RNA Biology

RNA Chaperones, RNA Annealers and RNA Helicases

Lukas Rajkowitsch, Doris Chen, Sabine Stampfl, Katharina Semrad, Christina Waldsich, Oliver Mayer, Michael F. Jantsch, Robert Konrat, Udo Bläsi & Renée Schroeder

To cite this article: Lukas Rajkowitsch, Doris Chen, Sabine Stampfl, Katharina Semrad, Christina Waldsich, Oliver Mayer, Michael F. Jantsch, Robert Konrat, Udo Bläsi & Renée Schroeder (2007) RNA Chaperones, RNA Annealers and RNA Helicases, RNA Biology, 4:3, 118-130, DOI: 10.4161/rna.4.3.5445

To link to this article: http://dx.doi.org/10.4161/rna.4.3.5445

Copyright © 2007 Landes Bioscience

Published online: 30 Nov 2007.

Submit your article to this journal

Article views: 613

View related articles

Citing articles: 123 View citing articles
Review

RNA Chaperones, RNA Annealers and RNA Helicases

Lukas Rajkowitsch
Doris Chen
Sabine Stampfl
Katharina Semrad
Christina Waldsich
Oliver Mayer
Michael F. Jantsch
Robert Konrat
Udo Bläsi
Renée Schroeder*

Max F. Perutz Laboratories; University of Vienna; Vienna, Austria

*Correspondence to: Renée Schroeder; Max F. Perutz Laboratories; Department of Biochemistry; University of Vienna; Dr. Bohrgasse 9/5; Vienna A-1030 Austria;
Email: renee.schroeder@univie.ac.at

Manuscript accepted: 12/19/07

Previously published online as a E-publication:
http://www.landesbioscience.com/journals/rnabiology/article/5445

KEY WORDS
RNA chaperone activity, RNA folding, RNA annealing

ACKNOWLEDGEMENTS
Work in the authors' lab is financed by the Austrian Science Fund (FWF) Special Research Program (SFB) F17 “Modulators of RNA Fate and Function” and by the FWF grant T261-B12 to K.S.

ABSTRACT

RNA molecules face difficulties when folding into their native structures. In the cell, proteins can assist RNAs in reaching their functionally active states by binding and stabilizing a specific structure or, in a quite opposite way, by interacting in a non-specific manner. These proteins can either facilitate RNA-RNA interactions in a reaction termed RNA annealing, or they can resolve non-functional inhibitory structures. The latter is defined as “RNA chaperone activity” and is the main topic of this review. Here we define RNA chaperone activity in a stringent way and we review those proteins for which RNA chaperone activity has been clearly demonstrated. These proteins belong to quite diverse families such as hnRNPs, histone-like proteins, ribosomal proteins, cold shock domain proteins and viral nucleocapsid proteins. DExD/H-box containing RNA helicases are discussed as a special family of enzymes that restructure RNA or RNP in an ATP-dependent manner. We further address the different mechanisms RNA chaperones might use to promote folding including the recently proposed theory of protein disorder as a key element in triggering RNA-protein interactions. Finally, we present a new website for proteins with RNA chaperone activity which compiles all the information on these proteins with the perspective to promote the understanding of their activity.

INTRODUCTION

RNA misfolding. Most RNAs rely on well-defined 3D structures for their functions, but folding of RNA molecules tends to be a difficult process. The multitude of possible RNA base-pairings together with the high stability of RNA duplexes can result in alternative secondary and tertiary structures that are thermodynamically as stable as the functional, native structure.1 Additionally, the RNA folding landscapes are not smooth but are interspersed with many intermediates, and some of these intermediates can represent kinetic traps that significantly slow down the folding process.2 This can be described in a kinetic partitioning model, in which only a small fraction of the RNAs folds directly into the native state whereas the remaining molecules are trapped in long-lived, misfolded conformations.3,4 In vitro, these barriers to efficient folding are very dominant, as has been demonstrated for the folding of the Tetrahymena group I intron ribozyme.2,5,7 In vivo, it appears that RNAs fold efficiently, and it is generally believed that specialized proteins guide RNAs to their native state and/or resolve kinetic folding traps.8 For example, it has been demonstrated that the T4 phage derived group I td intron requires translation for efficient folding in vivo. In the absence of translation, exon sequences base-pair with intron sequences forming a kinetic trap, which substantially slows down splicing.9 This kinetic folding trap can be resolved by proteins with RNA chaperone activity.10 Additionally, it was demonstrated that the Tetrahymena group I intron also misfolds in the cell, and that the molecules are partitioned into a native and an inactive population depending on the exon sequence context.11 Yet, in vivo folding pathways are not well studied, and little is known about how RNAs assemble into functional particles.12,13

We have only sparse knowledge about what happens to misfolded RNAs in vivo. It can be assumed that RNA quality control pathways are in place that recruit non-functional, misfolded transcripts for degradation. For example, in E. coli a mutant pre-tRNA is degraded by the exoribonuclease polynucleotide phosphorylase after polyadenylation of the RNA.14 In yeast, hypomethylated tRNA and an aberrant form of a 5S rRNA that misfolds are modified by a specialized polyA polymerase, Ttf4p, before degradation.15 In higher eukaryotes, the Ro autoantigen has been implicated in non-coding RNA quality control, as it specifically binds to a misfolded form of the
Over the last decade, a heterogeneous group of proteins has emerged, whose members do not share a common sequence, motif or fold but the property to destabilize RNA structures and to resolve RNA-RNA interactions. They promote folding of RNAs by accelerating the escape from kinetic folding traps and prevent RNAs from being trapped in non-functional conformations. This activity is termed RNA chaperone activity and is at the focus of this review (Table 1). Proteins with RNA chaperone activity belong to established protein families and have been assigned distinct roles in diverse cellular processes such as the regulation of transcription, RNP assembly and stabilization, RNA export, virus replication and histone-like nucleoid structuring. So far no protein has been characterized whose primary function is to resolve non-specifically misfolded RNAs in the cell, and the term “RNA chaperone” has been used to refer to proteins with RNA chaperone activity in general.

