The NMR structure of the engineered halophilic DnaE intein for segmental isotopic labeling using conditional protein splicing

Heikkinen, Harri August

2022-05


http://hdl.handle.net/10138/344416
https://doi.org/10.1016/j.jmr.2022.107195

cc_by
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
The NMR structure of the engineered halophilic DnaE intein for segmental isotopic labeling using conditional protein splicing

Harri A. Heikkinen 1, A. Sesilja Aranko 2, Hideo Iwai *

Institute of Biotechnology, University of Helsinki, PO Box 65, Helsinki, FIN-00014, Finland

1. Introduction

Modern protein NMR studies typically require enrichment of stable isotopes such as 15N and 13C for triple-resonance NMR spectroscopy to expand frequency dimensions. Despite the expansion into 15N- and 13C-dimensions, NMR resonance assignments increasingly become complex as proteins become larger. Severe NMR signal overlaps hinder efficient NMR analysis of proteins, such as three-dimensional structure determination. Since many proteins are multi-domain proteins, it is also logical to analyze a domain or segment of interest to alleviate the NMR signal overlaps [1]. However, dissection of a domain from the intact full-length proteins could obscure structural features in the full-length context. Segmental isotopic labeling incorporating NMR active or inactive isotopes into a segment or protein could circumvent NMR signal overlaps in full-length proteins, yet enabling conventional triple-resonance NMR techniques [2,3]. Thus, segmental isotopic labeling is a robust labeling scheme for protein NMR spectroscopy and has opened a new horizon in protein NMR [2,3]. Segmental isotopic labeling has been demonstrated using protein trans-splicing (PTS), expressed protein ligation or intein-mediated protein ligation (EPL/IPL), or enzymatic protein ligation using sortase and asparaginyl endopeptidase (Fig. 1) [2–7]. Their applications have been demonstrated in various NMR studies of proteins [8,9,10]. However, additional labor-intensive procedures of segmental isotopic labeling have been a bottleneck. Moreover, segmental isotopic labeling by EPL/IPL and PTS have intrinsic limitations. These include lower tolerances of amino-acid types at the ligation junction, thereby reducing the final yield and/or leaving non-native sequence as a scar [11].

Thus, the utility of segmental isotopic labeling could diminish when modifications of the native protein sequence need to be introduced in the labeled protein. Another issue with protein ligation using PTS is related to the solubilities of split protein fragments and/or split intein fragments indispensable for segmental isotopic labeling [4]. Split polypeptide fragments from one protein, including an intein catalyzing protein splicing, often become insoluble, particularly when forming one globular domain [3,4] (Fig. 2). Therefore, segmental isotopic labeling usually requires refolding of
the target protein to be assembled, thereby complicating segmental isotopic labeling [3,4,12]. In vivo segmental isotopic labeling was developed to circumvent this problem using refolding in living cells but could complicate NMR analysis due to isotopic scrambling during the cell growth [13].

Our goal is to develop efficient ways for conveniently producing segmentally isotopic labeled proteins. We previously demonstrated highly soluble inteins from extremely halophilic inteins to overcome the solubility issue of split intein fragments for protein ligation by PTS as well as intein-mediated protein purification [14,15]. Halophilic inteins from extremely halophilic organisms seem to be generally inactive under low salinity but could be activated by increasing the salt concentration [14–17](Fig. 2). This salt-dependent conditional protein splicing (CPS) provides a way to regulate the protein splicing reaction by adjusting the salt concentration, opening new possibilities for controlled protein ligation (Fig. 2c). However, for efficient protein splicing, natural halophilic inteins such as MCM2 intein from *Halorhabdus utahensis* (*Hut*MCM2) require 3–4 M NaCl, which might not be suitable for some target proteins [14,15]. Therefore, we asked whether it could

Fig. 1. Methods for segmental isotopic labeling of proteins. (a) Expressed protein ligation (EPL)/Intein-mediated protein ligation (IPL) via native chemical ligation (NCL). (b) Protein trans-splicing (PTS) using split inteins. (c) Enzymatic ligation using sortase-mediated ligation (SML) and asparaginyl endopeptidase (AEP) mediated ligation (AML).

