Nucleic Acids Research, 1999, Vol. 27, No. 19 **3811–3820**

Cloning and biochemical characterization of *Bacillus subtilis* YxiN, a DEAD protein specifically activated by 23S rRNA: delineation of a novel sub-family of bacterial DEAD proteins

Karl Kossen and Olke C. Uhlenbeck*

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA

Received June 15, 1999; Revised August 2, 1999; Accepted August 10, 1999

ABSTRACT

DEAD, DEAH and DExH proteins are involved in almost every facet of RNA biochemistry. Members of these protein families exhibit an RNA-dependent ATPase activity and some possess an ATP-dependent RNA helicase activity. Although genetic studies have identified specific functions for certain DEx$^{D}/_{H}$ proteins from which an RNA substrate can be reasonably inferred, only DbpA from *Escherichia coli* has been shown to exhibit significant RNA specificity *in vitro*. Here we describe the characterization of YxiN from *Bacillus subtilis*, the second DEx$^{D}/_{H}$ protein to show significant RNA specificity as an isolated, homogenous protein. The ATPase activity of YxiN, like that of DbpA, is stimulated by a 154 nt fragment of 23S rRNA. YxiN has a 2 nM apparent binding constant for this fragment, yet its ATPase activity shows 1800-fold RNA specificity. Along with the conserved motifs shared among all DEAD proteins, YxiN and DbpA have a conserved C-terminal extension. This extension is highly conserved in several additional DEAD proteins. We propose that the C-terminus identifies a protein sub-family whose members bind 23S rRNA and that proteins of this family are likely to function in rRNA maturation/ribosome biogenesis or an unappreciated aspect of translation.

INTRODUCTION

Members of the DEAD, DEAH and DExH (collectively DEx$^{D}/_{H}$) families of proteins are involved in most biochemical processes involving RNA. Recent data indicate that in yeast seven such proteins participate in pre-mRNA splicing (1), 14 have been identified in rRNA maturation/ribosome assembly (2) and two participate in translation (3,4). Additional DEx$^{D}/_{H}$ proteins have been identified in nuclear mRNA export (5,6) and mRNA degradation (7,8). The large number of DEx$^{D}/_{H}$ proteins in eukaryotes and the comparatively large number in bacteria suggests that there may be several additional processes which are facilitated by these proteins. Forty DEx$^{D}/_{H}$ proteins account for 0.5% of the open reading frames (ORFs) in yeast (2,9), an estimated 43 are found in *Caenorhabditis elegans* (10) and seven are present in *Escherichia coli*.

Despite the large amount of genetic data implicating the DEx$^{D}/_{H}$ proteins in biochemical processes involving RNA, their molecular role in these processes has remained elusive. Many of the aforementioned biochemical processes involve the rearrangement or disruption of RNA structure (11,12). This fact, in conjunction with biochemical studies which demonstrate that an increasing number of DEx$^{D}/_{H}$ proteins unwind duplex nucleic acids in an ATP-dependent manner (13–16), has led to the widely held hypothesis that the DEx$^{D}/_{H}$ proteins catalyze the isomerization of RNA structure (1,17). The recent work of Staley and Guthrie supports this conclusion by demonstrating that partial blockages in the pre-mRNA splicing pathway are introduced by the stabilization of dynamic RNA duplexes. Moreover, mutation of a spliceosomal DEAD protein exacerbates defects introduced in this manner (18).

DEAD proteins share eight highly conserved sequence motifs, which are flanked by variable N- and C-terminal extensions (2,19,20; Fig. 1A). The unique N- and C-terminal extensions of DEAD proteins appear to enable the specific functions of individual proteins (21). In the present paper we show that YxiN from *Bacillus subtilis* and several additional bacterial DEAD proteins share a 70 amino acid C-terminus similar to that of DbpA (Fig. 1B). The ATPase activity of YxiN, like that of DbpA, is specifically stimulated by a fragment of 23S rRNA. To the best of our knowledge, YxiN is only the second DEAD protein that exhibits significant RNA specificity as an isolated, homogeneous protein. We propose that the unique C-termini of YxiN and DbpA target these proteins to 23S rRNA, and that proteins bearing this domain form a novel sub-family of bacterial DEAD proteins. The presence of seven highly conserved basic amino acids in the C-terminal domain suggests that this extension may be involved in RNA binding.

MATERIALS AND METHODS

Database search and sequence alignment methods

The ORF for YxiN was identified using the BLAST search algorithm running on the SubtiList server (22). Searches against bacterial genome sequences were performed using

*To whom correspondence should be addressed. Tel: +1 303 492 6929; Fax: +1 303 492 3586; Email: olke.uhlenbeck@colorado.edu

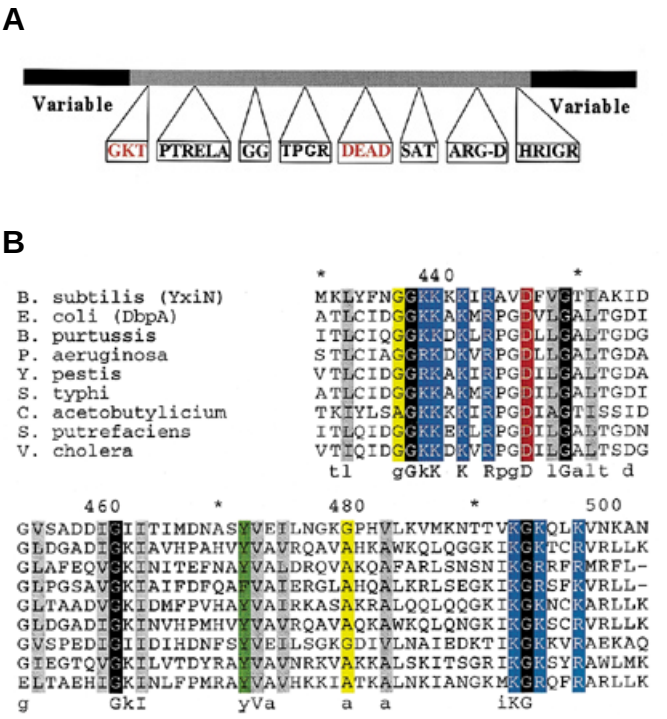

Figure 1. (**A**) Schematic depiction of motifs conserved among DEAD proteins. Shown is an idealized DEAD protein with eight conserved amino acid sequence motifs and variable N- and C-terminal extensions. Modified versions of the Walker A and B motifs, involved in the binding and hydrolysis of ATP respectively, are shown in red. Motifs shown are adopted from Luking *et al.* (12) and Fuller-Pace (20). For an alternative representation of these motifs see de la Cruz *et al.* (2). (**B**) Multiple sequence alignment of the C-terminal domains of YxiN, DbpA and seven putative homologs. The approximate location in the complete protein sequence is indicated above the alignment and a consensus is shown below. In 73 positions shown without gaps, 11 are identical and an additional 13 are conserved. Note the conservation of seven basic amino acids. For the alignment of all nine proteins there is 15% identity and 33% conservation. Conserved amino acid groupings used in this alignment and in the calculation of conservation are: E, D and C, red; G, A and S, yellow; H, K and R, blue; F, Y, W and H, green; I, L, V and A, gray; conserved glycines are shown in black. Less restrictive amino acid groupings including hydrophobic, small and polar were not considered.

either the TFASTXY algorithm running at the Pittsburgh Supercomputing Center or the gapped BLAST algorithm at The National Center for Biotechnology Information (23,24). Preliminary sequence data were obtained from The Institute for Genomic Research Website at http://www.tigr.org . Multiple sequence alignments were produced using MSA (25). Alignments were manipulated using GeneDoc, provided courtesy of Karl B. Nicholas (http://www.cris.com/%7EKetchup/genedoc.shtml). Structural predictions were performed on the UCLA-DOE fold recognition server (http://fold.doe-mbi.ucla.edu/) (26).

