246–254 Nucleic Acids Research, 2001, Vol. 29, No. 1

ARED: human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins

Tala Bakheet, Mathias Frevel[1], Bryan R. G. Williams[1], William Greer and Khalid S. A. Khabar[1,2,*]

Department of Biostatistics, Epidemiology and Scientific Computing (Bioinformatics Section) and [1]Department of Biological and Medical Research, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia and [2]Department of Cancer Biology, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH 44195, USA

Received August 16, 2000; Revised and Accepted October 2, 2000

ABSTRACT

The adenylate uridylate-rich elements (AREs) mediate the rapid turnover of mRNAs encoding proteins that regulate cellular growth and body response to exogenous agents such as microbes, inflammatory and environmental stimuli. However, the full repertoire of ARE-containing mRNAs is unknown. Here, we explore the distribution of AREs in human mRNA sequences. Computational derivation of a 13-bp ARE pattern was performed using multiple expectation maximization for motif elicitations (MEME) and consensus analyses. This pattern was statistically validated for the specificity towards the 3′-untranslated region and not coding region. The computationally derived ARE pattern is the basis of a database which contains non-redundant full-length ARE-mRNAs. The ARE-mRNA database (ARED; http://rc.kfshrc.edu.sa/ared) reveals that ARE-mRNAs encode a wide repertoire of functionally diverse proteins that belong to different biological processes and are important in several disease states. Cluster analysis was performed using the ARE sequences to demonstrate potential relationships between the type and number of ARE motifs, and the functional characteristics of the proteins.

INTRODUCTION

Adenylate uridylate-rich element (ARE)-containing genes include a number of the early response genes that regulate cell proliferation and responses to exogenous agents. AREs have been functionally attributed to a restricted number of mRNAs such as certain hematopoietic cell growth factors (e.g. granulocyte–monocyte colony stimulating factor, GM-CSF), interleukins, interferons, TNF-α and some proto-oncogenes (e.g. c-fos, k-ras and pim-1); Table 1 shows a list of previously known ARE-containing mRNAs. These gene products have been shown to play an important role in different cancers and inflammation states (1–3). However, the full repertoire of ARE-containing mRNAs, referred to hereafter as ARE-mRNAs, is not known, and hence the range of possible consequences of dysregulation of the ARE-mRNA decay pathways remains undetermined. In this paper, we explore the repertoire of human ARE-mRNAs using a bioinformatics approach with the compilation of ARE-mRNA sequences in a database.

BACKGROUND: BIOLOGY OF AU-RICH ELEMENTS

A common trait of the ARE-mRNAs is that they are short-lived and thus rapidly disappear once their critical role in gene regulation ceases. ARE-mediated changes in mRNA stability are important in processes that require transient responses such as cellular growth, immune response, cardiovascular toning and external stress-mediated pathways. Thus, stabilization of the ARE-mRNAs can cause a prolonged response that may subsequently lead to a diseased state. Identifiable AREs in the 3′UTR of the mRNA were noted such as the pentameric and nonameric sequences of AUUUA and UUAUUUAUU. The minimal ARE sequence has been shown to be the nonamer rather than the pentamer (4–7). Shaw and Kamen (8) showed that the stable β-globin mRNA could be rendered unstable when its 3′UTR was replaced with the GM-CSF multiple ARE-containing 3′UTR. Several groups have described ARE-binding proteins that influence the ARE-mRNA stability. Among the well-characterized proteins are the mammalian homologs of ELAV (embryonic lethal abnormal vision) proteins including AUF1, HuR and He1-N2 (9–11). The zinc-finger protein tristetraprolin has been identified as another ARE-binding protein with destabilizing activity on TNF-α, IL-3 and GM-CSF mRNAs (12,13). Other non-ARE regions of certain mRNAs, including c-myc, histone and the transferrin receptor, have also been implicated in the stability of mRNA. The recent emphasis in literature is on signaling mechanisms, notably the role of stress activated protein kinases, p38 MAP kinase, in the ARE-mediated stabilization of certain ARE-mRNAs, e.g. IL-8, cyclooxygenase-2 and vascular endothelial growth factor (14–16).

*To whom correspondence should be addressed at: Head, Interferon and Cytokine Research Unit (MBC-03), Senior Scientist, Department of Biological and Medical Research, PO Box 3354, MBC-03, Riyadh 11211, Saudi Arabia. Tel: +966 1 442 7876; Fax: +966 1 442 7858; Email: khabar@kfshrc.edu.sa

Table 1. A list of previously known ARE-mRNAs

Gene name/function	Ref.
Early lymphocyte activation antigen CD69	(34)
6-phosphofructo-2-kinase (PFK-2)/fructose-2,6-bisphosphatase	(35)
B-cell leukemia/lymphoma2 oncogene (Bcl-2)	(27)
c-fos proto-oncogene (transcription)	(36)
CHOP/Growth arrest and DNA-damage inducible factor	(37)
c-myb proto-oncogene (transcription)	(38)
c-myc proto-oncogene (transcription)	(39)
Cyclin D1 (cell cycle)	(40)
Cyclooxygenase (inflammation)	(15)
Endothelin-2 (vascular toning)	(41)
Epidermal growth factor receptor	(42)
Estrogen receptor α	(43)
Fibroblast growth factor 2	(44)
Granulocyte colony stimulating factor	(38,45)
Glucose transporter 1	(46)
Granulocyte monocyte colony stimulating factor	(8,14)
Gro-α, growth-regulated gene	(47)
Inducible nitric oxide synthase	(48)
Interferon-α (immune response, innate)	(38,49)
Intereferon-αAA[a]	(49)
Interferon-α1	(38,49)
Interferon-α1B	(49)
Interferon-αF	(38,49)
Interferon-αG	(38,49)
Interferon-αH	(38,49)
Interleukin-1α (inflammation)	(50)
Interferon-β (immune response, innate)	(51,52)
Interferon-γ (immune response, adaptive)	(53)
Interleukin-1β (inflammation)	(54)
Interleukin-10 (immune response)	(55)
Interleukin-2 (immune response, adaptive)	(56,57)
Interleukin-3 (hematopoiesis)	(12)
Interleukin-4 (immune response)	(38)
Interleukin-6 (immune response)	(14)
Interleukin-8 (inflammation)	(14)
Interleukin-11 (adipogenesis inhibitory factor)	(58)
Lymphotoxin (immune response, inflammation)	(38)
K-ras proto-oncogene (signal transduction)	(59)
Leukemia inhibitory factor (LIF)	(60)
Macrophage colony stimulating factor	(61)
Macrophage chemotaxis protein-1	(62)
Macrophage inflammatory protein-α	(63)
Macrophage inflammatory protein-β	(55)
Macrophage inhibitory protein-2α	(64)
Mda-7 (melanoma differentiation-associated gene)	(65)

