J. Basic Microbiol. **42** (2002) 2, 91 – 103

([1]Duke University, Department of Biology, Durham, NC, USA 27708; [2]ETH Zürich, Institut für Mikrobiologie, Schmelzbergstr. 7, CH-8092 Zürich, Switzerland; [3]Georg-August-Universität Göttingen, Institut für Forstbotanik, Büsgenweg 2, D-37077 Göttingen, Germany; [4]Present address: National Institutes of Health, Building 10/8N306, 9000 Rockville Pike, Bethesda, MD, USA 20892-1766; [5]Present address: Universität Freiburg, Plant Biotechnology, Sonnenstr. 5, D-79140 Freiburg, Germany)

The *pab1* gene of *Coprinus cinereus* encodes a bifunctional protein for *para*-aminobenzoic acid (PABA) synthesis: implications for the evolution of fused PABA synthases

TIMOTHY Y. JAMES[1], ROBERT P. BOULIANNE[2, 4], ALAN P. F. BOTTOLI [2], JOSÉ D. GRANADO[2, 5], MARKUS AEBI[2] and URSULA KÜES[2, 3]*

(Received 15 November 2001/Accepted 07 December 2001)

The *pab1* gene of the basidiomycete *Coprinus cinereus* encodes PABA synthase, necessary for *para*-aminobenzoic acid production. The *C. cinereus* protein is bifunctional with an N-terminal glutamine amidotransferase domain and a C-terminal chorismate amination domain. In most bacteria, these two functions are encoded in separate genes (e.g., *pabA* and *pabB* of *E. coli*). Fused PABA synthases have so far been detected in actinomycetes, *Plasmodium falciparum*, fungi and *Arabidopsis thaliana*. Phylogenetic analysis shows that the fused PAB sequences form a tight group that also includes uncharacterized PabB homologues from several bacteria. Unfused bacterial PabA proteins group with the glutamine amidotransferase subunits of bacterial anthranilate synthases, independent of organismal systematics, indicating a complex and perhaps independent evolutionary origin. In contrast, unfused PabB group and fused PabA/B proteins form a monophyletic group on a branch separate from the chorismate amination subunits of anthranilate synthases, probably reflecting a need for recognition of different positions in the common substrate chorismate.

PABA (*para*-aminobenzoic acid) is a substrate of 7,8-dihydropteroate synthase (DHPS) in the de novo biosynthesis of folate (GREEN *et al.* 1996, VINNICOMBE and DERRICK 1999), important for the formation of purines, thymidylate, serine, methionine, glycine and formylmethionyl-tRNA (COSSINS and CHEN 1997). Bacteria, the protist *Plasmodium falciparum*, fungi, and plants are able to synthesize both PABA and folate in contrast to animals (HERRMANN 1995, COSSINS 2000).

PABA is generated in a two step reaction: (1) Chorismate + glutamine → ADC (4-amino-4-deoxychorismate) + glutamate and (2) ADC → PABA + pyruvate. In *Escherichia coli*, the products of the three genes *pabA*, *pabB* and *pabC* have been shown to catalyze these reactions. PabA and PabB associate with one another to form the ADC synthase (Step 1), commonly called PABA synthase. PabA (PABA synthase component I) is a PabB dependent glutaminase (glutamine amidotransferase) that generates ammonia from glutamine. The aminodeoxychorismate synthase or chorismate aminase PabB (PABA synthase component II) then uses the ammonia to aminate chorismate, producing ADC. Separately of PabA and PabB, the ADC lyase PabC finally generates PABA from ADC with the release of pyruvate (Step 2; GREEN *et al.* 1996). In most bacteria *pabA* and *pabB* are isolated genes, but *pabA* and *pabB* homologues are found as one fused gene in a number of actinomycetes (e.g., *Streptomyces griseus*; CRIADO *et al.* 1993). Likewise, all of the eukaryotes analyzed so far:

* Corresponding author Prof. Dr. KÜES; e-mail: ukuees@gwdg.de

WILEY-VCH Verlag Berlin GmbH, 13086 Berlin, 2002 0233-111X/02/0205-0091 $ 17.50+.50/0

the protist *Plasmodium falciparum* (TRIGLIA and COWMAN 1999), the ascomycetous fungi *Aspergillus fumigatus* (BROWN *et al.* 2000), *Saccharomyces cerevisiae* (EDMAN *et al.* 1993) and *Schizosaccharomyces pombe* (GenBank T40823), and the higher plant *Arabidopsis thaliana* (GenBank AAC79592) possess a single gene for ADC formation, whose 5′ and 3′ half are homologous to *E. coli pabA* and *pabB*, respectively.

Chorismate is also utilized in tryptophan synthesis in bacterial and eukaryotic microorganisms, and plants (HÜTTER *et al.* 1986, NICHOLS 1996). The first product in this pathway, anthranilate = *ortho*-aminobenzoic acid, is formed when chorismate is aminated to give ADIC (2-amino-4-deoxychorismate) and then aromatized in a two-step reaction similar to that leading to PABA. In bacteria, TrpE (anthranilate synthase component I = AS COI, a chorismate aminase), TrpG (anthranilate synthase component II = AS COII, a glutaminase or glutamine amidotransferase) and TrpD (the ADIC lyase) perform these reactions. TrpE, TrpG and TrpD are functional paralogues of PabB, PabA and PabC, respectively (NICHOLS 1996). *trpE*, *trpG* and their eukaryotic homologues are typically independent genes, but occasionally the genes are found fused with each other as in *Rhizobium meliloti, Deinococcus radiodurans*, and *Azospirillum brasiliense* (BAE *et al.* 1989, DE TROCH *et al.* 1997, GenBank F75547). In other bacteria, *trpG* is fused to *trpD* (*E. coli*, *Salmonella typhimurium*, *Thermatoga maritime*; Nichols 1996). In fungi, *trpG* is fused with *trpC* or often to *trpF*, other genes acting in the tryptophan biosynthesis pathway (SCHECHTMAN and YANOFSKY 1983, KOS *et al.* 1985). Remarkably, *Acinetobacter calcoaceticus*, *Pseudomonas acidovorans* and *Bacillus subtilis* use the same amphibolic glutamine amidotransferase in the synthesis of both anthranilate and PABA (BUVINGER *et al.* 1981, KAPLAN *et al.* 1984, SLOCK *et al.* 1990).

In this paper, we report the *Coprinus cinereus* Pab1 sequence, the first PABA synthase from a basidiomycetous fungus. Like other eukaryotes, *C. cinereus* Pab1 is bifunctional, containing glutamine amidotransferase (PabA) and chorismate aminase (PabB) domains. The apparently universal fusion of *pabA* and *pabB* genes in eukaryotes prompted us to test the hypothesis that bifunctional PABA synthases have arisen from a single evolutionary event. We analyze phylogenetic relationships between Pab1 of *C. cinereus* with eukaryotic and bacterial PABA and anthranilate synthases.

Materials and methods

***Coprinus* strains, plasmids and transformation:** Monokaryon PG78 (*A6 B42 trp1.1,1.6 pab1*) and homokaryon AmutBmut (*A43mut B43mut pab1*) were used in transformation according to the protocol of GRANADO *et al.* (1997). Transformation mixtures were plated on regeneration medium (GRANADO *et al.* 1997) and incubated at 37 °C. Tryptophan (100 mg/l) was added as needed. Plasmid pST17, a pUC13-derivative kindly supplied by L. A. CASSELTON, contains the *pab1*$^{+}$ gene from *C. cinereus* monokaryon JV6 on a 5.6 kb genomic *Pst*I fragment (MUTASA *et al.* 1990). A 3.8 kb *Nru*I-*Bam*HI subfragment was subcloned into the *Hin*dII and *Bam*HI sites of pTZ19R and pTZ18R (both PHARMACIA) to give pPAB1-1 and pPAB1-2, respectively (this study, GRANADO *et al.* 1997). All DNA manipulations were performed following standard protocols (SAMBROOK *et al.* 1989). Plasmids were propagated in *E. coli* XL1-Blue (Stratagene).

