BIOINFORMATICS

Vol. 14 no. 3 1998
Pages 271–278

Systematic genomic screening and analysis of mRNA in untranslated regions and mRNA precursors: combining experimental and computational approaches

Thomas Dandekar[1,2], Katrin Beyer[3], Peer Bork[1,2], Mary-Rose Kenealy[1], Kostas Pantopoulos[1], Matthias Hentze[1], Vera Sonntag-Buck[1], Gilles Flouriot[1], Frank Gannon[1], Walter Keller[3] and Sonja Schreiber[1]

[1]European Molecular Biology Laboratory, Postfach 102209, D-69012 Heidelberg, [2]Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rösslestr. 10, 13122 Berlin, Germany and [3]Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Received on December 1, 1997; revised on December 8, 1997; accepted on December 10, 1997

Abstract

Motivation: *The untranslated regions (UTRs) of mRNA upstream (5′UTR) and downstream (3′UTR) of the open reading frame, as well as the mRNA precursor, carry important regulatory sequences. To reveal unidentified regulatory signals, we combine information from experiments with computational approaches. Depending on available knowledge, three different strategies are employed.*
Results: *Searching with a consensus template, new RNAs with regulatory RNA elements can be identified in genomic screens. By this approach, we identify new candidate regulatory motifs resembling iron-responsive elements in the 5′UTRs of HemA, FepB and FrdB mRNA from Escherichia coli. If an RNA element is not yet defined, it may be analyzed by combining results from SELEX (selective enrichment of ligands by exponential amplification) and a search of databases from RNA or genomic sequences. A cleavage stimulating factor (CstF) binding element 3 of the polyadenylation site in the mRNA precursor serves as a test example. Alternatively, the regulatory RNA element may be found by studying different RNA foldings and their correlation with simple experimental tests. We delineate a novel instability element in the 3′UTR of the estrogen receptor mRNA in this way.*
Availability: *Strategy, methods and programs are available on request from T.Dandekar.*
Contact: *dandekar@embl-heidelberg.de*

Introduction

We have previously developed a general methodology to screen genomic databases for regulatory RNA motifs (Dandekar *et al.*, 1991; Dandekar and Hentze, 1995). After collecting a number of representative examples, a consensus RNA structure and essential sequence features are defined and translated into a search program to screen large genomic databases for new, further examples of this functional RNA motif. We have recently identified new eukaryotic iron-responsive elements (IREs; Gray *et al.*, 1996) and Rev-response-element-like structures in human mRNA (Dandekar and Koch, 1996).

We present here a more general treatment of regulatory structures in untranslated regions of mRNA and identify several novel elements with different function, both in the 5′UTR and 3′UTR of mRNA, or further downstream in the pre-mRNA precursor. Exploitation of knowledge available from experiments is essential for success. A consensus search can be relaxed in the requirements for RNA loop region residues according to available experimental data. New computationally identified RNA structures which may vary from the known consensus are identified (shown for prokaryotic IRE-like motifs). If less information is available on the exact RNA element, such as for mRNA 3′end processing protein recognition sites, information from SELEX (selective enrichment of ligands by exponential amplification) experiments is applied and translated into new templates for

© Oxford University Press

search programs. Genomic counterparts of the artificial selected RNA structures can then be searched. If even less is known about the structural requirements of a complex RNA element, correlation between stem–loop structures in RNA folds and experimental observations can be successful, as shown for a destabilizing element in the estrogen receptor mRNA.

Materials and methods

Computer programs were written in VAX-PASCAL. Testing steps and verification methods for consensus searches involved controls as previously described (Dandekar and Hentze, 1995). For instance, as a positive control for the IRE search, a carefully selected set of known IREs (Dandekar *et al.*, 1991; Gray *et al.*, 1996) had to be successfully retrieved by the program. Negative controls included failure to retrieve tRNA, rRNA or other highly structured RNA as false positives in database runs. The database screened was the EMBL database (Rodriguez-Tome *et al.*, 1996), Release 48. The IRE consensus (Hentze and Kühn, 1996) is shown in Figure 1.

The search for a 3′ processing motif starts with identification of a polyadenylation site (AAUAAA) in the sequence and then a search up to 50 nt downstream of the site for the 3′ processing site motif consisting of YAG (match all nucleotides; Y indicates pyrimidine) followed by UGYUAG [0–3 nucleotides (nt) interspersed]. If there were at least two matches in the latter, the program looked for GCUGNCC (0–5 nt interspersed; N indicates any nucleotide) and whether an A was present 3 nt before the last sub-motif.

