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

Heikkinen, Harri August

2022-05

Heikkinen , H A , Aranko , S & Iwaï , H 2022 , ' The NMR structure of the engineered halophilic DnaE intein for segmental isotopic labeling using conditional protein splicing ' , Journal of Magnetic Resonance , vol. 338 , 107195 . https://doi.org/10.1016/j.jmr.2022.107195

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

Journal of Magnetic Resonance 338 (2022) 107195

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

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 Iwaï *

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

ARTICLE INFO

Article history:
Received 21 December 2021
Revised 2 March 2022
Accepted 13 March 2022
Available online 18 March 2022

Keywords:
NMR structure
Protein splicing
Halophilic inteins
Segmental isotopic labeling
Conditional protein splicing
Protein engineering

ABSTRACT

Protein *trans*-splicing catalyzed by split inteins has been used for segmental isotopic labeling of proteins for alleviating the complexity of NMR signals. Whereas inteins spontaneously trigger protein splicing upon protein folding, inteins from extremely halophilic organisms require a high salinity condition to induce protein splicing. We designed and created a salt-inducible intein from the widely used DnaE intein from *Nostoc punctiforme* by introducing 29 mutations, which required a lower salt concentration than naturally occurring halo-obligate inteins. We determined the NMR solution structure of the engineered salt-inducible DnaE intein in 2 M NaCl, showing the essentially identical three-dimensional structure to the original one, albeit it unfolds without salts. The NMR structure of a halo-obligate intein under high salinity suggests that the stabilization of the active folded conformation is not a mere result of various intramolecular interactions but the subtle energy balance from the complex interactions, including the solvation energy, which involve waters, ions, co-solutes, and protein polypeptide chains.

1. Introduction

Modern protein NMR studies typically require enrichment of stable isotopes such as ^{15}N and ^{13}C for triple-resonance NMR spectroscopy to expand frequency dimensions. Despite the expansion into ^{15}N- and ^{13}C-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

Abbreviations: PTS, protein *trans*-splicing; EPL, expressed protein ligation; IPL, intein-mediated protein ligation; SML, sortase mediated ligation; AEP, asparaginyl endopeptidase; AML, asparaginyl endopeptidase mediated ligation; CPS, conditional protein splicing; NCL, native chemical ligation; *Hut*, *Halorhabdusutahensis*; *Npu*, *Nostoc punctiforme*; GB1, the B1 domain of IgG binding protein A.

* Corresponding author.
E-mail addresses: sesilja.aranko@aalto.fi (A.S. Aranko), hideo.iwai@helsinki.fi, iwai@alumni.ethz.ch (H. Iwaï).
[1] Present address: VERIFIN, Department of Chemistry, University of Helsinki, FIN-00014 Helsinki, Finland.
[2] Current Affiliation: School of Chemical Technology, Aalto University, P.O. Box 6100, 00076 Espoo, Finland.

https://doi.org/10.1016/j.jmr.2022.107195
1090-7807/

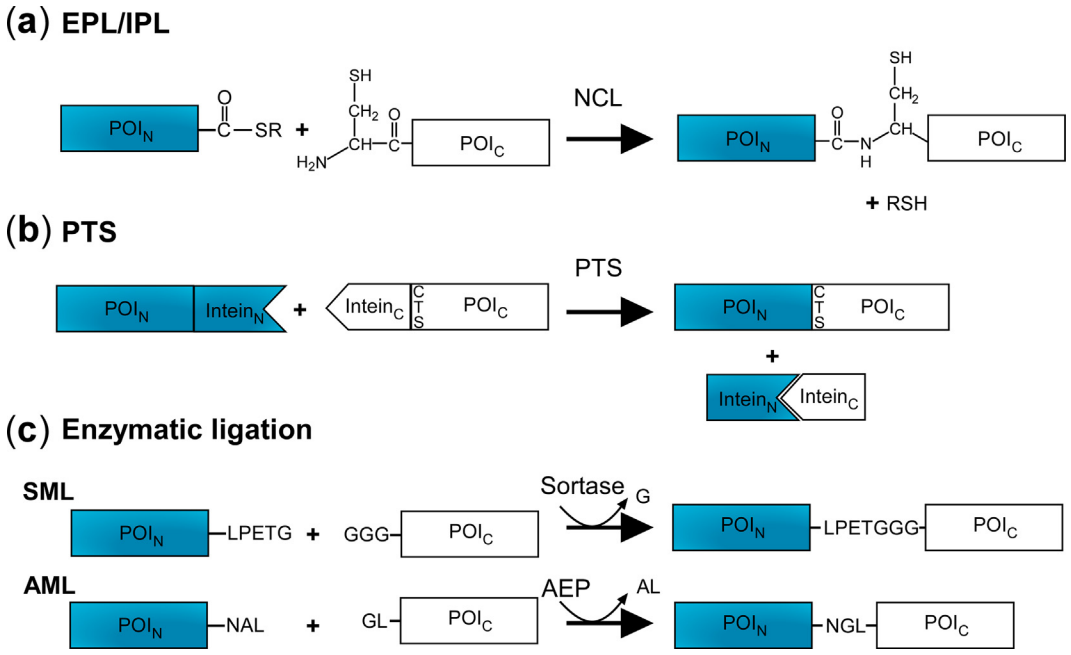

Fig. 1. Methods for segmental isotopic labeling of proteins. (**a**) Expressed protein ligation (EPL)/Intein-mediated protein ligation (IPL) via native chemical ligation (NCL). (**b**) Protein *trans*-splicing (PTS) using split inteins. (**c**) Enzymatic ligation using sortase-mediated ligation (SML)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

