ON THE PLAUSIBILITY OF A UV TRANSPARENT BIOCHEMISTRY

CHARLES S. COCKELL[1]* and ALESSANDRO AIRO[2]
[1] *British Antarctic Survey, High Cross, Cambridge, U.K.;* [2] *Free University Berlin, Institute for Biochemistry, Berlin, Germany*
(author for correspondence, e-mail: csco@bas.ac.uk)*

(Received 5 June 2001; accepted in revised form 22 January 2002)

Abstract. Some molecules, particularly aromatics, have high molar extinction coefficients at wavelengths in the damaging ultraviolet radiation region of the spectrum between 200 and 400 nm. Thus, under a UV radiation flux in which these wavelengths are represented, it could be argued that a selection pressure would exist for a UV transparent biochemistry in which they were not represented. This hypothesis is explored using data made available from proteomics, focusing particularly on tryptophan, against which a selection pressure could exist on present-day Earth as a result of its absorbance shoulder at wavelengths greater than 290 nm. The abundance of tryptophan in whole proteomes is lower than expected from the degeneracy of the genetic code. A lower usage of tryptophan is found in the cytochrome c oxidase polypeptide I of UV-exposed organisms compared to nocturnal and subterranean organisms, but not in ATP synthase chain A. Examination of the amino acid composition of photolyase, an enzyme that requires exposure to light to function, shows that the tryptophan abundances exceed those of the total proteome of most organisms and the abundances expected from the degeneracy of the genetic code. This is also true for cytochrome c oxidase, another enzyme that makes extensive use of the electron transfer properties of tryptophan. We suggest that the selection pressure for the use of tryptophan caused, among other factors, by the uses of delocalised pi-electrons that this aromatic provides in active sites and binding motifs outweighs the selection pressure for UV transparency. This trade-off explains the lack of conclusive evidence for a UV transparent selection pressure. We suggest that this trade-off applies to the stacked pi-electrons of DNA. It offers a solution to the long-standing paradox of why the macromolecule responsible for the faithful replication of information has high absorbance in the damaging UV radiation region of the spectrum.

Keywords: Archean, DNA, photolyase, pi-electron, tryptophan, UV radiation

1. Introduction

Ultraviolet (UV) radiation is not a ubiquitously deleterious physical factor for life. Many organisms use this part of the spectrum to enhance visual acuity (e.g., Yokoyama and Shi, 2000) and insofar as it can cause mutations, it has been discussed as a possible engine for diversity in microorganisms (Rothschild, 1999). However, it is true to say that the most biologically important effects of UV radiation are negative. It has been shown to be involved in damage ranging from the formation of DNA lesions (Helbling *et al*., 2001; Nara *et al*., 2001) to cataracts in certain metazoans that possess eyes (Michael *et al*., 2000).

Origins of Life and Evolution of the Biosphere **32:** 255–274, 2002.

TABLE I

Required modifications to terrestrial biochemistry to assemble a UV-transparent organism. Here 'UV-transparent' means a molar extinction coefficient of $\leq$100 M^{-1} cm^{-1} at wavelengths >230 nm and <400 nm. To achieve transparency down to 200 nm imposes even greater constraints than those shown here (e.g., removal of methionine and histidine from protein sequences and evolution of genetic bases with no absorbance between 200 and 230 nm, which are currently unknown)

Biochemical structure	Present configuration	UV-transparent organism	Biochemical and physiological limitations imposed by transparency
Genetic material	Absorbency in all five bases in the order A > G > U > T > C. Maximum value of ε of 13 400 M^{-1} cm^{-1} at 261 nm in adenine.	Use of guanazole, urazole and other heterocyclic compounds able to base pair, but possessing UV absorbance $\leq$100 M^{-1} cm^{-1} at wavelengths >230 nm.	Transparent genetic information can still act as a code, but its ability to interact with other molecules and respond to environmental cues via electron transfer is limited.
Proteins	UV absorbance >230 nm by phenylalanine, tyrosine, tryptophan and disulphide bonds (see Figure 1).	Primary structures contain no aromatic amino acids and secondary, tertiary or quaternary structures no disulphide bonds.	Active sites can bind substrates, but have limited ability to mediate reactions. Lack of disulphide bonds means that protein secondary and tertiary structure is dependent upon hydrogen bonding in structures such as anti-parallel beta sheets with proline used to generate hairpins.
Carbohydrates and sugars	Most have no significant UV absorption above 210 nm	No changes required	Widespread use of sugars and carbohydrates in cell processes in the place of proteins and other molecules would result in greater transparency.
Lipids	Unsaturated lipids create conjugated electron systems that absorb UV radiation.	All lipids saturated to minimize conjugation.	Lack of membrane fluidity. Particularly a problem for organisms at high pressures and extreme temperatures where membrane fluidity is important.
Other molecules	Some individual molecules have high UV absorbance. ATP and other nucleoside triphosphates have ε of 451 to 507 M^{-1} cm^{-1} at different wavelengths. ATP has ε of 507 M^{-1} cm^{-1} at 260 nm.	Selection for molecules with no UV absorbance. Would require novel energy capturing system not using ATP.	Many small molecules involved in charge transfer, energy capture etc. are UV absorbing. Would impose severe constraints on cell metabolism, signalling, and other intracellular and also intercellular processes in metazoans.

A question then arises as to whether UV radiation can impose on an exposed surface-dwelling biota a selection pressure for biochemical transparency to UV radiation. Table I summarizes the modifications to present-day terrestrial biochemistry that would be required to build an organism transparent to UV radiation above 230 nm. Apart from sugars, which are mostly UV transparent, fundamental changes are required. The changes are structurally possible. The fact that some RNA precursors that may have prebiotic significance, such as guanazole, are essentially transparent at wavelengths >230 nm has already been discussed in the context of the high UV flux postulated for early Earth (Kolb *et al*., 1994). In the absence of high pO_2 (>0.01 PAL [Present Atmospheric Level]) the ozone column may not have existed in the Archean and in the absence of other atmospheric UV absorbers such as sulfur or a hydrocarbon smog (Kasting *et al*., 1989; Sagan and Chyba, 1997) organisms could have been exposed to a UV flux with a DNA-weighted irradiance some three orders of magnitude higher than on present-day Earth (Cockell and Horneck, 2001) and wavelengths as low as 195 nm could have penetrated to the surface of the early Earth. This is, for example, the case for present-day Mars, which is a CO_2-dominated atmosphere lacking an ozone column over most latitudes during most of the year with its consequences for surface DNA-weighted irradiances (Cockell *et al*., 2000).

