Protein Splicing and Auto-cleavage of Bacterial Intein-like Domains Lacking a C'-flanking Nucleophilic Residue*†

Bacterial intein-like (BIL) domains are newly identified homologs of intein protein-splicing domains. The two known types of BIL domains together with inteins and hedgehog (Hog) auto-processing domains form the Hog/intein (HINT) superfamily. BIL domains are distinct from inteins and Hogs in sequence, phylogenetic distribution, and host protein type, but little is known about their biochemical activity. Here we experimentally study the auto-processing activity of four BIL domains. An A-type BIL domain from Clostridium thermocellum showed both protein-splicing and auto-cleavage activities. The splicing is notable, because this domain has a native Ala C'-flanking residue rather than a nucleophilic residue, which is absolutely necessary for intein protein splicing. B-type BIL domains from Rhodobacter sphaeroides and Rhodobacter capsulatus cleaved their N’ or C’ ends. We propose an alternative splicing mechanism for the A-type BIL domains. After an initial N-S acyl shift, creating a thioester bond at the N’ end of the domain, the C’ end of the domain is cleaved by Asn cyclization. The resulting amino end of the C’-flank attacks the thioester bond next at the N’ end of the domain. This aminolysis step splices the two flanks of the domain. The B-type BIL domain cleavage activity is explained in the context of the canonical intein-protein-splicing mechanism. Our results suggest that the different HINT domains have related biochemical activities of proteolytic cleavages, ligation and splicing. Yet the predominant reactions diverged in each HINT type according to their specific biological roles. We suggest that the BIL domain cleavage and splicing reactions are mechanisms for post-translational generating protein variability, particularly in extracellular bacterial proteins.

Received for publication, April 26, 2004, and in revised form, May 17, 2004
Published, JBC Papers in Press, May 18, 2004, DOI 10.1074/jbc.M404562200

Bareket Dassa, Haim Haviv, Gil Amitai, and Shmuel Pietrokovski‡
From the Department of Molecular Genetics, the Weizmann Institute of Science, Rehovot, Israel 76100

EXPERIMENTAL PROCEDURES

Bacterial Strains and DNA Primers—Rhodobacter sphaeroides 2.4.1 (Rsp) genome was a kind gift from Dr. Steven L. Porter (University of Oxford). Rhodobacter capsulatus (Rca) MD1 genome was a kind gift from Dr. Fevzi Daldal (University of Pennsylvania), and Clostridium thermocellum (Cth) genome was a kind gift from Dr. Ying Tsai (University of Rochester). The following BIL domains were cloned: BIL4-Cth

---

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
† To whom correspondence should be addressed. Tel.: 972-8-9344108; E-mail: shmuel.pietrokovski@weizmann.ac.il.
‡ To whom correspondence should be addressed. Tel.: 972-8-9342747; Fax: 972-8-9344108; E-mail: shmuel.pietrokovski@weizmann.ac.il.

The abbreviations used are: BIL, bacterial intein-like; B, BIL, Rsp, Rhodobacter sphaeroides; Rca, Rhodobacter capsulatus; Cth, C. thermocellum; HINT, Hog/intein; MS, mass spectrometry; MALDI, matrix-assisted laser desorption/ionization; M, maltose-binding protein; C, Chitin-binding domain.
**Activity of Intein-like Domains Lacking a C'-nucleophile**