The 3′UTR mRNA motif in the 3′UTR of the estrogen receptor mRNA is shown in detail in Figure 3.

Experimental

Cell lysates were prepared from *Escherichia coli* (Maniatis *et al.*, 1989). Autoradiograms of the radioactively labeled RNA structure were obtained after gel chromatography in the presence or absence of proteins binding the RNA element from *E.coli* cell extracts. Gel-retardation assays of IRE candidate structures were carried out as described previously (Gray *et al.*, 1996). T7 promotor constructs (T7 promotor and antisense of the RNA element as shown in Figure 1) were used to obtain RNA of the different candidate structures.

SELEX experiments (Türk, 1997) depended on purified cleavage stimulation factor (CstF) and subsequent selection of RNA ligands. Detailed protocols for the purification of cleavage stimulating factor CstF, selection of RNA ligands, and cleavage assays are described in Beyer *et al.* (1997).

Construction of the various deletions in the 3′UTR of the human estrogen receptor mRNA followed standard molecular biology techniques, as well as the determination of their stability by Northern analysis (Maniatis *et al.*, 1989) and transfection assays (Keaveney *et al.*, 1993).

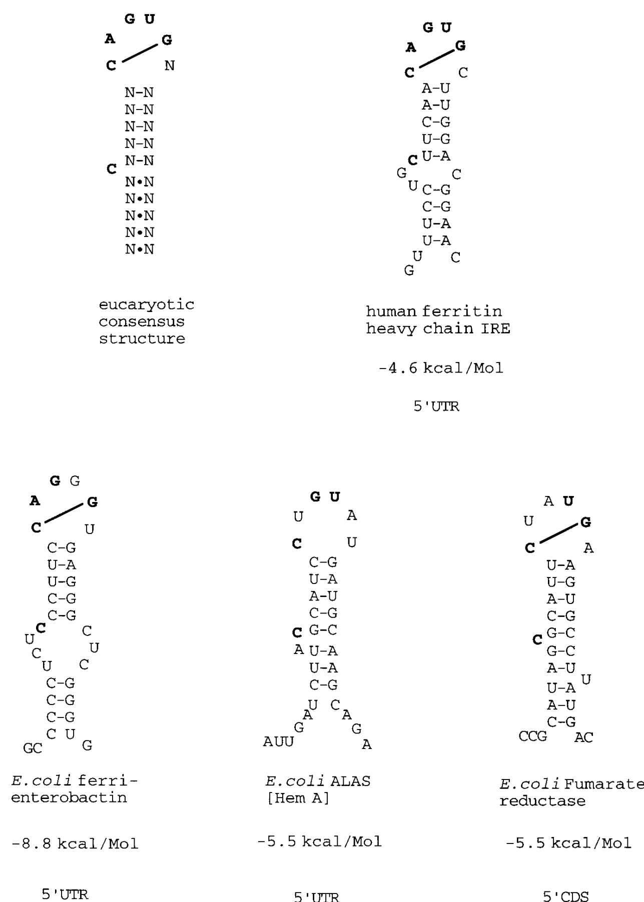

Fig. 1. New candidate iron-responsive elements found by computer screen of prokaryotic genomic sequences involving a relaxed loop consensus. Primary sequence and predicted secondary structure are shown. Conserved nucleotides in the loop and the bulged C are shown in bold. Name and functional position within the mRNA are indicated for each mRNA element. Energy for each structure according to Zucker and Stiegler (1981) is given at the bottom. **Top left:** Consensus structure for eukaryotic IREs (Hentze and Kühn, 1996). After the loop (with an additional interaction between the 5′ C and the G 1 nt before the 3′ end of the loop, sketched by a diagonal) follows top helix, a bulged C and a bottom helix. **Top right:** A known and verified control IRE in the 5′UTR of mRNA, human heavy chain for ferritin (Hentze *et al.*, 1987). **Bottom:** New identified prokaryotic IREs. Note their stable energies (bottom line) and the deviations from the loop consensus nucleotides. Exact positions in the mRNA (given is the position of the first C in the IRE-loop): *E.coli* ferrienterobactin transport protein (fepB; EMBL:ECFEPB) at nt 260, the coding sequence starts at nt 414; *E.coli* aminolevulinic acid synthase IRE at nt 1276 (EMBL:ECHEMA), the open reading frame starts at nt 1330; *E.coli* fumarate reductase (EMBL:ECAMPCFR) at nt 2660, note that the start of the coding sequence is at 2583, but extends to nt 3317 (so in fact slightly more 5′; however, the compact organization of this genomic region has already been noted earlier; Cole *et al.*, 1982).

