

第 4 章

Cumene 資化菌 *Pseudomonas fluorescens* IP01 株の cumene 代謝酵素遺伝子の単離と機能解析

4-1. 緒言

これまでにダイオキシン類の1つである Co-PCB の微生物分解に関しては、PCB の基本骨格をなす biphenyl 資化菌の biphenyl 分解系が分子レベルで詳細に研究されてきた。序論でも述べたように、当研究室においては biphenyl 資化菌から精製されたメタ開裂酵素と加水分解酵素が、単環芳香族化合物である cumene の中間代謝物 3-isopropylcatechol 及びメタ開裂物質 HOMODA に対しても分解活性を持っていることを明らかにしており[Ishigooka et al., 1986; Omori et al., 1986]、biphenyl 分解系酵素遺伝子と cumene 分解系酵素遺伝子の分子進化上の関連性や、cumene 分解系遺伝子も PCB に対して活性を持っているのかなどについて興味を持たれていた。

そこで本章では、cumene から HOMODA までの分解に関与する酵素遺伝子群 (*cumA1A2ORF3A3A4BC*) を cumene 資化菌 *P. fluorescens* IP01 株から単離し、これが Co-PCB である 4,4'-dichlorobiphenyl を分解可能であることを示すとともに、*bph* 遺伝子群と類似な構造をとることを明らかにした。

4-2. 材料及び方法

4-2-1. 菌株、プラスミド、試薬、培地、及び培養条件

本章で使用した菌株及びプラスミドを Table 4-1 に示した。*E. coli* JM109 株 [Yanish-Perron et al., 1985] は遺伝子ライブラリー及びサブクローニングを行ったプラスミドの宿主として用いた。遺伝子ライブラリーの作成には、ベクターとして pUC118 [Vieira and Messing, 1987] を使用した。通常の *E. coli* 用の培地として 2×YT 培地 (Table 2-2) あるいは LB 培地 (Table 2-3) [Sambrook et al., 1989] を用いた。IP01 株用の培地としては CFMM 培地を用いた (Table 2-4)。*E. coli* の選択培地には Ap を終濃度 50 µg/ml となるよう添加した。Cumene は濾過滅菌したものを培地および休止菌体反応液に終濃度 0.1% (vol/vol) となるよう添加した。試薬は、*cis*-benzene dihydrodiol については Aldrich から購入し、それ以外については市販のもので最も純度の高いものを購入し使用した。本研究で使用した試薬の調整法は細胞工学実験プロトコル [東京大学医科学研究所制癌研究部編, 1993]、クローニングとシーケンス [杉浦昌弘 編, 1989]、もしくは Sambrook らの成書 [Sambrook et al., 1989] に従った。培養は IP01 株については 30°C、*E. coli* については特に明記しないかぎり 37°Cで行った。プレート作製時には細菌培地の寒天を終濃度 1.6% (wt/vol)になるよう添加した。

4-2-2. Cumene 資化菌のスクリーニング法

スクリーニング法は、cumene を唯一の炭素源とした土壌からの集積培養を行った。

【操作】

- ① 0.1% (vol/vol) の cumene を含む 5 ml の CFMM 培地 (Table 2-4) に、一般土壌サンプルを添加し、30°C で 1 週間振とう培養 (300 strokes/min) を行う。
- ② 菌の生育が認められた培養液を 5 ml の同培地に 50 µl 植え継ぎ、再び 30°C で 1 週間振とう培養を行う。
- ③ 菌の生育が認められた培養液を適宜希釈した後、CFMM プレートに塗布する。シャーレ内に cumene を加えた容器を置き、周囲をパラフィルム等で密封し、蒸気を炭素源として供給する。
- ④ 30°C で一晩培養し、生育したコロニーを①と同じ条件で液体培養し、cumene 資化能を確認する。
- ⑤ 培養液をブイオンプレートに塗布して単菌であることを確認した後、再度その単一コロニーを①と同じ条件で液体培養し、生育が認められた株を cumene 資化菌とした。

4-2-3. IP01 株の同定

本菌株の同定は Bergey's Manual [Krieg and Holt 1984] に従って行った。

4-2-4. IP01 株の生育基質特異性の検討

生育基質特異性の検討に用いた芳香族化合物のうち、常温で液体の化合物については、濾過滅菌したものを CFMM 培地に終濃度 0.1% (vol/vol) となるよう添加した。常温で固体の化合物については、DMF もしくは DMSO に溶解し (0.1g/ml)、濾過滅菌したものを CFMM 培地に終濃度 0.1% (vol/vol) となる

Table 4-1. Bacterial strains and plasmids

| Strains or plasmids | Relevant characteristics | Source or reference |
|----------------------------|--|----------------------------|
| Bacterial strains | | |
| <i>P. fluorescens</i> IP01 | Cum ⁺ <i>a</i> | This study |
| <i>E. coli</i> JM109 | <i>recA1, endA1, gryA96, thi, hsdR17, supE44, relA1, Δ(lac-proAB)/F'[traD36, proAB⁺, lacI^q, lacZ ΔM15]</i> | Yanish-Perron et al., 1985 |
| Plasmids | | |
| pUC118 | Ap ^r , <i>lacZ</i> , pMB9 replicon, M13IG | Yanish-Perron et al., 1985 |
| pIP101 | Ap ^r , pUC118 with 8.2-kb <i>EcoRI</i> insert of strain IP01 DNA | This study |
| pIP201 | Ap ^r , pUC118 with 5.2-kb <i>HindIII</i> insert of strain IP01 DNA | This study |
| pIP103 | Ap ^r , pUC118 with 10.8-kb <i>EcoRI</i> insert of strain IP01 DNA | This study |
| pIP104 | Ap ^r , pUC118 with 10.3-kb <i>EcoRI-SacI</i> fragment from pIP103 insert | This study |
| pIP105 | Ap ^r , pUC118 with 7.5-kb <i>EcoRI-BamHI</i> fragment from pIP103 insert | This study |
| pIP106 | Ap ^r , pUC118 with 9.0-kb <i>StuI</i> fragment from pIP103 insert | This study |
| pIP107 | Ap ^r , pUC118 with 5.5-kb <i>HindIII</i> fragment from pIP103 insert | This study |
| pIP107D | Ap ^r , pUC118 with 3'-end deleted fragment of pIP107 insert | This study |

^aCum⁺ indicates the ability to grow on cumene as a sole source of carbon and energy.

よう添加し、30°Cで2日間振とう培養（300 strokes/ min）を行った。

4- 2- 5 . IP01 株の totalDNA の調製

Cumene を含む CFMM 培地中で生育させた IP01 株の菌体を用いて 3- 2- 2 に示した方法に従って行った。

4- 2- 6 . アガロースゲルからの DNA の回収

本章でアガロースゲルからの DNA の回収は 2- 2- 5 に示した方法に従って行った。

4- 2- 7 . 初発酸化酵素遺伝子のショットガンクローニング

芳香族化合物初発酸化酵素の中には、本来の基質の他にインドールの酸化能も合わせ持つものが存在し、インドールが酸化されると自動的な酸化により濃紺色のインジゴが生成することが知られている [Ensley et al., 1983]。一方、*E. coli* を LB 培地や 2 x YT 培地といった栄養培地で培養すると、*E. coli* のトリプトファンナーゼにより培地中のトリプトファンからインドールが生産される。よって、インドール酸化が可能な酵素遺伝子を導入された *E. coli* は、LB 培地または 2 x YT 培地プレート上で培養すると、インジゴ生成によりコロニーが濃紺色を呈する (Fig. 4- 1)。本章では、この方法を cumene 酸化酵素遺伝子のクローニングに利用した。

【操作】

- ① IP01 株の total DNA を *EcoRI* もしくは *HindIII* を個別に用いて切断する。
- ② ①の DNA 溶液を、2- 2- 7 に示した方法で脱リン酸化した pUC118 の各サイトに ligation し、2- 2- 9 に示した方法に従い *E. coli* を形質転換する。
- ③ Ap 50 µg/ ml、IPTG 0.1 mM、X- gal 20 µg/ ml を含むプレートに培養液を塗布した後、37°Cで一晩培養し、インジゴ色を呈するクローンを取得した。

4- 2- 8 . 休止菌体反応

本章においては 50 mM phosphate buffer (pH 7.0) を菌体の洗浄と反応 buffer として使用した。IP01 株休止菌体及び CumABC を発現する *E. coli* による酸化反応に用いた基質は、常温で液体の化合物については、濾過滅菌したものを休止菌体反応液に終濃度 0.1% (vol/ vol) となるよう添加した。常温で固体の化合物については、DMF もしくは DMSO に溶解し (0.1g/ ml)、濾過滅菌したものを休止菌体反応液に終濃度 0.1% (vol/ vol) となるよう添加した。IP01 株は 30°C、*E. coli* は 37°C で各々 18 時間反応を行った。

4- 2- 9 . High- performance liquid chromatography (HPLC) 分析

本章では、以下に示す条件で HPLC 分析を行った。

| | |
|----------|------------------------------|
| カラム | ODSpak F- 411A (昭和電工) |
| detector | Shodex UV/ VIS M- 315 (昭和電工) |

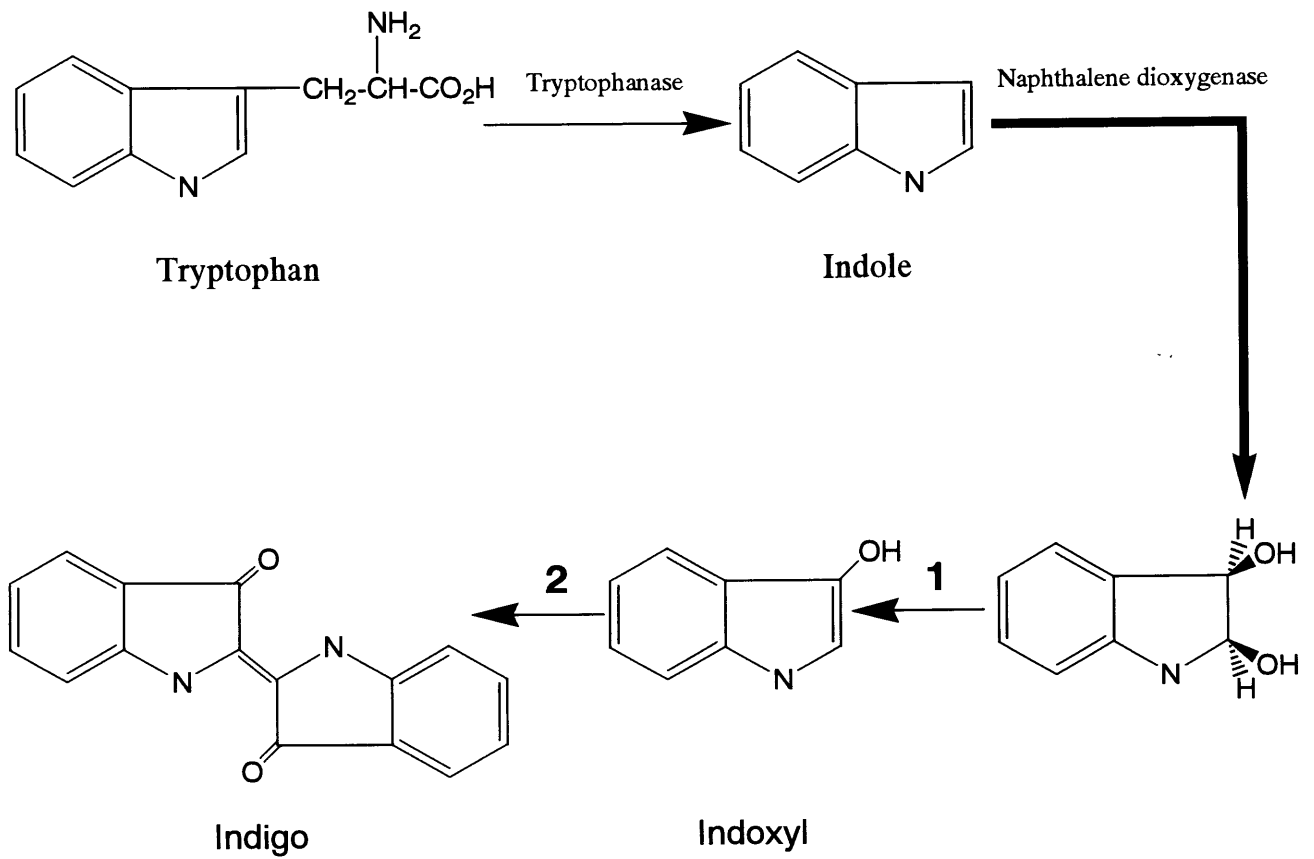


Fig. 4-1. Proposed pathway for indigo biosynthesis in a recombinant *E. coli* strain [Ensley et al., 1983]. 1: spontaneous elimination of water, 2: air oxidation.

| | |
|----|--|
| 波長 | 254 nm |
| 溶媒 | 70% acetonitrile/ 10 mM H ₃ PO ₄ |

4- 2- 1 0 . DNA 塩基配列決定法

本章で DNA 塩基配列の決定は 2- 2- 15 に示した方法に従って行った。

4- 2- 1 1 . DNA 塩基配列の accession number

本章で決定した塩基配列は、DDBJ、EMBL、GenBank nucleotide sequence database に登録しており、accession number は D37828 である。

4-3. 結果

4-3-1. IP01 株の同定及び諸性質

土壌サンプルから集積培養により単離された cumene 資化菌 IP01 株について、Bergey's Manual [Krieg and Holt 1984] に従って同定を行った。IP01 株はグラム陰性桿菌、運動性、オキシダーゼ陽性、カタラーゼ陽性など Table 4-2 に示す特徴を有しており、蛍光色素の蓄積や炭素源の資化性から、典型的な *Pseudomonas fluorescens* と同定された。

また、cumene 以外の生育基質特異性を検討したところ、IP01 株は toluene、propylbenzene、*n*-butylbenzene、isobutylbenzene に良好に生育したことから、炭素鎖 1～4 程度の側鎖を持つ 1 置換 benzene 類を生育基質とする菌であることが明らかとなった (Table 4-3)。Benzene についてはその毒性からか資化することができず、また側鎖の大きな芳香族化合物についても炭素源としては利用できなかった。

一方、cumene で生育させた IP01 株の休止菌体を用いて各種芳香族化合物に対する共酸化能を調べたところ、生育基質よりもさらに広い基質特異性を持つことが明らかとなった (Table 4-3)。炭素鎖 5～8 程度の側鎖を持つ 1 置換 benzene 類に加え、Co-PCB の基本骨格をなす biphenyl に対しても分解が可能であることが確認された。

4-3-2. Cumene 酸化酵素遺伝子の取得

IP01 株の total DNA を制限酵素 *EcoRI*、*HindIII* で個別に切断した後、pUC118 の対応する制限酵素サイトに挿入してゲノムライブラリーを作製した。*E. coli* JM109 株をこのゲノムライブラリーで形質転換し、約 10,000 コロニーからインジゴ色を呈する 3 種のクローンを取得した。それぞれのプラスミド (pIP101、pIP103、pIP201 と命名) について各種制限酵素で切断することにより制限酵素地図を作製したところ (Fig. 4-2)、これら 3 つの遺伝子はそれぞれ重複がなく、異なる遺伝子であることが確認された。

pIP101、pIP103、pIP201 のそれぞれを保持する *E. coli* を用いて cumene を基質とした休止菌体反応を行った。pIP103 を保持する *E. coli* においてのみ、反応溶液がメタ開裂物質 HOMODA の生成に起因すると考えられる黄色を呈し、かつ HPLC 分析により cumene の減少が確認された。他のプラスミドを保持する *E. coli* については、基質に対する有意な分解は認められなかった。そこで、pIP103 についてさらに詳細な解析を行うこととした。

4-3-3. pIP103 にコードされる分解系酵素群の基質特異性

取得された cumene 分解系酵素遺伝子群を保持する *E. coli* (pIP103) を用いて休止菌体反応を行い、その分解基質特異性を親株 IP01 株の芳香族分解能と比較した (Table 4-3)。その結果、取得された遺伝子にコードされている分解系酵素の基質特異性は、

Table 4-2. Identification of strain IP01.

| | Characteristic |
|---------------------------------------|----------------|
| Morphology | rod |
| Gram staining | — |
| Spore | — |
| Motility | + |
| Flagella | polar, >1 |
| Oxidase | + |
| Catalase | + |
| OF test | + |
| Fluorescent pigment | + |
| Water-soluble pigment | + |
| Accumulation of PHB | — |
| Protocatecuete, <i>ortho</i> cleavage | + |
| Growth at 40°C | + |
| Arginine dihydrolase | + |
| Denitrification | — |
| Gelatin hydrolysis | — |
| Starch hydrolysis | — |
| Utilization of | |
| Glucose | + |
| Trehalose | — |
| Mannitol | + |
| Inositol | + |
| Geraniol | — |
| L-Valine | + |
| β-Alanine | + |
| DL-Arginine | + |
| Testosterone | — |
| Quinone type | Q-9 |

Table 4-3. Substrate specificities of *P. fluorescens* IP01 and *E. coli* JM109 transformed with the cloned cumene degradation genes.

| Substrate | <i>P. fluorescens</i> IP01 | | <i>E. coli</i> JM109 (pIP103) |
|------------------------------|----------------------------|---------------------------|---|
| | Growth ^a | Co-oxidation ^b | <i>meta</i> -Cleavage compound ^c |
| Benzene | — | + | + |
| Toluene | ++ | + | + |
| Ethylbenzene | + | + | + |
| <i>n</i> -Propylbenzene | ++ | + | + |
| <i>n</i> -Butylbenzene | ++ | + | + |
| <i>n</i> -Pentylbenzene | — | + | + |
| <i>n</i> -Hexylbenzene | — | + | + |
| <i>n</i> -Heptylbenzene | — | + | + |
| <i>n</i> -Octylbenzene | — | + | + |
| <i>n</i> -Dodecylbenzene | — | — | — |
| <i>o</i> -Xylene | — | — | — |
| <i>p</i> -Xylene | — | — | — |
| Stylene | — | + | + |
| Allylbenzene | — | + | + |
| α -Methylstyrene | — | + | + |
| β -Methylstyrene | — | + | + |
| Isopropylbenzene (Cumene) | ++ | + | + |
| <i>sec</i> -Butylbenzene | + | + | + |
| <i>tert</i> -Butylbenzene | + | + | + |
| Isobutylbenzene | ++ | + | + |
| <i>p</i> -Cymene | — | — | — |
| <i>m</i> -Diisopropylbenzene | — | — | — |
| <i>p</i> -Diisopropylbenzene | — | — | — |
| Biphenyl | — | + | + |
| 4,4'-Dichlorobiphenyl | NT ^d | NT | + |

^aCells were cultivated at 30°C for 72 h in CFMM containing 0.1% (wt/vol) each substrate. ++, Good growth; +, moderate growth; —, no growth.

^b+ indicates that the metabolites of the substrate were detected by HPLC. — means that no metabolites of the substrate were detected.

^c+ indicates that the culture medium turned yellow in the resting cell reaction of the substrate using the recombinant strain *E. coli* JM109 (pIP103; *cumA 1A 2A 3A 4BC*). — means that the color of the reaction solution did not change.

^dNT indicates 'Not Tested'.

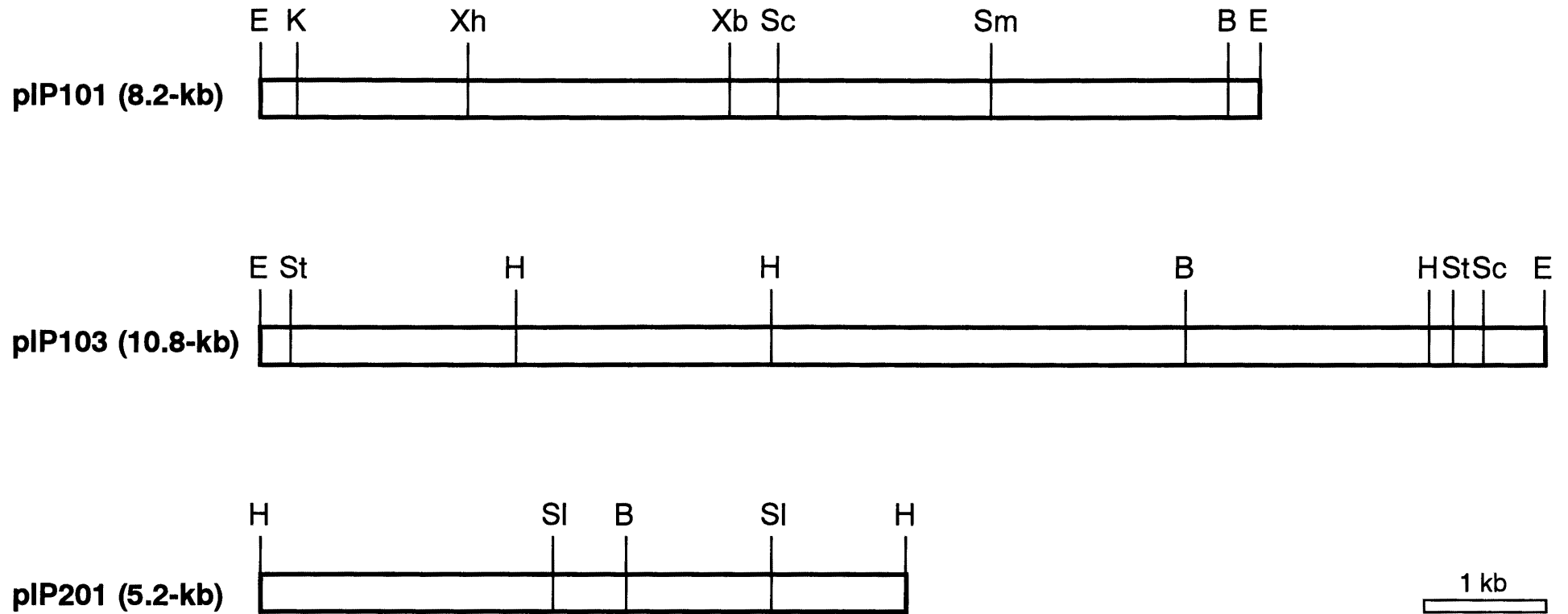


Fig. 4-2. Restriction maps of the insert DNA fragment of pIP101, pIP103, and pIP201. Restriction site designations: B, *Bam*HI; E, *Eco*RI, H, *Hind*III; K, *Kpn*I; Sc, *Sac*I; Sl, *Sal*I; St, *Stu*I; Xb, *Xba*I, Xh, *Xho*I.

親株の共酸化能を調べた結果とよく一致することが明らかとなった。また、分解が認められた基質については必ずメタ開裂まで反応が進み、dead-end product などの蓄積は認められなかった。また、Co-PCB の一つである 4,4'-dichlorobiphenyl も相当するメタ開裂物質まで変換されることが明らかとなった。

4-3-4. 初発酸化酵素活性、脱水素酵素活性、メタ開裂酵素活性を示す領域の限定

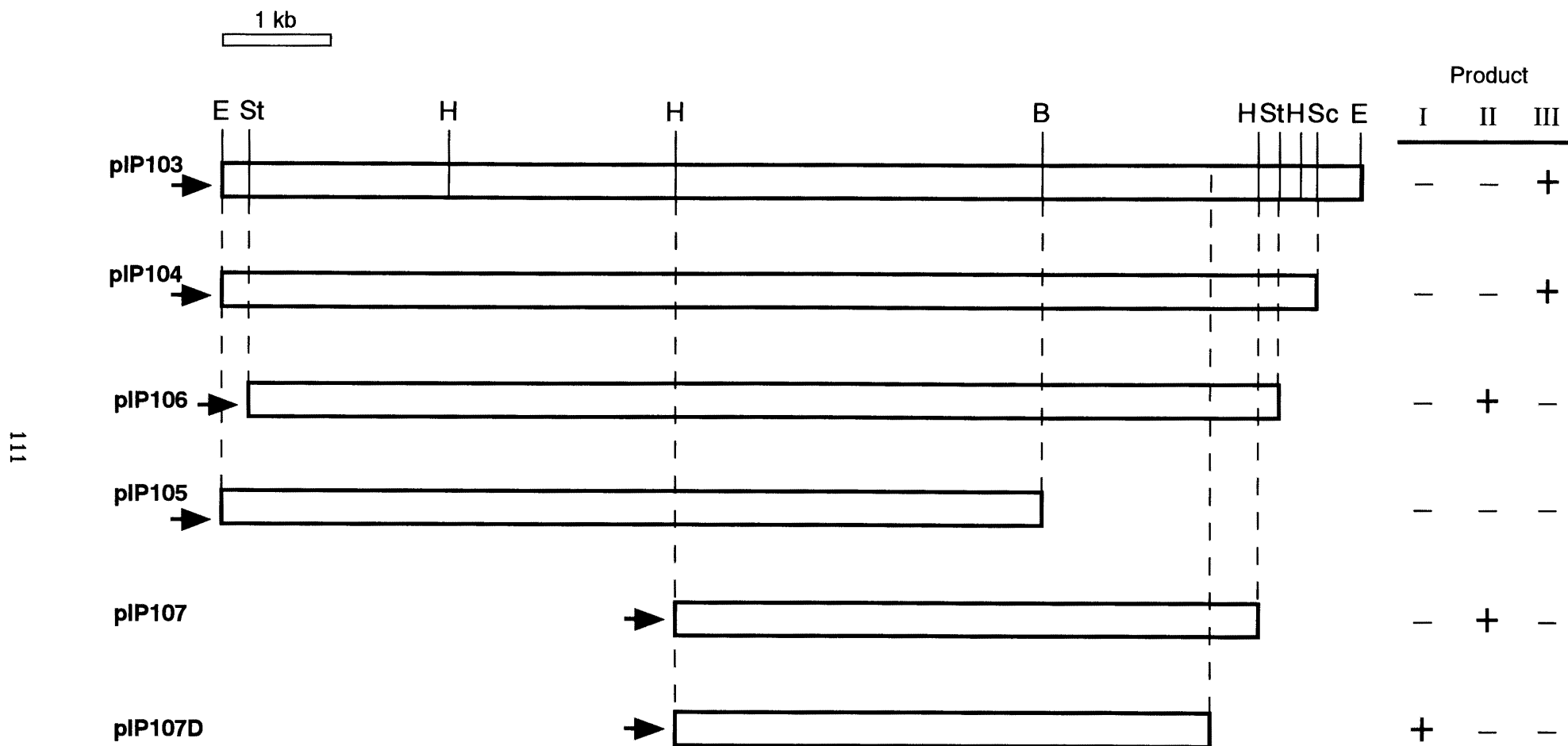
pip103 の cumene 分解に関与する各酵素活性の発現に必要な領域の同定を試みた。pip103 から各種サブクローンを構築し、それらを保持する *E. coli* を用いて cumene または benzene を基質とした休止菌体反応を行った (Fig. 4-3)。その結果、詳細なデータは示さないが、1) pip104 で cumene を基質とした場合に反応溶液がメタ開裂物質の生成に起因する黄色を呈したこと、2) pip107 で cumene を基質とした場合に反応溶液は黄色にならず、さらに GC-MS 分析により TMS 化された 3-isopropylcatechol と推定される化合物 [m/z 296 (M^+), 281 ($M^+ - CH_3$), 208, 193, 73] が検出されたこと、3) pip107D で benzene を基質とした場合に *cis*-benzene dihydrodiol が蓄積したことから、cumene を相当するメタ開裂物質まで分解する酵素遺伝子群は、約 6.8-kb の *HindIII*-*SacI* 断片内に位置することが予想された。また、cumene を 3-isopropylcatechol まで分解する酵素遺伝子群は約 5.5 kb の *HindIII* 断片内に存在することも示された。

一方、代謝産物の同定結果から、これら酵素群による cumene の分解経路は、芳香環の初発酸化による 2 酸素原子添加反応後、カテコール体を経由してメタ開裂されるというオーソドックスな分解系であることも示された (Fig. 4-4)。

4-3-5. Cumene 代謝系酵素遺伝子群 (*cum* 遺伝子) の塩基配列の決定

Cumene の分解系酵素がコードされていると予想された約 6.5-kb の *HindIII* 断片内に関して各種サブクローンと deletion mutant を構築し、各プラスミドインサート DNA の両方の strand について塩基配列を決定した。決定した塩基配列についてコンピュータ解析を行ったところ、7 個の ORF (上流から ORF1~ORF7 と命名) が認められ、cumene から HOMODA への分解系酵素遺伝子群は Fig 4-4 に示す遺伝子構造をとっていることが明らかとなった (決定した塩基配列と 7 個の ORF の推定アミノ酸配列及び推定 SD 配列 [Shine and Dalgarno 1975] を付図 5 [論文の末尾に添付] に示した)。7 個の ORF は全て同じ向きにコードされており、polycistronic な転写を受けていると考えられた。また、今回塩基配列を決定した領域の GC 含量は約 53%と、*P. fluorescens* total DNA の平均 GC 含量 (約 59-61%) に比べるとかなり低い値を示した。

また興味深いことに、ORF7 終止コドンの約 50 bp 下流に、それ自体対称な GCATGC (制限酵素 *SphI* の認識サイトと同じ配列) が、間に C を挟んで 4 回繰り返す配列が存在した (付図 5)。さらにそのモチーフの上流と下流にも、中央から折り返すと相補す



111

Fig. 4-3. Restriction map and deletion analysis of genes encoding enzymes involved in cumene degradation. The boxes indicate the extent of individual subclones constructed in plasmid vector pUC118. The arrows indicate the orientation of the *lac* promoter of this vector. The accumulation of products I, II, and III from resting cell reactions using *E. coli* transformed with each subcloned plasmid are indicated by + or -. Products I, II, and III are 2,3-dihydro-2,3-dihydroxycumene, 3-isopropylcatechol, and HOMODA, respectively. Restriction site designations: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; Sc, *Sac*I; St, *Stu*I.

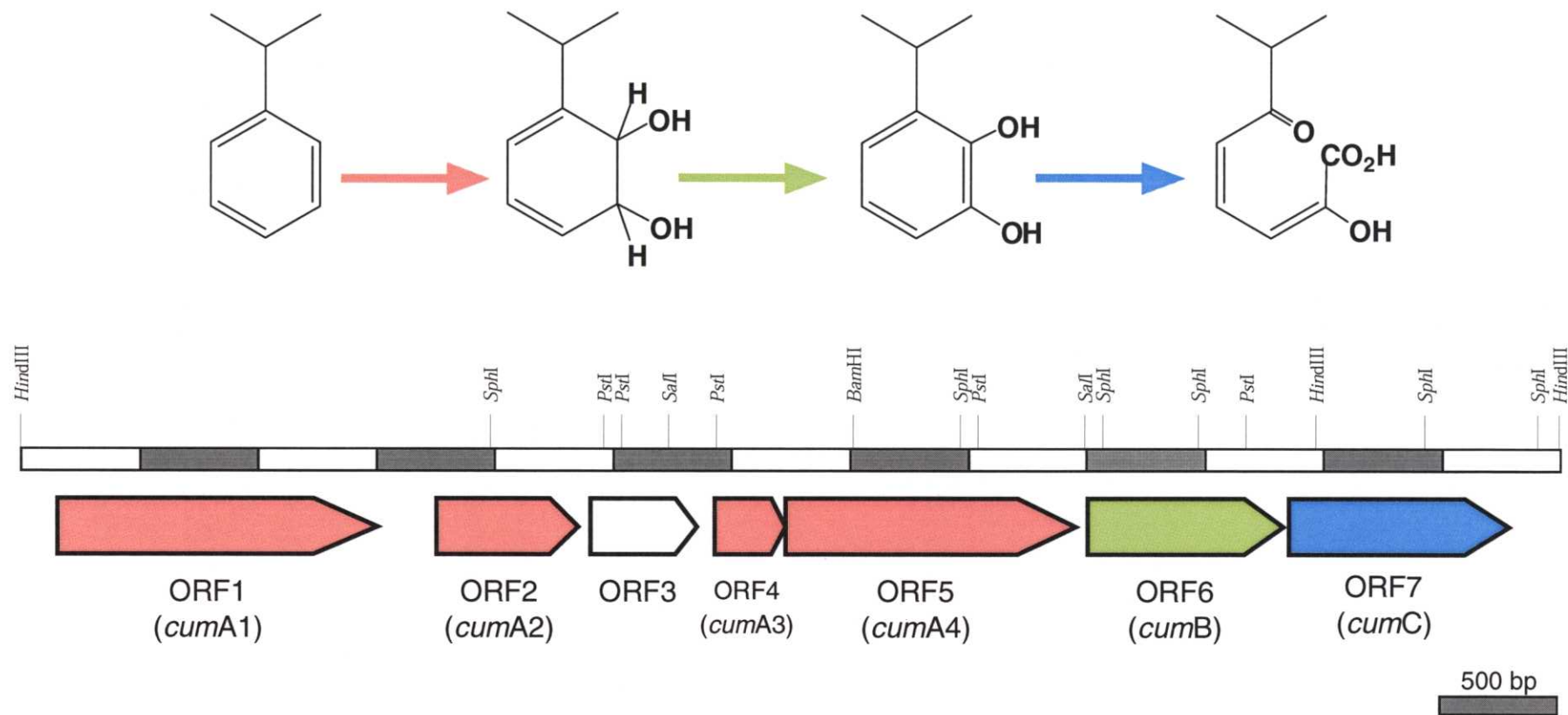


Fig. 4-4. Physical map of the 6,508-bp *Hind*III fragment derived from pIP103.

る TGA、GCA が存在した。このような配列が、mRNA の二次構造の形成に関与し、なんらかの調節、特に転写終結のシグナルになっている可能性が推測される。

4-3-6. 各 ORF 推定アミノ酸配列の相同性検索

7 個の ORF の推定アミノ酸配列と相同性を持つ配列を SWISS PROT protein database より検索したところ、ORF1, 2, 4, 5 については biphenyl 分解系の BphA1A2A3A4 や toluene 分解系の TodC1C2BA をはじめとする芳香族化合物分解系の初発酸化酵素 (multicomponent dioxygenase)、ORF6 については BphB や TodD などの脱水素酵素 (*cis*-diol dehydrogenase)、ORF7 は BphC や TodE などのメタ開裂酵素 (extradiol dioxygenase) と高い相同性を示した (Table 4-4)。そこで ORF1, 2, 4, 5, 6, 7 をそれぞれ、*cumA1*, *cumA2*, *cumA3*, *cumA4*, *cumB*, *cumC* と命名した。*cumA1A2ORF3A3A4BC* は、*Burkholderia* (formerly *Pseudomonas*) sp. LB400 株 [Erickson and Mondello 1992, Hofer et al., 1993] や *Pseudomonas pseudoalcaligenes* KF707 株 [Taira et al., 1992] の biphenyl dioxygenase、2,3-dihydro-2,3-dihydroxybiphenyl dehydrogenase、2,3-dihydroxybiphenyl 2,3-dioxygenase をコードする *bphA1A2ORF3A3A4BC* 遺伝子群と、相同性だけでなく遺伝子構造も非常に類似していることが明らかとなった。

CumA1 には、Rieske-type の [2Fe-2S] クラスタールとの結合に関与するアミノ酸配列である CXHX₁₆₋₁₇CXXH [Rieske et al., 1964] が存在していた。また、反応に直接関与すると考えられる二価鉄イオンの結合領域と考えられる配列 [Jiang et al., 1996] が存在するなど、multicomponent dioxygenase system の terminal oxygenase large subunit の特徴を有していた (Fig. 4-5A)。CumA3 にも、Rieske-type の [2Fe-2S] クラスタールとの結合に関与するアミノ酸配列 [Rieske et al., 1964] が認められ、multicomponent dioxygenase system の ferredoxin であることが示された (Fig. 4-5B)。また CumA4 には、multicomponent dioxygenase system の ferredoxin reductase に見られる FAD および NAD の結合部位 [Mason and Cammack 1992] と考えられる共通配列 GXGX₂GX₃A [Niedle et al., 1992] が完全に保存されていた (付図 5)。

一方、cumene dioxygenase をコードする CumA1A2 と CumA3A4 の間の位置に存在する ORF3 の推定アミノ酸配列は、biphenyl 分解系に見られる LB400 株の ORF1 [Ericson and Mondello 1992]、KKS102 株の ORF4 [Kikuchi et al., 1994]、KF707 株の ORF3 [Taira et al., 1992] とそれぞれアミノ酸レベルで 70、60、58% の相同性を示した。KF707 株において ORF3 が biphenyl の初発酸化に必須ではないことが示されているが [Taira et al., 1992]、その機能については現在までのところ明らかになっていない。

さらに、CumC には、二価鉄イオンの結合に関与すると考えられている 3 つの His (H) 残基、1 つの Tyr (Y) 残基 [Hofer et al., 1993] が保存されていた (付図 5)。KF707

Table 4-4. GC content and amino acid sequence similarities between degradation enzymes of *P. fluorescens* IP01 and those of other strains^a

| Enzymes | <i>P. fluorescens</i> IP01 | | | <i>P. putida</i> F1 | | | <i>Burkholderia</i> sp. LB400 | | | | |
|--------------------------------|----------------------------|---------|-------------------|---------------------|---------|-------------------|-------------------------------|-------|---------|-------------------|---------------------------|
| | | G+C (%) | No. of amino acid | | G+C (%) | No. of amino acid | Identity ^b (%) | | G+C (%) | No. of amino acid | Identity ^b (%) |
| Dioxygenase | | | | | | | | | | | |
| ISP large subunit ^c | CumA1 | 50.9 | 459 | TodC1 | 60.4 | 450 | 64.5 | BphA1 | 62.2 | 458 | 72.1 |
| Small subunit | CumA2 | 49.6 | 186 | TodC2 | 58.4 | 189 | 51.5 | BphA2 | 58.1 | 213 | 59.2 |
| Unknown ORF | ORF3 | 59.5 | 153 | — | — | — | — | ORF1 | 57.9 | 139 | 57.7 |
| Ferredoxin | CumA3 | 52.7 | 109 | TodB | 56.8 | 107 | 56.1 | BphA3 | 56.1 | 109 | 78.0 |
| Reductase | CumA4 | 57.9 | 411 | TodA | 63.7 | 410 | 51.2 | BphA4 | 68.9 | 408 | 72.9 |
| Dihydrodiol dehydrogenase | CumB | 53.4 | 276 | TodD | 60.4 | 274 | 57.8 | BphB | 61.0 | 277 | 72.9 |
| Extradiol dioxygenase | CumC | 52.5 | 312 | TodE | 58.2 | 291 | 46.9 | BphC | 60.3 | 298 | 56.7 |

^aThe amino sequences of the enzymes are from the GenBank DNA database.

^bCompared with *P. fluorescens* IP01 enzymes.

^cISP: Iron-sulfur protein.

Large subunit of terminal oxygenase component

| | | CXH | CXXH | |
|----------------|----|--------------------------------------|---------------|-----|
| CumA1 (IP01) | 96 | FLNQCRHRGMRIERSDFGNAKSFTCTYHG | | 124 |
| BphA1 (LB400) | 96 | FLNQCRHRGMRICRSDAGNAKAFTCSYHG | | 124 |
| BphA1 (KF707) | 96 | FLNQCRHRGMRICRSDAGNAKAFTCSYHG | | 124 |
| BphA1 (KKS102) | 96 | FLNQCRHRGMRICRSDAGNAKAFTCTYHG | | 124 |
| TodC1 (F1) | 96 | FLNQCRHRGMRICRADAGNAKAFTCSYHG | | 124 |
| NahAc (PpG7) | 77 | FLNVCRHRGKTLVSVEAGNAKGFVCSYHG | | 105 |
| DoxB (C18) | 77 | FLNVCRHRGKTLVSVEAGNAKGFVCSYHG | | 105 |
| PahAc (OUS82) | 77 | FLNVCRHRGKTLVNAEAGNAKGFVCSYHG | | 105 |
| | | *** ***** | ***** * * *** | |

Ferredoxin

| | | CXH | CXXH | |
|----------------|----|--------------------------------------|-------------|----|
| CumA3 (IP01) | 40 | QDRCTHGDWSLSEGGYLEGDIVECSLHMG | | 68 |
| BphA3 (LB400) | 40 | QDRCTHGDWSLLDGGYLEGDVVECSLHMG | | 68 |
| BphA3 (KF707) | 40 | QDRCTHGDWSLSDGGYLEGDVVECSLHMG | | 68 |
| BphA3 (KKS102) | 40 | QDRCTHGEWSLSDGGYLEGDVVECSLHMG | | 68 |
| TodB (F1) | 40 | QDTCTHGDWALSD-GYLDGDIVECTLHFG | | 67 |
| NahAb (PpG7) | 42 | DNLCTHGSARMSD-GYLEGREIECPLHQG | | 69 |
| DoxA (C18) | 42 | DNLCTHGSARMSD-GYLEGREIECPLHQG | | 69 |
| PahAb (OUS82) | 42 | DDLCTHGAARMSD-GYLEGREIECPLHQG | | 69 |
| | | *** * * * * | *** * * * * | |

Fig. 4-5. Highly conserved regions containing two Cys-His pairs in CumA1 and CumA3 shown in comparison with the corresponding amino acid sequences of Bph [Erickson and Mondello, 1992; Taira et al., 1992; Fukuda et al., 1994], Tod [Zylstra and Gibson, 1989], Nah [Simon et al., 1993], Dox [Denome, 1993], Pah [Takizawa et al., 1994], multicomponent dioxygenases. Two Cys-His pairs showed bold character and the amino acids identical in all proteins are indicated by asterisks.

