Commentary
DOI 10.1007/s12038-014-9414-3

Biochemistry and evolutionary biology: Two disciplines that need each other†

Biochemical information has been crucial for the development of evolutionary biology. On the one hand, the sequence information now appearing is producing a huge increase in the amount of data available for phylogenetic analysis; on the other hand, and perhaps more fundamentally, it allows understanding of the mechanisms that make evolution possible. Less well recognized, but just as important, understanding evolutionary biology is essential for understanding many details of biochemistry that would otherwise be mysterious, such as why the structures of NAD and other coenzymes are far more complicated than their functions would seem to require. Courses of biochemistry should thus pay attention to the essential role of evolution in selecting the molecules of life.

[Cornish-Bowden A, Peretó J and Cárdenas ML 2014 Biochemistry and evolutionary biology: Two disciplines that need each other. *J. Biosci.* **39** 13–27] **DOI** 10.1007/s12038-014-9414-3

1. Introduction

In a recent commentary in this journal, Valdecasas *et al.* (2013) argued that 'there is no common ground between science and religion', a sentiment with which we agree wholeheartedly. Unfortunately, however, evolution is a scientific topic that arouses so much hostility in some religious readers that some editors consider that one should avoid treading on their sensibilities. Needless to say we disagree, and in this article we shall discuss a proposition based on the well-known statement of Theodosius Dobzhansky (1973) that 'nothing makes sense in biology except in the light of evolution', which could have served as the title of this article: very little in *biochemistry* makes sense except in the light of evolution; very little in evolution makes sense except in the light of biochemistry. It is now widely recognized, at least by biochemists if not by more traditionally minded palaeontologists, that an understanding of evolution requires knowledge of biochemistry. It is less well recognized, however, that knowledge of evolution is necessary for understanding biochemistry.

Without biochemistry we would know nothing about the mechanisms of the changes that allow evolution, and we would not understand why evolution occurs at all – why do organisms not remain fixed for all time? Moreover, it is the biochemical evidence that has produced the modern consensus that the human and chimpanzee lines separated much more recently than the anatomical evidence suggested, and that they are more closely related to one another than either is to the gorilla. Although there is now broad acceptance that biochemistry is needed for a full appreciation of evolution, we shall argue here that the converse is also true, that there is a great deal in biochemistry that makes little sense if not seen in the light of evolution. It follows then that biochemistry ought to be taught with a strong emphasis on the evolutionary record. However, although most modern textbooks of biochemistry make some mention of evolution, it tends to be confined to a chapter on its own. A few books, such as that of Berg *et al.* (2012), integrate ideas of evolution into their presentation of the whole subject, but these are exceptional.

Keywords. Biochemistry; biological design; evolution; LUCA; NAD; RNA world

†This paper is dedicated to the memory of Professor Tito Ureta (1935–2012), in recognition of his great contributions to evolutionary biochemistry.

Published online: 27 January 2014

One of the triumphs of modern evolutionary investigation was the discovery of *Tiktaalik roseae* on Ellesmere Island, in the Canadian Arctic (Daeschler *et al.* 2006). Its importance is not so much that it demonstrated the existence of intermediate species between bony fish and tetrapods – something no biologist doubted – but that knowledge of phylogeny predicted when it should have lived, geology predicted where fossils ought to be located, and exploration of one of the most inhospitable regions of the world confirmed both predictions. Although biochemistry certainly played a part in the predicted date of 375 million years ago, this is not a particularly biochemical example. However, there is a less famous example from protein chemistry that similarly confirmed the relevance of biochemistry to evolution 30 years ago, when Penny *et al.* (1982) showed that similar phylogenetic trees were obtained for 11 species using sequence data from five different proteins. In fact, this is just one example of how studies with biochemical data have illuminated the whole of evolutionary science in recent decades. This has not always been well understood: for example, after examining the controversy over the dating of the separation of the various primate species, Lewin (1987) summarized the view of palaeontologists in the words 'The biochemistry is wrong because it doesn't agree with the fossils. Period.'

2. What biochemistry can tell us about evolution

2.1 *Classifying organisms*

Before protein and later gene sequences became available, classification of organisms and construction of phylogenetic trees depended almost entirely on anatomy and fossils. The presence of the same bones, organized in the same way, in such apparently different organisms as bats, cats and whales, provides a strong suggestion that these are more closely related to one another than, say, to birds, sharks or bony fish. This sort of homology was well understood as long ago as the early 19th century, when Étienne Geoffroy St. Hilaire (1830) argued that 'philosophically speaking, there is just one animal',[1] or even earlier, when Denis Diderot (Anonymous 1754) wrote that 'When one considers the animal kingdom, and notices that among the quadrupeds, there is not a single one that does not have the functions and parts – above all the interior parts – entirely similar to those of another quadruped, does not one willingly believe that there was never anything but a single animal prototype of all the animals, of which Nature has only lengthened, shortened, modified, multiplied or obliterated certain organs?' However, anatomy can only take us so far: it can tell us nothing about the relationships between animals, fungi and plants, and very little about those between distantly related organisms within these kingdoms.

The other great non-biochemical source of information about phylogeny has been the fossil record. However, although this has been very effective for tracing the evolution of organisms with abundant fossils, such as horses, it is almost useless for some species that interest us more: hominin fossils are extremely sparse, and chimpanzee and gorilla fossils are even more sparse, the first chimpanzee fossil being reported as recently as 2005 (McBrearty and Jablonski 2005). So, although gross morphology suggested a date around 30 million years ago for the separation of human ancestors from those of the African apes, it has become clear, first from protein data (Sarich and Wilson 1967), and now amply confirmed by genome data, that it was much more recent, between 5 and 8 million years ago.

Compared with fossils, biochemical data are extremely abundant, and become more so every year. Moreover, the biological clocks of different proteins are spread over a huge range: proteins like the fibrinopeptides, the peptides that are discarded when fibrinogen is hydrolysed to fibrin, evolve very fast, and can be used for classifying closely related organisms; globins and cytochromes evolve more slowly, and are useful for more distantly related organisms; histones evolve so slowly that they can be used to study relations between plants and animals. For histones it is probably the case that all allowable structures were found very early in evolution and new ones would be lethal.[2]

The possibility of using protein sequences to study evolutionary relationships was first suggested – almost as a throw-away line – in Francis Crick's review of protein synthesis (Crick 1958), and developed

[1] It would be tempting to assume that by 'animal', Geoffroy meant 'mammal', but in fact he meant it much more generally, recognizing, for example, similarities in anatomy between mammals, fish, birds and even insects and spiders (Stott 2012).

[2] This does not mean that no other histone structures could function, only that there is no possible evolutionary route to them.

more thoroughly by Émile Zuckerkandl and Linus Pauling (Zuckerkandl and Pauling 1962, 1965). They have now been used in practice for this for about half a century, starting with the classification of mammals on the basis of fibrinopeptides (Doolittle and Blombäck 1964), and followed by classification of a much broader range of organisms, from mammals to yeasts, made with cytochrome c (Fitch and Margoliash 1967). However, as early as 1969 Carl Woese recognized, in a letter to Crick quoted by Woese and Goldenfeld (2009), that the molecule of choice would be ribosomal RNA, on account of its presence in large amounts and with the same function in all self-replicating cells, together with the relative ease of extracting it. Later Woese and Fox (1977) used ribosomal RNA to demonstrate the existence of three (not two) primary kingdoms of life, and it is now very widely used for constructing the tree of life, with, of course, vastly more data than were available in 1977.

