Available online at www.sciencedirect.com
ScienceDirect
Phytochemistry 68 (2007) 853–864
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Plant progesterone 5β-reductase is not homologous to the animal enzyme. Molecular evolutionary characterization of P5βR from *Digitalis purpurea*

Isabel Gavidia [a,*], Rosa Tarrío [b], Francisco Rodríguez-Trelles [b], Pedro Pérez-Bermúdez [a], H. Ulrich Seitz [c]

[a] *Department of Plant Biology, Faculty of Farmacy, University of Valencia, Av. V.A. Estellés s/n, 46100 Burjasot, Spain*
[b] *Grupo de Medicina Xenómica, Hospital Clínico Universitario, Universidade de Santiago de Compostela, 15706 Santiago, Spain*
[c] *Center of Plant Molecular Biology (ZMBP), University of Tübingen, Germany*

Received 5 October 2006; received in revised form 8 November 2006
Available online 20 December 2006

Abstract

Plants of the genus *Digitalis* produce cardiac glycosides, i.e. digoxin, which are widely used for congestive heart failure. Progesterone 5β-reductase (P5βR) is a key enzyme in the biosynthesis of these natural products. Here, we have carried out the purification and partial amino acid sequencing of the native P5βR from foxglove (*Digitalis purpurea*), and isolated a cDNA encoding this enzyme. Similarly to other steroid 5β-reductases, the recombinant P5βR catalyzes the stereospecific reduction of the Δ^4-double bond of several steroids with a 3-oxo,$\Delta^{4,5}$ structure. The gene encoding P5βR is expressed in all plant organs, and maximally transcribed in leaves and mature flowers. P5βR belongs to the short-chain dehydrogenase/reductase (SDR) superfamily, bearing no structural homology to its mammalian counterpart, which is a member of the aldo-keto reductase (AKR) superfamily. A similar situation occurs with 3β-hydroxy-Δ^5-steroid dehydrogenase (3βHSD), the gene immediately preceding P5βR in the cardenolide pathway, which suggests that the entire route has evolved independently in animals and plants. P5βR is retained only in plants, where it is ubiquitous, and a few distantly related bacterial lineages after its diversification from the last universal common ancestor. Evolutionary conserved changes in its putative active site suggest that plant P5βR is a member of a novel subfamily of extended SDRs, or a new SDR family.

Keywords: *Digitalis purpurea*; Plantaginaceae; Cardenolide biosynthesis; Steroid metabolism; Gene expression; Molecular evolution; Analogous enzymes; P5βR; SDR family

1. Introduction

The genus *Digitalis*, a member of the Plantaginaceae, comprises several therapeutic species, as they are a source of cardiac glycosides. Ecologically, the advantages of cardenolides for *Digitalis* are uncertain. As for many other secondary metabolites, a role as deterrents to herbivores has been suggested (Harborne, 1982; Malcolm and Zalucki, 1996).

Results with labelled and unlabelled precursors from the last three decades suggest a hypothetical pathway for cardenolide biosynthesis. Up to date, only four reactions of the genin biosynthesis have been described at biochemical level (see Fig. 1): (i) the transformation of phytosterols to pregnenolone (Pilgrim, 1972), (ii) progesterone formation from pregnenolone (Finsterbusch et al., 1999), and (iii) the sequential reductions of progesterone to 5β-pregnane-3,20-dione (Gärtner et al., 1990) and 5β-pregnan-3β-ol-20-one (Gärtner and Seitz, 1993). In animal tissues, all these reactions of the steroid metabolism, except the cholesterol side-chain-cleaving reaction, are catalysed by

* Corresponding author. Tel.: +34 963544929; fax: +34 963544926.
E-mail address: isabel.gavidia@uv.es (I. Gavidia).

0031-9422/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2006.11.019

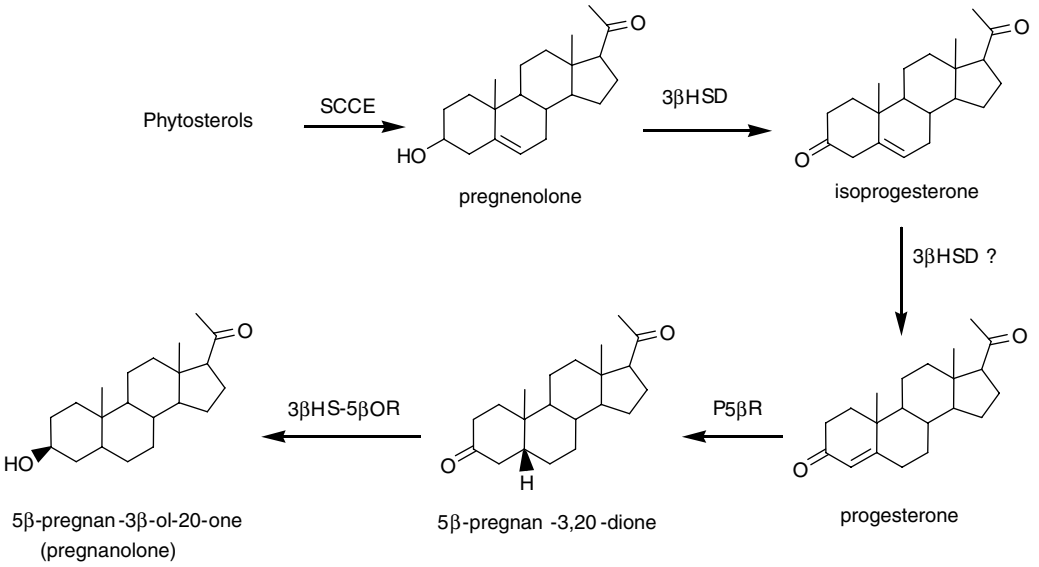

Fig. 1. Putative initial steps and enzymes involved in the biosynthetic pathway of cardenolides.

enzymes of the aldo-keto reductase (AKR) superfamily (Penning et al., 2001).

The steps of the cardenolide pathway until progesterone formation are similar to those of cholesterol metabolism towards steroid hormones in animals; the enzymes involved convert a variety of substrates and can also be found in plants that do not accumulate cardenolides (Seitz and Gärtner, 1994). Progesterone reduction is a divergent intersection point of distinct metabolic routes in both animals and plants. In this reaction, progesterone is metabolised by stereospecific double bond reductases producing 5α- or 5β-derivatives.

A steroid 5α-reductase gene was for the first time described in *Arabidopsis thaliana* (*DET2*), and further investigations have shown that the 5α-reductase gene is well conserved in the plant kingdom. *DET2* encodes an enzyme that reduces different steroid substrates (i.e. progesterone, testosterone, etc.). The enzyme is involved in the biosynthesis of the plant growth hormones brassinosteroids (Li et al., 1996, 1997). In parallel to this participation in primary metabolism, the 5α-reduction of progesterone in the cardenolide-producing species leads to uzarigenin derivatives (Gavidia et al., 2002b), that are also used therapeutically. The steroid 5α-reductase activity is encoded by descendants of the same ancestral gene in animals and plants, which belongs to the AKR superfamily (Li et al., 1997).

The alternative double bond reduction of progesterone is catalysed by a 5β-reductase, which has been proposed to have a key function in the cardenolide-producing branch, forming the required 5β-configured pregnanes (Gärtner and Seitz, 1993). This is the only one of the known enzymes of this route exclusively present in cardenolide forming cultures, and its activity has been directly correlated with cardenolide biosynthesis (Gärtner et al., 1994; Stuhlemmer and Kreis, 1996).

Cloning of the gene that encodes this key enzyme was one goal of our research on cardenolides. In order to achieve this objective, two strategies were adopted. Based on the use of orthologous genes we screened a cDNA library of foxglove (*Digitalis purpurea*) using the cDNA encoding the Δ^4-3-ketosteroid 5β-reductase of rat liver (Onishi et al., 1991) as a probe. As a result of this study we reported the cloning and expression of two AKR genes (Gavidia et al., 2002a). These proteins reduced the ketone group, at position 3 and 20, of some steroids, but they were not active on the Δ^4-double bond of the compounds assayed.

In the following we describe the parallel approach performed based on the purification of the native progesterone 5β-reductase from *D. purpurea* leaves and the subsequent partial amino acid sequencing of the protein. We used the peptide information for screening cDNA and genomic libraries prepared from leaves of *D. purpurea*. We report the isolation of *P5βR*, the gene encoding progesterone 5β-reductase, as well as its heterologous expression, structural and molecular organisation and expression pattern. In addition, we address the evolutionary history of the *P5βR* gene in connection with the possibility that the cardenolide pathway emerged independently in animals and plants.

