Technical Journal Club

September 26, 2017

Patrick Schürch

nature biotechnology

Global analysis of protein structural changes in complex proteomes

Yuehan Feng[1,5], Giorgia De Franceschi[1,2,5], Abdullah Kahraman[1,5], Martin Soste[1], Andre Melnik[1], Paul J Boersema[1], Patrizia Polverino de Laureto[2], Yaroslav Nikolaev[3], Ana Paula Oliveira[4] & Paola Picotti[1]

Introduction

The cell regulates protein function by modulating its…

- expression levels,
- protein-protein interactions
- and chemical modifications.

These changes are routinely measured on a global level by MS-based proteomic based techniques.

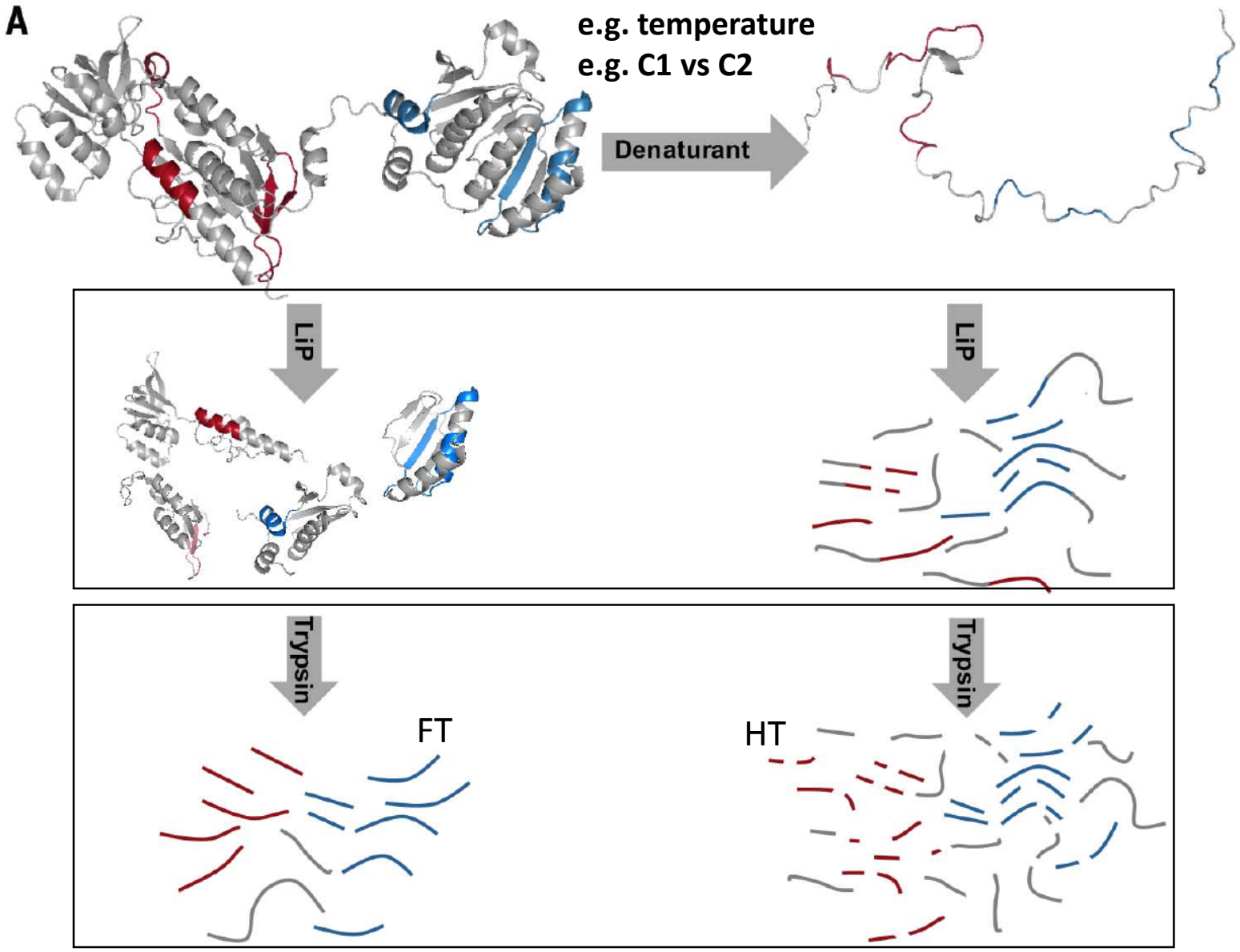
- One exception remains the assessment of global changes in protein folds.

The structure of proteins influences their function and can be triggered by binding of ligands, interactions with other proteins, mutations or environmental stimuli.

- specific proteins undergo structural changes that result in the formation of insoluble deposits (amyloidoses and NDD)
- complex biologicals matrices

LiP-SRM is a new method that allows to study structural changes of proteins on a global level in complex biological environments.

Workflow of the analysis of protein structure (based on LiP-MS)



LiP = limited/native proteolysis (Proteinase K or Thermolysin)

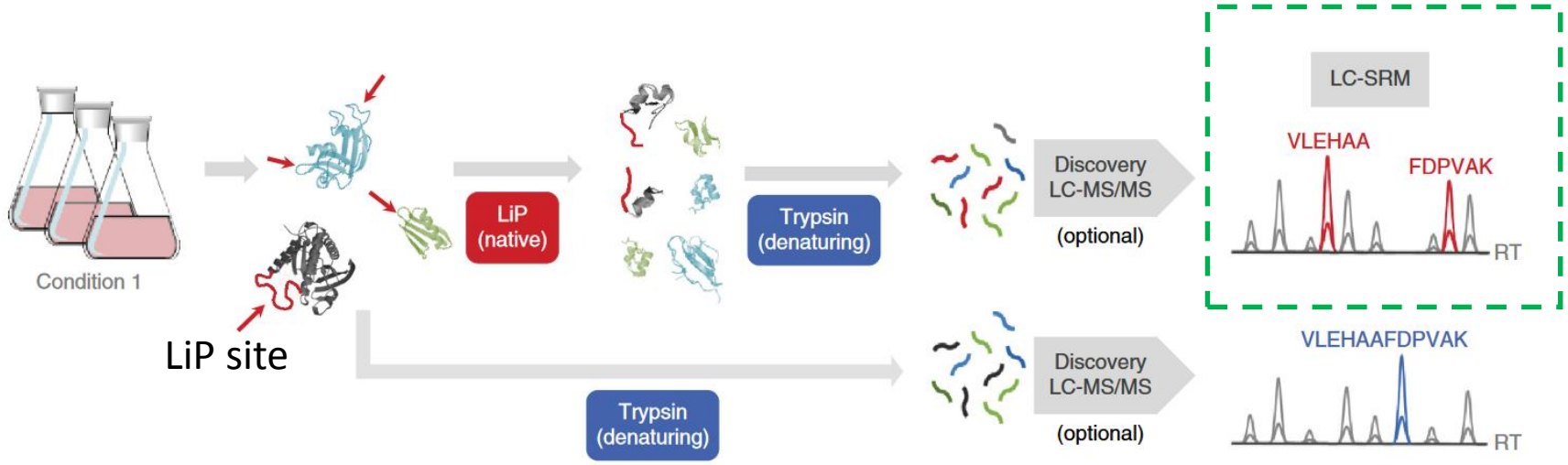
- Mechanical proteome extraction
- Low E/S ratio
- **Structure-specific proteolytic patterns**
- LiP cleavage occurs preferentially at locally **unstructured segments** (FT vs. HT peptides)

Trypsin digest

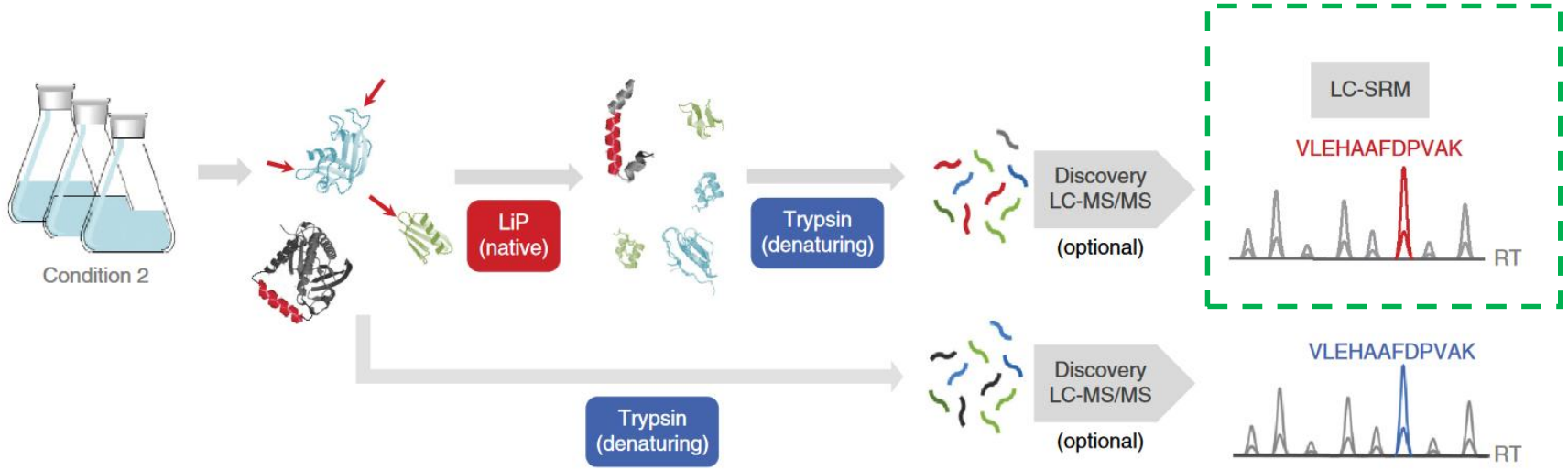
- **Denaturing** conditions
- **Not structure-specific**
- Generating peptides for **bottom-up proteome analysis**
- Controls: trypsin-only digested samples