Table 1. *Continued*

Gene name/function	Ref.
Monocyte Chemotactic Protein-3	(66)
MYCN (transcription)	(67)
Nerve growth factor	(49,68)
Platelet-derived growth factor/c-sis protooncogene	(69)
Pim-1 proto-oncogene (singal transduction)	(70)
Plasminogen activator inhibitor type 2	(71)
Thioredexin reductase (metabolism, redox)	(72)
Tissue factor (thromboplastin, coagulation)	(73)
Tumor necrosis factor (inflammation, others)	(2,6,74)
Urokinase-type plasminogen (uPA) receptor	(75)
Urokinase-type plasminogen activator	(24)
Vascular endothelial growth factor	(16)

The list has been compiled using PubMed literature search according to criteria described in the text. Gene name/function were given according to the most characterized function in conjunction with AREs. If the gene name does not correspond to a biological function, a function category in brackets is given. This list was used as a training set for computational derivation of ARE motif.
[a]There are more than 20 IFN-α subtypes.

COMPUTATIONAL DERIVATION OF THE ARE MOTIF AND A NON-REDUNDANT HUMAN ARE-mRNA DATABASE

Sequence retrieval and analysis utilized the GCG-Wisconsin Package (Genetics Computer Group/Oxford Molecular Company, Madison, WI) and additional programs written in the Practical Extraction and Report Language (PERL). A total of 36 951 human mRNA/cDNA sequences were extracted from GenBank Release 113 (National Center for Biotechnology Information, NCBI) using Lookup program (GCG codes) that was used to find mRNA or cDNA in the Definition Field along with *Homo sapiens* in the Organism Field (Source) in GenBank entries. Subsequently, a PERL code was written to extract the sequences that contain the field CDS in the Features Table to exclude those that do not have CDS; this resulted in 27 403 CDS-containing mRNA/cDNA sequences. This file was used as the input to another PERL program that extracted sequences as with complete CDS, i.e. without ambiguous CDS such as <, >, complement or join. The output was 15 148 sequences as full-length CDS-containing mRNA/cDNA file. The 3′UTRs were constructed using Assemble program (GCG codes) which extracted the sequences downstream of CDS (i.e. >CDS). This step was necessary because most of the GenBank records lack the 3′UTR as an annotated Feature key—this limitation of annotation was previously noted (17)—despite the fact that this information can be extracted computationally from CDS Feature as executed here. The UNIX command, Stream Editor (Sed), was used to remove sequences that have no 3′UTR. A resultant list of 13 057 human full-length CDS/3′UTR-containing mRNA sequences was finally compiled.

The minimum functionally relevant ARE motif has been previously reported to be UUAUUUA (A/U)(A/U). However, this consensus was based only on a limited number of AU-rich mRNAs such as GM-CSF, c-fos, TNF-α etc. (5–7,9). In the present analysis, a list of GenBank entries (Table 1) that belong to 57 previously recognized ARE-containing mRNAs was used as a training set for the computational derivation of the ARE motif. The selection of the ARE-mRNAs in this training list was based on either of two criteria: (i) mRNAs in which their AREs in the 3′UTR have been experimentally determined to affect their mRNA turnover, e.g. GM-CSF, c-fos, TNF-α and IFN-β; or (ii) mRNAs without available experimental evidence of an ARE-mediated mRNA decay but that contain AREs in their 3′UTR sequence and the mRNAs are transiently induced such as the case of IFN-α subtypes and k-ras. The 3′UTR sequences of these entries were extracted computationally as described above, and cleaned from the poly(A) tails using a PERL code to reduce its recognition as a pattern. The 57 3′UTRs were then plugged into the MEME (multiple expectation maximization for motif elicitations) program which finds conserved ungapped short motifs within a group of related, unaligned sequences (18). MEME yielded the motif pattern UAUUUAWW, which is very similar to the previously reported nonamer UUAUUUAWW. Next, we performed a consensus analysis around this motif, which resulted in the pattern WWWUAUUUAUWWW with a certainty level of 75% at each position. Subsequently, this 13-bp pattern was used by the FindPattern analysis (GCG codes). The stringency was decreased by allowing one mismatch in each direction of the nucleotides flanking the core pattern (UAUUUAU), in order to allow maximum recovery from the search. This step was performed on the 3′UTRs of the full-length CDS/3′UTR-containing mRNA list (13 057 sequences). The resulting subset of sequences was made minimally redundant using the CLEANUP program (19) with the parameters of 90% similarity and 90% overlap, which produced an output file that contained

Table 2. Statistical characteristics of the computationally-derived ARE pattern in 3′UTR and CDS

	3′UTR				CDS				*P* value[f]
	No.[a]	Finds[b]	Mean[c] ± S.D.	%[d]	No.	Finds	Mean ± S.D.	% sp[e]	
Mismatch = 0	276	349	1.3 ± 0.7	31	2	3	N/A	>99	N/A
Mismatch = 1	736	3670	4.9 ± 6	82	27	50	1.85 ± 3	96	0.0001
Mismatch = –2	897	9781	10.9 ± 12	100	98	233	2.37 ± 3.7	89	0.0001

