Amyloids

International Edition: DOI: 10.1002/anie.201506772
German Edition: DOI: 10.1002/ange.201506772

Untangling a Repetitive Amyloid Sequence: Correlating Biofilm-Derived and Segmentally Labeled Curli Fimbriae by Solid-State NMR Spectroscopy

*Tobias Schubeis, Puwei Yuan, Mumdooh Ahmed, Madhu Nagaraj, Barth-Jan van Rossum, and Christiane Ritter**

Abstract: *Curli are functional bacterial amyloids produced by an intricate biogenesis machinery. Insights into their folding and regulation can advance our understanding of amyloidogenesis. However, gaining detailed structural information of amyloids, and their tendency for structural polymorphisms, remains challenging. Herein we compare high-quality solid-state NMR spectra from biofilm-derived and recombinantly produced curli and provide evidence that they adopt a similar, well-defined β-solenoid arrangement. Curli subunits consist of five sequence repeats, resulting in severe spectral overlap. Using segmental isotope labeling, we obtained the unambiguous sequence-specific resonance assignments and secondary structure of one repeat, and demonstrate that all repeats are most likely structurally equivalent.*

Amyloids have long been considered the result of protein misfolding, and are tightly associated with neurodegenerative diseases, such as Alzheimer's or Parkinson's disease.[1,2] However, many more proteins can adopt this quaternary structure in vivo or in vitro, and amyloid formation is utilized by organisms from all domains of life to pursue diverse biological tasks.[3,4] These functional amyloids are not a result of uncontrolled protein misfolding, and therefore provide naturally occurring model systems suitable for studying amyloid structure, propagation, and control.[5]

The morphology and biophysical characteristics of all amyloid fibrils are remarkably similar, although their primary structures differ strongly. Owing to their fibrillar nature, high-resolution structural information of amyloids has only become available recently through the application of solid-state NMR (ssNMR) spectroscopy.[6–8] Highest quality sample preparation is essential for a successful structure determination by ssNMR. Amyloids are typically fibrillized under in vitro conditions, which can result in heterogeneity or polymorphisms. For disease-associated amyloids, such polymorphisms occur frequently in vitro but also in vivo and are believed to form the basis for prion-like strain phenomena.[9,10] Functional amyloids, on the other hand, consist of an evolutionary optimized sequence. This is believed to favor a single molecular structure under physiological conditions, although polymorphisms could be generated in vitro.[11]

Among functional amyloids, the biogenesis of curli is one of the best understood.[12,13] Curli are found in the extracellular matrix of Enterobacteriaceae like *Escherichia coli* and function in biofilm formation, host adhesion, and invasion. The major curli subunit, CsgA, consists of a 22-residue flexible N-terminal export signal,[14] followed by five imperfect repetitive units (R1–R5) that form the amyloid core (Figure 1a). The glutamine and asparagine sidechains are suggested to form hydrogen bonds defining a stable strand–loop–strand structure with parallel β-sheets perpendicular to the fibril axis (Figure 1b).[15] A more recent structural model was computed using contact information from multiple sequence alignments as structural restraints.[16] Furthermore, CsgA has been studied in vitro by various biophysical techniques.[17] Native curli have been extracted from bacterial biofilms to obtain 1D ^{13}C ssNMR data at natural abundance.[18] ssNMR studies of recombinant CsgA fibrils pointed towards a β-solenoid structure, but the reported spectra exhibited rather broad, poorly resolved resonances, impeding unambiguous assignment and detailed structural investigation.[19]

Herein, we report a strategy to make this highly repetitive protein amenable to a detailed analysis by high-resolution ssNMR, while at the same time retaining the correlation to the functional fold. To this end, we extracted uniformly ^{13}C,^{15}N-labeled native curli from biofilms produced by the *E. coli* strain MC4100 grown on modified M9 agar plates. Fibrils were purified to homogeneity by a modification of established procedures (Figure S1 in the Supporting Information).[20] This approach allowed us to record ssNMR spectra of isotope-labeled amyloid directly isolated from its native environment, which can thus be expected to have retained its native structural properties.