**Table I**

| Primer name | Primer sequence | Restriction site | Flank amino acids |
|-------------|-----------------|------------------|-------------------|
| BIL1-Rep    |                 |                  |                   |
| 5p-Nrep-b1l | GAATTGATCCGTAGGCCAACCTGAGCTGG | EcoRI | +14 |
| 3p-Nrep-b1l | TCTAAGGGGCAAGGGAGTCCTTCCGTT | XbaI | +52 |
| BIL2-Rep + flanks | GAAATGATCCGTAGGCCAACCTGAGCTGG | EcoRI | +32 |
| 5p-rsp-b1l2 | TCTAAGGGGCAAGGGAGTCCTTCCGTT | XbaI | +9  |
| 3p-rsp-b1l2 | TCTAAGGGGCAAGGGAGTCCTTCCGTT | XbaI | +1  |
| BIL2-Rep-no flanks | TCTAAGGGGCAAGGGAGTCCTTCCGTT | EcoRI | +1  |
| 5p-rsp2-bi-only +1 | TCTAAGGGGCAAGGGAGTCCTTCCGTT | XbaI | +1  |
| 3p-rsp2-only +1 | TCTAAGGGGCAAGGGAGTCCTTCCGTT | BamHI | +1  |
| BIL4-Cth   |                 |                  |                   |
| 5p.BILA Cth | AAAAGGATCCCTGCTTTGTCGCGACCAGTG | XbaI | +1  |
| 3p.408.BILA Cth | AAAATCTAGATGCTATCAGGACAC | BamHI | +36 |
| 1522-Rca   |                 |                  |                   |
| 5p.BIL     | GATCCAACTACGATCCGACGACC | XbaI | +1  |
| 1522-108bp | TCTAAGGACCATACGTCCTCAAGGCCTG | XbaI | +35 |
| 3p.BIL     |                 |                  |                   |
| 1522 +105bp|                 |                  |                   |

*Number of residues flanking the BIL domain.*

(NCBI gi code 23020817); BIL1-Rep (NCBI gi code 22959584); BIL2-Rep (NCBI gi code 22959191); and 1522-Rca (1). The BIL domains were amplified by PCR using the primers in Table I and cloned between two protein tags, the maltose-binding protein (M) upstream and the chitin-binding domain (C) downstream. The chimeric protein, M-B-C, containing the maltose-binding protein (M) from Escherichia coli and a downstream cbd gene coding for chitin-binding domain (C) from Bacillus circulans.

Functional Assay of Protein-splicing and Cleavage Activity—The coding sequence of different BIL domains (B) was cloned in-frame between two protein tags, the maltose-binding protein (M) upstream and the chitin-binding domain (C) downstream. The chimeric protein, M-B-C, was overexpressed and extracted in *E. coli* bacteria as described previously (1). Protein extraction buffer contained 20 mM Tris, pH 7.4, 200 mM NaCl, 1 mM EDTA, and 1 mM sodium azide.

Heat Purification of BIL4-Cth Domain—The supernatant of E. coli cell lysate overexpressing the BIL4-Cth construct was heated in extraction buffer to 37 °C for 10 min. Soluble proteins were separated on SDS-PAGE (Fig. 1). Western Blot Analysis was used to identify Western Blot Analysis—Sequence searches used the BLAST programs (10) and the BLIMPS program for block-to-sequence searches (11). Block multiple sequence alignments and phylogenetic analysis were conducted as described by Amita et al. (1). Protein motifs were detected using the InterProScan tool (www.ebi.ac.uk/interpro/scan.html).

RESULTS

To characterize the proteolytic activity of new A- and B-type BIL domains, each BIL domain (B) was cloned in-frame between two protein tags, maltose-binding protein (M) upstream and chitin-binding domain (C) downstream. Protein products of each chimeric gene (M-B-C) were examined in *vivo* and *in vitro* by various methods. To characterize the BIL domain activity in its native protein context, some of the domains were cloned with their full or partial native flanks, whereas others were cloned only with single residue flanks.

Protein Splicing and Cleavage of an A-type BIL Domain with Ala +1 Residue—BILA4-Cth is one of the 23 A-type BIL domains identified in the thermophilic bacterium *Cth* (1). It is typical of most A-type BIL domains to have all of the intein protein-splicing active site residues with the exception of the C'-flanking nucleophile (supplemental Fig. S3). Instead of Cys, Ser, or Thr invariably present in inteins, BILA4-Cth is followed by an Ala +1 residue. This is the residue present in 18% A-type BIL domains (fraction calculated as weighted average of putative active domains).

