doi:10.1016/j.jmb.2005.05.002 J. Mol. Biol. (2005) 350, 319–337

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Solution Structure of the Symmetric Coiled Coil Tetramer Formed by the Oligomerization Domain of hnRNP C: Implications for Biological Function

Stefanie R. Whitson, Wallace M. LeStourgeon and Andrzej M. Krezel*

Department of Biological Sciences, 465 21st Ave. South Vanderbilt University, Nashville TN 37232, USA

During active cell division, heterogeneous nuclear ribonucleoprotein (hnRNP) C is one of the most abundant proteins in the nucleus. hnRNP C exists as a stable tetramer that binds about 230 nucleotides of pre-mRNA and functions *in vivo* to package nascent transcripts and nucleate assembly of the 40 S hnRNP complex. Previous studies have shown that monomers lacking or possessing mutant oligomerization domains bind RNA with low affinity, strongly suggesting a cooperative protomer-RNA binding mode. In order to understand the role of the oligomerization domain in defining the biological functions and structure of hnRNP C tetramers, we have determined the high-resolution NMR structure of the oligomerization interface that is formed at the core of the complex, examining specific molecular interactions that drive assembly and contribute to the structural integrity of the tetramer. The determined structure reveals an antiparallel four-helix coiled coil, where classically described knobs-into-holes packing interactions at interhelical contact surfaces are optimized so that side-chains interdigitate to create an even distribution of hydrophobic surfaces along the core. While the stoichiometry of the complex appears to be primarily specified by occlusion of hydrophobic surfaces, particularly the interfacial residue L198, from solvent, helix orientation is primarily determined by electrostatic attractions across helix interfaces. The creation of potential interaction surfaces for other hnRNP C domains along the coiled coil exterior and the assembly of oligomerization interfaces in an antiparallel orientation shape the tertiary fold of full-length monomers and juxtapose RNA-binding elements at distal surfaces of the tetrameric complex in the quaternary assembly. In addition, we discuss the specific challenges encountered in structure determination of this symmetric oligomer by NMR methods, specifically in sorting ambiguous interatomic distance constraints into classes that define different elements of the coiled coil structure.

Keywords: hnRNP C; coiled coil; tetramerization domain; symmetric oligomer; NMR

*Corresponding author

Introduction

Heterogeneous nuclear ribonucleoprotein (hnRNP) C is extremely abundant in nuclei during cell division, functioning to actively package and chaperone nascent pre-mRNA transcripts *via* nucleation of 40 S hnRNP particles.[1] When isolated from native hnRNP complexes, hnRNP C proteins exist as $C1_3C2$ tetramers,[2] where C1 and C2 are alternatively spliced gene products that differ by a 13 amino acid residue insertion at C1 G106.[3] hnRNP C tetramers wrap ~230 nucleotide increments of

Abbreviations used: hnRNP, heterogeneous nuclear ribonucleoprotein; RRM, RNA recognition motif; bZLM, basic region zipper-like motif; CTD, C-terminal domain; CLZ, hnRNP C leucine zipper-like oligomerization domain; Trx, thioredoxin; SDS-PAGE, sodium-dodecyl sulfate polyacrylamide gel electrophoresis; M_r, molecular mass; kDa, kilodalton; MDa, megadalton; NOE, nuclear Overhauser effect; RMSD, root-mean-squared-deviation; Lac, lactose operon; Rop, repressor of primer; SFC, neuronal synaptic fusion complex; PP2C, protein phosphatase 2C.

E-mail address of the corresponding author: andrzej.m.krezel@vanderbilt.edu

elongating transcript, three tetramers assembling into a 19 S triangular complex intermediate containing ~700 nucleotides of RNA[4] that directs stoichiometric binding of three $A2_3B1$ and three $A1_3B2$ tetramers to accomplish 40 S particle formation.[5]

Since the physiologically active form of hnRNP C is a tetramer, oligomerization and organization of subunits is critical to *in vivo* RNA-binding activity. Multiple hnRNP C domains are implicated in modulation of RNA-binding affinity, and the intrinsic ability to package defined lengths of pre-mRNA is directed by the orientation and juxtaposition of these elements relative to one another.[6] Primary RNA-binding determinants have been suggested to be contained in both the RNA recognition motif (RRM, residues 8–87)[7–9] and the basic region zipper-like motif (bZLM, residues 140–179).[10–12] Moreover, the acidic C-terminal domain (CTD, residues 208–290) may regulate RNA binding through phosphorylation on specific serine residues.[13–15] However, high affinity hnRNP C–RNA binding interactions only result through positive cooperativity that is conferred by the oligomerization of hnRNP C to mediate association of multiple RNA-binding elements.[9,10,12] In this respect, a 28 residue sequence (hnRNP C leucine zipper-like oligomerization domain, CLZ, residues 180–207) that immediately follows the basic high-affinity nucleic acid binding region has been demonstrated to drive oligomerization *in vivo*[12] and spontaneously form an antiparallel coiled coil *in vitro*.[6]

In order to understand the role of the oligomerization domain in defining the functions and conformational integrity of hnRNP C tetramers, we have determined the molecular contact interfaces formed in CLZ tetramers that drive assembly of the antiparallel coiled coil. Formation of α-helical coiled coils is a common oligomerization mechanism.[16–18] These structures are stabilized by continuous interhelical contacts formed between hydrophobic faces of amphipathic helices. For each seven residues (heptad) along the helix (traditionally denoted *a*, *b*, *c*, *d*, *e*, *f* and *g*[19]) the contact geometry is repeated for the ideal coiled coil. Residues occupying *a* and *d* heptad positions are hydrophobic and form interhelical contacts in the coiled coil core; additional contact surfaces are provided by residues at interfacial *e* and *g* heptad positions.

Assembly of CLZ tetramers is governed by the properties of the residues that compose their polypeptide sequence; as revealed by the variability across coiled coil structures, a pattern of hydrophobic residues at *a* and *d* heptad positions is implicitly insufficient to determine stability or direct a specific conformation. Investigations to define a particular sequence that triggers coiled coil formation reveal that any sequence that provides stability beyond a threshold allows assembly of subunits.[20] Factors that promote α-helical secondary structure, such as helical propensities of side-chains[21] and intrahelical electrostatic interactions,[22–24] as well as polypeptide chain length,[25–28] have been shown to influence the structural integrity of helical bundles. Van der Waals, hydrophobic and electrostatic interactions between residues at interhelical contact interfaces primarily determine the stability, helix orientation and stoichiometry of coiled coil structures.[17,18,29] While the efficiency and favorability of side-chain packing at core heptad positions has been shown to influence both helix orientation[29–31] and stoichiometry[32–34] of coiled coils, these effects are likely more minor determinants of conformational specificity in CLZ oligomers. However, burial of hydrophobic surfaces at core and interfacial heptad positions enhances stability[34,35] and influences the oligomerization state[36–38] of helical bundles, in this respect playing a significant role in directing assembly of CLZ helices into a tetramer. Moreover, interhelical electrostatic attractions resulting from the net charge distribution along helices and compensation for localized charge on side-chains at interfacial positions[38–40] appear to primarily determine the orientation of helices in the CLZ four-helix coiled coil.

Herein, we report the high-resolution nuclear magnetic resonance (NMR) solution structure of the CLZ tetramer, examining specific molecular interactions that define the antiparallel coiled coil and contribute to stability of hnRNP C tetramers. Oligomerization interactions of this domain significantly and directly influence *in vivo* and *in vitro* hnRNP C function through orientation and juxtaposition of RNA-binding and interdomain interaction elements. In addition, we discuss our method for resolving the specific challenges encountered in structure determination of this symmetric oligomer by NMR techniques.

Results and Discussion

Determination of CLZ coiled coil stoichiometry

Full-length hnRNP C monomers assemble into tetrameric complexes *in vivo*[2] and *in vitro*.[41] The 28 residue peptide used in our analyses is identical to the hnRNP C oligomerization domain with the exception of the L180I mutation that resulted from subcloning into the particular expression vector used. Oligomerization domain constructs similar to that used in studies described here have been demonstrated to have a molecular mass consistent with the mass of a tetramer.[6] The NMR line widths in spectra obtained with our construct were consistent with the expected molecular mass of a tetramer of ellipsoidal shape as well (13 kilodalton (kDa), 10–15 Hz for ^{1}H spins covalently bonded to ^{15}N nuclei with an estimated correlation time of 6–8 ns).[42] However, because it is inherently impossible to determine the stoichiometry of our CLZ construct or any symmetric oligomer by NMR methods without *a priori* knowledge of the monomer molecular mass,[43] chemical cross-linking and

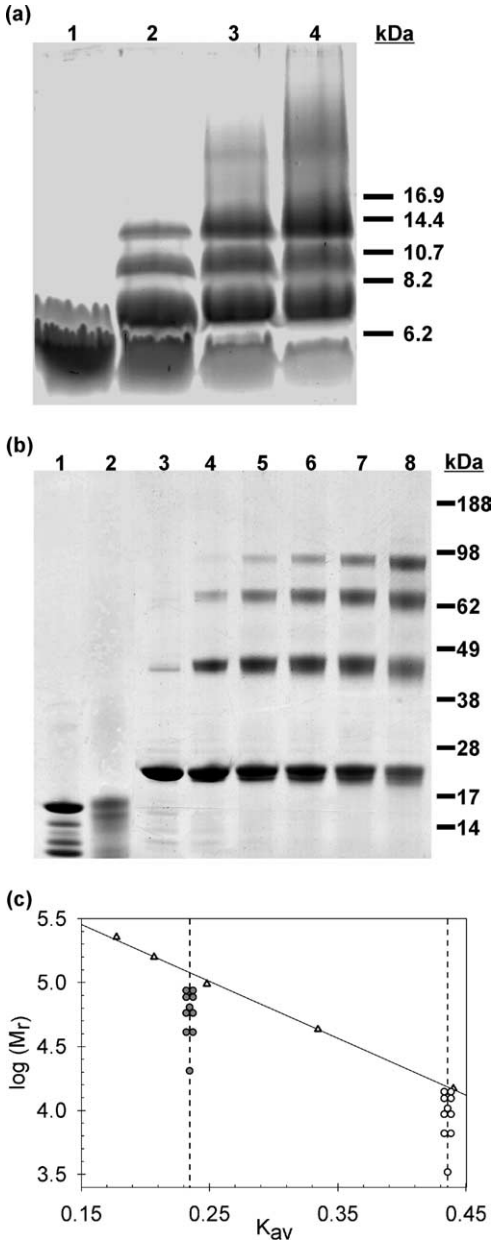

Figure 1. SDS-PAGE analysis and size-exclusion chromatography analyses of native and cross-linked CLZ tetramers in solution. (a) Cross-linking of CLZ oligomers with EDC. (1) No EDC, (2) 1.75 mM EDC, (3) 7.0 mM EDC, (4) 15.0 mM EDC. Gel lanes were overloaded in order to facilitate visualization of protein. The calculated mass of CLZ monomers is 3.3 kDa, dimers 6.6 kDa, trimers 9.9 kDa, and tetramers 13.2 kDa. Positions of molecular mass markers are shown on the right (M_r expressed in kDa). (b) Cross-linking of Trx-CLZ oligomers with BS^3. Cross-linking of 50 μM Trx fusion partner produced only monomers, confirming that oligomerization is CLZ-dependent. (1) Trx, no BS^3. (2) Trx, 10 mM BS^3, (3) Trx-CLZ, no BS^3, (4) Trx-CLZ, 0.2 mM BS^3, (5) Trx-CLZ, 0.5 mM BS^3, (6) Trx-CLZ, 1 mM BS^3, (7) Trx-CLZ, 2 mM BS^3, (8) Trx-CLZ, 5 mM BS^3. The calculated mass of Trx-CLZ monomers is about 20 kDa, dimers 40 kDa, trimers 60 kDa, and tetramers 80 kDa. Positions of molecular mass markers are shown on the right (M_r expressed in kDa). (c) Migration of Trx-CLZ and CLZ oligomers in size-exclusion chromatography experiments. The standard curve is plotted over the linear range of the column and the experimentally derived partition coefficient (K_{av}) of CLZ and Trx-CLZ are shown as dotted lines. The logarithm of the calculated mass of each Trx-CLZ (filled circles) and CLZ (open circles) oligomer is indicated by the position of oligomer symbols. Trx-CLZ eluted with an approximate mass 5.8 times the monomer and CLZ at an approximate mass 4.6 times the monomer.

size exclusion chromatography experiments were performed to verify complex stoichiometry in solution. Analyses employed both the free tetramerization domain and the 20 kDa thioredoxin (Trx)-CLZ fusion construct used for protein expression.

