



## Solution structure of DnaE intein from *Nostoc punctiforme*: Structural basis for the design of a new split intein suitable for site-specific chemical modification

Jesper S. Oeemig<sup>1</sup>, A. Sesilja Aranko<sup>1</sup>, Janica Djupsjöbacka, Kimmo Heinämäki, Hideo Iwai<sup>\*</sup>

Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, P.O. Box 65, Helsinki FIN-00014, Finland

### ARTICLE INFO

#### Article history:

Received 31 January 2009

Revised 13 March 2009

Accepted 25 March 2009

Available online 1 April 2009

Edited by Hans Eklund

#### Keywords:

Intein

Protein splicing

NMR spectroscopy

Protein ligation

Nuclear spin relaxation

Chemical modification

Protein *trans*-splicing

### ABSTRACT

**Naturally split DnaE intein from *Nostoc punctiforme* (*Npu*) has robust protein *trans*-splicing activity and high tolerance of sequence variations at the splicing junctions. We determined the solution structure of a single chain variant of *NpuDnaE* intein by NMR spectroscopy. Based on the NMR structure and the backbone dynamics of the single chain *NpuDnaE* intein, we designed a functional split variant of the *NpuDnaE* intein having a short C-terminal half (C-intein) composed of six residues. In vivo and in vitro protein ligation of model proteins by the newly designed split intein were demonstrated.**

© 2009 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

### 1. Introduction

Protein *trans*-splicing is a promising tool for post-translational ligation of protein fragments [1–3]. Protein splicing is a self-catalytic reaction catalyzed by an intervening sequence in a host protein, termed an intein, which ligates the flanking protein sequences, termed exteins, by a peptide bond and concomitantly excises itself from the host protein [3]. Protein splicing in *trans* could ligate two foreign peptide chains that are fused with either N- or C-terminal fragments (N- or C-inteins) of a split intein [4–6]. Protein *trans*-splicing system has opened many applications including segmental isotopic labelling of proteins, protein cyclization, in vivo protein engineering, and site-specific chemical modifications [6–15]. However, protein ligation by protein *trans*-splicing can be significantly modulated by the junction sequences as well as by the extein sequences, which could limit the applications of protein *trans*-splicing [16]. Fusion proteins with artificially split intein fragments often become insoluble because of the unstructured split fragments [6,17]. Therefore, recovering and refolding of the two precursor fragments from insoluble fractions are often necessary in order to perform protein ligation using protein *trans*-splicing with artificially split inteins

[6,17]. In contrast, naturally occurring split inteins from cyanobacteria such as the DnaE intein from *Nostoc punctiforme* (*Npu*) usually do not require denaturation/renaturation steps for protein *trans*-splicing [4,7,16]. *NpuDnaE* intein has also higher tolerance of sequence variations at the splicing junctions and superior splicing activity to the widely used DnaE intein from *Synechocystis* sp. PCC6803 (*Ssp*) [16]. Therefore, *NpuDnaE* intein might be more suitable for many biotechnological applications than other inteins. However, the C-terminal fragment of a split intein (C-intein or Int<sub>C</sub>) of the naturally occurring split *NpuDnaE* intein consists of 36 residues (*NpuInt*<sub>C36</sub>), which could be too long to be chemically synthesized and could disturb the solubility of the precursor fragment fused to the C-intein. We have been interested in identifying shorter functional C-inteins that still retain robust protein *trans*-splicing activity. A functional C-intein consisting of 15 residues (*NpuInt*<sub>C15</sub>) was previously identified in our group by shortening the C-intein systematically [18,19]. The systematic approach was applied because no three-dimensional structure of DnaE intein was available when the project was started [18].

To rationally design new split inteins with even shorter C-inteins, we determined the NMR structure of a single chain variant of *NpuDnaE* intein. Moreover, we investigated the backbone dynamics by NMR relaxation measurements in order to identify locations with internal conformational fluctuations that might be important for introducing a new split site. Based on the NMR

<sup>\*</sup> Corresponding author. Fax: +358 9 191 59541.

E-mail address: [hideo.iwai@helsinki.fi](mailto:hideo.iwai@helsinki.fi) (H. Iwai).

<sup>1</sup> These authors equally contributed to this work.

structure and the  $^{15}\text{N}$  relaxation analysis, we designed a new split intein from *NpuDnaE* intein and tested the protein *trans*-splicing activity in vivo as well as in vitro.

## 2. Materials and methods

### 2.1. Construction of *NpuInt*<sub>NAC6</sub>/*NpuInt*<sub>C6</sub>

The plasmid (pHYDuet93) for the expression of an N-terminal His-tagged B1 domain of IgG binding protein G (H<sub>6</sub>-GB1) fused with *NpuInt*<sub>NAC6</sub> was constructed from pSKDuet16 containing the gene of the single chain variant *NpuDnaE* intein and two GB1s as extens by replacing the codon of residue 132 with a stop codon by two oligonucleotides: #HK254: 5'-CACTCAAAAATTAAGCTTACCTTCTAATTGTTTC and #HK255: 5'-GAAGCTAAAGCTTAATTTT-GAGTGCAAAATTATG [16,18]. This plasmid contains RSF origin and a T7 promoter that is inducible by isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) [7]. *NpuInt*<sub>C6</sub> was constructed as a fusion protein with GB1. The gene of *NpuInt*<sub>C6</sub>-GB1 was amplified from pMH-BAD14C by a pair of the two oligonucleotides: #SZ015: 5'-TGC-CAAGCTTATTCCGTTACGGTG and #HK256: 5'-ACTAGTCATATGGGCTTCATAGCTTCTAATTG or #HK112: 5'-CTAAAGCTTAATGATGATGATGATG and #HK256 and cloned into pSKBAD2 by using NdeI and HindIII sites, which resulted in pSABAD109 (*NpuInt*<sub>C6</sub>-GB1) or pHYBAD94 (*NpuInt*<sub>C6</sub>-GB1-H<sub>6</sub>) [16]. These vectors contain ColE1 origin and an arabinose promoter for the expression.

