Catalysis in the nitrilase superfamily
Charles Brenner

Recently, we defined the nitrilase superfamily as consisting of 12 families of amidases, N-acyltransferases and presumptive amidases, in addition to the family of plant and bacterial nitrilases for which the superfamily was named. A novel Glu-Lys-Cys catalytic triad, found at the crystallographically defined Nit active site of worm NitFhit, was postulated to constitute the catalytic residues for all members of the superfamily. Recent experimental results confirm the essentiality of the catalytic triad residues and suggest the biochemical functions of additional branches and sub-branches of the nitrilase superfamily.

Introduction

Condensation and hydrolysis of carbon–nitrogen linkages are central to biology. Crystal structures of ribosomes [1] and nonribosomal peptide synthetases [2] have provided unprecedented insights into peptide bond formation, whereas the study of proteolysis is supported by many thousands of biochemical, structural and genetic investigations [3]. Although peptide bonds are the most obvious class of carbon–nitrogen bonds that are formed and cleaved, there is a menagerie of other carbon–nitrogen linkages whose metabolism is less well understood. In particular, a range of nitriles (R-C≡N), acid amides [R-C(=O)-NH₂], secondary amides [R-C(=O)-NH-R'], and N-carbamyl amides [R-NH-C(=O)-NH₂] are hydrolytic substrates of nine branches of the nitrilase superfamily, whereas N-acylation of polypeptide N termini is mediated by members of another branch of the superfamily.

Not setting out to classify nitrilases, amidases and N-acyltransferases, but rather to identify invertebrate homologs of the human tumor suppressor protein Fhit [4], my colleagues and I found that flies and worms encode their Fhit homolog [4] as a natural fusion protein with a member of the nitrilase superfamily [5]. Although the `Nit' domain of fly and worm NitFhit was easily recognizable by BLAST [6] as related to nitriase, it was not initially clear whether Nit orthologs would be found in vertebrates and fungi, which were the branches of eukarya known to encode Fhit homologs. By comprehensive examination of gene fusions in sequenced genomes, it was argued that proteins fused in some forms of life that are separate in others represent `Rosetta stones', which decode previously hidden cellular or biochemical interactions between proteins unrelated by sequence [7–9]. Biological significance of such fusion events can be assigned higher confidence with evidence that the genes are coordinately expressed in organisms in which they are not fused and that orthologs of the fused genes are both present or both absent in other genomes [9]. Remarkably, after cloning human and murine NIt1 genes, which are orthologs of the Nit domain of invertebrate NitFhit fusion proteins, the expression patterns of murine nit1 and murine nit1 were shown to be nearly identical [5]. Identification of homologs of the Nit domain of invertebrate NitFhit in the same eukaryotic phyla [10**] as those containing Fhit homologs [4] strengthened the case that Nit and Fhit function in the same biological pathway. The crystal structure of NitFhit showed that a Nit tetramer binds two Fhit dimers and that the C-terminal strand conserved in Nit homologs mediates physical interactions with Fhit [10**]. Although substrates for Nit have not been identified, the crystal structure of worm NitFhit [10**] and the crystal structure of an N-carbamyl-D-amino acid amidohydrolase from Agrobacterium [11**] demonstrated that a novel Glu-Lys-Cys catalytic triad (Figure 1) is positioned by a novel α-β-β-α sandwich protein fold (Figure 2) to mediate catalysis in the nitrilase superfamily.

Troubled by database annotations that are inconsistent and inaccurate, we examined all of the 176 available nonredundant protein sequences, as of November 2000, that belong to the nitrilase superfamily and classified them as members of 13 different branches, most of which have one or more member of known substrate specificity and function [12**]; see also the online supplement at http://genomebiology.com/content/supplementary/gb-2001-2-1-reviews0001-S1.htm).

This analysis revealed several salient points. Although essentially every member of the superfamily has discernible similarity to every other member, there are at least 12 prototypical sequences that 'pull out' smaller, mutually exclusive subsets of sequences with higher significance [12**]. Second, only one branch of the nitrilase superfamily contains members that have nitrile substrates, whereas the majority of branches of the nitrilase superfamily consist of amidases with distinct specificities [12**]. Third, although no member of the nitrilase superfamily has been shown to possess peptidase activity, several branches consist of enzymes with amide hydrolysis or amide condensation activities that act on protein sidechains or protein N termini [12**]. Fourth, protein fusion events have been common in the nitrilase superfamily and these fusions do assist in functional classification [12**]. Unfortunately, sequence database annotations for the 176 members of the nitrilase superfamily are as incorrect today as they were at the time of publication and have been joined by additional...
sequences that were incorrectly classified as nitrilases, amidases, β-ureidopropionases and so on without the benefit of the 13-branch classification [12••]. Here, I provide an update on structural and enzymatic data that pertain to all members of the nitrilase superfamily, and review new information about several branches of the superfamily.

