Use of protein trans-splicing to produce active and segmentally ^{2}H, ^{15}N labeled mannuronan C5-epimerase AlgE4

Edith Buchinger,[1,2,3] Finn L. Aachmann,[2] A. Sesilja Aranko,[3] Svein Valla,[2] Gudmund Skjåk-BræK,[2] Hideo Iwaï,[3] and Reinhard Wimmer[1]*

[1]Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Aalborg DK-9000, Denmark
[2]Department of Biotechnology, NOBIPOL, Norwegian University of Science and Technology, Trondheim 7491, Norway
[3]Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, Helsinki FIN-00014, Finland

Received 12 April 2010; Revised 29 May 2010; Accepted 2 June 2010
DOI: 10.1002/pro.432
Published online 15 June 2010 proteinscience.org

Abstract: Alginate epimerases are large multidomain proteins capable of epimerising C5 on β-D-mannuronic acid (M) turning it into α-L-guluronic acid (G) in a polymeric alginate. *Azotobacter vinelandii* secretes a family of seven epimerases, each of which is capable of producing alginates with characteristic G distribution patterns. All seven epimerases consist of two types of modules, denoted A and R, in varying numbers. Attempts to study these enzymes with solution-state NMR are hampered by their size—the smallest epimerase, AlgE4, consisting of one A- and one R-module, is 58 kDa, resulting in heavy signal overlap impairing the interpretation of NMR spectra. Thus we obtained segmentally ^{2}H, ^{15}N labeled AlgE4 isotopomeres (A-[^{2}H, ^{15}N]-R and [^{2}H, ^{15}N]-A-R) by protein *trans*-splicing using the naturally split intein of *Nostoc punctiforme*. The NMR spectra of native AlgE4 and the ligated versions coincide well proving the conservation of protein structure. The activity of the ligated AlgE4 was verified by two different enzyme activity assays, demonstrating that ligated AlgE4 displays the same catalytic activity as wild-type AlgE4.

Keywords: trans-splicing; inteins; protein ligation; alginate epimerases

Introduction

Production of isotopically enriched proteins that are suitable for structural and functional studies by nuclear magnetic resonance (NMR) is still a major limiting step. Structural studies of biomolecules by NMR have made tremendous progress mainly due to improved recombinant protein expression and ^{13}C, ^{15}N labeling, and/or deuteration. New ways of producing labeled proteins and nucleotides have stimulated the development

Abbreviations: DTT, Dithiothreitol; EDTA, Ethylenediaminetetraacetic acid; EPL, Expressed Protein Ligation; FPLC, Fast Protein Liquid Chromatography; G, α-L- guluronic acid; GdmCl, Guanidium chloride; GSH, reduced Glutathione; HEPES, N-2-Hydroxyethylpiperazine-N′-2-ethanesulfonic acid; HSQC Heteronuclear Single Quantum Coherence; IMAC, Immobilized Metal Affinity Chromatography; Int_C, C-terminal part of intein; Int_N, N-terminal part of intein; IPTG, Isopropyl β-D-1-thiogalactopyranoside; M, β-D-mannuronic acid; ME, Mercaptoethanol; NCL, Native Chemical Ligation; NMR, Nuclear Magnetic Resonance; *Npu*, *Nostoc punctiforme;* PTS, Protein *Trans*-Splicing; SAIL, Stereo-Array Isotope Labelling; SDS-PAGE, Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis; TCEP, *Tris*(2-carboxyethyl)phosphine; TRIS, *Tris*-(hydroxymethyl)aminomethane; TROSY, Transverse Relaxation Optimized Spectroscopy; *Amp*®, Ampilicin resistance gene; *Kan*®, Kanamycin resistance gene.

Additional Supporting Information may be found in the online version of this article

Grant sponsors: Academy of Finland; Grant numbers: 118385 and 131413; Grant sponsor: Danish Agency for Science, Technology and Innovation, KMB project number 182695/I40 and the Norwegian Research Council; The NMR laboratory at Aalborg University is supported by the Obel and SparNord foundations.

**Correspondence to:* Reinhard Wimmer, Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Aalborg DK-9000, Denmark. E-mail: rw@bio.aau.dk

