Journal of the Korean Magnetic Resonance Society **2010**, *14*, 45-54

Biochemical and Structural Characterization of HP1423 (Y1423_HELPY) from *Helicobacter pylori*

Ji-Hun Kim[1], Ki-Young Lee[1], Sung-Jean Park[2], and Bong-Jin Lee[1*]

[1]*Research Inst. of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul, 151-742, Korea*
[2]*Gachon University of Medicine and Science, Incheon, Korea*
Received April 21, 2010

Abstract : HP1423 (Y1423_HELPY) is a conserved hypothetical protein from *H. pylori* strain 26695. However, Sequence Blast result indicates that HP1423 belongs to S4 (PF01479) superfamily. According to Pfam database, the S4 domain is a small domain consisting of 60-65 amino acid residues, that probably mediates binding to RNA. In this study, we report the sequence-specific backbone resonance assignment of HP1423, which has 84 amino acid residues. We could assign unambiguously about 88% of all $^{1}H_{N}$, ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and $^{13}C{=}O$ resonances. We could not detect the resonances from residues 15-20, and disappearance of these peaks seems to be related with the intermediate-conformational exchange. These assigned NMR peaks of HP1423 can be used for studying the role of protein dynamics in millisecond timescale, and Protein-RNA binding.

Keywords : *Helicobacter pylori*, NMR, HP1423, Backbone assignment, Secondary structure

INTRODUCTION

Helicobacter pylori is the most famous gram-negative, flagellated bacterium that infects half of world's population. Helicobacter pylori can inhabit various regions of the stomach and duodenum with or without the clinical symptoms for a life time. It is an important human pathogenic bacterium related to the development of diverse gastric disease such as chronic gastritis, peptic ulcer and even stomach cancer.[1,2,3] Until now, the genomes

* To whom correspondence should be addressed. E-mail : lbj@nmr.snu.ac.kr.

of three strains 26695, J99 and HPAG1 have been sequenced.

HP1423 is classified as hypothetical protein from Helicobacter pylori strain 26695. According to Pfam database (www.pfam.sanger.ac.uk), HP1423 is in the large S4 Pfam (PF01479); there are thousands of sequences in this Pfam and over 30 structural representative to date. The S4 domain is a small domain consisting of 60-65 amino acid residues that was found in the bacterial ribosomal protein S4,[4] eukaryotic ribosomal S9, two families of pseudouridine synthases,[5] a novel family of predicted RNA methylases, a yeast protein containing a pseudouridine synthetase, a deaminase domain, bacterial tyrosyl-tRNA synthetases,[6] and a number of uncharacterized small proteins[7,8] that may be involved in translation regulation. The S4 domain forms the αL-RNA binding motif which is generally found in several RNA binding protein family. The S4 domain probably mediates the binding of αL-RNA binding motif to RNA.[9]

Here, we report the sequence-specific backbone resonance assignments, biochemical and structural characterization of HP1423.

EXPERIMENTAL

Protein sample preparation

The predicted ORF of HP1423 was cloned into the *NdeI/XhoI* site of pET-21a(+) expression vector(Novagen), yielding an C-terminally 6-histidine-tagged protein ($HP1423H_6$). The recombinant protein was transformed into *Escherichia coli* strain BL21 (DE3) host cell which is suitable for large-scale protein production. The cells were grown at 37℃ in M9 media containing $^{15}NH_4Cl$ and $^{13}C_6$-glucose (Cambridge Isotope Laboratories, Inc.) as nitrogen and carbon sources respectively. When the OD_{600} of cells reached at 0.5, IPTG was added to adjust final concentration to 0.5mM. After further cell growth for 4hrs at 37℃, the cells were harvested, and disrupted by sonication. HP1423 protein was purified by applying the supernatant onto a nickel affinity column (Qiagen; 3ml of resin per liter of cell culture) previously equilibrated with the binding buffer. The first purified sample was concentrated, followed by the buffer exchange to final buffer condition (20mM NaH_2PO_4/Na_2HPO_4 pH=5.8 , 150mM NaCl, 15mM Glycine, 1mM DTT, 1mM EDTA) via

superdex-75(Pharmacia) gel column. The NMR sample was ~1mM ^{15}N- and ^{15}N/^{13}C-labled protein prepared by 90% H_2O/10% D_2O buffer solution.