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

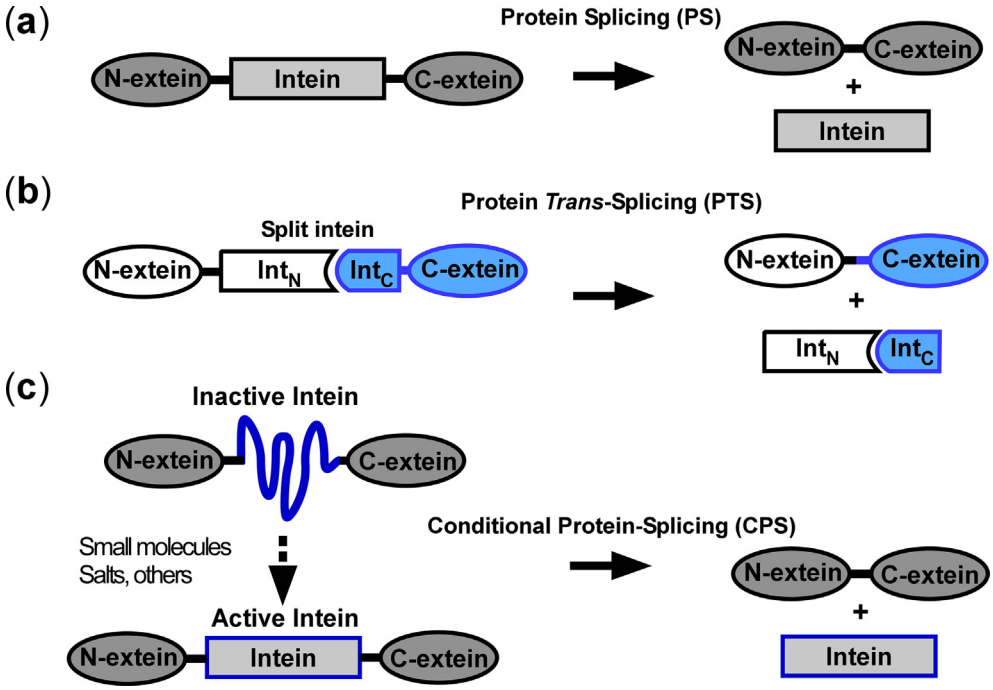

Fig. 2. (**a**) Protein splicing (PS) in *cis*. (**b**) Protein splicing in *trans* by split inteins (PTS). (**c**) Conditional protein splicing (CPS).

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 NpuDnaE 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 proteins, which is seemingly the primary mechanism for halo-adaptation [20,22]. Millet *et al*. increased Glu (E) and Asp (D) residues in the mesophilic IgG-binding domain of the protein L from *Streptococcus magnus* (ProtL) [20]. The successful conversion from a mesophilic protein to a halo-obligate protein has thus been previously demonstrated using ProL [20]. We applied a similar approach to introduce Asp (D) and Glu (E) on a single-chain variant of the highly robust split *Npu*DnaE intein [23,24]. *Npu*DnaE intein possesses high protein *trans*-splicing activity as well as tolerates many amino-acid types at the splicing junctions, making this intein very attractive for various biotechnological applications [23,25]. First, we introduced size-preserving mutations from Asn (N) to Asp (D) and from Gln (Q) to Glu (E) on the surface residues of the *Npu*DnaE intein [24].

Additionally, we replaced positive residues of Lys (K) and Arg (R) with Asp (D) or Glu (E) on the surface of *Npu*DnaE intein (Fig. 3a, Table S1). In total, we introduced 15 mutations on *Npu*DnaE intein and termed *Npu*DnaE_DE (Fig. 3). We tested the *cis*-splicing of the designed *Npu*DnaE_DE intein using the B1 domain of IgG binding protein A (GB1) as the extein, as previously described (Fig. 3b) [26]. To our surprise, these 15 mutations on the surface of *Npu*DnaE intein did not reduce the *cis*-splicing activity of *Npu*DnaE intein, producing the *cis*-spliced product and excised intein immediately after the protein expression (Fig. 3b). Proteins seem to tolerate massive mutations on the protein surface [27,28]. This result indicated that *Npu*DnaE intein was stable enough to accommodate these mutations on the surface and retained both active structure and protein-splicing function. The reported thermal unfolding midpoint of 84 °C for *Npu*DnaE intein supports the large free energy difference (ΔG) between the unfolded and folded states of *Npu*DnaE intein [29].

2.2. Design of the salt-inducible NpuDnaE 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*DnaE_DE, aiming to destabilize *Npu*DnaE_DE. 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*DnaE_DE (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.

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_DE. 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). The protein splicing of the engineered salt-inducible intein was slower than the original mesophilic intein but similar to other natural halophilic inteins [14,16].

2.3. NMR analysis of NpuDnaE_DEST

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 their NMR studies. First, we performed resonance assignments and determined the structure of *Npu*DnaE_DEST using NMR spectroscopy. The [^{1}H, ^{15}N]-HSQC spectrum of the C1A variant of *Npu*DnaE_DEST without any salt showed a typical [^{1}H, ^{15}N]-HSQC spectrum for an unfolded protein with highly overlapped peaks having ^{1}H chemical shift between 7.8 and 8.5 ppm (Fig. 4a and Supplemental Figure S1). We also recorded [^{1}H, ^{15}N]-HSQC spectrum of *Npu*DnaE_DEST(C1A) in the presence of 2 M NaCl, which displayed well-dispersed ^{1}H-^{15}N correlation peaks and almost no visible minor unfolded conformation (Fig. 4b; Supplemental Figure S2). The NMR spectrum suggested that 2 M NaCl induced the formation of the splicing-active three-dimensional structure. We thus determined the NMR structure of the C1A variant of *Npu*DnaE_DEST intein 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].

