Functional Proteomic and Structural Insights into Molecular Recognition in the Nitrilase Family Enzymes[†,‡]

Katherine T. Barglow,[§,||] Kumar S. Saikatendu,[⊥] Michael H. Bracey,[§,||] Ruth Huey,[⊥] Garrett M. Morris,[⊥] Arthur J. Olson,[⊥] Raymond C. Stevens,[⊥] and Benjamin F. Cravatt*[,§,||]

The Skaggs Institute for Chemical Biology and Departments of Chemical Physiology and Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received September 18, 2008

ABSTRACT: Nitrilases are a large and diverse family of nonpeptidic C−N hydrolases. The mammalian genome encodes eight nitrilase enzymes, several of which remain poorly characterized. Prominent among these are nitrilase-1 (Nit1) and nitrilase-2 (Nit2), which, despite having been shown to exert effects on cell growth and possibly serving as tumor suppressor genes, are without known substrates or selective inhibitors. In previous studies, we identified several nitrilases, including Nit1 and Nit2, as targets for dipeptide−chloroacetamide activity-based proteomics probes. Here, we have used these probes, in combination with high-resolution crystallography and molecular modeling, to systematically map the active site of Nit2 and identify residues involved in molecular recognition. We report the 1.4 Å crystal structure of mouse Nit2 and use this structure to identify residues that discriminate probe labeling between the Nit1 and Nit2 enzymes. Interestingly, some of these residues are conserved across all vertebrate Nit2 enzymes and, conversely, not found in any vertebrate Nit1 enzymes, suggesting that they are key discriminators of molecular recognition between these otherwise highly homologous enzymes. Our findings thus point to a limited set of active site residues that establish distinct patterns of molecular recognition among nitrilases and provide chemical probes to selectively perturb the function of these enzymes in biological systems.

Nitrilases are ubiquitous enzymes found in all species of plant, animal, and fungi, as well as many prokaryotes, with more than 150 known members that constitute 13 distinct branches that all share a common CN hydrolase domain (*1*). Despite their name, only the first branch of the nitrilase family, found exclusively in bacteria, actually hydrolyzes nitriles; the others cleave (or form) small molecule C−N bonds such as acid amides, secondary amides, ureas, and carbamates (*2, 3*). All nitrilases appear to share a common catalytic mechanism that involves a highly conserved cysteine nucleophile, a glutamate base, and a active site lysine that completes the catalytic triad (*2*). Mutation of any of these three residues leads to an inactive enzyme (*4*).

Several branches of the nitrilase superfamily are found in mammalian genomes. The vanins/biotinidases/pantetheinases are involved in vitamin synthesis (*5, 6*), and the glutaminase domains of glutamine-dependent NAD synthases generate ammonia from glutamine for NAD synthesis (*7, 8*). β-Ureidopropionases (Upβ) catalyze the hydrolysis of *N*-carbamoyl-β-alanine (CarβAla)[1] to β-alanine, which is a central step in the pyrimidine catabolic pathway (*9*). Mutation of Upβ in humans leads to a buildup of CarβAla, which exerts neurotoxic effects resulting in severe neuropathology and movement disorders (*9*).

All mammalian species also contain two uncharacterized nitrilases, termed Nit1 and Nit2, which share ∼40% sequence identity and a lower degree of homology with Upβ (Figure 1A). Nit1 knockout mice show a variety of phenotypes, including accelerated cell proliferation, enhanced survival of cells in response to DNA-damaging agents, and increased incidence of *N*-nitrosomethylbenzylamine-induced tumors (*10*). Nit1 overexpression decreases cell viability and increased caspase-3-dependent apoptosis. In combination, these data point to a possible role for Nit1 as a tumor suppressor gene. Overexpression of Nit2 in HeLa cells leads to a sharp decrease in cell proliferation and growth. In contrast to Nit1, this effect does not occur through activation of apoptosis but by the induction of G2 arrest (*11*). Interestingly, the *Nit2* gene is located at a genomic locus susceptible to allelic instability, which was detected in 12−38% of tested human primary tumors, indicating that Nit2 may also play a role in carcinogenesis (*11*). Finally, Nit1 and Nit2 show very distinct tissue distributions, with Nit2 being highly expressed in liver

---

[†] This work was supported by the National Institutes of Health (Grant CA087660 to B.F.C. and Grants R24-CA095830 and R01-GM069832 to A.J.O.), the Skaggs Institute for Chemical Biology, and a Howard Hughes Medical Institute predoctoral fellowship (to K.T.B.).

[‡] The coordinates for the mNit2 structure are available from the Protein Data Bank (id code 2w1v).

[*] To whom correspondence should be addressed: e-mail, cravatt@scripps.edu; phone, 858-784-8633; fax, 858-784-8023.

[§] The Skaggs Institute for Chemical Biology, The Scripps Research Institute.

[||] Department of Chemical Physiology, The Scripps Research Institute.

[⊥] Department of Molecular Biology, The Scripps Research Institute.

[1] Abbreviations: α-CA, α-chloroacetamide; CarβAla, *N*-carbamoyl-β-alanine; Nit1, nitrilase-1; Nit2, nitrilase-2; Upβ, β-ureidopropionase.

