Interaktionen und Modifikationen von RNAs und Proteinen

(Modul 10-202-2208; Spezialvorlesung)

Jörg Fallmann

Institute for Bioinformatics
University of Leipzig

04.05.2018

Central Dogma of Gene Expression

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{protein}$$

- only about 2% of human DNA is protein coding
  about 80% is transcribed

With great power comes great responsibility

A key mechanism for the survival of each cell and as a consequence whole organisms, is tight and correct regulation of gene expression

With great power comes great responsibility

A key mechanism for the survival of each cell and as a consequence whole organisms, is tight and correct regulation of gene expression

This includes localization, processing and turnover of all forms of nucleic acids (and proteins) in a cell

With great power comes great responsibility

A key mechanism for the survival of each cell and as a consequence whole organisms, is tight and correct regulation of gene expression

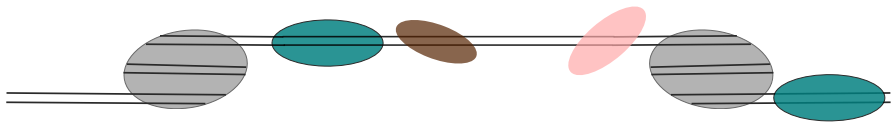
This includes localization, processing and turnover of all forms of nucleic acids (and proteins) in a cell

With growing complexity of organisms, so grows the complexity of regulatory mechanisms