The ARE-mRNA list of 897 was verified against 3′UTR and CDS for the specificity and database coverage of the 13-bp pattern under different search stringency (e.g. with one mismatch and two mismatches in nucleotides flanking the conserved core) used for computational compilation of ARED.
[a]Number of mRNA sequences retrieved by the search.
[b]The number of ARE patterns found in each subset.
[c]Mean of finds of the 13-bp ARE pattern per 3′UTR or CDS.
[d]Percentage coverage = % (number of 3′UTR with ARE pattern /total 897 mRNA sequences).
[e]Percentage specificity (% sp) = 1 – (CDS containing the pattern/total 897 mRNA sequences).
[f]*P* values indicate statistical significance between the mean of 13-bp ARE pattern per ARE-mRNA using an unpaired *t*-test with Welch correction (used because of the significantly different variances as verified by F test, $P < 0.0001$).
N/A = not applicable due to the small number of finds.

the longest available sequences. Approximately 17% redundancy in the ARE-mRNA list was computationally removed. A total of 897 minimally redundant sequences (~8% of the human mRNA sequences analyzed) were finally obtained and subsequently termed the 'ARE-mRNA database' (ARED). This database was stored as flat GenBank files and imported for further analysis into the commercial Vector NTI software version 5.5 (InforMax, Bethesda, MD).

In order to validate the specificity/sensitivity of the computationally-derived 13-bp ARE pattern in the 3′UTR (e.g. in comparison to CDS), we searched for the pattern in the CDS (Assemble and FindPattern GCG codes) in ARED (Table 2). The data show that the 13-bp ARE pattern with two mismatches (which was used for ARED build-up) was highly selective (89% specificity) towards 3′UTR when compared to CDS ($P < 0.0001$). The selectivity can also be increased to 96%, though at the expense of ARED coverage (82%; Table 2). Another distinguishable feature of the 13-bp pattern in typical ARE-mRNAs is that a great number of ARE-mRNAs (~40% of total ARE-mRNAs) have continuous patterns of AUUUA ($n > 1$) with the predominant pattern of WWWUAUUUAUUUAWW.

ARE CLUSTERING OF ARE-mRNAs

The ARE-mRNA database was further clustered into five groups depending on the number of motifs in the ARE stretch (Table 3). Clustering was performed in such a way that, for example, Group I included not only exact five or more continuous ARE pentamers but also those with 10% ambiguity, so that a stretch of NUUUAUUUAUUUAUUUAUUUN would fall in this category. This process was verified by a phylogenic tree relationship using Clustal-W alignment of ARE stretches and their variations. As could be expected, this analysis showed that the lower the number of ARE motifs in a group, the higher the number of sequences that were included, and apparently the more functionally diverse the corresponding ARE genes (Table 3). This may indicate evolutionary conservation within gene families containing long ARE stretches. It has been shown that the number of ARE motifs correlates with the turnover of ARE-mRNAs such as GM-CSF (20). Furthermore, the database also shows that not all cytokines or oncogene mRNAs belong to the ARE family as widely believed. Approximately 90% of the ARE-mRNAs compiled in the new database have not previously been recognized as such; hence only 10% represent previously known ARE-mRNAs. Based on this analysis we estimate the proportion of human mRNAs that contain functional AREs to be in the vicinity of 8%. This suggests, given an estimated number of genes in the human genome of ARE-mRNAs of 35 000–120 000 (21), that the number of ARE-containing genes is between 2800–9600.

BIOLOGICAL DIVERSITY AND INFERENCES IN DISEASE: EXAMPLES

The large number of ARE genes discovered by computational means show that they encode a large and wide variety of gene products in addition to the regulators of cell proliferation or the inflammatory/immune response. Group I (Table 3) contains many secreted proteins including GM-CSF, IL-1, IL-11, IL-12 and Gro-β that affect the growth of hematopoietic and immune cells (22). Although TNF-α is both a pro-inflammatory and anti-tumor protein, there is experimental evidence that it can act as a growth factor in certain leukemias and lymphomas (23). Unlike Group I, Groups II–V contain functionally diverse gene families comprising immune response, cell cycle and proliferation, inflammation and coagulation, angiogenesis, metabolism, energy, DNA binding and transcription, nutrient transportation and ionic homeostasis, protein synthesis, cellular biogenesis, signal transduction, and apoptosis. Only a few cases of ARE mRNAs coding for receptors were previously recognized: notably, urokinase-type plasminogen receptor and epidermal growth factor receptor (Table 1), but a significant number of receptors (69 mRNAs) mediating several biological responses including several hormone receptors were found in ARED. Interestingly, among the known apoptosis-inducing receptors, only TRAIL 2 and its closely-related death receptor 5, fall into the family of ARE mRNAs (Group II). Intriguingly, in some cases both the receptor and its ligand belong to ARED, for example, IL-10/IL-10R and urokinase-type plasminogen/u-PA receptor; indicating the tightly regulated processes of inflammation and cell adhesiveness, respectively (24). Among adhesion