LiP-SRM workflow

Conformation of protein XY is sensitive to Proteinase K digestion



…not sensitive to Proteinase K digestion



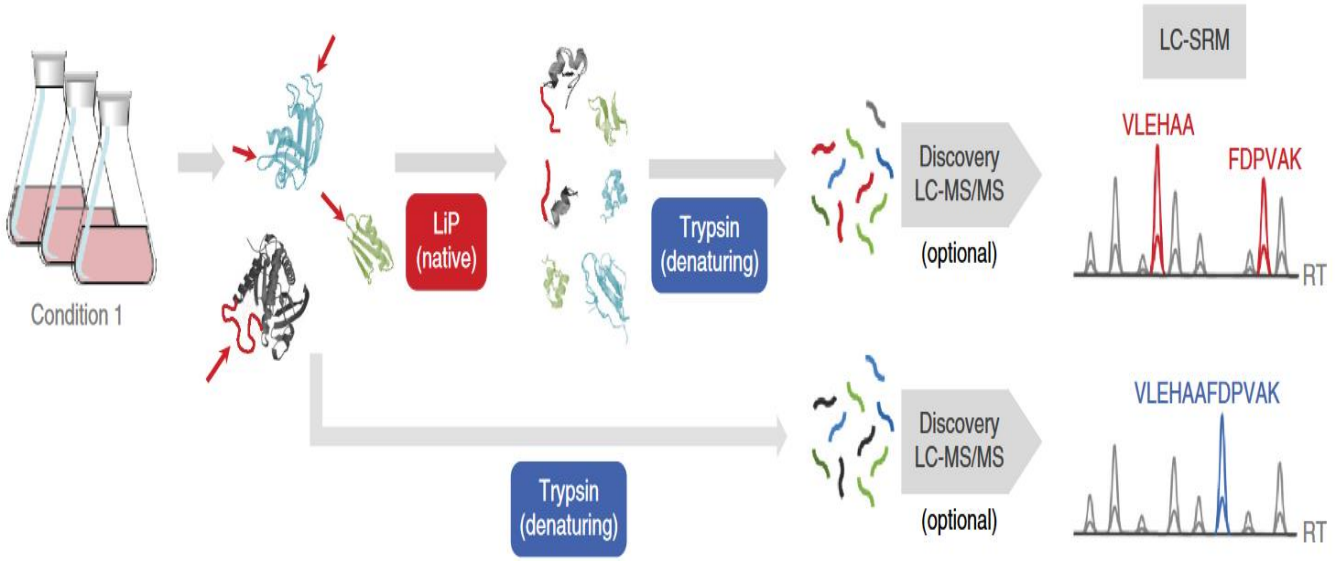
Initial proteolysis is dictated by structural properties of the substrate. A fully tryptic peptide containing a LiP cleavage site will be detected in the trypsin control and replaced by two half-tryptic halves in the sample subjected to LiP.

1) LiP (=limited/native proteolysis)

- LiP digestion
- Trypsin digestion

2) MS

- exploiting the high sensitivity and background filtering capabilities of **selected reaction monitoring** (SRM) MS
- «Precursor ion» ABC+ => MS-1 => AB and «product ion» C+ => MS-2



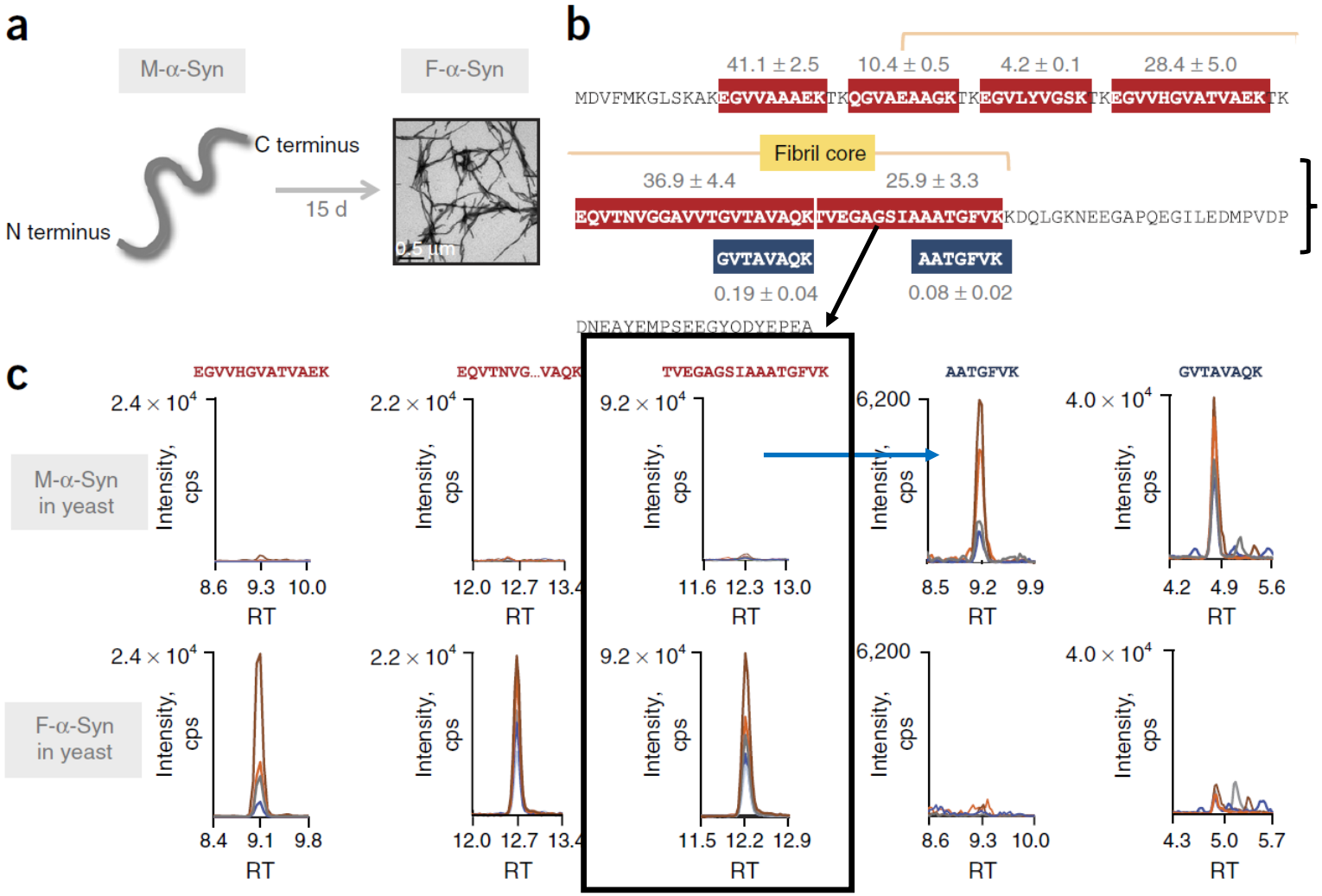
Fully tryptic peptides that are less abundant in doubly-digested sample probably contain LiP sites

Follow-up SRM analysis of all possible **half-tryptic peptides** that could be generated by internal cleavage

Identify peptides that are specific to a given protein conformation i.e. **«conformotypic peptides»**

Using liP-SRM to measure amyloid-forming protein

- Alpha-Syn is predominantly unfolded under physiological conditions => **M**-α-Syn (**m**onomeric)
- Parkinson's Disease: α-Syn switches to a β-sheet-rich fold and polymerizes into **f**ibrillar, amyloid aggregates => **F**-α-Syn
- **yeast proteome** spiked with either M- or F-α-Syn => SRM peaks obtained for **conformotypic peptides** in the cell extract compared

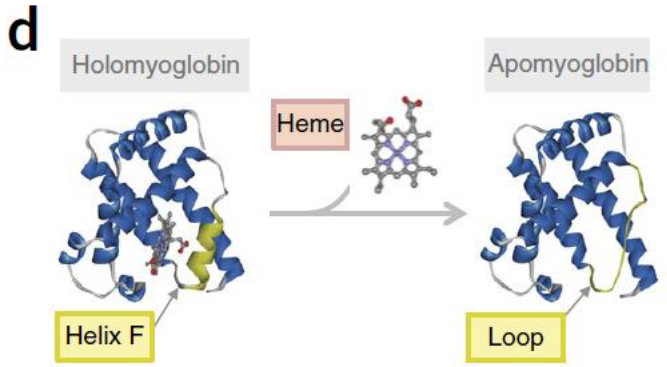


Fully tryptic peptides that map to the amyloid β-sheet-rich core of F-α-Syn.

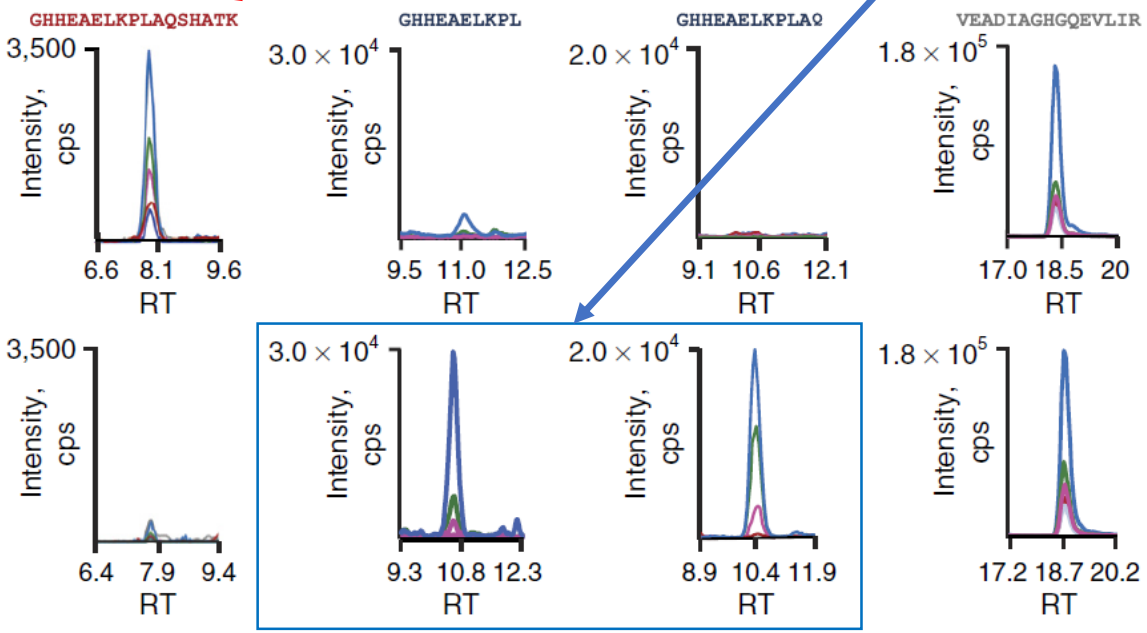
- These regions are protected from PK in F-α-Syn and 26-37 x more abundant.
- In M-α-Syn they are disordered, resulting in half-tryptic peptides.

Using LiP-SRM to detect subtle structural transitions

- Holomyoglobin => globular, with an α-helical fold
- Apomyoglobin => Based on NMR data, 7/8 helices remain structured, but helix F (13 amino acid) becomes disordered
- Spiked yeast proteome with either Holo- or Apomyoglobin



Fully tryptic peptide that encompasses helix F: only detected in Holomyoglobin-containing sample at a similar intensity as in Trypsin control



Cleavage site that remains the same in holo- or apo- conf.

