Redox potential linked to water loss from proteins in evolution and development

Jeffrey M. Dick [1,*]

[1] Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Ministry of Education, School of Geosciences and Info-Physics, Central South University, Changsha, China

[*] Correspondence: jeff@chnosz.net

Preprint compiled on January 29, 2021

Abstract: As gene sequences change through evolution, and as the abundances of different proteins change through development, the distinct elemental composition of the proteins at different times can be represented as an overall chemical reaction. Compositional and thermodynamic analysis of these reactions leads to novel insight on biochemical changes and enables predictions of intensive variables including redox potential. The stoichiometric hydration state refers to the number of H_2O (n_{H_2O}) in theoretical reactions to form the proteins from a set of thermodynamic components. By analyzing published phylostratigraphy and transcriptomic and proteomic datasets, I found that n_{H_2O} of proteins decreases on evolutionary timescales (from single-celled organisms to metazoans) and on developmental timescales in *Bacillus subtilis* biofilms. Moreover, values of n_{H_2O} computed for a developmental proteome of fruit flies are aligned with organismal water content from larva to adult stages. I present a thermodyamic model for the equilibrium chemical activity of target proteins in a genomic background. Conditions that maximize the activity of the target proteins are found by optimizing the values of water activity and oxygen fugacity, which are then combined to calculate effective values of redox potential (Eh). The effective Eh values during evolution range between values reported for mitochondria, the cytosol, and extracellular compartments. These results suggest a central role for water, and water activity, in the biochemistry of evolution and development.

Keywords: geobiochemistry; gene ages; biofilm development; *Drosophila melanogaster*; water activity; redox potential

1. Introduction

Evolutionary developmental biology (evo-devo) addresses the evolution of developmental mechanisms across organisms. Even in organisms with similar genetic makeup, differences in gene regulation lead to phenotypic variation and as a result affect the evolvability of particular biological lineages [1]. A well known example from evo-devo regards the spatial and temporal expression patterns of homeobox (Hox) genes, which regulate the morphogenesis of animals [2]. In these systems, there is a need to better understand the expression dynamics of large numbers of genes and proteins. This question has motivated the production of genome-wide transcriptomic and proteomic datasets in model organisms including animals and bacteria [3,4]. However, integration of these datasets with primary biochemical features has lagged. Water content is one of the most basic biochemical characteristics that changes through development [5,6]. In this study, I apply a compositional analysis to proteomic data to gain information about the stoichiometry of water in biochemical reactions at the proteome level, which can be related to measured water content. Subsequently, a thermodynamic analysis reveals the influence of water activity on the relative stabilities of proteins and leads to a new model for biochemical redox potentials over evolutionary and developmental time scales.

Water activity can be defined as the water vapor pressure in a system divided by the vapor pressure of pure water at the same temperature (e.g. [7–9]). In origin-of-life research, geological environments with low water activity have been proposed to reduce or overcome the energetic barriers to polymerization of biomolecules in an aqueous environment [10]. This concept has received renewed attention recently for serpentinizing systems, which could not only provide redox disequilibria to drive the abiotic synthesis of various organic molecules, but also provide pore spaces with reduced water activity [11]. Among extant organisms, some species of fungi and bacteria are capable of growth at environmental water activities (a_w) of 0.65 or lower [12]; a lowering of extracellular a_w is also associated with decreased cytoplasmic water content and activity [13,14]. The control of water activity is important for food microbiology [8], and the characterization of water activity in extraterrestrial settings helps to identify promising targets for exploration in astrobiology [7,15].

Examples from other areas of biology highlight the relevance of dynamic water content in biological systems. There is a progressive loss of water during development from prenatal to adult forms in mammals [6,16,17]. Conversely, relatively high water content has been recognized for over a century as a biochemical characteristic of cancer tissue [18–22], and some authors have highlighted parallel trends of higher water content in both cancer and embryonic tissue [23–25]. Models are available to predict water activity from the water content of foods [8], but there is a pressing need to relate observations of water content in normal and pathological states to water activity ([26]; see also [24, p. 189]), which is a better thermodynamic measure of the potential for hydration in biochemical reactions.

Although biomolecular conformations – i.e. the structures resulting largely from non-covalent interactions – have for some time been known to be affected by reduced water activity associated with the macromolecular crowding that is a characteristic of cellular interiors [9,27,28], the possibility that water activity might also be a controlling factor in covalent biochemical reactions has been largely overlooked. This assumption explicitly enters many biochemical and geobiochemical models that set a_{H_2O} = 1 in thermodynamic calculations [29–31].

In contrast, many petrological studies use thermodynamic models that link water activity with observable mineral assemblages in rocks. The consideration of both water activity and oxygen fugacity (a thermodynamic measure of oxidation potential) is essential for understanding melting and magmatic processes [32] as well as lower-temperature metasomatic processes including serpentinization [33]. Notably, there is a straightforward conversion between oxygen fugacity and redox potential expressed in the Eh scale. For instance, partial pressures of O_2 between 10^{-40} and $10^{-83.1}$ correspond to Eh values between approximately -70 and -410 mV at pH 7, 25 °C, and a_{H_2O} = 1 [34, p. 176].

A basic assumption for geochemistry is that knowledge of the initial and final states of a system is sufficient to describe a process that can be represented by thermodynamic models [35, p. 50]. Similarly, some models for the energetics of biomass synthesis in early Earth systems consider overall anabolic reactions from inorganic precursors without reference to the actual mechanisms that may be involved [36]. It follows that from a geochemical perspective, changes in protein identity and abundance – whether through evolution, ontogeny, or in cell culture experiments – would be best represented as an overall chemical reaction. This type of approach could be described as "geobiochemistry" to distinguish it from the textbook version of biochemistry. The traditional biochemical approach conceptualizes proteins as "molecular machines" that catalyze and control metabolic reactions [30,37], but does not step back to take a broader view of chemical changes in the proteome itself.

By applying a compositional analysis to proteomic data, it can be shown that water is consumed as a reactant in the differential expression of proteins of many cancer types compared to normal tissues [26]. This observation suggests that increasing water activity may be a driver for the observed biochemical changes at the proteome level, but a thermodynamic model should be developed to quantify this prediction.

In this study, I further develop the biological motivation and theoretical foundation for applying thermodynamics to patterns of protein occurrence and abundance. This is done through compositional and thermodynamic analysis of proteins associated with phylostrata, which represent the origin of genes at particular times in evolution. I also explore patterns of protein expression in developmental model systems, including biofilms and fruit flies. The common pattern that emerges is water loss from proteins over various timescales. The results are consistent with the hypothesis that physicochemical conditions exert a major influence on proteome dynamics during evolution and development.

2. Results

2.1. Compositional analysis of proteins in evolution

Several studies have associated patterns of gene expression in cancer with phylogenetically earlier genes [38,39]. The phylostratigraphic analysis used in these studies assigns ages of genes based on the latest common ancestor whose descendants have all the computationally detected homologs of that gene. To analyze the evolutionary trends of oxidation and hydration state of proteins, I used 16 phylostrata (PS) for human protein-coding genes given by Trigos et al. [38]. The mean lengths of proteins coded by genes in each phylostratum are plotted in Fig. 1A. There is an initial rise in protein length, leading up to Eukaryota, which is consistent with the previously reported greater median protein length in eukaryotes than prokaryotes [40]. The large decrease of protein length in later phylostrata could in part be an artifact of BLAST-based homology searches [41]. Other studies have detected generally shorter sequences for younger genes; whether this reflects a significant source of phylostratigraphic bias should be kept in mind [42].

The abundances of elements in the primary sequence of proteins (C, H, N, O, S) can be represented as 5-dimensional compositional vector. To visualize particular projections of this compositional space, it is helpful to use compositional metrics with meaningful chemical definitions. The metrics used here are carbon oxidation state (Z_C) and stoichiometric hydration state (n_{H_2O}). The carbon oxidation state represents the average charge on all carbon atoms in the molecule, given nominal charges of the other atoms (H^{+1}, N^{-3}, O^{-2}, S^{-2}). Assuming that the heteroatoms (N, O, S) are bonded only to H or C, and not to each other, the carbon oxidation state of amino acids and proteins can be computed directly from the elemental abundances (note that this assumption precludes consideration of disulfide bonds and some types of post-translational modifications) [43,44]. In contrast, the stoichiometric hydration state is the coefficient on H_2O in mass-balanced reactions representing the theoretical formation of the protein from a set of thermodynamic components (also termed basis species). So that n_{H_2O} and Z_C can be viewed as independent variables, it is important to choose basis species where their covariation is reduced. Accordingly, the basis species glutamine, glutamic acid, cysteine, H_2O, and O_2 (denoted "QEC") were chosen for this analysis [26,44].