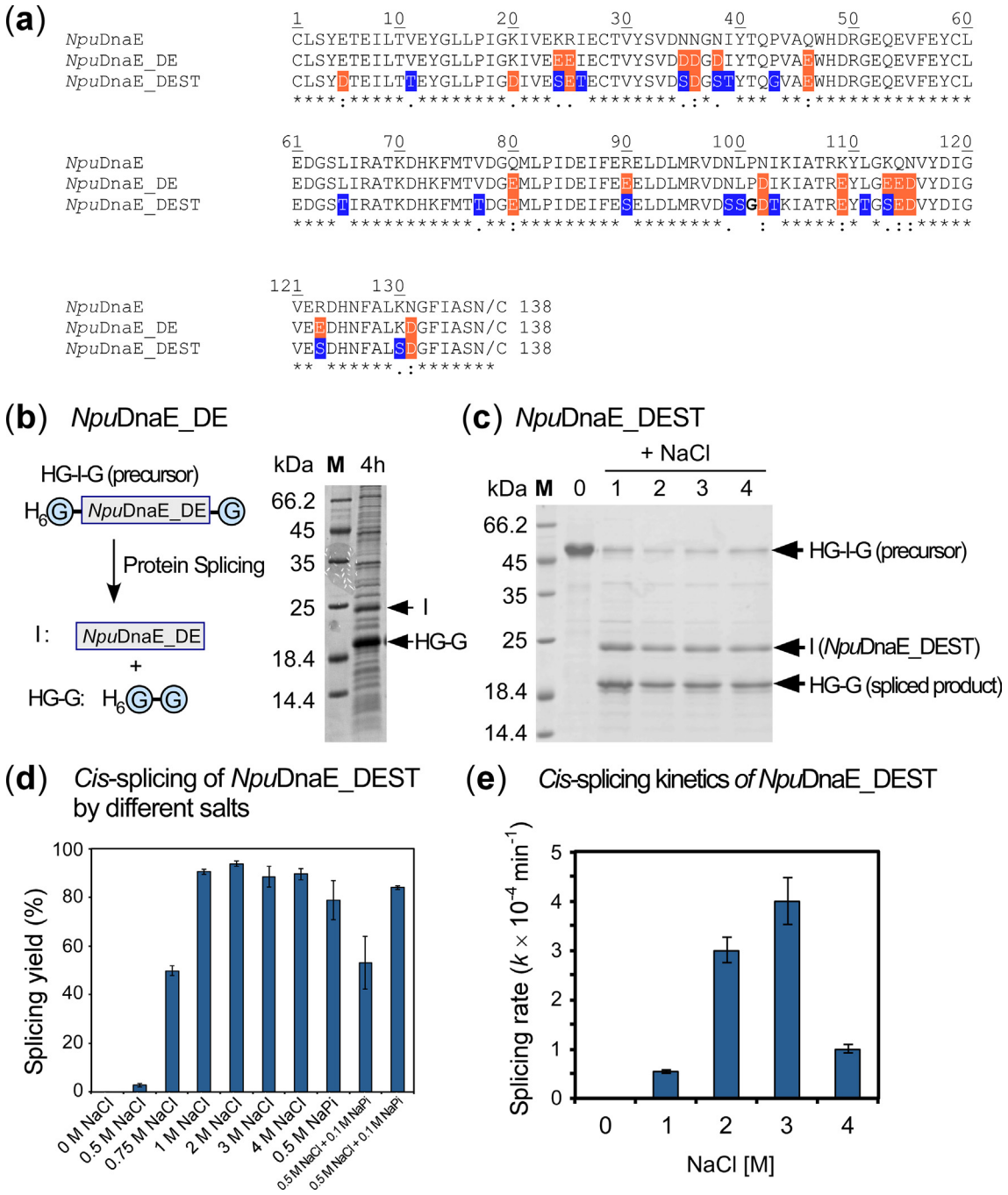

Fig. 3. Design and a *cis*-splicing assay of engineered *Npu*DnaE intein. (**a**) Primary structures of designed halophilic inteins aligned to the original *Npu*DnaE intein. Mutations by Glu (E) and Asp (D) are highlighted in orange. Additional mutations by Ser (S) and Thr (T) are highlighted in blue. (**b**) *In vivo cis*-splicing of *Npu*DnaE_DE using the two GB1 domains (G) as N- and C-exteins. Total cell lysate after 4-hour induction was analyzed by SDS-PAGE. (**c**) *Cis*-splicing of *Npu*DnaE_DEST (see the text) under different salt concentrations was analyzed by SDS-PAGE. The bands corresponding to the precursor (HG-I-G), excised intein (I), and the spliced product (HG-G) were indicated by arrows. (**d**) *Cis*-splicing yields of *Npu*DnaE_DEST under different salt conditions after 24-hour incubation. (**e**) *Cis*-splicing rates of *Npu*DnaE_DEST under different concentrations of NaCl.

2.4. Engineering of split NpuDnaE_DEST

We demonstrated salt-inducible *cis*-splicing of *Npu*DnaE_DEST (Fig. 3c and 3d). Next, we tested whether *Npu*DnaE_DEST could also be used for *trans*-splicing, i.e., salt-inducible *trans*-splicing for protein ligation. *Npu*DnaE_DEST was split into the N- and C-terminal fragments (Int^N and Int^C) at the naturally split-site of *Npu*DnaE 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 *Npu*DnaE_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 (H_6G–GH_6) (Fig. 5c). Thus, the split version of *Npu*DnaE_DEST was indeed capable of protein splicing in *trans*.

