

NMR Structure of the R-module A PARALLEL β -ROLL SUBUNIT FROM AN AZOTOBACTER VINELANDII MANNURONAN C-5 EPIMERASE*

Received for publication, September 13, 2005, and in revised form, December 13, 2005. Published, JBC Papers in Press, January 3, 2006, DOI 10.1074/jbc.M510069200

Finn L. Aachmann^{†§}, Britt I. G. Svanem[§], Peter Güntert[¶], Steffen B. Petersen^{||}, Svein Valla^{§1}, and Reinhard Wimmer^{‡2}

From the [‡]Department of Life Sciences, Aalborg University, Sohngaardsholmsvej 49, DK-9000 Aalborg, Denmark, [§]Department of Biotechnology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway, [¶]Tatsuo Miyazawa Memorial Program, RIKEN Genomic Sciences Center, 1-7-22 Suehiro, Tsurumi, Yokohama 230-0045, Japan, and ^{||}Institute of Physics and Nanotechnology, Aalborg University, Skjernvej 4A, DK-9000 Aalborg, Denmark

In the bacterium *Azotobacter vinelandii*, a family of seven secreted and calcium-dependent mannuronan C-5 epimerases (AlgE1–7) has been identified. These epimerases are responsible for the epimerization of β -D-mannuronic acid to α -L-guluronic acid in alginate polymers. The epimerases consist of two types of structural modules, designated A (one or two copies) and R (one to seven copies). The structure of the catalytically active A-module from the smallest epimerase AlgE4 (consisting of AR) has been solved recently. This paper describes the NMR structure of the R-module from AlgE4 and its titration with a substrate analogue and paramagnetic thulium ions. The R-module folds into a right-handed parallel β -roll. The overall shape of the R-module is an elongated molecule with a positively charged patch that interacts with the substrate. Titration of the R-module with thulium indicated possible calcium binding sites in the loops formed by the nonarepeat sequences in the N-terminal part of the molecule and the importance of calcium binding for the stability of the R-module. Structure calculations showed that calcium ions can be incorporated in these loops without structural violations and changes. Based on the structure and the electrostatic surface potential of both the A- and R-module from AlgE4, a model for the appearance of the whole protein is proposed.

Alginates are unbranched copolymers of 1 \rightarrow 4-linked β -D-mannuronic acid (M)³ and its C-5 epimer α -L-guluronic acid (G) (1, 2). Alginates are initially produced as a linear polymer of M-subunits only. Thereafter, single M-subunits in the polymeric chain are converted to G-subunits by enzymes called mannuronan C-5 epimerases. The mechanical and chemical properties of alginates depend on the composition and sequence of the two subunits, in particular, whether they are arranged in MM-, MG-, or GG-blocks. GG-blocks are rigid and form gels with divalent cations such as calcium. Recently, MG-blocks were suggested also to form gels with calcium (3). Poly(M) regions, however,

are more flexible without the ability to form gels with cations (4, 5). The relative amounts and distribution of M and G vary extensively among different species of both brown algae (6–8) and bacteria (9, 10)

Alginate is for brown algae essentially the same as cellulose is for trees and plants. The stripes of the algae are mainly G-rich, giving them the stiffness to work as a skeleton, whereas leaves are M-rich showing flexibility (11).

In *Azotobacter* sp., alginate is used as a capsular polysaccharide (12, 13), likewise in *Pseudomonas* sp. (14–16). In *Azotobacter vinelandii*, the composition and function of the alginates vary significantly depending on the environmental conditions.

Vegetatively growing cells produce alginates that form a loose capsule structure that is easily released into the growth medium. These alginates are typically M-rich (17), but under certain conditions of environmental stress, the cells enter a resting stage designated the “cyst” stage. The cysts are surrounded by a rigid alginate-containing protective coat, and due to the expression of a family of mannuronan C-5 epimerases (the AlgE family, see next paragraph), this coat also contains alginates with GG-blocks and strong gel-forming properties (17, 18).

In all known alginate-producing organisms, the alginates are modified at the polymer level by mannuronan C-5 epimerases converting M- to G-residues (19). The periplasmic epimerase AlgG is a part of alginate biosynthesis and is conserved in all known alginate-producing bacteria (20). Recently, the AlgG epimerase structure from *Pseudomonas aeruginosa* was suggested to contain a right-handed β -helix (21). This protein has, in addition, been shown to be essential for the protection of the alginate polymer against degradation by an alginate lyase, AlgL (22). Additionally, the bacterium *A. vinelandii* encodes a family of seven secreted and calcium ion-dependent mannuronan C-5 epimerases (AlgE1–7) (23). These epimerases consist of two types of structural modules, designated “A” (~385 amino acids each, with one or two copies) and “R” (~155 amino acids each, with one to seven copies), and additionally, a C-terminal signal peptide in the last of the R-modules (24, 25). The A-modules alone are catalytically active, but their reaction rates are significantly increased when covalently bound to at least one R-module (26). The R-module of AlgE4 is not catalytically active on its own (26). The N-terminal region of each R-module consists of four to seven copies of a nonapeptide with the consensus sequence LXG-GAGXDX, which is involved in calcium binding (24, 26). The individual sequence, number, and distribution of A- and R-modules give the individual epimerases their characteristic modes of action on this substrate. The R-module appears to have a significant role in the reaction catalyzed by the A-module by reducing the Ca²⁺ concentration needed for full activity and by enhancing the reaction rate (26). Recent experiments have also shown that the R-module influences the product specificity of

* This work was financed by the Danish Technical Research Council (STVF) and the Norwegian Research Council (NFR). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. The atomic coordinates and structure factors (code 2AGM) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ To whom correspondence may be addressed: Dept. of Biotechnology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway. Tel.: 47-73-59-86-94; Fax: 47-73-59-12-83; E-mail: svein.valla@biotech.ntnu.no.

² To whom correspondence may be addressed: Dept. of Life Sciences, Aalborg University, Sohngaardsholmsvej 49, DK-9000 Aalborg, Denmark. Tel.: 45-96-35-85-18; Fax: 45-98-14-18-08; E-mail: rw@bio.aau.dk.

³ The abbreviations used are: M, β -D-mannuronic acid; G, α -L-guluronic acid; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; RTX, repeats in toxins.

the attached A-module, *i.e.* the degree of GG- or MG-blocks produced (27).