The concept of UV radiation acting as a selection pressure at the biochemical level is not new. Cockell (1998) points out that the UV selection pressure is even exerted at the basic chemical level, mediated by pi-electron systems. Because conjugated molecules absorb UV radiation at wavelengths greater than approximately 190 nm, primarily on account of their π to π^* electronic transitions (and to a some extent n to π^* transitions), then it is apparent that damage will be caused by the presence of these bonds and that these types of chemical bonds would evolve within UV screening compounds in particular structural configurations to prevent UV damage in the first place, providing an elegant example of Darwinian selection pressures operating at a most fundamental chemical and even quantum mechanical level.

The possibility that a selection pressure for UV transparency might have been exerted on the genetic apparatus on the early Earth, resulting in UV transparent base pairs, cannot be readily tested because we do not have access to genetic material of this age. The fact that phylogenetic trees can be constructed based on the present-day architecture does not mean that DNA and RNA were the same in the Archean because the base pairs could have changed chemical characteristics, whilst preserving their phylogenetic relationships. A similar argument applies to proteins, as the chemical characteristics of amino acids attached to transfer RNA's could have changed over the last 4.0 Gyr without the t-RNA's themselves undergoing much change.

However, another intriguing possibility for testing the hypothesis of a selection pressure for UV transparency results from the properties of tryptophan. Tryptophan has a λ_{max} at 280 nm, but the shoulder extends into the region of UV radiation seen

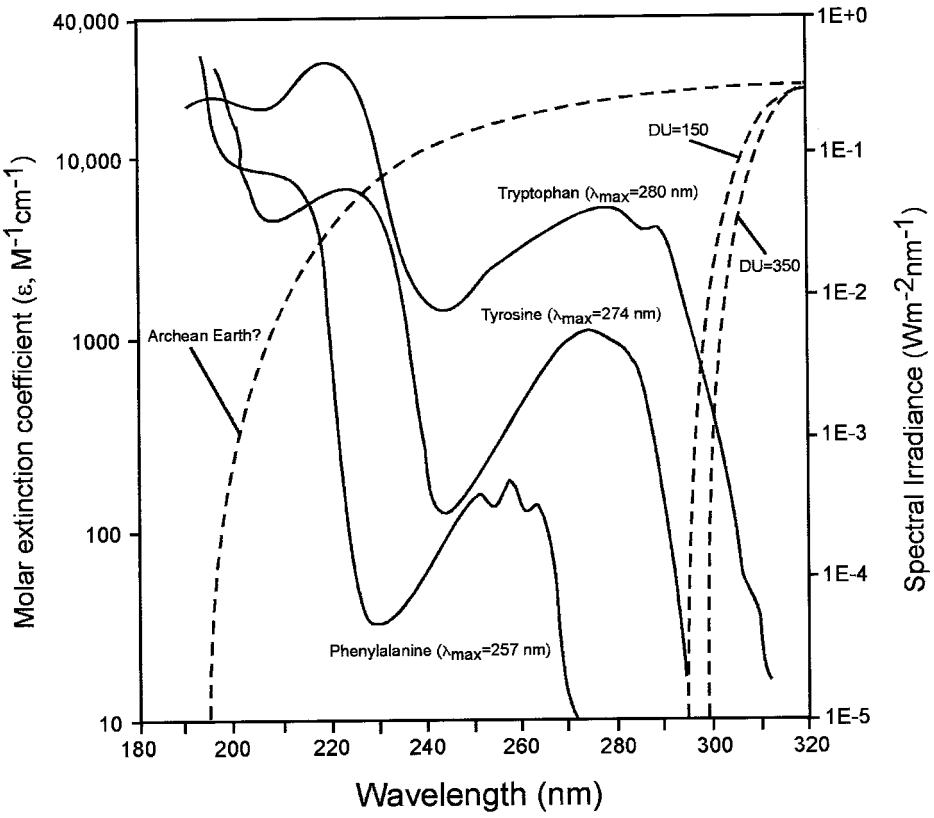

Figure 1. The absorption spectra of the three aromatic amino acids (adapted from Wetlaufer (1962) and Du *et al.* (1998)) at neutral pH. The molar extinction coefficient at λ_{max} is 5050 for tryptophan, 1440 for tyrosine and 220 M^{-1} cm^{-1} for phenylalanine. Also shown are the spectral irradiances for three atmospheric conditions; an ozone column abundance of 350 Dobson Units, the high end for ozone abundance for the undepleted polar stratosphere, a column abundance of 150 DU, which constitutes an approximately 50% depletion over the poles and the spectral irradiance that might have been received on the surface of the Archean (3.8–2.5 Ga ago) Earth assuming no UV absorbers other than CO_2 in the early atmosphere.

even on the present-day Earth (Figures 1 and 2). The absorbance of proteins in the UVB region (280–320 nm) is primarily determined by tryptophan and to a lesser extent tyrosine at wavelengths less than 290 nm. The absorbance of proteins in the UVB region is approximately proportional to the number of these residues in the protein. In Figure 2 the absorbance of the blue-copper electron transfer protein, azurin, and the hydrolytic enzyme, lysozyme, are shown. Azurin contains one tryptophan and lysozyme six. The absorbance of lysozyme is actually 8.8 times that of azurin at 280 nm, which is partly accounted for by the fact that lysozyme also contains one more tyrosine residue than azurin (a total of 3), which contributes towards absorbance at 280 nm.