We next tested the orthogonality between the split *Npu*DnaE_DEST and *Npu*DnaE intein in which different combinations of Int^N and Int^C between *Npu*DnaE_DEST and *Npu*DnaE were tested

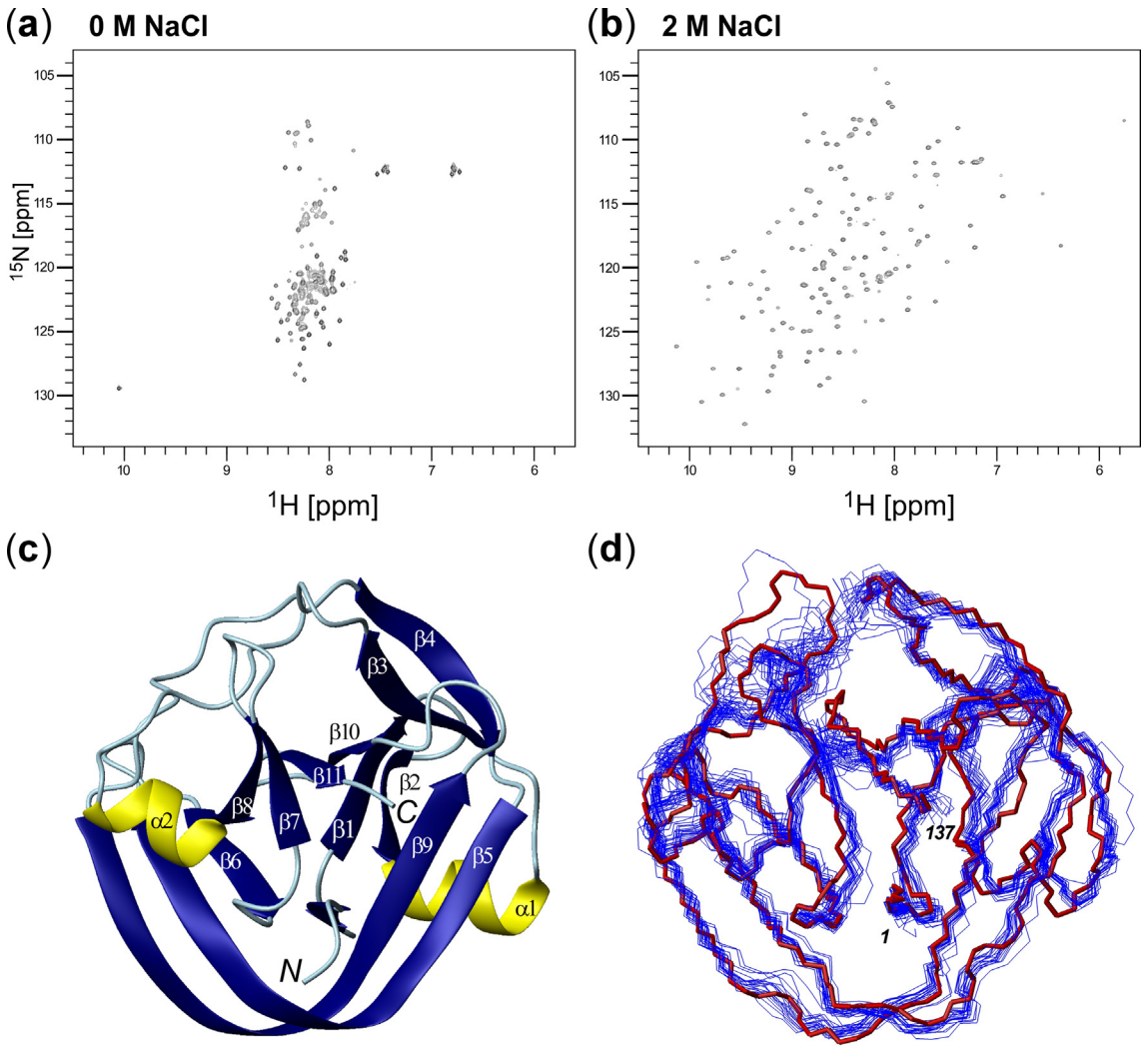

Fig. 4. NMR analysis of *Npu*DnaE_DEST. (**a**) [^{1}H, ^{15}N]-HSQC spectrum of *Npu*DnaE_DEST without any salt and (**b**) with 2 M NaCl. (**c**) A cartoon drawing of the NMR structure of *Npu*DnaE_DEST (C1A). (**d**) A superposition of the 20 NMR conformers of *Npu*DnaE_DEST and *Npu*DnaE (PDB: 2keq) [24].

(Fig. 5d). Among the four possible combinations, only the pair of Int^{N} and Int^{C} from *Npu*DnaE_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 *Npu*DnaE 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 *Npu*DnaE 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 *Npu*DnaE intein capable of protein splicing in *cis* and *trans* by adjusting the salt concentration.

Table 1
Structural statistics for the 20 energy-minimized NMR conformers of *Npu*DnaE_DEST.

Completeness of resonance assignments (%) [a]	
Backbone	96.3
Side chain	91.6
Aromatic	56.3
Distance restraints	
Total	1760
Sequential ($\|i-j\| \leq 1$)	864
Medium range ($1 < \|i-j\| < 5$)	182
Long range ($\|i-j\| \geq 5$)	714
No. of restraints per residue	12.6
No. of long-range restraints per residue	5.1
Residual restraint violations	
Average no. of distance violation per structure	
0.1-0.2 Å	3.7
0.2-0.5 Å	0.1 (max 0.27)
>0.5 Å	0
Average no. of dihedral angle violations per structure >5°	0
Model quality [b]	
Rmsd backbone atoms (Å)	0.7
Rmsd heavy atoms (Å)	1.1
Rmsd bond lengths (Å)	0.014
Rmsd bond angles (°)	2.2
MolProbity Ramachandran statistics [b]	
Most favored regions (%)	95.5
Allowed regions (%)	4.3
Disallowed regions (%)	0.2
Global quality scores (raw/Z score) [b]	
Verify3D	0.47/0.16
ProsaII	0.69/0.17
PROCHECK (ϕ-ψ)	−0.59/−2.01
PROCHECK (all)	−0.57/−3.37
MolProbity clash score	2.08/1.17
Model contents	
Ordered residue ranges	4–98, 107–134, 136–139.
PDB ID (BMRB ID)	7QIL (34695)

[a] Calculated from the expected number of resonances, excluding highly exchangeable protons (N-terminal, Lys, amino and Arg guanidino groups, hydroxyls of Ser, Thr, and Tyr), carboxyl groups of Asp and Glu, and non-protonated aromatic carbons. Backbone: H^N, N^H, C^α, C^β, H^α, C′.
[b] Calculated using PSVS version 1.5 [33].

