nature structural & molecular biology

Crystal structure of an intramolecular chaperone mediating triple–β-helix folding

Eike C Schulz[1], Achim Dickmanns[1], Henning Urlaub[2], Andreas Schmitt[1], Martina Mühlenhoff[3], Katharina Stummeyer[3], David Schwarzer[3], Rita Gerardy-Schahn[3] & Ralf Ficner[1]

Protein folding is often mediated by molecular chaperones. Recently, a novel class of intramolecular chaperones has been identified in tailspike proteins of evolutionarily distant viruses, which require a C-terminal chaperone for correct folding. The highly homologous chaperone domains are interchangeable between pre-proteins and release themselves after protein folding. Here we report the crystal structures of two intramolecular chaperone domains in either the released or the pre-cleaved form, revealing the role of the chaperone domain in the formation of a triple–β-helix fold. Tentacle-like protrusions enclose the polypeptide chains of the pre-protein during the folding process. After the assembly, a sensory mechanism for correctly folded β-helices triggers a serine-lysine catalytic dyad to autoproteolytically release the mature protein. Sequence analysis shows a conservation of the intramolecular chaperones in functionally unrelated proteins sharing β-helices as a common structural motif.

Proteins fold into their functional three-dimensional structure based on the information encoded in the residue sequence[1]. In all domains of life, various strategies evolved to assist folding processes in order to prevent immediate misfolding and aggregation of the nascent polypeptide chain. In the crowded cellular environment, protein folding is often aided by molecular chaperones. Molecular chaperones differ in size, function and energy dependence; however, all have in common to bind to the unfolded state of the protein in order to facilitate or mediate assembly of the correct three-dimensional structure[2,3]. In contrast to molecular chaperones, the intramolecular chaperones (IMCs) constitute a different class of chaperones. As part of the polypeptide chain, the IMC is typically cleaved off the target protein after the folding process is completed. Two classes of IMCs can be distinguished: class I IMCs assist the protein to fold into the correct tertiary structure, whereas class II IMCs are involved in quaternary structure assembly[4]. In contrast to many molecular chaperones, no evidence for an ATP-driven cleavage reaction could be found in IMCs. An example of a class II intramolecular chaperone has been identified in viral tailspike and fiber proteins. These proteins are functionally unrelated but share a highly conserved chaperone domain at their C terminus (C-terminal intramolecular chaperone domain, CIMCD), which is cleaved at a conserved position in an autoproteolytic reaction[5]. It was shown that the covalent linkage between the CIMCD and N-terminal pre-protein is necessary for correct folding, indicating that an *in trans* function of the chaperone is impossible[5]. Furthermore, it could be shown that some of the chaperone domains are exchangeable between the different pre-proteins[5,6]. While the three-dimensional structures of the CIMCDs are as yet unknown, the crystal structure of N-terminally truncated mature endoNF$_{246–910}$ shows that the homotrimeric enzyme comprises a triple β-helix involved in substrate recognition[7]. This triple β-helix and the related triple–β-spiral motif have been identified in various proteins, which often play a role as virulence factors[8]. In triple β-helices, three polypeptide chains wind around a common threefold symmetry axis, conferring an extraordinary stability to the protein. The rigid elongated shape allows triple β-helix–comprising proteins to protrude from a pathogen's surface in order to interact with flexible host cell receptors, like lipopolysaccharides[9]. However, proper assembly of triple–β-helical folds poses to be difficult in the absence of a trimerization domain[10]. Hence, most triple β-helices depend on a C-terminal extension for trimerization and correct assembly[11].

Here we present the crystal structures of two representatives of a large group of systematically, functionally and structurally similar intramolecular chaperones: the *Escherichia coli* phage K1F endosilidase CIMCD and the *Bacillus subtilis* phage GA-1 neck appendage protein CIMCD. Furthermore, we show that these chaperone domains are structurally conserved among a wide group of evolutionarily unrelated viruses and that they exclusively fold triple β-helices. Our results provide the structural basis for self-cleavage and its control for correct protein folding as well as providing mechanistic insights into the folding strategy, which might serve as a model for many clamp-like binding chaperones.

RESULTS

Structure of GP12-CIMCD

The crystal structure of the C-terminal domain of the neck appendage protein GP12 (GP12-CIMCD) was solved by means of single isomorphous

[1]Abteilung für Molekulare Strukturbiologie, Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen, Göttingen, Germany. [2]Bioanalytische Massenspektrometrie, Max Planck Institut für biophysikalische Chemie, Göttingen, Germany. [3]Institut für Zelluläre Chemie, Medizinische Hochschule Hannover, Hannover, Germany. Correspondence should be addressed to R.F. (rficner@uni-goettingen.de).

Received 7 May 2009; accepted 20 November 2009; published online 31 January 2010; doi:10.1038/nsmb.1746

Table 1 Data collection, phasing and refinement statistics

	GP12-CIMCD	GP12-CIMCD SeMet derivative	endoNF-CIMCD
Data collection			
Space group	P3	P6	P1
Cell dimensions			
a, *b*, *c* (Å)	84.04, 84.04, 41.45	84.12, 84.12, 40.8	62.65, 79.950, 109.58
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	81.71, 76.53, 87.11
Resolution (Å)	20–2.2 (2.24–2.20)	20–2.85 (2.88–2.85)	20–2.6 (2.66–2.60)
Wavelength (Å)	0.933	0.9783	0.9202
R_{sym}	0.036 (0.464)	0.0891 (0.479)	0.041 (0.38)
$I / \sigma(I)$	27.18 (3.34)	24.69 (4.73)	15.8 (2.18)
Completeness (%)	99.7 (100)	99.5 (99.7)	95.4 (95.3)
Redundancy	6.01	11.1	1.95
Refinement			
Resolution (Å)	42.2–2.2		27.6–2.6
No. reflections	16,515		59,958
R_{work} / R_{free}	0.1970 / 0.2322		0.2089 / 0.2545
No. atoms	2,074		12,670
Protein	1,979		12,123
Ligand / ion	10		41
Water	85		326
B-factors (Å^2)	48.85		55.04
Protein	48.99		54.78
Ligand/ion	51.24		74.49
Water	45.22		51.93
R.m.s. deviations			
Bond lengths (Å)	0.009		0.003
Bond angles (°)	1.386		0.727

Values in parentheses indicate the specific values in the highest resolution shell.

replacement including anomalous scattering (SIRAS) using a selenomethionine (SeMet) derivatized protein crystal, and the structure was refined at a resolution of 2.2 Å (**Table 1**). The asymmetric unit contains two monomers, which belong to different molecules. A crystal symmetry operation generates the other two monomers of the functional homotrimeric protein.

