Journal of the Korean Magnetic Resonance Society **2015**, *19*, *36*-41
DOI 10.6564/JKMRS.2015.19.1.036

Backbone NMR Assignments of an Uncharacterized Protein, SF1002 from *Shigella flexneri* 5a M90T

Yoo-Sup Lee[1], Won-Su Yoon[1], Il Chung[2], Ka Young Chung[3], Hyung-Sik Won[4] and Min-Duk Seo[1,2,*]

[1]Department of Molecular Science and Technology, Ajou University, Suwon, Gyeonggi 443-749, Republic of Korea
[2]College of Pharmacy, Ajou University, Suwon, Gyeonggi 443-749, Republic of Korea
[3]School of Pharmacy, Sungkyunkwan University, Suwon, Gyeonggi 440-746, Republic of Korea
[4]Department of Biotechnology, Research Institute (RIBHS) and College of Biomedical and Health Science, Konkuk University, Chungju, Chungbuk 380-701, Republic of Korea

Received April 12, 2015; Revised May 18, 2015; Accepted May 31, 2015

Abstract The causative agent of shigellosis, *Shigella flexneri*, is a Gram-negative anaerobic bacterial pathogen that causes one of the most infectious bacterial dysenteries in humans. It originates infection by invading cells of the colonic epithelium using a type III secretion system. Despite *S. flexneri* is closely linked with the human disease, structural study is very deficient. Here, we have initiated NMR study of SF1002 which is the uncharacterized protein from *S. flexneri* strain 5a M90T. Based on a series of triple resonance spectra, sequence-specific assignments of the backbone amide resonances of SF1002 could be completed. This NMR study would contribute to the structural genomics of *S.flexneri*.

Keywords *Shigella flexneri*, backbone NMR assignments, structural genomics

Introduction

Members of the genus *Shigella* are gram-negative, facultative anaerobic bacteria that cause one of the most infectious of bacterial dysenteries, shigellosis.[1] These bacteria are physiologically similar to *Escherichia coli* which are non-motile, non-spore forming, rod-shaped bacteria.[2] This genus contains four species, including *Shigella flexneri*, which itself comprises 14 different serotypes and subserotypes. *Shigella* species invade the colonic and rectal epithelium in human including many other primates, causing the acute mucosal inflammation. In particular, *S. flexneri* causes more fatalities than any other *Shigella* species in most developing countries.[3] Antibiotics can be widely used to treat shigellosis by the reduction of the bacterial secretion. However, *S. flexneri* is increasingly becoming antibiotic resistant recently.[4-5]

The genomes of most *Shigella* strains, including *S. flexneri*, have been completely sequenced.[6] However, structural studies of proteins from *S. flexneri* are very deficient for their importance. In the present study, we selected SF1002 as a target which is the uncharacterized protein from *S. flexneri* strain M90T. SF1002 consists of 75 amino acid residues with a molecular weight of 8.524 kDa. The determination of three-dimensional structure of SF1002 would be

* Address correspondence to: **Min-Duk Seo,** Department of Molecular Science and Technology, Ajou University, Suwon, Gyeonggi 443-749, Republic of Korea, Tel: 82-31-219-3450; Fax: 82-31-219-3435; E-mail: mdseo@ajou.ac.kr

essential for its functional identification and will contribute to the structural genomics of *S. flexneri*. Furthermore, structural study of SF1002 can contribute to developing the antibiotics against shigellosis caused by infection of *shigella*. Here, we report the backbone resonance assignments of SF1002, as the first step of NMR-based structural investigation.

Experimental Methods

Gene cloning- The gene encoding SF1002 was amplified by polymerase chain reaction (PCR) using the genomic DNA of *S. flexneri* strain 5a M90T as a template. The forward and reverse oligonucleotide primers were 5'-G GAA TTC **CAT ATG** CCA ACT CAA GAA GCG-3' and 5'-CCG CCG **CTC GAG** TTA AAC GTT TTT ACG TT-3', respectively, where the *Nde*I and *Xho*I restriction enzyme cleavage sites are shown in bold. The reverse primer did not contain a stop codon for convenient target purification with C-term hexa-histidine tag. The amplified PCR products were cleaved by restriction enzyme (*Nde*I and *Xho*I) and ligated into same enzyme cleaved expression vector pET21a (Novagen). The recombinant plasmid was confirmed by DNA sequencing and transformed into *Escherichia coli* strain BL21(DE3) codon plus competent cells.

