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 specify the biochemical functions of additional branches and sub-branches of the nitrilase superfamily.

Addresses
Structural Biology and Bioinformatics Program, Kimmel Cancer Center, Philadelphia, PA 19107, USA; e-mail: brenner@dada.jci.tju.edu

Current Opinion in Structural Biology 2002, **12**:775–782

0959-440X/02/$ – see front matter

Abbreviations
DAG diacylglycerol
NAD nicotinamide adenine dinucleotide

Introduction

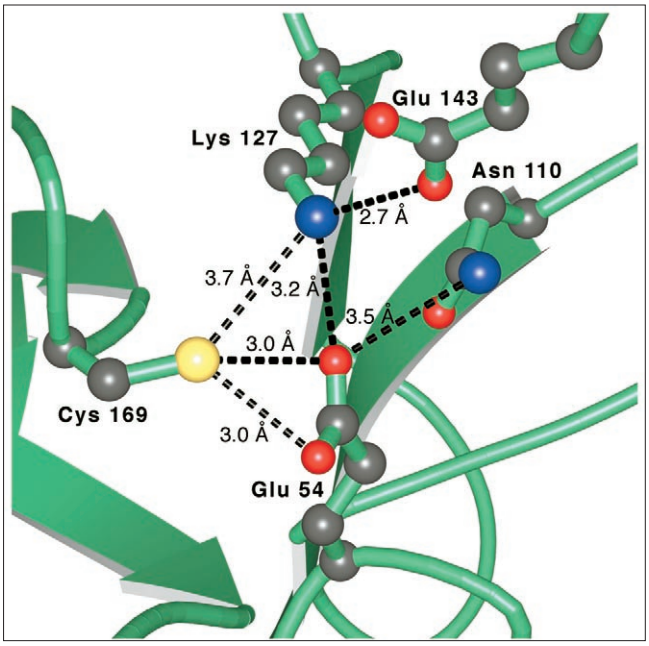
Condensation and hydrolysis of carbon–nitrogen linkages are central to biology. Crystal structures of ribosomes [1] and nonribosomal peptide synthetases [2] have provided unprecedented insights into peptide bond formation, whereas the study of proteolysis is supported by many thousands of biochemical, 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)-NH_2], secondary amides [R-C(=O)-NH-R′], and *N*-carbamyl amides [R-NH-C(=O)-NH_2] 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 nitrilase, 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 *fhit* 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).

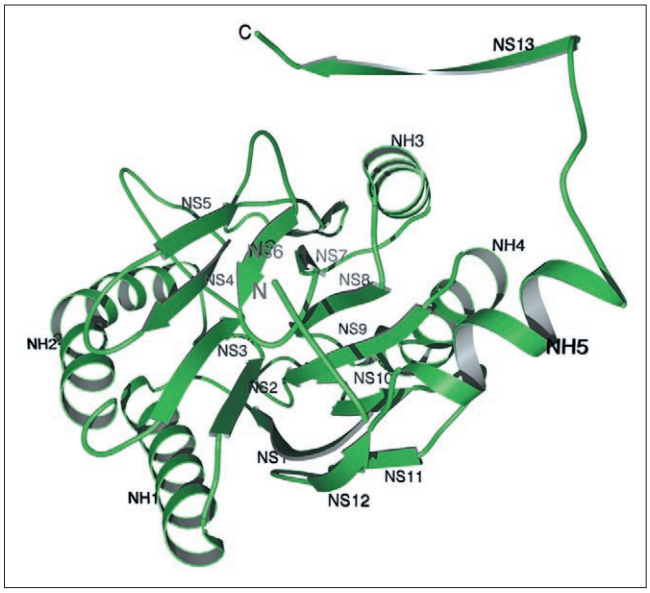
This analysis revealed several salient points. Although essentially every member of the superfamily has discernible similarity to every other member, there are at least 12 prototypical sequences that 'pull out' smaller, mutually exclusive subsets of sequences with higher significance [12••]. Second, 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

Figure 1



Active site of Nit. Reproduced from [10••].

Figure 2



Ribbon drawing of a Nit monomer. Reproduced from [10••].