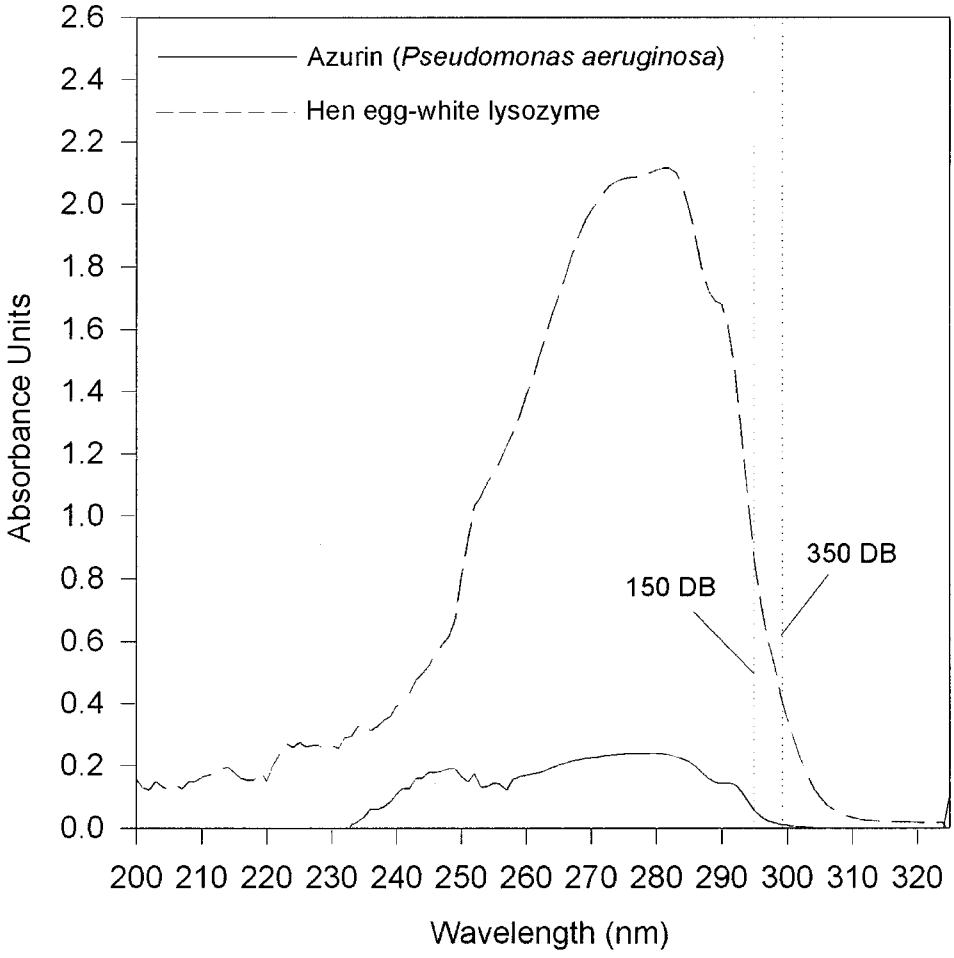

Figure 2. The absorbance characteristics in the UV-C (200–280 nm) and UV-B (280–320 nm) regions of 1 mg mL^{-1} of azurin (a blue-copper electron transfer protein from *Pseudomonas aeruginosa*) that possesses one tryptophan residue and 1 mg mL^{-1} hen egg-white lysozyme (a polysaccharide hydrolyzing enzyme) that possesses six tryptophan residues. The vertical dashed lines indicates the lower cut-off of UV radiation seen on present-day Earth under an ozone column abundance of 150 and 350 Dobson Units (see Figure 1 caption for the explanation for these values). On Archean Earth wavelengths as low as 200 nm may have reached the surface of the Earth.

The correlation between the number of tryptophan residues in a protein and the damage caused by absorbed UV radiation is not a simple linear relationship because it depends upon a number of factors. It will depend upon which UV-absorbing residues are critical for function, their microenvironment and the tertiary structure of a given protein. The quantum yield of damage in specific residues and types of residue will vary and it will be dependent upon the factors listed above. In a range of proteins the quantum yield of tryptophan inactivation by UV radiation at

254 nm varied from 0.016 to 0.18 (Grossweiner, 1976). Quantum yields have not been measured for present-day environmentally relevant UV ranges.

In some cases absorption of UV radiation causes disulphide bond breakage to occur, often mediated through energy transfer from aromatic amino acids. In egg hen-white lysozyme irradiated with 254 nm light, eighty-seven percent of the incident light is absorbed by tryptophan residues (Grossweiner, 1976) and 6% by disulphide bonds, but an equal number of tryptophan and disulphide bridges are destroyed, showing the higher quantum yield of disulphide bond breakage at 254 nm. Flash photolysis studies on the plant protease papain, lysozyme and other enzymes implicate tryptophan as the primary chromophore of UV-induced damage (Grossweiner, 1976), particularly at wavelengths >295 nm. Even if the tryptophan residue itself is not affected, energy transfer from tryptophan to non-absorbing sites can promote damage within the protein (Grossweiner, 1976).

The UV-absorbing properties of tryptophan have made it an important component of natural UV screening compounds. It acts as a precursor for the biosynthesis of kynurechine with a λ_{max} at 310 nm and 3-hydroxy kynurenine (λ_{max}, 368nm) (Takikawa *et al*., 2001) found in fish and mammalian eyes. The condensation of two tryptophan residues in cyanobacteria results in the formation of the UVA-screening compound, scytonemin (Proteau *et al*., 1993). Although these structural modifications increase the wavelength of maximum absorption, tryptophan itself is synthesized as a UV filter in the aqueous humour of avian eyes (Ringvold *et al*., 2000). This provides evidence that the absorbance shoulder of tryptophan is strong enough to provide a selection pressure *for* the use of this amino acid as a UV filter. This being the case, it is reasonable to look for evidence for selection *against* this absorbance in situations where it could be damaging, such as when tryptophan is incorporated into important enzymes and proteins.

It is the purpose of this article to examine the hypothesis that a selection pressure can exist for a UV transparent biochemistry.

2. Testing the Hypothesis

If we accept the hypothesis that UV radiation can act as a selection pressure for UV transparency then we can make some predictions about what we would expect to see manifested in protein sequences. Here we list four of these possible predictions and then examine each in more detail.

Prediction 1

A UV selection pressure would be stronger against amino acids with higher molar extinction coefficients in the UV region. Tryptophan, tyrosine and phenylalanine are the only amino acids with significant absorption >210 nm and so they primarily concern us here. Based on the molar extinction coefficients shown in Figure 1 we would predict the abundance of the UV-absorbing aromatic amino

acids to be tryptophan < tyrosine < phenylalanine under a UV environment that covered all wavelengths from 200 to 300 nm at roughly equal intensity (as could have been seen on Archean Earth in the absence of atmospheric UV absorbers). On present-day Earth wavelengths <295 nm are generally not encountered (except under ozone-depleted conditions), so we would at least predict an abundance tryptophan < tyrosine = phenylalanine. Note that the lack of UV-absorption in tyrosine at wavelengths > 295 nm is not true in all present-day environments. At a pH > 12 the λ_{max} for tyrosine shifts up to ~295 nm and the molar extinction coefficient, ε increases from ~60 to ~2500 M^{-1} cm^{-1}, so that a significant ε of ~1000 M^{-1} cm^{-1} exists at 305 nm. One might argue that tyrosine UV absorbance will be more environmentally problematic for organisms in basic environments, but most microbial alkalophiles maintain an internal cystosolic pH of ~7.2 by active proton transport (Alberts *et al.*, 1994) (although the microenvironmental protonation state of specific residues within proteins can be substantially different).

Prediction 2
The prevalence of tryptophan might be less in organisms exposed to sunlight compared to organisms that are either nocturnal, subterranean or have UV protective coatings such as fur, and are not exposed to UV radiation.

Prediction 3
The prevalence of tryptophan in enzymes specifically designed to be exposed to sunlight such as the deep branching UVA/blue-light activated enzyme photolyase (Kanai *et al.*, 1997), would be lower compared to enzymes that are not required to be exposed to light.

Prediction 4
If the postulate that wavelengths of UV radiation as low as 200 nm existed on the surface of early Earth is correct, we might predict all three amino acids to be less prevalent in deep branching organisms compared to organisms that did not exist on Archean Earth (with certain caveats discussed later).