2. Results and discussion

2.1. Isolation of the cDNA encoding P5βR

In a previous work, Gärtner and coworkers purified the native protein progesterone 5β-reductase and sequenced various peptide fragments (Gärtner et al., 1994). Since they did not obtain a gene fragment using degenerated primers, a new purification of the native protein was performed. The

size of the peptide fragments sequenced ranged from 9 to 30 amino acids (Fig. 2), which included the peptides (SLAE and FYYDLEDIMLEXV) reported by Gärtner et al. (1994). A comparison of these amino acid sequences with those available at that moment in the GenBank/EMBL sequence data bases indicated similarity to a wounding stress protein (AWI31) from *A. thaliana* (Yang et al., 1997).

For the isolation of the *P5βR* cDNA a pair of specific primers for *A. thaliana AWI31*, based upon two internal peptide sequences of progesterone 5β-reductase, were designed. The primer sequences were as follows: 5′-ACATGTCTTCTACGTCACCTGG and 5′-CTTCGCCTTGTCAATCCAAGAG. A 850 bp fragment was amplified from *D. purpurea* genomic DNA, this PCR product was used to screen a *D. purpurea* cDNA library at high stringency. One of the cDNA clones isolated, truncated at the 5′ region, contained the sequences of all the peptides obtained from the purified enzyme except one. 5′-RACE PCR reaction was performed using specific internal primers in order to get the full-length sequence. The PCR-amplicon obtained by 5′-RACE had a single in frame ATG initiation codon located at 11 nucleotides from the 5′-end, and the overlapping portion of the two sequences was identical. The entire sequence thus constructed was 1357 bp long and contained a 1167 bp long open reading frame which encodes 389 amino acid residues (Fig. 2). The calculated molecular weight (Mr 43,963 Da) was in agreement with that of the purified enzyme (Mr 43 kDa) on SDS–PAGE. The resulting clone was designated *P5βR*.

The screening of the *D. purpurea* genomic library led to the isolation of the *P5βR* gene. The analysis of the sequence of this genomic clone demonstrated *P5βR* to be a full-length cDNA clone because of the presence of an in frame stop codon before the first putative ATG. This analysis also revealed the presence of an unique intron (76 bp long) located at 40 bp from the ATG codon.

The sequence of *P5βR* from *D. purpurea* has been employed as a model for PCR amplification and cloning of the gene fragments corresponding to the putative homologous P5βR proteins in other *Digitalis* spp. (Herl et al., 2006a; see also the Erratum of Herl et al., 2006b). The identity of these sequences with P5βR from *D. purpurea* varied

```
                    M  S  W  W  W  A  G  A  I  G  A  A  K  K  K  L  E  E  D  D  A  P  P  K  H
TAGCTGAAAAAATGAGCTGGTGGTGGGCTGGAGCGATCGGCGCTGCAAAGAAAAAGTTGGAAGAAGATGACGCACCGCCAAAGCAT
 S  S  V  A  L  I  V  G  V  T  G  I  I  G  N  S  L  A  E  I  L  P  L  A  D  T  P  G  G
TCGAGCGTGGCGCTGATAGTTGGGGTAACTGGAATCATCGGCAACAGCCTGGCGGAGATCCTGCCGCTGGCCGACACCCCCGGCGGT
 P  W  K  V  Y  G  V  A  R  R  T  R  P  A  W  H  E  D  N  P  I  N  Y  I  Q  C  D  I  S
CCGTGGAAGGTATACGGCGTCGCCCGCCGCACCAGGCCCGCCTGGCATGAGGATAATCCGATCAATTACATCCAGTGCGACATATCC
 D  P  D  D  S  L  A  K  L  S  P  L  T  D  V  T  H  V  F  Y  V  T  W  A  N  R  S  T  E
GATCCAGATGACTCCCTAGCCAAGCTGTCGCCTCTGACTGATGTTACCCACGTGTTCTACGTTACCTGGGCTAATCGATCCACCGAA
 P  E  N  C  E  A  N  S  K  M  F  R  N  V  L  D  A  V  I  P  N  C  P  N  L  K  H  I  S
CCAGAAAACTGCGAAGCCAATAGCAAAATGTTCAGGAACGTGCTTGATGCAGTTATCCCTAATTGCCCCAATTTGAAGCACATCTCA
 L  Q  T  G  R  K  H  Y  M  G  P  F  E  S  Y  G  K  I  E  S  H  D  P  P  Y  T  E  D  M
TTGCAGACTGGGAGGAAGCATTACATGGGACCATTTGAATCCTACGGGAAAATAGAATCCCATGATCCACCCTACACTGAGGATATG
 P  R  L  K  Y  I  N  F  Y  Y  D  L  E  D  I  M  L  E  E  V  E  K  K  E  G  L  T  W  S
CCCAGGTTGAAGTACATTAACTTTTACTATGATTTAGAGGATATTATGCTTGAGGAGGTGGAGAAGAAGGAGGGTTTGACTTGGTCA
 V  H  R  P  G  N  I  F  G  F  S  P  Y  S  M  M  N  L  V  G  T  L  C  V  Y  A  A  I  C
GTTCATCGTCCAGGGAACATATTCGGGTTTTCTCCATATAGTATGATGAATTTGGTGGGCACCCTGTGTGTTTATGCAGCTATTTGC
 K  H  E  G  K  V  L  R  F  P  G  C  K  A  A  W  D  G  Y  S  D  C  S  D  A  D  L  I  A
AAACACGAGGGGAAGGTTTTGAGGTTTCCTGGTTGTAAGGCTGCGTGGGATGGGTACTCGGATTGCTCTGATGCAGATTTGATAGCG
 E  H  H  I  W  A  A  V  D  P  Y  A  K  N  E  A  F  N  V  S  N  G  D  V  F  K  W  K  H
GAGCATCATATTTGGGCGGCAGTGGATCCTTATGCAAAAAATGAGGCCTTTAATGTGAGTAATGGAGATGTGTTTAAATGGAAGCAT
 F  W  K  V  L  A  E  Q  F  G  V  E  C  G  E  I  G  E  G  V  D  L  K  L  Q  D  L  M  K
TTTTGGAAGGTGTTGGCGGAGCAGTTTGGGGTAGAGTGTGGAGAGTATGGAGAAGGGGTGGATTTGAAATTGCAGGATTTGATGAAG
 G  K  E  A  V  W  E  E  I  V  R  E  N  G  L  T  P  T  K  L  K  D  I  G  I  W  W  F  G
GGGAAGGAGGCGGTTTGGGAGGAAATCGTGAGGGAGAATGGGTTGACACCTACGAAACTGAAGGATATCGGGATTTGGTGGTTTGGT
 D  V  I  L  G  N  E  C  F  L  D  S  M  N  K  S  K  E  H  G  F  L  G  F  R  N  S  K  N
GATGTTATACTTGGGAATGAATGTTTTCTGGATAGTATGAATAAGAGCAAGGAGCATGGCTTTTTGGGATTTAGGAACTCCAAGAAT
 A  F  I  S  W  I  D  K  A  K  A  Y  K  I  V  P
GCGTTCATTTCTTGGATTGACAAGGCAAAAGCTTACAAGATTGTTCCTTGACATGGTTTCTCTTCAGTTTATCTTTGGTAATAATGT
GTTAGTTCACCAACTGTGTGGTGTTTTGGTGATTTCTGGCCTTGATGGTTGTGTCTTCTTTTATTATTGTATGATGTTTGTGAACTG
AATTGAAGCTTTTAATATCTGTCAGATTAATTAATGTTTAAAAAAAAAAAAAA
```

Fig. 2. cDNA and deduced amino acid sequence of *P5βR* from *D. purpurea*. The 10 peptide fragments sequenced from the native P5βR are underlined. GenBank Accession No. AJ310673.

between 96% and 99%, in all cases the 15 and 13 bp located at 5′ and 3′ termini, respectively, of the coding region have not been taken in account as they were used as primers for obtaining the mentioned sequences and, therefore, they are identical to the *D. purpurea P5βR* sequence.

2.2. Structural features of P5βR

Bl2seq (BLAST 2 sequences tool for local alignment at NCBI) pairwise alignment of P5βR against the available mammalian 5β-reductases (AKR1d1) does not detect significant similarity. In line with this outcome, the best-fit structural model for P5βR (*E*-value: 3.5e − 26) corresponds to GDP mannose-4,6-dehydratase of *Escherichia coli* (GMD; PDB code 1db3). 1db3 is a member of the Tyrosine-dependent oxidoreductase family, also known as short-chain dehydrogenases/reductases or SDR family, which is included within the Rossmann fold superfamily. The Rossmann fold, is formed by an open, parallel, six-stranded β-sheet with α-helices on both sides of the sheet. The domain forms a hydrophobic pocket which binds the dinucleotide cofactor (NAD, NADP, FAD). The domain contains a βA-αA-βB motif which has been used as a fingerprint for predicting the coenzyme-binding domain.

