Rapid Reports

Solution NMR Structure of the NlpC/P60 Domain of Lipoprotein Spr from *Escherichia coli*: Structural Evidence for a Novel Cysteine Peptidase Catalytic Triad[†]

James M. Aramini,[*,‡,||] Paolo Rossi,[‡,||] Yuanpeng J. Huang,[‡,||] Li Zhao,[‡,||] Mei Jiang,[‡,||] Melissa Maglaqui,[‡,||] Rong Xiao,[‡,||] Jessica Locke,[‡,||] Rajesh Nair,[⊥,||] Burkhard Rost,[⊥,||] Thomas B. Acton,[‡,||] Masayori Inouye,[§,||] and Gaetano T. Montelione[*,‡,§,||]

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854, and Northeast Structural Genomics Consortium

Received June 6, 2008; Revised Manuscript Received July 22, 2008

ABSTRACT: *Escherichia coli* Spr is a membrane-anchored cell wall hydrolase. The solution NMR structure of the C-terminal NlpC/P60 domain of *E. coli* Spr described here reveals that the protein adopts a papain-like $\alpha+\beta$ fold and identifies a substrate-binding cleft featuring several highly conserved residues. The active site features a novel Cys-His-His catalytic triad that appears to be a unique structural signature of this cysteine peptidase family. Moreover, the relative orientation of these catalytic residues is similar to that observed in the analogous Ser-His-His triad, a variant of the classic Ser-His-Asp charge relay system, suggesting the convergent evolution of a catalytic mechanism in quite distinct peptidase families.

Bacterial lipoproteins make up a diverse class of membrane-associated proteins that play important roles in a wide range of biological processes and in bacterial pathogenesis (*1*). In general, lipoproteins are comprised of an N-terminal signal peptide sequence that is cleaved within a conserved lipobox at an invariant cysteine, which in turn is covalently anchored to the bacterial membrane via a diacylglycerol thioether linkage (*1*−*3*). The *spr* gene of *Escherichia coli* encodes a 188-residue precursor of a predicted periplasmic surface outer membrane lipoprotein (SWISS-PROT entry SPR_ECOLI; NESG target entry ER541), consisting of a 26-residue N-terminal signal peptidase II recognition sequence culminating in the membrane anchoring cysteine. Its C-terminal domain is a member of the large NlpC/P60 protein domain family (*4*) (Pfam entry PF00877; MEROPS peptidase clan CA, family C40) comprised of more than 1300 sequences (Pfam 22.0) predominantly from bacteria and featuring both secreted and predicted membrane-bound lipoproteins. While the exact substrate specificity of *E. coli* Spr has not yet been established, the C40 peptidase/NlpC/P60 protein domain families are classified as γ-D-glutamyl-L-diamino acid-endopeptidases which hydrolyze specific peptide linkages in bacterial cell walls and are intimately involved in cell wall hydrolysis during cell growth and division or cell lysis/invasion (*4, 5*). Accordingly, members of this class of peptidases are potential targets for antibiotic drug discovery.

Here, we present the solution NMR structure of the C-terminal 126-residue NlpC/P60 domain of *E. coli* Spr, Spr[37−162] (molecular mass of 14.4 kDa, pI 10.2), which corresponds to residues 63−188 in the Spr precursor (Figure 1A). The protein adopts a fold whose scaffold is common to the papain superfamily of cysteine peptidases (*4*). The highly conserved active site of the protein reveals a novel Cys-His-His catalytic triad that appears to be unique to the NlpC/P60 family of enzymes, present in most members of this domain family, including the unpublished structures of two homologues from cyanobacteria, *Nostoc punctiforme* protein 53686717 and *Anabaena variabilis* Q3M7N3 (Figure 1A). To the best of our knowledge, this report represents the first structural description of an active site Cys-His-His catalytic triad in a peptidase, and the first detailed experimental structural perspective on this large and biologically important protein domain family.

The solution NMR structure determination of *E. coli* Spr[37−162] (PDB entry 2K1G; BMRB entry 15603), including the cloning, expression, and purification of ^{13}C- and ^{15}N-enriched protein samples, was performed following standard protocols of the Northeast Structural Genomics (NESG) Consortium (*6*) (see the Supporting Information for a complete description of the methods used in this work, as well as a summary of the NMR data and structural statistics for this study). The protein is monomeric in solution, based on gel filtration chromatography, static light scattering, and ^{15}N relaxation data. The structure of *E. coli* Spr[37−162]

[†] This work was supported by a grant from the National Institute of General Medical Sciences Protein Structure Initiative (U54-GM074958).