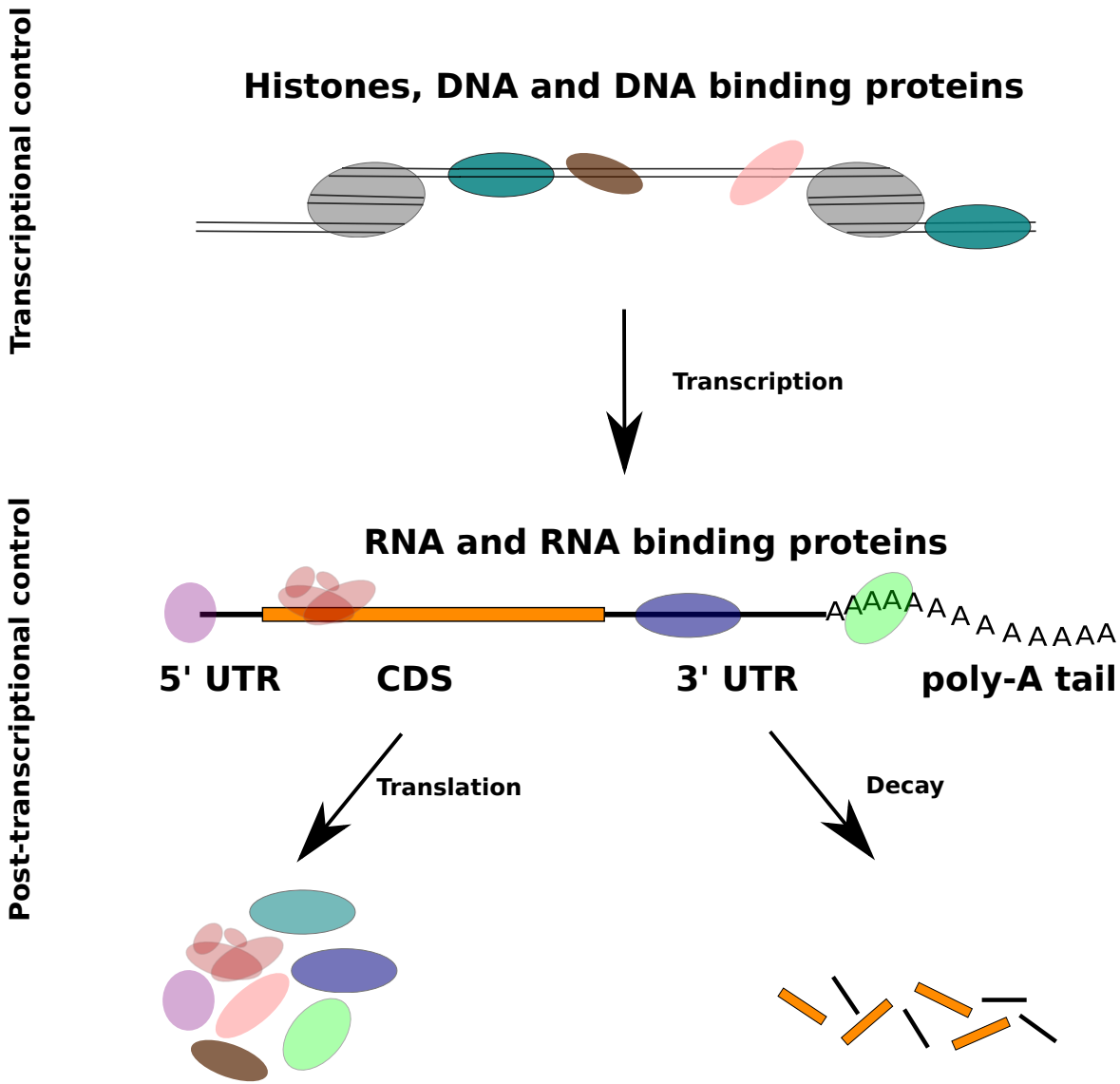
Gene Expression Regulation Simple

Transcriptional control

Histones, DNA and DNA binding proteins



Gene Expression Regulation Simple



RNA cycle of life

Synthesis

RNA-polymerase transcription from DNA template

RNA cycle of life

Synthesis

RNA-polymerase transcription from DNA template

From here on post-transcriptional

RNA cycle of life

Synthesis

RNA-polymerase transcription from DNA template

Processing

Via various complexes

de-/capping, de-/adenylation, splicing, modification, ...

RNA cycle of life

Synthesis

- RNA-polymerase transcription from DNA template

Processing

- Via various complexes
  de-/capping, de-/adenylation, splicing, modification, ...

Final un-/stable RNA molecule

RNA cycle of life

Synthesis

RNA-polymerase transcription from DNA template

Processing

Via various complexes
de-/capping, de-/adenylation, splicing, modification, ...

Final un-/stable RNA molecule

Available RNA depending on ratio $\frac{synthesis}{decay}$

Affects all sort of RNA (mRNA, tRNA, miRNA, lncRNA, ...)

Affects all sort of RNA (mRNA, tRNA, miRNA, lncRNA, ...)

Major part of RNA half-life control is performed directly by

Affects all sort of RNA (mRNA, tRNA, miRNA, lncRNA, ...)

Major part of RNA half-life control is performed directly by

proteins or **protein-complexes**

Interaction as the key

For regulation to take place, there has to be interaction

This means direct contact between RNA and protein

Interaction as the key

For regulation to take place, there has to be interaction

This means direct contact between RNA and protein

Or the lack thereof

RNA-protein interactions (RPIs)

Interplay between ribo-nucleic acids (RNA), either coding (mRNA) or non-coding (ncRNA), and RNA binding proteins (RBPs)

RNA-protein interactions (RPIs)

Interplay between ribo-nucleic acids (RNA), either coding (mRNA) or non-coding (ncRNA), and RNA binding proteins (RBPs)

Extreme versatility of RNA molecules (sequence and structural features) is reflected by an equal complexity of RNA binding domains and binding preferences of proteins

RNA-protein interactions (RPIs)

Interplay between ribo-nucleic acids (RNA), either coding (mRNA) or non-coding (ncRNA), and RNA binding proteins (RBPs)

Extreme versatility of RNA molecules (sequence and structural features) is reflected by an equal complexity of RNA binding domains and binding preferences of proteins

Remember the many structures for single molecule of RNA

RNA binding proteins (RBPs)

Hundreds of RNA binding proteins (RBPs) have been shown to be involved in virtually all aspects of (post-transcriptional) gene expression regulation

RNA binding proteins (RBPs)

Hundreds of RNA binding proteins (RBPs) have been shown to be involved in virtually all aspects of (post-transcriptional) gene expression regulation

A manually curated collection of more than 1.500 RBPs in human highlights their vast number and potential for interaction and regulation

RNA binding proteins (RBPs) II

Regulation is usually initiated by direct interaction between RBP and target RNA, requiring more or less specific sequence motifs and accessible binding sites

RNA binding proteins (RBPs) II

Regulation is usually initiated by direct interaction between RBP and target RNA, requiring more or less specific sequence motifs and accessible binding sites