Therefore, it is essential to study the structure of such an alginate epimerase and its modules in detail to achieve insight into its function. Knowledge of the three-dimensional structure of these proteins opens a range of further investigations concerning the metal binding properties and the substrate binding of alginate epimerases. Of special interest is the role of the R-module(s) after secretion and its/their influence on the function of the A-module(s).

Unsuccessful attempts have been made to crystallize the smallest epimerase (Alge4 consisting of one A- and one R-module). However, the A-module has been produced and crystallized separately, and recently the structure of this module was solved by x-ray crystallography (28). The A-module belongs to the family of pectate lyase folds. It folds as a single-stranded, right-handed parallel β -helix. This β -helix consists of 12 complete turns, made up of four β -strands each, except for the last turn at the C terminus, where the helix narrows down to three β -strands per turn. An N-terminal amphiphatic α -helix forms a cap to this part of the protein (28). Here, we report the structure of the R-module containing the C-terminal signal sequence and its possible calcium and alginate binding sites.

MATERIALS AND METHODS

Sample Preparation—Cloning, production, and purification of the uniformly $^{13}\text{C}/^{15}\text{N}$ -labeled R-module (167 amino acids) and the sample conditions for the NMR measurements were described previously (29). Samples for NMR studies contained 1.6 mM R-module in 20 mM HEPES buffer with 50 mM CaCl_2 at pH 6.9 in 95% $\text{H}_2\text{O}/5\%$ D_2O or 99.9% D_2O . All NMR experiments were recorded at 298 K on a Bruker DRX600 spectrometer equipped with a 5-mm *xyz*-gradient TXI(H/C/N) probe.

NMR Spectroscopy—The sequence-specific resonance assignments and the triple resonance NMR experiments used to obtain them are described in Ref. 29. Three-dimensional ^{13}C - and ^{15}N -edited $^1\text{H},^1\text{H}$ -NOESY spectra were recorded in D_2O and H_2O , respectively. The ^{15}N - ^1H heteronuclear NOEs were calculated from two independently measured and integrated ^{15}N -HSQC spectra as the ratio of the peak volumes with and without ^1H saturation. Nuclear magnetic relaxation (T_1 and T_2) measurements of ^{15}N were obtained by exponential fitting of the peak intensities in ^{15}N -HSQC spectra acquired with different relaxation delays (30, 31).

Metal Binding—The displacement of calcium ions in the R-module by paramagnetic thulium was achieved by adding small amounts of the aliquots of a 4.7 mM TmCl_3 solution for the first three steps until an R-module:calcium:thulium molar ratio of 1:25:0.5 was reached. Thereafter, aliquots of 47 mM TmCl_3 were added for the next thirteen steps until a (R-module:calcium:thulium) molar ratio of 1:25:10.5 was reached. Finally, aliquots of 470 mM TmCl_3 solution were added in three steps to a final ratio of (R-module:calcium:thulium) 1:25:30.75. In all titration steps, thulium was added directly to the uniformly ^{15}N -labeled R-module solution in the NMR sample tube in small volumes so that the increase in sample volume could be neglected. A ^{15}N -HSQC spectrum was recorded for each of the 19 titration steps.

Chemical shift changes in two-dimensional HSQC spectra were quantified by calculation of an absolute change in chemical shift as,

$$\Delta_{\text{HSQC}} = \sqrt{\Delta\nu_{\text{H}}^2 + \Delta\nu_{\text{N}}^2} \quad (\text{Eq. 1})$$

where $\Delta\nu_{\text{H}}$ is the chemical shift change of H^{N} atoms expressed in Hz, and $\Delta\nu_{\text{N}}$ is the chemical shift change of N atoms expressed in Hz.

Alginate Binding—The interaction between the R-module and alginate was investigated by adding small amounts of a 10 mM solution of

pentameric β -D-mannuronic acid (M_5), dissolved in a buffer identical to the NMR sample, to a solution of 0.15 mM $\text{U-}^{15}\text{N}$ -labeled R-module in 50 mM CaCl_2 , 20 mM HEPES, pH 6.9. M_5 was added in six portions until an R-module:alginate molar ratio of 1:1.2 was reached. In all titration steps, alginate was added in small volumes so that the increase in sample volume could be neglected. A ^{15}N -HSQC spectrum was recorded for each titration step, and spectral changes were quantified according to Equation 1.

Structure Calculation—NOESY cross-peaks were identified, integrated, and assigned in the aforementioned NOESY spectra using the program XEASY (32). The CALIBA (33) subroutine in CYANA was used to convert cross-peak intensities from NOESY spectra into distance constraints. Backbone torsion angle restraints were obtained from secondary chemical shifts using the program TALOS (34). On the basis of this input, the structure was calculated using the torsion angle dynamics program CYANA (35). Weak restraints on (ϕ, ψ) torsion angle pairs and on side chain torsion angles between tetrahedral carbon atoms were applied temporarily during the high temperature and cooling phases of the simulated annealing schedule to favor allowed regions of the Ramachandran plot and staggered rotamer positions, respectively. Structure calculations were started from 100 conformers with random torsion angle values. The 20 conformers with the lowest final CYANA target function values were embedded in a water shell of 8 Å thickness and energy-minimized against the AMBER force field (36) with the program OPALp (37).

Incorporation of Calcium Ions into the Structure Calculation—Additional structure calculations were performed with calcium ions incorporated into the structure. The calcium binding loops were determined from the thulium titration experiments. The binding geometry was inferred from alkaline metalloproteases from *P. aeruginosa* (Protein Data Bank (PDB) code 1KAP) and *Serratia marcescens* (PDB code 1SAT) (38, 39), two proteins with highly similar calcium binding motifs.

The calcium ion van der Waals radius was set to 1.12 Å (40). The positions of the Ca^{2+} ions were restrained by upper limit distance constraints between the respective Ca^{2+} ions and protein atoms using the calcium binding β -roll motifs from *P. aeruginosa* and *S. marcescens* (38, 39) as templates. One-half of the octahedral coordination is made up by the backbone oxygen atoms of G_2 (2.5 Å upper distance limit), G_4 (2.5 Å), and the side chain oxygen $\text{O}\delta^1$ of D_6 (2.5 Å). The coordination sphere is completed by the backbone oxygen atoms from G_1' (2.5 Å) and X_3' (2.5 Å) and the side chain oxygen $\text{O}\delta^2$ of D_6' (3.2 Å) of the neighboring GGXGX loop. To define the two last calcium binding sites, only distance constraints to the first half of the coordination sphere and D_6' were incorporated, because it was found in the x-ray structure of *P. aeruginosa* that the rest of the coordination sphere was made up by water ligands (39). Our structure calculations including Ca^{2+} ions are not intended to serve as a direct determination of the structural details of calcium binding, but merely to demonstrate the compatibility of our data with the known calcium binding mode of the nonapeptide motif.