10.1021/bi801786y CCC: $40.75

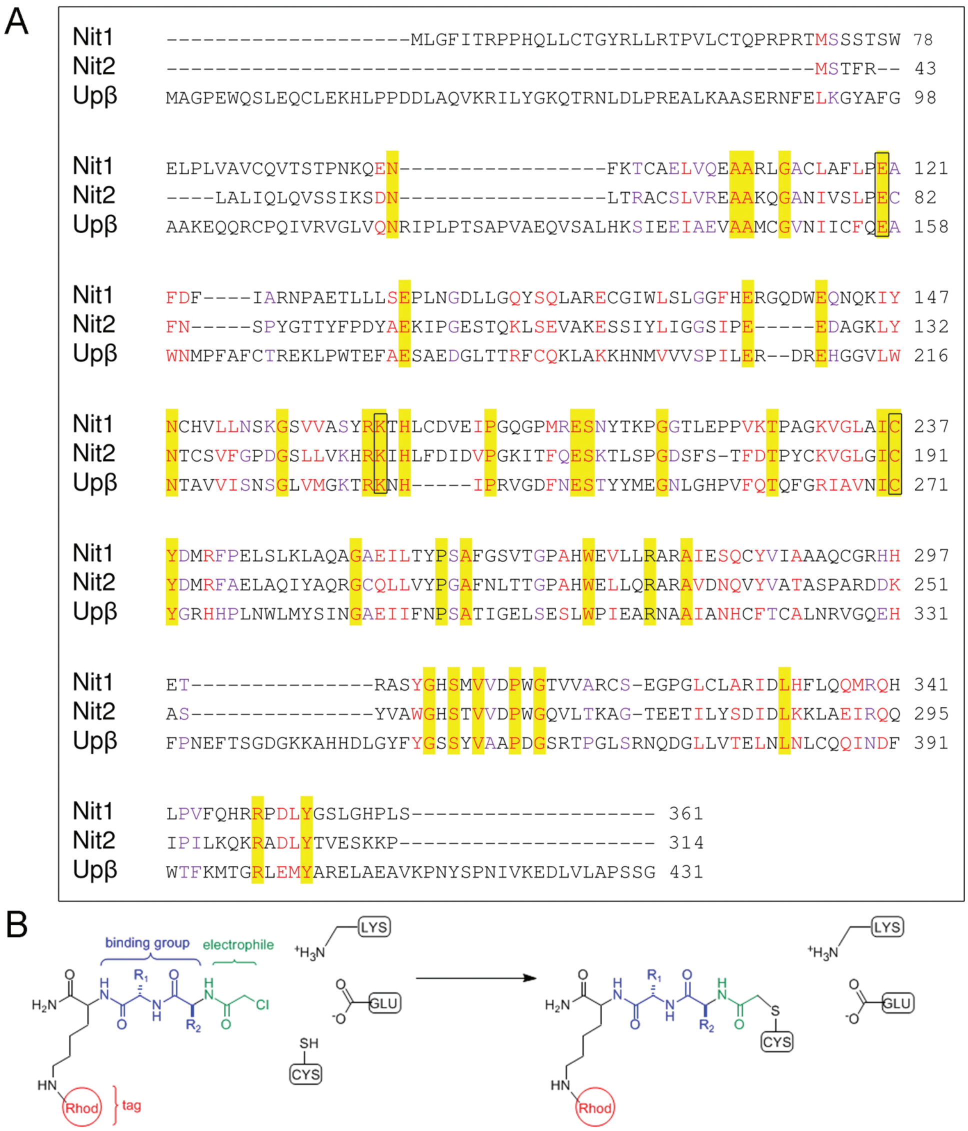

FIGURE 1: Mammalian nitrilase enzymes. (A) Sequence alignment of the murine Nit proteins with Upβ. Identical residues are highlighted; similar residues shown in red (highly similar) and purple (similar). Catalytic residues are boxed. Figure generated by CLUSTALW (*37*). (B) Schematic of labeling of a nitrilase catalytic cysteine by a dipeptide−α-chloroacetamide (α-CA) activity-based probe. R_1 and R_2, amino acid side chains; Rhod, carboxytetramethylrhodamine fluorophore for detection of labeled proteins.

and kidney, as well as fertilized eggs and oocytes, and Nit1 showing a more ubiquitous expression across mouse tissues (*10*).

Taken together, these previous studies suggest that Nit1 and Nit2 may have potentially overlapping but likely distinct biological functions. However, molecular probes (substrates or inhibitors) that can distinguish between Nit1 and Nit2 are lacking. Toward this end, and to achieve a more detailed understanding of the structural features that guide molecular recognition among members of the nitrilase family, we report here the use of activity-based proteomics (*12, 13*) and X-ray crystallography to identify key active site residues that discriminate small-molecule interactions for the Nit1 and Nit2 enzymes. The conservation of some of these residues across the entire mammalian Nit1 and Nit2 families suggests that they may also provide the basis for recognition of distinct sets of endogenous substrates.

MATERIALS AND METHODS

Cloning and Expression of mNit2. The coding sequence of mouse nitrilase-2 (mNit2) was amplified by PCR from a mouse liver cDNA library using the following primers: sense primer 5′-ATG TCT ACT TTC CGC CTG GCC CTC ATA CAG CTT CAA G-3′ and antisense primer 5′-TCAAGGCTTCTTTGATTCCACTGTATAGAGGTCTGCTCG-3′, to give an approximately 800 bp band that was cloned directly into pTrcHis-Topo (Invitrogen). The construct was confirmed by sequencing and then expressed in the BL21(DE3) strain of *Escherichia coli* cells as a N-terminal hexahistidine fusion with the gene 10 leader sequence, giving a protein product with a mass of 34.9 kDa. Cells were grown in LB media with 75 mg/L carbenicillin to an OD_{600} of 0.5 with shaking at 37 °C, and protein expression was induced with 1 mM IPTG. Cells were harvested 6 h later by centrifugation and frozen at −80 °C.

Purification of mNit2. Cells were lysed by stirring for 1 h at 4 °C in a solution of 10 mM Hepes/10 mM NaCl at pH 8 (buffer A) supplemented with 1 mg/mL lysozyme and 1 μg/mL DNase I. The lysate was then sonicated with a 50 W top sonicator for 1.5 min using 0.5 s pulses and centrifuged at 10000*g* for 10 min. To the supernatant was added Talon cobalt affinity resin (Clontech; 400 μL of slurry/g of cell paste), and the mixture was rotated at room temperature for 1 h. Beads were collected by centrifugation at 700*g* for 1 min, then washed twice in batch with buffer A, and applied to a 1 cm column. The column was washed with buffer A (10 mL/400 μL of resin slurry) and then with buffer A/1 M NaCl (5 mL/400 μL of resin slurry). The bound protein was eluted by the addition of 200 mM imidazole (3 mL/500 μL of resin slurry) to the capped column and incubation for 20 min. The eluted protein was concentrated to 0.5 mL using an Amicon centrifugal filter device (Millipore) and injected on a Superose 6 10/300 GL column (GE Healthcare). Following isocratic elution in buffer A, the purified protein was stored at 4 °C. Concentration of purified protein was determined by measuring the A_{280}, where a 1 mg/mL solution was estimated to have an absorbance of 1.23. These values correlated well with values determined by a Bio-Rad D_c protein assay kit. These conditions yielded highly pure (>99% by SDS−PAGE) mNit2 at approximately 15 mg of protein/L of culture. The protein appears to be a monomer in solution based on size-exclusion chromatography results, with occasional observations of a small amount of dimeric complex (Supporting Information Figure S1).

Crystallization of mNit2. Preliminary crystal screens for hanging drop vapor diffusion were set up in 24-well trays using the Index screen from Hampton Research at both 12 and 24 mg/mL purified mNit2. The hexahistidine tag and gene 10 leader sequence were not removed prior to crystallization. Spindles of crystals were obtained within 3 days at room temperature from a reservoir solution of 50 mM calcium chloride, 100 mM Bis-Tris, pH 6.5, and 30% (v/v) polyethylene glycol monomethyl ether 550 (PEG MME 550). This condition was optimized by varying the concentration of $CaCl_2$ and PEG MME 550 as well as the pH and buffer conditions. Optimal conditions, which gave large (up to 0.5 mm) X-shaped crystals, were 250 mM $CaCl_2$, 100 mM Hepes, pH 8.0, and 30% (v/v) PEG MME 550, with 12 mg/mL mNit2. Drops consisted of 1 μL of reservoir solution and 1 μL of mNit2 solution (in buffer A).

Data Collection and Analysis. Crystals were flash-frozen in liquid nitrogen without further cryoprotection. X-ray diffraction data were obtained at the Stanford Synchrotron Radiation Laboratory (SSRL) beamline 11.1 using a MAR325 detector. Data was collected at 1° oscillation for 180°. Crystals of mNit2 diffracted to a resolution of 1.4 Å and belonged to the space group *P*32, with unit cell parameters $a = 90.807$ Å, $b = 90.807$ Å, $c = 54.087$ Å, $\alpha = 90°$, $\beta = 90°$, and $\gamma = 120°$. The monochromatic beam wavelength was 1.000 Å. Diffraction images were processed, integrated, and scaled using HKL2000 (*14*) in hexagonal lattice *P*3. Systematically attempting molecular replacement in all the Laue groups showed the actual space group to be *P*32. Initial phases were obtained using Phaser (*15*) using an all-atom partially trimmed monomer of yeast NIT3 nitrilase (PDB id 1F89) as the query model and reflection data range of 20.0−2.5 Å. Phase improvement, model bias removal, and automated model building by ARP/wARP led to an 85% complete initial model with two chains in the asymmetric unit. The missing residues were built in manually, and the model was improved by alternating rounds of manual building in Coot and refining in the Refmac5 (*16*) module of the CCP4 suite of programs (*17*). Both anisotropic temperature factors and TLS (the three vibrational motion descriptors *t*ranslation/*l*ibration/*s*crew) parameters were included during refinement. Optimal TLS parameters were computed using the TLSMD server (*18*). Statistics of data collection, structure refinement, and stereochemical quality of the model are summarized in Table 1. The coordinates for the mNit2 structure are available from the Protein Data Bank (id code 2w1v).

