

Cite this: *Chem. Sci.*, 2023, 14, 5204

All publication charges for this article have been paid for by the Royal Society of Chemistry

# Structural and biochemical analysis of a novel atypically split intein reveals a conserved histidine specific to cysteine-less inteins†

Tim Pasch, <sup>a</sup> Alexander Schröder, <sup>a</sup> Sabrina Kattelman, <sup>a</sup> Miriam Eisenstein, <sup>b</sup> Shmuel Pietrokovski, <sup>\*b</sup> Daniel Kümmel <sup>\*a</sup> and Henning D. Mootz <sup>\*a</sup>

Protein *trans*-splicing mediated by a split intein reconstitutes a protein backbone from two parts. This virtually traceless autoproteolytic reaction provides the basis for numerous protein engineering applications. Protein splicing typically proceeds through two thioester or oxyester intermediates involving the side chains of cysteine or serine/threonine residues. A cysteine-less split intein has recently attracted particular interest as it can splice under oxidizing conditions and is orthogonal to disulfide or thiol bioconjugation chemistries. Here, we report the split PolB16 OarG intein, a second such cysteine-independent intein. As a unique trait, it is atypically split with a short intein-N precursor fragment of only 15 amino acids, the shortest characterized to date, which was chemically synthesized to enable protein semi-synthesis. By rational engineering we obtained a high-yielding, improved split intein mutant. Structural and mutational analysis revealed the dispensability of the usually crucial conserved motif N3 (block B) histidine as an obvious peculiar property. Unexpectedly, we identified a previously unnoticed histidine in hydrogen-bond forming distance to the catalytic serine 1 as critical for splicing. This histidine has been overlooked so far in multiple sequence alignments and is highly conserved only in cysteine-independent inteins as a part of a newly discovered motif NX. The motif NX histidine is thus likely of general importance to the specialized environment in the active site required in this intein subgroup. Together, our study advances the toolbox as well as the structural and mechanistic understanding of cysteine-less inteins.

Received 5th March 2023

Accepted 23rd April 2023

DOI: 10.1039/d3sc01200j

rsc.li/chemical-science

## Introduction

Split inteins reconstitute a host protein from two separate pieces to a single polypeptide chain. This peptide ligation reaction, termed protein *trans*-splicing, proceeds through one or two covalent thioester or ester intermediates on conserved cysteine, serine, or threonine residues and joins the intein flanks with a peptide bond (see Fig. 1). Because inteins excise themselves during the reaction, the ligation is virtually traceless, with the remaining +1 catalytic residue at the ligation site being part of the host protein (Fig. 1).<sup>1–3</sup> Transplanting split inteins into a heterologous sequence context of choice enables a plethora of protein engineering applications.<sup>2–7</sup> Therefore, functionally robust split inteins are being urgently sought after. The split inteins' specific properties and sequence features can be critical for the particular protein ligation challenge. For

example, short split N- or C-terminal intein fragments (Int<sup>N</sup> or Int<sup>C</sup>) are amenable to solid-phase peptide synthesis and enable semi-synthetic protein *trans*-splicing.<sup>8–11</sup>

Cysteine-less inteins are inteins lacking cysteine residues, first and foremost at their 1 and +1 positions which are involved in the formation of linear and branched intermediates, with the +1 position formally representing the first C-extein residue (Fig. 1). Cysteine-less inteins can catalyze the protein *trans*-splicing reaction under oxidizing conditions and without adding reducing agents. Under these conditions, the more abundant cysteine-dependent counterparts can be impaired in their efficiency due to partial or quantitative oxidation of their cysteine thiol moieties. This is particularly critical for split intein precursors that have been separately produced. Thus, cysteine-less split inteins significantly expand our toolbox for protein engineering applications, for example, to preserve critical disulfide bridges in proteins of interest (POI)<sup>12</sup> or for novel combinations with thiol-dependent bioconjugation reactions.<sup>12–15</sup> However, cysteine-less split inteins are very rare. The Aes123 PolB1 intein and its engineered CL variant provide the only such reported and characterized examples with useful activity.<sup>12</sup> The previously reported Neq Pol intein was the first discovered naturally split cysteine-less intein,<sup>16</sup> however, as

<sup>a</sup>Institute of Biochemistry, University of Münster, Corrensstr. 36, 48149 Münster, Germany. E-mail: Henning.Mootz@uni-muenster.de

<sup>b</sup>Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: shmuel.pietrokovski@weizmann.ac.il

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3sc01200j>

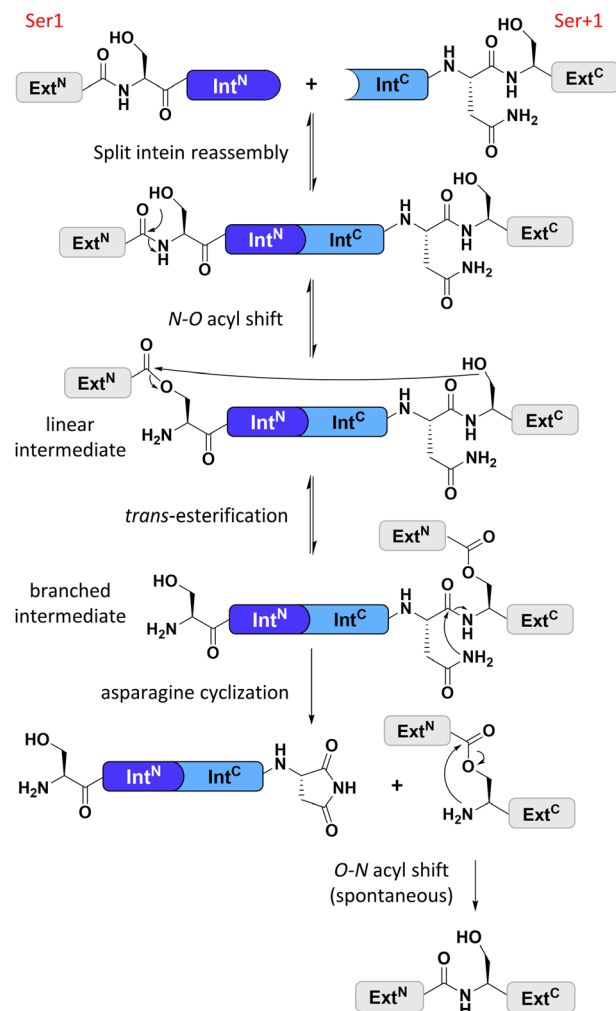


Fig. 1 General protein splice mechanism of a cysteine-less split intein. Note that such inteins operate with Ser1 and Ser+1 (Thr+1) residues at the two splice junctions to form the linear and branched ester intermediates.  $\text{Ext}^{\text{N}}$ ,  $\text{Ext}^{\text{C}}$  = N- and C-terminal exteins, respectively.

a thermophilic protein it required temperatures of at least 50–70 °C to show reasonable activity.<sup>16</sup> The other published example of a cysteine-less split intein is the Psp-GBD Pol intein, which was artificially generated from its contiguous, *cis*-splicing parent intein. Likely because of this engineering, the separate split intein precursors did not assemble properly and required a denaturation–renaturation cycle for refolding into the active form.<sup>15,17</sup> The high temperature and the required refolding preclude the Neq Pol and the Psp-GBD Pol inteins, respectively, from a practical utility for most applications.

