

Oxygen content of transmembrane proteins over macroevolutionary time scales

Claudia Acquisti^{1†}, Jürgen Kleffe² & Sinéad Collins¹

We observe that the time of appearance of cellular compartmentalization correlates with atmospheric oxygen concentration. To explore this correlation, we predict and characterize the topology of all transmembrane proteins in 19 taxa and correlate differences in topology with historical atmospheric oxygen concentrations. Here we show that transmembrane proteins, individually and as a group, were probably selectively excluding oxygen in ancient ancestral taxa, and that this constraint decreased over time when atmospheric oxygen levels rose. As this constraint decreased, the size and number of communication-related transmembrane proteins increased. We suggest the hypothesis that atmospheric oxygen concentrations affected the timing of the evolution of cellular compartmentalization by constraining the size of domains necessary for communication across membranes.

One of the major transitions in macroevolution was the appearance of eukaryotic cells between 2.1 and 1.8 billion years ago^{1–3}. Cellular compartmentalization by membranes that are impermeable to large or charged molecules requires transport and communication across intracellular membranes. Eukaryotes devote more proteins to roles in communication than prokaryotes; this innovation involved a shift in the dominant secondary structures of transmembrane proteins⁴. Protein secondary structure is largely determined by hydrophobicity⁵, where oxygen and nitrogen are vital to forming hydrophilic residues. Transmembrane protein topology is further influenced by charge, where positively charged amino acids are more prevalent in cytoplasmic domains and negatively charged amino acids are more prevalent in extracellular domains^{6–8}. This implies that changes in protein atomic composition may occur in parallel with changes in protein function. Traditionally, functional changes were thought to be associated with changes in amino acid sequence⁹, but an alternative approach is to consider proteins at the atomic level. This may be appropriate when large fluctuations in the elemental components of proteins occur through changes in absolute abundance, relative abundance, or form. In this case, nutritional constraints, metabolic optimization and chemical properties such as redox state may have important roles in protein evolution.

The atomic content of biomolecules has a role in evolution

Several examples of stoichiometric constraints on evolutionary and ecological outcomes have been reported recently. For example, variation in the atomic content of proteins in cyanobacterial light-harvesting proteins and microbial sulphur assimilatory enzymes correlates with nutrient availability^{10,11}. Similarly, the carbon content of proteomes differs between species and correlates with genomic G+C content, which may reflect carbon availability in natural habitats¹². The nitrogen content of proteins is lower in plants than in animals and is related to gene expression levels in plants¹³. These studies indicate that physiology, proteomes and genomes may bear

detectable ecological imprints over macroevolutionary time scales, and that ancient habitat composition may affect current proteome composition. Furthermore, recent advances in the field of ecological stoichiometry have shown that the relative abundances of phosphorus to carbon or nitrogen can influence ecological outcomes. This may occur through stoichiometric constraints on growth rate, which can lead to variation in life-history traits that subsequently affect species–species interactions^{14,15}. The relative abundances of nutrients have been associated with many macroevolutionary innovations including the appearance of winged insects¹⁶ and the timing of the Cambrian explosion¹⁷.

Molecular oxygen was introduced relatively quickly into the atmosphere by the ancestors of cyanobacteria about 2.2 billion years ago¹⁸ and has varied between about 15 and 35% over the time that eukaryotic cells have been present¹⁹. Here we address how atmospheric oxygen concentrations may have constrained transmembrane protein composition and structure, and then suggest a functional interpretation of these constraints at a cellular level. We explore the possible macroevolutionary consequences of this—namely, the timing of the appearance of eukaryotic cells. First, we characterize the oxygen content and topology of the entire set of predicted transmembrane proteins from 19 organisms (listed in Supplementary Table 1). We then show how the oxygen content of transmembrane proteins varies with respect to atmospheric oxygen concentration over the past 3.5 billion years, and we suggest a mechanism of how this could have constrained the timing of evolution of cellular compartmentalization.

Oxygen content of transmembrane proteins in prokaryotes and eukaryotes

To investigate how atomic content changed over macroevolutionary time scales, we calculated the mean side-chain density for carbon, hydrogen, oxygen and nitrogen content for the full, predicted proteomes of 19 organisms. Carbon and hydrogen content density functions have a nearly gaussian distribution. In contrast, we found that

¹Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Köln, Germany. ²Institute of Molecular Biology and Biochemistry, Charité Campus Benjamin Franklin, Arnimallee 22, 14195 Berlin, Germany. [†]Present address: Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona State University, Tempe, Arizona 85287-5301, USA.

