Average oxidation state of carbon in proteins

Jeffrey M. Dick[1]

[1]Department of Chemistry and Department of Applied Geology, Curtin University, Perth, WA, Australia
email: j3ffdick@gmail.com

Summary

The degree of oxidation of carbon atoms in organic molecules depends on the covalent structure. In proteins, the average oxidation state of carbon (Z_C) can be calculated as an elemental ratio from the chemical formula. To investigate oxidation-reduction (redox) patterns, groups of proteins from different subcellular locations and phylogenetic divisions were selected for comparison. Extracellular proteins of yeast have a relatively high oxidation state of carbon, corresponding with oxidizing conditions outside of the cell. However, an inverse relationship between Z_C and redox potential occurs between the endoplasmic reticulum and cytoplasm; this trend is interpreted as resulting from overall coupling of protein turnover to the formation of a lower glutathione redox potential in the cytoplasm. In Rubisco homologues, lower Z_C tends to occur in organisms with higher optimal growth temperature, and there are broad changes in Z_C in whole-genome protein compositions in microbes from different environments. Energetic costs calculated from thermodynamic models suggest that thermophilic organisms exhibit molecular adaptation to not only high temperature but also the reducing nature of many hydrothermal fluids. A view of protein metabolism that depends on the chemical conditions of cells and environments raises new questions linking biochemical processes to changes on evolutionary timescales.

Keywords: oxidation state, redox potential, subcellular location, protein metabolism, protein evolution

1 Introduction

Chemical reactions involving the transfer of electrons, known as oxidation-reduction or redox reactions, are ubiquitous in cellular and environmental systems [1, 2]. In the cell, the oxidation of thiol groups in proteins to form disulfides has the potential to regulate (activate or inhibit) enzymatic function [3]. Because these reactions are reversible on short timescales, a regulatory network known as redox signalling is made possible by reactions of small-molecule metabolites, including glutathione (GSH) and reactive oxygen species [4]. On timescales of metabolism, complex oxidation-reduction reactions are required for the formation (anabolism) and degradation (catabolism) of proteins and other biomolecules. Although many individual steps in biomass synthesis are irreversible, much biomass is ultimately recycled through endogenous metabolism [5]. On longer timescales, forces outside of individual cells and organisms sustain the redox disequilibria between inorganic and/or organic species that provide the energy source for metabolisms suited to a multitude of environments [6]. In turn, the actions of organisms can alter the redox conditions on Earth; the oxygenation of the atmosphere and oceans over geological time has a biogenic origin, and changed the course of later biological evolution [7].

Through evolution, the sequences of genes, and their protein products, are progressively altered. The elemental stoichiometry (chemical formula) and standard Gibbs energy of the molecules have a primary impact on metabolic requirements for energy and elemental resources. The energetic cost for synthesis of biomass is a function not only of the composition of the biomass, but also of environmental parameters including temperature and the concentrations of metabolic precursors. Temperature and oxidation-reduction potential have profound effects on the relative energetic costs of formation of different amino acids [8] or proteins [9]. These energetic costs are sensitive to differences in the elemental compositions of biomolecules. To a first approximation, a shift to a more reducing environment alters the energetics of reactions in a direction that favours the formation of relatively reduced chemical compounds. In a field test of this principle, metagenomic sequences for the most highly reduced proteins were found in the hottest and most reducing zones of a hot spring [10, 11].

The purpose of this study is to investigate a particular stoichiometric quantity, the average oxidation state of carbon (Z_C, defined below), as a comparative tool for identifying compositional patterns at different levels of biological organization. By comparing a quantity derived from the elemental compositions of proteins, this study addresses one aspect of biochemical evolution. However, the questions raised here differ in important respects from conventionally defined biochemistry and evolutionary biology. Biochemical studies are most often concerned with the functions of molecules [12], including enzymatic catalysis and non-covalent interactions involved in the structural conformation of proteins and binding of ligands. Studies in molecular evolution often place emphasis on the historical relationships between sequences, but not their physical properties [12]. Combining these viewpoints, most current work assumes that structural stability of proteins is the primary criterion for molecular adaptation to high temperature [13]. In contrast, in this study, more attention is given to the stoichiometric and energetic demands of the reactions leading to protein formation. Because material replacement of proteins depends on metabolic outputs [14], the compositional differences among proteins have significant consequences for cellular organization and metabolism.

The following questions have been identified: 1) How does the relationship between Z_C of amino acids and corresponding codons relate to the origin or form of the genetic code? 2) How do the differences in Z_C between membrane proteins and others compare with properties of amino acids, e.g. hydrophobicity, known to favour localization to membranes? 3) How are the differences in Z_C of proteins among eukaryotic subcellular compartments related to differences in redox potential? 4) How are the differences in Z_C in families of redox-active proteins related to the standard reduction potentials of the proteins? 5) How are the differences in Z_C among different organisms, both in terms of bulk (genome-derived) protein composition and for homologues of a single family (Rubisco), related to environmental conditions, especially temperature and redox potential?

In the Results each problem is briefly introduced, the empirical distribution of Z_C is described, and a discussion is developed to explore how the patterns reflect biochemical and evolutionary constraints. These short discussions, corresponding to topics (1)-(4), should be regarded as preliminary, and probably incomplete interpretations. The discussions are limited because there is no complete conceptual framework that links the biochemical reactions with the evolutionary processes that are implicit in all of the comparisons. The final section of the Results goes into more detail for the phylogenetic comparisons (topic 5) by examining the relative Gibbs energies of formation of proteins in environments of differing redox potential.

2 Methods

Throughout this study, "reducing" and "oxidizing" are used in reference to oxidation-reduction potential, tied to a particular redox couple or to environmental conditions, often expressed as millivolts on the Eh scale. "Reduced" and "oxidized" are used to refer to variations in the oxidation state of carbon.

The formal oxidation state of a carbon atom in an organic molecule can be calculated by counting -1 for each carbon-hydrogen bond, 0 for each carbon-carbon bond, and -1 for each bond of carbon to O, N or S [8, 15]. In photosynthesizing organisms, and autotrophs in general, the carbon source is CO_2, having the highest oxidation state of carbon (+4). The products of photosynthetic reactions include proteins and other biomolecules with a lower oxidation state of carbon. Even if the molecular structure is unknown, analytical elemental compositions can be used in calculations of the average oxidation state of carbon in biomass [16, 17]. Because any gene or protein sequence corresponds to a definite canonical (nonionized, unphosphorylated) chemical formula, the average oxidation state of carbon in these biomolecules is easily calculated.

In amino acids and proteins, the average oxidation state of carbon (Z_C) can be calculated using

$$Z_C = \frac{-n_H + 3n_N + 2n_O + 2n_S + Z}{n_C}, \quad (1)$$

where Z is the charge on the molecule, and n_C, n_H, n_N, n_O and n_S are the numbers of the subscripted elements in the chemical formula of the molecule. The coefficients on the terms in the numerator derive from formal charges of atoms other than C, as follows: H (+1), N (-3), O (-2), S (-2). Negative formal charges reflect greater electronegativities of these elements compared to carbon. If two thiol groups react to form a disulfide bond, the oxidation states of the two affected sulfur atoms change from -2 to -1. Although H_2 is produced in this reaction, the oxidation state of carbon in the protein remains constant. It follows that equation (1) is applicable only to chemical formulas of proteins in which the N, O, and S are all fully reduced (bonded only to H and/or C).

The Z in equation (1) ensures that ionization by gain or loss of a proton, having an equal effect on Z and n_H, does not change the Z_C. Likewise, gain or loss of H_2O, which affects equally the values $2n_H$ and n_O, does not alter the average oxidation state of carbon [15]. Accordingly, the Z_C of a peptide formed by polymerization of amino acids (a dehydration reaction) is a weighted average of the Z_C in the amino acids, where the weights are the number of carbon atoms in each amino acid. As an example, the Z_C of hen egg white lysozyme, having a chemical formula of $C_{613}H_{959}N_{193}O_{185}S_{10}$, is 0.016. This protein is oxidized compared to many other proteins, which commonly have negative values of Z_C.

To aid in reproducibility, data files of protein sequences or amino acid composition, except large files available from public databases, and computer program files for the calculations are provided in the Supporting Information. The calculations and figures were generated using the R software environment [18] together with the CHNOSZ package [19].

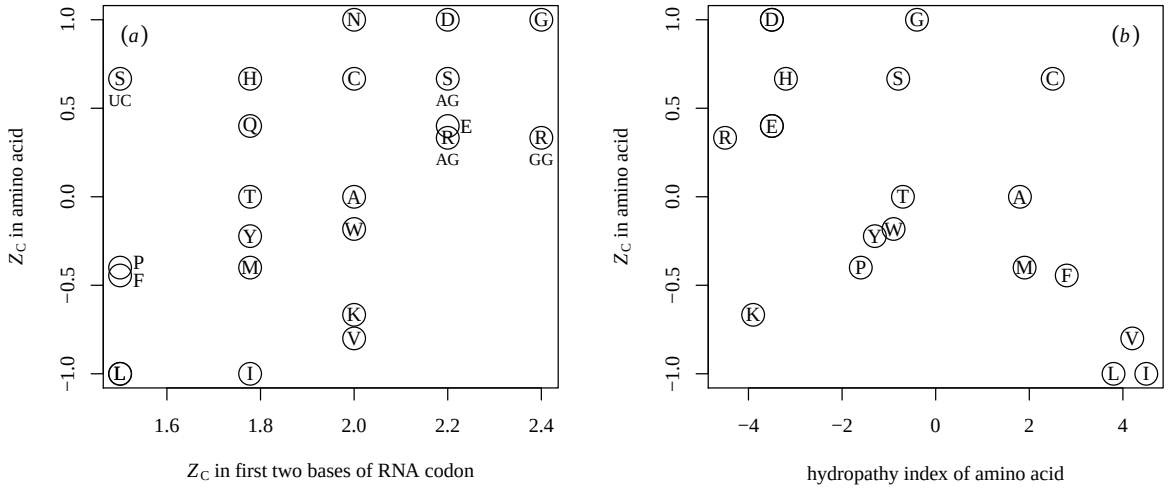

Figure 1: Average oxidation state of carbon (Z_C) in amino acids compared with (*a*) Z_C in first two bases of the corresponding RNA codons and (*b*) hydropathy index of the amino acids taken from Kyte and Doolittle, 1982 [24]. Standard one-letter abbreviations for the amino acids are used to identify the points. In (*a*), the different codon compositions for serine (S) and arginine (R) are indicated by letters below the symbols, and some amino acid labels are shifted for readability. In (*b*), labels for asparagine (N) and glutamine (Q) are omitted for clarity; they plot at the same positions as aspartic acid (D) and glutamic acid (E), respectively.