Results

5′ Untranslated region

Relaxed consensus search. To complement previous search results for new IREs in eukaryotes, we present here a systematic analysis of prokaryotic sequences. IREs are regulatory motifs in the untranslated region of mRNAs. In the 5′UTR, their recognition by specific protein factors (IRPs) stops translation of the subsequent open reading frame (Hentze and Kühn, 1996). The binding affinity of the protein factor to the IRE changes according to iron content. Prokaryotic sequences of the EMBL data library were screened by search programs written in VAX-PASCAL (Release 48; 32 047 entries, 59 717 667 nt). The eukaryotic consensus IRE motif is shown in Figure 1 (top, left). It consists of a specific stem–loop structure with a second helix attached after a bulged C, and several obligatory nucleotides in the loop region. In the top helix, base pairing is more strictly required than in the bottom helix, where 3 bp appear sufficient. This consensus is then translated into a search program. Negative filters exclude incompatible nucleotides and unstable secondary structures. However, available experimental results indicated that the loop consensus (Figure 1, shown in bold for each IRE and the consensus) was not as strictly required as previously thought and hence the search was modified accordingly (non-bold nucleotides in loop regions of Figure 1 illustrate that deviations from the consensus were allowed. The CUGUA in HemA has not been found to exist in eukaryotic IREs; on the other hand, the exact recognition requirements of prokaryotic IRPs may deviate from the eukaryotic ones.) Candidate mRNA structures passing the consensus structure requirements were next examined for correct position within the mRNA (Table 1; Figure 1). The remaining RNA structures were checked in detail for biological evidence implicating involvement in iron metabolism. Three prokaryotic RNA structures remained, after all filters, as potential IRE candidate structures: FepB, FrdB and HemA mRNAs from *E.coli*. Figure 1 compares them to known IREs in structure and predicted stability. The prokaryotic IREs are interesting candidate structures for their similarity to the consensus, their predicted folding energy, and considering the known function of the encoded protein. The fumarate dehydrogenase is an iron–sulfur cluster protein and converts succinate to fumarate, similar to succinate dehydrogenase, but under anaerobic conditions in *E.coli* (Cole *et al.*, 1982). We showed for several iron–sulfur cluster proteins the translational regulation of their synthesis by IREs, including the mammalian mitochondrial aconitase and succinate dehydrogenase mRNA from *Drosophila* (Gray *et al.*, 1996). The aminolevulinic synthase mRNA in *E.coli* encodes the first enzyme in heme synthesis. Heme requires iron as the central metal ion to function, and translational regulation by iron levels has already been observed and confirmed for eukaryotic IREs known for murine and human eALAS (Dandekar *et al.*, 1991; Melefors *et al.*, 1993). The *E.coli* ferrienterobactin is involved in prokaryotic iron transport. Several eukaryotic mRNAs involved in iron transport or storage contain well-characterized IREs such as ferritin (in the 5′UTR) and the transferrin receptor (IREs in the 3′UTR) (reviewed in Hentze and Kühn, 1996).

Table 1. Reducing the number of RNA candidate structures from a database search in the 5′UTR

Example: Output of a database search for a regulatory element in the 5′UTR of mRNA		
Step	Searching for	Total sequences identified
Primary sequence	Specific nucleotides	Tens of thousands (in many genes)
Complete regulatory structure	Multipartite element	Several hundred (depending on motif complexity and database size)
1st filter	RNA?	(alternative: non-transcribed, DNA)
	mRNA?	Several sequences
2nd filter	Position?	(alternative: coding region, downstream)
	5′UTR?	Few sequences
3rd filter	Known?	(from literature; positive controls)
	New?	Selected candidates for experimental testing (e.g. only three structures left in the case of the presented search for prokaryotic iron-responsive elements)

