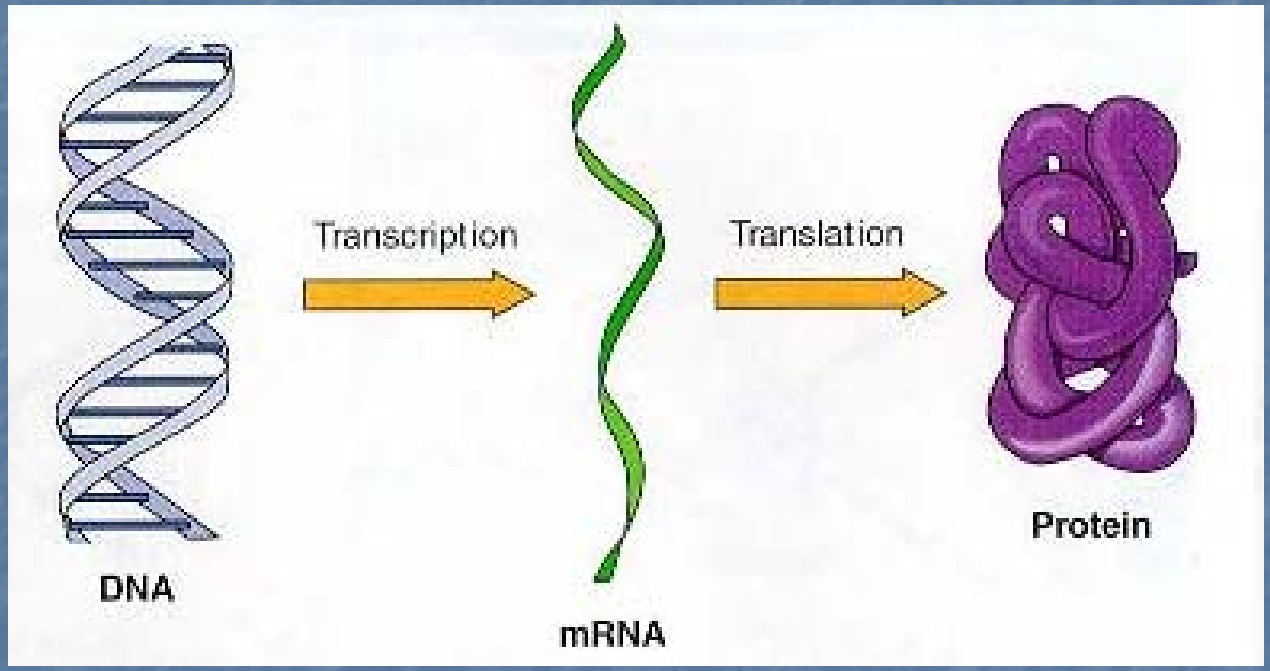
Controlled chaos

Tight regulation of unstructured proteins

Tõnu Margus

Journal Clubs in Bioinformatics March 2009

Sequence-structure-function paradigm

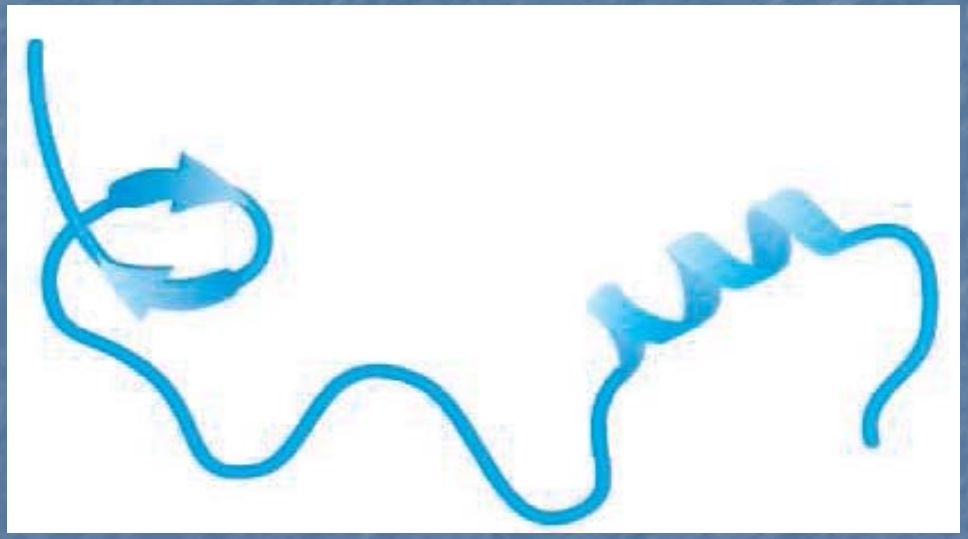


Proteins are traditionally viewed as rigid or semi-rigid "blocks," whose specificity and catalytic power are determined by the unique 3D structure

Intrinsically disordered proteins

IUPs

- Proteins that are wholly disordered & contain lengthy disordered segments, when alone in solution



- Such disordered proteins are abundant, diverse, vital, dynamic, and chaotic
- The concept of ill-structured but functional proteins have raised many questions

The advantage of lack of the structure!

- an increased interaction surface area
- conformational flexibility to interact with several targets
- the presence of molecular recognition elements that fold upon binding
- accessible posttranslational modification sites
- the availability of short linear interaction motifs

The main functions are related with

- transducing intracellular signals
- regulating processes including the cell division cycle,
- recognizing various binding partners (e.g., ligands, other proteins, and nucleic acids).
- These are complementary to the common catalysis and transport activities of proteins with well-defined, stable three dimensional structures

Abundance

- Intrinsically disordered regions are highly abundant in nature
- >50% of eukaryotic proteins likely contain at least one disordered region ≥30 amino acids in length
- >20% of eukaryotic proteins are expected to be mostly disordered

% depends from criteria; Gsponer J. 30% highly unstructured
30% highly structured

Evolutionary aspects

- Unstructured parts of protein were often found to appear in result of
  - alternative splicing (absent in one species, but present another)
  - new segments that become added to proteins (conversion of noncoding DNA into coding)
- Appearing of unstructured proteins are often associated with human diseases
- IUP sequences are evolving faster than highly structured sequences
- The rates and patterns of amino acid substitutions within intrinsically disordered proteins over evolutionary time are distinct from those within structured proteins

Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation

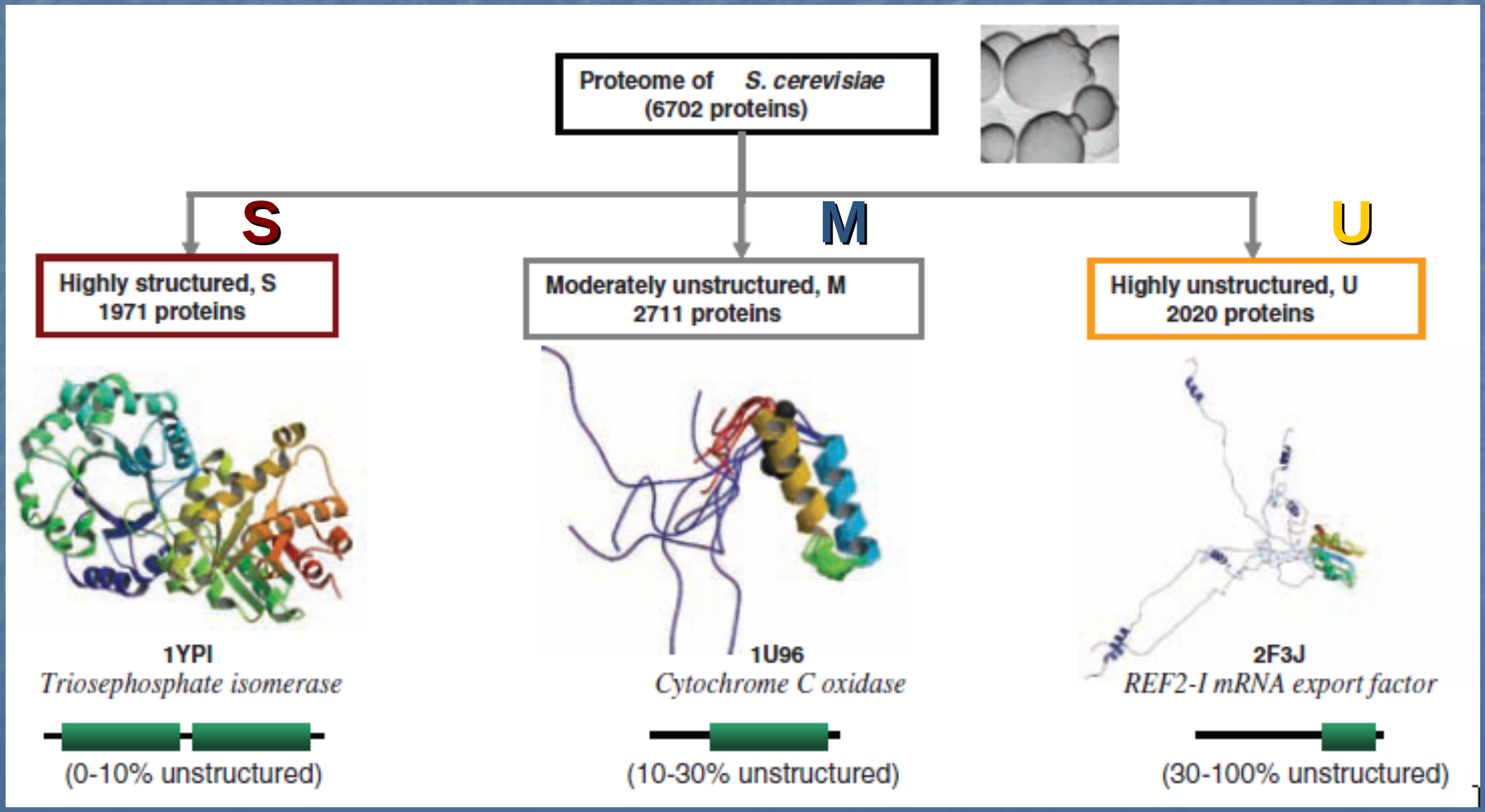
Jörg Gsponer,[1]* Matthias E. Futschik,[1,2,3] Sarah A. Teichmann,[1] M. Madan Babu[1]*

Altered abundance of several intrinsically unstructured proteins (IUPs) has been associated with perturbed cellular signaling that may lead to pathological conditions such as cancer. Therefore, it is important to understand how cells precisely regulate the availability of IUPs. We observed that regulation of transcript clearance, proteolytic degradation, and translational rate contribute to controlling the abundance of IUPs, some of which are present in low amounts and for short periods of time. Abundant phosphorylation and low stochasticity in transcription and translation indicate that the availability of IUPs can be finely tuned. Fidelity in signaling may require that most IUPs be available in appropriate amounts and not present longer than needed.

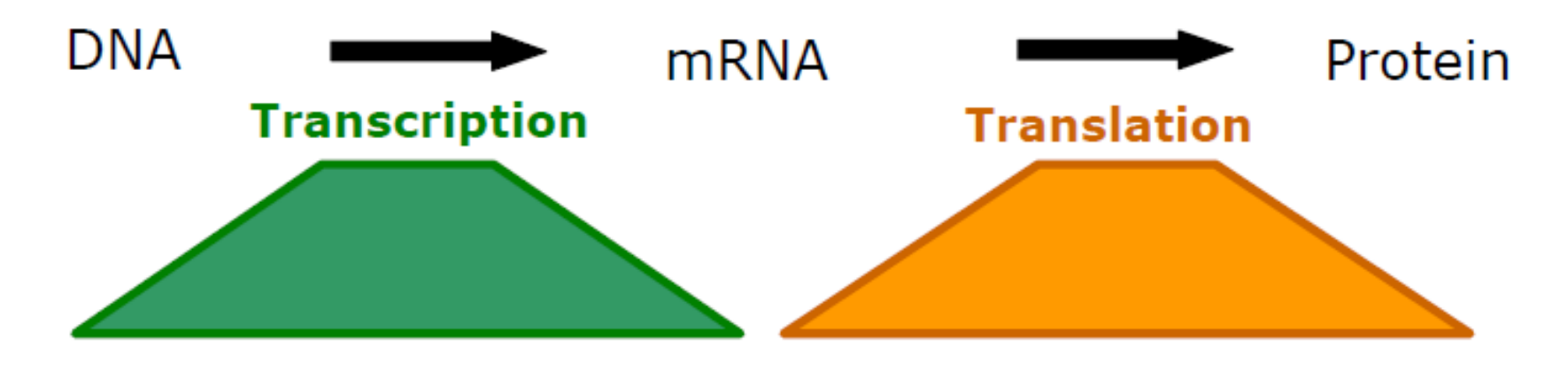
Science vol. 322 28. Nov 2008

Grouping proteins in the proteome of yeast

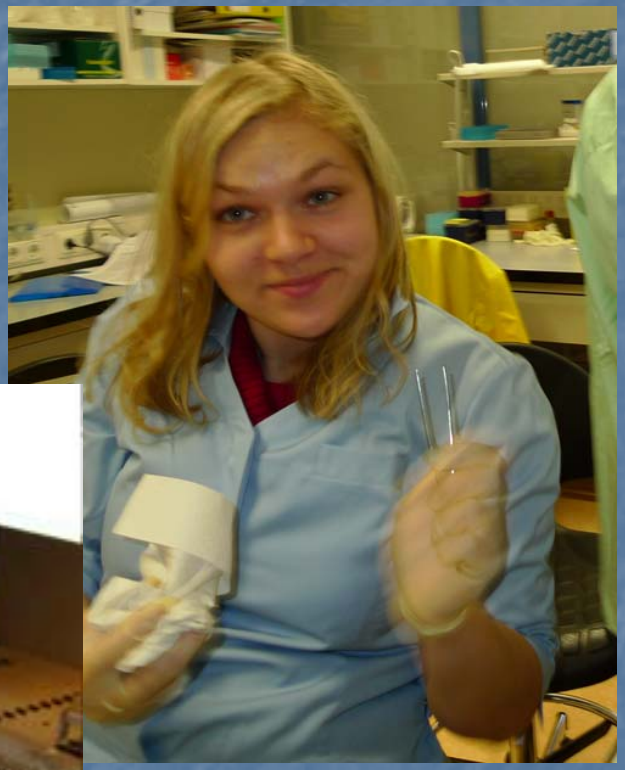
Using the Disopred2 software

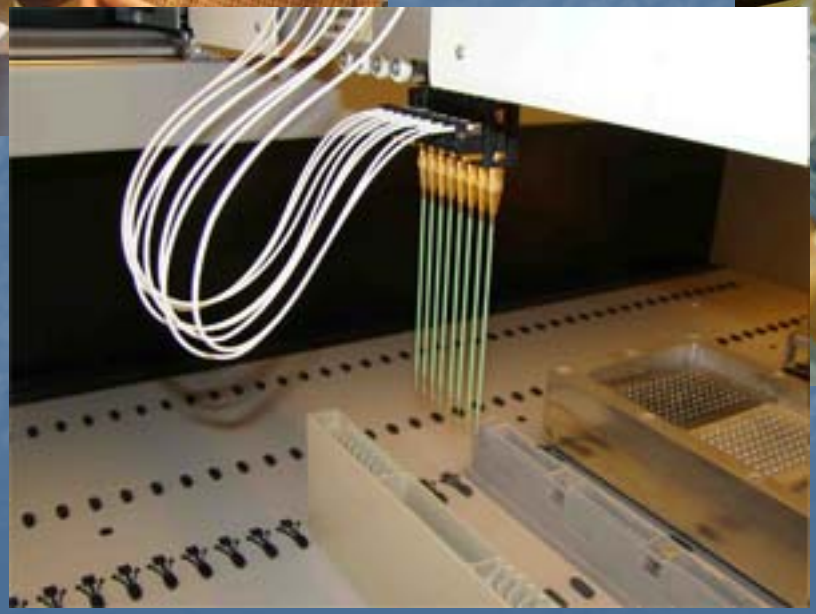


What to measure?