Sequence and phylogenetic analysis: Sequencing of the *C. cinereus pab1*$^{+}$ gene was carried out with ABI PRISM terminator reactions, and DNA subfragments cloned in pTZ18R or pTZ19R were sequenced using either M13 universal and reverse primers or custom made primers. Sequencing reactions were analysed with an ABI PRISM 310 genetic analyzer. Because Pab1 from *C. cinereus* represents the fusion of two prokaryotic homologues, two separate protein sequence alignments were used in phylogenetic analysis. One alignment consisted of PabA, TrpG, and the homologous regions of fused proteins. This alignment spanned amino acids 7–211 of *C. cinereus* Pab1. The second alignment was for the regions 454–735 of *C. cinereus* Pab1 and aligned the PabB and TrpE homologues or homologous regions. Alignments were constructed using the program ClustalX with

gap opening and gap extension penalties = 10.0 and the BLOSUM 30 protein weight matrix (THOMPSON *et al.* 1997). Alignments were adjusted manually using GeneDoc (NICHOLAS *et al.* 1997), and regions that were considered unalignable were discarded from phylogenetic analysis, leaving 132 aligned amino acids for the PabA alignment and 240 residues for the PabB alignment. Phylogenetic reconstruction used the PHYLIP software package (FELSENSTEIN 1993). Trees were created using the neighbor-joining algorithm applied to a pairwise distance matrix generated using the categories model and the HALL characterization of amino acids. Statistical support for groupings was estimated using 100 bootstrap replicates. The sequences used in the analysis and their GenBank accession numbers are as follows.

AS COI genes: *Aquifex aeolicus* (O66849), *A. thaliana* (JQ1685), *Archaeoglobus fulgidus* (O28669), *Arthrobacter globiformis* (P96556), *Azospirillum brasilense* (P26922), *Bacillus pumilis* (P18267), *B. subtilis* (P03963), *Bordetella bronchiseptica* (AAF76162), *Buchnera aphidicola* (Q9ZER9), *Campylobacter jejuni* (CAB74182), *Clostridium thermocellum* (P14953), *Corynebacterium glutamicum* (P06557), *E. coli* (CAA23666), *Haloferax volcanii* (P33975), *Helicobacter pylori* (Q9ZJU5), *Lactococcus lactis* (Q02001), *Leptospira biflexa* (P20463), *Methanococcus jannaschii* (Q58475), *Mycobacterium tuberculosis* (O06127), *Neisseria meningitides* (Q9XAZO), *Nicotiana tabacum* (7D1990), *Oryza sativa* (BAA82095), *Pseudomonas aeruginosa* (P20580), *Pyrococcus abysii* (A75163), *Ruta graveolens* (AAA74900), *S. cerevisiae* (P00899), *S. typhimurium* (A39812), *S. pombe* (T40974), *Spirochaeta aurantia* (P21690), *Streptomyces coelicolor* (T35072), *Sulfobolus solfataricus* (Q06128), *Thermococcus kodakaraensis* (Q9XGB3), *T. maritima* (Q08653), *Thermus thermophilus* (P05378).

Fused *AS COI-AS COII* genes: *A. brasilense* (P50872), *D. radiodurans* (F75547), *R. melloti* (P15395).

AS COII genes: *A. aeolicus* (E70349), *A. thaliana* (J02340), *A. fulgidus* (Q28670), *Brevibacterium lactofermentum* (P06558), *B. aphidicola* (AAC31216), *Cyanidium caldarium* chloroplast (T11982), *Cyanophora paradoxa* cyanelle (P48261), *Haemophilus influenzae* (P71381), *H. volcanii* (P33974), *H. pylori* (F71836), *L. lactis* (Q02003), *L. biflexa* (P20441), *Methanobacterium thermoautotrophicum* (O27693), *M. jannaschii* (Q57690), *Nitrosomonas europaea* (BAA83385), *Porphyra purpurea* chloroplast (P51362), *P. aeruginosa* (P20576), *Pseudomonas putida* (P00901), *Pyrococcus abyssi* (H75162), *Rhodobacter sphaeroides* (AAD09117), *S. typhimurium* (P00905), *Serratia marcescens* (P00900), *Shigella dysenteriae* (P00906), *S. coelicolor* (T36305), *S. solfataricus* (Q06129), *Thermus thermophilus* (P05379), *Vibrio parahaemolyticus* (P22101).

Other fused *AS COII* genes: *AS COII-trpC*: *S. cerevisiae* (AAA35176). *AS COII-trpC-trpF*: *Aspergillus niger* (P18483), *Cochliobolus heterostrophus* (Q92411), *Neurospora crassa* (P00908), *Penicillium chrysogenum* (P24773), *Phanerochaete chrysosporium* (P25170), *Phycomyces blakesleeanus* (P20409), *Pichia angusta* (P09575), *S. pombe* (T39468). *AS COII-trpD*: *E. coli* (P00904), *T. maritima* (CAA52203).

pabA genes: *E. coli* (P00903), *Klebsiella aerogenes* (P06194), *Mycobacterium leprae* (T10008), *Mycobacterium tuberculosis* (C70699), *S. typhimurium* (P06193), *Serratia marcescens* (P06193), *Streptomyces lividans* (P27627), *Vibrio cholerae* (AAF95760).

Fused *pabA-pabB* genes: *A. thaliana* (AAC79592), *A. fumigatus* (AAD31929), *C. cinereus* (AAF89583), *P. falciparum* (AAD38122), *S. cerevisiae* (P37254), *S. pombe* (T40823), *S. griseus* (P32483), *Streptomyces pristinaespiralis* (AAC44866), *Streptomyces venezuelae* (AAB30312).

pabB genes: *A. aeolicus* (F70398), *B. subtilis* (P28820), *E. coli* (P05041), *H. influenzae* (F64187), *K. aerogenes* (P12679), *Lactococcus lactis* (P27029), *S. typhimurium* (P12680), *S. lividans* (P27630), *V. cholerae* (AAF94462).

Other *pabB* genes (*pabB-pabC?*): *Campylobacter jejuni* (CAB73127), *H. pylori* (A71951).

Results

Characterization of the C. cinereus pab1 gene

In order to map the location of the *pab1* gene on plasmid pST17 as closely as possible, pST17 was individually digested with various restriction enzymes known to cut in the chromosomal DNA insert and subsequently transformed into *pab1*$^-$ strain PG78. Cleavage of pST17 with enzymes *Pst*I, *Nru*I, *Mlu*I, and *Bam*HI did not have a negative effect on the

transformation rate of PG78 to PABA prototrophy, suggesting that the respective restriction sites were not located within the *pab1* gene. In contrast, digestion with either *Nco*I, *Mun*I, *Ssp*I, *Nar*I, *Pvu*II, *Bgl*II, *Hin*dIII and *Sma*I clearly reduced the number of transformants relative to the uncut control (Fig. 1). Therefore, we suspected the gene to reside within the 3.8 kb-sized *Nru*I-*Bam*HI fragment. Constructs pPAB1-1 and pPAB1-2 containing this sequence conferred PABA prototrophy to monokaryon PG78 and to homokaryon Amut-Bmut (GRANADO *et al.* 1997, this study) confirming that the *pab1* gene is present within the subcloned fragment. The complete sequence of the fragment (GenBank accession number AF166096) was found to be 3829 bp in length. The *pab1*-coding region, deduced from protein alignments with PABA synthases from other organisms (Fig. 2), extends from nucleotides 365 to 2724, interrupted by two short introns at positions 424–471 and 2335–2387. The length of both the introns (48 and 53 bp, respectively) are typical for *C. cinereus*. Their 5′ splice junctions (intron 1: GTCAGT, intron 2: GTGTGA) and their 3′ splice junctions (intron 1: TAG, intron 2: CAG) match the *C. cinereus* consensus sequences (SEITZ *et al.* 1996, BOULIANNE *et al.* 2000). CTCA branch-acceptor sequences are present 14 bases and 20 bases upstream of the 3′ splice junction of intron 1 and intron 2, respectively.