Fig. 1A reveals distinct evolutionary patterns of oxidation state and hydration state of proteins. Z_C forms a strikingly smooth hump between PS 1 and 11 then increases rapidly to the maximum at PS 14, followed by a smaller decline to PS 16, which corresponds to *Homo sapiens*. In contrast, n_{H_2O} shows an overall decrease through time, although there are notable positive jumps between PS 3 and 4 and PS 7 and 8.

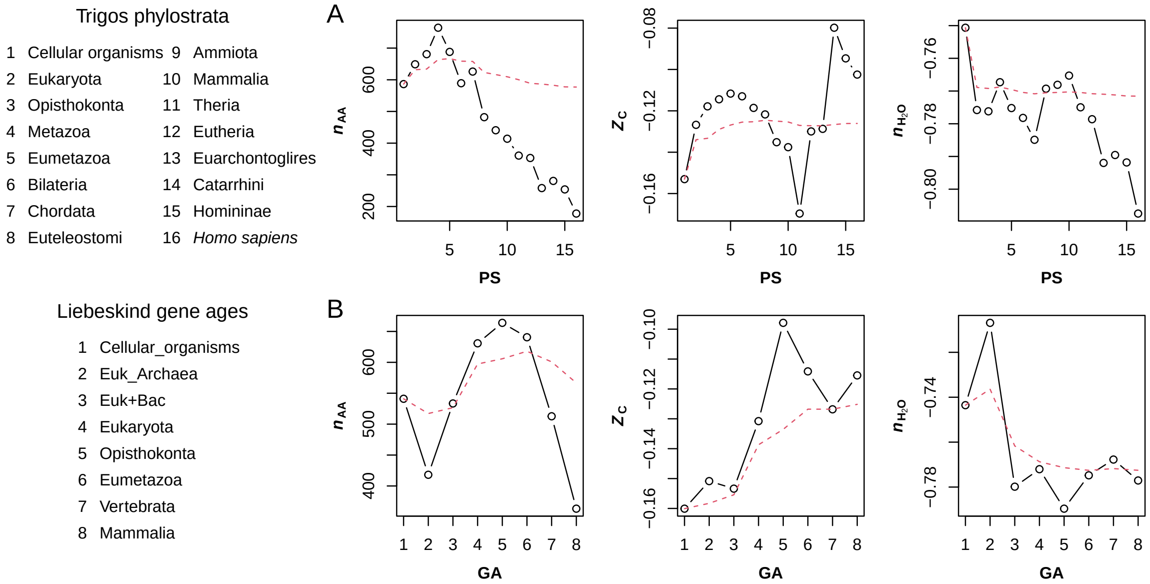

Figure 1. Compositional analysis of proteins in evolution. (**A**) Mean values of n_{AA}, Z_C, and n_{H_2O} of proteins for all protein-coding genes in each phylostratum (PS) given by Trigos et al. [38]. The points stand for the mean values for individual phylostrata, and the red line indicates the cumulative mean starting from PS 1. (**B**) Compositional metrics calculated using gene ages (GA) given by Liebeskind et al. [45]. The latest gene age is Mammalia, which corresponds to Trigos PS 10.

I also considered proteins grouped into eight gene ages (GA) reported by Liebeskind et al. [45] based on consensus tables for different age-estimation algorithms. Compared to the Trigos phylostrata, the Liebeskind gene ages have three steps between cellular organisms and Eukaryota, providing a greater resolution in earlier evolution, and stop at Mammalia, which corresponds to Trigos PS 10. Keeping in mind the different resolutions and scales of the Trigos phylostrata and Liebeskind gene ages, the two datasets show similar maxima for Z_C and protein length near Eumetazoa (or Opisthokonta, which is not one of the Trigos phylostrata), and an overall decrease of n_{H_2O} during evolution (Fig. 1B).

2.2. Theoretical prediction of relative stabilities of proteins

Up to now, we have seen the chemical composition of proteins represented in terms of selected compositional metrics: oxidation and hydration state. How can these metrics be related to environmental conditions: oxidation and hydration potential? Here I describe a thermodynamic method for predicting the environmental oxidation and hydration potentials that stabilize particular proteins compared to others. Stability in this context refers not to protein conformation (i.e. 3-dimensional structure determined by non-covalent interactions), but to the energetics of the overall *formation* of proteins from the basis species. The theoretical formation reactions considered here represent the distinct elements and covalent bonds (i.e. specific amino acid residues) in the primary sequences of different proteins.

To perform this calculation, I used the chemical affinity (A), which is the opposite of the non-standard Gibbs energy change of the reaction [46, p. 143] and can be computed from (e.g. [47])

$$A = -\Delta G = 2.303RT\log(K/Q) \tag{1}$$

where K is the equilibrium constant and Q is the activity product for the formation reaction for a particular protein. The right-hand side is multiplied by the natural logarithm of 10 (≈2.303) so that all other logarithmic values are common logarithms. Because it includes the chemical activities of all the species in the reaction, Q incorporates the sensitivity to oxidation and hydration potential, which are represented by oxygen fugacity ($\log f_{O_2}$) and water activity ($\log a_{H_2O}$). On the other hand, K is a function of the standard Gibbs energy of the reaction and therefore of temperature and pressure.

An example of a balanced formation reaction is shown below for chicken egg-white lysozyme (UniProt: LYSC_CHICK).

$$0.5144\mathrm{Gln} + 0.3892\mathrm{Glu} + 0.0780\mathrm{Cys} = \mathrm{C_{4.752}H_{7.434}N_{1.496}O_{1.434}S_{0.078}} + 0.8794\mathrm{H_2O} + 0.4713\mathrm{O_2} \tag{2}$$

The chemical formula of the whole protein ($C_{613}H_{959}N_{193}O_{185}S_{10}$) is divided by the length (129) to give the per-residue formula that is a product of the reaction. The only other species in the reaction are the basis species chosen to project the elemental composition into chemical space: glutamine ($C_5H_{10}N_2O_3$), glutamic acid ($C_5H_9NO_4$), cysteine ($C_3H_7NO_2S$), H_2O, and O_2. The standard Gibbs energy ($\Delta G_f°$) of the protein calculated using amino acid group additivity is also divided by the protein length to give the per-residue $\Delta G_f°$, which is combined with the standard Gibbs energies of the other species in the reaction to calculate the standard Gibbs energy of reaction ($\Delta G_r°$), and from that, $\log K$. By using the `subcrt()` function in the CHNOSZ package [48], the $\log K$ for Reaction 2 is computed to be -39.84. The methods for amino acid group additivity for proteins are described in [49]. It would be possible to include pH effects in this model by considering ionization of protein sidechain and terminal groups [49,50], but in order to focus on the contributions of hydration and oxidation potential, the present calculations are only concerned with proteins treated as neutral species.

Calculation of the chemical affinities requires values for the activities of all species in the reaction. $\log a_{H_2O}$ and $\log f_{O_2}$ are used here as exploratory variables, so they were assigned a range of values at equal intervals in order to construct a 2-dimensional grid that is used for plotting the relative stabilities of the proteins. The chemical activities of the other basis species were assigned using mean concentrations of amino acids in human plasma [51]; expressed as logarithms of concentrations in mol/l, these are -3.2 for glutamine, -4.5 for glutamic acid, and -3.6 for cysteine (-3.6) [52]. The initial (non-equilibrium) activity of the per-residue formula for each protein was set to unity; the equilibrium activities were calculated as described next.