### 2.2. Protein preparations

A uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labelled sample of the single chain *NpuDnaE* intein was produced and purified as previously described [21]. A 2 mM doubly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labelled NMR sample was used for the NMR measurements required for the structure calculation. The  $^{15}\text{N}$  relaxation measurements were performed with  $^{15}\text{N}$ -labelled sample. *NpuInt*<sub>NAC6</sub> fused with an N-terminally His-tagged GB1 and *NpuInt*<sub>C6</sub> fused with an C-terminally His-tagged GB1 were individually expressed in *Escherichia coli* ER2566 (New England Labs, Ipswich, USA) and purified by Immobilized Metal Ion Affinity Chromatography (IMAC) with Profinia protein purification system (BioRAD, USA). The purified proteins were dialyzed against 10 mM Tris, 500 mM NaCl and 1 mM EDTA, pH 7.0.

### 2.3. NMR measurements

NMR measurements were performed at  $^1\text{H}$  frequency of 600 MHz on Varian INOVA spectrometer equipped with a triple resonance cryogenic probe head or on Varian INOVA 800 MHz spectrometer equipped with a triple resonance probe head. All the NMR experiments were performed at 298 K. The  $T_1$ ,  $T_2$ , and  $^{15}\text{N}\{^1\text{H}\}$ -NOE measurements were performed at  $^1\text{H}$  frequency of 600 MHz at 298 K. The  $T_1(^{15}\text{N})$  relaxation times were determined with the following  $T_1$  relaxation delays: 0, 50, 100, 150, 200, 300, and 500 ms [22,23]. The  $T_2(^{15}\text{N})$  relaxation times were obtained based on a CPMG-type sequence with the interval of 625  $\mu\text{s}$  between  $^{15}\text{N}$  180° pulse in the CPMG cycles with the following  $T_2$  relaxation delays: 10, 30, 50, 70, 90, 110, 150, and 190 ms.  $^{15}\text{N}\{^1\text{H}\}$ -NOEs were determined by the comparison of the peak volumes between the HSQC spectra with and without  $^1\text{H}$  saturation of 2.5 s with special care for water-suppression [22,23,34]. The volumes of the peaks were analyzed and fitted by the program ANALYSIS [24]. The structure calculation was performed based on the resonance assignments by the automatic NOE assignment algorithm implemented in the program CYANA [25,26]. The structures were calculated based on the NOE peaks obtained from 3D  $^{15}\text{N}$ -resolved [ $^1\text{H}, ^1\text{H}$ ]-NOESY and 3D  $^{13}\text{C}$ -resolved [ $^1\text{H}, ^1\text{H}$ ]-NOESY spectra

with mixing times of 80 and 70 ms, respectively. Peaks were picked and integrated using the program ANALYSIS [24]. The NOE peak lists together with the assigned chemical shift values were used as inputs for the automatic NOE assignments and for the structure calculation [25,26]. At the last cycle, 3154 distance restraints was used for the structure calculations starting from 100 random conformations. The 20 best CYANA conformers with the lowest values of the CYANA target function were subjected to the energy minimization by the program AMBER [27,28].

