Biochemistry

pubs.acs.org/biochemistry Article

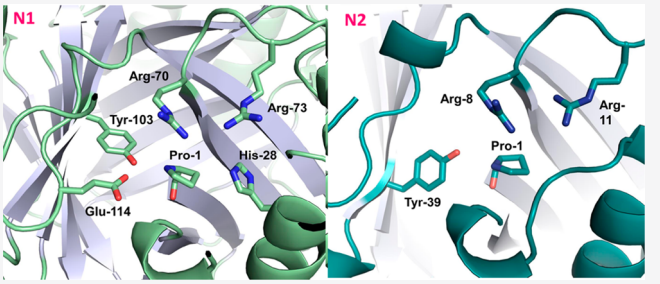
Kinetic and Structural Analysis of Two Linkers in the Tautomerase Superfamily: Analysis and Implications

Bert-Jan Baas, Brenda P. Medellin, Jake A. LeVieux, Kaci Erwin, Emily B. Lancaster, William H. Johnson, Jr., Tamer S. Kaoud, R. Yvette Moreno, Marieke de Ruijter, Patricia C. Babbitt, Yan Jessie Zhang, and Christian P. Whitman*

Cite This: *Biochemistry* 2021, 60, 1776−1786 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The tautomerase superfamily (TSF) is a collection of enzymes and proteins that share a simple $\beta-\alpha-\beta$ structural scaffold. Most members are constructed from a single-core $\beta-\alpha-\beta$ motif or two consecutively fused $\beta-\alpha-\beta$ motifs in which the N-terminal proline (Pro-1) plays a key and unusual role as a catalytic residue. The cumulative evidence suggests that a gene fusion event took place in the evolution of the TSF followed by duplication (of the newly fused gene) to result in the diversification of activity that is seen today. Analysis of the sequence similarity network (SSN) for the TSF identified several linking proteins ("linkers") whose similarity links subgroups of these contemporary proteins that might hold clues about structure−function relationship changes accompanying the emergence of new activities. A previously uncharacterized pair of linkers (designated N1 and N2) was identified in the SSN that connected the 4-oxalocrotonate tautomerase (4-OT) and *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD) subgroups. N1, in the *cis*-CaaD subgroup, has the full complement of active site residues for *cis*-CaaD activity, whereas N2, in the 4-OT subgroup, lacks a key arginine (Arg-39) for canonical 4-OT activity. Kinetic characterization and nuclear magnetic resonance analysis show that N1 has activities observed for other characterized members of the *cis*-CaaD subgroup with varying degrees of efficiencies. N2 is a modest 4-OT but shows enhanced hydratase activity using allene and acetylene compounds, which might be due to the presence of Arg-8 along with Arg-11. Crystallographic analysis provides a structural context for these observations.



Linkers are proteins that have sequence similarity with other proteins in different subgroups in a sequence similarity network (SSN).[1] Although these linking sequences are contemporary ones, they can provide hints about intermediate or transitional features resulting in the divergence from a common ancestor. Characterization of the proteins coded by linking sequences reveals changes in structure and function that might have occurred in the evolution of proteins in one cluster from the proteins in another cluster. The potential value of examining relationships among such contemporary linkers was demonstrated in the analysis of the SSN for the tautomerase superfamily (TSF).[1]

Members of the tautomerase superfamily are constructed from a $\beta-\alpha-\beta$ structural unit that is found as a single unit (58−84 amino acids) or two consecutively fused units (110−150 amino acids).[1−3] The experimentally characterized members have a catalytic amino-terminal proline (Pro-1) that functions as a general base or acid, depending on the pK_a of the prolyl nitrogen.[3−5] This simple motif has been repurposed by nature many times over to catalyze a variety of functionally diverse enzymatic reactions, including tautomerization,[6] hydrolytic dehalogenation,[7,8] hydration,[7] decarboxylation,[9] aldol condensation,[10] and oxygenation.[11] Members also have biological activities where the best characterized one is the macrophage migration inhibitory factor (MIF) that functions as a pro-inflammatory cytokine.[12]

A global analysis of the TSF collected more than 11000 nonredundant sequences of proteins found throughout the biosphere.[1] These sequences were sorted into five subgroups [4-oxalocrotonate tautomerase, 4-OT, 5-(carboxymethyl)-2-hydroxymuconate isomerase, CHMI, malonate semialdehyde decarboxylase, MSAD, *cis*-3-chloroacrylic acid dehalogenase, *cis*-CaaD, and MIF] in the SSN that aligned with the characterized activities of the title enzymes.[1] Although the SSN showed several linkers connecting the subgroups, four

Received: March 25, 2021
Revised: May 2, 2021
Published: May 21, 2021

https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786

linker enzymes along a similarity path between the 4-OT and *cis*-CaaD subgroups were chosen for initial characterization as a proof of principal (Figure 1).[1,13] The structure-based comparison of these linkers identified transitions between them that were paralleled in the kinetic analysis and suggested how the *cis*-CaaD-like enzymes might evolve from a 4-OT-like ancestor.[1,13]

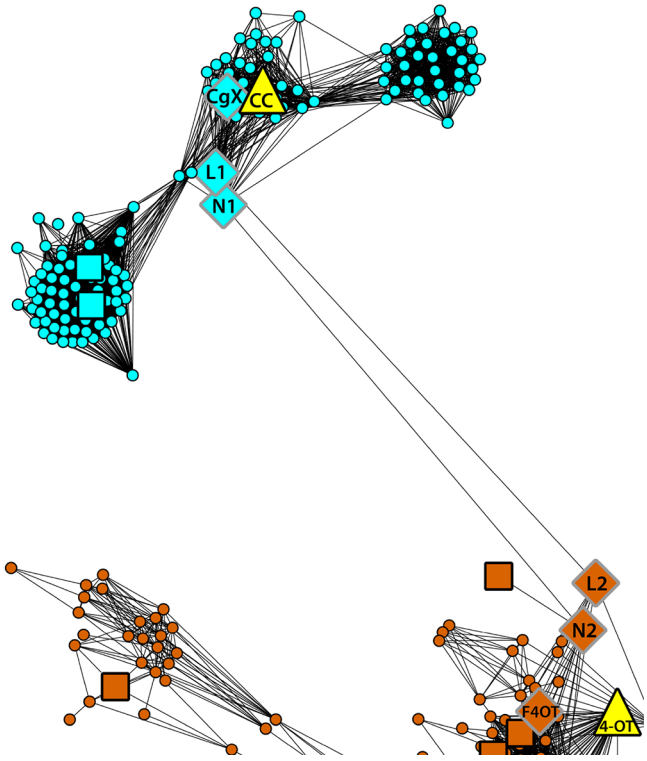


Figure 1. Close-up of the 50% representative sequence similarity network of the TSF showing representative nodes that connect the 4-OT and *cis*-CaaD subgroups (modified from ref 1). The representative nodes of the 4-OT subgroup are colored brown (bottom), where 4-OT is designated by the yellow triangle, and those in the *cis*-CaaD subgroup are colored cyan, where *cis*-CaaD (CC) is designated by the yellow triangle (top). The other nodes that contain biochemically characterized enzymes are labeled using the abbreviations below. Large unlabeled squares designate representative nodes that contain structurally characterized protein(s). Two previously uncharacterized linker nodes, N1 and N2, are the targets of this study. The structural changes associated with a series of hops along the similarity pathway between the 4-OT and *cis*-CaaD subgroups are summarized below and described in detail elsewhere.[1] Abbreviations: F4OT, fused 4-OT; L2, linker 2; L1, linker 1; CgX, a *cis*-CaaD homologue.