The relative stabilities of proteins (all represented by per-residue formulas) were quantified using the Boltzmann distribution written as [50]

$$\frac{a_i}{\sum a_i} = \frac{e^{A_i/RT}}{\sum e^{A_i/RT}} \tag{3}$$

where a is activity, A is affinity, and i designates a single protein in a system of any number of proteins. For convenience in the calculation, the total activity in the result ($\sum a_i$) is fixed at unity; this in combination with the assumption of unit activity coefficients means that the a_i for each protein represents its fractional degree of formation in equilibrium with all other proteins. The predominant protein is the one with the highest predicted activity, but all the proteins in the equilibrium model have finite values of activity. The next step is to find the conditions that maximize the activity of particular target proteins in a model system.

2.3. *Maximizing stabilities of target proteins on a genomic background*

In order to thermodynamically characterize the changes in protein composition between phylostrata, model proteins for each phylostratum (referred to here as *PS model proteins*) were generated by computing the mean amino acid composition of all proteins in each phylostratum. Equilibrium calculations for the 16 PS model proteins for the Trigos phylostrata are displayed on a $\log a_{\mathrm{H_2O}} - \log f_{\mathrm{O_2}}$ diagram (Fig. 2A). Predominance fields for only few proteins are visible on the diagram; the other model proteins with lower activity are “hiding” under the plane of the diagram. As indicated both for the predominant proteins (by the colored fields in the diagram) and for all 16 PS model proteins (by the points), the activities maximize at extreme values of $\log a_{\mathrm{H_2O}}$ and $\log f_{\mathrm{O_2}}$; those for the predominant proteins actually maximize at infinite values of $\log a_{\mathrm{H_2O}}$ and/or $\log f_{\mathrm{O_2}}$. Therefore, it is not possible to use this model system to find particular values of water activity and oxygen fugacity that characterize each phylostratum.

What happens if we add to the system many different human proteins in addition to the 16 PS model proteins? It is likely that some of the human proteins will be more stable than the PS model proteins, so that none of the latter will predominate. The large number of human protein sequences (e.g. 16,974 in the Trigos phylostrata dataset that can be mapped to the UniProt database) prevents running the complete calculation on a laptop computer with 16 GB of RAM, so random samples of human proteins were used here. The calculated predominance diagram for a system of 16 PS model proteins together with 200 randomly sampled human proteins reveals that a relatively small number

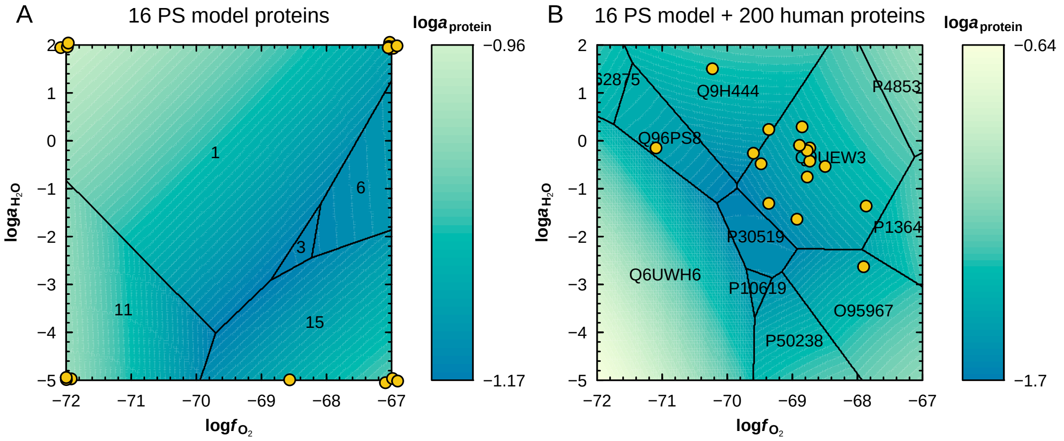

Figure 2. Strategy for deriving theoretical values of water activity (log$a_{\mathrm{H_2O}}$) and oxygen fugacity (log$f_{\mathrm{O_2}}$) that maximize the predicted equilibrium activity of PS model proteins. The colors represent the activities of the predominant proteins and the lines represent equal activities for the predominant proteins. (**A**) Equilibrium calculations include 16 PS model proteins derived from the Trigos phylostrata. Only five proteins predominate at different ranges of log$a_{\mathrm{H_2O}}$ and log$f_{\mathrm{O_2}}$; all the others have lower activity. The yellow circles indicate the location for the maximum activity of all of the PS model proteins; a small amount of jitter is added to uncover overlapping points. (**B**) Equilibrium calculations include the same 16 PS model proteins (target proteins) and 200 randomly sampled proteins from the human proteome (background proteins). All the predominant proteins are predicted to come from the background population, and are labeled with their UniProt IDs. The target proteins have lower activities that are maximized at the conditions indicated by the yellow circles.

of human proteins predominate in equilibrium. The 16 PS model proteins are among the less stable proteins, and their activities are maximized at particular values of log$a_{\mathrm{H_2O}}$ and log$f_{\mathrm{O_2}}$ (Fig. 2B).

2.4. Maximum activity analysis applied to phylostrata

The method described above was used to predict optimal values of water activity and oxygen fugacity to maximize the activity of the PS model proteins (the *target proteins*) against the background of thousands of human proteins. The values of log$a_{\mathrm{H_2O}}$ and log$f_{\mathrm{O_2}}$ that maximize the activity of 16 PS model proteins are shown in Fig. 3A and B. In a single calculation, a random sample of 2000 human proteins was used. The sampling and equilibrium calculations were performed 100 times; the means of all runs are represented by the red lines in Fig. 3A and B. The mean values of log$a_{\mathrm{H_2O}}$ are also plotted in Fig. 3C, with droplines at Cellular organisms, Eukaryota, Eumetazoa, and Mammalia to indicate coinciding gene ages in the Liebeskind dataset.

It is not surprising to find that the trends in hydration and oxidation potentials depicted in Fig. 3A and B are similar to those in the corresponding compositional metrics, $n_{\mathrm{H_2O}}$ and Z_{C}, in Fig. 1. For instance, the earliest phylostratum (cellular organisms) has both the highest $n_{\mathrm{H_2O}}$ and log$a_{\mathrm{H_2O}}$, and PS 1–10 hover around moderate values of these quantities. According to the thermodynamic model, this corresponds to about log$a_{\mathrm{H_2O}}$ = 0, indicated by the horizontal dashed line in Fig. 3C. After PS 10 (Mammalia), there is a decrease in both $n_{\mathrm{H_2O}}$ and log$a_{\mathrm{H_2O}}$. However, although PS 16 (*Homo sapiens*)