株の BphC において、site- direct mutagenesis により His²¹⁰ が活性に必要不可欠なアミノ酸残基であることが示されているが [Taira et al., 1992]、alignment の結果、CumC にも相当する His (H) 残基が確認された。

4-4. 考察

本章では、cumene 資化菌 *P. fluorescens* IP01 株において、cumene から HOMODA への分解代謝に関与する酵素遺伝子群 (*cumA1A2ORF3A3A4BC*) を取得し、塩基配列を決定した。加えて、本代謝系酵素群が、Co-PCB である 4,4'-dichlorobiphenyl やその基本骨格を成す biphenyl に対しても分解活性を持つことを明らかにした。この結果から、PCB 分解菌として詳細な遺伝子解析が進んでいる biphenyl 分解系の *bph* 遺伝子群と IP01 株の *cum* 遺伝子群との相同性や遺伝子構造の関連性について興味を持たれた。

その後の研究により、現在 IP01 株の cumene 代謝系酵素遺伝子群については、全代謝系遺伝子の構造 (*cumA1A2ORF3A3A4BCEGFHD*) が決定され、この構造は *P. putida* F1 株の toluene 代謝系 *tod* 遺伝子群や *Burkholderia* sp. LB400 株の biphenyl 代謝系 *bph* 遺伝子群と類似していることが明らかとなっている [野尻, 1998]。特に *bph* 遺伝子群とは類縁性が高く、例えば、1) 加水分解酵素 CumD を除く全ての代謝系酵素群は、IP01 株の生育基質でもある toluene の分解系遺伝子群 (*tod*) よりも *bph* 遺伝子群の各酵素と相同性が高い、2) *bph* 遺伝子群に隣接して存在する機能未知の ORF と相同性の高い ORF3 が *cum* 遺伝子群にも存在する、3) LB400 株や KF707 株の *bph* 遺伝子群とは各 ORF の大きさや遺伝子群の構造も非常に類似しているといった特徴を有しており、両遺伝子群の分子進化上の関連が示唆された (Fig. 4-6)。

しかしながら、*cum* 遺伝子群と *bph* 遺伝子群の間で相同性や構造の類似性が見られるものの、IP01 株は biphenyl を生育基質とすることができない。この原因として、biphenyl あるいはその代謝産物が *cum* 遺伝子群の発現誘導を行えない可能性も考えられるが、IP01 株においては既に加水分解酵素 CumD が biphenyl のメタ開裂物質 HOPDA に対してほとんど活性を示さないことが、酵素学的諸性質及びタンパク質の立体構造から説明されている [佐久, 2000]。CumD と BphD [Nandhagopal et al., 1997] の基質結合ポケットは非常によく似た形をしているが、基質結合ポケットに存在するアミノ酸残基の側鎖を詳細に見ていくと、CumD の方が側鎖の大きなアミノ酸残基が込み入っており、全体的にポケットの空間が狭くなっていることが明らかとなった。よってこの基質結合ポケットの狭さにより、HOPDA が CumD の基質ポケットに入りにくいことが推測されている。今後、活性中心を構成するアミノ酸残基の 1 つ Ser (S) 残基を Ala (A) 残基に置換した CumD 変異体酵素 [池田, 1998] を用いて CumD 基質複合体の結晶構造が解明されれば、基質認識メカニズムが明らかになると予想される。

一方、本研究が開始されるまでに、cumene 資化菌 *P. putida* RE204 株の cumene 代謝に関与する遺伝子群がプラスミド上にコードされており、かつ約 20-kb の DNA 領域が頻繁に欠失することで cumene 資化能を失うという報告がなされていた [Eaton and Timmis 1986a, 1986b]。また、別の cumene 資化菌 *Rhodococcus erythropolis* BD2 株

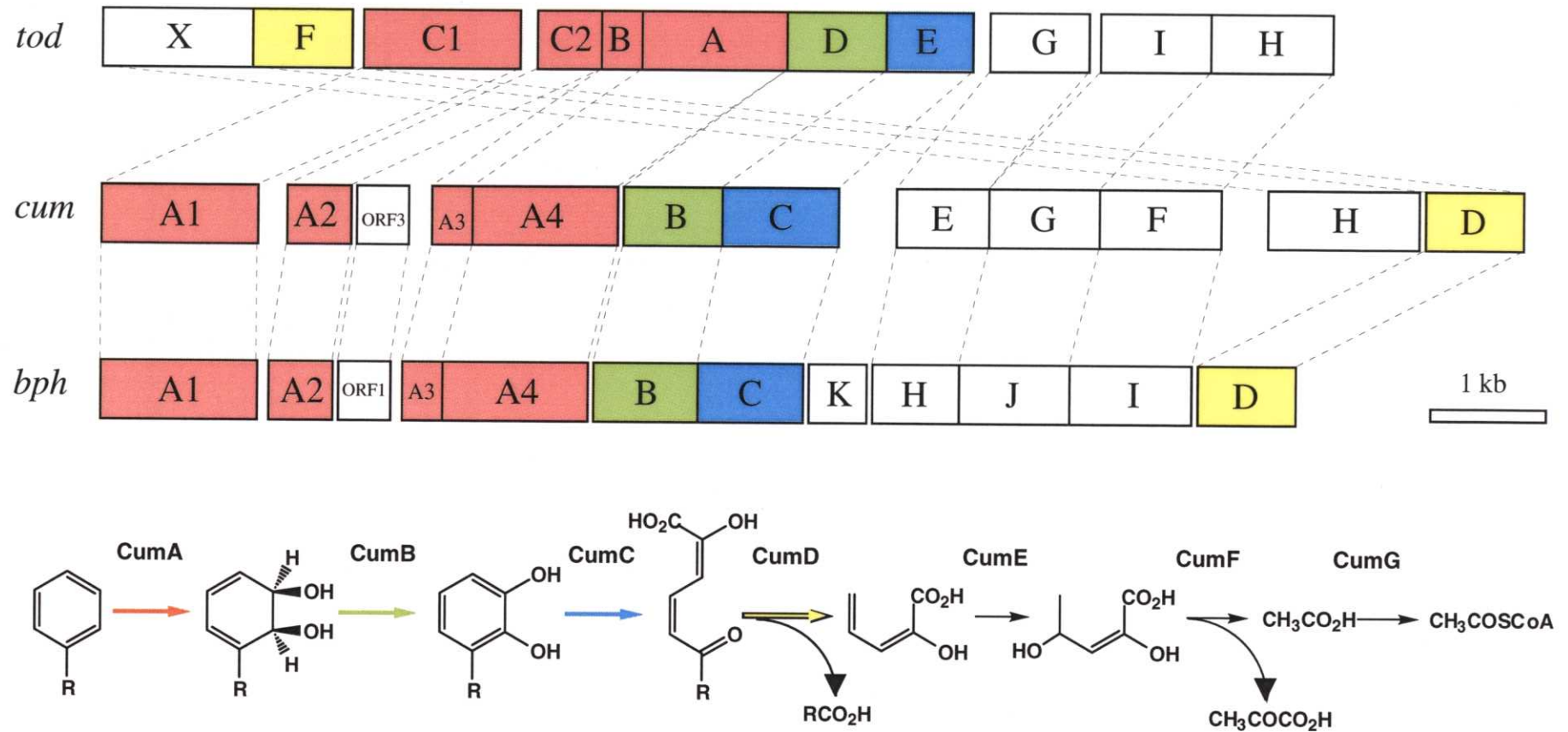


Fig. 4-6 Comparison of the DNA regions containing the genes involved in the degradation of cumene, toluene (*tod* gene cluster from *P. putida* F1), and biphenyl (*bph* gene cluster from *Burkholderia* sp. LB400). Genes are indicated by boxes and corresponding letters. [野尻, 1998より引用]

においても、cumene 代謝系遺伝子群は伝達性の線状プラスミドにコードされており、プラスミド欠損変異株が cumene 資化能を失うという報告がなされていた [Dabrock et al., 1994]。さらに我々が、IP01 株の *cumA1A2ORF3A3A4BC* 遺伝子群の塩基配列を発表した直後に [Aoki et al., 1996]、ドイツで単離された cumene 資化菌 *Pseudomonas* sp. JR1 株の同酵素遺伝子群の塩基配列が報告され、*cumABC* とアミノ酸レベルで 94-100% の高い identity を示すことが明らかとなった [Pflugmacher et al., 1996]。このように、1) いくつかの cumene 代謝系遺伝子群がプラスミド上にコードされていること、2) 日本とドイツといった地理的にかけ離れた場所において、100%に近い相同性を持つ分解系遺伝子群を保持した異なる cumene 資化菌が単離されたこと、3) IP01 株の *cum* 遺伝子群の平均 GC 含量が約 53%と、*P. fluorescens* total DNA の平均 GC 含量 (約 59-61%) に比べるとかなり低い値を示したことなどから、IP01 株の *cum* 遺伝子群が他の微生物から転移してきた可能性が示唆されていた。

現在までに、*P. fluorescens* IP01 株 [Aoki et al., 1996; 野尻, 1998]、*P. putida* RE204 株 [Eaton et al., 1998] 及び *Pseudomonas* sp. JR1 株 [Pflugmacher et al., 1996; accession number AF155505] については cumene 代謝系遺伝子群とその周辺領域について塩基配列が決定されており、これら代謝系酵素は互いに相同性が高く、類似の遺伝子構造をとることも明らかとなった (Fig. 4-7)。しかしながら興味深いのは、各 cumene 代謝系遺伝子群の周辺領域に多数の IS 様因子や direct repeat が存在していることである。RE204 株においては、1006-bp もの長さの direct repeat が約 20 kb 離れて存在し、これらに挟まれた cumene 代謝系遺伝子が頻繁に欠失することが知られている。また、RE204 株の *ipb* 遺伝子群上流には複数の IS が存在するが、そのうち制御遺伝子 (*ipbR*) を挟む形で存在している IS1543 及び IS1544 は、IP01 株の IS1406 と DNA レベルで約 67-68%の相同性を示している。一方 JR1 株においても、*ipb* 遺伝子群上流に複数の IS と相同性のある ORF が存在するが、そのうち *ipbR* 遺伝子の直下流に存在する IS は、IP01 株の IS1406 と 100%同一であった。JR1 株と IP01 株がほぼ同じ遺伝子を有していることを考えると、まだ解析は行っていないものの IP01 株の上流もおそらく同じ遺伝子構造をとっており、2つの全く同じ IS に代謝系遺伝子が挟まれた複合トランスポゾン様構造を形成しているものと推測される。これらの結果は、IP01 株を含めた *cum* (*ipb*) 遺伝子群が“動く遺伝子”上にコードされることを示唆するものである。今後は、*cum* 遺伝子群がプラスミド上にコードされているのかなど、さらなる解析が必要である。

さらに、cumene 代謝系酵素の初発酸化酵素は、地下水や井戸水の汚染で近年問題となっている trichloroethylene (TCE) も効率よく分解することが知られている [Dabrock et al., 1992, 1994; Pflugmacher et al., 1996; Berendes et al., 1998]。実際に IP01 株の *CumA1A2A3A4* についても、難分解性有機塩素化合物である TCE を分解することが確

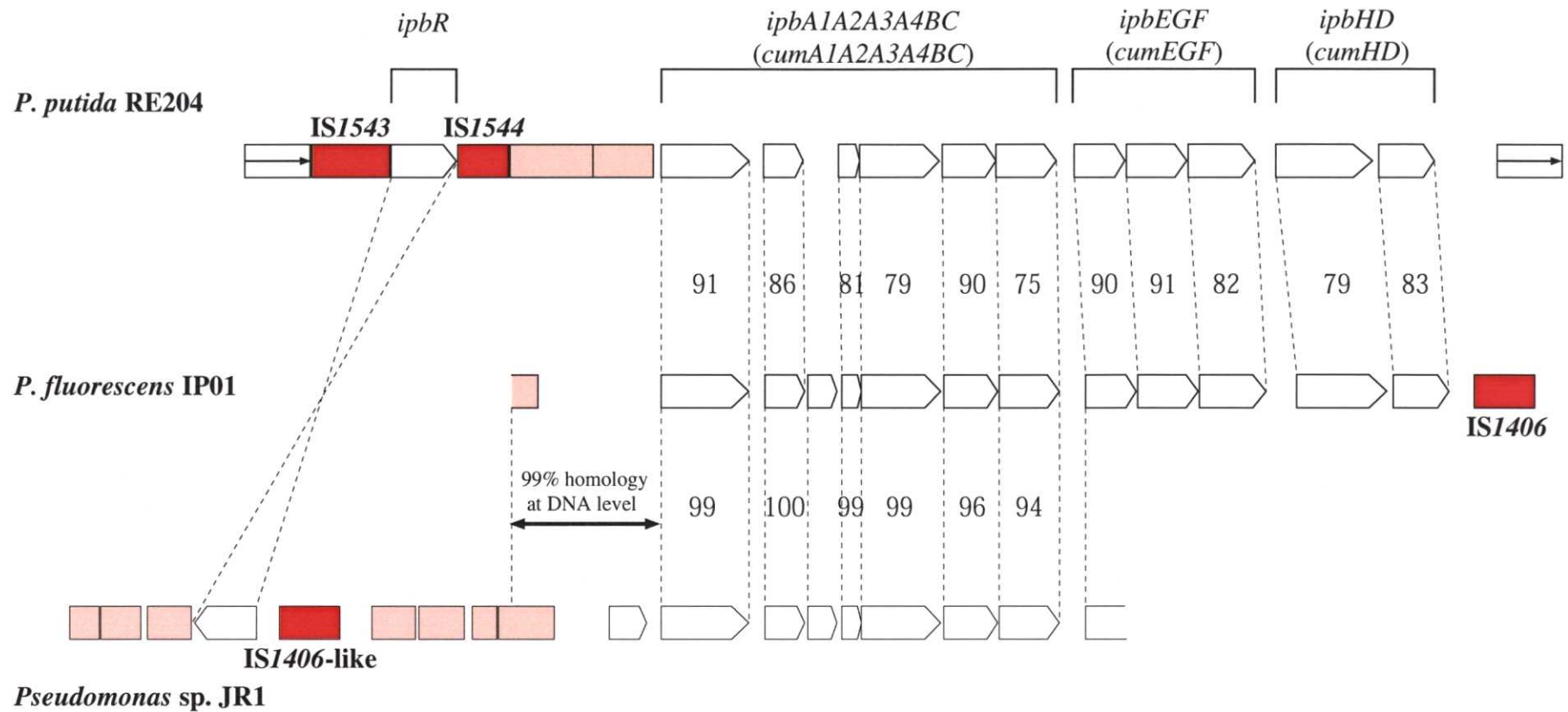


Fig. 4-7. Comparison of *ipb* (*cum*) gene clusters and their flanking regions from *P. putida* RE204, *P. fluorescens* IP01, and *Pseudomonas* sp. JR1. ORF are indicated by pentagons or boxes. DNA region homologous to ISs or transposases represent shaded boxes. Black boxes represent homologous IS elements found in these three gene clusters. The boxed arrows indicate the 1007-bp direct repeat.

認されている [Takami et al., 1999, 2000]。このように *cum* 遺伝子群が、cumene や toluene といった芳香族化合物だけでなく、PCB や TCE に対しても分解能を持っていることを考えると、今後この *cum* 遺伝子が“動く”あるいは異種微生物に転移するメカニズムを解明することは、PCB や TCE 汚染のバイオレメディエーションを高効率で行う技術を開発する上でも重要であると思われる。

総括と展望

ダイオキシン汚染はきわめて深刻な環境問題を引き起こしており、その汚染除去を最終的な目的として、ダイオキシン類の分解能を有する微生物が数多く取得・研究されている。本研究においては、ダイオキシン類と類似の化学構造を有する DF、CAR 及び cumene を資化する 3 種類の細菌から、ダイオキシン分解系酵素遺伝子群、またはその周辺の遺伝子領域を取得し、構造・機能解析を行うことを目的とした。

まず、第 2 章では、DF 資化菌 *Terrabacter* sp. DBF63 株から PCR 法により、DF 代謝に関与する初発酸化酵素 (DFDO) の oxygenase component (*dbfA1A2*) をコードする遺伝子を取得した。また DFDO の基質特異性を検討した結果、DF や DD だけでなく、数種類の 1～3 塩素化ダイオキシンも分解可能であることが示された。しかしながら、DFDO は基質特異性の高い酵素で、CAR、xanthene、phenoxathiin といった基質に対しては angular dioxygenation を触媒せず、naphthalene のような PAHs に対して *cis*-dihydroxylation を行わないことも明らかとなった。第 3 章では、CAR 資化菌 *Pseudomonas* sp. CA10 株において既に解析されていた CAR/DD 分解系遺伝子群 *carAaAaBaBbCAcORF7Ad* の周辺領域を単離・解析し、*carAd* のすぐ下流域に HPD の代謝に関与する酵素遺伝子群 (*carDFE*) が、*carAa* の約 21 kb 上流に AN から CAT の変換に関与すると推測される酵素遺伝子群 (*antABC*) が存在することを明らかにした。また、CAT 代謝に関与する酵素遺伝子群 (*catRBCA*) が chromosome 上に存在し、*car* 及び *ant* 遺伝子群はメガプラスミド pCAR1 上に存在することも示された。さらに、CAR 代謝に関与する酵素遺伝子群の全遺伝子構造を明らかにしていく過程で、4 つの互いに相同性の高い IS を同定した。これらを詳細に解析することで、*car* 遺伝子群が現在のユニークな遺伝子構造を形成するに至った遺伝子再構成の痕跡を見出すこともできた。第 4 章では、cumene 資化菌 *Pseudomonas* sp. IP01 株から、cumene 分解代謝系遺伝子群 (*cumA1A2ORF3A3A4BC*) を単離し、これが cumene から HOMODA への変換だけでなく、Co-PCB である 4,4'-dichlorobiphenyl から相当するメタ開裂物質への変換も行うことが示された。また、*cum* 遺伝子群が、biphenyl 資化菌の *bph* 遺伝子群と相同性が高く、遺伝子構造も非常に類似していることが明らかとなった。

第 2 章の研究によって、世界で初めてグラム陽性細菌から angular dioxygenase の oxygenase component (*DbfA1A2*) を単離することに成功した。*DbfA1* は、現在までに angular dioxygenase component として報告されている CA10 株の *CarAa* [Sato et al., 1997a] や *Sphingomonas wittchii* RW1 株の *DxnA1* [Armenguard et al., 1998] とは分子系統樹上では異なるグループに属していることが明らかとなった (Fig. 2-8)。加えて、各 angular dioxygenase においては基質特異性も異なっており、*DbfA1A2* と *CarAaAcAd* とでは塩素化ダイオキシンに対する反応性の違いも確認されていることから [羽部ら、

2001b]、本酵素の基質認識機構に興味を持たれている。現在当研究室においては、DbfA1A2 の結晶化に向けて研究が進められており、将来、本酵素の立体構造や塩素化ダイオキシンとの反応メカニズムに関して有用な知見が得られるものと期待される。今後、閉鎖系での bioremediation を有効かつ強力に行う手法の1つとして、このような分解系酵素の X 線結晶構造解析等から得られる基質認識機構、反応機構についての基礎的知見を集積し、それをタンパク質工学的な手法を用いた酵素の改変へとフィードバックすることで、さらに有用な分解酵素を造成していくことが重要であると考えられる。

また第3章の研究によって、*car* 及び *ant* 遺伝子群の周辺領域に、組換えや転移など遺伝子構造の変化を起こすのに重要な役割を果たしている IS が複数存在することが明らかとなり、*ant* 遺伝子群は IS5*car*2 と IS5*car*3 に挟まれた複合トランスポゾン様構造をとっていることも示された。一連の有機化合物の輸送に関与していると推測される ORF24 と ORF30 が、この複合トランスポゾンによって分断されていることや、IS5*car*2 とそれに続いて存在する *antA* 遺伝子の 5'末端部分が transposition を起こした痕跡が見出されたことから (Fig. 3-8; Fig. 3-13)、これら IS は実際に動いており、遺伝子構造変化を引き起こしているものと推測される。さらに、この IS5*car*2 の one-ended transposition により IS5*car*1 が形成された結果、全 *car* 遺伝子群を含む IS5*car*1 と IS5*car*4 に挟まれた領域もまた複合トランスポゾン様の構造をとることになる。この複合トランスポゾンについても、他の領域に転移しうるのか興味深い。現在当研究室において、*car* 遺伝子群を含む領域が相同組換え等の機構により欠失する現象が確認されているため詳細な解析を急いでいるが [前田ら, 2001]、このような現象は、*car* 遺伝子群がレプリコン上の他の領域にも転移してダイオキシン分解能を増強する可能性を示唆している。一方、CAR/DD 分解系遺伝子群はメガプラスミド pCAR1 上に局在していることも明らかとなった。芳香族化合物分解系遺伝子群がプラスミド上にコードされている例は数多く存在し、これらプラスミドが接合伝達により異種細菌に水平伝播する報告も多い。実際当研究室においても、CAR/DD 分解系遺伝子群がコードされている pCAR1 と類似のメガプラスミドが異種細菌間で転移することが既に明らかとなっている [羽部ら, 2001a]。応用的な観点からすると、ダイオキシン分解系遺伝子群の異種微生物への転移は、bioaugmentation の問題点の1つであった「汚染サイトに接種された分解菌が新しい環境中で適応できず分解能を発揮できない」という問題点を解決する糸口となるかもしれない。汚染サイトに接種された分解菌が土着の菌群にプラスミドを転移すれば、たとえ接種した分解菌が死滅したとしても、ダイオキシン分解能を付与された土着の菌を活性化させる (biostimulation) ことで、汚染土壤の修復を行うことが可能となる。以上のことから、今後は IS や pCAR1 の転移能に関して機能解析を行うとともに、pCAR1 がどのような性質のプラスミドなのか詳細な解析を行

っていく必要があると考えられる。

また、第4章より得られた結果から、PCBの基本骨格をなす biphenyl 資化菌だけでなく、単環芳香族化合物である cumene の分解系酵素が PCB 分解能を有することが明らかとなった。このように biphenyl とは化学構造の異なる物質の代謝系酵素が PCB を分解する報告が Gilbert らによってなされており、*Arthrobacter* sp. B1B 株において cumene と化学構造の類似な *l*-carbone、limonene、*p*-cymene などの代謝系酵素が PCB 共代謝を行うことが示されている [Gilbert and Crowley 1997]。これらの結果は、植物由来の terpenoids を使用することによって土壤細菌における PCB 生分解能を増強できる可能性を示しており、それ自身人体に有害でかつ水溶性の低い biphenyl に代わる PCB 分解酵素誘導基質を利用した bioremediation の新たなアプローチであると考えられる。ダイオキシン分解に関しても現在 DF や CAR といった人体に有害でかつ難溶性の基質を使用しているため、これらを土壤添加物として利用することはできない。しかし同様の手法を用いてダイオキシン分解酵素を誘導する無害な化合物をスクリーニングし、汚染土壤にその基質を繰り返し添加することで、土着の分解菌の活性を高めることが可能になると期待される。既に当研究室においても、CAR の中間代謝物である AN が塩素化ダイオキシン分解系である *car* 遺伝子群を誘導することがノーザンハイブリダイゼーション解析により明らかにされており [浦田ら, 2001]、現在、AN の bioremediation への適用の可能性を検討している。

以上、本研究で得られた基礎的知見を基に今後は、1) タンパク質工学的手法によるダイオキシン分解活性が強化された改変酵素の創製、2) ダイオキシン類分解系遺伝子群の“動く”メカニズムの解明、3) 共代謝によりダイオキシン類を分解できる無害な基質の利用といった課題について研究を展開させることで、bioremediation 技術開発の実用化に向けた基盤を築けるものと期待される。

参考文献

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.
- Aoki, H., T. Kimura, H. Habe, H. Yamane, T. Kodama, and T. Omori. 1996. Cloning, nucleotide sequence, and characterization of the genes encoding enzymes involved in the degradation of cumene to 2-hydroxy-6-oxo-methylocta-2,4-dienoic acid in *Pseudomonas fluorescens* IP01. *J. Ferment. Bioeng.* **81**: 187-196.
- Arai, H., T. Ohishi, M. -Y. Chang, and T. Kudo. 2000. Arrangement and regulation of the genes for *meta*-pathway enzymes required for degradation of phenol in *Comamonas testosteroni* TA441. *Microbiology.* **146**: 1707-1715.
- Arai, H., T. Yamamoto, T. Ohishi, T. Shimizu, T. Nakata, and T. Kudo. 1999. Genetic organization and characteristics of the 3-(3-hydroxyphenyl)propionic acid degradation pathway of *Comamonas testosteroni* TA441. *Microbiology.* **145**: 2813-2820.
- Armengaud, J., J. Gaillard, and K. N. Timmis. 2000. A second [2Fe-2S] ferredoxin from *Sphingomonas* sp. strain RW1 can function as an electron donor for the dioxin dioxygenase. *J. Bacteriol.* **182**: 2238-2244.
- Armengaud, J., B. Happe, and K. N. Timmis. 1998. Genetic analysis of dioxin dioxygenase from *Sphingomonas* sp. strain RW1: catabolic genes dispersed on the genome. *J. Bacteriol.* **180**: 3954-3966.
- Armengaud, J., K. N. Timmis, and R. -M. Wittich. 1999. A functional 4-hydroxysalicylate/hydroxyquinol degradative pathway gene cluster is linked to the initial dibenzo-*p*-dioxin pathway genes in *Sphingomonas* sp. strain RW1. *J. Bacteriol.* **181**: 3452-3461.
- Armengaud, J., and K. N. Timmis. 1997. Molecular characterization of Fdx1, a putidaredoxin-type [2Fe-2S] ferredoxin able to transfer electrons to the dioxin dioxygenase of *Sphingomonas* sp. RW1. *Eur. J. Biochem.* **247**: 833-842.
- Armengaud, J., and K. N. Timmis. 1998. The reductase RedA2 of the multi-component dioxin dioxygenase system of *Sphingomonas* sp. RW1 is related to class-I cytochrome *P*₄₅₀-type reductases. *Eur. J. Biochem.* **253**: 437-444.
- Arnett, C. M., J. V. Parales, and J. D. Haddock. 2000. Influence of chlorine substituents on rates of oxidation of chlorinated biphenyls by the biphenyl dioxygenase of *Burkholderia* sp. strain LB400. *Appl. Environ. Microbiol.* **66**: 2928-2933.
- Arthur, A., E. Nimmo, S. Hettle, and D. Sherratt. 1984. Transposition and transposition immunity of transposon Tn3 derivatives having different ends. *EMBO J.* **3**: 1723-1729.
- Ausubel, F. M., R. Brent, R. E. Kingston, D.D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. 1990. *Current Protocols in Molecular Biology*. John Wiley and Sons, Inc., N.Y.
- Avila, P., F. de la Cruz, E. Ward, and J. Grinstead. 1984. Plasmids containing one inverted repeat of Tn21 can fuse with other plasmids in the presence of Tn21 transposase. *Mol. Gen. Genet.* **195**: 288-293.
- Batic, C. J., D. P. Ballow, and C. C. Correll. 1991. Phthalate dioxygenase reductase and related flavin-iron-sulfur

- containing electron transferases, p. 543-556. In F. Muller (ed.), *Chemistry and Biochemistry of Flavoenzymes*. CRC Press, Boca Raton, Fla.
- Berendes, F., N. Sabarth, B. Averhoff, and G. Gottschalk.** 1998. Construction and use of an *ipb* DNA module to generate *Pseudomonas* strains with constitutive trichloroethene and isopropylbenzene oxidation activity. *Appl. Environ. Microbiol.* **64**: 2454-2462.
- Bertini, I., F. Capozzi, A. Dikiy, B. Happe, C. Luchinat, K. N. Timmis.** 1995. Evidence of histidine coordination to the catalytic ferrous ion in the ring-cleaving 2,2',3-trihydroxybiphenyl dioxygenase from dibenzofuran-degrading bacterium *Sphingomonas* sp. strain RW1. *Biochem. Biophys. Res. Commun.* **215**: 855-860.
- Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1513-1523.
- Bosch, R., E. Garcia-Valdes, and E. R. B. Moore.** 1999a. Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. *Gene.* **236**: 149-157.
- Bosch, R., E. Garcia-Valdes, and E. R. B. Moore.** 2000. Complete nucleotide sequence and evolutionary significance of a chromosomally encoded naphthalene-degradation lower pathway from *Pseudomonas stutzeri* AN10. *Gene.* **245**: 65-74.
- Bosch, R., E. R. B. Moore, E. Garcia-Valdes, and D. H. Pieper.** 1999b. NahW, a novel, inducible salicylate involved in mineralization of naphthalene by *Pseudomonas stutzeri* AN10. *J. Bacteriol.* **181**: 2315-2322.
- 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.
- Bressler, D. C., and P. M. Fedorak.** 2000. Bacterial metabolism of fluorene, dibenzofuran, dibenzothiophene, and carbazole. *Can. J. Microbiol.* **46**: 397-409.
- Bünz, P. V., and A. M. Cook.** 1993. Dibenzofuran 4,4a-dioxygenase from *Sphingomonas* sp. strain RW1: angular dioxygenation by a three-component enzyme system. *J. Bacteriol.* **175**: 6467-6475.
- Bünz, P. V., R. Falchetto, and A. M. Cook.** 1993. Purification and characterization of two isofunctional hydrolases [EC 3.7.1.8] in the degradative pathway for dibenzofuran in *Sphingomonas* sp. strain RW1. *Biodegradation.* **4**: 171-178.
- Coco, W. M., R. K. Rothmel, S. Henikoff, and A. M. Chakrabarty.** 1993. Nucleotide sequence and initial functional characterization of the *clcR* gene encoding a LysR family activator of the *clcABD* chlorocatechol operon in *Pseudomonas putida*. *J. Bacteriol.* **175**: 417-427.
- Collinthework, W. L., P. J. Chapman, and S. Dagley.** 1973. Stereospecific enzymes in the degradation of aromatic compounds by *Pseudomonas putida*. *J. Bacteriol.* **113**: 922-931.
- Dabrock, B., M. Keßeler, B. Averhoff, and G. Gottschalk.** 1994. Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism. *Appl. Environ. Microbiol.* **60**: 853-860.

- Dabrock, B., J. Riedel, J. Bertram, and G. Gottschalk.** 1992. Isopropylbenzene (cumene) – a new substrate for the isolation of trichloroethene-degrading bacteria. *Arch. Microbiol.* **158**: 9-13.
- Denome, S. A., D. C. Stanley, E. S. Olson, and K. D. Young.** 1993. Metabolism of dibenzothiophene and naphthalene in *Pseudomonas* strains: complete DNA sequence of an upper naphthalene catabolic pathway. *J. Bacteriol.* **175**: 6890-6901.
- Eaton, R. W.** 1996. *p*-Cumate catabolic pathway in *Pseudomonas putida* F1: cloning and characterization of DNA carrying the *cmt* operon. *J. Bacteriol.* **178**: 1351-1362.
- Eaton, R. W.** 1997. *p*-Cymene catabolic pathway in *Pseudomonas putida* F1: cloning and characterization of DNA encoding conversion of *p*-cymene to *p*-cumate. *J. Bacteriol.* **179**: 3171-3180.
- Eaton, R. W., O. V. Selifonova, and R. M. Gedney.** 1998. Isopropylbenzene catabolic pathway in *Pseudomonas putida* RE204: nucleotide sequence analysis of the *ipb* operon and neighboring DNA from pRE4. *Biodegradation.* **9**: 119-132.
- Eaton, R. W., and K. N. Timmis.** 1986a. Characterization of a plasmid-specified pathway for catabolism of isopropylbenzene in *Pseudomonas putida* RE204. *J. Bacteriol.* **168**: 123-131.
- Eaton, R. W., and K. N. Timmis.** 1986b. Spontaneous deletion of a 20-kilobase DNA segment carrying genes specifying isopropylbenzene metabolism in *Pseudomonas putida* RE204. **168**: 428-430.
- Engesser, K. H., W. Fietz, P. Fischer, P. Schulte, and H. -J. Knackmuss.** 1990. Dioxygenolytic cleavage of aryl ether bonds: 1,2-dihydro-1,2-dihydroxy-4-carboxybenzophenone as evidence for initial 1,2-dioxygenation in 3- and 4-carboxybiphenyl ether degradation. *FEMS Microbiol. Lett.* **69**: 317-322.
- Engesser, K. H., V. Strubel, K. Christoglou, P. Fisher, and H. G. Rast.** 1989. Dioxygenolytic cleavage of aryl ether bonds: 1,10-dihydro-1,10-dihydroxyfluoren-9-one, a novel arene dihydrodiol as evidence for angular dioxygenation of dibenzofuran. *FEMS Microbiol. Lett.* **65**: 205-210.
- Ensley, B. D., B. J. Ratzkin, T. D. Osslund, M. J. Simon, L. P. Wackett, and D. T. Gibson.** 1983. Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science.* **222**: 167-169.
- Erickson, B. D., and F. J. Mondello.** 1992. Nucleotide sequencing and transcriptional mapping of the genes encoding biphenyl dioxygenase, a multicomponent polychlorinated biphenyl-degrading enzyme in *Pseudomonas* strain LB400. *J. Bacteriol.* **174**: 2903-2912.
- Eulberg, D., S. Lakner, L. A. Golovlwva, and M. Schlömann.** 1998. Characterization of a protocatechuate catabolic gene cluster from *Rhodococcus opacus* 1CP: evidence for a merged enzyme with 4-carboxymuconolactone-decarboxylating and 3-oxoadipate enol-lactone-hydrolyzing activity. *J. Bacteriol.* **180**: 1072-1081.
- Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* **39**: 783-791.
- Fortnagel, P., H. Harms, R. -M. Wittich, W. Francke, S. Krohn, and H. Meyer.** 1989a. Cleavage of dibenzofuran and dibenzodioxin ring systems by a *Pseudomonas* bacterium. *Naturwissenschaften.* **76**: 222-223.
- Fortnagel, P., H. Harms, R. -M. Wittich, S. Krohn, H. Meyer, V. Sinnwell, H. Wilkes, and W. Francke.** 1990.

- Metabolism of dibenzofuran by *Pseudomonas* sp. strain HH69 and the mixed culture HH27. *Appl. Environ. Microbiol.* **56**: 1148-1156.
- Fortnagel, P., R. –M. Wittich, H. Harms, S. Schmidt, S. Franke, V. Sinnwell, H. Wilkes, and W. Francke. 1989b. New bacterial degradation of the biaryl ether structure. Regioselective dioxygenation prompts cleavage of the ether bonds. *Naturwissenschaften.* **76**: 523-524.
- Fukuda, M., Y. Yasukochi, Y. Kikuchi, Y. Nagata, K. Kimbara, H. Horiuchi, M. Takagi, and K. Yano. 1994. Identification of the *bphA* and *bphB* genes of *Pseudomonas* sp. strain KKS102 involved in degradation of biphenyl and polychlorinated biphenyls. *Biochem. Biophys. Res. Commun.* **202**: 850-856.
- Gallegos, M. –T., R. Schleif, A. Bairoch, K. Hofmann, and J. L. Ramos. 1997. AraC/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* **61**: 393-414.
- Gerlt, J. A., and P. G. Gassman. 1992. Understanding enzyme-catalyzed proton abstraction from carbon acids: details of stepwise mechanisms for β -elimination reactions. *J. Am. Chem. Soc.* **114**:5928-5934.
- Gilbert, E. S., and D. E. Crowley. 1997. Plant components that induce polychlorinated biphenyl biodegradation by *Arthrobacter* sp. strain B1B. *Appl. Environ. Microbiol.* **63**: 1933-1938.
- Habe, H., T. Kimura, H. Nojiri, H. Yamane, and T. Omori. 1996. Cloning and nucleotide sequence of the genes involved in the *meta*-cleavage pathway of cumene degradation in *Pseudomonas fluorescens* IP01. *J. Ferment. Bioeng.* **81**: 247-254.
- Haddock, J. D., J. R. Horton, and D. T. Gibson. 1995. Dihydroxylation and dechlorination of chlorinated biphenyls by purified biphenyl 2,3-dioxygenase from *Pseudomonas* sp. strain LB400. *J. Bacteriol.* **177**: 20-26.
- Halden, R. U., and D. F. Dwyer. 1997. Biodegradation of dioxin-related compounds: a review. *Bioremediation J.* **1**: 11-25.
- Halden, R. U., B. G. Halden, and D. F. Dwyer. 1999. Removal of dibenzofuran, dibenzo-*p*-dioxin, and 2-chloro dibenzo-*p*-dioxin from soils incubated with *Sphingomonas* sp. strain RW1. *Appl. Environ. Microbiol.* **65**: 2246-2249.
- Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**: 557-580.
- Happe, B., L. D. Eltis, H. Poth, R. Hedderich, K. N. Timmis. 1993. Characterization of 2,2',3-trihydroxybiphenyl dioxygenase, an extradiol dioxygenase from the dibenzofuran- and dibenzo-*p*-dioxin-degrading bacterium *Sphingomonas paucimobilis* strain RW1. *J. Bacteriol.* **175**: 7313-7320.
- Harayama, S., and M. Rekik. 1993. Comparison of the nucleotide sequences of the *meta*-cleavage pathway genes of TOL plasmid pWW0 from *Pseudomonas putida* with other *meta*-cleavage genes suggests that both single and multiple nucleotide substitutions contribute to enzyme evolution. *Mol. Gen. Genet.* **239**: 81-89.
- Harayama, S., M. Rekik, K. –L. Ngai, and L. N. Ornston. 1989. Physically associated enzymes produce and metabolite 2-hydroxy-2,4-dianoate, a chemically unstable intermediate formed in catechol metabolism via *meta* cleavage in *Pseudomonas putida*. *J. Bacteriol.* **171**: 6251-6258.
- Harms, H., H. Wilkes, V. Sinnwell, R. –M. Wittich, K. Figge, W. Francke, and P. Fortnagel. 1991.

- Transformation of 3-chlorodibenzofuran by *Pseudomonas* sp. HH69. FEMS Microbiol. Lett. **81**: 25-30.
- Harms, H., R. -M. Wittich, V. Sinnwell, H. Meyer, P. Fortnagel, and W. Francke. 1990. Transformation of dibenzo-*p*-dioxin by *Pseudomonas* sp. strain HH69. Appl. Environ. Microbiol. **56**: 1157-1159.
- Hartnett, C., E. L. Niedle, K. -L. Ngai, and L. N. Ornston. 1990. DNA sequences of genes encoding *Acinetobacter calcoaceticus* protocatechuate 3,4-dioxygenase: evidence indicating shuffling of genes and of DNA sequences within genes during their evolutionary divergence. J. Bacteriol. **172**: 956-966.
- Helin, S., P. C. Kahn, B. L. Guha, D. G. Mallows, and A. Goldman. 1995. The refined X-ray structure of muconate lactonizing enzyme from *Pseudomonas putida* PRS2000 at 1.85 Å resolution. J. Mol. Biol. **254**: 918-941.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene. **28**: 351-359.
- Hofer, B., S. Backhaus, and K. N. Timmis. 1994. The biphenyl/polychlorinated biphenyl-degradation locus (*bph*) of *Pseudomonas* sp. LB400 encodes four additional metabolic enzymes. Gene. **144**: 9-16.
- Hofer, B., L. D. Eltis, D. N. Dowling, and K. N. Timmis. 1993. Genetic analysis of a *Pseudomonas* locus encoding a pathway for biphenyl/polychlorinated biphenyl degradation. Gene. **130**: 47-55.
- Hoier, H., M. Schlömann, A. Hammer, J. P. Glusker, H. L. Carrell, A. Goldman, J. J. Stezowski, and U. Heinemann. 1994. Crystal structure of chloromuconate cycloisomerase from *Alcaligenes eutrophus* JMP134 (pJP4) at 3 Å resolution. Acta Crystallogr. Sect. **D50**:75-84.
- Ishigooka, H., Y. Yoshida, T. Omori, and Y. Minoda. 1986. Enzymatic dioxygenation of biphenyl-2,3-diol and 3-isopropylcatechol. Agric. Biol. Chem. **50**: 1045-1046.
- Ka, J., and J. M. Tiedje. 1994. Integration and excision of a 2,4-dichlorophenoxyacetic acid-degradative plasmid in *Alcaligenes paradoxus* and evidence of its natural intergenic transfer. J. Bacteriol. **176**: 5284-5289.
- Kasuga, K., H. Nojiri, H. Yamane, T. Kodama, and T. Omori. 1997. Cloning and characterization of the genes involved in the degradation of dibenzofuran by *Terrabacter* sp. strain DBF63. J. Ferment. Bioeng. **84**: 387-399.
- Katti, S. K., B. A. Katz, and H. W. Wyckoff. 1989. Crystal structure of muconolactone isomerase at 3.3 Å resolution. J. Mol. Biol. **205**:557-571.
- Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, T. Yoshizawa, Y. Nakamura, F. T. Robb, K. Horikoshi, Y. Masuchi, H. Shizuya, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. DNA Res. **5**: 55-76.
- Keim, T., W. Francke, S. Schmidt, and P. Fortnagel. 1999. Catabolism of 2,7-dichloro- and 2,4,8-trichlorodibenzofuran by *Sphingomonas* sp. strain RW1. J. Ind. Microbiol. Biotechnol. **23**: 359-363.
- Kikuchi, Y., Y. Yasukochi, Y. Nagata, M. Fukuda, and M. Takagi. 1994. Nucleotide sequence and functional analysis of the *meta*-cleavage pathway involved in biphenyl and polychlorinated biphenyl degradation in

- Pseudomonas* sp. strain KKS102. J. Bacteriol. **176**: 4269-4276.
- Kimura, N., A. Nishi, M. Goto, and K. Furukawa.** 1997. Functional analyses of a variety of chimeric dioxygenases constructed from two biphenyl dioxygenases that are similar structurally but different functionally. J. Bacteriol. **179**: 3936-3943.
- Kimura, T., Y. Zhang, T. Kodama, T. Omori.** 1996. Isolation and characterization of Tn5-induced mutants deficient in carbazole catabolism. FEMS Microbiol. Lett. **135**: 65-70.
- Krieg, N. R., and J. G. Holt.** 1984. Bergey's manual of systematic bacteriology, vol. 1. Williams & Wilkins, Baltimore.
- Kumamaru, T., H. Suenaga, M. Mitsuoka, T. Watanabe, and K. Furukawa.** 1998. Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. Nat. Biotechnol. **16**: 663-666.
- Kurkela, S., H. Levaslaiho, E. T. Plava, and T. H. Teeri.** 1991. Cloning, nucleotide sequence and characterization of genes encoding naphthalene dioxygenase of *Pseudomonas putida* NCIB9816. Gene. **73**: 355-362.
- Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London). **227**: 680-685.
- Lau, P. C. K., H. Bergerron, D. Labbé, Y. Wang, R. Brousseau, and D. T. Gibson.** 1994. Sequence and expression of the *todGIH* genes involved in the last three steps of toluene degradation by *Pseudomonas putida* F1. Gene. **146**: 7-13.
- Leveau, J. H., and J. R. van der Meer.** 1996. The *tfdR* gene product can successfully take over the role of the insertion element-inactivated *tfdT* protein as a transcriptional activator of the *tfdCDEF* gene cluster, which encodes chlorocatechol degradation in *Ralstonia eutropha* JMP134 (pJP4). J. Bacteriol. **178**: 6824-6832.
- Lipscomb, J. D., and A. M. Orville.** 1992. Mechanistic aspects of dihydroxybenzoate dioxygenase, p. 243-298. In H. Sigel and A. Sigel (ed.), Metal Ions in Biological Systems, Vol. 28. Degradation of environmental pollutants by microorganisms and their metalloenzymes. Marcel Dekker, Inc., New York, N.Y.
- Mahillon, J., and M. Chandler.** 1998. Insertion Sequences. Microbiol. Mol. Biol. Rev. **62**: 725-774.
- Masai, E., K. Sugiyama, N. Iwashita, S. Shimizu, J. E. Hauschild, T. Hatta, K. Kimbara, K. Yano and M. Fukuda.** 1997. The *bphDEF meta*-cleavage pathway genes involved in biphenyl/polychlorinated biphenyl degradation are located on a linear plasmid and separated from the initial *bphABC* genes in *Rhodococcus* sp. strain RHA1. Gene. **187**: 141-149.
- Masai, E., A. Yamada, J. M. Healy, K. Kimbara, M. Fukuda, and K. Yano.** 1995. Characterization of biphenyl catabolic genes of gram-positive polychlorinated biphenyl degrader *Rhodococcus* sp. strain RHA1. Appl. Environ. Microbiol. **61**: 2079-2085.
- Mason, J. R., and R. Cammack.** 1992. The electron-transport proteins of hydroxylating bacterial dioxygenases. Annu. Rev. Microbiol. **46**: 227-305.
- McKay, D. B., M. Seeger, M. Zielinski, B. Hofer, and K. N. Timmis.** 1997. Heterologous expression of biphenyl dioxygenase-encoding genes from a gram-positive broad-spectrum polychlorinated biphenyl degrader and

- characterization of chlorobiphenyl oxidation by the gene products. *J. Bacteriol.* **179**: 1924-1930.
- Monna, L., T. Omori, and T. Kodama. 1993. Microbial degradation of dibenzofuran, fluorene, and dibenzo-*p*-dioxin by *Staphylococcus auriculans* DBF63. *Appl. Environ. Microbiol.* **59**: 285-289.
- Mötsh, S., and R. Schmitt. 1984. Replicon fusion mediated by a single-ended derivatives of transposon Tn1721. *Mol. Gen. Genet.* **195**: 281-287.
- Narayanasamy, N., T. Senda, T. Hatta, A. Yamada, E. Masai, M. Fukuda, and Y. Mitsui. 1997. Three-dimensional structure of microbial 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid (HPDA) hydrolase (BphD enzyme) from *Rhodococcus* sp. strain RHA1, in the PCB degradation pathway. *Proc. Japan. Acad.* **73(B)**: 154-157.
- Niedle, E. L., C. Hartnett, L. N. Ornston, A. Bairoch, M. Rekik, and S. Harayama. 1991. Nucleotide sequence of the *Acinetobacter calcoaceticus* *benABC* genes for benzoate 1,2-dioxygenase reveal evolutionary relationships among multicomponent oxygenases. *J. Bacteriol.* **173**: 5385-5395.
- Nojiri, H., J. -W. Nam, M. Kosaka, K. Morii, T. Takemura, K. Furihata, H. Yamane, and T. Omori. 1999. Diverse oxygenations catalyzed by carbazole 1,9a-dioxygenase from *Pseudomonas* sp. strain CA10. *J. Bacteriol.* **181**: 3105-3313.
- Okstad, O. A., M. Gominet, B. Purnelle, M. Rose, D. Lereclus, and A. B. Kolsto. 1999. Sequence analysis of three *Bacillus cereus* loci carrying PlcR-regulated genes encoding degradative enzymes and enterotoxin. *Microbiology.* **145**: 3129-3138.
- Olivera, E. R., B. Minambres, B. Garcia, C. Muniz, M. A. Moreno, A. Ferrandez, E. Diaz, J. L. Garcia, and J. M. Luengo. 1998. Molecular characterization of the phenylacetic acid catabolic pathway in *Pseudomonas putida* U: the phenylacetyl-CoA catabolon. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 6419-6424.
- Omori, T., K. Sugiyama, H. Ishigooka, and Y. Minoda. 1986. Purification and some properties of a 2-hydroxy-6-oxo-6-phenyl-hexa-2,4-dienoic acid hydrolyzing enzyme from *Pseudomonas cruciviae* S93 B1 involved in the degradation of biphenyl. *Agric. Biol. Chem.* **50**: 931-937.
- Otaka, E., and T. Ooi. 1989. Examination of protein sequence homologies. 5. New perspective on evolution between bacterial and chloroplast-type ferredoxins inferred from sequence evidence. *J. Mol. Evol.* **29**: 246-254.
- Ouchiya, N., Y. Zhang, T. Omori, and T. Kodama. 1993. Biodegradation of carbazole by *Pseudomonas* spp. CA06 and CA10. *Biosci. Biotechnol. Biochem.* **57**: 455-460.
- Pflugmacher, U., B. Averhoff, and G. Gottschalk. 1996. Cloning, sequencing, and expression of isopropylbenzene degradation genes from *Pseudomonas* sp. strain JR1: identification of isopropylbenzene dioxygenase that mediates trichloroethene oxidation. *Appl. Environ. Microbiol.* **62**: 3967-3977.
- Pollard, J. R., and T. D. H. Bugg. 1998. Purification, characterization and reaction mechanism of monofunctional 2-hydroxypentadienoic acid hydratase from *Escherichia coli*. *Eur. J. Biochem.* **251**: 98-106.
- Powell, J. A. C., and J. A. C. Archer. 1998. Molecular characterisation of a *Rhodococcus ohp* operon. *Antonie van Leeuwenhoek.* **74**: 175-188.
- Ramos, J. L., E. Duque, P. Godoy, and A. Segura. 1998. Efflux pumps involved in toluene tolerance in

- Pseudomonas putida* DOT-T1E. J. Bacteriol. **180**: 3323-3329.
- Reverchon, S., W. Nasser, and J. Robert-Baudouy.** 1991. Characterization of *kdgR*, a gene of *Erwinia chrysanthemi* that regulates pectin degradation. Mol. Microbiol. **5**: 2203-2216.
- Rieske, J. S., D. H. MacLennan, and R. Coleman.** 1964. Isolation and properties of an iron-protein from the (reduced coenzyme Q)-cytochrome C reductase complex of the respiratory chain. Biochem. Biophys. Res. Commun. **15**: 338-344.
- Romero-Steiner, S., R. E. Parales, C. S. Harwood, and J. E. Houghton.** 1994. Characterization of the *pcaR* regulatory gene from *Pseudomonas putida*, which is required for the complete degradation of *p*-hydroxybenzoate. J. Bacteriol. **176**: 5771-5779.
- Saito, A., T. Iwabuchi, and S. Harayama.** 2000. A novel phenanthrene dioxygenase from *Nocardioides* sp. strain KP7: expression in *Escherichia coli*. J. Bacteriol. **182**: 2134-2141.
- Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406-425.
- Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 5463-5467.
- Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sato, S., J. -W. Nam, K. Kasuga, H. Nojiri, H. Yamane, and T. Omori.** 1997a. Identification and characterization of genes encoding carbazole 1,9a-dioxygenase in *Pseudomonas* sp. strain CA10. J. Bacteriol. **179**: 4850-4858.
- Sato, S., N. Ouchiyama, T. Kimura, H. Nojiri, H. Yamane, and T. Omori.** 1997b. Cloning of genes involved in carbazole degradation of *Pseudomonas* sp. strain CA10: Nucleotide sequences of genes and characterization of *meta*-cleavage enzymes and hydrolase. J. Bacteriol. **179**: 4841-4849.
- Schell, M. A.** 1993. Molecular Biology of the LysR family of transcriptional regulators. Annu. Rev. Microbiol. **47**: 597-626.
- Schmid, A., B. Rothe, J. Altenbuchner, W. Ludwig, and K. -H. Engesser.** 1997. Characterization of three distinct extradiol dehydrogenase involved in mineralization of dibenzofuran by *Terrabacter* sp. strain DPO360. J. Bacteriol. **179**: 53-62.
- Schreiner, G., T. Wiedmann, H. Schimmel, and K. Ballschmiter.** 1997. Influence of the substitution pattern on the microbial degradation of mono- to tetrachlorinated dibenzo-*p*-dioxins and dibenzofurans. Chemosphere. **34**: 1315-1331.
- Seah, S. Y. K., G. Labbé, S. R. Kaschabek, F. Reifenrath, W. Reineke, and L. D. Eltis.** 2001. Comparative specificities of two evolutionarily divergent hydrolases involved in microbial degradation of polychlorinated biphenyls. J. Bacteriol. **183**: 1511-1516.
- Seah, S. Y. K., G. Labbé, S. Nerdinger, M. R. Johnson, V. Snieckus, and L. D. Eltis.** 2000. Identification of a serine hydrolase as a key determinant in the microbial degradation of polychlorinated biphenyls. J. Biol. Chem. **275**:

15701-15708.