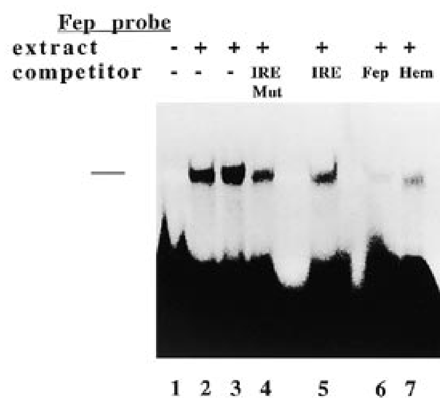

Fig. 2. Gel-retardation experiment of a new candidate structure found by genomic screening (IRE-like structure in FepB mRNA). The different IREs tested were obtained from T7 promotor constructs (see Materials and methods). Lane 1, FepB RNA without protein extract; lanes 2 and 3, FepB RNA shifted with *E.coli* lysate (3: only lane where iron, 100 µM ferric citrate, was added to the medium during growth). In lanes 4–7, FepB RNA is shifted with *E.coli* lysate and competed with cold competitor RNA. A point-mutated human ferritin IRE (Hentze *et al.*, 1987) is utilized in lane 4, unmutated human ferritin IRE (Hentze *et al.*, 1987) in lane 5, the putative Fep B IRE in lane 6 and *E.coli* HemA IRE in lane 7.

The IRE candidate structures were tested experimentally in gel-retardation assays. For each of the prokaryotic IREs, extracts from *E.coli* cells grown in L-broth contain a protein(s) that forms a complex with an IRE probe of the candidate structure. Complex formation is reduced by competition with an excess of the unlabeled candidate IRE. Figure 2 illustrates this for the Fep IRE (lanes 1, no extract; lanes 2 and 3, putative IRE binding in the presence of *E.coli* extract). Controls shown include successful competition by cold IRE (lane 6; no bandshift) as well as cross-competition by other prokaryotic candidate IREs (lane 7) and eukaryotic IREs (lanes 4 and 5, human ferritin IRE). Putative IRE binding in extracts from *E.coli* cells grown in the presence of iron (100 µM ferric citrate) are shown in lane 3 (see Discussion).

Once candidate RNA structures are identified and analyzed, detailed experimental characterization can be carried out. Starting from known RNA structures, new RNA candidate structures with potentially similar function can be identified by a modified, relaxed consensus search combined with experimental testing. The 3′UTR may be searched by identical programs, only the screening filters (Table 1) have to be changed accordingly. However, our strategy has to be more strongly modified if the regulatory RNA element is less well characterized. This is demonstrated in the next two examples.

3′ Untranslated region and downstream of the polyadenylation site

Motif search complementing a SELEX approach. If no genomic template structures are available, different genomic search strategies for regulatory RNA must be sought. As an alternative to approaches such as neuronal networks (Ogura *et al.*, 1997), language-based approaches (Trifonov, 1996) and other non-consensus search methods (e.g. Tiwari *et al.*, 1997), we present here the combination of artificial *in vitro* evolution and genomic screening. The computer-based genomic screen delineates how close the *in vitro* selection procedure comes to the situation *in vivo* (at least according to the genome content).

An example is the case of downstream elements required for polyadenylation. The pre-mRNA 3′end processing factor CstF was purified from HeLa cells and from calf thymus, and used in filter binding assays. RNA populations with high binding affinity to CstF were selected by several rounds of SELEX (Türk, 1997). The sequence of the selected RNAs was determined by cloning and sequencing. These *in vitro* results were then compared with data from natural organisms by genomic screening procedures. Conserved RNA features were determined first and next translated into a simple but powerful search program identifying polyadenylation sites and looking for a selected tri-partite RNA motif (see Materials and methods) downstream of the polyadenylation site in genomic sequences. Complete results of the genomic search are shown in Table 2. To identify high-score hits, only cases where there were more than two matches at the important first part of the motif, YAG, were screened further. As a second filter, more than two matches in the next 5 part of the motif were required for further screening (steps in the distribution in Table 2). To identify low-score hits in the 3 part of the motif (e.g. to identify new protein factors interacting with the downstream part of the RNA motif and perhaps CstF), our search can be easily modified (protocols available from T.D.).

