Characterization of Member of DUF1888 Protein Family, Self-cleaving and Self-assembling Endopeptidase

Jerzy Osipiuk, Rory Mulligan, Monireh Bargassa, John E. Hamilton, Mark A. Cunningham, and Andzej Joachimiak

From the 1Midwest Center for Structural Genomics and Structural Biology Center, Biosciences, Argonne National Laboratory, Argonne, Illinois 60439 and the 2Department of Physics and Geology, University of Texas-Pan American, Edinburg, Texas 78539

Background: The Shewanella oneidensis SO1698 belongs to a protein family with unknown function.

Results: The protein has three activities; it cleaves its own peptide backbone, it forms an internal cross-link through Lys-98, and it self-assembles into a covalently linked oligomer.

Conclusion: The conserved Asp-Pro motif activates water that hydrolyzes the protein backbone.

Significance: The protein represents a new family of endopeptidases.

The crystal structure of SO1698 protein from Shewanella oneidensis was determined by a SAD method and refined to 1.57 Å. The structure is a β sandwich that unexpectedly consists of two polypeptides; the N-terminal fragment includes residues 1–116, and the C-terminal one includes residues 117–125. Electron density also displayed the Lys-98 side chain covalently linked to Asp-116. The putative active site residues involved in self-cleavage were identified; point mutants were produced and characterized structurally and in a biochemical assay. Numerical simulations utilizing molecular dynamics and hybrid quantum/classical calculations suggest a mechanism involving activation of a water molecule coordinated by a catalytic aspartic acid.

As part of a broader initiative for protein structure determination, we have investigated several proteins from Shewanella oneidensis strain MR-1 and report here the crystal structure of the 125-residue SO1698 protein. The Gram-negative bacterium S. oneidensis (1) is noteworthy because it can be cultured under both aerobic and anaerobic conditions. Under anaerobic conditions, the organism derives energy from the reduction of heavy metals and demonstrates a remarkable tolerance for normally toxic metals, such as chromium and uranium. As a result, there is significant interest in utilizing the S. oneidensis for bioremediation purposes; metals in reduced oxidation states typically demonstrate reduced solubilities, inhibiting their transport through soils and minimizing their environmental impact. Because S. oneidensis can be cultured under aerobic conditions, large scale production and handling is simplified. The SO1698 protein belongs to a domain of unknown function family (DUF1888) and is assigned to HMM-Pfam PF08985 (2), consisting principally of uncharacterized proteins from several Shewanella species (Fig. 1) and classified as a cupredoxin-like fold. Its specific biological function is unknown at this time. In our studies of the protein, we have identified three notable biological activities. First, the protein cleaves its own peptide backbone in a pH-dependent fashion, like other aspartic peptidases. Second, it forms an internal cross-link through the side chain of Lys-98 at the site of cleavage. Third, at low pH, the protein self-assembles into an oligomeric state, most likely hexameric, which displays unusual stability in denaturing conditions. These uncommon activities lead us to believe that the SO1698 protein is a member of a new family of endopeptidases using the Asp-Pro (DP) motif, and we shall refer to it as the DP-EP protein.

The DP-EP protein does not appear to possess exogenous peptidase activity. We were not able to find any substrate for DP-EP protein proteolysis other than the DP-EP protein itself (data not shown). BLAST analysis of the protein finds homologous proteins predominantly in other strains of Shewanella, as is illustrated in Fig. 1. There are, however, similar proteins identified in Alteromonas (Alteromonas bacterium) and Alteromonadales (Alteromonadales macleodii). In all of these proteins, we observe a conserved SXDP sequence after the final β strand, where the nonconserved residue X is predominantly a proline. In our studies, mutation of the leading proline residue (P115A) does not affect the autoproteolytic function. The DP-EP protein does not have any close structural homologs. DALI analysis (3) shows human coagulation factor VIII (693-amino acid protein, Protein Data Bank entry 2R7E) as the closest homolog with a Z-score (the strength of structural similarity

* This work was supported, in whole or in part, by National Institutes of Health Grant GM074942. This work was also supported by the United States Department of Energy, Office of Biological and Environmental Research, under Contract DE-AC02–06CH11357. This work has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”), Argonne, a United States Department of Energy Office of Science laboratory, is operated under Contract DE-AC02–06CH11357. This work was also supported by the United States Department of Energy, Office of Biological and Environmental Research, under Contract DE-AC02–06CH11357. This work has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”), Argonne, a United States Department of Energy Office of Science laboratory, is operated under Contract DE-AC02–06CH11357.

** This article contains supplemental Table S1, Figs. S1–S5, and Movie S1.

The atomic coordinates and structure factors (codes 3N5S, 3NJF, 3NJG, 3NJH, 3NJL, 3NJQ, 3NJK, 3NJU, 3NJM, and 3NJN) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

1 Supported by National Science Foundation FaST program Grant HRD-0703584, administered by the Department of Educational Programs at Argonne National Laboratory. To whom correspondence may be addressed: Dept. of Physics and Geology, University of Texas-Pan American, 1201 W. University Dr., Edinburg, TX 78539. Tel.: 956-665-2171; Fax: 956-665-2423; E-mail: cunningham@utpa.edu.

2 To whom correspondence may be addressed: Midwest Center for Structural Genomics and Structural Biology Center, Biosciences, Argonne National Laboratory, 9700 S. Cass Ave., Argonne, IL 60439. Tel.: 630-252-3926; Fax: 630-252-6126; E-mail: andrzejj@anl.gov.

