Molecular Phylogenetics and Functional Evolution of Major RNA Recognition Domains of Recently Cloned and Characterized Autoimmune RNA-Binding Particle

Erhan Süleymanoğlu

Medical Faculty, Vienna Biocenter, Institute of Biochemistry, University of Vienna, Vienna, Austria.

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are spliceosomal macromolecular assemblages and thus actively participate in pre-mRNA metabolism. They are composed of evolutionarily conserved and tandemly repeated motifs, where both RNA-binding and protein-protein recognition occur to achieve cellular activities. By yet unknown mechanisms, these ribonucleoprotein (RNP) particles are targeted by autoantibodies and hence play significant role in a variety of human systemic autoimmune diseases. This feature makes them important prognostic markers in terms of molecular epidemiology and pathogenesis of autoimmunity. Since RNP domain is one of the most conserved and widespread scaffolds, evolutionary analyses of these RNA-binding domains can provide further clues on disease-specific epitope formation. The study presented herein represents a sequence comparison of RNA-recognition regions of recently cloned and characterized human hnRNP A3 with those of other relevant hnRNP A/B-type proteins. Their implications in human autoimmunity are particularly emphasized.

Key words: hnRNP proteins, RNA folding, RNA-protein interactions, molecular evolution

Introduction

The enclosure of eukaryotic genomes within the nuclear envelope evolutionarily generated the necessity to transport macromolecules selectively between nucleus and cytoplasm. Following their synthesis in cytoplasm, histones and nucleic acid polymerases have to reach the nucleus, while specifically cytoplasmic proteins have to be kept out. Mature mRNA, tRNA and rRNA molecules and their associated proteins have to follow the opposite route, while their immature precursors have to be kept in (*1–5*). Eukaryotic mRNA is enzymatically metabolized and compacted with proteins within nuclei to generate functional messenger ribonucleoprotein (mRNP) particles (*6*, *7*). The control of protein biosynthesis involves regulation of intranuclear functions with participation of specific proteins, many of which seem to enter from the cytoplasm as a time-specific event (*3–7*). The heterogeneous nuclear ribonucleoproteins (hnRNPs) comprise a group of important regulators engaged in these cellular processes (*8–10*).

Extensive research efforts have been devoted to the cloning and molecular characterization of numerous ribonucleoproteins (RNPs) in the context of their active role in mRNA biogenesis and metabolism (*1, 2, 9–11*), as well as in terms of their involvement as autoantigens in human diseases (*11*). RNA-protein interactions are considered as primary macromolecular forces governing gene expression. Hence, emphasis is put on the roles of various RNA-binding motifs and their subsequent roles in governing cellular activities, both in health and pathology (*9*). Recently, there has been considerable interest in the participation of these RNA-binding particles in triggering the immune response in human autoimmune disorders (*11*).

Substantial research both at cDNA and protein level have been performed on immunochemical features of human autoantigens (http://www.zoo.uni-heidelberg.de/mol_evol/MB/ana_base.html). Of particular interest is the group of mammalian hnRNPs, which constitutes a part of spliceosome, which itself is an autoimmune target (*12*). Novel nucleotide sequences are continuously reported, which afterwards are used to delineate homology among various members based on previously determined nucleotides and protein sequences.

hnRNPs of diverse origin possess a common well-known structural motif. However, the functional chemistry of these domains remains unknown. The mechanism of contribution of numerous hnRNPs with a similar structural RNP motif (*13*) to the induc-

E-mail: erhan@e-mail.dk;
erhan@mail-online.dk

tion of different immune reactions is sequence-specific (*14*). Phylogenetic approach is essential to find functionally important genomic sequences based on detection of their high degree of conservation across different species. Such approach shows the level of improvement of the prediction of gene-regulatory elements in the human genome. This necessitates the study of the degree of homology of RNA recognition motifs (RRM) among these proteins, requiring an evolutionary computation. Having considered the importance of submitting new sequences for further functional characterization, the newly cloned and expressed cDNA of previously unknown member of the hnRNP A/B family of proteins (Figure 1) is presented here. Its RNA-binding properties and tissue-specific gene expression profiles were recently determined (*15*). Based on the concept of correlation between sequences and RNA-binding modes, a systemic search was performed for nucleic acid association by evaluating sequence conservation using multiple sequence alignments search tools. This study continues phylogenetic results obtained from previous larger data sets (*16*).

Fig. 1 The general 2xRNA-binding domain (RBD)—glycine structure of hnRNP A3. The space between the two adjacent RBDs is occupied by inter-RNA recognition motif linker fragment (IRL). Amino acids 1–209 comprise both RBD1 and RBD2. RBD1 alone is composed of fragments of amino acids 1–112, while RBD2 is from amino acids located in positions 112–209. Glycine-rich domain contains amino acids numbered 209–296.

Results

The aim of the presented work herein was to isolate novel cDNA sequences with important functional implications in human pathology. Our efforts have been devoted to the cloning and subsequent tissue-specific gene expressions of numerous human RNPs from the hnRNP A/B family of proteins (Figure 1; ref. *15*). The objective was to search for molecular basis of autoimmunity by applying comparative analysis of the sequences of diverse autoantigens. In this context, evolutionary computation approach could give us major clues on how evolutionarily conserved mRNA transport machinery fails are linked to development of human autoimmune disorders. We were mainly interested in cDNAs, which might encode the yet undescribed hnRNP B2. The need for this was based on two observations. On the one hand, autoantibodies directed against hnRNP A2 crossreact with hnRNP B1 and hnRNP B2. Since hnRNP B1 is an alternatively spliced variant of hnRNP A2, this suggests that hnRNP B2 might be an alternatively spliced form of hnRNP A2/B1. However, no attempts to clone a cDNA encoding hnRNP B2 were successful so far. On the other hand, cDNAs closely related to hnRNP A1 and hnRNP A2 have been previously isolated from a human fetal brain library and from a *Xenopus laevis* library, respectively. Their close relationships with hnRNP A2 suggested that one of these cDNAs might actually encode hnRNP B2 (*15*).

