Structural insights into the evolution of the pantothenate-biosynthesis pathway

C.M.C. Lobley*, F. Schmitzberger*, M.L. Kilkenny†, H. Whitney†, H.H. Ottenhof†, E. Chakauya†, M.E. Webb‡, L.M. Birch‡, K.L. Tuck‡, C. Abell‡, A.G. Smith† and T.L. Blundell*

*Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, U.K., †Department of Plant Sciences, Downing Street, Cambridge CB2 3EA, U.K., and ‡University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

Abstract

Pantothenate is synthesized in bacteria, fungi and plants, and as vitamin B5 is a dietary requirement in animals. The three-dimensional structures of the four Escherichia coli enzymes involved in the production of pantothenate have been determined. We describe the use of comparative analyses of the sequences and structures to identify distant homologues of the four enzymes in an attempt to understand the evolution of the pathway. We conclude that it is likely to have evolved via a patchwork mechanism, whereby the individual enzymes were recruited separately.

Introduction

Pantothenate is a necessary precursor to coenzyme A and phosphopantetheine, the prosthetic group of the acyl carrier protein, both of which are vital to a multitude of metabolic processes [1]. Also known as vitamin B5, it is a dietary requirement in animals. However, pantothenate is synthesized by bacteria, fungi and plants. The enzymes that are required to produce pantothenate in Escherichia coli have been extensively studied by X-ray crystallography ([3–5], and F. von Delft, T. Inoue, A.S. Saldanha, H.H. Ottenhof, F. Schmitzberger, L.M. Birch, V. Dhanaraj, M. Witt, C. Abell, A.G. Smith and T.L. Blundell, unpublished work) and kinetic analysis [6–8], and are attractive anti-microbial, fungicide and herbicide targets.

In E. coli, four enzymes are involved in the production of pantothenate (Scheme 1). Ketopantoate hydroxymethyltransferase (KPHMT; EC 2.1.2.11, SwissProt P31057) converts α-ketosovalerate into ketopantoate using 5,10-methylene tetrahydrofolate. Subsequently ketopantoate is reduced to pantoate by ketopantoate reductase (KPR; EC 1.1.1.169, SwissProt P77728) using NADPH as the hydrogen donor. Concomitantly l-aspartate is converted into β-alanine by aspartate decarboxylase (ADC; EC 4.1.1.11, SwissProt P31664). The ATP-consuming condensation of β-alanine and pantoate catalysed by pantoate synthetase (PS; EC 6.3.2.1, SwissProt P31663) completes the pathway. This simple, well-studied pathway is ideal for the investigation of evolution not only of the enzymes individually but also of the pathway as a whole.

To date, all theories of enzyme-pathway evolution have assumed that the first organisms could survive solely on the prebiotic soup – the heterotrophic origin of life. The possible chemical compositions of the prebiotic soup and early atmosphere are the subject of some debate, but the potential availability of pantothenate and its precursors has been shown experimentally [9].

The Horowitz hypothesis (retrograde evolution) assumes that when a metabolite from the prebiotic soup is exhausted, existing enzymes mutate to give a new enzyme that can catalyse the synthesis of the necessary metabolite. When that intermediate becomes limiting the first enzyme will again mutate and a pathway will begin to emerge to produce the required metabolite. If this were the case, each enzyme in a pathway would be homologous [10]. This appears to have happened in the tryptophan-biosynthesis pathway, where three sequential enzymes have similar structures and active sites. All three proteins are members of the same superfamily in the SCOP (Structural Classification of Proteins) database [11], implying a common ancestor.

The patchwork hypothesis, proposed by Ycas in 1974 [12] and adapted by Jenson in 1976 [13], relies on the existence of a pool of enzymes with low specificity. These enzymes were then recruited to a pathway as required, and the specificity was honed to give the highly specific enzymes we see today. This cannot apply to the very first pathways, since it requires the emergence of protein biosynthesis [12,13].

These classical views of evolving pathways were complemented by theoretical models of gene duplication, pioneered by Susumu Ohno [14], in which functionally redundant paralogous genes were assumed to be produced by whole-genome duplication, non-homologous recombination or through transpositions, thereby freeing one from selective constraints and allowing rapid evolution, initially by neutral evolution. Such redundancy will normally be eliminated by mutations causing one copy to become non-functional, but

Key words: enzyme pathway, evolution, pantothenate, X-ray crystallography.