Identifying novel cysteine-less split inteins requires locating the active site serine or threonine residues at the 1 and +1 positions. However, it is unclear how the nucleophilic side chains at those positions are selected in intein evolution. In a survey of all *cis* and *trans*-splicing inteins of the intein database, only 10% exhibited only serine or threonine residues at both these positions.<sup>18</sup> The +1 position appears more permissive for site-directed mutagenesis than the 1 position, as a couple of examples of C+1S substitutions with retained splicing activity

were reported.<sup>19,20</sup> In contrast, the Cys1 residue appears much more refractory to changing the side chain thiol to an alcohol moiety, as for several inteins a C1S mutation was shown to result in an inactive intein.<sup>18,20</sup> No attempt to convert a cysteine-dependent intein into a cysteine-less intein useful for preparative purposes has been successful so far. Obviously, inteins exhibit a high level of specialization around the active site, with a Ser1 residue requiring a dedicated activation mechanism that the active site of a Cys1 intein cannot provide.

Artificially splitting continuous *cis* inteins is another route to obtain split inteins. Indeed, several *cis* inteins without cysteine at the 1 and +1 positions are known, either with or without cysteines at other non-conserved positions. However, folding and assembly problems observed for the above-mentioned artificially split Psp-GBD Pol intein<sup>17</sup> underline the potential drawbacks of this strategy and have also been observed for several other artificially split inteins.<sup>21–23</sup> Therefore, identifying naturally occurring split inteins is considered the more attractive route to well-behaved variants<sup>19,24,25</sup> and thus also the most promising path to powerful novel cysteine-less split inteins.

Here, we report the novel and atypically split PolB16 OarG intein and its biochemical and structural characterization. By replacement of two non-conserved cysteines, it could be mutated into a completely cysteine-less variant and further engineered for efficient splicing properties. Strikingly, we uncovered a new conserved and indispensable histidine residue, while the typically crucial motif N3 histidine is superfluous in the PolB16 OarG intein. The region of this newly identified histidine is conserved and present in other cysteine-less inteins, and forms the first signature trait specific for them, apart from their splice junctions serine or threonine residues. The PolB16 OarG intein utilizes a novel mechanism for activating the Ser1 residue in the protein-splicing pathway.

## Results and discussion

### Identification of the split PolB16 OarG intein and its utility for cysteine-less protein *trans*-splicing applications

We identified the split PolB16 OarG intein, from now on also simply termed PolB16 intein, in sequence data from a sheep gut metagenome.<sup>26</sup> The intein genes are inserted in a gene for a putative PolB-type DNA-polymerase. Naturally occurring split inteins typically exhibit a longer  $\text{Int}^{\text{N}}$  of about 100 amino acids (aa) and a shorter  $\text{Int}^{\text{C}}$  of about 35–40 aa. However, atypically split inteins with a short  $\text{Int}^{\text{N}}$  fragment and correspondingly longer  $\text{Int}^{\text{C}}$  have recently been identified.<sup>20,27,28</sup> The PolB16 intein carries an atypical split position close to the N-terminal end of the encoded split intein domain to give  $\text{Int}^{\text{N}}$  and  $\text{Int}^{\text{C}}$  fragments of 15 and 168 aa (Fig. 2), respectively. Most notably, both residues at the 1 and +1 positions are serine. Furthermore, the intein overall exhibits the typical conserved motifs of a class 1 intein<sup>29</sup> (Fig. 2) that represent conserved residues across inteins, of which some are involved in catalyzing the steps of the protein splicing pathway. This analysis suggested the PolB16 intein splices through linear and branched oxoester intermediates (Fig. 1). The  $\text{Int}^{\text{C}}$  fragment contains two cysteines within motifs N3 and C2,<sup>30</sup> (previously also referred to as motifs B and



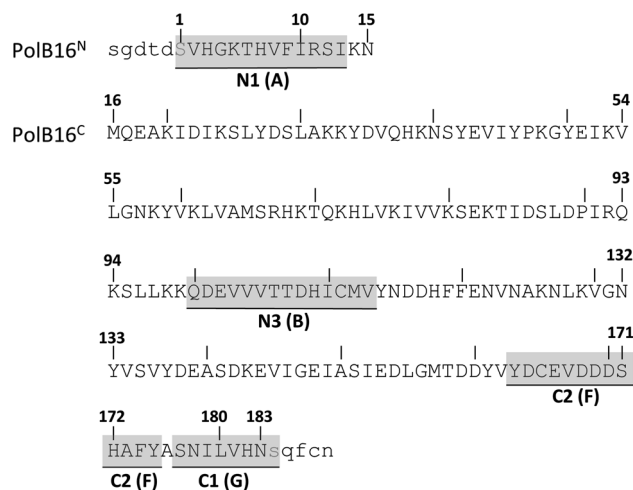


Fig. 2 Amino acid sequence of the split PolB16 intein. Shown are the Int<sup>N</sup> (PolB16<sup>N</sup>) and Int<sup>C</sup> (PolB16<sup>C</sup>) precursors with 5 extein residues each (in lower cases). Highlighted are the conserved motifs (grey) with the older block nomenclature in brackets, as well as the two catalytic serines at positions 1 and +1 (blue).

F)<sup>31</sup> however, at non-essential and only moderately conserved positions that usually harbor hydrophobic residues.

For a biochemical characterization, we produced recombinant constructs of the PolB16 intein fragments in *E. coli*. Each fragment contained 5 native extein residues, fused to maltose-binding protein (MBP) and enhanced green fluorescent protein (eGFP) to give MBP-Int<sup>N</sup>-H<sub>6</sub> (1) and Int<sup>C</sup>-eGFP-H<sub>6</sub> (2), respectively. Upon mixing these purified model proteins, we observed virtually complete turnover of the Int<sup>C</sup> precursor (2) at

pH 7.0 and 25 °C (Fig. 3A and S1B†). We made similar findings for the mutated Int<sup>C</sup> precursor Int<sup>C</sup>(C111A, C165A)-eGFP-H<sub>6</sub> (3), suggesting that both non-conserved cysteines can be replaced without impairing intein activity (Fig. 3). In this fully cysteine-less Int<sup>C</sup> precursor the cysteine in the +4 position of the flanking extein region was also replaced with alanine.

We varied temperature, salt concentration and pH conditions to reveal the optimal activity of the cysteine-less PolB16 split intein in terms of efficiency at pH 6.0 and 25 °C (Fig. S2†). Fig. 3B shows that the splice product MBP-eGFP-H<sub>6</sub> was formed with 89% yield under these conditions. As a by-product of C-terminal cleavage, eGFP-H<sub>6</sub> was found with 9.5% yield (Fig. 3C). Importantly, these protein *trans*-splicing assays led to the same results in the absence and presence of reducing agents like DTT and TCEP. At pH 7 we found the wild-type split PolB16 intein (1 + 2) to be twice as fast as its cysteine-less variant (1 + 3) with a reaction constant of  $0.16 \pm 0.02 \times 10^{-3} \text{ s}^{-1}$  compared to  $0.77 \pm 0.01 \times 10^{-4} \text{ s}^{-1}$ , respectively, assuming pseudo-first-order conditions. At pH 6 the rates increase by a factor of 4–6-fold with reaction constants of  $0.57 \pm 0.05 \times 10^{-3} \text{ s}^{-1}$  vs.  $0.43 \pm 0.03 \times 10^{-3} \text{ s}^{-1}$ . The corresponding half-life times are  $t_{1/2} = 73 \text{ min}$  and  $150 \text{ min}$  at pH 7, as well as  $20 \text{ min}$  and  $27 \text{ min}$  at pH 6, respectively. The improved rates and splice products yields observed at pH 6 compared to pH 7 likely stem from the beneficial effect of the lower pH to form the linear intermediate in equilibrium with the peptide bond (Fig. 1), mediated by protonation of the liberated  $\alpha$ -amino group of Ser1.<sup>32</sup>

