J Biomol NMR (2014) 59:185–197
DOI 10.1007/s10858-014-9838-y

ARTICLE

Solid state NMR chemical shift assignment and conformational analysis of a cellulose binding protein facilitated by optimized glycerol enrichment

Hadar Ivanir · Amir Goldbourt

Received: 28 March 2014 / Accepted: 4 May 2014 / Published online: 14 May 2014

Abstract Magic-angle spinning solid-state NMR has been applied to study CBM3b–Cbh9A (CBM3b), a cellulose binding module protein belonging to family 3b. It is a 146-residue protein having a unique nine-stranded β-sandwich fold, in which 35 % of the structure is in a β-sheet conformation and the remainder of the protein is composed of loops and unstructured regions. Yet, the protein can be crystalized and it forms elongated needles. Close to complete chemical shift assignment of the protein was obtained by combining two- and three-dimensional experiments using a fully labeled sample and a glycerol-labeled sample. The use of an optimized protocol for glycerol-based sparse labeling reduces sample preparation costs and facilitates the assignment of the large number of aromatic signals in this protein. Conformational analysis shows good correlation between the NMR-predicted secondary structure and the reported X-ray crystal structure, in particular in the structured regions. Residues which show high B-factor values are situated mainly in unstructured regions, and are missing in our spectra indicating conformational flexibility rather than heterogeneity. Interestingly, long-range contacts, which could be clearly detected for tyrosine residues, could not be observed for aromatic phenylalanine residues pointing into the hydrophobic core, suggesting possible high ring mobility. These studies will allow us to further investigate the cellulose-bound form of CBM proteins.

Keywords Solid-state NMR · Magic-angle spinning · Cellulose-binding protein · Resonance assignment · Isotopic labeling · [2-^{13}C] glycerol

Introduction

Cellulose, the main structural component of plant cells and the most abundant carbohydrate polymer in nature, cannot be easily degraded. Anaerobic microorganisms have evolved a wide variety of enzymes, which enable the degradation of plants cell wall polysaccharides (Bayer et al. 2004). These enzymes can be found in the free state or in a large extracellular enzyme complex called cellulosome. The cellulosome is composed of two main protein groups: catalytic modules of glycoside hydrolases (GHs), which catalyze the hydrolysis of the glycosidic linkage between carbohydrate units, and different non-catalytic modules that serve many different roles (Bayer et al. 2004). Carbohydrate-binding module (CBM) is a large family of non-catalytic proteins that serve as scaffoldin-borne of the cellulosome multi-enzyme. The role of CBM is to recognize and bind specifically to carbohydrates and in that manner concentrate the enzymes onto the substrate for efficient hydrolysis of polysaccharides in general and cellulose in particular. In recent years there is increasing interest related to the mechanism of cellulosome degradation of biomass, since simple sugars suitable for fermentation to biofuels are released in this process (Demain et al. 2005). As CBM plays an important role in locating the crystalline surface of cellulose and therefore being one of the rate-limiting factors in the degradation process, various spectroscopic methods have been used for biological and structural studies of CBM proteins. Examples are Defocused Orientation and Position Imaging (DOPI) (Dagel

Electronic supplementary material The online version of this article (doi:10.1007/s10858-014-9838-y) contains supplementary material, which is available to authorized users.

H. Ivanir · A. Goldbourt (✉)
Raymond and Beverly Sackler Faculty of Exact Sciences, School of Chemistry, Ramat Aviv, 69978 Tel Aviv, Israel
e-mail: amirgo@post.tau.ac.il

et al. 2011), quantum dots fluorescence (Liu et al. 2009), X-ray crystallography (Szabo et al. 2001; Lammerts van Bueren and Boraston 2004; Boraston et al. 2006) and solution NMR (Simpson et al. 2002; Santos et al. 2012). X-ray and solution NMR provided 3D structures of various CBM proteins, some of them in complex with polysaccharides of several sugar rings or sugar derivatives.

To date there are 66 known families of CBM proteins, which are classified according to amino-acid sequence homology. The family 3 of CBM proteins have a typical fold of nine-stranded β-sandwich with jelly-roll topology (Yaniv et al. 2011). Those proteins possess a planar array of aromatic amino-acids on their surface that form stacking interactions with the glucose rings of crystalline cellulose chains resulting in strong van-der Waals interactions, thus stabilizing the structure. In addition, the side chains of polar amino acid residues may form hydrogen bonds with the sugar ligand, helping to stabilize the interaction. Since the cellulose complex cannot be crystallized or dissolved, magic-angle spinning (MAS) NMR could be utilized in order to obtain detailed structural characterization of such systems.

Magic-angle spinning solid-state NMR has proved its great potential as a structural elucidation technique, as a tool for probing protein dynamics and for characterizing specific interactions between biomolecules (Böckmann 2006; Reif 2012; Weingarth and Baldus 2013). Initially, structures of a few small (<80 a.a) microcrystalline proteins such as SH3 (Castellani et al. 2002), GB1 (Zhou et al. 2007), kaliotoxin (Lange et al. 2005) and ubiquitin (Zech et al. 2005) were determined by MAS NMR. Further developments in sample preparation, sparse labeling schemes, multidimensional pulse-sequences, proton detection techniques (Huber et al. 2011; Ward et al. 2011; Asami and Reif 2013) and the utilization of paramagnetic reagents (Bertini et al. 2010b; Jaroniec 2012; Sengupta et al. 2012) enabled progress towards structure determination of microcrystalline proteins with higher molecular weight (Bertini et al. 2010a) or complexity (Yan et al. 2013), protein aggregates and amyloids (Schütz et al. 2011), membrane proteins (Sharma et al. 2010; Verardi et al. 2011; Ahuja et al. 2013) and macromolecular complexes (Loquet et al. 2012). Also, site-specific resonance assignments of an increasing amount of proteins with over 100 amino-acids were deposited in the Biological Magnetic Resonance Bank (BMRB) (Ulrich et al. 2008), although 70 % of all MAS NMR based assigned proteins are shorter than 100 amino acids (Fig. S1 in the Supplementary Material). Nonetheless, the number of structures of large proteins is relatively small (Balayssac and Bertini 2008; Jehle et al. 2010; Shahid et al. 2012; Knight et al. 2012; Park et al. 2012; Yan et al. 2013), and this number is even smaller for MAS NMR-based structures that can be compared to existing crystallographic structures.

The assignment of small proteins can be safely obtained from uniformly labeled samples using a set of complementary multi-dimensional experiments such as NCACX/NCOCX/CONCX. For larger proteins, extensive assignments become difficult as a result of spectral congestion and resonance overlap. It has been shown before that several selective ^{13}C labeling techniques can facilitate the assignment of solid proteins since they reduce spectral congestion and eliminate scalar couplings thus narrowing the spectral lines. Examples are the use of [2-^{13}C]-glycerol and [1,3-^{13}C]-glycerol (LeMaster and Kushlan 1996) to obtain 'checkboard' labeling; [1-^{13}C] glucose and [2-^{13}C] glucose (Hong 1999; Lundström et al. 2007) that provide mainly the assignments of methyl, Cα, and Cβ carbons; obtaining a carbonyl-methyl labeling pattern using either pyruvate (Tugarinov and Kay 2005); induction of the Entner Doudoroff (ED) pathway in *E. coli* (Refaeli and Goldbourt 2012) or acetate labeling of pseudomonas (Eddy et al. 2013). Such bacteria use the ED pathway for glucose catabolism (Entner and Doudoroff 1952; Goldbourt et al. 2007). The commonly reported method for ^{13}C sparse labeling with glycerol is based on the protocol published by Lemaster et al., in which protein overexpression is applied in a minimal medium containing the partially labeled glycerol precursors.

In this manuscript we present an extensive assignment of the 146 amino-acids protein CBM3b from cellobiohydrolase 9A (Cbh9A) by means of MAS solid-state NMR by utilizing common multi-dimensional experiments applied on fully enriched samples, and on a sparse labeled sample, which was produced using an optimized glycerol-based preparation protocol. This protocol combines the approach of LeMaster for glycerol-based enrichment with that of Marley et al. (2001) for the production of high yields of sparsely labeled proteins using minimal amounts of selectively labeled precursors. The structure of the protein has lately been determined by means of X-ray crystallography (Yaniv et al. 2012) and shows a particularly high content of unstructured regions (∼65 % of the sequence), which is quite unique among microcrystalline proteins that have been studied thus far by MAS NMR.

