Volume 257, number 2, 373–376 FEB 07796 November 1989

Secondary structure prediction for RNA binding domain in RNP proteins identifies $\beta\alpha\beta$ as the main structural motif

A. Ghetti, C. Padovani, G. Di Cesare and C. Morandi

Istituto di Scienze Biologiche, Università di Verona, Strada le Grazie, 37134 Verona, Italy

Received 23 August 1989

In eukaryotic cells transcript processing is strictly dependent upon binding of specific proteins. Nuclear RNA binding proteins share a common domain, which is involved in RNA binding. In order to characterize RNP-RNA interactions we have performed a secondary structure prediction based both on statistical algorithms and comparative analysis of different proteins. A high conservation for secondary structure propensity between different RNPs was observed.

Ribonucleic protein; Structure prediction

1. INTRODUCTION

In recent years, it has been demonstrated that eukaryotic mRNAs are always associated in the cytoplasm as well as in the nucleoplasm with specific proteins which play an important role in transcript metabolism [1,2].

After transcription by Pol II, the synthesized heterogeneous nuclear RNA (hnRNA) tightly associates with a set of proteins called 'heterogeneous nuclear ribonucleic proteins' (hnRNP) to form a NA/protein complex in a 'nucleosome-like' structure, sedimenting at 200–250 S.

The hnRNA intimately associated with the hnRNP in the nucleus undergoes several modifications, such as capping at the 5′ end, polyadenylation at the 3′ end and splicing, before being secreted through the nuclear pores into the cytoplasm [3–6].

The nuclear RNA binding proteins (hnRNP) include at least 20 protein species, only partially characterized, which are for the majority basic and with molecular masses ranging from 32 to 120 kDa [7].

Recently several cDNA corresponding to these proteins have been isolated and sequenced. The first was characterized by Riva and coworkers, during a research on single-stranded DNA binding proteins (ssDBP) in eukaryotic cells [8]. This work demonstrated that human single-stranded DNA binding protein UP1 (identified by Alberts, see [9]) is a degradation product of hnRNP protein A1. It has been hypothesized that hnRNP proteins contain distinct domains and one of these (in the specific case the one corresponding to UP1 polypeptide) confers with the protein the affinity for single-stranded nucleic acids. Further analysis of UP1 sequence domain revealed that it contains a tandem duplicate of 90 residues.

Later works identified the UP1's subdomain in other nuclear proteins from different sources: polyadenylate binding proteins [10–12], C1 hnRNP [13], nucleolin [14], proteins bound to small nuclear RNAs (snRNP) [15–17] and others.

A common characteristic of these proteins is their ability to bind single-stranded nucleic acids in vitro and RNA in vivo.

The 90 residue sequence, now called 'RNA binding domain', is strikingly conserved from yeast to man and it has also recently been found in plants [18]. RNA binding domain contains two short stretches, whose degree of conservation through evolution is total; these regions contact RNA as shown by Merril and coworkers using UV crosslinking [19].

It has generally been assumed that hnRNP proteins bind RNA regardless of nucleic acid sequence. Recent data suggest that this could not be true. Swanson and Dreyfuss demonstrated that in vitro A1 protein has a specific affinity for sequences located at the 3′ end of mammalian introns [20]. Other, more indirect evidence comes from studies on development in *Drosophila*; it has been reported that sex determination during development is regulated at the level of RNA processing of particular transcripts. The proteins controlling this alternative splicing contain the RNA binding domain that could interact with transcripts in a sequence-specific way, thus locating different splicing sites [21–23].

In view of all these data it would be of great interest to define the three-dimensional structure of RNA bin-

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

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

Fig.1. Secondary structure prediction for RNA binding domain of RNP proteins. Aligned sequences are shown together with secondary structure prediction; below each sequence the first row represents the structure prediction according to the method of Garnier et al. [26] and the second row the prediction following Chou and Fasman [25]. Asterisk indicates α-helix while double line is for β-sheet. (a) and (b) indicate the two RNP consensus sequences. The order of sequences is the following: (1) human A1 hnRNP domain 1 (dom.1) [32]; (2) human hnRNP dom.2 [32]; (3) *Drosophila* A1 hnRNP dom.1 [33]; (4) *Drosophila* A1 hnRNP dom.2 [33]; (5) human C1 hnRNP [13]; (6) *Drosophila* tra-2 gene product [23]; (7) *Drosophila* sx1 gene product dom.1 [22]; (8) *Drosophila* sx1 gene product dom.2 [22]; (9) human snRNP 70K U1 [15]; (10–13) yeast polyadenylate binding protein dom.1–4 [10,11]; (14–17) human polyadenylate binding protein dom.1–4 [12]; (18–21) hamster nucleolin dom.1–4 [14]; (22) abscisic acid-inducd protein [18].

ding domain, in order to understand the binding mechanism and to elucidate the molecular aspects of alternative splicing.

Unfortunately, since X-ray diffraction maps of RNP proteins are not available at the moment, we can only infer some characteristics of the structure from the analysis of their amino acid sequences.