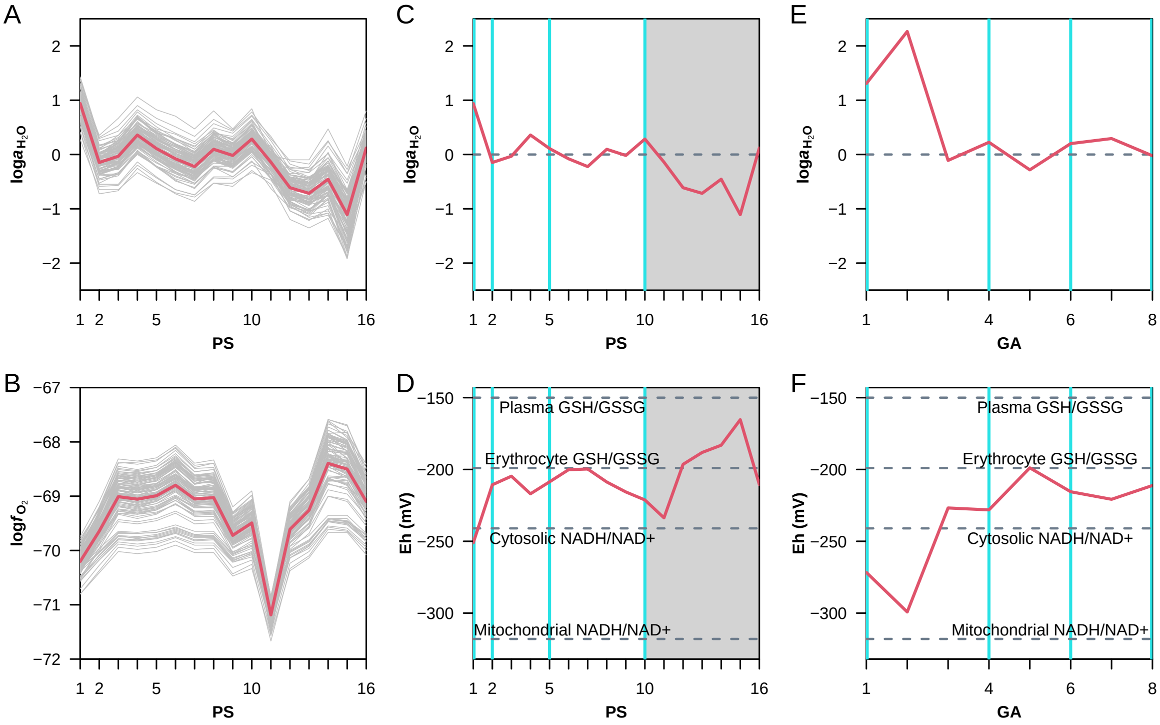

Figure 3. Analysis of optimal water activity and oxygen fugacity for phylostrata, and calculation of theoretical redox potential (Eh). (**A**) and (**B**) Values of $\log a_{H_2O}$ and $\log f_{O_2}$ that maximize the activity of PS model proteins for the Trigos phylostrata in equilibrium with each other and 2000 randomly sampled proteins from the human proteome. Each thin gray line represents a calculation for one random sample, and thick red lines show the means for 100 calculations. (**C**) Comparison of mean values of $\log a_{H_2O}$ with pure water (dashed horizontal line at $\log a_{H_2O}$ = 0). Light blue vertical lines represent phylostrata (Trigos PS 1, 2, 5, and 10) that correspond to gene ages in the Liebeskind dataset (Cellular organisms, Eukaryota, Eumetazoa, and Mammalia). The shaded gray area represents post-Mammalia ages, which are not available in the Liebeskind dataset. (**D**) Values of Eh calculated from the mean values of $\log a_{H_2O}$ and $\log f_{O_2}$ using Eqs. (5–6). Dashed horizontal lines represent redox potentials for different cellular compartments and reactions [53,54]. (**E**) and (**F**) Analogous calculations for $\log a_{H_2O}$ and Eh using the Liebeskind gene ages; vertical lines at GS 1, 4, 6, and 8 represent Cellular organisms, Eukaryota, Eumetazoa, and Mammalia.

has the lowest n_{H_2O} of any phylostratum (Fig. 1A), the thermodynamic model predicts a small rise in $\log a_{H_2O}$ for PS 16 compared to PS 15. Looking at the oxidation trends, PS 11 (Theria) has both the lowest Z_C and lowest $\log f_{O_2}$.

It should be kept in mind that oxygen fugacity is a thermodynamic quantity that implies nothing about the actual mechanism of oxidation or reduction [35]. For instance, in geological systems where there is no free O_2, the processes that actually provide oxygen come from other reactions in the environment [55]. A practical use of oxygen fugacity is as a thermodynamic parameter that can be used to calculate other parameters that are easier to measure [56, p. 245]. Likewise, the theoretical values of $\log a_{H_2O}$ and $\log f_{O_2}$ shown here simply represent thermodynamic measures of hydration and oxidation potential that maximize the activities of the target proteins. The practical value of these parameters is demonstrated below by combining them to calculate Eh, which is a more common scale of redox potential in biochemistry.

2.5. Effective redox potential and implications for early cellular evolution

Effective values of redox potential (Eh) can be obtained by considering the half-cell reaction for H_2O:

$$\mathrm{H_2O} = 0.5\mathrm{O_2} + 2\mathrm{H}^+ + 2e^- \tag{4}$$

At equilibrium, $K = Q$, where K and Q are the equilibrium constant and activity product. For Reaction 4, this gives

$$\mathrm{pe} = 0.25 \log f_{\mathrm{O_2}} - \mathrm{pH} - 0.5 \log a_{\mathrm{H_2O}} - 0.5 \log K \tag{5}$$

where $\mathrm{pe} = -\log a_{e^-}$ and $\mathrm{pH} = -\log a_{\mathrm{H^+}}$. Values of Eh can then be calculated using

$$\mathrm{Eh} = \frac{2.303RT\mathrm{pe}}{F} \tag{6}$$

where R, T, and F are the gas constant, temperature, and Faraday constant.

Fig. 3D shows theoretical values of Eh for the PS model proteins calculated using Eqs. (5–6) with pH = 7. The effective Eh is a composite variable; it is elevated by either increasing $\log f_{\mathrm{O_2}}$ or decreasing $\log a_{\mathrm{H_2O}}$. As with $\log f_{\mathrm{O_2}}$, Eh exhibits a broad hump between PS 1 and 11, but the whole profile is tilted up, reflecting the overall evolutionary decrease of $\log a_{\mathrm{H_2O}}$. Several measurements for selected redox pairs in cells and plasma are shown for comparison [53,54]. The PS 1–11 hump begins and ends close to the cytosolic Eh of the NADH/NAD$^+$ redox pair and maximizes near the Eh for cytosolic GSH/GSSG measured in erythrocytes. Between PS 12 and 15 there is a rapid rise toward Eh values characteristic of GSH/GSSG in plasma, followed by a return in PS 16 to the redox potential for cytosolic GSH/GSSG.

The first four Liebeskind consensus gene ages [45] are cellular organisms (GA 1), Euk_Archaea (GA 2; the common ancestor of Eukaryota and Archaea), Euk+Bac (GA 3; genes present only in Eukaryotes and Bacteria, representing horizontal transfer to Eukaryotes), and Eukaryota (GA 4), and therefore provide greater resolution over these stages of evolution than the Trigos phylostrata. Figs. 3E and F show the results for the maximum activity analysis applied to the Liebeskind gene ages. There is an increase of both $n_{\mathrm{H_2O}}$ (Fig. 1B) and $\log a_{\mathrm{H_2O}}$ between GA 1 and 2, and these values are higher than those for all subsequent gene ages. The latter hover near $\log a_{\mathrm{H_2O}} = 0$ and have effective Eh values that range between cytosolic NADH/NAD$^+$ and GSH/GSSG (Fig. 3F).

The Euk_Archaea group is of particular interest because it represents the common ancestor of Eukaryota and Archaea. Although proteins coded by Euk_Archaea genes (GA 2) are slightly more oxidized than those in GA 1 (Fig. 1B), their higher $n_{\mathrm{H_2O}}$ – and optimal $\log a_{\mathrm{H_2O}}$ – means that they are stabilized by lower redox potentials, close to -300 mV. This redox potential approaches that of the NADH/NAD$^+$ system in mitochondria [54]. The low redox potential suggested by these results could reflect a reductive overall cellular physiology, typical of archaeal cells, that operated before the endosymbiotic transfer of mitochondria [57]. Subsequently, subcellular conditions outside of the mitochondria could become more oxidizing; the release from a reductive chemistry might explain the rise in effective Eh at GA 3 and later.

2.6. Compositional and thermodynamic analysis of biofilm development

Temporal patterns of gene and protein expression during development have been documented for a growing number of organisms. A recent study reports data for *Bacillus subtilis* biofilms, which have been compared to developing embryos [4]. Those authors combined gene and protein abundances with phylostrata assignments to compute a transcriptome age index (TAI) and proteome age index (PAI) for timepoints in the biofilm development. Here I just used the reported abundances to calculate the weighted mean amino acid composition for proteins at each developmental stage, which are used as the model proteins for the compositional and thermodynamic analysis described below.

Futo et al. [4] described three periods of biofilm growth: early (6H–1D), mid (3D–7D), and late (1M–2M). Timepoints of 2D and 14D are regarded as transitional stages between these periods, and are marked by vertical lines in Fig. 4. In the early period of development, there is a steep decline in the mean protein length. Note that this is computed simply by combining the lengths of canonical protein sequences from the UniProt database with normalized gene or protein expression values reported by Futo et al. [4]; no phylostrata values are used in this or any of the following calculations. The late stage of biofilm development, where only transcriptomic data are available, shows another steep decline in mean length of the corresponding proteins (Fig. 4A).