Table 3. Diversity of ARE mRNA sequences

Group I Cluster (AUUUAUUUAUUUAUUUAUUUA)		
MMSET type I (MMSET)	AF071594	Cell growth/differentiation
GDP-L-fucose:β-D-galactoside 2-α-l-fucosyltransferase	M35531	Metabolism, carbohydrate
Cellular growth-regulating protein	L10844	Cell growth
Gro-β (melanoma stimulating growth factor)	M57731	Cell growth
Pim-1	M16750	Signal transduction, oncogenes
Neuoron-specific γ-2 enolase	M22349	Development/differentiation
Nuclear matrix protein NRP/B (NRPB)	AF059611	DNA transcription/differentiation
Natural killer cell stimulatory factor (IL-12)	M65290	Hematopoiesis, immune response
Granulocyte–macrophage colony stimulating factor (GM-CSF)	M11220	Hematopoiesis
Adipogenesis inhibitory factor (IL-11)	X58377	Hematopoiesis, metabolism
Natural resistance-associated macrophage protein	D50402	Immune response
Interleukin 1-β	M15330	Inflammation
Tumor necrosis factor	X01394	Inflammation
Group II Cluster (AUUUAUUUAUUUAUUUA) stretch		
Apoptosis-inducing (TRAIL) receptor 2	AF016849	Apoptosis
Death receptor 5 (DR5)	AF012535	Apoptosis
K-ras oncogene protein	M54968	Signal transduction
Thioredoxin reductase	S79851	Metabolism, redox
Interleukin-10 receptor	U00672	Immune regulation, receptors
Tyrosine kinase (ELK1) oncogene	M25269	Transcriptional regulation
6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (PFKFB)	AF056320	Metabolism, carbohydrate
Inducible 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (IPFK-2)	AF109735	Metabolism, carbohydrate
MOP1, basic helix–loop–helix PAS	U29165	Transcriptional regulation
Cytokine-inducible SH2-containing protein (G18)	AF132297	Signal transduction, inhibitors
K-Cl cotransporter KCC4	AF10536	Nutrient transport
Dishevelled 1 (DVL1)	AF006011	Signal transduction
Guanine nucleotide regulatory factor (LFP40)	U72206	Cellular motility/biogenesis
Zinc finger containing protein ZNF157 (ZNF157)	U28687	DNA transcription
Tubulin-folding cofactor C	U61234	Cellular motility/biogenesis
Interferon(LyIFN-α-1)	E00102	Immune response, innate
α-interferon Gx-1	E00124	Immune response, innate
Inteferon-α	V00542	Immune response, innate
Leukocyte interferon-α, clone pIFN105	M28585	Immune response, innate
Interferon α J	E00052	Immune response, innate
Messenger RNA for human leukocyte interferon	V00542	Immune response, innate
Angiotensin/vasopressin receptor AII/AVP	AF054176	Signal transduction, receptors
c-fos proto-oncogene	V01512	Transcriptional regulation
Group III Cluster (WAUUUAUUUAUUUAW) stretch		
Interferon (IFN-α-M1)	M27318	Immune response, innate
B-cell leukemia/lymphoma 2 (bcl-2) proto-oncogene	M13994	Apoptosis
Condoroitin 6-sulfotransferase	AB017915	Metabolism, sulphate
Sodium bicarbonate cotransporter 3 (SLC4A7)	AF047033	Metabolism, transport
Cyclooxygenase-2 (Cox-2)	M90100	Inflammation
Neuralized mRNA	AF029729	Development, neural
Gro (growth regulated) gene	J03561	Cellular growth
Epiregulin	D30783	Cellular growth
A-kinase anchor protein	U17195	Signal transduction
CREB-binding protein	U47741	Transcriptional regulation
L-type amino acid transporter subunit LAT1	AF104032	Nutrient transport

Table 3. *Continued*

Zinc finger DNA-binding motifs (IA-1)	M93119	Transcriptional regulation
Interleukin 3	M20137	Hematopoiesis
E16	M80244	Signal transduction, receptors
2-oxoglutarate dehydrogenase	D10523	Metabolism, energy
SIRP-β1	Y10376	Signal transduction, inhibitors
Putative sulphate transporter (PDS)	AF030880	Nutrient transport
Interleukin 2 (IL2)	U25676	Immune response, adaptive
Post-synaptic density protein 95 (PSD95)	U83192	Neural regulation
Dihydrolipoamide dehydrogenase-binding protein	AF001437	Metabolism, energy
Type 2 iodothyronine deiodinase	AF093774	Metabolism, hormones
RGP G-protein 3	U27655	Signal transduction
α-Endosulfine	AF067170	Nutrient transport, regulation
Tumor necrosis factor α inducible A20 zinc finger protein	M59465	Transcriptional regulation
Vascular endothelial growth factor	AF022375	Cellular growth, angiogenesis
Inhibitor of apoptosis protein-1 (MIHC)	AF070674	Apoptosis, regulation
A28-RGS14p	U70426	Signal transduction
TNFR2-TRAF signaling complex protein	L49432	Transcriptional regulation
Fibrillin-2	U03272	Cellular motility/biogenesis
Farnesylated-proteins converting enzyme 1	Y13834	Protein modification
D-1 dopamine receptor	X58987	Signal transduction, neural
HEB helix–loop–helix protein (HEB)	M80627	Transcriptional regulation
IFN-ω 1	A12140	Immune response, innate
Interferon β	X04430	Immune response, innate
Lymphotoxin	E01275	Immune response
Musashi/Nrp-1	AB012851	Development, neural
Thiamine carrier 1 (TC1)	AF153330	Metabolism, nutrient transport
Transcription factor (HTF4A)	M83233	Transcription
Phospholipase C-β-2	M95678	Signal transduction
Onconeural ventral antigen-1 (Nova-1)	U04840	Development/differentiaion, neural
Protein tyrosine kinase mRNA	M59371	Signal transduction
Tyrosine kinase receptor p145TRK-B (TRK-B)	U12140	Signal transduction, receptors
Protein tyrosine phosphatase	U27193	Signal transduction, inhibitors
Transcriptional regulatory protein p54	AF045451	Transcriptional regulation
cAMP phosphodiesterase PDE7 (PDE7A1	L12052	Signal transduction
Retinoic acid receptor γ 1	M38258	Transcriptional regulation, receptors
Tissue factor (thromboplastin)	A19048	Inflammation
c-sis/platelet derived growth factor 2 (PDGF2)	AF022375	Cellular growth, oncogenes
Endothelial leukocyte adhesion molecule-1 (ELAM-1)	M30640	Inflammation
Cluster IV Group (WWAUUUAUUUAWW) stretch		
175 sequences available with other group at http://rc.kfshrc.edu.sa/ared		
Cluster V Group (WWWWAUUUAWWWW) stretch		
582 sequences available with other group at http://rc.kfshrc.edu.sa/ared		

The ARE-mRNAs were clustered into five groups containing five, four, three and two pentameric repeats, while the last group contains only one pentamer within the 13-bp ARE pattern. Functional categories were assigned whenever possible according to NCBI-COG functional annotation (76)—in addition to the categories: inflammation, immune response, Development/Differentiation—using an extensive PubMed literature search. Those ARE-mRNAs, which do not have known biological function in Groups I–III are not shown in the table. Members of the training list (Table 1) are indicated in bold.

molecules, the endothelial leukocyte adhesion molecule-1 (ELAM-1, Group III) appears to be unique among other adhesion molecules in sense that it belongs to ARED.