How does each of these four predictions stand up to test and what, if anything, can we learn from the outcomes of the predictions? Over the last decade advances have been made in the sequencing of whole organismal genomes and proteins. This data allows us to examine these predictions. In this article protein sequences were obtained from SWISS-PROT, the annotated protein sequence database maintained by the Department of Medical Biochemistry at the University of Geneva and the European Bioinformatics Institute (May 2001). Additional sequences were also obtained from TrEMBL, a collection of sequences held by the same institutes, but not yet incorporated into SWISS-PROT. The accuracy of the sequences cannot be verified, but any errors in the sequences are unlikely to be greater than the variations imposed by the sample size of organisms studied. Non-redundant sequences were

TABLE II
Percentage of aromatic amino acids in the total proteome of selected organisms

Species	Tryptophan (W)	Tyrosine (Y)	Phenylalanine (F)	Total
Archaea				
Thermoplasma acidophilum	0.85	4.64	4.70	10.19
Methanobacterium thermoautotrophicum	0.84	3.22	3.65	7.71
Halobacterium NRC-1	1.09	2.55	3.12	6.76
Archaeoglobus fulgidus	1.04	3.64	4.59	9.27
Pyrococcus abyssi	1.18	3.83	4.35	9.36
Mean	1.00±0.13	3.43±0.63	4.08±0.55	8.66±1.24
Eubacteria				
Thermotoga maritima	1.10	3.58	5.19	9.87
Deinococcus radiodurans	1.35	2.30	3.16	6.81
Haemophilus influenzae	1.14	3.15	4.47	8.76
Escherichia coli	1.52	2.86	3.89	8.27
Mean	1.28±0.17	2.97±0.46	4.18±0.75	8.43±1.10
Eukaryotes				
Arabidopsis thaliana	1.27	2.88	4.32	8.47
Caenorhabditis elegans	1.11	3.20	4.96	9.27
Drosophila melanogaster	1.00	2.96	3.54	7.50
Homo sapiens	1.22	2.72	3.66	7.60
Mean	1.15±0.10	2.94±0.17	4.12±0.57	8.21±0.72
Total mean	1.13±0.18	3.19±0.59	4.12±0.64	9.45±1.08

used and only complete proteome or protein sequences were used. In the case of Table II a random selection of organisms was chosen from each kingdom from the Proteome Analysis Database as described by Apweiler *et al*. (2001). For Tables III to V a random selection of organisms were selected from the sequence list.

The analysis of amino acid composition was undertaken using the statistical methods described previously (Brendel *et al*., 1992). Here, abundance data for the three amino acids, tryptophan (W), tyrosine (Y) and phenylalanine (F) are

presented for selected proteins. The total aromatic amino acid composition is also provided.

3. Prediction 1

Table II shows the prevalence of the three aromatic amino acids in the whole proteomes of a variety of organisms from the archaea to multi-cellular eukaryotes. The mean for each group as well as the mean for all the organisms is shown at the bottom of the table. Of the three amino acids, tryptophan has the lowest mean abundance, being 65% less abundant than the mean abundance of tyrosine in all the organisms considered. Tyrosine is also found in lower abundance than phenylalanine, with a mean 22% less abundance in the proteomes of the organisms considered.

The difference in the mean abundances of the different aromatic amino acids is significant when evaluated by one-way ANOVA ($F_{2,37}$, $p < 0.01$). The relative abundances are in the same order of decreasing extinction coefficients at λ_{max} shown in Figure 1. This is the sequence of abundances that we would expect if a UV selection pressure did exert an influence on amino acid content. However, the difference between tyrosine and phenylalanine would only be expected if wavelengths below 295 nm penetrated to the surface of the Earth at some point in its past, resulting in a preferential selection against tyrosine.

On the basis of Table II then, the data is consistent with our predictions on the prevalence of the aromatic amino acids under a UV selection pressure, given certain caveats for tyrosine and phenylalanine, but it is completely consistent with our predictions on the prevalence of tryptophan. What about the expected patterns of amino acid usage based on the characteristics of the genetic code?

We can estimate how the abundance of a given amino acid compares to the abundance expected from random mutation processes by calculating the probability that a random single-point mutation in the DNA code will result in the generation or removal of that amino acid in a given protein of known primary sequence. Our calculations include nonsense and missense mutations, either transitions or transversions, but we do not include frameshift mutations. The degeneracy of the genetic code can be used to calculate these probabilities. Tryptophan is coded for by one codon (a triplet of bases which codes for an amino acid) and tyrosine and phenylalanine by two codons each (either codon can be used to code for the amino acid). Thus, it is a relatively simple task, for example, to calculate the probability that a single random mutation in a sequence with a given abundance of tryptophan codons will cause a non-tryptophan codon to turn into a tryptophan codon and a tryptophan codon to turn into a non-tryptophan codon.

For tryptophan, using the mean abundance across all the proteomes examined, we can calculate that the probability that a random single-point mutation in a DNA sequence with this abundance will result in loss of a tryptophan is 0.0113, but the

probability of a mutation of a non-tryptophan residue into a tryptophan is 0.0154. In other words, there is less use of tryptophan in the proteomes than would be expected. This result would be consistent (although not necessarily caused by) a UV selection pressure against its use.

Given the degeneracy of the genetic code we can calculate the abundance at which the probability of loss and gain of a tryptophan residue is the same. It is about 1.5%. This percentage is very similar to the abundances in the proteomes of many organisms (Table II). It results in two conclusions. Firstly, that it is not possible to assign the low abundance of tryptophan to a selection pressure against its use in many organisms because the abundance is roughly what would be expected from amino acid codon use and secondly, that a small variation from this percentage is impossible to assign either to a UV selection pressure or other factors that might vary the abundance of tryptophan (some of which are discussed later).

We can perform the same calculation with tyrosine and phenylalanine using the mean proteome abundances. In the case of tyrosine the probability of mutational loss of a tyrosine is ∼20% greater than the probability of gaining one. This percentage is ∼54% for phenylalanine. These percentages suggest that the abundance of these two amino acids is higher than would be expected based on mutational considerations, suggesting that they are useful to organisms and selected for.

4. Prediction 2

The second prediction that there would be less tryptophan in UV-exposed organisms compared to UV-shielded organisms can be tested by examining a protein for which sequences are available in both exposed and shielded organisms (shielded includes surface-dwelling organisms that have fur or scales and therefore have the majority of their enzymes protected from UV radiation). Here, we have examined the protein cytochrome c oxidase (polypeptide I). This enzyme is the terminal enzyme in the respiratory chain. It is located in the inner membrane of mitochondria and bacteria and catalyses the reduction of oxygen to water. The enzyme is therefore rather essential and it is deep-branching. Table III provides the analysis of the amino acid composition of cytochrome c oxidase polypeptide I for various organisms. When a comparison is made between exposed and shielded organisms by a student *t*-test, only tryptophan shows a significant ($p < 0.01$) difference, being lower for UV exposed organisms compared to UV shielded organisms. Given that the wavelengths that reach the surface of present-day Earth would be expected to select against tryptophan, then this result is consistent with the prediction. However, an identical analysis using the same organisms was undertaken for ATP synthase chain A, part of the enzyme responsible for the production of ATP from ADP and Pi (data not shown). In this enzyme there was no significant difference at the 95 or 99% confidence level in the abundance of the separate aromatic amino acids between the shielded and UV exposed organisms.