Cloning of *yxiN* using two potential start codons

The *yxiN* ORF contains two potential start codons separated by 33 nt (27). The correct start codon could not be determined *a priori*, so two clones were created for the inducible expression of either protein in *E.coli*. The longer gene was amplified via PCR using the upstream primer TATCGATCACGCCGTCTGCATATGAGAACATCTGCAAGAGACAGGAG and a common downstream primer (CTCCAGTGGATCCTTCATCATTTATTCGCTTTATTCACCTTCAGCTG). The shorter gene was amplified using the upstream primer CGCGACGCGCATATGAGTCATTTTAAAAACTATCAAATCAGTCATGAC and the common downstream primer. In each case pelleted cells of *B.subtilis* strain 168 (Bacillus Genetic Stock Center, Columbus, OH) were used directly as a template for PCR. Both primer sets introduce *Nde*I and *Bam*HI restriction sites, which were used to clone into the pET-3a expression vector. Clones were sequenced using T7 initiator and terminator primers, as well as three internal primers.

Isolates of two PCR reactions deviate from the published gene sequence at position 1091. The published gene sequence codes for the shorter of the two proteins (28,29) and identifies this region as having a series of three cytosines followed by five adenosines. We find that this region consistently amplifies as two cytosines followed by six adenosines. As we have well-resolved sequences of this region from multiple primers and independent PCR amplifications, we conclude that this position is properly identified as an adenosine. This change results in the substitution of a glutamine for a proline at position 364 in the short version of YxiN.

Expression and purification of YxiN

YxiN was expressed and purified by a method similar to that employed in the purification of DbpA (30). Briefly, BL21 (DE3) pLysS cells were transformed and plated on LB agar containing 50 µg/ml ampicillin and 30 µg/ml chloramphenicol. An individual colony was used to inoculate a 30 ml overnight culture in LB medium, which was subsequently diluted into 3 l of YTG medium containing the same antibiotics. This culture was grown at 37°C to an OD_{600} of 0.2, then transferred to a 25°C shaker and allowed to cool for 1 h prior to induction with IPTG (0.5 mM). The induced culture was grown for 5 h, then pelleted and stored frozen. Cells were resuspended in 120 ml of cold lysis buffer (50 mM Tris, pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 1 mM benzamidine and 1 mM PMSF). The cell suspension was then sonicated, and cleared by centrifugation at 13 000 *g* for 10 min. The soluble fraction was centrifuged for an additional 1 h at 186 000 *g*. The supernatant was collected and nucleic acid was precipitated by addition of a 1/10 vol of 2% poly(ethyleneimine) in 50 mM Tris, pH 7.5, and 200 mM NaCl. A white precipitate was removed by centrifugation at 46 000 *g* for 15 min. Crude YxiN was precipitated in 70% $(NH_4)_2SO_4$ followed by centrifugation at 17 000 *g* for 30 min. Precipitated proteins were resuspended in 10 ml of buffer A (20 mM MOPS pH 7.5, 200 mM NaCl, 1 mM DTT and 10% glycerol) and dialyzed against 3 l of buffer A plus 1 mM PMSF. The protein was purified on a Pharmacia FPLC system using a HiLoad SP Sepharose column. Loosely bound proteins were eluted in ~2 void volumes of buffer A. Proteins retained by the column were eluted with a linear 200–1000 mM NaCl gradient. The elution of YxiN at 570 mM NaCl was partially overlapped by an 80 kDa contaminant. This contaminant was effectively removed by diluting fractions enriched in YxiN with buffer A and running them over the same column a second time. YxiN was purified further with a Superdex 75 gel filtration

column equilibrated in buffer A supplemented with NaCl to 0.5 M. Both the long and short versions of the protein were purified effectively by this method, and each ran as a single band on a Coomassie stained protein gel. The protein was concentrated by dialysis against a 10% solution of polyethyleneglycol (≥35 kDa; Fluka) in 20 mM MOPS pH 7.5, 200 mM NaCl and 1 mM DTT. The glycerol concentration was brought to 55%, prior to storage at –70°C. Protein concentrations assume extinction coefficients of 22 810 and 21 090 M^{-1} cm^{-1} at 280 nm for the long and short forms of the protein, respectively (31).

RNA substrates

Yeast total RNA, yeast tRNA, poly(A) and *E.coli* rRNA were purchased from Boehringer Mannheim. Other rRNAs were prepared by *in vitro* transcription with T7 RNA polymerase, purified, and stored as described previously (30,32). *Escherichia coli* 23S rRNA was transcribed from plasmid PCW1, following linearization with *Afl*II (33). *Escherichia coli* 23S rRNA domain V (nt 2040–2625) was transcribed from *Bam*HI-linearized plasmid ECV.7, a generous gift of Rachel Green (Johns Hopkins University). Nucleotide 2041 of the domain V construct was changed to a guanosine to provide a strong start site and run-off from the *Bam*HI cut site introduces four non-native nucleotides (CUAG). A similar construct was created for the transcription of *B.subtilis* 23S rRNA domain V (nt 2069–2653). This sequence was amplified from plasmid pDK105, which codes for *B.subtilis* 23S rRNA domains IV and V (34), and cloned into pUC19 for transcription following *Bam*HI digestion. Two guanosines are introduced at the 5'-end of the transcript and four non-native nucleotides associated with the *Bam*HI cut site (GAUC) are found at the 3'-end. *Escherichia coli* 23S rRNA domains I–IV were amplified via PCR from plasmid PCW1. The primers used in this amplification introduced *Bam*HI and *Hin*dIII sites for cloning into pUC19. Four non-native nucleotides associated with the *Hin*dIII cut site (AGCU) are introduced into the resultant transcript. *Escherichia coli* 23S rRNA domain VI was cloned via an analogous method and can also be transcribed following linearization with *Hin*dIII.