How to get the data?



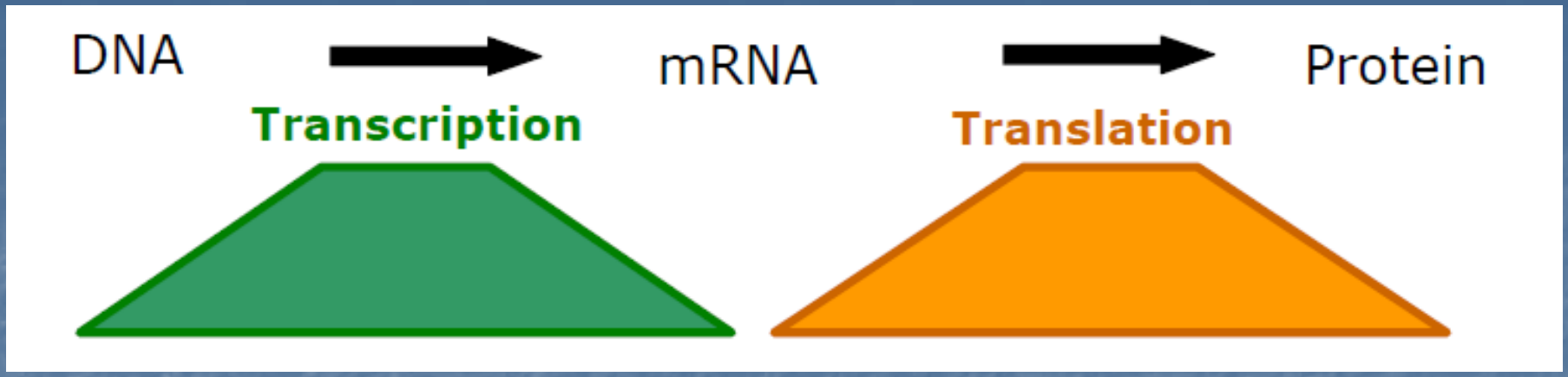


No need, the data is already there …

Table S1: Compendium of datasets used in our study

Type of information *[source]*	Description of the method used to obtain the data
Transcriptional noise (S11) (S10)	The presence or absence of TATA box was used as a proxy for transcriptional noise. The authors identified the genes in yeast and humans that have a TATA box by scanning promoter regions using a position weight matrix.
Transcriptional complexity (S1)	The transcriptional regulatory network for yeast was constructed by compiling high-confidence protein-DNA interactions from several published ChIP-chip experiments that involved 156 transcription factors and over 5000 genes across several conditions.
Transcriptional rate (S2)	Transcriptional rates for yeast grown in YPD were calculated by the authors based on the transcript abundances and mRNA half-lives. These were in turn determined by obtaining and comparing transcript levels of the wild-type and the temperature sensitive RNA polymerase *rpb1-1* mutant strains using an Affymetrix microarray.
Transcript abundance (S2) (S12)	Transcript abundances for yeast grown in YPD (*S. cerevisiae*) and Edinburgh minimal medium (*S. pombe*) were determined by using an Affymetrix high density oligonucleotide array.
Transcript half-life (S3) (S12) (S14)	Transcript half-lives were determined by obtaining transcript levels over several minutes after inhibiting transcription. This was done using the temperature sensitive RNA polymerase *rpb1-1* mutant *S. cerevisiae* strain, by adding 1,10-phenanthroline to *S. pombe* and by adding Actinomycin D to human HepG2 cells.

They used ~ 100 published articles to gather all related data !!!



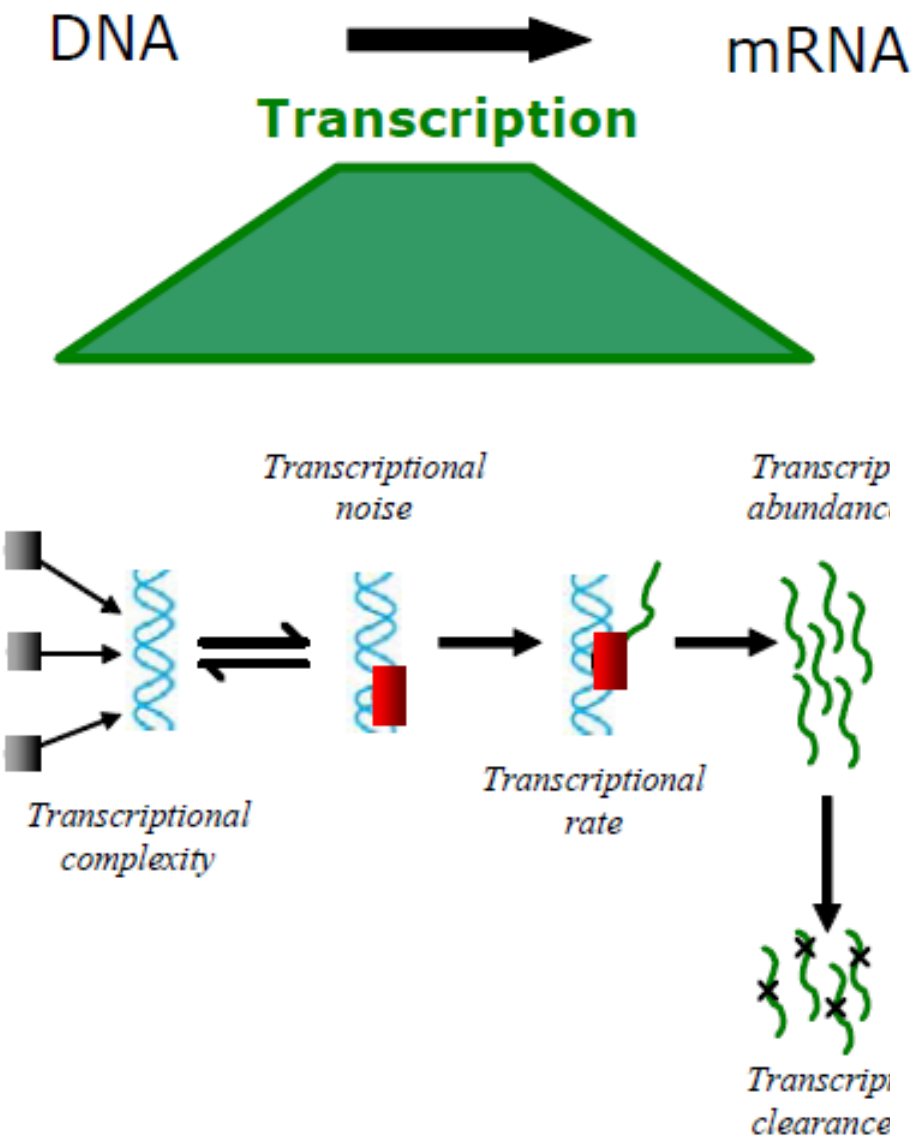
Parameters, related with regulation of gene dosage