RESULTS

The solution structure of the R-module was calculated on the basis of NOE upper distance limits and TALOS-derived (34) backbone torsion angle restraints (Table 1). The coordinate and constraint files were deposited in the PDB data base (accession code 2AGM).

Fig. 1 shows the sequence and the location of the secondary structure elements, whereas Fig. 2, A and C, show the three-dimensional structure. The structured N-terminal end of the R-module is a β -roll defined by three amino acids forming a short β -strand and six amino acids forming a loop to the next short β -strand followed by six amino acids

Structure of the R-module

TABLE 1
Characterization of energy-minimized NMR structures

	Without calcium ions	With calcium ions
Input data for the structure calculation		
Total number of NOE distance constraints	1,576	1,576
Intraresidual	575	575
Short range	418	418
Medium range	130	130
Long range	453	453
Number of upper distance limits for calcium ions		34
Torsion angle constraints ^a	138	138
Structure statistics, 20 conformers^b		
CYANA target function value (\AA^2) ^c	3.47 ± 0.53	4.60 ± 1.21
Maximal distance constraint violation (\AA)	0.13 ± 0.01	0.14 ± 0.01
Maximal torsion angle constraint violation (°)	3.88 ± 0.60	3.95 ± 1.37
AMBER energies (kcal/mol)		
Total	-4,537 ± 161	-7,769 ± 201
van der Waals	-397 ± 20	-149 ± 24
Electrostatic	-5,289 ± 167	-8,745 ± 180
PROCHECK Ramachandran plot analysis		
Residues in favored regions (%)	69.1	72.8
Residues in additionally allowed regions (%)	27.7	24.7
Residues in generously allowed regions (%)	2.9	2.2
Residues in disallowed regions (%)	0.3	0.3
Root mean square deviation to the average coordinates (\AA)^d		
N, C α , C' (5–145)	1.13	1.19
N, C α , C' (secondary structure) ^e	0.66	0.67
Heavy atoms (5–145)	1.66	1.68
Heavy atoms (secondary structure) ^e	1.20	1.21

^a Obtained from secondary chemical shifts using the program TALOS (34).

^b The values given are the average and standard deviation over the 20 energy-minimized conformers with the lowest CYANA (version 2.1) target function values that represent the NMR solution structure.

^c The average target function values for the 20 best CYANA conformers before energy minimization.

^d Average coordinates of the 20 energy-refined conformers after superposition for best fit of the N, C α , and C' atoms of the residues indicated in parentheses.

^e Residues 6–8, 15–17, 24–26, 33–35, 42–44, 51–53, 61–63, 66–68, 71–73, 81–83, 100–105, 111–115, and 127–129.

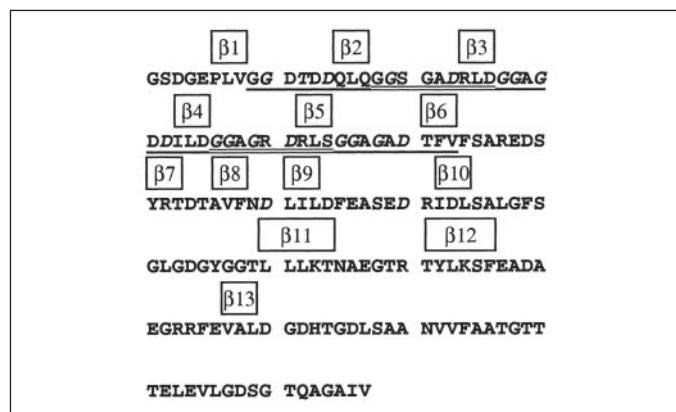


FIGURE 1. The amino acid sequence of the R-module with the secondary structure elements indicated. The individual RTX repeats are indicated by alternating single and double underlining. *Italic letters* indicate amino acids involved in coordination of calcium ions.

forming the next loop, resulting in an 18-amino-acid β -roll unit. This is reflected by 19.5% of all long range NOEs being observed between atoms that are 18 residues apart in the sequence. This structure is repeated three times (β_1 6–8, β_2 15–17, β_3 24–26, β_4 33–35, β_5 42–44, β_6 51–53) making up three turns of the β -roll (Figs. 1 and 3). Thereafter, a loop-out forms an anti-parallel β -hairpin (β_7 61–63, β_8 66–68). The polypeptide chain folds back and makes a new turn elongating the β -roll (β_9 71–73, β_{10} 81–83). Thereafter, a longer loop bulges out, followed by a less well defined region (91–100). The structure ends with two anti-parallel β -strands (β_{11} 100–104, β_{12} 111–114). The last β -strand (β_{13} 127–129) is situated in between the roll and the anti-parallel β -sheet (β_{11} – β_{12}), parallel to the first and anti-parallel to the latter. Finally, the polypeptide again forms a somewhat less well ordered loop structure from residue 130 onward. From amino acid residue 145 to the C termi-

nus (residue 167), no ordered secondary structure exists. This is in agreement with heteronuclear NOE measurements (data not shown) pointing at a more mobile polypeptide chain.