the density function for oxygen content is bimodal in prokaryotes and nearly gaussian in eukaryotes, with the exception of *Caenorhabditis elegans* and *Giardia lamblia* (Fig. 1; Supplementary Figs 1, 2). The same pattern occurs for nitrogen content density, although it is less pronounced (Supplementary Figs 3–5). To test if the low-oxygen peak of the bimodal distributions was associated with a particular subset of proteins, we extracted the transmembrane proteins from the full proteome for each organism using a transmembrane protein topology prediction method that was based on a hidden Markov model (TMHMM)²⁰. We assigned the remainder of the full proteome minus the putative transmembrane proteins to the ‘non-transmembrane protein’ group. Comparing the oxygen content density functions for the non-transmembrane with those for the transmembrane protein set shows that the distributions of the two sets are significantly distinct from each other (chi-squared test, $P < 0.0001$; Fig. 1).

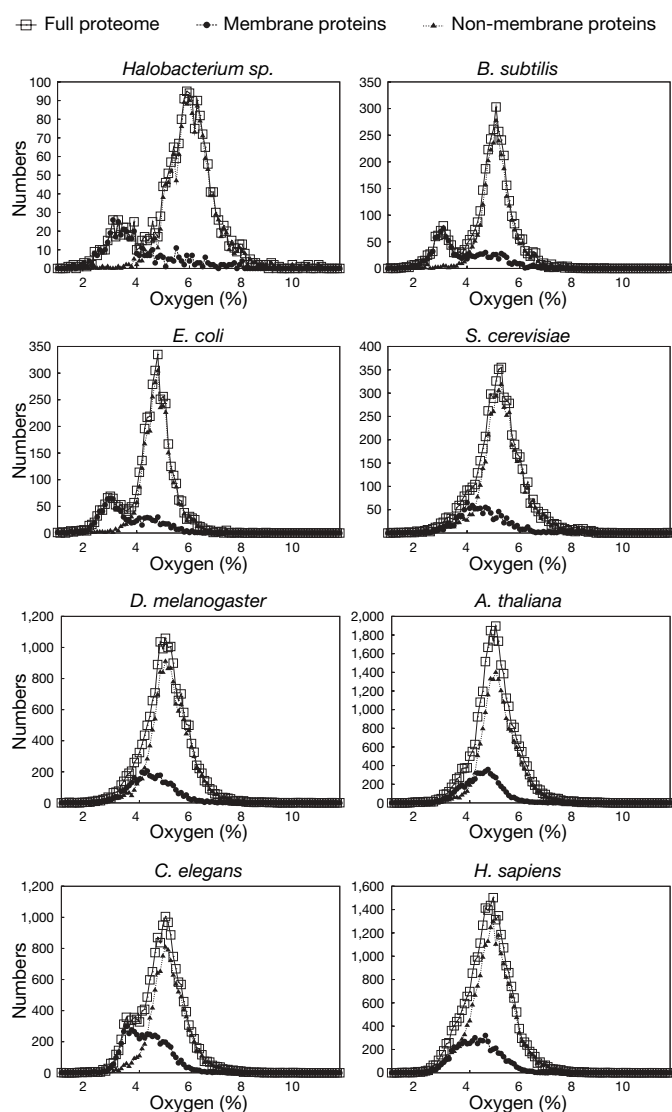


Figure 1 | Oxygen content density functions on full proteome, transmembrane and non-transmembrane protein sets. For each proteome (*Halobacterium sp.*, *Bacillus subtilis*, *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Caenorhabditis elegans* and *Homo sapiens*) the oxygen content density was calculated as the percentage of oxygen atoms found in residue side chains for each protein, and the percentages plotted as a histogram. The oxygen content density histograms were plotted separately for the transmembrane and non-transmembrane protein sets. Sample sizes given in Supplementary Table 1.

Transmembrane proteins constitute a separate group in terms of oxygen content density in eukaryotic and prokaryotic proteomes, although the difference is more pronounced in prokaryotes. The difference in mean oxygen percentages of transmembrane and non-transmembrane proteins is 0.73 in eukaryotes, whereas it is 1.33 in prokaryotes. Transmembrane proteins have lower oxygen contents than non-transmembrane proteins in each taxonomic domain (Cochran-Cox 1-tailed t -test, $P < 0.0005$). Although eukaryotic proteomes are larger than prokaryotic ones, the fraction of transmembrane proteins is largely conserved: in most genomes, 20–30% of all genes encode transmembrane proteins²¹. This suggests that the differences in oxygen content density distributions between prokaryotes and eukaryotes are not attributable to gross differences in proteome composition. However, individual transmembrane proteins in eukaryotes tend to be longer than in prokaryotes²², and often

