Evolution of the protein stoichiometry in the L12 stalk of bacterial and organellar ribosomes

Iakov I. Davydov1,*, Ingo Wohlgemuth2,3,*, Irena I. Artamonova4,5,6, Henning Urlaub3,7, Alexander G. Tonevitsky8,9 & Marina V. Rodnina2

The emergence of ribosomes and translation factors is central for understanding the origin of life. Recruitment of translation factors to bacterial ribosomes is mediated by the L12 stalk composed of protein L10 and several copies of protein L12, the only multi-copy protein of the ribosome. Here we predict stoichiometries of L12 stalk for >1,200 bacteria, mitochondria and chloroplasts by a computational analysis, and validate the predictions by quantitative mass spectrometry. The majority of bacteria have L12 stalks allowing for binding of four or six copies of L12, largely independent of the taxonomic group or living conditions of the bacteria, whereas some cyanobacteria have eight copies. Mitochondrial and chloroplast ribosomes can accommodate six copies of L12. The last universal common ancestor probably had six molecules of L12 molecules bound to L10. Changes of the stalk composition provide a unique possibility to trace the evolution of protein components of the ribosome.
Ribosomes—the ancient, ubiquitous protein factories—consist of RNA ribosomal and ribosomal proteins (r-proteins). The sequence and structure of rRNA is extremely well conserved, which allows using the rRNA sequence for construction of phylogenetic trees and analysis of phylogenetic relationships and evolution. Also the sequence and the stoichiometry of r-proteins, which are usually present in one copy per ribosome, are highly conserved. The only r-protein that is present in multiple copies is L12, a protein of the large ribosomal subunit, which is part of the so-called L12 stalk of the ribosome (in bacteria and archaea, or P-stalk in eukaryotes). The L12 stalk entails ribosomal protein L10 (or its archaean and eukaryotic orthologue P0 in the P-stalk) and multiple copies of protein L12 (P1/P2 in eukaryotes). The L12 stalk is required for translation, because it recruits to the ribosome the auxiliary translation factors, in particular translational GTPases, such as initiation factor 2, elongation factors Tu and G, and release factor 3 or their eukaryotic homologues, and accelerates GTPase activity of some of them. The mitochondrial homologue of L12, MRPL12, not only has an important role in translation but also regulates the transcription in mitochondria. Variations in the L12 sequence between species provide the specificity for the interactions with translation factors: the replacement of the L12–L10 stalk in *Escherichia coli* ribosomes with its eukaryotic counterpart (P0/P1/P2 stalk) makes the ribosome to bind eukaryotic, rather than bacterial elongation factor. The evolutionary and functional advantage of having multiple L12 copies is not known; reducing the L12 copy number to just two copies affects both translation efficiency and accuracy.

Ribosomes from *E. coli* have four molecules of the L12 protein arranged in two dimers bound to one molecule of L10 (refs 17–20). The 4:1 stoichiometry was also found in ribosomes from *Bacillus subtilis* and *Bacillus stearothermophilus*. In contrast, ribosomes from *Agrobacterium tumefaciens* and from a number of thermophilic bacteria, such as *Thermotoga maritima, Thermus thermophilus* or *Thermus aquaticus* contain six copies of L12 per L10 (refs 9,21,23), suggesting that the 4:1 stoichiometry of L12: L10 is not universal. This raised the questions of whether stoichiometries other than 4:1 and 6:1 exist in bacteria, whether a given stoichiometry is related to a particular taxonomic group or a specific living environment, or which evolutionary processes resulted in a given stoichiometry. Biochemical and mass spectrometry methods that were used so far to quantify the L12 copy numbers are not suitable for a large-scale screening of hundreds of species. Here we used computational tools to predict the ribosomal L12 stalk composition for a wide range of bacteria, mitochondria and chloroplasts. The predictions for the 8:1 stoichiometry, which has not been seen so far, are validated by mass spectrometry.