Table 1: Crystallographic Data Collection and Structure Determination Statistics for mNit2

space group	*P*32
unit cell parameters	
size (Å)	$a = 90.807$, $b = 90.807$, $c = 54.087$
angle (deg)	$\alpha = 90$, $\beta = 90$, $\gamma = 120$
wavelength (Å)	1.0000
resolution range (Å)	78.57−1.40 (1.49−1.40)
total no. of observations	896905
unique no. of reflections	92543 (6449)
completeness score (%)	99.07 (93.67)
redundancy	2.7 (1.8)
mean $I/(\sigma)$	41.35 (2.57)
R_{sym} on I	0.031 (0.284)
refinement statistics	
resolution range (Å)	40−1.4 (1.43−1.40)
R_{cryst}	0.128 (0.222)
R_{free}	0.179 (0.279)
bond length (Å) (rmsd)	0.008
bond angle (deg) (rmsd)	1.25
av B value (Å^2)	12.1
validation and stereochemistry	
most favored (%)	98.6
generously allowed (%)	1.4

Preparation of Mutants of mNit2. Mutants were prepared by either Quikchange mutagenesis or overlap extension PCR (for mutation primer sequences, see Supporting Information). For Quikchange reactions, 125 ng of each primer was used per reaction. PCRs were performed with 40 ng of mNit2-pTrcHis template and Turbo Pfu polymerase (Stratagene). Overlap extension PCRs were performed with Phusion polymerase (New England Biolabs) with the following primers: 5′-ATGTCTACTTTCCGCCTGGCCCTCATAC-AGCTTCAAG-3′ paired with the reverse mutation primer and 5′-TCAAGGCTTCTTTGATTCCACTGTATAGAGG-TCTGCTCG-3′ paired with the forward mutation primer. These PCRs were individually gel purified, then combined pairwise, and subjected to 30 PCR cycles without primer. The products from these PCRs were directly used in a final PCR with the primers above to give the mutant product. These products were either directly cloned into pTrcHis-Topo (invitrogen) or subjected to a second round of PCR with the following primers: 5′-GGCTCGAGATGTCTACTTT-CCGCCTGGCCCTCATAC-3′ containing an *Xho*I restriction site and 5′-GGGGTACCTCAAGGCTTCTTTGATTCCA-CTGTATAGAGG-3′ containing a *Kpn*I restriction site. These PCR products were subcloned into PCR-Topo2.1 (Invitrogen) and then transferred into pTrcHisA (Invitrogen). All clones were confirmed by sequencing.

Cloning, Expression, and Purification of mUpβ. The coding sequence of mouse ureidopropionase-β (mUpβ) was amplified by PCR from a mouse liver cDNA library and subcloned into pTrcHisA (Invitrogen). The construct was confirmed by sequencing and then expressed in the BL21(DE3) strain of *E. coli* cells as an N-terminal hexahistidine fusion with a gene 10 leader sequence, giving a protein product with a mass of 48.3 kDa. Cells were grown in LB media with 75 mg/L carbenicillin to an OD_{600} of 0.4 with shaking at 37 °C, and protein expression was induced with 1 mM IPTG. Cells were harvested 6 h later by centrifugation and frozen at −80 °C. Cells were lysed by stirring for 1 h at 4 °C in a solution of 10 mM Hepes/10 mM NaCl at pH 8 (buffer A) supplemented with 1 mg/mL lysozyme and 1 μg/mL DNase I. The lysate was then sonicated with a 50 W top sonicator for 1.5 min using 0.5 s pulses and then centrifuged at 10000*g* for 10 min. To the supernatant was added Talon cobalt affinity resin (Clontech, 150 μL of slurry/g of cell paste), and the mixture was rotated at room temperature for 1 h. Beads were collected by centrifugation at 700*g* for 1 min, then washed twice in batch with buffer A, and applied to a 1 cm column. The column was washed with buffer A (10 mL/150 μL of resin slurry) and then with buffer A/1 M NaCl (5 mL/150 μL of resin slurry) and then with 10 mM imidazole (5 mL/150 μL of resin slurry). The bound protein was eluted by the addition of 200 mM imidazole (3 mL/500 μL of resin slurry) to the capped column and incubation for 20 min. The eluted protein was concentrated to 0.5 mL using an Amicon centrifugal centrifugal filter device (Millipore) and injected on a Superose 6 10/300 GL column (GE Healthcare). Following isocratic elution in buffer A supplemented with 1 mM TCEP (essential to maintain the protein as a single monomeric peak on the column and also to maintain protein stability for overnight storage or concentration past 8 mg/mL), the purified protein was stored at 4 °C. Concentration of purified protein was determined by measuring the A_{280}, where a 1 mg/mL solution was estimated to have an absorbance of 1.28. These values correlated well with values determined by a Bio-Rad D_c protein assay kit. These conditions yielded highly pure (>95% by SDS−PAGE) mUpβ at approximately 2 mg/L of culture.

Expression of mNit1. The coding sequence of mouse nitrilase-1 (mNit1) was amplified by PCR from a mouse liver cDNA library and subcloned into pcDNA3.1 (Invitrogen) and transiently transfected into COS7 cells using Fugene (Roche).

Molecular Modeling of the Probe−mNit2 Interaction. The adducts of the Leu-Arg-α-CA and Leu-Glu-α-CA probes with the α-carbon and γ-sulfur atoms of the side chain of Cys191 were built using the ChemDraw utility of ChemBio3D Ultra 3D (version 11.0; CambridgeSoft) and energy-minimized using standard MM2 parameters (adapted by CambridgeSoft from Molecular Mechanics; Burkert, U., and Allinger, N. L., ACS Monograph 177, American Chemical Society, Washington, DC, 1982). These probe−cysteine adducts were set up for covalent docking using AutoDockTools (ADT) version 1.4.4 (http://autodock.scripps.edu/resources/adt) (*19, 20*), the graphical user interface to AutoDock, allowing all non-amide bonds to be rotatable. The cysteine β-carbon atom in the adduct was changed to a covalently binding atom type, and the mNit2 loop-shifted model was modified by deleting the β-carbon and γ-sulfur from the side chain of Cys191. Atomic affinity maps, an electrostatic potential map, and a covalent affinity map were computed using AutoGrid4 (*20*). The covalent affinity map was calculated using a spherically symmetric inverted Gaussian-shaped potential energy with a half-width of 4.0 Å and an energy penalty of 1000.0 kcal/mol, with the minimum energy value centered on the coordinates of the β-carbon atom of Cys191 from the original pdb file (i.e., 16.759 69.024 23.132). In subsequent dockings, the covalently binding atom type in the adduct seeks out this position, effectively binding the ligand to the receptor. AutoDock 4 (*20*) was used to perform 500 independent dockings, each one starting from random initial positions. The Lamarckian Genetic Algorithm was used with a population size of 150 and a maximum number of energy evaluations of 25 million, while default values were accepted for all of the other parameters. Results were clustered using ADT, using an rmsd tolerance of 2.0 Å.

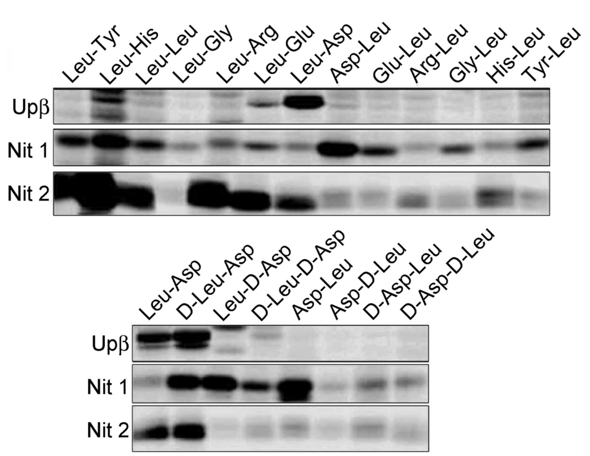

FIGURE 2: Probe reactivity profiles for murine nitrilase family enzymes. Upβ and Nit1 are shown using 1 mg/mL transiently transfected COS7 soluble proteome; Nit2 (which, in our hands, did not express in COS7 cells) is shown using 2.5 μM purified protein. All probe labeling was done at 20 μM, with a 1 h incubation at room temperature. Fluorescent images are shown in gray scale.