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 *Npu*DnaE intein with ED mutations (*Npu*DnaE_DE) was purchased from Integrated DNA Technologies, BVBA (Leuven, Belgium) as plasmid pIDTSMART-KAN-GeneSyn11 and cloned into pSKDuet16 (addgene #41684) using *Bam*HI and *Kpn*I restriction sites, resulting in pJODuet107 [26]. The gene of *Npu*DnaE intein with DE and ST mutations (*Npu*DnaE_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 *Npu*DnaE_DEST gene from pIDTSMART-KAN-GeneSyn21 into pSKDuet16 using *Bam*HI and *Kpn*I, resulting in pBHDuet134. For NMR structural analysis, the gene of *Npu*DnaE_DEST with C1Amutation and a stop codon after the intein sequence was cloned into pHYRSF53 (addgene #64696) [45] using two oligonucleotides of I754: 5′-GTGGATCCG GAGGAGCTTTAAGCTATGACACGGAAA and SK187: 5′-ATCAAGCT TAATTAGAAGCTATGAAGCC. The resulted plasmid was named pBHRSF137.

The N-terminal split fragment (Int^N) of *Npu*DnaE_DEST was amplified from pBHDuet134 by PCR using the two oligonucleotides, SK092: 5′-ACGGATCCTGTTTAAGCTATGAAACGGAAA TATTG, I766: 5′-ATGAAGCTTAATCACCGCTACTATCAACC and cloned into pSKDuet16 (addgene #41684), resulting in pSADuet825 [26]. The C-terminal split fragment (Int^C) of *Npu*DnaE_DEST was amplified buy PCR #I767: 5′-AACATATGACTAAAATAGCCACACGTGAATA TAC: HK983: 5′-GGAATTCACCTTCCGTTACGGTGTAGGTT 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 *Npu*DnaE_DEST was tested by the purified precursor with *Npu*DnaE_DEST flanked by two B1 domains of the IgG-binding protein produced using the plasmid pBHDuet134 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 *Npu*DnaE intein and the engineered split *Npu*DnaE_DEST, two N- and C-precursors with the split *Npu*DnaE intein in the plasmids pMHBAD14 (addgene #42304) and pSKDuet01 (addgene #12172) were used. The plasmids of pSADuet825 and pSABAD824 from the split *Npu*DnaE_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