I used the model proteins (i.e. mean amino acid compositions) at each timepoint to compute values of Z_{C} and $n_{\mathrm{H_2O}}$. The proteome-based model proteins exhibit decreasing $n_{\mathrm{H_2O}}$ in the early to mid developmental period (6H to 7D) (Fig. 4B). Except for an initial rise starting from the liquid culture (LC), the time course for the transcriptome-based model proteins also exhibits decreasing $n_{\mathrm{H_2O}}$. In the late developmental period, the transcriptome-based model proteins show an even greater decrease in $n_{\mathrm{H_2O}}$; unfortunately, proteomic data for this period are not available for comparison. Through early and mid development there is relatively little variation in Z_{C} of the model proteins, but the transcriptome-based model proteins become much more oxidized in the late period (Fig. 4C).

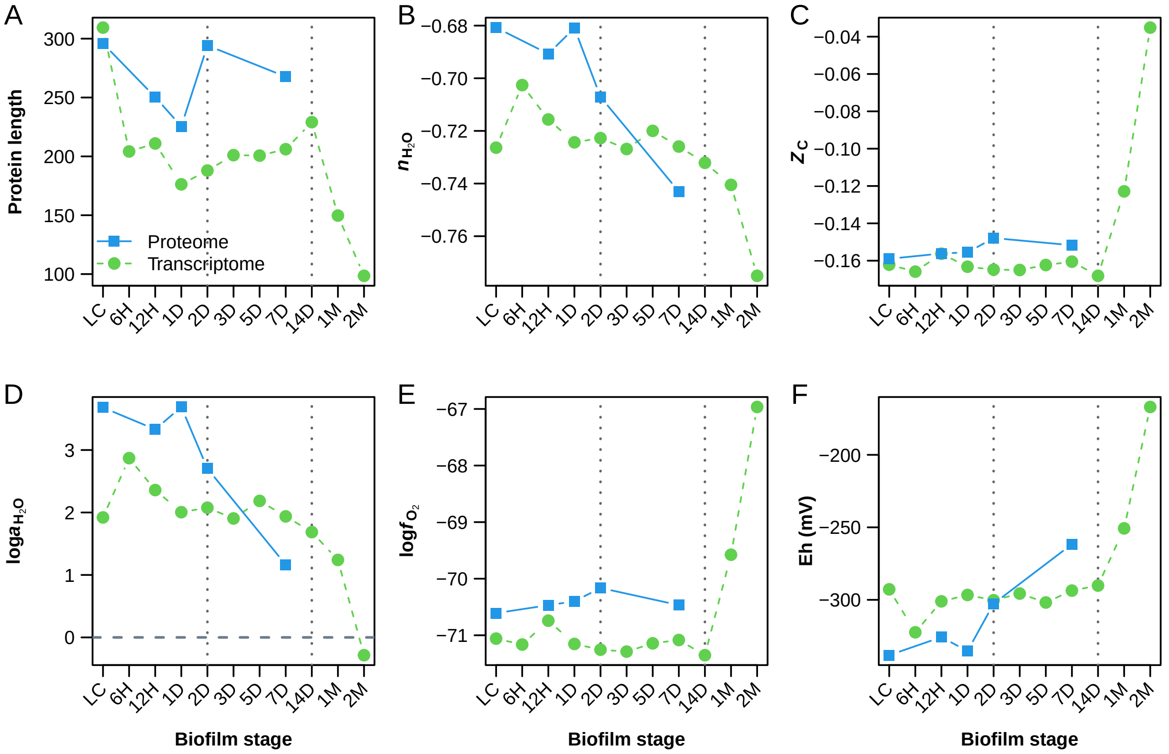

Figure 4. Compositional and thermodynamic analysis of model proteins in developing *B. subtilis* biofilms. Dotted lines represent transitions between early and mid periods (2D) and mid and late periods (14D). Model proteins at each timepoint have amino acid compositions that are the expression-weighted mean for reported normalized expression levels of genes or proteins [4]. (**A**) Protein length; (**B**) stoichiometric hydration state (n_{H_2O}); (**C**) carbon oxidation state (Z_C). (**D**) and (**E**) Optimal values of $\log a_{H_2O}$ and $\log f_{O_2}$ that maximize the equilibrium activity of the model proteins against a background of proteins from the human proteome. These are mean values for 100 samples, each consisting of 2000 randomly selected proteins from the human proteome. (**F**) Effective redox potential (Eh) computed from Eqs. (5)–(6) and the optimal values of $\log a_{H_2O}$ and $\log f_{O_2}$.

The trends of n_{H_2O} are closely reflected in values of $\log a_{H_2O}$ computed from the thermodynamic analysis. The values of $\log a_{H_2O}$ are positive at all but the last timepoint (Fig. 4D). This implies that the biofilm requires an elevated internal hydration potential for early growth, but the internal conditions of the biofilm eventually reach a near-equilibrium state with the aqueous medium, where the activity of H_2O is very close to unity.

Optimal values of $\log a_{H_2O}$ were combined with those of $\log f_{O_2}$ (Fig. 4E) to compute effective values of Eh using Eqs. (5)–(6). The effective values of Eh rise through development, particularly in the latter stages (Fig. 4E), which could be an indication of the attainment of more oxidizing conditions within the biofilm. The hypothesis that aging biofilms progressively become more oxidized could be tested with oxygen microelectrode and/or redox potential measurements. These types of measurements have been reported for some species such as *Geobacter sulfurreducens* [58], but I was unable to find any in the literature for aging *B. subtilis* biofilms.

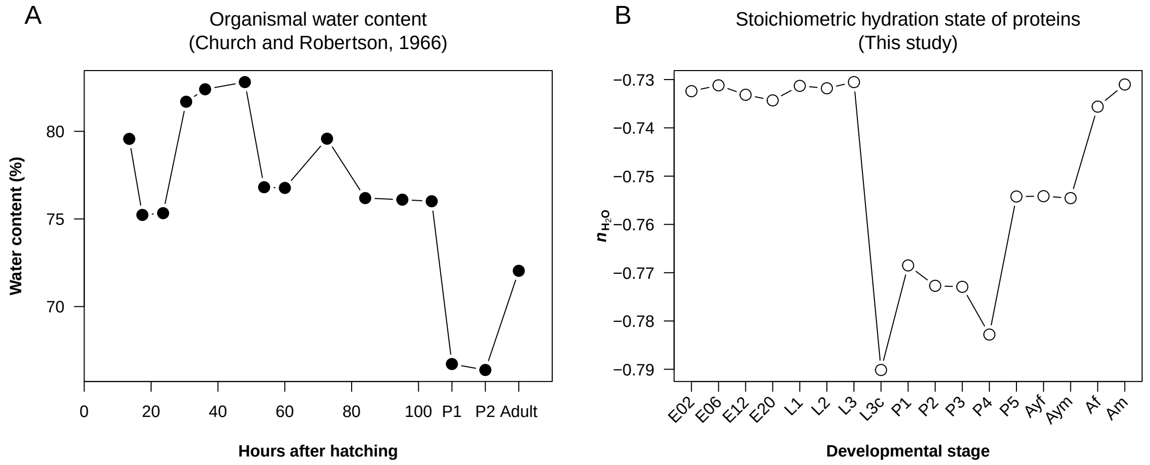

Figure 5. (**A**) Organismal water content during growth of fruit flies (*D. melanogaster*) from larvae (hours after hatching) to pupal and adult stages, replotted from Figure 4 of [5]. (**B**) Stoichiometric hydration state of proteins computed from the developmental proteome of [3]. Labels starting with "E", "L", "P", and "A" refer to embryogenesis, larval, pupal, and adult stages. "Ay" stands for young adult, and "f" and "m" indicate female and male adults.

2.7. Organismal water content and stoichiometric hydration state of proteins in development of fruit flies

The fruit fly, *Drosophila melanogaster*, is a widely used invertebrate model organism in genetics and developmental biology. A recent study provides proteomic data for developmental stages of *D. melanogaster*, including embryogenesis, larva, pupa, and adults [3]. The changes of water content and other biochemical constituents in the development of *D. melanogaster* from larval to adult stages, when grown on chemically defined axenic medium, were reported in [5]. As larvae progress through different instars (i.e. a few days post-hatching), the water content shows some variation around 80%. The water content then decreases sharply to 66% in the prepupal stage [5]. After this, the data of [5] show a rise of water content to greater than 70% in adults (Fig. 5A).