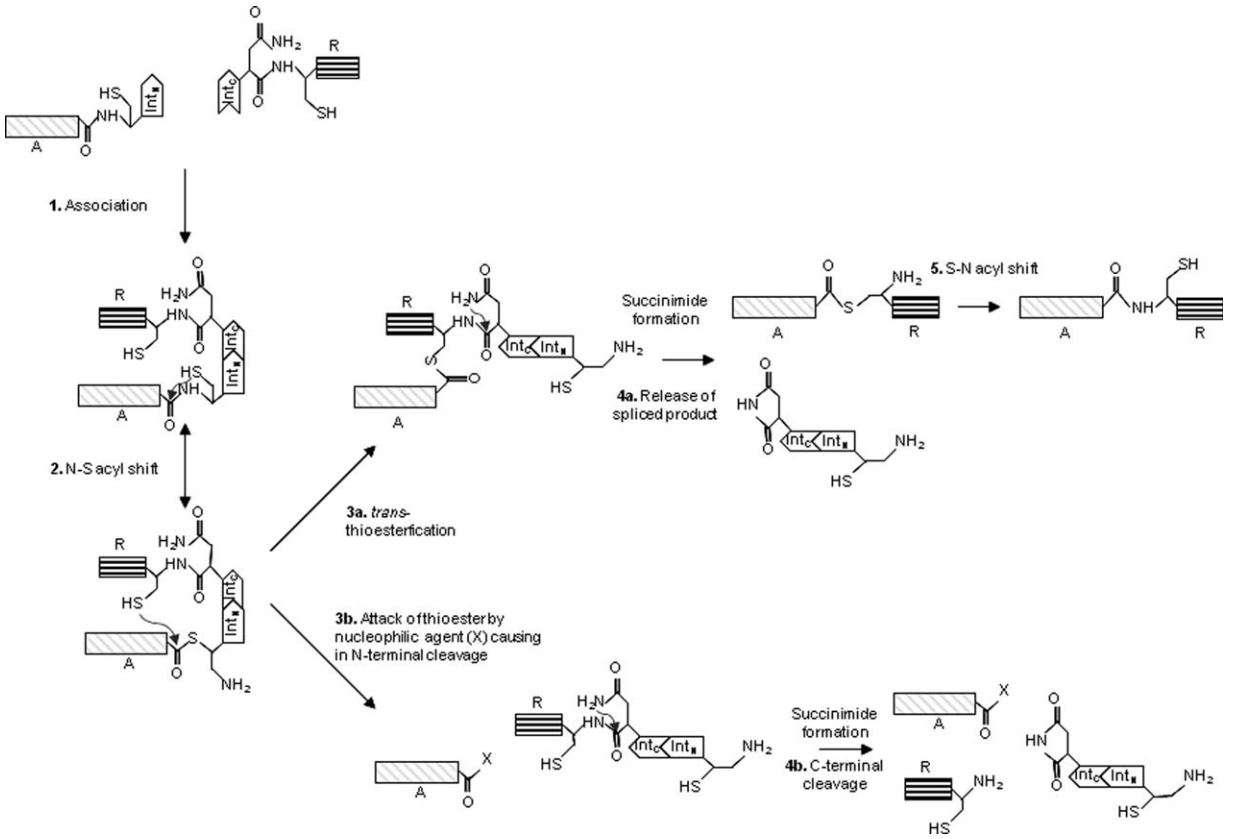

Figure 1. Mechanism of the trans-splicing and trans-cleavage reactions. The reaction steps are: 1. Association of the two intein domains, 2. Attack by Cys1 of intein results in a reactive thioester, 3. either (a) attack of the N-terminal thioester by the first cysteine residue in the C-terminal extein to the intein yields a branched thioester or (b) attack of the thioester by nucleophilic reagents—also water (X) yielding in N-terminal cleavage, 4. Cyclization of the C-terminal asparagine residue results in (a) spliced product or (b) C-terminal cleavage, 5. *S-N* acyl rearrangement restores a native peptide bond.

of novel NMR experiments.[1,2] It has also expanded the scope of NMR with biological macromolecules such as larger proteins by new labeling technology. Whereas NMR structure determination of globular domains of below 20 kDa has increasingly become a routine procedure, resonance assignment of larger proteins can be very time-consuming. A variety of labeling methods such as methyl labeling,[3,4] selective isotopic labelling[5] and stereo-array isotope labeling (SAIL)[6] have been developed and exploited for NMR studies of larger proteins. One of the potentially powerful labeling methods is segmental isotopic labeling, where a segment or a domain in a protein are selectively labeled with stable isotopes.[7,8] Isotopic labeling of a segment or domain in a large protein not only simplifies the spectral complexity but also allows investigation of a region of interest in an intact protein with conventional triple-resonance NMR experiments.[9] Segmental isotopic labeling has been achieved by native chemical ligation (NCL),[10] expressed protein ligation (EPL),[11,12] protein *Trans*-splicing (PTS)[7,9] and enzymatic ligation by sortase and V8 protease.[13,14] One frequently used reaction is called native chemical ligation where an N-terminal cysteine reacts with C-terminal thioester via *trans*-thioesterification and *S-N* acyl shift to a peptide bond. A family of proteins called inteins perform a similar reaction. Inteins are intervening peptide sequences, which excise themselves post-translationally and ligate two flanking N-and C-terminal segments (exteins) via a peptide bond.[15–20] In EPL the intein variant is cleaved by thiol reagents resulting in an extein with a C-terminal thioester. Together with the other extein with an N-terminal cysteine, the two fragments then ligate like in NCL. PTS is an intein mediated method, where the two parts of the split intein associate and perform the peptide ligation (Fig. 1). PTS is a very robust method for ligation and in the last few years *in vivo* and *in vitro* ligation of two segments has been established[21] and recently also three fragments could be ligated.[22] Here, we describe an application of segmental isotopic labeling by protein *trans*-splicing to an enzyme, alginate C5-epimerase AlgE4. This enzyme belongs to a family of seven secreted, structurally related, Ca^{2+} dependent mannuronan C5-epimerases (AlgE1–7) produced by *Azotobacter vinelandii*.[23] These epimerases catalyse epimerisation around C-5 of β-D-mannuronic acid (M) to α-L-guluronic acid (G) at the polymer level in the alginate

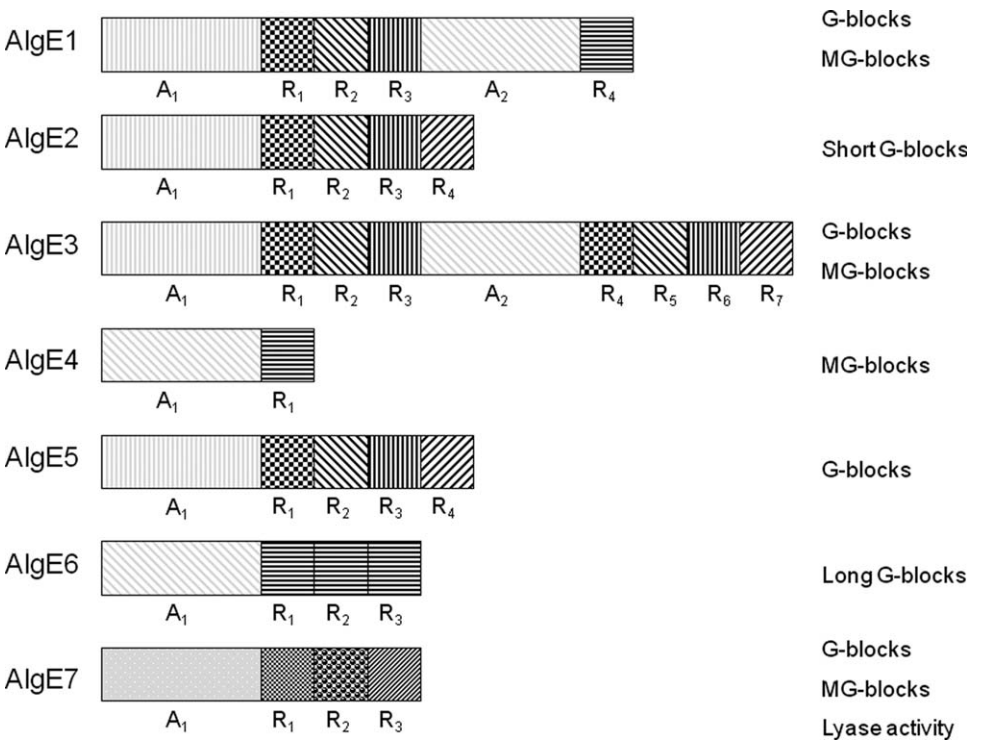

Figure 2. *Azotobacter vinelandii* expresses a family of extracellular alginate epimerases which only consists of two different modules named A- and R-module. Only the A-modules are catalytically active but the R-modules enhance the activity significantly if bound to an A-module. The different members of the family show different epimerisation products, given on the right side of the figure, AlgE7 acts also as a lyase.[23]

polysaccharide. Each AlgE epimerase produces a unique epimerisation pattern of M and G subunits, and AlgE4 produces an alternating structure (MG-blocks).[24]

