite their importance to cell biology and medicine, however, no structural data at atomic resolution currently exist for integral membrane proteins of the inner nuclear membrane.

Analysis of the sequences of integral proteins of the inner nuclear membrane has identified a common structural domain of about 50 amino acids in three human and two *Caenorhabditis elegans* proteins [16]. As this domain is present in LAP2, emerin, and MAN1, it has been termed LEM (Figure 1). All LAP2 isoforms share an N-terminal segment composed of such a LEM domain (residues 108–151) that is connected to a highly divergent LEM-like domain by a linker that is probably unstructured (residues 1–47). The LEM domain of LAP2β is mostly included in the region between amino acid residues 67 and 137, whose deletion abolishes the interaction of LAP2β with the nonspecific DNA binding protein BAF [14]. Furthermore, the N-terminal LEM-like domain of LAP2β is found in the region between amino acid residues 1 and 85, which is involved in chromosome binding [6].

Here, we report the three-dimensional solution structures of the LEM and LEM-like domains of LAP2β, as determined by proton Nuclear Magnetic Resonance

Key words: emerin; inner nuclear membrane proteins; lamin; LAP2; LEM domain; NMR

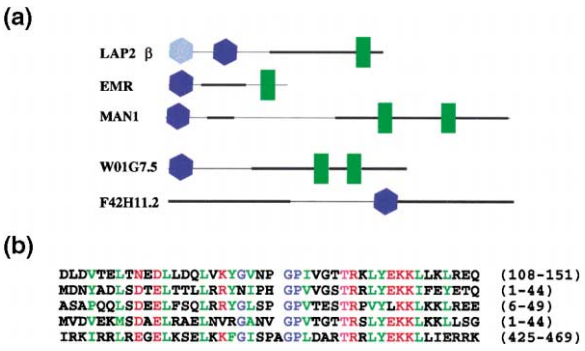

Figure 1. Sequence Analysis of Proteins Containing a LEM Module

(a) A schematic organization of proteins containing a LEM module [16]. Three human inner nuclear membrane proteins, LAP2 isoform β (LAP2 β), emerin (EMR), and MAN1 are displayed, as are two *C. elegans* proteins, the membrane-anchored W01G7.5 and the soluble F42H11.2. Dark blue hexagons represent LEM domains, light blue hexagons represent LEM-like domains, green rectangles represent transmembrane segments, and thin black lines represent sequence regions of low complexity.

(b) Multiple alignment of the LEM domains of human LAP2 (residues 108–151), emerin (residues 1–44), and MAN1 (residues 6–49), as well as *C. elegans* W01G7.5 (residues 1–44) and F42H11.2 (residues 425–469) [16]. Conserved hydrophobic residues are displayed in green, conserved charged residues are displayed in red, conserved polar and neutral residues are displayed in magenta, and conserved glycines and prolines are displayed in blue.

(NMR) and molecular modeling. We show that both domains, even though they are highly divergent (they share only 18% identity), have similar three-dimensional structures, composed of a three-residue N-terminal α helix and two large parallel α helices interacting through a set of conserved hydrophobic amino acids. The two structures are analyzed in terms of topohydrophobic positions, electrostatic surface, and conserved patches of amino acids. Searches for protein domains displaying the LEM structural motif, which identified domains of bacterial multienzyme complexes involved in protein-protein interactions, are presented.

Results

Oligomerization States of LEM-Like and LEM Domains

Analytic ultracentrifugation was used in order to characterize the oligomerization state of the two N-terminal domains of LAP2. At the equilibrium, masses of 6430 ± 580 and 6230 ± 760 Da were measured for monomer theoretical masses of 6272 and 6533 Da, respectively. Thus, the LEM-like and LEM domains of LAP2 do not self-associate at pH 6.3 and at a concentration lower than 50 μM.

Secondary Chemical Shift Analysis

Sequence analyses of the LEM-like and LEM sequences by AGADIR [17] or PHD [18] predict the presence of two large α helices in both peptides. Consistently, analysis of the Hα secondary chemical shifts of LEM-like and

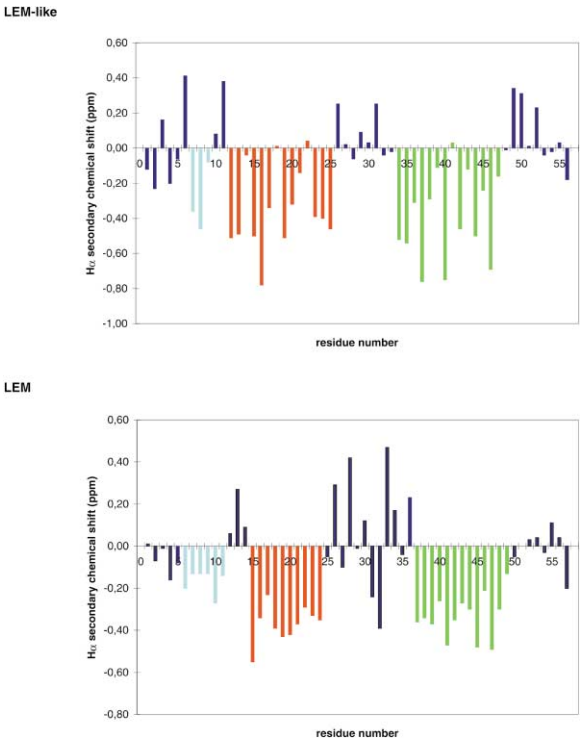

Figure 2. Secondary Hα Chemical Shifts of LAP2 LEM-Like and LEM Domains as a Function of Residue Number

Consecutive residues showing a secondary chemical shift lower than −0.1 ppm are predicted to form a helix [37]. The bars corresponding to the short N-terminal helix are colored in cyan. The bars corresponding to the two main helices are colored in red and green, successively.