SDRs are homo-oligomers composed of subunits 250–350 amino acid residues in length, catalysing NAD(P)(H)-dependent oxidation/reduction reactions, which agrees with the observation that the molecular mass of denatured P5βR (43 kDa) is far lower than the apparent mass of the native protein (280 kDa) determined by Gärtner et al. (1994). The N-terminal region binds the coenzymes NAD(H) or NADP(H), whereas the C-terminal region embodies the substrate binding part. SDRs operate on a variety of substrates, including steroids. Until now, five SDR families have been identified (Kallberg et al., 2002). All of them exhibit reduced primary sequence conservation, but for several diagnostic motifs (1–7 in Fig. 3a). The 'classical' and 'extended' families are the most common. The extended family comprises isomerases, lyases, and several oxidoreductases, with a primary structure of ≈350 residues, which are sorted in three subfamilies, namely eD1, eD2, and eP1 (Kallberg et al., 2002). 1db3, the best-fit structural model for P5βR, is a member of the eP1 subfamily.

The alignment of P5βR and 1db3 (Fig. 3a) reveals several salient features. The residues corresponding to the diagnostic motifs of the extended SDR family (Kallberg et al., 2002) show similar numbers of matches to the consensus in the two sequences. Specifically, the characteristic Rossmann fold motif, GxxGxxG (where x denotes any amino acid; motif 1 in Fig. 3a), which is involved in the coenzyme-binding domain, is well conserved. The same is observed for the region binding the adenine ring of the coenzyme, DhxD (h, hydrophobic residue; motif 2 in Fig. 3a), and, in general, for most of the residues with suggested structural roles. Despite the substantial degree of resemblance at significant sites, P5βR lacks four residues in the region of the YxxxK motif (6 in Fig. 3a), which is typically conserved in the extended family and throughout most of the SDR superfamily. Various studies indicate that this motif forms part of the active site, Y acting as the catalytic base, and K interacting with the nicotinamide ribose and lowering the pK_a of the Y–OH (Oppermann et al., 2003). Interestingly, the changes to the YxxxK consensus motif in P5βR seem to form part of a novel, previously undescribed motif presumably retaining the same ancestral function. Multiple alignment of 10 representative P5βRs at the corresponding Rossmann fold and YxxxK motifs (Fig. 3b) indicates that these two regions have indeed evolved under similar patterns of strong purifying selection since the diversification of the tree of life (see below for a detailed discussion of the evolutionary history of the *P5βR* gene). The two regions are conserved even in more rapidly evolving paralogs not shown in the figure. Conservancy of the NFYYxxED motif on such a large evolutionary scale attests for its functional significance, presumably as a proxy of the typical YxxxK motif. If confirmed, this would classify P5βR as prototype of a new extended subfamily or a new SDR family.

The highest-scoring functional domain detected in the Pfam and COG databases for P5βR is a nucleoside-diphosphate-sugar epimerase domain (COG0451). There are several epimerases in the extended family. Specifically, two have been described from *Arabidopsis*, namely 3,5-epimerase/4-keto reductase and GDP-mannose-3′,5′-epimerase, which are multifunctional enzymes carrying, respectively, two and three functions at the same active site. However, P5βR only exhibits 5β-reductase activity (see below).

2.3. Heterologous expression of P5βR

To determine whether the *D. purpurea P5βR* encodes a functional steroid 5β-reductase, a full-length *P5βR* cDNA was overexpressed in *E. coli* as fusion protein with MBP or His-tag. The recombinant P5βRs were purified to apparent homogeneity by affinity chromatography, using an amylose resin column or a His-binding resin column, from the extracts of bacteria transformed with pMal-P5βR or pQE-P5βR, respectively. The purified proteins had the expected size, however steroid 5β-reductase activity was only detected for the purified P5βR-MBP protein. GC and MS analyses demonstrated that the purified P5βR-MBP possesses significant steroid 5β-reductase activity catalysing the reduction of progesterone to 5β-pregnane-3,20-dione (Fig. 4a). This reaction required NADPH as a cofactor and was irreversible, which is in agreement with the results reported by Gärtner et al. (1994) for the native protein. Like mammalian steroid 5β-reductases (Charbonneau and Luu-The, 2001) and the native P5βR (Gärtner et al., 1994), the recombinant form can catalyse the 5β-reduction of several steroids with a 3-oxo,$\Delta^{4,5}$ structure, and the highest substrate specificity was obtained with progesterone. Nevertheless, the enzyme catalyses less efficiently the transformation of other substrates. The relative specific activity

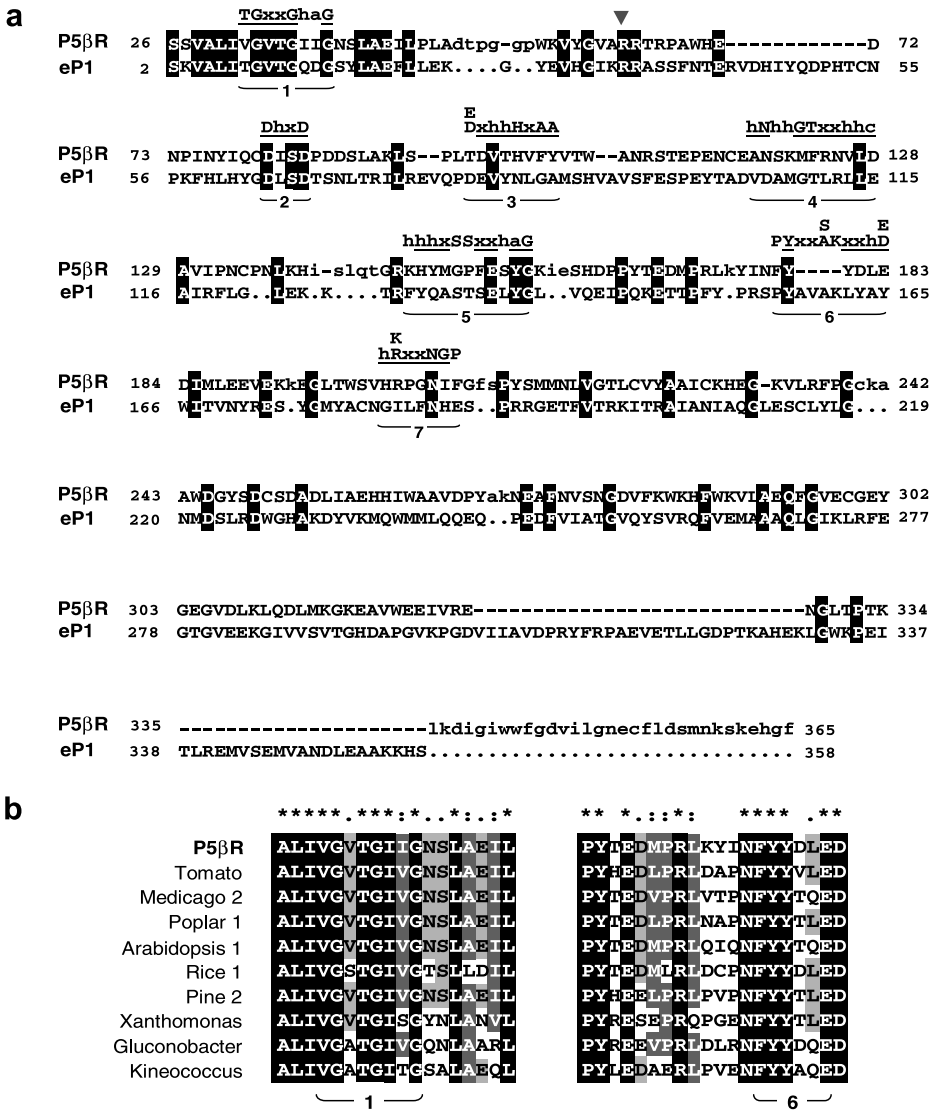

Fig. 3. (a) Pairwise alignment of P5βR against the extended SDR subfamily eP1 represented by *E. coli* GDP-mannose 4,6-dehydratase (PDB code 1db3), obtained using the SUPERFAMILY method (http://supfam.mrclmb.cam.ac.uk/SUPERFAMILY/) (Gough et al., 2001). SUPERFAMILY is a database of identified domains within proteins of known structure using Hidden Markov Models (HMM), relying on the Structural Classification of Proteins (SCOP) database (Andreeva et al., 2004; Murzin et al., 1995). Shaded columns indicate identical amino acids. eP1 diagnostic motifs (1–7) are shown above the alignment. The motifs are '1' structural role in coenzyme binding region, '2' adenine ring binding of coenzyme, '3' structural role in stabilising central β-sheet, '4–6' parts of the active site, and '7' structural role, reaction direction (see Kallberg et al., 2002). Letters above the motifs denote alternative amino acids at the corresponding positions. Underlined positions in the motifs indicate similar residues in P5βR. 'a' represents aromatic (F, W, Y, H), 'c' charged (D, E, H, K, R), 'h' hydrophobic (I, L, V, C, A, G, M, F, Y, W, H, T, P), 'p' polar (Y, W, H, K, R, E, Q, D, N, S, T), and 'x' any residue. (b) Alignment of P5βR and representative similar sequences at the diagnostic motif 1 and 6 regions. Multiple alignment was done in CLUSTALW 1.81, using the BLOSUM weight matrix and default gap open and extension penalties. Black, dark-grey, and light-grey shading denotes: 100% (*), ⩾80% (:), ⩾60% (.) amino acid conservation.

calculated for testosterone and cortisol were threefold and fourfold lower, respectively, than that found for progesterone as substrate.

