Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 4125; doi:10.3390/ijms19124125

Article

Muller's ratchet and ribosome degeneration in the obligate intracellular parasites *Microsporidia*

Sergey V. Melnikov[1], Kasidet Manakongtreecheep[1,#], Keith Rivera[2], Arthur Makarenko[2], Darryl Pappin[2], and Dieter Söll [1,3]*

[1] Department of Molecular Biophysics and Biochemistry, Yale University;
[2] Cold Spring Harbor Laboratory, Cold Spring Harbor;
[3] Department of Molecular Biophysics and Biochemistry, Department of Chemistry, Yale University;
Present address: Massachusetts General Hospital Cancer Center, Harvard Medical School

* Correspondence:
dieter.soll@yale.edu; Tel.: +1-203-432-6200

Abstract: *Microsporidia* are fungi-like parasites that have the smallest known eukaryotic genome, and for that reason they are used as a model to study the phenomenon of genome decay in parasitic forms of life. Similar to other intracellular parasites that reproduce asexually in an environment with alleviated natural selection, *Microsporidia* experience continuous genome decay driven by Muller's ratchet – an evolutionary process of irreversible accumulation of deleterious mutations, which leads to gene loss and miniaturization of cellular components. Particularly, *Microsporidia* have remarkably small ribosomes in which the rRNA is reduced to the minimal enzymatic core. To better understand the impact of Muller's ratchet on RNA and protein molecules in parasitic organisms, particularly regarding their ribosome structure, we have explored an apparent effect of Muller's ratchet on microsporidian ribosomal proteins. Through mass spectrometry, analysis of microsporidian genome sequences and analysis of ribosome structure from non-parasitic eukaryotes, we found that massive rRNA reduction in microsporidian ribosomes appears to annihilate binding sites for ribosomal proteins eL8, eL27, and eS31, suggesting that these proteins are no longer bound to the ribosome in microsporidian species. We then provided an evidence that protein eS31 is retained in *Microsporidia* due to its non-ribosomal function in ubiquitin biogenesis. To sum up, our study illustrates that while *Microsporidia* carry the same set of ribosomal proteins as non-parasitic eukaryotes, some of ribosomal proteins are no longer participating in protein synthesis in *Microsporidia* and they are preserved from genome decay by having extra-ribosomal functions.

Keywords: Muller's ratchet; genome decay; ribosome; rRNA expansions; rudimentary proteins.

1. Introduction

It is well documented that parasitic lifestyle causes continuous genome decay manifested in progressive gene loss and accumulation of deleterious mutations in essential genes [1-3]. Yet we are only beginning to understand how genome decay impacts structure of individual proteins and nucleic acids in parasitic organisms.

Among numerous forms of life that inhabit our planet, there is a group of single-cell pathogens that can grow and reproduce only inside other living cells. These organisms evolve under conditions radically different from those of free-living species: they live in small population with little competition for nutrients and shelter, which weakens natural selection; they undergo repeated population bottlenecks upon transition from one host cell to another, which favors genetic drifts; and they proliferate asexually, which prevents elimination of toxic mutations through recombination of genetic material [3]. Consequently, obligate intracellular parasites evolve under conditions that favor gradual and irreversible accumulation of deleterious mutations. In the long-term, this irreversible

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 4125; doi:10.3390/ijms19124125

process, known as Muller's ratchet, causes genome decay reflected in massive loss of non-essential genes and increasing number of deleterious mutations in essential genes in parasite genomes [3, 4].

The genome decay caused by Muller's ratchet can be observed at every level of cellular organization, from organelles to individual proteins and nucleic acids. For instance, if we take genome size in bacterial species, more than over 4,000 protein-coding genes are typically found in species with free-living lifestyle; approximately ~2,000-3,000 protein-coding genes in species with hybrid lifestyle with elements of free-living and parasitism, less than ~1,500 of protein-coding genes are found in most parasitic species with strictly intracellular lifestyle. In some extreme cases of long-term intracellular parasitism, the total number of genes can be as low as 500 protein-coding genes [5]. The remaining essential genes in parasitic organisms undergo truncations and accumulate deleterious mutations that compromise protein and RNA folding, reduce enzymatic activity, and possibly reduce specificity of molecular interactions inside parasitic cells [6, 7].

An increasing interest to the phenomenon of genome decay stems from a growing evidence that this erosive evolutionary process might be used as an "Achilles' heel" to create new therapies against parasitic infections [7-10]. For instance, massive loss of genes involved in nutrient biosynthesis, nutrient sensing, stress response, and quality control, can make parasitic cells vulnerable to stressors and factors that require activity of quality control systems [7, 8]. Or accumulation of deleterious mutations in essential proteins and nucleic acids may make their structures more sensitive to heat and more dependent on the activity of molecular chaperones, suggesting hypersensitivity of parasites to heat shock and chaperone inhibitors [9-13]. Importantly, the genome decay in parasitic forms of life is observed in intracellular pathogens from all kingdoms of life, regardless of their evolutionary origin [2, 5]. Hence, if therapeutic targeting of decayed proteins and nucleic acids can be found for one group of parasites, it can be potentially applied to a much broader range of parasitic species.