**Results**

**L12 copy number in bacteria.** L12 entails two domains, the amino-terminal domain (NTD) and the carboxy-terminal domain, which are connected by a flexible hinge region. The globular carboxy-terminal domain interacts with translation factors, whereas the α-helical NTD is responsible for L12 dimerization and binding to L10. L10 contains eight α-helices and four β-sheets, and the L12 NTD dimers bind to consecutive elements in the C-terminal helix of L10, which comprises α-helical segments separated by bends (Fig. 1). Although in all bacterial species studied so far each segment in helix α8 binds an L12 dimer, the sequence conservation between the segments is too low to allow for a reliable identification of L12 binding in other organisms. However, the copy number of L12 per ribosome depends on the length of helix α8, that is, the number of consecutive helical elements. Thus, secondary structure predictions can be used to locate helix α8 and estimate the number of potential L12-binding segments. This structure-guided approach was first used to predict the 6:1 ratio of L12:L10 in ribosomes from *T. maritima* and then to analyse bacterial, archaean and eukaryotic stalks. Sequence analyses of 28 bacterial, species suggested 4:1 and 6:1 stoichiometries in bacteria with a tendency towards 6:1 for thermophilic organisms, which may, however, be due to the limited number of sequences studied.

To predict L12 copy numbers, >2,000 sequences of r-protein L10 were downloaded from the UniProt database and grouped into 754 characteristic clusters (see Methods), representing >1,200 different species. For each of these sequences, the length of helix α8 was predicted (see Methods). The distribution of predicted lengths has a clearly bimodal shape with major peaks corresponding to lengths of helix α8 of 32–33 and 40–42 amino acids (Fig. 2a); a few examples for longer α8 helices of around 55 amino acids are also seen. The difference between the two major modes corresponds to exactly one binding segment for an L12 dimer, suggesting that the species with the shorter helix α8 can bind two L12 dimers and those with the longer helix α8 have three L12-binding segments, and thus may have six copies of L12 per L10. The predictions were consistent with the experimentally determined L12 copy numbers for all organisms, where biochemical or mass spectrometry data are available ( Supplementary Table S1). Phylogenetic analysis suggests that the species with two or three L12-binding segments on L10 are evenly distributed among the tree (Supplementary Figs S1–S4), with 686 sequences with 3 segments and 535 with 2 (Supplementary Datasets S1, S2).

Comparison of the sequences of the individual L12-binding segments of L10 provides an insight into the evolution of helix α8 (Fig. 2b,c). For the three L12-binding segments from *T. maritima* similarity between segments 2 and 3 is significantly higher (motif similarity score 1.18) than between segments 1 and 2 (0.27) or 1 and 3 (0.45). This suggests that the additional, third L12-binding segment has evolved via a duplication of the second segment and subsequent sequence divergence. Sequence clustering analysis (Supplementary Fig. S5) suggests that the first L12-binding segment is more conserved than the other binding segments. Although less conserved than the first segment, the distal segments show a significant degree of sequence similarity to adjacent segments, suggesting that they emerged from consecutive duplication events.
Bacteria with eight copies of protein L12 per ribosome. Although the length of helix $a_8$ in most bacteria is below 47 amino acids, a few organisms have a particularly long helix $a_8$ comprised of about 55 amino acids with a characteristic pattern of segments that are separated by bends (Fig. 3), suggesting that these ribosomes may bind 8 copies of L12 (Supplementary Table S2 and Supplementary Fig. S6). Most of the organisms that appear to have four potential L12 dimer binding sites are cyanobacteria (Fig. 4), except for two species in the Roseiflexus family.