In some enzymatic reactions, the two main chromatographic peaks, progesterone (P, RT = 29.32) and its 5β-reduced product (5β, RT = 27.50), were accompanied by two minor peaks with intermediate retention times (Fig. 4b). MS analyses allowed to identify these compounds as $\Delta^{5,6}$ isoprogesterone (RT = 27.88) and 17-isoprogesterone (RT = 28.33). As discussed above, some SDR proteins could show different activities (i.e. reductase and epimerase) linked to the same active site. Therefore, it was necessary to determine whether, or not, the $\Delta^{5,6}$ isomer and the 17-epimer of progesterone appeared as products of the enzyme activity. Our experimental data proved that the formation of these products is not related with the activity of the recombinant P5βR, since both peaks also appeared in the control chromatograms from reactions without enzyme (Fig. 4c).

We can conclude that the formation of the isomer and epimer occurred during the steps of organic solvent evaporation in the process of pregnane extraction. In this concern, we point out that similar chromatograms were

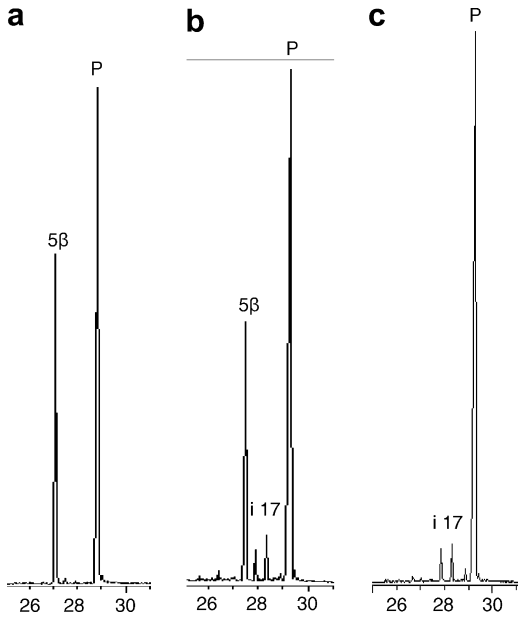

Fig. 4. GC–MS analysis of the reaction catalysed by recombinant P5βR: (a) standard enzyme assay using progesterone (P) as substrate and where the only product obtained was identified as 5β-pregnan-3,20-dione (5β); (b) standard enzyme assay where isoprogesterone (i) and 17-isoprogesterone (17) were detected; (c) control reaction without enzyme.

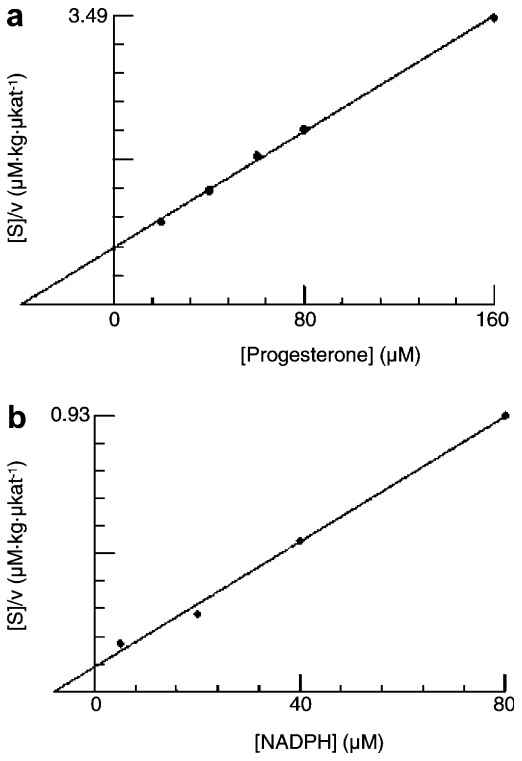

Fig. 5. Biochemical characterisation of the recombinant P5βR. The purified enzyme (20 μg) was incubated as described in Section 3 with varying concentrations of progesterone (0–160 μM) or NADPH (0–80 μM) to determine their respective kinetic data. Activity was calculated based on the reduction of absorbance at 340 nm. Hanes plots for the substrate (a) and cofactor (b).

reported by Herl et al. (2006a) in their work on functional studies of 5β-POR from *Digitalis lanata*. Surprisingly, they affirmed that the minor compounds seen in the chromatograms did not represent pregnane isomers as they could deduce from the MS spectra of their experiments.

To determine the kinetic properties of expressed P5βR protein, 20 μg of recombinant protein was incubated at 30 °C with varying amounts of NADPH (0–80 μM) or progesterone (0–160 μM). The apparent K_m value for progesterone was determined to be 39.1 ± 1.1 μM with a V_{max} of 57.0 ± 2.5 μkat kg^{-1} (Fig. 5a), while the apparent K_m value for NADPH was 7.9 ± 0.3 μM with a V_{max} of 94.7 ± 3.2 μkat kg^{-1} (Fig. 5b). These values are comparable to those calculated for the native P5βR, $K_m = 34$ μM for progesterone and $K_m = 6$ μM for NADPH (Gärtner et al., 1994).

A similar enzymatic function has been attributed to a cDNA (5β-POR) isolated from *D. lanata* (Herl et al., 2006a). The isolation of this cDNA was based in the use of primers designed from the *D. purpurea P5βR* sequence (see the Erratum published by Herl et al., 2006b). Since both primers included part of the coding region, the resulting cDNA is not a real full length for *D. lanata*. Data of the kinetic properties, determined for the 5β-POR expressed from the generated cDNA, showed lower V_{max} and higher K_m values than those obtained for the heterologous and native P5βR proteins from *D. purpurea*.

2.4. Gene expression analysis

The expression pattern of the gene *P5βR* in several *D. purpurea* tissues was evaluated by semi-quantitative RT–PCR. PCR amplification was performed using specific primers for *P5βR* and for actin, as a control for constitutive expression, (Fig. 6a). RNA transcripts of the gene were present in all the organs tested (mature flowers, leaves, stem at the base of the rosette, vascular bundle at the basal part of the leaf, and roots). They were more abundant in leaves and flowers and the lowest amount was present in the vascular bundle.

Cardenolides accumulate mainly in leaves, albeit they are also present in the other organs of the plant. Flowers, shoots and roots have, respectively, 40%, 10% and 10%, of the cardenolide content in basal leaves (Vogel and Luckner, 1981). These authors also observed that the cardenolide content in flowers is much higher in developed that in opening flowers, which could explain the small difference in *P5βR* transcript abundance between flowers and leaves detected in our study. Therefore, the reported expression levels of *P5βR* gene correlate with the spatial accumulation of cardenolides in *Digitalis* plants.