One clinically relevant model to study extreme genome decay can be observed in the fungi-related pathogens *Microsporidia*. These organisms have the smallest known eukaryotic genome, with many microsporidian species comprising as little as ~1,800 genes, which is ~3 times smaller than genomes of non-parasitic fungi, such as the yeast *Saccharomyces cerevisiae* [14, 15]. *Microsporidia's* miniature bodies are made of a single cell whose size is comparable to that of *Escherichia coli*. All known *Microsporidia* species are parasites; several species, including *Enterocytozoon bieneusi*, *Encephalitozoon cuniculi*, and *Trachipleistophora hominis*, are found in ~10-40% HIV-positive humans, where they cause chronic infections of the gastrointestinal tract, eye and kidney [16]. Due to lack of effective treatment against these pathogens, they are included in the list of priority emerging pathogens by the National Institute of Allergy and Infectious Diseases [17].

As is the case in other intracellular parasites, regressive changes in *Microsporidia* can be observed at every scale of their organization, from their overall cell structure to the structure of individual macromolecules [14, 15]. The most impressive regressions are observed in the protein synthesis machinery, particularly in the ribosome. Compared to ribosomes from other eukaryotes, whose molecular weight varies between ~3.3 MDa and ~4.5 MDa, microsporidian ribosomes are expected to weigh only ~2.2 MDa, which is even smaller than ~2.5 MDa ribosomes from bacteria [18-20]. This remarkable reduction is achieved primarily by massive degeneration of microsporidian rRNA, particularly degeneration of rRNA expansions (rRNA segments that are missing in bacteria, but present in most eukaryotes, where many of them are thought to be involved in regulation of protein synthesis). For instance, in most microsporidian species, the 18S rRNA length varies around ~1,250-1,350 bases, compared to ~1,800-1,850 bases in non-parasitic eukaryotes [19-22]. This is especially impressive given the fact that even in bacterial ribosomes the 16S rRNA comprises ~1,600 bases. It is therefore assumed that microsporidian ribosomes represent the minimal machinery of protein synthesis stripped down from regulatory and quality control-related components and debased to the minimal enzymatic core [14, 19].

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 4125; doi:10.3390/ijms19124125

To better understand the phenomenon of molecular degeneration in parasitic forms of life, we explored in this study the fate of ribosomal proteins in microsporidian parasites. Particularly, we asked how massive loss of rRNA expansions could have affected the ribosomal proteins that are bound to rRNA expansions in ribosomes from non-parasitic eukaryotes. By using comparative analysis of microsporidian genome sequences and analysis of structures of fungal ribosomes, we showed that massive reduction of microsporidian rRNA appears to eliminate the interface between ribosomal RNA and several ribosomal proteins, including eS31, eL8, and eL27. These ribosomal proteins appear to lose their attachment to the ribosome and turn into free-standing proteins in microsporidian parasites. We further used mass spectrometry analysis of proteins from a microsporidian *E. cuniculi* to show that eS31, eL8, and eL27 proteins are still abundantly expressed in the microsporidian proteome. For protein eS31, we show that its preservation in the microsporidian proteome is likely related to the extraribosomal function of this protein in ubiquitin biogenesis. In sum, our study illustrates that despite seeming conservation in parasites with highly reduced genomes, some components of the protein synthesis machinery could have lost their original role in protein synthesis and are retained in parasites by having non-protein synthesis related activities.

2. Materials and Methods

Analysis of rRNA-protein contacts evolution in the ribosome structure. To illustrate rRNA degeneration in microsporidian parasites, we compared secondary structure diagrams of rRNA from the microsporidian *Encephalitozoon cuniculi* and the yeast *Saccharomyces cerevisiae.* For both species, the diagrams of 18S, 5.8S and 25S rRNA were retrieved from the Comparative RNA Website (http://www.rna.icmb.utexas.edu/) and from the previous studies of *E. cuniculi* rRNA degeneration [19, 20]. Subsequently, the crystal structure of the 80S *S. cerevisiae* ribosome was used to map rRNA segments that are present in *S. cevevisiae* but degenerated in *E. cuniculi*.

The rRNA-protein interface was calculated by using the yeast 80S ribosome structure (pdb id 5dgv) and an online tool to measure protein interfaces in macromolecular assemblies jsPISA from the CPP4 package (http://www.ccp4.ac.uk/pisa/) [23]. Conservation of protein sequences was estimated by using multiple sequence alignments with ClustalO [24]. Microsporidian species that were used in the analysis shown in **Figure 2B** are *Nosema ceranae*, *Edhazardia aedis*, *Spraguea lophii*, *Trachipleistophora hominis*, *Vavraia culicis*, *Vittaforma corneae*, *Nematocida ausubeli*, *Nematocida parisii*, *Nematocida displodere*, *Enterocytozoon hepatopenaei*, *Encephalitozoon cuniculi*, *Mitosporidium daphnia*, *Pseudoloma neurophilia* (note that in some species ubiquitin-eS31 is annotated as ubiquitin-S27a or ubiquitin-carboxy-terminal extension protein). Yeast sequences that were used in the analysis shown in **Figure 2A** are *Kluyveromyces lactis*, *Schizosaccharomyces pombe*, *Scheffersomyces stipites*, *Exophiala dermatitidis*, *Coniosporium apollinis*, *Candida auris*, *Pseudozyma antarctica*, *Ogataea parapolymorpha*, *Wickerhamomyces ciferrii*, *Hanseniaspora uvarum*, *Pichia kudriavzevii*, *Tetrapisispora phaffii*, *Torulaspora delbrueckii*.