Cross-linking experiments performed with the same CLZ peptide used for NMR samples employed a zero-length (amino-carboxy) cross-linker (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride, EDC), which requires the close proximity of amino and carboxy groups that is needed for salt-bridge formation and can be predicted from the structure reported herein. Because chemical cross-linking reactions are second-order with short-lived reactive species and because reactive groups are mobile side-chains, cross-linking produces time-dependent mixtures of products. As a result, sodium-dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separated cross-linked tetramers in solution appear as a ladder containing tetrameric and all lower-order oligomeric species. The true stoichiometry of the cross-linked protein is detected as the highest order oligomer that accumulates with increasing doses of cross-linker.[44] Covalently linked CLZ tetramers accumulate in an EDC-concentration-dependent manner, the abundance of the tetramer far exceeding any non-specific larger covalent aggregates (Figure 1(a)). Interestingly, even in the absence of cross-linker, dimeric CLZ is detectable by SDS-PAGE, indicating that inter-subunit interactions are extremely stable. Cross-linking studies using Trx-CLZ fusion protein and a long-length (spacer distance of 11.4 Å) amino-reactive cross-linker (bis(sulfosuccinimidyl)suberate, BS^3), yielded similar results (Figure 1(b)). In this instance, use of the fusion protein affords simpler detection at low concentrations and produces a more native context of the hnRNP C oligomerization domain in that it is preceded N-terminally by a large globular domain and unstructured linker. Since the protein concentration was ten times lower in these experiments than for those using CLZ alone, high cross-linker concentrations produced no detectable large covalent aggregates. Furthermore, these cross-linking studies revealed that the dissociation constant of CLZ oligomers is smaller than 50 μM; thus, CLZ was observed as a tetramer in NMR experiments, where monomer concentration was ~2 mM.

Our cross-linking results could be interpreted as

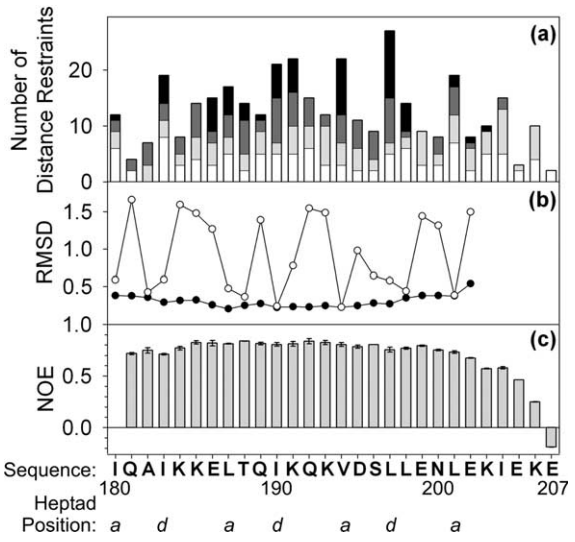

Figure 2. Analysis of NMR experimental results. (a) NOE-derived non-redundant distance restraints used in CLZ structure calculations. The number of intraresidue (white), sequential (light gray), medium-range (dark gray) and long-range (black) distance restraints are shown for each residue in the CLZ sequence. (b) RMS deviations at each residue for the final ensemble of 20 conformers relative to the average structure. The RMSDs of backbone N, C^{α} and C′ atoms (black circles) and heavy atoms (white circles) show the periodicity expected from the peptide sequence and heptad position. RMSDs are not shown for unstructured residues K203–E207. (c) $^{15}N\{^{1}H\}$ steady-state NOE for backbone amides of the CLZ sequence. Smaller values indicate increasing mobility within the molecule, characteristic of an unstructured region. Error bars represent values from duplicate experiments.

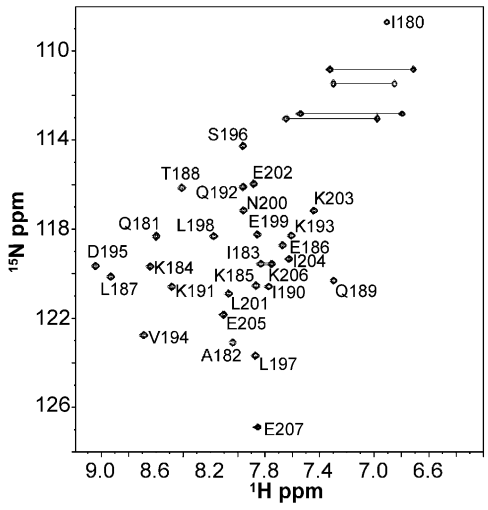

Figure 3. 2D $^{1}H–^{15}N$ HSQC spectrum of CLZ tetramers. Sample was 2 mM in CLZ monomer, pH 6.0, 27 °C. Backbone H^{N} resonance assignments are labeled in one letter amino acid code with residue number. Horizontal lines connect side-chain amide proton pairs of Asn and Gln.

an equilibrium between tetrameric and other lower-order oligomeric species; we therefore complemented this approach with gel filtration analysis. Size exclusion chromatography of CLZ and Trx-CLZ oligomers confirmed the presence of a single oligomeric species in solution at 10 μM concentration of monomers. SDS-PAGE analysis of the eluted fractions verified the correct monomeric mass. The elution volumes of both protein oligomers were slightly greater than that for ideal globular tetramers, but consistent with what would be expected given the ellipsoidal shape of an antiparallel coiled coil structure and migration properties of these oligomers (Figure 1(c)). Gel filtration studies performed by Shahied *et al.* using a nearly identical oligomerization domain construct revealed an oligomer with 3.8 times the molecular mass of a monomer.[6] Our size exclusion experiments conclusively show that no detectable oligomers smaller than a tetramer form in solution. Collectively, the preponderance of evidence presented here as well as the extensive literature addressing the stoichiometry of C protein and the oligomerization domain in a biological context indicates that CLZ oligomers are observed as tetramers in our NMR studies.

Structure determination

Circular dichroism spectroscopy,[10] NMR chemical shift indices and cross-peak patterns in 2D $^{1}H–^{1}H$ nuclear Overhauser effect (NOE) spectra[6] predict mostly α-helical structure for peptide constructs that isolate the oligomerization domain of hnRNP C. Inspection of the CLZ sequence could suggest that all 28 residues are likely α-helical, as the Ile, Leu and Val residues that occupy *a* and *d* coiled coil heptad positions would form the hydrophobic face of an amphipathic helix predicted to comprise the tetramer core. However, chemical shift index analyses[45] using resonance assignments for H^{α}, C^{α}, C^{β} and C′ nuclei revealed an α-helical segment composed only of residues Q181–E202; the remainder of the sequence, including I204, is predicted to have a random coil structure. The algorithm STABLECOIL[46] also predicted significantly higher coiled coil stability for the first 23 amino acid residues than the C-terminal residues. In addition, only three medium and long-range NOE interactions were observed for K203–E207 (Figure 2); during structure calculations, residues in this region sampled a broad range of conformations with backbone and heavy atom root-mean-squared deviations (RMSDs) of 1.3–7.0 Å and 3.3–7.8 Å, respectively. The $^{15}N\{^{1}H\}$ steady-state heteronuclear NOE measurements showed residues 181–202 with NOE/no NOE ratios above 0.6 and residues 203–207 showing rapidly decreasing values. Therefore, a 22 residue segment that comprises only three heptads is of requisite length in the CLZ sequence to adopt a coiled coil fold, in contrast to the original 28 residue sequence (four heptads) proposed.[6,10] While the oligomerization

Table 1. Average internuclear distances for the ensemble of 20 NMR conformers corresponding to selected cross-peaks observed in NOE spectra

Interacting nuclei	Heptad position	Intrahelical distance (Å)	α1–α2 Interhelical distance (Å)	α1–α4 Interhelical distance (Å)
K191 H^{α}–V194 H^{β}	*e-a*	2.6(±0.1)	11.1(±0.1)	4.3(±0.1)
K191 H^{α}–V194 $H^{\gamma 2}$	*e-a*	4.6(±0.6)	8.8(±0.1)	5.6(±0.7)
I190 $H^{\gamma 2}$–V194 H^{N}	*d-a*	4.8(±0.7)	6.2(±0.7)	6.3(±0.7)
L187 H^{N}–L198 $H^{\delta 2}$	*a-e*	18.4(±0.6)	11.7(±0.8)	4.8(±0.6)
I183 H^{α}–L197 $H^{\delta 2}$	*d-d*	20.5(±1.1)	3.6(±0.9)	10.2(±1.1)

Standard deviations are given in parentheses.

interface in coiled coils is typically longer, three heptads were previously found to be sufficient for tetramer assembly of the lactose operon (Lac) repressor.[28]

2D ^{1}H–^{15}N heteronuclear single quantum coherence (HSQC) spectra[47] revealed well-resolved cross-peaks that correlated all directly attached backbone and side-chain amide ^{1}H–^{15}N nuclei pairs (Figure 3). Only one resonance for each ^{1}H–^{15}N nucleus pair in the CLZ sequence was observed. All scalar and dipolar coupling experiments also unambiguously produced one set of cross-peaks for each residue of the sequence. This observation is consistent with a structure where CLZ tetramers are composed of symmetrically packed subunits. However, because equivalent nuclei of symmetric subunits have degenerate chemical shifts, an independent verification of the oligomeric state of CLZ in solution by biochemical methods was essential as a basis for structural calculations of the tetramer. The orientation of helices in the tetramer was determined through examination of NOE-derived internuclear distances between nuclei positioned remotely in the peptide sequence. For example, accounting for both intra- and intermonomer origins, the observed cross-peak between L187 H^{N} and L198 $H^{\delta 2}$ requires a separation of at least 18 Å for a parallel orientation of CLZ helices; however, an antiparallel orientation of helices positions these nuclei within the observable range for NOE cross-relaxation (Table 1). Therefore, the single set of cross-peaks observed in all CLZ spectra and the interhelical distance constraints derived from NOE spectra collectively define a tetramer with helices oriented antiparallel to both adjacent helix partners (Figure 4).

As evidenced by the rarity of symmetric oligomers solved by NMR methods in the Protein Data Bank,[48] the inherent complexities associated with assignment of observed NOEs[43] provide additional challenges in structural determination. For a symmetric helical tetramer, each cross-peak could potentially represent a combination of multiple nuclear interaction types, including intrahelical and interhelical interactions across α1–α2 type, α1–α3 type and α1–α4 type interfaces (helix designators α1, α2, α3 and α4 are shown in Figure 4). In order to determine the class of nuclear interaction that produced each observed cross-peak in CLZ spectra, distance restraints were carefully examined based on models of an α-helical monomer in the context of an antiparallel tetramer, with consideration for the position of each residue on the helical wheel (Figure 4). α1–α3 interactions are typically negligible as a result of the $1/r^{6}$ proportionality of cross-peak intensity to the internuclear distance (r) and the diameters of α-helices. Moreover, because this type of interhelical interaction occurs between nuclei in helices that are oriented in parallel, such interactions cannot be distinguished from intrahelical interactions in NMR spectra. Thus, with few exceptions, intraresidue, sequential and medium-range distance constraints could be assigned as intramonomer. In the context of an antiparallel

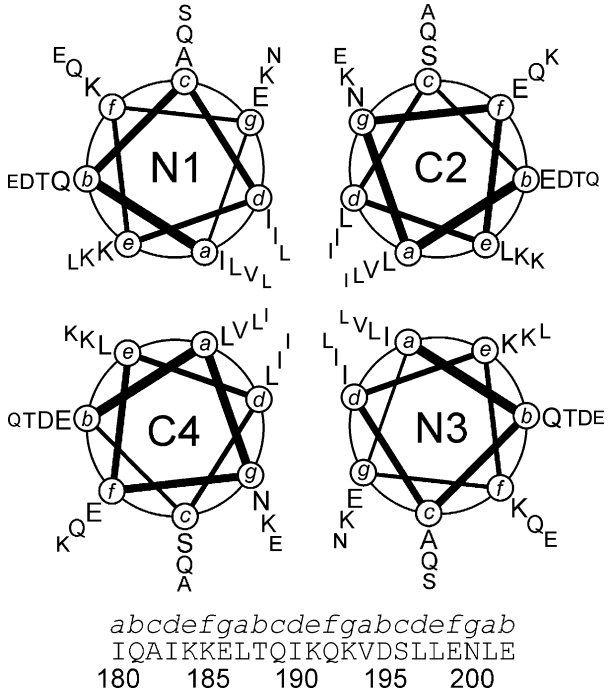

Figure 4. Antiparallel helical wheel diagram of the CLZ tetramer, residues I180–E202. Each helix is labeled with an arbitrarily assigned number (1–4) and N or C denoting the closest polypeptide terminus, where helical axes are oriented perpendicular to the page. Heptad positions are indicated by circled letters *a*, *b*, *c*, *d*, *e*, *f* and *g* and are connected by lines (thickest for those closest) to follow the sequential order of residues (in one letter amino acid code). Smaller font amino acids are farther from the viewer. Residue numbers are as given in sequence below.

