Journal of the Korean Magnetic Resonance Society **2007**, *11*, 73 - 84

Purification and Backbone Assignment of the Hypothetical Protein MTH1821 from *Methanobacterium Thermoautotrophicum H*

Sooyoung Kwak, Woonghee Lee, Joon Shin, Sunggeon Ko and Weontae Lee*

Department of Biochemistry, HTSD-NMR National Research Laboratory,
College of Science, Yonsei University, Seoul 120–749, Korea
Received April 5, 2007

Abstract : MTH1821 (UniProtKB/TrEMBL ID O27849) is a 96-residue hypothetical protein from the open reading frame of *Methanobacterium thermoautotrophicum H* one of the target organisms of structural genomics pilot project. Proteins which contain conserved sequence compared with MTH1821 have not been discovered yet and the functional and structural information for MTH1821 is not available. Here, we present the sequence-specific backbone resonance using multidimensional heteronuclear NMR spectroscopy and propose the secondary structure using GetSBY software. The backbone resonances of N, HN, C_α, C_β, CO and H_α which are necessary for a prediction of secondary structure by GetSBY were assigned about 98% (557/568). The secondary structure of MTH1821 confirmed that it is comprised of four strand regions and two helical regions. This report will provide a valuable resource for the calculation solution structure of MTH1821 and for the other hypothetical protein that is targeted for structural-based functional discovery.
Keywords : MTH1821, NMR, Backbone resonance assignment, *Methanobacterium thermoautotrophicum*

INTRODUCTION

Structural proteomics approach has been considered as one of the major initiative project in the post-genomics era. The large number of protein structures determined by NMR spectroscopy and X-ray crystallography has provided valuable inceptions for protein folding as well as their biological functions. We initiated structural proteomics

*To whom correspondence should be addressed. E-mail : wlee@spin.yonsei.ac.kr

pilot project for model organism, *Methanobacterium thermoautotrophicum H* which is a thermophilic archaebacterium with an optimal growth temperature of 65℃ and a generation time of about 5 hours. It has been determined by a whole-genome shotgun sequencing approach in 1997, and its genome includes 1,885 open reading frames (ORFs) which have been confirmed. Therefore several hypothetical proteins from *Methanobacterium thermoautotrophicum H* have been selected as structural genomics targets. MTH1821 is a 96-residue hypothetical protein from the open reading frame of *Methanobacterium thermoautotrophicum H* with a molecular weight of 11.5 kDa and a calculated pI of 4.27. BLAST searches found some proteins that have sequence identities with MHT1821. TrEMBL A5UKH2 (*Methanobrevibacter smithii*), Q28UK2 (*Jannaschia sp.*), A5ALD2 (*Vitis vinifera*), Q1KUQ1 (*Cleome spinosa*), A0HPA5 (*Leifsonia xyli subsp. xyli*), and Q54NX7 (*Dictyostelium discoideum AX4*) are putative uncharacterized proteins like MTH1821. A2Q2P2 (*Medicago truncatula*) possesses a zinc finger and Q6AEC6 (*Leifsonia xyli subsp. xyli*) is a GTP-binding protein era homolog. However, proteins which contain conserved sequence compared with MTH1821 have not been discovered yet and the functional and structural information for MTH1821 is not available. Here to annotate the biochemical and biological functions of MTH1821, biophysical techniques and database analysis based on structural information have also been used. We assigned the sequence-specific backbone resonance using multidimensional heteronuclear NMR spectroscopy and predicted secondary structure using GetSBY software we developed. NMR spectroscopy is unique in being able to clarify the atomic structure of macromolecules in solution, provided that highly concentrated solutions (about 1m*M*, or 15 mg/ml for a 15 kDa protein) can be obtained. GetSBY is a protein secondary structure prediction software and use chemical shifts of N, HN, C_α, C_β, CO and H_α in backbone of protein.