Preparation of protein extracts from microsporidian spores. Microsporidian spore production and protein extraction was adapted from [25]. The *E. cuniculi* isolate GB-M1 used for genome sequencing was cultured in Madin-Darby canine kidney or human foreskin fibroblast cells as previously described [26]. After collection of culture supernatants, parasites were sedimented (5,000g, 5 min) and heated at 65°C in a 1% SDS solution. A spore-rich fraction was recovered after several washes in water to remove host cell debris. Parasites were disrupted in a lysis buffer (10^9-10^{10} cells/mL) containing 100 mM DTT, 4% CHAPS and 0.2% SDS, by repeated cycles of freezing-thawing and sonication (Deltasonic 1320, 300 W, 28 kHz). Proteins of broken cells were then extracted with a solution containing 7 M urea, 2 M thiourea, 100 mM DTT, 4% CHAPS and 0.2% SDS for 5 h at room temperature. After centrifugation (16,100g, 5min) the supernatant was collected. The protein sample was characterized by SDS-PAGE on a 12% polyacrylamide gel and stored at -20°C.

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 4125; doi:10.3390/ijms19124125

Tryptic digestion of microsporidian protein extracts. Tris(2-carboxyethyl)phosphine (TCEP) was added to final concentration of 5mM. Samples were then heated to 55°C for 20 min, allowed to cool to room temperature and methyl methanethiosulfonate (MMTS) added to a final concentration of 10 mM. Samples were incubated at room temperature for 20 min to complete blocking of free sulfhydryl groups. Lysates were acidified with phosphoric acid to a final concentration of 1.2% and added to an S-TrapTM containing 6x lysate volume of s-trapping buffer (90% methanol, 100 mM TEAB). S-TrapTM was spun down at 4,000g for 30 seconds to remove buffer, washed with 0.2 mL of S-trapping buffer and spun again to remove all buffer. 2 µg of sequencing grade trypsin (Promega) in 125 µL of 50 mM TEAB was then added to the S-TrapTM and they were digested overnight at 37°C. After digestion the peptides were eluted from the column with subsequent applications of 50 mM TEAB, 0.2% formic acid in water and 0.2% formic acid in 50% acetonitrile. Peptides were dried in vacuo. Peptides were then reconstituted in 50 µL of 0.5M TEAB/70% isopropanol and labeled with 8-plex iTRAQ reagent for 2 hours at room temperature essentially according to [27]. Labeled samples were then acidified to pH 4 using formic acid, combined and concentrated in vacuum until ~10 µL remained [28, 29].

Mass spectrometry. An Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) equipped with a nano-ion spray source coupled to an EASY-nLC 1200 system (Thermo Scientific) was used. The LC system was configured with a self-pack PicoFrit™ 75-µm analytical column with an 8-µm emitter (New Objective, Woburn, MA) packed to 25 cm with ReproSil-Pur C18-AQ, 1.9 µM material (Dr. Maisch GmbH). Mobile phase A consisted of 2% acetonitrile; 0.1% formic acid and mobile phase B consisted of 90% acetonitrile; 0.1% formic acid. Peptides were then separated using the following steps: at a flow rate of 200 nL/min: 2% B to 6% B over 1 min, 6% B to 30% B over 84 min, 30% B to 60% B over 9 min, 60% B to 90% B over 1 min, held at 90% B for 5 min, 90% B to 50% B over 1 min and then flow rate was increased to 500 nL/min as 50% B was held for 9 min. Eluted peptides were directly electrosprayed into the Fusion Lumos mass spectrometer with the application of a distal 2.3 kV spray voltage and a capillary temperature of 300°C. Full-scan mass spectrum (Res=60,000; 400-1600 m/z) were followed by MS/MS using the "Top N" method for selection. High-energy collisional dissociation (HCD) was used with the normalized collision energy set to 35 for fragmentation, the isolation width set to 1.2 and a duration of 10 seconds was set for the dynamic exclusion with an exclusion mass width of 10 ppm. We used monoisotopic precursor selection for charge states 2+ and greater, and all data were acquired in profile mode.

Database searching to detect protein identity in microsporidian bulk protein extracts. The total of 44,547 peptides were used in the analysis. Peaklist files were generated by Mascot Distiller (Matrix Science) [30]. Protein identification and quantification was carried using Mascot 2.450 against the Uniprot_Microsporidia database (2,041 sequences; 694,360 residues). Methylthiolation of cysteine and N-terminal and lysine iTRAQ modifications were set as fixed modifications, methionine oxidation was set as variable. Trypsin was used as cleavage enzyme with two missed cleavages allowed. Mass tolerance was set at 5 ppm for intact peptide mass and 0.07 Da for fragment ions. Search results were rescored to give a final 1% FDR using a randomized version of the same Uniprot Microsporidia database.

3. Results

3.1. Massive loss of rRNA expansion segments degenerates binding sites for ribosomal proteins in microsporidian ribosomes.

We first asked how the loss of rRNA expansions could have affected rRNA-protein contacts in microsporidian ribosomes. To anwer this questions, we analyzed rRNA-protein contacts in the 80S ribosome structure from yeast *S. cerevisiae*, which are ones of the closest organisms to *Microsporidia* among species with known ribosomes structure .

As shown previously [19, 20], microsporidian rRNA lacks all major rRNA expansion segments (**Figure 1**). Mapping of rRNA deletions on the 3D ribosome structure showed that microsporidian

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 4125; doi:10.3390/ijms19124125

ribosomes additionally lack a few conserved rRNA segments, including helices h15, h16, h17 and h33 in the 18S rRNA, which are otherwise conserved in species from bacteria to eukaryotes (**Figure 1A**).