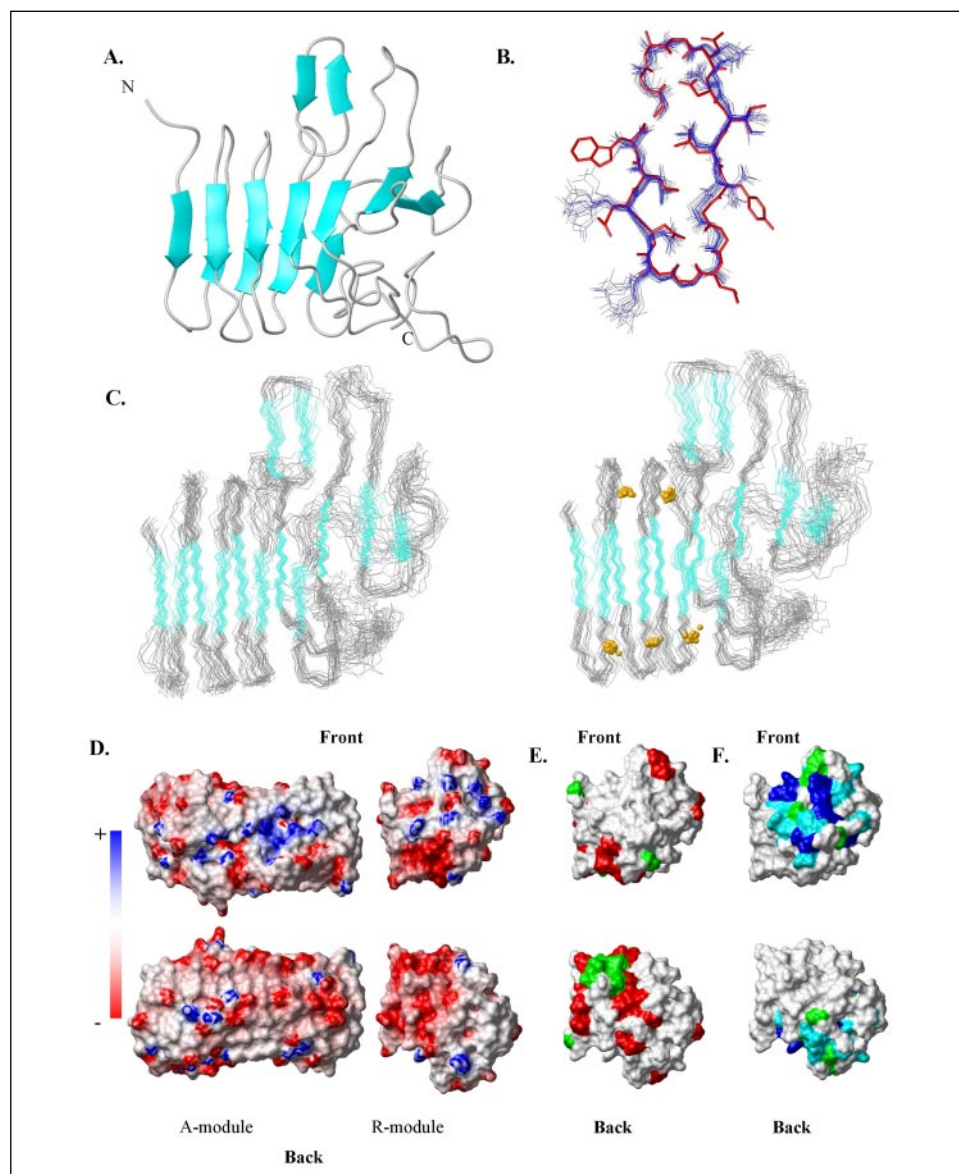
Remarkably, only 27% of the amino acids of the R-module are in regular secondary structure elements (17% in parallel and 10% in anti-parallel β -strands), whereas 73% of the residues form a coil structure. This is corroborated by circular dichroism measurements (results not shown). Altogether, the protein forms an elongated structure along the axis of the β -roll with a small groove at the front side. The electrostatic surface potential of the R-module shows a positively charged patch formed by arginine and lysine side chains along the small groove at the front side of the β -roll. Aspartic acid and glutamic acid residues form negatively charged patches along the turns of the β -roll (Fig. 2D, right).

To obtain further information about the intramolecular dynamics of the R-module, ^{15}N - ^1H -NOEs as well as T_1 and T_2 of ^{15}N were measured (data not shown). The ^{15}N - ^1H -NOEs show increased flexibility for the N-terminal residues 1–7 and a highly mobile chain from residue 145 onward. The same tendency is reflected in the ^{15}N T_1 and T_2 relaxation times. For the structured part of the R-module, the T_1 : T_2 ratio is a direct measure of the correlation time of the overall rotational tumbling of the molecule. The average T_1 : T_2 ratio for the protein was 12.1 ± 1.1 , which, assuming a spherical particle, corresponds to an overall rotational correlation time $\tau_c = 10.7 \pm 0.5$ ns (30).

The R-module has been shown to bind calcium ions (26). To obtain further information about calcium binding sites, it was attempted to prepare a metal-free apo-form of the R-module by adding EDTA. However, this resulted in immediate precipitation, so that no sample of apo-R-module could be obtained. The importance of Ca^{2+} ions for the stability of the R-module is also underlined by the fact that NMR samples of the R-module at 5 mM CaCl_2 showed visible signs of degradation in the NMR spectra after 48 h at room temperature, whereas the NMR samples at 50 mM CaCl_2 were stable for months. Thus, because an apo-R-module was beyond reach, a

FIGURE 2. **The NMR structure of the AlgE4 R-module.**

A, ribbon drawing of the energy-refined R-module structure with the lowest residual CYANA target function, viewed from the front and side of the β -roll. The secondary structure elements are parallel strands β_1 (6–8), β_2 (15–17), β_3 (24–26), β_4 (33–35), β_5 (42–44), β_6 (51–53), β_9 (71–73), β_{10} (81–83), and β_{13} (125–129), and the anti-parallel strands β_7 (61–63), β_8 (66–68), β_{11} (100–104), and β_{12} (111–114) all shown in cyan. **B**, structure of heavy atoms in the loop region of residues 27–44 for the 20 best energy-refined conformations of the R-module (*thin lines*) and residues 351–368 of *P. aeruginosa* (*thick lines*). **C**, bundle of the 20 energy-refined conformations of the R-module in the region of amino acid residues 5–145. Shown are the R-module without calcium ions (*left*) and the R-module with calcium ions incorporated in the loops (*right*). β -sheet regions are colored in cyan and calcium ions are depicted as orange spheres. The coordinate files of the calcium-free form were deposited in the PDB data base with accession code 2AGM. **D**, electrostatic surface potential of the AlgE4 A- (*left*) (28) and R-modules (*right*). Red and blue, respectively, denote regions of negative and positive potential on the protein surface. **E**, Shown is the interaction with Tm^{3+} . Green patches denote amino acids, with $\Delta_{\text{HSQC}} > 20$ Hz at an R-module:calcium:thulium ratio of 1:25:3.75, and red patches denote amino acids, with $\Delta_{\text{HSQC}} > 100$ Hz at an R-module:calcium:thulium ratio of 1:25:25.75. **F**, interaction with pentameric M_5 alginate. Blue surface patches indicate amino acids, with $\Delta_{\text{HSQC}} > 60$ Hz, cyan patches indicate amino acids, with $30 \text{ Hz} < \Delta_{\text{HSQC}} < 60$ Hz. Green surface patches indicate amino acids that change significantly upon titration with M_5 but only where this change could not be quantified throughout the titration because of signal overlap. Δ_{HSQC} values were calculated according to Equation 1. Structure models were prepared with the program MOLMOL (60).