## 3. Results

### 3.1. NMR structure of a single chain variant of *NpuDnaE* intein

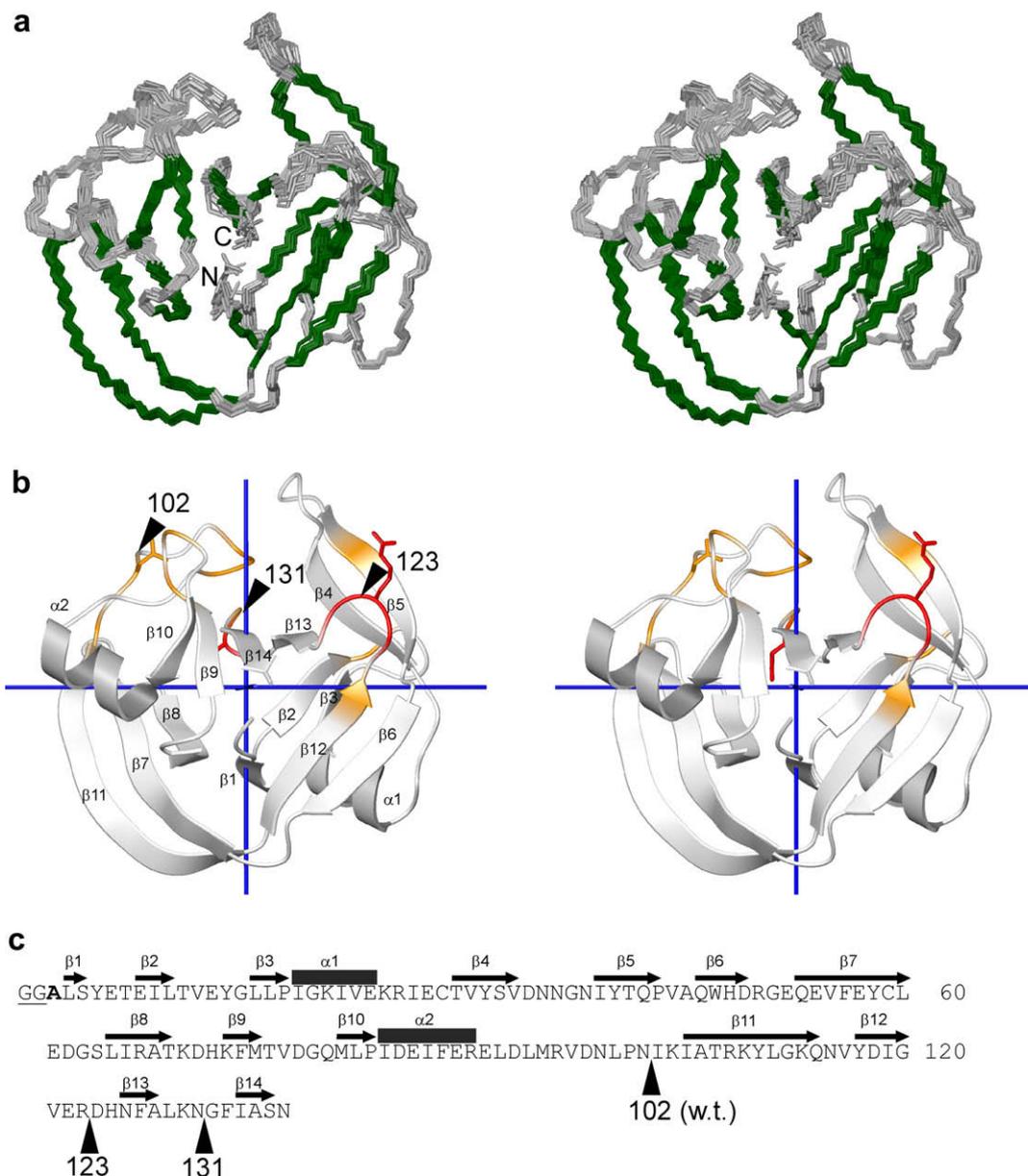
A superposition of the 20 energy-minimized conformers is displayed with an RMSD value of 0.49 Å, representing the well-defined structure of the single chain *NpuDnaE* intein (PDB entry: 2KEQ, Fig. 1a, Table 1). The structure is composed mostly of  $\beta$ -strands with two short  $\alpha$ -helices. The overall structure has a disk shape of a diameter of 19–22 Å with 14 Å thickness, resembling a horseshoe-like fold common in Hedgehog/intein (HINT) domains [29,30]. A single chain variant of *NpuDnaE* intein was constructed by connecting 102 residues of the N-intein (*NpuInt*<sub>NAC36</sub>) and 35 residues of the C-intein without the first methionine of the C-intein (*NpuInt*<sub>C36</sub>). The connected region forms a loop without regular secondary structures, which might suggest that a few residues in the vicinity of the naturally occurring split site could be removed without loss of the function (Fig. 1). The structure of the single chain *NpuDnaE* intein has a compact globular structure without any highly disordered region. The largest disordered region is a loop between  $\beta$ 12 and  $\beta$ 13, where the resonances of the backbone amide groups as well as the side-chains are missing and thereby lacking substantial NOEs for this region.

### 3.2. Relaxation data

Nuclear spin relaxation measurement of  $^{15}\text{N}$  spins of proteins is a popular approach for studies of global protein motions and internal backbone mobility [31–34]. We hypothesized that the residues indicating higher internal mobility are suitable locations for a new split site and recorded a set of  $^{15}\text{N}$  relaxation measurements (Fig. 2a, b and d). Assuming an isotopic model, a global correlation time,  $\tau_c$  can be derived from  $T_1/T_2$  ratios for all the residues [33,34]. Residues having conformational exchanges that shortened  $T_2$  values or internal motions contributing  $T_1$  could be identified by comparing the  $T_1/T_2$  ratios for individual backbone amide groups with the average  $T_1/T_2$  ratio,  $\langle T_1/T_2 \rangle$ . By using the program DASHA, the average  $T_1/T_2$  ratio,  $\langle T_1/T_2 \rangle = 8.0 \pm 0.4$  was obtained at  $^1\text{H}$  frequency of 600.0 MHz after disregarding the residues for which  $T_1/T_2$  ratio is outside of the standard deviation (Fig. 2c), which results in an apparent global rotational correlation time of  $8.5 \pm 0.5$  ns [35]. The disregarded residues are likely to have conformational exchanges or faster internal motions. The conformational exchanges were suggested for the following residues: 10, 40, 76, 78, 96–98, 102, 104, 120, and 132 (Fig. 1b and Fig. 2c). These residues are located in loops except for residue 40 (Fig. 2). Residues 96–98, 102, and 104 are located in the loop connecting the N- and C-inteins to make the single chain variant, suggesting higher conformational fluctuations in this region. Thus, the internal backbone flexibility might be a key feature for designing novel functional split inteins.