Deletions in rRNA in microsporidian ribosomes occur on the ribosome suface, where rRNA expansions extensively interact with ribosomal proteins (**Figure 1A,B**). Our analysis of rRNA-protein interactions in the 80S yeast ribosome showed that loss of rRNA expansions should eliminate at least 1/3 of the protein-ribosome interface for eight ribosomal proteins (**Figure 1B**). In one extreme case, truncation of rRNA expansions ES27L and ES31L in the 25S rRNA appears to eliminate more than 85% interface between proteins eL27 and other ribosomal components. Similarly, degeneration of helix h33 in the 18S rRNA appears to eliminate more than 80% interface between protein eS31 and other ribosomal components. And more than 50% of interface between protein eS7 and other ribosomal components appears to get lost due to degeneration of ES3S and ES6S rRNA expansions in microsporidian 18S rRNA. Given this massive loss of protein-rRNA interactions, it is plausible that some ribosomal proteins, particularly eS31 and eL27, represent stand-alone rather than ribosome-bound proteins in microsporidian species.

3.2. Proteins that appear to lose association with the ribosome also lose their sequence conservation. We next asked if the loss of rRNA expansion in *Microsporidia* is accompanied with loss of corresponding ribosomal proteins that are bound to rRNA expansions in ribosomes from non-parasitic eukaryotes. Through analysis of microsporidian genome sequences, we found that even proteins eS7, eS31, and eL27 are still retained in microsporidian species, although in some species they have undergone significant truncations (**Figure 2A,B**). This is especially prominent in protein eS31, which sequence is truncated by ~70% in some microsporidian species, including *Edhazardia aedis* and *Vittaforma corneae* (**Figure 2B**). However, despite massive truncations in ribosomal proteins in some microsporidian species, even ribosomal proteins that have lost the corresponding rRNA expansion segmets are still present in microsporidian species.

We next analyzed sequence conservation of ribosomal proteins eS7, eS31, and eL27. We anticipated that, in *Microsporidia*, these proteins should be highly variable because they are no longer involved in interactions with rRNA. Indeed, we found that these apparent free-standing ribosomal proteins have poorly conserved sequences in microsporidian species (**Figure 2A,B**). One especially notable example is protein eS31. Apart from its role in protein synthesis, eS31 has an additional function in ubiquitin biogenesis: in eukaryotes, eS31 is produced as a fusion with ubiquitin, and ubiquitin is being cleaved off from eS31 during ribosome biogenesis. We found that in *Microsporidia* the ubiquitin sequence remains more than 90% invariant in microsporidian species, whereas the eS31 sequence carries only six invariant residues, illustrating extremely poor conservation of eS31 in microsporidian species (**Figure 2B**). Notably, four of these six invariant residues represent cysteines that are critically required for eS31 folding (**Figure 2C**). Particularly, this example of eS31 conservation in microsporidian species indicates that some ribosomal proteins could have lost their original role in protein syntheis and are preserved in microsporidia for their extra-ribosomal functions.

3.3. Apparent rudimentary ribosomal proteins are abundantly expressed in microsporidian cells. We finally asked if the genes coding for degenerated ribosomal proteins represent indeed functional genes and not pseudogenes. To address this, we used mass spectrometry and analyzed protein exctracts from microsporidian *E. cuniculi* spores (**Supplementary Data S1**). We found that all the apparent free-standing proteins, including eS31, eL27, and eS7, were observed among the most frequently detected proteins in the microsporidian proteome. For instance, protein eS31 (annotated as "Similarity to monoubiquitin/carboxy-extension protein fusion" in the *Enciphalitozoon cuniculi* genome due to poor eS31 sequence conservation) was found among the top 200 most frequently detected proteins (**Supplementary Data S1**). This analysis revealed that, despite the apparent loss of ribosomal association, the ribosomal proteins are still abundantly expressed in microsporidian cells.

Peer-reviewed version available at *Int. J. Mol. Sci.* **2018**, *19*, 4125; doi:10.3390/ijms19124125