We explored the rare and extremely short Int<sup>N</sup> fragment of the PolB16 intein and prepared an Int<sup>N</sup> precursor by solid-phase peptide synthesis (SPPS) with 5(6)-carboxyfluorescein (CF) as a synthetic fluorophore in a short N-extein (CF-SGDTD-Int<sup>N</sup>

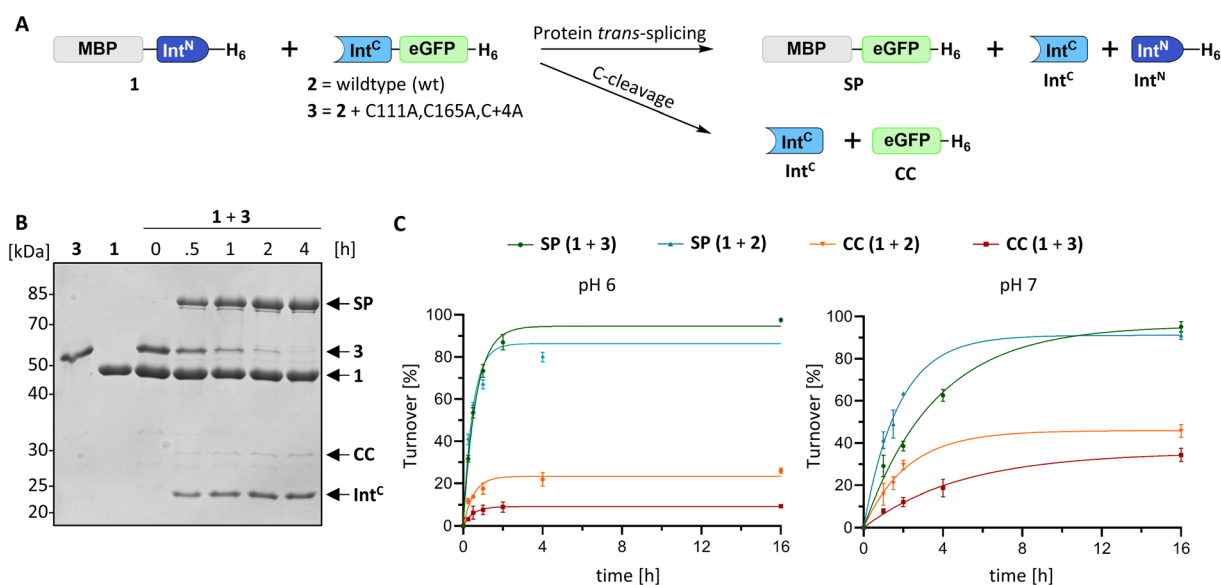
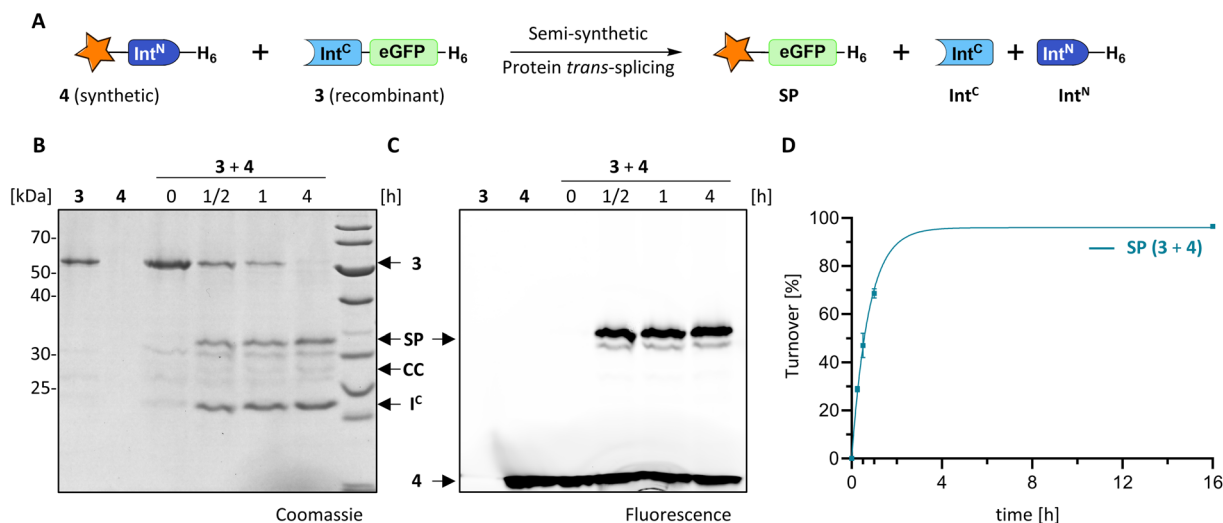


Fig. 3 Protein *trans*-splicing of the PolB16 intein and its engineered cysteine-less mutant. (A) Scheme of the reactions. (B) Reactions were performed with the Int<sup>N</sup> (10  $\mu\text{M}$ ) and Int<sup>C</sup> (5  $\mu\text{M}$ ) precursors at 25 °C and pH 6.0. Shown is a representative coomassie-stained SDS-PAGE gel analysis with the Int<sup>C</sup> precursor mutant devoid of any cysteines. (C) Densitometric analyses of the reactions under the same conditions as in (B), but at the indicated pH values ( $n = 3$  for each reaction; error bars represent standard deviations). SP = splice product. CC = C-terminal cleavage product.



**Fig. 4** Semi-synthetic protein *trans*-splicing with the PolB16 intein. (A) Scheme of the reaction. Synthetic CF-SGDTD-Int<sup>N</sup> (4) was prepared by SPPS. (B and C) SDS-PAGE analysis using coomassie-staining or fluorescence, respectively, of the reaction between 4 (15  $\mu$ M) and 3 (5  $\mu$ M) at 25 °C and pH 6. (D) Densitometric analysis of the reaction ( $n = 4$ ; error bars represent standard deviations) shown in (B). SP = splice product.