To verify the L12:L10 stoichiometry in a species with a particularly long helix $a_8$, we applied a mass spectrometry approach to determine the composition of the L12 stalk of ribosomes purified from *Athrospira platensis*. We chose *A. platensis* as a model organism because it can be easily grown in laboratory cultures26, and the ribosomes could be prepared using established methods27. In contrast to other mass spectrometry approaches that analyse intact macromolecules21, the present approach28 allows the absolute quantification of L12:L10 stoichiometry based on the ratio of peptides obtained by protease digestion, the concentrations of which reflect the concentration of their precursor proteins29 (Fig. 5), provided protein digestion is complete30 (Supplementary Fig. S7). To achieve precise

![Graph](image-url)

**Figure 2 | L12-binding segments of helix $a_8$ of bacterial L10 proteins.** 
(a) Distribution of the predicted length of helix $a_8$ among L10 proteins. (b,c) Motif logos for the L12-binding segments in helix $a_8$ of L10 sequences, which have more than 50% identity with *E. coli* (b) and more than 40% identity with *T. maritima* L10 (c). Here and later the numbering starts from the fifth amino acid of $a_8$ helix, which corresponds to the start of the first L12-binding segment9.

![Graph](image-url)

**Figure 3 | Protein L10 from Synechococcus sp. strain JA-3-3Ab** 
(a) Secondary structure predictions of the secondary structure elements (top, $a$-helices; middle, $b$-sheets; bottom, random coil) of L10. (b) The confidence of $a$-helix formation prediction for a given amino acid position in L10; the maximum confidence is set to 1. Amino acid residues are numbered (x axis) starting from the N-terminal fMet. (c) Sequence logos of the L12-binding motifs in helix $a_8$ of L10 from Cyanobacteria.
mass spectrometer of purified ribosomes that were digested by chromatography tandem mass spectrometry on an Orbitrap L10) were chosen based on the initial analysis by liquid standards to quantify the mass spectrometric signal intensities in endogenous peptides to be analysed. Aqua peptides added in same sequence and physico-chemical properties as the quantification peptides, Aqua peptides) were used that had the colour.

The predicted L12:L10 stoichiometry is indicated by orange (6:1) or blue (8:1) colour.

To obtain reliable results, three to four SRM transitions were chosen for the quantification of each peptide (Fig. 5b–d and Supplementary Table S3). The linearity of the sample signal response in the SRM experiments was established by titrations of each Aqua peptide over four orders of magnitude at constant concentration of endogenous peptides (Supplementary Fig. S8a,b). We varied the concentration of either Aqua or sample peptides (Supplementary Fig. S8c,d) within a concentration range around a fixed concentration of sample or Aqua peptides, respectively. Knowing the concentrations of the two peptides of L12 and the three peptides of L10 in the digestion mix, we calculated the L12:L10 stoichiometry from six possible combinations of peptides for each titration point. As we found no systematic bias for any of the combinations, we averaged over all titration points and all technical and biological replicates. This approach yielded a L12:L10 ratio of 4.1 ± 0.6 for E. coli (based on 702 independent values; Fig. 5e and Supplementary Dataset S3), in agreement with the published values. For purified ribosomes from A. platensis the L12:L10 ratio was 7.7 ± 1.1, suggesting close to four L12 dimers attached to L10, in accordance with the particularly long z8 helix of that organism, as predicted by the computational approach.