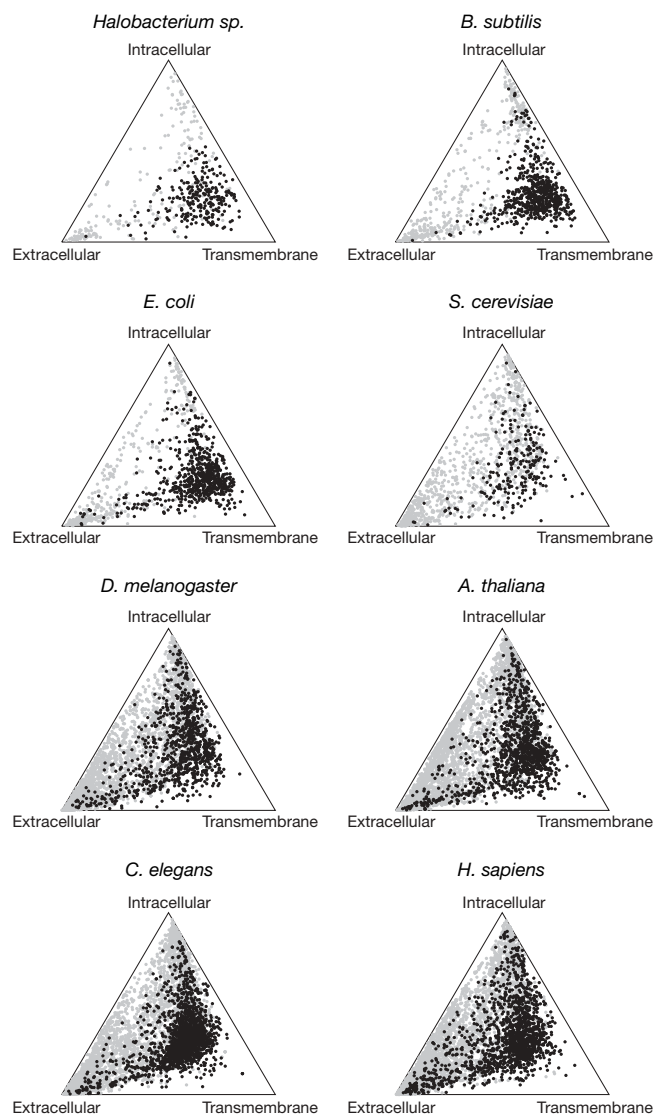


Figure 2 | Ternary diagrams of compositional data for transmembrane, extracellular and intracellular domains for the entire predicted transmembrane protein set. Each individual protein is represented by a three-dimensional vector of components (extracellular, intracellular and transmembrane), the sum of which is 1. The magnitude of a component is equal to the length of the perpendicular axis leading to the edge opposite the vertex having the same identity as that component. The coordinates of each data point show topology, and colour shows oxygen content (dark colour, $[O] < 3.9\%$; light colour, $[O] \geq 3.9\%$). Sample sizes given in Supplementary Table 1.

lengthen either the intra- or extracellular domains relative to prokaryotes, rather than simply expanding the entire protein²³.

Oxygen content and transmembrane protein topology

To describe the relationship between transmembrane protein oxygen content and their topology, and to further characterize differences between the prokaryotic and eukaryotic transmembrane proteins, the topology of each transmembrane protein was determined for the entire proteome of each organism using TMHMM²⁰ (Fig. 2; Supplementary Figs 6–7). Two clusters can be seen within each diagram. One is made up of proteins with a high proportion of transmembrane domains that have low oxygen content, and the second is made up of proteins with a lower proportion of transmembrane domains with higher oxygen content. The distributions of the low-oxygen and high-oxygen proteins in the ternary diagrams differ significantly from each other (multivariate likelihood ratio test for compositional data²⁴, $P = 0.00001$). The overlap between protein sets defined either by oxygen content or by topology is at least 80% in all cases, indicating that there is good agreement between oxygen con-

tent and topology when the two are independently estimated. Our distribution of topologies is consistent with previous work showing that transmembrane proteins have a tendency to either form many transmembrane domains with short connecting loops or few transmembrane domains with large extracellular domains, but not both²². In addition, our compositional data show differences in transmembrane protein topology between eukaryotes and prokaryotes, with prokaryotes having a higher proportion of transmembrane-domain-rich proteins. Similarly, there is a threefold difference in the number of low-oxygen transmembrane proteins between prokaryotes and eukaryotes, whereas there is a tenfold difference in the number of high-oxygen transmembrane proteins, indicating that high-oxygen proteins dominated by intra- or extracellular domains are preferentially added as proteomes increase in size over time. The preferential addition of high-oxygen proteins over time shows that some transmembrane proteins may have been oxygen-limited at the time they evolved, or that there was selection against the use of oxygen in external transmembrane domains. This indicates that changes in transmembrane protein oxygen content may have been connected with changes in protein structure and function over geological time.