Fig. 2. (a) Protein splicing (PS) in cis. (b) Protein splicing in trans by split inteins (PTS). (c) Conditional protein splicing (CPS).
lower the salt concentration needed for activating protein splicing by protein engineering.

Here, we report the conversion of a well-characterized mesophilic intein, the DnaE intein from *Nostoc punctiforme* (*Npu*DnaE), into a halo-obligate intein and its NMR structure under a high salt condition. In addition, we further demonstrated salt-induced protein trans-splicing of the engineered halo-obligate DnaE intein for protein ligation.

2. Results

2.1. Design of the salt-inducible *Npu*DnaE intein, version 1

Proteins from halophilic organisms are generally more acidic than other organisms because of increased acidic amino acids in the proteins [18,19]. The abundance of aspartate (D) and glutamate (E) residues results in a decrease in the solvent-accessible area of the proteins [18,19]. The abundance of aspartate (D) and glutamate (E) residues results in a decrease in the solvent-accessible area of the proteins [18,19].

We chemically synthesized the gene of *Npu*DnaE_DEST and cloned it into a plasmid for the expression of a precursor protein containing *Npu*DnaE_DEST with the N- and C-exteins of GB1. Indeed, *Npu*DnaE_DEST did not splice when the precursor protein was expressed in *E. coli* and could be purified as a precursor protein (Fig. 3c). This result suggests the tertial structure of *Npu*DnaE_DEST was presumably disrupted in the absence of a high concentration of salts in contrast to the first version of *Npu*DnaE_DEST. Next, we examined the salt effects on the purified precursor protein with *Npu*DnaE_DEST (Fig. 3c). Indeed, *Npu*DnaE_DEST was able to splice upon adding >1 M NaCl (Fig. 3c). Thus, we successfully demonstrated the conversion of the well-characterized *Npu*DnaE intein into a halo-obligate intein, which requires a much less salt concentration than inteins from extremely halophilic archaea [14]. We also tested different salts for cis-splicing of *Npu*DnaE_DEST, resulting in similar results as other salt-dependent halophilic inteins (Fig. 3d) [14,16]. This observation suggests that other co-solutes, such as sucrose, could also induce cis-splicing of *Npu*DnaE_DEST at different concentrations. The splicing kinetics was strongly dependent on the salt concentration suggesting the equilibrium between unfolded and folded states. Generally, a higher salt concentration induces a faster cis-splicing, although 4 M NaCl reduced the apparent kinetics probably due to precursor precipitations (Fig. 3e).

2.2. Design of the salt-inducible *Npu*DnaE intein, version 2

Our initial attempt to convert the *Npu*DnaE intein into a halo-obligate intein was unsuccessful. Thus, we revisited the amino-acid composition of two halo-obligate inteins from extremely halophilic archaea (Supplemental Table S1). We noticed that Ser (S) and Thr (T) residues are also abundant among the two halo-obligate inteins as observed for other proteins [19]. Therefore, we introduced Ser and Thr residues into *Npu*DnaEDEST, aiming to destabilize *Npu*DnaEDEST. In addition, we selected seven hydrophobic Val, Ile, or Leu residues with 23–48% solvent-accessible areas for introducing Thr residues (Supplemental Table S2). We anticipated that these mutations on hydrophobic residues by Thr could play critical roles in destabilizing *Npu*DnaE intein.

Furthermore, we replaced the hydrophobic residue of Leu100 in the loop with Ser residue. Other Ser residues were introduced on the surface by replacing Glu or Asp on *Npu*DnaEDEST (Fig. 3a). In total, we introduced 29 mutations on the original cis-splicing *Npu*DnaE intein and termed the newly designed *Npu*DnaE intein, *Npu*DnaE DEST. We expected that these 29 mutations, including partly buried hydrophobic residues, would destabilize *Npu*DnaE intein sufficiently to make it deficient in protein splicing.