Abbreviations used: KPHMT, ketopantoate hydroxymethyltransferase; KPR, ketopantoate reductase; ADC, aspartate decarboxylase; PS, pantothenate synthetase; P4M, phosphoethanolamine mutase; IL, isocitrate lyase; CENDH, N-(1-carboxyethyl)-L-ornvaline dehydrogenase; AHR, acetohydroxyisomerase; G3PCT, glucose-3-phosphate cytidylyltransferase.

*To whom correspondence should be addressed (e-mail carina@cryst.bioc.cam.ac.uk).
it can sometimes give rise to selectively advantageous new functions. These ideas have recently been tested by analyses of newly available genome sequences [15–17]. All agree that gene conversion is a rare event other than in multigene families. The analyses show less support for neutral or relaxed evolution than had been expected from the Ohno model, although there is evidence for some relaxation of constraints in the very early phases of divergence [15]. Two hypotheses have been presented to explain the data. The subfunctionalization hypothesis assumes that, as natural selection does not 'know' which gene is under natural selection, both accumulate mutations faster by selectively neutral evolution, some of which may be individually deleterious, so that after a time none of the paralogues is able to substitute for the ancestor [15]. An alternative hypothesis assumes there is initially 'gene sharing' so that the original gene encodes two functions, not dissimilar to the model of Ycas. Paralogues produced by gene duplication can then assume separate functions to the selective advantage of the organism. Kondrashov et al. [16] found evidence that two copies of the gene usually provide
Materials and methods

BLAST [27,28] and ClustalX [29] were used, respectively, to identify and align proteins homologous to the enzymes of the pantothenate biosynthesis pathway. The catalytic residues were identified from the X-ray structures and the conserved residues in the sequence alignments.

The structures of the four enzymes were compared with the proteins of both the PDB [22], using the Dali server (http://www.ebi.ac.uk/dali) [30], and the SCOP database (http://scop.mrc-lmb.cam.ac.uk/scop/) [11]. This enabled the identification of other proteins with the same fold, family or superfamily. Sequence-structure searches were carried out using the Fugue server (http://www-cryst.bioc.cam.ac.uk/~fugue/prfsearch.html) [31], 3D-PSSM (http://www.shg.bio.ic.ac.uk/) [32] and the Bioinbgui server (http://www.cs.bgu.ac.il/~bioinbgui/) [33] to identify distant homologues. Similar proteins once identified were aligned structurally using Comparer [26]. These alignments were annotated using JoY [34]. Structural superpositions were made using MNYFIT [35].

The conservation between the sequences was investigated, paying particular attention to the catalytic residues, by visual inspection of the multiple-sequence and structure alignments. The conservation of fold was assessed by inspection of the superposed co-ordinates of the structures.

Results

Occurrence of pan genes

A BLAST search shows that all four genes, panB, panC, panD and panE, are found in a diverse range of eubacteria. This is not true for archaea or eukaryotes. In archaea only the panB and panE genes have been identified. In plants and fungi panB and panC have been definitively identified, but panD is certainly not present. Sequences that are reminiscent of panE need further investigation to establish their identities.

KPHMT

KPHMT is a decamer with (βα)8 barrel protomers (F. von Delft, T. Inoue, A.S. Saldanha, H.H. Ottenhof, F. Schmitzberger, L.M. Birch, V. Dhanaraj, M. Witty, C. Abell, A.G. Smith and T.L. Blundell, unpublished work). The structure differs from the classic (βα)8 barrel by the presence of an additional N-terminal helix and the absence of the α-helix between strands β7 and β8. The active site, with co-ordinated Mg2+ ion, is located at the C-terminus of the barrel. The ubiquity and diversity of the (βα)8 barrels make their evolutionary relationships difficult to define, in particular with respect to distinguishing instances of convergent and divergent evolution. The most similar proteins, phosphoenolpyruvate mutase (PEPM; PDB code 1pym) and isocitrate lyase (ICL; PDB codes 1dq and 1igw), were analysed further (Table 1). Although the sequence identity between these proteins is relatively low, the structure-based sequence alignment shows a remarkable similarity in their secondary-structure elements and conservation of the active-site residues.

