FEBS Letters

FEBS PRESS science publishing by scientists

RESEARCH LETTER

The three-dimensional structure of the Vint domain from *Tetrahymena thermophila* suggests a ligand-regulated cleavage mechanism by the HINT fold

Hideo Iwaï[1], Hannes M. Beyer[1,*], Julia E. M. Johansson[1], Mi Li[2,3] and Alexander Wlodawer[2]

1 Institute of Biotechnology, University of Helsinki, Finland
2 Center for Structural Biology, National Cancer Institute, Frederick, MD, USA
3 Basic Science Program, Frederick National Laboratory for Cancer Research, MD, USA

Correspondence
H. Iwaï, Structural Biology and Analytics, Orion Pharma, Orionintie 1, FI-02100 Espoo, Finland
Tel: +358 50 966 2493
E-mail: hideo.iwai@orionpharma.com

Present address
*Institute of Synthetic Biology, Heinrich-Heine-University Düsseldorf, Germany

(Received 12 November 2023, revised 13 December 2023, accepted 3 January 2024, available online 13 February 2024)

doi:10.1002/1873-3468.14817

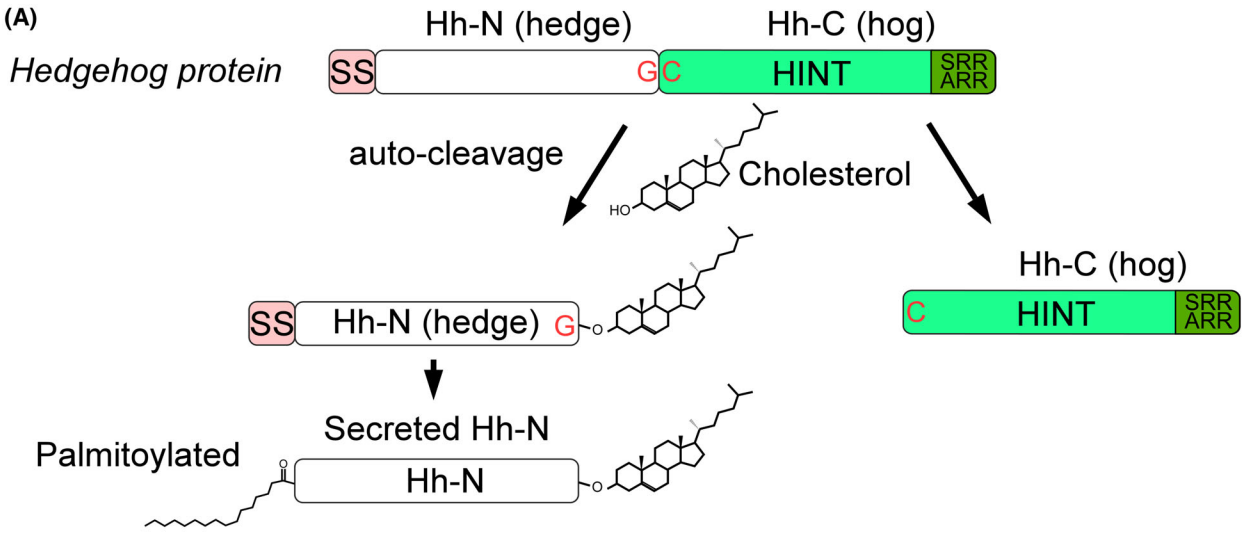
Edited by Hee-Jung Choi

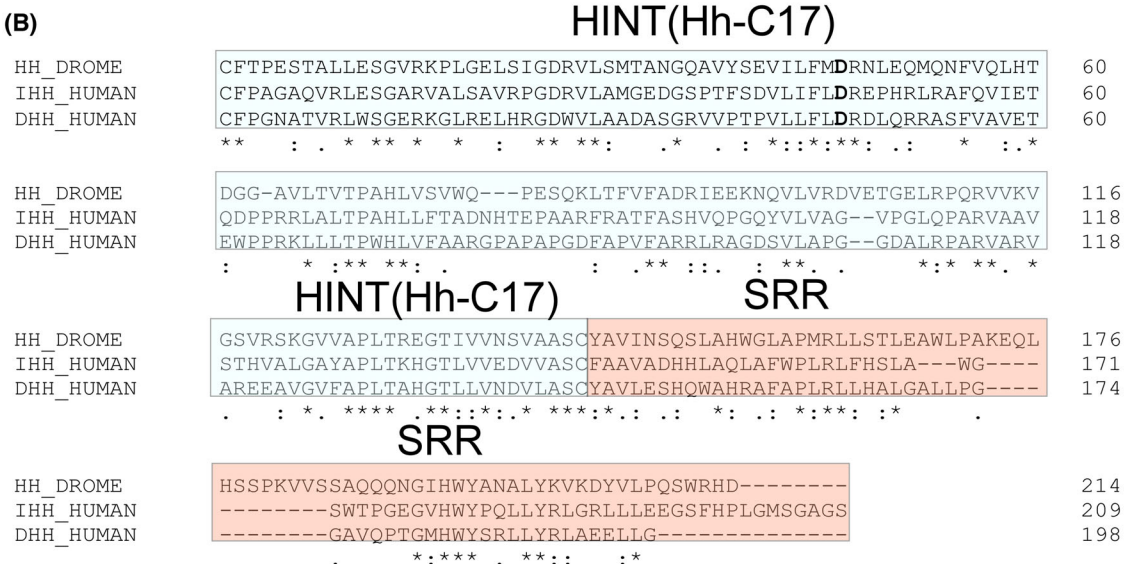
Vint proteins have been identified in unicellular metazoans as a novel hedgehog-related gene family, merging the von Willebrand factor type A domain and the Hedgehog/INTein (HINT) domains. We present the first three-dimensional structure of the Vint domain from *Tetrahymena thermophila* corresponding to the auto-processing domain of hedgehog proteins, shedding light on the unique features, including an adduct recognition region (ARR). Our results suggest a potential binding between the ARR and sulfated glycosaminoglycans like heparin sulfate. Moreover, we uncover a possible regulatory role of the ARR in the auto-processing by Vint domains, expanding our understanding of the HINT domain evolution and their use in biotechnological applications. Vint domains might have played a crucial role in the transition from unicellular to multicellular organisms.

