Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs

Rui-Ming Xu[1]*, Lana Jokhan[1], Xiaodong Cheng[1], Akila Mayeda[2] and Adrian R Krainer[2]

Background: Heterogeneous nuclear ribonucleoprotein (hnRNP) A1 is one of the most abundant core proteins of hnRNP complexes in metazoan nuclei. It behaves as a global regulator of alternative pre-mRNA splicing by antagonizing the activities of several serine/arginine-rich splicing factors (SR proteins), resulting in the activation of distal alternative 5′ splice sites and skipping of optional exons. Purified hnRNP A1 has nucleic acid annealing activity. The protein also shuttles continuously between the nucleus and the cytoplasm, a process mediated by signals within its C-terminal glycine-rich domain. The N-terminal region of human hnRNP A1, termed unwinding protein 1 (UP1), contains two RNA-recognition motifs (RRMs), RRM1 and RRM2. Understanding the structural elements by which hnRNP A1 interacts with RNA will have broad implications for studies of RNA processing.

Results: The crystal structure of UP1 has been determined to 1.9Å resolution. Each RRM independently adopts the characteristic RRM fold, consisting of a four-stranded antiparallel β-pleated sheet and two α helices packed on one side of the β sheet. The two RRMs are antiparallel and held in close contact, mainly by two Arg–Asp ion pairs. As a result, the two four-stranded β sheets are brought together to form an extended RNA-binding surface. A segment of the linker connecting the two RRMs is flexible in the absence of bound RNA, but the general location of the linker suggests that it can make direct contacts with RNA. Comparison with other RRM structures indicates that a short 3_{10} helix, found immediately N-terminal to the first β strand in RRM1, may interact with RNA directly.

Conclusions: The RRM is one of the most common and best characterized RNA-binding motifs. In certain cases, one RRM is sufficient for sequence-specific and high affinity RNA binding; but in other cases, synergy between several RRMs within a single protein is required. This study shows how two RRMs are organized in a single polypeptide. The two independently folded RRMs in UP1 are held together in a fixed geometry, enabling the two RRMs to function as a single entity in binding RNA, and so explaining the synergy between the RRMs. The UP1 structure also suggests that residues which lie outside of the RRMs can make potentially important interactions with RNA.

Introduction

Nascent transcripts synthesized by RNA polymerase II are bound by a group of nuclear proteins, known as heterogeneous nuclear ribonucleoproteins (hnRNPs) [1]. During its entire nuclear lifetime, pre-mRNA exists in this bound form. The hnRNP complex contains more than twenty proteins, and the six most abundant ones, A1, A2, B1, B2, C1 and C2, are known as the 'core' proteins. The proteins of the hnRNP complex are involved in diverse aspects of pre-mRNA metabolism.

One of the most studied core hnRNP proteins is hnRNP A1; it has been shown to catalyze the base-pairing of complementary single-stranded polynucleotides [2–5]. The RNA-annealing activity of hnRNP A1 may directly facilitate base-pairing interactions between snRNA and pre-mRNA. Furthermore, hnRNP A1 can interact stably with U2 and U4 small nuclear ribonucleoprotein particles (snRNPs) *in vitro* [6]. The interaction of hnRNP A1 with U2, together with the association between hnRNP A1 and certain 3′ splice sites *in vitro* [7], suggested a potential role for hnRNP A1 in influencing the binding of U2 snRNA to the branch site in the early stages of spliceosome assembly [6].

Most mammalian pre-mRNAs contain multiple introns, that allow different mRNA molecules to be generated from the same pre-mRNA by alternative splicing. Numerous genes have been identified that utilize alternative

splicing to regulate their expression during development or differentiation. Both *in vitro* and *in vivo*, hnRNP A1 has been shown to influence alternative 5′ splice-site selection. It antagonizes the activity of splicing factors of SR (serine/arginine-rich) protein family, such as SF2/ASF and SC35, leading to the activation of distal 5′ splice sites and promoting alternative exon skipping [8–12].

Although hnRNP A1 is predominantly a nuclear protein, it has also been shown to shuttle continuously between the nucleus and the cytoplasm [13]. A 38-amino acid domain near the C terminus, which has been shown to be necessary and sufficient for the nuclear localization of hnRNP A1 [14,15], is also an export signal [16] and is therefore responsible for this shuttling activity. Combined with the observation that hnRNP A1 binds to poly(A)⁺ RNA in both the nucleus and the cytoplasm, shuttling suggests that hnRNP A1 may be involved in transporting mature mRNA through the nuclear pores to the cytoplasm.

Human hnRNP A1 consists of a single polypeptide chain of 320 amino acids. An N-terminal proteolytic fragment spanning the first 196 amino acids, known as unwinding protein 1 (UP1) [17,18], contains two RNA-recognition motifs (RRM, also known as RNA-binding domain, RBD; and ribonucleoprotein consensus sequence, RNP-CS) separated by a short linker. The C-terminal region of hnRNP A1 is particularly rich in glycine residues and includes several Arg–Gly–Gly (RGG) tri-peptide repeats that also constitute an RNA-binding motif [19]. The RRM is an ancient and extremely common RNA-binding motif, within which the RNP-2 hexamer and RNP-1 octamer submotifs are highly conserved (reviewed in [20]). The tertiary structures of several isolated RRMs have been determined — the crystal and NMR structures of the first of two RRMs in the U1 snRNP A polypeptide (U1A) [21,22], its co-crystal structure with a bound hairpin RNA [23], its NMR structure as a complex with the 3′ untranslated region (UTR) of its own mRNA [24], the NMR structure of the RRM of hnRNP C1/C2 with and without bound RNA [25,26] and the NMR structure of the second RRM of the *Drosophila* Sex-lethal protein [27].

Early studies showed that although both the N-terminal UP1 domain and the C-terminal glycine-rich domain of hnRNP A1 are capable of binding to single-stranded nucleic acids, the intact protein exhibits more stable and highly cooperative binding, with cooperativity resulting from self-association between the C-terminal domains of separate protein molecules [28–31]. Both domains are required for the alternative-splicing activity of hnRNP A1, and mutations in conserved surface residues of either RRM abolish activity, although general RNA binding by the purified protein is not substantially affected [32]. Maximal RNA-binding and annealing activities require the intact protein, although the UP1 and C-terminal domains each

show reduced, but detectable, activity in these assays [3,4, 30,32,33]. Although hnRNP A1 has general RNA-binding properties and binds to most pre-mRNA transcripts in the nucleus, purified hnRNP A1 binds with higher affinity to short RNAs containing one or more copies of the motif UAGGGA/U ([34]; I Watakabe, A Hanamura and ARK, unpublished results). The physiological significance of this high-affinity binding remains controversial [35].

We report the crystal structure of UP1 at 1.9 Å resolution, as determined by the method of multiple isomorphous replacement with anomalous diffraction (MIRAS) from two mercurial derivatives. The crystal structure shows that the folding of both RRMs closely resembles that of the previously determined individual RRM structures. The two independently folded RRMs are held rigidly by two Arg–Asp salt bridges, and are positioned with the C-terminal α helices of each RRM adjacent and antiparallel to each other, such that the two four-stranded β sheets form an extended RNA-binding surface. The inter-RRM linker loop is located on the surface between the two β sheets, and its position suggests that it is capable of contacting bound RNA directly. A short 3_{10} helix immediately N-terminal to the first β strand of the first RRM occupies a spatial position corresponding to regions involved in critical RNA contacts in other RRM-containing proteins.

Results
Overall structure
The UP1 structure reported here consists of 170 amino acids (Pro7–Arg92 and Gly99–Ser182) out of 196 total residues in the expressed UP1 fragment. Three regions are not modeled in the structure due to poor electron density. These regions include the first six residues at the N terminus, the last 14 residues at the C terminus and six internal residues (Glu93–Pro98). These poorly defined segments of UP1 are presumably flexible and therefore disordered in the crystal lattice; the terminal residues are not absent from the purified recombinant protein, because mass spectrometry analysis is consistent with the presence of all 196 residues (data not shown).

