

## Minireview

# The logic of metabolism and its fuzzy consequences

**Antoine Danchin\*** and **Agnieszka Sekowska**

*Building G1, AMABiotics SAS, 2 rue Gaston Crémieux, Evry 91000, France.*

### Summary

**Intermediary metabolism molecules are orchestrated into logical pathways stemming from history (L-amino acids, D-sugars) and dynamic constraints (hydrolysis of pyrophosphate or amide groups is the driving force of anabolism). Beside essential metabolites, numerous variants derive from programmed or accidental changes. Broken down, variants enter standard pathways, producing further variants. Macromolecule modification alters enzyme reactions specificity. Metabolism conform thermodynamic laws, precluding strict accuracy. Hence, for each regular pathway, a wealth of variants inputs and produces metabolites that are similar to but not the exact replicas of core metabolites. As corollary, a shadow, *paralogous metabolism*, is associated to standard metabolism. We focus on a logic of paralogous metabolism based on diversion of the core metabolic mimics into pathways where they are modified to minimize their input in the core pathways where they create havoc. We propose that a significant proportion of paralogues of well-characterized enzymes have evolved as the natural way to cope with paralogous metabolites. A second type of denouement uses a process where protecting/deprotecting unwanted metabolites – conceptually similar to the procedure used in the laboratory of an organic chemist – is used to enter a completely new catabolic pathway.**

### Introduction

Metabolism is at the heart of life. Between life and death, there exists a state where life is in limbo, dormancy, but a dormant seed or spore will be considered alive only when

a flux of matter and energy resumes, initiating metabolism. This state of flux has striking but uncharted consequences at the border between permanence and change. To visualize the quandary, an old historical myth provides a vivid illustration of the question raised, when describing the return of Theseus, the king of Athens, from Crete: *The ship wherein Theseus and the youth of Athens returned had thirty oars, and was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting in new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same* [Plutarch 75 ACE (translated by John Dryden, 1994)]. Despite the variety in the kind and age of swapped planks, something of the ship remained unaffected regardless of the passing of time. This mysterious attribute lies in the *information* carried by the relationships that link the planks together (Gell-Mann and Lloyd, 1996). In the same way, while we commonly focus on the material make up of living organisms, life is a science of dynamic relationships between a variety of objects from very concrete molecules, where the arrangement of unique chemical bonds is cardinal, to the fairly abstract domain of regulation.

Life began with carbon-based chemistry and ended up with the flourishing of genes and gene expression (Dyson, 1985). Life as we know it programs via its genome the building up and management of compartmentalization (the cell's envelope, its nanomachines such as the ribosome or adenosine triphosphate (ATP) synthase, and a variety of other contraptions) and metabolism (chemical fluxes linking catalytic centres together, constructing and breaking down molecules for the manufacture, storage, salvage, maintenance of the cell's building blocks, coupled to energy management). For decades, intermediary metabolism was commonly considered as a boring haphazard collection of chemicals. It was taught in classes where students tried to remember by heart lists of compounds not really understood as following much logic. Yet, there is logic in the design of metabolism, and this logic begins to be deciphered in a complete revival of the topic. To describe this logic would ask for a whole

Received 3 July, 2013; revised 2 August, 2013; accepted 26 August, 2013. \*For correspondence. E-mail [antoine.danchin@normalesup.org](mailto:antoine.danchin@normalesup.org); Tel. (+331) 6087 5306; Fax (+331) 6087 5303.

textbook. We point out here some of the rules that are beginning to emerge, focusing on a seldom-described facet of metabolism, paralogous metabolism, that results from the unavoidable presence of errors in the unfolding of metabolic processes. We discuss paralogous metabolism as a shadow of core metabolism that operates on variants of core metabolites. It is both a way to dispose of potentially toxic variants, and making use of the inevitability of errors, it initiates an exploration of novel pathways that may evolve into beneficial functions.