The active sites of these enzymes all involve the co-ordination of an Mg2+ ion and the sphere of co-ordination is largely conserved. Such exact duplication of the functional site implies homology between the proteins.

There are 24 superfamilies with the (βα)8 barrel fold in SCOP, each of which is thought to have a common ancestor. PEPM and ICL are classified in the phosphoenolpyruvate/ pyruvate superfamily in the SCOP database. Further members of this superfamily include 2-dehydro-3-deoxygalactarate aldolase from E. coli, phosphoenolpyruvate carboxylase from E. coli, pyruvate kinase from E. coli and pyruvate phosphate dikinase from Clostridium symbiosum. Inspection of the active sites of these enzymes suggests that the superfamily can be subdivided such that KPHMT, PEPM and ICL form one group and 2-dehydro-3-deoxygalactarate aldolase, pyruvate kinase, phosphoenolpyruvate carboxylase and pyruvate phosphate dikinase form a second group. ICL is involved in the glyoxylate cycle, which enables the conversion of acetyl-CoA into sugars. Since KPHMT is required in the
Table 1 | Sequence and structure database searches

The percentage sequence identity was taken from the JoY output and the Z scores were taken from each of the programs used. In Dali, proteins giving a Z score of less than 2.0 are considered to be structurally dissimilar. In Fugue, a Z score of greater than 6 means the proteins are certain to be related, scores greater than 4 show that the proteins are likely to be related, greater than 3.5 is marginal, greater than 2 is a guess and less than 2 is uncertain. N/A indicates a protein that was not identified using a particular method and therefore there is no related Z score. The results for other servers are not shown, since no extra data were gained from them. Abbreviations are explained in the text except: FDH-H, formate dehydrogenase H; FDH-N, formate dehydrogenase N (α-subunit); NAP, dissimilatory nitrate reductase; NSF, N-ethylmaleimide-sensitive fusion protein; VAT-N, vasolin-containing protein-like ATPase; P97, membrane fusion ATPase p97 N-terminal domain; DDGA, 2-dehydro-3-deoxy-galactarate aldolase; PPC, phosphoenolpyruvate carboxylase; PK, pyruvate kinase; PPDK, pyruvate phosphate dikinase; SCHAD, human short-chain 1-3-hydroxyacyl CoA dehydrogenase; AHIR, acetohydroxy isomeroreductase; 6-PGDH, 6-phosphogluconate dehydrogenase; UDPGDH, UDP-glucose dehydrogenase; G-3-PDH, glycerol-3-phosphate dehydrogenase; P97, membrane fusion ATPase p97 N-terminal domain; Nmn AT, nicotinamide mononucleotide adenylyltransferase; Y-tRS, E-tRS, W-tRS, M-tRS and R-tRS, tyrosine, glutamate, tryptophan, methionine and arginine tRNA synthetases, respectively.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Potential homologue</th>
<th>PDB code</th>
<th>Sequence identity (%)</th>
<th>Fugue Z score</th>
<th>Dali Z score</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPHMT</td>
<td>PEPM</td>
<td>1pym</td>
<td>13.5</td>
<td>14.47</td>
<td>17.1</td>
</tr>
<tr>
<td>ICL</td>
<td></td>
<td>1gwd</td>
<td>16.5</td>
<td>3.35</td>
<td>14.6</td>
</tr>
<tr>
<td>DDGA</td>
<td></td>
<td>1dxe</td>
<td>11.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PK</td>
<td></td>
<td>1e0t</td>
<td>16.1</td>
<td></td>
<td>11.8</td>
</tr>
<tr>
<td>PPDK</td>
<td></td>
<td>1kbl</td>
<td>10.6</td>
<td></td>
<td>9.9</td>
</tr>
<tr>
<td>PPC</td>
<td></td>
<td>1gb4</td>
<td>12.9</td>
<td></td>
<td>6.2</td>
</tr>
<tr>
<td>KPR (N-terminal domain)</td>
<td>CENDH</td>
<td>1bg6</td>
<td>15</td>
<td>19.09</td>
<td>17.1</td>
</tr>
<tr>
<td></td>
<td>SCHAD</td>
<td>2dhb</td>
<td>11</td>
<td></td>
<td>14.7</td>
</tr>
<tr>
<td></td>
<td>6-PGDH</td>
<td>2pgd</td>
<td>13</td>
<td>18.42</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>UDPGDH</td>
<td>1dli</td>
<td>15</td>
<td>12.90</td>
<td>13.1</td>
</tr>
<tr>
<td></td>
<td>AHIR</td>
<td>1qmg</td>
<td>16</td>
<td></td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>G-3-PDH</td>
<td>1vey</td>
<td>17</td>
<td>21.61</td>
<td>12.2</td>
</tr>
<tr>
<td>KPR (C-terminal domain)</td>
<td>CENDH</td>
<td>1bg6</td>
<td>12</td>
<td>2.76</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>G-3-PDH</td>
<td>1evey</td>
<td>15</td>
<td></td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>6-PGDH</td>
<td>2pgd</td>
<td>15</td>
<td></td>
<td>2.8</td>
</tr>
<tr>
<td>ADC</td>
<td>DMSO reductase</td>
<td>1dmr</td>
<td>10.2</td>
<td></td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>FDH-H</td>
<td>1aa6</td>
<td>15.2</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>FDH-N</td>
<td>1kqf</td>
<td>8.1</td>
<td></td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>Trimethylamine N-oxide reductase</td>
<td>1tmo</td>
<td>10.3</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Arsenite oxidase</td>
<td>1q8k</td>
<td>10.9</td>
<td></td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>NAP</td>
<td>2nap</td>
<td>8.1</td>
<td></td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>NSF</td>
<td>1qcs</td>
<td>5.3</td>
<td></td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>VAT-N</td>
<td>1cz4</td>
<td>16.5</td>
<td>3.79</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td>P97</td>
<td>1e32</td>
<td>11.2</td>
<td>3.64</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>Barwin</td>
<td>1bw4</td>
<td>9.9</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td>PS (N-terminal domain)</td>
<td>Y-IRS</td>
<td>2ts1</td>
<td>7.5</td>
<td>9.3</td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td>G3PCT</td>
<td>1c0z</td>
<td>20.2</td>
<td>8.8</td>
<td>6.03</td>
</tr>
<tr>
<td></td>
<td>PPAT</td>
<td>1b6t</td>
<td>18.9</td>
<td>8.6</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>E-IRS</td>
<td>1gln</td>
<td>11.5</td>
<td>8.5</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>ATP sulphurylase</td>
<td>1g8f</td>
<td>13.9</td>
<td>8.5</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Nmn AT</td>
<td>1f9a</td>
<td>15.2</td>
<td></td>
<td>4.53</td>
</tr>
<tr>
<td></td>
<td>W-IRS</td>
<td>1djr</td>
<td>7.5</td>
<td></td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td>Y-IRS</td>
<td>1h3f</td>
<td>13.3</td>
<td></td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td>M-IRS</td>
<td>1a8h</td>
<td>9.7</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>R-IRS</td>
<td>1bs2</td>
<td>11.5</td>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>