3.2. Figures

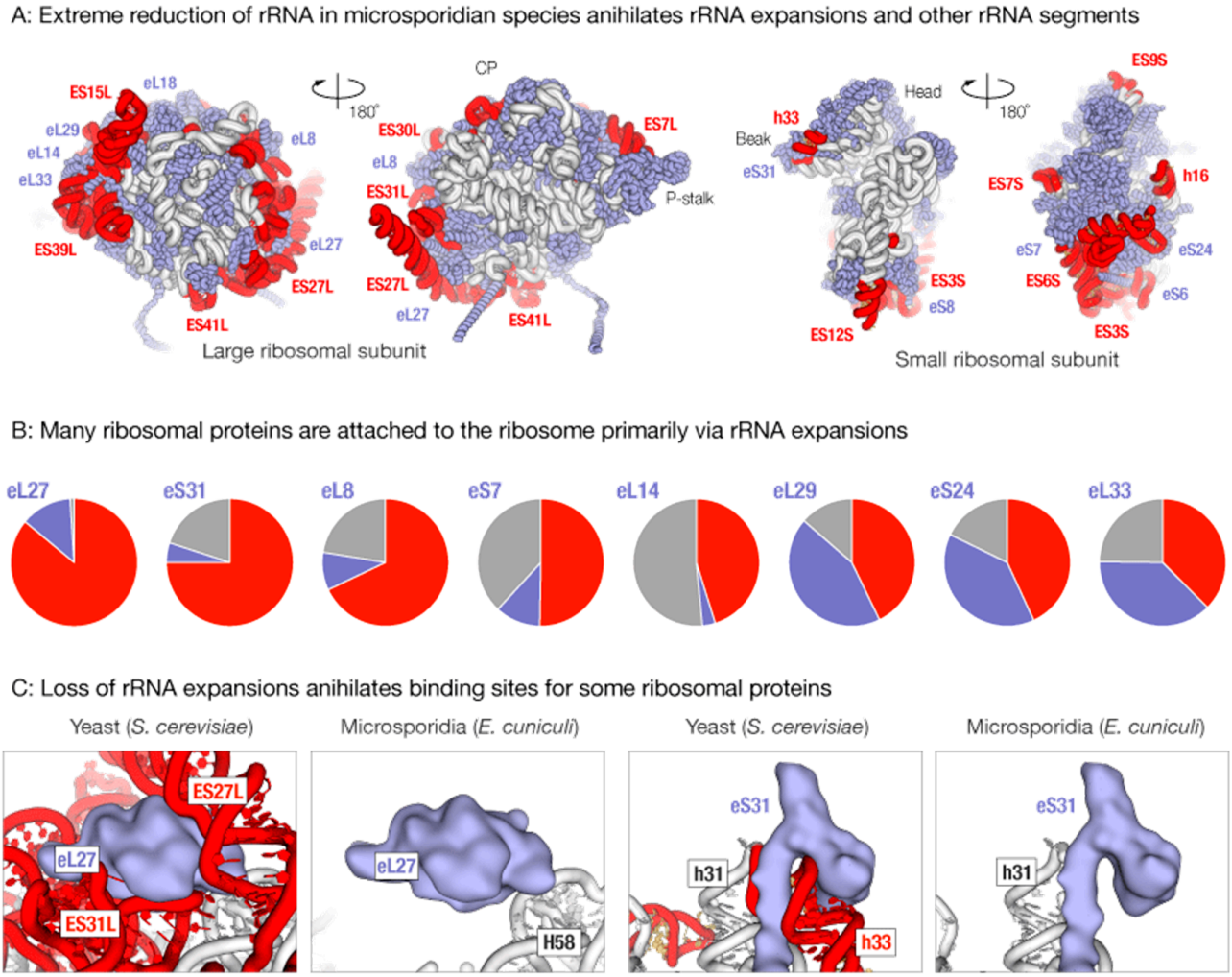

Figure 1. Extreme reduction of microsporidian rRNA appears to annihilate many rRNA-protein contacts, transforming some ribosomal proteins into ribosome-bound or free-standing rudiments with no apparent biological function. (**a**) The panel illustrates massive loss of rRNA expansion segments in microsporidian ribosomes. It shows the ribosome structure from yeast *Saccharomyces cerevisiae* – a non-parasitic fungi which is one of the closest non-parasitic relatives of Microsporidia for which there is a known ribosome structure – in which the rRNA segments that are missing in microsporidian species are highlighted in red. The rRNA segments that are conserved between the yeasts and *Microsporidia* are shown in grey, and ribosomal proteins are shown in blue. (**b**) The diagrams summarize the rRNA-protein and protein-protein interface for eight ribosomal proteins that form interact primarily with rRNA expansion segments in the 80S yeast ribosome. (**c**) The panels zoom on the structure of the 80S ribosomes to illustrate that protein eS31 is anchored to the ribosome via its contact with helix h16 in the 16S rRNA, and protein eL27 is sandwiched between the rRNA expansion segments ES27L and ES31L in the 25S rRNA. Because h16, ES27L and ES31L are fully degenerated in microsporidian rRNA, microsporidian proteins eL27 and eS31 are predicted to no longer be components of protein synthesis machinery and rather be free-standing proteins in *Microsporidia*.

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 4125; doi:10.3390/ijms19124125

