Functional Analysis of the Validamycin Biosynthetic Gene Cluster and Engineered Production of Validoxylamine A

Linquan Bai,1 Lei Li,1,4 Hui Xu,1,4 Kazuyuki Minagawa,2,4 Yi Yu,1, Yirong Zhang,1 Xiufen Zhou,1 Heinz G. Floss,3 Taifo Mahmud,2,* and Zixin Deng†**

1Laboratory of Microbial Metabolism and School of Life Science & Biotechnology Shanghai Jiaotong University Shanghai 200030 China
2Department of Pharmaceutical Sciences Oregon State University Corvallis, Oregon 97331
3Department of Chemistry University of Washington Seattle, Washington 98195

Summary

A 45 kb DNA sequencing analysis from Streptomyces hygroscopicus 5008 involved in validamycin A (VAL-A) biosynthesis revealed 16 structural genes, 2 regulatory genes, 5 genes related transport, transposition/integration or tellurium resistance; another 4 genes had no obvious identity. The VAL-A biosynthetic pathway was proposed, with assignment of the required genetic functions confined to the sequenced region. A cluster of eight reassembled genes was found to support VAL-A synthesis in a heterologous host, S. lividans 1326. In vivo inactivation of the putative glycosyltransferase gene (valG) abolished the final attachment of glucose for VAL-A production and resulted in accumulation of the VAL-A precursor, validoxylamine, while the normal production of VAL-A could be restored by complementation with valG. The role of valG in the glycosylation of validoxylamine to VAL-A was demonstrated in vitro by enzymatic assay.

Introduction

Validamycin A (VAL-A; Figure 1A), a weakly basic C7N-aminocyclitol-containing antibiotic, was first isolated from Streptomyces hygroscopicus var. limonius, and later from S. hygroscopicus var. jingangensis 5008 (hereafter S. hygroscopicus 5008 or strain 5008). It is widely used, especially in Asia, to control sheath blight disease of rice plants and dumping-off of cucumber and Rhizoctonia solani [1]. In the presence of VAL-A, normal extension of the main fungal hyphae is switched to an abnormal branching at the tips, and further development of those growing fungi is repressed [2]. VAL-A was also found to have insecticidal activity, most likely resulting from its strong inhibition of trehalase, the trehalose-degrading enzyme. Trehalose is a primary carbohydrate in fungi (8%–10% of the dry cell weight of *P. sasakii*) and is recognized as a characteristic blood sugar of insects. The enzyme trehalase appears to play an important metabolic role in these organisms in generating glucose for energy supply or for other physiological purposes. Its inhibition results in disruption of the glucose supply system of the fungus, leading to growth abnormality and death. Although in vitro experiments revealed that the pseudodisaccharide, validoxylamine A, is more active than VAL-A against trehalase [3], results of in vivo experiments suggest the opposite. It is proposed that the presence of the glucose moiety on VAL-A is essential for its efficient entry into fungal mycelia, in which it is hydrolyzed by an α-glucosidase to yield the active pharmacophore, validoxylamine A [4]. VAL-A is also used as the source of valienamine, a pharmacologically important precursor for the production of the antidiabetic drug, Voglibose (Basen), whose mechanism of action is to inhibit α-glucosidase in the intestine, comparable to another antidiabetic agent, acarbose.

The biosynthesis of the C7N-aminocyclitol moiety has been investigated to some extent in Actinoplanes species. SE50/110 (acarbose-producer) [5–7], in *S. hygroscopicus* var. *limonius* (VAL-A-producer) [8–10], and, more recently, in *Microtetraspora spiralis* (pyralomicin producer) [11]. Feeding experiments with a number of isotopically labeled potential intermediates to *S. hygroscopicus* var. *limonius* [8, 10, 12] demonstrated that 2-epi-5-epi-valiolone, 5-epi-valiolone, valienone, and valiendole were incorporated into VAL-A, leading to a proposed biosynthetic pathway to VAL-A, as shown in Figure 1B. Similar feeding experiments were also performed with Actinoplanes species SE50/110 [6, 7] and *Microtetraspora spiralis*, but with surprisingly different incorporation patterns. Only 2-epi-5-epi-valiolone and 5-epi-valiolone were found to be efficiently incorporated into pyralomicin [11], whereas 2-epi-5-epi-valiolone was the only one of these compounds incorporated into acarbose.

The gene cluster for acarbose biosynthesis has been cloned and a number of the proteins involved have been characterized, resulting in a proposed biosynthetic pathway to acarbose [13]. The pathway is initiated by a cyclization of sedoheptulose 7-phosphate, a C7 sugar phosphate involved in the pentose phosphate pathway, catalyzed by the dehydroquinase synthase-like protein, AcbC, to give 2-epi-5-epi-valiolone (Figure 1B) [5]. The latter compound is then phosphorylated to its 7-phosphate derivative catalyzed by the kinase AcbM. Furthermore, the product of acbO, which was found in the same operon as acbC and acbM, was identified as a 2-epi-5-epi-valiolone 7-phosphate 2-epimerase [14]. Efficient incorporation of 2-epi-5-epi-valiolone into both VAL and acarbose strongly suggested that both biosynthetic pathways share the same cyclization step. However, further downstream the pathways to the two compounds may be different, as feeding experiments with the producing strains gave two different patterns. This leaves many open questions surrounding the biosynthesis of acarbose and VAL-A, which cannot be addressed by conventional feeding experiments alone.

*Correspondence: taifo.mahmud@oregonstate.edu (T.M.); zxdeng@sjtu.edu.cn (Z.D.)
*These authors contributed equally to this work.
Using the acbC gene from the acarbose cluster as a heterologous probe, we have successfully cloned and identified the VAL-A biosynthetic genes from a genomic library of *S. hygroscopicus* 5008. Sequencing of a 6 kb BamHI fragment (Figure 2A), in vivo gene inactivation, and in vitro biochemical characterization resulted in the identification of 2-epi-5-epi-valiolone synthase (ValA), a homolog of AcbC, paving the way for further sequencing of the entire cluster [15].

