GENETIC AND CHEMICAL STUDIES RELATING TO α-AMINOISOBUTYRIC ACID (AIB) METABOLISM

By

Jeffrey A. Bickmeier

RECOMMENDED:

Advisory Committee Chair

Departmental Head

APPROVED:

Dean, College of Natural Science & Mathematics

Dean of Graduate School

September 23, 2004
Date

GENETIC AND CHEMICAL STUDIES RELATING TO α-AMINOISOBUTYRIC ACID (AIB) METABOLISM

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

In Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Jeffrey A. Bickmeier, B.S.

Fairbanks, Alaska

December 2004

Abstract

The non-protein amino acid α-aminoisobutyric acid (Aib) occurs naturally in soil. *Burkholderia cepacia* utilizes Aib as a sole nitrogen source by virtue of a 2,2-dialkylglycine decarboxylase, which is encoded by the *dgdA* gene. In this study, 30 prokaryotes were cultured on minimal media containing Aib as the sole nitrogen source and identified by 16S rRNA gene sequencing. Several attempts were made to clone *dgd* genes from these organisms. Plasmid libraries were constructed in *E. coli* from DNA isolated from four Aib-utilizing bacteria. Libraries were screened for clones which demonstrated the ability to grow on Aib containing media and therefore contained cloned *dgd* genes. Sequence analysis of plasmids isolated from Aib utilizing clones revealed no similarity to *dgd* genes from *B. cepacia.*

Also, 4-methyl-2-(4-nitrophenyl)-5(4H)-oxazolone was alkylated at the 4-position with various alkyl halides. This was the key step in synthesis of disubstituted N-*p*-nitrobenzoyl-DL-glycines, which are precursors to 2,2-dialkylglycines. Synthesis was performed in multiple steps. N-*p*-nitrobenzoyl-DL-alanine was synthesized by aminolysis of DL-alanine with benzoyl chloride. The 4-methyl-2-(4-nitrophenyl)-5(4H)-oxazolone was synthesized by dehydration of N-*p*-nitrobenzoyl-DL-alanine with acetic anhydride. Alkylation of 4-methyl-2-(4-nitrophenyl)-5(4H)-oxazolone was performed by adding base, which made the oxazolone enolate, and alkyl halide, which participated in S_N2 reactions with the enolate. Substituted oxazolones were hydrolyzed into disubstituted N-*p*-nitrobenzoyl-DL-glycines with DCl.

Table of Contents

List of Figures

List of Schemes

List of Tables

Acknowledgments

This research was supported by a Genome Diversity grant from Alaska EPSCoR, a grant from the University of Alaska Foundation, and by the UAF Department of Chemistry and Biochemistry. The *M. smegmatis* strain was kindly provided by Dr. Keith Derbyshire of the New York State Department of Health Wadsworth Center.

I would like to thank my research advisor, Dr. John Keller, for his support and encouragement throughout my coursework and research. Much thanks to my committee members Dr. Tom Clausen and Dr. Larry Duffy for their assistance. I would also like to acknowledge the invaluable help of Sheila Chapin for her incredible knowledge of UAF bureaucratic minutia, Marlys Schneider for her seemingly constant availability to help me track down chemicals, Dr. Richard Stolzberg and Dr. Tom Clausen for helping me become an honorary "Junior Gas Chromatographer", and Hua Zhang for her help with DNA sequencing.

I would also like to single out the following current/former students: Hong Hong Sun, Lilly Allen-Daley, Trent Volz, Emily Reiter, Chance Riggins, Jane LeBlond, Tanya Bailey, Tammy Thompson, Adrienne Orr, Erik Harrington, and Shane Coleman. In an academic or leisure capacity I count them as great allies.

Lastly I would like to thank my wife, my friends, and my family. I could not have started this project without the love and support I received over the course of my entire life. Because of this fortune I dedicate my thesis to them.

Chapter 1. Overview and Stratagy

Introduction

This is a two-part thesis. Part 1 describes the isolation of α-aminoisobutyric acid (Aib)-utilizing soil organisms, their identification through 16S rRNA gene sequencing, and progress toward the cloning, sequencing, and characterization of their Aib-utilization genes. Part 2 describes the synthesis of organic compounds related to Aib, specifically, α,α-disubstituted N-acyl amino acids.

This introduction describes the history and background of 2,2-dialkylglycine decarboxylase (DgdA) research, the significance of DgdA and its regulatory protein DgdR, the genes that encode them (*dgdA* and *dgdR*), the use of 16S rRNA gene sequencing as a means of identifying prokaryotes, the cloning and expression of genes from various genera of prokaryotes in *E. coli,* and the use of oxazalones in synthesis of amino acids and related compounds. Areas of interest involving *dgdA* and *dgdR* research, potential problems, and benefits of this investigation are also discussed.

Molecular Biology

2,2-dialkylglycine decarboxylase. Aib is an extremely stable, non-coded amino acid also known as 2-methylalanine (2MA) (Figure 1). It is distinguishable from alanine in that it contains a second methyl group in place of the alanine α-hydrogen. Aib occurs naturally in soils as a major component of enzymatically synthesized antibiotic peptides made by fungi[1] and as a result of carbonaceous meteor earth striking events.[2] Pyridoxal phosphate (PLP)-dependent metabolism of Aib by *Burkholderia* (formerly *Pseudomonas*) *cepacia* was first reported in the classic study by Aaslestad and Larson in which the bacterium was shown to utilize Aib as a sole nitrogen source.[3] Subsequently, purification and characterization of a vitamin B6 dependant 2,2-dialkylglycine decarboxylase (DgdA) isolated from *B. cepacia* was reported by Bailey and Dempsey.[4]

DgdA function. Decarboxylation and succedent transamination of 2,2-dialkylglycines by DgdA proceeds via two half-reactions. The reaction initiates with the PLP bonded as an imine to the amino group of the Lys272 side chain in the active site, and then proceeds with the transaldimination of the cofactor with the substrate (Scheme 1). The first reaction, which has the effect of transferring the substrate amino group to the enzyme-bound cofactor, results in the production of pyridoxamine phosphate (PMP), acetone, and carbon dioxide (Scheme 2a). The second half reaction is a reversible transamination in which the PMP amino group reacts with pyruvate resulting in a renewal of PLP and the

production of L-alanine (Scheme 2b). Binding of a potassium ion near the enzyme active site is required for catalysis.[5]

Characterization of DgdA. Purification of DgdA to homogeneity was first performed by Bailey and Dempsey.[4] Later, Lamartiniere et al showed that the functional enzyme has a molecular mass of 180 kDa and is composed of four identical subunits of about 45 kDa.[6, 7] Keller et al first cloned the *dgd* genes and showed the gene sequence encodes a 434-residue subunit with significant homology to aminotransferases, but not to other decarboxylases.[8] This has led to the suggestion that *dgdA* evolved from a gene that encoded an aminotransferase and secondarily developed the ability to decarboxylate, rather than a decarboxylase that later evolved aminotransferase ability. Finally, crystal structures determined to 2.1 Å resolution were reported by Toney et al, which revealed details about the DgdA structure.[9] The expected tetrameric protein was shown to be a dimer of dimers, with the closest association occurring between monomers within the dimer portion of the enzyme (Figure 2a). As the irreducible functional unit, each dimer contains two active sites composed of residues from both monomers. The interactions between the two dimers are much less involved and seem to convey no apparent functional advantage.

The DgdA structure includes two metal ion binding sites per monomer, one with binding specificity for a sodium ion, and another that binds unspecifically to Group I cations but requires bound potassium or cesium for the enzyme to function properly[10] (Figure 2b). It has been shown that the bound potassium or cesium ion participates in a

hydrogen bond network extending from the ligand to residues in the adjacent active site, which forms a specific structure necessary for enzyme function.[9] Also interesting is the apparent bifunctionality of the active site that carries out both decarboxylation and transamination. This is considered unusual in that most multifunctional enzymes contain a discrete active site for each function.

DgdR regulates the transcription of *dgdA*. In cloning, expressing, and sequencing the *dgdA* gene from *B. cepacia*, Keller et al also discovered an open reading frame (ORF) divergently transcribed 78 bp away from the *dgdA* transcription start site.[8] An investigation into the sequence of this upstream ORF, called *dgdR*, and a study of its gene product revealed a predicted 294-amino acid DgdR protein that was found to negatively regulate the metabolism of Aib.[11] DgdR shows significant homology to the LysR family of transcriptional regulators, and includes the putative N-terminal helix-turn-helix (HTH) DNA binding motif typical of this protein family. DNA footprinting assays by Allen-Daley et al showed that the DgdR protein binds to the promoter region of *dgdA* where RNA polymerase also binds prior to initiation of transcription. In the presence of the inducer Aib, the area of DNA covered by DgdR was reduced and the conformation of the DNA-DgdR complex was altered.[11]

A hypothetical molecular mechanism by which DgdR regulates the transcription of *dgdA,* based on the experimental evidence, was proposed by Allen-Daley et al. These steps include, a) The DgdR protein remains bound to the promoter of *dgdA*, blocking its transcription, b) Aib enters the cell and binds to DgdR, c) Aib induces a conformation

change in the protein that includes releasing a portion of the *dgdA* promoter, d) Relaxation of DgdR allows RNA polymerase access to the *dgdA* promoter and in turn, the cell expresses DgdA enzyme and metabolizes Aib.[11] Presumably, this continues until the supply of Aib is exhausted which permits DgdR to cover the *dgdA* promoter once again.

LysR-type transcriptional regulation. LysR-type transcriptional regulators (LTTRs) were first reported by Henikoff et al. in 1988[12] and are now identified as probably the largest family of prokaryotic regulatory proteins. Since its initial discovery, this group has accrued hundreds of members that span a diverse range of genera, and regulate an abundance of different genes and regulons.[13] While functionally diverse, this protein class is distinguished by similar characteristics. LTTRs encode proteins that are coinducer-responsive transcriptional activators and/or repressors; range in size from 276 to 324 residues per monomer; bind to regulated DNA targets both in the presence and absence of inducer; are divergently transcribed from a promoter that is close to and often overlaps the promoter of the regulated target gene; and also control transcription of their own gene due to the proximity of the promoters.[13] Some exceptions from this paradigm have been reported and include systems lacking divergent promoters that still negatively autoregulate, systems that do not negatively autoregulate, and/or systems that positively autoregulate.[13]

The structures of LTTRs are best described as tetrameric or dimeric DNA binding oligomers that share a common domain configuration, including a helix-turn-helix (HTH)

DNA binding domain within residues 1-66, a coinducer responsive domain between residues 122-270, and a conserved C-terminal domain.[13] The most conserved domain is the helix-turn-helix at the N terminus, which among all LTTRs has about a 40% similarity of sequence.[13, 14] The coinducer regions of LTTRs have much more divergent sequences than that of the helix-turn-helix domain, which is likely due to the diversity of inducers across the family.[13, 14] Highly divergent sequences between functionally different LTTRs presumably allows for the accommodation of different specific coinducers. Finally, the C-terminal protion of the protein is implicated in oligomerzation of LTTR monomers. Thus deletions in this conserved region hamper the ability of LTTRs to tetramerize and, as a result, hinder proper binding to DNA.[14]

DgdR structure. Structurally, it is unknown which residues are important to the binding of DgdR to DNA or its inducer Aib, although some structural information maybe obtained by sequence alignments made with other LTTRs. Until recently, crystal structures of LTTRs were available only for the C-terminal fragments of the *E. coli* OxyR and the *Klebsiella aerogenes* CysB proteins.[15, 16] Crystallization of DgdR and most other LTTRs has been unsuccessful due to the insoluble nature of these proteins, leaving until recently the OxyR and CysB fragments as the only structures to reference. These models provide diagrams for how LTTRs are structurally organized around their respective coinducer binding domains, but offer no structural information regarding N-terminal HTH binding domains. However, in 2003 Muraoka et al illucidated the crystal structure of the full CbnR protein from *Ralstonia eutropha* (Figure 3).[17] This is the first LTTR

crystal structure that includes the HTH binding domain and now provides a realistic model for DNA binding by a LysR biomolecule.

Figure 3 shows how the CbnR tetramer complexes with DNA. CbnR binds to DNA at a recognition-binding site (RBS) and an activation-binding site (ABS), which bends the DNA. The ABS sequence contains part of the promoter region for the regulated gene that, when bound by CbnR, prevents the regulated gene from being transcribed. Upon binding with an inducer, the ABS site is released, or partially released and the regulated gene promoter may be accessed by a RNA polymerase for transcription. While there is no crystal structure of a LysR complexed with DNA, the arrangement of the HTH domains, which are the most conserved regions across LTTRs, now allows for modeling of other LTTRs with their deduced binding sites.

Expression of *dgdA*, *dgdR*, and exogenous genes in *E. coli*. The feasibility of expressing genes cloned from a wide range of prokaryotes in *E. coli* has been demonstrated repeatedly.[18] The genera identified in this study: *Bacillus*, *Arthrobacter*, *Corynebacterium*, *Rhodococcus*, *Stenotrophomonas*, *Pseudomonas*, and *Pantoea*, all have genes that have been cloned and expressed in *E. coli*.[8, 19-25] Therefore, it is reasonable to assume that *E. coli* RNA polymerase is capable of recognizing most promoters from these organisms and producing gene products. Furthermore, two genes of interest in this study, *dgdA* and *dgdR* from *B. cepacia*, have already been cloned and expressed in *E. coli*. In the study by Keller et al, the *B. cepacia dgdA* was expressed by *E. coli*, and allowed high level production of the *dgdA* gene product, DgdA.[8] Expression

of *dgdR* was also shown when Allen-Daley et al used the presence and absence of the inducer, Aib, to control the expression of a reporter gene oriented adjacent to a *dgdA* promoter and divergent from the *dgdR* gene.[11] For these reasons *E. coli* was chosen as host organism for cloning and expressing of cloned *dgd* genes.

***dgd* interest.** The primary interest of the dgd metabolic system has shifted from an investigation into mechanism and structure of the enzyme DgdA, to the regulatory protein DgdR. Generation of a structure based on the *B. cepacia* DgdR has been unsuccessful due to the insolubility of the protein.[26] It is unknown at this time whether DgdR from other organisms will prove to be more soluble. Furthermore, the ability of DgdR to regulate gene expression is also of interest. It has already been shown in expression assays to control the expression of a reporter gene oriented adjacent to a *dgdA* promoter.[11] Potentially, the Aib-controlled expression of any gene could be achieved once properly aligned to a *dgdA* promoter and a *dgdR*. This system is especially attractive in that it would be activated by the simple, stable, Aib molecule. This observation led to the patenting of *dgdR*, DgdR and the *dgd* operon.[27, 28]

Besides the interest in understanding the molecular organization and manipulation of DgdR, there is also interest in the role, if any, that *dgd* genes play in soil ecology. At this time it is unknown how and if *dgd* genes are important to bacterial soil ecosystems and how widespread they are. Currently, genome sequencing of various soil organisms has provided the majority of new, putative *dgd* gene sequences and their phylogenetic

distribution. The identification of Aib utilizing organisms in multiple genera is significant as it would suggest a possible ancient lineage of the *dgd* genes.

16S rRNA gene sequencing as a method of microorganism identification. Classical identification of bacteria is performed by a series of systematic tests that reveal the morphological and catalytic traits characteristic of a given organism's genus, species, or strain. If enough tests are performed, the researcher can identify any prokaryote with a high degree of confidence. While there are many tests available, the drawback to this approach is that it is slow and difficult to reproduce.[29] However, with the advent of DNA sequencing technologies, gene sequences common to all bacteria can be compared, and features of an organism's sequence can be used as a basis of identification. The 16S rRNA subunit gene, which is found in all prokaryotes and archaebacteria, has been the gene of choice in most sequence-based identifications. Degenerate primers, such as those designed in a phylogeny study performed by Weisburg et al, amplify an approximately 1500 bp fragment in virtually all prokaryotes.[30] The National Center for Biotechnology Information (NCBI) hosts a database that stores every recorded sequence of DNA that has been published.[31] Their search engine uses the BLAST algorithim[32] and provides the sequences and their source information that are the closest match to a given query, allowing the researcher to identify the organism whose 16S rRNA gene sequence is most similar. Amplified 16S rRNA gene fragments have been the measure of diversity in a number of studies on uncultured bacteria from environmental sources.[33, 34]

Potential Problems. The problems that this research encountered were based primarily in the assumptions made regarding *dgd* containing organisms. Aside from *Burkholderia*, no *dgd* genes have been cloned in *E. coli*. Despite the number of foreign genes that can be expressed in *E. coli*, it is unknown if the *dgd* genes in the organisms in this study are clonable, or if they can be expressed by *E. coli*. Finally, it is also unknown if another system exists for the utilization of Aib as a nitrogen source. Although culturing prokaryotes that utilize Aib without the benefit of a *dgd* system would not provide any new *dgd* containing organisms, it would provide organisms utilizing a novel method for the breakdown of Aib.

Organic Synthesis

An autobiographical note. My interest in synthesis of 2,2-dialkylglycines stems from my experiences in the UAF organic chemistry laboratory course Chemistry 324. In it, I and other students attempted to alkylate various *para*-substituted 4-methyl-2-phenyl-5(4H)-oxazolones **1** at the 4 position. Briefly, this was accomplished by creating a nucleophilic enolate **2** from **1** (Figure 4) and using it in S_N2 reactions with various alkyl halides. It was observed by our class that enolate formation using the *p*-nitro substituted oxazolone **2a** was accompanied by an intense deep blue color. This intense color then faded as the reaction proceeded, a characteristic that suggested the use of color as an indicator of reaction progress.[26] After completing the course, I continued to work on a project investigating the use of **2a** as a reaction indicator in alkylation experiments and as a convenient precursor leading to various dialkylglycine molecules. Later, I worked on the other project described in this thesis, the attemped cloning of *dgd* genes from bacteria isolated from Aib containing media.

The C4-alkylation of 5(4H)-oxazolones has acquired importance as a key step in the synthesis of alpha-substituted dialkyl amino acids.[35, 36] α-Di-substituted amino acids are important substituents of antibiotic peptides,[37] induce and stabilize different types of conformations when incorporated into small peptides,[38] and, in terms of this research, act as inducers of the LysR-type transcription regulator, DgdR, which regulates expression of the dialkylglycine decarboxylase gene *dgdA* in certain bacteria.[11] The arrival at substituted 5(4H)-oxazolones is the acme of a multistep synthesis (Scheme 3) that

proceeds by a) acylation of alanine or serine with an acid halide to an N-acyl amino acid **3**, b) dehydrating cyclization of the N-acyl amino acid to a 5(4H)-oxazolone **4**, and c) alkylation of oxazolone at the 4 position to give **5**. A review of the synthetic steps in Scheme 3 and the background regarding the use of **1a** in synthesis follows. As the compound of interest for this work is **1a,** this review will focus on N-benzoyl amino acid derivatives.

Aside from being used as intermediates in the synthesis of alpha disubstituted amino acids, N-acyl amino acids have been synthesized for other purposes including, for example, as alternate substrates for Acylase I from porcine kidney,[39] and for use in diabetes therapy as potential inhibitors of aldose reductase.[40] Approaches for synthesis of N-benzoyl amino acid derivatives often rely on acid halide chemistry, however methods differ for dealing with the HCl evolved in the course of the reaction.[39, 40] In this study evolved HCl was simply entrained by flow of argon through the reaction vessel, rather than neutralizing HCl with a base as in Schotten-Baumann conditions.

Oxazolones have been synthesized since the late 19th century and have been used for various applications as valuable intermediates in organic synthesis.[41-44] Numerous approaches exist for their synthesis, the most general being the employment of acetic anhydride as the dehydrating reagent.[45, 46] Such harsh reaction conditions are acceptable in synthesis of N-benzoyl derivatives largely due to the stability the phenyl group lends the compound. Other, more labile N-acyl amino acids decompose readily in the presence of acetic anhydride.

Alkylation of 5(4H)-oxazolones generally utilizes an S_N2 substitution mechanism where the 5(4H)-oxazolone enolate participates as a nucleophile and an alkyl halide is used as an electrophile. A review of the subject revealed two approaches that utilized this method, both of which were based on N-benzoyl amino acids and reported good yields of product. The first used phase transfer conditions to facilitate alkylation, with tetrabutyl ammonium chloride acting as the phase transfer catalyst.[35] The advantage of a phase transfer system is that the reaction does not require an inert atmosphere. The second, more involved, approach was performed in DMF solvent relying on inverse addition of a NaH dispersion under argon to generate the desired enolate for alkylation.[36]

Utilization of **1a** as a substrate in the synthesis of heterocyclic systems and synthesis of alpha-substituted serines has been documented.[41-44] Also, intense colors associated with the compound have also been reported. In a study by O'Brien and Niemann, they reported upon addition of **1a** to an aqueous alkali solution an intense red-violet color evolved.[45] However, **1a** was not used in alkylation reactions involving use of its enolate **2a** as a reaction indicator until the present study.

Significance of Dgd research. There are several possible impacts of the research: 1) It would provide multiple organisms that contain *dgd* genes. Identification of these organisms would illustrate how widespread these genes are throughout prokaryotes and illustrate the importance of DgdA metabolism in soil ecology. 2) Sequence information of new *dgdA* and *dgdR* genes would provide the basis for multiple sequence alignments between new genes and the *B. cepacia dgd* cassette. This will provide informative data

on DgdR and LysR-type structure and function. Also, additional *dgdA* sequences gathered from cloning experiments would also increase knowledge of the enzyme it encodes. By comparing the *B. cepacia dgdA* with other functioning 2,2-dialkylglycine decarboxylase gene sequences, residues important to decarboxylation may be better understood. 3) Finally, the flexibility of the DgdR binding site could be probed using different 2,2-dialkylglycines, most of which do not occur in nature. Besides the synthesis of unreported compounds, the reaction scheme developed here permits the alkylation of N-benzoyl substituted oxazolones while providing a built in reaction indicator and would provide a possible key step in the synthesis of different 2,2-dialkylglycines. These compounds in turn could then be used in binding assays with the DgdR protein, or tested as inducers or inhibitors in the regulation of genes engineered to be under the control of a DgdR.

Works Cited

1. Bruckner, H., and Przybylski, M. (1984) *J. Chromatogr. 296*, 263-75.

2. Cronin, J. R., and Pizzarello, S. (1997) *Science 275*, 951-5.

3. Aaslestad, H. G., and Larson, A. D. (1964) *J. Bacteriol. 88*, 1296-303.

4. Bailey, G. B., and B., D. W. (1967) *Biochemistry 6*, 1526-32.

5. Toney, M. D., Hohenester, E., Cowan, S. W., and Jansonius, J. N. (1993) *Science 261*, 756-9.

6. Lamartiniere, C. A., Itoh, H., and Dempsey, W. B. (1971) *Biochemistry 10*, 4783-4787.

7. Dempsey, W. B. (1969) *J. Bacteriol. 97*, 182-5.

8. Keller, J. W., Baurick, K. B., Rutt, G. C., O'Malley, M. V., Sonafrank, N. L., Reynolds, R. A., Ebbesson, L. O., and Vajdos, F. F. (1990) *J. Biol. Chem. 265*, 5531-9.

9. Toney, M. D., Hohenester, E., Keller, J. W., and Jansonius, J. N. (1995) *J. Mol. Biol. 245*, 151-79.

10. Hohenester, E., Keller, J. W., and Jansonius, J. N. (1994) *Biochemistry 33*, 13561-70.

11. Allen-Daley, E., Sun, H., Bray-Hallan, S. T., and Keller, J. W. (2002). Manuscript in preparation.

12. Henikoff, S., Haughn, G. W., Calvo, J. M., and Wallace, J. C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6602-6.

13. Schell, M. A. (1993) *Annu. Rev. Microbiol. 47*, 597-626.

14. Lochowska, A., Iwanicka-Nowicka, R., Plochocka, D., and Hryniewicz, M. M. (2001) *J. Biol. Chem. 276*, 2098-107.

15. Choi, H., Kim, S., Mukhopadhyay, P., Cho, S., Woo, J., Storz, G., and Ryu, S. (2001) *Cell 105*, 103-13.

16. Tyrrell, R., Verschueren, K. H., Dodson, E. J., Murshudov, G. N., Addy, C., and Wilkinson, A. J. (1997) *Structure 5*, 1017-32.

17. Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K., and Senda, T. (2003) *J. Mol. Biol. 328*, 555-66.

18. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998) *Chem. Biol. 5*, R245-9.

19. Brady, S. F., Chao, C. J., Handelsman, J., and Clardy, J. (2001) *Org. Lett. 3*, 1981-4.

20. Baurick, K. B. Cloning and Expression of Pseudomonas Cepacia 2,2-Dialkylglycine Decarboxylase in Escherichia Coli. M.S., University of Alaska, Fairbanks, AK, 1987.

21. Brandsch, R., Faller, W., and Schneider, K. (1986) *Mol. Gen. Genet. 202*, 96-101.

22. Bresler, M. M., Rosser, S. J., Basran, A., and Bruce, N. C. (2000) *Appl. Environ. Microb. 66*, 904-8.

23. Chou, J. C., Mulbry, W. W., and Cohen, J. D. (1998) *Mol. Gen. Genet. 259*, 172-8.

24. Trott, S., Burger, S., Calaminus, C., and Stolz, A. (2002) *Appl. Environ. Microb. 68*, 3279-86.

25. Rondon, M. R., Raffel, S. J., Goodman, R. M., and Handelsman, J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 6451-5.

26. Keller, J. W. (1999). Personal Communication.

27. Keller, J. W. (1993). *Repressor Protein Gene for Regulating Expression of Polypeptides and its use in the Preparation of 2,2-Dialkylglycine Decarboxylase of Pseudomonas cepacia.* Patent No. 5210025.

28. Keller, J. W. (1994). *Repressor Protein and Operon for Regulationg Expression of Polypeptides and its use in the Preparation of 2,2-Dialkyglycine Decarboxylase of Pseudomonas cepacia.* Patent No. 5356796.

29. Guerrero, R. (2001) *Int. Microbiol. 4*, 103-9.

30. Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991) *J. Bacteriol. 173*, 697-703.

31. http://www.ncbi.nlm.nih.gov/.

32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403-10.

33. Dunbar, J., Takala, S., Barns, S. M., Davis, J. A., and Kuske, C. R. (1999) *Appl. Environ. Microb. 65*, 1662-9.

34. Hiorns, W. D., Methe, B. A., Nierzwicki-Bauer, S. A., and Zehr, J. P. (1997) *Appl. Environ. Microb. 63*, 2957-60.

35. Gelmi, M. L., Pocar, D., and Rossi, L. M. (1984) *Synthesis*, 763-5.

36. Obrecht, D., Bohdal, U., Ruffieux, R., and Muller, K. (1994) *Helv. Chim. Acta 77*, 1423-9.

37. Brewer, D., Mason, F. G., and Taylor, A. (1987) *Can. J. Microbiol. 33*, 619-25.

38. Obrecht, D., Spiegler, C., Schönholzer, P., and Müller, K. (1992) *Helv. Chim. Acta 75*, 1666-95.

39. Chenault, H. K., Dahmer, J., and Whitesides, G. M. (1989) *J. Am. Chem. Soc. 111*, 6354-64.

40. Benvenuti, S., Severi, F., Costantino, L., Vampa, G., and Melegari, M. (1998) *Il. Farmaco. 53*, 439-42.

41. Kamiñski, Z. J., Leplawy, M. T., and Zabrocki, J. (1973) *Synthesis*, 792-3.

42. Sain, B., Baruah, J. N., and Sandhu, J. S. (1982) *J. Heterocycl. Chem. 19*, 1511-4.

43. Sain, B., and Sandhu, J. S. (1986) *J. Heterocycl. Chem. 23*, 1007.

44. Sain, B., Singh, S. P., and Sandhu, J. S. (1992) *Tetrahedron 48*, 4567-78.

45. O'Brien, J. L., and Niemann, C. (1957) *J. Am. Chem. Soc. 79*, 80-5.

46. Song, S., Jung, B., Lee, K., and Kim, H. (1994) *Bull. Korean Chem. Soc. 15*, 520-1.

Figure 1.

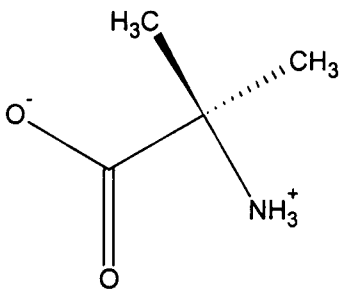
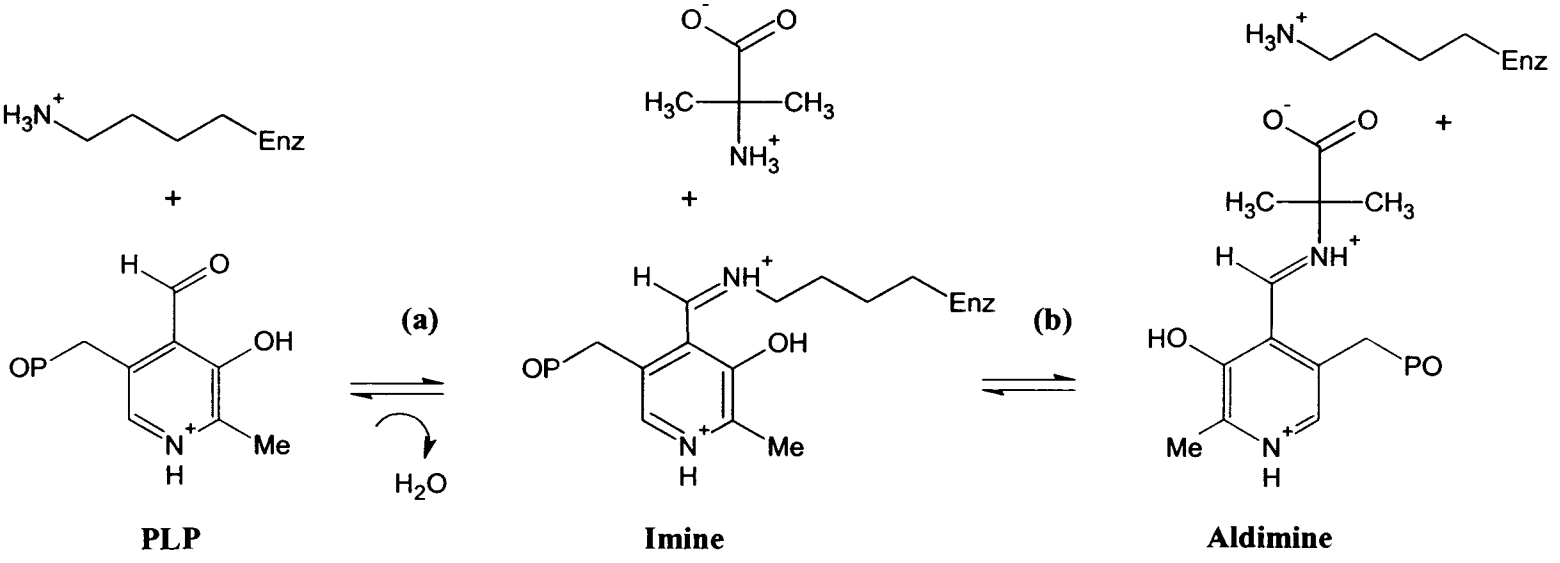


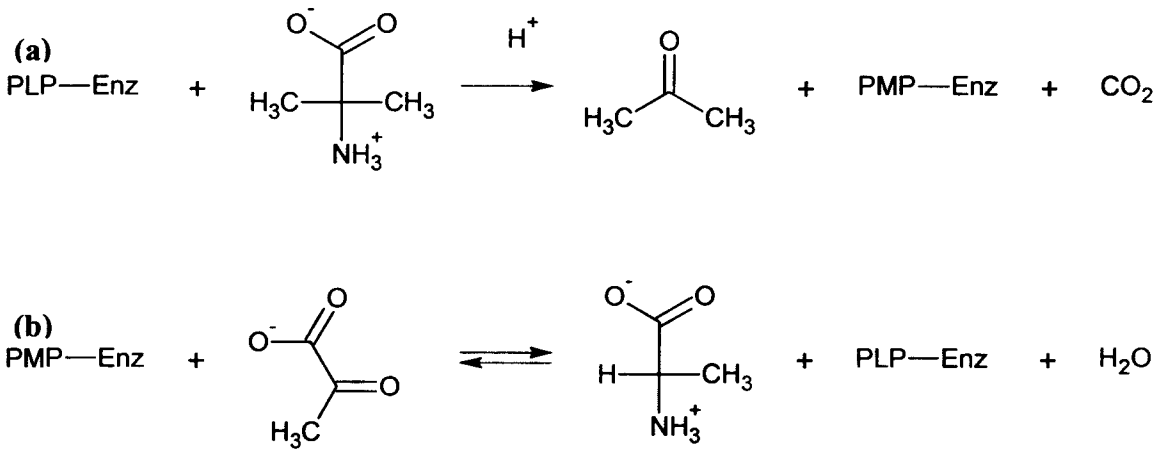
Figure 1. α-Aminoisobutyric acid (Aib).

Scheme 1.



Scheme 1. The interaction between PLP and DgdA. (a) Imine formation between the enzyme and the PLP cofactor. (b) External aldimine formation between PLP and the substrate, in this case. Aib.

Scheme 2.



Scheme 2. Reactions catalyzed by DgdA. (a) The irreversible decarboxylation reaction. (b) The reversible transamination reaction.

Figure 2.

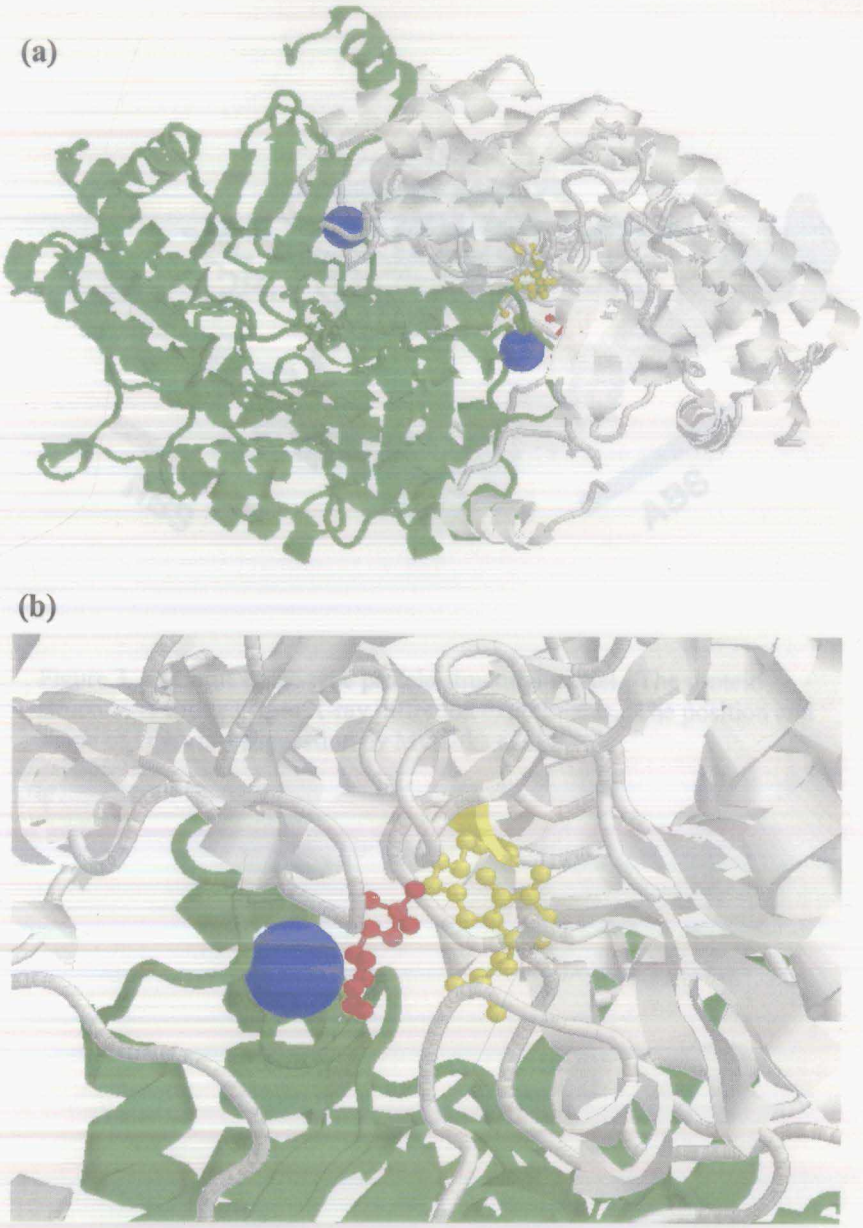


Figure 2. The DgdA dimer structural model.[9] (a) Shows the interaction between monomers. (b) Shows the active site. Morpholino-ethanesulfonic acid (MES), which mimics the substrate is red; the cofactor PLP bound to Lys272, is yellow; and potassium, which is shown larger in proportion to MES and PLP, is blue.