Protein can fold into regular structures such as the α-helix, the β-sheet, and turns and loops. These secondary structures are well defined and contribute to form the final protein structure. And protein structure dictates function, given that the specificity of active sites and binding sites depends on the precise three-dimensional conformation. Therefore this study is a foundation for the calculation solution tertiary structure of MTH1821.

EXPERIMENTAL

Sample preparations

MTH1821 gene was acquired from *Methanobacterium thermoautotrophicum H* genomic DNA by PCR amplification and subcloned into the pET-13b expression vector (Novagen Inc.) with a thrombin cleavage site, and the vector contains additional hexa-histidine affinity tag (His-tag) at the N-terminus. MTH1821 protein was over-expressed in *E.coli* BL21 (DE3) pLysS strain (Novagen Inc.). In order to obtain a $^{15}N/^{13}C$-labeled protein, cells were grown on M9 minimal media with $^{15}NH_4Cl$ and ^{13}C-gluscose at 37℃ until the optical density reached 0.6 at 600 nm. Thiamine 0.1% and ampicillin 5% were added in the media to support optimal growth. Then protein expression was induced with 1 m*M* isopropyl β-D-thiogalactoside (IPTG) for 12hours at 25℃. Cells were harvested by centrifugation at 6000 x *g* for 30 minutes, resuspended in 20 ml binding buffer (100 m*M* sodium chloride, 20 m*M* HEPES, in pH 7.6) and lysed by the sonicator. The lysate was centrifuged at 14000 x *g* for 30 minutes to gain the soluble supernatant, and a Ni^{2+} affinity chromatography (Amersham Pharmacia) was accomplished. The proteins bound to the column were washed with 40 m*M* imidazole washing buffer and eluted with 300 m*M* imidazole elution buffer. The eluted protein was gone through gel filtration chromatography (Superdex 75). The final NMR sample was approximately 1.0-1.5 m*M* concentration with 100 m*M* sodium chloride and 20 m*M* HEPES buffer containing 0.002% NaN_3 at pH 7.6. And for NMR signal detection, 10% D_2O was added.

NMR spectroscopy

All NMR experiments data for assignments was acquired on Bruker DRX 500 MHz spectrometer at 298 K. Assignments of backbone resonances were derived from spectra of 2D-[^1H-^{15}N] HSQC (Kay et al., 1992), 3D-CBCA(CO)NH (Muhandiram and Kay, 1994), HNCACB (Wittekind et al., 1993), HNCA (Ikura et al., 199Oa), HNCO, and HNHA (Kuboniwa et al., 1994). 3D- ^{15}N-edited NOESY (τ_m = 150ms) was also used for tentative assignment. These achieved FIDs were processed by using NMRPipe/NMRDraw

(Delaglio et al., 1995) and the spectra were analyzed by using Sparky program (Goddard and Kneller, 2003).

Prediction of secondary structure

On the basis of the assigned chemical shifts about backbone atoms, secondary structure of MTH1821 was detected by GetSBY (Get Secondary structure By PACSY) software. Chemical shifts of N, HN, C_α, C_β, CO and H_α were required to execute a GetSBY.