Here we describe the sequencing and functional analysis of 45 kb of DNA, which led to the identification of 8 out of 27 genes as essential for validamycin biosynthesis through reconstituted heterologous gene expression in *S. lividans* 1326. We also demonstrate the glycosyltransferase function of the identified valG gene, first by its in vivo inactivation and functional complementation in strain 5008, second by in vitro enzymatic characterization of the ValG protein expressed in *Escherichia coli*.

**Figure 1. Valienamine-Containing C,N Amino-cyclitols and Proposed Biosynthetic Pathways to VAL-A and Acarbose**

(Top) Chemical structures of VAL-A, acarbose, and pyralomicin 1a. Dashed-boxed regions show the cyclitol moiety shared by all three compounds. (Bottom) Previously proposed biosynthetic pathways to VAL-A [8] and acarbose [13, 14].

**Figure 2. Cloning Strategy for Sequencing and Genetic Organization of val Gene Cluster**

(A) Cloning strategy for sequencing of the 45 kb region. B, BamHI; C, ClaI; E, EcoRI.

(B) Genetic organization of the val gene cluster. Genes proposed to be involved in validamycin biosynthesis are represented as black arrows. The previously sequenced 6.0 kb BamHI fragment was indicated by double-arrowed line.
Table 1. Deduced Functions of the Open Reading Frames in the *val* Gene Cluster

<table>
<thead>
<tr>
<th>Protein</th>
<th>aa</th>
<th>Proposed Function</th>
<th>Homologs, Origin</th>
<th>Identity, Similarity (%)</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1</td>
<td>400</td>
<td>Integrase</td>
<td>Urd-int, <em>Streptomyces fradiae</em></td>
<td>88, 94</td>
<td>AA00216</td>
</tr>
<tr>
<td>ORF2</td>
<td>391</td>
<td>Transposase</td>
<td>SAV1204, <em>Streptomyces avermitilis</em></td>
<td>90, 92</td>
<td>BAC68914</td>
</tr>
<tr>
<td>ORF3</td>
<td>276</td>
<td>Transposase</td>
<td>SCO7080, <em>Streptomyces coelicolor</em></td>
<td>91, 93</td>
<td>CAB98752</td>
</tr>
<tr>
<td>ValO</td>
<td>221</td>
<td>Phosphatase/phosphohexomutase</td>
<td>COGO637, <em>Thermobifida fusca</em></td>
<td>39, 58</td>
<td>ZP_00293893</td>
</tr>
<tr>
<td>ValN</td>
<td>331</td>
<td>Cyclic reductase</td>
<td>Gln0396, <em>Gloeobacter violaceus</em></td>
<td>29, 39</td>
<td>BAC88310</td>
</tr>
<tr>
<td>ValM</td>
<td>424</td>
<td>Aminotransferase</td>
<td>DRA0029, <em>Deinococcus radiodurans</em></td>
<td>39, 54</td>
<td>NP_285533</td>
</tr>
<tr>
<td>ValL</td>
<td>492</td>
<td>Validoxylamine A 7-phosphate synthase</td>
<td>Rxy102001033, <em>Rubrobacter xylanophilus</em></td>
<td>30, 47</td>
<td>ZP_00187712</td>
</tr>
<tr>
<td>ValK</td>
<td>324</td>
<td>Epimerase/dehydratase</td>
<td>APE1179, <em>Aeropyrum pernix</em></td>
<td>31, 51</td>
<td>BAA08165</td>
</tr>
<tr>
<td>ValA</td>
<td>414</td>
<td>2-epi-5-epi-valiolone synthase</td>
<td>AcbC*, <em>Actinoplanes sp.</em></td>
<td>48, 65</td>
<td>CAAT7208</td>
</tr>
<tr>
<td>ValB</td>
<td>373</td>
<td>Valienol-1-phosphate nucleotidylyltransferase</td>
<td>GigC, <em>Bacillus halodurans</em></td>
<td>33, 50</td>
<td>BAA04806</td>
</tr>
<tr>
<td>ValC</td>
<td>351</td>
<td>Cyclic ketolase</td>
<td>GlkC, <em>Bacillus cereus</em></td>
<td>29, 44</td>
<td>ZP_00238533</td>
</tr>
<tr>
<td>ValD</td>
<td>451</td>
<td>Glyoxalase</td>
<td>NocADRAFT_3188, <em>Nocardiodes sp.</em> J5614</td>
<td>40, 49</td>
<td>ZP_00052713</td>
</tr>
<tr>
<td>ValE</td>
<td>331</td>
<td>Oxidoreductase</td>
<td>SAV4805, <em>S. avermitilis</em></td>
<td>45, 56</td>
<td>BAC72517</td>
</tr>
<tr>
<td>ValF</td>
<td>401</td>
<td>Oxidoreductase</td>
<td>SAV1825, <em>S. avermitilis</em></td>
<td>80, 87</td>
<td>BAC65856</td>
</tr>
<tr>
<td>ValG</td>
<td>422</td>
<td>Glycosyltransferase*</td>
<td>KfoC*, <em>Escherichia coli</em></td>
<td>35, 49</td>
<td>BAC00523</td>
</tr>
<tr>
<td>ValH</td>
<td>396</td>
<td>Transport protein</td>
<td>SAV6900, <em>S. avermitilis</em></td>
<td>52, 66</td>
<td>BAC74611</td>
</tr>
<tr>
<td>ValI</td>
<td>611</td>
<td>Glycosyl hydrolase</td>
<td>SAV6503, <em>S. avermitilis</em></td>
<td>69, 81</td>
<td>BAC74214</td>
</tr>
<tr>
<td>ValJ</td>
<td>305</td>
<td>Oxidoreductase</td>
<td>SAV4805, <em>S. avermitilis</em></td>
<td>43, 55</td>
<td>BAC72517</td>
</tr>
<tr>
<td>ORF4</td>
<td>204</td>
<td>Unknown</td>
<td>MII4476, <em>Mesorhizobium loti</em></td>
<td>39, 55</td>
<td>NP_105539</td>
</tr>
<tr>
<td>ORF5</td>
<td>347</td>
<td>Unknown</td>
<td>Smc00549, <em>Sinorhizobium meliloti</em></td>
<td>28, 47</td>
<td>NP_385900</td>
</tr>
<tr>
<td>ValP</td>
<td>293</td>
<td>Regulatory protein</td>
<td>SCO3723, <em>S. coelicolor</em></td>
<td>49, 63</td>
<td>CAD55330</td>
</tr>
<tr>
<td>ValQ</td>
<td>163</td>
<td>Regulatory protein</td>
<td>SCO3723, <em>S. coelicolor</em></td>
<td>45, 59</td>
<td>CAD55330</td>
</tr>
<tr>
<td>ORF6</td>
<td>191</td>
<td>Tellurium resistance protein</td>
<td>SAV896, <em>S. avermitilis</em></td>
<td>92, 97</td>
<td>BAC68806</td>
</tr>
<tr>
<td>ORF7</td>
<td>451</td>
<td>Unknown</td>
<td>SCO3725, <em>S. coelicolor</em></td>
<td>72, 81</td>
<td>CAB76972</td>
</tr>
<tr>
<td>ORF8</td>
<td>290</td>
<td>Unknown</td>
<td>SAV7288, <em>S. avermitilis</em></td>
<td>52, 68</td>
<td>BAC74979</td>
</tr>
<tr>
<td>ORF9</td>
<td>308</td>
<td>Esterase</td>
<td>Bcepa03001004, <em>Burkholderia cepacia</em></td>
<td>47, 59</td>
<td>ZP_00218391</td>
</tr>
<tr>
<td>ORF10</td>
<td>349</td>
<td>Cytochrome P450 hydroxylase</td>
<td>P450-SU2, <em>Streptomyces griseolus</em></td>
<td>51, 61</td>
<td>AAA26025</td>
</tr>
</tbody>
</table>