Expression and purification- The cells were grown at 37℃ in M9 minimal medium, which was supplemented with [^{15}N]NH_4Cl and [^{13}C]glucose, as the sole source of nitrogen and carbon, respectively, to produce the isotope-[^{13}C/^{15}N]-enriched protein. When the A_{600} of cell growth reached about 0.6, protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.5 mM. After 4 hrs induction, cells were harvested by centrifugation and resuspended in lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM Imidazole, 5% glycerol, 10 mM β-mercaptoethanol, 0.2% NP40, pH 7.2). The cells were disrupted by sonication on ice and supernatant was loaded onto a HisTrap FF column (GE Healthcare) pre-equilibrated with lysis buffer. The bound protein was eluted using a gradient from 10 to 500 mM imidazole in the same buffer. Fractions containing SF1002, were concentrated to about 2 ml and loaded onto a HiLoad 16/60 Superdex 75 column (GE Healthcare) that had been equilibrated with the final buffer (20 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, pH 6.5). Finally, the purified solution was concentrated to 1.3 mM for NMR measurements, as estimated using the molar absorptivity (17990 $M^{-1}cm^{-1}$ at 280 nm) predicted from the amino acid sequence. The purified protein was judged to be >98% pure by SDS-PAGE (Figure 1) and the monomeric state of the purified SF1002 was confirmed by the gel-filtration analysis.[6-7]

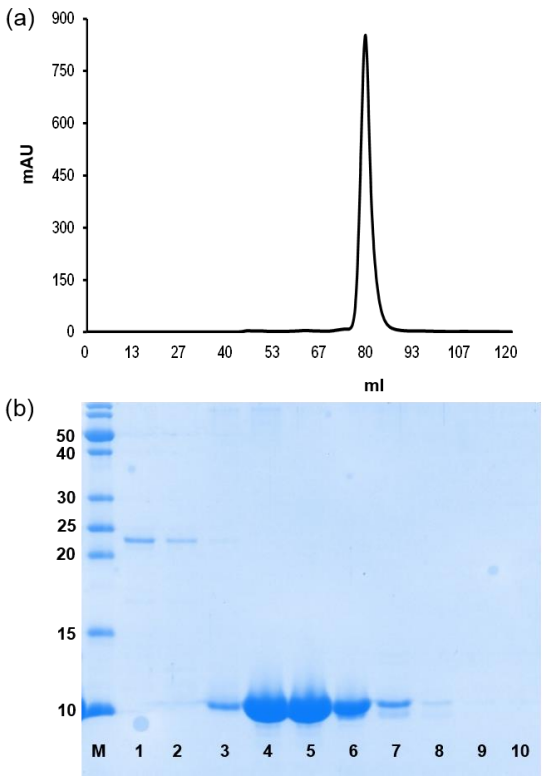

Figure 1. Purification of SF1002. (a) Elution profile of SF1002 from a Hiload 16/60 Superdex 75 column (43 ml void volume and 120 ml total bed volume). (b) SDS-PAGE of SF1002 in peak fractions after gel-permeation chromatography. Lane 5 corresponds to the highest peak fraction. Lane 4 to 6 were collected to ensure protein purity.

Lane M contains the molecular weight markers (labelled in kDa).

