Identification of important chemical groups of the *hut* mRNA for HutP interactions that regulate the *hut* operon in *Bacillus subtilis*

T. S. Kumarevel[1,2], S. C. B. Gopinath[1], S. Nishikawa[1], H. Mizuno[2] and P. K. R. Kumar[1,*]

[1]Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), Central 6, Tsukuba, Ibaraki 305-8566, Japan and [2]Department of Biochemistry, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan

ABSTRACT

HutP is an RNA binding protein that regulates the expression of the histidine utilization (*hut*) operon in *Bacillus* species by binding to *cis*-acting regulatory sequences on *hut* mRNA. We recently solved the HutP crystal structure, which revealed a novel fold where three dimers are arranged in a 3-fold axis to form the hexamer. We also identified a minimal RNA binding element sufficient for HutP binding: three UAG trinucleotide motifs, each separated by 4 nt, located just upstream of the terminator. In the present study we have identified important RNA chemical groups essential for HutP interactions, by combining an *in vitro* selection strategy and analyses by site-specific base substitutions. These analyses suggest that each HutP molecule recognizes one UAG motif, where the first base (U) can be substituted with other bases, while the second and third bases (A and G) are required for the interactions. Further analyses of the chemical groups of the A and G bases in the UAG motif by modified base analogs suggested the importance of the exocyclic NH_2 group in these bases. Also, in this motif, only the 2′-OH group of A is important for HutP recognition. Considering the important chemical groups identified here, as well as the electrostatic potential analysis of HutP, we propose that Glu137 is one of the important residues for the HutP–RNA interactions.

INTRODUCTION

Eubacteria are known to utilize transcriptional termination as a method for gene regulation. A frequently used pathway involves the formation of a potential terminator structure upstream of the coding region, which functions in concert with activated cellular factors to terminate at that site or to allow transcription to proceed through it. In one case, the quiescent state allows transcription, and the binding of an activated factor activates the termination complex, causing transcription to abort. In the other, the ground state of the mRNA possesses a specific secondary structure capable of triggering the RNA polymerase to pause, prompting the release of both the polymerase and RNA transcript from the DNA template (1). In this case, the binding of a *trans*-acting factor alleviates this block, allowing transcriptional read-through. This second process is generally referred to as an attenuation or anti-termination mechanism of regulation. The distinction between these two pathways is the end result of interactions between the terminator and the activated protein: is transcription terminated or allowed to continue? In the attenuation process, the regulatory protein, activated by the regulatory molecule, pauses the transcription at the terminator structure, which otherwise permits the read-through of the transcription apparatus. The best example for this kind of regulation is the tryptophan biosynthetic operon (*trpEDCBA*) of *Bacillus subtilis* (2). In contrast to this mechanism, the anti-termination process requires the activated protein to bind to the pre-existing terminator structure, to allow the RNA polymerase to transcribe the downstream genes. An example of this is the BglG/SacY family of anti-termination proteins (3,4). Another example in this family is HutP, which regulates the histidine utilization (*hut*) operon in *B.subtilis* (5,6). The target sequences for these anti-terminator proteins comprise either single- or double-stranded regions of their respective mRNAs.

The *hut* operon consists of five structural genes, *hutH*, *hutU*, *hutI*, *hutG* and *hutM*, which are located far downstream from the promoter and a positive regulatory gene, *hutP*, located just downstream from the promoter (7–10). The nucleotide sequence located in between the *hutP* gene and structural genes forms a stem–loop terminator structure (6) (Figure 1a). This terminator structure is attenuated by HutP after it is activated by the binding of L-histidine. Recently, *cis*-acting regulatory sequences were identified that are essential for the full attenuation of the terminator, located just upstream of and partially overlapping the putative terminator sequence (11). HutP protein only interacts with the regulatory sequence on *hut* mRNA after it is allosterically activated by L-histidine. The HutP protein is also found in other *Bacillus* species such as *Bacillus anthracis* (12), *Bacillus cereus* (13) and *Bacillus holodurans* (14). The primary sequence of HutP is ~60% conserved between these species.

---

*To whom correspondence should be addressed. Tel: +81 29 861 6085; Fax: +81 29 861 6095; Email: pkr-kumar@aist.go.jp

Figure 1. A proposed structure of *hut* mRNA terminator and crystal structure of HutP. (**a**) The proposed *hut* mRNA terminator structure. The nucleotide regions +498 to +572 are proposed to form a terminator structure. (**b**) A view of the HutP dimer along the non-crystallographic 2-fold axis. HutP is shown in a ribbon diagram with labels for α-helices and β-strands (molecule A, green; molecule B, red), and HBN is shown as a 'ball-and-stick' model colored by atom type (nitrogen, blue; carbon, green; oxygen, red). The N- and C-termini of HutP are indicated by N and C, respectively. (**c**) A view of the HutP hexamer along the 3-fold axis with HBN, which is necessary for its biological function. The E137 residue, a potential residue proposed for RNA recognition is shown as a ball-and-stick model.

We, recently, solved the crystal structure of HutP in the activated form (15). The structure revealed a novel protein fold; two HutP monomers ford a dimer and three dimers are arranged in a crystallographic 3-fold symmetry to form a

hexamer (Figure 1b and c). The histidine analog is located between the two monomers. The imidazole ring of the L-histidine is buried in a hydrophobic pocket made up by surrounding residues from two adjacent molecules. The amino and carboxy moieties of the L-histidine analog have an extensive hydrogen bond network within the hydrophobic pocket. Interestingly, we could identify the key residues involved in the activation process and proposed an activation mechanism for the HutP protein.

The *hut* mRNA appears to form two alternative secondary structures that depend upon the presence of the activated HutP protein (11). The regions that fold into alternative structures are located between +459 to +572 of *hut* mRNA. The *hut* mRNA appears to form a stable terminator structure (+498 to +572) in the absence of activated HutP, whereas, the region between +459 to +537 seems to form a destabilized RNA structure (ribonucleic anti-terminator, RAT) upon binding to the activated HutP (11). Our recent analysis suggested a modified secondary structure, in which the segment of nucleotides 498 to 514 does not form a hairpin, but is bound to the HutP protein (15).