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-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-β-α sandwiches on top of each other. The other twofold axis doubles the width of the entire α-β-β-α-α-β-β-α sandwich, 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-β-α-α-β-β-α structure. Although other proteins, such as DNase I [14], have grossly similar α-β-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••].

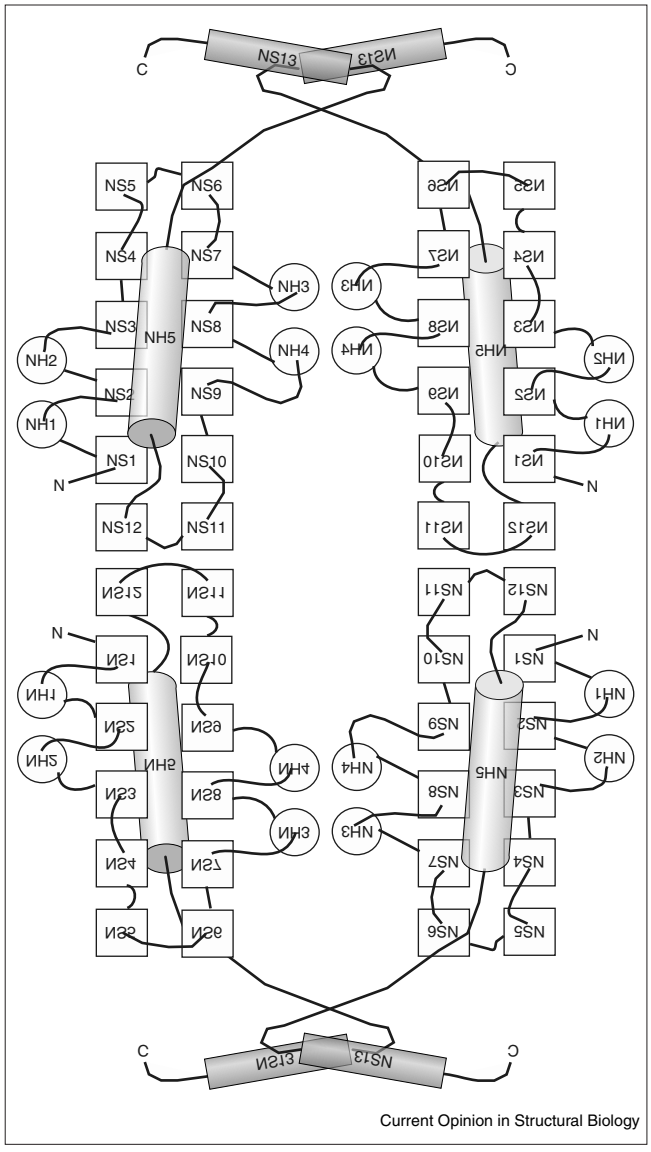
Branch 1 nitrilases: insights into specificity

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 *Nicotania*, and reported that Nit4 constitutes a distinct subbranch 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 acid 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 function 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)-CO_2^-] 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

Figure 3



Schematic structure of the Nit tetramer at the core of the NitFhit crystal structure [10••]. β Strands are numbered NS1 through NS13. α Helices are numbered NH1 through NH5. NS13 is conserved only in branch 10 and is found associated with Fhit domains [10••].

Table 1.

Branches of the nitrilase superfamily.