### 3.3. Design of a new split intein

The NMR solution structure and the backbone dynamics of *NpuDnaE* intein were analyzed in order to rationally design new



**Fig. 1.** (a) Stereo view of the superimposed 20 energy-minimized CYANA conformers, representing the three-dimensional structure of the single chain variant of *NpuDnaE* intein in solution. The locations of the N- and C-termini are indicated.  $\beta$ -Strands are highlighted in green. (b) Ribbon drawing of the lowest energy conformer of the single chain variant of *NpuDnaE* intein. Calculated axes of the structure are displayed by the three lines, indicating the center of the structure. The horizontal line is the longest axis. The shortest axis of the structure is perpendicular to the plane of the page. Side-chains of the preceding residues of the split sites indicated by filled triangles in (c) are displayed. The residues of which amide groups were not detectable in the NMR spectra were colored in red. The residues indicating conformational exchanges are colored in orange. (c) Primary structure of the single chain variant of *NpuDnaE* intein used for the structure determination. The locations of the secondary structures are indicated above the amino acid sequence. C1A mutation is indicated in bold. Filled triangles indicate the functional split sites that were experimentally tested. The program MOLMOL was used to produce (a) and (b) [40].

split inteins that could constitute a shorter and functional C- or N-intein for protein ligation by protein *trans*-splicing [18,20]. Because there was no three-dimensional structure of *SspDnaE* intein available when the project started [37], we previously shifted the split site of *SspDnaE* systematically toward the C-terminus and discovered that only the C-terminal 16 residues were required for the functional C-intein of *SspDnaE* [18]. This split site was also applied to *NpuDnaE* intein and demonstrated that the C-terminal 15 residues of *NpuDnaE* intein were sufficient for efficient protein *trans*-splicing [18]. This split site is located in the loop between  $\beta$ 12 and  $\beta$ 13 of the *NpuDnaE* intein structure (Fig. 1b). It is noteworthy that residue 120 indicates conformational exchange from the relaxation data and is also located in the same loop, where reso-

nances of the backbone amide groups of residues 122–124 could not be observed presumably due to internal fluctuations that shortened the  $T_2$  relaxation times [21]. From the relaxation analysis, we found a few locations with reduced  $T_2$  relaxation times presumably due to conformational exchanges, which are around residue 10, residues 76–78, residues 96–104, and residue 132 (Fig. 2). These locations are coincident with the loops with higher disorder in the NMR structure (Fig. 2e). The naturally occurring split site of *NpuDnaE* intein is precisely located in the loop where a number of residues (residues 96–104) indicate conformational exchanges. The NMR structure of *NpuDnaE* intein and the relaxation data might suggest a potential split site just after residue 97. In fact, naturally split intein in *Nanoarchaeum equitans* family

**Table 1**

The experimental NMR data for the structure calculation and the structural statistics of the 20 energy-minimized conformers of *NpuDnaE* intein.

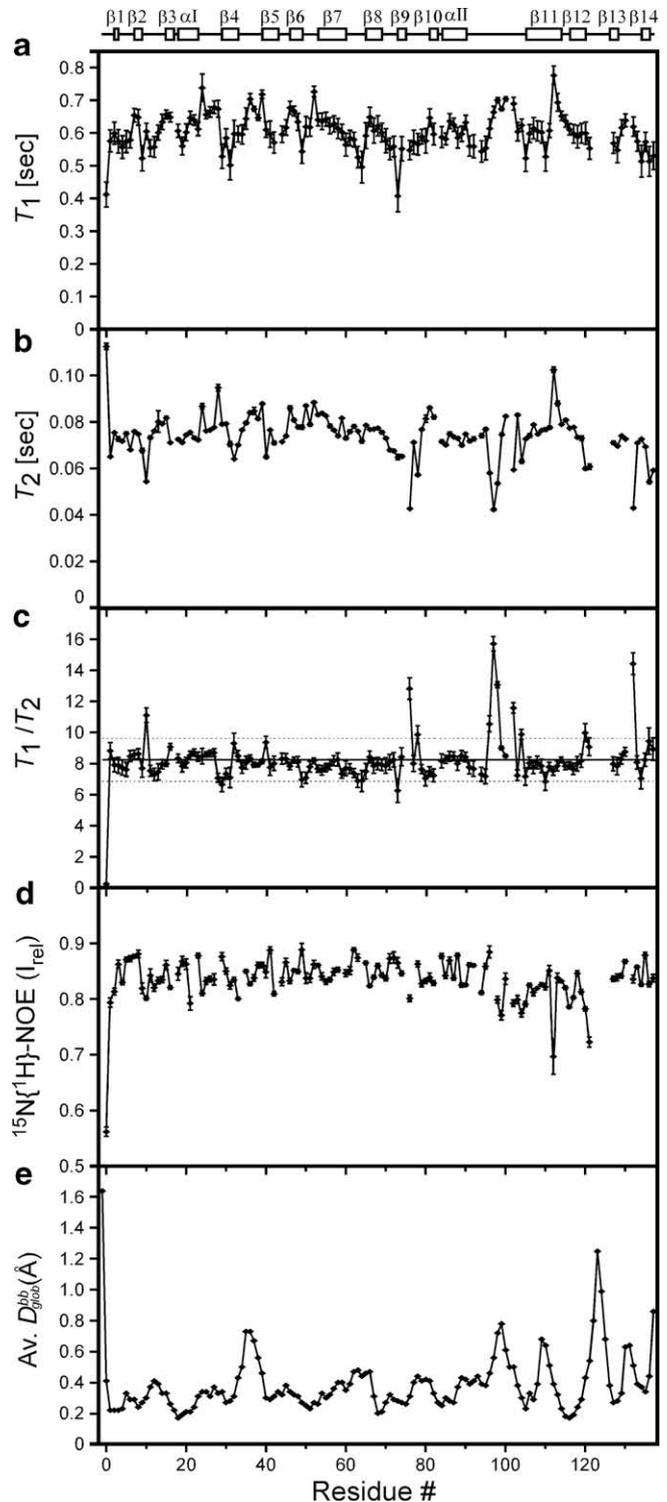
Quantity	Value
Upper distance limits	3154
Short range NOE ( $ i - j  \leq 1$ )	1498
Medium range NOE ( $1 <  i - j  < 5$ )	433
Long range NOE ( $ i - j  \geq 5$ )	1223
Residual Cyana target function ( $\text{\AA}^2$ )	$0.37 \pm 0.17$
<i>Amber energies (kcal/mol)</i>	
Total	$-5711.78 \pm 27.91$
van der Waals	$-1173.35 \pm 9.06$
Electrostatic	$-9503.76 \pm 240.66$
<i>Residual NOE violation</i>	
Number $\geq 0.1$ \AA	$4 \pm 2$
Maximum (\AA)	$0.24 \pm 0.07$
<i>Residual dihedral angle violations</i>	
Number $\geq 2.5^\circ$	$1 \pm 1$
Maximum ( $^\circ$ )	$3.38 \pm 0.24$
<i>RMSD from ideal geometry</i>	
Bond length (\AA)	$0.0098 \pm 0.0001$
Bond angles ( $^\circ$ )	$1.979 \pm 0.016$
<i>RMSD to mean coordinate</i>	
Backbone 1-137 (\AA)	$0.45 \pm 0.07$
Heavy atoms 1-137 (\AA)	$0.94 \pm 0.07$
<i>Ramachandran plot statistics<sup>a</sup></i>	
Most favoured regions (%)	87.5
Additional allowed regions (%)	12.4
Generously allowed regions (%)	0.1
Disallowed regions (%)	0.0