To isolate the searched cDNA, human liver and brain cDNA expression libraries were screened by PCR using primers complementary to 5′- and 3′-untranslated regions of the FBRNP cDNA. The isolated sequence seemed to encode the full-length protein. Interestingly, however, it was not completely homologous to the FBRNP cDNA. Since the obtained new sequence shared close identity to the *Xenopus laevis* hnRNP A3 cDNA sequence (Entrez; accession number L02956), the protein was termed human hnRNP A3.

Nucleotide sequence comparisons between FBRNP, *Xenopus laevis* hnRNP A3 and our newly determined human hnRNP A3 proteins revealed that extensive sequence conservation exist in RNA-binding regions. The differences observed here were mainly at the third position of the codon triplet. The majority of sequence variations were seen at the Gly-rich domain, composed of amino acids at positions 211-373. These sequences were observed more at nucleotide level, as expected, compared to the translated protein sequences (Figure 2). Only protein sequences are shown for brevity.

Identification of various nucleic acid-binding domains of diverse hnRNPs was achieved by cloning and sequencing of cDNAs encoding these motifs. In general, all known human hnRNP proteins contain at least one RNA-binding module and one another auxiliary domain fragment. The RNA-binding motifs contain the RNP consensus sequences (CS-RBD), the RNA recognition motif (RRM; ref. *10*, *13*, *17*), the RNP-80 motif, the RGG box (*18*), and the KH domain (*19*). RNP domain is the most common feature in these RNPs. This domain is found in hnRNPs in various amounts, ranging from 1 (in hnRNP C) to 4 (*e. g.* in Poly A-binding protein; ref. *10*). Figure 1 shows the general modular structure of hnRNP A/B type of RNP particles. Their general structure is composed of two domains: the first 195 residues comprise the so-called UP1 domain, containing two canonical RNA-recognition motifs (RRM 1 and RRM 2), each of which is comprised from the conserved RNP-2 and RNP-1 submotifs. The Gly-rich C-terminal domain

Fig. 2 Deduced protein sequences of *Xenopus laevis* hnRNP A3, FBRNP, and human hnRNP A3. The sequences are aligned and displayed using the CLUSTALW programme, as mentioned in Materials and Methods. The alignment spans 373 amino acid residues. Dashes (–) represent apparent deletions or insertions; colons (:) mark semi-conservative amino acid exchanges; periods (.) depict more distantly located residues; and asterisks (*) show conserved residues. Amino acids of FBRNP and *Xenopus laevis* hnRNP A3, respectively, differing from human hnRNP A3, are shown in bold type.

comprises an RGG box and a nuclear localization motif. This motif contributes to protein-protein interaction patterns, as well as to subcellular localization (*18*). Two conserved solvents exposed Phe residues at the centre of the β-sheet in each RRM-contacted RNA. The least conserved 3′ loop in U1A is engaged in extensive RNA interactions, whose conformation changes upon RNA binding (*17*, *18*, *20*). The RNP domain interacts with a flexible single-strand RNA and the β-sheet provides a large surface for extensive interaction with nucleotides. Regions outside of the RRM may also play important roles in RNA binding (*10*). Identification of these motifs as RNA-binding domains has been used for prediction of this activity in various proteins of yet undescribed function possessing these domains.

As seen in Figure 2, the newly isolated human hnRNP A3 cDNA encodes a protein of 296 amino acids (a.a.). The calculated molecular weight of 32 kDa was also confirmed electrophoretically. Amino acids numbered 1–98 comprise RBD1, 112–209 comprise RBD2, 209–296 comprise the Gly-rich domain, while 99–111 contain inter-RRM linker (IRL) segments. To our surprise, comparison with the previously reported FBRNP cDNA sequence, encoding a 269 a.a. protein, revealed only 85% identity. Therefore, it was assumed that the currently presented cDNA encodes a novel and yet undescribed protein. Despite the close homologies among the RNA-binding regions (95%), remarkable differences can be seen in the C-terminal domain (Figure 2), where deletions and insertions are apparent. Thus, the FBRNP is 27 a.a. shorter and there are fewer conserved residues in the C-terminal part. Both cDNAs also show high homologies to the cDNA encoding hnRNP A3 from *Xenopus laevis* except for a stretch of 14 a.a. at the RBD1 N-terminal part. The RNA-binding regions of the three proteins are almost identical, while the auxiliary domains are less conserved. The *Xenopus laevis* protein shows 75% homology with FBRNP and 83% with the newly cloned human hnRNP A3 protein, respectively (Figure 2). The reduced length of the human hnRNP A3, as compared to the *Xenopus laevis* homologue (296 vs. 373 a.a.), is compatible with differences observed between *Xenopus laevis* and rat hnRNP A1, which are composed of 365 and 320 residues, respectively (*15*).

