Intramolecular proton transfer in the cyclization of geranylgeranyl diphosphate to the taxadiene precursor of taxol catalyzed by recombinant taxadiene synthase

David C Williams¹, Brian J Carroll², Qingwu Jin³, Christopher D Rithner⁴, Steven R Lenger⁴, Heinz G Floss², Robert M Coates³, Robert M Williams⁴ and Rodney Croteau¹

Background: The committed step in the biosynthesis of the anticancer drug taxol in yew (Taxus) species is the cyclization of geranylgeranyl diphosphate to taxa-4(5),11(12)-diene. The enzyme taxadiene synthase catalyzes this complex olefin cation cyclization cascade involving the formation of three rings and three stereogenic centers.

Results: Recombinant taxadiene synthase was incubated with specifically deuterated substrates, and the mechanism of cyclization was probed using MS and NMR analyses of the products to define the crucial hydrogen migration and terminating deprotonation steps. The electrophilic cyclization involves the ionization of the diphosphate with closure of the A-ring, followed by a unique intramolecular transfer of the C11 proton to the re-face of C7 to promote closure of the B/C-ring juncture, and cascade termination by proton elimination from the β-face of C5.

Conclusions: These findings provide insight into the molecular architecture of the first dedicated step of taxol biosynthesis that creates the taxane carbon skeleton, and they have broad implications for the general mechanistic capability of the large family of terpenoid cyclization enzymes.

Introduction

The diterpenoid taxol [1] (paclitaxel, Figure 1) is now well-established as a potent chemotherapeutic agent, showing excellent activity against a range of cancers and is currently approved for treating refractory ovarian, metstatic breast, and non-small-cell lung cancers, as well as Kaposi’s sarcoma [2]. (Paclitaxel is the generic name for Taxol, a registered trademark of Bristol-Myers Squibb. Because of its greater familiarity, ‘taxol’ is used throughout.) The limited supply of the drug from the original source, the bark of the Pacific yew (Taxus brevifolia Nutt.; Taxaceae), prompted intensive efforts to devise alternative means of production [3,4]. These efforts have yielded a commercially viable semisynthesis of taxol and its analogs from advanced taxane diterpenoid (taxoid) metabolites that are more readily available from yew [5,6]. However, with increasing applications in chemotherapy, both in treatment of additional cancer types and for earlier disease interven-
tion, the supply and cost of these drugs will remain important issues [7]. Total syntheses of taxol have been achieved by several elegant routes [8–13] but the yields are too low to be practical, and it is clear that in the foreseeable future the supply of taxol and its synthetically useful progenitors must rely on biological methods of production [7]. The development of improved biological processes must be based upon a detailed understanding of the pathway for taxol biosynthesis, the enzymes which catalyze the sequence of reactions and their mechanisms of action, and the genes encoding these enzymes, especially those responsible for slow steps of the pathway.

Taxadiene synthase from Taxus species catalyzes the first committed step in the biosynthesis of taxol and related taxoids by the cyclization of the universal diterpenoid precursor (E,E,E)-geranylgeranyl diphosphate (GGPP, 2) to the parent olefin (6), which undergoes an extended series
of oxygenation and acylation reactions (Figure 1) [14,15]. This enzyme, which has been isolated from both yew saplings [16] and cell cultures [17], catalyzes a slow, but apparently not rate limiting, step in the taxol biosynthetic pathway [17]. Initial assessment of the mechanism of taxadiene synthase [18] indicates that the reaction involves the ionization and cyclization of GGPP to a transient verticillyl intermediate (3), proposed to have the 11R configuration to allow intramolecular transfer of the C11 proton to C7 (4) to initiate transannular B/C-ring closure to the taxenyl cation (5), followed by deprotonation at C5 to yield taxa-4(5),11(12)-diene (6). No other details of the stereochemical mechanism of this novel enzyme are known.

A cDNA encoding the taxadiene synthase from *T. brevifolia* has been obtained by a homology-based PCR cloning method [19]. A truncated version of the enzyme, in which the plastidial targeting peptide of the preprotein has been deleted, has been functionally overexpressed in *Escherichia coli* and shown to resemble the native enzyme in kinetic properties [20]. This advance has made available sufficient amounts of the enzyme, and thus the enzyme product, to permit a more detailed study of the electrophilic cyclization reaction cascade.

