The Homodimerization of *Thalictrum tuberosum O*-Methyltransferases by Homology-based Modelling

Heejung Yang, Joong-Hoon Ahn, Karp Joo Jeong, Sangsan Lee,[†] and Yoongho Lim[*]

Bio/Molecular Informatics Center, Konkuk University, Seoul 143-701, Korea
[†]Supercomputing Center, Korea Institute of Science and Technology Information, Daejeon 305-350, Korea

Two *O*-methyltransferases, OMTII-1 and OMTII-4 of meadow rue *Thalictrum tuberosum* showed a high sequence identity. Of 364 amino acids only one residue is not the same, which is Tyr21 or Cys21. Even if the 21st residues in these OMTs are not included in the binding sites of the enzymes, binding affinities of the enzyme homodimers over the same substrate are very different. While the binding affinity of one homodimer over caffeic acid is 100%, that of the other is 25%. Authors tried to predict the three-dimensional structures of *Thalictrum tuberosum O*-methyltransferases using homology-based modelling by a comparison with caffeic acid *O*-methyltransferase, and explain the reason of the phenomenon mentioned above based on their three dimensional structural studies. In the enzyme homodimer, the better binding affinity may be caused by the shorter distance between the 21st residue and the binding site of the other monomer.

Key Words : Modelling, Molecular dynamics, Homology, Methyltransferase, *Thalictrum tuberosum*

Introduction

Since secondary metabolites of plants play an important role in improving disease resistance of plants against pathogens as well as promoting human health, an increasing number of research groups are studying their biosynthetic pathways.[1] Among many enzymes participated in their biosynthetic pathways, methyltransferases are ubiquitous and create more diverse class of metabolites. Secondary metabolites such as lignin, suberin, flavonoids, antocyanins, and isoflavonoids which are synthesized *via* phenylpropanoid pathway are also modified by methyltransferases.[2,3] Because they show diversity and play an important role in plant growth and development as well as the interactions of plants with their environment, many studies on methyltransferases have been carried out.[4] Especially, *O*-methyltransferases (OMTs) are more common. OMTs can be classified based on their functions: (1) The first group methylates phenyl-propanoids in the lignin biosynthesis pathway. (2) The second group methylates flavonoids, including chalcones, flavonols, and flavones. (3) The third group methylates alkaloids and other chemicals such as sugars and scent compounds.[5] In all three groups, OMTs transfer the methyl group of *S*-adenosyl-L-methionine (SAM) to the hydroxyl group of phenolic acceptors, and form their methyl ether derivatives and *S*-adenosyl-L-homocysteine (SAH) as products.[6] *O*-methylation not only reduces the chemical reactivity of their phenolic hydroxyl groups but also increases their lipophilicity and their intracellular com-partmentation.[7] Therefore, to know the function of OMTs at a molecular level, many genes encoding OMTs have been isolated and are characterized. In contrast with mammalian OMTs, plant OMTs show narrow substrate specificities as well as position-specific activities, so that the homology-based comparison such as BLAST could not give enough information about the biological function of each OMTs such as substrate and product.[8] Molecular modelling method may help us understand the reason why plant OMTs show narrow substrate specificities as well as position-specific activities.[9,10]

It is known that methyl jasmonate can induce an alkaloid, berberine in cell cultures of meadow rue *Thalictrum tubero-sum*.[11] Four cDNA encoding *O*-methyltransferases were isolated. Among these, two cDNAs named OMT II-1 and OMT II-4 showed high sequence homology. Of 364 amino acids only one residue is not the same, which is Tyr21 of OMT II-1 and Cys21 of OMT II-4. Even though both enzymes have the same substrate, caffeic acid, their binding affinities are different. The binding affinity values over caffeic acid of enzyme homodimers were reported by Frick and Kutchan.[11] While in the case of OMT II-1 homodimer the binding affinity over caffeic acid is 100%, that of OMT II-4 homodimer is 25%. Of course, if the different amino acid is positioned at the binding site, this phenomenon may be explained. Compared with residues neighboring the binding site of caffeic acid *O*-methyltransferase (COMT) which shows a high homology over 74% with OMT II, their binding site can be predicted but the 21st residue is not included there.[8] Authors tried to predict the three-dimensional (3D) structures of *Thalictrum tuberosum* OMTs using homology-based modelling. A template for homology-based modelling was COMT deposited in Protein Data Bank (PDB code: 1KYZ.pdb).[12] Based on their 3D structural studies, we tried to explain the reason why two enzyme homodimers show very different binding affinity over caffeic acid.