A comparable correlation between gene transcript abundance and cardenolide accumulation could not be established for *3βHSD*, the other gene involved in the production of these secondary metabolites that has been cloned and its expression pattern studied. Lindemann et al. (2000) found that the *3βHSD* gene is active predominantly in roots, although much lower levels of expression

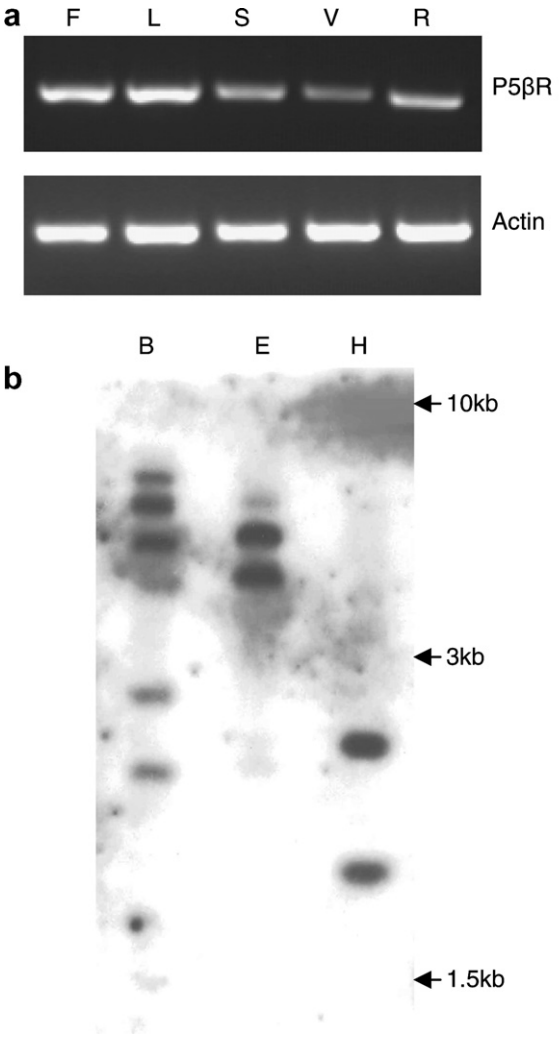

Fig. 6. Expression and genomic organisation of the *P5βR* gene. (a) RT–PCR analysis of *P5βR* expression in F (flowers), L (leaves), S (stems), V (vascular bundles) and R (roots). RT–PCR of the actin gene was used as a loading control. (b) Southern blot analysis of genomic DNA from *D. purpurea*. Samples of 20 μg DNA digested with *BamH*I (B), *Eco*RI (E) and *Hind*III (H) were loaded onto each lane. The blot was hybridised with the 850 bp cDNA probe isolated from *P5βR*.

were also detected in flowers, shoots and leaves of *D. lanata*. This expression pattern is unexpected if it is assumed that *3βHSD* is part of the general plant steroid metabolism. Nevertheless, these results are in agreement with the low levels of 3βHSD enzyme activity found in *D. lanata* leaves by Lindemann and Luckner (1997).

The results on the *3βHSD* gene and ours are the first data concerning the expression of genes involved in cardenolide biosynthesis. The data raise some questions about the sites of biosynthesis and accumulation of cardenolides in *Digitalis* plants. 3βHSD and P5βR are two consecutive enzymes of the putative pathway for cardenolide biosynthesis. From the expression patterns of their corresponding genes it can be hypothesised that the initial steps of this route, until progesterone formation (see Fig. 1), occur mainly in root cells, whereas the subsequent, more specific steps take place in aerial organs. This idea is also supported by the enzyme activity data of side chain cleaving enzyme (SCCE) in leaves gathered by Lindemann and Luckner (1997), who found that the activity of this enzyme, which produces pregnenolone from phytosterols, is even lower than that exhibited by 3βHSD. The connection between ground and aerial organs would be the long-distance transport of intermediary compounds, namely progesterone, via xylem. Confirmation of this hypothesis would require isolation of the genes encoding the phytosterol SCCE and the 3β-hydroxysteroid-5β-oxidoreductase (3βHS-5βOR) enzymes (see Fig. 1). Knowledge of these genes would open the possibility of investigating the spatial patterns and the coordinated regulation of the first steps in cardenolide biosynthesis.

2.5. Genomic complexity of the P5βR gene family

To evaluate the number of *P5βR*-related genes in *D. purpurea*, we made a DNA gel blot using *P5βR* as a probe (Fig. 6b). Genomic DNA was digested with either *BamH*I, *Eco*RI or *Hind*III and probed at high stringency. Two strong and two weaker hybridising bands were seen in *Eco*RI digests, two bands in digests with *Hind*III, and five or six bands with *BamH*I. Since there is a *BamH*I site within the *P5βR* open reading frame, these results indicate that this gene is a member of a small multi-gene family of presumptive two *P5βR* copies, and at least other two copies with related sequences to the *P5βR* gene exist in the genome of *D. purpurea*.

The number of genes coding for enzymes that catalyse the same reaction is an important factor of metabolic regulation. From the genomic Southern analysis, we assume that the organisation of genes encoding P5βR corresponds to that of ARs (aldose reductases) in *D. purpurea* (Gavidia et al., 2002a) where two copies of the gene encoding the two enzymes AR1 and AR2 exist with the same function but different enzyme activity. Other gene families involved in secondary metabolism have shown differential expression or changes in substrate specificity among the members of the gene family (Facchini and De Luca, 1995; Milkowski et al., 2004).

2.6. Origins and evolution of the P5βR gene

A comparison of all archaeal, bacterial, and eukaryotic complete and partial genomes in the public domain with the P5βR amino acid sequence indicates that the *P5βR* gene is phylogenetically circumscribed to higher plants (~70/80% average amino acid sequence identity/similarity to P5βR), where it is widespread, and isolated cases of distantly related, either facultative or obligate aerobic bacteria, including a few α and γ proteobacteria and one actinobacterium (~35/50%) (Fig. 7). There were no significant matches among archaea, protists, animals and fungi, but this result might, in part, reflect unequal representation of taxa in the sequence databases. Plant and bacterial

P5βRs clearly form two separate clusters, which suggests that the gene was already present in the last universal common ancestor – as opposed to having been transferred intracellularly to the eukaryotic nucleus from proto-organellar endosymbionts – and was subsequently lost in many lineages. The phylogenetic tree derived from P5βR places *Pseudomonas* more closely related to the α-proteobacterium *Aurantimonas* than to the remaining γ-proteobacteria; the bootstrap support for this clade (68%) is however low to justify a potential implication of horizontal gene transfer. It remains to be determined why the *P5βR* gene exhibits such a sparse phylogenetic distribution within prokaryotes.

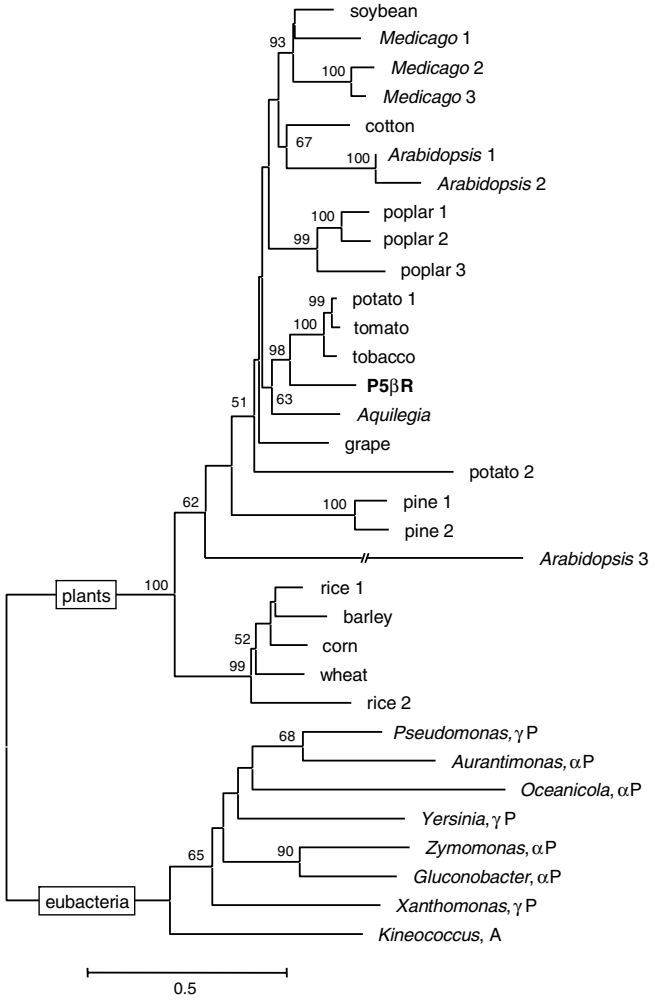

Fig. 7. Unrooted Neighbor-Joining (NJ) tree based on the Jones–Taylor–Thornton matrix distances assuming gamma-distributed rates among sites ($\alpha = 1.20 \pm 0.09$; referred to as JTT + dG model) for the P5βR amino acid data. The JTT + dG model was chosen among alternative models of protein evolution using the likelihood-ratio test criterion, adopting as initial hypothesis for model fitting the condensed distance tree topology obtained using *MEGA* version 3.1 (Kumar et al., 2004) with default parameters (see Rodríguez-Trelles et al., 2001). Bootstrap values (based on 1000 pseudo replications) greater than 50% are given on the respective nodes. Branch lengths are proportional to the corresponding numbers of amino acid replacements per site. Sequence data were gathered by translated BLAST (tblastn) against all complete and partial genomes currently available through the National Center for Biotechnology Information (NCBI), Department of Energy (DOE)/Joint Genome Institute (JGI), and The Institute for Genomic Research (TIGR) databases, using P5βR as the query and expect similarity threshold $E \leqslant 10^{-50}$.