**Results and discussion**

**Chain termination: C5-deprotonation**

To examine the stereochemistry of the C5-deprotonation of the terminal taxenyl C4-carbocationic intermediate of the reaction sequence (5, Figure 1), the appropriate deuterium labeled substrate, (4R)-[4-2H1]GGPP (7, Figure 2) (84 mol% 2H), was prepared and incubated at preparative scale with the purified recombinant taxadiene synthase. The olefinic products of the reaction were isolated and subjected to GC-MS analysis for comparison to the prod-

---

**Figure 1. Cyclization of geranylgeranyl diphosphate to taxol.**

**Figure 2. Cyclization of deuterium labeled GGPP.**
ucts identically obtained from the recombinant enzyme using unlabeled substrate. The taxa-4(5),11(12)-diene (6) enzymatically synthesized from the deuterium labeled acyclic precursor bore no detectable deuterium compared to the product prepared from the unlabeled precursor, thereby indicating that the C4 β-hydrogen is selectively and entirely lost in the terminating deprotonation step (see Figure 2). Interestingly, the use of the alternate substrate also permitted observation of the phenomenon of isotopically sensitive branching [21,22], in which the deuterium-dependent slowing of the C5-deprotonation of the common taxenyl cation intermediate promoted alternative deprotonations for terminating the reaction cycle. With the unlabeled substrate, taxadiene synthase yields principally taxa-4(5),11(12)-diene (6, 94%), with lesser amounts (V5%) of the 4(20),11(12)-isomer (8) resulting from deprotonation from the C20 methyl group, and negligible proton loss from the bridgehead methine (C3) to give the 3(4),11(12)-isomer (9) [20]. With the deuterated substrate, the proportion of taxa-4(5),11(12)-diene (6) in the enzymatic cyclization products decreased to 77.6%, with a concomitant increase in the proportions of the 4(20),11(12)-isomer (8) to 11.1% and of the 3(4),11(12)-isomer (9) to 9.2%; the latter two isomers fully retained the deuterium at C5.

Although there was relatively little change in the absolute rate of formation of the 4(20)-isomer, partitioning of the taxenyl carbocation to the 3(4)-isomer was substantially enhanced in response to the primary deuterium kinetic isotope effect on the C5 methylene deprotonation. An overall rate suppression resulting from the deuterium substitution on total olefin formation of about 50% was also observed due to C10 deuterium substitution, yielding k_H/k_D = 2.0 similar to that previously noted for the C5-deprotonation). Evaluation of the mass spectrum revealed that >99% of the deuterium originally present at C10 of GGPP (C11 of the intermediate) had migrated to C7 of the taxane skeleton as assessed by shift of the C-ring fragment ion at m/z 122 to m/z 123 (and of the parent ion from m/z 272 to m/z 273), thereby demonstrating the complete fidelity of the intramolecular proton transfer.

The olefin fraction of the enzymatic reaction products was purified by a combination of open column silica gel chromatography (with pentane) and reversed phase (C18) HPLC (with acetonitrile) to afford ~10 µg of pure C7-deuterated taxa-4(5),11(12)-diene (6) following transfer to deuterobenzene and removal of residual pentane used to partition the product from the HPLC solvent. The complete assignment of the 1H NMR spectrum of taxa-4(5),11(12)-diene in CDCl3 has been described previously [14] and these assignments have been confirmed in C6D6 for the present work. Since earlier work had indicated that deuterium from [10-^2H1]GGPP should reside at C7 of taxadiene following transfer from C11 of the verticillyl intermediate [18], the assignments most relevant to defining the regiochemistry of the product include those for the C-ring H5, H6 (α and β), H7 (α and β), and the H19 methyl protons.