The BILA4-Cth M-B-C precursor was overexpressed in *vivo* as a double-tagged protein, and its products were detected and analyzed. Putative protein-splicing products, the excised BIL domain and the ligated M-C flanks, and the M- and M'-cleavage products were detected. These products were identified by Western blotting of total cell lysates and affinity-purified proteins separated on SDS-PAGE (Fig. 1A). Relative quantities of products were calculated according to measurements taken from Coomassie Blue-stained SDS gels of amyloose-purified proteins and total lysates (supplemental Fig. S1). Only trace amounts of the M-B-C precursor were detected under all of the separation procedures, indicating an efficient processing. Spliced product M-C comprised 20–25% of the final products, whereas C'-cleavage product M-B comprised ~5% of the final products. M and B proteins comprised most of the final products, indicating that they were generated by a combination of N'- and C'-cleavages. The final amount of B protein was much larger than the amount of the M-C-splicing product. This finding implies that both protein splicing and cleavage at its N' and C' ends released the B protein. The C product was not identified in the gels, perhaps because of cellular degradation.

To characterize the putative splicing product using MALDI-
Activity of Intein-like Domains Lacking a C'-nucleophile

To examine whether the Tris cell extraction and protein purification buffer promoted cleavage and splicing of the M-B-C precursor, the extraction and purification procedures were repeated using different buffers (Bio-Tris propane, HEPES, sodium phosphate, and borate). Same products and relative amounts were observed with all of these control buffers (data not shown).

In Vivo and In Vitro Cleavage Activities of B-type BIL Domains—B-type BIL domains are more heterogeneous in sequence than A-type domains (1). To characterize their activity, we cloned three different B-type BIL domains into the double-tagged system (described above): the two BIL domains present in \textit{R. sphaeroides} termed BIL1-Rsp and BIL2-Rsp and one of the 14 BIL domains present in \textit{R. capsulatus} termed 1522-Rea. The conserved C'-sequence motif of B-type BIL domains is distinct from those motifs in other known HINT domains (1). The C'-end of the cloned BIL1-Rsp and 1522-rea is typical of B-type BIL domains, whereas BIL2-Rsp has an atypical C'-end (supplemental Fig. S3).

\(N\)'-Cleavage of B-type BIL1-Rsp—BIL1-Rsp, a B-type BIL domain from \textit{R. sphaeroides}, was cloned between M and C tags with its native N'-14 residue and C'-51 residue flanks and overexpressed in \textit{E. coli} cells. M-B-C precursor M and B-C \(N\)'-cleavage products were identified by Coomassie Blue staining and Western blotting of total lysate and affinity-purified protein samples (Fig. 2). To verify the nature of the \(N\)'-cleavage product, B-C, the band was micro-sequenced. The resulting sequence (XPTPGT) corresponded to the predicted \(N\)'-end of the BIL domain, which also includes Cys-1, which usually cannot be detected by this method (supplemental Table S-I).

An additional 58-kDa band was co-purified with the M-B-C precursor. Its analysis suggests that the band might include more than a single protein species. Both anti-M and anti-C antibodies reacted with this band. However, the peptide mapping of the band identified peptides from both the M tag and the \textit{E. coli} GroEL chaperone protein. Additionally, no peptides from the B and C domains were identified (data not shown).

Intact mass of the band identified a mass of 58.317 kDa corresponding to GroEL and an additional unidentified protein mass of 65.175 kDa (Table IV). As a control, we checked a cross-reaction of anti-C antibodies with purified GroEL protein (supplemental Fig. S2B). Anti-C antibodies showed reactivity toward GroEL, probably because of their polyclonal nature.

GroEL chaperone was detected in protein samples purified the following affinity columns: on amylose; chitin; and amylose followed by chitin. This indicates a tight and specific binding of GroEL with the precursor and/or protein products. The association of GroEL with unfolded proteins is reversible to some extent upon incubation with ATP-Mg-K (12). Such incubation of washed protein samples bound on chitin reduced but did not eliminate the amount of GroEL eluted from chitin (supplemental Fig. S2B).