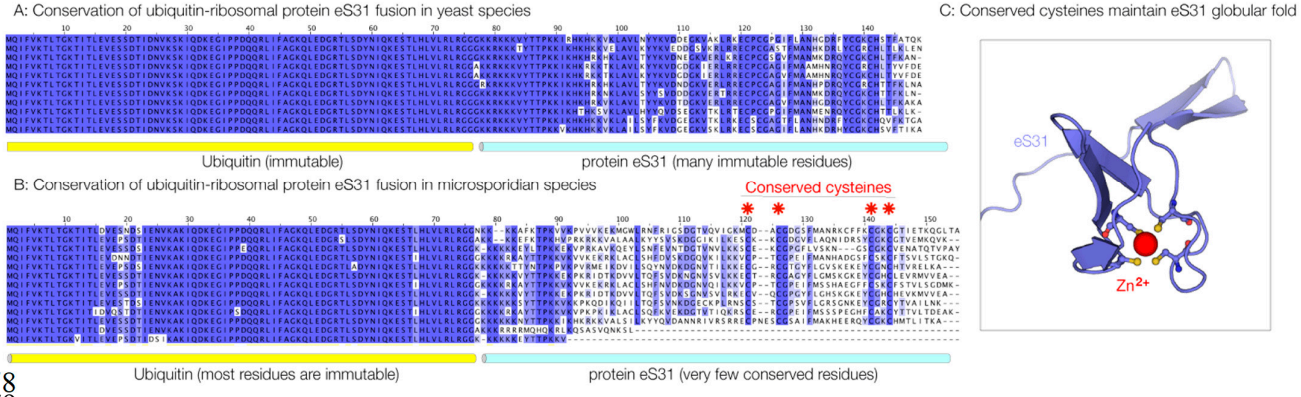

Figure 2. In parasites *Microsporidia*, ribosomal protein eS31 turns into a poorly conserved peptide with only a few residues remaining conserved due to their critical role in protein folding. (**a**) In most eukaryotes, protein eS31 is produced as a fusion with ubiquitin. During ribosome biogenesis, ubiquitin-eS31 fusion is being cleaved into ubiquitin and eS31. The panel illustrates sequence conservation of ubiquitin-eS31 in yeast species, where ubiquitin sequence remains immutable, and eS31 sequence has more than 80% identity between yeast species. (**b**) In *Microsporidia*, the ubiquitin sequence remains highly conserved, with only ~10% of sequence showing moderate variations, whereas the eS31 sequence becomes into highly variable, with only six residues being conserved between microsporidian species. Four of these residues (all cysteines) coordinate a zinc ion in the middle of eS31 globular domain. (**c**) Crystal structure of eS31 protein shows that the folding of eS31 critically depends on the presence of a zinc-finger motif, which is typically present in structure of small proteins where the globular domain is too small to form stable globule via hydrophobic interactions. The fact that the cysteine residues remain among a very few immutable residues in microsporidian eS31 indicating the presence of selective pressure to maintain the eS31 folding intact. High conservation of these cysteine residues in *Microsporidia* suggests that there is selective pressure to keep eS31 properly folded, possibly to avoid formation of protein aggregates that would intoxicate microsporidian cells or would compromise ubiquitin biosynthesis. It appears from this example that, despite the fact that S31 have lost its original role in protein synthesis, it is still retained in the microsporidian proteome for its extraribosomal role in ubiquitin biogenesis.

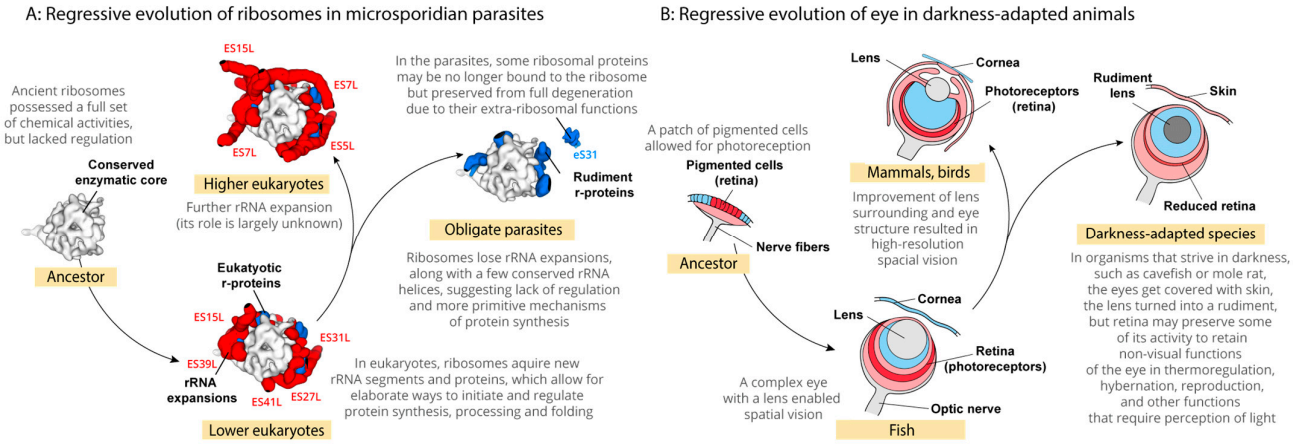

Figure 3. Degeneration of microsporidian ribosomes illustrates a remarkable similarity of regressive evolution at the scale of individual proteins and nucleic acids, and at the scale of organs of animal body. (a) Our study suggests that upon transition to parasitic lifestyle, microsporidian ribosomes not only have lost their rRNA expansions, but also transformed several ribosomal proteins from active players of ribosome assembly into dysfunctional rudiments. For instance, in non-parasitic eukaryotes, many ribosomal proteins are buried in the ribosome interior where they stabilized interactions between the ancient core of the ribosome and rRNA expansions. In *Microsporidia*, these proteins (i.e. eS6, eS8, eS24, eL8, eL14, eL18, eL39) are expected to be exposed on the ribosomal surface where they have no apparent biological function. Other proteins, such as eS31, eL27 and, possibly, eS7 appear to become free-standing proteins that are no longer associated with the ribosome in

Microsporidia. However, eS31, eL27, and eS7 proteins are not eliminated from microsporidian proteomes, suggesting these proteins may be retained in *Microsporidia* for their extra-ribosomal functions. (**b**) A schematic path of eye evolution shows that similar regressive changes occur during animal adaptation to darkness as observed in cavefish and mole rats. Importantly, although the eye of darkness-adapted animals is blind and is filled with many rudiments (e.g. the non-functional degenerated lens), it retains the light-sensitive retina that fulfils non-visual functions of light perception to control animal hibernation, mating and thermoregulation [34, 35].

4. Discussion

In the present study we explore an apparent effect of Muller's ratchet on the structure of microsporidian ribosomes. Particularly, we analyze ribosomal proteins that stabilize and coordinate activity of rRNA expansions in eukaryotic ribosomes. We show that massive loss of rRNA expansion in microsporidian ribosomes appears to annihilate the binding sites for a few ribosomal proteins (eS7, eS31, and eL27) suggesting these proteins are no longer part of the protein synthesis machinery, but rather free-standing proteins that are retained in microsporidian species for extra-ribosomal functions, or simply awaiting their total extinction forced by genome decay.