Materials and methods

Protein expression, purification, and sample preparation

A pET-28a(+) plasmid bearing the wild-type CBM3b protein, containing resistance to kanamycin antibiotics and an N-terminal purification His-tag, was obtained from the group of Prof. Frolow in Tel Aviv University. Protein expression, purification and crystallization were based on known procedures (Yaniv et al. 2011, 2012) modified to

suite the solid-state NMR sample preparation requirements. The plasmid was transformed into *Escherichia coli* BL21(DE3) competent cells (Novagene). The cultures were grown in 1 liter of LB rich media containing 50 μg/ml kanamycin until the optical cell density at 600 nm (OD600) reached ~0.8, and then harvested by centrifugation. $^{13}C/^{15}N$ uniformly labeled samples were obtained by suspending the pelleted cells in 250 ml of fully labeled minimal media (M9 stock solution (Marley et al. 2001), BME vitamins (sigma), metal trace (Cai et al. 1998), 2 mM $MgSO_4$, 50 μg/ml kanamycin, 4 g/l $^{13}C_6$ D-glucose and 1 g/l $^{15}NH_4Cl$). Consequently, the pellets were four-fold concentrated with regard to the LB rich-media. ^{13}C sparse labeling was obtained by suspending the pellets in 167 ml minimal media (six-fold concentration) containing 8 g/l [2-^{13}C]-glycerol and 8 g/l $NaH^{13}CO_3$. This particular procedure for glycerol growth optimization is further discussed in the text. After incubation at 37 °C for 1 h at 225 rpm, 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added and the protein was induced. After 13h, the cells were harvested and lysed using a microfluidizer (Microfluidics™, M-110L).

CBM3b was first isolated by metal affinity chromatography using TALON™ resin charged with cobalt ions (Clonthech) according to the manufacturer's instructions and then purified by fast protein liquid chromatography (FPLC) using Superdex 75 16/60 column (GE Healthcare) operating on an ÄKTA prime system (GE Healthcare). The purified protein was concentrated to a final concentration of 27–35 mg/ml using a Millipore centrifugal filter (3,000 MWCO). Crystallization by the sitting drop method was performed using Crystal PEG/ION screen condition 48 (20 % PEG 3350, 0.2 M diammonium hydrogen citrate) from Hampton Research screens as was reported to be the preferred crystallization condition for obtaining single-crystals (Yaniv et al. 2012). Needle crystals grew in 96 sitting drop wells containing 10 μl protein and 10 μl crystallization buffer (condition #48). After 3 days at constant temperature of 18 °C, primary long needle crystals were observed using an optical microscope. Maximal amount of crystals was obtained after 2 weeks. The crystals were collected from the sitting drops by a pipette. The BCA assay of the supernatant revealed that ~50 % of the protein was crystallized. The needle-like crystals were transferred into a 4 mm ZrO_2 rotor by series of centrifugations.

NMR data was acquired using three samples; two fully labeled samples, belonging to two different preparation batches with yields of 40 mg (sample 1) and 50 mg (sample 2) per liter of starting culture, respectively, which provided consistent chemical shifts, and a sparsely labeled sample, obtained with a yield of ~40 mg per liter of starting culture.

MAS SSNMR spectroscopy

Experiments were carried out on a Bruker Avance-III spectrometer operating at a magnetic field of 14.09 T (^1H Larmor frequency of 600.0 MHz) equipped with a Bruker BioSpin's triple-resonance 4 mm E-free probe operating at ^1H, ^{13}C and ^{15}N frequencies. The experiments were conducted on a ~20 mg protein sample at a temperature of 10–15 °C depending on the particular spinning rate (calibrated by $^{207}PbNO_3$). ^1H–$^{13}C/^{15}N$ magnetization transfer was obtained using a linear 10 % ^1H-ramped cross-polarization (CP) (Schaefer and Stejskal 1976; Metz et al. 1994). During evolution and acquisition times swept-frequency two-pulse phase modulation (swf-TPPM) (Chandran et al. 2008) ^1H decoupling of ~80 kHz was applied. A homonuclear proton driven spin diffusion (PDSD) (Szeverenyi et al. 1982) two-dimensional (2D) ^{15}N–^{15}N correlation experiment was carried out using a mixing time of 2 s. Three data sets of ^{13}C–^{13}C correlation spectra were acquired for the uniformly labeled sample using dipolar-assisted rotational resonance (RAD/DARR) (Takegoshi et al. 2001; Morcombe et al. 2004) with mixing times of 15, 100 and 250 ms. Additional data sets were acquired on a [2-^{13}C] glycerol sample using mixing times of 15 ms and 200 ms. Heteronuclear 2D N-CX spectra were obtained using double cross-polarization (DCP) (Baldus et al. 1998) and 3D NCACX/NCOCX spectra were acquired by using DCP followed by DARR mixing. Power levels (ν_1) of ~50 kHz were used for the ^{13}C and ^{15}N channels during CP and hard pulses. During DCP the conditions were $\nu_1^N = 2.5\nu_R, \nu_1^C = 1.5\nu_R$ or $\nu_1^N = 3.5\nu_R, \nu_1^C = 2.5\nu_R$. The difference in transfer efficiency is further discussed in the text. The spinning rate during the acquisition of DARR spectra was set to 13.5 kHz and for all other experiments 11 kHz, which gave better ^{13}C–^{15}N magnetization transfer. Full tabulated experimental parameters can be found in the Supplementary Material. The data were processed using Topspin versions 2.0 and 3.2 (Bruker, Karlsruhe, Germany) and NMRPipe (Delaglio et al. 1995) software and analyzed using Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco).

Results and discussion

Site-specific sequential assignment

The majority of ^{13}C and ^{15}N site-specific assignments were carried out on a ^{13}C, ^{15}N-uniformly labeled sample using 2D homonuclear ^{13}C–^{13}C and ^{15}N–^{15}N correlations, 2D ^{13}C–^{15}N correlations, and 3D heteronuclear NCACX/NCOCX (Pauli et al. 2001) correlation experiments. Three-dimensional experiments were imperative for the assignment since CBM3b is a large protein (146

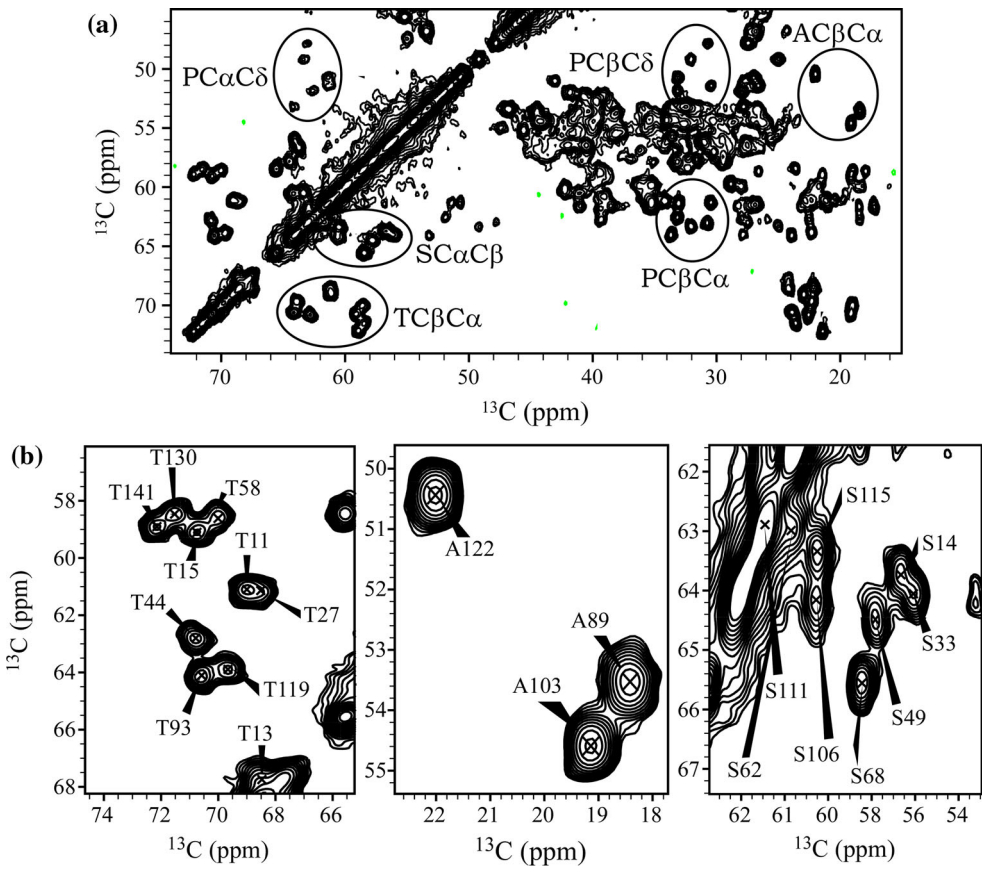

Fig. 1 **a** A ^{13}C–^{13}C DARR15 (15 ms mixing time) spectrum exhibits the well resolved Ala, Thr, Ser and Pro amino-acids (*circled* in the main spectrum). **b** Expansion of the Cα–Cβ region of Thr, Ala and Ser in the DARR spectrum and their site-specific assignment. As shown in Table S2 of the Supplementary Material, all threonines and alanines in the sequence are unambiguously assigned, as well as eight out of nine serines (S72 is not observed in any one of the spectra probably due to high mobility)

residues + additional 20 residues belonging to a histidine purification tag), and 2D spectra had a large degree of spectral congestion, in particular in the backbone region. Consequently, 3D experiments provided sequential-walks through adjacent residues, short mixing time DARR experiments were used mainly for the assignment of side-chains, and long-mixing time DARR experiments validated and completed many of these assignments. Spectral congestion and ambiguities, still existent in these experiments due to the size of the protein, were then resolved using a sparsely labeled sample grown using 2-^{13}C glycerol.