---
[*]Corresponding author. Tel: +82-2-450-3760; Fax: +82-2-453-3761; E-mail: yoongho@konkuk.ac.kr

Materials and Methods

Target and template proteins. The amino acid sequence of OMT II-1 was obtained from GenBank (Accession No. AF064693; protein ID, AAD29841). It is composed of 364 amino acids. The primary sequence of OMT II-4 was obtained from GenBank (Accession No. AF064696; protein ID, AAD29844) too. There is only one different residue between OMT II-1 and OMT II-4, which is Tyr21 and Cys21, respectively. A template for homology-based modelling was COMT deposited in Protein Data Bank (PDB code: 1KYZ.pdb).[12] The number of amino acids of COMT is 365. A comparison of OMT II-1, OMT II-4, and COMT is shown in Figure 1. 1KYZ.pdb includes three polymer chains such as A, C, and E. Among the three chains, the chains A and C consisted of a dimer and the chain E exists as a monomer. Since homodimers are studied in this work, the chain A contained in COMT dimer was chosen as a template. The 3D structure of SAH produced from SAM by the enzyme was borrowed from 1KYZ.pdb. The 3D structure of the substrate, caffeic acid was adopted from 1KOU.pdb which is the complex of photoactive yellow protein and caffeic acid.[13]

Molecular modelling. The molecular modelling calculations were carried out on an O2 R12,000 Silicon Graphics work-station. The modelling software was InsightII (Accelrys, San Diego, USA.). The forcefield used for molecular dynamics (MD) and energy minimization was cvff provided by Accelrys. Homology module included in InsightII was used for homology modelling.[14] The sequence of the template protein was extracted and aligned with the target protein. The structurally conserved regions (SCRs) were determined manually. With SCRs, the sequence was aligned. The loops and the variable regions not to be included in SCRs were built. The conformations of side chains were determined by conformational search of rotamers. The protein was embedded in a 5 Å shell of 2,153 water molecules to imitate aqueous solvent conditions. The assembled molecules were subject to energy minimization by InsightII/Discover module. Steepest descents were carried out until maximum derivative of 1.0 kcal/molÅ, and conjugate gradients were followed until maximum derivative of 0.1 kcal/molÅ. After energy minimization, MD was performed at 300 K for 1,000 psec with 1 fsec each step. The output conformers were collected at every 4 psec. 250 conformers were saved in the history file. The energy profile was analyzed using InsightII/Analysis module. For a statistical evaluation of the conformer, PROCHECK was applied.

Docking of caffeic acid and SAH. The substrate, caffeic acid and the cofactor, SAH were docked into the protein using InsightII/Docking module on a Silicon Graphics O2 R12,000 workstation. A ligand and protein assembly was created using a grid docking method. First, a grid was prepared surrounding the protein. After setup the grid, the ligand was docked. The ligand and protein assembly was soaked in a 5 Å shell of 2,153 water molecules to imitate aqueous solvent conditions, and was subject to energy minimization. After energy minimization, MD was performed at 300 K for 1,000 psec with 1 fsec each step.

Homodimerization. In order to obtain a homodimer, the same enzyme was used as the ligand and the protein. Like docking a ligand and protein assembly, a grid was prepared surrounding the protein. After setup the grid, another protein monomer considered as a ligand was docked. A protein and another protein assembly was subject to energy minimization. After energy minimization, MD was performed at 300 K for 500 psec with 1 fsec each step.