Transcript availability = **Rate of transcription** − **Rate of degradation**

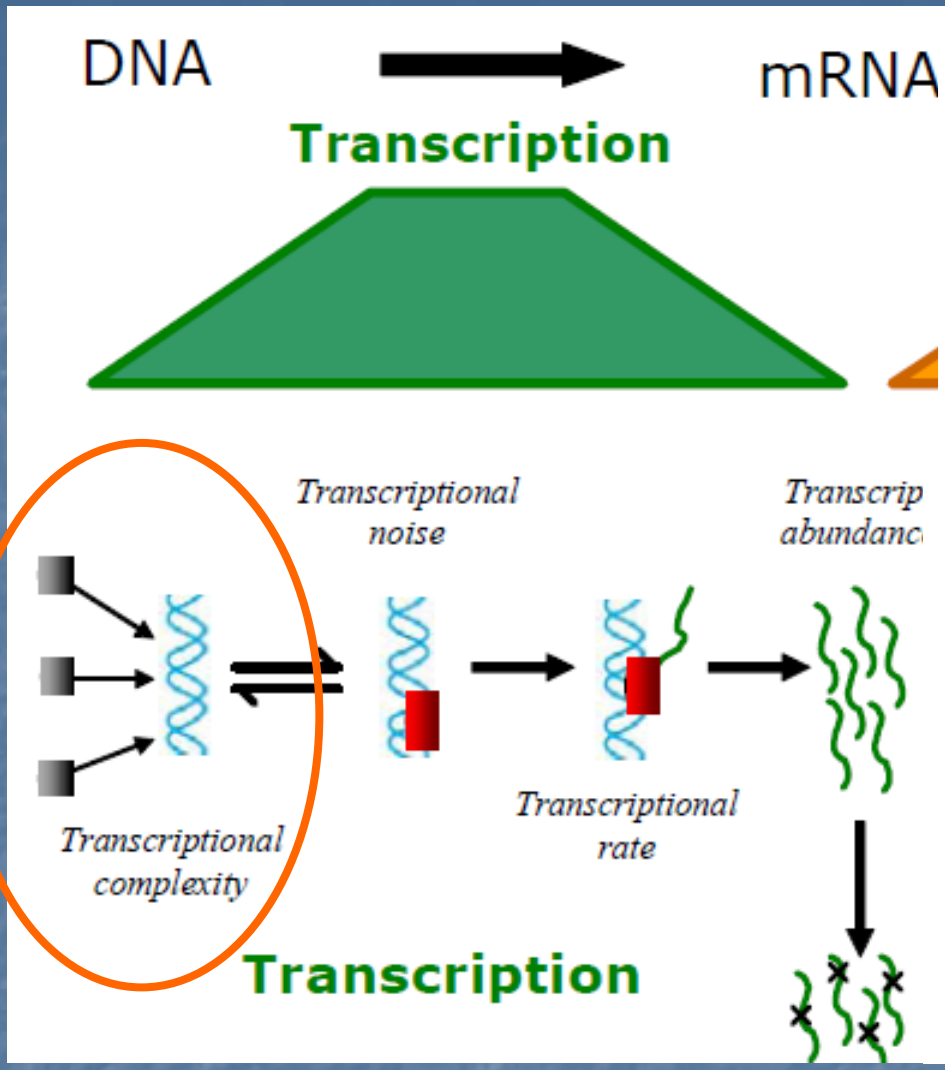
Mõõdetav suurus

Kaudselt hinnatav

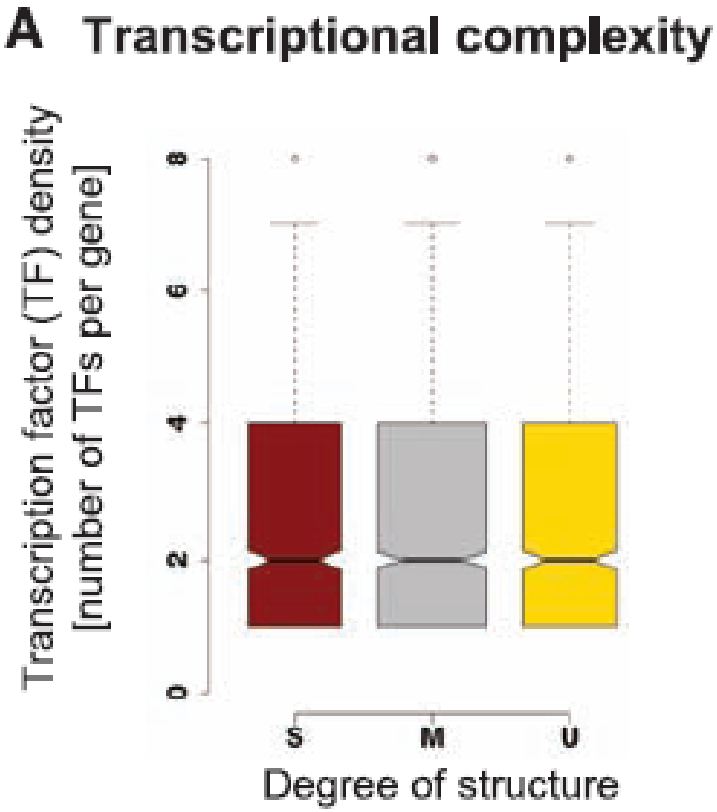
Kaudselt hinnatav ja ka otseselt mõõdetav



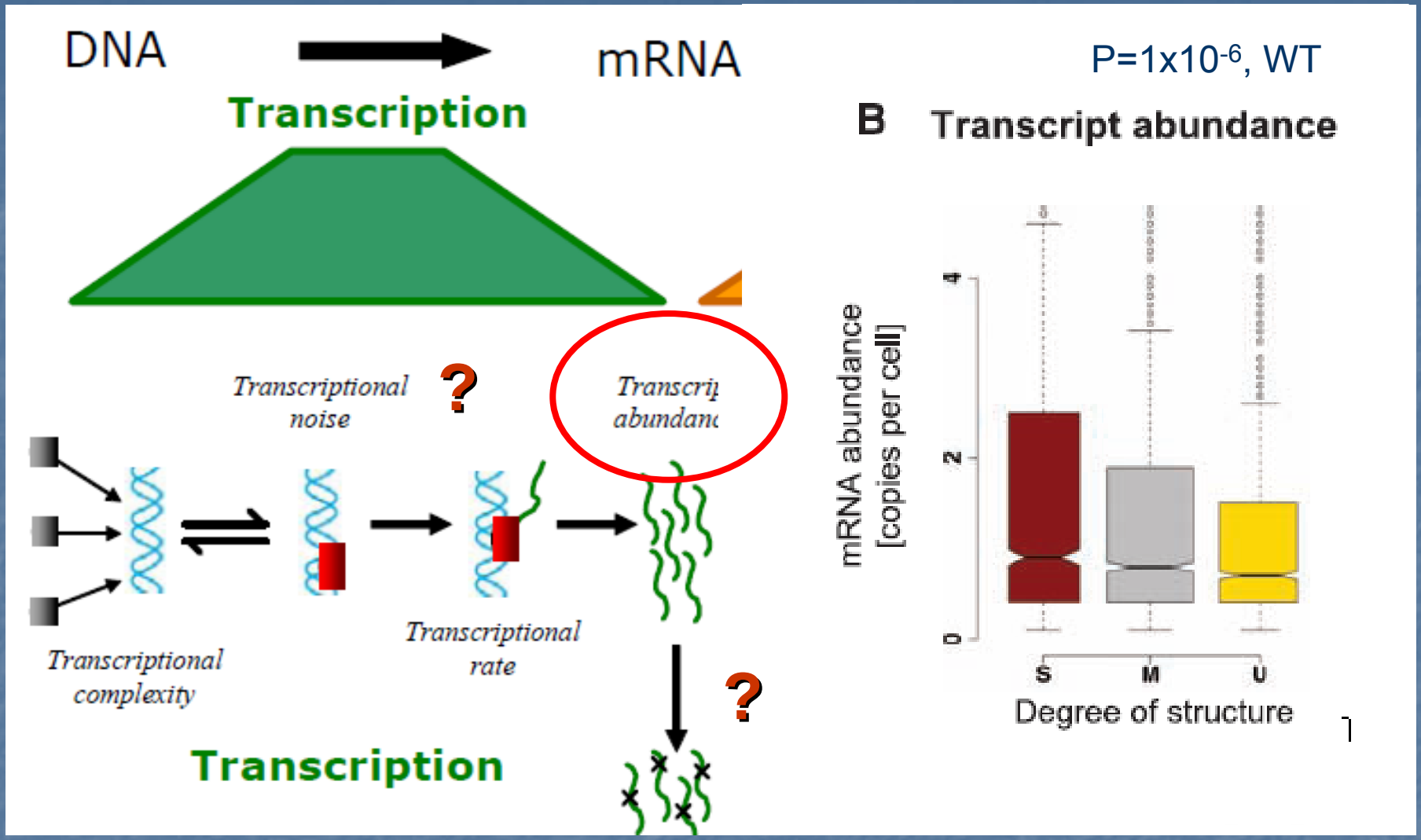
Because the steady-state amount of mRNA could be affected by the rate at which the transcripts are produced or degraded, they investigated whether the transcriptional rate or the degradation rate were different for the transcripts that encode ***highly structured*** and ***unstructured proteins***