LEM suggests the presence of a short N-terminal helix and two large helices (Figure 2).

Proton-Proton Distance Restraints

On the NOESY spectra, 2136 (1202 in H_2O and 934 in D_2O) and 2574 (1629 in H_2O and 945 in D_2O) peaks were analyzed for LEM-like and LEM, respectively. In the case of LEM-like, at the end of the assignment and structure calculation procedure, 42 peaks remained unassigned. Within these peaks, 26 correspond to unknown chemical shifts, probably characteristic of minor conformations. The chemical shifts of the remaining 16 peaks could be assigned, but the corresponding proton-proton distance was higher than 7 Å on the three-dimensional structures (i.e., no reasonable assignment that is consistent with the set of already assigned peaks, and thus the three-dimensional structure, was found). On the basis of the 2092 other peaks, 964 restraints were generated. Using an error of ±25%, 60 restraints led to systematic violations higher than 0.5 Å and thus were not used. The structures were calculated on the basis of the 904 remaining distance restraints, which comprised 102 ambiguous restraints at the end of the calculations. The mean number of unambiguous distance restraints per residue yields 14.3 but is very different from one residue to another. It is particularly high for the leucine residues buried in the core of the protein (79 for Leu19), but it is

Table 1. Experimental Restraints and Structural Statistics for LEM-like and LEM Domains

	LEM-like[a]	LEM
Number of Experimental Distance Restraints		
Unambiguous	802	857
Ambiguous	102	174
Rmsd from distance restraints	0.090 ± 0.002	0.062 ± 0.0014
Number of violations higher than 0.5 Å	1.2 ± 0.9	0.8 ± 0.3
Rmsd from Ideal Values		
Bond (Å)	0.0064 ± 0.0001	0.0045 ± 0.0001
Angle (°)	0.980 ± 0.017	0.720 ± 0.020
Energy (kcal/mol)		
Bond	38.0 ± 1.8	19.0 ± 1.0
Angle	245.2 ± 8.8	134.7 ± 8.7
vdW[b]	71.7 ± 5.9	43.2 ± 6.0
nOe[c]	382.7 ± 17.7	207.3 ± 9.6
Ramachandran Analysis[d]		
Residues in favored regions	52.8%	57.3%
Residues in additional allowed regions	37.7%	34.9%
Residues in generously allowed regions	8.7%	5.3%
Residues in disallowed regions	0.9%	2.4%
Coordinate Precision (Å)		
For residues	3–48	6–48
On backbone atoms	0.39 ± 0.05	0.40 ± 0.07

[a] All values are averaged on the ten X-PLOR structures.
[b] The van der Waals energy is calculated with a repel function and the parallhdg parameters.
[c] The values of the square-well nOe are calculated with force constants of 50 kcal/mol Å^2.
[d] Calculated with Procheck-nmr [36].

very low at the C terminus, which adopts a random coil structure (<6 for Ala54, Gly55, and Thr56).

In the case of LEM, at the end of the calculation procedure, 67 peaks remained unassigned. Seven peaks correspond to unknown chemical shifts, probably characteristic of minor conformations. The chemical shifts of the remaining 60 peaks could be assigned, but the corresponding proton-proton distance was higher than 7 Å on the three-dimensional structures (i.e., no reasonable assignment that is consistent with the set of already assigned peaks, and thus the three-dimensional structure, was found). On the basis of the 2507 other peaks, 1069 restraints were generated. Restraints leading to systematic violations higher than 0.5 Å were not used. The structures were calculated on the basis of the 1031 remaining distance restraints, which comprised 174 ambiguous restraints at the end of the calculations. The mean number of unambiguous distance restraints per residue yields 15 but is very different from one residue to another. It is particularly high for the leucine residues buried in the core of the protein (68 for Leu44), but it is very low at the N and C termini, which adopt a random coil structure (<7 for Ser54, Ser56, and Ser57).

Structural Statistics

Analysis of the ten final structures (Table 1) of LEM-like and LEM shows that no distance violation larger than 0.55 Å is present. Furthermore, the covalent geometry is respected, as evidenced by the low rmsd value for bond lengths and valence angles. The values of the van der Waals energy is small, indicating that there is no incorrect nonbonded contacts. The Ramachandran plot confirms the good quality of the structures, as no residues are systematically in the disallowed region. Over the ten structures, the percentages of residues in the most favored and additional allowed region are 99.2% and 97.5% for LEM-like and LEM, respectively.

Backbone Structure

A backbone superposition of the ten lowest energy structures of LEM-like and LEM is shown in Figure 3. The conformation of the backbone (C, N, and Cα atoms) is well defined in both proteins, except in the N-terminal and C-terminal regions. In the well-defined segments (residues 3–48 for LEM-like and 6–48 for LEM), the rmsd with respect to the mean coordinate is close to 0.4 ± 0.1 Å. In both structures, a three-residue N-terminal α helix and two large α helices, named helix 1 and helix 2, are observed (Figure 4). Inspection of the Ramachandran maps shows that the three-helical segments in LEM-like are Pro7–Leu10, Lys12–Asn22, and Val35–Leu44, and in LEM, the three-helical segments are Asp9–Glu12, Asn15–Tyr25, and Gly34–Leu47. These helices are characterized by numerous hydrogen bonds. The hydrogen bonds present in more than 80% of the helical structures are the following: in LEM-like, helices 1 and 2 present 6 and 7 i → i + 4 hydrogen bonds, respectively; in LEM, helices 1 and 2 present 5 and 8 i → i + 4 hydrogen bonds, respectively.