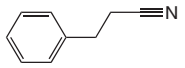
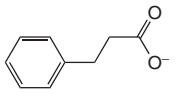
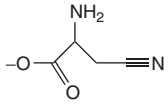
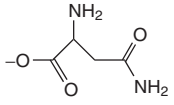
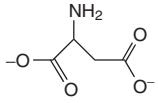
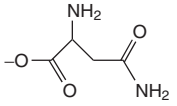
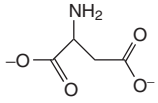
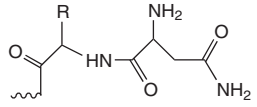
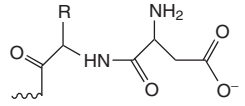
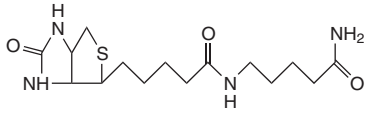
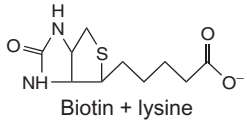
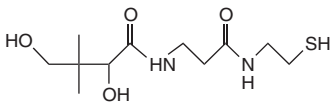
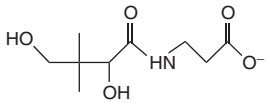
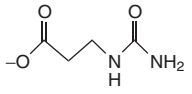
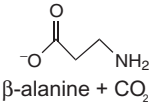
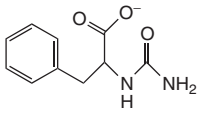
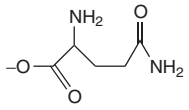
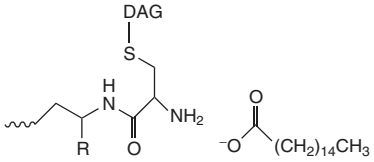
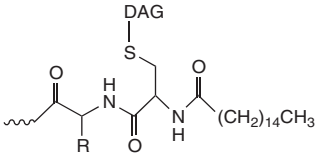
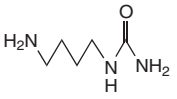
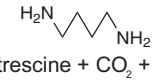
Branch number Enzyme name Prototypical gene NCBI accession code PDB code [Key references]	Substrates	Products
1.1 Auxin-producing nitrilase Arabidopsis NIT1 P32961 [25•]	 Phenylpropionitrile + $2H_2O$	 Phenylacetic acid + NH_4^+
1.2 Aliphatic nitrilase Rhodococcus nitA Q03217 [16]	 Acrylonitrile + $2H_2O$	 Acrylic acid + NH_4^+
1.4 β-cyano-L-alanine hydratase/nitrilase Arabidopsis NIT4 P46011 [29••]	+H_2O  β-cyano-alanine +$2H_2O$	 Asparagine  Aspartic acid + NH_4^+
2 Aliphatic amidase Pseudomonas amiE P11436 [20•]	 Asparagine + H_2O	 Aspartic acid + NH_4^+
3 N-terminal amidase Saccharomyces NTA1 NP_012596 [36]	 N-terminal asparagine + H_2O	 N-terminal aspartic acid + NH_4^+
4.1 Biotinidase human BTD P43251 [37]	 Biocytin + H_2O	 Biotin + lysine
4.2 Pantetheinase/vanin mouse vnn-1 NP_035834 [40,42•]	 Pantethine + H_2O	 Pantothenic acid + cysteamine
5 β-Ureidoproprionase rat BAS Q02348 [44]	 N-carbamyl β-alanine + $2H_2O$	 β-alanine + CO_2 + NH_4^+

Table 1 (cont.).

Branches of the nitrilase superfamily.

Branch number Enzyme name Prototypical gene NCBI accession code PDB code [Key references]	Substrates	Products
6 Carbamylase Agrobacterium DCase JW0082 1ERZ [11••]	 N-carbamyl-D-phenylalanine + $2H_2O$	D-phenylalanine + CO_2 + NH_4^+
7 and 8 Glutamine-dependent NAD synthetase Saccharomyces QNS1 7 P71911 8 NP_011941 [12••]	 Glutamine + H_20	Glutamic acid + NH_3
9 Apolipoprotein N-acetyltransferase Pseudomonas cutE Q9ZI86 [12••,29••]	 N-terminal DAG-modified cysteine + palmitate	 N-palmitylated protein + H_2O
10 Nit humanNIT1 NP_005591 1EMS [5,10••]	Unknown	Unknown
11 N-carbamyl putrescine amidohydrolase Pseudomonas AguB G83608 [12••,50, this review]	 N-carbamyl putrescine + $2H_2O$	 Putrescine + CO_2 + NH_4^+
12 Branch 12 Sphingomonas IdhX JC4601 [12••]	Unknown, possible protein substrate	Unknown

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].

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 pantethine to pantothenic acid (vitamin B_5) 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)-NH_2] react with the appropriate amidases in a manner that consumes two water molecules to produce the corresponding primary amine (R-NH_2) 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 Fhit by virtue 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://genomebiology.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

1. Moore PB, Steitz TA: **The involvement of RNA in ribosome function.** *Nature* 2002, **418**:229-235.

2. Keating TA, Marshall CG, Walsh CT, Keating AE: **The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains.** *Nat Struct Biol* 2002, **9**:522-526.