Notably, not all cyanobacteria included in the computational analysis have a long helix z8, and those with the long helix z8 do not form a monophyletic group (Fig. 4). Given the high sequence similarity in helix z8 for cyanobacteria that have four L12-binding segments (Supplementary Fig. S6), an independent acquisition of the additional segment seems highly unlikely. This suggests that the duplication of the L12-binding segment in helix z8 occurred in a common ancestor of the cyanobacteria. Furthermore, Bayesian ancestral state inference21 (see Methods) suggests that the common ancestor of cyanobacteria had a long helix z8, which could accommodate four L12 dimers (calculated probability 0.86). Notably, this kind of analysis does not take into account the sequence similarity among L12-binding motifs, providing an independent indication for a high L12 copy number on the ribosomes of the ancestral cyanobacteria. The shorter L10 variants in some cyanobacteria then would be the result of a later loss of one L12-binding segment. Of the four predicted L12-binding segments of L10 from cyanobacteria, segment 1 is similar (1.36) to segment 1 of helix z8 of L10 from T. maritima sequence family (Figs 2c and 3b), which further confirms high conservation between the first binding segments across species. Segments 3 and 4 from cyanobacteria have very similar sequences (1.8), but have significantly lower similarity scores to segments 1 and 2 (0.18–0.72).
L12 copy number in mitochondria and chloroplasts. With the same computational approach as described above, we analysed the length of helix $a_8$ in the mitochondrial homologue of L10, MRPL10, and in L10 from chloroplasts using all sequences that could be unambiguously assigned as organellar L10 in eukaryotic genomes. In general, mitochondrial ribosomes are quite different from bacterial ones, as they contain many more proteins and less rRNA. The fact that the closest MRPL10 orthologue is L10 from alphaproteobacteria reflects the alphaproteobacterial origin of mitochondria. We analysed MRPL10 sequences from 82 eukaryotes (Supplementary Table S4). The prediction of three L12-binding segments of MRPL10 indicated that mitochondrial ribosomes bind six molecules (three dimers) of L12 (Fig. 6).

Emergence of the L12-binding segments. The phylogenetic analysis suggests two possible evolutionary scenarios for the change of the L12 copy number, that is, the acquisition or the loss of helix $a_8$ fragments of L10 (Fig. 7a,b). The two possibilities can be distinguished by analysing the phylogenetic tree of L10. In cases where many species predicted to have three L12-binding segments cluster closely together, whereas only a few species in the group have shorter helices $a_8$, we classified the event as a segment loss. Vice versa, if a few members of a phylogenetically related group have a longer helix $a_8$ indicative of binding of six copies of L12, whereas the majority has a shorter helix $a_8$, the event can be classified as acquisition of a segment. Phylogenetic analysis indicated a loss of 11 amino acids in helix $a_8$ of L10 in Deinococcus, Opitutaceae, Veillonella, Dichelobacter and other species (Supplementary Table S5). All these chloroplast L10 sequences have three L12-binding segments.
structure prediction. The centre of the ribosome, which appears to be a relic from that time, is built of RNA, and its activity is not appreciably affected by the r-proteins (for review, see ref. 39). The time of adding r-proteins is controversial. It seems reasonable to assume that the emergence of the L12 stalk proteins is coupled to the evolution of the translation factors and leads to higher speed and fidelity of translation, which is consistent with the important role of L12 in the recruitment of translation factors to the modern ribosome. Very early in evolution, probably only one copy of L12 was present on the ancestral ribosome (Fig. 7c). The earliest event that altered the L12 copy number must have been the emergence of the L12 NTD dimerization surface and its binding surface on L10; however, the traces of these stages of ribosome evolution eroded away. It is likely that this first sole binding segment was duplicated early in evolution. In fact, none of the organisms for which genomic data is available so far has L10 with helix a8 so short as to account for only one L12 dimer binding.

Figure 6 | L10 from human mitochondrial protein MRPL10. (a) Secondary structure prediction. (b) The confidence of α-helix formation prediction. (c) Sequence logos of the L12-binding segments in helix a8.

Methylocutea. In contrast, an acquisition of 11 amino acids seemed to have occurred in Clostridiales (Fig. 7b) and in Nitratiruptor. In general, it appears that the loss of a L12-binding segment is the more frequent event. Bayesian approach to ancestral state inference suggested a higher probability (0.67) for helix a8 in L10 of the last bacterial common ancestor to accommodate six copies of L12, rather than four copies (probability 0.33; organisms with eight copies of L12 bound to L10 are too rare to be included in the analysis).