NMR spectroscopy

All NMR spectra were recorded at 298K on Bruker AVANCE 500, 600 (equipped with a cryoprobe) spectrometers. The backbone and side chain assignments were performed by using 3D-HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HBHA(CO)NH, ^{15}N-TOCSY-HSQC, C(CO)NH-TOCSY spectra. Aromatic ring resonances were assigned by using 3D ^{15}N-NOESY-HSQC (mixing time 120ms), ^{13}C-NOESY-HSQC (mixing time 100ms). Chemical shifts were referenced to DSS externally. NMR spectra were processed using the program NMRPipe/nmrDraw[10] and analyzed with the program NMRView[11]. The secondary structure was predicted from the values using Chemical Shift Index (CSI)[12] and the backbone torsion angles were obtained using TALOS program[13] which uses a combination of ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$, $^{1}H_\alpha$ and ^{13}C=O chemical shift of triplet of adjacent residues. The ^{1}H-^{15}N steady-state heteronuclear NOE values were determined from the spectra recorded with (NOE experiment) and without (NONOE) experiment.

RESULTS AND DISCUSSION

The molecular weight of HP1423 is 10.4kDa calculated from the protein sequence. However, in spite of relatively small protein, we couldn't find resonance peaks corresponding to residues from 15 to 20 in its ^{1}H-^{15}N HSQC spectrum. This disappearance of some resonance peaks would be possible through homo or hetero multimerization. Thus, we determined the protein molecular organization using gel-permeation chromatography. As shown in Fig. 1, the molecular weight of HP1423 was determined to be approximately 11kDa, which value corresponds to the monomeric state of the protein. As a result, the disappearance of peaks in HSQC spectrum was not related to multimerization of protein. The protein-metal or protein-nucleotides complex could not be considered as a cause of disappearance in considering existence of EDTA in NMR sample buffer and UV spectrum in the 220 to 330 nm region(data are not shown).

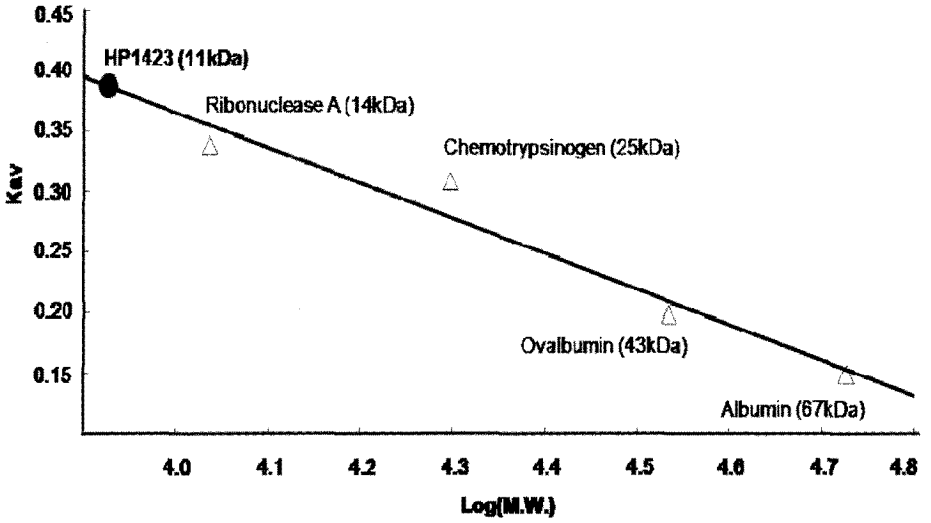

Figure 1. Apparent molecular weight of HP1423, as determined by gel filtration analysis. Open triangle indicates the protein standard included in Low Gel Filtration Calibration Kit (Amersham Biosciences) used. The filled circle means the K_{av} value of HP1423, of which molecular weight was calculated using the standard curve.