2.2 *Why do organisms evolve?*

Biochemistry is also useful for understanding a quite different aspect of evolution: why does it happen at all? No organism 'wants' to evolve. On the contrary, everything we know about DNA replication and translation tells us that there are many mechanisms to ensure constancy from one generation to another. Yet, organisms do change, so the constancy is not perfect. Biochemistry tells us that the replication and translation mechanisms depend on base pairing, which depends on thermodynamic interactions between the bases, which are not infinitely strong. So, although AT and CG pairs form stronger interactions than, say, GT or AG pairs, the difference is not enormous (figure 1), and the latter two pairs produce only minor distortion of the DNA helix. Likewise, in decoding on ribosomes, GU pairs (corresponding to GT in DNA) are a common cause of translation infidelity (Demeshkina *et al.* 2013).

There are similar biochemical considerations to explain why mistakes occur in charging amino acids to transfer RNA: threonine and valine, for example, are similar enough in shape and size for threonine to be charged on $tRNA^{Val}$ about once in 300 events. If this and other similar errors remained uncorrected, few protein molecules would be synthesized correctly. The 'double sieve' mechanism of Fersht and Kaethner (1976) allows 'wrong' aminoacyl-tRNA molecules to be hydrolysed with a much higher frequency than 'right' molecules, and brings down the error rate to an acceptable level. Error correction is expensive, however, as it must be paid for by hydrolysis of ATP; so, living systems need to compromise between high accuracy and low cost. The outcome is that genomes cannot remain constant from generation to generation, and so evolution is inevitable.

2.3 *Collagen*

An extreme illustration of the expense involved in overcoming errors is provided by the procollagen cycle in animals, in which defective collagen molecules are immediately hydrolysed. In human muscle, for example, as many as 95% of newly synthesized procollagen molecules are immediately discarded and not converted into collagen for incorporation into muscle (Mays *et al.* 1991). In industrial processes a rejection rate as high as this could only be tolerated for extremely expensive items, and would be unthinkable for a mass-produced item, as collagen (by far the most abundant protein in the human) surely is. This is yet another example of the fact that the living organisms that we see cannot be regarded as the result of careful design, but rather they are the best that an evolutionary process has managed to produce.

Collagen also illustrates a different question of evolution. Collagen-related diseases are not usually life-threatening, but they have a major impact on the quality of life, not only for humans but for other large animals in the wild, such as elephants. They appear to result from the incapacity to synthesize glycine in sufficient quantities to allow a high enough rate of turnover of collagen to meet the requirements for a healthy life (Meléndez-Hevia *et al.* 2009). But to say this raises as many questions as it answers, and illustrates why the teaching of evolutionary biology and of biochemistry must feed on one another. The fundamental biochemical problem results from the fact that biosynthesis of glycine from serine is stoichiometrically constrained by the production of C_1 units for transfer to folic acid (Meléndez-Hevia and de Paz-Lugo 2008). Surplus glycine molecules can be transformed into C_1 units, but no known metabolic reaction can transform surplus C_1 units into glycine. Evolution provides the key to understanding this. Collagen appeared with the Metazoa (animals) – all animals require collagen, and collagen is not found in any other

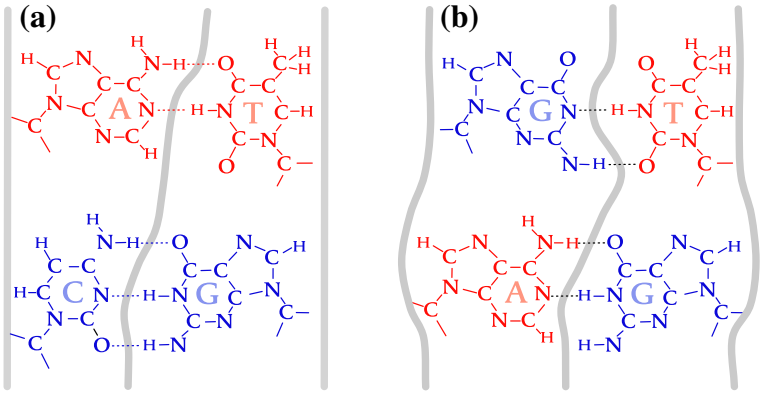

Figure 1. Base-pairing. (**a**) The 'standard' pairing scheme not only allows two hydrogen bonds in A = T and three in C≡G, but it also allows these pairs to be formed without distorting the overall structure. (**b**) Other pairing, such as G=T or A=G, is also possible, however, with two hydrogen bonds in each pair, without producing enormous distortion of the overall structure.

organisms – but the first animals were small and aquatic, and, in accordance with allometric principles (Galilei 1638; Thompson 1945), required much less collagen in proportion to their size than large terrestrial animals. The mechanisms for regulating production of C_1 units and glycine therefore evolved in organisms that had no difficulty for producing enough collagen for their needs: the problem only arose with large terrestrial animals, which have inherited a regulatory system that is very poorly adapted to their needs. Moreover, animals in general show very little evidence of a capacity to evolve new metabolic pathways, and so by the time the large animals appeared they had almost no possibility of correcting an ancient problem.

2.4 *Dominance and recessivity*

The mechanistic basis of the genetic property of recessivity in diploid organisms remained a mystery until it was explained by Kacser and Burns (1981). Many years earlier Ronald Fisher (1934) had suggested an explanation in terms of 'modifier genes', whereas Sewall Wright (1934) preferred an explanation in terms of biochemistry. Fisher's theory is obscure and difficult to understand, but it was generally accepted by geneticists until Kacser and Burns (1981) showed that an explanation much closer to Wright's must be correct (Cornish-Bowden and Nanjundiah 2006). In other words, knowledge of biochemical principles is needed to understand a property that has been fundamental in genetics since the time of Gregor Mendel.

2.5 *Reconstructing ancient proteins and experimental evolution*

At present it is not possible to recreate extinct organisms, although there are hopes that it will be possible to recreate the mammoth, which became extinct relatively recently. However, biochemical structures are not so constrained, and various ancestral proteins with the structures deduced by phylogenetic analysis have been synthesized and found to have the expected properties. A recent example is provided by resuscitation of an ancestral corticosteroid receptor (Kohn *et al.* 2012) that has shed light on understanding drug selectivity for modern steroid receptors.

Another powerful approach to evolutionary dynamics and mechanisms is experimental evolution in the laboratory, using fast-replicating microorganisms and genomic analysis techniques (Buckling *et al.* 2009). Studies of experimental evolution were made even during Darwin's lifetime (Dallinger 1878), but did not become widespread until much more recently. Probably the longest such project is the one running in Richard Lenski's laboratory with *E. coli* cultures for the past 25 years, during which the bacterial populations have gone through more than 55000 generations. The possibility to keep frozen samples of the culture allows the replay of the experiments and the study of the role of contingency and genomic changes in the evolutionary process. A remarkable recent result on the emergence of a novel trait is the identification after 31000 generations of a clone of *E. coli* able to use citrate as a carbon

source under aerobic conditions (Blount *et al.* 2012). It is worth remembering that the inability to grow aerobically on citrate is one of the criteria used by bacteriologists to define *E. coli* as a species, and thus it is legitimate to ask whether Lenski and coworkers have witnessed the emergence of a true new bacterial species forcing the biochemical machinery to adapt to a new carbon source. Furthermore, experimental evolution is not restricted to microorganisms: it has obvious applications in biotechnology and opens new prospects in science education, such as allowing observation of evolution in real time (Kawecki *et al.* 2012).