3. Rawlings ND, O'Brien E, Barrett AJ: **MEROPS: the protease database.** *Nucleic Acids Res* 2002, **30**:343-346.

4. Brenner C: **Hint, Fhit, and GalT: function, structure, evolution, and mechanism of three branches of the histidine triad superfamily of nucleotide hydrolases and transferases.** *Biochemistry* 2002, **41**:9003-9014.

5. Pekarsky Y, Campiglio M, Siprashvili Z, Druck T, Sedkov Y, Tillib S, Draganescu A, Wermuth P, Rothman JH, Huebner K *et al.*: **Nitrilase and Fhit homologs are encoded as fusion proteins in *Drosophila melanogaster* and *Caenorhabditis elegans.*** *Proc Natl Acad Sci USA* 1998, **95**:8744-8749.

6. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.

7. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D: **Detecting protein function and protein-protein interactions from genome sequences.** *Science* 1999, **285**:751-753.

8. Enright A, Iliopoulos I, Kyrpides N, Ouzounis C: **Protein interaction maps for complete genomes based on gene fusion events.** *Nature* 1999, **402**:86-90.

9. Marcotte E, Pellegrini M, Thompson M, Yeates T, Eisenberg D: **A combined algorithm for genome-wide prediction of protein function.** *Nature* 1999, **402**:83-86.

10. •• Pace HC, Hodawadekar SC, Draganescu A, Huang J, Bieganowski P, Pekarsky Y, Croce CM, Brenner C: **Crystal structure of the worm NitFhit Rosetta stone protein reveals a Nit tetramer binding two Fhit dimers.** *Curr Biol* 2000, **10**:907-917.

This paper satisfied the phylogenetic profiles criterion [9] for Nit and Fhit as Rosetta stone proteins, defined the α-β-β-α protein fold and the Glu-Lys-Cys catalytic triad that are found throughout the nitrilase superfamily, and showed how Nit and Fhit fit together.

11. •• Nakai T, Hasegawa T, Yamashita E, Yamamoto M, Kumasaka T, Ueki T, Nanba H, Ikenaka Y, Takahashi S, Sato M *et al.*: **Crystal structure of *N*-carbamyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases.** *Structure* 2000, **8**:729-737.

Independent of [10••], this paper defined the α-β-β-α protein fold and the Glu-Lys-Cys catalytic triad that are found throughout the nitrilase superfamily.

12. •• Pace HC, Brenner C: **The nitrilase superfamily: classification, structure and function.** *Genome Biology* 2001, **2**:1-9.

This paper classified the nitrilase superfamily into 13 branches, showing that only branch 1 contains nitrilases, whereas the other branches perform amide hydrolysis and condensation chemistry. Seven of the thirteen branches were functionally defined, in part, by the presence of Rosetta stone [7,8] fusions. In the most striking example, branches 7 and 8 were inferred to be glutaminases for associated NAD synthetase domains.

13. Wang WC, Hsu WH, Chien FT, Chen CY: **Crystal structure and site-directed mutagenesis studies of N-carbamoyl-D-amino-acid amidohydrolase from *Agrobacterium radiobacter* reveals a homotetramer and insight into a catalytic cleft.** *J Mol Biol* 2001, **306**:251-261.

14. Lahm A, Suck D: **DNase I-induced DNA conformation. 2 Å structure of a DNase I-octamer complex.** *J Mol Biol* 1991, **222**:645-667.

15. Stevenson DE, Feng R, Dumas F, Groleau D, Mihoc A, Storer AC: **Mechanistic and structural studies on *Rhodococcus* ATCC 39484 nitrilase.** *Biotechnol Appl Biochem* 1992, **15**:283-302.

16. Kobayashi M, Komeda H, Yanaka N, Nagasawa T, Yamada H: **Nitrilase from *Rhodococcus rhodochrous* J1. Sequencing and overexpression of the gene and identification of an essential cysteine residue.** *J Biol Chem* 1992, **267**:20746-20751.

17. Bork P, Koonin EV: **A new family of carbon-nitrogen hydrolases.** *Protein Sci* 1994, **3**:1344-1346.

18. Novo C, Tata R, Clemente A, Brown PR: ***Pseudomonas aeruginosa* aliphatic amidase is related to the nitrilase/cyanide hydratase enzyme family and Cys166 is predicted to be the active site nucleophile of the catalytic mechanism.** *FEBS Lett* 1995, **367**:275-279.