The minimal RAT that binds HutP was mapped previously to the nucleotide region from +459 to +537 of the *hut* mRNA. Our recent structural (15) and biochemical studies (11) revealed that HutP forms a hexamer, as a trimer of dimers (Figure 1b and c). Based on these results, we searched for repeated sequences within the minimal 79mer RAT element, and found three UAG repeats that are each separated by a 4 nt spacer region. To determine whether this is the HutP binding region, we synthesized a short RNA (21mer RAT) representing the region from +496 to +516 of the *hut* mRNA, and analyzed its binding to the activated HutP (15). The equilibrium binding studies revealed that the 21mer RAT binds to activated HutP as efficiently as the 79mer RAT. Thus, the 21mer RNA represents the minimal RNA element sufficient for HutP recognition. The bases located in the spacer regions in between the UAG motifs could be replaced with other bases, but the 4 nt spacing was optimal for HutP recognition. The substitution of the spacer nucleotides by four U bases in between the UAG binding motif increased the affinity for HutP. Nevertheless, the functional groups important for the HutP–RNA interaction, within both the UAG motifs and the HutP protein, remained obscure.

In the present study, we reveal the RNA identity elements for the *hut* mRNA–HutP interactions. We achieved this by a combination of both an *in vitro* selection strategy and a site-specific base substitution interference analysis. We found that in the conserved UAG motif, the initial U could be replaced with other bases, while the A in the second position and the G in the third position were progressively more important for HutP binding. These studies identified the core recognition motif as XAG for the HutP interaction, where X indicates any base. Site-specific base modification analyses further revealed that the 6-amino moiety of the A base and the 2-amino moiety of the G base, at the second and third positions, respectively, are the key RNA functional groups for HutP recognition. Further analyses on the chemical groups of RNA within the UAG motif revealed only one 2′-OH of A that is important for HutP recognition. By combining various analyses, such as base modifications of RNA, electrostatic potential and mutational analyses of HutP, and database

analyses of protein–RNA complexes, we now propose an elegant model for the HutP–RNA interactions.

MATERIALS AND METHODS

Overexpression and purification of HutP

The nucleotide sequence encoding HutP, containing the Val 51 Ile mutation, was amplified and cloned into the pET5a vector (Promega). The resultant plasmid was transformed into *Escherichia coli* strain BL21 (DE3), and the HutP protein was overexpressed at mid-log phase by the addition of IPTG (1 mM) and purified as described previously (16).

RNA synthesis

The 79mer RNA was enzymatically synthesized by *in vitro* transcription using T7 RNA polymerase on a synthetic DNA template. Initially, the template DNA containing the region from +459 to +537 of RAT was chemically synthesized. To generate the double-stranded DNA and to add the T7 promoter to the template, we synthesized two additional sets of primers (forward primer, 5′-*AGTAATACGACTCACTATAGGG*ACATTCGGCGTTGGAATTATCA-3′; and reverse primer, 5′-CTTCTGTTTTCACGCATAGCCCCCTAT-3′). Using the template DNA and primers, we prepared the dsDNA template, which now contained a T7 promoter (in italics), using a commercial PCR kit (Ex*Taq* kit, Takara, Japan). The reaction mixture was cycled at 94°C for 1.15 min, 50°C for 1.0 min, and 72°C for 1.15 min for 10 cycles. The resulting PCR product was precipitated, and ∼2 μg were used for *in vitro* T7 transcription in the presence of [α-^{32}P]CTP (Amersham). Transcription was performed at 37°C for 3 h, using an Ampliscribe transcription kit (Epicenter Technologies, Madison, WI). After transcription, the products were treated with 2 U of DNase I (RNase free) for 10 min at 37°C, to remove the template DNA, and were mixed with an equal volume of 2× urea buffer (7 M urea, 50 mM EDTA, 90 mM Tris–borate containing 0.05% bromophenol blue). The reaction mixtures were denatured at 90°C for 2 min and fractionated on a 12% polyacrylamide gel containing 7 M urea. The RNA band was excised, and the RNA was eluted from the gel. The RNA was concentrated by ethanol precipitation, redissolved in water and then quantitated by the absorbance at 260 nm.

The minimal RAT binding RNA elements shown in Table 1 were chemically synthesized on an RNA/DNA synthesizer (Model 394; Applied Biosystems Inc.) using phosphoramidite chemistry. All amidites were purchased from Glen Corporation. The functional groups of the synthesized RNAs were deprotected by established protocols (ABI Manual), and the RNA was purified on a 20% acrylamide gel containing 7 M urea. We labeled these RNAs at the 5′ end with [γ-^{32}P]ATP (Amersham) in the presence of T4 polynucleotide kinase (Takara, Japan), and recovered the labeled products after fractionation by 20% PAGE.

Filter binding assay

The RAT RNAs (both the shorter and longer ones) for the HutP binding studies were used in a filter binding assay similar to that reported previously (15). Various concentrations (10–1330 nM final) of purified HutP were incubated in binding

Table 1. 21mer RNA and its variants (all sequences given 5′–3′)

3mer	UAG
4mers	UAGA, AUAG
7mer	UUUAGUU
14mer	UUUAGUUUUUAGUU
21mer	CAUAGAUCUUAGACGAUAGGG
21mer (UAG→GAU)	CAGAUAUCUGAUACGAGAUGG
4U-spacer	UUUAGUUUUUAGUUUUUAGUU
U-1 base mutated to A, C, G	UU**A**AGUUUU**A**AGUUUU**A**AGUU
	UU**C**AGUUUU**C**AGUUUU**C**AGUU
	UU**G**AGUUUU**G**AGUUUU**G**AGUU
A-2 base mutated to G, C, U	UUU**G**GUUUUU**G**GUUUUU**G**GUU
	UUU**C**GUUUUU**C**GUUUUU**C**GUU
	UUU**U**GUUUUU**U**GUUUUU**U**GUU
Atomic mutations at A-2	UUUA$_{2AP}$GUUUUUA$_{2AP}$GU-UUU-A$_{2AP}$GUU
G-3 base mutated to A, C, U	UUUAA**U**UUUUAA**U**UUUUAA**U**UU
	UUUA**C**UUUUUA**C**UUUUUA**C**UUU
	UUUA**U**UUUUUA**U**UUUUUA**U**UUU
Atomic mutations at G-3 in all three sites	UUUA**I**UUUUUA**I**UUUUUA**I**UUU
Individual G-3 site mutations	UUUA**I**UUUUUAGUUUUUAGUU
	UUUAGUUUUUA**I**UUUUUAGUU
	UUUAGUUUUUAGUUUUUA**I**UU
UAG motif mutated to dU, dA, dG	UU**dU**AGUUUU**dU**AGUUUU**dU**AGUU
	UUU**dA**GUUUUU**dA**GUUUUU**dA**GUU
	UUUA**dG**UUUUUA**dG**UUUUUA**dG**UU