Example sequences found are shown in Table 3. They identify genomic sequences resembling the consensus in several parts. In most cases where a potential polyadenylation site is present, some match with the *in vitro* selected RNA consensus structure can be found. The peak fraction of all hits has 10 matches. However, if the exact position of the element within the mRNA is considered (e.g. directly after the first AAUAAA), as well as its absence in non-mRNA encoding sequences (intergenic DNA, introns, other RNA types), the specificity of this motif is not sufficiently high. The SELEX experiment selected this motif as one with high affinity to the 3′ processing protein factor CstF. Our genomic search indicates that this particular motif is not the exact RNA element used *in vivo*, otherwise there would be more high-score matches and only a few matches would be found

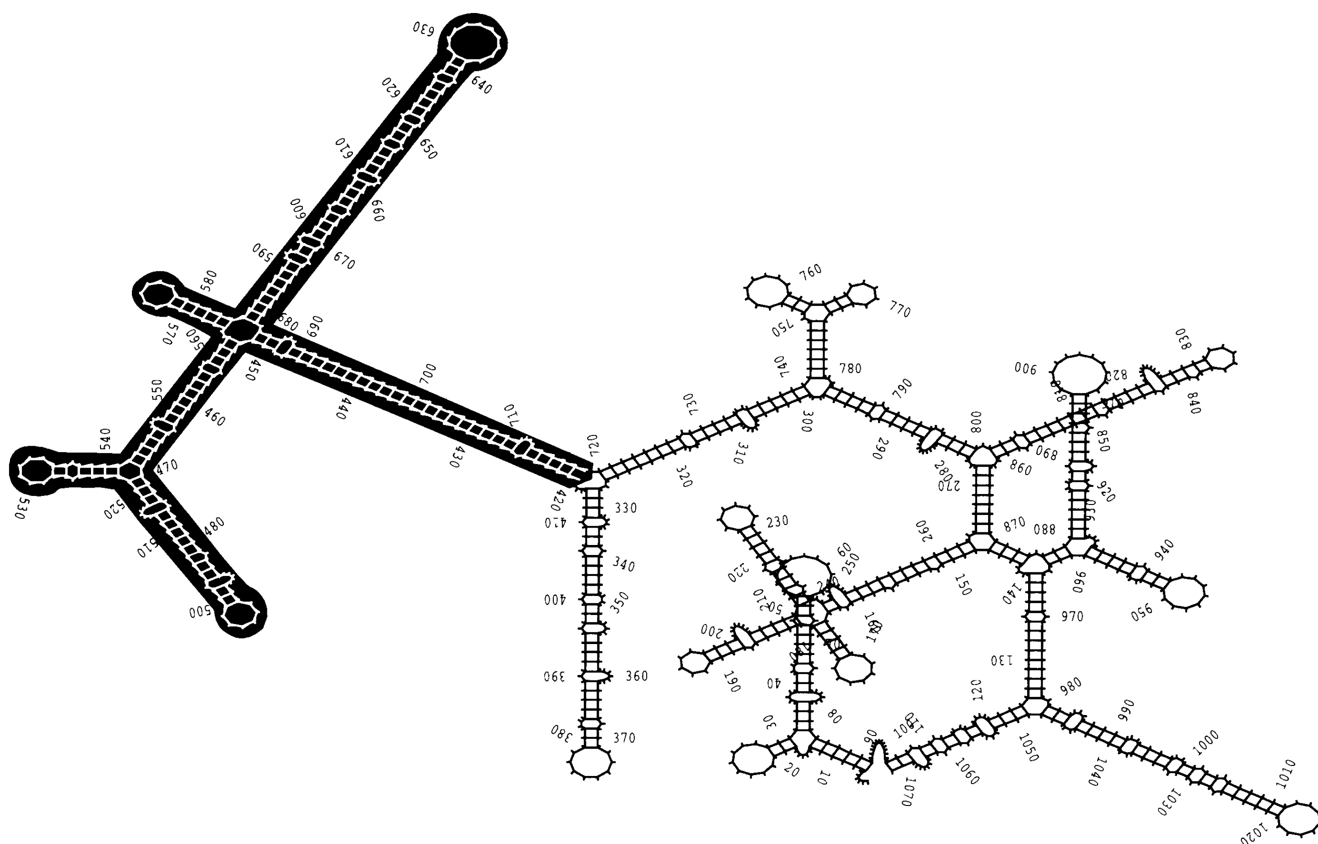

Fig. 3. RNA structure in the 3′UTR of the estrogen receptor mRNA (mRNA from entry EM:HSERR). Shown are nt 1375–2446 of the 3′UTR. The element marked on the left (nt 420–720 of this fragment) is the region suggested to contain an RNA instability element. The appearance/disappearance of the two long stems (base pairing of nt 420–450 against 682–720, as well as the second stem of 584–628 against 638–680) in folding different subsets of the mRNA correlates with the instability of the respective construct (Table 4).