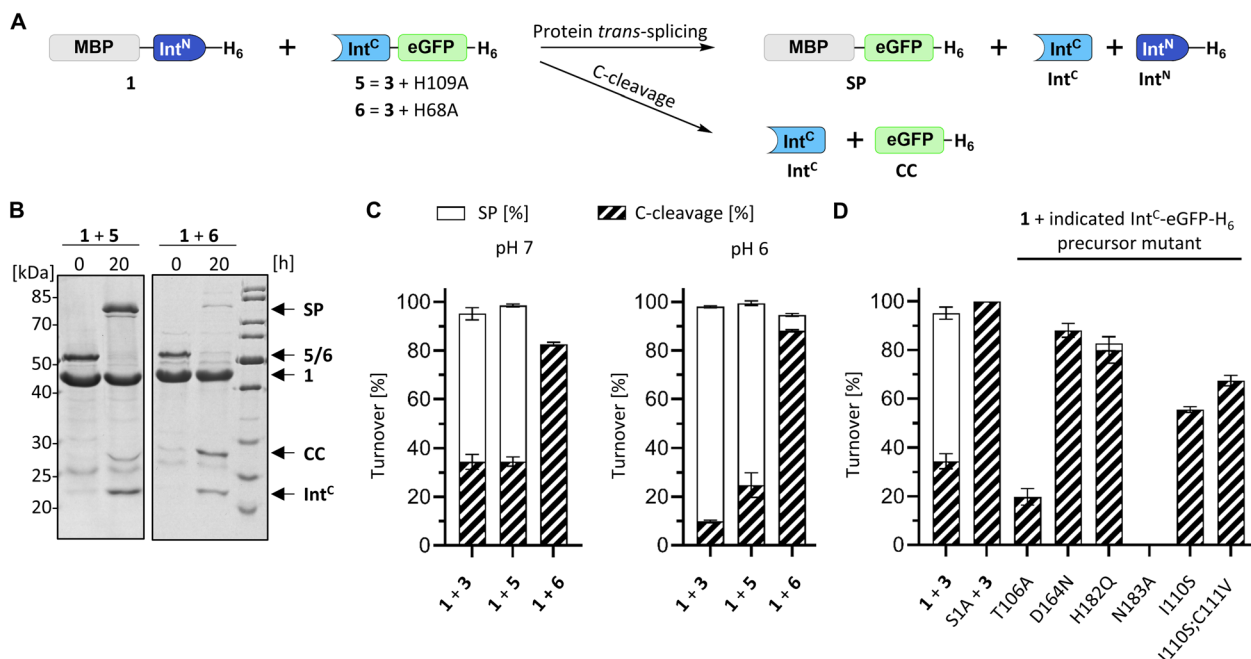
construct 4; Fig. S3†). Fig. 4 shows that upon incubation with the Int<sup>C</sup> precursor 3 virtually quantitative formation of the semi-synthetic splicing product was obtained. The rate of  $0.38 \pm 0.01 \times 10^{-3} \text{ s}^{-1}$  was comparable to the fully recombinant intein.

As PolB16 intein is only the second characterized cysteine-less split intein with ambient temperature activity, we also demonstrated that it is orthogonal to the previously described Aes123/CL intein (Fig. S4†),<sup>12</sup> thus enabling possibilities for

dual-splicing applications. Together, these findings establish the engineered, cysteine-less PolB16 intein as a valuable addition to the protein chemistry toolbox.

#### The typically critical motif N3 histidine is dispensable in the PolB16 intein

We next examined the origins of the observed C-terminal cleavage. At pH 7.0 the level of this side reaction is higher



**Fig. 5** Mutational analysis of the PolB16 intein. (A) Scheme of the PTS reactions. Reaction were performed by mixing the Int<sup>N</sup> (10  $\mu$ M) and Int<sup>C</sup> precursors (5  $\mu$ M) at 25 °C at the indicated pH. (B) Representative SDS-PAGE analysis of reactions of the indicated mutants at pH 6. Shown are coomassie-stained gels. (C and D) Diagram representation of the reaction yields of the total turnover as the sum of the SP (white) and CC (diagonally striped) ( $n = 3$ ; error bars represent standard deviations). See Fig. S5† for the SDS-PAGE analyses of the remainder of the mutants.



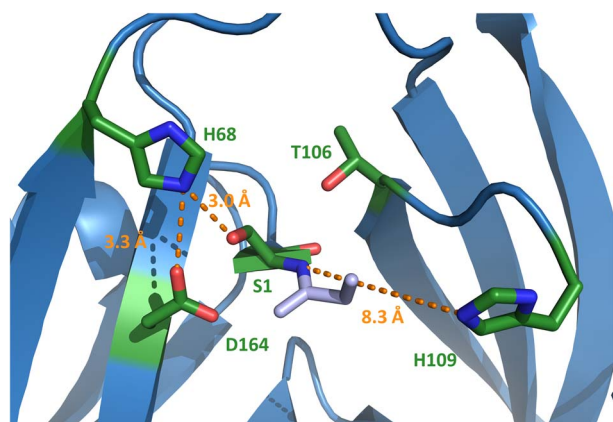


Fig. 6 Crystal structure of the N-terminal active site of the PolB16 intein. The discussed and conserved residues of splicing motifs are highlighted in green. The intein was crystallized with S1A and N183A substitutions. For this representation, the side chain of Ser1 was modeled back into the structure using PyMol. Selected distances are given in Ångström (orange). Motif N3 histidine 109 is in unusual position.

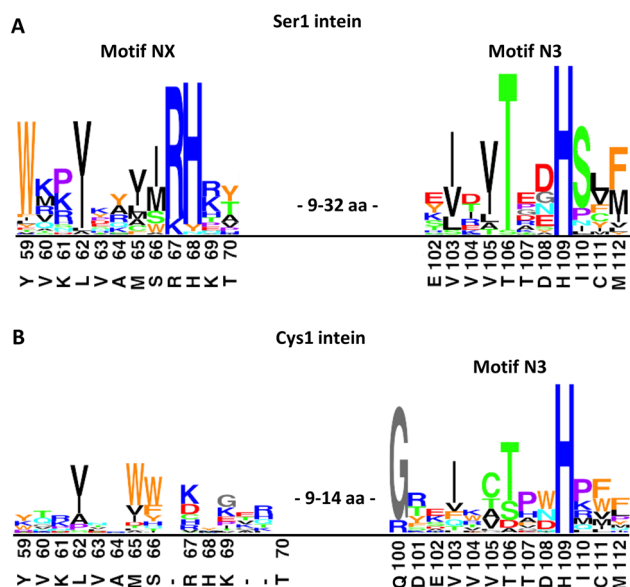


Fig. 7 Sequence logo motifs N3 and NX in different intein types. The height of each individual amino acid represents their abundance at the given position. Amino acids were colored based on their properties. Corresponding positions and residues in PolB16 are shown below. (A) 58 Ser1 inteins from the InBase<sup>43</sup> database that also have a Ser+1 or Thr+1 at the C-terminal splice junction. (B) 19 representative Cys1 inteins of known structure. PolB16 is part of the data for (A), but not for (B). Fig. S7† shows the multiple sequence alignments used to create these logos and the intervening regions between the NX and N3 regions.

than typically observed for naturally split inteins. C-cleavage is caused by premature asparagine cyclization. Given the absence of any detectable N-cleavage as a potential initial trigger,<sup>33</sup> we suspected that the asparagine cyclization step is partially decoupled from the initial N-O acyl shift within the protein

splicing pathway.<sup>34–36</sup> To probe the effect of a completely impaired initial N-O acyl shift we mutated the motif N3:10 histidine (His109), which is typically involved in this step by weakening the N-terminal scissile bond.<sup>37–39</sup> Surprisingly, no significant adverse impact was observed with the Int<sup>C</sup>(H109A)-eGFP-H<sub>6</sub> mutant (construct 5), despite the highly conserved nature and usually critical role of this residue in other inteins. The obtained splicing and C-cleavage yields remained nearly unchanged relative to the unmutated cysteine-less Int<sup>C</sup> precursor (Fig. 5A–D). At pH 7 the splicing rate of the mutant precursor even increased by nearly 3-fold ( $0.24 \pm 0.01 \times 10^{-3} \text{ s}^{-1}$ ), whereas at pH 6 a nearly 4-fold decrease was observed compared to the unmutated precursor ( $0.11 \pm 0.01 \times 10^{-3} \text{ s}^{-1}$ ). Beyond the surprising dispensability of the motif N3:10 histidine, these findings hinted at unusual molecular properties.