The purified HP1423 was very unstable and precipitated in a concentration-dependent manner at high concentration. This aggregation problem could be solved via the addition of the reducing agent (DTT or β ME) in the protein solution. One or both of cysteins in HP1423 containing two cystein amino acids, may form an unfavorable inter or intra disulfide bond which causes precipitation of the protein at high concentration. Actually, one cystein residue is close to the region showing peak disappearance and the other is located at the L-loop region.

^{1}H-^{15}N HSQC spectrum of ^{15}N riched HP1423 showed good dispersion of peaks but some peaks of residues from 15 to 20 were not detected (Fig. 2). To find out the missed peaks, sequence specific labelings about Ala and Val were performed (Fig 3).

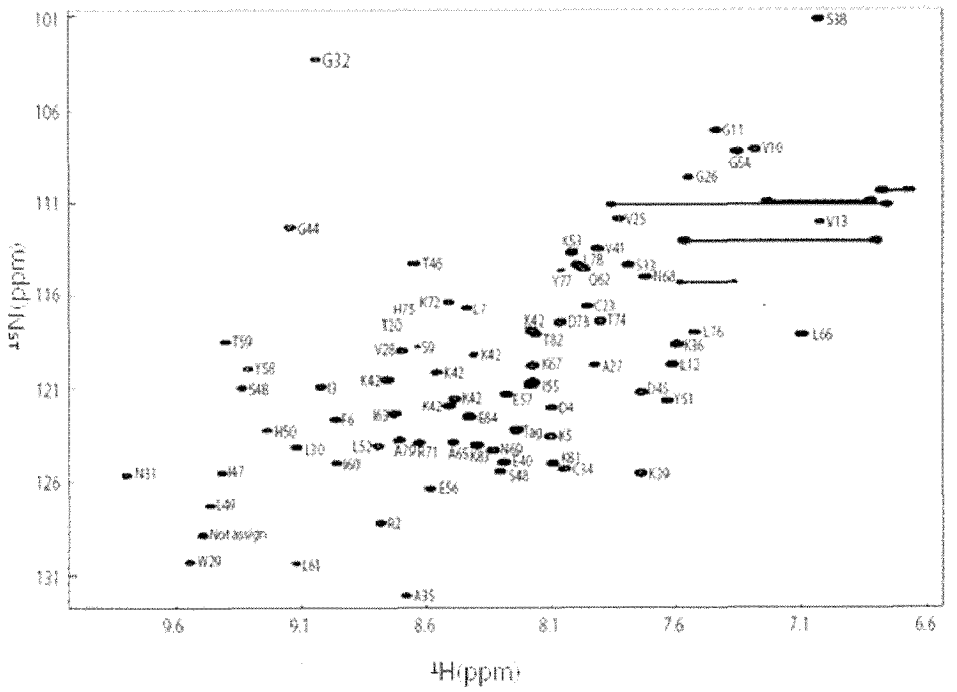

Figure 2. 2D-^{1}H and ^{15}N HSQC spectrum of HP1423. The each resonance in the spectrum is labeled with assigned amino acid residues.

However, the results of these experiments were not helpful for finding missed peaks. Consequently, 88% of all $^{1}H_N$, ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$ and ^{13}C=O resonances were assigned (Table 1). The secondary structure of HP1423 was determined using CSI and TALOS. The secondary structure elements were defined at the region where showed a consensus as same second structure on the results from both of CSI and TALOS. The secondary structure of HP1423 is composed of four β-strands and one α-helix. The β-strands correspond to residues 26-30, 44-50, 55-59, and 78-80. The α-helices correspond to residues 3-9.