Red region is fully tryptic in holo-conf. but is not protected from proteinase K in apo-conf. (half-tryptic)

Using liP-SRM to detect subtle structural transitions

Controls

- LiP patterns were qualitatively and quantitatively independent of the abundance of target protein substrates in complex proteome extracts.
- Different E/S ratios, protease incubation times and types of protease were tested.

Can conformotypic peptides be used to calculate the amount of each conformation in a mixture of two conformational states?

=> Mixtures of HoloMb (FT peptide) and ApoMb (HT peptide) containing either 90%, 50% or 10% ApoMb.

Supplementary Table 6. Comparison of the apoMb/holoMb mixture derived from the SRM measurement of conformotypic peptides for holo- and apoMb

	Peptide	Average % Apo, in 10% Holo Mix	Average % Apo, in 50% Holo Mix	Average % Apo, in 90% Holo Mix
Calculated via FT	GHHEAELKPLAQSHATK	93.3 +/- 0.5	48.9 +/- 2.1	19.3 +/- 3.5
Calculated via HT	GHHEAELKPL	72.2 +/- 1.5	63.5 +/- 2.4	10.4 +/- 4.5
	GHHEAELKPLAQSHATK and GHHEAELKPL, combined	82.8 +/- 11.3	56.2 +/- 8.0	14.8 +/- 6.0

Apomyoglobin 90% 50% 10%

Analyzing protein conformational changes on a large scale

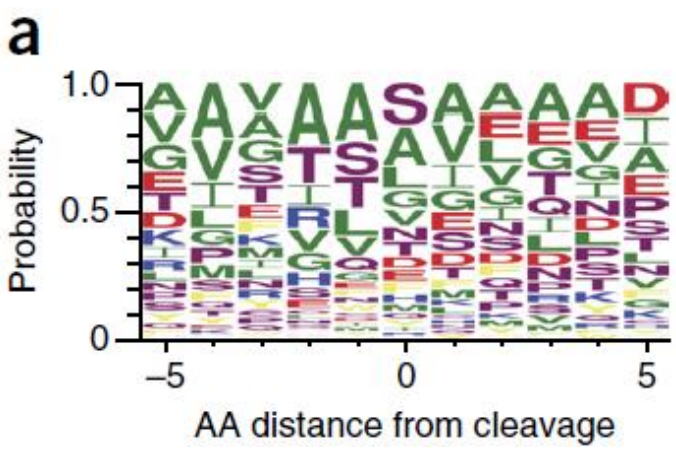
Is the LiP-based approach suitable unbiased identification of proteins that undergo structural transitions upon a given stimuli?

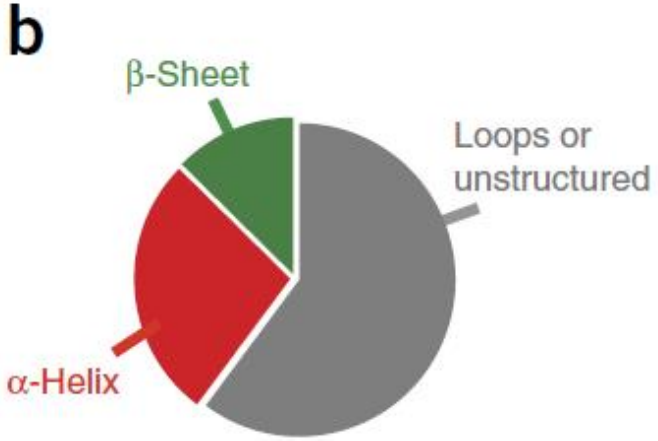
⇒ the metabolic transition from glucose- to ethanol-based growth in yeast
⇒ 12 samples (2 conditions, biological triplicate, each sample subjected to LiP/TRP or TRP-only digestion) analyzed by LC-MS/MS followed by SRM-based validation

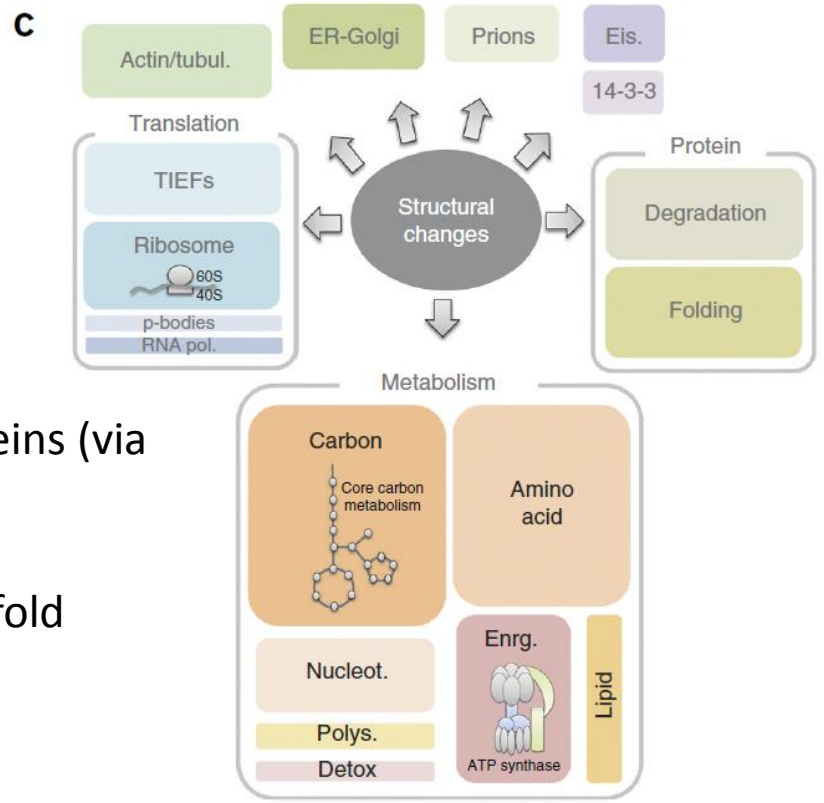
The shotgun analysis identified **21'899 peptides** mapping to **1'622 unique proteins** (FDR 1%)

- **4'267 LiP sites** mapping **to 1'001 proteins** were identified (PK vs. trypsin-only)
- The proteins associated with these LiP sites had copy numbers/cell in a range of < 2 x10E3 → 1 x 10E6

Analyzing protein conformational changes on a large scale







No obvious **consensus sequence** of **proteinase K** was observed (a)

They **mapped the HT peptides** to the available structures of the corresponding proteins (via protein data bank and homology models) (b)

586 HT peptides (mapping to **283 proteins**) had an altered abundance of at least 2-fold (after normalization for protein abundance and intracellular proteolytic activity) (c)

Conformational changes in core carbon metabolism proteins

<u>LiP-SRM analysis of carbon metabolism</u> upon a shift from glycolytic to gluconeogenic growth was done for *S. cerevisiae*

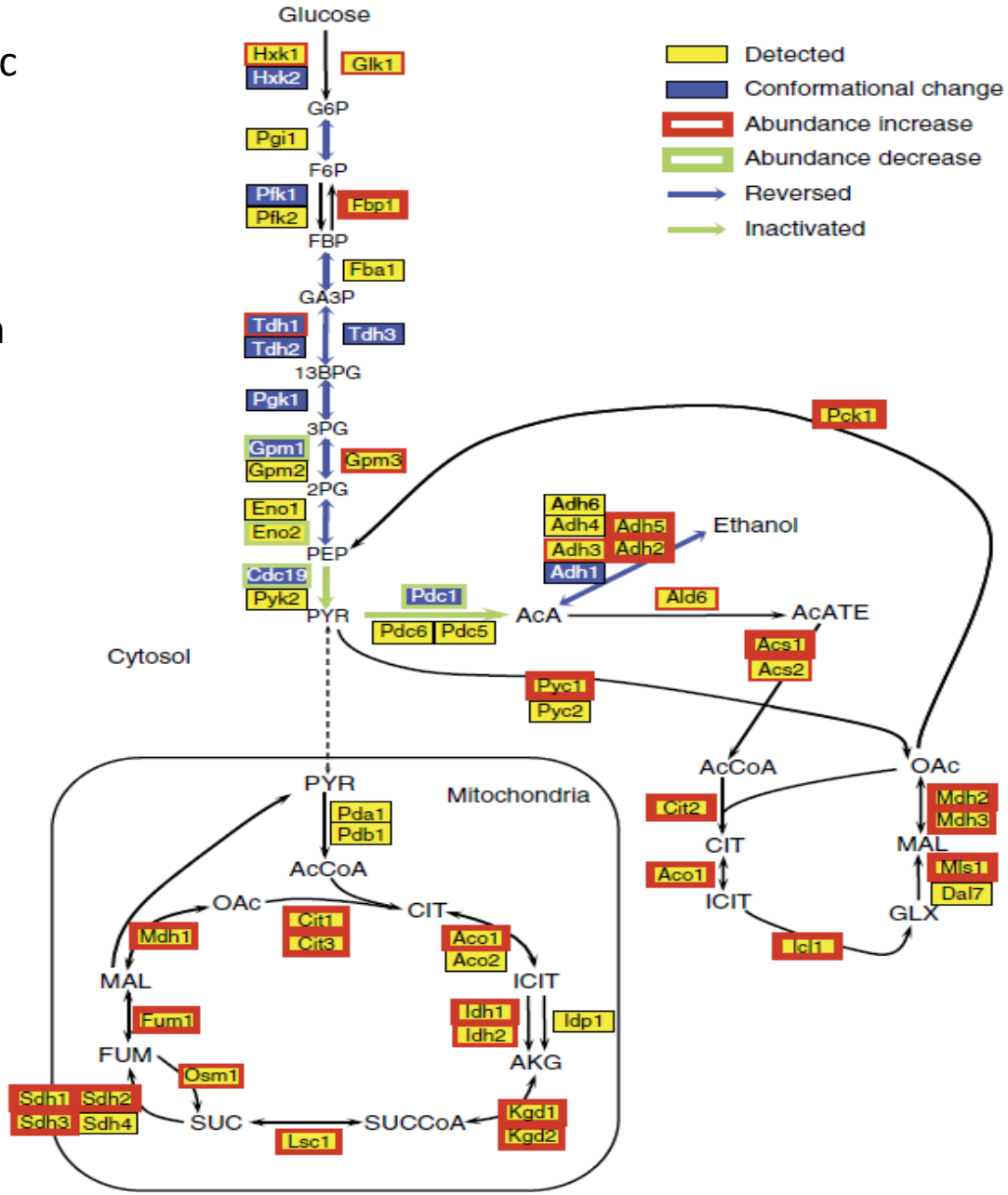
1. Enzymes of **TCA** and **glyoxylate cycles**: protein levels "↑"

2. Enzymes involved in **glycolysis** and **ethanol production**: protein levels "="
3. Levels of 19 half-tryptic LiP peptides changed significantly between the two conditions i.e. **structural transition**

⇒ These HT peptides mapped to 10 enzymes involved in either **glycolysis** or **ethanol production**.