RESULTS AND DISCUSSION

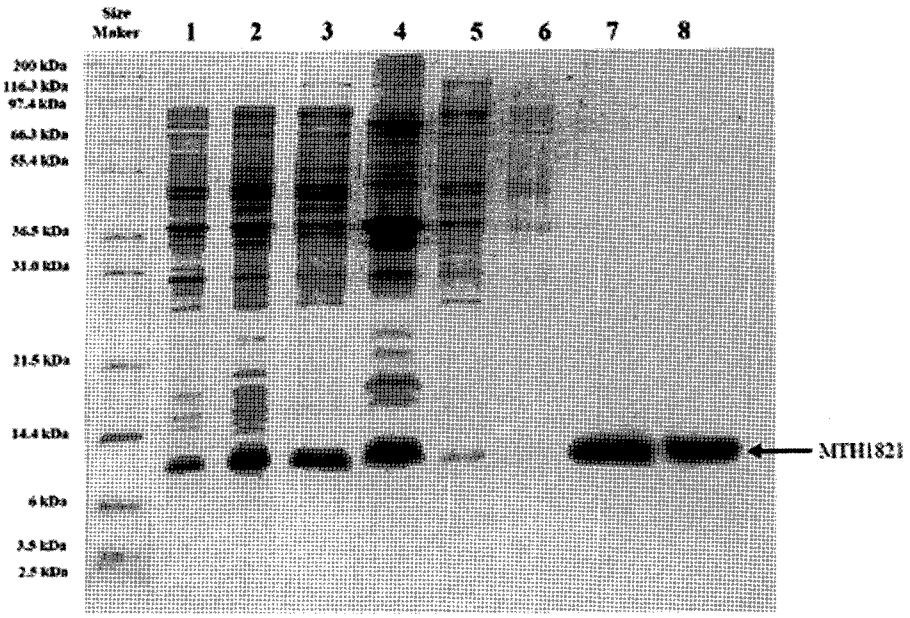

Fig. 1. (A) SDS-PAGE analysis at each step of purification.

(1), Before induction; (2), after induction by IPTG; (3), supernatant after centrifugation; (4), precipitant after centrifugation; (5), flow through in Ni^{2+} affinity chromatography step; (6), after washing; (7), elution sample; (8), after gel filtration.

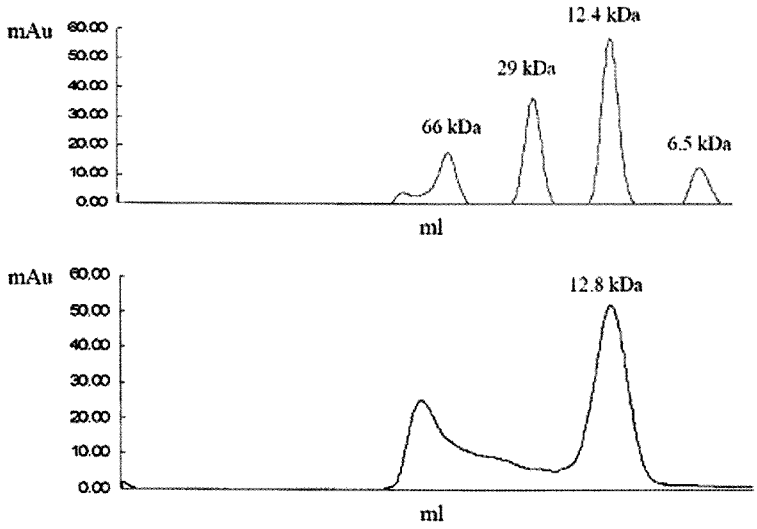

Fig. 1. (B) Final step of purification is gel filtration on Superdex 75. A calibration curve (top) was used to estimate the molecular weight of an MTH1821, the peak of MTH1821 appears as 12.4 kDa molecular weight (bottom).