**Figure 3.** Stereochemistry of intramolecular hydrogen migration.
The relatively small amount of biosynthetic [2H]taxadiene available (\(~10\) µg) precluded the use of two-dimensional NMR techniques. Consequently, 1D DPFGSE-TOCSY [23] experiments were done to explore ways of discriminating among the various spin systems in the olefin, with the hope that an unambiguous excitation of the H7 protons would result. The presence of H5 at \(5.38\) ppm, alone in this region of the spectrum, provided an entry point to define the C-ring system. Figure 4 illustrates the results of the 1D DPFGSE-TOCSY experiment, following selective excitation of H5 and a 65 ms mixing period. The most notable features of this spectrum are the doublet of doublets at \(1.19\) ppm (assigned to H7β), the multiplet at \(1.76\) (H7α), the singlet at \(1.73\) (C20 methyl), and the broad triplet and doublet at \(2.11\) and \(1.94\) (H6L and H6K, respectively). Since the interpretation of the result of deuterium incorporation into taxadiene rests upon the correct stereochemical assignments for the C7 protons, it was imperative to confirm these assignments.

A two-dimensional NOESY-NMR experiment (with a 400 ms mixing time) was conducted for this purpose (Figure 5). Through-space correlations among protons on the β- (top) face of the molecule were examined. Cross-peaks were observed correlating the diagonal signal at \(0.93\) ppm (C19 methyl) with signals at \(1.19\) (C17 methyl), at \(1.39\) and \(1.81\) (H9L, H9K), \(2.11\) (H6L), \(1.72\) (H2L), and \(1.19\) (H7L). No cross-peak was observed between the C19 signal and the signal at \(1.76\) ppm, which is consistent with the latter assignment as the H7α proton. Correlations with the diag-

---

**Figure 4.** Double pulsed field gradient spin echo-selective TOCSY of 6 (C6D6) produced from 10(b) and of unlabeled 6(a). The H5 signal at \(5.38\) was selectively excited in each case followed by a 65 ms mixing period. The predominant signals following the mixing period are those of the C-ring. Note the presence of the H7α proton signal in a at \(1.76\) but its absence in b following deuterium atom transfer. Other features to note are the collapse of the H7β, 6α, and 6β proton fine structure in b as compared with a.
onal signal for the H7\(\beta\) proton (\(\delta1.19\) ppm) were carefully examined. As expected, there was a strong cross-peak found at \(\delta1.76\) ppm assigned to the H7\(\alpha\) proton. There were also other cross-peaks at \(\delta0.93\) (C19 methyl), \(\delta1.39\) (H9\(\alpha\)), \(\delta1.94\) and \(\delta2.11\) (H6\(\alpha\) and H6\(\beta\), respectively). In contrast, the signal assigned to the H7\(\alpha\) proton (\(\delta1.76\) ppm) had fewer correlations, including those for H3 (\(\delta2.62\)) and the C18 methyl (\(\delta1.60\)), which is consistent with the stereochemical assignment to the \(\alpha\)- (bottom) face of the olefin.

With completion of the regio- and stereochemical NMR assignments, the enzymatically derived, deuterium labeled taxadiene sample was examined. Residual pentane (from the final solvent partitioning step) was removed from the sample by transfer to and repeated evaporation from \(\text{C}_{6}\text{D}_{6}\); otherwise, pentane signals completely dominate the 1D \(^1\text{H}\) NMR spectrum in the crucial \(\delta0.5-\delta1.5\) ppm region (Figure 4). By this means, the pentane signal intensities were reduced sufficiently to permit repetition of the 1D DPFGSE-TOCSY experiment, this pulsed field gradient technique being especially efficient at de-phasing those signals not correlated with the selected proton resonance. The 1D DPFGSE-TOCSY spectrum of the deuterium labeled olefin, with selective excitation of the H5 proton at \(\delta5.38\) ppm, is clear of artifacts including residual pentane signals (Figure 4b) and appears remarkably similar to that of the unlabeled standard (Figure 4a) with a few important exceptions. The signal for the H7\(\alpha\) proton (1.76 ppm), clearly present in the standard, has disappeared in the labeled olefin. The signal at \(\delta1.19\) ppm, assigned to the H7\(\beta\) proton, has collapsed to an apparent doublet (\(J=6\) Hz) consistent with the loss of its geminal H7\(\alpha\) partner. Finally, the NMR patterns for both of the H6 protons have been altered. These results are entirely consistent with transfer of the deuterium atom from C10 of the acyclic

**Figure 5.** NOESY NMR spectrum of 6 (\(\text{C}_{6}\text{D}_{6}\)). Inset a is a 500 MHz two-dimensional NOESY spectrum. Insets b and c are slices of the 2D spectrum taken through the methyl H19 and H7\(\beta\) positions, respectively.
precursor exclusively to the H7α position of the taxadiene product.

