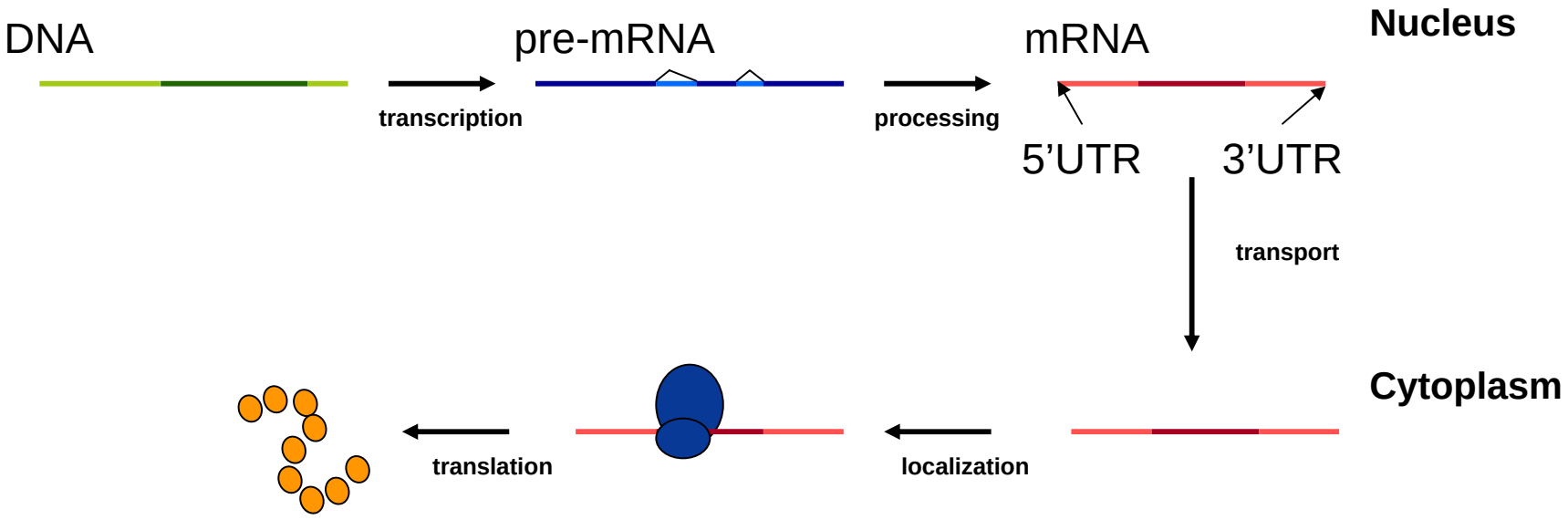
A journey through regulatory features of UTRs of eukaryotic mRNAs