Structure and enzymology

Independently, the structures of the branch 10 enzyme worm NitFhit [10••] and the branch 6 enzyme Agrobacterium N-carbamyl-D-amino acid amidohydrolase [11••] indicated that these enzymes consist of a tetramer of compact α-β-α sandwiches in which pairs of α helices form the bread and two sheets of six β strands form the meat layers (Figure 3). One twofold rotation axis stacks two α-β-α sandwiches on top of each other. The other twofold axis doubles the width of the entire α-β-α sandwiches such that each layer of bread and meat is more extensive. An additional Agrobacterial N-carbamyl-D-amino acid amidohydrolase crystal structure [13] is essentially identical to the first one [11••]. However, a second member of the branch 10 Nit proteins, the yeast Nit3 protein, solved by the New York Structural Genomics Consortium (http://www.rcsb.org/pdb/ PDB code 1F89), is dimeric, conserving the left-right twofold in Figure 3 that creates an α-β-α-α-β-α structure. Although other proteins, such as DNase I [14], have grossly similar α-β-α architecture, they are topologically distinct and are not related by descent.

Although no structure of a nitrilase superfamily member is yet available bound to a substrate analog, our group [10••] and that of Nakai [11••] concluded that the entire superfamily is likely to utilize a novel catalytic triad consisting of glutamic acid, lysine and cysteine. Historically, characterization of a nitrilase from Rhodococcus had indicated that it is a thiol enzyme that proceeds through covalent intermediates [15] and mutagenesis had identified the cysteine nucleophile [16]. Sequence gazing suggested [17,18] and mutagenesis confirmed [19] the essentiality of the corresponding cysteine residue in other branches of what we now consider to be the nitrilase superfamily [12••]. More recently, having threaded the aliphatic amidase from Pseudomonas to the structure of the Nit domain of NitFhit, Brown and co-workers [20•] tested the hypothesis [10••,11••] that glutamic acid and lysine residues are catalytically essential. Although Glu59Gln and Lys134Arg mutants were inactive, these substitutions appeared to affect protein stability [20•]. Nonetheless, the catalytic essentiality of glutamic acid and lysine were demonstrated by construction of stable, inactive mutants containing Glu59Gln and Lys134Asn substitutions [20•].

Earlier, we reviewed the catalytic mechanism of members of the nitrilase superfamily, and the geometrically interesting observation that enzymes in this superfamily attack either the cyano carbon of a linear nitrile substrate or the planar carbonyl carbon of an amide substrate and proceed through one or two tetrahedral intermediates toward products [12••]. It bears emphasis that, although members of the nitrilase superfamily have substrates in common with nitrile hydratases [21], signature amidases [22], N-terminal nucleophile hydrolases and triad amidotransferases [23], these enzymes are structurally and mechanistically unrelated. Substrates, products, prototypical gene names,
accession numbers and structures are summarized for 12 of the 13 branches of the nitrilase superfamily in Table 1. The thirteenth branch of the superfamily remains a group of outlier sequences with no prototype, Rosetta stone fusion or any characterization [12••].

**Branch 1 nitrilases: insights into specificity**

Branch 1 of the superfamily is the only branch to contain members that act primarily on nitriles. Although it was found that these enzymes convert indole-3-acetonitrile acid to the auxin indole-3-acetic acid [24], it has now been shown that the *Arabidopsis* enzymes Nit1, Nit2 and Nit3 strongly prefer 3-phenylpropionitrile [25••,26], whose product, phenylacetic acid, is found in nasturtium and has auxin activity [27]. Analysis of the *Arabidopsis NIT3* promoter indicates that its mRNA accumulates in response to sulfur starvation, probably via accumulation of O-acetylserine [28]. Because Nit3 nitrilase activity produces auxins that stimulate root development, these data suggest that sulfur starvation initiates a regulatory loop that allows roots to reach fresh supplies of sulfur [28].