The structure reveals that UP1 folds into two stable, independently folded, subdomains — subdomain I, which comprises residues 7–92, and subdomain II, corresponding to residues 99–182 (Fig. 1). Within subdomains I and II reside the highly conserved RRM1 (residues 15–89) and RRM2 (residues 106–180), respectively. Both RRM1 and RRM2 adopt the characteristic RRM fold, as previously determined from the single RRM structures of several RNA-binding proteins using both crystallographic and NMR methods [21,22,25,27]. The secondary structure and folding topology of the RRM1 of hnRNP A1 have also been previously determined by NMR methods [36]. The RRM fold consists of two α helices and four β strands in the order β_1-α_A-β_2-β_3-α_B-β_4 from the N terminus to the

Figure 1

Overall folding of UP1. **(a)** Ribbon diagram of the UP1 structure, viewed from the front (β-sheet side) of the molecule. The conserved RNP-2 and RNP-1 submotifs are colored cyan and purple, respectively. Red indicates the disordered portion of the linker region; its placement is not based on electron density, but was included for clarity. The N terminus starts at Pro7, and the C terminus ends at Ser182. **(b)** Side view of the overall folding of UP1. Diagrams were generated with the Ribbons program [67].
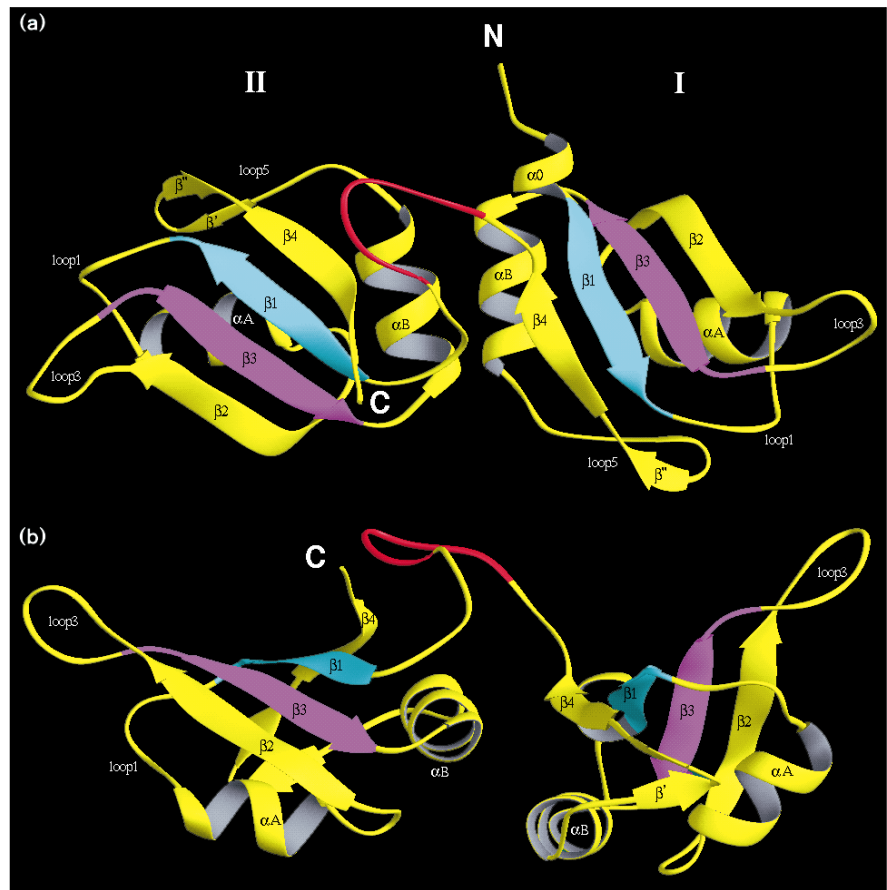

C terminus, following the nomenclature previously used for the U1A RRM [21] (Fig. 2a). The loops between the α helices and the β strands are designated sequentially from 1 to 5, and we use a roman numeral prefix to distinguish structural elements from the two RRMs. The four conserved β strands form an antiparallel β sheet (Fig. 1), with the spatial order β_4-β_1-β_3-β_2. The highly conserved RNP-2 hexamer and RNP-1 octamer sequences are located on the two central strands, β_1 and β_3, respectively. In each RRM, the two α helices are located on the same side of the β sheet, where they lie nearly perpendicular to each other. The RRM fold is stabilized by extensive hydrophobic interactions between the β sheet and the α helices. In addition to the above common, conserved secondary structure elements, features found in UP1 include a short 3_{10} helix, designated α_0, immediately preceding RRM1, and in both RRMs, two very short β strands, β′ and β″, at the tip of the segment that is usually designated as loop 5. This segment of each RRM therefore adopts a common β-turn-β conformation. An analogous β-turn-β feature has been described in the hnRNP C structure [25] and can also be seen in the high-resolution structure of the U1A–RNA complex [23].

RRM1 and RRM2 are found in opposite orientations in the UP1 structure and they can be related by an approximate twofold symmetry around an axis located between I-α_B and II-α_B and perpendicular to the plane shown in Figure 1a. Helices I-α_B and II-α_B are adjacent in space and oriented in antiparallel fashion. The two β sheets form an extended surface with eight β strands, with the four α helices all located on the same side of this surface. In addition to the covalent joining of RRM1 and RRM2 through the connecting loop, which is partially disordered in the crystal structure, major contacts between the two RRMs are mediated by charged residues located in, or near, I-α_B, I-β_4 and II-α_B.

Comparison of RRM structures

RRM1 and RRM2 of hnRNP A1 have a high degree of sequence homology (Fig. 2a), with 35% residue identity, 59% similarity and no gaps. This high degree of conservation is reflected in the three-dimensional structure of UP1. The entire Cα chains of RRM1 (Lys15–Ala89) and RRM2 (Lys106–Ala180) can be superimposed with a root mean square (rms) deviation of 1.49Å (Fig. 2b). The most pronounced conformational differences between the two

Figure 2
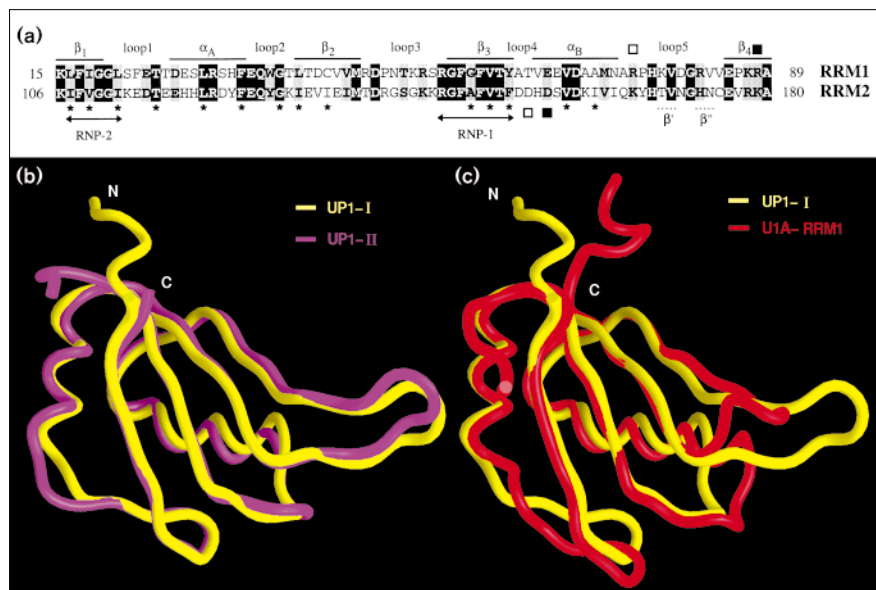
Comparison of RRM structures.
(a) Structure-based sequence alignment of the two hnRNP A1 RRMs, RRM1 and RRM2.

Identical amino acids are shown as white letters on a black background, and similar ones are shown as black letters on a gray background. The secondary structure is indicated at the top, except for β′ and β″ of both RRMs, which are shown at the bottom. The RNP-2 and RNP-1 submotifs are shown by double-headed horizontal arrows. The asterisks indicate hydrophobic core residues conserved among RRMs [20]. Open and filled boxes mark the two pairs of aspartate and arginine residues involved in inter-RRM salt bridges. (b) Superposition of the Cα chains of the two subdomains of UP1. Subdomain I (which includes RRM1) is shown in yellow and subdomain II (which includes RRM2) in magenta. The figure was generated with the program GRASP [68]. (c) Superposition of the subdomain I Cα backbone (yellow) with that of the U1A N-terminal RRM (red). The rms deviation is 1.0 Å, calculated for 48 Cα atoms corresponding to all the residues located in β_1–β_4, α_A, α_B, loop 2 and loop 4 in RRM1 of UP1. When the same set of Cα atoms was used for least-squares alignment of RRM1 with RRM2 in UP1, the resulting rms deviation was 0.97 Å. The U1A coordinates are from the U1A–RNA complex crystal structure [23].

