Modeling by homology of RNA binding domain in A1 hnRNP protein

Andrea Ghetti[1], Martino Bolognesi[2], Fabio Cobianchi[3] and Carlo Morandi[1]

[1]*Istituto di Scienze Biologiche, Università di Verona, Strada Le Grazie 37134, Verona, Italy,* [2]*Dipartimento di Genetica e Microbiologia, Sezione di Cristallografia, Università di Pavia, Pavia, Italy* and [3]*Istituto di Genetica Biochimica ed Evoluzionistica, CNR, Pavia, Italy*

Received 10 October 1990; revised version received 8 November 1990

Eukaryotic nuclear RNA binding proteins share a common sequence motif thought to be implicated in RNA binding. One of the two domains present in A1 hnRNP protein, has been modelled by homology in order to make a prediction of the main features of the RNA binding site. Acylphosphatase (EC 3.6.1.7) was selected as template for the modeling experiment. The predicted RNA binding site is a β-sheet containing the two RNP consensus sequences as well as lysines and arginines conserved among the family.

RNA binding domain; Acylphosphatase; Modeling by homology

1. INTRODUCTION

Nuclear RNA binding proteins, although heterogeneous in size and charge [1-4], are marked by at least two features: (i) they are able to bind RNA both in vivo and in vitro; and (ii) they include an RNA binding domain (RNA-BD) of about 80-100 amino acids, also called 'RNP motif' or RNP-80 motif [5].

Notwithstanding a substantial lack of knowledge on most of their functions, recent evidence suggests a sequence specific interaction of several RNPs with RNA. Similarly to the specific interaction of poly(A) binding protein with poly(dA) or poly(rA) stretches [6], or to the specific interaction of snRNPs with their cognate snRNAs [7,8], some specific interaction of hnRNP proteins with hnRNA have also been postulated [9-12].

RNP motifs of all RNP proteins, contain two short stretches highly conserved, called consensus sequence RNP1 and RNP2 [5]. Merril and coworkers have shown by UV crosslinking that the aromatic side chains of RNP1 and RNP2 of A1 hnRNP, react with poly(dT) oligomers added to the mixture [13]. This observation suggests that the single stranded nucleic acid might interact directly with twe two RNP consensus sequences.

In addition to specific interactions between RNPs and RNA, such as the hnRNPs/hnRNA or snRNPs/snRNA complexes, RNA sequence recognition by RNPs is also thought to be relevant for RNA processing [5-16]. For the above reasons it would be of primary interest to shed some light on the structural bases of such interaction.

Correspondence address: C. Morandi, Istituto di Scienze Biologiche, Università di Verona, Strada Le Grazie, 37134 Verona, Italy

As a preliminary approach, in the absence of information regarding RNP structures at the molecular and the conformational level, we have undertaken a homology modeling study on one RNA binding domain of A1 hnRNP protein.

2. MATERIALS AND METHODS

The second RNP motif, from the N-terminus, in A1 hnRNP proteins (we call this domain A1 RNA-BD II), was modelled according to the following procedure. Template structure for modeling A1 RNA BD II was selected by screening the protein sequence data bank CHPR11 from CILEA-CNR, using the Bishop's program 'Pepscan' setting the lower score value to 5. 420 sequences were retrieved; further selection was performed saving only those protein sequences for which a three-dimensional structure was available. In this subset only one sequence, corresponding to the enzyme acylphosphatase, showed the secondary structure pattern $\beta\alpha\beta\beta\alpha\beta$, which has been predicted for the RNP motif [17]. Acylphosphatase was aligned with A1 RNA-BD using the Needleman-Wunsch program [18]; RNP motifs alignment was performed as previously described [17]. Modeling was performed using the program FRODO run on Evans and Sutherland PS 300 system [19].

Circular dichroic spectra were recorded on a Jasco model 600 spectropolarimeter. Measurements were conducted at room temperature using cells with optical pathlength ranging from 0.0045 to 0.001 cm. Acylphosphatase was dialyzed against 20 mM phosphate buffer, pH 5.8, and concentrated at various extents between 0.4 mM and 4 mM. UP1 was prepared from the A1 hnRNP protein as described [20]. UP1 fragment was dialyzed against 20 mM phosphate buffer, pH 7.4. Protein concentration was determined spectrophotometrically by the Bradford assay method at 595 nm. α-Helix content was estimated using the program of Siegel et al. [32].