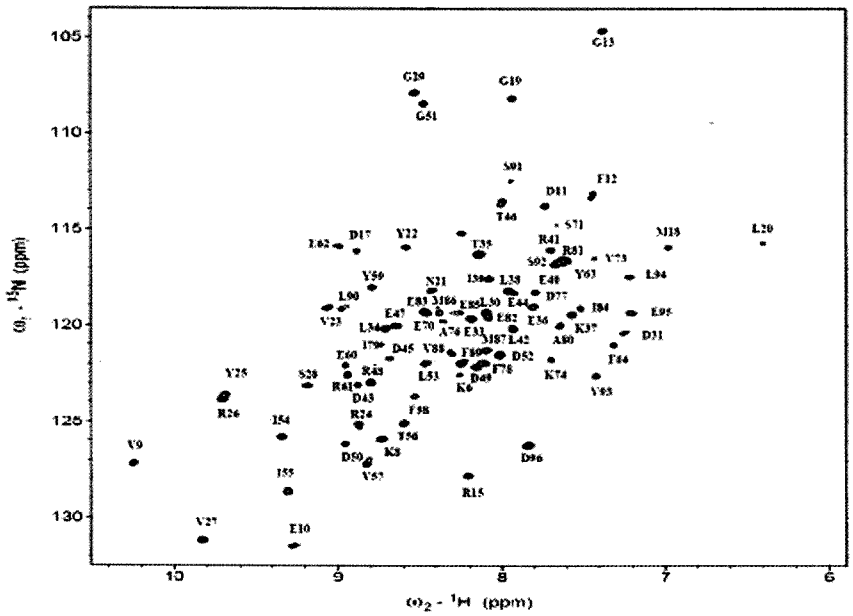

Fig. 2. 2D-[^{1}H-^{15}N] HSQC spectrum of MTH1821 recorded on a Bruker DRX 500 MHz spectrometer at 25℃. Each number represents the amino acid residue number in the amino acid sequence of MTH1821. Some residues were not appeared in [^{1}H-^{15}N] HSQC

In SDS-PAGE and gel filtration results, purified MTH1821 is not ascertained as 11.5 kDa but about 12.8 kDa, because the recombinant MTH1821 protein didn't treated with thrombin (Fig. 1). As shown in first peak of gel filtration, purified MTH1821 seems to form soluble aggregates, but SDS-PAGE band of the protein is appeared in native size because of the denaturation by SDS which is an ionic detergent. To estimate the molecular weight of an MTH1821 in gel filtration, a calibration curve relating elution volume and molecular weight was constructed by determining the elution volume of proteins with known molecular weights.

A 2D-[^{1}H-^{15}N] HSQC spectrum of MTH1821 at 25℃ is shown in Fig. 2. The peaks of six histidines and some amino acid residues which are recognized by thrombin are also displayed in [^{1}H-^{15}N] HSQC. Sequential assignment was achieved by HNCACB, CBCACONH and HNCA. The sequential assignment from N21 to G29 is displayed in HNCACB strip plots (Fig. 3. A). Strip plot is relevant to the ^{15}N plane of the indicated