Keywords: hedgehog; HINT; inteins; Vint; von Willebrand factor type A domain

Protein-splicing domains catalyzing self-excision and concomitant ligation of the flanking protein sequence are termed inteins, *int*ervening prot*ein*s [1,2]. The catalytic domain of protein splicing has a common HINT (Hedgehog/INTein) fold, sharing the same fold between the C-terminal auto-processing domain of hedgehog and inteins [3]. Inteins are prevalent across three kingdoms of life but found only in unicellular organisms [4]. In contrast, the hedgehog (Hh) protein is a signaling molecule that plays a vital role in intercellular communications during developmental processes, organ patterning, tissue and stem cell maintenance, and cell differentiation in multi-cellular organisms [5]. The aberrant Hh signaling is also linked to various diseases, including developmental defects and cancer [6]. The full-length Hh protein is a 45-kDa precursor protein composed of two domains: an N-terminal domain of 19 kDa (Hh-N, hedge) comprising the signaling activity and a catalytic auto-processing domain of 25 kDa (Hh-C, hog) with structural similarity to inteins (Fig. 1) [3,7]. The precursor hedgehog protein undergoes auto-processing through self-cleavage, followed by signal sequence cleavage and entry into the secretory pathway. Hh-C is the auto-processing domain responsible for both the peptide bond cleavage and the attachment of a cholesterol moiety to its carboxyl terminus (Fig. 1A) [8]. In contrast to inteins, Hh-C (hog) does not catalyze protein splicing but contains an additional sterol-recognition region (SRR) at the C-terminus of the catalytic HINT domain,

Abbreviations
ARR, adduct-recognition region; HINT, Hedgehog/INTein; IMAC, immobilized metal ion affinity chromatography; IPTG, isopropyl-β-D-1-thiogalactopyranoside; PDB, Protein Data Bank; RMSD, root mean square deviation; SRR, sterol-recognition region; SUMO, yeast SMT3 protein; Ulp1, ubiquitin-like-specific protease 1; vWA, von Willebrand factor type A domain.





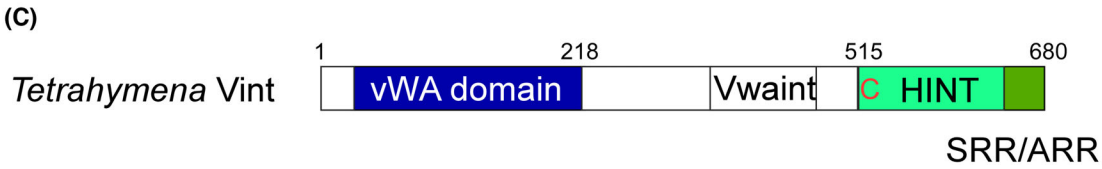


Fig. 1. The domain structures of hedgehog protein and *Tetrahymena* Vint protein. (A) Hh protein undergoes lipid modification steps to become a mature Hh-N protein. The C-terminal auto-processing domain (Hh-C) catalyzes the attachment of cholesterol moiety at the C-terminus of Hh-N and the concomitant N-terminal cleavage at the N-terminal end of Hh-C. (B) The sequence alignment of Hh-C from Drosophila and human hedgehog proteins. Hh-C is the auto-processing catalytic domain having the HINT fold. The SRR of Hh-C highlighted in pink indicates the presumable cholesterol-binding region of which the three-dimensional structure is unknown. (C) The domain structure of the Vint protein in *Tetrahymena thermophila*. The Vint (vWA domain and Hint domain proteins) consists of a vWA domain (von Willebrand factor type A domain), a new domain between the vWA and Vint domain termed Vwaint, and intein-like domain (HINT) followed by a putative sterol/adduct-recognition region (SRR/ARR).

which recognizes cholesterol molecules for their attachment to Hh-N [3–10] (Fig. 1A). The structural basis of the cholesterol attachment to Hh-N *via* the SRR has yet to be delineated [11,12].

Hh-N domains of hedgehog proteins are highly conserved signaling molecules that are crucial in body organization. Nematodes lack a bonafide hedgehog protein but contain hedgehog-related proteins with auto-processing domains [8–10]. Extensive genomic sequence analysis identified Hh-related proteins containing the von Willebrand factor type A domain (vWA) fused to a HINT domain distinct from bacterial-intein-like (BIL) proteins and inteins [9,10]. This specific group of HINT domain-containing Hh-related proteins was termed Vint (*V*WA domain and the H*int*-like domain) proteins [9]. The novel HINT domain associated with vWA domains was correspondingly identified in tetrahymena, fungi, and several other protists [9,10]. The HINT-containing Vint domain also contains a C-terminal putative adduct-recognition region (ARR), reminiscent of the SRR in hedgehog proteins. However, its role in a possible adduct addition reaction remains unknown (Fig. 1) [9,10].

We report here the first three-dimensional structure of the Vint domain, including the ARR, from *Tetrahymena thermophila* (*Tth*Vint). The three-dimensional structure of the Vint domain revealed the first atomic structure of the adduct/sterol recognition region (ARR/SRR), widely identified among Hh-related and hedgehog proteins. We further investigated the possible functional role of the Vint domain in Hh-related proteins, which might include a regulatory function of auto-processing by the Vint domain. The regulation by the ARR could potentially provide new insights into the functional roles of the Hint superfamily and a unique opportunity to control the auto-processing of the Vint domain as a biotechnological tool.

Methods

Construction of vectors bearing the Vint domain from *T. thermophila* SB210

The gene encoding the HINT domain of the Vint protein (residues 512–680) from *T. thermophila* SB210 (ciliates) was codon-optimized for *Escherichia coli* expression based on the protein sequence deduced from TTHERM_00471620 using the genetic code of tetrahymena. This step was necessary because of the alternative codon usage in *T. thermophila* [13]. The codon-optimized DNA for *E. coli* expression was chemically synthesized and cloned into the pIDTSMART-KAN vector, which was purchased from Integrated DNA Technologies Inc. (Leuven, Belgium). We fused the N-terminal yeast SMT3 (SUMO) protein with the Vint (512–680) to facilitate protein purification [14]. All the variants with various mutations and deletions were constructed and summarized in Table S1.