3. RESULTS

The RNA binding domain that we have modelled by homology is one of the two RNP motifs present in the A1 hnRNP protein [21,22]; the reason for such a choice

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

```
0   STARPLKSVDYEVFGRYQGVCFRMYAEDEARKIGVVGWVKNTS------KGTVTGFVQGP-EEKVNSMKSWLSKVGSPSSRIDRTNFSNEKTISKLEYSNFSVRY
        o| o o  o o      o  oo| oo o| o  oo oo        |   ||o    oo  |oo o             |  o| o|| |o  |
2 RPGAHLTVKKIFVGGIKE---DTEEHHLRDYFEQYGKIEVIEIMTDR---GSGKKRGFAFVTF-DDHDSVDKIVIQKYHTVNGH---NCEVRKALSKQEMASAS
          *** *          *   *  ** *  * *            ***** *      *  **      *          *   *
1 SPKEPEQLRKLFIGGLSF---ETTDESLRSHFEQWGTLTDCVVMRDPNTKRS---RGFGFVTY-ATVEEVDAAMNARPHKVDGR---VVEPKRAVSREDS
3 SITEPEHMRKLFIGGLDY---RTTDENLKAHFEKWGNIVDVVVNKDPRTKRSS---GFGFITY-SHSSMIDEAQKSRPHKIDGR---VVEPKRAVPRQD
4 SPNAGATVKKLFVGALKD---DHDEQSIRDYFQHFGNIVDINIVIDKETGKK---RGFAFVEF-DDYDPVDKVVLQKQHQLNGK---MVDVKKALPKQN
5 KTDPRSMNSRVFIGNLNTL--VVKKSDVEAIFSKYGKIVGCSVH----------KGFAFVQYVNERNARAAVAGEDGRMIAGQ---VLDINLAAEPKVNRG
6 EHPQASRCIGVFG--LNT---NTSQHKVRELFNKYGPIERIQMVIDAQTQRS---RGFCFIYFEKLSDARAAKDSCSGIEVDGR---RIRVDFSITQRA
7 NDPRASNT-NLIVNYLPQ---DMTDRELYALFRAIGPINTCRIMRDYKTGYSF---GYAFVDFTSEMDSQRAIKVL-NGITVRN---KRLKV
8 PGGESIKDTNLYVTNLPR---TITDDQLDTIFGKYGSIVQKNILRDKLTGRP---RGVAFVRYNKREEAQEAISAL-NNVIPEG---GSQPLSVR
9 AQGDAFK-T-LFVARVNY---DTTESKLRREFEVYGPIKRIHIVYNKGSEGSGKPRGYAFIEYEHERDMHSAYKHAD-GKKIDG---RRVLVDVERTVKG
10 SVENSSA-S-LYVGDLEP---SVSEAHLYDIFSPIGSVSSIRVCRDAITKTST---GYAYVNFNDHEAGRKAIEQL-NYTPIKG---RLCRIMWSQ
11 LRKKGSG--NIFIKNLHP---DIDNKALYDTFSVFGDILSSKIATDE-NGKS---KGFGFVHFEEEGAAKEAIDAL-NGMLLNG---QEIYVAPHL
12  ETKAHY-TNLYVKNINS---ETTDEQFQELFAKFQPIVSASLEKDAD-GK-L--KGFGFVNYEKHEDAVVKAVEALNDSELNG---EKLYVGRANKKNERM
13 EKMAKYQGVNLFVKNLDD---SVDDEKLEEEFAPYGTITSAKVMRTE-NGKS---KGFGFVCFSTPEEATKAITEK-NQQIVAG---KPLYVAIAQRKDVR
14 DRITRYQGVNLYVKNLDD---GIDDERLRKEFSPFGTITSAKVMME--GRKS---KGFGFVCFSSPEEATKAVTEM-NGRIVAT---KPLYVALAQRKEERQ
15 GARAKEF-TNVYIKNFGE---DMDDERLKDL---FGPALSVKVMTDESGK--L--KGFGFVSFERHEDAQKAVDEM-NCEL-NG---KQIYVGRAQKKVERE
16 PSLRKSGVGNIFIKNLDK---SIDNKALYNTFSAFGNILSCKVVCD-ENG-S---KGYGFVHFETQCAAERAIEKM-NGMLLND---RKVF-GRFKSRKERE
17 PSAPSYIMASLYVCDLHP---DVTEAMLYEKFSPAGPILSIRVCRDMITRRSL---GYAYVNFQQPADAERALDTM-NFDVIKG----K-FVRIMWSQRD
18 EGSESTTPFNLFIGNLNP---NKSVAELKVAISEPFAKNDLAVV-DVRTGTNR--K-FGYVDFESAEDLEKALELT-GLKVFGN---EIKLEKPK
19 DSKKVRAARTLLAKNLSF---NITEDELKEVFEDALEI-RL-VSQD---GKS---KGIAYIEFKSEADAEKNLEEKQGAEI-DG---RSVSLYYTGEKGQ
20 NSTWSGESKTLVLSNLSY---SATEETLQEVFEKATFI-KVPQ-NQQ--GKS---KGYAFIEFASFEDAKEALNSCNKMEIEGG---RTIRLELQGPRG
21  NARSQPSKTLFVKGLSE---DTTEETLKESFEGSVRA-RIVT--DRETGSS---KGFGFVDFNSEEDAKAAKEAMEDGEIDGN---KVTLDWAKPKGE
22  MAAADVEYRCFVGGLAW---ATSNESLENAFASYGEILDSKVITDRETGRS---RGFGFVTFSSENSMLDAIENMNGKELDGR---NITVNQAQSR
```