### PolB16 intein crystal structures show the hallmarks of a class 1 intein but the region around the motif N3 His109 is unresolved

To better understand the molecular details of the splicing mechanism with a dispensable motif N3:10 histidine, we determined crystal structures of the PolB16 intein. To this end, we genetically fused the Int<sup>N</sup> and Int<sup>C</sup> precursors, each with 10 extein residues, with a 3 amino acid linker (GSH). Ser1 and Asn183 at the splice junctions were mutated to alanine to block all splicing activity. We obtained protein crystals in this format for both the wild-type sequence and the cysteine-less double mutant (C111A, C165A) and could solve these at resolutions of 1.85 and 2.6 Å, respectively. Both structures were virtually identical (0.185 Å RMSD, root mean square deviation), consistent with the negligible functional impact of mutating the non-conserved cysteine residues (Fig. S6†). The structures show the canonical horseshoe fold of inteins with an insertion in the region aa T84–D101 that explains the longer Int<sup>C</sup> precursor compared to the minimal intein fold and other inteins (Fig. S7†). No sufficient electron density was found for a part within this region, D86–S95, suggesting it is mobile or less structured, as well as for the termini represented by extein residues (except Asp-1) and the ultimate N183A.

Surprisingly, electron density was also missing for the stretch I110–N127 (H109–N127 for the cysteine-less variant), which involves the residues directly following the aforementioned catalytic motif N3:10 residue His109 and the next layer of amino acids typically wrapped around this part of the active site (Fig. S6†). The last well-resolved amino acids, Asp108 and His109, show a conformation of the polypeptide chain that departs from the canonical intein structure. As a consequence, His109 is placed in an unusual position, remote from the scissile bond (Fig. 6 and S7C†), which is incompatible with its usual catalytic role. This finding explains the dispensable role of the His109 residue for catalysis. Thr106 as the last residue found in the usual conformation (Fig. S7C†) is the highly conserved motif N3:7 residue that plays a critical role in most inteins by twisting the upstream scissile bond in a strained conformation to support the acyl shift reaction.<sup>40</sup> Other key residues known to be involved in the canonical splicing mechanism were found at



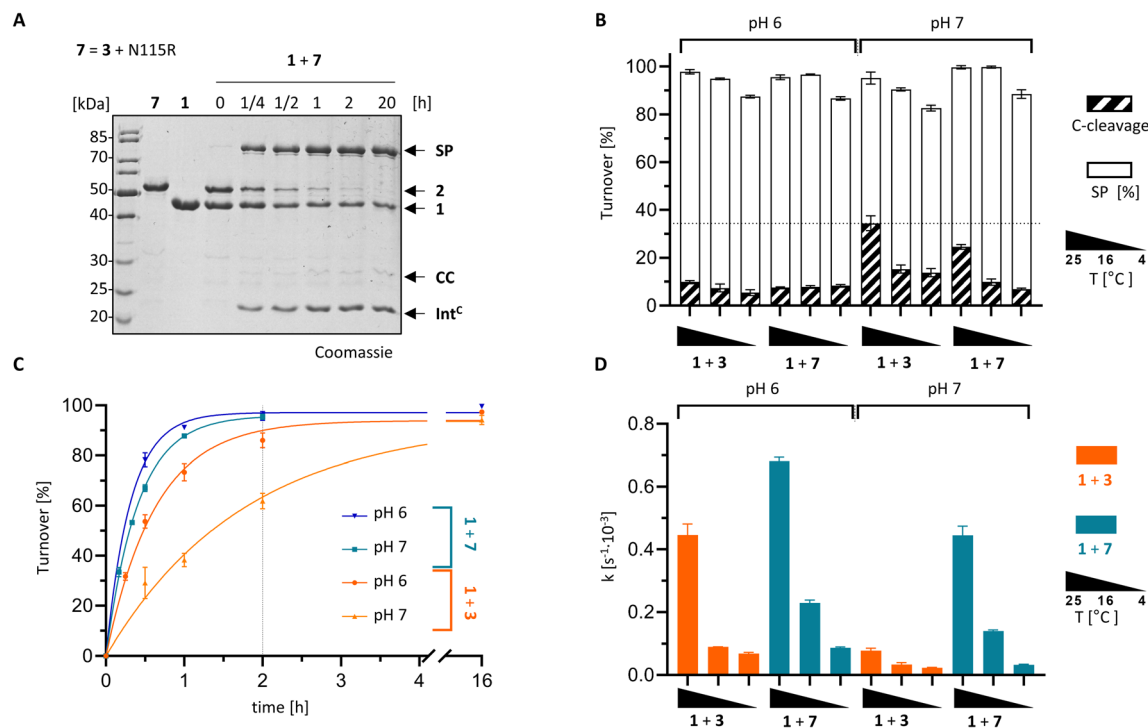


Fig. 8 Rationally improved PolB16 intein with N115R mutation. PTS reactions of the Int<sup>C</sup> precursors CL-Int<sup>C</sup>(C111A, N115R, C165A, C+4A)-eGFP-H<sub>6</sub> (7) and Int<sup>C</sup>-eGFP-H<sub>6</sub> (3) were performed according to the scheme shown in Fig. 3A using MBP-Int<sup>N</sup>-H<sub>6</sub> (1) at 10 μM and the Int<sup>C</sup> precursors at 5 μM. (A) Representative coomassie-stained SDS-PAGE gels of the reaction at 25 °C and pH 6. (B) Diagram of the total turnover as the sum of the SP (white) and CC (diagonally striped). Integrated into the same diagram is splicing rate (black; see y-axis on the right hand side) for the indicated split intein combinations ( $n = 3$ ). (C) Time-courses of the reactions at 25 °C. (D) Kinetic rates of the turnover reactions as determined by densitometric gel analysis. Error bars in (B–D) represent standard deviations ( $n = 3$ ).

their usual locations, including the motif C2 Asp164, the motif C1 His182 as well as the mutated S1A at the splice junction (Fig. 6 and S7C†). Probing several conserved motif residues by site-directed mutagenesis (S1A, T106A, D164N, H182Q, and N183Q) revealed either entirely or almost completely impaired splicing activities, either at the expense of increased cleavage reactions or resulting in decreased intein turnover altogether (Fig. 5D and S5†). Together, these findings corroborated the canonical roles of the conserved residues. Only the unusual positioning and functional dispensability of the motif N3:10 residue His109 suggested a mechanistic deviation.

#### Discovery of the novel motif NX with a histidine residue conserved only in cysteine-less inteins

Trying to explain the unusual functional and structural detachment of motif N3:10 residue His109 from the splicing pathway and the intein's active site, we noted in the crystal structures that His68 is positioned in the vicinity of the upstream splice junction. It approaches the N-terminal scissile bond from the opposite side as the motif N3:10 histidine (Fig. 6, S7C and D†). To our knowledge, a histidine at this position has not been probed by site-directed mutagenesis so far. We found that a H68A mutant of the Int<sup>C</sup> precursor (construct 6) was indeed completely incapable of protein *trans*-splicing at pH 7.0, while only residual activity of about 5% was observed at pH 6.0 (Fig. 5B and C), indicating a critical role for this residue.