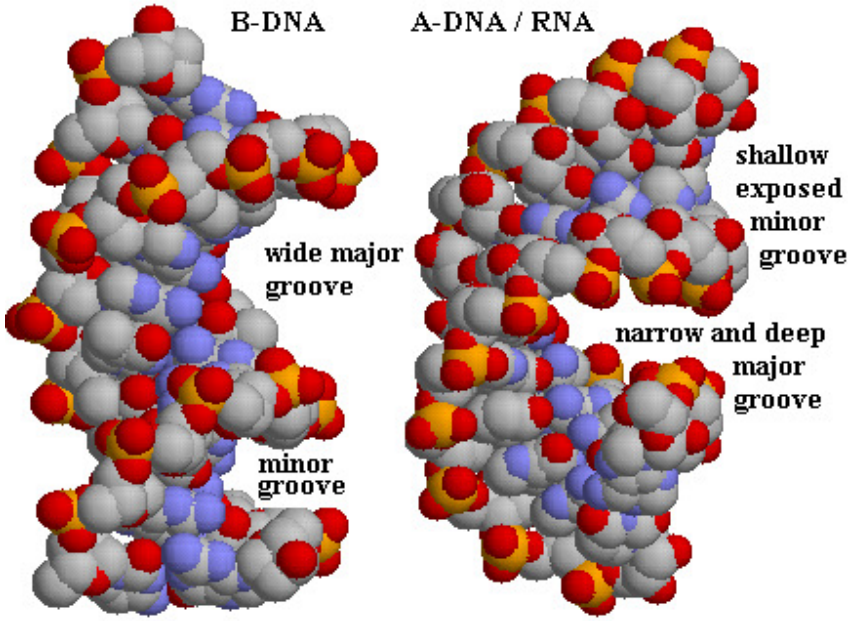
Many of the known RBPs seem to prefer single stranded binding regions, although some have been shown to interact with structured RNA sites

Binding RNA vs. binding DNA

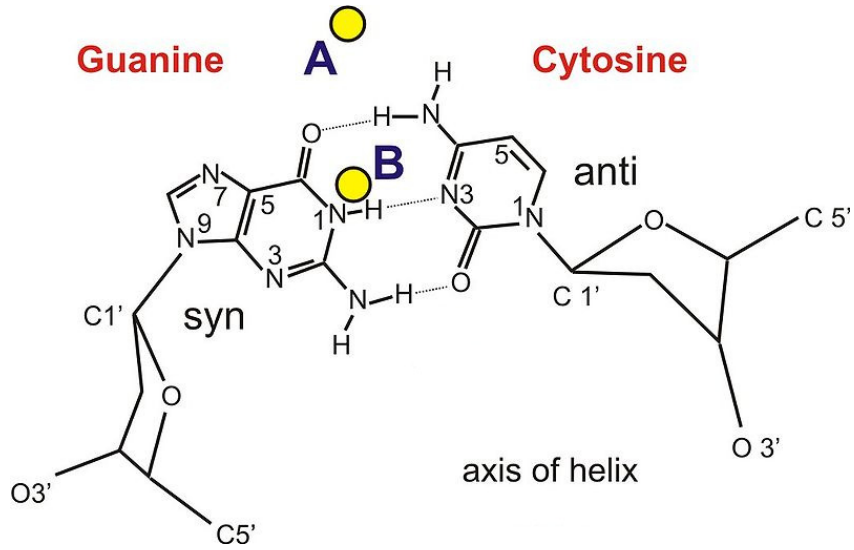
Feature	**RNA**	**DNA**
chemical nature	2'-OH	2'-desoxy
sequence	U	T
base paring rule	U-A & U-G	only T-A
structure	stem-loop structures	long double helix
helix geometry	A-form	usually B-form
function	information processing	information storage

B-form and A-form double-helix

overall shape



axis position



Same nucleotide sequences and **Same** cWW (cis-Watson-Crick/Watson-Crick) base pairs can form very different helical structures

RNA-Protein Interactions

Every RNA is in complex with unspecifically or specifically binding proteins. Togehter they form **R**ibo**N**ucleo**P**rotein Complex (RNPs).

RNAs fold into stem-loop structures
$\rightarrow$ many different types of interactions

RNA-Protein Interactions

Every RNA is in complex with unspecifically or specifically binding proteins. Togehter they form **R**ibo**N**ucleo**P**rotein Complex (RNPs).