The overall structure of GP12-CIMCD has a jellyfish-like outline with a central threefold symmetry axis and mainly consists of α-helices. It comprises a quite globular core with a diameter of approximately 50 Å and an extended loop region, reminiscent of tentacles, protruding some 45 Å from the center. The central part of the core is a slightly twisted three-helix bundle with a length of 25 Å, forming the trimerization interface. From the tip of the tentacles to the C-terminal end of the three-helix bundle, GP12-CIMCD has an extension of ~70 Å. α-helices and a short pair of antiparallel β-strands surround the three-helix bundle in an almost perpendicular orientation and are wrapped around the base of the tentacles (**Fig. 1**). Each monomer displays an α1-β1-α2-β2-α3-β3-β4-α4 fold (**Supplementary Fig. 1**). The N-terminal residues, preceding α1, are interacting with α2 of the neighboring monomer. The elements α1-β1-α2-β2-α3 form a flap wrapping around the central core formed by the α4 helices of the three subunits. The tentacle includes residues 688–720 comprising two antiparallel β-strands (β3 and β4) connected by an extended loop, which entirely contains polar side chains. For this polar area of the CIMCD, markedly increased *B*-factors can be observed, indicating an increased flexibility. Strand β4 of the tentacle directly passes into the central helix (α4). This central helix forms tight interactions with the helices from the other two polypeptide chains by multiple hydrophobic interactions.

However, the structure of GP12-CIMCD does not provide insight into its role in the folding process of the tailspike protein. Furthermore, it does not allow deduction of the autoproteolytic cleavage mechanism of the CIMCD, since it is not longer part of the intact pre-protein. An MS analysis of the GP12-CIMCD shows that the cleavage occurs N-terminally to a conserved serine residue (Ser620 in GP12) (**Supplementary Fig. 2**). In concert with these results, the serine residue is still present in the structure of GP12-CIMCD. In order to gain insight into the details of the cleavage reaction and the structural basis for CIMCD-mediated folding, the interaction between the CIMCD and the pre-protein needed to be examined. Because a noncleavable mutant of GP12 ($GP12_{S620A}$) failed to crystallize, we designed a truncated version of a noncleavable mutant of endoNF ($endoNF_{S911A}$), containing the triple β-helix but lacking the N-terminal catalytic sialidase domain, based on the three-dimensional structure of endoNF[5,7].

Structure of pre-cleavage endoNF-CIMCD

The N-terminal deletion of the first 784 residues of $endoNF_{S911A}$ yielded a protein comprising the triple–β-helix domain and the CIMCD of endoNF incapable of autoproteolytic cleavage. The crystal structure of this pre-cleavage endoNF-CIMCD was solved by molecular replacement using the triple–β-helix domain of $endoNF_{785–910}$ and the core of GP12-CIMCD lacking the tentacles as search models, and the structure was refined at a resolution of 2.6 Å (**Table 1**). The asymmetric unit contains six monomers forming two trimers.

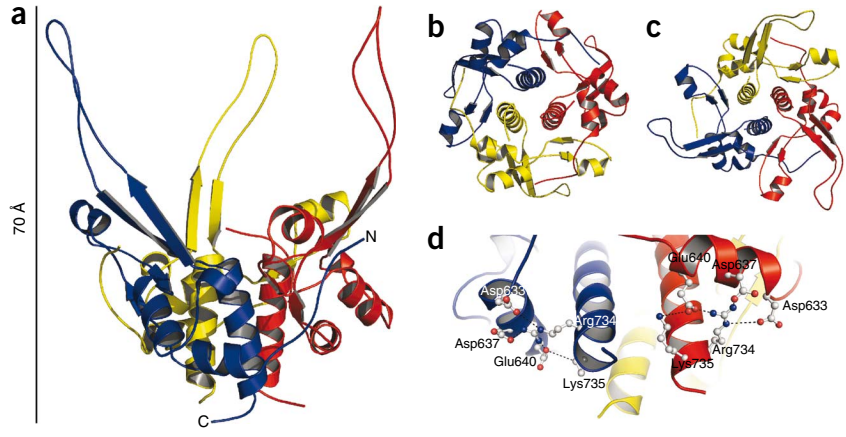

Figure 1 Structure of GP12-CIMCD. The structure of GP12-CIMCD exhibits a compact globular structure with three loops emanating approximately 45 Å from the core. The homotrimer is predominantly composed of α-helical subunits (colored in red, blue and yellow) with the loop region fixed at their base by a pair of antiparallel β-strands. (**a**) Overall side view. (**b**) Bottom view. The overall circular shape shows the twisted three-helix bundle in the center. (**c**) Top view. (**d**) Arg734 located near the C terminus in central helix is closely interacting with Asp633, Asp637 and Glu640, stabilizing the monomer.