Multiple alignment of RNA binding domains of many different proteins reveals a pattern of strong conservation through the course of evolution. Such a rigid conservation of protein sequence could be reasonably correlated to the tertiary structure maintenance.

Starting from this assumption we have utilized in a comparative way, secondary structure prediction algorithms to make proposals about some structural properties of RNA binding domain.

2. MATERIALS AND METHODS

Computer-aided algorithms were performed on an IBM personal computer PS/2 M 50.

Sequences were aligned first using the Needleman and Wunsch [24] algorithm; alignment was then refined manually.

Helix and sheet propensity were calculated employing the Chou and Fasman [25] and the Garnier-Osguthorpe-Robson [26] methods. Prediction algorithms by Garnier and coworkers were tested with several decision constant combinations, ranging from 0 to −88 without finding any appreciable variation in secondary structure prediction.

When prediction uncertainties were found utilizing Chou and Fasman method, only the higher score was considered.

3. RESULTS AND DISCUSSION

The rationale of our work is based on two main considerations. (i) Secondary structure prediction algorithms used in this work are those of Chou and Fasman [25] and Robson-Osguthorpe-Garnier [26]. Accuracy of prediction reached by these methods individually varies from 50% to 60% [27,28] but more attainable results may be obtained through the simultaneous application of the different algorithms to the same sequence ([27,28] and references therein). (ii) We assumed that highly similar sequences from different proteins belonging to the same family conserve the general pattern of secondary structure. Thus, comparing the secondary structure predictions for different RNA binding domains, we deduced a secondary structure consensus.

Applying the two prediction algorithms to each of the 22 sequences examined, we have drawn the table shown in fig.1. It seems to demonstrate a good conservation in secondary structure predisposition at a defined region of RNA binding domain.

The results of structure prediction shown in fig.1 can be better visualized in the barret diagram depicted in fig.2 where percent frequencies of secondary structure propensity are plotted for each residue.

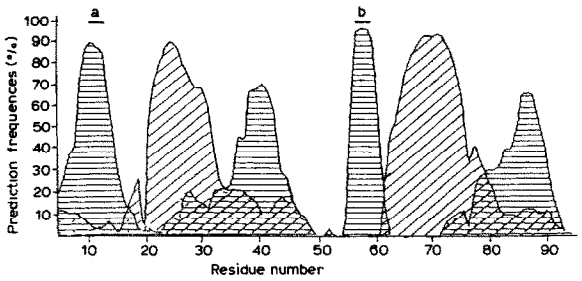

Fig.2. Secondary structure conservation in RNA binding domain of RNP proteins. Percent frequencies of predicted α-helices and β-sheet are plotted for each residue position along the domain. Horizontal shading was used for β conformation frequencies; diagonal shading was used for α conformation frequencies. (a) and (b) indicate RNP consensus sequences.

The domain displays a striking conservation for secondary structure propensity at the defined regions: the pattern of secondary structure prediction summarized in fig.2 suggests the presence of two α-helices and 4 β-sheets in the RNA binding domain of RNP. The general feature of the domain seems to be $\beta\alpha\beta$ supersecondary structure: this structure appears to be repeated twice in the domain.

For the two RNP consensus sequences for which a direct interaction with RNA by means of aromatic stacking has been demonstrated [19], our results predict a β-sheet conformation. It has to be noted that in a single-stranded DNA (ssDNA) binding protein belonging to another protein family, protein gp5 of fd phage, X-ray crystallography demonstrated that protein-ssDNA interaction is due mainly to aromatic stacking of DNA's bases and phenylalanine and tyrosines protruding from a β-sheet [29]. For the predicted α-helices, wheel plots have been drawn (data not shown) and almost all display an amphipathic pattern.

Recently, with a different approach, Chan and coworkers [30] identified a putative α-helix located in a region partially overlapping the first helix we have hypothesized. According to their model, this helix should be few residues shifted toward the C terminal and longer, if it is to be compared to our prediction.

The region between the second and third presumptive β-sheet is most subjected to insertions/deletions: such a sequence plasticity is very commonly found in loops or coil structures.

From the observation reported in this work 3 main considerations can be drawn: (i) the RNA binding domain of RNP proteins shows an extensive conservation of secondary structure propensity, according to the two algorithms utilized; (ii) the domain seems to belong to the α/β class of globular proteins as defined by Levitt and Chothia [31]; and (iii) the two RNP consensus sequences are predicted to be in β conformation. The same conformation was found in another protein family (gp5), in the region interacting with single-stranded DNA.

In the gp5 protein family, aromatics protruding from the β sheets confer to the polypeptide a general affinity for ssDNA; similarly, the two RNP consensus sequences, containing aromatic residues and predicted in β conformation, could also be involved in aspecific binding of RNA. The model outlined here is hypothetical in nature. It will be necessary to collect experimental data from NMR, X-ray diffraction and circular dichroism to check its validity. However, it represents a valuable reference for studies aimed at defining the nature of binding of RNP to RNA.