Several metabolic processes including carbohydrate, amino acid and nucleic acid metabolism are represented in ARED (Table 3). For example, Group III contains several transporters

and enzymes necessary for nutrient and nucleotide transport. Enzymes that belong to different functional categories are also present throughout ARED. A large number of transcription factors and DNA binding proteins including zinc finger proteins (at least 30 mRNAs) were present in ARED. Since many early response genes are triggered by signal transduction, many mRNAs in ARED belong to this category (more than 50 entries). In addition, ARED contains a significant number of large (>4 kb) ARE-mRNAs that have no known gene function (~100 sequence entries; 25). It is apparent from ARED that many ARE-mRNAs mediate transient processes including hormone response, apoptosis, immune response and inflammation, cellular growth, etc. Thus, this compilation of ARE-containing mRNAs may offer further insights into the biology of ARE and relationships to disease processes; examples are given below.

Several human cancer and inflammatory diseases, including certain B-cell lymphomas, neuroblastoma, and chronic inflammatory conditions have been linked to ARE defects (reviewed by Vassalli and co-workers, 3). Removal of the ARE stretch correlates with the increased oncogenicity of proto-oncogene c-fos (Group II, Table 3) (26). Bcl-2 mRNA (Group III, Table 3) is an ARE-containing mRNA in which its increased stability may lead to overproduction of the antiapoptotic BCL-2 protein likely responsible for neoplastic transformation of follicular B-cell lymphoma (27). The development of two specific pathologies of chronic inflammatory arthritis and Crohn's-like inflammatory bowel disease were seen in mice rendered mutant in their TNF gene (Group I) by deleting the AT-rich region (2).

Increased production of hematopoietic growth factors such GM-CSF acting as autocrine growth factors, due to defects in ARE-mediated stability, may contribute to the pathogenesis of leukemia (28,29). TRAIL 2 is known to be involved in apoptosis of tumor cells but not normal cells (30), thus, prolongation of TRAIL2 mRNA stability may cause unwanted apoptotic events instead of transient apoptotic events in normal cells. Other notable examples of the role of receptors in disease are the angiotensin/vasopressin receptor (Group II) and D1 dopamine receptor (Group III), which may be involved in hypertension (31,32). The vascular endothelial growth factor/vascular permeability factor (Group III) which affects angiogenesis is thought to promote the cancer survival process (33).

The present identification and clustering of new ARE-mRNAs together with previously known ARE-mRNAs may facilitate our understanding of how these genes are regulated and how they may possibly be involved in disease processes. Thus, the investigation of regulatory pathways and pathological processes that involve ARE-mediated action may benefit greatly from the bioinformatics approach described here, and the availability of the ARE-mRNA database.

ACCESS AND FUTURE DIRECTION

We are planning to expand the list of ARE-mRNAs to include joined CDS from GenBank entries and to refine redundancy. It is our intent to provide at least one updated release of ARED per year. The web site http://rc.kfshrc.edu.sa/ared contains flat files of ARED database and their cluster groups. Also, database description and tables are available at the same site. The ARED files are in GenBank flatfile view (i.e. nucleotide sequence with annotations) that can be either downloaded or directly viewed when opened with Internet browsers.