AlgE1–7 consist of a unique combination of two different modules, designated A and R, and a C-terminal peptide presumably involved in translocation of the protein. The A- and R-modules consist of ~385 and 155 amino acids, respectively. In all the AlgE enzymes there is an A-module located N-terminally, and AlgE1 and AlgE3 have one additional A-module in their sequences. The number of R-modules varies from 1 to 7[25] (Fig. 2). All A-modules and all R-modules share extensive sequence similarities, indicating that the *algE* gene family was generated by a series of gene duplication events.[26,27]

AlgE4 is the smallest of the alginate epimerases and has the composition A-R. The structures of the AlgE4 A- and R-module have been solved separately by X-ray crystallography and NMR, respectively.[28,29] Both proteins showed a highly unusual structure consisting mainly of parallel β-sheets making up a four stranded β-helix and a two stranded β-roll, respectively. The A-modules are catalytically active on their own.[30] The R-modules do not posses any catalytic activity but strongly enhance the reaction rate if at least one R-module is linked to an A-module. Furthermore, the R-module was also shown to interact with alginate oligomers.[29] The strength of the interaction depends on the relative content of M and G.

The aim of this study is to simplify further investigations into the overall structure and substrate binding of active AlgE4 by using segmentally labeled AlgE4, both A-[^{2}H, ^{15}N]-R and [^{2}H, ^{15}N]-A-R. The size of AlgE4 (57.7 kDa) makes it necessary to fully deuterate the NMR observable domain. We have chosen to use the naturally split intein of *Nostoc punctiforme* (*N. punctiforme*) for PTS as it shows high tolerance of sequence variations at the splicing junctions, high splicing activity with foreign exteins, and a high solubility.[31,32]

Results

Cloning and production

To obtain a segmentally labeled AlgE4, the epimerase was divided into two parts within the intermediate region connecting the A- and the R-module. The A-module was cloned upstream of the N-terminal part of the naturally split intein *Npu*DnaE intein (Int_N), while the R-module was cloned between the C-terminal *Npu*DnaE intein (Int_C) and a His-tag for purification. The short amino acid sequence KCFNG around the splicing site was used to obtain optimal splicing. This gives rise to a slightly different amino acid sequence of the produced AlgE4. In addition, the C-terminal 20 residues were omitted, as they are

Figure 3. Part of the sequence alignment of native AlgE4 and the segmentally labeled AlgE4 construct. The last 20 amino acids (underlined) of the wild-type AlgE4 are known to be unstructured and were exchanged to a His-Tag for purification. The rectangle shows the splicing site, where few amino acids were exchange to obtain optimal ligation.

known to be unstructured, and a C-terminal His-tag was included. Figure 3 illustrates the differences between the sequences of wild-type and ligated AlgE4.

Expression studies showed that the A-Int_N is soluble if the *Escherichia coli* cells were induced at low temperature (15°C, over night). However, the expression of Int_C-R always resulted in insoluble fractions that had to be refolded from 6 *M* Guanidium chloride (GdmCl) before protein ligation. On average, 24 mg (0.45 μmol) of purified A-Int_N was obtained from 1 L growth medium. The yield of the purified Int_C-R was 7 mg (0.30 μmol).

Reducing agent/protein ligation

The refolded R-module turned out to be a dimer, where dimerisation occurred by the formation of a disulfide bridge between the single cysteine residues from two Int_C-R molecules. This could be seen from comparing SDS-PAGE gels run under reducing and nonreducing conditions (data not shown). As the presence of the free cysteine is required for the ligation to proceed (see Fig. 1), the dimeric Int_C-R had to be reduced before PTS. As previously reported, the choice of reducing agents can have a significant effect on the *trans*-splicing efficiency.[33] For the ligation tests, purified A-Int_N and Int_C-R were used. To facilitate the purification, a construct in which an N-terminal His-tag was attached to A-Int_N was used, while Int_C-R had a C-terminal His-tag. Both proteins were purified by Immobilized Metal Affinity Chromatography (IMAC). For the ligation, concentrations of both parts were adjusted to 0.01 m*M*. Four different thiol-agents were tested (DTT, mercaptoethanol, reduced Glutathione (GSH) and cysteamine), at different temperatures (room temperature, 30 and 37°C), at different final concentrations (2.5, 5, and 10 m*M*) and at two different pH values (7 and 8). For one set of experiments the buffer was exchanged to HEPES buffer. None of the experiments gave a satisfying result; in all cases ligation was very slow. The conditions yielding the highest levels of ligated product were: room temperature, pH 7 and 5 m*M* DTT, mercaptoethanol or cysteamine. For Glutathione as reducing agent, the optimal condition for ligation was room temperature, pH 8 and 5 m*M* concentration. Even in the best cases, the amount of cleaved product exceeded the amount of ligated AlgE4. (Fig. 4A,C). Higher temperature and pH 8 resulted mainly in cleavage of A-Int_N. Changing to HEPES buffer had no effect on the ligation, the addition of Ca^{2+}-EDTA completely blocked ligation. The ligation tests also showed that neither A-Int_N nor Int_C-R were stable in solution at room temperature. Cleavage of the N-terminal intein part has been described before,[34–36] the instability of Int_C-R was probably caused by traces of proteolytic enzymes, as the degradation of Int_C-R could be prevented by adding protease inhibitors.

While none of the thiol-based reducing agents yielded a satisfying level of ligation, the *Tris*(2-carboxyethyl)phosphine (TCEP) reducing agent gave high yields of *trans*-splicing without cleavage [Fig. 4(B,D)]. Thus, we continued the work using TCEP as reducing agent to initiate ligation.

In vitro *ligation of A-[^{15}N]-R*

Protein ligation was initiated by adding TCEP to a final concentration of 5 m*M*. *Trans*-splicing was very robust—the ligation could be performed fast in crude cell extract without preliminary purification [Fig. 4(B,D)]. Within one hour, ligation yields up to 90% were reached.