<sup>a</sup> As determined by PROCHECK [41,42].

B-type DNA polymerase (*NeqPol* intein) has the split site after residue 98 [39]. Conformational exchange was also suggested for residue 10, which is in the proximity of the functional split site found in *SspDnaB* intein [36]. The coincidence between the functional split sites and the locations with conformational exchanges inspired us to introduce a new split site after residue 131 because residue 132 indicates conformational exchange and the backbone amide resonance of the preceding residue (residue 131) was not detectable in the HSQC spectra. These two residues are located in the turn between  $\beta 13$  and  $\beta 14$ .

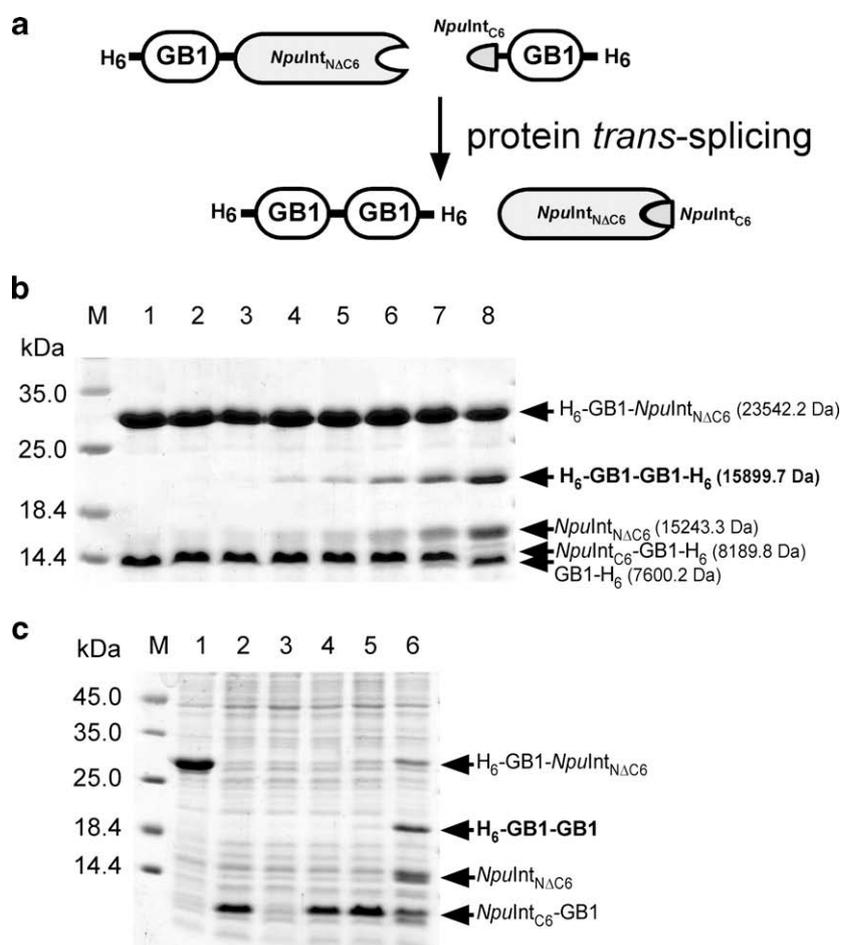
#### 3.4. Protein trans-splicing by *NpuInt<sub>NAC6</sub>/NpuInt<sub>C6</sub>*