3 Results

3.1 Comparison of Z_C of amino acids with hydropathy and properties of codons

In contemplating the ancient origin of the genetic code, the chemical similarities of respective codons and amino acids have been used to argue for coevolution (shared biosynthetic pathways) [20] or a tendency toward similar physicochemical properties. The possible advantages that were identified for similar physicochemical properties include enhancing the steric interactions between amino acids and codons [21], or increasing the similarity between different amino acids resulting from a single DNA base mutation in order to maintain protein structure [20].

In the genetic code, the first two bases (a "doublet") are more indicative of the amino acid than the third position of the codon [21]. The Z_C of amino acids are compared with the values calculated for the corresponding RNA nucleobase doublets in figure 1*a*. Some of the doublets, e.g. UU (phenylalanine, leucine), CU (leucine), UC (serine) and CC (proline) have identical Z_C (in this case, 1.5), leading to only 5 possible values of Z_C for the doublets. The overall relationship suggested by figure 1*a* is loose correlation between the Z_C of amino acids and of the RNA doublets. The most highly reduced amino acids, leucine (L) and isoleucine (I), are coded for by doublets having the two lowest Z_C values.

The increase in Z_C going from leucine to alanine to glycine (figure 1) is reflected in the metastability fields of these amino acids, which occur in order of increasing oxidation potential, or oxygen fugacity [22]. Metastable equilibrium refers to the equalization of the energies of reactions to form the amino acids; it is a partial equilibrium because the amino acids generally remain unstable with respect to inorganic species. Likewise, the relative Gibbs energies of formation reactions of amino acids differ considerably between hydrothermal (hot, reducing) and surface (cool, oxidizing) environments [8]. These patterns

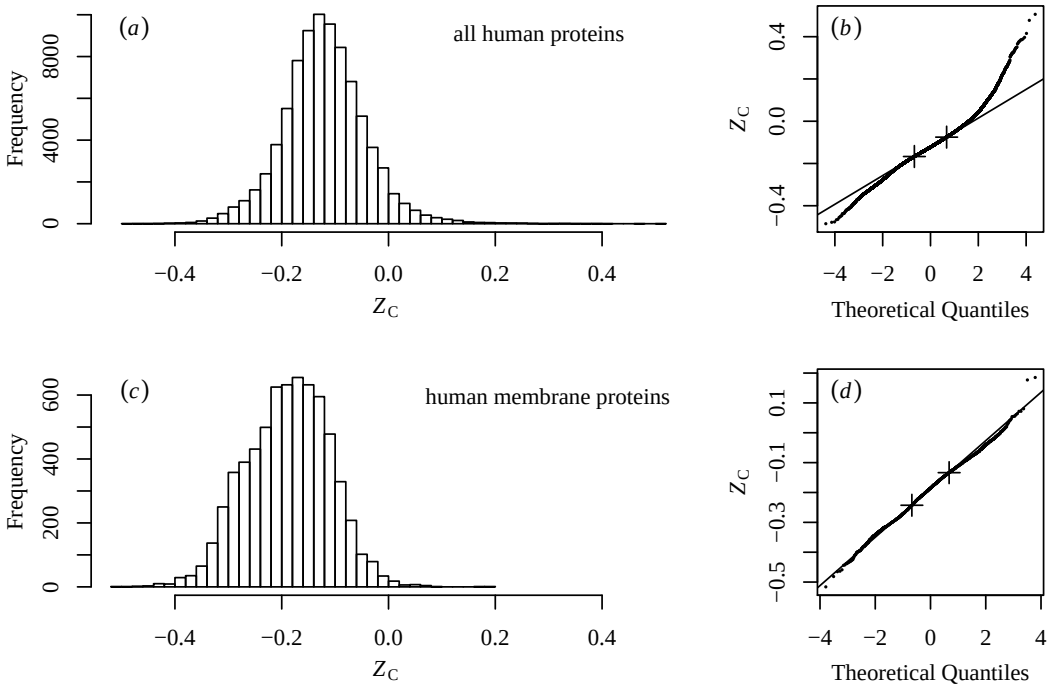

Figure 2: Average oxidation state of carbon shown in histograms and normal probability plots for (*a*-*b*) all human proteins and (*c*-*d*) human membrane proteins. Only proteins of sequence length greater than or equal to 50 amino acids are considered. In the normal probability plots the lines are drawn through the 1st and 3rd quartiles, indicated by the crosses.

support the possibility of metastable equilibria in hydrothermal environments between amino acids and nucleobase sequences that are paired in the genetic code. Tests of the potential for these states, carried out using Gibbs energies available at high temperature [9, 23], could reveal thermodynamic constraints on the energetics of abiotic or early biosynthesis independent of the arguments based on similarities in protein structure and biosynthetic pathways [20, 21].

The hydropathy index, based on the relative hydrophobicity and hydrophilicity of amino acids [24], is commonly used for identifying probable membrane-spanning domains of proteins. In figure 1*b*, Z_C is compared with the hydropathy values for individual amino acids. The three most hydropathic amino acids, isoleucine, leucine and valine, are also the three with the lowest Z_C. Therefore, membrane proteins with hydrophobic domains are likely to be more reduced than other proteins. The following sections examine the actual differences in human and yeast proteins.

3.2 Differences in Z_C of membrane proteins

The lipid (fatty acid) components of membranes are reduced relative to many other biomolecules, including amino acids, nucleotides, and saccharides (see figure 1 of Amend *et al.*, 2013 [25]). Proteins that are embedded in membranes tend to contain more hydrophobic amino acids, which enhance solubility of proteins in the membrane environment [24] and generally are relatively reduced (figure 1*b*).

To compare membrane proteins with other human proteins, sequences for all human proteins were taken from the UniProt database [26], and sequences for predicted membrane proteins were taken from all FASTA sequence files provided in Additional File 2 of Almén *et al.*, 2009 [27]. Only sequences at least 50 amino acids in length were considered. The distribution of Z_C of all human proteins (figure 2*a;*

n=83994) is centred on -0.123 (median), -0.120 (mean). In the Z_C of human membrane proteins, the distribution is shifted to lower values (figure 2*c*; n=6627, -0.186 median, -0.189 mean). The mean value is lower for membrane proteins than for all human proteins (Student's t-test: $p < 2.2 \times 10^{-16}$). Thus, the proteins located in the membranes are, on average, more reduced than other proteins in humans. It is tempting to speculate that the coexistence of reduced proteins with other relatively reduced biomolecules (lipids) reflects a compositional similarity that would contribute to energy optimization if metabolic pathways for proteins and lipids were operating under common redox potential conditions.

The observed distributions of Z_C are each compared with a normal distribution in normal probability plots (theoretical quantile-quantile (Q-Q) plots) in figure 2*b*,*d*. The steeper trends in the low- and high-quantile range of figure 2*b* indicate that the distribution of Z_C of human proteins has relatively long tails, especially at high Z_C, compared to a normal distribution. Although an asymmetry is apparent in the uneven shape of the histogram in figure 2*c* and the wiggles in figure 2*d*, the overall distribution of Z_C of the membrane proteins more closely resembles a normal distribution. Comparisons with normal distributions have implications, through the central limit theorem [28], for assessing the impact of many small-scale, independent effects in evolution on the chemical composition of organisms or their components. This theme, however, is not developed further here; instead, the overall differences in Z_C of proteins in subcellular compartments are considered next.

3.3 Subcellular differences in Z_C of proteins and comparison with subcellular redox potential

For some model organisms, including *Saccharomyces cerevisiae* (yeast), the identities of proteins associated with subcellular compartments are now available in databases. Here, calculations of Z_C of proteins and a comparison with independent measurements of redox potential are used to investigate the oxidation-reduction features and dynamics of cellular structure.

In a previous study, the limiting conditions for chemical transformations among proteins in subcellular compartments were quantified theoretically as a function of redox potential and hydration state [29]. In that study, the locations of proteins were taken from the the "YeastGFP" study of Huh *et al.*, 2003 [30]. That dataset has the advantage that relative abundances of many of the proteins are available, but it is limited to 23 named locations in the cell. In order to consider more cellular components (including the membranes), a more extensive reference proteome is used in this study. This proteome is based on the current *Saccharomyces* Genome Database (SGD) [31] annotations combined with the Gene Ontology (GO) [32] vocabulary for the "cellular component" aspect, which describes many organelles and membranes within the cell. Major cell components were selected for comparison, and the Z_C was calculated for protein products of the genes, as summarized in table 1. The median values are also portrayed in the drawing of a yeast cell in figure 3.

It is apparent that the membrane proteins are highly reduced. However, not all membranes are equal; the proteins in the nuclear and inner and outer mitochondrial membranes are less reduced than those in the plasma membrane, and the endoplasmic reticulum (ER) membrane has very highly reduced proteins. Among the organellar proteins considered, the ER has the most reduced proteins, followed by vacuoles and mitochondria; the cytoplasmic proteins are moderately reduced. The proteins in the nucleus, bud neck and bud are more oxidized than in the other compartments. The most oxidized proteins in the system are the extracellular ones. The relative Z_C of the proteins in the ER, mitochondrion, cytoplasm, and nucleus are consistent with the previous study in which the calculations took account of the abundances of proteins [29].