However, yields of natively purified proteins are usually low, and recombinant material is typically required for advanced isotope-labeling strategies. Recombinant CsgA forms fibrils under a broad range of conditions.[21] To yield samples with the highest degree of homogeneity required for

[*] Dr. T. Schubeis, P. Yuan, Dr. M. Ahmed, Dr. M. Nagaraj, Dr. C. Ritter
Laboratory of Macromolecular Interactions
Helmholtz-Zentrum für Infektionsforschung
Inhoffenstrasse 7, 38124 Braunschweig (Germany)
E-mail: christiane.ritter@helmholtz-hzi.de
Dr. M. Nagaraj, Dr. B.-J. van Rossum
NMR-Supported Structural Biology
Leibniz-Institut für Molekulare Pharmakologie (FMP)
Robert-Rössle-Strasse 10, 13125 Berlin (Germany)
Dr. M. Ahmed
Department of Physics, Faculty of Science, Suez University
Suez, 43533 (Egypt)
Supporting information and ORCID(s) from the author(s) for this article are available on the WWW under http://dx.doi.org/10.1002/anie.201506772.

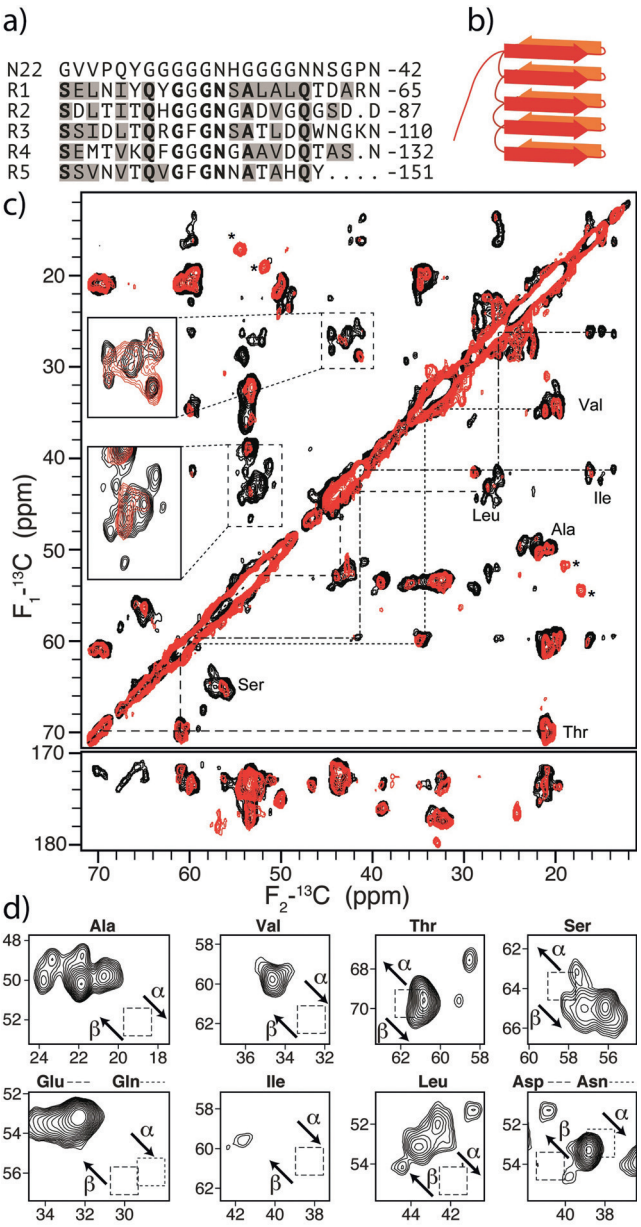

Figure 1. Native and recombinant curli. a) Primary structure of *E. coli* CsgA. Bold letters represent conserved residues over the different repeats, gray-highlighted letters are residues that are potentially adopting β-strand conformation. b) Model of the predicted strand–loop–strand motif. c) 2D ^{13}C–^{13}C spectra of native curli (red) and recombinant CsgA fibrils (black) recorded at 21 T with DARR mixing of 50 ms, insets are expanded regions with slightly different processing to emphasize weak signals, non-protein peaks are marked with an asterisk. d) Cα-Cβ regions of the ^{13}C–^{13}C correlation of recombinant CsgA for several amino acids. The dashed squares indicate random coil regions, and the arrows indicate expected trends for the cross-peaks owing to α-helices and β-strands.[22]

ssNMR spectroscopy, it was essential to use completely monomeric CsgA, devoid of preformed oligomers, for in vitro fibrillization, and to avoid any drying of the sample. Electron micrographs of our preparations are shown in Figure S2.