To further examine the importance of His68, we constructed a multiple sequence alignment of the subgroup of known class 1 inteins that lack cysteines at the intein junction positions, *i.e.*, using only Ser1/Ser+1 and Ser1/Thr+1 inteins (Fig. 7A and S7A†). These 58 inteins are highly diverse in sequence (Fig. S7A†) and are of 16 different intein alleles (*e.g.*, integration points<sup>41</sup>). Strikingly, this analysis revealed that the histidine corresponding to His68 in the PolB16 intein is highly conserved in this cysteine-independent subset of inteins. In contrast, when aligning sequences of 19 representative inteins of known structure of the Cys1 subset (containing Cys+1, Ser+1 or Thr+1 at the C-terminal splice junction), including 15 different alleles, we found no histidine or another residue at this position to be conserved (Fig. 7B, S7B and C†), despite the fact that their backbone folds of the region around the histidine are very similar to the PolB16 intein. Thus, it appears that the newly identified histidine residue was so far overlooked, because in mixed multiple sequence alignments of Ser1 and Cys1 inteins, in which the former group represents the smaller subset of known sequences, it remained obscured. We term the new sequence motif with the unveiled conserved histidine in Ser1 intein the NX motif. It represents a new signature motif specific to class 1 inteins with oxyester chemistry (Fig. 7).

The NX motif histidine residue is invariably preceded by an arginine or lysine residue, which interacts with a conserved glutamic acid (Glu43 in the PolB16 intein) in a preceding β-

strand (Fig. S7A and D†). Two of the examined sequences have tyrosine or asparagine residues instead of histidine, indicating that even in the Ser1 subset the motif NX histidine is not completely conserved, like all important intein residues except for those directly involved in the (thio)ester intermediates.

We next examined the role of motif NX histidine in two other Ser1 inteins (Fig. S8†). Its mutation to alanine in the previously reported Psp Pol-1 maxi-intein<sup>17,42</sup> as well as Mvu-M7 Pol-3 mini-intein, taken from Inbase<sup>43</sup> and reported here for the first time, virtually completely impaired splicing, suggesting a critical role. Of note, in both these inteins also the motif N3 histidine was critical, in contrast to the PolB16 intein. This latter histidine is absolutely conserved in all the Ser1 inteins we examined (Fig. 7 and S7†). Together, these findings underline the general importance of the newly discovered motif NX histidine contribution to catalysis in Ser1 inteins.

### Structure-guided engineering improves protein *trans*-splicing properties of the PolB16 intein

With the crystal structure as a basis, we attempted to rationally improve the splicing properties of the PolB16 intein. We focused on the surprisingly unstructured region around the N3 motif. Positions Ile110 and Cys111, following the conserved His109, are typically occupied in class 1 inteins by proline or a polar amino acid and a large hydrophobic or aromatic amino acid, respectively (Fig. 7).<sup>29</sup> However, attempts to redesign this sequence stretch with I110S and I110S/C111V mutations, inspired by the corresponding residues of the Aes123 intein,<sup>12</sup> led to inactive variants that only showed increased C-cleavage activity (~0–90%; Fig. S5D and S5†).

We then analyzed a structural model of the PolB16 intein calculated using the AlphaFold algorithm.<sup>44</sup> This model fitted our experimental data very well (Fig. S6†), interestingly however, it also predicted the typical intein fold for the region I110–N127 that was found unstructured in our crystals, albeit with only modest confidence scores (Fig. S9†). The lowest confidence was found in the turn between two  $\beta$ -sheets involving residues N115–H118. Guided by the multiple sequence alignment of Ser1/Ser(Thr+1) inteins (Fig. S7A†), we introduced an N115R substitution into the PolB16 intein. To our delight, this mutant Int<sup>C</sup> precursor Int<sup>C</sup>(C111A, N115R, C165A)–eGFP-H<sub>6</sub> (7) exhibited reduced C-terminal cleavage at pH 7.0 and significantly higher protein *trans*-splicing rates at both pH 7.0 and 6.0 ( $0.45 \pm 0.02 \times 10^{-3} \text{ s}^{-1}$  and  $0.68 \pm 0.01 \times 10^{-3} \text{ s}^{-1}$ ; corresponding to  $t_{1/2} = 26$  and 17 min, respectively; see Fig. 8). The rate at pH 7.0 and 25 °C represents a 6-fold improvement over the wild-type intein. This successful engineering result also further supported the idea that the region following motif N3 Thr106, including His109 and the unresolved stretch I110–N127, is attenuating the activity of the PolB16 intein.

## Conclusions

The novel atypically split PolB16 OarG intein reported herein is only the second reported mesophilic cysteine-less split intein. Although the native intein is not fully cysteine-less *per se*, its key

catalytic residues contain no cysteine. Through its Ser1 and Ser+1 residues at the splice junctions, it forms oxyesters in the linear and branched intermediates of the protein splicing pathway. After replacing two non-conserved cysteines, we showed protein *trans*-splicing efficiency of  $\geq 90\%$  at pH 6.0 and, importantly, without having treated the intein precursors at any time with reducing agents. A significantly improved rate and slightly improved yields were achieved with the rationally engineered N115R mutant variant. These results also suggest that further improvement of the PolB16 intein will be possible. While the splicing efficiencies of the current engineered PolB16 intein mutant are useful for the preparative manipulation of purified proteins, its splicing rates are probably not ideal to trigger rapid biological processes occurring on the second-to-minute timescale in real time.<sup>5,45–47</sup> For reactions at 37 °C the increasing tendency of the intein to form the C-terminal cleavage product at higher temperatures should be considered (Fig. S2†). Yet, protein *trans*-splicing also occurs at 37 °C (Fig. S2†) and a co-expression experiment of the PolB16 intein precursors in live *E. coli* cells indicated the compatibility of protein *trans*-splicing with a complex cellular environment (Fig. S10†), suggesting the intein will also be active in mammalian cells. Here we have focused on 25 °C as a useful temperature to modify purified proteins. The PolB16 intein is also fully orthogonal to the cysteine-less CL intein.<sup>12</sup> Furthermore, the PolB16 intein is the first atypically split cysteine-less intein. Its Int<sup>N</sup> fragment of only 15 aa is the shortest characterized so far<sup>28</sup> and enables simple accessibility of synthetic and chemically functionalized Int<sup>N</sup> precursors by SPPS for semi-synthetic protein splicing. Collectively, the PolB16 intein holds excellent potential for various protein engineering applications.

Our solved PolB16 intein crystal structures provide the first structural insight into a mesophilic cysteine-less split intein. A crystal structure of the thermophilic split Neq Pol intein was previously determined.<sup>48</sup> Interestingly, in both these structures, the sequence region surrounding the motif N3:10 histidine was missing, suggesting an unfolded character. Surprisingly, our mutational analysis revealed this histidine to be dispensable for catalysis in the PolB16 intein. These findings seem to show a structural and functional correlation for the PolB16 intein. A similar argument cannot be made for the Neq Pol intein as no mutational studies have been reported for the latter, perhaps because of its poor behavior in recombinant protein production and purification.<sup>16,48</sup> Typically, in other Ser1 inteins the motif N3 histidine is essential,<sup>17</sup> as we have also shown here for the newly characterized Mvu-M7 Pol-3 mini-intein.

The remainder of the conserved active site residues in the PolB16 intein as well as our structural and biochemical analysis strongly suggest this intein follows the regular oxyester intermediates of a class 1 protein splicing pathway. Other *cis*-inteins with Ser1/Ser+1(Thr+1) residues have previously been functionally characterized<sup>42</sup> and show similar active site arrangements, *i.e.*, the PI-TkoII intein<sup>51</sup> and the MjaTFIIB intein, respectively.<sup>52</sup>

Surprisingly, despite this large body of knowledge on both Ser1 and Cys1 inteins, we newly identified an NX motif with



a highly conserved histidine as the first signature sequence of the Ser1/Ser+1(Thr+1) subgroup of inteins. Its occurrence in inteins that utilize oxyester chemistry in their intermediates (and its absence in other inteins) has been overlooked so far. While in the PolB16 intein the motif NX histidine is critical and the motif N3 histidine is dispensable, the high conservation of both histidines in the Cys-independent Ser1 inteins suggests important roles for both these histidines in general (Fig. S7A and D†). This notion was further corroborated by the two other Ser1 inteins we have tested. Thus, the PolB16 intein seems to be a peculiar outlier with regard to the dispensable motif N3 histidine and represents another example for the tremendous diversity of variations of the protein splicing mechanism. The motif NX histidine appears to reflect the specialized environment evolved for the active site of inteins with a Ser1 residue.