*With confirmed function.

**Results**

**Overall Architecture of the Sequenced Region**

A total of 45,168 bp of DNA sequence (deposited in GenBank under accession number DQ164098), including the previously reported 5877 bp of DNA (AY753181), was determined using overlapping or linking fragments (Figure 2A). Overall GC content of this region is 68.7%—somewhat lower than the typical GC content as exemplified by *S. coelicolor* A3(2) [16] and *S. avermitilis* [17] (~72%). In silico analysis of the sequence revealed the presence of 27 complete open reading frames (ORFs), whose organization is shown in Figure 2B and proposed functions are listed in Table 1. Structural and regulatory genes likely involved in VAL-A biosynthesis, designated as *valA-Q* and indicated as black arrows in Figure 2B, are distributed in the middle part of the sequenced region, among which subcluster *valA-valJ* and *valK-valN* seem to be transcribed in opposite directions.

The *valA-valJ* Subcluster

*valA*, whose deduced protein sequence shows significant homology to AcbC from the carbace biosynthetic gene cluster, was confirmed to encode ValA [15]. The deduced amino acid sequence of ValB shows 33% identity with GlgC, a glucose-1-phosphate adenyllyltransferase from *Bacillus halodurans* C-125 [18]. ValC shows 30% identity with AcbM, the 2-epi-5-epi-valiolone 7-kinase from the carbace cluster, whose demonstrated function is to phosphorylate 2-epi-5-epi-valiolone to form 2-epi-5-epi-valiolone 7-phosphate [19]. Additionally, ValC shows a weaker homology to glucokinase from a *Bacillus* species (29% identity/44% similarity).

Full-length ValD (451 aa) is homologous to a putative glyoxalase/bleomycin resistance protein/dioxygenase domain from *Nocardiodes sp.* J5614. Two domains of ValD, one at Gly35-Glu194 and the other at Gly220-Ala382, showed 35.5% identity/50% similarity to each other.

Both ValE (331 aa) and ValJ (305 aa) show homology to 2-oxoglutarate and Fe(II)-dependent oxygenases, whose enzymatic functions are known to catalyze a variety of reactions typically involving the oxidation of an organic substrate with an oxygen molecule, such as hydroxylation, desaturation, or oxidative ring closure [20]. Noticeably, ValE and ValJ showed significant homology (79.8% identity at nucleotide [nt] level and 66.3% identity at the protein level), and mainly differed from each other at the C terminus.

ValF is highly homologous with putative oxidoreductases from *S. avermitilis* and *S. coelicolor* (76% identity/82% similarity). In addition, a ValF-related glucose-fructose oxidoreductase from *Zymomonas mobilis* is known to catalyze the reductive conversion of fructose to sorbitol [21], suggesting that ValF might be involved in the keto reduction of intermediate compounds in VAL-A biosynthesis.

ValG exhibits considerable homology to chondroitin polymerase KfoC from *E. coli* K4, which is known to...
catalyze the polymerization of the chondroitin backbone in the assembly of capsule polysaccharide K4 antigen [22]. However, ValG (422 aa) is 264 aa shorter than KfoC (686 aa), but 250 aa residues of its N-terminal domain, which is required for sugar donor binding, are very similar to known glycosyltransferases. In addition, one conserved short motif of DGS and two DXD motifs, which have been proposed to facilitate NDP-sugar binding and Mn$^{2+}$ ion binding [23], respectively, are present in ValG.

The deduced product of valH is homologous to putative transport protein SAV6900 from S. avermitilis and the fucose permease FucP from Mesorhizobium sp. BNC1 (33% identity and 50% similarity), whose functions are known to be cotransport of fucose, glucose, or galactose with H$^+$ [24]. Analysis of ValH by using the TMHMM Server v. 2.0 [25] suggests that 11 transmembrane helices span the length of the protein.

ValM seems to be closely related to the putative glucosamylase SAV6503 from S. avermitilis, which belongs to the glycosyl hydrolase family 15 [26]. Glucoamylases known to catalyze the hydrolysis of α-glucosidic linkages to release β-D-glucose from the nonreducing ends of starch and related polysaccharides [27]. However, no obvious function could be assigned in the pathway for VAL-A biosynthesis, except that D-glucose could be a precursor for VAL-A.