After 2 hours of ligation, the ligated product was purified by IMAC. The eluted fractions contain both the segmentally labeled AlgE4 (A-[^{15}N]-R) and the precursor [^{15}N]-Int_C-R, as both compounds carry

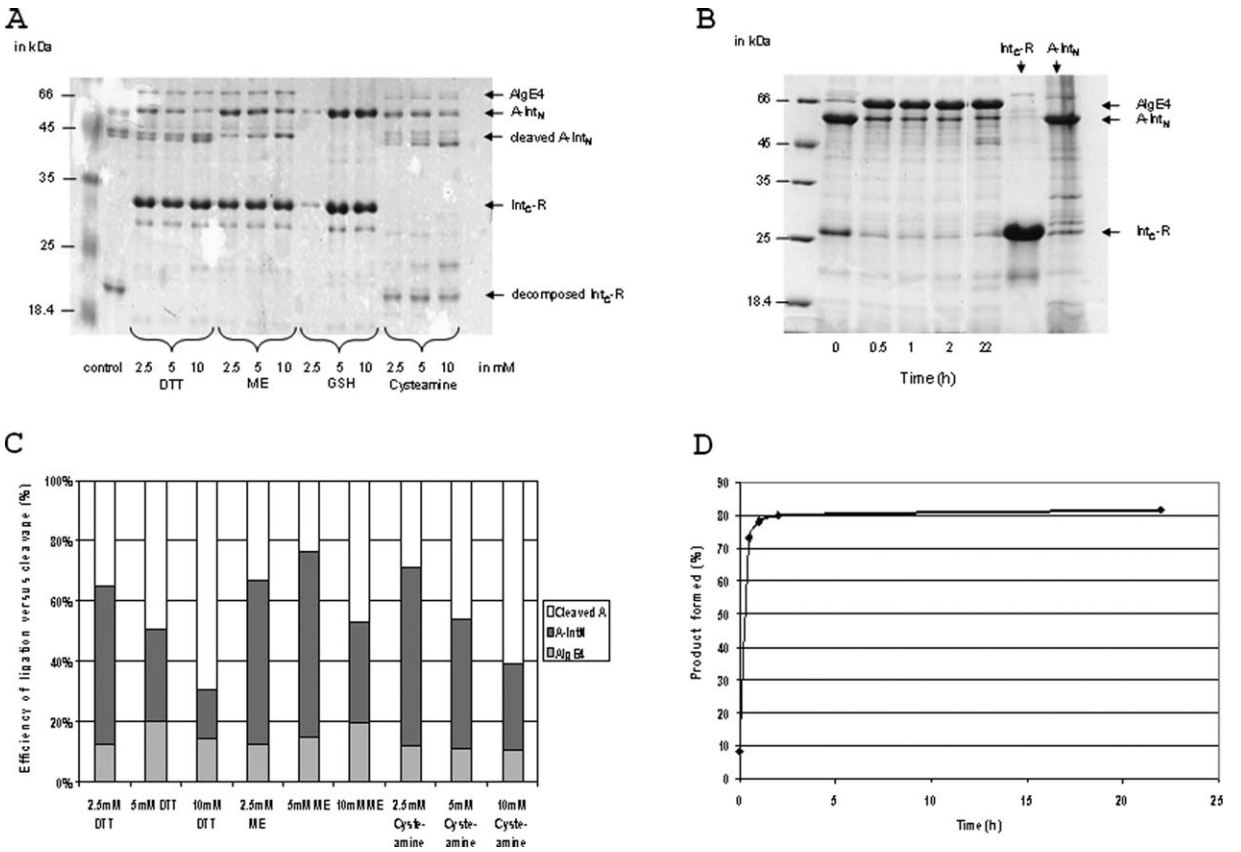

Figure 4. (A) Overview of the ligation with different reducing agents after 2 days ligation at room temperature in Tris buffer at pH 7 without protease inhibitor. A-Int$_N$ or Int$_C$-R was purified before ligation tests. The sample without any reducing agent showed that the A-Int$_N$ or Int$_C$-R is not stable without protease inhibitor. The reducing agents were Dithiothreitol (DTT), Mercaptoethanol (ME), Glutathione reduced (GSH) and Cysteamine. (B) SDS-PAGE of the ligation mixture containing A-Int$_N$ and Int$_C$-R. Both proteins were not purified before ligation although Int$_C$-R had to be refolded. To avoid degradation of Int$_C$-R, protease inhibitor was added and the reaction started after the addition of TCEP. Samples for SDS-PAGE were taken at the given time point. (C) Analysis of the gel lanes from panel A by ImageJ. The sum of A-module containing polypeptides (A-IntN, A (cleaved) and AlgE4) is normalized to 100%. Glutathione is inactive at pH 7 and was therefore omitted. (D) Formation (by ligation) of AlgE4 as a function of time. Analysis of the gel lanes from panel B by ImageJ.

the C-terminal His-tag. Therefore, the eluted fractions had to be further purified by gel-filtration. The amount of purified segmentally labeled A-[^{15}N]-R was ~15 mg (0.25 μmol) from A-Int$_N$ and Int$_C$-R produced in 1 L medium each.

In vitro *ligation of A-[^{2}H, ^{15}N]-R and [^{2}H, ^{15}N]-A-R*

The line width of the A-[^{15}N]-R was broader than that of the R-module alone. For NMR studies, segmentally labeled A-[^{15}N]-R is therefore only of limited use.

To obtain sharper signal of the R-module and due to the size of the A-module two different segmentally deuterated and [^{15}N]-labeled AlgE4 (A-[^{2}H, ^{15}N]-R and [^{2}H, ^{15}N]-A-R) were obtained. The protein ligation seemed to proceed slower with deuterated material, independent of which domain was labeled. Therefore the ligation reaction was allowed to continue for 16 h. The yield of the purified A-[^{2}H, ^{15}N]-R and [^{2}H, ^{15}N]-A-R was 3 mg (0.05 μmol) from A-Int$_N$ and Int$_C$-R produced in 1 L growth medium each. TROSY spectra of [^{2}H, ^{15}N]-A-R, A-[^{2}H, ^{15}N]-R and [^{2}H, ^{15}N]-AlgE4 were overlaid to confirm the correct fold of segmentally labeled AlgE4. Figure 5 shows the TROSY-NMR spectra of wild-type AlgE4 and the two segmentally labeled variants.