Table 2. Structural statistics for the final 20 energy minimized conformers representing the NMR ensemble of CLZ structures

Distance restraints (*non-redundant*)	
Intraresidue (per helix)	108
Intrahelix (per helix)	
Sequential	97
Medium range (2–4 residues)	90
Interhelix (long range, ≧5 residues)	
α1–α2-type (per tetramer)[a]	60
α1–α4-type (per tetramer)[a]	68
Hydrogen bonds (per helix)	19
Total distance restraints (per tetramer)	1384
Angle restraints (*per helix*)	
ϕ	22
ψ	22
χ_1	10
χ_2	7
CYANA target function (Å^2)	2.79 (±0.29)
Residual distance restraint violations (Å) (*upper limit*) (*per structure*)	
Number of violations ≧0.2 Å	3 (±2)
Sum of violations	8.6 (±0.7)
Maximum violation	0.37 (±0.12)
Residual angle restraint violations (*degrees*) (*per structure*)	
Number of violations ≧5°	0
Sum of violations	44.7 (±2.8)
Maximum violation	1.37 (±0.18)
AMBER energies (*kcal/mol*)	
Input CYANA structures	−2421.5 (±697.9)
Final energy minimized structures	−4855.6 (±21.9)
Sum of distance penalties	17.4 (±3.0)
RMS deviation from ideal geometry	
Bond lengths (Å)	0.0102 (±0.0001)
Bond angles (degrees)	1.964 (±0.025)

Standard deviations are given in parentheses.
[a] α1–α2-type and α1–α4-type restraints are described in Materials and Methods.

orientation of α-helices, all long-range interactions between nuclei more than one helix turn apart are exclusively interhelical, and as a result of their position in the heptad repeat, most were unambiguously defined as exclusively α1–α2 or α1–α4 type *a priori*. In this respect, observed cross-peaks between *a* and *e* position nuclei defined internuclear interactions along α1–α4 type interfaces; similarly, long-range interactions arising from cross-relaxation of nuclei at *d* and *g* heptad positions defined α1–α2 type interfaces (Table 1).

However, because spins of each oligomer subunit have identical chemical shifts, effectively only one set of spins is observed, so that intrahelical interactions are not inherently distinguished from interhelical ones. Thus, internuclear distances between residues that lie near symmetry axes, occupying core or interfacial positions in middle heptads of the helix (I190, K191, K193 and V194) were ambiguous, the observed cross-peaks between them potentially containing contributions from both intra- and interhelical interactions. For example, while the observed K191 H^{α}–V194 H^{β} cross-peak possibly contains both intramonomer and α1–α4-type interhelical contributions, the internuclear distances observed following initial structure calculations revealed that 96% of the cross-peak intensity is contributed by intrahelical interactions. In contrast, intrahelical interactions between K191 H^{α} and V194 $H^{\gamma 2}$ account for only ∼80% of cross-peak intensity; the remaining 20% results from interactions across α1–α4-type interfaces. In the most complex scenario (interactions between I190 $H^{\gamma 2}$ and V194 H^{N} nuclei), internuclear distances were within the observable range for NOEs for intrahelical, α1–α2-type and α1–α4-type interactions (Table 1).

Ambiguous distance constraints were assigned to a particular class of interaction through iterative torsion angle dynamics calculations that refined the structure. Results of final structure calculations are given in Table 2. No α1–α3-type distance constraints were included in final structure calculations. In fact, an analysis of α1–α3-type interatomic distances in the final structure indicates that only five unique interactions could possibly contribute greater than 5% of resonance intensity to observed cross-peaks in NOESY spectra, interactions which did not correspond to any intrahelical constraints included in calculations. Ramachandran plot analyses evaluating the relative frequencies of occurrence of CLZ Q181-E202 (ϕ,ψ) angle pairs in secondary structure elements show that 100% fall in the most favored α-helical regions. Moreover, RMS deviation from ideal bond and angle geometry, hydrogen bond energies, ω,ϕ,ψ, and χ torsion angles for each residue and circular variances of core residue torsions indicated high quality of the structure. The energy-minimized (AMBER, version 8) ensemble of NMR conformers had average backbone and heavy atom RMSDs of 0.3 Å and 0.9 Å, respectively, for residues that compose the coiled

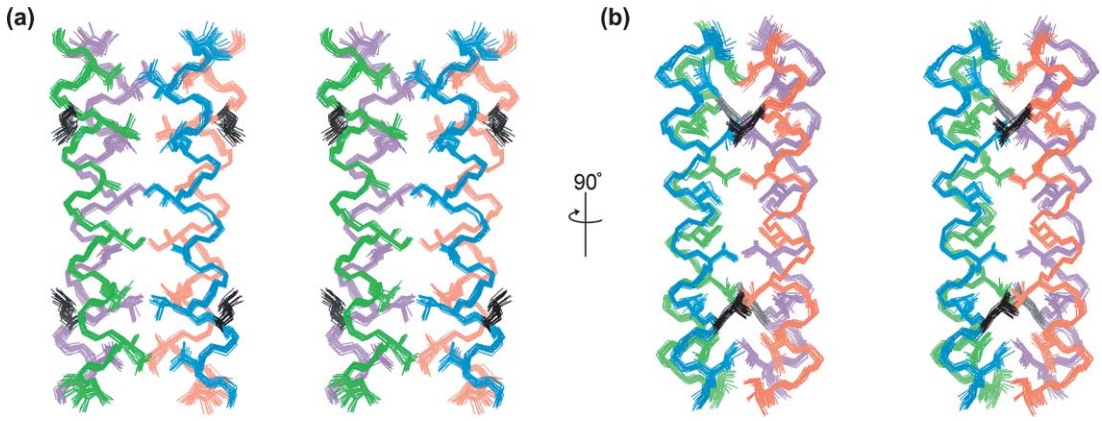

Figure 5. Two 90°-rotated stereo representations of the ensemble of 20 energy minimized NMR conformers, residues I180–E202 viewed along the (a) α1–α2 and (b) α1–α4 side. Structures were superimposed on Q181–E202 N, C^{α} and C' atoms. Side-chain heavy atoms of all residues occupying *a* and *d* heptad positions and L198 (L198 in black) are also shown.

coil (I180–E202). Residues that occupy *a* and *d* core heptad positions and form tetrameric interfaces are significantly better defined, having a heavy atom RMSD of 0.4 Å (Figure 5).

Topology of helices in CLZ tetramers

Helices of hnRNP C oligomerization domains associate to form an antiparallel four-helix coiled coil, where slight left-handed supercoiling allows helices to maintain contact interfaces along the length of the tetramer. Classically described coiled coil knobs-into-holes packing interactions are formed at residues occupying core *a* and *d* heptad positions as well as *e* and *g* heptad position residues that are tightly packed in helical interfaces. The geometry of α-helical subunits that compose the tetramer is actually more similar to ideal α-helices than to helices that form ideal left-handed coiled coils. On average, each turn of CLZ helices contains 3.57 residues; thus, the hydrophobic surface created by *a* and *d* position side-chains drifts slightly around each helix. In combination with a large helical radius of curvature (545 Å), the distance that helices could remain in contact is significantly shortened. While the ideal left-handed supercoil arrangement allows for helices to remain in contact over much longer distances by more pronounced

Table 3. Geometric parameters that describe the coiled coil formed by hnRNP C oligomerization domains

Monomer parameters	
Residues/turn[a]	3.57
Rise/residue[a] (Å)	1.50
Pitch (Å)	5.35
Radius of curvature[a] (Å)	545
Radius[b] (R_1) (Å)	1.5
Length[b] (Å)	33.7
Coiled coil parameters	
Pairwise helix crossing angles[b] (degrees)	
α1–α2	4.7
α1–α4	10.95
Interhelical axis distances[b] (Å)	
α1–α2	9.7
α1–α4	7.9
Residues/supercoil turn[a]	262
Radius[a] (R_0) (Å)	7.2
ϕ_C[a] (degrees)	25.8
Superhelix crossing angle[a] (degrees)	13.2
RMSD from ideal coiled coil geometry[a] (Å)	2.3
Buried surface area[b] (Å^2)	4674
Average number of complete knobs into holes packing interactions (per helix)[c]	7

[a] Values were determined by fitting atomic coordinates of the average CLZ tetramer structure (I183–E199) to ideal supercoil geometry. R_1 is the vector from the helix axis to the C^{α} of residues in *a* heptad positions; R_0 is the vector connecting each helix axis to the superhelix axis; ϕ_C is the angle between R_1 and R_0 for the same helix. Pairwise helix crossing angles describe the angle between helix axes of adjacent helices, and superhelix crossing angle describes the average angle of helix axes with respect to the superhelix axis.

[b] Values were determined with program defined functions of MOLMOL using the lowest energy structure.

[c] Packing interactions for knobs surrounded by four residues within 7.4 Å that form a complete hole were determined with SOCKET using the lowest energy structure.

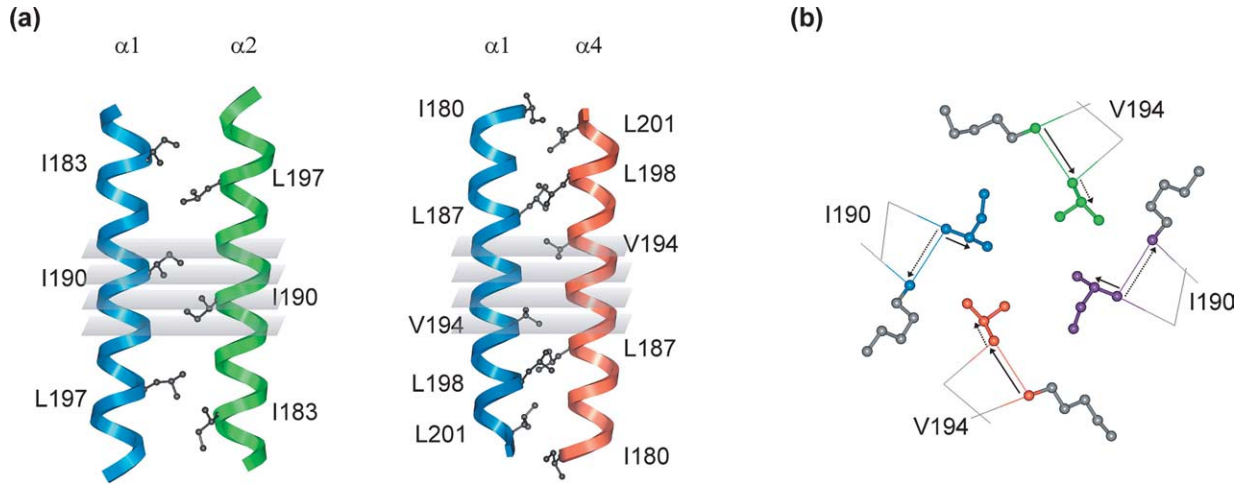

Figure 6. Packing interactions of hydrophobic residues in the lowest energy conformer of the NMR ensemble of structures, residues I180–E202. (a) Stacking of hydrophobic residues inside the core. At the α1–α2 interface, all *d* position residues are displayed. At the α1–α4 interface, all *a* position residues are displayed in addition to an *e* position hydrophobic residue, L198, which is buried in the tetrameric core and packs against L187 of the adjacent helix. Side-chain layers created by V194 (α2,α4), I190 (α1,α3), I190 (α2,α4) and V194 (α1,α3) are indicated with planes. (b) Packing angle geometries for V194 (*a* position) and I190 (*d* position). C^{α} backbone trace and side-chain heavy atoms of *a* and *d* heptad position residues are displayed; side-chains of *e* and *g* position residues that form the sides of packing cavities are colored in gray. Vectors connecting C^{α} atoms of holes and vectors along C^{α}–C^{β} bonds show that V194 packs perpendicularly with respect to I190 and K191 and I190 packs parallel with respect to K193 and V194.

helical unwinding (3.5 residues/turn) and curvature, helical bundles in globular proteins are generally shorter and straighter,[29] consistent with CLZ helices composing the core of a larger hnRNP C tetrameric complex.