I and A$_{2AP}$ denote ribo-inosine and 2-amino purine, respectively.

buffer (15 mM HEPES, pH 7.5, 30 mM NaCl, 5 mM MgCl$_2$, 10 μg of yeast tRNA), and L-histidine was added to a final concentration of 10 mM. The concentrations of HutP were based on the hexamer (96 000 Da). To this reaction mixture, RNA was added to a final concentration of 20 nM. The amount of RNA retained on the filter was quantitated (radioactivity) by a bio image analyzer BAS 2500 (Fuji Film). The equilibrium dissociation (K_d) constants were determined for different RNA–protein complexes using a non-linear, curve-fitting algorithm (GraphPad Prism 2.0, GraphPad Software).

In vitro selections

We synthesized the template oligo with randomized bases at the sites of the three UAGs in the 64mer RAT RNA: 5′-TTAATCACATATGAAACAGCCCANNNATCTNNNAC-GANNNGGGGCTATGCGTGAAAAACG-3′. To create an RNA library, this was first amplified to add a T7 promoter (in italics), with the forward primer 5′-AG*TAATACGACTCACTATAGG*AATTAATCACATATGAAACAGCCCA-3′ and the reverse primer 5′-CGTTTTTCACGCATAGCCCC-3′. The dsDNA product was prepared using a commercial PCR kit (Ex*Taq* kit), using cycling conditions of 94°C for 1.15 min, 50°C for 1.0 min and 72°C for 1.15 min for 12 cycles. The resulting PCR product was precipitated and used for *in vitro* T7 transcription at 37°C for 3 h, using the Ampliscribe transcription kit (Epicenter Technologies). After the transcription, the products were treated and purified as described in the RNA synthesis section above, except that an 8% polyacrylamide gel was used instead of a 20% gel. For each selection cycle, the RNA pool was denatured in binding buffer at 90°C for 2 min and then re-folded at room temperature, after which the activated HutP was added and the mixture was equilibrated for an additional 15 min at room temperature. The HutP protein was activated by incubating the desired concentration at room

temperature in the presence of 10 mM L-histidine in binding buffer. The ratio of RNA to protein was varied with the selection cycle, as indicated in Table 2. To separate the free and complexed RNAs from the mixture, the total reaction mixture was filtered through a nitrocellulose membrane (HAWP filter, 0.45 μM, 13.0 mm diameter, Millipore) which was pre-wet with the binding buffer. The membrane was washed once with 1 ml of binding buffer containing L-histidine. The RNA species complexed with the protein, which was bound to the filter, were eluted by denaturing the protein in a buffer containing 7 M urea and heating the contents at 90°C. The eluted RNAs were concentrated, precipitated with ethanol and resuspended in binding buffer. The resultant RNAs were used to generate a new RNA population for selection using reverse transcription, PCR and *in vitro* T7 transcription, as described previously (17). We carried out a total of seven selection rounds. In each cycle we applied more stringent conditions by decreasing the concentration of activated HutP relative to RNA, to ensure the selection and enrichment of the highest affinity binders in the pool. The binding ability of selected RNA pools (see Results and discussion) was assayed using ^{32}P body labeled RNA. The pools were quantitated by A_{260}, and the filter binding assays were performed as described above. Since the RNA pool from the seventh selection cycle showed only marginally higher binding ability to the

Table 2. Concentrations of RNA pool and HutP protein used during *in vitro* selections

Selection	Protein (μM)	RNA (μM)	Ratio (protein:RNA)	No. of PCR cycles[a]	Filter binding assay[b] (% of pool RNA bound to HutP protein)
1	20	10	2:1	12	0.5
2	2	2	1:1	10	Nd
3	2	2	1:1	10	7.4
4	1	1	1:1	10	Nd
5	0.5	1	1:2	8	29
6	0.5	5	1:10	6	Nd
7	0.5	10	1:20	4	46

[a]Number of PCR cycles carried out to amplify a product that is visible on a 4% Nusieve agarose in the presence of ethidium bromide staining.
[b]The filter binding assay was carried out in 100 ul of binding buffer containing HutP (2.0 uM), the RNA pool (40 nM) and tRNA (400 nM) only for the first, third, fifth and seventh round selections.

HutP than that from the fifth cycle, the seventh cycle RNA pool was converted into dsDNA by PCR, cloned into a TA vector (pCRII, Invitrogen, Japan K.K.), and then transformed into *E.coli* TOP10 cells. Plasmids from the positive colonies were isolated using the Qiaprep kit (Qiagen, K.K., Tokyo) and sequenced.

Gel-shift Assay

To evaluate the protein–RNA interaction ratios in solution, we carried out a gel-shift assay. The 79mer RNA was internally labeled as described above. The purified HutP was incubated in binding buffer (15 mM HEPES, pH 7.5, 30 mM NaCl, 5 mM MgCl$_2$, 10 μg of yeast tRNA) at pre-determined concentrations, and the RNA (79mer) probe was mixed at various concentrations in the presence of 10 mM L-histidine. The resulting reaction mixture was incubated for 15 min at room temperature, and then mixed with 2 μl of 90% glycerol before loading onto the gel. The free RNA and HutP–RNA complexes were resolved using a 10% polyacrylamide gel (running buffer, 0.5× TBE; constant voltage, 15 mA; temperature, 4°C). The resulting autoradiograms were analyzed with a bio image analyzer, BAS 2500 (Fuji Film).

Sequencing of the selected RNAs

Plasmids were sequenced using the conditions recommended for the ABI PRISM 3100 (Applied Biosystems, Foster City, CA).