Seeger, M., K. N. Timmis, and B. Hofer. 1995. Conversion of chlorobiphenyls into phenylhexadienoates and benzoates by the enzymes of the upper pathway for polychlorobiphenyl degradation encoded by the *bph* locus of *Pseudomonas* sp. strain LB400. *Appl. Environ. Microbiol.* **61**: 2654-2658.

Seeger, M., M. Zielinski, K. N. Timmis, and B. Hofer. 1999. Regiospecificity of dioxygenation of di- to pentachlorobiphenyls and their degradation to chlorobenzoates by the *bph*-encoded catabolic pathway of *Burkholderia* sp. strain LB400. *Appl. Environ. Microbiol.* **65**: 3614-3621.

Shine, J., and L. Dalgarno. 1975. Determination of cistron specificity in bacterial ribosomes. *Nature.* **254**: 34-38

Shingler, V., J. Powlowski, and U. Marklund. 1992. Nucleotide sequence and functional analysis of the complete phenol/3,4-dimethylphenol catabolic pathway of *Pseudomonas* sp. strain CF600. *J. Bacteriol.* **174**: 711-724.

Simon, M. J., T. D. Osslund, R. Saunders, B. D. Ensley, S. Suggs, A. Harcourt, W. -C. Suen, D. L. Cruden, D. T. Gibson, and G. J. Zylstra. 1993. Sequence of genes encoding naphthalene dioxygenase in *Pseudomonas putida* strains G7 and NCIB9816-4. *Gene.* **127**: 31-37.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503-517.

Strubel, V., K. H. Engesser, P. Fischer, and H. -J. Knackmuss. 1991. 3-(2-Hydroxy-phenyl)catechol as substrate for proximal *meta* ring cleavage in dibenzofuran degradation by *Brevibacterium* sp. strain DPO1361. *J. Bacteriol.* **173**: 1932-1937.

Sunnarborg, A., D. Klumpp, T. Chung, and D. C. Laporte. 1990. Regulation of the glyoxylate bypass operon: cloning and characterization of *iclR*. *J. Bacteriol.* **172**: 2642-2649.

Taira, K., J. Hirose, S. Hayashida, and K. Furukawa. 1992. Analysis of *bph* operon from the polychlorinated biphenyl-degrading strain of *Pseudomonas pseudoalcaligenes* KF707. *J. Biol. Chem.* **267**: 4844-4853.

Takami, W., M. Horinouchi, H. Nojiri, H. Yamane, and T. Omori. 1999. Evaluation of trichloroethylene degradation by *E. coli* transformed with dimethyl sulfide monooxygenase genes and/or cumene dioxygenase genes. *Biotechnol. Lett.* **21**: 259-264.

Takami, W., H. Nojiri, H. Yamane, and T. Omori. 2000. Degradation of trichloroethylene by recombinant *E. coli* in continuous culture. *Biotechnol. Lett.* **22**: 211-216.

Takizawa, N., N. Kaida, S. Torigoe, T. Moritani, T. Sawada, S. Satoh, and H. Kiyohara. 1994. Identification and characterization of genes encoding polycyclic aromatic hydrocarbon dioxygenase and polycyclic aromatic hydrocarbon dihydrodiol dehydrogenase in *Pseudomonas putida* OUS82. *J. Bacteriol.* **176**: 2444-2449.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids*

Res. **24**: 4876-4882.

Thomson, N. R., W. Nasser, S. McGowan, M. Sebaihia, G. P. C. Salmond. 1999. *Erwina cartovor*a has two KdgR-like proteins belonging to the IclR family of transcriptional regulators: identification and characterization of the RexZ activator and the KdgR repressor of pathogenesis. *Microbiology*. **145**: 1531-1545.

Treadway, S. L., K. S. Yanagimachi, E. Lankenau, P. A. Lessard, G. Stephanopoulos, and A. J. Sinskey. 1999. Isolation and characterization of indene bioconversion genes from *Rhodococcus* strain I24. *Appl. Microbiol. Biotechnol.* **51**: 786-793.

Van der Ploeg, J., M. Willemsen, G. van Hall, and D. B. Janssen. 1995. Adaptation of *Xanthobacter autrophicus* GJ10 to bromoacetate due to activation and mobilization of the haloacetate dehalogenase gene by insertion element IS1247. *J. Bacteriol.* **177**: 1348-1356.

Vieira, J., and J. Messing. 1987. Production of single-stranded plasmid DNA. *Methods Enzymol.* **153**: 3-11.

White, O., J. A. Eisen, J. F. Heidelberg, E. K. Hickey, J. D. Peterson, R. J. Dodson, D. H. Haft, M. L. Gwinn, W. C. Nelson, D. L. Richardson, K. S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J. J. Vamathevan, P. Lam, L. McDonald, T. Utterback, C. Zalewski, K. S. Makarova, L. Aravind, M. J. Daly, K. W. Minton, R. D. Fleischmann, K. A. Ketchum, K. E. Nelson, S. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser. 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286**: 1571-1577.

Widada, J., 2000. Ph.D. thesis. The University of Tokyo, Tokyo, Japan.

Wilkes, H., R. -M. Wittich, K. N. Timmis, P. Fortnagel, and W. Francke. 1996. Degradation of chlorinated dibenzofurans and dibenzo-*p*-dioxins by *Sphingomonas* sp. strain RW1. *Appl. Environ. Microbiol.* **62**: 367-371.

Wittich, R. -M., C. Strömpl, E. R. B. Moore, R. Blasco, and K. N. Timmis. 1999. Interaction of *Sphingomonas* and *Pseudomonas* strains in the degradation of chlorinated dibenzofurans. *J. Ind. Microbiol. Biotechnol.* **23**: 353-358.

Wittich, R. -M., H. Wilkes, V. Sinnwell, W. Francke, and P. Fortnagel. 1992. Metabolism of dibenzo-*p*-dioxin by *Sphingomonas* sp. strain RW1. *Appl. Environ. Microbiol.* **58**: 1005-1010.

Yabuuchi, E., H. Yamamoto, S. Terakubo, N. Okamura, T. Naka, N. Fujiwara, K. Kobayashi, Y. Kosako, and A. Hiraishi. 2001. Proposal of *Sphingomonas wittichii* sp. nov. for strain RW1 (T) known as a dibenzo-*p*-dioxin metabolizer. *Int. J. Syst. Evol. Microbiol.* **51**: 281-292.

Yamada, A., H. Kishi, K. Sugiyama, T. Hatta, K. Nakamura, E. Masai, and M. Fukuda. 1998. Two nearly identical aromatic compound hydrolase genes in a strong polychlorinated biphenyl degrader, *Rhodococcus* sp. strain RHA1. *Appl. Environ. Microbiol.* **64**: 2006-2012.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*. **33**: 103-119.

Zylstra, G. J., and D. T. Gibson. 1989. Toluene degradation by *Pseudomonas putida* F1: nucleotide sequence of the *tod* *C1C2BADE* genes and their expression in *Escherichia coli*. *J. Biol. Chem.* **264**: 14940-14946.

池田直弘, 1998年, 東京大学修士論文.

浦田雅章, 野尻秀昭, 吉田貴子, 羽部浩, 大森俊雄, 2001 年, 日本農芸化学会 2001 年度大会講演要旨集, p210.

春日和, 1994 年, 東京大学博士論文.

佐久敬, 2000 年, 東京大学修士論文.

杉浦昌弘編, 1989 年, クローニングとシーケンス, 農村文化社.

東京大学医科学研究所制癌研究部編, 1993 年, 細胞工学実験プロトコール, 秀潤社.

中井誠一郎, 2001 年, 東京大学修士論文.

野尻秀昭, 1998 年, 東京大学博士論文.

野尻秀昭, 浦田雅章, 吉田貴子, 羽部浩, 大森俊雄, 2000 年, 日本農芸化学会 2000 年度大会講演要旨集, p110.

羽部浩, 新谷政己, 吉田貴子, 野尻秀昭, 大森俊雄, 2001 年 a, 日本農芸化学会 2001 年度大会講演要旨集, p355.

羽部浩, 鄭鎮成, Onruthai Pinyakong, 李宗勳, 春日和, 吉田貴子, 野尻秀昭, 大森俊雄, 2001 年 b, 日本農芸化学会 2001 年度大会講演要旨集, p260.

前田香奈, 野尻秀昭, 吉田貴子, 羽部浩, 大森俊雄, 2001 年, 日本農芸化学会 2001 年度大会講演要旨集, p355.

門奈理佐, 1993 年, 東京大学修士論文.

*Bam*HI (ORF8)

| | | |
|------|---|------|
| 1 | <u>GGATCC</u> CCCGACCGTAGTCGTCGTCGGAGCGTCGGTCGCCGGCGTGCGGACGGCCAGTC | 60 |
| 3 | D P R T V V V V G A S V A G V R T A Q S | 22 |
| 61 | TCTGCGTACCGAGGGCTTCGGCGGTTCGCATCGTCCTCGTGGGCGACGAGCCGGACCTGCC | 120 |
| 23 | L R T E G F G G R I V L V G D E P D L P | 42 |
| 121 | CTATGACAAGCCGCCGCTGTCCAAGCAATTCTTGGCCGGCTGTTGGTCCGTTGAACGGAT | 180 |
| 43 | Y D K P P L S K Q F L A G C W S V E R I | 62 |
| 181 | CGCGCTGCTGACCGCCGAGGCGGGCAGCAGTCGGGATCGAGCTGCGTCTGGGGGTCGC | 240 |
| 63 | A L L T A E A A A A V G I E L R L G V A | 82 |
| 241 | GGCGCAGCGTCTGGACGTCGCCGGCCGGCAGGTGGTGCCTCATTGATGGGAGCACGATCGA | 300 |
| 83 | A Q R L D V A G R Q V V L I D G S T I D | 102 |
| 301 | CTACGACGCTGCGGTGCTCGCCACCGGTGCTGCCGCGCGCCCCGCCCGTGGTTCGGCCAC | 360 |
| 103 | Y D A A V L A T G A A A R P A P W S A T | 122 |
| 361 | ATCGGGGGTGCATGTGCTGCGTTCATGGCCGACGCACGCGCGCTGCGTGACGACCTGGA | 420 |
| 123 | S G V H V L R S M A D A R A L R D D L E | 142 |
| 421 | GGTCGACCGGCCAGTTGTGATCATCGGCGGGGGCTTTATCGGCGCGGAGGTGGCAGCCAC | 480 |
| 143 | V D R P V V I I GGGF IGAEVA A T | 162 |
| 481 | CGCGCACGCCCTGGGCAGAGAAGTCAACGTCGTCGATCCCCTTTCCGCTCCGATCGGACG | 540 |
| 163 | A H A L G R E V T V V D P L S A P I G R | 182 |
| 541 | CGTCGTCGGCCGAGAGGTTCGGAGAGTTGTTTCGTCGACCTGCACACCCGGCACGGCGTGAC | 600 |
| 183 | V V G R E V G E L F V D L H T R H G V T | 202 |
| 601 | CACGAAGTTCGGGGTGGGTGTGGAGTCGGTCGAGGGCGAAGCCGGTGACCTAACGGTGCA | 660 |
| 203 | T K F G V G V E S V E G E A G D L T V Q | 222 |
| 661 | GTTGACCGACGGTTCGCAGTTGGCGGGCGGGGACCGTCGTTGGTTCGGGATCGGGCGCGTTC | 720 |
| 223 | L T D G S Q L A A G T V V V G I G A V P | 242 |
| 721 | CAACGACGGCTGGCTGGCGGGCTCGGGCCTCCTCGTGGACGACGGCGTCTGCGACGA | 780 |
| 243 | N D G W L A G S G L L V D D G V V C D E | 262 |
| 781 | GCACTGCCGGGCGGTTCGACGCCCGACGTGTACGCCGTCGGTGACGTGCTGCGGTGGTT | 840 |
| 263 | H C R A V D A A D V Y A V G D V L R W F | 282 |
| 841 | CCACCCGGGACACGGAGAGCATGTGCGAGTCGAGCACTGGACCAACGCGGTTGAGTCCGC | 900 |
| 283 | H P G H G E H V R V E H W T N A V E S A | 302 |
| 901 | TGCCGTGGTGGCACACAACATCGCCCACCCGAGGAGCTGCGGTCCCCTGCCCCGGTTCGA | 960 |
| 303 | A V V A H N I A H P E E L R S H C P V E | 322 |
| 961 | GTACGTGTGGAGCGACCAGTACGACTGGAAGATTTCAGATCGTCGGCCGGCCGGTTGGGGC | 1020 |
| 323 | Y V W S D Q Y D W K I Q I V G R P V G A | 342 |
| 1021 | GACCCGGCACACGCTGGTGGGGGACCTGACTAGTGCATCCCCGCGGAGCGGGCGGTTTA | 1080 |
| 343 | T R H T L V G D L T S A H P R G A A V Y | 362 |
| 1081 | CACCGATGACAGCGGGGACCTGCGGGGAGCGGTTCGCGGTTAACTGGCCCAAGGCGTTGCT | 1140 |
| 363 | T D D S G D L R G A V A V N W P K A L L | 382 |
| 1141 | GGTCTGCCGGAGGGCGGTGGGTGCGGGCGCGTTCGACGACGTCGTCGCCAGCTCGA | 1200 |
| 383 | V C R R A V G A G A S F D D V V A Q L D | 402 |
| 1201 | CCGGCTGTCGGCCGCGGGGGCGGCGCCGGCATGAGCGCGGGTACGCCGCTGAGCGGCAG | 1260 |
| 403 | R L S A A G A A P A * | 412 |
| 1 | (ORF7) M S G G Y A A E R Q | 10 |
| 1261 | CTCGTCGCGGATGCATGTCCGGTCCCGCAGCTCGTGGGCTGTCCGACGGGTTCCTGGGC | 1320 |
| 11 | L V A D A C R V A A A R G L S D G F L G | 30 |
| 1321 | CACGTCAGCCTGCGGATCGACGAGGAACGCTTGTGATCCGCTGCCGCGGCCCTCAGGAG | 1380 |
| 31 | H V S L R I D E E R L L I R C R G P Q E | 50 |
| 1381 | CGGGCTTGGCCTGGACGACGGCCGCCGACGTCACCTCGTTGACCAGGGCGGGCGCCCT | 1440 |
| 51 | R G L A W T T A A D V H L V D Q G G A P | 70 |

Appendix. Fig. 1. The nucleotide and deduced amino acid sequences of the 8,179-bp *Bam*HI fragment (pDF01). The sense DNA strand is shown. The putative SD sequences are indicated by double lines under the sequence. Astarisks indicate stop codons. The amino acids which may constitute NAD-binding site [Niedle et al., 1992] with the conserved motif GXGX₂GX₃A are indicated by shaded boxes.

1441 GGAGCTCCGGGCGAGCTCGATGGATGGTCTCCACCGAACGAATTGCCGCTGCATGTGGAG 1500
71 G A P G E L D G W S P P N E L P L H V E 90
1501 GTGCTGCGTTCCCGGCCGGAGGCCCGCTCAGTCGTCCACGTGCACCCGCGGGCGGTAGTC 1560
91 V L R S R P E A A S V V H V H P R A V V 110
1561 GCAGCAGACCTGGCTGGCCTCGCGATCCGCCCATCGTTCGGTTCGACATCCCGGGC 1620
111 A A D L A G L A I R P I V G A F D I P G 130
1621 GCCAAGCTGGCCGCCGGGGGTGCCCGTCTACCCGCGTGGCGTGTCTCGTTCCGGAACCGT 1680
131 A K L A A G G V P V Y P R G V L V R N R 150
1681 GGCCTGGCCTTGGAGATGGTGGCGGCCATGGGTAAGCGGCCGGTTCGTCTGTCTCCGCGGG 1740
151 G L A L E M V A A M G K R P V V V L R G 170
1741 CACGGCTTGACTAGCAGCGGTGGATCGGTCCAGGAAGCGGTCTTCGGGGCGATCAGCGTG 1800
171 H G L T S S G G S V Q E A V L R A I S V 190
1801 GACTCTCTGGCCCGGCTGTCCCTTACAGATCGCCGCGGGGAGGGGCGCAGGCGGATCTG 1860
191 D S L A R L S L Q I A A A G G A Q A D L 210
1861 CCGGCCGAGGACTTGGCGGGCGTGCACCGACCTCGGACCGGGCTTCAACGAGGGCGTGGCG 1920
211 P A E D L A A L P D L G P G F N E G V A 230
1921 TGGCGGCATGAAGTCGCCAGGATGGGGTGACCGCCGGGGCGCAGGTCCTCGGCGTCGGTC 1980
231 W R H E V A R M G * 239
1981 CCTGCGCAGCAGACACCCCGTAGCCTCAGACCCATGAGAGCGATGTCGAACCTGCCGCCG 2040
1 (ORF6) M R A M S N L P P 9
2041 GCACCGGAGTTTGGCTCGGGCGCCGACGTACCCGATCGAGTCGGTGGACAACGCACTCCGG 2100
10 A P E F G S A P Q Y P I E S V D N A L R 29
2101 GTGGTGCTGCTGCTGGGGGAGCACCCAGCCTGCGGCTGACGGACGTCAGCAAGTACCTG 2160
30 V V L L L G E H P S L R L T D V S K Y L 49
2161 GGGGTGGCATCGTCCACTGCCACCGGCTGCTTGCATGCTCCAGTACCGCGGCTACGTC 2220
50 G V A S S T A E R L L A M L Q Y R G Y V 69
2221 CGGCAGGATGCCGCGACGCGCAGCTACGTCCCGGGCCGACCTGGACTCGCTGGCCTTC 2280
70 R Q D A A T R S Y V P G P T L D S L A F 89
2281 GGTCTGCTGCGCCGACTCGACGCCCGGGACCGGGCACGCCCGGTGCTCGAGCGGCTCAAC 2340
90 G L L R R L D A R D R A R P V L E R L N 109
2341 AACGATCTGGGGGAGACGGTGCACCTGGGCCGGCTCGAGGGGAACGAGGTCCACTTTATC 2400
110 N D L G E T V H L G R L E G N E V H F I 129
2401 GATTCCATCGAGAGTGATCGGGCCCTGCGAGTGGCCAACCGGCTGGGGCGCAGCATGCC 2460
130 D S I E S D R A L R V A N R L G R S M P 149
2461 GCCATTGCACTTCCACCGCAAGGCGCTGTTGGCCATGCTGACTGACGATGAGCTACAT 2520
150 A H C T S T G K A L L A M L T D D E L H 169
2521 CTGCTTTATCCCGACGAGCGGCTCATCCAGCTCACCCCGCATTCGCTAGGCACCCGGACC 2580
170 L L Y P D E R L I Q L T P H S L G T R T 189
2581 GAGCTTCTCGACGCGGTGCGCACGATCCGTTCCCGGGGTTTCGCGACGAGCAAGGAGGAG 2640
190 E L L D A V R T I R S R G F A T S K E E 209
2641 AGCGAGGAGGGTGTGCGCTCCCTCGCCGTCGCGCTGGCCTCGACCCGCTCTCCGCGGTTG 2700
210 S E E G V A S L A V A L A S T R S P R L 229
2701 GCGGTCAACGCCTCGCTGCCGCTCAGTCGCATGAGCTCGGGCACCAGCAGGACATCCTC 2760
230 A V N A S L P L S R M S S A T E Q D I L 249
2761 ACTCGGCTGGTAGCTGCGGGCGGAGGATCGACCACCTCCTGCTCTGATCGGCGTGGTCA 2820
250 T R L V A A A E E I D H L L L * 264
2821 GTCCGAGGTGCGCTGGGCCTCAGGCGGGACTGGGCTGCCCGCCGTCGCGTCCAGCCGGT 2880
CAGGCTCCAGCCGACCCGGAGTCCGCCCTGACCCGACGGGCGGCAGGGCCAGGTCGGCCA
304 * A P S P Q G G D R D L R 293
2881 TGCCGGTCCCCTGGACGTGGCGCACGAAGGACCGCACCGCCGGGGCGATGCCTCCTTCGC 2940
ACGGCCAGGGGACCTGCACCGCGTGCTTCTGGCGTGGCGGCCCGCTACGGAGGAAGCG
292 N G T G Q V H R V F S R V A P A I G G E 273

Appendix. Fig. 1. Continued. The putative helix-turn helix (residues 42-61) and IclR family signature motifs [Thomson et al., 1999] (residues 202-223) in ORF6 product are indicated by shaded boxes.

2941 GTGACCAGTAGATCCACAGCGGGGCGCTCATCGGGGTGTGCTCGTCCGACCAGCGCTGGCC 3000
 CACTGGTCATCTAGGTGTGCCCCGCGAGTAGCCCCACACGAGCAGCTGGTCGCGACCGG

272 R S W Y I W L P A S M P T H E D V L A P 253

3001 ATCTCGCCGAGTCGTCGGCACTGAGGTTGTGTCGGCGATGACGGCCACGCCAGCCCGG 3060
 TAGAGCGGCTCAGCAGCCGTGACTCCAACAGCAGCCGCTACTGCCGGTGCGGGTGCGGGCC

252 W R A S D D A S L N D D A I V A V G L G 233

3061 CCCGGGCGAGCGCAACGAGTGTGGTGGCGTTGACGCTCTCCGACGTGATGTCGAGCCCGA 3120
 GGGCCCGCTCGCGTTGCTCACACCACCGCAACTGCGAGAGGCTGCACTACAGCTCGGGCT

232 A R A L A V L T T A N V S E S T I D L G 213

3121 AGCCCCCTGCCCGCAGCGCCGGCTCCAGCAGGCGCCGGGTGAGGCTATCGGGCGGACCCGG 3180
 TCGGGGGACGGGCGTCCGGCCGAGGTCGTCGGCGGCCACTCCGATAGCCGCCTGGGCC

212 F G G A R L A P E L L R R T L S D A S G 193

3181 TGAGCACCGGCTGGCCACGCAATTCAGCGACCGAGACGTGGCTCCGGCCCCGCCACGCGT 3240
 ACTCGTGGCCGACCGGTGCGTTAAGTCGCTGGCTCTGCACCGAGGCGGGGGCGGTGCGCA

192 T L V P Q G R L E A V S V H S R G R W A 173

3241 GGTCGTCGGAGGTCACCACCGCCAGCCGACGTCGCCGAGCTGGTGGCCCGCAAGACGCT 3300
 CCAGCAGCCTCCAGTGGTGGCGGTGCGCGTGCAGCGGCTCGACCACCGGGCGTTCGCGA

172 H D D S T V V A L R V D G L Q H G A L R 153

3301 GGTCGGCCTGCCCGAGGCCCGTCCAGAGATCGACGTCCCCGTTAAGCACCTGTTGATCG 3360
 CCAGCCGGACGGGCTCCGGGCAGTGTCTAGCTGCAGGGGCAATTCGTGGACAAGCTAGC

152 Q D A Q G L G T V L D V D G N L V Q E I 133

3361 CGGGCAGGCCCGAGTACTCGTGGACGGCGACGTGGATAGCGGGATGCGCCGACCGGAAGG 3420
 GCCCGTCCGGGCTCATGAGCACCTGCCGCTGCACCTATCGCCCTACGCGGCTGGCCTTCC

132 A P L G S Y E H V A V H I A P H A S R F 113

3421 TGCCGATCAGGGGCGCGAGGAACCGTGGCACGTGCGGATATGCGCAGGCCACCGTGACGA 3480
 ACGGCTAGTCCCCGCGCTCCTTGGCACCGTGCACGCCTATACGCGTCCGGTGGCACTGCT

112 T G I L P A L F R P V H P Y A C A V T V 93

3481 CGCCCGCTCGACCGCGAGCGAGTTCGTGGGCAAACCTGGGAGACCGACTCCGCTCGATTCA 3540
 GCGGGCGAGCTGGCGCTCGCTCAAGCACCCGTTTGACCCTCTGGCTGAGGCGAGCTAAGT

92 V G A R G R A L E H A F Q S V S E A R N 73

3541 GCAGGTAGTCGGCGCGGGGCGAGCAATTGATTCGCCGACCGGTGAGCTGTACTCCGCGGC 3600
 CGTCCATCAGCCGCGCCCCGTCGTTAACTAAGGGCCGTGGCCACTCGACATGAGGCGCCG

72 L L Y D A R P L L Q N G A G T L Q V G R 53

3601 CGGACCTCTCGAAGAGGGGCGATGCCAGCTCCGCCTCCAGCGCCTTGACCGAACGCCACA 3660
 GCCTGGAGAGCTTCTCCCGCTACGGGTGCGAGGCGGAGGTCGCGGAACTGGCTTGCGGTGT

52 G S R E F L A I G L **E A E L A K V S R H** 33

3661 GCGAGGGCTGCGGTACGTCGAGGGTGGCGGGCGGCGGAGGAACGACCCCTCGCGGGCCA 3720
 CGCTCCCCGACGCCATGCAGCTCCCACCGCCGCGCGCTCCTTGCTGGGGAGCGCCCGGT

32 **L S P Q P V D L T A A A R L F** S G E R A 13

3721 CCGCATGGAGATACCTGAGTTGACGCAGCTCCACGCACGGCTCCTAGTTACACCAGTGAA 3780
 GCGGTACCTCTATGGACTCAACTGCGTCGAGGTGCGTGCCGAGGATCAATGTGGTCACTT

12 V A H L Y R L Q R L E M (ORF5) 1

3781 TAGCTCGATCCCAGAATAGACCGTAGACAGCGGTGATGAGGGTCATACGCTCCTACGTG 3840

3841 TACTGCAAGGCAGAAAGCCGTCCTACCGTGAGGAGAATTGGATGTCGGAGTCTGGTGGAG 3900
 1 (ORF4) M S E S G G G 7

3901 GAACCGTGGCCACTGCGCGGCAGCGTCAGCTGGTCGAGCGGGCACTCGGTGAGTGGCAGG 3960
 8 T V A T A R Q R Q L V E R A L G E W Q G 27

3961 GCGAGGTGGCGGGTCGAGTCATCGTTGTCACCGGTGGAGCCCGGGCATCGGCCGTTCCC 4020
 28 E V A G R V I V V T G G A R G I G R S L 47

4021 TCTGCGAGGGACTGTTGCGCGCCGGCGCCAAGGTCGTTGACGCGGACCTGACGTGGGACG 4080
 48 C E G L L R A G A K V V A A D L T W D D 67

Appendix. Fig. 1. Continued. The putative helix-turn helix motif [Schell 1993] (residues 18-42) in ORF5 product is indicated by shaded boxes.

4081 ACGCTGACGACTTCCGCAAGCAGCTGGAGTCGGACGGCAGCGGCATGGCCGTCGACATGG 4140
68 A D D F R K Q L E S D G S G M A V D M D 87
4141 ACATCACCGACGACGACGCCCTCGACGCCGCCGAGACGCTGTAATCGACAGGTTCCGGGA 4200
88 I T D D D A L D A A R D A V I D R F G T 107
4201 CCGTCGACGTCCCTAGTGAACAACCGTCTCTGGTCTCCGAAACCCTCTTCCCGCCACAG 4260
108 V D V L V N N A S L V S E T L F P P T G 127
4261 GCCACCGGAACACCCTTGACACGACCGACCGCGACTGGGAGGTGATGTTCCGGCGTCAACG 4320
128 H R N T L D T T D R D W E V M F G V N V 147
4321 TGTTCCGGCACGCTGAAGGCCATCCGCCGGTTCATCGAGCCGATGCGGGCCAGCAGCGGG 4380
148 F G T L K A I R R F I E P M R A Q Q R G 167
4381 GCAGCATCGTCAACGTGCTCAGCAGTGGCGTCTCGCCGTCGCCGCCGGCGGGCGGCTACC 4440
168 S I V N V V S S G V L A V A A G G G Y H 187
4441 ACGGGCTGCGCCCGTGGACGGTGCAGATGCCTTACCAGGCCACCAAGGCCCGCGTGATGG 4500
188 G L R P W T V E M P Y Q A T K A A V M A 207
4501 CGCTGACCTTCTACCTCGCCGAGGAGGTGCGCGGCGACGGGGTCCGGTTAACGCGATCA 4560
208 L T F Y L A E E V R G D G V A V N A I M 227
4561 TGCCGGGGCACACCCGAGCGTCTGGTTCGACGCCACCGCCCGGGCCTTCAACGAGCAGG 4620
228 P G H T R A S W F D A T A R A F N E Q G 247
4621 GCATCGCCTACTTCATGCGCCCGCGATTCCCGAGCACCTGCTGCCGATCTCGCTCTTCC 4680
248 I A Y F M R P A I P E H L L P I S L F L 267
4681 TCGCAGCCCAGGACAGCGCGGGGGCGTCCGGGGCCTCTACTACGTGCCGGAGTGGAACT 4740
268 A A Q D S A G A S G R L Y Y V P E W N Y 287
4741 ACGACCACGGCTACGGGGACTACGCCGCGTGGCAGGACCATGAGCTGCCTCCGGACATGG 4800
288 D H G Y G D Y A A W Q D H E L P P D M E 307
4801 AGGAGATCTACAGCCGGCTGGAAGCCGCGACCCCGTCATACGAGCGGGCCGGCGTGGCCC 4860
308 E I Y S R L E A A T P S Y E R A G V A H 327
4861 ACCTCCCTTCGACGCGCAGGGCGCCCTGTACGCCGGGCATGGCTAACCTCGGGGCC 4920
328 L P F D A Q G A L Y A A G M A N L G A Q 347
4921 AGAACAGTTGGACCAGCAATGACAGCGCTCAGTGATACCGCAGACGCCAGACGATCCGC 4980
348 N S W T S N D S A Q * 357
4981 TTACAGGTGGGCCGTTCCGGTCCGCCCTCCCATCTCACCTCACGCAACCCGTCAGACCCC 5040

5041 GAGAGGACCCGATCATGACCAGCATTAGCGAACGCCCCGTTGACGTGGCCTCCGCAGGAC 5100
1 (dbfA1) M T S I S E R P V D V A S A G L 16
5101 TGCACGACCTGGTGAAGCCCGACGAGGGCGCCGTCAGCCGAACGGTCTTCGTCGACGAGG 5160
17 H D L V K P D E G A V S R T V F V D E A 36
5161 CAATCTACCGCAAGGAATTGGACCGGGTATTCACCAAGACGTGGTTGTTTCATCGGCCACG 5220
37 I Y R K E L D R V F T K T W L F I G H E 56
5221 AGTCCCAGTTGAGTGAGCCGGGCGACTACCTGACGAACTTCATGGGCGAGGACCCTGTGA 5280
57 S Q L S E P G D Y L T N F M G E D P V I 76
5281 TCGCCACTCGCGGTGCGGACGGAGTGATCCGGGTGATGCTCAACTCCTGCGCCCACCGGG 5340
77 A T R G A D G V I R V M L N S C A H R G 96
5341 GCATGGCCGTTGTGTAGCACCGACGCCGGGTCCCTCGAAGTTCTTCCGCTGCCCTACCACG 5400
97 M A V C S T D A G S S K F F R C P Y H G 116
5401 GCTGGACCTACAGCAACAACGGCGATCTGATCGGTGTACCGCGCGCAGATACCGTCTACC 5460
117 W T Y S N N G D L I G V P R A D T V Y H 136
5461 ATGGCGAGCTGGACAAGTCGCGGCTAGGTCTGAAGCCGTTCCGCGGGTGGAGAACTACA 5520
137 G E L D K S R L G L K A V P R V E N Y K 156
5521 AGGGCTTCATCTTCGCCAATTGGGACGAGGACGCCATCCCGCTGGTGGACTACCTCGGCG 5580
157 G F I F A N W D E D A I P L V D Y L G A 176
5581 CTGACCAGCTCTGGTATCTGGACCTGGCCTTCGAGGCGCCGCTCGGCGGGCTCGAGGTGA 5640
177 D Q L W Y L D L A F E A P L G G L E V I 196
5641 TCGGCCCCACGATGAAGTTCCGGATCAAGGCCAACTGGAAGCTGGCGGGCGGAGAACTTCG 5700
197 G P T M K F R I K A N W K L A A E N F A 216

Appendix. Fig. 1. Continued.

5701 CCGGCGACGACTACCACGTGCTCTACACACATGGGTCGGCCTTCCAGATCGGCTTCCTCC 5760
217 G D D Y H V L Y T H G S A F Q I G F L P 236
5761 CCGATTACGACACGCTCGGCGACTACATCGCACATTTTCGACCACGGCCACGGGATGGGCG 5820
237 D Y D T L G D Y I A H F D H G H G M G D 256
5821 ACATCGGCAAGCCCGGCCGGGCTATCAGAACGACGTCGGGATGGCTCAGTTCCTCGGGC 5880
257 I G K P G R A Y Q N D V G M A Q F L G P 276
5881 CGGAGGCGATCGAGTACGTCAACGCCGTGCACGAGCGGCTCAAGGCCCGGGTCTCCCCGC 5940
277 E A I E Y V N A V H E R L K A R V S P L 296
5941 TGCAGGCGGAGATGCACGGGCTCGGTTCAGGGCAATATCTTCCCGAACCTGTCATGGATCA 6000
297 Q A E M H G L G Q G N I F P N L S W I K 316
6001 AGTTCGGCGTCTTCCACGTCTTCCGGGCTCTTCCAATGGCACCCGAGGGGACCGGGTGAGA 6060
317 F G V F H V F G L F Q W H P R G P G E I 336
6061 TCGAGGTCTGGCAGACGGCGCTCTTCGACCGCGACGCGCCGAGTTCGGTCAAGGACGTCG 6120
337 E V W Q T A L F D R D A P Q S V K D V A 356
6121 CCCGCACCCAGATGTCCAGGAGAATGCCGCGGCCGGGATCTTCGGCCAGGACGATGGCG 6180
357 R T Q M S Q E N A A A G I F G Q D D G E 376
6181 AGAACTTCGAGCAAATCACCGAGTCCGCCCGGGGTGGTCTCCAGACCCGGGATTTCA 6240
377 N F E Q I T E S A R G V V S Q T R D F N 396
6241 ACTACGCGATGGGCCTGGGGCACGAGGGCGAGATCCACGAGGAGGGATACCCCGGCCATT 6300
397 Y A M G L G H E G E I H E E G Y P G H L 416
6301 TGGGGCCCCACTACTCGGAGCAGAACCACCGCAACTTCTATCGCTACTGGCTCGAACTCA 6360
417 G P H Y S E Q N H R N F Y R Y W L E L M 436
6361 TGACCACCCCGGAGAGCAGAAATGACCACCTTCCCTCACCGATCTGCGCCGCGACGTCGA 6420
437 T T P G E Q K * 443
1 (dbfa2) M T T S L T D L R R D V E 13
6421 GGAATTCCTCTACAGCGAGGCCAAGATGCTCGACGAGCAACGCTACGACGAGTGGCTCGA 6480
14 D F L Y S E A K M L D E Q R Y D E W L D 33
6481 CCTGTTCACCGAGGATGTCCATTACTGGATGCCGATCACGGAGACCCGTGAGGTGCGGCA 6540
34 L F T E D V H Y W M P I T E T R E V R Q 53
6541 GCACCGGGACCACGTCCCCGGCGAGTGGTTCGCTCATGGAGGAGGACGCCCGCTTCTCGC 6600
54 H R D H V P G E W S L M E E D A R F L A 73
6601 CAAGCGGATGGAGCGCTTGGCCGGTGGCCTTGGCCACTCGGAGCAGCCGCGGTGCGGGAC 6660
74 K R M E R L A G G L A H S E Q P R S R T 93
6661 GCGTCGGTTCATCAGCAACGTCTGGTCAACCCCGGGCCGGACGGCGACCTGGTAGCGGA 6720
94 R R F I S N V L V T P G P D G D L V A E 113
6721 GTGCAACTTCATCGTCTTCCAGTCCCGGGCCAACTCCGAGCAGTTCCTTCGTGCGCTC 6780
114 C N F I V F Q S R R A N S E Q F F V G S 133
6781 CCGTCGCGACCGGATCGTGACCTCCGGTGAGAAGTGGAAAGATCGCCGAACGGACCGTGCT 6840
134 R R D R I V T S G E N W K I A E R T V L 153
6841 CCTCGACCACCGGGTGTGCCCCGTGCCATCTCCATCTTCTTCTGAGGCCAGGCGGCTG 6900
154 L D H R V L P R A I S I F F * 167
6901 ACCTCGACGGCAGTCCGGGACGTCGGGGCTGATGGCGCACCCAGAGCCGAGTTCTTGGCTGC 6960

6961 GCACCGGCTCTGCCGACGGCGAGCTCGTAGTCAGCAACGAGTTCGCCCGGGTGCGGGTAC 7020

7021 GCGCCGACCGAGACGGCAACGACCTTCGCCTGGCCATCCGTAGCCTGCGCACCGGCCGCG 7080

7081 AGGTCTTCCTTGACGCGCTGCAGCTGGAGAGCCTCACTTGGCTCGACGAGAGGGCCTACA 7140

7141 CGACATTGCTCACCGAGCCGTTCCGACCGGAGTGACCGACCATCCCGCACCGGCCGACG 7200

7201 GCCGAACAGAGAGGAGCCTCAACGTGACGAGCGCTGAAGGGGGAAACGGCCAGCCCCAGG 7260
1 (ORF3) M T S A E G G T A Q P Q G 13

Appendix. Fig. 1. Continued.

| | | |
|------|--|------|
| 1 | GATATCGAAGCTCGCACCGTTGGTGGCCTTGACCGCGCCAAAGGTTTTTTCCAGGGCCTG | 60 |
| 61 | GACCGACAGTTGCGGGTTCATGGGCGTACCTCGTGTTGGGCGATTGCGGTTGGCGCGGCA | 120 |
| 121 | TGAGCTTGCGCCGCCATGGCCAGGGGCTGAGCAGCAAGCCGTAGAGGCCTTTTTTGCCGA | 180 |
| 181 | ACAGCACCACCACCAGCAGCAACGGGCGAAGATCAGCAGCCAGTGCTCGGTCCACAGGC | 240 |
| | TGTCGTGGTGGTGGTTCGTCGTTGCCCGGCTTCTAGTCGTGGTCCAGAGCCAGGTGTCCG | |
| 571 | * C G T S P G C A | 564 |
| 241 | TGAGCACTGTTCCAGGCCAGGTACACGGCGGCGCCGAGGACCGGCCCCAGCAGGGTGCC | 300 |
| | ACTCGTGACAAGGTCCGGGTCCATGTGCCCGCGGCTCCTGGCCGGGGTTCGTCACAGG | |
| 563 | S C Q E L G L Y V A A G L V P G L L T G | 544 |
| 301 | GACGCCCGGAGGATGACCATCGCCATCAGTTCGCCGGACTTGTGCCAGGCGCCATGTC | 360 |
| | CTGCCGGGCTCCTACTGGTAGCGGTAGTCAAGCGGCTGAACACGGTCCGCGGGTACAG | |
| 543 | V G G L I V M A M L E G S K H W A G M D | 524 |
| 361 | CGGGCTGACGAACATCGCGTAGTTGGCCCCAGAGAATGCCCGCCAGGCCCCCGACACC | 420 |
| | GCCCGACTGCTTGTAGCGCATCAACCGGGTCTCTTACGGGCGGTCCGGGGGGGCTGTGG | |
| 523 | P S V F M A Y N A W L I G A L G G G V G | 504 |
| 421 | GGCGATGACAAACGCCCCGAGGCGATAGCGCAGCGGGTGCAGGCCAGGCTGATGCTGCG | 480 |
| | CCGCTACTGTTTGGGGGCTCCGCTATCGCGTCGCCACGTCCGGGTCCGACTACGACGC | |
| 503 | A I V F A G L R Y R L P H L G L S I S R | 484 |
| 481 | CCGTTCGCTTTGCTTGAGGCCGCGCAGCACGTAGCCGAAACGCGAATTGACCAGGCGCCG | 540 |
| | GGCAAGCGAAACGAACCTCCGGCGCGTTCGTGCATCGGCTTGGCGTTAACTGGTCCGCGGC | |
| 483 | R E S Q K L G R L V Y G F R S N V L R R | 464 |
| 541 | GCAGAACAGCAGCCAAGCCACCAACAGGCCAGGCACAGGTAGTAGAACTGCACCGGGTT | 600 |
| | CGTCTTGTGTCGTTCCGGTGGTGTCCGGGTCCGTGTCATCATCTTGACGTGGCCCAA | |
| 463 | C F L L W A V L L G L C L Y Y F Q V P N | 444 |
| 601 | ATCCAGGTCGATGCCCGGCGAGGGTGTGCGCTCGCTGAGCAAATGCCGTCGTCGCCGCC | 660 |
| | TAGGTCAGCTACGGGCCGTCCCACAACCGGAGCGACTCGTTTTACGGCAGCAGCGGGCGG | |
| 443 | D L D I G P L T N R E S L L I G D D G G | 424 |
| 661 | ATACAGCGACAGCGAACTGAGGACGAAATAGAGCATCTGGCCGAACGCCAGGGTAATCAT | 720 |
| | TATGTCGCTGTCGCTTACTCCTGCTTTATCTCGTAGACCGGCTTGGCGTCCCATTAGTA | |
| 423 | Y L S L S S L V F Y L M Q G F A L T I M | 404 |
| 721 | GATGAATTGCACCCCGCTGGTACGCAGCGACAGGTAGCCACCACCACGGCAAACAGCGC | 780 |
| | CTACTTAACGTGGGGCGACCATGCGTTCGCTGTCATCGGGTGGTGGTGGCCTTTGTCGCG | |
| 403 | I F Q V G S T R L S L Y G V V V A F L A | 384 |
| 781 | ACACACCAGCAGCGACAGCGGCCAAACGAGCAGCGCGCTGTTGGAACCGTCCCAGCCCCA | 840 |
| | TGTGTGGTTCGCTGTCGCGGTTTGGCTCGTCGCGGACAACCTTGGCAGGGTCCGGGT | |
| 383 | C V L L S L P W V L L A S N S G D W G W | 364 |
| 841 | CAGCGGCAAGGCCTGGGAGGTGTGAAAACCGATCACCCACCACGTAGCCGCCAGGCC | 900 |
| | GTCGCCGTTCCGGACCTCCACACTTTTGGCTAGTGGGGTGGTGCATCGGCGGGTCCGG | |
| 363 | L P L A Q S T H F G I V G V V Y G G L G | 344 |
| 901 | GAAGAAGGCCGCGTGGCCGAAGCTAATCATCGCGCCATAGCCGATCAGCAGGTCCAGGCT | 960 |
| | CTTCTTCCGGCGCACCGGCTTCGATTAGTAGCGGGTATCGGCTAGTCGTCCAGGTCCGA | |
| 343 | F F A A H G F S I M A G Y G I L L D L S | 324 |
| 961 | GGCGGCGCCATGCCGTAGATCGCCAGGCGCGTAAACAGGCTGACCAGAAAGGGTTCGTG | 1020 |
| | CCGCCCGGTACGGCATCTAGCGGTCCGCGCATTTGTCCGACTGGTCTTTCCCAAGCAC | |
| 323 | A A A M G Y I A L R T F L S V L F P E H | 304 |
| 1021 | CAGCAGCGAGGCCAGCAGCGGCGAGGAGGGCGAACGTCGCCAGGCAAAGCAGGTGATCAG | 1080 |
| | GTCGTCGCTCCGGTTCGTCGCGGCTCCTCCCGCTTGCAGCGGTCCGTTTCGTCCAGCTAGTC | |
| 303 | L L S A L L P L L A F T A L C L L D I L | 284 |

Appendix. Fig. 2. The nucleotide and deduced amino acid sequences of the 44,266-bp *EcoRV-EcoRI* DNA region (accession number AB047548). The sense DNA strand is shown. The putative SD sequences are indicated by double lines under the sequence. Astarisks indicate stop codons.