In a stepwise series of "hops" along the similarity pathway, the first hop connected the short 4-OT to a fused 4-OT (F4OT), which represents the fusion of two $\beta-\alpha-\beta$ structural units.[1] The next hop connected the fused 4-OT to linker 2 (also a fused 4-OT) designated L2 (in Figure 1). Both linker 2 and fused 4-OT showed structural and kinetic similarities with other members of the 4-OT subgroup, including the conservation of key active site residues and robust activities using 2-hydroxymuconate (2-HM) and phenylenolpyruvate (PP), di- and monocarboxylate substrates, respectively.[1,4−6] The hop from linker 2 to linker 1 (designed L1 in the *cis*-CaaD subgroup) introduced active site features associated with *cis*-CaaD-like properties and a loss of those associated with the 4-OT-like subgroup. Notably, tautomerase activity with the dicarboxylate substrate was lost, but that with the monocarboxylate substrate was retained. The final series of hops (linker 1 to CgX, a *cis*-CaaD homologue that functions as an inefficient *cis*-CaaD, and CgX to *cis*-CaaD designated CC in Figure 1) complete the transition.[1]

Besides the linker 1 and linker 2 representative nodes linking the 4-OT and *cis*-CaaD subgroups, the SSN shows a second edge that also connects these subgroups. The representative nodes for this second edge are designated N1 and N2 (Figure 1). Hence, N1 (in the *cis*-CaaD subgroup) and N2 (in the 4-OT subgroup) were chosen for sequence, kinetic, mechanistic, and structural analysis. N1 has the key components of the catalytic machinery (Pro-1, His-28, Arg-70, Arg-73, Tyr-103, and Glu-114) found in *cis*-CaaD.[3,8] Crystallographic analysis of N1 reveals that the active site mirrors those of two other characterized *cis*-CaaD subgroup members along with *cis*-CaaD. Not surprisingly, kinetic analysis and NMR analysis show that N1 functions similarly with pronounced hydration activity, but no 4-OT activity. Like the others, tautomerase activity is exhibited with a monocarboxylate substrate, phenylenolpyruvate. N2 lacks Arg-39, a key residue for 4-OT activity, although Pro-1 and Arg-11 are present.[1,3,14] Kinetic analysis shows that 4-OT activity is diminished in comparison to those of other characterized members of the 4-OT subgroup. Covalent modification of the Pro-1 nitrogen by 2-oxo-3-pentynoate (2-OP) reflects a low pK_a value, which is also a property shared with other 4-OT subgroup members.[3−5] However, unlike other characterized members of the subgroup, hydratase activity is enhanced in N2, which might be due to the presence of Arg-8 along with Arg-11. Crystallographic analysis of N2 provides a structural context. The weight of the experimental evidence highlights the importance of Arg-8 in the evolution of a 4-OT-like protein to *cis*-CaaD.[15]

EXPERIMENTAL PROCEDURES

Materials. Chemicals, biochemicals, buffers, and solvents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO), Fisher Scientific, Inc. (Pittsburgh, PA), EMD Millipore, Inc. (Billerica, MA), or Bio-Rad Laboratories, Inc. (Hercules, CA), unless otherwise noted below. Phenylenolpyruvate (PP) was purchased from Fluka Chemical Corp. (Milwaukee, WI) as the crystalline free acid, which exists exclusively in the enol form. Propiolic acid (98%) and 2-butynoate were purchased from Sigma-Aldrich and purified further prior to use by distillation and recrystallization, respectively. 2-Hydroxymuconate (2-HM),[6] 2-oxo-3-pentynoate (2-OP),[14] and 2,3-butadienoate[16] were synthesized according to the indicated references. The diethylaminoethyl (DEAE) Sepharose fast flow resin and the PD-10 Sephadex G-25 M buffer-exchange column were obtained from GE Healthcare Bio-Sciences (Pittsburgh, PA). The Econo-Column chromatography columns and the Bio-Gel P60 gel beads used for size-exclusion chromatography were obtained from Bio-Rad Laboratories, Inc. The Spectra/Chrom LC column was obtained from Spectrum Laboratories, Inc. (Rancho Dominguez, CA). The Amicon stirred cell concentrators and the ultrafiltration membranes (10000 Da, MW cutoff) were purchased from EMD Millipore, Inc. InstantBlue for staining SDS−PAGE gels was obtained from C.B.S. Scientific Company, Inc. (Del Mar, CA).

Bacterial Strains, Plasmids, and Growth Conditions. *Escherichia coli* strain BL21-Gold(DE3) was obtained from

https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786

Table 1. Crystallographic Data Collection and Refinement Statistics

	N1	N2
	Data Collection	
space group	$P2_12_12_1$	$R32$
cell dimensions		
a, *b*, *c* (Å)	41.36, 87.82, 120.71	78.84, 78.84, 182.95
α, β, γ (deg)	90.00, 90.00, 90.00	90.00, 90.00, 120.00
resolution (Å)	50.00−2.45 (2.49−2.45)[a]	50.00−1.65 (1.68−1.65)[a]
wavelength (Å)	0.97648	0.97648
R_{sym}/R_{pim}	0.127 (0.649)[a]/0.064 (0.368)[a]	0.057 (0.616)[a]/0.018 (0.290)[a]
$CC_{1/2}$[b]	0.886 (0.645)[a]	0.962 (0.767)[a]
I/σ	12.2 (1.6)[a]	50.2 (1.9)[a]
completeness (%)	90.4 (82.1)[a]	99.6 (96.7)[a]
redundancy	4.5 (3.9)[a]	10.6 (5.3)[a]
	Refinement	
resolution (Å)	31.80−2.45 (2.53−2.45)[a]	38.00−1.65 (1.71−1.65)[a]
no. of unique reflections	15278 (1336)[a]	26626 (2544)[a]
R_{work}	0.1954 (0.2439)[a]	0.1752 (0.2436)[a]
R_{free}[c]	0.2633 (0.3320)[a]	0.1865 (0.2668)[a]
no. of atoms		
protein	3396	984
ligand/ion	0	0
water	69	91
B-factor (Å^2)		
protein	33.9	25.4
ligand/ion	0	0
water	35.8	34.4
root-mean-square deviation		
bond lengths (Å)	0.010	0.006
bond angles (deg)	1.25	0.73
Ramachandran plot (%)		
favored	97.64	100
allowed	1.65	0
outliers[d]	0.71	0
Molprobity score[e]	1.7[e]/98th percentile[f]	0.89[e]/100th percentile[f]

[a]Values for the corresponding parameters in the outermost shell in parentheses. [b]$CC_{1/2}$ is the Pearson correlation coefficient for a random half of the data; the two numbers represent the lowest- and highest-resolution shell, respectively. [c]R_{free} is the R_{work} calculated for ∼10% of the reflections randomly selected and omitted from refinement. [d]There are three Ramachandran outliers in N1, which correspond to aspartate residues (Asp-13 chain A, Asp-13 chain B, and Asp-13 chain C) all with strong electron density. [e]The MolProbity score is calculated by combining the Clashscore with rotamer and Ramachandran percentage and scaled on the basis of X-ray resolution. [f]The percentage is calculated with the 100th percentile as the best and the 0th percentile as the worst among structures of comparable resolution.

Agilent Technologies (Santa Clara, CA). The genes encoding N2 (4-OT subgroup) from *Gammaproteobacteria bacterium* SG8_31 (UniProt entry A0A0S8FF56) and N1 (*cis*-CaaD subgroup) from *Corynebacterium halotolerans* YIM 70093 (UniProt entry M1NLA4) were codon-optimized for expression in *E. coli*, synthesized, and cloned into expression vector pJ411 by ATUM (Newark, CA). Cells were grown overnight (∼16 h) at 37 °C in Luria-Bertani (LB) medium, supplemented with 25 mM Na_2HPO_4 and 25 mM KH_2PO_4, buffered to pH ∼6.75, 5 mM Na_2SO_4, 2 mM $MgSO_4$, and kanamycin (Kn, 30 μg/mL).

General Methods. Steady-state kinetic assays were performed on an Agilent 8453 diode-array spectrophotometer at 22 °C. Nonlinear regression data analysis was performed using Grafit (Erithacus Software Ltd., Staines, U.K.). Protein concentrations were determined by the Waddell method.[17] SDS−PAGE was carried out on denaturing gels containing 12% polyacrylamide.[18] Mass spectrometry was conducted by The University of Texas at Austin Proteomics Facility using an Optimize Technologies protein trap in line with a Thermo Orbitrap Fusion mass spectrometer. The data were collected using the ion trap detector or nanoflow liquid chromatograph and deconvoluted using Thermo Protein Deconvolution software. Samples were prepared as described elsewhere.[7]

Discovery of N2 and N1. The SSN for the TSF was constructed as described elsewhere.[1] N2 and N1 are representative nodes for a second representative edge connecting the 4-OT and *cis*-CaaD subgroups. N1 was included in the phylogenetic tree of the 4-OT and *cis*-CaaD linkers, and both were included in the linker control experiments described elsewhere.[1]

Expression and Purification of N2 and N1. N2 and N1 were expressed and purified as follows. Fresh BL21Gold(DE3) transformants containing expression vector pJ411-N2 or pJ411-N1 were collected from a LB/Kn plate and used to inoculate a preculture (LB/Kn medium, 30 μg/mL, 10 mL). After being grown for ∼6 h at 37 °C, the preculture was used to inoculate 1 L of the LB supplemented medium. After being grown overnight at 37 °C while being shaken (250 rpm), the culture was incubated at 22 °C for 1 h and subsequently made 0.5 mM in isopropyl β-D-thiogalactoside to induce expression. After being incubated for 5 h at 22 °C, the cells were harvested

https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786

by centrifugation (10 min, 2500*g*) and resuspended in 10 mM Na_2HPO_4 buffer (pH 7.3, buffer A) to yield a total volume of ∼15 mL. The cell suspension was chilled on ice for 15 min, after which the cells were disrupted by sonication for 3 min at 50% duty cycle/50% output using a Vibra-Cell sonicator model VC250B (Sonics & Materials, Inc., Danbury, CT). Unbroken cells and debris were removed by centrifugation (45 min, 18000*g*). The clear supernatant was subsequently applied to a DEAE Sepharose column (∼15 mL bed volume), which had previously been equilibrated in buffer A. The column was washed with buffer A (3 × 15 mL), and retained proteins were subsequently eluted with a linear gradient of buffer A made 500 mM in NaCl. Typically, both N2 and N1 eluted around 150−200 mM NaCl.