Sequential-walk assignment starting points are residues such as Ser, Ala, Thr, Gly and Pro, which have unique and well-isolated ^{13}C chemical-shifts (their 2D ^{13}C peak positions in CBM3b are shown in Fig. 1) and are known as 'anchors'. In addition, residues that appear in small numbers in the amino-acid sequence (in this case, Trp, Cys, Met) can also be identified uniquely. As shown in Table S2, CBM3b contains only three alanine residues (Ala89, Ala103, Ala122), hence their identification was straightforward, and enabled the sequential walk assignment of the consecutive residues using the NCACX/NCOCX spectra. Figure 2 shows as an example a strip plot connecting A89-E95, which contains also the anchor residues glycine and threonine (another stretch starting with Ala103 appears in Fig. S3 of the Supplementary Material). Similarly, other stretches of amino acids were identified using Ser, Thr, Pro and Gly as starting points. Due to the unique ring structure of the proline residue, in many cases it is very hard to detect in 3D experiments; however, with some modification of experimental parameters we managed to observe site-specifically all six prolines in CBM3b, as discussed below.

CBM3b contains only a single tryptophan residue (W139), hence its chemical shifts could be easily obtained from the 2D DARR spectrum through its unique Cγ shift (111.9 ppm) and its cross peaks with other ring carbon signals such as Cδ2 (129.4 ppm), Cζ3 (123.2 ppm), etc., which are well isolated from all other aromatic residues. The assignment of Cys8, one of two assigned cysteine residues in CBM3b, is shown in the K3-E9 strip plot appearing in the Fig. S3, and the single Met is undetected since it is located in a highly mobile region in the N-terminal part, adjacent to the purification tag.

For large proteins such as CBM3b, the probability of backbone peak overlap is high also in 3D spectra. Therefore, in many cases sidechain cross-peaks in 3D spectra must be used in order to resolve overlapping N–Cα–Cβ signals. In order to observe such longer-range signals, polarization transfers during the experiments must be optimized

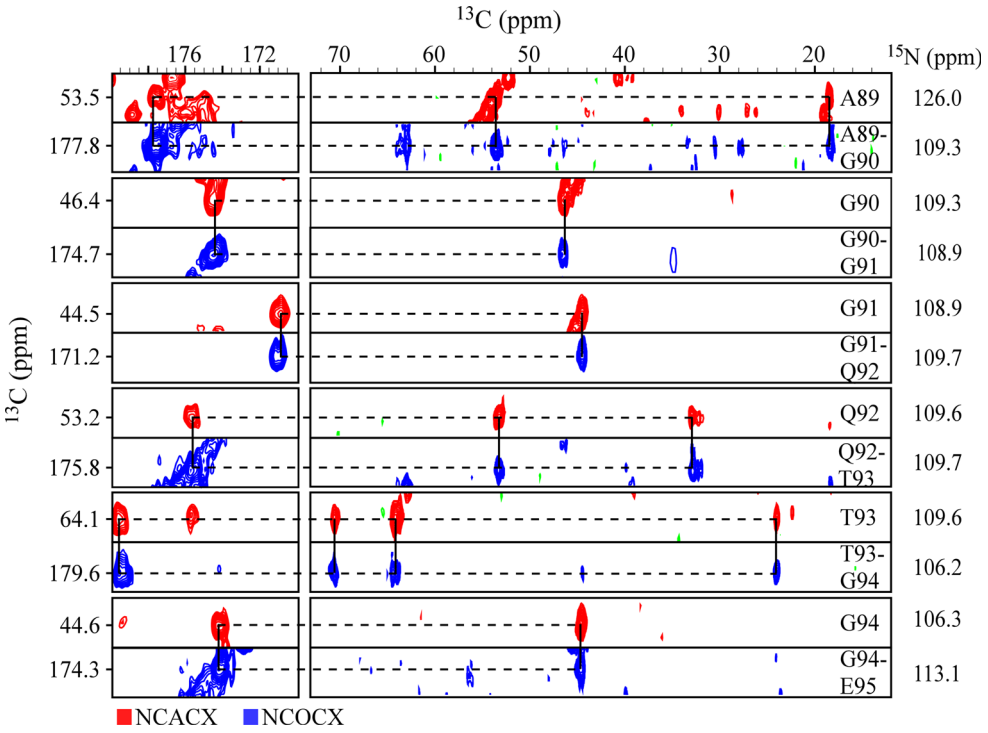

Fig. 2 Strip plots from NCACX (*red*) and NCOCX (*blue*) 3D correlation experiments showing the sequential-walk along residues 89–95. The Cα and C′ shifts from NCACX and NCOCX strips, respectively, are indicated on the *left* and ^{15}N shifts are shown on the *right* of each strip. The *dashed horizontal lines* link peaks within the strip while the *vertical solid lines* link the C'_i–N_i, $C\alpha_i$–N_i, CX_i–N_i correlation peaks in NCACX spectrum with C'_i–N_{i+1}, $C\alpha_i$–N_{i+1}, CX_i–N_{i+1} in NCOCX spectrum, respectively

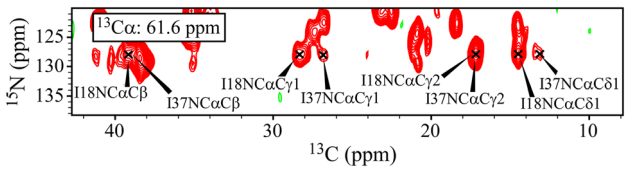

Fig. 3 An NCACX strip cut at a particular Cα value (61.6 ppm), showing the sidechain-based assignment of I18 and I37 in the 3D spectrum. In this example both isoleucines have identical N, Cα and Cβ chemical shifts and the carbonyl region exhibits large degree of congestion. Only due to the presence of the Cγ1 and Cδ1 cross-peaks, which exhibit different chemical shift values for those two residues, the assignment could be carried out. The site-specific assignment of Cδ1 of those Isoleucines was obtained using the NCOCX experiment (not shown)

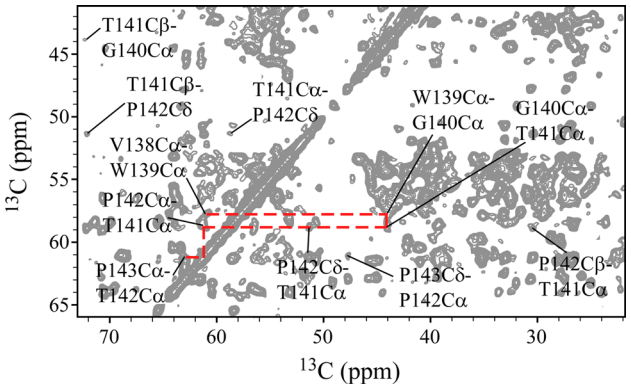

Fig. 4 Validation of site-specific assignment of V138-P143 is shown on a DARR100 spectrum. The *dashed line* links the Cα carbons of those residues. In addition, contacts between the sidechains of those residues further reinforce the assignment

(optimized N–C transfer conditions are discussed below). Figure 3 demonstrates a case, in which the N–Cα–Cβ cross-peaks of I18 and I37 overlap; however, site resolution was obtained due to the presence of Cγ1 and Cδ1 cross-peaks in the 3D spectrum, which have well resolved chemical shift values. Since those sidechain shifts were resolved in NCOCX spectra as well, they could unambiguously report on the correct assignment. A similar strategy was used in order to resolve several pairs of amino-acids, for which the Cα and Cβ chemical shifts are in the same range, e.g., Glu and Gln, Tyr and Phe, and Asp and Asn as illustrated in Fig. S4 of the Supplementary Material.

