Isopeptide Bonds

DOI: 10.1002/anie.201004340

NMR Spectroscopic and Theoretical Analysis of a Spontaneously Formed Lys–Asp Isopeptide Bond**

*Robert M. Hagan, Ragnar Björnsson, Stephen A. McMahon, Benjamin Schomburg, Vickie Braithwaite, Michael Bühl, James H. Naismith, and Ulrich Schwarz-Linek**

In memory of Peter Welzel

Isopeptide bonds, which are amide bonds involving the ε-amino group of lysine, were first identified to result from heat treatment of proteinaceous materials and fibrinogen activation.[1] Their central role in ubiquitination is their most widely known biological occurrence.[2] Self-generated isopeptide bonds between side chains of lysine and asparagine residues have emerged as a hallmark of surface proteins of Gram-positive bacteria after their discovery in the major pilin subunit of *Streptococcus pyogenes*.[3] Subsequently isopeptides were identified in proteins known to form, or be associated with, pili.[4–8] All bacterial isopeptides are found in β-sheet domains resembling the CnaA or CnaB folds of protein Cna from *Staphylococcus aureus*.[9,10] CnaA and CnaB domains are predicted to occur in thousands of bacterial surface proteins, and isopeptide bonds emerge as a very common posttranslational modification underpinning Gram-positive pilus formation and stability. In bacterial pilus proteins isopeptide bond formation depends on a catalytic glutamate or aspartate residue. It is thought to require location of the isopeptide triad (Lys, Asn, catalytic carboxyl group) within the hydrophobic core.[3] We have analyzed the isopeptide formed by spontaneous amidation of an aspartate side chain in a CnaB fold of a protein that has not previously been implicated in bacterial pili.

FbaB is a fibronectin-binding protein of invasive *S. pyogenes* strains.[11] It contains several intrinsically disordered sequence repeats that form high-affinity complexes with the human target protein.[12] The function of all other domains of FbaB, including two predicted CnaB domains (Figure 1a), is unknown. One of these CnaB domains, referred to from here on as CnaB2, is found embedded in the natively unfolded fibronectin-binding repeats. The structure of CnaB2 and the presence of the isopeptide were identified by X-ray crystallography.[13] While many proteins have been found or predicted to contain Asn–Lys isopeptides,[3] several potential Asp–Lys CnaB domains can now be added to the rapidly

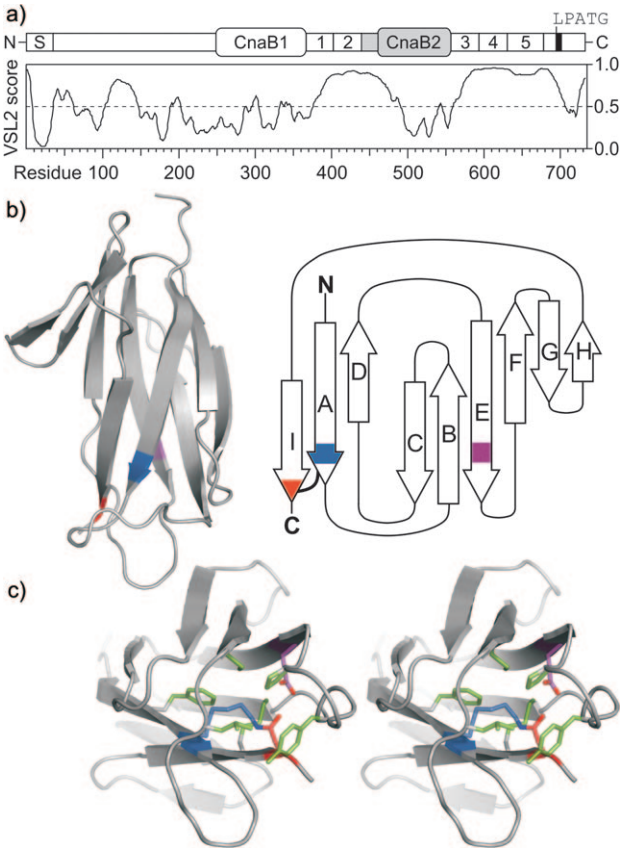

Figure 1. a) Domain organization of FbaB. S signal sequence, 1–5 fibronectin-binding repeats, LPATG cell-wall anchor. CnaB domains were predicted by Pfam[21] and PHYRE.[22] The construct used in this study is shaded gray. The bottom panel shows a VSL2 disorder prediction[23] for FbaB. Regions with a score above 0.5 are predicted to be intrinsically disordered. b) Ribbon representation and topology of CnaB2 (PDB code 2X5P). The positions of isopeptide triad residues are marked blue (K470), purple (E516), and red (D556). c) Stereo view of the isopeptide triad (colors as in (b)) and surrounding hydrophobic residues (green).