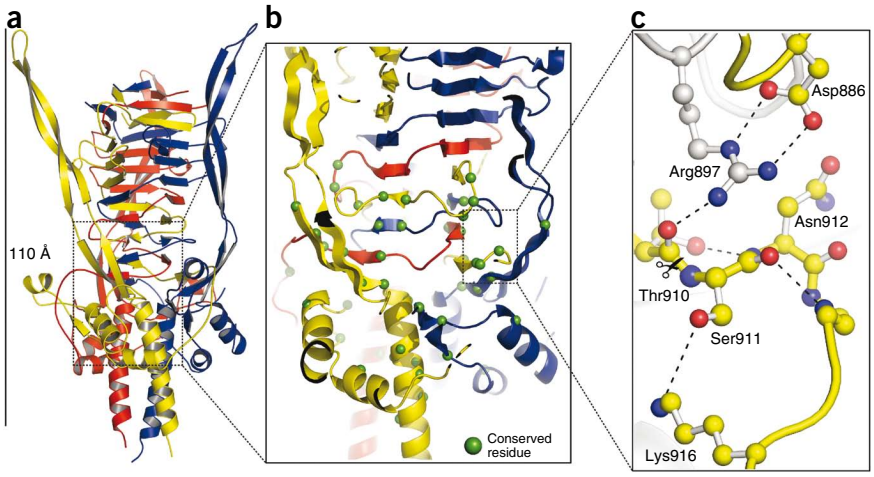

Figure 2 Structure of pre-cleavage endoNF-CIMCD. (**a**) Overall side view. The core region of the homotrimer is formed by α-helices, whereas the extended loops are composed of antiparallel β-strands, aligning in a shallow groove on the surface of the triple β-helix. (**b**) Conserved residues found in pre-cleavage endoNF-CIMCD, GP12-CIMCD and other proteins are indicated as green spheres. Notably, the conserved residues can be found encompassing secondary-structure elements. Furthermore, the conserved residues are not only present in the CIMCD but also N-terminally of the cleavage site in the triple β-helix. (**c**) Close-up of the cleavage site. Ser911 has been modeled into the structure to illustrate the proximity of this residue to the cleavage position between Thr910 and Ser911. Scissors indicate the scissile peptide bond. Arg897, conserved in the pre-protein, forms the oxyanion pocket directly above the carbonyl oxygen of Thr910, stabilizing the transition state to enable cleavage. Arg897 can only obtain its position after folding is completed; hence, it acts as a sensor for correct folding and as a trigger for cleavage.

The overall organization of the pre-cleavage endoNF-CIMCD (**Fig. 2**) shows two clearly distinct domains, the fully folded triple–β-helix stalk and the chaperone domain. The stalk domain essentially adopts the fold of an intertwined triple β-helix, which was previously described[7], whereas the chaperone domain adopts the same architecture found for GP12-CIMCD (see above). A central twisted three-helix bundle (α4) of 35 Å length, surrounded by α-helices and two short antiparallel β-strands (α1-β1-α2-β2-α3-α3a) forms the core with an approximate diameter of 50 Å. Three tentacles protrude some 70 Å (β3-β3a-β3b-β4) out of the center, giving rise to an overall extension of about 110 Å from the tip of the tentacles to the C terminus. The tentacles in pre-cleavage endoNF-CIMCD are almost twice as long as the tentacles of GP12-CIMCD and contain two twisted antiparallel β-strands (β3a-β3b) (**Fig. 2a**). The tentacles include residues 984–1035 and bind to the β-helix domain at a shallow groove on the surface (**Fig. 3**). Notably, a sialic acid binding site was also previously identified in this area, indicating that sialic acid binding is only possible after release of the CIMCD[7]. Similar to GP12-CIMCD, the tentacles of endoNF-CIMCD comprise a high proportion of polar residues, facing the outside while the non-polar residues predominantly point to the triple β-helix domain. Only five residues of the tentacles (Thr993, Asp994, Tyr1020, His1026 and Arg1035) form hydrogen bonds to the triple β-helix. Additional to the interaction with the β-helix, Arg1035 forms a hydrogen bond to the backbone atoms in β3.

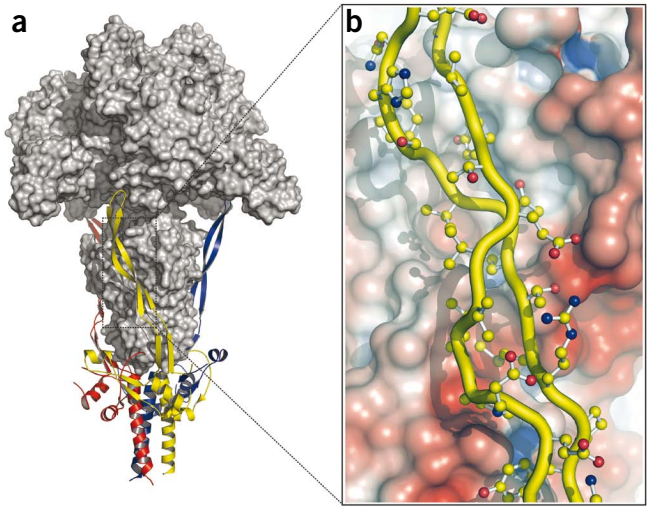

Figure 3 EndoNF before cleavage. (**a**) The figure shows a superposition of the crystal structure of endoNF and pre-cleavage endoNF-CIMCD. The CIMCD is displayed as a cartoon. EndoNF is displayed as a gray surface representation. (**b**) A close-up view of the interaction of the tentacles with the surface of the triple–β-helix domain of endoNF-CIMCD. The tentacles align along the stalk domain of pre-cleavage endoNF-CIMCD in a shallow groove on the surface. The tentacles are displayed as yellow cartoons, residues are shown in ball-and-stick mode and the surface is colored according to the electrostatic surface potential (APBS)[28].

The cleavage site

A major difference to GP12-CIMCD can be observed at the N-terminus of the chaperone domain of endoNF. As cleavage is inhibited by the alanine substitution of Ser911, it is possible to identify residues apparently involved in the autocatalytic cleavage reaction. The protein backbone at this position displays an S-shaped outline stabilized by the strictly conserved residue Thr910 forming a hydrogen bond to the backbone amide of Asn912. The S-shaped backbone brings the side chain of Ala911 into close proximity to the strictly conserved residue Lys916. Located next to the cleavage site between Thr910 and Ala911 lies Arg897, which forms a salt bridge to Asp886 and a hydrogen bond to the backbone carbonyl of Thr910 (**Fig. 2c**).