Fig. 1. Sequence alignment of horse muscle acylphosphatase (EC 3.6.1.7) and A1 RNA-BD II. The alignment includes 21 RNA binding domains from other RNP proteins. The sequences are marked as follows; 0, horse muscle acylphosphatase; 2, A1 RNA-BD II; sequences 1 and 3–22 are marked as described in [17]. Identities between acylphosphatase and A1 RNA-BD II are indicated by "|"; similar residues are evidenced by "o". Hydrophobic amino acids conserved among RNP proteins are marked by asterisks.

is related to the availability of a large amount of A1 protein since the corresponding cDNA has been cloned in an expression vector [23]. Modeling by homology is at the moment at the only chance for studying the structure of the A1 hnRNP protein, since this RNP is highly insoluble, restricting therefore the access to X-ray crystallography or NMR spectroscopy. The primary structure of A1 hnRNP protein has been characterized from human and rat [21,22], and the two sequences were found to be identical. The protein comprises two RNP motifs tandemly arranged at the N-terminal, followed by a C-terminal glycine-rich region. The enzyme acylphosphatase from horse muscle (ACP) (EC 3.6.1.7), whose structure has recently been resolved by

Fig. 2. C-α backbone of the modelled A1 RNA-BD II. Lysines, arginines and aromatic residues of the two RNP consensus sequences are shown in stereo view.

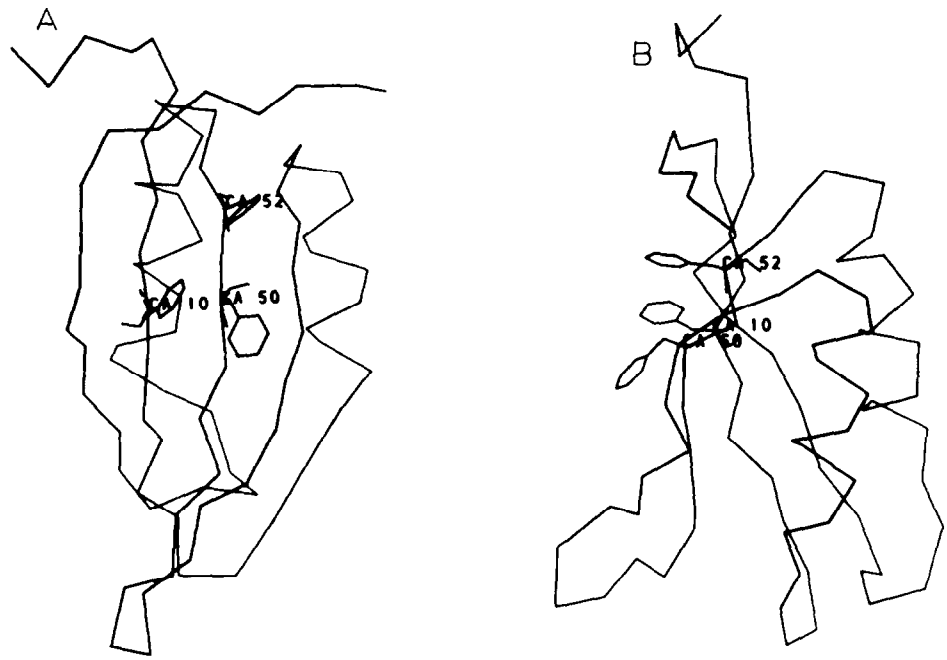

Fig. 3. C-α backbone of the modelled A1 RNA-BD II. Aromatic residues of the two RNP consensus sequences are shown. (A), front view of the β-sheet; (B), lateral view.