Ernst Mayr (1961) suggested that biologists ask two fundamental kinds of questions. Functional biologists enquire about proximate causes or mechanisms: how does it work? This aspect includes most of biochemistry and molecular biology. On the other hand, evolutionary biologists ask about ultimate or historical causes: why does it work the way it does? Dean and Thornton (2007) have proposed that resurrecting extinct proteins and experimental evolution have allowed the merging of biochemistry and evolution in the so-called functional synthesis, permitting the empirical evaluation of the molecular mechanisms by which genomic changes produce new phenotypes.

2.6 *Promiscuity and enzyme evolvability*

High substrate specificity is a universal property of enzymes that allows the performance of many simultaneous reactions in metabolism. Fidelity is crucial in many cases and costly proofreading mechanisms have emerged (e.g. replication of DNA, synthesis of aminoacyl-tRNAs). Although evolution has selected against noise and stochasticity, there are also many examples of lack of specificity, multifunctionality and promiscuity (Tawfik 2010). The inherent flexibility of proteins and the impossibility of absolute specificity allow the opportunism of evolutionary tinkering (Jacob 1977).

2.7 *Coevolution of metabolism and environment*

Cells are open chemical systems. Thus, the use of chemicals present in the medium and the release of products by metabolism considerably influence the chemical composition of the environment. Since its very inception, life has chemically influenced the environment, and at the same time, the changing chemistry of the atmosphere or the oceans has shaped the opportunities of metabolism. One of the most studied cases of chemical interdependence between geochemistry and biochemistry has been the presence of molecular oxygen O_2 in the atmosphere (Peretó 2011, and references therein). The primitive atmosphere was almost completely devoid of O_2. The appearance of cyanobacterial ancestors with the ability to split water to obtain electrons for photosynthesis initiated the massive production of O_2 as a byproduct with recognizable geological landmarks (e.g. the banded-iron formations), dramatic atmospheric changes (e.g. the marginalization of methane and the accumulation of O_2) and ample metabolic opportunities with the emergence of O_2-utilizing enzymes (e.g. cytochrome oxidase in respiratory chains, oxidases in steroid biosynthesis, etc.), many of them derived from anaerobic orthologues (Raymond and Blankenship 2004; Ducluzeau *et al.* 2009). The use of O_2 as a substrate had an enormous impact on metabolism and the evolution of eukaryotic life (Raymond and Segrè 2006).

3. What evolutionary biology can tell us about biochemistry

More examples could be given to illustrate the contributions biochemistry is making to evolutionary biology, but now we shall concentrate on the converse point, that knowledge of evolutionary biology can shed light on points of biochemistry that would otherwise be puzzling. This is not true, of course, for the whole of biochemistry: the kinetics and thermodynamics of enzyme-catalysed reactions can be taught without any reference to evolution, and the same is largely true of enzyme mechanisms: in the words of Jeremy Knowles (1991), enzyme catalysis is ‘not different, just better’; evolution has brought enzyme catalysis to a high degree of perfection, but without violating general chemical principles. It is partly for this reason that 100 years after the fundamental concepts of enzyme kinetics were developed by Michaelis and Menten (1913), they have required some extensions, but almost no revision (Cornish-Bowden 2013; Deichmann *et al.* 2014). Likewise, 50 years after the concept of allosteric regulation was introduced (Monod *et al.* 1963), it remains valid as a mechanistic idea (Cárdenas 2013), although discussion of its

physiological role in metabolic regulation does require some consideration of evolution (Hofmeyr and Cornish-Bowden 2000).

3.1 *Structure of coenzymes*

Enzyme catalysis depends in many cases on non-protein coenzymes with widely differing structures, such as metal ions, iron-sulphur centres, quinones and so on. Some of them, such as ATP, NAD, FAD, coenzyme A and *S*-adenosylmethionine, include what appear to be fragments of RNA, and Benner (2009) has strongly argued that these should be seen in the light of history, as remnants of a time when life on Earth used RNA as the only encoded biopolymer. For example, the structure of NAD is illustrated in figure 2. At first sight it seems absurdly complicated, far more than it needs to be to fulfil its chemical function, and anyone seeing it for the first time is bound to wonder how one could be expected to remember it. That is because it *is* absurdly complicated. The nicotinamide portion shown in red is essential for the chemical function, but it accounts for fewer than a quarter of the total number of atoms. The rest plays no part in the chemical mechanisms of enzymes using NAD as coenzyme, and although it has a role in specific binding, this function could surely have been fulfilled by a simpler structure. Moreover, the unused part of the molecule is the part mainly responsible for the lability of NAD, thereby adding greatly to the difficulties for developing biomimetic catalysts. Put in these terms, it seems impossible to explain why such a complicated structure was needed, but that is to think with the mind of an engineer or designer, and not with the mind of a tinker, who works by putting together bits and pieces that are available and more or less suitable for some purpose (Jacob 1977). Seen in this light it becomes clear why NAD resembles a degraded fragment of RNA and not a molecule designed for a specific purpose.

This knowledge contributes to a better understanding of biochemistry in two ways. First, the evolutionary perspective allows one to see that the structure is not totally arbitrary but reflects the history, in this case as a relic of an ancient form of RNA or even of a precursor of RNA; second, recognition of the RNA-like structure allows an appreciation that the structure of NAD has similarities with other structures in biochemistry.

This is by no means a unique example, and many other molecules in biochemistry, such as folic acid, have features that make little sense in terms of their functions, but make perfect sense in terms of evolutionary history. In this context we may ask whether the tripeptide glutathione (figure 3), which is not produced on the ribosome but by specialized enzymes, is in some sense a relic of an ancient protein. Clearly a *coded* system resembling the ribosomal production of protein offers the only way of producing the wide array of different proteins that we know today, but that does not exclude the possibility that less versatile systems were tried and later discarded after the origin of life.

Cyclic peptides are also produced by specialized enzymes, but in this case the enzymes are enormous and catalyse reactions of great complexity – cyclosporin synthetase, for example, catalyses a reaction with 29 substrates and at least 40 steps (Lawen and Zocher 1990) – making it implausible that cyclosporin could be a relic of an ancient molecule. This synthesis differs from processes in fatty acid metabolism that consist of a small set of reactions repeated many times, as each step of cyclosporin synthesis is essentially a different reaction.