#### *The logic of biomolecules' atomic composition*

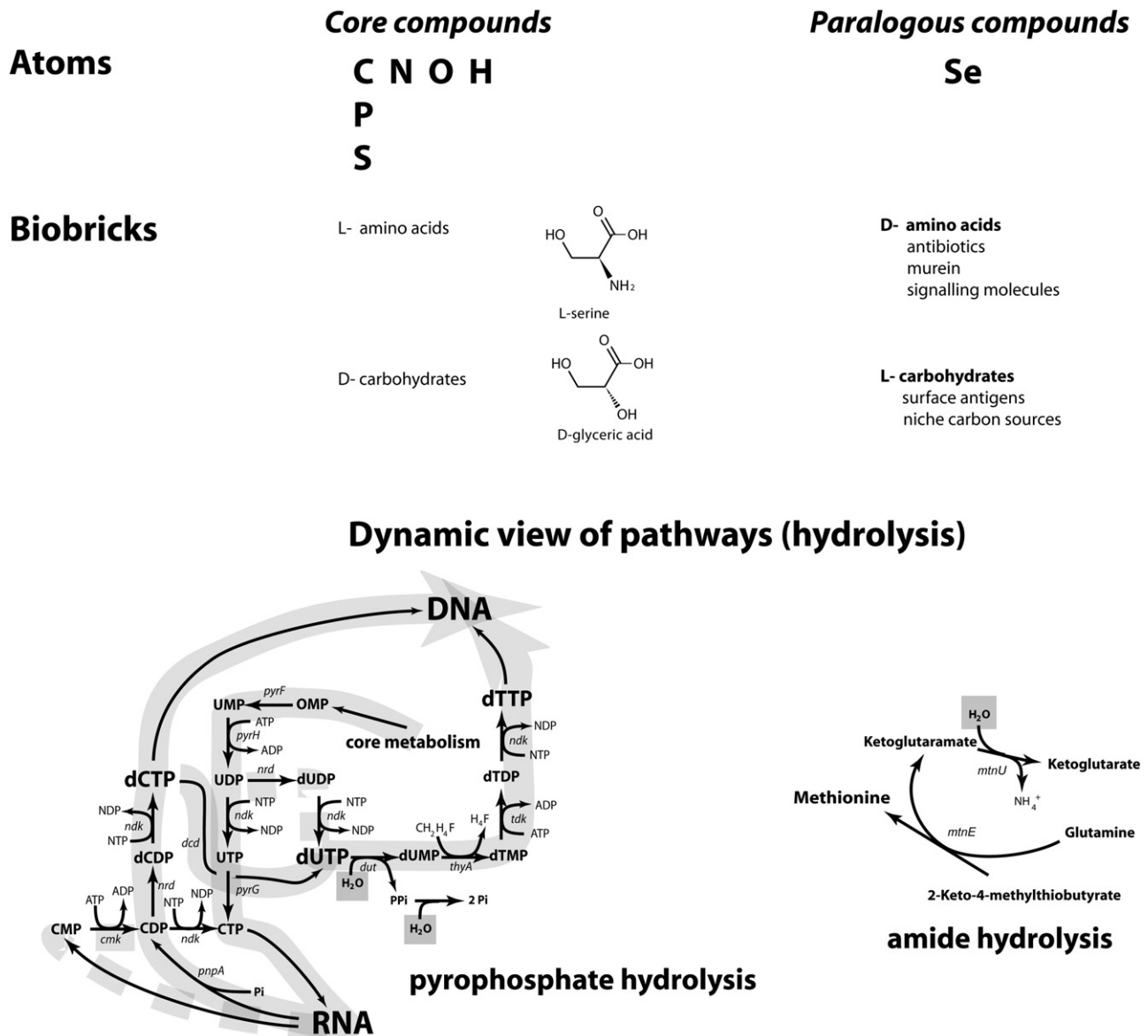
Present in most scenarios depicting the origins of life (Dyson, 1985), the chemicals that make up metabolism can be understood with textbook knowledge: why are the elements we find in living organisms in limited number? Why are they those we know? The short answer is straightforward. Life develops around 300 K, with water as its bathing medium and key metabolite. A core property of life's components is that beside a limited number of small molecules (a few tens of atoms), it is made of macromolecules, giant polymers obtained by elimination of a water molecule between a small alphabet of basic building blocks, amino acids and nucleotides. This seemingly simple arrangement allows management of information; while the backbone of these polymers is invariable, their side chains can be arranged in an unlimited variety of combinations. This informational view must however reckon with much physicochemical constraints. Making long-lived molecules, and a fortiori polymers, implies forming stable covalent bonds. A covalent bond materializes when electrons share their presence between two or more atoms' nuclei. This has a strong logical implication in terms of physics. Only the atoms belonging to the first rows in the periodic table of elements can make covalent interactions that are both varied and stable. The best candidates for making stable covalent links are, beside hydrogen, some of the atoms present in the second row of the periodic table. With its tetrahedral propensity for making covalent bonds, carbon is the atom that best explores the three dimensions of space. Three atoms are excluded for cosmological reasons: lithium, beryllium and boron are rare in the universe (Bernas *et al.*, 1967). Fluorine is also excluded because carbon-fluorine bonds are too stable to be easily exchanged (Furuya *et al.*, 2008). This limits the basic atoms of life to hydrogen, carbon, nitrogen and oxygen (typically, the atoms combined in extraterrestrial molecules, see, e.g. Pizzarello *et al.*, 2011).

Further down in the table, the atoms are progressively less prone to make covalent bonds that are both diversified and stable. They become involved in the building up of electrostatic bonds (much weaker than covalent bonds)

and in electron transfers (redox reactions). Several play critical roles in the biological catalytic reactions needed to construct, modify and destroy covalent bonds, and produce electrostatic interactions (metals) and more or less complicated electron transfers (transition metals and complex anions such as molybdate or tungstate). Two notable exceptions remain, sulphur (the higher homologue of oxygen) and phosphorus (the higher homologue of nitrogen). Phosphorus has been retained because of a unique chemical property of phosphates in water; these ionic structures can easily polymerize or form phosphoesters by dehydration. Furthermore, whereas their hydrolysis is exothermic, they remain stable at 300 K in water because phosphate hydrolysis requires a considerable energy of activation. This allowed phosphates to become a unique energy store and currency while keeping the ability to be retained in macromolecules as phosphodiester bonds (Westheimer, 1987). Sulphur is involved in electron transfers, in formation of energy-rich thioesters and via its affinity for iron, in making iron-sulphur clusters that probably had a seminal role in the origin of life on Earth, where it is quite abundant (Wickramasinghe, 1973). In a way, not so different from the build up of man-made crafts such as Theseus boat, these atoms are combined into all the components of life, following certain informational rules. Obeying these rules, however, cannot be strictly accurate. They have a fuzzy character associated to the inevitable thermal fluctuations. The appearance of metabolic shadows resulting in mimicry or paralogy plays a prominent role, as discussed later. At the level of atoms, this inevitable outcome is visible with some metals (those with an octahedral coordination sphere tend to swap roles, e.g. cobalt or manganese replacing magnesium) and at least one atom, selenium, sometimes mimics and replaces sulphur. As it is the rule with the way living organisms trap information (Binder and Danchin, 2011), this property, despite being the source of selenium toxicity, was recruited in a positive outcome of evolution as the 21st proteinogenic amino acid, selenocysteine (Johansson *et al.*, 2012). Selenocysteine is thus the first paralogous metabolite (Fig. 1), highlighting the positive and negative roles of the cognate paralogous pathways as a function of the distribution of selenium in natural and anthropized environments and accumulation via the food web (Daniels, 1996; Hladun *et al.*, 2012).

#### *Elements of the logic of metabolism*

As a start point, and among many choices, we illustrate some features of textbook metabolism that have seldom been understood as following consistent biochemical logic. This will allow us to understand how this logic opens up an entirely new world of metabolic options.



**Fig. 1.** The logic of metabolism and its consequences. We summarize here some of the logic of metabolism: atoms required to make stable covalent bonds, manage energy and electron transfers, with selenium as a paralogous atom of sulphur; the L-amino acids in proteins impose D-carbohydrates, and subsequently, paralogous metabolism creates novel functions for D-amino acids and L-carbohydrates; finally, metabolism is a dynamic process where the ubiquitous presence of water is used to drive reactions forward. Pyrimidine nucleotides biosynthesis (see text for details) requires pyrophosphate hydrolysis to prevent uracil to enter DNA. Salvage of methionine uses an irreversible step where 2-ketoglutaramate is produced and subsequently hydrolysed into 2-ketoglutarate.