Piotrowski, Weiler and colleagues, having investigated the substrate specificity of *Arabidopsis* enzymes Nit1, Nit2 and Nit3 [25••], also purified Nit4 from *Arabidopsis* and *Nicotiana,* and reported that Nit4 constitutes a distinct sub-branch of nitrilases that hydrolyze β-cyano-L-alanine to a –60:40 mixture of asparagine and aspartic acid [29••]. In deference to their reservation of the plant gene name *NIT4* for cyanoalanine hydratase/nitrilase orthologs, we designate this sub-branch ‘1.4’ in Table 1. Substantial production of asparagine from cyanoalanine [29••] was surprising in light of the classical literature on nitrilase from *Rhodococcus,* which indicated that this enzyme neither reacts with nor produces a significant amount of the corresponding amide [15,30]. Lack of acid amide production is due to the covalent mechanism: the enzyme does not add water to a nitrile, but rather forms an enzyme-linked thioimidate intermediate [15]. Because ammonia is a better leaving group than the enzyme [12••], the first enzyme-dependent water addition does not produce an acid amide but rather the acylenzyme, hydrolysis of which produces the acid product [15]. Cyanoalanine is produced by enzymatic reaction of cyanide with cysteine [31] and thus Nit4 apparently functions in a detoxification pathway [29••]. Because the auxin-producing nitrilases specifically to produce the acid product, whereas Nit4 functions to consume cyanoalanine, Nit4 may have lost selective pressure to perform the second water addition. Partitioning of a water-derived proton to the enzyme rather than to the ammonia leaving group would produce asparagine (cyanoalanine hydratase activity) rather than aspartic acid (cyanoalanine nitrilase activity).

The potential of nitrilases to convert cyanohydrins [R-CH(OH)-C≡N] that contain a racemic mixture of *R* and *S* isomers specifically to a single isomer of the corresponding α-hydroxy carboxylic acid [R-CH(OH)-CO₂⁻] has been exploited [32]. For the nitrilases that are stereospecific for a cyanohydrin stereoisomer, a high yield of the acid is obtained because, under the conditions of the assay, the isomer that is not a substrate undergoes racemization [33]. By cloning nitrilase genes from uncultured environmental samples and selecting for novel enzymatic activities [34], the Diversa group reports that more than 200 new nitrilases have been characterized — some specific for either of certain enantiomeric substrates — and has published substrate specificity for three novel sequences [35••]. We BLASTed [6] these sequences against the publicly available nonredundant protein databases on August 9, 2002 and found no homolog that is more than 57% identical to the newly reported nitrilases. Because substantially similar
<table>
<thead>
<tr>
<th>Branch number</th>
<th>Enzyme name</th>
<th>Prototypical gene</th>
<th>NCBI accession code</th>
<th>PDB code</th>
<th>Substrates</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Auxin-producing nitrilase</td>
<td>Arabidopsis NIT1</td>
<td>P32961</td>
<td>[25]</td>
<td>Phenylpropionitrile + 2H₂O</td>
<td>Phenylacetic acid + NH₄⁺</td>
</tr>
<tr>
<td>1.2</td>
<td>Aliphatic nitrilase</td>
<td>Rhodococcus nitA</td>
<td>Q03217</td>
<td>[16]</td>
<td>Acrylonitrile + 2H₂O</td>
<td>Acrylic acid + NH₄⁺</td>
</tr>
<tr>
<td>1.4</td>
<td>β-cyano-alanine hydratase/nitrilase</td>
<td>Arabidopsis NIT4</td>
<td>P46011</td>
<td>[29]</td>
<td>β-cyano-alanine + H₂O</td>
<td>Asparagine + NH₄⁺</td>
</tr>
<tr>
<td>2</td>
<td>Aliphatic amidase</td>
<td>Pseudomonas amiE</td>
<td>P11436</td>
<td>[20]</td>
<td>Asparagine</td>
<td>Aspartic acid + NH₄⁺</td>
</tr>
<tr>
<td>3</td>
<td>N-terminal amidase</td>
<td>Saccharomyces NTA1</td>
<td>NP_012596</td>
<td>[36]</td>
<td>N-terminal asparagine + H₂O</td>
<td>N-terminal aspartic acid + NH₄⁺</td>
</tr>
<tr>
<td>4.1</td>
<td>Biotinidase</td>
<td>human BTD</td>
<td>P43251</td>
<td>[37]</td>
<td>Biocytin + H₂O</td>
<td>Biotin + lysine</td>
</tr>
<tr>
<td>4.2</td>
<td>Pantetheinase/vanin</td>
<td>mouse vnn-1</td>
<td>NP_035834</td>
<td>[40,42]</td>
<td>Pantethine + H₂O</td>
<td>Pantothenic acid + cysteamine</td>
</tr>
<tr>
<td>5</td>
<td>β-Ureidopropionase</td>
<td>rat BAS</td>
<td>Q02348</td>
<td>[44]</td>
<td>N-carbamylβ-alanine + 2H₂O</td>
<td>β-alanine + CO₂ + NH₄⁺</td>
</tr>
</tbody>
</table>
molecules do not populate public databases, it is not yet possible to characterize additional sub-branches of nitrilases in the manner of Nit4 orthologs [29••]. These observations suggest that the public sequences are a tiny fraction of what exists in nature and that, in time, it will be possible to sort nitrilase sequences into a large number of specificity groups.