Discussion
The origin and evolution of ribosomes are central questions for understanding the emergence of life. It is generally assumed that the ribosome emerged from the RNA world, when the proteins did not exist yet, and RNA acted as catalyst in chemical reactions (for recent reviews see refs 34,35). In fact, the peptidyl transferase centre of the ribosome, which appears to be a relic from that ancient time, is built of RNA, and its activity is not appreciably affected by the r-proteins (for review, see ref. 39). The time of adding r-proteins is controversial. It seems reasonable to assume that the emergence of the L12 stalk proteins is coupled to the evolution of the translation factors and leads to higher speed and fidelity of translation, which is consistent with the important role of L12 in the recruitment of translation factors to the modern ribosome. Very early in evolution, probably only one copy of L12 was present on the ancestral ribosome (Fig. 7c). The earliest event that altered the L12 copy number must have been the emergence of the L12 NTD dimerization surface and its binding surface on L10; however, the traces of these stages of ribosome evolution eroded away. It is likely that this first sole binding segment was duplicated early in evolution. In fact, none of the organisms for which genomic data is available so far has L10 with helix a8 so short as to account for only one L12 dimer binding. Rather, the length of helix a8 in most organisms is consistent with binding of two or three L12 dimers, largely independent of the taxonomic group or the living conditions of bacteria. The only exceptions from this ‘four or six copies’ rule are found in cyanobacteria, which show a very long helix a8, indicating binding sites for eight copies of L12. Organisms with a higher potential copy number were not identified. The ribosomes from mitochondria and chloroplasts have three L12-binding segments on L10, suggesting the presence of six L12 copies on those ribosomes. Notably, in all cases where the L12:L10 stoichiometry was determined experimentally, the L12 copy number per ribosome is in excellent agreement with the computationally predicted value (Supplementary Table S1), providing a strong support for the validity of the present computational analysis.

Bayesian inference of the ancestral state using the distribution of species with different L12 copy numbers and the sequence analysis of the individual L12-binding segments of L10 from cyanobacteria suggest that in the last bacterial common ancestor the L12 copy number was at least six, whereas the last common ancestor of cyanobacteria had eight copies of L12 per ribosome. As also archaeal ribosomes can bind six copies of L12 (refs 14,21), it is likely that the hypothetical last universal common ancestor had six copies of L12 (Fig. 7c). In this scenario, modern bacterial species with four L12 copies emerged through a loss of one of the three L12-binding segments of helix a8 of L10, whereas, chloroplasts and those cyanobacteria that have six copies of L12 lost one of four segments.

One interesting observation is that the sequence of the first L12-binding segment is more evolutionarily conserved than the sequences of the other segments. One can hypothesize that the primary structure of the first segment is particularly important for L12 recruitment or the function of the ribosome stalk, thereby limiting the sequence variations, whereas deviations in sequences of other binding segments apparently can be tolerated better. Incidentally, the binding of the L12 dimer to the first segment of helix a8 is stabilized by interactions with the L10 NTD, which are not seen for the more distal dimers. This additional structural constraint may explain the higher degree of sequence conservation in the proximal segment of L10. The distal domains appeared as a result of sequential duplications followed by divergence in the primary sequence of the segments, while retaining their ability to bind L12. The high sequence similarity between the last two binding segments of L10 from cyanobacteria suggests that the additional binding segment resulted from a recent duplication event.

The reasons for acquiring multiple copies of L12 and for variable copy numbers of L12 are not known. The high copy number does not correlate with thermophilicity, in contrast to suggestions based on analysis of a limited number of bacterial
species\textsuperscript{14,21}. Multiple copies of the protein may increase the encounter frequency with translation factors, thereby facilitating their recruitment to the ribosome. Kinetic evidence suggests that a complete removal of L12 from \textit{E. coli} ribosomes decreases the rate of elongation factor Tu binding to the ribosome more than tenfold\textsuperscript{9}. On the other hand, ribosomes that have only one L12 dimer are active in translation \textit{in vitro}\textsuperscript{41}. An \textit{E. coli} strain with mutant L10 that can only bind a single L12 dimer is viable, but shows growth defects and is less efficient than the wild-type strain in initiation and elongation\textsuperscript{15}. Ribosomes with four L12 copies (from \textit{E. coli}) perform all translation functions with the same speed and efficiency as those with six copies (from \textit{M. smegmatis}), both with homologous and heterologous factors\textsuperscript{42}. It appears that adaptive selection tends to maintain the stoichiometry in the L12 stalk between four and eight copies of L12. Variations in this range are tolerated without appreciably affecting the activity of the ribosomes, whereas reducing or increasing the copy number beyond these limits would probably lead to reduced fitness of bacteria.