Results and Discussion

Three dimensional structures of OMT II-1 and OMT II-4. The 3D structure of OMT II-1 obtained by homology-based modelling was compared with that of COMT. While COMT includes 20 α-helices and 9 β-sheets, OMT II-1 has 16 and 12, respectively. However, since the superimposition of backbones of two proteins gives RMSD value of 0.9 Å, it can be considered that they have a high similarity.

The structure of OMT II-1 was evaluated using PROCHECK. Among 303 residues except Gly and Pro only one residue, Thr346 was found in disallowed regions of Ramachandran plot. The statistical analysis of Ramachandran plot showed that 89.4% are in the most favored regions, 8.9% in

```
          HHHHH  HHHHHHH      HHHHHHHHHHH H  HHHHHHH       BB HH HHHH
OMTII-1  MGSTENNHNL TPAEEEEEEA YLHAMQLASA SVLPMVLKAA IELDVLEIIA KAGKGAYVAP  60
OMTII-4  MGSTENNHNL TPAEEEEEEA CLHAMQLASA SVLPMVLKAA IELDVLEIIA KAGKGAYVAP  60
COMT     MGSTG-ETQI TPTHISDEEA NLFAMQLASA SVLPMILKSA LELDLLEIIA KAGPGAQISP  59

          HHHHHHHH HHHHHH   B  BBBBBB       BBBBBBB H HHHHH          HHHHHH
OMTII-1  SEIASQLSTS NSQAPTVLDR MLRLLASYKV LTCNLRNLED GGVERLYGLA PVCKFLVKNE 120
OMTII-4  SEIASQLSTS NSQAPTVLDR MLRLLASYKV LTCNLRNLED GGVERLYGLA PVCKFLVKNE 120
COMT     IEIASQLPTT NPDAPVMLDR MLRLLACYII LTCSVRTQQD GKVQRLYGLA TVAKYLVKNE 119

          H  HHHHHHH       HHHHH      HHHHHH     HHHH    H HHHHHHHHHHH HHHHHHHHHH
OMTII-1  DGVSMAPLVL MNQDKVLMES WYHLKDAVLD GGIPFNKAYG MTAFEYHGTD PRFNKVFNRG 180
OMTII-4  DGVSMAPLVL MNQDKVLMES WYHLKDAVLD GGIPFNKAYG MTAFEYHGTD PRFNKVFNRG 180
COMT     DGVSISALNL MNQDKVLMES WYHLKDAVLD GGIPFNKAYG MTAFEYHGTD PRFNKVFNKG 179

          HHHH        BBBBB      HHHHHHH H  BBBBB B            BBBBB
OMTII-1  MADHSTITMK KLLELYKGFE GLKSVVDVGG GTGATVNMIV TKHPTIKGIN FDLPHVIDDA 240
OMTII-4  MADHSTITMK KLLELYKGFE GLKSVVDVGG GTGATVNMIV TKHPTIKGIN FDLPHVIDDA 240
COMT     MSDHSTITMK KILETYTGFE GLKSLVDVGG GTGAVINTIV SKYPTIKGIN FDLPHVIEDA 239

                      BB       HHHHHHHH HHHHHH    BBBBB BB        HHH
OMTII-1  PAYPGVEHIG GDMFVSVPKG DAIFMKWILH DWSDEHSVKF LKNCYESIPA DGKVIIVESV 300
OMTII-4  PAYPGVEHIG GDMFVSVPKG DAIFMKWILH DWSDEHSVKF LKNCYESIPA DGKVIIVESV 300
COMT     PSYPGVEHVG GDMFVSIPKA DAVFMKWICH DWSDEHCLKF LKNCYEALPD NGKVIVAECI 299

          HHHHHHHHH HHH     B BHHHHHHHHH HH  BBB     BBBB B
OMTII-1  LPVFPETNLA AHTCFQLDNI MLAHNPGGKE RTEKDFKALS VKAGFTGFKV VCGAFGSWVM 360
OMTII-4  LPVFPETNLA AHTCFQLDNI MLAHNPGGKE RTEKDFKALS VKAGFTGFKV VCGAFGSWVM 360
COMT     LPVAPDSSLA TKGVVHIDVI MLAHNPGGKE RTQKEFEDLA KGAGFQGFKV HCNAFNTYIM 359

OMTII-1  EFCK-- 364
OMTII-4  EFCK-- 364
COMT     EFLKKV 365
```