RRMs occur in two regions. The first region is at the C-terminal end of the α_B helices. This difference is induced by Pro76 in RRM1, which displaces the Cα atom of Arg75 by more than 3.5 Å away from the corresponding position, Lys166, in RRM2. The second region of difference is the tip of loop 3 that is tilted and twisted in RRM1, with respect to that in RRM2. A proline residue (Pro49) is located at the tip of this loop in RRM1. However, the difference in orientation between the two loops may not be entirely due to the presence of Pro49, because the loop 3 regions of both RRMs are involved in crystal packing, which may alter the conformation of the loops.

The previously determined high resolution crystal structure of the N-terminal U1A RRM bound to stem-loop II of U1 snRNA ([23]; PDB code 1urn) is particularly informative with regard to RRM–RNA interactions, and it is used here for comparison with the UP1 RRMs. The Cα chain of the U1A RRM can be superimposed extremely well with those of UP1 RRM1 and RRM2 (Fig. 2c). Within the RRM fold, three regions of major differences between the U1A RRM and the two UP1 RRMs were found. Firstly, helix α_A of U1A is one turn longer at its N terminus than the corresponding helices in UP1. Secondly, the orientation of loop 5 is appreciably different in U1A than in the UP1 RRMs. This difference is perhaps attributable to the insertion of a single residue, Met72, at the end of α_B of U1A and to the presence of two proline residues, Pro76 and Pro81, in loop 5 of U1A. Thirdly, loop 3 is shorter in U1A than in the UP1 RRMs; the latter loops are in an extended conformation,

whereas the U1A loop 3 has the features of a turn. This particular loop in U1A is involved in extensive interactions with RNA, and its conformation changes upon RNA binding [23]. Loop 3 is also the least conserved region among different RRMs (reviewed in [20]).

Interactions between RRM1 and RRM2
Direct interactions between the two RRMs of UP1 are primarily mediated by two pairs of arginine–aspartic acid salt bridges. Arg75 forms a salt bridge with Asp155, whereas Arg88 interacts with Asp157 via a charge interaction and a hydrogen bond (Fig. 3). In RRM1, Arg75 and Arg88 are located at the beginning of loop 5 and at the end of β_4, respectively. In RRM2, Asp155 and Asp157 are located in loop 4 and at the beginning of α_B, respectively. Interestingly, Arg75 is immediately followed by a proline residue, which induces a sharp turn, whereas the residue homologous to Arg75 in RRM2, Lys166, is followed by a tyrosine residue. This proline-induced sharp turn is among the most pronounced differences between RRM1 and RRM2 (Fig. 2c). The proline residue is conserved among all known hnRNP A and B proteins, but it is not found in the RRMs of unrelated proteins. This feature may have been selected during evolution to enable Arg75 to form a salt bridge with Asp155, because an analogous interaction could not be modeled when RRM1 was substituted with RRM2 (data not shown).

In addition to these direct ionic interactions, two well ordered water molecules (W17 and W48, with temperature

factors of 10Å^2 and 11Å^2, respectively) were found to provide additional stability for the packing of the two RRMs. W17 forms hydrogen bonds with the mainchain oxygen of Pro86, the guanidino moiety of Arg75 and the Nε2 of the indole ring of His156. W48 also makes three hydrogen bonds with the backbone amide of Arg88, Oδ1 of Asp157 and Oδ2 of Asp160. Both Asp157 and Asp160 are within the α$_B$ helix in RRM2. Additional hydrogen-bonding networks through successive water molecules are present; however, they are less stable than those mediated by W17 and W48, as indicated by the higher thermal factors of the water molecules involved, and therefore are less significant in mediating RRM1–RRM2 interactions. Hydrophobic interactions appear not to play a major role in mediating the inter-RRM contacts. Only one hydrophobic residue, Met72, is present within the interface between the two subdomains, in addition to another isolated pair (Leu13 and Ile164) located at the outer edge of the interface. Although the distance between these two hydrophobic residues is indicative of a weak interaction, the presence of hydrophobic residues at these two positions is a phylogenetically conserved and unique feature of proteins closely related to hnRNP A1.

RNA-binding mechanisms

Although the present structure does not include bound RNA, a great deal can be inferred about the UP1–RNA interactions from the UP1 structure, in combination with previous findings from biochemical studies of hnRNP A1 and UP1 and structural studies of single RRM–RNA interactions in U1A and hnRNP C. It is known that the RRM β sheet provides a general RNA-binding platform, whereas the determinants of RNA-binding specificity lie outside the conserved structural elements. In the case of the single RRM of hnRNP C, as characterized by NMR, short stretches of amino acids at both the N and C termini of the RRM, as well as residues located on the β sheet, exhibit large perturbations in chemical shifts upon ligand binding [26], which underline the importance of these regions in RNA-binding. In the structures of the complexes of the U1A N-terminal RRM with a 21-nucleotide RNA hairpin and with the 3′ UTR of its own mRNA [23,24], three regions were shown to be involved in extensive interactions with RNA. These are the four-stranded β sheet, loop 3 and a C-terminal helix, α$_C$. The most noticeable β sheet surface residues involved in RNA-binding are two conserved aromatic residues, one in the RNP-1 submotif and the other in the RNP-2 submotif. The aromatic rings stack with single-stranded bases of the bound RNA. Residues located in loop 3 are key determinants of the U1A RNA-binding specificity. In addition, α$_C$ undergoes a large conformational change upon RNA binding and makes critical contacts with the RNA. Two amino acids, Glu19 and Lys22, from the neighboring loop 1 also make important RNA contacts.

Figure 3
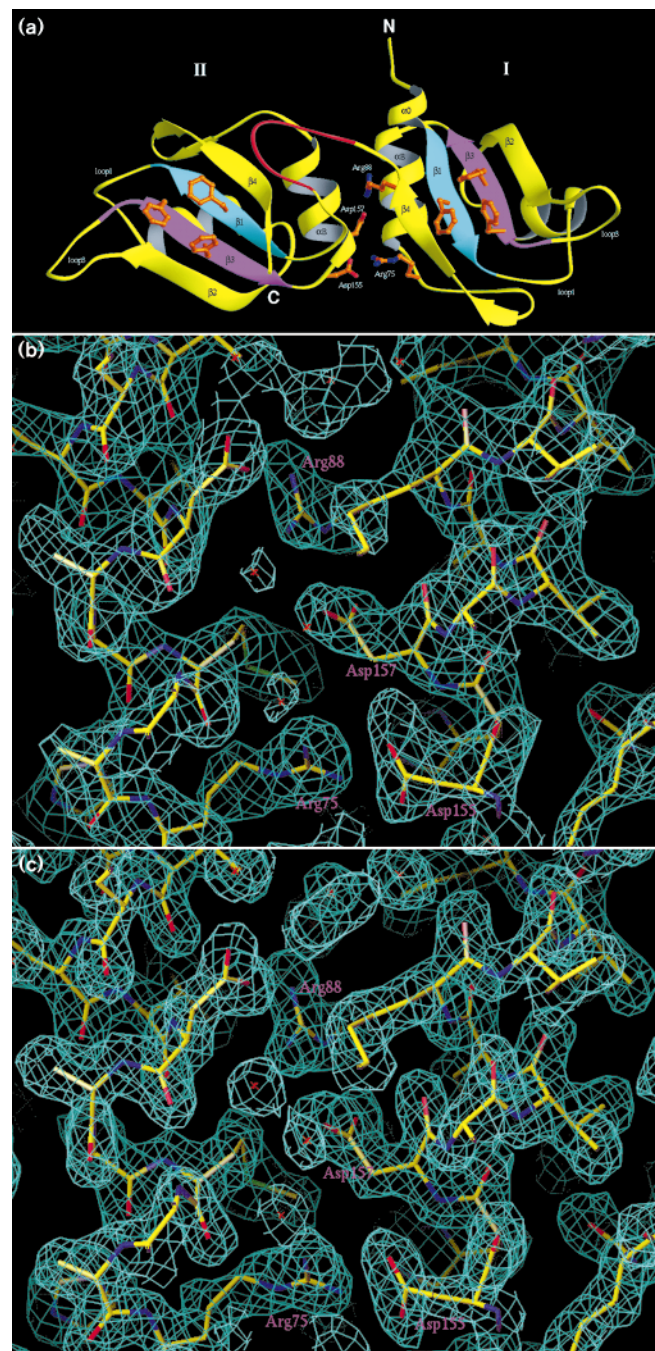
Interactions between the UP1 subdomains. **(a)** Ribbon diagram showing the inter-RRM salt bridges; the conserved solvent-exposed phenylalanines that are involved in RNA binding are also shown. Arginine, aspartate and phenylalanine residues are shown as ball-and-stick models, with carbon atoms shown in orange, nitrogen in blue and oxygen in red. **(b)** MIRAS-phased electron-density map of the region covering the two salt bridges. The refined model is shown in stick representation; red crosses indicate water molecules. The electron-density map is contoured at 1σ level. **(c)** 2F$_o$–F$_c$ electron-density map of the same region as in (b), as calculated from the refined model. The map is calculated using reflections from 6Å to 1.9Å and contoured at 1σ level.