Figure 3.

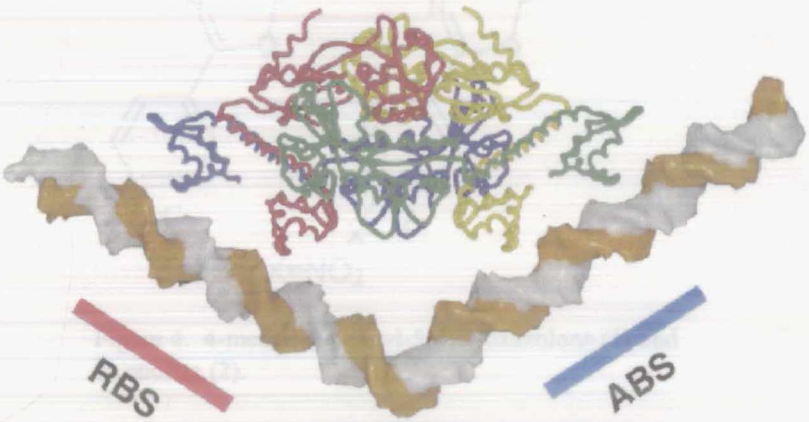


Figure 3. A CbnR LysR-type protein structural model. The protein structure is derived from X-ray diffration experiments. The position and shape of DNA is a suggestion by Muraoka et al.[17]

Figure 4.

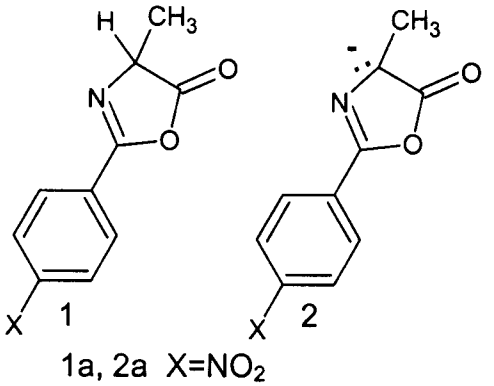
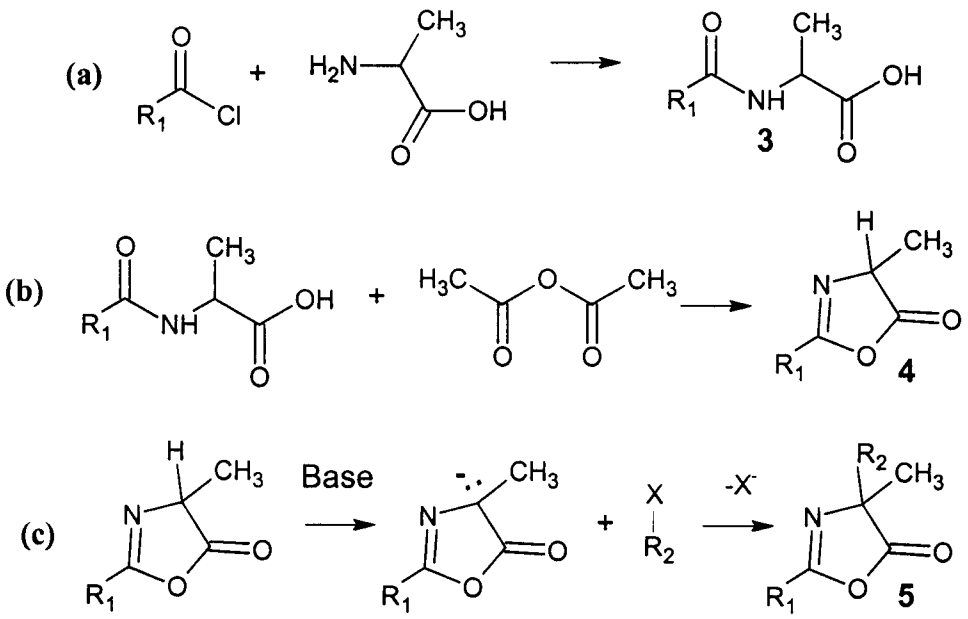


Figure 4. 4-methyl-2-phenyl-5(4H)-oxazolone **(1)** and its enolate **(2)**.

Scheme 3.



Scheme3. Synthesis of substituted oxazolone. (a) Amidization of an amino acid with an acid halide. (b) Cyclization of the N-acyl amino acid to an oxazalone. (c) Alkylation of the oxazolone at the 4 position. The R_1 and R_2 substituents used in this study are shown in Scheme 4 in Chapter 3.

Chapter 2. Molecular Biology

Introduction

The goal of this study was to isolate Aib metabolism genes from soil bacteria by cloning DNA in a host organism that is dgd negative. This seemed feasible for two reasons: 1) the *dgdA* and *dgdR* genes from the soil microorganism *B. cepacia* have already been cloned and conveyed the dgd+ phenotype to the *E. coli* host organism, and 2) organisms cultured from soil for this project that displayed the dgd+ phenotype also have many examples of genes expressed in *E. coli*. This suggested that *E. coli* RNA polymerase was capable of recognizing most promoters and generating gene products from the bacterial species identified by this study. In the event, a number of soil bacteria were identified that apparently utilize Aib, however a concerted attempt to clone dgd genes was not successful.

Materials and Methods

Miscellaneous. All chemicals and media used were of reagent grade or better. Water used was purified to resistivity greater than 10 MΩ-cm using a Milli-Q (type I) reagent grade water purification system. Microcentrifugations were performed in an Eppendorf centrifuge models 5414 or 5415. All glassware, media, and buffers were autoclaved in a Tuttnauer Brinkmann 3870 model autoclave for 20 min at 120 psi and 120°C. Non-autoclaveable media and antibiotics were filtered through VWR 20-μm membrane filters. Agarose gel electrophoresis was performed in a Bio rad wide mini-sub cell where voltage was applied with an EC-103 power supply (E-C Apparatus Corporation).

The DNA standard used in all electrophoresis gel runs was the Promega 1-kb DNA ladder which contained markers 10 kb, 8 kb, 6 kb, 5 kb, 4 kb, 3 kb, 2.5 kb, 2 kb, 1.5 kb, 1 kb, 0.75 kb, 0.5 kb, and 0.25 kb long. The *B. cepacia* strain and d_6-Aib were obtained from Dr. John Keller. The *Mycobacterium smegmatis* strain ATCC# 19420 was obtained from Keith Derbyshire at the New York State Department of Health Wadsworth Center.

Simulations of restriction digestions were performed *in silico* with the web program Web Cutter.[1] Assembly of overlapping sequencing products was performed on the CAP: Assemblage de séquences á INFOBIOGEN web site.[2] Multiple sequence alignments and construction of plasmid models were carried out on the OMIGA computer program (Oxford Molecular Group). BLAST and open reading frame (ORF)

searches on DNA sequences were performed at the National Center for Biotechnology Information (NCBI) website.[3]

Media: Organisms were cultured with selective and non-selective media including;

Aib medium: 3.8×10^{-3} g/L $FeCl_2\bullet4H_2O$, 3.6×10^{-2} g/L $MgCl_2\bullet6H_2O$, 9.5 g/L $K_2HPO_4\bullet3H_2O$, 3 g/L KH_2PO_4, 4 g/L Aib, 53.5 g/L NaOAc (or 40 g/L glycerol), and 87 mg/L K_2SO_4. Potassium salts in this media were autoclaved separately and the glycerol was filter sterilized.

d_6-Aib medium: Identical to Aib medium with d_6-Aib replacing the Aib.

NH_4 medium: Identical to Aib medium with 1 g/L of ammonium chloride replacing the Aib.

Nutrient broth (NB): 8 g/L powder (Difco).

Terrific broth (TB): 12 g/L bacto-tryptone, 24 g/L bacto-yeast extract, 5 g/L glycerol, 12.54 g/L $K_2HPO_4\bullet3H_2O$, 3 g/L KH_2PO_4, with potassium salts autoclaved separately.

Luria-Bertani broth (LB): 10 g/L tryptone, 5 g/L yeast, and 10 g/L NaCl. LB/amp contained additional 100 μg/mL ampicillin. In identifying recombinant clones LB/amp/X-gal/IPTG plates were overlaid with 40 μL of 2% X-gal in DMF and 7 μL of 20% IPTG in water and allowed to dry prior to spreading the plates with outgrowth.

M9/Aib medium: Per liter, to 750 mL of sterile H_2O (cooled to 50°C or less), add; 200 mL 5×M9/Aib salts, 2 mL 1M $MgSO_4$, 2 g/L glucose, 100 μL 1M $CaCl_2$, 1 mL

100 mg/mL leucine, 10 µL 100 mg/mL thiamine, sterile deionized water to 980 mL. 5×M9/Aib salts: 64 g/L $Na_2HPO_4 \cdot 7H_2O$, 15 g/L KH_2PO_4, 2.5 g/L NaCl, 2.0 g/L Aib.

M9/NH_4 medium: Identical to M9/Aib medium with 1 g/L of ammonium chloride replacing the Aib.

SOC broth: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 10 mL 250 mM KCl. Add sterile deionized water to above to 900 mL. Adjust pH of above to 7.0 and add water to 970 mL. Autoclave for 30 minutes, cool to RT, and add 10 mL sterile 1 M $MgCl_2$ and 20 mL filter-sterilized 1 M glucose.

Collection of soil. Soil was collected from 10 separate sites in the Fairbanks North Star Borough. Due to the variety of soil types, the method of soil collection was limited to the avoidance of plant materials and other inclusions such as rocks. Typically, soil was scooped into a plastic bag with a sterile spoon and stored at 4°C.

Enumeration and selection of bacteria that may utilize Aib as the sole nitrogen source. From each sample, 1 gram of soil and 50 mL of Aib broth was added to a sterile, foil-capped 125 mL Erlenmeyer flask. Once inoculated, flasks were then either shaken at 200 rpm at 37°C or incubated at room temperature with no shaking. Two dilutions of Aib cultures were performed to remove microorganisms from any nitrogen sources provided from the soil. This was done by removing 1 mL of culture after two days and transferring it to fresh Aib broth and repeating again after two days.

Isolation of Soil Microorganisms. Organisms cultured in Aib broth were isolated by either streaking or spreading culture onto plates containing Aib agar, terrific broth agar (TB), Luria-Bertani agar (LB), or Nutrient agar (NA). Isolated colonies were transferred to terrific broth for enumeration and plated again onto Aib agar plates for identification. An organism was considered an Aib metabolizer if, under these conditions, it grew on the final Aib plate.

Isolation of genomic DNA. Isolation of DNA from Aib utilizing organisms was performed with Qiagen DNeasy Tissue Kits using the manufacture's protocol for isolation of bacterial DNA. Isolates were grown overnight in culture tubes containing LB, then 2×10^9 cells were removed. The volume of broth that contained 2×10^9 cells was calculated by measuring the OD_{600} of culture aliquots on a Hewlett Packard 8452A diode array spectrophotometer. According to the manufactures handbook, an OD of 1 corresponded to a concentration of 1×10^9 cells per mL of *E. coli*. Isolated DNA was electrophoresed in 1% agarose gels containing ethidium bromide and the resulting bands were visualized by an Alpha Imager 2200 (Alpha Innotech) system under UV transillumination. Quantification of DNA was performed by comparing the fluorescent intensities of bands to that of the 1-kb DNA ladder. The basis of the comparisons were the integrated band intensities calculated by the Alpha Imager 2200 software.

Identification of bacterial isolates via 16S rRNA gene sequencing. Degenerate primers designed to amplify a ~1500 bp fragment of 16S rRNA genes from a diverse

range of bacterial genera were used in amplification and cycle sequencing reactions; forward primer: JAB1 5'-AGAGTTTGATCCTGGCTCAG-3', reverse primer: JAB2 5'-ACGGCTACCTTGTTACGACTT-3'.[4] All polymerase chain reactions (PCR) were performed on a P18282 Perkin Elmer DNA thermal cycler. A typical amplification was performed in a 500-μL microcentrifuge tube. The reactions were assembled in tubes on ice, and included 100 ng of bacterial genomic DNA, 1 μL of 50 μM forward primer stock solution, 1 μL of 50 μM reverse primer stock solution, 0.5 μL of Failsafe enzyme (Epicentre), 15 μL of Failsafe pre-mix J (Epicentre), sterile deionized water to a total reaction volume of 30 μL, and 30 μL of mineral oil. The reaction tube was briefly spun in a microcentrifuge to fully separate phases and was introduced to the thermal block at 80°C. The following cycles were carried out: one cycle of 2 min at 95°C; 30 cycles of 1 min at 95°C, 1 min at 54°C, 1.5 min at 72°C; and one cycle of 4 min at 72°C. In some cases following the final cycle the reaction was held at 4°C overnight, or until the samples could be stored at -20°C.

Aliquots of PCR amplification reactions were electrophoresed and visualized under UV on 1% agarose gels containing ethidium bromide. Gels were typically photographed and fluorescing bands were compared to a 1-kb ladder (Promega) and analyzed as the genomic DNA above. PCR reactions showing bands migrating at approximately 1500 bp were chosen for cycle sequencing reactions. Primers, primer dimers, and Taq polymerase were removed from the PCR amplification reactions by QIAquick PCR purification kits (Qiagen).

Sequencing of amplified fragments was performed by using the ABI Prism BigDye Terminator v3.0 Ready Reaction Cycle Sequencing Kit (Applied Biosystems). A typical cycle sequencing reaction was performed by adding to a 500-μL Eppendorf tube on ice: 4.0 μL of Terminator Ready Reaction Mix, 40 ng of purified PCR reaction product, 3.2 pmol of primer, deionized water to 20 μL, plus 40 μL of mineral oil. After spinning briefly in a microcentrifuge, the tube was introduced to the PCR thermal block at 80°C and was subjected to the following thermal cycles: 30 sec at 96°C, 15 sec at 50°C, and 4 min at 60°C, all for 25 cycles. If necessary, after 25 cycles the reaction was held at 4°C overnight, or until it could be stored at -20°C.

Extension products were purified by using Centri-sep spin columns (Princeton separations) loaded with a Sephadex G-50 slurry. The slurry was made by mixing 10 g of fine grade G-50 powder and transferring it to a 250 mL Erlenmeyer flask with stir-bar. To this was added 150 mL of deionized water and the resulting slurry was stirred for 20 min. Centri-sep columns were loaded with 650 mL of slurry, placed in collection tubes and spun for 3 min at 2500 rpm in a microcentrifuge. The resulting G-50 column was loaded with only the aqueous phase of the cycle sequencing reaction mixture. The mineral oil layer was removed prior to Centi-sep purification by freezing the entire reaction mixture at -70°C and aspirating the oil layer, leaving the frozen aqueous phase. The column, loaded with the extension products and placed in a fresh collection tube, was spun for 3 min at 2500 rpm. The resulting eluate was evaporated under vacuum in a centrivap concentrator (Labconco) and the pellet (often invisible) was stored at -20°C.

Pellets were sent to the UAF Core facility for nucleic acid analysis where sequencing was performed on an ABI 3100 Genetic analyzer equipped with a 50- or 80-cm, 16-pin capillary array. Resulting sequences were provided in the form of electropherogram files that were analyzed and edited using the Chromas computer program v.2.13 (Technelysium). Sequences were submitted to the National Center of Biotechnology Information (NCBI) sequence database in a BLAST search and the top five matches were recorded.[3] The sequence length used in the BLAST search was dependent on the quality of sequence obtained for the forward and reverse primed cycle sequencing reactions. The quality sequence region was defined as the region containing unambiguous base calls, or ambiguous calls that could be corrected through a visual examination of the electropherogram. Ambiguous calls occur when the basecaller program cannot assign a base at a given position based on the raw data collected from the sequencer. Quality sequence extending from both ends that allowed for good sequence overlap was used to construct contigs that were used in BLAST searches. If sequence in both directions could not be obtained, then sequence in the one direction was used. Contig construction of overlapping sequences were put together using the OMIGA sequence alignment computer program v-1.1 (Oxford Molecular).

Construction of plasmid libraries from genomic DNA.

Partial restriction digestion of genomic DNA of Aib utilizing organisms. To generate clonable fragments 4-6 kb in size, genomic DNA isolated from one soil microorganism or pooled from several organisms was partially digested by the enzyme Sau3A I

(recognition and overhang 5' GATC 3'). Before digesting genomic DNA, a scaled down optimization reaction was performed. A typical optimization reaction was conducted as follows: To a 500-μL Eppendorf tube on ice was added 5 μl of 10x restriction buffer, 5 μL of 10x BSA, approximately 500 ng of DNA, 2 μL of 0.125 U/μL Sau3A I restriction enzyme (Promega), and deionized water to a final volume of 50 μL, the mixture was spun briefly and incubated at 37°C. To each of nine 500-μL vials was added 5 μl aliquots from the incubating reaction mixture taken at t = 0, 5, 10, 15, 20, 30, 45, 60, 90, and 120 min, and were transferred to a 70°C bath for 15 min to kill remaining enzyme activity. All aliquots were loaded onto an ethidium bromide-containing 1% agarose gel, run at 70 V for 70 min, and visualized under UV. The reaction time that maximized the amount of partially digested DNA in the desirable 4 to 6 kb range was used subsequently in a scaled-up reaction.

Because the genomic DNA and the reagents used in restriction digestion are stored frozen, and multiple freeze-thaw cycles were found to negatively effect the reproducibility of a scaled-up reaction (data not shown), reagents were stored on ice between an optimization and a scaled up reaction. A typical scaled up reaction was performed immediately following the visualization of DNA from the optimization reaction on an electrophoresis gel. A typical reaction was performed as follows: To a 500-μL Eppendorf tube on ice was added 30 μL of 10x restriction buffer, 30 μL of 10x BSA, approximately 30 μg of DNA, 12 μL of 0.125 U/μL Sau3A I restriction enzyme, and deionized water to a final volume of 300 μL. The mixture was spun briefly and incubated at 37°C for the time determined in the optimization reaction. The tube was

then transferred to 70°C for 20 min to inactivate the restriction enzyme. The digested DNA was precipitated with 0.5 vol of 3M sodium acetate (pH 5.2) and 3.5 vol of 95% ethanol, briefly vortexed, stored at -70°C overnight, and spun at maximum speed in a microcentrifuge for 30 min. While visualizing the pellet, the supernatant was carefully removed with a micropipette, the pellet washed with 1 mL of 70% ethanol, and spun at maximum speed in a microcentrifuge for 15 min. After again carefully removing the supernatant the pellet was allowed to dry for 5 min under vacuum in a centri-vap concentrator and was resuspended in 30 μL of deionized water.

Collection of 4 to 6 kb restriction fragments. Restriction fragments three to six kb in length were collected as follows. To the 30 μL sample of restriction digested fragments, 6 μL of 6x loading dye was added. This 36 μL solution was loaded onto a 0.7% agarose gel. The gel ran for 30 min at 70 V in parallel with a 1 kb ladder, and was soaked in a 1 μg/mL ethidium bromide solution for 15 min. Taking care not to expose the gel to UV light for more than 30 sec, the section of gel containing 3-6 kb fragments, as corresponding to the 1 kb ladder, was sliced out and put in a 1.5-mL Eppendorf tube.

Elution of DNA from the gel slice was performed using a protocol derived from Kurien et al.[5] An Eppendorf tube containing the gel slice and a micro-pestle (Kontes) along with a 2-mL Eppendorf tube containing 0.1×concentrated TE (10 mM Tris pH = 8.0, 1mM EDTA) were placed in a 45°C bath and allowed to warm for 5 min. The gel slice was first crushed by the pestle and 100 μL of the 45°C TE was added to it. The crushed gel was homogenized for 1 min by the pestle. The tube was then spun for 30 s at

maximum speed in a microcentrifuge and the supernatant was saved. This was repeated once with another 100 μL of the TE and the two supernatants were combined. A 5 μL aliquot was taken and loaded on a 1% agarose gel and run at 70 V for 70 min in parallel with a 1-kb ladder to confirm the proper size, and to calculate the amount of DNA eluted. The volume of the DNA solution was reduced by spinning it under vacuum in a 261931 centri-vap concentrator (Labconco) and was stored at -70°C. This was repeated for other partial restriction digestion products until fractions pooled constituted 120 ng of restricted DNA in 15 μL of solution, enough product to perform ligations.

Ligation of partial digestion fragments into pUC19 and transformation of electrocompetent *E. coli* with ligation products. Pooled partially-digested restriction fragments were precipitated with sodium acetate and 95% ethanol and resuspended in an appropriate volume so that the concentration of the resulting DNA solution was at least 20 ng/μL. Ligation of DNAs to dephosphorylated, BamHI-digested pUC19 (Fermentas) was performed with the Epicentre Fast Link DNA ligation kits. An aliquot of ligation products was analyzed on an ethidium bromide-containing 1% agarose gel.

Transformation of electrocompetent *E. coli* with ligation products was performed by electroporating Epicentre TransforMax EC100 electrocompetent cells with a Bethesda Research Laboratories Cat Series 1600 Cell-Porator. A typical electroporation was performed as follows: To a 500-μL Eppendorf tube on ice was added 19 μL of electrocompetent cells and 1 μL of ligation mixture and then briefly spun in a microcentrifuge. This cell suspension was added a poration cuvette on ice by carefully

pipetting so it was suspended between the cuvette bosses. The cell suspension was electroporated with 2.06 kV and immediately transferred to an 8 in. culture tube containing 2 mL of SOC broth and incubated for 1 h at 37°C. Fifty to 100 μL of outgrowth was then spread onto M9/Aib plates and LB/amp plates; the latter were overlaid with 40 μL of 2% X-gal in DMF and 7 μL of 50 mM IPTG. All inoculated plates were then incubated at 37°C overnight, or until colonies of 1-2 mm in diameter appeared.

Amplification and storage of plasmid libraries. Amplification and storage of plasmid libraries was performed by the method reported of Vogeli.[6] To a sterile Erlenmeyer flask held at 45°C in a water bath was added outgrowth from electroporation experiments and 45°C molten, sterile agarose in LB broth. These were mixed so the final agarose concentration would be 0.7% by weight and the final titer would be ~1200 CFU/mL. From this mix, 2.5 mL aliquots were poured onto 100-mm LB/amp plates, were allowed to solidify, and were incubated at 37°C overnight. Large, unwanted colonies growing on the surface of the top agarose were removed prior to agarose scraping by flushing with LB broth and gentle rubbing with a glass rod. Surface colony cells suspended in the LB broth were collected with a glass pipette and stored in 12% glycerol at -70°C. The colony-containing top agarose was scraped off with a sterile spatula and collected in the barrel of a sterile, plastic 60 mL syringe. Collected top agarose was mixed with an equal volume of LB broth and passed through syringe needles of decreasing size until the resulting paste could pass through a 22 gauge needle. This paste was mixed with 0.5 vol

of sterile G-25 Sephadex (Sigma) slurry autoclaved in LB broth. To several Econo columns (Bio-Rad), containing a 2 mL bed of Sephadex G-25 slurry, approximately 12 mL of slurry-paste mix was added and the entire column was placed in a 15 mL centrifuge tube. Tubes were centrifuged at 1000 rpm in an IEC Centra CL2 tabletop centrifuge equipped with a fixed angle rotor. Eluted cell pellets were resuspended in 0.25 mL of LB, and were stored at -70°C in 12% glycerol.

Characterization of plasmid libraries. By virtue of ampicillin resistance derived from pUC19, colonies of plasmid containing *E. coli* were identified by their growth on LB/amp plates. Because of α-complementation of β-galactosidase fragments between non-recombinant plasmids and their *E. coli* hosts, non-recombinant plasmid-containing colonies on plates overlaid with X-gal were identified by their blue color, while recombinant colonies were identified by their white color.

The average size of the recombinant plasmids was determined by isolating plasmid from 6 to 14 randomly selected white colonies. Randomly picked colonies were inoculated in 2 mL cultures of TB/amp in eight in. test tubes, shaken at 250 rpm overnight at 37°C, and 1.5 mL of culture was spun down for plasmid DNA isolation. Plasmid DNA was isolated using Bio-rad Plasmid Mini prep kits and was performed by following the kit protocol. A 3 μL aliquot of each plasmid DNA isolated was digested with Hind III and EcoR I and was analyzed on a 1% agarose gel. The AlphaImager 2200 documentation and analysis system was used to infer fragment sizes of the UV-visualized bands in comparison to the 1-kb ladder. Total size of the insert was calculated by adding

up all the fragments and subtracting 2600 bp. Total insert sizes were recorded and the mean size of all the inserts was calculated.

Possible *dgd* containing clones were identified by directly plating electroporation outgrowth onto M9/Aib plates, or by replica plating recombinant colonies from TB/amp plates onto M9/Aib plates and incubating at 37°C. A typical replica plate was made using a replica plating block equipped with a fresh, sterile 15 cm×15 cm square of velveteen fixed to the block by a collar. A TB/amp plate with recombinant colonies was then placed upside down on the block so that the colonies on the plate were in direct contact with the velveteen. After removing the TB/amp plate, a fresh M9/Aib plate was placed against the inoculated velveteen. This transferred all colonies from the TB/amp plate that contacted the fabric to the M9/Aib plate. Replica plates were incubated at 37°C and checked daily for growth for one week. Colonies that grew up on M9/Aib plates were transferred to fresh M9/Aib plates, and TB/amp culture tubes.

Sequencing of inserts. Plasmids found to convey the DGD phenotype to *E. coli* were treated with the EZ::TN <Kan-2> (Epicentre) transposon insertions. Following the kit protocol, electrocompetent *E. coli* were transformed with transposon-containing plasmid DNA and outgrowth from the electroporation was plated on kanamyacin-containing LB plates, and incubated at 37°C overnight. Single colonies were picked, grown in a kanamyacin-containing supplemental broth, and the plasmid DNA was isolated. This plasmid DNA was used in cycle sequencing reactions with primers designed to complement the inserted transposon and sequence away from it in opposite directions. A

typical reaction was performed by adding to a 500-μL microfuge tube on ice: 500 ng of plasmid, 5 μL of terminator ready reaction mix (Applied Biosystems), 3.2 pmol of primer, and deionized water to 20 μL. The tube was introduced to the thermal block at 80°C and participated in cycle sequencing thermal cycles as above. Sequences derived from cycle sequencing experiments were assembled into contigs constituting the entire plasmid insert using the contig builder CAP : Assemblage de séquences à INFOBIOGEN.[2]

Confirmation of dgd phenotype.

Growth in Aib-containing media and assays for the Aib metabolism product acetone. To a sterile, 500-mL Erlenmeyer flask was added 50 mL of either Aib, NH_4, M9/NH_4, or M9/Aib medium. The medium was innoculated with 200 μL of frozen, preserved cells that had been thawed on ice. Cultures were grown at 37°C with 200 rpm shaking; after 12 h, or upon achieving an OD_{600} greater than 0.2, 2 mL of media were transferred to two, sterile 500 mL Erlenmeyer flasks each containing 50 mL of fresh media. At 1-7 h intervals 1.5 mL aliquots from both flasks were taken, 750 μL of which was stored in 2 mL microcentifuge tubes on ice at 4°C. The other 750 μL was diluted with an equal amount of deionized water and the OD_{600} was measured. An exception was *M. smegmatis*, cutures which were not assayed for cell density, only acetone production. The *M. smegmatis* grew only in floating clumps that could not be suspended in media. *M. smegmatis* in Aib medium was therefore grown at 37°C without shaking and aliquots were taken at 6-12 h intervals and stored on ice for later acetone determination.

When growth entered stationary phase, the aliquots stored on ice were assayed for acetone using GC-FID. Microcentrifuge tubes containing culture aliquots were centrifuged at 14,000 rpm for 2 min and 200 μL of the supernatant was transferred to a 1.5-mL gas chromotography (GC) sample vial. Flame ionization detection (FID) analysis of vial headspace (50 μL) was performed on a Hewlett Packard 6890 series GC system equipped with an Altech EC-1 30-m column with 0.25-μm of stationary film. Samples were run at a column temperature of 40°C, a front inlet temperature of 150°C, and a mobile phase (He) flow of 1.5 mL/min. Peak integration and analysis of spectra were performed by the Hewlett Packard HP GC Chemstation software.

Confirmation that acetone production was due to Aib metabolism was obtained by growing organisms in media that contained d_6-Aib as a sole nitrogen source. GC mass spectrum (GC-MS) analysis of d_6-Aib culture headspace was performed on a Hewlett Packard 5890 series II plus gas chromatograph with a Hewlett Packard 3972 series mass selective detector. Samples for analysis by GC-MS were prepared as above, except 100 μL of culture supernatant was added to GC vials instead of 200 μL. Aib medium spiked with d_6-acetone and acetone were prepared as standards.

Conversion of peak areas to acetone concentrations. To each of five 1.5 mL microcentrifuge tubes was added 1 mL of Aib broth containing increasing concentrations of acetone. The tubes were centrifuged for 2 min at 14,000 rpm and 200 μL of the supernatants were transferred to 1.5 mL GC vials and treated as above. Final acetone concentrations measured in vials were 0.68, 1.4, 2.0, 2.7, and 3.4 mM.

Results

This section contains the background, rational, and results of experiments directed toward cloning *dgd* genes from the cultured soil organisms. This includes culturing *dgd*+ soil organisms, identifying them by 16S rRNA gene sequencing, analysis of Aib metabolism using GC-FID and GC-MS, attempted cloning of *dgd* genes, and analysis of the recombinant DNA isolated from clones grown on an Aib containing screening medium.

The final sections of the results will focus on sequence analysis of the *dgd* and putative *dgd* genes and their gene products. A multiple sequence alignment of DgdA and putative DgdA sequences is analyzed to elucidate residues important to the decarboxylase function. A second alignment is performed on the *B. cepacia* DgdR and putative DgdR proteins to identify conserved residues. Finally, a diagram constructed of the intergenic region between the *B. cepacia dgdA* and *dgdR* genes is presented along with an annotated multiple sequence alignment of this region across all *dgd* gene and putative *dgd* gene containing organisms.

The isolation and identification of Aib-utilizing organisms was performed by the author and Tammy Thompson. Results presented in Tables and Figures attributed specifically to Tammy Thompson are noted.

Culturing *dgd*+ soil organisms. When Aib-containing minimal medium was inoculated with soil from various Fairbanks sites, many distinct microorganisms were observed (Table 1). A microorganism was designated as distinct based on colony morphology (e.g.

a pink colony is distinct from a yellow one) when the organism was grown on TB agar plates. Of the 69 distinct organisms, 43 were selected for use in cloning experiments and stored in 12% glycerol at -70°C. Under light microscopy, ten isolates; 1, 6, 7, 14, 15, 18, 24, 29, 35, and T12, by virtue of their size and morphological similarities to fungi, were determined to be eukaryotic and were not further characterized, leaving 33 preserved soil organisms.

Genomic DNA preparation. In order to perform PCR experiments and construct a plasmid library, genomic DNA was required from the organisms of interest. DNA was extracted using the DNeasy Tissue Kit (Qiagen) that provides DNA with a typical fragment size of 30-50 kb. DNA of this length was important because partial restriction of genomic DNA to ~ 4-6 kb fragments for use in cloning was employed (details appear later in this section). When partially restricting genomic DNA, the starting material should be 3 to 8 times longer than the desired insert length.[7, 8] The reason is that once digested, restricted DNA that is too small contains many fragments that are restricted on one end and sheered on the other, thus making them unclonable. Due to the presence of one compatible sticky end, these segments also act as competitive inhibitors during ligation reactions. DNA could not be isolated from organisms 12, 20, and 32, for unknown reasons. These were not further characterized, leaving 30 stored organisms with accompanying extracted DNA stocks.

Sequencing 16S rRNA genes. As previously mentioned, 16S rRNA gene sequencing was the principal strategy for identifying the prokaryotes cultured in this study. To provide template for cycle-sequencing reactions, degenerate forward and reverse primers (JAB1 and JAB2, respectfully) were used to amplify a ~ 1500 bp fragment from the 16S rRNA gene. These primers were based on oligos fD1 and rP2, described by Weisburg et al,[4] which were designed to amplify all but the last thirty bases at the 3' end of the *E. coli* 16S rRNA gene. Primers fD1 and rP2 differ from JAB1 and JAB2 in that the former are 37mers that include 17 and 16 extra bases on the 5' end, respectively, for use as linker sequences for subcloning. The linker sequences were omitted from JAB1 and JAB2 and did not affect the length of the sequenced region of the product.

Figure 5 shows an agarose electrophoresis gel of the PCR-amplified 16S rRNA gene fragments from various Aib-utilizing organisms isolated. The amplified fragments align with the ladder band that corresponds to 1500 bp. PCR products of this length were amplified from every isolate, none of which, upon gel electrophoresis, showed any other amplified products. Cycle sequencing reactions were performed on the PCR product DNA, with JAB1 or JAB2 as sequencing primers.

Organisms were identified by sequencing the PCR-amplified 16S rRNA gene fragments and comparing the sequences obtained using the BLAST algorithm[9] with 16S rRNA sequences in the NCBI Genbank database (Table 2). The 30 isolates cultured from Aib containing media and identified in this study originate from eight genera, including: *Agrobacterium* (one isolate), *Arthrobacter* (nine isolates), *Bacillus* (five isolates),

Corynebacterium (five isolates), *Pantoea* (one isolate), *Pseudomonas* (four isolates), *Rhodococcus* (three isolates), and *Stenotrophomonas* (two isolates).

Many of the sequences were identical to each other, therefore those isolates were considered the exact same organism, this included isolates 4, 11, and Y2; isolates 17, 22, and 33; isolates 3 and 9c; isolates 23 and 31; and isolates 9b and 21. Also, the top match for many isolates was the same, even when the 16S rRNA sequences from those isolates did not exactly match. The 30 sequenced 16S rRNA genes matched closest to 17 separate sequences from NCBI (Table 2).

Growth of organisms for cloning or *dgd* expression studies. The ability of selected organisms to grow on Aib containing media is shown in Table 3. Organisms on hand that contained a sequenced *dgdA* and *dgdR* gene included *B. cepacia*, *Agrobacterium tumefaciens*, and *Mycobacterium smegmatis*. Organisms 8, 17, 26, 31, and F were used in cloning experiments or in Aib metabolism assays. *B. cepacia*, *A. tumefaciens*, and organism 17 grew on liquid and solid Aib media. Organisms 8, 31, and F grew robustly on Aib containing plates, however, they did not grow in equivalent liquid media. Organisms 26 and *M. smegmatis* grew robustly on Aib plates, but produced only heterogeneous clumps in liquid Aib medium or on the surface of unshaken culture. The electrocompetent *E. coli* strain EC100, the selected host for cloning experiments, did not grow on any media containing Aib as a sole nitrogen source.

Acetone production by dgd+ organisms. The metabolic activity of the DgdA enzyme in the presence of Aib and pyruvate results in the production of CO_2, acetone, and alanine. Because acetone easily volatizes, detection of the ketone in the headspace of culture aliquots by gas chromatography with flame ionization detection (GC-FID) was used for confirming DgdA activity in liquid culture. In GC-FID, compounds elute from the GC column, are ionized by a flame, and a detector records the magnitude of the current (pA) produced by the ions. The resulting chromatogram shows each peak with the height in pA and the width in seconds.

In order to relate acetone concentration (mM) in Aib containing media to GC-FID peak areas (pA*s), standards (0.68, 1.4, 2.0, 2.7, 3.4, 6.8, 10.2, and 13.6 mM acetone) were made and a standard curve was generated (Figure 6). The standards were based on the maximum and minimum acetone concentrations detected in Aib-containing cultures of *B. cepacia* by Baurick and represent the expected concentration range of acetone in the mature culture.[10] The chromatograms were generated by injecting onto the GC-FID 50 μL samples of headspace vapor above 200 μL of standard. Figure 7 shows the results of headspace injections from 0.68 mM (Figure 7a) and 13.6 mM (Figure 7b) acetone standards and from a t = 40 h organism 17 Aib culture (Figure 7c). The slope of the standard curve, which had a R^2 of 0.99, indicated a peak area of one pA*s corresponded to a 0.3 mM acetone solution. A similar experiment that measured standard acetone concentration in an Aib culture of organism 17 showed accuracy of 90% (data not shown).