NMR experiments and backbone assignment- Conventional 2D-[^{1}H/^{15}N]HSQC and a series of triple resonance spectra {HNCACB, HN(CO)CACB, HNCO and HN(CA)CO} of the [^{13}C/^{15}N]SF1002 were acquired at 298 K on a Bruker Biospin Avance 600 spectrometer equipped with a cryoprobe. All NMR spectra were processed using NMRPipe/NMRDraw software and analyzed with NMRView J program. Chemical shifts were referenced indirectly using TALOS server (spin.niddk.nih.gov/bax/software/TALOS) for ^{1}H, ^{15}N and ^{13}C.[8]

Results and Discussion

About 29 mg of the purified [^{15}N/^{13}C] SF1002 was finally obtained from 1 L culture. In order to obtain the optimal NMR spectrum, 2D-[^{1}H/^{15}N] HSQC spectra of SF1002 were measured at three different pHs (8.4, 7.2 and 6.5) at 298 K. Spectral dispersion was significantly increased at lower pH (data not shown). As showing in Figure 2, the 2D-[^{1}H/^{15}N] HSQC spectrum of SF1002 was obtained at pH 6.5 with a well-dispersion, indicating that the protein is well structured.

For sequence-specific peak assignments of the backbone amide resonances in the 2D-[^{1}H/^{15}N] HSQC spectrum, we first verified peak clusters of individual resonances by collecting related peaks in the triple resonance spectra. Because all of the backbone NMR signals originating from the same residue appear the same ^{1}H^{N} and ^{15}N chemical shifts in individual spectra, the corresponding peaks could be combined into a peak cluster. As an example for the A27 residue in Figure 3, a single peak cluster is composed of ideally 10 peaks with the same amide ^{1}H^{N}/^{15}N chemical shift in the 2D-[^{1}H/^{15}N] HSQC and four kinds of triple resonance spectra. Thus, each peak cluster additionally contains chemical shift information about the inter- and intra-residue carbon atoms: ^{13}Cα(i), ^{13}Cα(i-1), ^{13}Cβ(i) and ^{13}Cβ(i-1) from the HNCACB spectrum, ^{13}Cα(i-1) and ^{13}Cβ(i-1) from the HN(CO)CACB spectrum, ^{13}CO(i) and ^{13}CO(i-1) from the HN(CA)CO spectrum, and ^{13}C(i-1) from HNCO spectrum. The ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shift values were also useful for prediction of the spin systems of individual clusters.

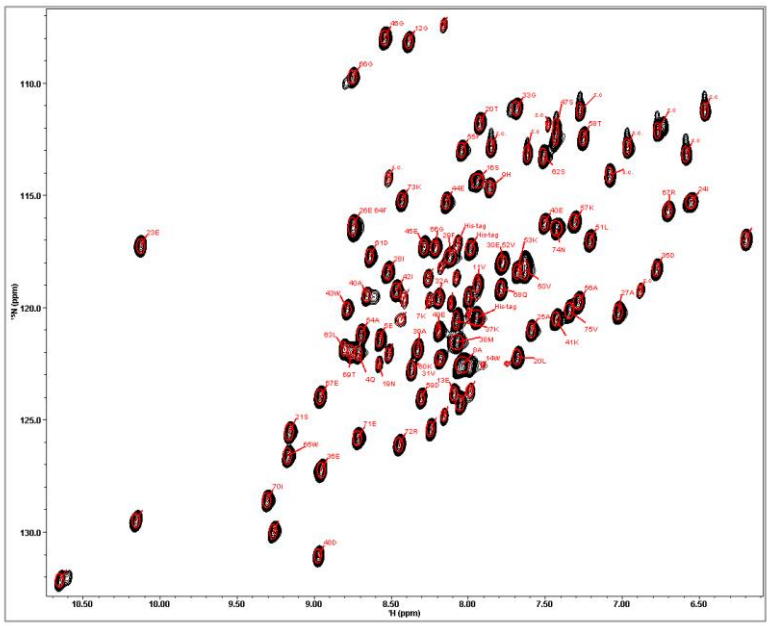

Figure 2. 2D-[^{1}H/^{15}N] HSQC spectrum of SF1002. Each resonance in the spectrum is labeled with assigned amino acid residues.