19. Grifantini R, Pratesi C, Galli G, Grandi G: **Topological mapping of the cysteine residues of N-carbamyl-D-amino-acid amidohydrolase and their role in enzymatic activity.** *J Biol Chem* 1996, **271**:9326-9331.

20. • Novo C, Farnaud S, Tata R, Clemente A, Brown PR: **Support for a three-dimensional structure predicting a Cys-Glu-Lys catalytic triad for *Pseudomonas aeruginosa* amidase comes from site-directed mutagenesis and mutations altering substrate specificity.** *Biochem J* 2002, **365**:731-738.

This paper confirmed the essentiality of glutamic acid and lysine [10••,11••] in the catalytic mechanism of the nitrilase superfamily.

21. Huang W, Jia J, Cummings J, Nelson M, Schneider G, Lindqvist Y: **Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold.** *Structure* 1997, **5**:691-699.

22. Patricelli MP, Cravatt BF: **Clarifying the catalytic roles of conserved residues in the amidase signature family.** *J Biol Chem* 2000, **275**:19177-19184.

23. Zalkin H, Smith JL: **Enzymes utilizing glutamine as an amide donor.** *Adv Enzymol Relat Areas Mol Biol* 1998, **72**:87-144.

24. Bartling D, Seedorf M, Mithofer A, Weiler EW: **Cloning and expression of an *Arabidopsis* nitrilase which can convert indole-3-acetonitrile to the plant hormone, indole-3-acetic acid.** *Eur J Biochem* 1992, **205**:417-424.

25. • Vorwerk S, Biernacki S, Hillebrand H, Janzik I, Muller A, Weiler EW, Piotrowski M: **Enzymatic characterization of the recombinant *Arabidopsis thaliana* nitrilase subfamily encoded by the NIT2/NIT1/NIT3-gene cluster.** *Planta* 2001, **212**:508-516.

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

26. Osswald S, Wajant H, Effenberger F: **Characterization and synthetic applications of recombinant AtNIT1 from *Arabidopsis thaliana.*** *Eur J Biochem* 2002, **269**:680-687.

27. Ludwig-Muller J, Cohen JD: **Identification and quantification of three active auxins in different tissues of *Tropaeolum majus.*** *Physiol Plant* 2002, **115**:320-329.

28. Kutz A, Muller A, Hennig P, Kaiser WM, Piotrowski M, Weiler EW: **A role for nitrilase 3 in the regulation of root morphology in sulphur-starving *Arabidopsis thaliana.*** *Plant J Cell Mol Biol* 2002, **30**:95-106.

29. •• Piotrowski M, Schonfelder S, Weiler EW: **The *Arabidopsis thaliana* isogene NIT4 and its orthologs in tobacco encode beta-cyano-L-alanine hydratase/nitrilase.** *J Biol Chem* 2001, **276**:2616-2621.

This paper showed that Nit4 orthologs constitute a new sub-branch of nitrilases with specificity for cyanoalanine and that some nitrilases do not exclusively produce an acid product.

30. Harper DB: **Characterization of a nitrilase from *Nocardia* sp. (*Rhodochrou*s group) N.C.I.B. 11215, using p-hydroxybenzonitrile as sole carbon source.** *Int J Biochem* 1985, **17**:677-683.

31. Blumenthal SG, Hendrickson HR, Abrol YP, Conn EE: **Cyanide metabolism in higher plants. 3. The biosynthesis of beta-cyanolanine.** *J Biol Chem* 1968, **243**:5302-5307.

32. Yamamoto K, Oishi K, Fujimatsu I, Komatsu K: **Production of R-(-)-mandelic acid from mandelonitrile by *Alcaligenes faecalis* ATCC 8750.** *Appl Environ Microbiol* 1991, **57**:3028-3032.

33. Inagaki M, Hiratake J, Nishioka T, Oda J: **One-pot synthesis of optically active cyanohydrin acetates from aldehydes via lipase-catalyzed kinetic resolution coupled with in situ formation and racemization of cyanohydrins.** *J Org Chem* 1992, 57:5643-5649.