Figure 1. A comparison of the primary sequences of OMT II-1, OMT II-4, and COMT. (Bold characters denote Tyr21 of OMT II-1 and Cys21 of OMT II-4. H and B denote α-helix and β-sheet, respectively.)

Table 1. A comparison of α-helices and β-sheets included in OMT II-1, OMT II-4, and COMT (top: α-helices; bottom: β-sheets)

OMT II-1	OMT II-4	COMT
Asn6 - Glu17	Asn6 - Glu17	
Tyr21 - Ser31	Cys21 - Ser31	Glu18 - Ser28
Pro34 - Ala40	Pro34 - Ala40	Ser30 - Leu42
Ile49 - Lys54	Ile49 - Lys54	Asp43 - Ala51
		Ser58 - Ser64
Ile63 - Thr76	Ile63 - Thr76	Asp72 - Tyr87
Asp100 - Arg105	Asp100 - Arg105	Thr110 - Val116
Phe115 - Asp121	Phe115 - Asp121	
Ser124 - Leu130	Ser124 - Leu130	Ile124 - Gln132
Lys135 - Glu139	Lys135 - Glu139	Asp133 - Glu138
		Ser139 - Tyr141
His143 - Val148	His143 - Val148	His142 - Gly150
Gly152 - Phe155	Gly152 - Phe155	Ile152 - Tyr158
Gly160 - His184	Gly160 - His184	Thr161 - His166
		Asp169 - Tyr195
Ser204 - Gly211	Ser204 - Gly211	Gly212 - Thr217
		Ile218 - Tyr222
		Leu232 - Asp238
Ile263 - His276	Ile263 - His276	Ile267 - Trp271
		Ser272 - Leu287
Glu298 - Thr313	Glu298 - Thr313	Ser307 - Asn324
Leu322 - Thr332	Leu322 - Thr332	
		Gln332 Ala342

OMT II-1	OMT II-4	COMT
Leu46 - Glu47	Leu46 - Glu47	
Arg80 - Ala86	Arg80 - Ala86	
Thr92 - Leu98	Thr92 - Leu98	Leu90 - Thr96
		Val102 - Leu108
Leu193 - Lys197	Leu193 - Lys197	Leu204 - Val207
Val216 - Thr221	Val216 - Thr221	
His235 - Asp239	His235 - Asp239	Gly227 - Asp231
Asp252 - Met253	Asp252 - Met253	Val245 - Gly249
Asn283 - Ser287	Asn283 - Ser287	Val262 - Phe263
Pro289 - Ala290	Pro289 - Ala290	Lys292 - Leu300
Ile320 - Met321	Ile320 - Met321	
Lys337 - Leu339	Lys337 - Leu339	Val349 - Ala353
Gly347 - Val351	Gly347 - Val351	Thr356 - Leu362

additional allowed regions, and 1.3% in generously allowed regions. G-factor is 0.11.

OMT II and substrate assembly. The substrate, caffeic acid and cofactor, SAM binding sites of OMT II-1 and OMT II-4 were determined based on a comparison of the corresponding residues with binding sites of COMT (Table 2). In the case of the substrate binding site, Ile316 of COMT is switched with Leu317 in OMT II-1 and OMT II-4. However, in the case of the SAM binding site, all residues are agreed with each other. Even though OMT IIs form a substrate assembly, a comparison of OMT II-1 with OMT II-4 shows that there is no difference in both the substrate and the cofactor binding sites. Therefore, the different binding affinity over caffeic acid of OMT II-1 and OMT II-4 cannot be explained by their primary sequences composing the binding sites.