RNAs fold into stem-loop structures
$\rightarrow$ many different types of interactions

proteins with **RNA binding domains (RBDs)** bind

single stranded RNA (ssRNA)
short motifs in unpaired regions
commonly in **sequence specific** manner

RNA-Protein Interactions

Every RNA is in complex with unspecifically or specifically binding proteins. Togehter they form **R**ibo**N**ucleo**P**rotein Complex (RNPs).

RNAs fold into stem-loop structures
$\rightarrow$ many different types of interactions

proteins with **RNA binding domains (RBDs)** bind

single stranded RNA (ssRNA)
short motifs in unpaired regions
commonly in **sequence specific** manner
double stranded RNA (dsRNA)
paired regions of stems
commonly in **sequence unspecific** manner

RNA binding domains (RBDs)

The versatility of RBPs makes it hard to predict interaction partners from amino-acid sequence alone.

RNA binding domains (RBDs)

The versatility of RBPs makes it hard to predict interaction partners from amino-acid sequence alone.

Most forms of protein interactions require specific, conserved (tertiary) structures with certain amino-acid content, known as domains

RNA binding domains (RBDs)

The versatility of RBPs makes it hard to predict interaction partners from amino-acid sequence alone.

Most forms of protein interactions require specific, conserved (tertiary) structures with certain amino-acid content, known as domains

RBPs contain RNA-binding domains (RBDs)

Although these domains are very specific and employ different interaction mechanisms, they can share some features that enable RNA-protein interactions

RNA binding domains (RBDs)

Protein domains can fold independently, and this fold plays a crucial role in the specificity of RNA recognition

RNA binding domains (RBDs)

Protein domains can fold independently, and this fold plays a crucial role in the specificity of RNA recognition

Hydrogen bonds with backbone and specific interactions between nucleotides and amino-acids are common for sequence specific binding

RNA binding domains (RBDs)

Protein domains can fold independently, and this fold plays a crucial role in the specificity of RNA recognition

Hydrogen bonds with backbone and specific interactions between nucleotides and amino-acids are common for sequence specific binding

Electrostatic interactions and stacking contribute to the affinity of a protein to RNA

RNA binding domains (RBDs)

Protein domains can fold independently, and this fold plays a crucial role in the specificity of RNA recognition

Hydrogen bonds with backbone and specific interactions between nucleotides and amino-acids are common for sequence specific binding

Electrostatic interactions and stacking contribute to the affinity of a protein to RNA

In general, RBDs are 35-90aa in size and interact with a small stretch of nucleotides (3-5nt)

RNA binding domains (RBDs)

Protein domains can fold independently, and this fold plays a crucial role in the specificity of RNA recognition

Hydrogen bonds with backbone and specific interactions between nucleotides and amino-acids are common for sequence specific binding

Electrostatic interactions and stacking contribute to the affinity of a protein to RNA

In general, RBDs are 35-90aa in size and interact with a small stretch of nucleotides (3-5nt)

To increase affinity and specificity often in combination with other RBDs in same protein, highlighting the modularity of most proteins

RNA binding domains (RBDs)

Protein domains can fold independently, and this fold plays a crucial role in the specificity of RNA recognition

Hydrogen bonds with backbone and specific interactions between nucleotides and amino-acids are common for sequence specific binding

Electrostatic interactions and stacking contribute to the affinity of a protein to RNA

In general, RBDs are 35-90aa in size and interact with a small stretch of nucleotides (3-5nt)

To increase affinity and specificity often in combination with other RBDs in same protein, highlighting the modularity of most proteins

Some RBPs do not contain a canonical RBD

Sequence-specific vs. sequence-unspecific binding

Sequence-specificity can be achieved via two strategies

H-bonds between protein backbone and RNA bases
$\rightarrow$ highly dependent of protein fold (hydrophobic sidechains looking towards RNA, almost no intramolecular stacking of RNA bases, instead intermolecular with sidechains, RNA bases not exposed to solvent, very rigid and specific scaffold)

