Journal of the Korean Magnetic Resonance Society **2006**, *10*, 96 - 104

NMR Structural Analysis and 3D Homology Modelling of APG8a from *Arabidopsis thaliana*

Young Kee Chae

Department of Chemistry and Recombinant Protein Expression Center (RPEC), Sejong University, 98 Gunja-Dong, Gwangjin-Gu, Seoul, Korea 143-747
Received March 12, 2006

Abstract : The gene coding for APG8a (At4g21980), a protein from *Arabidopsis thaliana*, is involved in the autophagy process. The protein is an interesting candidate for structure determination by NMR spectroscopy. Toward this end, APG8a has been produced recombinantly in *Escherichia coli* and typical NMR experiments such as ^{15}N-HSQC, HNCA, HN(CO)CA, CBCA(CO)NH, HCCH-TOCSY, HNCO were performed. The backbone resonances, HN, N, CA, CB, and C' were sequence-specifically assigned, and the secondary structures including 3 α helices and 4 β strands were deduced based on the assignments. Due to the intrinsic flexibility or the effect of the denaturant, the backbone resonances were not fully observed. Since the structure calculation by NMR data was not possible, the 3-dimensional model was built based on the sequence homology, and compared with the NMR results. The overall structure of the model could explain and complement the NMR derived secondary structures.
Keywords : NMR, modelling, homology, protein, structure

INTRODUCTION

APG8a from *Arabidopsis thaliana* is a single-chain protein of 122 amino acids with Mr = 13.9 kDa. The protein is involved in the autophagy process, which plays a key role in recycling during starvation and senescence.[1] By the ATP-dependent reaction cascade, APG8a becomes conjugated with phosphatidylethanolamine.[2] Prior to activation, the APG8a precursor is processed by the APG4 protease, which removes the five C-terminal residues and exposes a glycine residue (G117) as the C-terminus. After successful overexpression, purification, and refolding of APG8a,[3] we proceeded to the structural study

To whom correspondence should be addressed. E-mail : ykchae@sejong.ac.kr

and finished the sequence-specific assignments.[4] Since the NMR data were not complete enough, we decided to model the structure rather than to go for more data collection. This protein had sequence homology to proteins (Fig. 1) such as GATE-16 from *Bos taurus*,[5] Gaba(A) Receptor Associated Protein from *Homo sapiens*,[6] and from *Rattus norvegicus*,[7] Lc3 from *Rattus norvegicus*,[8] and Microtubule-Associated Protein Light Chain-3 from *Homo sapiens*.[9] In this article, we report the effort and result of the structural study of APG8a by NMR spectroscopy and homology modeling.

MATERIALS AND METHODS

Sample Preparation

The [^{13}C, ^{15}N]-doubly labeled sample was prepared as described elsewhere.[3] The protein sequence included the first 117 amino acids of APG8a with a consideration that the last 5 residues, 118 to 122, were to be removed upon maturation, and had additional Gly-Ser dipeptide at the N-terminus which was the remnant of the thrombin cleavage. Total number of amino acids was thus 119. The NMR sample contained 1 mM APG8a in 20 mM sodium phosphate buffer pH 7.0 with 1% glycerol, 0.5 M urea, 300 mM NaCl, 10 % D_2O, and 1 mM DSS. NMR samples were replaced after experiments every 2 days.

NMR Experiments.

As reported previously,[4] [^{15}N]-HSQC, [^{13}C]-HSQC, HNCA, HN(CO)CA, and HNCO[11,12] data were collected on a Bruker DMX-750 spectrometer. CBCA(CO)NH data were collected on a Varian Inova 800 spectrometer. All experiments were performed at 25 °C. The acquired data contained 1024 (^{1}H) x 128 (^{15}N) complex points for [^{15}N] HSQC, 1024 (^{1}H) x 256 (^{13}C) for [^{13}C] HSQC, 1024 (^{1}H) x 40 (^{15}N) x 40 (^{13}C) for HNCA and HN(CO)CA, 1024 (^{1}H) x 40 (^{15}N) x 32 (^{13}C) for HNCO, and 512 (^{1}H) x 32 (^{15}N) x 62 (^{13}C) for CBCA(CO)NH. The internal standard, DSS, was used to reference the spectra.