The quaternary structure of OMT II-1 and OMT II-4. In order to explain the binding difference in OMT II-1 and OMT II-4 in other way, we turn our attention to the quaternary structure of these proteins. The homodimer of OMT II-1 was obtained by docking OMT II-1 into another OMT II-1 (Figure 2). The N-terminals from each subunit lie in the opposite direction as found in COMT. It was known that COMT consists of two polymers A and C which are dimerized by contact of the N-terminal domain. To compare OMTII-1 homodimer with COMT homodimer, they are superimposed and its RMSD value is 8.1 Å (Figure 3). A comparison of OMTII-1 monomer with COMT monomer gives the RMSD value of 7.3 Å. That is, the case of dimmer shows better RMSD than that of monomer. As shown in Table 1, the N-terminal domain of COMT includes three helices between Glu18 and Ala51. However, OMT II-1
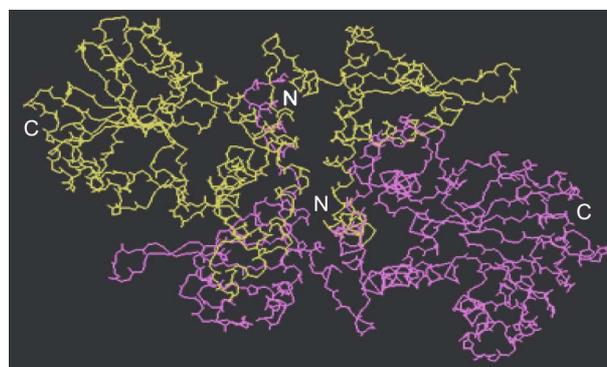

Figure 2. The structure of OMT II-1 homodimer. (white, a monomer; grey, another monomer)

Table 2. A comparison of residues neighboring binding sites of OMT II-1, OMT II-4, and COMT

Enzyme	substrate binding site												
COMT	Met130	Asn131	Leu136	Ala162	His166	Phe172	Phe176	Met180	His183	Ile316	Ile319	Met320	Asn324
OMTII-1	Met131	Asn132	Leu137	Ala163	His167	Phe173	Phe177	Met181	His184	Leu317	Ile320	Met321	Asn325
OMTII-4	Met131	Asn132	Leu137	Ala163	His167	Phe173	Phe177	Met181	His184	Leu317	Ile320	Met321	Asn325

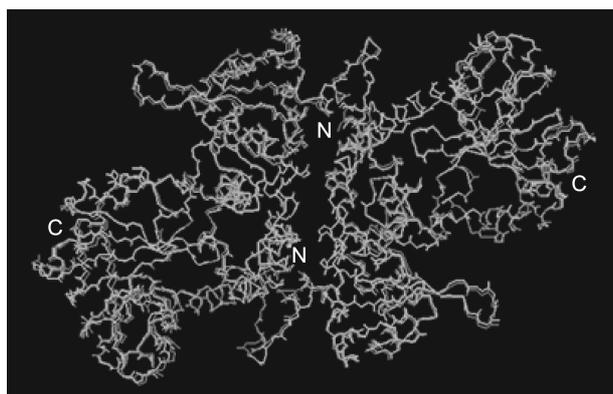| Enzyme	SAM binding site												
COMT	Asp206	Val207	Gly208	Gly209	Gly210	Thr211	Gly212	Asp231	Leu232	Asp251	Met252	Phe253	Lys265	Trp271
OMTII-1	Asp207	Val208	Gly209	Gly210	Gly211	Thr212	Gly213	Asp232	Leu233	Asp252	Met253	Phe254	Lys266	Trp272
OMTII-4	Asp207	Val208	Gly209	Gly210	Gly211	Thr212	Gly213	Asp232	Leu233	Asp252	Met253	Phe254	Lys266	Trp272