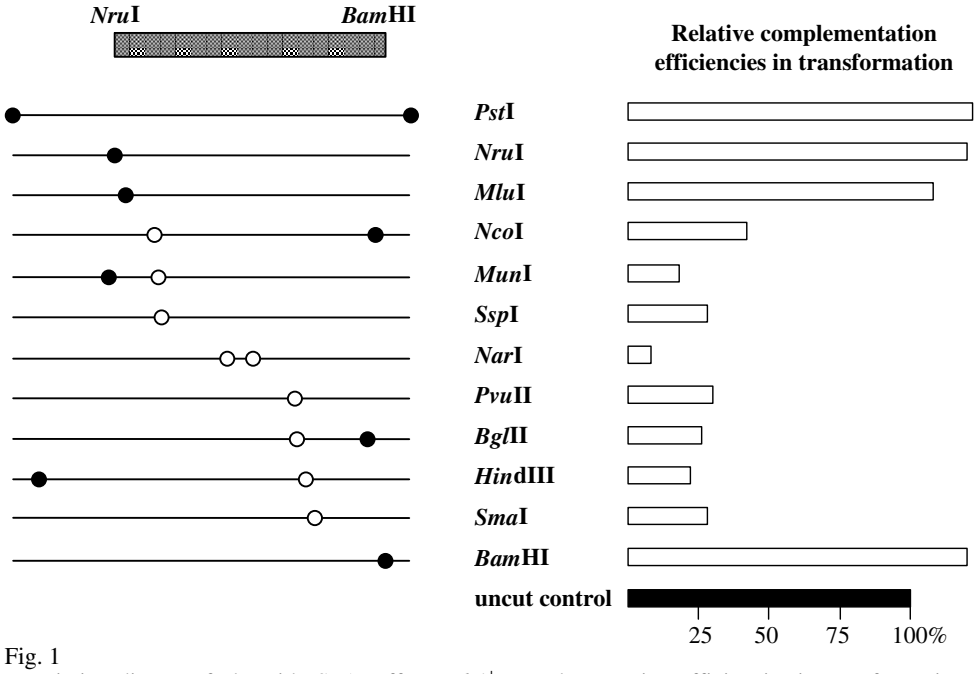


Fig. 1
Restriction digests of plasmid pST17 effect *pab1*$^+$ complementation efficiencies in transformation of the *pab1*-auxotrophic monokaryon PG78. Plasmid DNA were digested with the enzymes indicated, purified and aliquots of 1 µg transformed into strain PG78. From three independent series of transformations, the total numbers of transformants per DNA sample were determined and compared in percentage to that obtained in parallel experiments with 1 µg of uncut purified pST17 DNA (black bar = 100%; total numbers of transformants with uncut pST17 were 653, 382 and 164, respectively). The grey-shaded bar at the top left shows the 3.8 kb *Nru*I-*Bam*HI fragment used to construct pPAB1-1 and pPAB1-2. The lines below represent the 5.6 kb *Pst*I *C. cinereus* fragment present in pST17 and the positions of mapped recognition sites for restriction enzymes. Open circles indicate those restriction sites that appear to interfere with complementation activity, filled circles those sites where restriction enzyme digestion seems not to interfere

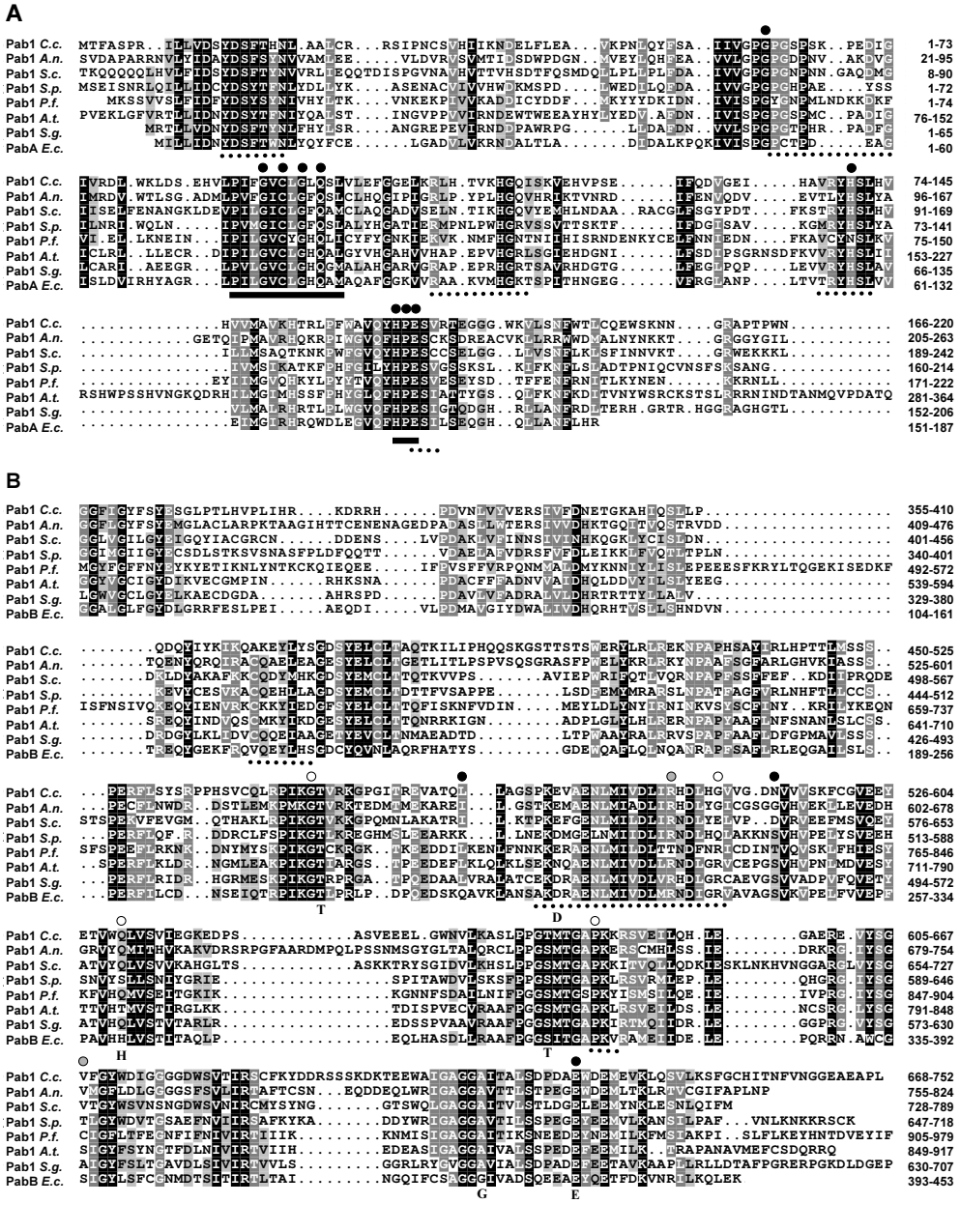
With the exception of *Sma*I, all restriction enzymes found to impede *pab1* function in the transformation assay (Fig. 1) have a recognition site within the deduced *pab1* coding region or in intron 1 (*Nco*I). *Sma*I cuts 65 bp downstream of the TGA stop codon and probably interferes with messenger termination. Terminator signals are not yet defined in *C. cinereus*. However, 64 bp downstream of the *Sma*I site is the motif TTTAAGCTTAGCAATTAAG-CAT followed by four potential poly(A) sites (CA di-nucleotides) within the next 24 bases. This motif resembles known poly(A) site positioning elements of the yeasts *S. pombe* (TTTGCTATAGTAATTT; HUMPHREY *et al.* 1994) and *S. cerevisiae* (AATAAA; TTAA-GAAC; GUO and SHERMAN 1996).