NMR Data Processing and Analysis

As previously reported,[4] the raw NMR data were processed by using NMRPipe[13] on

a PC Linux system. The final matrices contained 1024 (^{1}H) x 256 (^{15}N) real points for [^{15}N] HSQC, 2048 (^{1}H) x 512 (^{13}C) for [^{13}C] HSQC, 1024 (^{1}H) x 128 (^{15}N) x 128 (^{13}C) for HNCA, HN(CO)CA, and HNCO, and 512 (^{1}H) x 128 (^{15}N) x 256 (^{13}C) for CBCA(CO)NH. The processed data were converted to the format compatible with Sparky.[14] Backbone assignments were obtained manually. The secondary structure was deduced from the chemical shift index.[15]

Homology Modelling

Homology modeling was performed by using Modeller[16] on a PC Linux system. The template was GATE-16 from *Bos taurus*[5] which showed the highest sequence similarity to APG8a and whose structure was reported. The primary sequences were first aligned against each other, and the models were built. 100 structures were generated, and the one with smallest "Modeller Objective Function" was selected. The selected model was visualized with MOLMOL (Koradi et al., 1996), and evaluated and analyzed with PROSAII (Sippl, 1993) and PROCHECK (Laskowski et al., 1993).

RESULTS AND DISCUSSIONS

Spectral Assignments

The resonances from amide protons (HN), amide nitrogens (N), alpha carbons (CA), beta carbons (CB), and carbonyl carbons (C') were deposited at BioMagResBank (BMRB accession no. 6610). Probably due to the intrinsic flexibility or the denaturant (0.5 M urea) added to provide better solubility, resonances from several parts of the protein were not observed. The residues whose amide protons were not observed were residues 1, 4 to 13, 39 to 41, 48, and 83 which comprised 13.6 % (16/117) of the whole sequence. However, most of the unobserved resonances (11 out of 16) originated from the N-terminus, which could be considered plausible because it is the region where such phenomenon is generally observed.

Secondary Structural Elements

The secondary structural elements were deduced based on the consensus chemical shi-

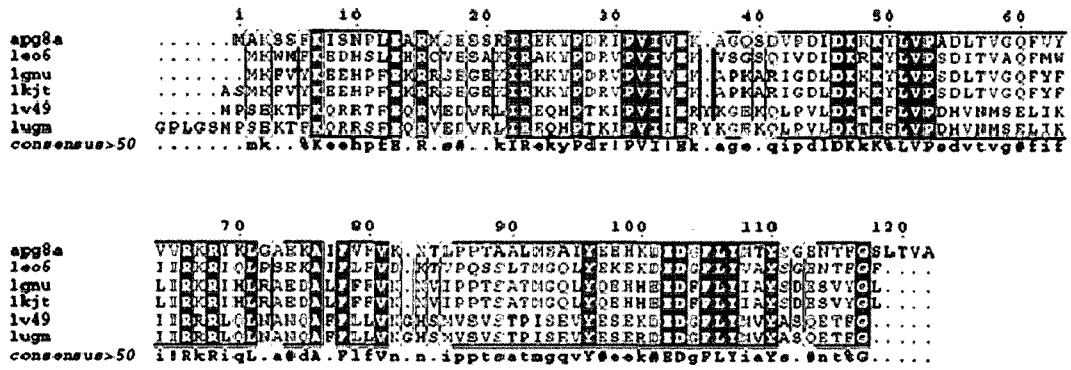


Fig. 1. Multiple alignment of sequences homologous to APG8a. The alignment was done at the MultAlin site (http://prodes.toulouse.inra.fr/multalin/multalin.html). The boxed plot was created at the ESPript site (http://espript.ibcp.fr/ESPript/ESPript/).