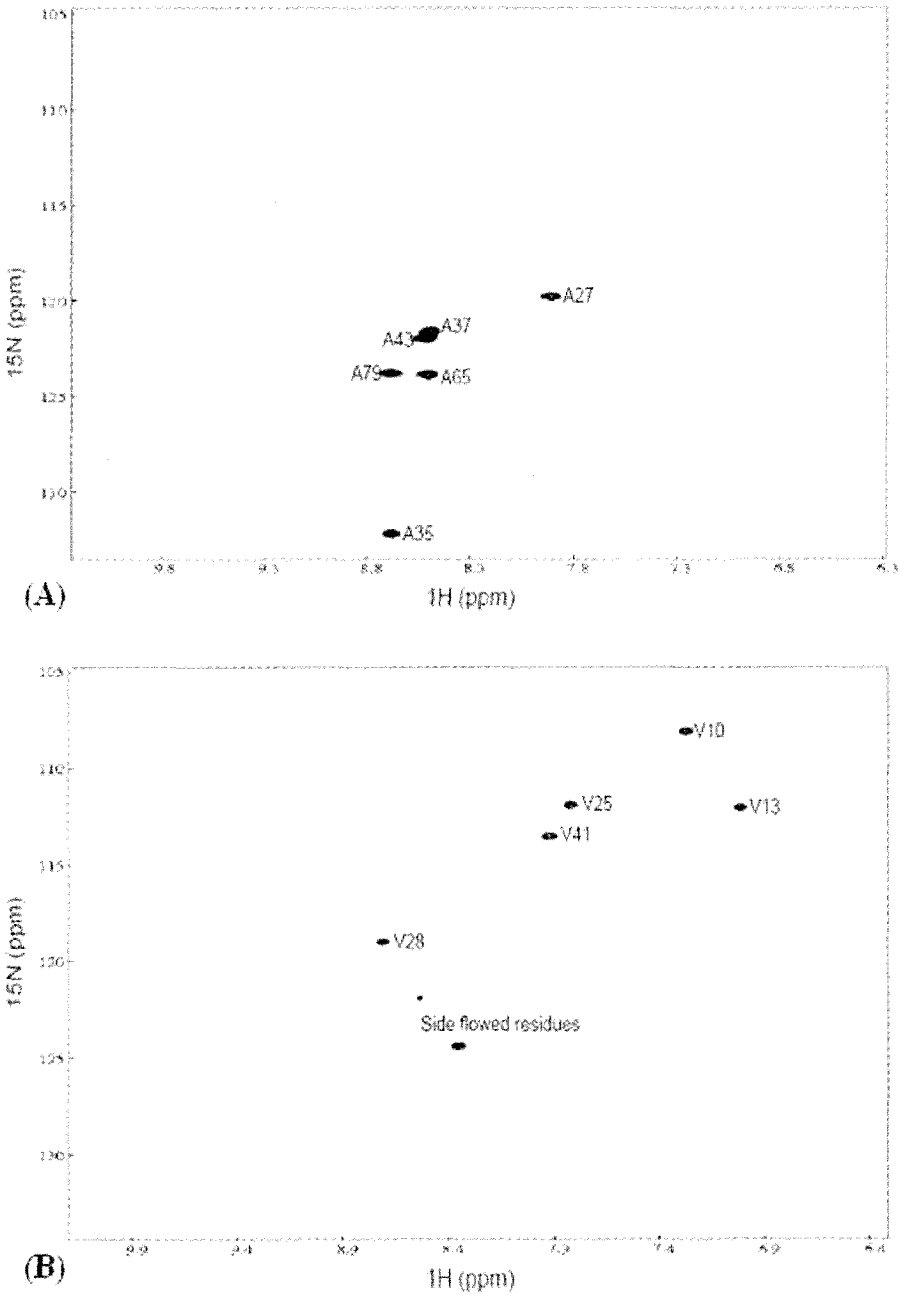

Figure 3. (A) Ala Residues specific ^{15}N labeled HSQC spectrum of HP1423 (B) Val Residues specific ^{15}N labeled HSQC spectrum of HP1423.

Interestingly, the peak disappearance region was predicted to form α helix from secondary prediction program from TIGR server (www.tigr.org) and comparison with the α L-motif region of Hsp15[14] which is unknown protein belonging to S4 superfamily (Fig. 4).