Kristin Missal

Bioinformatic, Institute of Computer Science, Univ. Leipzig, Germany

Bled, 2004

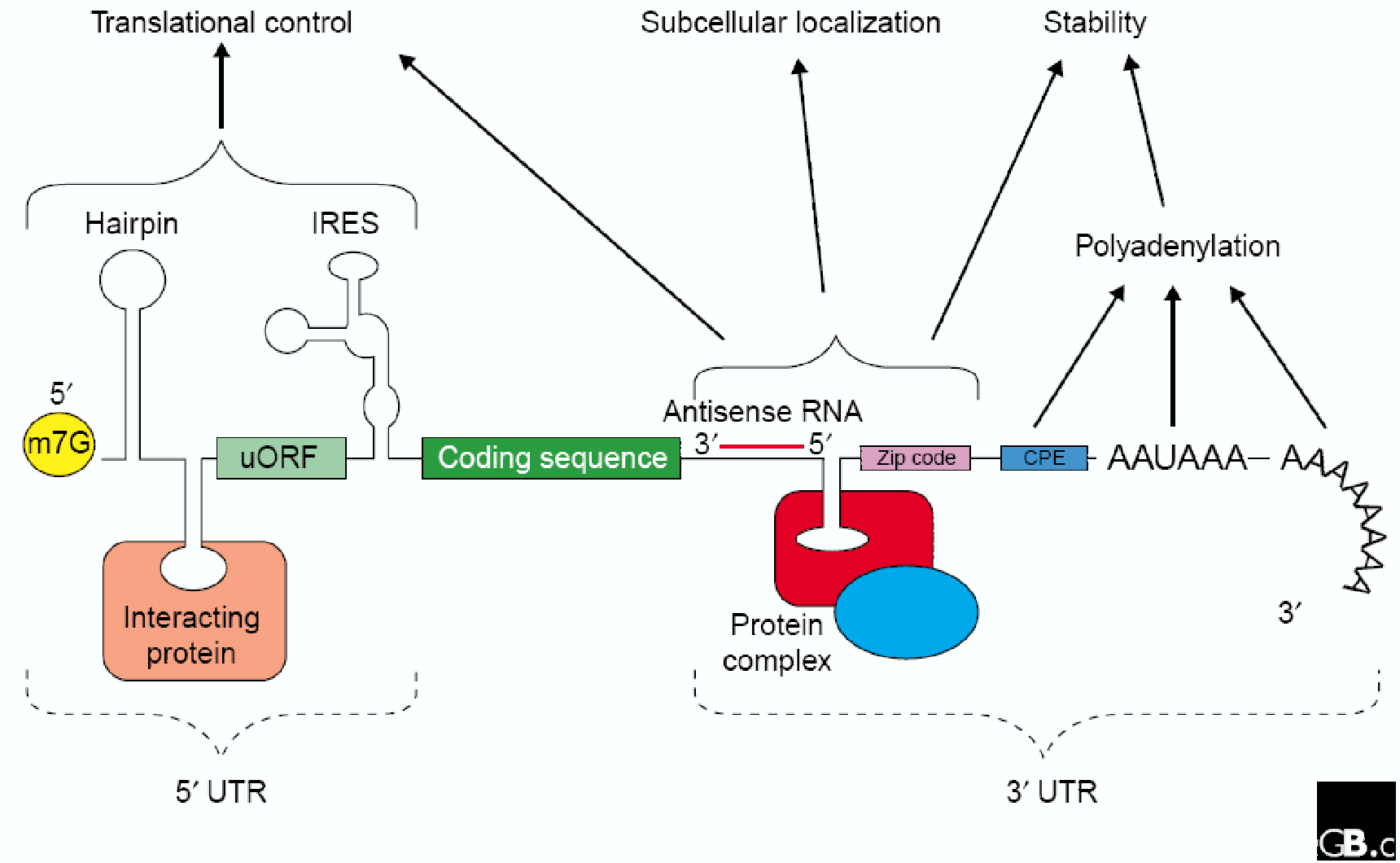
Regulation of gene expression



Transcriptional control: Whether a gene is transcribed or not and to what extend.

Post-transcriptional control: Controlling the fate of transcribed molecules.

Regulation by UTRs



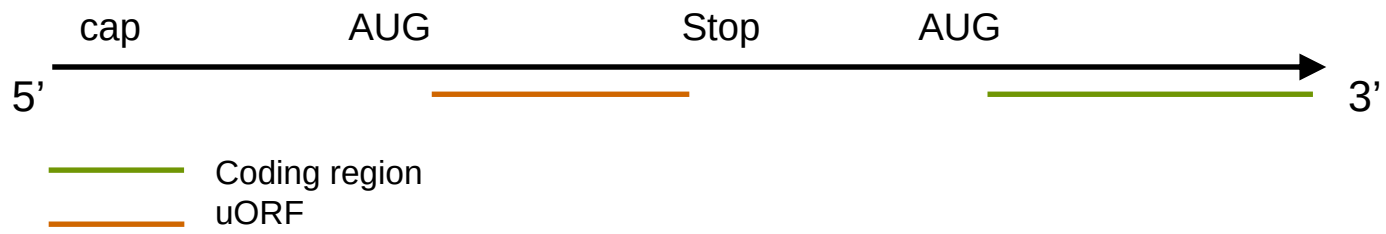
Mignone et al. 2002: Genome Biology 3(3):reviews0004.1-0004.10