To test our hypothesis that the split site can be introduced in the loops indicating conformational exchanges from the NMR analysis, we constructed a new split *NpuDnaE* intein, *NpuInt<sub>NAC6</sub>/NpuInt<sub>C6</sub>* by splitting the single chain *NpuDnaE* intein after residue 131. The C-intein of the newly engineered intein has the C-terminal six residues of the *NpuDnaE* intein and is called *NpuInt<sub>C6</sub>*. The N-intein of the newly engineered intein has the N-terminal 131 residues of the single chain *NpuDnaE* intein. This N-terminal intein fragment is called *NpuInt<sub>NAC6</sub>* because it is the N-terminal fragment (N-intein) of the single chain *NpuDnaE* intein lacking the C-terminal six residues. In order to test the protein ligation by the newly designed split *NpuDnaE* intein (*NpuInt<sub>NAC6</sub>/NpuInt<sub>C6</sub>*), each fragment of the new split intein (*NpuInt<sub>NAC6</sub>/NpuInt<sub>C6</sub>*) was fused with GB1 (Fig. 3a). The N-terminal precursor protein ( $H_6$ -GB1-*NpuInt<sub>NAC6</sub>*) contains an N-terminal His-tag, GB1, and *NpuInt<sub>NAC6</sub>* with the total theoretical molecular weight of 23542.2 Da. The C-terminal precursor protein contains *NpuInt<sub>C6</sub>*, GB1, and a C-terminal His-tag with the theoretical molecular weight of 8189.8 Da. Protein ligation by *NpuInt<sub>NAC6</sub>/NpuInt<sub>C6</sub>* was initiated by mixing the two precursors in vitro in the presence of 0.5 mM TCEP and monitored by SDS-PAGE after the mixing. The ligated product of



**Fig. 2.** Plots of the relaxation times (a)  $T_1(^{15}\text{N})$  and (b)  $T_2(^{15}\text{N})$ , (c)  $T_1/T_2$  ratios, (d)  $^{15}\text{N}\{^1\text{H}\}$ -NOEs, and (e) the average global displacement for the backbone atoms (C, N, C $^\alpha$ ) among the 20 conformers. The standard deviation for all the  $T_1/T_2$  ratios is indicated by broken lines in (c). The secondary structures are indicated above the plots.

$H_6$ -GB1-GB1- $H_6$  was formed with the kinetic constant of  $5.2 \pm 0.2 \times 10^{-5}$  ( $\text{s}^{-1}$ ) at the apparent molecular weight of about 20 kDa (Fig. 3b). This kinetic constant is comparable to the trans-splicing kinetic constant of  $6.6 \pm 1.3 \times 10^{-5}$  ( $\text{s}^{-1}$ ) reported for *SspDnaE* although the comparison between trans-splicing kinetics could not be directly compared because of the differences in the



**Fig. 3.** In vitro and in vivo protein ligation by *NpuInt*<sub>NΔC6</sub>/*NpuInt*<sub>C6</sub>. (a) Schematic drawing of protein *trans*-splicing by the designed split intein. (b) In vitro protein ligation by *NpuInt*<sub>NΔC6</sub>/*NpuInt*<sub>C6</sub>. M, Marker; lane 1, 0 min after mixing; lane 2, 5 min; lane 3, 10 min; lane 4, 30 min; lane 5, 1 h; lane 6, 2 h; lane 7, 4 h; lane 8, 24 h. (c) In vivo protein ligation by the dual vector system. M, Marker; lane 1, induction only with IPTG; lane 2, induction only with L-arabinose; lane 3, before induction; lane 4, 1 h after induction only with L-arabinose; lane 5, 1 h after the additional induction with IPTG (2 h after the induction with L-arabinose); lane 6, 2 h after the dual induction with both IPTG and L-arabinose.

extein sequences [18,38]. The molecular mass of 15899.7 Da was obtained for the ligated product by MALDI-TOF mass-spectrometry (Supplementary Fig. S1), which is in good agreement with the expected molecular mass (15899.2 Da). Despite the short length of six residues, the protein ligation was efficiently performed by *NpuInt*<sub>NΔC6</sub>/*NpuInt*<sub>C6</sub>. The C-terminal precursor fragment was mostly consumed after 24 h but the yield was estimated to be about 50% due to the cleavage reaction. Additionally, we tested in vivo protein ligation by *NpuInt*<sub>NΔC6</sub>/*NpuInt*<sub>C6</sub> by using the dual vector system with two different promoters previously developed in our group, which allows to control the expression of the two precursors individually by the addition of either IPTG or L-arabinose (Fig. 3c) [7]. Upon the dual induction with both IPTG and L-arabinose, the ligated product (H<sub>6</sub>-GB1-GB1) was produced as observed as a band of the apparent molecular weight of 19 kDa in the SDS-PAGE (lane 6, Fig. 3c). Thus, *Int*<sub>C6</sub> peptide from *NpuDnaE* could indeed be used for protein ligation in vivo as well in vitro.