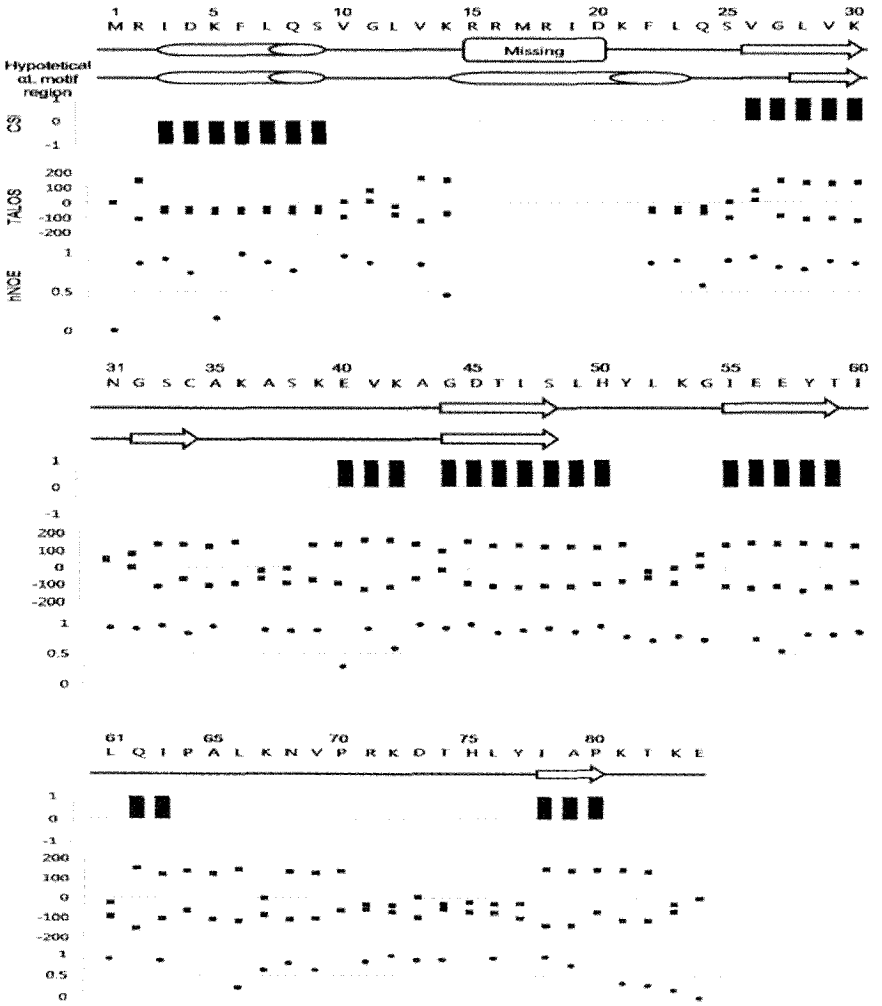

Figure 4. Secondary structure and local flexibility of HP1423. Summary of backbone resonance assignment of HP1423. Consensus-CSI shows the secondary structure predicted from the chemical shift indices of ^{1}H and ^{13}C. The values "-1" represents for the α-helical and "1" for β-strand tendency. TALOS showed the predicted backbone dihedral angels (phi, psi) with Filled quadrangle. The determined secondary structure is displayed with the cylinder for α-helix and the arrows for β-

strand. Secondary structure prediction of αL-motif region is predicted using the second structure prediction program from sequence of HP1423. The secondary structure and sequence of αL-motif region of Hsp15 which is also belonging to S4 superfamily, is represented below the secondary structure prediction graph. The result of Heteronuclear ^{1}H-^{15}N NOEs experiment is shown as filled circles.

These disappeared peaks might be caused by conformation exchange processes on microsecond to millisecond time scales in solution under equilibrium conditions[15]. High frequency motions of backbones represented by results of hetero-NOE experiment indicated that the backbone of HP1423 is generally rigid on picoseconds to nanosecond time scales. According to above mentioned experiments, probably this peak disappearance is caused by slow conformational exchange between α -helix and random-coil. This flexible character may be important for α helix to recognize RNA.