This page has been reformatted electronically. Pages 19452–19461 from THE JOURNAL OF BIOLOGICAL CHEMISTRY, VOL. 287, NO. 23, June 1, 2012. Published by the American Society for Biochemistry and Molecular Biology. Published in the U.S.A.
in S.D. values above expected) and r.m.s. deviation (of superimposed atoms) equal to 5.3 and 3.0 Å, respectively. Visual inspection of structure similarity excludes coagulation factor as a structural homolog for the DP-EP protein. Another match from DALI analysis was a human T cell receptor (Protein Data Bank entry 2IAN) that shows some visual similarity due to its sandwich structure.

**EXPERIMENTAL PROCEDURES**

**Protein Cloning, Expression, and Purification**—The ORF of the DP-EP protein from *S. oneidensis* was amplified from genomic DNA with KOD DNA polymerase using conditions and reagents provided by Novagen (Madison, WI). The gene was cloned into pMCSG7 (5) vector using a modified LIC protocol (6). This process generated an expression clone producing a fusion protein with an N-terminal His6 tag and a tobacco etch virus protease recognition site. A selenomethionine derivative of the expressed protein was prepared as described by Walsh et al. (7) and purified using standard procedures on an AKTAxpress semiautomated purification system (GE Healthcare) (8). The concentration of the purified protein was determined utilizing an ND-1000 spectrophotometer system (NanoDrop Technologies). The fusion tag was then removed by adding recombinant tobacco etch virus protease at a ratio of 1:75 and incubated for 48 h at 4 °C. The cleaved protein was then separated on a nickel-NTA-agarose nickel charged resin column (Qiagen Inc.). The purified protein solution was dialyzed in a crystallization buffer (20 mM HEPES, pH 8.0, 250 mM NaCl, 2 mM dithiothreitol) for 24 h and concentrated using a Centricon Plus-20 concentrator with a normal molecular weight limit of 5,000 (Millipore Corp.).

The purified DP-EP protein runs as a double band on 20% SDS-PAGE, indicating probable partial truncation of the protein (Fig. 4). Additionally, an oligomeric complex is observed on the SDS-PAGE that is most likely a hexamer.

**Size Exclusion Chromatography**—Size exclusion chromatography was performed on a Superdex-200 10/300GL column using FPLC (GE Healthcare). The column was pre-equilibrated with crystallization buffer (20 mM HEPES, pH 8.0, 250 mM NaCl, 2 mM DTT) and calibrated with protein standards: thyroglobulin (669 kDa), ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa), conalbumin (75 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carboxyanhydrase (29 kDa), ribonuclease A (13.5 kDa), and blue dextran (2,000 kDa). A 200-μl DP-EP protein sample at 4.7 mg/ml (or its D37A mutant) was injected into the column. The chromatography was carried out at room temperature at a flow rate of 0.3 ml/min. The calibration curve of $K_{av}$ versus log molecular weight was prepared using the equation, $K_{av} = (V_e - V_o)/(V_t - V_o)$, where $V_e$ represents the elution volume for the protein, $V_o$ is column void volume, and $V_t$ is total bed volume.

**Protein Crystallization**—The DP-EP native protein was crystallized using vapor diffusion in sitting drops and a CrystalQuick® standard profile-LBR round bottom plate (Greiner Bio-One North America, Inc.). A 0.4-μl droplet of protein (199 mg/ml) was mixed with a 0.4-μl droplet of crystallization reagent and allowed to equilibrate over 135 μl of crystallization reagent. The nanopipetting was performed using the Mos-
DUF1888 Family Self-cleaving Endopeptidase

quito® nanoliter liquid handling system (TTP LabTech). The finished plate was then incubated at 18 °C within a RoboIncubator® automated plate storage system (RoboDesign International, Inc.). Automated crystal visualization was utilized in locating crystals, Minstrel III® (RoboDesign International, Inc). These crystals were cryoprotected and flash-frozen in liquid nitrogen. The native protein crystallized in space group H32 with cell dimensions of \( a = b = 100.0 \, \text{Å} \) and \( c = 100.9 \, \text{Å} \), \( \alpha = \beta = 90^\circ \), \( \gamma = 120^\circ \). Mutant proteins were treated in the same manner as the native protein.

**Data Collection and Structure Determination**—Diffraction data were collected at 100 K at the 19-1D and 19-BM beamlines of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. The SAD data at 0.9792-Å wavelength were collected from a single (0.1 × 0.04 × 0.05-mm) selenomethionine-labeled protein crystal. The structure was determined by SAD phasing using the HKL3000 (9) suite incorporating SHELXC, SHELXD, SHELXE (10), MLPHARE, and SOLVE/RESOLVE (11) programs. The initial model was completed by using ARP/wARP (12), and manual modeling was completed using COOT (13). The structure was refined to 1.57 Å using REFMAC (14) in CCP4 (15). The final R-factor was 14.5% with the \( R_{free} \) factor of 16.8% with all data (Table 1). Data collection for mutant proteins followed the same procedures as for the native protein. Crystallographic data for the mutant proteins are summarized in supplemental Table S1.

**Validation and Deposition**—The stereochemistry of the structure was checked with PROCHECK (16) and the Ramachandran plot. The main chain torsion angles for all residues are in allowed regions. Atomic coordinates and experimental structure factors of the native structure have been deposited in the Protein Data Bank (Protein Data Bank entry 3N55). The following mutant structures have also been deposited: Y26F (3NJF), D37A (3NJH), K98A (3NJG), S114A (3NNJ), P115A (3NNJ), D116A (3NNK at pH 5.5; 3NJJ at pH 7.5), P117A (3NJM), and Q118A (3NNJ).

**Active Site Mutagenesis**—Site-directed mutagenesis was performed to identify residues essential for autoproteolysis. The original clone in pMCSG7 plasmid was used as a DNA template to generate eight site-directed mutants using the QuikChange site-directed mutagenesis method and reagents (Stratagene, La Jolla, CA). The plasmids were verified as carrying the desired mutation by DNA sequencing analysis, and the modified proteins were overproduced. The following mutants were constructed: Y26F, D37A, K98A, S114A, P115A, D116A, P117A, and Q118A. All mutants were crystallized at various pH conditions.