* To whom correspondence should be addressed. Telephone: (732) 235-5321. Fax: (732) 235-5633. E-mail: jma@cabm.rutgers.edu and guy@cabm.rutgers.edu.

[‡] Rutgers, The State University of New Jersey.

[⊥] Columbia University.

[§] University of Medicine and Dentistry of New Jersey.

[||] Northeast Structural Genomics Consortium.

10.1021/bi8010779 CCC: $40.75
Published on Web 08/21/2008

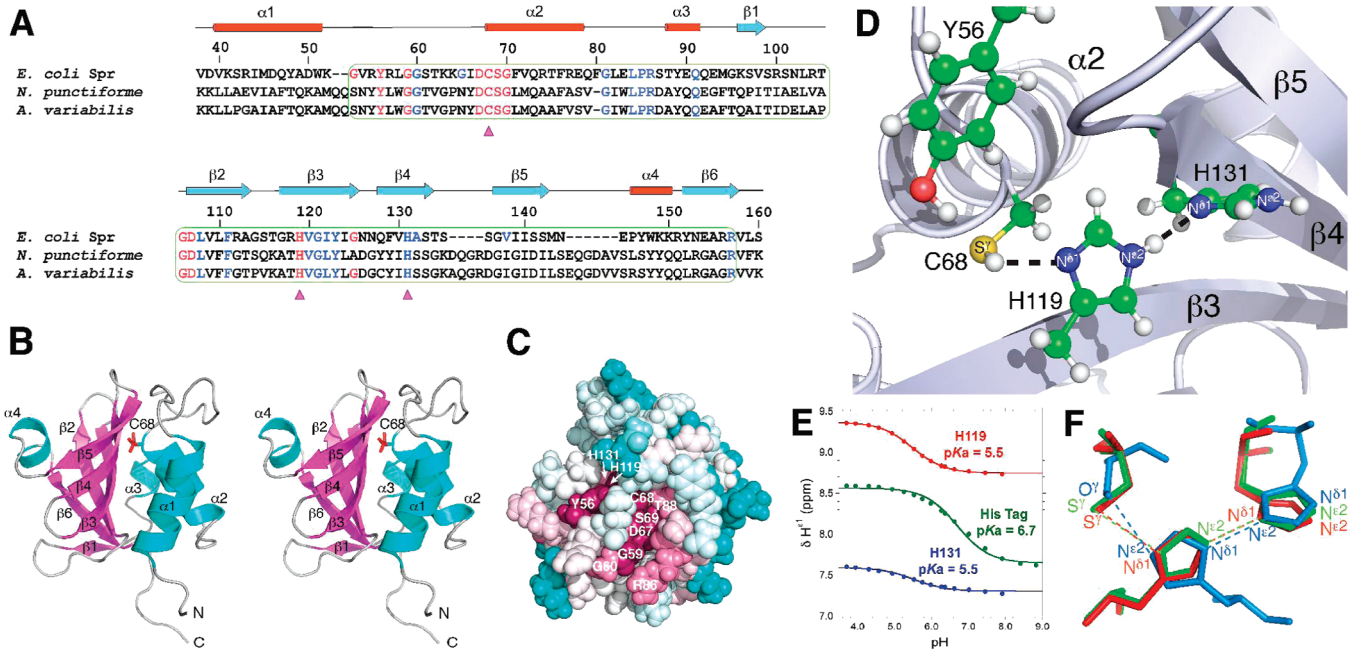

FIGURE 1: (A) Structure-based sequence alignment (*8*) of the regions encompassing the NlpC/P60 domains (outlined in green) in *E. coli* Spr (residues 38−160), *N. punctiforme* protein 53686717 (residues 94−227), and *A. variabilis* Q3M7N3 (residues 94−227). The sequence numbering for mature membrane-anchored *E. coli* Spr (corresponding to residues 63−188 in the full-length precursor) and the secondary structural elements found in its solution NMR structure (PDB entry 2K1G) are shown above the alignment. Amino acid residues in *E. coli* Spr conserved in >80 and >60% of the entire NlpC/P60 (PF00877) protein domain family (1349 sequences; Pfam release 22.0) are colored red and blue, respectively. Residues in the Cys-His-His catalytic triad are denoted below the alignment with magenta triangles. (B) Ribbon stereoview of the lowest-energy (CNS) conformer from the final solution NMR structure ensemble of *E. coli* Spr[37−162]. The α-helices and *β*-strands are colored cyan and magenta, respectively. The side chain of the catalytic cysteine (C68) is shown. (C) Consurf (*7*) image showing the conserved residues in the active site of *E. coli* Spr[37−162]. Residue coloring, reflecting the degree of residue conservation over the entire NlpC/P60 protein domain family, ranges from magenta (highly conserved) to cyan (variable). (D) View into the active site of *E. coli* Spr[37−162] showing the conserved Cys-His-His catalytic triad and flanking tyrosine. Juxtaposed heavy atoms in the triad and secondary structural elements are labeled. (E) Plots of histidine $H^{\varepsilon 1}$ chemical shift as a function of pH for U-^{13}C,^{15}N *E. coli* Spr[37−162] at 298 K, obtained by two-dimensional (2D) $^{1}H-^{13}C$ HSQC NMR: red for H119, blue for H131, and green for the C-terminal hexa-His tag (control). pK_a values for each are given. (F) Superposition of the active site residues of *E. coli* Spr[37−162] (PDB entry 2K1G; red), *N. punctiforme* protein 53686717 (PDB entry 2EVR; green), and human cytomegalovirus serine peptidase (PDB entry 1WPO; blue) (*18*). The key side chain atoms involved in these catalytic triads are labeled. For the superposition, the C^{α} and S^{γ} atoms of C68, the $N^{\delta 1}$ atom of H119, and the $N^{\delta 1}$ atom of H131 in the *E. coli* Spr[37−162] structure were superimposed with the equivalent atoms in the other structures. Hydrogen bond distances: $S^{\gamma}-N^{\delta 1}$, 3.83 ± 0.46 Å for 2K1G, 3.60 Å for 2EVR; $N^{\varepsilon 2}-N^{\delta 1}$, 2.81 ± 0.05 Å for 2K1G, 2.71 Å for 2EVR; $O^{\gamma}-N^{\varepsilon 2}$, 3.07 Å for 1WPO; $N^{\delta 1}-N^{\varepsilon 2}$, 3.09 Å for 1WPO. All structure figures were made using PyMol 1.1 (http://www.pymol.org).