in non-mRNA sequences. The computational approach thus identifies limitations of the SELEX procedure and the *in vitro* selection result. Further experiments are required to clarify the 3′-processing motif better. Experimental details and additional SELEX-based searches, including further motifs and their refinement by additional experimental data, can be found in Beyer *et al.* (1997).

The above example shows how genomic RNA motif searches can be extended to RNA motifs where no template is known by exploiting experimental information from *in vitro* experiments. At the same time, they compare the RNA motif found *in vitro* with the RNA signals as they are encoded in the genome.

Detailed analysis of a complex RNA stability motif in the 3′UTR. The example above considered a linear RNA motif. However, more complex RNA motifs in untranslated mRNA can be delineated. In this further extension of our combined strategy, different RNA folds are studied and correlated with experimental results. This template-free strategy is illustrated in Table 4. The long 3′UTR of human estrogen receptor mRNA contains a signal mediating rapid degradation of the mRNA (Keaveney *et al.*, 1993; Kenealy *et al.*, 1996). The behavior of this long untranslated region was now studied experimentally by testing different domains (indicated in the left column) and measuring the mRNA half-life (middle column). To identify the regulatory RNA motif, each RNA segment was folded independently, as well as different segments (F1–F3). The stem–loop structures obtained were compared with each other and the folding of the complete 3′UTR. Only the complex structure with two long stem–loops shown in Figure 3 correlated exactly (Table 4, right column) with the appearance and disappearance of the mRNA instability conferred by the different versions of the 3′UTR. New experi-

ments will test the structure delineated and the processing of this long 3′UTR. Furthermore, the element delineated can now be used as a template for further genomic searches as indicated above, both to reveal similar elements in other RNAs and to test the utilization of such signals in the rest of the genome.

Table 2. Hits for the RNA motif search on a SELEX 3′-processing motif in the EMBL database (Release 48)

	phg	vrl	pro	org	fun	pln	inv	vrt	mam	rod	pri	hum2
0–2 matches	86	1591	1764	949	978	2115	2072	1266	1188	3564	2784	1077
3 matches	3	63	25	8	14	13	39	36	58	107	103	44
4 matches	2	40	63	17	25	34	65	85	44	176	143	46
5 matches	53	1163	831	401	604	494	763	332	303	828	882	338
6 matches	0	95	7	3	3	7	9	14	8	33	28	15
7 matches	1	526	40	24	23	50	90	44	54	128	98	49
8 matches	34	1207	529	421	642	445	674	191	153	475	532	195
9 matches	81	3048	1790	1283	1334	1074	1514	505	393	1264	1362	512
10 matches	125	3965	2698	1408	2086	1335	2038	710	603	1701	2194	1046
11 matches	104	1788	2047	806	1542	788	1872	489	358	1324	1458	566
12 matches	27	418	563	220	544	187	857	212	124	541	542	223
13 matches	15	59	134	37	89	44	121	78	13	83	117	45
14 matches	0	0	2	4	6	3	17	3	2	21	9	7

The total number of hits for each organism category in the EMBL data library and a particular number of matches to the RNA motif is shown. The motif is given in Materials and methods, examples are found in Table 3.

Table 3. Processing motif 3′ of the polyadenylation site in the mRNA precursor

SELEX experiment; example sequence from one of the selected clones (B2):

uagguguuagaaggcuggcc

Consensus sequence comparing all sequenced clones with high affinity to CstF derived from SELEX:

yagN_{0-3}ugyuagN_{0-5}gcugncc

(plus an A three nucleotides before the third part of the motif)

Genomic examples with similarities to the consensus[a]:

EM:HSORF41 downstream of the polyadenylation site:

1940 **aauaaa**,acauaauagaauuuccuaaugagauauaucuuuauacuuaaa,**cag,**cuu,uuuuagaggugaguuu

EM:HSORFE1 downstream of the polyadenylation site:

1200 **aauaaa**,cauggccaugaacucagggccauugggauucucccucuggcagccc, **cag**,g,ugggagcgcugggaggg