The assumption that the presented clone encodes the human counterpart of frog hnRNP A3 is further reinforced, when conserved residues in the auxiliary domains of the A/B polypeptides are considered. Thus, a stricking homology is apparent at the C-terminus, which is well conserved between frog and human hnRNP A3 with 15 of the last 18 a.a. being identical. Both proteins end with a triplet RRF, which is also present in hnRNP A1, but neither in FBRNP nor in hnRNP A2/B1. Moreover, a glutamic acid located at the boundary of RBD2 and the auxiliary domains of *Xenopus laevis* and human hnRNP A3 (a.a. position 207) is substituted by valine in FBRNP, a further indication that the novel protein is indeed the human counterpart of *Xenopus laevis* hnRNP A3. Interestingly, this result is reported also for the murine hnRNP mBx protein, which is highly homologous to both human FBRNP and *Xenopus laevis* hnRNP A3 (*21*).

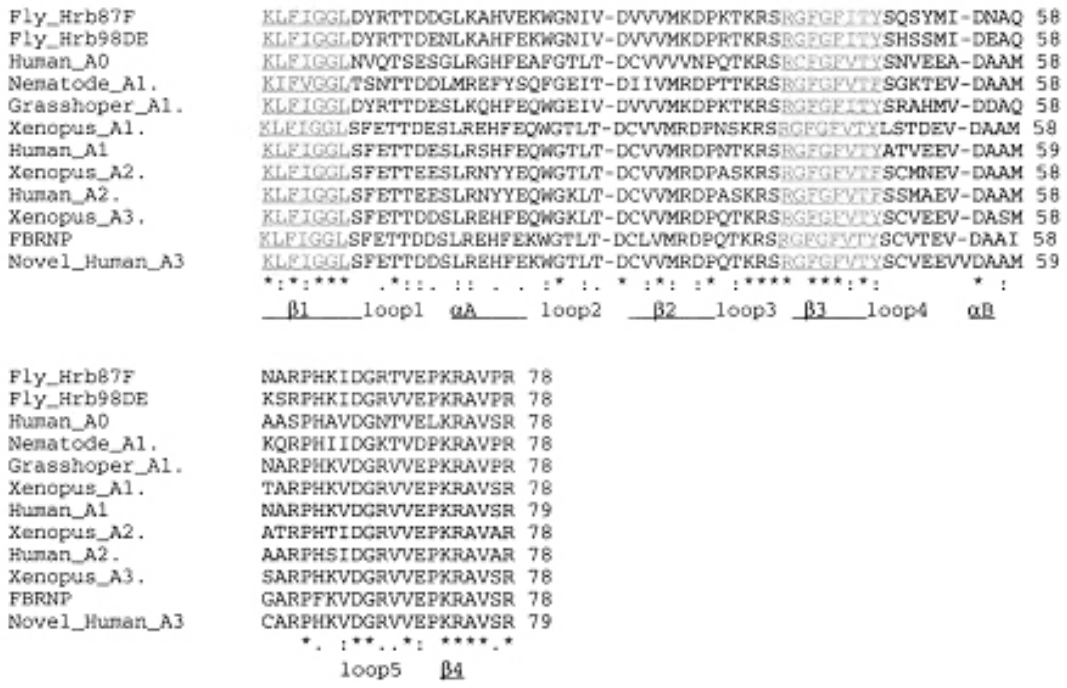
RRMs of hnRNP A/B type proteins share high degree of sequence homology. Most of the conformational differences between the two RRMs occur either at the C-terminal end of the α_B helices, or at the tip of loop 3, which is tilted and twisted in RRM1, as opposed to RRM2 (Figure 3). Loop 3 is the least conserved region among different RRMs. The RNP domain interacts with a flexible single-strand RNA and the β-sheet provides a large surface for extensive interaction with nucleic acids. The two aromatic side chains of RNP1 and one aromatic side chain of RNP2 provide a convenient template for base stacking of the RNA with neighbouring protein side chains, forming hydrogen bonds with the stacked bases. Although the major groove of the A-type RNA helix is too narrow and deep to provide a site for sequence-specific association, an RNA loop with exposed nucleotides provides a large surface for protein binding. However, the RNA-binding is not a unique feature of the β-sheets. Despite the evolutionary conservation of the RRMs, which are necessary for both general and sequence-specific nucleic acid binding, regions outside of the RRMs may also play important roles in RNA binding (*10*). For instance, the flexibility of the linker sequence connecting both RRMs, resulting from two pairs of Arg and Asp involved in IRL salt bridges and the ordered residues, creates a position which is highly probable for this sequence to be involved in direct RNA binding. Figure 3 shows the RRM sequences of 12 hnRNP A/B type of proteins, extending the previous data set of Mayeda *et al* (*16*). There is extensive sequence homology between these various proteins, which are conserved throughout evolution from insects to man. The sequences are 48–92% identical. Each of the RRMs forms an ungapped alignment with both RRMs of hnRNP A1. In all of the shown proteins, the RRMs are connected by a highly conserved IRL segments (Figure 1).

The evolutionary links between the RRMs of human hnRNP A3 and other hnRNP A/B proteins are depicted in a phylogenetic tree (Figure 4). The sequences of RRM1 and RRM2 cluster in two separate groupings, representing insect and vertebrate proteins. The patterns of branching for RRM1 and RRM2 sequences are nearly identical, except for the placement of the branch representing the minor human variant hnRNP A0. The almost identical pattern of branching for both RRMs suggests that these fragments have evolved in parallel. The division of insect and vertebrate RRM sequences takes place on two separate branches. Obviously, each of the insect and vertebrate proteins is equally distant from each other. Apparently, this fact indicates independent gene duplications of ancestral hnRNP A/B-like protein, in agreement with Mayeda *et al* (*16*).