Side Chains

The hydrophobic core of LEM-like and LEM domains is formed by a large number of leucine residues, assisted

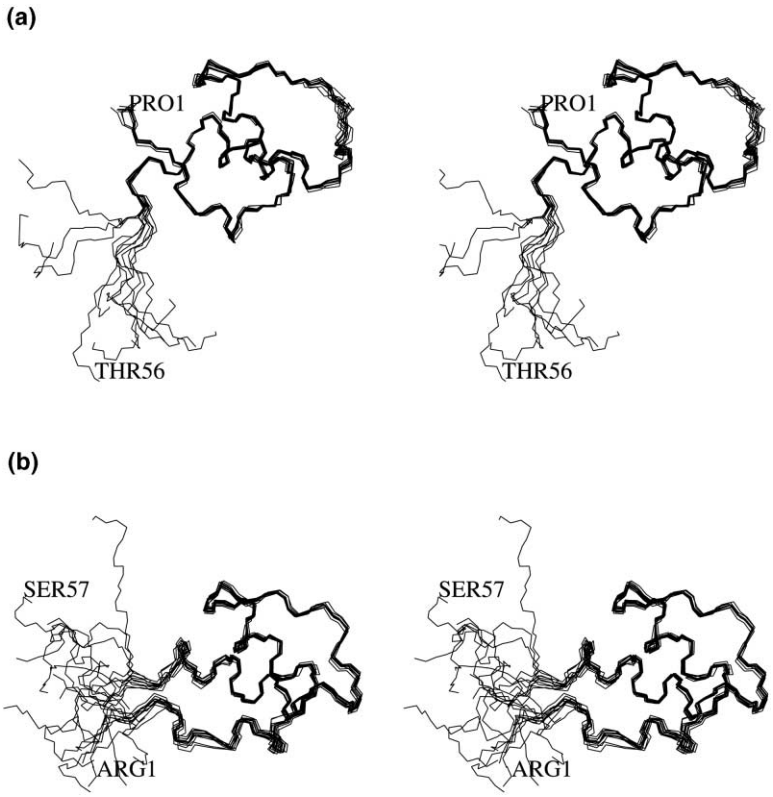

Figure 3. Three-Dimensional Stuctures of LAP2 LEM-Like and LEM Domains

(a) Stereo view of the ten final LEM-like domain backbone structures.
(b) Stereo view of the ten final LEM domain backbone structures.

by additional valine, isoleucine, phenylalanine, and tyrosine residues. In LEM-like, seven leucines are buried and interact with Phe3, Val24, Val35, Tyr36, Val37, and Tyr40 to stabilize the domain conformation. In LEM, six leucines are buried and, together with Val10, Tyr25, Val27, Ile32, and Tyr40, stabilize the domain conformation.

A few polar side chains are also buried in the LEM-like and LEM structures. In LEM-like, Glu5, Lys16, Asn22, and Gln31 are less than 20% solvent accessible. Lys16 is, in particular, salt-bridged to Asp13 in helix 1, and Gln31 is hydrogen-bonded to the backbone of Asp34 at the N terminus of helix 2. In LEM, Asp6, Glu21, Thr36, Gln41, and Arg48 are less than 20% solvent accessible. Glu21 is salt-bridged to Arg48, thus stabilizing the relative positioning of the two main helices. Similarly, interactions of Thr36 with Leu39 and Gln41 with Lys38 stabilize helix 2.

Discussion

LEM-Like and LEM Domains Share the Same Fold

Figure 4 shows that the LEM-like and LEM domains of LAP2β have very similar three-dimensional structures. Superposition of the LEM-like and LEM structures is optimized using the alignment displayed in Figure 5a.

Comparison of the Ramachandran plots for both proteins shows that they share, beginning from residue 7

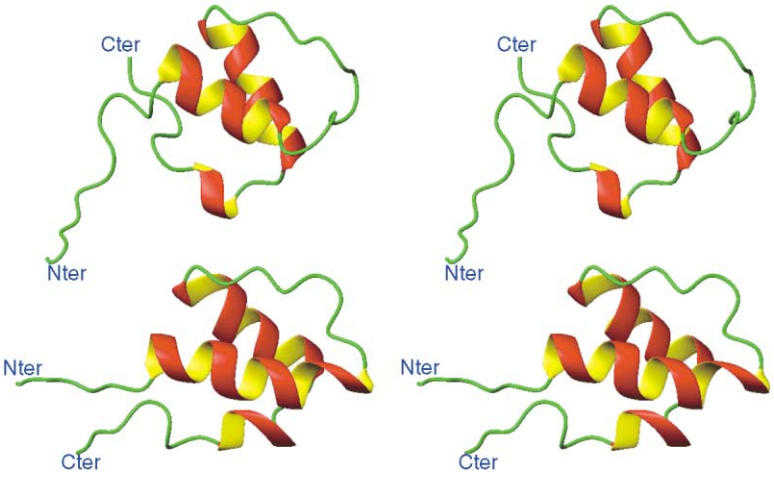

Figure 4. Stereo View of the Averaged Structures of LAP2 LEM-Like and LEM Domains

(a) The three helical segments correspond to Pro7-Leu10, Lys12-Asn22, and Val35-Leu44 in LEM-like domains.
(b) The three helical segments correspond to Asp9-Glu12, Asn15-Tyr25, and Gly34-Leu47 in LEM domains.
The ribbons are calculated using MOLMOL [38].