To demonstrate dgd activity in organisms that contain a known or putative *dgdA* gene, cultures of *B. cepacia*, *M. smegmatis*, and isolate 17 were grown in a minimal salts medium that contained Aib as the sole nitrogen source. These were assayed periodically for culture density and acetone. The hypothesis was, if the Aib is being metabolized in culture, then as the cell density of each culture increases, acetone should also increase. The OD_{600} cell density values of 750 μL aliquots of *B. cepacia* and organism 17 cultures measured over a 120 h period are shown in Figure 8a. After ~ 20 hr of incubation with shaking, both *B. cepacia* and 17 entered log phase growth which lasted until t = 60 hr, when stationary growth was observed.

Acetone production by *B. cepacia*, *M. smegmatis*, and organism 17 is shown in Figure 8b. To assay for acetone, 750-μL aliquots of Aib-containing cultures were centrifuged to remove cells and 200 μL of the supernatant was used for headspace analysis on a GC-FID. Acetone concentration increased in organism 17 and *B. cepacia* cultures until t = 60 hr, often which it decreased. Presumably, this is due to acetone being lost to evaporation. The acetone concentration of the Aib containing *M. smegmatis* culture also increased until t = 120 hr, when data was no longer taken.

These results indicate in the Aib containing medium the acetone concentration was at least 16 mM for the three cultures grown in this experiment. The Aib concentration in medium used for these assays was 19 mM, and therefore 84% of the Aib was metabolized to acetone. No acetone production was detected in cultures containing ammonium chloride as the sole nitrogen source.

GC-MS analysis of Aib and d_6-Aib containing cultures. To further identify acetone as the component of culture headspace that resulted in the GC-FID peak eluting at 1.70 min, GC-MS analysis was conducted on headspace vapors. Several samples included an Aib-containing medium (Aib-medium), a late log phase (t = 60 h) organism 17 culture grown in Aib medium (Aib-17), and an Aib medium standard to which acetone (3.4 mM) was added (Aib-Spike). Aliquots of 750 μL were centrifuged and 100 μL of supernatant was transferred to individual GC sample vials, then 50-μL injections of headspace vapors were made from each vial onto the GC-MS.

Figure 9 shows the results of Aib-medium headspace analysis. The chromatogram (Figure 9a) has one peak and is identical to a GC-FID trace of an empty sample vial (data not shown). The mass spectrum of the 1.940 min peak (Figure 9b) contains one major peak at m/z = 44 and is identical to the spectrum generated from CO_2 (Figure 9c).

Figure 10 shows the results of the Aib-17 GC-MS analysis. The chromatogram (Figure 10a) features two peaks with elution times of 2.0 and 2.25 min. The mass spectrum of the first peak was identical to CO_2. The mass spectrum of the second peak (Figure 10b), taken at 2.298 min, is very similar to the mass spectrum of acetone (Figure 10c), which has two peaks; the molecular ion at m/z = 58, and the methyl loss ion at m/z = 43. Analysis of the headspace of Aib-Spike generated a chromatogram and spectrum similar to Aib-17.

Acetone is present in low concentrations in most cell types as a result of acetoacetate decarboxylation,[11] therefore a method was developed to distinguish

endogenous acetone from acetone produced by Aib metabolism. In Aib metabolism by *B. cepacia* DgdA the α-carbon and methyl groups of the Aib amino acid correspond respectively to the carbonyl and methyl groups of the resulting acetone. Therefore, if d_6-Aib were used as the sole nitrogen source in a *B. cepacia* inoculated medium it would result in the production of d_6-acetone which has a mass spectrum clearly distinguishable from that of H_6-acetone. To ensure that the acetone detected in the headspace of Aib containing cultures originated from Aib substrate metabolized by a DgdA and not from another source, d_6-acetone was assayed in several samples, a) a late log phase (t = 60 hr) culture of organism 17 grown in d_6-Aib medium (d_6-Aib-17), b) a d_6-Aib medium made 3.4 mM d_6-acetone (d_6-Aib-Spike), and c) a d_6-Aib containing uninoculated medium (d_6-Aib-medium). The resulting chromatograms (Figure 11a) showed two peaks: one at 2.0 min for CO_2 and the second at 2.28 min for d_6-acetone. The mass spectrum of 2.28 min (Figure 11b) is very similar to the mass spectrum of pure d_6-acetone (Figure 11c). Both spectra show peaks at m/z = 64 and m/z = 46, which correspond to the molecular ion and the CD_3 loss ion of d_6-acetone, respectively. Headspace analysis of d_6-Aib-Spike produced chromatograms and spectra similar to late log phase organism 17 culture in d_6-Aib medium (data not shown). Also, headspace analysis of uninoculated d_6-Aib medium produced chromatograms and spectra similar to uninoculated Aib medium.

DgdA activity in transformed EC100. *E. coli* EC100 cells and the pUC19 plasmid vector were used in cloning and transformation experiments. To show that EC100 cells are able to express cloned *dgd* genes and produce a functional DgdA enzyme, EC100

were transformed with pUC19C7, a derivitive of pUC19 containing a 2.9-kb sequence of *B. cepacia* DNA.[12] Plating the electroporated EC100 cells onto LB/amp/X-gal/IPTG agar provided many white colonies, this indicated the bacteria were transformed and maintaining an insert containing plasmid. One clone designated EC7 was picked and tested for its ability to grow on M9 media with NH_4Cl or Aib as nitrogen sources. The results are shown in Figure 12a and shows that EC7 grows up to moderate cell density on both media. The same liquid cultures were tested for acetone production (Figure 12b). Little or no acetone was detected in the M9/NH_4 media and 5 mM acetone was detected in M9/Aib. The M9/Aib medium contained 20 mM Aib which indicates about a 20% utilization of the available Aib by EC7. This figure is a conservative estimate because: a) some acetone may have evaporated, and b) this was not a direct measure of Aib concentration.

Attempted cloning of dgd genes.

Sau3A I restriction enzyme partial digestion of *dgd*+ genomic DNA. The cloning strategy for this project included partially restricting isolated genomic DNA to ~ 4-6 kb fragments by the four-base cutter Sau3A I (5' overhang GATC). Complete digestion of genomic DNA by a four-base cutter, assuming an equal distribution of A, T, G, and C nucleotides, would generate 250 bp fragments on average. Therefore the reaction time was limited to allow for the generation of longer fragments. The two *dgd* genes in *B. cepacia*, *S. coelicolor*,[13] *M. smegmatis*,[14] *A. tumefaciens*,[15] *Burkholderia fungorum*,[16] *Burkholderia cenocepacia*,[17] and *Leuconostoc mesenteriodes*[18] are contained within an

approximately ~2300-bp region. Therefore, it is reasonable to assume that an uncharacterized *dgd* cassette would fit within a 4-kb restriction fragment.

Organisms were selected for use in cloning experiments based on the best growth rate on Aib plates compared to other isolates. Also, DNA from some isolates was pooled in order to simulate the mixes of DNAs that would be isolated directly from soil. DNA pooled from organisms 8, 26, and 31 (Pool) and DNA from organism F were used in partial Sau3A I restriction enzyme digests. Typical results of digestions are shown in Figure 13a. The digested DNA showed a wide range of sizes, 10 kb to 250 bp. The sections of gel containing fragments in the 4-6 kb range were cut out and eluted from the gel. Figure 13b shows that eluted fragments align with the 4-6 kb bands on the sizing ladder. Eluted restriction fragments of this size were used in subsequent ligation reactions.

Cloning of restricted genomic DNA. In order to generate a plasmid library of genomic DNA in the EC100 *E. coli* cells, restriction digested, size selected Pool and F DNA fragments were ligated to BamH I digested, dephosphorylated pUC19. The results of a typical ligation experiment are shown in Figure 13c. The recombinant plasmids appear at ~ 6-9 kb, which is consistent with addition of the pUC19 vector, 2.6kb, to the ligated Pool or F fragments, ~ 4-6 kb.

EC100 *E. coli* cells were then transformed, via electroporation, with F, or Pool ligation products. The results of typical transformation experiments are shown in Table 4. The transformation frequency of EC100 cells porated in the presence of pUC19 was

1.62×10^9 CFU/μg of vector DNA. Resulting outgrowth cultures Pool and F had titers of 9.9×10^3 CFU/mL and 7.2×10^3 CFU/mL, and white colony frequencies of 88% and 78%, respectfully. White colonies, which arise when using a vector and host compatible with blue/white α-complementation screening,[19] are of interest because they presumably contain plasmid with an exogenous DNA insert. These results are very close to the transformation efficiency of 1×10^{10} CFU/mL of EC100 electrocompetent cells reported by the supplier.

The results from plasmid preparations and restriction analysis of 14 clones from the above transformation are shown in Figure 14a. Lanes 1-7 and 9-15 contain isolated recombinant plasmids digested with BamH I. The average size of the inserts was 3640 bp, with inserts ranging from 2000 to 5400 bp. The results from plasmid preparations from white colonies transformed by pUC19-containing F fragments are shown in Figure 14b. The average size of the 6 Sau3AI generated F fragment inserts was 3950 bp, ranging from 2500 to 7000 bp. Determining the size of the inserts is important in that they should all be 4-6 kb in length in order to accomidate *dgd* genes. These insert ranges are wider than expected from the initial size selection, but still very useful for cloning a 2.3 kb target.

Using the average insert size, the number of clones that needed to be screened was calculated, because if too little of the library was screened there was less probability that a *dgd* clone would be found. The minimum number of clones to screen was calculated by using $N_{min} = \ln(1-P)/\ln(1-((I-X)/(n\bullet G)))$ from Gabor et al.[20] P is the desired probability to find a given gene once, I is the average insert size, X is the size of the *dgd*

genes, n is the number of genomes per the desired gene, and G is the average size of the genomes used. In this study N_{min} was calculated for P = 0.90, 0.95, and 0.99 for both libraries F and Pool, I and X was determined experimentaly, n was 1 in the F library and 1, 2, or 3 in the Pool library, and G was set at 5000 kb, which is an above average genome size for bacteria.

Table 5 summerizes the results. The maximum N_{min} calculated for Pool and F libraries was 4.9×10^4, and 1.3×10^4 colonies. When calculating the N_{min} for Pool, as a precaution, it was assumed that only one of the three isolates whose DNA comprises the Pool library contained *dgd* genes. Both N_{min} values were calculated with a 99% probability of the genes being cloned. The total number of clones screened in the Pool and F libraries were calculated at 1.0×10^5 and 4.4×10^4 colonies, respectively which constitutes two and three times the number of clones calculated above.

Recombinant clones cultured on Aib-containing media. The two libraries of recombinant plasmids were then screened for the dgd+ phenotype. Libraries were screened three ways: a) plating a volume of cells on M9/Aib agar, b) inoculating 100 mL of liquid M9/Aib with a volume of cells originating from F or Pool, and c) replica plating LB amp plates grown up with library cells onto M9/Aib plates. If clones contained *dgd* genes, then it would be expressed in its EC100 host and would result in colony or culture growth from one of the above screening methods.

Six clones were isolated from M9/Aib media. Outgrowth of EC100 transformed with pUC19 inserted with Pool DNA fragments generated one colony containing a

recombinant plasmid. Outgrowth of EC100 transformed with pUC19 inserted with F DNA fragments generated five colonies containing recombinant plasmid. All isolates grew on M9/Aib agar plates, but would not grow in liquid M9/Aib media over seven days at room temperature. Figure 15 shows the insert originating from Pool DNA fragments (plasmid C) was about ~ 6000 bp. Figure 16 shows the inserts originating from F DNA fragments were ~ 4300 and ~ 2200 bp, respectively. Because the 4300 bp inserts showed the same restriction pattern and were collected from the same plate, it was assumed that they were all the same clone. This was referred to as plasmid F. Likewise, the 2200 bp inserts all came from the same plate, showed the same restriction patterns, and were collectively renamed plasmid FC

Sequence analysis of recombinant clones. Plasmids C, F, and FC were sequenced using the EZ::TN <Kan-2>, a transposon mutagenesis kit (Epicentre), which inserts transposons into the recombinant plasmid. Sequencing reactions may then be carried out with transposon-specific primers. The entire plasmid sequence was constructed using the CAP : Assemblage de séquences á INFOBIOGEN online computer program[2] which assembled sequences from the fragments generated from randomly inserted transposons, the known sequence of the vector pUC19, and fragments generated by priming off pUC19 or a sequenced transposon fragment into a gap.

The final map of plasmid C is shown in Figure 17. It is comprised of bidirectional sequence from seven transposons, the pUC19 vector, and a sequence primed from fragment R4. Maps of F and FC were similar (not shown). The F map was

constructed with 12 transposon sequences, pUC19, and a sequence primed from pUC19. The FC map was constructed with eight transposons sequences, pUC19, a sequence primed from pUC19, and a sequence primed from a trasponson fragment.

Plasmids C, F, and FC contained inserts of 5496, 4202, and 2470 bp, and thus are large enough to accommodate a pair of *dgd* genes. An analysis of insert sequences showed each contained multiple open reading frames, some approximately the same length as the 1302 bp *B. cepacia dgdA* gene. BLAST searches were conducted on the insert DNA sequences, the results of which are reported in Table 6. Two types of searches were performed: a) BLASTn, which compares the nucleotide sequence of each insert to all nucleotide sequences contained on the NCBI database, and b) BLASTx, which translates the submitted nucleotide sequence across the six possible reading frames and compares the deduced amino acid sequences to all peptides in the database. NCBI bl2seq BLAST analysis, which scores the similarity of two sequences against each other, was also performed on the three inserts in pairwise combinations. This did not show any sequence similarity between inserts.

All significant hits resulting from searches conducted on inserts were nucleotide or peptide sequences that originated from high G +C gram positive bacteria. Insert C produced four hits, one using BLASTn and three using BLASTx. The single BLASTn hit was an exact match of DNA sequence from *Arthrobacter aurescens*, and the BLASTx hits were matches or similar to known transposases from *Arthrobacter* or closely related genera. Searches conducted on plasmid F resulted in three hits, one using BLASTn and two using BLASTx. The region of the insert which produced the BLASTn hit was

encompassed in the sequence that produced the most significant BLASTx hit, both of which match the nucleotide or protein ppGTPpp synthase sequence. The other insert F BLASTx match was with an ATPase from *Thermobifida fusca*. BLASTn and BLASTx searches on the FC plasmid insert did not result in any hits of prokaryotic nucleotide or protein sequences. Also, none of the insert sequences matched the sequences of known *dgd* genes or proteins.

Multiple sequence alignment analysis of *dgd* gene containing organisms. Until recently, the only *dgdA* and *dgdR* sequences available were those from *B. cepacia*. Due to the increasing number of whole-genome sequencing projects, a number of DgdA and putative DdgA and DgdR peptide sequences have now become available. Alignment of these sequences with the *B. cepacia* dgd proteins could reveal residues that are essential to the enzymatic or regulatory functions, respectively. This was the original intent of this project. The following section describes multiple sequence alignment analysis of the *B. cepacia* DgdA with 10 putative DgdAs, and transaminases from *Bacillus halodurans*, *Pyrococcus abyssi*, and *Psuedomonas syringae* (Figure 18). The latter three were chosen because they are the proteins most similar to *B. cepacia* DgdA, but which are not dialkyglycine decarboxylases. In a subsequent section, alignment between the *B cepacia* DgdR protein and 6 putative DgdR sequences is presented (Figure 19).

Alignment of DdgA proteins. The X-ray crystal structure of DgdA by Toney et al showed that DgdA is a dimer of dimers, each containing two active sites composed of

residues from both monomers.[21] Each dimer also contains two metal ligand binding sites: one with specificity for sodium (Na^+), and another that binds alkali metal unspecifically but requires bound potassium or cesium for the enzyme to function properly (K^+/Cs^+) (Figure 2b).[22] It has been shown that the bound potassium or cesium atom participates in a hydrogen bond network extending from the binding site to residues in the adjacent active site. This provides specific structures necessary for enzyme function.[23]

The percent identity between the *B. cepacia* DgdA and the putative DgdAs in the alignment is summarized in Table 7. Phylogenetically, the organisms most closely related to *B. cepacia* contained putative DgdA sequences that were the most similar to *B. cepacia* DgdA. The major exceptions were some fungal DgdAs were more similar to the *B. cepacia* DgdA than *L. mesenteriodes*.

Analysis of the alignment of potential DgdA sequences focused on residues that were identified by Toney et al as crucial to enzyme activity or metal binding.[21] The strategy was to compare the conservation of given residues among the putative DgdAs, and then to add transaminases to the alignment and recheck the positions. This would allow a better assessment of the importance of a given residue to decarboxylation function. Table 8 summerizes the results. From the DgdA alignment, the residues at the active site and K^+/Cs^+ binding site were more highly conserved than the residues at the Na^+ binding site. Of five residues at the Na^+ site, only one was conserved, Leu102. Comparably, 15 of 19 and 6 of 7 residues were conserved at the active and K^+/Cs^+ binding sites, respectively.

After adding the three transaminases to the alignment, the conservation of residues decreased. Of the 15, six, and one positions of high similarity at the active, K^+/Cs^+, and Na^+ sites, respectively, 10, three, and zero positions were still highly conserved. The residues which were highly conserved in the DgdA alignment and not conserved when transaminases were added were: Gln52, Ser80, Met141, Ala245, Gln394 from the active site, and Leu78, Ser80, Asp307 from the metal ligand binding site.

In order to find other amino acids that may help convey the unique decarboxylase activity of the DgdA, a model of the DgdA active site was constructed (Figure 20) by modifying a PDB crystal structure file of the *B. cepacia* DgdA.[21] Residues close to or in the active site that were conserved in the DgdAs but not the transaminases were highlighted yellow and residues that were conserved by the DgdAs and the transaminases were highlighted red. Also, because the DgdA active site is composed of residues located on different chains, the separate chains are colored white and blue. Figure 20 shows that residues that are conserved by DgdAs and transaminases are concentrated near the cofactor half of the active site. Residues that are conserved by DgdAs and not in transaminases are concentrated near the substrate half of the active site and near or part of the metal ligand binding site.

DgdR. Including *B. cepacia*, the seven *dgdA*-containing prokaryotic organisms also contain a divergently transcribed ORF that is homologous to *dgdR*. The putative DgdRs were predicted by the NCBI ORF finder to be LysR-type transcriptional regulators with

characteristic helix-turn-helix DNA binding motifs at the N-terminal.[3] Multiple sequence alignment of the translated DgdR sequences (Figure 19) revealed that they are not as similar as the DgdA proteins (Table 9).

The most conserved region of the putative DgdR proteins was from amino acid 20 to 91, which is within the helix-turn-helix motif; the proteins are about 50% similar in this region. After that, similarity decreased. Of the remaining 245 positions 51 were similar across the alignment, with the highest concentration of similarity occurring between positions 254 and 270 (71%).

In an attempt to investigate the similarity observed between positions 254 and 270 in Figure 20, an alignment was performed between the seven DgdR sequences and the OxyR, CysB, and CbnR protein sequences.[24-26] The latter three were chosen because they are the only LTTRs with available X-ray crystal structures. From the alignment, the regions homologous to residues 254 to 270 were noted for these LTTRs. Next, models were constructed of OxyR (Figure 21a), CysB (Figure 21b), and CbnR (Figure 21c) which highlighted the residues that were homologous from this region of interest. The similar residues in all models occurred near or within a proposed inducer binding pocket.[24-26]

Intergenic region of *dgd* and putative *dgd* containing organisms. The intergenic region between the divergent *dgdA* and *dgdR* genes presumably contains the promoter sequences for the two genes as well as the sites DgdR binds to for *dgd* transcriptional regulation. This section contains an analysis of the nucleotide sequence conducted on the

B. cepacia dgd intergenic region and a multiple sequence alignment of it with the putative *dgd* intergenic regions. The stratagy was to find conserved DgdR binding sites or other features that would point to similar functions of these genes.

The *B. cepacia dgd* intergenic region. The current mechanism of transcriptional regulation by LysR-type transcriptional regulators (LTTRs) is comprised of the following steps: 1) the LTTR binds to a sequence of approximately 50-60 bp which contains characteristic T-N_{11}-A recognition sequences and overlaps the -35 region of the transcriptional start site of the regulated gene; 2) binding of the LTTR causes a bend in the target DNA; 3) the bend is relaxed upon an inducer binding, which allows RNA polymerase to form a complex with the promoter and initiate transcription.[27, 28]

In *B. cepacia*, the DgdR binding site is situated near the DgdR start codon and extends into the intergenic sequence that separates the *dgdR* gene from the divergently transcribed *dgdA*. In a study conducted by Allen-Daley et al, the organization of the *B. cepacia dgd* intergenic region was deduced (Figure 22).[29] DNase I footprinting showed that the *B. cepacia* DgdR protects a 63-bp region covering the -10 and -35 regions of the *dgdA* promoter. Upon addition of Aib, the footprint was shortened on the down-stream end (*dgdA* coding strand) to 48 bp. Sites of enhanced cleavage within the protected region were observed at positions 23, 35, and 53 on the *dgdA* coding strand, and positions 29 and 30 of the *dgdR* coding strand. This indicates those bases were in a different conformation than the rest of the protected region. Addition of Aib abolished

cleavage at positions 29 and 30 and greatly reduced cleavage at positions 53, 35, and 23 which indicates a change in the DNA-DgdR configuration.

In an attempt to supplement the current organizational model of the *B. cepacia* *dgd* intergenic region, the DNase I protected region was scrutinized for T-N_{11}-A inverted dyads, which are characteristic of LTTR binding sites.[28] Two sites fit the T-N_{11}-A criteria (Figure 22) and constitute four, 4-base sites each separated by 10 bp (about one turn of B-DNA). In addition, another larger inverted repeat, AAGGTA-N_{13}-TACCTT, was found in the protected region and appears upstream one turn from the furthest upstream T-N_{11}-A sequence. This distribution of inverted dyads is similar to that of other LTTRs.[30, 31]

Multiple sequence alignments of *dgd* intergenic sequences from all *dgd* cassette containing organisms. There is low sequence consensus among the DgdR proteins (Table 9). Similarly, the length and sequence (Table 9 and Figure 23) of the intergenic regions separating the putative *dgd* genes (the region with which DgdR would interact) differs even between closely related genera. A multiple sequence alignment (Figure 23) of the nucleotide sequence of all the *dgd* and putative *dgd* cassettes was performed and the regions corresponding to the *B. cepacia* intergenic region were analyzed.

B. cenocepacia and *B. fungorum* are phylogenetically the most similar organisms to *B. cepacia* in this study. Their putative intergenic regions most closely resemble *B. cepacia*. This includes two T-N_{11}-A sequences and an upstream larger inverted dyad, AGNTA-N13-TANCT. Conversely, *A. tumefaciens*, which is also gram-negative, does

not have the same intergenic structure. This short intergenic region, 50 bp, contains three dyads: an ATC-N_9-GAT overlapping 12 bases of a GAT-N_9-ATC followed by a downstream T-N_{11}-A sequence.

The two putative *dgd* intergenic sequences originating from the high G+C, gram positive organisms, *M. smegmatis* and *S. coelicolor,* also did not display similar organization to the *B. cepacia* intergenic region, or to each other. In *M. smegmatis*, the region contains three T-N_{11}-A recognition sequences, the two most upstream overlapping by 11 bp. In *S. coelicolor*, the region contains only one T-N_{11}-A sequence. However, this sequence is flanked, on either side, by a 5-bp (upstream) and a 6-bp (downstream) sequence 22 bp away. Both sequences are repeated within the T-N_{11}-A sequence.

The low G+C, gram positive *L. mesenteroides* intergenic region contains six T-N_{11}-A sequences, which is not surprising considering that this contains the largest intergenic region (124 bp) and with the highest T+A content (78%) of the group. The sequences were clustered into two groups of three and typically had a TTNA-N_7-TNAA structure. Perhaps, with such a long intergenic space, the distance between the *dgdA* promoter and the *dgdR* promoters requires one DgdR to interact with each of two regulatory regions (one for the regulation of each promoter).

Discussion

This section includes: 1) possible explanations for inability to clone *dgd* genes from selected isolates, 2) discussion of multiple sequence alignments of *dgdA, dgdR* and their gene products, 3) a review of the possible ancient lineage of the *dgd* genes, 4) possible further study to be performed, and 5) conclusions.

Review. The experimental results suggest that the 30 organisms isolated from soil are capable of subsisting on media where the only nitrogen source available is from Aib. The identification of these organisms phylogenetically places them in a diverse distribution of genera among all prokaryotes and includes; *Agrobacterium, Arthrobacter, Bacillus, Corynebacterium, Pantoea, Pseudomonas, Rhodococcus,* and *Stenotrophomonas*. Of the isolated Aib users, I was able to confirm DgdA activity in isolate 17, a Corynebacterium, by assaying for acetone, a product of Aib metabolism.

A concerted effort was made to clone *dgd*-utilization genes from several isolates, however these attempts were not successful. Plasmid libraries were constructed from genomic DNA partially restricted to ~4 kb by the Sau3A I restriction enzyme and ligated into a pUC19 vector. These libraries, maintained in an *E. coli* host, contained cloned fragments of a size capable of accommodating a *dgd* gene cassette. Screening of the libraries for recombinant clones capable of subsisting on Aib media produced colonies maintaining 8, 26, or 31 (Pool) DNA which sustained one insert, and from organism F which sustained two cloned inserts. Upon sequencing the inserts, the DNA and translated

protein sequences bore no significant resemblance to the known *dgdA* and *dgdR* gene sequences or DgdA and DgdR protein sequences.

Sequence analysis was conducted on the *B. cepacia dgd* genes and their translated proteins. A multiple sequence alignment of the *B. cepacia* DgdA and multiple putative DgdAs highlighted conserved regions of the protein. Similar analysis of the *B. cepacia* DgdR showed very little amino acid consensus across species at the residue level. The *B. cepacia* intergenic region between *dgdA* and *dgdR* was scrutinized and putative DgdR binding sites were observed near the *dgdA* promoter region. A multiple sequence alignment of the *B. cepacia dgd* intergenic region and putative *dgd* gene containing organisms revealed analogous putative DgdR binding sites. However, only the most closely related sequences seem to be organized similarly.

Analysis of *dgd* gene cloning effort.

The difficulty in isolating *dgd* genes may be due to a fundamental conflict between a property of the organisms and the cloning or culturing strategy. The results of many experiments conducted for this project suggest that if *dgd* genes were present in the organisms and were clonable, the experiments would have succeeded: 1) Genomic DNA isolated was high molecular weight and of good quality. 2) DNA was restricted to a size that could be cloned at high efficiency, yet large enough to accommodate the *dgd* genes. 3) The host was transformed at a high frequency and a high ratio of colonies maintained insert containing plasmid. 4) The average size of cloned inserts was large enough to

carry a *dgd* cassette. Also, the number of clones that were screened from each library indicated that there was a >99% chance the *dgd* genes would have been cloned.

The *E. coli* host and the Aib media used in library screening probably did not effect the outcome of the experiments. In control experiments I showed that the host, when unmodified or maintaining pUC19, formed no colonies on the Aib media (results not shown). Also, the host, when transformed with a plasmid carrying the *B. cepacia dgd* genes, grew robustly on Aib media.

This raises the question, why are the genes not present and/or clonable? There are several possible reasons for the negative *dgd* cloning results:

1. Sau3A I restriction sites are too frequent or too infrequent. Sau3A I is a four base cutter and in a genome composed of 50% G+C will cut the DNA, on average, every 250 bp. For the purposes of this study, Sau3A I was used to partially digest isolated genomic DNA leaving some restriction sites uncleaved and resulting in ~ 4 kb fragments. If there is a higher concentration of Sau3A I sites within a given gene than the rest of the entire genome, the probability that a partial digestion of genomic DNA would leave the gene uncut decreases. Conversely, if there is a very low concentration of Sau3A I sites on either side of the desired gene, and none within the gene, the likelihood that a partial digestion would leave a fragment too large (> 10 kb) to be cloned in pUC19 increases. Moreover, the organisms used in the cloning experiments: *Pantoea*, *Arthrobacter*, and *Rhodococcus,* have genomes with a greater than 50% G+C content: 52%, 65%, and 67%,

respectively. Genomes with higher G+C content would drive down the frequency of Sau3A I sites (5' GATC 3') available for cleavage.

2. There are no *dgd* genes in the cultured organisms. It is possible that these organisms utilize a different pathway for Aib metabolism, and therefore do not contain *dgd* genes. While the gene product of *dgdA* is the only known enzyme that allows metabolism of Aib, it is not necessarily the only one. It is possible that other genes have evolved to perform this function. If the genetic pathway which allows these organisms to utilize Aib as a nitrogen source requires the concerted effort of multiple genes, the strategy of cloning small fragments may result in recombinant plasmid with only fragments of all the genes required. This would render the plasmid unable to convey the *dgd* phenotype to their host.

It is unlikely that these organisms are forsaking the available Aib and fixing nitrogen in the environment provided. If they do contain the necessary *nif* genes, anaerobic conditions are required for nitrogen fixation to occur.[32] Also, most nitrogen fixing bacteria are obligate anaerobes in symbiotic relationships with plant roots. Free living bacteria which fix nitrogen in aerobic conditions do so with specialized structures[33] (*Anabenea*) or a specialized respiratory system[34] (*Azotobacter*) and are exclusive to those or closely related genera. Of the organisms involved in the cloning experiments only organism 26 was identified as being closely related to a nitrogen fixing species, *Pantoea agglomerans*, which has been found in associations with sweet potatoes.[35]

Some of these organisms may not contain any gene or complement of genes which allow them to use Aib as a nitrogen source. The rationale used during the isolation of these organisms was that if an isolate consistently colonized an Aib agar plate, it contained the *dgd* genes. This was demonstrated by the organisms in Table 2 in the Results section. A stricter requirement for identifying Aib users would have been to show the decarboxylation of Aib. In the method used, this could only be shown in isolates that grew consistently in liquid Aib media. Some of the isolated organisms demonstrated growth in liquid Aib media, including organisms 8, 31, and F, only to lose that ability later although they retained their ability to grow on Aib agar. Consequently, decarboxylation of Aib in liquid media was shown in organism 17, but it was not possible to carry out the same experiments on 8, 31, 26 and F.

3. Organisms isolated were in a symbiosis with another organism(s). It is possible, but unlikely, that these organisms are part of a symbiosis with a decarboxylating species. If there was a symbiosis occurring between various organisms colonizing Aib plates, cloning the gene or genes responsible for Aib metabolism would become more complicated. Instead of cloning one Aib metabolizing gene or gene pathway per genome, there would be only one gene or pathway present per multiple genomes. This would increase the number of clones required to screen in order to fully cover the amount of genetic material cloned. For this scenario to occur, the DNA of the decarboxylating species would have to been isolated with the other symbiotic species. However this is

unlikely because 16S rRNA sequencing did not reveal any overlapping bases, as would be expected from a co-amplified, mixed sample.

Another possibility is these organisms are part of a symbiosis with fungi. Genetic material from a fungal symbiote would not have been isolated with a bacterial counterpart when executing the method used with the Qiagen DNeasy kits and therefore would not be cloned. Also, because 16S rRNA PCR and sequencing experiments used primers which were prokaryotic specific, the homologous fungal genes would not be amplified or sequenced. To date, the literature contains no mention of a symbiosis occurring between the organisms used in this study and soil fungi. Also, light microscopy did not reveal a fungal presence in the bacterial cultures.

Why were transformed *E. coli* observed to colonize selective Aib plates when the plasmids maintained did not contain a *dgdA*? The three sequenced inserts, which were isolated from picking transformed colonies from Aib plates, did not show any sequence similarity or contain genes coding for *dgd*-like proteins. Also, unlike the EC7 clone, which contained a *B. cepacia dgd* cassette,[12] none of the transformants could grow on liquid Aib media. Also puzzling was the fact that of the five colonies that resulted from transformation experiments performed with organism F DNA, three were observed on one Aib plate, and two on another. The inserts were shown by sequencing to be identical in each plate group.

It is possible that after the EC100 cells were electroporated, some transformants experienced an enhanced growth rate in outgrowth media. Maybe, when transferred to

Aib plates (see methods), an inhomogenous cluster of cells was smeared on the surface of the agar and formed multiple colonies which utilized nitrogen hoarded from the enriched outgrowth media.

DgdA alignment analysis.

Glutamine 52 and Tyrosine 301. Gln52 has been specifically implicated in decarboxylation function of the *B. cepacia* DgdA.[36, 37] Tony et al hypothesized that DgdA decarboxylation involves hydrogen bonding between the amide side chain of Gln52 and the substrate α-carboxylate.[37] From the multiple sequence alignment of DgdAs and transaminases conducted in this study, Gln52 was one of the few residues conserved across all the putative DgdAs and missing from all the aminotransferases (Figure 18). Woon[36] analyzed the importance of Gln52 on decarboxylation and showed that mutations at that position almost exclusively had a lower decarboxylation rate than the wild type.

In addition to Gln52, Tyr301 has been speculated to be involved in DgdA decarboxylation by hydrogen bonding to Gln52.[36, 37] Presumably, this orients Gln52 to interact effectively with the substrate. Also, it has been postulated that Tyr301 adds to the hydrophobicity of the active site. Tony et al speculated that the relatively hydrophobic active site of the *B. cepacia* DgdA in comparison with other transaminases may help convey its decarboxylase activity.[23] As pointed out by Woon, residues at equivalent positions with Tyr301 DgdA in transaminases are mostly glycines and

phenylalanines.[36] While phenylalanine provides the hydrophobic characteristic similar to the aromatic ring of Tyr301, phenylalanine and glycine could not participate in hydrogen bonding to Gln52.

In the multiple sequence alignment of the *B. cepacia* and putative DgdAs disscused above, there is some diversity at the homologous position of Tyr301: of the 13 DgdAs: seven had tyrosine, three phenylalanine, and two leucine. Phylogenetically, the distribution of this specific residue is random. The four Proteobacteria species (*A. tumefaciens*, *B. cepacia*, *B. cenocepacia*, and *B. fungorum* (the latter two are not shown in the alignment)), the low G+C gram negative bacteria, (*L mesenteroides*); and the fungi, *S. pombe*, *M. graminicola*, and *N. crassa*, contain tyrosine. The high G+C gram positive bacteria, *M. smegmatis* and *S. coelicolor*, and the fungi, *M. grisea,* contain phenylalanine. The fungi, *G. zeae* and *A. nidulans,* contain leucine. Until the decarboxylating ability of these putative DgdA enzymes can be assessed, or a substitution mutation study can be performed onTyr301 from the *B. cepacia* DgdA, it is unknown if Tyr301 contributes to the decarboxylation rate. If the putative DgdAs that lack a Tyr301 decarboxylate dialkylglycines comparably to the *B. cepacia* DgdA, these results would suggest the *p*-OH group of Tyr301 is not essential to decarboxylation.

DgdA exclusive residues. The *B. cepacia* DdgA active site model (Figure 20) highlights the residues near or in the active site that are highly conserved in both putative DgdAs and closely related transaminases (red) or only the DgdAs (yellow). One distinct characteristic of this model is the localized distribution of the two classes of residues (red

and yellow) around the active site. The transaminase-specific residues are concentrated on the cofactor side of the active site, while the decarboxylase-specific residues are positioned on the substrate side.

This polarized distribution possibly illustrates how the DgdA enzyme diverged from an ancestral transaminase. DgdA and related transaminases require bound PLP in the active site to mediate transaminations. In order to meet this requirement the portion of the active site important to PLP binding is conserved in transaminases and all PLP-dependent enzymes. The distinguishing characteristic that differentiates transaminases is the variety of substrates they act upon. Therefore, the substrate side of the active site is not conserved in different transaminases. In the case of DgdA, this divergence includes a specific metal-ion binding station essential to shaping the active site for proper enzyme function.

DgdR alignment analysis.