Sequence-specific vs. sequence-unspecific binding

Sequence-specificity can be achieved via two strategies

H-bonds between protein backbone and RNA bases
$\rightarrow$ highly dependent of protein fold (hydrophobic sidechains looking towards RNA, almost no intramolecular stacking of RNA bases, instead intermolecular with sidechains, RNA bases not exposed to solvent, very rigid and specific scaffold)

H-bonds between aa-sidechains and RNA bases (intramolecular RNA base stacking, RNA bases exposed to solvent)

Sequence-specific vs. sequence-unspecific binding

Sequence-specificity can be achieved via two strategies

- H-bonds between protein backbone and RNA bases
  $\rightarrow$ highly dependent of protein fold (hydrophobic sidechains looking towards RNA, almost no intramolecular stacking of RNA bases, instead intermolecular with sidechains, RNA bases not exposed to solvent, very rigid and specific scaffold)

- H-bonds between aa-sidechains and RNA bases (intramolecular RNA base stacking, RNA bases exposed to solvent)

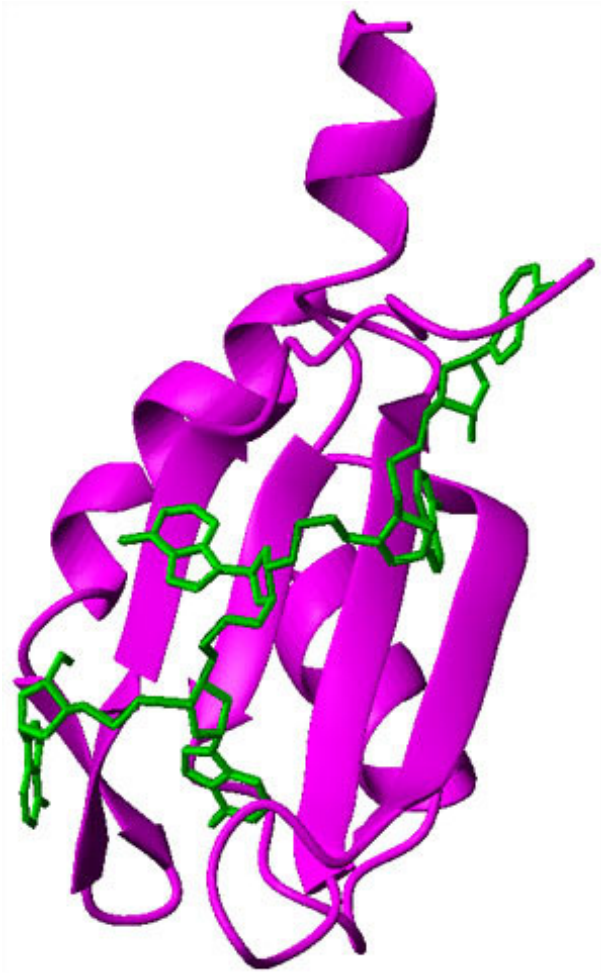
RBPs often employ a mix of strategies to bind their targets
$\rightarrow$ makes prediction of target sequences challenging

For some RBPs specific sequence preferences are known, for others they can be guessed from the available RBDs or do not exist

RRM – RNA-recognition motif

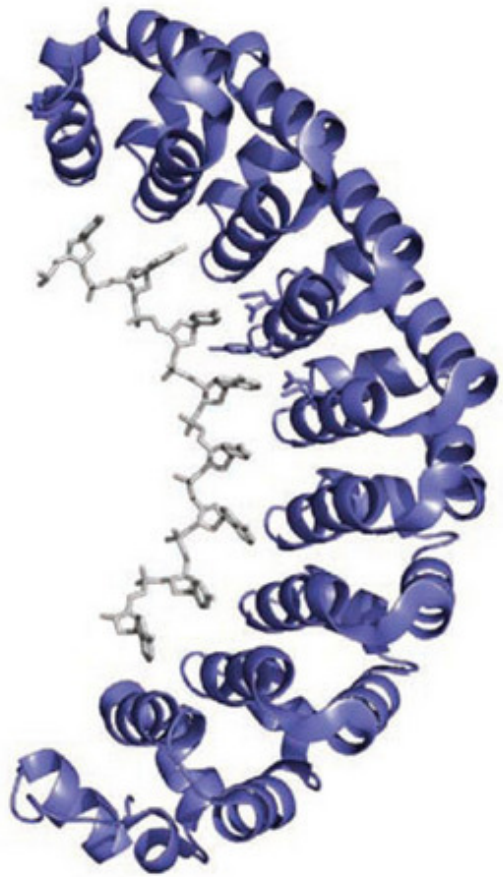
- most common
- binds ssRNA
- sequence specific
- 4 anti-parallel β-sheets packed against 2 α-helices
- 4-6nt interaction
- often in tandems/triplets
- very versatile