34. Short JM: **Recombinant approaches for accessing biodiversity.** *Nat Biotechnol* 1997, **15**:1322-1323.

35. • DeSantis G, Zhu Z, Greenberg WA, Wong K, Chaplin J, Hanson SR, Farwell B, Nicholson LW, Rand CL, Weiner DP *et al.*: **An enzyme library approach to biocatalysis: development of nitrilases for enantioselective production of carboxylic acid derivatives.** *J Am Chem Soc* 2002, **124**:9024-9025.
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.

36. Baker RT, Varshavsky A: **Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway.** *J Biol Chem* 1995, **270**:12065-12074.

37. Hymes J, Wolf B: **Biotinidase and its roles in biotin metabolism.** *Clin Chim Acta* 1996, **255**:1-11.

38. Pomponio RJ, Hymes J, Reynolds TR, Meyers GA, Fleischhauer K, Buck GA, Wolf B: **Mutations in the human biotinidase gene that cause profound biotinidase deficiency in symptomatic children: molecular, biochemical, and clinical analysis.** *Pediatr Res* 1997, **42**:840-848.

39. Swango KL, Demirkol M, Hünter G, Pronicka E, Sykut-Cegielska J, Schulze A, Wolf B: **Partial biotinidase deficiency is usually due to the D444H mutation in the biotinidase gene.** *Hum Genet* 1998, **102**:571-575.

40. Aurrand-Lions M, Galland F, Bazin H, Zakharyev VM, Imhof BA, Naquet P: **Vanin-1, a novel GPI-linked perivascular molecule involved in thymus homing.** *Immunity* 1996, **5**:391-405.

41. Maras B, Barra D, Dupre S, Pitari G: **Is pantetheinase the actual identity of mouse and human vanin-1 proteins?** *FEBS Lett* 1999, **461**:149-152.

42. • Pitari G, Malergue F, Martin F, Philippe JM, Massucci MT, Chabret C, Maras B, Dupre S, Naquet P, Galland F: **Pantetheinase activity of membrane-bound Vanin-1: lack of free cysteamine in tissues of Vanin-1 deficient mice.** *FEBS Lett* 2000, **483**:149-154.
This paper allows branch 4 of the nitrilase superfamily to be further classified into biotinidases and pantetheinases.

43. Fujii T, Ahn JY, Kuse M, Mori H, Matsuda T, Isobe M: **A novel photoprotein from oceanic squid (*Symplectoteuthis oualaniensis*) with sequence similarity to mammalian carbon-nitrogen hydrolase domains.** *Biochem Biophys Res Commun* 2002, **293**:874-879.

44. Kvalnes-Krick KL, Traut TW: **Cloning, sequencing, and expression of a cDNA encoding beta-alanine synthase from rat liver.** *J Biol Chem* 1993, **268**:5686-5693.

45. Zalkin H: **NAD synthetase.** *Methods Enzymol* 1985, **113**:297-302.

46. Cantoni R, Branzoni M, Labo M, Rizzi M, Riccardi G: **The MTCY428.08 gene of *Mycobacterium tuberculosis* codes for NAD$^+$ synthetase.** *J Bacteriol* 1998, **180**:3218-3221.

47. Willison JC, Tissot G: **The *Escherichia coli* efg gene and the *Rhodobacter capsulatus* adgA gene code for NH_3-dependent NAD synthetase.** *J Bacteriol* 1994, **176**:3400-3402.

48. Tokunaga M, Tokunaga H, Wu HC: **Post-translational modification and processing of *Escherichia coli* prolipoprotein *in vitro*.** *Proc Natl Acad Sci USA* 1982, **79**:2255-2259.

49. Rogers SD, Bhave MR, Mercer JFB, Camakaris J, Lee BTO: **Cloning and characterization of *cutE*, a gene involved in copper transport in *Escherichia coli*.** *J Bacteriol* 1991, **173**:6742-6748.

50. Nakada Y, Jiang Y, Nishijyo T, Itoh Y, Lu CD: **Molecular characterization and regulation of the aguBA operon, responsible for agmatine utilization in *Pseudomonas aeruginosa* PAO1.** *J Bacteriol* 2001, **183**:6517-6524.