EM:HSORFP downstream of the polyadenylation site:

4356 **aauaaa**, ugugcaaacugugcaaacugcaucuauaggaaacauuuuugugauuacggaauacuu,**uag**,u,ugauugc,gcugaaaau

EM:HSP1BX (human secretory protein (P1.B) mRNA downstream of the polyadenylation site:

439 **aauaaa**, ggucccaugcuccacccgagga,**cag**,u,ucuucgugccug [3′end]

[a]Examples from human genomic sequences are shown. The nucleotide position according to the complete entry is given for each example considering the U in the aauaaa polyadenylation site of the mRNA precursor. The first three nucleotides of the region similar to the RNA motif are indicated in bold, separators identify further parts of the motif.

Table 4. Testing of different mRNA constructs

Construct		mRNA degradation (half-life in h)	mRNA element present[a] (computed folding)
Complete 3′UTR		2–3	yes
UTR1	1–1375	14	no
UTR2	1375–2446	3–4	yes
UTR3	2446–3476	18	no
UTR4	3476–end	10	no
UTR2 a	1375–2009	11	no
UTR2 b	2010–2149	>20	no
UTR2 c	2150–2446	18	no
UTR2 a + b	1375–2149	6	yes
UTR2 b + c	2010–2446	20	no
Control foldings			
F1	1–2000	n.d.	no
F2	1501–3501	n.d.	yes
F3	2501–4300	n.d.	no

n.d., half-life not determined experimentally.
[a]The RNA element is shown in Figure 3.

Discussion

A consensus search for RNA elements in untranslated portions of mRNA can be extended to further examples and a less strict consensus, exploiting experimental information. The new prokaryotic candidate IREs identified in the 5′UTR by our approach can be folded into a structure like the IRE consensus and have similar stability. Furthermore, biological evidence as to the function of the encoded proteins suggests a connection with iron metabolism. All these data, including the first gel-retardation experiments of the radiolabeled *in vitro* transcribed putative IREs with *E.coli* extract, suggest these IREs to be genuine iron-responsive elements. However, more experiments are necessary to clarify and confirm their functionality. This would include further competition experiments, change in association with polysomes and swapping experiments to render another mRNA iron dependent after insertion of the putative IRE (Dandekar *et al.*, 1991). It is also necessary to obtain more direct evidence for regulation by cellular iron content. Some changes in binding when comparing cell extracts grown in media with different iron content have been observed in preliminary experiments, but these have to be repeated and analyzed further. The elements identified have variations in the loop region from the tight consensus known for many IREs from eukaryotes (see Figure 1; Figure 2, competition in lanes 4 and 5), notable HemA. A further step in validating their function as genuine IREs will require isolation of the binding protein partner from *E.coli* to show that a true IRP equivalent is present in *E.coli*. The RNA elements revealed here could be used for affinity selection of protein interaction partners or in other experiments aimed at the identification of their prokaryotic binding protein.

Apart from identifying novel RNAs carrying a regulatory RNA motif, the approach is also applicable to other searches involving stem–loop structures in the 5′UTR or 3′UTR of mRNA or the mRNA precursor.

We next demonstrate protocols on how a search strategy can be extended to RNA motifs where no genomic template is available. Template delineation by SELEX experiments and critically analyzing the SELEX result by computer-based genomic searches is one approach. Correlating RNA folds with experimental observations was a second method applied to identify a more complex RNA structure in the 3′UTR.

Although the SELEX RNA pool analyzed here by our approach has high binding affinity, the computer-based genomic search procedure is able to show that the particular RNA motif selected is not the exact signal used in the genomic sequences. Several different RNA motifs seem to be utilized in 3′mRNA processing and partly overlap (Beyer *et al.*, 1997). High affinity of the RNA motif obtained from the SELEX experiment for the binding partner (here CstF, the cleavage stimulating factor) is a necessary but not sufficient condition for efficient function *in vivo*.

The two long stem–loops delineated from the complex RNA instability element in the estrogen receptor mRNA probably indicate another protein interaction. Further studies will reveal the exact primary and secondary structure re-

quirements of the RNA element, including detailed analysis of mRNA processing and instability.