The DgdR alignment shown in Figure 19 indicates that the most conserved region is within the first 90 amino acids. This is identified in each sequence as a helix-turn-helix motif. Beyond that position the sequence diverges considerably. Comparing the *B. cepacia* DgdR singularly to each of the putative DgdRs (Table 9), the most similar sequences originated from the most closely related organisms, this ranged from nearly identical, *B. cenocepacia* (91%), to very little identity, *L. mesenteroides* (13%). In contrast, the *B. cepacia* DgdA was 95% to 53% identitcal in these organisms (Table 7).

An explanation for this discrepancy could be that decarboxylation by DgdA relies on rigid amino acid organization and sequence, while the DgdR function could be plastic enough to withstand large changes in amino acid content and still regulate effectively. This subject is further explored in a later section.

Intergenic analysis.

A mechanism for *B. cepacia* transcriptional regulation. Based on the mechanism proposed by Allen-Daley et al[29] and the location of the inverted dyads observed in the *B. cepacia dgd* intergenic region (Figure 22), an explanation for DgdR regulation of *dgdA* transcription follows. In the absence of a dialkylglycine, DgdR binds to the upstream half of the AAGGTA-N_{13}-TACCTT site and the most downstream T-N_{11}-A site. This blocks transcription of *dgdA* until a dialkylglycine is present. Bound dialkylglycine induces a conformation change in DgdR which releases, or partially releases, the downstream T-N_{11}-A site and interacts with the upstream T-N_{11}-A site. Upon releasing the downstream T-N_{11}-A site, the -10 promoter element is free to interact with RNA polymerase and initiation of transcription can commence.

An attractive aspect of this model is that it accounts for the enhanced DNase I cleavage observed by Allen Daley et al[29] in DgdR footprinting assays. Most of the sensitive loci were found within the upstream T-N_{11}-A sequence which were abolished or partially abolished upon addition of the inducer. This is corroborated by the proposed regulatory mechanism of the BenM LTTR that controls aromatic compound degradation

in *Acinetobacter sp.* ADP1.[31] The BenM binds to an intergenic region that contains three T-N_{11}-A sites spaced approximately one DNA turn apart from each other. During DNase I footprinting, the bases located within the middle T-N_{11}-A site experience enhanced cleavage in absence of the inducers, benzoate and/or *cis,cis*-muconate, which was abolished or reduced upon inducer addition. The model proposed in that study mirrors this one; the middle site is looped out in absence of inducer leaving it vulnerable to DNase I cleavage and is covered upon addition of the inducer(s).

There are a couple of experiments that could be conducted in order to better understand the *B. cepacia* intergenic region. In one, the *dgdR* transcription start site could be identified. This would show the position where transcription commences and could reveal the -10 and -35 promoter sequences relative to the DgdR DNase I footprint.

Second, putative DgdR binding sites could be used as loci for substitution mutations. This would show specifically which bases are significant to DgdR binding in the presence and absence of inducer and perhaps could explain the significance of the larger AAGGTA-N_{13}-TACCTT dyad. The downstream TACCTT sequence from the dyad is located entirely within the upstream T-N_{11}-A site and partially overlaps the putative -35 box of the *dgdA* promoter. The location of this unusually large dyad in the DNase I footprint is conspicuous and warrants further study.

Multiple sequence alignment of the intergenic region of putative *dgd* containing organisms. Based on the alignment shown in Figure 24, the intergenic regions, which are the DNA sequences between the two *dgd* (or putative *dgd*) genes, are not conserved

to the same extent as the *dgdA* and *dgdR* genes themselves: 2.6% identity versus 34.0% and 10.4%, respectively Lack of consensus in intergenic sequence and T-N_{11}-A site organization indicates the regulatory mechanism of DgdR may differ between species. This is corroborated by two observations:

1) The amino acid sequence of DgdR is not highly conserved. When comparing the gram negative and gram positive putative DgdR amino acid sequences they are more similar to hundreds of other LTTRs than to each other (data not shown). This could be explained by a concerted divergence of the intergenic region and *dgdR* sequence in ancestoral organisms. In this senerio, the most successful organisms would be those experiencing mutations within the intergenic region and *dgdR* that are complementary. These two regions of DNA have accommodated a great deal of change as demonstrated by multiple sequence alignments (Figures 19 and 22) and presumably still can regulate the transcription of *dgdA*.

2) The consensus of defining characteristics of LTTRs (see introduction) describes variations observed in their regulatory mechanisms.[28] Specifically, differences in how various LTTRs interact with proposed binding sites have been observed.[27] Also, there are homologous LTTR genes from different species that are positioned differently relative to the gene(s) under their regulation.[38]

Further work is necessary to fully characterize these intergenic regions, this would include: 1) DNase I footprinting assays, which would show where DgdR binds, 2) primer extension analysis of *dgdA* and *dgdR* mRNA which could find transcription start points, and 3) mutation analysis of putative DgdR binding sites, which could emphasize their

importance to transcriptional regulation. The results of these experiments could enhance the current model.[29]

Putative *dgdA*-containing organisms and an ancient lineage for bacterial genes related to metabolism of Aib.

In a search for homologs of *dgd* genes, the amino acid sequence of the *B. cepacia* DgdA protein was submitted to the National Center for Biotechnology Information (NCBI) in a BLAST search.[3] Twelve peptide sequences from multiple prokaryotic and eukaryotic genera were returned with considerable homology (>50%) to the DgdA amino acid sequence (Table 7). With the exception of *B. cepacia* and *Mycosphaerella graminicola,* whose *dgdA* genes have been cloned and sequenced,[12, 39] all amino acid sequences were products of translated predicted genes from sequencing projects. Therefore, none of these putative DgdAs has been purified, cloned, characterized, or shown to allow their respective source organisms the ability to utilize AIB as a sole nitrogen source.

Of the sequences of DgdA homologs returned from the search, six were from prokaryotic organisms and six were from eukaryotic organisms. An analysis of the sequence upstream from the prokaryotic putative *dgdA* gene start codon revealed open reading frames that initiated divergently 50-124 bp upstream. Translations of these ORFs gave peptide sequences with high similarity to proteins from the LTTR superfamily. In a similar analysis of the DNA upstream from the eukaryotic *dgdA* promoters, translated ORF sequences did not show significant similarity to the LTTRs or to each other. This

was consistent with previous findings in that LysR-type transcriptional regulation and their corresponding genes are exclusive to prokaryotes and archaea.[28]

Considering the organisms identified in this study, and the bacterial species that contain putative *dgdA* and *dgdR* genes, a total of 12 genera are represented in the prokaryotic domain that potentially utilize Aib as a nitrogen source. These genera occur in the Proteobacteria (three from this study and two sequence based), the high G+C gram positive bacteria (three from this study and two sequence based), and the low G+C gram positive bacteria (one from this study and one sequence based).

A phylogram of 49 bacterial species (Figure 24) representing all major prokaryotic groups was constructed based on 16S rRNA sequences, nine of which were sequenced from Aib-utilizing bacteria isolated in this study. The 40 other 16S gene sequences were taken from the NCBI site, 28 representing major prokaryotice groups, eight representing the genera of organisms isolated in this study, and four representing genera with *dgd* genes discovered by genome sequencing. This phylogeny is rooted with *Aquifex*, which is thought to be one of the earliest bacteria to diverge from the common ancestor to all prokaryotes. The branch point indicated in Figure 24 shows the point during evolution when an ancient organism, which evolved into all the progenitor groups, obtained the *dgd* genes. These groups include all gram positive and gram negative bacteria, which makes up the majority of known bacterial species, and excludes only the groups which branched off very early in prokaryotic evolution.

Lateral transfer of *dgd* genes. Another explanation for the distribution of *dgd* genes throughout the prokaryotic domain is lateral gene transfer. Unlike vertical transfer, where progeny inherit the genes of their ancestors, laterally transferred genes pass between species and are incorporated into the recipient's genome. This type of gene transfer has been identified as a major mechanism for evolutionary change in prokaryotes and archeae and is especially implicated in the distribution of genes responsible for specialized metabolism and antibiotic resistance.[40] The genetic content of some bacterial species is estimated to contain 17% laterally transferred sequences. Considering that *dgdA* codes for the specialized metabolism of Aib, a component of some fungal antibiotics, it is reasonable to question the proposed pedigree of the *dgd* genes and consider their wide distribution may be via lateral transfer.

One method for determining whether a given gene was laterally transferred or conventionally passed on is to compare G+C content of the gene in question to that of the entire genome.[40] The G+C content of known and putative prokaryotic *dgd* genes is shown in Table 10. These genes are identical, or nearly identical, in G+C content compared to the G+C content of each whole genome, which range from 37% to 72% G+C. Interestingly, the intergenic region of each organism has lower G+C content than either *dgd* gene or their entire genome. This is probably due to the presence of A+T rich promoter, upstream element, and T-N_{11}-A sequences required for the proper transcriptional regulation and recruitment of RNA polymerase. These results indicate that horizontal gene transfer probably did not contribute to the distribution of the *dgd* genes in these organisms, at least recently.

Future *dgdA* and *dgdR* related work. Besides the experiments suggested above, there are other experiments that could be conducted relating to *dgd* and putative *dgd* genes. Below is the rational for performing *dgd* related work and suggestions for further experiments.

DgdR crystal structure. The purpose of this study, to clone *dgd* genes from soil organisms, was inspired by the appearance of putative *dgd* genes resulting from the whole genome sequencing of *A. tumefaciens* and *S. coelicolor*. By accruing as many *dgd* sequences as possible, inferences could be made concerning the structure and function of the *dgd* coded proteins. This was especially attractive in regard to DgdR, which at this time has no crystal structure.

At the time this research was being conducted there was also no crystal structure available for any LTTR. Currently, there are three, the truncated inducer binding domains of CysB and OxyR, and the entire structure of the CbnR biomolecule.[24-26] Further investigation into the sequences of DgdR proteins could still be very useful toward the understanding of DgdR and other LTTR structures. The DgdR proteins from the organisms that have sequence available show very little identity, as does the intergenic regions to which DgdR binds. Accumulating as many different DgdR sequences as possible would better illustrate the diversity this regulator displays across bacteria and would highlight residues essential to its function. This in turn could help illustrate the plasticity of DgdR function on structurally different intergenic sequences.

Exogenous regulation of foreign genes by a *dgdA* transcriptional regulator. Currently, the *B. cepacia* DgdR protein is being investigated for its possible use as a transcriptional regulator of foreign genes. This system is especially attractive in that it would be activated by the simple, stable, Aib molecule. This is an observation that led to the patent of *dgdR*, DgdR and the *dgd* operon.[41, 42] It is possible that the eukaryotic transcriptional regulator, or a putative DgdR from another species, could perform transcriptional regulation of foreign genes more efficiently. The *B. cepacia* DgdR has been shown to only negatively regulate the transcription of genes under its control.[29] Considering the diversity demonstrated across the DgdRs and intergenic regions, and the observation that many other LTTRs are positive regulators, another DgdR protein may be a better candidate to control the expression of exogenous genes.

Strategy for finding more *dgd* containing organisms. In the course of completing this project many other organisms have been revealed to contain a putative *dgdA* and/or *dgdR* (Tables 7 and 9) which have increased our understanding of how widespread the genes are distributed amongst prokaryotes. Considering that the *dgdA* gene is found in both eukaryotes and prokaryotes, and assuming that lateral gene transfer has not occurred, the ancient organism that gave rise to all domains of life must have carried a *dgdA* gene. Thus far *dgdA* has only been shown in soil bacteria and fungi (Table 7). If this gene was present in the predecessor to all current life on earth, then it should be found in organisms occupying other niches besides soil. This study focused on soil, perhaps future searches

for dgd+ organisms could be extended to marine and freshwater communities or thermal vents and hot springs.

Also, uncultureable organisms could be screened for *dgd* genes. It has been estimated that less than 1% of all soil microorganisms are cultureable,[43, 44] which leaves an incredible abundance of genetic material left unexplored. Considering all the *dgd* containing organisms currently uncovered by genome sequencing and this study, it stands to reason that even more are not reachable using common culturing techniques. Many methods[45] and industrial kits exist for the isolation of high quality DNA from soil which could in turn be used in the construction of plasmid or cosmid libraries that could be screened for *dgd* genes.

Conclusions. The primary goal of this research was to culture previously unknown Aib utilizing organisms and to clone the genes that imparted that ability. The sequencing of the Aib utilization genes was to provide multiple *dgdA* and *dgdR* gene sequences that could be compared to the *B. cepacia dgd* genes. Although cloning the Aib metabolizing genes was not successful, a number of dgd-type genes were revealed from various genome sequencing projects. These provided putative *dgdA* and *dgdR* sequences that were analyzed here.

It is unknown what role the *dgd* genes play in soil ecology at this time. However, the occurrence of Aib in microbial soil communities as a part of antibiotic peptides synthesized by fungi and the discovery of a wide range of organisms that contain *dgd* genes suggests a function for *dgd* in bacteria-fungi interaction. There is much relevant

work that can be explored related to *dgd* genes, and continued research could provide novel and/or useful knowledge regarding; prokaryotic evolution, genetic engineering, and the understanding of DgdR and LTTR transcriptional regulation.

Works Cited

1. http://www.firstmarket.com/cutter/cut2.html.

2. http://www.infobiogen.fr/services/analyseq/cgi-bin/cap_in.pl.

3. http://www.ncbi.nlm.nih.gov/.

4. Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991) *J. Bacteriol. 173*, 697-703.

5. Kurien, B. T., Kaufman, K. M., Harley, J. B., and Scofield, R. H. (2001) *Anal. Biochem. 296*, 162-6.

6. Vogeli, G., Horn, E., Laurent, M., and Nath, P. (1985) *Anal. Biochem. 151*, 442-4.

7. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, Vol. 1, 3rd ed., Cold Spring Harbor Laboratory Press, New York.

8. Frischauf, A. M. (1987) *Methods Enzymol. 152*, 183-9.

9. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J Mol. Biol. 215*, 403-10.

10. Baurick, K. B. Cloning and Expression of Pseudomonas Cepacia 2,2-Dialkylglycine Decarboxylase in Escherichia Coli. M.S., University of Alaska, Fairbanks, AK, 1987.

11. Petersen, D. J., and Bennett, G. N. (1990) *Appl. Environ. Microb. 56*, 3491-8.

12. Keller, J. W., Baurick, K. B., Rutt, G. C., O'Malley, M. V., Sonafrank, N. L., Reynolds, R. A., Ebbesson, L. O., and Vajdos, F. F. (1990) *J Biol. Chem. 265*, 5531-9.

13. Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A. (2002) *Nature 417*, 141-7.

14. http://www.tigr.org/.

15. Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B. S., Cao, Y., Askenazi, M., Halling, C., Mullin, L., Houmiel, K., Gordon, J., Vaudin, M., Iartchouk, O., Epp, A., Liu, F., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Flanagan, C., Crowell, C., Gurson, J., Lomo, C., Sear, C., Strub, G., Cielo, C., and Slater, S. (2001) *Science 294*, 2323-8.

16. http://genome.jgi-psf.org/draft_microbes/burfu/burfu.home.html.

17. http://www.sanger.ac.uk/Projects/B_cenocepacia/.

18. http://genome.jgi-psf.org/draft_microbes/leume/leume.home.html.

19. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, Vol. 1, 3rd ed., Cold Spring Harbor Laboratory Press, New York.

20. Gabor, E. M., de Vries, E. J., and Janssen, D. B. (2004) *Environ. Microbiol. In Press.*

21. Toney, M. D., Hohenester, E., Cowan, S. W., and Jansonius, J. N. (1993) *Science 261*, 756-9.

22. Hohenester, E., Keller, J. W., and Jansonius, J. N. (1994) *Biochemistry 33*, 13561-70.

23. Toney, M. D., Hohenester, E., Keller, J. W., and Jansonius, J. N. (1995) *J. Mol. Biol. 245*, 151-79.

24. Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K., and Senda, T. (2003) *J. Mol. Biol. 328*, 555-66.

25. Tyrrell, R., Verschueren, K. H., Dodson, E. J., Murshudov, G. N., Addy, C., and Wilkinson, A. J. (1997) *Structure 5*, 1017-32.

26. Choi, H., Kim, S., Mukhopadhyay, P., Cho, S., Woo, J., Storz, G., and Ryu, S. (2001) *Cell 105*, 103-13.

27. McFall, S. M., Chugani, S. A., and Chakrabarty, A. M. (1998) *Gene 223*, 257-67.

28. Schell, M. A. (1993) *Annu. Rev. Microbiol. 47*, 597-626.

29. Allen-Daley, E., Sun, H., Bray-Hallan, S. T., and Keller, J. W. (2002). Manuscript in preparation.

30. Goethals, K., Van Montagu, M., and Holsters, M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 1646-50.

31. Bundy, B. M., Collier, L. S., Hoover, T. R., and Neidle, E. L. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 7693-8.

32. http://soils1.cses.vt.edu/ch/biol_4684/Cycles/nfix.html.

33. Golden, J. W., and Yoon, H. S. (2003) *Curr. Opin. Microbiol. 6*, 557-63.

34. Ureta, A., and Nordlund, S. (2002) *J. Bacteriol. 184*, 5805-9.

35. Asis, C. A., Jr., and Adachi, K. (2004) *Lett. Appl. Microbiol. 38*, 19-23.

36. Woon, S.-T. Wild-Type and Mutant 2,2-Dialkyglycine Decarboxylases: The Catalytic Role of Active Site Glutamine 52 Investigated by Site-Directed Mutagenesis and Computer Analysis. Ph.D., University of Alaska, Fairbanks, AK, 1998.

37. Toney, M. D., Pascarella, S., and De Biase, D. (1995) *Protein Sci. 4*, 2366-74.

38. Ramachandran, S., Magnuson, T. S., and Crawford, D. L. (2000) *DNA Seq. 11*, 51-60.

39. Adachi, K., Nelson, G. H., Peoples, K. A., DeZwaan, T. M., Skalchunes, A. R., Heiniger, R. W., Shuster, J. R., Hamer, L., and Tanzer, M. M. (2003) *Curr. Genet. 43*, 358-63.

40. Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000) *Nature 405*, 299-304.

41. Keller, J. W. (1993). *Repressor Protein Gene for Regulating Expression of Polypeptides and its use in the Preparation of 2,2-Dialkylglycine Decarboxylase of Pseudomonas cepacia.* Patent No. 5210025.

42. Keller, J. W. (1994). *Repressor Protein and Operon for Regulationg Expression of Polypeptides and its use in the Preparation of 2,2-Dialkyglycine Decarboxylase of Pseudomonas cepacia.* Patent No. 5356796.

43. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998) *Chem. Biol. 5*, R245-9.

44. Rondon, M. R., Goodman, R. M., and Handelsman, J. (1999) *Trends Biotechnol. 17*, 403-9.

45. Zhou, J., Bruns, M. A., and Tiedje, J. M. (1996) *Appl. Environ. Microb. 62*, 316-22.

Table 1. Soil sampling sites and numbers of isolates.

Sample Location	# of distinct organisms cultured
Harvard Circle (HC)[a]	17
Non-Permafrost Soil (NP)[a]	12
Under Tree (UT)[b]	12
Chena River (CR)[a]	9
Garbage Dump (GD)[a]	9
Greenhouse (GH)[b]	4
Villanova Drive (VD)[b]	2
Garden (G)[b]	1
Permafrost Soil (PS)[a]	1
Tanana River (TR)[a]	1
Tanana Slough (TS)[a]	1

[a]Isolated by Tammy Thompson.
[b]Isolated by Jeff Bickmeier.

Figure 5.

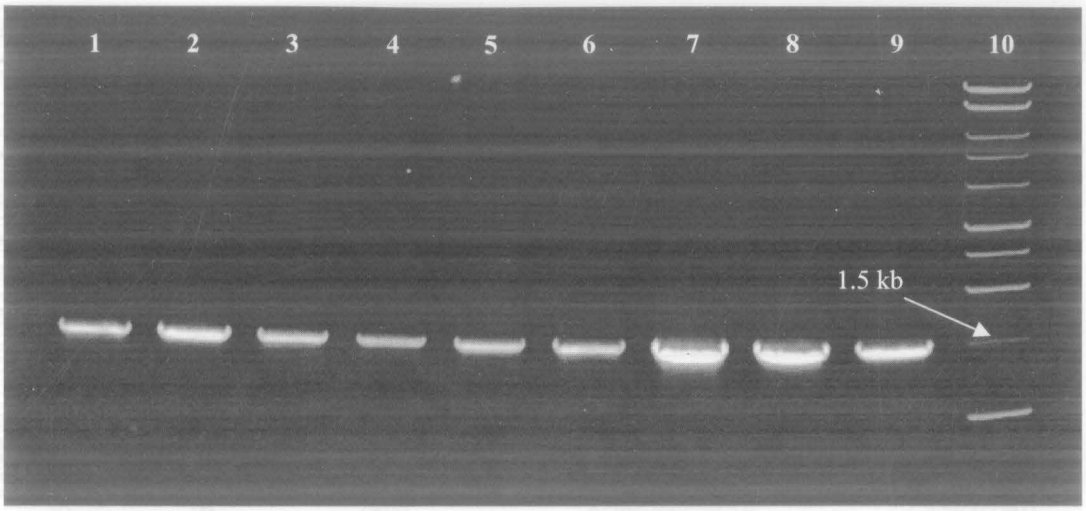


Figure 5. PCR amplification of prokaryotic 16S rRNA genes. Lanes 1-2: Isolate 16, Lanes 3-4: Isolate 19, Lanes 5-6: Isolate 23, Lane 7: A mix of isolate 23 and 36, Lanes 8-9: Isolate 36. Lane 10: 1kb ladder. See methods section for electrophoresis details.

Table 2. Identification of sequenced 16S rRNA genes of Aib-utilizing isolates.

Isolate	Source[e]	Sequence Length	Top Match[f]	Sequence Similarity
N[c]	VD	852[b]	*Agrobacterium tumefaciens*	852/852
2[d]	HC	1379	*Arthrobacter aurescens*	1363/1379
F[c]	VD	806[b]	*Arthrobacter histidinolovorans*	805/806
16[c]	UT	1409	*Arthrobacter nicotianae*	1394/1409
8[c]	NP	1431	*Arthrobacter oxydans*	1429/1432
4[d]	GD	1364	*Arthrobacter protophormiae*	1364/1364
11[d]	GD	636[a]	*Arthrobacter protophormiae*	636/636
Y2[c]	VD	1393	*Arthrobacter protophormiae*	1393/1393
19[c]	CR	1392	*Arthrobacter sp.*	1356/1367
36[c]	HC	1406	*Arthrobacter sp.*	1393/1401
37[c]	NP	1409	*Bacillus aquamarinus*	1397/1410
13[d]	UT	752[a]	*Bacillus pumilus*	672/755
10[d]	CR	764[b]	*Bacillus sp.*	763/764
28[c]	HC	1458	*Bacillus sp.*	1434/1459
30[c]	HC	801[a]	*Bacillus sp.*	800/801
34[d]	GH	802[a]	*Corynebacterium variabilis*	800/802
27[c]	GD	1369	*Corynebacterium variabilis*	1366/1370
33[d]	GD	779[b]	*Corynebacterium variabilis*	779/779
17[d]	GD	556[b]	*Corynebacterium variabilis*	556/556
22[d]	GD	803[b]	*Corynebacterium variabilis*	803/803
26[c]	HC	805[a]	*Pantoea agglomerans*	778/805
3[d]	CR	1418	*Pseudomonas libaniensis*	1417/1418
9c[d]	HC	747[a]	*Pseudomonas libaniensis*	699/700
5[d]	GH	249[b]	*Pseudomonas sp.*	236/249
G5[c]	G	630[a]	*Pseudomonas tolaasii*	608/632
23[c]	NP	880[a]	*Rhodococcus sp.*	874/880
25[c]	HC	744[b]	*Rhodococcus sp.*	735/738
31[c]	HC	759[a]	*Rhodococcus sp.*	753/759
9b[d]	HC	708[a]	*Stenotrophomonas maltophilia*	747/747
21[d]	GD	657[b]	*Stenotrophomonas maltophilia*	633/657

[a]Sequence derived from forward primed reaction.
[b]Sequence derived from reverse primed reaction.
[c]Sequenced by Jeff Bickmeier.
[d]Sequenced by Tammy Thompson.
[e]See Table 1 for definitions.
[f]Sequences with the same top match are highlighted yellow.

Table 3. Growth of selected microorganisms[a] on Aib-containing media.

Organism	Growth on Aib Containing Plates	Growth in Aib Liquid Media	Growth on Aib Liquid Media Surface
B. cepacia	+	+	-
A. tumefaciens	+	+	-
M. smegmatis	+	-	+
E. coli EC100[b]	-	-	-
8	+	-	-
17	+	+	-
26	+	-	+
31	+	-	-
F	+	-	-

[a]Of the 30 organisms identified using 16S rRNA sequencing, isolates 8, 26, 31, and F were used in cloning experiments and isolate 17 was used in Aib metabolism assays.
[b]Cloning host.

Figure 6.

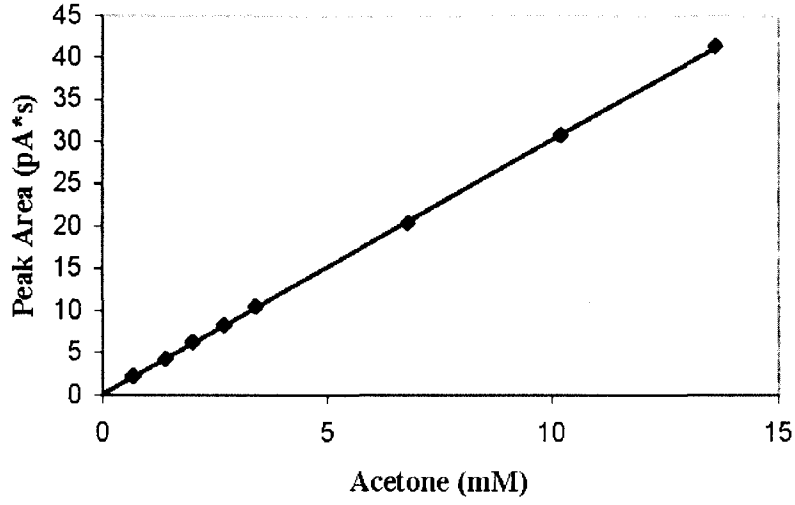


Figure 6. Standard curve of acetone concentrations. The equation of the line is $y = 3.02x + 0.06$ and has a R^2 value of 0.99.

Figure 7.

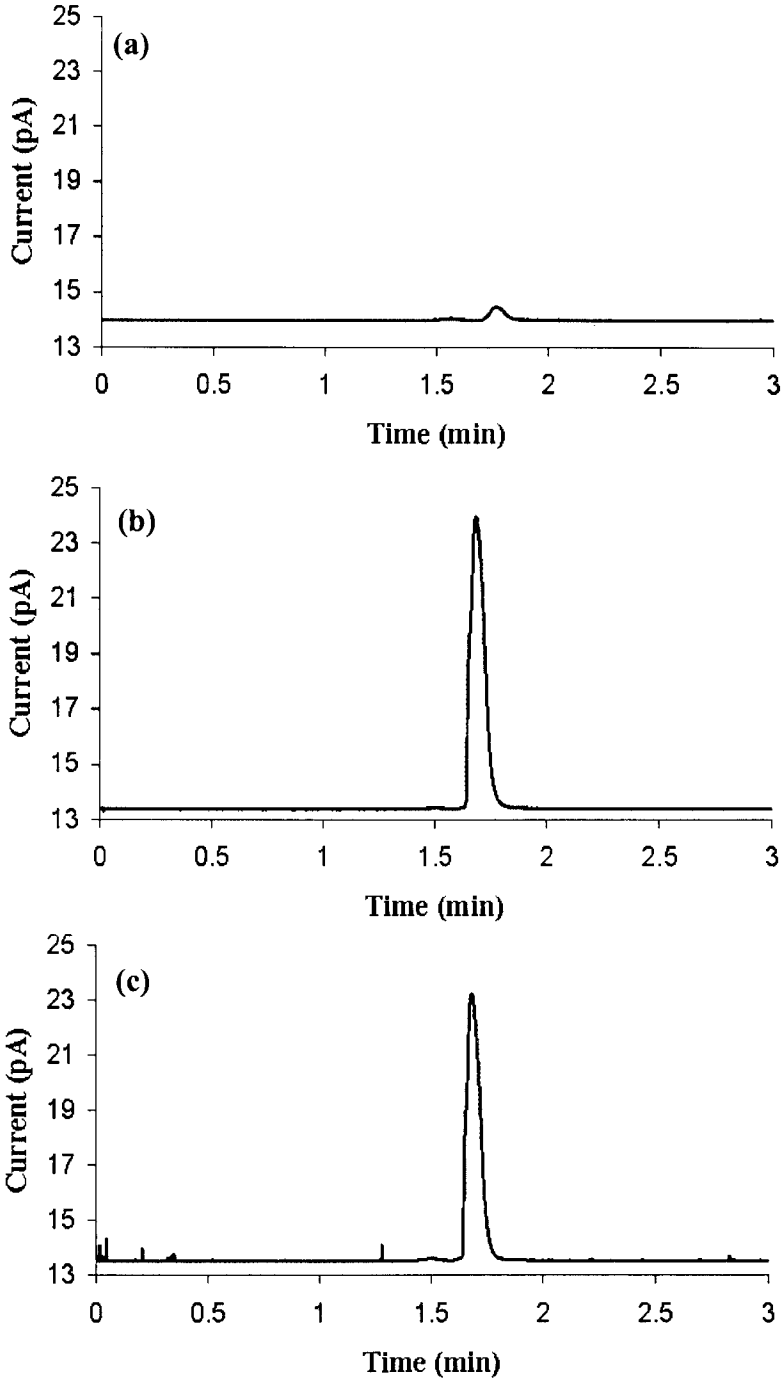


Figure 7. Typical results of GC-FID headspace analysis. (a) Chromatogram of headspace from Aib medium adjusted to 0.68 mM acetone. (b) Chromatogram of headspace from Aib medium adjusted to 13.6 mM acetone. (c) Chromatogram of organism 17 Aib culture headspace at t = 40 h.

Figure 8.

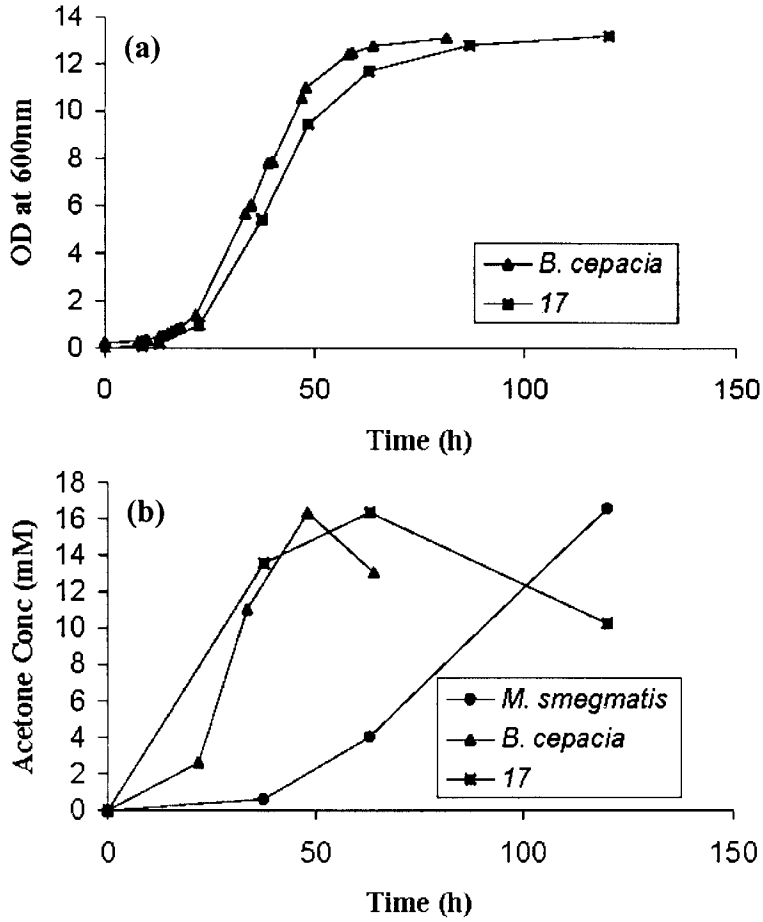


Figure 8. Cell growth and acetone concentration of cultures. Cell growth of *B. cepacia* and organism 17 cultures are shown in (a). Acetone concentration of *B. cepacia*, organism 17, and *M. smegmatis* cultures are shown in (b).

Figure 9.

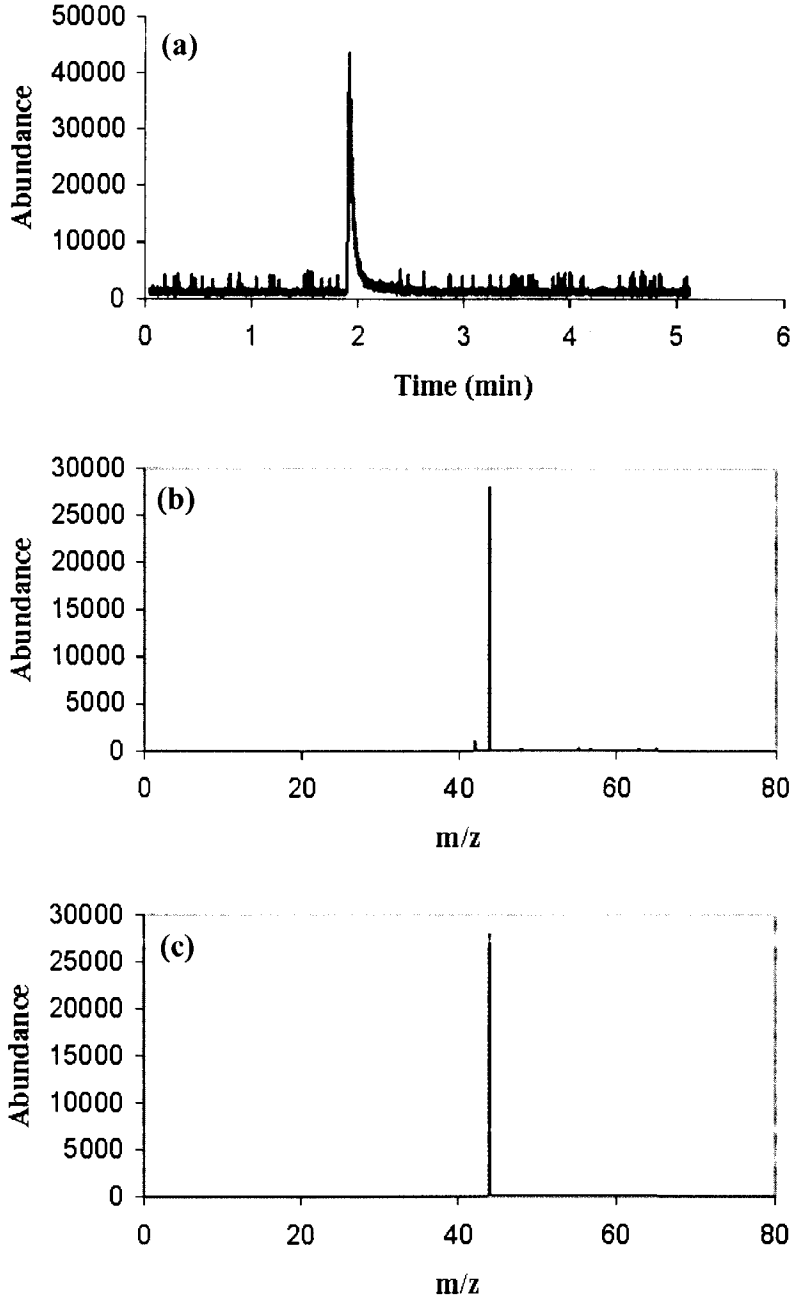


Figure 9. GC-MS analysis of Aib medium. (a) Gas chromatogram of uninoculated Aib media headspace. (b) Mass spectrum of uninoculated Aib media headspace taken at 1.940 min. (c) Mass spectrum of CO_2.

Figure 10.