The valK-valN Subcluster

The valK-valN genes are predicted to be transcribed in the opposite direction to the valA-valJ genes. BLAST analysis of ValK showed significant homology with dTDP-4-dehydrorhamnose reductase (RmlD) from Aeropyrum pernix, a member of the reductase/epimerase/dehydrogenase/dehydratase protein superfamily putatively involved in the biosynthesis of L-rhamnose from glucose-1-phosphate [28]. ValK contains a strictly conserved (P-X$_2$-Y-X$_2$-K-X$_n$-E) motif that is characteristic for this superfamily and another conserved (STDNVF) motif that is unique for RmlD-type proteins [29]. Interestingly, the conserved NAD(P) binding Wierenga motif (G-X$_2$-G-X$_2$-G) [30] is missing from the N terminus of ValK. NAD(P)$^+$-independent sugar epimerases from E. coli and other microorganisms have been reported previously [29, 31]. The reaction mechanism involves substrate to metal binding and a cis enediyolate-stabilized intermediate [32, 33]. However, no significant sequence similarity was found between ValK and the NAD(P)$^+$-independent sugar epimerases.

ValL shows significant homology to the putative trehalose 6-phosphate synthase Rxy021033 from Rubrobacter xylanophilus and lower homology to its counterpart OtsA in E. coli. In E. coli, the trehalose 6-phosphate synthase (OtsA) and the trehalose 6-phosphate synthase (OtsB) are responsible for the biosynthesis of trehalose from glucose-6-phosphate and UDP-glucose, with net retention of the anomeric configuration of the donor sugar [28].

ValM is closely related to pyridoxal 5’-phosphate-dependent (PLP) aminotransferases, such as the putative 4-aminoacrylate aminotransferase from Deinococcus radiodurans and the acetylornithine aminotransferase from Methanosarcina mazei. Consistent with other aminotransferases, the conserved Lys268 of ValM is believed to be the PLP binding residue for the Schiff base linkage. In addition, other conserved residues important for PLP binding (Asp239, Gin242, and Thr322) and catalysis (Glu206 and Arg349) are also present [34].

ValN shows significant homology to the zinc-dependent threonine dehydrogenase, Krao06004313, from Kineococcus radiotolerans and AcbL from the acarbose pathway of Actinoplanes sp. SE50/110 (29% identity/39% similarity). AcbL, a putative cytidyl dehydrogenase was proposed to be involved in the reduction of the C-1 keto group of valienone 7-phosphate to give 1-epi-valienone 7-phosphate [19].

Putative Regulatory Genes

ValP and valQ are predicted to encode a two-component regulatory system consisting of a separate histidine kinase (sensor) and a Sigma B PP2C-like phosphatase (response regulator) [35], as ValP showed the highest homology with the response regulator phosphatase domain and ValQ with the sensor kinase domain of SCO3723 from S. coelicolor A3(2). The existence of valP and valQ as two independent genes separated by only 28 bp seems unique in the val gene cluster, since most of the counterpart proteins, exemplified by SCO3723, are known to exist as single peptide with two domains conforming to ValP and ValQ, respectively. It is not known yet whether the independent organization of ValP and ValQ could have anything to do with the induction of Sigma B activity and subsequent validamycin production in response to environmental stress, such as increased temperature and/or high salt concentration. However, we have consistently observed an apparent increase in production of VAL-A and enhanced transcriptions of both valP and valQ in parallel when the producer was grown at 37ºC rather than at 30ºC (Lei Li and Linquan Bai, unpublished data).

Boundaries of the Cluster and Other Most Likely Unrelated Genes

Beyond valN, three transposable-like or integrase-like ORFs (ORF1–3) and valO, which exhibits homology to a predicted phosphatase/phosphohexomutase from Thermobifida fusca, were detected. To determine the left boundary of the biosynthetic gene cluster, valO, which is located 764 bp downstream of valN and most likely monocistronically transcribed, was inactivated.

Through double crossover recombination mediated by the 2.5 kb left flanking sequence and 4.0 kb right flanking sequence, an internal 588 bp fragment was replaced by $aac(3)\gamma$ via RedDirect methodology [36]. The mutant ZYR-5, confirmed by PCR amplification, was found to produce a comparable amount of VAL-A as the wild-type, suggesting that valO is not involved in the biosynthesis of VAL-A. Thus, the left-border of the val gene cluster was defined.

Of the seven ORFs flanking the putative two-component regulatory genes, valP and valQ, ORF 4, 5, 7, and 8 seem to encode proteins with unknown biological functions. Moreover, none of the detected homologies of ORF 6, 9, and 10, with a tellurium resistance protein, a group of esterases, and a cytochrome P450, seem to be relevant to validamycin biosynthesis. It is thus likely that all of the genes required for VAL-A biosynthesis...
are located between valN and valQ in a ~ 30 kb region (Figure 2B).

**Heterologous Production of Validoxylamine A and VAL-A in** S. *lividans* 1326

valABC seems to transcribe as one operon and valKLNM as another operon transcribed in the opposite direction (Figure 2B). The two operons were reassembled to have seven genes in the same orientation under the control of a constitutive PermE connectionString} promoter (Figure 3A), and a construct (pJTU757) was introduced into a heterologous host, S. *lividans* 1326, to see if the biosynthetic intermediate, validoxylamine A, can be produced. Since the 251 bp fragment to the N terminus of valK may still contain its own promoter, the recombinant, XH-6, was thus initially fermented at 37°C instead of at 30°C, as commonly set for S. *lividans*.

Even though bioassay with chloroform-extracted fermentation broth of XH-6 showed almost no inhibitory activity against *P. sasakii*, and HPLC analysis gave no obvious peak corresponding to validoxylamine A because of the lower yield, mass spectrometry analysis through direct injection of the sample did reveal the presence of validoxylamine A (Figure 3A). The typical fragmentation of validoxylamine A (m/z 336.1) under positive-ion mode gives validamine (m/z 178.1). The isolated compound with m/z of 336.3 (MS$_{1}$) from XH-6 was indeed fragmented to give a peak with m/z of 178.1 (Figure 3A, MS$_{2}$). These data demonstrate that only seven genes, valABCKLMN, from the two original operons are required for the synthesis of validoxylamine A. Fermentation of XH-6 at 30°C, followed by MS analysis, also revealed the production of validoxylamine A at low yield (data not shown).