⇒ Correlated with **metabolic flux data** (reversion or inactivation), not observed in TCA enzymes

Strikingly, these **structural changes** and **flux alterations** mostly occurred in enzymes that underwent no or minor abundance changes (not observed for the TCA or glyoxylate enzymes).

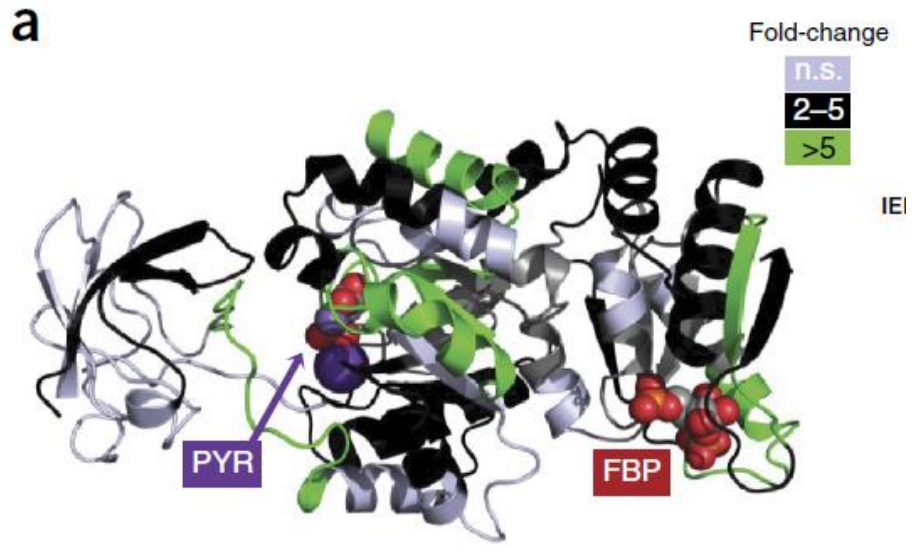


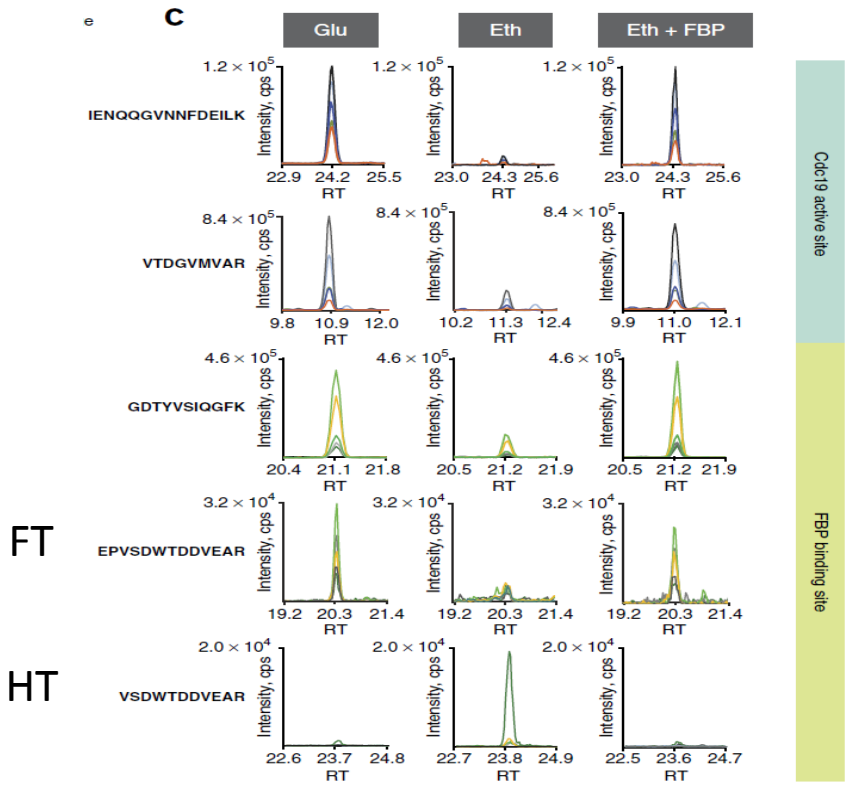
Structural changes of Cdc19 upon a switch from glucose- to ethanol-based metabolism

Reaction catalyzed by the **pyruvate kinase** (Cdc19) has to be inactivated upon the switch from glucose to ethanol.

The observed **structural changes** were due to **allosteric inactivation** since the peptides with the largest fold changes mapped to the **active site** and the **binding site of an allosteric regulator** of Cdc19 (fructose-1,6-biphosphate = FBP).

The concentration of **FBP** is 100-fold higher in glucose than in ethanol.





- FC of fully tryptic peptides in EtOH vs. Glu
- **Cdc19 is resistant to LiP when active**

Addition of FBP into EtOH medium reversed the «EtOH»-proteolytic pattern to that observed in glucose (Glu + excess FBP not shown)

Discussion

Limitations

1) The **resolution is lower** than that of NMR or X-ray crystallography, but can be improved by long separation gradients and chromatographic columns.

2) LiP-SRM will produce an **average description** of the conformational properties. If the protein is represented by a mixture of different **confomers**…
⇒ derive **conformotypic peptides for each conformation** separately (purified, *in vitro*)
⇒ the quantification of such markers in a heterogenous sample serves as an estimate of the conformational composition (see Myoglobin data) *in vivo*

3) Conformations are **not sensitive to LiP** cleavage
- Use for multiple digestion times/conditions

4) Soluble proteins vs. insoluble proteins

5) **Posttranslational modification** or **protein-protein interactions** might affect proteolytic patterns due to **sterical hinderance**
- e.g. binding partner blocks proteinase K
- Trypsin control is done for both/all conditions

6) **Loss of subcellular compartmentalization** during cell lysis
⇒ “artificial” interactions that would not occur *in vivo*

RESEARCH ARTICLE

PROTEOMICS

Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability

Pascal Leuenberger,[1,2] Stefan Ganscha,[2,3] Abdullah Kahraman,[4] Valentina Cappelletti,[1] Paul J. Boersema,[1] Christian von Mering,[4] Manfred Claassen,[3] Paola Picotti[1]*

Introduction

Temperature profoundly influences the physiology of cells

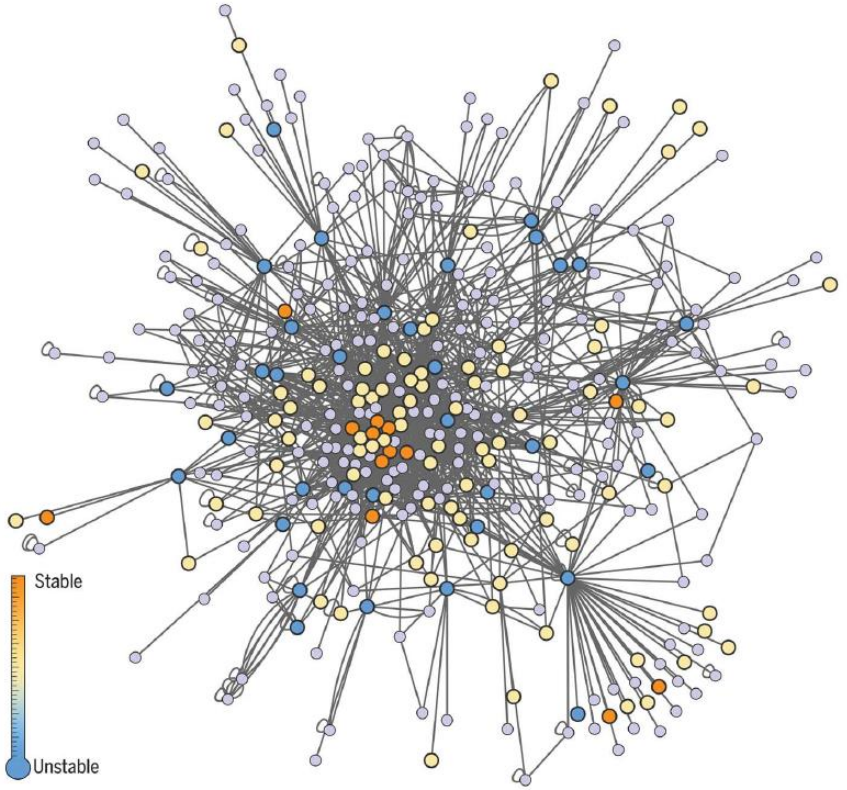
- Spermatogenesis, sex determination in amphibians, fever etc.

Understanding **thermotolerance** will contribute to various biological, clinical and biotechnological implications.

Thermal sensitivity of cells is caused by protein denaturation.

Scenario 1: **Global loss** of protein structures.

Scenario 2: **Loss of key proteins** (hubs and bottlenecks).

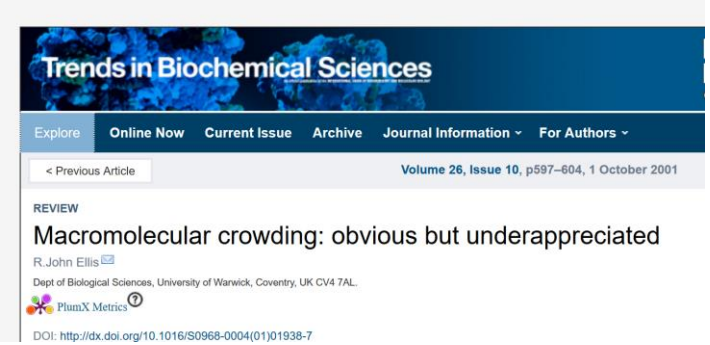


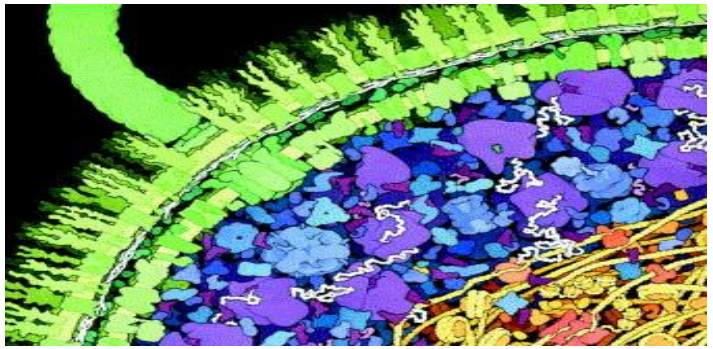
Protein-protein interaction network of *E. coli*. Node color indicates protein thermostability. Blue, unstable; yellow, medium-stable; orange, stable; gray, not measured. At the temperature of thermal cell death of *E. coli*, a subset of highly connected protein nodes involved in key cellular processes undergoes temperature-induced denaturation.