RESULTS AND DISCUSSION

Reactivity Profiles of Nit Proteins with Dipeptide−α-CA Activity-Based Probes. In previous studies (*21*), we discovered that multiple mammalian nitrilases (Upβ, Nit1, and Nit2) are targets of dipeptide−α-chloroacetamide (α-CA) activity-based proteomic probes. The binding group (the dipeptide) of α-CA probes facilitates interactions with structured sites in proteins, where the electrophilic CA group reacts with local nucleophiles. α-CA probes also contain a rhodamine and/or biotin group to enable detection, quantification, and enrichment/identification of protein targets directly from complex proteomes (*22*).

For nitrilase enzymes, the catalytic cysteine reacts with α-CA probes to form a stable thioether bond (Figure 1B) (*21*). Upβ was found to react strongly and exclusively with probes containing L-aspartate in the second position (*21*), which bear striking structural similarity to the endogenous Upβ substrate CarβAla (*9*). These data support a model where α-CA probes act as substrate mimetics for Upβ. Whether these probes also inform on important features of molecular recognition for other nitrilases (e.g., Nit1, Nit2), however, remained unknown. To explore this premise in more detail, we compared the reactivity profiles of Upβ, Nit1, and Nit2 with a library of structurally diverse dipeptide−α-CA probes.

Murine nitrilases were recombinantly expressed in *E. coli* (Upβ, Nit2) or COS7 cells (Nit1), treated with a panel of rhodamine-tagged dipeptide−α-CA probes (20 μM) for 1 h, and then analyzed by SDS−PAGE and in-gel fluorescence scanning (Figure 2). Consistent with previous results, Upβ

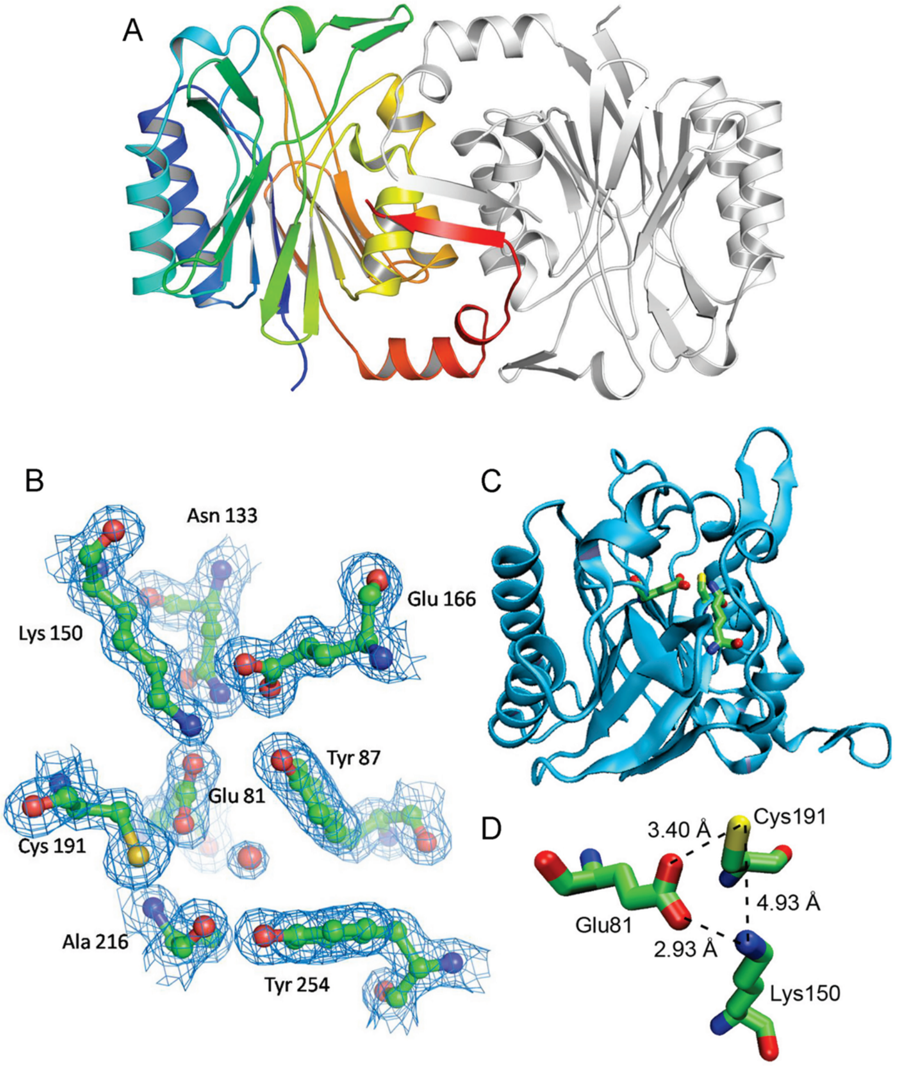

FIGURE 3: Structure and active site of mouse Nit2. (A) The mNit2 structure. One monomer is shown in gray; the other is colored by topology and sequence, from blue (N-terminus) to red (C-terminus). (B) The mNit2 active site. Electron density is shown as blue grids; putative atom positions are shown as sticks and balls. (C) Location of the active site within the Nit2 monomer. (D) Catalytic triad with distances shown between key catalytic atoms.

showed a highly restricted probe selectivity profile. In contrast, Nit1 and Nit2 exhibited more promiscuous labeling profiles, reacting with larger, yet distinct subsets of the probe library. Nit1 reacted most strongly with the D-Leu-Asp-, Leu-D-Asp-, and Asp-Leu-α-CA probes but showed detectable labeling with most members of the α-CA library. Interestingly, the Leu-D-Asp- and Asp-Leu-α-CA probes did not react with either Nit2 or Upβ. Nit2 showed a more restricted probe labeling profile, reacting most strongly with probes, such as Leu-Arg- and Leu-Glu-α-CA, that showed little or no labeling with Nit1 or Upβ. Overall, Nit2 preferred to react with probes containing a hydrophobic X_1 group and charged or bulky X_2 groups (where X_2 is the amino acid closest to the α-CA electrophile).

These data thus revealed that Nit1, Nit2, and Upβ, despite sharing rather high sequence homology, can be readily distinguished by their respective reactivity profiles with the α-CA probe library. Considering further that the preferred probes previously discovered for Upβ were found to mimic the structure of this enzyme's endogenous substrate (*9*), we wondered whether the distinct probe reactivity profiles of Nit1 and Nit2 might also provide insights into molecular recognition for these enzymes. We therefore next sought a mechanistic explanation for these differences by obtaining the first high-resolution crystal structure of a mammalian nitrilase.

Crystal Structure of Mouse Nit2. Mouse Nit2 was crystallized and its structure determined to 1.4 Å resolution with excellent data collection and refinement statistics [Table 1; the coordinates are available from the Protein Data Bank (id code 2w1v)]. The final refined model consisted of two crystallographically independent, almost identical monomers in the asymmetric unit and appeared grossly similar to other experimentally determined protein structures of the nitrilase superfamily (*23*). Both monomers showed well-defined electron density from residues 41−312 with no chain breaks. The topology of one monomer along with the adjacent dimer showing interfacial packing is presented in Figure 3A. Each monomer is approximately 41 × 34 × 34 Å and has a four-layered α/β/β/α architecture roughly similar to other nitrilase enzymes from bacteria (*24, 25*) and *Drosophila* (*26*), with a central solvent excluded pair of β-sheets of 12 β-strands. The two β-sheets (with six strands each) are sandwiched between two helical layers, one on either side. The first β-sheet is made up of strands 1, 2, 3, 6, 7, and 16 while the second consists of strands 10−15. Additionally, a β-hairpin