[*] Dr. R. M. Hagan, Dr. S. A. McMahon, B. Schomburg, V. Braithwaite, Prof. J. H. Naismith, Dr. U. Schwarz-Linek
Biomedical Sciences Research Complex, University of St Andrews
North Haugh, St Andrews KY16 9ST (UK)
Fax.: (44) 1334-462595
E-mail: us6@st-andrews.ac.uk
R. Björnsson, Prof. M. Bühl
School of Chemistry, University of St Andrews
North Haugh, St Andrews KY16 9ST (UK)

[**] We thank Dr. D. Uhrin (Edinburgh) and Dr. T. Lebl for advice on NMR spectroscopy, Dr. C. Botting for MS, M. Taylor for help with thermal shift assays, and Dr. S. Talay (Braunschweig) for her kind gift of *S. pyogenes* DNA. M.B. and R.B. thank EaStCHEM for support and for access to the Research Computing Facility and Dr. H. Früchtl for technical assistance. This work was funded in part by a Value in People award of the Wellcome Trust.

Supporting information for this article, including additional experimental details, is available on the WWW under http://dx.doi.org/10.1002/anie.201004340.

growing list of isopeptide proteins (Figure S1 in the Supporting Information).[14] CnaB2 is unusual as it occurs isolated in sequence and represents an independently folded domain ideally suited to address questions relating to the structural role of isopeptide bonds and the mechanism of their formation. Using NMR spectroscopy, circular dichroism (CD) spectroscopy, and thermal shift assays, we show that the isopeptide bond has a dramatic effect on protein dynamics and stability. We present a mechanism for isopeptide formation that is based on quantum mechanical/molecular mechanical (QM/MM) calculations.

In CnaB2 the isopeptide is formed by residues K470 and D556. The side chain of E516 appears in a position similar to the ones observed for catalytic side chains of other CnaB-like domains.[3,5] The isopeptide cross-links the parallel N-terminal and C-terminal β strands and is surrounded by hydrophobic groups (Figure 1b,c). CnaB2 and mutants K470A, E516Q, and D556 A, expected to lack the isopeptide bond, were expressed in unlabeled, ^{15}N-, or ^{13}C,^{15}N- (CnaB2, E516Q) labeled forms. As all previously found bacterial isopeptide bonds are formed by asparagine, a mutant D556N was also generated. Isopeptide-lacking mutants differed in electrophoretic mobility from native CnaB2 (Figure S2 in the Supporting Information). Absence of the isopeptide in mutant E516Q supports a role for E516 in catalysis. The D556N mutant gave rise to two bands in SDS-PAGE at positions corresponding to proteins without and with an isopeptide bond. A time course of the D556N mutant shows that isopeptide bond formation was complete after 24 h (Figure S2c in the Supporting Information, confirmed by MS). ^{1}H,^{15}N HSQC spectra of all mutants were typical for folded β-sheet proteins of approximately 13 kDa (Figure S3 in the Supporting Information). Spectra of CnaB2 and D556N in which isopeptide bond formation was complete were indistinguishable, thus reflecting their convergence to the same product from different start points. Although ammonia is a better leaving group than water, isopeptide bond formation in D556N was comparatively slow. Therefore it appears that the environment of the isopeptide bond is optimized for the particular reactive residue (acid or amide). A detailed study of this environment in Lys–Asn and Lys–Asp isopeptide proteins will be required to deconvolute the key factors in the surrounding site.