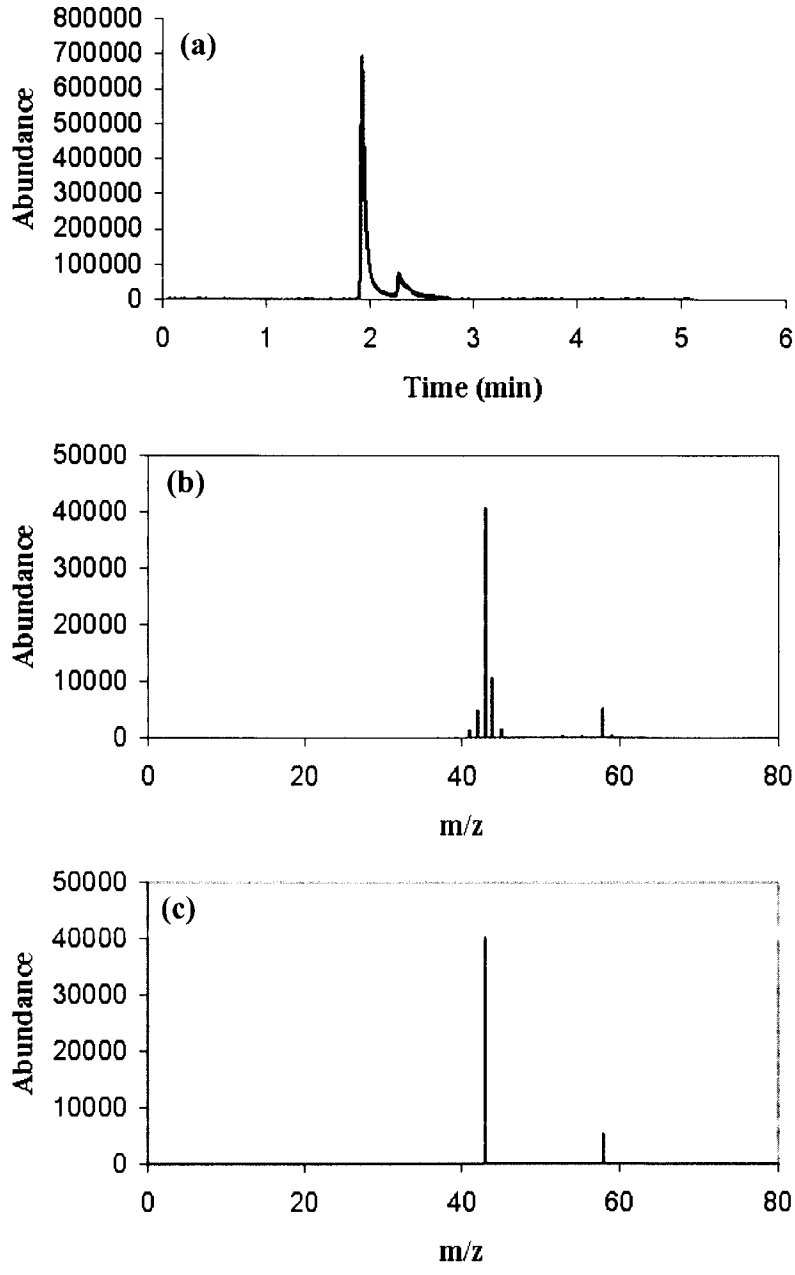


Figure 10. GC-MS headspace analysis of Aib bacterial culture. (a) Gas chromatogram of organism 17 Aib culture headspace. (b) Mass spectrum of organism 17 Aib culture headspace taken at 2.298 min. (c) Mass spectrum of acetone.

Figure 11.

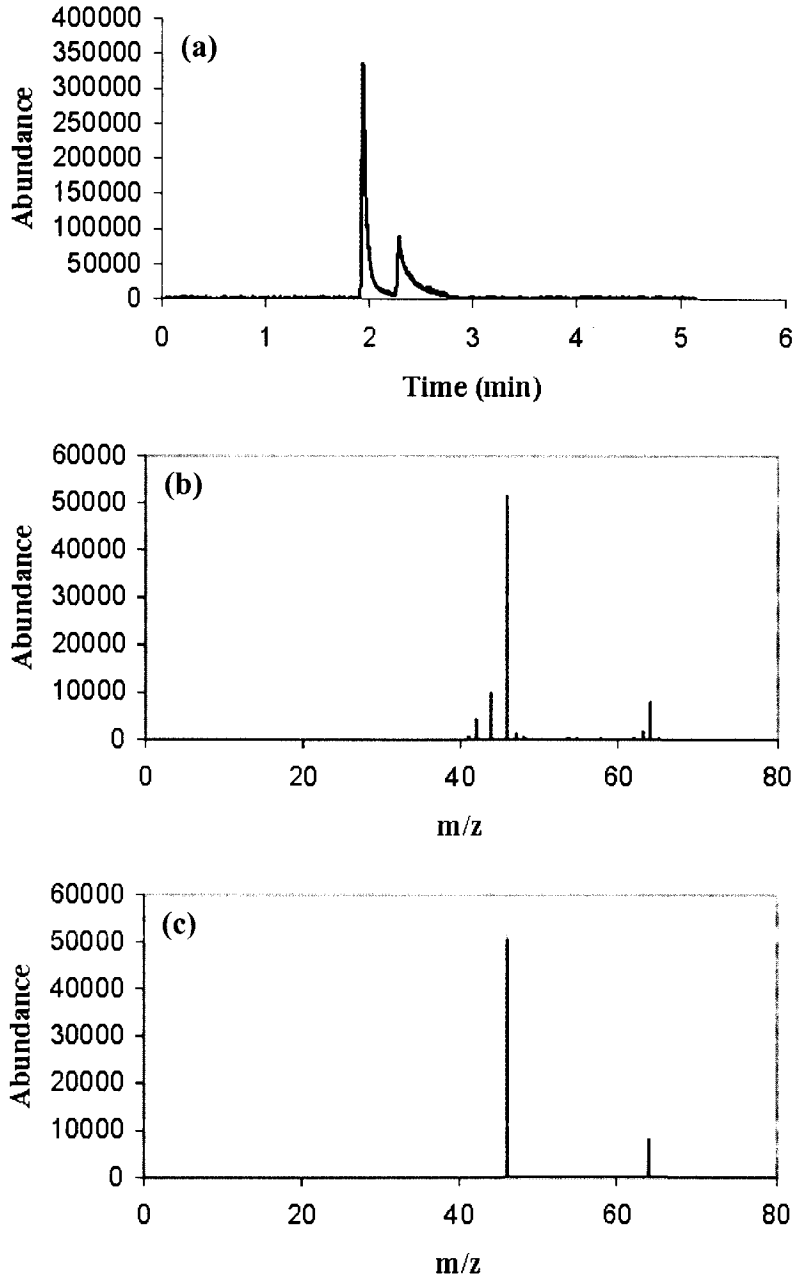


Figure 11. GC-MS headspace analysis of d_6-Aib bacterial culture. (a) Gas chromatogram of organism 17 d_6-Aib culture headspace. (b) Mass spectrum of organism 17 d_6-Aib culture headspace taken at 2.284 min. (c) Mass spectrum of d_6-acetone.

Figure 12.

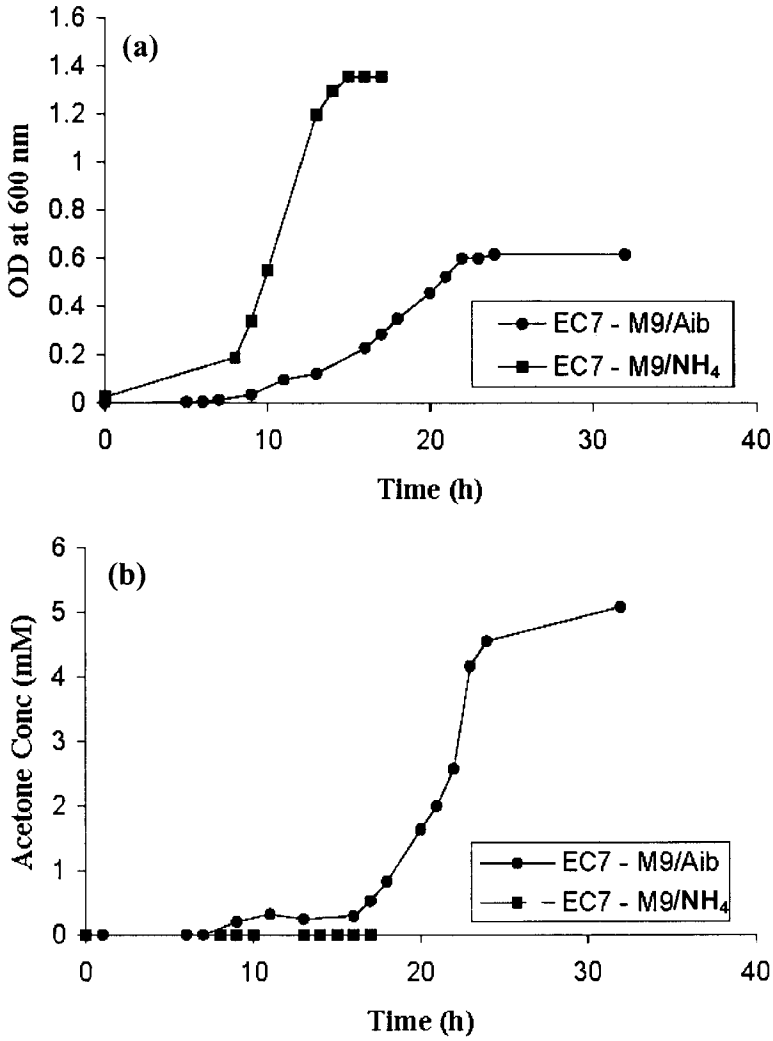


Figure 12. Cell growth (a) and acetone concentration (b) of EC7 bacterial culture. EC7 was grown in minimal salts NH_4Cl or minimal salts Aib media.

Figure 13.

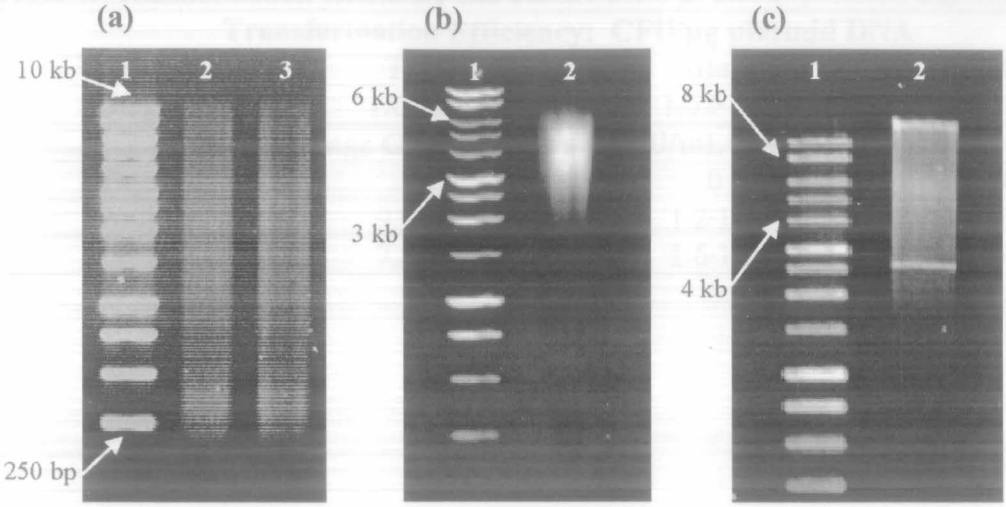


Figure 13. Agarose electrophoresis of partial Sau3A I-restricted genomic DNA. (a) Lane 1: 1kb ladder, Lanes 2-3: Pool genomic DNA partially digested by restriction enzyme Sau3AI. (b) Lane1: 1kb ladder, Lane 2: Pool DNA fragments excised from agarose gel and used in ligation reactions. (c) Lane 1: 1kb ladder, Lane 2: An aliquot taken from a typical ligation reaction involving Pool DNA fragments and pUC19. Arrows indicate size of a given DNA band.

Table 4. Transformation efficiency and culture titers of electroporation experiments.

Transformation Efficiency: CFU/μg plasmid DNA			
Vector	**Total**	**Blue**	**White**
pUC19	$1.62 \cdot 10^9$	$1.62 \cdot 10^9$	0
Average Culture Titers: CFU/mL outgrowth			
Water	0	0	0
Pool	$9.9 \cdot 10^3$	$1.2 \cdot 10^3$	$8.7 \cdot 10^3$
F	$7.2 \cdot 10^3$	$1.6 \cdot 10^3$	$5.6 \cdot 10^3$

Figure 14.

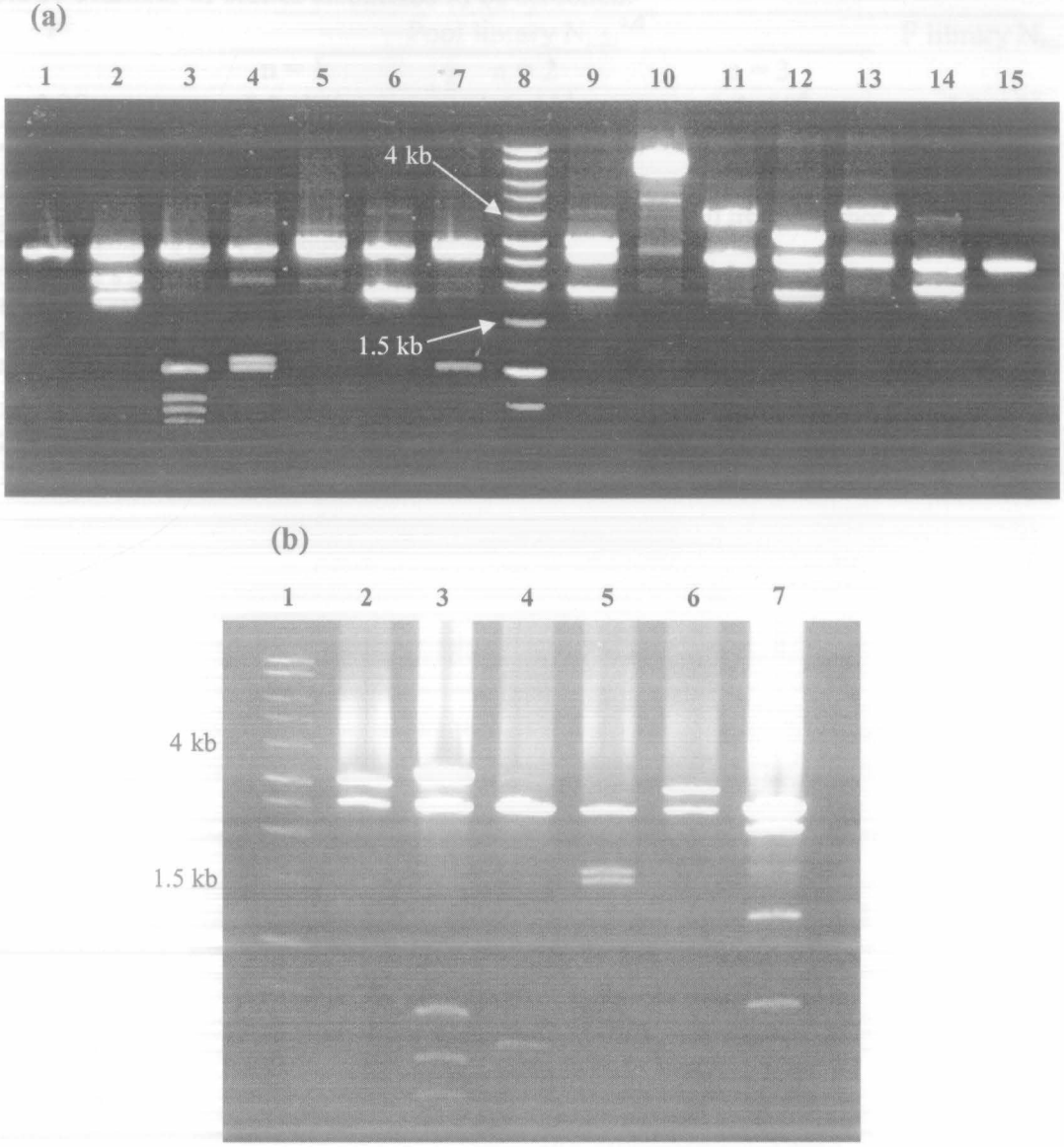


Figure 14. Sau3A I-digestion of recombinent plasmids. (a) Inserts originated from Pool DNA fragments, the average insert size was 3640 bp. (b) Inserts originated from F fragments and whose average insert size was 3950 bp.

Table 5. Number of clones calculated to be screened.[a]

P[b]	Pool library N_{min}[c,d]			F library N_{min}[e]
	n = 1	n = 2	n = 3	
0.90	8.2×10^3	1.6×10^4	2.5×10^4	6.8×10^3
0.95	1.1×10^4	2.1×10^4	3.2×10^4	8.8×10^3
0.99	1.6×10^4	3.3×10^4	4.9×10^4	1.3×10^4

[a] N_{min} was calculated using the equation $N_{min} = \ln(1-P)/\ln(1-(I-X)/(n\bullet G)))$ from reference 20.
[b] P = the probability a *dgd* was cloned.
[c] Three isolate's DNAs compose the Pool fragments, therefore the N_{min} was calculated in the circumstance there was one set of *dgd* genes per 1, 2, or 3 genomes.
[d] The total number of white colonies screened from Pool libraries was 1.0×10^5.
[e] The total number of white colonies screened from F libraries was 4.4×10^4.

Figure 15.

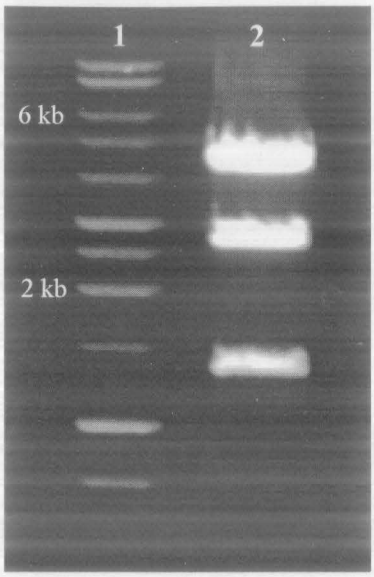


Figure 15. Isolation of plasmid C. Lane 1: 1kb ladder. Lane 2: Recombinant plasmid DNA isolated from a colony growing on M9/Aib containing media and digested with Hind III. Approximant size of insert DNA is 6 kb. See methods section for electrophoresis details.

Figure 16.

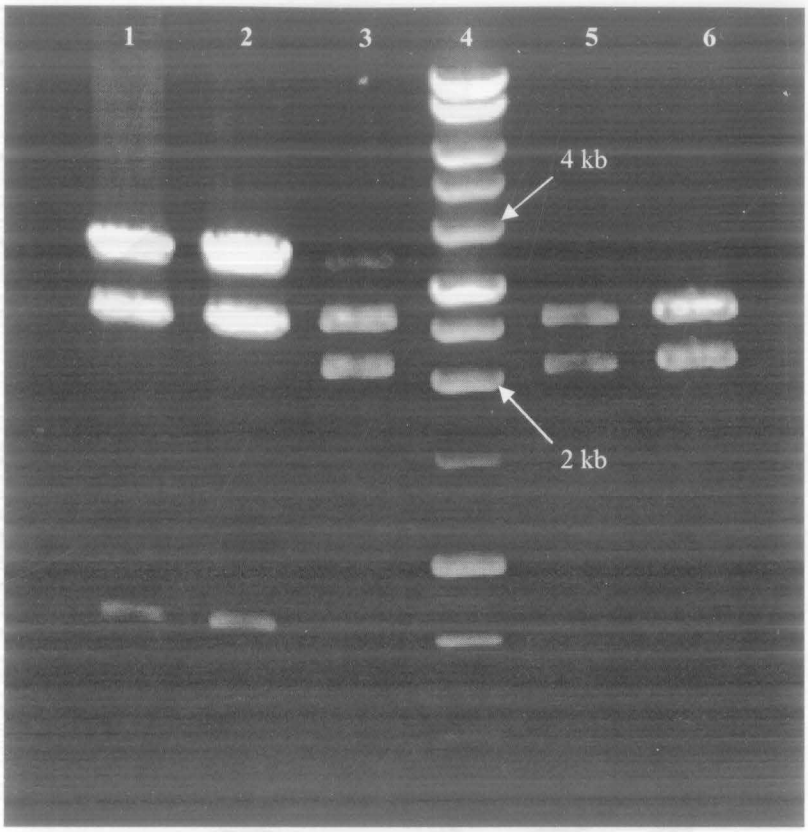


Figure 16. Isolation of plasmids F and FC. Lanes 1-3,5-6: Plasmid isolated from colonies grown on M9/Aib/leu plates and digested with Hind III. Lane 4: 1 kb ladder. Inserts from lanes 1&2 are approximately 4.3 kb and whose colonies of origin were picked from the same plate. Inserts from lanes 3,5&6 are approximately 2.2 kb and whose colonies of origin were picked from the same plate. See methods section for electrophoresis details.

Figure 17.

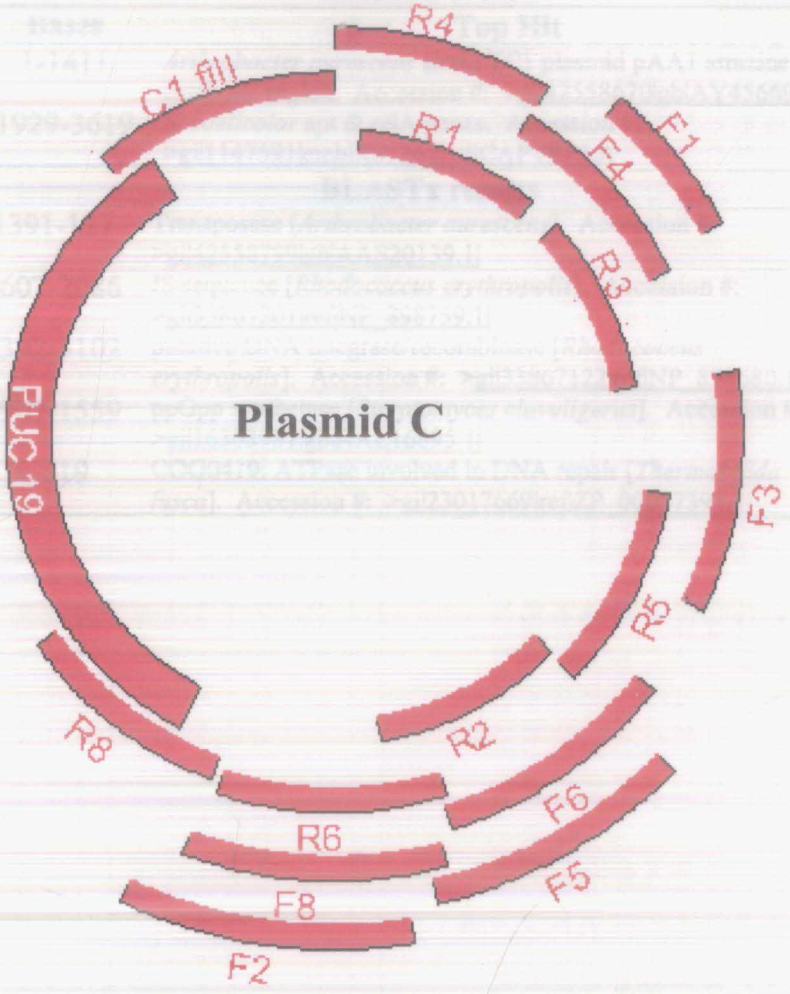


Figure 17. The plasmid C map. The 8179-bp map was constructed from sequences gathered from seven randomly inserted trasposons labeled above 1-6 and 8. R and F denote whether sequencing from a transposon was a forward or reverse primed reaction. C1 fill was a sequence generated from designing a primer to sequence beyond R4 and into pUC19. pUC19 is the thickest red bar.

Table 6. Identification of ORFs in cloned DNAs.

Insert	Bases	Top Hit	E value
		BLASTn results	
C	1-1411	*Arthrobacter aurescens* strain TC1 plasmid pAA1 atrazine catabolic region. Accession #: >gi|42558670|gb|AY456696.1|	0.0
F	1929-3619	*S. coelicolor* apt & relA genes. Accession #: >gi|1147591|emb|X87267.1|SCAPTRELA	1×10^{-45}
		BLASTx results	
C	1391-127	Transposase [*Arthrobacter aurescens*]. Accession #: >gi|42558799|gb|AAS20139.1|	0.0
	1607-2086	IS sequence [*Rhodococcus erythropolis*]. Accession #: >gi|33867201|ref|NP_898759.1|	1×10^{-41}
	3392-3102	putative DNA integrase/recombinase [*Rhodococcus erythropolis*]. Accession #: >gi|33867122|ref|NP_898680.1|	2×10^{-13}
F	3907-1559	ppGpp synthetase [*Streptomyces clavuligerus*]. Accession #: >gi|16303981|gb|AAL16895.1|	0.0
	4-819	COG0419: ATPase involved in DNA repair [*Thermobifida fusca*]. Accession #: >gi|23017669|ref|ZP_00057393.1|	6×10^{-59}

Figure 18.

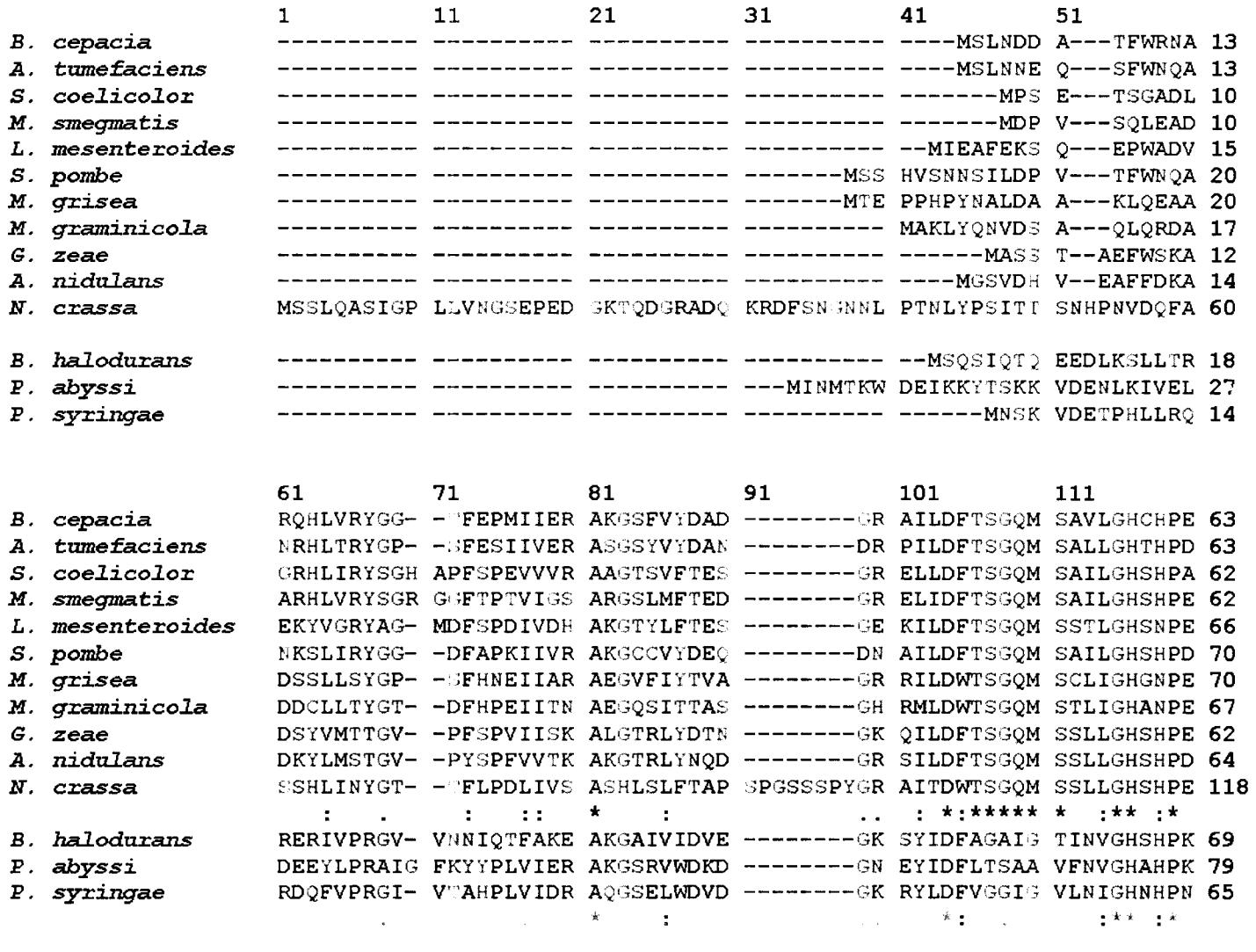

1 11 21 31 41 51
B. cepacia ---------- ---------- ---------- ---------- ----MSLNDD A---TFWRNA 13
A. tumefaciens ---------- ---------- ---------- ---------- ----MSLNNE Q---SFWNQA 13
S. coelicolor ---------- ---------- ---------- ---------- -------MPS E---TSGADL 10
M. smegmatis ---------- ---------- ---------- ---------- -------MDP V---SQLEAD 10
L. mesenteroides ---------- ---------- ---------- ---------- --MIEAFEKS Q---EPWADV 15
S. pombe ---------- ---------- ---------- -------MSS HVSNNSILDP V---TFWNQA 20
M. grisea ---------- ---------- ---------- -------MTE PPHPYNALDA A---KLQEAA 20
M. graminicola ---------- ---------- ---------- ---------- MAKLYQNVDS A---QLQRDA 17
G. zeae ---------- ---------- ---------- ---------- ------MASS T--AEFWSKA 12
A. nidulans ---------- ---------- ---------- ---------- ----MGSVDH V--EAFFDKA 14
N. crassa MSSLQASIGP LLVNGSEPED GKTQDGRADQ KRDFSNGNNL PTNLYPSITI SNHPNVDQFA 60
B. halodurans ---------- ---------- ---------- ---------- --MSQSIQTQ EEDLKSLLTR 18
P. abyssi ---------- ---------- ---------- ---MINMTKW DEIKKYTSKK VDENLKIVEL 27
P. syringae ---------- ---------- ---------- ---------- ------MNSK VDETPHLLRQ 14
61 71 81 91 101 111
B. cepacia RQHLVRYGG- -TFEPMIIER AKGSFVYDAD --------GR AILDFTSGQM SAVLGHCHPE 63
A. tumefaciens NRHLTRYGP- -SFESIIVER ASGSYVYDAN --------DR PILDFTSGQM SALLGHTHPD 63
S. coelicolor GRHLIRYSGH APFSPEVVVR AAGTSVFTES --------GR ELLDFTSGQM SAILGHSHPA 62
M. smegmatis ARHLVRYSGR GGFTPTVIGS ARGSLMFTED --------GR ELIDFTSGQM SAILGHSHPE 62
L. mesenteroides EKYVGRYAG- MDFSPDIVDH AKGTYLFTES --------GE KILDFTSGQM SSTLGHSNPE 66
S. pombe NKSLIRYGG- -DFAPKIIVR AKGCCVYDEQ --------DN AILDFTSGQM SAILGHSHPD 70
M. grisea DSSLLSYGP- -SFHNEIIAR AEGVFIYTVA --------GR RILDWTSGQM SCLIGHGNPE 70
M. graminicola DDCLLTYGT- -DFHPEIITN AEGQSITTAS --------GH RMLDWTSGQM STLIGHANPE 67
G. zeae DSYVMTTGV- -PFSPVIISK ALGTRLYDTN --------GK QILDFTSGQM SSLLGHSHPE 62
A. nidulans DKYLMSTGV- -PYSPFVVTK AKGTRLYNQD --------GR SILDFTSGQM SSLLGHSHPD 64
N. crassa SSHLINYGT- -TFLPDLIVS ASHLSLFTAP SPGSSSPYGR AITDWTSGQM SSLLGHSHPE 118
B. halodurans RERIVPRGV- VNNIQTFAKE AKGAIVIDVE --------GK SYIDFAGAIG TINVGHSHPK 69
P. abyssi DEEYLPRAIG FKYYPLVIER AKGSRVWDKD --------GN EYIDFLTSAA VFNVGHAHPK 79
P. syringae RDQFVPRGI- VTAHPLVIDR AQGSELWDVD --------GK RYLDFVGGIG VLNIGHNHPN 65

Figure 18 (cont).

```
                    121         131         141         151         161         171
B. cepacia          IVSVIGEYAG KLDHLFSGML S-RPVVDLAT RLANITPP-- -GLDRALLLS TGAESNEAAI 119
A. tumefaciens      IVATVDRQMR SVAHLFSGML S-RPVVELAS RLAALAP--- -GLDRVQLLT TGAESNEAAI 118
S. coelicolor       IVSTVREQVA HLDHLHSGML S-RPVVELAR RLAGTLPA-- -PLEKALLLT TGAEANEAAV 118
M. smegmatis        IVATVQSQIA HLDHLYSGML S-RPVIELSR RLAETLPP-- -TLDKVLLLS TGAESNEAAV 118
L. mesenteroides    IVTTLQQTVQ RLDHLYSGML S-RPNINLSK NIASKTGE-- -NLKKVIPLS TGSEVNEAAL 122
S. pombe            ITACIEKNLP KLVHLFSGFL S-PPVVQLAT ELSDLLPD-- -GLDKTLFLS TGGEANEAAL 126
M. grisea           VVSVIHQHAT SLDHLFSGML S-PPVINLAA ELKKLLPP-- -GLERSMFLS TGGESNEAAI 126
M. graminicola      IVETIHAHAR NLDHLFSGML S-PPVISLAK RLTSVTPP-- -GLDKALFLS TGGESNEAAI 123
G. zeae             VVEVVKHYVS ELDHLLSNMI T-HPVVNLAE RLAKFLPS-- -PLQKSFFLN TGSESTEAAI 118
A. nidulans         IVEVVKQYIG ELDHLLSNMI T-HPVVDLAE RLARFLPA-- -PLEKSFFLN TGSESTEAAI 120
N. crassa           IVSVISSHAS SLDHLFSGML S-PPVLNLAK RLTSVLPD-- -GLDRAMFLS TGGESNEAAI 174
                    :.  :       : ** *.:: :  * :.*:  .:           *.:   *. **.* .***:
B. halodurans       VVQAVQQQAD QFIHTGFNVM MYESYIELAE RLAALAPG-- GFEKKVLLQN SGAEAVENAV 127
P. abyssi           VVEAIKEQVD KFLNYTIGYL YTEPPVRLAE LLSEMTPG-- DFEKKVTFGF SGSDAVDSSI 137
P. syringae         VVKAIQAQLS KVTHACFQVA SYQPYLDLAK RLSLMIAGQS GIDHKAVFFT SGAEAVENAV 125
                    :.  :         :              : *:   :          .:      :*.:  : ::

                    181         191         201         211         221         231
B. cepacia          RMAKLVTGKY EIVGFAQSWH GMTGAAASAT --YSAGRKGV GPAAVGSFAI PAPFTYRP-- 175
A. tumefaciens      RMAKLVTGGH EVVAFAQSWH GMTGAAAAAT --YSAGRRGY GPVAAGSLVI PAPNSYRP-- 174
S. coelicolor       RMAKLVTGRH EIVSFARSWH GMTQAAANAT --YSAGRKGY GPAAPGNFAL PVPDRYRP-- 174
M. smegmatis        RMAKLVTGKH EIVSFARSWH GMTQAAANAT --YSAGRRGY GPAAPGNFAL PVPHRFRP-- 174
L. mesenteroides    RMAKLVTGKF EVVSFNKSWH GVTQASAGAT --YASARKSG APTSPGQLSI PTPYTYRP-- 178
S. pombe            RMAKVYTNKY ECVAFSSSWH GVTGGAASLT --FAAARRGY GPALPGSYTI PEPNPKLS-- 182
M. grisea           RMAKCVTGRW EIVGLSASWH GMTGGAAGAQ --YHSGRRGI GPMIPGNLML PAPNTYRS-- 182
M. graminicola      RLAKFYTGKF EIVGLASSWH GVTSGAIAAQ --YHAGRAGY GPNMPGNIAL PTPNSYRS-- 179
G. zeae             KIAKCYTGKF EIVAFSASYH GLTQGSGSAT --YSAGRSRG GPTMPGSLAF PAPYGYRS-- 174
A. nidulans         KIAKVYTGNF EVIAFAASYH GLTQGSGSVT --YSAGRRRG GPVMPGALAF PAPYAYRS-- 176
N. crassa           KMAKTYTGKF EVVGLGASWH GVTAQANSVQ --YHAGRRVG WPLMPGGLML PSPNAYRCHQ 232
                    ::**  *.   * :.:  *:* *:*  :       : :.*     *   *   : * *
B. halodurans       KMARKYTGRQ GIVSFSRGFH GRTLMTMTMT SKVKPYKFGF GPFAPEVYKA PYPYEYRRPE 187
P. abyssi           KASRAYTKKV HIISFRHSYH GMTYGALSVT GIVDEKVKSI VQPMSNVHIV DYPDPYRNPW 197
P. syringae         KIARARTNRP AIISFRGGFH GRTLLGTTLT GMSQPYKQNF GPMAPEVFHT PYPNEYRG-- 183
                    : ::  *      : :  .:* * *                              *
```