Each RRM in UP1 can bind RNA separately, albeit weakly [33]. The common RRM fold suggests that the general RNA-binding mechanism through the four-stranded β-sheet has been evolutionarily conserved. Indeed, UV-crosslinking and mutational studies have pointed to the involvement of the conserved phenylalanines in RNA binding. Phe17 and Phe59 in RRM1 and the corresponding pair in RRM2, Phe108 and Phe150, have been shown to UV-crosslink to ^{32}P-labeled p[dT]$_8$ [28]. These aromatic residues are essential components of the conserved RNP submotifs. In the co-crystal structure of the U1A RRM with bound RNA, the homologous residues, Tyr13 (corresponding to Phe17 and Phe108 in UP1) and Phe56 (corresponding to Phe59 and Phe108 in UP1), base stack with a cytosine and an adenine, respectively. These base-stacking interactions may be an evolutionarily conserved feature of RRM–RNA interactions. The equivalent position of a third pair of phenylalanines in UP1, Phe57 and Phe148, is occupied by Gln54 in U1A. Gln54 stabilizes three key residues involved in RNA binding by hydrogen bonding to the hydroxyl group of Tyr13 and the mainchain oxygens of Lys50 and Arg52 [23]. Double-substitution mutations of hnRNP A1, in which either Phe57 and Phe59 of RRM1 or Phe148 and Phe150 of RRM2 are replaced by a pair of aspartic acids, result in a complete loss of the alternative splicing activity of hnRNP A1, although general RNA binding is not impaired [32]. In the UP1 structure, the three phenylalanines in the RNP submotifs of each RRM cluster together (Fig. 3a) with approximately equal pairwise distances. The high degree of sequence and structural homology between the two RRMs suggests that the RNA-binding mechanisms for each of them should be similar. However, the two RRMs appear to function non-equivalently in RNA binding. The asymmetric roles of RRM1 and RRM2 in RNA binding are illustrated by the selection of different high affinity RNA consensus sequences in iterative *in vitro* RNA-binding selection (SELEX) experiments using hnRNP A1 lacking either RRM1 or RRM2 [34]. The different RNA-binding properties of the two hnRNP A1 RRMs may be explained by the differences of two surface residues in β_2 and two residues in loop 3. In β_2, Asp42 and Val44 in RRM1 are replaced by Val133 and Glu135 in RRM2, respectively. At the tip of loop 3, Pro49 and Asn50 in RRM1 are replaced by Arg140 and Gly141 in RRM2, respectively. Another potentially important difference between RRM1 and RRM2 is within loop 1 — Phe23 of RRM1 is replaced by Glu114 in RRM2; the corresponding residue in U1A is also a glutamate, Glu19, which makes important contacts with RNA. These different residues may affect RNA binding to a varying degree.

Although the conserved RRM features are crucial for both general and sequence-specific RNA binding, regions outside of the RRM may also play important roles in RNA binding. In the UP1 structure, the linker sequence connecting RRM1 and RRM2 appears to be flexible, and six residues in its middle are disordered. Nevertheless, the ordered residues in the structure place the linker in a position highly likely to be involved in direct RNA interactions (see Discussion section). In addition, structural comparisons of UP1 with the U1A and hnRNP C RRMs suggest that the short N-terminal 3_{10} helix, α_0, in subdomain I could potentially play a prominent role in interacting with RNA. This suggestion is based on the observation that structural elements in hnRNP C [26] and U1A [23,24], found in similar spatial locations as α_0 in UP1, are essential for RNA binding. The sidechains of Glu11 and Gln12 in α_0 of UP1 are above the two central β strands, β_3 and β_1, respectively (on the exposed side of the β sheet) and point to the central RNP submotifs in RRM1. In fact, α_0 and the residues immediately C-terminal to the β_4 strand in subdomain I appear to form a 'wall' on one side of the β-sheet plane. This wall would be expected to guide the RNA through the plane in a direction diagonal to the β strands in subdomain I of hnRNP A1 (Fig. 1a).

Although full-length hnRNP A1 is quite basic, with a calculated isoelectric point of 10.2, most of the positive charges are derived from the C-terminal RGG-box motif. The UP1 domain has a calculated isoelectric point of 8.3, and electric charges are not evenly distributed on the protein surface. The β-sheet side of the protein surface is more positively charged than the α-helix side (Fig. 4). This pattern of charge distribution is consistent with the RNA-binding role of the β sheets.

Discussion
Structural basis of RNA-binding specificity and RRM synergy in hnRNP A1
RNA–protein interactions are undoubtedly an essential feature of the mechanism by which hnRNP A1 regulates alternative splicing and associates with nascent transcripts in the nucleus. Some of the properties of hnRNP A1 appear somewhat paradoxical. On one hand, the ability of hnRNP A1 to associate with most, if not all, nascent transcripts *in vivo* [1], to bind in an ATP-independent manner to RNA with or without splice sites, as a purified protein or in nuclear extracts [7,37,38], and to promote duplex formation of complementary single-stranded RNAs [2–5], indicates that hnRNP A1 binds to RNA indiscriminately. On the other hand, it is difficult to imagine how indiscriminate binding can result in the highly specific concentration-dependent activation of distal alternative 5′ splice sites *in vitro* and in transfected cells [9,11,12]. Indeed, hnRNP A1 mutants have been described that retain the general RNA-binding capacity, but are inactive in alternative splicing [32]. In fact, hnRNP A1 is capable of interacting with RNA in a sequence-specific manner. It has recently been demonstrated by SELEX that hnRNP A1 binds with high affinity to RNAs that share the hexamer consensus sequence UAGGGA/U [34]; this sequence bears some resemblance to portions of the vertebrate 5′ and 3′ splice site consensus sequences. Both the UP1 domain and the

Figure 4

Electrostatic potential distribution of the UP1 surface. **(a)** Front view (β-sheet side). Positively and negatively charged potential are indicated in blue and red, respectively. The surface was calculated using a probe radius of 1.4 Å, and the potential is displayed at a $-15\,K_BT$ to $+15\,K_BT$ scale, where K_B is the Boltzmann constant. **(b)** Rear view (α-helix side). Diagrams were generated with the program GRASP.
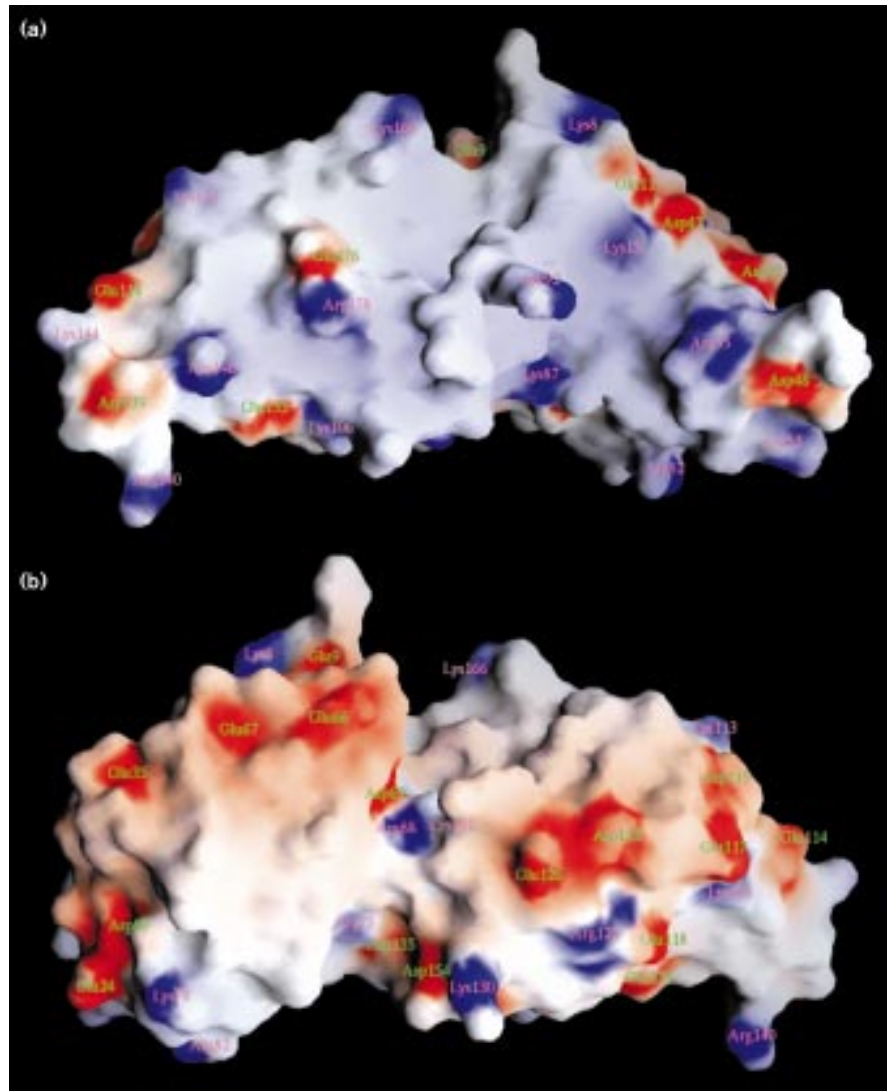
C-terminal glycine-rich domain have been shown to separately bind single-stranded DNA or RNA [30,32,33]. One view is that the RGG-box motif, or the entire C-terminal glycine-rich domain in which the RGG-box motif resides, is not a determinant of RNA-binding specificity, but its role is to increase the overall binding affinity of hnRNP A1. This view is supported by the observation that hnRNP A1[B], an alternatively spliced isoform of hnRNP A1 with a longer C-terminal glycine-rich domain, binds strongly to RNA irrespective of sequence ([32]; I Watakabe and ARK, unpublished data). Thus, the recognition of specific sequences for preferential binding by hnRNP A1 appears to be achieved by the UP1 domain, and both RRMs are required for this high affinity binding [34].