NMR [24], contains 98 amino acid residues, and displays a secondary structure pattern, $\beta\alpha\beta\beta\alpha\beta$, extremely similar to that predicted for RNA-BD [17]. The elements of secondary structure are arranged to form a β-sheet lying on two α-helices. Fig. 1 shows the alignment of ACP with A1 RNA-BD II and 22 other RNP motifs. The sequence identity between ACP and RNA-BD is 13.5%; on the other hand, considering amino acid similarity, the score rises to 38%. It is interesting to note that several (13/20) of the hydrophobic residues conserved within the RNP family are also present in ACP. A1 RNA-BD II was modelled using the alignment shown in Fig. 1, and the atomatic coordinates of ACP (kindly provided by Dr A. Pastore). First, residues comprised in secondary structure elements were replaced in the ACP structure with the corresponding residues of A1 RNA-BD II protein. Subsequently loop regions were substituted accounting for insertion/deletion; modified loops were modelled manually without moving secondary structure elements. Only few tight contacts between side chains have been adjusted by rotating these groups (see Discussion). The modeling of the core residues is particularly satisfactory: the resulting structure maintains a clear amphypathic character in its

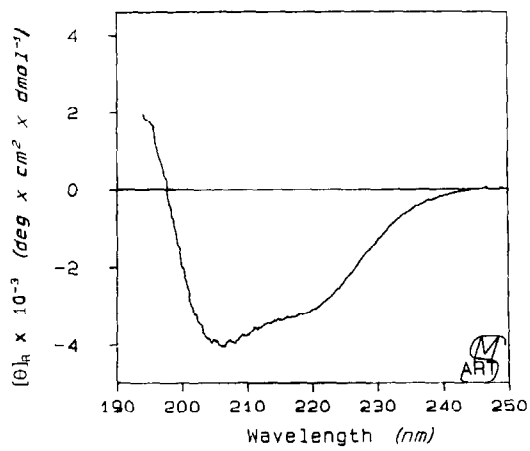

Fig. 4. Circular dichroic spectrum, at 25°C, of 60 μM UP1 polypeptide in 20 mM sodium phosphate, pH 7.5.

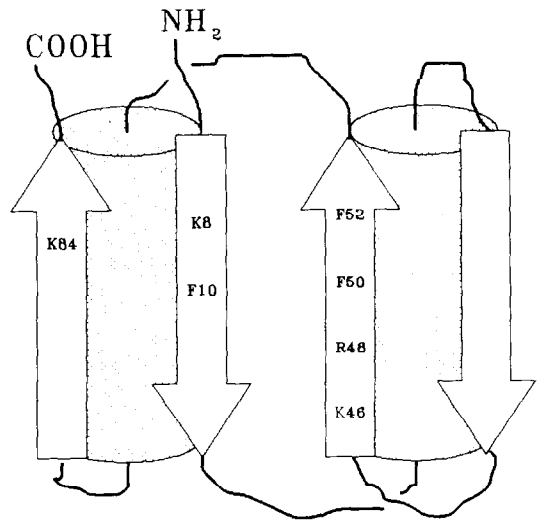

Fig. 5. Schematic drawing of hypothetical structure of RNP's RNA binding domain. Residues expected to play a role in RNA binding are evidenced.

secondary structure elements, supporting that the ACP topology is conserved in A1 RNA-BD II.

All the side chains oriented toward the interior of the molecule are non-polar, with the only exception of H58 (numeration referred to the sequence domain A1 RNA-BD II).

The β-sheet surface is rich in positively charged groups: two arginines and four lysines protrude from the β-sheet (Fig. 2). Four of these six residues (K8, K46, R48 and K84) are well conserved in the RNP family.

Also protruding from the beta sheet are the three phenylalanine residues belonging to the two RNP consensus sequences (Fig. 3A,B). Positively charged residues protruding from the beta sheet are clustered in two regions which are located on opposite sides relative to three aromatics F10, F50, F52, occupying the central portion of the beta sheet (Fig. 2).