TABLE III

Percentage of aromatic amino acids in Cytochrome c oxidase polypeptide 1 with a comparison between nocturnal/subterranean/fur-shielded and surface UV exposed eukaryotes

Species	Tryptophan (W)	Tyrosine (Y)	Phenylalanine (F)	Total
Eubacteria and Archaea				
Halobacteriium halobium	3.50	3.00	7.80	14.3
Thermus aquaticus	4.30	3.90	6.60	14.8
Bacillus PS2	2.90	2.60	11.20	16.7
Synechococcus vulcanus	3.80	3.80	6.90	14.5
Cyanidium caldarium	2.90	3.80	8.70	15.4
Mean	3.48±0.54	3.42±0.52	8.24±1.65	15.14±0.86
Eukaryotes				
UV-shielded organisms				
Dasypus novemcinctus (armadillo)	3.30	3.90	8.20	15.4
Ascaris suum (pig roundworm)	2.50	5.10	9.30	16.9
Balaenoptera musculus (blue whale)	3.50	3.70	8.30	15.5
Lumbricus terrestris (earthworm)	3.10	3.10	8.80	15.0
Oryctolagus cuniculus (rabbit)	3.30	3.50	8.40	15.2
Capra hircus (goat)	3.40	3.40	5.20	12.0
Rattus norvegicus (rat)	3.30	3.90	8.20	15.4
Talpa europaea (European mole)	3.30	3.70	8.20	15.2
Mean	3.18±0.28	3.79±0.55	8.07±1.14	15.07±1.28
UV-exposed organisms				
Locusta migratoria (migratory locust)	3.10	3.10	7.40	13.6
Artemia sanfranciscana (brine shrimp)	2.70	2.90	8.00	13.6
Haemophilus influenzae	2.30	3.80	8.40	14.5
Aspergillus nidulans	2.60	3.70	8.50	14.8
Oryza sativa (rice)	2.70	3.40	9.40	15.5
Arabidopsis thaliana (mouse-ear cress)	2.70	3.60	9.30	15.6
Zea mays (maize)	2.70	3.40	9.30	15.4
Helianthus annuus (sunflower)	0.90	5.40	3.60	9.9
Mean	2.46±0.62	3.66±0.71	7.99±1.78	14.11±1.76

Although one protein does support the hypothesis and the other one does not, several facts should be pointed out. Firstly, many of the 'exposed' organisms produce UV-screening compounds. In many cases the cuticle of insects is UV opaque (Buck and Callaghan, 1999) and plants respond to UV radiation by the production of flavonoids (Day, 1993). Enzymes from these organisms may not necessarily receive higher UV radiation than some of the 'shielded' organisms. Secondly, the premise that shielded organisms will, over time, gather a higher percentage of aromatic amino acids compared to exposed ones does not necessarily hold. As described under Prediction 1, the prevalence of tryptophan in whole proteomes is apparently underrepresented when mutation and the degeneracy of the genetic code are taken into account, but by only $\sim$15%. In the absence of any other selection pressure it is not necessary the case that the prevalence of tryptophan will increase to above 1.5% when the probability of a single-point mutation resulting in the loss of this residue will exceed the probability of generation. Therefore, in UV shielded organisms it is difficult to differentiate a selection pressure against tryptophan from the low abundance expected from the degeneracy of the genetic code. In the case of cytochrome c oxidase polypeptide I the difference in abundance between shielded and exposed organisms is only 23%.

5. Prediction 3

Is tryptophan abundance significantly lower in an enzyme known to be exposed to UV radiation compared to the whole proteome? We examined the tryptophan abundance in photolyase, responsible for dipyramidine dimer removal in DNA. The enzyme is activated by UVA and blue light at wavelengths between 350 and 500 nm and so therefore we know that in species where this photoreactivation pathway is active, the enzyme is exposed to UV light.

The abundance of tryptophan in all the photolyases examined (Table IV) is statistically greater than the mean of the total proteomes of the organisms examined here (t-test, $p < 0.01$). The mean tryptophan abundance in the photolyase enzymes across all genera examined is 2.93%, compared to 1.13% for the total proteomes (note that these are not the same species as we do not have the total proteome sequences for many species for which we have photolyase sequences). The tryptophan abundances in photolyase are also much higher than the abundances expected from the degeneracy of the genetic code. The reasons for this are likely to be similar to the reasons for the high abundance in cytochrome c oxidase polypeptide I (Table III). The function of photolyase is to cleave pyramidine dimers in DNA. The dimers, which form in DNA primarily by UV irradiation, are split by electron transfer. Although the functions of all of the tryptophans in photolyase have not been assigned, many have (Sancar, 1994). It is known that Trp222, Trp271 and Trp338 in *E.coli* photolyase are involved in the binding to the cofactor FAD (flavin adenine dinucleotide) and Trp306 is involved in electron transfer to FAD (Park

TABLE IV

Percentage of aromatic amino acids in photolyase from a diversity of organisms. In the case of archaea and bacteria, yeasts, and some eukaryotes the photolyases are classified as deoxyribodipyramidine photolyases in the SWISS-SPOT and TreEMBL databases. For other eukaryotes, (6–4) refers to photolyases that cleave pyramidine (6,4)pyramidone photoproducts and (CPD) refers to enzymes that cleave cyclobutane pyramidine dimers

Species	Tryptophan (W)	Tyrosine (Y)	Phenylalanine (F)	Total
Archaea				
Methanobacterium thermoautotrophicum	2.0	4.9	4.0	10.9
Halobacteriium halobium	2.5	4.0	4.0	10.5
Sulfolobus acidocaldarius	2.0	6.8	4.8	13.6
Mean	2.17±0.23	5.23±1.17	4.27±0.38	11.67±1.38
Eubacteria				
Thermus aquaticus	3.6	3.8	3.3	10.7
Esherichia coli	3.2	3.0	4.0	10.2
Synechococcus leopoliensis	3.5	2.9	3.7	10.1
Salmonella typhimurium	3.1	3.3	4.9	11.3
Vibrio cholerae	3.0	5.4	4.6	13.0
Mean	3.28±0.23	3.68±0.91	4.10±0.58	11.06±1.06
Eukaryotes				
Saccharomyces cerevisiae (Naker's yeast)	2.8	5.5	4.4	12.7
Carassius ayratus (deoxyribodipyramidine) (goldfish)	2.7	3.6	3.6	9.9
Brassica hirta (deoxyribopyramidine) (white mustard)	4.0	2.6	3.6	10.2
Homo sapiens (human photolyase-like sequence)	2.9	3.4	4.4	10.7
Arabidopis thaliana (6–4) (mouse-ear cress)	2.8	4.5	5.0	12.3
Arabidopis thaliana (CPD) (mouse-ear cress)	2.8	3.6	3.4	9.8