For comparison with Z_C of proteins, the values of subcellular redox potential (Eh, in mV) listed in table 2 were compiled from literature sources [1, 33–40]. Measurements of reduced and oxidized glutathione

Table 1: Summary of Z_C of proteins in subcellular locations of yeast. Numbers of proteins (n) in SGD associated with the indicated GO terms are listed. The numerators of the fractions denote membrane-associated proteins that are also listed as "integral to membrane" (GO:0016021); only these proteins were used in the calculations of Z_C.

cellular component	GO term	n	median Z_C	mean Z_C
cytoplasm	GO:0005737	2245	-0.136	-0.127
nucleus	GO:0005634	2073	-0.129	-0.121
mitochondrion	GO:0005739	1077	-0.164	-0.159
endoplasmic reticulum	GO:0005783	435	-0.191	-0.192
nucleolus	GO:0005730	263	-0.137	-0.128
Golgi apparatus	GO:0005794	215	-0.160	-0.167
cellular bud neck	GO:0005935	153	-0.111	-0.108
vacuole	GO:0005773	175	-0.164	-0.163
extracellular region	GO:0005576	95	-0.096	-0.098
cellular bud tip	GO:0005934	96	-0.113	-0.110
endoplasmic reticulum membrane	GO:0005789	283/338	-0.209	-0.211
plasma membrane	GO:0005886	224/427	-0.200	-0.188
mitochondrial inner membrane	GO:0005743	143/218	-0.184	-0.184
vacuolar membrane	GO:0005774	100/145	-0.205	-0.196
Golgi membrane	GO:0000139	76/121	-0.203	-0.200
nuclear membrane	GO:0031965	54/67	-0.161	-0.139
mitochondrial outer membrane	GO:0005741	51/92	-0.176	-0.177

(GSH and GSSG) in whole-cell extracts have been interpreted as reflecting cytoplasmic redox potential, but redox-sensitive green fluorescent protein (roGFP) probes [33] provide more specific data for subcellular locations. The data are not in all cases acquired from yeast, but it has been noted that cytoplasmic Eh values based on roGFP are similar for different model organisms [36].

The redox potentials in the vacuole and extracellular space are less well constrained than other locations. Under stress response, high amounts of GSSG, but not GSH, are sequestered in vacuoles [41]. A conservative lower range for the Eh of vacuoles (-160 to -130 mV) was calculated by taking a value of 80% GSSG and computing Eh from the GSH-GSSG equilibrium at concentrations of 1–10 mM GSH (see equation 21 and figure 4 of Ref. [35]). The redox potential would be higher if the GSSG/GSH ratio were in fact greater than 80/20. A high redox potential is also implicated by the presence of ferric iron species in vacuoles [42]. Extracellular redox state can vary greatly, but in aerobic organisms and laboratory culture it is likely to be generally oxidizing compared to subcellular compartments.

The values in table 2 are not comprehensive, and should be taken as a rough guide, but even with the uncertainties, comparison with the interquartile range of Z_C of the proteins reveals some trends (figure 4*a*). The difference in both Z_C and Eh is positive going from any subcellular compartment, except for vacuoles, to the extracellular space. This pattern has an intuitive explanation: by evolutionary adjustment to optimize proteins for their environment, the inside of the cell, which is more reducing, would be expected to have more reduced proteins compared to the outside.

The more surprising trend in figure 4*a* is an inverse relationship between Z_C and Eh of the cytoplasm and ER. Does this contrast have any biochemical significance? The ER is a component of the secretory pathway, which transports proteins to membranes and to outside the cell [38]. Let us conjecture that the populations of proteins in the ER and cytoplasm are connected through common metabolic intermediates – their formation and degradation are part of the recycling of biomass through endogenous metabolism [5], also implied by metabolic closure [14]. It follows that the formation of proteins of a higher Z_C in the cytoplasm entails the loss of electrons from proteins into metabolic pathways. Perhaps these pathways ultimately transfer these electrons to the formation of GSH in the cytoplasm.

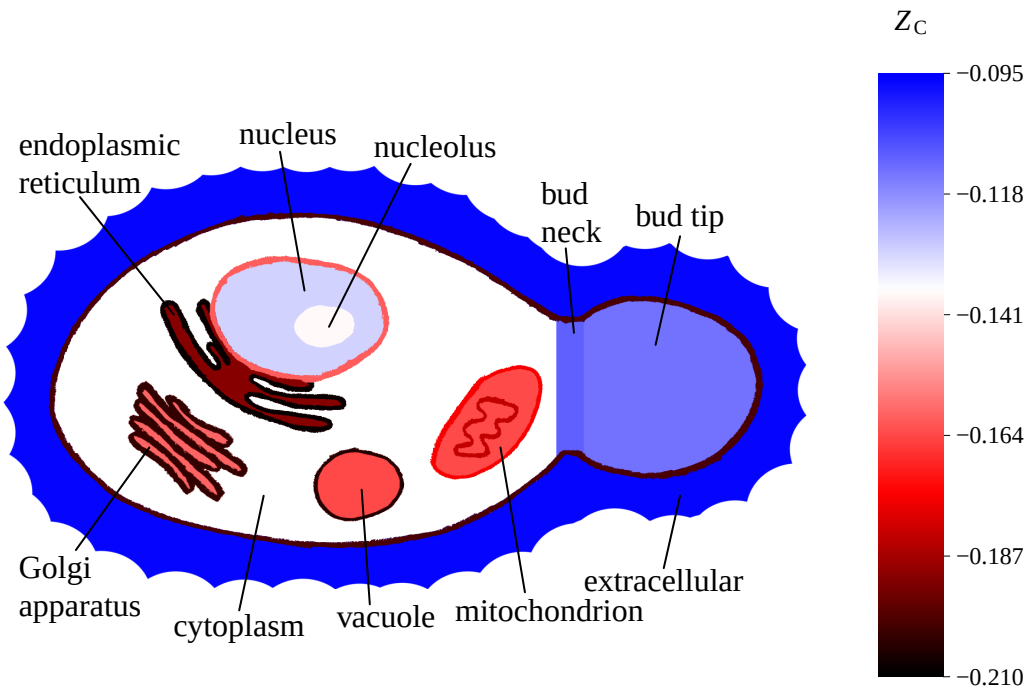

Figure 3: A schematic drawing of a yeast cell showing median values of the average oxidation state of carbon in proteins from selected subcellular locations listed in table 1. All cellular components listed in the table are represented in the drawing. The colour scale is adjusted so that the cytoplasm has a neutral hue (white), and locations with relatively oxidized and reduced proteins are depicted by blue and red colours, respectively. Darker red colours are used for the more reduced groups of proteins in some of the membranes such as the ER, Golgi, vacuolar and plasma membranes. The nuclear and inner and outer mitochondrial membranes are shown with lighter reds because their proteins are relatively oxidized compared to those in the other membranes. Components not separated by membranes (or with membranes not shown) include the nucleolus, bud neck and bud tip.

Table 2: Values of Eh compiled from literature sources, for yeast cells or culture except as noted. The ranges account for variation among different cell types, experimental techniques and published values, and are used to construct figure 4.

location	range (mV)	references
mitochondrion	-360 to -255	Ref. [33]: roGFP probe (-360 mV; human HeLa). Ref. [34]: rxYFP in matrix (-296 mV) and intermembrane space (-255 mV).
cytoplasm	-320 to -240	Ref. [35]: GSH in proliferating mammalian cells (-240 mV). Ref. [36]: roGFP probe of GSH (-320 mV).
endoplasmic reticulum	-208 to -133	Ref. [37]: NYTC peptide (-185 to -133 mV; murine hybridoma CRL-1606). Ref. [38]: roGFP (-208 mV; human HeLa)
vacuole	-160 to >-130	See text.
extracellular	-150 to >160	Ref. [1]: Aerobic (160 mV) and anaerobic (90 mV) cultures. Ref. [39]: Very high-gravity fermentation (-150 mV). Ref. [40]: *H. sapiens* plasma (-140 mV).

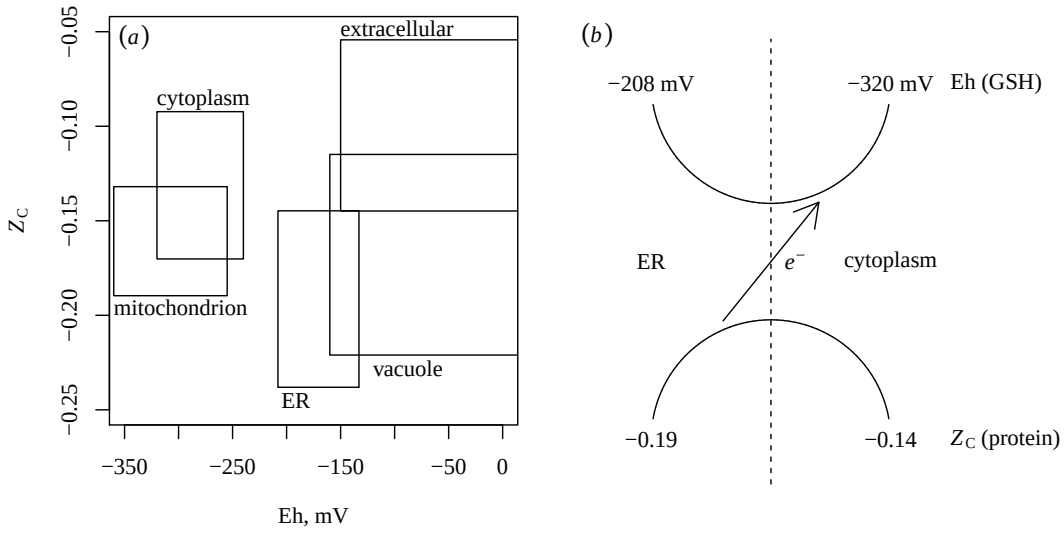

Figure 4: The plot (*a*) compares average oxidation state of carbon in proteins from different subcellular locations of yeast with Eh values taken from glutathione (GSH-GSSH) or other redox indicators (see table 2). The heights of the boxes indicate the interquartile ranges of Z_C values, and the widths represent the ranges of Eh listed in table 2. The scheme (*b*) invokes electron transfer to account for the contrasts in redox potential of GSH/GSSG and Z_C of proteins between ER and cytoplasm (see text).

This scheme is depicted in figure 4*b*. The vertical dashed line represents a physical (but not impermeable) boundary between ER and cytoplasm; the curved lines represent reactions and transport within glutathione and protein systems that cross the compartments, and the arrow represents a linkage between glutathione and protein systems, which is effectively a coupled oxidation-reduction reaction. The scheme represents a redox mass-balance interpretation of the overall stoichiometric relationships, not a mechanism applied to individual GSH or protein molecules; the complete picture of metabolic connectivity in the cell is certainly more complex. Note that this scheme refers to the relative oxidation states of carbon of the proteins, not the oxidation of protein thiol groups to form disulfides. Disulfide bond formation takes place during folding and secretion of proteins, and may also contribute to glutathione metabolism [4]. Although there is growing detail of the pathways of glutathione metabolism in the cell, including compartmentalization between ER and cytoplasm, it is not known how they connect with non-thiol systems (e.g. [40]). Integrating the oxidation-reduction requirements of protein metabolism into existing metabolic models may help to complete the balance sheet of redox interactions in the cell.