1081 GACTCGACGGGAACATGGGGTTCTCCTCAAGGCACGGGCGGGCAGCAGGCCGCGGGGGCG 1140
CTGAGCTGCCCTTGTACCCCAAGAGGAGTTCCGTGCCCCCGTCGTCCGGCGCCCCCGC
283 V R R S C P T R R L A R A P L L G R P R 264
1141 GACCAGCAACACCACGGCCATGACGATGTAGGCGCTGGCGGAAATCAGCCCCCGCTCAG 1200
CTGGTTCGTTGTGGTGCCTGACTGCTACATCCGCGACCGCCTTTAGTCGGGGCGGCGAGTC
263 V L L V V A M V I Y A S A S I L G G S L 244
1201 GGTGCCGCTGACTTCGGCGGGCAACAGTTGGTCCAGCAATTGCGGGATGTAGGCACGCCC 1260
CCACGGCGACTGAAGCCGCCCCGTTGTCAACCAGGTCGTTAACGCCCTACATCCGTGCGGG
243 T G S V E A P L L Q D L L Q P I Y A R G 224
1261 CAGGCCATCGACCATGCCGATCAGCAAGGCCCGGCCAGGGCGCCGCGCACCGAGCCGAC 1320
GTCCGGTAGCTGGTACGGCTAGTCGTTCCGGGGCCGGTCCCGCGGGCGCTGGCTCGGCTG
223 L G D V M G I L L A G A L A G R V S G V 204
1321 ACCGCCGACCACGATGACCACGAAGGTGGTAATCAAGACCTTTTCACCCATGCCGATTTTC 1380
TGGCGGCTGGTGTACTGGTGTCCACCATTAGTTCTGGAAAAGTGGGTACGGCTAAAG
203 G G V V I V V F T T I L V K E G M G I E 184
1381 CACCGACAGCAAGGGTGCGGCCATGAACCCGCGGAGGCCGCAAAGCACGGAGCCGAACAC 1440
GTGGCTGTCGTTCCACGCCCCGTTACTGGGCGCTCCGGCGTTTCGTGCCTCGGCTTGTG
183 V S L L P A A M F G A L G C L V S G F V 164
1441 GAACACCAGGGTGTACAGGCGTGCGATGTTACCCCCAGGGCGCCGACCATTTTCGTGGTC 1500
CTTGTGGTCCCACATGTCCGCACGCTACAAGTGGGGGTCCCAGGGCTGGTAAAGCACCAG
163 F V L T Y L R A I N V G L A G V M E H D 144
1501 GTCGGCCCCGCGCGGATCAGCATGCCAGGCGCGTGTGGTTTTATCAGCAGCCACATGGC 1560
CAGCCGGGGCCGCGCCTAGTCGTACGGTCCGCGCACACCAAATAGTCGTCCGGTGTACCG
143 D A G A R I L M G L R T H N I L L W M A 124
1561 GATCGCCACCACGGCCCCGGCGGCGATGAACAGCAAGCGGCTGACCGGGTACATCAACC 1620
CTAGCGGTGGTGCCTGGGGCCGCGCTACTTGTGCTCCGCGACTGGCCCATGTAGTTGGG
123 I A V V A G A A I F L L R S V P Y M L G 104
1621 CGGCAGCACTTCGACGAAGCTCGAATACCACTCGGGCGTGGGCATGGCCGGTGGGCTGCG 1680
GCCGTGCTGAAGCTGCTTCGAGCTTATGGTGAGCCCGCACCCGTACCGGCCACCCGACGC
103 P L V E V F S S Y W E P T P M A P P S R 84
1681 GCCGAACAGCACGGTGATTAATCGTTGGTGAAGAACCACCAGGGTGGGCCAGTA 1740
CGGCTTGTGCTGCCACTAATTGAGCAACCACTTCTTGTGGTGGCGGTCCCACCCGGTCAT
83 G F L V T I L E N T F F V V A L T P W Y 64
1741 CCTGGTCGAGGTGATCGCTGTTGTAGAGGCGGCGCATGATGCTGGTTTCGATAAGCAGCG 1800
GGACCAGCTCCACTAGCGACAACATCTCCGCCGCTACTACGACCAAAGCTATTCGTGCG
63 R T S T I A T T S A A C S A P K S L C R 44
1801 CGTATAGGGCCGAGCGGATGATCCGCTGCCAACCAGGCGAGCCAGAACGAGCCGCTCAGG 1860
GCATATCCCGGCTCGCCTACTAGGCGACGGTTGGGCGGCTCGGCTTGTCTCGGCGAGTCC
43 T Y P R A S S G S G V R R A L V L R E P 24
1861 ACGCGGTGTAGCCGGCAGAACGCGCCGACCATGTAGAAGGCGCCGTGGGCCAGGTTGAT 1920
TGCGCCACATCCGGCCGTCTTGCAGGGCTGGTACATCTTCCGCGGCACCCGGTCCAACCTA
23 R P T P R C F A G V M Y F A G H A L N I 4
1921 GACGCCCATGATGCCGAAGATCAGCGTCAGGCCGGCAGGCCAGCAAAAACAGAAGGGCG 1980
CTGCGGGTACTACGGCTTCTAGTCGAGTCCGGCCGTCGGGTCGTTTTTTGTCTTCCCGC
3 V G M (ORF31) 1
1981 CTTTATTGCAGGCCATTGAGGATTTGTTTCGAGTAGCAGCATGCGGGGAATCCTACGCAAG 2040
2041 GTGCCTGGCGGGTAAGGCCGCCAGGCACGGGACCGGAAGGTCGTTACAGCGCGCACTCG 2100
CACGGACCGCCATTCCGGCGGTCCGTGCCCTTCCAGCGAATGTGCGCGGTGAGC
386 * L A C E 383
2101 GCGGCGAAGCTGTCCACTTGCCTTGGCGATGGTCTGCACCGGCACCTGGACCAACTTG 2160
CGCCGCTTCGACAGGTGAACGCGGAACCGCTACCAGACGTGGCCGTGGACCTGGTTGAAC
382 A A F S D V Q A K A I T Q V P V Q V L K 363

Appendix. Fig. 2. Continued.

2161 CCGTTGGCATCGGCCTCTACCCGAGCAAATACCAGTCGATGACCGCGTGTGGTTTTTTG 2220
GGCAACCGTAGCCGGAGATGGGCGTCGTTTATGGTCAGCTACTGGCGCACGACCAAAAAC
362 G N A D A E V R L L Y W D I V A H Q N K 343
2221 CCGAAGGCGAAGTCGCCACGGATCGAGTCGAACGTGACGTTGCGCAGAGCACTGCGCAAG 2280
GGCTTCCGCTTCAGCGGTGCCTAGCTCAGCTTGCACTGCAACGCGTCTCGTGACGCGTTC
342 G F A F D G R I S D F T V N R L A S R L 323
2281 GCCTCGGGGTCTTTGAGGTTGCCGTCCACGGTTTTTCAGGGCGGCGCCGATCAGGCGGGCG 2340
CGGAGCCCCAGAACTCCAACGGCAGGTGCCAAAAGTCCCGCCGCGGCTAGTCCGCCCCG
322 A E P D K L N G D V T K L A A G I L R A 303
2341 GTGTCGTAGGTTTGGCGGGCGTACACGGTGGGCACGCGCTTGTACTTATCGTTGAAGGCC 2400
CACAGCATCCAAACGCGCCGCATGTGCCACCCGTGCGCGAACATGAATAGCAACTCCGG
302 T D Y T Q A A Y V T P V R K Y K D N F A 283
2401 TTGACGAAGGCTGGTTGGCCGGGTTGGTGGAGCCGGGTTCCACAGGGTGACGATGTTTC 2460
AACTGCTTCCGGACCAACCGGCCCAACCACCTCGGCCCAAGGTGTCCCCTGCTACAAG
282 K V F A Q N A P N T S G P N W L T V I N 263
2461 ATGCCCTTGGCCACATCACCGGTGGCGGCGAGCATGCGTTCGTCCATGGAGAACACCGGC 2520
TACGGGAACCGGTGTAGTGGCCACCGCGCTCGTACGCAAGCAGGTACCTCTTGTGGCCG
262 M G K A V D G T A A L M R E D M S F V P 243
2521 ACCACCATCGGGATGCTCTTGTTCAGGCCCGAGCTGCCGTACTGCTTGGCGAAGTTGATG 2580
TGGTGGTAGCCCTACGAGAACAAGTCCGGGCTCGACGGCATGACGAACCGCTTCAACTAC
242 V V M P I S K N L G S S G Y Q K A F N I 223
2581 CCGGCACCACCAGGGTGGAACTGGAAAATCGCGTCAGGCGCCAAGGAGCGCACCCGGGCC 2640
GGCCGTGGTGGTCCACCTTGACCTTTTAGCGCAGTCCGCGGTTCCCTCGCGTGGGCCGG
222 G A G G P H F Q F I A D P A L S R V R A 203
2641 AGCTCCACCGAGAAATCCAGTTGGTTGAGCTTGGTGTAGATCTCGGTGACTTCACCCTTG 2700
TCGAGGTGGCTCTTTAGGTCAACCAACTCGAACCACATCTAGAGCCACTGAAGTGGGAAC
202 L E V S F D L Q N L K T Y I E T V E G K 183
2701 AAGGTGCGCTTGAAGCCTTCCAGGGCATCGCGGCCACCCTGGTAGTTGGGGGCGAGGATG 2760
TTCCACGCGAACTTCGGAAGGTCCCGTAGCGCCGGTGGGACCATCAACCCCGCTCCTAC
182 F T R K F G E L A D R G G Q Y N P A L I 163
2761 ACCATTTTCTTGTAGCCCAGGTCGTTGGCGGCGACGCCGGCCATTTTCGTGGATGGTGTG 2820
TGGTAAAAGAACATCGGGTCCAGCAACCGCCGCTGCGGCCGGTAAAGCACCTACCACAGC
162 V M K K Y G L D N A A V G A M E H I T D 143
2821 TTCTGGTACGAGGCCACGAAGTAGTGGGGCTTGCACTGTTTGGCCGGCCAGGCTGGAGGGG 2880
AAGACCATGCTCCGGTGTTCATCACCCCGAACGTGACAAAACGGCCGGTCCGACCTCCCC
142 N Q Y S A V F Y H P K C Q K G A L S S P 123
2881 CCGGTGTTGTTGCTGATGTAGAAGCCGTCCTTGGTTCGCGGTGTTGACCACCGCCGCGAGC 2940
GGCCACAACAACGACTACATCTTCGGCAGGAACCAGCGCCACAACCTGGTGGCGGCGGTG
122 G T N N S I Y F G D K T A T N V V A A L 103
2941 ACGTTGGAGAAGATCACCCCGGTGTACAGCGAGATGCCGTCCAGGCTCATGCGGTGATG 3000
TGCAACCTCTTCTAGTGGGGCCACATGTCGCTCTACGGCAGGTCCGAGTACGCCAGCTAC
102 V N S F I V G T Y L S I G D L S M R D I 83
3001 ATCTGCTTGGCGCGCGCCAGTTGCAGGCCATCGTCCTCCACCAGCAGGCTCACCGGTACG 3060
TAGACGAACCGCGCGCGGTCAACGTCCGGTAGCAGGAGGTGGTTCGTCCGAGTGGCCATGC
82 I Q K A R A L Q L G D D E V L L S V P V 63
3061 CCGCCGAGCTTGGCGCCCTCCTGGTCGATGGCCAGTTGAAAGCCATCGCGCGAGTCTTCC 3120
GGCGGCTCGAACGGCGGGGAGGACCAGCTACCGGTCAACTTTCGGTAGCGCGCTCAGAAGG
62 G G L K G G E Q D I A L Q F G D R S D E 43
3121 CCCAGGTAACCGGCCGGGGTGGAGAGGGTGGTGATGAAGCCGATGCGCACCGGGTCTTGG 3180
GGGTCCATTGGCCGGCCCCACCTCTCCACCCTACTTTCGGCTACGCGTGGCCCCAGAACC
42 G L Y G A P T S L T T I F G I R V P D Q 23
3181 GCCCTGGGCCGCAAGGCTGGCCAGGCAACTGGCGCCAGGAGTGTACCCACGAGTTGTTTT 3240
CGGGACCCGGGCTCCGACCGGTCCGTTGACCGGGTCCCTCACATGGGTGCTCAACAAAA
22 A R P R L A P W A V P A L L T G V L Q K 3

Appendix. Fig. 2. Continued

3241 TTCATAGTGCCGCTTTAGGGATGGTCTGAAGTAGCCATGTATTTCCCCGGCATAACCACC 3300
AAGTATCACGGCGGAAATCCCTACCAGACTTCATCGGTACATAAAGGGGCCGTATGGTGG
2 K M (ORF30) 1
3301 TAGCGCAACTTTTAAACTCGCCAAATGATCAAAAAACAATCAATTCGACGAGTATTTCTT 3360
3361 ACGTTTTTCGCCACCCGGACCCACTCCGGTGGCCATTTCCCGCATTACAGGGCACCTCT 3420
TGCAAAAGCGGTGGGCTGGGGTGAGGCCACCGGTAAAGGGCGTAATGTCCGCGTGGAGA
326 * L R V E 323
3421 CCTGCATTTCGTCAGTAAATGTCGGCGCGCCATCCACAGGTCGACAGAGCGAACAGTGTG 3480
GGACGTAAGCAGTCATTTACAGCCGCGCGGTAGGTGTCCAAGCTGTCTCGCTTGTACAG
322 G A N T L L H R R A M W L N S L A F L T 303
3481 ACCAGTTGCGCCGTGTTTTTGGCCAAGCCGCGAAAGCGCACCTTGGTGTAAACCGAACTGG 3540
TGGTCAACGCGGCACAAAACCGGTTTCGGCGCTTTTCGCGTGGAACCACATTGGCTTGACC
302 V L Q A T N K A L G R F R V K T Y G F Q 283
3541 CGCTTGATTACCCGGAACGAATGCTCGACCTTGGCGCGCACCTGAGCCTTGGCCTTTTCA 3600
GCGAACTAATGGGCCTTGCTTACGAGCTGGAACCGCGCGTGGACTCGGAACCGGAAAAGT
282 R K I V R F S H E V K A R V Q A K A K E 263
3601 ATCTTGCCTTGGCTTTGTACAGCACGCTGCGCTTATCGAGTTTTTTGTAGGTGCTGCGG 3660
TAGAACGCGAACCGAAACATGTCGTGCGACGCGAATAGCTCAAAAACATCCACGACGCC
262 I K R K A K Y L V S R K D L K K Y T S R 243
3661 CGTGCCGCCACCTGCCAGATAACTTCACGTCCTTCATGCTCGGGCCGCTTTTCGACACCG 3720
GCACGGCGGTGGACGGTCTATTGAAGTGCAGGAAGTACGAGCCCGGCGAAAAGCTGTGGC
242 R A A V Q W I V E R G E H E P R K E V G 223
3721 GTGTAGCCTGCATCGGCGCAGACCACGTTTTTCGTCGCCATGCAGCAATTTGTCCACCTGG 3780
CACATCGGACGTAGCCGCGTCTGGTGCAAAAGCAGCGGTACGTCGTTAAACAGGTGGACC
222 T Y G A D A C V V N E D G H L L K D V Q 203
3781 GTGACATCCGCCACATTGGCCGCCGTGCCACTACGCTGTGCACCAGACCCGATTCGTCA 3840
CACTGTAGGCGGTGTAACCGGCGGCACGGGTGATGCGACACGTGGTCTGGGCTAAGCAGT
202 T V D A V N A A T G V V S H V L G S E D 183
3841 TCGGCGCCGATGTGGGCCTTCATGCCAAAATAATACTGGTTTCCCTTCTTGGTCTGGTGC 3900
AGCCGCGGCTACACCCGGAAGTACGGTTTTTATTATGACCAAAGGGAAGAACCAGACCAG
182 D A G I H A K M G F Y Y Q N G K K T Q H 163
3901 ATTTCCGGGTGCGCTTGCCTGCTTGTCTTGGTTCGAACTGGGCGCATTGATCAGCGTT 3960
TAAAGGCCAGCGGAACGGCAGGAACAAGAACCAGCTTGACCCGCGTAAGTACGCAA
162 M E P D R K G D K N K T S S P A N I L T 143
3961 GCATCGACGATGGTGCCTTGGCGCAGCGACAGGCCGCGGTGCGCCAGATAGCCATTGATG 4020
CGTAGCTGCTACCACGGGACCGCGTCCGCTGTCCGGCGCCAGCGGGTCTATCGGTAAGTAC
142 A D V I T G Q R L S L G R D G L Y G N I 123
4021 ACAGCGAGGATGCCGGCCGCGCAGCTCGTGTTCCTCCAGCAAGCGGCGGAAGTTGAGAATG 4080
TGTCGCTCCTACGGCCGGCGGTGCGAGCACAAAGAGGTCGTTTCGCGCCTTCAACTCTTAC
122 V A L I G A A L E H K E L L R R F N L I 103
4081 GTGGTTTCGTGGGGATGCGCTCCAGGTTTCAGCCCCGCAACTGGGCGCAGGATCGTGGTT 4140
CACCAAAGCAGCCCTACGCGAGGTCCAAGTCGGGCGGTTTGACCGCGTCCTAGCACCAA
102 T T E D P I R E L N L G A F Q R L I T T 83
4141 TCGTACAGCGCTTCCATCGCTGGATCGCTGTAGCCGAACCAGTCTGCAGCAGATGC 4200
AGCATGTGCGGAAGGAGGTAGCGACCTAGCGACATCGGCTTGGTCAAGACGTCGTCTACG
82 E Y L A E E M A P D S Y G F W N Q L L H 63
4201 ACACGCAGCATCGCCATCAACGGGTAGGCCGGACGGCCACCTTCACCTTTCGGATAATGT 4260
TGTGCGTCGTAGCGGTAGTTGCCCATCCGGCCTGCCGGTGGAAAGTGGAAAGCCTATTACA
62 V R L M A M L P Y A P R G G E G K P Y H 43
4261 GGCTCGATCAAAGCAATCAAGCCCTTCACGGCACCCCGATCCATCTCGATCAGGAAC 4320
CCGAGCTAGTTTCGTTAGTTTCGGGAAGGTGCCGTGGTGGGCTAGGTAGAGCTAGTCCTTG
42 P E I L A I L G K W P V V R D M E I L F 23

4321 AACTCCTTGC GGGTTTGCTTGC GCTTGC CAGCGTACTCGGCGT CGGCGAAGGTCATCTGC 4380
 TTGAGGAACGCCCAAACGAACGCGAACGGTCGCATGAGCCGCAGCCGCTTCCAGTAGACG
 22 L E K R T Q K R K G A Y E A D A F T M Q 3
 4381 TTCATCGGGAAACTCGGTGGGTGGGGTGC GCGGTATTTTGC CAAATCAGAAAGTCTTTTTC 4440
 AAGTAGCCCTTTGAGCCACCCACCCAGCGCCATAAAACGGTTTAGTCTTTCAGAAAAAG
 2 K M (tnpA3) 1
 4441 AGAGTTTCCTTAGTGAAGCCCGCATCGCACGGGCTGTAGTTATTAGTGGAGGGCAGTAAA 4500

4501 AAATTTCCGGGTGCCTGGAATGTTTTGTGGGGGGGCTTCCGTGTCAGGTGTTGCTCTGGGT 4560
 TTTAAAGCCACGGACCTTACAAAACACCCCGCGAAGGCACAGTCCACAACGAGACCCA
 342 * T N S Q T 338
 4561 GAATTTCTCGTAGTACAGGCGGGTCTGGTCCAGGGCCTGGGTGTCGAGCCATTGCTTGAT 4620
 CTTAAAGAGCATCATGTCCGCCAGACCAGGTCCCGGACCCACAGCTCGGTAACGAACTA
 337 F K E Y Y L R T Q D L A Q T D L W Q K I 318
 4621 CGACTCGACCATCGGCGGGCGCCCGCACAGGTACATGTCCAGGGCCTGGTTCGCGGAACTC 4680
 GCTGAGCTGGTAGCCGCCCGGGCGTGTCCATGTACAGGTCCCGGACCAGCGCCTTGAG
 317 S E V M P P P G C L Y M D L A Q D R F E 298
 4681 GGCCAGGTCGAAGTGCTCGGTGATGTAGCCGCGCTTGCCGTGCCATGTCCGGGCTGGGGT 4740
 CCGGTCCAGCTTACGAGCCACTACATCGGGCGCAACGGCACGGTACAGGCCCGACCCCA
 297 A L D F H E T I Y G R K G H W T R A P T 278
 4741 CGCTGATGAGCCGGGGTGTAAAGCGAAAGCCCGGGATGCGTTCCGGCGTAGGCCGCGATGCG 4800
 GCGACTACTCGGCCCCACATTCGCTTTCCGGGCCCTACGCAAGCCGCATCCGGCGCTACGC
 277 A S S G P H L R F G P I R E A Y A A I R 258
 4801 CTGGGTTTCGCACAGGTCCGCGGGCTTGC GCACGCCGTAGTACAGGTGCACCGGGTGGCC 4860
 GACCCAAAGCGTGTCCAGGCGCCGCAACGCGTGC GGCATCATGTCCACGTGGCCACCGG
 257 Q T E C L D A A N R V G Y Y L H V P H G 238
 4861 GCAGCCGCGCGCTCGGCGATTTCTGTCGAGCATGCCAAAAAGGCCGACAGGGGTTTGCC 4920
 CGTCGGCGGGCGGAGCCGCTAAAGCAGCTCGTACGGGTTTTTCCGGCTGTCCCCAAACGG
 237 C G G R E A I E D L M G L F A S L P K G 218
 4921 GCCGGCGACGAACACCAGCGGCTTGTGACGTGGCGCAGGTAGAAGGCCCCAGGGGCGC 4980
 CGGCCGCTGCTTGTGGTCCGCCAACAGCTGCACCGCGTCCATCTTCCGGGGGTCCCCGCG
 217 G A V F V L P K D V H R L Y F A G L P A 198
 4981 TTCCAGCAGGATTTTCGTCGCCGATCAGGCTGCGTTCCGCGGATGTAGTTGCTCATCAGCCC 5040
 AAGGTCGTCCTAAAGCAGCGGGCTAGTCCGACGCAAGCGCCTACATCAACGAGTAGTCGGG
 197 E L L I E D G I L S R E R I Y N S M L G 178
 5041 GTCGGGCAGCAAACGGATCAAGAACTGCAACTGGTTGCCGCTGTTGGGGCGGTTGGCGAA 5100
 CAGCCCGTTCGTTTGCCTAGTTCTTGACGTTGACCAACGGCGACAACCCCGCCAACCGCTT
 177 D P L L R I L F Q L Q N G S N P R N A F 158
 5101 GGAGTACGAGCGCTGGCCGTGGGTGCCGGGTATTTGCAGGCGGGCGTATTGGCCGGGCGAG 5160
 CCTCATGCTCGCGACCGGCACCCACGGCCATAAAACGTCCGCCCGCATAACCGGCCCGTCT
 157 S Y S R Q G H T G P I Q L R A Y Q G P L 138
 5161 GAAATCCAGTTGCTGGCCATCGGCGCCGGCGTCCAGGTGCAGGATCGCGGTGTTGGGCGA 5220
 CTTTAGGTCAACGACCGGTAGCCGCGGCCGAGGTCCACGTCCTAGCGCCACAACCCGCT
 137 F D L Q Q G D A G A D L H L I A T N P S 118
 5221 AACCTGCTCCACCGCGCGCACGATGCCCTTGAGCCGGCCTGGCCCTGGGGCATTGCACAG 5280
 TTGGACGAGGTGGCGCGGTGCTACGGGAACTCGGCCGGACCGGGACCCCGTAACGTGTC
 117 V Q E V A R V I G K L R G P G P A N C L 98
 5281 GCTGGAGGCGAAGTCAAGTAGAACGCGGCGTCCGAGCTGACCCGGGTCTGGCAGGTGAG 5340
 CGACCTCCGCTTACGCTTTCATCTTGC GCGCAGCCTCGACTGGGCCCAGACCGTCCACTC
 97 S S A F D F Y F A A D S S V R T Q T L 78

Appendix. Fig. 2. Continued. The conserved Cys residues (CX₄CX₂CX₃₁C) which may coordinate a [2Fe-2S] cluster in plant-type iron-sulfur proteins [Otake and Ooi 1989] are indicated by shaded boxes.

5341 CATCTTGC GTTGGCGCAGGTCTTCGGCCGACAGCGCTTCGTCGTCGACATAGTCTGGCT 5400
GTAGAACGCAACCGCGTCCAGAAGCCGGCTGTCGCGAAGCAGCAGCTGTATCAGGACCGA
77 M K R Q R L D E A S L A E D D V Y D Q S 58
5401 GTACTGGCCCCGATTTCGACGCGCCCTGGCAGGTGCCGCACACGCCCTCGCGGCAGTCCAG 5460
CATGACCGGGCTAAGCGTTCGCGGGGACCGTCCACGGCGTGTGCGGGAGCGCCGTCAGGTC
57 Y Q G S E C R G Q T G V G E R D L 38
5461 GGGGATGTTGATGCCGTTGCGCAACGCCGCGTTCGAGCAGGATTTTCGTGGCTCTGTACCGG 5520
CCCCTACAACCTACGGCAACCGGTTGCGGGCGCAGCTCGTCCTAAAGCACCGAGACATGGCC
37 P I N I G N R L A A D L L I E H S Q V P 18
5521 GAAGAACAGGGTTTTGCGCGTTCGGCAAACTGAAGGCGACCTTGTGGCTCATGATCCAGAC 5580
CTTCTTGTCCCAAACGGCAGCCGTTTTGACTTCCGCTGGAACACCGAGTACTAGGCTCTG
17 F F L T K G D A F S F A V K H S M (antC) 1
5581 TTCCGCCAGTCAGAGGTGATAGAAATCGAGCACCGAATTGATGGTGTGCTTGGAGCAGCAG 5640
AAGGCGGTTCAGTCTCCACTATCTTTAGCTCGTGGCTTAACTACCACAGCAACTCGTCGTC
163 * L H Y F D L V S N I T D N L L L 148
5641 CACGTGCTTGGCGGGCGATCTTCCAGCTCTCGCCATCGGGCCGACAGGCGGTAGGTGGCGCG 5700
GTGCACGAACGCCCGCTAGAAGGTTCGAGAGCGGTAGCCCGGGCGTCCGCCATCCACCGCGC
147 V H K R A I K W S E G D P R L R Y T A R 128
5701 CCCGAAGAAGTCTCGGGCGTTTGCAGGCGGTAGTAGAGGGTGTGCCAGTTGACTTGCAC 5760
CGGCTTCTTGACGAGCCGCCAAACGTCGCCATCATCTCCACACGGTCAACTGAACGTG
127 G F F Q E A T Q L R Y Y L T H W N V Q V 108
5761 CTCCAACCTCATCGTCGCCAGCGGCGCATGCGCACGTTGCTGAGCAGGTGCAGGGTGGCG 5820
GAGGTTGAGTAGCAGCCGTCGCCGCGCTACGCGTGAACGACTCGTCCACGTCCCACGC
107 E L E D D A L P A I R V N S L L H L T R 88
5821 CGGCATGGGCACGGTGGAGGCGGATTTGCCGGTTCGGATGCGGAACACCGGTCTTCCAG 5880
GCCGTACCCGTGCCACCTCCGCCTAAACGGCCACGCCTACGCCTTGTGGGCCAGAAGGTC
87 P M P V T S A S K G T R I R F V R D E L 68
5881 GCCGCCCCGGTTGGGGTAGTAGATCAGCGACATGGCCCCGTTTCGGGTCGCGGGTGTACTC 5940
CGGCGGGGCCAACCCCATCATCTAGTCGCTGTACCGGGCAAAGCCCAGCGCCACATGAG
67 G G R N P Y Y I L S M A R K P D R T Y E 48
5941 GTGCTCCGAGTCCCATTCGCGGCAGGTGGAATTCGCTGTGCTGCTGAACATCGACAGGTA 6000
CACGAGGCTCAGGGTAACGCCGTCACCTTAAGCGACAGCAGCGACTTGTAGCTGTCCAT
47 H E S D W Q P L H F E S D D S F M S L Y 28
6001 GGTGTCCAGTCTTGGGCGTCGCAGGCCTCGGCGTTGCGGTACAGAAATTGCTCGACGCG 6060
CCACAGGGTCCAGGACCCGCAGCGTCCGGAGCCGCAACGCCATGTCTTTAACGAGCTGCGC
27 T D W D Q A D C A E A N R Y L F Q E V R 8
6061 GTATTGCAGTTGCGAATTCATGTTTCACACCTCACGCAGTTTCAACGGTTGCTCTTGCTCG 6120
CATAACGTCAACGCTTAAGTACAAGTGTGGAGTGCCTCAAAGTTGCCAACGAGAACGAGC
7 Y Q L Q S N M (antB) 1
467 * V E R L K L P Q E Q E 457
6121 GCCTTGGCGGGCCAGGCCTTCGAGCAAAAAGCGTGCAGTTGCCGTTGTTGGTTGACGTAG 6180
CGGAACGCCCGGTCCGGAAGCTCGTTTTTCGCGACGGTCAACGGCACAACCAACTGCATC
456 A K R A L G E L L F R Q W N G H Q N V Y 437
6181 AGGCCCTCGTGGGTGAACTCAGTGC CGGTCAATACCGGCGAGATGCCGATGGTTTCGCTG 6240
TCCGGGAGCACCCACTTGAGTCACGGCCAGTTATGGCCGCTCTACGGCTACCAAAGCGAC
436 L G E H T F E T G T L V P S I G I T E S 417
6241 TTGGCCGTGGGGCCTTTCGACCCATTTGCCACGCCCGGGAGATGTCGCTCCAGCGCTCC 6300
AACCGGCACCCCGGAAGCTGGGTAACGGGTGCGGGGCCCTCTACAGCGAGGTGCGGAGG
416 N A T P G E V W K G V G R S I D S W R E 397
6301 AGGCGTGCCTGGAAGCCGCGCTGGGCTTCGCGGAACCTCCACCAGGTCATCGGGGGTGGCC 6360
TCCGCACGGACCTTCGGCGGACCCGAAGCGCCTTGAGGTGGTCCAGTAGCCCCCACGGG
396 L R A Q F G R Q A E R F E V L D D P T G 377

| | | |
|------|--|------|
| 6361 | ATGCCGGAGACGTTGAAGAAGTCCTCGAACTGGCGGATGCGGTTTTCCCGGTGGGCATCC | 6420 |
| | TACGGCCTCTGCAACTTCTTCAGGAGCTTGACCGCCTACGCCAAAAGGGCCAGCCGTAGG | |
| 376 | M G S V N F F D E F Q R I R N E R D A D | 357 |
| 6421 | GACTCGCCCACCACGCCCAGGCAGAAGCTGTTGATCTCGGTCTTGTTCAGGCAATCGGG | 6480 |
| | CTGAGCGGGTGGTGGGGTCCGTCTTCGACAACCTAGAGCCAGAACAAGGTCCGTTAGCCC | |
| 356 | S E G V V G L C F S N I E T K N W A I P | 337 |
| 6481 | CGGATGATGCGCAACTGCGAGCTGATCTGGTCGAGAAAGAACATGCTGGGGTAGATGTTT | 6540 |
| | GCCTACTACGCGTTGACGCTCGACTAGACCAGCTCTTCTTGTACGACCCCATCTACAAG | |
| 336 | R I I R L Q S S I Q D L F F M S P Y I N | 317 |
| 6541 | AGGTTGCGCAGGCGGTGCATCATCCACTGCGCCTTGCCCTGGCCGTACTCTTGGAGCAGG | 6600 |
| | TCCAACGCGTCCGCCACGTAGTAGGTGACGCGGAACGGGACCGGCATGAGAACCTCGTCC | |
| 316 | L N R L R H M M W Q K A G Q G Y E Q L L | 297 |
| 6601 | CGCGGCATGATGGTGGCGTAGCCGGGGCGCACGCTGGGGTTGGGCATGTCGCTGAACAGC | 6660 |
| | GCGCCGTACTACCACCGCATCGGCCCGCGTGGCAGCCCCAACCCGTACAGCGACTTGTGCG | |
| 296 | R P M I T A Y G P R V S P N P M D S F L | 277 |
| 6661 | ACGCTGTGGCCGTTGTGGAAAGCGAACCAGCCGTCGTCGGTTTTCTTGTGCGCCGGCGCCG | 6720 |
| | TGCGACACCGGCAACACCTTTTCGCTTGGTCGGCAGCAGCCAAAGGAACAGCGCCGCGGC | |
| 276 | V S H G N H F A F W G D D T E K D G A G | 257 |
| 6721 | AGCTTGCTGTAGTCCAGGGTGTGCGCGCTGGTCAGGCCTTTTTTCGCTGTTGACCTGCTGG | 6780 |
| | TCGAACGACATCAGGTCCCACAGCCGCGACCAGTCCGGAAAAAGCGACAACCTGGACGACC | |
| 256 | L K S Y D L T D A S T L G K E S N V Q Q | 237 |
| 6781 | CGGTGCTGCACGGTGGCCACGTAGTTGTAGTGCACGGTGCTGACGTGATAACCGTCCAGG | 6840 |
| | GCCACGACGTGCCACCGGTGCATCAACATCACGTGCCACGACTGCACTATTGGCAGGTCC | |
| 236 | R H Q V T A V Y N Y H V T S V H Y G D L | 217 |
| 6841 | CCGTTTTTCGTTTTGTCAGCTTCCAGTTGCCGTCGTAGGTGTAGGCCGATTTGCCCGGCAGC | 6900 |
| | GGCAAAAGCAAAACGTGCAAGGTCAACGGCAGCATCCACATCCGGCTAAACGGGCGGTGCG | |
| 216 | G N E N Q L K W N G D Y T Y A S K G P L | 197 |
| 6901 | ACCTCCAGCTCACCGGTGGGTGATTGGGCGACCATCATGTGCAAGAACAACCTTTGGCGTGC | 6960 |
| | TGGAGGTGCGAGTGGCCACCCACTAACCCGCTGGTAGTACAGCTTCTTGTGAAACCGCAGC | |
| 196 | V E L E G T P S Q A V M M D F F V K A D | 177 |
| 6961 | CCCAAAAATTCCTGGAGGCTGTCCTGGCCGTCGACGTCGAGGCTGATGAACACAAAGCCC | 7020 |
| | GGGTTTTTAAGGACCTCCGACAGGACCGGCAGCTGCAGCTCCGACTACTTGTGTTTTCGGG | |
| 176 | G L F E Q L S D Q G D V D L S I F V F G | 157 |
| 7021 | TTGTAGCTCTCGATGCGCGCTTTTTTCAGGCCACGGGTGGCCTTGTCGAACCCTTCGGGG | 7080 |
| | AACATCGAGAGCTACGCGCGAAAAAAGTCCGGTGGCCACCGGAACAGCTTGGGAAGCCCC | |
| 156 | K Y S E I R A K K L G R T A K D F G E P | 137 |
| 7081 | TACTCGCCGGGGGCTTTGACCTTGACCAGGCGGCCGTCGCTTTTGTAGCACCAGGCGTGG | 7140 |
| | ATGAGCGGCCCCCGAAACTGGAACCTGGTCCGCCGGCAGCGAAAACATCGTGGTCCGCACC | |
| 136 | Y E G P A K V K V L R G D S K Y C W A H | 117 |
| 7141 | AACGGGCAGGTGAAGGTGGACTGGTTGCCCTTGCCGACGCGGGTCAGGGTGGTGCCGCGG | 7200 |
| | TTGCCCGTCCACTTCCACCTGACCAACGGGAACGGCTGCGCCCAGTCCCACCAGGCGCC | |
| 116 | F P C T F T S Q N G K G V R T L T T G R | 97 |
| 7201 | TGCTGGCAGGCGTTGATAAGGGCATTGAGCTGGCCCTCGCCGTCGCGGTAATAATCATC | 7260 |
| | ACGACCGTCCGCAACTATTCCCGTAACTCGACCGGGAGCGGCAGCGCCATTATTAGTAG | |
| 96 | H Q C A N I L A N L Q G E G D R T I I M | 77 |
| 7261 | GGCTGGCGCCCGGCGCGCATCGTCATGAAGTCATGCTTGGTGGCGATTTTCGCTTTCGTGG | 7320 |
| | CCGACCGCGGGCCGCGGTAGCAGTACTTTCAGTACGAACAACCGCTAAAGCGAAAGCACC | |
| 76 | P Q R G A R M T M F D H K N A I E S E H | 57 |
| 7321 | CAGGCGTAGATCCAGTTCCTTCTCGAAGATCAACTCCATTTCCAGATCAAACAAGTCCGGC | 7380 |
| | GTCCGCATCTAGGTCAAGAAGAGCTTCTAGTTGAGGTAAAGGTCTAGTTTGTTCAGGCCG | |
| 56 | C A Y I W N K E F I L E M E L D F L D P | 37 |
| 7381 | TCGGTGAATATATCCCGGGCAATCCGGAAGACGCCGTCGGCGGGGCGAAAGTCCAGGCAG | 7440 |
| | AGCCACTTATATAGGGCCCGTTAGGCCTTCTGCGGCAGCCGCCCGCTTTCAGGTCCGTC | |
| 36 | E T F I D R A I R F V G D A P R F D L C | 17 |

Appendix. Fig. 2. Continued.

7441 CCTTCAATAAAGTTTTTCCATTGTTTCGACGCTTCTTGCCTGCTCATTAAATCGATACCTC 7500
GGAAGTTATTTCAAAAAGGTAACAAGCTGCGAAGAACGTGACGAGTAATTAGCTATGGAG

16 G E I F N K W Q E V S R A S S M (anta) 1
7501 GTTGCATCGGGCTAATTCGTGCATTCAATAAATACGCTGTGCCTGCATTAAATAAGTTGC 7560

7561 GGGGGGTGGGTCTATCCGCTTTATCGGTGAACACAGGCCATAAAATTGGCCGCCCGGCG 7620

7621 CGGCCCTGGCCGTGGGCGATAACGCCAAGTAAGTGGATACGAGTCTCGAATTAACAGGCT 7680

7681 AGGGATGGTCTGAAGTAGTTGTGTATTTCTGGCTGACTTCAGCCCCTGCCGATTTTAAAA 7740

7741 GCGGCAAATCGATCAAAAACGCTCAGATCACCCGCTTATTTCTGTGTTTTTAGCCGCCGC 7800

7801 GAGCACCTCGCGGCAGCCATTTCCACATTACAGGCGCACCTCTCCTGCATTGCGCCAGTA 7860
CTCGTGGAGCGCCGTGCGTAAAGGGTGAATGTCCGCGTGGAGAGGACGTAAGCGGTCAT

326 * L R V E G A N A L 318

7861 AATGTCGGCGCGCCATCCACAGGTTTCGACAGCGCGAACAGCGTCACCAGTTGCGCCGTGT 7920
TTACAGCCGCGCGGTAGGTGTCCAAGCTGTGCGGCTTGTGCGAGTGGTCAACGCGGCACA

317 L H R R A M W L N S L A F L T V L Q A T 298

7921 TCTTCGCCAGACCACGGAAGCGCACCTTACATAACCGAACTGGCGCTTGATCACCCGGA 7980
AGAAGCGGTCTGGTGCCTTCGCGTGAAGTGTATTGGCTTGACCGGAACTAGTGGGCCT

297 N K A L G R F R V K V Y G F Q R K I V R 278

7981 ACGGATGCTCGACCTTGGCTCGTACCTGAGCCTTGGCCTTCTCAATCTTGCAGCGGCTT 8040
TGCCACGAGCTGGAACCGAGCATGGACTCGGAACCGGAAGAGTTAGAACGCGCGCGAA

277 F P H E V K A R V Q A K A K E I K R A A 258

8041 TGACAAAAGCACTGCGCTTATCAAGCTTCTTGTAAAGTACTACGGCGGGCGGCGATCTGCC 8100
ACATGTTTTCGTGACGCGAATAGTTTGAAGAACATTCATGATGCCGCCCGCGCTAGACGG

257 K Y L A S R K D L K K Y T S R R A A I Q 238

8101 AGATGACCTCCCGCCTTTCATGTTTCAGGGCGTTTCTCGACGCCGGTGAACCCGCGTCGG 8160
TCTACTGGAGGGCCGGAAGTACAAGTCCCAGAAAGAGCTGCGGCCACATTGGGCGCAGCC

237 W I V E R G E H E P R K E V G T Y G A D 218

8161 CACCGACCATGTTCTCATCGCCGTGCGAGCAGCTTATCGACCTGGGTGACATCGGCCACGT 8220
GTGGCTGGTACAAGAGTAGCGGCACGTCGTCGAATAGCTGGACCCACTGTAGCCGGTGCA

217 A G V M N E D G H L L K D V Q T V D A V 198

8221 TGGCCGCCGTGCCACCACGCTGTGCACCAAGCCCGACTCATCATCGACACCGATGTGCG 8280
ACCGCGGCACGGGTGGTGCACACGTTGCGGGCTGAGTAGTAGCTGTGGCTACACGC

197 N A A T G V V S H V L G S E D D V G I H 178

8281 CCTTCATGCCGAAGTAGTACTGGTTACCCTTCTTGGTCTGATGCATTTCCGGGTGCGGCT 8340
GGAAGTACGGCTTCATCATGACCAATGGGAAGAACCAGACTACGTAAAGGCCAGCGCGA

177 A K M G F Y Y Q N G K K T Q H M E P D R 158

8341 TACCGTCTTGTCTTGGTTCGAACTCGGCGCATTGATCAGCGTGGCATCGACGATCGTGC 8400
ATGGCAGGAACAAGAACCAGCTTGAGCCGCGTAACTAGTCGCACCGTAGCTGCTAGCACG

157 K G D K N K T S S P A N I L T A D V I T 138

8401 CTTGACGCAGCGACAGCCCTCGGTGCGCCAGATAGCCATTGATGACAGCCAAGATGCCGG 8460
GAACTGCGTCTGCTGTCGGGAGCCAGCGGGTCTATCGGTAACACTGTCGGTCTACGGCC

137 G Q R L S L G R D G L Y G N I V A L I G 118

8461 CGGCTAGTTCGTGCTTCTCCAGCAAACGACGGAAGTTGAGGATGGTGGTTTCATCGGGGA 8520
GCCGATCAAGCACGAAGAGGTCGTTTGGCTGCCTTCAACTCCTACCACCAAAGTAGCCCT

117 A A L E H K E L L R R F N L I T T E D P 98

8521 TACGCTCCAAGCTCAGCCCCGGAAGTGGCGCAGGATGGTGGTCTCGTACAGCGCCTT 8580
ATGCGAGTTTCGAGTCGGGGCGCTTGACCGCGTCTACCACCAGAGCATGTCGCGGAGAA

97 I R E L S L G A F Q R L I T T E Y L A E 78

Appendix. Fig. 2. Continued.