(a)

```
1       PEFLEDPSVLTKDKLKSELVANNVTLPAGEQRKDVYVQLYLQHLTARNRPPLPAGT 56
103 RQEDKDDLDVTELTNEDLLDQLVKYGVNPGP  IVGTT RKLYEKKLLKLREQGTESRSS 159
```

(b)

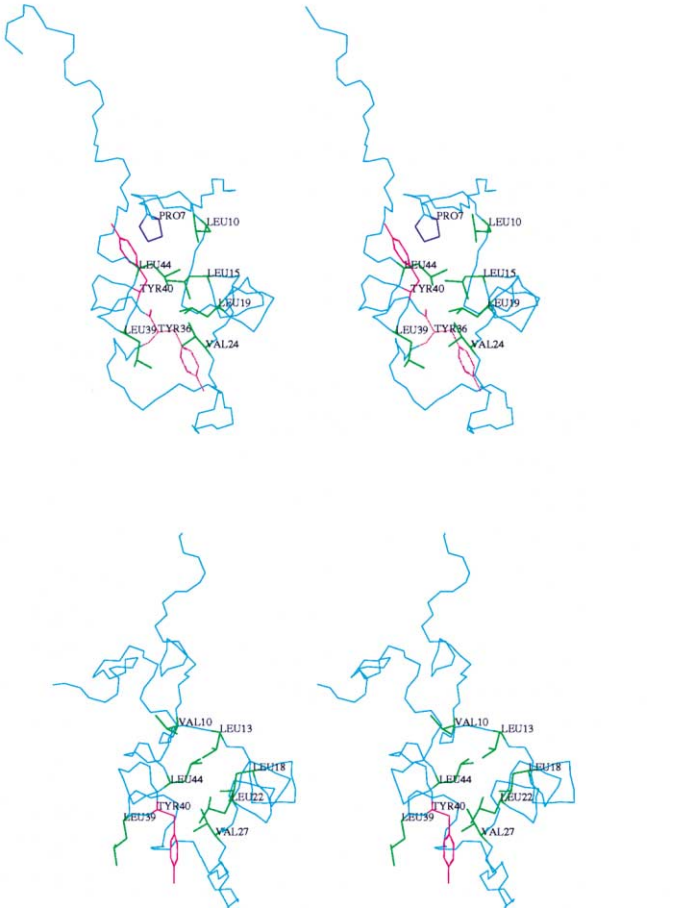

Figure 5. Conservation of Hydrophobic Residues within LEM-Like and LEM Domains

(a) Alignment of LEM-like (1–56) and LEM (103–159) sequences of LAP2 on the basis of their three-dimensional structures. Strict topohydrophobes (i.e., the conserved hydrophobic residues within the LEM family) are displayed in magenta, topohydrophobes conserved in four of the five LEM sequences are displayed in red, and hydrophobic residues that are not topohydrophobes are colored in cyan.

(b) Positioning of the topohydrophobic residues common to LEM-like and LEM domains. Averaged three-dimensional backbone structures of LEM-like and LEM domains are presented on the upper and lower views, respectively. Side chains of topohydrophobic residues common to LEM-like and LEM domains are displayed in green, except the side chain of Tyr40, which is colored in pink. The side chain of Tyr36 of LEM-like domains, which plays a structural role similar to that of Tyr40 in LEM domains, is also displayed and colored in violet.

in LEM-like and 10 in LEM, an N-terminal three-residue α helix; two residues that are mainly in β conformation; a large α helix of 11 residues; a residue that is in L conformation; two more residues in β conformation; a large loop that is specific to each protein and contains three insertions in favor of LEM-like, from residue 26 in LEM-like and 29 in LEM; a second large α helix of ten residues for LEM-like (35–44) and 15 residues for LEM (34–48); and, finally, a disordered tail. The rmsd between the backbone atoms of residues 7–25 and 35–44 in LEM-like and 10–28 and 35–44 in LEM is equal to 2.3 Å. Thus, the two three-dimensional structures are particularly close in these regions.

The positioning of the two main helices is conserved in LEM-like and LEM. The N-terminal three-residue helix is located at a similar position, but the structure and the positioning of the segment before this helix vary from one protein to the other. This is due, in particular, to the different positioning in the sequence of the leucine residue preceding the N-terminal helix, which interacts with hydrophobic residues of this helix. The loop between the two main helices contains an AGEQ and PGPI turn in LEM-like and LEM, respectively. However, the turn is positioned further from the protein core in LEM-like, because of the three-residue insertion. Finally, the second main helix is longer in LEM, probably due to the presence of additional hydrophobic residues at positions 45 and 47, which stabilize the α-helical conformation.

Role of the Topohydrophobic Residues in the LEM Fold

The LEM domains are characterized by a set of conserved residues [16] (Figure 1). A search for topohydrophobic residues (i.e., the conserved hydrophobic residues [19]) on the basis of the 5 known sequences of LEM domains shows that 11 residues are hydrophobic in at least 4 of the LEM sequences. In the LEM of LAP2β, they correspond to: Val10 (Pro in Man1), Leu 13, Leu18, Leu 22, Tyr25 (Arg in W01G7.5), Val27 (Ala in W01G7.5), Ile32, Leu39, Tyr40, Leu44, and Leu47 (Glu in F42H11.2).

Eight of these 11 residues are also hydrophobic in

LEM-like (Figure 5a). Only Tyr25 at the C terminus of helix 1, Ile 32 in the loop between the two main helices, and Leu47 at the C terminus of helix 2 have no hydrophobic counterparts in LEM-like. The eight topohydrophobic residues shared by LEM and LEM-like are displayed in Figure 5b. They play critical roles in the fold of the domains; the interaction between Pro7 (Val10 in LEM) and Leu10 (Leu13) stabilizes the N-terminal helix; Leu15 (Leu 18), Leu19 (Leu22), and Val24 (Val27) form the hydrophobic side of helix 1, which interacts with helix 2; Leu39 (Leu39 and Tyr 40) and Leu44 (Leu44) form the hydrophobic surface of helix 2, which interacts with helix 1; in LEM-like, Tyr 40, on the opposite side of helix 2, interacts with the topohydrophobic residues of the N-terminal helix. It is interesting to note that Tyr40, which is a topohydrophobic residue, plays a different role in the two structures.