CnaB2 and E516Q were sequentially assigned using a standard triple-resonance NMR spectroscopy approach. The mutation did not affect the overall structure, as evident from the similarity of ^{13}C chemical shifts (Figure S4a in the Supporting Information). The isopeptide NH signal, found at an unusual upfield ^{1}H chemical shift of $\delta = 6$ ppm (Figure 2a), was readily identified by a set of resonances in the HNCACB experiment that distinguish it from backbone amide groups (Figure 2c). HNCACB spectra show strong signals for Cα/Cβ resonances of residue *i* and weak signals for Cα/Cβ of residue *i*−1. Assignment of the weak cross-peaks in the K470ζ strip of the HNCACB spectrum to D556 Cα and Cβ was verified by the CBCA(CO)NH experiment, which connects amide NH groups with Cα and Cβ of the preceding residue. The assignment resulting from NMR spectroscopy experiments unequivocally identified the presence of the isopeptide and linked it to D556, further validating the X-ray structure[13] and MS (Figure S2d in the Supporting Information) results.

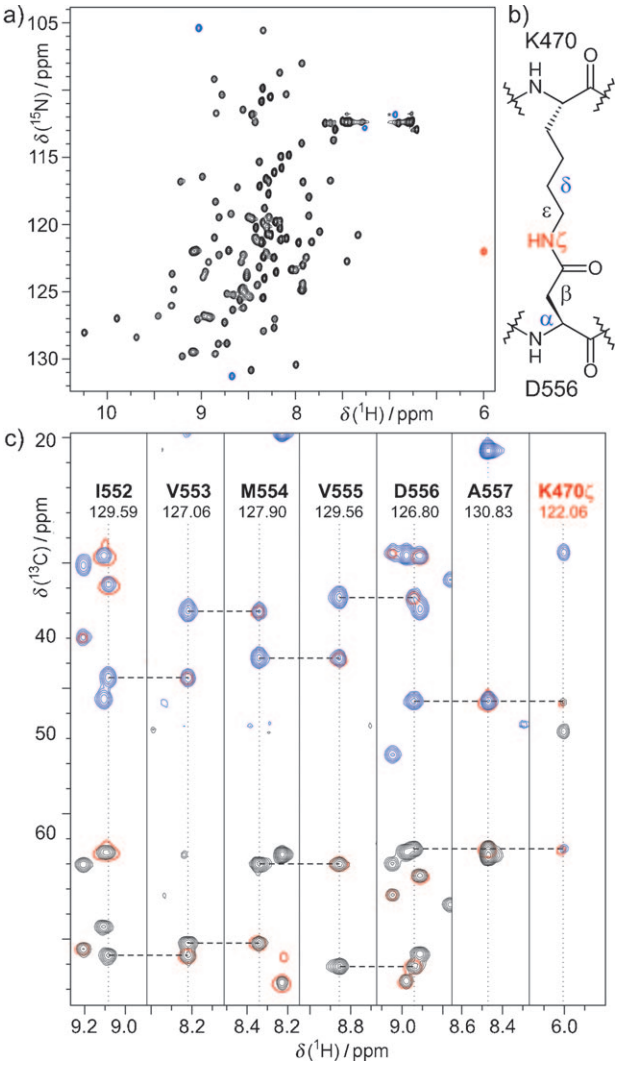

Figure 2. a) ^{1}H,^{15}N HSQC spectrum of CnaB2. The isopeptide NH cross-peak is shown in red. Blue signals are aliased in the ^{15}N dimension. b) Scheme of the isopeptide bond and the resonances observed in the K470ζ HNCACB strip. c) Strips from HNCACB and CBCA(CO)NH (single red contours) spectra for residues in β strand I and the isopeptide (K470ζ). ^{15}N chemical shifts (ppm) are shown on top. HNCACB Cα and Cβ cross-peaks appear in black (positive phase) and blue (negative phase), respectively. The phases of the D556 resonances in the K470ζ strip are swapped. The strong signals in this strip correspond to K470 Cδ (blue) and Cε (black). Dashed horizontal lines mark the "backbone walk" of the sequential assignment.