The stoichiometric hydration state of proteins computed for the fly developmental proteome is plotted in Fig. 5B. The $n_{\mathrm{H_2O}}$ is nearly constant during embryogenesis and three instars of larva (L1, L2, L3). There is a sudden drop to much lower $n_{\mathrm{H_2O}}$ in stage L3c, which is described as "L3 crawling larva" [3]. The pupa collected on different days (P1 to P5) exhibit a somewhat higher $n_{\mathrm{H_2O}}$, but still lower than the embryos. The $n_{\mathrm{H_2O}}$ rises in young adults and then again in old adults, which have $n_{\mathrm{H_2O}}$ values close to those of the embryos and early larva.

The strong decrease of proteomic $n_{\mathrm{H_2O}}$ in the crawling larva (L3c) is aligned with the decrease of water content in the prepupal stage, which is when the larva leaves the medium [5]. Likewise, the rise of water content in the adult fly to higher values (Fig. 5A), but less than that of the larva, is likely reflected in the trend of $n_{\mathrm{H_2O}}$ for young adults (Fig. 5B). No distinction was made between young and old adults in Ref. [5], so it is not possible to compare the continued rise of $n_{\mathrm{H_2O}}$ with their water content data. Overall, the $n_{\mathrm{H_2O}}$ computed from the proteome of developing *D. melanogaster* appears to be tightly coupled to organismal water content.

The somewhat higher $n_{\mathrm{H_2O}}$ of proteomes of adult males than those of females can be compared with water contents of 7 to 42 day old *D. melanogaster*, expressed as a percentage of total mass, calculated using data from Figure 4 of [59]. The percent water content for females ranges from 64.1 to 66.2, and that for males from 66.3 to 68.4. Similarly, in humans, adult males have a higher percent water content then females [17]. These observations suggest that the higher $n_{\mathrm{H_2O}}$ of proteomes of adult male flies in Fig. 5B reflects actual physiological differences between the sexes.

3. Discussion

The main findings of this study can be grouped into two themes: compositional and thermodynamic analysis. The results show that proteomes may be biochemically connected to the environment through multiple thermodynamic components (O_2 and H_2O) that vary over a range of timescales.

3.1. Compositional analysis

The compositional analysis uncovers decreasing stoichiometric hydration state of proteins in evolution, as represented by phylostratigraphic ages, and in development of *Bacillus subtilis* biofilms. The pattern for the developmental proteome of *Drosophila melanogaster* is not a uniform decrease, but suggests a more cyclical nature. The large decrease in $n_{\mathrm{H_2O}}$ at the crawling larva stage (L3c) is aligned with the measured prepupal decrease in organismal water content [5]. This strong association between measured water content and $n_{\mathrm{H_2O}}$ values computed from the proteome supports the biological relevance of the compositional analysis performed in this study.

By considering proteins as chemical species, this study emphasizes the importance of water loss as a major feature of both evolutionary and developmental processes. A recent analysis of proteins coded by environmental metagenomes points to decreased $n_{\mathrm{H_2O}}$ for particle-associated compared to free-living fractions from marine and freshwater samples [44]. Additionally, in cell culture, lower $n_{\mathrm{H_2O}}$ is characteristic of the proteomes of cells grown as 3D spheroids or aggregates, compared to 2D monolayers [26]. Taken together, these findings support the hypothesis that water loss, as measured by the stoichiometric hydration state of proteomes, is a feature of cell-cell interactions, and possibly more rigid cellular or multicellular structures.

3.2. Thermodynamic analysis

The main theoretical advance in this study follows from the observation that both O_2 and H_2O are involved in the water half-cell reaction. This is the rationale for using not only oxygen fugacity but also water activity as thermodynamic variables in the derivation of an effective redox potential expressed as Eh.

The analysis described here is related to thermodynamic models in geochemistry that involve "perfectly mobile components", which are represented by chemical potentials, instead of components defined by bulk composition [33,60]. By treating the chemical potentials of O_2 and H_2O as exploratory variables, optimal values can be found to maximize the predicted chemical activity of the target model proteins in a genomic background. These potentials are then combined with the law of mass action for water half-cell reaction to calculate an effective redox potential (Eh). A notable finding is that the model protein for the putative common ancestor of eukaryotes and Archaea is characterized by effective Eh values that approach those of the NADH/NAD$^+$ redox couple in present-day mitochondria. With further developments, this method could lead to a new way of looking at the evolutionary trends of protein composition that can uncover clues about subcellular chemical conditions in the past.

A possible concern about the thermodynamic analysis is the large range of water activity values in the model. The theoretical values of a_{H_2O} reach much lower values than saturated salt solutions (e.g. saturation of NaCl corresponds to 0.755 water activity [12]). At the other extreme, the theoretical values can be greater than 1, which represents an unphysical condition since pure water has unit activity. It may be possible to obtain $a_{H_2O} > 1$ in molecular dynamics simulations of mixture of H_2O and organic media due to oversaturation of H_2O and cluster formation in a nonpolar solvent, but such results were regarded as anomalous by other authors [61].

One finding that may serve as a "reality check" is that the systems analyzed here tend toward $a_{H_2O} = 1$ with time (see Fig. 3E and Fig. 4D). This tendency suggests a buffering effect, whereby after a relatively long time (in either evolution or development) proteomes naturally adjust toward equilibrium in a largely aqueous system.

4. Materials and Methods

All figures were created in R [62] using the contributed packages canprot version 1.1.0 [26] (available on the Comprehensive R Archive Network (CRAN)) and CHNOSZ [48] (development version > 1.4.0, which is available at https://r-forge.r-project.org/projects/chnosz/). Specifically, functions in the canprot package were used for calculating compositional metrics from amino acid compositions of proteins, and CHNOSZ was used for the thermodynamic calculations. The code to make the figures for this paper is available in the "evdevH2O.R" file and "evdevH2O" vignette in the JMDplots package version 1.2.5 [63].

Phylostrata were obtained from the supporting information of Trigos et al. [38] and the "main_HUMAN.csv" file of Liebeskind et al. [45,64]. Liebeskind et al. did not use phylostrata numbers, so gene ages 1 (oldest) to 8 (most recent) were assigned here (see Fig. 1B) corresponding to the names in the "modeAge" column of the source file. The Ensembl gene identifiers in the Trigos dataset were converted to UniProt accession numbers using the UniProt mapping tool [65]; in the case of duplicate UniProt accession numbers, the first matching phylostratum was used. The files with phylostrata assignments and UniProt IDs are available in the canprot package.

Transcriptomic and proteomic data for growing *B. subtilis* biofilms were taken from Supplementary file S10 of [4], specifically the tables named "Input values for calculating TAI" and "Input values for calculating PAI". The values were used without modification. Data for the *Drosophila* developmental proteome were extracted from Supplemental Table S1 of [3]. The values in the columns for imputed $\log_2$ LFQ intensity were exponentiated, then mean values were computed for each time point (4 replicates). For both the *B. subtilis* and *Drosophila* datasets, protein IDs were mapped using the UniProt mapping tool [65]. The canonical protein sequences were downloaded from UniProt, and the `read.fasta()` function in the CHNOSZ package was used to generate the amino acid compositions of the proteins. These were combined with the transcriptomic or proteomic abundances to compute weighted means for amino acid composition at each time point. The JMDplots package has the computed mean amino acid compositions, which are used as input for making the figures in this paper. The intermediate files (transcriptomic or proteomic abundances with UniProt mappings, and amino acid compositions of proteins) and R scripts to generate the mean amino acid compositions are available separately (https://github.com/jedick/devodata).