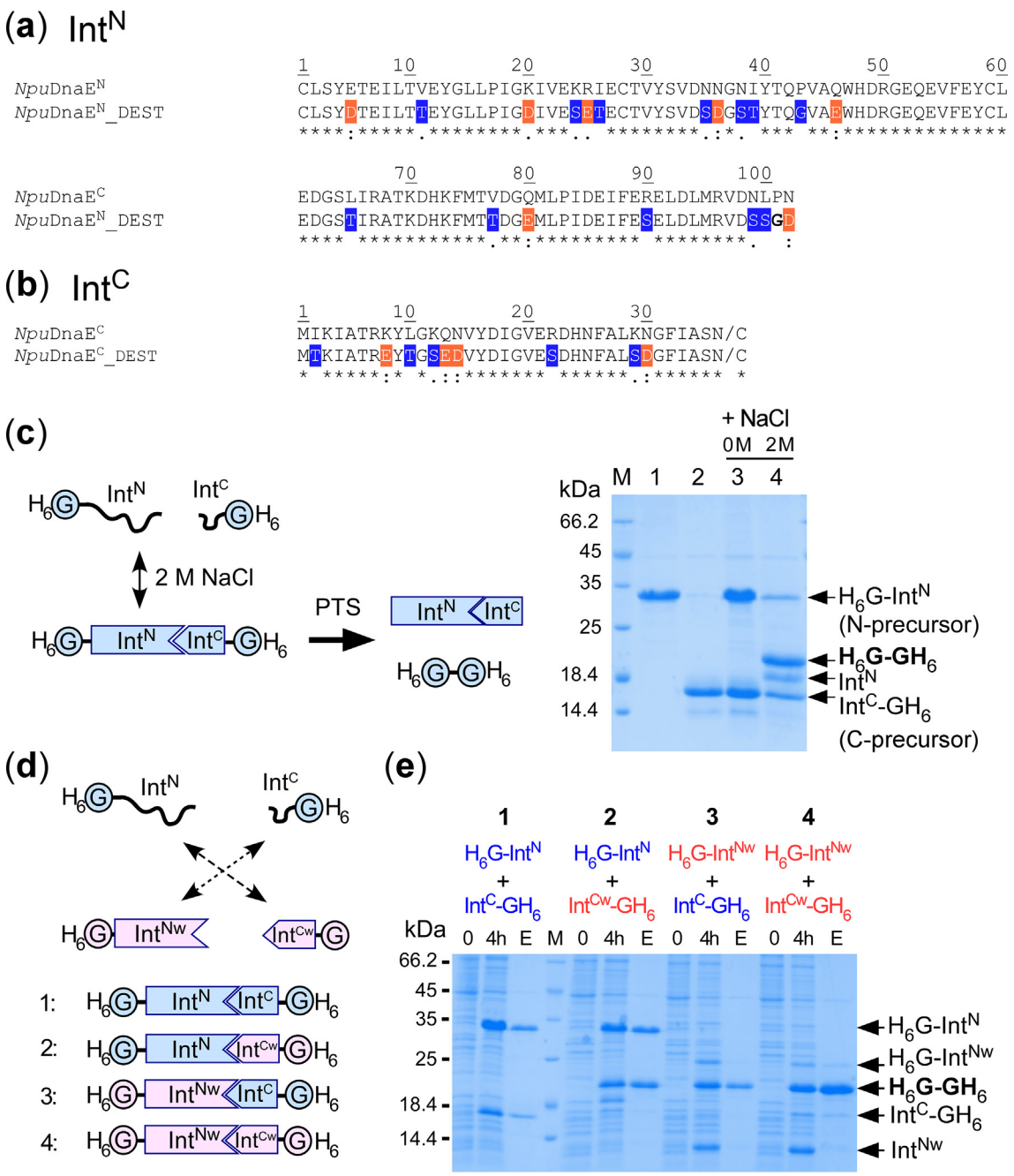

Fig. 5. Split *Npu*DnaE_DEST for *trans*-splicing (**a**) The N-terminal split fragment (Int^N) and (**b**) the C-terminal split fragment (Int^C) of *Npu*DnaE_DEST aligned to the wild-type split *Npu*DnaE intein (*Npu*DnaE^N and *Npu*DnaE^C). Mutations by Glu (E) and Asp (D) are highlighted in orange. Additional mutations by Ser (S) and Thr (T) are highlighted in blue. (**c**) Schematic drawing of salt-inducible *trans*-splicing and the SDS-PAGE analysis of *trans*-splicing by the split *Npu*DnaE_DEST. Lane 1, the N-terminal precursor; lane 2, the C-terminal precursor; lane 3, a reaction mixture of the N- and C-precursors after overnight incubation without any salt; lane 4, a reaction mixture after overnight incubation of the N- and C-precursors in the presence of 2 M NaCl. (**d**) Schematic drawings of the four combinations between split *Npu*DnaE_DEST and *Npu*DnaE inteins. (**e**) SDS-PAGE analysis of the cross-activities *in vivo* between the two split inteins. The N- and C-precursors were co-expressed by the induction with arabinose and IPTG. 0, 4 h, and E indicate samples before induction, after 4-hour protein induction, and the elution fraction from Ni-NTA spin columns, respectively. H_6G-GH_6 is the spliced product by *trans*-splicing.

25 μg/ml kanamycin [13,23]. The cells were grown at 37 ℃ 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 *Npu*DnaE_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 ℃ 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% ^{13}C, 100%^{15}N]- and [20% ^{13}C, 100% ^{15}N]-labeled *Npu*DnaE_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 *Npu*DnaE_DEST in 2 M NaCl, the following 2D and 3D experiments were used: [^{1}H, ^{15}N]-HSQC, BEST-HNCO, BEST-HNCA, BEST-HNCACB, BEST-HNCACO, BEST-HNCOCA, CBCA(CO)NH, CC(CO)NH, HBHA(CO)NH, HCC(CO)NH, ^{15}N-edited [^{1}H, ^{1}H]-TOCSY, and HNHA [47,48]. ^{1}H and ^{13}C assignments for aliphatic side-chain were based on [^{1}H, ^{13}C]-HSQC, HCCH-COSY, HCCH-TOCSY, ct-[^{1}H, ^{13}C]-HSQC, ^{13}C-edited [^{1}H, ^{1}H]-NOESY, and ^{15}N-edited [^{1}H, ^{1}H]-NOESY. For backbone resonance assignments of *Npu*DnaE_DEST without 2 M NaCl were performed based on the following 2D and 3D experiments: [^{1}H, ^{15}N]-HSQC, [^{1}H, ^{13}C]-HSQC, HNCO, HNCA, HNCACB, HNCACO, CBCA(CO)NH, CC(CO)NH, HCC(CO)NH, and intra-HNCA [47]. The spectra were processed with TopSpin 3.2 and analyzed using CcpNmr Analysis 2.4.2 software [49]. Three-dimensional NMR structures were calculated using CYANA 3.97 software, based on the automated NOE analysis algorithm [50,51]. Upper distance restraints were derived from the 3D ^{15}N- and ^{13}C-edited [^{1}H, ^{1}H]-NOESY spectra with a 60-ms mixing time. No additional hydrogen bond restraints were used. All NMR spectra were recorded at 303 K on a Bruker Avance III HD NMR spectrometer equipped with a cryogenically cooled TCI probe head at the ^{1}H frequency of 850 MHz using 3 mm or 5 mm NMR tubes. The final energy minimization in explicit waters was performed for the 20 best CYANA conformers with the lowest CYANA target function using AMBER14 [52]. The structures were validated with PSVS 1.5. [33]. The structures were visualized with MOLMOL [53].

Funding

This work was funded in part by TEKES (project # 1311/31/2014), the Academy of Finland (decision numbers: 131413, 137995, 277335), TaNeDS Europe 2017 grant program from Daiichi Sankyo Co., Ltd, and Novo Nordisk Foundation (NNF17OC0027550). HiLIFE-INFRA supported the NMR facility.

Declaration of Competing Interest

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.

Acknowledgments

This article was inspired by the work of Prof. Kurt Wüthrich, who provided incomparable training while H.I. was a Ph.D. student in his stimulating and exceptional research group at ETH Zürich. We thank B. Haas, S. Ferkau, and Dr. J. S. Oeemig for technical help in protein and plasmid preparations. The Finnish Biological NMR Center is supported by Biocenter Finland and HiLIFE-INFRA.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmr.2022.107195.

References

[1] C. Vogel, M. Bashton, N.D. Kerrison, C. Chothia, S.A. Teichmann, Structure, function and evolution of multi-domain proteins, Curr. Opin. Struct. Biol. 14 (2) (2004) 208–216, https://doi.org/10.1016/j.sbi.2004.03.011.

[2] R. Xu, B. Ayers, D. Cowburn, T.W. Muir, Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies, Proc. Natl. Acad. Sci. USA 96 (2) (1999) 388–393, https://doi.org/10.1073/pnas.96.2.388.

[3] T. Otomo, N. Ito, Y. Kyogoku, T. Yamazaki, NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation, Biochemistry. 38 (49) (1999) 16040–16044, https://doi.org/10.1021/bi991902j.

[4] M. Muona, A.S. Aranko, V. Raulinaitis, H. Iwaï, Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro, Nat. Protoc. 5 (3) (2010) 574–587, https://doi.org/10.1038/nprot.2009.240.