The topology description above is normalized and reflects differences in proportional contributions of inside (intracellular), outside (extracellular) and transmembrane domains to total protein length. However, it does not describe how the absolute sizes of these domains differ systematically between organisms. Within sets of high- and low-oxygen transmembrane proteins, we estimated the mean number of residues devoted to the inside, outside and transmembrane domains per proteome, on the basis of TMHMM (Fig. 3). Transmembrane proteins are either composed of many transmembrane segments with few short loops, or few transmembrane segments with large intra- or extracellular loops²². These topologies correspond to channels and receptors, respectively. We also calculated the oxygen concentration of each domain (Fig. 4), finding that high-oxygen transmembrane proteins (receptors) had higher oxygen content in their external than in their internal and transmembrane domains, or compared with any domain of the low-oxygen transmembrane proteins (Cochran-Cox 1-tailed t -test, corrected $P = 0.005$). When the domain lengths are plotted against the time of appearance of class,

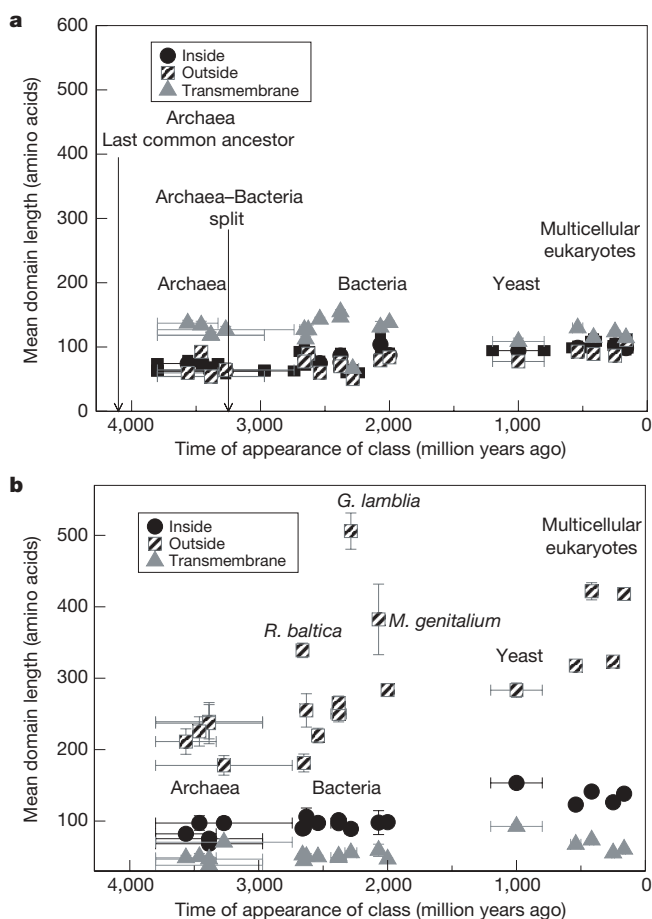


Figure 3 | Mean domain length versus time of appearance of class. **a**, Low-oxygen transmembrane proteins ($[O] < 3.9\%$). Mean values for domain lengths of outside and transmembrane domains do not correlate with time (Spearman's coefficients: inside, $r_s = -0.542$, $P = 0.009$; outside, $r_s = -0.229$, $P = 0.17$; transmembrane, $r_s = 0.217$, $P = 0.19$). The ranges of lengths are: $60 < \text{inside} < 109$; $50 < \text{outside} < 111$; $66.5 < \text{transmembrane} < 154.7$ amino acids. **b**, High-oxygen transmembrane proteins ($[O] \geq 3.9\%$). Inside, outside and transmembrane domain lengths correlate with time (Spearman's coefficients: inside, $r_s = -0.802$, $P = 0.00002$; outside, $r_s = -0.740$, $P = 0.0001$; transmembrane, $r_s = 0.606$, $P = 0.003$). The ranges of lengths are: $68 < \text{inside} < 150$; $177 < \text{outside} < 505$; $38 < \text{transmembrane} < 92$ amino acids). Points show mean \pm s.e.m. of the length, and the range of time of appearance^{30–34}. Sample sizes given in Supplementary Table 1.