The conservation of the RRM fold is consistent with the proposal that the RRM is an ancient, evolutionarily conserved, RNA-binding motif [20]. Individual RRMs may have preferences for different RNA sequences, due to differences in surface amino acids found outside the conserved RNP submotifs, although still on the β sheet, in loop 3 and in certain cases in loop 1. These differences probably contribute to, but may not be sufficient to account for, the RNA-binding specificity of hnRNP A1. An important feature of the RNA-binding properties of hnRNP A1 is that the two RRMs somehow act in a concerted fashion to give rise to the overall RNA-binding characteristics of the wild-type protein. The UP1 structure shows that the two independently folded RRMs are in close enough contact to form a single RNA-binding unit. The two RRMs are held together by two pairs of Asp–Arg salt bridges and by water-mediated interactions. These residues involved in interactions between the two subdomains are unlikely to make direct RNA contacts; however,

they may be critical for RNA binding, as they can proba-
bly affect the relative orientation of the two β sheet RNA-
binding surfaces. These specific inter-RRM interactions
may be a significant source of the synergistic behavior dis-
played by the hnRNP A1 RRMs. The two RRMs are
specifically orientated, such that the linker connecting
them is in a position to make potentially important direct
interactions with RNA. It is likely that the disordered
portion of the linker, which is poorly defined in the UP1
crystal structure, becomes structured upon binding RNA.
The linker has previously been suggested to be important
for the RNA-binding specificity of hnRNP A1 [34].

It has been observed that RNA binding by both UP1 and
hnRNP A1 is highly sensitive to salt concentration [33,
35,39]. The salt bridges observed in the UP1 structure may
partially account for the salt sensitivity. Because the two
bridges in the UP1 structure are partially exposed to the
solvent, it may be expected that an environment of high
ionic strength could have a strong destabilizing effect on
the packing of the RRMs, by disrupting the Arg–Asp elec-
trostatic interactions. Interestingly, it has been observed
that salt concentration can influence alternative splicing *in
vitro*, with increasing ionic strength having an opposite
effect on alternative 5′ splice site utilization to that of
adding hnRNP A1 ([9] and references therein). The activa-
tion of proximal 5′ splice sites upon increasing salt concen-
tration is consistent with destabilization of hnRNP A1
binding by disruption of the inter-RRM salt bridges. The
specific residues involved in salt bridges between the two
RRMs are highly conserved among the members of the
hnRNP A/B protein family, but not in more distantly
related RRMs (S Munroe, personal communication). This
observation strongly suggests that the structural basis for
achieving a defined orientation of the two RRMs relative to
each other has been phylogenetically conserved among
members of the hnRNP A/B family of proteins. Indeed,
several members of the hnRNP A/B family have been
shown to have splicing activities closely related to those of
hnRNP A1 [32,40]. The strong phylogenetic conservation
of the charged residues, together with the observed effects
of ionic strength on hnRNP A1 binding and function,
strongly suggest that the observed salt bridges reflect the
inherent structure of the protein and are not induced by
crystal packing.

An unexpected finding from the UP1 structure is that the
residues immediately N-terminal to β_1 in RRM1 fold into
a 3_{10} helix and occupy a similar spatial position to that of
the α_C helix at the C terminus of the U1A RRM. In U1A,
α_C undergoes a large conformational change upon ligand
binding and makes direct RNA contacts. An N-terminal
region of hnRNP C in a similar spatial position also shows
direct RNA binding, as observed by NMR. Deletions of
this region in hnRNP C and of α_C in U1A strongly affect
RNA binding [41–43]. The UP1 structure suggests that

Figure 5
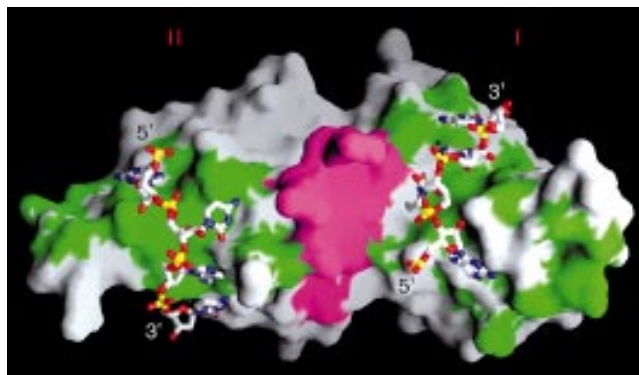
Potential RNA-binding sites on UP1. The positions corresponding to
residues that contact RNA in the U1A N-terminal RRM [23] were
mapped onto the UP1 β-sheet surface and are shown in green. The
ordered residues within the inter-RRM linker are shown in magenta. A
four-nucleotide segment of single-stranded RNA from the U1A–RNA
complex [23] is superimposed on each UP1 RRM, with the 5′ to 3′
polarity indicated. This segment of RNA contacts the conserved
RNP-2 and RNP-1 submotifs in the two central β strands of the U1A
N-terminal RRM. The shortest possible path for RNA bound to UP1,
according to this model, would involve joining the 3′ end of the RNA
segment shown at the left to the 5′ end of the RNA segment shown at
the right by a 4–5 nucleotide loop that crosses the linker region.
Nucleotide atoms are shown with carbon in white, nitrogen in blue,
oxygen in red and phosphorus in yellow.

the N-terminal 3_{10} helix in hnRNP A1 also makes impor-
tant RNA contacts. It is interesting to note that residues
that lie outside of the conserved RRMs and that are not
part of other recognizable RNA-binding motifs [44,45],
can nevertheless interact with RNA. In fact, the deter-
minant of the RNA-binding specificity of hnRNP C lies
entirely outside of the RRM, immediately C-terminal to
it [41]. In UP1, residues located at the C-terminal end
(186–196), but outside, of RRM2 have been implicated
in RNA binding on the basis of comparative binding
studies [33].

As a first step in modeling RNA binding by the UP1
domain of hnRNP A1, we have mapped the positions cor-
responding to residues in the U1A RRM that are involved
in contacts with its specific RNA ligand [23] on the surface
of each UP1 RRM (Fig. 5). One copy of the 4-nucleotide
single-stranded segment of U1 snRNA that contacts the
conserved RNP-1 and RNP-2 submotifs in the two central
β strands of the U1A RRM has also been superimposed on
the corresponding regions of each of the UP1 RRMs. It
should be noted, however, that the N-terminal RRM of
U1A is extremely specific in its binding, whereas hnRNP
A1 and many other RRM-containing proteins have con-
siderably less sequence specificity. It can be seen in
Figure 5 that the two sets of green patches, which repre-
sent amino acids whose U1A counterparts are involved in
RNA binding, are separated by the region occupied by the

inter-RRM linker (colored in magenta; six disordered residues not shown). According to this model, the conserved aromatic residues will base stack with the RNA and the negatively charged phosphate backbone would face outward. In one possible path for a single molecule of RNA bound by a UP1 monomer, the RNA traverses the linker region and binds to both RNP submotifs in each RRM. Because the 5′ to 3′ polarity of each RNA segment is assumed to be the same as in the U1A–RNA complex, it can be seen that a shorter length of RNA is required if subdomain I binds an RNA region downstream of the region bound by subdomain II rather than vice versa. With the RNA maximally stretched out, a loop of about four to five nucleotides would be required to join the 3′ end of the RNA segment bound by RRM2 to the 5′ end of the segment bound by RRM1; the total length of 12–13 nucleotides fits well with the size of the measured occluded site (14 ± 2) of UP1 bound to poly-etheno-adenosine [33]. One interesting scenario would be that the positively charged RGG motifs neutralize the phosphate backbone of the RNA, which is thus sandwiched between the extended β sheet surface and the C-terminal domain. More complex models, involving a more complicated RNA path, multiple simultaneous RNA targets, and/or protein multimerization, are of course possible. The validity of such models and the answers to many questions regarding RNA binding await the elucidation of the structure of hnRNP A1–RNA complexes.