Line width

For NMR studies, it is very important to have narrow lines, as broad lines have a deleterious effect on both resolution and signal-to-noise ratio. The peaks of the segmentally labeled AlgE4 (A-[^{15}N]-R) are expected to be broader than the peaks of the [^{15}N]-R-module alone due to the size difference. This was confirmed by comparing the line widths of 14 peaks from both spectra. The average line width in the HSQC-spectrum of segmentally labeled AlgE4 (A-[^{15}N]-R) was 17.35 ± 1.7 Hz which is significantly broader than the average line width of 10.1 Hz ± 0.9 Hz in the HSQC of the R-module alone. Deuteration and the use of TROSY reduced the line width of

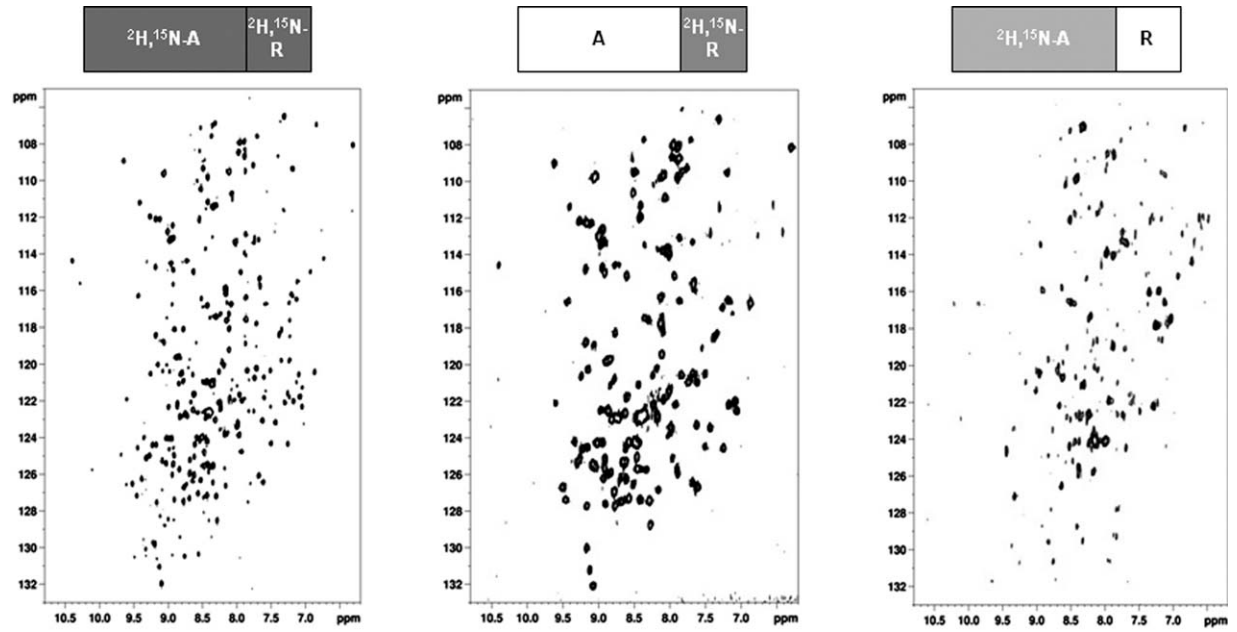

Figure 5. TROSY NMR spectra of [2H, 15N]-AlgE4, A-[2H, 15N]-R and [2H, 15N]-A-R.

the segmentally, labeled AlgE4 (A-[^{2}H, ^{15}N]-R) significantly to 11.2 Hz ± 1.3 Hz.

Activity studies

Two different epimerisation tests verified the activity of the segmentally labeled AlgE4. The NMR-based assay,[23] whose result is shown in Figure 6, certified that A-[^{15}N]-R epimerizes poly-M-alginate to the typical MG-pattern characteristic for AlgE4.[24] As the A-module alone is also active,[30] this is not necessarily a proof for the presence of active AlgE4. However, the epimerisation rate of the A-module alone is significantly lower than that of full-size AlgE4. Hence, a determination of the specific activity of the ligated product is needed to prove that the ligation product is as active as the native protein. By help of a tritium-release assay,[26] it was proven that ligated AlgE4 was able to epimerise poly-M into the expected MG-blocks with comparable specific activity to that of native AlgE4. Wild-type AlgE4 showed a specific activity of 887 ± 140 counts μg^{-1} $hrs^{-1,}$ while *in vitro* ligated AlgE4 showed an activity of 1300 ± 88 counts μg^{-1} hrs^{-1}.

Discussion

Previously, structural studies of AlgE4[28,29] were conducted either on the A- or R-module alone. These gave useful information on the reaction mechanism of the A-module as well as indications of the importance of the R-module for an active epimerase. But no information about the interactions between the two domains, functional aspects and their orientation with respect to each other in the presence and absence of alginate oligomers were gained from these studies. Yet, this information could explain the puzzling effect that the R-module connected to an A-module enhances the reaction rate although it has no activity on its own. Analysis of domain-domain interactions by NMR is difficult with conventional uniform labeling due to the size of AlgE4. Thus segmental isotopic labeling of individual domains in the full-length protein is a very attractive possibility to

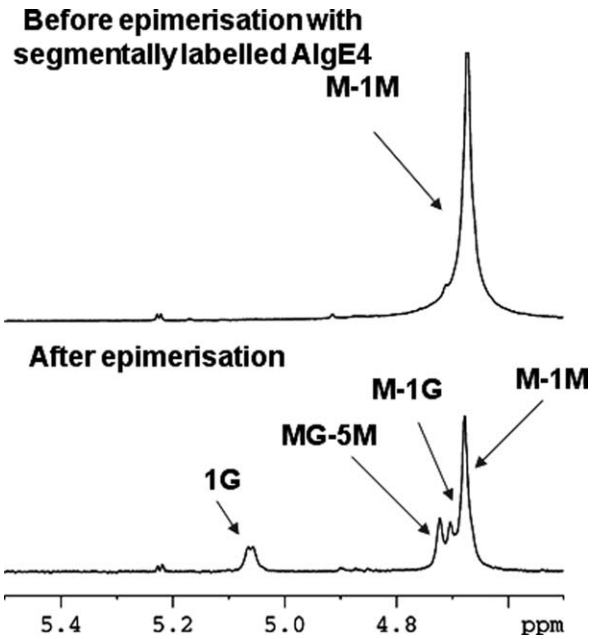

Figure 6. 1D NMR spectra of an alginate sample before (upper) and after (lower) epimerisation with ligated AlgE4. Poly-M-alginate was epimerized by AlgE4 over night, partially hydrolysed, freeze-dried and redissolved in D_2O. The NMR spectrum confirms the typical MG-pattern produced by AlgE4.