Many of these proteins have been characterized for their RNA binding properties and for their RNA chaperone activity in different assays, but we have only limited knowledge about their mode of action. Due to the lack of similarity between the individual protein families, the mechanisms they employ are likely to be diverse. However, some general principles can be derived from their shared features. Proteins with RNA chaperone activity bind RNAs only weakly and with low specificity, suggesting that the interaction of these proteins with RNA is transient and mainly of electrostatic nature. In line with this notion, for example a mutant version of the E. coli protein StpA with reduced RNA binding capacity displays a higher RNA chaperone activity than the wild-type, indicating that strong binding leads to RNA stabilization and is detrimental to RNA chaperone activity. The transient mode of action also entails that once the RNA is correctly folded, proteins with RNA chaperone activity are no longer required for the RNA to maintain its native structure. This has been shown for StpA and S12 in vitro assays by means of protein digestion between the RNA folding step and the activity assay.

Due to their weak and non-specific binding, RNA chaperones require an excess of protein over RNA, at least in RNA chaperone activity assays in vitro. However, if this coverage exceeds the optimal concentration, the function of the RNA can be impaired, probably because the bound protein prevents RNA interactions. Moreover, many of the proteins discussed in this review show a capability to oligomerize, such as the histone-like proteins H-NS and StpA as well as NCp7. While the dimer or oligomer could constitute the functional protein unit, homotypic protein-protein interactions could also be beneficial for RNA chaperone activity.

RNA chaperone activity does not require external energy input such as derived from ATP hydrolysis to resolve RNA structures and to displace RNA-bound proteins. Only few are processive, whereas most of the well characterized RNA helicases are non-processive “unwindases” such as DED1, elf4A, Cyt-19 and Mss116p (reviewed in ref. 36).

### Table 1: Definition of terms related to protein-assisted RNA restructuring

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA chaperone (protein with RNA chaperone activity)</td>
<td>A protein that binds transiently and non-specifically to RNA and resolves kinetically trapped, misfolded conformers. RNA chaperone activity entails the disruption of RNA-RNA interactions and the loosening of RNA structures. The interaction with the protein is needed for the unfolding of the RNA but not to maintain its structure. The protein does not require ATP-binding or hydrolysis for its activity. Examples: StpA, L1, S12, S199, S122.</td>
</tr>
<tr>
<td>RNA annealer (protein with RNA annealing activity)</td>
<td>A protein that accelerates annealing of complementary RNAs. It can act by binding to one or both RNAs thereby enhancing the local concentration and the probability of RNA-RNA interactions (molecular crowding, Example: Hfq). A protein with matchmaker activity also alters the structure of the RNA upon binding, thereby rendering it annealing-competent (Example: gBP21).</td>
</tr>
<tr>
<td>RNA helicase</td>
<td>A member of the DEd/Hbox protein family that utilizes energy derived from ATP hydrolysis to resolve RNA structures and to displace RNA-bound proteins. Only few are processive, whereas most of the well characterized RNA helicases are non-processive “unwindases” such as DED1, elf4A, Cyt-19 and Mss116p (reviewed in ref. 36).</td>
</tr>
<tr>
<td>Specific RNA-binding protein</td>
<td>A protein ligand (protein cofactor) that binds specifically to an RNA and thereby contributes to its native folding and thermodynamic stability. The continuous interaction with the protein is required to maintain the functional RNA structure. Examples: Cyt-18, CBP24, LtrA.</td>
</tr>
</tbody>
</table>

RNA chaperones appear to have diverse strategies to accelerate RNA folding, and the elucidation of these strategies is just at the beginning.

**RNA annealers.** In vitro annealing of short, unstructured RNAs occurs by itself, but this reaction can be accelerated by annealing-active proteins up to the diffusion limit. Molecular crowding might be one of the mechanisms proteins with RNA annealing activity (RNA annealers) employ, which results in an increase in the local RNA concentration and thereby enhances the probability of RNA annealing. This can be achieved by simultaneous binding of both RNAs. Longer RNAs, however, can form stable structures that have to be opened prior to hybridization with a complementary RNA, as has been shown in a systematic approach. Proteins with matchmaker activity can induce a conformational change in the RNA upon binding resulting in the exposure of residues relevant for intermolecular RNA interactions.

Similarly, the RNA chaperone activity of a protein can release an RNA from its folding trap before an actual annealing event. This activity is different from the above mentioned RNA annealing activities, which require continuous binding of the protein to the...
RNA until annealing has been completed. However, it can be difficult to discriminate between RNA chaperone and annealing activity experimentally, and a protein can have one or both of them.\(^{32,31,33,35}\) The special case of annealing of protein-bound guide RNAs with target RNAs and its most prominent example, human RISC, will be discussed in “Guide RNA targeted RNA annealing.”

**RNA helicases.** RNA helicases are thought to be involved in most if not all ATP-dependent processes involving rearrangement of structured RNAs (reviewed in refs. 36–38). They are RNA-dependent ATPases with high similarity to DNA helicases, and many RNA helicases can be grouped into the DEAD-box, DEAH and the DEExH subfamilies. In particular the action of non-processive DEAD-box proteins that behave like “unwindinges” has apparent similarities to RNA chaperone activity. RNA helicases, however, differ from RNA chaperones not only because they share common motifs between them but also because of their distinct mode of action. In contrast to RNA unfolding by transient binding of RNA chaperones, RNA helicases are ATPase-dependent enzymes that require external energy to unwind double-stranded RNAs, to remodel RNPs and to promote the splicing of large ribozymes. Importantly, while a number of RNA helicases also show RNA annealing activity no RNA chaperone activity (as defined in this review) has yet been reported for this class of proteins.