Restriction at the *Mlu*I site present at position 148 of the sequence did not interfere with gene function in transformation (Fig. 1). This site therefore defines the maximal length of the core promoter to 215 bp (Fig. 3). The *pab1* promoter fragment has 53% identity to the core promoter of the *C. cinereus trp1* gene for tryptophan synthase (maximal length: 321 bp; SKRZYNIA *et al.* 1989; Fig. 3), possibly reflecting common regulation of genes in the related metabolic pathways for PABA and tryptophan synthesis. Both core promoters are AT-rich and have no CAAT, TATA or other known regulatory nucleotide motifs (Fig. 3). The overall GC content of the *pab1* promoter region (47%) is low compared to the 52% GC content of the whole sequence, the 54% GC content of the *pab1* coding region and to the overall *C. cinereus* genomic GC content of 55% (WEBER *et al.* 1986).

The Pab1 protein of C. cinereus is bifunctional

The deduced protein product Pab1 is 752 amino acids long (Fig. 2). In BLAST searches, the whole protein shows highest identity to the bifunctional PABA synthases of the ascomycetous fungi *A. fumigatus*, *S. pombe* and *S. cerevisiae* (33–35% identity, 49–51% similarity) and of the plant *A. thaliana* (34% identity, 50% similarity). The *C. cinereus* protein is slightly more distantly related to the fused PABA synthases of *P. falciparum* and *Streptomyces* spp. (e.g., *S. griseus*), with 30–31% identity and 47–48% similarity. High similarity is usually found within the first 200 and last 400 amino acids of the proteins (Fig. 2). The region between these domains varies in length and degree of similarity between the different organisms suggesting a linker/spacer function for this region (not shown).

The conserved N-terminal regions of *C. cinereus* Pab1 and the fused PABA synthases of other organisms align with PabA of *E. coli* (35% identity, 53% similarity to *C. cinereus*) (Fig. 2A), and their conserved C-terminal regions align to PabB of *E. coli* (30% identity, 48% similarity to *C. cinereus*), (Fig. 2B). In accordance with a glutamine amidotransferase function, *C. cinereus* Pab1 has a GATASE motif (glutamine amidotransferase class-I active site; WENG and ZALKIN 1987), including an important cysteine (ROUX and WALSH 1993), at positions 88–99 (PIFGVCLGLQSL) in the N-terminal, PabA-homologous part of the protein (Fig. 2A). Moreover, all nine conserved amino acids from the glutamine domain present in the "triad"-family of amidotransferases (G^{59}-G^{84}-C^{86}-G^{88}-Q^{90}-H^{143}-H^{181}-P^{182}-E^{183}; numbers refer to amino acid positions in guanosine monophosphate synthetase of *E. coli*; ZALKIN and SMITH 1998) are found in Pab1 of *C. cinereus* (Fig. 2A). The amino acids corresponding to C^{86}, H^{181} and E^{183} have been shown in *E. coli* PabA to be required for catalytic activity (ROUX and WALSH 1993). Other strongly conserved regions in the N-terminal part of fused PABA synthases and *E. coli* PabA are also shared with the glutamine amidotransferase subunit of anthranilate synthases. Several of these sequences in *S. solfataricus* and *S. typhimurium* AS COII have recently been shown by crystallography to interact with their chorismate aminase partner protein AS COI (KNÖCHEL *et al.* 1999, MOROLLO and ECK 2001, Fig. 2A). Likewise, most of the sequences in the AS COI proteins interacting with the AS COII subunits (KNÖCHEL *et al.* 1999, MOROLLO and ECK 2001) are highly conserved in the C-termini of the fused PABA synthases and in PabB of *E. coli* (Fig. 2B). In contrast, only 2–4 of the 6 residues known to be catalytically important in *S. typhimurium* AS COI



(CALIGIURI and BAUERLE 1991, KNÖCHEL *et al.* 1999; Fig. 2B) are present within the PABA synthase sequences shown in Fig. 2B. One of the two amino acids conserved in all enzymes (an E close to the C-terminus) is also essential for PabB catalytic function in *E. coli*; the second (the T closest to the N-terminus) is adjacent to a residue conserved in PABA and anthranilate synthases that has been shown to be essential for PabB catalytic function in *E. coli* (Fig. 2B; RAYL *et al.* 1996). Most of the other residues that have been associated

◀ Fig. 2
Sequence alignments of the conserved N-terminal glutamine amidotransferase domain (A) and the conserved C-terminal chorismate amination domain (B) in fused PABA synthases to PabA and PabB of *E. coli*, respectively. (A) The conserved GATASE motif is underlined (WENG and ZALKIN 1987). The nine conserved amino acids from the glutamine domain present in "triad"-amidotransferases are indicated above the sequences with filled circles (ZALKIN and SMITH 1998). Dotted lines indicate regions that in the related AS COII proteins have been implicated in interaction with AS COI (KNÖCHEL *et al.* 1999, MOROLLO and ECK 2001). (B) Residues shown in PabB of *E. coli* to be essential for catalytic activity are marked by circles above the sequences. Open circles: class 1 residues needed for glutamine-dependent and NH_3-dependent reactions and interaction with PabA; black circles: class 2 residues needed for glutamine-dependent and NH_3-dependent reactions; grey-shaded circles: class 3 residues needed for glutamine-dependent reactions and interaction with PabA (RAYL *et al.* 1996). Amino acids shown below the sequences (T, D, H, T, G, E) are those catalytically important in *S. typhimurium* (CALIGIURI and BAUERLE 1991). The dotted line indicates regions that in the related AS CO I proteins have been implicated in interaction with AS COII (KNÖCHEL *et al.* 1999, MOROLLO and ECK 2001). *C.c.: C. cinereus, A.n: A. nidulans, S.c.: S. cerevisiae, S.p.: S. pombe, P.f.: P. falciparum, A.t.: A. thaliana, S.g.: S. griseus, E.c.: E. coli*

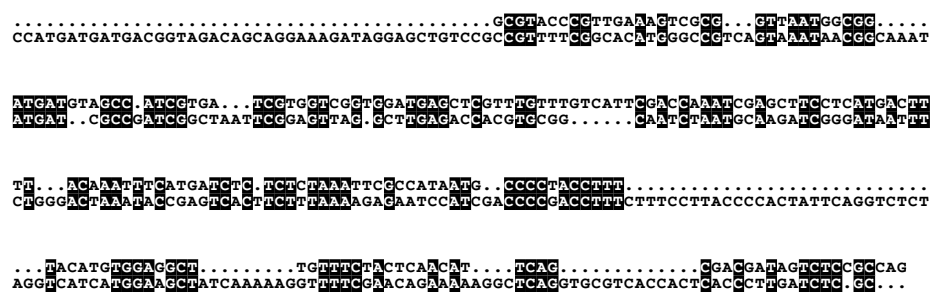


Fig. 3
Sequence alignment of core promoters of *C. cinereus* genes *pab1* and *trp1*

with catalytic function in *E. coli* (RAYL *et al.* 1996) are not well conserved amongst the PABA synthases (Fig. 2B) and thus may be involved in the maintenance of the proper structure of *E. coli* PabB and its catalytic domain. Indeed, some of these residues are known to be critical for the native polypeptide structure, as evidenced by failure to associate with PabA when mutated (RAYL *et al.* 1996, Fig. 2B). It is not clear yet from analyzing the amino acid sequences what causes PabB and AS COI proteins to react differentially with their common substrate chorismate.