RESULTS AND DISCUSSION

HutP binding analysis with shorter RNAs

Previously, our studies suggested that a 21mer RNA, composed of three UAG motifs with four spacer nucleotides within the RAT, is sufficient for efficient HutP recognition (15). Based on these studies, the UAG appears to be the key motif for interactions with HutP; however, it was not clear whether the UAG motif (trimer RNA) alone is sufficient for HutP binding. To test this, we synthesized shorter RNAs: 3mer (UAG), 4mer (AUAG/ UAGA) and 7mer (UUUAGUU) RNAs containing either 3′ and/or 5′ phosphates, and analyzed them for binding to activated HutP (Figure 2). None of the
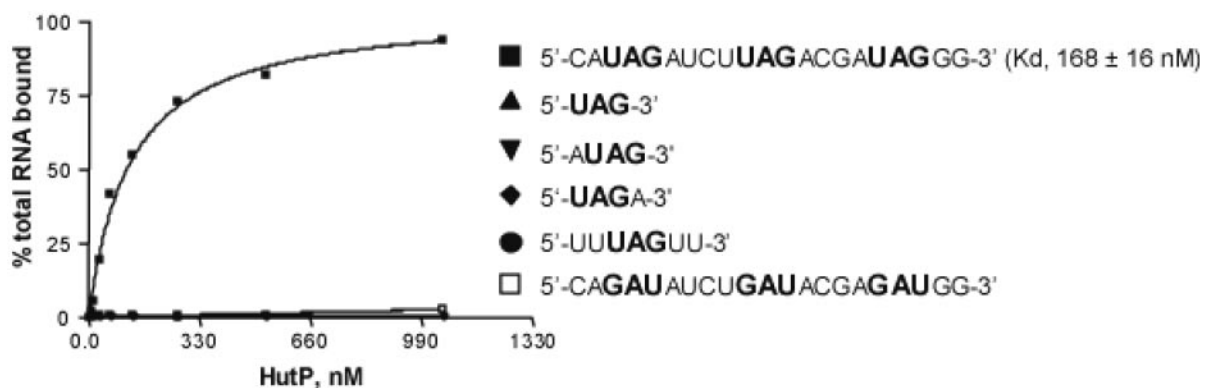

Figure 2. HutP binding analysis with shorter RNAs. The binding analysis was carried out using 20 nM of 5′ end labeled RNA in binding buffer (15 mM HEPES, pH 7.5, 30 mM NaCl, 5 mM MgCl$_2$, 10 μg of yeast tRNA) containing 10 mM of L-histidine. The HutP protein was initially activated with 10 mM of L-histidine, and binding reactions were carried out in the presence of various protein concentrations, ranging from 10 to 1330 nM.

shorter RNAs, including the 7mer, bound to HutP, suggesting that cooperative binding is important, probably to enhance the affinity of an RNA that contains three UAG motifs. However, when the three UAG motifs were substituted with the reverse sequence (UAG → G̲A̲U̲), the binding ability of this RNA was completely lost (Figure 2). This binding analysis clearly shows that the bases in the trimer are recognized specifically by the HutP protein. To substantiate further the importance of each of the bases, as well as to ascertain the tolerance for base substitutions in the UAG motif, we adopted an *in vitro* selection strategy. This allowed us to analyze many possible variations in a single experiment. For this, we completely randomized all three UAG sites, and presented them in the 64mer RAT backbone rather than the 21mer RAT subdomain, to maintain the complexity of folding for the longer RNA, which is assumed to be much more complicated than that of the shorter RNA.

In vitro selection strategy

We carried out a total of seven rounds of selection and amplification to enrich for RNA species that bind specifically to the activated HutP protein. In each successive round, we applied more stringent selection conditions, decreasing the absolute concentration of protein and the ratio of activated HutP to the RNA pool (Table 2). To avoid the selection of RNAs that bind to the filter, we pre-filtered the pool prior to each selection cycle. To monitor the progression of the enrichment for HutP binders, we analyzed the binding ability of the RNA pool after the first, third, fifth and seventh cycles, using a filter binding assay. For this, the activated protein was captured on nitro-cellulose membranes, along with any bound labeled pool RNAs. The binding analysis clearly indicates that as the cycles of selections progressed, the binding ability of the pool improved and reached ∼46% binding at saturation, at the end of the seventh selection cycle (Table 2). Also, to test the specificity of the binding behavior for the selected pool, we carried out a filter binding analysis in the presence of a 10-fold excess of total tRNA (from *E.coli*), and the binding was unaffected by the presence of the non-specific tRNA molecules. Since the seventh cycle RNA pool has the highest binding ability to activated HutP, we cloned and sequenced from this pool to determine if the fittest sequences for HutP binding fell into obvious classes. Interestingly enough, the majority of these RNAs (70%) conserved UAG at all three randomized positions (Classes I and II, Figure 3). The remaining 30% of the sequences contained mutations at only one nucleotide between any of the three triplets, predominantly at the U base. This was substituted with all other bases (Class III). These *in vitro* genetic selections clearly confirmed that HutP binds to the RNA motif containing three UAGs with 4 nt spacers in between them. We analyzed the binding activity for the first base variants, and found that they have a similar affinity for the protein as that of the wild-type 79mer RAT. This suggests that the first U could be substituted with any other base, without much effect on the binding. These results revealed that HutP prefers to bind to RNA that possesses three UAG motifs, but the binding interaction itself is not significantly affected when the first base of one UAG motif is not U. In these studies, we did not attempt to substitute bases in the spacer region, but the predominant clones after seven rounds of selections possessed AUAU and ACGA in the spacer regions. We believe these sequences were generated

Figure 3. *In vitro* selection of *hut* mRNA and conserved sequences from the selected RNA pool. RNA sequences found in the selected RNA pool. Mutations in the 64mer RAT are indicated by bold and italic for the UAG and spacer regions, respectively. The sequences were separated into three classes, based on the conservation of the UAG motif in all three positions and the presence of a mutation in the spacer regions.

at either the RT or PCR step. Nevertheless, the RAT containing this sequence has about 2-fold higher binding affinity for HutP as compared to the wild-type RAT, but the same affinity as the RAT with all of the bases in the spacer region substituted by U's.

The *in vitro* selection analysis also confirmed that the 21mer RNA possesses all of the necessary elements for HutP binding, including the three UAG motifs with the 4 nt spacer regions in between them (15). Since our previous analysis suggested that the optimum spacer length was 4 nt, in the current studies, we used a base substitution analysis instead of a base deletion analysis, in order to maintain the appropriate distance for optimal binding.

Evaluation of the U base at the first position in the UAG motif

In the selections, we obtained U substitutions only in one of the UAG motifs in RAT, and not in two or all three motifs. To evaluate the importance of this base in all three UAG sites, we systematically substituted the U base with A, G or C in all three triplets in the 21mer RAT. The binding analysis clearly revealed that all three variants of RAT retained the same binding ability as the unmodified 21mer RAT (Figure 4a), suggesting that the U base at the first position in any of the motifs could be replaced with other bases.