**Specific binding proteins.** Contrary to the action of RNA chaperones, proteins can also assist in RNA folding by interacting with RNAs specifically, recognizing and binding a distinct sequence or structure. These ligands can guide RNA folding through stabilization of the functionally active state or intermediate structures. Much is known about the interaction of these proteins with RNA, and many high-resolution structures of RNA-protein complexes are available, with the X-ray crystal structures of ribosomes being one of many invaluable sources of information.\(^{39}\) Furthermore, important advances are made examining the folding and assembly of domains of ribosomal RNAs with their specific proteins.\(^{40}\)

A well-studied example of a specific RNA binding protein is the group I intron splicing factor Cyt-18, whose binding to a defined structural element leads to a compaction of the intron structure and to tertiary structure stabilization.\(^{21,41,42}\) In another case, a protein ligand might stabilize RNA-RNA interactions such as kissing loop complexes resulting in extended duplex formation.\(^{43}\) Recently, a paper has been published, in which the encounter of the CBP2 protein with its specific target RNA is monitored by single molecule FRET (fluorescence resonance energy transfer). The interaction of the CBP2 protein with the intron RNA during its initial contact causes structural fluctuations within the RNA, resulting in short lived conformations, which are not detected during non-protein assisted folding.\(^{44}\) These initial movements of the RNA allow speculations about what happens when proteins with RNA chaperone activity contact RNAs for which they do not show specific binding affinity. One can easily imagine that proteins search RNAs for a specific ligand, and none is found. As a consequence they dissociate from the RNA, but during the encounter process the RNA and possibly also the protein have undergone slight dynamic movements that might have changed their conformations.\(^{22,35}\)

**ASSAYS FOR RNA CHAPERONE ACTIVITY**

The RNA chaperone activity of proteins can be determined using a wide variety of assays.\(^{26,45}\) They reflect the different properties of these proteins, and not all proteins are active in all assays. In a simple set-up, short RNAs are employed to monitor the RNA chaperone-mediated helix destabilization of double-stranded RNAs, which can result in strand dissociation and strand displacement (Fig. 1A). Similarly, the annealing of RNAs of varying length, sequence and structure has been used to examine the RNA chaperone activity of proteins. However, results from this kind of assay can be ambiguous since many publications lack a detailed investigation of the mode of action, which can involve non-RNA chaperone activity such as RNA crowding or matchmaker activity.\(^{31,35}\)

To monitor complex refolding events, group I intron cis- or trans-splicing or hammerhead ribozyme-mediated RNA cleavage can be assessed. Proteins with RNA chaperone activity increase the fraction of RNA molecules in the native, catalytically active conformation (Fig. 1B–D).\(^{18,46}\) In addition, new approaches to monitor single molecules include force spectroscopy and single molecule stretching using optical tweezers, and advances in time-resolved NMR suggest that this will become a powerful tool for measuring RNA folding in real-time.\(^{47,50}\)

These in vitro assays employ model RNA substrates or ribozyme systems, because RNA chaperones typically show non-specific RNA binding properties, and because the natural target RNAs might not be known. These assays are well suited for the general assessment of a candidate protein and to search for RNA chaperone activity mutants. However, comparing the activity of different proteins in a single assay harbors potential pitfalls, since typically purified proteins are tested with a unique set of RNAs and in the absence of their biological partners. Therefore, proteins have also been tested in vivo, mainly in a “folding trap assay,” which examines RNA chaperone dependent group I intron splicing in the cell\(^ {10,51}\) (Fig. 1E), and in a transcriptional assay that observes RNA chaperone assisted destabilization of a terminator stem (Fig. 1F).\(^ {52}\)

Some assays, especially those that measure helix destabilization, have often been performed using DNA as substrate and hence refer to this action as nucleic acid chaperone activity.\(^ {53}\) There are no systematic studies, which compare the activities of these proteins on RNA and DNA. We therefore concentrate here only on measurements performed with RNA, unless otherwise stated.

**CLASSIFICATION OF PROTEINS WITH RNA CHAPERONE ACTIVITY**

In this review, proteins with RNA chaperone activity are presented according to the family of their primary and well-defined function. Only those proteins for which extensive analyses have been reported will be discussed in this review, and hence, our protein listing is not complete. However, we will publish a more comprehensive compilation on our RNA chaperone activity website (www.projects.mpib.ac.at/rnachaperones), which we aim to update regularly, and we will also include evidence for proteins that did not show RNA chaperone activity in the respective assays.

**Heterogeneous nuclear ribonucleoproteins (hnRNPs).** Heterogeneous nuclear ribonucleoproteins are proteins that associate with RNA polymerase II transscripts (pre-mRNAs/hnRNAs), and in human cells more than 20 hnRNPs have been characterized to date. hnRNPs contain common RNA binding motifs like RGG boxes (arginine glycine boxes), RRMs (RNA recognition motives), hnRN K homology (KH)-domains and zinc finger domains.\(^ {54,55}\) Many hnRNPs shuttle between the nucleus and the cytoplasm and are involved in the regulation of gene expression at various levels like transcription, pre-mRNA splicing, 3'-end formation, mRNA...
Proteins That Restructure RNA

Figure 1. See legend, page 122.
Proteins That Restructure RNA

Figure 1. Assays for RNA chaperone activity. (A) RNA chaperone activity (RCA) on double-stranded RNA can result in helix destabilization (duplex unwinding), in full strand dissociation (RNA melting) or—in the presence of a competitor RNA—in strand displacement (strand exchange). (B) Pre-RNAs containing the thymidylate synthase group I intron have to fold correctly to undergo splicing. For the cis-splicing assay, the purified transcript is folded by heat-renaturing, and the reaction is initiated by addition of a guanosine cofactor. Proteins with RCA significantly increase the population of molecules with a catalytically active, splicing competent conformation. (C) To further strengthen the requirement for RNA-RNA interactions and correct folding, the pre-RNA can be transcribed in two pieces, which have to anneal for the formation of an active ribozyme. Trans-splicing is initiated by an exogenous guanosine cofactor that is ligated 5′ to the 5′-intron part in the first transesterification step. The second reaction step leads to ligated exons and intron release. Proteins with RCA facilitate this reaction, especially at low temperatures. (D) The hammerhead ribozyme assay monitors the cleavage of a substrate RNA that has to anneal to the ribozyme. RCA may resolve hammerhead ribozyme complexes and is beneficial for the release of the cleaved product and for multiple turnover. Thus, two activities—annealing and strand displacement—can be traced in one assay. (E) For the folding trap assay, a mutant of the thymidylate synthase gene containing a stop codon upstream from the 5′ splice site in exon 1 pre-RNA is expressed within E. coli cells. The premature stop codon prevents the ribosome from resolving an aberrant base pairing between 3′-terminal intron and exon 1 sequences, which precludes the folding of a native intron structure. Co-expression of RNA chaperones partly alleviates this splicing deficiency by resolving the misfolded structure. The RNA then folds correctly and the intron splices. (F) Transcription termination can be caused by the formation of a stem-loop in the nascent transcript, followed by a poly(U) stretch. This prevents the RNA polymerase from reaching a downstream reporter gene [chloramphenicol acetyltransferase, CAT], making the cells chloramphenicol sensitive. With RNA chaperones overexpressed in the cell, the terminator stem is ‘melted’, CAT is transcribed and the cells become chloramphenicol resistant (CmR).