#### 4. Discussion

Here we reported the NMR solution structure of a single chain *NpuDnaE* intein and demonstrated that the NMR structure and the <sup>15</sup>N relaxation measurements could be an effective tool for designing new functional split sites in inteins. The engineered functional split inteins derived from *NpuDnaE* intein in our group include *NpuInt*<sub>NΔC15</sub>/*NpuInt*<sub>C15</sub> and *NpuInt*<sub>NΔC6</sub>/*NpuInt*<sub>C6</sub> [18,19].

The C-inteins of these new split inteins constitute β-strands in the center of *NpuDnaE* intein (Fig. 1b). The C-terminal 15 residues of *NpuDnaE* intein form two anti-parallel β-strands buried exactly in the center of the horseshoe-like fold of HINT domain (Fig. 1b). These two strands resemble two fingers (β13 and β14 strands) inserted into the core of the structure. The six residues of *NpuInt*<sub>C6</sub> form the last β-strand (β14), which resembles one finger inserted in the center instead of the two fingers of *Int*<sub>C15</sub>. The simplicity of these secondary structures of *Int*<sub>C15</sub> and *Int*<sub>C6</sub> compared with the wild-type *Int*<sub>C36</sub>, which is intertwined into the remaining part of *NpuDnaE* intein structure, might be advantageous for protein *trans*-splicing reaction in some foreign contexts [18,19]. The simpler structure of *Int*<sub>C15</sub> and *Int*<sub>C6</sub> may allow the faster interaction between *Int*<sub>N</sub> and *Int*<sub>C</sub> although the detailed analysis of the interactions remains to be investigated. The newly designed split DnaE intein, *NpuInt*<sub>NΔC6</sub>/*NpuInt*<sub>C6</sub> can be used for in vivo as well as in vitro protein ligation. Moreover, the six residues of *NpuInt*<sub>C6</sub> are expected to be easily synthesized or fused to any other proteins as a ligation tag. The short length of *Int*<sub>C6</sub> peptide makes it affordable to synthesize even with a chemical modification. Particularly, *NpuInt*<sub>C6</sub> has the sequence of GFIAASN lacking any reactive primary amine except for the N-terminus, making it suitable for site-specific modification of ε-amino group in a lysine when the N-terminus is blocked. Therefore, the newly designed split DnaE of *NpuInt*<sub>NΔC6</sub>/*NpuInt*<sub>C6</sub> might be of practical use for site-specific chemical modifications of proteins. The short length of *Int*<sub>C6</sub> is also

expected to affect the solubility of the fused proteins much less than the original Int<sub>C</sub> bearing 36 residues. This would make it more suitable for many applications such as segmental isotopic labelling.

Recent advances in computational approaches and algorithms opened new possibilities to computationally design new enzymes with new functions from a defined scaffold. However, dynamical aspects of protein structures have not been taken into consideration for rational design of proteins. We here demonstrated the use of dynamical structures for the rational design of a novel functional split intein. This rational approach to design new split intein by the combination of the solution NMR structures and <sup>15</sup>N relaxation measurements could be generally applicable to other inteins as well as other enzymes.

## Acknowledgments

The authors thank Dr. K. Pääkkonen for his help in the initial structure calculation. This work is supported by the grant from the Academy of Finland (118385). J.S.O. gratefully acknowledges support from the Centre for International Mobility (CIMO).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2009.03.058.