Figure 18 (cont).

```
                        241         251         261         271         281         291
B. cepacia              RFER------ --NGA----- ---------- ---------- -----YDYLA ELDYAFDLID 197
A. tumefaciens          RFAKP----- --DGT----- ---------- ---------- -----NDWQT ELDDAFALID 197
S. coelicolor           SVVGA----- --DGE----- ---------- ---------- -----LDWRR QLDLGFEMID 197
M. smegmatis            DITDE----- --LGE----- ---------- ---------- -----LDWRR QLDLGFDLID 197
L. mesenteroides        NFFNS----- --EGE----- ---------- ---------- -----YDWHK ELDYGFEMVD 201
S. pombe                PFRDA----- --KGN----- ---------- ---------- -----YDWQK ELDYSFYMLD 205
M. grisea               IFR------- HADGT----- ---------- ---------- -----YDWET ELDHGWAMVD 205
M. graminicola          IFR------- HPDGS----- ---------- ---------- -----HDWKT ELEYGFDLVD 202
G. zeae                 PFRKA----- --DGS----- ---------- ---------- -----WDWET EMDFGWSLID 197
A. nidulans             PFKKA----- --DGS----- ---------- ---------- -----YDWEA ELDFGWSMID 199
N. crassa               GFRKREEKGL KQDGASGGEE ESKKGDADEK ENGKKNQGEN EEEWEYDWEA EMLYGWRLID 292
                         .          *                                    *     ::  .: ::*
B. halodurans           GMSED----- ---------- ---------- ---------- --AYHAYVLQ EFIQFMNVEV 210
P. abyssi               NIDGY----- ---------- ---------- ---------- --ENPSELAN RALDEVEKKI 220
P. syringae             ---------- ---------- ---------- ---------- --VTTEVALA ALHELLATQV 201

                        301         311         321         331         341         351
B. cepacia              RQSSGNLAAF IAEPILSSGG IIELPDGYMA ALKRKCEARG MLLILDEAQT GVGRTGTMFA 257
A. tumefaciens          NQSTGSLAAF IAEPILSSGG ILELPAGYLA ALKLKCEERG MLLILDEAQT GVGRTGHMFA 257
S. coelicolor           AQSVGSLAAC LVEPILSSGG VVELPPGYLA ALADKCRERG MLLILDEAQT GLCRTGDWYA 257
M. smegmatis            AQSVGSLAAC LIEPILSSGG IIELPPGYLA ALAQKCRERE MLLIVDEAQT GLCRTGDWYA 257
L. mesenteroides        AQSVGSLAAC IVEPIVSGGG ILVPPKGYLA ALKEKCRERG MLLIFDEAQT SLGRTGEWFA 261
S. pombe                KQSTGSLACM IVETILSTGG IIELPQGYLK ALKKKCEERG MLLIIDEAQT GIGRTGSMFS 265
M. grisea               AASSGSLAAV ILEPILSSAG MHALPDGYLA AMKAHCERRG MLLILDEAQT AIGRAGDMFA 265
M. graminicola          KQSCGSLAAL ILEPILSSGG MLVLPPGYLK AVKQHCERRG MLLIIDEAQT GIGRAGDMFA 262
G. zeae                 RQSVGSLAAF IMEPILSTGG ILDLPTGYLK RMQDECRKRE MLIIMDEAQT GVGRTGKMFA 257
A. nidulans             RQSVGSIAAF IMEPILSTGG ILDPPKGYFK RMVEECRKRG ILVIMDEAQT GVGRTGQMFA 259
N. crassa               QQSCGSLAAV IVEPIQSSGG MHVLPHGYLR RLKTECEKRG MLLIVDEAQT GIGRTGEMVA 352
                          * *.:*.  : *.* * .* :   * **:   :  .*. *  :*:*.***** .: *:*   :
B. halodurans           APES--VAAV VMEPVQGEGG FIVPSKSFVQ GVYRYCKEHG ILFIADEIQT GFARTGHYFA 268
P. abyssi               KELNGDVAGI ILEPIQGDAG VVIPPLEFIK GLKKLTDEYG MVFIDEEVQT GMGRTGKWWA 280
P. syringae             APDR--VAAI LIEPIQGDGG FLTAPVEFLK ALRALTEQHG IVLILDEIQT GFGRTGKWFG 259
                              :*   : * :  . *   .  :.   :        ::.* :* ** .. *:*   .
```

Figure 18 (cont).

```
                    361         371         381         391         401         411
B. cepacia          CQRDG----V TPDILTLSKT LGAGLPLAAI VTSAAIEERA HELGYLFYTT HVSDPLPAAV 313
A. tumefaciens      FQRDG----V TPDILTLSKT LGAGLPLAAV MTTEAIEQKA FEKGYLFYTT HVSDPLLAAV 313
S. coelicolor       FEHEG----V VPDILTLSKT LGAGLPLAAV LTSAEIEQRA HERGFLFFTT HVNDPLPAAV 313
M. smegmatis        FEHDG----V VPDILTLSKT LGAGLPLAAV ITSSEIEQRA QEREFLFFTT HVNDPLPAAV 313
L. mesenteroides    YQYDD----V EPDILTLSKT LGSGLPLAAL VTTDKIEKKA HEKGFLFYTS HVNDPLVAAV 317
S. pombe            FEHHG----I VPDILTLSKS LGAGTALAAV ITSEEIEKVC YDNGFVFYTT HASDPLPAAI 321
M. grisea           FQHPS-NGGV VPDVLTLSKT LGNGLPLSAV VTSDAIAARA RGEGFLFFTT HVNDPLPAAV 324
M. graminicola      FQHFAEDDGV VPDILTLSKT LGNGLPLSAV VTSNKIASFA KENNFLFYTT HINDPLPAAV 322
G. zeae             FEHEEG---V VPDILALSKT LGCGLPLASV STTSEIERGC KEAGFLWLTT HLNDPLTAAV 314
A. nidulans         FEYDG----I VPDILALSKT LGCGLPLASV STTAEIAKGC KEAGFLWLST HINDPLTAAV 315
N. crassa           INHDG----V VPDILTLSKT LGNGLPLSAV VTSHAIADVC AERDFLFYTT HVNDPLPAAV 408
                     :         :  **:*:***:* * * .*:::  *:  *   .      ::: :: * .*** **:
B. halodurans       SEHFE----I EPDLLTVSKS LGAGVPISGV IGRREVLDEA SPG--ELGGT YSGSPLGCAA 322
P. abyssi           IEHFE----V VPDLLVSAKA LGGGMPISAV VGRAEIMDSV PSP--LFVFT HVGHAVNASA 334
P. syringae         FEHAG----I QPDLVTVAKS LAGGMPLSGV VGRAEIMDAP LPG--GLGGT YGGNALSCAA 313
                     :         :   *::  : : *  *  ::.:      :                 : :  .: .:

                    421         431         441         451         461         471
B. cepacia          GLRVLDVVQR DGLVARANVM GDRLRRGLLD LMERFDCIGD VRGRGLLLGV EIVKDRRTKE 373
A. tumefaciens      GVTVLDVVER EGLLAAAIGR GARLRAGLVS LQQRFECVGD VRGRGLLMGL EIVADRQTKA 373
S. coelicolor       GNTVLDVLVR DRLDERARRL GAALREGLDK LAARHEVVGD VRGRGLLLGM ELVGDQVLGE 373
M. smegmatis        GLTVLDVLSR DRLDVRAREL GQRLRAGLDE LAGRHPIVAD VRGRGLLLGM ELSAEDTPGV 373
L. mesenteroides    GCTVMNIIER DGLCQVTKEK GNYLHDGLQK LVDEYDIVGD ARGRGLLQGL EIIKSKSNKE 377
S. pombe            GSTVLKVVKR DGLVEKAKIS GELLRSDLLR LKDKHPLIVD VRGLGLLQGI EIASCTDPSK 381
M. grisea           GLKVLQIVQR DGLVQRSRRL GERLHAGLRR LQERYACIGD VRGRGLMAGV EIVSDR-ETK 383
M. graminicola      GDKVLEIVIR DGLVERSREM GKVFQERLRA LQSRYGCIGD VRGRGLMAGV EIVADR-VTK 381
G. zeae             GDKVLEIVER DGICQKASER GQQLRAGLEK LQQKYWCIGD LRGRGLLQGI EIIADPKTKA 374
A. nidulans         GNKVLEVVER DGIARRAAER GAQLREGLVK LQQKYWCIGD VRGRGLLQGI EIISDPETRA 375
N. crassa           GDKVLEIVVR DDLVSHARRM GEILHSGLNQ LKKRYACIGD VRGRGLMAGV EIVEDRRKGK 468
                    *  *:.:: * : :   :    *  ::  *   *  ..  : *  ** **: *: *:
B. halodurans       ALAVLDIIEE EGLNERARKL GQVVMDKMAQ LQEQYECIGH IRGLGAMCAM EVVKDRHTKE 382
P. abyssi           AIATIKVIKE EKLVERAKEL GDYALKRFRE LQEEYPIIGD VRGKGLMIGV DIVKEG-TKD 393
P. syringae         ALAVIDTYEQ DGLLARGEQL GEHLRAGLLR LKDRYACIGD VRGTGFMLAM ELTKNDAARS 373
                     .  .:.   . : :       *      :  *  ..  :  . ** * : .: ::
```

Figure 18 (cont).

```
                          481         491         501         511         521         531
B. cepacia                P-ADGLGAKI  TRECMNLGLS  MNIVQLPGMG  GVFRIAPPLT  VSEDEIDLGL  SLLGQAIERA 432
A. tumefaciens            P-GFALGAKI  MEEAMRRGLS  MNIVKLPGMG  GVFRIAPPLT  VSDEEIDHGI  EIMSDSIQAA 432
S. coelicolor             GGADRLGAAV  TRRCFELGLH  MNIVQLPGMG  GIFRIAPPLT  ASDDEIARGV  AVLDQALTDA 433
M. smegmatis              S-ADQLGAVV  TRRCFELGLH  MNIVQLPGMG  GTFRLAPPLT  ATDNEIDRAL  EIMDEALAYA 432
L. mesenteroides          R-SEFIGDEI  TKRCYKLGLH  MNIVNLPGMG  GVFRIAPPLT  VSYEELDSGL  AILEQSIKSV 436
S. pombe                  P-SDFLGTVI  GDKCLELGMN  CNIVHLRGIG  GVFRIAPPLT  VTDEEIHKAI  EIFDSALTFT 440
M. grisea                 APGMELADRL  GKRMMELGVS  ANLATLASFG  GAFRIAPPIT  ISEEELDLGL  GFMEETLRST 443
M. graminicola            EGDIDLGAAI  GKRMTENGLW  AQLSTMASFG  GVFRLAPPLT  TTDEELMAGL  DIIEEAFAST 441
G. zeae                   P-GADLGQAI  SDKAMELGLS  CNVVNLPGMG  GVFRLAPAVT  VTAEEIEQGL  EILDKSFGAV 433
A. nidulans               P-GPELGQAV  SDQAMTKGLS  CNVVNLPGMG  GVFRLAPPVT  VTAEEIEEGL  AILDEAFGDV 434
N. crassa                 EPGLELAKRI  GDRAYELGLW  CNLSTHPSFG  GTFRIAPPIT  ISEKEVREGL  AVLEEAFRGV 528
                             :.  :    .    *:   ::    .:* * **:**.:*    .*:  .:  .: .::
B. halodurans             P-DKELTAKI  VSEAGNRGLL  L--LSAGQYG  NVLRFLMPIV  ITDEQLNNGF  AILEEALATA 439
P. abyssi                 P-NRELAQKI  CWRAWEKGLI  I--ITFGKHG  NVLRIAPPLT  ISKEDFDRGI  EIIEEAIKDA 450
P. syringae               P-DADLNQKV  IDQARIRGLL  V--IKCGVYR  NVLRFLAPLV  TTEQQIDEAL  SILDAALARV 430
                             :   :       :                   :*:  :     .:     .:  ::

                          541         551         561
B. cepacia                L---------  ----------  --- 433
A. tumefaciens            SGA-------  ----------  --- 435
S. coelicolor             ADAL------  ----------  --- 437
M. smegmatis              REPAVHRS--  ----------  --- 440
L. mesenteroides          QNDINLGNL-  ----------  --- 445
S. pombe                  AKEFSGSY--  ----------  --- 448
M. grisea                 EGSMPTGAIR  K---------  --- 454
M. graminicola            PGTKPLYERD  VAAAYDSQVD  ARL 464
G. zeae                   LELRGQMTAA  A---------  --- 444
A. nidulans               LKTWSASESD  SFLGGLFK--  --- 452
N. crassa                 EGTLPLY---  ----------  --- 535

B. halodurans             IVNR------  ----------  --- 443
P. abyssi                 INGKVPDDVI  KELRAW----  --- 466
P. syringae               LKSS------  ----------  --- 434
```

Figure 18 (cont).

Figure 18. Mutiple sequence alignment of the *B. cepacia* DgdA, putative DgdAs, (upper 10) and related transaminases (lower 3). Numbers at the top of the alignment keep a running tally of the residues in the entire alignment and numbers at the ends of each line reports residue numbers of specific sequences. Colored letters catagorize the amino acids; red letters represent residues with hydrophobic side chains; green, polar side chains; pink, positively charged side chains; and blue, negatively charged side chains. Symbols designate the level of consensus at a given position in the alignment; a "*" denotes the residue is identical in all sequences at that position; a ":" denotes that conserved substitutions have been observed at a position; and a "." denotes that semi-conserved substitutions have been observed. The first line of symbols refer to the 11 sequences above it, the lower line of symbols refer to all 14 aligned sequences.

Figure 19.

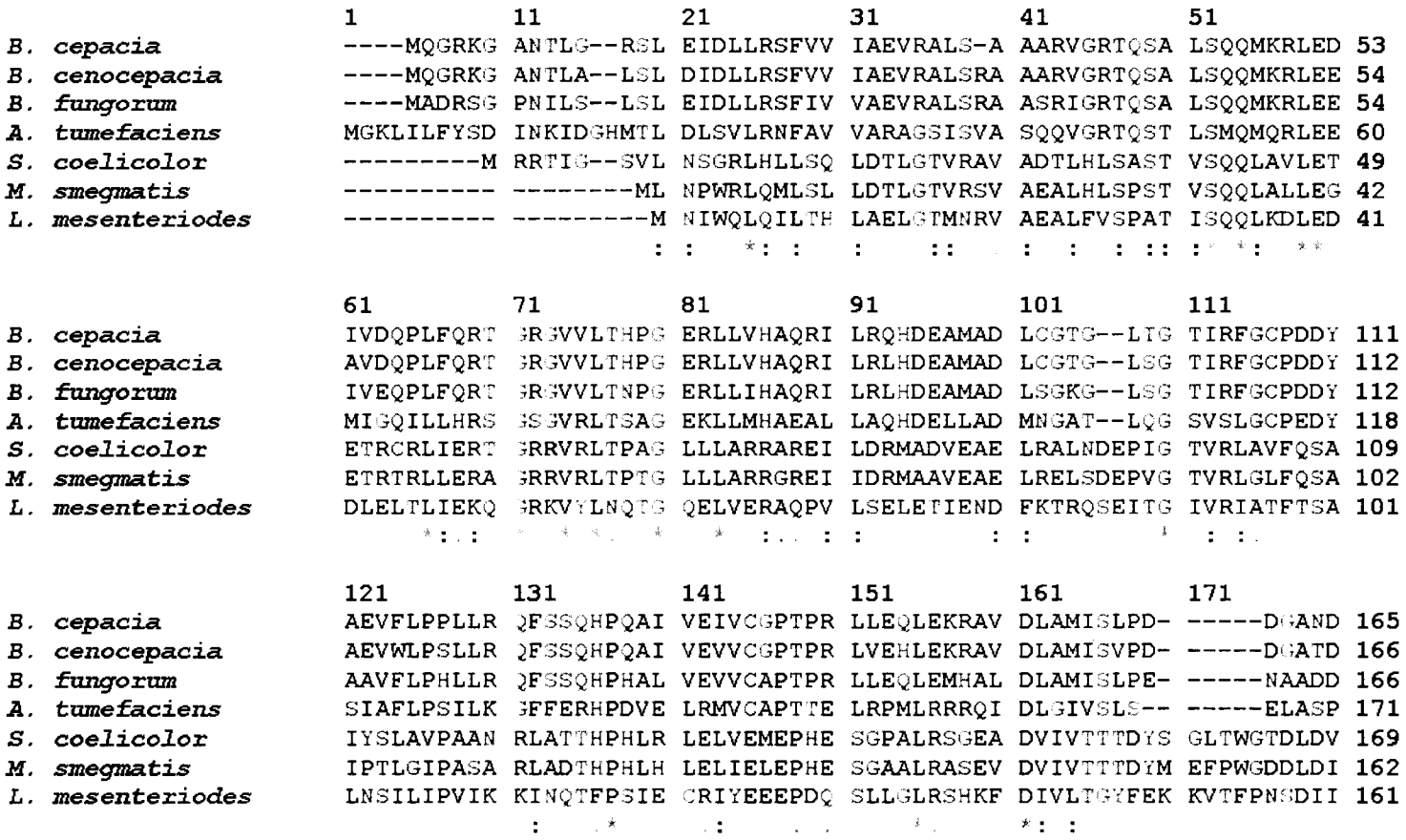

1 11 21 31 41 51
B. cepacia ----MQGRKG ANTLG--RSL EIDLLRSFVV IAEVRALS-A AARVGRTQSA LSQQMKRLED 53
B. cenocepacia ----MQGRKG ANTLA--LSL DIDLLRSFVV IAEVRALSRA AARVGRTQSA LSQQMKRLEE 54
B. fungorum ----MADRSG PNILS--LSL EIDLLRSFIV VAEVRALSRA ASRIGRTQSA LSQQMKRLEE 54
A. tumefaciens MGKLILFYSD INKIDGHMTL DLSVLRNFAV VARAGSISVA SQQVGRTQST LSMQMQRLEE 60
S. coelicolor ---------M RRTIG--SVL NSGRLHLLSQ LDTLGTVRAV ADTLHLSAST VSQQLAVLET 49
M. smegmatis ---------- --------ML NPWRLQMLSL LDTLGTVRSV AEALHLSPST VSQQLALLEG 42
L. mesenteriodes ---------- ---------M NIWQLQILTH LAELGTMNRV AEALFVSPAT ISQQLKDLED 41
61 71 81 91 101 111
B. cepacia IVDQPLFQRT GRGVVLTHPG ERLLVHAQRI LRQHDEAMAD LCGTG--LTG TIRFGCPDDY 111
B. cenocepacia AVDQPLFQRT GRGVVLTHPG ERLLVHAQRI LRLHDEAMAD LCGTG--LSG TIRFGCPDDY 112
B. fungorum IVEQPLFQRT GRGVVLTNPG ERLLIHAQRI LRLHDEAMAD LSGKG--LSG TIRFGCPDDY 112
A. tumefaciens MIGQILLHRS GSGVRLTSAG EKLLMHAEAL LAQHDELLAD MNGAT--LQG SVSLGCPEDY 118
S. coelicolor ETRCRLIERT GRRVRLTPAG LLLARRAREI LDRMADVEAE LRALNDEPIG TVRLAVFQSA 109
M. smegmatis ETRTRLLERA GRRVRLTPTG LLLARRGREI IDRMAAVEAE LRELSDEPVG TVRLGLFQSA 102
L. mesenteriodes DLELTLIEKQ GRKVYLNQTG QELVERAQPV LSELETIEND FKTRQSEITG IVRIATFTSA 101
121 131 141 151 161 171
B. cepacia AEVFLPPLLR QFSSQHPQAI VEIVCGPTPR LLEQLEKRAV DLAMISLPD- -----DGAND 165
B. cenocepacia AEVWLPSLLR QFSSQHPQAI VEVVCGPTPR LVEHLEKRAV DLAMISVPD- -----DGATD 166
B. fungorum AAVFLPHLLR QFSSQHPHAL VEVVCAPTPR LLEQLEMHAL DLAMISLPE- -----NAADD 166
A. tumefaciens SIAFLPSILK GFFERHPDVE LRMVCAPTTE LRPMLRRRQI DLGIVSLS-- -----ELASP 171
S. coelicolor IYSLAVPAAN RLATTHPHLR LELVEMEPHE SGPALRSGEA DVIVTTTDYS GLTWGTDLDV 169
M. smegmatis IPTLGIPASA RLADTHPHLH LELIELEPHE SGAALRASEV DVIVTTTDYM EFPWGDDLDI 162
L. mesenteriodes LNSILIPVIK KINQTFPSIE CRIYEEEPDQ SLLGLRSHKF DIVLTGYFEK KVTFPNSDII 161

Figure 19 (cont).

```
                        181         191         201         211         221         231
B. cepacia              DIIRREQLVW IGYPGLEPAH FDP--LPLAL SD------PD TLDHIAACDA LHRAGRDYRV 217
B. cenocepacia          DVIRREPLVW IGYPGLEPAH FDP--LPLAL SD------RD TLDHLAACDA LNRAGRDYRV 218
B. fungorum             DIIRREPLVW VGCPGMDSAH FDP--LPLAV SD------PD TLDHVAACEA LQRAGRAYRI 218
A. tumefaciens          EIIRREDFVW VANS-SRPAI LDRPVLPLAL SA------PA TLDYRAACDV MGASNRRYRV 224
S. coelicolor           MSLGSDPVLL VLPNGHPLAA RAAVDLAACK EETWACDRPQ SYMADLTVRL CRESGFEPRV 229
M. smegmatis            VAIGTDPIVL VLPQDHRLGR RPVVDLATCS DETWVCDRPD SYMAELTLRL CRQSGFEPRV 222
L. mesenteriodes        STNIGEDKLL VVSSKDDKSI FPIKHFSDLA SKKWIIEPKN TYLSNHIIQM CRENLFDPDV 221
                             :  :  :               :.                :                   :

                        241         251         261         271         281         291
B. cepacia              AYASSSLAGL IALVRSGQAF AVMTQTAVPA ---DLAIVNG DPRLPPLPAV GITLKF-DRK 273
B. cenocepacia          AYASSSLAGL IALVRSGQAF AVMTQTAVPA ---DLAIVHG DPRLPPLPAV GITLRF-DRK 274
B. fungorum             AYASSSLAGL IALVRSGQAF AVITQTAVAP ---DLHILNA DAALPPLPTI GITLKF-ERK 274
A. tumefaciens          AFASNSLAGL IAIARSGHAI SVFTRTAVPP ---DLFVITD G--LPTLPTI GLAVEL-GES 278
S. coelicolor           ACRFSNYLML LRHVETTGSI ALLPALAVTA ---DHAVVTR RLSPPVHRNV AVVVRRGAPQ 286
M. smegmatis            AGRFSTVSLL VHHVQTQGSV ALLPALAVGP ---EQKVLTR ELTLPVHRNV AAAFRR-SAA 278
L. mesenteriodes        VSEIHSYDAI KSLVKSQLGI SILPSLALTA GTDMNGIFIH DFDIKSKRQI YLLERKPQKR 281
                              .      :      :    :::.  *: .        :                :

                        301         311         321         331
B. cepacia              RPSHLTAAFA EHIRAVLPML ---------- ------ 293
B. cenocepacia          RPSHLTAAFA EHIRQTLPLL ---------- ------ 294
B. fungorum             RPSHLVTVFA EHIRLTLPLL ---------- ------ 294
A. tumefaciens          RPSLIAKALA KHVRSTLPRL ---------- ------ 298
S. coelicolor           RAAVNAVITA LRDHPEIEAL SAPAPSREGR ETGPRS 322
M. smegmatis            RPAVHAVVDA LRDHPEIPAL TER------- ------ 301
L. mesenteriodes        IPAIKIVSQM ITDQAIAVLK NFN------- ------ 304
                         .:          :
```

Figure 19 (cont).

Figure 19. Mutiple sequence alignment of the *B. cepacia* DgdR with putative DgdRs. Numbers at the top of the alignment keep a running tally of the residues in the entire alignment and numbers at the ends of each line reports residue numbers of specific sequences. Colored letters catagorize the amino acids; red letters represent residues with hydrophobic side chains; green, polar side chains; pink, positively charged side chains; and blue, negatively charged side chains. Symbols designate the level of consensus at a given position in the alignment; a "*" denotes the residue is identical in all sequences at that position; a ":" denotes that conserved substitutions have been observed at a position; and a "." denotes that semi-conserved substitutions have been observed.

Table 7. Sequence identity between *B. cepacia* DgdA and putative DgdAs.

Organism[a]	Length (a.a.)	% Identity
Burkholderia cepacia	**433**	-
Burkholderia cenocepacia	**433**	**95**
Burkholderia fungorum	**433**	**93**
Agrobacterium tumefaciens	**435**	**69**
Streptomyces coelicolor	**437**	**61**
Mycobacterium smegmatis	**440**	**60**
Magnaporthe grisea	454	57
Mycosphaerella graminicola	464	55
Schizosaccharomyces pombe	448	55
Neurospora crassa	535	54
Leuconostoc mesenteroides	**445**	**53**
Gibberella zeae	444	52
Aspergillus nidulans	452	51

[a]Bold names are bacteria, the others are fungi.

Table 8. Conservation of various DgdA residues.

Residue position[a]		11 DgdAs[b]			11 DgdAs + 3 transaminases[b]		
B.cepacia	Align	Match	Strong	None	Match	Strong	None
K^+/Cs^+ binding							
His77	134	X				X	
Leu78	135	X					X
Ser80	137	X					X
Thr303	410		X			X	
His304	411	X				X	
Val305	412			X			X
Asp307	414	X					X
Na^+ binding							
Ala95	153			X			X
Thr98	156			X			X
Pro99	157			X			X
Leu102	163	X					X
Asp103	164			X			X
Active Site							
Gln52	109	X					X
Phe79	136			X			X
Ser80	137	X					X
Gly111	172	X			X		
Ala112	173			X			X
Asn113	174	X				X	
Trp138	199		X			X	
His139	200	X			X		
Met141	202		X				X
Glu210	313	X			X		
Ser215	318			X			X
Asp243	346	X				X	
Ala245	348	X					X
Gln246	349	X			X		
Lys272	379	X			X		
Tyr301	408			X			X
Thr303	410		X			X	
Gln394	502		X				X
Arg406	514	X			X		

[a]The first column gives the residue position of DgdA in *B. cepacia*, the Align column gives the position in the multiple sequence alignment in Figure 20.

[b]Level of conservation is shown by placing an X in one of three boxes: Match, Strong, and None, which correspond to the "*" and ":" symbols from the alignment in Figure 20, respectively (the "." Symbol was omitted because there was not any semi-conserved substitutions observed for the residues listed here). The boxes are listed twice, once for the alignment of the 11 DgdA (or putative DgdAs), and again with the addition of transaminases from *B. halodurans, P. abyssi, and P. syringae.*

Figure 20.

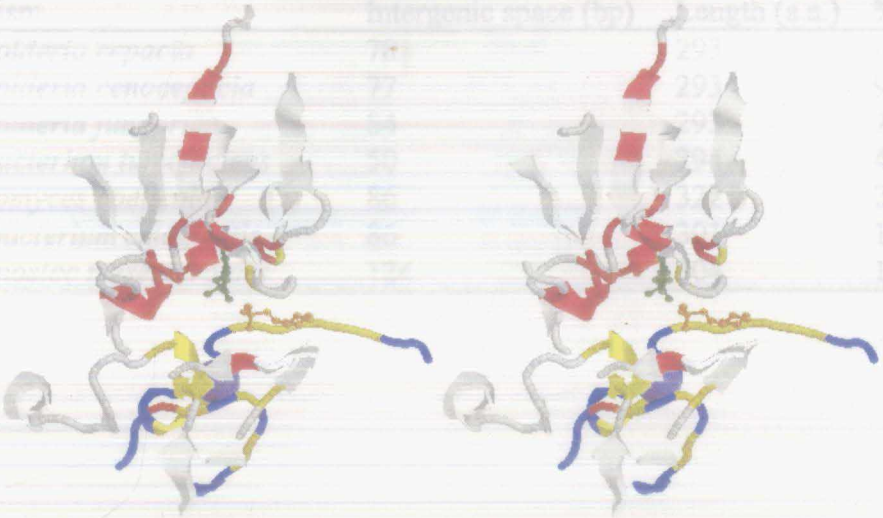


Figure 20. Stereo model of the *B. cepacia* DgdA active site. This model is based on the crystal structure by Toney et al.[21] The separate chains which make up the active site are white and blue, the PLP cofacter is green, the K^+ metal ion is purple, and a MES molecule (which mimics the substrate) is orange. Highlighted residues are conserved in a multiple sequence alignment of the *B. cepacia* DgdA, 10 putative DgdA sequences and 3 closely related transaminase sequences. Red residues are conserved across the DgdA and transaminase sequences, while the yellow residues are only conserved in the DgdA sequences.

Table 9. Sequence identity between *B. cepacia* DgdR and putative DgdRs.

Organism	Intergenic space (bp)	Length (a.a.)	% Identity
Burkholderia cepacia	78	293	-
Burkholderia cenocepacia	77	293	91
Burkholderia fungorum	84	293	78
Agrobacterium tumefaciens	50	294	43
Streptomyces coelicolor	86	322	21
Mycobacterium smegmatis	86	301	18
Leuconostoc mesenteroides	124	304	13

Figure 21.

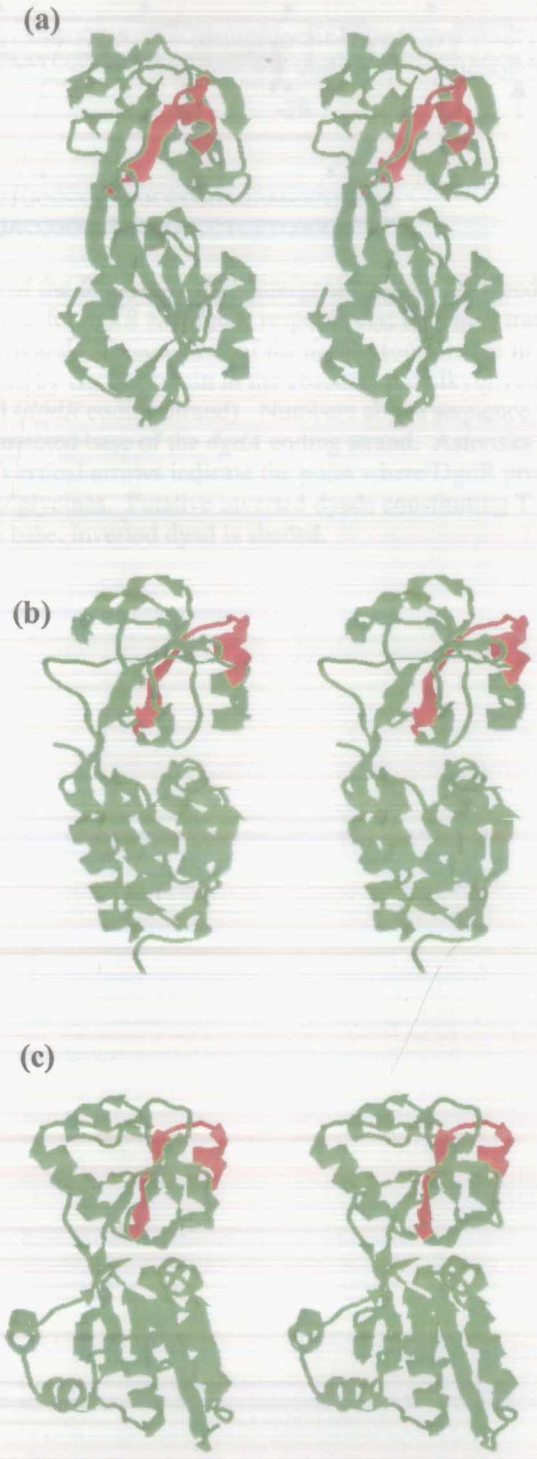


Figure 21. Stereo models of the (a) CbnR, (b) OxyR, and (c) CysB transcriptional regulators.[24-26] Red residues are homologous to a conserved region of the DgdR protein that was observed from the multiple sequence alignment performed in Figure 19.

Figure 22.

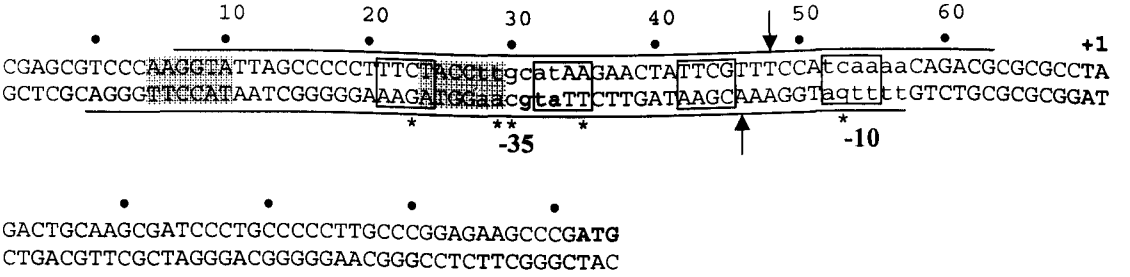


Figure 22. Sequence of the *B. cepacia dgd* intergenic region. Deduced by Allen Daley et al.[29] Bases in bold are the start codons for *dgdR* and *dgdA*, respectively; the *dgdA* transcription start point is designated with a +1, possible -10 and -35 sequences of the *dgdA* promoter are in lowercase. The regions protected from DNase I digestion by DgdR protein in the absence of dialkylglycines are overlined (*dgdA* coding strand), or underlined (*dgdR* coding strand). Numbers above sequence refer to entire protected region starting at the first protected base of the *dgdA* coding strand. Asterisks indicate bases that experience enhanced cleavage. Vertical arrows indicate the point where DgdR protection is shortened to upon introduction of dialkylglycines. Putative inverted dyads constituting T-N_{11}-A LTTR recognition sequences are boxed; a larger, 6 base, inverted dyad is shaded.

Figure 23.