Since valG is the only gene found in the sequenced region encoding a glycosyltransferase, it was tempting to coexpress this gene with valABCKLMN for the possible production of VAL-A in S. *lividans*. The new recombinant XH-9 harboring valABCKLMNG showed inhibitory activity in bioassay after fermentation at 37°C for 7 days, or at 30°C for 11 days (data not shown), suggesting the accumulation of VAL-A. The typical fragmentation pattern of VAL-A (m/z 498.2) under positive-ion mode is the loss of glucose, to give validoxylamine A (m/z 336.1), which is further fragmented to validamine (m/z 178.1). Detailed analysis by mass spectrometry of a 7-day fermented broth of XH-9 at 37°C clearly indicated the presence of VAL-A with m/z of 498.2, which was sequentially fragmented into validoxylamine A (m/z 336.1) and validamine (m/z 178.1) in the MS/MS experiment (Figure 3B).
The 11 day fermentation broth of XH-9 at 30°C was also analyzed by MS/MS, and an obviously similar fragmentation pattern for VAL-A could be observed (data not shown).

The successful heterologous production of validoxylamine A and VAL-A in *S. lividans* 1326 not only confirmed the necessity for VAL-A biosynthesis of the 8 structural genes from the 27 ORFs found in the sequenced region, but also revealed the identity of *valG* as a glycosyltransferase that transfers activated glucose to validoxylamine A as the final step in VAL-A biosynthesis.

**Targeted Replacement of valG Abolishes VAL-A Biosynthesis**

LL-1 (Figure 4B) was one of the confirmed mutants, with an internal 806 bp region of valG (from nt number 112 to 918) replaced by an *aac(3)IV* from the strain 5008 (Figure 4A). Bioassay indicated that the LL-1 fermentation broth had retained inhibitory activity, albeit at a reduced capacity (Figure 4C), but the presence of VAL-A (retention time of 9.7 min) could not be detected by HPLC analysis. Despite the fact that, under normal conditions, the wild-type strains also produced validoxylamine A (retention time of 6.5 min), a significantly increased accumulation of validoxylamine A was observed in LL-1 (Figure 4D). The abolished production of VAL-A and the increased accumulation of validoxylamine A in LL-1 strongly suggest that ValG catalyzes the glycosylation of validoxylamine A to yield VAL-A. The observed weak inhibitory activity agrees with the previous finding that validoxylamine A has much lower in vivo inhibitory activity than VAL-A [4].

**Complementation by Cloned valG Restores VAL-A Production**

When pJTU612, a pJ101-derived plasmid with an intact valG, was introduced into mutant LL-1, the culture broth of the thiostrepton-resistant exconjugant (LL-101) was found to have regained the inhibitory activity (Figure 4C) in bioassay. HPLC analysis unambiguously demonstrated the presence of VAL-A with the retention time of 9.7 min. However, the presence of valG under the strong constitutively expressed *PermE* promoter in LL-101 was not sufficient for the conversion of all
Validamycin Biosynthetic Gene Cluster

In Vitro Glycosylation of Validoxylamine A to VAL-A Using Recombinant ValG

Further proof to support the glycosyltransferase function of valG was obtained from its heterologous expression in E. coli, which gave rise to a 49 kDa soluble polyhistidine-tagged protein. Affinity purification on an Ni-NTA spin column (QIAGEN) or a BD TALON column gave a protein that was >75% pure as judged by SDS-polyacrylamide gel electrophoresis (Figure 5A). To test the catalytic activity of ValG, the enzymatic reaction was carried out using validoxylamine A and UDP-, GDP-, ADP-, or TDP-glucose as substrates. The reactions gave rise to a product that has the same Rf value as authentic VAL-A on TLC (Figure 5B). On the other hand, incubation with cell-free extracts of E. coli harboring empty pRSET-B vector gave no product (data not shown). The product was confirmed to be VAL-A by ESI-MS and NMR. UDP-glucose was the most efficient glycosyl donor for the ValG reaction, as judged by TLC (Figure 5B), whereas GDP-glucose was less efficient and ADP-glucose was much less efficient. To our surprise, TDP-glucose was the least efficient donor for the reaction among the activated sugars tested.

Discussion

Earlier feeding experiments with isotopically labeled potential intermediates led to a proposal that the biosynthesis of VAL-A is initiated by the cyclization of D-sedoheptulose 7-phosphate to form 2-epi-5-epi-valiolone, which subsequently epimerizes at C-2 to give 5-epi-valiolone, followed by dehydration between C-5 and C-6 to generate valienone. The conversion of valienone to validone takes place by anti-addition of hydrogen followed by transamination, with the α-nitrogen of glutamate as the probable nitrogen source, to form validamine. Analysis of an ~45 kb region of the S. hygroscopicus 5008 chromosome revealed 16 potential structural genes, which include genes deduced or proven to encode ValA, epimerase/dehydratase (ValK), reductase (ValN), transaminase (ValM), and glycosyltransferase (ValG) (Table 1). These enzymes fit nicely with the proposed scheme of VAL-A biosynthesis resulting from the feeding experiments. Despite some discrepancies between the biosynthesis of acarbose and VAL-A, cloning of the acarbose biosynthetic gene cluster and subsequent biochemical characterization of several enzymes involved in the biosynthesis has provided additional information for the elucidation of VAL-A biosynthesis [13]. In the case of acarbose, 2-epi-5-epi-valiolone, the cyclization product of AcbC is phosphorylated by the kinase AcbM to give 2-epi-5-epi-valirole 7-phosphate. Further conversion to 5-epi-valirole 7-phosphate is catalyzed by the epimerase AcbO. The presence of ValC, a homolog of AcbM, in the val gene cluster suggests the involvement of phosphorylated intermediates in VAL-A biosynthesis as well. However, since 2-epi-5-epi-valirole, 5-epi-valirole, valienone, and validone were efficiently incorporated into VAL-A, the kinase (ValC) may utilize valienone and/or validone as substrate. Alternatively, the kinase might have broad substrate specificity, activating 2-epi-5-epi-valirole, 5-epi-valirole, valienone, and validone, and making it possible for them to be incorporated into the biosynthetic pathway.