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Laland, K.N.; Uller, T.; Feldman, M.W.; Sterelny, K.; Müller, G.B.; Moczek, A.; Jablonka, E.; Odling-Smee, J. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings of the Royal Society B: Biological Sciences* **2015**, *282*, 20151019. doi:10.1098/rspb.2015.1019.
2. De Robertis, E.M. Evo-devo: variations on ancestral themes. *Cell* **2008**, *132*, 185–195. doi:10.1016/j.cell.2008.01.003.
3. Casas-Vila, N.; Bluhm, A.; Sayols, S.; Dinges, N.; Dejung, M.; Altenhein, T.; Kappei, D.; Altenhein, B.; Roignant, J.Y.; Butter, F. The developmental proteome of *Drosophila melanogaster*. *Genome Research* **2017**, *27*, 1273–1285. doi:10.1101/gr.213694.116.
4. Futo, M.; Opašić, L.; Koska, S.; Čorak, N.; Široki, T.; Ravikumar, V.; Thorsell, A.; Lenuzzi, M.; Kifer, D.; Domazet-Lošo, M.; Vlahoviček, K.; Mijakovic, I.; Domazet-Lošo, T. Embryo-like features in developing *Bacillus subtilis* biofilms. *Molecular Biology and Evolution* **2020**, *38*, 31–47. doi:10.1093/molbev/msaa217.
5. Church, R.B.; Robertson, F.W. A biochemical study of the growth of *Drosophila melanogaster*. *Journal of Experimental Zoology* **1966**, *162*, 337–351. doi:10.1002/jez.1401620309.
6. Friis-Hansen, B. Water distribution in the foetus and newborn infant. *Acta Paediatrica* **1983**, *72*, 7–11. doi:10.1111/j.1651-2227.1983.tb09852.x.
7. Tosca, N.J.; Knoll, A.H.; McLennan, S.M. Water activity and the challenge for life on early Mars. *Science* **2008**, *320*, 1204–1207. doi:10.1126/science.1155432.

8. Syamaladevi, R.M.; Tang, J.; Villa-Rojas, R.; Sablani, S.; Carter, B.; Campbell, G. Influence of water activity on thermal resistance of microorganisms in low-moisture foods: A review. *Comprehensive Reviews in Food Science and Food Safety* **2016**, *15*, 353–370. doi:10.1111/1541-4337.12190.
9. Jonchhe, S.; Pandey, S.; Emura, T.; Hidaka, K.; Hossain, M.A.; Shrestha, P.; Sugiyama, H.; Endo, M.; Mao, H. Decreased water activity in nanoconfinement contributes to the folding of G-quadruplex and i-motif structures. *Proceedings of the National Academy of Sciences* **2018**, *115*, 9539–9544. doi:10.1073/pnas.1805939115.
10. Pace, N.R. Origin of life: Facing up to the physical setting. *Cell* **1991**, *65*, 531–533. doi:10.1016/0092-8674(91)90082-A.
11. Lamadrid, H.M.; Rimstidt, J.D.; Schwarzenbach, E.M.; Klein, F.; Ulrich, S.; Dolocan, A.; Bodnar, R.J. Effect of water activity on rates of serpentinization of olivine. *Nature Communications* **2017**, *8*, 16107. doi:10.1038/ncomms16107.
12. Stevenson, A.; Cray, J.A.; Williams, J.P.; Santos, R.; Sahay, R.; Neuenkirchen, N.; McClure, C.D.; Grant, I.R.; Houghton, J.D.; Quinn, J.P.; Timson, D.J.; Patil, S.V.; Singhal, R.S.; Anton, J.; Dijksterhuis, J.; Hocking, A.D.; Lievens, B.; Rangel, D.E.N.; Voytek, M.A.; Gunde-Cimerman, N.; Oren, A.; Timmis, K.N.; McGenity, T.J.; Hallsworth, J.E. Is there a common water-activity limit for the three domains of life? *ISME Journal* **2015**, *9*, 1333–1351. doi:10.1038/ismej.2014.219.
13. Chirife, J.; Fontan, C.F.; Scorza, O.C. The intracellular water activity of bacteria in relation to the water activity of the growth medium. *Journal of Applied Bacteriology* **1981**, *50*, 475–479. doi:10.1111/j.1365-2672.1981.tb04250.x.
14. Record, Jr., M.T.; Courtenay, E.S.; Cayley, D.S.; Guttman, H.J. Responses of *E. coli* to osmotic stress: Large changes in amounts of cytoplasmic solutes and water. *Trends in Biochemical Sciences* **1998**, *23*, 143–148. doi:10.1016/S0968-0004(98)01196-7.
15. Stevenson, A.; Burkhardt, J.; Cockell, C.S.; Cray, J.A.; Dijksterhuis, J.; Fox-Powell, M.; Kee, T.P.; Kminek, G.; McGenity, T.J.; Timmis, K.N.; Timson, D.J.; Voytek, M.A.; Westall, F.; Yakimov, M.M.; Hallsworth, J.E. Multiplication of microbes below 0.690 water activity: implications for terrestrial and extraterrestrial life. *Environmental Microbiology* **2015**, *17*, 257–277. doi:10.1111/1462-2920.12598.
16. Logan, J.E.; Himwich, W.A. Animal tissues and organs: water content. In *Biology Data Book*, 2nd ed.; Altman, P.L.; Dittmer, D.S., Eds.; Federation of American Societies for Experimental Biology: Bethesda, Maryland, 1972; Vol. 1, pp. 392–398.
17. Calcagno, P.L.; Hollerman, C.E.; Jose, P.A. Total body water: man. In *Biology Data Book*, 2nd ed.; Altman, P.L.; Dittmer, D.S., Eds.; Federation of American Societies for Experimental Biology: Bethesda, Maryland, 1972; Vol. 3, pp. 1986–1989.
18. Cramer, W. On the biochemical mechanism of growth. *Journal of Physiology* **1916**, *50*, 322–334. doi:10.1113/jphysiol.1916.sp001758.
19. Downing, J.E.; Christopherson, W.M.; Broghamer, W.L. Nuclear water content during carcinogenesis. *Cancer* **1962**, *15*, 1176–1180. doi:10.1002/1097-0142(196211/12)15:6<1176::AID-CNCR2820150614>3.0.CO;2-F.
20. Saryan, L.A.; Hollis, D.P.; Economou, J.S.; Eggleston, J.C. Nuclear magnetic resonance studies of cancer. IV. Correlation of water content with tissue relaxation times. *JNCI: Journal of the National Cancer Institute* **1974**, *52*, 599–602. doi:10.1093/jnci/52.2.599.
21. Ross, K.F.A.; Gordon, R.E. Water in malignant tissue, measured by cell refractometry and nuclear magnetic resonance. *Journal of Microscopy* **1982**, *128*, 7–21. doi:10.1111/j.1365-2818.1982.tb00433.x.
22. Li, D.; Yang, Z.; Fu, A.; Chen, T.; Chen, L.; Tang, M.; Zhang, H.; Mu, N.; Wang, S.; Liang, G.; Wang, H. Detecting melanoma with a terahertz spectroscopy imaging technique. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2020**, *234*, 118229. doi:10.1016/j.saa.2020.118229.
23. Winzler, R.J. The chemistry of cancer tissue. In *The Physiopathology of Cancer*, 2nd ed.; Homburger, F., Ed.; Hoeber-Harper: New York, 1959; pp. 686–706.
24. Olmstead, E.G. *Mammalian Cell Water*; Lea & Febiger: Philadelphia, 1966.
25. McIntyre, G.I. Cell hydration as the primary factor in carcinogenesis: A unifying concept. *Medical Hypotheses* **2006**, *66*, 518–526. doi:10.1016/j.mehy.2005.09.022.
26. Dick, J.M. Water as a reactant in the differential expression of proteins in cancer. *Computational and Systems Oncology* **2020**, *1*, e1007. doi:10.1002/cso2.1007.

27. Miyoshi, D.; Karimata, H.; Sugimoto, N. Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions. *Journal of the American Chemical Society* **2006**, *128*, 7957–7963. doi:10.1021/ja061267m.

28. Gao, M.; Gnutt, D.; Orban, A.; Appel, B.; Righetti, F.; Winter, R.; Narberhaus, F.; Müller, S.; Ebbinghaus, S. RNA hairpin folding in the crowded cell. *Angewandte Chemie-International Edition* **2016**, *55*, 3224–3228. doi:10.1002/anie.201510847.

29. Alberty, R.A. Calculation of equilibrium compositions of biochemical reaction systems involving water as a reactant. *Journal of Physical Chemistry B* **2001**, *105*, 1109–1114. doi:10.1021/jp003515l.