**Proteolytic Activity**—To test the pH dependence of the autoproteolytic reaction, protein samples were incubated at 30 °C for 2.5 h in 50 mM MES buffer at different pH values. The transformation of the expected 13.7 kDa band into a truncated 12.7 kDa band was analyzed by SDS-PAGE. To test endopeptidase activity, bovine serum albumin (BSA) was submitted to the autoproteolytic reaction as described above.

**BLAST Analysis**—BLAST analysis (17) of the target sequence gi 24373266 from *S. oneidensis* MR-1 identified the following homologous sequences: gi|217973998 (Shewanella baltica OS223), gi|126173752 (S. baltica OS155), gi|153000037 (S. baltica OS185), gi|117921197 (Shewanella sp. ANA-3), gi|113970916 (Shewanella sp. MR-4), gi|120599486 (Shewanella sp. W3-18-1), gi|114561978 (Shewanella frigidimarina NCIMB 400), gi|212633520 (Shewanella piezotolerans WP3 fragment 80–136 and 177–253), (Shewanella amazonesis SB2B), gi|119471885 (A. bacterium TW-7), and gi|196155910 (A. macleodi “Deep ecotype”), gi|239996420 (A. macleodi ATCC 27126), and gi|163748478 (Shewanella benthica KT99).

**Molecular Dynamics**—Molecular dynamics simulations were performed with the code NAMD (18), developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. The CHARMM 27 force field (19) parameters were used. Analysis of dynamics trajectories was conducted with the program VMD (20), which was also used for figure preparation. Initial coordinates for protein atoms were obtained from the crystal structures. Missing atom coordinates (hydrogens) were defined with the PSFGEN module of VMD, which was also used to solvate the protein and neutralize the total charge of the simulation model. The solvation box was extended for 10 Å beyond the protein. With a 1 Å cut-off for electrostatic interactions, the box dimensions ensured a “diffuse” simulation model, in which no protein atoms interacted directly with protein atoms in the periodic images. Electrostatic interactions were computed using a smooth particle mesh Ewald method (21), with a grid size of ~1 Å spacing. Simulations began with a small amount of minimization (~1,000 steps) and then utilized typically 100,000 steps of NVT dynamics and 100,000 steps of NPT dynamics run with 1-fs time steps to equilibrate the system. Constant temperature was maintained with a Langevin method (22), and constant pressure conditions were enforced through a modified version of the Langevin piston (23) and Hoover (24, 25) methods. Typical production runs were of 1-ns duration and were conducted using NPT dynamics with 2-fs time steps, recording coordinate information at 1-ps intervals.

Simulations were conducted for monomeric, dimeric, and hexameric conformations of the protein, with the majority of simulations utilizing the dimeric form. In these cases, there were ~28,000 atoms in the simulation, with a simulation box size of ~50 × 65 × 80 Å. For the hexameric simulations, the model included ~63,000 atoms in a box of dimensions 95 × 101 × 70 Å. We observed noticeable differences in structure between the monomeric and dimeric forms but no appreciable differences between dimeric and hexameric forms.

In the molecular dynamics simulations that began from the native crystal structure, we observed that the initial minimization steps restored the cleaved peptide bond. Subsequent trajectories were compared with crystal structures of the noncleaving mutants: the r.m.s. deviation difference between Ca atoms of the crystal structure and average simulation structures was generally in the range of 1.5–1.7 Å, with the largest differences arising from mobile loops. The simulations were conducted in a diffuse limit, where protein atoms were not subjected to packing forces from other protein atoms, thereby accounting for much of the observed differences in structure.
Quantum Mechanics/Molecular Mechanics (QM/MM) Studies—To study energetics along the proposed reaction pathway, we performed hybrid QM/MM calculations with the program NWChem (26). We utilized a nudged elastic band (NEB) method (27) implemented recently. Initial coordinates for reactant states were taken from snapshots of the dynamics trajectories in which the attacking water molecule was positioned in what we would describe as a near attack conformation (28). The quantum partitions for all simulations included residues 114–118 and an attacking water (supplemental Fig. S4); hydrogen atoms were used as link atoms. Atoms beyond 15 Å from the target carbonyl carbon atom of Asp-116 were frozen in place, and only atoms within that spherical region were allowed to move. Reactant states were defined by a process in which the original model was optimized using the density functional method B3LYP (29) and a 3–21g* basis set to speed convergence to a coarse estimate of the state geometry. All subsequent QM/MM calculations used the 6–31g** level of theory. One of the water protons was then constrained to be 1 Å from the N nitrogen atom of Pro-117, and, if necessary, the N–C distance between the two leaving groups was constrained to be 3.1 Å. The geometry optimizations were then repeated. All atom constraints were then removed, and the model geometry was reoptimized.

The NEB simulations required an initial guess for states along the reaction pathway. These initial guesses were provided by optimizations with constraints applied to key atom distance. In the native protein simulation, for example, the distance between one of the water protons and the O=C oxygen atom of Asp-116 was first shortened stepwise in 0.2-Å increments to a final value of 1 Å. The O–C distance between the water oxygen atom and the C carbon atom of Asp-116 was then shortened stepwise in 0.2-Å increments from its initial value to 1.35 Å. The water proton was then moved stepwise to the N nitrogen atom of Pro-117, and finally the C–N distance was pushed out stepwise to 3 Å. This initial guess for the pathway resulted in 29 replicas (beads) of the system. The NEB methodology then was used to define an optimized pathway, one that provided a minimum energy path from reactant to product state. Free energies were obtained by running 20 ps of molecular dynamics to estimate the change in free energy from one bead to the next.