\(C\)'-Cleavage of B-type BIL2-Rsp in Vivo, In Vitro, and in Cell-free Systems—BIL2-Rsp was cloned between M and C tags with one native flanking residue at either end (N'-Leu and C'-Pro) and overexpressed in \textit{E. coli} and in a cell-free system. In both systems, the main product was the M-B-C precursor with small amounts of M-B- and M-cleavage products (Fig. 3A). An additional band of \(\sim 70\) kDa appeared above the precursor band when expressed in \textit{vivo}. This band was identified as the \textit{E. coli} DnaK chaperone protein. It was not detected in the overexpressed control protein, M-C. Identity of the above products was verified by Western blotting, N-terminal sequencing of the M-B-C band, MALDI-MS peptide mapping of the M-B and DnaK bands, and MALDI MS intact mass analysis of the M-B band (Fig. 3A, Table IV, and data not shown). This last
isolated M-B-C precursor was incubated in vitro at 4 or 37 °C in two different environments of pH 7.4 and 8.5. No products of the precursor were detected under any of these conditions.

**Species and Protein Host Distribution of BIL Domains**—BIL domains were identified originally in species from Gram-negative α, β, and γ Proteobacteria and from Gram-positive Actinobacteria and the Bacillus/Clostridium group (1). Further data base searches now broaden the taxonomic range of BIL domains to major bacterial divisions and lineages (supplemental Table S-II). A-type BIL domains were found in δ Proteobacteria, Cyanobacteria, Spirochaetes, Planctomycetes, and Verrucomicrobia. B-type BIL domains were found in α Proteobacteria, Rhizobium, and Silicibacter species.

Sequence analyses of over a hundred identified BIL flanks reconfirmed our previous observation of the nature of the BIL domain hosts. BIL domains are present in homologs of known and predicted secreted proteins. This is exemplified by *Streptomyces avermitilis*, *Verrucomicrobiu*, and *Gloeobacter* A-type BIL domains that are found downstream of long (400–5400 residues) rhs core elements. Rhs elements are composite genetic elements, and their cores are believed to be cell-surface ligand-binding proteins (14). The BIL domains are present in the hyper-variable core extension region that can be shuffled between the core and downstream open-reading frame regions.

**DISCUSSION**

In this study, we show that a typical A-type BIL domain is capable of protein splicing without a C'-nucleophilic +1 residue and that B-type BIL domains can cleave their N' or C' ends. Both types of domains are not uncommon, appearing in diverse bacterial divisions. These findings reflect the auto-processing nature of intein-like domains. We explain the N'- and C'-cleavage of B-type BIL domains by reactions occurring in the canonical intein protein-splicing mechanism and propose an alternative pathway for A-type BIL domains splicing. Our results suggest that the biochemical activities of the BIL domains are distinct from inteins, and their native biological function is probably protein modification by splicing and cleavage activity.

**Protein-splicing Mechanism without a Nucleophile +1 Residue**—Intein protein-splicing mechanism was largely determined by mutational analysis of a few representative intein domains (2, 6, 15–18). This allowed the delineation of the biochemical reactions of protein splicing and supported splicing as the native activity of inteins. Other evidence for the nature of interin activity are the high efficiency of intein protein-splicing, intein distribution in species and host proteins, and the function of intein genes as selfish genetic elements (19).

Currently, the accepted mechanisms for intein protein-splicing require a Cys, Ser, or Thr +1 residue at the intein immediate C'-flank. This nucleophilic +1 residue is crucial for the trans-esterification step and for the final acyl rearrangement (Fig. 4, steps 2A and 4A). In inteins with N'-Ala-1, the nucleophilic +1 residue directly attacks the peptide bond at the intein N' end (16). Mutating the intein active site residues, including the +1 nucleophilic residue, abolishes splicing or leads to cleavage of the intein C', N' end, or both (15, 20).