Nucleotides 2454–2606 of *E.coli* 23S rRNA were previously cloned between the *Eco*RI and *Xba*I sites of pUC19 (30). Following linearization with *Xba*I, this clone can be used to transcribe nt 2454–2606 of *E.coli* 23S rRNA with three non-native nucleotides at the 3'-end. The corresponding secondary structural element in *B.subtilis* 23S rRNA is composed of nt 2481–2634. This sequence was amplified from pDK105 with the introduction of a T7 promoter and cloned between the *Eco*RI and *Xba*I sites in pUC19. *Escherichia coli* 23S rRNA nt 2503–2583 were transcribed from plasmid A1, provided by Randall Story (Caltech). The aforementioned sequence is cloned between the *Kpn*I and *Pst*I sites of pUC19 and can be transcribed following linearization with *Bsp*120I. The 2503 position has been changed to a guanosine to provide a strong transcription start and the transcript is terminated by two non-native cytosines. The corresponding secondary structural element in *B.subtilis* 23S rRNA (nt 2531–2611) was cloned and transcribed in a similar fashion.

The Group I intron L-21 was transcribed from plasmid pTL-21 following linearization with *Sca*I (35).

Coupled spectroscopic ATPase assay

The ATPase activity of YxiN was monitored using a coupled spectroscopic ATPase assay (30,36). The conditions are identical to those used in the study of DbpA, except that the KCl concentration of the assay buffer was raised to 175 mM. The resultant buffer contains 50 mM HEPES-K pH 7.5, 175 mM KCl, 10 mM $MgCl_2$, 100 μM DTT, 200 μM NADH, 1 mM phospho-(enol)pyruvate, 13 μg/ml lactate dehydrogenase and 23 μg/ml pyruvate kinase. Kaleidagraph (Synergy Software) was used to fit plots of A_{338} versus time and rates were calculated using an extinction coefficient of 6220 M^{-1} cm^{-1} for NADH. A buffer blank taken in the absence of the variable component (RNA or ATP) was subtracted from all data points prior to analysis. All kinetic data were collected at 37°C on a Cary 1E multicell UV-Vis spectrophotometer.

The ATP kinetic parameters, $K_{m,\ ATP}$ and k_{cat}, were derived from titrations of 10 nM enzyme with ATP·Mg in the presence of saturating concentrations of RNA. Plots of rate versus ATP concentration were fitted using the Michaelis–Menton equation. All reported ATP kinetic constants are taken from the average of at least two independent determinations. Variance in ATP kinetic parameters was never >15%.

RNA kinetic parameters were derived from titrations of 10 nM enzyme with RNA in the presence of 5 mM ATP·Mg. As the enzyme concentration often exceeded the $K_{app,\ RNA}$, plots of ATPase rate versus RNA concentration were fitted using a fully general kinetic equation (37):

$$v = \frac{V_{max}}{2[E_t]}\left([E_t] + K_{app,RNA} + [S_t] - \sqrt{([E_t] + K_{app,RNA} + [S_t])^2 - 4[S_t][E_t]}\right) \quad \mathbf{1}$$

where v is the observed rate, V_{max} is the maximal rate, $[E_t]$ is the concentration of YxiN, $[S_t]$ is the total concentration of the activating RNA and $K_{app,\ RNA}$ is the apparent binding constant of this RNA. The equation takes into account the fact that the added RNA ($[S_t]$) will not accurately reflect the concentration of free RNA, as it is tightly bound by relatively high concentrations of YxiN. Reported values of k_{max} are equal to $V_{max}/[E_t]$.

Data were fitted to equation **1** by non-linear regression using Kaleidagraph, with a full set of partial derivatives (i.e. partials with respect to V_{max}, $[E_t]$ and K_{app}). The ability of the program to fit model data was tested rigorously and it was determined that the program converged on the correct value provided that the error in duplicate data points did not exceed 10%. Actual deviation of duplicate data points, calculated for multiple independent data sets, averaged no more than 4%. Reported RNA kinetic parameters represent the average of at least two data sets, with the exception of constants for weak activators of ATP hydrolysis, which occasionally represent single determinations (see below). The variance in kinetic parameters derived from independent data sets was ~27%. Although this variance is high, no attempt is made to compare closely related values of k_{max} or $K_{app,\ RNA}$.

RNAs which weakly stimulated the ATPase activity of YxiN failed to yield standard saturation plots when the enzyme was titrated with RNA. In these cases the enzyme concentration was increased to 70 nM and titrated with large quantities of RNA. After subtraction of the RNA-free background rate, data were plotted as ATPase activity (ATP min^{-1} $YxiN^{-1}$) versus RNA concentration. Activity was a function of added RNA and the slope of a linear fit was taken as k_{max}/K_{app}.

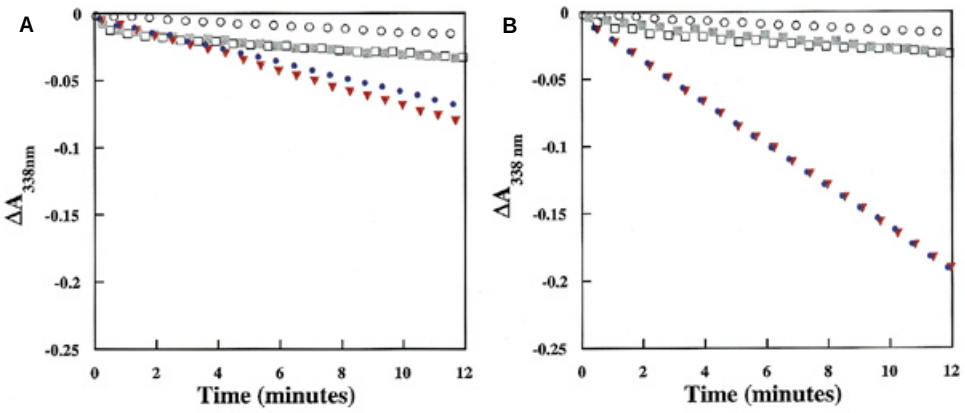

Figure 2. ATPase activity of the long (**A**) and short (**B**) versions of YxiN in the presence of several RNA cofactors. ATPase activity was monitored using an enzymatically coupled ATPase assay, in which the hydrolysis of ATP leads to the concomitant oxidation of NADH and loss of absorbance at 338 nm. Open circle, no protein control; open square, no RNA; gray square, poly(A); blue circle, *B.subtilis* domain V RNA; red triangle, *E.coli* rRNA.

The variance of $K_{app, RNA}$ with KCl concentration was measured at 125, 175, 250 and 350 mM KCl. *Bacillus subtilis* 2481–2634 nt RNA (154mer) was used in these experiments. A YxiN concentration of 2 nM was used for the first two KCl concentrations, but was subsequently increased to 5 nM. Kinetic curves were fitted as described for $K_{app, RNA}$ above.

The background rate of ATP hydrolysis was determined in standard buffer with 5 mM ATP·Mg, varying the YxiN concentration between 0 and 2.5 μM. The ATPase rate was a linear function of enzyme concentration and the background rate of ATP hydrolysis was determined from the slope of a plot of ATPase activity versus concentration of YxiN.