Figure 3. A superimposition of OMTII-1 homodimer (white) on COMT homodimer (grey).

appears to have one more helix that is composed of twelve amino acids between Asn6 and Glu17. Since the helix structure is not flexible, the presence of one more helix in OMT II-1 can make wider contact area and it could facilitate dimer formation. The driving forces for dimerization are electrostatic interactions and van der Waals interactions. As listed in Table 1, having one more helix in OMT II-1 results in more van der Waals force for dimer formation. Comparing the residues showing van der Waasls interaction, included in opposing α-helices of each N-terminal, OMT II-1 has nine more residues than COMT. It would confer more van der Waals force when OMT II-1 forms dimer than when COMT does. In addition, while there are six continuing Glu residues in the first helix of OMT II-1, there are only two Glu residues in COMT. As a result, it may cause stronger electrostatic interactions. Taken together, OMT II-1 is predicted to form dimer.[11] Two OMT cDNA clones were isolated from freshwater weed *Chrysosplenium americanum*.[15] They have only three residue difference but they show a large substrate specificity. While one OMT (NCBI Accession No.: P59049) shows 100% binding affinity for quercetin, another OMT (NCBI Accession No.: Q42653) does 38%. Of course, three residues are included in the binding sites. Therefore, this phenomenon can be explained based on the dimmer interface interaction.

OMT II-4 homodimer was also obtained by the same method as applied on OMT II-1. As shown in Figure 1, Tyr21 of OMT II-1 is switched with Cys21 in OMT II-4. However, the comparison of helices of OMT II-1 with those of OMT II-4 shows that there is no difference (Table 1). While OMT II-1 homodimer showed the binding affinity of 100% over caffeic acid, OMT II-4 homodimer did 25%.[11] Since OMT IIs are predicted to form dimer, the quaternary structures of OMT IIs cause the substrate binding differences.

The characteristic in the dimer formation of COMT is that Ser28 sticks out into the binding site of the other monomer.[16] At the same time, the residues being close to Ser 28 move closer to the binding site of the other monomer. It helps to exclude solvent in the substrate binding pocket. In the cases of OMT IIs, Ser29 matches Ser28 of COMT. The second helix of OMT IIs is composed of twelve residues between
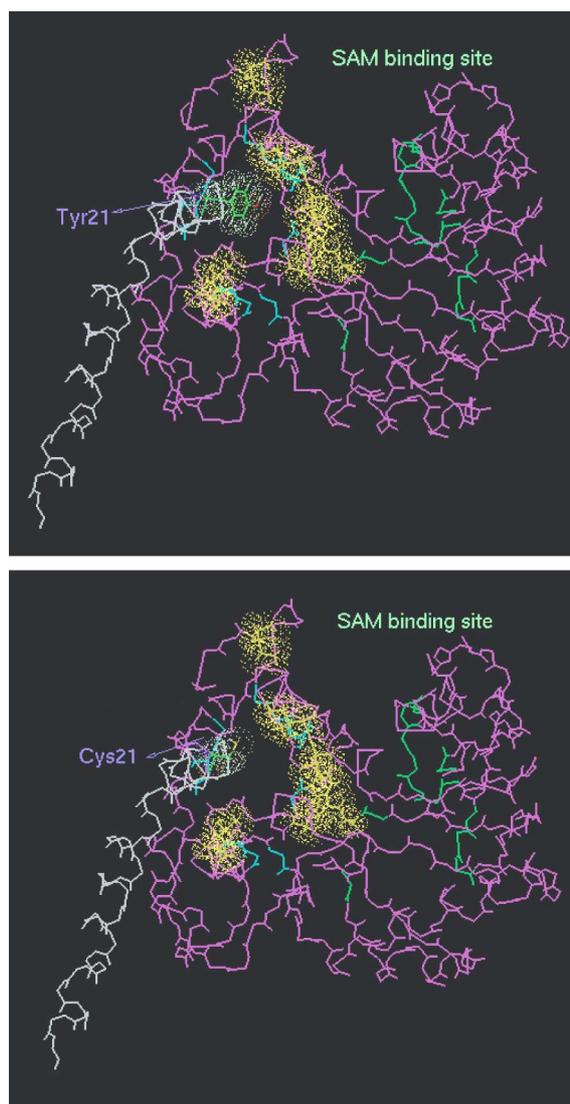