Conservation of the Buried Residues in the LEM Domains

Consistent with their structural role, the topohydrophobic residues are particularly buried. Within the 12 positions corresponding to residues buried (i.e., less than 20% solvent accessible) in both LEM-like and LEM domains, 9 are occupied by topohydrophobic residues shared by LEM domains. Only seven of these are hydrophobic in LEM-like (the exceptions are Asn22 at the C terminus of helix 1 and Gln31 at the N terminus of helix 2). The three nontopohydrophobic positions corresponding to buried residues in LEM-like and LEM are Phe3/Asp6, Leu26/Pro29, and Val35/Thr36. Interestingly, Pro29 is a proline in three of the five LEM sequences, and Thr36 is strictly conserved in all the LEM domains. Finally, a few positions are buried in only one domain. For example, Tyr36 in LEM-like, which has no equivalent in LEM, is buried and may play the role of the topohydrophobic Tyr40 in LEM.

Electrostatic Properties of the LEM Domains

A comparison of the five LEM domain sequences also shows that a set of nonhydrophobic residues is conserved [16]. Interestingly, most of these residues are not conserved in LEM-like, suggesting, in particular, different electrostatic properties for LEM and LEM-like domains.

In the LEM domain of LAP2β, the negatively charged clusters are located at the N and C termini interface (Asp4, Asp6, Asp7, and Glu 49) and on the solvent-exposed surface of helix 1 (Glu16, Asp17, and Asp20), whereas the positively charged clusters are found around Tyr25 (Lys5, Lys24, and Arg48) and on the solvent-exposed surface of helix 2 (Lys42, Lys43, and Lys46). Only one of these clusters, comprising Lys42, Lys43, and Lys46, is mostly conserved within the LEM family. This cluster is included in a continuous surface that is highly conserved within the LEM family, which is formed by residues Thr36, Arg37, Leu39, Tyr40, Glu41, Lys42, Lys43, and Leu44 and is located on the solvent-exposed surface of helix 2.

In the LEM-like domain, a negatively charged cluster is found at the N terminus (Glu2, Glu5, and Asp6), and a small positively charged cluster is found at the C terminus (Arg47 and Arg49). However, most of the charged side chains are involved in salt bridges: in helix 1, Asp13 interacts with Lys16, and Lys14 interacts with Glu18; in the loop between the two main helices, Asp34 is closed to Lys12 and Lys33.

Search for Domains Displaying Two Parallel α Helices

LEM and LEM-like of LAP2β are small structural domains (43 and 46 residues are structured in LEM and LEM-like, respectively) stabilized only by intramolecular electrostatic interactions and hydrophobic contacts. There are few structural domains of less than 50 residues containing no disulfide bond or metal ion binding motif in the Protein Data Bank [20]. To our knowledge, only two are mainly α-helical, the head piece domain of chicken villin [21] and the homologous E3/E1p and E3 binding domains of the dehydrogenase multienzyme complexes of bacteria [22–24]. Interestingly, the domains of the multienzyme complexes show two large parallel α helices. Structural alignment of the two main helices of these domains, referenced as 1BBL (domain from *E. coli* [22]) and 2PDE (domain from *Bacillus stearothermophilus* [23]) in the Protein Data Bank, with those of the LEM domain of LAP2β (Figure 6a) yields an rmsd of 1.9 and 2.3 Å, respectively. Thus, the positioning of the two helices are similar in the three domains (Figure 6b). A search for the eight topohydrophobic residues characteristic of LEM-like and LEM domains in 1BBL and 2PDE sequences (Figure 6a) shows that seven of these are conserved, if we enlarge the group of strictly hydrophobic residues to alanine and threonine. The nonconserved topohydrophobic position corresponds to Val10, which has no equivalent in 1BBL and 2PDE.

Figure 6b also shows that large differences exist between 1BBL, 2PDE, and LEM domains. First, the well-structured N-terminal segment of LEM domains, which contain a three-residue α helix, is not present or unstructured in the 1BBL and 2PDE domains. Second, the large loop found between the two main helices is orientated completely differently in the two domain families. It has been described as containing a β sheet in the BBL domain. No β structure is observed in LEM domains. It also contains a 3_{10} helix in both 1BBL and 2PDE. No 3_{10} helix is found in LEM domains. Finally, a turn is found in this loop for 1BBL and 2PDE, at a position corresponding to the deletions in the LEM sequences.

The identity between 1BBL and the LEM domain of LAP2β is only of 5 residues within 42 (12%), and the similarity is of 20 residues (48%). In the case of 2PDE, the identity is again of 5 residues within 42 (12%), and the similarity is of 22 residues (52%). Thus, an analogy between 1BBL, 2PDE, and LEM domains is very difficult to predict on the basis of sequence analysis only. However, a clear analogy exists from a structural point of view.