titration was performed substituting Ca^{2+} ions with a paramagnetic ion. We chose the trivalent lanthanide ion, thulium, which has a very similar ionic radius as calcium and therefore can be expected to cause only minor structural changes (41, 42). Titration with thulium resulted in chemical shift changes and vanishing signals. The results are summarized in Figs. 2E and 3. The chemical shift changes observed during thulium titration are more pronounced for the amide protons than for the backbone nitrogen atoms, which argues for minimal structure changes, because ^{15}N shifts are more sensitive to small structural changes (43). These changes are mainly associated with amino acid residues situated in the loops of the β -roll and in the vicinity of negative charges on the surface of the protein. Thulium ions also bind to the negatively charged residues in the C-terminal disordered coil region of the R-module. Attempts to achieve a total replacement of calcium with thulium were performed by first unfolding the R-module in urea, followed by refolding in a buffer containing thulium ions instead of calcium. This resulted in precipitation regardless of whether urea was diluted slowly or quickly out of the sample (results not shown). Also, during titration of the R-module with thulium, the protein precipitated at a thulium concentration of ~ 6 mM. Again, the presence of calcium proved to be vital for maintaining the structural stability of the R-module.

Titration of the alginate pentamer M_5 to the R-module resulted in specific changes in the spectra. The chemical shift changes occurring upon titration with M_5 are plotted against the sequence in Fig. 4, whereas Fig. 2F shows the location of the amino acids with changing chemical shifts on the surface of the R-module. The binding event must be rather strong, because chemical shift changes essentially stopped after having reached a 1:1 molar ratio of protein and M_5 . Not surprisingly, chemical shift changes upon titration with M_5 occurred in the vicinity of the patch of positively charged amino acids on the surface of the R-module.

DISCUSSION

Here we have described a study of the R-module, which is one of two types of structural modules forming the family of seven secreted mannuronan C-5 epimerases in *A. vinelandii*. We have solved the NMR structure of the R-module from AlgE4 containing a C-terminal signal sequence. This structure opens a rational understanding of the secretion of AlgE4 and its function in relation to the A-module. Calcium is important both for the enzymatic activity of the epimerases and for gel formation in alginates. Therefore, the metal binding property has been investigated using the para-

Structure of the R-module

FIGURE 3. Observed chemical shift changes upon titration of the R-module with thulium.

Shown are Δ_{HSQC} values (calculated according to Equation 1) at R-module:calcium:thulium ratios of 1:25:3.75 (top) and 1:25:25.75 (bottom) plotted as a function of the amino acid sequence. Amino acids for which shifts could not be observed after thulium titration are as follows: 2, 4, 5, 7, 9, 12, 13, 14, 17, 30, 31, 34, 35, 47, 48, 51, 55, 58, 59, 74, 76, 78, 79, 80, 86, 94, 120, 121, 134, 153, 154, 156, 157, 159, 160, and 167. Underlined amino acid numbers indicate residues in a turn or coil structure. Doubly underlined amino acid numbers indicate residues in one of the loops of the repeat region. Secondary structure elements in the polypeptide chain are plotted along the top of the graph.

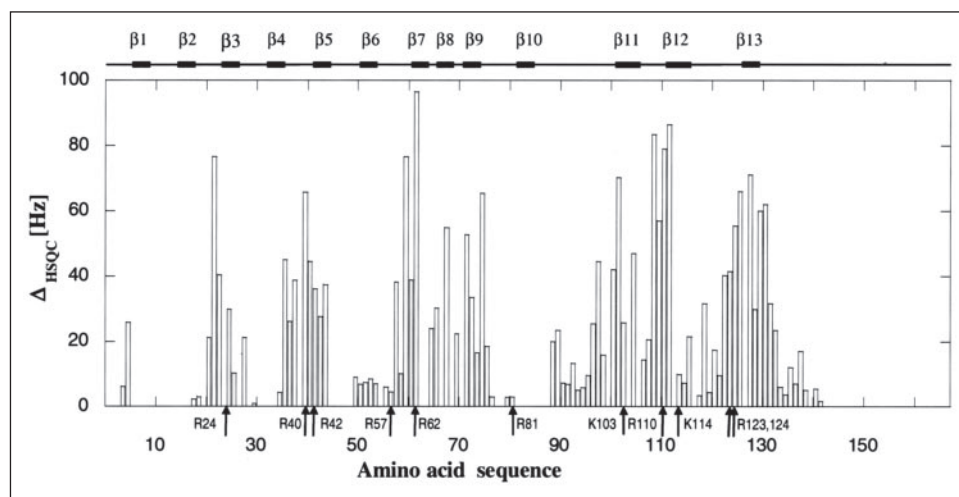
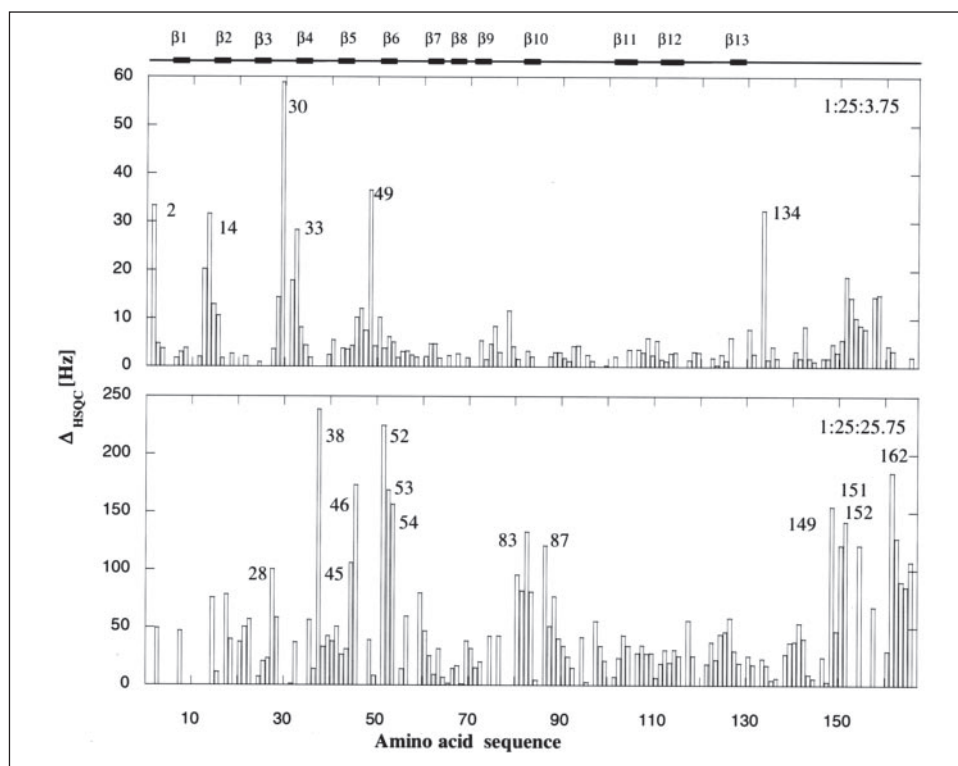