Figure 4. A comarison of the N-terminal domain of OMT II-1 homodimer backbone (top) with that of OMT II-4 homodimer backbone (bottom). (Dots denote the residues showing van der Waals interaction with Tyr21 or Cys21.)

the 21st residue and the 31st. Tyr21 and Cys21 are the starting residue of the helix. Therefore, the residues being close to Ser29 might move closer to the substrate binding pocket as those of COMT. To clarify this conjecture, the residues showing van der Waals interaction with Tyr21 of OMT II-1 were investigated in the other monomer. Eight residues of the other monomer were found, which were Phe115, Met125, His184, Ile187, Thr188, Thr313, Phe355, and Gly356. Unlike OMT II-1, only three residues were observed in Cys21 of OMT II-4, which were Phe115, Met125, and Phe355. This phenomenon can be explained by the reason why Tyr21 being closer to the binding site of the other monomer can help to keep out solvent molecules from the binding site and make the substrate access the binding site easier. As a result, the shorter distance between Tyr21 and the binding site of the other monomer probably causes the better binding affinity.

References

1. Dixon, R. A. *Nature* **2001**, *411*, 843.
2. Maury, S.; Geoffroy, P.; Legrand, M. *Plant Physiol.* **1999**, *121*, 215.
3. Zhong, R.; Morrison, W. H. III; Himmelsbach, D. S.; Poole, F. L.; Ye, Z. H. *Plant Physiol.* **2000**, *124*, 563.
4. Schaller, H.; Bouvier-Navé, P.; Benveniste, P. *Plant Physiol.* **1998**, *118*, 461.
5. Wang, J.; Pichersky, E. *Arch. Biochem. Biophy.* **1999**, *368*, 172.
6. Ibrahim, R. K.; Bruneau, A.; Bantignies, B. *Plant Mol. Biol.* **1998**, *36*, 1.
7. Muzac, I.; Wang, J.; Anzellotti, D.; Zhang, H.; Ibrahim, R. K. *Arch. Biochem. Biophy.* **2000**, *375*, 385.
8. Eckardt, N. A. *Plant Cell*. **2002**, *14*, 1185.
9. Moon, J. K.; Kim, J.; Rhee, S.; Kim, G.; Yun, H.; Chung, B.; Lee, S.; Lim, Y. *Bull. Korean Chem. Soc.* **2002**, *23*, 1545.
10. Choe, J.; Chang, S. *Bull. Korean Chem. Soc.* **2002**, *23*, 48.
11. Frick, S.; Kutchan, T. M. *Plant J.* **1999**, *17*, 329.
12. Wilmouth, R.; Turnbull, J.; Welford, R.; Clifton, I.; Prescott, A.; Schofield, C. *The Protein Data Bank. Structure (London)* **2002**, *10*, 93.
13. Van Aalten, D. M. F.; Crielaard, W.; Hellingwerf, K. J.; Joshua-Tor, L. *Acta Crystallogr. Sect. D* **2002**, *58*, 585.
14. Yoon, E. Y. *Bull. Korean Chem. Soc*. **2001**, *22*, 293.
15. Gauthier, A.; Gulick, P. J.; Ibrahim, R. K. *Arch. Biochem. Biophy.* **1998**, *351*, 243.
16. Zubieta, C.; Kota, P.; Ferrer, J.; Dixon, R. A.; Noel, J. P. *Plant Cell* **2002**, *14*, 1265.