Protein expression and purification of the Vint domain

The Vint domain (512–680) with C514A (C1A) mutation was expressed as an N-terminally His-tagged SUMO-fusion protein in the *E. coli* strain T7 Express (NEB, Ipswich, MA, USA) [14]. The cells were transformed with plasmid pBHRSF123 (Table S1) and grown at 37 °C in 2 L of LB medium supplemented with a final concentration of 25 $\mu g \cdot mL^{-1}$ kanamycin. The SUMO-fusion protein was induced with a final concentration of 1 mM isopropyl-β-D-thiogalactoside (IPTG) for 2.5 h when OD_{600} reached 0.6. The induced cells were harvested by centrifugation at 4700 ***g*** for 10 min, 4 °C and lysed in 20 mL Buffer A (50 mM sodium phosphate, pH 8.0, 300 mM NaCl) using continuous passaging through an EmulsiFlex-C3 homogenizer (AVESTIN, Ottawa, ON, Canada) at 15 000 psi for 10 min at 4 °C. Lysates were cleared by centrifugation at 38 000 ***g*** for 60 min at 4 °C and loaded on a 5-mL HisTrap HP column (Cytiva, Marlborough, MA, USA). The column was extensively washed with 50 mM imidazole in Buffer A and eluted with a linear imidazole gradient to 300 mM in Buffer A. The elution fractions containing the SUMO-fusion protein were collected and dialyzed overnight against 2 L phosphate-buffered saline. The SUMO-fusion protein was digested with yeast ubiquitin-like specific protease (Ulp1) at 30 °C for 4 h [14]. The reaction mixture was again applied to a 5-mL HisTrap HP column to remove the cleaved SUMO domain, Ulp1 protease, and the uncleaved fusion protein. The flow-through fractions from immobilized metal ion affinity chromatography (IMAC) were dialyzed against 2 L of Milli-Q water, concentrated to 14.6 $mg \cdot mL^{-1}$ for crystallization, and stored at −70 °C.

Protein expression of Vint variants

Small-scale expressions of the SUMO-fusion proteins with various mutations were performed as described above in 5 mL LB medium cultures of *E. coli* T7 Express (NEB) using the plasmids listed in Table S1. The cells were harvested by centrifugation and lysed with B-PER Bacterial Protein Extraction Reagent (Thermo Fisher Scientific, Waltham, MA, USA). The supernatant was loaded after the centrifugation onto a Ni-NTA spin column (Qiagen, Hilden, Germany) and eluted with 250 mM imidazole in Buffer A [15].

Crystallization of the Vint_C1A domain and diffraction data collection and processing

Diffracting crystals for Vint_C1A were obtained at room temperature by mixing 100 nL concentrated protein

(14.6 mg·mL^{-1}) with 100 nL mother liquid (2% PEG 400, 2.4 M ammonium sulfate, 0.1 M HEPES pH 6.8). Glycerol was used as a cryoprotectant. Diffraction data were collected at beamline ID23-1 at ESRF (Grenoble, France), which was equipped with a Pilatus detector. Data were processed with XDS using AUTOPROC at the nominal resolution of 1.80 Å (Table S2). An initial model was obtained using the PHASER molecular replacement protocol, with the solution structure of the bacterial intein-like domain from (PDB: 2LWY) selected as a search model [16,17]. The model was further improved using PHENIX AUTOBUILD and COOT and refined to 1.80 Å using PHENIX [18,19]. The final model was built with COOT, followed by rounds of refinement using REFMAC5 and PHENIX [18,20]. The entire polypeptide chain could be traced into the electron density map, although the linker connecting two β-strands in the ARR was traced only approximately. The quality of the final structure was validated using the MolProbity web server [21]. The coordinates and structure factors have been deposited in the Protein Data Bank with accession number 8R2C for *Tth*Vint.

Results

The three-dimensional structure of *Tth*Vint

Vint domains were identified as a novel HINT gene family in many eukaryotes encoding a vWA domain fused to a distinct HINT domain [9]. As the first step toward understanding the biological function of Vint proteins and the evolutionary relationship with inteins and hedgehog proteins, we determined the three-dimensional structure of the Vint domain by X-ray crystallography at the 1.8-Å resolution (Fig. 1).

Vint proteins were originally identified by their sequence homology to hedgehog proteins. Consequently, the structure of *Tth*Vint domain also showed the highest structural similarity to Hh-C (PDB code: 1AT0) from *Drosophila melanogaster* with an RMSD of 2.0 Å for 105 residues of Hh-C with a *Z*-score of 14.0, as calculated by the Dali server [22]. BIL domain (PDB code: 2LWY [16]) had the second-highest *Z*-score of 13.6 for 109 residues, confirming that the Vint domain is structurally related to the HINT superfamily. Interestingly, the HINT domain of *Tth*Vint is even smaller than the smallest reported intein (120 residues) and has a distinct additional region at the C-terminus, presumably recognizing an adduct [10]. The ARR of *Tth*Vint corresponds to the SRR of Hh-C, of which no structural information exists in the previously reported three-dimensional structure of Drosophila Hh-C [7]. The ARR of Vint domain is located on the ventral side of the horseshoe-crab shape of the HINT fold, where the N-terminus of Vint is also on the same side [23] (Fig. 2A). The ARR consists of a single α-helix and three β-strands. The three-dimensional structure of the ARR from *Tth*Vint is distinct from the structural models of the SRR of hedgehog proteins, which feature α-helical domains constituting the sterol-binding site [12,24]. The ARR structure implies that the adduct is unlikely a sterol molecule as in hedgehog proteins. It is noteworthy that we could model five sulfate ions present in the crystallization medium into the clear electron density (Fig. 3A).

The interaction with the ARR in the Vint domain

The observation of sulfate ions at the interface between ARR and the HINT structure prompted us to speculate on the identity of potential naturally occurring adducts or molecules recognized by the ARR of Vint proteins (Fig. 3A). Vint proteins were identified as associated with the vWA domain, a well-characterized domain family that plays a significant role in cell adhesion within extracellular matrix proteins [25]. Therefore, we decided to use a heparin sepharose column to examine the potential interactions between the Vint domain and heparin sulfate, one of the glycosaminoglycans (GAGs) in cell surfaces and the extracellular matrix (Fig. 3B). Indeed, we observed the binding of the Vint domain to a heparin column, which could be effectively eluted using a sodium chloride gradient. To confirm the binding, we mutated Arg154, a residue involved in the observed interactions in the crystal structure with sulfate ions, to Ser. Indeed, we observed that the R154S variant eluted at a lower salt concentration from the heparin column (Fig. 3B). The R152M mutation further reduced binding to the heparin column (Fig. 3B). The mutational studies on the interacting residues with sulfate ions provided strong evidence supporting an interaction between the Vint domain and heparin. Heparin or sulfated glycosaminoglycans could thus serve as naturally occurring interacting partners, possibly engaging in an adduct reaction.