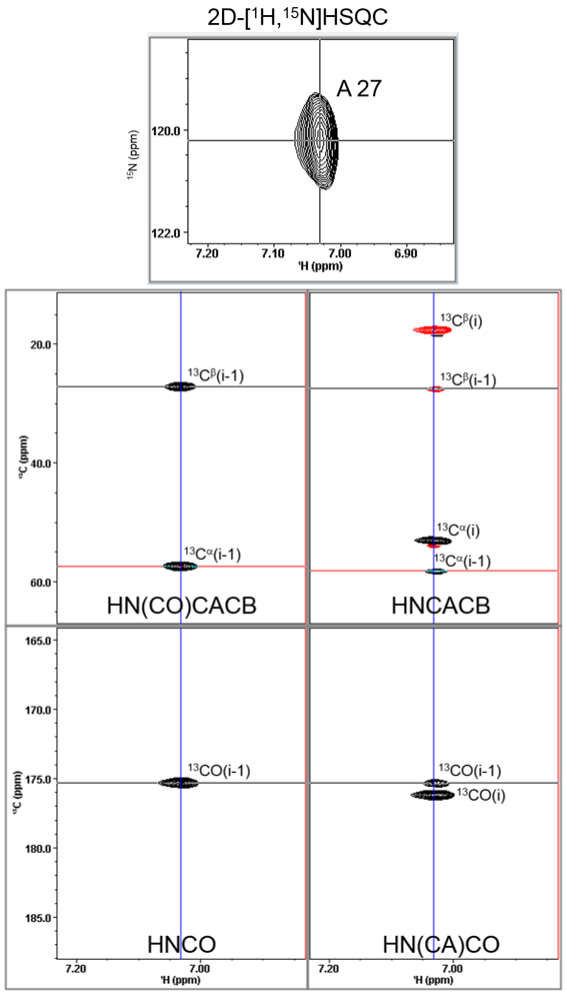

Figure 3. Verification of a peak cluster. A selected residue taken from the 2D-[^{1}H/^{15}N] HSQC spectrum (Figure 2) was enlarged in upper panel and the other strip plot for the triple resonance spectra were taken from the slices with the ^{15}N chemical shift. The x-axis of the individual spectrum represents the ^{1}H chemical shift. The y-axes represent the ^{13}C chemical shift in the triple resonance spectra, while ^{15}N chemical shifts in the 2D-[^{1}H/^{15}N] HSQC spectrum. The 10 peaks with the same ^{1}HN and ^{15}N chemical shift are combined into peak cluster.

In this way, totally 84 peaks in the 2D-[^{1}H/^{15}N] HSQC spectrum was confirmed as backbone amide resonances with reasonable patterns of peak clusters. Then, sequence-specific assignments could be achieved by the sequential linking of the peak clusters. Finally, all expected backbone [^{1}H/^{15}N] correlations have been assigned, not including two proline residues (P2 and p22) and the N-terminal residues which could not be detected. The additional residues, H10 and hexa-histidine tag at the C-terminal are also unassigned due to the signal overlap and the lack of sequential connectivity, despite clear observation of peak cluster. SF1002 is 75 amino acids except eight non-native residues at the C-terminus (LEHHHHHH). In total, all the available resonances except N-terminal two residues (M1 and T3) and H10 could be completely assigned. The finally assigned backbone amide $^{1}H^{N}$, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts are summarized in table 1.

Table 1. Chemical shifts (ppm) of $^{1}H^{N}$, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO of SF1002 at 298 K and pH 6.5. All chemical shifts were referenced indirectly using TALOS server (NA, not available; ND, not detected).

a.a.	$^{1}H^{N}$	^{15}N	^{13}Cα	^{13}Cβ	^{13}CO
M1	NA	NA	NA	NA	NA
P2	NA	NA	ND	ND	ND
T3	ND	ND	59.195	67.551	171.944
Q4	9.154	121.08	56.133	26	175.645
E5	8.798	121.136	57.268	25.938	175.317
A6	7.968	123.956	53.469	15.312	178.723
K7	7.76	118.816	57.133	29.69	177.489
A8	7.912	120.729	52.721	15.222	176.991
H9	7.452	110.586	54.196	25.032	169.947
H10	ND	ND	54.852	22.552	172.733
V11	7.698	116.848	66.6	28.013	174.608
G12	8.478	103.727	46.84	NA	175.883
E13	7.836	126.943	59.694	26.506	176.501
W14	7.974	122.284	62.919	27.678	178.384
A15	9.178	114.032	56.273	14.036	178.931
S16	7.562	113.118	62.086	59.024	181.919
L17	7.138	120.641	56.598	39.427	175.539
R18	5.158	109.39	51.234	25.338	174.485
N19	7.72	124.952	53.742	35.037	168.477
T20	7.628	107.44	58.518	75.378	167.457
S21	10.018	132.23	57.966	59.866	179.201