To compare sample quality and to identify spin systems both in native and recombinant curli, we used short mixing-time 2D-^{13}C-^{13}C DARR spectra (Figure 1c), resulting mostly in intra-residue correlations. Strikingly, the spectra show very little dispersion, especially in the Cα region where most cross peaks are congested within a 3–4 ppm range. Even though all types of amino acids can be identified according to their characteristic chemical shift patterns, almost no isolated signals can be distinguished. Although the signal intensities were significantly weaker in the native sample because of lower yields, the majority of signals can be identified in both spectra (see inserts in Figure 1c and Figure S3). Exceptions are few a non-protein peaks in the native spectra that are attributable to a lipid contamination. In addition, a few leucine and isoleucine side-chain resonances and some sequential contacts measured for recombinant CsgA are absent in the spectrum of the native sample, likely a result of the lower signal intensity. Importantly, however, those signals present in both spectra overlap very well. This situation suggests that recombinant CsgA adopts essentially the same structure as native curli.

Closer examination of the isoleucine and the alanine Cβ cross peaks permits the identification of all three isoleucine and up to eight (out of nine) of the alanine spin systems (Figure S4). The few isolated peaks are remarkably sharp with intrinsic ^{13}C line widths as narrow as 0.35 ppm, which compares well with line widths reported for the HET-s prion domain[23] and is indicative of high sample homogeneity. In addition, these residues are well spread within the protein sequence (see Figure 1a). Together, these findings suggest that CsgA adopts a well-defined conformation.

The Cα and Cβ chemical shifts are a good monitor for secondary structure. Even though individual residues cannot be assigned owing to spectral overlap, the shifts are congested in areas indicative of a predominant β-sheet topology as expected for amyloid fibrils (Figure 1d). Amino acid types which have chemical shifts corresponding to all the detected Cα–Cβ cross peaks define two potential β-strand core regions that are conserved among all repeats (Figure 1a). A more detailed analysis of the secondary structure requires a sequence-specific resonance assignment. Considering that the amyloid-forming core of CsgA consists of about 110 residues, the spectra are remarkably empty. This situation could be explained by both the sequence repetition and in addition a structural repetitive arrangement of the protein, where each sequence repeat faces a similar chemical environment in the regular β-strand stacking. The overlap can to some extent be reduced through nitrogen-edited 3D correlations (data not shown), but not sufficiently to attempt a complete chemical shift assignment.

We introduced segmental isotope labeling to obtain unambiguous resonance assignments of the first repeat (R1) within the context of full-length CsgA. We recently presented an efficient way to apply intein-mediated protein trans-splicing to amyloid fibrils in vitro.[24] A complementary pair of fusion proteins containing residues 21–69 or 71–151 of CsgA linked to the N- or C-terminal domain of the split intein *Npu*DnaE, respectively, was generated (Figure 2a). Thr70 was replaced with the catalytic cysteine of the intein and no

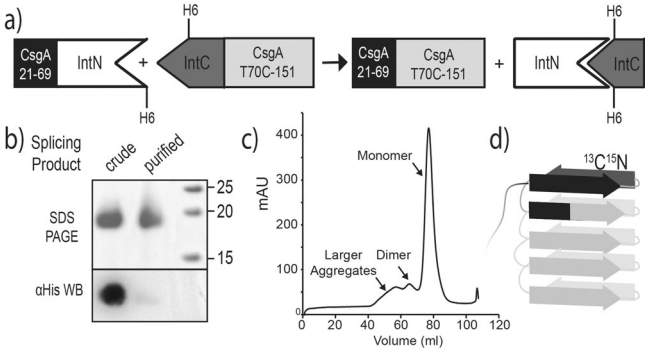

Figure 2. Preparation of segmentally labeled CsgA fibrils. a) CsgA—intein fusion proteins and their splicing reaction. The isotope-labeled fragment is shown in black. b) SDS-PAGE and anti His_6-tag western blot of crude and purified segmentally labeled CsgA. c) Superdex 200 size-exclusion chromatogram of segmentally labeled CsgA. d) Amyloid model with indication of the labeled segment (black).