The fractions from the DEAE Sepharose column containing the target protein were pooled and made 1 M in $(NH_4)_2SO_4$ by the addition of the appropriate volume of a 3.2 M $(NH_4)_2SO_4$ stock solution in buffer A. The sample was incubated on ice for 2 h, after which precipitates were removed by centrifugation (20 min, 18000*g*). The clear supernatant was then loaded onto a Phenyl Sepharose column that had previously been equilibrated in buffer A, made 1 M in $(NH_4)_2SO_4$. Unbound proteins were removed by washing with 3 column volumes of buffer A, made 1 M in $(NH_4)_2SO_4$. Retained proteins were subsequently eluted with a linear gradient of buffer A. N2 and N1 typically eluted around 600 and 300 mM $(NH_4)_2SO_4$, respectively.

The fractions containing the target protein were pooled and concentrated to a volume of ∼3 mL using an Amicon concentrator equipped with a 3 kDa cutoff membrane filter. This sample was subsequently applied to a size-exclusion column (Bio-Gel P60 beads, 500 mL bed volume), which had previously been equilibrated in buffer A, and proteins were eluted with buffer A (0.5 mL/min) at 22 °C. The protein-containing fractions (5 mL) were analyzed by SDS−PAGE, and those containing near-homogeneous target protein were pooled and concentrated to ∼20 mg/mL using an Amicon concentrator. Aliquots were flash-frozen in liquid nitrogen and stored at −80 °C until further use.

Light Scattering Experiments. The light scattering experiments were carried out at 25 °C on a DAWN HELEOS-II multiangle light scattering photometer with an Optilab T-rex refractometer detector and a Wyatt QELS dynamic light scattering detector (MALS-QELS system) (Wyatt Technology, Santa Barbara, CA), as described elsewhere.[19] Samples were delivered through a TSK-GEL $G300PW_{XL}$ size-exclusion column (7.8 mm × 300 mm, pore size of 300 Å) (Tosoh Bioscience LLC, King of Prussia, PA) connected to a Shimadzu LC-20AD HLPC system (model WTC-030S5). The solvent was 10 mM NaH_2PO_4 buffer (pH 7.3), and the flow rate was 0.4 mL/min. Molar mass moments in grams per mole were determined with 20 μL samples at concentrations of 25 and 2.5 μM. Molar masses, peak concentrations, and hydrodynamic radii were determined with Astra 6 software (Wyatt Technology).

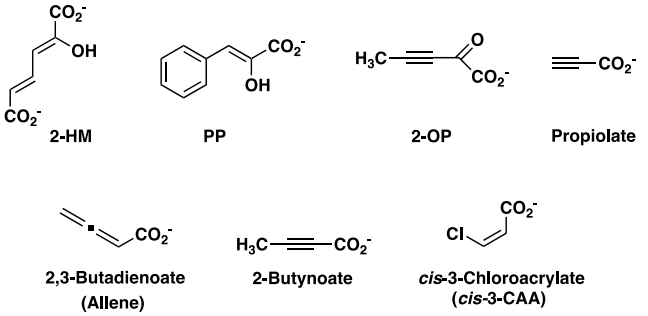
Crystallization. Initial crystallization conditions for N1 and N2 were identified using sparse-matrix screening with a Phoenix crystallization robotic system (Art Robbins Instruments). After systematic optimization of the hits, reproducible N1 crystals grew under the condition consisting of 100 mM Tris buffer (pH 7.0), 35% PEG 3550, and 200 mM Li_2SO_4 with protein at 70 mg/mL at 4 °C. N2 crystals grew in vapor diffusion sitting drops under conditions of 40% 2-methyl-2,4-pentanediol (MPD), 4% PEG 8000, and 1% sodium cacodylate with protein at 20 mg/mL mixed with mother liquor in a 1:1 ratio at 4 °C. All crystals were cryoprotected with mother liquor supplemented with 30% glycerol and then vitrified in liquid nitrogen.

Data Collection, Processing, Structure Determination, and Refinement. X-ray diffraction data for N1 and N2 crystals were collected at Advanced Photon Source (APS) beamline 23-ID-B. Diffraction data were indexed, integrated, and scaled using HKL2000. The phase was determined for N1 and N2 using PHASER from the PHENIX software suite.[20] Both structures were determined by molecular replacement using as search models Cg10062 [Protein Data Bank (PDB) entry 3N4G] for N1 and the 4-OT homologue TomN (PDB entry 3RY0) for N2.[21] Structure refinement was carried out using PHENIX Refine. The TLS parameter was included in the refinement of all structures. The final structures were evaluated during and after refinement using Molprobity.[22] Data collection and refinement statistics for all structures are summarized in Table 1. Figures were prepared using PyMol (The PyMOL Molecular Graphics system, version 1.8, Schrodinger, LLC).[23]

Steady-State Kinetics. The enol−keto tautomerization of 2-hydroxymuconate (2-HM) and phenylenolpyruvate (PP) was assayed at 22 °C in 20 mM Na_2HPO_4 buffer (pH 7.3) using various enzyme concentrations (0.6−2.25 μM) depending on the catalytic efficiency of each enzyme.[21,24] Stock solutions of 2-HM (50 mM) and PP (100 mM) were prepared by dissolving the appropriate amount of the free acid in absolute ethanol. The ketonization of 2-HM by N2 was monitored by following the depletion of the enol form at 330 nm (ε = 934 M^{-1} cm^{-1}) using substrate concentrations ranging from 26 to 300 μM. The ketonization of PP by both N2 and N1 was monitored by following the depletion of the enol form at 306 nm (ε = 3675 M^{-1} cm^{-1}) using substrate concentrations ranging from 170 to 1030 μM. An appropriate quantity of enzyme [from an ∼20 mg/mL stock solution in 10 mM Na_2HPO_4 buffer (pH 7.3)] was diluted in sodium phosphate buffer (780 μL in a 2 mm path length quartz cuvette), and the assay was initiated by the addition of a small aliquot (3−25 μL) of the substrate (either 2-HM or PP) from a stock solution using a positive displacement pipet. At all substrate concentrations, the non-enzymatic rate was subtracted from the enzymatic rate of ketonization.

The hydration of propiolate, 2,3-butadienoate, and 2-oxo-3-pentynoate (2-OP) (Scheme 1) by both N2 and N1, as well as the hydrolytic dehalogenation of *cis*-3-chloroacrylate by N1, was monitored according to procedures published elsewhere,[25−27] with the following modifications. The hydration

Scheme 1. Compounds Used in This Work



https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786

of 2,3-butadienoate was monitored at 230 nm (ε = 790 M^{-1} cm^{-1}) and 250 nm (ε = 250 M^{-1} cm^{-1}) for N1 and N2, respectively.

NMR Analysis. The protocol for monitoring the reaction of N1 or N2 with 2,3-butadienoate (allene), 2-butynoate, and propiolate by ^{1}H NMR spectroscopy followed that described elsewhere.[25,26] Accordingly, an amount of each compound (4 mg) was added to 100 mM Na_2HPO_4 buffer (pH ∼9.2) containing 30 μL of dimethyl sulfoxide-d_6 (DMSO-d_6) (final total volume of 600 μL). The final pH of the solution was adjusted to 8.0. Subsequently, an aliquot of N1 or N2 was added to the mixture such that the final amount of protein was 0.5 mg. For overnight NMR runs (i.e., 18 h), product formation was determined using a Bruker AVANCE III 500 MHz spectrometer. For shorter NMR runs, product formation was monitored every 3 min for 10 intervals on a Varian (Palo Alto, CA) DirectDrive 600 MHz spectrometer. The initial spectrum was recorded 3 min after mixing. DMSO-d_6 was used as the lock signal and for standardization of the chemical shifts (at 2.49 ppm). The chemical shifts for the products are reported elsewhere.[7,26,27] The approximate amount of product in the mixture was determined by integration, as reported elsewhere.[26] NMR signals were analyzed using SpinWorks version 3.1.6 (Copyright 2009 Kirk Marat, University of Manitoba, Winnipeg, MB).