CLZ tetramer structure deviates from ideal coiled coil geometry principally as a consequence of unequal dimensions and straighter helices that break the pseudo 4-fold symmetry, resulting in a small superhelix crossing angle and large superhelical pitch (391 Å). In this respect, the helix axis separations and pairwise helix crossing angles across α1–α2-type and α1–α4-type interfaces are similar but not identical (Table 3), the large radius of curvature of CLZ α-helices requiring helix axes to adjust closer to a 180° crossing angle in order to maintain the contact interface. In general, differences in the pairwise dimensions and crossing angles of four-helix bundles can be correlated to various effects specific to sequence and interhelical interactions, such as the size of side-chains that occupy core and interfacial positions in the assembly or asymmetric hydrophobic and/or electrostatic interactions between the two sides.

Although four-helix bundle assemblies are a frequently occurring protein folding motif in the structural database, few of these structures form the required knobs-into-holes packing interactions that define coiled coil packing. Like CLZ tetramers, helices in the Lac repressor[49] and repressor of primer (Rop)[50] four-helix coiled coils associate antiparallel to adjacent helices. In contrast, the four helix interfaces of neuronal synaptic fusion complex (SFC)[51] and GCN4-pLI tetramers[32] are formed by assembly of helices that are oriented in parallel. While these structures all exhibit the left-handed supercoiling of helices that is typical of coiled coils composed from sequences having a heptad repeat, the specific interactions that are formed at helix interfaces introduce variation in the topology of coiled coil structure. The 34 Å length of the CLZ coiled coil is somewhat short in comparison to the length of other four-helix coiled coils, where four-helix interaction surfaces range in length from ~30 Å in the Lac repressor to ~85 Å in SFC. Moreover, the superhelix crossing angle in the CLZ coiled coil (Table 3) is also slightly less than for other four-helix coiled coils (which ranges between 17° for Rop and 31° for SFC), being intimately linked with the large radius of CLZ helix curvature. The 7.2 Å supercoil radius (R_0) of CLZ tetramers is consistent with that of GCN4-pLI (7.5 Å) and SFC (7.0 Å), but varies more significantly from that of Rop (6.6 Å) and the Lac repressor (6.7 Å) as a result of overrepresentation of small Ala residues at core and interfacial positions in Rop and Lac repressor sequences. The variance of these helical parameters across coiled coil structures illustrates that small adjustments within the supercoil can accommodate a range of sequence-specific interactions in maintaining the coiled coil packing that is important to stability of these structures.

Contributions of specific packing interactions to CLZ coiled coil stability

Knobs-into-holes packing interactions of core and interfacial residues at interhelical contact surfaces of the CLZ supercoil are described by the projection of side-chains from one helix into cavities on adjacent helices created by four side-chains that

are related in sequence by $i-3$, i, $i+1$ and $i+4$.[52] At the α1–α4 interface, the side-chains of *a* position residues extend into cavities created by *a*, *d*, *e* and *a* side-chains (forming the bottom, two sides and top of the hole, respectively) and *e* position side-chains by *e*, *a*, *b* and *e* side-chains of the opposing helix; similarly, knobs from *d* and *g* position residues on α1 have packing interactions with *dgad* and *gcdg* holes on α2 and *vice versa*. Adjacent CLZ subunits are slightly offset along the long axis of the complex in order to accommodate the stacking of side-chains from the same heptad position vertically along the superhelical axis that is directed by an antiparallel orientation of helices. Helix offsets at the α1–α2 and α1–α4 interfaces are about 0.2 heptad units, similar to vertical offsets of helices in the Lac repressor (0.22),[29] such that the turns of helix backbones interdigitate with respect to neighboring helices, rather than lying directly across from each other. Consequently, alternating layers of *a* and *d* position side-chains in the tetrameric core contain two identical side-chains from non-neighboring helices. This layering pattern continues along the length of the coiled coil core (Figure 6).

For coiled coil structures with no helix offset, such as the classically described leucine zipper and all coiled coils with parallel helices, or where helix offsets are about 0.5, such as in Rop, the turns of adjacent helix backbones lie directly across from each other so that one side-chain from each helix composes each layer in the coiled coil core. For a tetramer composed of identical subunits, helices oriented in parallel have layers formed by four identical residues while an antiparallel orientation of helices generates mixed layers composed of two identical *a* and two identical *d* position residues. Assuming variation among residue types at *a* and *d* heptad positions along the core, antiparallel helix orientations might provide more efficient packing interactions as a result of the greater flexibility afforded by mixed layers.[30,31] In contrast, specific packing interactions of the I, L and V residue side-chains that occupy *a* and *d* positions in the CLZ coiled coil likely make only minor contributions to determining helix orientation in the tetramer because they are not of significantly different size. However, the interdigitation of side-chains described by the offset of neighboring CLZ helices allows slightly tighter packing to create an even distribution of hydrophobic side-chains over the length of the tetrameric core, maximizing contact surfaces for van der Waals interactions.

Effects of hydrophobic surface burial and packing angle geometries on CLZ oligomerization

In light of the multitude of factors that have been shown to play a role in determining the conformational specificity of coiled coils,[17,20,32,53] there is likely not a sole determinant of the tetrameric stoichiometry of CLZ helices. However, using a structural approach, we examined the possible contributions of hydrophobic surface burial and packing angle geometries to the stability of tetramer formation in the context of other coiled coil structures. It has been previously proposed that the angle at which specific side-chain knobs orient with respect to their corresponding holes can determine the stoichiometry of coiled coils.[32] Defined by the angle between a vector along the C^{α}–C^{β} bond of each knob and the C^{α}–C^{α} vector connecting the sides of the corresponding hole, packing angle geometries in dimers are parallel ($\sim 30^{\circ}$) for *a* position residues and perpendicular ($\sim 90^{\circ}$) for *d* position residues, and are perpendicular for *a* position residues and parallel for *d* position residues in four-helix coiled coils; acute packing angles ($\sim 60^{\circ}$) are observed at both *a* and *d* positions for trimeric coiled coils.[32] In accordance, the side-chains of *a* position residues in CLZ tetramers pack with perpendicular geometry and those of *d* position residues pack parallel to their respective holes (Figure 6). Therefore, in the CLZ tetrameric core, mixed *a*/*d* side-chain layers have mixed packing geometries. In spite of significant differences in the composition of directly opposing and surrounding side-chains with respect to their heptad position, directionality and helix offset, packing angle geometries are conserved at heptad positions regardless of the orientation of helices within the four-helix coiled coil, since for symmetrically spaced amphipathic α-helices, *a* position C^{α}–C^{β} bonds always point directly towards an adjacent helix and *d* position knobs always point away from their respective helix interface.

Because packing angle geometries at core heptad positions are different depending on the number of helices that compose a coiled coil, it has been suggested that the identity of residues occupying *a* and *d* positions can determine the complex stoichiometry as a result of the preferences of their most favored rotamer conformation for specific packing geometries.[32] This hypothesis is largely supported by thermodynamic stability analyses of GCN4[32] and GCN4-derived dimeric and trimeric coiled coils,[34,35] where residues having β-branched side-chains are energetically favored in positions that require parallel knobs-into-holes packing arrangements and Leu residues in positions that exhibit perpendicular packing. Despite these preferences, the identity of residues at *a* and *d* heptad positions in the CLZ sequence is mixed; β-branched residues occupy positions having both parallel (I183 and I190) and perpendicular (V194) packing geometries, and Leu residues occupy both parallel (L197) and perpendicularly (L187 and L201) packed core positions. Although the majority (66%) of these residues follow the rotamer preferences for favorable packing energetics in GCN4 and GCN4-derived dimeric and trimeric peptides, χ_1 and χ_2 side-chain dihedral angles of all CLZ *a* and *d* position residues having significant order ($S > 0.8$) fall in favorable energy minima regardless of their packing geometries. Thus, assembly of the tetrameric CLZ coiled coil does not appear to be entirely

governed by conformational preferences of core heptad position residues for particular packing geometries. This is consistent with a weaker correlation of rotamer preferences and packing angle geometries in thermodynamic analysis of the four-helix coiled coil GCN4-pLI.[33]

In fact, across coiled coil structures, the covalent geometries and polarities of residues that form the sides of opposing packing holes are equally as important as the specific rotamer preference of knob residues to the assembly of helices in a particular oligomeric state and orientation. The mixed perpendicular and parallel packing geometries of antiparallel coiled coils allows increased flexibility in side-chain selection to provide more efficient packing interactions in the core than discrete *a* or *d* layers of parallel coiled coils, antiparallel structures actually containing a broader spectrum of amino acids than parallel ones.[53] In addition, packing angle geometries can deviate significantly across coiled coil structures.[36,53] Minimization of packing energies at core positions in coiled coil oligomers can define the stoichiometry and topology of helices in the coiled coil by increasing the stability past a threshold for coiled coil assembly. However, while specific knobs-into-holes interactions of CLZ helices are maximized to achieve optimal packing as a consequence of the stoichiometry and orientation of helices in the coiled coil, the influence of rotamer preference on the observed stoichiometry may be secondary to more influential effects.

The entropic gain in protein surroundings that is produced by occlusion of aliphatic side-chains from solvent is probably the most influential driving force behind any complex assembly and is likely a significant determinant of CLZ coiled coil stoichiometry. Formation of the coiled coil interface buries nearly 4700 Å^2 of surface area (approximately 1170 Å^2 per monomer), typical of biologically relevant protein–protein interfaces.[54] Based on the ratio of buried surface areas in helices of GCN4 dimers and GCN4-pLI tetramers,[32] if CLZ coiled coils were dimeric, each helix would only bury ~650 Å^2, well below the threshold for high-affinity protein–protein interactions. Assembly of CLZ helices results in complete burial of hydrophobic residues that occupy middle heptad *a* and *d* positions in the tetramer core. In addition, significant burial of *e* and *g* position residues at helix interfaces is achieved, *e* position residues that lie along α1–α4-type interfaces being slightly more buried (84%) than *g* position residues along α1–α2-type interfaces (77%), directly correlating with shorter helix axis separations along α1–α4-type interfaces.

Analysis of other four-helix coiled coils reveals that the proclivity of a protein sequence for a particular stoichiometry appears to be driven in part by the polarity of residues at core and interfacial heptad positions. While a buried *a* position Asn residue specifies a preference for dimer stoichiometry in native GCN4 sequences,[32] non-polar *e* position Ala residues direct higher order oligomerization in the Lac repressor.[36] Mutation of charged residues at *e* and *g* heptad positions to Ala[55] or *e* position Ala residues to charged amino acids[38] has been shown to result in a switch between dimer and tetramer formation in coiled coils. Although the primary interactions of L198 occur only along α1–α4-type interfaces, this *e* position residue may be important in directing the stoichiometry of CLZ helical assemblies. While the hydrophobic surface in coiled coil helices that is created by *a* and *d* position residues spans about 29% of individual helix circumference, the hydrophobic surface created by addition of an *e* position residue, L198, locally extends to around 43% of the helix circumference. This suggests that changes in the juxtaposition of helices away from a dimer geometry (where the core is formed only by *a* and *d* residues) are required in order to minimize solvation energies of all three residues, because burial of the extended hydrophobic surfaces cannot be satisfied by interaction with only one opposing helix. In addition, since CLZ helices are oriented antiparallel to each other and the total thermodynamic contribution of *e* position Leu residues in coiled coils has been proposed to be similar to Lys-Glu salt bridges in the same location,[20] L198 likely makes significant contributions to the stability of the tetramer. Interestingly, a sequence similar to the CLZ LLENLE (*d-b* heptad positions) sequence, LLEKLN in Rop (also *d-b*, the second L corresponding to L198 in CLZ), forms interfacial interactions along one side of the supercoil in the antiparallel four-helix coiled coil formed by two identical polypeptide chains.