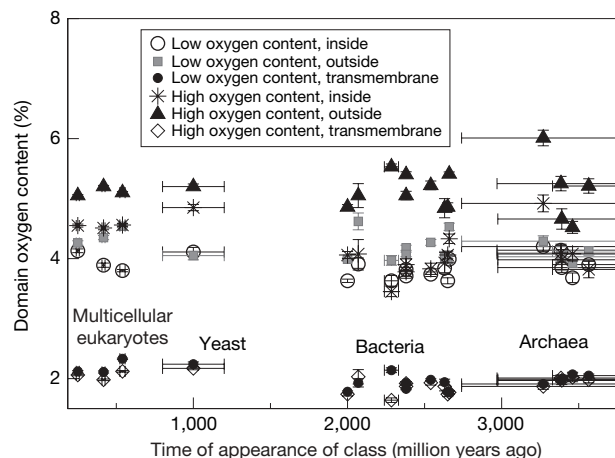


Figure 4 | Inside, outside and transmembrane domain oxygen content versus time of appearance of class. For each transmembrane protein with low-oxygen ($[O] < 3.9\%$) or high-oxygen ($[O] \geq 3.9\%$) content, the oxygen content of different domains was estimated using TMHMM²⁰. Each point represents the mean \pm s.e.m. of oxygen content, and the range of time of predicted first appearance^{30–34}. The mean oxygen content of specific domains does not correlate with the time of appearance of class (Spearman's coefficients: low-oxygen domains, inside, $r_s = -0.125$, $P = 0.9$; outside, $r_s = -0.22$, $P = 0.37$; transmembrane, $r_s = -0.404$, $P = 0.11$; and high-oxygen domains, inside, $r_s = -0.350$, $P = 0.15$, outside, $r_s = 0.107$, $P = 0.65$, transmembrane, $r_s = -0.071$, $P = 0.77$). Sample sizes given in Supplementary Table 1.

high-oxygen transmembrane proteins show a rapid increase in outside relative to inside domains (Fig. 3b). In contrast, low oxygen transmembrane proteins do not show any obvious difference in rates of increase, and mean values of inside and outside domains change relatively slowly over time in low-oxygen transmembrane proteins (Fig. 3a). However, the mean oxygen concentration of the external domains did not change directionally over time in either the high- or low-oxygen transmembrane proteins (receptors or channels) (Spearman's coefficient: low-oxygen proteins $r_s = -0.22$, $P > 0.37$; high-oxygen proteins $r_s = 0.107$, $P = 0.65$), meaning that the charge per unit length remained relatively constant (Fig. 4). This is consistent with charge density being important for the insertion of domains into cellular membranes^{6–8}, and suggests that the total oxygen content of a domain must be altered by changing the length of that domain. The relatively rapid changes in the size of the oxygen-rich external domains coincide with increasing organismal complexity, whereas changes in the nitrogen-rich internal domains are not as pronounced (Fig. 3). Differential rates of evolution have previously been observed in the outside and inside domains of chemokine receptors²³, which is consistent with the general trend we found for all high-oxygen transmembrane proteins. Interestingly, the parasitic bacteria *Mycoplasma genitalium* has much longer external domains than expected, which are similar in length to eukaryotes, perhaps as a result of co-evolution with eukaryotic hosts (Fig. 3). The compartmentalized prokaryote *Rhodospirillum rubrum* also has longer extracellular domains than the other prokaryotes, indicating that external domains in transmembrane proteins may be important to this compartmentalization (Fig. 3).

In our data, the high-oxygen transmembrane proteins show an increase in receptors relative to channels in eukaryotes, coupled with a faster rate of growth of external than internal domains over millions of years. This is consistent with previous observations that the percentage of total open reading frames encoding transport proteins is lower in eukaryotes than in prokaryotes⁴. This increase is correlated with the transitions from prokaryotic to eukaryotic cells, and from unicellular to multicellular organisms^{25,26}. One possible explanation for different domain growth rates is that the domains may have different elemental requirements, with the positively charged internal domains requiring more nitrogen, and the negatively charged external domains requiring more oxygen. There is a similar, yet weaker signal in the overall per residue nitrogen content, which suggests that oxygen levels were more important to the changes in transmembrane protein topology seen here than changes in nitrogen levels. Still, it does not exclude the possibility that the ratio of the two elements had a role.

Taken together, data in Fig. 3 and in Fig. 4 show that the oxygen content of extracellular domains is higher in receptors than it is in channels; this is evident from increases in both the density of oxygen content and the total length of external domains. This points towards a key role for oxygen in the increase in abundance and size of receptors over time. In terms of cellular function, a faster increase in the size of external domains indicates an increase in the proportion of communication-related proteins over time. The differential changes in domain length between high- and low-oxygen transmembrane proteins, together with previous work, suggest that external loops of transmembrane proteins are under different selective constraints than internal loops. This constraint may have prevented the formation of large external domains, limiting communication across membranes when atmospheric oxygen concentrations were low.