Experiments on the connections between redox conditions and protein metabolism at the subcellular level can help elucidate the possible effects of coupling of protein metabolism to the glutathione redox system. If the net transfer of high-Z_C proteins to the cytoplasm is stopped, then a decrease in GSH/GSSG redox potential in the ER relative to the cytoplasm would be expected based on the scheme shown in figure 4. This prediction is consistent with the outcome of experiments showing that puromycin-induced halting of protein synthesis causes a decrease in the redox potential monitored by roGFP in the ER [43]. A further untested implication of this hypothesis is that in ER-stress experiments the Z_C of the protein population in the ER would increase. Metabolism of proteins might also interact with redox pathways other than the oxidation and reduction of glutathione. A linkage of this type has been documented in plants, where degradation of aromatic and branched-chain amino acids was identified as a source of electrons for the mitochondrial electron transport chain [44].

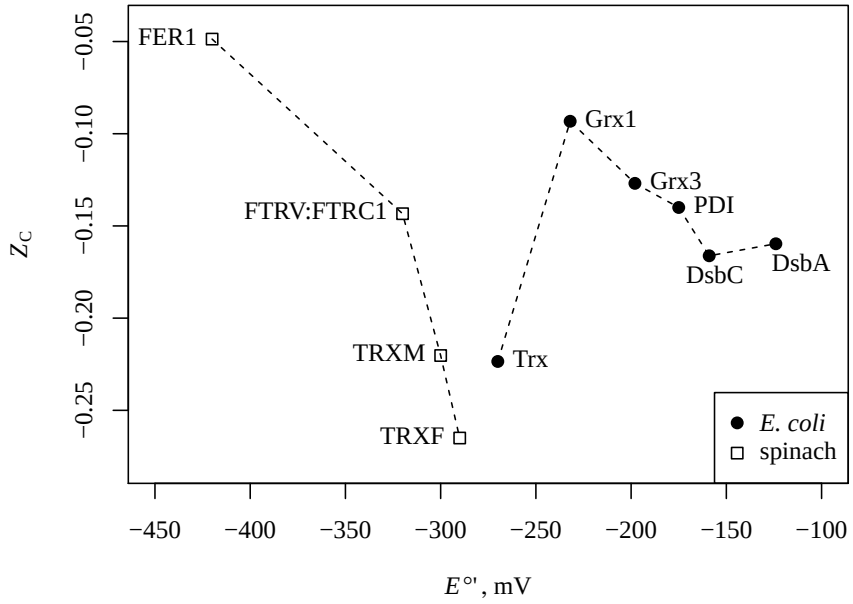

Figure 5: Average oxidation state of carbon in proteins compared with standard reduction potentials for ferredoxin (FER1), ferredoxin/thioredoxin reductase (FTRV:FTRC1 dimer), thioredoxin *f* (TRXF) and *m* (TRXM) from spinach [47] and thioredoxin (Trx), glutaredoxin 1 (Grx1) and 3 (Grx3), protein disulfide isomerase (PDI), and thiol:disulfide interchange protein A (DsbA) and C (DsbC) from *E. coli* [48].

3.4 Comparison of Z_C with standard reduction potentials of proteins

In plants, the oxidation and reduction of the iron-sulfur cluster in ferredoxin and the thiol/disulfide groups in ferredoxin-thioredoxin reductase and thioredoxin are coupled to form the ferredoxin/thioredoxin system [45] ("system" here refers to the set of interacting proteins, and is not a system in the thermodynamic sense). Ferredoxin has the lowest standard reduction potential (midpoint potential, $E^{\circ\prime}$ at 25 °C) in this system, and its iron-sulfur cluster is reduced by light energy through photosystem I. The oxidation of ferredoxin coupled to the reduction of thioredoxin is catalysed by ferredoxin/thioredoxin reductase. Reduced thioredoxin can disseminate the redox signal via reduction of disulfide groups in other proteins, activating their enzymatic functions. Glutaredoxins are another group of thiol/disulfide proteins that also interact with disulfide bonds in proteins and, unlike thioredoxin, are reduced by glutathione [46].

The standard reduction potentials of proteins in the ferredoxin/thioredoxin system in spinach [47] and of thioredoxin and the glutaredoxin system in *E. coli* [48] are compared with Z_C in figure 5. In the ferredoxin-thioredoxin chain of spinach, there is a strong decrease in Z_C with increasing $E^{\circ\prime}$. In the glutaredoxin system of *E. coli* and the associated proteins protein disulfide isomerase (PDI) and thiol:disulfide interchange protein (DsbA), there is a smaller decrease in Z_C with increasing $E^{\circ\prime}$. Thioredoxin in *E. coli*, which has Z_C and $E^{\circ\prime}$ that are similar to thioredoxin in spinach, does not follow the trend apparent in the glutaredoxin and related proteins.

From figure 5 it appears that the Z_C of proteins involved in some parts of the redox signalling networks are inversely correlated with their standard reduction potential. Do systematic changes in amino acid composition implied by Z_C impact the chemistry of the active site? The atomic environment surrounding the redox-active sites affects reduction potential [49], and changing very few residues in the vicinity of

the active site can affect the function [50]. However, these or similar proximal effects may not provide a complete explanation for the trends in Z_C in figure 5, which apply to the entire protein sequences. A speculative explanation is offered here. In general, any oxidation of a protein molecule may involve loss of an electron from the active site or from a covalent bond distal to the active site. In proteins with higher Z_C, the degree of oxidation of amino acids is greater, making the further loss of electrons from covalent bonds more difficult. The covalent oxidation of proteins ultimately leads to their degradation, disrupting the function of redox signalling networks. Therefore, it may be advantageous for low-$E^{\circ\prime}$ active sites (those that have a greater potential to lose electrons on signalling timescales), to be associated with high-Z_C proteins (those in which the covalent bonds have lost a greater number of electrons on evolutionary timescales). The break in the pattern by thioredoxin in *E. coli*, and the scattered distribution of Z_C and $E^{\circ\prime}$ of many other redox-active proteins that are not shown, presumably indicate that these hypothetical relations are applicable only to closely interacting chains of redox-active proteins and not the entire cell.

There are dual hypotheses here: first, that high-Z_C proteins have a lower tendency for irreversible oxidative degradation, and second, that reversible reduction potential and tendency for irreversible oxidation, which are different biochemical properties of the same molecule, are jointly tuned by evolution. One implication of the first hypothesis is that thioredoxin in spinach, which has a relatively low Z_C, would be more easily covalently oxidized, and therefore may have a higher turnover rate than ferredoxin.

3.5 Phylogenetic variation in Z_C of proteins and comparison with optimal growth temperature

A comparison of Z_C of the combined proteins from selected microbial genomes is shown in figure 6. The sets of proteins shown on the left-hand side of figure 6 correspond to those organisms whose scientific names contain the indicated substring. In many cases, the names of the organisms reflect their environments and/or metabolic strategies. Examples of the matching genus names are *Natronobacterium*, *Haloferax*, *Rhodobacter*, *Acidovorax*, *Methylobacterium*, *Chlorobium*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Methanococcus*, *Thermococcus*, *Pyrobaculum*, *Sulfolobus*. Most terms, however, match more than one genus (e.g. *Pyrobaculum* and *Pyrococcus*). On the right-hand side of figure 6 are shown genera containing many groups with clinical and technological relevance; by the numbers of points it is apparent that their representation in RefSeq is greater than that of the environmental microbes.

A general trend toward lower Z_C in proteins in organisms from hot environments (e.g. represented by Thermo and Pyro in the names) is apparent. Organisms with the highest Z_C inhabit saline evaporative waters (Natr, Halo), while other aquatic organisms (e.g. Rhodo) have less highly oxidized proteins. Within a given genus (right-hand side of plot), the clusters of Z_C tend to be tighter, reflecting conserved compositional trends. *Streptomyces*, common in soils, has the highest Z_C of the genera shown here. *Buchnera* is notable because its proteins are highly reduced, including one example (*Buchnera aphidicola* BCc) with the lowest Z_C of proteins within the entire dataset. *B. aphidicola* BCc is the primary endosymbiont of the cedar aphid (*Cinara cedri*) and, at the time of sequencing, had one of the smallest known bacterial genomes [51]. Being relatively closely related [52], *Mycoplasma* and *Clostridium* also have relatively low Z_C. *Mycoplasma* are known for their small genomes and dependence on metabolic products of the host; the low Z_C may be a constraint imposed by growth in reducing intracellular or intraorganismal environments.

We now turn to a comparison of homologues of a specific protein. Rubisco is an essential enzyme for carbon fixation. Sequence comparison of homologues (related sequences that appear in different organisms) has provided the basis for many phylogenetic studies [53]. Major divergent forms of the enzyme are Forms I and II, found in aerobic organisms, and Form III and "Rubisco-like proteins", found in anaerobic organisms [54]. The organisms listed in table 3 have in common the occurrence of Rubisco in their genome. Forms I, II or III were included in this comparison, but Rubisco-like proteins were