Contributions of electrostatic interactions to tetramer assembly and stability

While the contributions of Coulombic forces to stability are significantly less than the energetic contributions of van der Waals packing effects and desolvation of hydrophobic residues, interactions between charged and uncharged polar side-chains play defining roles in the conformational specificity of coiled coil assembly. Disregarding the influence of any other effects on the orientation of helices, interchain electrostatic attractions of helix dipoles favor an antiparallel orientation of helices.[39] However, charge density in coiled coil structures is generally high, especially along solvent-exposed interfaces of interhelical contacts, providing numerous opportunities for favorable electrostatic attractions between opposing charges.[17] In fact, antiparallel orientations of helices are the most commonly described topology for four-helix bundles, and while a parallel orientation of helices is the most common orientation in two and three-helix coiled coils,[36] naturally occurring domains consisting only of parallel four-helix bundles are rare in the Protein Data Bank.[48]

Sixty nine percent of CLZ *b*, *c*, *e* and *g* position residues have side-chains that are either charged or polar. Optimal compensation of localized charges

and the overall charge distribution along each helix establishes an antiparallel orientation of CLZ helices. Examination of the charge distribution along CLZ helices reveals a concentration of net positive charge (+3) at the N-terminal end and net negative charge (−3) at the C-terminal end of each helix when helices are divided by a plane at V194; as a result, assembly of helices in an antiparallel orientation clearly offers greater opportunity for charge compensation to enhance coiled coil stability than a parallel orientation of helices. Based on interatomic distances in the final structure, several electrostatic and hydrogen-bonding interactions can be identified that drive assembly and promote the stability of the antiparallel tetramer. At the center of the structure, an *e-b* salt-bridge interaction between K191 and D195 provides local charge compensation at α1–α4-type interfaces (Figure 7). However, while this favorable ionic attraction contributes to tetramer stability, it likely does not specify antiparallel helix orientation, as a *g-b* interfacial interaction that could occur if helices were oriented in parallel would potentially provide a favorable interaction between K193 and D195.

Therefore, charge–charge interactions near the ends of the sequence are likely more influential in favoring an antiparallel orientation of helices. In this respect, several opportunities for favorable ionic side-chain interactions are present at helix interfaces around the antiparallel tetramer. At α1–α4-type interfaces, a *b-e* interhelical interaction between K184 and E202 would provide local charge compensation at the ends of the coiled coil. An interaction between side-chains of two *g* position residues, E186 and N200, can stabilize α1–α2-type interfaces through formation of a hydrogen bond. In addition, K193 is located proximally to both E186 and Q189 on α1–α2-type interfaces. Side-chains of all these residues sample a broad range of conformational space (Figure 2) and no direct NOE evidence for interaction was observed in the spectra. However, because E186 has multiple potential interhelical interactions at the ends of the coiled coil, it is probable that it plays a role in favoring an antiparallel orientation of helices.

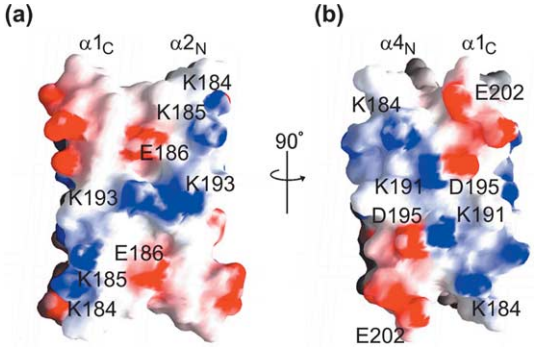

Figure 7. Electrostatic surface potential representations of CLZ coiled coils, residues Q181–E202. (a) α1–α2 side and (b) α1–α4 side are displayed with the C terminus of α1 (denoted as $\alpha1_C$) and the N terminus of the adjacent helix at the top. Molecular surfaces were generated using GRASP and shaded according to electrostatic surface potential for side-chain atoms with full charge. The spectrum of red, white and blue color spans from −7.8 through 0 to +11.4 kT/e.

Electrostatic and hydrogen-bonding interactions within individual helices promote the intrahelical stability that is an essential component of the structural integrity of the tetramer. In addition to the expected hydrogen bonds that form between *i* and *i*+4 residues of all α-helices, a hydrogen-bonding interaction occurs between T188 $H^{\gamma1}$ and the K184 backbone oxygen to provide local intrahelical stability. This interaction may be enhanced as a result of the local hydrophobic environment created by interhelical packing proximal to L198 (Figure 4). Similarly, a large percentage of threonine side-chain hydroxyl protons in transmembrane helices have similar hydrogen-bonding interactions with *i*−3 or *i*−4 backbone carbonyl oxygen atoms.[56] Q181, Q189, Q192 and N200 are all positioned proximally to Lys residues within CLZ helical subunits, potentially providing additional hydrogen-bonding interactions that contribute to intrahelical stability. Interestingly, these polar residues are located at solvent accessible positions in the CLZ tetramer and are spaced equidistant from the center of the tetramer. Thus, in the context of a symmetric hnRNP C complex where other hnRNP C domains are packed around the coiled coil core, these residues may instead play roles in protein–protein interdomain interactions.

As a consequence of antiparallel CLZ helix orientation, a six residue basic ridge composed by K184, K185 and K193 is formed spanning α1–α2-type interfaces of the tetramer (Figure 7). Although K184 and K193 are potentially involved in interhelical contacts, this basic ridge of residues may provide an interaction surface for the acidic C-terminal domain of hnRNP C that follows the oligomerization domain in sequence. While engineered salt bridges on the solvent-exposed surface of proteins and α-helical peptides generally contribute only 0.2–0.5 kcal/mol to stability, ionic interactions that are buried in the core of a globular protein have been estimated to be as high as 3–5 kcal/mol,[57] and hydrogen bond strengths are much higher in hydrophobic environments than in aqueous solvent.[58] Therefore, the contributions of hydrogen-bonding and electrostatic interactions of CLZ residues are not only important in oligomerization of helices in the coiled coil, but also in providing potential protein–protein interaction interfaces with other domains packed against CLZ in the context of full-length globular hnRNP C tetramers.

Functional biological evidence for the importance of specific residues to CLZ coiled coil structure

Sequences of the oligomerization domain of hnRNP C homologs are well conserved, exhibiting 70–96% identity with residues that form the coiled

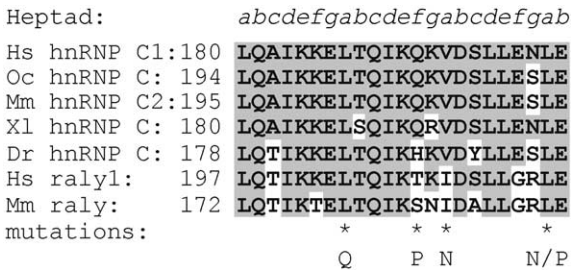

Figure 8. Effects of specific amino acid substitutions on tetramer stability. Sequence alignment of CLZ homologs. Sequences are well conserved in positions that stabilize unique tetramer structure. Sequences are as follows: *Homo sapiens* hnRNP C1 isoform (Genbank accession no. P07910);[3,89] *Oryctolagus cuniculus* hnRNP C (Genbank accession no. AAC61695);[90] *Mus musculus* hnRNP C2 isoform (Genbank accession no. Q9Z204);[91] *Xenopus laevis* hnRNP C (Genbank accession no. A31765);[92] *Danio rerio* hnRNP C (Genbank accession no. AAQ97793); *Homo sapiens* raly1 (Genbank accession no. AAO85516); *Mus musculus* raly (Genbank accession no. A47318).[93] Identical residues are highlighted in gray. Substitutions at sites denoted by stars disrupt tetramer assembly.[9]

coil in the oligomerization domain of human hnRNP C (Figure 8). Oligomerization interfaces formed by these sequences likely have the same topology as CLZ tetramers because residues that are essential to the conformational specificity and that contribute the majority of stability to coiled coil assembly are conserved. Specifically, the sequence determinants that define coiled coil stoichiometry by desolvation of hydrophobic surfaces, including the *e* position residue L198 and residues occupying core *a* and *d* heptad positions, are identical, with the exception of a conservative V194I substitution in raly proteins. Charged residues occupying *b*, *e* and *g* positions that lie at helix interaction interfaces and are critical to maintaining structural integrity (K191 and D195) and the distribution of charge that promotes an antiparallel assembly of helices are also well conserved. Moreover, most amino acid substitutions are conservative with respect to side-chain character and are not predicted to contribute greatly to the overall stability of the four-helix coiled coil.

In contrast, protein–protein oligomerization interactions in full-length hnRNP C are disrupted by randomly generated and engineered L187Q, Q192P, L201P and V194N/L201N mutations.[9] Burial of polar Gln and Asn residues at *a* heptad positions likely increases the energetic costs of desolvation for these specific residues, in addition to disturbing van der Waals packing and solvation of opposing hydrophobic residues that form cavities for knobs-into-holes interactions along the core. As a result of very low helical propensity,[21] substitution of a Pro residue near the center of each helix (Q192P) likely causes conformational rearrangements that cannot accommodate formation of the coiled coil interface; moreover, although the resulting helix kink that would be induced by L201P mutation lies at the termini of tetrameric interfaces, the shortened helices are apparently not of requisite length for coiled coil assembly. In this respect, our torsion angle dynamics modeling studies of these proline mutant CLZ tetramers revealed an eight to tenfold increase in the target energy function as a result of steric clashes.

The antiparallel coiled coil domain: implications to hnRNP C complex architecture and function

The orientation and stoichiometry of helices in the CLZ oligomer interface is elemental in defining domain architecture and biological functions of hnRNP C. The thermodynamic stability that is achieved by coiled coil assembly, experimentally evidenced in the high melting temperature of CLZ tetramers (>89 °C),[10] is the result of added energetic contributions from burial of hydrophobic surfaces, van der Waals packing and charge–charge interactions of CLZ residues that form critical interhelical contacts. hnRNP C tetramers retain structural integrity under conditions of 2 M NaCl, 2% deoxycholate, pH between 5.5 and 11.5, EDTA containing buffers, β-mercaptoethanol, or at 95 °C in 2 M guanidinium,[2] consistent with CLZ sequences containing the primary sequence determinants of hnRNP C oligomerization and being buried in the core of the tetrameric complex that is observed as four-lobed spherical or slightly ellipsoidal particles in electron micrographs.[59]

High-affinity hnRNP C–RNA binding interactions only occur when protein oligomerization mediates the association of multiple RNA-binding elements. hnRNP C constructs of lower order stoichiometry exhibit reduced RNA-binding affinity both *in vivo* and *in vitro*.[9,10,12] Furthermore, only tetrameric hnRNP C constructs contain an *in vitro* occluded nucleotide binding site size that is large enough to correlate with ~230 nucleotide lengths of RNA shown to be bound to hnRNP C tetramers *in vivo*.[10,41] Because the sequence elements within both the RRM and bZLM domains that are implicated in direct protein–RNA interactions are all N-terminal in sequence to CLZ, an antiparallel orientation of the helices at the oligomerization interface spaces RNA-binding domains at distal ends of the coiled coil, allowing long lengths of RNA to wrap peripherally around the complex and contact each RNA-binding surface. In this respect, the occluded nucleotide binding site sizes of mutant hnRNP C oligomer constructs that differ with respect to stoichiometry and orientation of monomers relative to each other[10] support an antiparallel assembly of hnRNP C monomers in the native tetrameric complex. Symmetric spacing of binding surfaces in the hnRNP C tetramer is also consistent with the primary function of hnRNP C as a packaging protein for nascent pre-mRNA transcripts; preferential binding to specific nucleotide sequences is necessarily precluded because sequences repeated at regular intervals that correspond to lengths appropriate for symmetrical