### Branches 2 and 3: aliphatic amidases and proof of Glu-Lys-Cys

The prototypical branch 2 enzyme, the aliphatic amidase from *Pseudomonas*, now has a predicted three-dimensional structure [20•] based on the structure of the Nit domain of NitFhit [10••]. The major difference between branch 2 amidases and branch 3 amidases is that branch 2 enzymes hydrolyze small-molecule substrates, such as the carboxamide sidechain of an asparagine, whereas branch 3 enzymes are specific for conversion of an asparagine or glutamine residue to the corresponding acid residue, when that residue is at the N terminus of a polypeptide [36]. Branch 3 enzymes are thus responsible for converting what Varshavsky has termed tertiary destabilizing residues to secondary destabilizing residues for subsequent ubiquitination and proteolysis [36].

### Table 1 (cont.).

Branches of the nitrilase superfamily.

<table>
<thead>
<tr>
<th>Branch number</th>
<th>Enzyme name</th>
<th>Prototypical gene</th>
<th>NCBI accession code</th>
<th>PDB code</th>
<th>[Key references]</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Carbamylase</td>
<td>Agrobacterium DCase JW0082</td>
<td>1ERZ</td>
<td>[11••]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-carbamyl-D-phenylalanine + 2H₂O</td>
</tr>
<tr>
<td>7 and 8</td>
<td>Glutamine-dependent NAD synthetase Saccharomyces QNS1 7 P71911 8 NP_011941</td>
<td></td>
<td></td>
<td>[12•]</td>
<td>Glutamine + H₂O</td>
</tr>
<tr>
<td>9</td>
<td>Apolipoprotein</td>
<td>N-acetyltransferase Pseudomonas cutE Q9ZI86</td>
<td>[12•,29•]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-terminal DAG-modified cysteine + palmitate</td>
</tr>
<tr>
<td>10</td>
<td>Nit</td>
<td>humanNIT1</td>
<td>NP_005591</td>
<td>1EMS</td>
<td>Unknown</td>
</tr>
<tr>
<td>11</td>
<td>N-carbamyl putrescine amidohydrolase Pseudomonas AgdB GB3608</td>
<td></td>
<td>[12•,50, this review]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-carbamyl putrescine + 2H₂O</td>
</tr>
<tr>
<td>12</td>
<td>Branch 12</td>
<td>Sphingomonas IdhX JC4601</td>
<td></td>
<td>[12•]</td>
<td>Unknown, possible protein substrate</td>
</tr>
</tbody>
</table>
Branch 4: from vitamins to calamari
On the basis of sequence similarity, we classified 18 sequences as members of the biotinidase/vanin branch of the nitrilase superfamily [12••]. Biotinidase is the enzyme responsible for recycling biotin (vitamin H) from biocytin (i.e. lysine-conjugated biotin) and from lysyl biotinylated proteins [37]; biotinidase deficiency in humans has been characterized molecularly [38,39]. Vanins are related extracellular enzymes first characterized as antigens on hematopoietic precursor cells involved in thymic homing and adhesion [40]. Sequence information [41] and data from a mouse mutant prove that vanin-1 encodes pantetheinase, the enzyme that hydrolyzes pantetheine to pantothenic acid (vitamin B₅) plus the antioxidant cysteamine [42•]. Thus, the biotinidases (branch 4.1) and the pantetheinases (branch 4.2) are secondary amidases with similar specificities that are involved in vitamin recycling. Additionally, although the significance is not yet clear, investigators working on a 60 kDa squid photoprotein have sequence data showing that it is a branch 4 enzyme [43].

Branches 5 and 6: β-ureidopropionases and carbamylases
N-carbamyl compounds [R-NH-C(=O)-NH₂] react with the appropriate amidases in a manner that consumes two water molecules to produce the corresponding primary amine (R-NH₂) plus carbon dioxide and ammonia. Branch 5 enzymes, the β-ureidopropionases, function in pyrimidine base catabolism [44], whereas the branch 6 enzymes are primarily known as bacterial N-carbamyl-D-amino acid amidohydrolases, now of known structure [11••,13].