8581 CCATCGCTGGATCGCTGTAACCGAACCAGTTTTGCATCAGGTGCACGCGCAGCATCGCCA 8640
GGTAGCGACCTAGCGACATTGGCTTGGTCAAAACGTAGTCCACGTGCGCGTCGTAGCGGT
77 E M A P D S Y G F W N Q M L H V R L M A 58
8641 TCAGCGGATAGGCCGGACGACCGCCTTACCCTTGGGATAATGCGGTTTCGATCAAGGCAA 8700
AGTCGCCTATCCGGCCTGCTGGCGGAAGTGGGAACCTATTACGCCAAGCTAGTTCCGTT
57 M L P Y A P R G G E G K P Y H P E I L A 38
8701 TCAAACCTTCCACGGCACCACCCGATCCATCTCGATCAGGAACAACCTTTGCGGGTTT 8760
AGTTTGGGAAGGTGCCGTGGTGGGCTAGGTAGAGCTAGTCCTTGTGAGAAACGCCAAA
37 I L G K W P V V R D M E I L F L E K R T 18
8761 GTTTGCCTTACCGGCGTACTCGGCGTCAGCGAAGGTCATCTGCTTCATGGGGAAACTCA 8820
CAAACCGGAATGGCCGCATGAGCCGCAGTCGCTTCCAGTAGACGAAGTACCCCTTTGAGT
17 Q K R K G A Y E A D A F T M Q K M (tnp2) 1
8821 GCGGGTGAATCCGGGTATTTTGCCAAAATCTGGAAGTCTTCTTCAGAGTTCCCTAGAG 8880
8881 AGTCTGGCGCCATTGGCGCTTAATCACTGGCGATGCGGGTTTGCCCGGTGGCCGGCCCGG 8940
8941 CATCGACCCTAACGACTCATGCTTAAAAATTTAATAAAGAAACAACCTGCCAACTATTGGG 9000
9001 GCGGCTACTTTGGTTCGCACCGCTACTTAATGTGCCCGCACATTAAAAGAGAGGGGATAAGT 9060
9061 TATGTTTCAAGCGAGTTGGTTTACCCGTGGTATGCAAACGGCCAGCCTGGCGGCATTAAT 9120
1 M F Q A S W F T R G M Q T A S L A A L M 20
(ORF24)
9121 GACCCCGATGCTGGCCCAGGCGGACTTCGTTGAAGACAGCCAGTTGGCCCTGGGGTTGCG 9180
21 T P M L A Q A D F V E D S Q L A L G L R 40
9181 TAACTTTTATATCGACCGCGACTTCAAGGGTGACAACCCGAAGAACTCCCGGGTTCGGCAG 9240
41 N F Y I D R D F K G D N P K N S R V G S 60
9241 TTGGACCCAGGGTTTTGACCTGCGCTTCAACTCCGGCTACACCGAAGGCACCCTGCAATT 9300
61 W T Q G F D L R F N S G Y T E G T L Q F 80
9301 TGCCCTCGACGCCTCGATGCAATACGCCTACCGGCTCGACGGCGGTGGCGGCCGCGGCC 9360
81 A L D A S M Q Y A Y R L D G G G G R G P 100
9361 GGACAGCATCATCCCCTACAGCACCCGCCCGCAGGAGCAGGCCACCGACTACGGGCGCGG 9420
101 D S I I P Y S T R R Q E Q A T D Y G R G 120
9421 CGGCTACACGGCGAAAGTGCCTATTCCAAGACCGAGCTTAAATACGGCGAACACCGGCC 9480
121 G Y T A K V R Y S K T E L K Y G E H R P 140
9481 GACGCTGCCGGTGGCCTACACCGACGATTCGCGGCAATTGGTGACCACCTACCACGGCAC 9540
141 T L P V A Y T D D S R Q L V T T Y H G T 160
9541 GCTGATCGAATCCAAGGAGGTGGACAAACTCACCTGACCGGCGGGCGCTTACCAGAAAT 9600
161 L I E S K E V D K L T L T G G R F T E I 180
9601 CAGCGGCCGCGAATCCTCCAACAGCGAAAAAATGTACCTGTTCAACGGCCCCGATGTGCA 9660
181 S G R E S S N S E K M Y L F N G P D V Q 200
9661 GCGGCGCAGTGACGGCTTGAACCTTCGGTGGCGCCAGTTACGCCTTACGCCCGGCCCTGAC 9720
201 R R S D G L N F G G A S Y A F T P A L T 220
9721 CGGCACCTACTTTTACGGGCAACTGGAAGACATCTACCGCCAGCATTACCTGGGCGCCAC 9780
221 G T Y F Y G Q L E D I Y R Q H Y L G A T 240
9781 CTACCTGGCCGACCTGGGCAGTGGCTATGCGCTCAAGACCGACCTGCGCTATTTCAATAA 9840
241 Y L A D L G S G Y A L K T D L R Y F N N 260
9841 CAAGGAAGACGGCGATGCCCTTTATGGGCGCATCGACAACCGCTCCTACGGCGCCATGAC 9900
261 K E D G D A L Y G R I D N R S Y G A M T 280
9901 CACCCTGCGCAAGGGCGCCACTCCTTTGGCCTGGGCTACCAGCGCATGCTCGGCGACAG 9960
281 T L R K G A H S F G L G Y Q R M L G D S 300
9961 CAGTTCCCCACCCTCAATGGCTACGCCCCGACCGGTACCTGGTGAACCTGGTGGCGGGT 10020
301 T F P T L N G Y A P Q P Y L V N W S A V 320

Appendix. Fig. 2. Continued.

10021 TGCCTTCGTCAAGCCCAACGAAAAATCCTGGCAGGCGCGCTATGACTACGACTTCGCCGG 10080
321 A F V K P N E K S W Q A R Y D Y D F A G 340
10081 CCTCGGCCTGCCGGGCTGCGCCTGATGACGCGGTACATGCGTGGCAGCGATATCGACCG 10140
341 L G L P G L R L M T R Y M R G S D I D R 360
10141 CGGCGCCCTGAGCGACAACAGCGAGTCGGAGCGCAATATCTACCTCTCCTACGTGGTCCA 10200
361 G A L S D N S E S E R N I Y L S Y V V Q 380
10201 GGAGGGGCGCTGAAGAACCCTCGGCGTAGAGCTGCGCAACATCAACGTCAAGTTCGCCCA 10260
381 E G P L K N L G V E L R N I N V K F R H 400
10261 CGGCAATGACTTCGACGAAAACCGCCTGATTACCACCTACACCTGGAAATTCTGGTAAAC 10320
401 G N D F D E N R L I T T Y T W K F W * 418
10321 CGGTGCCGCCGGCCGGTGGGGTAACACCCGCCGGCCGATCTGGGCAGGTAATACCCAACC 10380

10381 CCGCCGCTGGGGAATACCTGCCCCGGCGCACCGGCGGCTCATCTGTGTGCGTTTTTGCAGC 10440

10441 GCCACTGTGTGGTTTGGGGCGCGTGGGGAAGCGGGAATCCTGGCGTTGTGCTATCGCTG 10500

10501 GCCGATTAGTCGCTGAATGATATCCGCATATTTTCGCAAGGGCCCCGCGGCCCCACGCG 10560

10561 GCATGAAGATTGCTGTTTGACCGTAAGACGTGCCGTCAGAGCGGTACAAAAAACAGCA 10620

10621 AACAGCCGTGGGTGCCCGTGATGAGTACAAGCCTTGATGTGCATGTTTCGAGACATTCGT 10680
1 (ORF23) M M S T S L D V H V R D I R 14
10681 ATAGAGCGCTATGACCTCAATGGCGCGGACAGTGGATGGCCGGGATCTGCGGGCCGCAC 10740
15 I E R Y D L N G A R Q W M A G I C G P H 34
10741 CTGCTCAAGACCCAGTACCCCAACCGCATCCAGTTCATCACAGCGCCAATGTGCTCAAG 10800
35 L L K T Q Y P N R I Q F H H S A N V L K 54
10801 TCGATGTCCACCATCCTGGGGGTGATCGAGTACGGCACGGACGTGACGGTGGGCATCGAA 10860
55 S M S T I L G V I E Y G T D V T V G I E 74
10861 GACGTGAAACACCTCAACAGCTACAGCCTCAGCCTGCCCTGGTGGGCGAGCAGCACCTG 10920
75 D V E H L N S Y S L S L P L V G E Q H L 94
10921 AGCAAGGACGGCTGCCAGGTGACCTCCGACCGCGACAGCGGATCATCGTCTCGCCCCAC 10980
95 S K D G C Q V T S D R D S A I I V S P H 114
10981 GAAAACCAGGAAGTACCATCAATGGCGACTGCCGCAAGGTGCAGGTGGTCATCACCCGC 11040
115 E N Q E L T I N G D C R K V Q V V I T R 134
11041 GCGGCCATGCGCAAATCCCTGGAAGACATGCTGCAACGGCCCTGGAAACGCCCTGCGG 11100
135 A A M R K S L E D M L Q R P L E T P L R 154
11101 TTCGAGCCGGTGTGATGGACGCGGTCAACGGCGCGTCTGCGTCTGGTGGCGCATGGCGCGC 11160
155 F E P V M D A V N G A S A S W W R M A R 174
11161 TATTTTCATCGGCGAACTGGAGCGCAGCCGCGAGCTGTACGACCAGACCTTTTTCACCCGT 11220
175 Y F I G E L E R S R E L Y D Q T F F T R 194
11221 GATATCGAAAGTTTCGCTTATCAAGGGCTTGATCCTCGCCCAGCCCAACAACACTCCGCAC 11280
195 D I E S S L I K G L I L A Q P N N Y S D 214
11281 GAATTGCGCGACGTGCTCGGCGTGAAGCTGCCGCACTACCTGATCCGCGCCAAGCAGTAC 11340
215 E L R D V L G V K L P H Y L I R A K Q Y 234
11341 ATCCACGATAACGCCCGGAGGTGCTGCACCTGGAGGACATCGAGGCGATCTCCGGGGTG 11400
235 I H D N A R E V L H L E D I E A I S G V 254
11401 TCGCGGTTCAAGCTGTTTCAAGGCTTCAAGAAATACTTCGGCATGTGCCCCATGGCCTAC 11460
255 R F K F E G K K Y F M S P M A Y 274
11461 CTGAAAAAGTACCGCCTCGACGCCGTACGCCAGGAAATCCTCGAAGACAACCTCGGCGCGC 11520
275 L K K Y L D A V R Q E I L E D N S A R 294
11521 AATATTTTCGGTGTGATCGCCATGGGCTGGGGCTTCAACCACCTGGGCCGTTTCTCCAGCGAG 11580
295 N S V M G W N H L G R S S E 314

Appendix. Fig. 2. Continued. Seventeen amino acid residues are well conserved among the AraC/XylS family regulatory proteins, and this sequences was conserved in at least 60% of aligned proteins. thirteen amino acid residues were conserved in ORF23 (shaded boxes).

11581 TACCGCAAGTTGTTTCGACGAGACGCCAGCAGCACCTGCAACGCAACGAAGCCC GCCG 11640
315 Y **K** L F D E T **S** S T L Q R N E A R R 334
11641 AACCGGTCGCTTTGAGCCATGGCCGGGTTAGCGGGGGCGGGGCTCGACCTGTGCGGGCTG 11700
335 N R S L * 338
1 (ORF22) M A G L A G A G L D L S R L 14
11701 TGCCAAACGCCGGTGTTCACCAGCCAATCGCGCCAGGAAAGCCATCGGCTGCTGGCCCGT 11760
15 C Q T P V F T S Q S R Q E S H R L L A R 34
11761 GAGTTGGTCGAGCACGACCTGAGTTGGCGCGCAGGCACGGTCGATACGGCGTTCTTTTCGC 11820
35 E L V E H D L S W R A G T V D T A F F R 54
11821 GCGCAAGTGGGGCGCTGCCAACTGTTTGTGTTGCGTTATGGCGCGGAGGTTGAGGTGCGG 11880
55 A Q V G R C Q L F V L R Y G A E V E V R 74
11881 CCCC GGCCGTTACCGATTTCTGTGCTGATGCAAATGCCCTGCGCGGGCGCGCCAGTTTC 11940
75 P R P F T D F V L M Q M P L R G R A S F 94
11941 GAATGCGACGGCTTGGCCCTGGAGCTGGAAGCGGGGCAGGTGGCGGTGCTCTCGCCCAAC 12000
95 E C D G L A L E L E A G Q V A V L S P N 114
12001 CGCCAGGCGCGGCTGGTGTGGCACC CGGTTGCGAACAGCTATTGCTCAAGGTGCCGCGC 12060
115 R Q A R L V W H R G C E Q L L L K V P R 134
12061 GCGTTGTTGCGCCTGATGCAGGAACGGCTGGTGGACGAAGGTGGGCTGGCCGCGGATTTTC 12120
135 A L L R L M Q E R L V D E G G L A A D F 154
12121 AGCCTGCCGCCAGCCTACCGGCTCGACACGGGCCAGGCCCGTGGCTGGCTACGTTTGGTG 12180
155 S L P P A Y R L D T G Q A R G W L R L V 174
12181 GATGACCTGCTGGAACA ACTGGGGGGCCAGCGCCGGCGAACCGGCCCATGGCCACTGGCT 12240
175 D D L L E Q L G G Q R R R T G P W P L A 194
12241 GCGGCACCTGGAAGAAAGCCTGCCGCTGTTTTTGTCTACCCACCCGCCGAAAGTCAAGG 12300
195 A A P G R K P A A V F A H P P A E S R R 214
12301 TTTGGCCTTGGCCGGGTCCCCGGCCAGCGCCAGGCGTTGCGCCATTGCACAAGATCGAA 12360
215 F G L G R V P G P A P G V A P L H K I E 234
12361 GGCTTGATGCGCAGCCGCCTGGGGGAAGCGCTGTCCCTCACCGAACTGGCGGCCACGGCG 12420
235 G L M R S R L G E A L S L T E L **A** T A 254
12421 GGGATGAGCGTGCAGCCTCAACACCCCTGTGCCAGCGCTGCTACGGGCAAACCCCATG 12480
255 G M **V** R S **N** T L C Q R C Y **Q** T P M 274
12481 GAGCGCCTGCGCAACTGCGGCTCGAAGCCGTGCGCGAGCTTGC GCGATGGCCGCGAGCTG 12540
275 E R L R N C G S K P C A S L R D G R E L 294
12541 AGCATCACCGAAGTCGCCCTGGAACACGTTCCGCCACCTCGGCCGTTCCGCCGCTACTAC 12600
295 S **T** E **L** E H V R P P R P **A** R Y Y 314
12601 CGCGAACGCTTCGGCGAGCTGCCGCACCAGACTATTGCCAGTAAGTGACCGGGGGCGCG 12660
315 **E** R F **E** L **H** Q T I A Q * 328
12661 GTGGCCTGGCCGCACATTGCGCTATCACCCCTCACTGAATTGTGGGAGCAGTGCCTGACGC 12720

12721 TCATCTCACCAGCACCCGCGCTCCACTGTGGGAGCAAGGGTCTTACACCGACCATTTCGTC 12780

12781 ACCGCTACTCAGTGCCGCCCTCGCGAGCAGGCTCGCTCCCACAGAGGAAAAATTAAATCC 12840

12841 CAGCCACAGCACCGTTCAATTGTGGGAGCGGGCCTGCCCGGATGACGGCCTATCAGCCG 12900

12901 CAGCAAATGTATCGCCTGGCTTGGACCCATTCGCGAGCAGGCTCGCTCCCACAGAGGAAA 12960

12961 AATTAATCCCAGCCACCGCACCGTTCAATTGTGGGAGCGGGCCTGCCCGGATGGCGGC 13020

13021 CTATCAGCCGAGCAAATGTATTGCCGCTACTGGACCCATCGCGGGCAGGCCGTTCCACA 13080

13081 GCGTTTGCAGCCTTTACGGGGAAGGCTGGCGCTAGCGTTGCGGCTGTCCGCCCATTTG 13140

Appendix. Fig. 2. Continued. Eleven amino acid residues out of 17 amino acid residues conserved among the AraC/XylS family were conserved in ORF22 (shaded boxes).

13141 CCCATCAGGGTGTGGAGGGCAGTTGTTTCGTCGAGCAGTTTCCTCGCCGTTTCTACCCCG 13200
13201 GCCACCGGCAGGCCGGTGGGGTTCGAGGAGGCCAATTTGCACGAGGACGCTGGCCTGGTCC 13260
13261 CAGTAGATGTGCTCATGGCACAGCTTGTGCGCCGCGAACTTGATGACCCCCAGCATGGGG 13320
13321 ATTTCCACGTATTTGCCGGTGGGGGCGACGTTGGGCAGCATCCAGTCGATTTGCTGGTG 13380
13381 TGGGTAAAGCACATGACGAATTCATCGACTATCTGCGTGGCCCCACGGTGCGGGAGATG 13440
13441 GGCACCAGTTTCATGTCCGGCGGGTTGGCGTGGACAAAATGGTTCTGATAGAAGCGGCTC 13500
13501 AGTTCGCGGTAGCCCACGCCCCGGTCATGGTGGGAATGTGGTTGACGTAGGGCTCGGCG 13560
TCAAGCGCCATCGGGTGCGGGGGCCAGTACCACCCTTACACCAACTGCATCCCGAGCCGC
273 * P P F T T S T P S P 264
13561 ACCATGGTGCATGGTGC CGCGAACATCGCGCAGGCGAATTCAAACTCGATGTGTTTTT 13620
TGGTACCAGCGTACCAGCGGCCTTGTAGCGCGTGC CGCTTAAGTTTGAGCTACACAAAAA
263 S W P R M T A P V D R V A F E F E I H K 244
13621 CCCACAGTGC GGACAGGTTGTAGTGC GGCGCGATGGCCTGCCTCAGCGCGCGATGCTGC 13680
GGGTGTCACGCCTGTCCAACATCACGCCCCGGCTACCGGACGGAGTCGCGCCGCTACGACG
243 E W L A S L N Y H P G I A Q R L A A I S 224
13681 GTTGGTGC GCCATCAGGGCGGGCGGGTTTGTGTAGTGCATGCCGCCGTGGCGGGCGAAGG 13740
CAACCACGCGGTAGTCCC GCCCCAAACAACATCACGTACGGCGGCACCGCCCCGCTTCC
223 R Q H A M L A A P K N Y H M G G H R A F 204
13741 CGTGGTCCACCTCGGGGTAGGTGTAGATCTCGACGTTGGGCCGCGCGCCAGGCGTTGGC 13800
GCACCAGGTGGAGCCCCATCCACATCTAGAGCTGCAACCCGGCGCGGGTCCGCAACCG
203 A H D V E P Y T Y I E V N P R A G L R Q 184
13801 TGATGAGGGCGCGGGCATTGGCGTGCAGTAGCCGTCCTGTTTCGGCGAAGTGCAGCACCA 13860
ACTACTCCC CGCGCCCGTAACCGCAGCGTCATCGGCAGGACAAGCCGCTTACGTCGTGGT
183 S I L A R A N A D C Y G D Q E A F H L V 164
13861 GGTGGCCCTTGATGCGCTCGGCTTCGTCCAGCAGGTGCTCGATGCCCATGCCGTAGTAGC 13920
CCACCGGGAAC TACGCGAGCCGAAGCAGGTGCTCCACGAGCTACGGGTACGGCATCATCG
163 L H G K I R E A E D L L H E I G M G Y Y 144
13921 CCACCGCGCAGGCCACGTCGGTGC GGGTGCGGGTTCAGGTAGGCGAGCTTGCCGCCCATGC 13980
GGTGGCGCGTCCGGTGCAGCCACGCCAGCGCCAGTCCATCCGCTCGAACGGCGGGTACG
143 G V A C A V D T R T A T L Y A L K G G M 124
13981 AGAAACCCACGTAGCCCAGGCCGCTGGCTTCGACCTCGGGCAGCTCACGCAGGGCTTTGA 14040
TCTTTGGGTGCATCGGGTCCGGCGACCGAAGCTGGAGCCC GTCGAGTGCCTCCGAAACT
123 C F G V Y G L G S A E V E P L E R L A K 104
14041 AGCTGGCGTCGATGTCCTCGACGGCCTTGTCTATGTCCAGGCGCTGGAACAGGTGATGG 14100
TCGACCGCAGCTACAGGAGCTGCCGGAACAGATACAGGTCCGCGACCTTGTCCAGCTACC
103 F S A D I D E V A K D I D L R Q F L D I 84
14101 CGCGGGCGAAGTCCGGCTTCGGTGTAGCCCAGGTCCACCCCAGGTCCAGGCGCCAGAACA 14160
GCGCCCCGCTT CAGCCGAAGCCACATCGGGTCCAGGTGGGGTCCAAGGTCCGCGGTCTTGT
83 A R A F D A E T Y G L D V G P E L R W F 64
14161 GGTCTGGCACCAAGGCGACGTAGCCTTCTTCGGCGTAGAGGTGCGCCACTTCGCGCATGT 14220
CCAGACCGTGGTTCCGCTGCATCGGAAGAAGCCGCATCTCCAGCCGGTGAAGCGCGTACA
63 L D P V L A V Y G E E A Y L D A V E R M 44
14221 TGGCATTGATGCCGAAAATCTCCTGGCCGATCACCACCCCCGGCCCTTGCCGCTGGCGG 14280
ACCGTAACTACGGCTTTTAGAGGACCGGCTAGTGGTGGGGGCCGGGGAACGGCGACCGCC
43 N A N I G F I E Q G I V V G P G K G S A 24
14281 GGAGCGCCAGGTAACCGCTGAAACTGCCGCTGCCGTCGCTGGCCTGTACCTTCATCATCT 14340
CCTCGCGGTCCATTGGCGACTTTGACGGCGACGGCAGCGACCGGACATGGAAGTAGTAGA
23 P L A L Y G S F S G S G D S A Q V K M M 4

Appendix. Fig. 2. Continued.

14341 TGCCCATGGTGTACCTCGTTTTTCGACTCATCAGGTGTTGTTGTTTTAGTTGCCTACTGG 14400
ACGGGTACCACAATGGAGCAAAAAGCTGAGTAGTCCACAACAACAAAATCAACGGATGACC
3 K G M (ORF21) 1
269 * Q 269
14401 TCACGCAAACGCTCGAACCAGGTGCGTGGCGACAGGCCCTGGTCGGGGCCGAGGATCGAG 14460
AGTGCCTTTGCGAGCTTGGTCCACGCACCGCTGTCCGGGACCAGCCCCGGCTCCTAGCTC
268 D R L R E F W T R P S L G Q D P G L I S 249
14461 AAACCGCCGTCCACCGGGATATCGACGCCGGTGATCCACGACGCACCTGGCTGCACAGG 14520
TTTGGCGGCAGGTGGCCCTATAGCTGCGGCCACTAGGTGCTGCGTGGGACCGACGTGTCC
248 F G G D V P I D V G T I W S A G Q S C L 229
14521 AAGGCCACTACCTGGCCGATTTCCCTCGCCGCGGCCACGCGCCCCGAGCGGGTGAAAGGTA 14580
TTCCGGTGATGGACCGGCTAAAGGAGCGGCGCCGGGTGCGCGGGCTCGCCCACTTTCCAT
228 F A V V Q G I E E G R G V R G L P H F T 209
14581 GCACCGATCCGGTCGGCCACGGCCACATCGCCGCCACTGAGGCTGGCCACCGAGGGCGAC 14640
CGTGGCTAGGCCAGCCGGTGCCGGTGTAGCGCGGGTACTCCGACCGGTGGCTCCCCGCTG
208 A G I R D A V A V D G G S L S A V S P S 189
14641 CAGGTCCAGGCCGGCGATACGCTGATGACCCGGATCGCCGCGGGCGCCAGGCTCACCGCG 14700
GTCCAGGTCCGGCCGCTATGCGACTACTGGGCCTAGCGGCGGGCCGCGGTCCGAGTGGCGC
188 W T W A P S V S I V R I A A P A L S V A 169
14701 AGGTTGCGGGTCAGTTGCAGGAGCCCGGCCTTGGACGCTGGGTACAGGGCGCGGCCAGCG 14760
TCCAACGCCCAGTCAACGTCTCTCGGGCCGGAACCTGCGACCCATGTCCCGCGCCGGTCCG
168 L N R T L Q L L G A K S A P Y L A R G A 149
14761 GCGCCGAACCTGCGCCGGTGCTGCCGATGTTGATGACCGTGCCGCGCCCGTCATGTGG 14820
CGCGGCTTGAACGGCGGCCACGACGGCTACAATACTGGCACGGCGGGCGGAGTACACC
148 A G F K G G T S G I N I V T G G G T M H 129
14821 GGCGCGGCTTTTTGACGAAGATCGCCGCGGACACCAGGTTGACGTTGAGGGTCTTGAGC 14880
CCGCGCCGAAAAACGTGCTTCTAGCGGGCTGTGGTCCAACCTGCAACTCCCAGAACTCG
128 P A A K Q V F I A A S V L N V N L T K L 109
14881 CATTGCGCACGGTCGGACAGCAGCCCCTGGTCGTATACACGCAGGCGTTGTTGACCAGG 14940
GTAACGCGTGCCAGCCTGTCTGTCGGGGACCAGCAGTATGTGCGTCCGCAACAACTGGTCC
108 W Q A R D S L L G Q D D Y V C A N N V L 89
14941 ATGTCGAGGCGGCCGAACGCGGCCAGGGTGGCGGCGATGCAGGCATCGATATCGTCTGTCG 15000
TACAGCTCCGCGGCTTGCGCCGGTCCCACCGCCGCTACGTCCGTAGCTATAGCAGCAGC
88 I D L R G F A A L T A A I C A D I D D D 69
15001 CGCTCGACGTCGCAGGCTGCAAAACGGCAGCGTTCACCCAATGCCTCGGCCAGGGCCGCG 15060
GCGAGCTGCAGCGTCCGACGTTTTGCGCTCGCAAGTGGGTTACGGAGCCGGTCCCAGGCGC
68 R E V D C A A F R C R E G L A E A L A A 49
15061 CCCGTGTGCGGTTGCGTCCCACCAAGGTGACGCTGGCGCCGGCACGCAGCAGTTGGCGG 15120
GGGCACAGCGCCAACGCAGGGTGGTTCCTACTGCGACCGCGCCGTCGCTCGTCAACCGCC
48 G T D R N R G V L T V S A G A R L L Q R 29
15121 GCAATGTCGGTGCCCAGGCCCTGGGTGCTGCCGGTGACGATCGCCACCTGGCCTTGCAGG 15180
CGTTACAGCCACGGGTCCGGGACCCACGACGGCCACTGCTAGCGGTGGACCGGAACGTCC
28 A I D T G L G Q T S G T V I A V Q G Q L 9
15181 TCGGCAGTCGGCAGGGGCTTCATGGCCGATCCCCAGGGTGTGGCGACATCGCTGAGGG 15240
AGCCGTCAGCCGTCCCCGAAGTACCGGCTAGGGGGTCCCACGACCGCTGTAGCGACTCCC
8 D A T P L P K M (ORF20) 1
15241 TGGCGATCAGGCTCATGCTCTGCAGCGCTGCGCGGTGGGTTTTAGGGTTGGCGGCGGCGC 15300
15301 AGGCATCGGCCAGGACCGTGACCTCGAAACCGCAATCGACGGCATGGCGCACGGTGCTTT 15360
15361 CCACCACCGAATGGGTGCTACGCCCGCCACCAGCAATTGGCGTGCCTCGAGCTGGCGCAG 15420
15421 CAGGGTGTCCAGGGGCGTGCCGTAGAAGGCGTTGATGCGGGTGTGGGTGACGACAAACTC 15480

Appendix. Fig. 2. Continued.

15481 GGTGGCGCTGGCCAAGGGTTGCAGGGCGTCTGGAAGTGCAGCGCCCCAGGTGCCTTCGCG 15540
15541 CACCGCACCGCTGGCCGCCACGGCGCGCAGGATCGGGCGGTTCTGCACCAGGTGGCGTA 15600
15601 GTCCGGGCGATAGGCCACGCGCACATGAATGATCGGCCAGCCCCGGGCGGGGCGCCGTG 15660
15661 TTGCAGGCGCATGGCCGCGTCCAGCACGGTGGTGGGGCCGGGTCGTGGCGGCCAGGCC 15720
15721 GACGCGGATGCGCCCATCGGGGTGCAGCACGTCTGTTCTGGTAGTGCAGGGCAAGAACGGT 15780
15781 CGGGCGCCCCATCAGACAAATACCTCCAGGTTGGCGAACACGTCGTTGCAGTTGATGAGG 15840
GCCCCGGGGTAGTCTGTTTATGGAGGTCCAACCGCTTGTGCAGCAACGTCAACTACTCC
177 * V F V E L N A F V D N C N I L 163
15841 TCGACGCGCTTGAGCTGGATCTTCCAGCCGCCCCCTCTGGCTGGCCAGTTTGTAGGTA 15900
AGCTGCGCGAACTCGACCTAGAAGGTGCGGGGGGGAGGACCGACCGGTCAAACATCCAT
162 D V R K L Q I K W G G G E Q S A L K Y T 143
15901 TAGCGCCCGCCAATTGGCGTTGCACCAGGCGCCATTCGCTGACCTGGAAGGTGGCGCTC 15960
ATCGCGGGCCGGTTAACCGCAACGTGGTCCGCGGTAAGCGACTGGACCTTCCACCGCGAG
142 Y R G A L Q R Q V L R W E S V Q F T A S 123
15961 ACCACCAGCAGGCCCTGGCGGTGCGAGGCTTCGACCATCACGTTGCCACCAGGTGGGCG 16020
TGGTGGTCTCGGGACCGCCAGCTCCGGAAGCTGGTAGTGCAACGGGTGGTCCACCCGC
122 V V L L G Q R D L G E V M V N G V L H A 103
16021 GTGCGGGTGGGCGGTTGTTGCGACCAGTTGCGCGGTTTTCACGCGGGCGGATGCGCACC 16080
CACGCCACCCGCCAACAACGCTGGTCAACGCGCCAAAAGGTGCGCCGCCTACGCGTGG
102 T R T P P Q Q S W N R P N E V R R I R V 83
16081 TCGCGCAGCATGCGGTTTTCCAGAACAGCGAAATATGCTCGAACGGGTGGCCTGGCCG 16140
AGCGGTCGTACGCCAAAAGGGTCTTGTGCGCTTTATACGAGCTTGCCCAACCGGACCGGC
82 E R L M R N E W F L S I H E F P N A Q G 63
16141 TGCTGCTGCGGGCGCCAGTACATGCCGCTGTGCGGTGAACAAGCGGTCCCATTTCCAGAAG 16200
ACGACGACGCCGCGGGTTCATGTACGGCGACAGCCACTTGTTCGCCAGGGTAAGGGTCTTC
62 H Q Q P A W Y M G S D T F L R D W E W F 43
16201 CGCCGCTCGTCCAGCAGGCGCGCTTCGCGGTAGAGGAACTGTTCCAGCTCACGCAGGGTG 16260
GCGGCGAGCAGGTCTCGCGCGAAGCGCCATCTCCTTGACAAGGTGAGTGCCTCCAC
42 R R E D L L R A E R Y L F Q E L E R L T 23
16261 TCGCCGGGCACGTGGGCCGGGGCCAGGTTGTCCGTGGCGACGTCGAAGGTGCGATCAAAC 16320
AGCGGCCCGTGCACCCGGCCCCGGTCCAACAGGCACCGCTGCAGCTTCCACGCTAGTTG
22 D G P V H A P A L N D T A V D F T R D F 3
443 * V 443
16321 AGCATGGGTGCTCTCCTTGGCGATCAGGGTGGCGGTGATGTAGTGTCTCCACGCAGCAA 16380
TCGTACCCACGAGAGGAACCGCTAGTCCCACCGCCAGTACATCAGGAAGGTGCGTCTTT
2 L M (ORF19) 1
442 A H T S E K A I L T A T M Y H K W A A F 423
16381 CTGGTTGCGCATGGGCAGCTCGCTGGTGGCGGTTGGTGGAGATTGCGCCGTGCGGCAGCCC 16440
GACCAACGCGTACCCGTCGAGCGACCACGGCAACCACCTCTAACGCGGCAGCCCGTGGG
422 Q N R M P L E S T G N T S I A G D P L G 403
16441 CTGGTGGTGGCGGATAACGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 16500
GACCAGCCACGGCCGTATTGCTACGACCGACTAGGTAAGCGGGCGGGCCAGTACAACGG
402 Q D T G A Y R H Q S I W E G G R T M N G 383
16501 TTCCTGGCAACGGGCGTAGACCTCCACGTCATCGGGCATGACGTTGGAGGAGGGTGGAGTT 16560
AAGGACCGTTGCCCCGATCTGGAGGTGACGTTAGCCCGTACTGCAACCTCCTCCCACTCAA
382 E Q C R A Y V E V D D P M V N S S P S N 363
16561 GATGACGTTGGCGTAGGTCAGCGCCCGGTGCAACACCGCATCCGGCGCGCCTTTGCAGCG 16620
CTACTGCAACCGCATCCAGTCGCGGGCCAGCTTGTGGCGTAGGCGCGGAAACGTCGC
362 I V N A Y T L A R D F V A D P A G K C R 343

Appendix. Fig. 2. Continued.

16621 GAACGTCTGGATTTCCACCAGGGTCTGGTCCACGGCCATCGGGCGGATGATCCGAAACTG 16680
CTTGACAGACCTAAAGGTGGTCCCAGACCAGGTGCCGGTAGCCCGCCTACTAGGCTTTGAC
342 F T Q I E V L T Q D V A M P R I I R F Q 323
16681 CATGAACACGGTATGGGGCGAGCCGCTGCCGTAGATCACCGTGTGTGGCGGTTTCATGCC 16740
GTACTTGTGCCATACCCCGCTCGGCGACGGCATCTAGTGGCACAACACCCGCCAAGTACGG
322 M F V T H P S G S G Y I V T N H R N M G 303
16741 CAGGATGGCGCGGGCGCGTTCCTCGCCATAGGCCGCGCTCAGGCACTCGAAGTGGGCCTT 16800
GTCCTACCGCGCCCCGCGCAAGAAGCGGTATCCGGCGCGAGTCCGTGAGCTTCACCCGGAA
302 L I A R A R E E G Y A A S L C E F H A K 283
16801 GGACACCTCGTCGCGCTCGGCGGGCGGCTGGGTGCAAGATCGCTTCCATGTAGCCATGGCC 16860
CCTGTGGAGCAGCGGAGCCGCCGCCGACCCAGCTTCTAGCGAAGGTACATCGGTACCGG
282 S V E D R E A A A P D F I A E M Y G H G 263
16861 GTTGTGGTAAGCGCGCAGTTCGAGTTTTTCCAGAAGTCGTAGGGCTCGCCGTTGCCGTC 16920
CAACACCATTTCGCGCGTCAAGCTCAAAAAGGGTCTTCAGCATCCCAGCGGCAACGGCAG
262 N H Y A R L E L K E W F D Y P E G N G D 243
16921 CATGATGAGCAGCTCGAACGGCATCTCGCCCAGGCTCTGCGCCTCGCTCGGGGCCGCGTC 16980
GTACTACTCGTCGAGCTTGCCGTAGAGCGGGTCCGAGACCGGAGCGACGCCCGGCGCAG
242 M I L L E F P M E G L S Q A E S R A A D 223
16981 CACCGAGGATTCGTGGGTGACCGGGCGTGCATGGTGTGTCGTCAGGTTCTCGTAGAAGAC 17040
GTGGCTCCTAAGCACCCACTGCGCCCGCACGTACCACAGCACGTCCAAGAGCATCTTCTG
222 V S S E H T V R A H M T D H L N E Y F V 203
17041 CTTCCAGTTGGAGCGCTGCACCACCCGAAAAATACCGCCGGCCACTTCCACTTACCCAC 17100
GAAGTCAACCTCGCGACGTGGTGGGCCTTTTATGGCGGCCGGTGAAGGTGAAGTGGGTG
202 K W N S R Q V V R F I G G A V E V E G V 183
17101 CGGCGAGCGGTGCGAGAGGTTGTCGATGGAGGTGATGACGCCGCCGAGAAAGCCCTTGAG 17160
GCCGCTCGCCAGCGTCTCCAACAGCTACCTCCACTACTGCGGCGGCTCTTTCGGGAAGTC
182 P S R D C L N D I S T I V G G L F G K L 163
17161 GTCCGGCCCATGGGGGCTTTGGCTGGCGAACACGAAGCCGCGATAGCTGTGACCCGTCG 17220
CAGGCCGGGTACCCCCGAAACCGACCGCTTGTGCTTCGGCGCTATCGACAGCTGGGCACG
162 D P G H P S Q S A F V F G R Y S D V R A 143
17221 CAGGCCACCATGGAGAAATCCGGGTGGCTGGGGTTGAAGCAGGTGCCCTCGAAGGCGTT 17280
GTCCGGGTGGTACCTCTTTAGGCCACCGACCCCAACTTCGTCCACGGGAGCTTCCGCAA
142 L G V M S F D P H S P N F C T G E F A N 123
17281 TTTAGCGGTGCCGAGAGGTGCTGGCCGTCGAGCTTGAAGGTCCAGGCGTGGTACGGGCA 17340
AAAGTCGCCACGGCTCTCCACGACCCGGCAGCTCGAACTTCCAGGTCCGCACCATGCCCGT
122 K L P A S L H Q G D L K F T W A H Y P C 103
17341 GCGGAAGAACTTGCCGACGTTGCCGTGCGCTTCGGCTACCAGCTTGGCGCCCTTGTGCGG 17400
CGCCTTCTTGAACGGCTGCAACGGCAGCGGAAGCCGATGGTTCGAACCGCGGGAACACGCC
102 R F F K G V N G D G E A V L K A G K H P 83
17401 GCAGCGGTTGTACAGCACGTTGACGCGTTTTGTCGGCGCCACGCACCAGGATCACGTCTTG 17460
CGTCGCCAACATGTGTCGTCGCAACTGCGCAACAGCCGCGGTGCGTGGTCTTAGTGCAGAAC
82 C R N Y L V N V R K D A G R V L I V D Q 63
17461 ATCGCCGATGCGGGTGGTGTGGTAGTCGCCGGGGTTTTTACCTGGCTCTCGTGACCCAC 17520
TAGCGGCTACGCCACCACACCATCAGCGGCCCAAAAAGTGGACCGAGAGCACTGGGTG
62 D G I R T T H Y D G P N K V Q S E H G V 43
17521 GTAGATCCAGGCCTGGCCGTAGATACGCTGCATTTCCAGGGCAAACAACACTGAGGGTCCGGT 17580
CATCTAGGTCCGGACCGGCATCTATGCGACGTAAAGTCCCGTTTGTGACTCCAGCCA
42 Y I W A Q G Y I R Q M E L A F L Q P D T 23
17581 GTAGACGCTTTTGTGCACGCTGTCATCGCGCACAGGGCGGCGATCTGTTGTTGTTGACTCC 17640
CATCTGCGAAAACAGTGCAGCAGTAGCGCGTGGTCCCAGCGCTAGACAAGCAAGCACCC
22 Y V S K H V S D D R V L A A I Q E N T P 3
17641 AATCATGGTGTCTCCTTTTAACCAGCCATAAACAGGCGCGTGGGGTGGCCACGCACTTAC 17700
TTAGTACCACAGAGGAAAATTGGTCCGGTATTTGTCCGCGCACCCACCGGTGCGTGAATG

2 I M (ORF18)

* 1

Appendix. Fig. 2. Continued.

17701 AAGTCGAGCACCAGCCGAGCGCTGGAGGCGCGGAGACACAGGTGCAGAAACGGCCGTCG 17760
TTCAGCTCGTGGTTCGGCTCGCGACCTCCGCGCGCTCTGTGTCCACGTCTTTGCCGGCAGC
315 L D L V L R A S S A R S V T C F R G D 296
17761 CTGCGCTCGCGCGCTGACAGGCAGATGTCGCGGTGCTCGGCGCTGCCTTCGAGCACCTGG 17820
GACGCGAGCGCGGACTGTCCGTCTACAGCGCCACGAGCCGCGACGGAAGCTCGTGGACC
295 S R E R A S L C I D R H E A S G E L V Q 276
17821 GCCACGCAGACCCCGCAGTCGCCACGGCGGCAATCGAACAGCGGGTCCAGGTGGCGTCG 17880
CGGTGCGTCTGGGGCGTCAGCGGTGCCGCGCTTAGCTTGTGCGCCAGGTCCAGCCGACG
275 A V V G D G R R D F L P D L D A D 256
17881 AGCATGGCCTCCAGCACACTCTTGCCCGCAGGCACCTGGAGGGTGATGCCGCTGCTGCGC 17940
TCGTACCGGAGGTGCTGTGAGAACGGGCGTCCGTGGACCTCCACTACGGCGACGACGCG
255 L M A E L V S K G A P V Q L T I G S S R 236
17941 AGTTCACCTTCAAACTGGCGTCGCCCCGCGCGTGCAGGCTGCCGGTGAACAATTCGCTG 18000
TCAAGGTGAAGTTTTGACCGCAGCGGGCGCCGACGCTCCGACGGCCACTTGTTAAGCGAC
235 L E V E F S A D G A A D L S G T F L E S 216
18001 TGTACGTGGCTGTGCTCCAGCCCAACTCACGGGCGCTGTGCAGCAGCTGGCGATCAGG 18060
ACATGCACCGACAGCAGGGTTCGGGTTGAGTGCCTCGGCGACACGTCGTGCGACCGCTAGTCC
215 H V H S D D W G L E R A S H L V S A I L 196
18061 GCTTTGGGCGCCGACACATAGAGCTGGCGGCCGTTATGGGGGGTGGCCAGTAGGCTGGCC 18120
CGAAACCCGGGCGTGTGTATCTCGACCGCCGCAATACCCCCACCGGTCATCCGACCGG
195 A K P G C V Y L Q R G N H P T A L L S A 176
18121 AGGGGCAGGCGCCGCGCGGGTTCGCGCCGCTGATCCAGCAGTGGGCGCTGCCAGGGCC 18180
TCCCCGTCCGCGGCGCGGCCAGCGGGCGGCGACTAGGTGCTCACCCGCGACGGGTCCCGG
175 L P L R R A P D G G S I W C H A S G L A 156
18181 CGGACTTCTTCCAGGTATGCCATGCCCTGGGCCTCGCGCCCGGCGTAGTGCAGGGTGAAG 18240
GCCTGAAGAAGGTCCATACGGTACGGGACCCGAGCGCGGGCCGCATCACGTCCCACTTC
155 R V E E L Y A M G Q A E R G A Y H L T F 136
18241 GGCTGCTGGCCGGCGGCAAGCGCACGGGCCATGCCAGCATCGGGGTGATGCCAATGCCG 18300
CCGACGACCGGCGCCGCTTCGCGTGCCTCGGTACGGGTCGTAGCCCCACTACGGTTACGGC
135 P Q Q G A A L A R A M G L M P T I G I G 116
18301 CCGGCGATCAACAGGTAGTCGCTGGCCTGCGTCACCAAGGGGAAATGATTCTTCGGTGCC 18360
GGCCGCTAGTTGTCCATCAGCGACCGGACGCGAGTGGTTCCTTACTAAGAAGCCACGG
115 G A I L L Y D S A Q T V L P F H N K P A 96
18361 CGCACCTGTAGCGGCTGGCCGACCTGCAATTGGTGGATCCAGCGGAACCGCCGCTGCTC 18420
GCGTGGACATCGCCGACCGGCTGGACGTTAACCACCTAGGTGCGGCTTGGCGGCGACGAG
95 R V Q L P Q G V Q L Q H I W R S G G S S 76
18421 TGTTCTTCCAGTTGCACGGCGATCTCGTAGTGGGCGTGGCCCGGCGAGGTTGACCAGGGAA 18480
ACAAGAAGGTCAACGTGCCGCTAGAGCATACCCGACCGGGCCGTTCAACTGGTCCCTT
75 Q E E L Q V A I E Y H A H G P L N V L S 56
18481 TAGGCGCGCCATTGGGGCTTGTGCCCCGCAATCTGGATCTCGATGTGCGCCCTGGAGTG 18540
ATCCGCGCGGTAACCCGAACGACGGGCGTTAGACCTAGAGCTACACGCGGGGACCTCAC
55 Y A R W Q P K S G A I Q I E I H A G P T 36
18541 AAGCCGGGCAGGGGCTGCTGGCATCCCCACCAGGATCAGGCGGCGAATGCGCGGGGTGA 18600
TTCGGCCCGTCCCGACGACCGTAGGGGGTGGTCTTAGTCCGCGGCTTACGCGCCCACT
35 F G P L P Q Q C G G G P D P P S H A P H 16
18601 GCGGGTCGATTGCTTCGATACGCAAGGACAGGCTGGGCATGCACGATTCCTGAAACGGGG 18660
CGCCCAGCTAAGCAAGCTATGCGTTCTGTCCGACCCGTACGTGCTAAGGACTTTGCCCC
15 A P R N T R Y A L V P Q A H M (ORF17) 1
18661 CGAGCAGGGCCCGCCGTTGAAGTCGCCCTCACTGTAGGCACGGCGCGGGGGCAAAAAG 18720
18721 CCGATATCGGCAAGTCGTTAGCCGCATTTTGCCTAAATGCGCGCCAGGGCTGGGGGCACT 18780

Appendix. Fig. 2. Continued. The conserved Cys residues (CX₄CX₂CX₂₈C) which may coordinate a [2Fe-2S] cluster in plant-type iron-sulfur proteins [Otake and Ooi 1989] are indicated by shaded boxes.