Cold shock domain proteins (CSPs). At low temperature, the RNA folding problem becomes more dominant, because non-native structures, which represent kinetic traps, are more stable. Therefore, it is not surprising that many bacterial proteins, which accumulate upon cold shock, have RNA chaperone activity. Cold shock proteins (CSPs) are small proteins that contain a single nucleic acid binding domain, the cold shock domain (CSD). In E. coli nine CSPs have been identified, which are differentially regulated. The E. coli proteins CspA and CspE have been studied most carefully with respect to their RNA chaperone activities. They function as transcription antiterminators or translational enhancers at low temperature, where RNA secondary structures become inhibitory. An in vivo assay (Fig. 1F) detects read-through past a terminator stem, which was positioned upstream of a chloramphenical resistance gene. When CspA, CspC or CspE are overexpressed, read-through can be detected suggesting that these proteins resolve the terminator stem.

CspA is a small hydrophilic protein consisting of only 70 amino acids. The structure of CspA is almost identical to the cold shock domain and contains five antiparallel β strands forming a closed five-stranded β-barrel. Two RNA-binding motifs (RNP1 and RNP2) are located on the β2 and β3 strands, respectively (Fig. 2). The protein has an overall negative surface with a positively charged aromatic nucleic acid binding area. Binding of the RNA to the protein involves stacking of the aromatic side chains with the RNA bases, but mutation of three centrally located residues in CspE does not affect RNA binding. Most interestingly, the mutant lost the ability to melt nucleic acids, to antiterminate transcription, and to acclimate to cold.

The CSD protein structure motif is wide-spread. It was first detected in several proteins, which bind oligonucleotides and oligosaccharides and was hence termed the OB-fold. The E. coli ribosomal protein S1 consists of six copies of the CSD motif, which was then also named S1 motif. It shows RNA chaperone activity which is discussed below in the context of ribosomal proteins. The OB/CSD/S1 motif is also found in the E. coli translation initiation factor 1 (IF1), which displays RNA chaperone activity in several assays. IF1 is active in trans-splicing, it rescues the splicing of a misfolded intron in vivo, and it resolves a terminator stem in vitro and in vivo.

CspA has further striking similarity (43% sequence identity) to the cold shock domain of the Y-box factors, which are eukaryotic nucleic acid binding proteins involved in transcriptional and translational regulation. One well-studied member is the major messenger ribonucleoprotein particle protein p50, also known as YB-1, which is a general regulator of translation by modulating mRNA structural rearrangements and packaging. The protein promotes non-specific strand displacement and annealing of both
RNA and DNA. In plants, the wheat protein WCSP1 as well as CSDP1 and CSDP2 (ATCSP2) from *A. thaliana* were reported to complement a quadruple *E. coli* cspA, cspB, cspE and cspG mutant, and they can melt double-stranded DNA. WCSP1 was also shown to induce transcriptional read-through in vivo and CSDP1 enhanced RNase activity in vitro, but it would be interesting to analyze their activity in RNA chaperone activity assays requiring transient, non-specific binding to RNA. Interestingly, CspS from psychrophilic and from uncultured archaea (with 36–59% amino acid identity to *E. coli* CspA) as well as an archaeal structural CSD fold analog have very recently been shown to complement the *E. coli* quadruple csp mutant as well. While they still await further characterization regarding their RNA chaperone activity, these results indicate that functional cold shock domain proteins can be found in all domains of life.

**Ribosomal proteins.** In the ribosome the RNA moiety encompasses the catalytic activity necessary for peptide bond formation whereas the main function of the proteins lies within supervising correct assembly and providing a stable yet dynamic scaffold. Thus it was not surprising that many ribosomal proteins from the small and the large subunit of *E. coli* showed strong RNA chaperone activity in vitro. A decade ago, screening for *E. coli* factors that enhance group I intron splicing in vitro revealed that many ribosomal proteins of the small ribosomal subunit display this property, and S12 was the most effective in increasing splicing in vitro. It possesses a long unstructured N-terminal tail, locates to the interface of the 30S subunit close to the decoding site in an area otherwise devoid of protein. It was shown to have strong RNA chaperone activity both in vitro and in vivo. Ribosomal protein S1 is another well-studied small subunit protein with RNA chaperone properties. S1 is the largest ribosomal protein, and locates to the junction of head, platform and body of the 30S subunit where it is the only ribosomal protein that is not stably associated with the ribosome. It is involved in mRNA-binding to the small ribosomal subunit, it disrupts RNA secondary structures in vitro and induces structural changes in the mRNAs during translation initiation in vivo. In line with this role, S1 shows strong RNA strand displacement activities but no annealing activity in vitro.

Proteins from the large ribosomal subunit have been screened systematically for their RNA chaperone activity, and nearly a third of the 34 *E. coli* proteins were found to be active in a trans-splicing assay. Of these, L1 has been studied further, and its RNA chaperone activity was found to be conserved in all three domains of life. The X-ray structure of the ribosome revealed that L1 is one of the ribosomal proteins that folds into a compact globular structure. However, most other ribosomal proteins that show RNA chaperone activity in vitro have at least partially unstructured segments. L3 and L24 which both have long unstructured extensions show high RNA chaperone activity in vitro and are the only ribosomal proteins capable of initiating 50S ribosomal subunit assembly.

S1 has been repeatedly co-purified with the annealing-active *E. coli* protein Hfq, raising the possibility that they work complementarily, and S1’s interaction with the RNA polymerase might be part of the linkage between transcription and translation. Several other ribosomal proteins also have extra-ribosomal functions, and until today it has not been shown if the RNA chaperone activity of ribosomal proteins plays an exclusive role in ribosome assembly and during translation or if it might be involved in additional cellular processes.