RRM in pink, RNA (5nt) green



zinc finger (CCCH/CCHC/C2H2)

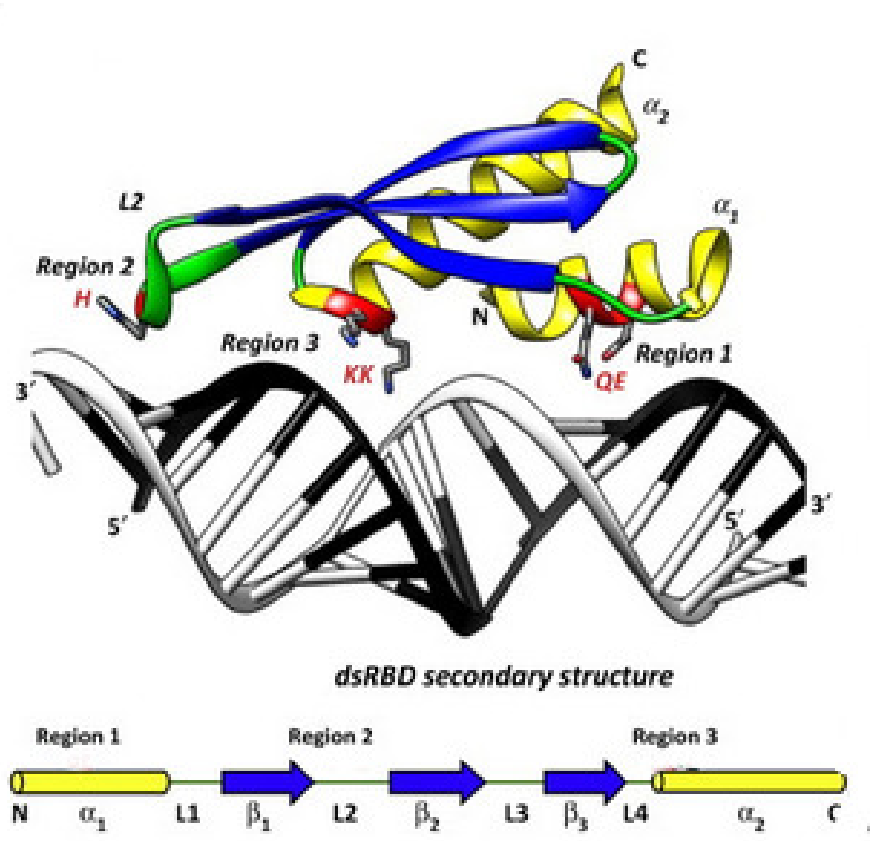
- C2H2 bind DNA, CCCH RNA, CCHC viral RNA
- coordinate zinc as common property
- unspecific binding to dsRNA
- specific binding to ssRNA
- RNA bases bulged out
- Intermolecular stacking and H-bond between RNA and backbone



9 zinc fingers (blue), RNA (8nt) grey

dsRBD – double-stranded RNA Binding Domain

- 3 anti-parallel β-sheets
- α-helices on N- and C-terminus
- unspecific binding to dsRNA via α-helices and loop between β-sheets
- contacts backbone
- sequence-specific binding of minor groove (ADAR2)



β-sheets (blue), A-form RNA

Protein interaction elements

Interaction between RNA and proteins depends on both, a protein domain that recognizes the target RNA, as well as RNA elements that are recognizable by the protein domain.

Protein interaction elements

Interaction between RNA and proteins depends on both, a protein domain that recognizes the target RNA, as well as RNA elements that are recognizable by the protein domain.