Evolutionary origin of bifunctional PABA synthase genes

Anthranilate synthases, much like PABA synthases, are also composed of a single large protein or as two individual protein components (CRAWFORD 1989, NICHOLS 1996). Utilizing both chorismate and glutamine as substrates, PABA and anthranilate synthases catalyze similar enzymatic reactions and strong sequence similarities suggest a common ancestor for both enzymes (GONCHAROFF and NICHOLS 1984, NICHOLS 1996). Moreover, in some organisms (e.g., *B. subtilis*) component II supplies glutamine amidotransferase activity for both kinds of enzymes (CRAWFORD 1989, NICHOLS 1996).

The BLAST search performed with the *C. cinereus* Pab1 sequence revealed homology to the PABA synthases and also, at similar levels, to bacterial, algal and plant anthranilate synthases (not shown). We performed a phylogenetic analysis to investigate whether *pab* genes formed a monophyletic group (i.e., are descended from a single common ancestor)

versus the alternative that the *pab* genes have arisen multiple times from anthranilate synthase-like ancestors. We analyzed separately those protein domains in Pab1 of *C. cinereus* that were homologous to PabA and those homologous to PabB (Figs. 4 and 5).

In the PabA trees, all the fused PABA synthases, being either of actinomycetes or eukaryotic origin, grouped into a single monophyletic lineage (Fig. 4). In contrast, unfused PabA genes were found together with various bacterial AS COII genes in three separate

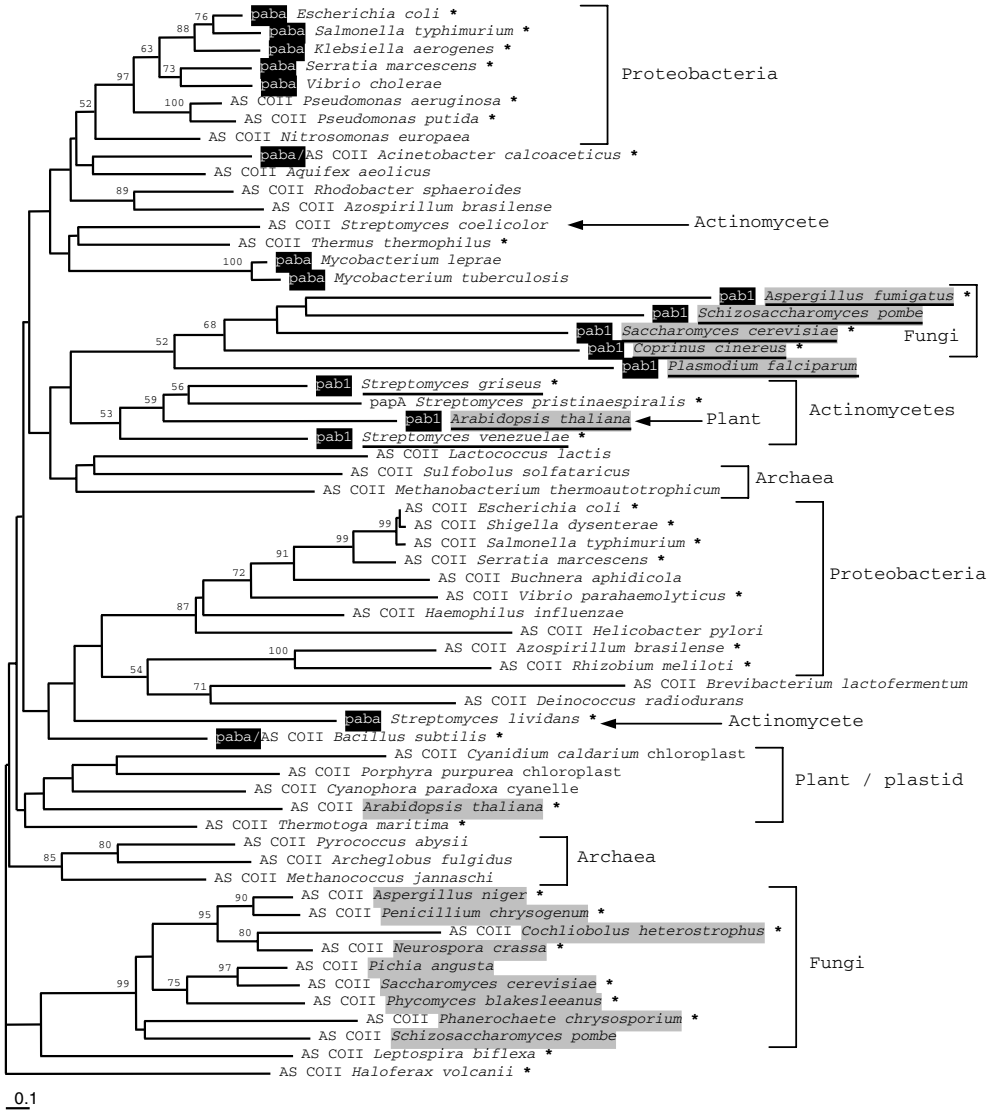


Fig. 4
Neighbor-joining phylogram of the relationships among PabA and AS COII (TrpG) proteins. Genes functioning in PABA synthesis are shown as white text on black background. Eukaryotic sequences are indicated using grey highlighting. PabA-PabB protein fusions are underlined. Asterisks indicate sequences for which functional studies have verified gene function. Numbers above branches indicate percent bootstrap support using 100 replicates