**Histone-like proteins from bacteria.** The *E. coli* protein StpA has been identified as a suppressor of a splicing-defective mutant of the phage T4 thymidylate-synthase gene. StpA shows high homology (58% amino acid identity) to the histone-like *E. coli* protein H-NS, which is involved in nucleoid structure formation and acts as a global transcription regulator. H-NS and StpA are small basic proteins that preferentially bind to bent DNA, and they compact the *E. coli* chromosome and control the expression of a large number of genes.

StpA shows RNA chaperone activity superior to H-NS in a variety of assays in vitro and in vivo. The protein promotes strand displacement, trans- and cis-splicing of the td group I intron. It also rescues misfolded introns in vivo and is active in RNA annealing. Of all RNA chaperones, StpA’s activity has been studied to the greatest extent and structural evidence for its mode of action has been provided. The presence of StpA results in an increased accessibility of bases involved in tertiary structure elements within an intron RNA to the methylating agent DMS. By loosening the tertiary fold of the RNA, the RNA chaperone resolves kinetic traps and enables further conformational searches and refolding of the RNA into a splicing competent structure. Intron mutants with an already destabilized tertiary structure are sensitive to StpA, suggesting that the structural stability of the RNA determines whether StpA is beneficial or detrimental to folding. StpA binds only weakly to RNA with a preference for unstructured RNA, and a mutant protein with reduced RNA binding capacity has stronger RNA chaperone activity, indicating a reverse correlation between binding and unfolding activities. So far, no target RNA for StpA has been found despite the application of methods as sensitive as genomic SELEX, which is in line with the transient binding of the protein and its low affinity for RNA. In accordance with the definition of RNA chaperones, StpA can be removed after resolving the kinetic trap, and the RNA can proceed to fold into its native structure independent of the RNA chaperone.

StpA and H-NS have a bipartite structure with the N-terminal domain being involved in protein-protein interactions while the C-terminal domain mediates nucleic acid binding. The structure of the N-terminal domain of histone-like proteins has been determined using NMR and crystallography, and heteronuclear NMR spectroscopy revealed a defined but highly disordered, flexible chain for the C-terminal domain. Structure predictions indicate a low overall compactness of *E. coli* StpA, with the C-terminal domain harboring preformed regions exposed to the solvent. These regions of disorder are implicated to be sites of RNA chaperone activity and indeed, the C-terminal domain alone is capable of strand displacement, cis- and trans-splicing. It is the site of homo- and heterodimerization. StpA proteins with an N-terminal L30P mutation lose the ability of the wild-type protein to dimerize as well as to anneal two short RNAs and to bind them simultaneously. Therefore, dimerization via the N-terminal part and the ability of molecular crowding appear to be required for StpA’s RNA annealing activity.

**Viral nucleocapsid proteins (NCps).** The nucleocapsid (NC) proteins of retroviruses and retroelements are small basic proteins processed from the Gag polyprotein by viral proteases. In the context of the Gag polyprotein of HIV-1, the NC has 10-fold stronger nucleic acid binding affinity than the processed peptides NCp15, NCp9 and NCp7, but its nucleic acid destabilizing activity is negligible. This is consistent with the primary function of
Gag being a nucleic acid binding and packaging protein, while the processed nucleocapsids, which can also serve as nucleic acid chaperones, are essential for viral replication.\textsuperscript{110} HIV-1 NCp7 was among the first proteins to be reported to accelerate the annealing of an RNA substrate to the hammerhead ribozyme as well as to catalyze the dissociation of the RNA products from the ribozyme.\textsuperscript{118} NCp7 also very efficiently rescued the splicing of a group I intron mutant in vivo.\textsuperscript{10} The nucleic acid chaperone activity of the NC was shown to be the result of a combination of its strong RNA annealing and its duplex destabilizing activities.\textsuperscript{111,112}

Most NC proteins contain one or two CCHC zinc fingers which form a specific RNA-binding motif flanked by short regions rich in basic amino acids and proline.\textsuperscript{110} Binding of NC to single- and double-stranded nucleic acids has a strong electrostatic component due to the four positively charged amino acid residues of the N-terminus, and the aromatic residues of the zinc fingers are involved in partial stacking with unpaired bases.\textsuperscript{113,114} Monitoring single molecule DNA stretching and force induced melting, binding rate constants for nucleic acids to NC were determined, and the fast on/off rates were found to correlate with NC’s ability to promote strand annealing. Mutational studies on the zinc fingers revealed a good correlation between the ability of NC to accelerate strand annealing and its ability to rapidly bind and dissociate from nucleic acids.\textsuperscript{109}

The hepatitis C virus (HCV) is a positive strand RNA virus, which encodes a single polyprotein that is processed into many viral proteins. Among these is the structural core protein, which is thought to coat the RNA genome with several hundred copies. This core protein is the major antigen of HCV, and it was found to have DNA annealing and strand displacement activity. The protein also promotes RNA dimerization via kissing loop formation.\textsuperscript{115} Similarly, the Hepatitis delta virus (HDV) encodes a single protein, the hepatitis delta antigen (HDAg). It binds RNA, shows RNA-chaperone activity in a hammerhead cleavage assay, and it is important for viral replication and packaging. HDV RNAs of genomic and antigenomic strands each contain a ribozyme, the HDV ribozymes, which are essential for generating monomeric RNAs during replication. HDAg was shown to enhance the self-cleavage activity of these ribozymes.\textsuperscript{116}

Similar to retroviruses, retroelements or LTR-retrotransposons contain Gag structural proteins, which lack the CCHC-zinc finger RNA-binding domains but like the retroviral nucleocapsids are involved in reshaping the retroviral RNAs for replication. For example, the yeast TY1 element contains a C-terminal region, which anneals the primer tRNA to the 5’ PBS, and it mediates TY1 RNA dimerization and promotes cDNA synthesis.\textsuperscript{117} Also the Drosophila retrotransposon Gypsy contains a non-processed Gag protein, which has a C-terminal domain responsible for tRNA primer annealing and which is indispensable for cDNA synthesis.\textsuperscript{118} To which extent these proteins also have non-specific RNA chaperone activity remains to be demonstrated.