Functional Role of LEM Domains

1BBL and 2PDE domains are involved in protein-protein interactions in bacterial multienzyme complexes. They interact by their first large α helix, and, in particular, through polar and positively charged residues, with

(a)

```
1BBL    YASLEEQNNDALSPAIRRLLAEHNLDASAIKGTGVGGRLTREDVEKHLAKA
2PDE              VIAMPSVRKYAREKGVDIRLVQGTGKNGRVLKEDIDAFLAGGA
LEM   RQEDKDDLDVTELTNEDLLDQLVKYGVNPGPIVGTTR  KLYEKKLLKLREQGTESRSS
                  a      hhhhhhhh                 hhhhhhh  b
```

(b)

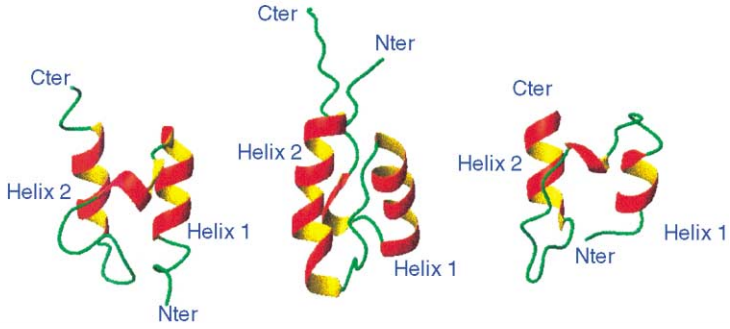

Figure 6. Structure Comparison of the LAP2 LEM Domain with Bacterian Multienzyme Complex Domains

(a) Alignment of the two domains of the multienzyme complexes, referenced as 1BBL and 2PDE in the Protein Data Bank, onto the sequence of the LEM domain of LAP2β. The consensus secondary structure is noted under the three sequences, using "h" as the symbol of α-helical conformation. On the same line, "a" and "b" indicate the beginning and the end, respectively, of the sequences used for the calculations of identity and similarity percentages.

(b) Ribbon representations of the E3/E1p (left view) and E3 binding (right view) domains of the dehydrogenase multienzyme complexes of bacteria [22, 23], as compared to the LEM domain of LAP2.

other enzymes of the complexes [24]. Their structural analogy with LEM domains is consistent with the role of the protein-protein interaction domain described by now for LEM domains [9, 14].

In particular, deletion of the region 67–137, a segment overlapping the LEM domain of LAP2β, abolishes the interaction of LAP2β with the nonspecific DNA binding protein BAF [14]. In the presence of DNA, BAF forms a large oligomeric nucleo-protein complex [25]. Thus, it has already been suggested that BAF may be associated with other proteins in the cell and may be part of a multiprotein-DNA complex. Several oligomerization modules involved in DNA compaction have been structurally elucidated and are described as helical protein domains, using their helices as dimerization elements to form homo- and heterodimers [26]. This type of homo- and heterodimerization is proposed for BAF [25] and can be extended to α-helical BAF binding proteins such as the LEM domain of LAP2β. The LEM-like and LEM domains of LAP2 do not self-associate. However, it should be tested whether LAP2 LEM domain can interact with other domains adopting the LEM fold, as for example, LAP2 LEM-like domain.

Interestingly, LEM-like and LEM domains of LAP2β share only ten common residues, and most of them are buried hydrophobic residues. The biochemical nature of the solvent-accessible residues of LEM-like and LEM is completely different, suggesting that, whereas LEM-like and LEM domains belong to the same protein and share the same LEM fold, they target different protein surfaces. Such a phenomenon has already been reported for HP1 proteins, which are involved in gene silencing via the formation of heterochromatic structures. These proteins are composed of two structurally related domains, an N-terminal chromo domain and a C-terminal shadow chromo domain, connected by a flexible linker. Mouse HP1β was shown to be dimeric, the interaction being mediated by the shadow chromo domain, with the chromo domains moving independently of each other at the end of flexible links [27].

Biological Implications

Integral membrane proteins of the inner nuclear membrane are involved in chromatin organization and postmitotic reassembly. Mutations in one of these proteins, called emerin, have been described in X-linked Emery-Dreifuss muscular dystrophy [15]. Analysis of the sequences of these proteins has identified a common structural domain of about 50 amino acids in the human LAP2, emerin, and MAN1 and in two *C. elegans* proteins [16]. In particular, all LAP2 isoforms share an N-terminal segment composed of such a LEM domain that is connected to a highly divergent LEM-like domain by a linker that is probably unstructured.

We report the first three-dimensional structures of domains belonging to a protein anchored in the inner nuclear membrane, the LEM-like and LEM domains of LAP2. Our results show that the LEM and LEM-like domains of LAP2 share a common fold, which is characterized by a set of eight topohydrophobic residues critical for the stabilization of the three-residue N-terminal helix and the two large helices, as well as for their relative positioning. Such a parallel positioning of two α helices is also found in the E3/E1p and E3 binding domains of the bacterial dehydrogenase multienzyme complexes, which are structural domains of about 40 amino acids involved in protein-protein interactions.

The five LEM domains described by Lin et al. [16], all of which probably adopt a common fold, exhibit a conserved surface of eight residues on the solvent-accessible side of helix 2. Such a mainly positively charged surface is proposed to be involved in the LEM biological function, that is, in the interaction of the LEM

domain with a protein partner. The biochemical nature of the solvent-accessible residues of LEM and LEM-like domains is not conserved. The LEM and LEM-like domains are proposed to interact with different regions of a common biological target or to have different biological targets.

Experimental Procedures

Sample Preparation

The LEM-like and LEM domains were predicted to include residues 1–47 and 108–151 of LAP2β, respectively [16]. Both domains were extended in order to obtain well-structured proteins that are soluble at a millimolar concentration. The final sequences chosen for the LEM-like and LEM structural studies correspond to residues 1–56 and 103–159, respectively, of LAP2β.

The chemical synthesis of both proteins was carried out in solid phase using the Fmoc strategy on an Applied Biosystems 431A. The proteins were purified by HPLC on a semipreparative Vydac C18 column from Merck, and their purity was checked on the corresponding analytic column in the same solvent conditions. Their molecular weights were measured by electrospray mass spectrometry and were found to be consistent with the expected sequences.