The functional role of the ARR in the Vint domain

For protein expression in *E. coli*, we fused the Vint domain to a SUMO domain containing an N-terminal hexahistidine tag with a three-residue wild-type sequence at the N-terminal junction preceding the Vint domain (Fig. 4A) [14,26]. The expressed fusion protein immediately underwent self-cleavage *via* N-S acyl shift at the first Cys residue, yielding two bands in an SDS-PAGE gel used to monitor the overexpression in *E. coli* (Fig. 4B) [26]. This observation indeed suggests that the Vint domain is an auto-processing domain,

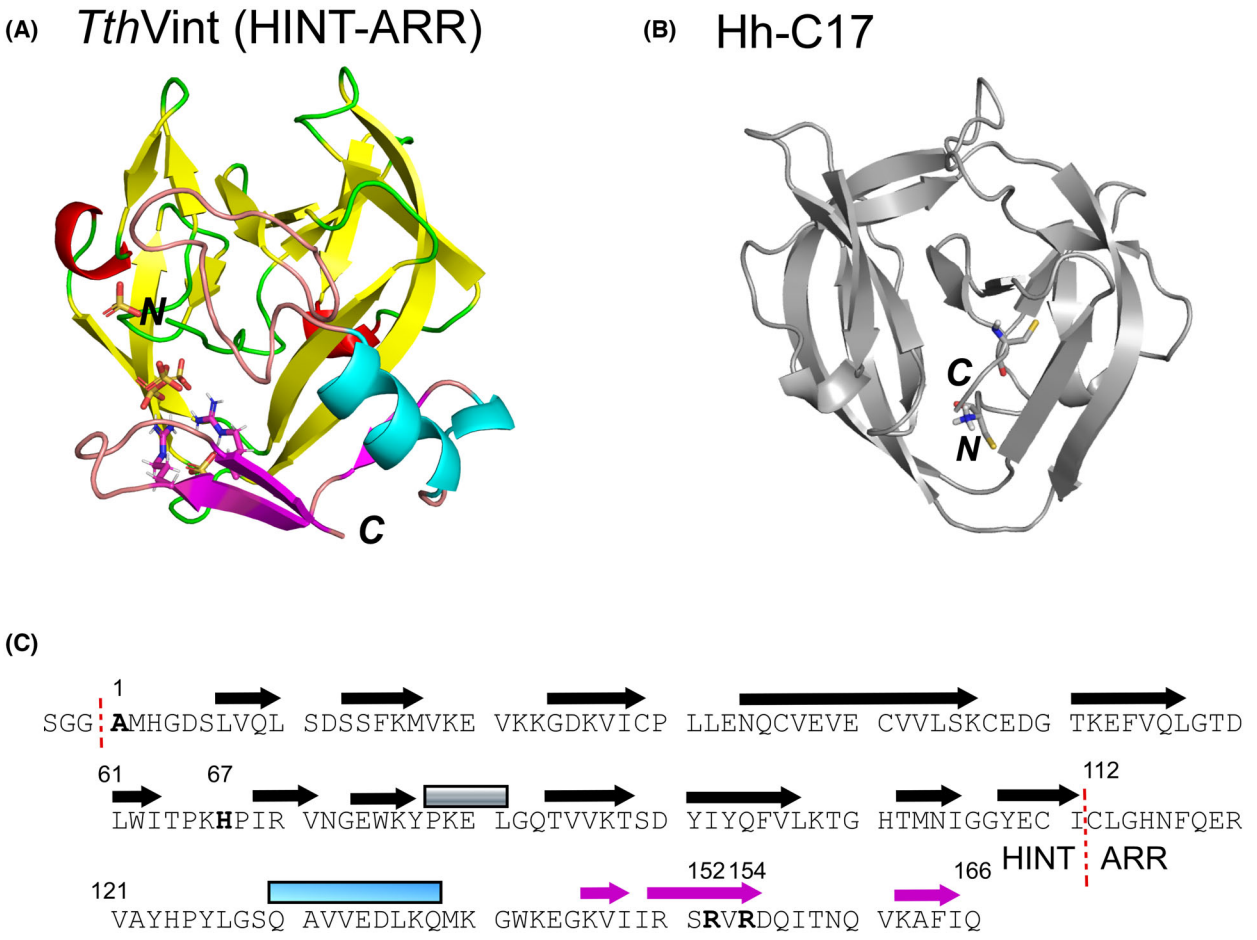


Fig. 2. The three-dimensional structures of *Tth*Vint and Drosophila Hh-C. (A) A cartoon model of *Tth*Vint. β-sheets and α-helices of the HINT fold are shown in yellow and red, respectively. The helix and β-sheets in the ARR/SRR are shown in cyan and magenta, respectively. The modeled four sulfate ions are displayed as stick models with two interacting arginine residues. N and C denote the N- and C-termini, respectively. (B) Cartoon representation of the three-dimensional structure of Drosophila Hh-C without SRR. N and C denote the N- and C-termini, respectively. (C) The secondary structures of the Vint domain along the amino acid sequence. The first residue of the catalytic cysteine of the HINT fold was mutated to Ala (C1A). The active site His residue, the first Ala residue, and two Arg residues interacting with sulfate ions are in bold. Vertical broken lines indicate the boundaries of the HINT fold. The numbers above the amino acid sequence indicate the number from the first residue of the HINT fold. (D) The structure-based sequence alignment between Hh-C (PDB: 1AT0) and Tetrahymena Vint domain.