Our study illustrates that automated analysis of parasitic genomes may significantly underestimate the extent of genome decay in parasitic species. We show here that some components of a parasitic cell that are 'defined' by automated annotations as functional constituents of the protein synthesis machinery may in fact represent rudimentary proteins that are retained in parasites for their non-ribosomal function, as is the case of microsporidian protein eS31 that is involved in ubiquitin biogenesis. Thus, our study illustrates the presence of evolutionary intermediates between the genes and pseudogenes in parasitic genomes. These genes are not yet pseudogenes because they are actively expressed into proteins; yet the proteins produced from these genes are not fully active functionally as they lack at least one of their original biological activities in non-parasitic species.

Regarding the ribosome function, it is pertinent to mention that the case of *Microsporidia* shows that the eukaryotic ribosome can function *in vivo* in complete absence of rRNA expansion segments. Furthermore, our study illustrates that microsporidian ribosomes not only lack rRNA expansion segments, but a few rRNA segments that are conserved between eukaryotes and bacteria, particularly helix h16 in the 18S rRNA. During protein synthesis, helix h16, among other factors, is thought to promote ribosomes self-organization into polysomes, in which each individual ribosome is properly spaced and oriented to allow access of regulatory proteins and factors of co-translational protein processing and processing. In eukaryotes, helix h16 also participates in the initiation of protein synthesis by helping the ribosome to find the initiation codon in the messenger RNA [31-33]. It is tempting to imagine that lack of this helix in microsporidian ribosomes leads to a less organized structure of polysomes and more primitive forms of protein synthesis initiation, or even occasional codon recognition errors during the start-codon search in messenger RNA.

Curiously, regressive changes in microsporidian rRNA and ribosomal proteins strikingly resemble regressive evolution in the macroscopic world (**Figure 3A, B**). For instance, in a classical study of eye degeneration blind mole rats *Spalax ehrebegi*, extreme regression of eye was shown to eliminate many anatomical features required for spatial vision, including the eye to past of a primitive eye, which lacks the capacity of focused vision, including transparent cornea and functional lens (**Figure 3B**) [34, 35]. In mole rats, cornea is replaced by a layer of epithelium that buries the eye under the skin, and functional lens is degenerated into a non-functional rudiment in the middle of the eye structure (**Figure 3B**). And yet not all of the eye functions are lost in mole rats: at the bottom of the eye ball there is still a layer of functional photosensitive cells of retina. These cells allow blind mole rates to perceive light and thereby regulate such vital aspects of animal activity as hibernation, thermoregulation, and reproduction. The presence of light-sensitive retina in otherwise degenerated eye of mole rats made these animals a unique model to test non-visual functions of eye, suggesting

that microsporidian parasites may potentially turn into a unique model to study extra-ribosomal functions of ribosomal proteins.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: Mass spectrometry analysis of bulk protein extracts from spores of human pathogen *Encephalitozoon cuniculi*.

Author Contributions: S.V.M. and D.S. devised the study. S.V.M., K.M. performed analysis of ribosome sequences and structures, K.R., A.M. performed analysis of mass spectrometry and identified the spores, S.V.M. and D.S. analyzed the data and wrote the manuscript.

Funding: This study was funded by the National Institute of General Medical Sciences (NIGMS) at the National Institutes of Health, grant number R35GM122560 and by the U.S. Department of Energy (DOE), grant number DE-FG02-98ER20311 (both to D.S.). This work was performed with assistance from the Cold Spring Harbor Laboratory Mass Spectrometry Shared Resource, which is supported by the National Cancer Institute Cancer Center Support Grant 5P30CA045508 (to D.J.P.).