TABLE IV
(continued)

Species	Tryptophan (W)	Tyrosine (Y)	Phenylalanine (F)	Total
Eukaryotes (continued)				
Xenopus laevis (6–4) (African clawed frog)	3.4	3.5	4.6	11.5
Drosophila melanogaster (6–4) (Fruit fly)	2.8	3.5	4.4	10.7
Brachydanio rerio (6–4) (Zebrafish)	3.3	3.1	4.8	11.2
Oryzias latipe (deoxyribodipyramidine) (Medaka fish)	2.6	2.6	5.0	10.2
Monodelphis domestica (type not specified) (opossum)	2.8	3.6	5.5	11.9
Mus musculus (mouse)	2.8	3.5	4.1	10.4
Mean	2.97±0.38	3.58±0.75	4.40±0.61	10.96±0.92

et al., 1995). Trp277 and Trp384 are involved in dimer binding and Trp277 is responsible for electron transfer to the dipyramidine dimer (Park *et al*., 1995) that results in it splitting back to the original two pyramidines.

The high tryptophan abundance in photolyase is because of the requirement for these electron transfer reactions made possible by delocalised electrons. A similar argument can be advanced for cytochrome c oxidase. The mean abundance of tryptophan in cytochrome c oxidase polypeptide 1 is 2.97% across all the genera examined, above the value at which the probability of a single-point mutation resulting in the loss of this residue will exceed the probability of generation. Cytochrome c oxidase, like photolyase (Table III), is involved in electron transfer reactions. Tryptophan 121 of subunit II, for example, is known to be the electron entry site in the enzyme (Witt *et al*., 1998).

The high tryptophan abundances in these enzymes, and particularly photolyase, illustrate that a biochemical trade-off might exist in enzymes between the potentially damaging result of the UV absorbance of conjugated or de-localized pi-electron systems and the beneficial role of these systems in electron transfer (interestingly the monomerization of dimers can be catalysed by photosensitised free tryptophan in solution). Important roles are also assigned for tyrosine and phenylalanine residues in photolyase, as for other enzymes (see for example Pai *et*

TABLE V
Percentage of aromatic amino acids in RNA Polymerase Sigma Factor of various bacteria

Species	Tryptophan (W)	Tyrosine (Y)	Phenylalanine (F)	Total
Chloroflexus aurantiacus	0.50	2.00	1.50	4.00
Thermotoga maritima	1.60	4.70	4.70	11.00
Deinococcus radiodurans	1.30	3.90	3.10	8.30
Thermus aquaticus	0.70	1.90	2.40	5.00
Nostoc punctiforme	0.80	2.60	2.11	5.51
Synechocystis sp.	1.60	2.20	2.40	6.20
Vibrio parahaemolyticus	0.90	2.20	2.50	5.60
Escherichia coli	0.70	2.00	2.50	5.20
Mean	1.01±0.40	2.69±0.97	2.65±0.88	6.35±2.10

al. (1990) and Befort *et al*. (1996)). The high abundances of tryptophan in a light exposed enzyme compared to the whole proteome suggests that evolution weighs in favour of biochemical versatility against UV transparency.

6. Prediction 4

The final prediction is that deep-branching organisms might show evidence of an intense Archean UV environment. In Table V, we provide an aromatic amino acid sequence analysis for RNA polymerase sigma factor, an essential and well conserved protein involved in transcription in eubacteria. Some of the organisms represented such as *Thermotoga, Chloroflexus and Thermus* are proposed to be deep-branching. *Chloroflexus* has specifically been examined for resistance to the high UV radiation proposed for Archean Earth (Pierson *et al*., 1993). The mean tryptophan abundance is not significantly lower than in *Esherichia coli* and individually many of these organisms have higher abundances.

In cytochrome c oxidase, the abundances of tyrosine and phenylalanine are not statistically lower in the archaea and eubacteria examined ($p < 0.01$) compared to the eukaryotes, many of which are deep-branching and tryptophan is statistically *higher* in the archaea.

In the total proteomes the differences between the abundances of the aromatic amino acids between archaea and eukaryotes are not statistically significant ($p < 0.01$) either individually or considered as a total aromatic amino acid content. The results obtained here imply that no record of a UV selection pressure against UV absorbing amino acids can be found in the archaea in the proteins examined. There

are a number of possible explanations for this as well as those presented in the next section. Many of the archaea examined are hot-vent organisms (e.g., *Pyrococcus* and *Archaeoglobus* spp.) and not exposed to UV radiation. One would not expect to find a difference. One could argue that maybe the biologically effective irradiances of UV radiation were not higher on early Earth than today, perhaps because of the presence of atmospheric UV absorbers (Kasting *et al.*, 1989; Sagan and Chyba, 1997) or the presence of oxygen in the early atmosphere (Towe, 1996), and this is the reason for the lack of evidence. Similarly to the arguments presented earlier, the importance of tryptophan in enzymatic functions and protein structure-function might also result in its presence, regardless of any other selection pressures. Finally, in the absence of a UV selection pressure since the formation of an ozone column $\sim$2.25 Ga one would expect that random mutations would, over time, bring the number of aromatic amino acids at least up to a percentage consistent with the degeneracy of the genetic code.

7. Other Factors in the Selection of Protein UV Transparency

The lack of evidence of UV transparency in proteins is likely to be compounded by a number of factors. Firstly, aromatic amino acids have more functions than just electron transfer reactions. Their hydrophobicity makes them an important contributor to protein stability and tertiary structure (e.g., Georis *et al.*, 2000). They also take part in Van der Waals interactions and thus in substrate binding within active sites (Raine *et al.*, 1995), quite apart from the reactions that occur at the active site.

The selection pressure for UV transparency will depend upon the turnover time of enzymes which is found to vary from several minutes to over 160 hr depending upon enzyme, tissue, developmental state and organism (Price and Stevens, 2000). Proteins that turn over rapidly are less likely to be affected by UV damage than those that linger in the cell for a long time. We know little about the turnover time of photolyase or any of the enzymes in the diversity of organisms discussed here. Nor do we know about their rates of UV damage, so it is impossible to speculate on the damage versus turnover times of these proteins and thus what pressures these parameters might place on UV transparency.

The selection pressure will also depend upon the location and concentration of proteins. Proteins that are freely moving in the cytosol with high abundance will have more copies available to do the required job and so UV damage might be less important than, for example, proteins that are in a fixed location in just a few copies, such as the reaction centers of the photosynthetic apparatus. We know little of how translation levels of proteins relate to UV damage and thus efficacy of protein function. However, de novo synthesis of the D1 and D2 reaction centers in the cyanobacterium *Synechocystis* sp. has been demonstrated in response to UV radiation (Sass *et al.*, 1997). The evolutionary significance of this must be

interpreted with caution, but it is consistent with the notion that in some instances up-regulation of protein synthesis is the way in which organisms deal with UV damage in critical proteins because the importance of the UV chromophores in other functions prevents the evolution of UV transparency as a response to UV damage.

