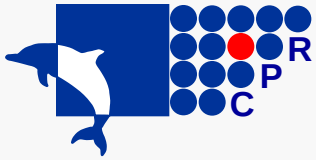
Biological insights from large-scale protein copy number measurements



David O'Connor

Centre for Proteomic Research

University of Southampton, UK

Plan of talk

- The case for absolute quantification
- Which absolute quantification method?

  Use of a data-independent acquisition approach
- What can you do with such data?

  Case history – *Chlamydia trachomatis*

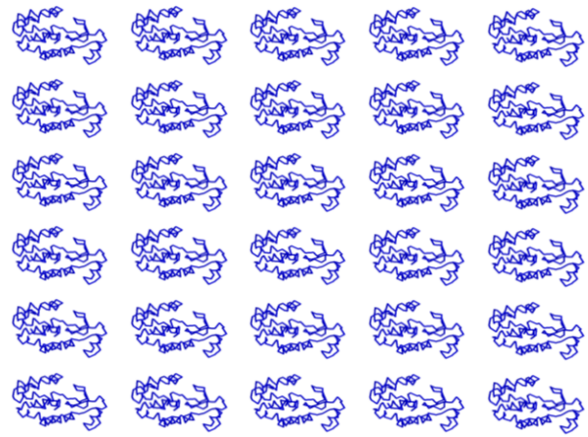
Better Quantification

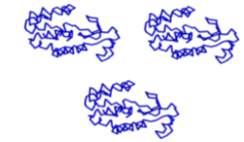
Absolute versus relative quantification

Advantages of Absolute Quantification

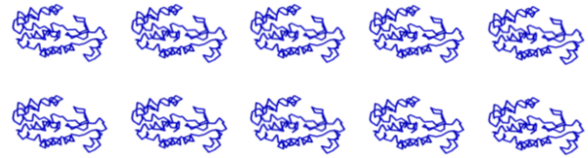
Measuring numbers of molecules/cell gives more information

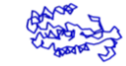
[Protein]
State 1





[Protein]
State 2



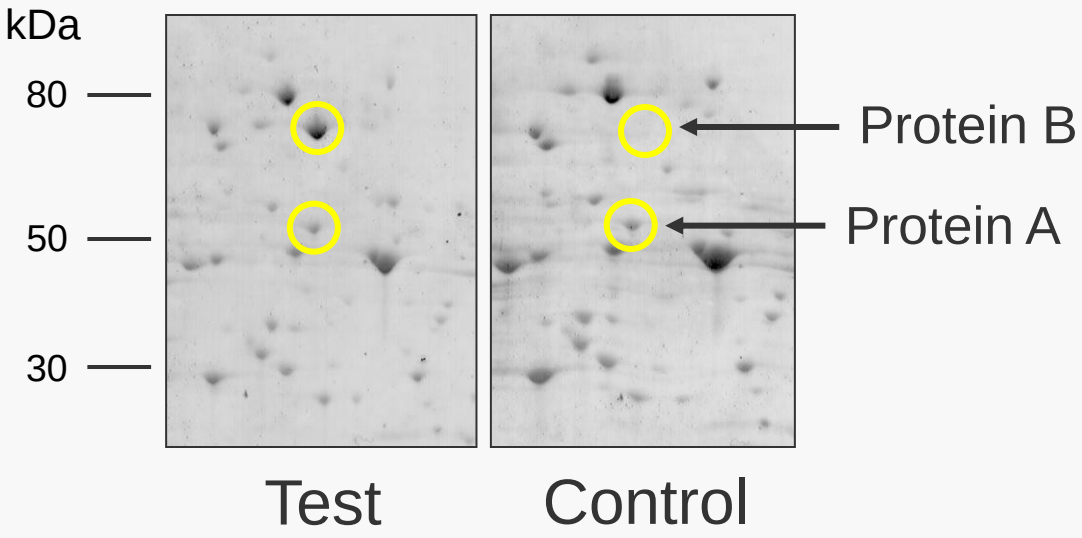


Significant

Noise?

Advantages of Absolute Quantification

Data loss associated with relative quantification (2-D gels, SILAC, iTRAQ etc.)



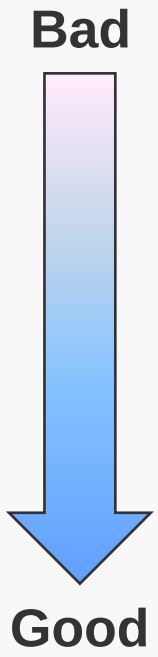
	Spot Volume		Test / Control
	Test	Control	
Protein A	100.1	125.2	0.8
Protein B	2160.3	n.d.	?

Advantages of Absolute Quantification

Ranking proteins in terms of molecules/cell can be useful...

RANK	PROTEIN	MOLECULES/CELL
1	Protein A	1×10^7
2	Protein B	1×10^6
3	Protein C	1×10^5
4	Protein D	1×10^4
5	Protein E	1×10^3
. . .	. . .	. . .
25	Protein Y	<10
26	Protein Z	<10

POTENTIAL AS DRUG TARGET?



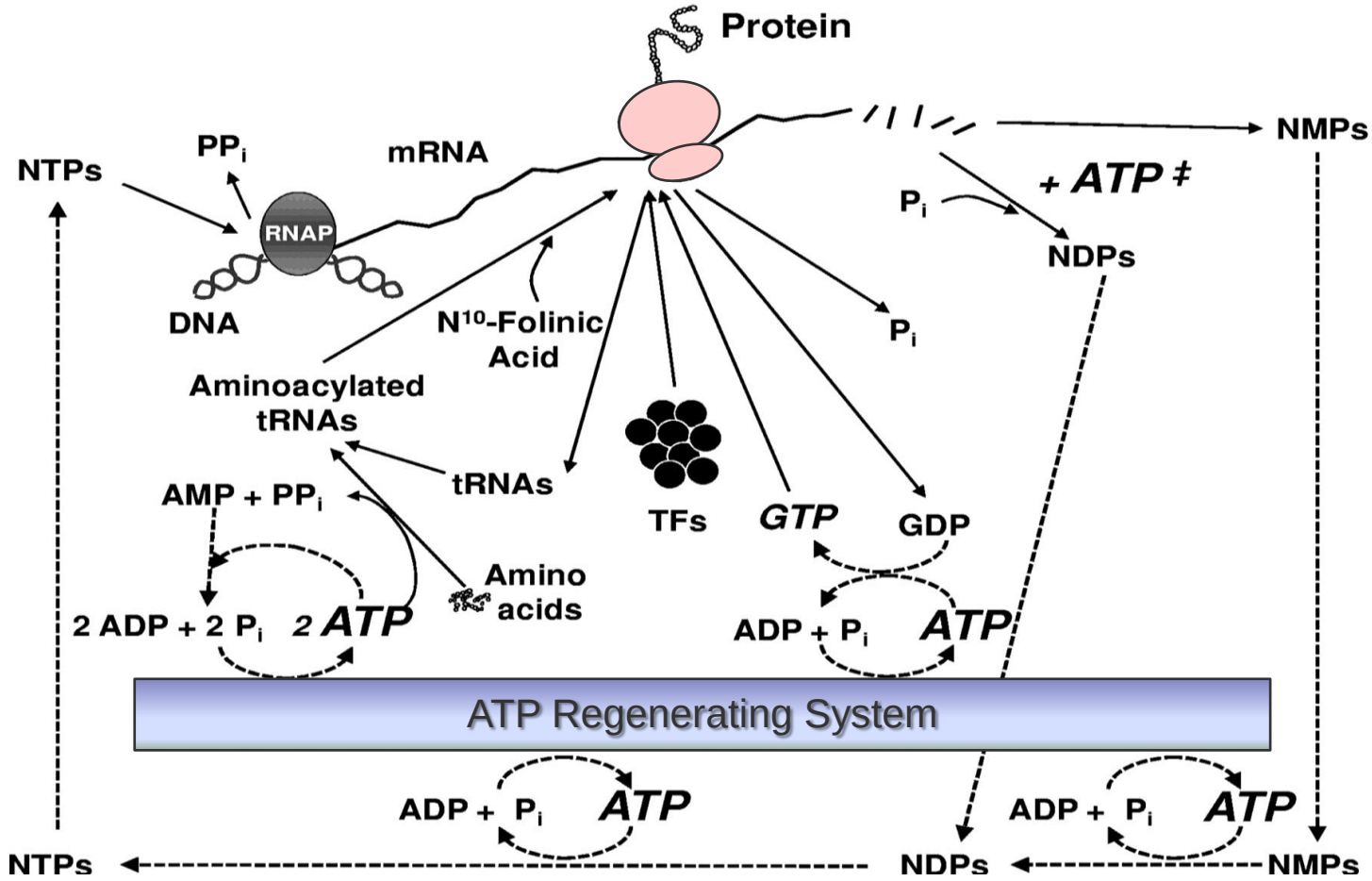
Advantages of Absolute Quantification

Identifying and ranking factors that determine protein abundance...

E.g. codon usage, length, hydrophilicity, pI, [mRNA], location of gene in genome etc.