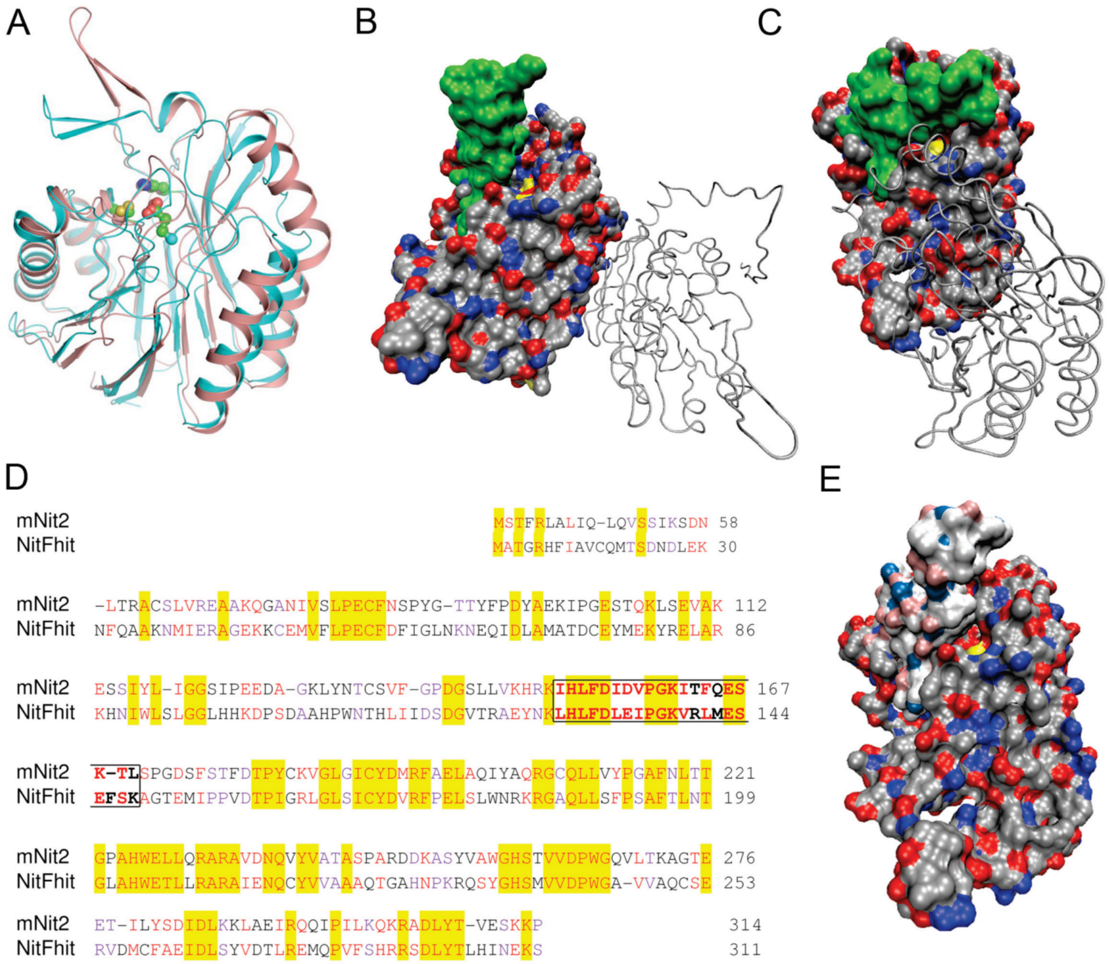

FIGURE 4: Distinct conformations for a loop region covering the active site of nitrilase enzymes. (A) Overlay of the structures of mNit2 and the Nit domain of worm NitFhit, showing the overall similar fold with a significant shift in a loop region covering the active site. mNit2 is shown in blue; worm NitFhit is shown in pink. (B) Surface rendition, colored by atom type, of the Nit domain of worm NitFhit structure. Residues of the active site loop (128−148) are shown in green; the active site cysteine can be seen in yellow. The second chain of the crystallographic dimer is shown as a backbone C_α trace. (C) Surface rendition of mNit2. Residues of the active site loop region (151−171) are shown in green; the active site cysteine is shown in yellow and largely inaccessible to solvent. The second chain of the crystallographic dimer is shown as a backbone C_α trace. (D) Sequence alignment of mNit2 and worm NitFhit. Identical residues are highlighted; similar residues are shown in red and purple. The residues that made up the dynamic loop over the active site are shown in bold and boxed. Figure generated by CLUSTALW (*37*). (E) Loop-shifted homology model of mNit2. The protein is shown as a molecular surface colored by atom type. The loop region is shown in lighter silver/blue/red; the yellow visible in the upper middle of all structures is the sulfur of the active site cysteine.

composed of strands 4 and 5 and a type II′ β-turn is positioned as a lid at one edge of the two β-sheets. DALI (*27*) comparison of mouse Nit2 with known protein structures of PDB clearly identified it to be a member of the carbon−nitrogen hydrolase fold of the structural classification database SCOP (*28*).

The active site of mNit2 (Figure 3B,C) closely resembles that of the Nit domain of NitFhit (*23*) and yeast Nit3 (*29*). The catalytic cysteine sulfur is 3.4 Å from the oxygen of the catalytic Glu81, which is 2.93 Å from the ε-nitrogen of Lys150, completing the catalytic triad (Figure 3D).

A Potentially Dynamic Loop Covering the Active Site Cleft. Comparison of the mNit2 structure with the structures of other nitrilase enzymes shows the overall similarity of the fold (Figure 4A). There is, however, a significant shift in a loop region that covers the active site cavity (Figure 5B,C). This loop is in the "open" position for yeast Nit3 (*29*) and worm NitFhit (*23*) (Figure 4B) and in a "closed" position for the agrobacterium Dcase (*25*) and the mNit2 structure (Figure 4A,C). mNit2 is overall approximately 35% identical and 55% homologous to the Nit domain of worm NitFhit (Figure 4D). Interestingly, an even higher level of similarity (43% identity, 80% homology) is observed in the loop region (residues 151−171 in mNit2 and residues 128−148 in NitFhit), despite this portion of the proteins differing most between the two structures. This difference in loop conformations appears to reflect the distinct dimer interfaces observed in the mNit2 and NitFhit structures. In the Nit domain of NitFhit, the dimer interface is on the opposite face of the protein from the active site (Figure 4B), whereas, for mNit2, the dimer interface lies directly in front of the active site, with dimeric contacts in the loop region (Figure 4C). Whether these two dimeric interfaces reflect distinct mechanisms for regulation of mNit2 and NitFhit or simply represent alternative modes for crystal packing remains unknown. In either case, the high level of sequence conservation displayed by the dynamic active site lid suggests an important role for this loop region in nitrilase function.

Molecular Modeling of α-CA Probes into the mNit2 Structure. We were unable to obtain diffracting crystals of