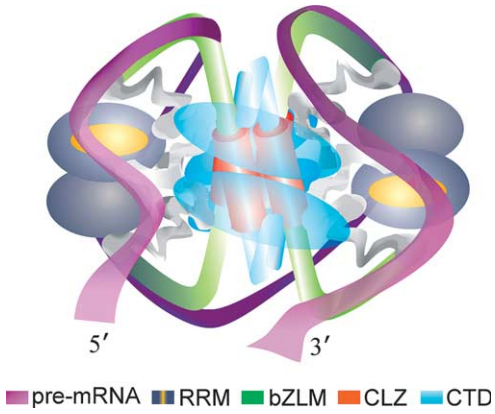

Figure 9. Model of the domain architecture of hnRNP C tetramer in complex with RNA. Residues 8–87 correspond to the RNA recognition motif (RRM), 140–179 correspond to the high-affinity RNA binding basic region zipper-like motif (bZLM), 180–202 correspond to the leucine zipper-like oligomerization domain (CLZ) and 208–290 correspond to the acidic C-terminal domain (CTD). RRM consensus binding motifs are colored orange. Residues 88–139 (including the region corresponding to the C1/C2 alternative splice site) are colored gray.

spacing around an hnRNP C tetramer are not found in the nuclear pool of pre-mRNA.[41]

The model for hnRNP C tertiary domain organization and quaternary tetramer assembly presented (Figure 9) is designed to be consistent with experimentally determined functions of specific primary and secondary structure elements. We have proposed a model of this complex previously;[6] however, our refined model, in particular, has a more symmetric set of interactions between RNA and hnRNP C and among various hnRNP C domains. Because hnRNP C-packaged RNA is susceptible to nuclease activity, is required for 19 S complex formation and interacts with associated hnRNP monoparticle proteins,[1] pre-mRNA must be peripherally wrapped around hnRNP C tetramers. Primary sequence determinants of hnRNP C for protein–nucleotide and protein–protein interactions must also be solvent-exposed, while elements that form interdomain contacts must be buried. In this respect, although CLZ domain residues are responsible for formation of the primary oligomerization interface in hnRNP C tetramers, residues located in other domains that have been shown to influence the integrity of the tetrameric complex must be oriented appropriately, the simplest mechanism of which is by juxtaposition of these elements against the CLZ core.

Early cross-linking data indicated that *in vivo*, hnRNP C tetramers exist as heterotetramers that are each composed of three C1 and one C2 protein.[2] If true, then some mechanism to ensure heterotetramer assembly must exist. The additional 13 residue splice site insertion does not appear to significantly influence tetramer assembly, since both C1 and C2 proteins expressed in bacterial cells form stable homotetramers.[41] However, it is possible that heterotetramers are slightly more stable than homotetramers due to packing of the additional amino acids into the quaternary fold.

Favorable electrostatic attractions of charges along the length of oligomerization domain interfaces primarily direct the antiparallel domain organization of hnRNP C tetramers. However, the gross separation of charge along CLZ helices is extended in the context of full-length hnRNP C by the basic region that is N-terminal to CLZ and the acidic C-terminal domain. Because deletion of C-terminal domain (CTD) residues 241–290 results in a dimeric complex with subunits oriented in parallel,[10] but inclusion of residues 1–263 restores antiparallel tetramer assembly,[6] sequence elements in the CTD may play important roles in maintaining structural integrity and conformational specificity of the hnRNP C complex. The mechanism by which this would occur is unknown, however, it is possible that the stability of the native oligomer is strengthened through interactions between charged elements in CTD and CLZ domains. Therefore, the electrostatic charge distribution at surfaces of the coiled coil exterior may potentially direct interdomain interactions in the context of full-length globular hnRNP C tetramers.

In order to examine possible interdomain interaction interfaces between the CTD and CLZ coiled coil tetramer, sequences of predicted secondary structural elements[60,61] in the CTD were modeled onto the atomic coordinates for corresponding structural elements in the C-terminal region of protein phosphatase 2C (PP2C, residues 298–368)[62] (PDB accession no. 1A6Q). The fold in this region of PP2C is described by an orthogonal array of helices and was predicted to be structurally similar to hnRNP C residues 184–290 by a protein domain homology search.[63] Based on this homology model, CTD residues K216 and K219 pack proximal to E199 and E202 in the CLZ domain of the same hnRNP C monomer. In addition, D245, D246, E249, E250, D256 and D257 lie within or proximal to CTD helical regions that span the basic ridge of residues across CLZ α1–α2-type interfaces formed by K184, K185 and K193. Possible interdomain interactions between these CTD and CLZ residues are supported by site-specific mutagenesis studies where substitution of either repeating Lys or Ala residues at E249-E250 and D256-D257 disrupted tetramer assembly.[6] Therefore, interaction of CTD elements from both ends of the antiparallel coiled coil with N and C-terminal regions of two adjacent CLZ helices stabilizes the antiparallel tetramer through charge–charge interactions around the CLZ coiled coil.

In the quaternary structure of hnRNP C tetramers, serine residues in the CTD that are phosphorylated to modulate RNA-binding interactions with hnRNP C[13–15] must be accessible to kinases and phosphatases but proximal to elements that stabilize protein–RNA complexes. Also, monomeric constructs containing both bZLM and RRM

domains have higher RNA-binding affinity and larger occluded binding site sizes than shorter N-terminal constructs.[10] Therefore, the RNA-binding sequence elements located within the bZLM[10,11,64,65] and RRM domains[9] must be located on the surface of the globular tetrameric complex in order to be accessible for interaction with nucleotide bases and spatially oriented such that an extended RNA binding interface is formed across each monomer surface. Elements of the RRM that are potentially involved in mediation of 19 S complex formation[11] must also be positioned at the exterior of each hnRNP C tetramer.

Finally, *in vivo*, each hnRNP C tetramer packages ~230 nucleotide lengths of pre-mRNA,[5,41] however a spherical complex having a diameter of ~90 Å[59] can only wrap a maximum of 125 nucleotides around its latitudinal midpoint. Therefore, nascent transcripts must effectively circumscribe the complex twice, forming contacts with two antiparallel subunits for each wrap around the tetramer; the simplest and most efficient way to curve RNA around a spherical complex and maintain the four inflection points that are necessarily required to wrap twice about the tetramer is with a baseball seam topology, which divides the complex into two parts with equal areas. In this manner, pre-mRNA transcripts contact each binding surface with the same polarity. In addition, because RNA wraps around hnRNP C tetramers so that initial contacts and exits are on the same side of the tetramer, assembly of RNA-bound hnRNP C tetramers into a 19 S triangular complex may be facilitated.

Materials and Methods

High performance liquid chromatography (HPLC) solvents were from Burdick and Jackson (Muskegon, MI). Trifluoroacetic acid (TFA) (HPLC grade), BS^3 and EDC were from Pierce (Rockford, IL). All buffers, detergents and salts were from Sigma (St. Louis, MO) or Fisher Scientific (Hampton, NH). NMR solvents and $^{13}C/^{15}N$ labeled compounds were from CIL (Andover, MA). All chemicals were ACS reagent grade or better.

Protein expression and purification

DNA encoding the CLZ sequence (I180–E207) was subcloned into the pET-32(LIC) vector (Novagen, Madison, WI) and transformed into BL21(DE3) *Escherichia coli* cells. The expressed protein construct included N-terminal fusion sequences of Trx and a His_6 purification tag that were removed by recombinant enterokinase (Novagen) digestion after initial purification steps. For ^{15}N and $^{15}N/^{13}C$-labeled samples, bacterial cultures were grown in M9 minimal media with $^{15}NH_4Cl$ as nitrogen and [^{13}C]glucose as carbon sources and supplemented with 50 μg/ml ampicillin (Sigma). Cultures reaching an absorbance at 600 nm of about 1.0 were induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (Fisher Scientific) and grown for four additional hours. Cells were collected by centrifugation and lyzed with freeze-thaw and sonication in the presence of 1% octyl-β-D-thioglucopyranoside, 5 mM imidazole, 500 mM NaCl, 20 mM Tris–HCl (pH 7.9).

Bacterial cell lysates were applied to a metal-chelating Sepharose column (Pharmacia, Piscataway, NJ) charged with Ni^{2+}. Fusion proteins that bound to the column following washes with 500 mM NaCl, 30 mM imidazole, 20 mM Tris (pH 7.9) were eluted with 100 mM EDTA, 500 mM NaCl, 20 mM Tris (pH 7.9), dialyzed against H_2O (pH ~8) and lyophilized to dryness. Optimal cleavage of the N-terminal fusion sequences was achieved in the presence of 2 M urea, 50 mM NaCl, 20 mM Tris (pH 7.7) over a period of 72 hours at room temperature. CLZ was isolated from fusion sequences by additional Ni^{2+} affinity chromatography in the presence of 2 M urea. Final purification was performed with reversed-phase HPLC using a C8 column and a water/acetonitrile gradient elution from 0–90% acetonitrile, 0.1% TFA. The final product was vacuum concentrated, frozen, lyophilized and redissolved in H_2O or ^2H_2O (pH 6.0).

Gel filtration experiments

Gel filtration experiments were performed using a Superose 12 10/30 column (Pharmacia) equilibrated with 50 mM KH_2PO_4, 150 mM NaCl (pH 7.0). Eluted proteins were detected by monitoring absorbance at 214 nm. Column volume (V_t) was 23.56 ml, and void volume (V_o) 8.52 ml as measured by dextran 2000 (2 MDa) migration. A standard chromatographic mobility curve was generated using catalase (four subunits, M_r ~226 kDa), aldolase (four subunits, M_r ~158 kDa), phosphorylase B (single subunit, M_r ~97 kDa), ovalbumin (M_r ~43 kDa) and thioredoxin (M_r ~14.8 kDa). The partition coefficient (K_{av}) for each protein was calculated by $K_{av} = (V_e - V_o)/(V_t - V_o)$, where V_e was the experimentally derived elution volume of individual species. In gel filtration experiments for CLZ and Trx-CLZ oligomers, samples of each 1 ml fraction that exhibited significant absorbance at 214 nm were analyzed by SDS-PAGE performed under reducing conditions. CLZ or Trx-CLZ was present only in the major peak of each elution profile corresponding to the oligomeric state.

Cross-linking experiments

Samples containing either 250 μM CLZ prepared in 50 mM KH_2PO_4, 150 mM NaCl (pH 7.0), 50 μM Trx-CLZ or N-terminal fusion sequences from the Trx-CLZ construct prepared in 20 mM *N*-(2-hydroxyethyl)piperazine-*N*′−2-ethanesulfonic acid (Hepes) buffer (pH 7.4) were used in the cross-linking studies. Cross-linking was initiated by addition of 0–15 mM EDC to the CLZ samples and reactions proceeded for one hour at room temperature before quenching with 50 mM hydroxylamine (final concentration). Experiments with the fusion protein used 0–5 mM BS^3 for Trx-CLZ or 0–10 mM BS^3 for Trx in Hepes buffer (pH 7.4). These reactions proceeded for 30 minutes at room temperature and were quenched by addition of 200 mM glycine (pH 7.4) for five minutes at room temperature. Samples were denatured in electrophoresis sample buffer (62.5 mM Tris (pH 6.8), 2% (w/v) SDS, 10% (v/v) glycerol, 50 mM dithiothreitol, 0.001% (w/v) bromphenol blue) at 95 °C for five minutes. SDS-PAGE was performed using either a 16% (w/v) Tricine-SDS gel or a 4–12% gradient gel (Gibco/Invitrogen, Carlsbad, CA). Proteins were visualized by staining the gel with Coomassie Blue-R250 or Colloidal Coomassie Blue G-250 (Invitrogen).