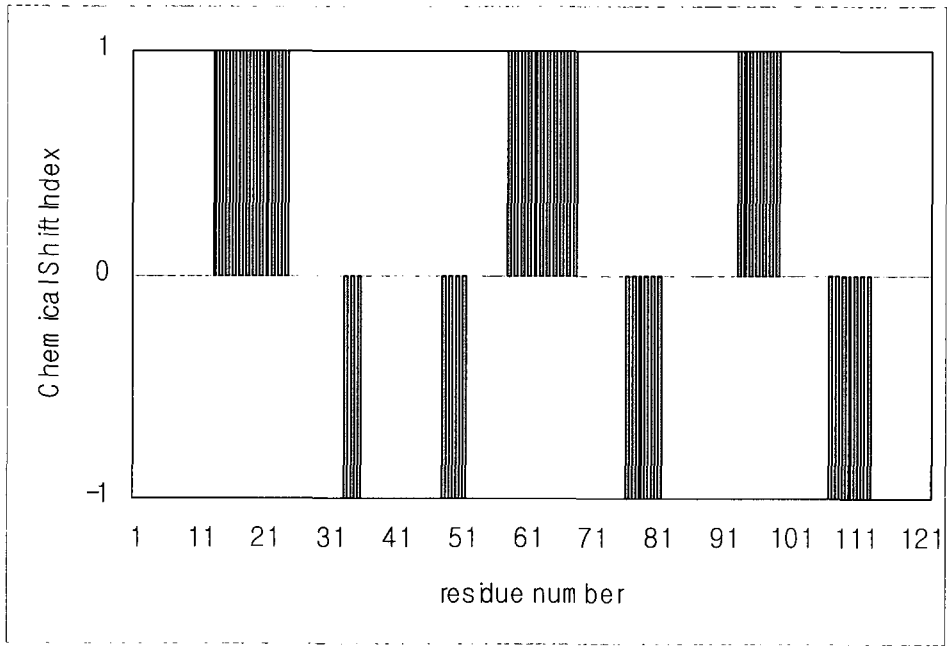


Fig. 2. Prediction of secondary structures by Chemical Shift Index (CSI) based on H^N, H^α, N, C^α, C^β, and C′ chemical shifts and the protein sequence.

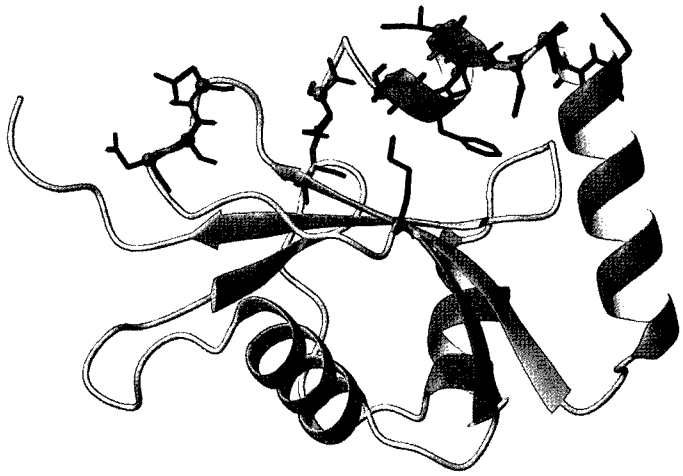
ft index using three resonance types, C^{α}, C^{β}, and C'. As shown in Fig. 2, there are 3 helices and 4 strands. The residues 13 to 24, 58 to 68, and 93 to 99 constituted the three helices, and the residues 33 to 35, 48 to 51, 75 to 81, and 107 to 113, the four strands. As a whole, those elements were arranged in a α-β-β-α-β-α-β topology. This topology was almost identical with that of the homologous proteins except the additional helix at the N-terminus where NMR signals were not detected in APG8a.

Comparative Modelling

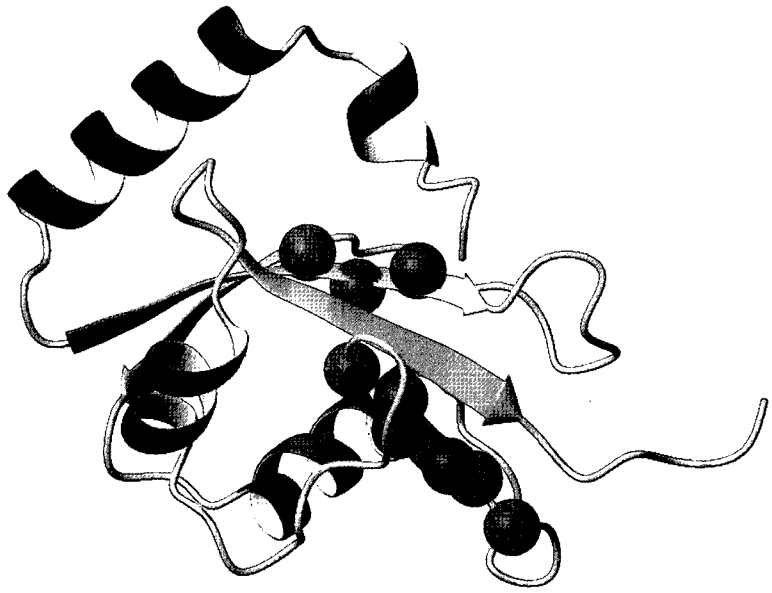
The 3D model of APG8a was built against the structure of GATE-16 from *Bos taurus*. Among the 100 calculated model structures, one with the lowest "Modeller Objective Function" was chosen. The values ranged from 467.7 to 835.8 with the average of 561.1. The selected structure was examined with PROSAII. In Fig. 4a, the combined energy was plotted against residue numbers. The energies of both APG8a and 1EO6 were below zero with window size of 20, which means the model structure was reasonable. The overall shapes of the combined energy are similar in the two structures although APG8a showed generally higher values from residues 60 to 100. The model structure was checked with PROCHECK, and the Ramachandran plot (Fig. 4b) showed none of the residues were in the disallowed regions, which confirmed the reasonability of the model.