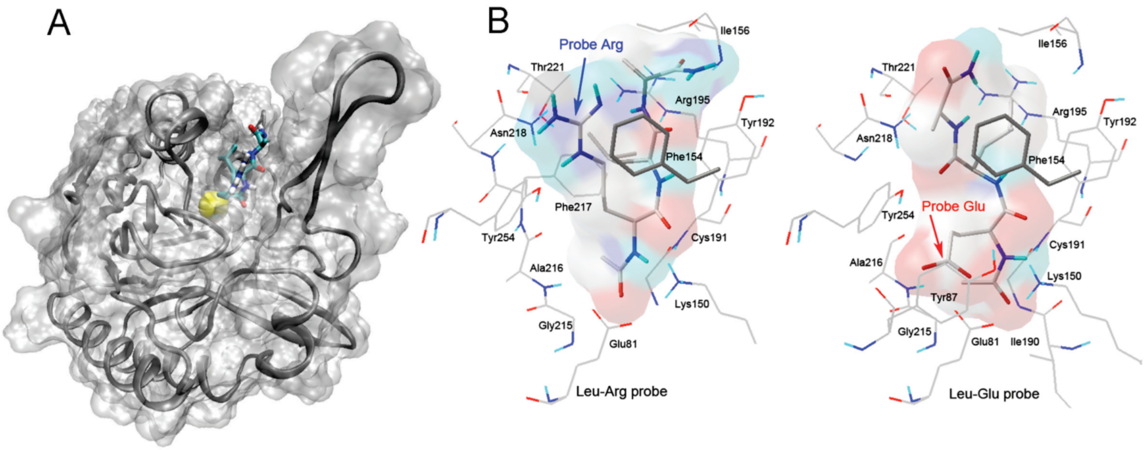

FIGURE 5: Modeling of the Leu-Arg-α-CA and Leu-Glu-α-CA probes into the mNit2 active site. (A) Depiction of the active site cleft within the entire structure. mNit2 protein (with loop shifted, upper right) is shown as a C_α ribbon inside a translucent molecular surface; probe (Leu-Arg-α-CA) is shown in ball-and-stick format. The lowest energy docked conformation of a representative member of the largest cluster (70) of Leu-Arg-α-CA is shown and can be seen covalently bound to Cys191 (sulfur atom shown in yellow) beneath the molecular surface. (B) Close-up view of the active site cleft with residues shown. Leu-Arg-α-CA (left) and Leu-Glu-α-CA (right, both lowest energy members of the largest cluster) are shown in ball-and-stick format with its molecular surface colored gray, except where atoms are in contact with the protein, in which case these parts of the surface are colored by atom type. The side chains in the protein that are within 5 Å of the probe are shown as lines and labeled by residue name and number. Note the different modeled position of the probe X_2 position residues (Arg and Glu, respectively). This figure (part B) was created using AutoDockTools (ADT) (*38*).

mNit2 in complex with dipeptide−α-CA probes, possibly reflecting the fact that large structural rearrangements would be required to accommodate probe access to the enzyme active site. Indeed, initial attempts to model the probe into the mNit2 active site failed, as the "closed" loop conformation prevented interactions of probe molecules with the catalytic cysteine. Given the high homology of the loop region between mNit2 and NitFhit, as well as our data showing that α-CA probes can label recombinant mNit2 in solution, we hypothesized that the loop was trapped in a single, closed conformation in the crystal structure but otherwise capable of accessing the open conformation in solution. We therefore created a loop-shifted model for mNit2 based on the loop position in NitFhit (Figure 4E) and used the AutoDock platform (*19, 20*) to form covalent adducts between this form of mNit2 and dipeptide−α-CA probes. From a gel analysis of mNit2 labeling (Figure 2 and Supporting Information Figure S2), we chose two probes that showed strong labeling with mNit2, Leu-Glu- and Leu-Arg-α-CA for covalent docking into the Nit2 active site. Covalent adducts of the probes with the β-carbon and γ-sulfur atoms of the side chain of Cys191 were built using a chemical modeling program and energy-minimized by molecular dynamics simulation. These probe−cysteine adducts were set up for covalent docking with the mNit2 loop-shifted model modified by deletion of the β-carbon and γ-sulfur from the side chain of Cys191. The results for 500 independent dockings of each probe were clustered using an rmsd tolerance of 2.0 Å (Supporting Information Figure S3). Larger clusters indicate convergence of stochastic trials onto a single low-energy conformation and thereby suggest reliable docking results. The Leu-Glu-α-CA probe showed a single large cluster of 154 members, with the next largest cluster being significantly smaller (<60), while the Leu-Arg-α-CA probe showed two low-energy clusters of 68 and 70 members with all other clusters less than 45 (Supporting Information Figure S3). Of the two Leu-Arg-α-CA clusters, only the larger (70-member) cluster showed probe binding in a mode consistent with our mutagenesis data (see below) and therefore was used in the analysis shown here.

The docking analysis revealed two related, but distinct binding modes for positively charged (Leu-Arg-α-CA) and negatively charged (Leu-Glu-α-CA) probes (Figure 5). In both cases, the probes bind in a cleft lined by Ala216, Tyr87, Phe217, Tyr254, Asn218, Thr221, Phe154, Arg195, and Tyr192, with Cys191, the catalytic residues Lys150 and Glu81, and Gly215 lying at the bottom of the binding cleft. When overlaid (Figure 6A), the Leu-Arg-α-CA probe has the Arg residue in an extended conformation sitting directly above a highly aromatic patch consisting of tyrosines 192, 254, and 87 and phenylalanines 217, 154, and 164 (Figure 6B) and forming a cation−π interaction to Phe154 (Figure 6C). In the case of the Leu-Glu-α-CA probe, the Glu residue appears to fold down (Figure 6A), placing it beneath the aromatic patch in an area rich in potential hydrogen bond donors/acceptors (Figure 6D). In both cases, the Leu residue of the probes appears to sit in a hydrophobic region around Phe217 (Figure 6A, lower right). This analysis also points to the existence of an "aromatic cage" surrounding the active site (Figure 6B,C), consisting of residues Tyr192, Phe154, Phe164, Tyr254, and Phe217. Such "aromatic patches" have been previously described in proteins such as acetylcholinesterase (*30*) and are thought to stabilize positive charges via a network of cation−π interactions and accommodate versatile interaction modes with substrates and inhibitors.

Mutation of Candidate Active Site Residues Involved in Probe Recognition. To confirm the model for probe binding suggested by docking analysis, we made a number of mutants of key residues in the active site cleft and evaluated these mutant enzymes for their respective probe labeling patterns. Most mutants expressed at levels comparable to wild-type mNit2 and had identical size-exclusion chromatography traces (Supporting Information Figure S4), suggesting folded proteins free of aggregates. A mNit2 mutant bearing an alanine in place of the catalytic cysteine Cys191 was used as a control to define background probe labeling, which was