18781 CATCACTGTGGGAGCTGGCCCAGGTGCCCATAGGGGGGAGGTATTGGCGGGCGGCTATGC 18840

18841 ACCGGCTAAAGTCGGCATTTTTTTTGACGGAGGAACGAACATGCAGGCATTCAACAGTGGC 18900
TGGCCGATTTTCAGCCGTAAAAAACTGCCTCCTTGCTTGTACGTCCGTAAGTTGTCACCG
132 * L R C K K S P P V F M C A N L L P 116
18901 TGGGGACAATGGCGGCTGGGGGCGCTGCCGGGCTTCTGCTCGTTATGGCGCCGCCAGG 18960
ACCCCTGTTACCGCCGACCCCCGGCGACGGCCCCGAAGACGAGCAATACCGCGGCGGGTCC
115 Q P C H R S P A A A P S R S T I A G G L 96
18961 CCGAGGTGGCGCCGCCGGTGC CGCGCCGAGATCACAGGCTTGTGGACTTTGTTCGAGCACA 19020
GGCTCCACCGCGGCGGCCACGCGCGGCTCTAGTGTCCGAACGACCTGAAACAGCTCGTGT
95 G L H R R R H A G L D C A Q Q V K D L V 76
19021 GCGGCTGCCAGTTCGTGCGCAACGGCACCGCCTACCCGGCCCCGCAGGCGCGCGAGCATT 19080
CGCCGACGGTCAAGCACGCGTTGCCGTGGCGGATGGGCCGGGGCGTCCGCGCGCTCGTAA
75 A A A L E H A V A G G V R G R L R A L M 56
19081 TGCAGAAAAAACTGACCTACCTGGAAGGCAAGAAACGGATCAACAGCGCCGAGGACTTTA 19140
ACGTCTTTTTTACTGGATGGACCTTCCGTTCTTTGCCCTAGTTGTTCGCGGCTCCTGAAAT
55 Q L F F Q G V Q F A L F P D V A G L V K 36
19141 TCGAACTGGCCGCCACCCAAAGCAGCATGAGCGGCAAGCCTTACCAAGTGC GTTGCCCAA 19200
AGCTTGACCGGGCGGTGGGTTTCGTCTACTCGCCGTTCCGGAATGGTTCACGCAACGGGTT
35 D F Q G G G L A A H A A L R V L H T A W 16
19201 CGGGCTCGCAACCAGCGGGCACCTGGTTGAACACCGAGTTGCACAGGCAACGCGCTCAGT 19260
GCCCAGCGTGGTTCGCGCCGTGGACCAACTTGTGGCTCAACGTGTCCGTTGCGCGAGTCA
15 R A R L W R A G P Q V G L Q M (ORF16) 1
19261 AGTCAGCGGGTTGACCCGCGCTCGGTGCGCAGTGGCCGAAAGGGTGCCGGGCGCAACAC 19320

19321 CTAGTGGGGCGGCTGCGTACCCCCGCGGGGAGAGCCCGCTTGCTACCAGCAGCCTTTAC 19380

19381 CCAACACCTACGCACGCTGGGTATTTACAGGTCCTTGCTAACCGCCGAGGCCTGGGCAC 19440

19441 TTAGGCGGGGAAGTGCTGGCGGTTACAACACCTAACTAACAGGCCGTGTAGGGCGTTTCT 19500

19501 CTAATTCATGTAGGCCTCATATTTAAACGCTCCACCCCCCTCAATCCTCCCCGCAACTA 19560

19561 GGCACAATGCCGCCGCGCCCTGCTGCCTGGCGCGCTTACATCCCAACGCTGT TAGCCTC 19620

19621 TTTGGCGGGTTATGGCCGACTGGCTGAACTGCGAATTTGCAACCTTATGCCTGGTTTGG 19680

19681 ATGTCACGCCCATGAACCCCACTGCCCCGCTCATCAGCGATCAAACACGCCTGTATTCCCT 19740

19741 TGCAAACGCCACGTCCATGGCCCCCTGCACGTTGAGCGCTGGGCCGGCCGCGAGGCGC 19800

19801 TTTCCGAGCTGTACCCTGGGACGTGCTGGCCCTGAGCGCCGACCATCGCCTACCCCTCG 19860

19861 AGGCGATGCTCGGTGCGCCATTACCTTGCTCACGGCCCTGGCCGATGGCAGCCACAGCC 19920

19921 GACGCAGCGGTTTGTCTAGCCGTGTCGCCCGCCTGGGCAGCGACGGCTCCCTGACCCGCT 19980

19981 ACCACCTGACCCTGGTGCCCTGGCTCTGGGTGTTGAGCCAGAGCTTTCGCTCCCGGGTGT 20040

20041 TCCAAGACAGCACCGTGCAGCAGATCGTCAGCCAGGTCTTCGCCGGTTACGAGGCCTACG 20100

20101 CCCGTTGGCGTTTACCCCTGGACGCCGACAGTTACTGAGCCACGTGCGCCCCGCGCAGTT 20160

20161 ACTGCGTGCAGTACCGCGAAACCGATTTGACTTTGTGTCAGCGGCTACTGGCCGAAGAGG 20220

Appendix. Fig. 2. Continued.

20221 GCCTGGGCTATTGCTTTGTGCGAGGACCCACAAGCACCGGCCGGCCACTGCCTGGTCATCT 20280
 20281 TTGGCCAGAGCCCCGGGCTGCCTGAAGACCTCACCTCGGCCAGCGCTACCAGCGCGCCGA 20340
 20341 CGCGGTGGAGGAGCAAGACACCGTGCAGTTCATGGGCCAGCTCCATCGCCTGAGCCACCG 20400
 20401 GCGCATCACCTGCTGAGCAGCGACTACAAAACCCGCCAGGCCATCAGCGCCTCGCCGCC 20460
 20461 CCGTCGGCCCCGAGCGCGAGGGCCTGGAAAGCTACGACCCGGTGGGGCGATTACGCCTTTG 20520
 20521 CCACCCTGAGCGAGGCCGAGCATTACGTACGTTTGCACAGCGAGGCCCTGCAGGTCGATA 20580
 20581 AGTCCTTCTGGCACGGCCGCGCCAACGTGCGCAGTGCCCCGCGCCGGAACACGCTTTGAGC 20640
 20641 TGTTGCACGCGCCTTGGGTCCACACGGCGCACCTGCGCGGGCAGGACCACTTCTTCA 20700
 20701 CCCAGGTGCACAGTTACGGCATCAACAACCTGGAAAGCGGCGCAGCGCTGTTGGATGCCG 20760
 20761 ACCTAACAACCCGCCTGCAACTCGACCACCTCGACCCCCAGGTGCTGGCCGAAGCCCAGC 20820
 20821 GCAGCGTCCTTTGCCCAGCAGTTCAACGCCGTGGGGCACAACCAACCGTGGCGCCCCGACC 20880
 1 (ORF15) M G H N Q P W R P T 10
 20881 CTGGCCGATGGCACCGGCCAGCGCCTCAACCCGGTGGCCACCGCCCACGGCCCACAAACG 20940
 11 L A D G T G Q R L N P V A T A H G P Q T 30
 20941 GCCATCGTGCCTCGGCGCCGACGGGCAAACCACGGCGGGCAGTGGCAGCCCCTTGCACTGC 21000
 31 A I V V G A D G Q T T A G S G S P L H C 50
 21001 GATGCCATGGGCGGGGTGCGGGTGCCTTCCATTGGCAGGCCGAAGACGACCGTGGCGCC 21060
 51 D A M G R V R V R F H W Q A E D D R G A 70
 21061 TGCTGGGTGCGGGTTCGGCCAAGCCTTGGCGGGCCAGGGGTTCGGCGCGCAGTTCTTGCCC 21120
 71 C W V R V G Q A L A G Q G F G A Q F L P 90
 21121 CGTATCGGCCAGGAAGTGCTGGTGCCTTTCATCAACGGCGACATCGACCGGCCGATCATT 21180
 91 R I G Q E V L V H F I N G D I D R P I I 110
 21181 TTCAAGGCCCTGTACAACGGCCGCGGGCAAGGCGGTGGCCCCACACCGGGCGGCGCC 21240
 111 F K A L Y N G R G E G G V A P T P G G A 130
 21241 CCCGGCAAAGCAACCTGGCGGTGTATGCCAAAGCCAACGACCAACGCCCCAGCGCCGAA 21300
 131 P G Q S N L A V Y A K A N D Q R P S A E 150
 21301 GCCAACCTCGCCGGGGGCAACAGCCCGCGTGGCACGGCGGTGGCGGGCGGTGAAACACAG 21360
 151 A N L A G G N S P A W H G G G G G E T Q 170
 21361 CACCGCAACGCGGGCGGCGCTGTCGGGCTTCAAGACCCAAGGCTTCGATGGCCAAGGCCAT 21420
 171 H R N A A A L S G F K T Q G F D G Q G H 190
 21421 AACCAACTGGTCTTTCGACGACAGCGACCAACAAGGCCGATCCAGATGGCCACCACCCAA 21480
 191 N Q L V F D D S D Q Q G R I Q M A T T Q 210
 21481 GCCCACAGCCAACTCAACCTCGGCCACCTGATCCACCAGGCCGGCAACTACCGGGGCGAGT 21540
 211 A H S Q L N L G H L I H Q A G N Y R G S 230
 21541 TTTGCGGGCAGGGGTTGAACTGCGCACCGACGCCTACGGGGCGGTGCGCGCCGCCAAA 21600
 231 F R G Q G F E L R T D A Y G A V R A A K 250
 21601 GGCTTGCTGCTCAGCACCTACGCCATCGAGCCGCACACCCCGGCGAGGCGATGCCAGCGCC 21660
 251 G L L L S T Y A I E P H T P A G D A S A 270
 21661 GCGCAAGCGCTGCTGCAACAACAACGCGACCTCGCCGCACGCTTCGACCTGGCCGCGCAG 21720
 271 A Q A L L Q Q Q R D L A A R F D L A A Q 290
 21721 ATCCACCACACCGTGCCCTGGCCAGCCACCGCGCGTACACCAGGCCCAACTCAGCGCG 21780
 291 I H H T V P L A S H R G V H Q A Q L S A 310
 21781 CTGATCGCCGATCAAGCACCGCTCCAGGCCCTGCACACCAGCCTCAGCGCCACCGTCACC 21840
 311 L I A D Q A P L Q A L H T S L S A T V T 330

Appendix. Fig. 2. Continued.

21841 GGCGAACGACTCGCCCAGGCCAGCGCCGCGCCTTTAAAGCGCGACAGCCACCCACCCCA 21900
331 G E R L A Q A S A A A F K A R Q P P T P 350
21901 TCCCCACAGCGCCGACCCCTTGCTCACCCCTCAGCGCCGCGACAGCATCGCCAAAATCG 21960
351 S P T A P T P C S P S A P A T A S P K S 370
21961 CCGGCCAATCACTGCAATGGACCGCCGCGCAAACCTCACCCCTGGGCAGCGGGCGCCGACA 22020
371 P A N H C N G P P R K P S P W A A A P T 390
22021 CCCACCTGCCACCGGCCAAACCTGCGCCTGCACAGCGGCCAAGCCATCGCCGCCCTGG 22080
391 P T C P P A K P C A C T A A K P S P P W 410
22081 CCGGCGTCGCAACAGGCCGACGGCATCGGCCCTGGCCCTGATCGCCAGCGCCGGGGCGGT 22140
411 P A S Q Q A D G I G P G P D R Q R R G G 430
22141 GGACATCCAGGCGCAACAGAGCGACATCCACCTCAGCGCCCAAGACGACCTGAAAGATTC 22200
431 G H P G A T E R H P P Q R P R R P E R F 450
22201 ATACTCGATCGAGGCAGCGGCACAAATGGCCGCGGGCACCACCCTGCACCTGGCCGTGGA 22260
451 I L D R G S G T N G R G H H P A P G R G 470
22261 GGGCGGCGCCAGCCTGACCCTGGAAGGCGGCCAACATCACCTTCGCCTGCCCGGGCAGCAT 22320
471 G R R Q P D P G R R Q H H L R L P G Q H 490
22321 CAAGGTGCATGCCGGCAAACGCCAGTTCAGCGGCCCCAGCCGTTCCAGTTTGGCTTGCC 22380
491 Q G A C R Q T P V Q R P Q P F P V W L A 510
22381 CCTGTTTCCACAATCGGTGTGTGTGGCGTGCCTGTAAAGCGGCGCGCAACGGCGCCCC 22440
511 P V S T I G V C G V P V K S G A Q R R P 530
22441 CTATGCGCAGTTGCAGTAATGACCCCGGCCTTTAGACCTTTATGACCCCGCAGGCGGACG 22500
531 L C A V A V M T P A F R P L * 544
22501 GCGCCGGCAGCGACCTGTTGAACTGGATGCAAGCGCAACCCCAAGCGCCTTATTGGGGCGT 22560

22561 TGGTGGATGTGGCGTTACTGGGGCGGGCACCCCTTCAAGGCTGCGGCGCGCCCATCACT 22620
1(ORF14)M W R Y W A R H P S R L R R A A I T 18
22621 GGGCACCCGTCAATGCCCTGGCGAACACGCCCTGCAAGCGTTCGGATGAGGTGGCCCCG 22680
19 G H P S M P W R T R P C K R S D E V A P 38
22681 CATCTGATTTGCCCTGGGGCGACCCCCTCAAGAAGCCAGGCACAAATCGCGCAATTACT 22740
39 H L I C P G R H P S R T P G T N R A I T 58
22741 GGACACTCACCGCAGGCACACCCGCCCTCAGCTTCTTGCGCAGCCGCCTACCCCGCACAA 22800
59 G H S P Q A H P P S A S C A A A Y P A Q 78
22801 GCCCTGCAACTGCTATTTCGGGTACTTGGCCAAGGCCAGGTTTCAGGGCCGGGCAAGGGCC 22860
79 A L Q L L F G Y L A K A Q V Q G R A R A 98
22861 GTGCATCCTGCGCTTTGCCAACAGCCGTATCCTGCCGGGCCTGCTGCAAACCCCTTGCCCC 22920
99 V H P A L C Q Q P Y P A G P A A N P C P 118
22921 CAGCCAACAACGGCGGTAGCCGACATCGTTCAGGACTGGCGAGGCTGTGATCGATACGG 22980
119 Q P T T A R S R H R S G L A R L * 134
22981 CCACGTTTCAGGGCTGGCTTGCCCTGCACGAGGCAGTACCAGCCGCACAGATAGACACCCG 23040

23041 CCCCCACCTGCACCTGTGCCCGGAACAACTGGCTGCGCTGCTGACACAGGCAGAAGCCGA 23100

23101 CGCGGTGTTTTACCAATTAACGCAGCAGCATCCAGAAGTCTCCCCAACAACAACGCGC 23160

23161 AGCCTTCCATCAGCGCCTCGCTACTTTGCCTCTCAACGGCCAGTGAAATGAACCTGTCCA 23220

23221 CCGAACAAGACCGCTGGCTGTTTGTGCTGCTGAGTTTGAACCTACGGCAATGAGTTCTATC 23280

23281 GCTGCCCAGCACTTGCCCCAACCTGGGATGTTATTGGCGAACAAGGCGGAGTTTGCAAG 23340

23341 AACAAACACAAGGGTGGAGCGCTGAAACGTTGCAGGCATTGATGGATTTCAAGCCTCCGC 23400

23401 CAACTGTGCCTGATGGGCCAAACCCCTTTAATTGGAGGCCTGCAAATGAGCGGATTTGA 23460
1 (ORF13) M S G F E 5

Appendix. Fig. 2. Continued.

23461 ATTTGGCGGGTACCGCCGGCTCCGACGGTACAGCGCATTAAATAGTGGTTCGCCGTTACTTG 23520
6 F G G Y R R L R R Y S A L I V V A V T W 25
23521 GTGGGTTGTAATGGGGAAGAACCGGACAAAGGCCCGCGTCCGCCAGTATCAGTGGTTAC 23580
26 W V V M G K N R T K A P R P P V S V V T 45
23581 AACACACCCGAAGATTATATTCATCGATTCTTTATTGATGGGTTCGTGGGGGGGAACGGT 23640
46 T T P K I I F I D S L L M G R G G G T V 65
23641 TTTTGCCTATAGCGGAGGGGGCGGTTTCGTATGCTGTATCAACTATCCTCGCQAATGGCG 23700
66 F A Y S G G G G F V C C I N Y P R E W R 85
23701 GCCTGGCCTCAACAGCCACCGTCCGATGGAGTACTTCAAGTTCATTCCGGATGGCCCTC 23760
86 P G L N S H R P M E Y F K F H S G W P S 105
23761 CAGGAGTCACCTGGCATGAGCTGGTAGTGCCAATTGAGTATTACGATAAGTTGGGCTCCC 23820
106 R S H L A * 110
23821 GTTTAAATGTGCACTTTTGGCCAGGTCACACAGTGCCGGCTGTTGATATGGAAACGCGCT 23880

23881 CCGCCGACTCCCAAGGGCTATCGCGGGCCGGCTGCTCCGATCAAAACCTGCTGACTGGCCT 23940

23941 TACTGACAATAATTACTTAGGAATTAACCTCCATGAATGCGTTTCAAATTGGCAGGCAACA 24000
1 (ORF12) M N A F Q I G R Q H 10
24001 TCTGCGTCAACTATTAAGTTTACTCGTGCTGGCCGTAGTATTGCTTGTGGCTGTAAGGC 24060
11 L R Q L L S L L V L A V V L L V G C K A 30
24061 GGAAGAACCGGATAAAGGCCCTGCGACTGCCAGTATCAGTGCCTATAACCACACCGCTGA 24120
31 E E P D K G P A T A S I S A Y N H T A D 50
24121 CTATATACATGAATTTTATGTAGATGGAGCGTGGGGGGTAACCTCGCGTGCTTATGGCGG 24180
51 Y I H E F Y V D G A W G G N S R A Y G G 70
24181 TGGGGGAGGTTTTGTATGTTGCATTAGCTATCCTCGCGAATGGCGCCCTGACTTAACCGC 24240
71 G G G F V C C I S Y P R E W R P D L T A 90
24241 TACTGTTTCGGTGGAGTACCTCAAGTTC AATACCAAAAACCTTTTCTGGCGAAACGTGGCA 24300
91 T V R W S T S S S I P K T F S G E T W H 110
24301 TGAATTGGTAGTGCCGATTGAGTACTACGATAAGTTGGGTTCACTGAATGTGCACTTTTT 24360
111 E L V V P I E Y Y D K L G S L N V H F L 130
24361 ACCAAATCACCAAGTGC GACTACTGATCTGGAACGGTTCAGCCGGTTCCAAAGGTTATCG 24420
131 P N H Q V R L L I W N G S A G S K G Y R 150
24421 CGGCCAGACGCGCCGGAAAAACCTGCTGACTGGCCTTACTGACAATAATCACTTAGGAA 24480
151 G P D A P E K P A D W P Y * 163
24481 TTAACCTCCATGAATGCGTTTCAAATTGGCAGGCAATACCTGCGTCAACTATTAAGTTTAC 24540
1 (ORF11) M N A F Q I G R Q Y L R Q L L S L L 18
24541 TCGTGCTAGCGGTATTGCTTGTGGCTGTAAGGCGGAAGAGCCGGACAAAGGCCCTGCGA 24600
19 V L A V L L V G C K A E E P D K G P A T 38
24601 CTGCCAGTATCAGTGCCTATAACCACACGGCGGACTACATTCATCAGTTTTATATCGACG 24660
39 A S I S A Y N H T A D Y I H Q F Y I D G 58
24661 GCACTTGGGGAGGAAACGCACGTGCTTATGGCGGGGGGGAGGTTTTGTATGTTGCATTA 24720
59 T W G G N A R A Y G G G G G F V C C I S 78
24721 GCTATCCTCGCGAATGGCGCCCTGACTTAACCGCTACTGTGCGGTGGAGTACTTCAGCTT 24780
79 Y P R E W R P D L T A T V R W S T S A S 98
24781 CAAAGCCCAACCCCTATACGGGAACTACCTGGCACGAAAAAGTAGTGCCGATTGAGTATT 24840
99 K P N P Y T G T T W H E K V V P I E Y Y 118
24841 ACGACAAATTGGGCTCGTTGAATGTACATTTCTTACCCAATAACCAAGTAAGGCTACTGA 24900
119 D K L G S L N V H F L P N N Q V R L L I 138
24901 TCTGGAACGGATCGGCCGGCTCCAAGGGGTATAGAGGGCCAGATGCTCCGGAAAAACCTG 24960
139 W N G S A G S K G Y R G P D A P E K P A 158
24961 CTGGCTGGCCAAAGACTCTTTATCTGGAGGATCTGAAATGACTGAATTTCAACCTGGTTG 25020
159 G W P K T L Y L E D L K * 170
25021 GTACAAGGTCCGACGCAAGCTGGATAACTCCCTGGACTGTGCTACTTCGATACACCGCA 25080

Appendix. Fig. 2. Continued. The DRs are indicated by shaded boxes.

25081 TACCTACAGCGCTTATTGCTGGCTAGAGCCCAATGGCAATCAACGGCTGTGGTGGGATGC 25140

25141 CCGGCAGCACCCACCGGCCCGTGCCATGGCGTGGCAGCCACGTTTCCAACGGGTTGCCGA 25200

25201 CCCACTGTGCGATTGCCTCCTGGGATGCCTATGCCTGCACCCTGGGCAACTCACAGATGGA 25260

25261 GCCGCCGATTCCCTATCTAAGCCAGTCGGCTGGTCAGCGTTGGTTATTAAGGAAACTCTGA 25320

25321 AGAAGACTTCCAGATTTTGGCAAAATACCCGGATTCCACCCGCTGAGTTTCCCCATGAAG 25380
1 (tnpA1) M K 2

25381 CAGATGACCTTCGCTGACGCCGAGTACGCCGGTAAGCGCAAACAAACCCGCAAAGAGTTG 25440
3 Q M T F A D A E Y A G K R K Q T R K E L 22

25441 TTCCTGATCGAGATGGATCGGGTGGTGCCGTGGAAGGGTTTGATTGCCTTGATCGAACCG 25500
23 F L I E M D R V V P W K G L I A L I E P 42

25501 CATTATCCCAAGGGTGAAGGCGGTCGTCGGCCATCCGCTGATGGCGATGCTGCGCGTG 25560
43 H Y P K G E G G R P A Y P L M A M L R V 62

25561 CACCTGATGCAAACTGGTTTCGGTTACAGCGATCCAGCGATGGAAGAGGCGCTGTACGAG 25620
63 H L M Q N W F G Y S D P A M E E A L Y E 82

25621 ACCACCATCCTGCGCCAGTTCGCGGGGCTGAGCTTGGAGCGTATCCCCGATGAAACCACC 25680
83 T T I L R Q F A G L S L E R I P D E T T 102

25681 ATCCTCAACTTCCGTCGTTTGGCTGGAGAAGCACGAACTAGCCGCCGGCATCTTGGCTGTC 25740
103 I L N F R R L L E K H E L A A G I L A V 122

25741 ATCAATGGCTATCTGGGCGACCGAGGGCTGTGCGCTGCGCCAAGGCACGATCGTCGATGCC 25800
123 I N G Y L G D R G L S L R Q G T I V D A 142

25801 ACGCTGATCAATGCGCCGAGTTCGACCAAGAACAAGGACGGTAAGCGCGACCCGAAATG 25860
143 T L I N A P S S T K N K D G K R D P E M 162

25861 CATCAGACCAAGAAGGGTAACCAGTACTACTTCCGGCATGAAGGCGCACATCGGTGTCGAT 25920
163 H Q T K K G N Q Y Y F G M K A H I G V D 182

25921 GATGAGTCGGGCTTGGTGCACAGCGTGGTGGGCACGGCGCCAACGTGGCCGATGTCACC 25980
183 D E S G L V H S V V G T A A N V A D V T 202

25981 CAGGTCGATAAGCTGCTGCACGGCGATGAGAACATGGTCGGTGCCGACGCGGGTTACACC 26040
203 Q V D K L L H G D E N M V G A D A G Y T 222

26041 GCGGTCGAGAAACGCCCTGAACATGAAGGCCGGGAGGTCATCTGGCAGATCGCCGCCCGC 26100
223 G V E K R P E H E G R E V I W Q I A A R 242

26101 CGTAGTACTTACAAGAAGCTTGATAAGCGCAGTGCTTTGTACAAAGCCGCGCGCAAGATT 26160
243 R S T Y K K L D K R S A L Y K A A R K I 262

26161 GAGAAGGCCAAGGCTCAGGTACGAGCCAAGGTCGAGCATCCGTTCCGGGTGATCAAGCGC 26220
263 E K A K A Q V R A K V E H P F R V I K R 282

26221 CAGTTCGGTTATGTGAAGGTGCGCTTCCGTGGTCTGGCGAAGAACACGGCGCAACTGGTG 26280
283 Q F G Y V K V R F R G L A K N T A Q L V 302

26281 ACGCTGTTGCGCTGTCGAACCTGTGGATGGCGCGCCGACATTTACTGGCGAATGCAGGA 26340
303 T L F A L S N L W M A R R H L L A N A G 322

26341 GAGGTGCGCCTGTAATGTGGGAAATGGCTGCCGCGAGGTGCTCGCGGGCTAAAAACAC 26400
323 E V R L * 326

26401 AGAAATAAGCGGGTGATCTGAGCGTTTTTGGATCGATTTGCCGTTTTTAAAATCGGCAGGG 26460

26461 GCTGAAGTCAGCCAGAAATACACAACACTTTCAGACCATCCCTAGCCTGTTAATTCGAGA 26520

26521 CTCGTATCCAGTACTTGGCGTTATCGCCACGGCCAGGGCCGCGCCGGGCGCCAATTT 26580

26581 TATGGCCTGTGTTACCGATAAAGCGGATAGACCCAACCCCCGCAACTTATTTAATGCA 26640

26641 GGCACAGCGTATTTATTGAATGCACGAATTAGCCCCGATGCAACGAGGTATCGATTAATGA 26700

Appendix. Fig. 2. Continued.

26701 GCAGTGCAAGAAGCGTCGAACAATGGAAAACTTTATTGGATGCGCGCCGGGCGCCAGCC 26760
1 (ORF9) M E K L Y W M R A G R Q P 13
26761 GATGATTATTACCCGCGAGGGCGAGGGCCAGCTCAATGCCCTTATCAACGCCTGCCAGCA 26820
14 M I I T R D G E G Q L N A L I N A C Q H 33
26821 CCGCGGCACCACCCTGACCCGCGTGGCAAGGGCAACCAGTCCACCTTCACCTGCCCGTT 26880
34 R G T T L T R V G K G N Q S T F T C P F 53
26881 CCACGCCTGGTGTACAAAAGCGACGGCCGCTGGTCAAGGTCAAAGCCCCGGCGAGTA 26940
54 H A W C Y K S D G R L V K V K A P G E Y 73
26941 CCGGAAGGGTTCGACAAGGCCACCCGTGGCCTGAAAAAGCGCGCATCGAGAGCTACAA 27000
74 P E G F D K A T R G L K K A R I E S Y K 93
27001 GGGCTTTGTGTTTCATCAGCCTCGACGTCGACGGCCAGGACAGCCTCCAGGAATTTTTGGG 27060
94 G F V F I S L D V D G Q D S L Q E F L G 113
27061 CGACGCCAAAGTGTCTTCGACATGATGGTTCGCCCAATCACCCACCGGTGAGCTGGAGGT 27120
114 D A K V F F D M M V A Q S P T G E L E V 133
27121 GCTGCCGGGCAAATCGGCCTACACCTACGACGGCAACTGGAAGCTGAAAACGAAAACGG 27180
134 L P G K S A Y T Y D G N W K L Q N E N G 153
27181 CCTGGACGGTTATCACGTCAGCACCGTGCACCTACAACCTACGTGGCCACCGTGCAGCACCG 27240
154 L D G Y H V S T V H Y N Y V A T V Q H R 173
27241 CCAGCAGGTCAACAGCGAAAAAGGCCTGACCAGCGCCGACACCCTGGACTACAGCAAGCT 27300
174 Q Q V N S E K G L T S A D T L D Y S K L 193
27301 CGGCGCCGGCGACAAGGAAACCGACGACGGCTGGTTCGCTTCCACAACGGCCACAGCGT 27360
194 G A G D K E T D D G W F A F H N G H S V 213
27361 GCTGTTTCAGCGACATGCCCAACCCAGCGTGGCGCCCGGCTACGCCACCATCGGCCGCGC 27420
214 L F S D M P N P S V R P G Y A T I G R A 233
27421 GGAGGAGACAATCGAAATGACTAAAATCACAAGTACTTTTGCATCTATGCAGACGTCAGG 27480
234 E E T I E M T K I T S T F A S M Q T S G 253
27481 CGTTAAGGCAGTACACGTTGCTTTGCTAGGAACCCCTGTAGCGTGCTTTACAACCAACGT 27540
254 V K A V H V A L L G T L V A C F T T N V 273
27541 TGCTGCCAGTGACTTGCCCACTGTAAACCTGGGTTGACGAGTTTTCTCGACGGTGGCGCC 27600
274 A A S D L P T V N L G S T S F L D G A P 293
27601 ACCATCGGGTCCGGGTCTGTATTTCACTCAATATCTACAGCGCTACAGCAGCCATCGGCT 27660
294 P S G P G L Y F T Q Y L Q R Y S S H R L 313
27661 CCTAGATCAAGCTGGCAATAGAGTGCCGCTTCCAAGAACGGATCTACGATTAGACGTATC 27720
314 L D Q A G N R V P L P R T D L R L D V S 333
27721 AATAACGCAGTTTGTATACTTGACAAATCAAGACATTCTGGGCGGGAAGCTAGCTTTCGA 27780
334 I T Q F V Y L T N Q D I L G G K L A F D 353
27781 TTTTCTTCTCCCTTGGGTGGTTCGACAAGCGCATGACGACGGACTGAACAATGCAGTGCT 27840
354 F L L P W V V D K R I D D G L N N A V L 373
27841 TAACGCTGACCGTGGTGTGGCGACCTACATTCGGCCCGGCTATTCAGTGGGCGCCGGT 27900
374 N A D R G V G D L T F G P A I Q W A P V 393
27901 GATGGGTCTAGGGGGCCTCGGTATGCCCATCGTGTGCAATTCCAGTTCCTTCTGCCAAC 27960
394 M G P R G P R Y A H R V E F Q F L L P T 413
27961 GGGTAGCTATGACAGACTCAAGGCGATCAATCCCGGCTCCAACCTTCTGGTCTTTCAATCC 28020
414 G S Y D R L K A I N P G S N F W S F N P 433
28021 TTACTGGGCAGGTACTGTGTGGATGAATCCCAACTGGACCACATCCTGGCGCCTTCACTA 28080
434 Y W A G T V W M N P N W T T S W R L H Y 453
28081 TCTCTGGAATGGAAAAACACCGATCCTGGCCCATCATATGGTCCTGGCGCTACCTCGGT 28140
454 L W N G K N T D P G P S Y G P G A T S V 473
28141 GCAGGCAGGACAGGCAGTGCACGCAACTTCACGACAGAATATGCGCTCACGCCTCAGTT 28200
474 Q A G Q A V H A N F T T E Y A L T P Q F 493
28201 CCGTCTGGGAATTAACGGTTATTGGTTGAATCAAATTACCGACACCAAGATTAATGGTCG 28260
494 R L G I N G Y W L N Q I T D T K I N G R 513
28261 AGATGTTTCCGGTCGTCGTGAGAAGGTCTGGTCAATCGGCCAGGAGCAGTTTTTCAGCCT 28320
514 D V S G R R E K V W S I G P G A V F S L 533

Appendix. Fig. 2. Continued.

28321 TTCTCCTCACGACCATTTGTTCTCCTGAATGCATATTTTCGAACAAGACGCTGAGAATAGGCC 28380
534 S P H D H L F L N A Y F E Q D A E N R P 553
28381 GGAAGGAAAAAGAATGCAGTTACGCTGGGTTTCATCACTTTAAGTAAATGGTGAACCTAGA 28440
554 E G K R M Q L R W V H H F K * 567
28441 CGCCTATGGGCCAAGATCCCAAGGCACGAGTTTATGCAGGCAGCGATTCAATGAAAAAGG 28500

28501 AGATGGACGTGGCGAACGTTGATGAGGCAATTTTAAAAAGAGTAAAAGGCTGGGCGCCCT 28560
M A N V D E A I L K R V K G W A P Y
(carAa start codon)
.....
.....

34621 ATTCATTTTGACGCATTTTTCTAACACCTATGGAGAGTTGCAAGAATATGGAGCCACAAC 34680
I H F D A F F * (carAd stop codon)
1 (carD) M E P Q Q 5
34681 AAATTGACTCGCTCGGAAACGAGCTCTATGAAGCGCTGATTAAGCGTACCCCGCTAAGTC 34740
6 I D S L G N E L Y E A L I K R T P L S P 25
34741 CGCTTAGTTTCGCGAGGTTTTGATATCAGCATTGAAGATGCCTACCAAATCCAGCAACGTA 34800
26 L S S R G F D I S I E D A Y Q I Q Q R M 45
34801 TGACGGCGCTGCGAATTCAAAATGGCGAAAAAATCATCGGACGGAAGATTGGTGTAAACAA 34860
46 T A L R I Q N G E K I I G R K I G V T S 65
34861 GCAAGGCAGTGATGAACATGCTTGGTGTGCACCAACCGGACTTCGGCACCTTGACGGACG 34920
66 K A V M N M L G V H Q P D F G T L T D G 85
34921 GCATGGTTTTGCAGCTCTGCGGAACCAATACGTATGGAAAACCTTATTCAGCCCAAAGCTG 34980
86 M V C S S A E P I R M E N L I Q P K A E 105
34981 AAGGTGAAATTGCCTTCGTCATGAAGAAGGACCTAATGGGGCCGGGTGTCACTGCAGCTG 35040
106 G E I A F V M K K D L M G P G V T A A D 125
35041 ATGTATTGGCGGCAACAGAAGGCGTCATGGCTTGTTCGAGATTGTCGACTCACGTATCA 35100
126 V L A A T E G V M A C F E I V D S R I K 145
35101 AGAACTGGGAAATTAAGATTCAAGACACCGTGGCTGACAATGCATCGTGGCGTTTTTG 35160
146 N W E I K I Q D T V **ADNAS** C G V F V 165
35161 TTCTCGGCGACCGGCTGGTTGACCCCGCAAGTATAGATCTTTTCTTGTGCGGGATGGTCT 35220
166 L G D R L V D P A S I D L S L C G M V L 185
35221 TAGAGAAAAACGGTGAAATAGTGGTCACTGGAGCTGGCGCCGCCACCATGGCCTCACCTG 35280
186 E K N G E I V V T G A G A A T M A S P V 205
35281 TCAACGCTATGGTTTGGCTAGCGAATACTCTAGGACGACTTGGGATTCTCTGAAGGCTG 35340
206 N A M V W L A N T L G R L G I P L K A G 225
35341 GCGAGGTCGTTCTCTCTGCGCATTAGGTGCAATGGTGCCCGTTCAGGCAGGTGACAATT 35400
226 E V V L S G A L G A M V P V Q A G D N L 245
35401 TACGGGTGGCGATTGGCGGCATTGGCGGTTGCACGGTTCGATTTCGTCTAACGAACATTG 35460
246 R V A I G G I G G C T V R F V * 260
35461 ATACAAATATGCCAGCTTCCTATCAAAGAAAGCTGGTGTAGGACGAGCTTCTATCGTC 35520

35521 TAGATTGCTTTGATCAATAGTTACGTTTTAAGGTGGAATGTAATGGATGAAAAACAGAAT 35580
1 (ORF33) M D E K T E F 7
35581 TTGCATGGGACGATTCCCTACTTGTAGGGCACCATGCCATGGATATGGTTCACCGTGAGT 35640
8 A W D D S Y L L G H H A M D M V H R E F 27
35641 TTGTTGACCTGTTGCAGGCAGCCTTGACGGCACCGAATGCAGAACTGCCAACGTGCTTG 35700
28 V D L L Q A A L T A P N A E L P N V L A 47
35701 CCGACATCGCCGTCACGCAGAAGGTCACTTTGCACAGGAGAACGAATGGATGGAGAGCC 35760
48 D I A R H A E G H F A Q E N E W M E S Q 67
35761 AAGGGTTTCCAGGTGGCGACTGCCATATAGACGAGCACAAAGTACTGGCTTCTATCT 35820
68 G F P G G D C H I D E H N K V L A S I Y 87

Appendix. Fig. 2. Continued. The pentapeptide (ADNAS) suggested to have either a catalytic role or structural role in HPD hydratase [Lau et al., 1994] is indicated by shaded box.

35821 ATGACGTGCAGGAAAAAGTAATTC AAGGTGATATTGCAATAGGACGAAAACCTGGCGAAA 35880
88 D V Q E K V I Q G D I A I G R K P G E S 107
35881 GUCCTGTCCGAGTGGTTCCCCGACCATGCGTCTTTTATGGATTCTGCTCTCGCAGCGTGG 35940
108 P V R V V P R P C V F Y G F C S R S V A 127
35941 CTAGTAAAACGTACGTACGGTGGCTCGCCGATTGTTCTACGACGTGCAATTCAAAGATCT 36000
128 S K T Y V R W L A D C S T T C N S K I L 147
36001 TAAAGTGTGCTAGTTGTTACGTAAGAATTATTTGAAAGCAGAAAAGTTTATTTTGTGAA 36060
148 K C A S C S R K N Y F E S R K F I L * 165
36061 TATTAATATGTTCCGATGAATCATAAAGTTAAATAAATGCAAATTTTGTTCGGAAGTCC 36120

36121 ACAAAGTCTTTTGAAAAAAGAACCGATATAGCTCAAAAAGATGAAGAAGGCCATAAG 36180

36181 CCTATTGCGGAGAGCATAGCCAGTGGCGATGGAAACTGCCCCAGTTCAAAAGAGTGGCTC 36240

36241 GTCCTCTACAATGTATCGTTGTTTCTAGAAGCTTTTATAGACTTGCTCAGCGCATTTC 36300

36301 AGGCGGTTAAGAATGCAAAAAGGCTGATCTAAAATTGCGCGAAAATAATTCGCTTTTATTA 36360

36361 ACGGTTTTGGTGTGAGTGTGCAAAAATCATGTATTTGCGGAACATGCGTATTGGTACTCG 36420
1 (ORF34) M Y L R N M R I G T R 11
36421 ACTGTGGACAAGTTTCGGGATAATTCTCGGGATTTTGGTGTGATGGTCAGAATCGGCAA 36480
12 L W T S F G I I L G I L V L M V R I G N 31
36481 TGGTCTTAATGCGAAAAACAAGCAAAAGCTCCTCAATGCATGGAAGTATCCAACAAGAA 36540
32 G L N A K N K Q K L L N A L E V S N K K 51
36541 ATCAATTAGCATTGGTGTAAATGAAGAGCGCGTTGACCGAGAAAAGTGAATTGTCCGCAC 36600
52 S I S I G V M K S A L T E K S V I V R T 71
36601 CATCAGCCGGCAGTCTGACATTGAGAACATTCAAAACATCAACGCGCGGGTTGCCGAAGA 36660
72 I S R Q S D I E N I Q N I N A R V A E E 91
36661 AAATAATCGCTACGCCGTTGCAAATAATCAGTTAACAGCGCTTGGTCTAAATGAACGCGA 36720
92 N N R Y A V A N N Q L T A L G L N E R E 111
36721 GAAGTTCATAATTGACCAAATCAATCGTCTTGACACCGAGATCCTTGGCCATTTTGTATA 36780
112 K F I I D Q I N R L D T E I L G H F D K 131
36781 GGTCACGAAGCAGATCCTTGGCTCTGACGCCGAAGGGGCAGAAAAAACCATTTTACTAA 36840
132 V T K Q I L A S D A E G A E K T I F T N 151
36841 CATTGACGGGTTGGTGC AAAGTGTATTGGCCGAGATGGACAAACTTGTAGCTTTGCAGGA 36900
152 I D G L V Q S V L A E M D K L V A L Q D 171
36901 TTCCTCAGCTTCTGAAATCCTGGCGACCTCAGTGGTAGATGACGATCGTCTCATGGTGT 36960
172 S S A S E I L A T S V V D D D R L M V L 191
36961 AACTGCTTTTATCGGAACGATTTGTTTGGCTTATCGGTGGCGCTTTTCGCTGGATTATTAC 37020
192 T A F I G T I C L L I G G A F A W I I T 211
37021 GCGTAGTATTACTAAACCCCTAAATGCTGGTCTGTTTGAATAACAATTTTTAAGGAGTAC 37080
212 R S I T K P L N A G L F E I Q F L R S T 231
37081 TCAACGTGGCAGATACAGGCAGAGCAAAGGTAGCAATTATCGGGTCGGGTAACATCGGTA 37140
232 Q R G R Y R Q S K G S N Y R V G * 247
1(carF)M A D T G R A K V A I I G S G N I G T 19
37141 CTGACCTCATGATTAAGATCCTGCGTCATGGAAAGTTTTTAGAAATGGGCGCGTTGGTGG 37200
20 D L M I K I L R H G K F L E M G A L V G 39
37201 GCATCGATCCTGCTTCAGATGGCTTAGAGCGCGCAAAGCGCATGGGTGTAGCAACAACAG 37260
40 I D P A S D G L E R A K R M G V A T T A 59
37261 CTGGCGGTATCGAGGGACTTCTTGAGATGCCAGGATTTTCTGATATACGCATTGCCTTTG 37320
60 G G I E G L L E M P G F S D I R I A F D 79
37321 ATGCAACGTCTGCTGGCGCGCATACCGACACAACAAATGCTACAGGAAAAAGGTGTGC 37380
80 A T S A G A H H R H N K L L Q E K G V R 99