*Structural logic: L-amino acids and D-sugars* It has been widely noticed that life is built up using L-amino acids and D-sugars. A prevailing, generally unconscious vitalist view assumes that the amino acid chiral bias cannot be contingent but must result from a particular causal process (or even design) that needs to be explained (see for example a physicists' view, Boyd *et al.*, 2013). However, the cost of maintaining symmetry makes it futile to look for an enlightening cause of a primary break of symmetry. The simplest logic here is that once a contingent break of symmetry

has occurred, it constrained further breaks (Gleiser *et al.*, 2008). The chemical world of life was doomed to become asymmetrical.

However, subsequent related questions open up: is there a link between the two chiral preferences? Does the L- choice for amino acids impose the D- choice for carbohydrates? Prebiotic carbon chemistry formed a great many specific chiral molecules in racemic proportions at the onset of their synthesis. Maintaining symmetry can be extremely costly. At some point a symmetry-breaking

imbalance in racemic mixtures led to systematic accumulation of a particular class of stereoisomers, L-amino acids for example. In the same way as continental Europeans drive on the right, it would be difficult to change the rule. The start point – to drive on the left or on the right – may have been contingent, although being right-handed certainly had a say in the initial rule of riding left. We can start on either side, but once chosen, one must keep it as it is to avoid conflicts, unless an external command comes to reorganize the whole. Once chiral molecules became involved in their own synthesis or salvage – this must be so because reproduction entails some level of autocatalysis (Dyson, 1985; Markovitch *et al.*, 2012) – they were trapped into their first chiral imbalance. A restricted alphabet of proteinogenic amino acids was then picked up, constrained by the requirements of ubiquitous metabolic pathways and by the availability of carbon, nitrogen and sulphur, as well as overall metabolic costs (Akashi and Gojoberi, 2002; Pascal *et al.*, 2006; Kaleta *et al.*, 2013). Twenty (22, with selenocysteine and pyrrolysine) L-amino acids constitute the basic building blocks of proteins. Retained during evolution, they are now recognized by all enzyme cavities in a highly space-constrained way. As a consequence, most proteases and peptidases are active on chains made of the L-enantiomers, not their D- counterparts. As a common functional consequence of established constraints, unavoidable errors opened the logic for a selection pressure leading to protease-insensitive protective structures that evolved towards containing the resistant D-isomers (Werle and Bernkop-Schnurch, 2006) and accounting for the structure of antibiotics and murein.

A further selection pressure was also set in motion. It explains the ubiquity of D-carbohydrates. Translation is at the core of life. Any interference with the process is deleterious, as seen in the role of the majority of antibiotics. In particular, molecules that mimic proteinogenic amino acids display some toxicity (see, e.g. Bence and Crooks, 2003; Sakurama *et al.*, 2009). Replacing the alpha-amino group with a hydroxyl group creates a molecule that is a potentially toxic mimic of a proteinogenic amino acid. As a case in point, L-glycerate is similar to L-serine and competes with it in many reactions (Fig. 1). Hence, D-glycerate is the preferred stereoisomer, L-glycerate being generally toxic (Zhu and Lin, 1987). D-glycerate-related molecules are pivotal in carbohydrate metabolism. It follows that the majority of carbohydrates are of the D- stereoisomer type. This stereochemical constraint has consequences for the make-up of nucleic acids. Its logic explains why most biologically relevant carbohydrates, ribose and deoxyribose, in particular, are of the D- type. Selection of a bias has the consequence that it creates opportunities for using the symmetrical bias in novel functions. As for L-amino acids errors in this D-

chirality bias triggered a selection pressure that resulted in the formation of L-sugars such as L-rhamnose [e.g. in rhamnolipids (Aguirre-Ramirez *et al.*, 2012)]. These compounds created an opportunity for the emergence of novel metabolic pathways as they were subsequently accumulated in the environment. They are even used as a preferred carbon source by some organisms (Kim *et al.*, 2012).

*The logic of metabolism's dynamics: driving reactions forward* The view just presented is static. Yet, metabolism is a dynamic process. Beside the logic that clusters molecules along with common specific structural properties, the time-dependent direction of biochemical processes must also follow a logical ordering. We illustrate this logic with examples displaying the key role of hydrolysis.

#### *Hydrolysis of pyrophosphate drives anabolism*

The great majority of biochemical reactions is reversible. This is even the case of obviously anabolic reactions such as that carried out by DNA polymerase (Lahiri *et al.*, 2013). Still, most organisms keep building up new biomass. A forward direction must be imposed on major anabolic reactions despite reversibility. Can we identify the forces driving this process? Consuming energy to go forward is likely, but sparing as much energy as possible is important as well. Phosphate-related bonds play a key role in this domain because of the high activation energy necessary for hydrolysis. The upshot of this faculty is that in notable salvage processes, phosphorolysis is used instead of hydrolysis to preserve energy-rich phosphate bonds. This accounts for the presence of polynucleotide phosphorylase in the RNA degradosome. While the enzyme can work as a polymerase without template, it is mainly used as a source of the ribonucleoside diphosphates that are the precursors of the deoxyribonucleotides (Danchin, 2009; Becket *et al.*, 2012). Indeed, synthesis of deoxyribonucleotides uses ribonucleoside diphosphates, not triphosphates, as substrates. dNDPs are subsequently phosphorylated by nucleoside diphosphokinase via a reversible reaction driven by the excess of ATP over adenosine diphosphate (ADP). This organization has an unexpected consequence: de novo pyrimidine biosynthesis goes through uridine diphosphate (UDP) as an intermediate, but uridine triphosphate (UTP) is the direct precursor of cytidine triphosphate (CTP), so that no cytidine diphosphate (CDP) is formed (Fig. 1). If this was not balanced by other pathways, U would readily go into DNA, which would lack C (Nitschke *et al.*, 1998). The only way to prevent U entering DNA during replication is to keep the deoxyuridine triphosphate (dUTP) concentration as low as possible. This is achieved via the action of an essential enzyme, dUTPase, that



forms deoxyuridine monophosphate (dUMP) [the precursor of deoxythymidine monophosphate (dTMP)] and pyrophosphate (PPi) (Persson *et al.*, 2001). Because the reaction is reversible, this can only be met if the PPi concentration is kept low. This target is brought about by pyrophosphatase, demonstrating that PPi hydrolysis is a key metabolic driving force (Fig. 1). A similar answer is established for transcription and replication (Lahiri *et al.*, 2013): in the presence of PPi, the polymerases go backwards.