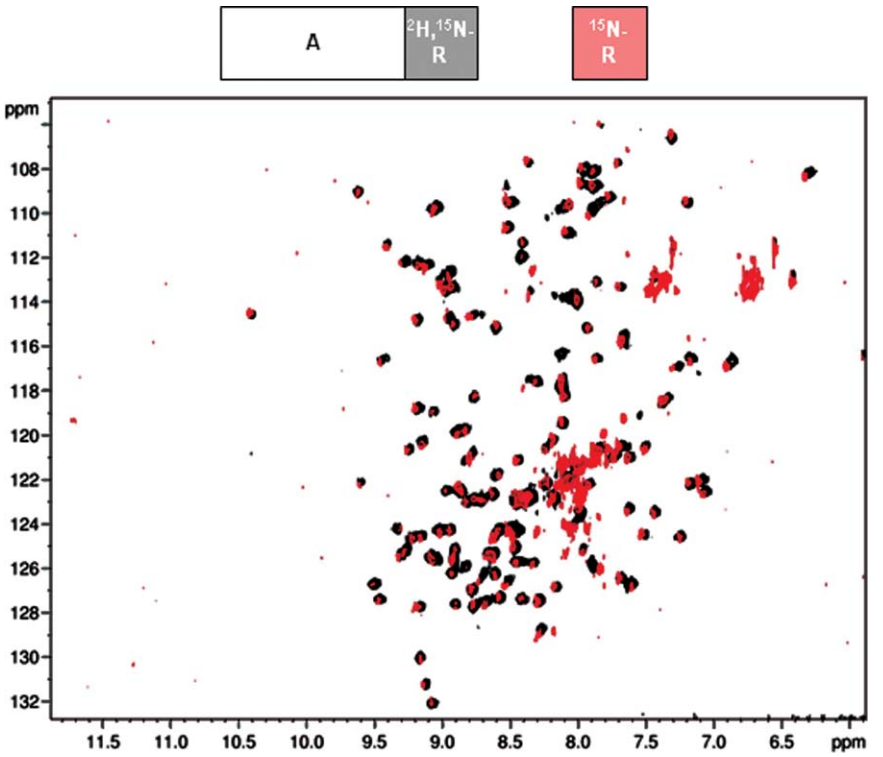

Figure 7. TROSY-NMR of A-[^{2}H, ^{15}N]-R (black) and [^{1}H,^{15}N]-HSQC of [^{15}N]-R-module (red). Extra peaks occurring in the spectrum of the R-module around 8 ppm/122 ppm stem from the unstructured, C-terminal signal sequence that is present in the R-module, but not in the Int$_C$-R construct (Fig. 3).

analyze the structure-function relationship of the individual AlgE4 modules. Figure 7 shows an overlay of NMR spectra of the R-module alone and segmentally labeled AlgE4 (A-[^{2}H, ^{15}N]-R). It demonstrates clearly that the overall structure of the R-module is essentially the same, as most chemical shifts are identical. Only a few residues in the N-terminus of the R-module show changes in chemical shifts. This result suggests that the two domains of AlgE4 do not interact with each other in solution, at least in the absence of substrate. For *trans*-splicing the natural split intein *Npu*DnaE was used as it shows high robustness in *trans*-splicing with non-native exteins, as long as few, here only three amino acids around the splicing site are kept from the native exteins.

In this study, the choice of reducing agent had a significant effect on protein ligation. Without any reducing agent the reaction was blocked as the single cysteins of two Int$_C$-R molecules formed a disulfide bond. Most of the reducing agents tested here were thiol agents. While they reduce disulfide bridges, they also perform a nucleophilic attack on the high energetic thioester occurring after the *N-S* acyl shift as an intermediate of the ligation reaction and thus cause cleavage of the A-module—especially, at higher concentrations of reducing agent (Fig. 1). TCEP is a good alternative, as it has the necessary reduction power to reduce the dimer but has no possibility for a nucleophilic attack. The ligation was performed in the crude cell extract without initial purification. TCEP is probably most often the best choice as reducing agent for *trans*-splicing but there are cases where the reducing agent used has no effect on the amount of ligated product.[33]

We also attempted an *in vivo* ligation following a protocol from the literature.[21] However, all attempts at *in vivo* ligation yielded only minute or undetectable amounts of ligated AlgE4 (data not shown). We ascribe this to the lower solubility of Int$_C$-R, its tendency to misfold when produced at higher temperatures and its susceptibility to proteolytic cleavage. The amounts produced would by no means suffice for NMR spectroscopy, however, we could determine the specific activity of the *in vivo* ligated AlgE4 and found it to be 975 $\pm$ 126 counts μg^{-1} hrs^{-1}, that is quite close to the specific activity of the native AlgE4.

To get the narrowest line width possible, a segmentally labeled, deuterated AlgE4 (A-[^{2}H, ^{15}N]-R) was produced. It narrowed the line width of full-length AlgE4 signals in a TROSY experiment to a value only slightly bigger than the line width obtained from the R-module alone in a standard HSQC. Additionally, also the complementary segmentally labeled deuterated AlgE4 ([^{2}H, ^{15}N]-A-R) was produced. Alginate epimerisation tests confirm the activity of the ligated AlgE4. Altogether, segmentally labeled AlgE4 is a strong tool for a better understanding of the structure-function relationship of the epimerases.

Materials and Methods

Vectors for in vivo *and* in vitro *protein ligation*

The vectors used for expression of the segmentally labeled proteins have been deposited to the GenBank with accession numbers HM070247 (pSABAD92A), HM070248 (pEBDuet23A) and HM070249 (pEBDuet28A).

pSABAD92A constructs

The plasmid pSABAD92A encodes a fusion protein consisting of the C-terminal fragment of DnaE intein from *Nostoc punctiforme* (Int_C) and the R-module of the alginate epimerase AlgE4 (residue 385–533). The last 20 residue (534–553) of AlgE4 were exchange to a C-terminal His-tag for purification (EFHHHHHH).

The construct pSABAD92A has a ColE1 origin for replication and the arabinose promoter. pSABAD92A has also the ampilicin resistance gene (Amp^R). (For detailed cloning steps and the vector maps see Supporting Information).

pEBDuet23A constructs

The plasmid pEBDuet23A encodes a fusion protein consisting of the A-module of AlgE4 (residues 1–379) and the N-terminal part of the DnaE intein from *Nostoc punctiforme* (Int_N). The vector has also the kanamycin resistance gene (Kan^R), RSF origin and the expression of the fusion gene is tightly controlled by *T7*/*lac* promoter.

pEBDuet28A constructs

pEBDuet28A has also the kanamycin resistance gene (Kan^R), RSF origin and the *T7*/*lac* promoter. The fusion protein, consisting of an N-terminal His-Tag for purification, the A-module of AlgE4 and the N-terminal fragment of the DnaE intein from *Nostoc puntiforme,* is expressed after adding IPTG.

Buffers and expression media

1 L LB-medium contained 10 g Tryptone, 5 g Yeast and 5 g NaCl. The pH was adjusted to 7.2 with 4 *M* NaOH and the medium was sterilized by autoclaving.

For 1 L M9-medium 7.2 g $Na_2HPO_4{\cdot}2H_2O$, 3 g KH_2PO_4, 0.5 g NaCl and 1 g NH_4Cl were dissolved in 1 L H_2O. The pH was adjusted to 7.4 and autoclaved. Before the expression 2 mL of 1 *M* $MgSO_4$, 20 mL of trace metal, 5 mL MEM Vitamins 100x (Invitrogen) and 2 g glucose dissolved in 10 mL H_2O was added.