RESULTS

Identification of potential homologs of DbpA

DEAD proteins share a set of eight conserved amino acid sequence motifs. Variable extensions found to either side of these conserved motifs are thought to facilitate the unique functions of individual DEAD proteins (21). The DEAD box protein DbpA is unique among the known DEAD proteins in that its ATPase activity is specifically stimulated by bacterial 23S rRNA. DbpA has a short N-terminus with 25 amino acids preceding the first conserved sequence motif. A C-terminus extending 120 amino acids beyond the last characteristic DEAD motif was, therefore, considered most likely to be responsible for the RNA specificity of DbpA. A series of database searches conducted in an attempt to find DEAD proteins with a C-terminus similar to DbpA identified a similar domain in the hypothetical protein YxiN from *B.subtilis*. The cloning, expression and characterization of YxiN allowed us to test the hypothesis that the C-terminal domain defines the unique RNA specificity of DbpA.

Cloning, expression and purification of YxiN

The gene for the expression of YxiN contains two potential start codons, which produce proteins differing in length by 11 amino acids. Based on the conservation of Shine–Dalgano sequences, the internal AUG appears to be the correct start codon (28). However, Fuller-Pace and co-workers previously noted that expression of DbpA initiates from a GUG start codon 75 nt upstream of an in-frame AUG and that the shorter form of the protein is inactive (38). Because of the ambiguity in identification of the proper start codon for YxiN, and the significant effect of start site selection noted with DbpA, both long and short versions of the gene encoding YxiN were cloned via PCR into the pET-3a expression vector. This vector allows for inducible overexpression in *E.coli*. Both short and long versions of YxiN expressed well in BL21 (DE3) pLysS cells and could be purified to ≥95% homogeneity.

ATPase activity and RNA specificity of YxiN

Stimulation of the ATPase activity of YxiN by a number of RNAs was examined using an enzymatically coupled spectroscopic assay (30,36). In this assay the hydrolysis of ATP is coupled to the oxidation of NADH with a concomitant loss of absorbance at 338 nm. Figure 2A and B shows the response of either the long or short version of YxiN to various RNA cofactors. It is clear that 23S rRNA and transcripts of domain V can stimulate the ATPase activity of YxiN. RNA homopolymers, yeast total RNA and calf thymus DNA have no stimulatory activity. The results shown in Figure 2 demonstrate that the long form of YxiN is ~2-fold less active than the short form. Moreover, when stored under identical conditions the long form of the protein was unstable relative to the short form. The shorter form of YxiN was, therefore, used in all subsequent experiments.

The RNA-stimulated ATPase activity of the short form of YxiN was examined thoroughly, in an attempt to quantitatively define the RNA specificity and kinetic parameters of this activity. In this analysis, the enzyme was treated as an ATPase with an essential RNA cofactor. This treatment is technically correct, as we have no evidence to show that the enzyme alters the RNA, whereas ATP is clearly a substrate. Moreover, the rate of ATPase activity in the absence of RNA is exceedingly low (0.03 ATP min^{-1} $YxiN^{-1}$). Treating the RNA as an essential

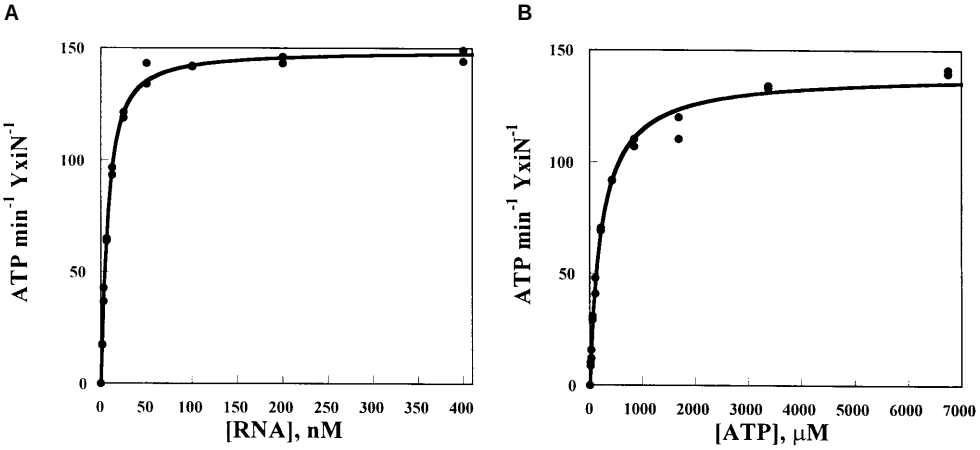

Figure 3. ATPase activity as a function of added RNA (**A**) or ATP (**B**). The factor that is not varied is present at a constant, saturating concentration. Plot (A) is the fit to a fully general kinetic equation, as described in Materials and Methods, and yields values for k_{max} and $K_{app, RNA}$. Plot (B) is fit to the Michaelis–Menton equation and yields values of k_{cat} and $K_{m, ATP}$. *Bacillus subtilis* domain V RNA was used as a cofactor in each case.

Table 1. Comparative steady-state kinetic parameters for the interaction of YxiN and DbpA with various RNA cofactors

RNA	YxiN		DbpA[a]	
	k_{max} (min^{-1})	K_{app} (nM)	k_{max} (min^{-1})	K_{app} (nM)
No RNA	0.03		0.25	
E.coli 16S + 23S native	170	4.1	580	24
23S T7 transcript	163	18	530	33
E.coli domain V	156	2.8	391	13
B.subtilis domain V	167	5.4	–	–
E.coli 153mer	155	2.9	580	11
B.subtilis 154mer	170	2.8	–	–
E.coli 81mer	103	3.2	540	41
B.subtilis 81mer	142	5.8	–	–

[a]Data from Tsu and Uhlenbeck (30; and in preparation). Note that data for YxiN and DbpA were collected at different ionic strengths due to the tighter binding of YxiN to its RNA cofactor. Specifically, data for YxiN were collected at 175 mM KCl as opposed to 50 mM for DbpA.

activator for ATP hydrolysis allows the extrapolation of the maximal ATPase activity in the presence of a given RNA (k_{max}) and the apparent binding constant of the activating RNA (K_{app}) from the titration of RNA into a solution containing a constant amount of protein and saturating ATP (Fig. 3A). These terms are analogous to k_{cat} and K_m used to describe the interaction of an enzyme with a true substrate, such as ATP. k_{cat} and K_m for ATP were derived from titrations of YxiN with ATP·Mg in the presence of saturating RNA (Fig. 3B).

Initial attempts to determine the affinity of YxiN for 23S rRNA were unsuccessful as the lowest concentrations of RNA used (<1 nM) fully stimulated ATPase activity. This result suggested that YxiN interacts very tightly with RNA. However, it was not possible to determine the apparent binding constant under the conditions previously used for DbpA as the lowest amount of YxiN for which an activity can be reliably measured (2 nM) greatly exceeded the K_{app}. In the presence of 175 mM KCl the interaction of YxiN with RNA was weaker. Even under these conditions, the binding of 23S rRNA to YxiN remained so tight that the enzyme concentration approached, and often exceeded, the apparent binding constant. This situation results in a significant amount of the RNA being sequestered by the enzyme, such that the assumption of substrate (activator) excess implicit in Michaelis–Menton kinetics is no longer valid. RNA saturation curves were therefore fitted to a fully general kinetic equation (equation **1**).