FIGURE 4. Observed chemical shift changes upon titration of the R-module with pentameric M_5 alginate. Shown are Δ_{HSQC} values (calculated according to Equation 1) between R-module in the absence of alginate and at an R-module: M_5 ratio of 1:1.2 plotted versus the amino acid sequence. Secondary structure elements in the polypeptide chain are plotted along the top of the graph. In the bottom, the position of positively charged side chains is indicated.

magnetic cation Tm^{3+} to observe possible calcium binding sites in the protein. The structure of the R-module belongs to the class of all- β proteins and shows a rare, right-handed parallel β -roll. Similar structures have only been found in the C-terminal part of the metalloproteases from *P. aeruginosa* psychrophilic *P. aeruginosa*, *S. marcescens*, and *Serratia sp E-15* (SMP_E-15) (PDB codes 1KAP, 1G9K, 1SAT, and 1SRP) (38, 39, 44, 45). The N-terminal β -roll of the R-modules contains a characteristic motif of nonapeptide repeats with the consensus sequence LXGGAGXD X_n (24), which is nothing but a cyclic permutation of the consensus sequence GGXGXDX X_n (U = a large hydrophobic amino acid) that Baumann *et al.* (39) found in the β -roll of the metalloproteases. This nonapeptide repeat motif was originally found in a number of toxins from different Gram-negative bacteria. It was denoted RTX (repeats in toxin) (46, 47). Later, it had also been found in other non-toxin related proteins, such as metalloproteases, iron-regulating proteins, and lipases, including L3 lipases (48,

49). A common feature for these RTX motif-containing domains is that they are located after the N-terminal catalytic domain and that they have a C-terminal secretion signal for translocation through a type I secretion system or ATP binding cassette transporter (48, 50, 51). Mapping the consensus sequence of the β -roll onto the structure, the XUX motif forms the β -strand. The large hydrophobic amino acids (U) are buried in the hydrophobic core of the β -roll, where they manifest themselves by numerous NOESY cross-peaks between the protons of these side chains. The large hydrophobic amino acid is leucine in the first five β -strands and phenylalanine in the sixth β -strand (52). Thereafter follows Phe-54, which breaks up the tightly packed β -roll. Subsequently, only one more turn is added to the β -roll. This is also observed in the x-ray structures of the metalloproteases. The introduction of β -roll-breaking aromatic side chains can be a way to control or stop the elongation of the β -roll.

The GGXGXD motif forms the loop between the two β -strands in the

two-layered β -sandwich. The loops have an interesting pattern of NOEs between H^N and $H^{\alpha 2,3}$ of G_1 to H^δ or H^ϵ of U_8 and from H^β of D_6 to H^δ or H^ϵ of U_8 . The side chain of D_6 is buried in the β -roll, and the backbone carboxyl oxygen atoms from G_2 , G_4 , G_1' and X_3' are oriented toward the interior of the loop. Such buried negative charges in a protein are energetically highly unfavorable. The electron-rich environment as well as the presence of a cavity between the loops suggests a metal binding site, as a metal binding would neutralize the unfavorable unpaired charge in the hydrophobic core. Thulium titration data show calcium replacement in the loops between residues 8–14, 26–32, 45–50, and residue 80, whereas indication of a weaker interaction was observed for the loops 18–23 and 36–41 and residue 69. The calculated structure of a loop region of the R-module was overlaid with a calcium binding loop from the β -roll in metalloprotease *P. aeruginosa* (Fig. 2B). The high structural identity of the loop regions suggests that the R-module binds calcium in essentially the same way as the metalloproteases *P. aeruginosa* and *S. marcescens*. The incorporation of calcium ions into the structure calculation of the R-module did only result in a minor increase of the target function (Fig. 2D; Table 1). Even though this is no direct evidence for the calcium binding mode of the R-module, it shows that a calcium binding mode equal to *P. aeruginosa* is fully compatible with our experimental data. Exchange of calcium with thulium proved to happen more readily compared with the corresponding substitution in the metalloproteases, where it could not be achieved. Only the two most C-terminal calcium ions at the end of the β -roll in the metalloproteases could be replaced. This was ascribed to the fact that water molecules substitute for some of the protein ligands in the coordination of these ions (39). Neither could the calcium ions be exchanged from a synthetic β -roll (39, 45, 52). This dissimilarity could be due to the R-module being a truncated protein. The lack of the N-terminal part compared with the metalloproteases could make the β -roll more flexible. The negatively charged patch at one side of the R-module and a similar negatively charged patch on one side of the A-module could function as cation storage (e.g. calcium, barium, strontium) for the AlgE4 epimerase to either conserve the protein integrity and/or facilitate gel formation after epimerization of the alginate, to drive the reaction (53, 54).