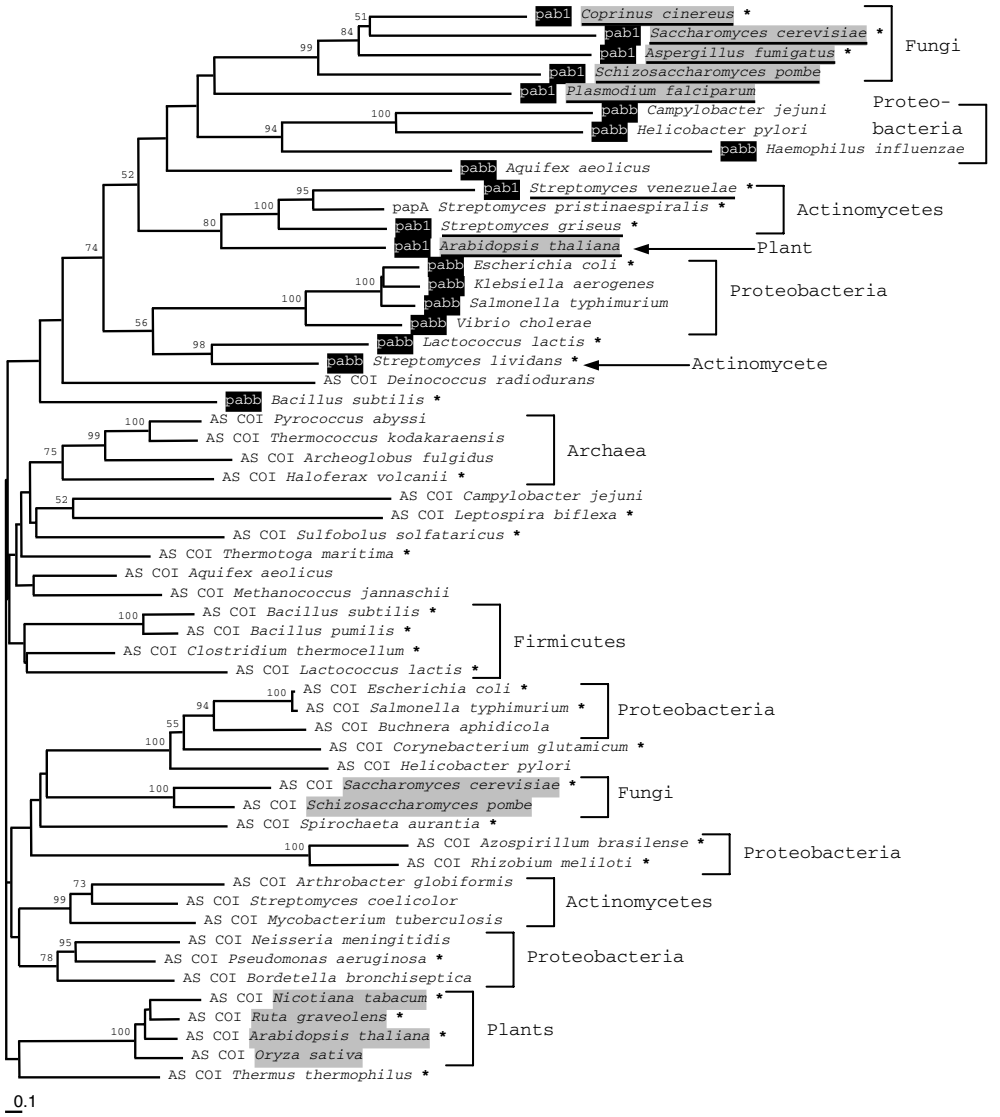


Fig. 5
Neighbor-joining phylogram of PabB and AS COI (TrpE) proteins. For explanation of details, see legend of Fig. 4

clades, each unrelated to the clade containing fused Pab genes. Of particular interest, the PabA proteins of *E. coli* and other proteobacteria group into a well-supported clade that appears to have arisen from a group of non-PAB synthase genes, including AS COII proteins of *Pseudomonas* spp. but also the amphibolic AS COII-PabA protein of *Acinetobacter calcoaceticus* (Fig. 4). These results suggest multiple independent origins of the glutamine amidotransferase component of PABA synthase (PabA) from AS COII-like ancestors.

In the PabB neighbor-joining phylogeny, the PabB proteins are nearly monophyletic with the single exception that the fused anthranilate synthase (AS COII-AS COI) protein of

D. radiodurans is basal but within the clade containing all of the PabB proteins (Fig. 5). Since a functional analysis of the gene product so far is lacking, it is possible that the *D. radiodurans* protein also functions in PABA synthesis. Within the PabB clade, the fused Pab genes do not appear monophyletic because solitary PabB proteins of four bacteria cluster within the fused Pab proteins of eukaryotes and actinomycetes. However, because the relationships between the four bacterial PabB proteins and the fused Pab proteins are not statistically supported by bootstrapping, it is unclear whether the bacterial PabB proteins are truly derived from within the fused Pab proteins, or rather the fused Pab proteins form a monophyletic group. To evaluate these alternative hypotheses, we used the maximum likelihood SH test of SHIMODAIRA and HASEGAWA (1999), as implemented in the PAML software package (YANG 1997), to estimate the likelihoods of two alternative trees. The neighbor-joining tree in which PabB proteins are not monophyletic is shown in Fig. 5; an alternative tree in which the fused Pab proteins are monophyletic was made by reconnecting the branch containing Pab1 of actinomycetes and *Arabidopsis* onto the branch uniting fungal and *Plasmodium* fused Pab proteins. Surprisingly, the results of the SH test indicated that a PabB phylogeny grouping all fused Pab proteins as monophyletic has a significantly higher likelihood than the phylogeny shown in Figure 5 ($P < 0.05$; results not shown). Thus, it is suggested that fused Pab proteins have a single evolutionary origin.

Discussion

In *C. cinereus* two genes are known to contribute to para-aminobenzoic acid production (NORTH 1990). In this study, we have established the *pab1* gene sequence from *C. cinereus*. We found the gene to encode a bifunctional PABA synthase with an N-terminal glutamine amidotransferase domain and a C-terminal chorismate amination domain. Bifunctional enzymes have previously been described in actinomycetes and other eukaryotes. Our phylogeny analysis suggests that these fused proteins have a common origin. The fused genes also group together with unfused *pabB* genes. In contrast, *pabA* genes are far more heterogenous and are found in three separate clades amongst various bacterial AS COIIs genes, irrespective of whether fused or not to other genes of the tryptophan biosynthesis pathway. A recent report of *pabB* in *Streptomyces venezuelae* suggested a phylogeny that differs significantly from that presented here (CHANG *et al.* 2001). Viewing the whole protein sequences of fused and unfused *pabB* proteins and fused and unfused AS COIIs in their phylogenetic analysis, CHANG *et al.* (2001) propose a superfamily of anthranilate/PABA synthases in which *pabB* genes are not monophyletic with respect to AS COII genes. However, their analysis also showed that fused PABA synthases are monophyletic, with several of the fused genes from *Streptomyces* spp. having diverged to function in secondary metabolism. Out phylogenetic analyses include many more sequences than CHANG *et al.* (2001). In contrast to their results, we find that *pabB* genes appear to be monophyletic, and it is the *pabA* genes that are clearly not monophyletic and have multiple origins.

Perhaps the most surprising result from the molecular phylogenies in our study was the close relationship of fused *pab* genes in actinomycetes with the fused *pab* gene reported from *A. thaliana*. The unfused PabA and PabB proteins from *S. lividans* are also included in the phylogeny, and these proteins group separately from the fused Pab proteins of other *Streptomyces* spp. Possibly, the fused Pab proteins in actinomycetes have been acquired via horizontal gene transfer from a plant or another eukaryotic lineage, because the unfused proteins of *S. lividans* group more according to organismal phylogeny (i.e., group with other gram-positive eubacteria) than the fused Pab proteins. An alternative hypothesis is that following a duplication and subsequent fusion of the *pabA-pabB* genes, the unfused *pab* genes have been selectively lost in most, but not all, of the actinomycetes. In this scenario, the fused *pab* genes of actinomycetes would then have given rise to eukaryotic PABA syn-

thases. The presence of duplications of *pab* genes in the history of the actinomycetes is witnessed by the close relationship of the protein PapA from *S. pristinaespiralis* to the fused Pab proteins of actinomycetes (Figs. 4 and 5). The *S. pristinaespiralis* PapA protein obviously has recently diverged to perform a function unrelated to PAB synthesis – synthesis of 4-amino 4-deoxychorismic acid precursors of the antibiotic pristinamycin I (BLANC *et al.* 1997). A second example, just released after completing our analysis, is found in *S. venezuelae.* Next to the fused *pab1* gene the bacterium has two extra genes, one falling in the *pabA* class and the other in the *pabB* class. Unlike the fused gene, the unfused *pabA* and *pabB* do not function in chloramphenicol synthesis (CHANG *et al.* 2001) Such recent divergence in function highlights the difficulty in making protein functional assignments based on sequence similarity.

How are gene fusions among PABA synthase and anthranilate synthase components created, and are these fusions relevant to the function or phylogeny of these genes? It could be hypothesized that these gene fusions may not act to increase cooperativity among subunits of the biosynthesis genes because, for example, in the most frequent gene fusion (*trpC-trpF*), the two domains both fold and function independently of one another (NICHOLS 1996). Gene fusions could often be generated by weak or deleted transcription termination signals between adjacent genes, resulting in the formation of fused transcripts. Subsequent deletion of a stop codon in the N-terminal protein of the fused transcript could lead to translation read-throughs into the second protein coding region and hence a fused protein. Importantly, the *trp* genes of most prokaryotes are typically organized into operons (CRAWFORD 1989), and gene fusions uniting previously distinct but adjacent coding regions could be responsible for creating the observed fusion proteins. Although the *pab* genes of *E. coli* and *S. typhimurium* are unlinked, in at least three bacteria, *S. lividans* (ARHIN and VINING 1993), *S. venezuelae* (CHANG *et al.* 2001) and *C. jejuni* (PARKHILL *et al.* 2000), *pabA* and *pabB* genes are adjacent on the chromosome.