RNA ANNEALERS

The Sm-like protein Hfq. More than 30 years ago, the \textit{E. coli} host factor I/Q (Hfq) was first identified as being required for phage QB replication,\textsuperscript{119} but it was its role in cellular physiology and the characterization of a broadly pleiotropic phenotype of an \textit{E. coli} \textit{hfq} mutant strain that reignited interest in this highly conserved protein.\textsuperscript{120} Subsequently, Hfq was found to be involved in stability control of several mRNAs and small regulatory RNAs, in riboregulation of target mRNAs by sRNAs and to act as a virulence factor in several bacterial pathogens.\textsuperscript{120-134}

Hfq is a heat-stable and basic protein, and zone sedimentation demonstrated that it forms stable hexamers in solution.\textsuperscript{135,136} Electron microscopic studies of the \textit{E. coli} Hfq protein as well as crystal structures of Hfq homologs revealed that it has a hexameric ring-shaped structure and that it belongs to the large family of Sm- and Sm-like proteins.\textsuperscript{121,122,137} These proteins are involved in RNA processing in eukaryotes and bind to various RNAs, primarily recognizing short U-rich stretches, known as Sm sites.\textsuperscript{138} An A/U-rich region preceded or followed by a stem-loop structure has likewise emerged as a common RNA binding motif for Hfq.\textsuperscript{125,139-141} Currently, it is poorly understood how the Hfq-hexamer interacts with RNA substrates. The X-ray structure of \textit{S. aureus} Hfq complexed with a single-stranded oligoribonucleotide revealed that it binds to the protein in a circular conformation around the central basic cleft of the hexameric ring.\textsuperscript{137} Mutational studies as well as FRET assays suggest that one \textit{E. coli} Hfq-hexamer provides at least two binding surfaces, and that an Hfq-hexamer can bring together RNAs in a pair-wise fashion.\textsuperscript{31,142-144} A recent paper describes that in addition to the two binding sites in the inner rim cavity and on the distal site of the hexamer, the long unstructured C-terminal extension of the \textit{E. coli} Hfq is also required for riboregulation, possibly by providing an additional binding platform for mRNAs.\textsuperscript{142}

Most data on Hfq-RNA interactions come from studies on small non-coding \textit{E. coli} RNAs (ncRNAs) that interact with target mRNAs. Hfq has been shown to stimulate in vitro annealing of sp0r24 RNA with galK mRNA,\textsuperscript{121} of OxyS with hflA mRNA,\textsuperscript{122} of RyhB with sodB mRNA\textsuperscript{128,140} and of SgrS with ptsG mRNA.\textsuperscript{145} For some of these interactions, it was shown that Hfq is no longer required after the annealing event, discounting the possibility that Hfq stabilizes these RNA complexes.\textsuperscript{122,128} Hfq could induce structural changes in the RNAs upon binding, thereby mediate these interactions. Recent in vitro studies using RNase footprinting and circular dichroism techniques did not reveal structural changes in the sRNAs RyhB\textsuperscript{140} and DsrA,\textsuperscript{139,146} while a FRET-based assay detected a conformational change in the DsrA RNA upon Hfq binding.\textsuperscript{142} Hfq did induce structural changes in the 5’-untranslated regions of the RyhB target sodB mRNA\textsuperscript{140} and of the MicA target \textit{ompA} mRNA,\textsuperscript{123} which could facilitate the interaction of the mRNAs with their sRNAs. At least for \textit{ompA} mRNA it was shown that the structural changes induced by Hfq prevailed upon proteolysis of the protein.\textsuperscript{123} Hfq disrupts a preformed complex between a stretch of the rpoS 5’UTR and the 5’ part of DsrA in vitro,\textsuperscript{147} but it did not facilitate strand displacement of non-specifically bound RNAs.\textsuperscript{35} Although expression of Hfq rescued the splicing defect of an intron mutant in vivo,\textsuperscript{123} RNA chaperone activity for Hfq in defined in vitro assays with non-target RNAs remains to be demonstrated. The lack of a non-specific RNA dissociation capacity is most likely rooted in the apparent primary function of Hfq, which is the stimulation of stable RNA-RNA base-pairing to affect the translational output.

Moreover, recent reports about an ATPase activity of Hfq suggested that external energy sources might fuel Hfq’s activity.\textsuperscript{94} However, the preparation of Hfq that facilitated the dissociation of specifically bound RNA was determined to be ATP-free, and addition of ATP did not change Hfq’s inability to non-specifically displace RNA strands.\textsuperscript{55,147} Hence, further investigations will be required to corroborate this finding and to assign a function.
Clustering of basic amino acids occurs predominantly at the C-terminal end. The Ro autoantigen is the main Ro RNPs. These proteins are associated with rheumatic autoimmune diseases like Systemic Lupus Erythematosus and Sjögrens syndrome. Ro RNPs are composed of the 60 kDa Ro protein, small cytoplasmic RNAs of unknown function (so-called Y RNAs), the La autoantigen, and are probably temporarily associated with several hnRNP proteins like hnRNP L and hnRNP K. Most interesting is the fact that Ro is conserved among various species like nematodes, cyanobacteria, and eubacteria. D. radiodurans, an eubacterium, tolerates unusually high doses of UV irradiation and its Ro ortholog, the Rsr protein, associates with Y-like bacterial small non-coding RNAs. Rsr mutants are more sensitive to UV irradiation than the wild type. Since the eukaryotic Ro protein also binds misfolded non-coding RNAs, it has been suggested that Ro RNPs are involved in a special RNA quality control pathway.

Several high resolution X-ray crystal structures of the Ro autoantigen complexes with non-coding RNAs give a first insight into how Ro might recognize an RNA (Fig. 3). The 60 kDa protein is composed of two domains, a large α-helical domain with HEAT repeats and a von Willebrand Factor A domain, which resembles a von Willebrand Factor A domain and is a potential ligand binding site. The HEAT repeats form a ring-shaped arrangement with a central cavity. Ro 60 therefore resembles its shape the bacterial Sm-like protein Hfq, and it forms a ring similar in structure and size to the hexameric Hfq. Moreover Ro 60—like Hfq—accelerates RNA annealing but it is unable to promote cis- or trans-splicing and shows only minor activity in an in vivo assay.