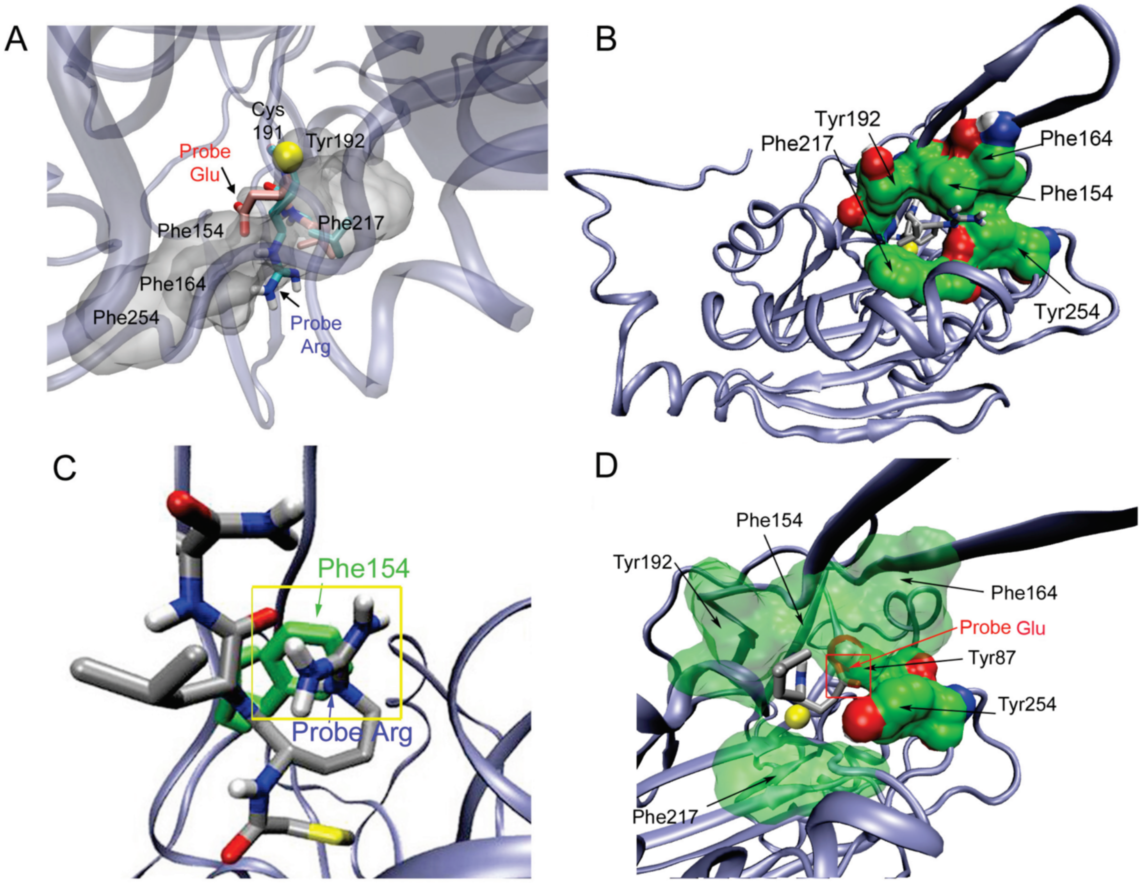

FIGURE 6: Specific interactions of Leu-Arg- and Leu-Glu-α-CA probes modeled in the mNit2 active site. (A) Overlay of the Leu-Arg- (blue) and Leu-Glu- (red) α-CA probes (representative members of the largest cluster). The "aromatic ring" region is shown in translucent gray; the remainder of the protein is shown as a C_α ribbon. The Arg of the Leu-Arg-α-CA probe sits above the aromatic region; the Glu of the Leu-Glu-α-CA probe sits beneath this region. (B) The aromatic "ring" around the active site interacting with the Leu-Arg-α-CA probe. Nit2 is shown as a C_α ribbon, and the modeled Leu-Arg-α-CA probe is colored as in part B, with the guadinium (blue/white) pointed toward the viewer. Shown in green as space-filling models are residues Tyr192, Phe154, Phe164, Tyr254, and Phe217. The active site cysteine is visible in yellow inside the "ring". (C) Predicted cation−π interaction of Phe154 with the Leu-Arg-α-CA probe. Nit2 is shown as a C_α ribbon; Phe154 is shown in green, and the Leu-Arg-α-CA probe is shown in ball-and-stick format colored by atom, with carbon in gray. The guanidinium of the probe aligns with the aromatic group of Phe154, forming a cation−π interaction (yellow box). (D) Same as part C but with the Leu-Glu-α-CA probe. Potential H-bond partners (Tyr87, Tyr254) are shown in solid green, with the Glu residue of the probe (gray/red) sitting between the two tyrosine hydroxyls; the remainder of the aromatic ring residues are in translucent green.

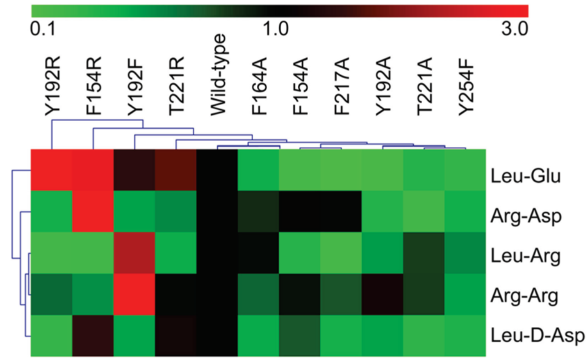

FIGURE 7: Probe labeling analysis of mNit2 point mutants. Labeling experiments were carried out in triplicate with 2.5 μM protein and 20 μM probe for 1 h at room temperature and then analyzed by SDS−PAGE, and the band intensities were quantified. The plot represents hierarchical clustering analysis of average labeling intensities ($n = 3$), where intensities have been normalized to the intensity of the wild-type (WT) protein and reported as linear fold change (where red represents stronger and green represents weaker labeling than wild type).

essentially negligible (data not shown). The labeling pattern of each mutant with a subset of five α-CA probes was measured and compared to the labeling pattern of wild-type mNit2 (Figure 7).

Several tested mutants show an overall loss or decrease of labeling for both positively and negatively charged probes. Alanine mutants of the residues that make up the aromatic "ring", including the Phe154Ala, Phe164Ala, Phe192Ala, and Phe217Ala, fell into this category (Figure 7), consistent with a model where both types of probes interact with these aromatic residues. These labeling patterns could also reflect a general role for the aromatic ring in stabilizing the active site structure of mNit2. Similarly, the Tyr254Phe mutant showed a decrease in labeling with all probes, although this decrease was more substantial for the Glu- and Asp-containing probes (5-fold) than for the Arg-containing probes (2-fold), consistent with the modeling analysis, which pointed to the hydroxyl of Tyr254 as a potential hydrogen bond donor and stabilizer of negatively charged probes.

Other mutants showed differential effects on specific subsets of the probe library. The Tyr192Arg mutant, for example, displayed a striking (>6-fold) increase in labeling with the Leu-Glu-α-CA probe but decreased labeling with all other probes. The Arg mutant, which introduces a positively charged residue into the active site, could form a salt bridge to the Glu probe, increasing labeling while, at the same time, disrupting the aromatic interactions apparently needed for binding of Arg-containing probes and potentially introducing a repulsive charge interaction. The selectivity of this mutant for Glu over Asp in the probe X_2 position may indicate that the carbon chain of the Asp residue is not long enough to interact positively with Tyr192Arg. The Thr221Arg mutant also showed a selective increase in labeling with the Leu-Glu- and Leu-D-Asp-α-CA probes. This residue is postulated to lie underneath the aromatic patch and is a potential hydrogen bond donor that may contribute to stabilization of negatively charged probes. The Tyr192Phe mutant displayed an intriguing labeling pattern, exhibiting a strong increase in labeling with Leu-Arg- and Arg-Arg-α-CA probes. The enhanced labeling with positively charged probes could be explained by an improvement in aromatic π-stacking in the largely hydrophobic aromatic patch resulting from deletion of the tyrosine hydroxyl group.

Finally, the Phe154Arg mutant was found to show dramatically increased reactivity with the Leu-Glu- and Arg-Asp- (and, to a lesser extent, Leu-D-Asp-) α-CA probes but diminished labeling of the Leu-Arg- and Arg-Arg-α-CA agents (Figure 7). This is consistent with the docking results, which implicated a key cation−π interaction with Arg-containing probes. The mutation of Phe154 to Arg would ablate this interaction and at the same time introduce charge−charge repulsion to further impair reactivity with X-Arg-α-CA probes (and conversely improve reactivity with X-Glu/Asp-α-CA probes). Intriguingly, Phe154 is a cysteine in Nit1 (Figure 1), which does not label strongly with X-Arg-α-CA probes (Figure 2). Examination of sequences for Nit1 and Nit2 across species reveals that this Phe154 to Cys mutation is completely conserved for all vertebrate Nit1/Nit2 enzymes, including those in birds and ruminants as well as rodents and primates.

Of the other four residues that, along with Phe154, make up the aromatic patch, only two are conserved in Nit1 (Tyr192 and Phe217), while the others are nonaromatic residues. Thus, Phe164 and Tyr254 may, along with Phe154, contribute to the selective labeling of Nit2 over Nit1 with X-Arg-type α-CA probes and help to establish differences in molecular recognition for Nit1 and Nit2. Tyr254, in fact, aligns to an Arg in mNit1, which shows strong labeling with negatively charged probes. Finally, Phe164 in the mouse Nit2 protein aligns to Tyr142 in a nitrilase enzyme from *Rhodococcus rhodochrous*, and this residue has recently been described to be a key determinant of substrate recognition (*31*).

Taken together, these data suggest that probe binding in mNit2 is driven by a select group of residues, including those that make up the aromatic patch, particularly Phe154 and Tyr192. The conservation of these residues across the vertebrate Nit2 family suggests that their role in molecular recognition could potentially extend to substrate binding and selectivity. Furthermore, the conversion of Phe154 to cysteine in Nit1 enzymes points to differences in the molecular recognition of homologous subclasses of nitrilases in the mammalian proteome that we hypothesize may reflect distinct biochemical functions.