In our study, the major products of BIL4-Cth expression were N'- and C'-cleavages, whereas protein splicing was ap-
proximately a quarter to a fifth of the A-type BIL domain activity with almost complete processing of the precursor. Most probably, the initial cleavage activity was at the C-end, producing the M-B and C products, followed by additional N'-cleavage of the M-B product, producing the M and B products. This is supported by the relative amounts of the final products and the absence of the B-C product.

Our results show protein splicing of an A-type BIL with conserved sequence features closely related to inteins including all of the active site residues apart from the $\text{C}^-\text{H}$ residue. Hence,

FIG. 2. $\text{N}'$-cleavage activity of BIL1-Rsp. Protein products from E. coli overexpression of M-B-C construct with B-type BIL1-Rsp were eluted from amylose (A), chitin (C), or both (A+C) affinity columns or analyzed in total cell lysate (T). Proteins were separated on SDS-PAGE and either stained with Coomassie Blue or detected by anti-M, anti-C, or anti-GroEL (Anti-G) antibodies. See “Results” for discussion of GroEL cross-detection by anti-C antibodies.

FIG. 3. $\text{C}'$-cleavage activity of BIL2-Rsp. A, left, protein products from E. coli overexpression of M-B-C construct with B-type BIL2-Rsp were eluted from either amylose (A) or chitin (C) affinity columns or analyzed in total cell lysate (T). Proteins were separated on SDS-PAGE and either stained with Coomassie Blue or detected by anti-M, anti-C, or anti-DnaK antibodies. A, right, proteins translated in vitro in a cell-free system were labeled with $\text{[35S]}\text{Met}$.

B, in vitro incubation of a purified precursor at 4 °C.

FIG. 4. Canonical and proposed protein-splicing mechanisms for inteins and A-type BIL domains. The intein/BIL domain is marked as a black rectangle flanked by an N-terminal flank (N) and a C-terminal flank (C). Right, canonical intein protein-splicing mechanism. Left, proposed protein-splicing mechanism of A-type BIL domains lacking a C'-nucleophilic residue.
we propose a modified protein-splicing mechanism for A-type BIL domains. The mechanism is similar to the canonical protein-splicing mechanism of inteins, only differing in the nature of the nucleophilic attack on the thioster bond in the N’ end at the BIL domain.

Our suggestion includes the following steps of protein splicing in A-type BIL domains (Fig. 4). (i) A thioster is formed at the N’ end of the domain by the N-S acyl shift (Fig. 4, step 1) by attack of the thiol group of the conserved Cys-1 residue on the carbonyl group of the peptide bond N-terminal to Cys-1. This reaction is the same as the first step of canonical intein protein splicing (15, 18, 20). (ii) Concomitantly, the conserved Asn residue at the C’ of the domain undergoes cyclization into an aminosuccinimide ring, cleaving the peptide bond at the domain C’ end (Fig. 4, step 2B). This step generates two intermediate products: the N’-flank covalently connected to the BIL domain by a thioster bond and the detached C’-flank. This reaction also occurs in intein protein splicing but only after ligation of the two intein flanks (Fig. 4, step 3A) (5, 21). In inteins, premature Asn cyclization results in C’-cleavage and no splicing (22). Although the timing of Asn cyclization is tightly controlled in inteins, it can still occur when other steps of the splicing are blocked by mutations at the N’- and/or C’-splice junction (17, 23–25). (iii) The free N terminus of the C’-flank performs an aminolysis reaction of the labile thioster bond next at the N’ junction of the domain formed in step i. This reaction ligates the two BIL domain flanks with a peptide bond and releases the BIL domain from its N’-flank. This step probably occurs immediately after step ii to prevent the dissociation of the C’-flank from the N’-flank and BIL domain. (iv) Finally, the BIL domain C’-aminosuccinimide ring hydrolyzes into Asn or iso-Asn, similarly to inteins (Fig. 4, step 4) (26).