Scientific question paper 2

Large-scale stability analyses so far have used data from **purified proteins**.

- **Heterogenous data sets** (experimental conditions, different technologies etc.).
- The **influence of the cellular matrix** on protein stability/structure was not taken into account (chaperones, molecular crowding, binding events etc.).
- These factors would also influence current estimations of **intrinsically disordered proteins (IDPs).**





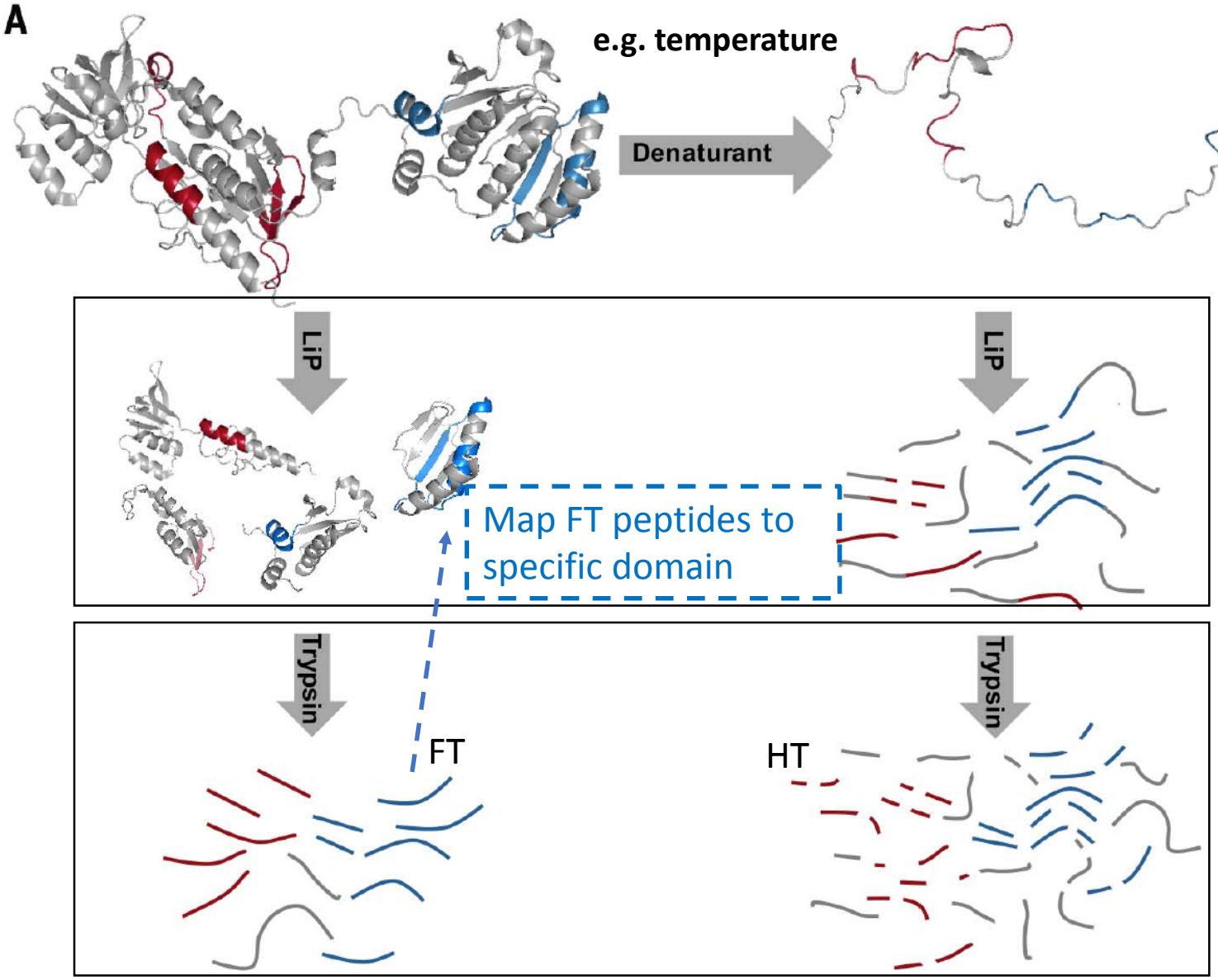
www.cell.com

Goal: Evaluate thermal stabilities/structural changes of proteins on a proteome-wide scale directly in a complex biological matrix!

⇒ LiP-MS on proteomes of *Escherichia Coli, Saccharomyces cerevisiae, Thermus thermophilus* and human cells

⇒ Evaluate *in vivo* determinants of protein thermostability, prevalence of IDPs and evolutionary conservation of protein thermal stability

Workflow of the analysis of protein stability (based on LiP-MS)



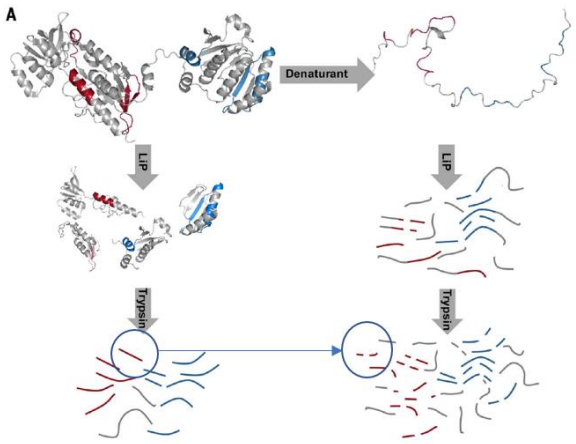
LiP (Proteinase K or Thermolysin)

- Native proteome extraction (mechanical)
- **structure-specific** proteolytic patterns
- LiP cleavage occurs preferentially at locally **unstructured segments** (FT vs. HT peptides)
- Specific sites are protected from PK in the folded form

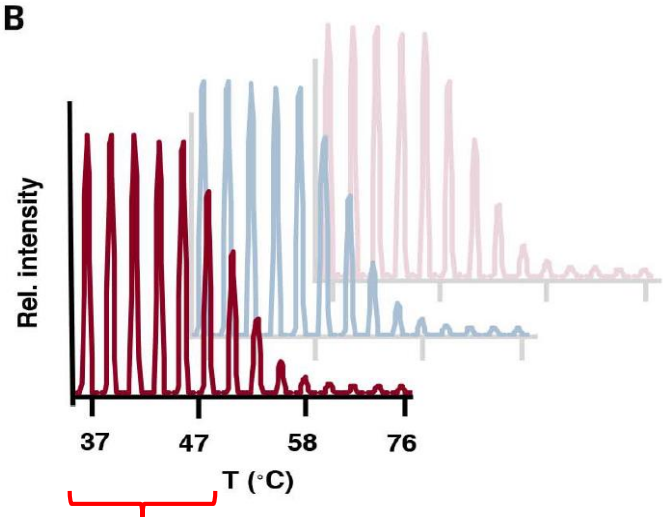
Trypsin digest

- Denaturing conditions
- Not structure-specific
- Controls: trypsin-only digested samples

Extracted proteomes are subjected to 14 temperature conditions prior to LC-MS/MS analysis of proteolytic pattern (fully tryptic, FT / half-tryptic, HT peptides are quantified).

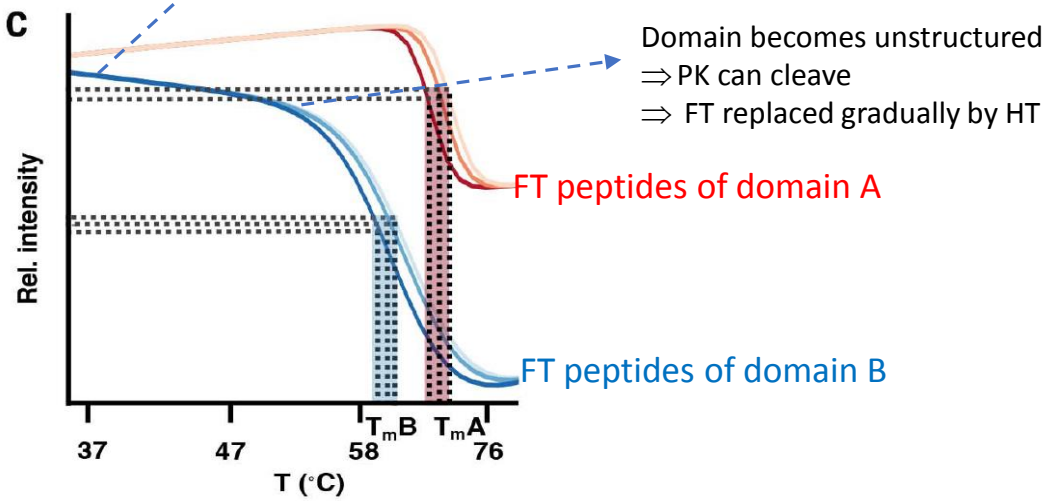


- 14 temperature conditions (37-76°C)
- peptides are quantified via shotgun MS
- Map peptides to domains



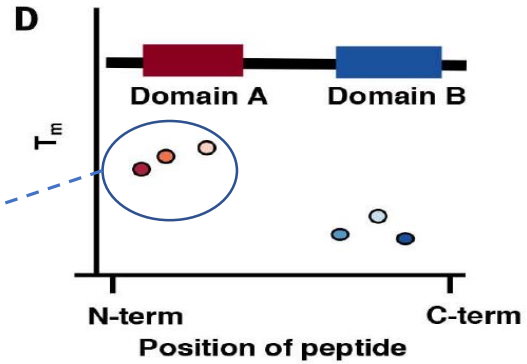
FT peptide (protected from PK)

Domain structured
High abundance of FT peptide (protection from PK)

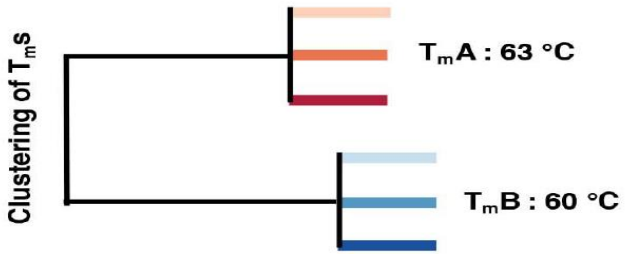


Signal intensity of each FT vs. time => apparent Tm

Domain A contains several FT with their resp. Tms (should be in similar range…)