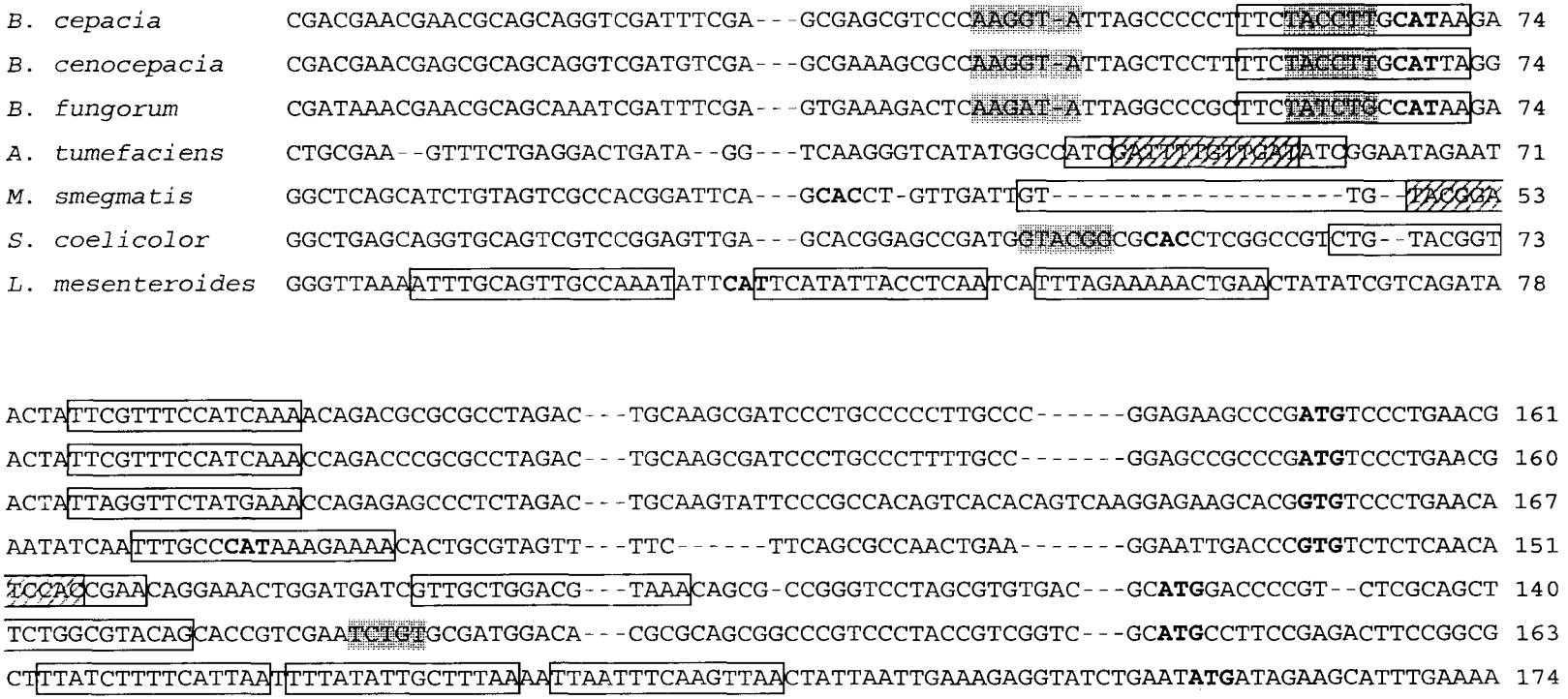


Figure 23. Multiple sequence alignment of the *dgdA-dgdR* intergenic regions from seven organisms. Bases in bold are start codons for the *dgdR* and *dgdA* genes, respectively. Boxed regions contain a putative T-N_{11}-A LTTR recognition sequence; areas where T-N_{11}-A sequences overlap are filled with diagonal slashes. Areas of interest that do not contain a T-N_{11}-A sequence are shaded grey. In the three *Burkholderia* sequences, the shaded areas highlight an inverted dyad of 6 bases interrupted by 13 bases. In *S. coelicolor*, the T-N_{11}-A sequence is flanked by repeated sequences 22 bp away.

Figure 24.

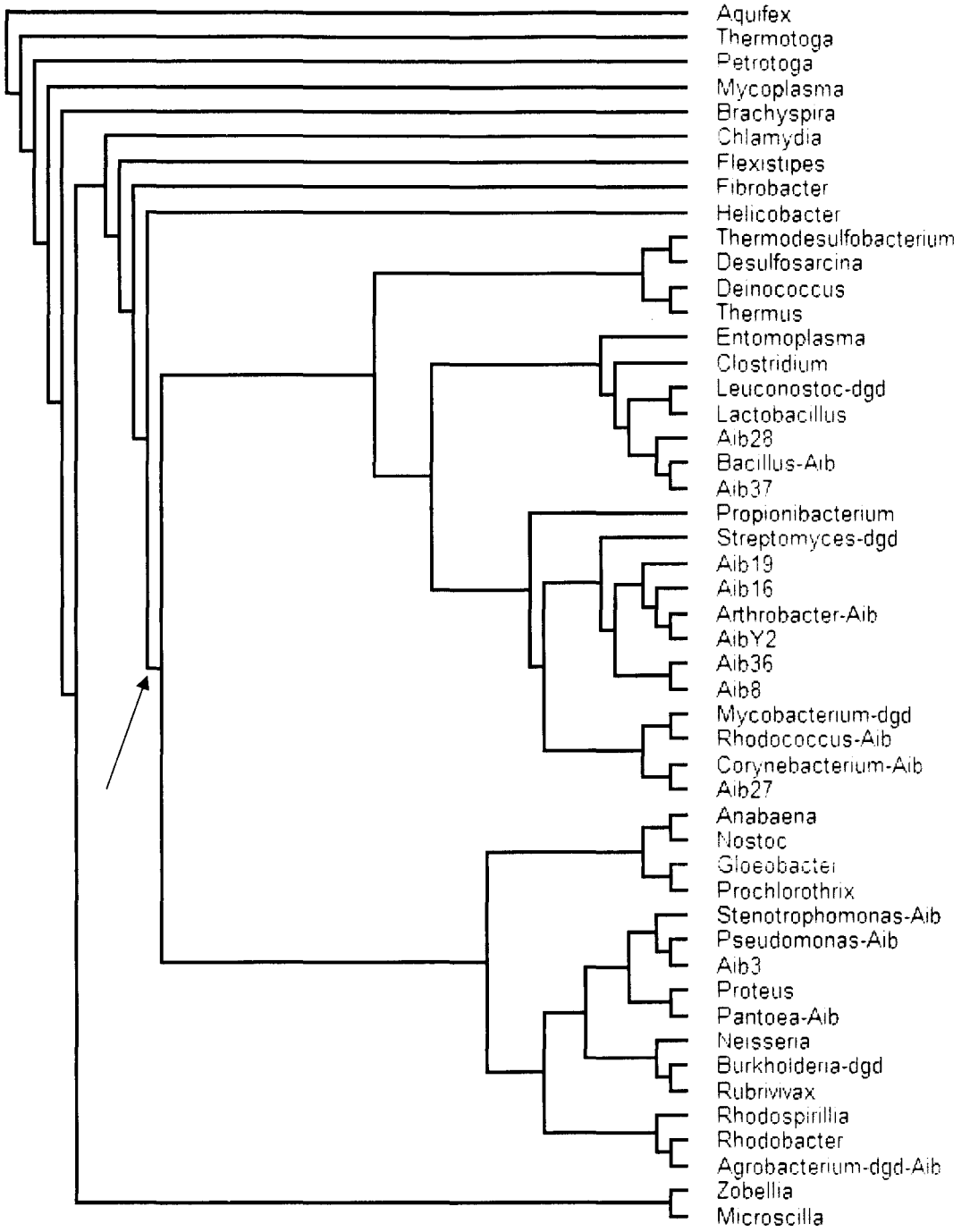


Figure 24. Phylogram of 49 bacterial genera, including nine Aib-metabolizing isolates. Based on complete or nearly-complete sequences of 16S rRNA genes (~1400-bp) from representative species and rooted at *Aquifex*. "Aibn" are isolates grow on Aib plates as the sole nitrogen source. "x-Aib" genera are the most closely related to organisms isolated in this work and have 16S rRNA genes sequenced. "y-dgd" organisms contain *dgdA* and *dgdR* genes as determined by cloning and sequencing, or by whole genome sequencing. The arrow indicates the point during evolution where ancestral *dgd* genes may have originated.

Table 10. G+C Content of *dgd* genes.

Organism	% G+C			
	Whole genome	*dgdA* gene	Intergenic region	*dgdR* gene
Leuconostoc mesenteroides	37	39	22	34
Agrobacterium tumefaciens	62	60	44	63
Burkholderia fungorum	62	63	48	63
Burkholderia cepacia	67	69	58	68
Burkholderia cenocepacia	67	69	60	68
Mycobacterium smegmatis	69	68	55	69
Streptomyces coelicolor	72	74	67	74

Chapter 3. Organic Synthesis

Introduction

This section describes progress toward the synthesis of disubstituted N-acyl amino acids, which are precursors to α-disubstituted glycines. Alpha-di-substituted amino acids are important substituents of antibiotic peptides,[1] they are able to induce and stabilize different types of conformations when incorporated into small peptides;[2] and they are useful, in terms of this research, as potential inducers of the LysR type transcription regulator, DgdR, which regulates expression of the dialkylglycine decarboxylase gene *dgdA* in certain bacteria.[3] Specifically, this section focuses on the C4-alkylation of 5(4H) oxazolones, **2** in Scheme 4, with various alkyl halides and the use of the 4-methyl-2-(4-nitrophenyl)-5(4H)-oxazolone enolate (**2a**) as an intense blue indicator for the alkylation reaction. Briefly, the synthetic steps investigated are shown in Scheme 4 and include: a) conversion of alanine to N-*p*-nitrobenzoyl-DL-alanine **1** using *p*-nitrophenyl benzoyl chloride, b) dehydration and cyclization of **1** using acetic anhydride to form 4-methyl-2-(4-nitrophenyl)-5(4H)-oxazolone **2**, c) formation of enolate **2a** and alkylation of **2a** at the 4-position with various alkyl halides to make **3a-c**, and d) ring opening of **3** with acid to form a N-*p*-nitrobenzoyl disubstituted glycines, **4a-c**.

Materials and Methods

Miscellaneous. All chemicals and media used were of reagent grade or better. Water used was purified to resistively greater than 10 MΩ with a Milli-Q type I reagent grade water purification system. Melting points of compounds were taken on a Laboratory Devices Inc. Mel-Temp 3.0. Nuclear magnetic resonance (NMR) spectroscopy was performed on a Varian 300 MHz NMR spectrometer and spectra were analyzed either on the Varian VNMR software computer program version 6.1, revision C or the Advanced Chemistry Development Inc. ACD / Spec Viewer software, version 3.50.

General procedures

Synthesis of N-*p*-nitrobenzoyl-DL-alanine (1). Based on protocol from a study by Chenault et al.[4] To a 100-mL round-bottom flask equipped with magnetic stirrer, hot plate, reflux condenser, and Ar inlet was added 3.68 g (41.2 mmol) DL-alanine, 50 mL acetonitrile (HPLC grade), and 8.05 g (43.4 mmol) of *p*-nitrobenzoyl chloride (Aldrich). Prior to adding reagents, the flask was briefly dried by heating under a slow stream of Ar. The mixture was refluxed until the solids had dissolved and no acidic gas was detected in the Ar outlet stream using wet pH paper (12 h). Upon completion of the reaction, the flask was cooled to give 4.34 g of crystallized amide. Evaporation of remaining acetonitrile gave a total of 9.3 g (95% yield): mp 166-167 °C (lit[5] 166-167 °C); ^{1}H NMR (d_6-acetone, 300 MHz) δ 1.54 (d, J = 7.6 Hz , 3H, -CH_3), 4.69 (dq, J = 7.6, 7.6 Hz, 1H, -

CH), 8.16 (d, J = 9.4 Hz, ArH, 2H), 8.28 (s, -NH, 1H), 8.32 (d, J = 9.4 Hz ArH, 2H) (Figure 25).

Synthesis of 4-methyl-2-(4-nitrophenyl) 5(4H)-oxazolone (2). Based on a protcol from a study by Gelmi et al.[6] To a 50-mL round-bottom flask immersed in a 70 °C water bath, equipped with thermometer, magnetic stirrer, hot plate, and condenser, was added 4.3 g (18 mmol) of **1** and 13 mL of acetic anhydride. Reagents were stirred for 30 min during which time the solid completely dissolved. The reaction flask was cooled to -20°C overnight, where 1.5 g of crystallized product was suction filtered, then dried *in vacuo*. The filtrate was reduced *in vacuo* and subsequent cooling gleaned a total of 3.8 g (95% yield) of **2**: mp 127-129 °C (lit[7] 127-129 °C) ; ^{1}H NMR (d_6-acetone, 300 MHz) δ 1.56 (d, J = 7.7 Hz, -CH_3, 3H), 4.70 (q, J = 7.7 Hz, -CH, 1H), 8.25 (d, J = 8.9 Hz, ArH, 2H), 8.43 (d, J = 8.9 Hz, ArH, 2H) (Figure 27).

The C4-alkylation of 2 to make 3. Alkylation reactions of **2** utilized six alkyl halides in separate reactions: allyl bromide, benzyl bromide, bromopropane, crotyl chloride, iodomethane, and propargyl bromide. In the following sections the descriptions of the alkylation reactions will relate a typical method that used a specific alkyl halide.

The phase transfer alkylation reaction (Method A). Based on a protocol from a study by Gelmi et al.[6] To a 25 mL rb flask equipped with magnetic stirrer was added 3.5 mL dichloromethane, 0.221 g **2**, and 0.2 mL propargyl bromide. A solution of 0.383 g

sodium carbonate monohydrate and 0.034 g tetrabutylammonium bromide in 2.3 mL water was added. Upon adding the tetrabutylammonium bromide a characteristic intense blue color appeared. Stirring continued for 17 min upon which the blue color faded to pale yellow. To this solution 1.00 mmol of the internal standard 2-methoxynaphthalene was added and stirring continued until dissolved. Phases were separated, the aqueous phase was extracted once with 5 mL dichloromethane, and the combined organic phases extracted once with 5 mL water (final pH = 6.5). The organic phase was dried over anhy magnesium sulfate, evaporated *in vacuo*, to give 0.42 g crude yellowish solid. A 10 mg sample of crude solid was added to 0.7 mL of d_6-acetone for NMR purposes. Integrals indicated a 75% yield. A portion of crude product was purified using a Harrison Research 89E Chromatotron centrifugal thin-layer chromatograph eluted with ethyl acetate and methylene chloride. Collected fractions were evaporated for NMR spectroscopic analysis.

Alternatively, a similar reaction was performed under positive pressure of argon. This increased the time required for the intense blue reaction solution to fade from 17 min to overnight. Integration of NMR peaks from product spectra indicated a 75% yield.

Alkylation via inverse addition of NaH (Method B). Based on a protocol from a study by Obrecht et al.[2] To a flame dried 50-mL three-necked flask equipped with magnetic stirrer, ice bath, and under positive pressure of dry Ar, was added 15 mL of distilled DMF, 1.105 g of **2** (5.00 mmol), and 0.865 mL of benzyl bromine (10.0 mmol). Under constant stirring, 240 mg of 57% NaH dispersion in oil was added to the solution in small

portions. Upon adding NaH a characteristic deep blue color appeared indicating the presence of the enolate. Stirring continued for 30 min at 10-15°C and continued at room temperature until blue color faded to a clear yellow solution. To this solution was added 5 mmol of the internal standard 2-methoxynaphthalene and stirring continued until the compound was dissolved. The remainder of the solution was poured onto ice (25 g), 0.1 N aq. HCl (15 mL), and AcOEt (50 mL), the org. phase washed with H_2O (2x15 mL) and brine (25 mL), dried ($MgSO_4$), and evaporated to provide a yellow crude solid. NMR data from a 50 μL aliquot in d_6-acetone indicated 72% yield. A portion of crude product was purified using a Harrison Research 89E Chromatotron centrifugal thin-layer chromatograph with ethyl acetate and methylene chloride from which fractions were evaporated for ^{1}H NMR spectroscopic analysis.

Compounds synthesized by both Methods A and B:

4-methyl-4-propargyl-2-(4-nitrophenyl) 5-oxazolone (3a): mp 132-134 °C; ^{1}H NMR (d_6-acetone, 300 MHz) δ 1.61 (s, -CH_3, 3H), 2.54 (t, J = 2.3 Hz, -CH, 1H), 2.87 (dd, J = 2.3, 2.3 Hz, -CH_2-, 2H), 8.29 (d, J = 8.8 Hz, ArH, 2H), 8.48 (d, J = 8.8 Hz, ArH, 2H) (Figure 28).

4-methyl-4-allyl-2-(4-nitrophenyl) 5-oxazolone (3b): mp 132-134 °C; ^{1}H NMR (d_6-acetone, 300 MHz) δ 1.56 (s, -CH_3, 3H), 2.64 (m, -CH_2-, 2H), 5.15 (m, -CH_2, 2H), 5.69 (m, -CH-, 1H), 8.25 (d, J = 8.8 Hz, ArH, 2H), 8.43 (d, J = 8.8 Hz, ArH, 2H) (Figure 29).

4-methyl-4-benzyl-2-(4-nitrophenyl) 5-oxazolone (3c): yellow oil; ^{1}H NMR (d_6-acetone, 300 MHz) δ 1.64 (s, -CH_3, 3H), 3.20 (d, J = 13.5 Hz, -CH_2-, 1H), 3.25 (d, J = 13.5 Hz, -CH_2-, 1H), 7.18 (s, ArH, 4H), 7.35 (m, ArH, 1H) (Figure 30).

Ring opening of 3 to form 4. To an NMR tube containing 10 mg of **3** dissolved in 0.7 mL of d_6-acetone, one drop of a 20% solution of DCl in D_2O was added, then an NMR spectrum was obtained immediately.

Compounds resulting from ring opening of 3.

N-*p*-nitrobenzoyl-2-methyl-2-propargyl-DL-glycine (4a): ^{1}H NMR (d_6-acetone, 300 MHz) δ 1.66 (s, -CH_3, 3H), 2.48 (t, J = 2.9 Hz, -CH, 1H), 2.86 (dd, J = 17.7, 2.9 Hz, -CH_2-, 1H), 3.32 (dd, J = 17.7, 2.9 Hz, -CH_2-, 1H), 8.12 (d, J = 8.8 Hz, ArH, 2H), 8.32 (d, J = 8.8 Hz, ArH, 2H) (Figure 32).

N-*p*-nitrobenzoyl-2-methyl-2-allyl-DL-glycine (4b): ^{1}H NMR (d_6-acetone, 300 MHz) δ 1.61 (s, -CH_3, 3H), 2.80 (dd, J = 13.9, 6.3 Hz, -CH_2-, 1H), 2.94 (dd, J = 13.9, 6.3 Hz, -CH_2-, 1H), 5.13 (m, -CH_2, 2H), 5.85 (m, -CH-, 1H), 8.12 (d, J = 8.7 Hz, ArH, 2H), 8.33 (d, J = 8.7 Hz, ArH, 2H) (Figure 33).

N-*p*-nitrobenzoyl-2-methyl-2-benzyl-DL-glycine (4c): ^{1}H NMR (d_6-acetone, 300 MHz) δ 1.60 (s, -CH_3, 3H), 3.37 (d, J = 13.5 Hz, -CH_2-, 1H), 3.59 (d, J = 13.5 Hz, -CH_2-, 1H), 7.23 (m, ArH, 5H), 8.04 (d, J = 8.8 Hz, ArH, 2H), 8.32 (d, J = 8.8 Hz, ArH, 2H) (Figure 34).

Results and Discussion

Reaction of alanine and N-*p*-nitrobenzoyl chloride to give 1. Compound **1** was prepared by refluxing N-*p*-nitrobenzoyl chloride and DL-alanine in acetonitrile under positive pressure of argon. The NMR spectrum of **1,** shown in Figure 25, is consistent with the proposed structure. The α-hydrogen is coupled to its neighboring methyl group and the single amido hydrogen. This produces a doublet of quartets which appear as a quintuplet at δ 4.69 ppm where J = 7.6 Hz. The coupled methyl appears as a doublet at δ 1.53 ppm where J = 7.6 Hz. In the aromatic region the *o*-hydrogens appear as a doublet at δ 8.32 ppm and appear to partially obscure the amido hydrogen peak at δ 8.28 ppm. At δ 8.16 ppm another doublet corresponds to phenyl *m*-hydrogens. Integrals of peaks indicated a 2 to 1 ratio of the phenyl *m*-hydrogens to the α-hydrogen quintuplet, a 3 to 1 ratio of methyl to the α-hydrogen, and a 3 to 2 ratio of phenyl *o*-hydrogens to *m*-hydrogens. A COSY performed on the same sample (Figure 26) confirms the α-hydrogen couples to the methyl doublet and the amido hydrogen. Compound **1** of this purity was used in subsequent reactions.

Dehydration of 1 with acetic anhydride gave 2. Preparation of **2** was performed by dehydrating the N-(4-nitrobenzoyl)-DL alanine with acetic anhydride. Pure amide in acetic anhydride was heated for 30 minutes at 70°C under constant stirring. Crystallization of **2** was carried out by cooling the reaction mixture to -20°C and filtering the crystals. Figure 27 shows an NMR spectrum of **2**, which is similar to the spectrum of

1. As in **1**, the methyl hydrogens appear as a doublet at δ 1.56 ppm and the phenyl *o*- and *m*-hydrogens appear as doublets at δ 8.42 and 8.26 ppm, respectively. However, the cyclization of **1** to **2** resulted in loss of water and left the α-hydrogen coupled to one less hydrogen, replacing the doublet of quartets depicted in Figure 25 with a quartet at δ 4.70 ppm. Product of this purity crystallized out of the completed reaction at -20°C and was used in subsequent reactions.

Distinction between alkylation method A and B is reaction time and percent yield. Both methods investigated were used to perform alkylations of **2** employing the alkyl halides: propargyl bromide, allyl bromide, and benzyl bromide to give compounds **3a**, **3b**, and **3c**, respectively. Method A involved adding **2** and an alkyl halide to dichloromethane and combining it with vigorous stirring to an aqueous solution of tetrabutylammonium bromide and sodium carbonate monohydrate. Method B was performed by addition of NaH to a DMF solution of **2** and an alkyl halide at 10-15°C, under stirring and positive argon flow. The results in Table 11 show that method A is superior to B in that it has faster reaction times and higher percent yields. In both methods products were purified using thin layer chromatography (Figures 28-30).

The empirical data of compounds **3a-c** are shown in Table 12, these data are similar. This includes the melting points of compounds **3a** and **3b**, both 132-134°C, and the ^{1}H-NMR spectra of all three compounds (Figures 28-30). Like **1** and **2**, **3** has *o* and *m* aromatic hydrogens which appear as doublets at δ 8.42 and 8.26 ppm respectively, and a

methyl singlet at δ 1.56 ppm. The ^{1}H-NMR data of the substituents on compounds **3a-c** are shown in Table 12.

The quantification of reaction times shown in Table 11 was based on the visual assessment of when the blue enolate had disappeared. The reaction began with the addition of base and resulted in rapid formation of an intense blue color. As the reaction progressed, the blue color faded to green and finally, pale yellow. This indicated that the enolate was no longer present. To determine percent yield an unreactive compound, 2-methoxynaphthalene was added to the reaction mixture in an equimolar quantity equal to **2** after the blue color had completely faded. This mixture was worked up normally and analyzed by NMR spectroscopy. By comparing the peak area of the 2-methoxynaphthalene methyl group (δ 3.9) to the methyl group of the substituted oxazalone (δ 1.67) the percent yield of the reaction could be determined (Figure 31).

Alkylation of 1 is dependant on the reactivity of the alkyl halide used. Table 11 shows the alkyl halides used that successfully participated in the reaction. Other, less reactive, alkyl halides used were crotyl chloride, iodomethane, and bromopropane; these did not successfully alkylate **2a**. The major distinction between successful and unsuccessful alkylations was not reaction time. In experiments that employed method A or B the time taken for the blue enolate to fade fell in the range reported in Table 11 (6 to 17 min for method A, 1 to 3 hr for B) regardless of which alkyl halide was used (data not shown). The major observation that distinguished failed reactions from successful ones was the coloration of the separated phases in the phase transfer experiment. Work up of

the phase transfer reaction mixture involved separating the aqueous and organic phases. The observation was made that a "successful" reaction was followed by a yellow organic phase and a clear aqueous phase during workup. Conversely, a "failed" reaction always had a clear organic phase and a yellow aqueous phase. This inspired the hypothesis that during the reaction or workup the newly synthesized **3** spontaneously hydrolyzed to **4**, which was soluble in water. Unfortunately, analysis of aqueous phases in the above "failed" reactions did not detect **4**.

Hydrolysis of substituted 1 gave an amide. Once it was established by NMR that the reaction product was the alkylated oxazalone, it was investigated whether **3** could be hydrolyzed back to an amide analogous to N-(4-nitrobenzoyl)-DL alanine. This was attempted by taking a sample of **3**, dissolving it in d-6 acetone and adding DCl to the sample. To show the difference in **3** before and after the addition of DCl, consecutive ^{1}H NMR spectrum were taken immediately before and after DCl addition. The results of hydrolysis reactions performed on **3a**, **3b**, and **3c**, to **4a**, **4b**, and **4c** are respectfully shown in Figures 32-34. These spectra are very similar to the corresponding spectra taken of **3a**, **3b**, and **3c** (Figure 28-30). However, a structural change is evident by the significant change of chemical shift in the diastereotopic hydrogens.

Breaking the N-bond of compounds **4a-c** to form di-substituted amino acids were not attempted in this study. However, in a similar study conducted by John Keller, breaking the N-bond of a propargyl substituted N-*p*-chlorobenzoyl-alanine was ventured.[8] This compound was refluxed in an acetic acid/hydrochloric acid/water solution and the

subsequent product was analyzed by nuclear magnetic resonance spectroscopy. The results of the reaction showed that 75% of the benzoyl-free, di-substituted amino acid had its propargyl group hydrolyzed to a ketone, presumably from the harsh reaction conditions. It is unknown how this approach would affect the compounds synthesized in this study, but seems likely that the propargyl substituted *p*-nitro analog would be hydrolyzed in the same fashion.

Conclusions. This research provides groundwork for three further manipulations of these compounds. These include: (1) the participation of other electrophilic compounds in S_N2 type reactions with the oxazolone **2.** For this research only six alkyl halides were used in alkylation experiments with the oxazolone **2**, three of which were unsuccessful. Other alkyl halides could be used to establish the reactive or steric requirements needed of the attacking chemical species. (2) Second, different *p*-substituted phenyl oxazolones could be used in alkylation reactions. The corresponding oxazolone enolate used in this study, **2a**, gave a distinctive blue color, perhaps the synthesis and use of other *p*-substituted phenyl oxazolones would provide different colored enolates that could also be used as reaction indicators. (3) Lastly, an attept should be made to breakdown **4** to a dialkylglycine. Once hydrolysis of the substituted oxazolones **3** to **4** is completed, a method for breaking the amide N-bond between the *p*-nitrobenzoyl substituent and the disubstituted glycine may be investigated. After separated and purified, the novel synthesized dialkylglycine may be used in assays with DgdR as an alternate coinducer. It

could also be used as an alternative substrate or inhibitor for the dialkylglycine decarboxylase enzyme DgdA.

Works Cited

1. Brewer, D., Mason, F. G., and Taylor, A. (1987) *Can. J. Microbiol. 33*, 619-25.

2. Obrecht, D., Spiegler, C., Schönholzer, P., and Müller, K. (1992) *Helv. Chim. Acta 75*, 1666-95.

3. Allen-Daley, E., Sun, H., Bray-Hallan, S. T., and Keller, J. W. (2002). Manuscript in preparation.

4. Chenault, H. K., Dahmer, J., and Whitesides, G. M. (1989) *J. Am. Chem. Soc. 111*, 6354-64.

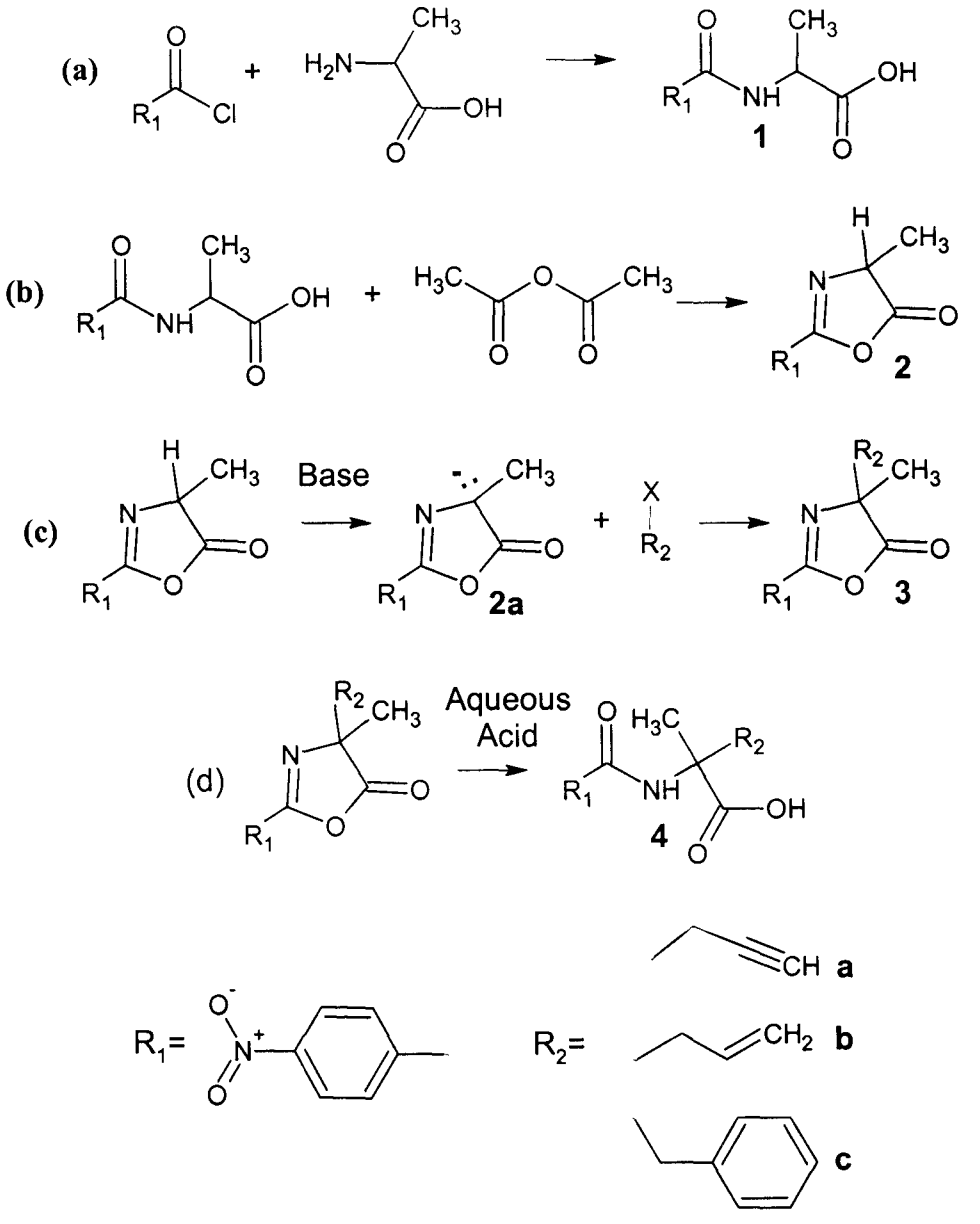
5. Benvenuti, S., Severi, F., Costantino, L., Vampa, G., and Melegari, M. (1998) *Il. Farmaco. 53*, 439-42.

6. Gelmi, M. L., Pocar, D., and Rossi, L. M. (1984) *Synthesis*, 763-5.

7. O'Brien, J. L., and Niemann, C. (1957) *J. Am. Chem. Soc. 79*, 80-5.

8. Keller, J. W. (2004). Personal Communication.

Scheme 4.



Scheme 4. Synthesis of N-*p*-nitrobenzoyl disubstituted glycines. (a) The amidization of alanine with p-nitrophenyl benzoyl chloride to form N-p-nitrobenzoyl-DL-alanine **1**. (b) The dehydrating cyclization of **1** using acetic anhydride to from 4-methyl-2-(4-nitrophenyl) 5(4H)-oxazolone **2**. (c) The alkylation of **2** at the 4 position with various alkyl halides to make **3**, and (d) the ring opening of **3** with acid to form a N-p-nitrobenzoyl disubstituted glycine 4.

Figure 25.

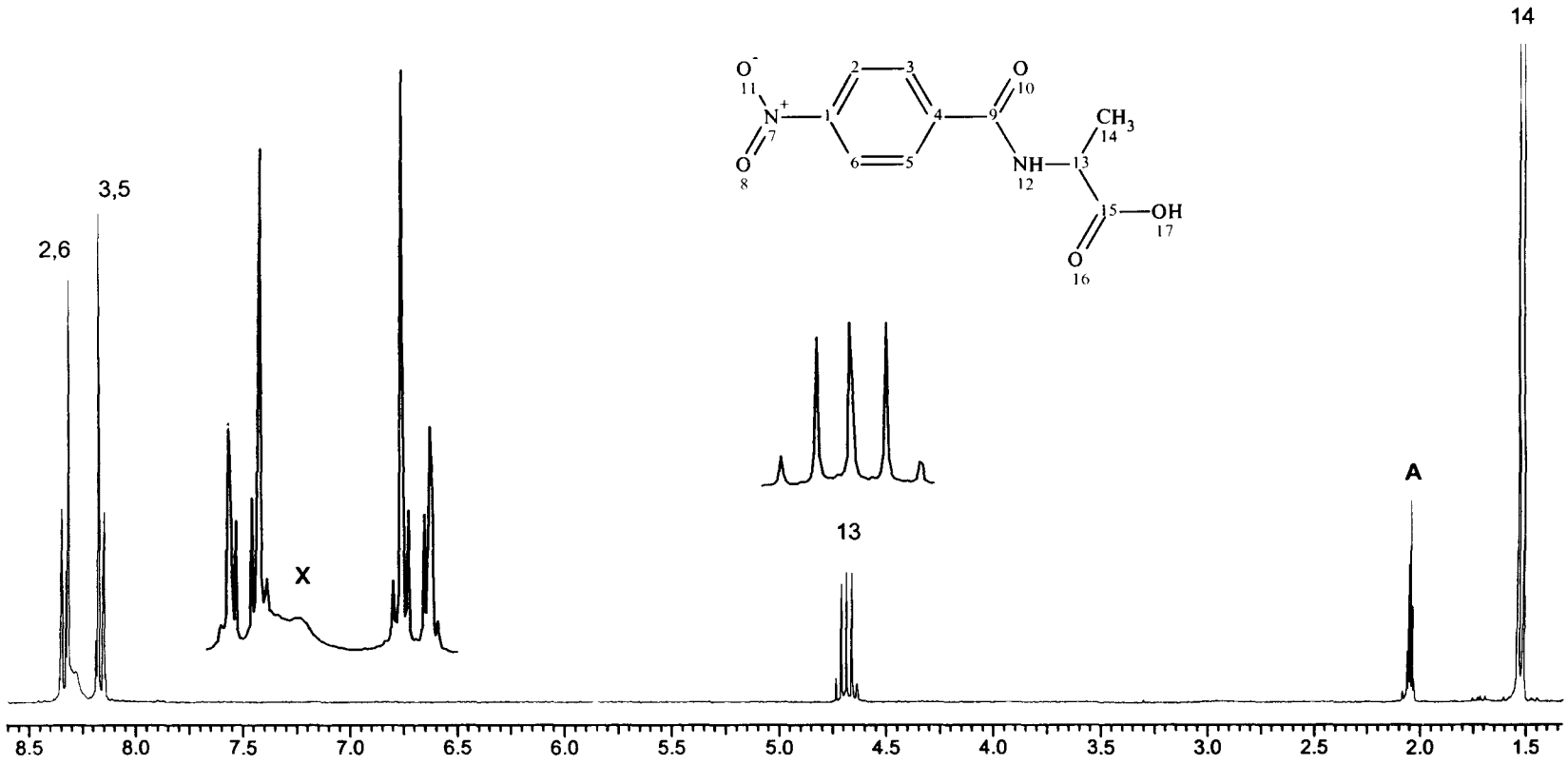


Figure 25. ^{1}H NMR spectrum of N-*p*-nitrobenzoyl-DL-alanine. The **X** marks the peak corresponding to the amido hydrogen, **A** marks the solvent d6-acetone.

Figure 26.