[5] M.A. Refaei, A. Combs, D.J. Kojetin, et al., Observing selected domains in multi-domain proteins via sortase-mediated ligation and NMR spectroscopy, J. Biomol. NMR. 49 (1) (2011) 3–7, https://doi.org/10.1007/s10858-010-9464-2.

[6] L. Freiburger, M. Sonntag, J. Hennig, J. Li, P. Zou, M. Sattler, Efficient segmental isotope labeling of multi-domain proteins using Sortase A, J. Biomol. NMR. 63 (1) (2015) 1–8, https://doi.org/10.1007/s10858-015-9981-0.

[7] K.M. Mikula, I. Tascón, J.J. Tommila, H. Iwaï, Segmental isotopic labeling of a single-domain globular protein without any refolding step by an asparaginyl endopeptidase, FEBS Lett. 591 (9) (2017) 1285–1294, https://doi.org/10.1002/1873-3468.12640.

[8] K.K. Frederick, V.K. Michaelis, M.A. Caporini, et al., Combining DNP NMR with segmental and specific labeling to study a yeast prion protein strain that is not parallel in-register, Proc. Natl. Acad. Sci. USA 114 (14) (2017) 3642–3647, https://doi.org/10.1073/pnas.1619051114.

[9] Y. Shiraishi, M. Natsume, Y. Kofuku, et al., Phosphorylation-induced conformation of β2-adrenoceptor related to arrestin recruitment revealed by NMR, Nat. Commun. 9 (1) (2018) 194, https://doi.org/10.1038/s41467-017-02632-8.

[10] N. Chu, T. Viennet, H. Bae, et al., The structural determinants of PH domain-mediated regulation of Akt revealed by segmental labeling, Elife 9 (2020) e59151, https://doi.org/10.7554/eLife.59151.

[11] Y. Minato, T. Ueda, A. Machiyama, I. Shimada, H. Iwaï, Segmental isotopic labeling of a 140 kDa dimeric multi-domain protein CheA from Escherichia coli by expressed protein ligation and protein trans-splicing, J. Biomol. NMR. 53 (3) (2012) 191–207, https://doi.org/10.1007/s10858-012-9628-3.

[12] T. Otomo, K. Teruya, K. Uegaki, T. Yamazaki, Y. Kyogoku, Improved segmental isotope labeling of proteins and application to a larger protein, J. Biomol. NMR. 14 (2) (1999) 105–114, https://doi.org/10.1023/a:1008308128050.

[13] S. Züger, H. Iwai, Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies, Nat. Biotechnol. 23 (6) (2005) 736–740, https://doi.org/10.1038/nbt1097.

[14] A. Ciragan, A.S. Aranko, I. Tascon, H. Iwaï, Salt-inducible protein splicing in cis and trans by inteins from extremely halophilic archaea as a novel protein-engineering tool, J. Mol. Biol. 428 (2016) 4573–4588, https://doi.org/10.1016/j.jmb.2016.10.006.

[15] A. Ciragan, S.M. Backlund, K.M. Mikula, H.M. Beyer, O.H.S. Ollila, H. Iwaï, NMR structure and dynamics of TonB investigated by scar-less segmental isotopic labeling using a salt-inducible split intein, Front. Chem. 8 (2020) 136, https://doi.org/10.3389/fchem.2020.00136.

[16] A.S. Aranko, H. Iwaï, The Inducible Intein-Mediated Self-Cleaving Tag (IIST) System: A Novel Purification and Amidation System for Peptides and Proteins, Molecules 26 (19) (2021) 5948, https://doi.org/10.3390/molecules26195948.

[17] G. Ortega, A. Laín, X. Tadeo, B. López-Méndez, D. Castaño, O. Millet, Halophilic enzyme activation induced by salts, Sci Rep. 1 (2011) 6, https://doi.org/10.1038/srep00006.

[18] S. Fukuchi, K. Yoshimune, M. Wakayama, M. Moriguchi, K. Nishikawa, Unique amino acid composition of proteins in halophilic bacteria, J. Mol. Biol. 327 (2) (2003) 347–357, https://doi.org/10.1016/s0022-2836(03)00150-5.

[19] S. Paul, S.K. Bag, S. Das, et al., Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes, Genome Biol. 9 (2008) R70, https://doi.org/10.1186/gb-2008-9-4-r70.

[20] X. Tadeo, B. López-Méndez, T. Trigueros, A. Laín, D. Castaño, O. Millet, Structural basis for the aminoacid composition of proteins from halophilic archea, PLoS Biol. 7 (12) (2009) e1000257, https://doi.org/10.1371/journal.pbio.1000257.

[21] G. Ortega, T. Diercks, O. Millet, Halophilic protein adaptation results from synergistic residue-ion interactions in the folded and Unfolded States, Chem. Biol. 22 (2015) 1597–1607, https://doi.org/10.1016/j.chembiol.2015.10.010.

[22] D. Madern, C. Ebel, G. Zaccai, Halophilic adaptation of enzymes, Extremophiles. 4 (2) (2000) 91–98, https://doi.org/10.1007/s007920050142.

[23] H. Iwai, S. Züger, J. Jin, P.H. Tam, Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme, FEBS Lett. 580 (7) (2006) 1853–1858, https://doi.org/10.1016/j.febslet.2006.02.045.

[24] J.S. Oeemig, A.S. Aranko, J. Djupsjöbacka, K. Heinämäki, H. Iwaï, Solution structure of DnaE intein from Nostoc punctiforme: structural basis for the design of a new split intein suitable for site-specific chemical modification, FEBS Lett. 583 (9) (2009) 1451–1456, https://doi.org/10.1016/j.febslet.2009.03.058.

[25] J.S. Oeemig, H.M. Beyer, A.S. Aranko, J. Mutanen, H. Iwaï, Substrate specificities of inteins investigated by QuickDrop-cassette mutagenesis, FEBS Lett. 594 (20) (2020) 3338–3355, https://doi.org/10.1002/1873-3468.13909.