Aminolysis reaction, involving an attack of the C’-amine on a N’-ester, was proposed previously to occur in intein protein splicing (27, 28). A detailed analysis of representative inteins established the canonical protein-splicing mechanism and ruled out aminolysis as part of the process (15, 20). Considering our experimental results and the various residues in the +1 position of A-type BIL domains, we suggest that these domains protein splice with an aminolysis reaction.

Recently, aminolysis was proposed as part of a peptide-splicing activity of the proteasome that generates the displayed variant antigenic peptides (8). The cleaved peptides within the proteasome are attached transiently from the C’ end to ester bonds (21). Vigneron et al. (8) suggest that the N’ end of another cleaved peptide from the same protein attacks this bond in an aminolysis reaction, ligating the two peptides. Aminolysis also occurs in other biological reactions, including the attachment of myristate to the N’ end of proteins by N-myristoyltransferase (29).

Why are inteins integrated upstream to Cys, Ser, or Thr residues when, as we show here, protein-splicing can proceed with other residues in this position? Being able to successfully integrate in a wider range of sites seems highly advantageous for selfish genetic elements such as inteins (19, 30). We believe the answer to this question is related to the differences between the mechanisms for protein splicing in inteins and in A-type BIL domains. The intein domain and its flanks remain covalently attached until ligation of the flanks and release of the intein (Fig. 4). In our proposed mechanism for A-type BIL domains, the C’-flank is detached from the BIL and its N’-flank before its ligation to the N’-flank. This may lead to a higher frequency of N’- and C’-cleavage side products. Such partial splicing in inteins will reduce the amount of mature (spliced) host proteins, which are typically conserved, and crucial proteins, and might negatively affect cell survival. Perhaps even more harmful is the possible dominant-negative effect of the cleaved byproducts of intein hosts. In contrast, partial splicing of BIL domains (i.e. N’- and/or C’-cleavage) may serve for increasing the protein host variability (1).

Our results, together with previous reports of other atypical intein protein-splicing mechanism (9), show that this activity can proceed by several alternative and partially overlapping biochemical reactions. Thus, the canonical intein protein-splicing mechanism may need to be expanded, or its scope may need to be limited. Aminolysis and perhaps other atypical mechanisms may be the way some inteins and other HINT domains protein-splice.

Cleavage Mechanisms of B-type BIL Domains—The B-type BIL domains were found by us to auto-catalytically cleave their N’ or C’ ends. This activity is analogous to inteins protein-splicing side reactions and is common in N-terminal rearrangements of auto-processing proteins (2). Both intein and BIL domains have conserved Cys or Ser in position 1 whose thiol or hydroxyl groups are essential for the acyl rearrangement at the N’ terminus. Thus, the N’-peptide-bond of BIL1-Rsp could be converted into a thioster through the N-S acyl shift, similarly to inteins (Fig. 4, step 1). In inteins, this reaction is followed by trans-esterification of the thioster by the side chain of the +1 residue, forming a branch intermediate and leading to splicing product formation. Such products were not obtained in the BIL1-Rsp precursor expression, suggesting that the labile thioster was hydrolyzed by water or by an external nucleophile. We do not exclude the possibility that this cleavage was coupled to ligation of the upstream flank with an external nucleophile, similar to the attachment of cholesterol to Hedge domain upstream to the Hog HINT domain. Such a ligation would modify the M tag and assign it with a higher mass. One of the BIL1-Rsp yet uncharacterized products may correspond to this putative product.

A previously proposed mechanism for C’-cleavage of the Chy R1 intein mutant (9) and for PolII intein (31) can explain the C’-auto-cleavage of BIL2-Rsp. According to this finding, an attack of the BIL domain Ser-I hydroxyl group on a peptide bond carbonyl at the C’ region of the domain would form an ester bond through the N-O acyl shift, which in turn can be hydrolyzed, detaching the BIL domain from its C’-flank (9). This proposed mechanism is independent of a C’-nucleophilic residue. Assuming that BIL domains have the HINT fold, their N’ end is in a position to cleave their C’ region.