These different protocols to extend genomic searches are presented as an alternative to a number of other strategies for nucleic acid motif searches available (e.g. Gutell *et al.*, 1992; Ogura *et al.*, 1997; Tiwari *et al.*, 1997; Trifonov, 1997; further protocols are reviewed in Dandekar and Hentze, 1995). Our protocols are particularly useful to exploit and combine experimental data with computational prediction.

References

Beyer,K., Dandekar,T. and Keller,W. (1997) RNA-ligands selected by cleavage stimulation factor (CstF) contain distinct sequence motifs that function as downstream elements in 3′-end processing or pre-mRNA. *J. Biol. Chem.*, **272**, 26769–26779.

Cole,S.R., Grundström,T., Jaurin,B., Robinson,J. and Weiner,J.H. (1982) Location and nucleotide sequence of frdB, the gene coding the iron-sulphur protein subunit of the fumarate reductase of *Escherichia coli*. *Eur. J. Biochem.*, **126**, 211–216.

Dandekar,T. and Hentze,M.W. (1995) Finding the hairpin in the haystack. *Trends Genet.*, **11**, 45–50.

Dandekar,T. and Koch,G. (1996) DNA and mRNA sequence of the immune protective DNA ligase I gene match the rev response element of HIV. *DNA Seq.*, **6**, 119–121.

Dandekar,T., Stripecke,R., Gray,N.K., Goosen,B., Constable,A., Johansson,H.E. and Hentze,M.W. (1991) Identification of a novel iron-responsive element in murine and human erythroid delta-aminolevulinic acid synthase. *EMBO J.*, **10**, 1903–1909.

Gutell,R.R., Power,A., Hertz,G.Z., Putz,E.J. and Stormo,G.D. (1992) Three dimensional constraints on the higher order structure of RNA: Continued development and application of computational sequence analysis methods. *Nucleic Acids Res.*, **20**, 5785–5795.

Gray,N.K., Pantopoulos,K., Dandekar,T., Ackrell,B.A.C. and Hentze,M.W. (1996) Translational regulation of mammalian and *Drosophila* citric acid cycle enzymes via iron-responsive elements. *Proc. Natl Acad. Sci. USA*, **93**, 4925–4930.

Hentze,M.W. and Kühn,L.C. (1996) Molecular control of vertebrate iron metabolism: Circuits operated by iron, nitric oxide, and oxidative stress. *Proc. Natl Acad. Sci. USA*, **93**, 8175–8182.

Hentze,M.W., Caughman,S.W., Rouault,T.A., Barriocanal,J.G., Danc,A., Harford,J.B. and Klausner,R.D. (1987) Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science*, **238**, 1570–1573.

Keaveney,M., Parker,M.G. and Gannon,F. (1993) Identification of a functional role for the 3′ region of the human oestrogen receptor gene. *J. Mol. Endocrinol.*, **10**, 143–152.

Kenealy,M.R., Flouriot,G., Pope,C. and Gannon,F. (1996) The 3′UTR of the human estrogen receptor gene post-transcriptionally reduces mRNA levels. *Biochem. Soc. Trans.*, **24**, 107S.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Melefors,Ö., Goosen,B., Johansson,H.E., Stripecke,R., Gray,N.K. and Hentze,M.W. (1993) Translational control of 5-aminolevulinate synthase mRNA by iron-responsive elements in erythroid cells. *J. Biol. Chem.*, **268**, 5974–5978.

Ogura,H., Agata,H., Xie,M., Odaka,T. and Furutani,H. (1997) A study of learning splice sites of DNA sequence by neural networks. *Comput. Biol. Med.*, **27**, 67–75.

Rodriguez-Tome,P., Stoehr,P.J., Cameron,G.N. and Flores,T.P. (1996) The European Bioinformatics Institute (EBI) databases. *Nucleic Acids Res.*, **24**, 6–12.

Tiwari,S., Ramachandran,S., Bhattacharya,A., Bhattacharya,S. and Ramaswamy,R. (1997) Prediction of probable genes by Fourier analysis of genomic sequences. *Comput. Applic. Biosci.*, **13**, 263–270.

Trifonov,E.N. (1996) Interfering contexts of regulatory sequence elements. *Comput. Applic. Biosci.*, **12**, 423–429.

Türk,C (1997) Using the SELEX combinatorial chemistry process to find high affinity nucleic acid ligands to target molecules. *Methods Mol. Biol.*, **67**, 219–230.

Zuker,M. and Stiegler,P. (1981) Optimal computational folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.