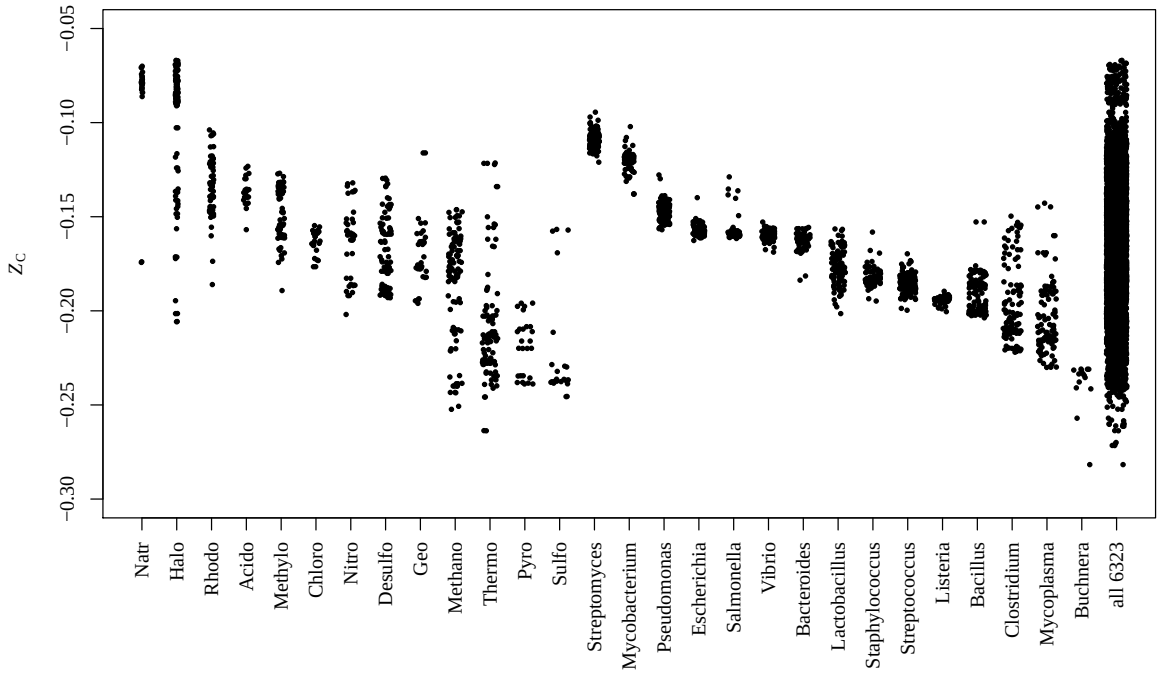

Figure 6: Average oxidation state of carbon in total combined proteins from sequenced microbial genomes. Sequences of microbial proteins were taken from NCBI RefSeq release 61. Only organisms with total sequenced protein length greater than 100,000 amino acids were used, leaving 6323 organisms. The group on the left-hand side is identified by substring matches in the scientific name of the organism; the terms were chosen to emphasize environmental variation. The group on the right-hand side consists of the indicated genera, emphasizing organisms of clinical and biotechnological relevance. The final category represents total proteins from all microbial genomes meeting the minimal size requirement, many of which are not shown in the other categories.

Table 3: Names of species, optimal growth temperatures (T_{opt}) and UniProt [26] accession numbers (IDs) for the large subunit of ribulose bisphosphate carboxylase (Rubisco). Literature references for T_{opt} are indicated in brackets. Abbreviations: A – Archaea; B – Bacteria; E – Eukaryota. The numbers are used to identify the points in figure 7 (duplicated numbers occur in different temperature ranges).

number	domain	species	T_{opt}, °C	reference	ID
1	E	*Phaeocystis antarctica*	0–6	[55]	G9FID8
2	B	*Octadecabacter antarcticus* 307	4–10	[56]	M9R7V1
3	A	*Methanolobus psychrophilus*	18	[57]	K4MAK9
4	A	*Methanococcoides burtonii*	23	[58]	Q12TQ0
5	B	*Bradyrhizobium japonicum*	25–30	[59]	Q9ZI34
6	B	*Thiobacillus ferrooxidans*	28–30	[60]	P0C916
7	E	*Zea mays*	30	[61]	P00874
8	B	*Mariprofundus ferrooxydans*	30	[62]	Q0EX22
9	B	*Desulfovibrio hydrothermalis*	35	[63]	L0RHZ1
1	A	*Methanosarcina acetivorans*	35–40	[64]	Q8THG2
2	B	*Acidithiobacillus caldus*	45	[65]	F9ZLP0
3	E	*Cyanidium caldarium*	45	[66]	P37393
4	B	*Sulfobacillus acidophilus*	45–50	[67]	P72383
5	B	*Pseudomonas hydrogenothermophila*	52	[68]	Q51856
6	B	*Synechococcus* sp. (strain JA-2-3B'a(2-13))	50–55	[69]	Q2JIP3
7	A	*Methanosaeta thermophila*	55–60	[70]	A0B9K9
8	B	*Thermosynechococcus elongatus*	57	[71]	Q8DIS5
9	B	*Clostridium clariflavum*	60	[72]	G8LZL2
1	B	*Bacillus acidocaldarius*	65	[73]	F8IID7
2	B	*Thermotoga lettingae*	65	[74]	A8F7V4
3	B	*Thermomicrobium roseum*	70	[75]	B9KXE5
4	A	*Archaeoglobus fulgidus*	76	[76]	O28635
5	A	*Methanocaldococcus jannaschii*	85	[77]	Q58632
6	A	*Thermofilum pendens*	85–90	[78]	A1RZJ5
7	A	*Staphylothermus marinus*	85–92	[79]	A3DND9
8	A	*Pyrococcus horikoshii*	98	[80]	O58677
9	A	*Pyrococcus furiosus*	100	[81]	Q8U1P9

excluded; however, some that were tested were found to have considerably lower Z_C. The selection of organisms was made in order to represent a variety of optimal growth temperatures (T_{opt}) as reported in the studies cited in the table [55–81].

A comparison between T_{opt} and Z_C of Rubisco is presented in figure 7. The Z_C of Rubisco are somewhat higher than the bulk protein content of the organisms; compare for example the values for *Pyrococcus horikoshii* and *P. furiosus* (the highest-temperature points labelled 8 and 9 in figure 7*a*) with the range of values for "Pyro" in figure 6. At lower temperatures (0 to 50 °C), the differences between domains of life are most apparent; Rubiscos of the Bacteria in this sample set are more oxidized than those of Archaea and Eukaryota. There is a tendency for the Rubiscos of the Archaea to have lower Z_C; this appears to be characteristic of anaerobic methanogenesis and Form III Rubiscos. An interesting exception is the high-Z_C Rubisco of *Methanosaeta thermophila*; this organism grows on acetate to produce both CH_4 and CO_2 [70]. The major pattern that emerges is that higher temperatures are associated with a lower average oxidation state of carbon in proteins. As outlined below, a decrease in oxidation state of carbon in the covalent structure of the proteins confers energetic savings in hot, reducing environments.

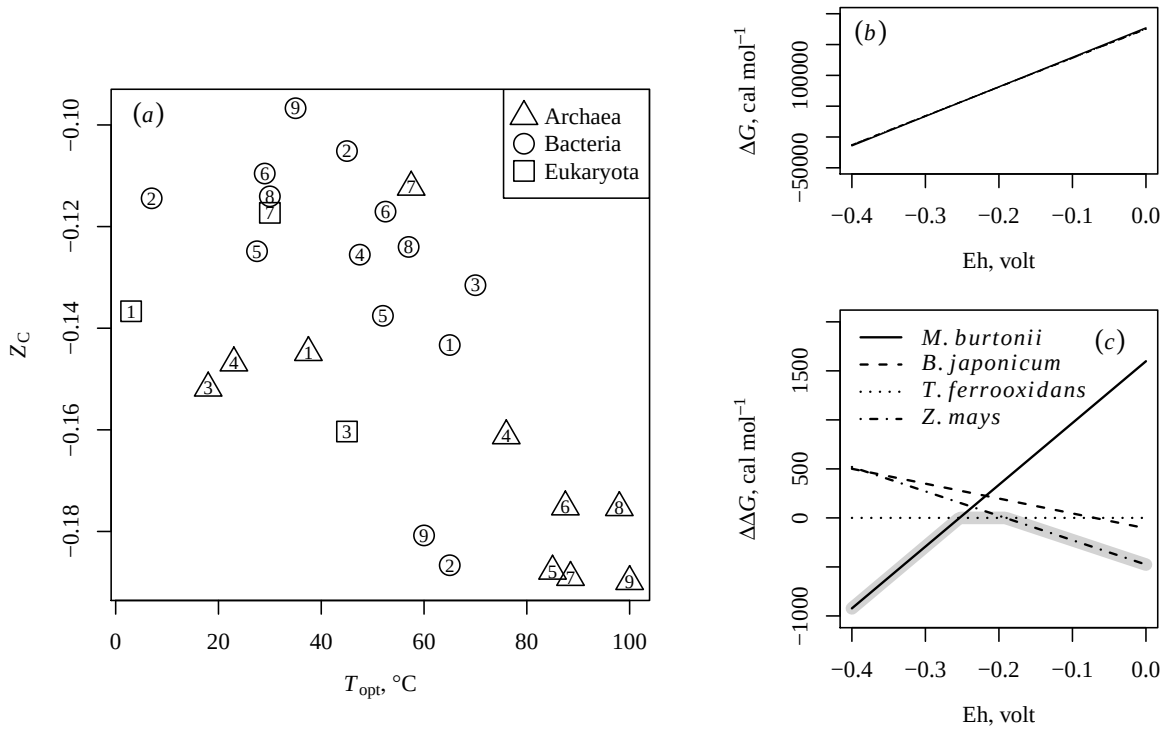

Figure 7: (*a*) Average oxidation state of carbon in Rubisco compared with optimal growth temperature (T_{opt}) of organisms. Numbers are used to identify the organisms (see table 3). (*b*-*c*) Gibbs energies of formation reactions, per residue, of selected Rubisco from organisms in the 23–30 °C range of optimal growth temperature (points labelled 4–7). (*b*) Total Gibbs energies of individual reactions as a function of Eh and (*c*) difference between the reaction for *T. ferrooxidans* and the others are shown. The grey highlight indicates the protein with the lowest ΔG along the range of Eh values.

3.6 Beyond Z_C: energetics of protein formation as a function of environmental redox potential

To a first approximation, energetic considerations predict that more reducing conditions tend to favour formation of proteins with relatively lower Z_C, and vice versa. To assess the directionality and magnitude of chemical forces on the evolutionary transformations of proteins, the energetics of reactions can be calculated using thermodynamic models. Because the timescales of evolution are much longer than transformations of biomolecules during metabolism, a discussion of the assumptions underlying the application of thermodynamic theory to biochemical evolution is warranted.

The calculation of equilibrium provides a quantitative description of the state of a system in an energy minimum. The assumption of equilibrium is the foundation for many models of inorganic processes in geochemistry. Although the metabolic formation of proteins and other biochemical constituents proceeds in a non-equilibrium manner, using the equilibrium state as a frame of reference makes it possible to compare the energetics of living systems quantitatively ("how far from equilibrium").

In energetic terms, adaptation can be defined as a problem of optimal efficiency, with trade-offs between energy utilization and power [82, p. 140]. Overall minimization of biomass synthesis costs can be expected from the energy utilization standpoint. The links between energetics and evolutionary outcome implicitly depend on the proposal that greater fitness is associated with mutations that lower the synthesis costs of proteins for a given function (e.g. [83]).