In the attempt to obtain independent experimental evidence for such a model, preliminary CD measurements were carried out on both ACP and purified fragments of the A1 proteins. As an example, the CD spectrum of the A1 fragment 1-184, comprising the two RNP motifs, is reported in Fig. 4. The spectrum is almost identical to the spectra of the two independent RNA binding domains (data not shown). A reliable comparison with the CD properties of ACP is difficult since we found that in the latter case the CD spectrum is strongly concentration-dependent. At 10^{-5} M concentration the CD patterns of the two polypeptides are similar but not identical. This could be an indication that, under these experimental conditions, there are differences among the two structures. This is definitely possible, due to the low sequence homology between A1 RNA-BD II and ACP (see Discussion below).

4. DISCUSSION

A tentative model for an RNP motif from hnRNP protein A1 has been described. The sequence of the RNA binding domain was used in model building employing the atomic coordinates of acylphosphatase. Although identity between ACP and A1 RNA-BD II sequences is low, the hydrophobic sites conserved in RNP family are well maintained in the alignment between the two proteins; similarity of the hydrophobic cores seems to be one of the major determinants of structure homology [24-26]. Furthermore, examples are known of protein sharing very low sequence identity and yet showing similar structures [25-28].

It has been shown that, in general, the lower the homology between sequences the lower the similarity between tertiary structures. According to Chothia and Lesk [30], in the case we have considered, the real structure of A1 RNA-BD II could diverge significantly from ACP starting model (having a root mean square deviation of the C-α atoms larger than 2 Å. Due to the low sequence identity and its implications we did not use a rigorous modeling procedure for loops and side chains. At this stage we only focus on the overall topology of the structure and on the general features of the RNA binding site(s).

Aromatic side chains of the two RNPs, highly conserved and strongly implicated in RNA binding, protrude from the same side of the molecule. Binding of A1 protein to RNA or single stranded DNA is salt sensitive; and this has also been observed for many other RNPs. This means that ionic interactions, likely between phosphate groups and Lys and Arg side-chains, occur at the binding site. The presence of the conserved lysines and arginines in proximity of the aromatic residues implicated in RNA binding is in keeping with the structural features of the proposed model.

Recently, binding of U1A and U1B'' to their cognate snRNA, has been characterized using chimaeric proteins [31]. Scherly and colleagues, have shown that the isolated RNP motif is still capable of specific binding; moreover, they have identified a stretch of amino acids just preceding RNP1 consensus sequence (corresponding in A1 RNA-BD II to residues 34-45) determining binding specificity. This findings is in agreement with our model since the 34-45 segment comprises one of the four β-strands and a loop, adjacent to the aromatic groups of RNP1 consensus sequence. In conclusion, although the model presented remains largely speculative, in relation to its general features, namely the folding topology and overall disposition of residues in space, some suggestive observations can be made.

A four-strand antiparallel β-sheet seems to be a good candidate for binding site, since it includes the two RNP consensus sequences, exposing the aromatic side chains known to contact RNA. Furthermore it is rich in lysines and arginines and comprises a region which seems to be involved in binding specificity in RNP proteins.

According to this model, it cannot be ruled out that in some proteins of the RNP family, also the stretches corresponding to residues 6-12 and 82-88 of the A1 RNA-BD II, which are part of in the beta sheet, might play some role in RNA binding.

Due to the limitations of the CD results in the condition we could test, the model remains hypothetical in nature. Sound experimental support to our hypothesis must rest on a more detailed conformational analysis by NMR and CD on the fragments of A1 protein. Work is in progress in this direction.

Acknowledgements: We are indebted to Dr E. Peggion for assistance in analysis of CD spectra and helpful discussion. We thank Dr A. Lesk for useful discussions. Acylphosphatase was a generous gift from Dr G. Ramponi and Dr M. Stefani. We also acknowledge Dr A. Pastore for giving us the atomic coordinates of ACP, Dr S. Riva for useful suggestions and G. Di Cesare for helping us in the computer operations. This work has been supported by the Italian Ministry of Education (40% funds), by the CNR 'Progetto Finalizzato Ingegneria Genetica', and by the 'Progetto Finalizzato Biotecnologie e Biostrumentazione', CNR Italy.