Control of translation efficiency

Leaky scanning:

Occurrence of upstream AUGs correlates with a long 5'UTR and with weak start codon context of first AUG codon

Down-regulate translation:



Control of translation efficiency

Role of secondary structure in 5'UTR:

Inhibitory effects of very stable secondary structures

Internal ribosome entry site (IRES):

Is a mechanism of translation initiation alternative to the conventional 5'-cap dependent ribosome scanning.

Common structural motif: A Y-type stem-loop structure followed by the AUG triplet or followed by additional stem-loop structures and the AUG triplet. (Le, and Maizel, 1997: Nucleic Acids Res.25,362-69)

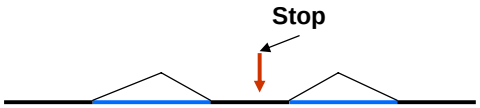
Control of mRNA stability

Changes in rate of mRNA degradation may alter the amount of protein in a cell.

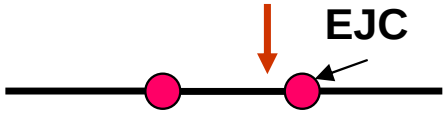
Mechanisms of mRNA degradation:

- **AU-rich elements** in 3' UTRs affect rate of shortening of the poly(A)-tail (Deadenylation)
- Removal of the cap at the 5' end (Decapping)
- Nonsense mediated mRNA decay (Decapping)

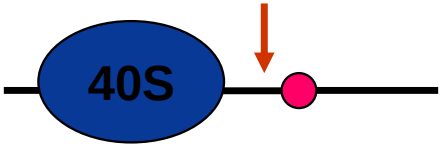
Nonsense-mediated mRNA decay



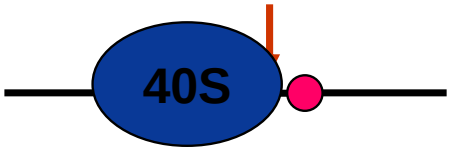
Pre-mRNA processing



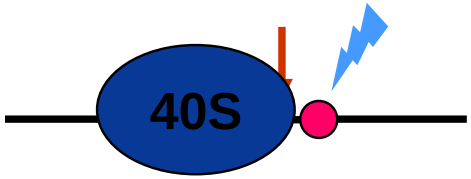
Spliceosome deposits exon junction complexes (EJCs) at site of intron removal



With first round of translation, the ribosome displaces the EJCs



If ribosome reaches a stop codon upstream of the final EJCs, last EJCs will remain bound



Recruiting of a decapping enzyme through interactions between EJC proteins and release factors triggers rapid mRNA decay

Functional analysis of UTRs

- UTRs with their *cis*-acting elements have critical role in many aspects in regulation of gene expression

- Functional elements share common motifs

- Identifying common motifs may lead to new sequence regions important for regulation of gene expression

$\Rightarrow$ Need of general analysis of features in primary and secondary structure of UTRs

Functional analysis of UTRs

UTRdb: Specialized database of 5' and 3' UTR sequences of eukaryotic mRNAs (Pesole, et al. 2002: Nucleic Acids Res. 30, 335-340)

- Generated by parsing EMBL/GenBank DB entries and cleaning from redundancy
- Additional information like number of exons in corresponding gene region, presence of repetitive elements and occurrence of regulatory elements
- Cross-referencing to primary DB entry and to corresponding 5' or 3' UTR
- Current release (24th October 2003 - against EMBL release 75) includes 62163 for homo sapiens, 32538 entries for mouse and 9557 for rattus norvegicus

Functional analysis of UTRs

UTRsite: A collection of functional sequence patterns located in 5' or 3' UTR sequences (Pesole, et al. 2002: Nucleic Acids Res. 30, 335-340)

- Generated on basis of information reported in literature
- Description of biological role of functional element
- Current release (30th July 2003) includes 31 entries

Functional analysis of UTRs

Common oligonucleotides: The WordUP algorithm finds oligonucleotide motifs which may be involved in regulatory activity (Pesole et al. 1992: Nucleic Acids Res. 20, 2871-2875).