Highly conserved residues within the cavity interact with single-stranded RNA, and a basic patch on the outer surface of the toroid binds the Y-RNA. More recently, a co-crystal of Ro 60 with a fragment from the 5S rRNA revealed yet another basic RNA-binding platform surrounding the central cavity. Ro 60 has an extended surface that can interact with misfolded RNAs with helices and a 3′ terminal single-stranded tail of at least five nucleotides. Due to its capacity to bind to many misfolded non-coding RNAs, it has been proposed that Ro 60 uses its extensive RNA-binding surface to scavenge RNAs that fail to bind their specific RNA-binding proteins.

Double-stranded RNA binding domain (dsRBD) proteins. The double-stranded RNA-binding domain (dsRBD) is a highly conserved protein motif of about 70 amino acids in length that can be found in proteins of different functions from E. coli to human. Prominent dsRBD-containing proteins include the RNase III family of proteins comprising Drosha and Dicer, the Staufen proteins involved in RNA localization, the double-stranded RNA-dependent kinase PKR, and the RNA-editing protein family ADAR, to name a few.

dsRBDs have been shown to exhibit strand annealing activity with RNAs that do not anneal by themselves. The Xenopus laevis protein Xlrbpa is the homologue of mammalian TRBP, a protein involved in RISC assembly, and the protein contains three dsRBDs that bind dsRNA with different affinities. Interestingly, the individual dsRBDs and their combinations have different characteristics in RNA binding and annealing, and these are not correlated. The full-length protein containing all three dsRBDs is the strongest RNA-binder but has only moderate strand annealing activity and shows no strand displacement activity (Rajkowitsch L, Jantsch MF, unpublished results). The isolated second dsRBD, in contrast, shows both strand annealing and displacement activities. A lack of correlation between overall RNA-binding and RNA annealing activity can also be seen for the three dsRBDs of the Xenopus RNA-editing enzyme ADAR1. Here the first dsRBD shows strong annealing activity but almost no RNA-binding activity in vitro (Rajkowitsch L, Jantsch MF, unpublished results). The second dsRBD, in contrast, displays little strand annealing activity but is a strong RNA-binder.

dsRBDs specifically recognize the A-form helix adopted by double-stranded RNA. The crystal structure of a dsRBD complexed with a synthetic double-stranded RNA demonstrates the lack of base-specific interactions between the domain and its bound RNA. Instead, the dsRBD interacts with the phosphate backbone where the characteristic distances of major and minor grooves are measured. Sequence-specific recognition of target RNAs can be achieved by a combination of different mechanisms. On the one hand, the helical structure of substrates can be disrupted by internal bulges leading to a bent helix that can be accommodated by some dsRBDs. On the other hand, some dsRBDs can specifically interact with a hairpin-structure where the terminal loop region of the RNA interacts with the most N-terminal RNA-binding surface of the dsRBD. Finally, multiple dsRBDs are frequently found in several proteins. Therefore, a combination of the two above-mentioned mechanisms can lead to an “induced fit” resulting in specific binding of a longer RNA harboring a combination of structural elements. Like other RNA-binding domains the dsRBD is basic in nature with a calculated pI ranging from 8.3 to 9.5, depending on the domain investigated. Clustering of basic amino acids occurs predominantly at the C-terminal α-helix which is also essential for RNA-binding. Interestingly, in some domains this C-terminal helix is poorly conserved and consistently shows only a weak interaction with RNA. Recent evidence indicates, however, that these poorly conserved domains might serve novel functions such as protein-protein interaction.
Guide RNA targeted RNA annealing. Short RNAs can form lasting complexes with proteins and direct the resulting RNP complex to target sequences, which subsequently leads to RNA annealing. The RNA-induced silencing complex (RISC) is such a complex, where an siRNA is tightly bound to an argonaute protein. Target recognition is facilitated by a non-specific interaction of the argonaute protein with single-stranded RNA regions.\textsuperscript{165} Formation of the correct siRNA—target RNA complex results in target RNA cleavage and consecutively, in repression of gene expression.

Another well-characterized example can be found in kinetoplastid organisms such as the African trypanosomes. Expression of mitochondrial genes requires uridylate residues to be inserted or deleted at defined sites in pre-mRNAs. These editing sites are specified in trans by guide (g)RNAs. In the matchmaker model of the gBP21-mediated RNA annealing, the guide RNA-binding protein gBP21 first forms a stable 1:1 complex with the guide RNA. Notably, the 5’ end of the guide RNA becomes unstructured and is presented in an annealing-ready state, and ionic bonds between the gRNA and the protein reduce the net negative charge of the RNA facilitating binding of the target pre-mRNA. Upon annealing of the two RNAs the protein dissociates due to reduced affinity for the annealed product.\textsuperscript{35} This model has been confirmed by the X-ray crystal structure revealing that one part of the gRNA is buried and the bases of another part of the gRNA, which needs to interact with the target RNA, are exposed to the solvent.\textsuperscript{166}

A third example for a protein-assisted RNA annealing process occurs during biogenesis of the small subunit in eukaryotes. Processing of the precursor rRNA (pre-rRNA) to release the 18S rRNA precursor requires the formation of short duplexes between the U3 small nucleolar RNA (snoRNA) and the pre-rRNA at multiple sites. In yeast, the interaction of these otherwise unstable base-pairings has been reported to be stabilized by the proteins Imp3p and Imp4p within the context of a large RNP complex, the small subunit processome (SSUP).\textsuperscript{167} Imp4p also appears to restructure the U3 snoRNA to promote annealing with the 5’ end of the 18S rRNA. In summary, these three examples use slightly different strategies, but they do have in common that one RNA is bound more specifically and the bases are displayed in a way that facilitates interaction.