[26] S. Ellilä, J.M. Jurvansuu, H. Iwaï, Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in Escherichia coli, FEBS Lett. 585 (21) (2011) 3471–3477, https://doi.org/10.1016/j.febslet.2011.10.005.
[27] H. Iwai, G. Wider, K. Wüthrich, NMR structure of a variant 434 repressor DNA-binding domain devoid of hydroxyl groups, J. Biomol. NMR 29 (3) (2004) 395–398, https://doi.org/10.1023/B:JNMR.0000032609.72759.41.
[28] M.S. Lawrence, K.J. Phillips, D.R. Liu, Supercharging proteins can impart unusual resilience, J. Am. Chem. Soc. 129 (33) (2007) 10110–10112, https://doi.org/10.1021/ja071641y.
[29] A.S. Aranko, J.S. Oeemig, T. Kajander, H. Iwaï, Intermolecular domain swapping induces intein-mediated protein alternative splicing, Nat. Chem. Biol. 9 (10) (2013) 616–622, https://doi.org/10.1038/nchembio.1320.
[30] A.S. Aranko, S. Züger, E. Buchinger, H. Iwaï, In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins, PLoS One 4 (4) (2009) e5185, https://doi.org/10.1371/journal.pone.0005185.
[31] A.E. Kelly, H.D. Ou, R. Withers, V. Dötsch, Low-conductivity buffers for high-sensitivity NMR measurements, J. Am. Chem. Soc. 124 (40) (2002) 12013–12019, https://doi.org/10.1021/ja026121b.
[32] M. Takeda, K. Hallenga, M. Shigezane, et al., Construction and performance of an NMR tube with a sample cavity formed within magnetic susceptibility-matched glass, J. Magn Reson. 209 (2) (2011) 167–173, https://doi.org/10.1016/j.jmr.2011.01.005.
[33] A. Bhattacharya, R. Tejero, G.T. Montelione, Evaluating protein structures determined by structural genomics consortia, Proteins. 66 (4) (2007) 778–795, https://doi.org/10.1002/prot.21165.
[34] S.N. Timasheff, Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components, Proc. Natl. Acad. Sci. USA 99 (15) (2002) 9721–9726, https://doi.org/10.1073/pnas.122225399.
[35] J.K. Lanyi, Salt-dependent properties of proteins from extremely halophilic bacteria, Bacteriol Rev 38 (3) (1974) 272–290.
[36] G. Otting, E. Liepinsh, K. Wüthrich, Protein hydration in aqueous solution, Science 254 (5034) (1991) 974–980, https://doi.org/10.1126/science.1948083.
[37] V. Dötsch, G. Wider, G. Siegal, K. Wüthrich, Salt-stabilized globular protein structure in 7 M aqueous urea solution, FEBS Lett. 372 (2–3) (1995) 288–290, https://doi.org/10.1016/0014-5793(95)01004-x.
[38] K. Pervushin, G. Wider, H. Iwai, K. Wüthrich, NMR structures of salt-refolded forms of the 434-repressor DNA-binding domain in 6 M urea, Biochemistry 43 (2004) 13937–13943, https://doi.org/10.1021/bi048496a.
[39] E. Liepinsh, H. Rink, G. Otting, K. Wüthrich, Contributions from hydration of carboxylate groups to the spectrum of water-polypeptide proton-proton Overhauser effects in aqueous solution, J. Biomol. NMR 3 (2) (1993) 253–257, https://doi.org/10.1007/BF00178268.
[40] J. Qvist, G. Ortega, X. Tadeo, O. Millet, B. Halle, Hydration dynamics of a halophilic protein in folded and unfolded states, J. Phys. Chem. B. 116 (10) (2012) 3436–3444, https://doi.org/10.1021/jp3000569.
[41] M. Belfort, Mobile self-splicing introns and inteins as environmental sensors, Curr. Opin. Microbiol. 38 (2017) 51–58, https://doi.org/10.1016/j.mib.2017.04.003.
[42] G.R. Buel, K.J. Walters, Can AlphaFold2 predict the impact of missense mutations on structure?, Nat Struct. Mol. Biol. 29 (1) (2022) 1–2, https://doi.org/10.1038/s41594-021-00714-2.
[43] K. Inomata, A. Ohno, H. Tochio, et al., High-resolution multi-dimensional NMR spectroscopy of proteins in human cells, Nature 458 (7234) (2009) 106–109, https://doi.org/10.1038/nature07839.
[44] A.G. Sobol, G. Wider, H. Iwai, K. Wüthrich, Solvent magnetization artifacts in high-field NMR studies of macromolecular hydration, J. Magn. Reson. 130 (2) (1998) 262–271, https://doi.org/10.1006/jmre.1997.1287.
[45] F. Guerrero, A. Ciragan, H. Iwaï, Tandem SUMO fusion vectors for improving soluble protein expression and purification, Protein Expr. Purif. 116 (2015) 42–49, https://doi.org/10.1016/j.pep.2015.08.019.
[46] H.A. Heikkinen, S.M. Backlund, H. Iwaï, NMR structure determinations of small proteins using only one fractionally 20% 13C- and uniformly 100% 15N-labeled sample, Molecules 26 (3) (2021) 747, https://doi.org/10.3390/molecules26030747.
[47] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, Prog. Nucl. Magn Reson Spectrosc 34 (1999) 93–158.
[48] E. Lescop, P. Schanda, B. Brutscher, A set of BEST triple-resonance experiments for time-optimized protein resonance assignment, J. Magn. Reson. 187 (2007) 163–169.
[49] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 59 (2005) 687–696.
[50] P. Güntert, C. Mumenthaler, K. Wüthrich, Torsion angle dynamics for NMR structure calculation with the new program DYANA, J. Mol. Biol. 273 (1997) 283–298.
[51] P. Güntert, Automated structure determination from NMR spectra, Eur. Biophys. J. 38 (2009) 129–143, https://doi.org/10.1007/s00249-008-0367-z.
[52] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[53] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graph 14 (1996) 51–55.