Whereas *Hut*MCM2 intein requires 3–4 M NaCl to fold into an active conformation, *Npu*DnaE_DEST requires a lower concentration of NaCl for the splicing activity [14]. Thus, structural characterization of *Npu*DnaE_DEST by NMR spectroscopy has become easier because a high salt concentration drastically reduces NMR sensitivity due to the ionic conductivity and dielectric losses, especially with cryogenic probes [31,32]. Thus, we aimed to lower the salt concentration required for the activity of a halo-obligate intein to enable our NMR studies. First, we performed resonance assignments and determined the structure of *Npu*DnaE_DEST using NMR spectroscopy. The [1H, 15N]-HSQC spectrum of the C1A variant of *Npu*DnaE_DEST without any salt showed a typical [1H, 15N]-HSQC spectrum for an unfolded protein with highly overlapped peaks having 1H chemical shift between 7.8 and 8.5 ppm (Fig. 4a and Supplemental Figure S1). We also recorded [1H, 15N]-HSQC spectrum of *Npu*DnaE_DEST(C1A) in the presence of 2 M NaCl, which displayed well-dispersed [1H, 15N] correlation peaks and almost no visible minor unfolded conformation (Fig. 4b; Supplemental Figure S2). The NMR spectrum suggested that 2 M NaCl was sufficient for the folding of the splicing-active three-dimensional structure. We thus determined the NMR structure of the C1A variant of *Npu*DnaE_DEST in the presence of 2 M NaCl (Table 1). The *Npu*DnaE_DEST structure has the common HINT(Hedgehog/INTein) fold similar to the original *Npu*DnaE intein (Fig. 4c) [24]. The r.m.s.d. between *Npu*DnaE_DEST and the single-chain variant of *Npu*DnaE intein for the backbone atoms of residues 1–137 was 0.9 Å (Fig. 4d), confirming that the designed *Npu*DnaE_DEST was successful in retaining the functional structure as we desired [24]. The most notable deviations were observed for the loop where the natural split *Npu*DnaE intein was connected to make a single-chain variant of *Npu*DnaE intein (Fig. 4d) [24].
2.4. Engineering of split NpuDnAE_DEST

We demonstrated salt-inducible cis-splicing of NpuDnAE_DEST (Fig. 3c and 3d). Next, we tested whether NpuDnAE_DEST could also be used for trans-splicing, i.e., salt-inducible trans-splicing for protein ligation. NpuDnAE_DEST was split into the N- and C-terminal fragments (IntN and IntC) at the naturally split-site of NpuDnAE intein (Fig. 5a and 5b) [23]. We used the N- and C-terminally His-tagged GB1 as N- and C-extein, respectively (Fig. 5c). Two N- and C-precursors were independently expressed and purified. Trans-splicing using the split NpuDnAE_DEST was tested in vitro by mixing the two precursors and incubating overnight in the absence or presence of 2 M NaCl (Fig. 5c). While no trans-splicing reaction was observed without any salt, 2 M NaCl induced trans-splicing reaction producing the ligated product (H6G–GH6) (Fig. 5c). Thus, the split version of NpuDnAE_DEST was indeed capable of protein splicing in trans.

We next tested the orthogonality between the split NpuDnAE_DEST and NpuDnAE intein in which different combinations of IntN and IntC between NpuDnAE_DEST and NpuDnAE were tested (Fig. 3).
Among the four possible combinations, only the pair of IntN and IntC from NpuDnaE_DEST did not react when the two precursors were co-expressed in vivo (Fig. 5d). We observed trans-splicing of the other combinations of the split inteins in which one of the two split intein fragments was the wild-type fragment from NpuDnaE intein. Only the combination containing the designed 29 mutations was inactive, suggesting that other combinations form the functional structure. The orthogonality test confirmed that all the 29 mutations were indispensable for the salt dependence and supports that the high stability of NpuDnaE intein makes it highly tolerant of sequence variations.