Acknowledgments: We would like to thank Dr. Catherine Texier (France) and Dr. Louis M. Weiss for providing us the samples of microsporidian protein extracts. We also thank the members of Dieter Söll lab for valuable discussions and critical feedback during preparation of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Moran, N. A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **2002**, *108*:583-586.
2. McCutcheon, J. P., and N. A. Moran. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **2011**, *10*:13-26.
3. Andersson, S. G., and C. G. Kurland. Reductive evolution of resident genomes. *Trends Microbiol* **1998**, *6*:263-268.
4. Muller, H. J. The relation of recombination to mutational advance. *Mutat Res* **1964**, *106*:2-9.
5. Moran, N. A., and G. M. Bennett. The tiniest tiny genomes. *Annu Rev Microbiol* **2014**, *68*:195-215.
6. Moran, N. A. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* **1996**, *93*:2873-2878.
7. Melnikov, S. V., K. D. Rivera, D. Ostapenko, A. Makarenko, N. D. Sanscrainte, J. J. Becnel, M. J. Solomon, C. Texier, D. J. Pappin, and D. Soll. Error-prone protein synthesis in parasites with the smallest eukaryotic genome. *Proc Natl Acad Sci U S A* **2018**, *115*:E6245-E6253.
8. Gill, E. E., and N. M. Fast. Stripped-down DNA repair in a highly reduced parasite. *BMC Mol Biol* **2007**, *8*:24.
9. Shonhai, A., A. G. Maier, J. M. Przyborski, and G. L. Blatch. Intracellular protozoan parasites of humans: the role of molecular chaperones in development and pathogenesis. *Protein Pept Lett* **2011**, *18*:143-157.
10. Wust, P., B. Hildebrandt, G. Sreenivasa, B. Rau, J. Gellermann, H. Riess, R. Felix, and P. M. Schlag. Hyperthermia in combined treatment of cancer. *Lancet Oncol* **2002**, *3*:487-497.
11. Lambert, J. D., and N. A. Moran. Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* **1998**, *95*:4458-4462.
12. Escarmis, C., C. Perales, and E. Domingo. Biological effect of muller's ratchet: distant capsid site can affect picornavirus protein processing. *J Virol* **2009**, *83*:6748-6756.
13. Fares, M. A., A. Moya, and E. Barrio. Groel and the maintenance of bacterial endosymbiosis. *Trends Genet* **2004**, *20*:413-416.
14. Katinka, M. D., S. Duprat, E. Cornillot, G. Metenier, F. Thomarat, G. Prensier, V. Barbe, E. Peyretaillade, P. Brottier, P. Wincker, F. Delbac, H. El Alaoui, P. Peyret, W. Saurin, M. Gouy, J. Weissenbach, and C. P. Vivares. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **2001**, *414*:450-453.
15. Peyretaillade, E., H. El Alaoui, M. Diogon, V. Polonais, N. Parisot, D. G. Biron, P. Peyret, and F. Delbac. Extreme reduction and compaction of microsporidian genomes. *Res Microbiol* **2011**, *162*:598-606.
16. Didier, E. S., and L. M. Weiss. Microsporidiosis: not just in AIDs patients. *Curr Opin Infect Dis* **2011**, *24*:490-495.
17 NIH: National Institute of Allergy and Infectious Diseases. Available online: https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens (accessed on 17 November, 2018)
18. Melnikov, S., A. Ben-Shem, N. Garreau de Loubresse, L. Jenner, G. Yusupova, and M. Yusupov. One core, two shells: Bacterial and eukaryotic ribosomes. *Nat Struct Mol Biol* **2012**, *19*:560-567.
19. Peyretaillade, E., C. Biderre, P. Peyret, F. Duffieux, G. Metenier, M. Gouy, B. Michot, and C. P. Vivares. Microsporidian encephalitozoon cuniculi, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a lsu rrna reduced to the universal core. *Nucleic Acids Res* **1998**, *26*:3513-3520.
20. Vossbrinck, C. R., J. V. Maddox, S. Friedman, B. A. Debrunner-Vossbrinck, and C. R. Woese. Ribosomal rna sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* **1987**, *326*:411-414.
21. Weiss, L. M., X. Zhu, A. Cali, H. B. Tanowitz, and M. Wittner. Utility of microsporidian rRNA in diagnosis and phylogeny: A review. *Folia Parasitol (Praha)* **1994**, *41*:81-90.
22. Dong, S., Z. Shen, L. Xu, and F. Zhu. Sequence and phylogenetic analysis of SSU rRNA gene of five *Microsporidia*. *Curr Microbiol* **2010**, *60*:30-37.
23. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res* **2015**, *43*:W314-319.
24. Sievers, F., and D. G. Higgins. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* **2014**, *1079*:105-116.
25. Brosson, D., L. Kuhn, F. Delbac, J. Garin, P. Vivares C, and C. Texier. Proteomic analysis of the eukaryotic parasite *Encephalitozoon cuniculi* (*Microsporidia*): A reference map for proteins expressed in late sporogonial stages. *Proteomics* **2006**, *6*:3625-3635.
26. Delbac, F., F. Duffieux, D. David, G. Metenier, and C. P. Vivares. Immunocytochemical identification of spore proteins in two *Microsporidia*, with emphasis on extrusion apparatus. *J Eukaryot Microbiol* **1998**, *45*:224-231.

27. Ross, P. L., Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, and D. J. Pappin. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004**, *3*:1154-1169.

28. Gilar, M., P. Olivova, A. E. Daly, and J. C. Gebler. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci* **2005**, *28*:1694-1703.

29. Wang, Y., F. Yang, M. A. Gritsenko, Y. Wang, T. Clauss, T. Liu, Y. Shen, M. E. Monroe, D. Lopez-Ferrer, T. Reno, R. J. Moore, R. L. Klemke, D. G. Camp, 2nd, and R. D. Smith. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human Mcf10a cells. *Proteomics* **2011**, *11*:2019-2026.

30. Perkins, D. N., D. J. Pappin, D. M. Creasy, and J. S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*:3551-3567.

31. Passmore, L. A., T. M. Schmeing, D. Maag, D. J. Applefield, M. G. Acker, M. A. Algire, J. R. Lorsch, and V. Ramakrishnan. The eukaryotic translation initiation factors eIF1 and eIF1a induce an open conformation of the 40S ribosome. *Mol Cell* **2007**, *26*:41-50.

32. Pisareva, V. P., A. V. Pisarev, A. A. Komar, C. U. Hellen, and T. V. Pestova. Translation initiation on mammalian mRNAs with structured 5'utrs requires dexh-box protein Dhx29. *Cell* **2008**, *135*:1237-1250.

33. Myasnikov, A. G., Z. A. Afonina, J. F. Menetret, V. A. Shirokov, A. S. Spirin, and B. P. Klaholz. The molecular structure of the left-handed supra-molecular helix of eukaryotic polyribosomes. *Nat Commun* **2014**, *5*:5294.

34. Sanyal, S., H. G. Jansen, W. J. de Grip, E. Nevo, and W. W. de Jong. The eye of the blind mole rat, *Spalax ehrenbergi*. Rudiment with hidden function? *Invest Ophthalmol Vis Sci* **1990**, *31*:1398-1404.

35. de Jong, W. W., W. Hendriks, S. Sanyal, and E. Nevo. The eye of the blind mole rat (*Spalax ehrenbergi*): Regressive evolution at the molecular level. *Prog Clin Biol Res* **1990**, *335*:383-395.