Atmospheric oxygen levels, cellular communication and compartmentalization

If atmospheric oxygen levels constrained transmembrane protein composition over very long time scales, it is expected that the two should covary. We found that the atomic composition of transmembrane proteins does scale with atmospheric oxygen levels over macroevolutionary time scales. Figure 5a shows the mean oxygen contents of high- and low-

oxygen transmembrane proteins separately in 19 organisms plotted against the atmospheric oxygen concentration at the time the classes of these organisms first appeared. Archaea, and, to a lesser extent, other prokaryotes, show a similar mean oxygen content of high-oxygen transmembrane proteins and a lower mean oxygen content of low-oxygen transmembrane proteins relative to eukaryotes (Spearman's coefficients: low-oxygen transmembrane proteins, $r_s = 0.601$, $P = 0.0032$; high-oxygen transmembrane proteins, $r_s = -0.114$, $P = 0.33$). This suggests that oxygen availability influenced transmembrane protein composition, and that older taxa exclude oxygen from low-oxygen proteins to a greater extent than do younger taxa.

The selective use of oxygen by older taxa indicates that atmospheric oxygen concentrations may have limited the size and/or number of high-oxygen proteins that were produced in ancient proteomes. Figure 5b shows the ratio of the numbers of high- and low-oxygen transmembrane proteins. In contrast with the composition of individual transmembrane proteins (Fig. 5a), where single-celled compartmentalized organisms (*Saccharomyces cerevisiae*, *R. rubrum* and *G. lamblia*) are intermediate between prokaryotes and multicellular eukaryotes, these organisms show the same properties as other eukaryotes in terms of the proportion of the proteome devoted to high-oxygen transmembrane proteins, regardless of whether they are prokaryotic or eukaryotic (Fig. 5b). In addition, eubacteria and

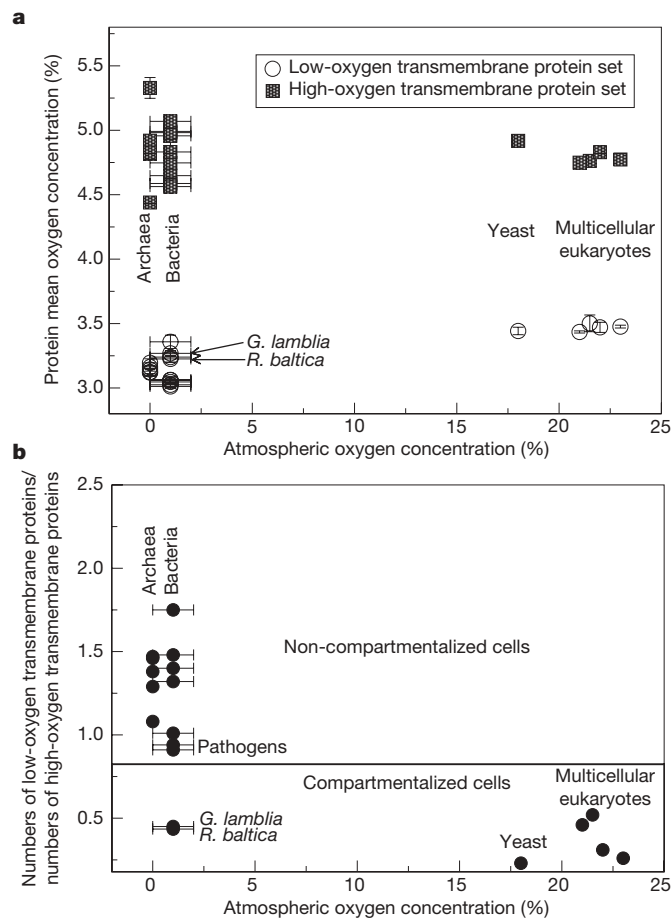


Figure 5 | Mean proteome oxygen content. **a**, Mean oxygen content of amino acid side chains in high-oxygen ($[O] \geq 3.9\%$) and low-oxygen ($[O] < 3.9\%$) transmembrane proteins plotted against the atmospheric oxygen level at the time of appearance of the organism's class. Each point represents the mean \pm s.e.m. of the oxygen content and the range of atmospheric oxygen concentrations at the time of appearance^{18,19}. **b**, Ratio of the absolute numbers of low- to high-oxygen transmembrane proteins per proteome versus atmospheric oxygen concentration at the time of appearance. Error bars show the range of atmospheric oxygen concentrations at the time of appearance^{18,19}. Sample sizes are given in Supplementary Table 1.

archaea are indistinguishable in this respect, and there is no systematic association between eukaryote complexity, proteome oxygen partitioning, and atmospheric oxygen levels. This suggests a very basic functional difference associated with proteome composition. Because the only two groups that can be distinguished by how oxygen is partitioned at the proteome level are compartmentalized and non-compartmentalized cells ($F_{(1,14)} = 13.25$, $P = 0.002$), the simplest interpretation of this observation is that cellular compartmentalization requires larger intra- and extracellular domains, probably in order to integrate cellular processes such as signalling and transport. In an oxygen-poor (reducing) atmosphere, it may not have been possible to produce a large enough number of these domains for communication in a compartmentalized cell. As such, atmospheric oxygen concentration may have affected the rate of increase in cellular complexity and the timing of the appearance of eukaryotic cells.