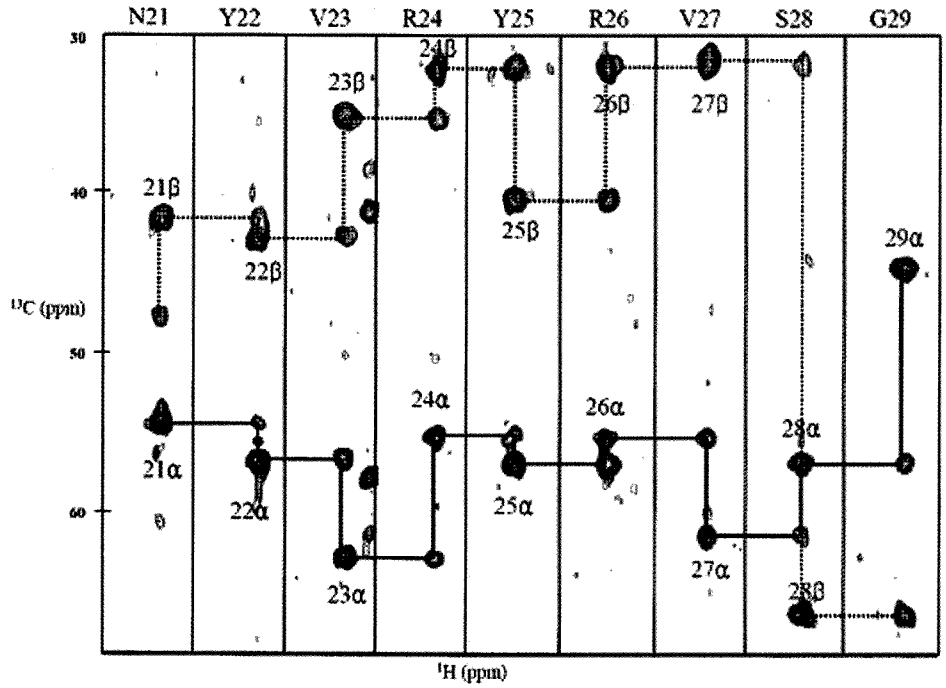

Fig. 3. (A) Strip plots from ^{15}N plane of the HNCACB spectrum shows C_α and C_β sequential connectivity.

amino acid residues and exhibits C_α and C_β sequential connectivity. The backbone resonances of N, HN, C_α, C_β, CO and H_α which are necessary for a prediction of secondary structure by GetSBY were assigned about 98% (557/568) (Table 1). GetSBY compared the assigned chemical shifts of six backbone atoms with the averaged chemical shift data as reference database derived from PACSY (Protein structure and chemical shift in Yonsei: PACSY is a relative database management system based on the PDB files of Protein Database Bank (PDB) which has information about the protein coordinates and on the chemical shift data of Bio Magnetic Resonance Bank (BMRB) which has information about the chemical shift of atoms in protein by NMR experiment) and then selected structure with the smallest difference of these values. The missing of chemical shift and the irrational structure were transacted by pass through a chain of smoothing process in GetSBY program. The secondary structure of MTH1821 confirmed that it is comprised of four strand regions and two helical regions (Fig. 3 B). Residue L5~F12, Y22~S28, D45~D50 and D52~Y59 constitute strands, and E36~R41 and K74~L94 constitute helices.

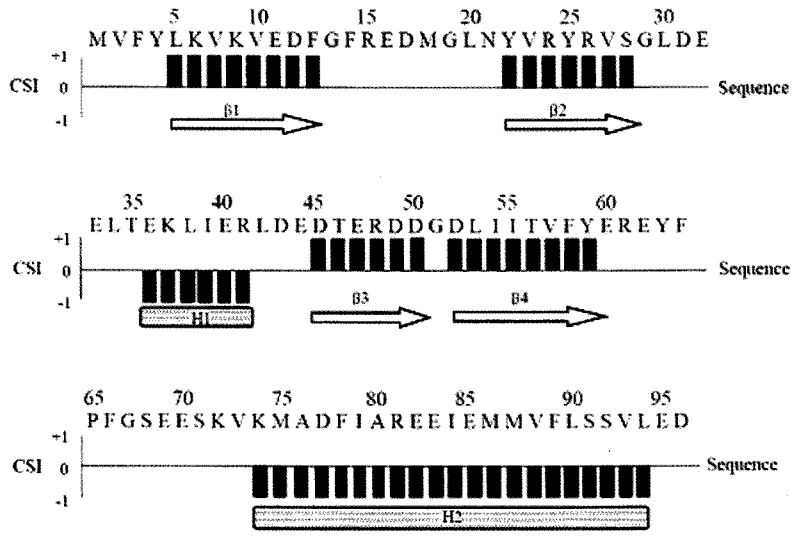

Fig. 3. (B) Protein secondary structure of MTH1821 predicted from GetSBY with backbone chemical shifts. X axis represents amino acid sequence and Y axis represents CSI (chemical shift index). '0' means the chemical shifts of the random coil structure. '+1' and '-1' mean the positive and negative deviation of chemical shifts from random coil values and refer to strand and helix structure separately.

According to BLAST search, amino acid sequence from V23 to E47 in MTH1821 displays sequence identity with A2Q2P2 (*Medicago truncatula*) possesses a zinc finger. GetSBY identified this region has secondary structures as helix and strand. There are 27% acidic amino acids such as aspartate and glutamate in full sequence of MTH1821 and the region from V23 to E47 contains 23% of whole acidic amino acids. The negative charges in this region may participate in binding to positively charged region of other molecules by formation of cation binding sites.