Comparison of the NMR data with the 3D Model

In Fig. 3a, the residues whose amide proton resonances were not observed (except pralines) were represented with the ball-and-stick model. Interestingly, the regions presumed to be disordered or flexible were clustered together in the three-dimensional space According to the model, there were 4 α helices and 4 β strands. The residues 5 to 9, 12 to 25 58 to 69, and 92 to 99 comprised 4 α helices, and 29 to 36, 49 to 53, 78 to 80, and 106 to 111, 4 β strands. The NMR data were insufficient to support the existence of the first a helix but the other helices showed good agreement with the model. As for the 4 β strands, the overall locations were the same in both NMR data and the model. However, the NMR data predicted much shorter one (3 residue long) for the first β strand than the model (7 residue long) and longer one (6 residue long) for the third strand than the model (3 residue long). This was depicted in Fig. 3b. In the model structure, the region flanking the third strand was

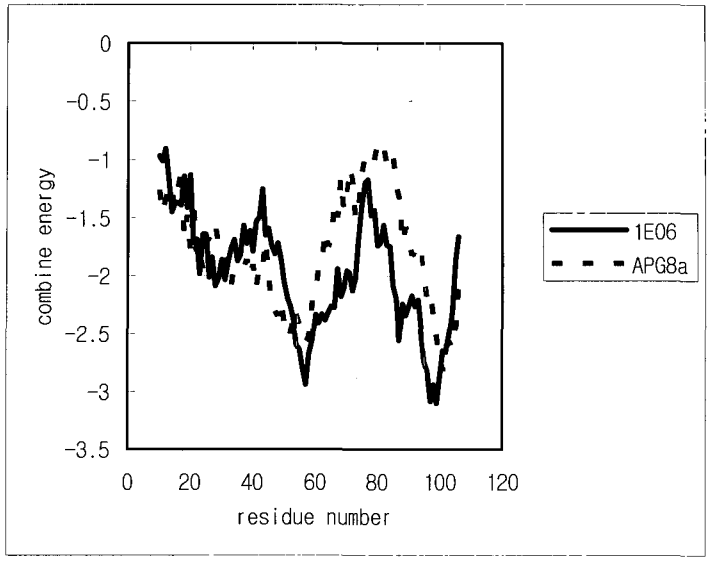


(a)