PPi is thus used as the pivotal element in anabolism. Besides being involved in the biosynthesis of all macromolecules (translation, transcription, replication, synthesis of polysaccharides used in storage, adhesion and protection . . .), it is a ubiquitous product of countless anabolic processes (Heinonen, 2001). It is found in the synthesis of amino acids (asparagine, arginine, cysteine, histidine . . .), coenzymes [coenzyme A, folate, nicotinamide adenine dinucleotide (NAD), riboflavin . . .], the ubiquitous key metabolite S-adenosylmethionine, phospholipids and murein. It is also involved in regulation (for example in synthesis of cAMP, cyclic 3',5' adenosine monophosphate). Modulation of the PPi level can be tuned as a key strategy for modulation of biological functions, including management of information. As expected, pyrophosphatase is ubiquitous and essential (Kajander *et al.*, 2013). The logic of its involvement is straightforward despite the fact that it appears to be wasting energy. It is likely that membrane pyrophosphatase – sometimes thought to generate PPi (Russell *et al.*, 2013), contrary to what is proposed here – is designed to recover some energy via creation of a proton gradient that may be used to drive ATP synthesis (Luoto *et al.*, 2011; Huang *et al.*, 2013; Kajander *et al.*, 2013).

#### *Irreversible deamidation prevents unwanted catabolism*

Analysis of seemingly wasteful or enigmatic reactions is a way to identify the underlying logic of metabolic steps. Eventhough reactions are reversible, the high concentration of water makes that hydrolysis will generally dominate reactions where water is involved. Here is another example. Most aminotransferases use an amino acid as the alpha-amino group donor. In many cases, aspartate or glutamate is the donor, as expected from their central place in metabolism. The reaction produces oxaloacetate or 2-ketoglutarate that nicely enters the TCA cycle. Yet, in some cases, *glutamine* is used as the alpha-amino group donor. This is surprising because the product of the reaction, 2-ketoglutaminate, is a reactive molecule that does not readily enter metabolism and is generally toxic. It can also spontaneously cyclize into the lactam form 2-hydroxy-5-oxoproline, the fate of which has not been explored in-depth. Paralogous metabolites such as

succinamate (Krasnikov *et al.*, 2009) or oxamate (Bojanowski *et al.*, 1964) are also produced under a variety of conditions, but not much is known about their fate despite some knowledge of omega-amidases in Eukarya (Mora, 1990; Jaisson *et al.*, 2009). In Bacteria, analysis of the main origin of 2-ketoglutaminate revealed that it is the result of a transamination reaction producing methionine (Berger *et al.*, 2003). This reaction is the final step of the methionine salvage pathway (Sekowska *et al.*, 2004). In *Bacillus subtilis*, this pathway is connected with an enzyme, MtnU, the omega-amidase that hydrolyses 2-ketoglutaminate to 2-ketoglutarate (Belda *et al.*, 2013). The presence of this reaction based on hydrolysis introduces an irreversible step that prevents methionine degradation, allowing the cell to save a costly metabolite (Fig. 1). More generally, this reaction is used in fungi to drive ammonium acquisition (Mora, 1990). Many more similar reactions, where hydrolysis is the driving force, will eventually be found in the future. Surprisingly, not much is known about the likely ubiquitous presence in the environment of these molecules or their cyclization product (2-hydroxy-5-oxoproline and succinimide, for example), and as a consequence, not much is known about metabolic pathways that would use them (Ogawa *et al.*, 1996).

#### *Fuzzy logic: potpourri of the metabolic consequences of the passing of time*

The logic exposed earlier assumes that pathways and metabolites are invariable. Yet, enzymes are macromolecules, and all macromolecules are prone to age with concomitant slips in activity. Many are modified post-transcription or post-translation. Some must also turnover to bring about regulatory functions. When not susceptible to repair, macromolecules are shredded into small pieces, producing some of the original building blocks, which are salvaged and recycled. Modified residues, which come either from programmed modifications or from age-related alterations, build up a collection of non-standard metabolites, the ultimate fate of which has seldom been explored. A further source of non-standard metabolites results from inaccuracies in the enzyme reactions. All these metabolites, whether resulting from programmed pathways, errors or ageing, are mimics of core metabolites. They may take their place in existing pathways, starting a further source of metabolic variants. This is the entry point of a variety of shadow pathways, generally structurally related to core pathways, that we named paralogous metabolism.

While the existence of programmed modification of biological compounds, as well as modifications related to senescence are widely acknowledged (but poorly explored in general, e.g. scanning PubMed or PubMed Central or even Google for '3-methyladenine degradation'

or 'catabolism' or 'breakdown' yields no result), the fact that enzyme activities may be tied to a systematic production of errors *in vivo* is just beginning to be explored [in the context of animal metabolic disorders (Van Schaftingen *et al.*, 2013)]. Yet, the idea that errors played a critical role in the development of life goes back to the reflections that resulted in the concept of evolution by natural selection. Evolution required that variants existed, and in a process dominated by reproduction, this meant that the progeny was not systematically an exact copy of the parents, but a similar copy (Dyson, 1985). 'Similar' implied that the process of reproduction was not always faithful. With the advent of genetics and the understanding of DNA as the carrier of the organism's blueprint, the reflection focused on the idea of mutations as errors during and after the process of DNA replication. A considerable body of studies was therefore devoted to the exploration of errors and proofreading at the level of DNA (Dalhus *et al.*, 2009). In contrast, not much work explored the ubiquitous presence of errors elsewhere in biological processes, except during translation (Hopfield, 1974; Pechmann *et al.*, 2013). Yet, exquisitely accurate and specific enzymes were not born fully grown and armed with the properties some [but by no means all (Yip and Matsumura, 2013)] display today. Ancient enzymes were fuzzy in the way they recognized their substrates (Kajander *et al.*, 2013). In the course of evolution, many but far from all, progressively became more and more specific (Nam *et al.*, 2012).