Protein Labeling Studies. The monoisotopic masses of native, unmodified N2 and N1 were determined by MS analysis using protein samples diluted to ∼25 μM in 10 mM NaH_2PO_4 buffer (pH 7.3) from ∼20 mg/mL stock concentrations. Samples to determine the potential covalent modification of N2 and N1 by 2-oxo-3-pentynoate (2-OP) were prepared as follows. A stock solution of 2-OP (60 mM) was prepared by dissolving 4 mg of the compound in 100 mM NaH_2PO_4 buffer (600 μL, pH ∼9) and adjusting the final pH to 7.0. Potential labeling reactions of 2-OP were initiated by the addition of ∼1 mg of protein (diluted from an ∼20 mg/mL stock solution) to 1 mL of 10 mM NaH_2PO_4 buffer (pH 7.3) containing 2-OP (6 mM). After incubation for 1 h at 22 °C, $NaBH_4$ (from a 1 M stock solution in deionized water) was added to a final concentration of 25 mM, and the resulting reaction mixture was incubated at 22 °C for 3 h. Subsequently, the reaction mixtures were buffer-exchanged into 10 mM NaH_2PO_4 buffer (pH 7.3) using a PD-10 column. The elution fractions containing protein were pooled, flash-frozen in liquid N_2, and stored at −80 °C until MS analysis.

RESULTS AND DISCUSSION

Discovery of N2 and N1 and Sequence Analysis. Linker proteins N1 and N2 were discovered in the course of the global analysis of the TSF where their representative nodes were connected through a representative linker edge between the *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD) and 4-oxalocrotonate tautomerase (4-OT) subgroups.[1] In that work, sequences from the representative nodes within the individual subgroups show a connection trajectory in which N1 sequences are connected to sequences comprising linker 1, and to those of other representative nodes of the *cis*-CaaD homologue (CgX), and *cis*-CaaD, but are not connected to those of linker 2 (in the 4-OT subgroup) (Figure 1). Sequences of the N2 representative node are not directly connected to those comprising canonical 4-OT or fused 4-OT representative nodes. Nor are they connected to those of linker 2 or linker 1. Of the characterized proteins, N2 sequences are connected only to those of N1.

Purification of N2 and N1 and MS and Light Scattering Analysis. Both N1 and N2 were overproduced in *E. coli* BL21(DE3) and purified to near homogeneity. The typical yield was ∼40 mg for N1 and ∼100 mg for N2 (per liter of culture). MS analysis of the purified N1 showed a single major signal corresponding to a monoisotopic mass of 16104 Da (calcd 16235 Da) in the reconstructed mass spectrum. MS of the purified N2 showed a single major signal corresponding to a monoisotopic mass of 7010 Da (calcd 7141 Da) in the reconstructed mass spectrum. The mass difference of 131 Da between the theoretical and observed mass indicates that the translation-initiating methionine was post-translationally removed, resulting in the mature and catalytically active proteins (144 amino acids for N1 and 64 amino acids for N2) with an N-terminal proline. Light scattering analysis of N1 showed a single peak (100% mass fraction) with a molar mass of 48700 Da (±3.7%). This species is likely the trimer form of N1 (calcd 48312 Da). Light scattering analysis of N2 showed a single peak (100% mass fraction) with a molar mass of 42720 Da (±3.7%). This species is likely the hexamer form of N2 (calcd 42060 Da).

Sequence and Structural Analyses of N1 and N2. The conserved residues in N2 and N1 were aligned with those of 4-oxalocrotonate tautomerase (4-OT), fused 4-OT, linker 2, linker 1, CgX, and *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD) (Figure 2). The sequence of N1 aligns well with the

Figure 2. Alignment of conserved N1 and N2 sequence motifs with the structure-based alignment of previously characterized 4-OT superfamily members[1] as shown in Figures 3 and 4. The 4-OT subgroup members are colored blue (with conserved residues boxed in blue), and those from the *cis*-CaaD subgroup are colored red (with conserved residues boxed in red). The numbers at the top correspond to the residue numbers in N2, and the numbers at the bottom correspond to those in N1.

conserved residues in the other members of the *cis*-CaaD subgroup (boxed in red). In N2, Arg-8 replaces Leu-8 (and Ile in fused 4-OT and linker 2) but aligns with Arg-70 (using the N1 numbering system) in linker 1, CgX, *cis*-CaaD, and N1 (boxed in red). In addition, Tyr-39 in N2 replaces the conserved Arg (Arg-39 in 4-OT) in the 4-OT subgroup. While detailed sequence analysis (Figure 1 and ref 1) clearly places N2 in the 4-OT subgroup, the motif alignment shown in Figure 2 suggests it also has "*cis*-CaaD-like" properties, particularly at the key position Arg-8 of the 4-OT subgroup (position Arg-70 of the *cis*-CaaD subgroup).

The structure of N1 was determined by X-ray crystallography at a resolution of 2.45 Å. One trimer exists in each asymmetric unit in space group $P2_12_12_1$, with 120° of rotation for each monomer. The crystal structure of N2 diffracts to 1.65 Å in space group $R32$. With two monomers in each asymmetric

https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786

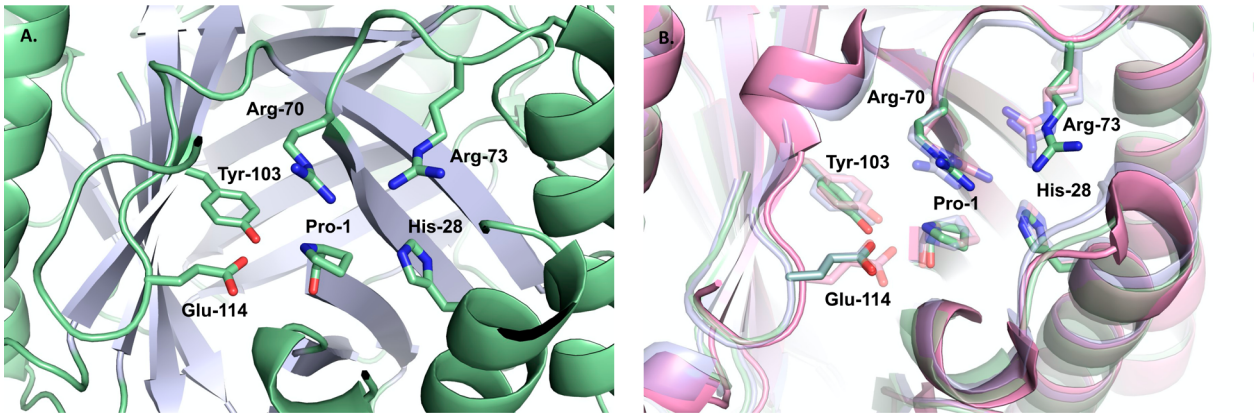


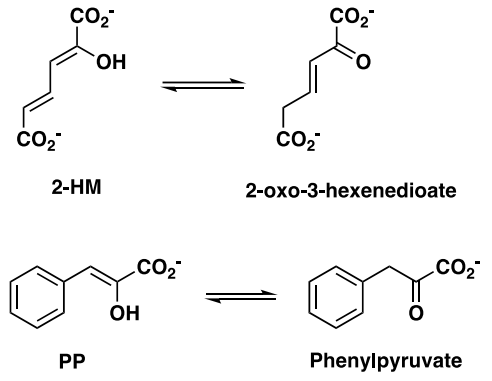
Figure 3. Crystal structures of N1. The overall structures are shown in ribbon diagrams with conserved residues shown as sticks. (A) Overall structure of N1 with putative active site residues labeled. The secondary structure is shown as a ribbon diagram with the β-strand colored light blue and α-helices colored lime green. (B) Superimposition of N1, CgX, and *cis*-CaaD. N1 is colored using the same color scheme as in A; however, CgX is colored lilac, and *cis*-CaaD is colored pink.

unit, crystallographic symmetry generates the functional unit of a homohexamer with six subunits.