Advantages of Absolute Quantification

Finding out where a cell is investing its energy



E.g. protein synthesis consumes ca. two-thirds of the total energy produced by a rapidly growing *Escherichia coli* cell

Jewett, M.C. *et al*. (2009) *J. Bacteriol*. 191: 1083-1091

Plan of talk

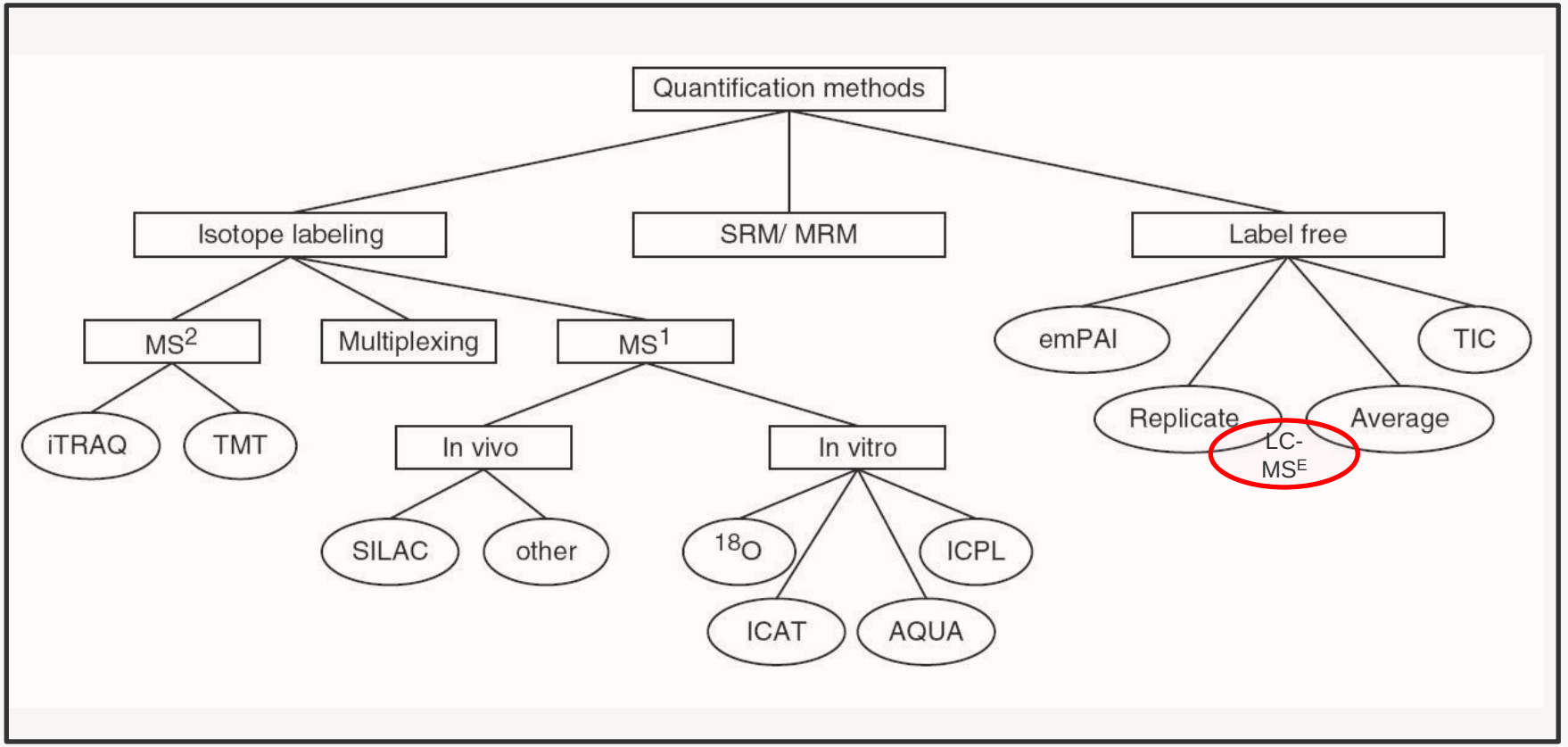
- The case for absolute quantification
- Which absolute quantification method?

  Use of a data-independent acquisition approach
- What can you do with such data?

  Case history – *Chlamydia trachomatis*

Which absolute quantification method?

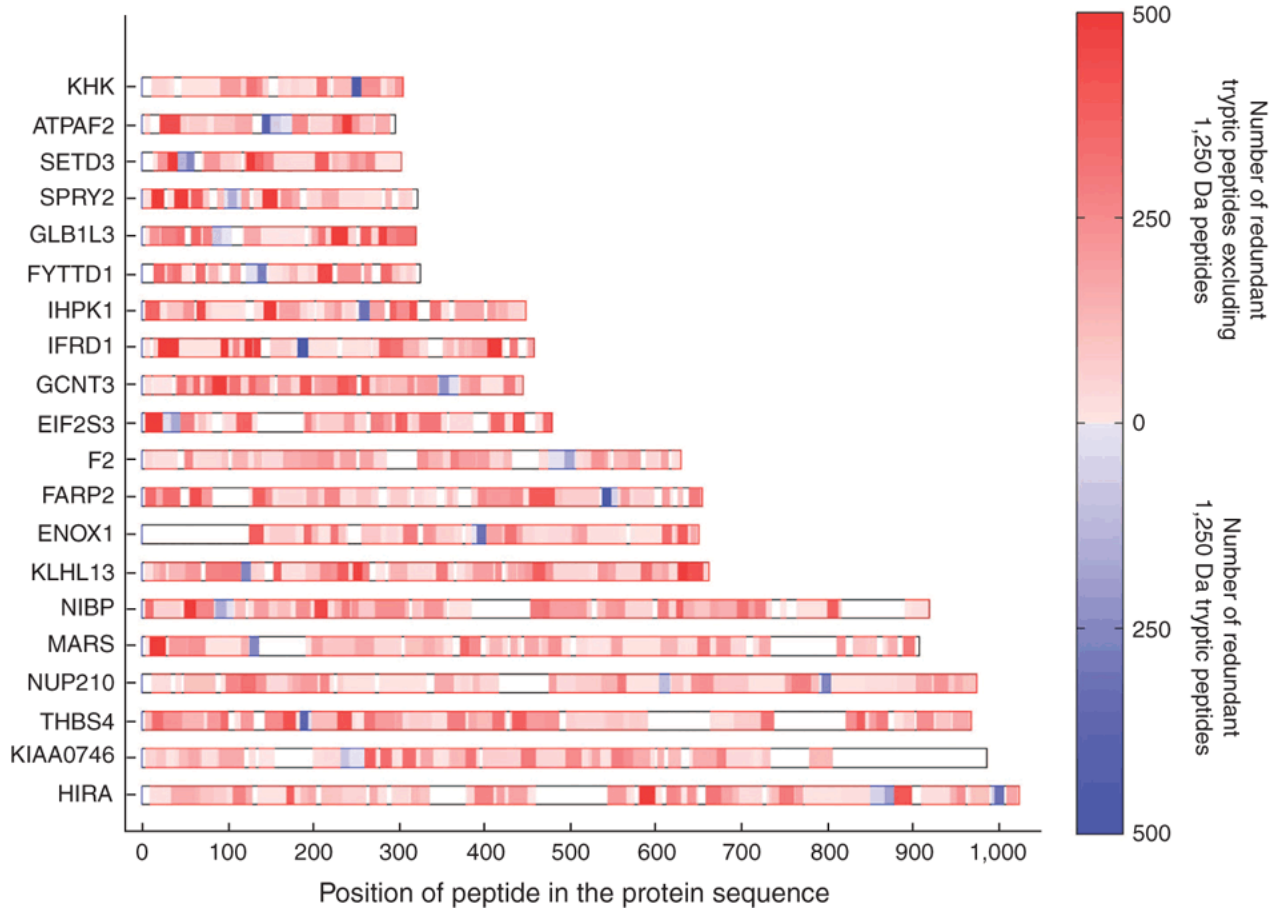
Use of a data-independent acquisition (DIA) strategy



Review: Vaudel, M. *et al*. (2010) Protein and peptide quantification: a map of the minefield *Proteomics* 10: 650-670.

Which absolute quantification method?