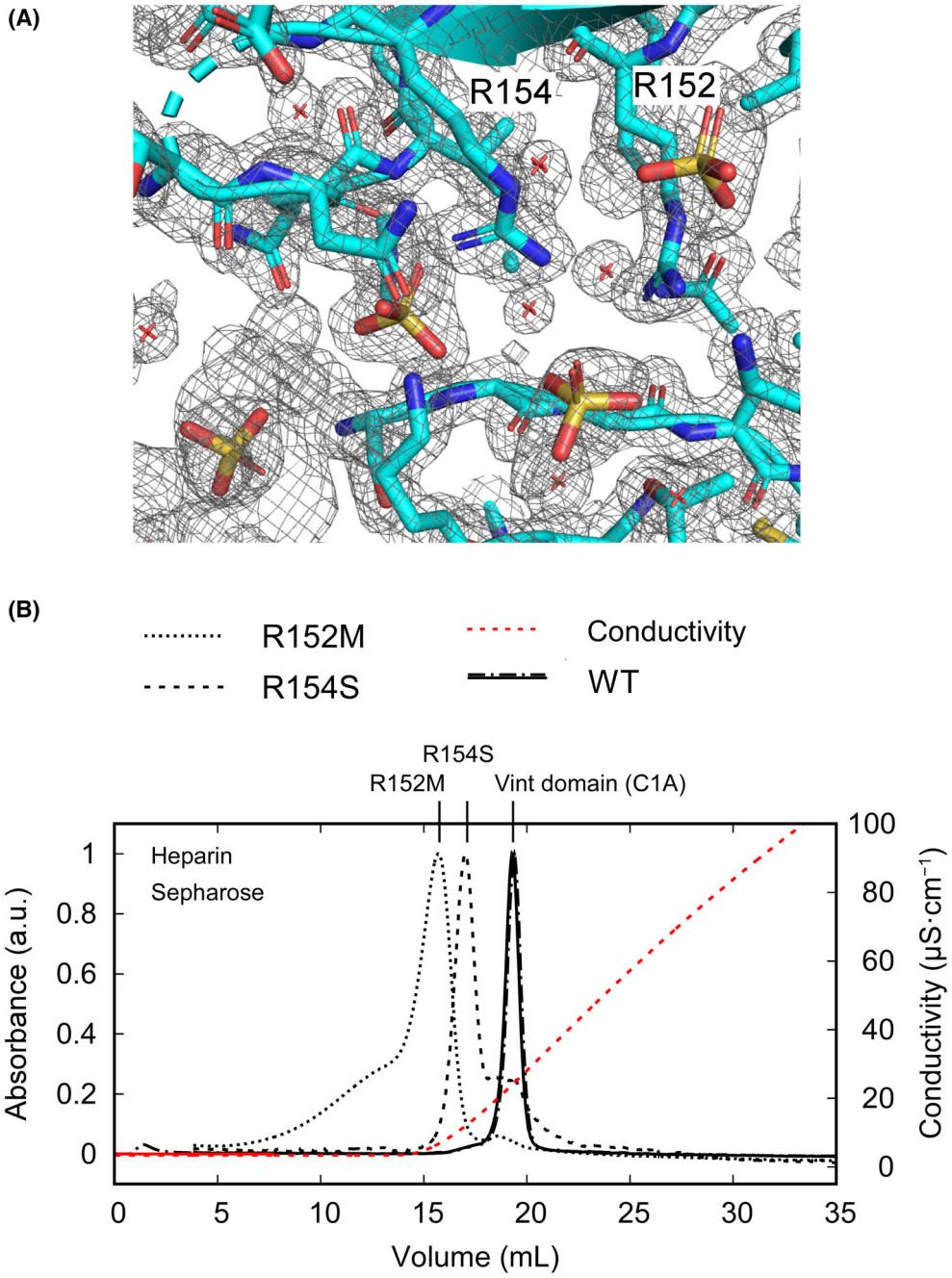


Fig. 3. Interactions between SRR/ARR and sulfate ions. (A) Electron density map around modeled sulfate ions. Residues R152 and R154, which were tested for their effect on the binding to a heparin sepharose column, are labeled. (B) Elution profiles of the Vint domain (C1A) and two variants bearing additional R152M and R154S mutations from a heparin sepharose column by a linear NaCl concentration gradient.

catalyzing an effective self-cleavage reaction in the tested construct. We then examined the active site residues in the HINT domain. In class 1 inteins and BILs, the first residue of the HINT domain is typically either cysteine or serine, involved in the initial N-S acyl shift [16,27]. This N-S acyl migration forms a thioester bond that can be cleaved or used for ligation *via* a nucleophile. To validate the auto-processing mechanism further, we mutated the first Cys residue (Cys515) to Ala, aiming to hinder the N-S acyl shift of the Vint domain (Fig. 4). As anticipated, the C1A mutation prevented self-cleavage during protein expression.

Consequently, the fusion protein bearing Vint with C1A mutation was purified without cleavage. This result