## References

- [1] Xu, M.Q. and Evans Jr., T.C. (2004) Recent advances in protein splicing: manipulating proteins in vitro and in vivo. *Curr. Opin. Biotechnol.* 16, 440–446.
- [2] Saleh, L. and Perler, F.B. (2006) Protein splicing in *cis* and in *trans*. *Chem. Rev.* 6, 183–193.
- [3] Paulus, H. (2000) Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem.* 69, 447–496.
- [4] Wu, H., Hu, Z. and Liu, X.Q. (1998) Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* 95, 9226–9231.
- [5] Mills, K.V., Lew, B.M., Jiang, S. and Paulus, H. (1998) Protein splicing in *trans* by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein. *Proc. Natl. Acad. Sci. USA* 95, 3543–3548.
- [6] Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y. and Nakamura, H. (1998) Segmental isotope labeling for protein NMR using peptide splicing. *J. Am. Chem. Soc.* 120, 5591–5592.
- [7] Züger, S. and Iwai, H. (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat. Biotechnol.* 23, 736–740.
- [8] Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C. and Benkovic, S.J. (1999) Production of cyclic peptides and proteins in vivo. *Proc. Natl. Acad. Sci. USA* 96, 13638–13643.
- [9] Evans, T.C., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L.X., Benner, J., Liu, X.Q. and Xu, M.Q. (2000) Protein *trans*-splicing and cyclization by a naturally split intein from the *dnaE* gene of *Synechocystis* species PCC6803. *J. Biol. Chem.* 275, 9091–9094.
- [10] Iwai, H., Lingel, A. and Plückthun, A. (2001) Cyclic green fluorescent protein produced in vivo using an artificially split PI-Pful intein from *Pyrococcus furiosus*. *J. Biol. Chem.* 276, 16548–16554.
- [11] Williams, N.K., Prosser, P., Liepinsh, E., Line, I., Sharipo, A., Littler, D.R., Curmi, P.M., Otting, G. and Dixon, N.E. (2002) In vivo protein cyclization promoted by a circularly permuted *Synechocystis* sp. PCC6803 DnaB mini-intein. *J. Biol. Chem.* 277, 7790–7798.
- [12] Mootz, H.D., Blum, E.S., Tyszkiewicz, A.B. and Muir, T.W. (2003) Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo. *J. Am. Chem. Soc.* 125, 10561–10569.
- [13] Kurpiers, T. and Mootz, H.D. (2007) Regioselective cysteine bioconjugation by appending a labeled cysteine tag to a protein by using protein splicing in *trans*. *Angew. Chem. Int. Ed. Engl.* 46, 5234–5237.
- [14] Schwartz, E.C., Saez, L., Young, M.W. and Muir, T.W. (2007) Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nat. Chem. Biol.* 3, 50–54.
- [15] Ludwig, C., Pfeiff, M., Linne, U. and Mootz, H.D. (2006) Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein *trans*-splicing. *Angew. Chem. Int. Ed.* 45, 5218–5221.
- [16] Iwai, H., Züger, S., Jin, J. and Tam, P.H. (2006) Highly efficient protein *trans*-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett.* 580, 1853–1858.
- [17] Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T. and Kyogoku, Y. (1999) Improved segmental isotope labeling of proteins and application to a larger protein. *J. Biomol. NMR* 14, 105–114.
- [18] Aranko, A.S., Züger, S., Buchinger, E., and Iwai, H. (2009) In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS ONE* 4, e5185.
- [19] Muona, M., Aranko, A.S. and Iwai, H. (2008) Segmental isotopic labelling of a multi-domain protein by protein ligation using protein *trans*-splicing. *ChemBioChem* 9, 2958–2961.
- [20] Iwai, H., Aranko, A.S. and Djupsjöbacka, J. (2008) Protein ligation using protein *trans*-splicing. *J. Pept. Sci.* 14 (Suppl.), 183.
- [21] Heinämäki, K., Oemig, J.S., Djupsjöbacka, J. and Iwai, H. (2009) NMR resonance assignment of DnaE intein from *Nostoc punctiforme*. *Biomol. NMR assign.* doi:10.1007/s12104-008-9137-1.
- [22] 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) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation. *Biochemistry* 33, 5984–6003.
- [23] Kay, L.E., Nicholson, L.K., Delaglio, F., Bax, A. and Torchia, D.A. (1992) Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T1 and T2 values in proteins. *J. Magn. Reson.* 97, 359–375.
- [24] Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687–696.
- [25] Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* 273, 283–298.
- [26] Herrmann, T., Güntert, P. and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* 319, 209–227.
- [27] Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., DeBolt, S., Ferguson, D., Seibel, G. and Kollman, P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91, 1–41.
- [28] 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) A 2nd generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197.
- [29] Hall, T.M., Porter, J.A., Young, K.E., Koonin, E.V., Beachy, P.A. and Leahy, D.J. (1997) Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell* 91, 85–97.
- [30] Johnson, M.A., Southworth, M.W., Herrmann, T., Brace, L., Perler, F.B. and Wüthrich, K. (2007) NMR structure of a K1ba intein precursor from *Methanococcus jannaschii*. *Protein Sci.* 16, 1316–1328.
- [31] Kay, L.E., Torchia, D.A. and Bax, A. (1989) Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 128, 8972–8979.
- [32] Clore, G.M., Driscoll, P.C., Wingfield, P.T. and Gronenborn, A.M. (1990) Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear <sup>15</sup>N-<sup>1</sup>H NMR spectroscopy. *Biochemistry* 29, 7387–7401.
- [33] Barbato, G., Ikura, M., Kay, L.E., Pastor, R.W. and Bax, A. (1992) Backbone dynamics of calmodulin studied by <sup>15</sup>N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry* 31, 5269–5278.
- [34] Luginbuhl, P., Pervushin, K.V., Iwai, H. and Wüthrich, K. (1997) Anisotropic molecular rotational diffusion in <sup>15</sup>N spin relaxation studies of protein mobility. *Biochemistry* 36, 7305–7312.
- [35] Orekhov, V.Yu., Nolde, D.E., Golovanov, A.P., Korzhnev, D.M. and Arseniev, A.S. (1995) Processing of heteronuclear NMR relaxation data with the new software DASHA. *Appl. Magn. Reson.* 9, 581–588.
- [36] Sun, W., Yang, J. and Liu, X.Q. (2004) Synthetic two-piece and three-piece split inteins for protein *trans*-splicing. *J. Biol. Chem.* 279, 35281–35286.
- [37] Sun, P., Ye, S., Ferrandon, S., Evans, T.C., Xu, M.Q. and Rao, Z.H. (2005) Crystal structures of an intein from the split *dnaE* gene of *Synechocystis* sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. *J. Mol. Biol.* 353, 1093–1105.
- [38] Martin, D.D., Xu, M.Q. and Evans Jr., T.C. (2001) Characterization of a naturally occurring *trans*-splicing intein from *Synechocystis* sp. PCC6803. *Biochemistry* 40, 1393–1402.
- [39] Choi, J.J., Nam, K.H., Min, B., Kim, S.J., Söll, D. and Kwon, S.T. (2006) Protein *trans*-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite *Nanoarchaeum equitans*. *J. Mol. Biol.* 356, 1093–1106.
- [40] Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14, 51–55.
- [41] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.
- [42] Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M. (1992) Stereochemical quality of protein structure coordinates. *Proteins* 12, 345–364.