Limitations of a data-dependent acquisition (DDA) strategy

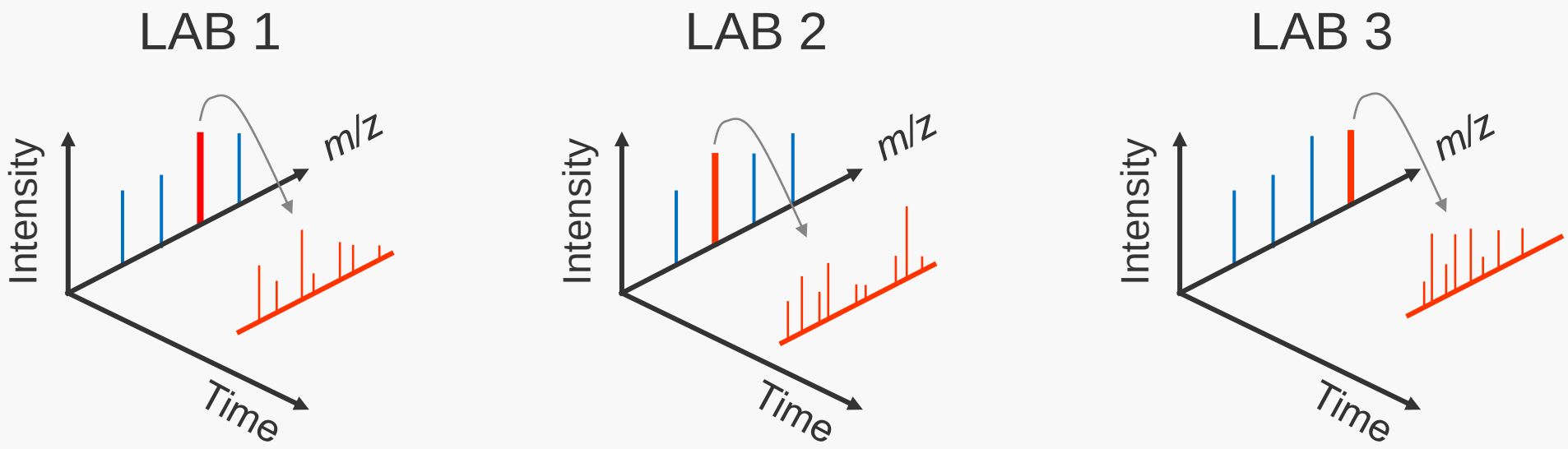


- Only 7 out of 27 labs identified all 20 proteins correctly
- Only <u>one</u> lab saw all proteotypic peptides – why?

Bell, A.W. *et al*. (2009) *Nature Methods*. 6: 423-430.

Which absolute quantification method?

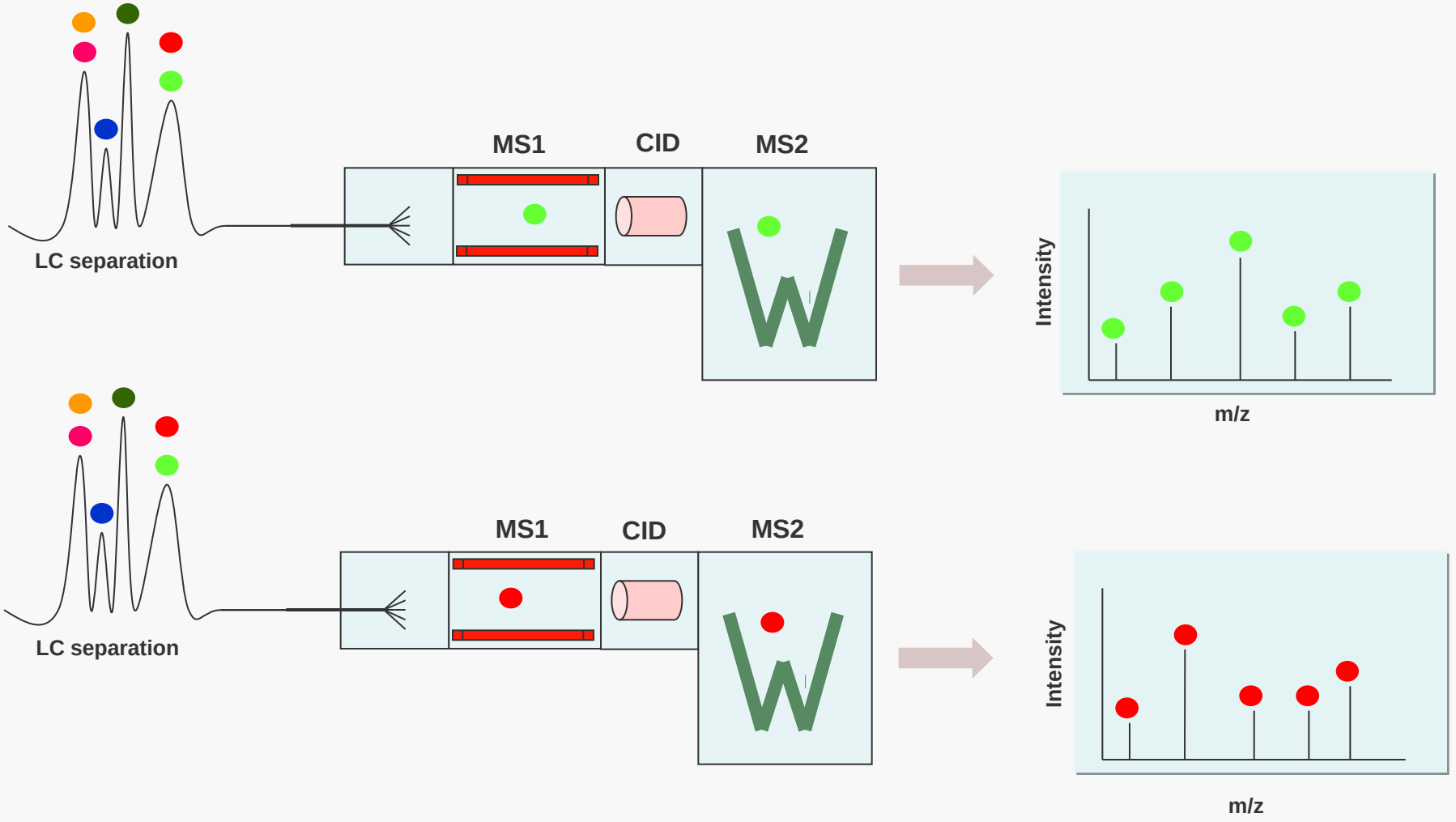
Limitations of a data-dependent acquisition (DDA) strategy



- Serial selection of precursor ions biases analysis to high abundance components
- Precursor ion scans are stochastic - different ions may be selected for fragmentation in different runs → irreproducibility
- Selection windows of 2-4 Da means additional precursor may be selected for fragmentation along with target ion → lower signal:noise

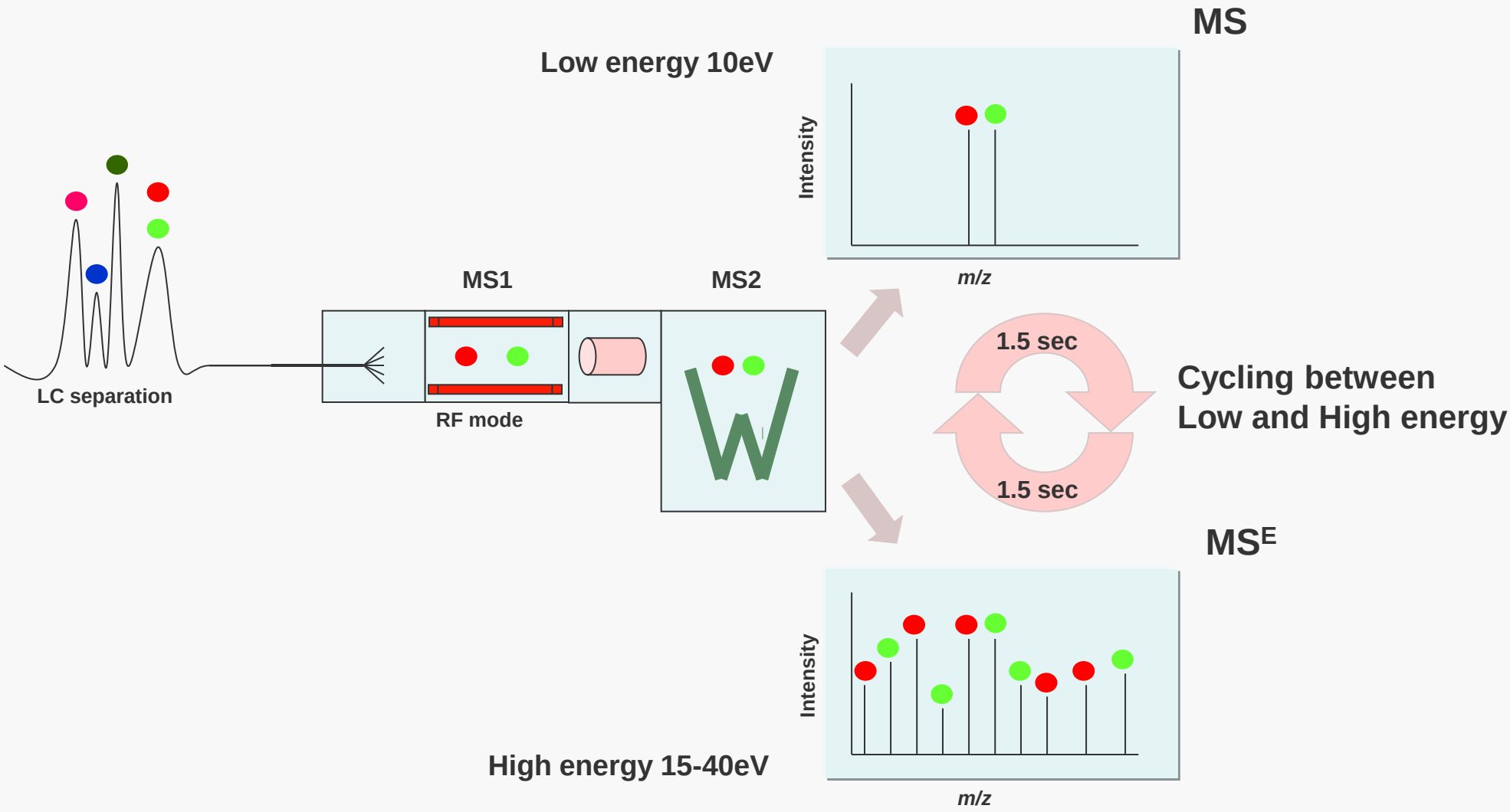
‘Traditional’ LC - Tandem Mass Spectrometry