Evaluation of the A base at the second position in the UAG motif

In our selection studies, we did not obtain enough variants in the selected RNA pool to evaluate substitutions at the second and third positions, the A and G bases, respectively. To address the importance of these bases in recognition, we prepared three 21mer RNAs containing site-specific substitutions at the second positions, where A was systematically replaced in all three sites by U, G or C bases. The binding analysis shows that the RNAs with A to G and A to C substitutions completely lost the ability to bind HutP. On the other hand, the A to U substitutions retained the affinity for HutP, although it
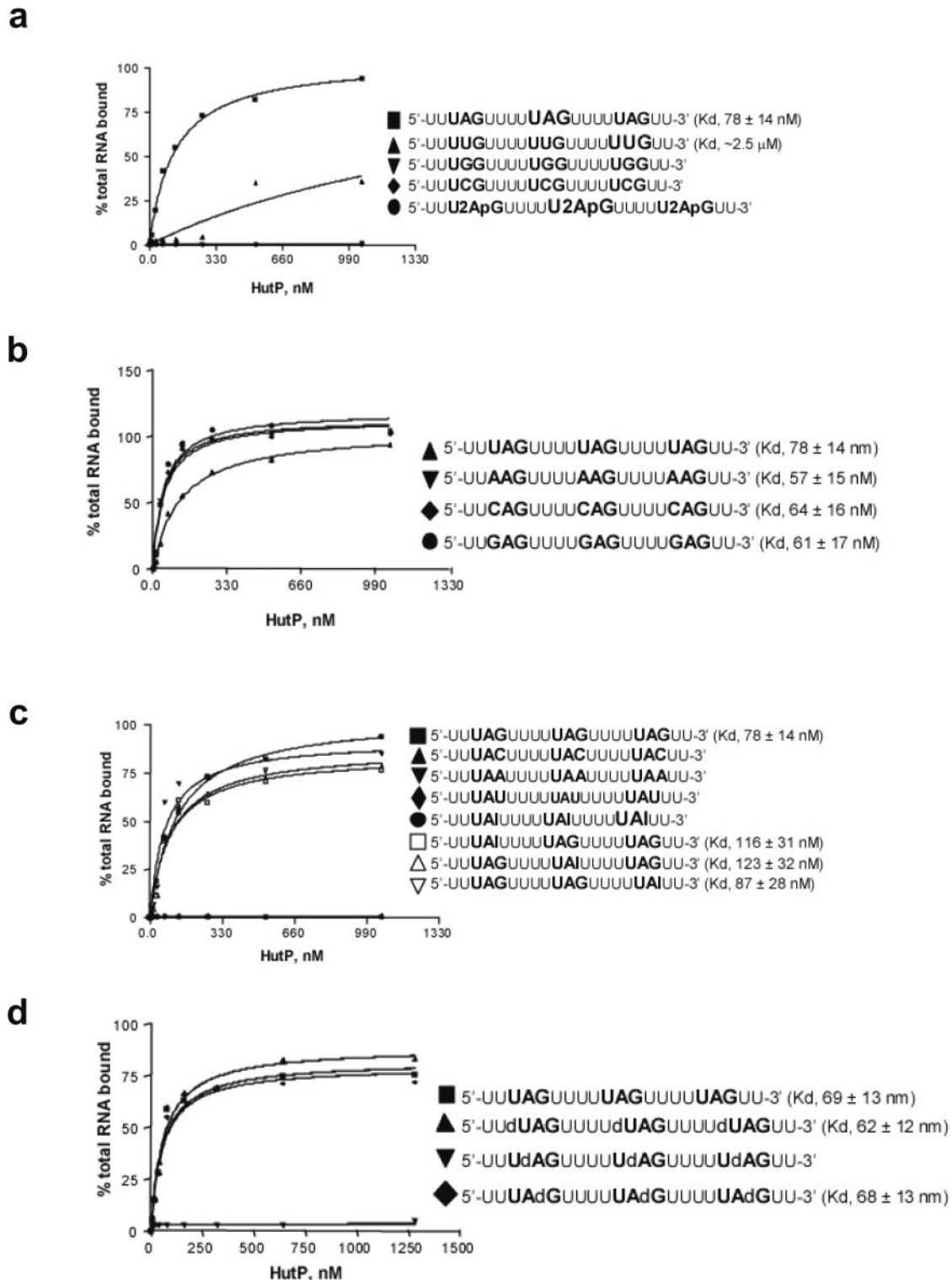
Figure 4. Effects of base substitutions and deoxy ribose substitutions at the UAG motif on HutP binding. (**a**) Effects of base substitutions at the UAG motif on HutP binding. RNAs containing substitutions (A or G or C) were prepared chemically, labeled at the 5′ end, and then used in the filter binding assay described in Figure 2. (**b**) Effects of base substitutions and the base analog at the second base of the UAG motif on HutP binding. RNAs containing substitutions (U or G or C), in addition to RNA containing the 2-amino purine analog, were prepared chemically, labeled at the 5′ end, and then used in the filter binding assay described in Figure 2. (**c**) Effects of base substitutions and the base analog at the third base of the UAG motif on HutP binding. RNAs containing substitutions (A or C or U), in addition to RNA containing ribo-inosine, were prepared chemically, labeled at the 5′ end, and then used in the filter binding assay described in Figure 2. To evaluate the positional importance for the UAG motifs' recognition, we prepared three RNAs containing inosine substitutions at the first, second and third UAG motif in the 21mer RAT. These RNAs were tested for HutP binding. (**d**) Effects of deoxy base substitutions at the UAG motif on HutP binding. To evaluate the 2′-OH group importance for the UAG motifs' recognition, three RNAs containing substitutions (dU or dA or dG), were prepared chemically, labeled at the 5′ end, and then used in the filter binding assay described in Figure 2.

was 30-fold lower ($K_d \sim 2.5$ µM) (Figure 4b). These site-specific substitutions as well as the earlier selection results suggest that the A base is mandatory at the second position for HutP recognition. Having determined the indispensability of A at this position, we wanted to identify the functional groups of this base involved in HutP binding. For this, we prepared shorter RNAs (21mers) containing the modified purine analog, 2-amino purine. The modified bases were incorporated by chemical synthesis at all three UAG motifs, at the second position. The modified RNA with 2-amino purine failed to bind to HutP, suggesting that the exocyclic NH_2 group of the A base plays an important role in HutP recognition (see Supplementary Material, Figure 7a). It is interesting to note that the orientation of the NH_2 functional groups of the A base at the second position requires the placement of an exocyclic NH_2 group at C6, and only at C6, since the addition of an amino group at C2, as in A_{2N} and G, abolishes binding.