For the analytic ultracentrifugation experiments, protein concentration was 10–50 μM, and the solvent was a buffer of 1mM sodium phosphate (pH 6.3). For the NMR experiments, protein concentration was 1–1.5 mM. The solvent was a buffer of 20 mM sodium phosphate (pH 6.3). 3-(trimethylsilyl)[2,2,3,3-$^{2}H_4$]propionate was added as a chemical shift reference. Two samples of each protein were prepared; the first sample was diluted in 90% H_2O, 10% D_2O, and the second sample was diluted in 100% D_2O.

Analytic Ultracentrifugation Experiments

Sedimentation equilibrium was performed at 298 K on a Beckman Optima XLA ultracentrifuge using an AN 60 Ti rotor and cells with a 12 mm optical path length. Sample volumes of 100 μl were centrifuged at 30,000 and 40,000 rpm. Radial scans of absorbance at 274 nm were taken at 3 hr intervals, and equilibrium was achieved after 60 hr. Data were analyzed using the XL-A/XL-2 software supplied by Beckman.

NMR Experiments

All experiments were carried out at 298 K on a Bruker 500 MHz or 600 MHz spectrometer. Two-dimensional DQF-COSY [28], TOCSY [29], NOESY [30], and off-resonance ROESY [31] experiments were recorded on both proteins. A DIPSI2 composite pulse was used for isotropic mixing during 80 ms in the TOCSY experiments. A set of NOESY experiments was carried out with mixing times of 60 ms, 100 ms, 140 ms, and 180 ms, using a preparation period of 2 s. To filter for pure exchange effects, off-resonance ROESY spectra were recorded by applying, during the mixing time, a radiofrequency irradiation characterized by an axis making an angle of 35° with the static magnetic field, an rf field power of 9800 Hz, and a duration of 100 ms. The water signal was suppressed by a WATERGATE sequence [32]. All the experiments were performed in hypercomplex mode. The spectra were recorded with 512 $t_1 \times 1024$ t_2 points (1024 $t_1 \times 4096$ t_2 for the DQF-COSY). Data processing was carried out using XWIN-NMR (Bruker) and FELIX (Biosym Technologies).

One-dimensional spectra were recorded before and after each set of experiments to ensure that the proteins were not structurally modified with time.

Proton Resonance Frequency Assignment

In order to assign each of the protein proton frequencies, spin systems were identified on the TOCSY spectra. Sequential assignment was essentially carried out on the basis of HN_i-$H\alpha_{i+1}$ and HN_i-HN_{i+1} nOe interactions.

In the LEM-like domain, all protons were assigned, except the NH of Glu2 on the backbone and the amine protons of Gln31 on the side chains. An additional minor conformation was detected for Lys12 and for the side chains of Phe3, Gln31, Tyr40, and His43.

In the LEM domain, all protons were assigned, except the NH of Gln2 and the NH, Hα1, and Hα2 of Gly51 on the backbone and the amine protons of Gln2, Asn15, and Gln50 on the side chains.

Hα chemical shifts were compared to those found for each amino acid in GGXA peptides [33].

Experimental Restraints

Proton-proton distance restraints were deduced from the analysis of the NOESY spectra recorded at different mixing times. The volumes of the nOe cross-peaks were integrated. For each peak, a build-up curve was constructed by fitting the experimental volumes to the following function of the mixing time: $f(\tau_m) = a\,\tau_m + b\,\tau_m^2$. The coefficient was taken as a build-up rate of the corresponding nOe. Calibration of these dipolar correlation rates was achieved on the basis of the known range of dαN distances. The errors made on the distances were evaluated to 25%. When comparison of the distances deduced from peaks found on both sides of the diagonal and in both solvents (H_2O and D_2O) showed an error larger than 25%, an error equal to twice their rmsd was used [34].

Structure Calculation

A semiautomated iterative assignment procedure was used to assign the nOe and to construct the three-dimensional structures simultaneously. This procedure is described in detail in Savarin et al. [35]. It starts with a linear structure of the protein and a distance restraint list reflecting all the possible assignments of each peak. At this stage, most of the nOe assignments are ambiguous, and the distance range is 1.8–5 Å for all the restraints. The nonambiguous long-range restraints are checked carefully, because at this stage, they are critical for the folding of the molecule. Then, a first set of 100 structures are calculated; the ten energetically most favorable structures are selected, and on the basis of these structures, a new nOe assignment is calculated using a cutoff of 7 Å (an assignment is used if the corresponding distance is less than 7 Å in at least four structures). After two such iterations, the experimental distances are introduced, and about 20 new iterations are calculated. When progressing in the iterations, ambiguous restraints are more and more uniquely assigned by choosing the closest proton pair in the three-dimensional structures. A force field adapted to NMR structure calculation (file topallhdg.pro and parallhdg.pro in X-PLOR 3.1) was used in this procedure. Finally, 200 structures were calculated, and the 10 best structures were selected to be analyzed.

These ten structures of lower energy were deposited at the Protein Data Bank (see Accession Numbers).

Acknowledgments

We thank Philippe Savarin and Gérard Battelier for great technical assistance, Cecilia Östlund for helpful input, and Jean-Claude Courvalin for invaluable discussions. H.J.W. was supported by a grant from the Muscular Dystrophy Association.