It was recognized 40 years ago that the ability of many enzymes to catalyse the metabolic flow of multiple substrates plays a key role in the evolution of new function (Jensen, 1976; O'Brien and Herschlag, 1999). This may account for the fact that core pathways in extant organisms still display functional promiscuity. For example, the biosynthesis of valine, isoleucine and pantothenate shares several enzymes (acetolactate synthase, acetohydroxy acid isomeroreductase, dihydroxy acid dehydratase, branched-chain amino acid aminotransferase), while these metabolites originate from pyruvate (valine and pantothenate) and 2-ketobutyrate (isoleucine) (Park and Lee, 2010). In the same way, serine and pyridoxal-phosphate synthesis share aminotransferase SerC that recognizes two different substrates (Drewke *et al.*, 1996) (Fig. 2).

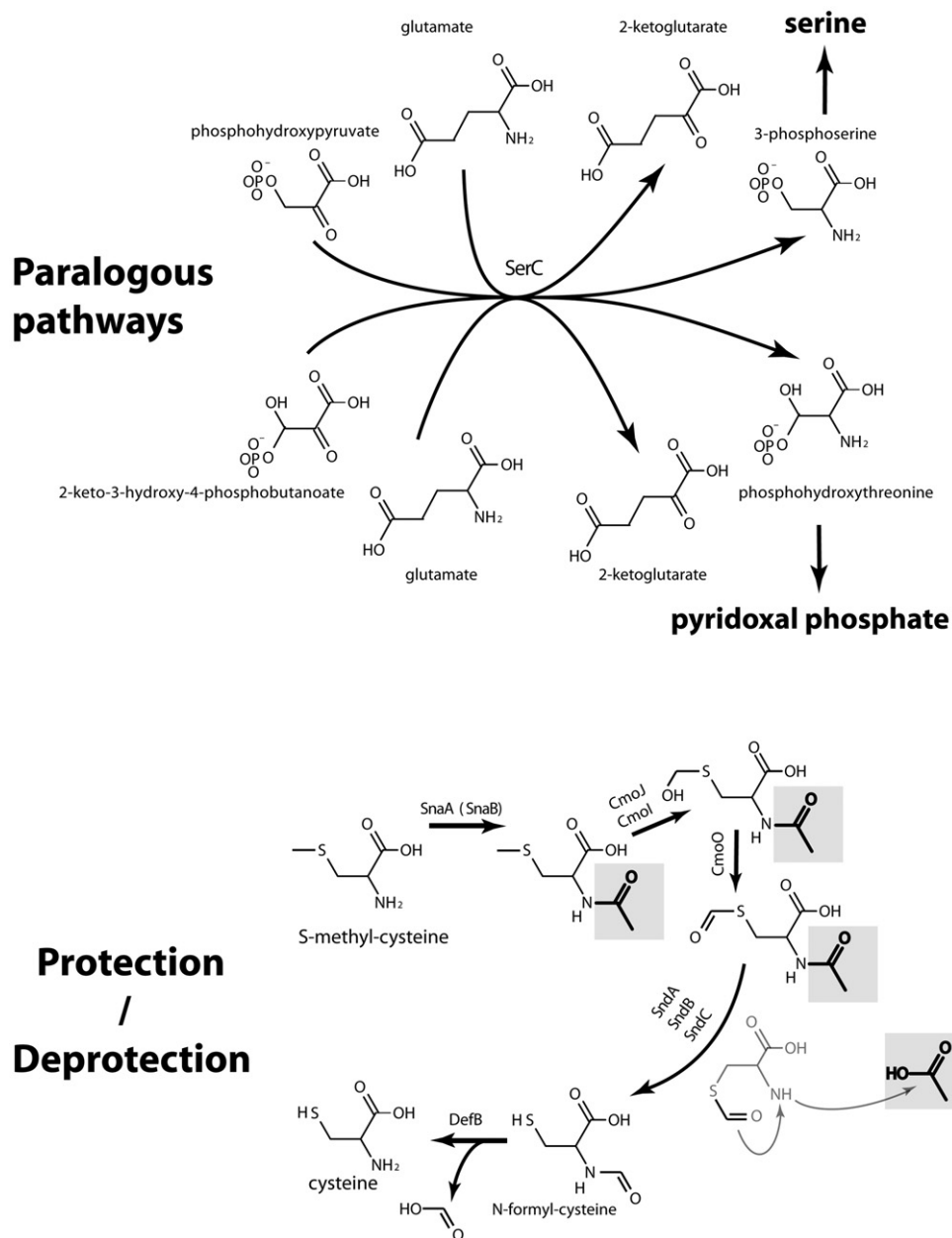
A variety of such housekeeping paralogous pathways exist in different organisms and enzyme activity moonlighting is the rule rather than the exception. Many explanations have been proposed to account for that situation, in particular 'demands for higher metabolic flux may provide an evolutionary selective pressure to enhance an enzyme's catalytic rate and reduce the required enzyme concentration. However, catalytic improvements for one substrate of a generalist enzyme can suppress other cata-

lytic activities' (Nam *et al.*, 2012). The outcome is that both enzyme substrates and enzymes are somewhat fuzzy. In short, cells keep making compounds similar to housekeeping metabolites. Because they are similar to normal compounds, they may easily take their place in some of the standard reactions of the cell and poison them. They may also be recruited for the construction of novel pathways that may be beneficial for the organism that is competing with others for resources in the environment.

This has considerable consequences as, from the chemical point of view, there is no difference between a regular substrate and a chemically related compound. It follows that errors and programmed modifications must be considered together as they are often chemically related and as the explicit and possibly anecdotal sequence of events that gave rise to a particular metabolite does not show up in its structure. We propose here scenarios for the becoming of these ubiquitous variants of core metabolites, *paralogous metabolites* for short. Briefly, the inevitability of errors in metabolic processes has created the conditions for recurrent recruitment of variants of the regular building blocks for novel functions in the course of evolution. This hypothesis entails the generation of a functional logic in the way these metabolites are processed. Two distinct pathways are emphasized here.