Table 1 lists values of k_{max} and K_{app} for the interaction of YxiN with a number of RNA constructs for which saturable

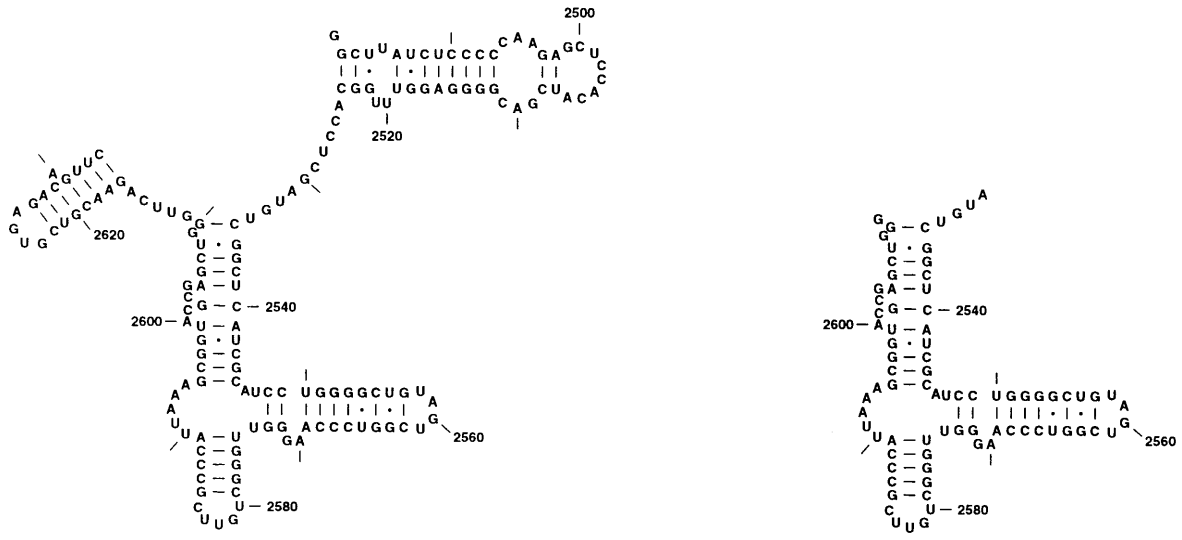

Figure 4. Secondary structures of the *B.subtilis* 154mer and 81mer. Structures are taken from the central wheel of *B.subtilis* 23S rRNA domain V. The secondary structures shown are derived from phylogenetic sequence analysis, as described in Schnare *et al*. (39) and Gutell *et al*. (40).

Table 2. Relative RNA Specificities of YxiN and DbpA

RNA	Relative specificity YxiN[a]	Relative specificity DbpA[b]
E.coli 16S + 23S	1.5	2.2[c]
23S T7 transcript	6.7	3.3[c]
E.coli domain V	1.1	1.8[c]
B.subtilis domain V	2.0	–
E.coli 153mer	1.1	1.0[c]
B.subtilis 154mer	1.0	–
E.coli 81mer	1.9	4.0[c]
B.subtilis 81mer	2.5	–
23S domains I–IV	1 800	3 300
23S domain VI	14 100	11 100
Group I Intron	20 600	Inactive
Poly(A)[d]	9 500	34 400
Poly(U)[d]	7 900	53 400
Yeast tRNA	1 500 000	1 500 000

[a][k_{max}/K_{app} (*B.subtilis* 154mer)]/[k_{max}/K_{app} (RNA)].
[b][k_{max}/K_{app} (*E.coli* 153mer)]/[k_{max}/K_{app} (RNA)].
[c]Data from Tsu and Uhlenbeck (30; and in preparation).
[d]Homopolymer concentrations are in 200mer units.

ATPase activity is observed. Two sub-fragments of domain V used to define the region of RNA responsible for the activation of DbpA are also listed, along with the *B.subtilis* analog of each. These RNAs are referred to by size; the secondary structures of the *B.subtilis* versions are shown in Figure 4 (39,40). From the steady-state kinetic parameters listed in Table 1 it is apparent that YxiN is an RNA-dependent ATPase, with an activity comparable to DbpA. Moreover, YxiN has an apparent binding constant for RNA that is exceedingly tight, even in the presence of high concentrations of KCl.

Bacterial rRNA and fragments taken from domain V of 23S rRNA strongly stimulate the ATPase activity of YxiN. All other RNAs tested to date failed to appreciably stimulate ATPase activity. This result is demonstrated quantitatively in

Table 2, which lists the ratios of apparent second order rate constants (discrimination factors) for various RNA cofactors. There is very little difference between the stimulatory activity of native rRNA and transcripts of 23S rRNA. Transcripts with the sequence of domain V from either *B.subtilis* or *E.coli* are also equivalent RNA cofactors. These results suggest that neither the modified nucleotides found in native rRNA nor the limited number of sequence differences between *E.coli* and *B.subtilis* 23S rRNA in this region have any effect on the ability of these RNAs to elicit ATPase activity.

From a further examination of Table 2 it is clear that transcribed 23S rRNA sequences lacking domain V are poor activators of ATP hydrolysis. YxiN shows a 14 000-fold discrimination between its target sequence in domain V and all of domain VI. There is an 1800-fold difference between stimulation by the 154 nt target sequence and the first four domains of *E.coli* 23S rRNA. In comparison of the second order rate constants for the domain I–IV and VI rRNAs, it is important to note that the concentration of these constructs are measured in 2013 and 279 nt units respectively. The stimulation of ATPase activity by non-specific RNAs is dependent on bulk RNA concentration. This dependence makes the apparent second order rate constant for ATPase activation by a non-specific RNA directly proportional to its length. Discrimination factors are inversely proportional to RNA length. The domain I–IV construct is, therefore, a much weaker activator of ATP hydrolysis than Table 2 suggests.

From the sub-fragments initially used to define the RNA responsible for stimulating the ATPase activity of DbpA, we note that the 154mer is able to fully stimulate the ATPase activity of YxiN. An 81 nt sub-fragment of this construct retains a tight apparent binding constant and elicits ATPase activity to a similar extent. From these results we conclude that the principle RNA target of YxiN is within the central wheel of domain V.

Affinity for ATP and dependence on the RNA cofactor

Table 3 lists the k_{cat} and K_m values for the interaction of YxiN with ATP in the presence of saturating concentrations of four RNAs. Consistent with the results of Table 1, each RNA elicits a similar high rate of ATP hydrolysis when both ATP and RNA are saturating (k_{cat}). All of the constructs which contain the fully active 154mer have a similar K_m for ATP (219–275 μM). However, the K_m for ATP in the presence of the 81mer is nearly 3-fold higher (681 μM). The observation of a significant change in $K_{m, ATP}$ without a change in k_{cat} implies that the affinity of YxiN for ATP may be dependent on the type of RNA bound. This preliminary result suggests that the interaction of RNA and ATP with YxiN may be coupled, an observation which has been previously noted for eIF-4A and the DNA helicase Rep (41,42).