Our quantitative analysis of purified ribosomes by mass spectrometry suggests that essentially all potential binding segments on L10 are occupied by L12, which is consistent with the earlier immunoblot analysis of other bacterial ribosomes, for example, from \textit{E. coli}, \textit{T. maritima}\textsuperscript{9} and \textit{M. smegmatis}\textsuperscript{4:1 6:1}.

\textbf{Figure 7 | Evolutionary scenarios for the emergence of the L12 copy number.} (a) Loss of an L12-binding segments in helix $a_8$ of L10. (b) Acquisition of an additional potential L12-binding segment. The phylogenetic trees are based on 16S rRNA sequences. The predicted L12:L10 stoichiometry is indicated by orange (6:1) or green (4:1) colour. (c) A schematic of a proposed evolutionary path. Dashed lines indicate those changes in the L12:L10 copy number for which only limited experimental evidence exists.
(unpublished data), and mass spectroscopy data. This suggests that the maximum possible L12:L10 occupancy is 10% of L12 per hour. Archaeal stalk may be less stable than in ribosome is very stable, allowing for the dissociation of less than 10% of L12 per hour. The expression of L10 and L12 is controlled by an autoregulation mechanism where the L10:L12 complex inhibits the translation of its own mRNA by binding to the leader preceding the coding sequence. In addition, although the L10 in the stalk is rather stable, excess free L10 is probably rapidly degraded, thereby maintaining matching relative concentrations of the two proteins in the cell. In archaea, the control mechanisms appear to be different, because the gene coding L10 and L12 as part of the L1 operon. However, the L10:L12 complex does not regulate the expression of the operon. In summary, the present combination of computational predictions and quantitative mass spectrometry provides an insight into the emergence of protein stoichiometry in the functionally important ribosomal L12 stalk and suggests a wide range of stalk compositions—with four, six or eight copies of L12 per L10—on ribosomes from bacteria, mitochondria and chloroplasts.

Methods

Sequence selection. The UniProt database was used as a source for amino acid sequences. L10 and L12 sequences were retrieved using family requests 'ribosomal protein L10P family' and 'ribosomal protein L12P family', respectively. Sequences with 90% sequence identities were clustered, and each cluster was represented by a single sequence in further analysis. Manual curation included the analysis of sequence completeness and similarity to other sequences. 754 L10 sequences formed the target sample that represented >2,000 individual sequences. Sequence logos were created using WebLogo server, using the chemistry colour scheme. A search of the NCBI database for organellar L10 sequences yielded 82 mitochondrial and 29 chloroplastic sequences that could be unambiguously annotated. The sequences of putative L10 of several plants were not included because of their unclear origin.