First, the very fact that paralogous metabolites are variants of key metabolites makes room for them to be processed by variants of the enzymes involved in the related core metabolic pathways. Homeomorphism, a common concept in chemistry, makes it plausible to consider that paralogues of core enzymes, continuously produced via the process of gene duplication followed by mutation (Kondrashov, 2012), may be recruited for these functions. This constitutes a novel class of pathways, *paralogous pathways*. Unquestionably, they will be difficult to identify *in silico* without explicit *in vivo* experiments [see for example the difficult identification of the highly similar arginases and agmatinases, often misannotated (Sekowska *et al.*, 2000)]. Indeed, the initial substrates and their variants are highly related, so that it is expected that the structure of the enzymes that process them be also related.

Because these enzymes are descended from the initial enzymes they are structurally similar. A great many enzyme activities that depend on co-factors such as pyridoxal phosphate or NAD are not fully characterized, and they make a sizeable proportion of the putative enzymes of unknown function coded in the many bacterial genomes that keep being sequenced [see for example the updated annotation of *B. subtilis* (Belda *et al.*, 2013)]. Many are likely involved in paralogous metabolism. In the same way, clusters of genes containing epimerases, mutases and dehydratases are likely to be involved in the



**Fig. 2.** Paralogous metabolism and its consequences. Synthesis of serine and pyridoxal phosphate use an enzyme in common, SerC. This moonlighting property of enzymes is widespread and is likely to be used in the construction of pathways built on paralogous metabolites. Such metabolites are ubiquitous and often toxic. We illustrate the case of S-methyl-cysteine, an analogue of methionine, which shows how protection via acetylation is used to prevent the metabolite to enter translation while being processed in an oxidative pathway. Upon deacetylation the molecule follows a straightforward path to yield cysteine (Chan *et al.*, 2013, this issue).

catabolism of related carbohydrates. This process of paralogous metabolism is widely spread, and it is not always linked to catabolism. For example, it has been positively recruited in the genomic islands that code for surface complex carbohydrates that are based on variants of core carbohydrate units (see for example McCallum *et al.*, 2012), contributing to antigenic variation. Here again, functional recruitment of pre-existing entities

is the rule. Variants of the enzymes that resulted in the synthesis of erroneous metabolites are obvious candidates to involve them in variant pathways. Hence, a natural way to evolve is to use paralogues of enzymes to implement variant – paralogous – pathways. As a matter of fact, this process has been activated repeatedly in the course of evolution for the modification of nucleic acids: the same enzyme activity of pseudouridine synthases

is used on different substrates (tRNAs, rRNAs) with enzymes that are structurally related (e.g. RluB, RsuA and possibly YjbC in *Escherichia coli*). It is also prominent in many organisms for the parallel degradation of pyruvate and branched-chain amino acid dehydrogenase complexes (Bunik and Degtyarev, 2008).

Based on a completely different logic, a second family of pathways has to be considered, especially when the paralogous compound of interest negatively interferes with a key pathway. In this situation, paralogous metabolites need to be exported or entered into catabolic pathways. Exported molecules enrich the environment in molecules that initiate a complete universe of possible nutrients. This opens up the discovery of novel activities or the recruitment of existing promiscuous activities (Carbonell *et al.*, 2011; Luo *et al.*, 2012), as discussed earlier. Yet, export may not be enough to save the cell from poisoning. For this reason, even before being exported, addition of a relevant chemical group on the molecule will prevent its input in important pathways and possibly tag it for a novel degradation pathway. This can be achieved by tagging paralogous metabolites with a ubiquitous chemical group, such as an acyl- or an alkyl-group (Sekowska *et al.*, 2001). Subsequently, the tagged compound may be exported [export safety valves used for other purposes are widespread in cells (Danchin, 2012)] or enter a specific catabolic pathway. We have illustrated the latter situation in an accompanying paper (Chan *et al.*, 2013), where the fate of S-methyl-cysteine, an analogue of methionine and branched-chain amino acids is unravelled (Fig. 2). The fascinating logic of the corresponding pathways is that the cell proceeds in a way quite similar to that of the organic chemist; it first protects the metabolite so that it does not go astray. Then, it submits it to specific catabolic reactions, and finally, it deprotects the modified metabolite, releasing a core building block for the cell to use.

#### Concluding remarks

A most intriguing question of biology is that of the emergence of novel functions. To answer it, we depicted metabolism as submitted to the logic of physicochemical constraints. We explored the functional consequences of the presence of variants in metabolism, resulting from reactions produced by errors or programmed modifications of biological compounds. Emergence of novel functions always comes from the creation of opportunities. Paralogous metabolites – the name we give to variants of regular metabolites – are continuously produced by living organisms. Had not they been input in novel metabolic pathways, they would progressively accumulate in the environment (this is indeed what happens with lignocellulosic waste). However, their very existence

creates an opportunity for degradation and salvage of usable elementary components. It also allows for exploration of novel useful functions. There is no function without a material support that implements it. Evolution by gene duplications creates the complementary opportunity via emergence of activities that are similar to those present in the parent standard pathways. When no valuable novel function has yet been discovered, another logical way to cope with unwanted waste is degradation to its ultimate components after modification of the unwanted metabolite to prevent it from poisoning regular pathways. This is indeed the solution that emerged in some situations, and we can hope that lessons taken from this particular process will help us conceive novel approaches to the destruction of man-made pollutants. Finally, while conforming to the established properties of matter and the local stability and relevance of known pathways, the logic of metabolism does not seem to be optimized in terms of what would be a human engineering design [but we may dispute that human engineering is really optimized (Verna, 2011)]. This implies that understanding optimization – in particular for managing accuracy and sparing energy and matter – will be at the core of next decade metabolic engineering.

#### Acknowledgements

We wish to thank anonymous reviewers and Eric Fourmentin for constructive comments. This work was supported by European Union's 7th Framework Programme Microme FP7-KBBE-2007-3-2-08-222886.

#### References

- Aguirre-Ramirez, M., Medina, G., Gonzalez-Valdez, A., Grosso-Becerra, V., and Soberon-Chavez, G. (2012) The *Pseudomonas aeruginosa* *rmlBDAC* operon, encoding dTDP-L-rhamnose biosynthetic enzymes, is regulated by the quorum-sensing transcriptional regulator RhlR and the alternative sigma factor sigmaS. *Microbiology* **158**: 908–916.
- Akashi, H., and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci USA* **99**: 3695–3700.
- Becket, E., Tse, L., Yung, M., Cosico, A., and Miller, J.H. (2012) Polynucleotide phosphorylase plays an important role in the generation of spontaneous mutations in *Escherichia coli*. *J Bacteriol* **194**: 5613–5620.
- Belda, E., Sekowska, A., Le Fevre, F., Morgat, A., Mornico, D., Ouzounis, C., *et al.* (2013) An updated metabolic view of the *Bacillus subtilis* 168 genome. *Microbiology* **159**: 757–770.
- Bence, A.K., and Crooks, P.A. (2003) The mechanism of L-canavanine cytotoxicity: arginyl tRNA synthetase as a novel target for anticancer drug discovery. *J Enzyme Inhib Med Chem* **18**: 383–394.