3.2 *Why is the structure of DNA so badly suited to its function?*

The structure of DNA is commonly represented, not only in popular articles, but also often in textbooks, in terms of block diagrams that look tidy but conceal much of the chemical reality. We have already mentioned this in section 2.2, but here we shall consider it from a somewhat different point of view. The three hydrogen bonds in a CG pair allow a fairly strong and specific interaction, but the AT pair, with only two hydrogen bonds, is far less satisfactory. This arrangement is certainly not the only possible one, and Yang *et al.* (2011) have shown that one could have a different base-pairing system – with six instead of four bases – that would work better. Again, the explanation is not that the existing structure of DNA is the best possible, but simply that it is a ‘frozen accident’, a structure that proved to work adequately in the early history of life, but then became impossible to improve because so many proteins depended on a fixed system. As suggested by Szathmáry (2003), the current four-letter genetic alphabet is a legacy of a primitive RNA world and does not represent an optimum under the evolutionary constraints and repair mechanisms of our DNA world.

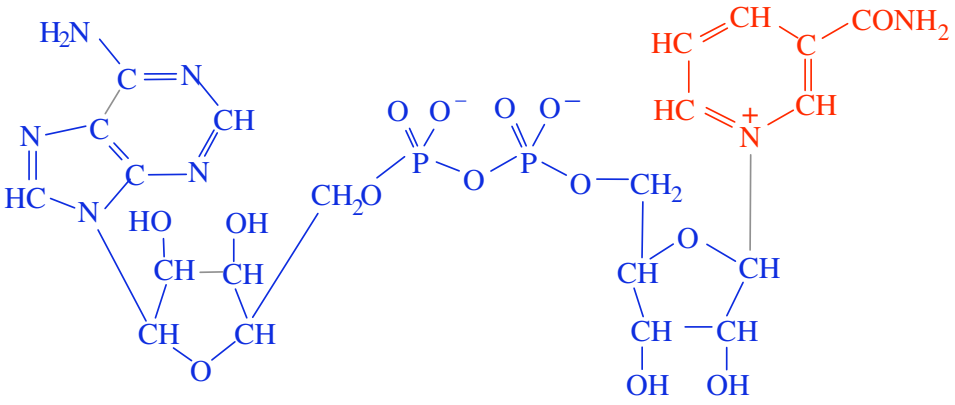

Figure 2. The structure of the coenzyme NAD. Only the 16 atoms shown in red appear necessary for the chemical function of the coenzyme. They account for about 23% of all the atoms in the molecule. The others, shown in blue, have roles to play in the specific binding of the molecule to enzyme molecules, but it is hard to believe that this binding function could not have been fulfilled with a simpler and less labile structure.

The 'universal genetic code' is now known not to be universal, as minor differences are found in mitochondria and some other systems, so it cannot be strictly fixed. Clearly a few changes must have occurred, although the initial instinct was to think this would be impossible, so how can it be explained? There are numerous examples in the standard code of multiple codons for the same amino acids, and there are wide variations in codon usage between organisms, with some codons used very little by some organisms. As an extreme example, the bacterium *Mycoplasma capricolum* does not use the codon CGG to code for arginine, or anything else (Oba *et al.* 1991). This means that assigning this codon to some other amino acid (whether one of the standard set or another) would have no effect on the synthesis of any arginine-containing protein. Such reassignment after falling into disuse is most likely to have occurred in very small genomes, which is why departures from the standard code are found in mitochondria and small bacteria.

3.3 *Configuration of amino acids*

All of the chiral amino acids found in proteins have the L-configuration. This makes sense if they arose progressively in early evolution, each 'new' amino acid being modified from one that existed previously. However, on a purely functional basis it is less easy to understand, as it would imply that no D-amino acid could fulfil its function better than the corresponding L-amino acid – not impossible, but not easy to believe. The evidence of an enantiomeric excess (L/D>1) observed in meteoritic amino acids suggests that prebiotic environments could be enriched in L-amino acids and that some chemical determinism played a role in the emergence of biochemical homochirality (Pizzarello and Lahav 2010; Pizzarello and Shock 2010; Blackmond 2011). Nonetheless, all chemical or physical mechanisms that have been suggested for generating chirality in a symmetric environment are extremely feeble (Guijarro and Yus 2009): for

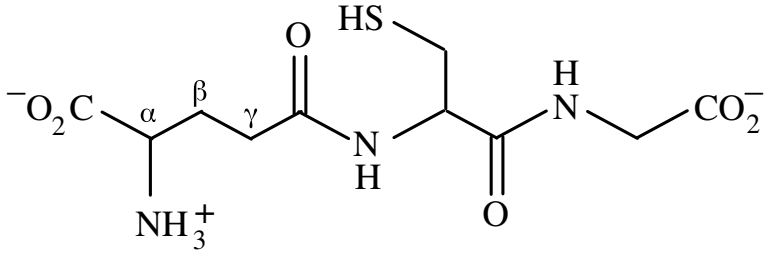

Figure 3. Glutathione (γ-glutamylcysteinylglycine). Can this be regarded as a fragment of an ancient protein?

example, the weak nuclear force (which is asymmetric) could account for an enantiomeric excess at 25°C of one part in 10^{17} and the circular polarization of the sun's light at dawn and dusk is equally unpromising, especially as it operates in opposite directions at the two times.

3.4 *The unity of biochemistry*

For a biologist familiar with the vast array of anatomies, habitats, behaviour patterns and other characteristics found in the living world, it appears very surprising that metabolism is, to a first approximation, the same for all organisms, with the same reactions: 'from the elephant to the butyric acid bacterium – it is all the same' (Kluyver and Donker 1925).[3] We now know that these reactions are catalysed by enzymes that are clearly homologous, even between species as distantly related as mammals, yeast and bacteria. The high degree of similarity between the sequences of hexokinase A from the rat and glucokinase from *Saccharomyces cerevisiae* seen along the two diagonals in figure 4 indicates beyond any reasonable doubt that the two proteins are homologous, i.e. they are derived from the same ancestral sequence, despite the vast gulf of time that separates yeasts from mammals. The fact that the horizontal axis in figure 4 is twice as long as the vertical axis reflects the fact that hexokinase A is a 100 kDa protein, whereas yeast glucokinase is a 50 kDa protein. The presence of two obvious diagonals (highlighted in red) is more interesting; and indicates that although hexokinase A is a monomeric protein, it is 'dimer-like', in the sense that the two halves of the molecule are very similar to one another. All this is easy to explain from an evolutionary perspective (Cárdenas *et al.* 1998), but difficult otherwise.

An objection might be raised that the similarity of sequences could derive not from homology but from an extremely restricted set of structures capable of catalysing the reaction. This may be true for the small region around the catalytic centre, but it can be ruled out as a general explanation by the results of shuffling experiments (Kolkman and Stemmer 2001). In such experiments fragments of the genes coding for the same protein in one or more species are reassembled in random orders, thereby generating a library of related sequences. Many of these have no catalytic activity, of course, as they will have a high probability of disrupting the sequence around the catalytic residues, but enough do have activity to make it clear that the number of sequences capable of fulfilling any particular function is very large, enough to rule out the hypothesis that the similarity between proteins from widely separated species is due to anything but descent from a common ancestor.

Another way to explain similarity without homology would be to invoke horizontal gene transfer, or transfer of a gene from one organism to a very distantly related one. This probably explains the similarity of sequence of the superoxide dismutases of various fishes, and a light-emitting bacterium, *Photobacterium leiognathi*, found in the light organ of a luminous fish, *Leiognathus splendens* (Cornish-Bowden 1985). However, such examples of transfer between very distantly related species are rare enough not to seriously undermine the usual type of explanation. In any case, it still implies homology: even if the fish and bacteria do not have a *recent* common ancestor, the similarity of the proteins still suggests a remote one.