Discussion

The independent evolution of individual tandemly repeated protein domains and their functional relevance has always being an intriguing issue. The RNP domain is commonly encountered scaffold among the nearly 350 different folds, that is, the favourable secondary arrangements of around 10,000 protein structures covered by the currently used databases (*22–25*). Comparisons of various protein structures, in combination with their nucleotide analyses serve as a clue for their evolution. The debate is whether these structures have evolved for achievement of specific functions or for thermodynamic stability and/or for kinetic folding reasons. The same structural topology can determine numerous activities, preserved through evolution. In biological systems, enzymes evolve by acquiring of new thermodynamic or kinetic properties by the already existing protein folds. Generally, only the overall folding pattern is conserved in protein groups, and as the sequence diverges, the structures deform. Among the widespread catalytic folds, such as the TIM barrel and the globins, the RRM domain attracts research efforts into delineating the advancements of evolution starting from all α-helical proteins *via* all β-sheet proteins and reaching a group of α/β-proteins. In this route, the α/β-barrel is the most frequently seen enzymatic fold and appears to be a selected topology for the directed evolution of new biocatalysts.

It is worth studying how the present α/β-barrels are linked evolutionarily to each other and through which way they have evolved from simpler ancestors. The topic becomes more attractive when these folds comprise a domain involved in a disease state, such as autoimmunity. The typical examples here are the va-



(A) RNA-recognition motif-1 (RRM1).

(B) RNA-recognition motif-2 (RRM2).

Fig. 3 Sequence alignments of sequences from the tandem RRM1 (A) and RRM2 (B) of 12 hnRNP A/B type of proteins. The conserved RNP-1 (left) and RNP-2 (right) submotifs are underlined. Conserved secondary structures are indicated below the RRM1 and RRM2 alignments. Dashes (–) represent apparent deletions or insertions, respectively; colons (:) mark homologous amino acid exchanges; periods (.) depict more distantly located residues; and asterisks (*) show conserved residues.

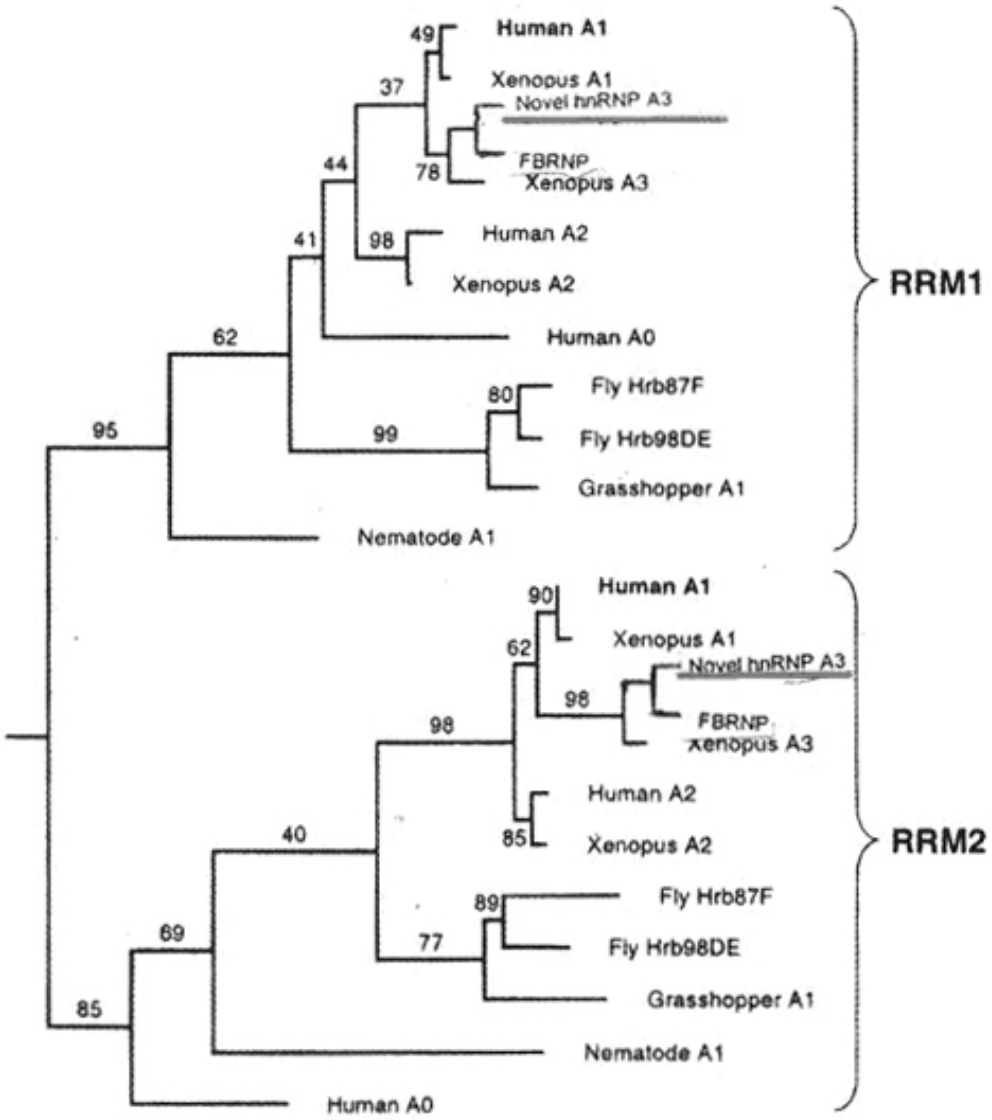


Fig. 4 Phylogenetic tree of the RRMs of hnRNP A/B-like proteins, based on Mayeda's data set (*16*). Numbers indicate the degree of conservation. The position of the newly cloned and sequenced human hnRNP A3 is underlined.

rious hnRNPs, acting as autoantigens (*9*, *11*). The rapidly growing information on structures and functions of numerous hnRNPs will be invaluable for understanding the pathological mechanisms by which these particles participate as targets of autoantibodies. Of particular interest is the determination of the role of RNA-protein folding patterns in development of disease-specific epitope formation. In this context, the evolution of selected nucleic acid-binding sequences and how this contributes to disease establishment remain unknown. Therefore, this study was devoted to presentation of evolutionary connections between the various hnRNPs and their RNA-binding domains. A phylogenetic tree is given to depict these motifs in the major group of these particles, including also the recently cloned and sequenced novel human hnRNP A3 (*15*), followed by remarks on the lastly selected sequences, relevant in diseases.