NMR experiments

Spectra used for assignments and structure determination were recorded at 27 °C, pH 6.0, 2 mM monomer protein concentration on a 600 MHz Avance spectrometer (14.098 T) (Bruker, Billerica, MA) or a 800 MHz Avance spectrometer (18.78 T) equipped with triple resonance gradient probes at Vanderbilt University Biomolecular NMR facility. Pulsed field gradient methods were used to suppress water signal.[66] Spectra were processed with XWINNMR software (Bruker) and analyzed with XEASY software.[67] ^{1}H chemical shifts were referenced directly to internal 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS); ^{15}N and ^{13}C references were determined with the appropriate frequency ratios.[68]

Backbone sequential resonance assignments were determined and verified using ^{1}H–^{15}N HSQC,[47] ^{1}H–^{13}C HSQC,[69] 3D CBCANH,[70] 3D CBCA(CO)NH,[71] 3D HNCA,[72] 3D HNCO,[73] 2D ^{1}H–^{1}H NOE spectroscopy (NOESY)[74] and 3D ^{15}N-edited NOESY-HSQC[75] experiments. Side-chain assignments were provided by 2D HCCH-TOCSY[76] in combination with 2D ^{1}H–^{13}C HSQC experiments and completed with 3D HCCH-COSY[77] and 3D ^{13}C-edited NOESY-HSQC[78,79] experiments. All NOESY experiments employed mixing times of 50 ms and 100 ms. The acquired spectra collectively allowed assignments of 93% of all nitrogen, carbon and hydrogen nuclei. Internuclear distances were derived from integral intensities of cross-peaks in natural abundance 2D ^{1}H–^{1}H NOESY, 3D ^{15}N-edited NOESY-HSQC and 3D ^{13}C-edited NOESY-HSQC spectra. Quantitative ^{3}J-coupling experiments[80] provided χ_1 torsion angle restraints for I180, I183, I190, T188 and V194 and the χ_2 angle of L197. ^{15}N{^{1}H} steady-state NOE experiments[81] were used to examine backbone dynamics.

Structure calculation and analysis

Structure calculations of monomeric and tetrameric CLZ were performed with CYANA[82] using standard corrections and modifications to account for nuclei that were not stereospecifically assigned. Although the traditionally defined upper-bound distance for observable NOE cross-relaxations is 5 Å,[74] the intensity of a few cross-peaks in NOE spectra were enhanced largely as a result of spin diffusion between closely packed methyl groups of aliphatic side-chains of core *a* and *d* position residues. In these cases (four distinct cross-peaks), 1 Å or less was added to upper-bounds distances in order to account for cross-peaks between protons having internuclear distances farther than 5 Å. For example, transfer of magnetization from L187 H^N through the spatially close L197 H^{δ} methyl group (~2 Å) to L197 H^{β} protons (~4.5 Å) allows cross-peak to be observed, even though the internuclear distance between L187 H^N and L197 H^{β} nuclei is outside the traditionally defined range for observable dipolar cross-relaxations. The data set used in torsion angle dynamics structure calculations contained calibrated interatomic upper distance bounds and ϕ, ψ, χ_1 and χ_2 dihedral angle restraints derived from analysis of chemical shift indices[45] and ^{3}J-coupling data. Based on chemical shift index analysis[45] and α-helical cross-peak patterns in 2D ^{1}H–^{1}H NOE spectra, hydrogen bonding distance restraints between $H^N(i+4)$ and $C'O(i)$ atoms for Q181–E202 were also included. Experimentally, residues L187 through L198 showed the slowest rates of hydrogen amide exchange (logarithmic protection factors on the order of 2 at 25 °C, pH 6), supporting the existence of hydrogen bond interactions.

Initial determination of CLZ helix conformation employed short and medium-range constraints that defined intrahelical distances, exclusive of any potentially ambiguous restraints arising from NOEs having both intra- and interhelical contributions. Calculation of CLZ monomer structures was performed in an iterative process, with evaluation of consistently violated internuclear distance and torsion angle constraints for correctness of assignment and successive inclusion of initially ambiguous restraints subsequently determined through generated helix structures to contain intramonomer contributions. Approximate contributions (% values) to observed cross-peak intensity were determined by calculating the ratio of cross-peak intensity (proportional to $1/r^6$) that would result from internuclear distances (r) for intramonomer, α1–α2 and α1–α4-type interactions. The final set of determined intramonomer restraints generated an ensemble of 20 CLZ monomer conformers that was well defined and had a low average target energy function (0.1 Å^2). For calculation of CLZ tetramer structure, four monomers were joined in sequence with 22 volumeless linker residues intervening between each monomer to allow exploration of all helix orientations and crossing angles in the tetramer. Interhelical distances defined by long-range constraints were examined on the basis of an antiparallel tetramer model in order to define the class of interhelical interactions for each constraint. α1–α2-type distance restraints included α1–α2, α2–α1, α3–α4 and α4–α3 distances; α1–α4-type distance restraints included α1–α4, α4–α1, α2–α3 and α3–α2 distances. The number of interhelical constraints was maximized through iterative torsion angle dynamics minimization and evaluation of initially ambiguous distance constraints to obtain the final set of upper bounds distances used in structural calculations of CLZ tetramers.

Initial calculations permitted generous conformational freedom (30°) for backbone dihedral angles within the range of angles allowed for α-helices. In order to require the experimentally evidenced symmetry of the tetrameric complex in distance geometry calculations, backbone and side-chain dihedral angles of each unique α-helical residue were analyzed in 40 monomers composing the ten lowest energy tetramer structures of 100 calculated. The circular variance of angles (here used synonymously with the angular order parameter) was determined using in-house scripts similar to those found in PROCHECK.[83] For those torsion angles exhibiting significant order ($S>0.8$), average angles were imposed in successive calculations with reduced conformational freedom (down to ±1°). The final torsion angle dynamics calculation generated 400 structures. Constrained energy minimization using the same distance restraints was performed for the 50 models with the lowest target energy function over 2000 steps with the generalized Born solvent model[84,85] in the SANDER module of AMBER (version 8).[86] Pairwise RMSDs for 20 energy-minimized structures were calculated by superimposition in Insight II (Accelrys, Inc., San Diego, CA); these structures were then superimposed on the structure having the lowest pairwise RMSD and averaged to produce the mean structure. Reported values are RMSDs from this average structure.

Visualization and analyses of structure were performed with Insight II, MOLMOL,[87] PROCHECK[83] and GRASP.[88] Specifically, calculation of helix axis separation, interhelical angles and buried surface area was performed in

MOLMOL, calculation of RMSD and analysis of dihedral angles for specific residues were performed with PROCHECK, and calculation of electrostatic potential at the molecular surface of CLZ tetramers was performed with GRASP. Coiled coil knobs-into-holes packing interactions were analyzed with SOCKET[53] using a 7.4 Å cutoff distance for interacting side-chains. Atomic coordinates were fit to the equation describing ideal coiled coil geometry using a script written by Mark Sales† to determine supercoil parameters. Helical offsets were measured by the displacement between C^{α} atoms of interdigitating *a* and *d* position residues along the superhelix axis. Molecular modeling of the CLZ and CTD domains was performed using standard tools available in Swiss-PdbViewer version 3.7.

Protein Data Bank accession code

Structural data, including the atomic coordinates of the 20 lowest energy conformers, NMR constraints and resonance assignments have been deposited with the RCSB Protein Data Bank (accession no. 1TXP).

Acknowledgements

We gratefully acknowledge Dr Kristin Whitson for assistance with cross-linking experiments as well as critical reading of the manuscript and Dr Jaison Jacob for assistance with quantitative ^{3}J coupling experiments. This work was supported by NIH Grant GM48567 to W.M.L.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.05.002

References

1. McAfee, J. G., Huang, M., Soltaninassab, S., Rech, J. E., Iyengar, S. & LeStourgeon, W. M. (1997). The packaging of pre-mRNA. In *Eukaryotic mRNA Processing—Frontiers in Molecular Biology Series* (Krainer, A. R., ed.), pp. 68–102, Oxford University Press, New York.
2. Barnett, S. F., Friedman, D. L. & LeStourgeon, W. M. (1989). The C proteins of HeLa 40S nuclear ribonucleoprotein particles exist as anisotropic tetramers of (C1)3 C2. *Mol. Cell Biol.* **9**, 492–498.
3. Merrill, B. M., Barnett, S. F., LeStourgeon, W. M. & Williams, K. R. (1989). Primary structure differences between proteins C1 and C2 of HeLa 40S nuclear ribonucleoprotein particles. *Nucl. Acids Res.* **17**, 8441–8449.
4. Conway, G., Wooley, J., Bibring, T. & LeStourgeon, W. M. (1988). Ribonucleoproteins package 700 nucleotides of pre-mRNA into a repeating array of regular particles. *Mol. Cell Biol.* **8**, 2884–2895.
5. Huang, M., Rech, J. E., Northington, S. J., Flicker, P. F., Mayeda, A., Krainer, A. R. & LeStourgeon, W. M. (1994). The C-protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. *Mol. Cell Biol.* **14**, 518–533.
6. Shahied, L., Braswell, E. H., LeStourgeon, W. M. & Krezel, A. M. (2001). An antiparallel four-helix bundle orients the high-affinity RNA binding sites in hnRNP C: a mechanism for RNA chaperonin activity. *J. Mol. Biol.* **305**, 817–828.
7. Gorlach, M., Wittekind, M., Beckman, R. A., Mueller, L. & Dreyfuss, G. (1992). Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J.* **11**, 3289–3295.
8. Gorlach, M., Burd, C. G. & Dreyfuss, G. (1994). The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J. Biol. Chem.* **269**, 23074–23078.
9. Wan, L., Kim, J. K., Pollard, V. W. & Dreyfuss, G. (2001). Mutational definition of RNA-binding and protein–protein interaction domains of heterogeneous nuclear RNP C1. *J. Biol. Chem.* **276**, 7681–7688.
10. McAfee, J. G., Shahied-Milam, L., Soltaninassab, S. R. & LeStourgeon, W. M. (1996). A major determinant of hnRNP C protein binding to RNA is a novel bZIP-like RNA binding domain. *RNA*, **2**, 1139–1152.
11. Shahied-Milam, L., Soltaninassab, S. R., Iyer, G. V. & LeStourgeon, W. M. (1998). The heterogeneous nuclear ribonucleoprotein C protein tetramer binds U1, U2, and U6 snRNAs through its high affinity RNA binding domain (the bZIP-like motif). *J. Biol. Chem.* **273**, 21359–21367.
12. Tan, J. H., Kajiwara, Y., Shahied, L., Li, J., McAfee, J. G. & LeStourgeon, W. M. (2001). The bZIP-like motif of hnRNP C directs the nuclear accumulation of pre-mRNA and lethality in yeast. *J. Mol. Biol.* **305**, 829–838.
13. Stone, J. R., Maki, J. L. & Collins, T. (2003). Basal and hydrogen peroxide stimulated sites of phosphorylation in heterogeneous nuclear ribonucleoprotein C1/C2. *Biochemistry*, **42**, 1301–1308.
14. Fung, P. A., Labrecque, R. & Pederson, T. (1997). RNA-dependent phosphorylation of a nuclear RNA binding protein. *Proc. Natl Acad. Sci. USA*, **94**, 1064–1068.
15. Mayrand, S. H., Dwen, P. & Pederson, T. (1993). Serine/threonine phosphorylation regulates binding of C hnRNP proteins to pre-mRNA. *Proc. Natl Acad. Sci. USA*, **90**, 7764–7768.
16. Burkhard, P., Stetefeld, J. & Strelkov, S. V. (2001). Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* **11**, 82–88.
17. Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375–382.
18. Kohn, W. D., Mant, C. T. & Hodges, R. S. (1997). Alpha-helical protein assembly motifs. *J. Biol. Chem.* **272**, 2583–2586.
19. McLachlan, A. D. & Stewart, M. (1975). Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J. Mol. Biol.* **98**, 293–304.
20. Lee, D. L., Lavigne, P. & Hodges, R. S. (2001). Are trigger sequences essential in the folding of two-stranded alpha-helical coiled-coils? *J. Mol. Biol.* **306**, 539–553.
21. O'Neil, K. T. & DeGrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, **250**, 646–651.
22. Huyghues-Despointes, B. M. & Baldwin, R. L. (1997).