Biochemical characterization

In previous studies, various conserved residues were exchanged to study their influence on cleavage or folding of endoNF, showing that Leu903 and Pro907 have no effect on the cleavage of the chaperone domain[5]. The crystal structure reveals that these two nonpolar side chains participate in stabilization of the hydrophobic core of the stalk domain, a common pattern in triple β-helices. Mutation of the residues flanking strand β2 (His954, Gly956) led to insoluble protein, without detectable cleavage of the CIMCD[5,6]. Both residues are situated in a pocket between the triple β-helix and the flap surrounding it, forming hydrogen bonds to strand β1. Because no obvious interaction between these residues and the triple β-helix is detectable, it is conceivable that they do not stabilize the core of the CIMCD but instead play an important role during the folding process. Substitution of Arg1035 by alanine leads to only partially soluble protein incapable

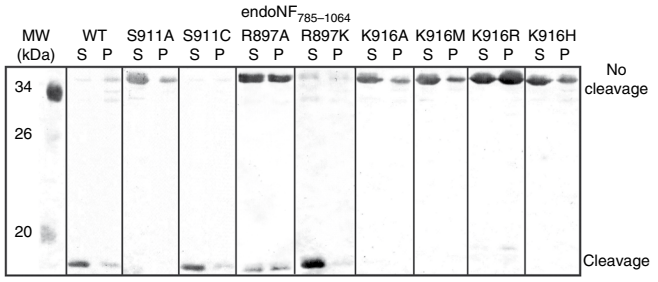

Figure 4 Biochemical characterization of the cleavage mechanism. The autoproteolytic cleavage reaction was analyzed by site-directed mutagenesis. The figure shows a western blot of a denaturing SDS-PAGE against the N-terminal StrepII tag of pre-cleavage endoNF-CIMCD. A functional cleavage reaction leads to a band at 15 kDa, corresponding to the released N-terminus; the wild-type protein ($endoNF_{785–1064}$) is shown as control. An inhibition of the cleavage reaction is indicated by a signal at 45 kDa, corresponding to the uncleaved full-length protein; the mutation reported to inhibit cleavage (S911A) is shown as control[5]. A mutation that affects the folding behavior of the triple β-helix leads to insoluble protein as reported previously[6] and is indicated by a signal in the insoluble fraction (P). S911C indicates that a different nucleophilic residue can also embrace the role of the strictly conserved Ser911. A positive charge next to the cleavage site is required for efficient cleavage, as indicated by R897A and R897K. The signal in the soluble fraction of R897K shows that a positive charge is suitable to enable the cleavage reaction. The point mutants K916A, K916M, K916R and K196H, which replace the amide group of Lys916, indicate that the presence of Lys916 is essential in order to activate Ser911.

of releasing the CIMCD[6]. Arg1035 is the terminal residue of strand β4 that forms hydrogen bonds to the side chain of Ser905 and to the backbone carbonyl of Asn906 in the triple β-helix. Furthermore, Arg1035 binds to Ala983 in the loop in front of strand β3.

Based on the CIMCD crystal structures, additional site-directed mutagenesis studies were performed to gain deeper insight into the cleavage reaction. Substituting Arg897 by alanine leads to mainly uncleaved and insoluble protein. However, the mutation R897K, which maintains the positive charge of the side chain, does not have an inhibitory effect on the cleavage reaction, leading to correctly folded soluble protein (**Fig. 4**). Thus, the positive charge plays a crucial role in the self-cleavage reaction. Arg897 forms a salt bridge to Asp886 and is located in direct proximity to the carbonyl oxygen of the scissile peptide bond between Thr910 and Ser911. This position is suitable to stabilize the intermediate oxyanion during peptide cleavage. The self-cleavage reaction depends on the catalytic residue Ser911. Substitution of Ser911 for cysteine shows that another nucleophilic residue can mimic the properties of Ser911 leading to cleaved protein, whereas substitution to alanine inhibits cleavage activity (**Fig. 4**). Hence, serine and cysteine, respectively, act as nucleophile, which has also been found in other proteins with self-cleavage activity for example, Ntn-hydrolases[12]. Activation of Ser911 depends on the presence of Lys916, as indicated by substitutions for alanine, methionine, arginine and histinde (K916A, K916M, K916R and K916H, respectively; **Fig. 4**). Mutation of Lys916 shows that this residue is crucial for the cleavage reaction, as substitution of the residue leads to uncleaved but soluble products (**Fig. 4**). In order to activate Ser911, Lys916 has to reside in the deprotonated state, which requires a considerably reduced pK_a value. The hydrophobic environment of Lys916 should result in a reduced pK_a value stabilizing its uncharged state. Similarly reduced pK_a values for lysine residues acting as general bases in peptide cleavage have previously been described for other self-cleaving proteins (for example, the *E. coli* LexA repressor[13,14]). Point mutations to histidine (K916H) or arginine (K916R) could not maintain cleavage activity (**Fig. 4**). The side chains of both residues do not match the spatial requirements necessary for an efficient activation process. However, unlike the histidine mutant, the arginine mutant is still a weak catalyst due to its size. Hence, the arginine substitution leads to an extremely small amount of product (**Fig. 4**). These effects of arginine or histidine substitutions have been previously described for other catalytic lysine-serine dyads, clearly placing the CIMCDs among the family of hydroxyl-amine proteases[15].