It has been argued previously that the propensity for some evolutionary changes can be modelled using

the related concepts of equilibration, energy minimization, and maximum entropy. In an evolutionary context, these concepts have often been defined by analogy to their definitions in thermodynamics [82, 84]. The current discussion instead considers the possibility of direct application of chemical thermodynamics (the geochemical approach) to formulate a quantitative description of patterns of protein composition. The major assumption used in the following discussion is that energetic demands of protein formation depend not only on the composition of the protein, but also the environmental conditions. Therefore, it is expected that environmental adaptation has left an imprint on chemical compositions of phylogenetically distinct proteins.

An example calculation is carried out here for selected Rubiscos from organisms with optimal temperatures in the range of 23–30 °C (table 3, numbers 4–7). The basis species (representing inorganic starting materials) and their chemical activities used for this example are CO_2 (10^{-3}), H_2O (1), NH_3 (10^{-4}), H_2S (10^{-7}) and H^+ (10^{-7}, i.e. $pH = 7$). The activity of the electron, representing the effects of the redox variable (Eh) through the Nernst equation, is left to vary. The Gibbs energies of the reactions to form the proteins were calculated as described previously [9, 19] and are plotted in figure 7*b*. In figure 7*b* it can be seen that the Gibbs energies of the reactions to form the proteins (ΔG), normalized by protein length, steadily increase (become less favourable) with increasing Eh, but the differences between the proteins can not be discerned easily. In figure 7*c*, the values of ΔG are shown relative to the reaction for *T. ferrooxidans*, giving a difference in Gibbs energy of formation ($\Delta\Delta G$) that can be used to assess the relative energetics of the reactions. Lower energies indicate the most stable protein (in the sense of chemical formation, not structural conformation), and these relative energies depend on the chemical conditions of the environment, measured in part by the oxidation-reduction potential.

By using the Gibbs energy calculations such as shown in figure 7*c*, one can make an assessment of which protein in a given redox environment demands lower energy for overall synthesis (i.e. is more stable) than other possibilities. Where two lines cross in figure 7*c*, the energies to form the two proteins are equal, representing a metastable (partial) equilibrium. In metastable equilibrium, the coexistence of proteins with equal energies of formation corresponds to a local energy minimum. This interpretation does not preclude the non-equilibrium character of the overall biosynthetic process, because the energies of formation remain non-zero (figure 7*b*).

The three least costly proteins, going from low to high Eh, are those from *M. burtonii*, *T. ferrooxidans*, and *Z. mays*. Methanogenic archaea such as *M. burtonii* [58] inhabit reducing environments, where Eh values as low as at least -400 mV have been documented [2]. In contrast, the metal-leaching activity of *T. ferrooxidans* is associated with an increase in oxidation-reduction potential (ORP, which carries the same units as Eh) [85] and, compared to *M. burtonii*, its Rubisco is more oxidized and is consequently more stable at higher Eh. The correspondence between redox conditions and lower ΔG of formation of the proteins provides thermodynamic evidence for environmental adaptation of the proteins.

The protein from *T. ferrooxidans*, which has the highest Z_C of the four considered, does not have the lowest value of $\Delta\Delta G$ in the most oxidizing (highest Eh) conditions. Instead, the Rubisco from *Z. mays*, even though it has lower Z_C, is calculated to relatively more stable at the highest Eh values considered. This result shows that comparing values of Z_C provides only an approximation of the dependence of the relative energetics of protein formation on redox changes; chemical thermodynamic models integrate more information about reaction stoichiometry and the effects of multiple environmental variables.

Previously, comparison of Rubisco from hot-spring organisms adapted to higher temperatures revealed an increase in frequency hydrophobic amino acids, interpreted as increasing the conformational stability of proteins at high temperature [86]. However, another basis for interpretations of thermophilic adaptation depends on the relative energetic costs of synthesis of amino acids [87]. The finding made in this study is that the high-temperature Rubiscos exhibit a shift toward lower Z_C (figure 7*a*). Higher temperatures are often associated with more reducing environments. For example, compared to seawater, activities

of dissolved hydrogen in hydrothermal fluids are higher, and mixed hydrothermal-seawater fluids have a reducing potential that favours formation of relatively reduced amino acids [8]. Redox potential is a major variable affecting the energetics of protein formation at different temperatures; therefore, adaptation to minimize biosynthetic costs in high-temperature environments is likely to have more than just a thermophilic (temperature dependent) aspect.

Although the stoichiometric comparisons are distinct from sequence-based phylogenetic analyses which are used to test for positive selection, it is conceivable that Z_C could in the future be incorporated into tree-based models of evolution that take account of physicochemical properties of amino acids in proteins [88]. However, as noted above, thermodynamic comparisons have greater power than compositional comparisons such as Z_C. The computation of metastable equilibrium takes account simultaneously of temperature, redox potential (expressed as Eh, activity of hydrogen, or oxygen fugacity) and other variables [19]. In another study, analysis of metagenomic and geochemical data led to a predicted metastable succession of proteins (with generally increasing Z_C) that could be aligned with a gradient of increasing oxidation potential and decreasing temperature in a flowing hot spring [10]. When grouped by taxonomic similarity, the Z_C in the hot-spring proteins, while becoming lower at high temperature, also spread over a broader range, leading to tighter constraints on the redox conditions suitable for metastable coexistence of the organisms [11].

Not only differences in the oxidation states of present-day environments, but also the oxygenation of Earth's atmosphere and oceans through geological time could have profound impacts on the energetics of biomass synthesis [7]. Adaptation to reduce these costs likely would lead to divergences in Z_C that are apparent across different taxa, while closer phylogenetic relationships should confer a similarity in Z_C. In common with the Rubiscos, comparison of the total proteins of microbes reveals a tendency for proteins in organisms associated with hot, as well as sulfidic and methanogenic environments, to be more reduced (figure 6).

4 Conclusions

Proteins are products of metabolism; their synthesis and degradation are part of the network of chemical reactions that sustains the living cell. Comparisons of the average oxidation state of carbon in proteins have provided a starting point for visualizing the compositional diversity of proteins in relation to redox chemistry in subcellular compartments and external environments. The large differences in Z_C of proteins in locations such as the ER and cytoplasm likely have consequences for the dynamics of oxidation-reduction reactions involving glutathione and other metabolites. Further insight may be gained by including the formation and degradation of proteins in kinetic and stoichiometric models of metabolic networks. Extension of these concepts to other phenomena entailing changes in both redox potential and protein expression, such as stress response to oxidizing agents and the cell cycle, can be envisioned. The deepest significance of the observed patterns lies in their emergence over evolutionary timescales. The inverse trend relating Z_C with standard reduction potentials in chains of redox-active proteins is a case where the chemical composition of the proteins may be tuned with the electron-transfer chemistry of the active sites. Compositional divergences among proteins are also apparent in phylogenetic comparisons, and here it is reasonable to conclude that correlations between oxidation state of carbon in proteins and the redox potential of the environment indicate some degree of energy savings conferred by evolution. The natural history of protein evolution is a result of processes that are both unpredictable (mutation events) and, to some extent, deterministic (selection for fitness in a given environment). By describing protein molecules in terms of chemical composition and energetics it will be possible to identify some of the forces that help to shape the occurrences of proteins in different cells and environments.

5 Supporting information

The supporting information is provided in a ZIP archive containing the following code and data files:

prep.R This file contains code used to prepare the data files for easier handling by the plotting functions.

plot.R This file contains code used to make the figures appearing in this paper. The functions, in the order of the figures (1–7), are amino(), human(), yeast(), potential(), midpoint(), phylo(), and rubisco(). The code is written in R [18] and depends on version 1.0.2 of the CHNOSZ package [19], available from the Comprehensive R Archive network (`http://cran.r-project.org`).

data/SGD_associations.csv For yeast genes, this table lists the accessions, SGDID, and the association to cellular components in the Gene Ontology, derived from gene_association.sgd.gz, protein_properties.tab and go_terms.tab downloaded from `http://www.yeastgenome.org` on 2013-08-24. All gene associations with the NOT qualifier were removed, as were those without a matching entry in protein_properties.tab (e.g. RNA-coding genes).

data/ZC_HUMAN.csv, ZC_membrane.csv Compilations of the values of Z_C for human proteins and human membrane proteins. Values in ZC_HUMAN.csv ware calculated from protein sequences in HUMAN.fasta.gz, downloaded from `ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/proteomes/HUMAN.fasta.gz` on 2013-08-24 (file dated 2013-07-24). Values in ZC_membrane.csv were calculated from protein sequences in all *.fa files in Additional File 2 of Almén *et al.*, 2009 [27].

data/codons.csv In the first column, the three-letter abbreviations for each of the RNA codons; in the second column, the names of the corresponding amino acids.

data/midpoint.csv List of protein names, UniProt IDs and standard midpoint reduction potentials used to make figure 5. Start and stop positions, taken from UniProt, identify the protein chain excluding initiator methionines or signal peptides.

data/protein_refseq.csv Amino acid compositions of total proteins in 6758 microbial genomes from RefSeq release 61, dated 2013-09-09. The gene identifier (gi) numbers of the sequences were assigned taxonomic IDs (taxids) using the RefSeq release catalogue. The amino acid compositions of the total proteins were calculated by averaging the compositions of all proteins for each taxid. The "organism" column contains the taxid used in NCBI databases, the "ref" column contains the names of the RefSeq files from which the amino acid sequences were taken (with start and end positions in parentheses) followed by the scientific name of the organisms in brackets, and the "abbrv" column contains the number of amino acids for that organism. Scientific names for the taxids at the species level were found using the names.dmp and nodes.dmp files downloaded from `ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz` on 2013-09-18.

data/rubisco.csv UniProt IDs for Rubisco and optimal growth temperatures of organisms (see table 3).

cell/*.png PNG images for each of the cellular components used to make figure 3.

fasta/midpoint/*.fasta FASTA sequence files for proteins shown in figure 5.

fasta/rubisco/*.fasta FASTA sequence files for each Rubisco identified in table 3.

Acknowledgements

Thanks to Svenja Tulipani and Katy Evans for their comments on an earlier version of the manuscript.

References

[1] Hewitt LF. 1950 *Oxidation-Reduction Potentials in Bacteriology and Biochemistry*, 6th edn. Edinburgh, Scotland: E. & S. Livingstone Ltd.