N1 is in the *cis*-CaaD subgroup and has the same set of six active set residues (Pro-1, His-28, Arg-70, Arg-73, Tyr-103, and Glu-114) as those found in *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD)[28] and the *cis*-CaaD homologue, CgX[26,29] (Figures 2 and 3A,B) except that linker 1 shows an alanine in place of the tyrosine at position 103. Importantly, the residues are positionally conserved in those structures, and the active sites are nearly superimposable (Figure 3B). On the basis of the structural similarities, it is anticipated that N1 will have activities similar to those of CgX and *cis*-CaaD using similar substrates with varying efficiencies. Note that the two arginine residues are to one side of the catalytic Pro-1.[1]

N2 belongs to the 4-oxalocrotonate tautomerase (4-OT) subgroup. Sequence analysis and structure analysis reveal that the active site configuration for canonical 4-OT activity is only partially conserved, where canonical 4-OT activity refers to that using 2-hydroxymuconate (2-HM), a dicarboxylate substrate (Schemes 1 and 2). In N2, Arg-11 is conserved but

Scheme 2. Enzyme-Catalyzed Tautomerization Reactions



Arg-39 is replaced by a tyrosine (Figures 2 and 4A,B). Moreover, Leu-8 (in 4-OT) is replaced with Arg-8 (Figure 4B). In 4-OT, the nucleophilic Pro-1 is sandwiched between two arginine residues (Arg-11 and Arg-39), but in N2, the two arginine residues are placed to one side of Pro-1 (Figure 4B). This property is a characteristic of *cis*-CaaD subgroup members.[1] In addition, a water molecule is located 2.8 Å from the prolyl nitrogen and 3.2 Å from the hydroxyl group of Tyr-39 and could play a role in hydration.

Kinetic Properties of N1 and N2. N1 and N2 were screened with seven known TSF substrates (Scheme 1) to assess their 4-OT-like and *cis*-CaaD-like properties.[1,21,24] The kinetic parameters are listed in Table 2. Phenylenolpyruvate (PP) and 2-hydroxymuconate (2-HM) assess tautomerase activity (a 4-OT-like property), whereas *cis*-3-chloroacrylate, propiolate, 2,3-butadienoate, 2-butynoate, and 2-OP uncover hydration potential (a *cis*-CaaD property).[1] The previous kinetic analysis of the enzymes linking the 4-OT and *cis*-CaaD subgroups showed a loss of 4-OT activity and an increase in hydration activity that culminates in the *cis*-CaaD-catalyzed hydrolytic dehalogenation.[1] 4-OT, fused 4-OT, and linker 2 (in the 4-OT subgroup) were found to be proficient tautomerases using 2-HM ($k_{cat}/K_m \sim 10^6-10^7$ M^{-1} s^{-1}) or PP ($k_{cat}/K_m \sim 10^5-10^6$ M^{-1} s^{-1}) (Schemes 1 and 2).[1] Linker 1, CgX, and *cis*-CaaD (in the *cis*-CaaD subgroup) show no detectable 4-OT activity with 2-HM, but a modest tautomerase activity using PP ($k_{cat}/K_m \sim 10^3-10^5$ M^{-1} s^{-1}).[1] Analysis of the active sites shows that in the 4-oxalocrotonate tautomerase subgroup members, Pro-1 is positioned between two arginine groups (positionally comparable to Arg-11 and Arg-39 in 4-OT) such that the di- and monocarboxylate substrates can be processed by 4-OT, fused 4-OT, and linker 2 (Figure 4B). However, in linker 1, CgX, and *cis*-CaaD, two arginine groups (positionally comparable to Arg-70 and Arg-73 in *cis*-CaaD) are to one side of Pro-1 so that 2-HM is not processed, but PP is (Figure 3B).

Kinetic analysis of N2 shows both substrates [phenylenolpyruvate (PP) and 2-hydroxymuconate (2-HM)] are processed, but activity with 2-HM is markedly decreased ($k_{cat}/K_m \sim 10^2$ M^{-1} s^{-1}), compared with those of other characterized 4-OT subgroup members ($k_{cat}/K_m \sim 10^6-10^7$ M^{-1} s^{-1}).[1] Notably, in N2, Pro-1 is not between two arginine residues because Arg-39 is replaced with a tyrosine (Figure 4A). Hence, this substitution can account for the decreased activity of N2 with 2-HM. However, like linker 1, CgX, and *cis*-CaaD, two arginines (Arg-8 and Arg-11) are to one side of Pro-1 (a *cis*-CaaD-like property) so that PP activity ($k_{cat}/K_m \sim 10^3$ M^{-1} s^{-1}) is observed but is also reduced compared to that of 4-OT.[1] The active site of N1 is similar to those of linker 1, CgX, and *cis*-CaaD where two arginine groups are to one side

https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786

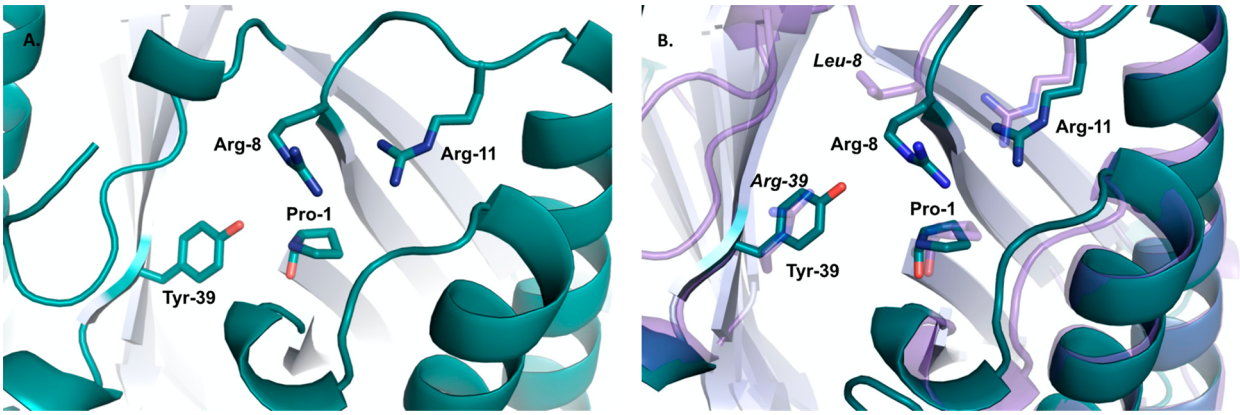


Figure 4. Crystal structures of N2. The overall structures are shown as ribbon diagrams with active site residues shown as sticks. (A) Overall structure of N2 with putative active site residues labeled. The secondary structure is shown as a ribbon diagram with β-strands colored silver and α-helices colored teal. The active site residues are numbered. (B) Superimposition on N2 and 4-OT. N2 is colored in the same color scheme as in panel A, but 4-OT is colored violet with the nonconserved residues labeled in italics.

of Pro-1 (Figure 3B). Hence, the monocarboxylate substrate (PP) is processed ($k_{cat}/K_m \sim 10^4$ M^{-1} s^{-1}), but the dicarboxylate substrate (2-HM) is not.

N1 shows activity with *cis*-3-chloroacrylate ($k_{cat}/K_m \sim 10^2$) and more significant activity with two other substrates, propiolate ($k_{cat}/K_m \sim 10^3$) and 2,3-butadienoate ($k_{cat}/K_m \sim 10^3$) (Scheme 3 and Table 2). The first two compounds are processed to malonate semialdehyde (MSA), and the third compound is converted to acetoacetate (Scheme 3). The reactions of N1 with propiolate and 2,3-butadienoate were examined further by ^{1}H NMR spectroscopy to determine if hydration is followed by decarboxylation. (Decarboxylation cannot be detected by the UV assay.) After 30 min, a reaction mixture of N1 and propiolate consisted of the unreacted substrate (83.9%), MSA (0.3%) and its hydrate (1.2%), and acetaldehyde (6.3%) and its hydrate (8.3%). The amount of decarboxylated product exceeds that expected for the non-enzymatic reaction. After 18 h, a reaction mixture of N1 and 2,3-butadienoate consisted of acetoacetate (97.6%) and acetone (2.4%). The amount of decarboxylated product is comparable to that expected for non-enzymatic activity (after 18 h). Hence, N1 catalyzes hydration/decarboxylation with propiolate, but only hydration with the 2,3-butadienoate.

These activities (hydration/decarboxylation vs hydration) parallel those of CgX, but at a much lower efficiency. For CgX, the reactions with propiolate and 2,3-butadienoate have k_{cat}/K_m values of $\sim 10^5$ (propiolate) and $\sim 10^3$ (2,3-butadienoate) (Table 2). For the allene, both CgX and N1 show high K_m values (780 and 3100 μM, respectively), which could be a function of the rigid and bulky allene system. For CgX, ^{1}H NMR analysis shows that both reactions are complete in 3 min: with prolionate, the product is mostly the decarboxylation product, acetaldehyde (75%) (i.e., hydration/decarboxylation), and with 2,3-butadienoate, the product is 100% acetoacetate (i.e., hydration).