Characterization of purified 40*S* hnRNP particles has shown that the core proteins of the hnRNP complex are bound to hnRNA as heterotetramers [37]. In particular, three hnRNP A1 monomers form a complex with one hnRNP B2 monomer; the primary sequence of hnRNP B2 remains unknown. Complexes of hnRNP that have been reconstituted *in vitro* from purified proteins and RNA show this stoichometry [37]. It has been shown that the glycine-rich C-terminal domain mediates self-association and interactions with other hnRNP proteins [46,31]. At the present time, the structural basis for these protein–protein interactions, and their effect on hnRNP A1 function, and RNA-binding affinity and specificity, are not known. Post-translational modifications of hnRNP A1 occur at its C-terminal domain; the known modifications include serine phosphorylation and arginine dimethylation [47–49]. Phosphorylation of hnRNP A1 can affect its RNA binding and annealing activities *in vitro* [47,49, 50]. Whether and how these post-translational modifications modulate hnRNP A1 activity in alternative splicing, and/or its subcellular distribution and nucleo-cytoplasmic shuttling remain to be determined.

Implications for other multi-RRM RNA-binding proteins

In proteins with two or more RRMs, a single RRM is sometimes sufficient for sequence-specific and high affinity RNA binding, as in the case of U1A. In many other cases, however, several RRMs are needed for stable and/or specific RNA binding. For example, hnRNP A1 requires two RRMs for its alternative splicing function [32]. The general splicing factor U2AF65 has three RRMs, all of which are required for sequence specific RNA binding and for constitutive splicing [51]. Poly(A)-binding protein (PABP) has four RRMs, which can act combinatorially for binding, although not all of them are required for poly(A) binding and *in vivo* function [52–54].

Cooperativity among RRMs is believed to be a general feature of proteins with multiple RRMs, as a survey of several such proteins indicated that the free energies of binding of the individual RRMs to RNA do not add up to the overall free energy of binding by the complete set of RRMs in each native protein [55]. In some cases, the linker that connects two RRMs in a protein is highly conserved in length and/or sequence among homologous proteins, such as in the hnRNP A/B family (S Munroe, personal communication) and in PABP [54]. This phylogenetic conservation suggests that the spacing between the RRMs can be critical for a proper binding geometry, and also that the linker residues may in some cases participate in sequence specific RNA contacts. But in other cases, the linker connecting the RRMs is dispensable, such as in SF2/ASF, in which deletion of the polyglycine linker has no effect on splicing activity [56], even though both RRMs are required for constitutive splicing [56,57].

In proteins in which cooperativity between RRMs is important, individual RRMs appear to function together as a single RNA-binding unit, giving rise to the RNA-binding properties characteristic of the intact protein. Thus, interactions among the RRMs may be as important as direct RNA–protein contacts, as these inter-RRM interactions are responsible for fixing the RRMs in a proper spatial arrangement for optimal RNA binding. The UP1 structure provides the first view of how two RRMs are spatially organized. The two RRMs in UP1 are held together mainly by two Arg–Asp salt bridges, forming a rigid structure that leaves the inter-RRM linker lying on the surface for potential RNA contacts. Other proteins that require two or more contiguous RRMs for RNA binding are likely to employ different mechanisms to orient the RRMs, because the residues involved in salt bridging in UP1 are not conserved in most RRMs (see alignment in [20]). However, an interesting question to be addressed in future studies is whether each pair of synergistic RRMs is always oriented in a manner similar to that seen in UP1 (i.e. with the two RRMs antiparallel and the RNA-binding surface of the two β sheets facing the same side). If so, different strategies may be employed to hold together the RRMs (e.g. hydrophobic interactions or salt bridges at slightly different locations).

Biological implications

**The nuclear protein, heterogeneous nuclear ribonucle-
oprotein A1 (hnRNP A1) plays important roles in a
number of cellular processes, including packaging of
pre-mRNA transcripts in the nucleus, control of gene
expression through regulation of alternative 5′ splice-site
selection, and promoting annealing of complementary
single-stranded nucleic acids. *In vitro* experiments have
shown that hnRNP A1 can act as an RNA chaperone
to ensure correct folding of biologically active RNAs,
such as ribozymes and tRNAs [58,59]. The hnRNP A1
protein has also been implicated in transporting mature
mRNA from the nucleus to the cytoplasm [13]. These
biological effects of hnRNP A1 principally result from
RNA–protein interactions, so the determination of the
structural basis for hnRNP A1's interactions with RNA
has broad implications for structure–function studies
of pre-mRNA processing.**

**The N-terminal region of hnRNP A1, termed the
unwinding protein 1 (UP1), contains two well character-
ized RNA-recognition motifs (RRMs); the C-terminal
region contains several RGG tri-peptide repeats that con-
stitute an additional RNA-binding motif. The hnRNA
A1 protein utilizes both motifs in interacting with RNA.
Two RRMs acting synergistically are required for activ-
ity and for stable, specific RNA binding. This require-
ment is also found in many proteins that have several
RRMs. The crystal structure of UP1 shows that the two
RRMs are held in close contact by two pairs of Arg–Asp
salt bridges, which result in the juxtaposition of the four-
stranded β sheet from each RRM to form an extended
RNA-binding surface. The position of the linker segment
connecting the two RRMs suggests that it has a potential
involvement in contacting RNA. The residues of a short
3_{10} helix, immediately preceding the first RRM, may also
participate in RNA binding, as they occupy a spatial
position analogous to regions of the U1A and hnRNP C
RRMs that are involved in critical RNA contacts with
their respective ligands.**

Materials and methods

Crystallization and data collection

The pET9c-UP1 plasmid was constructed by S Munroe and has been
described previously [32]. Recombinant human UP1 was expressed
and purified as previously described [32]. Crystals were grown in the
presence of 25% polyethylene glycol (PEG-4000) at pH 8.5. 20%
2-methyl-2,4-pentanediol (MPD) or 10–15% glycerol was added to
the crystallization buffer in anticipation of cryo-data collection [60]. The
crystals belong to space group P2$_1$ with unit-cell parameters a = 38.0 Å,
b = 43.9 Å, c = 55.8 Å, β = 94.2°.

All diffraction data were collected using a Mar imaging-plate detector
at the beamline X12C of the National Synchrotron Light Source,
Brookhaven National Laboratory. Crystals were frozen at 100°K using
an Oxford cryo-system during data collection. Native data at 1.9 Å reso-
lution were collected from two crystals at λ = 1.197 Å and λ = 1.15 Å,
respectively. Two mercurial derivative data sets at 2.5 Å resolution were
obtained from two different crystals, one of which was soaked in

p-chloromercuribenzene sulphonate (PCMBS) and the other in methyl
mercuric chloride (MeHgCl). The PCMBS data set was collected at
λ = 1.0 Å, whereas the MeHgCl data set was collected at λ = 1.0078 Å
to maximize the anomalous differences [60]. Data reduction statistics
for the native and the derivative crystals are shown in Table 1. All data
reduction were carried out using the HKL program suite [61].

Structure determination and refinement

Two mercury sites were identified by isomorphous and anomalous
difference Patterson maps for both the PCMBS and MeHgCl deriva-
tives. The spatial positions of the metal sites are very similar in the
two derivatives. However, the relative occupancy of the sites appears
to be different. Both derivative data sets were used for phasing, using
the PHASES suite of programs [62]. The initial 3 Å multiple isomor-
phous replacement with anomalous scattering (MIRAS) phases were
improved by solvent flattening [63], with a solvent content of 40%. A
new round of heavy-atom refinement was then performed with the

Table 1

Statistics from the crystallographic analysis of UP1.