To assess the impact of the isopeptide on protein dynamics, heteronuclear NOEs (hNOEs) were measured for CnaB2 and E516Q (Figure 3a). Overall the E516Q mutant showed slightly higher flexibility. Protein samples that had been exchanged into D_2O still gave rise to HSQC cross-peaks even after an incubation period of several weeks, indicating a particularly stable structure. For E516Q, 18 non-exchanging amide protons were exclusively located in β sheets. In CnaB2, 33 non-exchanging protons, including the isopeptide NH

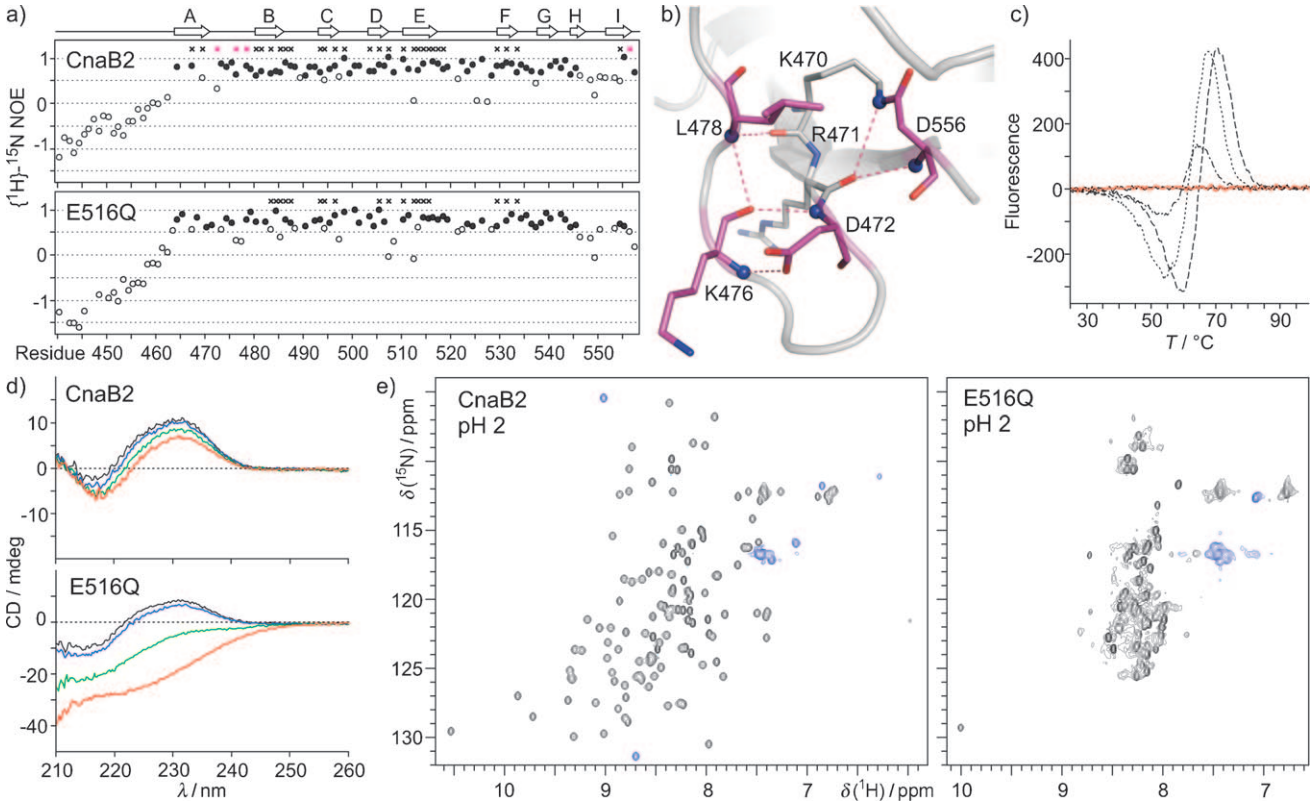

Figure 3. a) $\{^1H\}^{15}N$ heteronuclear NOEs of CnaB2 and the E516Q mutant. Open circles indicate NOE values smaller than 0.6, a threshold for NH groups found in a stable structured context.[24] The N-terminal 21 residues exhibit extensive motion on the pico- to nanosecond timescale. Non-exchanging amide protons are indicated by crosses. b) Network of H-bonds in the AB loop involving non-exchanging amide NH groups (N atoms shown as balls). Respective residues are shown as magenta sticks and correspond to magenta crosses in (a). c) Thermal shift assays. The minima correspond to denaturation temperatures of 53 °C (K470A, –•–•), 54 °C (E516Q, •••••), and 59 °C (D556A, ––––). CnaB2 (red solid line) and D556N (black solid line) did not undergo denaturation below 100 °C. d) Far-UV CD spectra of CnaB2 and E516Q were recorded at 20 (black), 37 (blue), 60 (green), and 80 °C (red). e) $^1H,^{15}N$ HSQC spectra of CnaB2 and E516Q at pH 2. Blue cross-peaks are aliased in the ^{15}N dimension.