N2 also shows mostly hydratase activity with both substrates. For propiolate, the k_{cat}/K_m is $\sim 10^4$, and for 2,3-butadienoate, the k_{cat}/K_m is ~67 (also with a high K_m value of 16100 μM). The reactions of N2 with propiolate and allene were examined further by ^{1}H NMR spectroscopy. After 35 min, a reaction mixture of N2 and propiolate consisted of the unreacted substrate (73%), malonate semialdehyde (8%), hydrate (19%), and trace amounts of acetaldehyde. After 35 min, a reaction mixture of N2 and allene consisted of the unreacted substrate (20.7%), acetoacetate (77.9%), and acetone (1.4%). The amount of decarboxylated product could be due to enzyme catalysis. Although a little more difficult to categorize, ^{1}H NMR analysis shows that N2 functions mostly as a hydratase with weak decarboxylase activity.

N1 and N2 also process 2-butynoate, as assessed by ^{1}H NMR spectroscopy. After 18 h, a reaction mixture of N1 and 2-butynoate consisted of acetoacetate (38.7%), acetone (11.1%), and starting material (50.2%). The amount of decarboxylated product is larger than that expected for non-enzymatic activity. After 18 h, a reaction mixture of N2 and 2-butynoate consisted of acetoacetate (78.6%), acetone (4.8%), and starting material (16.6%). The amount of decarboxylated product is that expected for non-enzymatic activity. The reaction of N2 is clearly faster than that of N1 (using 2-butynoate). The reaction of CgX with 2-butynoate is comparable, but much faster. After 39 min, the reaction mixture consists of mostly acetoacetate and a small amount of acetone (2.5%). The amount of acetone exceeds that expected for the non-enzymatic decarboxylation.[26]

The activity of N2 is faster than that of N1 using propiolate and comparable to that of N1 using the allene, as assessed by the kinetic parameters and ^{1}H NMR analysis. N1 and N2 also process 2-butyonate to acetoacetate (Scheme 3), as assessed by ^{1}H NMR analysis. Again, N2 is faster than N1. The significant hydration activity coupled with the modest tautomerase activity is typically associated with members of the *cis*-CaaD subgroup. Although the overall sequence similarity is significant enough for N2 to be placed in the 4-OT subgroup, the presence of *cis*-CaaD-like machinery at critical active site positions [two arginine residues to one side of Pro-1 (Figure 4B)] could cause the observed reaction preferences along with the presence of Tyr-39, which might position a water molecule to facilitate hydration. We have previously shown that replacement of Leu-8 with Arg-8 enhances the hydration activity of 4-OT (using *trans*-3-CAA).[15] Taken together, these observations provide experimental evidence for the transitional properties of N2, as suggested by the SSN analysis that identifies N2 as a linker protein between the 4-OT and *cis*-CaaD subgroups.

Incubation of N1 or N2 with 2-Oxo-3-pentynoate (2-OP). N1 and N2 were incubated in separate reaction mixtures

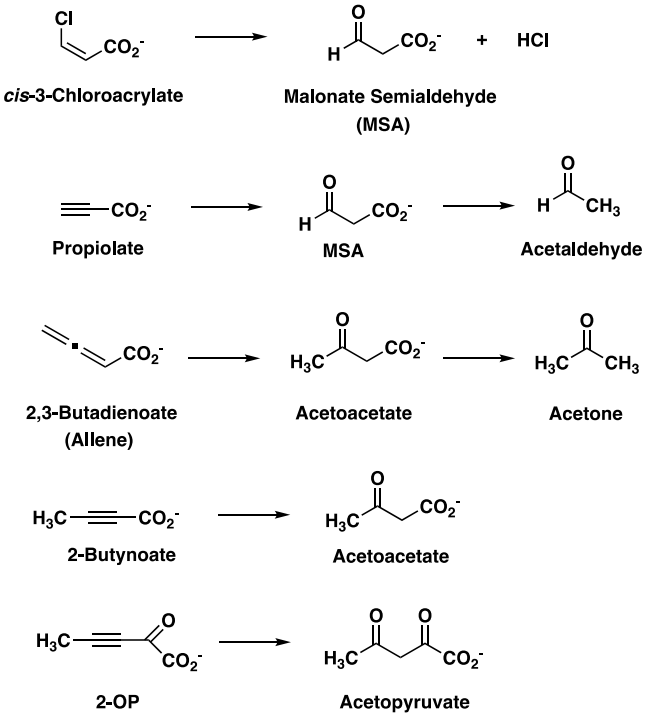
https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786

Table 2. Steady-State Kinetic Parameters for N1, N2, and CgX Using Substrates Shown in Scheme 1

	N1 (*cis*-CaaD subgroup)			N2 (4-OT subgroup)			CgX (*cis*-CaaD subgroup)		
substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
cis-3-CAA	0.14 ± 0.01	1270 ± 110	(1.0 ± 0.1) × 10^2	trace[a]			1.0 ± 0.1[d]	72000 ± 8500	14 ± 2
propiolate	0.12 ± 0.01	14.6 ± 1.6	(8.2 ± 0.9) × 10^3	1.20 ± 0.03[b]	20.2 ± 2.3	(5.9 ± 0.7) × 10^4	6.0 ± 0.2[d]	33 ± 5	(1.8 ± 0.2) × 10^5
2,3-butadienoate	23.0 ± 1.1	3100 ± 244	(7.4 ± 0.6) × 10^3	1.09 ± 0.05	16100 ± 1370	67 ± 7	4.0 ± 0.3[d]	780 ± 120	(5.1 ± 0.8) × 10^3
2-OP	0.08 ± 0.01	733 ± 59	(1.1 ± 0.1) × 10^2	ND[c,f]	ND[f]	ND[f]	0.33 ± 0.03[e]	6200 ± 750	(0.05 ± 0.1) × 10^3
PP	5.6 ± 0.3	386 ± 47	(1.4 ± 0.2) × 10^4	–	–	(6.7 ± 0.1) × 10^3	7.5 ± 0.5	410 ± 40	(1.8 ± 0.2) × 10^4
2-HM	ND[f]	ND[f]	ND[f]	0.48 ± 0.05	645 ± 95	(7.4 ± 1.0) × 10^2	ND[f]	ND[f]	ND[f]

[a]The activity is slightly lower than that of canonical 4-OT using *trans*-3-CAA, based on ^{1}NMR analysis.[30] [b]Canonical 4-OT catalyzes the hydration of propiolate, but very poorly. The following kinetic parameters were determined: k_{cat} = 6.7 × 10^{-3} s^{-1}, K_m = 19 mM, and k_{cat}/K_m = 3.5 × 10^{-1} M^{-1} s^{-1}. [c]N2 is covalently modified by 2-OP under the assay conditions [in 10 mM Na_2HPO_4 buffer (pH 7.3)]. Formation of acetopyruvate (Scheme 3), the product of hydration, is not detectable (by UV/vis), even after prolonged incubation times (16 h). [d]Data from ref 1. [e]Data from ref 25. [f]Not detected.

Scheme 3. Enzyme-Catalyzed Reactions with Various Substrates



with 2-OP at pH 7.3, as described in Experimental Procedures. After incubation for various amounts of time, the enzymes were recovered and analyzed by MS. The masses of the major signals are summarized in Table 3.

Table 3. Masses Corresponding to the Major Signals Observed in the Reconstructed Spectra of the Incubation Reaction Mixtures of N1 (native mass of 16104 Da) and N2 (native mass of 7010 Da) with 2-OP

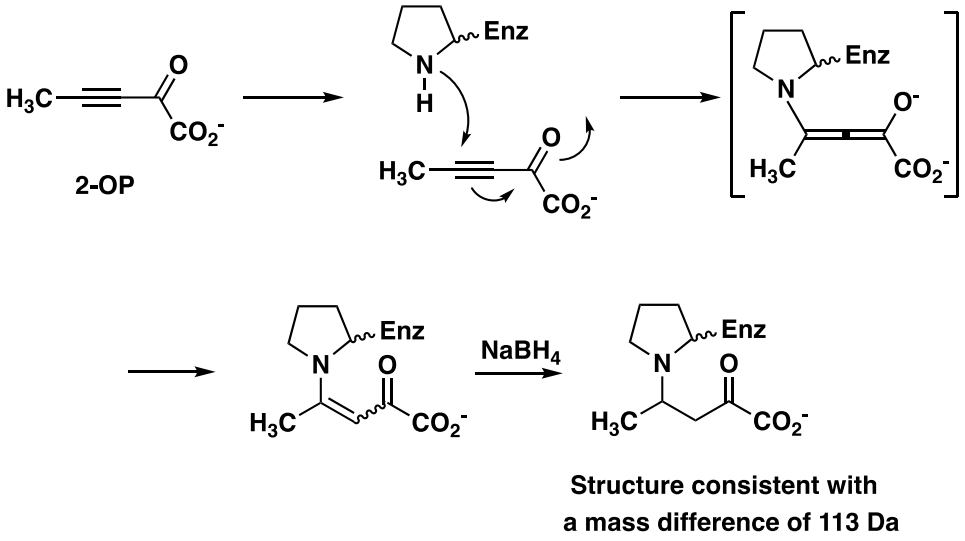
	N1		N2	
	observed mass (Da)	ratio[a]	observed mass (Da)	ratio
2-OP	16104	2:1	7010	1:1:1.5
	16217		7052[b]	
			7123	

[a]The ratio represents the relative signal intensity. [b]The +42 Da adduct is likely an artifact of the MS conditions.