Evaluation of the G base at the third position in the UAG motif

To address the importance of the G base at the third position, we prepared three RNAs where the G base was substituted by other bases (A/C/U) in all three UAG motifs in the small (21mer) RNA. The G base appears to be very important. Any substitution at this position caused a complete loss of HutP binding (Figure 4c), in contrast to the second position of UAG, which retained some binding ability with a U substitution. This correlates well with previous studies (18). To identify the potential functional group of the G base, we again used a modified base, in this case inosine, which lacks the exocyclic NH_2 group. We prepared a 21mer RNA containing inosine substituted specifically for the G base in all three UAG trinucleotides. The inosine-substituted RNA failed to bind to the HutP protein. This suggests that the exocyclic NH_2 group of the G base is also critical for HutP binding (see Supplementary Material, Figure 7b). To determine whether polarity for the G base recognition exists at the first, second or third UAG positions, we prepared three additional shorter RNAs, in which the G base is substituted by inosine in each of the three sites singly. All three variant RNAs showed slightly lower HutP binding affinity ($K_d \sim 87–123$ nM) as compared to the unmodified RNA. The lower affinity of these RNAs could originate from the loss of one G base interaction with the HutP. Taken together, the results of the *in vitro* selection and the site-specific substitution analyses clearly suggest that HutP recognizes each UAG motif in the terminator region of the *hut* mRNA. Also, it appears that at least two functional units of the UAG motif may be required to favor HutP binding. To test this notion, we synthesized a shorter 14mer RNA (5′-UUUAGUUUUUAGUU-3′) and tested it for binding. This RNA binds to HutP with 2-fold lower affinity ($K_d \sim 220$ nM) as compared to that of the 21mer RNA. However, further shortening to a 7mer RNA (5′-UUUAGUU-3′) abolished HutP binding, thus indicating that at least two UAG units are essential to achieve sufficient affinity for stable complex formation.

Evaluation of the 2′-OH in the UAG motif

From the above analyses, we identified that the A and the G bases were important in the UAG motif. However, we do not know how the HutP protein can differentiate the DNA from the
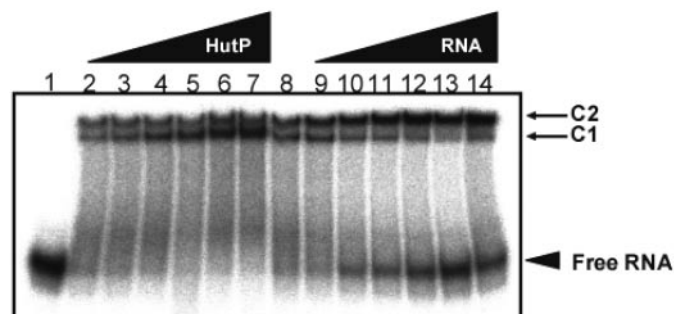

Figure 5. Gel mobility shift assay showing the HutP–79merRAT complex. All binding reactions were carried out in binding buffer in the presence of 10 mM L-histidine. Lane 1, labeled 79mer RNA 0.4 µM in the absence of HutP. Lanes 2 to 7, labeled 79mer RNA (0.4 µM) in the presence of increasing concentrations of HutP, as follows: lane 2, 0.62 µM; lane 3, 1.25 µM; lane 4, 2.5 µM; lane 5, 5.0 µM; lane 6, 7.5 µM; lane 7, 10.0 µM. Lanes 8 to 14 represent 1.25 µM of HutP with the labeled 79mer RNA (0.4 µM), plus increasing concentrations of unlabeled 79mer RNA, as follows: lane 8, 0.4 µM; lane 9, 0.8 µM; lane 10, 1.6 µM; lane 11, 2.4 µM; lane 12, 3.2 µM; lane 13, 4.0 µM; lane 14, 4.8 µM. The position of the free RNA is indicated by an arrow and the 1:1 and 1:2 complexes are indicated by C1 and C2, respectively.

RNA bases. In order to understand this issue, we have critically analyzed the 2′-OH interactions within the UAG motifs. To address the importance of the 2′-OH at the UAG motif, we prepared three RNAs where the UAG motif bases were substituted by other deoxy bases (dU/dA/dG) in all three UAG motifs in the small (21mer) RNA. The 2′-OH of the A appears to be very important. The deoxy A base substitution at this position caused a complete loss of HutP binding (Figure 4d), whereas the first and third position of UAG, which retained the binding ability with a dU and dG substitution. Thus, the critical contribution of the 2′-OH group of the A in the complex explains the ability of HutP–RNA complex to distinguish between the RNA and DNA complexes.

Stoichiometric analysis of HutP–RNA interactions by a gel mobility shift assay

In our earlier studies, we found that the HutP protein has two potential sites for a shorter RNA element (20mer) in the presence of L-histidine (15). However, the optimal ratio between HutP and its longer (79mer) RNA has not been established yet. For this, we carried out a gel-shift assay with the labeled 79mer RNA and increasing concentrations of activated HutP and RNA (unlabeled 79mer RNA). In both cases, we observed C1 (1:1) and C2 (1:2) ternary complexes (Figure 5). As revealed by the gel-shift assay, the 79mer RNA predominantly forms a 1:1 ratio in the presence of higher protein concentrations (Figure 5, lanes 2 to 7), while on the other hand, it predominantly forms a 1:2 ratio with higher RNA concentrations (Figure 5, lanes 8 to 14). The distance between the two shifts (C2 to C1) appeared to be the expected loss of one unit of 79mer RNA, $\sim$25 000 Da. From these analyses, it is clear that HutP has two potential sites for its cognate RNA, irrespective of the shorter (20mer) or longer (79mer) RNAs derived from the anti-terminator region (+459 to +537).