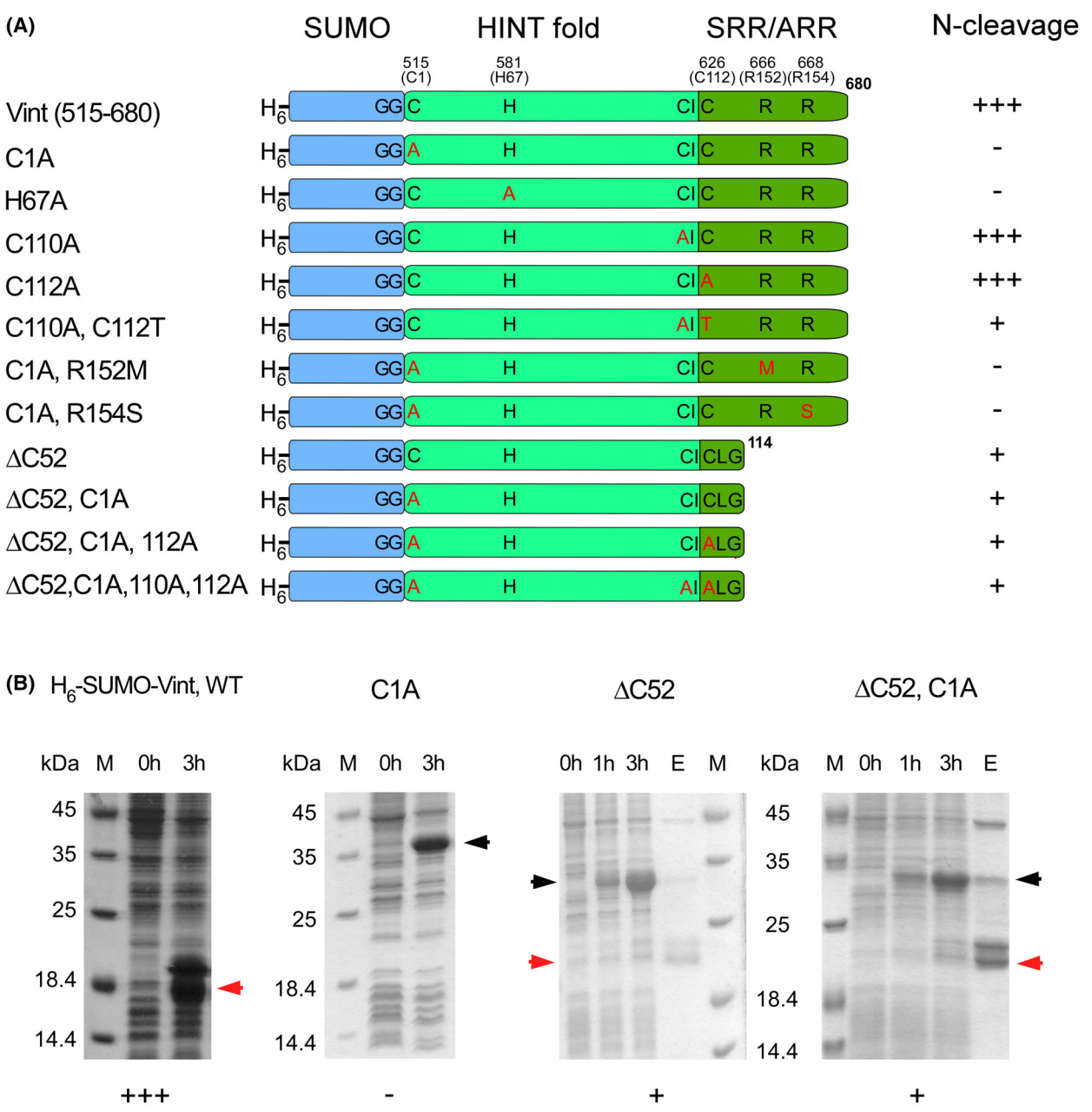


Fig. 4. The effect of mutations in the Vint domain. (A) A summary of the impact of various mutations and a deletion on the N-cleavage reactions by the Vint domain. Residues 514–680 of *Tth*Vint were fused with a yeast SUMO domain at the N-terminus with an N-terminal hexahistidine tag. The mutations are indicated in red. The residue numbers of Vint domains are shown. The numberings from the first Cys residue of the HINT fold used in this article are shown in brackets. ΔC52 indicates the deletion of the SRR/ARR after residue 114 from the Vint domain. "+ + +" designates observed immediate N-cleavage during the protein expression. Variants with no detectable cleavage are indicated by "−", whereas "+" represents the variants with weak cleavage during protein purification using the N-terminal His-tag. (B) SDS-PAGE analysis of the wild-type, the C1A variant, the deletion variant (ΔC52), and the deletion variant with C1A mutation (ΔC52 C1A). M, 0 h, 3 h, and E indicate molecular weight markers, before induction, 3 h after the protein induction, and the elution fraction from the Ni-NTA spin column, respectively. Arrowheads indicate the bands for the precursor protein. Red arrows indicate the bands of the cleaved Vint domain.

aligns with published reports on the N-S acyl shift as the initial reaction step during the auto-processing in class 1 inteins and BIL domains [1,16]. Next, we focused on His67, corresponding to the most conserved histidine in block B of inteins [28]. As anticipated, substituting His67 with alanine (H67A) abolished the self-cleavage

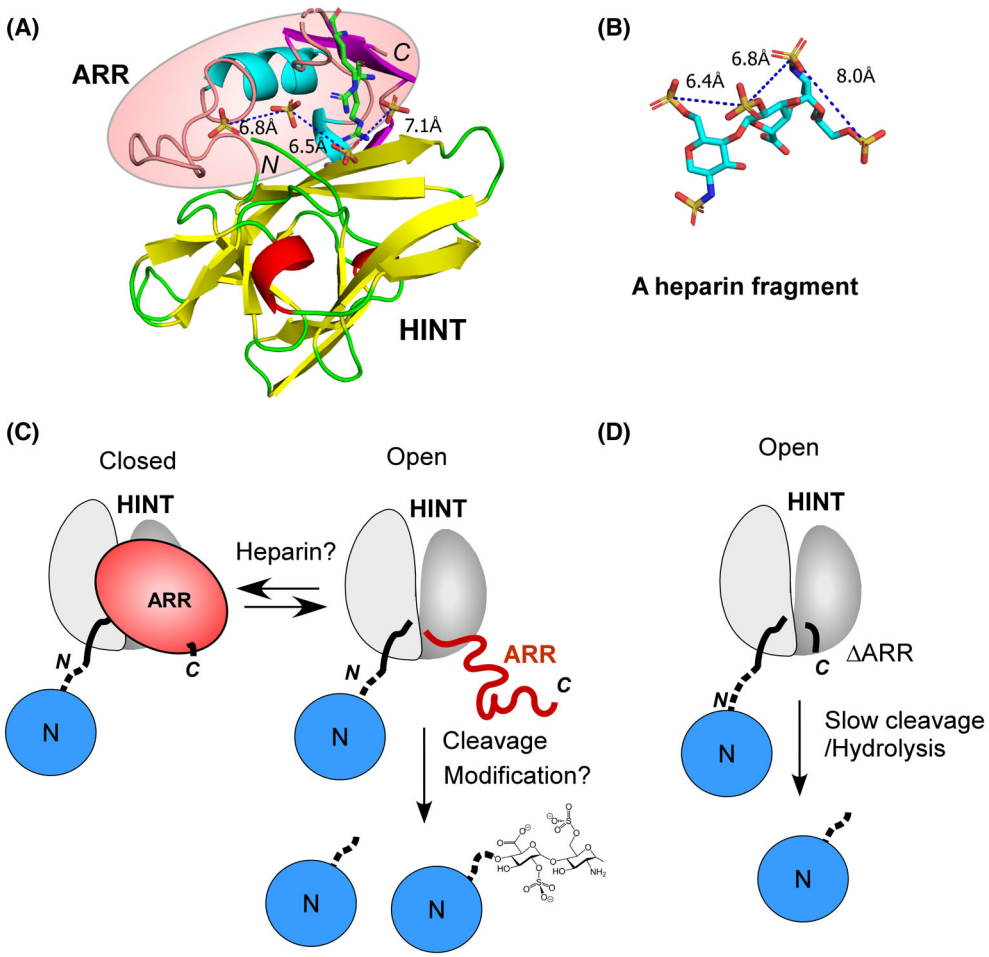


Fig. 5. A hypothetical model for regulating the auto-processing of the Vint protein by the HINT domain and the ARR/SRR. (A) The three-dimensional structure of the Vint domain with ARR. The ARR is highlighted with a pink oval. R152, R154, and three sulfate ions with the distances between them are shown as stick models. The β-sheet and α-helical turn of the HINT domain are shown as cartoons in yellow and red, respectively. The β-sheet and α-helix of ARR are colored in magenta and cyan, respectively. N and C indicate the N- and C-termini, respectively. (B) The solution structure of a heparin fragment (PDB: 3IRJ) is shown as a stick model with the distances between sulfur atoms. (C) A hypothetical regulation mechanism of the auto-processing of Vint proteins by adducts. (D) A possible unregulated cleavage by the ARR deletion.

of the Vint fusion protein (Fig. 4 and Fig. S1). However, when we introduced a single mutation at one of the two cysteine residues near the C-terminal end of the Vint domain, we observed no apparent effect on the auto-processive N-cleavage (Fig. 4). This finding implies that the side-chain thiol groups in these residues are not directly involved in the N-cleavage reaction.