Figure 1 shows the general 2xRBD-Gly type of structure of the A/B type of human hnRNPs. The two canonical RRM-1 and RRM-2, containing the conserved RNP-2 and RNP-1 submoieties, and the RGG C-terminal part constitute the main domains. The translated sequences of the recently characterized human hnRNP A3 particle is presented herein (Figure 2), solely to demonstrate the high degree of sequence homology among the different RRMs. Its protein sequence is compared to that of other two members of

this protein family—*Xenopus laevis* hnRNP A3 and FBRNP for brevity. More detailed motif comparisons are given afterwards (Figures 3 and 4). The selected three members in Figure 2 are functionally related quite distantly from each other, each having a buried core tryptophan and there are disulphide bonds on the highly conserved IRL-segments (*16*). Usually, conservation of activity is not expected because of the evolutionary pressure. The sequence comparisons of the three closely related hnRNP A3 members showed that the general need for a large buried hydrophobic moieties is well satisfied. This is evolutionarily highly conserved due to thermodynamic stability requirements. This represents a possible case, where structural constraints make evolution towards a different conformation unlikely. The high level of conservation of these residues is related to the interlocking interactions and restraints of the loop connecting the β-sheets, by providing the huge hydrophobic template necessary for core packing. In addition, having large hydrophobic inner part and suitable hydroxyls capable of hydrogen-bond formation, these residues (*e.g.* tyrosines) could be conserved for structural stability needs. Performing amino acid mutations, or substitutions at the sites thus removing hydrophobic packaging interactions, would provide further evidence for the core packing of hnRNP globule. Moreover, the role of other conserved residues is worth to approach. These are buried in the compacted core and are expected to participate in overall folding of the scaffold. However, to verify whether this is the case, more mutational analyses, combined with thermodynamic, kinetic, structural and evolutionary computation data are needed.

Amino acid sequence comparisons of RRM structures (Figure 3) depicted the hydrophobic core residues, conserved among different RRMs. The inserted sequence of the human hnRNP A3 shared extensive sequence homology (up to 90%) existing between RRMs of all these numerous proteins, which are conserved throughout evolution. A phylogenetic tree (Figure 4) depicted the evolutionary history of the RRMs of human hnRNP A3 and other hnRNP A/B proteins. The objective was to deduce whether the observed patterns of conservation resulted due to structural features for this type of fold architecture, or alternatively,were the consequence of divergent evolution. Parallel symmetric groupings of RRM1 and RRM2 in two different clusters, which belong to insect and vertebrate members is well seen. With the slight deviation of hnRNP A0, the rest of the RRM branch sequences are almost identical. This branching pattern is an indication of parallel proceeding evolution, in accordance with relevant studies (*16*, *26*). Each of the members of these two subgroups are equally distant from each other. These results suggest that these branching patterns originated from an ancestral hnRNP A/B-like protein probably through independent gene duplications. Phylogenetic analyses of this sort provide evidence for evolutionary history and origin of the modular structure of hnRNPs and their contemporary functional implications. This approach helps to delineate whether RNA-binding (RBD) and protein-protein (auxiliary) domains have evolved in parallel or follow another evolutionary history. In this context evolutionary computation on every newly characterized hnRNP particle becomes a valuable prerequisite for understanding its nowadays functions. The sequence of human hnRNP A3 was inserted in the data set at Figure 4 and becomes a further support of the evolutionary trend of these RNP family of proteins, appearing also in previous studies (*16*, *26*). These earlier studies proved that the origin of hnRNPs is a consequence of independent gene duplication. In their landmark work, Fukami-Kobayashi *et al* (*26*) showed that the ancestral gene of the hnRNPs had have two RNA-binding domains even before the divergence of invertebrates and vertebrates and that it diverged to contemporary hnRNP genes while preserving the tandemly repeated structures. The origin of natural selection of particular nucleotide sequences and their conservation from primitive RNA-based catalysis to eukaryotic chromosome evolution remain to be elucidated (http://manske.virtualave.net/genetik/vorlesung_ws 99/teil1-3/rna_und_die_evolution.htm; ref. *27*, *28*). In the light of "primordial RNA catalysis doctrine", there is now sufficient data to prove the substitution of ancestral RNAs by more efficient RNA-protein complexes with subsequent replacement of the latter by proteins (*27*). By using molecular modelling approach, it has been recently shown (*28*) that in this transition position, the two RNA-binding sites of the simulated proto-protein interact with target sites to form a stable RNP complex. Since the selected proteins were unable to resolve misfolded RNA, *in vivo* selected protein associated with the two RNA-binding sites, indicating that the protein facilitates the correct folding of the ribozyme. This provides evidence that these modelled self-splicing RNA-protein complexes can be considered as a primitive form of splicing factors or RNA chaperones, which contribute to correct RNA folding (*28*). In terms of molecular evolution of primitive catalysts, these self-splicing intron assemblies possibly became splicing RNPs, suggesting that an active RNP particles could have originated from a primordial ribozyme (*28*). The evolutionary driving force for this is the small structural cost, through which the protein binding compensates for some folding deficiencies in the RNA (*29*). This serves as another proof that the evolution from RNA to RNP-determined catalysis