	Native	PCMBS*	MeHgCl
Resolution (Å)	1.9	2.5	2.5
Observed reflections	123 655	24 320	27 627
Unique reflections	14 637	6295	6749
Completeness (%)	99.7	96.3	98.4
R$_{merge}$ (%)[†]	6.2	8.6	5.9
Phasing			
Hg sites		2	2
Phasing power at 3.0 Å[‡]			
isomorphous		2.21	1.97
anomalous		2.34	2.89
Overall figure of merit	0.71		
R$_{cullis}$[§]		0.573	0.557
R$_{Kraut}$[#]			
isomorphous		0.167	0.156
anomalous		0.175	0.176
Refinement			
Resolution range (Å)	6.0–1.9		
R factor (R$_{free}$)[¶]	19.7 (25.4)		
Reflections with |F| > 2σ	14 080		
Number of protein atoms	1366		
Number of water molecules	130		
Rms deviation			
bond lengths (Å)	0.012		
bond angles (°)	1.29		
dihedrals (°)	25.23		
improper angles (°)	1.06		

UP1 crystallizes in the space group P2$_1$, with cell parameters
a = 38.0 Å, b = 43.9 Å, c = 55.8 Å, β = 94.2°. *PCMBS is
p-chloromercuribenzene sulphonate. [†]R$_{merge}$ = Σ|I − < I >|/Σ< I >, where
I is the measured intensity and < I > is the averaged intensity of
multiple measurements of the same reflection. The summation is over all
the observed reflections. [‡]Phasing power = rms (< F$_H$ > / E), where F$_H$ is
the calculated structure factor of the heavy atoms and E is the residual
lack of closure. [§]R$_{cullis}$ = Σ||F$_{PH}$ − F$_P$| − F$_H$(calc)|/Σ|F$_{PH}$ − F$_P$|, where F$_{PH}$
and F$_P$ denote observed derivative and native crystal structure factors,
respectively, and F$_H$ denotes the calculated heavy atom structure factor.
[#]R$_{Kraut}$ = Σ||F$_{PH}$| − |F$_{PH}$(calc)||/Σ|F$_{PH}$| for isomorphous data, and
R$_{Kraut}$ = Σ(||F$_{PH+}$| − |F$_{PH+}$(calc)|| + ||F$_{PH−}$| − |F$_{PH−}$(calc)||)/Σ(|F$_{PH+}$| + |F$_{PH−}$|)
for anomalous data. [¶]R factor = Σ||F$_o$| − |F$_c$||/Σ|F$_o$|, where F$_o$ denotes
the observed structure factor amplitude and F$_c$ denotes the structure
factor calculated from the model. 10% of reflections were used to
calculate R$_{free}$.

Figure 6
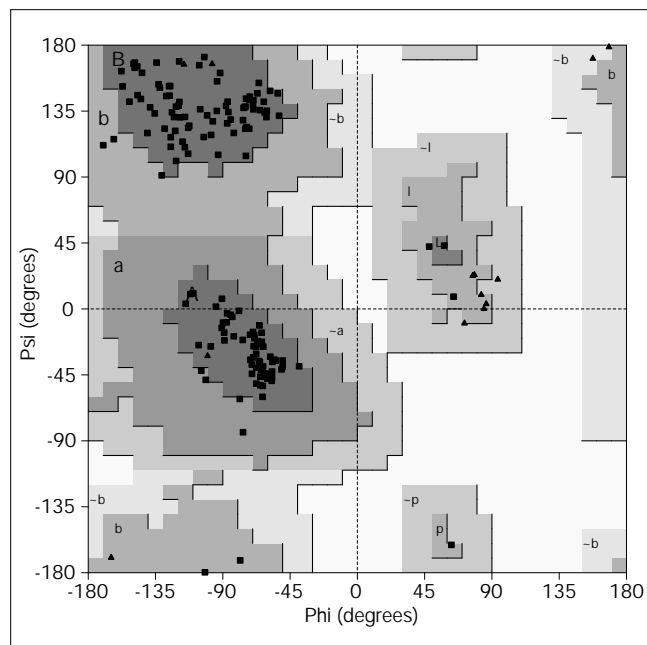
Ramachandran plot of the refined UP1 model. Glycine residues are shown by triangles and non-glycine residues by squares. Darkest grey shading represents the most favoured regions. The plot was generated with the program PROCHECK [66].

solvent-flattened phases held fixed as references. This procedure was repeated three times, and the resulting solvent-flattened phases were used to generate a skeleton representation of the electron-density map for chain tracing. The final phasing statistics are shown in Table 1.

Chain tracing and model building were performed using the graphics program O [64]. The electron density is of excellent quality (Fig. 3b), and nearly all the residues present in the final structure were built into the electron-density map in the first round. The resulting model was then refined against the diffraction data in the resolution shell 8.0–3.0 Å, using X-PLOR programs [65]. An initial round of positional refinement followed by simulated annealing, from a temperature of 3000°K to 300°K, reduced the crystallographic R factor from 45% to 29.1%. Multiple rounds of model rebuilding and refinement were carried out before refining the model against higher resolution data of 2.5 Å, 2.0 Å and finally 1.9 Å. The placement of ordered water molecules and the refinement of the temperature factors were done after the model had been satisfactorily refined at 2.0 Å resolution. During the X-PLOR refinements, the R_{free} value was monitored using 10% of the data. Refinement statistics are shown in Table 1. The refined model has good stereochemistry as assessed by the PROCHECK program [66]. The Ramachandran plot of the mainchain parameters (Fig. 6) shows 91.3% of the non-glycine, non-proline, residues within the most favored region and 8.7% in the additional allowed regions.

Accession numbers
The atomic coordinates of the refined UP1 model have been deposited with the Brookhaven Protein Data Bank, with the code 1up1. The coordinates can also be obtained by e-mail from RMX at xur@cshl.org.

References
1. Dreyfuss, G., Matunis, M.J., Piñol-Roma, S. & Burd, C.G. (1993). hnRNP proteins and the biogenesis of mRNA. *Ann. Rev. Biochem.* **62**, 289–321.
2. Kumar, A. & Wilson, S.H. (1990). Studies of the strand-annealing activity of mammalian hnRNP complex protein A1. *Biochemistry* **29**, 10717–10722.
3. Pontius, B.W. & Berg, P. (1990). Renaturation of complementary DNA strands mediated by purified mammalian heterogeneous nuclear ribonucleoprotein A1 protein: implications for a mechanism for rapid molecular assembly. *Proc. Natl. Acad. Sci. USA* **87**, 8403–8407.
4. Munroe, S.H. & Dong, X. (1992). Heterogeneous nuclear ribonucleoprotein A1 catalyzes RNA–RNA annealing. *Proc. Natl. Acad. Sci. USA* **89**, 895–899.
5. Portman, D.S. & Dreyfuss, G. (1994). RNA annealing activities in HeLa nuclei. *EMBO J.* **13**, 213–221.
6. Buvoli, M., Cobianchi, F. & Riva, S. (1992). Interaction of hnRNP A1 with snRNPs and pre-mRNAs: evidence for a possible role of A1 RNA annealing activity in the first steps of spliceosome assembly. *Nucleic Acids Res.* **20**, 5017–5025.
7. Swanson, M.S. & Dreyfuss, G. (1988). RNA binding specificity of hnRNP proteins: a subset bind to the 3′ end of introns. *EMBO J.* **7**, 3519–3529.
8. Fu, X.-D., Mayeda, A., Maniatis, T. & Krainer, A.R. (1992). General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5′ and 3′ splice site selection. *Proc. Natl. Acad. Sci. USA* **89**, 11224–11228.
9. Mayeda, A. & Krainer, A.R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**, 365–375.
10. Mayeda, A., Helfman, D.M. & Krainer, A.R. (1993). Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol. Cell. Biol.* **13**, 2993–3001.
11. Cáceres, J.F., Stamm, S., Helfman, D. & Krainer, A.R. (1994). Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Nature* **265**, 1706–1709.
12. Yang, X., Bani, M.R., Lu, S.J., Rowan, S., Ben-David, Y. & Chabot, B. (1994). The A1 and A1ᴮ proteins of heterogeneous nuclear ribonucleoparticles modulate 5′ splice site selection *in vivo. Proc. Natl. Acad. Sci. USA* **91**, 6924–6928.
13. Piñol-Roma, S. and Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355**, 730–732.
14. Siomi, H. & Dreyfuss, G. (1995). A nuclear localization domain in the hnRNP A1 protein. *J. Cell Biol.* **129**, 551–560.
15. Weighardt, F., Biamonti, G. & Riva, S. (1995). Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. *J. Cell Sci.* **108**, 545–555.
16. Michael, W.M., Choi, M. & Dreyfuss, G. (1995). A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* **83**, 415–422.
17. Herrick, G. & Alberts, B. (1976). Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. *J. Biol. Chem.* **251**, 2124–2132.
18. Williams, K.R., Stone, K.L., LoPresti, M.B., Merrill, B.M. & Planck, S.R. (1985). Amino acid sequence of the UP1 calf thymus helix-destabilizing protein and its homology to an analogous protein from mouse myeloma. *Proc. Natl. Acad. Sci. USA* **82**, 5666–5670.
19 Kiledjian, M. & Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* **11**, 2655–2664.
20. Birney, E., Kumar, S. & Krainer, A.R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**, 5803–5816.
21. Nagai, K., Oubridge, C., Jessen, T.H., Li, J. & Evans, P.R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **346**, 515–520.