Implications for Probe and Inhibitor Design. Our analysis of factors that determine probe−protein interactions in nitrilase enzymes also has broader implications for the design of covalent activity-based probes and inhibitors. In the field of activity-based proteomics, there is a need for both class-wide probes to broadly profile many members of large enzyme families (*32*−*34*) and target-specific probes to allow the monitoring and potential inhibition of individual enzymes (*35, 36*). As large-scale initiatives continue to solve structures of diverse protein families at ever-increasing rates, the potential exists for the rational, structure-guided design of activity-based probes with tailored reactivity profiles. As demonstrated here, the binding elements of activity-based probes play an important role in defining their target profile and can thus be used as a key factor to tune the selectivity of probes bearing a common electrophile.

ACKNOWLEDGMENT

The authors thank Justin Carlson for initial assistance with the crystallography, BinQing Wei for assistance with the loop-shifted model of mNit2, and the Cravatt, Stevens, and Olson laboratories for helpful discussion.

SUPPORTING INFORMATION AVAILABLE

A list of primers used for preparation of Nit2 mutants and Figures S1−S4. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Bork, P., and Koonin, E. V. (1994) A new family of carbon-nitrogen hydrolases. *Protein Sci. 3*, 1344–1346.
2. Brenner, C. (2002) Catalysis in the nitrilase superfamily. *Curr. Opin. Struct. Biol. 12*, 775–782.
3. Pace, H. C., and Brenner, C. (2001) The nitrilase superfamily: classification, structure and function. *Genome Biol. 2*, REVIEWS0001.
4. Novo, C., Farnaud, S., Tata, R., Clemente, A., and Brown, P. R. (2002) 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. 365*, 731–738.
5. Maras, B., Barra, D., Dupre, S., and Pitari, G. (1999) Is pantetheinase the actual identity of mouse and human vanin-1 proteins? *FEBS Lett. 461*, 149–152.
6. Pitari, G., Malergue, F., Martin, F., Philippe, J. M., Massucci, M. T., Chabret, C., Maras, B., Dupre, S., Naquet, P., and Galland, F. (2000) Pantetheinase activity of membrane-bound Vanin-1: lack of free cysteamine in tissues of Vanin-1 deficient mice. *FEBS Lett. 483*, 149–154.
7. Bieganowski, P., and Brenner, C. (2003) The reported human NADsyn2 is ammonia-dependent NAD synthetase from a pseudomonad. *J. Biol. Chem. 278*, 33056–33059.
8. Bieganowski, P., Pace, H. C., and Brenner, C. (2003) Eukaryotic NAD^+ synthetase Qns1 contains an essential, obligate intramolecular thiol glutamine amidotransferase domain related to nitrilase. *J. Biol. Chem. 278*, 33049–33055.
9. van Kuilenburg, A. B., Meinsma, R., Beke, E., Assmann, B., Ribes, A., Lorente, I., Busch, R., Mayatepek, E., Abeling, N. G., van Cruchten, A., Stroomer, A. E., van Lenthe, H., Zoetekouw, L., Kulik, W., Hoffmann, G. F., Voit, T., Wevers, R. A., Rutsch, F., and van Gennip, A. H. (2004) beta-Ureidopropionase deficiency: an inborn error of pyrimidine degradation associated with neurological abnormalities. *Hum. Mol. Genet. 13*, 2793–2801.
10. Semba, S., Han, S. Y., Qin, H. R., McCorkell, K. A., Iliopoulos, D., Pekarsky, Y., Druck, T., Trapasso, F., Croce, C. M., and Huebner, K. (2006) Biological functions of mammalian Nit1, the counterpart of the invertebrate NitFhit Rosetta stone protein, a possible tumor suppressor. *J. Biol. Chem. 281*, 28244–28253.
11. Lin, C. H., Chung, M. Y., Chen, W. B., and Chien, C. H. (2007) Growth inhibitory effect of the human NIT2 gene and its allelic imbalance in cancers. *FEBS J. 274*, 2946–2956.
12. Cravatt, B. F., Wright, A. T., and Kozarich, J. W. (2008) Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem. 77*, 383–414.

13. Evans, M. J., and Cravatt, B. F. (2006) Mechanism-based profiling of enzyme families. *Chem. Rev. 106*, 3279–3301.
14. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307–326.
15. Read, R. J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 57*, 1373–1382.
16. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 53*, 240–255.
17. The CCP4 suite: programs for protein crystallography (1994) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 760−763.
18. Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 62*, 439–450.
19. Morris, G. M., Goodsell, D. S., Huey, R., and Olson, A. J. (1996) Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des. 10*, 293–304.
20. Goodsell, D. S., Morris, G. M., and Olson, A. J. (1996) Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit. 9*, 1–5.
21. Barglow, K. T., and Cravatt, B. F. (2006) Substrate mimicry in an activity-based probe that targets the nitrilase family of enzymes. *Angew. Chem., Int. Ed. Engl. 45*, 7408–7411.
22. Barglow, K. T., and Cravatt, B. F. (2004) Discovering disease-associated enzymes by proteome reactivity profiling. *Chem. Biol. 11*, 1523–1531.
23. Pace, H. C., Hodawadekar, S. C., Draganescu, A., Huang, J., Bieganowski, P., Pekarsky, Y., Croce, C. M., and Brenner, C. (2000) Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers. *Curr. Biol. 10*, 907–917.
24. Sakai, N., Tajika, Y., Yao, M., Watanabe, N., and Tanaka, I. (2004) Crystal structure of hypothetical protein PH0642 from Pyrococcus horikoshii at 1.6Å resolution. *Proteins 57*, 869–873.
25. Wang, W. C., Hsu, W. H., Chien, F. T., and Chen, C. Y. (2001) 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. 306*, 251–261.
26. Lundgren, S., Lohkamp, B., Andersen, B., Piskur, J., and Dobritzsch, D. (2008) The crystal structure of beta-alanine synthase from Drosophila melanogaster reveals a homooctameric helical turn-like assembly. *J. Mol. Biol. 377*, 1544–1559.
27. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol. 233*, 123–138.
28. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol. 247*, 536–540.
29. Kumaran, D., Eswaramoorthy, S., Gerchman, S. E., Kycia, H., Studier, F. W., and Swaminathan, S. (2003) Crystal structure of a putative CN hydrolase from yeast. *Proteins 52*, 283–291.
30. Ariel, N., Ordentlich, A., Barak, D., Bino, T., Velan, B., and Shafferman, A. (1998) The "aromatic patch" of three proximal residues in the human acetylcholinesterase active centre allows for versatile interaction modes with inhibitors. *Biochem. J. 335*, 95–102.
31. Yeom, S. J., Kim, H. J., Lee, J. K., Kim, D. E., and Oh, D. K. (2008) A determinant residue of substrate specificity in nitrilase from Rhodococcus rhodochrous ATCC 33278 for aliphatic and aromatic nitriles. *Biochem. J.* (in press).
32. Greenbaum, D., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradszky, K. F., and Bogyo, M. (2002) Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics 1*, 60–68.
33. Liu, Y., Patricelli, M. P., and Cravatt, B. F. (1999) Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. U.S.A. 96*, 14694–14699.
34. Sieber, S. A., Niessen, S., Hoover, H. S., and Cravatt, B. F. (2006) Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nat. Chem. Biol. 2*, 274–281.
35. Arastu-Kapur, S., Ponder, E. L., Fonovic, U. P., Yeoh, S., Yuan, F., Fonovic, M., Grainger, M., Phillips, C. I., Powers, J. C., and Bogyo, M. (2008) Identification of proteases that regulate erythrocyte rupture by the malaria parasite Plasmodium falciparum. *Nat. Chem. Biol. 4*, 203–213.
36. Chiang, K. P., Niessen, S., Saghatelian, A., and Cravatt, B. F. (2006) An enzyme that regulates ether lipid signaling pathways in cancer annotated by multidimensional profiling. *Chem. Biol. 13*, 1041–1050.
37. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res. 31*, 3497–3500.
38. Sanner, M. F. (1999) Python: a programming language for software integration and development. *J. Mol. Graphics Model. 17*, 57–61.

BI801786Y