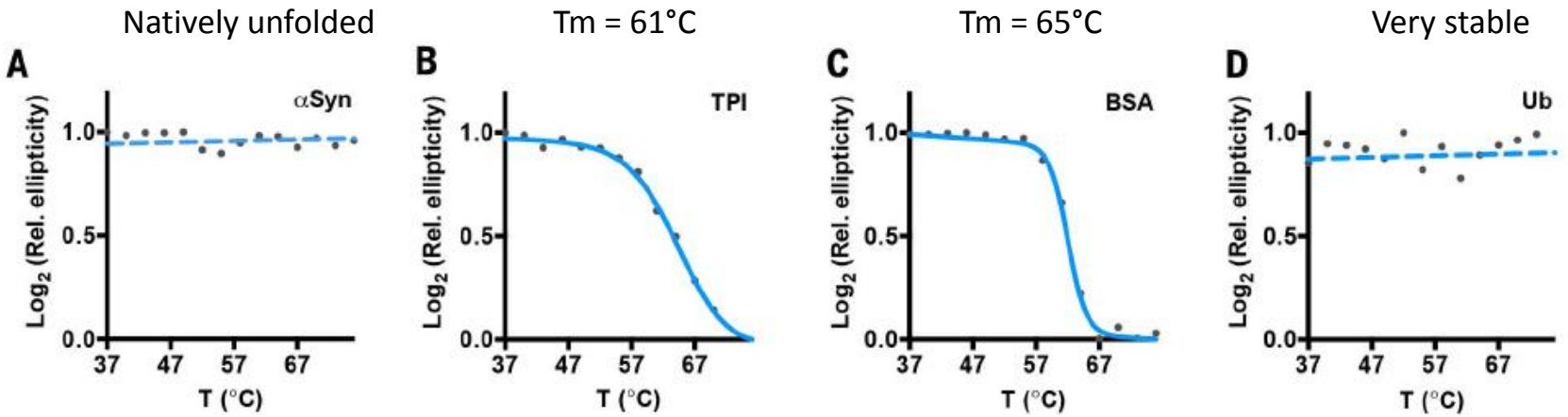
Domain A is more stable than B



Hierarchical clustering of peptide profiles

- When comparing overall Tm between proteins => use domain with lowest Tm

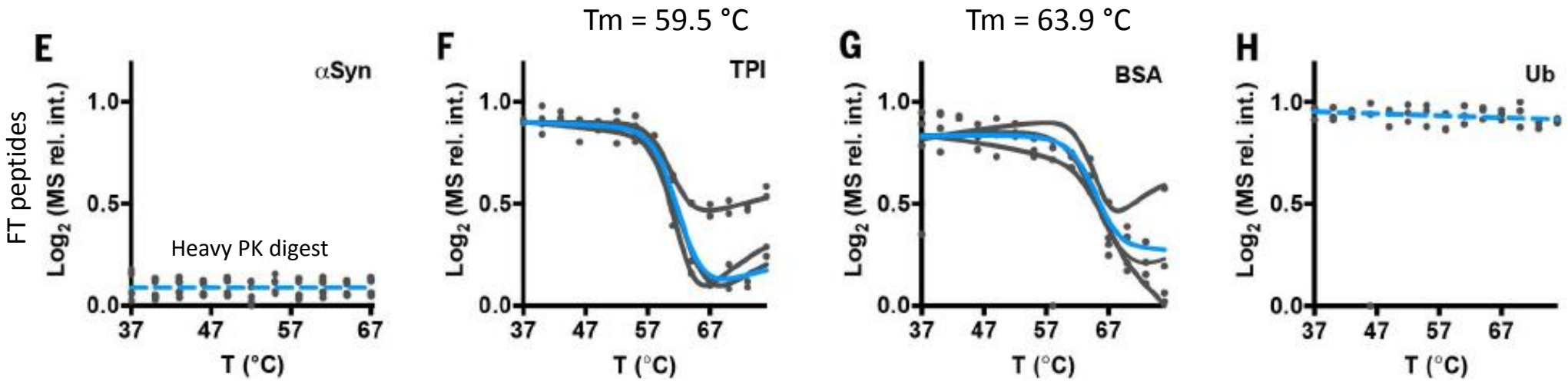
Experimental validation of approach



Circular dichroism analyses (*in vitro*)

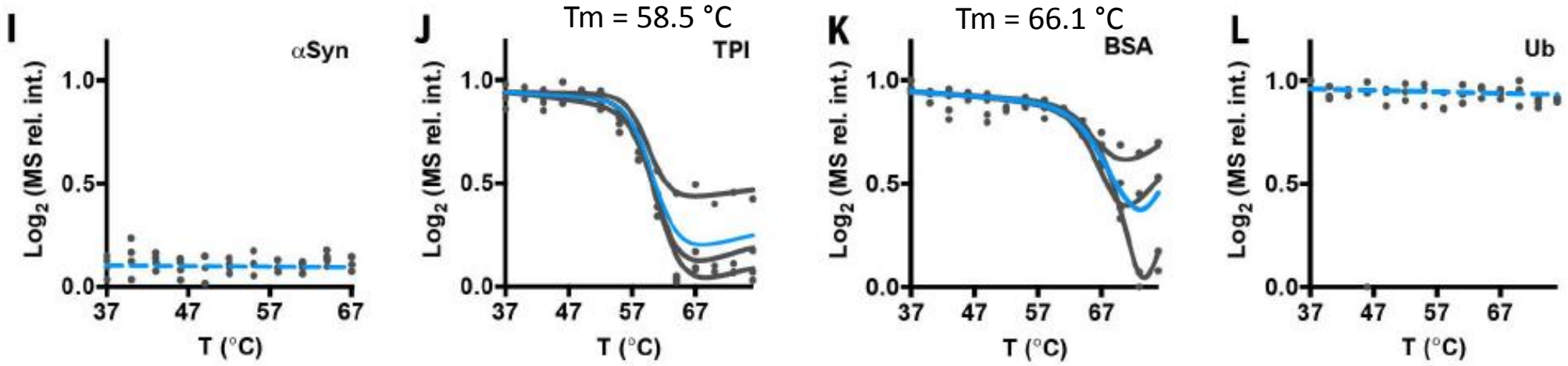
- 4 well-characterized single domain proteins
- data normalized
- Sigmoidal or linear model was chosen

Experimental validation of approach



LiP-MS analyses of **purified proteins, n=3**

- Grey = fit for each replicate (e.g. FT 1, FT2, FT3)
- blue = average of all measured peptides for one protein (FT 1-3 combined)



LiP-MS analyses of **purified proteins** spiked into an ***E. coli*** **background**

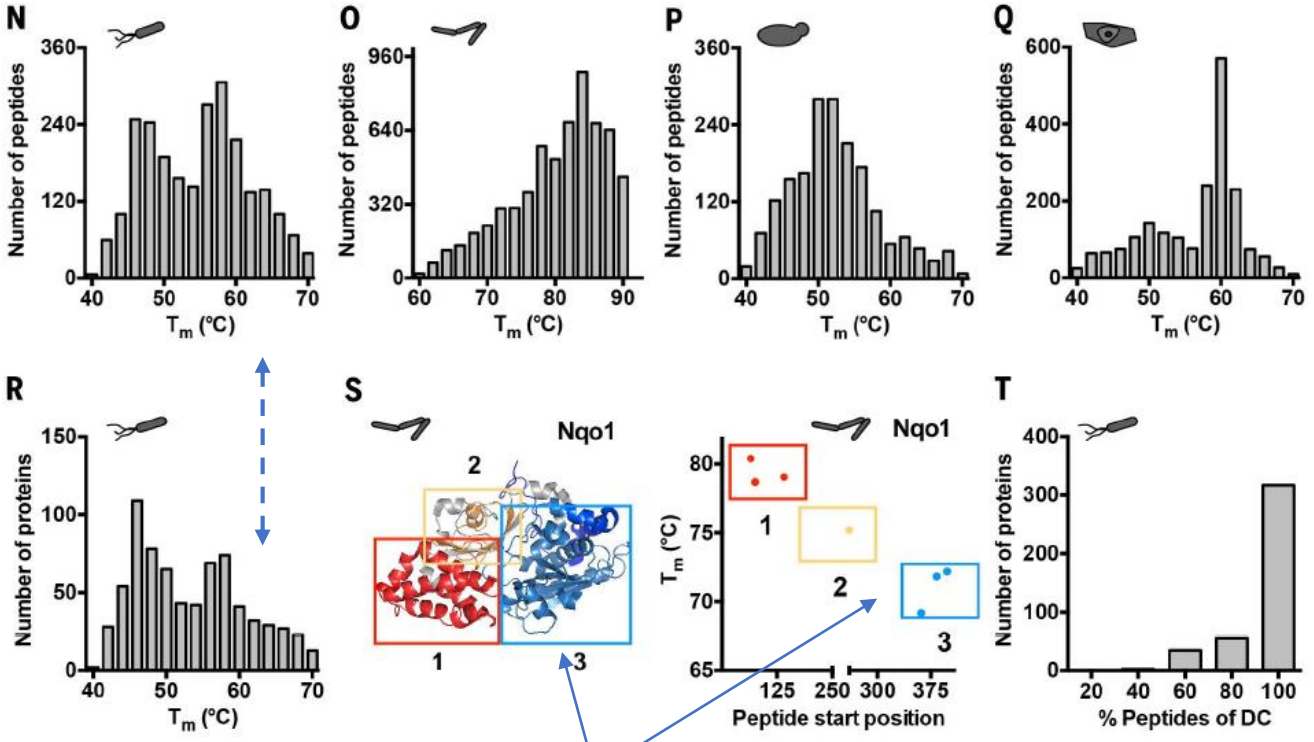
M

Organism	* Identified	** Quantified across T range	*** Unfolding behavior	High quality	
E. coli	2,153	1,663	1,211	730	OGT = 37°C
T. thermophilus	1,786	1,683	1,480	1,083	OGT = 60°C
S. cerevisiae	3,364	2,053	1,420	707	OGT = 30°C
Human	4,151	2,654	2,030	1,037	OGT = 37°C

*numbers of distinct proteins identified via MS (@37°C, trypsin only, n=3) => number decreases at higher temp.