**Transition state modeling**

Modeling of the intermediate C12 carbocation resulting from the C11–C15 closure in the initially formed macrocycle (Figure 6), with minimization of this structure using MacSpartan (AM1 basis set, overall molecular charge of +1, trivalent carbon at C12), reveals that the C11 proton is within ~2.2 Å of the C7 carbon and seemingly perfectly poised for the transannular migration to the rα-face at C7. Furthermore, inspection of the conformation of the C12 carbocation strongly suggests that it is unlikely that an amino acid at the enzyme active site could gain access to the C11 proton to mediate the C11–C7 transfer, since this proton is buried deep within the concave face of the 12-membered macrocycle. Based on both the experimental evidence and the conformational analysis, it appears that an unassisted intramolecular transfer is the most plausible mechanism for the initiation of the final ring closure step.

**Figure 6.** Molecular modeling of the C11 to C7 hydrogen migration converting the verticillen-12-yl carbocation to its 8-yl isomer prior to cyclization of the C-ring.

**Figure 7.** Synthesis of all trans-(4R)-[4-2H1]GGPP (8).
mediated by taxadiene synthase. Thus, the C7–C8 olefin serves as the Brønsted base that quenches the incipient carbocation at C12 and no active site enzyme base needs to be invoked for this process.

Of further interest in this enzymatic reaction is the facial bias of the final C8–C3 olefin cation cyclization. The established stereochimistry of the taxane B/C-ring junction dictates attack by the C3/C4 π-electrons on the C8 cation (formed immediately following the C11 to C7 proton migration) from the same face to which the newly installed proton at C7 has migrated (overall syn addition of H+ and C3 to the 7,8 double bond). Modeling of this process (Figure 6), which formally involves a criss-cross of bond formation across the C7/C8 double bond, proved to be insightful. The molecular model of the C8 carbocation suggests that, upon pyramidalization at C7 and fashioning of the C11/C12 olefin, the conformation of the 12-membered ring twists, relative to the C12-cationic center, thereby rocking the C3/C4 π-electrons into the correct facial orientation for capture of the C8 cation and establishing the trans-fused B/C-ring juncture.

There is precedent in enzymatic terpenoid cyclization reactions for apparently similar deprotonation–reprotonation steps involving transient formation of olefinic intermediates, with reprotonation of a distal double bond to initiate subsequent rearrangements or ring closures. For example, in the diterpene series, an intramolecular proton transfer in subsequent rearrangements or ring closures. For example, in the diterpene series, an intramolecular proton transfer in the diterpene series, an intramolecular proton transfer in

A remarkable olefin cation cascade

The overall stereochemical mechanism for the conversion of GGPP to taxadiene [45(11)-diene (Figure 3) is shown to involve the following five processes: (1) Walden inversion at the diphosphate-bearing carbon (C1) coupled with anti-periplanar C1–C14 and C11–C15 bond formation initiated by the ionization of the diphosphate group and leading to the verticillen-12-yl carbocation (3) (the stereochemical outcome at C1 and C15 was elucidated by separate enzymatic cyclizations of (R)-[1-3H1]- and (S)-[1-3H1]-GGPP and 600 MHz 1H NMR analyses of the resulting [2-3H1]- and [16,16,16-2H3]taxadienes; unpublished results, D.C.W., Q.J., R.M.C., R.C.); (2) conformational inversion of the A-ring accompanied by rotations of the macrocyclic loop; (3) a unique intramolecular transfer of the C11 proton to the n-face of C7 to generate the isomeric verticillen-8-yl carbocation; (4) bond rotations and closure of the B/C-ring juncture via capture of the C8 carbocation by the transannular C3/C4 π-electrons to form the penultimate taxen-4-yl intermediate; and (5) elimination of the 5-B proton from a twist boat conformer to form taxadiene.