represents an evolutionary design against misfolding rather than for the maintenance of a protein-binding site. Our preliminary biophysical results (*30*) suggest that this is the case with human hnRNP A3, as well. The current study presents the sequence conservation of mainly hydrophobic residues of human hnRNP A3, which form the hydrophobic packing cores and are essentially and evolutionarily conserved in other RNA-binding domains in the hnRNPs studied (Figures 2–4). This indicates that the RNA-binding domains have evolved from a common hnRNP A/B-type proto-protein, supporting further previous models (*16*, *26*). Interestingly, while the majority of models of RBD protein family built up to now are based on divergent evolutionary principles, another view claims that this is open to doubt in cases of similar, but more distantly related functions and structures (*31*). Their model suggests that even though the ancient RBDs could have developed separately by divergent evolution, these RBDs have evolved conserved RNP motifs with a similar structure and function on similar surfaces. The authors emphasize that RNP motifs are conserved both in scaffold architecture and in function, which provides an intriguing case of convergent evolution. Further advancements in sequence comparison techniques will help to understand whether these functional sites have arisen multiple times during evolution (http://online.itp.ucsb.edu/online/infobio01/higgs1/pdf/higgs1.pdf; ref. *22–24, 32*). Regarding the evolutionary conservation of hnRNP domains, two possibilities exist for explaining their nowadays existence—they are either hyperadaptable, or they may have developed features required to perform vital cellular functions (*32*). Despite the case with cold shock domain (*31*), hnRNP A/B proteins appear to possess eukaryotic origins, because prokaryotic homologs could not be detected until now. All these RBDs have similar dimerization roles leading to similar three-dimensional architectures, suggesting that they may have arisen from a common ancestor, which have diverged in sequence afterwards (*32*). These domains are found also in chloroplast protein sequences, implying the link with endosymbiont hypothesis, and also implys that the multicellularity in plants and animals did not evolved independently (*26*, *31*, *32*).

Domain evolution of hnRNPs rises the question that whether they are consequences of continuous or discontinuous evolution. Fukami-Kobayashi *et al* (*26*) showed that the RBD fused with the SR-rich domain proceeded the divergence of splicing factors, and that these two domains have arisen together thus conserving the fused domain organization. The existence of Gly-rich domain (Figure 1), also supports the proposal of domain organization of the unique origin of hnRNPs, followed by duplication and divergence of RBDs and structural transitions of the auxiliary domain (*18*). This indicates that the auxiliary domain shared by functionally related RNPs diverged in parallel together with the RBDs (*26*). It is thus implied, that these ancestral repeats must have oligomerized afterwards to adopt a similar structure seen in contemporary homologs (*32*).

In the light of the fact that hnRNP A/B particles are considered as disease genes, an intriguing issue becomes the use of protein domain data of various hnRNPs to delineate the link and etiology of diseases. As stated earlier, the majority of these gene activities related to autoantigenic hnRNPs are covered by relevant databases. These databases are useful for predicting functions of newly sequenced pathogenic genes. Within this respect, identification of human paralogous disease genes generates further pathological gene candidates to be sequenced, because paralogous genes are frequently encountered as mutated versions in similar diseases (*32*). The sequence comparison, however, may be insufficient for deduction of its functional role for patients diagnozed with the suspected disease. Therefore, we followed the suggestion of Ponting, C.S. and Russell, R.R. (*32*) and combined our sequence determination of human hnRNP A3 with its gene expression profiles, its molecular interaction features (*30*), and its tissue-specific gene expression patterns (*15*). Interestingly, our molecular characterization of human hnRNP A3 showed that while the recombinant hnRNP A3 with its 296 a.a. migrates as expected as a 32 kDa protein on SDS-PAGE analysis, it is recognized by the patients' sera as a 50 kDa protein. This is attributed to alternative splicing or due to tissue-specific expression of a highly related yet unknown crossreactive protein. Surprisingly, neither the 50 kDa nor the 32 kDa protein was detected in HeLanuclear extracts, further supporting the assumption that hnRNP A3 is not ubiquitously expressed as hnRNP A1 or hnRNP A2. Northern Blotting analysis showed the variable expression patterns of hnRNP A3 mRNA in human tissues. Thus, it is highly expressed in spleen, ovary, small intestine, lung, liver, skeletal muscle, kidney and pancreas, whereas expression in thymus, testis, colon and peripheral blood was hardly detectable. Expression in prostata, heart, brain and placenta was low, but clearly detectable (*15*). In addition, circular dichroism and fluorescence spectroscopic measurements demonstrated that the human hnRNP A3 protein is a stable particle—the free energy of unfolding of the full-length hnRNP A3 is 58 $^{\circ}$C, as shown by both urea and temperature denaturation (*30*). This particle increases its apparent stability upon interaction with RNA fragments $_{\mathrm{r}}(\mathrm{UUAGGG})_4$ to higher temperature values. The high binding affinity to this repeat indicates its preference for association with purine-rich consensus sequences and target-