Discussion

In this study, we have shown that transmembrane proteins can be divided into two groups according to their oxygen content. Independent topology prediction reveals these same two groups. We have shown that the proportion of receptors to channels increases over time and coincides with a change in cellular organization. In addition, older proteomes contain less oxygen per residue and produce fewer high-oxygen proteins. Taken together, this suggests that oxygen use was selected against in these proteomes. This constraint lessened over time as the concentration of atmospheric oxygen increased, which resulted in the extracellular domains of transmembrane proteins increasing in size over time faster than the internal domains. Consequently, we propose the following hypothetical mechanism: atmospheric oxygen concentration constrained the topology of ancient transmembrane proteins by limiting the number and size of external domains that could be formed.

Any mechanistic explanation of how atmospheric oxygen concentration limited the number and size of external domains is necessarily speculative. One possibility is that it was simply futile to exude large, oxygen-rich domains in a reducing atmosphere where oxidized amino acids could have been rapidly reduced. In this case, the use of oxygen-rich amino acids would have been selected against by natural selection because protein structure would have been more robust when fewer oxidized residues were exuded. Linking this to the timing of appearance of eukaryotic cells implies that the oxygen content is preferentially increased in receptors, and that this increase affects receptor function. This makes intuitive sense because the external domains of receptors required for communication have specific secondary and tertiary structures, many of which have some minimum size²³. This is consistent with the bias we found towards having both longer and more oxygen-dense external domains in receptors relative to channels, and with the fact that eukaryotic genomes encode more and larger receptors than do prokaryotes. This suggests that protein oxygen content itself is important, rather than being a proxy for some other property. A second possibility is metabolic limitation. There is less evidence for direct limitation, though the synthesis of tyrosine requires molecular oxygen; indirect limitation seems more likely. For example, the synthesis of amino acids with oxygen in their side chains requires less energy than that of other amino acids²⁷, so selectively excluding oxygen-rich amino acids entails an energetic cost. In addition, the synthesis of many hormones and neurotransmitters requires molecular oxygen, indicating that high levels of molecular oxygen may be needed for communication-related molecules in general.

Constraints on transmembrane protein topology may have played an important part in the timing of the appearance of compartmentalized cells. One of the limits inherent in using data about extant organisms to draw conclusions about constraints on their ancestors is that it is impossible to know how much adaptation to current habitat may bias these data. However, it is reasonable to assume that some historical imprint remains in sequence data. We used organisms with

a wide range of metabolism and cellular organization. We have used several different archaea and an extremophile bacterium (*Aquifex aeolicus*), such that the bias associated with any particular extreme of environment is taken into account. Many of the microbes have experienced high levels of oxygen relative to their ancestors for several million years, such that the differences we report are a conservative estimate of the extent to which oxygen may have historically limited proteome composition. This work adds to a growing body of literature connecting atmospheric oxygen levels with macroevolutionary changes, most recently with complexity in metabolic networks²⁸ and cell types²⁹. In order to understand the broad role of oxygen levels in major transitions, further investigation and cooperation between the fields of palaeoclimatology, evolutionary biology and bioinformatics are necessary. This promises to yield many interesting results that address fundamental relationships between macroevolutionary transitions and environmental change.

Received 8 June; accepted 14 November 2006.

Published online 20 December 2006.