Intriguingly, when we removed the ARR domain from the Vint domain (ΔC52) without any mutations, the fusion protein showed a slow cleavage during IMAC purification. Surprisingly, the cleavage profile remained unchanged even with additional C1A mutation (Fig. 4 and Fig. S1). The C1A variant with the ARR deletion might suggest that the ARR deletion variant employs a cleavage mechanism that significantly differs from the one used by class 1 inteins (Fig. 4B and Fig. S1) [31]. Nevertheless, removing the ARR offered crucial evidence, indicating that ARR plays a role in regulating the auto-processing self-cleavage reaction carried out by the Vint domain, in addition to the common catalytic residues. The ARR seemingly protects the Vint protein from undesired cleavages around the HINT domain (Fig. 5).

Discussion

Inteins, BILs, and Hh-C represent members of the HINT superfamily that have been extensively studied [10].

Despite various structural and biochemical studies of various HINT domains, these efforts have predominantly centered around inteins. Conversely, the Vint domain belonging to the HINT superfamily has remained relatively understudied after its discovery through genomic sequence analysis [8,9]. Moreover, studies of HINT domains in eukaryotes have been notably restricted.

For the first time, we experimentally elucidated the three-dimensional structure of the Vint domain and shed light on the structural role of the C-terminal putative ARR (Fig. 5A). The three-dimensional structure led us to hypothesize that the ARR could have a binding affinity for sulfated GAGs, such as heparin sulfate because the distances between observed sulfate ions (6.4–7.1 Å) in the Vint domain structure are close to the spaces between sulfate atoms in the solution structure of heparin sulfate (Fig. 5B) [29]. The crystal structure supports the potential interaction of the observed sulfate-binding sites with sulfated glycosaminoglycans. Indeed, when we mutated the two arginine residues close to the sulfate ions, the binding affinity decreased using a heparin sepharose column. Notably, when the ARR was removed, this led to a decrease in N-cleavage efficiency by the truncated Vint domain. The removal of ARR confirmed its critical role in orchestrating the active site residues within the Vint domain, thereby promoting effective N-cleavage by the HINT domain (Fig. 5C). Deletion variants and mutations within the Vint domain may imply that adduct binding could facilitate cleavage or trigger potential modification by the HINT domain, akin to Hedgehog proteins (Fig. 5D) [30]. While the ARR features an α-helix and three β-sheets, the SRR from Hh-C is predicted to include a helix–loop–helix structure [12,24]. Therefore, it is worth noting that the functional roles of the SRR and ARR may also differ significantly.

Although the HINT superfamily shares a common structural fold, the three-dimensional structure of *Tth*Vint highlighted the distinct variations at the residues responsible for the catalytic reaction. The variations in the active residues within the HINT domain provide supportive evidence for the hypothesis that the HINT fold represents a case of structural convergence, offering a catalytic local environment for transesterification and cleavage reactions [31,32]. The C1A mutation in the Vint domain without ARR still displayed weak cleavages, suggesting that the Vint domain might alternatively undergo a reaction mechanism similar to the mechanism found in class 2 and 3 inteins, both of which lack the first cysteine or serine residue [33,34]. The Vint domain could thus serve as additional support for the structural convergence of the HINT superfamily throughout evolution.

Inteins among the HINT superfamily have become essential biotechnological tools for *in vivo* protein ligation by split inteins, C-terminal protein amidation, protein purification, and various conditional protein ligation and cleavages as we and others have demonstrated [35–40]. Similarly, the auto-processing Hh-C was demonstrated to be used for cholesterol-conjugated protein production and represents a tool for targeting functional peptides and proteins to cell membranes by producing various cholesterol-conjugated proteins [30,41]. The Vint domain of Hh-related proteins might provide a novel facile system to control cleavage or modification by adducts (Fig. 5). A further understanding of the biochemical roles of Vint proteins *in vivo* could offer a crucial tool for comprehending cell-to-cell communication and cell adhesions within the extracellular matrix, and assisting in the synthesis or modification of natural and artificial extracellular matrices and biomaterials. Vint proteins could also be further exploited for identifying novel molecules with significant therapeutic potential by attenuating their functions.

Intriguingly, inteins within the HINT superfamily exclusively occur in unicellular organisms, while hedgehog proteins have been found in most of the metazoans [10]. Vint proteins found in unicellular metazoa could play an important role in cell-to-cell communication through conjugation with adducts. They could also potentially play a pivotal role in the transition from non-metazoa to multicellular animals during the evolution of animals. Exploring Hh-related proteins bearing Vint domains by further studies could provide insights into their *in vivo* functions and contribute to our understanding of the evolutionary process of HINT domains from unicellular to multicellular organisms.

Acknowledgements

We thank B. Haas, J. Tommila, J. Lahtinen. L. Knaapi for technical help in protein and plasmid preparations. We thank T. Kajander and V. Manole for their technical assistance at the crystallization facility. We acknowledge the beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF) *via* the Finnish BAG. This work was supported in part by the Academy of Finland (131413, 137995, 277335), Novo Nordisk Foundation (NNF17OC0025402 to HMB, NNF17OC0027550 to HI), and Biocenter Finland, HiLIFE-INFRA for crystallization and NMR facilities at the Institute of Biotechnology, in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, and in part with Federal funds from the National Cancer Institute, National Institutes

of Health, Department of Health and Human Services, under Contract No. 75N91019D00024. This contract number represents work performed within the scope of work of the severable FFRDC Bridge contract. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Author contributions

HI conceived, designed, and directed the project. HI, HMB, and JEMJ acquired and interpreted the data. ML and AW analyzed the crystallographic data. HI wrote the paper with inputs from all authors.

Peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/1873-3468.14817.

Data accessibility

The structural data that support these findings are openly available in the wwPDB at https://doi.org/10.2210/pdb8R2C/pdb.