Additional validation of the site-specific assignments relied on the acquisition of DARR spectra with longer mixing times (100 and 250 ms) and on $^{15}N_i$–$^{15}N_{i+1}$ correlations of adjacent residues using PDSD with a mixing time of 2 s. Examples for $C\alpha_i$–$C\alpha_{i+1}$ correlations are shown in Fig. 4, and NN correlations, appearing in Fig. S5 in the Supplementary Material (D113-I120 is indicated explicitly), show many additional sequential walks, which were further validated in many instances by the detection of $^{15}N_i$–$^{15}N_{i+2}$ correlations.

Observation and assignment of proline residues

As mentioned above, site-specific assignment of prolines is important for the sequential-walk as they constitute one of

the anchor amino-acids due to their unique ^{15}N chemical-shift (~140 ppm), which is well isolated from the rest of the backbone ^{15}N resonances (~105–125 ppm). Nonetheless, prolines are difficult to observe as a result of three main reasons; (1) the proline amide group lacks a directly bonded proton due to its ring structure making the ^{1}H–^{15}N magnetization transfer less efficient (Shahid et al. 2012); (2) polarization is transferred to two directly-bonded carbons, Cα and Cδ; (3) in 3D NCC experiments, during DCP, the ^{15}N channel is typically irradiated with low RF power at a frequency of 120 ppm. This allows efficient selective polarization transfer to neighboring carbons from the majority of the backbone nitrogens; however, the off-resonance shifts of prolines at ~140 ppm (and to some extent that of glycines ~100–105 ppm) reduce their transfer efficiency. We found that the sensitivity of proline excitation in the NCA spectrum (and of other residues as well) is improved (by a factor of ~1.5) when the power is increased without compromising the selectivity towards the aliphatic region. Indeed, more sidechain carbons are excited (Cβ, Cγ etc.), an effect which produces some peak aliasing in a 3D experiment; however, no carbonyl signals are generated and peak aliasing can be directed to desired regions (Goldbourt et al. 2007). In our hands, the use of a Hartman-Hahn matching condition with $\nu_1^N = 3.5\nu_R, \nu_1^C = 2.5\nu_R$ was superior to $\nu_1^N = 2.5\nu_R, \nu_1^C = 1.5\nu_R$ not only for NCO transfer (Shi and Ladizhansky 2012; Loening et al. 2012) but also for NCA transfer. This improvement in sensitivity is demonstrated in Fig. 5a, where an overlay of two NCA spectra, acquired at identical sample and offset conditions, is shown; a sensitive spectrum with $\nu_1^N = 3.5\nu_R, \nu_1^C = 2.5\nu_R$ (black) and a less efficient spectrum with $2.5\nu_R, \nu_1^C = 1.5\nu_R$ (red). Combining the data obtained from the improved 2D NCA (Fig. 5a) spectrum with DARR (Fig. 5b) and 3D NCACX/NCOCX enabled the site specific assignment of all six prolines, as shown in Fig. 5b. The improved sensitivity in a 3D NCOCX experiment is also clearly demonstrated in Fig. 5c, where the assignment of the three amino acids A122–F123–Q124 is obtained in a single strip of the improved DCP condition due to the existence of both N_{i+1}–C'_i–CX_{i+1} and N_{i+1}–C'_i–CX_{i-1} correlations in addition to the common N_{i+1}–C_i–CX_i.

Assignments and constraints with an optimized glycerol-based sample preparation

The commonly used method for ^{13}C sparse labeling is based on the protocol published by Bracken and co. (Marley et al. 2001) for high yields of uniform labeled proteins using glucose as the sole ^{13}C precursor. According to this procedure, cell mass is produced using unlabeled rich media followed by exchange into a minimal media with a four-fold reduction in its volume. When we applied this procedure to CBM3b, the protein yield was four-fold smaller with glycerol than the yield obtained with glucose. Following a series of optimizations, we found that 8 g/L glycerol instead of the commonly used concentration of 2 g/L improved the yield significantly; however, this procedure requires significant amounts of expensive glycerol. In order to increase the yield while using minimal amounts of labeled glycerol, optimizations of cell density in natural abundance minimal media were carried out and then applied to the labeled material.

Transformed competent BL21/DE3 *E. coli* cells were grown in 1 l of LB rich medium until the optical density reached ~0.8. Four different volumes from the same culture (100, 200, 300 and 400 ml culture) were pelleted by centrifugation, washed, and resuspended in a final volume of 50 ml unlabeled minimal glycerol media. Consequently, the pellets were two-fold, four-fold, six-fold and eight-fold concentrated respectively with regard to the LB rich-media. After an incubation period of 1 h with rapid shaking (220 rpm, 37 °C) the protein was induced (1 mM IPTG) overnight at similar conditions. The cell cultures containing the expressed protein were pelleted by centrifugation and then lysed. Protein yields, determined by the BCA method from cell lysates of the natural abundance glycerol minimal media growths are shown in Fig. 6. The highest normalized total yield was obtained for a six-fold concentrated media.

We prepared from a 1 l rich culture, a [2-^{13}C] glycerol (and ^{15}N) labeled sample using the optimized protocol (total minimal media 167 ml, 1.3 g [2-^{13}C] glycerol, a similar amount of ^{13}C-labeled sodium bicarbonate), and obtained a yield of 40 mg of clean CBM3b protein, similar to the yield normally obtained with glucose. Using this sample, we performed 2D and NCC 3D experiments, and obtained improved resolution and sensitivity following the elimination of half of the signals, of the J-couplings and of most of the strong homonuclear dipolar couplings. In particular, improved signals were obtained for the aromatic residues (eleven tyrosines and nine phenylalanines in CBM3b), which have reduced sensitivity due to strong homonuclear and scalar couplings in the fully labeled sample. For example, in the 3D spectra of the fully labeled sample very few aromatic peaks were detected (one of the exceptions, F123, is shown in Fig. 5c), and since the residues have very close Cα and Cβ chemical shift values, specific assignments of these amino acids were difficult to obtain. However, as shown in Fig. 7, the sparse sample produced many highly-resolved aromatic signals in both 2D and 3D spectra, allowing their unequivocal assignment.

At DARR experiments acquired with long mixing times, as shown in Fig. 8, aromatic residues exhibit broad featureless lines for the fully labeled sample, but highly