Our heterologous conditions of protein expression may alter the native activity of BIL domains. Overexpression in E. coli cells and changes in the domain context (BIL domain flanks), as well as in vitro conditions such as redox environment or temperature, may alter the protein in vivo fold and function. Nevertheless, in light of extensive experiments in other proteins and HINT domains, we assume that the BIL domain activity we observed is related to their native one. Improper folding of flanked B-type BIL domains may have triggered the overexpression of chaperones (DnaK, GroEL) (12). We propose that the chaperones, which were co-purified with B-type BIL but were absent in A-type BIL or the control vector, are not merely byproducts of the heterologous expression system. Chaperones may be involved in BIL domains proper folding, extracellular targeting, or biological activity. Attachment of chaperones to the BIL precursor may also spatially block its splicing activity.

Biological Roles of Different Types of HINT Domains—The HINT superfamily currently includes four separate families: inteins; Hogs; A-type BIL; and B-type BIL domains. All of the families are homologous and share sequence, structure, and biochemical properties (2, 4, 6, 32). Yet each family is distinct
in specific sequence features, protein host context, and biological roles. Members of each family can be diverse in sequence and are still found occasionally in new protein and phylogenetic contexts. It is likely that other HINT families will be discovered and characterized. Thus, identifying the family of a HINT domain can be an additional challenge to recognizing the domain as a HINT type.

Sequence motifs and structure folds characterizing the HINT superfamily and those specific to inteins, Hogs, and BIL domains have been described previously (1, 33, 34). Most inteins also include a central homing-endonuclease domain (35) not found in the other known HINT families. Inteins are also integrated in conserved positions of essential proteins. Both these features are a consequence of the selfish element nature of intein genes (19, 30). Hog domains are located upstream to the cholesterol-binding domain and downstream to the Hedgehog domains and to the Wart and Ground domains of nematodes (36). The role of Hog domains in hedgehog proteins and perhaps also in the nematode proteins is post-translational modification in the maturation process of their host protein.

Less information is available for the two known BIL domains. Nevertheless, the experimental and computational results we show in this work support our initial hypotheses. Most BIL domains are present in variable positions of non-conserved proteins. Many BIL host proteins also include motifs, repeats, and domains that characterize extracellular protein regions. We show here and in the first report of the BIL domains (1) that the biochemical activity of BIL domains includes protein splicing and auto-cleavage of their hosts. We suggest that the biological role of BIL domains is to increase the variability of their hosts, mainly in extracellular protein regions, by cis- and trans-splicing of proteins and other moieties to the hosts.

Acknowledgments—We thank Prof. Meir Wilcheck for supportive suggestions and Prof. Amnon Horovitz for samples of GroEL and its antibodies. We thank the Mass Spectrometry unit of the Weizmann Institute of Science (Rehovot, Israel) and The Smoler Protein Center, Department of Biology (Technion, Israel).