The most extended gene functional annotation strategy is based on two steps: the first step is to detect a homologous relationship between pairs of proteins (e.g., using the BLAST tool). The second step is to infer functional similarity from homology. Recent computational studies (Tian and Skolnick, 2003; George et al., 2005) indicate that precise function (i.e., that specified in the four digits of an E.C. number) diverges below 60% sequence identity, and that functional inferences implying transference of the first three digits of an E.C. number are unwarranted below identities of 40%. Either way, it seems unlikely that bacterial P5βR carry the same function as in plants, since sterol biosynthesis is almost completely absent in prokaryotes (Pearson et al., 2003; Volkman, 2004). It is more plausible that P5βR was newly recruited for its function in *Digitalis* at some point along the branch leading to the diversification of higher plants in Fig. 7.

In contrast to plants, transformation of progesterone to 5β-pregnane-3,20-dione in vertebrates is accomplished by a member of the AKR superfamily (Onishi et al., 1991; Charbonneau and Luu-The, 2001), which takes the form of a $(\beta/\alpha)_8$ TIM barrel fold. AKRs are monomeric proteins, about 320 amino acids in length, which bind NAD(P)(H) to metabolise a range of substrates. Vertebrates have several families of AKRs which are all thought to be derived by duplication from a single AKR gene at the origin of the phylum (Seery et al., 1998; reviewed in Hyndman et al., 2003). A homolog of the vertebrate *P5βR* gene has been identified in *Digitalis*, but the encoded protein is unable to catalyse the 5β-reduction of progesterone (Gavidia et al., 2002a). The evidence coming from primary structure points to the conclusion that P5βR is not an AKR but a SDR protein. Therefore, P5βR illustrates a nice case of a metabolic function for which animals and plants have independently recruited unrelated genes (an AKR versus a SDR, in animals and plants, respectively) (McCarty and Chory, 2000; Meyerowitz, 2002; Suzuki et al., 2004; Ausubel, 2005). The case of P5βR stands in contrast with the situation uncovered for the gene responsible for the 5α-reduction of sterols – an intermediate step in the brassinolide biosynthesis pathway – which is highly conserved across animals and plants (Li et al., 1997).

In addition to P5βR, 3β-hydroxy-Δ^5-steroid dehydrogenase (3β-HSD) also exemplifies an activity for which animals and plants have recruited evolutionarily unrelated enzymes. 3β-HSD is a single bifunctional enzyme that catalyses the interconversion of pregn-5-ene-3β-ol-20-one (pregnenolone) into pregn-4-ene-3,20-dione (progesterone) in a two-sequential-steps reaction: dehydrogenation of pregnenolone, and isomerisation of the resulting isoprogesterone (pregn-5-ene-3,20-dione) to progesterone. Both in *Digitalis* and mammals, this reaction is placed immediately before the conversion of progesterone to 5β-pregnane-3,20-dione by P5βR. In mammals, 3β-HSD and P5βR activities

occur also uncoupled at other points of the steroid hormone metabolism [see biochemical pathway charts for E.C.:1.1.1.145 and E.C.:1.3.99.6 enzymatic activities, respectively, in the ExPASy proteomics server at http://au.expasy.org/enzyme/]. Mammals were recently shown to synthesise their own digitalis-like compounds from cholesterol (albeit in tiny amounts, compared to *Digitalis*) in the adrenal gland (Qazzaz et al., 2004; Hamlyn, 2004), likely via 3β-HSD (Lichtstein et al., 1998), although a definitive argument for the involvement of pregnenolone and progesterone as intermediaries has still to be made. As in the case of P5βR, plant 3β-HSDs are members of the SDR family (Finsterbusch et al., 1999), unlike mammalian 3β-HSDs, which are AKRs (Luu-The et al., 1991).

Recruitment of analogous enzymes that have evolved independently of one another (as opposed to homologous enzymes that have descended from a common ancestral enzyme) for the same reaction in different phylogenetic lineages seems to be more common than previously realised (Galperin et al., 1998). In most situations, analogous enzymes are sandwiched between those that are universally conserved. Acquisition of entire pathways catalysed by a set of unrelated enzymes seems to be rare (Galperin et al., 1998). That two critical, consecutive steps of the cardenolide pathway are catalysed by enzymes that are not homologous between animals and plants suggests that this biosynthetic route has evolved largely independently in the two kingdoms.

The *P5βR* phylogenetic tree shown in Fig. 7 reflects a history of duplications in higher plants. Particularly noticeable is the case of monocots, whose P5βR genes all derive from an ancient duplicate originated before the origin of gymnosperms [the bootstrap support for the bifurcation monocots/(gymnosperms, dicots) in Fig. 7 raises to 97% after removing the *Arabidopsis* 3 sequence]. Similarly, *Arabidopsis* has three copies, one of which (*Arabidopsis* 3) predates the diversification of dicots and has evolved at a faster rate, perhaps evidencing an instance of neofunctionalisation (see Rodríguez-Trelles et al., 2003; Benderoth et al., 2006). Gene duplication followed by functional diversification is assumed to drive the bewildering diversity of plant secondary metabolism. Tandem duplicates often remain linked, forming clusters within the plant genomes (Ober, 2005). The two only plant genomes in Fig. 7 that have been completed (those of *Arabidopsis* and rice) have their respective *P5βR* paralogs placed on different chromosomes.

3. Experimental

3.1. Plant materials

Shoot cultures of foxglove (*D. purpurea*) were established as previously described by Gärtner and Seitz (1993). Every three weeks newly developed shoots were transferred to fresh liquid nutrient medium (Murashige and Skoog, 1962) supplemented with 3% glucose, 1 mg/l indoleacetic acid and 2 mg/l kinetin. Cultures were maintained in a growth chamber under a 14 h photoperiod at 21 °C on a rotary shaker (90 rpm). Alternatively, the complete plantlets formed were transplanted into soil and acclimatised to ex vitro conditions. Plants were grown under standard greenhouse conditions at a day/night temperature regime of 20/18 °C.

3.2. Purification of native enzyme and amino acid sequence analysis

In vitro 11-day-old leaves were used for progesterone 5β-reductase extraction and purification according to the protocols previously described by Gärtner et al. (1994). The eluted protein (18 μg) after SDS/PAGE was digested with Tosyl-L-phenylalanyl chloromethane treated trypsin and the peptide mixture was resolved by reverse phase HPLC. The different fractions were collected and applied to TFA treated glass-filter discs coated with polybrene and sequenced in a protein sequencer 494A (Applied Biosystems, Weiterstadt, Germany). HPLC fractions were also analysed by MALDI-TOF mass spectrometry (G2025A, Hewlett–Packard, Waldbronn, Germany). Ten peptide fragments were sequenced.

3.3. Cloning and sequencing of the P5βR cDNA

Genomic DNA, isolated from*D. purpurea* leaves according to Dellaporta et al. (1983), was used as a template for PCR amplification. A fragment of 850 bp was obtained when using a pair of specific primers for *A. thaliana AWI31*, based on the amino acid sequences THVFYVTW and ISWIDKAK for P5βR. This fragment was ^{32}P-labelled and used as a probe for screening the *D. purpurea* cDNA library (Gavidia et al., 2002a) under high stringency conditions. Nylon filter lifts were prehybridised and hybridised at 60 °C in 330 mM sodium phosphate buffer (pH 7), 7% SDS, 1 mM EDTA and 1% BSA. The positive clones were isolated, their cDNA inserts in vivo excised and subcloned into pBluescript SK(−), and then sequenced using an ABI 310 (Perkin–Elmer, Germany). Since the expected cDNA was truncated at the 5′ terminus, 5′-RACE-PCR reactions were carried out using a SMART RACE cDNA amplification Kit (Clontech) according to the supplier's protocol.