In conclusion, HP1423 would be a candidate for studying the slow conformational changes and the structural differences between apo and RNA binding form. The structural characterization of HP1423 will help us gain insights into their function and folding mechanism.

Table 1. Chemical shifts of ^{1}HN, ^{15}N, ^{13}CO, ^{13}Cα and ^{13}Cβ of HP1423. All chemical shifts were referenced to the frequency of the methyl proton resonance of DSS.

Num	AA	HN	N	CA	CB	CO	Num	AA	HN	N	CA	CB	CO
1	M	nd	nd	54.849	nd	172.18	43	A	8.534	122.06	53.734	17.672	177.99
2	R	8.76	128.17	56.899	34.547	178.39	44	G	9.122	112.42	44.689	NA	174.61
3	I	9.002	120.93	64.118	37.931	175.31	45	D	7.717	121.25	55.37	42.58	175.44
4	D	8.056	122.1	57.68	37.991	177.15	46	T	8.624	114.33	61.414	70.622	174.5
5	K	8.087	123.56	58.472	32.381	178.79	47	I	9.411	125.67	59.341	42.038	174.55
6	P	NA	NA	62.791	39.505	176.35	48	S	9.325	121.14	56.639	63.958	173.26
7	L	8.419	116.79	58.092	42.007	178.62	49	L	9.461	127.26	53.722	43.056	175.36
8	Q	7.213	116.7	57.894	28.021	179.12	50	H	9.223	123.5	54.406	27.834	173.86
9	S	8.629	118.97	62.595	ND	179.13	51	Y	7.576	121.67	56.831	39.393	176.96
10	V	7.259	108.08	61.122	30.631	175.25	52	L	8.826	124.35	58.408	41.474	179.54
11	G	7.427	107.24	46.011	NA	175	53	K	8.05	113.73	56.487	32.08	175.99
12	L	7.603	119.86	56.612	42.379	ND	54	G	7.322	108.47	56.455	NA	171.75
13	V	6.994	112.08	59.955	35.05	ND	55	I	8.201	120.98	60.771	39.68	176.67
14	K	8.604	121.26	58.754	33.618	ND	56	E	8.594	126.6	55.027	33.755	174.6
15	R	ND	ND	ND	ND	ND	57	E	8.287	121.25	54.773	33.655	175.37
16	R	ND	ND	ND	ND	ND	58	Y	9.297	119.94	56.957	43.685	174.63
17	V	ND	ND	ND	ND	ND	59	T	9.396	118.62	61.825	70.508	174.42
18	L	ND	ND	ND	ND	ND	60	I	8.946	124.91	63.335	37.499	176.12
19	A	ND	ND	ND	ND	ND	61	L	9.114	130.43	56.027	42.34	177.15
20	T	ND	ND	ND	ND	ND	62	Q	7.938	114.73	54.668	32.106	172.35
21	D	ND	ND	57.952	40.372	ND	63	I	8.745	122.48	55.883	37.102	175.44
22	M	8.087	118.35	59.472	33.773	178.58	64	P	NA	NA	63.223	31.784	175.25
23	C	7.949	116.75	61.9	26.847	178.77	65	A	8.512	123.92	51.616	19.255	177.13
24	N	8.566	120.25	55.438	37.686	177.62	66	L	7.062	118.3	54.113	43.748	176.48
25	V	7.807	111.96	61.547	31.761	176.93	67	K	8.183	119.99	57.845	32.891	176.18
26	G	7.974	109.91	46.939	NA	174.95	68	N	7.722	114.99	52.796	40.313	173.61
27	A	7.848	119.9	52.409	20.156	175.91	69	N	8.348	124.54	60.084	32.557	174.03
28	V	8.712	119.11	62.07	33.413	173.98	70	P	NA	NA	62.908	32.335	177.59
29	W	9.533	130.34	55.607	30.284	174.15	71	R	8.648	124.15	59.109	29.607	179.07
30	L	9.095	124.14	52.715	46.32	176.48	72	K	8.515	116.33	58.106	31.79	176.11
31	N	9.785	125.7	54.538	37.109	ND	73	D	8.043	117.42	54.276	42.756	177.45
32	G	9.033	103.42	45.834	NA	173.95	74	T	7.865	117.51	68.493	ND	176.64
33	S	7.768	114.44	57.096	65.589	172.24	75	H	8.578	116.35	57.381	28.157	174.3
34	C	8.045	125.5	61.005	26.12	174.08	76	L	7.473	118.07	55.187	40.646	177.37
35	A	8.686	132.32	51.23	21.05	176.06	77	Y	8.062	114.88	59.996	41.013	176.14
36	K	7.65	118.88	55.148	33.807	176.86	78	I	7.988	114.5	58.945	43.843	173.02
37	A	8.498	121.68	55.156	19.792	174.89	79	A	8.684	123.82	48.6	21.441	174.86
38	S	7.001	101.24	56.598	64.141	174.69	80	P	NA	NA	62.901	32.219	176.53
39	K	7.73	125.72	56.595	32.583	175.63	81	K	8.068	125.22	56.009	33.282	176.31
40	E	8.301	125.12	56.184	30.644	177.25	82	T	8.221	118.42	61.567	70.031	174.13
41	V	7.909	113.55	59.072	33.892	174.6	83	K	8.422	124.29	56.463	32.972	176.46
42	K	8.736	120.55	54.44	36.222	174.5	84	E	8.453	122.82	56.559	30.077	176.39