Finally the study here does not discriminate between functional aromatic amino acids and ones that have no specific function and could be replaced by a non-UV absorbing residue. Because studies on proteins focus on determining the functional role of particular residues, there is a paucity of data on these non-functional residues. A better study of UV transparency might focus on a studying the difference, for example, in the abundance of non-critical aromatics in the backbone of proteins in microorganisms exposed and not exposed to UV radiation.

8. Genetic UV Transparency?

There are some RNA precursors that are essentially transparent to UV radiation greater than 230 nm. Urazole, cytazole and guanazole are five-membered heterocyclic compounds that mimic uracil and cytosine and can undergo base pairing (Kolb *et al*., 1994). Urazole and Guanazole are UV transparent (Kolb *et al*., 1994) and it has been suggested that they could have been used in a genetic code that was less prone to UV-induced damage on early Earth. We do not have organisms that have genetic codes with these bases and so the same quantitative examination that we presented with proteins cannot be repeated.

One could construct an argument that the reason why DNA is UV-absorbing is because the present structure of DNA became locked into organisms that evolved in deep-sea hydrothermal vents where UV transparency was not required. Once they evolved onto the surface of the Earth it was too late to change such a fundamental part of organismal biochemistry.

However we speculate that a more fundamental trade-off confronted early life on Earth that is identical to the trade-off experienced by present-day organisms with the use of tryptophan in proteins. Whereas it is true that base structures can be found that are UV transparent (Kolb *et al*., 1994), the lack of UV absorbance >230 nm is because of the rigidity of their electron systems. They are limited in their electron transfer reactions compared to the delocalized pi-electron-containing base pairs found in the present structure of DNA. DNA is not merely a code. It interacts with proteins and other molecules to allow an organism to respond to its environment. The difference between a chimpanzee and a human is largely because of difference in the *expression* of genes, not so much because of the 1% difference in the genetic code itself.

Our knowledge of electron transfer reactions in DNA is still very much in its infancy. However, it is becoming apparent that the stacked pi-electron systems of DNA not only confer stability to the double helix (Saenger, 1984), but also allow

for charge transfer along the length of the DNA (Porath *et al*., 2000) of distances between 6 to 8 base pairs (Poy and Cornelio, 1998). This characteristic apparently allows damage to DNA to be repaired from a remote site in the DNA (Hall *et al*., 1996) and goes some way to explaining how damage recognition processes (not just for UV damage, but many other environmental assaults as well) can be so effective over an entire genome length (Rajski *et al*., 2000). Evidence suggests that the delocalised pi-electron systems in DNA are also involved in protein-DNA interactions (Wagenknecht *et al*., 2000) that allow DNA to interact with, and respond to, the intracellular (and thus ultimately extracellular) environment. Charge migration between tryptophan and DNA has also been demonstrated as a mechanism for peptide – DNA interactions (Wagenknecht *et al*., 2000).

The particle in the box model of electrons predicts that the greater the distance over which delocalized electrons have occupancy, the longer the wavelength of light they absorb. Thus, in simplistic terms it seems that an unfortunate coincidence of the physical world is that to have delocalized electrons that are free to move over molecular lengths and can thus take part in electron transfer reactions from molecule to molecule, it is necessary to accept the cost of absorbancy at wavelengths of UV radiation encountered on planetary surfaces (i.e. between 195–400 nm, depending on atmospheric composition). Although this energy can potentially be non-destructively dissipated, it is energetically difficult to do this with 100% efficiency, as the formation of DNA lesions and UV-induced protein damage attests.

For the nucleic acid world to make the transition to single-celled organisms that responded to their environment, the first successful self-replicating genetic machinery was the result of a fundamental biochemical concession in a system possessing delocalised pi-electrons with the concomitant cost of UV-absorbancy. The versatility of a genetic machinery possessing delocalised electrons and therefore UV absorbancy more than outweighed the cost of DNA repair and the production of UV-screening compounds. Ultimately this fact probably contributed to the evolution of UV screening and repair mechanisms. These mechanisms allow an organism to address the problem of UV radiation damage, whilst making use of the versatility of an electronic biochemistry that is also UV absorbing. This hypothesis would provide an explanation for the apparently paradoxical UV absorbance of the very machinery responsible for the faithful replication of information.

References

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D.: 1994, *Molecular Biology of the Cell*, Garland Publishing, New York, pp. 1294.

Apweiler, R., Biswas, M., Fleischmann, W., Kanapin, A., Karavidopoulou, Y., Kersey, P., Kriventseva, E. V., Mittard, V., Mulder, N., Phan, I. and Zdobnov, E.: 2001, Proteome Analysis Database: Online Application of Interpro and Clustr for the Functional Classification of Proteins in Whole Genomes, *Nucleic Acids Res.* **29**, 44–48.

Befort, K., Tabbara, L., Kling, D., Maigret, B. and Kieffer, B. L.: 1996, Role of Aromatic Transmembrane Residues of the Delta-Opoid Receptor in Ligand Recognition, *J. Biol. Chem.* **271**, 10161–10168.

Brendel, V., Bucher, P., Nourbaksh, I., Blaisdell, B. E. and Karlin, S.: 1992, Methods and Algorithms for Statistical Analysis of Protein Sequences, *Proc. Natl. Acad. Sci.* **89**, 2002–2006.

Buck, N. and Callaghan, T. V.: 1999, The Direct and Indirect Effects of Enhanced UV-B on the Moth Caterpillar *Epirrita Autumnata*, *Ecol. Bull.* **47**, 68–76.

Cockell, C. S.: 1998, Ultraviolet Radiation, Evolution and the π-Electron System, *Biol. J. Linnean Soc.* **62**(3), 449–457.

Cockell, C. S. and Horneck, G.: 2001, The History of the UV Radiation Climate of Earth – Theoretical and Space-Based Observations, *Photochem. Photobiol.* **73**, 447–451.

Cockell, C. S., Catling, D., Davis, W. L., Kepner, R. N., Lee, P. C., Snook, K. and Mckay, C. P.: 2000, The Ultraviolet Environment of Mars: Biological Implications Past, Present and Future, *Icarus* **146**, 343–359.

Day, T. A.: 1993, Relating UV-B Radiation Screening Effectiveness of Foliage to Absorbing-Compound Concentration and Anatomical Characteristics in a Diverse Group of Plants, *Oecologia* **95**, 542–550.

Du, H., Fuh, R. A., Li, J., Corkan, A. and Lindsey, J. S.: 1998, Photochemcad: A Computer Aided Design and Research Tool in Photochemistry, *Photochem. Photobiol.* **68**, 141–142.