22. Hoffman, D.W., Query, C.C., Golden, B.L., White, S.W. & Keene, J.D. (1991). RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc. Natl. Acad. Sci. USA* **88**, 2495–2499.

23. Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438.

24. Allain, F.H.-T., Gubser, C.C., Howe, P.W.A., Nagai, K., Neuhaus, D. & Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* **380**, 646–650.

25. Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G. & Mueller, L. (1992). ^{1}H, ^{13}C, and ^{15}N NMR assignments and global folding pattern of the RNA binding domain of the human hnRNP C proteins. *Biochemistry* **31**, 6254–6265.

26. Görlach, 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.

27. Lee, A.L., Kanaar, R., Rio, D.C. & Wemmer, D.E. (1994). Resonance assignments and solution structure of the second RNA-binding domain of Sex-lethal determined by multidimensional heteronuclear magnetic resonance. *Biochemistry* **33**, 13775–13786.

28. Merrill, B.M., Stone, K.L., Cobianchi, F., Wilson, S.H. & Williams, K.R. (1988). Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. *J. Biol. Chem.* **263**, 3307–3313.

29. Cobianchi, F., Karpel, R.L., Williams, K.R., Notario, V. & Wilson, S.H. (1988). Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1: large scale overproduction in *Escherichia coli* and cooperative binding to single-stranded nucleic acids. *J. Biol. Chem.* **263**, 1063–1071.

30. Kumar, A., *et al.*, & Wilson, S.H. (1990). Mammalian heterogeneous nuclear ribonucleoprotein A1: nucleic acid binding properties of the COOH-terminal domain. *J. Biol. Chem.* **265**, 17094–17100.

31. Casas-Finet, J.R., Smith Jr., J.D., Kumar, A., Kim, J.G., Wilson, S.H. & Karpel, R.L. (1993). Mammalian heterogeneous ribonucleoprotein A1 and its constituent domains: nucleic acid interaction, structural stability and self-association. *J. Mol. Biol.* **229**, 873–889.

32. Mayeda, A., Munroe, S.H., Cáceres, J.F. & Krainer, A.R. (1994). Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.* **13**, 5483–5495.

33. Shamoo, Y., Abdul-Manan, N., Patten, A.M., Crawford, J.K., Pellegrini, M.C. & Williams, K.R. (1994). Both RNA-binding domains in heterogeneous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. *Biochemistry* **33**, 8272–8281.

34. Burd, C.G. & Dreyfuss, G. (1994). RNA binding specificity of hnRNP A1, significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**, 1197–1204.

35. Abdul-Manan, N., O'Malley, S.M. & Williams, K.R. (1996). Origins of binding specificity of the A1 heterogeneous nuclear ribonucleoprotein. *Biochemistry* **35**, 3545–3554.

36. Garrett, D.S., Lodi, P.J., Shamoo, Y., Williams, K.R., Clore, G.M. & Gronenborn, A.M. (1994). Determination of the secondary structure and folding topology of an RNA-binding domain of mammalian hnRNP A1 protein using three-dimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry* **33**, 2852–2858.

37. 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.

38. Bennett, M., Piñol-Roma, S., Staknis, D., Dreyfuss, G. & Reed, R. (1992). Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly *in vitro*. *Mol. Cell. Biol.* **12**, 3165–3175.

39. Nadler, S.G., *et al.*, & Williams, K.R. (1991). Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative, UP1, with RNA and DNA: evidence for multiple RNA binding domains and salt-dependent binding mode transitions. *Biochemistry* **30**, 2968–2976.

40. Shen, J., Zu, K., Cass, C.L., Beyer, A.L. & Hirsh, J. (1995). Exon skipping by overexpression of a *Drosophila* heterogeneous nuclear ribonucleoprotein *in vivo*. *Proc. Natl. Acad. Sci. USA* **92**, 1822–1825.

41. Görlach, 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.

42. Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J. & Mattaj, I.W. (1989). Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J.* **8**, 4163–4170.

43. Lutz-Freyermuth, C., Query, C.C. & Keene, J.D. (1990). Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc. Natl. Acad. Sci. USA* **87**, 6393–6397.

44. Burd, C.G. & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621.

45. Mattaj, I.W. (1993). RNA recognition: a family matter? *Cell* **73**, 837–840.

46. Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S. & Biamonti, G. (1996). hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J. Mol. Biol.* **259**, 337–348.

47. Cobianchi, F., Calvio, C., Stoppini, M., Buvoli, M. & Riva, S. (1993). Phosphorylation of human hnRNP A1 abrogates *in vitro* strand annealing activity. *Nucleic Acids Res.* **21**, 949–955.

48. Rajpurohit, R., Lee, S.O., Park, J.O., Paik, W.K. & Kim, S. (1994). Enzymatic methylation of recombinant heterogeneous nuclear RNP protein A1. Dual substrate specificity for *S*-adenosylmethio-nine:histone-arginine *N*-methyltransferase. *J. Biol. Chem.* **269**, 1075–1082.

49. Municio, M.M., Lozano, J., Sanchez, P., Moscat, J. & Díaz-Meco, M.T. (1995). Identification of heterogeneous ribonucleoprotein A1 as a novel substrate for protein kinase C zeta. *J. Biol. Chem.* **270**, 15884–15891.

50. Idriss, H., Kumar, A., Casas-Finet, J.R., Guo, H., Damuni, Z. & Wilson, S.H. (1994). Regulation of *in vitro* nucleic acid strand annealing activity of heterogeneous nuclear ribonucleoprotein protein A1 by reversible phosphorylation. *Biochemistry* **33**, 11382–11390.

51. Zamore, P.D., Patton, J.G. & Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609–614.

52. Sachs, A.B., Davis, R.W. & Kornberg, R.D. (1987). A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol. Cell. Biol.* **7**, 3268–3276.

53. Nietfeld, W., Mentzel, H. & Pieler, T. (1990). The *Xenopus laevis* poly(A) binding protein is composed of multiple functionally independent RNA binding domains. *EMBO J.* **9**, 3699–3705.

54. Burd, C.G., Matunis, E.L. & Dreyfuss, G. (1991). The multiple RNA-binding domains of the mRNA poly(A)-binding protein have different RNA-binding activities. *Mol. Cell. Biol.* **11**, 3419–3424.

55. Shamoo, Y., Abdul-Manan, N. & Williams, K.R. (1995). Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Res.* **23**, 725–728.

56. Zuo, P. & Manley, J.L. (1993). Functional domains of the human splicing factor ASF/SF2. *EMBO J.* **12**, 4727–4737.

57. Cáceres, J.F. & Krainer, A.R. (1993). Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J.* **12**, 4715–4726.

58. Karpel, R.L., Swistel, D.G., Miller, N.S., Geroch, M.E., Lu, C. & Fresco, J.R. (1975). Acceleration of RNA renaturation by nucleic acid unwinding proteins. *Brookhaven Symp. Biol.* **26**, 165–174.

59. Herschlag, D., Khosla, M., Tsuchihashi, Z. & Karpel, R.L. (1994). An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J.* **13**, 2913–2924.

60. Jokhan, L., Dong, A.-P., Mayeda, A., Krainer, A.R. &. Xu, R.-M. (1997). Crystallization and preliminary X-ray diffraction studies of UP1, the two-RRM domain of hnRNP A1. *Acta Cryst. D*, in press.

61. Otwinowski, Z. (1993). Oscillation data reduction program. In *Data Collection and Processing, Proceedings of the CCP4 Study Weekend*. (Sawyer, L., Issacs, N. & Bailey, S., eds) pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.

62. Furey, W. & Swaminathan, S. (1997). PHASES: a program package for the processing and analysis of diffraction data from macromolecules. *Methods Enzymol.*, in press.

63. Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112.

64. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.

65. Brünger, A.T. (1992). X-PLOR. A system for X-ray crystallography and NMR. Version 3.1, Yale University Press, New Haven, CT, USA.

66. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structure coordinates. *J. Appl. Cryst. A* **42**, 140–149.

67. Carson, M. (1991). Ribbons 2.0. *J. Appl. Cryst.* **24**, 958–961.

68. Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296.