Potential RNA binding sites on HutP

In our previous studies, we defined the importance of the three UAG motifs and the requirement of spacer nucleotides
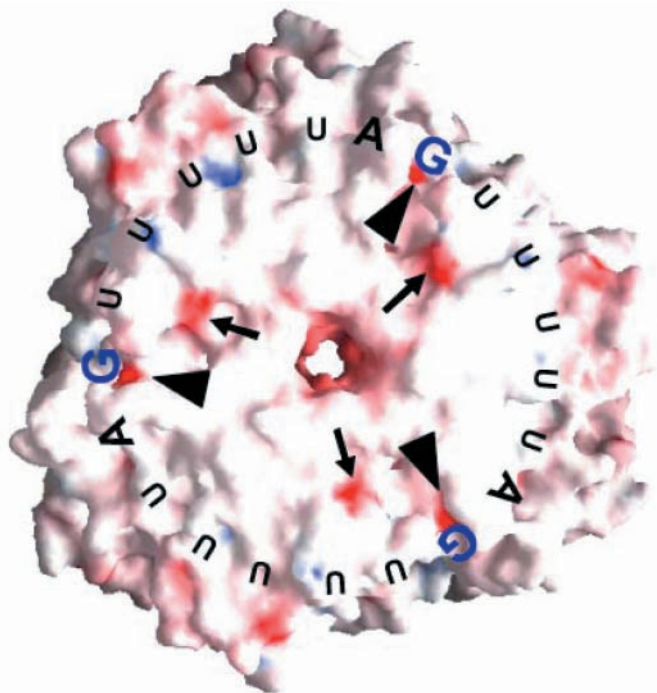
Figure 6. Potential RNA binding sites of HutP. Electrostatic potential of the molecular surface of hexameric HutP, based on the crystal structure of HutP. The electrostatic potential was calculated and visualized using GRASP (18). Basic regions are shown in blue and acidic regions are red. The electronegative potentials for the Glu55 residues are indicated by arrows and arrowheads, respectively. The important A and G bases for the HutP recognition are highlighted with bigger sized letters.

in between them (15). To substantiate our previous findings on the RNA identity elements, we carried out extensive mutational analyses to identify the functional groups within the UAG motif. The present work revealed that the NH_2 groups of the A and G bases greatly influenced the HutP–RAT binding. It appears from this analysis that the HutP interactions rely mainly on the base contacts within the RAT. Since the protein–RNA interactions responsible for binding in complexes are generally guided by the electrostatic potential, i.e. the donor–acceptor or acceptor–donor relationships in many complexes, we calculated the electrostatic potential surface for the hexameric HutP by GRASP (19), and searched for suitable acceptor sites that could accommodate the donor NH_2 group of an A and G base of the cognate RNA (Figure 6). The analysis shows that the HutP surface has very little electronegativity, which is seldom found in RNA binding proteins. Based on the above hypothesis, we identified two potential candidates (Glu55, Glu137) for the RNA recognition, and proposed that Glu137 may interact with the NH_2 group of the A or G base, based on the mutational analysis (15). A recent computational analysis on protein–RNA complexes suggested that Glu residues preferred to interact with G bases (propensity value, 1.43) rather than A bases (propensity value, 0.41). In addition, the formation of direct hydrogen bonds between the RNA bases and the amino acid residues of the protein suggests that the Glu residue preferred to interact with the G base, rather than the A base (20,21). Hence, we speculate that the NH_2 group of G interacts with Glu137, based

on the statistical analysis of existing protein–RNA complexes (Figure 6).

The other well-known example of U(G) AG recognition protein is TRAP, a 70-residue protein that regulates the expression of the L-tryptophan biosynthetic genes (trpEDCFBA) in several bacilli by binding to the 5′ non-coding leader region of the operon mRNA. TRAP binds to this RNA when activated by bound L-tryptophan to form a 'terminator' loop, which leads to transcription termination by preventing the formation of an 'anti-terminator' stem–loop structure (22–24). Moreover, TRAP can only bind to its target sequence with the nascent RNA transcript of the leader region of the operon when activated by L-tryptophan. The recognition sequence is organized as 11 GAG or UAG triplets, separated by two or three non-conserved nucleotides (25). The structure of TRAP complex with a 53-base RNA containing 11 GAG triplets separated by AU spacer has been solved and revealed the importance of the A2 and the G3 in the GAG triplets. Although no interaction between the protein and the phosphate backbone are observed, and only one hydrogen bond is established with the ribose 2′-OH of G3 in five of the 11 units (26). These hydrogen bonds seem to be critical, as RNA analogs with deoxyribose at G3 of all 11 repeats exhibit about 10 000-fold lower affinity to TRAP (27). Similar to the TRAP, for HutP also the A2 and the G3 bases are very important for the interactions with proteins, whereas the UAG triplets are separated by a 4 nt spacer in contrast to the TRAP, which requires a 2 nt spacer in between. More importantly, the 2′-OH of G3 is important for TRAP case whereas the 2′-OH of A2 is very critical for HutP binding. However, we have not analyzed the 2′-OH group interactions in the spacer regions, and these may provide some additional contacts. The notable differences are the TRAP recognizes 11 units of UAG/GAG for the interactions whereas the HutP protein requires 3 UAG motifs, and in contrast to TRAP, the hexameric HutP can interact with two RNA molecules. Taken together, it is interesting to observe that two kinds of non-homologous proteins recognize a similar motif (UAG) but with different specificity for their mode of action.