**3. Discussion**

Proteins exert their biological functions via their three-dimensional structures with their specific dynamics. It is known that solvent additives such as co-solvents modulate their biochemical activities by interacting with proteins as well as water molecules [34,36]. In other words, co-solvents or additives could control the protein functions in response to the solvent environment [34]. Chaotropic agents such as denaturants typically inactivate proteins by disrupting their three-dimensional structures. In the case of proteins from extremely halophilic organisms, the absence of a high salinity condition could make proteins unstructured and inactive without any chaotropic agents (Fig. 4a) [14,16,35]. Interestingly, urea-denatured proteins could also be refolded into a native conformation by co-solutes, such as high concentrations of various salts [37,38]. Both urea-unfolded and halo-obligate unfolded proteins are highly soluble without any aggregation [14,37]. The effects of co-solutes on both halo-obligate and urea-denatured proteins suggest that interactions with co-solutes could shift protein folding/unfolding equilibrium effectively, presumably due to preferential hydration or preferential co-solvents binding [34]. Negatively charged halophilic proteins are often considered to bind water molecules stronger and protect the structure against unfolding and aggregation by high salts with the hydration shell [22]. However, NMR studies do not support more substantial hydration around carboxyl groups common for halophilic proteins [39,40].

Because some inteins in nature act as regulatory sensors responding to environmental conditions [41], controlling protein-splicing has attracted various chemical and biotechnological applications, including segmental isotopic labeling [14,15]. We successfully demonstrated the conversion of a mesophilic intein to a halo-obligate salt-inducible intein by rationally designed mutations. Thus, we created a controllable NpuDnaE intein capable of protein splicing in cis and trans by adjusting the salt concentration.
Whereas simple charge replacements by introducing Asp and Glu on the protein surface were insufficient to make it salt-dependent, additional mutations on partially buried hydrophobic residues disrupted the folded conformation and made it salt-sensitive (Fig. 3). Cross-activities of the engineered split intein fragments indicated the requirement of all the mutations to be salt-dependent, suggesting the subtle balance of the free energy differences among the four combinations (Fig. 5e). Not only ionic salts but other additives such as sugars could activate halo-obligate inteins (Fig. 3d), suggesting the macromolecular crowding environment, such as cellular environments, could also affect the folding/unfolding equilibrium [16]. The three-dimensional structure of halo-obligate proteins under high salinity is not a mere result of various intramolecular interactions but the subtle energy balance between the different complex interactions, including the solvation, which involves waters, ions, co-solutes, and the polypeptide chains [21,22]. It is, therefore, unlikely that the current structural prediction algorithm like AlphaFold2 could predict the active or inactive conformation of proteins [42].

NMR spectroscopy can provide high-resolution three-dimensional structures of proteins under various solvent milieu [38], including cellular milieu [43], and is a powerful tool for investigating protein hydration [36,44]. Further NMR studies of solvent-protein interactions, including hydration shells, could shed light on understanding the halo-adaptation mechanism, protein aggregation, and protein solubility. Furthermore, the molecular interactions of proteins with solvent and co-solvents could play a critical role in developing novel biotechnological tools such as environmental and molecular sensors. Together with protein engineering to overcome technical difficulties in NMR [2–7,27], NMR will provide an almost unlimited range of protein studies for various proteins, including structured and disordered proteins.

4. Materials and methods

4.1. Constructions of plasmids for protein expression

The gene of NpuDnaE intein with ED mutations (NpuDnaE_DE) was purchased from Integrated DNA Technologies, BVBA (Leuven, Belgium) as plasmid pIDTSMART-KAN-Genesyn11 and cloned into pSKDuet16 (addgene #41684) using BamHI and KpnI restriction sites, resulting in pJDuet107 [26]. The gene of NpuDnaE intein with DE and ST mutations (NpuDnaE_DEST) was chemically synthesized and purchased from Integrated DNA Technologies as pIDTSMART-KAN-Genesyn21. The plasmid encoding the cis-splicing precursor protein with two GB1 domains as the exteins was constructed by cloning the NpuDnaE_DEST gene from pIDTSMART-KAN-Genesyn21 into pSKDuet16 using BamHI and KpnI, resulting in pBDuet134. For NMR structural analysis, the gene of NpuDnaE_DEST with C1A mutation and a stop codon after the intein sequence was cloned into pHYR5F53 (addgene #64696) [45] using two oligonucleotides of IT54: 5'-GTGGATCCGGAGGACCTTAAAGCTATGACACGGAAAA and SK187: 5'-ATCAAGCTTAATTAGAAGCTATGAAAA and cloned into pMHBAD14 (addgene #42304) [24], resulting in pSABAD824.