**number of proteins consistently quantified at all 14 temperature points

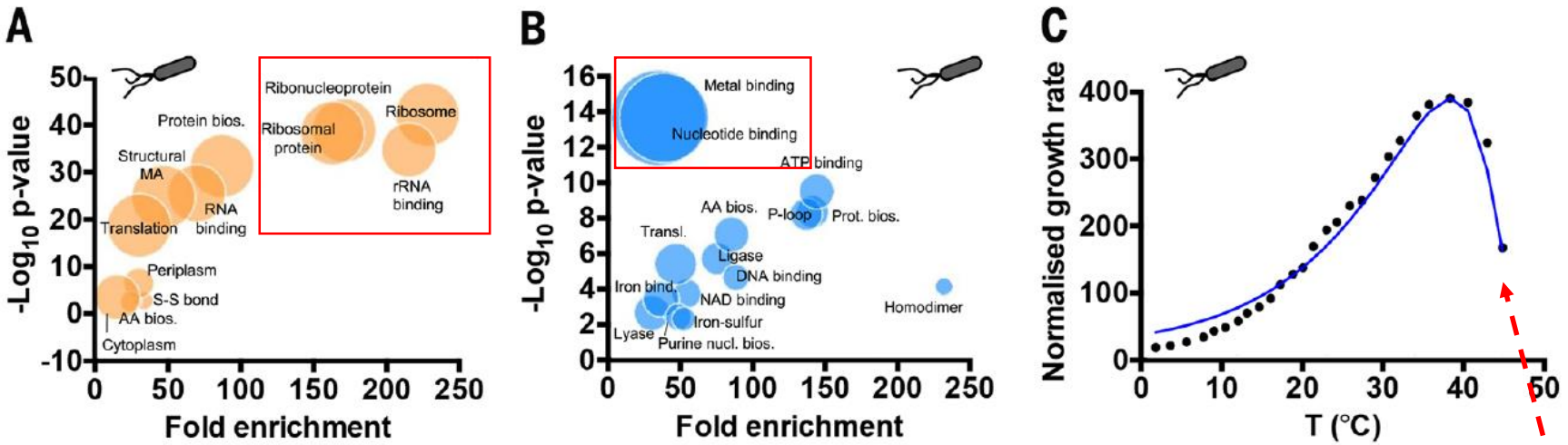
***proteins with sigmoidal melting curves



T: **Domain info** was available for 410 *E. coli* proteins (pfam.xfam.org) **Frequency of peptides** mapping to the same **domain** that were also associated within the same **Tm cluster** (i.e. DC, domain cluster)

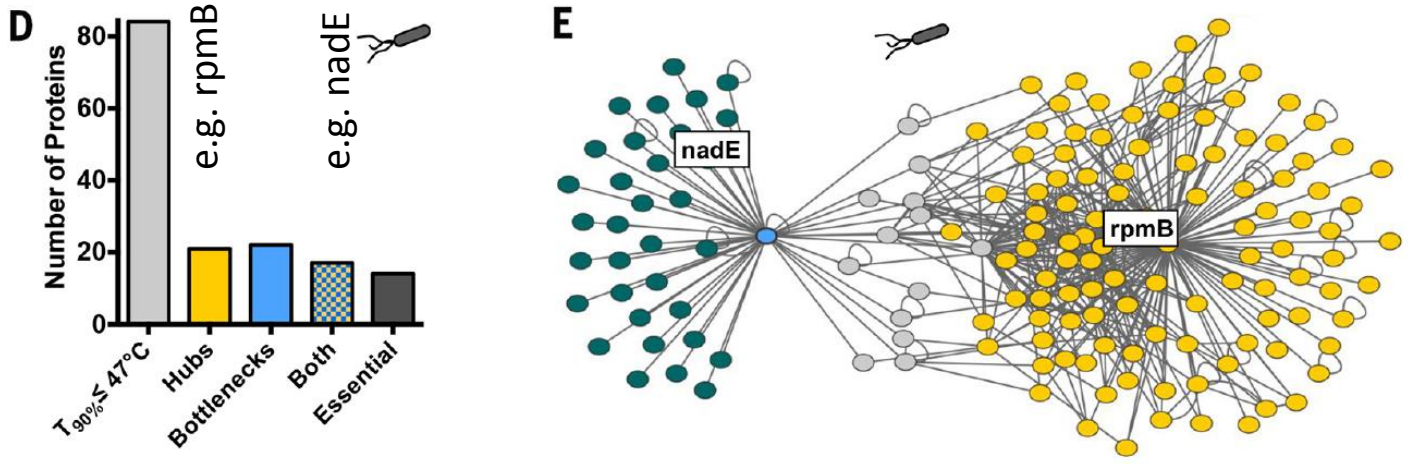
Tm clusters identify protein domains (can also be applied to proteins with unknown structures!)

Comparison of protein stabilities with structural/biological features in *E. coli*



Functional enrichment analysis for stable (A) and unstable (B) proteins in *E. coli*

- Clear separation of ontology terms, circle size reflects the number of proteins within group (tool DAVID, david.ncifcrf.gov)
- Unstable proteins were enriched for cofactor/DNA-binding proteins => high degree of conformational flexibility to enable interactions?



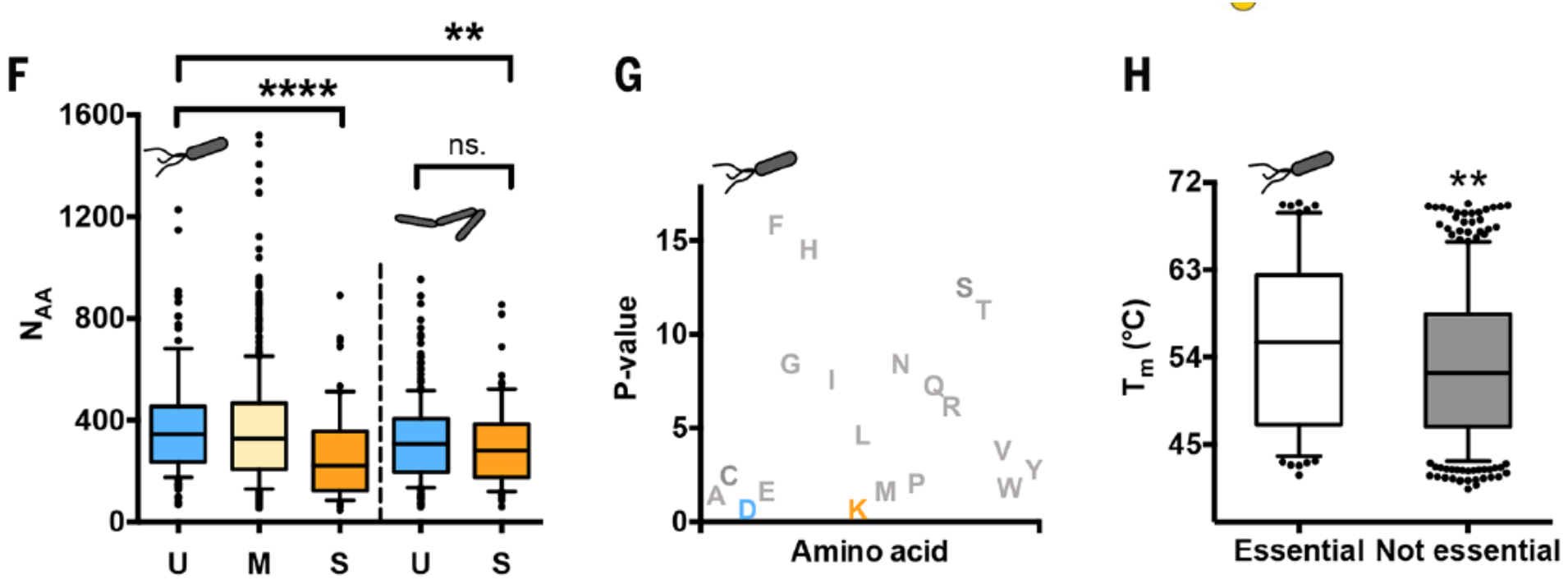
Transition midpoint @ 47°C
⇒ which proteins have their $T_{90\%}$ (i.e. 90% denatured) at this point?
⇒ determinants of temp.-induced cell death

Frequency of hubs, bottlenecks and essential genes in the set of *E. coli* proteins with T90% values <47°C

- Mainly proteins involved in protein, nucleic acid and fatty acid synthesis

Comparison of protein stabilities with structural/biological features in *E. coli*

Correlating stability with…



…Size

- Shorter => more stable
- U = unstable, 10 % of proteins with lowest Tm
- S = stable, 10 % with highest Tm

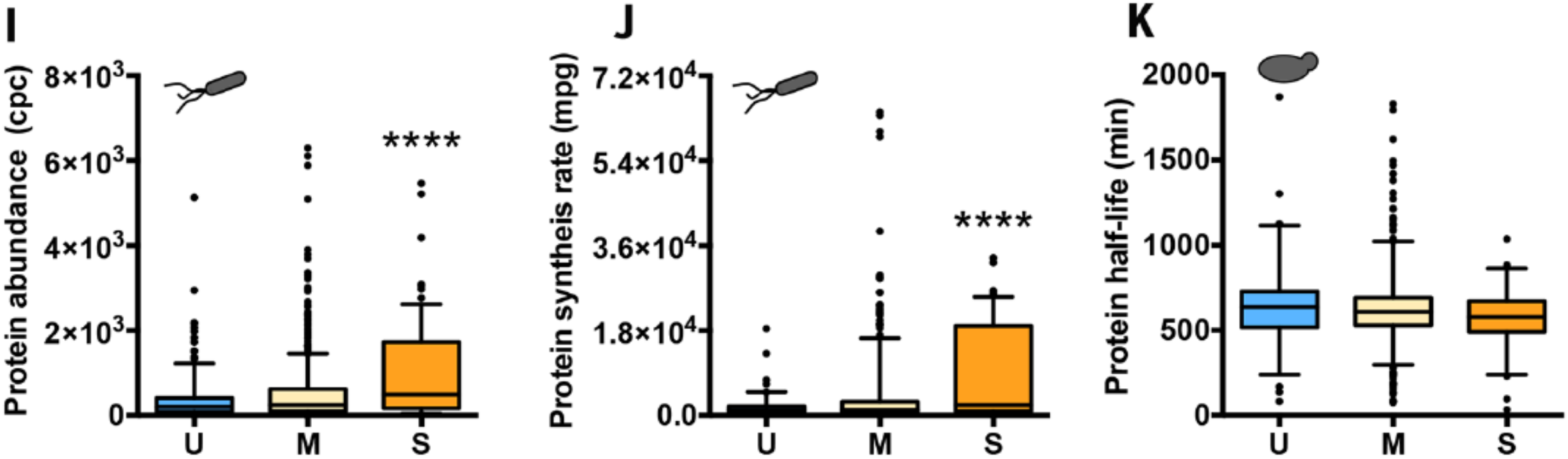
…amino acid composition

- Blue (aspartic acid) more abundant in U proteins
- Red (lysine) more abundant in S proteins
- More β-sheets in S proteins (not shown)

Wilcoxon test

Comparison of protein stabilities with structural/biological features in *E. coli*

Correlating stability with…



…abundance (copies per cell)