3.4. Genomic library construction and screening

A *D. purpurea* genomic library was constructed in a λ vector (Lambda FIX II/Xho/partial fill-in vector kit, Stratagene) according to the manufacturer's protocol. The resulting library (ca. 3×10^{10} pfu/ml) was screened by plaque hybridisation with a ^{32}P-labelled fragment (850 bp) of the *P5βR* cDNA. Nylon filters were prehybridised and hybridised at 60 °C in 330 mM sodium phosphate buffer (pH 7), 7% SDS, 1 mM EDTA and 1% BSA. After overnight incubation, the membranes were washed twice for 15 min in 2xSSC with 0.1% SDS at 60 and 40 °C, consecutively, and exposed to X-ray on X-OMAT AR films (Kodak) using intensifying screens at −80 °C for 3–4 days. Isolation and purification of

bacteriophage λ were performed according to standard procedures (Sambrook et al., 1989) with some modifications. Nucleotide sequences were determined from both ends using universal primers or synthetic oligonucleotides designed from several positions in the cDNA.

3.5. RNA extraction and expression analysis by RT–PCR

RNA was extracted from roots, stems at the base of the rosette, vascular bundle at the basal part of the leaf, leaves and mature flowers collected from mature *D. purpurea* plants. Extraction was performed according to the method described by Steimle et al. (1994), but for root samples this method was modified as described by Roca-Pérez et al. (2004). Before reverse transcription, total RNAs were treated with DNase I (Takara) according to the manufacturer's protocol. Five micrograms of total RNA was used for cDNA synthesis. Reverse transcription was performed using 50 units of SuperScript II RT (Invitrogen) in the presence of 40 units of recombinant ribonuclease inhibitor (Invitrogen). After first-strand synthesis, the RNA template was removed by digestion with 2 units of RNase H (Invitrogen). cDNA derived from 0.5 μg RNA was used for PCR reaction. PCR was performed using two specific primers, 5′-CCTACGGGAAAATA-GAATCCC-3′ and 5′-ACACCACACAGTTG GTGAAC-3′, and the following program: 94 °C for 5 min, 30 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min, and finally incubated at 72 °C for 10 min. The PCR products were analysed on a 1% agarose gel.

3.6. Heterologous expression

The full-length open-reading frame of *P5βR* cDNA was amplified by PCR using the primers 5′-AAGCATGC-GAATGAGCTGGTGGTGG GC-3′ and 5′-GGCCA-GATCTAGGAACAATCTTG-3′ to introduce *Sph*I and *Bgl*II restriction sites, respectively. The amplified fragment was cloned into pQE70 QIAexpress vector (Qiagen). The recombinant plasmid was used to transform *E. coli* M15/PREP4. Preparation and purification of the recombinant enzyme was essentially performed as described in Gavidia et al. (2002a). After induction of gene expression by addition of IPTG the cultures were maintained at 37, 28 and 15 °C for 4 h.

The *P5βR* cDNA was also subcloned into *Xmn*I-*Pst*I sites of pMAL-c2 (New England Biolabs). Here it was expressed as a fusion protein with maltose-binding-protein (MBP) at the N-terminus. The *Pst*I-ended fragment was amplified using a set of primers, 5′-ATGAGCTGGT GGTGGGCT-3′ and 5′-TTTTTCTGCAGTCAAG-GAA-CAATCTTGTAAGCTTTTG-3′, and *Pfu* DNA polymerase (NEED). The constructed vector was introduced into *E. coli* BL21 (DE3, pLys3) and cultured in LB medium supplemented with glucose (2 g/l) and ampicillin (100 mg/l). Gene expression was induced by addition of 0.3 mM IPTG and then cultured at 15 °C for 4 h. The cells were harvested by centrifugation and disrupted by sonication in buffer A (20 mM Tris–HCl pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT and 0.2 mM PMSF). After centrifugation at 12,000*g* for 10 min, soluble protein extracts were obtained. The MBP-tagged P5βR protein was then purified in buffer B (buffer A with 10 mM maltose) by amylose resin column chromatography (New England Biolabs) according to the supplier's protocol. The electrophoretic separation of proteins was performed on 12% polyacrylamide gels according to Laemmli (1970).

3.7. Enzyme assays and GC–MS analysis

After purification of the recombinant proteins, they were concentrated and the buffer changed for buffer C (50 mM Tris–HCl pH 7.5, 250 mM sucrose, 2 mM EDTA, 1mM DTT) using Vivaspin concentrators (VIVASCIENCE). Protein was quantified by the method described by Bradford (1976). In the assays of P5βR activity, the standard reaction mixture consisted, according to Gärtner et al. (1994), of 6.4 mM NADP+, 32.1 mM glucose 6-phosphate, 42 nkat glucose 6-phosphate dehydrogenase, 270 μM progesterone and 100 μg of recombinant P5βR protein, in a final volume of 750 μl in buffer C. The assays were incubated for 3 h at 30 °C and the reaction was terminated by adding CH_2Cl_2. The extraction and purification of the pregnanes were carried out as described by Gärtner and Seitz (1993). All steroid substrates employed were purchased from Sigma with the highest purity available.

Pregnanes were analysed by GC–MS using an Agilent 5973N mass spectrometer connected to an Agilent 6890N gas chromatograph with capillary column 19091S-4331 HP-5MSI (30 m long, 0.25 mm i.d. with 0.25 μm film thickness). Splitless injections were made at an injector temperature of 250 °C. The oven was programmed from 150 to 280 °C at 5 °C min^{-1} with a constant flow of 1 ml min^{-1} He. Mass spectra were compared with those of authentic standards.

Acknowledgements

We are grateful to Drs. Axel Hirner, Stefan Veit and Wolfgang Koch for helpful discussions. The European Commission supported this work with postdoctoral grants to I.G. (Contracts BIO4-CT97-5019 and QLK3-CT-1999-51296). R.T and F.R-T. have received support from contracts F.I.S. G03/184 and Ramón y Cajal, from the Ministerio de Sanidad y Consumo and Ministerio de Ciencia y Tecnología (Spain), respectively. This research was funded in part by the Generalitat Valenciana (Project GV05/127).

References

Andreeva, A., Howorth, D., Brenner, S.E., Hubbard, T.J., Chothia, C., Murzin, A.G., 2004. SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res. 32, D226–D229.

Ausubel, F.M., 2005. Are innate immune signaling pathways in plants and animals conserved?. Nat. Immunol. 6 973–979.

Benderoth, M., Textor, S., Windsor, A.J., Mitchell-Olds, T., Gershenzon, J., Kroymann, J., 2006. Positive selection driving diversification in plant secondary metabolism. Proc. Natl. Acad. Sci. USA 103, 9118–9123.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.

Charbonneau, A., Luu-The, V., 2001. Genomic organization of a human 5β-reductase and its pseudogene and substrate selectivity of the expressed enzyme. BBA-Gene Struct. Expr. 1517, 228–235.

Dellaporta, S.L., Wood, J., Hicks, J.B., 1983. A plant DNA-minipreparation: Version II. Plant Mol. Biol. Res. 1, 19–21.

Facchini, P.J., De Luca, V., 1995. Phloem-specific expression of tyrosine/dopa decarboxylase genes and the biosynthesis of isoquinoline alkaloids in opium poppy. Plant Cell 7, 1811–1821.

Finsterbusch, A., Lindemann, P., Grimm, R., Eckerskorn, C., Luckner, M., 1999. Δ^5-3β-Hydroxysteroid dehydrogenase from *Digitalis lanata* Ehrh.- a multifunctional enzyme in steroid metabolism? Planta 209, 478–486.

Galperin, M.Y., Walker, D.R., Koonin, E.V., 1998. Analogous enzymes: independent inventions in enzyme evolution. Genome Res. 8, 779–790.

Gärtner, D.E., Seitz, H.U., 1993. Enzyme activities in cardenolide-accumulating, mixotrophic shoot cultures of *Digitalis purpurea* L. J. Plant Physiol. 141, 269–275.

Gärtner, D.E., Keilholz, W., Seitz, H.U., 1994. Purification, characterization and partial peptide microsequencing of progesterone 5β-reductase from shoot cultures of *Digitalis purpurea*. Eur. J. Biochem. 225, 1125–1132.

Gärtner, D.E., Wendroth, S., Seitz, H.U., 1990. A stereospecific enzyme of the putative biosynthetic pathway of cardenolides. Characterization of a progesterone 5β-reductase from leaves of *Digitalis purpurea*. L. FEBS Lett. 271, 239–242.

Gavidia, I., Pérez-Bermúdez, P., Seitz, H.U., 2002a. Cloning and expresión of two novel aldo-keto reductases from *Digitalis purpurea* leaves. Eur. J. Biochem. 269, 2842–2850.

Gavidia, I., Seitz, H.U., Pérez-Bermúdez, P., Vogler, B., 2002b. LC-NMR applied to the characterisation of cardiac glycosides from three micropropagated *Isoplexis* species. Phytochem. Anal. 13, 266–271.