Production of CoA, it is reasonable to propose that it evolved before ICL. The second group within the superfamily is most closely related to PEPM, so PEPM is proposed to be evolutionarily younger than ICL or KPHMT.

KPR

KPR is a monomer, with two clearly defined domains. The N-terminal domain (residues 1–167) is based on the Rossmann fold, with five β-strands and three α-helices. The C-terminal
Figure 1 | JoY annotated sequence alignments

(a) E. coli KPHMT with sequence homologues and a consensus sequence followed by the structural homologues. Active-site residues that are strictly conserved between the sequences are shaded yellow. (b) N-terminal domain of E. coli KPR and its structural homologues. Boxed residues are conserved throughout the KPRs. (c) E. coli ADC. Boxed residues are conserved throughout the ADCs. The sequence alignment shown only shows those residues that the SCOP database indicates make up the fold. (d) N-terminal domain of E. coli PS and its homologues. The HIGH and KMSKS motifs have been boxed in all sequences, and the residues strictly conserved in all PS enzymes are boxed in 1iho. // indicates that an insertion in that sequence has been removed. In the tRNA synthetases the tRNA-binding domain, co-incident with the PS dimerization domain, has been removed. The key to JoY format can be found at: http://www-cryst.bioc.cam.ac.uk/~joy/joyman.htm.