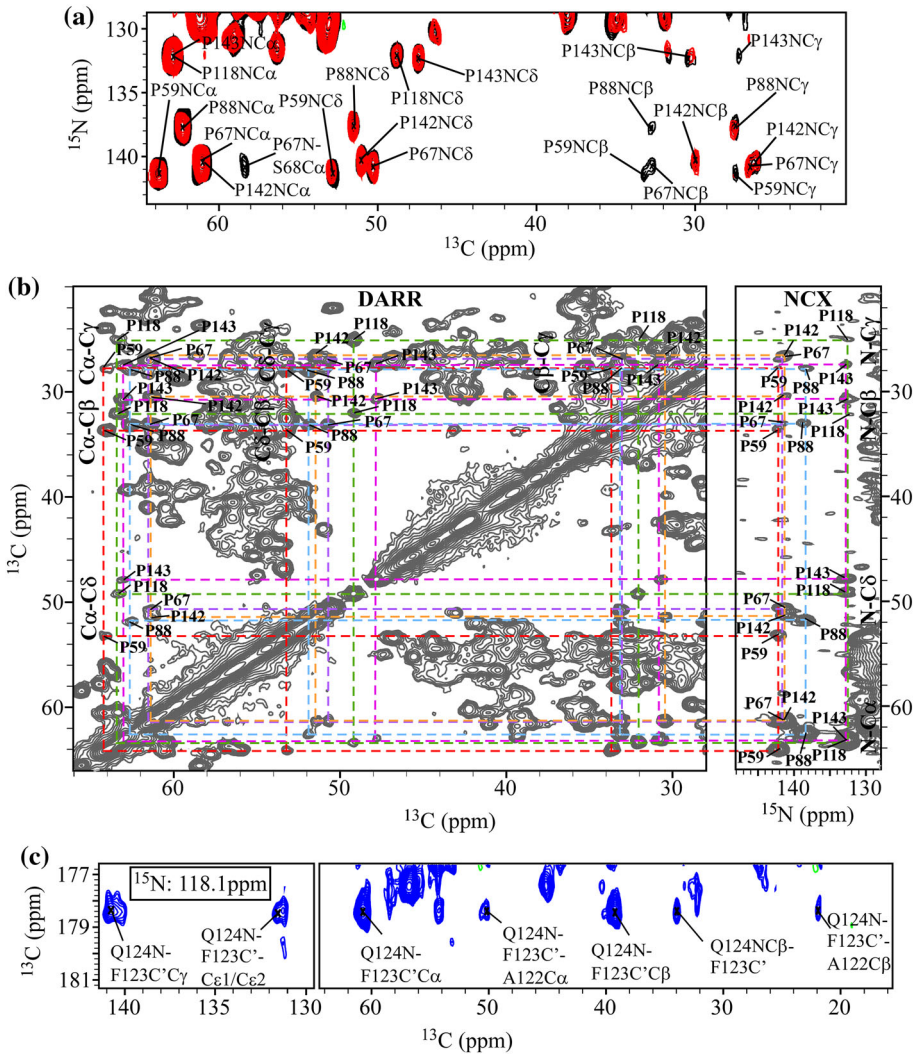

Fig. 5 **a** An overlay of the prolines region of two NCX spectra obtained with different matching conditions but similar irradiation offsets, $\delta(^{15}N) = 120$ ppm, $\delta(^{13}C) = 50$ ppm. The black spectrum obtained using $\nu_1^N = 3.5\nu_R, \nu_1^C = 2.5\nu_R$ exhibits an improved sensitivity in general and particularly in the prolines region (by a factor of 1.5), in comparison with the red spectrum, which was obtained with $\nu_1^N = 2.5\nu_R, \nu_1^C = 1.5\nu_R$. The spectra are plotted using 15 contour levels with the lowest contour drawn at a signal-to-noise ratio of 5.3. **b** The amino acid patterns of all six prolines, indicated by *colored dashed lines*, are shown by combining ^{13}C–^{13}C DARR (*left*) and NCX (*right*) spectra. **c** Strip plots from 3D NCOCX correlation experiments obtained using an N–CO polarization transfer condition of $\nu_1^N = 3.5\nu_R, \nu_1^C = 2.5\nu_R$. This condition enables to obtain not only the regular two-residue walk N_{i+1}–C'_i–CX_i but also N_{i+1}–C'_i–CX_{i+1} and the three-residue-walk N_{i+1}–C'_i–CX_{i-1}. Due to the strong polarization transfer, two additional peaks are observed in the aromatic region, which is otherwise only observed in a sparsely labeled sample, as discussed below

resolved signals in the sparse sample, allowing assignment validation, and the acquisition of long-range contacts and distance restraints.

Long-range correlations

It is well established that sparse labeling provides samples that are highly efficient for obtaining long-range contacts due to the reduction in the number of labeled carbons. In the case of CBM3b, long-range aromatic cross-peaks, which are otherwise missing from the uniformly labeled sample, were readily obtained. For example, the cross-peak at 47.8–131.4 ppm, which appears only in the DARR200 spectrum from the sparse sample shown in Fig. 8, implies on an inter-residue correlation between the Cδ of Pro143 and a tyrosine. From sequential assignment, this tyrosine must be assigned to Tyr42. Looking at the X-ray structure, Cγ of Tyr42 and Cδ of Pro143 have an inter-nuclear distance of 3.57 Å, in agreement with our assignment. Another contact with the Cε carbon of Tyr42 validates this observation.

Interestingly, the long-range correlation experiment shown in Fig. 8 indicates that while the tyrosine amino acids

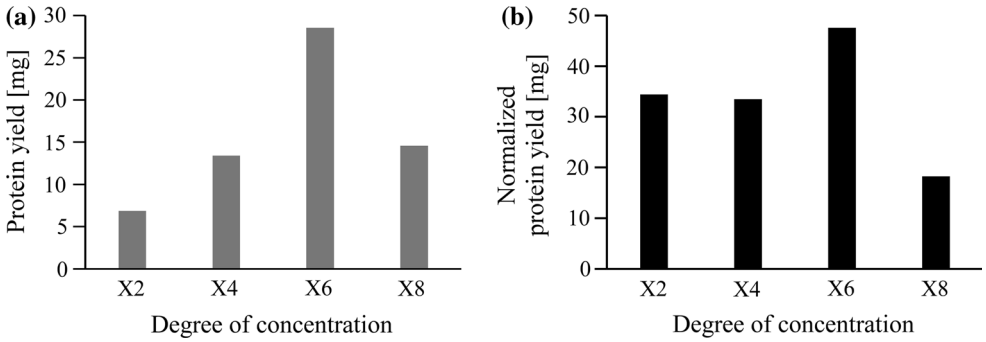

Fig. 6 Protein yields as a function of cell concentrations. **a** Total yields obtained for the 100, 200, 300 and 400 ml rich media cell culture concentrated by X2, X4, X6 and X8 respectively to 50 ml minimal glycerol media. **b** Normalized yield, per 1 l of initial culture. The best yield in both total and normalized quantities was obtained with the X6 concentration

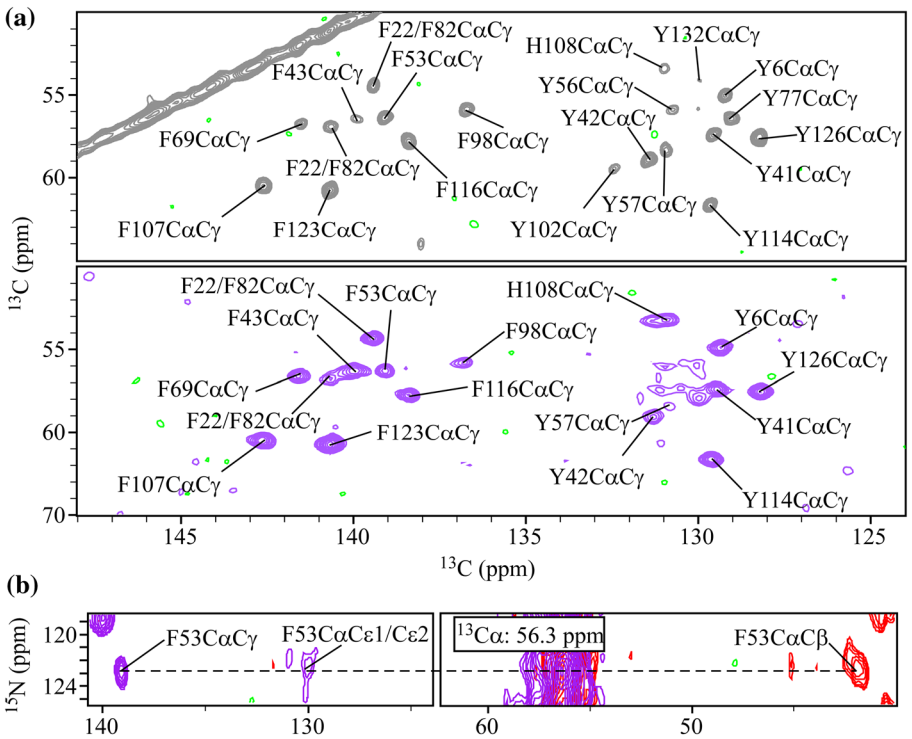

Fig. 7 Experimental spectra from a sparsely enriched CBM3b. *Top* **a** the assignment of 7/9 Phe and 9/11 Tyr residues shown through Cα-Cγ cross-peaks observed in DARR15. *Bottom* **a** 2D ^{13}C-^{13}C projection obtained from the 3D NCACX exhibits many of the Phe and Tyr cross-peaks shown in the DARR15 spectrum allowing their site-specific assignment. Two ambiguous Phe residues, unlabeled in the spectra above, belong to F22 and F82 and are not site-specifically assigned. One ambiguous weak Tyr signal with Cα-Cγ correlation peak at 56.4–129.0 ppm and one undetected signal belong to Y32 and Y77, which reside in or on the edge of highly mobile regions (Fig. 9 *below*). **b** Assignment of F53 aided by the acquisition of 3D NCACX data on a sparse labeled sample. The overlay of NCACX spectra from a fully labeled sample (*red*) and a 2-glycerol sample (*purple*) shows the existence of aromatic signals *(left)* in the sparse sample that are missing in the fully labeled sample

show a significant number of long-range contacts, phenylalanines (with the exception of F123) lack such long-range cross-peaks. We find a similar behavior, and even more pronounced, in the DARR250 spectrum of the fully labeled sample where the entire region between 20 and 30 ppm shows a broad signal. This region clearly indicates many inter-residue cross-peaks between Tyr and other residues. Yet, no such signals exist for phenylalanine Cγ signals. According to the X-ray structure, shown in Fig. 8c, most of the phenylalanines point into the hydrophobic core of the protein and only F107 and F123 are exposed to the solvent. It is likely that tyrosine amino acids are more rigid due to the creation of hydrogen bonds through their hydroxyl group, thus experiencing strong homonuclear dipolar couplings,