[2] Baas Becking LGM, Kaplan IR, Moore D. 1960 Limits of the natural environment in terms of pH and oxidation-reduction potentials. *J. Geol.* **68**, 243–284. See http://www.jstor.org/stable/30059218.

[3] Ziegler DM. 1985 Role of reversible oxidation-reduction of enzyme thiols-disulfides in metabolic regulation. *Annu. Rev. Biochem.* **54**, 305–329. (doi:10.1146/annurev.bi.54.070185.001513)

[4] Bindoli A, Rigobello MP. 2013 Principles in redox signaling: from chemistry to functional significance. *Antioxid. Redox Signaling* **18**, 1557–1593. (doi:10.1089/ars.2012.4655)

[5] Dawes EA, Ribbons DW. 1962 The endogenous metabolism of microorganisms. *Annu. Rev. Microbiol.* **16**, 241–264 (doi:10.1146/annurev.mi.16.100162.001325)

[6] Amend JP, Shock EL. 2001 Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol. Rev.* **25**, 2175–243 (doi:10.1016/S0168-6445(00)00062-0)

[7] Sleep NH, Bird DK. 2008 Evolutionary ecology during the rise of dioxygen in the Earth's atmosphere. *Phil. Trans. R. Soc. B* **363**, 2651–2664. (doi:10.1098/rstb.2008.0018)

[8] Amend JP, Shock EL. 1998 Energetics of amino acid synthesis in hydrothermal ecosystems. *Science* **281**, 1659–1662. (doi:10.1126/science.281.5383.1659)

[9] Dick JM, LaRowe DE, Helgeson HC. 2006 Temperature, pressure, and electrochemical constraints on protein speciation: group additivity calculation of the standard molal thermodynamic properties of ionized unfolded proteins. *Biogeosciences* **3**, 311–336. (doi:10.5194/bg-3-311-2006)

[10] Dick JM, Shock EL. 2011 Calculation of the relative chemical stabilities of proteins as a function of temperature and redox chemistry in a hot spring. *PLoS ONE* **8**, e22782. (doi:10.1371/journal.pone.0022782)

[11] Dick JM, Shock EL. 2013 A metastable equilibrium model for the relative abundances of microbial phyla in a hot spring. *PLoS ONE* **8**, e72395. (doi:10.1371/journal.pone.0072395)

[12] Harms MJ, Thornton JW. 2013 Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–571. (doi:10.1038/nrg3540)

[13] Sterner R, Liebl W. 2001 Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **36**, 39–106. (doi:10.1080/20014091074174)

[14] Cornish-Bowden A, Cárdenas ML, Letelier J-C, Soto-Andrade J. 2007 Beyond reductionism: metabolic circularity as a guiding vision for a real biology of systems. *Proteomics* **7**, 839–845. (doi:10.1002/pmic.200600431)

[15] Helgeson HC. 1991 Organic/inorganic reactions in metamorphic processes. *Can. Mineral.* **29**, 707–739. See http://canmin.geoscienceworld.org/content/29/4.toc.

[16] Masiello CA, Gallagher ME, Randerson JT, Deco RM, Chadwick OA. 2008 Evaluating two experimental approaches for measuring ecosystem carbon oxidation state and oxidative ratio. *J. Geophys. Res.: Biogeosci.* **113**, G03010. (doi:10.1029/2007JG000534)

[17] Kroll JH *et al.* 2011 Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol. *Nat. Chem.* **3**, 133–139. (doi:10.1038/NCHEM.948)

[18] R Core Team. 2013 R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. See http://www.R-project.org.

[19] Dick JM. 2008 Calculation of the relative metastabilities of proteins using the CHNOSZ software package. *Geochem. Trans.* **9**, 10. (doi:10.1186/1467-4866-9-10)

[20] Wong JT-F. 1981 Coevolution of genetic code and amino acid biosynthesis. *Trends Biochem. Sci.* **6**, 33–36. (doi:10.1016/0968-0004(81)90013-X)

[21] Woese CR, Dugre DH, Saxinger WC, Dugre SA. 1966 The molecular basis for the genetic code. *Proc. Natl Acad. Sci. USA* **55**, 966–974. (doi:10.1073/pnas.55.4.966)

[22] Shock EL. 1990 Do amino acids equilibrate in hydrothermal fluids? *Geochim. Cosmochim. Acta* **54**, 1185–1189. (doi:10.1016/0016-7037(90)90450-Y)

[23] LaRowe DE, Helgeson HC. 2006 Biomolecules in hydrothermal systems: calculation of the standard molal thermodynamic properties of nucleic-acid bases, nucleosides, and nucleotides at elevated temperatures and pressures. *Geochim. Cosmochim. Acta* **70**, 4680–4724. (doi:10.1016/j.gca.2006.04.010)

[24] Kyte J, Doolittle RF. 1982 A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132. (doi:10.1016/0022-2836(82)90515-0)

[25] Amend JP, LaRowe DE, McCollom TM, Shock EL. 2013 The energetics of organic synthesis inside and outside the cell. *Phil. Trans. R. Soc. B* **368**, 20120255. (doi:10.1098/rstb.2012.0255)

[26] The UniProt Consortium. 2012 Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40**, D71–D75 (doi:10.1093/nar/gkr981)

[27] Almén M, Nordström KJV, Fredriksson R, Schiöth HB. 2009 Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* **7**, 50. (doi:10.1186/1741-7007-7-50)

[28] Frank SA. 2009 The common patterns of nature. *J. Evol. Biol.* **22**, 1563–1585. (doi:10.1111/j.1420-9101.2009.01775.x)

[29] Dick JM. 2009 Calculation of the relative metastabilities of proteins in subcellular compartments of *Saccharomyces cerevisiae*. *BMC Syst. Biol.* **3**, 75. (doi:10.1186/1752-0509-3-75)

[30] Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003 Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691. (doi:10.1038/nature02026)

[31] Cherry JM *et al.* 2012 Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700–D705. (doi:10.1093/nar/gkr1029)

[32] Ashburner M *et al.* 2000 Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (doi:10.1038/75556)

[33] Hanson GT, Aggeler R, Oglesbee D, Cannon M, Capaldi RA, Tsien RY, and Remington SJ. 2004 Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators. *J. Biol. Chem.* **279**, 13044–13053. (doi:10.1074/jbc.M312846200)

[34] Hu JJ, Dong LX, Outten CE. 2008 The redox environment in the mitochondrial intermembrane space is maintained separately from the cytosol and matrix. *J. Biol. Chem.* **283**, 29126–29134. (doi:10.1074/jbc.M803028200)

[35] Schafer FQ, Buettner GR. 2001 Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. *Free Radic. Biol. Med.* **30**, 1191–1212. (doi:10.1016/S0891-5849(01)00480-4)

[36] Morgan B, Ezeriņa D, Amoako TNE, Riemer J, Seedorf M, Dick TP. 2013 Multiple glutathione disulfide removal pathways mediate cytosolic redox homeostasis. *Nat. Chem. Biol.* **9**, 119–125. (doi:10.1038/NCHEMBIO.1142)

[37] Hwang C, Sinskey AJ, Lodish HF. 1992. Oxidized redox state of glutathione in the endoplasmic reticulum. *Science* **257**, 1496–1502. (doi:10.1126/science.1523409)

[38] Birk J, Meyer M, Aller I, Hansen HG, Odermatt A, Dick TP, Meyer AJ, Appenzeller-Herzog C. 2013 Endoplasmic reticulum: reduced and oxidized glutathione revisited. *J. Cell Sci.* **126**, 1604–1617. (doi:10.1242/jcs.117218)

[39] Lin Y-H, Chien W-S, Duan K-J. 2010 Correlations between reduction-oxidation potential profiles and growth patterns of *Saccharomyces cerevisiae* during very-high-gravity fermentation. *Process Biochem.* **45**, 765–770. (doi:10.1016/j.procbio.2010.01.018)

[40] Go Y-M, Jones DP. 2008 Redox compartmentalization in eukaryotic cells. *Biochim. Biophys. Acta* **1780**, 1271–1290. (doi:10.1016/j.bbagen.2008.01.011)

[41] Queval G, Jaillard D, Zechmann B, Noctor G. 2011 Increased intracellular H_2O_2 availability preferentially drives glutathione accumulation in vacuoles and chloroplasts. *Plant, Cell Environ.* **34**, 21–32. (doi:10.1111/j.1365-3040.2010.02222.x)

[42] Singh A, Kaur N, Kosman DJ. 2007 The metalloreductase Fre6p in Fe-efflux from the yeast vacuole. *J. Biol. Chem.* **282**, 28619–28626. (doi:10.1074/jbc.M703398200)

[43] van Lith M, Tiwari S, Pediani J, Milligan G, Bulleid NJ. 2011 Real-time monitoring of redox changes in the mammalian endoplasmic reticulum. *J. Cell Sci.* **124**, 2349–2356. (doi:10.1242/jcs.085530)

[44] Araújo WL, Tohge T, Ishizaki K, Leaver CJ, Fernie AR. 2011 Protein degradation - an alternative respiratory substrate for stressed plants. *Trends Plant Sci.* **16**, 489–498. (doi:10.1016/j.tplants.2011.05.008)

[45] Schürmann P, Buchanan BB. 2008 The ferredoxin/thioredoxin system of oxygenic photosynthesis. *Antioxid. Redox Signaling* **10**, 1235–1273. (doi:10.1089/ars.2007.1931)

[46] Prinz WA, Åslund F, Holmgren A, Beckwith J. 1997 The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**, 15661–15667. (doi:10.1074/jbc.272.25.15661)

[47] Hirasawa M, Schürmann P, Jacquot J-P, Manieri W, Jacquot P, Keryer E, Hartman FC, Knaff DB. 1999 Oxidation-reduction properties of chloroplast thioredoxins, ferredoxin:thioredoxin reductase, and thioredoxin *f*-regulated enzymes. *Biochemistry* **38**, 5200–5205. (doi:10.1021/bi982783v)

[48] Åslund F, Berndt KD, Holmgren A. 1997 Redox potentials of glutaredoxins and other thiol-disulfide oxidoreductases of the thioredoxin superfamily determined by direct protein-protein redox equilibria. *J. Biol. Chem.* **272**, 30780–30786. (doi:10.1074/jbc.272.49.30780)

[49] Krause G, Lundström J, Lopez Barea J, Pueyo de la Cuesta C, Holmgren A. 1991 Mimicking the active site of protein disulfide-isomerase by substitution of proline 34 in *Escherichia coli* thioredoxin. *J. Biol. Chem.* **266**, 9494–9500.