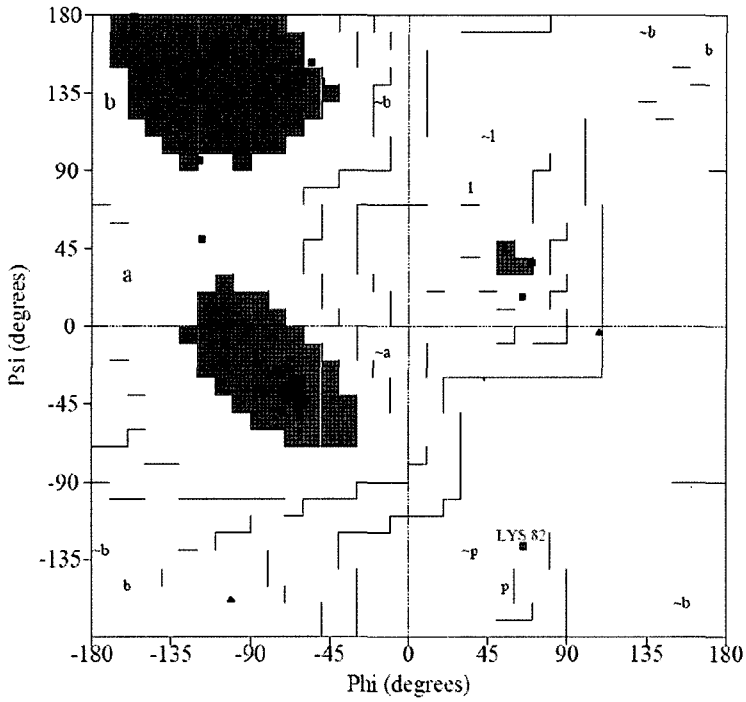


(b)

Fig. 3. A model structure with the lowest target value created with MOLMOL. (a) The unobserved residues were represented with the ball-and-stick model. (b) The α carbons of the predicted strand.



(a)



(b)

Fig. 4. (a) The PROSA energy plot of APG8 and 1EO6. (b) Ramachandran plot of the model APG8a.

quite extended, and it could be turned into the regular strand without spending much energy, or this could be the one of the structural differences between the known structure (1EO6) and APG8a. Since the NMR data lacked information on the proline residues, it was hard to reason why the first strand was predicted much shorter based on the NMR data. However, the first strand is very close to the regions that were not observed, and the flexibility or disorder of those regions might influence the regularity of that strand.

Acknowledgement

This work was supported by Korea Research Foundation Grant funded by Korea Government (MOEHRD, Basic Research Promotion Fund) (KRF-2005-003-C00111)

REFERENCES

1. J. H. Doelling, J. M. Walker, E. M. Friedman, A. R. Thompson, and R. D. Vierstra, *J. Biol. Chem.* **277**, 33105 (2002).
2. N. Mizushima, T. Noda, T. Yoshimori, Y. Tanaka, T. Ishii, M. D. George, D. J. Klionsky, M. Ohsumi, and Y. Ohsumi, *Nature*, **395**, 395 (1998).
3. Y. K. Chae, H. Im, Q. Zhao, J. H. Doelling, R. D. Vierstra, and J. L. Markley, *Protein Expr Purif.* **34(2)**, 280 (2004).
4. Y. K. Chae, K. Lee, and J. L. Markley. *J. Biomol. NMR*, **32(4)**, 337 (2005).
5. Y. Paz, Z. Elazar, and D. Fass, *J. Biol. Chem.* **275(33)**, 25445 (2000).
6. D. Knight, R. Harris, M. S. McAlister, J. P. Phelan, S. Geddes, S. J. Moss, P. C. Driscoll and N. H. Keep, *J. Biol. Chem.* **277(7)**, 5556 (2002).
7. V. N. Bavro, M, Sola, A. Bracher, M. Kneussel, H. Betz, and W. Weissenhorn, *EMBO Rep.* **3(2)**, 183 (2002).
8. K. Sugawara, N. N. Suzuki, Y. Fujioka, N. Mizushima, Y. Ohsumi, and F. Inagaki, *Genes Cells*, **9(7)**, 611 (2004).
9. T. Kouno, M. Mizuguchi, I. Tanida, T. Ueno, T. Kanematsu, Y. Mori, H. Shinoda, M. Hirata, E. Kominami, and K. Kawano, *J. Biol. Chem.* **280(26)**, 24610 (2005)

10. K. H. Gardner and L. E. Kay, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 357 (1998).
11. J. G. Pelton and D. E. Wemmer, *Annu. Rev. Phys. Chem.* **46**, 139 (1995).
12. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, *J. Biomol. NMR* **6**, 277 (1995).
13. T. D. Goddard, and D. G. Kneller, SPARKY 3.110, University of California, San Francisco (2004)
14. D. S. Wishart and B. D. Sykes, *Methods Enzymol.* **239**, 363 (1994).
    A. Šali and T. L. Blundell, *J. Mol. Biol.* **234**, 779 (1993).
15. R. Koradi, M. Billeter, and K. Wüthrich, *J. Mol. Graphics*, **14**, 51 (1996).
16. M. J. Sippl, *Proteins*, **17**, 355 (1993).
17. R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, *J. Appl. Cryst.* **26**, 283-291 (1993).