ing features towards deleterious G-tetrad structures. The latter fact suggests its active participation in alternative splicing—a common feature of well-known telomitic repeats interactions of hnRNP A1. This RNA fragment, hnRNP A3 particle recognition, can act by facilitating the splicing of alternative intron of the pre-mRNA. Since this represents a case of regulation of splice-site selection with further functional implication in human disease (*11*, *12*, *33*, *34*), epitope-mapping studies of human hnRNP A3 were carried out. hnRNP A3 possesses two major autoepitopes. The first one is comprised of both RBD1 and RBD2, and the second one is composed of RBD2 and certain parts of Gly-rich domain (*15*, *30*). This epitope recognition pattern differs from the epitope determined for the highly related hnRNP A1 and hnRNP A2, respectively. Thus, in hnRNP 2 the RBD2 was found to contain the major epitope, which was recognized by patients suffering from rheumatoid arthritis or systemic lupus erythenathosus. This particular region was also found to be essential for interaction with RNA and the patients' autoantibodies strongly inhibited RNA binding. As oppose to these, autoantibodies derived from patients diagnozed with mixed connective tissue disease recognized an epitope comprising RBD1 and RBD2. Interestingly, the major epitope of hnRNP A1 also comprises both RBDs, which were also targeted by patients with rheumatoid arthritis and systemic lupus erythemathosus. Taken together, these data confirm that rheumatoid arthritis, systemic lupus erythemathosus and mixed connective tissue disease are immunochemically linked by systemic autoimmunity to the functionally important RNA-binding regions of hnRNP A/B proteins. The observed trend brings the question of why so closely related and evolutionarily conserved RBD domains are differentially recognized by the autoantibodies. We proposed recently (*30*) the overall folding patterns of these domains, and the protein antigenicity arising from these patterns is the determining factor. The present study combined with previously determined sequences of U1A, hnRNP C, hnRNP A1 and *Drosophila* sex lethal (sxl) protein structures reveals that they have the same fold, but have different placing of the second and fourth β-strands. Individual RRMs have preferences towards various RNA sequences, due to differences in surface amino acids found outside the conserved RNP submotifs. In our opinion, this can explain the differential recognition of RRMs by patients' autoantibodies. In certain cases, the two RRMs somehow act in concerted fashion to give rise to the overall RNA- and antibody-binding characteristics, whereas in other cases the presence of only one RRM is sufficient for autoantibody and RNA recognition. Thus, RBD1 itself bound strongly to RNA fragments, however, the joining of RBD2 to RBD1 that increased the overall affinity indicats that RBD2 also strongly affects RNA-binding. This evidence indicates the possible existence of RNA conformation-specific autoantibodies.

Materials and Methods

Preparation of RNA-binding domains

To isolate the cDNA, human liver and brain cDNA libraries were screened by PCR using primers complementary to sequences in the 5′- and 3′-untranslated regions of the fetal brain (FBRNP) cDNA, as described (*15*). Screening of cDNA expression library resulted in isolation of a clone, which is a member of a 2xRNA-binding domain (RBD)—the glycine family of hnRNP proteins. A cDNA was isolated from a human liver library encoding a 296-a.a. human hnRNP A3 polypeptide, highly homologous to the fetal brain cDNA, as well as to hnRNPs A1 and A2. At the nucleotide level, the highest degree of similarity is shared with a cDNA from *Xenopus laevis*. The cDNAs encoding RNA-binding fragments of hnRNP A3 sequence were generated by PCR as deletion mutants, starting from the full-length human hnRNP A3. PCR mix was prepared by adding 660 μL water, 220 μL 10 mM dNTP mix (Pharmacia Biotech, Uppsala, Sweden), 220 μL 10× cloned *Pfu* polymerase buffer (Promega, Madison, USA), and 122 μL 25 mM $MgCl_2$ for twice. PCR primers used are shown on Table 1. 20 μL of each oligonucleotide (100 pmol/mL) was taken and added to 480 μL water. PCR reactions were run by taking 172 μL of the PCR mix (*i. e.*, 86 μL per reaction), and mixing this amount with 12 μL of both primers. The amplification reaction was performed in a BiometraR, TRIO-Thermoblock^{TM} PCR cycler. Denaturation was achieved by incubating at 94 $^\circ$C for 30 sec, while annealing temperature was 52 $^\circ$C for 1 min. Extension reaction was performed at 72 $^\circ$C for 1 min, 30 cycles. The amplified fragments were cloned into ligation independent cloning (LIC) vector (Novagen, Madison, USA).

DNA Sequencing

Clones containing the correct insert were sequenced afterwards. All the DNA sequencings were performed by the Vienna Biocenter oligo team (http://emb1.bcc.univie.ac.at/gem).

Evolutionary analysis

Nucleotides and deduced protein sequences of previously determined RNA-binding domains of RNPs from various species and human hnRNP A3

were aligned employing CLUSTALW programme (http://www2.ebi.ac.uk/clustalw/; ref. *22*), based on the experimental design of Mayeda *et al* (*16*).

Table 1 PCR Primers Used for Generation of RBD-Binding Domains of Human hnRNP A3 Protein

Primer	Sequence	Encoded fragment
1NA3LIC	5′ GACGACGACAAGATGGAGGTAAAACCGCCGCCTGGTCGC 3′	Full-length hnRNP A3
1CA3LIC	5′ GAGGAGAAGCCCGGTTTAGAACCTTCTGCTACCATATCC 3′	RBD1 and RBD2
CT111A3	5′ GAGGAGAAGCCCGGTTTAAACAGGTCTCTTTGGTTC 3′	RBD1
CT209A3LIC	5′ GAGGAGAAGCCCGGTTTACTGCATCTCTTGTTTAGA 3′	RBD2

Acknowledgements

I thank Dr. Richard Hrabal (Laboratory of NMR Spectroscopy, Institute of Chemical Technology, Prague, Czech Republic; http://www.vscht.cz/nmr) for the hospitality and generous financial support during my stay in his laboratory in Prague as a guest researcher and for the detailed supervision concerning our joint work on NMR studies of RNA-protein recognition and its significance in the control of stability of hnRNP particles.