RESULTS AND DISCUSSION

The native protein crystallizes readily and rhombohedral crystals diffract to high resolution. The crystal structure was determined by a single-wavelength anomalous diffraction method and refined to 1.57 Å. Crystallographic data are summarized in Table 1. The asymmetric unit of the protein structure contains a single protein monomer, as illustrated in schematic form in Fig. 2. The monomer forms a β sandwich consisting of two mixed β sheets. The structure shows distant similarity to convalin (Protein Data Bank entry 1DGR). Unexpectedly, the polypeptide chain near the C-terminal end of the protein is severed, as can be seen more readily in the electron density depicted in Fig. 3. Two distinct polypeptides are clearly visible at 1.57 Å resolution; the N-terminal fragment includes residues 1–116, and the C-terminal fragment includes residues 117–125. The C-terminal strand remains bound to the protein main body in a groove between strands β$_1$ and β$_2$.

In addition, the electron density maps clearly revealed that the Lys-98 side chain nitrogen atom is covalently linked to the peptide carbon atom of Asp-116 (Fig. 3). The major multimeric form of the protein in the crystal appears to be a dimer (Fig. 2) formed by two monomers related by 2-fold crystallographic symmetry. The dimerization interface is quite extensive and involves a number of direct and solvent-mediated interactions using main chain and side chain atoms. The total buried solvent-accessible area due to the dimer formation is about 1,243 Å$^2$. In the crystal, three dimers assemble into a hexamer (supplemental Fig. S1).

Proteolytic Activity and Oligomerization—We initially hypothesized that the DP-EP protein was an aspartic endopeptidase utilizing the aspartate residue present to cleave the peptide bond. There is a clear pH dependence evident, with an optimal pH between 6 and 6.2. Tests for endopeptidase activity using BSA protein did not show any degradation of BSA. Hence, the SO1698 protein seems to have only autoproteolytic activity.

In Fig. 4, we also observe an oligomeric band. The band corresponds to 57 kDa based on used molecular weight standards, which could represent a protein tetramer. However, the protein monomer migrates as a 10.5-kDa protein in the same gel instead of a 13.7 kDa band, and the cleaved monomer migrates as a 9-kDa protein instead of a 12.7 kDa band. This anomaly could be explained by a low protein pI value equal to 4.23. It has been observed that acidic proteins migrate on SDS-PAGE anomalously. Thus, the observed cross-linked band represents

**TABLE 1**

| Summary of the SO1698 protein crystallographic data |
|--------------------------------------------------|
| **Data collection**                                |
| Experiment                                       | SAD                               |
| Space group                                      | H 3 2                             |
| Unit cell dimensions                             |                                   |
| a, b, c (Å)                                      | 100.0, 100.0, 100.9               |
| α, β, γ (degrees)                                | 90, 90, 120                       |
| Protein molecule/asymmetric unit                 | 1                                 |
| Energy (keV)                                     | 0.9793                            |
| **Data processing**                              |
| Resolution range (Å)                             | 32.9–1.57                         |
| Unique reflections                               | 52008                             |
| Redundancy                                       | 12.7                              |
| Completeness (%)                                 | 98.8                              |
| I/σ (all)                                        | 44.1                              |
| R$_{merge}$ (%)                                   | 5.1                               |
| R$_{work}$ (%)                                   | 34.9                              |
| R$_{free}$ (%)                                    | 37.2                              |
| r.m.s. deviation, bond length (Å)                 | 0.019                             |
| r.m.s. deviation, bond angle (degrees)            | 1.819                             |
| Ramachandran outliers (%)                        | 0.00                              |
| Protein residues                                 | 119 of 128*                       |
| Water molecules                                  | 140                               |
| Zinc ion                                         | 1                                 |
| Acetate ion                                       | 1                                 |
| β-Mercaptoethanol molecule                       | 1                                 |
| **Refinement**                                   |
| Resolution range (Å)                             | 33–1.57                           |
| R$_{work}$ (%)                                   | 14.5                              |
| R$_{free}$ (%)                                    | 34.9                              |
| r.m.s. deviation, bond length (Å)                 | 0.019                             |
| r.m.s. deviation, bond angle (degrees)            | 1.819                             |

* Includes three N-terminal residues that are cloning artifacts.
protein hexamer based on SDS-polyacrylamide gels (expected molecular mass based on migration of cleaved monomer, is 54 kDa). The oligomerization state of SO1698 in solution was further investigated by size exclusion chromatography. Under native conditions, only one major chromatography fraction is observed. This faction corresponds to a 108-kDa homo-oligomer (supplemental Fig. S2). The SO1698 D37A mutant migrates with similar apparent molecular mass (104 kDa). In solution, the apparent molecular mass of SO1698 is larger than hexamers (82.2 kDa) observed in crystals. However, because the SO1698 is donut-shaped with dimensions of 80/1100/45 Å, it may appear larger than globular proteins. Therefore, we conclude that SO1698 is a hexamer both in the crystal and in solution.

Formation of the oligomer is also pH-dependent, forming only at acidic pH. The unusual stability of oligomers under denaturing conditions of SDS-PAGE could be explained by covalent bonding of hexamers; however, we did not observe any such connections in our structure. The two N-terminal residues are not visible in the crystal structure, so we tested truncation mutants of 4, 7, and 10 N-terminal residues to exclude the possibility that these residues participate in the covalent linkage. Each of these mutants exhibited biochemical properties and crystallized in the same fashion as wild-type protein (data not shown). Additionally, we tested multiple crystals of DP-EP protein crystallizing in different morphological forms and having different space groups. All displayed the same dimeric and hexameric assemblies.