Conservation of the CIMCD in viruses and bacteria

A database search using the GP12-CIMCD sequence as reference revealed a high number of homologous proteins. Notably, homologous proteins are found not only in bacteriophages but also in several bacteria such as *Vibrio cholerae* (**Supplementary Fig. 3**). Analysis of flanking sequences in these bacterial genomes shows that the genes containing a CIMCD are most likely phage related. Secondary-structure prediction and alignment show a high degree of structural conservation in the CIMCD itself, with a common sequel of predicted secondary structure motifs (α-β-α-β-α-α-β-β-α). These secondary-structure motifs perfectly match the arrangement found in the two crystal structures. The fold only differs in the variable length of the tentacles, ranging from 11 to 53 residues in total. This variability in tentacle length is also observed in the crystal structures of GP12-CIMCD and pre-cleavage endoNF-CIMCD. Notably, conserved residues and predicted 8–50 β-strands are found in the region N-terminal to the canonical cleavage site (**Supplementary Fig. 3**).

DISCUSSION

Folding of endoNF strictly depends on the presence of a C-terminal intramolecular chaperone domain (CIMCD), which is removed after proper assembly of the triple β-helix by autoproteolytic, ATP-independent cleavage[5,6]. EndoNF lacking the CIMCD exclusively forms insoluble protein aggregates during biosynthesis[6].

The first step of the folding reaction is obviously trimerization of the CIMCD, indicated by the fact that released CIMCDs, as well as separately expressed CIMCDs, exist as stable homotrimers independently of the presence of an N-terminal protein. In support of this hypothesis, it could be shown that functional folding of the triple β-spiral of the human adenovirus penton fiber can also be achieved if the C-terminal domain is replaced for the 'foldon' of bacteriophage T4 fibritin. Replacing the C-terminal β-sandwich domain of the adenovirus fiber for the minimum length foldon of the bacteriophage indicates that the C-terminal domain is required for trimerization rather than for folding of the N-terminal triple β-spiral[16–18]. In order to assemble into a complex quaternary structure, many homotrimeric adhesins depend on a C-terminal anchor as a trimerization domain. However, aside from their globular shape, the C-terminal extensions share no further structural similarity with the reported CIMCDs. The parallel–β-helix tailspike proteins (TSP) of the *Podoviridae* bacteriophages P22, Phi29, HK620, and Sf6 all contain a C-terminal trimerization domain that predominantly consists of β-strands, unlike the CIMCDs, which are mainly α-helical[19–23]. In the extensively studied P22 TSP, a combination of a triple β-helix, a β-prism and an antiparallel β-sheet interact in order to form a trimerization domain[21,22]. In the recently described neck appendages of bacteriophage Phi29, a C-terminal extension, which mainly consists of two β-barrels, aids maturation of the parallel–β-helix pre-protein in an ATP-driven reaction[23]. In HK620 and Sf6 TSP, a C-terminal

β-sandwich has been found; however, it is unknown whether these domains are responsible for the assembly of the parallel β-helix[19,20].

In the case of endoNF, mere trimerization is not sufficient to initiate spontaneous self-assembly. The structure of pre-cleavage endoNF-CIMCD suggests that the highly polar tentacles of the CIMCD support the formation of the triple β-helix after trimerization. This is supported by domain-swap experiments with the highly homologous TSP endoNE. The CIMCD of endoNE is substantially shorter than the CIMCD of endoNF, specifically lacking strands β3a and β3b. Notably, a chimeric protein combining the N-terminal part of endoNF and the CIMCD of endoNE (endoNF-E) showed no enzymatic activity despite high sequence similarity, indicating a failure of trimerization and correct folding[5]. This underlines the importance of the tentacles to the folding process and suggests that the length of the tentacles correlates to the length of the triple β-helix to be formed, thereby acting as a molecular ruler.

Within the tentacles, a strictly conserved residue, Arg1035, has an important role. Site-directed mutagenesis has shown that an alanine substitution of Arg1035 disturbs the interaction between the triple β-helix and both β-strands of the tentacle, resulting in an insoluble, incorrectly folded protein[6]. This sensitivity to spatial rearrangements is reflected in the conservation of secondary-structure elements and shows the important role of the tentacles in mediating proper triple–β-helix assembly. Moreover, this indicates that correct positioning of the tentacles is crucial to the folding of the triple β-helix. This framework of the homotrimeric CIMCD forces the three N-terminal polypeptide chains to fold into an intertwined triple β-helix.

Upon successful folding, the tentacles are attached to the outer surface of the triple β-helix, which finally leads to positioning of the side chain of Arg897 in close proximity to the carbonyl oxygen of the scissile peptide bond between the CIMCD and mature endoNF, triggering the cleavage reaction. Proteolytic cleavage efficiency is markedly increased by stabilization of the intermediate oxyanion at the scissile peptide bond. Stabilization of the negative charge is provided by the countercharge of Arg897 (**Figs. 2c** and **4**). The scissile peptide bond is further destabilized by the unfavorable geometry of the protein backbone in the pre-cleavage state. It is tempting to speculate that the hydrogen bond between Thr910 and Asn912 introduces a kink into the backbone that generates a high tension at the N terminus of the CIMCD. In the crystal structure of the released GP12-CIMCD, its N-terminal residues adopt a very different conformation, leading to a location at the outer rim of the core. This conformational change induced by the peptide bond cleavage might also force the dissociation of the mature protein and its CIMCD. Concludingly, consistent with the MS results, the structural and biochemical data clearly show that the cleavage process is based on a serine-lysine catalytic dyad. The cleavage reaction strictly depends on the correct orientation of all involved residues; however, this interplay is only achieved in a correctly folded protein. Therefore, the cleavage reaction acts as a control mechanism for a completely folded triple β-helix.