References

1. Singh, R. 2002. RNA-protein interactions that regulate pre-mRNA splicing. *Gene Expr.* 10: 79-92.
2. Fromont-Racine, M., *et al.* 2003. Ribosome assembly in eukaryotes. *Gene* 313: 17-42.
3. Jensen, T.H., *et al.* 2003. Early formation of mRNP: license for export or quality control? *Mol. Cell* 11: 1129-1138.
4. Stutz, F. and Izaurralde, E. 2003. The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends Cell Biol.* 13: 319-327.
5. Reed, R. 2003. Coupling transcription, splicing and mRNA export. *Curr. Opin. Cell Biol.* 15: 326-331.
6. Hieronymus, H. and Silver, P.A. 2003. Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. *Nat. Genet.* 33: 155-161.
7. Jensen, T.H. and Rosbash, M. 2003. Co-transcriptional monitoring of mRNP formation. *Nat. Struct. Biol.* 10: 10-12.
8. Mili, S., *et al.* 2001. Distinct RNP complexes of shuttling hnRNP proteins with pre-mRNA and mRNA: candidate intermediates in formation and export of mRNA. *Mol. Cell. Biol.* 21: 7307-7319.
9. Krecic, A.M. and Swanson, M.S. 1999. hnRNP complexes: composition, structure and function. *Curr. Opin. Cell Biol.* 11: 363-371.
10. Varani, G. 1997. RNA-protein intermolecular recognition. *Acc. Chem. Res.* 30: 189-195.
11. Conrad, K., *et al* (eds.). 2003. *Autoantigens, Autoantibodies, Autoimmunity.* Pabst Science Publishers, Lengerich, Germany.
12. Monneaux, F. and Muller, S. 2001. Key sequences involved in the spreading of the systemic autoimmune response to spliceosomal proteins. *Scand. J. Immunol.* 54: 45-54.
13. Varani, G. and Nagai, K. 1998. RNA recognition by RNP proteins during RNA processing. *Ann. Rev. Biophys. Biomol. Struct.* 27: 407-445.
14. Markovtsov, V., *et al.* 2000. Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol. Cell. Biol.* 20: 7463-7479.
15. Süleymanoğlu, E. 2001. Cloning and characterization of human heterogeneous nuclear ribonucleoprotein (hnRNP) A3—a novel member of the hnRNP A/B family. Ph.D. Thesis, Medical Faculty, Vienna Biocenter, Uiversity of Vienna, Vienna, Austria (http://www.arcs.ac.at/dissdb/rn036425).
16. Mayeda, A., *et al.* 1998. Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. *RNA* 4: 1111-1123.
17. Hall, K.B. RNA-protein interactions. 2002. *Curr. Opin. Struct. Biol.* 12: 283-288.
18. Nichols, R.C., *et al.* 2000. The RGG domain in hnRNP A2 affects subcellular localization. *Exp. Cell Res.* 256: 522-532.
19. Grishin, N.V. 2001. KH domain: one motif, two folds. *Nucleic Acids Res.* 29: 638-643.
20. Waggoner, S.A. and Liebhaber, S.A. 2003. Regulation of alpha-globin mRNA stability. *Exp. Biol. Med.* 228: 387-395.
21. Plomaritoglou, A., *et al.* 2000. Molecular characterization of a murine, major A/B type hnRNP protein: mBx. *Biochim. Biophys. Acta* 1490: 54-62.
22. Lesk, A.M. 2001. *Introduction to Protein Architecture*, pp. 165-193. Oxford University Press, Oxford, UK.

23. Page, R.D.M. and Holmes, E.C. 1998. *Molecular Evolution: A Phylogenetic Approach.* Blackwell Science Ltd., Oxford, UK.
24. Johnson, M.S. and Lehtonen, J.V. 2000. Comparison of protein three-dimensional structures. In *Bioinformatics: Sequence, Structure and Databanks, Practical Approach* (eds. Higgins, D. and Taylor, W.). Oxford University Press, Oxford, UK.
25. Anantharaman, V., *et al.* 2003. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr. Opin. Chem. Biol.* 7: 12-20.
26. Fukami-Kobayashi, K., *et al.* 1993. Evolutionary clustering and functional similarity of RNA-binding proteins. *FEBS Lett.* 335: 289-293.
27. Podgornaya, O.I., *et al.* 2003. Structure-specific DNA-binding proteins as the foundation for three-dimensional chromatin organization. *Int. Rev. Cytol.* 224: 227-296.
28. Atsumi, S., *et al.* 2001. Design and development of a catalytic ribonucleoprotein. *The EMBO J.* 20: 5453-5460.
29. Garcia, I. and Weeks, K.M. 2003. Small structural costs for evolution from RNA to RNP-based catalysis. *J. Mol. Biol.* 331: 57-73.
30. Süleymanoğlu, E. 2003. On some aspects of RNA-protein folding patterns in ribonucleoprotein particles and their implications in human autoimmune diseases. *Comptes rendus l'Academie bulgare des Sciences* 56: 47-54.
31. Graumann, P. and Marahiel, M.A. 1996. A case of convergent evolution of nucleic acid binding modules. *BioEssays* 18: 309-315.
32. Ponting, C.P. and Russell, R.R. 2002. The natural history of protein domains. *Annu. Rev. Biophys. Biomol. Struct.* 31: 45-71.
33. Lipes, B.D. and Keene, J.D. 2002. Autoimmune epitopes in messanger RNA. *RNA* 8: 762-771.
34. Faustino, N.A. and Cooper, T.A. 2003. Pre-mRNA splicing and human disease. *Genes and Develop.* 17: 419-437.