Branches 7 and 8: glutaminase domains of glutamine-dependent NAD synthetases
From the early days of the characterization of nicotinamide adenine dinucleotide (NAD) synthetases, it was observed that eukaryotic NAD synthetases can use glutamine as an ammonia source, whereas some purified bacterial NAD synthetases use ammonia but not glutamine [45]. Although it was clear from the sequence of a Mycobacterial NAD synthetase that a bacterial glutamine-dependent NAD synthetase contained an N-terminal extension [46] with respect to bacterial ammonia-dependent enzymes [47], no one identified a potential glutamine amidotransferase active site until we observed that possession of nitrilase-related domains correlated perfectly with glutamine dependence [12••]. It is interesting to note that, in most other branches of the nitrilase superfamily, function depends on consumption of a nitrile or production of an acid, whereas branch 7 and 8 Rosetta stone proteins apparently function to produce ammonia for the associated NAD synthetase active site. The function of these enzymes is under investigation in our laboratory.

Branch 9: apolipoprotein N-acyltransferases
Although most branches of the nitrilase superfamily consist of amidases, branch 9 is the only branch known to catalyze the reverse amidase (i.e. condensation) reaction in vivo. The reaction consists of palmitoylation of an N-terminal diacylglycerol (DAG)-modified cysteine residue [48,49]. We classified 28 homologous enzymes as nitrilase-related apolipoprotein N-acyltransferases by sequence analysis [12••].

Branch 11: a whiff of putrescine defines one of the four mysterious branches
Beyond the first nine branches, for which we were able to assign or confidently predict substrate specificity, we classified four additional branches [12••]. The branch 10 enzymes are implicated in the function of Rosetta stone relationships [5,10••]. The branch 12 enzymes are implicated in protein post-translational modification by virtue of our observation that they are sometimes fused with the RimI N-terminal acetyltransferases [12••]. Branch 13 remains a bin of nonfused outliers, without a single clear prototype, that is expected to be classified into multiple branches once functional information is available for some of its sequences [12••]. Branch 11 was classified as containing 13 sequences of unknown specificity [12••]. A new publication reported that the product of the Pseudomonas AguB gene is an N-carbamyl putrescine amidohydrolase that is related to β-ureidopropionases [50]. This enzyme is thought to function in the metabolism of arginine into spermidine and succinate [50]. Because the new publication provided neither the sequence nor the accession number of AguB [50], and databases have yet to refer to the AguB gene name, it took some detective work to determine that AguB is a new member of branch 11. These bacterial and plant enzymes, which now number more than 20, probably function as carbamylases for the production of putrescine or related amines. Although we earlier [12••] used a Helicobacterial enzyme (accession number F71901) to define branch 11, we now utilize AguB (accession number G83608).

Conclusions
The nitrilase superfamily is interesting because of the combinatorial manner in which nature has crossed chemical specificity (nitrile versus acid amide attack) with R and R’ specificity, and forward versus reverse enzyme chemistry. Enzymes in the superfamily serve diverse and important roles in biology, including synthesis of signaling molecules, vitamin and coenzyme metabolism, small-molecule detoxification and protein post-translational modification. Those who annotate genomes or wish to understand the function of newly identified members of the nitrilase superfamily should realize that most primary sequence annotations contain incorrectly assigned enzymatic activities and consult expertly curated resources, including Table 1 herein, reference [12••] and its online supplement at http://genomewebiology.com/content/supplementary/gb-2001-2-1-reviews0001-S1.htm as starting points for the prediction and analysis of biochemical and cellular functions.
Acknowledgements

Work was supported by National Cancer Institute grant P01CA77738.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest


This paper confirmed the essentiality of glutamic acid and lysine (10*,11**) in the catalytic mechanism of the nitrilase superfamily.


This paper showed that phenylpropionitriple is a better substrate than indoleacetonitrile for the prototypical plant nitrilases.


This paper showed that ATN4 orthologs constitute a new sub-branch of nitrilases with specificity for cyanalanine and that some nitrilases do not exclusively produce an acid product.


   This paper describes a research program to clone nitrilases with novel specificities from uncultured microbes and provides specificity information for three new sequences that are enantioselective for cyanohydrins.
   This paper allows branch 4 of the nitrilase superfamily to be further classified into biotinidases and pantetheinases.