In analogy to the proposed function of the dehydrogenase AcbL in acarbose biosynthesis, ValN is predicted to reduce the C-1 keto group of valirole-7-phosphate to give either valirole 7-phosphate or its epimer, 1-epi-valirole 7-phosphate (Figure 6). If valirole 7-phosphate is formed, a downstream coupling reaction, most likely catalyzed by the trehalose 6-phosphate synthase homolog (ValL), would occur with net retention of the C-1 configuration of the unsaturated cyclitol. However, if 1-epi-valirole 7-phosphate is formed, the coupling reaction must take place with inversion of the C-1 configuration. In acarbose biosynthesis, it was proposed that reduction of valirole 7-phosphate yields 1-epi-valirole 7-phosphate, and further phosphorylation of the latter compound by an unidentified kinase gives 1-epi-valirole...
1,7-diphosphate. However, while this hypothesis is attractive, there is no evidence for the presence of a second kinase in the validamycin cluster. The presence of valO, which encodes a protein homologous to phosphatase/phosphohexomutase, in the validamycin cluster conversely suggests the involvement of a phosphomutation step from the C-7 to the C-1 hydroxyl group, as commonly observed in sugar biosyntheses. However, inactivation of valO did not abolish the production of VAL-A, which leaves this part of the pathway unclear.

Nevertheless, the availability of the genetic information reported here, as well as results from our previous feeding experiments and biochemical analysis of several related enzymes, allows us to propose a more comprehensive pathway to VAL-A (Figure 6). In vivo inactivation and in vitro characterization of ValA confirmed the initiation step from D-sedoheptulose 7-phosphate to the cyclic product, 2-epi-5-epi-valionone [15]. Subsequently, epimerization at C-2 and dehydration at C-5/C-6 of 2-epi-5-epi-valionone by the proposed bifunctional ValK would give valienone. ValK has significant homology with the sugar epimerase/dehydratase, and it appears to be the only plausible candidate enzyme for the epimerization and dehydration reactions among the eight proteins known to be involved in VAL-A biosynthesis. In addition, ValK may not require NAD(P)+ for catalytic activity, which is similar to the L-ribulose-5-phosphate 4-epimerase and the dTDP-4-dehydrorhamnose 3,5-epimerase [29, 31]. However, no significant homology was found between the primary structures of ValK and the NAD+-independent sugar epimerases. This raises questions regarding the function and reaction mechanism of ValK, which prompts further investigations.

Prior to any further processing, valienone may be phosphorylated, presumably by the kinase, ValC, to give valienone 7-phosphate. Valienone 7-phosphate may serve as the branching point to be reduced to validone 7-phosphate on the one hand, and to valienol 7-phosphate or 1-epi-valienol 7-phosphate on the other, by the reductase ValN and/or other reductases/dehydrogenases. There are three oxidoreductase genes present in the cluster: valE, valF, and valJ. However, the results of the heterologous expression experiments in S. lividans suggest that these genes are not necessary for the biosynthesis of VAL-A. Even so, it may be that they are involved in the biosynthesis of validamycins B and/or G, the hydroxylated analogs of VAL-A. Transamination of the keto group of validone 7-phosphate catalyzed by the aminotransferase ValM would give validamine 7-phosphate. On the other hand, valienol 7-phosphate or its 1-epimer is either phosphorylated at C-1 by an unidentified kinase to generate the 1,7-diphosphate derivative or transformed into valienol 1-phosphate through catalysis by a phosphoglucomutase. The product is subsequently converted to an NDP-valienol derivative by the nucleotidyl transferase ValB. Condensation of validamine 7-phosphate with the NDP-valienol derivative would give phosphorylated validoxylamine A, whose conversion to validoxylamine A requires a dephosphorylation activity. The corresponding phosphatase has not been located in the sequenced region. However, an internal trehalose 7-phosphate phosphatase or a nonspecific sugar phosphatase may be able to carry out this hydrolytic reaction. Finally, validoxylamine A is converted to VAL-A biosynthesis through catalysis by the glycosyltransferase ValG by using UDP-glucose as the sugar donor.

Significance

VAL-A is a strong crop protectant produced by Streptomyces hygroscopicus and is widely used for the treatment of the sheath blight disease of rice plants and the vegetable dumping-off caused by the fungus.
Pellicularia sasakii. In addition, its degradation product, valienamine, is a pharmaceutically important precursor for the production of the anti-diabetic agent, voglibose. Despite its significant and wide applications in the biomedical and agricultural sectors, little is known about its biosynthesis. Sequencing and bioinformatics analysis of the validamycin biosynthetic gene cluster, as well as in vivo and in vitro characterization of a number of enzymes involved in VAL-A biosynthesis, provide important insights into the biosynthesis of this antibiotic. The data presented will also aid in the elucidation of the biosynthesis of other related aminocyclitol-containing natural products. In combination with knowledge about the acarboc bio- synthetic gene cluster, it may now be possible to design unique primers for the screening and cloning of other aminocyclitol biosynthetic gene clusters from various sources. Exemplified by the significant accumulation of valdoxyamine A in the glycosytransferase mutants, in-depth knowledge about VAL-A biosynthesis could potentially be used to establish a platform for the production of drug precursors (e.g., valienamine, validamine, valiolamine, and hydroxyvalidamine). Functional analysis of validamycin biosynthesis also sets the stage for creating novel analogs of this important crop protectant at hand, one can contemplate the possibility of expressing them in plant systems to endow crop plants with their built-in protection against pathogens such as P. sasakii.

Experimental Procedures

Bacterial Strains, Plasmids, Culture Techniques, and Media

*S. hygroscopicus* 5008 [37] and its derivatives were grown on solid SMF medium or in modified liquid TSB medium [38] (supplemented with 1% yeast extract and 10.3% sucrose). Fermentation in a genomic library and for gene inactivation [41]. pJTU472 contains *S. hygroscopicus* DNA sequencing was performed using a set of five fragments digested with the same restriction enzymes. An exconjugant having the correct plasmid, named pJTU612, was obtained by transformation and cloning of newly purified pJTU612 and the original one by restriction enzymes as described in [47].