proton, were found. Three of these (D472, K476, L478) are located in the A–B loop and undergo fast exchange in E516Q. In the crystal structure[13] they form a network of H-bonds involving the isopeptide NH proton, backbone atoms of K470 and D556, and a conserved acidic side chain (D472, Figure 3b). These H-bonds are obviously preserved in solution, but only in the presence of the isopeptide bond. The notion of this long-range effect of the isopeptide on protein dynamics is further supported by a comparison of NH chemical shifts of CnaB2 and E516Q. Almost all significant differences are located in loop regions (Figure S4b in the Supporting Information) including the A–B loop, and correlate with D_2O exchange data. Isopeptide bonds lend thermal and proteolytic stability to pilus proteins.[15] Denaturation temperatures of CnaB2 and mutants were determined in fluorescence-based thermal shift assays. Mutants lacking the isopeptide unfolded at temperatures between 53–59 °C, whilst native CnaB2 did not undergo noticeable denaturation below 100 °C (Figure 3c). This dramatic effect of the isopeptide was confirmed by CD spectroscopy (Figure 3d). At 80 °C, spectra of all mutants lacking the isopeptide were typical for polypeptides devoid of any secondary structure. The spectrum of CnaB2 was largely unaffected by temperature increase. An $^1H,^{15}N$ HSQC spectrum of CnaB2 at pH 2 showed signal dispersion and line widths similar to spectra recorded at pH 6. HSQC spectra of E516Q showed increasing line broadening from pH 4 and characteristics of denatured protein at pH 2 (Figure 3e). Taken together, these data establish that the isopeptide bond imparts a remarkable conformational, thermal, and pH stability to CnaB2.

To shed light on the mechanism of isopeptide bond formation and the role of the critical E516 residue, we performed QM/MM calculations.[16] Starting from X-ray coordinates,[13] a solvated model for CnaB2 with separated K470 and D556 side chains was prepared and equilibrated using classical molecular dynamics (MD) simulations. A full reaction profile was constructed by QM/MM optimizations of reactant, product, key intermediates, and transition states. As a detailed discussion of the full path is beyond the scope of this Communication, we only highlight the key findings. In the lowest energy structure before isopeptide bond formation the three critical groups are neutral. This result reflects a perturbation of pK_a values commonly observed for buried functional groups involved in enzyme catalysis.[17] The D556 carboxyl is H-bonded to both K470 and E516 (**1**, Figure 4a). It costs very little energy to direct the H-bond from K470 Nζ to

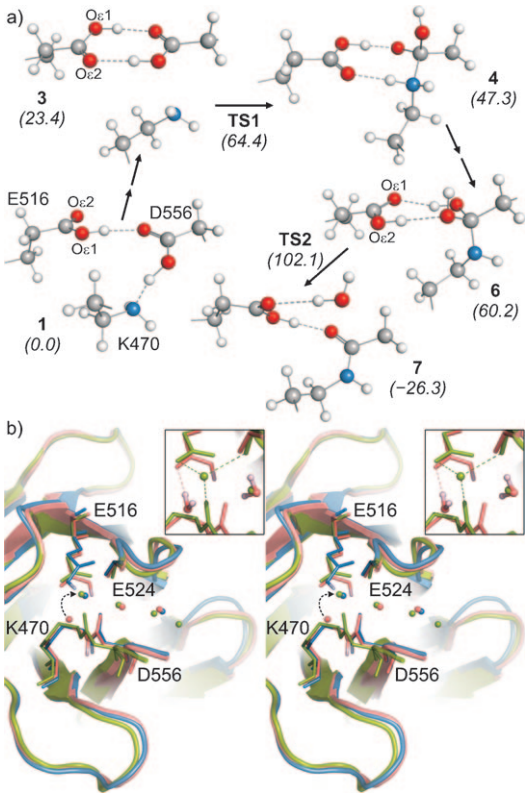

Figure 4. a) Salient intermediates of isopeptide bond formation (only key part of the QM region is shown); in parentheses: relative energies at the M06-2X/MM level (in kJ mol^{-1}). b) Overlay of the isopeptide region in crystal structure (green) and computational models after QM/MM (pink) and 1 ns MD (cyan). Water molecules are shown as balls. The arrow indicates movement of the eliminated water to the position observed in both the MD and the crystal structure. Inset: H-bonds of this water molecule in these two locations.