Remarkably, the NX motif region has the same backbone fold as in highly diverse Cys1 inteins (Fig. S7C†), indicating that the NX histidine and other residues within the motif utilize the general intein fold to modify protein splicing catalysis for the special case of Ser1.

What is the contribution of the motif NX histidine to catalysis? Soon after the discovery of intein protein splicing<sup>53,54</sup> in 1990 their catalytic mechanism was predicted to include an active site histidine by analogy to the catalytic triad mechanism of cysteine and serine proteases.<sup>55,56</sup> The few known examples prompted the suggestion of the conserved penultimate histidine to be this residue. However, computational analysis of the then 10 known intein sequences showed a newly discovered conserved histidine of motif N3 (then termed block B) as a better candidate for the theoretical active site histidine.<sup>31</sup> This was later verified by a genetic screen and intein structure determination.<sup>57,58</sup> However, while some work points to the N3 histidine as a general base for the nucleophilic side chain from residue 1,<sup>38,59</sup> there is also ample evidence for an alternative role in destabilizing and twisting the upstream scissile bond.<sup>37,39,49,50</sup> This evidence includes the notion that the N3 histidine is typically facing the peptide backbone and not the position 1 side chain. Given the different nucleophilicities of the cysteine and serine side chains, a higher degree of activation would be required for a Ser1 residue. For example, Ser1 inteins could have evolved to rely on the assistance of a histidine as a general base histidine, while Cys1 inteins have not. Similarly, cysteine and serine hydrolases exhibit specialization with differently tuned catalytic dyads and triads, in which the attacking nucleophilic side chains are not simply interchangeable.<sup>60</sup>

Our PolB16 intein structures showed the motif NX histidine in possible hydrogen-bond forming distance to the Ser1, when the side chain hydroxyl of the latter is modeled back in (3.0 Å; Fig. 6). Similar observations can be made for the other known crystal structures of Ser1 inteins.<sup>48,51,52</sup> Therefore, it is most tempting to postulate that the motif NX histidine assists in the initial N–O acyl shift by polarizing the Ser1 hydroxyl group to facilitate the attack on the backbone carbonyl carbon of the upstream scissile bond. A similar suggestion was made by Ribo and co-workers for the corresponding His61 of the thermophilic Neq Pol intein, however, without performing any site-directed mutagenesis to substantiate the hypothesis.<sup>48</sup> Alternatively,

these histidines might deprotonate or help position the conserved motif C2 aspartate (Asp164 in the PolB16 intein in a distance of 3.3 Å), which is implicated in the N–S (or N–O) acyl shift in other inteins. Furthermore, it might be involved in the transesterification step by helping in the attack of the Ser+1/Thr+1 nucleophile through further polarization of the Ser1 ester carbonyl group. The lack of the motif NX histidine in the GOS TerL intein,<sup>20</sup> the only known Ser1 intein with a more reactive Cys+1 nucleophile that may be less dependent on such additional catalysis, would be consistent with this latter idea. Future work is required to study the exact catalytic role of the motif NX histidine.

In summary, the cysteine-less split PolB16 intein holds great potential for versatile splicing applications. Our identification of the motif NX histidine helps explain the specialized catalysis of Ser1/Ser+1(Thr+1) inteins. It might also enable new approaches to engineer powerful, new cysteine-less inteins. In general, we expect cysteine-less inteins to prove very useful in the future for applications to be performed in the absence of reducing agents or in combination with other thiol-based chemistry schemes that could not be carried out with Cys-dependent inteins or alternative cysteine-dependent protein ligation tools.<sup>61,62</sup>

## Data availability

ESI is available comprising additional experimental and sequence data. Coordinates of the crystal structures of the wildtype and C111A/C165A mutant of the PolB16 intein have deposited at the Protein Data Bank under pdb accessions numbers 8CPN and 8CPO, respectively. Further original data is available from the authors upon request.†

## Author contributions

Conceptualization: T. P., S. P., H. D. M.; formal analysis: T. P., A. S., S. K., D. K., H. D. M.; funding acquisition: H. D. M.; investigation: T. P., A. S., S. K., M. E., S. P., H. D. M.; project administration: S. P., D. K., H. D. M.; supervision: H. D. M.; writing: T. P., S. P., D. K. and H. D. M.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We thank Stephan Wilmes for help with processing the structural data and Francine Perler (New England Biolabs) for providing a plasmid encoding the Psp GBD-Pol intein. We gratefully acknowledge partial funding of this project by the DFG (project grant MO1073/7-1 to H. D. M.).

## Notes and references

- 1 N. H. Shah and T. W. Muir, *Chem. Sci.*, 2014, 5, 446–461.
- 2 H. D. Mootz, *Chembiochem*, 2009, 10, 2579–2589.