One slice at a time



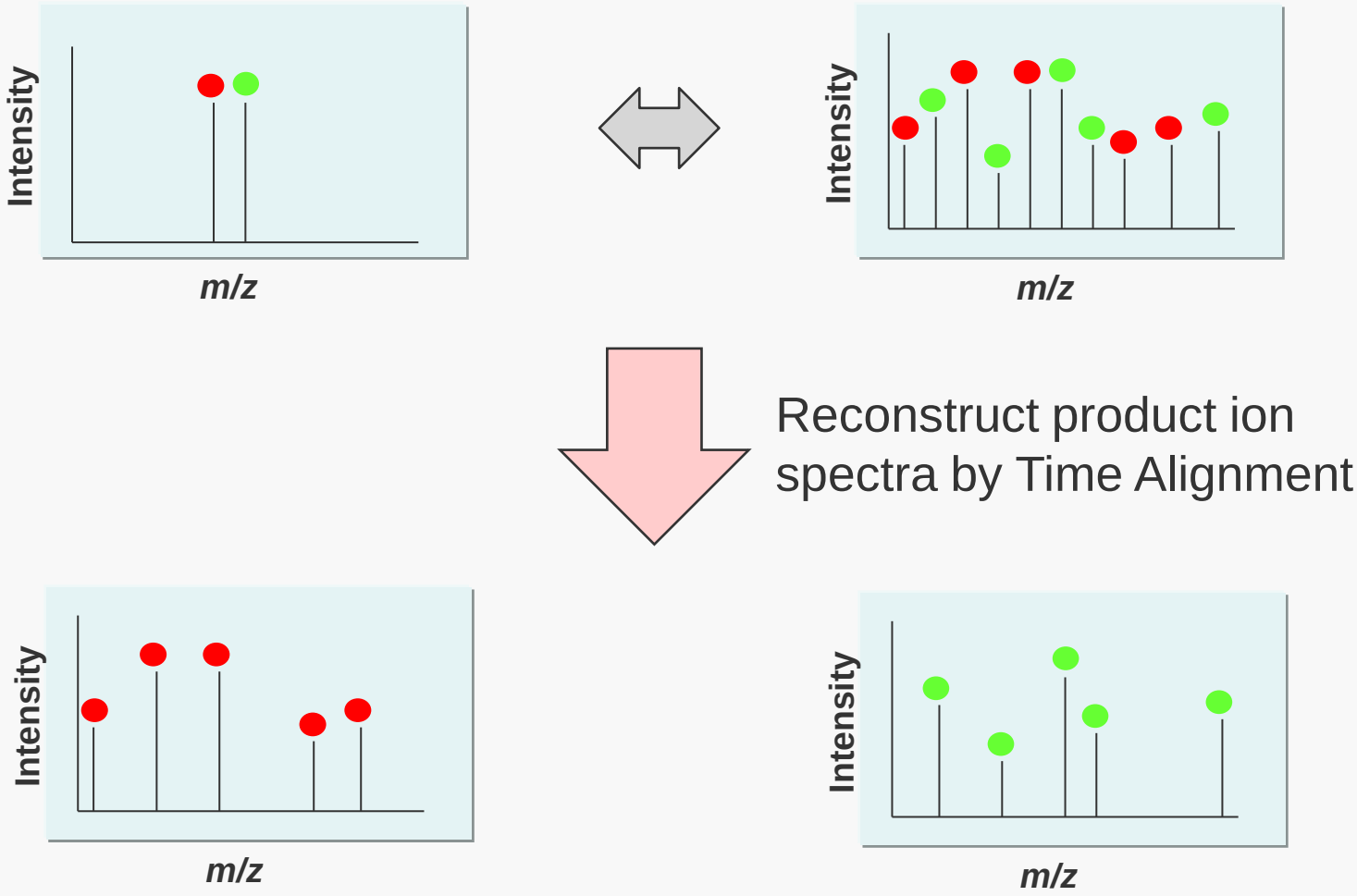
Label-free proteomics

Principle of LC-MS^E



Label-free proteomics

Principle of LC-MS^E



Plan of talk

- The case for absolute quantification
- Which absolute quantification method?

  Use of a data-independent acquisition approach
- What can you do with such data?

  Case history – *Chlamydia trachomatis*

Chlamydia trachomatis

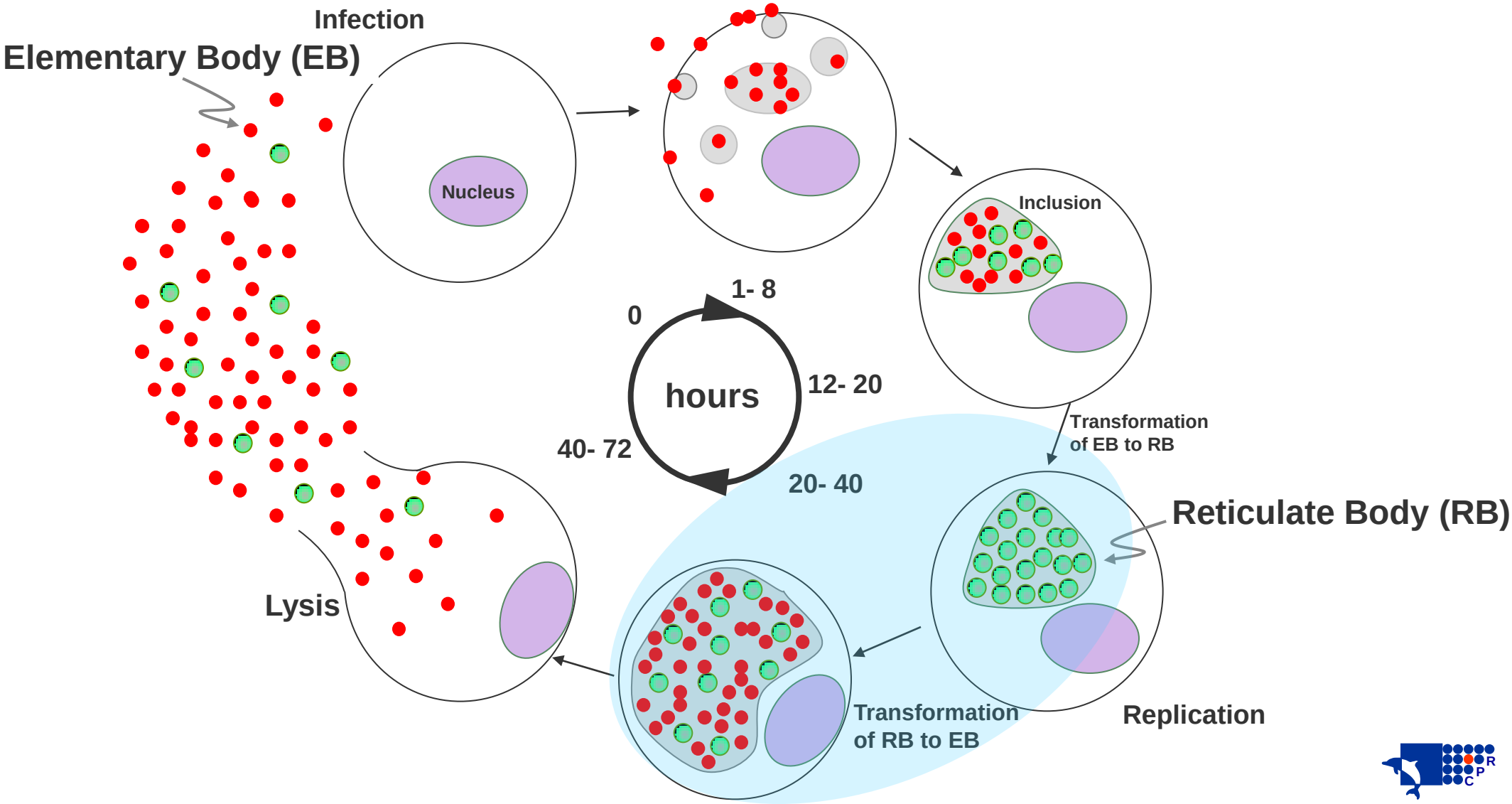
A widespread and important pathogen

http://www.nature.com/eye/journal/v19/n10/fig_tab/6701963f5.html

- Causes trachoma - the leading cause of preventable blindness
- ~84 million people have active infection
- Also major cause of genital tract infections – leads to pelvic inflammatory disease and tubal factor infertility