REFERENCES

1. Wilson,T. and Treisman,R. (1988) Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3′ AU-rich sequences. *Nature*, **336**, 396–399.
2. Kontoyiannis,D., Pasparakis,M., Pizarro,T.T., Cominelli,F. and Kollias,G. (1999) Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity*, **10**, 387–398.
3. Conne,B., Stutz,A. and Vassalli,J.D. (2000) The 3′-untranslated region of messenger RNA: a molecular 'hotspot' for pathology? *Nat. Med.*, **6**, 637–641.
4. Chen,C.Y. and Shyu,A.B. (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.*, **20**, 465–470.
5. Lagnado,C.A., Brown,C.Y. and Goodall,G.J. (1994) AUUUA is not sufficient to promote poly(A) shortening and degradation of an mRNA: the functional sequence within AU-rich elements may be UUAUUUA(U/A)(U/A). *Mol. Cell. Biol.*, **14**, 7984–7995.
6. Lewis,T., Gueydan,C., Huez,G., Toulme,J.J. and Kruys,V. (1998) Mapping of a minimal AU-rich sequence required for lipopolysaccharide-induced binding of a 55-kDa protein on tumor necrosis factor-α mRNA. *J. Biol. Chem.*, **273**, 13781–13786.
7. Zubiaga,A.M., Belasco,J.G. and Greenberg,M.E. (1995) The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation. *Mol. Cell. Biol.*, **15**, 2219–2230.
8. Shaw,G. and Kamen,R. (1986) A conserved AU sequence from the 3′-untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell*, **46**, 659–667.
9. Zhang,W., Wagner,B.J., Ehrenman,K., Schaefer,A.W., DeMaria,C.T., Crater,D., DeHaven,K., Long,L. and Brewer,G. (1993) Purification, characterization, and cDNA cloning of an AU-rich element RNA-binding protein, AUF1. *Mol. Cell. Biol.*, **13**, 7652–7665.
10. Levine,T.D., Gao,F., King,P.H., Andrews,L.G. and Keene,J.D. (1993) Hel-N1: an autoimmune RNA-binding protein with specificity for 3′ uridylate-rich untranslated regions of growth factor mRNAs. *Mol. Cell. Biol.*, **13**, 3494–3504.
11. Ma,W.J., Cheng,S., Campbell,C., Wright,A. and Furneaux,H. (1996) Cloning and characterization of HuR, a ubiquitously expressed Elav-like protein. *J. Biol. Chem.*, **271**, 8144–8151.
12. Stoecklin,G., Ming,X.F., Looser,R. and Moroni,C. (2000) Somatic mRNA turnover mutants implicate tristetraprolin in the interleukin-3 mRNA degradation pathway. *Mol. Cell. Biol.*, **20**, 3753–3763.
13. Carballo,E., Lai,W.S. and Blackshear,P.J. (2000) Evidence that tristetraprolin is a physiological regulator of granulocyte-macrophage colony-stimulating factor messenger RNA deadenylation and stability. *Blood*, **95**, 1891–1899.
14. Winzen,R., Kracht,M., Ritter,B., Wilhelm,A., Chen,C.Y., Shyu,A.B., Muller,M., Gaestel, M., Resch,K. and Holtmann,H. (1999) The p38 MAP kinase pathway signals for cytokine-induced mRNA stabilization via MAP kinase-activated protein kinase 2 and an AU-rich region-targeted mechanism. *EMBO J.*, **18**, 4969–4980.
15. Lasa,M., Mahtani,K.R., Finch,A., Brewer,G., Saklatvala,J. and Clark,A.R. (2000) Regulation of cyclooxygenase 2 mRNA stability by the mitogen-activated protein kinase p38 signaling cascade. *Mol. Cell. Biol.*, **20**, 4265–4274.
16. Pages,G., Berra,E., Milanini,J., Levy,A.P. and Pouyssegur,J. (2000) Stress-activated protein kinases (JNK and p38/HOG) are essential for vascular endothelial growth factor mRNA stability. *J. Biol. Chem.*, **275**, 26484–26491.
17. Pesole,G., Liuni,S., Grillo,G., Licciulli,F., Larizza,A., Makalowski,W. and Saccone,C. (2000) UTRdb and UTRsite: specialized databases of sequences and functional elements of 5′ and 3′ untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.*, **28**, 193–196.
18. Bailey,T.L. and Gribskov,M. (1998) Methods and statistics for combining motif match scores. *J. Comput. Biol.*, **5**, 211–221.
19. Grillo,G., Attimonelli,M., Liuni,S. and Pesole,G. (1996) CLEANUP: a fast computer program for removing redundancies from nucleotide sequence databases. *Comput. Appl. Biosci.*, **12**, 1–8.
20. Akashi,M., Shaw,G., Hachiya,M., Elstner,E., Suzuki,G. and Koeffler,P. (1994) Number and location of AUUUA motifs: role in regulating transiently expressed RNAs. *Blood*, **83**, 3182–3187.
21. Editorial (2000) The nature of the number. *Nature*, **25**, 127–128.
22. Witsell,A.L. and Schook,L.B. (1992) Tumor necrosis factor alpha is an autocrine growth regulator during macrophage differentiation. *Proc. Natl Acad. Sci. USA*, **89**, 4754–4758.