The structure analysis raises the question of whether the same structural pattern can be found in other C-terminal extensions. Not only residues involved in the cleavage reaction are conserved among the homologous CIMCDs; the crystal structures of GP12-CIMCD and pre-cleavage endoNF-CIMCD also show that at all positions that form crucial parts of the CIMCDs, conserved residues are found. Furthermore, these conserved residues participate in the assembly of the core of the three subunits via stabilization of the central three-helix bundle. The presence and arrangement of the conserved residues within the secondary-structure motifs indicate an identical tertiary and quaternary structure for all the homologous CIMCDs found in the database. Interestingly, the conservation of sequence elements is not restricted to the CIMCD but also expands into the triple β-helix, encompassing the residues upstream of the cleavage site and even residues predicted to be part of the first upstream β-strands. This indicates a conservation of tertiary structure elements: namely, β-helices within the N-terminal sequences. According to the high similarity of the CIMCD sequences, it is likely that all N-terminal sequences assemble into trimers. Hence, these data suggest that all CIMCDs homologous to endoNF-CIMCD and GP12-CIMCD act as intramolecular chaperones involved in the folding of triple β-helices.

An investigation for protein structures similar to those of the endoNF-CIMCD and GP12-CIMCD within the protein data bank (PDB) revealed no significant similarity to known three-dimensional structures (maximum Dali score = 3.5)[24]. However, the overall shape as well as the distribution of hydrophobic residues of endoNF-CIMCD and GP12-CIMCD clearly resembles other chaperones containing clamp-like binding sites for example, prefoldin[25] (PFD, PDB 1FXK) or the 17-kDa protein[26] (SKP, PDB 1SG2) (**Fig. 5**). These chaperones also display basket-like structures, which are formed by protrusions exposing from a central base. Despite the general similarity in shape, these chaperones completely differ in their fold with respect to that

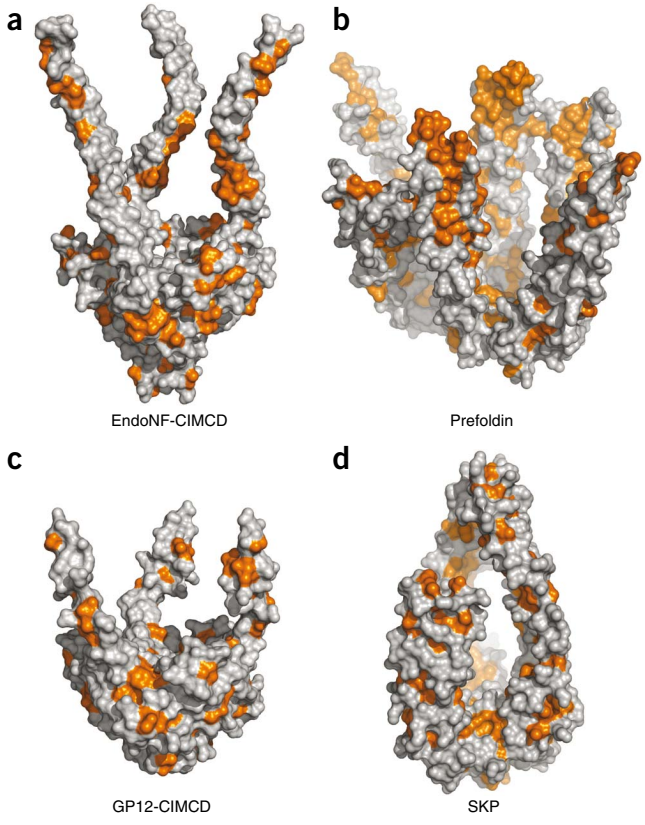

Figure 5 Comparison to other clamp-like binding proteins. (**a**) endoNF-CIMCD: C-terminal intramolecular chaperone domain of the endosialidase from bacteriophage K1F. (**b**) Prefoldin (PDB 1FXK[25]). (**c**) GP12-CIMCD: C-terminal intramolecular chaperone domain of the neck appendage protein GP12 of *Bacillus* phage GA1. (**d**) SKP: seventeen kilodalton protein (PDB 1SG2) (ref. 27). Hydrophobic regions are mapped as orange patches on the surface of the proteins. Interestingly, the hydrophobic areas of the CIMCD's and prefoldin are concentrated to the tip and center of the tentacles facing inside the 'basket'. SKP displays a more even distribution of hydrophobic areas at the outside and inside of the tentacles, suggesting a different folding pathway.

of the CIMCDs. The platform of PFD and SKP consists mainly of β-strands, while the tentacle-like protrusions are exclusively α-helical. In contrast to these two chaperones the CIMCDs express the opposite secondary-structure composition, with the tentacles being antiparallel β-strands and the core being mainly α-helical. For prefoldin, it has been shown that the tentacles are involved in substrate recognition and binding. To allow for binding to a large variety of substrates, a high flexibility of the tentacles was suggested. This allows accommodating proteins of various sizes and increases the ability of the chaperone to act on multiple targets[26].

The high degree of similarity in the structural framework suggests convergent evolution of folding strategies. Despite different origins, two different solutions evolved in chaperones to form basket-like frameworks that promote and control the folding process. Clamp-like binding sites are found in various chaperones, including PFD, SKP, Tim9-Tim10, trigger factor and HSP40/70[27], suggesting a common folding strategy for these chaperones. Binding to the long protrusions transiently stabilizes the target protein until the hydrophobic residues fold into a stable core. Upon folding, the interaction with the tentacles gets weaker and the chaperone releases the protein. Examples for this strategy have been found in all domains of life, which can now be extended to viruses and the CIMCDs conserved in triple β-helix–containing proteins.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Protein Data Bank: Coordinates and structure factors have been deposited with accession codes 3GUD (GP12-CIMCD) and 3GW6 (pre-cleavage endoNF-CIMCD).

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

The authors would like to thank the beamline staff scientists of European Synchrotron Radiation Facility, Dr. P. Tucker from the European Molecular Biology Laboratory for outstanding support during data collection at the Deutsches Elektronen Synchrotron and D. Gloth and J. Warweg for technical assistance.