As already mentioned earlier, the RTX protein family and the R-module contain a C-terminal secretion signal for translocation through ATP binding cassette transporters, where the proteins are secreted across the inner and outer cell membranes into the medium of Gram-negative bacteria. This secretion signal is not cleaved from the mature protein. In the literature (48, 55, 56), there are several suggestions as to what triggers translocation. For some RTX proteins, it has been suggested that the signal for secretion is located within the 60 amino acids at the C terminus, but no specific sequence responsible for secretion has been identified. Hui *et al.* (55) suggest that the recognition motif could be a sequence of secondary structure elements consisting of an amphiphilic α -helix followed by a charge-rich linker (8–10 residues) and another amphiphilic α -helix-positioned \sim 45 amino acid residues from the C terminus. Such structure elements are not found in the R-module and there is no indication of other secondary structure elements in the C-terminal part of the protein. It has even been suggested that a general signal for translocation does not exist for these RTX proteins (48). In the R-module of AlgE4, the 22 C-terminal amino acid residues 145–167 have a disordered conformation in solution. In the AlgE epimerase family, there are some common features for the C termini. First, there are short stretches of hydrophobic amino acids from residues 142–144, 155–156 (numbering according to the R-module of AlgE4), and at the very C-terminal end (25). Second, there are no positively charged amino acids found in the last 31 amino acids of the random coil C-terminal tail. This indicates the possibility that negative charges or a cation binding seg-

ment could trigger translocation through the ATP binding cassette transporter.

The function of the RTX domain in relation to the N-terminal catalytic domain has been proposed to facilitate translocation of larger proteins and to ensure structural integrity after translocation (48, 57, 58). This does not explain why AlgE2, AlgE5, AlgE6, and AlgE7 have three to four R-modules. Common for them all is that only the C-terminal R-module contains the secretion signal.

No experimental data are yet available on the structure of the whole AlgE4 protein consisting of one A-module followed by one R-module. Both the A- and the R-module exhibit characteristic, positively charged patches on their surfaces. This similarity, seen in the light of the nature of the protein's substrate, a negatively charged polymer, suggests that the whole AlgE4 protein might form a single elongated structure with a long, positively charged patch for substrate binding (Fig. 2D). Also from the point of view of the secondary structure elements, the R-module appears as a natural extension of the A-module. The C terminus of the A-module contains some prolines that might be the reason for the observed narrowing of a four-stranded to a three-stranded β -helix. Because 6 of a total of 12 prolines are found in the region between the A- and the R-module, one may speculate that the function of the prolines is to force the β -helix into the β -roll structure and thus form a transition from the A- to the R-module. Altogether, AlgE4 seems to consist of a peculiar, right-handed parallel β -helix structure that is narrowing down to a β -roll, a fold not previously reported in literature.

We have seen strong interactions between the R-module and an alginate M_5 -pentamer. This is in contrast to results achieved previously by atomic force microscopy, which showed that the A-module of AlgE4 binds to alginate stronger than the full-length protein, whereas the R-module was not observed to interact directly with the polymer (59). Our NMR data point at the interaction taking place in the vicinity of a cluster of positively charged side chains on the surface of the protein. The significance of these side chains for alginate binding is also underlined by the fact, that some of these positively charged amino acids, Arg-24, Arg-40, Arg-110, and Lys-114, are highly conserved throughout the 25 R-modules of *A. vinelandii*.

As mentioned, the activity of the A-module is stimulated by the R-module. This might possibly be linked to the alginate binding property of the R-module.

In the absence of a structure of the complete AlgE4 protein, the structures of the R- and A-modules provide vital information on the architecture of this protein. Similar to the A-module, the R-module folds as a parallel right-handed β -roll, and it is the first structure solved with NMR spectroscopy having this rare fold. It was demonstrated that calcium binding is crucial for the structural stability of the R-module. Thulium titration of the R-module has revealed the location of metal binding sites. Metal ions are bound in the loops homologous to other RTX proteins. Calcium binding can be important for epimerization, stabilization, and/or gel formation of alginate. Alginate titration showed specific and strong interactions between the R-module and a pentameric M_5 alginate. Knowledge about the structure and alginate interaction of the R-module provides the basis for better insight into the structure and catalytic activity of the *A. vinelandii* AlgE1–7 alginate epimerase family.

Acknowledgments—We are thankful to Lee O'Brien Andersen, Jakob Christiansen, Thomas M. Dantoft, Thomas Marius Nielsen, and Niels-Kristian Østergaard for performing the titration work with thulium. We thank Prof. Gudmund Skjåk-Bræk for a gift of pentameric M_5 alginate. The NMR laboratory at Aalborg University received financial support from the Obel Foundation.