Acknowledgements: We thank Dr P.F. Pignatti for generous support, and Dr S. Riva and Dr M. Bolognesi for useful suggestions and discussions. We also acknowledge Dr G. Gaudino and Dr S. Giordano for critical reading of the manuscript. This work has been supported by the Italian Ministry of Education (40% funds) and by the CNR 'Progetto Finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie'.

REFERENCES

[1] Dreyfuss, G. (1986) Annu. Rev. Cell. Biol. 2, 459–498.
[2] Knowler, J.T. (1983) Int. Rev. Cytol. 84, 103–153.
[3] Beyer, A.L., Miller, O.L. and McKnight, S.L. (1980) Cell 20, 76–84.
[4] Beyer, A.L. and Osheim, Y.N. (1988) Genes and Development 2, 754–765.
[5] Padgett, R.A., Grabowoski, P.I., Konarska, M.M., Seiler, S. and Sharp, P.A. (1988) Annu. Rev. Biochem. 55, 1119–1150.
[6] Fakan, S., Leser, G. and Martin, T.E. (1986) J. Cell. Biol. 103, 1153–1157.
[7] Swanson, M.S. and Dreyfuss, G. (1988) Mol. Cell. Biol. 8, 2237–2241.
[8] Riva, S., Morandi, C., Tsoulfas, P., Pandolfo, M., Biamonti, G., Merril, B., Williams, K.R., Multhaup, G., Beyreuther, K., Werr, H., Henrich, B. and Schafer, K.P. (1986) EMBO J. 6, 2267–2273.
[9] Herrick, G. and Alberts, B. (1976) J. Biol. Chem. 261, 2124–2132.
[10] Adam, S.A., Nakagawa, T.Y., Swanson, N.S., Woodruff, T.K. and Dreyfuss, G. (1986) Mol. Cell. Biol. 6, 2932–2943.
[11] Sachs, A.B., Bond, M.W. and Kornberg, R.D. (1986) Cell 45, 827–835.
[12] Grange, T., Martins de Sa, C., Odds, J. and Pictet, R. (1987) NAR 15, 4771–4787.
[13] Swanson, N.S., Nakawa, T.Y., LeVan, K. and Dreyfuss, G. (1987) Mol. Cell. Biol. 7, 1731–1739.
[14] Lapeyre, B., Amalric, F., Ghaffary, S.H., VenKatarama Rao, S.V., Dumbar, T.S. and Olson, M.O. (1986) J. Biol. Chem. 261, 9167–9173.
[15] Theissen, H., Etzerodt, M., Reuter, M., Schneider, C., Lottspellch, F., Argos, P., Luhrmann, R. and Philipson, L. (1986) EMBO J. 5, 3209–3217.
[16] Weiben, E.D., Rohleder, A.M., Nenninger, J.M. and Pederson, T. (1985) Proc. Natl. Acad. Sci. USA 82, 7914–7918.
[17] Rokeach, L.A., Haselby, J.A. and Hoch, S.O. (1988) Proc. Natl. Acad. Sci. USA 85, 4832–4836.
[18] Gomez, J., Sanchez-Martinez, D., Stiefel, V., Rigau, J., Puigdomenech, P. and Pages, M. (1988) Nature 334, 263–264.
[19] Merril, B.M., Stone, K.L., Cobianchi, F., Wilson, S.H. and Williams, K.R. (1988) J. Biol. Chem. 263, 3307–3313.
[20] Swanson, N.S. and Dreyfuss, G. (1988) EMBO J. 7, 3519–3529.
[21] Goralski, T.J., Edstrom, J.E. and Baker, B.S. (1989) Cell 56, 1011–1018.
[22] Bell, L.R., Maine, E.M., Schede, P. and Cline, T.W. (1988) Cell 55, 1037–1046.
[23] Amrein, H., Gorman, M. and Nothiger, R. (1988) Cell 55, 1025–1035.
[24] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 120, 97–120.
[25] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–147.
[26] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[27] Argos, P. (1988) in: Protein Structure (Creighton, T.E. ed.) pp.169–190, IRL Press.
[28] Taylor, W.R. (1987) in: Nucleic Acid and Protein Sequence Analysis (Bishop, M.J. and Rawlings, C.J. eds) pp.285–322, IRL Press.
[29] McPherson, A. and Brayer, G.D. (1985) in: Biological Macromolecules and Assemblies, vol.2 (McPherson, A. and Journak, F. eds) pp.323–392, John Wiley and Sons, New York.
[30] Chan, E.K.L., Sullivan, F.K. and Tan, E.M. (1989) NAR 17, 2233–2244.
[31] Levitt, M. and Chothia, C. (1976) Nature 261, 552–557.
[32] Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M.T., Ghetti, A., Riva, S. and Morandi, C. (1988) NAR 16, 3751–3770.
[33] Haynes, S.R., Rebbert, M.L., Mozer, B.A., Forquignon, F. and Dawid, I.G. (1987) Proc. Natl. Acad. Sci. USA 84, 1819–1823.