This is the first report describing the HutP binding core elements as an XAG motif with a 4 nt spacer length ($XAGX_5AG$-X_5AG; where X indicates any RNA base), which interacts in a cooperative manner. Also, the NH_2 groups of the A and G bases and 2′-OH of the A in the triplets were found to be essential for the HutP–RAT complex formation. Using this information, as well as the spacer length requirements, we have identified the potential RNA binding residues of HutP. Nevertheless, the ternary complex structure is essential, not only to confirm the key interactions proposed above, but also to reveal the mechanism of anti-termination and to add to our understanding of the transcriptional regulation of the *hut* operon in *Bacillus* species. Currently, these studies are in progress in our lab.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Richardson,J.P. and Greenblatt,J. (1996) Control of RNA chain elongation and termination. In Neidhardt,F.C., Curtiss,R.,III, Ingraham,J.L., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli* and *Salmonella: Cellular and Molecular Biology*. American society for Microbiology, Washington, DC, pp. 822–848.
2. Gollnick,P. (1994) Regulation of the *Bacillus subtilis* trp operon by an RNA-binding protein. *Mol. Microbiol.*, **11**, 991–997.
3. Rutberg,B. (1997) Antitermination of transcription of catabolic operons. *Mol. Microbiol.*, **23**, 413–421.
4. Henkin,T.M. (1996) Control of transcription termination in prokaryotes. *Annu. Rev. Genet.*, **30**, 35–57.
5. Oda,M., Katagai,T., Tomura,D., Shoun,H., Hoshino,T. and Furukawa,K. (1992) Analysis of the transcriptional activity of the hut promoter in *Bacillus subtilis* and identification of a *cis*-acting regulatory region associated with catabolite repression downstream from the site of transcription. *Mol. Microbiol.*, **6**, 2573–2582.
6. Wray,L.V.,Jr and Fisher,S.H. (1994) Analysis of *Bacillus subtilis* hut operon expression indicates that histidine-dependent induction is mediated primarily by transcriptional antitermination and that amino acid repression is mediated by two mechanisms: regulation of transcription initiation and inhibition of histidine transport. *J. Bacteriol.*, **176**, 5466–5473.
7. Chasin,L.A. and Magasanik,B. (1968) Induction and repression of the histidine-degrading enzymes of *Bacillus subtilis*. *J. Biol. Chem.*, **243**, 5165–5178.
8. Kimmhi,Y. and Magasanik,B. (1970) Genetic basis of histidine degradation in *Bacillus subtilis*. *J. Biol. Chem.*, **245**, 3545–3548.
9. Oda,M., Sugishita,A. and Furukawa,K. (1988) Cloning and nucleotide sequences of histidase and regulatory genes in the *Bacillus subtilis* hut operon and positive regulation of the operon. *J. Bacteriol.*, **170**, 3199–3205.
10. Yoshida,K., Sano,H., Seki,S., Oda,M., Fujimura,M. and Fujita,Y. (1995) Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the hut and wapA loci. *Microbiology*, **141**, 337–343.
11. Oda,M., Kobayashi,N., Ito,A., Kurusu,Y. and Taira,K. (2000) *Cis*-acting regulatory sequences for antitermination in the transcript of the *Bacillus subtilis* hut operon and histidine-dependent binding of HutP to the transcript containing the regulatory sequences. *Mol. Microbiol.*, **35**, 1244–1254.
12. Read,T.D., Peterson,S.N., Tourasse,N., Baillie,L.W., Paulsen,I.T., Nelson,K.E., Tettelin,H., Fouts,D.E., Eisen,J.A., Gill,S.R. *et al.* (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature*, **423**, 81–86.
13. Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B., Kapatral,V., Bhattacharyya,A., Reznik,G., Mikhailova,N., Lapidus,A. *et al.* (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature*, **423**, 87–91.
14. Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N., Fuji,F., Hirama,C., Nakamura,Y., Ogasawara,N. *et al.* (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res.*, **28**, 4317–4331.
15. Kumarevel,T.S., Fujimoto,Z., Karthe,P., Oda,M., Mizuno,H. and Kumar,P.K.R. (2004) Crystal structure of active HutP: A RNA binding protein that regulates *hut* operon in *Bacillus subtilis*. *Structure*, **12**, 1269–80.
16. Kumarevel,T.S., Fujimoto,Z., Padmanabhan,B., Oda,M., Nishikawa,S., Mizuno,H. and Kumar,P.K.R. (2002) Crystallization and preliminary X-ray diffraction studies of HutP protein: an RNA-binding protein that regulates the transcription of hut operon in *Bacillus subtilis*. *J. Struc. Biol.*, **138**, 237–240.
17. Yamamoto,R., Katahira,M., Nishikawa,S., Baba,T., Taira,K. and Kumar,P.K.R. (2000) A novel RNA motif that binds efficiently and specifically to the Tat protein of HIV and inhibits the trans-activation by Tat of transcription *in vitro* and *in vivo*. *Genes Cells*, **5**, 371–388.
18. Oda,M., Kobayashi,N., Fujita,M., Miyazaki,Y., Sadaie,Y., Kurusu,Y. and Nishikawa,S. (2004) Analysis of HutP-dependent transcription antitermination in the *Bacillus subtilis hut* operon: identification of HutP binding sites on *hut* antiterminator RNA and the involvement of the N-terminus of HutP in binding of HutP to the antiterminator RNA. *Mol. Microbiol.*, **51**, 1155–1168.
19. Nicholls,A., Bharadwaj,R. and Honig,B. (1993) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Biophys. J.*, **64**, A116–A125.
20. Kim,H., Jeong,E., Lee,S.W. and Han,K. (2003) Computational analysis of hydrogen bonds in protein-RNA complexes for interaction patterns. *FEBS Lett.*, **552**, 231–239.
21. Treger,M. and Westhof,E. (2001) Statistical analysis of atomic contacts at RNA–protein interfaces. *J. Mol. Recognit.*, **14**, 199–214.
22. Shimotsu,H., Kuroda,M.I., Yanofsky,C. and Henner,D.J. (1986) Novel form of transcription attenuation regulates expression of the *Bacillus subtilis* tryptophan operon. *J. Bacteriol.*, **166**, 461–471.
23. Babitzke,P. and Yanofsky,C. (1993) Recognition of *Bacillus subtilis trp* attenuation *in vitro* with TRAP, the *trp* rna-binding attenuation protein. *Proc. Natl Acad. Sci. USA*, **90**, 133–137.
24. Babitzke,P. (1997) Regulation of tryptophan biosynthesis: trping the TRAP or how *Bacillus subtilis* reinvented the wheel. *Mol. Microbiol.*, **26**, 1–9.
25. Babitzke,P., Stults,J.T., Shire,S.J. and Yanofsky,C. (1994) TRAP, the trp RNA binding attenuation protein of *Bacillus subtilis*, is a multisubunit complex that appears to recognize G/UAG repeats in the trpEDCFBA and trpG transcripts. *J. Biol. Chem.*, **269**, 16597–16604.
26. Antson,A.A., Dodson,E.J., Dodson,G., Greaves,R.B., Chen,X.P. and Gollnick,P. (1999) Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature*, **401**, 235–242.
27. Elliott,M.B., Gottlieb,P.A. and Gollnick,P. (1999) Probing the TRAP–RNA interaction with nucleoside analogues. *RNA*, **5**, 1277–1289.