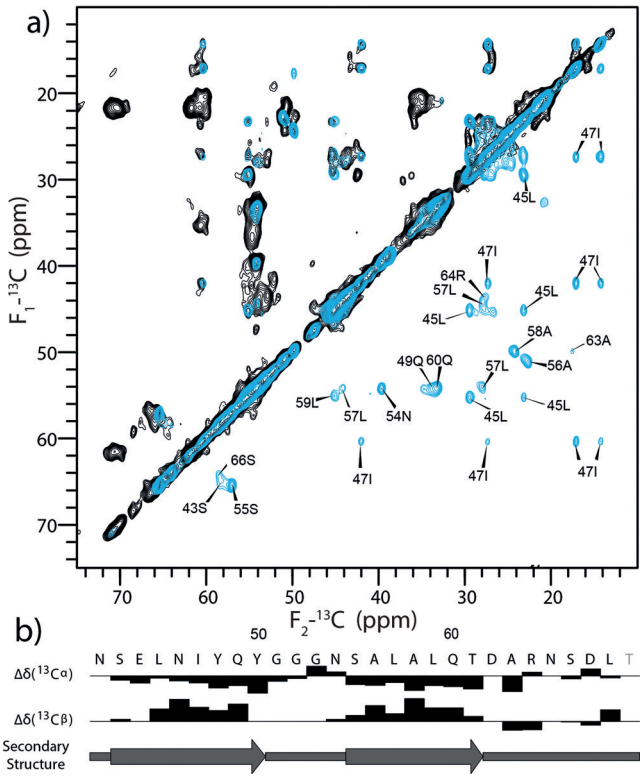

Figure 3. Resonance assignment of the structural repeat R1 of CsgA. a) ^{13}C–^{13}C solid–state NMR spectra of uniformly (black) and segmentally (cyan) labeled CsgA Fibrils. Cross peaks of the uniformly labeled sample are only depicted above the diagonal, resonance assignments are annotated below the diagonal. The spectra were recorded at 21 T with DARR mixing of 50 ms. b) Chemical shift deviation of random coil values and predicted secondary structure of the CsgA R1 repeat.

further linker residues were introduced at the splicing junction. Mixing of the precursor proteins under reducing conditions resulted in segmentally labeled CsgA that precipitated as amyloid fibrils during the splicing reaction (Figure S5). We checked the purity of this precipitate by SDS-PAGE and anti-His_6 western blotting, as the educts and the product migrated exactly at the same size. This revealed a strong co-precipitation of His_6-tagged educts (Figure 2b), making further purification steps necessary. Removal of His_6-tagged contaminants and a final size exclusion chromatography (Figure 2c) under denaturing conditions resulted in pure monomeric CsgA (Figure 2b), and homogenous fibrils. The labeling scheme is illustrated in Figure 2d. The spectra of segmentally and uniformly labeled CsgA are in excellent agreement (Figure 3a). All spin systems present in the isotope-labeled segment align perfectly with resonances of fully labeled CsgA, which strongly indicates that the Thr70Cys mutation did not alter the fold. The overall cross-peak intensity appears slightly improved, and the intrinsic line widths remained equally narrow (ca. 0.35 ppm). The spectral crowding is dramatically reduced, for instance only a single isoleucine and three alanine residues are labeled and detectable.

The overall similarity between the spectra of R1-labeled and fully labeled CsgA strongly supports the notion that the individual sequence repeats also form structurally repetitive units, thus explaining the high degree of signal overlap observed in the fully labeled spectra.

Residue-specific resonance assignments were obtained using 2D and 3D NCACX and NCOCX spectra. Except for Asp62 and Asn65, every residue in the segment 43–68 could be unambiguously assigned (Figure 3 and S6). The detection of a single set of well-resolved NMR resonances confirms the presence of a unique molecular conformation. A chemical shift analysis using CSI predicts the presence of two β-strands in the R1 repeat, spanning from Ser43 to Tyr50 and Ser55 to Thr61 (Figure 3b).[25] The short rigid glycine-loop between these two strands and the apparent spectroscopic equivalence of the sequence repeats, make a β-solenoid arrangement plausible. To gather further evidence for this, we recorded a 2D DARR spectrum with a long mixing time of 300 ms (Figure S7a). Segmental labeling within amyloid fibrils predominantly results in intra-molecular polarization transfer.[24] Apart from intra-residue correlations, around 170 new intra-molecular cross peaks were picked in the aliphatic region of the spectrum. The vast majority of these were assigned to sequential connectivities or medium-range transfers and thus served as an excellent verification of the sequential assignment.