Protein can fold into regular structures such as the α-helix, the β-sheet, and turns and loops. These secondary structures are well defined and contribute to form the final protein structure. In other words, determination of the protein tertiary structure is able to enhance by spatial alignment of determined secondary structure accurately. In addition, the protein tertiary structure detected exactly can provide an essential clue to identify its function, because protein structure dictates function, given that the specificity of active sites and binding sites depends on the precise three-dimensional conformation. This report will provide a valuable resource for the calculation solution structure of MTH1821 protein and for the other hypothetical protein that is targeted for structural-based functional discovery.

Acknowledgement

This study was supported by the Ministry of Science and Technology of Korea/the Korea Science and Engineering Foundation through the NRL program of MOST NRDP (M1-0203-00-0020); the Ministry of Education and Human Resource through the BK21 project (W.L.).

Table 1. Chemical shifts of N, HN, C_α, C_β, CO and H_α of MTH1821. 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as internal reference for proton chemical shifts. (unit: ppm)

	Residue	HN	N	CO	CA	CB	HA	
1	MET			175.5	55.49	32.75	4.283	
2	VAL	7.906	120.7	175.1	62.32	20.88	3.841	
3	PHE	7.761	121.6	174.5	56.98	40.25	4.504	
4	TYR	8.346	120.9	174.1	57.23	38.50	4.467	
5	LEU	7.439	122.0	176.0	54.70	43.61	4.495	
6	LYS	8.267	122.6	174.5	55.20	34.65	4.532	
7	VAL	8.223	123.4	173.9	60.88	34.32	4.830	
8	LYS	8.728	125.9	176.8	52.96	36.63	5.005	
9	VAL	10.250	127.2	175.1	64.51	31.38	3.827	
10	GLU	9.282	131.5	175.5	57.25	31.50	4.384	
11	ASP	7.738	113.8	172.7	53.22	43.04	4.409	
12	PHE	7.440	113.1	173.5	55.66	41.55	4.568	
13	GLY	7.383	104.7	171.1	46.24		3.991	
14	PHE	9.062	125.1	174.2	57.34	42.37	4.941	
15	ARG	8.213	127.8	ND	53.14	30.03	4.538	
16	GLU	ND	ND	177.4	59.18	29.39	3.705	
17	ASP	8.882	116.1	176.8	55.57	38.43	4.165	
18	MET	6.981	116.0	176.3	54.96	35.07	4.305	
19	GLY	7.936	108.2	173.6	46.47		3.954	3.665
20	LEU	6.404	115.7	176.0	53.19	47.63	4.800	
21	ASN	8.437	118.2	174.0	54.00	41.87	5.3800	
22	TYR	8.590	115.9	173.3	56.34	43.07	6.011	
23	VAL	9.072	119.1	173.9	62.32	35.80	4.316	
24	ARG	8.875	125.1	174.9	54.87	33.02	5.438	
25	TYR	9.688	123.6	174.1	56.86	41.09	5.038	
26	ARG	9.709	123.9	175.3	55.52	33.12	4.883	
27	VAL	9.827	131.2	175.4	61.31	32.75	4.942	
28	SER	9.196	123.1	173.6	56.93	66.25	5.162	
29	GLY	8.522	107.8	175.0	45.12		4.509	3.611
30	LEU	8.087	119.3	176.9	54.30	43.12	4.316	
31	ASP	7.251	120.4	177.5	52.55	41.52	4.654	
32	GLU	8.794	121.7	178.2	60.57	29.30	3.907	
33	GLU	8.186	119.7	179.9	59.69	29.37	3.960	
34	LEU	8.711	120.2	178.8	57.52	42.34	3.957	
35	THR	8.142	116.4	175.6	68.19	ND	3.551	
36	GLU	7.805	119.0	178.9	59.37	29.01	3.843	
37	LYS	7.568	119.5	178.1	59.48	32.89	3.972	
38	LEU	7.952	118.2	178.3	58.12	41.58	3.767	
39	ILE	8.082	117.6	178.4	65.85	38.07	3.496	
40	GLU	7.672	116.8	178.4	58.70	29.87	4.006	
41	ARG	7.704	116.1	177.1	57.36	32.09	4.170	
42	LEU	7.933	120.2	176.7	54.88	42.99	4.444	
43	ASP	8.876	123.1	176.5	54.40	40.93	4.689	
44	GLU	7.920	118.4	175.3	56.14	31.52	4.431	
45	ASP	8.680	121.7	175.6	55.63	41.45	4.498	
46	THR	7.995	113.5	174.1	59.62	72.77	5.300	
47	GLU	8.639	120.0	174.6	54.90	33.79	4.553	