Received: December 21, 2000
Revised: May 2, 2001
Accepted: May 7, 2001

References

1. Worman, H.J., Yuan, J., Blobel, G., and Georgatos, S.D. (1988). A lamin B receptor in the nuclear envelope. Proc. Natl. Acad. Sci. USA *85*, 8531–8534.
2. Senior, A., and Gerace, L. (1988). Integral membrane proteins specific to the inner nuclear membrane and associated with the nuclear lamina. J. Cell. Biol. *107*, 2029–2036.
3. Clements, L., Manilal, S., Love, D.R., and Morris, G.E. (2000). Direct interaction between emerin and lamin A. Biochem. Biophys. Res. Commun. *267*, 709–714.
4. Foisner, R., and Gerace, L. (1993). Integral membrane proteins of the nuclear envelope interact with lamins and chromosomes, and binding is modulated by mitotic phosphorylation. Cell *73*, 1267–1279.
5. Ye, Q., and Worman, H.J. (1994). Primary structure analysis and lamin B and DNA binding of human LBR, an integral protein

of the nuclear envelope inner membrane. J. Biol. Chem. *269*, 11306–11311.
6. Furukawa, K., Fritze, C.E., and Gerace, L. (1998). The major nuclear envelope targeting domain of LAP2 coincides with its lamin binding region but is distinct from its chromatin interaction domain. J. Biol. Chem. *273*, 4213–4219.
7. Collas, P., and Courvalin, J.C. (2000). Sorting nuclear membrane proteins at mitosis. Trends Cell Biol. *10*, 5–8.
8. Dechat, T., Vlcek, S., and Foisner, R. (2000). Review: lamina-associated polypeptide 2 isoforms and related proteins in cell cycle-dependent nuclear structure dynamics. J. Struct. Biol. *129*, 335–345.
9. Worman, H.G., and Courvalin, J.C. (2000). The inner nuclear membrane. J. Membr. Biol. *177*, 1–11.
10. Furukawa, K., Glass, C., and Kondo, T. (1997). Characterization of the chromatin binding activity of lamina-associated polypeptide (LAP)2. Biochem. Biophys. Res. Commun. *238*, 240–246.
11. Duband-Goulet, I., and Courvalin, J.C. (2000). Inner nuclear membrane protein LBR preferentially interacts with DNA secondary structures and nucleosomal linker. Biochemistry *39*, 6483–6488.
12. Ye, Q., and Worman, H.J. (1996). Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins homologous to Drosophila HP1. J. Biol. Chem. *271*, 14653–14656.
13. Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J.C., and Worman, H.J. (1997). Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR. J. Bio. Chem. *272*, 14983–14989.
14. Furukawa, K. (1999). LAP2 binding protein 1 (L2BP1/BAF) is a candidate mediator of LAP2-chromatin interaction. J. Cell Sci. *112*, 2485–2492.
15. Bione, S., et al., and Toniolo, D. (1994). Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. Nat. Genet. *8*, 323–327.
16. Lin, F., et al., and Worman, H.J. (2000). MAN1, an inner nuclear membrane protein that shares the LEM domain with lamina-associated polypeptide 2 and emerin. J. Biol. Chem. *275*, 4840–4847.
17. Munoz, V., and Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. Nat. Struct. Biol. *1*, 399–409.
18. Rost, B., Sander, C., and Schneider, R. (1994). PHD, an automatic mail server for protein secondary structure prediction. Comput. Appl. Biosci. *10*, 53–60.
19. Poupon, A., and Mornon, J.P. (1998). Populations of hydrophobic amino acids within protein globular domains: identification of conserved "topohydrophobic" positions. Proteins *33*, 329–342.
20. Bernstein, F.C., et al., and Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. *112*, 535–542.
21. McKnight, C.J., Matsudaira, P.T., and Kim, P.S. (1997). NMR structure of the 35-residue villin headpiece subdomain. Nat. Struct. Biol. *4*, 180–184.
22. Robien, M.A., et al., and Gronenborn, A.M. (1992). Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*. Biochemistry *31*, 3463–3471.
23. Kalia, Y.N., Brocklehurst, S.M., Hipps, D.S., Appella, E., Sakaguchi, K., and Perham, R.N. (1993). The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. J. Mol. Biol. *230*, 323–341.
24. Mande, S.S., Sarfaty, S., Allen, M.D., Perham, R.N., and Hol, W.G. (1996). Protein-protein interactions in the multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase. Structure *4*, 277–286.
25. Cai, M., et al., and Clore, G.M. (1998). Solution structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration. Nat. Struct. Biol. *5*, 903–909.
26. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251–260.
27. Brasher, S.V., et al., and Laue, E.D. (2000). The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. EMBO J. *19*, 1587–1597.
28. Rance, M., Sorensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., and Wuthrich, K. (1983). Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering . Biochem. Biophys. Res. Commun. *117*, 479–485.
29. Braunschweiler, L., and Ernst, R.R. (1983). Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. J. Magn. Reson. *53*, 521–528.
30. Kumar, A., Ernst, R.R., and Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem. Biophys. Res. Commun. *95*, 1–6.
31. Desvaux, H., Birlirakis, N., Wary, C., and Berthault, P. (1995). Improved versions of off-resonance ROESY. Mol. Phys. *86*, 1059–1073.
32. Piotto, M., Saudek, V., and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J. Biomol. NMR *2*, 661–665.
33. Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids (New York: John Wiley and Sons).
34. Gilquin, B., Lecoq, A., Desné, F., Guenneugues, M., Zinn-Justin, S., and Ménez, A. (1999). Conformational and functional variability supported by the BPTI fold: solution structure of the Ca2+ channel blocker calcicludine. Proteins *34*, 520–532.
35. Savarin, P., Zinn-Justin, S., and Gilquin, B. (2001). Variability in automated assignment of NOESY spectra and three-dimensional structure determination : a test case on three small disulfide-bonded proteins. J. Biomol. NMR *19*, 49–62.
36. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*, 477–486.
37. Wishart, D.S., Sykes, B.D., and Richards, F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry *31*, 1647–1651.
38. Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. *14*, 51–55.

Accession Numbers

The entry codes for LEM-like and LEM structures are 1h9e and 1h9f, respectively.