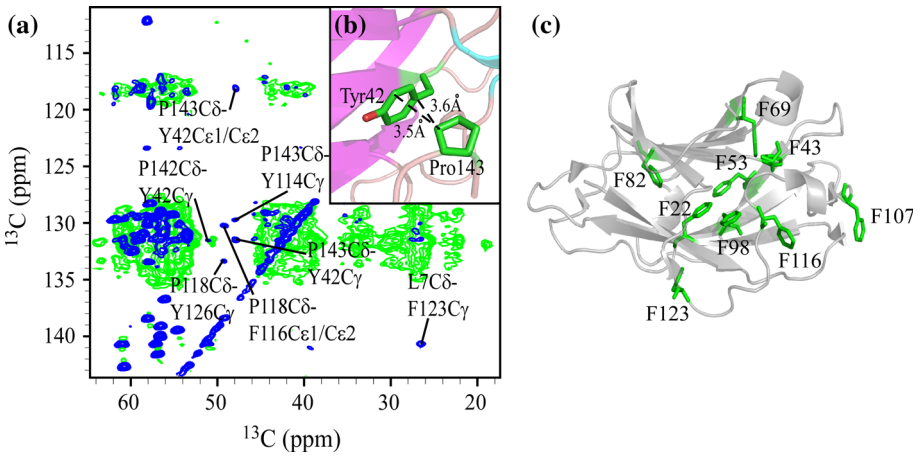

Fig. 8 **a** An overlay of the aromatic region extracted from long-mixing-time DARR spectra of a uniformly labeled sample (250 ms, *green*) and a [2-^{13}C] glycerol sample (200 ms, *blue*). The improved resolution in the sparsely labeled spectrum enables the assignment of the aromatic residues. For clarity, only Y42 and Y114 are indicated out of nine site-specifically assigned tyrosine residues. Several Pro-Tyr long-range cross-peaks, which could be detected only in the sparsely labeled sample, are indicated. **b** The proximity of P143 and Y42 is shown within the X-ray structure of CBM3b. The distance reported between P143Cδ and Y42Cγ is 3.6 Å, and between P143Cδ and to Y42Cε1 is 3.5 Å. **c** The nine phenylalanines are shown in the X-ray structure of CBM3b. With the exception of F107 and F123, which are exposed to the solvent, all seven phenylanines face into the hydrophobic core. The only long-range cross-peak involving Phe, shown in (**a**), is between F123Cγ and L7Cδ

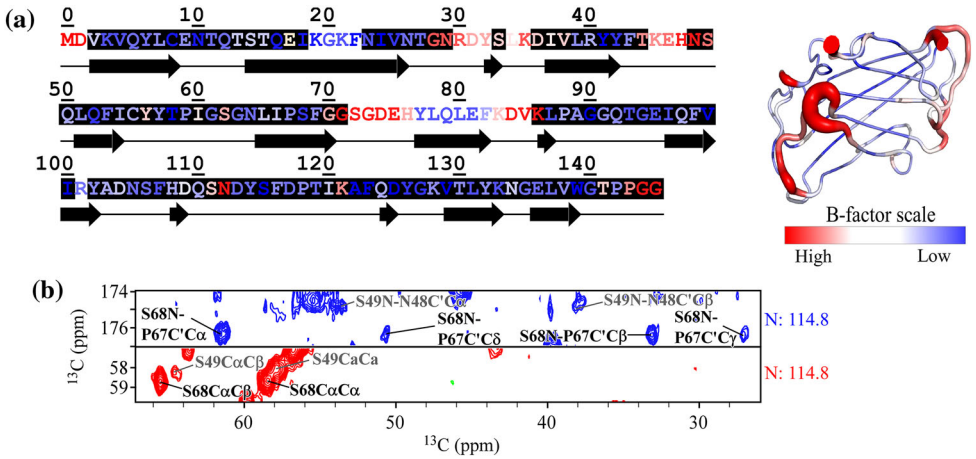

Fig. 9 X-ray B-factors and NMR assignments of CBM3b. **a** The amino-acids sequence of the protein (without the his-tag) and its secondary structure according to PDB model 2YLK (*left*) and the folded protein model (*right*), drawn using PyMOL (www.pymol.org), are colored according to the reported B-factor values. Highly dynamic regions are colored in *red* and rigid regions are colored *in blue*. Assigned residues are highlighted in *black*. **b** An example for the use of B-factor values in the assignment. The NCOCX strip (*top*, *blue*) exhibits the assignment of N48 and P67, both of which precede serine residues. Although both serines have a similar ^{15}N shift, the more intense signal is associated with the lower B-factor reported for S68

while phenylalanines form weak hydrophobic interactions and are more dynamic on this time scale. These observations will further be investigated and may point to some functional role of these residues.

X-ray B-factor values and protein assignments

X-ray B-factor values are related to dynamics or small heterogeneities (Halle 2002). Therefore, increased magnitudes of B-factor values are related to reduced signal intensities in the NMR spectra (Habenstein et al. 2011). Consequently, many weak or missing signals in the spectra of CBM3b, as shown in Fig. 9a, are associated also with large B-factor values. An interesting exception are the glycine residues; we observed all 15 glycine signals in the 2D NCA spectrum as shown in Fig. S6, including G73, G144 and G145, which are reported to have high mobility according to the X-ray B-factor values. Perhaps the lack of

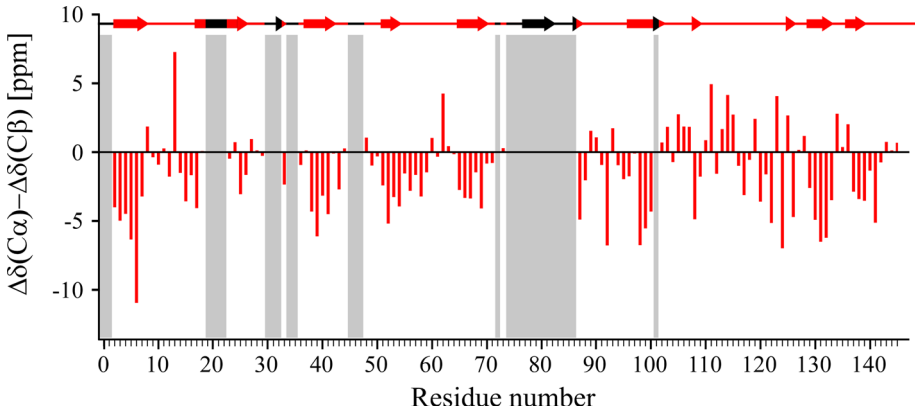

Fig. 10 Secondary chemical shifts of CBM3b. The secondary structure elements according to the crystal structure are shown on *top* with ssNMR-assigned regions colored in *red*. *Horizontal gray bars* indicated unassigned or undetected regions. Negative Δδ(Ca) − Δδ(Cb) values, which imply on β-sheet regions, are in excellent agreement with the X-ray data. Comparison of the secondary chemical shifts with the crystal structure shows an agreement of 91 % in the β-sheet regions

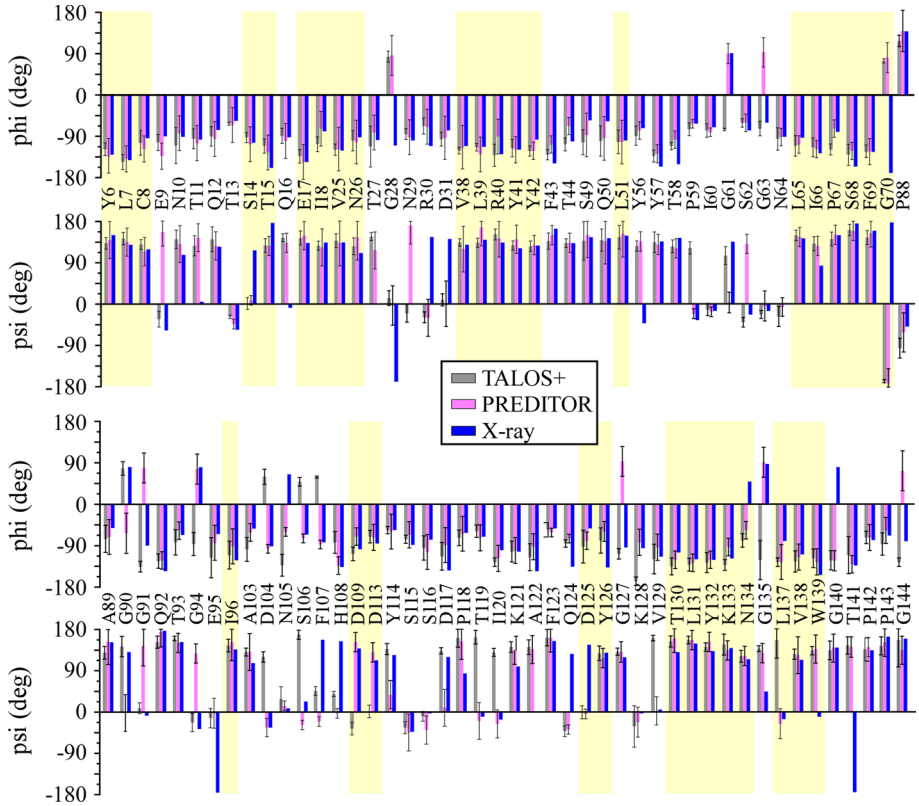

Fig. 11 A comparison of ψ and ϕ predicted by TALOS+ (*gray*) and PREDITOR (*pink*) and calculated from X-ray (*blue*) of 101 of the NMR-assigned residues. β-sheet regions are indicated by the *yellow-shaded* areas

sidechain atoms reduces to some extent the loss of signal due to spin diffusion, and the two methylene protons further aid in polarization transfer between an amide nitrogen and its attached alpha-carbon (the so-called PAIN-CP effect (Lewandowski et al. 2007)).