30. Haynie, D.T. *Biological Thermodynamics*, 2nd ed.; Cambridge Univesity Press: Cambridge, 2008. doi:10.1017/CBO9780511802690.

31. Shock, E.L.; Canovas, P.; Yang, Z.; Boyer, G.; Johnson, K.; Robinson, K.; Fecteau, K.; Windman, T.; Cox, A. Thermodynamics of organic transformations in hydrothermal fluids. *Reviews in Mineralogy and Geochemistry* **2013**, *76*, 311–350. doi:10.2138/rmg.2013.76.9.

32. Foley, S.F. A reappraisal of redox melting in the Earth's mantle as a function of tectonic setting and time. *Journal of Petrology* **2011**, *52*, 1363–1391. doi:10.1093/petrology/egq061.

33. Evans, K.A.; Powell, R.; Frost, B.R. Using equilibrium thermodynamics in the study of metasomatic alteration, with an application to serpentinites. *Lithos* **2013**, *168-169*, 67–84. doi:10.1016/j.lithos.2013.01.016.

34. Garrels, R.M.; Christ, C.L. *Solutions, Minerals, and Equilibria*; Harper & Row: New York, 1965.

35. Anderson, G.M.; Crerar, D.A. *Thermodynamics in Geochemistry: The Equilibrium Model*; Oxford University Press: New York, 1993. doi:10.1093/oso/9780195064643.001.0001.

36. Amend, J.P.; McCollom, T.M. Energetics of biomolecule synthesis on early Earth. In *Chemical Evolution II: From the Origins of Life to Modern Society*; Zaikowski, L.; Friedrich, J.M.; Seidel, S.R., Eds.; American Chemical Society, 2009; Vol. 1025, *ACS Symposium Series*, chapter 4, pp. 63–94. doi:10.1021/bk-2009-1025.ch004.

37. Burg, J.M.; Tymoczko, J.L.; Stryer, L. *Biochemistry*, 5th ed.; W. H. Freeman, 2002.

38. Trigos, A.S.; Pearson, R.B.; Papenfuss, A.T.; Goode, D.L. Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *Proceedings of the National Academy of Sciences* **2017**, *114*, 6406–6411. doi:10.1073/pnas.1617743114.

39. Zhou, J.X.; Cisneros, L.; Knijnenburg, T.; Trachana, K.; Davies, P.; Huang, S. Phylostratigraphic analysis of tumor and developmental transcriptomes reveals relationship between oncogenesis, phylogenesis and ontogenesis. *Convergent Science Physical Oncology* **2018**, *4*, 025002. doi:10.1088/2057-1739/aab1b0.

40. Brocchieri, L.; Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Research* **2005**, *33*, 3390–3400. doi:10.1093/nar/gki615.

41. Moyers, B.A.; Zhang, J. Further simulations and analyses demonstrate open problems of phylostratigraphy. *Genome Biology and Evolution* **2017**, *9*, 1519–1527. doi:10.1093/gbe/evx109.

42. Van Oss, S.B.; Carvunis, A.R. *De novo* gene birth. *PLOS Genetics* **2019**, *15*, e1008160. doi:10.1371/journal.pgen.1008160.

43. Dick, J.M. Average oxidation state of carbon in proteins. *Journal of the Royal Society Interface* **2014**, *11*, 20131095. doi:10.1098/rsif.2013.1095.

44. Dick, J.M.; Yu, M.; Tan, J. Uncovering chemical signatures of salinity gradients through compositional analysis of protein sequences. *Biogeosciences* **2020**, *17*, 6145–6162. doi:10.5194/bg-17-6145-2020.

45. Liebeskind, B.J.; McWhite, C.D.; Marcotte, E.M. Towards consensus gene ages. *Genome Biology and Evolution* **2016**, *8*, 1812–1823. doi:10.1093/gbe/evw113.

46. Denbigh, K. *The Principles of Chemical Equilibrium*, 4th ed.; Cambridge University Press: Cambridge, 1981. doi:10.1017/CBO9781139167604.

47. Solel, E.; Tarannam, N.; Kozuch, S. Catalysis: energy is the measure of all things. *Chemical Communications* **2019**, *55*, 5306–5322. doi:10.1039/C9CC00754G.

48. Dick, J.M. CHNOSZ: Thermodynamic calculations and diagrams for geochemistry. *Frontiers in Earth Science* **2019**, *7*, 180. doi:10.3389/feart.2019.00180.

49. Dick, J.M.; LaRowe, D.E.; Helgeson, H.C. Temperature, pressure, and electrochemical constraints on protein speciation: Group additivity calculation of the standard molal thermodynamic properties of ionized unfolded proteins. *Biogeosciences* **2006**, *3*, 311–336. doi:10.5194/bg-3-311-2006.

50. Dick, J.M.; Shock, E.L. A metastable equilibrium model for the relative abundances of microbial phyla in a hot spring. *PLOS One* **2013**, *8*, e72395. doi:10.1371/journal.pone.0072395.

51. Tcherkas, Y.V.; Denisenko, A.D. Simultaneous determination of several amino acids, including homocysteine, cysteine and glutamic acid, in human plasma by isocratic reversed-phase high-performance liquid chromatography with fluorimetric detection. *Journal of Chromatography A* **2001**, *913*, 309–313. doi:10.1016/S0021-9673(00)01201-2.
52. Dick, J.M. Chemical composition and the potential for proteomic transformation in cancer, hypoxia, and hyperosmotic stress. *PeerJ* **2017**, *5*, e3421. doi:10.7717/peerj.3421.
53. van 't Erve, T.J.; Wagner, B.A.; Ryckman, K.K.; Raife, T.J.; Buettner, G.R. The concentration of glutathione in human erythrocytes is a heritable trait. *Free Radical Biology and Medicine* **2013**, *65*, 742–749. doi:10.1016/j.freeradbiomed.2013.08.002.
54. Jones, D.P.; Sies, H. The redox code. *Antioxidants & Redox Signaling* **2015**, *23*, 734–746. doi:10.1089/ars.2015.6247.
55. Frost, B.R. Introduction to oxygen fugacity and its petrologic importance. In *Oxide Minerals*; De Gruyter, 1991; Vol. 25, *Reviews in Mineralogy*, pp. 1–10. doi:10.1515/9781501508684-004.
56. Anderson, G.M. *Thermodynamics of Natural Systems*, 2nd ed.; Cambridge University Press: Cambridge, 2005. doi:10.1017/CBO9780511840258.
57. Martin, W.F.; Sousa, F.L. Early microbial evolution: The age of anaerobes. *Cold Spring Harbor Perspectives in Biology* **2016**, *8*. doi:10.1101/cshperspect.a018127.
58. Babauta, J.T.; Nguyen, H.D.; Harrington, T.D.; Renslow, R.; Beyenal, H. pH, redox potential and local biofilm potential microenvironments within *Geobacter sulfurreducens* biofilms and their roles in electron transfer. *Biotechnology and Bioengineering* **2012**, *109*, 2651–2662. doi:10.1002/bit.24538.
59. Gibbs, A.G.; Markow, T.A. Effects of age on water balance in *Drosophila* species. *Physiological and Biochemical Zoology* **2001**, *74*, 520–530. doi:10.1086/322162.
60. Rumble, III, D. The role of perfectly mobile components in metamorphism. *Annual Review of Earth and Planetary Sciences* **1982**, *10*, 221–233. doi:10.1146/annurev.ea.10.050182.001253.
61. Wedberg, R.; Abildskov, J.; Peters, G.H. Protein dynamics in organic media at varying water activity studied by molecular dynamics simulation. *Journal of Physical Chemistry B* **2012**, *116*, 2575–2585. doi:10.1021/jp211054u.
62. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
63. Dick, J.M. JMDplots 1.2.5, 2021. doi:10.5281/zenodo.4477998.
64. Liebeskind, B.; McWhite, C.D.; Hines, K. Gene-Ages v1.0, 2016. doi:10.5281/zenodo.51708.
65. Huang, H.; McGarvey, P.B.; Suzek, B.E.; Mazumder, R.; Zhang, J.; Chen, Y.; Wu, C.H. A comprehensive protein-centric ID mapping service for molecular data integration. *Bioinformatics* **2011**, *27*, 1190–1191. doi:10.1093/bioinformatics/btr101.