None of this implies that there are no differences in biochemistry between different organisms: clearly there are, but it is evident that the general outline of the metabolic map is essentially the same and that there are some truly universal (and ancient) reactions catalysed by homologous enzymes. Homology can be detected between eukaryotes, bacteria and archaea, and is allowing current research into the characteristics of the last universal common ancestor ('LUCA').

The concept of LUCA is sometimes misunderstood: even people who research into its characteristics often talk as if it existed shortly after the origin of life, or as if it was in some way a special individual or species. Both of these ideas are false. LUCA clearly had an evolved metabolism and information-processing system, and must have been far more complex than the protoorganisms at the origin of life. The fact that it was not 'special' can be seen more easily in terms of the last common mammalian ancestor, because this concerns species that are familiar to everyone. As illustrated in figure 5, present knowledge indicates that this ancestor existed around 180 million years BP. The date of the common ancestor of the monotremes and all the others. However, the monotremes constitute a group with very few living descendants, in terms both of numbers of species and of numbers of individuals. It is by no means unlikely that they will all become extinct, even within the lifetimes of

[3] As discussed by Friedmann (2004), this aphorism is better known in various versions attributed to Monod, such as 'Anything that is true of *Escherichia coli* must be true of elephants, only more so'.

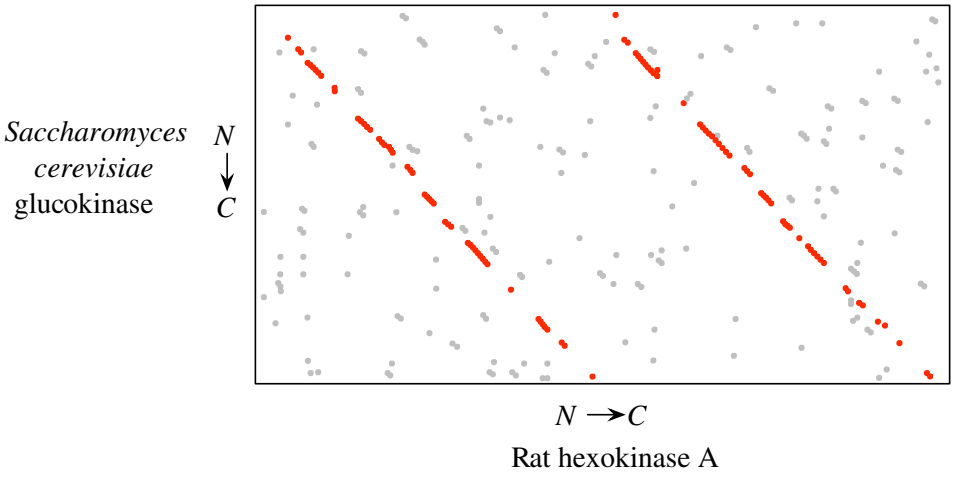

Figure 4. Comparison of the amino acid sequences of rat hexokinase A and yeast glucokinase (Cárdenas *et al.*, 1998). Each dot represents a locus in which the residues are identical in the two sequences, but is only shown if it occurs in a range of nine residues in which there are four or more such identities. Such filtering is necessary to avoid filling up the whole diagram with random identities. The colours used, red for the diagonals, interpreted as signal, and grey for the other points, interpreted as noise, are an *interpretation* of the data, not the data as such.

people alive today, and if that happens the last common mammalian ancestor will move to 140 million years BP, 40 million years closer to the present, even though the impact of such an extinction on the rest of the mammalian species will be almost negligible. We see, therefore, that far from being 'special' the definition of the last common mammalian ancestor depends arbitrarily on the survival of a small number of taxa. It is less likely, but not impossible, that all the marsupials will disappear, in which case the last common mammalian ancestor will have lived only 105 million years ago.

Similar considerations apply to LUCA. Indeed, it is not impossible that future research will reveal a few surviving species of microorganisms that separated from the rest of the tree of life well before the Eubacteria and Archaea separated from one another. If this happens, the date for LUCA will suddenly move backwards in time, with essentially no impact on our understanding of the species that we know at present.

3.5 *The structure of glycogen*

Discussion of the structure of glycogen, especially in an educational context, is complicated by the fact that the structure proposed in 1940 by Meyer and Bernfeld (1940) and still shown today in nearly

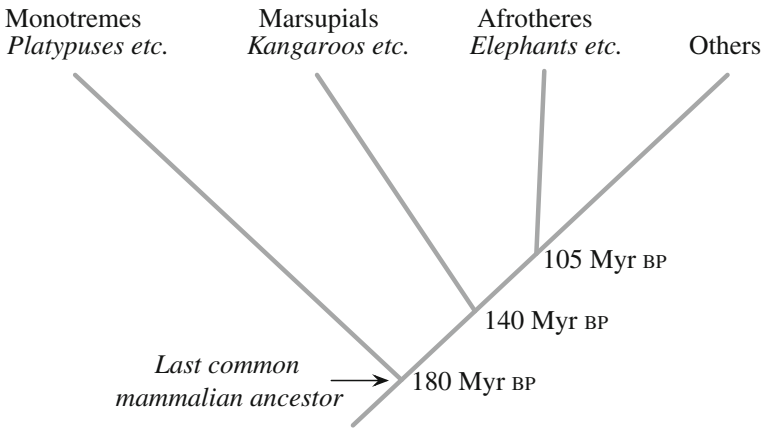

Figure 5. The last common mammalian ancestor, estimated to have lived about 180 million years ago.

all textbooks of biochemistry is wrong, and has been known to be wrong since Gunja-Smith *et al.* (1970) established the correct structure more than 40 years ago, apart from the presence of the protein glycogenin, which came later (Lomako *et al.* 1988). The essential points of the true structure (figure 6) are as follows:

1. Apart from a single C chain attached to the glycogenin molecule in the interior of the structure, glycogen consists of equal numbers of A chains, which are unbranched, and B chains, which are branched.
2. A complete molecule contains about 12 layers like skins on an onion, and all the A chains reach the surface.
3. Each chain is a polymer of 11–15 glucose units.
4. Each B chain contains two branches.