48	ARG	8.799	123.0	176.3	55.70	31.21	4.694	
49	ASP	8.157	122.2	174.8	54.04	43.33	4.685	
50	ASP	8.957	126.2	175.6	55.76	39.74	4.114	
51	GLY	8.477	108.4	174.2	45.63		4.056	3.806
52	ASP	8.014	121.6	175.2	53.23	42.22	5.040	
53	LEU	8.473	122.0	174.4	53.36	46.03	4.933	
54	ILE	9.344	125.8	176.7	58.45	37.62	5.286	
55	ILE	9.309	128.7	175.7	59.80	41.87	4.885	
56	THR	8.593	125.0	172.3	62.83	69.25	4.931	
57	VAL	8.825	127.3	174.1	61.05	34.26	4.027	
58	PHE	8.529	123.7	175.6	57.25	41.43	5.409	
59	TYR	8.793	118.0	175.8	56.98	43.67	4.821	
60	GLU	8.953	122.1	178.4	56.92	30.21	4.468	
61	ARG	8.940	122.5	178.7	59.71	30.25	4.124	
62	GLU	8.985	115.9	176.3	58.61	28.79	3.919	
63	TYR	7.600	116.7	174.9	57.30	39.80	4.500	
64	PHE	7.316	121.0	ND	53.42	39.56	4.150	
65	PRO			175.9	62.44	32.69	4.372	
66	PHE	8.203	125.6	175.1	55.01	38.01	4.426	
67	GLY	7.350	107.0	174.2	45.50		3.944	
68	SER	7.760	115.2	ND	57.56	64.62	*4.714*	
69	GLU	ND	ND	178.3	58.59	29.15	3.974	
70	GLU	8.456	119.3	177.4	58.36	29.55	4.005	
71	SER	7.671	114.8	ND	59.64	63.37	4.706	
72	LYS	ND	ND	178.7	58.64	32.38	3.988	
73	VAL	7.436	116.5	177.1	63.86	32.20	3.968	
74	LYS	7.705	121.8	177.3	55.74	32.15	4.468	
75	MET	7.619	121.0	177.7	58.48	31.90	4.205	
76	ALA	8.348	119.8	180.7	55.65	17.86	3.976	
77	ASP	7.796	118.3	178.6	56.90	40.52	4.370	
78	PHE	8.096	122.0	178.1	61.24	38.72	4.069	
79	ILE	8.736	121.0	177.0	65.62	37.56	3.474	
80	ALA	7.642	120.0	180.6	54.91	18.02	4.212	
81	ARG	7.619	116.7	178.8	59.13	29.88	4.009	
82	GLU	8.085	119.6	179.7	58.71	28.85	3.98	
83	GLU	8.476	119.3	178.9	60.52	29.32	3.695	
84	ILE	7.514	119.2	178.2	64.65	37.73	3.786	
85	GLU	8.240	119.4	180.3	59.39	28.87	3.947	
86	MET	8.388	119.4	177.7	56.61	30.58	4.493	
87	MET	8.090	121.3	178.7	60.36	31.97	3.762	
88	VAL	8.297	121.6	179.0	66.82	31.57	3.536	
89	PHE	8.247	122.0	178.3	60.89	38.92	4.318	
90	LEU	8.984	119.2	178.7	57.53	41.51	3.467	
91	SER	7.955	112.5	176.0	62.72	62.89	3.879	
92	SER	7.633	116.7	176.5	61.24	62.73	4.201	
93	VAL	7.423	122.7	177.4	64.98	31.41	3.559	
94	LEU	7.220	117.5	176.6	55.76	42.85	3.795	
95	GLU	7.202	119.4	175.6	57.39	30.53	4.097	
96	ASP	7.832	126.3		56.21	42.31	4.351	