Table 3. ATP kinetic parameters for YxiN

RNA	k_{cat} (min^{-1})	K_m (μM)
E.coli 16S + 23S	163	230
B.subtilis domain V	154	219
B.subtilis 154mer	180	275
B.subtilis 81mer	163	681

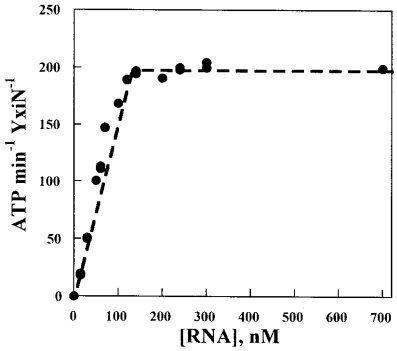

Figure 5. Titration of 120 nM YxiN with *B.subtilis* 154mer shows an apparent 1:1 stoichiometry. The exceedingly high concentration of protein drives the added RNA into an RNA–protein complex, stimulating the ATPase activity of the protein. The ATPase activity was monitored using the standard coupled assay. A breakpoint in the titration was reached at 106 nM added RNA, corresponding to an approximately 1:1 ratio of protein to RNA.

Protein activity and stoichiometry of the active complex

The stoichiometry of the YxiN–RNA complex was determined by titrating excess protein (120 nM) with the fully active 154mer (Fig. 5). The concentration of protein used is ≥40-fold above $K_{app, RNA}$, such that all of the added RNA will form a complex with the enzyme. The ATPase activity of YxiN increases in a linear fashion until a titration point is reached and levels off thereafter. Lines fitted to the ascending and linear portions of this plot reveal a titration point of 106 nM. This corresponds to an approximate 1:1 ratio of RNA to protein. This estimated stoichiometry is dependent on the accuracy of both protein and RNA concentrations and is insensitive to changes in stoichiometry that do not affect ATPase activity. With regard to the former concern, both the protein and RNA are highly purified and we have no reason to suspect that a systematic error in the determination of either concentration would bias the experimentally determined stoichiometry.

Monovalent salt dependence of the interaction of YxiN with RNA

The monovalent salt dependence of a protein–nucleic acid interaction can be used to place an upper limit on the number of ionic contacts involved in complex formation (43,44). The dependence of $K_{app, RNA}$ on KCl concentration was examined for the *B.subtilis* 154mer between 125 and 350 mM KCl. The expected linear variance in a log–log plot was observed (Fig. 6). A linear fit of this data has a slope of 5.4 ± 0.3, indicating that five to six ionic contacts are made between the *B.subtilis* 154mer and YxiN.

The dependence of $K_{app, RNA}$ on KCl concentration was previously determined for DbpA and the *E.coli* 154mer (30). Here a slope of 2.5 ± 0.6 was observed. Relative to DbpA, YxiN may form as many as three additional ionic interactions with the 154 nt RNA. These additional electrostatic contacts may contribute a large portion of the additional RNA binding energy observed with YxiN. Moreover, the monovalent salt dependence of YxiN allows for a prediction of $K_{app, RNA}$ under

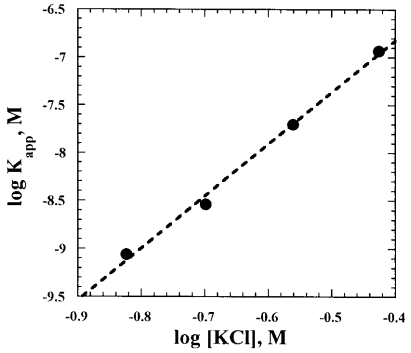

Figure 6. The dependence of $K_{app, RNA}$ on KCl concentration suggests that interaction of YxiN with its cognate 154mer RNA involves five to six ionic contacts. Shown is a plot of log $K_{app, RNA}$ versus log[KCl]. A linear fit of this data has a slope of 5.4 ± 0.3.

the buffer conditions used in the study of DbpA. Were it possible to study the interaction of YxiN with its cognate 154mer at 50 mM KCl, a $K_{app, RNA}$ of ~10 pM would be expected.

DISCUSSION

YxiN is the second DEx$^{D}/_{H}$ protein with a demonstrated RNA-specific ATPase activity in the absence of accessory proteins. It is clear from Table 2 that transcripts of 23S rRNA lacking domain V are unable to stimulate ATP hydrolysis. Therefore, the principle target of YxiN is within domain V of 23S rRNA, a finding similar to that previously obtained in our laboratory and that of Frances Fuller-Pace for DbpA (30,45). Assay of a series of rRNA sub-fragments identifies a 154 nt fragment of domain V which fully stimulates ATPase activity. This fragment is comprised of nt 2481–2634 of *B.subtilis* 23S rRNA (Fig. 4A) and is exactly analogous to that identified in *E.coli* rRNA for DbpA. Portions of this RNA have been shown to interact with tRNAs in the A and P sites of *E.coli* ribosomes, suggesting that it is near the heart of the peptidyltransferase center (46–48).

While the RNA specificity of YxiN is in some cases lower than that of DbpA (Table 2), this may reflect the fact that YxiN binds all RNAs more tightly. The tight interaction of YxiN with RNA may enhance the potential for non-specific RNA–protein interactions. Alternatively, the slightly lower RNA specificity of YxiN may reflect a lower k_{max}, such that non-specific RNAs which stimulate even a modest activity alter the ratio of apparent second order binding constants. Regardless of the reason for the minor difference in the RNA specificities of YxiN and DbpA, the RNA specificity of YxiN is very high when compared to other DEx$^{D}/_{H}$ proteins. Yeast PRP5, the next most specific DEx$^{D}/_{H}$ protein characterized to date, has a 5-fold RNA specificity in its ATPase activity (49). The low nanomolar apparent binding constant and 10^3- to 10^6-fold RNA specificity of YxiN are similar to those of other sequence-specific RNA binding proteins, such as ribosomal protein S15 (50), the splicing factor U1A (51) and the bacteriophage MS2 coat protein (52).

It is possible that the RNA specificity of ATP hydrolysis is greater than that of RNA binding. We employed competition experiments in an attempt to discriminate between RNA specificity at the level of ATP hydrolysis and specificity at the level of binding. If RNAs which do not stimulate the ATPase activity of YxiN bind to a significant extent, they may compete with the RNA-specific ATPase activity. In preliminary experiments we have shown that large quantities of RNA homopolymers have only a modest effect on ATPase activity stimulated by the 81mer (data not shown). These preliminary results suggest that specific RNA binding is at least partially responsible for the RNA specificity of ATP hydrolysis. However, it remains possible that non-specific RNAs are able to bind to YxiN without adversely affecting ATPase activity.