Although a single labeled repeat is not sufficient for a full structure calculation, we observed several medium-range and two long-range restraints defining the loop region from Glu49 to Ser55 (Figure S7a). These contacts define a tight turn between the two β-strands (Figure S7b).[26]

Residues 21–42 were not detectable in the cross polarization based spectra and are therefore probably flexible. Indeed, in a ^{1}H–^{13}C correlation excited by INEPT we could identify resonances accounting for all residue types present in this segment (Figure 4). Interestingly, we could also identify resonances corresponding to threonine, aspartate, and alanine, which are absent in the stretch 21–42, but rather are located in the loop region connecting R1 and R2. Although these residues are also detectable by cross polarization, this result suggests that this loop is potentially more flexible than the loop connecting the two β-strands within R1. Interest-

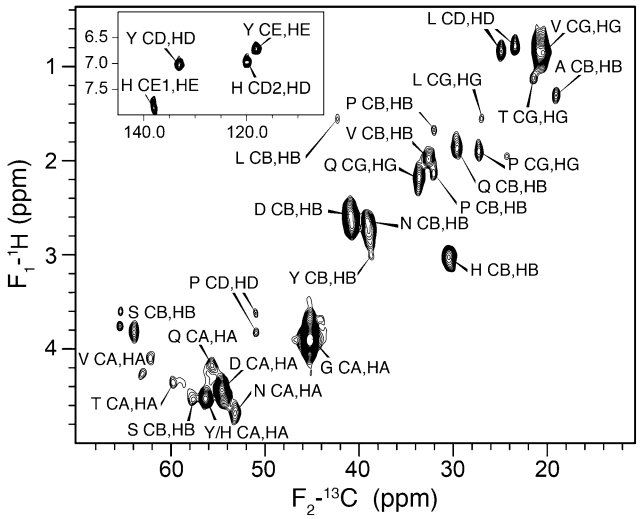

Figure 4. INEPT-detected residues in segmentally labeled CsgA fibrils, with amino-acid type assignments for both aliphatic (main figure) and aromatic (inset) regions.

ingly, we additionally observed two leucine side chain resonances. Assuming that the sequence repeats form structural repeats in which the conserved glutamine and asparagine residues lie on top of each other, it is plausible that the two leucine side chains are solvent accessible and might therefore be sufficiently flexible to allow detection in an INEPT spectrum. It is noteworthy in this respect that some leucine side chains were missing in the spectra of native curli.

In summary, we have provided ssNMR-based evidence for the structure of the functional amyloid curli. To our knowledge, we show herein for the first time high-resolution ssNMR spectra of a functional amyloid directly isolated from its native environment. The spectra of curli extracted from uniformly labeled biofilm and of recombinantly produced CsgA have a high degree of similarity, strongly supporting the integrity of the in vitro preparations. We employed segmental labeling of the single R1 repeat to obtain residue-specific assignments and to provide evidence for a structurally repetitive arrangement of the CsgA sequence repeats, which presumably form a well-defined β-solenoid structure. This study clearly demonstrates the prospects of segmental isotope labelling for the structure determination of proteins with sequential or structural repeats. Combining segmental labeling of different combinations of repeats with other labeling schemes—such as sparse labeling—should allow the full structure calculation of CsgA in curli fibrils by ssNMR.

Acknowledgements

We are grateful to Manfred Rohde for carrying out the EM analysis of curli, to Thorsten Lührs (SeNostic GmbH) and Johannes Spehr (HZI Braunschweig) for critical discussions. T.S. has been supported by a fellowship of the HZI graduate school, M.A. by a DFG grant RI 2210/1-1 and M.N. by a DFG grant RO 3496/3-1.