adopts a papain-like α+*β* fold comprised of four α-helices and a sheet of six antiparallel *β*-strands arranged in an ααα*ββββββαβ* topology (Figure 1B). As predicted for the entire NlpC/P60 domain family (*4*), a highly conserved catalytic cysteine (C68) occurs at the end of a helix (α2) and is packed against a *β*-sheet core featuring a conserved histidine (H119) from a *β*-strand (*β*3). A ConSurf (*7*) analysis of surface features conserved across the NlpC/P60 family demonstrates that highly conserved residues are clustered in a large groove, clearly identifying the active site and substrate binding site of this class of enzymes (Figure 1C). The groove is lined with numerous conserved polar and charged residues, culminating with the catalytic C68 and partially buried H119. Close inspection of the active site (Figure 1D) confirms that the predicted (*4*) third polar residue in the catalytic triad is indeed a second histidine in an adjacent strand (H131 in *β*4) that is conserved in more than 60% of the sequences in the NlpC/P60 domain family, where the frequency of polar residues at this sequence position follows a descending trend: His ≫ Asn > Glu > Gln > Asp. The phenol group of a highly (≈ 90%) conserved tyrosine (Y56) also juts into the active site and may play a role in modulating the nucleophilicity of C68 S^{γ} and/or binding to the substrate and resulting anionic tetrahedral intermediate during catalysis. Although this function in peptidases is often performed by backbone and side chain (Asn/Gln) amide moieties, there is precedent in the literature for stabilization of the oxyanion intermediate by a tyrosine OH group in certain serine peptidases, including prolyl oligopeptidase (*9*). We propose that the active catalytic network in this enzyme comprises the thiol group of C68, $N^{\delta 1}$ of H119, $H^{\varepsilon 2}$ of H119, and $N^{\delta 1}$ of H131, with both histidines adopting the $N^{\varepsilon 2}H$ tautomeric state revealed by NMR data. In this arrangement, H119 can act as a general base in catalysis, with the second histidine in the triad, H131, serving to properly orient the side chain of H119. A hydrogen bonding network stabilizing neutral (imidazole base) histidines, as shown in Figure 1D, would be expected to exhibit reduced histidine pK_a values. Indeed, pH titrations of *E. coli* Spr[37−162] monitored by NMR reveal pK_a values of 5.5 for both active site histidines (Figure 1E), compared with a pK_a of 6.7 for the histidines in the unstructured hexa-His tag, confirming our picture of this novel catalytic triad at neutral pH.