DDGA, 2-dehydro-3-deoxy-galactarate aldolase.
domain (168–303) has been described as family-specific and is entirely α-helical [5]. In SCOP the two domains have been classified separately, with the N-terminal domain being a member of the NAD(P)-binding Rossmann-fold-domain superfamily and the 6-phosphogluconate dehydrogenase-like N-terminal-domain family, and the C-terminal domain being a member of the 6-phosphoglucuronate dehydrogenase-like C-terminal-domain superfamily and the single-member KPR family.

A Dali search was carried out using the N- and C-termini separately and the sequences submitted for sequence-structure alignment. In all searches a large number
of proteins were identified with a structure similar to the N-terminal domain, including N-(1-D-carboxyethyl)-L-norvaline dehydrogenase (CENDH; PDB code 1bg6), human short-chain 1-3-hydroxyacyl-CoA dehydrogenase (2hdh), 6-phosphogluconate dehydrogenase (2pgd), UDP-glucose dehydrogenase (1dli), acetohydroxy isomeroreductase (AHIR; 1qmg) and glycerol-3-phosphate dehydrogenase (1evy). In contrast to this, only three proteins, CENDH, glycerol-3-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase, were found with any significant similarity in the C-terminal domain (Table 1).

From the alignments and structural superpositions, it was immediately apparent that the N-terminal domain has much greater structural similarity than the C-terminal domain (Figures 1 and 2). Interestingly, AHIR, a protein capable of catalysing the reduction of ketopantoate, is found only in the Dali search for the N-terminal domain. The difference in the C-terminal domain fold might be the cause of the lower reaction efficiency seen in AHIR. Putative homologues of KPR are all dehydrogenases, requiring NAD(P)(H) as a cofactor.

However, of the predicted catalytic residues in KPR, only the Gly-Xaa-Gly-Xaa-Xaa-Gly motif in the Rossmann fold is conserved. The complete lack of conservation in the active-site residues suggests that these enzymes diverged a considerable time ago.

**ADC**

ADC has been identified only in eubacteria, including pathogenic species such as *Mycobacterium tuberculosis* and *Yersinia pestis* (results not shown). It is tetrameric and only active after an autoprocessing reaction occurs between Gly-24

---

**Figure 2** | Structural superpositions

(a) KPR in blue (1KS9) and CENDH in red (1BG6) (43). (b) PS in green (1iho) (3) and G3PCT in orange (1coz) (40). (c) ADC in red (1AW8) and its structural homologues in blue. The image was made using Prekin and Mage and is in wide-eyed stereo (44).
and Ser-25 to produce a covalently attached pyruvoyl group to which L-aspartate is then bound [4]. The initial BLAST alignment of ADC shows a strong conservation of the residues implicated in the mechanisms of auto-processing and catalysis.

In 1999, Castillo et al. [36] observed that a six-stranded double-$\psi$ $\beta$-barrel was shared by a large group of proteins. Subsequently this group of proteins has been defined as the ADC-like superfamily in SCOP. This superfamily is divided into three families: the pyruvoyl-dependent ADC family has only one member, ADC; the formate dehydrogenase/DMSO reductase C-terminal-domain family has seven members; and the Cdc48 N-terminal-domain-like family has a further four members. The double-$\psi$ $\beta$-barrel is found as a circular permutation in aspartic proteinases [36].

From sequence and structure searches, a number of these proteins were identified, although no server detected all members of the superfamily, probably due to the paucity of sequence identity between the homologues (Table 1). From the sequence alignment and superposed models, it is obvious that the fold is well conserved (Figures 1 and 2), although none of the active-site residues is. The other enzymes in the superfamily do not use auto-processing, so the mechanism is also not conserved.

There are several other pathways, including the degradation of uracil, that produce $\beta$-alanine, so that the loss of functional ADC would not be detrimental to the organism. This implies that ADC, and possibly the pantothenate pathway, evolved before other pathways that produce $\beta$-alanine.

**PS**

Alignment of sequences of PS indicates that most of the conserved residues are in the active site (results not shown). Among the conserved residues were His-Xaa-Gly-His and Xaa-Xaa-Ser-Arg from the HIGH motif [37] and the less well-conserved KMSKS motif respectively [38,39]. These motifs are characteristic of class 1 aminoacyl-tRNA synthetases, indicating that PS and glucose-3-phosphate cytidyl transferase (G3PCT; 1coz) are homologues [3,40].