Trace metal for M9-medium. 0.1 g/L $ZnSO_4$, 0.8 g/L $MnSO_4$, 0.5 g/L $FeSO_4$, 0.1 g/L $CuSO_4$, and 1 g/L $CaCl_2$ unlabeled proteins were produced in LB-medium. For production of the ^{15}N-labeled proteins M9-medium was supplemented with 1 g $(^{15}NH_4)_2SO_4$. ^2H, ^{15}N labeled protein were expressed in M9-medium prepared with D^2O (99% D) and supplemented with 1 g $(^{15}NH_4)_2SO_4$ and 2 g U-^{2}H-D -glucose.

Lysis buffer contained 20 m*M* HEPES pH 6.9, 800 m*M* NaCl, 10 m*M* $CaCl_2$, 0.1% Triton X. Folding buffer contained 20 m*M* HEPES pH 6.9, 800 m*M* NaCl, 5 m*M* $CaCl_2$. The elution buffer consists of 20 m*M* HEPES pH 6.9, 800 m*M* NaCl, 250 m*M* imidazole, 5 m*M* $CaCl_2$. One-hundred milliliter 2 × SDS buffer contained 10 mL of 1.5 *M* TRIS (pH 6.8), 6 mL 20% SDS, 30 mL glycerol and 1.8 mg bromophenol blue. To 2 mL aliquots 100 μL 1 *M* DTT was added.

Expression of Int_C–R, [^{15}N]-Int_C–R and [^{2}H, ^{15}N]-Int_C–R

The *E.coli* cells with the plasmid pSABAD92A were grown in 1 L LB-medium, 1 L M9-medium supplemented with 1 g $(^{15}NH_4)_2SO_4$ or M9-medium prepared with 99% D_2O and supplemented with 1 g $(^{15}NH_4)_2SO_4$ and 2 g U-^{2}H-D-glucose. 100 μg/mL ampicillin was added. The cells were grown at 37 °C to an OD_{600} of 0.5–0.7 and induced with 0.2% (w/v) arabinose and further incubated for 3 h. The cells were harvested, resuspended in lysis buffer and stored at −20 °C for further purification.

Expression of A–Int_N and [^{2}H, ^{15}N]-A–Int_N

Cells harboring the plasmid (pEBDuet23A or pEBDuet28A used only for ligation test) for A–Int_N were grown in 1 L LB-medium or 1 L M9-medium in D_2O supplemented with 1 g $(^{15}NH_4)_2SO_4$ and 2 g U-^{2}H-D-glucose, respectively, containing 50 μg/mL kanamycin at 37°C to an OD_{600} of 0.5–0.7. The cell culture was incubated on ice for 5 min. The protein was induced with 1 m*M* IPTG at 15 °C and incubated over night. The cells were harvested and resuspended in lysis buffer.

Refolding of [^{15}N]-Int_C–R/[^{2}H,^{15}N]-int_C–R and [^{2}H, ^{15}N]-A–Int_N

The cells containing the expressed protein were thawed and lysed by sonication. The pellets were solubilised in 3 mL 20 m*M* HEPES pH 6.9, 6 *M* GdmCl, 800 m*M* NaCl, 5 m*M* $CaCl_2$ at 4 °C. The solution was diluted 10 times in folding buffer and was dialyzed against the same buffer at 4 °C.

In vitro *ligation tests*

For one reaction 100 μL of A-Int_N with a concentration of 0.1 m*M* were mixed with 100 μL of Int_C-R. The reactions were started by adding reducing agent. The effect of temperature (room temperature, 30 and 37°C), pH (7 and 8), concentration of reducing agents (2.5, 5, and 10 m*M*) as well as different reducing agents (DTT, mercaptoethanol, GSH and cysteamine) were tested. At selected time points 20 μL samples for SDS-PAGE analysis were collected. The ligation was stopped by adding aliquots of 2 × SDS-buffer to the samples, followed by heating at 95°C for 5 min. The samples were loaded on 12% SDS polyacrylamide gels.

In vitro *ligation of segmentally labeled AlgE4*

Cells containing A-Int_N were sonicated. The crude cell extract was mixed with the refolded R-Int_C, where one tablet of protease inhibitor ("complete EDTA-free" from Roche Diagnostics) and TCEP to a final concentration of 5 m*M* had been added to initiate the reaction. After incubation for 2 hours at room temperature, the segmentally labeled AlgE4 was purified from the reaction mixture.

Purification of ligated AlgE4

The ligation solution was loaded onto a 1 mL Ni^{2+} His-Trap™ FF crude (GE Healthcare) column equilibrated in folding buffer. The column was washed with 10 column volumes of this buffer. Ligated AlgE4 and Int_C-R were eluted by a linear gradient to the elution buffer. Ligated AlgE4 was then separated from Int_C-R by size-exclusion chromatography (Superdex 75 HR 10/30).

NMR measurements

Protein NMR spectra were recorded at 298 K on a BRUKER DRX 600 spectrometer equipped with a 5 mm xyz-gradient TXI (H/C/N) probe using TopSpin 1.3 and a BRUKER Avance 800 MHz spectrometer equipped with 4 channels operating at a field strength of 18.8 T, equipped with a 5 mm TCI Cryoprobe, using TopSpin 2.0. The sample buffer was 10 m*M* HEPES pH 6.9 and 25 m*M* $CaCl_2$.

NMR spectra of alginate samples were recorded at 363 K on a BRUKER DRX 400 spectrometer equipped with a 5 mm z-gradient DUL (H/C) probe.

To reduce the viscosity of the alginate samples for obtaining higher resolution of the NMR analyses, the alginate samples were degraded by mild acid hydrolysis[37] and dialyzed, freeze-dried and redissolved in D_2O (99.9% D) before the NMR measurement.

Measurement of epimerase activity by radioisotope assay and NMR

The two different activity tests were performed as reported.[30] The relative amount of MG after the epimerisation was calculated as described.[38]

Line width

A Lorentzian line shape function was fitted to slices taken through 14 different peaks with the line width as variable parameter.

$$I = \sum_{x} \frac{w^2}{(w^2 + (x - t)^2)}$$

where

w line width ppm
x point in ppm
t middle of the Lorentz curve in ppm.

Acknowledgments

The expert technical assistance by Hanne Krone Nielsen is gratefully acknowledged. The 800 MHz spectra were obtained using the Bruker Avance spectrometer of the Danish Instrument Center for NMR Spectroscopy of Biological Macromolecules. Furthermore, we thank Wenche I. Strand and Helga Ertesvåg for kind assistance in the epimerzation assays.