- Berger, B.J., English, S., Chan, G., and Knodel, M.H. (2003) Methionine regeneration and aminotransferases in *Bacillus subtilis*, *Bacillus cereus*, and *Bacillus anthracis*. *J Bacteriol* **185**: 2418–2431.
- Bernas, R., Gradsztajn, E., Reeves, H., and Schatzman, E. (1967) On the nucleosynthesis of lithium, beryllium, and boron. *Ann Phys* **44**: 426–478.
- Binder, P.M., and Danchin, A. (2011) Life's demons: information and order in biology. What subcellular machines gather and process the information necessary to sustain life. *EMBO Rep* **12**: 495–499.
- Bojanowski, R., Gaudy, E., Valentine, R.C., and Wolfe, R.S. (1964) Oxamic transcarbamylase of *Streptococcus allantoicus*. *J Bacteriol* **87**: 75–80.
- Boyd, R.N., Kajino, T., and Onaka, T. (2013) Supernovae, neutrinos and the chirality of amino acids. *Int J Mol Sci* **12**: 3432–3444.
- Bunik, V.I., and Degtyarev, D. (2008) Structure-function relationships in the 2-oxo acid dehydrogenase family: substrate-specific signatures and functional predictions for the 2-oxoglutarate dehydrogenase-like proteins. *Proteins* **71**: 874–890.
- Carbonell, P., Lecoindre, G., and Faulon, J.L. (2011) Origins of specificity and promiscuity in metabolic networks. *J Biol Chem* **286**: 43994–44004.
- Chan, C.M., Danchin, A., Marlière, P., and Sekowska, A. (2013) Paralogous metabolism: S-alkyl-cysteine degradation in *Bacillus subtilis*. *Environ Microbiol*, doi: 10.1111/1462-2920.12210 (in press).
- Dalhous, B., Laerdahl, J.K., Backe, P.H., and Bjoras, M. (2009) DNA base repair – recognition and initiation of catalysis. *FEMS Microbiol Rev* **33**: 1044–1078.
- Danchin, A. (2009) A phylogenetic view of bacterial ribonucleases. *Prog Mol Biol Transl Sci* **85**: 1–41.
- Danchin, A. (2012) Scaling up synthetic biology: do not forget the chassis. *FEBS Lett* **586**: 2129–2137.
- Daniels, L.A. (1996) Selenium metabolism and bioavailability. *Biol Trace Elem Res* **54**: 185–199.
- Drewke, C., Klein, M., Clade, D., Arenz, A., Muller, R., and Leistner, E. (1996) 4-O-phosphoryl-L-threonine, a substrate of the *pdxC*(serC) gene product involved in vitamin B6 biosynthesis. *FEBS Lett* **390**: 179–182.
- Dyson, F.J. (1985) *Origins of Life*. Cambridge, UK: Cambridge University Press.
- Furuya, T., Kuttruff, C.A., and Ritter, T. (2008) Carbon-fluorine bond formation. *Curr Opin Drug Discov Devel* **11**: 803–819.
- Gell-Mann, M., and Lloyd, S. (1996) Information measures, effective complexity and total information. *Complexity* **2**: 44–52.
- Gleiser, M., Thorarinson, J., and Walker, S.I. (2008) Punctuated chirality. *Orig Life Evol Biosph* **38**: 499–508.
- Heinonen, J.K. (2001) *Biological Role of Inorganic Pyrophosphate*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Hladun, K.R., Smith, B.H., Mustard, J.A., Morton, R.R., and Trumble, J.T. (2012) Selenium toxicity to honey bee (*Apis mellifera* L.) pollinators: effects on behaviors and survival. *PLoS ONE* **7**: e34137.
- Hopfield, J.J. (1974) Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci USA* **71**: 4135–4139.
- Huang, Y.T., Liu, T.H., Lin, S.M., Chen, Y.W., Pan, Y.J., Lee, C.H., et al. (2013) Squeezing at entrance of proton transport pathway in proton-translocating pyrophosphatase upon substrate binding. *J Biol Chem* **288**: 19312–19320.
- Jaisson, S., Veiga-da-Cunha, M., and Van Schaffingen, E. (2009) Molecular identification of omega-amidase, the enzyme that is functionally coupled with glutamine transaminases, as the putative tumor suppressor Nit2. *Biochimie* **91**: 1066–1071.
- Jensen, R.A. (1976) Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30**: 409–425.
- Johansson, A.L., Collins, R., Arner, E.S., Brzezinski, P., and Hogbom, M. (2012) Biochemical discrimination between selenium and sulfur 2: mechanistic investigation of the selenium specificity of human selenocysteine lyase. *PLoS ONE* **7**: e30528.
- Kajander, T., Kelloso, J., and Goldman, A. (2013) Inorganic pyrophosphatases: one substrate, three mechanisms. *FEBS Lett* **587**: 1863–1869.
- Kaleta, C., Schauble, S., Rinas, U., and Schuster, S. (2013) Metabolic costs of amino acid and protein production in *Escherichia coli*. *Biotechnol J* **8**: 1105–1114.
- Kim, S.M., Paek, K.H., and Lee, S.B. (2012) Characterization of NADP<sup>+</sup>-specific L-rhamnose dehydrogenase from the thermoacidophilic Archaeon *Thermoplasma acidophilum*. *Extremophiles* **16**: 447–454.
- Kondrashov, F.A. (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci* **279**: 5048–5057.
- Krasnikov, B.F., Nostramo, R., Pinto, J.T., and Cooper, A.J. (2009) Assay and purification of omega-amidase/Nit2, a ubiquitously expressed putative tumor suppressor, that catalyzes the deamidation of the alpha-keto acid analogues of glutamine and asparagine. *Anal Biochem* **391**: 144–150.
- Lahiri, I., Mukherjee, P., and Pata, J.D. (2013) Kinetic characterization of exonuclease-deficient *Staphylococcus aureus* PolC, a C-family replicative DNA polymerase. *PLoS ONE* **8**: e63489.
- Luo, J., van Loo, B., and Kamerlin, S.C. (2012) Catalytic promiscuity in *Pseudomonas aeruginosa* arylsulfatase as an example of chemistry-driven protein evolution. *FEBS Lett* **586**: 1622–1630.
- Luoto, H.H., Belogurov, G.A., Baykov, A.A., Lahti, R., and Malinen, A.M. (2011) Na<sup>+</sup>-translocating membrane pyrophosphatases are widespread in the microbial world and evolutionarily precede H<sup>+</sup>-translocating pyrophosphatases. *J Biol Chem* **286**: 21633–21642.
- McCallum, M., Shaw, S.D., Shaw, G.S., and Creuzenet, C. (2012) Complete 6-deoxy-D-altro-heptose biosynthesis pathway from *Campylobacter jejuni*: more complex than anticipated. *J Biol Chem* **287**: 29776–29788.
- Markovitch, O., Sorek, D., Lui, L.T., Lancet, D., and Krasnogor, N. (2012) Is there an optimal level of open-endedness in prebiotic evolution? *Orig Life Evol Biosph* **42**: 469–473. discussion 474.
- Mora, J. (1990) Glutamine metabolism and cycling in *Neurospora crassa*. *Microbiol Rev* **54**: 293–304.