P=0.55, Wilcoxon test



The number of transcription factors (TFs) that regulate a gene was comparable between the two groups (P=0.55, Wilcoxon test)



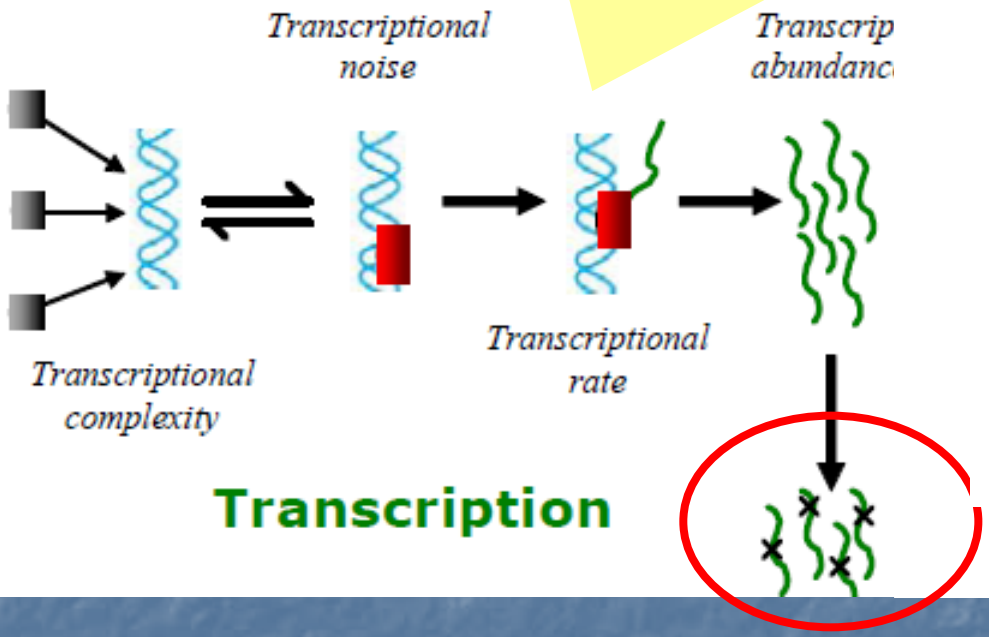
They see less mRNA for unstructured proteins

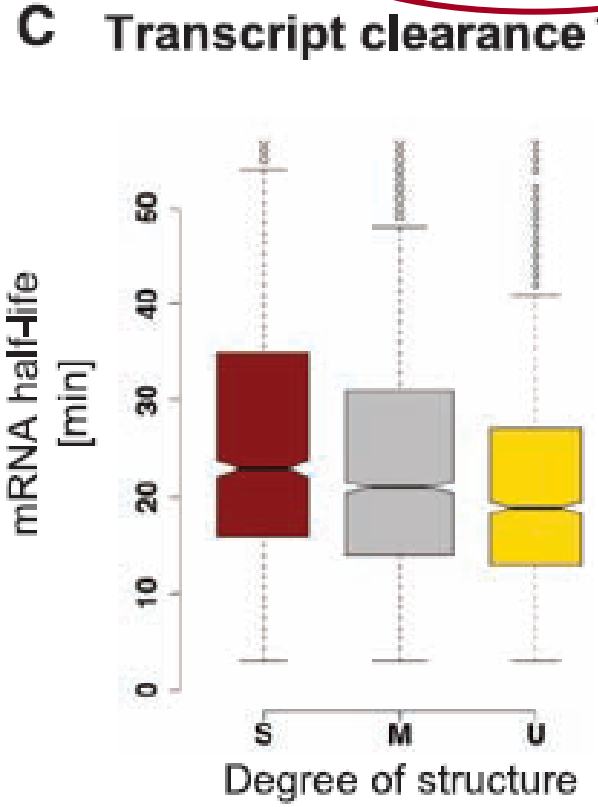
DNA → mRNA

P=1x10^{-16}, WT

Transcriptional rate

Highly Structured	2.4±0.1	
Mod. Structured	2.0±0.1	(0.04)
Unstr. Structured	**1.7±0.1**	(6x10^{-8})





Thus, differences in decay rates appear to be a major factor leading to differences in mRNA abundance (SOM S5)

mRNA degradation

- The two major pathways of mRNA decay are initiated by removal of the poly(A) tail

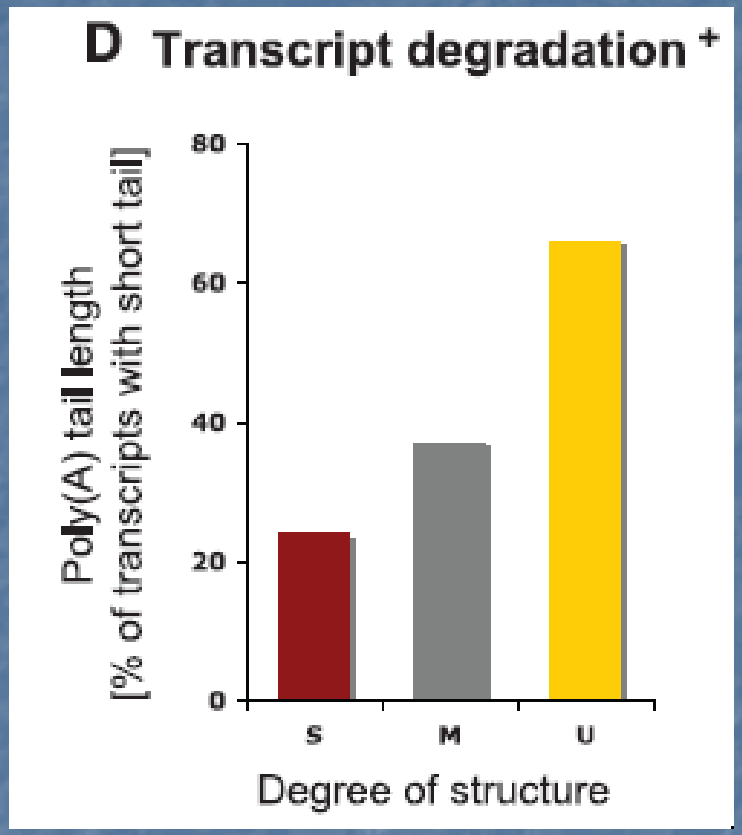
  They analyzed the presence of poly(A) tail