The structure of the NlpC/P60 domain of *E. coli* Spr presented here is quite similar to structures of homologous domains in two cell wall hydrolases from cyanobacteria, namely, *N. punctiforme* protein 53686717 (PDB entry 2EVR) and *A. variabilis* Q3M7N3 (PDB entry 2HBW), which are

81% identical in sequence to each other. Structural alignments by Dali (*10*) reveal significant structural similarity between the NMR structure of *E. coli* Spr[37−162] and these two crystal structures (Dali *Z* scores of 15.3 for 2EVR and 14.8 for 2HBW; C^{α} rmsds of 2.3 Å for 2EVR and 2.3 Å for 2HBW; sequence identities of 27% for 2EVR and 28% for 2HBW). These 234-residue proteins feature a separate N-terminal SH3-like β-barrel, followed by an NlpC/P60 domain which superimposes very well with the structure of *E. coli* Spr[37−162], in spite of the relatively low level of sequence identity between these bacterial proteins from two distinct phyla. Moreover, the relative orientations of the residues in the catalytic triad are practically identical in the three structures.

The *E. coli* Spr[37−162] structure is also distantly related to two CHAP (cysteine, histidine-dependent amidohydrolases/peptidases) domain family (*11*) (PF05257; MEROPS clan CA, family C51) cysteine peptidases both featuring a Cys-His-Glu catalytic triad, namely, the N-terminal amidase domain from *E. coli* glutathionylspermidine synthetase (PDB entry 2IOB) (*12*) and a secretory antigen from *Staphylococcus saprophyticus* (PDB entry 2K3A) (*13*) (Dali *Z* scores of 7.3 for 2IOB and 5.8 for 2K3A; C^{α} rmsds of 2.9 Å for 2IOB and 5.1 Å for 2K3A; sequence identities of 13% for 2IOB and 9% for 2K3A). In terms of modeling leverage, defined elsewhere (*14*), the *E. coli* Spr[37−162] structure has a total modeling leverage value of 482 structural models, and a novel leverage value of 30 models (based on UniProt release 12.8; PSI Blast $E < 10^{-10}$).

In conclusion, we have described the solution NMR structure of the NlpC/P60 domain of *E. coli* Spr and have identified a novel Cys-His-His catalytic triad in the active site of this cysteine peptidase. To the best of our knowledge, this is the first literature example of this constellation of Cys-His-His active site residues in cysteine peptidases. While β-ketoacyl-acyl carrier protein (ACP) synthases, so-called CHH enzymes, also feature one cysteine and two histidine active site residues (*15, 16*), the overall protein fold and topology as well as the geometric distribution of the histidines about the cysteine (both point toward the nucleophile) are quite different from those found in the NlpC/P60 domain family (not shown). Interestingly, the analogous rare Ser-His-His catalytic triad, a variation of the classic Ser-His-Asp paradigm first discovered in α-chymotrypsin four decades ago (*17*), exists in the homodimeric human cytomegalovirus serine peptidase (Pfam entry PF00716; MEROPS clan SH, family S21) (*18, 19*). In spite of the very different overall fold and opposite orientations of the histidine imidazole rings compared to the structure reported here (i.e., both histidines interact with the preceding residue in the triad via $N^{\varepsilon 2}$, not $N^{\delta 1}$), the geometries of the catalytic residues are remarkably similar in these Cys-His-His and Ser-His-His triads (Figure 1F). Hence, this appears to be an example of convergent evolution of a catalytic mechanism in disparate peptidase clans. Taken together with the structures for its two distantly related homologues from cyanobacteria, our structure provides a framework for future mutagenesis and biochemical studies on the key residues in the proposed active site (in particular, C68, H119, H131, and Y56), to shed further light on the mechanism of action of this large class of biologically important cell wall hydrolases.