Helix prediction algorithm. As the divergence in sequences precludes the unbiased identification of the potential L12-binding segments based on the sequence comparison, the L12-binding segments were located based on the length of helix X8. The PSIPRED software was used to predict secondary structures. A continuous helix in L10 was defined by a helix width of 3.0. The reaction mixture was then mixed in glass vials in 10% acetonitrile/0.1% formic acid in a final volume of 20 μl. Heavy isotope-labelled Aqua peptides (5 μM) were purchased from Thermo Scientific. SRM measurements were performed on a TSQ Vantage Triple Quadrupole mass spectrometer (Thermo Fisher Scientific). Chromatographic separations of peptides were performed on an Easy nLC II Nano LC system (Thermo Fisher Scientific). The peptides were loaded on a self-packed trap column (2 cm length, 150 μm inner diameter, packed with Reprosil-Pur 120 C18 material (Dr Maisch). The nanoLC was operated at a flow of 300 nl min⁻¹. The PSIPRED software was used to predict secondary structures. A continuous helix in L10 was defined by a helix width of 3.0. The reaction mixture was then mixed in glass vials in 10% acetonitrile/0.1% formic acid in a final volume of 20 μl. Heavy isotope-labelled Aqua peptides (5 μM) were purchased from Thermo Scientific. Absolute quantification by SRM. SRM measurements were performed on a TSQ Vantage Triple Quadrupole mass spectrometer (Thermo Fisher Scientific). Chromatographic separations of peptides were performed on an Easy nLC II Nano LC system (Thermo Fisher Scientific). The peptides were loaded on a self-packed trap column (2 cm length, 150 μm inner diameter, packed with Reprosil-Pur 120 C18 material (Dr Maisch) with a maximal flow of 10 μl min⁻¹ (buffer B: 2% acetonitrile/0.1% formic acid) limited by a pressure maximum of 200 bar. The trap column was washed with 50 μl under the same conditions. The peptides were eluted with a 0.5% gradient (5 min, 0% acetonitrile/0.1% formic acid to 60% acetonitrile/0.1% formic acid from an in-house packed trap column (14 cm length, 75 μm inner diameter, packed with Reprosil-Pur 120 C18 3 μm material (Dr Maisch). The nanoLC was operated at a flow of 300 nl min⁻¹. Q1 and Q3 were both set to unit resolution (0.7 full width at half maximum). A spray voltage of 1,650 V was used with a heated ion transfer setting of 270 °C for desolvation. The declustering voltage was kept constant at 10 V and the Chromatofor 15 was used. For each SRM transition, the collision energy was optimized by regular
injection of Aqua peptides varying the collision energy (± 5) around the theoretical value predicted by the skyline software. SRM transitions for the native peptides were obtained from Q3 mass transistion, using the expected mass differential from the standard peptides by the Skyline software (Supplementary Table S3). Schedules transitions were recorded in a time window of 4 min. The cycle time was 3 s and the average dwell time 200–500 ms per transition. The peptide ratios were obtained by peak integration of the individual transitions using the Skyline software (https://skyline.proteomex.com/software/skyline). The individual transitions show an s.d. of less than 5% and were therefore averaged to yield one peptide ratio. The fact that the individual transitions show identical ratios between endogenous and Aqua peptide, as well as same chromatographic retention time, together with the observation that the different peptides show similar ratios, clearly demonstrates that the detected SRM signal indeed derives from the targeted peptides. In each experiment, 78 individual complex stoichiometries were calculated (48 from the Aqua titration and 30 from the sample titration). For the determination of the complex stoichiometry, three independent biological samples were analysed, each of them in triplicate. Thus, the determined complex stoichiometry for each combination is the mean of 78 individual values for the complex stoichiometries that were generated based on 117 individual HPLC runs.

References
Acknowledgements

We thank Mikhail Gelfand for valuable suggestions, Wolfgang Wintemeyer for critical reading of the manuscript, Dmitry Los and Inna Maslova for providing A. platensis cells, and Christina Kothe, Sandra Kappler and Uwe Plessmann for expert technical assistance. This work was supported by Ministry of Education and Science of the Russian Federation (no. 16.522.12.2015 to I.I.D.), by the Max Planck Society (M.V.R., H.U. and I.W.), the Deutsche Forschungsgemeinschaft (M.V.R. and H.U.), and by the Russian Academy of Sciences ("Molecular and Cellular Biology" programme, to I.I.A.). Some of the results were acquired using the supercomputer SKIF MSU "Chebyshev" of the Moscow State University Supercomputer Complex. We thank V.V. Voevodin and S.A. Zhumatia for the opportunity to perform calculations on the supercomputer.

Author contributions

I.I.D. carried out computational analysis with the help of I.I.A., I.W. performed the mass spectrometric analyses under supervision of H.U., and all authors contributed to design of the experiments, data interpretation and writing the paper.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/