Acknowledgements

We thank AUSTEN R. D. GANLEY for a critical review of the manuscript. This work was supported by the Swiss National Foundation (grants 31-46'940.96, 31-46'940.96/2 and 31-59157.99) and the ETH Zürich. UK is sponsored by the Deutsche Umweltstiftung (DBU).

References

ARHIN, F. F. and VINING, L. C., 1993. Organization of the genes encoding *p*-aminobenzoic acid synthetase from *Streptomyces lividans* 1326. Gene, **126**, 129–133.

BAE, Y. M., HOLMGREN, E. and CRAWFORD, I. P., 1989. *Rhizobium meliloti* anthranilate synthase gene: cloning, sequence, and expression in *Escherichia coli*. J. Bacteriol., **171**, 3471–3478.

BLANC, V., GIL, P., BAMAJACQUES, N., LORENZON, S., ZAGOREC, M., SCHLEUNIGER, J., BISCH, D., BLANCHE, F., DEBUSSCHE, L., CROUZET, J. and THIBAUT, D., 1997. Identification and analysis of genes from *Streptomyces pristinaespiralis* encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I. Mol. Microbiol., **23**, 191–202.

BOULIANNE, R. P., LIU, Y., AEBI, M., LU, B. C. and KÜES, U., 2000. Fruiting body development in *Coprinus cinereus*: regulated expression of two galectins secreted by a non-classical pathway. Microbiology, **146**, 1841–1853.

BROWN, J. S., AUFAUVRE-BROWN, A., BROWN, J., JENNINGS, J. M., ARST, H. Jr. and HOLDEN, D. W., 2000. Signature-tagged and directed mutagenesis identify PABA synthetase as essential for *Aspergillus fumigatus* pathogenicity. Mol. Microbiol., **36**, 1371–1380.

BUVINGER, W. E., STONE, L. C. and HEATH, H. E., 1981. Biochemical genetics of tryptophan synthesis in *Pseudomonas acidovorans*. J. Bacteriol., **147**, 62–68.

CALIGIURI, M. G. and BAUERLE, R., 1991. Subunit communication in the anthranilate synthase complex from *Salmonella typhimurium*. Science, **252**, 1845–1848.

CHANG, Z., SUN, Y., HE, J. and VINING, L. C., 2001. *p*-Aminobenzoic acid and chloramphenicol biosynthesis in *Streptomyces venezuelae*: gene sets for a key enzyme, 4-amino-4-deoxychorismate synthase. Microbiology, **147**, 2113–2126.

COSSINS, E., 2000. The fascinating world of folate and one-carbon metabolism. Can. J. Bot., **78**, 691–708.

COSSINS, E. A. and CHEN, L., 1997. Folates and one-carbon metabolism in plants and fungi. Phytochemistry, **45**, 437–452.

CRAWFORD, I. P., 1989. Evolution of a biosynthetic pathway: the tryptophan paradigm. Ann. Rev. Genet., **43**, 567–600.

CRIADO, L. M., MARTIN, J. F. and GIL, J. A., 1993. The *pab* gene of *Streptomyces griseus*, encoding *p*-aminobenzoic acid synthase, is located between genes possibly involved in candicidin biosynthesis. Gene, **126**, 135–139.

DE TROCH, P., DOSSELAERE, F., KEIJERS, V., DE WILDE, P. and VANDERLEYDEN, J., 1997. Isolation and characterization of the *Azospirillum brasilense trpE(G)* gene, encoding anthranilate synthase. Curr. Microbiol., **34**, 27–32.

EDMAN, J. C., GOLDSTEIN, A. L. and ERBE, J. G., 1993. Para-aminobenzoate synthase gene of *Saccharomyces cerevisiae* encodes a bifunctional enzyme. Yeast, **9**, 669–675.

FELSENSTEIN, J., 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author.

GONCHAROFF, P. and NICHOLS, B. P., 1984. Nucleotide-sequence of *Escherichia coli pabB* indicates a common evolutionary origin of para-aminobenzoate synthetase and anthranilate synthetase. J. Bacteriol., **159**, 57–62.

GRANADO, J. D., KERTESZ-CHALOUPKOVÁ , K., AEBI, M. and KÜES, U., 1997. Restriction enzyme-mediated DNA integration in *Coprinus cinereus*. Mol. Gen. Genet., **256**, 28–36.

GREEN, J. M., NICHOLS, B. P. and MATTHEWS, R. G., 1996. Folate biosynthesis, reduction, and polyglutamylation. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology. Vol. 1, 2nd ed. (NEIDHARDT, F. C., CURTISS, R., III, INGRAHAM, J. L., LIN, E. C. C., BROOKS LOW, K., MAGASANIK, B., REZNIKOFF, W. S., RILEY, M., SCHAECHTER, M. and UMBARGER, H. E., Editors), pp. 665–673. ASM Press, Washington DC.

GUO, Z. and SHERMAN, F., 1996. 3′-end-forming signals of yeast mRNA. Trends Biochem. Sci., **21**, 477–481.

HERRMANN, K. M., 1995. The shikimate pathway – Early steps in the biosynthesis of aromatic compounds. Plant Cell, **7**, 907–919.

HUMPHREY, T., BIRSE, C. E. and PROUDFOOT, N. J., 1994. RNA 3' end signals of the *S. pombe ura4* gene comprise a site determining and efficiency element. EMBO J., **13**, 2441–2451.

HÜTTER, R., NIEDERBERGER, P. and DEMOSS, J. A., 1986. Tryptophan biosynthetic genes in eukaryotic microorganisms. Ann. Rev. Microbiol., **40**, 55–77.

KAPLAN, J. B., GONCHAROFF, P., SEIBOLD, A. M. and NICHOLS, B. P., 1984. Nucleotide sequence of the *Acinetobacter calcoaceticus trpGDC* gene cluster. Mol. Biol. Evol., **1**, 456–472.

KNÖCHEL, T., IVENS, A., HESTER, G., GONZALEZ, A., BAUERLE, R., WILMANNS, M., KIRSCHNER, K. and JANSONIUS, J. N., 1999. The crystal structure of anthranilate synthase from *Sulfolobus solfataricus*: functional implications. Proc. Natl. Acad. Sci. USA, **96**, 9479–9484.

KOS, A., KUIJVENHOVEN, J., WERNARS, K., BOS, C. J., VAN DEN BROEK, H. W. J., POUWELS, P. H. and VAN DEN HONDEL, C. A. M. J. J., 1985. Isolation and characterization of the *Aspergillus niger trpC* gene. Gene, **39**, 231–238.

MOROLLO, A. A. and ECK, M. J., 2001. Structure of the cooperative allosteric anthranilate synthase from *Salmonella typhimurium*. Nature Struct. Biol., **8**, 243–247.

MUTASA, E. S., TYMON, A. M., GÖTTGENS, B., MELLON, F. M., LITTLE, P. F. R. and CASSELTON, L. A., 1990. Molecular organisation of an *A* mating type factor of the basidiomycete fungus *Coprinus cinereus*. Curr. Genet., **18**, 223–229.

NICHOLS, B. P., 1996. Evolution of genes and enzymes of tryptophan biosynthesis. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology. Vol. 2, 2nd ed. (NEIDHARDT, F. C., CURTISS, R., III, INGRAHAM, J. L., LIN, E. C. C., BROOKS LOW, K., MAGASANIK, B., REZNIKOFF, W. S., RILEY, M., SCHAECHTER, M. and UMBARGER, H. E., Editors), pp. 2638–2648. ASM Press, Washington DC.