REFERENCES

[1] Dreyfuss, G. (1986) Annu. Rev. Cell. Biol. 2, 459-498.
[2] Theissen, H., Etazerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argoes, P., Luhrmann, R. and Philipson, L. (1986) EMBO J. 5, 3209-3217.
[3] Sillekens, P.T., Habets, W.J., Beijer, R.P. and van Venrooij, W.J. (1987) EMBO J. 6, 3841-3448.
[4] Habets, W.J., Sillekens, P.T.G., Hoet, M.H., Schalken, J.A., Roebroek, A.J.M., Leunissen, J.A.M., van den Ven, W.J.M. and van Venrooij, W.J. (1987) Proc. Natl. Acad. Sci. USA 84, 2421-2425.
[5] Bandziulis, R.J., Swanson, M.S., Dreyfuss, G. (1989) Genes Dev. 3, 431-437.
[6] Sachs, A.B., Bond, M.W. and Kornberg, R.D. (1986) Cell 45, 827-835.
[7] Scherly, D. et al. (1989) EMBO J. 8, 4163-4170.
[8] Lutz-Freyermuth, C. and Keene, J.D. (1989) Mol. Cell. Biol. 9, 2975-2982.
[9] Beyer, A.L. and Osheim, Y.N. (1988) Genes Dev. 2, 754-765.
[10] Swanson, M.S. and Dreyfuss, G. (1988) 7, 3519-3529.
[11] Mayrand, S.H. and Pederson, T. (1990) Nucleic Acids Res. 18, 3307-3318.
[12] Buvoli, M., Cobianchi, F., Biamonti, G. and Riva, S. (1990) Nucleic Acids Res. (submitted).
[13] Merril, B.M., Stone, K.L., Cobianchi, F., Wilson, S.H. and Williams, K.R. (1988) J. Biol. Chem. 263, 3307-3313.
[14] Golarski, T.J., Edstrom, J.E. and Baker, B.S. (1989) Cell 56, 1011-1018.
[15] Bell, L.R., Maine, E.R., Schedl, P. and Cline, T.W. (1988) Cell 55, 1037-1046.
[16] Amrein, H., Gorman, M. and Nothiger, R. (1988) Cell 55, 1025-1035.
[17] Ghetti, A., Padovani, C., Di Cesare, G. and Morandi, C. (1989) FEBS Lett. 257, 373-376.
[18] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 120, 97-120.
[19] Jones, T.A. (1978) J. Appl. Christallogr. 11, 268-272.
[20] Kumar, A., Williams, K.R. and Szer, W. (1986) J. Biol. Chem. 261, 11266-11273.
[21] Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M.T., Ghetti, A., Riva, S. and Morandi, C. (1988) Nucleic Acids Res. 16, 3751-3770.
[22] Cobianchi, F., Sen Gupta, D.N., Zmudzka, B.Z. and Wilson, S.H. (1986) J. Biol. Chem. 261, 3536-3543.
[23] Cobianchi, F., Karpel, R.L., Williams, K.R., Notario, V. and Wilson, S.H. (1988) J. Biol. Chem. 263, 1063-1071.
[24] Saudek, V., Williams, R.J.P. and Ramponi, G. (1989) FEBS Lett. 242, 225-232.
[25] Bowie, J.U., Clarke, N.D., Pabo, C.O., Sauer, R.T. (1990) Proteins 7, 257-264.
[26] Bashford, D., Chothia, C. and Lesk, A.M. (1987) J. Mol. Biol. 196, 199-216.
[27] Lesk, A.M. and Chothia, C. (1980) J. Mol. Biol. 136, 225-270.
[28] Sevcik, J., Sanishvili, R.J., Pavlovsky, A.G. and Polyakov, K.M. (1990) Trends Biochem. Sci. 15, 158-162.
[29] Leijonmark, M., Appelt, K., Badger, J., Liljas, A., Wilson, S.K. and White, S.W. (1988) Proteins 3, 243-251.
[30] Chothia, C. and Lesk, A.M. (1986) EMBO J. 5, 823-826.
[31] Scherly, D., Boelens, W., Dathan, N.A., van Venrooij, W.J. and Mattaj, I.W. (1990) Nature 345, 502-506.
[32] Siegel et al. (1980) Anal. Biochem. 104, 160-167.

NOTE ADDED IN PROOF

After the refinement of our model we learnt that the three dimensional structure of a protein belonging to the RNP family (U1A small nuclear ribonucleoprotein), has been determined (Nagai et al. Nature, in press; K. Nagai pers. comm.). A striking structural similarity between the putative RNA binding domain of U1A and acylphosphatase was observed at the level of Cα backbone (A. Lesk, pers. comm.; K. Nagai, pers. comm.).