*ND: not detected

Italic letters: tentative assignment

REFERENCES

1. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, *J Mol Biol* **215**, 403-10 (1990)
2. S. Grzesiek, H. Dobeli, R. Gentz, G. Garotta, A. M. Labhardt, and A. Bax, *Biochemistry* **31**, 8180-90 (1992)
3. K. L. Constantine, V. Goldfarb, M. Wittekind, M. S. Friedrichs, J. Anthony, S. C. Ng, and L. Mueller, *J Biomol NMR* **3**, 41-54 (1993)
4. D. S. Wishart, and B. D. Sykes, *J Biomol NMR* **4**, 171-80 (1994)
5. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, *J Biomol NMR* **6**, 277-93 (1995)
6. D. R. Smith, L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, J. N. Reeve, and et al., *J Bacteriol* **179**, 7135-55 (1997)
7. A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren, *Acta Crystallogr D Biol Crystallogr* **54,** 905-21 (1998)
8. J. L. Pennings, J. T. Keltjens, and G. D. Vogels, *J Bacteriol* **180**, (10), 2676-81 (1998)
9. H. Qian, M. S. Rogers, J. Schleucher, U. Edlund, E. E. Strehler, and I. Sethson, *Protein Sci* **7**, 2421-30 (1998)
10. D. Christendat, A. Yee, A. Dharamsi, Y. Kluger, A. Savchenko, J. R. Cort, V. Booth, C. D. Mackereth, V. Saridakis, I. Ekiel, G. Kozlov, K. L. Maxwell, N. Wu, L. P. McIntosh, K. Gehring, M. A. Kennedy, A. R. Davidson, E. F. Pai, M. Gerstein, A. M. Edwards, and C. H. Arrowsmith, *Nat Struct Biol* **7**, 903-9 (2000)
11. Z. Kelman, and J. Hurwitz, *J Biol Chem* **275**, 7327-36 (2000)
12. D. Christendat, V. Saridakis, Y. Kim, P. A. Kumar, X. Xu, A. Semesi, A. Joachimiak, C. H. Arrowsmith, and A. M. Edwards, *Protein Sci* **11**, 1409-14 (2002)
13. R. Vadrevu, C. J. Falzone, and C. R. Matthews, *Protein Sci* **12**, 185-91 (2003).
14. C. Zhang, and S. H. Kim, *Curr Opin Chem Biol* **7**, 28-32 (2003)

15. F. J. Blanco, A. Yee, R. Campos-Olivas, A. R. Ortiz, D. Devos, A. Valencia, C. H. Arrowsmith, and M. Rico, *Protein Sci* **13**, 1458-65 (2004)
16. C. H. Lee, J. W. Jung, A. Yee, C. H. Arrowsmith, and W. Lee, *Protein Sci* **13**, 1148-54 (2004)
17. O. Julien, I. Gignac, A. Hutton, A. Yee, C. H. Arrowsmith, and S. M. Gagne, *J Biomol NMR* **35**, 149-54 (2006)
18. T. D Goddard, D. G. Kneller, SPARKY 3, 3.110 ed., San Francisco: University of California 2004