NICHOLAS, K. B, NICHOLAS, H. B., JR. and DEERFIELD, D. W., II., 1997. GeneDoc: analysis and visualization of genetic variation. EMBNEW.NEWS, **4**, 14.

NORTH, J., 1990. *Coprinus cinereus*. In: Genetic Maps. Book 3. Lower Eukaryotes. Fifth Edition. (S. J. O'BRIEN, Editor), pp. 370–381. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

PARKHILL, J., WREN, B. W., MUNGALL, K., KETLEY, J. M., CHURCHER, C., BASHAM, D., CHILLINGWORTH, T., DAVIES, R. M., FELTWELL, T., HOLROYD, S., JAGELS, K., KARLYSHEV, A., MOULE, S., PALLEN, M. J., PENN, C. W., QUAIL, M., RAJANDREAM, M. A., RUTHERFORD, K. M., VANVLIET, A., WHITEHEAD, S. and BARRELL, B. G., 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature, **403**, 665–668.

RAYL, E. A., GREEN, J. M. and NICHOLS, B. P., 1996. *Escherichia coli* aminodeoxychorismate synthase: analysis of *pabB* mutations affecting catalysis and subunit association. Biochim. Biophys. Acta, **1295**, 81–88.

ROUX, B. and WALSH, C. T., 1993. *p*-aminobenzoate synthesis in *Escherichia coli*: Mutational analysis of 3 conserved amino-acid-residues of the amidotransferase PabA. Biochemistry, **32**, 3763–3768.

SAMBROOK, J., FRITSCH, E. F. and MANIATIS, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

SCHECHTMANN, M. G. and YANOFSKY, C., 1983. Structure of the tri-functional *trp-1* gene of *Neurospora crassa* and its aberrant expression in *Escherichia coli*. J. Mol. Appl. Genet., **2**, 83–99.

SEITZ, L. C., TANG, K., CUMMINGS, J. W. and ZOLAN, M. E., 1996. The *rad9* gene of *Coprinus cinereus* encodes a proline-rich protein required for meiotic chromosome condensation and synapsis. Genetics, **142**, 1105–1117.

SHIMODAIRA, H. and HASEGAWA, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol., **16**, 1114–1116.

SKRZYNIA, C., BINNINGER, D. M., ALSPAUGH, J. A., II and PUKKILA, P. J., 1989. Molecular characterization of TRP1, a gene coding for tryptophan synthetase in the basidiomycete *Coprinus cinereus*. Gene, **81**, 73–82.

SLOCK, J., STAHLY, D. P., HAN, C.-Y., SIX, E. W. and CRAWFORD, I. P., 1990. An apparent *Bacillus subtilis* folic acid biosynthetic operon containing *pab*, an amphibolic *trpG* gene, a third gene required for synthesis of *para*-aminobenzoic acid, and the dihydropteroate synthase genes. J. Bact., **172**, 7211–7226.

THOMPSON, J. D., GIBSON, T. J., PLEWNIAK, F., JEANMOUGIN, F. and HIGGINS, D. G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nuc. Acids Res., **24**, 4876–4882.

TRIGLIA, T. and COWMAN, A. F., 1999. *Plasmodium falciparum*: a homologue of *p*-aminobenzoic acid synthetase. Exp. Parasitol., **92**, 154–158.

VINNICOMBE, H. G. and DERRICK, J. P., 1999. Dihydropteroate synthase: an old drug target revisited. Biochem. Soc. Trans., **27**, 53–58.

WEBER, C. A., HUDSPETH, M. E. S., MOORE, G. P. and GROSSMAN, L. I., 1986. Analysis of the mitochondrial and nuclear genomes of two basidiomycetes, *Coprinus cinereus* and *Coprinus stercorarius*. Curr. Genet., **10**, 515–525.

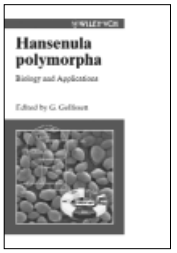
WENG, M. and ZALKIN, H., 1987. Structural role for a conserved region in the CTP synthetase glutamine amide transfer domain. J. Bacteriol., **169**, 3023–3028.

YANG, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS, **13**, 555–556.

ZALKIN, H. and SMITH, J. L., 1998. Enzymes utilizing glutamine as an amide donor. Adv. Enzymol., **72**, 87–144.

Mailing address: URSULA KÜES, Georg-August-Universität Göttingen, Institut für Forstbotanik, Büsgenweg 2, D-37077 Göttingen, Germany
Fax: ++49-551-392705; Phone: ++49-551-397024
E-mail: ukuees@gwdg.de

Pacesetters in Biotechnology!



Gellissen, G. (ed.)

Hansenula polymorpha

Biology and Applications

2002. 365 pages
Hardcover.
€ 119.-/£ 70,- / $ 95.-*

Aimed at researchers and scientists working in the field this book provides a comprehensive, up-to-date overview of the present status of Hansenula polymorpha research, applications and methods. It covers all aspects of the organism, ranging from systematics, genetics, methanol metabolism and peroxisomal function to its use as a technology platform for the production of recombinant proteins, and also includes a detailed chapter on laboratory methods.

Forthcoming!

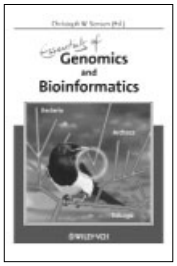
Hoppert, M.

Microscopic Techniques in Biotechnology

2002. 314 pages.
Hardcover.
€ 109.-/£ 65,- / $ 95.-*
ISBN 3-527-30198-4

Focusing on current applications, this book presents the various methods as well as their suitability and limitations for any specific problem. One particularly valuable feature is its coverage of all basic information on the structure of the objects relevant in biotechnology, allowing readers to choose the most suitable applications. They will also find in-depth background information on structure-function relationships, plus sample preparation with respect to a particular technique and the equipment required. The authors finish with an overview of the potential for devices and applications of future interest to biotechnology.

** €-Prices refer to Germany only.*



Sensen, C.W. (ed.)

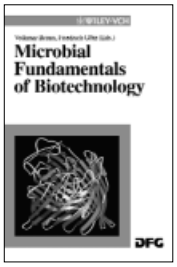
Essentials of Genomics and Bioinformatics

2002. 442 pages.
Softcover.
€ 69.-/£ 39.95 / $ 60.-
ISBN 3-527-30541-6

This book, written by internationally renowned experts, gives excellent coverage of genomics – one of the most rapidly developing areas in life science and treats bioinformatics as the most important related discipline to genomics.

It encompasses everything from nucleic acid technologies, proteomics and bioinformatics, applications in human health, agriculture and comparative genomics to the legal and ethical issues of the subject.

Get the concise version of the Biotechnology Volume 5b for your own desk now !



Braun, H., Götz, F. (eds.)

Microbial Fundamentals of Biotechnology

2001. 377 pages.
Hardcover.
€ 139.-/£ 80.- / $ 125.-
ISBN 3-527-30615-3

This title brings together scientists of different disciplines with common interests: to gather information and report research results in studies of micro-organisms and their metabolism and to present possibilities for finding new, biologically active metabolites and gives answers to questions: how to use the biosynthetic abilities of micro-organisms to produce new biologically active compounds, how to synthesise (microbiologically or chemically) new structures, and how to test their biological activities.

90642022_vi

Register now for the free
WILEY-VCH Newsletter!
http://www.wiley-vch.de/home/pas

WILEY-VCH · P.O. Box 10 11 61
69451 Weinheim, Germany · Fax: +49 (0) 62 01-60 61 84
e-mail: service@wiley-vch.de · **http://www.wiley-vch.de/**

WILEY-VCH