REFERENCES

1. Amitai, G., Belenkiiy, O., Dassa, B., Shainskaya, A., and Pietrokovski, S. (2003) Mol. Microbiol. 47, 61–73
2. Paulus, H. (2000) Annu. Rev. Biochem. 69, 447–496
3. Porter, J. A., Ekker, S. C., Park, W. J., von Kessler, D. P., Young, K. E., Chen, C. H., Ma, Y., Woods, A. S., Cotter, R. J., Koonin, E. V., and Beachy, P. A. (1996) Cell 86, 21–34
4. Perler, F. B. (1998) Cell 92, 1–4
5. Xu, M. Q., Comb, D. G., Paulus, H., Noren, C. J., Shao, Y., and Perler, F. B. (1994) EMBO J. 13, 5517–5522
6. Perler, F., Noren, C., and Wang, J. (2000) Angew. Chem. Int. Ed. Engl. 39, 450–466
7. Southworth, M. W., Yin, J., and Perler, F. B. (2004) Biochem. Soc Trans. 32, 250–254
8. Vigneron, N., Stroobant, V., Chapiron, J., Ooms, A., Degiovanni, G., Morel, S., Van Der Bruggen, P., Boon, T., and Van Den Eynde, B. J. (2004) Science 304, 587–590
9. Amitai, G., Dassa, B., and Pietrokovski, S. (2004) J. Biol. Chem. 279, 3121–3131
10. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
11. Henikoff, S., Henikoff, J. G., Alford, W. J., and Pietrokovski, S. (1995) Gene (Amst.) 163, 17–26
12. Bukau, B., and Horwich, A. (1998) Cell 92, 351–366
13. Coote, J. (1992) FEBS Lett. 28, 137–161
14. Yong Dong, W. (1998) J. Bacteriol. 180, 4102–4110
15. Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B., and Xu, M.-Q. (1996) J. Biol. Chem. 271, 22159–22168
16. Southworth, M. W., Benner, J., and Perler, F. B. (2000) EMBO J. 19, 5019–5026
17. Xu, M.-Q., and Perler, F. B. (1996) EMBO J. 15, 5146–5153
18. Romazelli, A., Shekhtman, A., Cowburn, D., and Muir, T. W. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 6397–6402
19. Pietrokovski, S. (2001) Trends Genet. 17, 465–472
20. Shao, Y., Xu, M. Q., and Paulus, H. (1996) Biochemistry 35, 3810–3815
21. Grell, M., and Huber, R. (2005) Int. J. Biochem. Cell Biol. 35, 606–616
22. Wood, D. W., Wu, W., Belfort, G., Derbyshire, V., and Belfort, M. (1999) Nat. Biotechnol. 17, 889–892
23. Cooper, A. A., Chen, Y. J., Lindorfer, M. A., and Stevens, T. H. (1993) EMBO J. 12, 2575–2583
24. Southworth, M. W., Amaya, K., Evans, T. C., Xu, M. Q., and Perler, F. B. (1999) BioTechniques 27, 110–120
25. Chong, S., Mintello, G. E., Zhang, A., Cantor, E. J., Liao, W., Xu, M. Q., and Benner, J. (1998) Nucleic Acids Res. 26, 5109–5115
26. Shao, Y., Xu, M. Q., and Paulus, H. (1995) Biochemistry 34, 10844–10850
27. Clarke, N. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 11084–11088
28. Perler, F. B., and Adam, E. (2000) Curr. Opin. Biotechnol. 11, 377–383
29. Farazi, T. A., Waksman, G., and Gordon, J. I. (2001) Biochemistry 40, 6335–6343
30. Liu, X. Q. (2000) Annu. Rev. Genet. 34, 61–76
31. Mills, K. V., Manning, J. S., Garcia, A. M., and Wuerdeman, L. A. (2004) J. Biol. Chem. 279, 20665–20691
32. Pietrokovski, S. (1998) Protein Sci. 7, 64–71
33. Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A., and Leahy, D. J. (1997) Cell 91, 85–97
34. Dalgaard, J. Z., Moser, M. J., Hughey, R., and Mian, I. S. (1997) J. Comput. Biol. 4, 153–214
35. Belfort, M., and Roberts, R. (1997) Nucleic Acids Res. 25, 3379–3388
36. Burglin, T. R. (1996) Curr. Biol. 6, 1047–1050
37. Mehlman, T., Benjamin, M., Merhav, D., Osman, F., Ben-Azuli, Y., Goldshleger, R., Karlshis, L., and Shainskaya, A. (2002) Proceedings of the 50th Conference of American Society for Mass Spectrometry, Orlando, June 2–6, 2002, American Society for Mass Spectrometry, Santa Fe, NM