37381 GCGTTATTGATTTAACGCCGGCCGCAATTGGTCCTTATGTTATTCCAGCTATTAATTTAG 37440
100 V I D L T P A A I G P Y V I P A I N L D 119
37441 ACGAGCAATTGTCCGCTAAGAATATCAACATGGTGACCTGCGGTGGGCAAGCTACTATTC 37500
120 E Q L S A K N I N M V T C G G Q A T I P 139
37501 CAATTGTGGCGGCGATTTCTAAAGTTGCGAAAGTACATTACGCGGAGATCGTTGCATCTA 37560
140 I V A A I S K V A K V H Y A E I V A S I 159
37561 TTTCTAGTAAATCTGCGGGACCTGGAACGCGCGCAAACATTGACGAGTTCCTGAAACTA 37620
160 S S K S A G P G T R A N I D E F T E T T 179
37621 CTTCGCGGGCGATTGAAGTTCTTGGCGGAGCATCGAAAGGAAAGGCGATTATTGTACTGA 37680
180 S R A I E V L G G A S K G K A I I V L N 199
37681 ATCCTGCAGAACC GCCTCTGATTATGCGTGACACAGTCTTTGCATTGTCTGACATTGCTA 37740
200 P A E P P L I M R D T V F A L S D I A N 219
37741 ATAAATCGGCCGTTGAGGACAGTATTCTGGAAATGGTGGAGAGGGTGAACCTTTATGTCC 37800
220 K S A V E D S I L E M V E R V N S Y V P 239
37801 CAGGGTATCGTTTGAACAAAAAGTCCAGTTCGATATCTTTGACGAGAAAAATGCACTTA 37860
240 G Y R L K Q K V Q F D I F D E K N A L N 259
37861 ACCTACCAGGTATTGGCAGAAAAGCAGGCTTAAAGACCTCTGTGTTCTTGGAGGTTGAGG 37920
260 L P G I G R K A G L K T S V F L E V E G 279
37921 GCGCGGCGCATTACCTACCAGCATAACGCTGGTAACTCGATATTATGACCTCGGCCGCGA 37980
280 A A H Y L P A Y A G N L D I M T S A A M 299
37981 TGGCCTGTGCTGAGCGCATGGCTCAAACACAATTGGTTGAAGAAACGGCGTGAGGCTAAC 38040
300 A C A E R M A Q T Q L V E E T A * 315
1 (carE) M R L T 4
38041 CATGAGCAATAAAAACTTTATATTTCTGACGTGACCTTGCCTGATGGTAGTCATGCGAT 38100
5 M S N K K L Y I S D V T L R D G S H A I 24
38101 TCGTCATCAGTATAGCCTTGACCACGTGCGTAAGATTTCTAAAGCTCTTGACGATGCACG 38160
25 R H Q Y S L D H V R K I S K A L D D A R 44
38161 AGTCGATTCAATCGAAGTAGCCCACGGTGATGGTTTGGCAGGTGCCAGCTTTAATTATGG 38220
45 V D S I E V A H G D G L A G A S F N Y G 64
38221 CTTTGGTGCGCATACGGACGTGGAGTGGATCGCTGCGGTTGCTGAAACGGTTAAGCATGC 38280
65 F G A H T D V E W I A A V A E T V K H A 84
38281 CAAGGTAGCAACCTTACTTCTTCCCGAATTGGTACCCTGCATGACTTGCCTGCCGCATA 38340
85 K V A T L L L P G I G T L H D L R A A Y 104
38341 CGATGCCGGAGCGCGCTAGTACGTGTAGCTACCCATTGTACCGAGGCAGATGTGTGCGAA 38400
105 D A G A R V V R V A T H C T E A D V S K 124
38401 GCAGCATATTGCTGCAGCACGCGAACTTGGTATGGATACCGTTGGCTTCTGATGATGAG 38460
125 Q H I A A A R E L G M D T V G F L M M S 144
38461 CCATATGACAACGCCACAGAAAGTTAGCGGAACAAGCCAAGCTAATGGAAGGCTATGGTGC 38520
145 H M T T P Q K L A E Q A K L M E G Y G A 164
38521 CACATGTATCTATGTCGTCGACTCCGGCGGTGCGCTAACAATGGATGGTGTCCGTGATCG 38580
165 T C I Y V V D S G G A L T M D G V R D R 184
38581 TTTTAGAGCCTTCCGTGAGGTGTTAAAGCCTGAAACAGAGCTGGGTATGCATGCTCATCA 38640
185 F R A F R E V L K P E T E L G M H A H H 204
38641 CAACCTTAGCTTGGGAGTGGCAAACCTCATTGTTGCCGTGGAGGAAGGCTGCGATCGAAT 38700
205 N L S L G V A N S I V A V E E G C D R I 224
38701 CGATGCTAGCTTAGCTGGTATGGGCGCAGGTGCGGGTAATGCGCCACTAGAAGTGTAT 38760
225 D A S L A G M G A G A G N A P L E V F I 244
38761 TGCAGCAGCCGATCGACTCAGTTGGAACCATGGTTGTGATGTATATGGATTGATGGATGC 38820
245 A A A D R L S W N H G C D V Y G L M D A 264
38821 CGCTGATGATATTGTCCGGCCGCTCAGGATCGCCCTGTGAGGGTTGACCGTGAAACGCT 38880
265 A D D I V R P L Q D R P V R V D R E T L 284
38881 GGCTCTCGGTTATGCTGGCGTATACTCAAGCTTCTTCCGGCATGCTGAAGCAGCTTCCGC 38940
285 A L G Y A G V Y S S F L R H A E A A S A 304

38941 AAAATATGGCATTAAAGACGGTTGATATTCTTGTGCGAGCTGGGCCGACGTCGAATGGTAGG 39000
305 K Y G I K T V D I L V E L G R R R M V G 324
39001 TGGTCAAGAGGACATGATTGTCGATGTCGCACTTGATTTAGCCAATAAAATAAACTTAACG 39060
325 G Q E D M I V D V A L D L A N K * 340
39061 CTTTTCGCACCCATAATGTCGTACGCTTTATGGGGTGTTCCTTCTGTTTTGCGGCCG 39120

39121 TTTTGGGAGGGATTTAATTATTCCGAAAAGAAGCCCATAATTAATCCGAAAATGCGGG 39180

39181 TTTAAGTGAGTTTAAACGAATCGAAGCGGGTTGTTCCGGAGGCTTTGAAGGCGGATTTTAT 39240

39241 TTCGGATGGTGGTAAAAAGGGCCATTGACGAAACGATTGGCGAAGCATTGGCGAAACGAT 39300

39301 TGGCGAAACTCCGGCTGCTGCGAGTTCCCGTGCAAACCTTTCACCTAGTCCGTCATTTCA 39360

39361 GTCTGTCAGCCTGGGAGCCAAGGGCCGCAAATCCGAGTATTCCCAAGACACCGACAACCA 39420

39421 GAGCAACGCCACTGATACCGCTCGGAAGGTATTGACTGCCCATAGCTAGTACCAAACCAA 39480

39481 TAAGGCCGGATATCACAGCCGCTTGGTAAGTCGTTCCCTTTTGAGCATTACGGGCCTGCT 39540

39541 GCTGAAAGAAGGCTAACACAGGCCGTCCGCAGTTCAGCACCCCGGTGCATTCCACCAGG 39600

39601 TGCGCTGACCGCACTGAAATGGGCAAGGAACAACATTGGGATCATCCTCACGCATGAAGG 39660

39661 GCGTCGGCGGCTCAACGATCAAATCCCATCTAGCCTTCTACCTGGACGTTGTTGTTCC 39720

39721 CTTCAATTTGTTGATTTCGCCATACATAGCCTCCAATGCCCTGCGAAGATAGCCCACCCAG 39780
GAAGTTAAACAACCTAAGCGGTATGTATCGGAGGTTACGGGACGCTTCTATCGGGTGGGTC
265 * N T S E G Y M A E L A R R L Y G V W 248
39781 TGGCGAAACTCTGCATCCGTCATGTGTCGCCGCGATTCCCTTGCCGGTAACAATCTCGCGC 39840
ACCGCTTTGAGACGTAGGCAGTACACAGCGCGCTAAGGAACGGCCATTGTTAGAGCGCG
247 H R F E A D T M H R R S E K G T V I E R 228
39841 TGCATCGTCGATAGTCGTCATTCACCCAAATTGAGCCAAGTAGTCCGTTTTCGATTTG 39900
ACGTAGCAGCGTATCAGCAGTAAGGTTGGGTTAACTCGGTTTCATCAGGCAAAGCTAAAAC
227 Q M T A Y D D N W G L Q A L Y D T E I K 208
39901 AAGAGATGGCCACGCGTGTGCATGTAGGCTTACCTCTAGCTGCCCTATCAATTCGTCCC 39960
TTCTCTACCGGTGCGCACACGTACATCCGAAGTGGAGATCGACGGGATAGTTAAGCAGGG
207 F L H G R T H M Y A E G R A A R D I R G 188
39961 TGCTGAGTGTTATTGAGCTTCGCTCTAGCCAATGGATGACTTCAGGCGCCATTGATTCG 40020
ACGACTCACAATAACTCGAAGCGGAGATCGGTTACCTACTGAAGTCCGCGGTAACCTAAGC
187 Q Q T N N L K A E L W H I V E P A M S E 168
40021 TCAAATAGCCAGATATCCCTCCAGCTCTTCGGAGCAAGACCAAATGCCTTTGCTAGTTCT 40080
AGTTTATCGGTCTATAGGGAGGTCGAGAAGCCTCGTTCTGGTTTACGGAAACGATCAAGA
167 D F L W I D R W S K P A L G F A K A L E 148
40081 CCGTATAGAGCGCCAAACTTGAACGATTTACCAAGGCGCTTGTGGCGGTATTCGTTAATC 40140
GGCATATCTCGCGGTTTGAACCTGCTAAATGGTTCCGCGAACACCGCCATAAGCAATTAG
147 G Y L A G F K F S K G L R K H R Y E N I 128
40141 TGCTTGATTAGGCTTTCCTATACGTGCCCTGAAAGCGGGTTCGGCCCTATAGATCCAAAT 40200
ACGAACTAATCCGAAAGATATGCACGGGACTTTCGCCCCAGCCGGGGATATCTAGGTTTA
127 Q K I L S E I R A R F A P D A G I S G F 108
40201 GGGGCGGCTATCTCAATTCGCGGCTTTGCTGGCGTGAACCGTCACGCGACCAACCGCA 40260
CCCCGCCGATAGAGTTAAGAGCCGCGAAACGACCGCACTTGGCAGTGCCTGGTTGGCGT
107 P A A I E I R P A K S A H V T V R G V A 88

Appendix. Fig. 2. Continued.

40261 CCGCCAATCTGCACGTTACCGCTGCCCTTGATTGACTGGTGAACCACCCTTTCCCAGCG 40320
GGCGGTTAGACGTGCAATGGCGACGGGAACTAAGTACCACCTGGTGGGAAAGGGGTCGC
87 G G I Q V N G S G K I S Q H V V R E G A 68
40321 ACCTGCGTATTGTTATCGCCCTCTACGGACTGCGATGGAAGAGACTTTCCCTTCGTGTC 40380
TGGACGCATAACAATAGCGGGAGATGCCTGACGCTACCTTCTCTGAAAGGGGAAGCACAG
67 V Q T N N D G E V S Q S P L S K G R R T 48
40381 TTGTTGGAAGGTAACGCAGGTGCATCCTGGATCATCTTCACAGCATCCAGAAAAGCAGCG 40440
AACAACTTCCATTGCGTCCACGTAGGACCTAGTAGAAGTGTCGTAGGTCTTTTCGTCGC
47 K N S P L A P A D Q I M K V A D L F A A 28
40441 CCTGGTGAAGGTGTGTCAGGCTTCTTCGACGACCGGCCTACTACAGGAAAGCGAATGATG 40500
GGACCACTTCCACACAGTCCGAAGAAGCTGCTGGCCGGATGATGTCCTTTTCGCTTACTAC
27 G P S P T D P K K S S R G V V P F R I I 8
40501 TTCGACTTTGTATTTGTCATGGTTTCTCCTACGCTGCGAGCTCGATAACCTGACGGATCG 40560
AAGCTGAAACATAAACAGTACCAAAGAGGATGCGACGCTCGAGCTATTGGACTGCCTAGC
7 N S K T N T M (ORF37) 1
40561 TGTCTGAATCAATGTGCTGCTGGGAGGAAAAAGAGCGGTACAACATCACGATGGCGCGGG 40620

40621 CTTTCTTTCTGCGGAAAGACTCTTATGCATCTTCGTCAGCAATAACTCCATGCCCTCAA 40680

40681 TAATCTGCGAGAGCAGCTCTTCGTCGATTGGTGTGTAGATGCGGCGTGAGCCTCTTCCA 40740

40741 TCCGGTGCCTGATAGAACATAAAGAACATCGAGCCCTTCTTCATGAGCGTTTCTAGGT 40800

40801 AGAGAGCATCGGGGAACCGCTCGCCATTCTCGTAGTTGAACTGGGTTTTCCGGCTTACTC 40860

40861 CCGTCAGCGCGCACATCTCAAGCTGCGAGAGACCCAAGCGGTGCGGCTCACTCTTCAGTC 40920

40921 TTTCCCAATGGTTTGATAATAAATTACCCATAATTTGGTAAATAGATACGCTATCGCT 40980

40981 TGACAATGTATAAATTTTTACCCATCATAGCAACAAATAACCCATTCTTACGCATTCTAC 41040

41041 TCTAGGTGACGACATGAAAACGCTCAGAACCCCCAACAAGTGTGGCGGACTTCGAACA 41100

41101 TCGCGGAATAACGATTTCCGGCTGGGCAAGGTCTCATGGTTTCTCGCGTGATGTTGTGCG 41160

41161 CTCTGTTTTATACGGGCGGTCAAAGGCCTCTACGGCTCCGGGCACAAGGTTGCGGTGCT 41220

41221 GTTGGGCCTTAAAGATGGCGTCATCGACGAACTGAAATAGGCGGTTGTAGCTATACCAGA 41280

41281 GATAACCTCAACAAATAGACCGCCATGAACCTGTCCGTACCGACCCCAAACACCTCGCGA 41340

41341 GCTCCCGCTATATCCGTATTAATGCGGTGAACAGCGATTGGCTGACCGCCGATGATTTTC 41400

41401 GCGAGGTTGGCCGGGATTGATATCCGCGCTGGTCAAATTGCGCTGCGAAAAGCTGTAGAT 41460

41461 GGCAAACCTTGGCGCGGATGCTATTTAGGGAACTCTGAAGAAGACTTCCAGATTTTGGC 41520

41521 AAAATACCCGGATTCCACCCGCTGAGTTTCCCCATGAAGCAGATGACCTTCGCTGACGCC 41580
1 (tnpA4) M K Q M T F A D A 9
41581 GAGTACGCCGGTAAGCGCAAACAAACCCGCAAAGAGTTGTTCTGATCGAGATGGATCGG 41640
10 E Y A G K R K Q T R K E L F L I E M D R 29
41641 GTGGTGCCGTGGAAGGGTTTGATTGCCTTGATCGAACCGCATTATCCCAAGGGTGAAGGC 41700
30 V V P W K G L I A L I E P H Y P K G E G 49

Appendix. Fig. 2. Continued.

41701 GGTCGTCCGGCCTATCCGCTGATGGCGATGCTGCGCGTGCACCTGATGCAAACTGGTTC 41760
50 G R P A Y P L M A M L R V H L M Q N W F 69
41761 GGTTACAGCGATCCAGCGATGGAAGAGGCGCTGTACGAGACCACCATCCTGCGCCAGTTC 41820
70 G Y S D P A M E E A L Y E T T I L R Q F 89
41821 GCGGGGCTGAGCTTGGAGCGTATCCCCGATGAAACCACCATCCTCAACTTCCGTCGTTTG 41880
90 A G L S L E R I P D E T T I L N F R R L 109
41881 CTGGAGAAGCACGAACTAGCCCGCGCATCTTGGCTGTCATCAATGGCTATCTGGGCGAC 41940
110 L E K H E L A A G I L A V I N G Y L G D 129
41941 CGAGGGCTGTGCTGCGCCAAGGCACGATCGTCGATGCCACGCTGATCAATGCGCCGAGT 42000
130 R G L S L R Q G T I V D A T L I N A P S 149
42001 TCGACCAAGAACAAGGACGGTAAGCGCGACCCGAAATGCATCAGACCAAGAAGGGTAAC 42060
150 S T K N K D G K R D P E M H Q T K K G N 169
42061 CAGTACTACTTCGGCATGAAGGCGCACATCGGTGTCGATGATGAGTCGGGCTTGGTGCAC 42120
170 Q Y Y F G M K A H I G V D D E S G L V H 189
42121 AGCGTGGTGGGCACGGCGGCAACGTGGCCGATGTCACCCAGGTCGATAAGCTGCTGCAC 42180
190 S V V G T A A N V A D V T Q V D K L L H 209
42181 GGCGATGAGAACATGGTCCGGTCCGACGCGGGTTACACCGCGTCGAGAAACGCCCTGAA 42240
210 G D E N M V G A D A G Y T G V E K R P E 229
42241 CATGAAGGCCGGGAGGTCATCTGGCAGATCGCCGCCCGCGTAGTACTTACAAGAAGCTT 42300
230 H E G R E V I W Q I A A R R S T Y K K L 249
42301 GATAAGCGCAGTGCTTTGTACAAAGCCGCGCAAGATTGAGAAGGCCAAGGCTCAGGTA 42360
250 D K R S A L Y K A A R K I E K A K A Q V 269
42361 CGAGCCAAGGTCGAGCATCCGTTCCGGTGATCAAGCGCCAGTTCGGTTATGTGAAGGTG 42420
270 R A K V E H P F R V I K R Q F G Y V K V 289
42421 CGCTTCCGTTGGTCTGGCGAAGAACACGGCGCAACTGGTGACGCTGTTCCGCGCTGTCGAAC 42480
290 R F R G L A K N T A Q L V T L F A L S N 309
42481 CTGTGGATGGCGCGCCGACATTTACTGGCGAATGCAGGAGAGGTGCGCCTGTAATGTGGG 42540
310 L W M A R R H L L A N A G E V R L * 326
42541 AAATGGCTGCCGCGAGGTGCTCGCGGCGCTAAAAACACAGAAATAAGCGGGTGATCTGA 42600

42601 GCGTTTTTGATCGATTTGCCGCTTTTAAAATCGGCAGGGGCTGAAGTCAGCCAGAAATAC 42660

42661 ACAACTACTTCAGACCATCCTTAGCTGCCCTTTCCGCCGTCTACATCATCGATGGCATAG 42720

42721 GCATTAATATTTCCGCCAGTATCGGTGTGACGCTCTATCCAAACGATAATGAGAATGCCG 42780
1 (ORF39) M T L Y P N D N E N A D 12
42781 ATACACTGCTACGTCATGCCGATCAGGCTATGTATAAGGCCAAACAAAGTGGCCGTAACC 42840
13 T L L R H A D Q A M Y K A K Q S G R N R 32
42841 GTTTCCATTTGTTTGTATGATCAAAAGATAAGATGGATAAATCAGCTTTTGTATCCGTAA 42900
33 F H L F D V S K D K M D K S A F D T V I 52
42901 TACGGGTTCCGCAAGCGTTACATGATGGTGAGTTATGCTTGCATTATCAACCAAAAATAA 42960
53 R V R Q A L H D G E L C L H Y Q P K I S 72
42961 GTTTGAGCAGTGGGGCAGTGATTGGTTTTGAAGCGCTATTGCGTTGGCAGCATCCGCGAG 43020
73 L S S G A V I G F E A L L R W Q H P R D 92
43021 ATGGACTTATATTACCGCAGTATTTTATACCGTTGATTGAGCAGAGCGATCTGATCATGG 43080
93 G L I L P Q Y F I P L I E Q S D L I M E 112
43081 AGATCGGCGAGTGGGTCATAGATCAGGCGCTCAGTCAAATAGAACAATGGGCGGACTTGG 43140
113 I G E W V I D Q A L S Q I E Q W A D L G 132
43141 GTCATTCGTGGTCTGTGAGTGTAACATTTTCAGCATTACATTTCAAAAAAGAAAACTTTG 43200
133 H S W S V S V N I S A L H F K K E N F V 152
43201 TAAAGTCCCTAAAGTTTTTGTGCTGGACAGCCACCCTAATGTTTTGCCGAGATGTTGGATA 43260
153 K S L K F L L D S H P N V L P Q M L D I 172
43261 TTGAGATTACCGAATCAGTGGTGATAGAACAATTATCACATGTCACCTCAGTGCCTCATTG 43320
173 E I T E S V V I E H L S H V T Q C L I A 192

Appendix. Fig. 2. Continued.

43321 CTTGCCAGGATCTGGGTGTGACATTTTCGTTGGACGACTTTGGAACCGGGTATTCATCGC 43380
193 C Q D L G V T F S L D D F G T G Y S S L 212
43381 TGAGCTACCTTAAACAACCTGCCACGCAGTCAATCAAGATAGACAAATCTTTTATACGCG 43440
213 S Y L K Q L P T Q S I K I D K S F I R D 232
43441 ACATACTCATCGATAAAGACAGCTTGGGGCTGACTAAGGCGATCATCGGCCTAGCGAAAT 43500
233 I L I D K D S L G L T K A I I G L A K S 252
43501 CGTTTAACCGCGAGGTTATTGCTGAAGGGGTGAAACGGTTGAGCACGAAGGTGTTGCTG 43560
253 F N R E V I A E G V E T V E H E G V A D 272
43561 ATGAACCTAGGTTGTGATGTAGCGCAAGGCTATGGTATTGCGAAACCCATGCCGGTAGAA 43620
273 E P R L * 276

43621 AAAGTGCTATCTTGGGTGGCGAAGTTTGTCCCTGCTCATTATCTAGAGGGGGTAATAGA 43680

43681 TAAACCGCACCGTCCACGAACCTGATTGCGTCGTCGCAACCTAGTGGATTGAGCGGTTAAT 43740

43741 TCATTTATCCACCGGTAACGATATACTGTGCGTCCAATGCCATGGATTGGTTATGCGCAT 43800

43801 GGCTCGACCGACTTTCCCCATTCGCTCGAAATATATTTTTTATGATACTATCCCAGCT 43860

43861 GTTATTTTCTGGAACAATGAAATAATTAATATTATCATAAGCAACAAATCGATAATAATA 43920

43921 AAGAGGATCATATATGTCCGATCAAGAGTTAATACTAGAGCAACTCCGTAACGTGAATTA 43980
1 (ORF40) M S D Q E L I L E Q L R N V N Y 16
43981 TGATAAGTTAAGCAAACAAGATATCGTAACAATGCTAGATGCAGGGGAACAGACTCTAGA 44040
17 D K L S K Q D I V T M L D A G E Q T L E 36
44041 AAATATGGCAGCATCAAGAAAACTGCCATTGATATAATAGCAGAGGCAATGAACCAAAC 44100
37 N M A A S R K T A I D I I A E A M N Q T 56
44101 AAAGAATACGCTTGAGGCACTTGATTATTCTGGAAAGGTAGAAGTTGAAATCAGGAAAAA 44160
57 K N T L E A L D Y S G K V E V E I R K N 76
44161 TCTCAGCATACTGAGAGAGCTACTGATGACTGGCGATCTATAGTCCCCTAGACTGGGCGA 44220
77 L S I L R E L L M T G D L * 89
44221 ATTAGCACAGGAGGTTGGCCTTGATTAGGGGAGCCTACAGAATTC 44266

| | | |
|------|--|---------|
| 1 | GAATTCATGCTCGACATGAACAGGTGGAGAACTACACGCTCTTCCCCTATCTTGAACCC | 60 |
| 61 | CTGGCGCCGGGGGCGCACACGGGCCTTTCTACGAACTGCGCGAGTACAACCTGGATCCCT | 120 |
| 121 | TCGGGCCTGACCCCCACCCTGGAAGGCTGGAAGAAGGCCGTGGGCCCCGCGCACCGGGGAA | 180 |
| 181 | GGTTATTCACAGGTCTACGCCGCCTTCTACGCCACCGACGGGCGTACCCCGCGCTACCTG | 240 |
| 241 | CACATCTGGCCCTACCGGAGCCTGGAACAACGCCTGGACGTGCGCACGCGAGCCGTGAG | 300 |
| 301 | GACGGCGTCTGGCCCCGGAAAACCCGGCCCCGAGCTGCGTGAAATGCATTCACCGCC | 360 |
| 361 | TACCTGCCGGCGAAGTTCTCGCCGCTGCGCTAGGCCGGTAACCCTTTGGTCTTTTCCTCG | 420 |
| 421 | GCCCGTGCCACTCCTTTGCGGGCCTTTTTTATTGCGACATACGGGCGGCATTTACCGT | 480 |
| 481 | AGGGGCGAATTCATTCGCCAAGCCGTTTCGTAGGATGGGTTGAGCGAAGCGATACCCATGC | 540 |
| 541 | GGGCCACGGTTGATGGGTATCGCAAGCTCAACCCATCCTACAAGAGCAGATTGCTACAGC | 600 |
| | CCCGGTGCCAACTACCCATAGCGTTTCGAGTTGGGTAGGATGTTCTCGTCTAACGATGTGC | |
| 293 | | * L 293 |
| 601 | CGCCCCCGCGACAACCTCTCGATCAGCGCCAGGCAATGGGTCAGCGCCGGGGAAGCGTCA | 660 |
| | GCGGGGGCGCTGTTGAGGAGCTAGTCGCGGTCCGTTACCCAGTCGCGGCCCTTCGCAGT | |
| 292 | R G R S L E E I L A L C H T L A P S A D | 273 |
| 661 | CCGGTGGCGGGCTGATGATGATGGGCGAGGTGGCCTGGGCCTCCTGCAGCTTGGCGTAG | 720 |
| | GGCCACGCCCGGACTACTACTACCCGCTCCACCGGACCCGGAGGACGTCGAACCGCATC | |
| 272 | G T R R S I I I P S T A Q A E Q L K A Y | 253 |
| 721 | CCGATGTCGGCGCGGTGCAGCACCTGCACCGAGGCCGGCACCAGGGTGATGCCGATGCCG | 780 |
| | GGCTACAGCCGCGCCACGTCGTGGACGTGGCTCCGGCCGTGGTCCCCTACGGCTACGGC | |
| 252 | G I D A R H L V Q V S A P V L T I G I G | 233 |
| 781 | GCGGCCACCAGGCCGATGGCGGTCTGCAGTTCGTTGGTCCACTGCGCCACCTCGATCTGC | 840 |
| | CGCCGGTGGTCCGGCTACCGCCAGACGTCAAGCAACCAGGTGACGCGGTGGAGCTAGACG | |
| 232 | A A V L G I A T Q L E N T W Q A V E I Q | 213 |
| 841 | AGGCCATAGGCCGCGAACAGCGCCAGCACGTGGTTCGGCATAGCTCGGCCGGGGGTTGCC | 900 |
| | TCCGGTATCCGCCGCTTGTGCGGTTCGTGCACCAGCCGTATCGAGCCGGCCCCAACGGG | |
| 212 | L G Y A A F L A L V H D A Y S P R P N G | 193 |
| 901 | GGGTAGAGCACGAAGCTCTCCTGGGCAAGGTGGTGAGGCTCACCGGGCCGGCAGCAGG | 960 |
| | CCCATCTCGTGCTTCGAGAGGACCCGTTCCAGCCACTCCGAGTGGCCCCGGGCGTCTGTC | |
| 192 | P Y L V F S E Q A L D T L S V P G P L L | 173 |
| 961 | GGATGCCCCGGCGGGCAGGGCGGCCACCAGCGGGTTCGAGGGTCAGCACCTGCTGCTGGATC | 1020 |
| | CCTACGGGCCGCCCCTCCCGCCGGTGGTTCGCCAGCTCCCAGTCGTGGACGACGACCTAG | |
| 172 | P H G A P L A A V L P D L T L V Q Q Q I | 153 |
| 1021 | GCCGGGTCGTTGATGTGGATGCGGCCGAAGCCGACGTCGATGCGTCCGGCCTTGAGGGAC | 1080 |
| | CGGCCAGCAACTACACCTACGCCGGCTTCGGCTGCAGCTACGCAGGCCGGAACCTCCCTG | |
| 152 | A P D N I H I R G F G V D I R G A K L S | 133 |
| 1081 | TCCACCTGTTGCAGGGTGGTCATTTCCGACAGGCCAGCTCCAGGCCGGCGTCTGCTGCGC | 1140 |
| | AGGTGGACAACGTCCACCAGTAAAGGCTGTCCGGGTCGAGGTCCGGCCGACGCGACGCG | |
| 132 | E V Q Q L T T M E S L G L E L G A D S R | 113 |
| 1141 | AGGCGGCGTATCAGATCGGGCAGTACGCCGTAGAGGGTGGAGGGGGCGAAGCCGATGCCG | 1200 |
| | TCCGCCGATAGTCTAGCCCGTCATGCGGCATCTCCACCTCCCCCGCTTCGGCTACGGC | |
| 112 | L R R I L D P L V G Y L T S P A F G I G | 93 |

Appendix. Fig. 3. The nucleotide and deduced amino acid sequences of the 8,918-bp *EcoRI-NotI* DNA region (accession number AB047272). The sense DNA strand is shown. The putative SD sequences are indicated by double lines under the sequence. Astarisks indicate stop codons.

1201 AGCCAGGCTTTTTCCCCGAACCAATACGCCGGGTATTGTACAAAACCTTGGTCAGCTGC 1260
TCGGTCCGAAAAAGGGGGCTTGGTTATGCGGCCATAACAGTGTGGAAACCAGTCGACG
92 L W A K E G S G I R R T N D C V K T L Q 73
1261 TCGAGGAGGGTGTCTGGAATGTTTCGTGGAAAAAGCGTCCGGCTTCGGTAAGCCGAGCGGT 1320
AGCTCCTCCCACGACCTTACAAGCACCTTTTTTCGAGGCCGAAGCCATTCGGCGTCGCCA
72 E L L T S S H E H F F R G A E T L R L P 53
1321 CGGCCACGTTCCAGCAGGGCCACCCCGAGTTCGTCTCGAGCTGCTGGATCTGCCGGCTC 1380
GCCGGTGCAAGGTCGTCCCGGTGGGGCTCAAGCAGGAGCTCGACGACCTAGACGGCCGAG
52 R G R E L L A V G L E D E L Q Q I Q R S 33
1381 AACGGGGGCTGCGCGATGTGCAGGCGTTCGCGGGCGGGGTGAAATTGAGGGTCTCGGCG 1440
TTGCCCCCGACGCGCTACACGTCCGCAAGGCGCCGCGCCACTTTAACTCCCAGAGCCGC
32 L P P Q A I H L R E A A R T F N L T E A 13
1441 AGCGCCTGGAAGTAGCGCAGGTGGCGAAGCTCCATGGTACTCCGGGGTGGTGGTGTTCG 1500
TCGCGGACCTTCATCGCGTCCACCCTTCGAGGTACCACTGAGGCCCCACCACCACAAGC
12 L A Q F Y R L H R L E M (catR) 1
1501 GGGTTCATCGATACCTTGCGGGTATCGTGCTAGACAAACTCTATATTGGACCCACGGTT 1560

1561 TGTCGCGGCCACTATCCGATACCAGAACAAGCAAACCTGATGGGTATTGAAGTGAGCCAA 1620
1 (catB) M S Q 3
1621 GTCCTGATCGAAAAGCGTTCGAGACGATCATCGTTCGACCTGCCGACCATCCGCCCGCACAAG 1680
4 V L I E S V E T I I V D L P T I R P H K 23
1681 CTGGCCATGCACACCATGCAGCACCAGACGCTGGTGGTGTATCCGCCTGCGCTGCTCCGAT 1740
24 L A M H T M Q H Q T L V V I R L R C S D 43
1741 GGCATCGAAGGTCTCGGCGAGTCCACCACCATCGGTGGCCTGGCCTACGGCAACGAGAGC 1800
44 G I E G L G E S T T I G G L A Y G N E S 63
1801 CCGGAAAGCATCAAGCAGAACATCGACAGCCACCTGGCTCCGCTGCTGATCGGCCAGGAC 1860
64 P E S I K Q N I D S H L A P L L I G Q D 83
1861 GCCGCAATATCAACGCCGCCATGCTGCGCCTGGACAAGGCCCAAGGGCAACACCTTC 1920
84 A A N I N A A M L R L D K A A K G N T F 103
1921 GCCAAGTCCGGCCTGAAAAGCGCCCTGCTCGACGCCAGGGCAAGCGCCTTGGCCTGCCG 1980
104 A K S G L E S A L L D A Q G K R L G L P 123
1981 GTCAGCGAGCTGCTCGGCGGGCCGCGTCCGTGACGGCCTGGAAGTGGCCTGGACCCTGGCC 2040
124 V S E L L G G R V R D G L E V A W T L A 143
2041 AGCGGCGACACCGCCAAGGACATCGCCGAAGCCCAGCACATGCTGGAATCCGCCGCCAC 2100
144 S G D T A K D I A E A Q H M L E I R R H 163
2101 CGCATCTTCAAGCTGAAGATCGGCGCCGGTGAAGTGGACCGGACCTCAAGCACGTGATC 2160
164 R I F L I G A G E V D R D L K H V I 183
2161 GCCATCAAGCAGGCCCTGGGCGACACCGCCAGCGTGC GCGTGGACGTC AACCAGGCCTGG 2220
184 A I K Q A L G D T A S V R V D V N Q A W 203
2221 GACGAATCCGTTGCCCTGCGCGCCTGCCGCATCCTCGGCGACAACGGCATCGACCTGATC 2280
204 D E S V A L R A C R I L G D N G I D L I 223
2281 GAGCAGCCCATCTCGCGCATCAACCGTTCGGGCCAGGTACGCCTGAACCAGCGCAGCCCC 2340
224 E Q P I S R I N R S G Q V R L N Q R S P 243
2341 GCGCCGATCATGGCCGACGAATCCATCGAAAAGCGTCAAGATGCCTTCAGCCTGGCCGCC 2400
244 A P I M A D E S I E S V E D A F S L A A 263
2401 GACGGCGCCGCCAGCATCTTCGCCCTGAAAATCGCCAAGAACGGCGGCCCGGTGCCGTG 2460
264 D G A A S I F A L K I A K N G G P R A V 283
2461 CTGCGCACCGCGCAGATCGCCGAGGCGGGCGGTATCGCCCTTTACGGCGGTACCATGCTG 2520
284 L R T A Q I A E A A G I A L Y G G T M L 303
2521 GAAGGTTCCATCGGCACCCTGGCCTCGGCCACGCCTTCGTGACCCTCAAGCAACTGACC 2580
304 E G S I G T L A S A H A F V T L K Q L T 323

Appendix. Fig. 3. Continued. The Asp (D) and Glu (E) residues involved in manganese coordination are indicated by black boxes with white character, and the Lys (K) and Glu (E) residues directly involved in the enzyme mechanism are indicated by shaded boxes (CatB).

2581 TGGCACACCGAGCTGTTCCGGCCCGCTGCTGCTGACCGAGGAAATCGTCACCGAAGCGCCG 2640
324 W H T L F G P L L L T E E I V T E A P 343
2641 GTCTACCGCGACTTCCAGCTGCACGTGCCGCGTACCCCGGGCCTTGGCCTGGCCCTGGAC 2700
344 V Y R D F Q L H V P R T P G L G L A L D 363
2701 GAAGAGCGCCTGGCGTTCTTCCGCCGCAAGTAAACAGGAGATCGATCCCATGCTGTTCCA 2760
364 E E R L A F F R R K * 373
1 (catC) M L F H 4
2761 CGTGAAGATGACCGTAAGACTTCCGGTTCGACATGGACCCGGCCAAAGCCGCCAAGCTCAA 2820
5 V K M T V R L P V D M D P A K A A K L K 24
2821 GGCCGACGAGAAGGAAGTGGCCAGCGCCTGCAGCGCGAGGGCAAGTGGCGCCACCTGTG 2880
25 A D K E L A Q R L Q R E G K W R H L W 44
2881 GCGCATCGCCGGTCACTACGCCAACTACAGCGTGTTCGACCTCGAAAGCGTCGAGGCACT 2940
45 R I A G H Y A N Y S V F D L E S V E A L 64
2941 GCACGACACCCTGATGCAGCTGCCGCTGTTCCCTACATGGAGATCGAGGTGGACGGTCT 3000
65 H D T L M Q L P L F P Y M E I E V D G L 84
3001 GTGCCGCCACCCGTCGTCGATTACGAAGACGATCGTTGATCGTTTCGCAACCGATAAAA 3060
85 C R P S I H E D D R * 96
3061 ACAAGATGAGGAAATAGCCATGACTATCAAACCTGTCCAGCACCGAAAGCGTTCAAAAGTT 3120
1 (catA) M T I K L S S T E S V Q K F 14
3121 CTTCCAGGAAGCCAGCGGCTTCAACAACGACCAGGGCAGCCAGCGCATGAAGAAGCTGGT 3180
15 F Q E A S G F N N D Q G S Q R M K K L V 34
3181 CAACCGCATCCTGCTCGACACCGTGAAGATCATCGAAGACCTGGAAGTCACCACCGAAGA 3240
35 N R I L L D T V K I I E D L E V T T E E 54
3241 ATTCTGGAAGGCGGTTCGATTACCTCAACCGCATGGGTGCCCGTGGCGAAGCCGGCCTGCT 3300
55 F W K A V D Y L N R M G A R G E A G L L 74
3301 GGTAGCGGGCCTGGGCCTGGAGCACTACCTCGACCTGCTGCTGGATGCCAGGACGAGGC 3360
75 V A G L G L E H Y L D L L L D A Q D E A 94
3361 GGTCGGCCTGTCCGGCGGCACCCACGCACCATCGAAGGCCCGCTGTACGTGGCCGGCGC 3420
95 V G L S G G T P R T I E G P L Y V A G A 114
3421 CCCGCTGTCCGAAGGCGAGACCCGCATGGACGACGGCACTGACGCCGGCACCGTGATGTT 3480
115 P L S E G E T R M D D G T D A G T V M F 134
3481 CCTTGAAGGCCAGGTGGTTCGATGCAGACGGCAAGCCGATCGCCGGCGCTACCGTCGACCT 3540
135 L E G Q V V D A D G K P I A G A T V D L 154
3541 GTGGCACGCCAACACCAAGGGCACCTACTCCTACTTCGACACCACCCAGTCGGACTACAA 3600
155 W H A N T K G T S Y F D T T Q S D Y N 174
3601 CCTGCGCCGCGCATCCTCACCGATGCCAACGGCCGCTACCGCGCCCGCAGCATCGTGCC 3660
175 L R R R I L T D A N G R Y R A R S I V P 194
3661 GTCCGGCTACGGCTGCCCGCCGGATGGCACCACCCAGGAAGTGTGAATCTGCTCGGTGC 3720
195 S G G C P P D G T T Q E V L N L L G R 214
3721 CCACGGCAACCGTCCGGCCACATCCACTTCTTCATCTCCGCGCCGGGCTACCGTCACCT 3780
215 H G N R P A I F F I S A P G Y R H L 234
3781 GACCACCCAGATCAACCTGTCCGGCGACGAGTACCTGTGGGACGACTTCGCTACGCCAC 3840
235 T T Q I N L S G D E Y L W D D F A Y A T 254
3841 CCGCATGGCCTGGTTCGGTGAAGTGCCTTACCAGAACGCGCCCGCCGCAAGGCCCG 3900
255 R D G L V G E V R F T E D A A A A K A R 274
3901 TGGTGTGGACGGCAGCCGCTTCGCCGAGATCGAGTTCGGTTTCCAGCTGCAGAAGCCCC 3960
275 G V D G S R F A E I E F G F Q L Q K A P 294
3961 GACCGCCGAGCAGGAAGCCATTCGCGACCGGTACGCGCAGAAGCCTGAGTTGTGATTGT 4020
295 T A E Q E A I R D R V R A E A * 309
4021 GGCGGGGGAGGCTGTTCTCCGCCATTTGAGCCGGCGTTTGGCCGGCTTTTTTATGGG 4080

Appendix. Fig. 3. Continued. The amino acid residues which were reported to contribute to the active site [Katti et al., 1989] are indicated by shaded boxes (CatC). The His (H) and Tyr (Y) residues reported to be involved in iron binding [Hartnett et al., 1990] are indicated by shaded boxes (CatA).

4081 CGGGGTATGGGGGAGGACACTCCTGTAGGGGCGAATTCATTCGCCAAGGGCAACGCAGTT 4140
4141 GCCCCTGAGGTTTCCCCGGGCAGACCTTCGGCCTGCTTGGCGAATGAATTCGCCCCACA 4200
4201 AGTAGCCTGCCAGCGTTGCCCAACTCCCCGGATAGCCATCCCCGACTATCCGGATTTC 4260
4261 CCGCGCCACCCTTTGCCGCTAATCCACGGACGACCAACCGGAGGTGCCCATGACCCGTC 4320
1 (ORF5) M T R P 4
4321 CCAACGAACCGGCCGAACCTGGCCGCCTACTACGACGTCCAGTACAACGCCCGGAGAGCG 4380
5 N E P A E L A A Y Y D V Q Y N A R E S V 24
4381 TCGAGGACTTCGACCAGTACCCGCGGCAGTACCGCGCGCTGAGCGATGACGCCCATGCC 4440
25 E D F D Q Y P R Q Y R A L S D D A H A R 44
4441 GGCTGCGCTGCATCAAGGACGTGGCCTACGGGCCGGGCGGGCGAGCGCCTGGATATCT 4500
45 L R C I K D V A Y G P G A G E R L D I F 64
4501 TCCCCGCCGAGCGACCCGACGCGCCGGTGTCTGCTGTTTCATCCATGGCGGCTACTGGCGG 4560
65 P A E R P D A P V L L F I H G G Y W R A 84
4561 CGCTGTCCAAGGCCGACTCGGCCTTCATGGCGCCGGCGCTGACCGCTGCCGGTGCCTGCG 4620
85 L S K A D S A F M A P A L T A A G A C V 104
4621 TGGTTGTGCTGGACTACGACCTGGCCCCGGCGGTGACCCTCGACCATATCGTCGACCAGA 4680
105 V V L D Y D L A P A V T L D H I V D Q T 124
4681 CCCGCCGCGCGCTGGCCTGGCTGCATCGGCACATCGCCGAGTTCGGCGGCGATCCGCAGC 4740
125 R R A L A W L H R H I A E F G G D P Q R 144
4741 GCCTCTACGCCAGCGGCAGCTCGGCGGGCGGCCACCTGACCGGCATGCTGCTGGCCGGCG 4800
145 L Y A S G S S A G G H L T G M L L A G G 164
4801 GCTGGCACGCCGACTACGGGGTGCCCGACAACGTGCTGCGCGGGGCATTGCCCATCAGCG 4860
165 W H A D Y G V P D N V L R G A L P I S G 184
4861 GCCTGTTTCGACCTGCGCCCGCTGCTGGAGACCCATATCAACGGCTGGATGGGCATGGACG 4920
185 L F D L R P L L E T H I N G W M G M D D 204
4921 ACGCGGCGGGCCCGGCGCAACAGCCCCGAGCTTCCAGCTGCCGACCCGGGGCGCGGAACTGG 4980
205 A A A R R N S P S F Q L P T R G A E L V 224
4981 TGATCAGCTATGGCGCGCTGGAGACGGCGGAGTTCCGCCCGCCAGTCCCACGAGTTCCTCG 5040
225 I S Y G A L E T A E F A R Q S H E F L E 244
5041 AGGCCTGGACCGCCCGTGGGCTGCCAGGTCGCTTCGTCGCGGCGCCGGGCAGGAACCACT 5100
245 A W T A R G L P G R F V A A P G R N H F 264
5101 TCGACGTGGTGTGCTGGAGCTGGGGCAGCCGGGGACGCCGCTGTACCAGGCGATGATCGAGT 5160
265 D V V L E L G Q P G T P L Y Q A M I E L 284
5161 TGATGGGGTTGAACCTGTAGGGGCGAATTCATTCGCCAAGGGCAGCGGAGCTGCCCTGCG 5220
285 M G L N L * 304
5221 GACTCCTTCGGGCAGACTGCGGTCTGCTTGGCGAATGAATTCGCCCTACGGGAAAACC 5280

5281 CGCCCCCAGGGCCGCACTTCACCCCAGCAGTTGACGCAGGCTTCCCGCCAGCCGCTGCTT 5340
GCGGGGGTCCC GGCGTGAAGTGGGGTTCGTAACCTGCGTCCGAAGGGCGGTGCGGCACGAA
149 * G L L Q R L S G A L R Q K 137
5341 CTCGTCGAGGCGCTGCATGATCCGCTGGACGATCTCGTAGCGGGCGGTGGCCACGAATTT 5400
GAGCAGCTCCGCGACGTACTAGGCGACCTGCTAGAGCATCGCCCGCCACCGGTGCTTAAA
136 E D L R Q M I R Q V I E Y R A T A V F K 117
5401 CATCTGGGCCAGGGCCTCGAGGGCGACGTCGGATTTTTCCAGGCGGTTGCGGGTGGAGCC 5460
GTAGACCCGGTCCCGGAGCTCCCGCTGCAGCCTAAAAAGGTCCGGCCACGCCACCTCGG
116 M Q A L A E L A V D S K E L G T R T S G 97
5461 GTCCACCTGGCCGAGGCCGCTGCGGGCGTCGCTGGAGCCCTGTTGCAGGCCGCCGACTAT 5520
CAGGTGGACCGGCTCCGGCGACGCCCGCAGCGACCTCCGGACAACGTCCGGCGGCTGATA
96 D V Q G L G S R A D S S A Q Q L G G V I 77

Appendix. Fig. 3. Continued.