Both N1 and N2 were covalently modified by 2-oxo-3-pentynoate (2-OP) as reflected by the signals at 16217 and 7123 Da, respectively, corresponding to an adduct with a mass of 113 Da (Scheme 4). 2-OP is used as a probe for the pK_a value of the prolyl nitrogen (of Pro-1) on the basis of its behavior with different TSF members.[3,31,32] In general, if the pK_a is low (<7), the enzyme is irreversibly inactivated by the formation of a covalent bond between the prolyl nitrogen and C-4 of 2-OP (Scheme 4).[14] If the pK_a is high (>9), the enzyme catalyzes a hydration reaction to form acetopyruvate at pH ∼7 (Scheme 3).[3,31,32] In 4-OT, the prolyl nitrogen of Pro-1 has a pK_a of 6.4, as measured by ^{15}N NMR spectroscopy,[5] and the enzyme is irreversibly inhibited by covalent modification of the prolyl nitrogen.[14] For N2, the results parallel those of 4-OT, suggesting that Pro-1 has a low pK_a value. Notably, most of the protein is labeled (∼71.4%) and there is no detectable hydration (see Tables 2 and 3). N1, on the contrary, is in the

https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786

Scheme 4. Mechanism for Inactivation by 2-OP and Reduction by $NaBH_4$



cis-CaaD subgroup. The pK_a of Pro-1 in *cis*-CaaD is 9.3, consistent with its role as a general acid in the dehalogenation reaction.[3,33] Hence, *cis*-CaaD catalyzes the hydration of 2-OP along with some inactivation.[3] N1 is also modified by 2-OP (~33.3%), but not as extensively as N2. The N1-catalyzed hydration of 2-OP could also decrease the amount of 2-OP for reaction with the prolyl nitrogen (Table 2). Nonetheless, if these observations can be viewed as rough comparisons of the pK_a values of the prolines, it suggests that they are different and the pK_a of Pro-1 in N1 is higher than the pK_a of Pro-1 in N2.

CONCLUSIONS

Two previously identified linker nodes (N1 and N2) in the SSN that connect the *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD) and 4-oxalocrotonate tautomerase (4-OT) subgroups[1] were subjected to sequence, structure, and kinetic characterization. Sequence analysis places N1 in the *cis*-CaaD subgroup and N2 in the 4-OT subgroup. N1 functions like the other characterized *cis*-CaaD subgroup members with both tautomerase (using a monocarboxylate, but not a dicarboxylate substrate) and hydratase activities (using *cis*-CAA, allene, and acetylene substrates). The active site of N1 mirrors those of the other *cis*-CaaD subgroup members and is consistent with the kinetic profile. N2, however, does not behave like other characterized 4-OT subgroup members but functions in a *cis*-CaaD-like manner. It has significantly reduced tautomerase activity (using mono- and dicarboxylate substrates) and enhanced hydratase activity (using allene and acetylene substrates). Structural analysis shows two substitutions that might convey *cis*-CaaD-like behavior. Leu-8 (in 4-OT) is replaced by an arginine, and Arg-39 is replaced by a tyrosine. As a result, Pro-1 is now not between two arginine residues, but on one side of them. This is a common feature of *cis*-CaaD subgroup members. Structural analysis further suggests that Tyr-39 might position a water molecule for fortuitous addition to the substrate. These active site changes are simple but were not predictable ones from sequence function patterns defined by large-scale analysis of more than 11000 TSF members, underscoring the difficulty of functional assignment of proteins.

ASSOCIATED CONTENT

Accession Codes

The atomic coordinates and structure factors have been deposited as Protein Data Bank entries 7M58 for N1 and 7M59 for N2. N1 from *C. halotolerans* YIM 70093 (UniProt entry M1NLA4). N2 from *G. bacterium* SG8_31 (UniProt entry A0A0S8FF56).

AUTHOR INFORMATION

Corresponding Author

Christian P. Whitman − *Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy and Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0002-8231-2483; Phone: (512) 471-6198; Email: whitman@austin.utexas.edu; Fax: (512) 232-2606

Authors

Bert-Jan Baas − *Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, United States*

Brenda P. Medellin − *Department of Molecular Biosciences, The University of Texas at Austin, Austin, Texas 78712, United States*

Jake A. LeVieux − *Department of Molecular Biosciences, The University of Texas at Austin, Austin, Texas 78712, United States*

Kaci Erwin − *Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, United States*

Emily B. Lancaster − *Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, United States*

William H. Johnson, Jr. − *Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, United States*

Tamer S. Kaoud − *Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, United States*

R. Yvette Moreno − *Department of Molecular Biosciences, The University of Texas at Austin, Austin, Texas 78712, United States*

https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786

Marieke de Ruijter − *Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, United States*
Patricia C. Babbitt − *Departments of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, California 94158, United States*
Yan Jessie Zhang − *Department of Molecular Biosciences and Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0002-9360-5388

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.1c00220

Funding
This research was supported by National Institutes of Health grants [GM-129331 to C.P.W. and Y.J.Z., GM-104896 and GM125882 to Y.J.Z., GM60595 to P.C.B., and S10 OD021508−01 (Bruker AVANCE III 500 MHz spectrometer)] and Robert A. Welch Foundation grants (F-1334 to C.P.W. and F-1778 to Y.J.Z.).

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The protein mass spectrometry analysis was conducted in the Institute for Cellular and Molecular Biology Protein and Metabolite Analysis Facility at The University of Texas at Austin. The authors thank Steve D. Sorey (Department of Chemistry, The University of Texas at Austin) for his expert assistance in the acquisition of the ^{1}H NMR spectra. Instrumentation and technical assistance for this work were provided by the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at The University of Texas at Austin. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract DE-AC02-05CH11231.

■ ABBREVIATIONS

APS, Advanced Photon Source; CHMI, 5-(carboxymethyl)-2-hydroxymuconate isomerase; *cis*-CaaD, CC, or *cis*-3-CAA, *cis*-3-chloroacrylic acid dehalogenase; CgX, *cis*-CaaD homologue designated Cg10062 from *Corynebacterium glutamicum*; DEAE, diethylaminoethyl; DMSO, dimethyl sulfoxide; MS, mass spectrometry or mass spectrometric; F4-OT, fused 4-oxalocrotonate tautomerase; 2-HM, 2-hydroxymuconate; Kn, kanamycin; LB, Luria-Bertani; MPD, 2-methyl-2,4-pentanediol; MIF, macrophage migration inhibitory factor; MSA, malonate semialdehyde; MSAD, malonate semialdehyde decarboxylase; MR, molecular replacement; MW, molecular weight; MAD, multiwavelength anomalous dispersion; NMR, nuclear magnetic resonance; 4-OT, 4-oxalocrotonate tautomerase; PP, phenylenolpyruvate; PEG, polyethylene glycol; PDB, Protein Data Bank; 2-OP, 2-oxo-3-pentynoate; SSN, sequence similarity network; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; *trans*-3-CAA, *trans*-3-chloroacrylic acid dehalogenase; TLS, translation-libration-screw-rotation; TSF, tautomerase superfamily; UV, ultraviolet.