References

1 Noren CJ, Wang J and Perler FB (2000) Dissecting the chemistry of protein splicing and its applications. *Angew Chem Int Ed Engl* **39**, 450–466.
2 Anraku Y, Mizutani R and Satow Y (2005) Protein splicing: its discovery and structural insight into novel chemical mechanisms. *IUBMB Life* **57**, 563–574.
3 Paulus H (2000) Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* **69**, 447–496.
4 Novikova O, Jayachandran P, Kelley DS, Morton Z, Merwin S, Topilina NI and Belfort M (2016) Intein clustering suggests functional importance in different domains of life. *Mol Biol Evol* **33**, 783–799.
5 Beachy PA, Cooper MK, Young KE, von Kessler DP, Park W-J, Hall TMT, Leahy DJ and Porter JA (1997) Multiple roles of cholesterol in hedgehog protein biogenesis and signaling. *Cold Spring Harb Symp Quant Biol* **62**, 191–204.
6 Briscoe J and Thérond P (2013) The mechanisms of hedgehog signalling and its roles in development and disease. *Nat Rev Mol Cell Biol* **14**, 416–429.
7 Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA and Leahy DJ (1997) Crystal structure of a hedgehog autoprocessing domain: homology between hedgehog and self-splicing proteins. *Cell* **91**, 85–97.
8 Porter JA, Ekker SC, Park WJ, von Kessler DP, Young KE, Chen CH, Ma Y, Woods AS, Cotter RJ, Koonin EV *et al.* (1996) Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell* **86**, 21–34.
9 Bürglin TR (2008) Evolution of hedgehog and hedgehog-related genes, their origin from Hog proteins in ancestral eukaryotes and discovery of a novel Hint motif. *BMC Genomics* **9**, 127.
10 Bürglin TR (2008) The hedgehog protein family. *Genome Biol* **9**, 241.
11 Purohit R, Peng DS, Vielmas E and Ondrus AE (2020) Dual roles of the sterol recognition region in hedgehog protein modification [published correction appears in Commun Biol. 2020 Jun 2;3(1):286]. *Commun Biol* **3**, 250.
12 Mafi A, Purohit R, Vielmas E, Lauinger AR, Lam B, Cheng Y-S, Zhang T, Huang Y, Kim SK, Goddard WA *et al.* (2021) Hedgehog proteins create a dynamic cholesterol interface. *PLoS One* **16**, e0246814.
13 Horowitz S and Gorovsky MA (1985) An unusual genetic code in nuclear genes of *Tetrahymena*. *Proc Natl Acad Sci USA* **82**, 2452–2455.
14 Guerrero F, Ciragan A and Iwaï H (2015) Tandem SUMO fusion vectors for improving soluble protein expression and purification. *Protein Expr Purif* **116**, 42–49.
15 Ellilä S, Jurvansuu JM and Iwaï H (2011) Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in *Escherichia coli*. *FEBS Lett* **585**, 3471–3477.
16 Aranko AS, Oeemig JS and Iwaï H (2013) Structural basis for protein trans-splicing by a bacterial intein-like domain: protein ligation without nucleophilic side-chains. *FEBS J* **280**, 3256–3269.
17 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* **40**, 658–674.
18 Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK and Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D* **58**, 1948–1954.
19 Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D* **66**, 486–501.
20 Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F and Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D* **67**, 355–367.
21 Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC (2010) MolProbity: all-atom

structure validation for macromolecular crystallography. *Acta Crystallogr D* **66**, 12–21.
22 Holm L, Laiho A, Törönen P and Salgado M (2023) DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci* **32**, e4519.
23 Beyer HM, Mikula KM, Kudling TV and Iwaï H (2019) Crystal structures of CDC21-1 inteins from hyperthermophilic archaea reveal the selection mechanism for the highly conserved homing endonuclease insertion site. *Extremophiles* **23**, 669–679.
24 Banavali NK (2020) The mechanism of cholesterol modification of hedgehog ligand. *J Comput Chem* **41**, 520–527.
25 Whittaker CA and Hynes RO (2002) Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell* **13**, 3369–3387.
26 Aranko AS, Züger S, Buchinger E and Iwaï H (2009) In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS One* **4**, e5185.
27 Mills KV, Johnson MA and Perler FB (2014) Protein splicing: how inteins escape from precursor proteins. *J Biol Chem* **289**, 14498–14505.
28 Pietrokovski S (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci* **3**, 2340–2350.
29 Khan S, Gor J, Mulloy B and Perkins SJ (2010) Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: new insight into heparin-protein complexes. *J Mol Biol* **395**, 504–521.
30 Vincent S, Thomas A, Brasher B and Benson JD (2003) Targeting of proteins to membranes through hedgehog auto-processing. *Nat Biotechnol* **21**, 936–940.
31 Beyer HM, Virtanen SI, Aranko AS, Mikula KM, Lountos GT, Wlodawer A, Ollila OHS and Iwaï H (2020) The convergence of the Hedgehog/INTein fold in different protein splicing mechanisms. *Int J Mol Sci* **21**, 8367.
32 Salemme FR, Miller MD and Jordan SR (1977) Structural convergence during protein evolution. *Proc Natl Acad Sci USA* **74**, 2820–2824.
33 Southworth MW, Benner J and Perler FB (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J* **19**, 5019–5026.
34 Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, Ishino Y, Pietrokovski S and Perler FB (2010) Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. *J Biol Chem* **285**, 2515–2526.
35 Wood DW and Camarero JA (2014) Intein applications: from protein purification and labeling to metabolic control methods. *J Biol Chem* **289**, 14512–14519.
36 Züger S and Iwai H (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat Biotechnol* **23**, 736–740.
37 Volkmann G and Iwaï H (2010) Protein trans-splicing and its use in structural biology: opportunities and limitations. *Mol Biosyst* **6**, 2110–2121.
38 Aranko AS and Iwaï H (2021) The inducible intein-mediated self-cleaving tag (IIST) system: a novel purification and amidation system for peptides and proteins. *Molecules* **26**, 5948.
39 Ciragan A, Aranko AS, Tascon I and Iwaï H (2016) Salt-inducible protein splicing in cis and trans by inteins from extremely halophilic archaea as a novel protein-engineering tool. *J Mol Biol* **428**, 4573–4588.
40 Heikkinen HA, Aranko AS and Iwaï H (2022) The NMR structure of the engineered halophilic DnaE intein for segmental isotopic labeling using conditional protein splicing. *J Magn Reson* **338**, 107195.
41 Zhang X, Xu Z, Moumin DS, Ciulla DA, Owen TS, Mancusi RA, Giner JL, Wang C and Callahan BP (2019) Protein-nucleic acid conjugation with sterol linkers using hedgehog autoprocessing. *Bioconjug Chem* **30**, 2799–2804.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. SDS-PAGE analysis of expression and purification of the Vint domain variants.

Table S1. Plasmids and oligonucleotides used in this study.

Table S2. Crystallographic data collection and refinement statistics.