- Puf family RNA-binding proteins, which affect transcript stability

  They analyzed puf5 binding to mRNA

Poly(A) tail length

P=1x10^{-16}, Fisher exact test



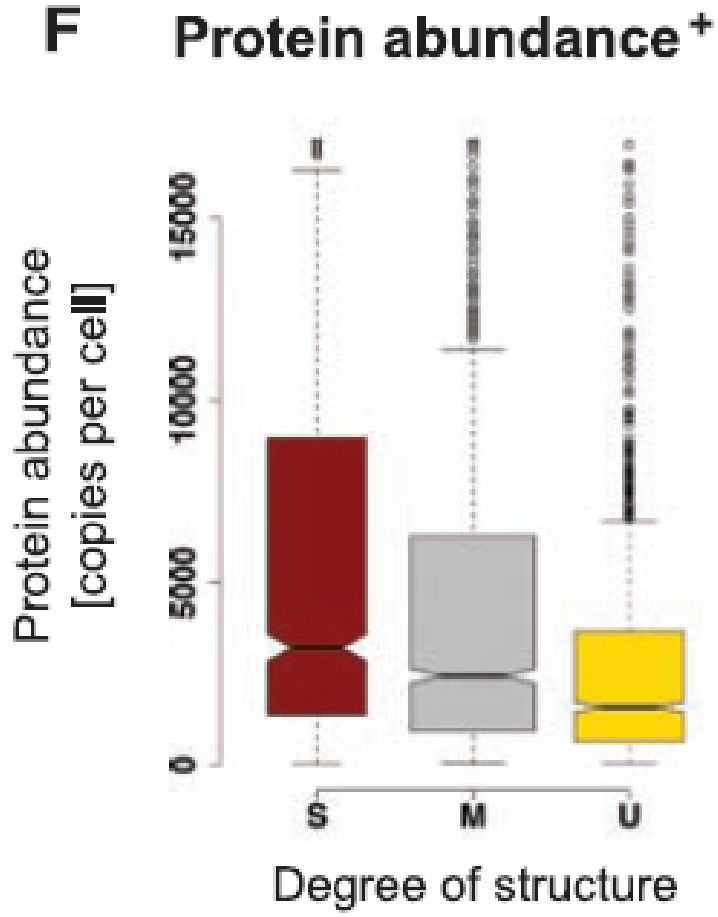
Puf5p binding

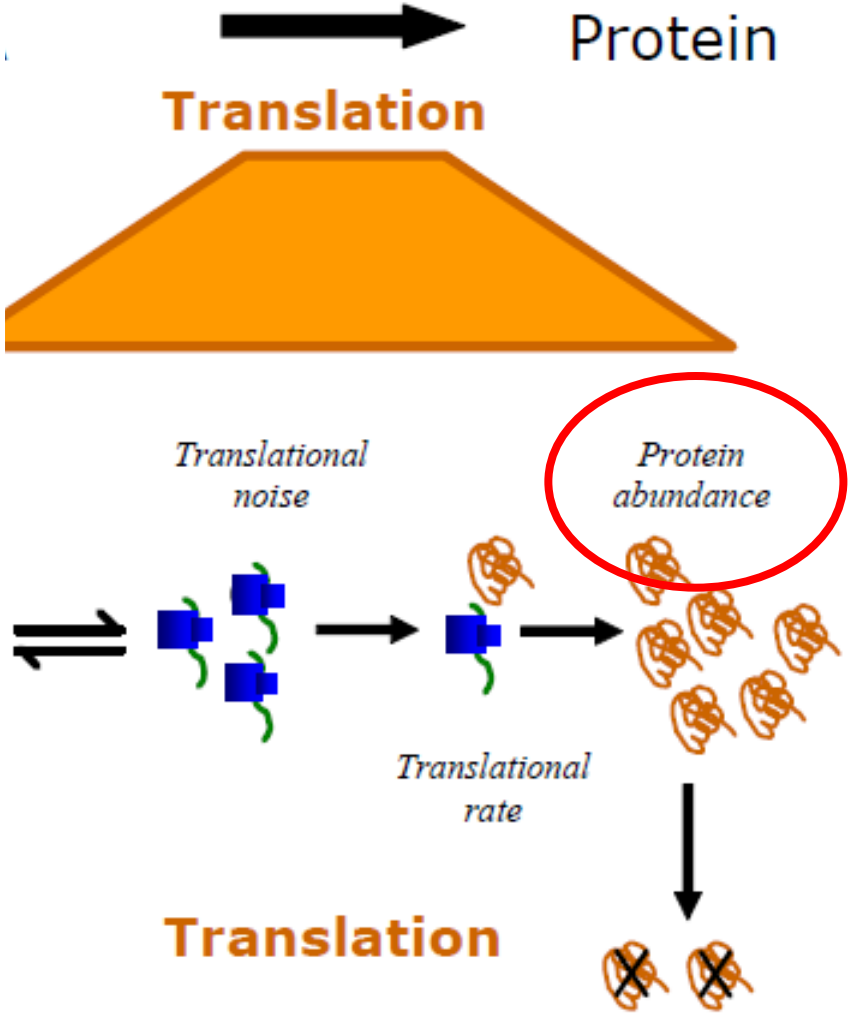
P = 5 × 10^{-10}

Puf5p binding was enriched for transcripts that encode highly unstructured proteins. In fact, 108 of the 224 transcripts bound by Puf5p encode highly unstructured proteins, a much greater number than expected by chance, which was 68 transcripts

Thus, poly(A) tail length and interaction with specific RNA-binding proteins may modulate the stability of transcripts encoding IUPs (SOM text S5)