While protein domains have been studied and characterized in detail, RNA elements crucial for interaction are in general understudied

Protein interaction elements on sequence level

Among the most prominent metazoan RNA sequence elements are AU-rich elements (AREs) and GU- or U- rich elements (GU/UREs)

Protein interaction elements on sequence level

Among the most prominent metazoan RNA sequence elements are AU-rich elements (AREs) and GU- or U- rich elements (GU/UREs)

Found in many mRNAs and ncRNAs, throughout all genic regions (UTR, CDS, intron, exon)

Protein interaction elements on sequence level

Among the most prominent metazoan RNA sequence elements are AU-rich elements (AREs) and GU- or U- rich elements (GU/UREs)

Found in many mRNAs and ncRNAs, throughout all genic regions (UTR, CDS, intron, exon)

Of course there are others (*e. g.* splice sites, poly-A, ...)

So what about structure?

For the interaction of RNAs and proteins, structure is key

So what about structure?

For the interaction of RNAs and proteins, structure is key

Proteins need structure to form binding domains

So what about structure?

For the interaction of RNAs and proteins, structure is key

Proteins need structure to form binding domains

RNA structure defines if a binding motif is accessible or not

So what about structure?

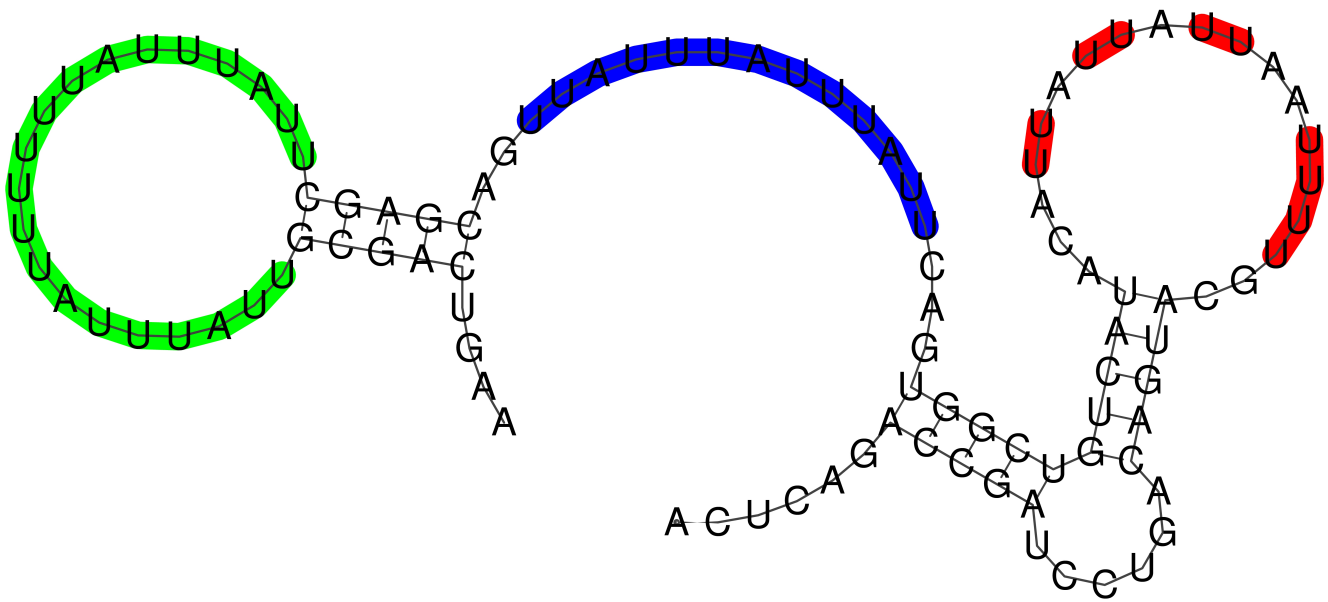
For the interaction of RNAs and proteins, structure is key

Proteins need structure to form binding domains

RNA structure defines if a binding motif is accessible or not

Sequence elements important for single-stranded RNA binding proteins, but they require the target RNA sequence to be unpaired or in an accessible structural context like the loop section of a hairpin loop

So what about structure?