Significance

Taxadiene synthase is a remarkable terpene cyclase that appears to function by binding and ionization its substrate GGPP to mediate an enantio- and face-selective polylefin cation cascade that involves the formation of three carbon–carbon bonds, three stereogenic centers, and the loss of hydrogen in ‘a single step’. The seemingly unassisted intramolecular proton transfer mechanism of taxadiene synthase is thus strikingly unusual in this regard, and suggests that this enzyme type is capable of mediating complex olefin cation cyclizations, with absolute stereochemical fidelity, by conformational control alone. These findings provide insight into the molecular architecture of the first dedicated step of taxol biosynthesis that creates the taxane carbon skeleton, and these observations have broad implications for the general mechanistic capability of the large family of terpenoid cyclization enzymes that catalyze similar electrophilic reaction cascades.

Materials and methods

Substrate preparation

All trans-GGPP (2, 98%) and [10-2H1]-GGPP (10, >90%; >99 mol% 2H) were prepared and purified as previously described [18]. All trans-(R)-[4-2H1]-GGPP (8, 93%; 84 mol% 2H) was prepared from previously described [28] ester A (25:1 E:Z ratio by 1H NMR) and the known [29–31] sulphone F (Figure 7). Ester A was reduced with LiA2H4 to dideuterated alcohol B, oxidized to monodeuterated aldehyde C by the Swern method, and reduced asymmetrically to D with (S)-alpine borane. Camphorate derivatization and 1H NMR analysis showed a 95:5 enantiomeric ratio for D [32]. Following conversion to the mesylate E and coupling [33] to sulphone F, the product G was obtained in 50% yield after purification. The sulphone group and benzyl ether protecting group were removed by reduction with Li in ENH2 at −78°C to give (R)-[4-2H1]-geranylgeranol H in 29% yield following AgNO3 silica gel argentation chromatography to remove a contaminating double bond isomer. To determine the enantiomeric purity of H, the asymmetric epoxidation [34] of the 2,3 double bond was carried out to differentiate the two protons adjacent to the epoxide ring by 1H NMR analysis. Integration of the relevant peaks at 1.51 and 1.35 ppm showed a ratio of 1:8 which, by assuming the same enantioselectivity as in the epoxidation of geran- 90%; alcohol H was converted to the corresponding diphosphate ester (30% yield) by established procedures [35,36]. A single component was observed upon cellulose TLC and no impurities were observed in the 1H and 31P NMR spectra; the latter (D2O, 162 MHz) displayed two doublets at −9.54 and −5.72 ppm (J=20.8 Hz).

Enzymatic conversions

The truncated (M60) version of recombinant taxadiene synthase, in which the plastidial transit peptide has been deleted, was overexpressed in E. coli and purified (>96%) as previously described [20]. Incubations were carried out in 3 ml of the standard assay buffer [20] containing 1 mg of enzyme and a 2 mM concentration of the appropriate substrate. After immediate extraction with pentane to remove any organ- ic soluble contaminants, the reaction mixture was maintained at 31°C until in excess of 100 μg of olefin product had been generated. Following incubation, the pentane-soluble reaction products were extracted for capillary GC-MS analysis [20] to establish overall conversion, the distri-
bution of taxadiene isomers, and the deuterium content of these olefinic products. In preparation for the NMR analysis of the product derived from [10-3H]-GGPP (10), the olefin mixture was purified by silica gel column chromatography with hexane, followed by reversed phase (C18) HPLC with acetonitrile. The concentrated acetonitrile eluent was partitioned between pentane and water, and the pentane phase was transferred to deuterobenzene (99.96 mol%) and repeatedly concentrated under N2 (without drying which results in decomposition [14]) to remove residual solvent, yielding ~10 μg of pure taxa-4(5),11(12)-dien (6) in a 500 μl sample volume.

MS and NMR spectrometry

Protocols for capillary GC-MS analysis of the taxadiene isomers have been described in detail elsewhere [14,18,20]. All NMR spectra were recorded on a Varian Unity-500 spectrometer at 25°C using a very sensitive 1H indirect detection probe. 1D double pulsed field gradient spin echo TOCSY (DPFGSE-TOCSY) spectra were obtained with a mixing time of 65 ms. 2D NOESY spectra were acquired using mixing times of 400 ms. 2D spectra were collected as 256 (f1) × 4096 (f2) complex points with a sweep width of 7 kHz in each dimension. Data were processed using Varian, INC. NMR software. The final data size, after linear prediction in f1 and zero-filling in both dimensions, was 1024 (f1) × 1024 (f2) complex points.

Acknowledgements

The authors gratefully acknowledge financial support from the National Institutes of Health.

References