George, R.A., Spriggs, R.V., Bartlett, G.J., Gutteridge, A., MacArthur, M.W., Porter, C.T., Al-Lazikani, B., Thornton, J.M., Swindells, M.B., 2005. Effective function annotation through catalytic residue conservation. Proc. Natl. Acad. Sci. USA 102, 12299–12304.

Gough, J., Hughey, R., Karplus, K., Chothia, C., 2001. Assignment of genome sequences using a library of hidden Markov models that represent all proteins of known structure. J. Mol. Biol. 313, 903–919.

Hamlyn, J.M., 2004. Biosynthesis of endogenous cardiac glycosides by mammalian adrenocortical cells: three steps forward. Clin. Chem. 50, 469–470.

Harborne, J.B., 1982. Introduction to Ecological Biochemistry, Second ed. Academic Press, New York.

Herl, V., Fischer, G., Müller-Uri, F., Kreis, W., 2006a. Molecular cloning and heterologous expression of progesterone 5β-reductase from *Digitalis lanata* Ehrh. Phytochemistry 67, 225–231.

Herl, V., Fischer, G., Müller-Uri, F., Kreis, W., 2006b. Erratum to "Molecular cloning and heterologous expression of progesterone 5β-reductase from *Digitalis lanata* Ehrh" [Phytochemistry 67 (3) (2006) 225–231]. Phytochemistry 67, 1554.

Hyndman, D., Bauman, D.R., Heredia, V.V., Penning, T.M., 2003. The aldo-keto reductase superfamily homepage. Chem. Biol. Inter. 143–144, 621–631.

Kallberg, Y., Oppermann, U., Jörnvall, H., Persson, B., 2002. Short-chain dehydrogenases/reductases (SDRs). Coenzyme-based functional assignment in completed genomes. Eur. J. Biochem. 269, 4409–4417.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.

Li, J., Biswas, M.G., Chao, A., Russell, D.W., Chory, J., 1997. Conservation of function between mammalian and plant steroid 5α-reductases. Proc. Natl. Acad. Sci. USA 94, 3554–3559.

Li, J., Nagpal, P., Vitart, V., McMorris, T.C., Chory, J., 1996. A role for brassinosteroids in light-dependent development of *Arabidopsis*. Science 272, 398–401.

Lichtstein, D., Steinitz, M., Gati, I., Samuelov, A., Deutsch, J., Orly, J., 1998. Biosynthesis of digitalis-like compounds in rat adrenal cells: hydroxycholesterol as possible precursor. Life Sci. 62, 2109–2126.

Lindemann, P., Finsterbusch, A., Pangert, A., Luckner, M., 2000. Partial cloning of a Δ^5-3β-hydroxysteroid dehydrogenase from *Digitalis lanata*. In: Okamoto, M., Ishimura, Y., Nawata, H. (Eds.), Molecular Steroidogenesis, Proceedings of the Yamada Conference LII, Frontiers Science Series, 29-XXIV. Universal Academy Press, Tokyo, pp. 333–334.

Lindemann, P., Luckner, M., 1997. Biosynthesis of pregnane derivatives in somatic embryos of *Digitalis lanata*. Phytochemistry 46, 507–513.

Luu-The, V., Takahashi, M., de Launoit, Y., Dummont, M., Lachance, Y., Labrie, F., 1991. Evidence for distinct dehydrogenase and isomerase sites within a single 3-beta-hydroxysteroid dehydrogenase 5-ene-4-ene isomerase protein. Biochemistry 30, 8861–8865.

Malcolm, S.B., Zalucki, M.P., 1996. Milkweed latex and cardenolide induction may resolve the lethal plant defence paradox. Entomol. Exp. Appl. 80, 193–196.

McCarty, D.R., Chory, J., 2000. Conservation and innovation in plant signalling pathways. Cell 103, 201–209.

Meyerowitz, E.M., 2002. Plants compared to animals: the broadest comparative study of development. Science 295, 1482–1485.

Milkowski, C., Baumert, A., Schmidt, D., Nehlin, L., Strack, D., 2004. Molecular regulation of sinapate ester metabolism in *Brassica napus*: expression of genes, properties of the encoded proteins and correlation of enzyme activities with metabolite accumulation. Plant J. 38, 80–92.

Murashige, T., Skoog, F., 1962. A revised medium for rapid growth and bioassay with tobacco tissue culture. Physiol. Plant. 15, 473–496.

Murzin, A., Brenner, S.E., Hubbard, T.J.P., Chothia, C., 1995. SCOP: a Structural Classification of Proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540.

Ober, D., 2005. Seeing double: gene duplication and diversification in plant secondary metabolism. Trends Plant Sci. 10, 444–449.

Onishi, Y., Noshiro, M., Shimosato, T., Okuda, K., 1991. Molecular cloning and sequence analysis of cDNA encoding Δ^4-3-ketosteroid 5β-reductase of rat liver. FEBS Lett. 2, 215–218.

Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X.Q., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B., Jornvall, H., 2003. Short-chain dehydrogenases/reductases (SDR): the 2002 update. Chem. Biol. Interact. 143, 247–253.

Pearson, A., Budin, M., Brocks, J.J., 2003. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. Proc. Natl. Acad. Sci. USA 100, 15352–15357.

Penning, T.M., Ma, H., Jez, J.M., 2001. Engineering steroid hormone specificity into aldo-keto reductases. Chem. Biol. Interact. 130, 659–671.

Pilgrim, H., 1972. "Cholesterol side-chain cleaving enzyme" Aktivität in Keimlingen und in vitro kultivierten Geweben von *Digitalis purpurea*. Phytochemistry 11, 1725–1728.

Qazzaz, H.M.A.M., Cao, Z., Bolanowski, D.D., Clark, B.J., Valdes, R., 2004. De novo biosynthesis and radiolabeling of mammalian digitalis-like factors. Clin. Chem. 50, 612–620.

Roca-Pérez, L., Boluda, R., Gavidia, I., Pérez-Bermúdez, P., 2004. Seasonal cardenolide production and *Dop5βr* gene expression in natural populations of *Digitalis obscura*. Phytochemistry 65, 1869–1878.

Rodríguez-Trelles, F., Tarrío, R., Ayala, F.J., 2001. Erratic overdispersion of three molecular clocks; GPDH, SOD, and XDH. Proc. Natl. Acad. Sci. USA 98, 11405–11410.

Rodríguez-Trelles, F., Tarrío, R., Ayala, F.J., 2003. Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the *xanthine dehydrogenase* gene. Proc. Natl. Acad. Sci. USA 100, 13413–13417.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Seery, L.T., Nestor, P.V., FitzGerald, G.A., 1998. Molecular evolution of the aldo-keto reductase gene superfamily. J. Mol. Evol. 46, 139–146.

Seitz, H.U., Gärtner, D.E., 1994. Enzymes in cardenolide-accumulating shoot cultures of *Digitalis purpurea* L. Plant Cell Tiss. Org. Cult. 38, 337–344.

Steimle, D.E., Schnitzler, J.P., Seitz, H.U., 1994. Modes of expression of phenylalanine ammonia-lyase and chalcone synthase in elicitor-treated carrot cell cultures. Acta Hort. 381, 206–209.

Stuhlemmer, U., Kreis, W., 1996. Cardenolide formation and activity of pregnane-modifying enzymes in cell suspension cultures, shoot cultures and leaves of *Digitalis lanata*. Plant Physiol. Biochem. 34, 85–91.

Suzuki, H., Sawada, S., Watanabe, K., Nagae, S., Yamaguchi, M., Nakayama, T., Nishino, T., 2004. Identification and characterization of a novel anthocyanin malonyltransferase from scarlet sage (*Salvia splendens*) flowers: an enzyme that is phylogenetically separated from other antocyanin acyltransferases. Plant J. 38, 994–1003.

Tian, W., Skolnick, J., 2003. How well is enzyme function conserved as a function of pairwise sequence identity? J. Mol. Biol. 333, 863–882.

Vogel, E., Luckner, M., 1981. Distribution of cardenolides in *Digitalis lanata*. Planta Med. 41, 161–165.

Volkman, J.K., 2004. Sterols and other triterpenoids: source specificity and evolution of biosynthetic pathways. Org.Geochem. 36, 139–159.

Yang, K.Y., Moon, Y.H., Choi, K.H., Kim, Y.H., Eun, M.Y., Guh, J.O., Kim, K.C., Cho, B.H., 1997. Structure and expression of the *AWI 31* gene specifically induced by wounding in *Arabidopsis thaliana*. Mol. Cells 7, 131–135.