P22	NA	NA	65.93	28.726	174.277
E23	11.886	114.126	63.534	25.514	180.413
I24	4.894	108.206	62.4	31.848	174.054
A25	7.044	115.13	56.749	18.93	174.773
E26	9.132	113.196	61.166	27.658	176.22
A27	5.912	116.222	55.325	18.72	177.065
I28	8.842	116.53	69	33.792	175.126
F29	7.942	111.388	65.528	32.352	175.855
E30	7.208	114.655	60.546	26.17	181.854
V31	8.188	122.918	67.114	28.099	176.001
A32	8.242	112.519	48.775	16.116	174.43
G33	7.09	113.706	47.044	NA	173.236
Y34	9.232	98.139	61.634	28.1	165.96
D35	5.194	114.472	49.812	41.254	173.379
E36	9.552	133.175	61.014	26.396	177.351
K37	7.906	118.822	59.527	27.636	180.661
M38	7.854	121.961	55.738	27.585	176.331
A39	8.502	119.661	55.337	12.552	177.079
E40	6.66	112.504	59.422	25.292	176.251
K41	6.6	119.182	60.135	29.766	178.093
I42	8.726	116.939	66.04	34.716	177.918
W43	9.396	114.678	66.119	25.32	177.264
E44	7.922	108.834	58.731	28.826	176.895
E45	8.228	112.612	57.346	30.198	177.07
G46	8.804	104.938	39.622	NA	166.955
S47	6.586	108.281	55.776	64.06	172.238
D48	9.574	139.134	58.626	37.606	176.553
E49	8.026	120.711	59.29	26.822	178.565
V50	7.088	114.166	67.528	29.21	173.835
L51	6.194	108.996	59.078	39.693	176.537
V52	7.378	114.692	68.36	29.189	179.395
K53	7.124	112.551	59.737	31.524	178.641
A54	9.25	118.067	55.603	17.394	179.349
F55	7.79	106.254	64.768	33.75	175.147
A56	6.418	115.447	51.827	15.208	176.721
K57	6.37	112.162	53.873	33.102	170.841
T58	6.266	108.829	53.414	66.904	165.198
D59	8.248	124.482	50.126	37.832	175.249
K60	8.49	123.955	55.865	33.076	175.268
D61	8.912	113.015	53.852	40.648	172.269
S62	6.722	109.661	53.118	65.036	166.491
L63	9.388	119.921	48.852	50.603	168.363
F64	9.192	112.246	51.338	41.172	171.555
W65	10.17	130.687	49.885	28.88	172.711
G66	8.154	124.296	48.166	NA	170.803
E67	9.564	126.745	55.306	27.05	172.725
Q68	7.302	116.66	52.317	30.708	172.065
T69	9.284	128.256	61.182	67.512	169.871
I70	10.408	133.408	56.292	39.3	170.492
E70	9.086	127.3	50.984	30.282	173.863
R72	8.634	129.908	59.914	29.354	176.539
K73	8.622	108.654	57.419	28.124	174.243
N74	6.484	112.775	51.94	36.633	173.907
V75	6.496	119.286	64.214	29.485	174.167
L76	7.626	118.061	52.86	50.686	177.075
E77	9.224	112.428	53.372	27.556	177.449