Georis, J., De Lemos Esteves, F., Lamotte-Brasseur, J., Bougnet, V., Devreese, B., Giannotta, F., Granier, B. and Frère, J.-M.: 2000, An Additional Aromatic Interaction Improves the Themostability and Themophilicity of a Mesophilic Family 11 Xylanase: Structural Basis and Molecular Study, *Protein Science* **9**, 466–475.

Grossweiner, L. I.: 1976, Photochemical Inactivation of Enzymes, *Curr. Topics Rad. Res.* **11**, 141–199.

Hall, D. B., Holmlin, R. E. and Barton, J. K.: 1996, Oxidative Damage Through Long-Range Electron Transfer, *Nature* **382**, 731–735.

Helbling, E. W., Buma, A. G. J., De Boer, M. K. and Villafane, V. E.: 2001, *In Situ* Impact of Solar Ultraviolet Radiation on Photosynthesis and DNA in Temperate Marine Phytoplankton, *Marine Ecol. Prog. Ser.* **211**, 43–49.

Kanai, S., Kikuno, R., Toh, H., Ryo, H. and Todo, T.: 1997, Molecular Evolution of the Photolyase-Blue-Light Photoreceptor Family, *J. Mol. Evol.* **45**, 535–548.

Kasting, J. F., Zahnle, K. J., Pinto, J. P. and Young, A. T.: 1989, Sulfur, Ultraviolet Radiation, and the Early Evolution of Life, *Orig. Life Evol. Biosph.* **19**, 95–108.

Kolb, V. M., Dworkin, J. P. and Miller, S. L.: 1994, Alternative Bases in the RNA World: The Prebiotic Synthesis of Urazole and its Ribosides, *J. Mol. Evol.* **38**, 549–557.

Michael, R., Vrensen, G. F. J. M., Van Marle, J., Lofgren, S. and Soderberg, P. G.: 2000, Repair in the Rat Lens after Threshold Ultraviolet Radiation Injury, *Invest. Opthalmol. Vis. Sci.* **41**, 204–212.

Nara, K., Nagashima, F. and Tasui, A.: 2001, Highly Elevated Ultraviolet-Induced Mutation Frequency in Isolated Chinese Hamster Cell Lines Defective in Nucleotide Excision Repair and Mismatch Repair Proteins, *Cancer Research* **61**, 50–52.

Pai, E. F., Krengel, U., Petsko, G. A., Goody, R. S., Kabsch, W. and Kittinghofer, A.: 1990, Refined Crystal Structure of the Triphosphate Conformation of H-Ras P21 at 1.35 Å Resolution: Implications for the Mechanism of GTP Hydrolysis, *EMBO J.* **9**, 2351–2359.

Park, H.-W., Kim, S.-T., Sancar, A. and Deisenhofer, J.: 1995, Crystal Structure of DNA Photolyase from *Escherichia Coli*, *Science* **268**, 1866–1872.

Pierson, B. K., Mitchell, H. K. and Ruff-Roberts, A. L.: 1993, *Chloroflexus aurantiacus* and Ultraviolet Radiation: Implications for Archean Shallow-Water Stromatolites, *Orig. Life Evol. Biosph.* **23**, 243–260.

Porath, D., Bezryadin, A., De Vries, S. and Dekker, C.: 2000, Direct Measurement of Electrical Transport through DNA Molecules, *Nature* **403**, 635–638.

Poy, C. D. and Cornelio, M. L.: 1998, Electron Migration in DNA Matrix: An Electron Transfer Reaction, *Ecletica Quimica* **23**, 99–109.

Price, N. C. and Stevens, L.: 2000, *Fundamentals of Enzymology*, Oxford University Press, Oxford, pp. 478.

Proteau, P. J., Gerwick, W. H., Garcia-Pichel, F. and Castenholz, R. W.: 1993, The Structure of Scytonemin, an Ultraviolet Sunscreen Pigment from the Sheaths of Cyanobacteria, *Experientia* **49**, 825–829.

Raine, A. R. C., Y, C.-C., Packman, L. C., White, S. A., Matthews, F. S. and Scrutton, N. S.: 1995, Protein Recognition of Ammonium Cations using Side-Chain Aromatics, *Prot. Sci.* **4**, 2625–2628.

Rajski, S. R., Jackson, B. A. and Barton, J. K.: 2000, DNA Repair: Models from Damage and Mismatch Recognition, *Mutation Res.* **447**, 49–72.

Ringvold, A., Anderseen, E. and Kjonniksen, I.: 2000, UV Absorption by Uric Acid in Diurnal Bird Aqueous Humor, *Investigative Opthalmology and Visual Science* **41**, 2067–2069.

Rothschild, L. J.: 1999, The Influence of UV Radiation on Protistan Evolution, *J. Euk. Micro.* **46**(5), 548–555.

Saenger, W.: 1984, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Sagan, C. and Chyba, C.: 1997, The Faint Young Sun Paradox: Organic Shielding of Ultraviolet-Labile Greenhouse Gases, *Science* **276**, 1217–1221.

Sancar, A.: 1994, Structure and Function of DNA Photolyase, *Biochemistry* **33**, 2–9.

Sass, L., Spetea, C., Mate, Z., Nagy, F. and Vass, I.: 1997, Repair of UV-B Induced Damage of Photosystem II via *De Novo* Synthesis of the D1 and D2 Reaction Centre Subunits in *Synechocystis* sp., PCC 6803, *Photosyn. Res.* **54**, 55–62.

Takikawa, O., Littlejohn, T. K. and Truscott, R. J. W.: 2001, Indoleamine 2,3-Dioxygenase in the Human Lens, the First Enzyme in the Synthesis of UV Filters, *Experimental Eye Res.* **72**, 271–277.

Towe, K. M.: 1996, Environmental Oxygen Conditions during the Origin and Early Evolution of Life, *Adv. Space Res.* **18**, (12)7–(12)15.

Wagenknecht, H. A., Stemp, E. D. A. and Barton, J. K.: 2000, DNA-Peptide Radicals Generated through DNA-Mediated Electron Transport, *Biochemistry* **39**, 5483–5491.

Wetlaufer, D. B.: 1962, Ultraviolet Spectra of Proteins and Amino Acids, *Adv. Prot. Chem.* **17**, 303–390.

Witt, H., Malatesta, F., Nicoletti, F., Brunori, M. and Ludwig, B.: 1998, Trytophan 121 of Subunit II is the Electron Entry Site to Cytochrome C Oxidase in *Paracoccus Denitrificans* – Involvement of a Hydrophobic Patch in the Docking Reaction, *J. Biol. Chem.* **273**, 5132–5136.

Yokoyama, S. and Shi, Y. S.: 2000, Genetics and Evolution of Ultraviolet Vision in Vertebrates, *FEBS Lett.* **486**, 167–172.