Active Site Mutagenesis—To analyze the autoproteolytic activity of the protein, we constructed eight catalytic site mutants: Y26F, D37A, K98A, S114A, P115A, D116A, P117A, and Q118A. All mutants were crystallized at various pH conditions, and their structures were determined at high resolution (range 1.25–2.45 Å). A summary of the crystallographic data is provided in supplemental Table S1. The crystal structures were analyzed for presence/absence of a bond between residues 116 and 117. A summary of the mutagenesis analysis is provided in Table 2. Two mutations (D116A and P117A) resulted in complete loss of autoproteolytic function, and one mutation (D37A) resulted in major loss of function. In addition, the S114A mutation displayed substantial reduction of the activity. The remaining mutations had no visible impact on self-cleavage.

These results suggest that the two aspartic acid residues (Asp-37 and Asp-116) may be involved in the self-cleaving reaction, as found in peptidases, where two acidic residues are utilized in the catalytic process (30–32). What we observe in the crystal structures contradicts this assumption. We suppose that the D116A structure represents the conformational state of the active site of the native protein less the catalytic carboxyl moiety. In the D116A structure, the carboxyl group of Asp-37 is ~6 Å from the amide carbon atom of Ala-116; this distance is seemingly too large for the carboxyl group of Asp-37 to play a direct role in the self-cleaving function of the protein. To understand how loss of the carboxyl group from Asp-37 leads to loss of self-cleaving function in the D37A mutant, we compared the D116A and D37A structures in Fig. 5. We observe that the mutation causes a significant reorganization of the active site.
The protein backbones were aligned on the C atoms from residue 60 to 100 (r.m.s. deviation = 1.16 Å). There is a register shift in the \( \beta \) strands in the D37A structure. In the alignment depicted in Fig. 5, the Gln-118 side chain has rotated from one side of the backbone in the D116A structure to the other in the D37A structure, hydrogen-bonding to the hydroxyl of Tyr-26 in the latter structure. The \( \mathrm{Ne}2 \) nitrogen atoms from the two Gln-118 residues are separated by 9.4 Å in this alignment. This twisting of the peptide backbone significantly alters the relative orientations of Asp-116 (Ala-116) and Pro-117 between the two structures. As a consequence, we believe that the carboxyl group of Asp-37 plays an important role in stabilizing the conformation of the active site in the native protein but does not play an active role in catalysis.

The results of enzyme loss of activity by site mutations observed in protein crystals were confirmed by SDS-PAGE, as shown in Fig. 6. Interestingly, the oligomerization reaction occurs even for mutants that do not possess the autoproteolytic function; therefore, these two activities seem independent and distinct. As can be seen in Fig. 6, the oligomerization reaction, similarly to native enzyme, is pH-dependent in the mutant proteins.

The structures of completely inactive D116A and P117A mutants are quite similar but display differences in the active site. In supplemental Fig. S3, the structures were aligned on the \( \mathrm{C} \) carbon atoms from residue 60 to 100 (r.m.s. deviation = 0.10 Å). Most side chains around the active site are found in conformations that are nearly equivalent. There is a noticeable rotation of the carbonyl oxygen atom from residue 116 between the two structures. We infer that this rotation is sufficient to make the nucleophilic attack of an activated water molecule energetically unfavorable. Therefore, Pro-117 seems to provide

**TABLE 2**

Mutagenesis analysis

| pH | Native | Y26F | D37A | K98A | S114A | P115A | D116A | P117A | Q118A |
|----|--------|------|------|------|-------|-------|-------|-------|-------|
| 4.5| +      | +    | +    | ±    | +     | −     | −     | −     | +     |
| 5.5| +      | −    | +    | ±    | +     | −     | −     | −     | +     |
| 6  | +      | +    | ±    | −    | ±     | −     | −     | −     | +     |
| 6.5| +      | −    | +    | ±    | −     | −     | −     | −     | +     |
| 7  | +      | −    | +    | ±    | −     | −     | −     | −     | +     |
| 7.5| +      | −    | +    | ±    | −     | −     | −     | −     | +     |
| 8  | +      | −    | +    | ±    | −     | −     | −     | −     | +     |
| 8.5| +      | −    | +    | ±    | −     | −     | −     | −     | +     |
| 9.5| ±      | −    | +    | ±    | −     | −     | −     | −     | −     |

**FIGURE 4.** pH dependence of DP-EP protein self-cleaving reaction. The self-cleaving function (cut/uncut) of the native protein demonstrates distinct pH dependence, with the optimum in the range of pH 6.0 – 6.2. The formation of a multimeric protein complex (hexamer) is also observed. Reaction samples in MES buffer at pH values indicated at the top of each lane were performed as described under “Experimental Procedures” and separated by SDS-PAGE. The “control” corresponds to the protein sample without incubation. The arrows indicate positions of protein oligomer and monomers, uncut (13.72 kDa) and cut (12.66 kDa).

**FIGURE 5.** Comparison of D116A and D37A mutant structures. The backbone and carbon atoms of the D116A and (D37A) mutant proteins are colored green and mauve, respectively. There is a significant reorganization of the active site; arrows indicate the movement of side chains between the structures. The Ala-116 side chain in the D116A structure is hidden beneath the Pro-117 residue from the D37A structure in this view.