P=1x10^{-16}, Wilcoxon test

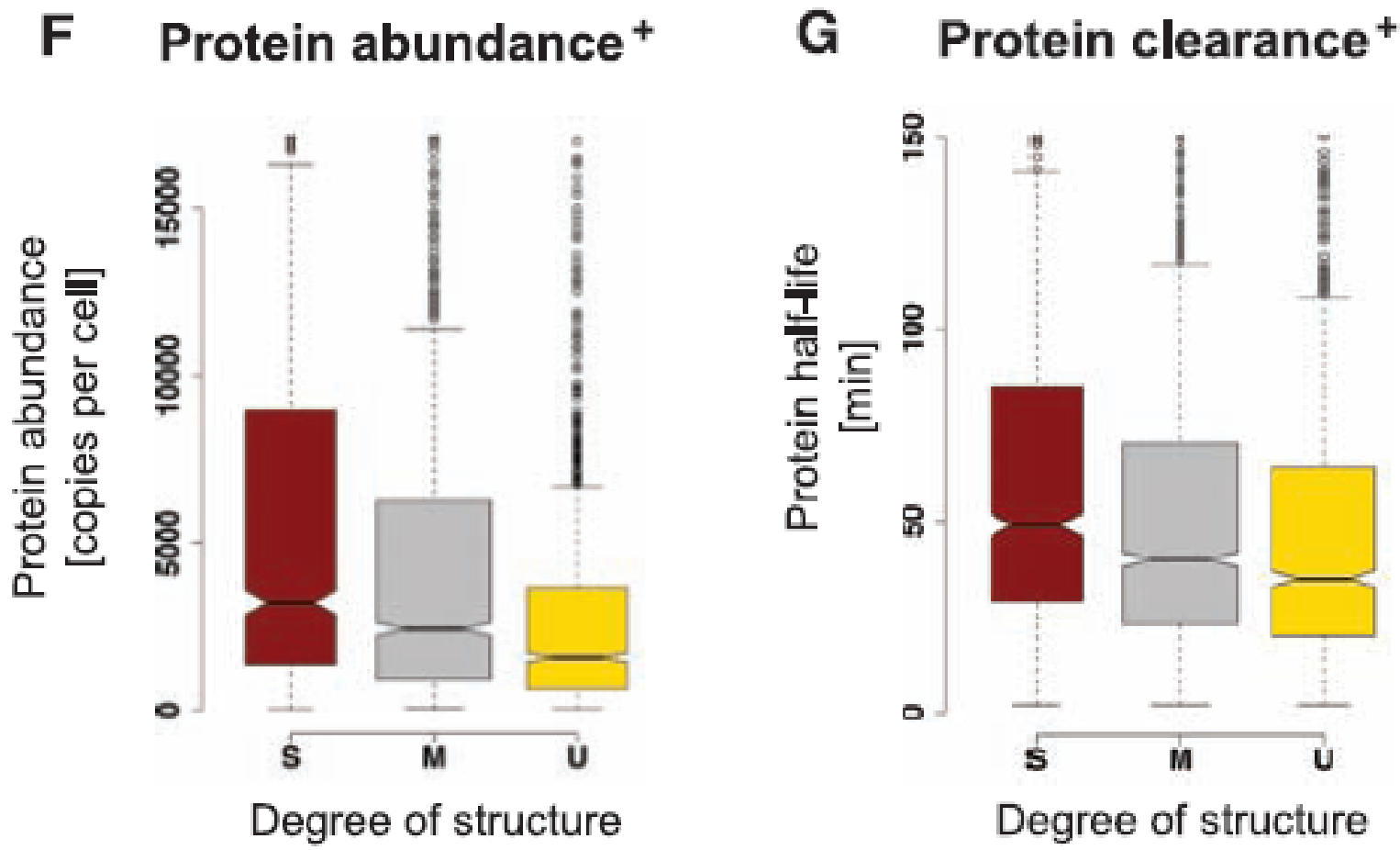




Unstructured proteins tend to be less abundant than structured proteins

P=1x10^{-16}, Wilcoxon test

P=1x10^{-15}, Wilcoxon test



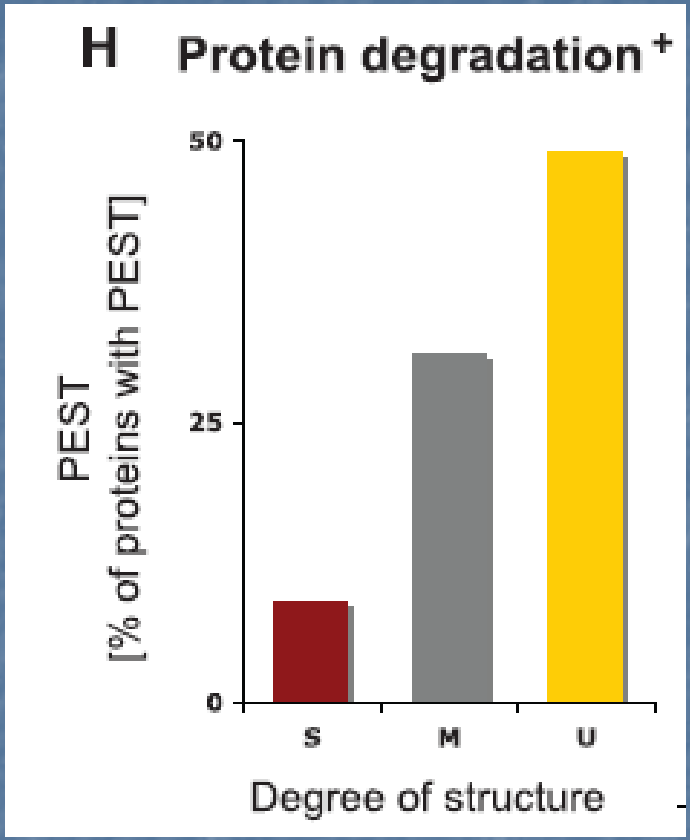
Unstructured proteins tend to be less abundant than structured proteins and they half life is shorter

Protein degradation

- Two pathways that mediate ubquitin proteasome–dependent degradation are:
  - the N-end–rule pathway and (certain amino acid at N-terminus leads to degradation pathway)
  - PEST–mediated degradation pathway (regions rich in proline, glutamic acid, serine, and threonin)