E516 Oε2 (**3**), so that the N atom is positioned for nucleophilic attack at the carbonyl center. This attack can proceed smoothly through transition state **TS1**, with a total barrier of only 64 kJ mol^{-1}, thus affording zwitterionic **4**. The tetrahedral intermediate **6** is reached via proton transfers from E516 to D556 and from K470 to E516 in an eight-membered transition state very similar in energy to **TS1**. This proton shuttle explains the pivotal role of catalytic carboxyl groups in isopeptide bond formation. A carboxamide group (as in E516Q) effectively shuts off this channel because it would require formation of an unfavorable C(OH)(=NH) moiety, thereby rationalizing the inability of the mutant to form an isopeptide. Direct water elimination from **6** to yield product **7** is computed to have the highest barrier of the pathway (**TS2**, 102 kJ mol^{-1}) and is thus indicated to be the rate-limiting step. The released water (after MD simulation starting from the QM/MM optimized product) is located in a position very close to where a water molecule is found in the crystal structure (Figure 4b), a remarkable agreement between computational and experimental models. This water molecule is stabilized by three H-bonds, and in the crystal structure is linked to the bulk solution by a linear array of three water molecules. This array likely represents a channel allowing the eliminated water to dissociate from the reaction center (Figure 4b). The proposed reaction mechanism explains the absolute requirement for a catalytic residue and a hydrophobic environment; it allows for the formation of isopeptide bonds from both asparagine and aspartic acid residues. The mechanism and the key intermediates are reminiscent of peptide bond formation in the peptidyl transferase center of the ribosome, which proceeds through a zwitterionic tetrahedral intermediate and depends on proton shuttling mediated by the 2′ OH group of adenosine 76 and potentially a water molecule in the P site.[18] These groups appear to play roles analogous to the two oxygen atoms in the carboxyl group of E516 in CnaB2.

In summary, we have applied a variety of experimental and computational techniques to study the structural role of an isopeptide bond and the mechanism of its spontaneous formation in a key building block of bacterial surface proteins. Our data imply that this type of bond is far more common than previously realized and give important insights into the purpose of isopeptide bonds and the structural requirements for their formation.

Experimental Section

Site-directed mutagenesis was carried out using the QuikChange protocol (Stratagene) according to supplier's guidelines. All proteins were expressed and purified as previously described for CnaB2.[13] Isotope labeling was achieved by expression in minimal media supplemented with U-^{13}C glucose and/or $^{15}NH_4Cl$.

Samples for NMR spectroscopy were prepared in 10 mM phosphate buffer, pH 6.0, containing 5% D_2O and 0.02% NaN_3. NMR spectroscopy experiments were performed at 25 °C on a Bruker DRX500 spectrometer equipped with a 5 mm TXIz probe. All spectra were processed with NMRPipe[19] and analyzed with CCPN Analysis 1.0.[20]

Thermal shift assays were carried out with protein samples at 20 μM in phosphate-buffered saline (PBS) in the presence of Sypro-Orange (Invitrogen) in a final volume of 50 μL. Fluorescence was recorded using a QPCR instrument (Stratagene) over a temperature range of 25–100 °C with a 1 °C min^{-1} gradient.

CD spectra were recorded using a Jasco J-270 Spectropolarimeter over 210–260 nm at protein concentrations of 10 μM in PBS. Each sample and a control (PBS) was scanned in sextuplicate at 20, 37, 60, and 80 °C.

QM/MM optimizations were performed at the B3LYP/6-31 + G-(d,p)/CHARMM level, followed by M06-2X/6-311 + G(3df,3pd)/CHARMM single points, using the ChemShell software.

Received: July 15, 2010
Published online: September 28, 2010

Keywords: computational chemistry · isopeptide bonds · NMR spectroscopy · protein structures

[1] R. S. Asquith, M. S. Otterburn, W. J. Sinclair, *Angew. Chem.* **1974**, *86*, 580–587; *Angew. Chem. Int. Ed. Engl.* **1974**, *13*, 514–520.
[2] C. M. Pickart, *Annu. Rev. Biochem.* **2001**, *70*, 503–533.
[3] H. J. Kang, F. Coulibaly, F. Clow, T. Proft, E. N. Baker, *Science* **2007**, *318*, 1625–1628.
[4] J. M. Budzik, L. A. Marraffini, P. Souda, J. P. Whitelegge, K. F. Faull, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10215–10220.