The protein backbones were aligned on the \( \mathrm{C} \) atoms from residue 60 to 100 (r.m.s. deviation = 1.16 Å). There is a register shift in the \( \beta \) strands in the D37A structure. In the alignment depicted in Fig. 5, the Gln-118 side chain has rotated from one side of the backbone in the D116A structure to the other in the D37A structure, hydrogen-bonding to the hydroxyl of Tyr-26...
DUF1888 Family Self-cleaving Endopeptidase

Numerical Simulations—The mechanism of the autoproteolytic activity is not immediately discernible from the crystal structure of the native protein. The scissile bond is cleaved between the Asp-116 and Pro-117 residues, and the distance in the crystal structure between the Asp-116 carbonyl carbon and Pro-117 nitrogen atoms is 6.2 Å, leaving little indication as to the structure of the Michaelis complex for the reaction. To reconstruct the organization of the active site prior to the autoproteolytic reaction, we conducted a series of molecular dynamics simulations of the native and mutant proteins, with the anticipation that our native protein simulations should result in a structure that is similar to the D116A mutant structure. Different simulation models were constructed that included monomer, dimer, or hexamer forms of the protein, with the majority of the simulations conducted with dimers. Significant differences in binding and orienting a water molecule in the vicinity of the scissile bond were observed between the monomer and dimer models; no such differences were noted between dimer and hexamer models. Different simulation models were constructed with initial coordinates from native and mutant crystal structures. The native protein simulations should result in a structure that is similar to the D116A mutant structure. The scissile bond is cleaved between the Asp-116 and Pro-117 residues, and the distance in the crystal structure between the Asp-116 carbonyl carbon and Pro-117 nitrogen atoms is 6.2 Å, leaving little indication as to the structure of the Michaelis complex for the reaction. To reconstruct the organization of the active site prior to the autoproteolytic reaction, we conducted a series of molecular dynamics simulations of the native and mutant proteins, with the anticipation that our native protein simulations should result in a structure that is similar to the D116A mutant structure. Different simulation models were constructed that included monomer, dimer, or hexamer forms of the protein, with the majority of the simulations conducted with dimers. Significant differences in binding and orienting a water molecule in the vicinity of the scissile bond were observed between the monomer and dimer models; no such differences were noted between dimer and hexamer models. Different simulation models were constructed with initial coordinates from native and mutant crystal structures. The native protein simulations began with the peptide backbone severed, which was then healed through a lengthy equilibration process. No significant differences were observed as a result of different starting configurations. Examination of the dynamics trajectories indicated the presence of oriented water molecules in two sites near the peptide bond carbon atom of Asp-116. During the simulations, the water molecules from the two potential sites exchanged reasonably frequently (~100–300 ps) with bulk water molecules, but the sites remained occupied almost continuously. In one site, the hydroxyl group from the residue Tyr-26 was associated with orienting the water (the water molecule 236 from the D116A mutant structure 3NJK occupies a comparable position). The Y26F mutant retains its autoproteolytic function, so we conclude that the attack of the peptide backbone must originate from the other site. The other site reflects a conformation that is consistent with a mechanism for hydrolytic cleavage of the peptide backbone that is sketched in Fig. 7. (The water molecule 147 from the P117A mutant structure 3NJM occupies a roughly comparable position.) We propose that the oriented water molecule (stage 1) interacts with the side chain carboxyl group of Asp-116; the carboxyl serves as a general base and accepts a proton (stage 2). The hydroxide ion attacks the backbone carbon atom (stage 3); the remaining proton transfers to the proline ring, and the backbone is severed (stage 4). We illustrate the location of the attacking water and organization of the active site in Fig. 8, in which one frame from a 2-ns simulation of the (dimeric) native protein is compared with the non-cleaving D116A mutant structure. The two structures were aligned (r.m.s. deviation = 1.12 Å) on the Cα carbon atoms from residue 60 to 100. The conformation of the active site residues observed in the simulation mimics that observed in the mutant protein. We take this as evidence that the molecular dynamics simulations provide a realistic representation of the active site conformations.

In this particular frame, the water molecule is oriented appropriately for a nucleophilic attack on the peptide backbone. The water molecule makes hydrogen bonds with the Oδ2 oxygen atom from Asp-116 (d(O–O) = 3.2 Å) and the O oxygen atom from Pro-117 (d(O–O) = 2.7 Å). The oxygen atom from the water is also positioned about 3.1 Å from the Cα carbon atom of Asp-116, in what we would categorize as a near attack conformation (28). This conformation is consistent with the mechanism for the autoproteolytic reaction that we illustrate in Fig. 7; the oriented water molecule is activated by the carboxyl group from the Asp-116 residue and subsequently attacks the backbone carbonyl carbon. The remaining proton transfers to the N nitrogen of the prolyl leaving group, and the peptide backbone is severed.
Aspartic peptidases are widely distributed among all organisms and are involved in many important biological processes (30–32). Aspartic peptidases are potential pharmaceutical targets in human diseases, including hypertension, cancer, Alzheimer disease, malaria, and AIDS. From the MEROPS database (33), aspartic peptidases form 15 families based on their sequence similarity. They are most active at acidic pH, although some aspartic peptidases are active at neutral pH and usually do not require cofactors. The typical aspartic peptidase catalytic center is formed by two aspartate residues that activate a water molecule that conducts a nucleophilic attack on the peptide bond. Several peptidases display self-cleaving activity that results in significant conformational change in the protein. This feature is observed in viral proteins, where autocatalytic cleavage is followed by assembly of mature proteins into infectious virions (34–36). Clan AB represents the aspartyl endopeptidases that are viral coat proteins from nodavirus and tetravirus. The catalytic residue of the Family A6 endopeptidases from Clan AB is Asp-75, and the catalytic domain includes an adjacent proline that is strictly conserved but occurs on the N-terminal side of the catalytic aspartic acid (PDX). The target site for the A6 peptidases is the Asn-363–Ala-364 bond that is located on a different strand of the protein. The mechanism and structure of the A6 peptidases are quite different from the DP-EP protein, and we believe that this protein is a member of a new family of endopeptidases using a Pro–Asp–Pro motif.