From the characterization of the interaction of YxiN with ATP (Table 3), we note that an apparent ATP binding constant of ~240 μM is similar to those of other DEx$^{D}/_{H}$ proteins. DbpA has a $K_{m, ATP}$ of ~120 μM (30), eIF-4A 330 μM (42) and PRP22 95 μM (16). The similarity of apparent ATP binding constants is not surprising, as proteins of this class interact with ATP via the same set of conserved sequence motifs. Further analysis of Table 3 reveals a possible dependence of $K_{m, ATP}$ on the bound RNA cofactor. Activation of ATP hydrolysis by the shortest RNA in the series shows a significantly higher $K_{m, ATP}$. The fact that k_{cat} remains the same in this case suggests that the change in $K_{m, ATP}$ reflects weakened ATP binding. An apparent change in the interaction with ATP with a change in the activating RNA suggests that the interaction of YxiN with RNA and ATP may be coupled. The coupled binding of ATP and nucleic acid was previously noted in eIF-4A and the DNA helicase Rep (41,42). In the present case, the possibility of coupling provides a plausible mechanistic model for RNA-specific ATPase activity. Namely, that ATP is only productively bound to the complex of YxiN and a specific RNA cofactor.

In the titration of YxiN with 154mer RNA, an approximate 1:1 stoichiometry was attained (Fig. 5). From this titration we conclude that the preparation of YxiN is fully active. Although the active complex appears to have a 1:1 stoichiometry, this does not necessarily imply that the functional form of the enzyme is monomeric. Indeed, most DNA helicases that have been studied in significant detail appear to act in oligomeric complexes (typically dimers or hexamers) (53). These complexes are proposed to provide the helicase with multiple nucleic acid binding sites, allowing rapid, processive movement of the complex. Equilibrium and velocity sedimentation data for the hepatitis C virus NS3 protein, however, suggest that the active form of this DExH helicase is monomeric (54). The implied function of most DEx$^{D}/_{H}$ proteins is the disruption of small stretches of double-stranded RNA. As this function does not appear to necessitate a high degree of processivity (3), the formation of higher order complexes may be unnecessary. Further examination of the aggregation state of YxiN and DbpA is, therefore, of great interest.

The unique substrate specificity of YxiN implies that it may function in ribosome biogenesis or in a heretofore unappreciated aspect of translation. Given that the affinity of an enzyme for its substrates and cofactors is typically matched to their intracellular concentration (55,56), the tight binding of 23S rRNA would appear to eliminate the large ribosomal subunit as a target. This proposition is in agreement with the observation that *E.coli* 50S and 70S ribosomes are poor activators of DbpA (30). However, it remains possible that the tight binding affinity reflects the fact that YxiN must displace ribosomal proteins which are tightly bound to this region of rRNA (48,57,58).

Thus, YxiN could bind to a sub-population of ribosomes and function in an unappreciated facet of translation. We consider it more likely, however, that YxiN interacts with pre-rRNA and functions in ribosome biogenesis. In this respect, YxiN may be analogous to one of the 14 $DEx^{D}/_{H}$ proteins which participate in the assembly of yeast ribosomes (2).

The ability to identify a homolog of DbpA based on the conservation of its C-terminus implies that additional homologs might be identified in a similar manner. We have searched the complete and partially complete genomes of several additional bacteria and have been able to identify seven additional DEAD proteins that share a similar C-terminus. Interestingly, four of the five Gram-negative bacteria with a published 5' DNA sequence appear to initiate at a GUG codon (38). The greater number of potential homologs allowed for the creation of a multiple sequence alignment that clearly defines the C-terminal domain (Fig. 1B). This alignment is extremely strong, with 15% identity and 33% conservation among all nine proteins over 73 amino acids. We propose that the presence of this domain defines a sub-family of DEAD proteins.

A careful inspection of the C-terminal domain reveals the conservation of seven basic amino acids, four glycines and numerous tiny amino acids (G, A and S). The preponderance of highly conserved basic amino acids in the C-terminus is consistent with this region being involved in RNA binding. The known function of YxiN and DbpA appears to necessitate that this domain be able to bind, or assist in the binding of, rRNA. It seems reasonable to propose that this C-terminal sequence folds as an independent domain and assists in the binding of the specific rRNA target.

The co-crystal structures of the related PcrA, Rep and HCV NS3 helicases with DNA oligos assist in the formation of a model for the interaction of YxiN with rRNA (59–61). In all three co-crystal structures the conserved sequence motifs fold into two similar domains that bind ATP and contact the backbone of the bound nucleic acid in a sequence non-specific fashion (62). The general features of the fold adopted by the conserved helicase motifs appear to be quite general. Indeed, the structural threading algorithm of Fischer and Eisenberg predicts that the conserved motifs of both YxiN and DbpA will adopt a similar two domain fold (26). Assuming that the conserved motifs fold into such a two domain structure, the C-terminal domain would be free to fold independently and facilitate the binding of rRNA. The localization of rRNA in this three domain complex would promote non-specific interactions of the conserved amino acid sequence motifs with RNA and elicit RNA-stimulated ATPase activity.

ACKNOWLEDGEMENTS

The authors wish to thank Andrew Feig, Evelyn Jabri, Kevin Polach and Chris Tsu for their helpful suggestions and comments in preparation of this manuscript. Special thanks are due to Rachel Green, Randall Story, Bernard Weisblum, Arthur Zaug, Daniel Ziegler and the Bacillus Genetic Stock Center for providing reagents and constructs. This research was supported in part by a grant from the Pittsburgh Supercomputing Center through the NIH National Center for Research Resources cooperative agreement, and we thank Hugh Nicholas of this organization for his support. The following organizations are recognized for their funding and release of sequence data: Beowulf Genomics, The Cystic Fibrosis Foundation, The Department of Energy, The National Institute for Allergy and Infectious Diseases, Pathogenesis and The Wellcome Trust. K.K. was supported in part by a grant from the Colorado Advanced Technology Institute.