Keywords: functional amyloid · intein · isotopic labeling · protein structures · solid-state NMR spectroscopy

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 14669–14672
Angew. Chem. **2015**, *127*, 14880–14884

[1] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
[2] C. Ross, M. Poirier, *Nat. Med.* **2004**, *10 Suppl*, S10–7.
[3] T. P. J. Knowles, M. Vendruscolo, C. M. Dobson, *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 384–396.
[4] J. Greenwald, R. Riek, *Structure* **2010**, *18*, 1244–1260.
[5] X. Wang, M. R. Chapman, *Prion* **2008**, *2*, 57–60.
[6] R. Tycko, R. B. Wickner, *Acc. Chem. Res.* **2013**, *46*, 1487–1496.
[7] G. Comellas, C. M. Rienstra, *Annu. Rev. Biophys.* **2013**, *42*, 515–536.
[8] A. K. Schütz, T. Vagt, M. Huber, O. Y. Ovchinnikova, R. Cadalbert, J. Wall, P. Güntert, A. Böckmann, R. Glockshuber, B. H. Meier, *Angew. Chem. Int. Ed.* **2015**, *54*, 331–335; *Angew. Chem.* **2015**, *127*, 337–342.
[9] B. H. Meier, A. Böckmann, *Curr. Opin. Struct. Biol.* **2014**, *30C*, 43–49.
[10] S. B. Prusiner, *Annu. Rev. Genet.* **2013**, *47*, 601–623.
[11] C. Wasmer, A. Soragni, R. Sabaté, A. Lange, R. Riek, B. H. Meier, *Angew. Chem. Int. Ed.* **2008**, *47*, 5839–5841; *Angew. Chem.* **2008**, *120*, 5923–5925.
[12] M. R. Chapman, L. S. Robinson, J. S. Pinkner, R. Roth, J. Heuser, M. Hammar, S. Normark, S. J. Hultgren, *Science* **2002**, *295*, 851–855.
[13] M. L. Evans, M. R. Chapman, *Biochim. Biophys. Acta Mol. Cell Res.* **2014**, *1843*, 1551–1558.
[14] L. S. Robinson, E. M. Ashman, S. J. Hultgren, M. R. Chapman, *Mol. Microbiol.* **2006**, *59*, 870–881.
[15] S. K. Collinson, J. M. Parker, R. S. Hodges, W. W. Kay, *J. Mol. Biol.* **1999**, *290*, 741–756.
[16] P. Tian, W. Boomsma, Y. Wang, D. E. Otzen, M. H. Jensen, K. Lindorff-Larsen, *J. Am. Chem. Soc.* **2015**, *137*, 22–25.
[17] X. Wang, D. R. Smith, J. W. Jones, M. R. Chapman, *J. Biol. Chem.* **2007**, *282*, 3713–3719.
[18] O. McCrate, X. Zhou, C. Reichhardt, L. Cegelski, *J. Mol. Biol.* **2013**, *425*, 4286–4294.
[19] F. Shewmaker, R. P. McGlinchey, K. R. Thurber, P. McPhie, F. Dyda, R. Tycko, R. B. Wickner, *J. Biol. Chem.* **2009**, *284*, 25065–25076.
[20] S. K. Collinson, L. Emödy, K. H. Müller, T. J. Trust, W. W. Kay, *J. Bacteriol.* **1991**, *173*, 4773–4781.
[21] M. S. Dueholm, S. B. Nielsen, K. L. Hein, P. Nissen, M. Chapman, G. Christiansen, P. H. Nielsen, D. E. Otzen, *Biochemistry* **2011**, *50*, 8281–8290.
[22] Y. Wang, O. Jardetzky, *Protein Sci.* **2002**, *11*, 852–861.
[23] A. B. Siemer, C. Ritter, M. Ernst, R. Riek, B. H. Meier, *Angew. Chem. Int. Ed.* **2005**, *44*, 2441–2444; *Angew. Chem.* **2005**, *117*, 2494–2497.
[24] T. Schubeis, T. Lührs, C. Ritter, *ChemBioChem* **2015**, *16*, 51–54.
[25] D. S. Wishart, *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 62–87.
[26] A. V Kajava, A. C. Steven, *Fibrous Proteins: Amyloids, Prions and Beta Proteins*, Elsevier, Amsterdam, **2006**.

Received: July 22, 2015
Published online: October 16, 2015