Notice that the real structure, with all of the chains of about the same length, and exactly two branches in each B chain – not one, not three, but two – is tidier and easier to visualize than the wrong structure, making it hard to understand why the latter has been so persistent in textbooks. The same characteristics are found in the glycogen molecules of many different organisms – not only mammals, but also fish, birds, invertebrates, bacteria, protozoans and fungi (Meléndez *et al.* 1997). All glycogens,

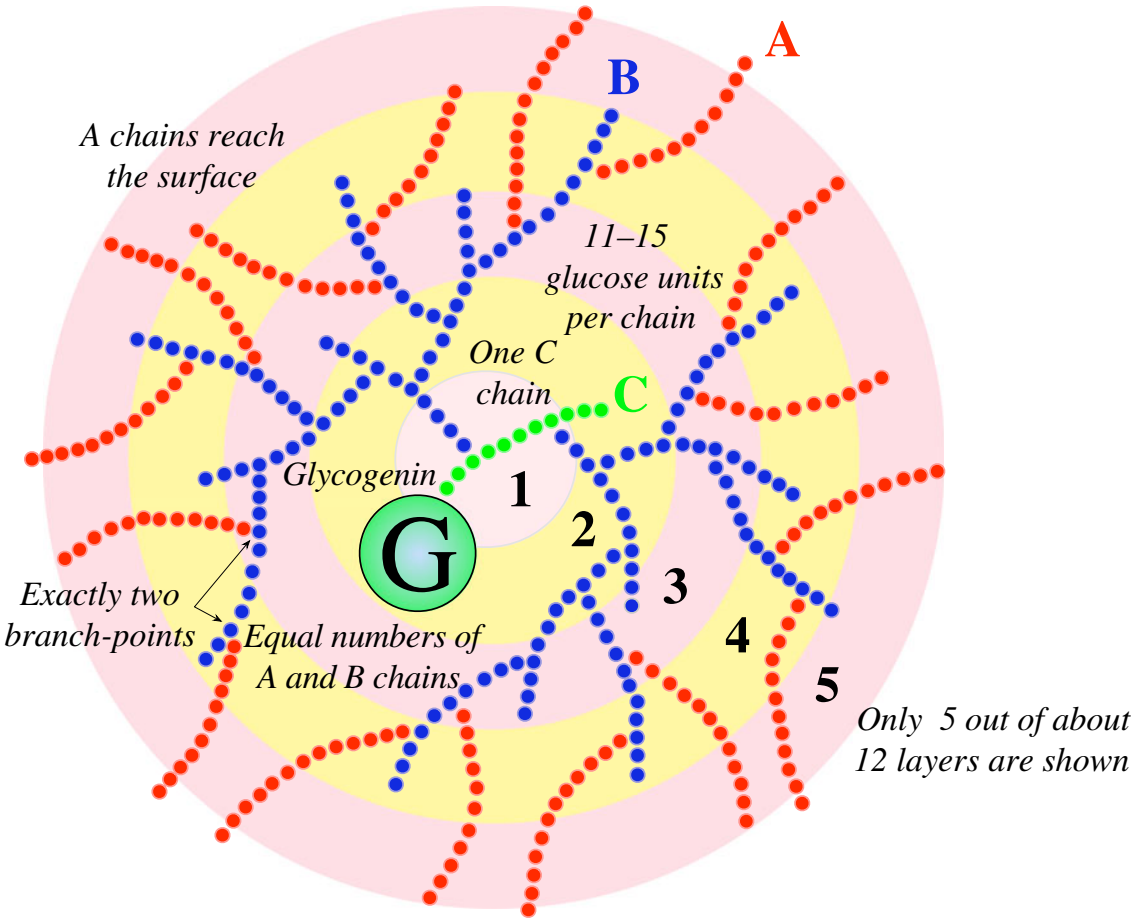

Figure 6. The structure of glycogen as it exists across the animal kingdom, in yeast and in bacteria. The A chains of glucose residues are shown in red, the B chains in blue, and the C chain in green. The protein glycogenin in the core of the polymer is labelled G. Notice that the layers become more tightly packed as one moves away from the centre: this imposes a limit of about 12 layers. However, to avoid making the diagram too complicated to understand, a structure of only five layers is shown.

whether from different tissues in mammals, or from bacteria or yeast, have the same structure, with very similar chain lengths, from 11 to 16.

This structure allows the equivalent of about 50000 glucose molecules to be packed into a single molecule in such a way that half of them are immediately available for retrieval when they are needed. If the numbers were different (for example, more or less than two branch-points in each B chain, or substantially longer or shorter chains), the space would be used much less efficiently, so the number of glucose molecules that could be stored, and the proportion of them available for immediate release, would be smaller, and Meléndez *et al.* (1997) showed that it is thus an *optimal* structure that has been obtained through natural selection.

It may not be immediately obvious why it is necessary to store such a large number of glucose molecules in a single glycogen molecule. If the glucose were simply free in solution, it would imply a maximum concentration of about 0.4 M, or about 80 times the actual blood-glucose concentration, and that in turn would imply large and intolerable variations in osmotic pressure when the concentration responded to changes in demand. The second point is that a structure that places half of the glucose molecules on the surface allows extremely rapid release of the stored glucose when the blood-glucose concentration falls.

3.6 *Natural selection and the neutral theory*

Many biologists, and even some biochemists, talk as if natural selection were the only important mechanism driving evolution. Although this is satisfactory to some extent for discussing evolutionary changes that can be regarded as clear examples of adaptations, i.e. changes that have some identifiable function, it is not at all believable for explaining molecular evolution, i.e. the accumulated changes in protein structure, as seen, for example, in figure 4: rats did not of course evolve from yeast, nor vice versa, but the similarity between the two enzymes compared indicates that they did evolve from a common ancestor, and the present day sequences reflect the changes that have occurred in the two lineages since they separated. Many arguments indicate that it would be absurd to regard every one, or even a significant number, of these changes as adaptations. Instead, they indicate that many changes can occur with no change in function, and the *neutral theory* (Kimura 1983) is based on the idea that most mutations that are fixed in the population are *neutral* – they have no effect on function one way or the other. In its modified form known as the *nearly neutral theory*, proposed by Ohta (1973) and extended by Ohta and Gillespie (1996), it is now widely accepted.

4. Some pitfalls

Although our view is that understanding of evolutionary biology and biochemistry depend on one another, with the implication that any modern course of biochemistry ought to be taught from an evolutionary perspective, one must be careful not to take this too far, and to suppose that every property in biochemistry is a consequence of evolution. In this section we shall consider some properties that do not depend in any way on natural selection.

4.1 *Kinetics and thermodynamics*

The thermodynamic principles that govern the interactions of biological molecules, whether base pairing in nucleic acids, binding of substrates to enzymes or anything else, are completely independent of any evolutionary considerations, and would be exactly the same if the materials of life were quite different from those that we know.

The same argument applies, but less strongly, to kinetic behaviour. Kinetic parameters can respond to evolutionary pressure (Cornish-Bowden 1976), and so the high catalytic activity and specificity of enzymes must be the result of natural selection. However, the elementary principles of enzyme kinetics are independent of evolution. Even if enzymes were made of RNA, or some other material, rather than protein, we should expect them to show the same properties of substrate saturation as we see with protein enzymes; in other words, we should still expect the Michaelis–Menten equation to serve as a starting point for

teaching the kinetic properties. Indeed, at the time this equation was established (Michaelis and Menten 1913) almost nothing was known about the chemical nature of enzymes.

4.2 *Control and regulation*

Although the regulatory properties of enzymes such as phosphofructokinase can be attributed to natural selection, many of the kinetic properties of metabolic pathways consisting of several enzymes are the result of mathematical necessity and could not be changed by selection. In particular, there is now no doubt that flux control in a system is shared by all of the enzymes in the system, although not usually shared equally (Kacser *et al.* 1995; Fell 1997). This means (with some complications due to the fact that shares can be negative) that the share of control held by any enzyme must be small, and that nearly all enzymes appear to be 'in excess'. The real role of the classical allosteric and cooperative mechanisms is to maintain homeostasis, i.e. to regulate metabolite concentrations, not fluxes (Cornish-Bowden and Cárdenas 2001). So, although some authors have tried to find some adaptive value in the existence of most enzymes 'in excess', there is none to be found.