5521 CTGGCGGATCTGCTCGCTGGCTTCGTTGGCGCGGGCGGGCGAGGTTGCGCACCTCGTCCGC 5580
GACCGCCTAGACGAGCGACCGAAGCAACCGCGCCCGCTCCAACCGGTGGAGCAGGCG
76 Q R I Q E S A E N A R R A L N R V E D A 57
5581 GACCACCGCGAAACCGCGCCCTGCTCACCGCGCGAGCCGCTTCGATGGCGGGCGTTCAA 5640
CTGGTGGCGCTTTGGCGCGGGACGAGTGGCCGCGCTCGGCGAAGCTACCGCCGCAAGTT
56 V V A F G R G Q E G A R A A E I A A N L 37
5641 GGCCAGCAGGTTGGTCTGCTTGGCGATCTCCTGGATGCTCCCGACTATCGAACCGATGCG 5700
CCGGTCTGTCACCAACAGACGAACCGCTAGAGGACCTACGAGGGCTGATAGCTTGGCTACGC
36 A L L N T Q K A I E Q I S G V I S G I R 17
5701 TTGCGACTGTTTCGCTGAGGTCGTTGAACACCTCGGTGATATGGCCCATGTAGCTGGCCAG 5760
AACGCTGACAAGCGACTCCAGCAACTTGTGGAGCCACTATAACGGGTACATCGACCGGTC
16 Q S Q E S L D N F V E T I H G M (ORF6) 1
5761 TTCGCTGATGGTGGCGCGGAGTCTCGGATGCTGTGGGTCGACAGCTGCAGGCATTCGCC 5820

5821 GCGCGGCGCGCCAGGTTGTTGGCCTGCTGCATCTCGGCCTCGCTCTCGGCGATGGAGCGG 5880

5881 CTGAGCAGTTCAGGTGCTGGAACAGCTGCTCGAAGGGGACGGCCGGAGCGGGCGCCACG 5940

5941 GTCTTTTCCACAGCGGCCTGGATAACCGCTGGCGGTGCCAGAGCCGGCACTTCGACCACT 6000

6001 ACTTCTATCGTCCGTGGCGGGCGCGCAGTGGAAAAGCCAACCCGCGCCAGGTTGCAGGCC 6060

6061 AGCCCCAGGCCAAGCAGCCAGGGGCTGGCCAGTACGAAGGCGATCAGACAAACAAGGCCG 6120

6121 CCACTGCCAAGCAGCAGGGCGGCCAGGGTTTTGCGGGACATGGAACGACTCCTTGTGGT 6180

6181 GGCCAGGACGTTTGTGGGGGCGATTTCAATCGCCAAGGGCAGCGAAGCTGCCCGGGGGT 6240

6241 TCGAAGGGCAGACCTGCGGCCTGCTTGGCGAATGAATTGCCCCCTACAGGTAGAGCAGGC 6300

6301 CGTGCCCAGGATGGGCAATGGCCCTCGGCCTTAGAACAGCGCCAGGGTGTAGTTGAGGAT 6360
GCACGGGTCCCTACCCGTTACCGGGAGCCGGAATCTTGTGCGGGTCCCACATCAACTCCTA
417 * F L A L T Y N L I 409
6361 CAGGCGGTTCTCGTCGATGTCGGTGCGGTAGTTGGAGCGCGGGTGACGTTGCGCACGCG 6420
GTCCGCCAAGAGCAGCTACAGCCACGCCATCAACCTCGCGCGCCACTGCAACGCGTGCGC
408 N R N E D I D T R Y N S R A T V N R V R 389
6421 GATGCCAGGCCCTTGAGCGAGCCGCTCTGGATCACGTAGCCGATGTCCAGGTGCGGTT 6480
CTACGGGTCCGGGAACCTCGCTCGGCGAGACCTAGTGCATCGGCTACAGGTCCAGCGCAAG
388 I G L G K L S G S Q I V Y G I D L D R E 369
6481 CCAGTCCTTGCCTTCGAAACCCTTGCCGGTGTGACGCTTGTACCGGTGATGTAGCGCAC 6540
GGTCAGGAACGGAAGCTTTGGGAACGGCCACAGCTGCAACAGTGGCCACTACATCGCGTG
368 W D K G E F G K G T D V N D G T I Y R V 349
6541 GGTGGTGGTCAGGCCGGGGATGCCGAGGGCGGCGAAGTCGTAGTCGTGGCGTACCTGCCA 6600
CCACCACAGTCCGGCCCCCTACGGCTCCCGCCGCTTACAGCATCAGCACCGCATGGACGGT
348 T T T L G P I G L A A F D Y D H R V Q W 329
6601 GGAGCGCTGGTTCGGCCGAGGCGAACTCGTAGGTGGGCACCTCGTTGCCAGTGGGCTGAT 6660
CCTCGCGACCAGCCGGCTCCGCTTGAGCATCCACCCGTGGAGCAACGGGTACCCGACTA
328 S R Q D A S A F E Y T P V E N G L P S I 309
6661 GTTGGCGAACACGCGGGGGAAGCCCTGGTCCCGAACATGGCCTGGTAACCGACGAAGAA 6720
CAACCGCTTGTGCGCCCCCTTCGGGACCAGCGGCTTGTACCGGACCATTGGCTGCTTCTT
308 N A F V R P F G Q D G F M A Q Y G V F F 289
6721 GGTGTTGCCGCGCGCTTGGCCGACAGCAGCAAGTAGAAGGCCTGGTTGTCGATCGAGCC 6780
CCACAACGGCGGCGGAACCGGCTGTGCTGCTTCATCTTCCGGACCAACAGCTAGCTCGG
288 T N G G R K A S L L L Y F A Q N D I S G 269

Appendix. Fig. 3. Continued.

| | | |
|------|--|------|
| 6781 | GATCAGCGCCTTGCCGTCCTCGTTGGAATCGTAGAAGCCGAGGTTGGCGCCCAGCACCCA | 6840 |
| | CTAGTCGCGGAACGGCAGGAGCAACCTTAGCATCTTCGGCTCCAACCGCGGGTCTGGGT | |
| 268 | I L A K G D E N S D Y F G L N A G L V W | 249 |
| 6841 | GTCGCCTACCGGCTCGGCGTGCTTAAGGCCGAGGTGGCGCTGGTTCGTAGATGTCTTCCAG | 6900 |
| | CAGCGGATGGCCGAGCCGCACGAATTCCGGCTCCACCGCGACCAGCATCTACAGAAGGTC | |
| 248 | D G V P E A H K L G L H R Q D Y I D E L | 229 |
| 6901 | CTGGCCGTACCAGGCGCTCAGGGTGGTGGCGTTGGCGTTGAAGGCGTAGTCGCCGCCGGC | 6960 |
| | GACCGGCATGGTCCGCGAGTCCCACCACGCCAACCGCAACTTCCGCATCAGCGGCGGCCG | |
| 228 | Q G Y W A S L T T R N A N F A Y D G G A | 209 |
| 6961 | GTAAGTTGAAGGCGTCGCCGGAGCCTGGCGCTGGGGCTGGTAGCCGACGATGGCCTGGAT | 7020 |
| | CATTCAACTTCCGCAGCGGCCTCGGACCGCGACCCCGACCATCGGCTGCTACCGGACCTA | |
| 208 | Y T S P T A P A Q R Q P Q Y G V I A Q I | 189 |
| 7021 | ATCGCTGCGGCCGGACTCGTTGCGCAGGCTGGTGGTGTTCAGGTGGCCGCCCTGCAGGGT | 7080 |
| | TAGCGACGCCGGCTGAGCAACGCGTCCGACCACCACAAGTCCACCGCGGGACGTCCCA | |
| 188 | D S R G S E N R L S T T N L H G G Q L T | 169 |
| 7081 | CAGGGCCGCGATCTCGCCGGAGGTGACGCTGACGCCCTGGTAGCTGGGCGGCAGCAGGCG | 7140 |
| | GTCCCGGCGCTAGAGCGGCCTCCACTGCGACTGCGGGACCATCGACCCGCCGTCGTCCGC | |
| 168 | L A A I E G S T V S V G Q Y S P P L L R | 149 |
| 7141 | GTTGTGCGTGAACACCAGCACCGGCACGTTGGGTTGCAGTTCGCCGACCTTCAGTTCGGT | 7200 |
| | CAACAGCGACTTGTGGTCTGGGCCGTGCAACCCAACGTCAAGCGGCTGGAAGTCAAGCCA | |
| 148 | N D S F V L V P V N P Q L E G V K L E T | 129 |
| 7201 | CTTGAGTAGCGCATCTTCCCGGCCACCCAGGCGGCTGTAGTCGTCGGCGGGCGCGCCG | 7260 |
| | GAACCTCATCGCGTAGAAGGGCCGGTGGGGTCCGCCGACATCAGCAGCCGCCCGCGCGGC | |
| 128 | K S Y R M K G A V G P P Q L R R R A R R | 109 |
| 7261 | TCCTCCTGCACCGGCAGCAGGCCGGTGTGCTGCGGTGCGGGCTGCTGTCGAGCTTGATG | 7320 |
| | AGGAGGACGTGGCCGTCGTCCGGCCACAACGACGCCAGCCCCGACGACAGCTCGAACTAC | |
| 108 | G G A G A A P R H Q Q P R P Q Q R A Q H | 89 |
| 7321 | CCGAGGGGTGCCGATGGCGTCGATACCGAAGCCAGCGGGCCCTGGGTGTAGCCGACTT | 7380 |
| | GGCTCCCCACGGCTACCGCAGCTATGGCTTCGGGTGCGCCGGGACCCACATCGGCCTGAA | |
| 88 | R P T G I A D I G F G L P G Q T Y G S K | 69 |
| 7381 | GAAGTTGAGGATGAAGCCCTGGGCCATTCTCGGCCTTGGACTGCTGGTTGGGCCCCGAC | 7440 |
| | CTTCAACTCCTACTTCGGGACCCGGTAAGGAGCCGGAACCTGACGACCAACCCGGGCTG | |
| 68 | F N L I F G Q A W E E A K S Q Q N P G V | 49 |
| 7441 | TATGTCGGAGAAGTCGCGGGTGAAGTAGTAGTTGCGCAGTTGCAGGGTGGCGGTGGAATC | 7500 |
| | ATACAGCCTCTTCAGCGCCGACTTCATCATCAACGCGTCAACGTCCCACCGCCACCTTAG | |
| 48 | I D S F D R S F Y Y N R L Q L T A T S D | 29 |
| 7501 | CTCGATGAAGCCGGACTCGGCGGCCTGGACGCAAAGGGGGAGGGCGCTCAGTAGGAGGGG | 7560 |
| | GAGCTACTTCGGCCTGAGCCGCCGGACCTGCGTTTTCCCCCTCCCGGAGTCATCCTCCCC | |
| 28 | E I F G S E A A Q V C L P L A S L L L P | 9 |
| 7561 | TAACGCACGGGACTTGTAGATCATCGCAAGGCTCTTCTTCTTCTGGTTTTCAAGGCGCAG | 7620 |
| | ATTGCGTGCCCTGAACATCTAGTAGCGTTC <u>GAGAAGA</u> AAGAAGACCAAAAGTCCGCGTC | |
| 8 | L A R S K Y I M (ORF7) | 1 |
| 7621 | CCATTGCGCCGCCGGCCCGCGAGGGGCCGAGGGTGATAGCTGCGTCCGGGTGTCCGGCTCG | 7680 |
| 7681 | TCGTGACCCGAGCCGGGATTCAGGCGCCGGTCAGACGCCGAGGTAGCGGTGCGACAGCT | 7740 |
| | AGCACTGGGCTCGGCCCTAAGTCCGCGGCCAGTCTGCGGCTCCATCGCCACGCTGTGCA | |
| 229 | * V G L Y R H S L | 222 |
| 7741 | CCGGTGTGCGCCGCAACTGTTCCGGCGTGCCCTGCCAGACCTGTCGGCCCTTCTCGATGA | 7800 |
| | GGCCACAGCGGCGGTTGACAAGGCCGCACGGGACGGTCTGGACAGCCGGGAAGAGCTACT | |
| 221 | E P T A A L Q E P T G Q W V Q R G K E I | 202 |
| 7801 | TGTGGTGGCGATCGACCACCCGCGCCATCTCCTTGAGGTTCTTGTGATCACCAGGATGG | 7860 |
| | ACACCACCGCTAGCTGGTGGGCGGGTAGAGGAACCTCCAAGAACAGCTAGTGGTCCCTACC | |
| 201 | I H H R D V V R A M E K L N K D I V L I | 182 |

Appendix. Fig. 3. Continued.

| | | |
|------|---|------|
| 7861 | TCTCGCCCTCGGCCTTGAGGGTGGCCAGGCAGTTCCAGATGGTCTCGCGGATCACCGGGG | 7920 |
| | AGAGCGGGAGCCGGAACCTCCACCGGTCCGTCAAGGTCTACCAGAGCGCCTAGTGGCCCC | |
| 181 | T E G E A K L T A L C N W I T E R I V P | 162 |
| 7921 | CCAGGCCCTCGGTGGCTTCGTTCGAGGATCAGCAGCTGCGGGTTGGTCATCAGCGCGGGG | 7980 |
| | GGTCCGGGAGCCACCGAAGCAGCTCCTAGTCGTTCGACGCCCAACCAGTAGTCGCGCGCCC | |
| 161 | A L G E T A E D L I L L Q P N T M L A R | 142 |
| 7981 | CCACCACCAGCATCTGCTGCTCGCCGCCGAGAGGGTCTGGACAGTTGCTCCTGGCGCT | 8040 |
| | GGTGGTGGTTCGTAGACGACGAGCGGGCGGCCTCTCCAGGACCTGTCAACGAGGACCGCCA | |
| 141 | A V V L M Q Q E G G S L T R S L Q E Q R | 122 |
| 8041 | CTTCCAGGCGCGGGAAGAGCTTGTAGATGCGCGCCAGGTCCCCTGCGCCACGGGTCG | 8100 |
| | GAAGGTCCGCGCCCTTCTCGAACATCTACGCGCGGTCCAGGGTGAACGGCGGTGCCCAGC | |
| 121 | E E L R P F L K Y I R A L D W K G G R T | 102 |
| 8101 | CGGTGGCCAGGAGGTTTTCTTACGCTCAGGCTGCCGAAGGCGGACGACCTTCCGGCA | 8160 |
| | GCCACCGGTCTCCAAAAGGAAGTGCAGTCCGACGGCTTCCGCGCTGCTGGAAGGCCGT | |
| 101 | A T A L L N E K V S L S G F A R R G E P | 82 |
| 8161 | CCAGGCCGAGGCCGCGCTGGGCGATGCGGAAGGGCAGGGCGCCGCGCATCTCCTTGCCGT | 8220 |
| | GGTCCGGCTCCGGCCGGACCCGCTACGCTTCCCGTCCCGCGGCGGTAGAGGAACGGCA | |
| 81 | V L G L G A Q A I R F P L A G R M E K G | 62 |
| 8221 | GGATGCGGATGCTGCCGCGCTGGGCTTGAGCAGGCCATGATGGATTTCACTGTGGTGG | 8280 |
| | CCTACGCTACGACGGGCGGACCCGAACCTCGTCCGGGTACTACCTAAAGTGACACCACC | |
| 61 | H I R I S G A S P K L L G M I S K V T T | 42 |
| 8281 | TCTTGCCCATGCCGTTGCGGCCGATCAGCGAGACCACCTCGCCCTGTTGATCTTCAGGT | 8340 |
| | AGAACGGGTACGGCAACGCCGGCTAGTCGCTCTGGTGGAGCGGGACAAGCTAGAAGTCCA | |
| 41 | T K G M G N R G I L S V V E G Q E I K L | 22 |
| 8341 | GCATGTCGAACAGCACCTGGCTCAGGCCGTAGCCGGCCTGCAATCCACTGACTTCAAGCA | 8400 |
| | CGTACAGCTTGTTCGTGGACCGAGTCCGGCATCGGCCGGACGTTAGGTGACTGAAGTTCGT | |
| 21 | H M D F L V Q S L G Y G A Q L G S V E L | 2 |
| 8401 | TGTTCTTCCCCAGGTAGGCCGCGCGGACCTGTGCGTTGCTGCGGACTTTCGTCGACGCTG | 8460 |
| | ACAAGAAGGGGGTCCATCCGGCGCGCCTGGACACGCAACGACGCCTGAAGCAGCTGCGAC | |
| 1 | M (ORF8) | 1 |
| 8461 | CCGGTGAGGATGGTCTGGCCGTAGACCAGGACGGTGATGCGGTTCGGCCAGGGCGAACACG | 8520 |
| | | |
| 8521 | GCGTCCATGTCGTGCTCCACCAGCAGGATGGCGTAGCGGCCCTTGAGCGACATCAGCAGC | 8580 |
| | | |
| 8581 | TCGGTCATGCGCGCGGACTCCTCCGCGCCATGCCGGCCATGGGCTCGTCCAGCAGCAGC | 8640 |
| | | |
| 8641 | ACCGAGGCTTCTTGGGCCAGCGCCAGGGCCAGCTCCAGTTGCCGGCGTTCCGCCGTGGGAC | 8700 |
| | | |
| 8701 | AGCTCGCTCACCAGGTGCTCGGGCGCGGGCCGAGGCCGACCTGGTCGAGGTACTGGCGG | 8760 |
| | | |
| 8761 | GCCGGGTCGAGCAGGCGGGTCCCGGCCGAGCGGACTCCACATGCCGAAGTGCGGCCTT | 8820 |
| | | |
| 8821 | CGCGGGCGGCGATGGCGACGGCGACGTTCTCCAGCAGGCTGAAGTCGGGGTAGAGCTGGC | 8880 |
| | | |
| 8881 | TGACCTGGAACGAGCGGGCCAGGCCAGGGCGGGCCG | 8918 |

Appendix. Fig. 3. Continued.

| | | |
|------|---|------|
| 1 | GAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAGCGGCAGCGGGTCCCTTCGGGATGCC | 60 |
| 61 | GGCGAGCGGGCGGACGGGTGAGTAATGCCTAGGAATCTGCCTGGTAGTGGGGGATAACGTT | 120 |
| 121 | CGGAAACGGACGCTAATACCGCATACTGCTACGGGAGAAAGTGGGGGATCTTCGGACCT | 180 |
| 181 | CACGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGGGGTAAAGGCTCACCAAGG | 240 |
| 241 | CGACGATCCGTAAC TGGTCTGAGAGGATGATCAGTCACACTGGAAC T GAGACACGGTCCA | 300 |
| 301 | GACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGCGAAAGCCTGATCCAGCC | 360 |
| 361 | ATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGC | 420 |
| 421 | AGTAAGTTAATACCTTGCTGTTTTGACGTTACCAACAGAATAAGCACCGGCTAACTTCGT | 480 |
| 481 | GCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGC | 540 |
| 541 | GCGCGTAGGTGGTTCAGCAAGTTGGAGGTGAAATCCCCGGGCTCAACCTGGGAACTGCCT | 600 |
| 601 | CCAAACTACTGAGCTAGAGTACGGTAGAGGGTAGTGGAATTTCTGTGTAGCGGTGAAA | 660 |
| 661 | TGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACTACCTGGACTGATACTGACAC | 720 |
| 721 | TGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA | 780 |
| 781 | CGATGTCGACTAGCCGTTGGGATCCTTGAGATCTTAGTGGCGCACTAACCGGATAAGTCG | 840 |
| 841 | ACCGCCTGGGGAGTACGGCCGCAAGGTTAAAAC TCAAATGAATTGACGGGGGCCGCACA | 900 |
| 901 | AGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCTGGCCTTGACAT | 960 |
| 961 | GCTGAGAACTTCCAGAGATGGATTGGTGCCTTCGGGAACTCAGACACAGGTGCTGCATG | 1020 |
| 1021 | GCTGTGCTCAGCTCGTGTGCTGAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTG | 1080 |
| 1081 | TCCTTAGTTACCAGCACGTTATGGTGGGCACTCTAAGGAGACTGCCGGTGACAAAACCGGA | 1140 |
| 1141 | GGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCTA | 1200 |
| 1201 | CAATGGTCGGTACAAAGGGTTGCCAAGCCGCGAGGTGGAGCTAATCCCATAAAACCGATC | 1260 |
| 1261 | GTAGTCCGGATCGCAGTCTGCAACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGTGA | 1320 |
| 1321 | ATCAGAATGTCACGGTGAATACGTTCCCGGGCCTTG TACACACCGCCCGTCACACCATGG | 1380 |
| 1381 | GAGTGGGTTGCTCCAGAAGTAGCTAGTCTAACCGCAAGGGGGACGGTTACCACGGAGTGA | 1440 |
| 1441 | TTCATGACTGGGGTG | 1455 |

Appendix. Fig. 4. A partial nucleotide sequence of 16S-rRNA gene from strain CA10 (accession number AB047273).

```

1 AAGCTTGTTTTTAATTCTAACAAGAGCATCTTGAAGGAGAGTATAGGTCATGAGTTCAAT      60
1 HindIII                                     (cumAl) M S S I      4
61 AATAAATAAAGAAGTGCAGGAAGCCCTTTGAAATGGGTGAAAAACTGGTCTGACGAGGA      120
5 I N K E V Q E A P L K W V K N W S D E E      24
121 GATTAAGCGCTCGTTGATGAGGAAAAGGGGTTGCTTGATCCACGTATTTTCTCTGATCA      180
25 I K A L V D E E K G L L D P R I F S D Q      44
181 GGATTTGTATGAGATCGAGCTTGAGAGGGTGTGCTCGATCCTGGCTGCTGCTGGGCA      240
45 D L Y E I E L E R V F A R S W L L L G H      64
241 CGAGGGGCACATTCCCAAAGCCGGGGATTATCTGACCACCTACATGGGTGAAGACCCAGT      300
65 E G H I P K A G D Y L T T Y M G E D P V      84
301 AATTGTAGTGAGGCAGAAAGACCGGAGCATTAAAGTCTTTTTAAACCAATGTCGGCATCG      360
85 I V V R Q K D R S I K V F L N Q C R H R      104
361 CGGTATGCGTATTGAGCGATCGGATTTTGGCAACGCAAAGTCATTTACCTGCACTTATCA      420
105 G M R I E R S D F G N A K S F T C T Y H      124
421 CGGGTGGGCCTATGACACCGCCGGTAATCTGGTCAATGTACCCTACGAGAAAGAGGCTTT      480
125 G W A Y D T A G N L V N V P Y E K E A F      144
481 TTGTGACAAAAAGAGGGTGACTGCGGGTTCGACAAGGCCGACTGGGGGCCGCTGCAAGC      540
145 C D K K E G D C G F D K A D W G P L Q A      164
541 GCGGGTGGATACTTACAAGGGGCTGATTTTTGCCAACCTGGGATAACCGAAGCCCCTGATTT      600
165 R V D T Y K G L I F A N W D T E A P D L      184
601 GAAGACCTATCTGAGCGATGCAACACCCTATATGGACGTGATGCTCGATCGGACCGAGGC      660
185 K T Y L S D A T P Y M D V M L D R T E A      204
661 AGTTACTCAGGTCATCACCGGTATGCAAAGACGGTAATCCCCTGTAACGGAAATTCGC      720
205 V T Q V I T G M Q K T V I P C N W K F A      224
721 CGCCGAGCAATTCTGTAGCGATATGTACCATGCGGGAACGATGGCGCATCTTTCAGGTGT      780
225 A E Q F C S D M Y H A G T M A H L S G V      244
781 ATTGTCCAGCCTCCCGCCTGAAATGGATTTGTCCCAAGTAAAGTTACCGTCAAGTGGGAA      840
245 L S S L P P E M D L S Q V K L P S S G N      264
841 TCAGTCCGGGCTAAGTGGGGTGGACATGGGACCGGCTGGTTCAATGACGATTTTCGCACT      900
265 Q F R A K W G G H G T G W F N D D F A L      284
901 TCTGCAAGCCATCATGGGTCCTAAGGTTGTCGATTACTGGACCAAAGGTCCAGCTGCTGA      960
285 L Q A I M G P K V V D Y W T K G P A A E      304
961 GCGTGCAAAGAGCGTCTGGGTAAAGTTCTTCCGGCTGATCGCATGGTTGCTCAGCATAT      1020
305 R A K E R L G K V L P A D R M V A Q H M      324
1021 GACCATTTTTCCGACATGCTCATTTCCTTCCGATCAATACAGTCCGTACTTGGCACCC      1080
325 T I F P T C S F L P G I N T V R T W H P      344
1081 ACGTGGCCCTAATGAGATCGAAGTTTGGTCTTTTCATCGTAGTGGATGCTGATGCACCTGA      1140
345 R G P N E I E V W S F I V V D A D A P E      364
1141 AGATATCAAGGAAGAATATCGTCGGAAAACATCTTCACCTTCAATCAAGGGGGAACCTA      1200
365 D I K E E Y R R K N I F T F N Q G G T Y      384
1201 CGAGCAGGACGATGGCGAAAACCTGGGTGGAGGTTTCAGCGGGGATTGCGCGGCTACAAGGC      1260
385 E Q D D G E N W V E V Q R G L R G Y K A      404
1261 TAGAAGTAGACCTCTTTGTGCCAGATGGGGGCGGGTGTGCCAAACAAGAACAACCCGGA      1320
405 R S R P L C A Q M G A G V P N K N N P E      424
1321 GTTTCCTGAAAAGACCAGCTACGTTTATAGCGAAGAAGCTGCGCGAGGGTTCTACCACCA      1380
425 F P G K T S Y V Y S E E A A R G F Y H H      444
1381 CTGGAGCCGCATGATGTCCGAGCCGAGTTGGGACACGCTAAAGTCTTGAGCAGATAAAGT      1440
445 W S R M M S E P S W D T L K S *      459
1441 GACCGAAAAAAGCAATCACTTTCATCGGGTTTCTACCGTGGTAGACAAGGGTTTAGCCTG      1500

1501 TTTTTTGGTTGCTGGAAGTGCCTAAGTGAATTGATTAACCTGGGTAAACCCCTGGCTTTG      1560

```

Appendix. Fig. 5. The nucleotide and deduced amino acid sequences of the 6,508-bp *HindIII* fragment of pIP103. The sense DNA strand is shown. The putative SD sequences are indicated by double lines under the sequence. Astarisks indicate stop codons.

1561 TCGGGGGTATTTACTCGGGTGCATTCCAAAATGTACAGCTGTGCGTTTGGTGATAATCGT 1620

1621 CATGCTATGGATTTGCTATTTGCATGAGCCGAGTGCAGGTGCGCCAACATATATACAGGA 1680

1681 AACTAATTATGACATCCGCTGATTTGACAAAACCCATCGAGTGGCCAGAAATGCCCGTCA 1740
1 (cuma2) M T S A D L T K P I E W P E M P V S 18

1741 GTCTTGAATTGCAAAATGCCGTTGAGCAATTCTACTATCGCGAAGCACAGTTGCTTGATT 1800
19 L E L Q N A V E Q F Y Y R E A Q L L D Y 38

1801 ATCAAACTATGAGGCCTGGCTGGCTTTACTGACCCAAGACATCCAATATTGGATGCCAA 1860
39 Q N Y E A W L A L L T Q D I Q Y W M P I 58

1861 TTCGTAATACTCATAACATCCCGGAATAAGGCGATGGAGTACGTGCCCCCGGGGTAATG 1920
59 R T T H T S R N K A M E Y V P P G G N A 78

1921 CCCATTTTGCAGGAGACGTATGAGAGCATGCGTGCAGCATTCGGGCGAGGGTTTCGGGGC 1980
79 H F D E T Y E S M R A R I R A R V S G L 98

1981 TTAAGTGGACTGAAGATCCACCGTCGCGCAGCCGGCACATTGTAAGCAACGTTATCGTCC 2040
99 N W T E D P P S R S R H I V S N V I V R 118

2041 GCGAAACTGAGAGTGCTGGTACTTTGGAAGTTAGTTCTGCGTTCCTTTGTTACCGTAATC 2100
119 E T E S A G T L E V S S A F L C Y R N R 138

2101 GATTGGAGCGTATGACGGACATCTATGTGCGGTGAGCGTCGAGATATTTTGTCCGTGTAA 2160
139 L E R M T D I Y V G E R R D I L L R V S 158

2161 GTGACGGGCTGGGATTCAAAATGCGCAAGCGAAGCATCTTGCTCGACCAGAGCAGGATTA 2220
159 D G L G F K I A K R T I L L D Q S T I T 178

2221 CAGCGAATAATCTCAGCCAGTTTTTCTAACTAGGGAATGCTGGCCACTTACCCTATACCC 2280
179 A N N L S Q F F * 186

2281 AGCCTATTCATGAGAGCGGCCTGAAAATGAAGAGGAGCTACCCGATAGCTACGCAAAC 2340
1 (ORF3) M K R S Y P I A T Q T N 12

2341 ATCGCGCTCGCCCTTTTCTGATCGCGATCGGTATCTTTTACTTGGCCAATCTCCTCGGGA 2400
13 R A R P F L I A I G I F Y L A N L L G T 32

2401 CTTTGCATTTTCAGCAGCCTGCGGCTGTTTCGGCATGATGTATTCGGGTGTGGATTTGCAGG 2460
33 L H F S S L R L F G M M Y S G V D L Q V 52

2461 TCGGCGCTCCGGTATTCACCCTGCTGCAGGATGCCTGGGCCGTAGTCGGGCTGCAGCTGG 2520
53 G A P V F T L L Q D A W A V V G L Q L G 72

2521 GGGGCACTGGGCTGGTTGCGTTGTGGGGCGCACGTACCCCGTGCCTTCATGGCGGTTG 2580
73 G T G L V A L W G A R Q P V R F M A V V 92

2581 TCCCGTGGTCATCGTCACGGAAGTGCTCGACGGTATCTGGGACTTGTACAGCATCGTTT 2640
93 P V V I V T E V L D G I W D L Y S I V W 112

2641 GGAGTCACGAAGCCATGTGGTTCCGGCTCCTGACGTTCCGCATCCACGTGGTGTGGATCG 2700
113 S H E A M W F G L L T F A I H V V W I V 132

2701 TCTGGGGTTACAGGTATGGCGCGTGTCTGCTCGCCGGTTCATCTGGCTTAACCGTCCCAA 2760
133 W G L Q V W R V S S R R S S G L T V P T 152

2761 CCTCCTGAATCTGTGGGCTGAATTGAACTCTGTCAATTTCCATGCGCGTCGACGTTATC 2820
153 S * 153

2821 CGGCCGCTTGCGCCAAGCTGGCTGCCGGATTTAATGAGTAGTTAAAGTTAGCTATAGAAA 2880

2881 CTCTGAAAAAGGCTTGACCTCATGAGATATCCAGTCTGCAGTCCGCGTGGTTACTGGCGT 2940

2941 GCATTTTCCGAGTGCCTACTTTTTTTCAGACCAACTCTATAATAAAGAGACAAAAAGAATGA 3000
1 (cuma3) M T 2

3001 CTTTTTCCAAAGTTTGTGAAGTATCTGATGTGCCGTCGGTGACGCCTTGCAGGTTGAAA 3060
3 F S K V C E V S D V P V G D A L Q V E S 22

3061 GTAAGGGCGAAGCCGTCGCGATTTTCAACGTCGATGGAGAGTTGTTTCGCAACACAGGACC 3120
23 K G E A V A I F N V D G E L F A T Q D R 42

3121 GTTGCATCATGGTACTGGTCCCTTGTCCGAAGGCGGCTACCTAGAGGGTGACATTGTGC 3180
43 C T H G D W S L S E G G Y L E G D I V E 62




Appendix. Fig. 5. Continued.

3181 AATGCTCGCTGCACATGGGTAGGTTCTGTGTCCGCACGGGCAAGGTAAAAGCAGCACCGC 3240
63 C S L H M G R F C V R T G K V K A A P P 82
3241 CCTGTGAGCCGCTGAAGATATATCCGATTCGAATAGATGGCAGCGATGTGTTTCGTAGACT 3300
83 C E P L K I Y P I R I D G S D V F V D F 102
3301 TTGATGCCGGGTATCTAGCGCCATGATTAATCAATCGTCATTATTGGTGCTGGCTTGGC 3360
103 D A G Y L A P * 109
1 (cumA4) M I K S I V I I **G A G L A** 13
3361 TGGCGCAACTGCCACTCGCTATCTTCGCGCCCAAGGATATCAGGGAAAGATCCATCTGGT 3420
14 **G A T A** T R Y L R A Q G Y Q G K I H L V 33
3421 CGGGGAGGAGTTGCATGTGGCTTACGATCGCCCCCTCCTTATCCAAGGACACCCTGTCAGG 3480
34 G E E L H V A Y D R P S L S K D T L S G 53
3481 AAAAGTGGTCGAACCACCCGCAATCCTGGATCCTTGTTGGTATGCATCGGCCGATATAGA 3540
54 K V V E P P A I L D P C W Y A S A D I D 73
3541 TCTCCATTTAGGTGTACGCGTGACCGGTATTGATGTGGTAAACCACCAGGTACTTTTCGA 3600
74 L H L G V R V T G I D V V N H Q V L F E 93
3601 ATCCGGTGACATTCTAGCCTACGACCGACTGCTATTAGCCACCGGCGCTCGCGCGCGGGC 3660
94 S G D I L A Y D R L L L A T G A R A R R 113
3661 TATGGCTATTACGGGAAGCGAGTTGGCCGGCATTACACCTTGCGTGACCGCGCCGACAG 3720
114 M A I T G S E L A G I H T L R D R A D S 133
3721 CCAGGCGCTGAGGCAGGCGCTTGAGCCGGGCCAGTCTCTGGTAATTGTGCGCGGTGGCCCT 3780
134 Q A L R Q A L E P G Q S L V I V **G G G L** 153
3781 GATCGGTTGCGAAGTGGCGACCACTGCTATTAATGCCGGTGCCACGTCCTGTTCTGGA 3840
154 **I G C E V A** T T A I N A G A H V T V L E 173
3841 GGCCGGGACGAAGTGCCTGTTGCGAGTGCTAGGCCGATCAACCGGGGCTGGTGTGCGAA 3900
174 A G D E L L L R V L G R S T G A W C R N 193
3901 CGAGTTGGAGCGTTTGGGTGTCCGGGTTGAACTGAACGCACAGGCAGCGCATTTCGAGGG 3960
194 E L E R L G V R V E L N A Q A A H F E G 213
3961 CGAGGGACACGTGCATGCCGTCGTTTGTGCCGATGGACGTCGGATAGCAGCTGGCACAGT 4020
214 E G H V H A V V C A D G R R I A A G T V 233
4021 TTTGGTGAGCATCGGTGCAGAACCAGCCGACGAAGTGGCACGTGCGGCCGGTATCGCATG 4080
234 L V S I G A E P A D E L A R A A G I A C 253
4081 TGAGCGCGGCGTGGTAGTTGACGCTACGGGTGCAAGCTCATGTCCTGCAGTATTCGCGGC 4140
254 E R G V V V D A T G A S S C P A V F A A 273
4141 AGGTGACGTAGCAGCCTGGCCCCTGAGGTCCGGTGAAGTGCCTCGCTGGAGACCTACCT 4200
274 G D V A A W P L R S G E L R S L E T Y L 293
4201 GAACAGCCACATGCAGGCTGAAACTGCCGCCGCGCCATGTTAGGCAAGTCTATCCCGGC 4260
294 N S H M Q A E T A A A A M L G K S I P A 313
4261 TCTTCAGGTGCCAACCTCTTGGACGGAGATTGCAGGGCATCGGATACAGATGGTTGGCGA 4320
314 L Q V P T S W T E I A G H R I Q M V G D 333
4321 CATCGAAGGCCCGGAGAAGTTGTCTTGC GCGGTAACGTCGAGAATGGTCAGCCGCTGGT 4380
334 I E G P G E V V L R G N V E N G Q P L V 353
4381 GCAGTTCAGGGTCTTGATGGTTCGCGTTGAAGCCGCAACGGCTATCAATGCCCCGGAAGA 4440
354 Q F R V L D G R V E A A T A I N A P E D 373
4441 TTTTCCCGTTGCAACCCGATTGGTGGCTGACCACATTCCTGTATCGGCCACAAAATTGCA 4500
374 F P V A T R L V A D H I P V S A T K L Q 393
4501 GGACGCTAGCTCTAACTTGC GGGATTTTATGAAAGCTAAAGCTGAGCGATGCGAGTGAAC 4560
394 D A S S N L R D F M K A K A E R C E * 411
4561 GTCGACTTATTCTTAATAACTAGAAGAGCTTAAAAGTGAAGCTAAAAGAAGAAGTTATAT 4620
1 (cumB) V K L K E E V I L 9
4621 TAATCACAGGGGGCGCATCTGGACTGGGGCATGCCCTTGTGGAACGGTTCGTGGCCGAAG 4680
10 I T G G A S G L G H A L V E R F V A E G 29

Appendix. Fig. 5. Continued. The amino acids which may constitute FAD- and NAD-binding site [Niedle et al., 1992] with the conserved motif GXGX₂GX₃A are indicated by shaded boxes.

| | | |
|------|--|------|
| 4681 | GTGCCAAGGTGGCTGTCCTCGATAAGTGTGCCGATCGACTTCAACAGTTGGAGTCTGATC | 4740 |
| 30 | A K V A V L D K C A D R L Q Q L E S D H | 49 |
| 4741 | ACGGCGAAGATGTGGTATGCATTGTCGGTGATGTGCGTTCAATGGAAGACCAGAACTGG | 4800 |
| 50 | G E D V V C I V G D V R S M E D Q K L A | 69 |
| 4801 | CTGCCAGTCGTTGTATTGCCAAGTTCGGGAGGATTGACACTCTGATCCCCAACGCCGCTA | 4860 |
| 70 | A S R C I A K F G R I D T L I P N A A I | 89 |
| 4861 | TCTGGGATTACAACACGGCACTTGTGCGATCTACCTGAAGACAGTATCGACAAAGCATTG | 4920 |
| 90 | W D Y N T A L V D L P E D S I D K A F D | 109 |
| 4921 | ATGAAGTATTTTCAGATTAATGTTAAGGGCTATATTTCTCGCCGTCAAGGCATGCCTGCCAG | 4980 |
| 110 | E V F Q I N V K G Y I L A V K A C L P A | 129 |
| 4981 | CCCTGGTTGCCAGCCGAGGCAGCGTGATCTGCACGATCTCGAATGCAGGTTTTTACCCCA | 5040 |
| 130 | L V A S R G S V I C T I S N A G F Y P N | 149 |
| 5041 | ATGGCGGTGGTCTCTTTACACGGCGACGAAGCACGCAGTGGTGGGGCTGGTGC CGGAGC | 5100 |
| 150 | G G G P L Y T A T K H A V V G L V R E L | 169 |
| 5101 | TGGCATTTCGAGTTGGCACCATACGTCCGGGTCAATGGTGTGGGGGTTGGCGGTATCAATA | 5160 |
| 170 | A F E L A P Y V R V N G V G V G G I N T | 189 |
| 5161 | CGGATTTGAGAGGACCTTGCTCGTTGGGAATGAGTGAGCAGTCGATATCGAATGTGCCGT | 5220 |
| 190 | D L R G P C S L G M S E Q S I S N V P L | 209 |
| 5221 | TGGCTGAGTTGCTGCAGGACGTACTGCCAATCGGCCGCCTGCCTGACGCCGAAGAATACA | 5280 |
| 210 | A E L L Q D V L P I G R L P D A E E Y T | 229 |
| 5281 | CAGGAGCTTATGTGTTTTTCGCCACGCGTGGGACATCTGCGCCCGCCACCGGTGCCTTAC | 5340 |
| 230 | G A Y V F F A T R G T S A P A T G A L L | 249 |
| 5341 | TGAATTATGACGGAGGCATGGGCGTTTCGCGGACTTTTTTCCGCAGTCGGAGGTAAGGACC | 5400 |
| 250 | N Y D G G M G V R G L F S A V G G K D L | 269 |
| 5401 | TGCTCGAAAACTTAACATTGATTGAGGAGATGAAAATGGGCATTAAGCTTGGGTTAC | 5460 |
| 270 | L E K L N I D * | 276 |
| 1 | (cumC) M G I K S L G Y | 8 |
| 5461 | ATGGGGTTCCTCTGTAAGTGATGTACCGGCATGGCGCTCGTTTCTCACCGAAAAAGTGGGT | 5520 |
| 9 | M G F S V S D V P A W R S F L T E K V G | 28 |
| 5521 | TTGATGGAGGTTGTTGGCTCCGATGAGAATGCCTTATACCGCATGGACTCACGCAGTCGG | 5580 |
| 29 | L M E V V G S D E N A L Y R M D S R S R | 48 |
| 5581 | CGGATTGCCGTGGAAAGGGGGGAGGCTGACGACCTAGCATTCCGCCGTTATGAAGTTGCC | 5640 |
| 49 | R I A V E R G E A D D L A F A G Y E V A | 68 |
| 5641 | AATCCGCTGGCCTTGAAGCTGATTACGGAGCGGTACGGGAGGCTGGTGTTCAGGTGAGG | 5700 |
| 69 | N P L A L K L I T E R L R E A G V Q V R | 88 |
| 5701 | ACGGCGACACTGAACTGGCAGAAAAGCGTGGCGTTGATGGAACCTGGTCTCTTATGAAG | 5760 |
| 89 | T G D T E L A E K R G V D G T W S L M K | 108 |
| 5761 | ATCCATTTGGAATGCCCACTGGAAATTTACTACGGGACTACCGAACTATTCGAGCAGCCT | 5820 |
| 109 | I H L E C P L E I Y Y G T T E L F E Q P | 128 |
| 5821 | TTCGTTTCTGGCACTGCTGTCACTGGGTTCTGACTGGTGACCGAGGAGCTGGGCATTAT | 5880 |
| 129 | F V S G T A V T G F L T G D R G A G Y | 148 |
| 5881 | TTTTATACTGTCCCGAATATTGAAGAAGGACTGGCTTTCTATACTGGCATGCTGGGTTTC | 5940 |
| 149 | F Y T V P N I E E G L A F Y T G M L G F | 168 |
| 5941 | CAGATGTCCGACGTCATTGATATAGCTATGGGTCCGGATATTACAGTGCGGGGATACTTT | 6000 |
| 169 | Q M S D V I D I A M G P D I T V R G Y F | 188 |
| 6001 | CTTCATTGCAACGGGCGCCACCACACAATGGCGATCGCGGAGGCTCCGTTACCCAACAGA | 6060 |
| 189 | L H C N G R H H T M A I A E A P L P N R | 208 |
| 6061 | GTTACCATTTTTTACAGAGTCCCTTGACGCTGGATGATGTAGGTCATGCGTACGACCGA | 6120 |
| 209 | V H F L Q S P L T L D D V G H A Y D R | 228 |
| 6121 | ATCGATGGATTGGGCGACAAATCTACCGACTCCAATCTTCGGGTGCCGGCAAATAGTGAT | 6180 |
| 229 | I D G L G D K S T D S N L R V P A N S D | 248 |
| 6181 | ATTAGGTCCAGCAGGATCACGGCGACGATCGGACGCCATGTCAACGATCACATGATTTCC | 6240 |
| 249 | I R S S R I T A T I G R V N D H M I S | 268 |

Appendix. Fig. 5. Continued. Three His (H) residues and one Tyr (Y) residue conserved in extradiol dioxygenase [Hofer et al., 1993] are indicated by shaded boxes.

| | | |
|------|---|------|
| 6241 | TTTTACGCTGAGACGCCGTCCGGGTTTGAGCTTGAGTTTGGTTGGGGCGCGCGCGACGTA | 6300 |
| 269 | F  A E T P S G F E L E F G W G A R D V | 288 |
| 6301 | GATGACCGGTCTTGGGTGATGACGAGGCACAAGCGCACGGCCATGTGGGGTCATAAATCT | 6360 |
| 289 | D D R S W V M T R H K R T A M W G H K S | 308 |
| 6361 | ATGCGTAATAAGTAAGTACTTGGTCCGGAGTTAATATTTACTGAGGTTTTGCACTTACAG | 6420 |
| 309 | M R N K * | 312 |
| 6421 | TCTGT        TCGTACAACCTCTGTCCTGGG | 6480 |
| 6481 | GTTAGATTTGATCAATTTTGATA <u>AAGCTT</u> | 6508 |

HindIII

Appendix. Fig. 5. Continued. The inverted repeat region found in the downstream of *cumC* is indicated by shaded boxes.

謝辞

本博士論文研究を行うにあたり、多くの方々から御指導、御協力を賜りました。

修士課程在学中から助手に採用され現在に至るまで、この様な素晴らしい研究の場と研究テーマを与えていただき、親身な御指導と常に暖かい御配慮をいただきました東京大学生物生産工学研究センター生物制御工学部門教授 大森俊雄先生に心より感謝申し上げます。同様に様々な場で貴重な御助言をいただき、かつ暖かく見守っていただきました同センター生物構造工学部門教授 山根久和先生に深く感謝いたします。また、直接御指導を賜り、折に触れ貴重な御教示、御配慮をいただきました同センター生物制御工学部門講師 野尻秀昭先生に深く感謝申し上げます。同様に様々な場でお世話になり、暖かい御配慮をいただいた同センター生物構造工学部門助教授 西山真先生、同センター生物制御工学部門助手 吉田貴子先生に感謝いたします。

本研究の共同研究者であり、折に触れ貴重な討論をしていただいたトヨタ自動車株式会社 木邑敏章博士、昭和電工株式会社 青木裕史氏、秋田県立大学助手 春日 和博士、韓国京畿大学助教授 李 宗勲博士に深く感謝申し上げます。また、共同研究者として精力的に研究をこなしてくださった鄭 鎮成氏、関口裕代氏、中井誠一郎氏、前田香奈氏、浦田雅章氏に深く感謝いたします。

また本研究をまとめる上で、南 貞媛博士、Jaka Widada 博士をはじめとする生物制御工学研究室の皆様、ロッテ中央研究所 吉田圭司郎氏には、日々の研究に際し貴重な御助言、御協力をいただいただけでなく、様々な御配慮をいただきました。この場を借りて厚く御礼申し上げます。

最後になりましたが、常に暖かく見守り、励ましながら研究生活を支えてくださいました両親をはじめとする家族に心から感謝いたします。

2001年 6月

羽部 浩