REFERENCES

1. Staley,J.P. and Guthrie,C. (1998) *Cell*, **92**, 315–326.
2. de la Cruz,J., Kressler,D. and Linder,P. (1999) *Trends Biochem. Sci.*, **24**, 192–198.
3. Rogers,G.W.,Jr, Richter,N.J. and Merrick,W.C. (1999) *J. Biol. Chem.*, **274**, 12236–12244.
4. Chuang,R.Y., Weaver,P.L., Liu,Z. and Chang,T.H. (1997) *Science*, **275**, 1468–1471.
5. Tseng,S.S., Weaver,P.L., Liu,Y., Hitomi,M., Tartakoff,A.M. and Chang,T.H. (1998) *EMBO J.*, **17**, 2651–2662.
6. Snay-Hodge,C.A., Colot,H.V., Goldstein,A.L. and Cole,C.N. (1998) *EMBO J.*, **17**, 2663–2676.
7. Py,B., Higgins,C.F., Krisch,H.M. and Carpousis,A.J. (1996) *Nature*, **381**, 169–172.
8. Anderson,J.S. and Parker,R. (1996) *Curr. Biol.*, **6**, 780–782.
9. Shiratori,A., Shibata,T., Arisawa,M., Hanaoka,F., Murakami,Y. and Eki,T. (1999) *Yeast*, **15**, 219–253.
10. Hodges,P.E., McKee,A.H., Davis,B.P., Payne,W.E. and Garrels,J.I. (1999) *Nucleic Acids Res.*, **27**, 69–73.
11. Madhani,H.D. and Guthrie,C. (1994) *Annu. Rev. Genet.*, **28**, 1–26.
12. Luking,A., Stahl,U. and Schmidt,U. (1998) *Crit. Rev. Biochem. Mol. Biol.*, **33**, 259–296.
13. Iost,I., Dreyfus,M. and Linder,P. (1999) *J. Biol. Chem.*, **274**, 17677–17683.
14. Laggerbauer,B., Achsel,T. and Luhrmann,R. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 4188–4192.
15. Wang,Y., Wagner,J.D. and Guthrie,C. (1998) *Curr. Biol.*, **8**, 441–451.
16. Wagner,J.D., Jankowsky,E., Company,M., Pyle,A.M. and Abelson,J.N. (1998) *EMBO J.*, **17**, 2926–2937.
17. Herschlag,D. (1995) *J. Biol. Chem.*, **270**, 20871–20874.
18. Staley,J.P. and Guthrie,C. (1999) *Mol. Cell*, **3**, 55–64.
19. Linder,P., Lasko,P.F., Ashburner,M., Leroy,P., Nielsen,P.J., Nishi,K., Schnier,J. and Slonimski,P.P. (1989) *Nature*, **337**, 121–122.
20. Fuller-Pace,F. (1994) *Trends Cell Biol.*, **4**, 271–274.
21. Wang,Y. and Guthrie,C. (1998) *RNA*, **4**, 1216–1229.
22. Moszer,I. (1998) *FEBS Lett.*, **430**, 28–36.
23. Pearson,W.R., Wood,T., Zhang,Z. and Miller,W. (1997) *Genomics*, **46**, 24–36.
24. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402.
25. Gupta,S.K., Kececioglu,J.D. and Schaffer,A.A. (1995) *J. Comput. Biol.*, **2**, 459–472.
26. Fischer,D. and Eisenberg,D. (1996) *Protein Sci.*, **5**, 947–955.
27. Fujita,Y. (1996) EMBL accession no. D83026.
28. Yoshida,K., Shindo,K., Sano,H., Seki,S., Fujimura,M., Yanai,N., Miwa,Y. and Fujita,Y. (1996) *Microbiology*, **142**, 3113–3123.
29. Yoshida,K., Sano,H., Seki,S., Oda,M., Fujimura,M. and Fujita,Y. (1995) SwissProt accession no. P42305.
30. Tsu,C.A. and Uhlenbeck,O.C. (1998) *Biochemistry*, **37**, 16989–16996.
31. Gill,S.C. and von Hippel,P.H. (1989) *Anal. Biochem.*, **182**, 319–326.
32. Milligan,J.F. and Uhlenbeck,O.C. (1989) *Methods Enzymol.*, **180**, 51–62.
33. Weitzmann,C.J., Cunningham,P.R. and Ofengand,J. (1990) *Nucleic Acids Res.*, **18**, 3515–3520.
34. Kovalic,D., Giannattasio,R.B., Jin,H.J. and Weisblum,B. (1994) *J. Bacteriol.*, **176**, 6992–6998.
35. Zaug,A.J., Grosshans,C.A. and Cech,T.R. (1988) *Biochemistry*, **27**, 8924–8931.
36. Bessman,M. (1963) *Methods Enzymol.*, **6**, 166–176.
37. Segel,I. (1993) *Enzyme Kinetics*. John Wiley & Sons, New York, NY, pp. 72–77.
38. Fuller-Pace,F.V., Nicol,S.M., Reid,A.D. and Lane,D.P. (1993) *EMBO J.*, **12**, 3619–3626.
39. Schnare,M.N., Damberger,S.H., Gray,M.W. and Gutell,R.R. (1996) *J. Mol. Biol.*, **256**, 701–719.
40. Gutell,R.R., Gray,M.W. and Schnare,M.N. (1993) *Nucleic Acids Res.*, **21**, 3055–3074.
41. Lohman,T.M. (1993) *J. Biol. Chem.*, **268**, 2269–2672.

42. Lorsch,J.R. and Herschlag,D. (1998) *Biochemistry*, **37**, 2180–2193.
43. Record,M.T.,Jr, Anderson,C.F. and Lohman,T.M. (1978) *Q. Rev. Biophys.*, **11**, 103–178.
44. Carey,J. and Uhlenbeck,O.C. (1983) *Biochemistry*, **22**, 2610–2615.
45. Nicol,S.M. and Fuller-Pace,F.V. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 11681–11685.
46. Joseph,S. and Noller,H.F. (1996) *EMBO J.*, **15**, 910–916.
47. Green,R., Switzer,C. and Noller,H.F. (1998) *Science*, **280**, 286–289.
48. Zimmerman,R.A., Thomas,C.L. and Wower,J. (1990) In Hill,W.E., Dahlberg,A., Garrett,R.A., Moore,P.B., Schlessinger,D. and Warner,J.R. (eds), *The Ribosome: Structure, Function and Evolution*. American Society for Microbiology, Washington, DC.
49. O'Day,C., Dalbadie-McFarland,G. and Abelson,J. (1996) *J. Biol. Chem.*, **271**, 33261–33267.
50. Batey,R.T. and Williamson,J.R. (1996) *J. Mol. Biol.*, **261**, 536–549.
51. Rimmele,M.E. and Belasco,J.G. (1998) *RNA*, **4**, 1386–1396.
52. LeCuyer,K.A., Behlen,L.S. and Uhlenbeck,O.C. (1995) *Biochemistry*, **34**, 10600–10606.
53. Lohman,T.M. and Bjornson,K.P. (1996) *Annu. Rev. Biochem.*, **65**, 169–214.
54. Porter,D.J., Short,S.A., Hanlon,M.H., Preugschat,F., Wilson,J.E., Willard,D.H.,Jr and Consler,T.G. (1998) *J. Biol. Chem.*, **273**, 18906–18914.
55. Fersht,A. (1985) *Enzyme Structure and Mechanism*. W.H. Freeman, New York, NY, pp. 311–346.
56. Hackney,D.D. (1990) In Boyer,P.D. (ed.), *The Enzymes*. Academic Press, New York, NY, Vol. 19, pp. 1–36.
57. Baranov,P.V., Kubarenko,A.V., Gurvich,O.L., Shamolina,T.A. and Brimacombe,R. (1999) *Nucleic Acids Res.*, **27**, 184–185.
58. Osswald,M., Greuer,B. and Brimacombe,R. (1990) *Nucleic Acids Res.*, **18**, 6755–6760.
59. Velankar,S.S., Soultanas,P., Dillingham,M.S., Subramanya,H.S. and Wigley,D.B. (1999) *Cell*, **97**, 75–84.
60. Korolev,S., Hsieh,J., Gauss,G.H., Lohman,T.M. and Waksman,G. (1997) *Cell*, **90**, 635–647.
61. Kim,J.L., Morgenstern,K.A., Griffith,J.P., Dwyer,M.D., Thomson,J.A., Murcko,M.A., Lin,C. and Caron,P.R. (1998) *Structure*, **6**, 89–100.
62. Korolev,S., Yao,N., Lohman,T.M., Weber,P.C. and Waksman,G. (1998) *Protein Sci.*, **7**, 605–610.