- More abundant => more stable

…synthesis rate

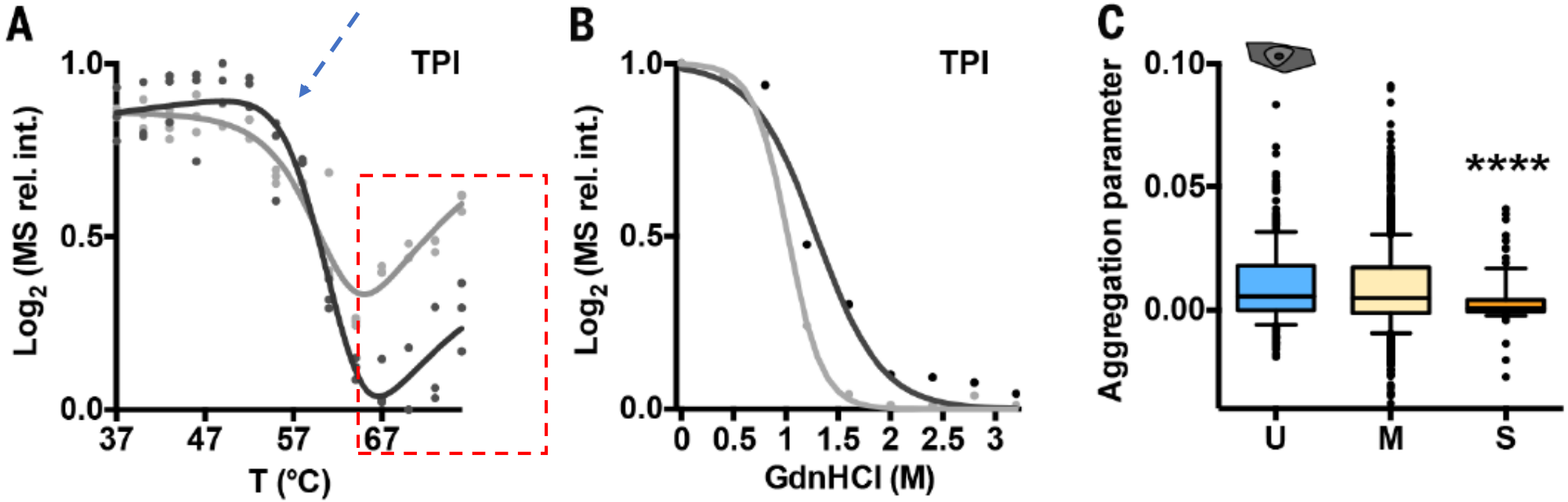
- Stable proteins have higher synthesis rate
- (ribosome profiling experiments)

…turnover rate

- Not available for *E. coli*
- Determinants of thermostability and proteolysis are not equal

Thermally stable proteins show a slightly reduced aggregation propensity

Temp. increases
⇒ unstructured domains
⇒ PK cleavage: amount of FT peptide reduced (here 2 different ones)

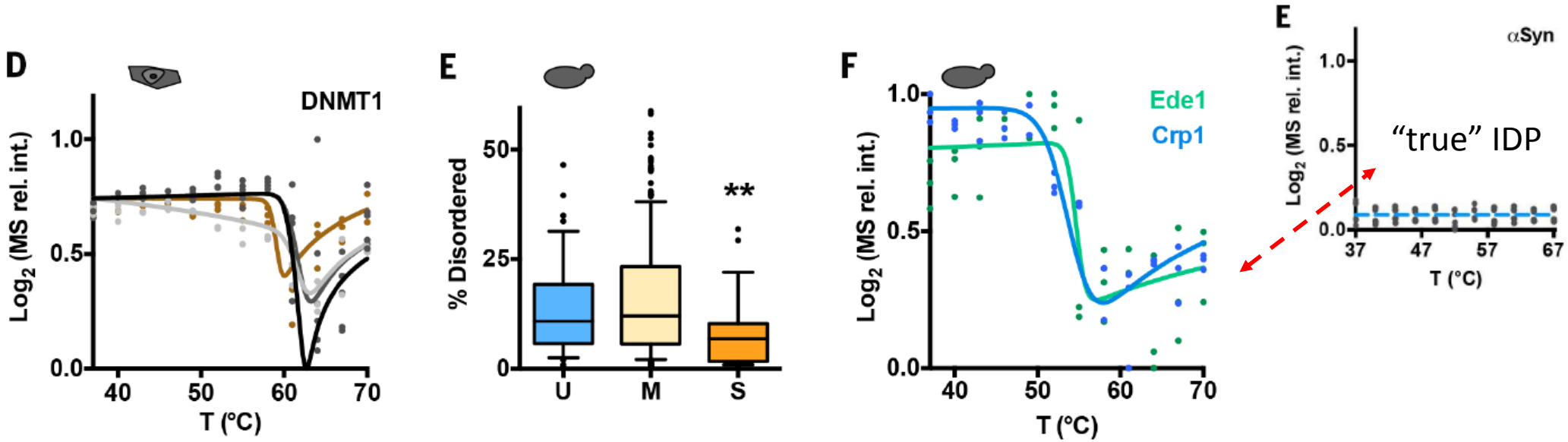


Slope => aggregation parameter

TPI spiked into *E. coli lysate* (A) together with protein denaturant GdnHCL (B)

- Signal intensities of FT peptides increase at high temperatures (here: 67°C) due to aggregation that protects unstructured domains from PK cleavage
- Propensity for aggregation: *T. thermophilus* < *E. coli* proteins (not shown)

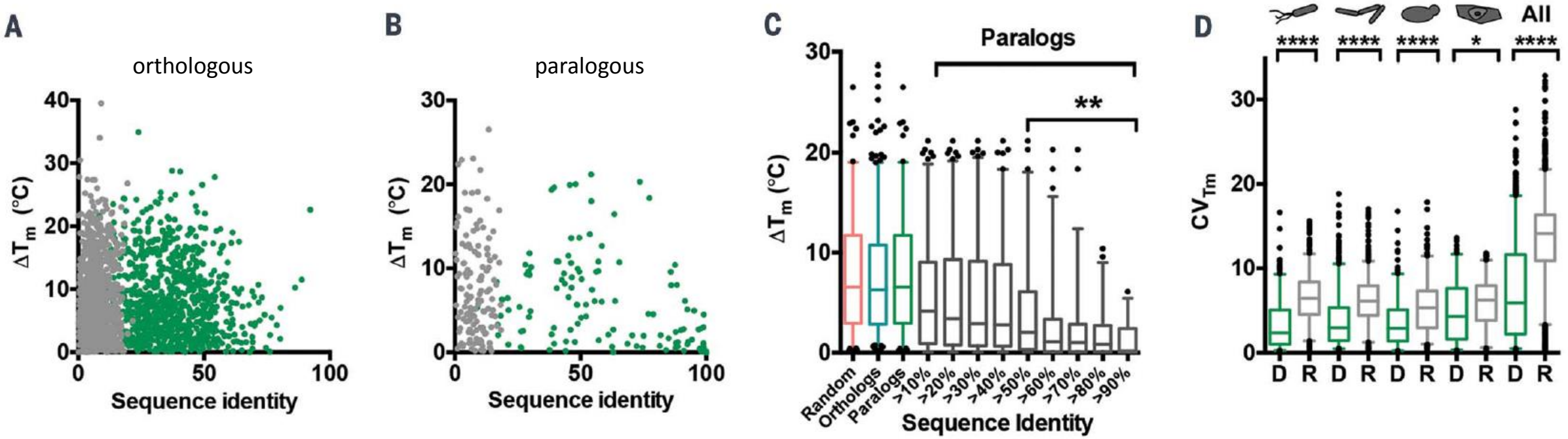
Thermally stable proteins aggregate less and are intrinsically less disordered



- Thermal profiles of 4 peptides shown from aggregation-prone human DNMT1. (D)
- Percent structural disorder for the different stability classes .(E)
- Thermal profiles of previously predicted* **IDPs** Ede1 and Crp1 revealed that they are no IDPs since they exhibit a **sigmoidal unfolding behavior** (i.e. they have a complex structure). (F)
- Actually, half of all previously predicted IDPs (*Gsponer et al., 2008, Science) had a structure *in vivo*

  => stabilization by chaperones, binding partners

Evolutionary conservation of protein and domain stability



- **A-C** conservation: Ortho-/paralogous protein* pairs (green) from all proteomes, randomly assigned pairs (gray, 1 dot = 1 pair).
  => Conservationg of Tms among the paralogous proteins (due to younger evolutionary age?)
- **D** Coefficient of varation from all peptides associated to homologous domain pairs (D) or random pairs (R).
  =>homologous domains are equally stable.

*eggNOG data base

Summary & Conclusion

Largest data-set **proteome-wide thermal denaturation profiles** obtained for *E. coli, S. cerevisiae, T. thermophilus* and human cells taking into account the vast influence of the cellular matrix.

Thermostable proteins…

…exhibit enriched Lysine residues and β-sheet structures,
…are shorter,
…are less prone to thermal aggregation than unstable proteins.

Domain thermostability was conserved within species and across organisms.

- However, thermal stability was not similar for orthologous proteins suggesting that Tms of proteins are determined by **reshuffling** of protein **domains**.

Proteome instability does not arise from the simultaneous and **generalized loss** of proteins.

- Only **a small subset of essential proteins** undergo denaturation leading to cell death.

Conlusion

«Translational robustness” theory: **Highly expressed proteins must tolerate translational errors** that would otherwise lead to the toxic accumulation of misfolded proteins.

> Missense errors in translation occur at rates of one per 10^3–10^4 codons (Kramer and Farabaugh, 2007; Ogle and Ramakrishnan, 2005; Parker, 1989); at an error rate of 5×10^{-4}, 18% of proteins expressed from an average length (~400-codon) gene contain at least one missense substitution. Roughly ~10–50% of
>
> Drummond et. al, Cell, 2009

⇒ Increasing the **thermodynamic stabilites** of the folds of **abundant proteins** will broaden the range of amino acid replacements that a protein can tolerate without triggering misfolding.

⇒ Data showed a direct relationship between **protein thermal stability** and **intracellular abundance** and an inverse correlation between **protein stability** and **aggregation/local misfolding**.

Intrinsically disordered proteins (IDPs) have been predicted to make up 30% of the proteome.

⇒ This study reveals that half of those undergo a two-state denaturation profiles, meaning they are globally or locally structured within the cells.

Outlook LiP-MS

Powerful tool for structural biology and synthetic biology.

Adds another layer to proteomic analyses (proteins are not just regulated via their expression levels).

Structure-based drug discovery (e.g. conformation-sensitive drugs).

The LiP-MS analysis yields conformotypic peptides that can be employed as «conformational biomarkers»

- E.g. the ratio of monomeric and fibrillar α-Syn in Parkinson's disease

Technology licensed to Biognosys (Schlieren).

BIOGNOSYS
NEXT GENERATION PROTEOMICS