In some cases the use of B-factor values can be helpful in the assignment procedure. Figure 9b shows assignment strip plots for residues P67-S68. In the NCACX strip corresponding to an ^{15}N shift of 114.8 ppm, there are two optional assignments for Ser68; however, only one of them has an intense Cβ peak, which implies that one of these serines is more mobile than the other. Since the B-factor value of Ser68 is relatively low it is assigned as shown while the weak signal is assigned to Ser49, which has a larger B-factor in the X-ray structure.

All chemical shifts were deposited in the BMRB under accession number 19831, and are shown in Table S3 of the Supplementary material.

The secondary structure of CBM3B

The conformation of CBM3b was obtained by calculating the chemical shift differences between the deviations of Cα and Cβ from random-coil values (Wang and Jardetzky 2002), and by backbone torsion-angles prediction using both TALOS+ (Shen et al. 2009) and PREDITOR (Berjanskii et al. 2006). The plot of $\Delta\Delta\alpha\beta = \Delta\delta(C\alpha) - \Delta\delta(C\beta)$ (Bousset et al. 2001) in Fig. 10 shows good agreement between the negative $\Delta\Delta\alpha\beta$ values and the β-sheet structure regions, while other regions have alternating positive and negative values and indicate loops or unstructured residues. As shown in Fig. 11, of the 146 amino acids in the sequence (excluding the flexible N-terminal His-tag), 101 torsion angle ϕ/ψ pairs were predicted by TALOS+ and PREDITOR. For TALOS+, 80 residues have good classification and for PREDITOR 90 residues were considered, which have confidence values ≥0.6. When compared to the published X-ray crystal structure, 87.5 %/70 % of the TALOS+/PREDITOR-predicted angles are in good agreement ($|\phi_{predicted} - \phi_{\text{X-ray}}| \leq 50°$) with the structure. When considering only the structured regions (shaded regions in Fig. 10), which are normally better predicted by these software tools, these numbers are increased to 95 and 90 % for TALOS+ and PREDITOR, respectively. These results indicate, that while torsion angles predicted by NMR shifts are very reliable for the determination of secondary structure elements, unstructured regions should be considered with care. Moreover, structure calculation should rely mostly on the acquisition of distance constraints, in particular in unstructured regions.

Conclusion

We presented the extensive site-specific assignment of the crystalline form of the 146 amino-acids globular protein CBM3b, obtained from data acquired on fully labeled samples and a sparsely glycerol-based sample, which was obtained under an optimized growth protocol. Utilizing the two complementary 3D NCACX/NCOCX spectra assisted by 2D homo- and heteronuclear correlation experiments, we obtain site-specifically 80 % of the backbone atoms (N, C′, Cα), 78 % of Cβ carbons and 43 % of all other side chain resonances (carbon- and nitrogen side-chains). Overall, for 117 amino acids at least one assignment was obtained site-specifically, and additional nine amino acids were completely identified in a residue-specific manner as described in table S2 of the supplementary material. The completion of the assignment was mostly limited by regions that undergo motional averaging on the NMR timescale, and are associated with weak or missing signals, which are highly correlated to regions with high reported B-factor values. Exceptions are glycine residues that can easily be detected in all regions.

Structurally, we find that the torsion-angles predicted from our assignments are in good agreement with the reported crystal structure, in particular in the β-sheet regions. Interestingly, many long-range contacts can be observed for aromatic tyrosine residues, but not for phenylalanines located in the interior of the protein, suggesting possible flexibility of their rings. Our extensive assignment of this protein will facilitate investigation of its structural properties and of similar proteins containing only a small percentage of common secondary structural elements. In particular, it will allow us to study the cellulose binding properties of such carbohydrate binding proteins.

References

Ahuja S, Jahr N, Im S-C et al (2013) A model of the membrane-bound cytochrome b5-cytochrome P450 complex from NMR and mutagenesis data. J Biol Chem 288:22080–22095. doi:10.1074/jbc.M112.448225

Asami S, Reif B (2013) Proton-detected solid-state NMR spectroscopy at aliphatic sites: application to crystalline systems. Acc Chem Res 46:2089–2097. doi:10.1021/ar400063y

Balayssac S, Bertini I (2008) Paramagnetic shifts in solid-state NMR of proteins to elicit structural information. Proc Natl Acad Sci 105:17284–17289. doi:10.1073/pnas.0708460105

Baldus M, Petkova AT, Herzfeld J, Griffin RG (1998) Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. Mol Phys 95:1197–1207. doi:10.1080/002689798166215

Bayer EA, Belaich J-P, Shoham Y, Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58:521–554. doi:10.1146/annurev.micro.57.030502.091022

Berjanskii MV, Neal S, Wishart DS (2006) PREDITOR: a web server for predicting protein torsion angle restraints. Nucleic Acids Res 34:W63–W69. doi:10.1093/nar/gkl341

Bertini I, Bhaumik A, De Paëpe G et al (2010a) High-resolution solid-state NMR structure of a 17.6 kDa protein. J Am Chem Soc 132:1032–1040. doi:10.1021/ja906426p

Bertini I, Emsley L, Lelli M (2010b) Ultrafast MAS solid-state NMR permits extensive 13C and 1H detection in paramagnetic metalloproteins. J Am Chem Soc 132:5558–5559. doi:10.1021/ja100398q

Böckmann A (2006) Structural and dynamic studies of proteins by high-resolution solid-state NMR. Comptes Rendus Chim 9:381–392. doi:10.1016/j.crci.2005.06.008

Boraston AB, Healey M, Klassen J et al (2006) A structural and functional analysis of alpha-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. J Biol Chem 281:587–598. doi:10.1074/jbc.M509958200

Bousset L, Belrhali H, Janin J et al (2001) Structure of the globular region of the prion protein Ure2 from the yeast *Saccharomyces cerevisiae*. Structure 9:39–46. doi:10.1016/S0969-2126(00)00553-0

Cai M, Huang Y, Sakaguchi K et al (1998) An efficient and cost-effective isotope labeling protocol for proteins expressed in shape *Escherichia coli*. J Biomol NMR 11:97–102. doi:10.7171/jbt.12-2302-004

Castellani F, Van Rossum B, Diehl A (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. Nature 420:23–26. doi:10.1038/nature01070

Chandran CV, Madhu PK, Kurur ND, Brauniger T (2008) Swept-frequency two-pulse phase modulation (SWf-TPPM) sequences with linear sweep profile for heteronuclear decoupling in solid-state NMR. Magn Reson Chem 46:943–947. doi:10.1002/mrc.2285

Dagel DJ, Liu YS, Zhong LL et al (2011) In situ imaging of single carbohydrate-binding modules on cellulose microfibrils. J Phys Chem B 115:635–641. doi:10.1021/jp109798p

Delaglio F, Grzesiek S, Vuister GW et al (1995) NMRPIPE—a multidimensional spectral processing system based on unix pipes. J Biomol NMR 6:277–293. doi:10.1007/bf00197809

Demain AL, Newcomb M, Wu JHD (2005) Cellulase, clostridia, and ethanol. Microbiol Mol Biol Rev 69:124. doi:10.1128/mmbr.69.1.124-154.2005

Eddy MT, Belenky M, Sivertsen AC et al (2013) Selectively dispersed isotope labeling for protein structure determination by magic angle spinning NMR. J Biomol NMR 57:129–139. doi:10.1007/s10858-013-9773-3

Entner N, Doudoroff M (1952) Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. J Biol Chem 196:853–862

Goldbourt A, Day LA, McDermott AE (2007) Assignment of congested NMR spectra: carbonyl backbone enrichment via the Entner-Doudoroff pathway. J Magn Reson 189:157–165. doi:10.1016/j.jmr.2007.07.011

Habenstein B, Wasmer C, Bousset L et al (2011) Extensive de novo solid-state NMR assignments of the 33 kDa C-terminal domain of the Ure2 prion. J Biomol NMR 51:235–243. doi:10.1007/s10858-011-9530-4

Halle B (2002) Flexibility and packing in proteins. Proc Natl Acad Sci USA 99:1274–1279. doi:10.1073/pnas.032522499