† http://ucxray.berkeley.edu/~mark/fitcc.html

Ion-pair and charged hydrogen-bond interactions between histidine and aspartate in a peptide helix. *Biochemistry*, **36**, 1965–1970.
23. Smith, J. S. & Scholtz, J. M. (1998). Energetics of polar side-chain interactions in helical peptides: salt effects on ion pairs and hydrogen bonds. *Biochemistry*, **37**, 33–40.
24. Burkhard, P., Kammerer, R. A., Steinmetz, M. O., Bourenkov, G. P. & Aebi, U. (2000). The coiled-coil trigger site of the rod domain of cortexillin I unveils a distinct network of interhelical and intrahelical salt bridges. *Struct. Fold. Des.* **8**, 223–230.
25. Su, J. Y., Hodges, R. S. & Kay, C. M. (1994). Effect of chain length on the formation and stability of synthetic alpha-helical coiled coils. *Biochemistry*, **33**, 15501–15510.
26. Lumb, K. J., Carr, C. M. & Kim, P. S. (1994). Subdomain folding of the coiled coil leucine zipper from the bZIP transcriptional activator GCN4. *Biochemistry*, **33**, 7361–7367.
27. Lau, S. Y., Taneja, A. K. & Hodges, R. S. (1984). Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *J. Biol. Chem.* **259**, 13253–13261.
28. Fairman, R., Chao, H. G., Mueller, L., Lavoie, T. B., Shen, L., Novotny, J. & Matsueda, G. R. (1995). Characterization of a new four-chain coiled-coil: influence of chain length on stability. *Protein Sci.* **4**, 1457–1469.
29. Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S. & Richardson, D. C. (1995). The Alacoil: a very tight, antiparallel coiled-coil of helices. *Protein Sci.* **4**, 2252–2260.
30. Monera, O. D., Zhou, N. E., Lavigne, P., Kay, C. M. & Hodges, R. S. (1996). Formation of parallel and antiparallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. *J. Biol. Chem.* **271**, 3995–4001.
31. Lovejoy, B., Choe, S., Cascio, D., McRorie, D. K., DeGrado, W. F. & Eisenberg, D. (1993). Crystal structure of a synthetic triple-stranded alpha-helical bundle. *Science*, **259**, 1288–1293.
32. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, **262**, 1401–1407.
33. DeLano, W. L. & Brunger, A. T. (1994). Helix packing in proteins: prediction and energetic analysis of dimeric, trimeric, and tetrameric GCN4 coiled coil structures. *Proteins*, **20**, 105–123.
34. Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T. & Hodges, R. S. (2000). Effects of side-chain characteristics on stability and oligomerization state of a *de novo*-designed model coiled-coil: 20 amino acid substitutions in position "d". *J. Mol. Biol.* **300**, 377–402.
35. Wagschal, K., Tripet, B., Lavigne, P., Mant, C. & Hodges, R. S. (1999). The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins. *Protein Sci.* **8**, 2312–2329.
36. Solan, A., Ratia, K. & Fairman, R. (2002). Exploring the role of alanine in the structure of the Lac repressor tetramerization domain, a ferritin-like Alacoil. *J. Mol. Biol.* **317**, 601–612.
37. Potekhin, S. A., Medvedkin, V. N., Kashparov, I. A. & Venyaminov, S. (1994). Synthesis and properties of the peptide corresponding to the mutant form of the leucine zipper of the transcriptional activator GCN4 from yeast. *Protein Eng.* **7**, 1097–1101.
38. Alberti, S., Oehler, S., von Wilcken-Bergmann, B. & Muller-Hill, B. (1993). Genetic analysis of the leucine heptad repeats of Lac repressor: evidence for a 4-helical bundle. *EMBO J.* **12**, 3227–3236.
39. Monera, O. D., Kay, C. M. & Hodges, R. S. (1994). Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. *Biochemistry*, **33**, 3862–3871.
40. Sheridan, R. P., Levy, R. M. & Salemme, F. R. (1982). Alpha-helix dipole model and electrostatic stabilization of 4-alpha-helical proteins. *Proc. Natl Acad. Sci. USA*, **79**, 4545–4549.
41. McAfee, J. G., Soltaninassab, S. R., Lindsay, M. E. & LeStourgeon, W. M. (1996). Proteins C1 and C2 of heterogeneous nuclear ribonucleoprotein complexes bind RNA in a highly cooperative fashion: support for their contiguous deposition on pre-mRNA during transcription. *Biochemistry*, **35**, 1212–1222.
42. Cavanagh, J., Fairbrother, W. J., Palmer, A. G. & Skelton, N. J. (1996). *Protein NMR spectroscopy*, Academic Press, San Diego pp. 16–19.
43. O'Donaghue, S. & Nilges, M. (1999). Calculation of symmetric oligomer structures from NMR data. In *Biological Magnetic Resonance: Structure, Computation and Dynamics in Protein NMR* (Krishna, M. & Berliner, L., eds), Vol. 17, pp. 131–161, Kluwer Academic/Plenam Publishers, New York.
44. Staros, J. V. (1982). N-hydroxysulfosuccinimide active esters: bis(N-hydroxysulfosuccinimide) esters of two dicarboxylic acids are hydrophilic, membrane-impermeant, protein cross-linkers. *Biochemistry*, **21**, 3950–3955.
45. Wishart, D. & Sykes, B. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239**, 363–392.
46. Tripet, B. & Hodges, R. S. (2001). STABLECOIL: An algorithm designed to predict the location and relative stability of coiled-coils in native protein sequences. In *Peptides: The Wave of the Future, Proceedings of the Second International/Seventeenth American Peptide Symposium* (Lebl, M. & Houghten, R., eds), pp. 365–366, Kluwer Academic Publishers, New York.
47. Bodenhausen, G. & Ruben, D. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters*, **69**, 185–189.
48. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al*. (2000). The Protein Data Bank. *Nucl. Acids Res*, **28**, 235–242.
49. Friedman, A. M., Fischmann, T. O. & Steitz, T. A. (1995). Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science*, **268**, 1721–1727.
50. Banner, D. W., Kokkinidis, M. & Tsernoglou, D. (1987). Structure of the ColE1 rop protein at 1.7 A resolution. *J. Mol. Biol.* **196**, 657–675.
51. Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature*, **395**, 347–353.
52. Crick, F. H. C. (1953). The packing of α-helices: simple coiled coils. *Acta Crystallog.* **6**, 689–697.
53. Walshaw, J. & Woolfson, D. N. (2001). Socket: a

program for identifying and analysing coiled-coil motifs within protein structures. *J. Mol. Biol.* **307**, 1427–1450.
54. Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198.
55. Krylov, D., Mikhailenko, I. & Vinson, C. (1994). A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. *EMBO J.* **13**, 2849–2861.
56. Eilers, M., Patel, A. B., Liu, W. & Smith, S. O. (2002). Comparison of helix interactions in membrane and soluble alpha-bundle proteins. *Biophys. J.* **82**, 2720–2736.
57. Kohn, W. D., Kay, C. M. & Hodges, R. S. (1998). Orientation, positional, additivity, and oligomerization-state effects of interhelical ion pairs in alpha-helical coiled-coils. *J. Mol. Biol.* **283**, 993–1012.
58. Pace, C. N. (2001). Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry*, **40**, 310–313.
59. Rech, J. E., LeStourgeon, W. M. & Flicker, P. F. (1995). Ultrastructural morphology of the hnRNP C protein tetramer. *J. Struct. Biol.* **114**, 77–83.
60. Kneller, D. G., Cohen, F. E. & Langridge, R. (1990). Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.* **214**, 171–182.
61. Garnier, J., Gibrat, J.-F. & Robson, B. (1996). GOR secondary structure prediction method version IV. *Methods Enzymol.* **266**, 540–553.
62. Das, A. K., Helps, N. R., Cohen, P. T. & Barford, D. (1996). Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 A resolution. *EMBO J.* **15**, 6798–6809.
63. Corpet, F., Gouzy, J. & Kahn, D. (1998). The ProDom database of protein domain families. *Nucl. Acids Res.* **26**, 323–326.
64. Nakielny, S. & Dreyfuss, G. (1996). The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals. *J. Cell Biol.* **134**, 1365–1373.
65. Siomi, H. & Dreyfuss, G. (1997). RNA-binding proteins as regulators of gene expression. *Curr. Opin. Genet. Dev.* **7**, 345–353.
66. Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
67. Bartels, C., Xia, T., Billeter, M., Güntert, P. & Wüthrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR*, **5**, 1–10.
68. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D. *et al*. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids: IUPAC-IUBMB-IUPAB inter-union task group on the standardization of databases of protein and nucleic acid structures determined by NMR spectroscopy. *J. Biomol. NMR*, **12**, 1–23.
69. Bendall, M., Pegg, D., Doddrell, D. & Field, J. (1983). Inverse DEPT sequence. Polarization transfer from a spin ½ nucleus to n spin ½ heteronuclei *via* correlated motion in the doubly rotating reference frame. *J. Magn. Reson.* **51**, 520–526.
70. Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293.
71. Grzesiek, S. & Bax, A. (1992). An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**, 201–207.
72. Bax, A. & Ikura, M. (1991). An efficient 3D NMR technique for correlating the proton and 15N backbone resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *J. Biomol. NMR*, **1**, 99–104.
73. Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.* **34**, 93–158.
74. Kumar, A., Ernst, R. & 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.
75. Marion, D., Kay, L., Sparks, S., Torchia, D. & Bax, A. (1989). Three-dimensional heteronuclear NMR of 15N labeled proteins. *J. Am. Chem. Soc.* **111**, 1515–1517.
76. Bax, A., Clore, G. & Gronenborn, A. (1990). 1H–1H correlation *via* isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C enriched proteins. *J. Magn. Reson.* **88**, 425–431.
77. Kay, L., Xu, G., Singer, A., Muhandiram, D. & Forman-Kay, J. (1993). A gradient-enhanced HCCH-TOCSY experiment for recording side-chain 1H and 13C correlations in H2O samples of proteins. *J. Magn. Reson. B*, **101**, 333–337.
78. Zuiderweg, E., Petros, A., Fesik, S. & Olejniczak, E. (1991). Four-dimensional [13C, 1H, 13C, 1H] HMQC-NOE-HMQC NMR spectroscopy: resolving tertiary NOE distance constraints in the spectra of larger proteins. *J. Am. Chem. Soc.* **113**, 370–372.
79. Vuister, G., Clore, G., Gronenborn, A., Powers, R., Garrett, D., Tschudin, R. & Bax, A. (1993). Increased resolution and improved spectral quality in four-dimensional 13C/13C-separated HMQC-NOESY-HMQC spectra using pulsed field gradients. *J. Magn. Reson.* **101**, 210–213.
80. Vuister, G., Wang, A. & Bax, A. (1993). Measurement of three-bond nitrogen-carbon J couplings in proteins uniformly enriched in 15N and 13C. *J. Am. Chem. Soc.* **115**, 5334–5335.
81. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G. *et al*. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
82. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
83. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
84. Bashford, D. & Case, D. A. (2000). Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**, 129–152.
85. Xia, B., Tsui, V., Case, D. A., Dyson, H. J. & Wright,

P. E. (2002). Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J. Biomol. NMR*, **22**, 317–331.
86. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. R., Cheatham, T. E., III, DeBolt, S. *et al*. (1995). AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Phys. Commun.* **91**, 1–41.
87. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 29–32 see also pp. 51–5.
88. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
89. Swanson, M. S., Nakagawa, T. Y., LeVan, K. & Dreyfuss, G. (1987). Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Mol. Cell Biol.* **7**, 1731–1739.
90. Jiang, W., Guo, X. & Bhavanandan, V. P. (1998). Four distinct regions in the auxiliary domain of heterogeneous nuclear ribonucleoprotein C-related proteins. *Biochim. Biophys. Acta*, **1399**, 229–233.
91. Williamson, D. J., Banik-Maiti, S., DeGregori, J. & Ruley, H. E. (2000). hnRNP C is required for postimplantation mouse development but is dispensable for cell viability. *Mol. Cell Biol.* **20**, 4094–4105.
92. Preugschat, F. & Wold, B. (1988). Isolation and characterization of a Xenopus laevis C protein cDNA: structure and expression of a heterogeneous nuclear ribonucleoprotein core protein. *Proc. Natl Acad. Sci. USA*, **85**, 9669–9673.
93. Michaud, E. J., Bultman, S. J., Stubbs, L. J. & Woychik, R. P. (1993). The embryonic lethality of homozygous lethal yellow mice (Ay/Ay) is associated with the disruption of a novel RNA-binding protein. *Genes Dev.* **7**, 1203–1213.

Edited by A. G. Palmer III

(Received 13 March 2005; received in revised form 28 April 2005; accepted 3 May 2005)