ND : Not Detected , NA : Not Available

Acknowledgement

This work is financially supported by Innovative Drug Research Center for Metabolic and Inflammatory Disease and 2009 BK21 Project for Medicine, Dentistry and Pharmacy. We thank NCIRF and KBSI for using NMR machines.

REFERENCES

1. T.L. Cover, M.J. Blaser, *Adv. Intern. Med.* ***41,*** 85-117, (1996).
2. S.B. Jang, C. Ma, S.J. Park, A.R. Kwon, B.J. Lee. *JKMRS* ***13,***117-125, (2009).
3. W.J. Kim, J.S. Lim, W.S. Son, H.C. Ahn, B.J. Lee, *JKMRS* ***12,*** 65-73, (2008).
4. A.P. Carter, W.M. Jr. Clemons, D.E. Brodersen, R.J. Morgan-Warren, B.T. Wimberly, *Nature.* ***407,*** 340-348, (2000).
5. A. Matte, G.V. Louie, J. Sivaraman, M. Cygler, S.K. Burley, *Acta Crystallogr. Sect. F :Struct. Biol. Cryst. Commun.* ***61(Pt 4),*** 350-354, (2005).
6. A. Yaremchuk, I. Kriklivyi, M. Tukalo, S.Cusack, *EMBO J.**21,*** 3829-3840, (2000).
7. B.L. Staker, P. Korber, J.C. Bardwell, M.A. Saper, *EMBO J.* ***19,*** 749-757, (2000).
8. L. Volpon, C. Lievre, M.J. Osborne, S. Gandhi, P. Iannuzzi, R. Larocque, A. Matte, M. Cygler, K. Gehring, I. Ekiel, *J. Bacteriol.**185,*** 4204-4210, (2003).
9. L. Aravind, E.V. Koonin, *J. Mol. Evol.* ***48,*** 291-302, (1999).
10. F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfifer, A. Bax, *J. Biomol. NMR* ***6,*** 277-293, (1995).
11. B.A. Johnson, *Methods Mol. Biol.* ***278,*** 313-352, (2004).
12. D.S. Wishart, B.D. Sykes, *J. Biomol. NMR* ***4,*** 171-180, (1994).
13. G. Cornilescu, F. Delaglio, A. Bax, *J. Biomol. NMR* ***13,*** 289-302, (1999).
14. B.L. Staker, P. Korber, J.C. Bardwell, M.A. Saper. *EMBO J.* ***19,*** 749-757, (2000).
15. K. Henzler-Wildman, D. Kern, *Nature* ***450,*** 964-972, (2007).