PS is composed of an N-terminal domain (1–176) containing a Rossmann fold and the dimerization domain, and a C-terminal domain (177–283), which is enzyme specific [3]. Sequence and structure searches identified homologues of the N-terminal domain, including several tRNA synthetases (tyrosine tRNA synthetase, 2ts1; glutamate tRNA synthetase, 1gln; tryptophan tRNA synthetase, 1d2r; methionine tRNA synthetase, 1a8h; arginine tRNA synthetase, 1bs2), phosphopantetheine adenyltransferase (1b6t) and G3PCT; and these were aligned structurally using Comparer. All are members of the nucleotidyl transferase superfamily in the SCOP database.

Again, even though there is low sequence identity, the structural superposition is striking, particularly in the core regions (Figures 1 and 2). Phosphopantetheine adenyllyltransferase and G3PCT are the ‘pure’ members of the family, with the minimum required structural elements. PS has an additional dimerization domain inserted between $\beta$3 and $\alpha$4, and the tRNA synthetases have a 3’-tRNA-binding domain inserted in a similar position. No significant results were found for the C-terminal domain.

It is reasonable to propose the tRNA synthetases were the earliest members of this family of proteins, since it is likely that RNA came before proteins, so the earliest proteins must have been those involved in converting RNA to proteins. It has also been shown that pantothenate could have been produced prebiotically [9], whereas the aminoacyl tRNAs are not. As PS contains insertions in the same region as the tRNA synthetases, it seems likely PS evolved directly from the tRNA synthetases.

**Conclusions**

From the comparative sequence and structural analysis carried out for each enzyme, we propose that the pantothenate pathway evolved via recruitment of enzymes in the manner of the patchwork theory. We have seen that in three of the four enzymes, namely KPR, ADC and PS, there is no conservation of active-site residues between the homologues, suggesting that the pathway was early to evolve, and the enzymes have almost completely lost sequence identity while retaining similar structures. Since the intermediates of the pathway are not used in other pathways, and the four enzymes are not evolutionarily related, the Horowitz hypothesis does not apply.

No evidence has been found that a pathway similar to the pantothenate pathway evolved first and was borrowed ‘en masse’ when pantothenate synthesis was required.

It is a reasonable assumption that the pantothenate pathway evolved early in the history of life, since pantothenate is essential for CoA and for the prosthetic group of acyl carrier protein. Since these substances are central to primary metabolism, when the prebiotic soup was exhausted of them, their synthesis would have become vital.

Initially four enzymes were required for the synthesis of pantothenate in all organisms. Since the pathway is not tightly regulated, the evolution of a transporter protein enabled the secretion of pantothenate into the environment [41]. Subsequently, as the transporter is reversible, many higher organisms have come to use pantothenate from the environment and have lost the ability to synthesize pantothenate. Between the two extremes we have the archaea with only two proteins from the pathway, and plants and fungi with two or three proteins from the pathway. These are likely to synthesize pantothenate, taking intermediates from other, more recently evolved, pathways.

By detailed evolutionary study of simple enzyme pathways such as this, we begin to understand how organisms evolved, giving us the complex biochemistry we see today.

---

We thank the Biotechnology and Biological Sciences Research Council (BBSRC) for funding of the structural studies of the pantothenate biosynthesis enzymes. C.M.C.L., H.W., H.H.O. and K.L.T.
thank the BBSRC for studentships. H.H.O. is also supported by Aventis Crop Science (U.K.). M.L.K. thanks the CCF, ORS, University of Cape Town and the Skye Foundation for funding. F.S. is funded by the Cambridge European Trust and the University of Vienna. E.C. is grateful to the CCT for a studentship. M.E.W. thanks the Engineering and Physical Sciences Research Council for a studentship. L.M.B. and M.E.W. thank Astex Technology for funding and support.

References

2 Reference deleted
34 Minguez, S., Deane, C.M., Blundell, T.L. and Overington, J.P. (1998) Bioinformatics 14, 617-622
44 Richardson, D.C. and Richardson, J.S. (1992) Protein Sci. 1, 3-9

Received 16 December 2002