Hong M (1999) Determination of multiple φ-torsion angles in proteins by selective and extensive 13 C labeling and two-dimensional solid-state NMR. J Magn Reson 401:389–401. doi:10.1006/jmre.1999.1805

Huber M, Hiller S, Schanda P et al (2011) A proton-detected 4D solid-state NMR experiment for protein structure determination. Chem Phys Chem 12:915–918. doi:10.1002/cphc.201100062

Jaroniec CP (2012) Solid-state nuclear magnetic resonance structural studies of proteins using paramagnetic probes. Solid State Nucl Magn Reson 43–44:1–13. doi:10.1016/j.ssnmr.2012.02.007

Jehle S, Rajagopal P, Bardiaux B et al (2010) Solid-state NMR and SAXS studies provide a structural basis for the activation of alphaB-crystallin oligomers. Nat Struct Mol Biol 17:1037–1042. doi:10.1038/nsmb.1891

Knight MJ, Pell AJ, Bertini I et al (2012) Structure and backbone dynamics of a microcrystalline metalloprotein by solid-state NMR. Proc Natl Acad Sci USA 109:11095–11100. doi:10.1073/pnas.1204515109

Lammerts van Bueren A, Boraston AB (2004) Binding sub-site dissection of a carbohydrate-binding module reveals the contribution of entropy to oligosaccharide recognition at "non-primary" binding subsites. J Mol Biol 340:869–879. doi:10.1016/j.jmb.2004.05.038

Lange A, Becker S, Seidel K et al (2005) A concept for rapid protein-structure determination by solid-state NMR spectroscopy. Angew Chem Int Ed Engl 44:2089–2092. doi:10.1002/anie.200462516

LeMaster D, Kushlan D (1996) Dynamical mapping of *E. coli* thioredoxin via 13C NMR relaxation analysis. J Am Chem Soc 7863:9255–9264. doi:10.1021/ja960877r

Lewandowski J, De Paepe G, Griffin R (2007) Proton assisted insensitive nuclei cross polarization. J Am Chem Soc 129:728–729. doi:10.1021/ja0650394

Liu YS, Zeng YN, Luo YH et al (2009) Does the cellulose-binding module move on the cellulose surface? Cellulose 16:587–597. doi:10.1007/s10570-009-9306-0

Loening NM, Bjerring M, Nielsen NC, Oschkinat H (2012) A comparison of NCO and NCA transfer methods for biological solid-state NMR spectroscopy. J Magn Reson 214:81–90. doi:10.1016/j.jmr.2011.10.012

Loquet A, Sgourakis NG, Gupta R et al (2012) Atomic model of the type III secretion system needle. Nature 486:276–282. doi:10.1038/nature11079

Lundström P, Teilum K, Carstensen T et al (2007) Fractional 13C enrichment of isolated carbons using [1-13C]- or [2- 13C]-glucose facilitates the accurate measurement of dynamics at backbone Calpha and side-chain methyl positions in proteins. J Biomol NMR 38:199–212. doi:10.1007/s10858-007-9158-6

Marley J, Lu M, Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. J Biomol NMR 20:71–75. doi:10.1023/A:1011254402785

Metz G, Wu XL, Smith SO (1994) Ramped-amplitude cross-polarization in magic-angle-spinning NMR. J Magn Reson A110:219–227. doi:10.1006/jmra.1994.1208

Morcombe CR, Gaponenko V, Byrd RA, Zilm KW (2004) Diluting abundant spins by isotope edited radio frequency field assisted diffusion. J Am Chem Soc 126:7196–7197. doi:10.1021/ja047919t

Park SH, Das BB, Casagrande F et al (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. Nature 491:779–783. doi:10.1038/nature11580

Pauli J, Baldus M, van Rossum B et al (2001) Backbone and side-chain C-13 and N-15 signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 tesla. ChemBioChem 2:272–281. doi:10.1002/1439-7633(20010401)2:4<272:aid-cbic272>3.0.co;2-2

Refaeli B, Goldbourt A (2012) Protein expression and isotopic enrichment based on induction of the Entner-Doudoroff pathway in *Escherichia coli*. Biochem Biophys Res Commun 427:154–158. doi:10.1016/j.bbrc.2012.09.031

Reif B (2012) Deuterated Peptides and Proteins: Structure and Dynamics Studies by MAS Solid-State NMR. 831:279–301. doi:10.1007/978-1-61779-480-3

Santos CR, Paiva JH, Sforca ML et al (2012) Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. Biochem J 441:95–104. doi:10.1042/bj20110869

Schaefer J, Stejskal EO (1976) Carbon-13 nuclear magnetic resonance of polymers spinning at the magic angle. J Am Chem Soc 98:1031–1032. doi:10.1021/ja00420a036

Schütz AK, Soragni A, Hornemann S et al (2011) The amyloid-Congo red interface at atomic resolution. Angew Chem Int Ed Engl 50:5956–5960. doi:10.1002/anie.201008276

Sengupta I, Nadaud PS, Helmus JJ et al (2012) Protein fold determined by paramagnetic magic-angle spinning solid-state NMR spectroscopy. Nat Chem 4:410–417. doi:10.1038/nchem.1299

Shahid SA, Markovic S, Linke D, van Rossum B-J (2012) Assignment and secondary structure of the YadA membrane protein by solid-state MAS NMR. Sci Rep 2:803. doi:10.1038/srep00803

Sharma M, Yi M, Dong H et al (2010) Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer. Science 330:509–512. doi:10.1126/science.1191750

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223. doi:10.1007/s10858-009-9333-z

Shi L, Ladizhansky V (2012) Intrinsically disordered protein. Analysis 895:153–165. doi:10.1007/978-1-61779-927-3

Simpson PJ, Jamieson SJ, Abou-Hachem M et al (2002) The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus xylanase*. Biochemistry 41:5712–5719. doi:10.1021/bi012093i

Szabo L, Jamal S, Xie H et al (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. J Biol Chem 276:49061–49065. doi:10.1074/jbc.M109558200

Szeverenyi NM, Sullivan MJ, Maciel GE (1982) Observation of spin exchange by two-dimensional Fourier transform 13C cross polarization-magic-angle spinning. J Magn Reson 47:462–475. doi:10.1016/0022-2364(82)90213-X

Takegoshi K, Nakamura S, Terao T (2001) C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR. Chem Phys Lett 344:631–637. doi:10.1016/s0009-2614(01)00791-6

Tugarinov V, Kay LE (2005) Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. ChemBioChem 6:1567–1577. doi:10.1002/cbic.200500110

Ulrich EL, Akutsu H, Doreleijers JF et al (2008) Biological magnetic resonance data bank. Nucl Acids Res 36:D402–D408. doi:10.1093/nar/gkm957

Verardi R, Shi L, Traaseth NJ et al (2011) Structural topology of phospholamban pentamer in lipid bilayers by a hybrid solution and solid-state NMR method. Proc Natl Acad Sci 108:9101–9106. doi:10.1073/pnas.1016535108

Wang Y, Jardetzky O (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci 11(4):852–861. doi:10.1110/ps.3180102.Structure

Ward ME, Shi L, Lake E et al (2011) Proton-detected solid-state NMR reveals intramembrane polar networks in a seven-helical transmembrane protein proteorhodopsin. J Am Chem Soc 133:17434–17443. doi:10.1021/ja207137h

Weingarth M, Baldus M (2013) Solid-state NMR-based approaches for supramolecular structure elucidation. Acc Chem Res 46:2037–2046. doi:10.1021/ar300316e

Yan S, Hou G, Schwieters CD et al (2013) Three-dimensional structure of CAP-gly domain of mammalian dynactin determined by magic angle spinning NMR spectroscopy: conformational plasticity and interactions with end-binding protein EB1. J Mol Biol 425:4249–4266. doi:10.1016/j.jmb.2013.04.027

Yaniv O, Shimon LJW, Bayer EA et al (2011) Scaffoldin-borne family 3b carbohydrate-binding module from the cellulosome of Bacteroides cellulosolvens: structural diversity and significance of calcium for carbohydrate binding. Acta Crystallogr D Biol Crystallogr 67:506–515. doi:10.1107/S0907444911011322

Yaniv O, Petkun S, Shimon LJW et al (2012) A single mutation reforms the binding activity of an adhesion-deficient family 3 carbohydrate-binding module. Acta Crystallogr D Biol Crystallogr 68:819–828. doi:10.1107/S0907444912013133

Zech SG, Wand AJ, McDermott AE (2005) Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. J Am Chem Soc 127:8618–8626. doi:10.1021/ja0503128

Zhou DH, Shea JJ, Nieuwkoop AJ et al (2007) Solid-state protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy. Angew Chem Int Ed Engl 46:8380–8383. doi:10.1002/anie.200702905