Similar self-cleavage has been reported previously for human mucin MUC2 (37) and precursor of heavy chain 3 (proH3) (38). In both cases, sequence Asp-Pro was essential for autocatalysis, and the reaction was accelerated at low pH. The “clip-and-link” activity was reported for two toxins from Gram-negative pathogens (FrpC of Neisseria meningitidis and ApxIVA of Actinobacillus pleuropneumoniae) containing RTX repeats (39). These proteins can autocleave an Asp–Pro bond in a pH–calcium–dependent manner and covalently link the generated fragment to itself or to another protein molecule by a newly formed isopeptide Asp–Lys bond. Therefore, RTX proteins display similar activities to DP-EP. They can cleave their own peptide backbone in a pH-dependent fashion, they form an internal cross-link through the Lys side chain at the site of cleavage, and they form self-cross-linked oligomers that are covalently connected. However, in contrast to RTX toxins, DP-EP self-assembly is independent of the autocatalysis and is not metal ion-de-
DUF1888 Family Self-cleaving Endopeptidase

dependent. The biological role for autocleavage of toxins is unknown.

The DP-EP protein does not appear to possess exogenous peptidase activity. We were not able to find any substrate for DP-EP protein proteolysis other than the DP-EP protein itself. BLAST analysis of the protein finds homologous proteins predominantly in other strains of Shewanella, as is illustrated in Fig. 1. There are, however, similar proteins identified in Alteromonas and Alteromonadaceae (gi239996420). In all of these proteins, we observe a conserved SDXP sequence in the C-terminal region of the protein (after the final β strand in DP-EP), where the non-conserved residue X is predominantly a proline. In our studies, mutation of the leading proline residue (P115A) does not affect the autoproteolytic function. The DP-EP protein does not have any close structural homologs. The DALI server shows human coagulation factor VIII (693-amino acid protein, Protein Data Bank entry 2R7E) as the closest homolog, with a Z-score (the strength of structural similarity in S.D. values above expected) and r.m.s. deviation (of superimposed atoms) equal to 5.3 and 3.0 Å, respectively. However, visual comparison of the two structures excludes coagulation factor as a structural homolog for the DP-EP protein. Another match from DALI analysis was a human T cell receptor (Protein Data Bank entry 21AN) that shows some visual similarity due to its β sandwich structure.

The Asp-Pro sequence appears to be critical for determining the self-cleaving function of the DP-EP protein, although it is not sufficient, as evidenced by the loss of self-cleavage function in the D37A mutant. We observe that the loss of rigidity associated with the trailing proline (P117A mutant) leads to loss of self-cleaving. The Asp-116 carbonyl group rotates in the P117A structure in such a way that the oriented water molecule no longer occupies a near attack conformation, and the nucleophilic attack on the peptide backbone becomes energetically unfavorable. We observe a similar motif in the methyltransferase from E. coli, for which the crystal structure was obtained (Protein Data Bank entry 2FPO) (40). As we illustrate in supplemental Fig. S5, the catalytic domain of this enzyme contains an Asp-Pro-Pro sequence. This suggests that the DP sequence is a conserved motif for catalytic activity.

We note that the A6 family of endopeptidases is found in viral coat proteins, and the autolytic activity in those proteins is assumed to play a role in capsid assembly and maturation. It is not apparent what role self-assembly might play in the Gram-negative Shewanella, but there are several possibilities, including (i) processing still unidentified protein(s), (ii) participation in assembly of larger microcompartments in bacteria, (iii) serving as a pH sensor, or (iv) serving as a signal transducer. In known genomes, the DP-EP gene is located between transposase (or phosphorus-oxygen lyase) and two-component response transcription regulator genes, suggesting that this protein may be involved in signal transduction.

Similar autocleavage has been reported for MUC2 and proH3 as well as FrpC and ApxIVA toxins. In the latter case, an internal cross-link and self-assembly is observed, but the cellular functions of these reactions are unclear.

The DP-EP protein has been annotated as “metal binding.” In fact, a metal binding site has been found in the structure capable of coordinating several different metals (magnesium, calcium, and zinc) that were present in the crystallization solutions. Several residues contribute to this site (Gly-71 (main chain carbonyl), Ile-72 (main chain carbonyl), and Asp-74, 89, 91, and Thr-93 (Protein Data Bank entries 3NJH, 3NJ5, and 3NJL)). The metal binding site is 22.7 Å from the catalytic Asp residue and is 12.0 Å from the C-terminal peptide that is being cleaved. Metal binding may contribute to stability in this region of the protein; however, in contrast to FrpC and ApxIVA toxins, the DP-EP self-cleavage is not metal-dependent. These various hypotheses about cellular and biochemical functions of DP-EP and other members of DUF1888 require further experimental investigation.

Acknowledgments—We thank the members of the Midwest Center for Structural Biology for support, in particular Frank Collart, James Adibullah, and Garrett Cobb for efforts in cloning proteins, and Marat Valley at Pacific Northwest National Laboratory for help with NWChem. We acknowledge the Texas Advanced Computing Center at the University of Texas at Austin for providing high performance computing resources that have contributed to the research results reported within this paper.