Secondary structural information was obtained on the basis of chemical shifts. The chemical shift is a good means of a sensitive measure of molecular conformation, backbone dihedral angle, hydrogen bond interactions, backbone dynamics and ring-flip rates. We employed CSI and TALOS+ programs to predict the secondary structure of SF1002. CSI has become a standard method for predicting secondary structure of protein by characterizing deviation of chemical shifts of certain nuclei in amino acid relative to their random coil values.[9] TALOS+ program was also used to get backbone dihedral angles (φ, ϕ) using chemical shift values.[8] Delta values ($\delta C\alpha$-$\delta C\beta$) of backbone carbon to random coil chemical shift were also used, which are consistent with the CSI/TALOS+ prediction. As shown in Figure 4, examination of ($\delta C\alpha$-$\delta C\beta$) plot and the results of CSI/TALOS+ indicate that SF1002 consists of five α-helixes and two β-strands

(α1–α2–α3–α4–α5–β1–β2).

In further study, NMR experiment and assignment of side chain will be performed in order to determine the three-dimensional structure of SF1002 which can be used to find its molecular function.

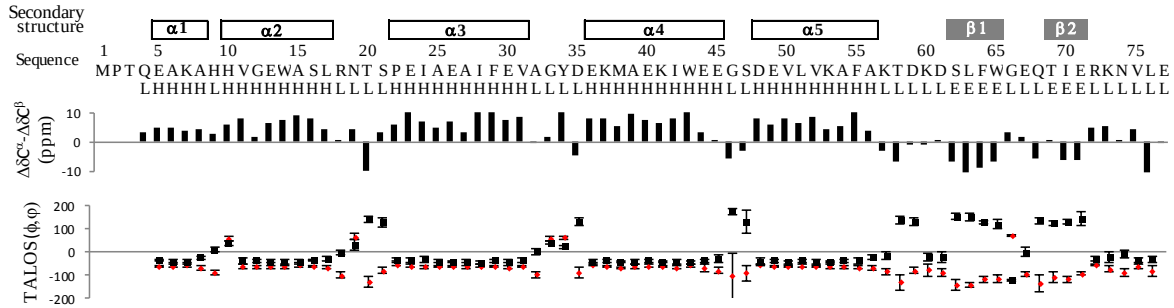

Figure 4. Determination of secondary structure of SF1002. Determined α-helix and β-strand regions are represented by empty and filled rectangle, respectively. Delta values (δCα-δCβ) of backbone carbon to random coil chemical shift were plotted. Backbone dihedral angles were calculated using TALOS+ server.[7]

Acknowledgements

This study made use of the NMR machine at the College of pharmacy, Gachon university. This work is supported by grants provided by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2014R1A1A2054691 and NRF–2012R1A1A1039738).

References

1. J. Wei, M. B. Goldberg, V. Burland, M. M. Venkatesan, W. Deng, G. Fournier, G. F. Mayhew, G. Plunkett III, D. J. Rose, A. Darling, B. Mau, N. T. perna, S. M. Payne, L. J. Runyen-Janecky, S. Zhou, D. C. Schwartz, F. R. Blattner, *Infect. Immun*. **71**, 2275 (2003)
2. Q. Jin *et al*., *Nucleic Acids Res*. **30**, 4432 (2002)
3. M. L. Bennish, B. J. Wojtyniak, *Rev. Infect. Dis*. **13**, S245 (1991)
4. S. Ashkenaze, I. Levy, V. Kazaronovski, Z. Samra, *J. Antimicrb. Chemother*. **51**, 427 (2003)
5. P. J. Sansonetti, *PLos Med*. **3**, 1465 (2006)
6. Y.-S. Lee, K.-S. Ryu, Y. Lee, S. Kim, K. W. Lee, H.-S. Won, *j. Kor. Magn. Reson. Soc*. **15**, 137 (2011)
7. Y.-S. Lee, H.-S. Ko, K.-S. Ryu, Y.-H. Jeon, H.-S. Won, *j. Kor. Magn. Reson. Soc*. **14**, 117 (2010)
8. N. T. Onodera, J. Ryu, T. Durbic, C. Nislow, J. M. Archibald, R. Rohde, *J. Bacteriol*. **194**, 3022 (2012)
9. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **44**, 213 (2009)
10. D. S. Wishart, B. D. Sykes, *J. Biomol. NMR* **4**, 171 (1994)