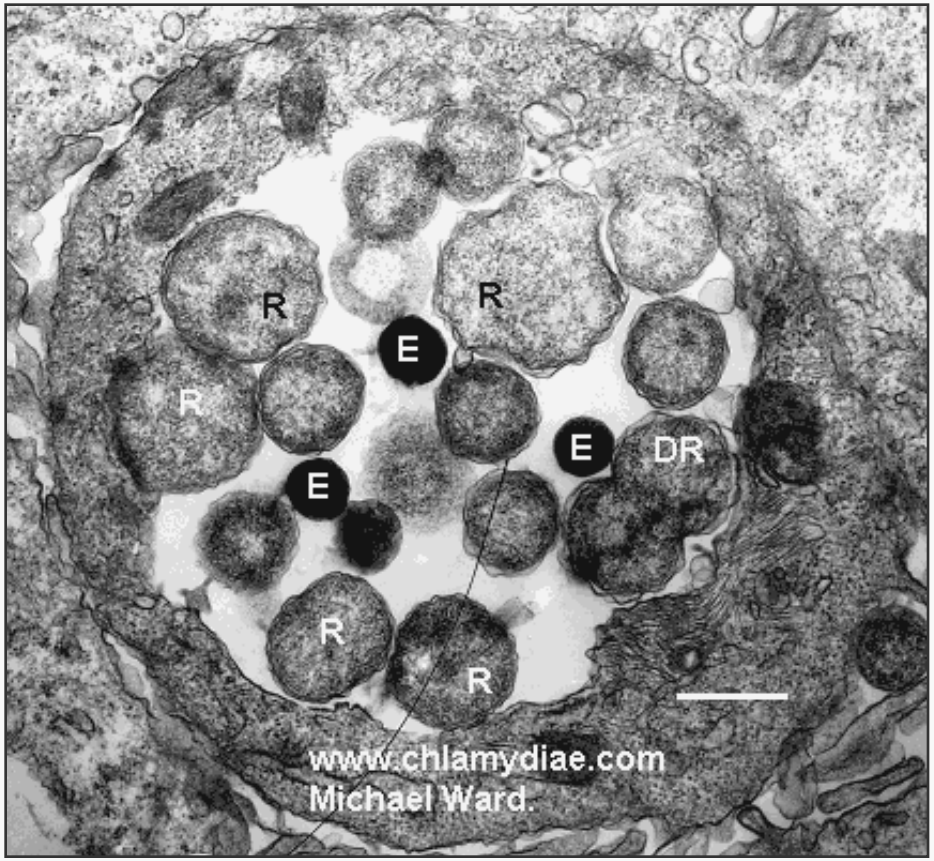
Life cycle of *Chlamydia trachomatis*

Elementary Bodies ↔ Reticulate Bodies



Chlamydia trachomatis

Elementary Bodies and Reticulate Bodies



EBs

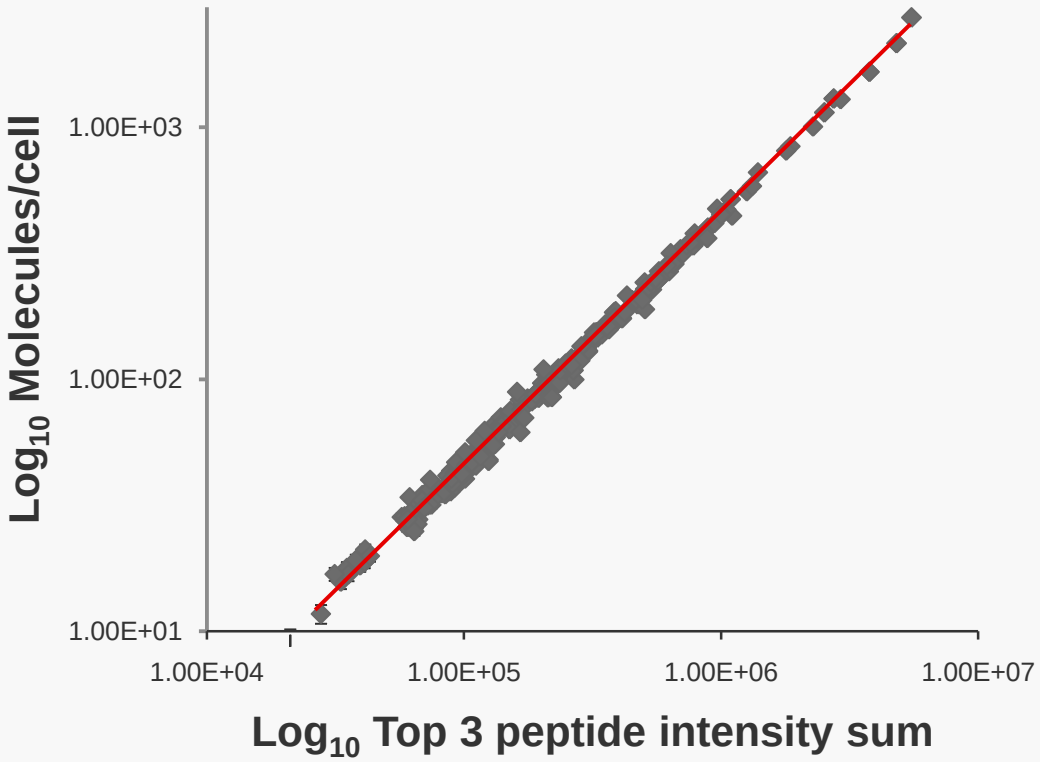
- Extracellular, infectious form
- Metabolically quiescent