- Han, T. M. & Runnegar, B. Macroscopic eukaryotic algae from the 2.1-billion-year-old Negaunee iron-formation, Michigan. *Science* **257**, 232–235 (1992).
- Schopf, J. W. & Klein, C. (eds) *The Proterozoic Biosphere: a Multidisciplinary Study* (Cambridge Univ. Press, New York, 1992).
- Knoll, A. H. Proterozoic and early Cambrian protists: evidence for accelerating evolutionary tempo. *Proc. Natl Acad. Sci. USA* **91**, 6743–6750 (1994).
- Ren, Q. & Paulsen, Q. T. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Comp. Biol.* **1**, e27 (2005).
- Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
- von Heijne, G. Net N–C charge imbalance may be important for signal sequence function in bacteria. *J. Mol. Biol.* **192**, 287–290 (1986).
- Sipos, L. & von Heijne, G. Predicting the topology of eukaryotic membrane proteins. *Eur. J. Biochem.* **213**, 1333–1340 (1993).
- von Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rules. *J. Mol. Biol.* **225**, 487–494 (1992).
- Tourasse, N. J. & Li, W. H. Selective constraints, amino acid composition, and the rate of protein evolution. *Mol. Biol. Evol.* **17**, 656–664 (2000).
- Mazel, D. & Marlière, P. Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. *Nature* **341**, 245–248 (1989).
- Baudouin-Cornu, P., Surdin-Kerjan, Y., Marlière, P. & Thomas, D. Molecular evolution of protein atomic composition. *Science* **293**, 297–300 (2001).
- Baudouin-Cornu, P., Schuerer, K., Marlière, P. & Thomas, D. Intimate evolution of Proteins. Proteome atomic content correlates with genome base composition. *J. Biol. Chem.* **279**, 5421–5428 (2004).
- Elser, J. J., Fagan, W. F., Subramanian, S. & Kumar, S. Signatures of ecological resource availability in the animal and plant proteomes. *Mol. Biol. Evol.* **23**, 1946–1951 (2006).
- Elser, J. J. *et al.* Biological stoichiometry from genes to ecosystems. *Ecol. Lett.* **3**, 540–550 (2000).
- Kay, A. D. *et al.* Toward a stoichiometric framework for evolutionary biology. *Oikos* **109**, 6–17 (2005).
- Hasenfuss, I. A possible evolutionary pathway to insect flight starting from lepismatid organization. *J. Zool. Syst. Evol. Res.* **40**, 65–81 (2002).
- Elser, J. J., Watts, J., Schampel, J. H. & Farmer, J. Early Cambrian food webs on a trophic knife-edge? A hypothesis and preliminary data from a modern stromatolite-based ecosystem. *Ecol. Lett.* **9**, 295–303 (2006).
- Knoll, A. H. *Life on a Young Planet: The First Three Billion Years of Evolution on Earth* (Princeton Univ. Press, Princeton and Oxford, 2003).
- Berner, R. A. Atmospheric oxygen over Phanerozoic time. *Proc. Natl Acad. Sci. USA* **96**, 10955–10957 (1999).
- Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
- Stevens, T. J. & Arkin, I. T. Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins* **39**, 417–420 (2000).
- Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**, 1029–1038 (1998).
- Liò, P. & Vannucci, M. Investigating the evolution and structure of chemokine receptors. *Gene* **317**, 29–37 (2003).
- Aitchison, J. *The Statistical Analysis of Compositional Data* 154–155 (Chapman and Hall, London, 1986).
- Tamames, J., Ouzounis, C., Sander, C. & Valencia, A. Genomes with distinct function composition. *FEBS Lett.* **389**, 96–101 (1996).
- Liu, J. & Rost, B. Comparing function and structure between entire proteomes. *Protein Sci.* **10**, 1970–1979 (2001).

27. Akashi, H. & Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA*. **99**, 3695–3700 (2002).
 28. Raymond, J. & Segre, D. The effect of oxygen on biochemical networks and the evolution of complex life. *Science* **311**, 1764–1767 (2006).
 29. Hedges, S. B., Blair, J. E., Venturi, M. L. & Shoe, J. L. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* **4**, 2 (2004).
 30. Martin, W. *et al.* Early cell evolution, eukaryotes, anoxia, sulfide, oxygen, fungi first (?), and a tree of genomes revisited. *IUBMB Life* **55**, 193–204 (2003).
 31. Douzery, E. J. P., Snell, E. A., Baptiste, E., Delsuc, F. & Philippe, H. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl Acad. Sci. USA*. **101**, 15386–15391 (2004).
 32. Sheridan, P. P., Freeman, K. H. & Brenchley, J. E. Estimated minimal divergence times of the major bacterial and archaeal phyla. *Geomicrobiol. J.* **20**, 1–14 (2003).
 33. Hedges, S. B. The origin and evolution of model organisms. *Nature Rev. Genet.* **3**, 838–849 (2002).
 34. Anderson, J. S. & Sues, H.-D. (eds) *Major Transitions in Vertebrate Evolution* (Indiana Univ. Press, Bloomington and Indianapolis, in the press).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** The authors would like to thank J. Anderson for help with the estimates of the time of appearance of the organisms used for this study, and D. Schomburg, R. Wüschiers, D. Bauer, A. Scialpi, M. Koornneef, H. Hillebrand, A. M. Tarchi, P. Bruni, T. Wiehe, B. Haubold and T. Rothery for valuable discussions.
- Author Contributions** C.A. initiated and devised the project; C.A., S.C. and J.K. analysed the data; and S.C. and C.A. wrote the manuscript.
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.A. (Claudia.Acquisti.1@asu.edu).