4.2. In vitro protein cis-splicing assays of NpuDnaE_DEST

Protein cis-splicing of NpuDnaE_DEST was tested by the purified precursor with NpuDnaE_DEST flanked by two B1 domains of the IgG-binding protein produced using the plasmid pBDuet134 and purified using immobilized metal affinity chromatography as described elsewhere [14,23,30]. The experiments were performed at 25 °C with a final concentration of 25 μM precursors in 50 mM Tris HCl, pH 7.4, together with a different final concentration of the salts. The samples were taken at the time points of 0 min, 3 min, 10 min, 30 min, 1 h, 3 h, 6 h, and 24 h for the kinetic analysis. Protein splicing was analyzed by SDS-PAGEs using 18% gels and Coomassie Blue staining.

4.3. Testing cross-activities between NpuDnaE and NpuDnaE_DEST

For testing the cross-activities between the wild-type split NpuDnaE intein and the engineered split NpuDnaE_DEST, two N- and C-precursors with the split NpuDnaE intein in the plasmids pMHBAD14 (addgene #42304) and pSKDuet01 (addgene #12172) were used. The plasmids of pSADuet825 and pSABAD824 from the split NpuDnaE_DEST were used for testing the cross-activities. The four combinations of pairs of two plasmids were transformed into E. coli strain T7 express (New England Biolabs) in 5 ml LB medium supplemented with 100 μg/ml ampicillin and
The cells were grown at 37 ºC and induced with a final concentration of 0.08% (w/v) arabinose and 0.5 mM IPTG when the OD 600 reached 0.5–0.6. The two precursor proteins were induced for four hours. The harvested cells were lysed with B-PER cell lysis buffer (Thermo fisher scientific) and purified using Ni-NTA spin columns (Qiagen). The total cell lysate and elution from the Ni-NTA column were analyzed by 18% SDS-PAGEs. For in vitro trans-splicing tests, the N- and C-precursors of split NpuDnaE_DEST were expressed using pSADuet825 or pSABAD824, respectively, and purified with Ni-NTA spin columns separately. The N- and C-precursors mixtures were incubated at 25 ºC for 19 h either in 0.5 mM TCEP, 2 M NaCl, or 0 M NaCl. The reaction mixture was precipitated by Trichloroacetic acid (TCA) to remove the salts and analyzed by 18% SDS-PAGEs.

### 4.4. Sample preparation of NpuDnaE_DEST (C1A) for NMR

[100% 13C, 100% 15N]- and [20% 13C, 100% 15N]-labeled NpuDnaE_DEST (C1A) was produced as His-tagged SUMO fusion using plasmid pBHRSF137 as previously described [45,46]. The purified
protein was dialyzed and concentrated into a 1.9 mM in 2 M NaCl, 20 mM sodium phosphate buffer, pH 6 or 1.6 mM in 20 mM sodium phosphate buffer, pH 6.0.

4.5. NMR spectroscopy and NMR structure determination

For the structure-determination of NpuDnaE_DEST in 2 M NaCl, the following 2D and 3D experiments were used: \(^{1}H,^{15}N\)-HSQC, BEST-HNCO, BEST-HNCA, BEST-HNCCCA, BEST-HNCCOA, CBCA(CO)NH, CC(CO)NH, HBBACA(CO)NH, HCC(CO)NH, \(^{15}N\)-edited \(^{1}H,^{1}H\)-TOCSY, and NHNA [47,48]. \(^{1}H\) and \(^{13}C\) assignments for aliphatic side-chain were based on \(^{1}H,^{13}C\)-HSQC, HCC-COSY, TCH-TOCSY, and -TCH-NOESY. The structures were visualized with MOLMOL [53].

Acknowledgments

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References