**DExD/H-box proteins (RNA helicases)**

The term RNA helicase is typically used for a large and ubiquitous group of RNA-dependent ATPases belonging to the helicase superfamilies, including the known RNA helicases.\textsuperscript{37} Members of the DExD/H-box protein family are believed to catalyze most if not all ATP-driven conformational changes in complexes that contain RNA.\textsuperscript{168} While some RNA helicases have been assigned to this family solely because of sequence homologies, RNA unwinding has been shown for more than 60 proteins, and their distinct mode of action has been recently reviewed.\textsuperscript{36} Only some helicases unwind double-stranded RNA in an unidirectional, processive manner and are also capable of displacing RNA-bound proteins. Quite contrary, a large number of DExD/H-box proteins behave rather like “unwindases” with non-processive activity.

Well-documented evidence for unwinding of short RNA duplexes by RNA helicases comes from observations of eIF4A helicase activity. This DEAD-box protein is considered the “godfather” of RNA helicases (reviewed in refs. 169 and 170), and it is required to resolve inhibitory mRNA secondary structures prior to translation initiation. Although it can be stimulated by RNA as well as by the translation initiation factor eIF4B, its “unwindase” activity is inversely related to the RNA duplex stability, limiting thereby the length of the RNA region being unwound in a single ATP-hydrolyzing cycle.\textsuperscript{171} Moreover, the ADP-bound form of eIF4A has a lower affinity for RNA, which would facilitate non processive cycles of unwinding and enzyme dissociation. It is proposed that this local destabilization might be sufficient to cause strand dissociation in most natural helicase RNA substrates such as short intra- and intermolecular duplexes or stems.\textsuperscript{172}

Despite similarities between this local RNA unwinding activity and the RNA unfolding activity of RNA chaperones, RNA helicases can be assigned to a class of its own. While RNA chaperones by definition are active in a non-specific, transient way, many DExD/H-box proteins require special RNA features such as 5’- or 3’- overhangs or a defined secondary structure for efficient loading and ATP hydrolysis.\textsuperscript{173} They categorically require ATP for their unwinding activity and to facilitate the catalysis of natural substrates such as self-splicing introns. In vivo, folding of these large ribozymes requires protein factors, e.g., the RNA helicase Cyt-19, a protein from *Neurospora crassa*. Cyt-19 was isolated as a group I intron splicing factor that functions in concert with the specific RNA-binding protein Cyt-18.\textsuperscript{174,175} In vitro, Cyt-19 stimulates self-splicing of non-cognate introns under near physiological conditions, indicating a broader RNA binding property.\textsuperscript{176} Another RNA helicase, Mss116p from yeast mitochondria, promotes splicing of a similar range of group I and group II introns in an ATP-dependent manner in vitro.\textsuperscript{177} Interestingly, Mss116p requires ATP to stimulate splicing of the a57 group II intron, but the RNA helicase activity of the protein is not required for this reaction.\textsuperscript{178} Furthermore, Mss116p but not Cyt-19 promotes the splicing of a *N. crassa* LSU-ORF group I intron in vitro in absence of ATP and this activity is stimulated in the presence of ADP.\textsuperscript{177} RNA helicases obviously discriminate between RNA substrates, and they appear to stimulate folding not only by ATP-dependent unwinding of RNA but in some cases rather via stabilization of on-pathway intermediates.

These two RNA helicases also display RNA annealing activity (partly up to the diffusion limit), a property they share with many other DExD/H-box proteins such as DED1\textsuperscript{30} and Ddx42p.\textsuperscript{179} The RNA annealing activity appears to be independent from the unwinding activity, since it does not require ATP, and it has been linked to the occurrence of specific arginine/glycine-rich clusters and domains.\textsuperscript{30,177,180-182} Interestingly, an enhancement of the RNA annealing activities in the presence of ADP was shown for the human Ddx42p,\textsuperscript{179} the DEAD-box protein DED1 from *Saccharomyces cerevisiae*\textsuperscript{30} and was suggested for Mss116p.\textsuperscript{177} Most fascinating was the observation that DED1 has the ability to balance RNA unwinding with varying ratios of ATP to ADP. The unwinding activity increases with ATP concentration, whereas the annealing activity decreases, and DED1 further distinguishes between different RNA substrates. Thus this RNA helicase is able to regulate RNA remodeling for different substrates as a function of ATP and ADP concentrations.\textsuperscript{183}

**RNA chaperone activity associated with intrinsically unstructured protein domains**

Many of the proteins with RNA chaperone activity show a high level of intrinsically disordered regions.\textsuperscript{184,185} These unstructured domains differ from rigid three-dimensionally structured proteins...
Whether this domain shows intrinsic unstructuredness remains to be shown, but forecasting functional domains using structural predictions will prove to be very useful for the investigation of proteins with RNA chaperone activity.

**The RNA chaperone activity website.** The RNA chaperone activity website (www.projects.mfpl.ac.at/RNAchaperones) is a resource containing all relevant information on proteins that are known to have RNA chaperone activity. The website presents the data on three different levels. The first level introduces the classes of proteins with RNA chaperone activity as described in this review. The second lists all the proteins of one class, the species from which an RNA chaperone is coming from and the major function(s) of each protein with RNA chaperone activity. Links from this list lead to the third level, where detailed information can be found including synonyms, DNA and amino acid sequences, accession numbers, domain structure, 3D structure (if available), binding motifs, data on experimental evidence for RNA chaperone activity and references. In addition, metastructural predictions are given, which are based on the calculation of residue-specific compactness of the protein. For this, the structural complexity of an individual residue is predicted in the context of the 3D fold (surface exposure). Wherever possible, links to other webpages or databases are provided. The data, which are stored in a relational database (MySQL) is manually curated, will be regularly updated and is freely accessible to all users. In addition, researchers are invited to contribute to this resource by sending further data on already displayed RNA chaperone activity or on new entries. An appropriate input form will be available. By creating the RNA chaperone activity website we hope to provide a repository, which promotes research on this diverse family of proteins.

**References**


www.landesbioscience.com RNA Biology 127
Escherichia coli


167. Gercekzi T, Correll C. Imp3 and Imp4 mediate formation of essential U3-precuror RNA (pre-rRNA) dimers, possibly to recruit the small subunit processome to the pre-rRNA. Proc Natl Acad Sci 2004; 101:15301-6.