References

1. Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R (2006) Novel 13C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. J Magn Reson 178:56–64.
2. Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94:12366–12371.
3. Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE (1996) Selective methyl group protonation of perdeuterated proteins. J Mol Biol 263:627–636.
4. Neri D, Szyperski T, Otting G, Senn H, Wüthrich K (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling. Biochemistry 28:7510–7516.
5. Kainosho M, Tsuji T (1982) Assignment of the three methionyl carbonyl carbon resonances in *Streptomyces subtilisin* inhibitor by a carbon-13 and nitrogen-15 double-labeling technique. A new strategy for structural studies of proteins in solution. Biochemistry 21: 6273–6279.
6. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440: 52–57.
7. Yamazaki T, Otomo T, Oda N, Kyogoku Y, Uegaki K, Ito N, Ishino Y, Nakamura H (1998) Segmental isotope labeling for protein NMR using peptide splicing. J Am Chem Soc 120:5591–5592.
8. Xu R, Ayers B, Cowburn D, Muir TW (1999) Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. Proc Natl Acad Sci USA 96:388–393.
9. Otomo T, Teruya K, Uegaki K, Yamazaki T, Kyogoku Y (1999) Improved segmental isotope labeling of proteins and application to a larger protein. J Biomol NMR 14: 105–114.
10. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. Science 266:776–779.
11. Severinov K, Muir TW (1998) Expressed Protein Ligation, a novel method for studying protein-protein interactions in transcription. J Biol Chem 273:16205–16209.
12. Zhao W, Zhang Y, Cui C, Li Q, Wang J (2008) An efficient on-column expressed protein ligation strategy: application to segmental triple labeling of human apolipoprotein E3. Protein Sci 17:736–747.
13. Mao H, Hart SA, Schink A, Pollok BA (2004) Sortase-mediated protein ligation: A new method for protein engineering. J Am Chem Soc 126:2670–2671.
14. Machova Z, Eggelkraut-Gottanka Rv, Wehofsky N, Bordusa F, Beck-Sickinger AG (2003) Expressed enzymatic ligation for the semisynthesis of chemically modified proteins. Angew Chem Int Edit 42:4916–4918.
15. Mootz HD (2009) Split inteins as versatile tools for protein semisynthesis. Chembiochem 10:2579–2589.

16. David R, Richter MP, Beck-Sickinger AG (2004) Expressed protein ligation. Method and applications. Eur J Biochem 271:663–677.
17. Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. Nat Methods 3:429–438.
18. Iwaï H, Züger S (2007) Protein ligation: applications in NMR studies of proteins. Biotechnol Genet Eng Rev 24: 129–145.
19. Paulus H (2000) Protein splicing and related forms of protein autoprocessing. Annu Rev Biochem 69:447–496.
20. Perler FB (2002) InBase: the intein database. Nucl Acids Res 30:383–384.
21. Züger S, Iwaï H (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nat Biotechnol 23: 736–740.
22. Busche AE, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Dötsch V, Iwaï H (2009) Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein. Angew Chem Int Ed Engl 48:6128–6131.
23. Ertesvåg H, Høidal HK, Schjerven H, Svanem BI, Valla S (1999) Mannuronan C-5-epimerases and their application for in vitro and in vivo design of new alginates useful in biotechnology. Metab Eng 1:262–269.
24. Høidal HK, Ertesvåg H, Skjåk-Bræk G, Stokke BT, Valla S (1999) The recombinant *Azotobacter vinelandii* mannuronan C-5-epimerase AlgE4 epimerizes alginate by a nonrandom attack mechanism. J Biol Chem 274: 12316–12322.
25. Ertesvåg H, Doseth B, Larsen B, Skjåk-Bræk G, Valla S (1994) Cloning and expression of an *Azotobacter vinelandii* mannuronan C-5-epimerase gene. J Bacteriol 176:2846–2853.
26. Svanem BI, Skjåk-Bræk G, Ertesvåg H, Valla S (1999) Cloning and expression of three new *Azotobacter vinelandii* genes closely related to a previously described gene family encoding mannuronan C-5-epimerases. J Bacteriol 181:68–77.
27. Ertesvåg H, Høidal HK, Hals IK, Rian A, Doseth B, Valla S (1995) A family of modular type mannuronan C-5-epimerase genes controls alginate structure in *Azotobacter vinelandii*. Mol Microbiol 16:719–731.
28. Rozeboom HJ, Bjerkan TM, Kalk KH, Ertesvåg H, Holtan S, Aachmann FL, Valla S, Dijkstra BW (2008) Structural and mutational characterization of the catalytic A-module of the mannuronan C-5-epimerase AlgE4 from *Azotobacter vinelandii*. J Biol Chem 283: 23819–23828.
29. Aachmann FL, Svanem BI, Güntert P, Petersen SB, Valla S, Wimmer R (2006) NMR structure of the R-module: a parallel beta-roll subunit from an *Azotobacter vinelandii* mannuronan C-5 epimerase. J Biol Chem 281:7350–7356.
30. Ertesvåg H, Valla S (1999)The A modules of the *Azotobacter vinelandii* mannuronan-C-5-epimerase AlgE1 are sufficient for both epimerization and binding of Ca2+. J Bacteriol 181:3033–3038.
31. Iwaï H, Züger S, Jin J, Tam PH (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. FEBS Lett 580:1853–1858.
32. Zettler J, Schutz V, Mootz HD (2009)The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Lett 583: 909–914.
33. Aranko AS, Züger S, Buchinger E, Iwaï H (2009) In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. PLoS One 4: e5185.
34. Martin DD, Xu MQ, Evans TC (2001) Characterization of a naturally occurring trans-splicing intein from *Synechocystis* sp. PCC6803. Biochemistry 40:1393–1402.
35. Nichols NM, Benner JS, Martin DD, Evans TC (2003) Zinc ion effects on individual Ssp DnaE intein splicing steps: regulating pathway progression. Biochemistry 42:5301–5311.
36. Nichols NM, Evans TC (2004) Mutational analysis of protein splicing, cleavage, and self-association reactions mediated by the naturally split Ssp DnaE intein. Biochemistry 43:10265–10276.
37. Ertesvåg H, Skjåk-Bræk G, (1999) Modification of alginate using mannuronan C-5-epimerases. In: Bucke C, Ed. Carbohydrate biotechnology protocols. Humana Press: Totowa, NJ, 71–78.
38. Grasdalen H, Larsen B, Smidsrød O (1979) A p.m.r. study of the composition and sequence of uronate residues in alginates. Carbohyd Res 68:23–31.