ACKNOWLEDGMENT

We thank G. V. T. Swapna, John Everett, Markus Fischer, and Alex Eletsky for valuable scientific discussion and insightful correspondence.

SUPPORTING INFORMATION AVAILABLE

Complete experimental methods used in this study, table of NMR and structural statistics (Table S1), NMR sequential connectivity map (Figure S1), static light scattering data (Figure S2), ^{15}N T_1 and T_2 relaxation data (Figure S3), stereoview of the final structural ensemble (Figure S4A), electrostatic surface potential of the lowest-energy conformer (Figure S4B), $^{1}H-^{15}N$ HMQC spectrum of Spr[37−162] (Figure S5), and superposition of the Spr[37−162] solution structure with its closest structural homologue (Figure S6). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Babu, M. M., Priya, M. L., Selvan, A. T., Madera, M., Gough, J., Aravind, L., and Sankaran, K. (2006) *J. Bacteriol. 188*, 2761–2773.
2. Juncker, A. S., Willenbrock, H., von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003) *Protein Sci. 12*, 1652–1662.
3. Tokuda, H., and Matsuyama, S. (2004) *Biochim. Biophys. Acta 1693*, 5–13.
4. Anantharaman, V., and Aravind, L. (2003) *Genome Biol. 4*, R11.
5. Layec, S., Decaris, B., and Leblond-Bourget, N. (2008) *J. Mol. Microbiol. Biotechnol. 14*, 31–40.
6. Acton, T. B., Gunsalus, K. C., Xiao, R., Ma, L.-C., Aramini, J., Baran, M. C., Chiang, Y.-W., Climent, T., Cooper, B., Denissova, N. G., Douglas, S. M., Everett, J. K., Ho, C. K., Macapagal, D., Rajan, P. K., Shastry, R., Shih, L.-Y., Swapna, G. V. T., Wilson, M., Wu, M., Gerstein, M., Inouye, M., Hunt, J. F., and Montelione, G. T. (2005) *Methods Enzymol. 394*, 210–243.
7. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003) *Bioinformatics 19*, 163–164.
8. Shindyalov, I. N., and Bourne, P. E. (1998) *Protein Eng. 11*, 739–747.
9. Fülöp, V., Böcskei, Z., and Polgár, L. (1998) *Cell 94*, 161–170.
10. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123–138.
11. Bateman, A., and Rawlings, N. D. (2003) *Trends Biochem. Sci. 28*, 234–237.
12. Pai, C.-H., Chiang, B.-Y., Ko, T.-P., Chou, C.-C., Chong, C.-M., Yen, F.-J., Chen, S., Coward, J. K., Wang, A. H.-J., and Lin, C.-H. (2006) *EMBO J. 25*, 5970–5982.
13. Rossi, P., Aramini, J. M., Xiao, R., Chen, C. X., Nwosu, C., Owens, L. A., Maglaqui, M., Nair, R., Fischer, M., Acton, T. B., Honig, B., Rost, B., and Montelione, G. T. (2008) *Proteins* (in press).
14. Liu, J., Montelione, G. T., and Rost, B. (2007) *Nat. Biotechnol. 25*, 849–851.
15. Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., and Lindqvist, Y. (1998) *EMBO J. 17*, 1183–1191.
16. Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., Lindquist, Y., and Larsen, S. (1999) *FEBS Lett. 460*, 46–52.
17. Matthews, B. W., Sigler, P. B., Henderson, R., and Blow, D. M. (1967) *Nature 214*, 652–656.
18. Tong, L., Qian, C., Massariol, M.-J., Bonneau, P. R., Cordingley, M. G., and Lagacé, L. (1996) *Nature 383*, 272–275.
19. Chen, P., Tsuge, H., Almassy, R. J., Gribskov, C. L., Katoh, S., Vanderpool, D. L., Margosiak, S. A., Pinko, C., Matthews, D. A., and Kan, C.-C. (1996) *Cell 86*, 835–843.

BI8010779