23. Liu,R.Y., Fan,C., Liu,G., Olashaw,N.E. and Zuckerman,K.S. (2000) Activation of p38 mitogen-activated protein kinase is required for tumor necrosis factor-alpha-supported proliferation of leukemia and lymphoma cell lines. *J. Biol. Chem.*, **275**, 21086–21093.
24. Montero,L. and Nagamine,Y. (1999) Regulation by p38 mitogen-activated protein kinase of adenylate- and uridylate-rich element-mediated urokinase-type plasminogen activator (uPA) messenger RNA stability and uPA-dependent *in vitro* cell invasion. *Cancer Res.*, **59**, 5286–5293.
25. Kikuno,R., Nagase,T., Ishikawa,K., Hirosawa,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O. (1999) Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res.*, **6**, 197–205.
26. Raymond,V., Atwater,J.A. and Verma,I.M. (1989) Removal of an mRNA destabilizing element correlates with the increased oncogenicity of proto-oncogene fos. *Oncogene Res.*, **5**, 1–12.
27. Capaccioli,S., Quattrone,A., Schiavone,N., Calastretti,A., Copreni,E., Bevilacqua,A., Canti,G., Gong,L., Morelli,S. and Nicolin,A. (1996) A bcl-2/IgH antisense transcript deregulates bcl-2 gene expression in human follicular lymphoma t(14;18) cell lines. *Oncogene*, **13**, 105–115.
28. Hoyle,P.E., Steelman,L.S. and McCubrey,J.A. (1997) Autocrine transformation of human hematopoietic cells after transfection with an activated granulocyte/macrophage colony stimulating factor gene. *Cytokines Cell. Mol. Ther.*, **3**, 159–168.
29. Paul,C.C., Mahrer,S., McMannama,K. and Baumann,M.A. (1997) Autocrine activation of the IL-3/GM-CSF/IL-5 signaling pathway in leukemic cells. *Am. J. Hematol.*, **56**, 79–85.
30. Kim,K., Fisher,M.J., Xu,S.Q. and el-Deiry,W.S. (2000) Molecular determinants of response to TRAIL in killing of normal and cancer cells. *Clin. Cancer Res.*, **6**, 335–346.
31. Murphy,M.B. (2000) Dopamine: a role in the pathogenesis and treatment of hypertension. *J. Hum. Hypertens.*, **14**, S47–S50.
32. Burrell,L.M., Risvanis,J., Johnston,C.I., Naitoh,M. and Balding,L.C. (2000) Vasopressin receptor antagonism–a therapeutic option in heart failure and hypertension. *Exp. Physiol.*, **85**, 259S–265S.
33. Adams,J., Carder,P.J., Downey,S., Forbes,M.A., MacLennan,K., Allgar,V., Kaufman,S., Hallam,S., Bicknell,R., Walker,J.J., Cairnduff,F., Selby,P.J., Perren,T.J., Lansdown,M. and Banks,R.E. (2000) Vascular endothelial growth factor (VEGF) in breast cancer: comparison of plasma, serum, and tissue VEGF and microvessel density and effects of tamoxifen. *Cancer Res.*, **60**, 2898–2905.
34. Santis,A.G., Lopez-Cabrera,M., Sanchez-Madrid,F. and Proudfoot,N. (1995) Expression of the early lymphocyte activation antigen CD69, a C-type lectin, is regulated by mRNA degradation associated with AU-rich sequence motifs. *Eur. J. Immunol.*, **25**, 2142–2146.
35. Chesney,J., Mitchell,R., Benigni,F., Bacher,M., Spiegel,L., Al-Abed,Y., Han,J.H., Metz,C. and Bucala,R. (1999) An inducible gene product for 6-phosphofructo-2-kinase with an AU-rich instability element: role in tumor cell glycolysis and the Warburg effect. *Proc. Natl Acad. Sci. USA*, **96**, 3047–3052.
36. Chen,C.Y., Chen,T.M. and Shyu,A.B. (1994) Interplay of two functionally and structurally distinct domains of the c-fos AU-rich element specifies its mRNA-destabilizing function. *Mol. Cell. Biol.*, **14**, 416–426.
37. Ubeda,M., Schmitt-Ney,M., Ferrer,J. and Habener,J.F. (1999) CHOP/GADD153 and methionyl-tRNA synthetase (MetRS) genes overlap in a conserved region that controls mRNA stability. *Biochem. Biophys. Res. Commun.*, **262**, 31–38.
38. Reeves,R. and Magnuson,N.S. (1990) Mechanisms regulating transient expression of mammalian cytokine genes and cellular oncogenes. *Prog. Nucleic Acid Res. Mol. Biol.*, **38**, 241–282.
39. Brewer,G. (1991) An A + U-rich element RNA-binding factor regulates c-myc mRNA stability in vitro. *Mol. Cell. Biol.*, **11**, 2460–2466.
40. Rimokh,R., Berger,F., Bastard,C., Klein,B., French,M., Archimbaud,E., Rouault,J.P., Santa Lucia,B., Duret,L., Vuillaume,M. *et al.* (1994) Rearrangement of CCND1 (BCL1/PRAD1) 3′ untranslated region in mantle-cell lymphomas and t(11q13)-associated leukemias. *Blood*, **83**, 3689–3696.
41. Saida,K., Hashimoto,M., Mitsui,Y., Ishida,N. and Uchide,T. (2000) The prepro vasoactive intestinal contractor (VIC)/endothelin-2 gene (EDN2): structure, evolution, production, and embryonic expression. *Genomics*, **64**, 51–61.
42. McCulloch,R.K., Walker,C.E., Chakera,A., Jazayeri,J. and Leedman,P.J. (1998) Regulation of EGF-receptor expression by EGF and TGF alpha in epidermoid cancer cells is cell type-specific. *Int. J. Biochem. Cell. Biol.*, **30**, 1265–1278.
43. Kenealy,M.R., Flouriot,G., Sonntag-Buck,V., Dandekar,T., Brand,H. and Gannon,F. (2000) The 3′-untranslated region of the human estrogen receptor alpha gene mediates rapid messenger ribonucleic acid turnover. *Endocrinology*, **141**, 2805–2813.
44. Touriol,C., Morillon,A., Gensac,M.C., Prats,H. and Prats,A.C. (1999) Expression of human fibroblast growth factor 2 mRNA is post-transcriptionally controlled by a unique destabilizing element present in the 3′-untranslated region between alternative polyadenylation sites. *J. Biol. Chem.*, **274**, 21402–21408.
45. Brown,C.Y., Lagnado,C.A., Vadas,M.A. and Goodall,G.J. (1996) Differential regulation of the stability of cytokine mRNAs in lipopolysaccharide-activated blood monocytes in response to interleukin-10. *J. Biol. Chem.*, **271**, 20108–20112.
46. Hamilton,B.J., Nichols,R.C., Tsukamoto,H., Boado,R.J., Pardridge,W.M. and Rigby,W.F. (1999) hnRNP A2 and hnRNP L bind the 3′UTR of glucose transporter 1 mRNA and exist as a complex in vivo. *Biochem. Biophys. Res. Commun.*, **261**, 646–651.
47. Sirenko,O.I., Lofquist,A.K., DeMaria,C.T., Morris,J.S., Brewer,G. and Haskill,J.S. (1997) Adhesion-dependent regulation of an A+U-rich element-binding activity associated with AUF1. *Mol. Cell. Biol.*, **17**, 3898–3906.
48. Rodriguez-Pascual,F., Hausding,M., Ihrig-Biedert,I., Furneaux,H., Levy,A.P., Forstermann,U. and Kleinert,H. (2000) Complex contribution of the 3′-untranslated region to the expressional regulation of the human inducible nitric-oxide synthase gene. Involvement of the RNA-binding protein HuR. *J. Biol. Chem.*, **275**, 26040–26049.
49. Caput,D., Beutler,B., Hartog,K., Thayer,R., Brown-Shimer,S. and Cerami,A. (1986) Identification of a common nucleotide sequence in the 3′-untranslated region of mRNA molecules specifying inflammatory mediators. *Proc. Natl Acad. Sci. USA*, **83**, 1670–1674.
50. Gorospe,M. and Baglioni,C. (1994) Degradation of unstable interleukin-1 alpha mRNA in a rabbit reticulocyte cell-free system. Localization of an instability determinant to a cluster of AUUUA motifs. *J. Biol. Chem.*, **269**, 11845–11851.
51. Peppel,K., Vinci,J.M. and Baglioni,C. (1991) The AU-rich sequences in the 3′-untranslated region mediate the increased turnover of interferon mRNA induced by glucocorticoids. *J. Exp. Med.*, **173**, 349–355.
52. Grafi,G., Sela,I. and Galili,G. (1993) Translational regulation of human beta interferon mRNA: association of the 3′ AU-rich sequence with the poly(A) tail reduces translation efficiency *in vitro*. *Mol. Cell. Biol.*, **13**, 3487–3493.
53. Gillis,P. and Malter,J.S. (1991) The adenosine-uridine binding factor recognizes the AU-rich elements of cytokine, lymphokine, and oncogene mRNAs. *J. Biol. Chem.*, **266**, 3172–3177.
54. Kastelic,T., Schnyder,J., Leutwiler,A., Traber,R., Streit,B., Niggli,H., MacKenzie,A. and Cheneval,D. (1996) Induction of rapid IL-1 beta mRNA degradation in THP-1 cells mediated through the AU-rich region in the 3′UTR by a radicicol analogue. *Cytokine*, **8**, 751–761.
55. Kishore,R., Tebo,J.M., Kolosov,M. and Hamilton,T.A. (1999) Cutting edge: clustered AU-rich elements are the target of IL-10-mediated mRNA destabilization in mouse macrophages. *J. Immunol.*, **162**, 2457–2461.
56. Lindstein,T., June,C.H., Ledbetter,J.A., Stella,G. and Thompson,C.B. (1989) Regulation of lymphokine messenger RNA stability by a surface-mediated T cell activation pathway. *Science*, **244**, 339–343.
57. Henics,T., Sanfridson,A., Hamilton,B.J., Nagy,E. and Rigby,W.F. (1994) Enhanced stability of interleukin-2 mRNA in MLA 144 cells. Possible role of cytoplasmic AU-rich sequence-binding proteins. *J. Biol. Chem.*, **269**, 5377–5383.
58. Yang,L. and Yang,Y.C. (1994) Regulation of interleukin (IL)-11 gene expression in IL-1 induced primate bone marrow stromal cells. *J. Biol. Chem.*, **269**, 32732–32739.
59. Quincoces,A.F. and Leon,J. (1995) Serum growth factors up-regulate H-ras, K-ras, and N-ras proto-oncogenes in fibroblasts. *Cell Growth Differ.*, **6**, 271–279.
60. Carlson,C.D., Bai,Y., Jonakait,G.M. and Hart,R.P. (1996) Interleukin-1 beta increases leukemia inhibitory factor mRNA levels through transient stimulation of transcription rate. *Glia*, **18**, 141–151.
61. Chambers,S.K. and Kacinski,B.M. (1994) Messenger RNA decay of macrophage colony-stimulating factor in human ovarian carcinomas in vitro. *J. Soc. Gynecol. Investig.*, **1**, 310–316.