AUTHOR CONTRIBUTIONS

E.-C.S. performed molecular cloning, protein purification and crystallization, data collection, integration and refinement, data analysis and writing of the manuscript; A.D., sequence comparison and alignments and writing of the manuscript; H.U., MS analysis; A.S., molecular cloning; M.M., K.S., D.S., feedback on the manuscript; R.G.-S., initiator of the project; R.F., data analysis, writing of the manuscript and acted as project leader.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
2. Fenton, W.A. & Horwich, A.L. GroEL-mediated protein folding. *Protein Sci.* **6**, 743–760 (1997).
3. Hartl, F.U. Molecular chaperones in cellular protein folding. *Nature* **381**, 571–579 (1996).
4. Chen, Y.J. & Inouye, M. The intramolecular chaperone-mediated protein folding. *Curr. Opin. Struct. Biol.* **18**, 765–770 (2008).
5. Mühlenhoff, M., Stummeyer, K., Grove, M., Sauerborn, M. & Gerardy-Schahn, R. Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. *J. Biol. Chem.* **278**, 12634–12644 (2003).
6. Schwarzer, D., Stummeyer, K., Gerardy-Schahn, R. & Muhlenhoff, M. Characterization of a novel intramolecular chaperone domain conserved in endosialidases and other bacteriophage tail spike and fiber proteins. *J. Biol. Chem.* **282**, 2821–2831 (2007).
7. Stummeyer, K., Dickmanns, A., Muhlenhoff, M., Gerardy-Schahn, R. & Ficner, R. Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F. *Nat. Struct. Mol. Biol.* **12**, 90–96 (2005).
8. Weigele, P.R., Scanlon, E. & King, J. Homotrimeric, β-stranded viral adhesins and tail proteins. *J. Bacteriol.* **185**, 4022–4030 (2003).
9. Kajava, A.V. & Steven, A.C. β-rolls, β-helices, and other β-solenoid proteins. *Adv. Protein Chem.* **73**, 55–96 (2006).
10. Papanikolopoulou, K. *et al.* Amyloid fibril formation from sequences of a natural β-structured fibrous protein, the adenovirus fiber. *J. Biol. Chem.* **280**, 2481–2490 (2005).
11. Mitraki, A., Papanikolopoulou, K. & Van Raaij, M.J. Natural triple β-stranded fibrous folds. *Adv. Protein Chem.* **73**, 97–124 (2006).
12. Oinonen, C. & Rouvinen, J. Structural comparison of Ntn-hydrolases. *Protein Sci.* **9**, 2329–2337 (2000).
13. Silaty, S.N. & Vu, H.K. The role of electrostatic interactions in the mechanism of peptide bond hydrolysis by a Ser-Lys catalytic dyad. *Protein Eng.* **4**, 919–922 (1991).
14. Dao-pin, S., Anderson, D.E., Baase, W.A., Dahlquist, F.W. & Matthews, B.W. Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry* **30**, 11521–11529 (1991).
15. Lin, L.L. & Little, J.W. Autodigestion and RecA-dependent cleavage of Ind- mutant LexA proteins. *J. Mol. Biol.* **210**, 439–452 (1989).
16. Papanikolopoulou, K., Forge, V., Goeltz, P. & Mitraki, A. Formation of highly stable chimeric trimers by fusion of an adenovirus fiber shaft fragment with the foldon domain of bacteriophage t4 fibritin. *J. Biol. Chem.* **279**, 8991–8998 (2004).
17. Papanikolopoulou, K. *et al.* Adenovirus fibre shaft sequences fold into the native triple β-spiral fold when N-terminally fused to the bacteriophage T4 fibritin foldon trimerisation motif. *J. Mol. Biol.* **342**, 219–227 (2004).
18. van Raaij, M.J., Mitraki, A., Lavigne, G. & Cusack, S. A triple β-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**, 935–938 (1999).
19. Barbirz, S. *et al.* Crystal structure of *Escherichia coli* phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related. *Mol. Microbiol.* **69**, 303–316 (2008).
20. Müller, J.J. *et al.* An intersubunit active site between supercoiled parallel β helices in the trimeric tailspike endorhamnosidase of *Shigella flexneri* Phage Sf6. *Structure* **16**, 766–775 (2008).
21. Steinbacher, S. *et al.* Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 A, fully refined structure of the endorhamnosidase at 1.56 A resolution, and the molecular basis of *O*-antigen recognition and cleavage. *J. Mol. Biol.* **267**, 865–880 (1997).
22. Steinbacher, S. *et al.* Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science* **265**, 383–386 (1994).
23. Xiang, Y. *et al.* Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. *Mol. Cell* **34**, 375–386 (2009).
24. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**, 2780–2781 (2008).
25. Siegert, R., Leroux, M.R., Scheufler, C., Hartl, F.U. & Moarefi, I. Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell* **103**, 621–632 (2000).
26. Korndörfer, I.P., Dommel, M.K. & Skerra, A. Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture. *Nat. Struct. Mol. Biol.* **11**, 1015–1020 (2004).
27. Stirling, P.C., Bakhoum, S.F., Feigl, A.B. & Leroux, M.R. Convergent evolution of clamp-like binding sites in diverse chaperones. *Nat. Struct. Mol. Biol.* **13**, 865–870 (2006).
28. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041 (2001).

ONLINE METHODS

DNA constructs. The plasmid encoding GP12 used for overexpression was previously described[6]. We generated an N-terminal deletion construct of endoNF$_{\Delta 1–784}$, with the point mutation S911A (denoted as pre-cleavage endoNF-CIMCD) and cloned it into a modified pET22b expression vector[6]. We performed point mutations (R897A, R897K, S911C, K916M, K916A, K916R and K916H) of endoNF using the Site-Directed Mutagenesis Kit (Stratagene) following the manufacturers instructions. We verified all constructs by DNA sequencing.