4.3 *Enthalpy–entropy compensation*

Another supposed phenomenon that is sometimes thought to have evolutionary significance comes from the fact that when the thermodynamic parameters $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$ are estimated from Arrhenius plots and plotted against one another, they typically show a virtually perfect linear correlation. For example, a whole book on biochemical adaptation was written around this idea (Hochachka and Somero 1984), and more recently, Gutfreund (1995) presented an example of such an 'adaptation', with a virtually perfect straight-line correlation between the Arrhenius parameters of ATPases from fishes from environments as different as the Antarctic Ocean and lakes of tropical Africa. However, this is simply a statistical artifact derived from the fact that the typical temperature range over which enzyme activity measurements can be made (say 5–40°C in the most favourable circumstances) is too small to yield meaningful estimates of both parameters (Cornish-Bowden 2002). A compensation plot is simply a plot of a measurement against itself after transforming it in two ways, and has no evolutionary significance whatsoever. The 'compensation temperatures' sometimes reported in these kinds of experiments depend only on the temperatures chosen for the measurements and are completely independent of the observations (Cornish-Bowden 2012).

5. Discussion

The central point that we have argued for in this article is the intimate relationship between biochemistry and evolution. Neither can be well understood without the other, and it follows, therefore, that neither should be taught without the other. Courses in evolutionary biology ought to include significant emphasis on the principles of biochemistry, and courses on biochemistry ought to stress the role evolution has played in making biochemistry what it is. Many aspects of protein evolution – many more than we have discussed in this article – are covered in the recent book by Ureta (2011).

We do not claim to have covered the whole field of biochemistry and evolution in this article, nor that we have given the only possible answers to all of the questions we have raised. On the contrary, we believe that there is room for controversy, and we hope, therefore, that we can stimulate discussion.

Acknowledgements

ACB and MLC acknowledge the support of the Centre National de la Recherche Scientifique. JP acknowledges the intellectual motivation of many students during more than 20 years of teaching evolutionary biochemistry and origins of life at the University of Valencia, as well as the financial support for his research on symbiosis to the Spanish Mineco (grant BFU2012-39816-C02-01). The impetus for writing this article came from three sessions at the IUBMB-FEBS-SEBBM Congress in Seville in 2012, and we thank the organizers for facilitating our participation in these.

References

Anonymous 1754 Pensées sur l'interprétation de la Nature

Benner SA 2009 *Life, the Universe and the scientific method* (Gainesville, FL: The Ffame Press)

Berg JM, Tymoczko JL and Stryer L 2012 *Biochemistry* (San Francisco: W. H. Freeman)

Blackmond DG 2011 The origin of biological homochirality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **36** 2878–2884

Blount ZD, Barrick JE, Davidson CJ and Lenski RE 2012 Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489** 513–518

Buckling A, Maclean RC, Brockhurst M A and Colegrave N 2009 The Beagle in a bottle. *Nature* **457** 824–829

Cárdenas ML 2013 Michaelis and Menten and the long road to the discovery of cooperativity. *FEBS Lett.* **587** 2767–2771

Cárdenas ML, Cornish-Bowden A and Ureta T 1998 Evolution and regulatory role of the hexokinases. *Biochim. Biophys. Acta* **1401** 242–264

Cornish-Bowden A 1976 The effect of natural selection on enzymic catalysis. *J. Mol. Biol.* **101** 1–9

Cornish-Bowden A 1985 The amino-acid sequences of the copper-zinc superoxide dismutases from swordfish and *Photobacter leiognathi* confirm the predictions made from the compositions. *Eur. J. Biochem.* **151** 333–335

Cornish-Bowden A 2002 Enthalpy–entropy compensation: a phantom phenomenon. *J. Biosci.* **27** 121–126

Cornish-Bowden A 2012 Enthalpy–entropy compensation as deduced from measurements of temperature dependence; in *Protein-ligand interactions* (ed) H Gohlke (Weinheim: Wiley–Blackwell) pp 33–43

Cornish-Bowden A 2013 The origins of enzyme kinetics. *FEBS Lett.* **587** 2725–2730

Cornish-Bowden A and Cárdenas ML 2001 Information transfer in metabolic pathways: effects of irreversible steps in computer models. *Eur. J. Biochem.* **268** 6616–6624

Cornish-Bowden A and Nanjundiah V 2006 The basis of dominance; in *The biology of genetic dominance* (ed) RA Veitia (Georgetown, Texas: Landes Bioscience) pp 1–16

Crick FHC 1958 On protein synthesis. *Symp. Soc. Exp. Biol.* **12** 138–163

Daeschler EB, Shubin NH and Jenkins FA Jr 2006 Devonian tetrapod-like fish and the evolution of the tetrapod body plan. *Nature* **440** 764–771

Dallinger DH 1878 On the life-history of a minute septic organism: with an account of experiments made to determine its thermal death point. *Proc. R. Soc. Lond.* **27** 332–350

Dean AM and Thornton JW 2007 Mechanistic approaches to the study of evolution: the functional synthesis. *Nat. Rev. Genet.* **8** 675–688

Deichmann U, Schuster S, Mazat, J-P and Cornish-Bowden A 2014 Commemorating the 1913 Michaelis–Menten paper *Die Kinetik der Invertinwirkung*: three perspectives. *FEBS J.* **281** 435–463

Demeshkina N, Jenner L, Westhof E, Yusupov M and Yusupova G 2013 New structural insights into the decoding mechanism: translation infidelity via a G · U pair with Watson–Crick geometry. *FEBS Lett.* **587** 1848–1857

Dobzhansky T 1973 Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* **35** 125–129

Doolittle RF and Blombäck B 1964 Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. *Nature* **202** 147–152

Ducluzeau AL, van Lis R, Duval S, Schoepp-Cothenet B, Russell MJ and Nitschke W 2009 Was nitric oxide the first deep electron sink? *Trends Biochem. Sci.* **34** 9–15

Fell D 1997 Understanding the control of metabolism (London: Portland Press)

Fersht AR and Kaethner MM 1976 Enzyme hyper-specificity – rejection of threonine by valyl-transfer-RNA synthetase by mis-acylation and hydrolytic editing. *Biochemistry* **15** 3342–3346

Fisher RA 1934 The possible modification of the response of the wild type to recurrent mutations. *Am. Nat.* **62** 115–126

Fitch WM and Margoliash E 1967 Construction of phylogenetic trees. *Science* **155** 279–284

Friedmann HC 2004 From '*butyribacterium*' to '*E. coli*' – An essay on unity in biochemistry. *Perspect. Biol. Med.* **47** 47–66

Galilei G 1638 Discorsi e dimostrazioni matematiche intorno a due nuove scienze attenenti alla mecanica e i movimenti locali (Leyden: Elzevir)

Geoffroy St Hilaire É 1830 Principes de philosophie zoologique: discutés en mars 1830 au sein de l'Académie royale des sciences (Paris: Pichon et Didier)