The validation of aminocyclitol biosynthetic gene clusters from various sources. Exemplified by the significant accumulation of valdoxyamine A in the glycosytransferase mutants, in-depth knowledge about VAL-A biosynthesis could potentially be used to establish a platform for the production of drug precursors (e.g., valienamine, validamine, valiolamine, and hydroxyvalidamine). Functional analysis of validamycin biosynthesis also sets the stage for creating novel analogs of this important crop protectant at hand, one can contemplate the possibility of expressing them in plant systems to endow crop plants with their built-in protection against pathogens such as *P. sasakii*.

PCR-amplified valG was cleaved as a 1.2 kb Ndel-EcoRI fragment and cloned into pJTU695 digested with the same restriction enzymes. The new plasmid, pJTU12, was introduced into *P. sasakii* into LL-1 through conjugation as previously described [15], and exconjugants were selected with 25 μg/ml thioestrepton. Further confirmation was carried out through plasmid isolation from 2-day-old mycelia, transformation into *E. coli*, and comparison between the newly purified pJTU612 and the original one by restriction enzymes digestion. An exconjugant having the correct plasmid, named LL-101, was investigated through fermentation, bioassay, and HPLC analysis, as previously described [15].

Targeted Disruption of valO

Shuttle vector pJTU695 for our complementation study was constructed as follows. The 1628 bp region between *Shp1* and *Nru1* of the *pET15b* (Novagen) was replaced first by an 868 bp Shp1-Ball fragment containing oriT from pSET152 [45]. A 163 bp Ndel-BglII region of the resulting plasmid was further replaced by a 4365 bp fragment containing the replication origin of pJR10 and ts, creating pJTU676. A region of pJTU676, with the *Perme* promoter, ribosomal binding site (RBS), and Ndel site, was replaced with a corresponding fragment from plB139 [46].

DNA Sequencing and Analysis

DNA sequencing was performed using a set of five fragments directly isolated from cosmids 4G8 and 17F2 [15], or cloned onto pBluescript SK(+) (Stratagene) or pSL301 (Invitrogen), which were originally connected or overlapping to cover an ~45 kb region with the validamycin biosynthetic gene cluster (Figure 2A). pHZ2232 with the 1.0 kb BamHI fragment was directly sequenced with T3 and T7 primers. Sequencing analysis of the 21 kb Clal fragment, a 19.6 kb EcoRI fragment (pHZ2239), a 6.0 kb BamHI fragment (pHZ2229), and 7.0 and 3.8 kb BamHI fragments (pJTU646) were as described in [47].

Inactivation of valG

A 2.4 kb EcoRI-HindIII fragment flanking the left of the nt 112 of valG and a 1.0 kb HindIII-EcoRI fragment flanking the right of the nt 918 of valG were coligated into EcoRI-digested pJ2925 to generate pJTU600. A HindIII fragment carrying the 1.4 kb aacQIV cassette from pJTU472 was inserted into the HindIII site between the 2.4 and 1.0 kb fragments to generate pJTU607, from which a 4.8 kb BglII fragment was excised for insertion into the BamHI-digested pHZ1358 to construct a vector (pJTU609). pJTU609 was then used for mediating gene replacement (Figure 4) to obtain a valG mutant (LL-1). The mutation was confirmed by PCR amplification using forward primer valG2-F (5′-AGAGGCGATCTGGTGGTGA-3′) and reverse primer valG2-R (5′-GGTAGATGACTCGCTCCCTC-3′).

Complementation of Mutant LL-1 with Cloned valG

PCR-amplified valG was cleaved as a 1.2 kb Ndel-EcoRI fragment and cloned into pJTU695 digested with the same restriction enzymes. The new plasmid, pJTU612, was introduced into *P. sasakii* via combinatorial biosynthesis and yield improvement through metabolic engineering. Finally, with all the genes necessary for the synthesis of this important crop protectant at hand, one can contemplate the possibility of expressing them in plant systems to endow crop plants with their built-in protection against pathogens such as *P. sasakii*. Sequencing analysis of the valG biosynthetic gene cluster, as well as in vivo and in vitro characterization of a number of enzymes involved in VAL-A biosynthesis, provide important insights into the biosynthesis of this antibiotic. The data presented will also aid in the elucidation of the biosynthesis of other related aminocyclitol-containing natural products. In combination with knowledge about the acarboc bio- synthetic gene cluster, it may now be possible to design unique primers for the screening and cloning of other aminocyclitol biosynthetic gene clusters from various sources. Exemplified by the significant accumulation of valdoxyamine A in the glycosytransferase mutants, in-depth knowledge about VAL-A biosynthesis could potentially be used to establish a platform for the production of drug precursors (e.g., valienamine, validamine, valiolamine, and hydroxyvalidamine). Functional analysis of validamycin biosynthesis also sets the stage for creating novel analogs of this important crop protectant at hand, one can contemplate the possibility of expressing them in plant systems to endow crop plants with their built-in protection against pathogens such as *P. sasakii*. Sequencing analysis of the valG biosynthetic gene cluster, as well as in vivo and in vitro characterization of a number of enzymes involved in VAL-A biosynthesis, provide important insights into the biosynthesis of this antibiotic. The data presented will also aid in the elucidation of the biosynthesis of other related aminocyclitol-containing natural products. In combination with knowledge about the acarboc bio- synthetic gene cluster, it may now be possible to design unique primers for the screening and cloning of other aminocyclitol biosynthetic gene clusters from various sources. Exemplified by the significant accumulation of valdoxyamine A in the glycosytransferase mutants, in-depth knowledge about VAL-A biosynthesis could potentially be used to establish a platform for the production of drug precursors (e.g., valienamine, validamine, valiolamine, and hydroxyvalidamine). Functional analysis of validamycin biosynthesis also sets the stage for creating novel analogs of this important crop protectant at hand, one can contemplate the possibility of expressing them in plant systems to endow crop plants with their built-in protection against pathogens such as *P. sasakii*. Sequencing analysis of the valG biosynthetic gene cluster, as well as in vivo and in vitro characterization of a number of enzymes involved in VAL-A biosynthesis, provide important insights into the biosynthesis of this antibiotic. The data presented will also aid in the elucidation of the biosynthesis of other related aminocyclitol-containing natural products. In combination with knowledge about the acarboc bio- synthetic gene cluster, it may now be possible to design unique primers for the screening and cloning of other aminocyclitol biosynthetic gene clusters from various sources. Exemplified by the significant accumulation of valdoxyamine A in the glycosytransferase mutants, in-depth knowledge about VAL-A biosynthesis could potentially be used to establish a platform for the production of drug precursors (e.g., valienamine, validamine, valiolamine, and hydroxyvalidamine). Functional analysis of validamycin biosynthesis also sets the stage for creating novel analogs of this important crop protectant at hand, one can contemplate the possibility of expressing them in plant systems to endow crop plants with their built-in protection against pathogens such as *P. sasakii*. Sequencing analysis of the valG biosynthetic gene cluster, as well as in vivo and in vitro characterization of a number of enzymes involved in VAL-A biosynthesis, provide important insights into the biosynthesis of this antibiotic. The data presented will also aid in the elucidation of the biosynthesis of other related aminocyclitol-containing natural products. In combination with knowledge about the acarboc bio- synthetic gene cluster, it may now be possible to design unique primers for the screening and cloning of other aminocyclitol biosynthetic gene clusters from various sources. Exemplified by the significant accumulation of valdoxyamine A in the glycosytransferase mutants, in-depth knowledge about VAL-A biosynthesis could potentially be used to establish a platform for the production of drug precursors (e.g., valienamine, validamine, valiolamine, and hydroxyvalidamine). Functional analysis of validamycin biosynthesis also sets the stage for creating novel analogs of this important crop protectant at hand, one can contemplate the possibility of expressing them in plant systems to endow crop plants with their built-in protection against pathogens such as *P. sasakii*.
Preparation of Cell-free Extracts and Purification of His-tagged ValG