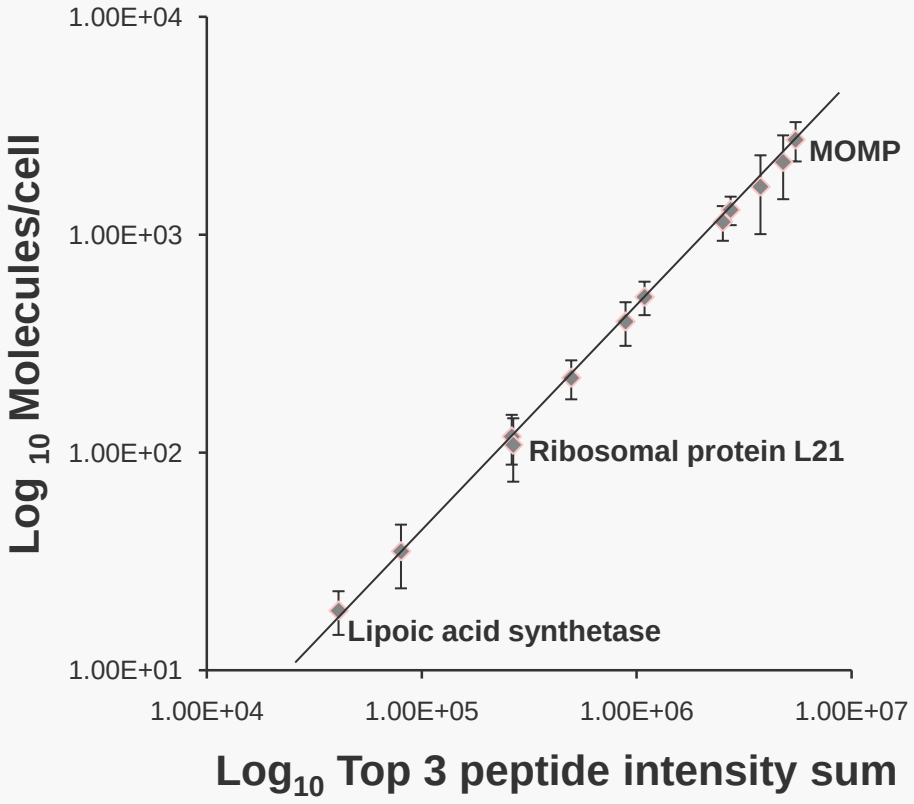
RBs

- Intracellular, non-infectious
- Active, replicating stage

Label-free proteomics

Dynamic range and reproducibility





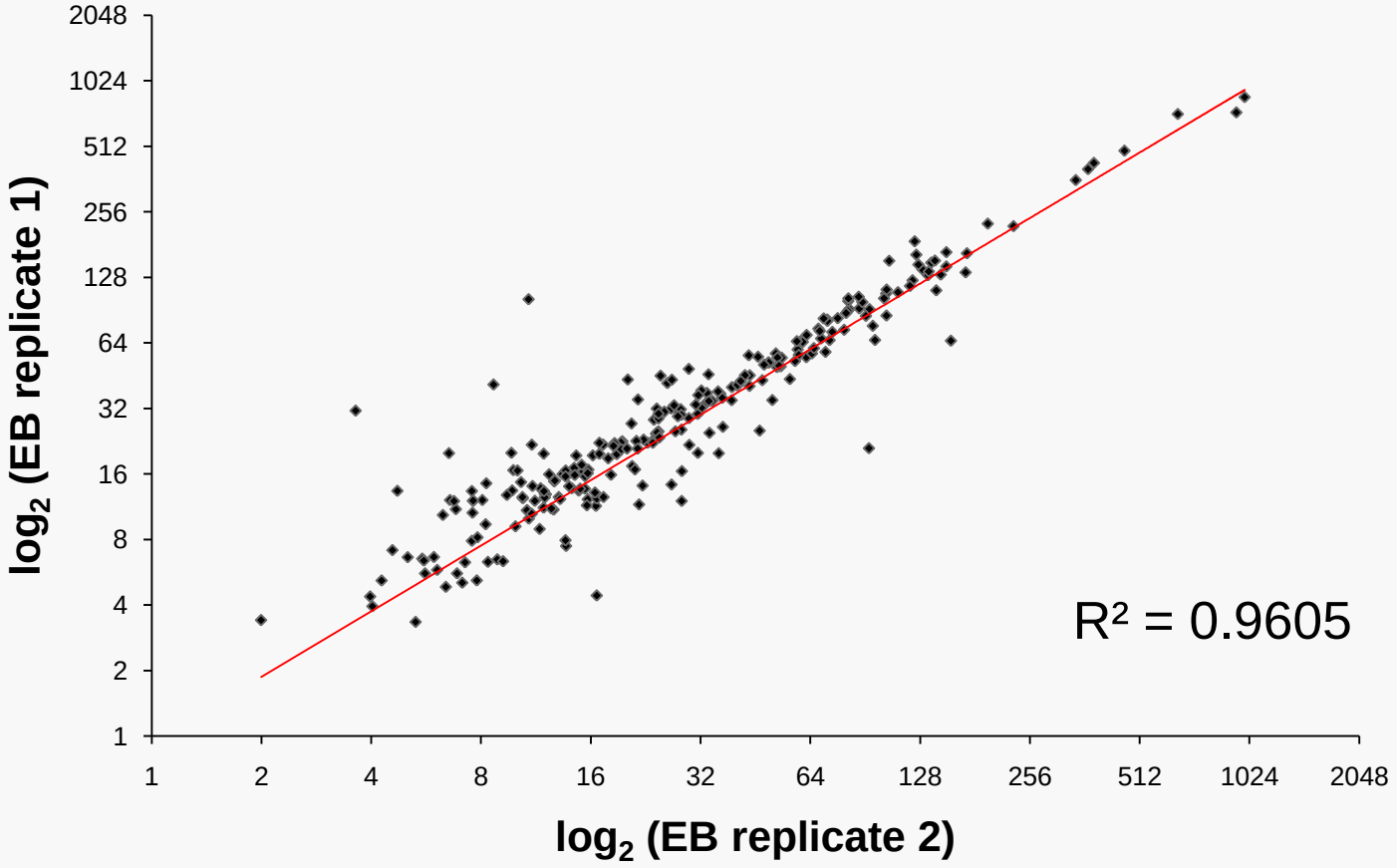
R^2 = 0.9967

Technical replicates: ~12% CV

Biological replicates: ~ 16% CV

Label-free proteomics

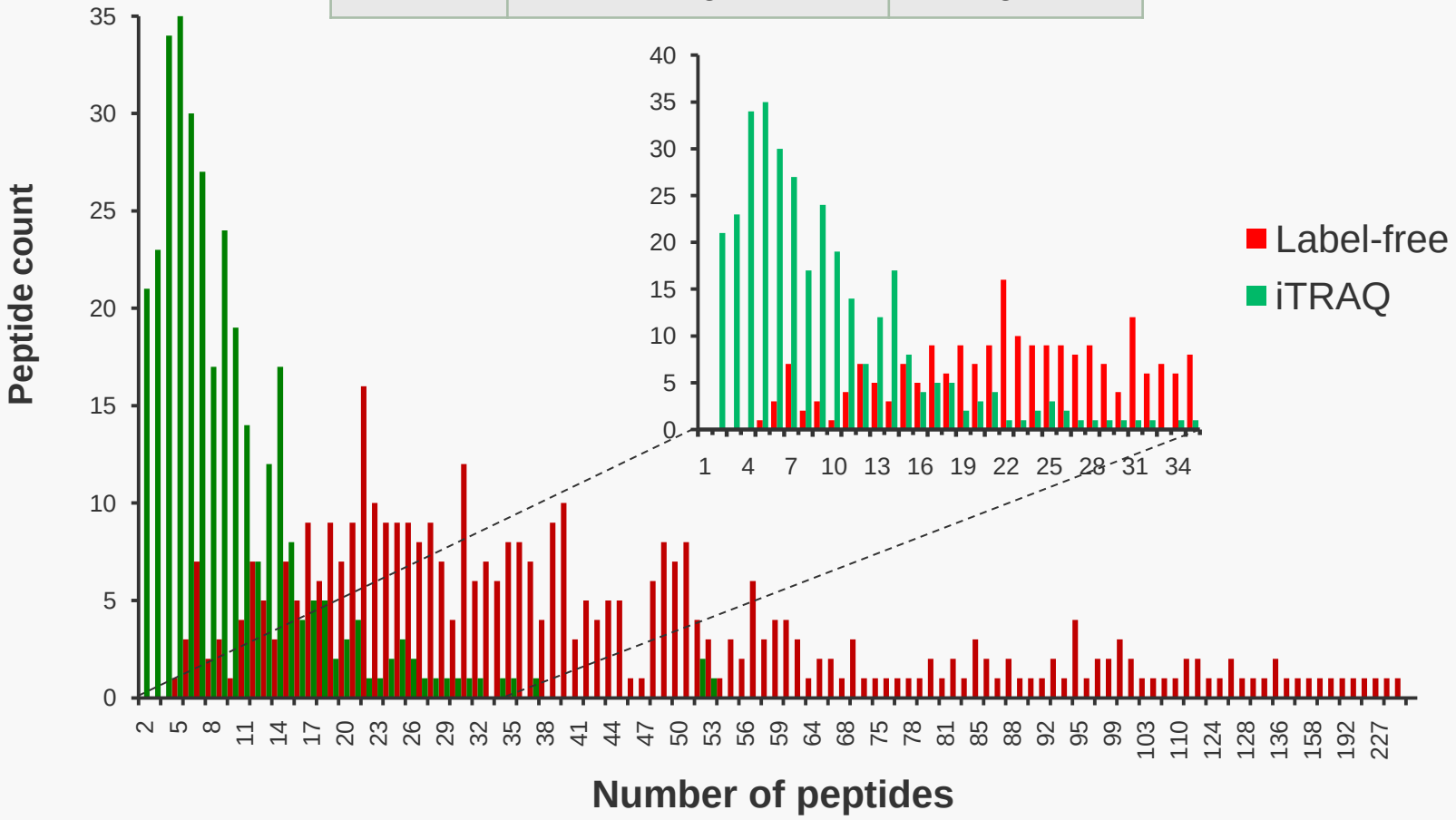
Dynamic range and reproducibility



Label-free proteomics

Peptides used to assign proteins – LC-MS^E vs. iTRAQ

	Average number of peptides per protein	Sequence coverage (%)
iTRAQ	10	26
LC-MS^E	46	64