Expression and purification. We expressed GP12 pre-protein with an N-terminal StrepII tag (IBA) and a C-terminal His$_6$-tag. Expression was performed in *E. coli* BL21(DE3) cells at 18 °C for 18 h. We separated the autoproteolytically cleaved N-terminal domain from its CIMCD by StrepII-tag affinity chromatography. We achieved purification by Ni^{2+}-NTA chromatography (100 mM HEPES, pH 7.5; 500 mM NaCl; 5 mM imidazole) (GE). We further purified the CIMCD by size-exclusion chromatography using a Superdex S-75 gelfiltraton column (10 mM Tris, pH 8.0, 50 mM NaCl) (GE). We concentrated pure CIMCD to 12.5 mg ml^{-1}, flash froze it in liquid nitrogen and stored it at −80 °C. We overexpressed the selenomethinonine (SeMet) derivative of GP12-CIMCD in *E. coli* B834(DE3) cells grown in M9-minimal medium supplemented with 50 mg l^{-1} D- and L-SeMet (Sigma) at 18 °C for 18 h and purified it as the wild-type protein. We performed expression of pre-cleavage endoNF-CIMCD as described above. We performed purification by StrepII-tag affinity chromatography (100 mM Tris, pH 7.5, 150 mM NaCl), before size-exclusion chromatography using a Superdex S-200 column (10 mM Tris, pH 7.5, 50 mM NaCl) (GE). We concentrated the protein to 10 mg ml^{-1}, flash froze it in liquid nitrogen and stored it at −80 °C.

Cleavage assay. We expressed active-site mutants of endoNF-CIMCD like the wild-type protein. After lysis and ultracentrifugation (20.000*g*), we re-dissolved the pellet in a buffer volume equal to the soluble fraction (100 mM Tris, pH 7.5; 150 mM NaCl; 1 mM EDTA). In order to immobilize the StrepII-tagged protein, we incubated the soluble fraction 1:5 (v/v) with Streptactin beads for 1.5 h at 20 °C. We washed beads two times with tenfold amount of buffer and re-dissolved them in 2 ml of buffer. We heat-denatured soluble and pelleted fractions (99 °C, 10 min) and applied them to polyacrylamide gel electrophoresis. We determined the presence of StrepII-tagged protein by western blot analysis using the Strep-AP Detection Kit (IBA) following the manufacturer's instructions.

Crystallization and data collection. We grew crystals of GP12-CIMCD and pre-cleavage endoNF-CIMCD in sitting drop vapor diffusion crystallization plates. We obtained crystals in 25–30% (w/v) PEG 2000, 0.1 M sodium citrate, pH 4.8; 0.1 M $(NH_4)_2SO_4$, 20 °C. Microseeding yielded crystals suitable for diffraction experiments. We crystallized the SeMet-derived protein accordingly. We grew crystals of pre-cleavage endoNF-CIMCD in 15% (w/v) PEG 8000, 0.1 M bis-Tris, pH 6.0, 0.15 M $CaBr_2$, 20 °C. We cryoprotected crystals of both proteins by a 1 min soak in mother liquor containing 12.5% (v/v) 2,3-butanediol and flash froze them (100 K) for data collection. We performed X-ray diffraction data collection of the native dataset of GP12-CIMCD to a resolution of 2.2 Å at beamline 14.2 at the European Synchrotron Radiation Facility (Grenoble, France). Crystals of GP12-CIMCD belong to spacegroup P3 but are merohedrally twinned and contain two monomers per asymmetric unit. We collected a three-wavelength MAD dataset of SeMet-derived GP12-CIMCD at the European Molecular Biology Laboratory beamline X13 (Deutsches Elektronen Synchrotron). We collected a native dataset of a pre-cleavage endoNF-CIMCD crystal to a resolution of 2.6 Å at EMBL beamline X13 at the Deutsches Elektronen Synchrotron. Crystals of pre-cleavage endoNF-CIMCD belong to spacegroup P1 and contain two molecules per asymmetric unit. We collected all datasets at 100 K.

Structure determination and refinement. We processed all diffraction images using the XDS suite[29]. We solved the structure of GP12-CIMCD by means of SIRAS using the native dataset and the derivative dataset at the peak wavelength in AUTOSHARP[30]. We built an initial model using RESOLVE[31] and manually completed the model using Coot[32]. As we obtained the native dataset from a merohedrally twinned crystal (apparent spacegroup P6), we refined the structure using the native dataset in spacegroup P3, applying the respective twin-operator in PHENIX[33–35]. We determined the structure of pre-cleavage endoNF-CIMCD by molecular replacement using the β-helix domain of endoNF$_{785–910}$, PDB 1V0E[7] and the structure of GP12-CIMCD as a search model in MOLREP[36] and PHASER[37]. We completed the structure in iterative cycles of building (Coot) and refinement (PHENIX) by applying automatically derived NCS operators and simulated annealing. Both structures have been refined to reasonable stereochemistry (**Table 1** and **Supplementary Fig. 1**).

Sequence and structure analysis. We determined homologous proteins by protein-protein BLAST using the NCBI BLAST server[38]. We performed multiple structure-based sequence alignments manually. We prepared figures using Pymol (http://www.pymol.org) and the ESPRIPT server[39,40].

Mass spectrometry. We diluted purified SeMet-derived GP12-CIMCD (concentrations 10 mg ml^{-1}, in buffer 10mM Tris, pH 8.0, 100 mM NaCl) 1:10 with 50% (v/v) acetonitrile, 0.2% (v/v) formic acid and analyzed it on a Q-ToF mass spectrometer (Ultima, Waters) using a static needle with a capillary voltage of 1,300 V. We recorded spectra from 400–1600 *m/z*, processed them with Mass Lynx software 3.0 (Waters) and deconvoluted them with MaxEnt software (Waters).

29. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
30. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
31. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).
32. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
33. Adams, P.D. *et al.* Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat.* **11**, 53–55 (2004).
34. Terwilliger, T.C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69 (2008).
35. Zwart, P.H. *et al.* Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435 (2008).
36. Lebedev, A.A., Vagin, A.A. & Murshudov, G.N. Model preparation in MOLREP and examples of model improvement using X-ray data. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 33–39 (2008).
37. McCoy, A.J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 32–41 (2007).
38. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
39. Gouet, P., Courcelle, E., Stuart, D.I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
40. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).

doi:10.1038/nsmb.1746