Identifying RNA-protein interactions

RPIs are central part of complex interactome $\rightarrow$ interplay and mechanisms are not simple to investigate

Identifying RNA-protein interactions

RPIs are central part of complex interactome $\rightarrow$ interplay and mechanisms are not simple to investigate

Interaction is not necessarily functional

Identifying RNA-protein interactions

RPIs are central part of complex interactome $\rightarrow$ interplay and mechanisms are not simple to investigate

Interaction is not necessarily functional

Proteins can interact with their target not only at specific sites, but in a probing manner known as diffusional search $\rightarrow$ further complicating interaction analysis

Identifying RNA-protein interactions II

In principle, investigating interactions requires some knowledge of the interaction partners

Identifying RNA-protein interactions II

In principle, investigating interactions requires some knowledge of the interaction partners

Need for specific probes, antibodies, cell-types or substrates

Experimental Methods

Experimental methods for the characterization of RNA-RBP interactions can generally be broken down into

in vitro assays, which means free from other interacting factors and under experimental conditions

in vivo approaches which capture a snapshot of RBP binding to RNAs at natural expression levels or after induction

in vivo VS *in vitro* approaches

in vitro approaches

use a synthetic RNA bait to capture RBPs from cellular extracts or vice versa

in vivo VS *in vitro* approaches

in vitro approaches

use a synthetic RNA bait to capture RBPs from cellular extracts or vice versa

in vivo approaches

technically more challenging but preserve the context of competing or assisting interactions

RNA vs protein centric methods

RNA-centric methods

Protein-centric methods

RNA vs protein centric methods

RNA-centric methods $\rightarrow$ Pull down RNA

Protein-centric methods $\rightarrow$ Pull down protein

RNA vs protein centric methods

RNA-centric methods $\rightarrow$ Pull down RNA $\rightarrow$ analyze protein

Protein-centric methods $\rightarrow$ Pull down protein $\rightarrow$ analyze RNA

RNA vs protein centric methods

RNA-centric methods

- Mass spectrometry (MS) identify RBPs bound to RNA
- Allow the identification of novel RBP interactions
- RBPs for which antibodies are hard to come by
- Requires the purification of enough protein mass (hard)

Protein-centric methods

RNA vs protein centric methods

RNA-centric methods

- Mass spectrometry (MS) identify RBPs bound to RNA
- Allow the identification of novel RBP interactions
- RBPs for which antibodies are hard to come by
- Requires the purification of enough protein mass (hard)

Protein-centric methods

- Co-IP of protein with crosslinked targets
- Require some knowledge about protein and specific antibodies for IP
- Can easily be applied in a high-throughput manner and require lower amounts of starting material (PCR)

Protein-centric methods

Purification of RNA-RBP complexes via target protein

Specific purification methods for protein *in vivo*
or way to express a tagged version *in vitro* $\rightarrow$ recombinant protein

Protein-centric methods

- Purification of RNA-RBP complexes via target protein
  - Specific purification methods for protein *in vivo*
    or way to express a tagged version *in vitro* $\rightarrow$ recombinant protein
  - Immunoprecipitation (IP) of the protein via specific antibodies
    - Most common
    - Quality and specificity of AB has huge impact on reliability

Protein-centric methods

- Purification of RNA-RBP complexes via target protein
  - Specific purification methods for protein *in vivo*
    or way to express a tagged version *in vitro* $\rightarrow$ recombinant protein
  - Immunoprecipitation (IP) of the protein via specific antibodies
    - Most common
    - Quality and specificity of AB has huge impact on reliability
  - Co-IP'd RNA is reverse transcribed into cDNA

Protein-centric methods

Purification of RNA-RBP complexes via target protein

- Specific purification methods for protein *in vivo*
  or way to express a tagged version *in vitro* $\rightarrow$ recombinant protein

- Immunoprecipitation (IP) of the protein via specific antibodies
  - Most common
  - Quality and specificity of AB has huge impact on reliability

- Co-IP'd RNA is reverse transcribed into cDNA

- PCR amplification (NOT POSSIBLE WITH PROTEINS!!!)
  - Detect interaction partners from less starting material

Protein-centric methods

Purification of RNA-RBP complexes via target protein

- Specific purification methods for protein *in vivo*
  or way to express a tagged version *in vitro* $\rightarrow$ recombinant protein

- Immunoprecipitation (IP) of the protein via specific antibodies
  - Most common
  - Quality and specificity of AB has huge impact on reliability

- Co-IP'd RNA is reverse transcribed into cDNA

- PCR amplification (NOT POSSIBLE WITH PROTEINS!!!)
  - Detect interaction partners from less starting material

- Sequencing

Protein centric assays