REFERENCES

1. Heidelberg, J. F., Paulsen, I. T., Nelson, K. E., Gaidos, E. J., Nelson, W. C., Read, T. D., Eisen, J. A., Sheshadri, R., Ward, N., Methé, B., Clayton, R. A., Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S., DeBoy, R. T., Dodson, R. J., Durkin, A. S., Haft, D. H., Kolonay, J. F., Madupu, R., Peterson, J. D., Umayam, L. A., White, O., Wolf, A. M., Vamathevan, J., Weidman, J., Impraim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T. R., McDonald, L., A., Feld­blyum, T. V., Smith, H. O., Venter, J. C., Nealon, K. H. & Fraser, C. M. (2002) Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis. Nat. Biotechnol. 20, 1118–1123

2. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrett, D., Bate­man, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R. R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Lonsdale, D., Silventoinen, V., Orchard, S. E., Pagni, M., Peyruc, D., Ponting, C. P., Selengut, J. D., Servant, F., Sigrist, C. J., Vaughan, R. & Zdobnov, E. M. (2003) The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res. 31, 315–318

3. Holm, L. & Rosenström, P. (2010) DALI server. Conservation mapping in 3D. Nucleic Acids Res. 38, W545–W549

4. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript. Analysis of multiple sequence alignments in PostScript. Bioinformatics 15, 305–308

5. Stols, L., Gu, M., Dieckmann, L., Raffen, R., Collart, F. R. & Donnelly, M. I. (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr. Purif. 25, 8–15

6. Dieckmann, L., Gu, M., Stols, L., Donnelly, M. I. & Collart, F. R. (2002) High throughput methods for gene cloning and expression. Protein Expr. Purif. 25, 1–7

7. Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R. & Joachimiak, A. (1999) Taking MAD to the extreme. Ultrafast protein structure determination. Acta Crystallogr. D Biol. Crystallogr. 55, 1168–1173

8. Kim, Y., Dementieva, I., Zhou, M., Wu, R., Lezondra, L., Quartery, P., Joachimiak, G., Korolev, O., Li, H. & Joachimiak, A. (2004) Automation of protein purification for structural genomics. J. Struct. Funct. Genomics 5, 111–1121

9. Otwinowski, Z., Borek, D., Majewsli, W. & Minor, W. (2003) Multipara-
metric scaling of diffraction intensities. *Acta Crystallogr.* A **59**, 228–234
10. Schneider, T. R. & Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779
11. Terviliger, T. C. (2003) SOLVE and RESOLVE. Automated structure solution and density modification. *Methods Enzymol.* **374**, 22–37
12. Morris, R. J., Perrakis, A. & Lamin, V. S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244
13. Emsley, P. & Cowtan, K. (2004) Coot. Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
14. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
15. Collaborative Computational Project, Number 4 (1994) The CCP4 suite. Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–776
16. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996) AQUA and PROCHECK-NMR. Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410
18. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802
19. Mackerrel, A. D., Basford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Waterman, M., Wierkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem.* **B 102**, 3586–3616
20. Humphrey, W., Dalke, A. & Schulten, K. (1996) VMD. Visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38
21. Esmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593
22. Brünger, A., Brooks, C. L. & Karplus, M. (1984) Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105**, 495–500
23. Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. (1995) Constant pressure molecular dynamics simulation. The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621
24. Hoover, W. G. (1985) Canonical dynamics. Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697
25. Hoover, W. G. (1986) Constant pressure equations of motion. *Phys. Rev. A* **34**, 2499–2500
26. Kendall, R. A., Apra, E., Bernholdt, D. E., Bylaska, E. J., Dupuis, M., Fann, G. I., Harrison, R. J., Ju, J. L., Nichols, J. A., Nieplocha, J., Staatsma, T. P., Windus, T. L. & Wong, A. T. (2000) High performance computational chemistry. An overview of NWChem a distributed parallel application. *Comput. Phys. Commun.* **128**, 260–283
27. Vallev, M., Yang, J., Adams, J. A., Taylor, S. S. & Weare, J. H. (2007) Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations. *J. Phys. Chem. B* **111**, 13455–13464
28. Bruce, T. C. & Benkovic, S. J. (2000) Chemical basis for enzyme catalysis. *Biochemistry* **39**, 6267–6274
29. Becke, A. D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100
30. Dunn, B. M. (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.* **102**, 4431–4458
31. Davies, D. R. (1990) The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 189–215
32. Cooper, J. B. (2002) Aspartic proteinases in disease. A structural perspective. *Curr. Drug Targets* **3**, 155–173
33. Rawlings, N. D., Barrett, A. J. & Bateman, A. (2010) MEROPS. The peptidase database. *Nucleic Acids Res.* **38**, D227–D233
34. Gallagher, T. M. & Rueckert, R. R. (1988) Assembly-dependent maturation cleavage in provirions of a smallicosahedral insect ribovirus. *J. Virol.* **62**, 3399–3406
35. Schneemann, A., Zhong, W., Gallagher, T. M. & Rueckert, R. R. (1992) Maturation cleavage required for infectivity of a nodavirus. *J. Virol.* **66**, 6728–6734
36. Wery, J. P., Reddy, V. S., Hosur, M. V. & Johnson, J. E. (1994) The refined three-dimensional structure of an insect virus at 2.8 Å resolution. *J. Mol. Biol.* **235**, 565–586
37. Lidell, M. E., Johansson, E. V. & Hansson, G. C. (2003) An autocatalytic cleavage in C terminus of the human MUC2 mucin occurs at the low pH of the late secretory pathway. *J. Biol. Chem.* **278**, 13944–13951
38. Thueson, M. & Fries, E. (2000) The low pH in trans-Golgi triggers autocatalytic cleavage of pre-α-inhibitor heavy chain precursor. *J. Biol. Chem.* **275**, 30996–31000
39. Osicka, R., Procházková, K., Sule, M., Linhartová, I., Havlíček, V. & Sebo, I. (2004) A novel “clip-and-link” activity of repeat in toxin (RTX) proteins from Gram-negative pathogens. *J. Biol. Chem.* **279**, 24944–24956
40. Lesnyak, D. V., Osipiuk, J., Skarina, T., Sergiev, P. V., Bogdanov, A. A., Edwards, A., Savchenko, A., Joachimiak, A. and Dontsova, O. A. (2007) Methyltransferase that modifies guanine 966 of the 16 S rRNA. *J. Biol. Chem.* **282**, 5880–5887