62. Bhattacharya,S., Giordano,T., Brewer,G. and Malter,J.S. (1999) Identification of AUF-1 ligands reveals vast diversity of early response gene mRNAs. *Nucleic Acids Res.*, **27**, 1464–1472.
63. Wang,S.W., Pawlowski,J., Wathen,S.T., Kinney,S.D., Lichenstein,H.S. and Manthey,C.L. (1999) Cytokine mRNA decay is accelerated by an inhibitor of p38-mitogen-activated protein kinase. *Inflamm. Res.*, **48**, 533–538.
64. Hartner,A., Goppelt-Struebe,M., Hocke,G.M. and Sterzel,R.B. (1997) Differential regulation of chemokines by leukemia inhibitory factor, interleukin-6 and oncostatin M. *Kidney Int.*, **51**, 1754–1760.
65. Madireddi,M.T., Dent,P. and Fisher,P.B. (2000) Regulation of mda-7 gene expression during human melanoma differentiation. *Oncogene*, **19**, 1362–1368.
66. Kondo,A., Isaji,S., Nishimura,Y. and Tanaka,T. (2000) Transcriptional and post-transcriptional regulation of monocyte chemoattractant protein-3 gene expression in human endothelial cells by phorbol ester and cAMP signalling. *Immunology*, **99**, 561–568.
67. Chagnovich,D. and Cohn,S.L. (1997) Activity of a 40 kDa RNA-binding protein correlates with MYCN and c-fos mRNA stability in human neuroblastoma. *Eur. J. Cancer*, **33**, 2064–2067.
68. Sherer,T.B., Neff,P.S., Hankins,G.R. and Tuttle,J.B. (1998) Mechanisms of increased NGF production in vascular smooth muscle of the spontaneously hypertensive rat. *Exp. Cell Res.*, **241**, 186–193.
69. Liang,P. and Pardee,A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, **257**, 967–971.
70. Wingett,D., Reeves,R. and Magnuson,N.S. (1991) Stability changes in pim-1 proto-oncogene mRNA after mitogen stimulation of normal lymphocytes. *J. Immunol.*, **147**, 3653–3659.
71. Maurer,F., Tierney,M. and Medcalf,R.L. (1999) An AU-rich sequence in the 3′-UTR of plasminogen activator inhibitor type 2 (PAI-2) mRNA promotes PAI-2 mRNA decay and provides a binding site for nuclear HuR. *Nucleic Acids Res.*, **27**, 1664–1673.
72. Gasdaska,J.R., Harney,J.W., Gasdaska,P.Y., Powis,G. and Berry,M.J. (1999) Regulation of human thioredoxin reductase expression and activity by 3′-untranslated region selenocysteine insertion sequence and mRNA instability elements. *J. Biol. Chem.*, **274**, 25379–25385.
73. Ahern,S.M., Miyata,T. and Sadler,J.E. (1993) Regulation of human tissue factor expression by mRNA turnover. *J. Biol. Chem.*, **268**, 2154–2159.
74. Gueydan,C., Droogmans,L., Chalon,P., Huez,G., Caput,D. and Kruys,V. (1999) Identification of TIAR as a protein binding to the translational regulatory AU-rich element of tumor necrosis factor alpha mRNA. *J. Biol. Chem.*, **274**, 2322–2326.
75. Wang,G.J., Collinge,M., Blasi,F., Pardi,R. and Bender,J.R. (1998) Posttranscriptional regulation of urokinase plasminogen activator receptor messenger RNA levels by leukocyte integrin engagement. *Proc. Natl Acad. Sci. USA*, **95**, 6296–6301.
76. Tatusov,R.L., Galperin,M.Y., Natale,D.A. and Koonin,E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.*, **28**, 33–36. Updated article in this issue: *Nucleic Acids Res.* (2001), **29**, 22–28.