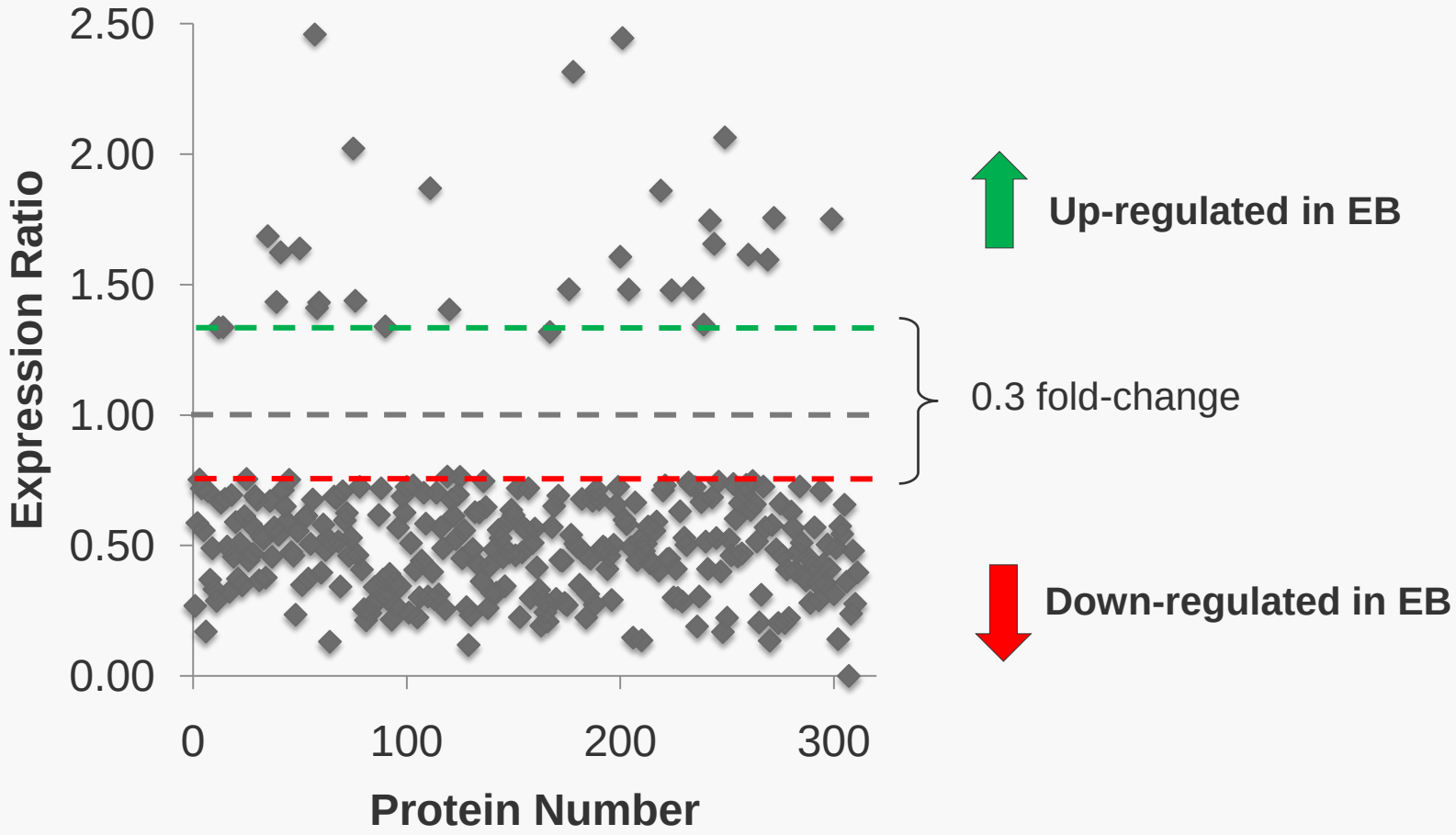
Label-free proteomics

Top ten most abundant proteins in EBs

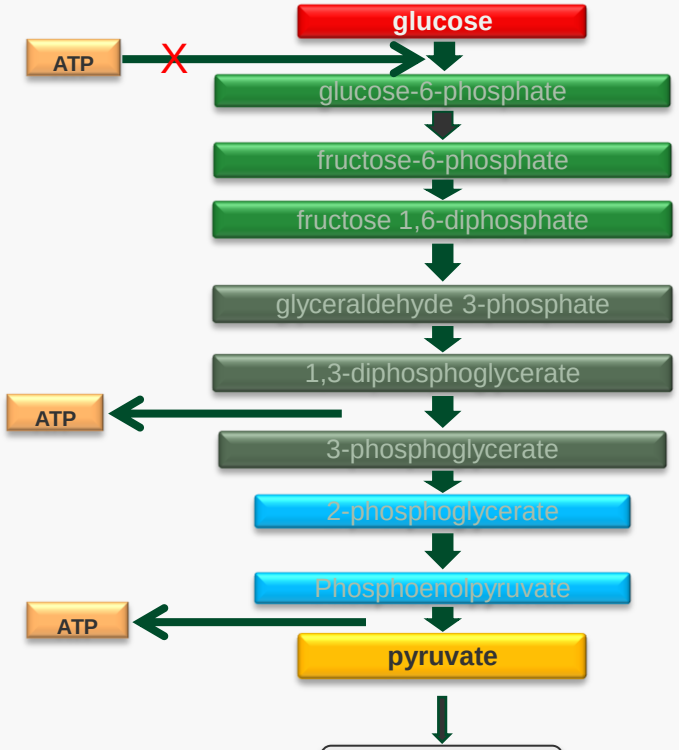
Locus	Gene name	Protein description	EB (molecules/cell)
CTL0050	*ompA*	major outer membrane protein	272,790
CTL0574	*tufA*	translation elongation factor Tu	215,611
CTL0652	*dnaK*	chaperone protein	166,008
CTL0365	*hsp60_1*	chaperonin GroEL	130,043
CTL0803	*mip*	peptidyl-prolyl cis-trans isomerase	129,190
CTL0847		conserved hypothetical protein	114,533
CTL0568	*rplL*	LSU ribosomal protein L12P (L7/L12)	100,628
CTL0887		putative exported protein	84,041
CTL0874		conserved hypothetical protein	80,739
CTL0488	*acpP*	acyl carrier protein	66,243