It assesses the statistical significance of each word of size w comparing the observed and expected number of sequences containing it.

Expected probability that sequence i contains oligomer s_k at least one time:

$$\pi_i(s_k) = 1 - e^{-\lambda_i} \tag{1}$$

λ_i is the average number of sequences containing oligomer s_k in sequence i:

$$\lambda_i = \underbrace{q_i(s_k)}_{\text{Probability that } s_k \text{ occurs in } i} \quad \underbrace{(L_i - w + 1)}_{\text{Maximal number of occurence of } s_k \text{ in } i} \tag{2}$$

WordUP

The statistical significance of the occurrence of oligomer s_k is verified by:

$$\chi^2(s_k) = \frac{(\overbrace{\sum_i p_i(s_k)}^{\text{Observed}} - \overbrace{\sum_i \pi_i(s_k)}^{\text{Expected}})^2}{\sum_i \pi_i(s_k)} \qquad (3)$$

Results:

	5'UTR		3'UTR	
w	**oligo**	χ^2	**oligo**	χ^2
6	CUGCAG	347.55	AAUAAA	4729.17
7	GGAGCCG	267.18	UGUAUUU	1802.74
8	GAAUUCGG	2316.47	UGUAUAUA	2917.89
9	GAAUUCCGG	4155.05	UACAGGCGU	3697.54

Only the most significant oligonucleotides are reported.

Functional analysis of UTRs

Common patterns: PatSearch is a more sophisticated pattern discovery algorithm. (Pesole et al. 2000: Bioinformatics 16, 439-450):

- It analyzes user submitted sequence collections for the presence of complex patterns.

- Definition of patterns is similar to regular expressions

$$\text{p1} = 4...4\text{p1p1}$$

  Pattern p1 will match any character sub-sequence that is made up of 3 repeats of the same 4 character sequence.

PatSearch

- Mismatch and/or mispairing below a user fixed threshold S is allowed

p1 = GSGC

	G	S	G	C
A	16	0	0	0
C	0	50	0	80
G	84	50	100	20
T	0	0	0	0

Match p1 against GACG: $84 + 0 + 0 + 20 = 104 > S$

PatSearch

- Pattern may include potential secondary structure elements

$$p2 = \sim p1$$

Pattern p2 matches the reverse complement of pattern p1.

$$p3 = 6...8\ 3...8 \sim p3$$

Pattern p3 matches a hairpin loop in which the stem comprises 6 to 8 nucleotides and the loop 3 to 8 nucleotides.

PatSearch

- *UTRsite* contains functional elements of 3' and 5' UTRs identified by *PatSearch*

Functional elements	UTR	UTRdb entries
Iron responsive element	3', 5'	121
Upstream ORF	5'	71 438
Internal ribosome entry site	3'	7 356
Class 2 AU-rich elements	3'	70

- Disadvantage: Pattern to search for must be known!

Comparative analysis

- Average length of 5'UTRs is more or less constant over taxonomic classes
- Average length of 3'UTRs is much more variable
- But length of 5' and 3' UTRs vary a lot within a species
- G and C content of 5'UTRs is greater than that of 3'UTRs
- Contain several types of repeats

Summary

- UTRs play important role in post-transcriptional regulation

- What has been already done?
  - Common oligonucleotides
  - Identification of known functional elements
  - Comparative studies

Outlook (first steps)

- Clustering primary structure depending on local alignments

- Analyzing secondary structures

THANK YOU!