[5] H. J. Kang, N. G. Paterson, A. H. Gaspar, H. Ton-That, E. N. Baker, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18427 – 18427.
[6] L. El Mortaji, R. Terrasse, A. Dessen, T. Vernet, A. M. Di Guilmi, *J. Biol. Chem.* **2010**, *285*, 12405 – 12415.
[7] N. Forsgren, R. J. Lamont, K. Persson, *J. Mol. Biol.* **2010**, *397*, 740 – 751.
[8] T. Izoré, C. Contreras-Martel, L. El Mortaji, C. Manzano, R. Terrasse, T. Vernet, A. M. Di Guilmi, A. Dessen, *Structure* **2010**, *18*, 106 – 115.
[9] J. Symersky, J. M. Patti, M. Carson, K. House-Pompeo, M. Teale, D. Moore, L. Jin, A. Schneider, L. J. DeLucas, M. Hook, S. V. Narayana, *Nat. Struct. Biol.* **1997**, *4*, 833 – 838.
[10] C. C. S. Deivanayagam, R. L. Rich, M. Carson, R. T. Owens, S. Danthuluri, T. Bice, M. Höök, S. V. L. Narayana, *Struct. Fold. Des.* **2000**, *8*, 67 – 78.
[11] Y. Terao, S. Kawabata, M. Nakata, I. Nakagawa, S. Hamada, *J. Biol. Chem.* **2002**, *277*, 47428 – 47435.
[12] U. Schwarz-Linek, J. M. Werner, A. R. Pickford, S. Gurusiddappa, J. H. Kim, E. S. Pilka, J. A. G. Briggs, T. S. Gough, M. Höök, I. D. Campbell, J. R. Potts, *Nature* **2003**, *423*, 177 – 181.
[13] M. Oke, L. G. Carter, K. A. Johnson, H. Liu, S. A. McMahon, X. Yan, M. Kerou, N. D. Weikart, N. Kadi, M. A. Sheikh, S. Schmelz, M. Dorward, M. Zawadzki, C. Cozens, H. Falconer, H. Powers, I. M. Overton, C. A. van Niekerk, X. Peng, P. Patel, R. A. Garrett, D. Prangishvili, C. H. Botting, P. J. Coote, D. T. Dryden, G. J. Barton, U. Schwarz-Linek, G. L. Challis, G. L. Taylor, M. F. White, J. H. Naismith, *J. Struct. Funct. Genomics* **2010**, *11*, 167 – 180.
[14] During revision of this manuscript, another Asp–Lys isopeptide bond was reported: J. A. Pointon, W. D. Smith, G. Saalbach, A. Crow, M. A. Kehoe, M. J. Banfield, *J. Biol. Chem.* **2010**, DOI: 10.1074/jbc.M110.149385.
[15] H. J. Kang, E. N. Baker, *J. Biol. Chem.* **2009**, *284*, 20729 – 20737.
[16] H. M. Senn, W. Thiel, *Angew. Chem.* **2009**, *121*, 1220 – 1254; *Angew. Chem. Int. Ed.* **2009**, *48*, 1198 – 1229.
[17] T. K. Harris, G. J. Turner, *IUBMB Life* **2002**, *53*, 85 – 98.
[18] T. M. Schmeing, K. S. Huang, D. E. Kitchen, S. A. Strobel, T. A. Steitz, *Mol. Cell* **2005**, *20*, 437 – 448.
[19] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277 – 293.
[20] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, P. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins Struct. Funct. Genet.* **2005**, *59*, 687 – 696.
[21] R. D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman, *Nucleic Acids Res.* **2010**, *38*, D211 – 222.
[22] L. A. Kelley, M. J. E. Sternberg, *Nat. Protoc.* **2009**, *4*, 363 – 371.
[23] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, A. K. Dunker, *Proteins Struct. Funct. Genet.* **2005**, *61*, 176 – 182.
[24] L. E. Kay, D. A. Torchia, A. Bax, *Biochemistry* **1989**, *28*, 8972 – 8979.