Cells were thawed and resuspended in disruption buffer (50 mM NaH2PO4, 300 mM NaCl, 10 mM imidazole [pH 8.0]). The suspension was sonicated three times for 25 s each, and cell debris was removed by centrifugation at 10,000 rpm for 10 min. The protein solution was applied to a BD TALON spin column (BD Biosciences) and centrifuged at 500 rpm for 2 min. The column was washed with washing buffer [twice with 50 mM NaH2PO4, 300 mM NaCl, 10 mM imidazole [pH 8.0], and once with 50 mM NaH2PO4, 300 mM NaCl, 100 mM imidazole [pH 8.0]]. The His-tagged protein was eluted with elution buffer (50 mM NaH2PO4, 300 mM NaCl, 500 mM imidazole [pH 8.0]) and dialyzed for 24 hr against 1 liter of dialysis buffer (25 mM Tris-HCl [pH 7.6], 10 mM MgCl2, 20 mM NH4Cl, and 0.5 mM DTT). Protein concentration was measured by the Bradford protein assay with bovine serum albumin as standard.

Enzyme Assay

The enzyme assay was carried out at 30°C for 3–2 hr in a 100 μl volume of 25 mM Tris-HCl (pH 7.6), 10 mM MgCl2, 20 mM NH4Cl, 15 mM NADP-glucose, 10 mM validoxylamine A, and 50 μl of protein solution (2.4 mg/ml protein). The reaction progress was monitored by TLC analysis. For a scale-up experiment, 3.35 mg of validoxylamine A was used. The reaction mixture was applied to a Microcon YM-10 (Amicon) spin column. The flow through was subjected to a cation-exchange (Dowex 50Wx8-200 [H+]) chromatography column. The column was washed with 20 ml of water, and the reaction product was eluted with 0.5 M aqueous NH4OH. Fractions containing the desired product were pooled and lyophilized to give 4.1 mg of VAL-A as a white powder.

Heterologous Production of Validoxylamine A and VAL-A

The valABC operon was first cloned downstream of the PermE promoter by inserting a 623 bp Ndel-FspI DNA fragment of valA and a 3280 bp FspI-BamHI fragment from pHZ2229 into pSL301 digested with NdeI and BamHI. Then, a 3280 bp NdeI-EcoRI fragment containing the complete valABC was cleaved from the resulting plasmid, pJTU755, and ligated into pBl139 digested with the corresponding enzymes, generating pJTU756. Subsequently, a 5811 bp EcoRI fragment, containing valKLMN from cosmid 17F, was inserted into EcoRI-digested pJTU756. Clones with valABC/KLMN transcribed in the same orientation were selected and named pJTU757.

Plasmid pJTU757 was integrated into the chromosome of XH-9, which was selected with thiostrepton and apramycin. Fermentation of XH-9 with plasmid pJTU757 resulted in another recombinant strain named XH-9, which was selected with thiostrepton and apramycin. Fermentation and bioassay analysis of XH-6 and XH-9 were done as previously described [15]. Tandem MS analysis was performed with fermentation broth extracted once with chloroform through directed injection on an Agilent 1100 series LC/MSD Trap system. The iontrap mass spectrometer was operated with the electrospray ionization source in the positive-ion mode. Drying gas flow was 10 liters/min, and nebulizer pressure was 50 psi. Drying gas temperature was 325°C. The fragmentation amplitude was varied between 1.0 and 1.8 V.

Acknowledgments

This work received support from the Ministry of Science and Technology (2003CB114205), the National Science Foundation of China, and the Ph.D. Training Fund of the Ministry of Education, and the Shanghai Municipal Council of Science and Technology. Work at Oregon State University was supported by NIH grant R01GM07258A. K.M. received postdoctoral training funds from Shionogi & Co., LTD, Osaka, Japan. The authors thank Dr. Patricia M. Flatt for helpful comments.

Received: November 4, 2005
Revised: January 25, 2006
Accepted: February 1, 2006
Published: April 21, 2006

References


Accession Numbers

The complete DNA and deduced protein sequences of the val gene cluster reported in this paper have been deposited in GenBank under the accession number DQ164098. Previously reported 5677 bp of DNA within the 45 kb sequenced region was deposited in GenBank under accession number AY753181.