PEST pathway

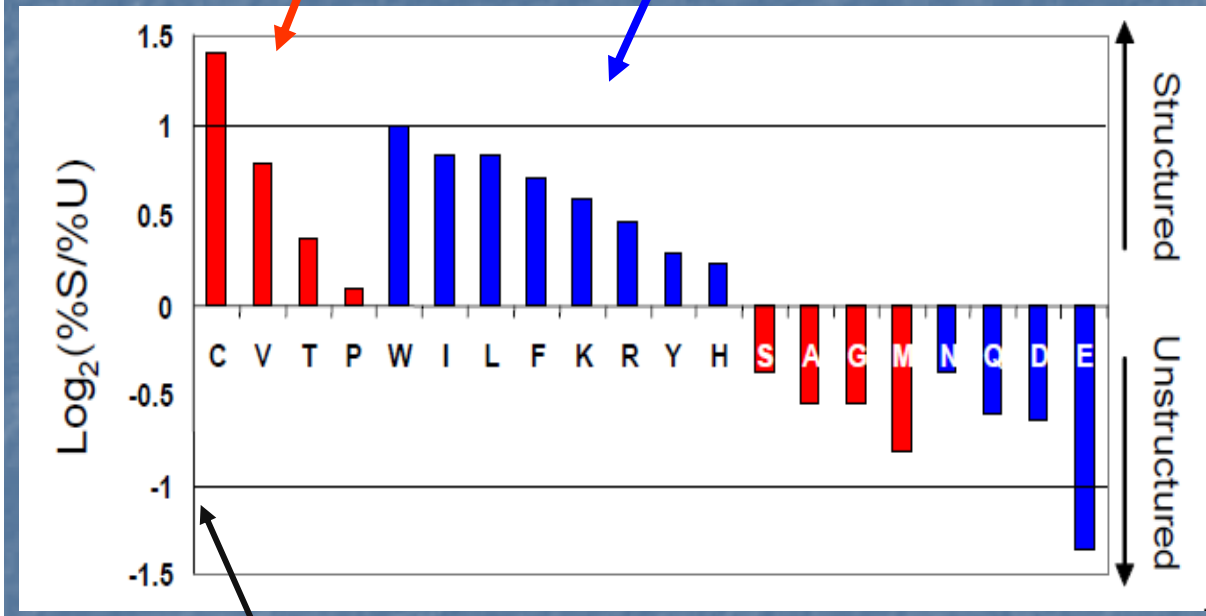
P=1x10^{-16}, Wilcoxon test



N-end-rule pathway

Pikk eluiga

Lühike eluiga

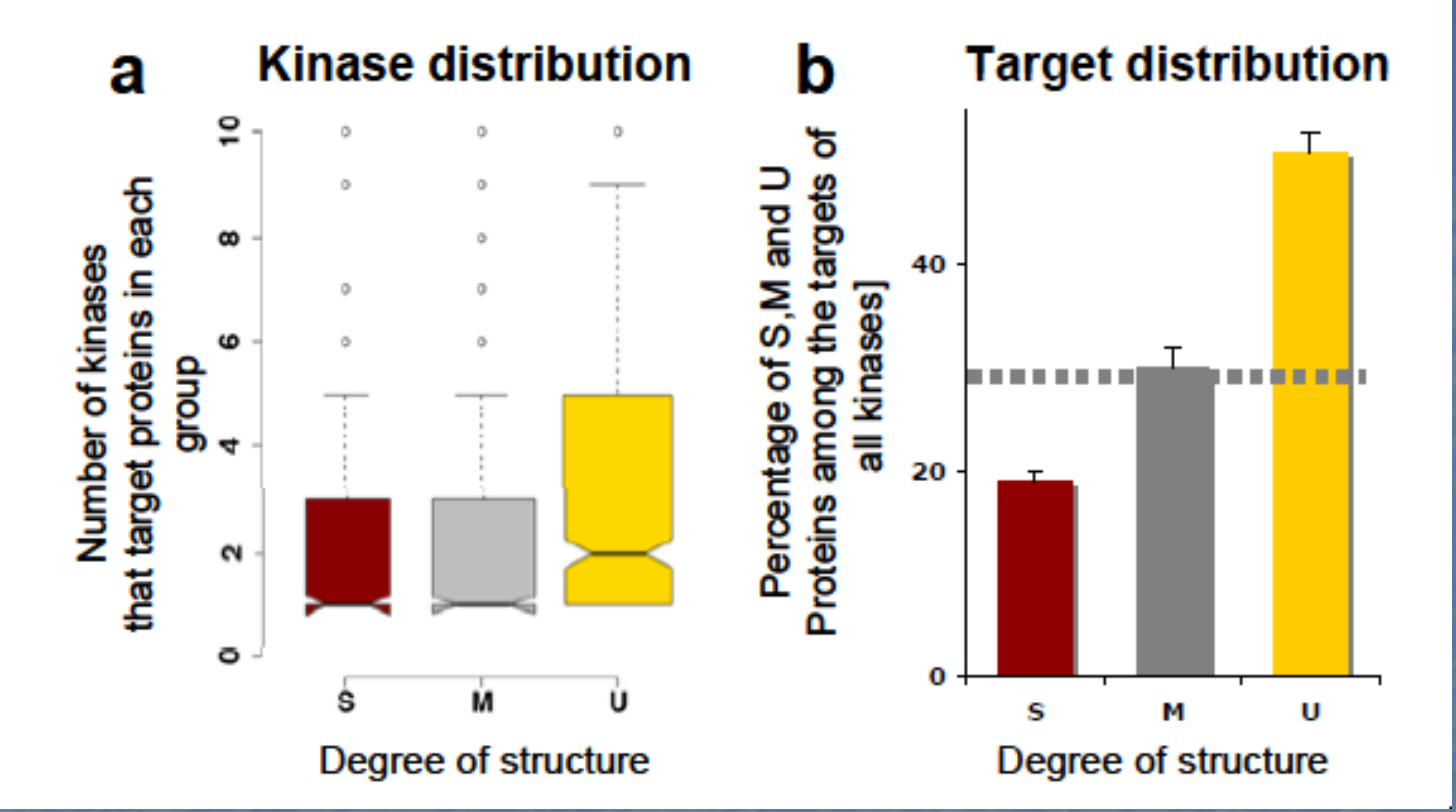


Olulisuse piirid +1 ja -1

Therefore, it appears that the availability of many IUPs is regulated via proteolytic degradation and a reduced translational rate.

Posttranslational modification

- For certain IUPs (for example, p27), posttranslational modifications such as phosphorylation can affect their abundance or half-life in a cell
- They analyzed the experimentally determined yeast kinase-substrate network to determine whether there is difference to be substrate for kinases for unstructured and structured proteins



1. highly unstructured proteins are on average substrates of twice as many kinases as are structured proteins ($P = 1 \times 10^{-12}$, Wilcoxon test)
2. On average, 51 ± 19% (SD) of all substrates of the kinases are highly unstructured,

   whereas only 19 ± 13% (SD) are highly structured

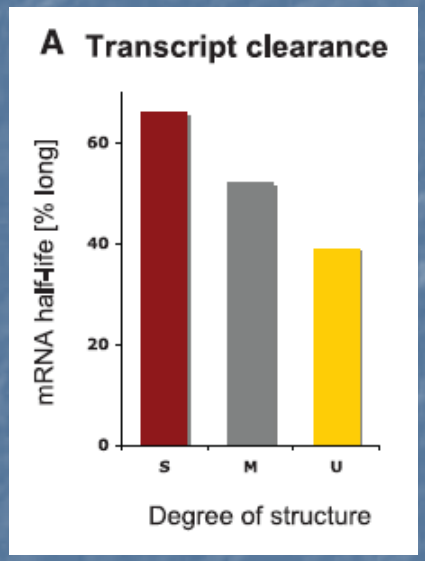
Kinases and cells response to different conditions

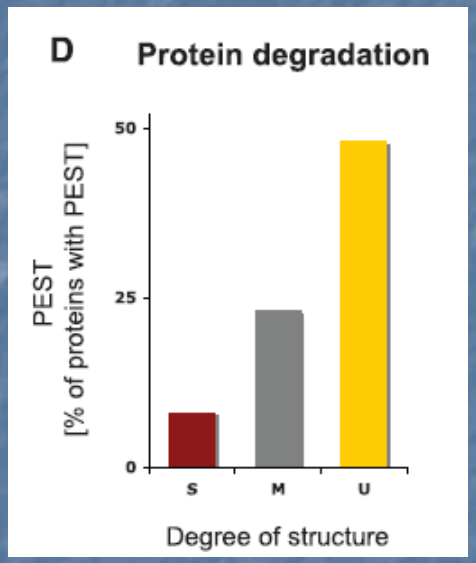
- They found that 85% of the kinases for which more than 50% of their substrates are highly unstructured are either regulated in a cell cycle–dependent manner (for example, Cdc28) or activated upon exposure to particular stimuli (for example, Fus3) or
  stress (for example, Atg1)

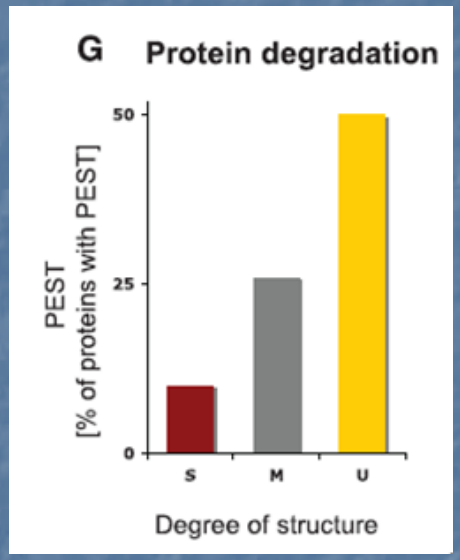
Saccharomyces cerevisiae

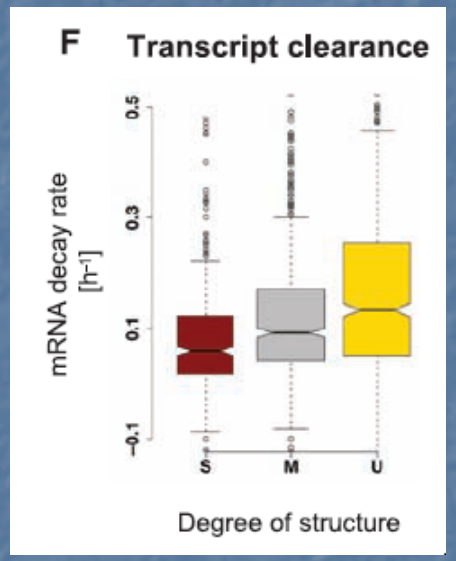
Universality

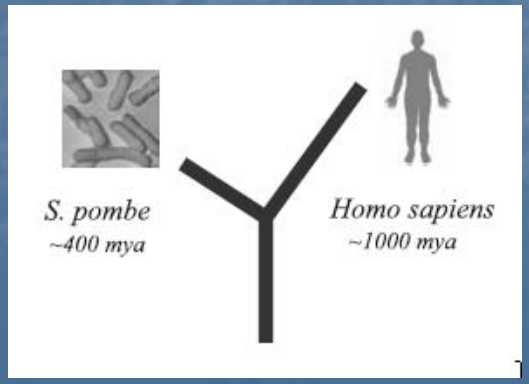
Schizosaccharomyces pombe & H. sapiens











Universality

Schizosaccharomyces pombe & H. sapiens

- Similar trends to those observed for S. cerevisiae were evident in these organisms
- Both unicellular and multicellular organisms appear to regulate the availability of IUPs
- The observed differences between structured and unstructured proteins were independent of:
  - the IUP prediction method used,
  - protein length,
  - localization within the major subcellular compartments,
  - different grouping of proteins, or
  - the number of interaction partners per protein

Conclusions

- IUPs are tightly regulated
  - Transcriptional level (synthesis and degradation)
  - Translation level and (synthesis and degradation)
  - Posttranscriptional modification (affects lifetime, localization and target)
- Their studies reveal an evolutionarily conserved tight control of synthesis and clearance of most IUPs
- The discovery was made possible by integrating multiple large-scale datasets that describe control mechanisms during transcription, translation, and post-translational modification with structural information on proteins

Tänan kuulamast !