REFERENCES

1. Fischer, F. G., and Dorfel, H. (1955) *Hoppe Seyler's Z. Physiol. Chem.* **302**, 186–203
2. Hirst, E. L., Percival, E., and Wold, J. K. (1964) *J. Chem. Soc.*, 1493–1499
3. Donati, L., Holtan, S., Morch, Y. A., Borgogna, M., Dentini, M., and Skjåk-Bræk, G. (2005) *Biomacromolecules* **6**, 1031–1040
4. Grant, G. T., Morris, E. R., Rees, D. A., Smith, P. J. C., and Thorn, D. (1973) *FEBS Lett.* **32**, 195–198
5. Stokke, B. T., Smidsrød, O., Bruheim, P., and Skjåk-Bræk, G. (1991) *Macromolecules* **24**, 6637–6645
6. Haug, A., Larsen, B., and Smidsrød, O. (1974) *Carbohydr. Res.* **32**, 217–225
7. Hellebust, J. A., and Haug, A. (1972) *Can. J. Bot.* **50**, 177–184
8. Smidsrød, O., and Draget, K. I. (1966) *Carbohydr. Eur.* **14**, 6–13
9. Skjåk-Bræk, G., Grasdalen, H., and Larsen, B. (1986) *Carbohydr. Res.* **154**, 239–250
10. Sherbrock-Cox, V., Russell, N. J., and Gacesa, P. (1984) *Carbohydr. Res.* **135**, 147–154
11. Frei, E., and Preston, R. D. (1962) *Nature* **196**, 130–134
12. Gorin, P. A. J., and Spencer, J. F. T. (1966) *Can. J. Chem.* **44**, 993–998
13. Cote, G. L., and Krull, L. H. (1988) *Carbohydr. Res.* **181**, 143–152
14. Linker, A., and Jones, R. S. (1964) *Nature* **204**, 187–188
15. Govan, J. R., Fyfe, J. A., and Jarman, T. R. (1981) *J. Gen. Microbiol.* **125**, 217–220
16. Fett, W. F., Osman, S. F., Fishman, M. L., and Siebles, T. S. (1986) *Appl. Environ. Microbiol.* **52**, 466–473
17. Sadoff, H. L. (1975) *Bacteriol. Rev.* **39**, 516–539
18. Hoidal, H. K., Svanem, B. I. G., Gimmestad, M., and Valla, S. (2000) *Environ. Microbiol.* **2**, 27–38
19. Valla, S., Li, J., Ertesvåg, H., Barbeyron, T., and Lindahl, U. (2001) *Biochimie (Paris)* **83**, 819–830
20. Rehm, B. H., Ertesvåg, H., and Valla, S. (1996) *J. Bacteriol.* **178**, 5884–5889
21. Douthit, S. A., Dlakic, M., Ohman, D. E., and Franklin, M. J. (2005) *J. Bacteriol.* **187**, 4573–4583
22. Gimmestad, M., Sletta, H., Ertesvåg, H., Bakkevig, K., Jain, S., Suh, S. J., Skjåk-Bræk, G., Ellingsen, T. E., Ohman, D. E., and Valla, S. (2003) *J. Bacteriol.* **185**, 3515–3523
23. Ertesvåg, H., Hoidal, H. K., Schjerven, H., Svanem, B. I. G., and Valla, S. (1999) *Metab. Eng.* **1**, 262–269
24. Ertesvåg, H., Doseth, B., Larsen, B., Skjåk-Bræk, G., and Valla, S. (1994) *J. Bacteriol.* **176**, 2846–2853
25. Ertesvåg, H., Hoidal, H. K., Hals, I. K., Rian, A., Doseth, B., and Valla, S. (1995) *Mol. Microbiol.* **16**, 719–731
26. Ertesvåg, H., and Valla, S. (1999) *J. Bacteriol.* **181**, 3033–3038
27. Bjerkan, T. M., Bender, C. L., Ertesvåg, H., Drablos, F., Fakhr, M. K., Preston, L. A., Skjåk-Bræk, G., and Valla, S. (2004) *J. Biol. Chem.* **279**, 28920–28929
28. Bjerkan, T. M. (2003) *Structure-Function Analyses of Mannuronan C-5 Epimerases*. Ph.D. thesis at the Department of Biotechnology, Norwegian University of Science and Technology, Trondheim, Norway
29. Aachmann, F. L., Svanem, B. I. G., Valla, S., Petersen, S. B., and Wimmer, R. (2005) *J. Biomol. NMR* **31**, 259
30. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry* **28**, 8972–8979
31. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33**, 5984–6003
32. Bartels, C., Xia, T.-H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) *J. Biomol. NMR* **5**, 1–10
33. Güntert, P., Qian, Y. Q., Otting, G., Müller, M., Gehring, W., and Wüthrich, K. (1991) *J. Mol. Biol.* **217**, 531–540
34. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13**, 289–302
35. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol.* **273**, 283–298
36. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc.* **117**, 5179–5197
37. Koradi, R., Billeter, M., and Güntert, P. (2000) *Comput. Phys. Commun.* **124**, 139–147
38. Baumann, U. (1994) *J. Mol. Biol.* **242**, 244–251
39. Baumann, U., Wu, S., Flaherty, K. M., and McKay, D. B. (1993) *EMBO J.* **12**, 3357–3364
40. Shannon, R. D. (1976) *Acta Crystallogr. Sect. A* **32**, 751–767
41. Bertini, I., Janik, M. B., Lee, Y. M., Luchinat, C., and Rosato, A. (2001) *J. Am. Chem. Soc.* **123**, 4181–4188
42. Mustafi, S. M., Mukherjee, S., Chary, K. V., Del Bianco, C., and Luchinat, C. (2004) *Biochemistry* **43**, 9320–9331
43. Walling, A. E., Pargas, R. E., and deDios, A. C. (1997) *J. Phys. Chem.* **A101**, 7299–7303
44. Aghajari, N., Van Petegem, F., Villeret, V., Chessa, J. P., Gerday, C., Haser, R., and Van Beeumen, J. (2003) *Proteins* **50**, 636–647
45. Hamada, K., Hata, Y., Katsuya, Y., Hiramatsu, H., Fujiwara, T., and Katsube, Y. (1996) *J. Biochem. (Tokyo)* **119**, 844–851
46. Felmler, T., Pellett, S., Lee, E. Y., and Welch, R. A. (1985) *J. Bacteriol.* **163**, 88–93
47. Felmler, T., and Welch, R. A. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5269–5273
48. Welch, R. A. (2001) *Curr. Top. Microbiol. Immunol.* **257**, 85–111
49. Amada, K., Haruki, M., Imanaka, T., Morikawa, M., and Kanaya, S. (2000) *Biochim. Biophys. Acta* **1478**, 201–210
50. Young, J., and Holland, I. B. (1999) *Biochim. Biophys. Acta* **1461**, 177–200
51. Omori, K., and Idei, A. (2003) *J. Biosci. Bioeng.* **95**, 1–12
52. Lilie, H., Haehnel, W., Rudolph, R., and Baumann, U. (2000) *FEBS Lett.* **470**, 173–177
53. Haug, A., and Larsen, B. (1971) *Carbohydr. Res.* **17**, 297–308
54. Larsen, B., and Haug, A. (1971) *Carbohydr. Res.* **20**, 225–232
55. Hui, D., Morden, C., Zhang, F., and Ling, V. (2000) *J. Biol. Chem.* **275**, 2713–2720
56. Duong, F., Lazdunski, A., and Murgier, M. (1996) *Mol. Microbiol.* **21**, 459–470
57. Letoffe, S., and Wandersman, C. (1992) *J. Bacteriol.* **174**, 4920–4927
58. Kwon, H. J., Haruki, M., Morikawa, M., Omori, K., and Kanaya, S. (2002) *J. Biosci. Bioeng.* **93**, 157–164
59. Sletmoen, M., Skjåk-Bræk, G., and Stokke, B. T. (2004) *Biomacromolecules* **5**, 1288–1295
60. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.* **14**, 51–55