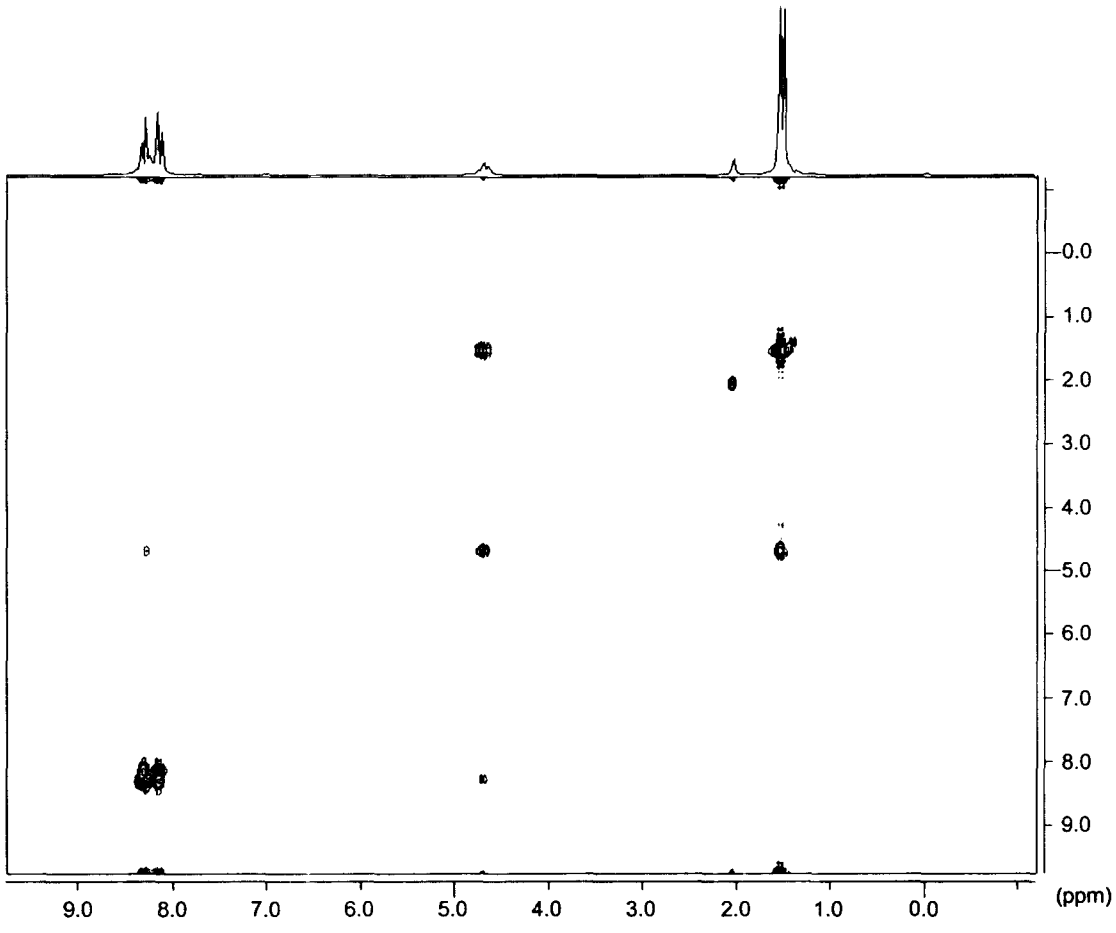


Figure 26. COSY spectrum of N-*p*-nitrobenzoyl-DL-alanine. This shows coupling between the α hydrogen pentet and the amido-hydrogen peak partially obscured by the aromatic doublets.

Figure 27.

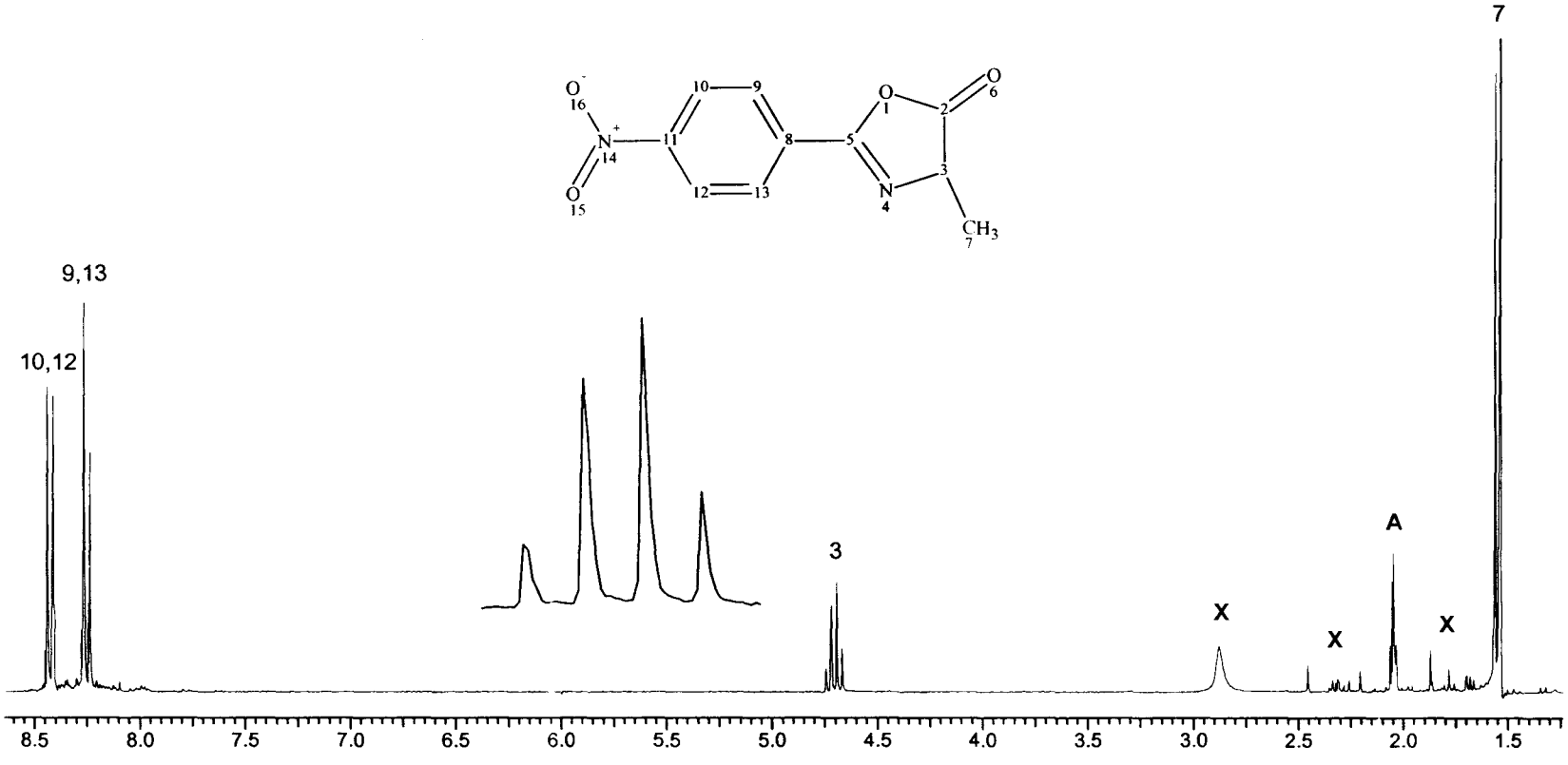


Figure 27. A ^{1}H NMR spectrum of 4-methyl-2-(4-nitrophenyl) 5(4H)-oxazolone. Xs mark impurities, A marks the solvent d6-acetone. The methyl singlet has been truncated to allow detail.

Table 11. Reaction times and yields of oxazolone substitution reactions.

Compound	R	X	Method	Reaction Time	% Yield[a]
3a	$CH{\equiv}C{-}CH_2{-}$	Br	A	17min	75
3b	$CH_2{=}CH{-}CH_2{-}$	Br	A	10min	80
3c	$C_6H_5{-}CH_2{-}$	Br	A	6min	92
3a	$CH{\equiv}C{-}CH_2{-}$	Br	B	73min	68
3b	$CH_2{=}CH{-}CH_2{-}$	Br	B	2 hours	67
3c	$C_6H_5{-}CH_2{-}$	Br	B	3 hours	72

[a]Percent yields are based on NMR integrals

Figure 28.

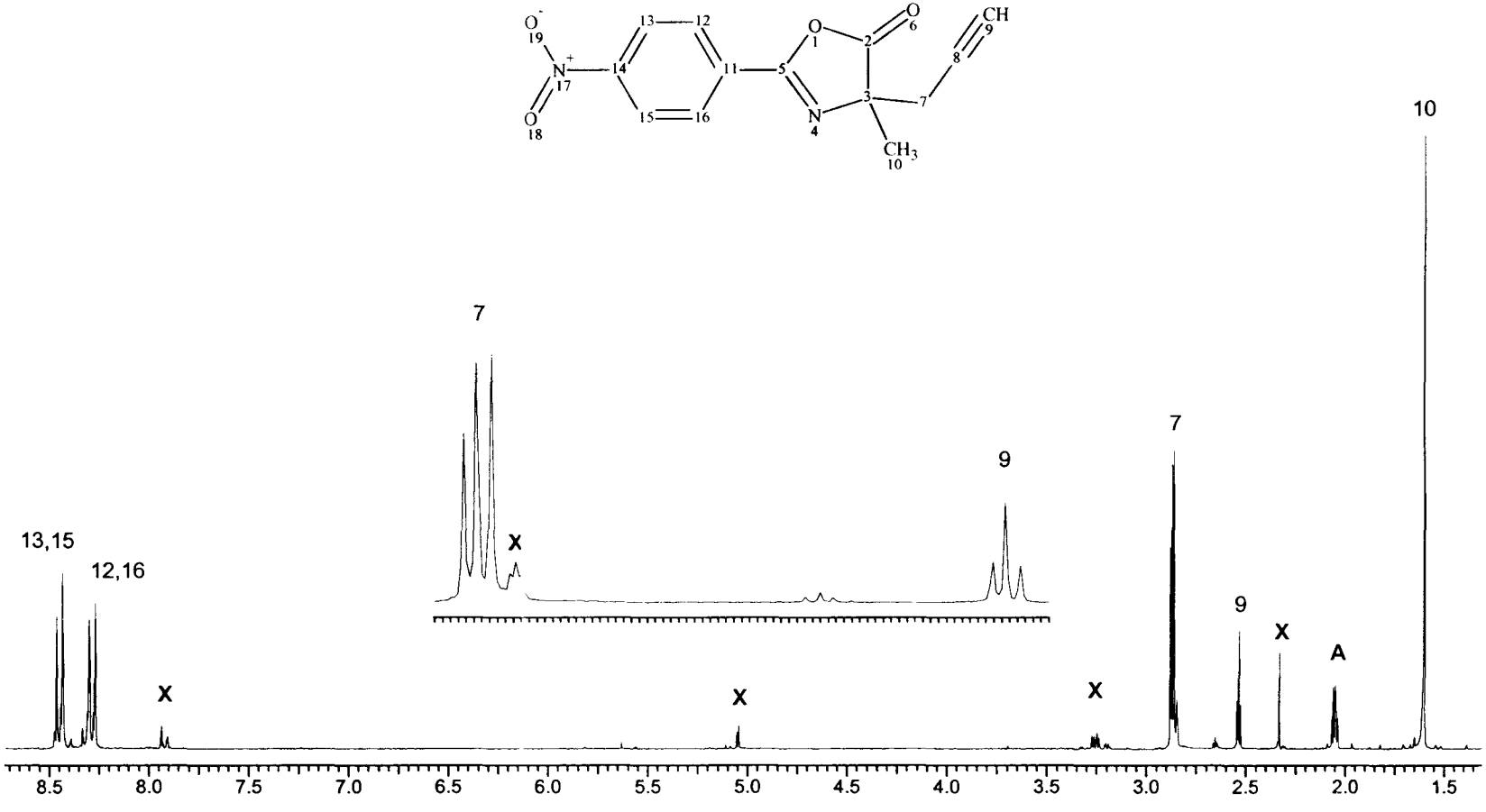


Figure 28. ^{1}H NMR spectrum of 4-propargyl-4-methyl-2-(4-nitrophenyl) 5-oxazolone. **X**s mark impurities, **A** marks the solvent d6-acetone. The methyl singlet has been truncated to allow detail.

Figure 29.

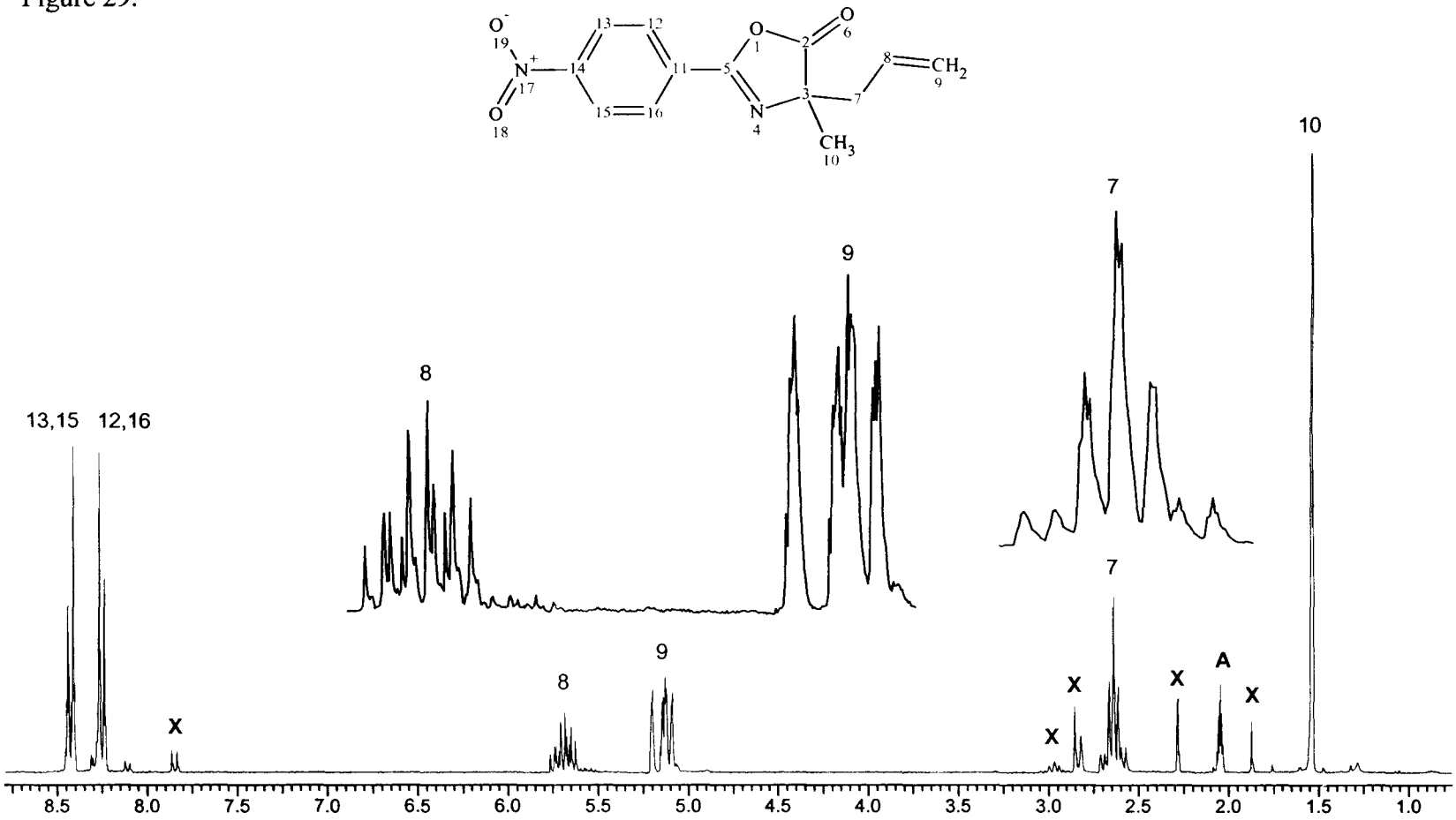


Figure 29. ^{1}H NMR spectrum of 4-allyl-4-methyl-2-(4-nitrophenyl) 5-oxazolone. Xs mark impurities, A marks the solvent d_6-acetone. The methyl singlet has been truncated to allow detail.

Figure 30.

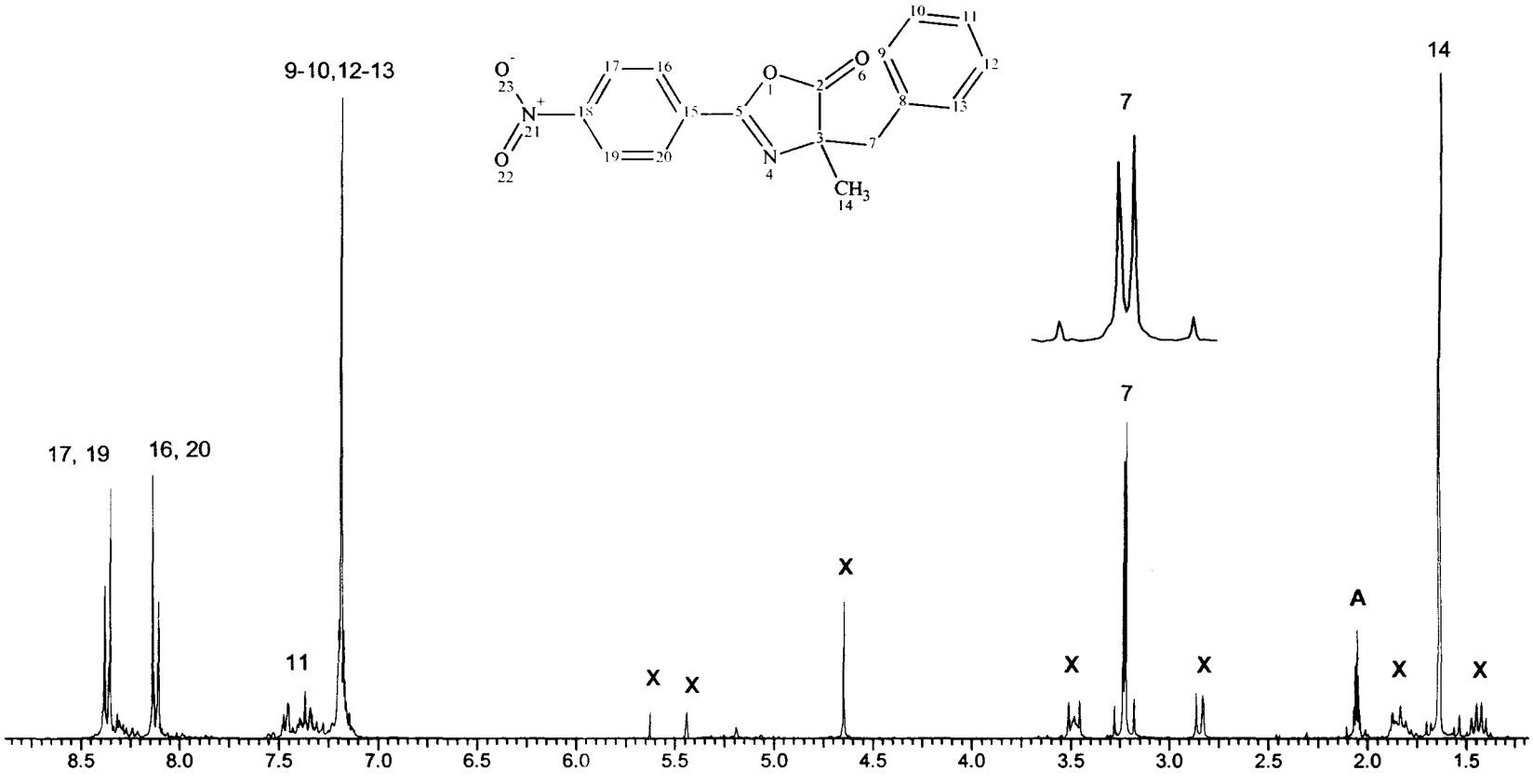


Figure 30. ^{1}H NMR spectrum of 4-benzyl-4-methyl-2-(4-nitrophenyl) 5-oxazolone. Xs mark impurities, A marks the solvent d6-acetone. The methyl singlet has been truncated to allow detail.

Table 12. NMR and melting point of synthesized compounds.

Compound	R	Melt Point °C	Non-aromatic ^{1}H-N.M.R. (d_6-acetone) δ [ppm]
3a	$CH{\equiv}C{-}CH_2{-}$	132-134	2.56 (t, 1H); 2.89 (d, 2H)
3b	$CH_2{=}CH{-}CH_2{-}$	132-134	2.65 (m, 2H), 5.15 (m, 2H), 5.70 (m, 1H)
3c	$C_6H_5{-}CH_2{-}$	oil	3.23 (q, 2H), 7.19 (m, 4H), 7.37 (m, 1H)

Figure 31.

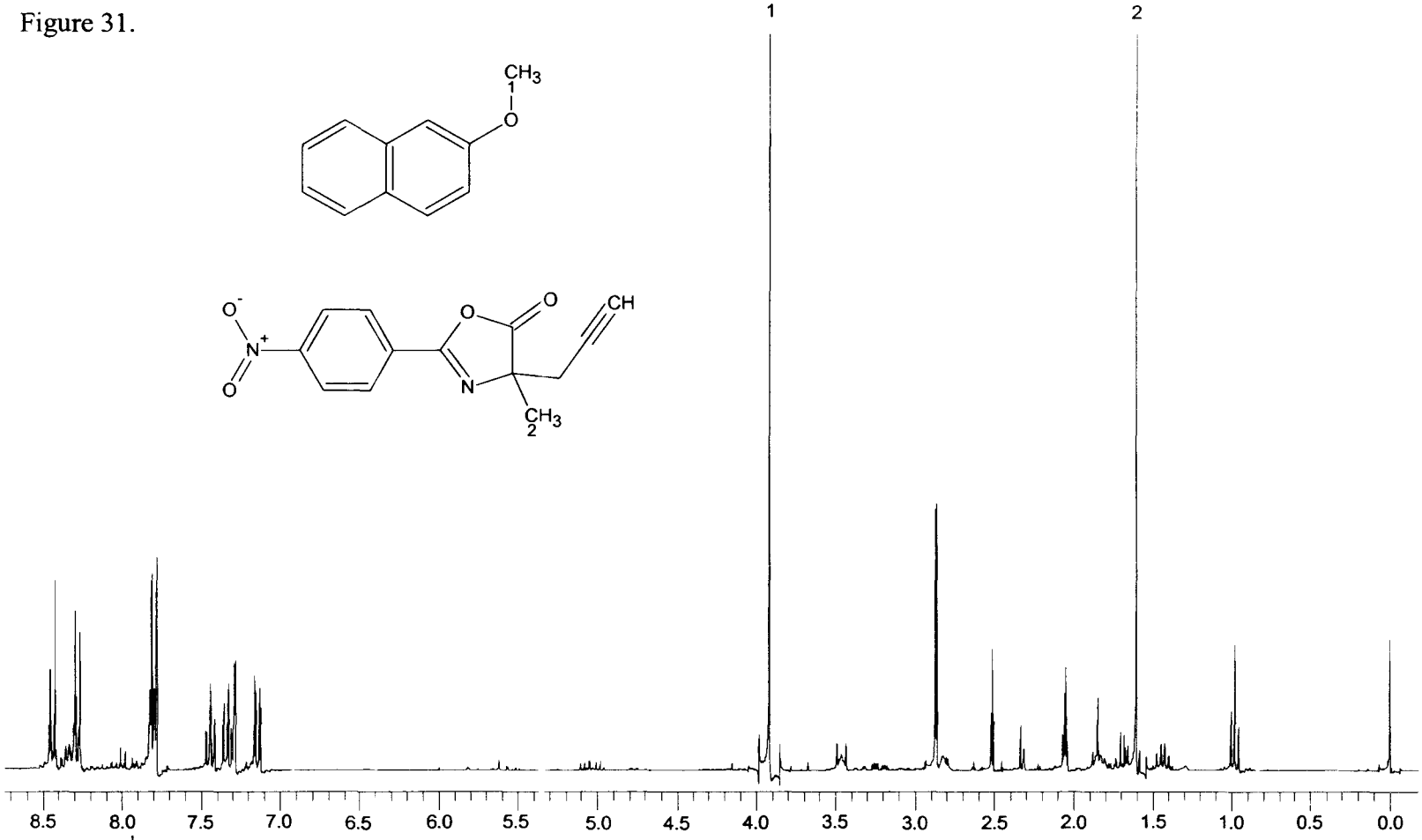


Figure 31. ^{1}H NMR spectrum of 4-propargyl-4-methyl-2-(4-nitrophenyl) 5-oxazolone **(3a)** and methoxynaphthalene. Because methoxynaphthalene and **2** were added in equimolar amounts, their methyl peaks in a ^{1}H NMR spectrum can be compared to determine percent yield. In this spectrum, the methoxynaphthalene methyl, 1, integrates to 1.0, where the **3a** methyl, 2, integrates to 0.6. Note: The tops of these peaks were truncated in order to show detail of the entire spectrum.

Figure 32.

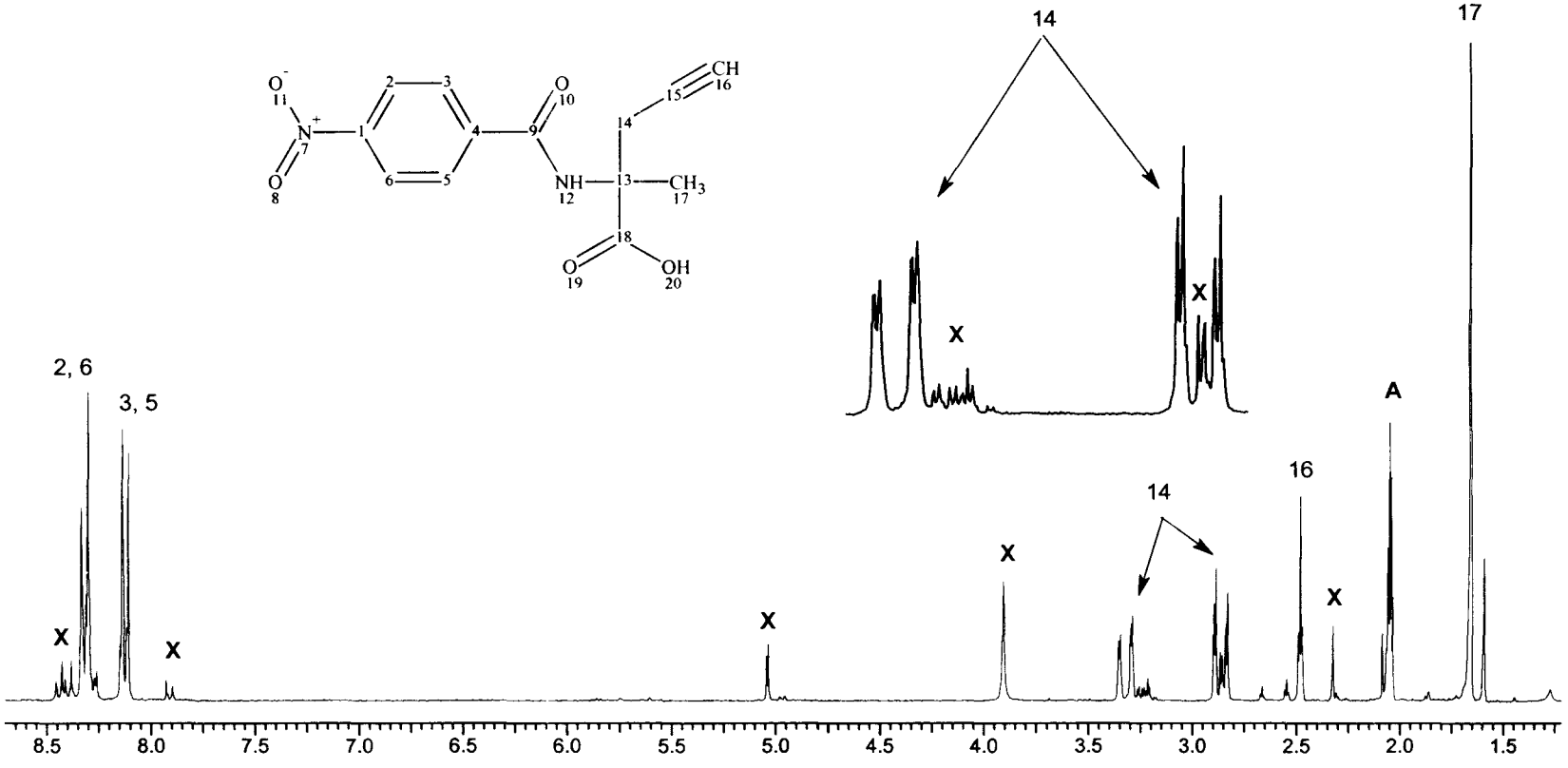


Figure 32. ^{1}H NMR spectrum of N-*p*-nitrobenzoyl-2-methyl-2-propargyl-DL-glycine. Xs mark impurities, A marks the solvent d6-acetone. The methyl singlet has been truncated to allow detail.

Figure 33.

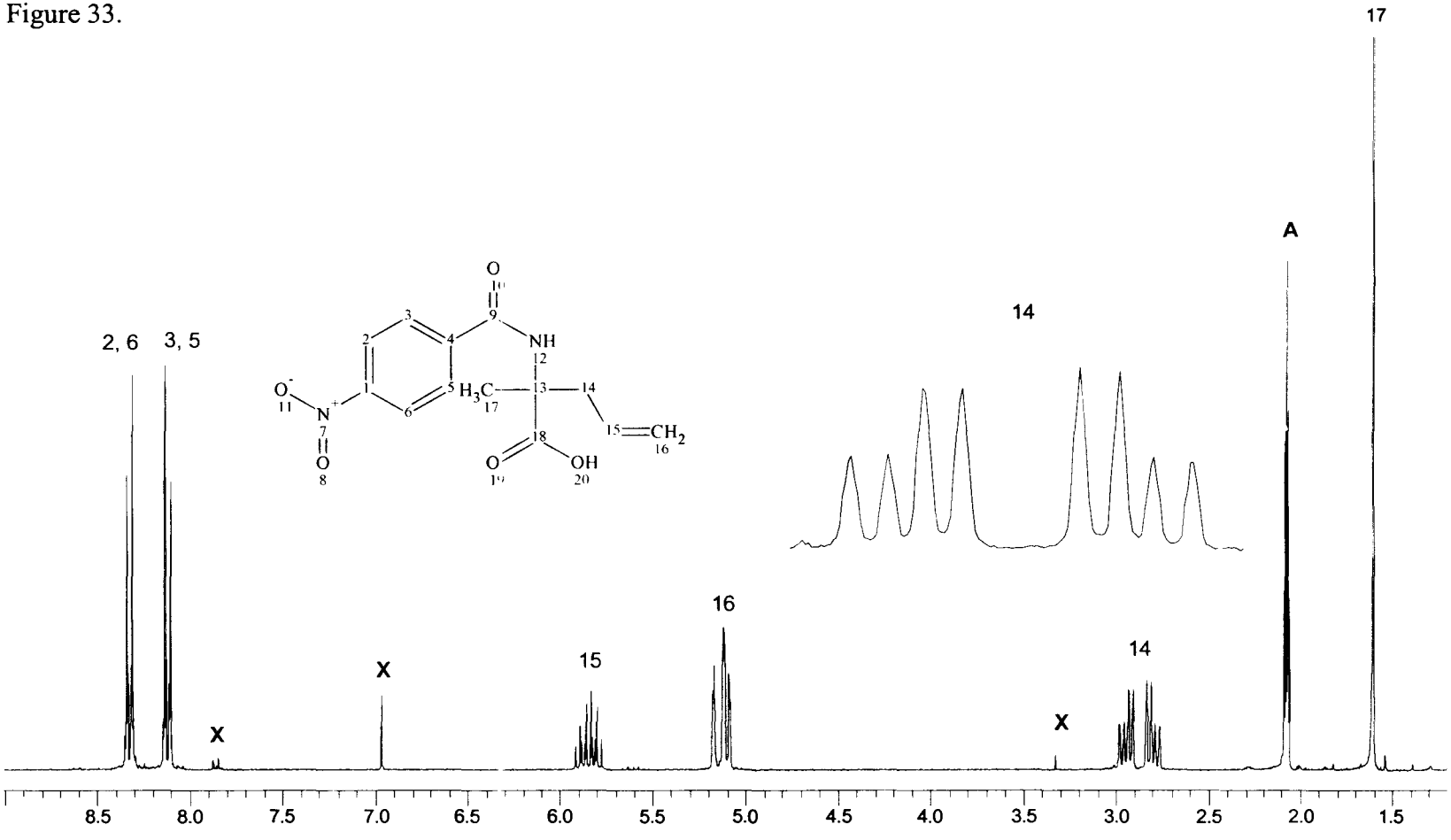


Figure 33. ^{1}H NMR spectrum of N-*p*-nitrobenzoyl-2-methyl-2-allyl-DL-glycine. Xs mark impurities, A marks the solvent d6-acetone. The methyl singlet has been truncated to allow detail.

Figure 34.

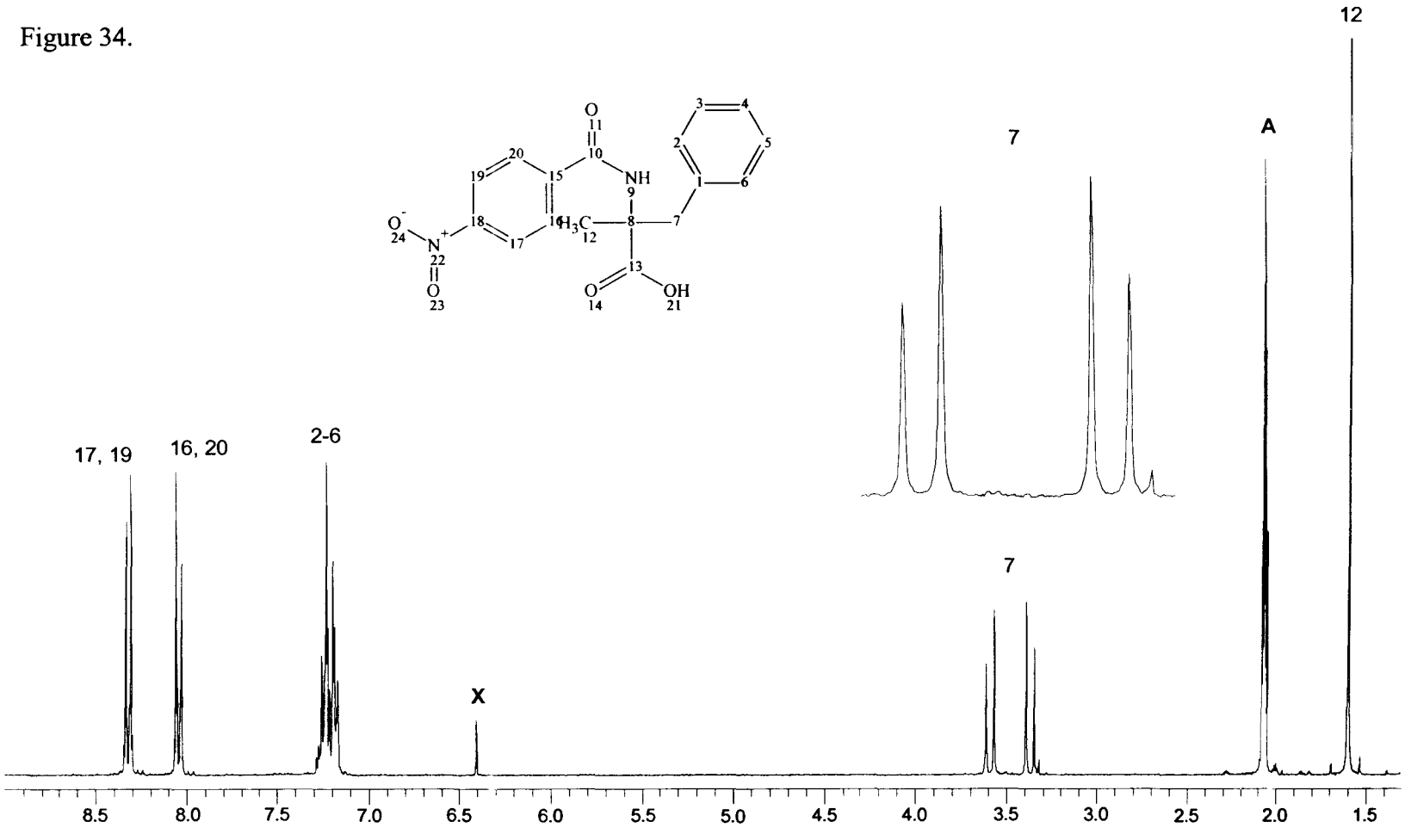


Figure 34. A ^{1}H NMR spectrum of N-*p*-nitrobenzoyl-2-methyl-2-benzyl-DL-glycine. Xs mark impurities, **A** marks the solvent d6-acetone. The methyl singlet has been truncated to allow detail.