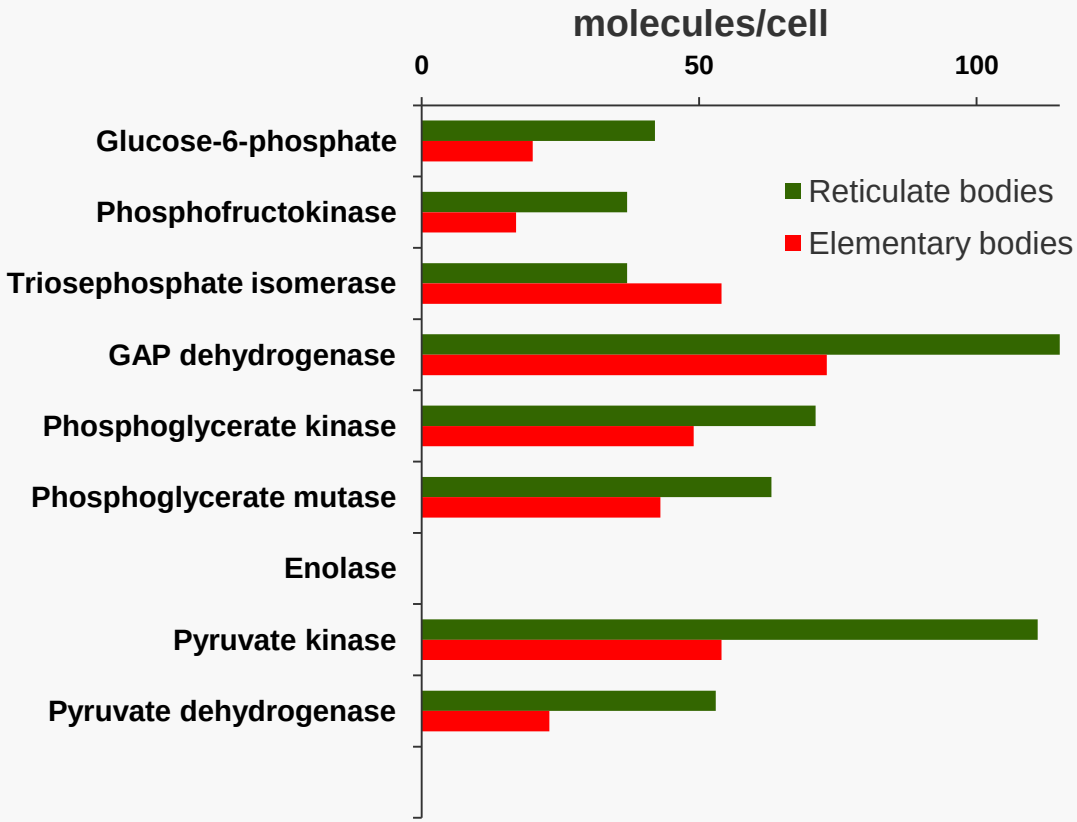
Label-free proteomics

Proteins that are differentially expressed between EBs and RBs

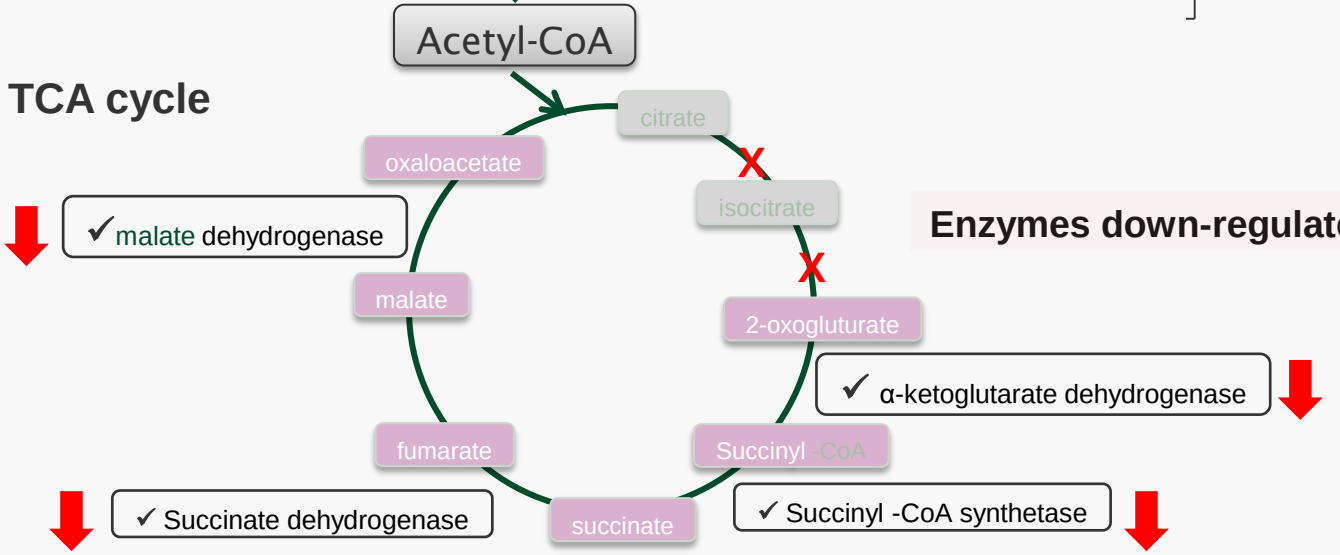


Glycolysis





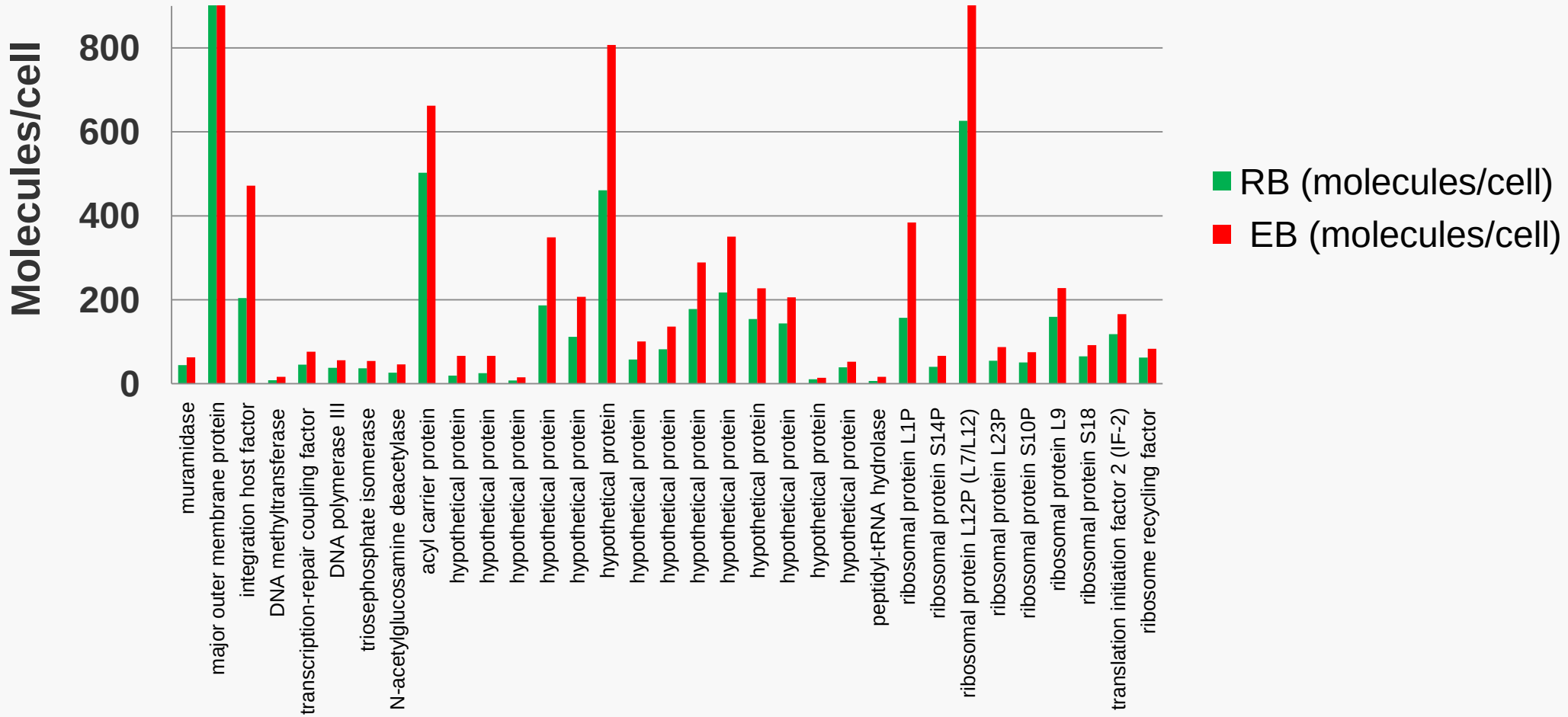
TCA cycle



Enzymes down-regulated in EBs

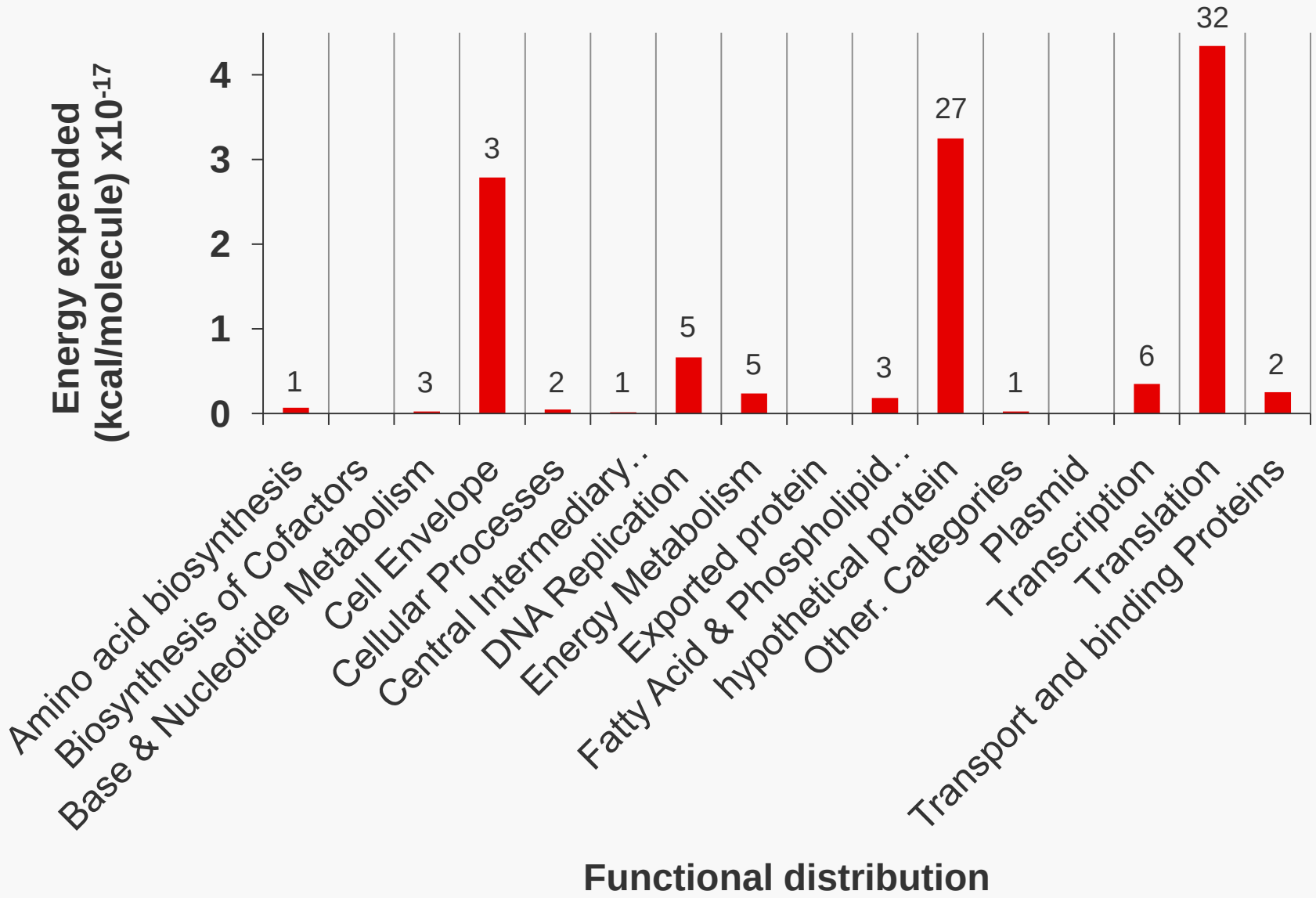
Label-free proteomics

Proteins that are differentially expressed between EBs and RBs



Where *Chlamydia trachomatis* invests its ATP

Energy expenditure by functional category



Quantification of the *Chlamydia trachomatis* proteome

Some conclusions

- Absolute quantification of most of predicted proteome in both RBs and EBs
- Rank order of expression reveals hitherto hypothetical proteins are among the most abundant in Chlamydia
- Dynamic expression range of >3 logs - 37 pg (AMP nucleosidase) to 29 ng (MOMP).
- EBs appear to have full complement of proteins even though metabolically quiescent
- Levels of most proteins are down in EBs but some accumulate (in anticipation of infection?)

Quantification of the *Chlamydia trachomatis* proteome

Some conclusions (cont.)

- LC-MS^E provides more extensive and robust qualitative and quantitative data relative to iTRAQ
- >71% of predicted *C. trachomatis* proteome is expressed during transition from RB to EB
- Absolute quantification data obtained for >62% of predicted proteome
- Differential expression data indicates C. *trachomatis* shuts down metabolic activity during the transition from RB to EB (e.g. glycolysis, TCA)
- Cell wall enzymes expressed in RBs – suggests novel role
- Majority of energy invested in protein translation machinery, one cell surface component and many hypothetical proteins

Label-free quantification

Some key challenges and issues

- Given sensitivity of detection is <10 molecules/cell, why is ‘only’ 71% of predicted proteome detected?
- Use of LC-MS^E for the quantification of PTMs?
- Faster cycling rate for MS^E (>10 Hz)?
- Multiplexing of LC-MS^E analyses?

Biological insights from large-scale protein copy number measurements

Acknowledgements

- Paul Skipp (CPR)
- Erika Parkinson (CPR)
- Pete Boyd (CPR)
- Shannon Pead (CPR)
- Ian Clarke (SGH)
- Lesley Cutcliffe (SGH)
- Jim Langridge (Waters)
- Chris Hughes (Waters)
- Therese McKenna (Waters)
- Simon Harris (Sanger)
- Helena Seth-Smith (Sanger)
- Nick Thomson (Sanger)

Funders: BBSRC, EU, MRC, NIHR, Unilever plc, Wellcome Trust