Guijarro A and Yus M 2009 *The origin of chirality in the molecules of life* (Cambridge: RSC Publishing)

Gunja-Smith Z, Marshall JJ, Mercier C, Smith EE and Whelan WJ 1970 Revision of the Meyer–Bernfeld model of glycogen and amylopectin. *FEBS Lett.* **12** 101–104

Gutfreund H 1995 *Kinetics for the Life Sciences* (Cambridge: Cambridge University Press)

Hochachka PW and Somero GN 1984 *Biochemical adaptation* (Princeton, NJ: Princeton University Press)

Hofmeyr JHS and Cornish-Bowden A 2000 Regulating the cellular economy of supply and demand. *FEBS Lett.* **476** 47–51

Jacob F 1977 Evolution and tinkering. *Science* **196** 1161–1166
Kacser H and Burns JA 1981 The molecular basis of dominance. *Genetics* **97** 639–666
Kacser H, Burns JA and Fell DA 1995 The control of flux. *Biochem. Soc. Trans.* **23** 341–366
Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I and Whitlock MC 2012 Experimental evolution. *Trends Ecol. Evol.* **27** 547–560
Kimura M 1983 *The neutral theory of molecular evolution* (Cambridge: Cambridge University Press)
Kluyver AJ and Donker HJL 1925 The unity in the chemistry of the fermentative sugar dissimilation processes of microbes. *Proc. K. Ned. Akad. Wet.* **28** 297–313
Knowles JR 1991 Enzyme catalysis – not different, just better. *Nature* **350** 121–124
Kohn JA, Deshpande K and Ortlund EA 2012 Deciphering modern glucocorticoid cross-pharmacology using ancestral corticosteroid receptors. *J. Biol. Chem.* **287** 16267–16275
Kolkman JA and Stemmer WPC 2001 Directed evolution of proteins by exon shuffling. *Nat. Biotechnol.* **19** 423–428
Lawen A and Zocher R 1990 Cyclosporin synthetase: the most complex peptide synthesizing multienzyme polypeptide so far described. *J. Biol. Chem.* **265** 11355–11360
Lewin R 1987 *Bones of contention: Controversies in the search for human origins* (New York: Simon and Schuster)
Lomako J, Lomako WM and Whelan WJ 1988 A self-glucosylating protein is the primer for rabbit muscle glycogen biosynthesis. *FASEB J.* **2** 3097–3103
McBrearty S and Jablonski NG 2005 First fossil chimpanzee. *Nature* **437** 105–108
Mayr E 1961 Cause and effect in biology. *Science* **134** 1501–1506
Mays PK, McAnulty RJ, Campa JS and Laurent GJ 1991 Age-related changes in collagen synthesis and degradation in rat tissues. Importance of degradation of newly synthesized collagen in regulating collagen production. *Biochem. J.* **276** 307–313
Meléndez R, Meléndez-Hevia E and Cascante M 1997 How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building *J. Mol. Evol.* **45** 446–455
Meléndez-Hevia E and de Paz-Lugo P 2008 Branch-point stoichiometry can generate weak links in metabolism: the case of glycine biosynthesis. *J. Biosci.* **33** 771–780
Meléndez-Hevia E, de Paz-Lugo P, Cornish-Bowden A and Cárdenas ML 2009 A weak link in metabolism: the metabolic capacity for glycine biosynthesis does not satisfy the need for collagen synthesis. *J. Biosci.* **34** 853–872
Meyer KH and Bernfeld P 1940 Research on starch. V Amylopectin *Helv. Chim. Acta* **23** 875–885
Michaelis L and Menten ML 1913 Kinetik der Invertinwirkung. *Biochem. Z.* **49** 333–369
Monod J, Changeux JP and Jacob F 1963 Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6** 306–329
Oba T, Andachi Y, Muto A and Osawa S 1991 CGG – an unassigned or nonsense codon in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. USA* **88** 921–925
Ohta T 1973 Slightly deleterious mutant substitutions in evolution. *Nature* **246** 96–98
Ohta T and Gillespie JH 1996 Development of neutral and nearly neutral theories. *Theor. Popul. Biol.* **49** 128–142
Penny D, Foulds LR and Hendy MD 1982 Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* **297** 197–200
Peretó J 2011 The origin and evolution of metabolisms; in *Origins and evolution of life: an astrobiological perspective* (ed.) M Gargaud, P López-García and H Martin (Cambridge: Cambridge University Press) pp 270–287
Pizzarello S and Lahav M 2010 On the emergence of biochemical homochirality: an elusive beginning. *Orig. Life Evol. Biosph.* **40** 1–2
Pizzarello S and Shock E 2010 The organic composition of carbonaceous meteorites: the evolutionary story ahead of biochemistry. *Cold Spring Harb. Perspect. Biol.* **2** a002105
Raymond J and Blankenship RE 2004 Biosynthetic pathways, gene replacement and the antiquity of life *Geobiology* **2** 199–203
Raymond J and Segrè D 2006 The effect of oxygen on biochemical networks and the evolution of complex life. *Science* **311** 1764–1767
Sarich VM and Wilson AC 1967 Rates of albumin evolution in primates. *Proc. Natl. Acad. Sci. USA* **58** 142–148
Stott R 2012 *Darwin's ghosts: In search of the first evolutionists* (London: Bloomsbury)
Szathmáry E 2003 Why are there four letters in the genetic alphabet? *Nat. Rev. Genet.* **4** 995–1001
Tawfik DS 2010 Messy biology and the origins of evolutionary innovations. *Nat. Chem. Biol.* **6** 692–696
Thompson DW 1945 *On growth and form* (Cambridge: Cambridge University Press)
Ureta T 2011 *Origen y Evolución de Proteinas y Enzimas* (Santiago, Chile: Editorial Universitaria)
Valdecasas AG, Boto L and Correas AM 2013 There is no common ground between science and religion. *J. Biosci.* **38** 181–187
Woese CR and Fox GE 1977 Phylogenetic structure of the prokaryote domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74** 5088–5090

Woese CR and Goldenfeld N 2009 How the microbial world saved evolution from the Scylla of molecular biology and the Charybdis of the modern synthesis. *Microb. Mol. Biol. Revs.* **73** 14–21

Wright S 1934 Fisher's theory of dominance. *Am. Nat.* **63** 274–279

Yang Z, Chen F, Alvarado JB and Benner SA 2011 Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **133** 15105–15112

Zuckerkandl E and Pauling L 1962 Molecular disease, evolution, and genetic heterogeneity; in *Horizons in biochemistry* (ed) M Kasha and Pullman B (New York: Academic Press) pp 189–225

Zuckerkandl E and Pauling L 1965 Evolutionary divergence and convergence of proteins; in *Evolving genes and proteins* (ed) V Bryson and HJ Vogel (New York: Academic Press) pp 97–166

Athel Cornish-Bowden[1,]*, Juli Peretó[2] and María Luz Cárdenas[1]

[1]*Unité de Bioénergétique et Ingénierie des Protéines, Centre National de la Recherche Scientifique, Aix-Marseille Université, 31 chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France*

[2]*Departament de Bioquímica i Biologia Molecular, Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Valencia, Spain*

**Corresponding author (Email, acornish@imm.cnrs.fr)*