[50] Ren GP *et al.* 2009 Properties of the thioredoxin fold superfamily are modulated by a single amino acid residue. *J. Biol. Chem.* **284**, 10150–10159. (doi:10.1074/jbc.M809509200)

[51] Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A. 2006 A small microbial genome: the end of a long symbiotic relationship? *Science* **314**, 312–313. (doi:10.1126/science.1130441)

[52] Rogers MJ *et al.* 1985 Construction of the mycoplasma evolutionary tree from 5S rRNA sequence data. *Proc. Natl Acad. Sci. USA* **82**, 1160–1164. (doi:10.1073/pnas.82.4.1160)

[53] Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. 2008 Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* **59**, 1515–1524. (doi:10.1093/jxb/erm361)

[54] Nisbet EG, Grassineau NV, Howe CJ, Abell PI, Regelous M, Nisbet RER. 2007 The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiology* **5**, 311–335. (doi:10.1111/j.1472-4669.2007.00127.x)

[55] Wang XD, Tang KW, Wang Y, and Smith Jr WO. 2010 Temperature effects on growth, colony development and carbon partitioning in three *Phaeocystis* species. *Aquat. Biol.* **9**, 239–249. (doi:10.3354/ab00256)

[56] Gosink JJ, Herwig RP, Staley JT. 1997 *Octadecabacter arcticus* gen. nov., sp. nov., and *O. antarcticus*, sp. nov., nonpigmented, psychrophilic gas vacuolate bacteria from polar sea ice and water. *Syst. Appl. Microbiol.* **20**, 356–365. (doi:10.1016/S0723-2020(97)80003-3)

[57] Zhang GS, Jiang N, Liu XL, Dong XZ. 2008 Methanogenesis from methanol at low temperatures by a novel psychrophilic methanogen, "*Methanolobus psychrophilus*" sp. nov., prevalent in Zoige wetland of the Tibetan plateau. *Appl. Environ. Microbiol.* **74**, 6114–6120. (doi:10.1128/AEM.01146-08)

[58] Franzmann PD, Springer N, Ludwig W, Demacario EC, Rohde M. 1992 A methanogenic archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov. *Syst. Appl. Microbiol.* **15**, 573–581. (doi:10.1016/S0723-2020(11)80117-7)

[59] Jordan DC. 1982 Transfer of *Rhizobium japonicum* Buchanan 1980 To *Bradyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants. Int. *J. Syst. Bacteriol.* **32**, 136–139. (doi:10.1099/00207713-32-1-136)

[60] Hallmann R, Friedrich A, Koops H-P, Pommerening-Röser A, Rohde K, Zenneck C, Sand W. 1992 Physiological characteristics of *Thiobacillus ferrooxidans* and *Leptospirillum ferrooxidans* and physicochemical factors influence microbial metal leaching. *Geomicrobiol. J.* **10**, 193–206. (doi:10.1080/01490459209377920)

[61] Miedema P. 1982 The effects of low temperature on *Zea mays*. *Adv. Agron.* **35**, 93–128. (doi:10.1016/S0065-2113(08)60322-3)

[62] Emerson D, Rentz JA, Lilburn TG, Davis RE, Aldrich H, Chan C, Moyer CL. 2007 A novel lineage of Proteobacteria involved in formation of marine Fe-oxidizing microbial mat communities. *PLoS ONE* **2**, e667. (doi:10.1371/journal. pone.0000667)

[63] Alazard D, Dukan S, Urios A, Verhé F, Bouabida N, Morel F, Thomas P, Garcia J-L, Ollivier B. 2003 *Desulfovibrio hydrothermalis* sp. nov., a novel sulfate-reducing bacterium isolated from hydrothermal vents. *Int. J. Syst. Evol. Microbiol.* **53**, 173–178. (doi:10.1099/ijs.0.02323-0)

[64] Sowers KR, Baron SF, Ferry JG. 1984 *Methanosarcina acetivorans* sp. nov., an acetotrophic methane-producing bacterium isolated from marine sediments. *Appl. Environ. Microbiol.* **47**, 971–978.

[65] Hallberg KB, Lindström EB. 1994 Characterization of *Thiobacillus caldus* sp. nov., a moderately thermophilic acidophile. *Microbiology* **140**, 3451–3456. (doi:10.1099/13500872-140-12-3451)

[66] Doemel WN, Brock TD. 1970 The upper temperature limit of *Cyanidium caldarium*. *Arch. Mikrobiol.* **72**, 326–332. (doi:10.1007/BF00409031)

[67] Norris PR, Clark DA, Owen JP, Waterhouse S. 1996 Characteristics of *Sulfobacillus acidophilus* sp. nov. and other moderately thermophilic mineral-sulphide-oxidizing bacteria. *Microbiology* **142**, 775–783. (doi:10.1099/00221287-142-4-775)

[68] Goto E, Kodama T, Minoda Y. 1977 Studies on hydrogen utilizing microorganisms. 5. Isolation and culture conditions of thermophilic hydrogen bacteria. *Agric. Biol. Chem.* **41**, 685–690.

[69] Allewalt JP, Bateson MM, Revsbech NP, Slack K, Ward DM. 2006 Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the Octopus Spring microbial mat community of Yellowstone National Park. *Appl. Environ. Microbiol.* **72**, 544–550. (doi:10.1128/AEM.72.1.544-550.2006)

[70] Kamagata Y, Kawasaki H, Oyaizu H, Nakamura K, Mikami E, Endo G, Koga Y, Yamasato K. 1992 Characterization of three thermophilic strains of *Methanothrix* ("*Methanosaeta*") *thermophila* sp. nov. and rejection of *Methanothrix* ("*Methanosaeta*") *thermoacetophila*. *Int. J. Syst. Bacteriol.* **42**, 463–468. (doi:10.1099/00207713-42-3-463)

[71] Yamaoka T, Satoh K, Katoh S. 1978 Photosynthetic activities of a thermophilic blue-green alga. *Plant Cell Physiol.* **19**, 943–954.

[72] Shiratori H, Sasaya K, Ohiwa H, Ikeno H, Ayame S, Kataoka N, Miya A, Beppu T, Ueda K. 2009 *Clostridium clariflavum* sp. nov. and *Clostridium caenicola* sp. nov., moderately thermophilic, cellulose-/cellobiose-digesting bacteria isolated from methanogenic sludge. *Int. J. Syst. Evol. Microbiol.* **59**, 1764–1770. (doi:10.1099/ijs.0.003483-0)

[73] Darland G, Brock TD. 1971 *Bacillus acidocaldarius* sp.nov., an acidophilic thermophilic spore-forming bacterium. *J. Gen. Microbiol.* **67**, 9–15. (doi:10.1099/00221287-67-1-9)

[74] Balk M, Weijma J, Stams AJM. 2002 *Thermotoga lettingae* sp nov., a novel thermophilic, methanol-degrading bacterium isolated from a thermophilic anaerobic reactor. *Int. J. Syst. Evol. Microbiol.* **52**, 1361–1368. (doi:10.1099/ijs.0.02165-0)

[75] Wu DY *et al.* 2009 Complete genome sequence of the aerobic CO-oxidizing thermophile *Thermomicrobium roseum*. *PLoS ONE* **4**, e4207. (doi:10.1371/journal.pone.0004207)

[76] Beeder J, Nilsen RK, Rosnes JT, Torsvik T, Lien T. 1994 *Archaeoglobus fulgidus* isolated from hot North Sea oil field waters. *Appl. Environ. Microbiol.* **60**, 1227–1231.

[77] Jones WJ, Leigh JA, Mayer F, Woese CR, Wolfe RS. 1983 *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol.* **136**, 254–261. (doi:10.1007/BF00425213)

[78] Zillig W, Gierl A, Schreiber G, Wunderl S, Janekovic D, Stetter KO, Klenk HP. 1983 The archaebacterium *Thermofilum pendens* represents a novel genus of the thermophilic, anaerobic sulfur respiring *Thermoproteales*. *Syst. Appl. Microbiol.* **4**, 79–87. (doi:10.1016/S0723-2020(83)80035-6)

[79] Fiala G, Stetter KO, Jannasch HW, Langworthy TA, Madon J. 1986 *Staphylothermus marinus* sp. nov. represents a novel genus of extremely thermophilic submarine heterotrophic archaebacteria growing up to 98 °C. *Syst. Appl. Microbiol.* **8**, 106–113. (doi:10.1016/S0723-2020(86)80157-6)

[80] González JM, Masuchi Y, Robb FT, Ammerman JW, Maeder DL, Yanagibayashi M, Tamaoka J, Kato C. 1998 *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles* **2**, 123–130. (doi:10.1007/s007920050051)

[81] Fiala G, Stetter KO. 1986 *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.* **145**, 56–61. (doi:10.1007/BF00413027)

[82] Wicken JS. 1987 *Evolution, Thermodynamics, and Information.* New York, NY: Oxford University Press.

[83] Akashi H, Gojobori T. 2002 Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **99**, 3695–3700. (doi:10.1073/pnas.062526999)

[84] Sella G, Hirsh AE. 2005 The application of statistical physics to evolutionary biology. *Proc. Natl Acad. Sci. USA* **102**, 9541–9546. (doi:10.1073/pnas.0501865102)

[85] Lombardi AT, Garcia Jr O. 2002 Biological leaching of Mn, Al, Zn, Cu and Ti in an anaerobic sewage sludge effectuated by *Thiobacillus ferrooxidans* and its effect on metal partitioning. *Water Res.* **36**, 3193–3202. (doi:10.1016/S0043-1354(02)00008-8)

[86] Miller SR. 2003 Evidence for the adaptive evolution of the carbon fixation gene *rbcL* during diversification in temperature tolerance of a clade of hot spring cyanobacteria. *Mol. Ecol.* **12**, 1237–1246. (doi:10.1046/j.1365-294X.2003.01831.x)

[87] McDonald JH. 2010 Temperature adaptation at homologous sites in proteins from nine thermophile-mesophile species pairs. *Genome Biol. Evol.* **2**, 267–276. (doi:10.1093/gbe/evq017)

[88] Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. 2003 TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* **19**, 671–672. (doi:10.1093/bioinformatics/btg043)