- 3 A. S. Aranko, A. Wlodawer and H. Iwai, *Protein Eng., Des. Sel.*, 2014, **27**, 263–271.
- 4 N. I. Topilina and K. V. Mills, *Mobile DNA*, 2014, **5**, 5.
- 5 B. Di Ventura and H. D. Mootz, *Biol. Chem.*, 2019, **400**, 467–475.
- 6 R. E. Thompson and T. W. Muir, *Chem. Rev.*, 2020, **120**, 3051–3126.
- 7 D. W. Wood and J. A. Camarero, *J. Biol. Chem.*, 2014, **289**, 14512–14519.
- 8 C. Ludwig, M. Pfeiff, U. Linne and H. D. Mootz, *Angew. Chem., Int. Ed. Engl.*, 2006, **45**, 5218–5221.
- 9 B. M. Lew, K. V. Mills and H. Paulus, *J. Biol. Chem.*, 1998, **273**, 15887–15890.
- 10 J. H. Appleby, K. Zhou, G. Volkmann and X. Q. Liu, *J. Biol. Chem.*, 2009, **284**, 6194–6199.
- 11 I. Giriat and T. W. Muir, *J. Am. Chem. Soc.*, 2003, **125**, 7180–7181.
- 12 M. Bhagawati, T. M. E. Terhorst, F. Fusser, S. Hoffmann, T. Pasch, S. Pietrokovski and H. D. Mootz, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 22164–22172.
- 13 T. Kurpiers and H. D. Mootz, *Angew. Chem., Int. Ed. Engl.*, 2007, **46**, 5234–5237.
- 14 A. L. Bachmann and H. D. Mootz, *J. Pept. Sci.*, 2017, **23**, 624–630.
- 15 S. Brenzel, M. Cebi, P. Reiss, U. Koert and H. D. Mootz, *Chembiochem*, 2009, **10**, 983–986.
- 16 J. J. Choi, K. H. Nam, B. Min, S. J. Kim, D. Soll and S. T. Kwon, *J. Mol. Biol.*, 2006, **356**, 1093–1106.
- 17 M. W. Southworth, E. Adam, D. Panne, R. Byer, R. Kautz and F. B. Perler, *Embo J.*, 1998, **17**, 918–926.
- 18 X. Qi, J. Wang, Q. Meng and X. Q. Liu, *Protein Pept. Lett.*, 2011, **18**, 1226–1232.
- 19 J. Zettler, V. Schütz and H. D. Mootz, *FEBS Lett.*, 2009, **583**, 909–914.
- 20 A. L. Bachmann and H. D. Mootz, *J. Biol. Chem.*, 2015, **290**, 28792–28804.
- 21 K. V. Mills, B. M. Lew, S. Jiang and H. Paulus, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 3543–3548.
- 22 W. Sun, J. Yang and X. Q. Liu, *J. Biol. Chem.*, 2004, **279**, 35281–35286.
- 23 S. Brenzel, T. Kurpiers and H. D. Mootz, *Biochemistry*, 2006, **45**, 1571–1578.
- 24 H. Iwai, S. Zuger, J. Jin and P. H. Tam, *FEBS Lett.*, 2006, **580**, 1853–1858.
- 25 P. Carvajal-Vallejos, R. Pallisse, H. D. Mootz and S. R. Schmidt, *J. Biol. Chem.*, 2012, **287**, 28686–28696.
- 26 W. Shi, C. D. Moon, S. C. Leahy, D. Kang, J. Froula, S. Kittelmann, C. Fan, S. Deutsch, D. Gagic, H. Seedorf, W. J. Kelly, R. Atua, C. Sang, P. Soni, D. Li, C. S. Pinares-Patino, J. C. McEwan, P. H. Janssen, F. Chen, A. Visel, Z. Wang, G. T. Attwood and E. M. Rubin, *Genome Res.*, 2014, **24**, 1517–1525.
- 27 I. V. Thiel, G. Volkmann, S. Pietrokovski and H. D. Mootz, *Angew. Chem., Int. Ed. Engl.*, 2014, **53**, 1306–1310.
- 28 M. Neugebauer, J. K. Bocker, J. C. Matern, S. Pietrokovski and H. D. Mootz, *Biol. Chem.*, 2017, **398**, 57–67.
- 29 K. Tori, B. Dassa, M. A. Johnson, M. W. Southworth, L. E. Brace, Y. Ishino, S. Pietrokovski and F. B. Perler, *J. Biol. Chem.*, 2010, **285**, 2515–2526.
- 30 S. Pietrokovski, *Protein Sci.*, 1998, **7**, 64–71.
- 31 S. Pietrokovski, *Protein Sci.*, 1994, **3**, 2340–2350.
- 32 Y. Shao and H. Paulus, *J. Pept. Res.*, 1997, **50**, 193–198.
- 33 J. Binschik, J. Zettler and H. D. Mootz, *Angew. Chem., Int. Ed. Engl.*, 2011, **50**, 3249–3252.
- 34 S. Frutos, M. Goger, B. Giovani, D. Cowburn and T. W. Muir, *Nat. Chem. Biol.*, 2010, **6**, 527–533.
- 35 D. D. Martin, M. Q. Xu and T. C. Evans Jr, *Biochemistry*, 2001, **40**, 1393–1402.
- 36 J. C. J. Matern, K. Friedel, J. Binschik, K. S. Becher, Z. Yilmaz and H. D. Mootz, *J. Am. Chem. Soc.*, 2018, **140**, 11267–11275.
- 37 A. Romanelli, A. Shekhtman, D. Cowburn and T. W. Muir, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 6397–6402.
- 38 Z. Du, P. T. Shemella, Y. Liu, S. A. McCallum, B. Pereira, S. K. Nayak, G. Belfort, M. Belfort and C. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 11581–11589.
- 39 J. Binschik and H. D. Mootz, *Angew. Chem., Int. Ed. Engl.*, 2013, **52**, 4260–4264.
- 40 A. K. Dearden, B. Callahan, P. V. Roey, Z. Li, U. Kumar, M. Belfort and S. K. Nayak, *Protein Sci.*, 2013, **22**, 557–563.
- 41 F. B. Perler, E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner and M. Belfort, *Nucleic Acids Res.*, 1994, **22**, 1125–1127.
- 42 M. Q. Xu, M. W. Southworth, F. B. Mersha, L. J. Hornstra and F. B. Perler, *Cell*, 1993, **75**, 1371–1377.
- 43 F. B. Perler, *Nucleic Acids Res.*, 2002, **30**, 383–384.
- 44 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583–589.
- 45 Y. David, M. Vila-Perello, S. Verma and T. W. Muir, *Nat. Chem.*, 2015, **7**, 394–402.
- 46 M. Bhagawati, S. Hoffmann, K. S. Hoffgen, J. Piehler, K. B. Busch and H. D. Mootz, *Angew. Chem., Int. Ed. Engl.*, 2020, **59**, 21007–21015.
- 47 T. Dhar and H. D. Mootz, *Chem. Commun.*, 2011, **47**, 3063–3065.
- 48 V. Gordo, D. Aparicio, R. Perez-Luque, A. Benito, M. Vilanova, I. Uson, I. Fita and M. Ribo, *Cell. Chem. Biol.*, 2018, **25**, 871–879.
- 49 K. Tori, M. Cheriyan, C. S. Pedamallu, M. A. Contreras and F. B. Perler, *Biochemistry*, 2012, **51**, 2496–2505.
- 50 K. Friedel, M. A. Popp, J. C. J. Matern, E. M. Gazdag, I. V. Thiel, G. Volkmann, W. Blankenfeldt and H. D. Mootz, *Chem. Sci.*, 2019, **10**, 239–251.
- 51 H. Matsumura, H. Takahashi, T. Inoue, T. Yamamoto, H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka and Y. Kai, *Proteins*, 2006, **63**, 711–715.



- 52 H. Iwai, K. M. Mikula, J. S. Oeemig, D. Zhou, M. Li and A. Wlodawer, *J. Mol. Biol.*, 2017, **429**, 3942–3956.
- 53 P. M. Kane, C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebel and T. H. Stevens, *Science*, 1990, **250**, 651–657.
- 54 R. Hirata, Y. Ohsumk, A. Nakano, H. Kawasaki, K. Suzuki and Y. Anraku, *J. Biol. Chem.*, 1990, **265**, 6726–6733.
- 55 R. A. Hodges, F. B. Perler, C. J. Noren and W. E. Jack, *Nucleic Acids Res.*, 1992, **20**, 6153–6157.
- 56 C. J. Wallace, *Protein Sci.*, 1993, **2**, 697–705.
- 57 M. Kawasaki, S. Nogami, Y. Satow, Y. Ohya and Y. Anraku, *J. Biol. Chem.*, 1997, **272**, 15668–15674.
- 58 X. Duan, F. S. Gimble and F. A. Quiocho, *Cell*, 1997, **89**, 555–564.
- 59 Z. Du, Y. Zheng, M. Patterson, Y. Liu and C. Wang, *J. Am. Chem. Soc.*, 2011, **133**, 10275–10282.
- 60 Y. Cen, W. Singh, M. Arkin, T. S. Moody, M. Huang, J. Zhou, Q. Wu and M. T. Reetz, *Nat. Commun.*, 2019, **10**, 3198.
- 61 H. Mao, S. A. Hart, A. Schink and B. A. Pollok, *J. Am. Chem. Soc.*, 2004, **126**, 2670–2671.
- 62 G. K. Nguyen, S. Wang, Y. Qiu, X. Hemu, Y. Lian and J. P. Tam, *Nat. Chem. Biol.*, 2014, **10**, 732–738.