- Nam, H., Lewis, N.E., Lerman, J.A., Lee, D.H., Chang, R.L., Kim, D., and Palsson, B.O. (2012) Network context and selection in the evolution to enzyme specificity. *Science* **337**: 1101–1104.
- Nitschke, P., Guerdoux-Jamet, P., Chiapello, H., Faroux, G., Henaut, C., Henaut, A., and Danchin, A. (1998) Indigo: a World-Wide-Web review of genomes and gene functions. *FEMS Microbiol Rev* **22**: 207–227.
- O'Brien, P.J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* **6**: R91–R105.
- Ogawa, J., Soong, C.L., Honda, M., and Shimizu, S. (1996) Novel metabolic transformation pathway for cyclic imides in *Blastobacter* sp. strain A17p-4. *Appl Environ Microbiol* **62**: 3814–3817.
- Park, J.H., and Lee, S.Y. (2010) Fermentative production of branched chain amino acids: a focus on metabolic engineering. *Appl Microbiol Biotechnol* **85**: 491–506.
- Pascal, G., Medigue, C., and Danchin, A. (2006) Persistent biases in the amino acid composition of prokaryotic proteins. *Bioessays* **28**: 726–738.
- Pechmann, S., Willmund, F., and Frydman, J. (2013) The ribosome as a hub for protein quality control. *Mol Cell* **49**: 411–421.
- Persson, R., Cedergren-Zeppeauer, E.S., and Wilson, K.S. (2001) Homotrimeric dUTPases; structural solutions for specific recognition and hydrolysis of dUTP. *Curr Protein Pept Sci* **2**: 287–300.
- Pizzarello, S., Williams, L.B., Lehman, J., Holland, G.P., and Yarger, J.L. (2011) Abundant ammonia in primitive asteroids and the case for a possible exobiology. *Proc Natl Acad Sci USA* **108**: 4303–4306.
- Plutarch 75 ACE (translated by John Dryden) (1994) Theseus. [WWW document]. URL <http://classics.mit.edu/Plutarch/theseus.1b.txt>.
- Russell, M.J., Nitschke, W., and Branscomb, E. (2013) The inevitable journey to being. *Philos Trans R Soc Lond B Biol Sci* **368**: 20120254.
- Sakurama, H., Takita, T., Mikami, B., Itoh, T., Yasukawa, K., and Inouye, K. (2009) Two crystal structures of lysyl-tRNA synthetase from *Bacillus stearothermophilus* in complex with lysyladenylate-like compounds: insights into the irreversible formation of the enzyme-bound adenylate of L-lysine hydroxamate. *J Biochem* **145**: 555–563.
- Sekowska, A., Danchin, A., and Risler, J.L. (2000) Phylogeny of related functions: the case of polyamine biosynthetic enzymes. *Microbiology* **146** (Pt 8): 1815–1828.
- Sekowska, A., Robin, S., Daudin, J.J., Henaut, A., and Danchin, A. (2001) Extracting biological information from DNA arrays: an unexpected link between arginine and methionine metabolism in *Bacillus subtilis*. *Genome Biol* **2**: RESEARCH0019.
- Sekowska, A., Denervaud, V., Ashida, H., Michoud, K., Haas, D., Yokota, A., and Danchin, A. (2004) Bacterial variations on the methionine salvage pathway. *BMC Microbiol* **4**: 9.
- Van Schaftingen, E., Rzem, R., Marbaix, A., Collard, F., Veiga-da-Cunha, M., and Linster, C.L. (2013) Metabolite proofreading, a neglected aspect of intermediary metabolism. *J Inherit Metab Dis* **36**: 427–434.
- Verna, D. (2011) Biological realms in computer science. In *10th SIGPLAN Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*. D'Hondt, T. (ed.). Portland, OR, USA: ACM, pp. 167–176.
- Werle, M., and Bernkop-Schnurch, A. (2006) Strategies to improve plasma half life time of peptide and protein drugs. *Amino Acids* **30**: 351–367.
- Westheimer, F.H. (1987) Why nature chose phosphates. *Science* **235**: 1173–1178.
- Wickramasinghe, R.H. (1973) Iron-sulphur proteins: their possible place in the origin of life and the development of early metabolic systems. *Space Life Sci* **4**: 341–352.
- Yip, S.H., and Matsumura, I. (2013) Substrate ambiguous enzymes within the *Escherichia coli* proteome offer different evolutionary solutions to the same problem. *Mol Biol Evol*. doi: 10.1093/molbev/mst105 (in press).
- Zhu, Y., and Lin, E.C. (1987) Loss of aldehyde dehydrogenase in an *Escherichia coli* mutant selected for growth on the rare sugar L-galactose. *J Bacteriol* **169**: 785–789.