■ REFERENCES

(1) Davidson, R., Baas, B.-J., Akiva, E., Holliday, G., Polacco, B. J., LeVieux, J. A., Pullara, C. R., Zhang, Y. J., Whitman, C. P., and Babbitt, P. C. (2018) A global view of structure-function relationships in the tautomerase superfamily. *J. Biol. Chem. 293*, 2342−2357.
(2) Murzin, A. G. (1996) Structural classification of proteins: new superfamilies. *Curr. Opin. Struct. Biol. 6*, 386−394.
(3) Poelarends, G. J., Veetil, V. P., and Whitman, C. P. (2008) The chemical versatility of the β-α-β fold: Catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell. Mol. Life Sci. 65*, 3606−3618.
(4) Stivers, J. T., Abeygunawardana, C., Mildvan, A. S., Hajipour, G., Whitman, C. P., and Chen, L. H. (1996) Catalytic role of the amino-terminal proline in 4-oxalocrotonate tautomerase: affinity labeling and heteronuclear NMR studies. *Biochemistry 35*, 803−813.
(5) Stivers, J. T., Abeygunawardana, Mildvan, A. S., Hajipour, G., and Whitman, C. P. (1996) 4-Oxalocrotonate tautomerase: pH dependences of catalysis and pKa values of active site residues. *Biochemistry 35*, 814−823.
(6) Whitman, C. P., Aird, B. A., Gillespie, W. R., and Stolowich, N. J. (1991) Chemical and enzymatic ketonization of 2-hydroxymuconate, a conjugated enol. *J. Am. Chem. Soc. 113*, 3154−3162.
(7) Wang, S. C., Person, M. D., Johnson, W. H., Jr., and Whitman, C. P. (2003) Reactions of *trans*-3-chloroacrylic acid dehalogenase with acetylene substrates: consequences of and evidence for a hydration reaction. *Biochemistry 42*, 8762−8773.
(8) Poelarends, G. J., Serrano, H., Person, M. D., Johnson, W. H., Jr., Murzin, A. G., and Whitman, C. P. (2004) Cloning, expression, and characterization of a *cis*-3-chloroacrylic acid dehalogenase: insights into the mechanistic, structural, and evolutionary relationship between isomer-specific 3-chloroacrylic acid dehalogenases. *Biochemistry 43*, 759−772.
(9) Poelarends, G. J., Johnson, W. H., Jr., Murzin, A. G., and Whitman, C. P. (2003) Mechanistic characterization of a bacterial malonate semialdehyde decarboxylase: identification of a new activity on the tautomerase superfamily. *J. Biol. Chem. 278*, 48674−48683.
(10) Zandvoort, E., Baas, B.-J., Quax, W. J., and Poelarends, G. J. (2011) Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: enamine formation and aldolase activity. *ChemBioChem 12*, 602−609.
(11) Baas, B.-J., Poddar, H., Geertsema, E. M., Rozeboom, H. J., de Vries, M. P., Permentier, H. P., Thunnissen, A.-M. W. H., and Poelarends, G. J. (2015) Functional and structural characterization of an unusual co-factor independent oxygenase. *Biochemistry 54*, 1219−1232.
(12) Taylor, A. B., Johnson, W. H., Jr., Czerwinski, R. M., Li, H.-S., Hackert, M. L., and Whitman, C. P. (1999) Crystal structure of macrophage migration inhibitory factor complexed with (*E*)-2-fluoro-p-hydroxycinnamate at 1.8 Å resolution: implications for enzymatic catalysis and inhibition. *Biochemistry 38*, 7444−7452.
(13) LeVieux, J. A., Baas, B.-J, Kaoud, T. S., Davidson, R., Babbitt, P. C., Zhang, Y. J., and Whitman, C. P. (2017) Kinetic and structural characterization of a *cis*-3-chloroacrylic acid dehalogenase homologue in *Pseudomonas* sp. UW4: A potential step between subgroups in the tautomerase superfamily. *Arch. Biochem. Biophys. 636*, 50−56.
(14) Taylor, A. B., Czerwinski, R. M., Johnson, W. H., Jr., Whitman, C. P., and Hackert, M. L. (1998) Crystal structure of 4-oxalocrotonate tautomerase inactivated by 2-oxo-3-pentynoate at 2.4 Å resolution: analysis and implications for the mechanism of inactivation and catalysis. *Biochemistry 37*, 14692−14700.
(15) Poelarends, G. J., Almrud, J. J., Serrano, H., Darty, J. E., Johnson, W. H., Jr., Hackert, M. L., and Whitman, C. P. (2006) Evolution of enzymatic activity in the tautomerase superfamily:

mechanistic and structural consequences of the L8R mutation in 4-oxalocrotonate tautomerase. *Biochemistry 45*, 7700−7708.
(16) Eglinton, G., Jones, E. R. H., Mansfield, G. H., and Whiting, M. C. (1954) Researches on acetylenic compounds. Part XLV. The alkaline isomerisation of but-3-ynoic acid. *J. Chem. Soc.*, 3197−3200.
(17) Waddell, W. J. (1956) A simple ultraviolet spectrophotometric method for the determination of protein. *J. Lab. Clin. Med. 48*, 311−314.
(18) Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*, 680−685.
(19) Kaoud, T. S., Devkota, A. K., Harris, R., Rana, M. S., Abramczyk, O., Warthaka, M., Lee, S., Girvin, M. E., Riggs, A. F., and Dalby, K. N. (2011) Activated ERK2 is a monomer in vitro with or without divalent cations and when complexed to the cytoplasmic scaffold PEA-15. *Biochemistry 50*, 4568−4578.
(20) Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 68*, 352−367.
(21) Burks, E. A., Yan, W., Johnson, W. H., Jr., Li, W., Schroeder, G. K., Min, C., Gerratana, B., Zhang, Y., and Whitman, C. P. (2011) Kinetic, crystallographic, and mechanistic characterization of TomN: elucidation of a function for a 4-oxalocrotonate tautomerase homologue in the tomaymycin biosynthetic pathway. *Biochemistry 50*, 7600−7611.
(22) Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 12−21.
(23) DeLano, W. L. (2002) *The PyMol molecular graphics system*, DeLano Scientific, San Carlos, CA.
(24) Burks, E. A., Fleming, C. D., Mesecar, A. D., Whitman, C. P., and Pegan, S. D. (2010) Kinetic and structural characterization of a heterohexamer 4-oxalocrotonate tautomerase from *Chloroflexus aurantiacus* J-10-fl: implications for functional and structural diversity in the tautomerase superfamily. *Biochemistry 49*, 5016−5027.
(25) Poelarends, G. J., Serrano, H., Person, M. D., Johnson, W. H., Jr., and Whitman, C. P. (2008) Characterization of Cg10062 from *Corynebacterium glutamicum*: implications for the evolution of *cis*-3-chloroacrylic acid dehalogenase activity in the tautomerase superfamily. *Biochemistry 47*, 8139−8147.
(26) Huddleston, J. P., Johnson, W. H., Jr., Schroeder, G. K., and Whitman, C. P. (2015) Reactions of Cg10062, a *cis*-3-chloroacrylic acid dehalogenase homologue, with acetylene and allene substrates: evidence for a hydration-dependent decarboxylation. *Biochemistry 54*, 3009−3023.
(27) Schroeder, G. K., Johnson, W. H., Jr., Huddleston, J. P., Serrano, H., Johnson, K. A., and Whitman, C. P. (2012) Reaction of *cis*-3-chloroacrylic acid dehalogenase with an allene substrate, 2,3-butadienoate: hydration via an enamine. *J. Am. Chem. Soc. 134*, 293−304.
(28) de Jong, R. M., Bazzacco, P., Poelarends, G. J., Johnson, W. H., Jr, Kim, Y.-J., Burks, E. A., Serrano, H., Thunnissen, A.-M.W.H., Whitman, C. P., and Dijkstra, B. W. (2007) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase: structural basis for substrate specificity and inactivation by (*R*)-oxirane-2-carboxylate. *J. Biol. Chem. 282*, 2440−2449.
(29) Protein Data Bank entries 3N4G and 3N4D.
(30) Wang, S. C., Johnson, W. H., Jr., and Whitman, C. P. (2003) The 4-oxalocrotonate tautomerase- and YwhB-catalyzed hydration of 3*E*-haloacrylates: implications for the evolution of new enzymatic activities. *J. Am. Chem. Soc. 125*, 14282−14283.
(31) Poelarends, G. J., Serrano, H., Johnson, W. H., Jr., Hoffman, D. W., and Whitman, C. P. (2004) The hydratase activity of malonate semialdehyde decarboxylase: mechanistic and evolutionary implications. *J. Am. Chem. Soc. 126*, 15658−15659.
(32) Azurmendi, H. F., Wang, S. C., Massiah, M. A., Poelarends, G. J., Whitman, C. P., and Mildvan, A. S. (2004) The roles of active-site residues in the catalytic mechanism of *trans*-3-chloroacrylic acid dehalogenase: a kinetic, NMR, and mutational analysis. *Biochemistry 43*, 4082−4091.
(33) Poelarends, G. J., Serrano, H., Johnson, W. H., Jr., and Whitman, C. P. (2004) Stereospecific alkylation of *cis*-3-chloroacrylic acid dehalogenase by (*R*)-oxirane-2-carboxylate: analysis and mechanistic implications. *Biochemistry 43*, 7187−7196.

https://doi.org/10.1021/acs.biochem.1c00220
Biochemistry 2021, 60, 1776−1786